Amino acid composition analysis of human secondary transport proteins and implications for reliable membrane topology prediction

Massoud Saidijam\textsuperscript{a}, Sonia Azizpour\textsuperscript{a} and Simon G. Patching\textsuperscript{b*}

\textsuperscript{a}Department of Molecular Medicine and Genetics, Research Centre for Molecular Medicine, School of Medicine, Hamadan University of Medical Sciences, Hamadan, Iran; \textsuperscript{b}School of BioMedical Sciences and the Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds LS2 9JT, UK

Communicated by Ramaswamy H. Sarma

(Received 18 October 2015; accepted 15 March 2016)

Secondary transporters in humans are a large group of proteins that transport a wide range of ions, metals, organic and inorganic solutes involved in energy transduction, control of membrane potential and osmotic balance, metabolic processes and in the absorption or efflux of drugs and xenobiotics. They are also emerging as important targets for development of new drugs and as target sites for drug delivery to specific organs or tissues. We have performed amino acid composition (AAC) and phylogenetic analyses and membrane topology predictions for 336 human secondary transport proteins and used the results to confirm protein classification and to look for trends and correlations with structural domains and specific substrates and/or function. Some proteins showed statistically high contents of individual amino acids or of groups of amino acids with similar physicochemical properties. One recurring trend was a correlation between high contents of charged and/or polar residues with misleading results in predictions of membrane topology, which was especially prevalent in Mitochondrial Carrier family proteins. We demonstrate how charged or polar residues located in the middle of transmembrane helices can interfere with their identification by membrane topology tools resulting in missed helices in the prediction. Comparison of AAC in the human proteins with that in 235 secondary transport proteins from \textit{Escherichia coli} revealed similar overall trends along with differences in average contents for some individual amino acids and groups of similar amino acids that are presumed to result from a greater number of functions and complexity in the higher organism.

\textbf{Keywords:} amino acid composition; drug targets; human transport proteins; membrane topology; structure

1. Introduction

Membrane proteins are coded by up to 30\% of the open reading frames in known genomes (Fagerberg, Jonasson, von Heijne, Uhlén, & Berglund, 2010; Liu & Rost, 2001; Wallin & von Heijne, 1998) and typical biological membranes consist of up to 50\% mass fraction of proteins (Johansson & Lindahl, 2009). They have important roles in many biological processes (e.g. transport of ions and molecules, control of transmembrane potential, generation and transduction of energy, signal recognition and transduction, catalysis of chemical reactions) and mutations in membrane proteins have been linked with a number of human diseases (Klepper & Voit, 2002; Kurze et al., 2010; Partridge, Therien, & Deber, 2002; Patching, 2015, 2016; Quadri et al., 2012; Ragona et al., 2012; Rosenbaum, Rasmussen, & Kobilka, 2009; Sanders & Myers, 2004; Shukla, Vaitiekunas, & Cotter, 2012; Striano et al., 2012; Suls et al., 2009; von Heijne, 2007; Watanabe et al., 2008; Weber et al., 2008). The molecular targets for around 50–60\% of current validated medicines are membrane proteins and they remain the principal target for new drug discovery (Bahar, Lezon, Bakan, & Shrivastava, 2010; Bakheet & Doig, 2009; Drews, 2000; Hopkins & Groom, 2002; Lundstrom, 2006; Overington, Al-Lazikani, & Hopkins, 2006; Rask-Andersen, Almén, & Schiöth, 2011) with many more likely to be identified following efforts to map the tissue-specific human proteome (Lindskog, 2015; Uhlén et al., 2015). Owing to the difficulties in applying the main biophysical techniques for high-resolution protein structure determination, X-ray crystallography and NMR spectroscopy, the number of structures of membrane proteins is still relatively few. They contribute only around 2.5\% of entries in the Protein Data Bank (PDB) (http://www.rcsb.org/pdb/home/home.do), thus limiting the amount of information available for traditional structure-based drug design. There is an almost infinite amount of information yet to be obtained about the structures, ligand interactions, molecular mechanisms, dynamics and multidimensional relationships of membrane proteins along with scope for development and application of a wide range of chemical, biochemical, biophysical,
molecular imaging and computational techniques to achieve this. Computational techniques include bioinformatics approaches, which can be used to explore, analyse and consolidate the huge amount of sequence information available for membrane proteins in the post genomic era. For a number of reasons, it is not feasible to apply experimental techniques to all membrane proteins, so bioinformatics methods are used to help in the identification of protein function, classification of proteins, investigation of evolutionary relationships, homology recognition and in the identification of potential drug targets. For example, amino acid sequence information can be used to identify homology, structural and functional domains and ligand and drug binding sites in membrane proteins.

