The Journal of Experimental Biology 209, 3383-3404
Published by The Company of Biologists 2006
doi:10.1242/jeb.02328

Comparative sequence analysis and tissue localization of members of the SLC6 family of transporters in adult *Drosophila melanogaster*

Matthew S. Thimgan[1,*], Jonathan S. Berg[2] and Ann E. Stuart[1]

[1]*Department of Cell and Molecular Physiology, University of North Carolina School of Medicine, Chapel Hill, NC 27599, USA and* [2]*Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza Houston, TX 77030, USA*

*Author for correspondence at present address: Department of Anatomy and Neurobiology, Washington University in St Louis Medical School, Box 8108 660, S. Euclid Avenue, St Louis, MO 63110, USA (e-mail: thimgan@pcg.wustl.edu)

Accepted 15 May 2006

Summary

The SLC6 family comprises proteins that move extracellular neurotransmitters, amino acids and osmolytes across the plasma membrane into the cytosol. In mammals, deletion of SLC6 family members has dramatic physiologic consequences, but in the model organism *Drosophila melanogaster*, little is known about this family of proteins. Therefore, in this study we carried out an initial analysis of 21 known or putative SLC6 family members from the *Drosophila* genome. Protein sequences from these genes segregated into either well-defined subfamilies, including the novel insect amino acid transporter subfamily, or into a group of weakly related sequences not affiliated with a recognized subfamily. Reverse transcription-polymerase chain reaction analysis and *in situ* hybridization showed that seven of these genes are expressed in the CNS. *In situ* hybridization revealed that two previously cloned SLC6 members, the serotonin and dopamine transporters, were localized to presumptive presynaptic neurons that previously immunolabelled for these transmitters. RNA for *CG1732* (the putative GABA transporter) and *CG15088* (a member of the novel insect amino acid transporter family) was localized in cells likely to be subtypes of glia, while RNA for *CG5226*, *CG10804* (both members of the orphan neurotransmitter transporter subfamily) and *CG5549* (a putative glycine transporter) were expressed broadly throughout the cellular cortex of the CNS. Eight of the 21 sequences were localized outside the CNS in the alimentary canal, Malpighian tubules and reproductive organs. Localization for six sequences was not found or not attempted in the adult fly. We used the *Drosophila* ortholog of the mammalian vesicular monoamine transporter 2, *CG33528*, to independently identify monoaminergic neurons in the adult fly. RNA for *CG33528* was detected in a limited number of cells in the central brain and in a beaded stripe at the base of the photoreceptors in the position of glia, but not in the photoreceptors themselves. The SLC6 localization observations in conjunction with likely substrates based on phylogenetic inferences are a first step in defining the role of Na/Cl-dependent transporters in *Drosophila* physiology.

Supplementary material available online at
http://jeb.biologists.org/cgi/content/full/209/17/3383/DC1

Key words: Na/Cl-dependent transporters, *in situ* hybridization.

Introduction

The SLC6 family of transporters, also referred to as the neurotransmitter:sodium symporter family (NSS) or sodium neurotransmitter transporter family (SNF), is composed of integral membrane transporter proteins. These transporters are critical for maintaining physiological homeostasis in animals by transferring compounds such as neurotransmitters, amino acids or osmolytes across the plasma membrane. SLC6 members transport compounds against their concentration gradient by coupling substrate uptake to the energy built up in the transmembrane Na^+ gradient. Some of these transporters also depend on the presence of extracellular Cl^- (reviewed in Amara and Arriza, 1993; Nelson, 1998). Substrates for members of the SLC6 family include: neurotransmitters such as GABA, serotonin, norepinephrine and dopamine; the amino acids glycine and proline, which in mammals are also used as neurotransmitters; and the osmolytes taurine and betaine. Recently, the broad substrate profiles B^{O+}, Bo and IMINO systems have been attributed to SLC6 transporters (Sloan and Mager, 1999; Broer et al., 2004; Takanaga et al., 2005). The SLC6 family also includes a group of transporters for which no substrates have been identified, known as the 'orphan'

neurotransmitter transporters (Amara and Arriza, 1993; Nelson, 1998).

The SLC6 gene family is defined by conserved structural features, including a predicted twelve-transmembrane domain topology, highly conserved amino acid residues, particularly in transmembrane (TM) domains 1, 2 and 4–8, and a large extracellular loop between transmembrane domains 3 and 4 that is predicted to be glycosylated (Amara and Arriza, 1993; Nelson, 1998). Crystal structural data have shown that many of the conserved residues in TM1, TM6 and TM8 are involved in substrate and sodium binding in a bacterial leucine transporter (Yamashita et al., 2005). Mice null for individual SLC6 transporters show abnormal physiological phenotypes. For instance, disruption of the dopamine transporter (DAT) gene leads to persistence of dopamine in the extracellular space and a hyperlocomotive phenotype equivalent to the effects of cocaine and amphetamines, which are known to inhibit the DAT (Giros et al., 1996). When other SLC6 transporters lack function, it can result in death or disease (Gomeza et al., 2003; Heller-Stilb et al., 2002; Quan et al., 2004; Tsai et al., 2004).

Despite their importance in mammalian physiology, there is little information about these proteins in *Drosophila*. Thus far, only four genes from the SLC6 family have been cloned from *Drosophila*: the genes responsible for the selective uptake of serotonin (*SerT*) (Corey et al., 1994; Demchyshyn et al., 1994) and dopamine (*DAT*) (Porzgen et al., 2001) and two orphan transporters *inebriated* (*ine*) (Burg et al., 1996; Soehnge et al., 1996) and *bloated tubules* (*blot*) (Johnson et al., 1999). The cellular pattern of expression of *SerT* RNA in the embryo is similar to that of serotonin immunolabelling (Demchyshyn et al., 1994); likewise *DAT* is expressed in a cellular pattern in the larva similar to that of dopamine immunolabelling (Porzgen et al., 2001). The other two sequences, Ine and Blot, do not yet have identified substrates but both are expressed in a variety of cells inside and outside the central nervous system (CNS) (Burg et al., 1996; Huang et al., 2002; Johnson et al., 1999; Soehnge et al., 1996).

To begin to address the role of SLC6 transporters in *Drosophila*, we used a bioinformatics approach to identify 21 *Drosophila* genes with similarity to known SLC6 transporters. Phylogenetic analysis revealed that these *Drosophila* SLC6 transporters segregated into four of the five previously recognized SLC6 subfamilies (Nelson, 1998) and helped to define a sixth subfamily, the insect amino acid transporters (IAAT) (Soragna et al., 2004; Boudko et al., 2005). Using *in situ* hybridization, we focused on localizing SLC6 transporters in the CNS of adult flies. Probes that did not label CNS cells were tested on a variety of other fly tissues as positive confirmation of the reagents. In the course of this work we also localized a vesicular monoamine transporter and, curiously, found that it labelled glial cells at the distal margin of the lamina cell body layer in addition to the expected localization in neurons in the brain in the monoaminergic pattern.

Materials and methods

Fly stocks

Drosophila yw^{67} or w^{1118} mutants, both of which lack eye pigment, were used in all *in situ* hybridizations.

Bioinformatic analysis, multiple sequence alignment and phylogenetic analysis

To identify candidate *Drosophila* genes that could encode Na^+/Cl^--dependent transporters, we initially performed BLASTP and TBLASTN (Altschul et al., 1997) searches of predicted proteins from the annotated *Drosophila* genome sequence (Adams et al., 2000). Well known Na^+/Cl^--dependent monoamine transporters from several organisms, including SerT, DAT and the norepinephrine transporter, were used in sequence searches. A comprehensive group of known or predicted amino acid transporters from vertebrates, insects and worm were assembled, and the full peptide sequences were evaluated by multiple sequence alignment using CLUSTALX (Thompson et al., 1997). The alignment of 84 SLC6 family transporters was used to create neighbor-joining phylogenetic trees in order to depict relationships between various members of the SLC6 family across multiple species. One thousand bootstrap trials were performed to evaluate the significance of the branch node patterns. We used a bootstrap value of >75% to define subfamilies. A separate alignment of the 21 putative SLC6 from *Drosophila* and the leucine transporter from *Aquifex aeolicus* [for which the crystal structure was recently published (Yamashita et al., 2005)] was generated to highlight conserved regions in the transmembrane domains. The final alignments depicted herein were manually adjusted and shaded using GeneDoc software (Karl B. Nicholas and Hugh B. Nicholas, 1997).

Riboprobe generation

Total RNA was isolated from *Drosophila* heads using the TRIzol reagent (Gibco, Carlsbad, CA, USA) and converted to cDNA using a poly-T primer and Superscript II reverse transcriptase (Gibco, Carlsbad, CA, USA). Portions of the cDNA for each candidate gene were amplified by polymerase chain reaction (PCR) and cloned into pGEM-T Easy (Promega, Madison, WI, USA). All clones were verified by sequencing (UNC sequencing facility). Digoxigenin (DIG)-labelled sense and anti-sense riboprobes were generated from linearized plasmids according to manufacturer's specifications using the DIG RNA labelling kit (Roche, Indianapolis, IN, USA). RNA probes averaging 1000 bp and ranging from 800–1200 bp were purified using three precipitations in 3× ethanol and 0.3 mol l^{-1} LiCl incubated at –80 for 2 h between each precipitation and resuspended in DEPC-treated water or hybridization buffer.

Tissue preparation

Drosophila were anesthetized with CO_2. The back third of the abdomen was cut off and the proboscis was removed or, in the case of abdominal sections, the front half of the fly was cut off to allow complete penetration of fixative. The flies were submerged in ice-cold 4% paraformaldehyde (PFA) in

phosphate-buffered saline (PBS), transferred to 10%, then to 20% sucrose in PBS, and incubated at 4°C overnight in each solution. Flies were then suspended in Optimal cutting temperature (Tissue-Tek, Torrance, CA, USA) reagent, frozen in isopentane cooled by liquid nitrogen, and sectioned on a cryostat microtome at 10–15 μm. Sections were collected on room temperature Superfrost Plus slides and stored at 4°C until *in situ* hybridization or antibody labelling could be performed.

In situ hybridization

The DIG method (Roche, Indianapolis, IN, USA) of *in situ* hybridization was used following the procedure of Nowicki and Burke with minor modifications (Nowicki and Burke, 2000). Briefly, slides containing sections of *Drosophila* heads were washed and post-fixed in 4% PFA in PBS for 30 min then rinsed twice for 5 min in PBS. After two 2 min washes in 2× sodium chloride/sodium citrate pH 4.5 (SSC), sections were incubated for 30 min in Tris-glycine buffer. Riboprobes, suspended in hybridization buffer at 25–100 ng μl^{-1}, were applied to the slides and incubated overnight at 65°C in a humidified chamber.

Sections were rinsed 3 times for 20 min each in 5× SSC at room temperature. Sections were transferred to a preheated solution of 20% formamide and 0.5× SSC (sol B) and incubated at 60°C for 40 min. Sol B was replaced with preheated sol B and the temperature of the solution was allowed to cool to 37°C. Sol B was again replaced with pre-heated sol B and the slides were incubated at 60°C for 30 min. Slides were transferred to 2× SSC at room temperature for 30 min. They were then incubated in 2% Boerhenger Blocking Reagent in a maleic acid buffer (block) for a minimum of 10 min. Sections were then incubated overnight in a humidified chamber at 4°C in anti-DIG antibody diluted 1:5000 in block.

Antibody was rinsed away with four washes of 10 min and one wash of 20 min in Tris-buffered saline (TBS) at RT. Sections were then incubated for 10 min in 100 mmol l^{-1} Tris with 500 μg ml^{-1} levamisole and 0.1% Triton X-100. Sections were placed in 0.131 mg ml^{-1} 5-bromo-4-chloro-3′-indolyphosphate p-toluidine (BCIP) and 0.5 mg ml^{-1} nitro-blue tetrazolium (NBT) in 10% polyvinyl alcohol for 1 h to 3 days. Sections were then rinsed in PBS and mounted with Glycergel (DAKO, Carpenteria, CA, USA). Sequential sections were probed by alternating between sense and anti-sense probes. Sense probes showed no specific label in the head but revealed non-specific labelling of the exterior margin of the eye.

Reverse transcription-polymerase chain reaction

Total RNA was isolated using the Qiagen RNeasy kit (Qiagen, Valencia, CA, USA) using the manufacturer's specifications. w^{1118} flies were dissected into head, thorax and abdomen and immediately frozen on dry ice. Tissue was manually homogenized and further lysed using the QiaShredder spin column. Total RNA was treated with DNase using the DNA-free kit (Ambion, Austin, TX, USA). 2.5 μg of total RNA was used for reverse transcription reaction. PCRs were run for 30 cycles at optimal temperature for the respective primer pairs. Exact primer sequences can be provided upon request.

Antibody staining

Flies were fixed as above and sectioned between 10–15 μm. Sections were permeabilized in PBS/0.5% Triton for 10 min and then washed in PBS. Endogenous fluorescence was quenched in 0.5% $NaBH_4$ in PBS for 10 min and the slides were washed in PBS. Sections were blocked in 5% normal goat serum (NGS) for 1 h, incubated overnight at 4°C in primary antibody in 5% NGS, rinsed in PBS, and incubated in an anti-goat secondary antibody Alexa 488 (Molecular probes, Carlesbad, CA, USA) at a dilution of 1:2000 in 5% NGS for 1 h. and rinsed again in PBS and mounted in Gel/Mount (Biomeda, Forest City, CA, USA). Repo antibody (Developmental studies hybridoma bank, Ames, IA, USA) was used at 1:1 and Neurexin IV (NrxIV, generous gift of M. Bhat)

Fig. 1. Twenty-one sequences in the *Drosophila* genome are homologous to SLC6 transporters. (A) Schematic illustration of the structure SLC6 family transporters (based on the crystal structure of Yamashita et al., 2005). Transmembrane domains are represented by grey rectangles and are numbered according to Yamashita et al., while intracellular and extracellular loops are represented by a thick dark line. Many metazoan SLC6 transporters have large N-terminal and C-terminal extensions, represented by a broken thick line, and most metazoan SLC6 transporters have a 10–50 amino acid extracellular loop between TM3 and TM4, indicated by a thick broken line. Conserved intracellular domains are indicated by white rectangles, and conserved extracellular domains are indicated by black rectangles. The arrows indicate the position of the beta sheet conformation predicted by Yamashita et al. Members of the orphan neurotransmitter subfamily have divergent extracellular loops between TM7 and TM8, and between TM11 and TM12 (dotted lines). (B) Multiple sequence alignment generated in ClustalX using 21 putative *Drosophila* SLC6 transporters and the leucine transporter from *Aquifex aeolicus*, manually adjusted to maximize comparison to the *Aquifex aeolicus* transporter. Protein domains are annotated at the top of each sequence block, indicating transmembrane (TM) regions, intracellular linker regions (IL), extracellular linker regions (EL) alpha helical structure (α), and beta-sheets (arrows); designation of the protein domains is based on the published alignment of Yamashita et al., with minor changes. The names of each protein sequence are on the left (see Table S1 in supplementary material for accession numbers), asterisks are used to mark 10 amino acid intervals, and the residue numbers are indicated at far right. Amino acid residues are shaded according to the degree of conservation (black=100%, dark grey=80%, light grey=60%; amino acids with similar chemical properties are considered equivalent for the purposes of determining conserved residues). Gaps are represented by dashes and residues removed from the alignment for space reasons are indicated by numbers in parentheses. Selected residues that were considered important for transporter function by Yamashita et al. are indicated by symbols at the bottom of each sequence block (○, charged residues at extracellular and cytoplasmic entrances; †,‡, residues important for coordinating sodium ions 1 and 2, respectively) and residues considered to be strictly conserved by Yamashita et al. are indicated as 'Invariant' on the bottom line of each sequence block to facilitate comparison between alignments.

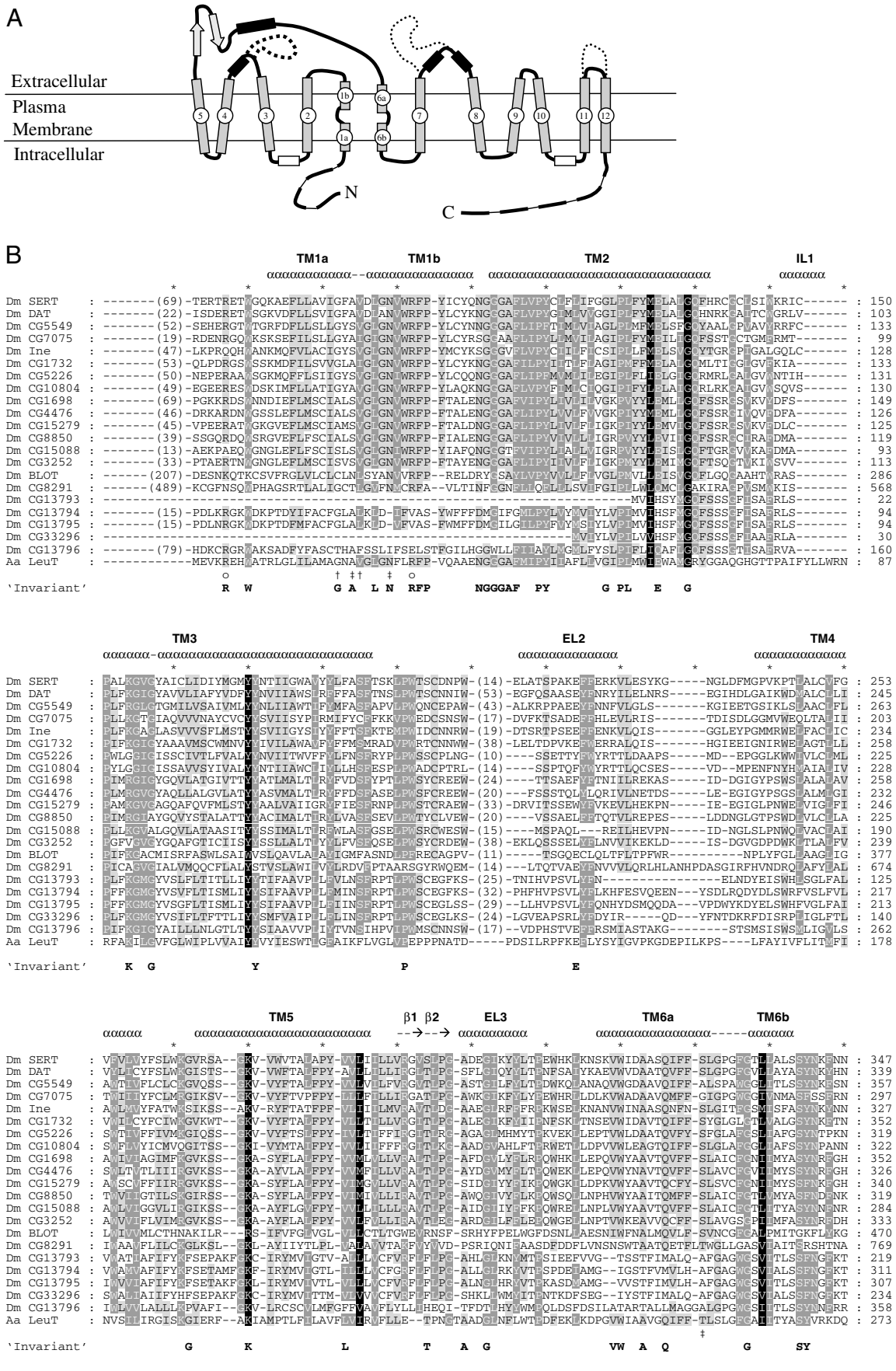


Fig. 1. For legend see previous page.

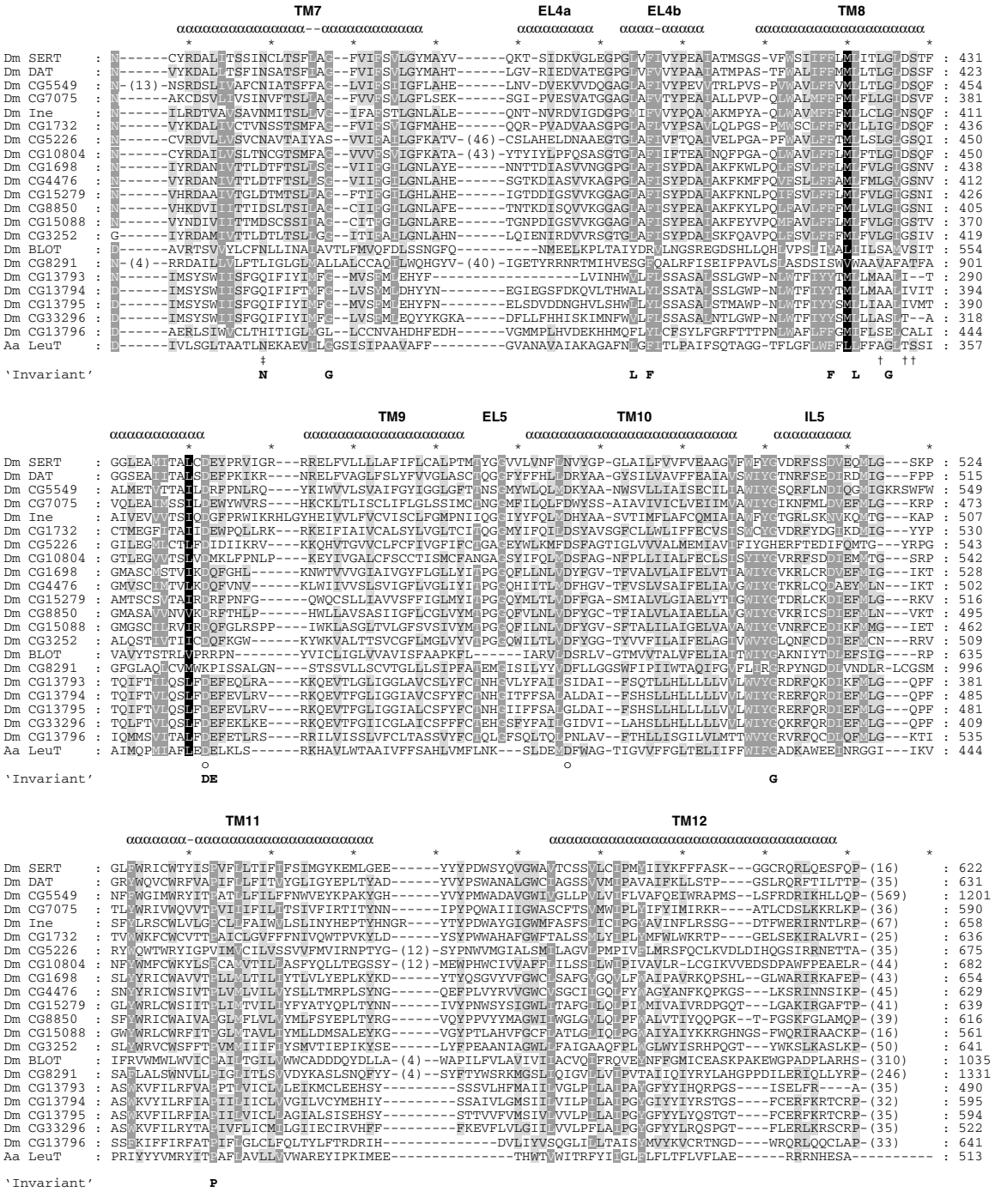


Fig. 1. Continued.

was used at a dilution of 1:500. Sections labelled with fluorescent markers were visualized on a Nikon Eclipse E800 microscope with a 20× Plan Apo objective with a numerical aperture (NA) of 0.75 or a 40× Plan Apo objective with an NA of 0.95 and imaged with a Hammamatsu ORCA-ER camera (Bridgewater, NJ, USA) and visualized with the software package Metamorph (Universal Imaging, Downingtown, PA, USA). Images were optimized for visualization and publication using Adobe Photoshop (Adobe, San Jose, CA, USA).

Results

The Drosophila genome contains 21 genes with homology to SLC6 transporters

We attempted to identify all SLC6 genes in the *Drosophila* genome using a bioinformatics approach with amino acid sequences for known SLC6 transporters as bait with which to search GenBank using the NCBI BLASTP and TBLASTN algorithms (Altschul et al., 1997). We identified a total of 21 genes that we consider to be putative members of the SLC6

family in *Drosophila* (listed in Fig. 1B and bolded and underlined in Fig. 2). All candidate sequences are predicted to have between ten and twelve TM domains (data not shown) and at least 20% identity to SerT or DAT. A multiple sequence alignment including each of the putative SLC6 family members in *Drosophila* with the amino acid sequence of a SNF6 family member from *Aquifex aeolicus*, $LeuT_{Aa,}$ for which the crystal structure was recently published (Yamashita et al., 2005), reveals that each of the candidate sequences contains multiple residues that are absolutely or highly conserved among all of the *Drosophila* sequences (Fig. 1B). A complete alignment of all transporter sequences used in our phylogenetic analysis is available (supplementary material, Fig. S1).

The highly conserved region in the vicinity of TM domains 1 and 2 (Lill and Nelson, 1998) was present in the majority of the candidate genes. For clarity in describing proteins occurring in multiple species, we will use a species prefix such as Dm for *Drosophila melanogaster*. Five of the candidate proteins (DmCG13793, DmCG13794, DmCG13795, DmCG13796 and DmCG33296) are more divergent at the N terminus, with DmCG13793 and DmGG33296 lacking TM1–TM2. The other three divergent candidate proteins have limited homology in the region of TM1 but have a significant degree of similarity

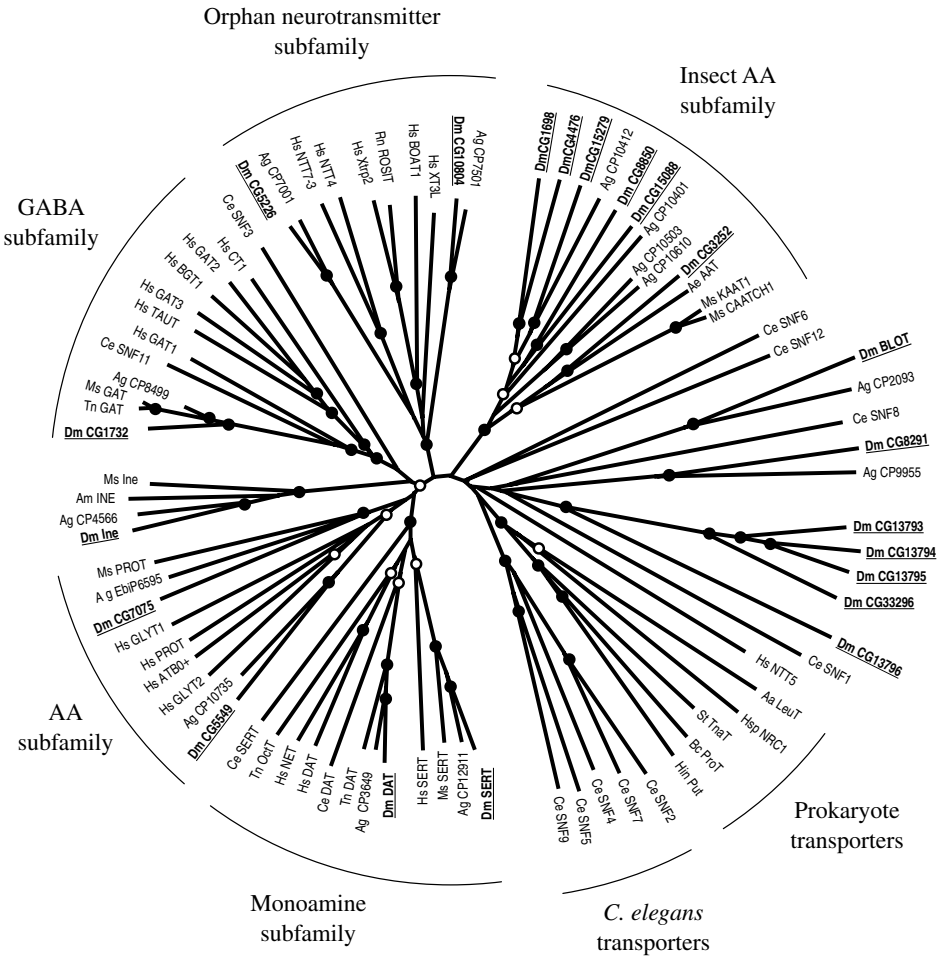


Fig. 2. Phylogenetic analysis assigns many *Drosophila* SLC6 proteins to previously identified subfamilies and a novel IAAT subfamily. An unrooted phylogenetic tree displaying *Drosophila* candidate amino acid (AA) sequences with cloned and predicted SLC6 transporter amino acid sequences from multiple organisms. Candidate *Drosophila* SLC6 transporters are labelled in bold and underlined. Prominently represented among the insect sequences are putative SLC6s from *Anopheles gambiae* (Ag), a mosquito known for carrying the malaria parasite and whose genome sequence was recently reported (Holt et al., 2002). We also have included known transporter sequences from insects such as the cabbage looper, *Trichoplusia ni* (Tn), the tobacco caterpillar *Manduca sexta* (Ms) and from mammals. See Table S1 in supplementary material for a list of the sequence names, abbreviations and accession numbers. Subfamilies (monoamine transporters, GABA transporters, orphan neurotransmitter transporters, insect amino acid transporters) are indicated. As expected, a number of sequences do not group with the previously identified subfamilies. Many of the *C. elegans* transporters are quite divergent and group with each other; likewise, the prokaryote transporters are most similar to each other but do not segregate into a recognized subfamily with eukaryote transporters. Several of the candidate *Drosophila* transporters form a small subgroup that does not convincingly group with any other transporters. Nodes that are identical in >90% of bootstrap trials are denoted by filled circles and nodes that are identical in 75–90% of bootstrap trials are denoted by empty circles. The complete multiple sequence alignment from which this phylogenetic tree was generated is available as Fig. S1 in supplementary material.

to SLC6 members throughout the remainder of their sequences (this is also the case for several of the more divergent *C. elegans* SNF genes). Based on the crystal structure of $LeuT_{Aa}$, the first transmembrane domain is likely to play a pivotal role in substrate binding (Yamashita et al., 2005). Therefore, these highly divergent sequences may lack functional transporter activity or have alternative roles. One possible explanation for the differences in the N-terminal region of the proteins is that most of the candidate genes are predicted *in silico* from genomic sequence and therefore are subject to the limitations of exon scanning algorithms. However, cDNA sequences for three of these candidate genes (DmCG13794, DmCG13795 and DmCG13796) have been deposited in the GenBank database and we have therefore included them in this analysis as tentative SLC6 family members.

Structural features identified in the candidate proteins may provide clues as to their function. For example, three candidates (Blot, Ine and DmCG8291) are predicted to have long intracellular N-terminal domains and three candidates (Blot, DmCG5549 and DmCG8291) are predicted to have large intracellular C-terminal domains. DmCG5549 has a short extension of the intracellular linker sequence between TM 6 and TM 7, while the orphan neurotransmitter subfamily members (including DmCG5226 and DmCG10804) are predicted to have extended extracellular linker 4a sequences between TM 7 and TM 8 as well as a short extension of the extracellular linker sequence between TM 11 and TM 12 (see Fig. 1A). Long intracellular terminal domains or loops may indicate intracellular regulation by protein-protein interactions or phosphorylation (Melikian, 2004) while extracellular loops are likely to be involved in substrate binding, the transport process and inhibitor binding (Zomot and Kanner, 2003).

Drosophila SLC6 transporters segregate into four previously recognized subfamilies and help define the novel insect amino acid transporter (IAAT) subfamily

In order to infer functional information about the putative *Drosophila* SLC6 family members, we compared them to a large number of known SLC6 sequences using phylogenetic analysis (Fig. 2). We aligned the 21 *Drosophila* sequences with an assembled collection of 63 putative neurotransmitter or amino acid transporters from other organisms: 25 sequences from other insects, 19 from vertebrates, 13 from *C. elegans* and 5 from prokaryotes. Fig. 2 displays a phylogenetic tree derived from the complete alignment of 84 SLC6 family members. Accession numbers for the sequences used are provided (supplementary material, Table S1) and the complete alignment is available (supplementary material, Fig. S1).

Some of the previously uncharacterized *Drosophila* sequences segregate into previously recognized subfamilies anchored by well-characterized transporters (Nelson, 1998). For example, DmCG1732 and its *Anopheles* homolog AgCP8499 are likely to function as GABA transporters as they are closely related to human GAT-1, *C. elegans* SNF11 [a recently identified GABA transporter (Mullen et al., 2006)], and the GABA transporters from *Trichoplusia ni* (Gao et al., 1999) and *Manduca sexta* (Mbungu et al., 1995). DmCG5549 and its *Anopheles* homolog AgCP10735 are related to the amino acid transporter subfamily, which includes the amino acid transporters for glycine and proline and a transporter with a broader substrate capacity, B^{O+} (Sloan and Mager, 1999). DmCG7075 and its *Anopheles* homolog, AgEbiG6595, are closely related to a proline transporter from *Manduca* (Sandhu et al., 2002), and these transporters segregate loosely with the glycine/proline subgroup and the monoamine transporters. Surprisingly, only the *Drosophila* DAT and SerT are present in the monoamine transporter subfamily. Since octopamine and histamine are also biogenic amines and the *Trichplusia ni* octopamine transporter is present in the monoamine subfamily, we expected that *Drosophila* candidate genes with these transport activities would be found in this subfamily. However, it appears that there are no close monoamine homologs in this subfamily.

Two candidate genes, DmCG5226 and DmCG10804 (and their *Anopheles* homologs AgCP7001 and AgCP7501, respectively) group with the previously defined subfamily of orphan neurotransmitter transporters that includes NTT4 and NTT7-3 as well as two other newly predicted mammalian orphan transporters (Nash et al., 1998). Interestingly, this family also includes a neutral amino acid transporter ($B^{O}AT$), mutations of which cause Hartnup disease (Broer et al., 2004; Kleta et al., 2004; Seow et al., 2004).

Our phylogenetic analysis identified a number of predicted genes that help define the novel insect amino acid transporter (IAAT) subfamily (Boudko et al., 2005), which to date is represented only in insects. Predicted sequences from *Drosophila melanogaster* (DmCG1698, DmCG3252, DmCG4476, DmCG8850, DmCG15088 and DmCG15279) and *Anopheles gambie* (AgCP10401, AgCP10412, AgCP10503 and AgCP10610) are clearly related to the *Manduca sexta* transporters KAAT1 (Castagna et al., 1998) and CAATCH1 (Feldman et al., 2000), which are competent to transport substrate amino acids using either K^+ or Na^+ as the driving ion. Sequence similarities among KAAT1, CAATCH1 and the other members of the IAAT subfamily suggest that many of the members of this subfamily possess this unique functional property.

A large number of predicted *Drosophila* and *Anopheles* SLC6 transporters segregate into poorly defined subgroups. For example, DmIne and AgCP4566 do not segregate nicely into a recognized subfamily. DmIne mediates a response to hypertonic solutions (Chiu et al., 2000) and has been proposed to comprise its own subfamily (Boudko et al., 2005). Other sequences, including a number of orphan transporters from *C. elegans*, vertebrate NTT-5 and putative *Drosophila* transporters (DmBlot, DmCG8291, DmCG13793, DmCG13794, DmCG13795 and DmCG33296) make up a loose assemblage of outliers, which do not appear to be affiliated with defined subfamilies. We have identified *Anopheles* homologs for the *Drosophila* Ine and Blot orphan transporters (AgCP4566 and AgCP2093, respectively) as well

as for the other outlier sequences, suggesting a conserved, if still unknown, function for these transporters.

Candidate gene expression in the head, thorax and abdomen of male and female flies

We used RT-PCR to determine a relative segmental expression profile for the SLC6 homologs in adult flies. mRNA was isolated from the head, thorax and abdomen of either males or females and amplified by RT-PCR to determine the presence or absence of transcript in each segment (Fig. 3). As a control for consistent starting cDNA template across all segments we used primers for the gene *rp49*, and as a control for a gene with enriched expression in the head and thorax we used primers for the gene for the histamine synthesizing enzyme, *histidine decarboxylase* (*HDC*) (Burg et al., 1993; Pollack and Hofbauer, 1991).

Six candidates were not expressed evenly across all body segments (Fig. 3). Four candidates, *SerT*, *DAT*, *CG1732* and *CG10804*, showed enriched expression in the head and thorax of both males and females. The cell bodies of the CNS are found in the head and thorax, and cells that immunolabel for serotonin (Valles and White, 1988), dopamine (Budnik and White, 1988; Nassel and Elekes, 1992) and GABA (Buchner et al., 1988) reside within the CNS. Expression of the orphan transporter, *CG10804*, was only found in the head and thorax, suggesting that *CG10804* functions in the CNS. Two of the candidates, *CG7075* and *CG4476*, were expressed more abundantly in the abdomen of the male and female fly, respectively.

The remaining fifteen SLC6 transporters showed approximately equivalent expression across the three body segments in both genders. This result suggests that these transcripts are expressed either in organs in each of the segments or in an organ present in all three segments. Although twelve of these transcripts displayed robust expression, three transcripts (*CG8850*, *CG13793* and *CG33296*) revealed only weak expression.

Genes involved in monoamine processing localize to a subset of cells in the adult CNS

Previous immunocytochemistry has revealed the location of serotonergic and dopaminergic cells in the adult fly CNS; these neurons are grouped in named clusters (Budnik and White, 1988; Valles and White, 1988). The riboprobes against *SerT* (Fig. 4A,B) and *DAT* (Fig. 4C,D) labelled CNS cells of roughly the same number and in the same general location as the immunolabelled cells. Due to limited resolution of the *in situ* hybridization technique and use of alternating sections between sense and anti-sense probes, we were only able to estimate the number of neurons that were labelled by each riboprobe. For the *SerT* probe, about 75 neurons labelled in cell clusters that approximated the position of the named clusters from previous serotonin immunolocalization studies in both the head and

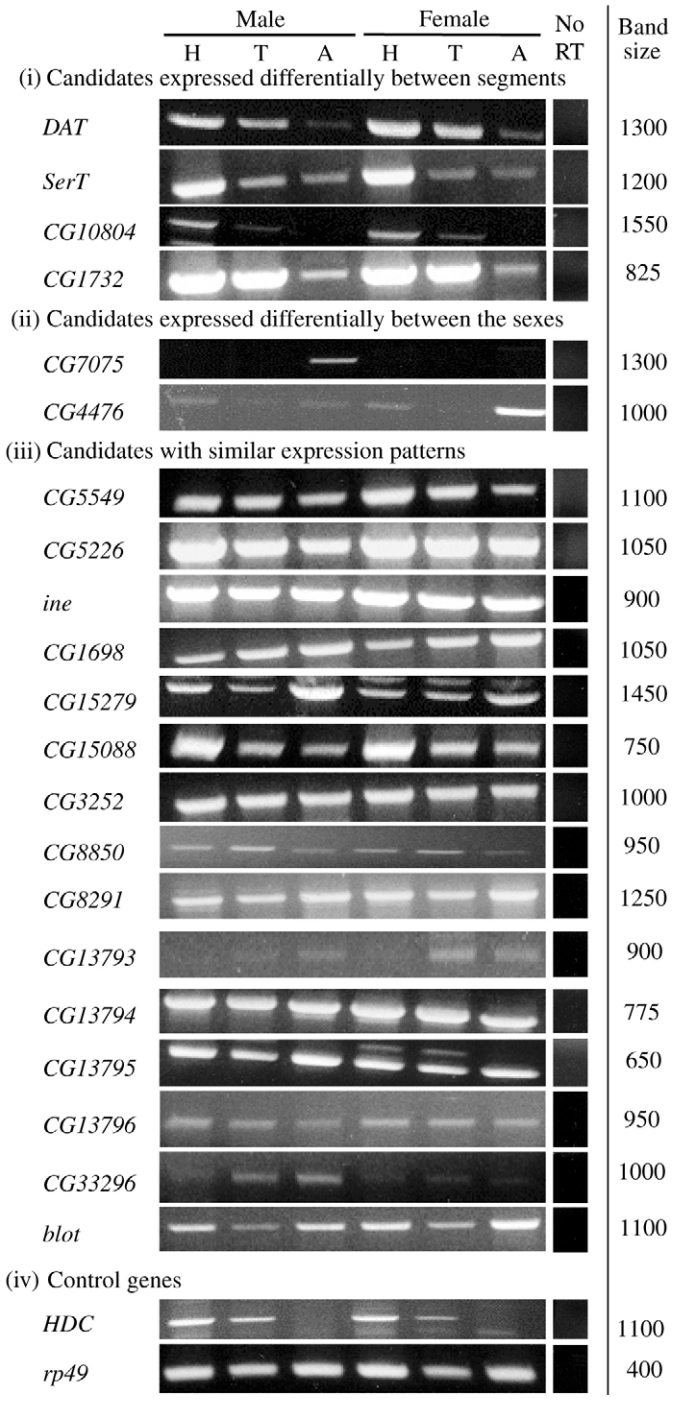


Fig. 3. The segmental mRNA expression profile of the SLC6 homologs in the head, thorax and abdomen of both male and female flies determined by RT-PCR. The head (H), thorax (T) and abdomen (A) were isolated from either male or female flies and RT-PCR was performed to determine presence or absence of transcript in a given body segment. Candidates are arranged from top to bottom as: (i) those expressed differentially in the three segments; (ii) those showing gender specificity; (iii) those expressed equally in the segments; (iv) control genes *HDC* and *rp49*. *HDC* is known to have enriched expression in the head and thorax and *rp49* controlled for equivalent starting template in the PCR reactions. The column labelled 'No RT' showed controls confirming that the amplicons were dependent on the presence of cDNA. Band size is measured in base pairs (bp).

thoracic ganglion. Labelled cells in the head (Fig. 4A) were found dorsally on either side of the midline in the position of the cluster called SP1 (arrows), ventrally in the position of the SE2 cluster (barbed arrow), and proximal to the optic lobe in the position of the LP2 cluster (notched arrow). The photoreceptor cells were never labelled. Cells in the thorax (Fig. 4B) are equivalent in position to Budnick and White's 'A1–7' neurons (arrows) (Budnik and White, 1988). Other sections showed neurons in positions of other named clusters and no cells were labelled in positions that substantially differed from the immunolabel. It is likely, therefore, that our *SerT* riboprobe is labelling serotonergic neurons.

The *DAT* riboprobe likewise labelled discrete cells in the pattern described for dopaminergic neurons using immunocytochemistry (Budnik and White, 1988). We identified up to 30 large cell bodies in a single preparation with alternating slides, approximately the number of large cell bodies labelled by the dopamine antibody. In the section shown in Fig. 3C, the riboprobe labelled cells in the positions of three named clusters: two dorsal clusters on either side of the midline corresponding to the DM cluster of neurons (arrows); neurons lateral to the dorsomedial (DM) cluster in the more dorsal DL1 cluster (white arrowheads); and cells more ventral in the DL2 cluster (carrot). We also detected *DAT* positive cells corresponding to the AbU (arrows) and ThL clusters (barbed arrow) in the thoracic ganglion (Fig. 4D). In other sections, we detected cells in other named clusters. Curiously, the riboprobe did not label a group of cells surrounding the medulla, the MC cell group that clearly labels for catecholamines [fig. 7A in Budnik and White (Budnik and White, 1988)], nor did it label a cluster of small cells along the midline referred to as the anteromedial (AM) neurons. Thus, as with the *SerT* riboprobe, the pattern of *in situ* label overlapped with that from dopamine immunohistochemistry, suggesting that, with the exception of the AM and MC cells, the *DAT* probe labels dopaminergic cells. It does not appear that the *DAT* probe labels octopaminergic in addition to dopaminergic cells, as we do not see *DAT* positive cells in the easily recognizable antennal lobe (AL) cluster of octopaminergic cells localized along the alimentary canal (Monastirioti et al., 1995). In addition, the *DAT* probe does not label octopaminergic cells in the larva and does not have a high affinity for octopamine in uptake experiments in *Xenopus* oocytes (Porzgen et al., 2001).

Although we did not find a *Drosophila* SLC6 homolog for either the plasmalemmal octopamine or histamine transporter, in the course of this study we localized an ortholog of a mammalian vesicular monoamine transporter 2 (VMAT2). In mammals, VMAT2 packages four biogenic amines (dopamine, serotonin, norepinephrine and histamine), and it is present in neurons using each of these four neurotransmitters (Peter et al., 1995). A *Drosophila* ortholog of the mammalian VMAT2 would be expected to label all monoaminergic cells, revealing cells that contain dopamine, serotonin, octopamine and histamine. The closest homologous sequence to *VMAT2* in *Drosophila* is *CG33528*, which is 56% identical (Greer et al., 2005).

A riboprobe designed against *CG33528* labelled on the order of 125 cells in the central brain (Fig. 4E,G) and thorax (Fig. 4F), which are in the anatomical position of named clusters of monoaminergic neurons. The riboprobes for *SerT* and *DAT* each labelled cells in the same location as the *CG33528* label (compare Fig. 4A,C with 4E), but there were consistently more cells labelled by the *CG33528* probe than for the *SerT* and *DAT* probes combined. This pattern may represent labelling of octopaminergic neurons since antibody labelling revealed that *CG33528* was found in serotonergic, dopaminergic and octopaminergic (Greer et al., 2005) neurons, but was not found in the histaminergic photoreceptors (Chang et al., 2006). We were able to detect *CG33528* label in both presumptive MC cells (open arrows in Fig. 4E), dopaminergic neurons present in the medullary cortex, and in the presumptive AL cells, octopaminergic cells that reside just lateral to the oesophagus (data not shown).

Curiously, we were unable to detect label in photoreceptors, which in the fly are histaminergic (Pollack and Hofbauer, 1991). To confirm that our *in situ* approach would work in the photoreceptors, we made a riboprobe to *HDC*, which is expressed in photoreceptors (Burg et al., 1993). The *HDC* riboprobe labelled the photoreceptor layer of adult head sections and in addition, about 20 discrete, bilateral cells in the central brain (Fig. 4J) as well as cells in the thoracic ganglion (data not shown) all in the anatomical positions consistent with immunolabelled histaminergic neurons (Pollack and Hofbauer, 1991).

Unexpectedly, the riboprobe for *CG33528* also labelled cells in a punctuated arc at the fenestrated layer at the base of the photoreceptors (Fig. 4G–I), a distinct region where the photoreceptor cell bodies form the axons that ultimately synapse in the optic lobes. This band of beaded label is less than 5 μm in width, the centers of the beads spaced 8–10 μm apart, and the label is intermixed with the fenestrations at the base of the photoreceptors that are present at the distal-most edge of the lamina cortex. Fingers of label occasionally penetrate into the photoreceptor layer (block arrows in Fig. 4H).

The cell types of this layer have been described for *Drosophila melanogaster* (Eule et al., 1995) and ultrastructurally in *Musca domestica* (Saint Marie and Carlson, 1983). There are two glial subtypes that reside at this distal margin of the lamina cortex, the fenestrated and pseudocartridge glia. In both species, these cells have their somata spaced approximately 10 μm apart; in *Musca*, both glial subtypes have been shown to wrap photoreceptor axons traversing this layer on their way to forming synapses in the optic lobes (Saint Marie and Carlson, 1983). The fenestrated glia, but not the pseudocartridge glia, have processes that invade the photoreceptor layers, similar to the *CG33528* labelling. When alternating sections were labelled with *CG33528* riboprobe (Fig. 4H) and an antibody to Neurexin IV (NrxIV; Fig. 4I), a component of septate junctions found on the plasma membrane of glia and not neurons (Baumgartner et al., 1996), the labelling pattern was similar. Like *CG33528*, the

NrxIV antibody labelled the base of the photoreceptors in a beaded fashion spaced approximately 10 μm apart (Fig. 3I). This labelling pattern suggests that *CG33528* is present in the distal-most glia of the lamina cell body layer, in particular the spacing and the fingers penetrating into the retina suggest that that these are the fenestrated glia.

Two SLC6 homologs label the CNS in a glial pattern

CG1732 (the putative GABA transporter) and *CG15088* (a member of the IAAT subfamily) are expressed in the CNS in a glial-like pattern (Fig. 5). Glia of the fly CNS form the barrier between hemolymph and the brain, wrap axons and isolate synaptic terminals in areas such as the lamina (Saint Marie and Carlson, 1983), among other roles. We have chosen to use the *Drosophila* glial terminology (Eule et al., 1995). In short, a layer of glia, called perineural glia (Fig. 5B, arrows), forms the outermost layer in the adult brain. Interior to the perineural glia, the somata of the subperineural glia (Fig. 4B, carrots) are

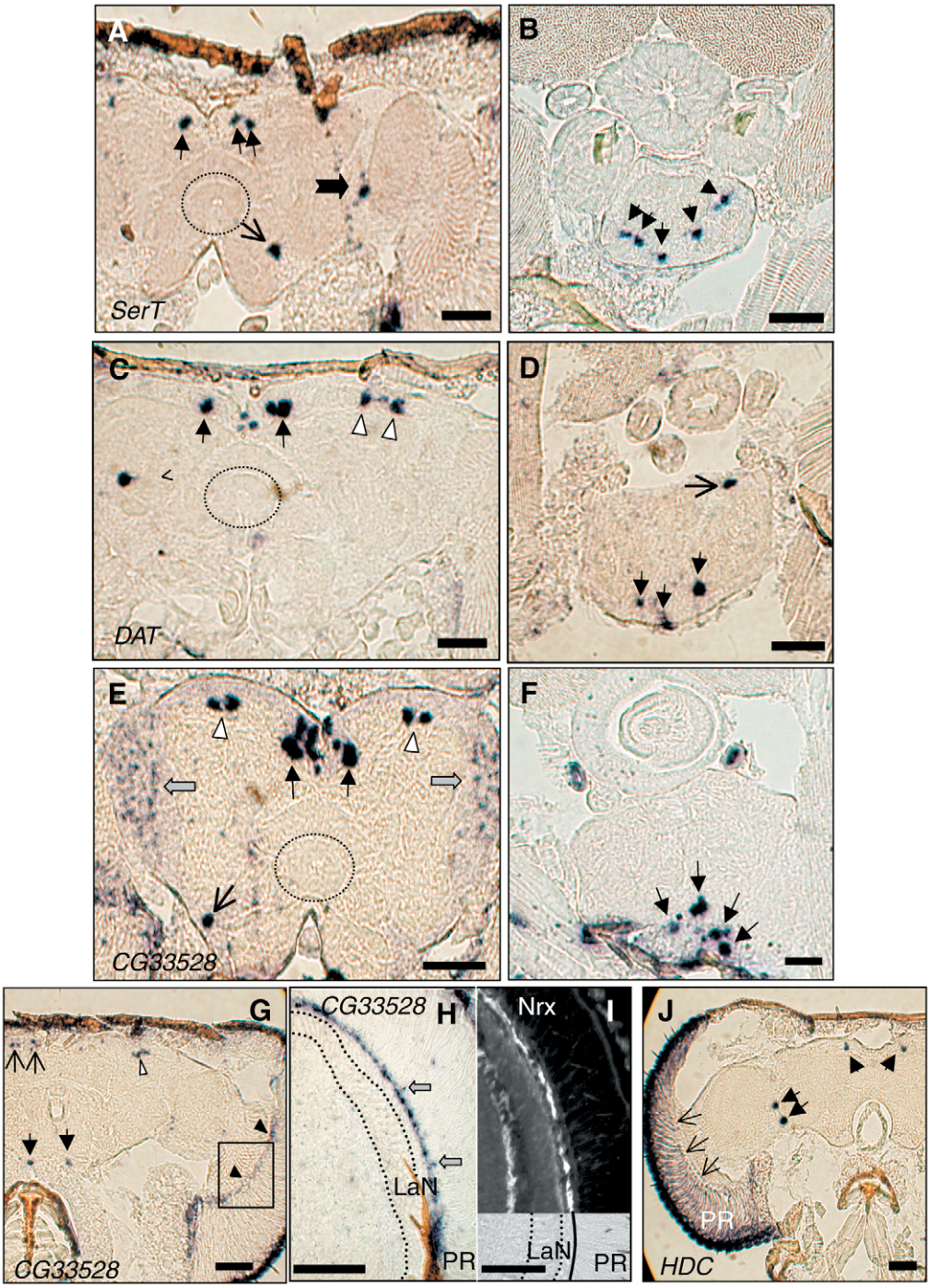


Fig. 4. For legend see next page.

interspersed with neuronal somata to form the cortex. Neuronal processes form the neuropil and neuropilar glia wrap the neuronal processes within the neuropil. As is true in general of invertebrate nervous systems, there are no neuronal somata within the neuropil.

The riboprobe *CG1732* labels a subset of regularly spaced cells that surround neuropil structures in the central brain (Fig. 5A,C) and ventral ganglion. Neuropil, including the subesophageal ganglion, the medulla and lobula (Fig. 5A), the antennal lobes and ventral nerve cord (data not shown), are surrounded by somata postive for *CG1732*. Between the medulla and lobula, *CG1732* labels a subset of giant glia at the inner chiasm (barbed arrow in Fig. 4A), a region that contains only glial and not neuronal cell bodies (Tix et al., 1997). This pattern is substantively different than that of GABA immunolabelled cells, which have cell bodies throughout the cortex surrounding the medulla and in a large cluster next to the lobula (Buchner et al., 1988).

CG1732 label (Fig. 5A,C) was compared in sequential sections to immunolabel for the glial nuclear marker, Repo (Fig. 5B,D) (Xiong et al., 1994). Repo labelled nuclei in positions of the perineural glia (arrows, Fig. 5B), the subperineural glia (carrots, Fig. 5B) and in regularly spaced intervals that form the outlines of several neuropilar structures. *CG1732* labelled only cells located between the cortex and neuropil (compare label surrounding the medulla and lobula in Fig. 5A with Fig. 5B). *CG1732* did not label perineural (arrows, Fig. 5A) or subperineural glia (carrots, Fig. 5A), which were labelled by Repo. Thus, the riboprobe for *CG1732* only labels the subset of glia that surround neuropil.

The label for *CG1732* had two additional, notable characteristics. First, there was a purple hue within the neuropil and not present throughout the cortex, which suggests that the mRNA for *CG1732* is possibly present within glial processes. Second, the purple hue is particularly dense in the distal one-third of the medulla (Fig. 5A,C), where the photoreceptors R7 and R8 make their first synapse. The dense label shows a banded pattern that runs parallel with the distal margin of the medulla and is visible in the left medulla (Fig. 5A). Repo-positive and *CG1732*-positive cells (notched open arrows in Fig. 5C,D) are located in similar positions and with similar spacing surrounding the neuropil. It appears that labelled processes emanate from regularly spaced cell bodies at the distal border of the medulla (Fig. 5C) and penetrate into the neuropil. The dense *CG1732*-labelling pattern ends abruptly at a consistent layer in the neuropil. Medulla neuropil glia are morphologically similar and localized similarly to these *CG1732* labelled cells [fig. 3A–C in Richardt et al. (Richardt et al., 2002)] and also send processes into the distal portion of the medulla, where the *CG1732* label is heaviest. *CG1732*-positive cells were not found in the cortex that surrounded the lamina or medulla.

Since *CG1732* is likely to be a GABA transporter and GABA uptake has been described in the lamina of the optic lobe (Campos-Ortega, 1974), we looked more closely at this region of the CNS. The optic lobes at higher magnification (boxed area of Fig. 5A,C) show that the lamina neuropil is lightly labelled with heavier label over a cluster of large cell bodies at the proximal border of the neuropil (arrowheads), where the outer chiasmatic and marginal glia both reside (Eule et al., 1995). The lamina neuropil has a weak purple tint that ends abruptly at the distal margin suggesting that cells that have processes that penetrate the lamina express this transporter. Epithelial glia would be the obvious candidate, as their processes surround the photoreceptor terminals in the laminar, and indeed in overdeveloped preparations, the cell bodies of the epithelial glia at the distal margin of the lamina appear to have faint label (data not shown). Nevertheless we were unable to definitively determine if the epithelial glia express *CG1732*.

The Repo antibody labels cells that both surround the neuropil and are intermixed with neuronal somata in the

Fig. 4. Localization of transcripts involved in monoaminergic neurotransmission using *in situ* hybridization. Representative frontal sections showing the labelling in the head (A,C,E,G–J) or the thorax (B,D,E). In (A,C,E) the ellipsoid body is circled with a dotted line to illustrate that sections are at approximately same depth in the head. (A,B) The *SerT* anti-sense riboprobe labels discrete cells in the head and in the thorax. Anatomical positions of the labelled cells in (A) are consistent with the SP1 (arrows), LP1 (notched arrow), and SE2 (barbed arrow) clusters of serotonergic neurons. Cells in (B) are present in the posterior portion of the ventral ganglion (arrows). (C,D) The *DAT* anti-sense riboprobe labels discrete cells in the head (C) and thorax (D). The anatomical position of the labelled cells suggests these cells are part of the DM (arrows), DL_1 (white arrowheads), and DL2 (carrot) clusters of neurons. The cells in (D) are present in the posterior portion of the ventral ganglion, and are likely to be part of the dorsal lateral cluster (barbed arrow) and the medial cluster (arrows). (E,F) The riboprobe for *CG33528* labels many cells (markers have remained consistent with A and C) throughout the head (E) and thorax (F) that are in similar anatomical locations as cells that label for the transporters, *SerT* and *DAT*. Gray open arrows denote labelling of the medullary layer (MC cells). (G) Wider view of *CG33528* labelling showing the cells of the brain that are labelled (arrows, barbed arrows and white arrowheads) and that a layer of cells at the base of the photoreceptors is also labelled (black arrowheads). Box denotes area of higher magnification in (H,I). (H,I) Serially sectioned preparation illustrating the region boxed in G but in a different preparation. (H) Section labelled with *CG33528* shows label right at the base of the photoreceptors (PR). Block arrows denote examples where the *CG33528* label has penetrated the photoreceptor layer. The dotted line represents the outline of the lamina neuropil (LaN). (I, top) Fluorescence image showing NrxIV antibody labelling glial septate junctions at the base of the photoreceptors in the same anatomical position as the *CG33528* label in (H). (I, bottom) For clarity, the brightfield image of the fluorescence image has been presented so that the morphological features (i.e. the characteristic striations of both the photoreceptor layer and the lamina neuropil) that define the different layers in optic lobes can be distinguished. The broken line outlines the lamina neuropil and the solid line denotes the base of the photoreceptors. (J) The riboprobe for the histamine synthesizing enzyme, *histidine decarboxylase* (*HDC*), labels cells in the central brain region (arrows) as well as the photoreceptor cell bodies (barbed arrows). In all panels dorsal is up. Scale bars, 50 μm.

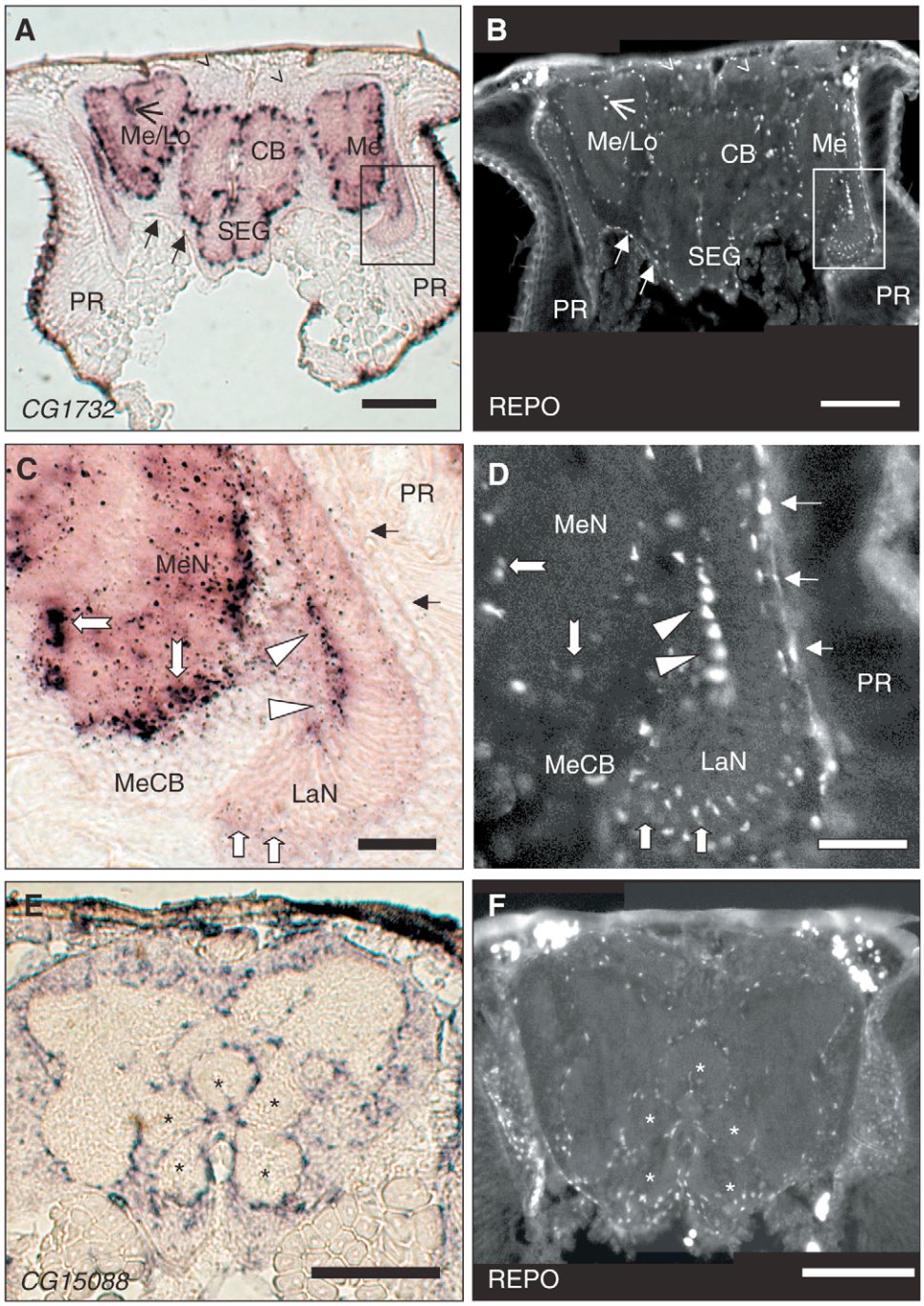


Fig. 5. *CG1732* and *CG15088* are expressed in a glial-like pattern. The various glial subtypes are shown in (B) and the corresponding area in a sequential section is shown in (A). Examples of perineural (arrow) and subperineural (carrot) glia are marked [nomenclature according to Eule et al. (Eule et al., 1995)]. Inner chiasmatic glia are marked by the barbed arrow. Neuropil are labelled Medulla/Lobula (Me/Lo), suboesophageal ganglion (SEG), central brain (CB), and photoreceptors (PR). (A) Frontal section of the fly head showing the brain and optic lobes. Cells labelled with the anti-sense riboprobe for *CG1732* surround the neuropils of both structures. The label occurs in regularly spaced intervals; the neuropils have a purple hue suggesting that the mRNA may be present in penetrating processes. (B) Sequential section to A labelled with the glial marker, Repo. Cells that express *CG1732* are likely to be glia because they occupy the same anatomical position and approximate spacing as cells that label with the nuclear glia marker Repo. Boxed areas in A and B denote areas shown at higher magnification in C and D, respectively. (C) Higher magnification of the optic lobes in A. In both C and D, arrowheads indicate labelled cells in either outer chiasm giant glia or marginal glia at the proximal margin of the lamina neuropil. Medulla neuropil (MeN), Medulla cell bodies (MeCB), and lamina neuropil (LaN) are denoted. Arrows show the location of glia at the distal margin of the lamina; open arrows show position occupied by epithelial glia. Notched arrows mark positions of medulla neuropil glia. Glia distal to the lamina neuropil do not label with *CG1732* and epithelial glia at the distal margin may faintly label for *CG1732*. (D) Higher magnification of B showing Repo label in the large glial nuclei of the outer chiasm glia (arrowheads) at the proximal margin of the lamina. Arrows show glia at the distal border of the lamina and open arrows show the epithelial glia; both label for Repo. Notched arrows mark the nuclei of medulla neuropil glia. (E) Pattern of label of *CG15088* probe. (F) the Repo antibody labels nuclei in similar sections from different preparations. Asterisks denote comparable neuropils. Like Repo, *CG15088* appears to label glia throughout the cortex. All sections are frontal sections with dorsal up; scale bars, 200 μm (A,B,E,F), 50 μm (C,D).

cellular cortex of both the lamina and the medulla. Prominent among the Repo-labelled nuclei are cells at the proximal border of the lamina (arrowheads, Fig. 5D) that are in the same position as *CG1732*-positive cells (arrowheads, Fig. 5C). Outside the neuropil borders, there exist Repo-labelled glia at the base of the photoreceptors (Fig. 5D, arrows) that are not labelled by *CG1732*, again indicating that that *CG1732* labels only a subset of glia.

CG15088 labels a different set of glia from *CG1732* (Fig. 5E). This putative SLC6 transporter is found within the IAAT subfamily rather than in a neurotransmitter transporter subfamily. *CG15088* is present in both a punctate and diffuse

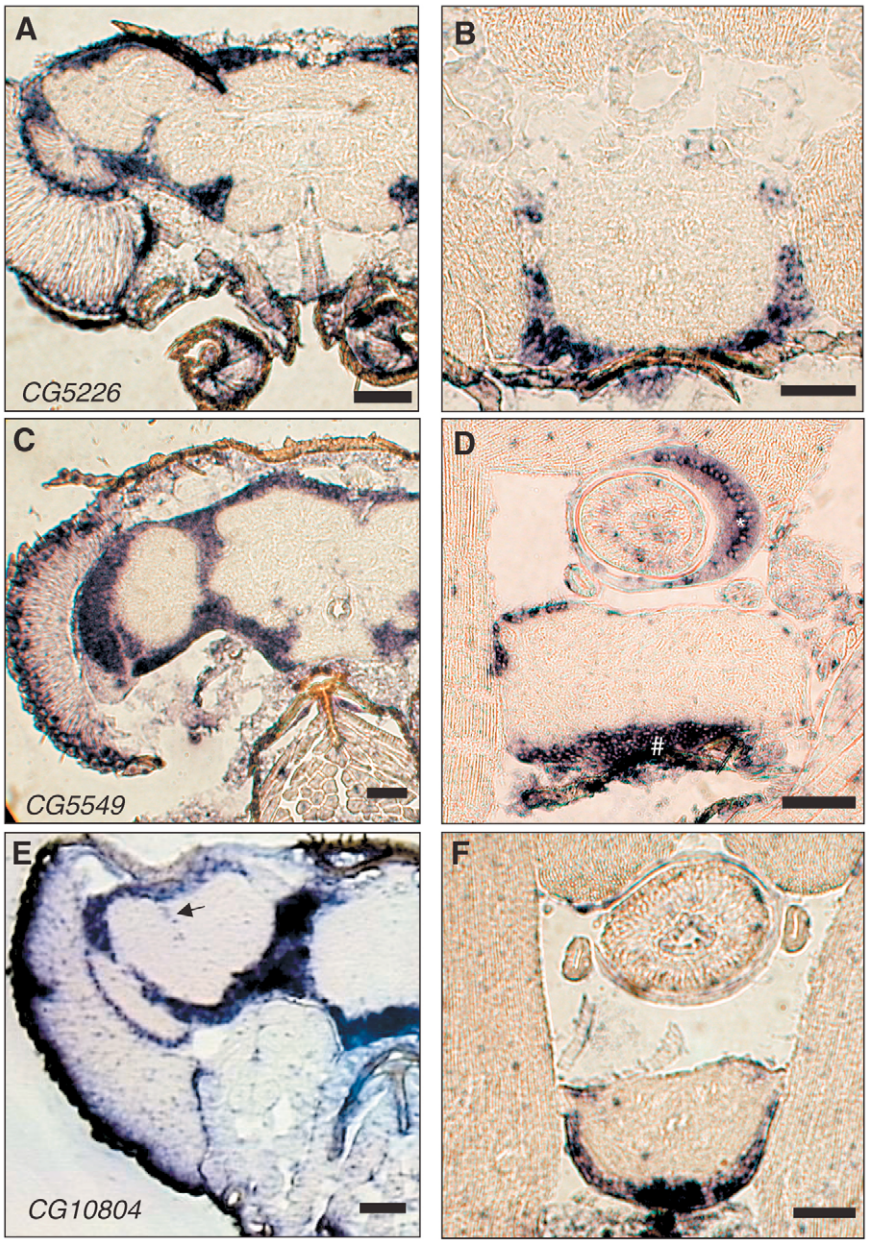


Fig. 6. *CG5226*, *CG5549*, and *CG10804* label cells broadly throughout the cell body layer in the adult head. (A) Representative horizontal section illustrating that the riboprobe for *CG5226* labels cells throughout the CNS cell body layer of the head (A) and representative frontal section showing labelling in the thorax (B). We did not detect glial labelling in the optic chiasm giant glia in addition to the cortical cell bodies. (C,D) Representative frontal sections showing the probe designed against *CG5549* labels the cellular cortex and the photoreceptors in the head (C) and CNS cells in the thoracic ganglion (#) and cardia (*) in (D). (E,F) The riboprobe for *CG10804* labels cells throughout the adult *Drosophila* head, including inner chiasm giant glia (arrow) (E) and CNS cells in the thorax (F). All panels are frontal sections with dorsal up; scale bars, 50 μm.

pattern throughout the cortex; the labelled cells line and define the neuropils at the inner border of the cortex and also are found throughout the cortex (Fig. 5E). This pattern is consistent with both that of subperineural and perineural glia. In contrast to *CG1732*, which labelled structures within the neuropil but not the cortex, the riboprobe for *CG15088* labels the cortex and does not label neuropil. *CG15088* label has a similar pattern to that of Repo-positive cells (Fig. 5F). Label from both probes surround neuropil structures such as the antennal lobes, ventral bodies and the ellipsoid body shown in Fig. 5E,F (asterisks), and both reagents label cells throughout the cortex. Also, we have detected *CG15088* label in the inner chiasm giant glia where only glial cell bodies are found (data not shown). These results suggest that *CG15088* may be expressed in all glia of the *Drosophila* cortex where it may play a role in transporting nutrient amino acids within the CNS.

Three Drosophila SLC6 homologs are broadly expressed in the CNS

CG5549, a candidate in the amino acid subfamily, and *CG5226* and *CG10804*, both segregating with the orphan neurotransmitter transporter subfamily, are expressed broadly throughout the cellular cortex in the head (Fig. 6A,C,E) and the thorax (Fig. 6B,D,F). The cortex comprises primarily neurons with glial cell bodies intermixed, but dense label throughout the head prevented us from distinguishing whether both types of cells were labelled. In the region of the inner optic chiasm between the medulla and lobula neuropil, only glial cell bodies reside (Tix et al., 1997), and we used this region to determine if at least this subset of glia was labelled by the probes. The riboprobes for *CG5549* and *CG10804* (but not *CG5226*) labelled large cell bodies at the inner optic chiasm (Fig. 6E, filled arrow). Of these three candidates, only *CG5549* labelled the photoreceptors. *CG5549* also labelled a cell layer in the cardia (asterisk in Fig. 6D), an organ at the transition between the foregut and midgut. The cardia is responsible for the creation of the peritrophic membrane, a four-layered extracellular barrier between the contents of the gut and the gut epithelium (King, 1988).

Eight SLC6 transporters were detected exclusively outside of the CNS

SLC6 members have been shown to transport compounds other than neurotransmitters, including amino acids and osmolytes, such as betaine and taurine. For *Drosophila* SLC6 transporters that had no detectable CNS expression, we extended our analysis to verify that the riboprobe would bind to transcripts expressed in other organs.