One of the basic parameters provided by the amino acid sequence that can be analysed relatively easily and quickly is that of amino acid composition (AAC), which quantifies the frequencies of the twenty different amino acid types in a protein. AAC was first correlated with protein structural and biological characteristics in 1982 (Nishikawa & Ooi, 1982). The large number of subsequent studies include calculation of AAC and the likelihoods of neighbouring residues for every Cα atom in 3718 mainly non-membrane proteins with crystal structures. This work demonstrated the absence of any preferential interactions between amino acids and that protein folding is simply dictated by frequencies of occurrences of individual amino acids (Mittal, Jayaram, Shenoy, & Bawa, 2010). A further outcome was a suggestion of ‘Chargaff’s rules for protein folding’, which was followed up by a number of interesting commentaries (Galzitskaya, Lobanov, & Finkelstein, 2011; Mezei, 2011; Mishra, 2011) and a more rigorous study for the probability of amino acid neighbours in the 3718 proteins with suggestions of evolved rules for protein folding (Mittal & Jayaram, 2011). An element of AAC analysis has been used to reveal the physical properties of proteins, including membrane proteins and their interactions with lipid bilayers. For example, AAC analysis was used to help demonstrate how transmembrane domains can have organelle-specific properties. Analysis of transmembrane domains in single-span (bitopic) proteins from fungi and vertebrates showed a conserved greater hydrophobic length in plasma membrane transmembrane domains compared with those that span endoplasmic reticulum and Golgi membranes. An asymmetric AAC of plasma membrane and Golgi transmembrane domains, differing between the two leaflets of the lipid bilayer, was also apparent (Sharpe, Stevens, & Munro, 2010). AAC has also been used as the basis for a large number of bioinformatics algorithms of varying complexity, including many specifically designed for and/or applied to membrane proteins (Cai & Chou, 2006; Chen & Li, 2013; Chou, 2009; Diao et al., 2008; Du, Gu, & Jiao, 2014; Wang et al., 2012). Analysis of the contents of individual types of amino acids and of groups of amino acids with similar physicochemical properties helps to distinguish membrane proteins from non-membrane proteins with the former usually having a higher content of hydrophobic residues (alanine, isoleucine, leucine, phenylalanine, valine) (Garrow, Agnew, & Westhead, 2005; Liu, Zhu, Wang, & Li, 2003; Pascal, Médigue, & Danchin, 2006). AAC analysis can be used alone or in combination with sequence alignments and other information to help identify and classify membrane protein types and to reveal secondary structure domains and sites of specific function, including those for ligand and drug binding (Bhasin & Raghava, 2004; Chen, Ou, Lee, & Gromiha, 2011; Gromiha & Yabuki, 2008; Lin & Ding, 2011; Mohabatkar, Mohammad Beigi, & Esmaeili, 2011; Ou, Chen, & Gromiha, 2010; Schaadt & Helms, 2012; Zia-Ur-Rehman & Khan, 2012). Indeed, a recent bioinformatics model was introduced for predicting membrane transport proteins and their substrate specificities using protein sequence information, including AAC (Mishra, Chang, & Zhao, 2014). This complements our AAC analysis of 235 secondary transport proteins from Escherichia coli, which revealed trends in content for individual amino acid types and for groups of amino acids with similar physicochemical properties that in some cases could be directly related to function and ligand specificity (Saidijam & Patching, 2015).

Secondary transporters have some common folds and shared transport mechanisms (Forrest, Krämer, & Ziegler, 2011; Shi, 2013) and are widespread in all organisms. In humans over 300 have been identified so far that transport a wide range of ions, metals and organic and inorganic solutes involved in energy transduction, control of membrane potential and osmotic balance, metabolic processes and in the absorption or efflux of drugs and xenobiotics. They are also emerging as important potential targets for the development of new drugs (Lin, Yee, Kim, & Giacomini, 2015) as well as being target sites for drug delivery to specific organs or tissues (Tashima, 2015), especially for enabling drugs to cross the blood–brain barrier (Geier et al., 2013; Mikitsh & Chacko, 2014; Patching, 2015, 2016; Patel, Goyal, Bhadada, Bhatt, & Amin, 2009; Sanchez-Covarrubias, Slosky, Thompson, Davis, & Ronaldson, 2014; Stenhjem, Hartz, Bauer, & Anderson, 2009). The largest groups of human secondary transport proteins are in the Mitochondrial Carrier family (Kuan & Saier, 1993; Haferkamp & Schmitz-Esser, 2012; Monné & Palmieri, 2014; Nury, Blesneac, Ravaud, & Pebay-Peyroula, 2010; Palmieri, 1994) and in the Major Facilitator Superfamily (Pao, Paulsen, & Saier, 1998; Reddy, Shlykov, Castillo, Sun, & Saier, 2012; Saier et al., 1999; Yan, 2013, 2015). Application of bioinformatics approaches to these proteins is important to contribute information for
understanding evolutionary relationships, defining functions and specific substrates, characterising structural domains, elucidating molecular mechanisms and identifying potential target sites for drugs. Hence, we have performed an AAC analysis of 336 human secondary transport proteins and used the results to look for correlations with classification, structural domains and specific substrates and/or function. The AAC results were compared with a phylogenetic analysis and predictions of membrane topology in the proteins, which are also based on the protein sequence. An important outcome is a demonstration how AAC can have implications for reliable membrane protein topology predictions. AAC in the human proteins was also compared with that in the secondary transport proteins from E. coli.