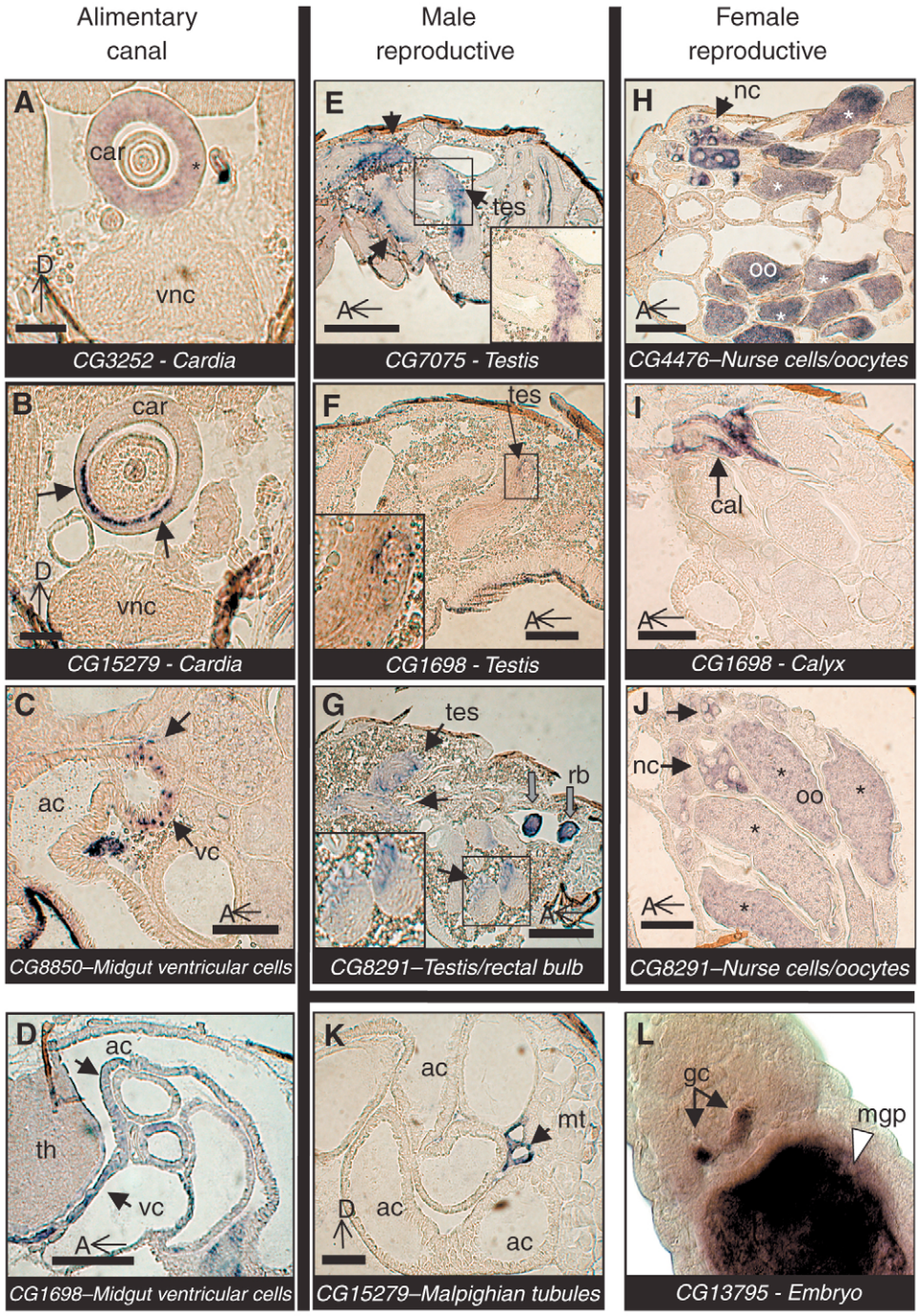


Fig. 7. Eight candidates label cells exclusively outside of the CNS. Panels show representative sections of a variety of tissues labelled using *in situ* hybridization, and are organized vertically by type of tissue labelled: (A–D) alimentary canal, (E–G) male reproductive system, (H–J) female reproductive system, and (K,L) other tissues. Frontal sections through the gut demonstrate that the riboprobe for *CG3252* (A) and *CG15279* (B) both label cells in the cardia (car; *), a structure at the transition between the foregut and the midgut. Ventral nerve cord is marked (vnc). (C,D) The anti-sense riboprobe for *CG8850* (C) and *CG1698* (D) both label ventricular cells (vc; filled arrows) that line the alimentary canal (ac) that runs from the thorax (th) into the abdomen. *CG8850* labels cells in a restricted domain of ventricular cells in the abdomen; neighbouring ventricular cells along the alimentary canal are not labelled. Sagittal sections reveal that *CG7075* (E), *CG1698* (F) and *CG8291* (G) label male reproductive tissue, likely the testis (tes) in the male abdomen (arrows). In addition, *CG8291* (G) labels the rectal bulb (rb; open arrows), a hindgut structure at the end of the alimentary canal. Insets show boxed areas at higher magnification. Horizontal (H) and sagittal (I,J) sections show that *CG4476* (H), *CG1698* (I), and *CG8291* (J) label structures in the female reproductive system. *CG4476* (H) and *CG8291* (I) label nurse cells (nc; arrows) and developing oocytes (oo; *). *CG1698* (I) labels a structure called the calyx (cal; arrows). (K) *CG15279* (mt; arrows) also labels two closely apposed Malpighian tubules. (L) Midgut precursor cells (mgp; arrowhead) and the Garland cells (gc; arrows) were labelled by *CG13795* in a stage 14 embryo. At later stages, only the Garland cells are labelled. Anterior of the embryo is in the upper left corner and this is a dorsal view of the embryo. Scale bars, 50 μm (A,B), 100 μm (C,F,I–K) and 200 μm (D,E,G,H). A, anterior; D, dorsal.

We determined if any of the candidate genes were expressed in other recognizable organs such as the alimentary canal, Malpighian tubules and/or reproductive organs.

Two candidates from the IAAT subfamily (*CG3252* and *CG15279*) were found in the cardia of the gut (Fig. 7A,B). The *Drosophila* cardia has been sub-divided longitudinally into 6 zones based on the morphological and intracellular characteristics of the constituent cells (King, 1988). *CG3252* is present in the long columnar cells that make up zone 5 of the cardia (Fig. 7A) and the label persists through zone 6 and into the posterior ventricular cells of the midgut (data not shown) where nutrient uptake is thought to occur. *CG15279* (Fig. 7B) is expressed in cells near the junction of the cardia and the ventricular cells but these cells are much shorter than the zone 5 cells. These cells are likely to be the cells of zone 6. *CG15279* label does not persist beyond the transition of the cardia to the intestine.

Three other candidates are also detected in cells along the

alimentary canal. Two candidates from the IAAT subfamily, *CG8850* and *CG1698*, are present in the digestive portion of the midgut. *CG8850* (Fig. 7C) is found in cells that line the gut lumen in a restricted portion; adjacent cells do not express *CG8850*. Midgut epithelial cells are responsible for the uptake of nutrients from consumed food (Miller, 1950). The *CG1698* (Fig. 7D) riboprobe labels a broader domain of the midgut that stretches from the thoracic into abdominal segments. *CG8291,* a candidate that does not segregate into a distinct subfamily, is expressed in the rectal bulb of the alimentary canal (Fig. 7G, arrows).

Fig. 7E–J shows tissues labelled by four transcripts found in the reproductive organs. *CG7075* is found exclusively in the testis of the male fly (Fig. 7E). The testes are recognizable by the elongated, striated appearance of the developing sperm within the organ (Miller, 1950). The riboprobes for *CG1698* (Fig. 7F,I) and CG8291 (Fig. 7G,J) both label the reproductive organs in both genders. *CG1698* labels the testis in the male (Fig. 7F) and a subset of cells in the female reproductive tract and potentially in the epidermis of the calyx, the region of the oviduct that meets the ovary (Fig. 7I). *CG8291* was detected in the testis (Fig. 7G, open arrows) and the nurse cells and oocyte in the female reproductive organ (Fig. 7J). Nurse cells furnish the developing oocyte with nutrients and maternally derived mRNAs (Mahajan-Miklos and Cooley, 1994). *CG4476* is expressed exclusively in the nurse cells and oocytes of the female reproductive system (Fig. 7H).

Only one candidate, *CG15279*, was found in the Malpighian tubules (Fig. 7K), the organ that carries out kidney-like function in the fly. The figure shows two closely apposed tubes that fit the morphological description of the cells and anatomical description of the Malpighian tubules (Miller, 1950). This apposition, and the cell morphology of the labelled structures, suggests that these structures are the Malpighian tubules somewhere after the stalk splits from one to two tubes. We were able to detect positive cells with these same characteristics in multiple regions of the abdomen in frontal sections (data not shown). Since the Malpighian tubules comprise a total of four branches that wind through the abdomen, it seems likely that the tubular cross-sections we observe belong to this system.

We detected *CG13795*, an outlier candidate sequence, in the embryo but not in the adult (Fig. 7I). The probe for *CG13795* labels the midgut precursors and the Garland cells at stage 14 and 15; as the gut matures to stage 16 it loses the expression of *CG13795*, but labelling of the Garland cell cluster, with its characteristic U-shape (Miller, 1950), is retained. Garland cells encircle the proventriculus, are known for their high endocytosis activity, and are proposed to carry out a liver-like function in cleaning the hemolymph of the fly (Kosaka and Ikeda, 1983).

Discussion

In this study, we set out to identify the full complement of *Drosophila* SLC6 transporters present in the *Drosophila* genome and determine their expression pattern in the adult CNS. Our search identified 21 sequences with sequence similarity to the SLC6 family, Table 1 lists these sequences and summarizes the overall findings. A majority of these sequences

Table 1. *Summary of Drosophila SLC6 family transporters*

Putative family	Transporter	Chromosome	Localization	RT-PCR expression
Serotonin	Dm SerT	60C8	CNS cells	Head/thorax
Dopamine	Dm DAT	53C7-C14	CNS cells	Head/thorax
Glycine	Dm CG5549	60A2-A3	Broad CNS	Equivalent
Proline	Dm CG7075	28C1	Male abdomen	Male abdomen
GABA	Dm CG1732	102D4	Glia	Head/thorax
‘Orphan’ NT transporters	Dm CG5226	55D4	Broad CNS	Equivalent
	Dm CG10804	3D4	Broad CNS	Head/thorax
	Dm Ine	24F4	Photoreceptors, CNS, digestive tract[1]	Equivalent
Insect AA transporters	Dm CG1698	46B3	Alimentary canal/reproductive cells	Equivalent
	Dm CG4476	67A2-A3	Female abdomen	Female abdomen
	Dm CG15279	35B6	Cardia/Malpighian tubules	Prominent in abdomen
	Dm CG8850	48F6	Ventricular cells	Equivalent
	Dm CG15088	55E11	Glia	Equivalent
	Dm CG3252	4F9	Cardia	Equivalent
Unclassified	Dm Blot	74B	Epithelium, CNS, Malpighian tubules[2]	Head/abdomen
	Dm CG8291	52D2	Reproductive cells	Equivalent
	Dm CG13793		Not done	Equivalent
	Dm CG13794		Not done	Equivalent
	Dm CG13795	28C2	Embryonic Garland cells/midgut	Equivalent
	Dm CG33296		Not done	Equivalent
	Dm CG13796	28C2	No *in situ*	Equivalent

[1](Soehnge et al., 1996; Burg et al., 1996).
[2](Johnson et al., 1999).

either segregated into four previously recognized subfamilies or helped to define a novel subfamily, the IAAT subfamily. Using *in situ* hybridization, we localized seven candidates to the CNS. Of these seven genes, two known genes (*SerT* and *DAT*) were expressed in a discrete pattern in the brain and thoracic ganglion predicted by known immunolabelling, two candidates (*CG1732* and *CG15088*) were expressed in a pattern consistent with glia, and three candidates (*CG5549*, *CG5226* and *CG10804*) were broadly expressed in the CNS. Neither phylogenetic analysis nor the *in situ* hybridization results suggested a candidate for an octopamine or histamine transporter despite the known presence of neurons containing these amines. Eight candidates labelled cells in the alimentary canal, reproductive organs and Malpighian tubules but were not detected in the CNS. We did not attempt or were unable to determine a pattern for the six remaining candidates for reasons discussed below.

Drosophila SLC6 have similar structural features to the leucine transporter from Aquifex aeoliscus

The crystal structure of a bacterial Na^+-dependent leucine transporter homologous with mammalian SLC6 was recently solved. Of particular interest, the structure reveals residues important for substrate binding, sodium binding and gating of the transporter. One major difference between our multiple sequence alignments and those of Yamashita et al. (Yamashita et al., 2005) is the inclusion of a larger number of sequences (see Fig. 1B and Fig. S1 in supplementary material). There are no absolutely invariant residues identified in our complete alignment of 84 putative transporters, and the only invariant residue in the *Drosophila* transporter alignment is a glycine residue at the end of TM2. We also find several residues that seem to be more conserved among metazoan transporters than would be suggested by the Yamashita alignment. The densest concentration of amino acid residues with direct contacts with the substrate and sodium were found in TM1, TM6 and TM8 with other critical residues scattered throughout the protein. Many of the known and candidate *Drosophila* SLC6 sequences contain these precise residues or have conservative amino acid substitutions, although there are clear differences among certain divergent transporter subgroups that may underlie differences in substrate specificity, ion binding and gating (see Fig. 1 and Fig. S1 in supplementary material).

Blot is perhaps one of the more divergent transporter sequences, yet it contains a conserved arginine responsible for gating and a conserved asparagine involved in sodium binding present in TM1, as well as many other conserved residues. In addition, Blot has been localized and the absence of Blot has a phenotype (Johnson et al., 1999), therefore Blot is a functional protein despite the disparity in sequence homology.

For CG33296 and CG13793–CG13796, the situation is not as clearcut, but our data suggests that they are expressed, as we have detected at least low expression by RT-PCR of each of these genes. Thus far, the predicted proteins for CG13793 and CG33296 are missing the critical TM1 and part of TM2. This could be a result of protein prediction errors or a common first exon that is shared between these proteins (see below). Another possibility is that these proteins have adopted a slightly different method of functioning from the ancestral SLC6 transporter. CG13794, CG13795 and CG13796 contain divergent TM1 sequences (though they still retain several highly conserved residues) throughout this and other segments of the alignment. For CG13795, we have demonstrated discrete expression in the Garland cells in late stage embryos, suggesting a specific role in these cells.

Sequences in the Drosophila genome help define a novel, large subfamily, IAAT

Our results confirm and expand a sixth, novel subfamily of SLC6 transporters, referred to as IAAT (Boudko et al., 2005), which we show includes six putative transporters from *Drosophila* and seven putative transporters from *Anopheles*. With the exception of minor differences in branching geometry and branching order at nodes with low bootstrap values, the clusters of transporters in the two trees are largely the same (Boudko et al., 2005). The sequence differences between the IAAT and other SLC6 subfamilies may underlie physiological differences in the function of these transporters. There are three cloned members of this subfamily, the *Manduca sexta* proteins known as potassium-coupled amino acid transporter-1 (KAAT1) (Castagna et al., 1998), the cation-anion-activated amino acid transporter/channel-1 (CAATCH1) (Feldman et al., 2000) and the *Aedes aegypti* amino acid transporter (AeAAT1) (Boudko et al., 2005). All three are competent to transport amino acids using either K^+ or Na^+ as the driving ion, especially at highly negative membrane potentials. In contrast, other SLC6 transporters use only Na^+ as their driving ion. The molecular basis of this ion specificity may relate to residues in the transporter responsible for coordinating these ions. In this regard, it is intriguing that the IAAT transporters all contain an alanine or serine substitution at the site of the glycine residue in TM1a reported to be involved in coordinating sodium ion Na1 (Yamashita et al., 2005). It should also be noted that other sodium coordinating residues or residues whose side chains reportedly interact with sodium ions in TM6a and TM8 are somewhat less conserved across the entire family of transporters.

The physiological difference in ion selectivity is thought to reflect the environment of the insect gut in which these transporters operate. In *Manduca*, goblet cells secrete high concentrations of potassium into the lumen of the gut; in addition, the columnar cells that take up the amino acids have an unusually high membrane potential across their luminal (apical) membrane (Harvey and Wieczorek, 1997). Nutrient uptake into the columnar cells is driven by both the high concentration of K^+ and the large negative membrane potential. All three of the cloned transporters are associated with the insect midgut cells: *KAAT1* was localized to columnar cells, *CAATCH1* was cloned from a library created from midgut epithelium, and *AeAAT1* was cloned from a posterior midgut library and was localized to various structures in the gut, including the cardia, posterior midgut and Malpighian tubules (Boudko et al., 2005).

Our *in situ* data, however, demonstrate that *Drosophila* IAAT subfamily members are expressed in more varied tissues than the gut. Confirming this, RT-PCR data suggests that five out of six *Drosophila* members of the IAAT subfamily are expressed in all three segments of the fly, the exception being *CG4476*, which is only weakly expressed in tissues outside of the female abdomen. *In situ* hybridization revealed that IAAT members were expressed in such diverse tissues as CNS (*CG15088*), midgut (*CG1698* and *CG8850*), cardia (*CG3252* and *CG15279*) and the Malpighian tubules (*CG15279*). As the brain is not expected to have a high potassium environment or unusually high membrane potentials as in *Manduca* gut, it is possible that the *Drosophila* transporters from the IAAT subfamily could use either Na^+ or K^+ as their driving ion, depending on ion availability in a given environment.

We can compare our IAAT localization results with two independent sources of information: microarray expression studies being carried out on isolated tissues and the embryonic *in situ* hybridization project being carried out by the Berkeley Drosophila Genome Project (BDGP) (Tomancak et al., 2002). Microarray data from the adult Malpighian tubules demonstrated that only the SLC6 transcript, *CG15279*, was enriched (30-fold) compared against the rest of the fly (Wang et al., 2004). This agrees with our *in situ* hybridization results (Fig. 7K).

BDGP has partially completed a systematic determination of embryonic expression for each annotated *Drosophila* gene using *in situ* hybridization. Patterns for four of the IAAT subfamily members have been released. *CG3252* was found in the embryonic hindgut and the Malpighian tubules, whereas we detected it in the cardia of the adult alimentary canal. *CG4476* was detected in early embryos, probably due to maternal contribution, and in the embryonic stomatogastric nervous system, germ cells and endocrine system. In the adult, we detected *CG4476* in nurse cells and developing oocytes of the female reproductive system, which confirms that mRNA for *CG4476* is maternally contributed to the embryo. BDGP reported expression of *CG1698* in the embryonic proventriculus, frontal ganglion and stomatogastric nervous system, but we detected *CG1698* in the reproductive organs and midgut of the adult. *CG8850* was identified in the embryonic Malpighian tubules by BDGP, but we found it expressed in cells lining the adult ventriculus in the midgut. These expression differences between the adult and the embryo are likely due to the specific requirements of the cells at the adult and embryonic stage.

Discrepancies exist between our RT-PCR and *in situ* localization observations. Indeed such discrepancies might be expected. The areas of detection differed between *in situ* hybridization and RT-PCR. Our *in situ* hybridization focused on the CNS and on other tissues in which we could clearly distinguish a signal, while for RT-PCR an entire body segment was used to derive the cDNA. Therefore, the RT-PCR may have detected gene expression in a tissue that we did not examine using *in situ* hybridization. Also, RT-PCR is the more sensitive technique; a transcript expressed at low levels but broadly throughout a tissue would be easier to pick up by RT-PCR.

Monoaminergic neurotransmitter transporters, SerT and DAT are found in presynaptic neurons in the adult CNS

We have localized *SerT* and *DAT* to neurons in the adult CNS that label in a pattern similar to the label for the respective neurotransmitter. Neurotransmitter transporters have been localized to presynaptic neurons, postsynaptic neurons and surrounding glia; *SerT* in the embryo (Demchyshyn et al., 1994) and *DAT* in the larva (Porzgen et al., 2001) have been localized to cells in the pattern of cells that immunolabelled for their respective neurotransmitter. Not every cell that labels for dopamine was positive for DAT labelling, indicating that all cells that release dopamine may not re-capture this compound. The localization of monoaminergic neurotransmitter transporters agrees with localization data from other species, which are found in the presynaptic neurons (Hoffman et al., 1998).

Amino acid transporter subfamily

There is only a single *Drosophila* SLC6 transporter (*CG5549*) that segregated into the amino acid subfamily. *CG5549* is expressed in each segment of the fly, as determined by RT-PCR, and is expressed broadly throughout the adult CNS, including the inner chiasm giant glia and the photoreceptors. Cloned members of this subfamily have substrates which include amino acids that double as neurotransmitters, such as proline or glycine (Malandro and Kilberg, 1996), or a broader substrate profile of cationic and neutral amino acids for the transporter B^{O+} (Sloan and Mager, 1999). The broad expression of *CG5549* in both neurons and glia of the CNS suggests a basic role for this candidate, such as nutrient uptake.

Orphan neurotransmitter transporter subfamily

The two candidates that segregate with the orphan transporters, *CG5226* and *CG10804*, both displayed a broad localization pattern throughout the CNS. Within the CNS, both *CG5226* and *CG10804* were detected in neurons. *CG10804* expression was also detected in the inner chiasmatic glia but we were unable to detect *CG5226* in this subset of glia. RT-PCR revealed expression differences. *CG10804* was enriched in the head and thorax, segments containing cell bodies of the CNS, whereas *CG5226* was evenly expressed across all segments. These expression differences suggest that these two orphan transporters may be carrying out broad but different functions in the fly, such as importing a necessary compound for cellular function or clearing a compound with a wide extracellular distribution. The embryonic localization by BDGP of *CG5226* (*CG10804* has not been released) shows that *CG5226* is expressed in cells throughout the more mature CNS of embryonic stages 13–16, but not in earlier CNS stages when the CNS is still developing. The similar localization pattern for *CG5226* in the embryo and adult suggests a similar role in more mature CNS cells. It has been predicted that substrates for

members of the orphan neurotransmitter transporter family will include amino acids (Boudko et al., 2005), based on phylogenetic analysis, and our localization data is consistent with this prediction. Interestingly, these transporters possess highly divergent extracellular linker sequences 2, 4a and 6, consistent with a role for these sequences in substrate selectivity.

GABA transporter subfamily

In *Drosophila*, uptake of the neurotransmitter GABA has been demonstrated in the lamina (Campos-Ortega, 1974), but the gene responsible for this uptake has not been identified. There is compelling evidence that *CG1732* is a GABA transporter (GAT) in *Drosophila*. It is the only candidate that segregated into the GABA subfamily, and it has significant identity with other known GABA transporters: *Manduca sexta* GAT (80%), *Trichplusia ni* GAT (81%), human GAT1 (59%) and human GAT3 (53%). By RT-PCR, *CG1732* is enriched in the head and thoracic segments but it is also present in the abdomen segment.

Evidence from other species suggests that GABA transporters can be present in both neurons and glia (Borden, 1996), but our data demonstrate that *CG1732* is expressed in a subset of glia. In *Drosophila*, the pattern of *CG1732* label only partially overlaps with the pattern of GABA immunolabelling in the optic lobes (Buchner et al., 1988). The GABA antibody labels approximately 1500 somata in the cortex of the medulla and, in the neuropil, there is faint label in the lamina, and denser label in the medulla. *CG1732* label is found in the neuropil but not in cell bodies of the cortex, suggesting that GABAergic neurons are not the primary source of *CG1732* expression in the medulla. Instead, the labelling pattern for *CG1732* is better represented by the immunolabelling pattern for the glial marker, Repo. In addition, glial expression of a *Drosophila* GABA transporter is supported by GABA uptake studies in the lamina of housefly and *Drosophila*, where ^{3}H-GABA was accumulated in the glia rather than neurons of the optic lobes (Campos-Ortega, 1974). These authors described uptake into glia of the distal border of the lamina (epithelial glia) and at the proximal border (marginal glia). We were not able to conclusively show expression of CG1732 in these cell types, but the lamina does show weak labelling in the lamina neuropil. Finally, in the embryo BDGP localized *CG1732* expression to a subset of cells in the ventral nerve cord that are in a position consistent with the channel glia. Thus, our evidence, combined with published findings, indicates that expression of *CG1732* is in a subset of glia in *Drosophila*.

In another insect, *Manduca sexta* (Umesh and Gill, 2002), MasGAT immunoreactivity coincided with GABA immunoreactivity (Homberg et al., 1987) in the neuropil but did not appear to be in coincident locations in the cortex. In adult *Manduca*, somata immunoreactive for GABA appear throughout the cortex of the optic lobe (Homberg et al., 1987), as is the case for *Drosophila*. Somata in this same area of cortex are not labelled by the antibody against MasGAT (Umesh and Gill, 2002) but somata near the neuropil are clearly labelled and often send a process into the nearest neuropil. In fact, some of these author's photographs reveal a band of dark staining surrounding the optic lobe neuropil, similar to *CG1732* labelling that we detect. These results suggest various possibilities. First, MasGAT is expressed in GABAergic neurons, but the final position of the MasGAT protein is at the terminals and not the cell bodies of GABAergic neurons. Therefore the antibody labelled only the terminals, leaving GABAergic cell bodies unlabelled. This explanation is possible though other somata were clearly labelled by the MasGAT antibody. It is also possible that only a small subset of neurons very near the neuropil express MasGAT. Another possibility is that MasGAT is expressed in glia similar to our conclusion in *Drosophila*.

Candidates unaffiliated with a defined subfamily

In our phylogenetic analysis, five candidates did not fall into any distinct subfamily, not even into the orphan subfamily. We have called these 'unclassified sequences' to distinguish these proteins from the previously defined orphan subfamily. We were able to localize these candidates outside the CNS.

Two unclassified candidates (*CG7075* and *CG8291*) are associated with the reproductive tissue. By *in situ* hybridization, *CG7075* is found in the testes of the male, and our RT-PCR results show that *CG70705* is expressed almost exclusively in the male abdomen, which supports our *in situ* localization results. In the adult, *CG7075* is likely to be a testes-specific gene because we did not detect it elsewhere in the fly. Also, *CG7075* was found as an EST (BF486171) from adult testes; testis expression was noted in a Flybase communication (Bazinet, 2000.7.10). *CG7075* was also found to be a testis-specific gene as determined by microarray analysis (Parisi et al., 2004).

CG8291 is localized in the reproductive tissue in both genders, in nurse cells and oocytes in the female and in the testes in the male. In a microarray study, Parisi et al also found that *CG8291* had enriched expression in the testis (Parisi et al., 2004). The *CG8291* riboprobe also labelled a structure along the alimentary canal, the rectal bulb, of both males and females. The rectal bulb is thought to be responsible for the breakdown of the peritrophic membrane, which forms a barrier between the contents of the gut and the gut epithelium and is generated by the cardia (Miller, 1950). Unfortunately, there is no clear common feature among the tissues in which *CG8291* is expressed which could lead to a hypothesis regarding *CG8291* function.

Interestingly, several of our unclassified candidate genes are organized in a cluster at the genetic region of 28C. The open reading frames for *CG13793–CG13796*, *CG33296* and *CG7075* all reside in what appears to be a region of genetic duplications, with up to 98% identity in portions of the repeated regions. *CG13795* is 92% identical to *CG13793* and 80–89% identical to *CG33296*, so it was difficult to confidently distinguish one sequence from another in this area. The organization of this region may indicate a unique relationship between these genes. One possibility is that a common exon

could be spliced to any one of the genes to create differentially functioning proteins. Alternatively, the subtle sequence differences could be significant for the timing or localization of candidate expression. This genetic region could form some sort of locus that possibly functions similarly to the cholinergic locus in *Drosophila*, where two sequences with different functions share the first exons (Kitamoto et al., 1998). This region could also represent an area of gene duplication, in which transporters with different substrate profiles are generated as postulated by Boudko et al. (Boudko et al., 2005).

Candidates not localized by in situ hybridization

Five identified candidates (*ine*, *blot*, CG13793, CG13794 and CG33296) were eliminated from the *in situ* hybridization study. Two of the candidates, *blot* and *ine*, have already been cloned and localized (Burg et al., 1996; Huang et al., 2002; Johnson et al., 1999) and were not pursued further. Three candidates (*CG13793*, *CG13794* and *CG33296*) were not pursued further because they did not have a predicted N terminus that agreed with other SLC6 homologs, and these candidates are located in a region likely with a genetic duplication. We approached CG13795 and CG13796 as representative sequences from this area. Neither CG13795 nor CG13796 were found in the adult, and therefore we did not pursue CG13793, CG13794 or CG33296 further. The fact that we were able to detect CG13795 in the embryo shows that the entire region cannot be attributed to pseudogenes and our RT-PCR results show that there is weak expression of each of these transcripts in the adult. Perhaps this genetic region represents an area where the genome has expanded and will generate an expansion of transporter population as described (Boudko et al., 2005).

Finally, one potential candidate with weak homology to SLC6 transporters (*CG31904*) was excluded from our study because homology extended only over a small fraction of the predicted protein. This small region is predicted to be fused with the adult cuticular protein 1 (which has no sequence similarity to SLC6 transporters), raising strong suspicions about an error in gene prediction.

Monoamine neurotransmitter transporters not found in our analysis

One surprising aspect of this study is that we did not find candidate sequences for the histamine and octopamine transporters in the monoamine subfamily. The absence of a *Drosophila* octopamine transporter (OAT) is perplexing since an OAT gene does exist in another insect, *Trichplusia ni* (Malutan et al., 2002), and, like *T. ni*, there are cells in *Drosophila* that immunolabel for octopamine (Monastirioti et al., 1995). Furthermore, octopamine influences behaviors such as grooming and locomotion (Yellman et al., 1997). The question remains whether *Drosophila* uses an octopamine transporter and if so how divergent the *Drosophila* sequence is from that of *T. ni*.

The absence of a histamine transporter sequence is similarly puzzling. Histamine is the neurotransmitter of *Drosophila* photoreceptors, and genetic evidence indicates that a mechanism to accumulate histamine into photoreceptors exists in *Drosophila* (Melzig et al., 1998), as it does in other species (Battelle et al., 1999; Stuart et al., 2002; Stuart et al., 1996). Flies lacking the enzyme that synthesizes histamine, histidine decarboxylase (HDC), are blind and have lost histamine immunolabelling in the optic lobe. These flies can then be fed histamine, which restores vision and histamine immunolabelling, suggesting that in the fly there is a process by which histamine can be accumulated in the photoreceptors and restore function to those cells. Based on data from another species (Stuart et al., 2002; Stuart et al., 1996), one would predict that the histamine uptake is dependent on Na^+ and Cl^-, as are the other SLC6 transporters.

Indeed, none of the SLC6 transporters were expressed in the pattern of histamine uptake, that is, in the photoreceptors and in cells located in the antennal lobe (Melzig et al., 1998), as would be expected in a direct uptake model. At the photoreceptor terminals, histamine may recycle in a more circuitous route through the surrounding glia (Borycz et al., 2002). Recent work on mutant flies has revealed that histamine may be conjugated to β-alanine into an intermediate compound, carcinine, by the enzyme Ebony. Ebony was localized by antibody to both epithelial and medulla neuropil glia that surround the photoreceptor terminals (Richardt et al., 2002). In this scheme, histamine released from the photoreceptors would be taken up into the glia and converted to carcinine. The carcinine then must be shipped to the photoreceptors where it would be catabolized into histamine and β-alanine by the enzyme tan. In this scheme the histamine transporter would reside in the surrounding glia and a separate transporter would exist in the photoreceptors to take up carcinine. However, we did not find an SLC6 homolog associated solely with either the photoreceptors or with their surrounding glia.

VMAT/CG33528

A sequence that is potentially involved with monoaminergic neurotransmission is the VMAT, *CG33528* (Greer et al., 2005). It is 44% identical to the mammalian VMAT2, which packages histamine into vesicles (Erickson et al., 1995). Our anti-sense riboprobe decorated a small number of cells in the central brain that are likely to be monoaminergic neurons (Monastirioti, 1999) and, unexpectedly, glial cells in a thin layer at the base of the photoreceptors. But this riboprobe did not label the photoreceptors, which are histaminergic and would be expected to express such a transporter. Immunocytochemistry also fails to find CG33528 in the photoreceptors or other histaminergic neurons although it co-localized with dopamine, 5-HT and octopamine neurons as expected (Chang et al., 2006).

Curiously, the *CG33528* label at the base of the photoreceptors is in glia, not neurons. Two types of glia (fenestrated and pseudocartridge) have been described in this region based on anatomy and enhancer traps (Eule et al., 1995; Saint Marie and Carlson, 1983). In the house fly, glia surround the photoreceptor axons as they traverse the lamina cortex.

Fenestrated glia contain large vesicles (~80–100 nm) and psuedocartridge glia contain even larger vacuoles (~0.2–2.0 μm); since these glia are thought to help form a type of blood–brain barrier (Auld et al., 1995), the vesicles and vacuoles may relate to this function. The monoamines histamine and serotonin are both present in the lamina, histamine in the photoreceptors and serotonin in nerve terminals that course amongst the monopolar cell bodies in the lamina cortex (Buchner et al., 1988; Pollack and Hofbauer, 1991). A vesicular transporter in glia is not without precedent. Mammalian astrocytes have been documented to express vesicular glutamate transporters and release glutamate from vesicles onto neurons in a regulated manner (Bezzi et al., 2004).

Is CG33528 involved in the recycling of histamine at the photoreceptor terminals? We think the anatomy argues against this possibility. First, neither the processes of the fenestrated nor pseudocartridge glia extend into the lamina neuropil where the photoreceptors synapse onto the monopolar cells (Saint Marie and Carlson, 1983). Epithelial glia reside in this region. Second, CG33528 has a much higher affinity for monoamines other than histamine (Greer et al., 2005). Given the juxtaposition of serotonin nerve endings with the glia expressing CG33528, perhaps this transporter is primarily concerned in some way with serotonin. Serotonergic terminals have been shown to penetrate into the lamina cortex (Buchner et al., 1988).

List of abbreviations

AA	amino acid
Ag	*Anopheles gambiae*
AL	antennal lobe
AM	midline cell
BDGP	Berkeley *Drosophila* genome project
blot	bloated tubules
CNS	central nervous system
DAT	dopamine transporter
Dm	*Drosophila melanogaster*
EST	expressed sequence tag
GABA	γ-aminobutyric acid
GAT	GABA transporter
HDC	histidine decarboxylase
IAAT	insect amino acid transporter
ine	inebriated
LaN	lamina neuropil
MC	medulla cell
MeN	medulla neuropil
Ms	*Manduca sexta*
NA	numerical aperture
NGS	normal goat serum
NrxIV	Neurexin IV
NSS	neurotransmitter:sodium symporter family
OAT	octopamine transporter
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PFA	paraformaldehyde
RT	room temperature
RT-PCR	reverse transcription-polymerase chain reaction
SerT	serotonin transporter
SNF	sodium neurotransmitter transporter family
TBS	Tris-buffered saline
TM	transmembrane
Tn	*Trichoplusia ni*
VMAT	vesicular monoamine transporter
VNC	ventral nerve cord

We would like to acknowledge the contributions of many people that made this work possible. Steve Crews provided support for this foray into fly biology. We thank David Krantz and Rapheal Romero for valuable discussions and sharing unpublished results. Antibodies were provided by Manzoor Bhat (Neurexin IV) and the Developmental Hybridoma Studies Bank (Repo; Iowa City, IA, USA). We appreciate the histology expertise of Kirk McNaughton and Barb Taylor. Alan Fanning and Jim Anderson allowed us use of their Nikon microscope and Husok Lee generously provided us with an HDC plasmid. This work was funded by the University Research Council and Medical Alumni Association of the University of North Carolina at Chapel Hill and NIH grant MH65501, both to A.E.S.

References

Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., Scherer, S. E., Li, P. W., Hoskins, R. A., Galle, R. F. et al. (2000). The genome sequence of *Drosophila melanogaster*. *Science* **287**, 2185-2195.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389-3402.

Amara, S. G. and Arriza, J. L. (1993). Neurotransmitter transporters: three distinct gene families. *Curr. Opin. Neurobiol.* **3**, 337-344.

Auld, V. J., Fetter, R. D., Broadie, K. and Goodman, C. S. (1995). Gliotactin, a novel transmembrane protein on peripheral glia, is required to form the blood-nerve barrier in Drosophila. *Cell* **81**, 757-767.

Battelle, B. A., Calman, B. G. and Hart, M. K. (1999). Cellular distributions and functions of histamine, octopamine, and serotonin in the peripheral visual system, brain, and circumesophageal ring of the horseshoe crab *Limulus polyphemus*. *Microsc. Res. Tech.* **44**, 70-80.

Baumgartner, S., Littleton, J. T., Broadie, K., Bhat, M. A., Harbecke, R., Lengyel, J. A., Chiquet-Ehrismann, R., Prokop, A. and Bellen, H. J. (1996). A Drosophila neurexin is required for septate junction and blood-nerve barrier formation and function. *Cell* **87**, 1059-1068.

Bazinet, C. (2000.7.10). FlyBase error report for CG7075 on Mon Jul 10 03:52:53 2000.

Bezzi, P., Gundersen, V., Galbete, J. L., Seifert, G., Steinhauser, C., Pilati, E. and Volterra, A. (2004). Astrocytes contain a vesicular compartment that is competent for regulated exocytosis of glutamate. *Nat. Neurosci.* **7**, 613-620.

Borden, L. A. (1996). GABA transporter heterogeneity: pharmacology and cellular localization. *Neurochem. Int.* **29**, 335-356.

Borycz, J., Borycz, J. A., Loubani, M. and Meinertzhagen, I. A. (2002). *tan* and *ebony* genes regulate a novel pathway for transmitter metabolism at fly photoreceptor terminals. *J. Neurosci.* **22**, 10549-10557.

Boudko, D. Y., Kohn, A. B., Meleshkevitch, E. A., Dasher, M. K., Seron, T. J., Stevens, B. R. and Harvey, W. R. (2005). Ancestry and progeny of nutrient amino acid transporters. *Proc. Natl. Acad. Sci. USA* **102**, 1360-1365.

Broer, A., Klingel, K., Kowalczuk, S., Rasko, J. E., Cavanaugh, J. and Broer, S. (2004). Molecular cloning of mouse amino acid transport system

B0, a neutral amino acid transporter related to Hartnup disorder. *J. Biol. Chem.* **279**, 24467-24476.

Buchner, E., Bader, R., Buchner, S., Cox, J., Emson, P. C., Flory, E., Heizmann, C. W., Hemm, S., Hofbauer, A. and Oertel, W. H. (1988). Cell-specific immuno-probes for the brain of normal and mutant *Drosophila melanogaster*. I. Wildtype visual system. *Cell Tissue Res.* **253**, 357-370.

Budnik, V. and White, K. (1988). Catecholamine-containing neurons in *Drosophila melanogaster*: distribution and development. *J. Comp. Neurol.* **268**, 400-413.

Burg, M. G., Sarthy, P. V., Koliantz, G. and Pak, W. L. (1993). Genetic and molecular identification of a Drosophila histidine decarboxylase gene required in photoreceptor transmitter synthesis. *EMBO J.* **12**, 911-919.

Burg, M. G., Geng, C., Guan, Y., Koliantz, G. and Pak, W. L. (1996). Drosophila *rosA* gene, which when mutant causes aberrant photoreceptor oscillation, encodes a novel neurotransmitter transporter homologue. *J. Neurogenet.* **11**, 59-79.

Campos-Ortega, J. A. (1974). Autoradiographic localization of ^{3}H-gamma-aminobutyric acid uptake in the lamina ganglionaris of *Musca* and *Drosophila*. *Z. Zellforsch. Mikrosk. Anat.* **147**, 415-431.

Castagna, M., Shayakul, C., Trotti, D., Sacchi, V. F., Harvey, W. R. and Hediger, M. A. (1998). Cloning and characterization of a potassium-coupled amino acid transporter. *Proc. Natl. Acad. Sci. USA* **95**, 5395-5400.

Chang, H. Y., Grygoruk, A., Brooks, E. S., Ackerson, L. C., Maidment, N. T., Bainton, R. J. and Krantz, D. E. (2006). Overexpression of the *Drosophila* vesicular monoamine transporter increases motor activity and courtship but decreases the behavioral response to cocaine. *Mol. Psychiatry* **11**, 99-113.

Chiu, C., Ross, L. S., Cohen, B. N., Lester, H. A. and Gill, S. S. (2000). The transporter-like protein inebriated mediates hyperosmotic stimuli through intracellular signaling. *J. Exp. Biol.* **203**, 3531-3546.

Corey, J. L., Quick, M. W., Davidson, N., Lester, H. A. and Guastella, J. (1994). A cocaine-sensitive *Drosophila* serotonin transporter: cloning, expression, and electrophysiological characterization. *Proc. Natl. Acad. Sci. USA* **91**, 1188-1192.

Demchyshyn, L. L., Pristupa, Z. B., Sugamori, K. S., Barker, E. L., Blakely, R. D., Wolfgang, W. J., Forte, M. A. and Niznik, H. B. (1994). Cloning, expression, and localization of a chloride-facilitated, cocaine-sensitive serotonin transporter from *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* **91**, 5158-5162.

Erickson, J. D., Eiden, L. E., Schafer, M. K. and Weihe, E. (1995). Reserpine- and tetrabenazine-sensitive transport of (3)H-histamine by the neuronal isoform of the vesicular monoamine transporter. *J. Mol. Neurosci.* **6**, 277-287.

Eule, E., Tix, S. and Fischbach, K. F. (1995). Atlas of glial cells in the adult optic ganglion.

Feldman, D. H., Harvey, W. R. and Stevens, B. R. (2000). A novel electrogenic amino acid transporter is activated by K^+ or Na^+, is alkaline pH-dependent, and is Cl^- independent. *J. Biol. Chem.* **275**, 24518-24526.

Gao, X., McLean, H., Caveney, S. and Donly, C. (1999). Molecular cloning and functional characterization of a GABA transporter from the CNS of the cabbage looper, *Trichoplusia ni*. *Insect Biochem. Mol. Biol.* **29**, 609-623.

Giros, B., Jaber, M., Jones, S. R., Wightman, R. M. and Caron, M. G. (1996). Hyperlocomotion and indifference to cocaine and amphetamine in mice lacking the dopamine transporter. *Nature* **379**, 606-612.

Gomeza, J., Hulsmann, S., Ohno, K., Eulenburg, V., Szoke, K., Richter, D. and Betz, H. (2003). Inactivation of the glycine transporter 1 gene discloses vital role of glial glycine uptake in glycinergic inhibition. *Neuron* **40**, 785-796.

Greer, C. L., Grygoruk, A., Patton, D. E., Ley, B., Romero-Calderon, R., Chang, H. Y., Houshyar, R., Bainton, R. J., Diantonio, A. and Krantz, D. E. (2005). A splice variant of the *Drosophila* vesicular monoamine transporter contains a conserved trafficking domain and functions in the storage of dopamine, serotonin, and octopamine. *J. Neurobiol.* **64**, 239-258.

Harvey, W. R. and Wieczorek, H. (1997). Animal plasma membrane energization by chemiosmotic H^+ V-ATPases. *J. Exp. Biol.* **200**, 203-216.

Heller-Stilb, B., van Roeyen, C., Rascher, K., Hartwig, H. G., Huth, A., Seeliger, M. W., Warskulat, U. and Haussinger, D. (2002). Disruption of the taurine transporter gene (*taut*) leads to retinal degeneration in mice. *FASEB J.* **16**, 231-233.

Hoffman, B. J., Hansson, S. R., Mezey, E. and Palkovits, M. (1998). Localization and dynamic regulation of biogenic amine transporters in the mammalian central nervous system. *Front. Neuroendocrinol.* **19**, 187-231.

Holt, R. A., Subramanian, G. M., Halpern, A., Sutton, G. G., Charlab, R., Nusskern, D. R., Wincker, P., Clark, A. G., Ribeiro, J. M., Wides, R. et al. (2002). The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science* **298**, 129-149.

Homberg, U., Kingan, T. G. and Hildebrand, J. G. (1987). Immunocytochemistry of GABA in the brain and suboesophageal ganglion of *Manduca sexta*. *Cell Tissue Res.* **248**, 1-24.

Huang, X., Huang, Y., Chinnappan, R., Bocchini, C., Gustin, M. C. and Stern, M. (2002). The *Drosophila* inebriated-encoded neurotransmitter/osmolyte transporter: dual roles in the control of neuronal excitability and the osmotic stress response. *Genetics* **160**, 561-569.

Johnson, K., Knust, E. and Skaer, H. (1999). bloated tubules (blot) encodes a *Drosophila* member of the neurotransmitter transporter family required for organisation of the apical cytocortex. *Dev. Biol.* **212**, 440-454.

King, D. G. (1988). Cellular organization and peritrophic membrane formation in the cardia (proventriculus) of *Drosophila melanogaster*. *J. Morphol.* **196**, 253-282.

Kitamoto, T., Wang, W. and Salvaterra, P. M. (1998). Structure and organization of the *Drosophila* cholinergic locus. *J. Biol. Chem.* **273**, 2706-2713.

Kleta, R., Romeo, E., Ristic, Z., Ohura, T., Stuart, C., Arcos-Burgos, M., Dave, M. H., Wagner, C. A., Camargo, S. R., Inoue, S. et al. (2004). Mutations in SLC6A19, encoding B0AT1, cause Hartnup disorder. *Nat. Genet.* **36**, 999-1002.

Kosaka, T. and Ikeda, K. (1983). Reversible blockage of membrane retrieval and endocytosis in the garland cell of the temperature-sensitive mutant of *Drosophila melanogaster*, shibirets1. *J. Cell Biol.* **97**, 499-507.

Lill, H. and Nelson, N. (1998). Homologies and family relationships among Na^+/Cl^- neurotransmitter transporters. *Meth. Enzymol.* **296**, 425-436.

Mahajan-Miklos, S. and Cooley, L. (1994). Intercellular cytoplasm transport during *Drosophila* oogenesis. *Dev. Biol.* **165**, 336-351.

Malandro, M. S. and Kilberg, M. S. (1996). Molecular biology of mammalian amino acid transporters. *Annu. Rev. Biochem.* **65**, 305-336.

Malutan, T., McLean, H., Caveney, S. and Donly, C. (2002). A high-affinity octopamine transporter cloned from the central nervous system of cabbage looper *Trichoplusia ni*. *Insect Biochem. Mol. Biol.* **32**, 343-357.

Mbungu, D., Ross, L. S. and Gill, S. S. (1995). Cloning, functional expression, and pharmacology of a GABA transporter from *Manduca sexta*. *Arch. Biochem. Biophys.* **318**, 489-497.

Melikian, H. E. (2004). Neurotransmitter transporter trafficking: endocytosis, recycling, and regulation. *Pharmacol. Ther.* **104**, 17-27.

Melzig, J., Burg, M., Gruhn, M., Pak, W. L. and Buchner, E. (1998). Selective histamine uptake rescues photo- and mechanoreceptor function of histidine decarboxylase-deficient *Drosophila* mutant. *J. Neurosci.* **18**, 7160-7166.

Miller, A. (1950). *Biology of Drosophila*. New York: John Wiley & Sons.

Monastirioti, M. (1999). Biogenic amine systems in the fruit fly *Drosophila melanogaster*. *Microsc. Res. Tech.* **45**, 106-121.

Monastirioti, M., Gorczyca, M., Rapus, J., Eckert, M., White, K. and Budnik, V. (1995). Octopamine immunoreactivity in the fruit fly *Drosophila melanogaster*. *J. Comp. Neurol.* **356**, 275-287.

Mullen, G. P., Mathews, E. A., Saxena, P., Fields, S. D., McManus, J. R., Moulder, G., Barstead, R. J., Quick, M. W. and Rand, J. B. (2006). The *Caenorhabditis elegans* snf-11 gene encodes a sodium-dependent GABA transporter required for clearance of synaptic GABA. *Mol. Biol Cell* **17**, 3021-3030.

Nash, S. R., Giros, B., Kingsmore, S. F., Kim, K. M., el-Mestikawy, S., Dong, Q., Fumagalli, F., Seldin, M. F. and Caron, M. G. (1998). Cloning, gene structure and genomic localization of an orphan transporter from mouse kidney with six alternatively-spliced isoforms. *Recept. Channels* **6**, 113-128.

Nassel, D. R. and Elekes, K. (1992). Aminergic neurons in the brain of blowflies and *Drosophila*: dopamine- and tyrosine hydroxylase-immunoreactive neurons and their relationship with putative histaminergic neurons. *Cell Tissue Res.* **267**, 147-167.

Nelson, N. (1998). The family of Na^+/Cl^- neurotransmitter transporters. *J. Neurochem.* **71**, 1785-1803.

Nowicki, J. L. and Burke, A. C. (2000). *Hox* genes and morphological identity: axial versus lateral patterning in the vertebrate mesoderm. *Development* **127**, 4265-4275.

Parisi, M., Nuttall, R., Edwards, P., Minor, J., Naiman, D., Lu, J., Doctolero, M., Vainer, M., Chan, C., Malley, J. et al. (2004). A survey of ovary-, testis-, and soma-biased gene expression in *Drosophila melanogaster* adults. *Genome Biol.* **5**, R40.

Peter, D., Liu, Y., Sternini, C., de Giorgio, R., Brecha, N. and Edwards, R. H. (1995). Differential expression of two vesicular monoamine transporters. *J. Neurosci.* **15**, 6179-6188.

Pollack, I. and Hofbauer, A. (1991). Histamine-like immunoreactivity in the visual system and brain of *Drosophila melanogaster*. *Cell Tissue Res.* **266**, 391-398.

Porzgen, P., Park, S. K., Hirsh, J., Sonders, M. S. and Amara, S. G. (2001). The antidepressant-sensitive dopamine transporter in *Drosophila melanogaster*: a primordial carrier for catecholamines. *Mol. Pharmacol.* **59**, 83-95.

Quan, H., Athirakul, K., Wetsel, W. C., Torres, G. E., Stevens, R., Chen, Y. T., Coffman, T. M. and Caron, M. G. (2004). Hypertension and impaired glycine handling in mice lacking the orphan transporter XT2. *Mol. Cell Biol.* **24**, 4166-4173.

Richardt, A., Rybak, J., Stortkuhl, K. F., Meinertzhagen, I. A. and Hovemann, B. T. (2002). Ebony protein in the *Drosophila* nervous system: optic neuropile expression in glial cells. *J. Comp. Neurol.* **452**, 93-102.

Saint Marie, R. L. and Carlson, S. D. (1983). The fine structure of neuroglia in the lamina ganglionaris of the housefly, *Musca domestica* L. *J. Neurocytol.* **12**, 213-241.

Sandhu, S. K., Ross, L. S. and Gill, S. S. (2002). Molecular cloning and functional expression of a proline transporter from *Manduca sexta*. *Insect Biochem. Mol. Biol.* **32**, 1391-1400.

Seow, H. F., Broer, S., Broer, A., Bailey, C. G., Potter, S. J., Cavanaugh, J. A. and Rasko, J. E. (2004). Hartnup disorder is caused by mutations in the gene encoding the neutral amino acid transporter SLC6A19. *Nat. Genet.* **36**, 1003-1007.

Sloan, J. L. and Mager, S. (1999). Cloning and functional expression of a human Na(+) and Cl(–)-dependent neutral and cationic amino acid transporter B(0+). *J. Biol. Chem.* **274**, 23740-23745.

Soehnge, H., Huang, X., Becker, M., Whitley, P., Conover, D. and Stern, M. (1996). A neurotransmitter transporter encoded by the *Drosophila inebriated* gene. *Proc. Natl. Acad. Sci. USA* **93**, 13262-13267.

Soragna, A., Mari, S. A., Pisani, R., Peres, A., Castagna, M., Sacchi, V. F. and Bossi, E. (2004). Structural domains involved in substrate selectivity in two neutral amino acid transporters. *Am. J. Physiol.* **287**, C754-C761.

Stuart, A. E., Morgan, J. R., Mekeel, H. E., Kempter, E. and Callaway, J. C. (1996). Selective, activity-dependent uptake of histamine into an arthropod photoreceptor. *J. Neurosci.* **16**, 3178-3188.

Stuart, A. E., Gebhardt, K. A., Vogel, S. N. and Rodriguez, O. (2002). Does the neurotransmitter transporter underlie adaptation at a histaminergic photoreceptor synapse? *Vis. Neurosci.* **19**, 307-319.

Takanaga, H., Mackenzie, B., Suzuki, Y. and Hediger, M. A. (2005). Identification of mammalian proline transporter SIT1 (SLC6A20) with characteristics of classical system imino. *J. Biol. Chem.* **280**, 8974-8984.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. and Higgins, D. G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**, 4876-4882.

Tix, S., Eule, E., Fischbach, K. F. and Benzer, S. (1997). Glia in the chiasms and medulla of the *Drosophila melanogaster* optic lobes. *Cell Tissue Res.* **289**, 397-409.

Tomancak, P., Beaton, A., Weiszmann, R., Kwan, E., Shu, S., Lewis, S. E., Richards, S., Ashburner, M., Hartenstein, V., Celniker, S. E. et al. (2002). Systematic determination of patterns of gene expression during *Drosophila* embryogenesis. *Genome Biol.* **3**, RESEARCH0088.

Tsai, G., Ralph-Williams, R. J., Martina, M., Bergeron, R., Berger-Sweeney, J., Dunham, K. S., Jiang, Z., Caine, S. B. and Coyle, J. T. (2004). Gene knockout of glycine transporter 1, characterization of the behavioral phenotype. *Proc. Natl. Acad. Sci. USA* **101**, 8485-8490.

Umesh, A. and Gill, S. S. (2002). Immunocytochemical localization of a *Manduca sexta* gamma-aminobutyric acid transporter. *J. Comp. Neurol.* **448**, 388-398.

Valles, A. M. and White, K. (1988). Serotonin-containing neurons in *Drosophila melanogaster*: development and distribution. *J. Comp. Neurol.* **268**, 414-428.

Wang, J., Kean, L., Yang, J., Allan, A. K., Davies, S. A., Herzyk, P. and Dow, J. A. (2004). Function-informed transcriptome analysis of *Drosophila* renal tubule. *Genome Biol.* **5**, R69.

Xiong, W. C., Okano, H., Patel, N. H., Blendy, J. A. and Montell, C. (1994). repo encodes a glial-specific homeo domain protein required in the *Drosophila* nervous system. *Genes Dev.* **8**, 981-994.

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

Yellman, C., Tao, H., He, B. and Hirsh, J. (1997). Conserved and sexually dimorphic behavioral responses to biogenic amines in decapitated *Drosophila*. *Proc. Natl. Acad. Sci. USA* **94**, 4131-4136.

Zomot, E. and Kanner, B. I. (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.

```
                        *       20        *       40        *       60        *       80
DmCG8291   : MSSKEQQAAGRDFKRNSNAYSSLPPTGTGAGCSGAALGSGTGTGKRARSKQHQLEHEQFGLSNSLSSESIRQACAYYDDE :   80

                        *      100        *      120        *      140        *      160
AmINE      : ------------------------------------MNATDSNSGVNGNDSNNTIAAKTNAHGNSVRLQSVIGKQSSC :   42
DmCG8291   : LSDEDLIEIQQTPQAFHQQTKQPLQLQPISFRLGDELGDERSAFCGSQEMQQVPLSTPIKQAAETDEALCVLSELDAILD :  160

                        *      180        *      200        *      220        *      240
AmINE      : VRKMSTMPEEEDASDDMDPDVIRIGNQHHYQHHHDADSDSPVEERGRLLSITDSTITAHHRKIAITPSSSSNLHHCNGLK :  122
MsIne      : ------------------------------------------------------------------------MNKVE :    5
DmCG8291   : VHDVSQLNGTCSSGSVNSGSDDDKVEDYLMDLDNYLEEMDNALNREDSLIIIDGHTSLKREPRTRTLPLSRKKKSSKKKS :  240

                        *      260        *      280        *      300        *      320
CeSERT     : ------------------------------------------------------------------------MLRWH :    5
TnOctT     : ----------------------------------------------------------------------MATLSGTG :    8
DmCG5549   : --------------------------------------------------------------------MGQSSKSGSAGVG :   13
HsGLYT2    : ------------------------------------------------------------MDCSAPKEMNKLPANSPEAAA :   21
HsATB0+    : ---------------------------------------------------------------MDKLKCPS-------- :    8
HsGLYT1    : ------------------------------------------------------------MSGGDTRAAIARPRMAAAHGPV :   22
AgCP4566   : ---------------------------------------------------------------------------Q :    1
AmINE      : KAIPRNDSGNSIHERAYSRLQQSRSGDQVGQPEARLLNRDNPQGILRSSARGIDSHSGGRYRCSSVRSVKSNPTMDEHHH :  202
MsIne      : SSTEAAAPSVAIHVEQHD--------DEQDKENSKLLSAHSPAPSITP--------SGQMRKVKSFSDTHKIRDVTTASG :   69
DmCG8291   : EDQEAAQGDFQREHQLRKTFSCSLRPTSQIASSSGSLETSTEPVMDVWRRQSMRRALEQEDTKATVAMEREEDDIPTLLV :  320
DmBLOT     : ----------------------------------------MSSSYKPRSSDEALAEQFSLARAGSGTKPPATPTYRQI :   38
CeSNF12    : ---------------------------------------MNGEWKSALRLQIEALAKRNELHRKSSVDEIKKRT :   35
HsNTT5     : ---------------------------------------------------------------------MKTEAQ :    6

                        *      340        *      360        *      380        *      400
DmSERT     : ----------------MDRSGSSDFAGAAATTGRSNPAPWSDDKESPNNEDDSNEDDGDHTTPAKVTDPLAPK--LANNE :   62
AgCP12911  : ------------------------------------------------------------------------------E :    1
MsSERT     : ------------------------------------------------------MPPSDAPPAPTAPPPDLPAT--TAQKS :   25
HsSERT     : ----------------METTPLNSQKQLSACEDGEDCQENGVLQKVVPTPGDKVESGQISNGYSAVPSPGAGDD--TRHSI :   63
CeSERT     : SVRRKQHQQLQAELSSGAASMLSAPESRRVSRSMSVKAPTASEYMPLSVADKPLTLTVSTSHSIDPNEPIAALGGLTPTK :   85
DmDAT      : ----------------------------------------------------------------MSPTGHISKSK :   11
AgCP3649   : ------------------------------------------------------------------------------ :    -
TnDAT      : ---------------------------------------------------------------------------MALK :    4
CeDAT1     : ------------------------------------------------------MQLVPTDDPDEKIGRTSNGMQNAT :   24
HsNET      : ----------------MLLAR------------------MNPQVQPENNGADTGPEQPLRARKTAELLVVKERNGVQC-- :   44
HsDAT      : ------------------MSK------------------SKCSVGLMSSVVAPAKEPNAVGPKEVELILVKEQNGVQLTS :   44
TnOctT     : VVTHTAPSYEEQRANPALLSRGASGTPGGRSVRDDGYCSASSTPRAFDNKSTKGSVVTLSCYKKEPKIQIEEECFYSEPN :   88
DmCG5549   : G-----------------------------------------------------------VGGGGGVVCPSTAIS :   29
AgCP10735  : ------------------------------------------------------------------------------ :    -
HsGLYT2    : AQGHPDGPCAPRTSPEQELPAAAAPPPPRVPRSASTGAQTFQSADARACEAERPGVGSCKLSSPRAQAASAALRDLREAQ :  101
HsATB0+    : ---------------------------------------------------------------------FFKCREKE :   16
HsPROT     : ------------------------------------------------------------MKKLQGAH---LRKPVTPD :   16
HsGLYT1    : APSSPEQVTLLPVQRSFFLPPFSGATPSTSLAES-------------------------VLKVWHGAY----NSGLLPQ :   72
DmCG7075   : ---------------------------------------------------------------------MANNQPPT :    8
AgEbiP6595 : ------------------------------------------------------------------------------ :    -
MsPROT     : ------------------------------------------------------------------------------ :    -
DmIne      : --------------------MPNRQDYDAQSSK--------------------HSEQSFFRFNR-VAKVPLGVDDDDSY :   38
AgCP4566   : QPQTTIARSGSYIFSDVG--GTPRLFSDATSIRSLASIGMGSTDGRRMVIRRVPNSPNELLTMIN-PPTPPDETYENESY :   78
AmINE      : QYHPPAASWASIVFSDGNGQTHVRTLPDSVSIRSLASIGMGSSNGRKLTIRRVLTSPSELLNMVHPPPSFEDDLSVCTSF :  282
MsIne      : AASARSLRPYEIVNTYPEGSESGTNNYGAPSVRSLASIGMGCTDGRKMVIRRVPTSPTELFHLVR-PPTPPDEDSASHES :  148
TnGAT      : ----------------------------------------MDTKNDG---------RSDDIELS----------AQGTSNK :   22
MsGAT      : ----------------------------------------METKNDS---------RSDDIELS----------AQGSGNK :   22
AgCP8499   : ------------------------------------------------------------------------------ :    -
DmCG1732   : ----------------------------------------MYTNSASDGDGGGDVCKHESIEMSKELGHT--TSLQSSSPT :   39
CeSNF11    : ------------------------------------------------------MCHYNLIHPAH-------WFFFLQFFR :   20
HsGAT1     : ----------------------------------------MATNGSKVADG----QISTEVSEAP----------VANDKPK :   28
HsGAT3     : --------------------------MTAEKALPLGNGKAAEEARESEAPGGGCSSGGAAPA------------RHPR :   40
HsGAT2     : ------------------------------------------MDSRVS----GTTSNGETKPV--------------YPV :   20
HsBGT1     : ------------------------------------------MDGKVAVQERGPPAVSWVPEE--------------GEK :   24
HsTAUT     : ------------------------------------------MATKEKLQCLKDFHKDILKPSP-----------GKSPG :   27
HsCT1      : ------------------------------MAKKSAENGIYSVSGDEKKGPLIAPGPDGAPAK------------GDGPV :   38
CeSNF3     : ------------------------------------------------------MGTSEHVPLPT----------DEAKAK :   17
DmCG5226   : ----------------------------------------MAATKIIDAP---RNGHEMAPLNTRARGDGTHGVTIVLTAP :   38
AgCP7001   : ----------------------------------------MTTKVDGVPGPRNNGHEMAPLNTRARGDGTHGVTIMLSPP :   40
HsNTT7-3   : ----------------------------------MPKNSKVVKRELD-DDVTESVKDLLSNEDAADDAFKTSELIVDGQE :   45
HsNTT4     : ----------------------------------MPKNSKVTQREHSSEHVTESVADLLALEEPVD--YKQSVLNVAGEA :   44
HsXtrp2    : ------------------------------------------------------MAHAPEPDPAAC-------------- :   12
RnROSIT    : ------------------------------------------------------------------------------ :    -
HsBOAT1    : ----------------------------------------MVRLVLPNPGLDARIPSLAELETIEQ-------------- :   26
HsXT3L     : ------------------------------------------------------------------------------ :    -
DmCG10804  : ------------------------------------------MSDAFDESSELEQDPPPFGSEASTSKAKAELVAIATA :   37
AgCP7501   : ------------------------------------------------------------------------------ :    -
DmCG1698   : ----------METTINNNNNDNGYMNQAYVGSLNASTISIPG---------AYKPNESSNGKKTLSEGSGSAPPPGHGD :   60
DmCG4476   : ------------------------MKDLKT------EILVSG---------GLPKLETS---RCAGSGEALLKPSLLPM :   37
DmCG15279  : ----------MDT--TQRN----HQNLAFVGDDGRSTAST---------------VEISTN-------SPALRDN--SD :   39
AgCP10412  : ----------MEG--RDNNGFIGDNSPSIAGQYRWTTPAAPNGVHVTHGGLALTDVELAVGGKTVPRPTPATGED--AG :   65
DmCG8850   : ----------------------MQGAQGSREGHPGSNDG-----------------------ISTVIYSAEGEEL :   30
DmCG15088  : ------------------------MTQLEISKK------------------------------------------- :    9
AgCP10401  : ---------------MATSNPAFENDEPVLTSDRTSHTATKPS----------LAAVTAKSSAKVNQFQQPPEQTTGGNT :   55
AgCP10503  : ------------------------------------------------------------------------------ :    -
AgCP10610  : ------------------------------------------------------------------------------ :    -
DmCG3252   : --------------------------MELKGVQPS-------------------------NGSSNGSGNGATNAA :   24
AeAAT      : -------------------------MPEIATISYPE--------------------SKKNDEANSSHGNGNGVVQLN :   32
MsKAAT1    : ------------------------MNDGQVNGGFESSEPKMEP-------------------KRSSQISLPPANLKATMD :   37
MsCAATCH1  : ------------------------MNDGQVNGGFESSEPKMEP-------------------KRSSQISLPPANNKAALD :   37
BcProT     : ------------------------------------------------------------------------------ :    -
```

```
HinSnfT    : ------------------------------------------------------------------------------- :    -
StTnaT     : ------------------------------------------------------------------------------- :    -
CeSNF2     : ------------------------MPTPVSTKKTIAPSTMKTTVAPPTMRTTMAPTTMKTTVAPPTMKTTRSPSTMKTT :   55
CeSNF7     : -----------------------------MAQSNALTTTTTGSSAPSVIPPTAPTASVKTKAP-----TKGP--VQPQ :   42
CeSNF4     : ------------------------MSTSNPDNSTDSPPDVSKTGDTYKSGTQTTNTDSLETKLSDDGPKKTVDLSLIPT :   55
CeSNF5     : ------------------------MADSGSNEEAMKRQAPSVKFDSPSKPDPQQVSQQSNQLSS----QKSIQQSTQSK :   51
CeSNF9     : -------------------------MTSP------TQRRSPS---DSSAAEDQRIATTIVRDSGS----SKEIEE---SQ :   39
DmCG13794  : ------------------------------------------------------------------------MIYETS :    6
DmCG13795  : ---------------------------------------------------------------------MVYETS :    6
DmCG13793  : ------------------------------------------------------------------------------- :    -
DmCG33296  : ------------------------------------------------------------------------------- :    -
DmCG8291   : ELPPRRDAEMRRCFSQGDCQASVAPTVGSQMLTEAHIFDNLLQTNARASSEEPRPRQYGRRLEGPPTPVRPLILAQSRPQ :  400
AgCP9955   : ------------------------------------------------------------------------------- :    -
DmBLOT     : SGAMATLPRHDNNHNHNDNSSRIRSDSTATSRTCVYIGGAPTGGGNGNEQISVTLDRRQPQRMSWLNMRRAKANETDLPT :  118
AgCP2093   : ------------------------------------------------------------------------------- :    -
CeSNF6     : ----------------------------------MSVSSNDPEQRNGRGMASGNNVDMSLYPPFIKQDAKLPDYTREGD :   45
CeSNF12    : VDMDKRIVKLRELVGSSANDAALFYLECVCHADETERLLNRGSSVGKEKKWKKVKRKTSSSVAPPLSRTISSLPGVSSPD :  115
AaLeuT     : ------------------------------------------------------------------------------- :    -
HsNTT5     : PSTSLLANTSWTGTVISDSVPGSQTWEDKGSLTRSATSWTSEAQVSAARVAEAQARTSQPKQISVLEALTASALNQKPTH :   86
DmCG13796  : -----------MKGICGEQLNNCMLLGDCENSEHGTELPGDFIGEPVETESGGGMGGSLALAARRIRSRQVHSMAVRTGS :   69
HspSnfT    : ------------------------------------------------------------------------------- :    -
CeSNF1     : ---------------------------------------------------------------------MWFEESSWSN :   10
CeSNF8     : -----------------------------------------------------------------------------MKPK :    4

                          *       420         *       440         *       460         *       480
DmSERT     : RILVVSV-------------------------------------------------------------------------- :   69
AgCP12911  : RSLVVAL-------------------------------------------------------------------------- :    8
MsSERT     : RSVVVSL-------------------------------------------------------------------------- :   32
HsSERT     : PATTTTLVAE----------------------------------------------------------------------- :   73
CeSERT     : EGRVAALRR------------------------------------------------------------------------ :   94
DmDAT      : TPTPHDNDNN----------------------------------------------------------------------- :   21
AgCP3649   : ------------------------------------------------------------------------------- :    -
TnDAT      : TPTPG---------------------------------------------------------------------------- :    9
CeDAT1     : LPIDGPVNTE----------------------------------------------------------------------- :   34
HsNET      : --LLAPRDG------------------------------------------------------------------------ :   51
HsDAT      : STLTNPRQSP----------------------------------------------------------------------- :   54
TnOctT     : KRLRTNSIKT----------------------------------------------------------------------- :   98
DmCG5549   : GCASGT-----------------------ALIQLDATGPGAGG------------------------------------- :   49
AgCP10735  : ------------------------------------------------------------------------------- :    -
HsGLYT2    : GAQASPPPGSSGPGNALHCKIPSLRGPEGD----ANVSVGKGTLERNNTPVVGWVNMSQSTVVLGTDGITSVLPGSVATV :  177
HsATB0+    : KVSASS---------------------------ENFHVGE---------------------------------------- :   29
HsPROT     : LLMTPS---------------------------DQGDVDLDV------------------------------------- :   31
HsGLYT1    : LMAQHSLAMA-----------------------QNGAVPSEAT------------------------------------- :   92
DmCG7075   : TNATRKDKRI----------------------------------------------------------------------- :   18
AgEbiP6595 : ------------------------------------------------------------------------------- :    -
MsPROT     : ------------------------------------------------------------------------------- :    -
DmIne      : LDMSDET-------------------------------------------------------------------------- :   45
AgCP4566   : GGMSDDSD-------------------------------------------------------------------------- :   86
AmINE      : DDAEDDAQD------------------------------------------------------------------------ :  291
MsIne      : DCEEEEED------------------------------------------------------------------------- :  156
TnGAT      : TSEVATKS------------------------------------------------------------------------- :   30
MsGAT      : PSDVAVKS------------------------------------------------------------------------- :   30
AgCP8499   : ------------------------------------------------------------------------------- :    -
DmCG1732   : TTDTSNKQQL----------------------------------------------------------------------- :   49
CeSNF11    : VIMVEATS-------------------------------------------------------------------------- :   28
HsGAT1     : TLVVKVQK-------------------------------------------------------------------------- :   36
HsGAT3     : VKRDK----------------------------------------------------------------------------- :   45
HsGAT2     : MEKKEED--------------------------------------------------------------------------- :   27
HsBGT1     : LDQEDED--------------------------------------------------------------------------- :   31
HsTAUT     : TRPEDEAEG------------------------------------------------------------------------- :   36
HsCT1      : GLGTPGGRL------------------------------------------------------------------------- :   47
CeSNF3     : ELEQSQHS-------------------------------------------------------------------------- :   25
DmCG5226   : QRNSVQSVD------------------------------------------------------------------------- :   47
AgCP7001   : DRSSIGSREGA----------------------------------------------------------------------- :   51
HsNTT7-3   : E-KDTDV--------------------------------------------------------------------------- :   51
HsNTT4     : GGKQKAV--------------------------------------------------------------------------- :   51
HsXtrp2    : ------------------------------------------------------------------------------- :    -
RnROSIT    : ------------------------------------------------------------------------------- :    -
HsBOAT1    : ------------------------------------------------------------------------------- :    -
HsXT3L     : ------------------------------------------------------------------------------- :    -
DmCG10804  : SADQQDI--------------------------------------------------------------------------- :   44
AgCP7501   : ------------------------------------------------------------------------------- :    -
DmCG1698   : SPSGV----------------------------------------------------------------------------- :   65
DmCG4476   : ALESS----------------------------------------------------------------------------- :   42
DmCG15279  : DQEAA----------------------------------------------------------------------------- :   44
AgCP10412  : RVPTT----------------------------------------------------------------------------- :   70
DmCG8850   : TINCE----------------------------------------------------------------------------- :   35
DmCG15088  : ------------------------------------------------------------------------------- :    -
AgCP10401  : SRTSALPEMNL----------------------------------------------------------------------- :   66
AgCP10503  : ------------------------------------------------------------------------------- :    -
AgCP10610  : ------------------------------------------------------------------------------- :    -
DmCG3252   : STEKT----------------------------------------------------------------------------- :   29
AeAAT      : AS------------------------------------------------------------------------------- :   34
MsKAAT1    : NIDDM---------------------------------------------------------------------------- :   42
MsCAATCH1  : NIDDT---------------------------------------------------------------------------- :   42
BcProT     : ------------------------------------------------------------------------------- :    -
HinSnfT    : ------------------------------------------------------------------------------- :    -
StTnaT     : ------------------------------------------------------------------------------- :    -
CeSNF2     : KEPSMTQTTKTVQSTITNPVSFLDS---------------------------------------EVTAKEYVKDGGAAR :   95
```

```
CeSNF7     : KKP----------TTVDN-------------------------------------------------------------------GR :   52
CeSNF4     : VSQNMEKDKD--NTTLDE-----------------------------------------------------------DRKFDTGVGIA :   82
CeSNF5     : IDPSRIDTKN---TTVTT-----------------------------------------------------------------TR :   68
CeSNF9     : ADPC-------------------------------------------------------------------------------- :   43
DmCG13794  : YDSGHKPFS---------------------------------------------------------------------------- :   15
DmCG13795  : YDTGHKPFS---------------------------------------------------------------------------- :   15
DmCG13793  : ---------------------------------------------------------------------------------- :    -
DmCG33296  : ---------------------------------------------------------------------------------- :    -
DmCG8291   : SAPTRVQMREPQLQDTPTHPIMSTCSELSSARSSRMPSPVSLPSDSSSSGSSSAEHDQEPDPVQTTTMCSASSTTPLEPL :  480
AgCP9955   : ---------------------------------------------------------------------------------- :    -
DmBLOT     : VTPSPQTTRSCISTVSSVVDSGGGRTTATSGRISSSGIVTLSGSNNTLSEIQGGYHDMDHDAVSKNFSISASAHNITTAS :  198
AgCP2093   : ---------------------------------------------------------------------------------- :    -
CeSNF6     : IEYPFEEITG--------------------------------------------------------------------------- :   55
CeSNF12    : VIQTIDVSALEDQTP-------------------------------------------------------QRPHWWDQFQLYR :  143
AaLeuT     : ---------------------------------------------------------------------------------- :    -
HsNTT5     : EKVQMTEKKES------------------------------------------------------------------------- :   97
DmCG13796  : AYDTLERP---------------------------------------------------------------------------- :   77
HspSnfT    : ---------------------------------------------------------------------------------- :    -
CeSNF1     : LRLTD------------------------------------------------------------------------------- :   15
CeSNF8     : KEVR-------------------------------------------------------------------------------- :    8

                         *       500         *       520         *       540         *       560
DmSERT     : --------T-ERT ET GQ-KAE   AV GFA D G       ICYQN  GA     C F IFG    FY  LA   FHR :  139
AgCP12911  : --------TGERR ET SQ-KAE   AV GFA D G       ICYQN  GA     C M LFG    FY  LA   FHR :   79
MsSERT     : --------TPARQ ET AK-KAE   AV GFA D G       ICYQN  GA     C M LFG    FF  LA   YHR :  103
HsSERT     : --------LHQGE ET GK-KVD   SV GYA D G       ICYQN  GA     T MAIFG    FY  LA   YHR :  144
CeSERT     : --------RSSMV DK AT-KME   AV GYA D G      SVCYKH  GA     F M MIG    FY  LV   FHR :  165
DmDAT      : --------SISDE ET SG-KVD   SV GFA D A       LCYKN  GA     G M VVG    FY  LA   HNR :   92
AgCP3649   : ----------EE ET SG-KVD   SV GFA D A       LCYKN  GA     G M VVG    FY  LA   FNR :   68
TnDAT      : ----------VGE ET GK-KVD   SV GFA D A       LCYKN  GA     C M VVG    FY  LA   FHR :   78
CeDAT1     : --------PKDPA EQ SG-KLD   SV GFA D G       LCFKN  GV     S M LLT    FY  LC   YYR :  105
HsNET      : ---------DAQP ET GK-KID   SV GFA D A       LCYKN  GA     T F IIA    FY  LA   YNR :  121
HsDAT      : --------VEAQD ET GK-KID   SV GFA D A       LCYKN  GA     L F VIA    FY  LA   FNR :  125
TnOctT     : --------EADDG ET GT-GAD   SI GFA D A       LCYRN  GA     T M VFGA   FY  LI   YNR :  169
DmCG5549   : ------AGESEHE GT TG-RFD   SL GYS G G       LCYNN  GA     T M VIA    MF  LSF  YAA :  122
AgCP10735  : ----------K GS TG-RYD   SL GYS G G       LCYSN  GA     T M VIA    MF  LS A YVG :   67
HsGLYT2    : ATQEDEQGDENKA GN SS-KLD   SM GYA G G       LAFQN  GA     L M ALA    FF  VS   FAS :  256
HsATB0+    : -------NDENQD GN SK-KSD   SM GYA G G       LTYSN  GA     A M ALA    FF  CS   FAS :  101
HsPROT     : --------DFAAH GN TG-KLD   SC GYC G G       RAYTN  GA     F M AIC    FF  LS   FSS :  102
HsGLYT1    : ------KRDQNLK GN GN-QIE   TS GYA G G       LCYRN  GA  F  F M IFC    FF  LSF  FAS :  165
DmCG7075   : --------ERDEN GQ KS-KSE   SL GYA G G       LCYRS  AA     L M ILA    FY  IL   FSS :   89
AgEbiP6595 : ------------------------------------------A     L M VLC    FF  TC   FSG :   29
MsPROT     : --------MSGATQDR GS-QLE   SC GYA G G       LCYRN  GA     F T IIC    VY  TT   FAS :   71
DmIne      : -------AQLKPRQQH AN-KMQ   AC GYS G G       MCYKS  GV     C I FICS   LF  LS   YTG :  117
AgCP4566   : ------LDNLKPR QH AN-KMQ   AC GYS G G       LCYKS  GV     F I LIC    LF  LA   YTG :  159
AmINE      : -----PGDYRQSR PH AN-KMQ   AC GYS G G       LCYKS  GV     F I IVC    LY  LS   FTR :  365
MsIne      : -----AAVHLKPR PF AN-KIQ   AC GYS G G       LCYKS  GA     F I LIC    LF  LA   YTA :  230
TnGAT      : ---------DLPE GS GS-KLD   SV GLA G G       LCYKN  GA     F T FLA    FF  LA   MLT :  100
MsGAT      : ---------NLPE GS AS-KLD   SV GLA G G       LCYKN  GA     F T FLA    FF  LA   MLT :  100
AgCP8499   : --------------------LD   SV GLA G G       LCYKN  GA     F T FLA    FF  LA   MLT :   59
DmCG1732   : -----VKIEQLPD GS SS-KMD   SV GLA G G       LCYKN  GA     I T FLA    FF  LA   MLT :  123
CeSNF11    : ----------VAE EQ SS-WAD   SC GYA G G       LCYQN  GA     C S VFC A  FI  TSW  LMS :   97
HsGAT1     : ------KAADLPD DT KG-RFD   SC GYA G G       LCGKN  GA     F T IFA    FL  CS   YTS :  109
HsGAT3     : ---------AVHE GH NN-KVE   SVAGEI G G       LCYKN  GA     V FFICC    FF  TA   FTS :  115
HsGAT2     : ---------GTLE GH NN-KME   SVAGEI G G       LCYKN  GA F   L F FTC    FL  TA   YTS :   97
HsBGT1     : ---------QVKD GQ TN-KME   SVAGEI G G       LCYKN  GA F   F FFFVC    FF  VA   YTS :  101
HsTAUT     : ---------KPPQ EK SS-KID   SVAGGF G G       LCYKN  GA     F F FGS    FF  II   YTS :  106
HsCT1      : ---------AVPP ET TR-QMD   SC GFA G G       LCYKN  GV     V IALVG    FF  IS   FMK :  117
CeSNF3     : ---------EEPD GQ TG-KFD   SM AYA G G       LCYKN  GS   V M FFCLAA   FL  VT   YLQ :   95
DmCG5226   : ------IPGNEPE AA SG-KMQ F SI GYS G G       LCQQN  GA     M M ILE    FL  LG   RMR :  120
AgCP7001   : ----GALPDGDPD AA SG-KMQ F SI GYS G G       LCQQN  GA     M M ILE    FL  LG   RMR :  126
HsNTT7-3   : ----EEGSEVEDE PA NS-KLQ   AQ GFS G G       LCQKN  GA     L L MVI    FF  LS   RIR :  126
HsNTT4     : ----EEELDAED- PA NS-KLQ   AQ GFS G G       LCQKN  GA     L L III    FF  LA   RIR :  125
HsXtrp2    : --------DLGDE PK DN-KAQ   SC GFA G G       LCQTY  GA     V A VFE    FH  LA   RLR :   83
RnROSIT    : ---------IEDE PK DN-KLQ   SC GFA G G       LCHTH  GA     F A VFE    FY  LA   RLR :   70
HsBOAT1    : --------EEASS PK DN-KAQ   TC GFC G G       LCQSH  GA     L L VLE    LY  FA   RLR :   97
HsXT3L     : ---------MEKA PL AN-SLQ  FAC SYA G G       LCQMY  GS     I M IVE    LY  LA   RMR :   70
DmCG10804  : ----VEGEKEGEE ES DS-KIM   AT GYA G G       LAQKN  GA     F M CIQ    FY  LA   RLR :  119
AgCP7501   : ----------DE ES DS-KWT   AT GYA G G       LAQKN  GA     F M LLQ    FY  LA   RLR :   68
DmCG1698   : -----EAGQPGKK DS NN-DIE   SC ALS G G       TALEN  GA     L V ILV K  YY  ML   FSS :  139
DmCG4476   : -----ELSPDRKA DN GS-SLE   SC ALS G G       TALEN  GA     L V FVV K  YY  ML   FSS :  116
DmCG15279  : -----KVPE---E AT GK-GVE   SC AMS G G       TALDN  GA     L V FLI K  YY  MV   FSS :  115
AgCP10412  : -----IIPEPNAD DQ GK-GVE   SC AMS G G       VALEN  GA     I V LLV K  YY  MI   FSS :  144
DmCG8850   : -----AESESSGQ DQ SR-GVE  FSC ALS G G       IALEN  GA     V V LLI R  YY  VI   FSS :  109
DmCG15088  : -----SEKPAEKPAEQ GN-GLE  FSC SLS G G       IAFQN  GT     L A LVI R  YY  IS   FTG :   83
AgCP10401  : -----MKNIHKVQ DK GK-DIE   SC ALS G G       TALEN  GA     L V LLV R  YY  ML S FSS :  140
AgCP10503  : ------------- EK TS-NIE T SC AYS GFG       TALDN  GA     L V FLI R  YY  MA   FCS :   66
AgCP10610  : ----------------------------------       TAFKN  GA     L V FII R  YY  MV   FSN :   44
DmCG3252   : -----DAEKPTAE TN GN-GLE   SC SVS G G       TAYEN  GA     I V FLI K  YY  MI   FTS :  103
AeAAT      : -----QPENAAQN PE SN-KLE   SC SMS G G       TAYEN  GA     L V FVI R  YY  MA   FTS :  108
MsKAAT1    : -----DLEAEPPE MV SN-NIE   SC ATS G G       IAYQN  GA     I V LLV K  YY  CV   FSS :  116
MsCAATCH1  : -----DLEAEPPE MV SN-NIE   SC ATS G G       IAYQN  GA     V V LLV K  YY  CV   FSS :  116
BcProT     : ----------MET QQ GT-RAG   FAA GSA G G       TAYEN  GA F   LFA LTT  S LAF FA  HRHR :   69
HinSnfT    : ---MAHSAPKAQK ET SG-RRA    AA GSA G G       TTYEN  GA     L A LTA    LF DYA  HRHR :   76
StTnaT     : ---------MEAQ DQ KS-RSG   FAT GAA G G F     MAYQN  GA     F A LTA    MI  FGF HKMR :   70
CeSNF2     : KAYRDAIQVENTD QA RG--IES   SS GQA G G      TVAYQN  TS   V I CAFVFA  AIH  FA   YAA :  173
CeSNF7     : RAYRNQVVVENVD QA RG--IESF SS GQA G G      TTAYKN  LS   A V CGILFA  AIH  FA   YAA :  130
CeSNF4     : DILRDQAKRAAMQ GSCETPVTNSANLC GMS G G      SRAYEN  SV     ITCALLF   GVY  FA   YQG :  162
CeSNF5     : PTLDTPPEEEEEK DG GN-SFE   TS GLA G G      TRAYNN  SA     LTCAFLF   AVYF FLT  YQG :  147
```