2. Methods

2.1. AAC analysis

A total of 336 secondary transport proteins from Homo sapiens were chosen for analysis from the Transporter Protein Database (http://www.membranetransport.org/) (as of December 2015), which is compiled and maintained by the laboratory of Ian Paulsen at Macquarie University (Ren, Chen, & Paulsen, 2007). At this time the Transporter Protein Database had 447 entries for human secondary transporters, which were manually assessed for sequence selection. Those that contained no entry information at all, represented truncated forms or isoforms of the same protein or that apparently represented discontinued records based on links to and/or searches in other databases were not included. A full list and other details of the chosen 336 proteins (protein number, protein name, source of sequence, UniProt accession number, transporter Family, known or putative substrate(s) and/or function, number of amino acid residues, molecular weight, putative number of transmembrane-spanning α-helices) are given in Supplementary Table S1. The sequence of each protein was taken from the Transporter Protein Database (http://www.membranetransport.org/) and checked for an exact match with a sequence in the Universal Protein Resource (UniProt, http://www.uniprot.org/), which was satisfied in the large majority of cases (308 out of 336). In other cases, the sequence used from the Transporter Protein Database was a truncated form or isoform of the same protein with the given UniProt entry. Each sequence was analysed using the ExPASy ProtParam tool (http://web.expasy.org/protparam/) (Gasteiger et al., 2005) to obtain a value for the content of each amino acid in the protein as a percentage of total amino acids. These values were transferred into a spreadsheet and used to construct the scatter plots in Figures 1 and 2, which show the percentage contents for individual amino acids and for combinations of amino acids with similar physicochemical properties, respectively, for all 336 proteins. A complete set of values for the percentage contents is given in Supplementary Table S2. The identification of percentage contents for individual amino acids and for combinations of amino acids that showed enhanced values was performed using a statistical test for upper outliers and the results are shown in Supplementary Table S3. The proteins with statistically high amino acid contents are labelled in Figures 1 and 2 and highlighted on the phylogenetic tree in Figure 3.

2.2. Membrane topology prediction

The number of putative transmembrane-spanning α-helices in each of the 336 proteins (Supplementary Table S1) was obtained by analysis of their sequences using the membrane topology prediction tool TOPCONS (http://topcons.cbr.su.se/) (Bernsel, Viklund, Hennerdal, & Elofsson, 2009), which has recently been updated to efficiently separate any N-terminal signal peptides from transmembrane regions and it can predict re-entrant loops (Tsirigos, Peters, Shu, Käll, & Elofsson, 2015). The number of predicted transmembrane helices and the pictures used in Figures 4 and 8 come from the consensus TOPCONS result. The membrane topology prediction tool TMHMM (http://www.cbs.dtu.dk/services/TMHMM/) (Krogh, Larsson, von Heijne, & Sonnhammer, 2001) was also used to produce pictures in Figure 8.

2.3. Phylogenetic analysis

The sequences for the 336 secondary transport proteins from H. sapiens listed in Supplementary Table S1 were aligned using the online multiple sequence alignment tool Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/) (Sievers et al., 2011). The resultant neighbour-joining phylogenetic tree was exported in Newick format and re-drawn using the online tool iTOL (Interactive Tree of Life, http://itol.embl.de/index.shtml) (Letunic & Bork, 2007, 2011).

3. Results and discussion

3.1. AAC in human secondary transport proteins

Over 300 human secondary transport proteins have been identified with sequence information at the protein level and they are relatively well characterised, some with experimental evidence of structure, structural domains, specific substrates and function. They are therefore an ideal choice of protein for performing a basic AAC analysis to look for trends that correlate with classification, structural features and substrate specificity. The content of each of the 20 different amino acid types expressed as
A percentage of total amino acids was calculated using the sequences of 336 human secondary transport proteins as described in Methods. The results are illustrated in the form of scatter plots segregated by transporter family for each individual amino acid (Figure 1) and for groups of amino acids with similar physicochemical properties (Figure 2). A complete set of values for the percentage contents is given in Supplementary Table S2 and average contents of amino acid types across all 336 proteins are indicated on the plots in Figures 1 and 2. Our previous analysis of AAC in 235 secondary transport proteins from *E. coli* identified some proteins with a statistically high content of individual amino acids or of groups of amino acids with similar physicochemical properties, which in some cases could be directly related to specific substrates or function (Saidijam & Patching, 2015).
therefore performed a statistical test for outliers in the human proteins (Supplementary Table S3) and proteins with statistically high contents are labelled on the plots of the results (Figures 1 and 2). Prediction of membrane protein topology is also based on protein sequence and contributes important information for classification of membrane proteins and for identification of structural domains (Tusnády & Simon, 2010). Hence, we performed a membrane topology prediction for all 336 human transporters as described in Methods to obtain values for the putative number of full transmembrane helices in each protein (Supplementary Table S1). This allowed identification of any correlations between the number of transmembrane helices and AAC in the proteins. We also performed a nearest-neighbour phylogenetic analysis of the 336 human transporters as described.
in Methods. The resultant phylogenetic tree (Figure 3) helps to confirm the family classifications and inter-relationships of the transporters and correlations with the AAC analysis and membrane topology predictions, which are also shown on the tree. The outcomes of the AAC analysis are discussed in the following sections under groupings of similar amino acids. Note that the number used after the word ‘protein’ in this article identifies the position of the protein in the lists given in Supplementary Tables S1 and S2.