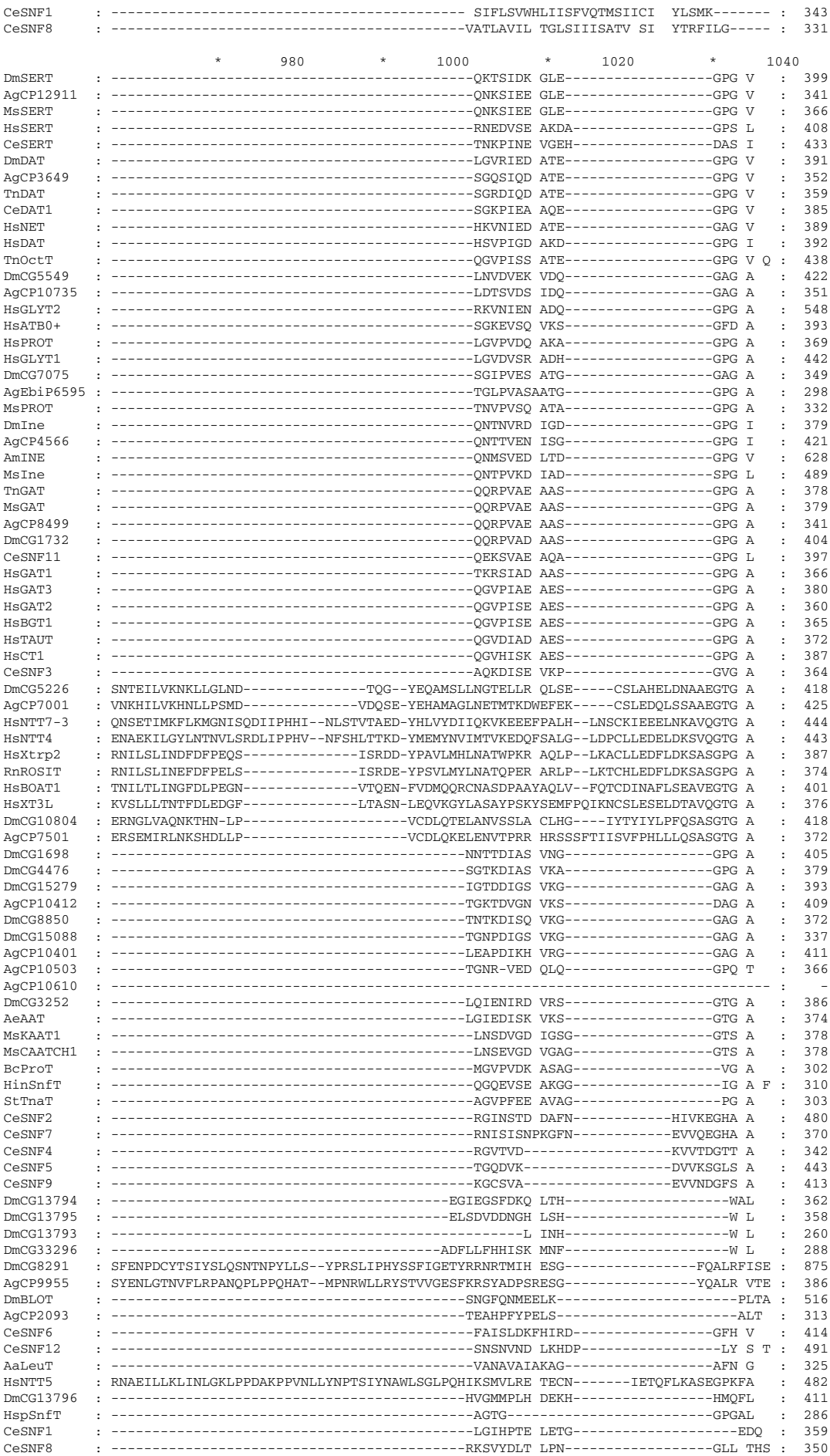
```
CeSNF9     : ------------- GA GN-QIE   ST GMA G G      TRAYNN  SA     VACALLF   AVY  FL   YHR :  109
DmCG13794  : ---------PDLK GK DK-PTD  FACFGLA K D-  VAS WFFFDM IFG    L Y VIYL   MV HSF   FSS :   84
DmCG13795  : ---------PDLN GK DK-PTD  FACFGLA K D-  VAS WMFFDM ILG    F Y SIYL   MV HSF   FSS :   84
DmCG13793  : ------------------------------------------------------------------MV HSY   FSS :   12
DmCG33296  : ------------------------------------------------------------ VIYL   LV HSF   FSS :   20
DmCG8291   : HQLQLLLREKCGFNSQ PH-AGSRT AL GCT G F  C  AVLTINF  N-   Q L LSVIF    LW  MC  AKIR :  558
AgCP9955   : --------------AR PH-AISRS ATTFCT G F  S  AVFSVHF AN-   Q L FSLLF    LW  MV  ARIR :   64
DmBLOT     : NAKLLPAVEDESN QTKCS-VFRG  LC CLN SYA  V   -RELDRY SA     V L FLV    VL  IS   FLG :  276
AgCP2093   : ----------RSLPQSRSS-IFRG  LC CLN TFA  V   -RELQLH SA     V L LLV    VL  IS   FLG :   68
CeSNF6     : ------VGDENRI GN SN-KSD   AV GFTAG GSF     LVFQH  AA     LCM CLAS   FF  MV   FSS :  128
CeSNF12    : RTDLLRFSKQNDR EL RT-QKD F SC GFM G GHTM   AKVYQH  GV F   LFS IFF    VF HLS   YTG :  222
AaLeuT     : ---------MEVK EH AT-RLGL  AMAGNA G G FL   VQAAEN  GA     I AFLLV    MW  WA  RYGG :   70
HsNTT5     : --------EVLLA PF SS-KTE   AQ GFS KPSC    A LWLNS  CS AAI IFM FLV    LF  MAA  SMR :  168
DmCG13796  : -------YRHDKC GR AKSADF FASCTHAFSS IFSELST GILHG WLL   A L G LFYS   FL  AF   FSS :  150
HspSnfT    : -----------MA DV HS-RLG   AA GSA G G       MTSEN  AA   V LGV LCV   GLLA FV  RRSR :   68
CeSNF1     : --------LVEEW DL PS-KIE   SA AYLFA T FLNL KLILEN  VA AAA A V GVLC  TMV  MS   VTG :   86
CeSNF8     : --------KDALP PE KS-WYDL FSV NLC G S FLI LAKVHEYR GA   A G I IML Y  LY  LI   FHR :   79

                        *       580         *       600         *       620         *       640
DmSERT     : C C SI KR C ALK ------  YAICLIDIY GM  NT  GWA Y LFA FT------SK   TS DNPWNTENCMQ- :  206
AgCP12911  : C C SI KR C ALK ------  YAICLIDIY GM  NT  GWA Y LFA FS------SE   TK GNPWNTENCSP- :  146
MsSERT     : C C TL KR C ALK ------  YAICMIDIY GM  NT  GWA Y LIA LASIN---SV   TS DNEWNTPLCTP- :  173
HsSERT     : N C SI RK C IFK ------  YAICIIAFY AS  NT  AWA Y LIS FT------DQ   TS KNSWNTGNCTN- :  211
CeSERT     : S C SI RK C LFR ------  YGICCICTF AI  NA  AQA Y AIV LSKIW--DSE  AS GNPWNTPRCSD- :  236
DmDAT      : K A TC GR V LFK ------  YAVVLIAFY DF  NV  AWS R FFA FT------NS   TS NNIWNTPNCRPF :  160
AgCP3649   : K A TC GR V LFK ------  YAVVLIAFY DF  NV  AWS R FFA FT------DS   TH SNAWNTVECKPF :  136
TnDAT      : K A TC GR V LFK ------  YAVVLIAFY DF  NV  AWA R FFA FT------TM   TN DNEWNTPACRPF :  146
CeDAT1     : K A TT GR C LFK ------  YCVILTAFY DF  NV  AWG H LYT FS------FN   AS NNSYNSPAC--Y :  171
HsNET      : E AATV K- C FFK ------  YAVILIALY GF  NV  AWS Y LFS FT------LN   TD GHTWNSPNCTDP :  188
HsDAT      : E AAGV K- C ILK ------  FTVILISLY GF  NV  AWA H LFS FT------TE   IH NNSWNSPNCSDA :  192
TnOctT     : Q P TL K- C LFK ------  FCAVMVAFY SF  NV  GWAFY LVS AR------SE   VH DNSWNTDQCWDS :  236
DmCG5549   : L P AV RRFC LFR ------  TGMILVSAI ML  NL  AWT F MFA FA------PV   QN EPAWSTKYCFSY :  190
AgCP10735  : L PAIL KR A ISE ------  YGMILVSGM ML  NV  AWT F MVV FE------DP   RG QHEWATVNCFSY :  135
HsGLYT2    : Q P SV K-AI ALQ ------C IAMLIISVL AI  NV  CYT F LFA FV------SV   GS NNPWNTPECKDK :  323
HsATB0+    : L P SV R- L LFQ ------  ITMVLISIF TI  NV  AYS Y MFA FQ------SE   KN S-SWSDKNCSRS :  167
HsPROT     : L P AV K- S LFK ------A AAMLLIVGL AI  NM  AYV F LFA LT------SD   EH GNWWNTELC--- :  166
HsGLYT1    : Q C GV R- S MFK ------  YGMMVVSTY GI  NV  CIAFY FFS MT------HV   AY NNPWNTHDCAG- :  231
DmCG7075   : T CTGM R- T LLK ------T IAQVVVNAYCVC  SV  SYP RMIFYCFF------KK   ED SNSWNTDDCVTA :  156
AgEbiP6595 : T C TV K- V LLK ------A IAIVVLNLICNA  IA  SYP L LWR LQ------AQ   ES GNAWNTPRCLEL :   96
MsPROT     : A C SV N- N LFK ------A YAVIVLNVIASI  SA  SYP L IYH MS------SP   QS GNSWNTVNCTEI :  138
DmIne      : R P GALGQ C LFK ------A LASVVVSFL ST  SV  GYS Y FFT FK------TE   ID NNRWNTPDCWVP :  185
AgCP4566   : R P GALGQ C LFK ------T LASVVVSFL ST  SV  AYA Y FFT FR------PE   TD SHRWNTPDCWIP :  227
AmINE      : R P GALGQ C LFK ------A LSSVVISFL ST HNV  AYA Y FFTAFR------AKQ  SD ENSWNTLACWLP :  433
MsIne      : H P GALSQ C LFK ------A LASVVISFL ST  AV  AWA Y FFT FK------TE   AS SNRWNTDQCWVP :  298
TnGAT      : I G GV K- A IFK ------  YAAAVMSCW NV  IV  AWA F FFM MR------SD   RN DNYWNTATCVNP :  167
MsGAT      : I G GV K- A IFK ------  YAAAVMSCW NV  IV  AWA F FFM MR------SD   RN DNYWNTATCVNP :  167
AgCP8499   : I G GV K- A IFK ------  YAAAVMSCW NV  IV  AWA F FFM LR------AD   RT DNAWNSINCVNP :  126
DmCG1732   : I G GV K- A IFK ------  YAAAVMSCW NV  IV  AWA F FFM MR------AD   RT NNWWNTVNCVSQ :  190
CeSNF11    : V G GM K- C IFK ------  IAAAVMAFW NI  IV  SWAAT LYN FTM-----SD   KN DHAWNTPNCRSE :  165
HsGAT1     : I G GV K- A MFK ------  LAAAVLSFW NI  IV  SWA Y LYN FT------TT   KQ DNPWNTDRCFSN :  176
HsGAT3     : E G TC RK C LFE ------  YATQVIEAH NV  II  AWA F LSNCFT------TE   AT GHEWNTENCVE- :  182
HsGAT2     : Q G TA RK C IFE ------  YASQMIVIL NV  II  AWA F LFS FT------ID   GG YHEWNTEHCME- :  164
HsBGT1     : Q S TA RK C LFQ ------  LASVVIESY NV  II  AWA F LFS FT------SE   TT NNFWNTEHCTD- :  168
HsTAUT     : E G TC EK C LFS ------  YASVVIVSL NV  IV  AWATY LFQ FQ------KE   AH NHSWNTPHCMED :  174
HsCT1      : A S NV N- C LFK ------  YASMVIVFYCNT  IM  AWGFY LVK FT------TT   AT GHTWNTPDCVEI :  184
CeSNF3     : K A EM L- C LFR ------  IGNVVIAFMCIA  CV  AWA F MIS IA------WVF  ET NNYWNDATCVTG :  162
DmCG5226   : L A GV NT H WLG ------  ISSCIVTLF AL  NV  TWVFF LFN FR------YP   SS P-LNGTG---FE :  184
AgCP7001   : L A GV NT H WLG ------  ISSCIVTLF AI  NV  TWCFY FFN FR------YP   AV PQLNGT-----D :  189
HsNTT7-3   : R S GV NY S KLG ------  FASCVVCYF AL  NV  GWS F FSQ FQ------QP   DQ P-LVKNASHTFV :  193
HsNTT4     : R S GV HY C RLG ------  FSSCIVCLF GL  NV  GWS F FFK FQ------YP   SE P-VVRNGSVAVV :  192
HsXtrp2    : K S GV TA S YLS ------  LGCVTLSFL SL  NT  AWV W LLN FQ------HP   SS P-PDLN--RTGF :  148
RnROSIT    : R S GV KT S YLG ------  LGCFSVSFL SL  NT  LWV W FLN FQ------HP   ST P-LDLN--RTGF :  135
HsBOAT1    : R S GV SS H ALK ------  LASMLTSFM GL  NT  SWI W LFN FQ------EP   SD P-LNEN--QTGY :  162
HsXT3L     : Q S GA RT S YLS ------  VASVVVSFF SM  NV NAWAFW LFH FQ------DP   SV P-LNGN--HTGY :  135
DmCG10804  : K A GV SQ S YLG ------  ISSAVVSYI AL  NT  AWC I LLH FE------SP   AD PTRLYKNFTYDH :  187
AgCP7501   : K A GV HE SAYLG ------  ISSAFVSYI AL  NT  AWC I LLH FE------TP   AE PKRLFKNFTYDI :  136
DmCG1698   : R S KV D-FS IMR ------  YGQVLATGI TT  AT  ALT R FVD FY------PT   SY REEWGTECLDSG :  206
DmCG4476   : R I QV D-FA LMR ------  YAQLLALGV AT  AS  ALT R FFD FA------SE   SF REEWGDGCVS-- :  181
DmCG15279  : R S KV D- C AMK ------  AGQAFQVFM ST  AA  AIIGR FIE FR------NP   ST RAEWGIHCINSA :  182
AgCP10412  : R S KV D- A IMR ------  YGQLFSVTA IT  SS  ALIAR MID FM------NP   AH RSEWQPNCIDSV :  211
DmCG8850   : R C RA D- A IMR ------ AYGQVYSTALATT  AC  ALT R LVA FS------EV   TY LVEWGKSCVATG :  176
DmCG15088  : R V KA D- A LLK ------ ALGQVLATAASIT  SS  ALT R WLA FG------SE   SR WESWGTDCHDGN :  150
AgCP10401  : R C NV D-AS AMR ------  VGQTYSTFI MT  AS  AVT R LIA FG------DP   SE NEAWNATCIDSR :  207
AgCP10503  : R C KI D- A AMR ------  VGQSVAMVVAMS  TP  AIT R LLL FS------SE   SK DHSWSRCSIRTF :  133
AgCP10610  : R C KV D- A AMR ------  VGQTVAIFT IT  AS  AVT R LVA FN------PE   AK DPTWPDCVDSSR :  111
DmCG3252   : Q T KI S- V GFV ------  YGQAFGTIC IS  SS  ALT Y LFV FQ------SE   SY RDEWTN-CVNSR :  169
AeAAT      : RSS KI E- S LFK ------  IGQLVGTTS VS  VS  ALT H IFA FA------SE   AT KDNWADNCVDS- :  174
MsKAAT1    : RNS KI S- S AMK ------T YAQAVGCGY LS  VV CGLC F LAM FQ------AT   AI QPEWEN-CVPSD :  182
MsCAATCH1  : RNS KV S- S AMK ------T YAQAAGCGY LS  VV CGLC Y LAM FQ------AT   AI QPEWEN-CVPSD :  182
BcProT     : GSAPLT FR S RAEF------  WWQMCVTFI ST  AV  AWS S TYFAITG--------A GK-------------- :  122
HinSnfT    : G APLS RRFS HFEV------F WWQMMVNVI GL  AV  GWAAS TYF FTG--------A GD-------------- :  129
StTnaT     : TAT TA KK NRRFEW------  WWQITVPVV VT  SV  SWS R LIF FTQ--------A GD-------------- :  123
CeSNF2     : KSPPAV RR M ALE ------  WMTCIVGAV GV  MI  SWI L IINIFYV-------NNMGR DNPWNQNNCYDG :  240
CeSNF7     : KSPPAA RR M ILE ------  WMTCLVGAI GV  MI  SWI L LFNIIYA-------TKIS--------------- :  183
CeSNF4     : RSPPFV RR M ALE ------F WVA---------------------------------------------------- :  183
CeSNF5     : KSPPVI RR R ILE ------  WMGVFVAAL AI  IV  SWISI MINICRG-----HFAL SH NNDWNNGTSCIT :  216
CeSNF9     : TTAPII RRFA ILQ ------  WMAVAVSSL AI  II  GWSTV MASIVMG-----HTSK NR GNSWNNAESCVD :  178
DmCG13794  : S F SA R- S FFK ------  YVSVFLTIS LI  SIFAAVP L MIN FR------PT   SCEGFKSWYNESDGK :  151
DmCG13795  : S F SA R- S FFK ------  YVSGFLTIS LI  SIFAAVP L IIN FR------PT   SCEGLSSWHNGSATL :  151
```

```
DmCG13793  : S F SA R- S LFK ------  YVSLFLTIT LI  TIFAAVP L VLN FR------PT   SCEGFKSWQNDTIAK :   79
DmCG33296  : S F AA R- A LFK ------  YVSIFLTFTTLI  SMF AIP L LIN FR------PT   SCEGLKSWHNRTDEY :   87
DmCG8291   : A P SM K- S ICA ------  IALVMQQCF AL STVS AWI V LRDVFPTA--ARSGYR QEMAFPYRYDASNAT :  629
AgCP9955   : G P TM R- S ICK ------  IALLLAQTL AL SAIS AWV V FRDAFIMR--NEK-YR QEPFEAYHSGRDNQS :  134
DmBLOT     : Q AAHT R-AS IFK ------ACMISRFASWLSAI VSLQAVLA A IGMFAS------ND   RE AGPVKLR----L :  339
AgCP2093   : Q SAHM R-AA FLK ------ASLVGRIASWLAAI TSMQSVIA L IGMLAF------KS  LRE SKSVTINPQAIY :  135
CeSNF6     : SAA SV K- V LFK ------  FAQVTISGFFAV  NI SAWT F LIN FS------FS   SN ANSWSGENCTLG :  195
CeSNF12    : QAANTA QR M IGS ------  WALVVIAIP AV  NI  AWA H FFQ AKGLLLG-DE   ET RDEWQLDNRCCN :  295
AaLeuT     : AQGHGTTPA FYLLWRNRFAKI  VFGLWIPLV AI  VY ESWT G AIKFLVG------L  EPPPNATDPDSILR-- :  142
HsNTT5     : Q G GV KI A WIG ------  YSSFMVCFI GL  NV NSWI F MSQ FQ------FP   EK PLTMNSSGFDPE :  236
DmCG13796  : S T SA R- A IFK ------  YAILLLNLGTLT  SIAAVVP I TVN IH------PV   MS NNSWNTQECSLH :  217
HspSnfT    : RSP GA SS SDSTLWR----VF HLSVVIAVV LS  SV GGWI R FGA LTG------------------------ :  119
CeSNF1     : RAP LA YH F VFR ------  VSQILFTMV LACMTKF STLCL LYYYFWTFHAGRSG   LN KAFPEFVSQPCR :  160
CeSNF8     : CSPWIFIRRCA ILQ ------F FMALVSAVT LYP QYS ARAFK LLSLARYRS---QD   ST GNWWNTETGHIV :  150

                        *       660         *       680         *       700         *       720
DmSERT     : ----------------------------------------VTS---ENFTEL-A----------------TSPAKE :  222
AgCP12911  : ----------------------------------------VTGRVNETVTS--T----------------VATATT :  164
MsSERT     : ----------------------------------------VTS--PQTNPN--S----------------STPAKE :  189
HsSERT     : ----------------------------------------YFSEDNITWTLHS----------------TSPAEE :  230
CeSERT     : ----------------------------------------DLNVTISRNGTPLT----------------TPSEE :  255
DmDAT      : ESQNASRVPVIG--NYSDLYAMGNQ----------------SLLYNETYMNGSSLDTSAVGHVE--------GFQSAASE :  214
AgCP3649   : GYINSTAATVVN--------RTG----------------VLAANVTSTVSTINDT--------------KFASAASE :  175
TnDAT      : EAIWESSANKSR----------------------------VRQNSSSTSLGSPPTT--------------PFTSAASE :  182
CeDAT1     : EPHWSEDGTAMC----------------------------RSANQSVSAEK----------------ISAAEE :  200
HsNET      : KLLNG-----------------------------------SVLGNHTKYSKYK----------------FTPAAE :  212
HsDAT      : HPGD------------------------------------SSGDSSGLNDTFG----------------TTPAAE :  215
TnOctT     : GRDNA-----------------------------------TNRTDVRYQGPLSH----------------FTPASE :  261
DmCG5549   : A--QADQCEAT----------------------------NGTYYLRTCHNATSAAENNITTLALGALKR----PPAEE :  234
AgCP10735  : E--EEDNCLAT----------------------------NRSYYMANCMSRE---QYRLLNVTLRTPRK----PPAEE :  176
HsGLYT2    : TKLLLDSCVIS----------------------------DHPKIQIKNSTFCMTAYPNVTMVNFTSQANKTFVSGSEE :  373
HsATB0+    : P--IVTHCNVSTV----------------------------NKGIQEIIQMNKSWVDINNFTCINGSEIYQPGQLP-SEQ :  216
HsPROT     : ----LEHRVSK----------------------------DGN--GALPLNLTCTV----------------SPSEE :  192
HsGLYT1    : ---VLDASNLT----------------------------NGSRPAALPSNLSHLLNHSLQRT----------SPSEE :  267
DmCG7075   : S-----DMGKQ----------------------------NSS----DVFKTSA----------------DE :  174
AgEbiP6595 : RG----DQHVELFLN----------------------------NSAMAIGDRYRTPA----------------DE :  123
MsPROT     : TGNS--SFFTS----------------------------NGS------ITTPE----------------DE :  157
DmIne      : QRKGIN-----------------------------------ASAPDTS----------------RTPSEE :  204
AgCP4566   : ERLKHN-----------------------------------LTRPDMS----------------RTPTEE :  246
AmINE      : TYSGIDN----------------------------------RTRPNAS----------------RTPAEE :  453
MsIne      : N---HN-----------------------------------HTKPNGS----------------QTPTEQ :  314
TnGAT      : YDRKNLTCWSTM----DMTTFCT----------------LNGKNISKAVLS----------------DPVKE :  203
MsGAT      : YDRKNLTCWSSLG---DMSTFCT----------------LNGRNVSKAVLS----------------DPVKE :  204
AgCP8499   : YDRKDLLCWETIGVNGTLTKICS----------------LNSNNVSMSDLA----------------DPVKE :  166
DmCG1732   : YERKNLHCWDKIIN-GTTQKVCS----------------VSALNITSLELT----------------DPVKE :  229
CeSNF11    : YVKIPCDSNRTIAEFFNVKVLTHDHIHEYKKQFFVGEK----MNWTVCSAADLSV----------------VSPVKE :  222
HsGAT1     : YS----------------------------------------MVNTTNMT----------------SAVVE :  191
HsGAT3     : FQKLNVSNY----------------------------------SHVSLQNAT----------------SPVME :  205
HsGAT2     : FQKTNGS-----------------------------------LNGTSENAT----------------SPVIE :  185
HsBGT1     : FLNHSGAG----------------------------------TVTPFENFT----------------SPVME :  190
HsTAUT     : TMRKNKSVW----------------------------------ITISSTNFT----------------SPVIE :  197
HsCT1      : FRHEDCANASLANL----------------------------TCDQLADRR----------------SPVIE :  212
CeSNF3     : KENFTELAR--------------------------------IKALVASAGGHT----------------QTSVEQ :  189
DmCG5226   : LEECAKS-------------------------------------------------------------SETTY :  196
AgCP7001   : VEECERS-------------------------------------------------------------SETAY :  201
HsNTT7-3   : EPECEQS-------------------------------------------------------------SATTY :  205
HsNTT4     : EAECEKS-------------------------------------------------------------SATTY :  204
HsXtrp2    : VEECQGS-------------------------------------------------------------SAVSY :  160
RnROSIT    : VQECQSS-------------------------------------------------------------GTVSY :  147
HsBOAT1    : VDECARS-------------------------------------------------------------SPVDY :  174
HsXT3L     : DEECEKA-------------------------------------------------------------SSTQY :  147
DmCG10804  : EPECVAS-------------------------------------------------------------SPTQF :  199
AgCP7501   : EPECVVS-------------------------------------------------------------SPTKY :  148
DmCG1698   : P-----------------------------------QEASRAT--SLAG----------------SGVR--TTSAEF :  228
DmCG4476   : ---------------------------------------ASGGQ--PLQG----------------QLSRNFSSSTQL :  202
DmCG15279  : P----------------------------------DASNWSQLESNDQRPQNYTMK---------SQNDRVITSSEW :  216
AgCP10412  : A----------------------------------GSA-----SPNDS----------------SSNRTLTSSSEL :  232
DmCG8850   : A-------------------------------------------TAANDSS----------------IVQGVSSAEL :  194
DmCG15088  : A-----------------------------------------QNSSGK----------------MSPAQL :  163
AgCP10401  : L---------------------------------------ITNMAEN----------------STATAVSSAEL :  226
AgCP10503  : L---------------------------------------DMRTLR----------------MDWPRIAGTF :  150
AgCP10610  : L---------------------------------------GSIALGP----------------NVTQPKTSADL :  130
DmCG3252   : PQEYVDNLLTGV----------------------------SLANESARNLSGI-----VANDETEKLQSSSEL :  209
AeAAT      : ---------------------------------------SLVVNEMRDS----------NSTSGQKVSSSQI :  197
MsKAAT1    : P---------------------------------------TLAAG----------------VGNITNGTSSAEL :  201
MsCAATCH1  : P---------------------------------------TLAAS----------------VNNITNGTSSAQL :  201
BcProT     : ------------------------------------------------------------------DTQS :  126
HinSnfT    : ---------------------------------------------------------------------KPID :  133
StTnaT     : ----------------------------------------------------------------------DPGT :  127
CeSNF2     : SEQSRCNKGLYSNMTANASSFLNTT----------------MAPLMSSGSKRIL---YINGKCQDATNMSVEVGTEQ :  298
CeSNF7     : ------------------------------------------------------------------------- :    -
CeSNF4     : ------------------------------------------------------------------------- :    -
CeSNF5     : MADQYLCKNHTKVMAN-STLWN----------------------SSLPIPDKMVYFNGACQD-----ANGTDVSTATEQ :  267
CeSNF9     : SGLQPLCASFNATGNETSPVWAN--------------------LSSEGRGKFYIYLNTSCYDKVDLEFNKTKMIAATEQ :  237
DmCG13794  : MTTCNMTIS-----------------------------EDLTDNSDNNTDFHR--------------PHFHVPSVL :  184
DmCG13795  : SSSCNKTL---------------------------------ILEDFSEYNRSHQ---------------LLHVPSVL :  180
DmCG13793  : WT-------------------------------------EEVAKETENDTYFYI--------------TNIHVPSVL :  105
DmCG33296  : PTLCHP-----------------------------------TIYKQHEN----YT--------------LGVEAPSRL :  112
DmCG8291   : GN-------------------------------------------------------------------LTQTVAE :  638
```

```
AgCP9955   : YR-------------------------------------------------------------------------LPDTVAD :  143
DmBLOT     : SGYLLTG-----------------------------------------------------------------------TSGQE :  351
AgCP2093   : NGYDVQQ------------------------------------------------------------------------SSGQE :  147
CeSNF6     : TRIQCKEMN-----------------------------------GTLLVNGSCIVEHASSNETTVIPLHDLGSIPSLK :  238
CeSNF12    : LHNLHSCFNSTN---------------------------------------------------------------SITAPEA :  314
AaLeuT     : ----------------------------------------------------------------------------PFKE :  146
HsNTT5     : CER------------------------------------------------------------------------TTPSIY :  245
DmCG13796  : ENYDVDDYRVDP-----------------------------------------------------------------HSTVE :  234
HspSnfT    : ------------------------------------------------------------------------------A :  120
CeSNF1     : EAGSITN---------------------------------------ITQINNRLN---------------TIQAESSMM :  185
CeSNF8     : FQNKCVN---------------------------------------ETKIK------------------MEAVESAEIQ :  172

                         *      740         *      760         *      780         *       800
DmSERT     :   ERKV ESYKG--------NGLDFMGPVKPT ALC FGV V VYFS W   RS--A  VV VTALA  V  I LLV G  :  292
AgCP12911  : TAMRQV EQYKS--------NGLDFMGPVKPS ALC FGV V VYFS W   RS--A  VV VTALA  V  L LLA G  :  234
MsSERT     :   ERNV EQHKS--------NGLDDMGPIKPS ALC FGV V VYFS W   RS--A  VV VTALA  V  L LLA G  :  259
HsSERT     :   TRHV QIHRS--------KGLQDLGGISWQ ALC MLI T IYFS W   KT--S  VV VTATF  I  S LLV GA :  300
CeSERT     :   LYKV EVQKS--------TGFDDLGGVKTS AVC LAV I VYFA W  PQS--S  IV VTATA  I  S LLI G  :  325
DmDAT      :   NRYI ELNRS--------EGIHDLGAIKWD ALC LIV L CYFS W   ST--S  VV FTALF  A  L LLI G  :  284
AgCP3649   :   NRYI ELDKS--------EGIHDLGAIKWD ALC LAV L CYFS W   ST--S  VV FTALF  A  L LLV G  :  245
TnDAT      :   NRAI ELQGS--------EGLHDLGPIKWD ALC FAV I CYFS W   ST--S  VV FTALF  A  L LLV G  :  252
CeDAT1     :   YKGF GLHEANAPNSHVIRSVTDLGNVRWD ALS FVV L CYFS W   HT--S  VV FTALF  V  G LFI G  :  278
HsNET      :   ERGV HLHES--------SGIHDIGLPQWQ LLC MVVVI LYFS W   KT--S  VV ITATL  F  F LLVHG  :  282
HsDAT      :   ERGV HLHQS--------HGIDDLGPPRWQ TAC VLVIV LYFS W   KT--S  VV ITATM  V  TALLL G  :  285
TnOctT     :   HRAV EMQNS--------EGLNDLGFPKWQ AIC GLV VTLYLS F   KS--S  VV MTATM  V  S LLA G  :  331
DmCG5549   :   NNFV GL--S--------KGIEETGSIKLS AAC FLA T VFLC C   QS--S  VV FTALF  V  V LFV G  :  302
AgCP10735  :   KNYF QI--S--------AGIEETGMINQR AIC LVA A VFLC S   KS--S  VV FTALF  V  AALFI G  :  244
HsGLYT2    :   KYFV KI--S--------AGIEYPGEIRWP ALC FLA V VYAS A   KT--S  VV FTATF  V  V LLI G  :  441
HsATB0+    :   NKVA QR--S--------SGMNETGVIVWY ALC LLA L VGAA F   KS--S  VV FTALF  V  L LLV GA :  284
HsPROT     :   SRYV HIQGS--------QGIGSPGEIRWN CLC LLA V VFLC L   KS--S  VV FTATF  L  L LLV G  :  262
HsGLYT1    :   RLYV KL--S--------DDIGNFGEVRLP LGC GVS L VFLC I   KS--S  VV FTATF  V  T LFV G  :  335
DmCG7075   :   HLEV RI--S--------TDISDLGGMVWEQLTA IIT I IYFC M   KS--V  VV FTVPF  L  F LLI GA :  242
AgEbiP6595 :   HNEI HI--S--------EGIEQMGGIVWP LLCDVIA I IFGC A   QS--V  VV CTATF  A  A LLV G  :  191
MsPROT     :   HRHL QV--S--------PNINHIGSIVAP FWCNLIC I VYLC CN  KS--V  IV FTVLF  V  S LFV G  :  225
DmIne      :   ENKV QISGG----------LEYPGMMRWE FAC ICA L VYFATW S KS--SA VR FTATF  V  I LMV A  :  272
AgCP4566   :   ENKV QISHG----------IEYPGAMRWE VAC VCA I VYFA W S KS--SA VR LTATL  V  V FLG S  :  314
AmINE      :   DNKV QISNG----------IEESGPLRWE IAC ITA I VYFS W S KS--SAQVR LTATL  L  V FLM S  :  521
MsIne      :   ERKV NMSAG---------IEYPGGMRWE AAC VCA V VYFA W S KS--SA VR ITTTL  L  I FLG S  :  382
TnGAT      :   ERRA QISSG----------IEHIGNIRWE AGT LLV V CYFC W   RW--T  VV FTALF  F  T LLI G  :  271
MsGAT      :   ERRA QISSG----------IEHIGNIRWE AGT LLV V CYFC W   RW--T  VV FTALF  F  T LLI G  :  272
AgCP8499   :   ERRT MISSG----------IDQVGSIRWE AGT LLV I CYFC W   KW--T  VV FTALF  F  T LLI G  :  234
DmCG1732   :   ERRA QISHG----------IEEIGNIRWE AGT LLV I CYFC W   KW--T  VV FTALF  V  T LLV G  :  297
CeSNF11    :   NHRV GISSG----------LENPGGIRWD ALF LLV I CYLC F   KW--T  VV ITASF  M  FCLLI G  :  290
HsGAT1     :   ERNMHQMTDG----------LDKPGQIRWP AIT AIA I VYFC W   GW--T  VV FSATY  I  I LFF G  :  259
HsGAT3     :   EHRV AISDG----------IEHIGNLRWE ALC LAA T CYFC W  TKS--T  VV VTATF  I  L LLI G  :  273
HsGAT2     :   ERRV KISDG----------IQHLGALRWE ALC LLA V CYFC W   KS--T  VV FTATF  L  V LLI G  :  253
HsBGT1     :   ERRV GITSG----------IHDLGSLRWE ALC LLA V CYFC W   KS--T  VV FTATF  L  V LLI G  :  258
HsTAUT     :   ERNV SLSPG----------IDHPGSLKWD ALC LLV L CFFC W   RS--T  VV FTATF  A  L LLV G  :  265
HsCT1      :   ENKV RLSGG----------LEVPGALNWE TLC LAC V VYFC W   KS--T  IV FTATF  V  V LLV G  :  280
CeSNF3     :   EKRV HDTGD----------ISEFGGIQWE FFI AAA L VYFA W   TQ--AR FV FCALF  V  F LLI G  :  257
DmCG5226   :   YRTT DAAPS---------MDEPGGLKWW VLC MLS T VFFI M   QS--S  VV FTSLF  I  T FFI G  :  264
AgCP7001   :   YRTT DAAPS----------IDETGGFKWW VLC MLS T VFFI M   QS--S  VV FTSLF  I  T FFI G  :  269
HsNTT7-3   :   YREA NISSS----------ISESGGLNWK TIC LAA V VCLA I   QS--S  II FSSLF  V  ICFLI AF :  273
HsNTT4     :   YREA DISDS----------ISESGGLNWK TLC LVA S VGMA V   QS--S  VM FSSLF  V  ACFLV G  :  272
HsXtrp2    :   YRQT NITAD----------INDSGSIQWW LIC AAS A VYMC I   ET--T  VI FTALF  L  T FLI G  :  228
RnROSIT    :   YRQT NITSD----------ISNTGTIQWK FLC VAC TTVYLC I   ES--T  VI FTALF  L  T FLI G  :  215
HsBOAT1    :   YRET NISTS----------ISDSGSIQWW LLC ACA S LYMCTI   ET--T  AV ITSTL  V  T FLI G  :  242
HsXT3L     :   YRKT NISPS----------LQENGGVQWEPALC LLA L VYLC L  TES--T  VV FTASL  C  I YLI G  :  215
DmCG10804  :   YRTT QCSES----------VDMPENFNYH AIA IVS F VYIC VQ  TS--S  IV MTAIF  V  I FFF G  :  267
AgCP7501   :   YRET QVSPS----------VNEPEQINYT ALA ITA S VYLC VQ  TE--SS IV ITAIF  V  I FFF G  :  216
DmCG1698   :   TNII REKASID---------DGIGYPSWS ALA AVA I IAGI F   KS--S  AS FLALF  V  L LLV A  :  297
DmCG44476  :  LQRIV NETDSLE---------EGIGYPSGS ALM GIS LTVTLI I   KS--S  AA VLALF  V  F LLV A  :  271
DmCG15279  :   VKEV HEKPNIE---------EGIGLPNWE VIG FIA SCVFFI R   KS--S  AS FLALF  V  G LLV A  :  285
AgCP10412  :   TKVV KELDGID---------DGIGLPDLR TLF VLS S VFLT I   KS--S  AS FLALF  V  T LLV AC :  301
DmCG8850   :   TQTV REPESLDD--------NGLGTPSWD VLC LAT V IGTI S   RS--S  AS FLALF  V  I LLI A  :  264
DmCG15088  : ---REI HEVPNID---------NGLSLPNWQ VAC AIA L IGGV I   KS--S  AA FLGVF  V  L LLL A  :  229
AgCP10401  :   VKDV KEASTIH---------DGIGTPDWR VLC LVP TCICLT I   KS--S  VA FLAIF  V  L LLI AC :  295
AgCP10503  : PLSCTLQ----------------GLGWPDGK VLC LVS S LVVI I   RS--T  AA FLAIF  V  F LLAHS  :  211
AgCP10610  :   RNTI HRAPLAE---------GLGWPDGK VLC LVS S LVVI I   RS--T  AA FLAIF  V  F LLAHS  :  198
DmCG3252   :   LNVV KEKLDIS---------DGVGDPDWK TLA FVA V IFLV M   KS--S  AA FLALF  V  F LLI A  :  278
AeAAT      :   LDIV KEKDSID---------DGIGAPDWK TLW SLA V IFLV V   KS--S  VA FLAIF  V  I ILV A  :  266
MsKAAT1    :   LRTV QQSDGIE---------GGLGAPIWY VLC FIA L VFGV A   KS--S  AA FLALF  V  ITLFITT :  270
MsCAATCH1  :   LRTV QQSDGIE---------GGLGAPIWY VLC FIA L VFGV A   KS--S  AA FLALF  V  ITLFITT :  270
BcProT     :  LFKEY HVAD----------KPGQFGGLVPE LIP ALV I VLGVAF   KKG-IEVVNRIFIPLLVV FL IVV A  :  195
HinSnfT    :   IGEF KMGDI---------KNGISFEFVGM TAP IAM I ALGV SM  QKG-IA VSSVLMPVLVV FM LVIYS  :  203
StTnaT     :   SSDF GITS----------GPLELGGLRWG FAA AVV FANYYISAQ  SGG-IE ACKIMTPFLIVA L FVI G  :  196
CeSNF2     :   SKSV RPSAG----------FTDFNSINLP LGA ALC V AAVV I   KA--I  LS VTVIL  F  VSLLICG  :  366
CeSNF7     : --KNFV KPSTG----------FTDFNSINVP LAA AVC V AIIS I   KA--I  LS YTVLL  A  IALMI G  :  249
CeSNF4     : --ANYI KPSKG----------LLDINSINWP FIA S---------FI --------- IS VTVIL  I  I LLI G  :  234
CeSNF5     :   MTYI QPSSG----------MLDFGGFNWP FAA SVC L TGLG L  AKI--M  IS VSVLV  V  V LFING  :  335
CeSNF9     :   FDYV KPTDG----------FLEFSSINFKSFFG NAC L VVLI W   GY--M  AS VIVTL  F  I LFF G  :  305
DmCG13794  :   LKHFESVQEENY------SDLRQDYDLSWRFVSLFVLV A VAFIFY FSETAMF  FIRYMVIGTLI  L CFG F  :  258
DmCG13795  :   QNHYDSMQQDAV------PDWYKDYELSWHFVGLFAII V IAFIFY FSETAKF  LIRYMVIVTLV  L CFV F  :  254
DmCG13793  :   N-------------------ELNDYEISWH SGLFALV ATIAFIFY FSEPAKF  CIRYMVIGTVA  L CFV F  :  166
DmCG33296  :   DYIR-----QDY------FNTDKRFDISRP IGLFTLS A IAIIFYHFSEPAKF  CIRYMVITTMV  V CFV F  :  181
DmCG8291   :   NVVV QRLHLANHPDAS--GIRFHVNDR-Q AFY ALI AAVFLI C   KS--L  LA IIYTL LVA A VTA F  :  713
AgCP9955   :   NGVV QRYYLAHQAVSS--SLPTGYAIRFQ AFN AII TFVFVA C   RS--L  ITIGVFVGA AA TAVCA F  :  219
DmBLOT     : CLQLTF TPFWR-------------NPLYFGL AAG IGL I VMLCTHNAKIL---R SI VFGLVGLV  CTLTGWE  :  415
AgCP2093   : CLKLTF RPVWR-------------SSLYFGL ALS ILL V SMVCTHS KIN---R SI LFGFVALV  IFETGWE  :  211
```

```
CeSNF6     :   QNDV MLSKG----------VDDFGTLNWY GLC LAC IAVFLC FQ  KS--S  VV VAVIV  I  T LLT L  :  306
CeSNF12    :   HSEV SLSTFG---------DFALGPLQSH VLS AAA L VFFG F   GS--IAQTMNVTATV  L  S LLL G  :  383
AaLeuT     :  LYSYIGVPKGD-----------EPILKPSLFAYIVFLITMF NVSI I   SKG-IE FAKIAMPTL I AVFLVI VF :  214
HsNTT5     :   YQQA KASDR----------IEDGGSPVYS VLPFFLC C VGAF IN  KS--T  VI VLVLL CF  VGFFI T  :  313
DmCG13796  :   RSMIASTAKG----------STSMSISWS LIG LSI L VLAL L P AF--I  VLRCSCVLM GFFVAVFLYL  :  301
HspSnfT    :   SNPGQYFATVS---------------YGLDAVGFHLLFLA TGLI V   KRG-IENATKLMMPAVVV  I LAAWAS :  184
CeSNF1     : Q LTTLERPSES----------IADFKDFQYP LIAQGAI IAIFVA CF  RF--V  TVAFVLMLA SS L FFL SA :  253
CeSNF8     :  LRNVIHRIGVS----------HMSFTNFNYG AYS FII LTISFVQN VNTR--TTPLI IPTVFA L CHFLLV S  :  240

                          *       820         *       840         *       860         *       880
DmSERT     : S P ---- DE  KYYLT E--------WHK KNSK  ID AS     ----LGPGF T  ALS  NKFNN CY------ :  350
AgCP12911  : T P ---- AE  RYYLT Q--------WDK NNSR  ID AS     ----LGPGF T  ALS  NKFNN CY------ :  292
MsSERT     : T P ---- TE  RYYLT E--------WHK QNSK  ID AS     ----LGPGF T  ALS  NKFNN CY------ :  317
HsSERT     : T P ---- WR  LFYLK N--------WQK LETG  ID AA     ----LGPGF V  AFA  NKFNN CY------ :  358
CeSERT     : L P ---- KN  YYYVT D--------FEK KDPA  SA AT     ----LGPGF V  ALS  NDFNN CY------ :  383
DmDAT      : T P ----SFL  QYYLT N--------FSA YKAE  VD AT     ----LGPGF V  AYA  NKYHN VY------ :  342
AgCP3649   : T P ----SAE  LYYLR N--------FDV YNAE  VD AT     ----LGPGF V  AYA  NKYHN VY------ :  303
TnDAT      : T P ----SAT  QYYLS N--------FEA TQPQ  VD AT     ----LGPGF V  AYA  NKYHN VY------ :  310
CeDAT1     : T P ----WQN  EYYLR N--------FEM KRPS  QD AT     ----LGPGF V  AYS  NDFHN VY------ :  336
HsNET      : T P ---- SN  NAYLHID--------FYR KEAT  ID AT     ----LGAGF V  AFA  NKFDN CY------ :  340
HsDAT      : T P ---- ID  RAYLSVD--------FYR CEAS  ID AT  C  ----LGVGF V  AFS  NKFTN CY------ :  343
TnOctT     : L P ---- TR  AYYLQ E--------LTR KDTQ  VD AV     ----VGAGF VH SYA  NTFHN CY------ :  389
DmCG5549   : T P ---- ST  LFYLT D--------WKQ ANAQ  GD AV    A----LSPAW G  TLS  NKFSN CYNCGTFE :  366
AgCP10735  : T P ---- YD  LYYVR D--------WDK RTAQ  GD AV    A----LSPAW G  TLA  NRFDN CY------ :  302
HsGLYT2    : T P ---- GA  WYFIT K--------WEK TDAT  KD AT     ----LSAAW G  TLS  NKFHN CY------ :  499
HsATB0+    : T E ---- SK  SYYIGAQS------NFTK KEAE  KD AT     ----LSVAW G  ALS  NKFKN CF------ :  344
HsPROT     : T P ---- WK  QFYLT Q--------FHH LSSK  IE AL     ----LGVGF G  TFA  NTFHQ IY------ :  320
HsGLYT1    : T E ---- FD  MYYLT Q--------WDK LEAK  GD AS     ----LGCAW G  TMA  NKFHN CY------ :  393
DmCG7075   : T P ---- WK  KFYLY E--------WHR LDLK  AD AV    G----IGPGW G  NMA  SSFRN AK------ :  300
AgEbiP6595 : T P ---- LE  KFYIM Q--------WSQ LNLR  AD AI     ----LGPGW G  NMA  NQFKN TK------ :  249
MsPROT     : T P ---- WK  MFYIL D--------WAQ AKPK  AD AT     ----LGPGW G  SMS  NKFHY NL------ :  283
DmIne      : T D ---- AE  RFFFR K--------WSE KNAN  IN AS N N ----LGITF S  SFA  NKYNN IL------ :  330
AgCP4566   : T E ---- DK  HYFFR N--------WEE GRAN  IN AA N N ----IGIAF S  SFA  NKYNN IL------ :  372
AmINE      : T E ---- DK  QFFFH R--------WEL ADAK  IY TA   N ----VGIAF S  CFA  NKFHNTIL------ :  579
MsIne      : T D ---- DG  RFFFK D--------WEL KQSRP VN AS   N ----IGIAF S  MFA  NRFDN FL------ :  440
TnGAT      : T P ---- MQ  KFYVM N--------MSK MESE  ID VT     ----YGLGL T  ALG  NKFTN VY------ :  329
MsGAT      : T P ---- ME  KFYVM N--------MSK LESE  ID VT     ----YGLGL T  ALG  NKFTN VY------ :  330
AgCP8499   : T P ---- FE  KFYVS N--------LSK AESE  ID VT     ----YGLGL T  ALG  NKFNN VY------ :  292
DmCG1732   : T P ---- LE  KFYII N--------FSK TNSE  ID VT     ----YGLGL T  ALG  NKFTN VY------ :  355
CeSNF11    : T E ---- GV  EFYLK D--------FSK LESK  VD VT     ----YGLGL A  ALG  NKFNN VY------ :  348
HsGAT1     : T P ---- KE  LFYIT N--------FRK SDSE  LD AT     ----YGLGL S  ALG  NSFHN VY------ :  317
HsGAT3     : T P ---- SE  KFYLY D--------LSR SDPQ  VD GT     ----YAICL C TALG  NNYNN CY------ :  331
HsGAT2     : T P ---- AQ  QFYLY N--------LTR WDPQ  MD GT     ----FAICL C TALG  NKYHN CY------ :  311
HsBGT1     : T P ---- YQ  IYYLK D--------LFR KDPQ  MD GT     ----FAICQ C TALG  NKYHN CY------ :  316
HsTAUT     : T P ---- GA  KFYLY D--------ITR EDPQ  ID GT     ----YAICL A TSLG  NKYKY SY------ :  323
HsCT1      : L P ---- LD  IYYLK D--------WSK GSPQ  ID GT     ----YAIGL A TALG  NRFNN CY------ :  338
CeSNF3     : T E ---- GT  YFYLK N--------ATR LDTA  KD GT     ----YGVGF A  ALG HNKFNH CF------ :  315
DmCG5226   : T R ---- GA  MHMYT K--------VEK LEPT  LD AT     ----FGLAF S  AFG  NTPKN CV------ :  322
AgCP7001   : T K ---- GA  MHMYT K--------VEK LEPT  LD AT     ----FGLAF S  AFG  NTPKN CV------ :  327
HsNTT7-3   : L N ----SID  RHMFT K--------LEI LEPK  RE AT    A----LGLGF G  AFS  NKRDN CH------ :  331
HsNTT4     : L R ---- VD  LHMFT K--------LDK LDPQ  RE AT    A----LGLGF G  AFS  NKQDN CH------ :  330
HsXtrp2    : T P ---- TK  IYLFT N--------MHI QNPR  LD AT     ----LSLAF GH AFA  NSPRN CQ------ :  286
RnROSIT    : T P ---- TE  TYLFT N--------MKI QNSR  LD AT     ----LSLAF GH AFA  NQPRN CE------ :  273
HsBOAT1    : T K ---- TN  VFLFT N--------VTE AQPDT LD GA     ----FSLAF G  SFS  NSVHN CE------ :  300
HsXT3L     : T H ---- TN  MYMFT K--------IEQ ANPKA IN AT     ----LGLGF S  AFA  NEPSN CQ------ :  273
DmCG10804  : T K ---- AD  AHLFT R--------WET LDPV  LE GT     ----LGLAF G  AFS  NPANN CY------ :  325
AgCP7501   : T K ---- SD  AHLFT R--------WES LEPV  LE GT     ----LGLAF G  AFS  NPANN CY------ :  274
DmCG1698   : T P ---- FD  LYFLR Q--------WHK LEPQ  YA VT     ----LAICF N  MYA  NRFGH IYR----- :  356
DmCG4476   : T P ---- YD  MYFLT Q--------WEK LEPQ  YN VT     ----LAVCF V  MYS  NRFGH VYR----- :  330
DmCG15279  : T P ----SID  YYFIK Q--------WGK LDPK  YA VT C   ----LSVCF N  MYS  NKFGH VHR----- :  344
AgCP10412  : T P ---- VD  VYFLK Q--------WDK YDPK  YA VT C   ----LSICF N  MYS  NKFRH VYR----- :  360
DmCG8850   : T P ---- WQ  VYFLK Q--------WSQ LNPH  YA IT     ----LAICF T  MYA  NDFNK VHK----- :  323
DmCG15088  : T P ---- ID  IYFFK Q--------WRE LNPL  YA VT     ----LAICF T  TYA  NNFNR VYN----- :  288
AgCP10401  : T E ---- AD  LYFIK Q--------WDR LEAK  YA VT     ----LTICF N  MYS  NRFHN VYRFVNAR :  359
AgCP10503  : S E ----SLE  KFFLT K--------WES FSAK  ME VT C   ----LSICF G  AYS  NNFSN VYRDAMII :  275
AgCP10610  : S E ----SLE  KFFLT K--------WES FSAK  ME VT C  ---------------------------------- :  233
DmCG3252   : T E ---- RD  LFFLE Q--------WGE LNPT  KE VV C   ----LAVGS P  MFA  NRFDHGIYR----- :  337
AeAAT      : T E ---- VD  IFFIK Q--------WGE LNPK  YS TT     ----LSVGM S  MFS  NNFHH IYR----- :  325
MsKAAT1    : L P ---- TD  LFFVT D--------WWK LELG  YA IT     ----LSVCT A  MFS  NGFKQ VYR----- :  329
MsCAATCH1  : I P ---- TD  LFFVT Q--------WAK LELG  YS VT     ----LTVCT P  MFS  NGFRH IYR----- :  329
BcProT     : T E ---- MQ  DAFFK D--------WSR FDGK  LA YG     ----LSLAF I  TYS  LPKNS TTN----- :  254
HinSnfT    : F P ---- AK  DALFT D--------WTK SNPS  IA YG     ----LSIGF I  TYA  LKKES LTG----- :  262
StTnaT     : T P ---- TY  NYFLN D--------FSK MDPG  VA YS     ----TTLAV V  AYA  VPEDS LAN----- :  255
CeSNF2     : S P ---- KD  YHLFG-------QTDFNY YKRET TA LQ  C  ----LSIGQ G  NIA  NKKTY WYR----- :  426
CeSNF7     : T E ---- ST  HFLFWGKPNEYGHYDFAG YDPD  IA LM  C  ----LSIGQ G  NIG  NKKTY WYR----- :  316
CeSNF4     : T E ---- GD  KYFLA-------RTDWVA YNFKT SA LT  C  ----LSIGF V  NTA  NKQNHRCFR----- :  294
CeSNF5     : FQD ----SGV  EMYFG-------TPNYTK YEQDT TE LK  C  ----LSVGH G  SLS  SPKRN IFR----- :  395
CeSNF9     : T D ---- ED  YYYLG-------NPDWSK FMPAT GE LK  C  ----LGIGY G  GMA  SKPNN CFR----- :  365
DmCG13794  : F P ---- LL  RKYVT S-----------PDEIA GIGSTFV VLHA---FGAGW S  ALS  NGFKT IMS----- :  315
DmCG13795  : F P ---- LN  HRYVT K-----------ASDMA GVVSTFI VLHA---FGAGW S  TLS  NGFKT IMS----- :  311
DmCG13793  : F P ---- HL  KNYMT S----------- EEFV GTSSTFI ALQA---FGAGW S  TLS  NGFKT IMS----- :  223
DmCG33296  : F P ----SHKL WMYIT N-----------TKDFSEGIYSTFI ALQA---FGAGW S  ALS  NGFKT IMS----- :  238
DmCG8291   : Y VD-----PSR QNIFAASD------FDDF VNSNS TA TQ T L ----WGLLGAS  AIT RSHTN---------- :  768
AgCP9955   : TFID-----YDS QNIFPATD------WQDFFLNSRS TS AQ T L ----WGLLGVS YALNCRSNRKGSCNV----- :  279
DmBLOT     : RNSF----SRHYFPELWGFDS--------NL AESN  FN LM  L  ----VNCGF A PMITGKFLYKG--------- :  470
AgCP2093   : TKSI----NEQYFPDLWPFHP--------ESFADST  FN LV  I  ----LNIGF A PVLTGKFLYKG--------- :  266
CeSNF6     : T D ----SLAA FYFLT K--------WEI MDLH  GE AV A   ----VSCCS G FTIA  SRFHN IYK----- :  365
CeSNF12    : S P ---- NK  TFLFTVDS--------TK WKWQ  KS AE    E----LGIDA P  SMAA SRYRN IYR----- :  442
AaLeuT     : L ETPNGT AD  NFLWT D--------FEK KDPG  IA VG     ----LSLGF A  TYA  VRKDQ--------- :  273
```

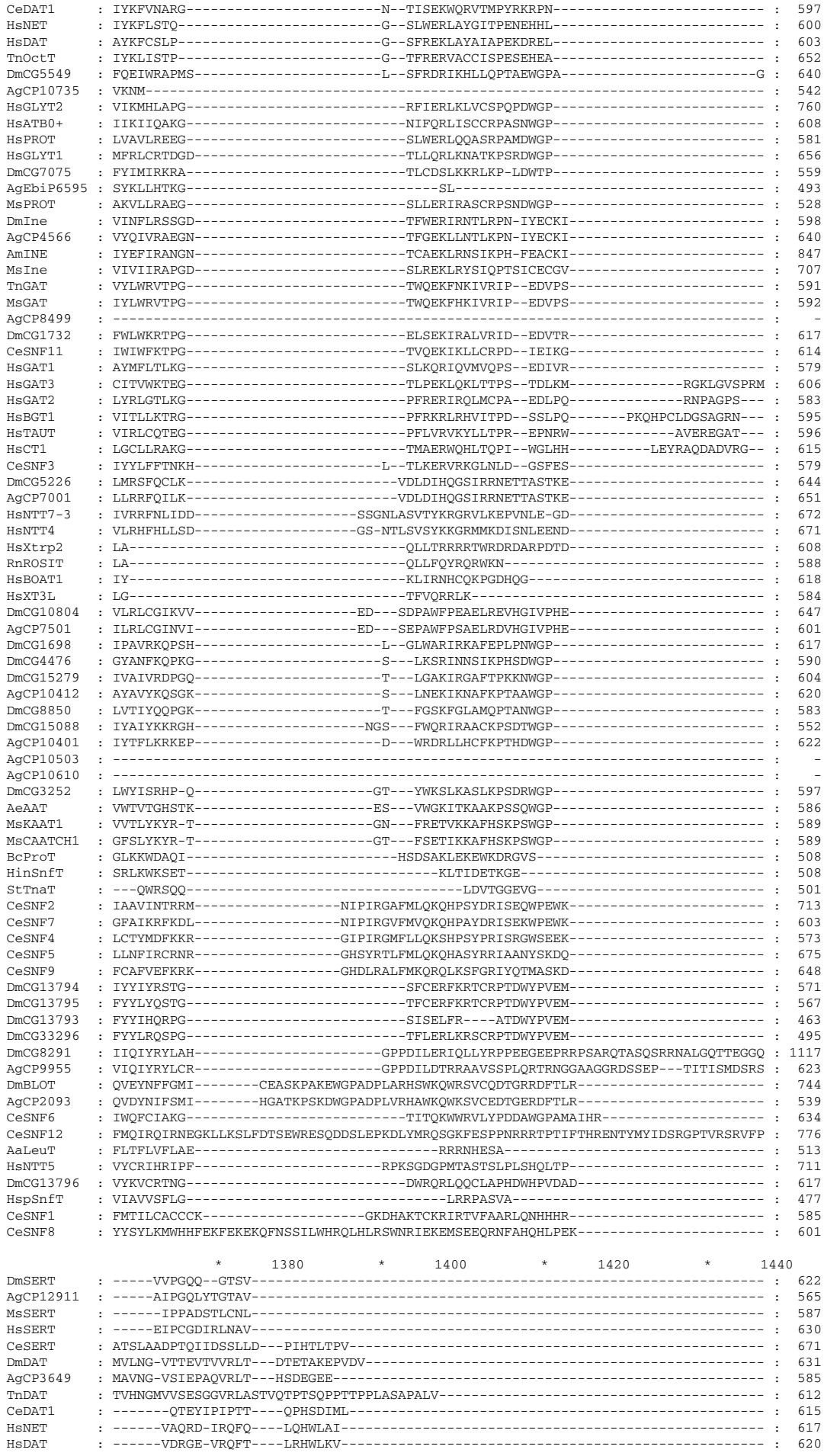
```
HsNTT5     : L E ---- KF  QQLVVAK--------ISD YNMS  SL GG  LSN----TGIGL S ASLA  MPQSN CLS----- :  372
DmCG13796  : IHEQ---ITFDT HYYWM -----------Q DSFDSILATARTALLMAGGALGPGW S  TLS  NNFRR--------- :  358
HspSnfT    : T D ---- GA  AFYLS D------VELLEADLLG LPA AG AL  ----LSLGS T  TYA  LDEER--------- :  241
CeSNF1     : T G ----VTEI EIYWKAT-------DWEK ADYQ  RL IE AILG----TGIGY AF TMS  MKRSN LVC----- :  313
CeSNF8     : H D ---- IE  QEMLRIE--------WAE LYSK  ID LNL VQ ----LSIGI GHTVMA  APPEK TFR----- :  299

                        *       900         *       920         *       940         *       960
DmSERT     : ----------------------------------------R ALITSSI CLTSFLAGFVI SV  YMAYV------- :  381
AgCP12911  : ----------------------------------------R ALLTSSI CLTSFLAGFVI SV  YMAQV------- :  323
MsSERT     : ----------------------------------------R ALITSSI CLTSFLAGFVI SV  YMAHV------- :  348
HsSERT     : ----------------------------------------Q ALVTSVV CMTSFVSGFVI TV  YMAEM------- :  389
CeSERT     : ----------------------------------------R AVTISII CATSFFSGCVV ST  YMSLL------- :  414
DmDAT      : ----------------------------------------K ALLTSFI SATSFIAGFVI SV  YMAHT------- :  373
AgCP3649   : ----------------------------------------K ALLTSLI SATSFIAGFVI SV  YMAHA------- :  334
TnDAT      : ----------------------------------------K AILTSVI SATSFVAGFVI SV  YMAHT------- :  341
CeDAT1     : ----------------------------------------V ALFTSFI CATSFLSGFVI SV  YMSCK------- :  367
HsNET      : ----------------------------------------R ALLTSSI CITSFVSGFAI SI  YMAHE------- :  371
HsDAT      : ----------------------------------------R AIVTTSI SLTSFSSGFVV SF  YMAQK------- :  374
TnOctT     : ----------------------------------------R CLVTTLV CFTSFFSGFVI TY  FMSHK------- :  420
DmCG5549   : LSTKSNS-------------------------------------R SLIVAFC IATSFFAGLVI SI  FLAHE------- :  404
AgCP10735  : ----------------------------------------R AIVVAIS ILTSFFAGFVI AI  FLAHE------- :  333
HsGLYT2    : ----------------------------------------R TLIVTCT SATSIFAGFVI SV  FMANE------- :  530
HsATB0+    : ----------------------------------------S AIVVCLT CLTSVFAGFAI SI  HMAHI------- :  375
HsPROT     : ----------------------------------------R TFIVTLG AITSILAGFAI SV  YMSQE------- :  351
HsGLYT1    : ----------------------------------------R SVIISIT CATSVYAGFVI SI  FMANH------- :  424
DmCG7075   : ----------------------------------------C SVLIVSI VFTSLLAGFVV SV  FLSEK------- :  331
AgEbiP6595 : ----------------------------------------L SIVIPIV CGTSILAGFVV SV  YFSHQ------- :  280
MsPROT     : ----------------------------------------RSSMIIPIV SATSIWAGFVV SV  FAAER------- :  314
DmIne      : ----------------------------------------R TVAVSAV MITSLLVGIFA ST  NLALE------- :  361
AgCP4566   : ----------------------------------------H TLAVSFV AITSLLVGIFA AT  NIALE------- :  403
AmINE      : ----------------------------------------I TMAVSLI GFSSLLVGIFS AT  NIALE------- :  610
MsIne      : ----------------------------------------H TVAVTLV AITSLIVGIFT AT  NIAFE------- :  471
TnGAT      : ----------------------------------------K ALIVCTV SSTSMFAGFVI SV  FMAHE------- :  360
MsGAT      : ----------------------------------------K ALIVCSV SSTSMFAGFVI SV  FMAHE------- :  361
AgCP8499   : ----------------------------------------K ALIVCTV SSTSMFAGFVI SV  FMAHE------- :  323
DmCG1732   : ----------------------------------------K ALIVCTV SSTSMFAGFVI SV  FMAHE------- :  386
CeSNF11    : ----------------------------------------KQALTVCFV SGTSVFAGFVI SF  FMATQ------- :  379
HsGAT1     : ----------------------------------------R SIIVCCI SCTSMFAGFVI SI  FMAHV------- :  348
HsGAT3     : ----------------------------------------R CIMLCCL SGTSFVAGFAI SV  FMAYE------- :  362
HsGAT2     : ----------------------------------------R CIALCFL SGTSFVAGFAI SI  FMSQE------- :  342
HsBGT1     : ----------------------------------------K CIALCFL SATSFVAGFVV SI  FMSQE------- :  347
HsTAUT     : ----------------------------------------R CMLLGCL SGTSFVSGFAI SI  FMAQE------- :  354
HsCT1      : ----------------------------------------K AIILALI SGTSFFAGFVV SI  FMAAE------- :  369
CeSNF3     : ----------------------------------------K AITMCFI GCTSITAGFAV SI  YMSHV------- :  346
DmCG5226   : ----------------------------------------R VLLVSVC AVTAIYASVVI AI  FKATVNVDRCVA :  360
AgCP7001   : ----------------------------------------R VLLVSVC AITAIYASVVI AI  YKATLGVDRCIQ :  365
HsNTT7-3   : ----------------------------------------F AVLVSFI FFTSVLATLVV AV  FKANVINEKCIT :  369
HsNTT4     : ----------------------------------------F AALVSFI FFTSVLATLVV AV  FKANIMNEKCVV :  368
HsXtrp2    : ----------------------------------------K AVVIALV RMTSLYASIAV SV  FKATNDYEHCLD :  324
RnROSIT    : ----------------------------------------K AVTIALV SMTSLYASITI SI  FKASNDYGRCLD :  311
HsBOAT1    : ----------------------------------------K SVIVSII GFTSVYVAIVV SV  FRATQRYDDCFS :  338
HsXT3L     : ----------------------------------------KHAIIVSLI SFTSIFASIVT SIY FKATFNYENCLK :  311
DmCG10804  : ----------------------------------------R AILVSLT CGTSMFAGVVV SV  FKATATFDRCTE :  363
AgCP7501   : ----------------------------------------R ALVVSVT CSTSMFAGVVV SV  FKATSIYDSCVE :  312
DmCG1698   : ---------------------------------------- ANIVTTL TFTSLLSGVII GI  NLAYE------- :  386
DmCG4476   : ---------------------------------------- ANIVTTL TFTSLLSGVII GI  NLAHE------- :  360
DmCG15279  : ---------------------------------------- AAIVTGL TMTSLLAGFTI GI  HLAHE------- :  374
AgCP10412  : ---------------------------------------- ATIVTSI TFTSLLAGCTI GI  HLAHV------- :  390
DmCG8850   : ---------------------------------------- VIIITTI SLTSILAGCII GI  NLAFE------- :  353
DmCG15088  : ---------------------------------------- IVIITTM SCSSIIAGCIT GI  NLARE------- :  318
AgCP10401  : -----------------------------------------DLL VTIVSIM TLTSMLAGLIV GV  HLAHV------- :  392
AgCP10503  : SWLDTFTSIVVGCIVFGVLENSKTNIALTLLQDPNISVSFPHS AMIISWL TFTSIVSGIVV GI  NIAHI------- :  348
AgCP10610  : ------------------------------------------------------------------------------- :    -
DmCG3252   : ---------------------------------------- AMIVTTL TLTSLLGGITI AI  NLAHN------- :  367
AeAAT      : ---------------------------------------- AMIVTTL TFTSLLGGMTI SI  NLAHN------- :  355
MsKAAT1    : ---------------------------------------- ALIVTTL TFTSFLSGCTI GI  NLAYE------- :  359
MsCAATCH1  : ---------------------------------------- AWIVTTL TFTSFLSGCTI GI  NLAYE------- :  359
BcProT     : ---------------------------------------- AFITGFA SGFELLAGIGV AA  FMANN------- :  284
HinSnfT    : -----------------------------------------SGLVVGFA SSFEVLAGIGV AA  FIATA------- :  292
StTnaT     : ---------------------------------------- AFITVFA SSFDFMAGLAV ST  YAAVT------- :  285
CeSNF2     : ---------------------------------------- VYIIVFL TLMSILGGITV ST  FLAYE------- :  456
CeSNF7     : ---------------------------------------- AYLVVLC TLMSLLGGTAV ST  FLAEQ------- :  346
CeSNF4     : ---------------------------------------- AIFLIIG TAMSLIGGAAV ST  YLARE------- :  324
CeSNF5     : ---------------------------------------- ALIVIIG TTMSLVGGGAV AT  YLAKA------- :  425
CeSNF9     : ---------------------------------------- ALIVIIG TMMSLVGGAAV ST  FLAKK------- :  395
DmCG13794  : ----------------------------------------YSWIISFGQIFIFTMFGLVS MLDHYYN--------- :  343
DmCG13795  : ----------------------------------------YSWIISFGQIFIYIMFGMVS MLEHYFN--------- :  339
DmCG13793  : ----------------------------------------YSWIISFGQIFIYIMFGMVS MLEHYF---------- :  250
DmCG33296  : ----------------------------------------YSWIISFGQIFIYIMFGLVS MLEQYYKGK------- :  268
DmCG8291   : -----------------------------------ANKAALRR AILLVLFTLIGLGLMALLALCCAQILWQHGYVYVPG :  813
AgCP9955   : ----------------------------------RTRRELRR AFLVAFITLTVLILAAVLGSAC QILNSRGYYYLPG :  324
DmBLOT     : ---------------------------------------- AVRTSVVYLCFNLLINAIAVTLFMVQFDLS------ :  501
AgCP2093   : ---------------------------------------- AIKTSFVYIFFNILITAIAVVFYVFQYNNSF----- :  298
CeSNF6     : ---------------------------------------- IWLVLIV VIVSLVGCLLT SA  FTCYE------- :  395
CeSNF12    : ---------------------------------------- SVLLVIM ALTSCLSGMVI SF  FIASE------- :  472
AaLeuT     : ---------------------------------------- IVLSGLTAATLNEKAEVILGGS SIPAAVAFFG--- :  307
HsNTT5     : ---------------------------------------- AFLVSVI LLTLLVFTSFN CV  FWATVITHRCCE :  409
DmCG13796  : ---------------------------------------- AERLSIWVCLTHITIGLMGLLCCNVAHDHFED---- :  391
HspSnfT    : -----------------------------------------SLAA GAAIAVL TCIGVLAGLVV PL FALVGD------- :  275
```