3.1.1. Hydrophobic residues

Amino acids with a hydrophobic (uncharged, non-polar) side chain are the most abundant type of residue found in the proteins, which are important for interaction with
the hydrophobic core of the membrane lipid bilayer. The average combined content of hydrophobic residues (alanine, isoleucine, leucine, phenylalanine, valine) in the 336 proteins is 40.5% (Figure 2 and Supplementary Table S2) and this increases to 48.6% when glycine residues are added. A relatively high content of hydrophobic residues is a major characteristic of membrane proteins that distinguishes them from non-membrane proteins and this trend is consistent with almost all proteins in this analysis. A few proteins do have a noticeably low content of combined hydrophobic residues, however. For example, the Calcium-Cation Antiporter family protein 42 (NP_004718, SLC24A1) has a combined hydrophobic residue content of 30.7%. Protein 42 is an ion exchanger in the visual transduction cascade, transporting one calcium ion and one potassium ion in exchange for four sodium ions, which controls the calcium concentration of outer segments during light and darkness (McKiernan & Friedlander, 1999). The protein is comprised of 1099 residues that form eleven putative transmembrane helices and the low hydrophobic residue content is explained by the presence of two large extramembrane domains.
In terms of individual types of hydrophobic amino acids, alanine residues have an average content of 8.2% and five proteins have a statistically high content. Glycine residues have an average content of 8.0% and two proteins have a statistically high content (Figure 1 and Supplementary Table S3). Of these, the Major Facilitator Superfamily protein 239 (NP_699188, SLC16A11) has a statistically high content of one or more specific residues (Figures 1 and 2, Supplementary Table S3) are highlighted with dots coloured to represent amino acids that are hydrophobic (alanine, isoleucine, leucine, phenylalanine, valine) or glycine (green), positively charged (arginine, lysine) (red), negatively charged (aspartic acid, glutamic acid) (blue), polar with hydroxyl or amido groups (asparagine, glutamine, serine, threonine, tyrosine) (pink), cysteine (yellow) and other residues (orange). The average numbers of predicted transmembrane spanning α-helices in each transporter family (Supplementary Tables S1 and S4) are also given to the nearest whole number (black numbers).

For example, crystal structures of the bovine mitochondrial ADP:ATP carrier (PDB 1OKC and 2C3E) contain six transmembrane helices (Nury et al., 2005; Pebay-Peyroula et al., 2003) and the sequence of this protein gives variable results when analysed by membrane topology prediction tools. This observation turns out to be a recurring trend that will be discussed further below.

In terms of individual types of hydrophobic amino acids, alanine residues have an average content of 8.2% and five proteins have a statistically high content. Glycine residues have an average content of 8.0% and two proteins have a statistically high content (Figure 1 and Supplementary Table S3). Of these, the Major Facilitator Superfamily protein 239 (NP_699188, SLC16A11) has a statistically high content of one or more specific residues (Figures 1 and 2, Supplementary Table S3) are highlighted with dots coloured to represent amino acids that are hydrophobic (alanine, isoleucine, leucine, phenylalanine, valine) or glycine (green), positively charged (arginine, lysine) (red), negatively charged (aspartic acid, glutamic acid) (blue), polar with hydroxyl or amido groups (asparagine, glutamine, serine, threonine, tyrosine) (pink), cysteine (yellow) and other residues (orange). The average numbers of predicted transmembrane spanning α-helices in each transporter family (Supplementary Tables S1 and S4) are also given to the nearest whole number (black numbers).

For example, crystal structures of the bovine mitochondrial ADP:ATP carrier (PDB 1OKC and 2C3E) contain six transmembrane helices (Nury et al., 2005; Pebay-Peyroula et al., 2003) and the sequence of this protein gives variable results when analysed by membrane topology prediction tools. This observation turns out to be a recurring trend that will be discussed further below.

In terms of individual types of hydrophobic amino acids, alanine residues have an average content of 8.2% and five proteins have a statistically high content. Glycine residues have an average content of 8.0% and two proteins have a statistically high content (Figure 1 and Supplementary Table S3). Of these, the Major Facilitator Superfamily protein 239 (NP_699188, SLC16A11) has a statistically high content of one or more specific residues (Figures 1 and 2, Supplementary Table S3) are highlighted with dots coloured to represent amino acids that are hydrophobic (alanine, isoleucine, leucine, phenylalanine, valine) or glycine (green), positively charged (arginine, lysine) (red), negatively charged (aspartic acid, glutamic acid) (blue), polar with hydroxyl or amido groups (asparagine, glutamine, serine, threonine, tyrosine) (pink), cysteine (yellow) and other residues (orange). The average numbers of predicted transmembrane spanning α-helices in each transporter family (Supplementary Tables S1 and S4) are also given to the nearest whole number (black numbers).
high content of both alanine and glycine residues and this appears to be balanced by a low content (1.5%) of isoleucines. Isoleucine residues have an average content of 6.2% and none of the proteins have a statistically high content (Figure 2 and Supplementary Table S3). With an average content of 12.6%, leucine residues are the most abundant of all residues by a clear margin and seven proteins have a statistically high content of leucines, the majority with a value above 20% (Figure 1 and Supplementary Table S3). Of these, the putative 12-helix Major Facilitator Superfamily protein 222 (NP_612440, MFSD3) (Figure 4(B)) has a leucine content of 25.2%, which is the highest value for any individual amino acid in all of the proteins. The high leucine content in protein 222 appears to be balanced by a very low content of isoleucines (.7%). It would be interesting to explore why this protein uses leucine residues to the almost exclusion of isoleucine residues. Phenylalanine and valine residues have average contents of 5.8% and 7.8%, respectively, and none of the proteins have statistically high contents of these (Figure 1 and Supplementary Table S3). The only trend for high contents of individual types of hydrophobic residues is in members of the Drug/Metabolite Transporter Superfamily (Figure 3). This property may be for provision of favourable interactions with lipophilic drug substrates. The combined content of hydrophobic residues in all proteins is relatively stable, however, with none having a statistically high value (Figure 2 and Supplementary Table S3). Hence, those proteins that have a statistically high content of one type of hydrophobic residue also tend to have lower contents of one or more other hydrophobic residues to balance the combined content, as demonstrated by the few examples described above.

3.1.2. Electrically charged residues (at physiological pH)

Basic and acidic residues with electrically charged side chains at physiological pH (arginine/lysine and aspartic
acid/glutamic acid, respectively) provide hydrogen bonding interactions that stabilise the structure of the protein, control its folding and insertion in the membrane and contribute favourable interactions with the hydrophilic surface of the membrane lipid bilayer and with substrates and ligands. Arginine residues have an average content of 4.2% and just one protein has a statistically high content (Figure 1 and Supplementary Table S3). This is protein 333 (NP_602297, SLC26A1), a sulphate:anion exchanger of the Sulphate Permease family with an arginine content of 8.5%. Clusters of high arginine contents are also apparent in the Anion Exchanger, Cation-Chloride Co-transporter, Chloride Carrier/Channel and Mitochondrial Carrier families, thus providing obvious links between amino acid content and specific substrates and function. Lysine residues have an average content of 3.8% and just one protein has a statistically high content (Figure 1 and Supplementary Table S3). This is protein 153 (NP_112581, SLC25A31), an ADP:ATP translocase of the Mitochondrial Carrier family with a lysine content of 8.6%. In the same way as other members of the Mitochondrial Carrier family, this protein gives variable results when its sequence is analysed by membrane topology prediction tools. Clusters of high lysine contents are also particularly noticeable in the Mitochondrial Carrier and Sulfate Permease families. The average combined content for arginine and lysine residues is 8.0% and five members of the Mitochondrial Carrier family have a statistically high combined content of these (Figure 2 and Supplementary Table S3). These include four ADP:ATP translocases that occupy a distinct branch of the family (Figure 3) and a separate tricarboxylate transporter. Indeed, the whole cluster of the Mitochondrial Carrier family (proteins 131-171) has a noticeably high content of combined positively charged residues (Figure 2 and Supplementary Table S3). A large majority of the Mitochondrial Carrier family proteins also give zero putative transmembrane helices when their sequences are analysed by the membrane topology prediction tool, so a possible link and an explanation for these observations is needed. The binding pocket in ADP:ATP translocases contains a number of basic residues that allow the strong binding of ADP or ATP (Nelson, Lawson, Klingenberg, & Douglas, 1993), which likely accounts for the overall high content of basic residues. For example, comparing the distribution of positively charged residues in protein 153 (NP_112581, SLC25A31) with analysis of its membrane topology suggests that these residues make a significant contribution to the missing transmembrane helices in topology predictions of Mitochondrial Carrier family proteins (Figure 4(C)). It is presumed that the positively charged residues in the binding sites, which appear in the middle of the transmembrane helices, cannot be handled by membrane topology prediction tools such that these helices are not identified properly.

Aspartic acid residues have an average content of 3.1% and 12 proteins have a statistically high content (Figure 1 and Supplementary Table S3). These include clusters of proteins from the Anion Exchanger, Cation Diffusion Facilitator and Mitochondrial Carrier families. Clusters of high contents of aspartic acid residues are also apparent for proteins in the Calcium-Cation Antiporter, Cation:Proton Antiporter-1 and Sulfate Permease families. Glutamic acid residues have an average content of 4.4% and 8 proteins have a statistically high content (Figure 1 and Supplementary Table S3), which include a cluster of 4 proteins from the Calcium-Cation Antiporter family. The highest content of glutamic acid residues (14.1%) is found in a 128-residue beta-subunit from the Organic Solute Transporter family (protein 296: NP_849190, SLC51B), which has one putative transmembrane helix. Protein 296 also has the highest content of combined negatively charged residues (18.8%), for which the average across all proteins is 7.4% (Figure 2 and Supplementary Table S2). Protein 296 functions as the intestinal basolateral transporter responsible for bile acid export from enterocytes into portal blood (Ballatori et al., 2005). Three proteins with statistically high combined contents of negatively charged residues in the Mitochondrial Carrier family also have statistically high combined contents of positively charged residues (Figure 3), thus balancing the overall charge. Clusters of proteins with high contents of combined negatively charged residues are also noticeable in the Anion Exchanger, Calcium-Cation Antiporter and Chloride Carrier/Channel families (Figures 2 and 3).