CeSNF1 : ---------- SIFLSVWHLIISFVQTMSIICI YLSMK------- : 343
CeSNF8 : ----------VATLAVIL TGLSIIISATV SI YTRFILG----- : 331
* 980 * 1000 * 1020 * 1040
DmSERT : ----------QKTSIDK GLE----------GPG V : 399
AgCP12911 : ----------QNKSIEE GLE----------GPG V : 341
MsSERT : ----------QNKSIEE GLE----------GPG V : 366
HsSERT : ----------RNEDVSE AKDA----------GPS L : 408
CeSERT : ----------TNKPINE VGEH----------DAS I : 433
DmDAT : ----------LGVRIED ATE----------GPG V : 391
AgCP3649 : ----------SGQSIQD ATE----------GPG V : 352
TnDAT : ----------SGRDIQD ATE----------GPG V : 359
CeDAT1 : ----------SGKPIEA AQE----------GPG V : 385
HsNET : ----------HKVNIED ATE----------GAG V : 389
HsDAT : ----------HSVPIGD AKD----------GPG I : 392
TnOctT : ----------QGVPISS ATE----------GPG V Q : 438
DmCG5549 : ----------LNVDVEK VDQ----------GAG A : 422
AgCP10735 : ----------LDTSVDS IDQ----------GAG A : 351
HsGLYT2 : ----------RKVNIEN ADQ----------GPG A : 548
HsATB0+ : ----------SGKEVSQ VKS----------GFD A : 393
HsPROT : ----------LGVPVDQ AKA----------GPG A : 369
HsGLYT1 : ----------LGVDVSR ADH----------GPG A : 442
DmCG7075 : ----------SGIPVES ATG----------GAG A : 349
AgEbiP6595 : ----------TGLPVASAATG----------GPG A : 298
MsPROT : ----------TNVPVSQ ATA----------GPG A : 332
DmIne : ----------QNTNVRD IGD----------GPG I : 379
AgCP4566 : ----------QNTTVEN ISG----------GPG I : 421
AmINE : ----------QNMSVED LTD----------GPG V : 628
MsIne : ----------QNTPVKD IAD----------SPG L : 489
TnGAT : ----------QQRPVAE AAS----------GPG A : 378
MsGAT : ----------QQRPVAE AAS----------GPG A : 379
AgCP8499 : ----------QQRPVAE AAS----------GPG A : 341
DmCG1732 : ----------QQRPVAD AAS----------GPG A : 404
CeSNF11 : ----------QEKSVAE AQA----------GPG L : 397
HsGAT1 : ----------TKRSIAD AAS----------GPG A : 366
HsGAT3 : ----------QGVPIAE AES----------GPG A : 380
HsGAT2 : ----------QGVPISE AES----------GPG A : 360
HsBGT1 : ----------QGVPISE AES----------GPG A : 365
HsTAUT : ----------QGVDIAD AES----------GPG A : 372
HsCT1 : ----------QGVHISK AES----------GPG A : 387
CeSNF3 : ----------AQKDISE VKP----------GVG A : 364
DmCG5226 : SNTEILVKNKLLGLND----------TQG--YEQAMSLLNGTELLR QLSE-----CSLAHELDNAAEGTG A : 418
AgCP7001 : VNKHILVKHNLLPSMD----------VDQSE-YEHAMAGLNETMTKDWEFEK-----CSLEDQLSSAAEGTG A : 425
HsNTT7-3 : QNSETIMKFLKMGNISQDIIPHHI--NLSTVTAED-YHLVYDIIQKVKEEEFPALH--LNSCKIEEELNKAVQGTG A : 444
HsNTT4 : ENAEKILGYLNTNVLSRDLIPPHV--NFSHLTTKD-YMEMYNVIMTVKEDQFSALG--LDPCLLEDELDKSVQGTG A : 443
HsXtrp2 : RNILSLINDFDFPEQS----------ISRDD-YPAVLMHLNATWPKR AQLP--LKACLLEDFLDKSASGPG A : 387
RnROSIT : RNILSLINEFDFPELS----------ISRDE-YPSVLMYLNATQPER ARLP--LKTCHLEDFLDKSASGPG A : 374
HsBOAT1 : TNILTLINGFDLPEGN----------VTQEN-FVDMQQRCNASDPAAYAQLV--FQTCDINAFLSEAVEGTG A : 401
HsXT3L : KVSLLLTNTFDLEDGF----------LTASN-LEQVKGYLASAYPSKYSEMFPQIKNCSLESELDTAVQGTG A : 376
DmCG10804 : ERNGLVAQNKTHN-LP----------VCDLQTELANVSSLA CLHG----IYTYIYLPFQSASGTG A : 418
AgCP7501 : ERSEMIRLNKSHDLLP----------VCDLQKELENVTPRR HRSSSFTIISVFPHLLLQSASGTG A : 372
DmCG1698 : ----------NNTTDIAS VNG----------GPG A : 405
DmCG4476 : ----------SGTKDIAS VKA----------GPG A : 379
DmCG15279 : ----------IGTDDIGS VKG----------GAG A : 393
AgCP10412 : ----------TGKTDVGN VKS----------DAG A : 409
DmCG8850 : ----------TNTKDISQ VKG----------GAG A : 372
DmCG15088 : ----------TGNPDIGS VKG----------GAG A : 337
AgCP10401 : ----------LEAPDIKH VRG----------GAG A : 411
AgCP10503 : ----------TGNR-VED QLQ----------GPQ T : 366
AgCP10610 : ---------- : -
DmCG3252 : ----------LQIENIRD VRS----------GTG A : 386
AeAAT : ----------LGIEDISK VKS----------GTG A : 374
MsKAAT1 : ----------LNSDVGD IGSG----------GTS A : 378
MsCAATCH1 : ----------LNSEVGD VGAG----------GTS A : 378
BcProT : ----------MGVPVDK ASAG----------VG A : 302
HinSnfT : ----------QGQEVSE AKGG----------IG A F : 310
StTnaT : ----------AGVPFEE AVAG----------PG A : 303
CeSNF2 : ----------RGINSTD DAFN----------HIVKEGHA A : 480
CeSNF7 : ----------RNISISNPKGFN----------EVVQEGHA A : 370
CeSNF4 : ----------RGVTVD----------KVVTDGTT A : 342
CeSNF5 : ----------TGQDVK----------DVVKSGLS A : 443
CeSNF9 : ----------KGCSVA----------EVVNDGFS A : 413
DmCG13794 : ----------EGIEGSFDKQ LTH----------WAL : 362
DmCG13795 : ----------ELSDVDDNGH LSH----------W L : 358
DmCG13793 : ----------L INH----------W L : 260
DmCG33296 : ----------ADFLLFHHISK MNF----------W L : 288
DmCG8291 : SFENPDCYTSIYSLQSNTNPYLLS--YPRSLIPHYSSFIGETYRRNRTMIH ESG----------FQALRFISE : 875
AgCP9955 : SYENLGTNVFLRPANQPLPPQHAT--MPNRWLLRYSTVVGESFKRSYADPSRESG----------YQALR VTE : 386
DmBLOT : ----------SNGFQNMEELK----------PLTA : 516
AgCP2093 : ----------TEAHPFYPELS----------ALT : 313
CeSNF6 : ----------FAISLDKFHIRD----------GFH V : 414
CeSNF12 : ----------SNSNVND LKHDP----------LY S T : 491
AaLeuT : ----------VANAVAIAKAG----------AFN G : 325
HsNTT5 : RNAEILLKLINLGKLPPDAKPPVNLLYNPTSIYNAWLSGLPQHIKSMVLRE TECN-------IETQFLKASEGPKFA : 482
DmCG13796 : ----------HVGMMPLH DEKH----------HMQFL : 411
HspSnfT : ----------AGTG----------GPGAL : 286
CeSNF1 : ----------LGIHPTE LETG----------EDQ : 359
CeSNF8 : ----------RKSVYDLT LPN----------GLL THS : 350

```
                        *      1060         *      1080         *      1100         *      1120
DmSERT     : V PE  ATM------SGSVF SI   L  IT   D TFGGLEAMI A C EYP--RVIGR---RRELFVLLLLAFIFLCA :  468
AgCP12911  : V PE  AMM------KGSVF SI   L  IT   D TFGGLEAMI A C EYP--RTIGR---RRELFVLILLGLIYICA :  410
MsSERT     : V PE  ATM------TGSVF AI   L  IT   D TFGGLEAVT A C EYP--RVLGR---HREVFVAVLLLFIYICA :  435
HsSERT     : T AE  ANM------PASTF AI   L  IT   D TFAGLEGVI A L EFP--HVWAK---RRERFVLAVVITCFFGS :  477
CeSERT     : V PQ  ATM------DYSCF SF   V  IT   D TFAGIEAFI GFC ESR---FLSKN---RKWFVLVICIIYYFLS :  501
DmDAT      : V PA  ATM------PASTF AL   M  LT   D SFGGSEAII A S EFP---KIKRN---RELFVAGLFSLYFVVG :  459
AgCP3649   : V PA  ATM------PGSIF AL   M  LT   D SFGGSEAII A S EFP---KIGRN---REIFVAVLFSLYFVVG :  420
TnDAT      : V PA  ATM------PGSTF AL   M  LT   D SFGGSEAII A S EFP---PIGRH---REIFVACLFTLYFIVG :  427
CeDAT1     : V PE  STM------PYAPF SV   L  MT   D SFGGSEAII G S EFP---ILKKN---REVFVGCLFAFYMVIG :  453
HsNET      : L PE  STL------SGSTF AV   V  LA   D SMGGMEAVI G A DFQ---VLKR---HRKLFTFGVTFSTFLLA :  457
HsDAT      : I PE  ATL------PLSSA AV   I  LT   D AMGGMESVI G I EFQ---LLHR---HRELFTLFIVLATFLLS :  460
TnOctT     : V PE  ATL------PGASL AM   F  IM   D GMGGLECVI G L QAR---ACGATWLRREHFTLIVVCVSFCVA :  509
DmCG5549   : V PEV TRL------PVSPV AV   V  LT   D QFALMETVT A L RFP---NLRQY---KIWVVLSVAIFGYIGG :  490
AgCP10735  : V PEL GKL------QMPRL SV   F  LT   D QFTLMETVV A L TFP---ALRRR---KIAVVAGVSIVGYLGG :  419
HsGLYT2    : V PE  TRL------PLSPF AI   L  LT   D MFATIETIV S S EFP--KYLRTH---KPVFTLGCCICFFIMG :  617
HsATB0+    : A PE  AQL------PGGPF SI   F  LT   D QFASIETIT T Q LFP--KVMKKM---RVPITLGCCLVLFLLG :  462
HsPROT     : V PQ  TML------PLSPF SF   F  LT   D QFAFLETIV A T EFP--YYLRPK---KAVFSGLICVAMYLMG :  438
HsGLYT1    : A PE  TLL------PISPL SL   F  IL   G QFCLLETLV A V EVGN-EWILQK---KTYVTLGVAVAGFLLG :  512
DmCG7075   : T PE  ALL------PVPQL AL   F  FL   D VFVQLEAIM S L EWY---WVRSH---KCKLTLISCLIFLGLS :  417
AgEbiP6595 : T PE  SML------PLSPL AC   S  FF   D MFVQIEAIV S M VLP---RLRPM---KLQVTAGTCFLLFLLS :  366
MsPROT     : T PAT TMM------PAPNL AIT  V  FF   D MFVTIEAIIAG L EFP---RFKSR---KRIIAFITCVVLFSFS :  400
DmIne      : V PQ  AKM------PYAQL AV   F  LC   N QFAIVEVVV S Q GFP--RWIKRHLGYHEIVVLFVCVISCLFG :  451
AgCP4566   : V PQ  AKM------PAAQL AV   F  LC   N QFAIVEVVV S Q GFP--RWIKRKLVYHELLVLIVCVVSFFAG :  493
AmINE      : L PQ  AKM------PASQV AV   F  VC S N QFAIVEVVV S Q GFP--KWVKRHLTCHEMLVLLICVISFLFG :  700
MsIne      : V PQ  AKM------PASQL AV   F FLC   N QFAIVEVVV S Q GFP--DMIRKRLVYHELLVLLVCAVSLLCG :  561
TnGAT      : A PS  LQL------PGAPL SC   F  LL   D QFCTMEGFI A I EWP--KLLRRR---KEIFIAITCVISYLVG :  447
MsGAT      : A PS  LQL------PGAPL SC   F  LL   D QFCTMEGFI A I EWP--KLLRRR---KEIFIAITCIISYLVG :  448
AgCP8499   : A PS  LQL------PGAPL SC   F  LL   D QFCTMEGFI A I EWP--HLLRKR---KEIFIAIVCMISYLVG :  410
DmCG1732   : V PS  LQL------PGSPM SC   F  LL   D QFCTMEGFI A I EWP--QLLRKR---KEIFIAIVCALSYLVG :  473
CeSNF11    : A PSG LQL------PYTQF SC   L  LF   D QFCTMEGFF A I EFP--QIRRKKY-GREIFVGVICVISYLIG :  468
HsGAT1     : A PE  TQL------PISPL AI   S  LM   D QFCTVEGFI A V EYP--RLLRNR---RELFIAAVCIISYLIG :  435
HsGAT3     : A PK  TMM------PLSPL AT   M  IF   D QFVCVESLV A V MYP--KVFRRGY-RRELLILALSVISYFLG :  451
HsGAT2     : A PR  VML------PFSPL ACC  F  VL   D QFVCVESLV A V MYP--HVFRKKN-RREVLILGVSVVSFLVG :  431
HsBGT1     : A PK  TMM------PLSQL SC   I  IF   D QFVCVECLV ASI MFP--RQLRKSG-RRELLILTIAVMCYLIG :  436
HsTAUT     : A PK  TMM------PLPTF SI   I  LL   D QFVEVEGQI S V LYP--SFLRKGY-RREIFIAFVCSISYLLG :  443
HsCT1      : A PR  TLM------PVAPL AA   F  LL   D QFVGVEGFI G L LLP--ASYYFRF-QREISVALCCALCFVID :  458
CeSNF3     : A PEVASNL------PMKQV AV   L  TI   D QVCMMEGLF A E AFP--ILRKYK----KQSLGIFCLFFFCIG :  432
DmCG5226   : V TQ  VEL------PGAPF AV   T  LS   G QIGILEGMLCT F ID------IIKRVKKQHVTGVVCLFCFIVG :  486
AgCP7001   : V TE  VEL------PGSPF AV   T  LS   G QIGILEGMLCT F ID------IFKRIRKPYVTGVVCAFCFSVG :  493
HsNTT7-3   : A TE  THF------PASPF SV   L  VN   G MFGTIEGIV P V T---------FKVRKEILTVICCLLAFCIG :  509
HsNTT4     : A TE  THF------PASPF SV   L  IN   G MIGTMAGIT P I T---------FKVPKEMFTVGCCVFAFLVG :  508
HsXtrp2    : V TETDLHM------PGAPV AM   G  FT   S MFGTVEALV P L VGV-----LPRWVPKEALTGLVCLVCFLSA :  456
RnROSIT    : V TE  LHM------PGASV SV   G  FT   S MFGNMEGVI P F MGI-----LPKGVPKETMTGVVCFICFLSA :  443
HsBOAT1    : V TE  TKM------PLSPL SV   I  FC   S MFGNMEGVVVP Q LRV-----IPPKWPKEVLTGLICLGTFLIG :  470
HsXT3L     : V TE  KNM------EVSQL SV   F  LM   G MLGNTAAIL P T SKI-----ISSHLPKEAISGLVCLVNCAIG :  445
DmCG10804  : I TE  NQF------PGAQL AV   L  FT   D QFGTLEGVV S V MK------LFPNLPKEYIVGALCFSCCTIS :  486
AgCP7501   : I TE  NQF------PAAQL AV   L  FT   D QFGTLEGVS S M MK------LFPNVPKEMITGALCMSCCVLS :  440
DmCG1698   : S PD  AKFK-----WLPQL SV   L  FV   G NVGMASCMS V K QFG---HLKNWT-----VVVGIAIVGYFLG :  472
DmCG4476   : S PD  AKFK-----MFPQV SL   A  FM   G NVGMVSCIM V K QFV---NVKLWI-----IVVSLSVIGFLVG :  446
DmCG15279  : S PD  AKFK-----NLPQI SV   L  FV   G NIAMTSCSV A R RFP---NFGQWQ-----CSLLIAVVSFFIG :  460
AgCP10412  : S PE  AKFE-----VLPQA SV   L  FV   G NVAMTSCVM V K QFP---RVRNWQ-----AATIIAICGVLLG :  476
DmCG8850   : S PE  AKFK-----YLPQL AV   F  LV   G NIGMASAVVNV K RF---THLPHWL-----LAVSASIIGFLCG :  439
DmCG15088  : S PE  AKFE-----YVPQM AV   L  FV   G TVGMGSCILRV R QFGL-RSPPIWK-----LASGLTVLGFSVS :  406
AgCP10401  : T PD  AKFT-----FWPQF AIA  L  FV   G NVGMATTIM V R RFP---QLQPAL-----VAVGIAIVGYGIG :  478
AgCP10503  : T PD  AKFD-----AAQNV AV   L FFL   G NTGIVTTIV A R RFP---ELPNWK-----VVISIAIYGFCCG :  433
AgCP10610  : ---------------------------------------------------------------------------- :    -
DmCG3252   : S PD  SKFQ-----AVPQL SV   F  FV   G IVALQSTIV I C QFK---GWKYWK-----VALTTSVCGFLMG :  453
AeAAT      : S PD  AKFD-----IVPQL SV   F  FV   G AVALHGAII AFW AFP---KRKYWH-----LALILSTIGFCTG :  441
MsKAAT1    : S PD  AKT------FQPQL SV   L  SV   G SVALLSTLN V M AFP---RVPTVY-----MSALSCTCGFLLG :  444
MsCAATCH1  : S PD  AKT------FQPQL SV   L  SV   G SVALLSTFN LAM AFP---RVPTVY-----MSAMTCSCGFLLG :  444
BcProT     : V PQI NELP------MSPL GV   LS TGA  T LISLAEVCFAA SEKFS---------LSREKTIGIMGTLLVLVS :  367
HinSnfT    : A PTI NKAP------FGEVLGM   GA TFAA T FISVIEVII A Q KIR---------ISRGKVTFIVGVPMMLVS :  375
StTnaT     : A PK  SMLPGPT--WLQSL GI   SA LLA  S SISQMESFA A I RFG---------VDRKKLLGWFSLIGFAFS :  372
CeSNF2     : V TE  AEMP------VPYL YA   I  FL   S EVVIVEIVC C A RFR------YLRERRWLTVLSVSAVFFLLG :  548
CeSNF7     : V TE  AQMP------YPFL YA   V  FL   S EIVIIETVC C A RFH------YLRQHRWITVVTVSTTFFLLG :  438
CeSNF4     : A PD  NQMP------FPGL SF   F  FL   S LLVIVEVMC C C KFP------RLREKRVFVVIGVCFLLFLCG :  410
CeSNF5     : V PE  TRMP------VPWL CF   L  FL  AS EIALVDVFC C Y QYP------RFRNRKWIVVIAWCSVLYCIG :  511
CeSNF9     : A PE  GRMP------LPEL SF   L  FF   S EVAYVNVFC S C QFV------NLRKRKWLVAVLWCLVLYVLG :  481
DmCG13794  : SSAT  SSLG------WPNL TF   T  LMAA IVITTQIFTVLQS F EFEVLRVRKQE------VTFGLIGGIAVCS :  430
DmCG13795  : SSAS  STMA------WPNL TF   S  LIAA IVMTTQIFTVLQS F EFEVLRVRKQE------VTFGLIGGIALCS :  426
DmCG13793  : SSAS  SSLG------WPNL TF   T  LMAA I--TTQIFTILQS F EFEQLRAKKQE------VTLGLIGGLAVCS :  326
DmCG33296  : SSAS  NTLG------WPNL TF   S  LLAS T--ATQLFTVLQS F EFEKLKERKQE------VTFGIICGLAICS :  354
DmCG8291   : I PAV SLAS--D--SISWV AA A ATFAGF  AQLCVMWKPIS A G STS--------------SVLLSCVTGLLLS :  937
AgCP9955   : LLPA  AAATP-E--HIAPI GM G LA VLF  GQLCVMWKPIAGA G APS--------------SILLSCVTGLLLG :  449
DmBLOT     : I DRV NGSREGDSHLLQHLVPS I A  ILSA V ITVAVYTSTRL PRRPN----------------YVICLIGLVVA :  580
AgCP2093   : I DR  S-VQESD-PTLQRLIPA S T  FISS V IVIYIYTSTRI RKHPN----------------YTVCLAGLVVA :  375
CeSNF6     : FLAE  AGVS------VAPL AG   I  LL VHA QMFVVETIV S C EYP-----ERLRRNRRHVLTTVCALFILLS :  483
CeSNF12    : V PGVTSFMY------WGGL AT   G  VLAA DAEFAWLEMIA AFM HFS-----MKNKAVENRLLAFLCLAGFFLG :  560
AaLeuT     : TLPAIFSQTAG------GTFLGF   F  FFA  T SIAIMQPMIAF E ELK---------LSRKHAVLWTAAIVFFSA :  390
HsNTT5     : S VE  SFLP------PSVF SF   L  LA   S AIGIMQGII P Q TFS------FFRKHTKLLIVGVFLLMFVCG :  550
DmCG13796  : C SYLFGRFT-----TTPNL AF   G  FLSE CALIIQMMSVI A F EFE-------TLRSRRILVISSLVFCLTAS :  479
HspSnfT    : GLAN FVAVP------YGRAVGA   A  SLAA S SISMLELPV F V EYG---------------VSRRRATALFVG :  345
CeSNF1     : W ML YVSY-------LPKL SG LLT CICTMFSVITILALSVL TFE AFG---ENWSKCCRRFFLALFVCLFCFGLT :  429
CeSNF8     : T PVL RQVA-----ANSWL VF   ST VV  WK CALIMKTITVT A IFN----IKMKTYFFDCLCIFVSMVIATSG :  421

                        *      1140         *      1160         *      1180         *      1200
DmSERT     : - PTMTYG VVLVNFL V GPG-LAILFVV--FV AAG F F  VDRFSS  EQ L SKPGLF----------------- :  527
AgCP12911  : - PTMTYG VYLVNFL V GPG-LAILFVV--FV AAG F F  VENFSA  EQ L KKPSLF----------------- :  469
```

```
MsSERT     : - PTTTYG VYLVD L V GPG-LAILFVV--FA AAG C V  VDRFSE  RT L HTPGWF----------------- :  494
HsSERT     : - VTLTFG AYVVK LEE ATG-PAVLTVA--LI A A S F  ITQFCR  KE L FSPGWF----------------- :  536
CeSERT     : -FPAISYG QFVIPFL E GVS-LSVLFIV--TC M A C F  VDQFSK  RA L FYPGIY----------------- :  560
DmDAT      : - ASCTQG FYFFH L R AAG-YSILVAV--FF A A S I  TNRFSE  RD I FPPGRY----------------- :  518
AgCP3649   : - ASCTQG FYFFQ L R AAG-YSILIAV--FF A A S I  TERFCN  KD I FAPGIY----------------- :  479
TnDAT      : - TSCTKG FYFFQ L R AAG-YSILVAV--FF A A S I  TERFCE  RD I FRPGLY----------------- :  486
CeDAT1     : - AMCTEG ILIMEWLII GTT-WGLLIAV--FC A V A I  LRQFVH  KE M FRPGNY----------------- :  512
HsNET      : - FCITKG IYVLT L T AAG-TSILFAV--LM A G S F  VDRFSN  QQ M FRPGLY----------------- :  516
HsDAT      : - FCVTNG IYVFT L H AAG-TSILFGV--LI A G A F  VGQFSD  QQ T QRPSLY----------------- :  519
TnOctT     : -CINVTPG IYMFH L T AAG-ISLLCSA--LF A A S F  LKRFSD  EE L FRPGLY----------------- :  568
DmCG5549   : - GFTTNS MYWLQ M K AAN-WSVLLIA--IS C L A I  SQRFLN  QG I KRSWFWN---------------- :  550
AgCP10735  : - IFVTNS MYWFQ F K AAN-WSVLLIA--IA C L G V  CERFIS  EE L KRSKWFH---------------- :  479
HsGLYT2    : -FPMITQG IYMFQ V T AAS-YALVIIA--IF L G S V  LQRFCE  EM I FQPNIF----------------- :  676
HsATB0+    : - VCVTQA IYWVH I H CAG-WGILIAA--IL L G I I  GNRFIE TEM I AKRWIFW---------------- :  522
HsPROT     : - ILTTDG MYWLV L D SAS-FGLMVVV--ITTC A TRV  IQRFCR  HM L FKPGLY----------------- :  497
HsGLYT1    : - PLTSQA IYWLL M N AAS-FSLVVIS--CIMC A M I  HRNYFQ  QM L FPPPLF----------------- :  571
DmCG7075   : -SIMCTNG MFILQ F W SSA-IAVIVIC--LV I M A I  IKNFML  EF L KRPTLY----------------- :  476
AgEbiP6595 : - ACVTQG MYLVQ L W SGS-ISVILVC--IV V T G I  CEQFVR  EF IDRKVERF----------------- :  425
MsPROT     : - ICNTEG LHVIG L SHVAI-LCVPLVC--AL I AAV T --ENFSF  LF T RPLRRI----------------- :  457
DmIne      : - PNIIQG IYYFQ M H AAS-VTIMFLA--FC M A A F  TGRLSK  KQ T KAPSFY----------------- :  510
AgCP4566   : - PNLIQG IYFFQ I H AAS-VSIMFLA--FF T A A F  INRLSK  KQ T RYPSFY----------------- :  552
AmINE      : - PNISQG IYFFQ I H AAS-ISIMFLA--FF V A S F  VRRLCN  KE T RVPSSY----------------- :  759
MsIne      : - PHIIHS IYVFQ M Y AAS-LSITYLA--FF V A A F  VGRLSR  KQ T RQPSLY----------------- :  620
TnGAT      : - SCISEG MYVFQ L S AVSGFCLLFLI--FF C S S A  VNRFYDG KE I YYPTIW----------------- :  507
MsGAT      : - SCISEG MYVFQ L S AVSGFCLLFLI--FF C S S A  VNRFYDG KE I YYPTIW----------------- :  508
AgCP8499   : - TCITEG MYVFQ L S AVSGFCLLFLM--FF C S S A  VDRFYDG KD I YYPLSW----------------- :  470
DmCG1732   : - TCITQG MYIFQ L S AVSGFCLLWLI--FF C S S C  VDRFYDG KD I YYPTVW----------------- :  533
CeSNF11    : - TTVTEG FYVFQ F F AASGWALLWLL--FF C A S SL IDRWYEH KS I YYPSAW----------------- :  528
HsGAT1     : - SNITQG IYVFK F Y SASGMSLLFLV--FF C S S F  VNRFYD  QE V SRPCIW----------------- :  495
HsGAT3     : - VMLTEG MYIFQ F S AASGMCLLFVA--IF C C G V  SNRFYD  ED I YRPPSL----------------- :  511
HsGAT2     : - IMLTEG MYVFQ F Y AASGMCLLFVA--IF S C A V  AKRFYD  ED I YRPWPL----------------- :  491
HsBGT1     : - FLVTEG MYIFQ F Y ASSGICLLFLS--LF V C S V  ADRFYD  ED I YRPWPL----------------- :  496
HsTAUT     : - TMVTEG MYVFQ F Y AASGVCLLWVA--FF CFV A I  GDNLYDG ED I YRPGPW----------------- :  503
HsCT1      : - SMVTDG MYVFQ F Y SASGTTLLWQA--FW C V A V  ADRFMD  AC I YRPCPW----------------- :  518
CeSNF3     : - PMVTHS SHWLT F A GASGYALLFVV--FF V G A G  AHNIRKA HE I VTLPKG----------------- :  492
DmCG5226   : -FIFCTGA EYWLK F S AGT-IGLVVVA--LM M A I I  HERFTE  FQ T YRPGRY----------------- :  545
AgCP7001   : - IFTTGA EYWLK F S AGT-IGLVVVA--LM M A I I  HEKFTE  YQ T IRPGLY----------------- :  552
HsNTT7-3   : - IFVQRS NYFVT F D SAT-LPLLIVV--IL N A C V  IDKFME  KD L FAPSRY----------------- :  568
HsNTT4     : - LFVQRS NYFVT F D SAT-LPLTLIV--IL N A A I  TKKFMQE TE L FRPYRF----------------- :  567
HsXtrp2    : -TCFTLQS NYWLE F N AAS-LNLLMLA--FL V G V V  MKRFCD  AW T RRPSPY----------------- :  515
RnROSIT    : - CFTLQS SYWLE F S AAS-LNLIIFA--FM V G IHV  IKRFCD  EW T RRPSLY----------------- :  502
HsBOAT1    : -FIFTLNS QYWLS L S AGS-IPLLIIA--FC MFS V V  VDRFNK  EF I HKPNIF----------------- :  529
HsXT3L     : - VFTMEA NYWFD F D AAT-LSLLLIV--LV T A C V  LRRFES  KA T RAVSWY----------------- :  504
DmCG10804  : - CFANGA SYIFQ M S AGN-FPLLIIA--LF C S S I  VRRFSD  EM T SRPNFY----------------- :  545
AgCP7501   : - CFANGA SYIFQ M S AGS-YTLLIIA--FF C G S I  LKRFAD  EL T SRPSLY----------------- :  499
DmCG1698   : - LYITPG QFLLN V Y GVT-FVALVLA--IF L T A I  VKRLCR  EF I IKTSLY----------------- :  531
DmCG4476   : - IYITPG QHIIT M FHGVT-FVSLVSA--IF L A G I  TKRLCQ AEY LNIKTSNY----------------- :  505
DmCG15279  : - MYITPG QYMLT V F GAS-MIALVLG--IA LYT G I  TDRLCK  EF L RKVGLY----------------- :  519
AgCP10412  : -SIYVTPG QYVLK V Y GAS-SIALVLA--IA L A G V  VDRLCK TEF L HRPNLY----------------- :  535
DmCG8850   : - VYMTPG QFVLN V F GCT-FIALVLA--IA L A G I  VKRICS  EF LNVKTSFY----------------- :  498
DmCG15088  : - VYMTPG QFILN V F GVS-FTALILA--IG L A A I  VNRFCE  KF M IETGWY----------------- :  465
AgCP10401  : - IYTTPG QYVLDFL F GAS-FVALVLA--VF MFTFA I  VSRICR  EF L IKTGLY----------------- :  537
AgCP10503  : - VYITP---------------------------------------------------------------------- :  439
AgCP10610  : ---------------------------------------------------------------------------- :    -
DmCG3252   : - VYVTPG QWILT V F GGT-YVVFILA--IF LAG V V  LQNFCD  EF CNRRVSLY----------------- :  512
AeAAT      : - VYITPG QWILD V H GGT-FLIYVLA--II M A F I  LDNWCN  EF VQRRVGLY----------------- :  500
MsKAAT1    : - VYCTPG QFILE V H GGT-FLVLFCA--IS LAG F I  LENLCL  EF M KTTGFY----------------- :  503
MsCAATCH1  : - VYCTPG QYILE V H GGT-FLVLFCA--IS LAG F I  LENLCL  EF L KKTGAY----------------- :  503
BcProT     : - VFATRG LMFLY G Y ANN-FGLITIA--LA V T GLILRRLPVYQ HANFVSDIKLRT----------------- :  426
HinSnfT    : V LFGTTT LPMLD F K VNY-FGIVAVA--FASL A -VANEKLGLLG HLNETSSFKVGF----------------- :  434
StTnaT     : -ALFATGA VHILD V H VGS-YAIAILG--LV A V G IM TARIR-EHVN TSDIRVGM----------------- :  430
CeSNF2     : - VMTTDA FYWFD F E STG-ISATFGT--AIMC T F C  TNHFRE  AV W EPESWLATCL----------GPA :  614
CeSNF7     : - VMTSDA FYWFD Y E SAG-VSATIGT--AIMC T C C  LDNFRA  RE WEEPESRLDKYL----------GPG :  504
CeSNF4     : - ILTTDA IYWFA F E GSG-FGALISA--TSMC I S I  MRNFRA  AS L RG-TRISNIF----------GHN :  475
CeSNF5     : - VFSTRA YYWFE F E AAG-FSSVCTV--VC L V M I  FRNVRD  TE V HARNKFTGAI----------GAH :  577
CeSNF9     : - VMVTDG FYWFI F E AAG-VSSCCAV--TA V I A V  RRNQLA  KE F ESKNKCTSWI----------GQH :  547
DmCG13794  : -FYFCTNH ITFFSALALDAIF--SHSLLH--LLLL V L I  RERFQR  EF L QPFASWK----------------- :  489
DmCG13795  : -FYFCTNH IIFFSALGLDAIF--SHSLLH--LLLL V L I  RERFQR  EF L QPFASWK----------------- :  485
DmCG13793  : - YFCTNH VLYFA LSIDAIF--SQTLLH--LLLL V L V  RDRFQK  KF L QPFASWK----------------- :  385
DmCG33296  : -FFFCTEH SFYFA LGIDVIL--AHSLLH--LLLL V L I  QKRFQR  EF L QPFASWK----------------- :  413
DmCG8291   : - PFATEM ISILYYV FLLGGSWFIPIIW--TA IFG FLIR RPYNGD  VNDLRLCGSMS----------------- :  997
AgCP9955   : - PLATEG IHIVHYL TILGAAWWLLLLW--IGHI A FLVR RPFTSDI VNDLQMLQSFS----------------- :  509
DmBLOT     : - ISFAAPKFLIAR L SRLVG-TMVVTAL--VF L A T I  AKNIYT  EFSI RPIFRV----------------- :  639
AgCP2093   : - AGLLCPNYIFPR L TRIVG-ALAVCAM--IF I L I V  SKNLYT  EFSL RPVLKA----------------- :  434
CeSNF6     : - PFCLSS LFWME LTQ VLT-WPLVVIA--FL C A N V  VDNMLD AKW V YWPPCY----------------- :  542
CeSNF12    : - PLCAQG IFVFHAIENLNAN-WNSFSLA--LLSVAI C V  IDNYLT  SA LRVPRIQISKATRLKEKLIYFFGPG :  636
AaLeuT     : H VMFLNKS---LDEM F AGT-IGVVFFG--LT L IFF I  ADKAWEE NRGGIIKVPRI----------------- :  447
HsNTT5     : - FFTRPS SYFIR LSD WIVFPIIVVVV---F T A S A  ARRFLA  TI L HPISPI----------------- :  609
DmCG13796  : S YFCTQL FSQLTQLPNLAVF-THLLISG---ILV MTT V  RVRFQC  QF L KTISSFK---------------- :  539
HspSnfT    : VFTVTGSATALSPG F FVAGTLVDLLMTAGLIGFL FTG VV QNAVTEYELGAGAVASALSTP-------------- :  410
CeSNF1     : - YFATNA RHAYE ATRSIKYCTIYFILT---A LFATA F CAHKLGK  HN MRSRCCSCFG-------------- :  490
CeSNF8     : - FLCFES PLYLT IQDIQTVATGIFITA---- LFI LEL  LDTIYG  HE M QPSNDQWSR----------AYT :  485

                        *      1220         *      1240         *      1260         *      1280
DmSERT     : ---WRIC TYIS VFLLT FIFSIMG------YK--------------EMLGEEYY  DWSYQVGWA TCSSV C  MYI :  584
AgCP12911  : ---WRIC KYIS TFLFC LIFSLLG------YE--------------EMLGEEYE  EWSVAAGWA TLSSV C  TYM :  526
MsSERT     : ---WRTC SYIS VFLLV FVFSVLA------HE--------------EMLGGEYT  SWSITVGWV TGTTVSC  LYI :  551
HsSERT     : ---WRIC VAIS LFLLF ICSFLMS------PP--------------QLRLFQYN  YWSIILGYC GTSSF C  TYI :  593
CeSERT     : ---WRVC T-CS VFISV FIMTVYNSS----FK--------------PIQMASYT  WWSVILGWF RLLSV A  VFA :  618
```

```
DmDAT      : ---WQVC RFVA IFLLF TVYGLIG------YE--------------PLTYADYV  SWANALGWC AGSSV M  AVA :  575
AgCP3649   : ---WRVC KFVA LFLLF IIYGLIG------HE--------------PLSYEDYV  SWANVLGWC AGSSM M  LMA :  536
TnDAT      : ---WRVC RFVA AFLLF TAYGLLD------WE--------------PLRYEGYV  GWANVLGWA AGSSV C  TVA :  543
CeDAT1     : ---WKFC SCAA FILLS ITSNFIN------YQ--------------ALTYQDYT  TAANVIGIIFALSGASF  LVG :  569
HsNET      : ---WRLC KFVS AFLLF VVVSIIN------FK--------------PLTYDDYI  PWANWVGWG ALSSM L  IYV :  573
HsDAT      : ---WRLC KLVS CFLLF VVVSIVT------FR--------------PPHYGAYI  DWANALGWV ATSSMAM  IYA :  576
TnOctT     : ---WRIC KFVS TFIIG VVFGLLY------QQ--------------PLQYQQYT  PWAVVLGWG ACSSI M  VVG :  625
DmCG5549   : -FFWGIM RYIT ATLLF LFFNWVE------YK--------------PAKYGHYV  MWADAVGWI GLLPV V FLVA :  609
AgCP10735  : -RFWVTV KYIT AMLLF LCFNWLQ------YK--------------PVAYGRYI  NWADVFGWI ALAPFGIAVIVA :  538
HsGLYT2    : ---WKVC AFVT TILTF LCFSFYQ------WE--------------PMTYGSYR  NWSMVLGWL LACSV W  IMF :  733
HsATB0+    : -LWWRAC FVIT ILLIA FIWSLVQ------FH--------------RPNYGAIP  DWGVALGWC IVFCI W  IMA :  581
HsPROT     : ---FRAC LFLS ATLLA LVYSIVK------YQ--------------PSEYGSYR  PWAELLGIL GLLSC M  AGM :  554
HsGLYT1    : ---FQIC RFVS AIIFF LVFTVIQ------YQ--------------PITYNHYQ  GWAVAIGFL ALSSV C  LYA :  628
DmCG7075   : ---WRIV QVVT VIIIF LITSIVF------IR--------------TITYNNIP  QWAIIIGWASCFTSV W  LYI :  533
AgEbiP6595 : ---WIVS KYIT IVLTF FTTTIVY------NT--------------EVSYNGFS  RWAIGVGWASCLVSI C  VFA :  482
MsPROT     : ---WMVL RYVILLILLV TLYSLLE------VS--------------SLA-------------GWF TLVSV C  IYA :  501
DmIne      : ---LRSC LVLG CLLFA WVLSLINYH----EP--------------TYHNGRYT  DWAYGIGWMFASFSL C  GYA :  569
AgCP4566   : ---LRFCLLIAV LLLIS WIFSLINYE----AP--------------TYHNGKYH  GWAHGLGWT ASASL C  SYA :  611
AmINE      : ---FRFC LIAA FLIMA WVFSLIDYE----PP--------------TYHNGEYK  WWAEAIGWG ASLSL C  AFA :  818
MsIne      : ---FRFC LIAS ALLLA WVASMVDYT----PP--------------SYR--QYQ  AWAQALGWI ASLSL C  VYA :  677
TnGAT      : ---WKFC VGFT AICIS FIFNLVQWK----PI--------------KY--MNYE  WWSHAFGWFTALSSM C  GYM :  564
MsGAT      : ---WKFC VGFT AICIS FIFNLVQWT----PI--------------KY--MNYE  WWSHAFGWFTALSSM C  GYM :  565
AgCP8499   : ---WKFC VITT AICV------------------------------------------------------------ :  484
DmCG1732   : ---WKFC CVTT AICLG FFFNIVQWT----PV--------------KY--LDYS  WWAHAFGWFTALSSM Y  LYM :  590
CeSNF11    : ---WKFC VFAT SVCFG LLFGLIKYQ----PL--------------RIDAYNYD  VWGHIFGWF SLSSM C  GYA :  587
HsGAT1     : ---WKLC SFFT IIVAG FIFSAVQMT----QL--------------TMG--NYV  KWGQGVGWL ALSSM L  GYM :  552
HsGAT3     : ---IKWC MIMT GICAG FIFFLIKYK----PL--------------KYN-NIYT  AWGYGIGWL ALSSM C  LWI :  569
HsGAT2     : ---IKYC LFLT AVCTATFLFSLIKYT----PL--------------TYN-KKYT  WWGDALGWL ALSSM C  AWS :  549
HsBGT1     : ---VKIS LFLT GLCLATFLFSLSKYT----PL--------------KYN-NVYV  PWGYSIGWF ALSSM C  LFV :  554
HsTAUT     : ---MKYS AVIT VLCVGCFIFSLVKYV----PL--------------TYN-KTYV  NWAIGLGWS ALSSM C  LVI :  561
HsCT1      : ---MKWC SFFT LVCMG FIFNVVYYE----PL--------------VYN-NTYV  WWGEAMGWAFALSSM C  LHL :  576
CeSNF3     : ---IEYV KFCA ATSLV FVFCVVYYH----PV--------------KYP-DGKD  FWANAFGWF SSCSM V  GYA :  550
DmCG5226   : ---WQWT RYIG VIMVC LVSSVVFMV----IR-NPTYGAWN---ADLGMIEQKS  NWVMGIALS ILAGV P  IVF :  614
AgCP7001   : ---WQMT RYIG AIMGC LLSSILSML----LR-NPTYGAWN---GDLGIVENKS  NWVMGIALA IIAGV P  VVF :  621
HsNTT7-3   : ---YYYM KYIS LMLLS LIASVVNMG----LS-PPGYNAW---IEDKASEEFLS  TWGLVVCVS VVFAI P  VVF :  637
HsNTT4     : ---YFYM KFVS LCMAV TTASIIQLG----VT-PPGYSAW---IKEEAAERYLY  NWAMALLIT IVVAT P  VVF :  636
HsXtrp2    : ---WRLT RVVS LLLT- FVAYIILLF----WK-PLRYKAWNPKYELFPSRQEKL  GWARAACVL SLLPV W  VAA :  586
RnROSIT    : ---WQVT RVVS MLLFG FLSYIVLLA----QS-SPSYKAWNPQYEHFPSREEKL  GWVQVTCVL SFLPS W  GIA :  574
HsBOAT1    : ---WQVT RVVS LLMLI FLFFFVVEV----SQ-ELTYSIWDPGYEEFPKSQKIS  NWVYVVVVI AGVPS T  GYA :  601
HsXT3L     : ---WKVM AGVS LLIVS FVFYLSDYI----LTGTLKYQAWDASQGQLVTKD---  AYALAVIGL VASST C  LAA :  574
DmCG10804  : ---WMFC KYLS CAMVT LLASFYQLL----TE-GSSYPAWI---GSKGATEGME  HWCIVVAFF ILSSI W  IVA :  614
AgCP7501   : ---WMLC KYIS IAMIT LVASFLELA----SE-GSSYPGWN---ALTGTTDQLE  HWCIVVAIL ILVSI W  GVA :  568
DmCG1698   : ---YRIC AVVT LLMLT LIYTLVLYE----PL--------------KYK-DYT- QSGVYVFGWC SAFGVGQ LFWA :  588
DmCG4476   : ---YRIC SIVT LVMLV LVYSLLTMR----PL--------------SYN-GQE-  LVYRVVGWC SGCIIGQ FYWA :  562
DmCG15279  : ---WRLC SIIT LIMTV LIYFYATYQ----PL--------------TYN-NIV-  NWSYSIGWL TAFGI Q  IWM :  576
AgCP10412  : ---WRLC RWIT LLMFV LIYNLVTLE----PL--------------MYK-QYV-  TVAYGIGWC FAFGL Q  IWA :  592
DmCG8850   : ---WRIC AIVA GLMFL LVYMLFSYE----PL--------------TYR-GVQ-  PVYYMAGWI WGLGV Q  FWA :  555
DmCG15088  : ---WRLC RFIT GLMTA LIYMLLDMS----AL--------------EYK-GVG-  TLAHVFGCF ATLGL Q  GWA :  522
AgCP10401  : ---WRIC GFIT TLLAA LLYHIATYE----TF--------------TFN-GYV-  DGMYAFGWC FAAGV Q  AWA :  594
AgCP10503  : ------------------------------------------------------------------------------ :    -
AgCP10610  : ------------------------------------------------------------------------------ :    -
DmCG3252   : ---WRVC SFFT VMMII FIYSMVTIE----PI--------------KYS-ELY-  EAANIAGWL FAIGAAQF LWG :  569
AeAAT      : ---WRLC GLIT LFMIA FIYSLVEYK---WP---------------TYS-GQQ-  GEALICGIF FIFGI Q LIWA :  557
MsKAAT1    : ---WRLC GIVT GMMII FIYALLSFE----NL--------------VFG-DFYV  VAGYVAGYM LFLGI L  IGI :  561
MsCAATCH1  : ---WRLC GVIT AIMTT FFYALLASN----NL--------------VFG-DNYV  TAGYVSGYL LFLGMTF  IGI :  561
BcProT     : --FWRVSLLVIT LMLGY LIDGTIQNIKKN----------------------YGD  TEFVVTYGWSIAAAF I AIII :  482
HinSnfT    : --FWRL-CIVLTSGVLAF LFSEGAKVFSEG----------------------YEG  NWFVNTFGWGMSISL V AFFL :  489
StTnaT     : --WWDVLVKYVT VLLGYNILSNFISEFREP----------------------YAG  TGALVLFGW VAIAMFGTSLFM :  486
CeSNF2     : AYFWTLI KFVT VCSLL MGLWIWEKKYPHKG-------------------DGKI  PVFDALGWF AVVPF V  GFA :  675
CeSNF7     : AYVWTIV RFVT ICCLV MFAWIWDKKYPFKG-------------------DSKK  PVFDAFGWF AALPFFV  VFA :  565
CeSNF4     : SPYYGFN MFIM IFGCV IFLTAN-RGYPFKG-------------------ESAV  VVFDIVGWL AVFPF M  VIA :  535
CeSNF5     : SWYFTAN MVIS SIALI VGLSFV-REYPYMG-------------------RHDI  AVFDIFGWF SFLPV I  IFM :  637
CeSNF9     : SPYFGFN MFVT VLGSI IVLASI-RQYPYQTD----------------ENKPNM  LAFDIVGWA CLLPMAL  IFA :  610
DmCG13794  : ----VYILRFIA IILIICLVVGILVCYM----------------------EHIYSSAIVLGMSII VILPI A  GYG :  542
DmCG13795  : ----VFILRFIA IILVICLLAGIALSIS----------------------EHSYSTTVVFVMSIV VVLPI A  GYG :  538
DmCG13793  : ----VFILRFVA PTLVICLVLEIKMCLE----------------------EHSYSSSVLHFMAII VGLPI A  AYG :  438
DmCG33296  : ----VFILRYTA IVFLICMILGIIECIR-------------------VHFF KEVFLVLGII VVLPF A  GYG :  466
DmCG8291   : -AFLALS NVLL IGLIT SVVDYKASLSNQFY----------------YWRGKSY TYWSRKMGSL QIGVL V  VTA : 1060
AgCP9955   : -AFVAFA NFLI IGLIF CILQYRLSNSAALF----------------NCN---Y  LWARQVGGF QVSFL L  VVM :  569
DmBLOT     : ---WMWL VICPAILTGI VWWCADDDQ---------------------YDLLAEYL RWAPILFVLAVIVIIAC QIFR :  695
AgCP2093   : ---WLFI GFIPVLLMAL CWWSITYYD--------------------NDLLIDY  KWLPVVFSLSVIVLLAC EISK :  490
CeSNF6     : -IFWKIL KFIC MVYLA LCFLWLDWN-------------------SIQYESYQ  YWSILTAWC ASFPL L  IVG :  601
CeSNF12    : GIYIKFSLCFIC VILTV LVASVLGYQR--------------------ISFAGRPI IDYEIVAWI MIGPL V  LVA :  696
AaLeuT     : ---YYYVMRYIT AFLAV LVVWAREYI---------------------------- KIMEETHWT WITRFYI GLFL :  495
HsNTT5     : ---FGWL PHLC VVLLI FVTMMVHLCMKPITYMSWDSS--------TSKEVLRP  PWALLLMIT FAIVI P  AYF :  678
DmCG13796  : ----IFFIRFAT IFLGLCLFQLTYLFTR-------------------------DRIHDVLIYVSQGLIL TAISYM :  587
HspSnfT    : ---WLYAVGVVI VFLIFTLVSGIAAALG-------------------------VSITAVELWGVTISQPR FAGLSV :  460
CeSNF1     : --HFFLY TYLL ILPAG AFLNALDYKFSSFS------------------APIHEWKYSEYVGIA AMVPL P  LYF :  549
CeSNF8     : WPIWKWR TTAPAFSIIFAVSAFFRPFPSPE--------------------NDLE YYFYDLIQFF LFFVI SPFGFA :  544

                          *      1300         *      1320         *      1340         *      1360
DmSERT     : IYKFFFASK-----------------------G--GCRQRLQESFQPEDNCGS-------------------------- :  612
AgCP12911  : VYKFLISS-------------------------G--NCNDRLRRTFKPEPLIPT-------------------------- :  553
MsSERT     : IYKLLITP-----------------------G--NCINRIKTIQRPE--VTS---------------------------- :  576
HsSERT     : AYRLIITP-----------------------G--TFKERIIKSITPET--PT---------------------------- :  618
CeSERT     : IIYLLSGT-----------------------G--TLYERFRWAITPQQRRNS-------------------------- :  645
DmDAT      : IFKLLSTPG---------------------------SLRQRFTILTTPWRDQQS-----------------------MA :  604
AgCP3649   : SYKLIVTPG---------------------------SLRQRFKILTTPWRDQQ-------------------------- :  562
TnDAT      : IYKLITAKG---------------------------SFVERLRLLTTPYADTEG-----------------------NG :  572
```



```
TnOctT     : ------IRGGAPVSRFS----WRHWLYV---------------------------------------------------------- :  670
DmCG5549   : RPCVNLHAERYQSQNYDGVTNTKILIEEEASPKQAANKKDRQLTLDVDVDVECLPLATVGKRPGAPKSMGKQGSSGNLQT :  720
AgCP10735  : ---------------------------------------------------------------------------------- :    -
HsGLYT2    : ------FLAQHRGERYKNMIDPLGT-----------------SSLGLKLPVKDLELGTQC-------------------- :  797
HsATB0+    : ------YLEQHRGERYKDMVDPK-------------------KEADHEIPTVSGSRKPE--------------------- :  642
HsPROT     : ------SLEENRTGMYVATLAGSQSPKPLMVHMRKYGGITSFENTAIEVDREIAEEEESMM------------------- :  636
HsGLYT1    : ------ALLEHRTGRYAPTIAPSPEDGFEVQPLHPDK-----AQIPIVGSNGSSRLQDSRI------------------- :  706
DmCG7075   : -------ADPQDRAEYEAFRRQR-------------------RMPGFMSDTDMESSK----------------------- :  590
AgEbiP6595 : ---------------------------------------------------------------------------------- :    -
MsPROT     : -------SEPEKRREWELLKKQK---------------------ADIFPLNDLDKY------------------------ :  556
DmIne      : -------CGEHHCEHDYPEQEQFMLAQEMATVYKPTNPHLLNLGQKCGYNAMQASPSHAEAGGPCGQ------------- :  658
AgCP4566   : -------CGEHHCDHDFPDEE---------------------------------------------------------- :  654
AmINE      : -------CGQEYCNESMHNFEEDMIKEQQD----PNSPIQLSISPPLSKSQT--------------------------- :  888
MsIne      : ------NGCDICCSESDPPDDKTVIN------------------------------------------------------ :  727
TnGAT      : ------LRTKMQAEEQAKHGCKA---------------------------------------------------------- :  608
MsGAT      : ------LRTKM----------------------------------------------------------------------- :  597
AgCP8499   : ---------------------------------------------------------------------------------- :    -
DmCG1732   : ------LREKMLREAYAKEIEFNSL------------------------------------------------------- :  636
CeSNF11    : ------AMENAENLELVEDFQNAI--------------------------------------------------------- :  632
HsGAT1     : ------PENGPEQPQAGSSTSKEAYI------------------------------------------------------- :  599
HsGAT3     : VTVNDCDAKLKSDGTIAAITEKETHF------------------------------------------------------- :  632
HsGAT2     : -------APATPRTSLLRLTELESHC------------------------------------------------------- :  602
HsBGT1     : -------FGPSPTREGLIAGEKETHL------------------------------------------------------- :  614
HsTAUT     : --------PYNSRTVMNGALVKPTHIIVETMM------------------------------------------------- :  620
HsCT1      : ------LTTLTPVSESSKVVVVESVM-------------------------------------------------------- :  635
CeSNF3     : ----PAKKNLVNNAEELKFIESSSQ--------------------------------------------------------- :  600
DmCG5226   : ---MIDNDDDNMSPDMPPQDSLAATRFTIGDFEN---------------------------------------------- :  675
AgCP7001   : ---MID-EDDVVSPDLPPPQSLAATRFTIGDFDV---------------------------------------------- :  681
HsNTT7-3   : ---DTSLIHGKIPSEMPSPNFGKNIYRKQSGSPTLDTA--PNGRYGIGYLMADIMPDMPESDL----------------- :  730
HsNTT4     : ---ETRFILSKVPSEAPSPMPTHRSYLGPGSTSPLETSGNPNGRYGSGYLLAS----TPESEL----------------- :  727
HsXtrp2    : ---MRPDTDTRPDTDMRPDTDMR----------------------------------------------------------- :  628
RnROSIT    : ----------THLESALKPQESRGC--------------------------------------------------------- :  603
HsBOAT1    : ------LVSTLSTASMNGDLKY------------------------------------------------------------ :  634
HsXT3L     : ----------RGDADPVA--------------------------------------------------------------- :  592
DmCG10804  : ---PTELERSIFCFNMDGTEGMCCPKYGLPEKSLEEEE------------------------------------------- :  682
AgCP7501   : ---PTDIEISLFCIRADGSEGLCCPIYGPREQPLDEEE------------------------------------------- :  636
DmCG1698   : -------SDPQTLKRYQLFVQ---EGNANALFRRSSIWHKIYDNIFG---------------------------------- :  654
DmCG4476   : -------SDPKKLMDYQMFLRNKDEENSKNVNRR-CVCYTAGDRIFG---------------------------------- :  629
DmCG15279  : -------SDPLLREQYHKEIE-----NELTPKRGQGIWAAIKQNIFG---------------------------------- :  639
AgCP10412  : -------IDPATHYEYKKFID-----ED---------------------------------------------------- :  636
DmCG8850   : -------LQTQKFEAYIIHRK-------READFKSPRGGYLFDNIFG---------------------------------- :  616
DmCG15088  : -------ANVQ-LDATI---------------------------------------------------------------- :  561
AgCP10401  : -------EDPELNAKYHESVY----KHEQSLPRDRSLVRRMFDNVFS---------------------------------- :  658
AgCP10503  : ---------------------------------------------------------------------------------- :    -
AgCP10610  : ---------------------------------------------------------------------------------- :    -
DmCG3252   : -------ANPEIRREWVIFKN--QKAAQRATQKDTSKLGFFWRKVANFCGSNK---------------------------- :  641
AeAAT      : -------CTEHTRKAWLKYKE--EAKSRRDDIIYAKNHSSIVRKLCVLFGMYDDFIQLDSTRL----------------- :  640
MsKAAT1    : -------RSPRLRREWMQFRI--EAKALR-QKMNTSRVKHLWYSITGAYRRNINK------------------------ :  634
MsCAATCH1  : -------RSPRERREWMQFKA--EAKALR-QKMNTSRVKHLWYSITGAYRRNIN------------------------- :  633
BcProT     : ---------------------------------------------------------------------------------- :    -
HinSnfT    : ---------------------------------------------------------------------------------- :    -
StTnaT     : ---------------------------------------------------------------------------------- :    -
CeSNF2     : ------KAIGDKLPEQEPGDDDLNMIDGLETKETWSFLYKFLSIRVFNKAKIVIETGKYLK-------------------- :  768
CeSNF7     : ------QKIGDKLPELEPGEDDL--EDGLEGSDD-DLFYEY---------------------------------------- :  635
CeSNF4     : ------RSVGAQLPSFEPGEEEFKSQEPGYGPV------------------------------------------------ :  600
CeSNF5     : ------LALQDQLPDKEPWDEE-----NVDLTDSESESRNAASGDVPIDDVATIDTSSTYHQVY---------------- :  728
CeSNF9     : ------RAKQKILPDREPWDDAPTGHQFVASPDFQSSSQSTRS-ERPSERSNYLNSAASVHSKTQ---------------- :  706
DmCG13794  : -------EHRQQYEEAVSNTDMTQQLTAGPN------------------------------------------------- :  595
DmCG13795  : -------EHRQQYEEAVDSTEMTHQLFEVTEEVN------------------------------------------------ :  594
DmCG13793  : -------EYRQQYGELVQKSEMTHQLFDRTEEVN------------------------------------------------ :  490
DmCG33296  : -------KHRQHYGDRVGNIEMTDHLFNGIEEVN------------------------------------------------ :  522
DmCG8291   : HDAQNDAPPKYTPPPSYTTATGARLAKLLRQSIRRSVRRVLGDSSRTRPVLSLDAESASPAAPPDYLTILTNPAGSSFNA : 1197
AgCP9955   : VRDADDAPPKYTPPPSYTTATGARIAKLLRNSIRRSVRRLMGETSGSAP--SVRQRTALPQTAPDSL------------- :  688
DmBLOT     : RRGTRDYTHSIKKGQYSSATKYGVQNGGQTQTPMHQQHWKSSTPGNSSPNYSGSMFGDSAIEEDISVDEFPGITQQQYVP :  824
AgCP2093   : RRGTKDYTNSIKKGQYTHSHKYGASS-----------NRNLSTTGSNSPNYSGSVFGDSAIEEDMSVEKYP--------- :  599
CeSNF6     : AEKFPLQIPEARRLLLPPEVEIASSRGVLQEEMPMSYDYNTSSAADVRSNRSTGHGATDVRSVAATNNTIPKFERETAI- :  713
CeSNF12    : LGASLDPYGWKAGRLRDRQQQIEETASNYSEEDSATTNSFMASTVKHNDDMELTLFGSPPAILGDDEKIMTTRFSESMPV :  856
AaLeuT     : ---------------------------------------------------------------------------------- :    -
HsNTT5     : -------SKEVQKEEILQVDETKYPSTCNVTS------------------------------------------------- :  736
DmCG13796  : --------NRRFYEEIMGISEMLVIDANANTT------------------------------------------------- :  641
HspSnfT    : ---------------------------------------------------------------------------------- :    -
CeSNF1     : ------DEKAQPIPRYTSNAPGYLLLPQAPLAEPEIYA------------------------------------------- :  617
CeSNF8     : ---ESEDLEEFWKLVYRPFFETPSHETVMMPIIPKSKEIVKPRQASKKRSKKRSKKDEFLTFETYYKEKGK--------- :  669

                        *      1460         *      1480         *      1500         *      1520
DmCG5549   : NAKNGTHKSPTTEISPSKQPLIPSEKDNKSANIKVLAAGTQSSGRECTTMSTFAGGGEKKSTKVNLATGATSAAAPLASS :  800
DmCG8291   : DPSPASSSSPESIEIDQRPVGYSQRSQSLGRKLHRSGASTLERRPYTAEDVVTILRSSVRHRQSQGGGNLVTASTLPRPP : 1277
DmBLOT     : FQASDTKQSRYSQRIRQTAQPQTQTQMRPPRESVEKHREVVYIRRLSDGGTGTGHATRIEISPSNESITYGNGKSNNNYS :  904
CeSNF12    : NYKCRNVEVPRIPNKLPQNMERMARKTRKKRSSPSASDPPVPTSPLPPPPKLQHCRSEPPMMNSKESHSPEIITPGDDSP :  936

                        *      1540         *      1560         *      1580         *      1600
DmCG5549   : NIATAATVAAATKSPPAAGVAPTTLVGIAATGTTVAAAANGKATLKTTISSGAVAATAAGGNKAASVAAATATVAAGKTP :  880
DmCG8291   : TAAMSTHLEDASFRSIENLVLNAEPPDRTPGVELELELEAGQAEECARNNTSVI-------------------------- : 1331
DmBLOT     : SAMSRNPLGRSTGCLAAPASRRRPACMSSGRPARWSMLTHTSCHLQRRPVERRITFAGGSSTSIRKSIPRSCEIRGGPGV :  984
CeSNF12    : SISNSSDDSSDDCFRRATVIRRKTSDDDAFTHFSTATAESISITPLDFPRQRSLSSVAIYDQEQKNGRSKVLSQLKRPKP : 1016

                        *      1620         *      1640         *      1660         *      1680
DmCG5549   : AAAMALNVAASVKVSQTDKTAPLKTTISASVATAPGGTKASTAPSVATAAVVSVSSGATVPAAASKAAAPLKATAPPPSL :  960
DmBLOT     : GNEAFTFKKTCCIQNKRICICNPYCIYGTPNYKLRITNHVHTESVSSSCEL---------------------------- : 1035
```

```
CeSNF12    : IDMPPK---------------------------------------------------------------------- : 1022

                       *      1700         *      1720         *      1740         *      1760
DmCG5549   : AAVNATQAAKSTPVAAASALVASAPPAKSTSASVAKTNTNAGSKPTATLSTCKSDAPAAAASVGVTNGKVTPSGGATSTA : 1040

                       *      1780         *      1800         *      1820         *      1840
DmCG5549   : QPAVNQKATAAAAAQKSQGATATATPTKTALNQKTIVPSAGSGKGDVPSSRSTDTPPPAKGAASSNISAKSTKAGKTQAN : 1120

                       *      1860         *      1880         *      1900         *      1920
DmCG5549   : ETSPTKSAVKPIKAGSTSPTKAVVSKPGASAAKTAAASAATATATAAGSSSSNKKAATSTAATLGSQAKTKGNASGAGTK : 1200

                                      DmCG5549   : K : 1201
```