The average difference in content of positively charged residues compared with negatively charged residues across all proteins (positive minus negative) is .6% (Figure 2 and Supplementary Table S2). Protein 247 (NP_696961, SLC22A47) and protein 296 (NP_849190, SLC51B) produce the largest negative values from this difference subtraction (−7.2 and −11.0%, respectively) and these proteins contain the highest contents of glutamic acid residues. A large majority of proteins in the Mitochondrial Carrier family produce noticeably high positive values, consistent with observations already discussed.

3.1.3. Polar uncharged residues with hydroxyl or amido groups

Polar uncharged residues with hydroxyl- or amido-containing side chains (serine/threonine and asparagine/glutamine, respectively) provide hydrophilic sites in the protein to facilitate internal contacts for protein stabilisation and to provide favourable interactions with the hydrophilic surface of the membrane lipid bilayer, the aqueous environment or with hydrophilic substrates and ligands.
Serine residues have an average content of 7.5% and 9 proteins have a statistically high content (Figure 1 and Supplementary Table S3). Whilst these include Mitochondrial Carrier protein 142 (NP_057696, SLC25A37), the large majority of this family tend to have a relatively low content of serines. Clusters of high serine contents are apparent in proteins of the Cation Diffusion Facilitator, Neurotransmitter: Sodium Symporter and Sulfate Permease families. Threonine residues have an average content of 5.5% and 13 proteins have a statistically high content (Figure 1 and Supplementary Table S3). These include protein 42 of the Calcium-Cation Antiporter family, 2 proteins from the Amino Acid/Auxin Permease family and seven Mitochondrial Carrier family proteins. Four of the Mitochondrial Carrier members are uncoupling proteins that occupy a distinct branch in the family (Figure 3). The average combined content of serine and threonine residues is 13.0% and five proteins have a statistically high combined content of these (Figure 2 and Supplementary Table S3). In addition to protein 42 (NP_004718, SLC24A1), these include a sodium-dependent amino acid transporter of the Amino Acid/Auxin Permease family (protein 9: NP_109599, SLC38A1) (Wang et al., 2000) and hENT3 of the Equilibrative Nucleoside Transporter family (protein 123: NP_060814, SLC29A4) (Baldwin et al., 2005). The highest combined content of serine and threonine residues (19.1%) is found in a sodium ion: phosphate symporter of the Major Facilitator Superfamily (protein 194: NP_005826, SLC17A2).

Asparagine residues have an average content of 3.0% and two proteins have a statistically high content (Figure 1 and Supplementary Table S3). These are a Cystinosin homologue of the Lyosomal Cystine Transporter family (protein 130: NP_004928, CTNS) (Chiaverini et al., 2012) and Sideroflexin-1 of the Mitochondrial Tricarboxylate Carrier family (protein 259: NP_073591, SFXN1). Glutamine residues have an average content of 3.4% and 11 proteins have a statistically high content (Figure 1 and Supplementary Table S3), which include a cluster of the Mitochondrial Carrier family. The highest content of glutamines (7.6%) is found in a sulphate/anion exchanger of the Sulphate Permease Family (protein 333: NP_602297, SLC26A1), which has four putative transmembrane helices based on the topology prediction, whilst published information for this protein suggests twelve transmembrane helices (Regeer, Lee, & Markovich, 2003). Indeed, the whole of the Mitochondrial Carrier and Sulfate Permease families have a relatively high content of glutamines. The average combined content of asparagine and glutamine residues is 6.4%. Whilst none of the proteins have a statistically high combined content of these, there is a trend for relatively high contents in the Mitochondrial Carrier and Sulfate Permease families (Figure 2 and Supplementary Table S3).

3.1.4. Functionalisled aromatic residues

In addition to providing hydrophobic interactions, histidine, tryptophan and tyrosine residues provide functionalised sites that stabilise the structure of the protein and interact with substrates and ligands. Such residues tend to have a low content in the proteins and be located only at specific sites of structural stabilisation and functionality. Histidine residues have an average content of 1.8% and eight proteins have a statistically high content (Figure 1 and Supplementary Table S3). These are dominated by a cluster of six zinc ion transporters of the Cation Diffusion Facilitator family for which protein 62 (NP_598003, SLC30A4) has the highest content of 7.2%. Protein 62 is 376-residue zinc transporter 7, which has six putative transmembrane helices and appears to facilitate zinc transport from the cytoplasm into the Golgi apparatus (Kirschke & Huang, 2003). Twenty-one out of the 27 histidine residues in this protein are located in the intracellular loop between transmembrane helices IV and V (Figure 4(D)). A high content of histidine residues is clearly correlated with a function of zinc ion transport, which we recognised in our AAC analysis of secondary transporters from E. coli (Saidijam & Patching, 2015), exemplified by the Cation Diffusion Facilitator family zinc exporter ZitB (Grass et al., 2001; Lee et al., 2002; Rahman et al., 2008). Clusters of relatively high histidine contents are also apparent in proteins of the Anion Exchanger and Sulfate Permease families. Tryptophan residues have an average content of 1.6%, which is the lowest average value for any individual type of amino acid. Two members of the Neurotransmitter:Sodium Symporter family (protein 272: NP_005620, SLC6A4 and protein 273: NP_009162, SLC6A14) contain statistically high contents of tryptophans with values of 3.8 and 3.9%, respectively (Figure 1 and Supplementary Table S3). Clusters of relatively high tryptophan contents are also apparent for the rest of the Neurotransmitter: Sodium Symporter family and for the Organo Anion Transporter family. This provides an obvious link for favourable interactions with their neurotransmitter and organic anion substrates, respectively. Tyrosine residues have an average content of 3.2% and five proteins have a statistically high content (Figure 1 and Supplementary Table S3). These include two members of the Mitochondrial Carrier family (protein 158: NP_112489, SLC25A28 and protein 160: NP_110407, MFTC) with contents of 6.2% and 6.3%, respectively. The highest content of tyrosines (7.7%) is found in protein 173 (NP_055528, LAPTM4A) of the putative four-helix Multidrug Endosomal Transporter family. In the C-termini of proteins from this family there are hydrophilic domains containing several tyrosine-based sorting motifs YXXHy, where Hy represents a bulky hydrophobic residue, which direct the proteins to specific intracel-
lular compartments (Hogue, Nash, Ling, & Hobman, 2002). Also, clusters of high tyrosine contents are particularly noticeable for proteins in the Drug/Metabolite Transporter, Neurotransmitter:Sodium Symporter and Solute:Sodium Symporter families.

3.1.5. Sulphur-containing residues

Native cysteine residues in membrane proteins provide stabilisation of structure through the formation of disulphide bonds, which are often found in extracellular domains, and they provide active sites for the binding of cationic substrates and ligands such as metal ions. Cysteine residues have an average content of 2.1% and 17 proteins have a statistically high content (Figure 1 and Supplementary Table S3). These include two members of the Cation Diffusion Facilitator family (protein 60: NP_061183, SLC30A10 and protein 61: NP_067017, SLC30A1) that transport manganese and zinc, respectively. The equal highest content of cysteine residues (6.2%) is found in two isoforms (Eggermont et al., 1997) of the Chloride Ion Channel family (protein 74: NP_068505, CLCN6 and protein 76: NP_068503, CLCN6), which both begin with the sequence ‘MAGCRGSLLCCCRWCCC CGE’. Other proteins with statistically high cysteine contents include a cluster of seven members of the putative twelve helix Organo Anion Transporter family (Figure 3). Indeed, the whole of the Organo Anion Transporter family contain a noticeably high content of cysteines, 10 of which are conserved in a large extracellular domain between putative transmembrane helices IX and X where they are involved in disulphide bonds and possibly required for proper surface expression of the proteins (Hagenbuch & Stieger, 2013; Hänggi, Grundschöber, Leuthold, Meier, & St-Pierre, 2006). Methionine residues have an average content of 2.9% and eight proteins have a statistically high content (Figure 1 and Supplementary Table S3). The highest content of methionines (5.7%) is in protein 172 (NP_006753, LAPTM5) of the Multidrug Endosomal Transporter family.

3.1.6. Proline residues

Proline residues play a number of structural and dynamic roles in membrane proteins. These include sites of helix kink, transmission elements of conformational changes during transport processes and participation in polar interactions with substrates and ligands (Deber & Therien, 2002). Proline residues have an average content of 5.1% and 9 proteins have a statistically high content (Figure 1 and Supplementary Table S3). The large majority of proline residues in these proteins are not located within the putative transmembrane helices. For example, in protein 195 (NP_006508, SLC7A28), a 613-residue putative 12-helix monocarboxylate transporter of the Major Facilitator Superfamily, 33 out of 57 proline residues are in an N-terminal intracellular region and only eleven within the predicted transmembrane helices (Figure 4E)). The highest content of proline residues (11.4%) is in the 140-residue putative thiamine/folate transporter of the Reduced Folate Carrier family (protein 306: NP_064546, C2orf83) that has a single putative transmembrane helix.

3.2. Comparison of AAC in human and bacterial proteins

Comparing the average amino acid contents in the 336 human proteins analysed in this work with those in 235 secondary transport proteins from E. coli (Saidijam & Patching, 2015) reveals a similar overall trend along with some differences in contents of individual amino acids and in groups of amino acids with similar physicochemical properties (Figure 5). All hydrophobic residues (alanine, isoleucine, leucine, phenylalanine, valine) and glycine have varying degrees of higher contents in the bacterial over the human proteins. Alanine is the individual amino acid having the largest difference with average contents of 10.8 and 8.2% in the bacterial and human proteins, respectively. The combined average contents of hydrophobic residues is 47.0% (55.9% with glycines) and 40.5% (48.6% with glycines) in the bacterial and human proteins, respectively. These differences may simply be a consequence of a greater requirement for non-hydrophobic residues in the human proteins to satisfy more complex functions, hence decreasing the content of hydrophobic residues. This agrees with the observation that all charged residues (arginine, lysine, aspartic acid, glutamic acid) have slightly higher average contents in the human over the bacterial proteins. The average difference in content between positively and negatively charged residues (positive minus negative) has a small positive value in both the bacterial and human proteins with a higher value in the bacterial proteins (1.6 and .6%, respectively). The average content of cysteine residues in the human proteins is approximately twice that in the bacterial proteins with values of 2.1 and 1.1%, respectively. This agrees with a study that demonstrated a general trend for higher cysteine contents in proteins from higher organisms, which is explained by an evolutionary introduction of a greater number of uses for cysteine residues (Miseta & Csutora, 2000), for example in CxxC motifs and derivatives with redox functions (Fomenko & Gladyshev, 2003). Methionine residues have a slightly higher average content in the bacterial proteins than in the human proteins with values of 4.0 and 2.9%, respectively. All polar uncharged residues with hydroxyl or amido groups (serine, threonine, asparagine, glutamine), functionalised aromatic residues (histidine, tryptophan, tyrosine) and proline residues
have relatively similar average contents in the bacterial and human proteins. In terms of deviation in contents of individual amino acids, the greatest value is seen for leucine residues in both bacterial and human proteins. The greatest differences in the deviations between the two sets of proteins are seen for cysteine, glutamic acid, leucine and lysine residues, which have higher values in the human proteins in all cases. The greatest differences in the deviations for groupings of amino acids with similar physicochemical properties are seen for the groups of charged residues with higher values for the human proteins. These differences in deviation presumably reflect the greater complexity and variety in functions of the human proteins over the bacterial proteins.