Table S1. *Sequence names, abbreviations, and accession numbers used for phylogenetic analysis*

Species	Sequence name	Abbreviation	GenBank accession no.
INSECTS			
Drosophila melanogaster	Serotonin transporter	Dm SERT	NP_523846
	Dopamine transporter	Dm DAT	NP_523763
	Inebrieted	Dm Ine	NP_523763
	Bloated tubules	Dm Blot	CAB53640
	CG1698	Dm CG1698	NP_610517
	CG1732	Dm CG1732	NP_651930
	CG3252	Dm CG3252	NP_572219
	CG4476	Dm CG4476	NP_648293
	CG5226	Dm CG5226	NP_611340
	CG5549	Dm CG5549	NP_611836
	CG7075	Dm CG7075	NP_609135
	CG8291	Dm CG8291	NP_611064
	CG8850	Dm CG8850	NP_610755
	CG10804	Dm CG10804	NP_726868
	CG13793	Dm CG13793	NP_609136
	CG13794 (orf)	Dm CG13794	AAY55020
	CG13795 (orf)	Dm CG13795	AAQ22494
	CG13796 (orf)	Dm CG13796	AAN71401
	CG33296	Dm CG33296	NP_995644
	CG15088	Dm CG15088	NP_611364
	CG15279	Dm CG15279	NP_609726
Anopheles gambiae	EbiP6595	Ag Ebi6595	XP_316980.2
	CP2093	Ag CP2093	XP_315020.2
	CP3649	Ag CP3649	XP_308462.2
	CP4566	Ag CP4566	XP_319694.2
	CP7001	Ag CP7001	XP_309367.2
	CP7501	Ag CP7501	EAA06912.3
	CP8499	Ag CP8499	XP_318854.2
	CP9955	Ag CP9955	XP_320247.2
	CP10401	Ag CP10401	XP_309846.2
	CP10412	Ag CP10412	XP_309840.2
	CP10503	Ag CP10503	XP_309844.2
	CP10735	Ag CP10735	XP_311463.2
	CP10610	Ag CP10610	XM_309845.2
	CP12911	Ag CP12911	XP_310113.2
Manduca sexta	Potassium AA transporter 1	Ms KAAT1	AAC24190.1
	CAATCH1	Ms CAATCH1	AAF18560.1
	GABA transporter	Ms GAT	AAA92342.1
	Proline transporter	Ms PROT	AAP94997.1
	Serotonin transporter	Ms SERT	AAN59781.1
	Inebriated homolog	Ms Ine	AAF21642.1
Trichoplusia ni	GABA transporter	Tn GAT	AAF70819.1
	Octopamine transporter	Tn OctT	AAL09578.1
	Dopamine transporter	Tn DAT	AAN52844.2
Apis mellifera	Inebriated homolog	Am INE	XP_395197.2
Aedes aegypti	Amino acid transporter	Ae AAT	AF543193.1
MAMMALS			
Homo sapiens	GABA transporter 1 (SLC6A1)	Hs GAT1	NP_003033.2
	Norepi transporter (SLC6A2)	Hs NET	NP_001034.1
	Dopamine transporter (SLC6A3)	Hs DAT	NP_001035.1
	Serotonin transporter (SLC6A4)	Hs SERT	NP_001036.1
	Glycine transporter 2 (SLC6A5)	Hs GLYT2	NP_004202.1
	Taurine transporter (SLC6A6)	Hs TAUT	NP_003034.2
	Proline transporter (SLC6A7)	Hs PROT	NP_055043.1
	Creatine transporter (SLC6A8)	Hs CT1	NP_005620.1
	Glycine transporter 1 (SLC6A9)	Hs GLYT1	NP_964012.2
	GABA transporter 3 (SLC6A11)	Hs GAT3	NP_055044.1
	Betaine transporter (SLC6A12)	Hs BGT1	NP_003035.2
	GABA transporter 2 (SLC6A13)	Hs GAT2	NP_057699.2
	AA transporter B0+ (SLC6A14)	Hs ATB0+	NP_009162.1
	NT transporter 7-3 (SLC6A15)	Hs NTT7-3	NP_877499.1
	NT transporter 5 (SLC6A16)	Hs NTT5	NP_054756.2
	Transporter Xtrp2 (SLC6A17)	Hs Xtrp2	NP_872438.1
	B0 AA transporter 1 (SLC6A19)	Hs BOAT1	NP_001003841.1
	Transporter XT3L (SCL6A20)	Hs XT3L	NP_064593.1
Rattus norvegicus	ROSIT	Rn ROSIT	U12973
WORM			
Caenorhabditis elegans	Dopamine transporter	Ce DAT1	NP_499043
	Serotonin transporter	Ce SERT	AAK84832
	SNF family transporter 1	Ce SNF1	NP_491491
	SNF family transporter 2	Ce SNF2	NP_492396
	SNF family transporter 3	Ce SNF3	NP_493910
	SNF family transporter 4	Ce SNF4	NP_496730
	SNF family transporter 5	Ce SNF5	NP_496735
	SNF family transporter 6	Ce SNF6	NP_497416
	SNF family transporter 7	Ce SNF7	NP_499702
	SNF family transporter 8	Ce SNF8	NP_502271
	SNF family transporter 9	Ce SNF9	NP_503077
	SNF family transporter 11	Ce SNF11	NP_505873
	SNF family transporter 12	Ce SNF12	NP_509527
PROKARYOTES			
Aquifex aeolicus	Leucine transporter	Aa LeuT	NP_214423.1
S. thermophilum	NaCl-dep. transporter	St TnaT	YP_074271.1
Bacillus cereus	Na/Proline symporter	Bc ProT	NP_833789.1
H. influenzae	SNF family transporter	Hin SnfT	ZP_00154505.1
Halobacterium sp.	SNF family transporter	Hsp SnfT	AAG19816.1