During the review process for this article we received the following comments from one of the referees: While the authors claim to have addressed the concerns raised in my initial review, unfortunately, the some of the treatment is very superficial with hand-waving arguments only. Figures 1, 2 and new Figure 4 are certainly an improvement over the previous version. However, while the authors are stressing on the importance of amino acid compositions, I am very unclear on their reluctance to allow a proper and more rigorous analysis to be performed. This is beyond the scope of and not appropriate for the current paper on human secondary transporters, but we may be interested in performing a proper comparative study for a future publication. In addition to the reasons given in the response with the previous revision, the results of Sharpe et al. are presented as normalised probabilities for amino acids at each individual residue position in the transmembrane domains of bitopic proteins and therefore not directly comparable with our AAC results on secondary transporters. The referee also refers to ‘soluble proteins and other bitopic proteins’, but the proteins in our AAC analysis of secondary transporters are polytopic and not bitopic. This information was included with approval by the journal Editor.

3.3. Implications of AAC for reliable membrane topology prediction

Earlier we saw how proteins with a statistically high content of positively charged residues tend to give a much lower number of putative transmembrane helices...
Figure 6. Comparison of amino acid content with the number of putative transmembrane helices in 336 human secondary transport proteins.

Notes: (A) Total number of residues in the proteins. (B) Combined content of hydrophobic residues (alanine, isoleucine, leucine, phenylalanine, valine) in the proteins, taken from Figure 2. (C) Combined content of positively charged residues (arginine, lysine) in the proteins, taken from Figure 2. (D) Combined content of negatively charged residues (aspartic acid, glutamic acid) in the proteins, taken from Figure 2. (E) Putative number of transmembrane helices in the proteins resulting from analyses of their sequences by the membrane topology prediction tool TOPCONS (http://topcons.cbr.su.se/) (Bernsel et al., 2009), which are also listed in Supplementary Table S1. (F) Transporter family grouping of the proteins, which corresponds with Supplementary Table S1. The vertical grey lines separate the 36 different families of transport proteins, which are listed in the legend for Figure 1.
Figure 7. Comparison of amino acid contents with the number of putative transmembrane helices in families of human secondary transport proteins.

Notes: Thirty-six different transporter families match those given in Supplementary Table S1 and in the legend for Figure 1. The black dots represent the average number of total residues in the proteins from each transporter family, which were calculated from the data shown in Figure 6(A). The red and blue dots represent the average combined contents of positively charged residues and negatively charged residues, respectively, in the proteins from each transporter family, which were calculated from the data shown in Figure 6(C) and (D). The grey columns represent the average number of predicted transmembrane spanning helices in the proteins from each transporter family, which were calculated from the data shown in Figure 6(E). The error bars represent the standard deviations of the means. Average values and standard deviations for the data shown in this Figure are given in Supplementary Table S4.

than expected when their sequences are analysed by a membrane topology prediction tool. This trend is especially evident for the whole of the Mitochondrial Carrier family (proteins 139-171) when the number of putative transmembrane helices is compared with AACs in all 336 human proteins and the 36 transporter families (Figures 6 and 7). Indeed, Mitochondrial Carriers are the only family with a statistically high average content of positively charged residues, which aligns with a statistically low number of predicted transmembrane helices (Supplementary Table S4). Based on the total number of residues and on the hydrophobic residue content in these proteins, several transmembrane helices are expected and crystal structures of Mitochondrial Carrier family proteins have confirmed this to be the case, for example, the six transmembrane helices typically found in ADP:ATP translocases (Monné & Palmieri, 2014; Ruprecht et al., 2014). Comparison of hydrophobicity in crystal structures and amino acid sequences of a bovine mitochondrial ADP:ATP translocase (PDB 1OKC; Pebay-Peyroula et al., 2003) and the human glucose transporter GLUT1 (protein 196: NP_006507, SLC2A1) (PDB 4PYP; Deng et al., 2014) illustrates a higher continuous intensity of hydrophobic residues in the transmembrane helices of GLUT1 (Figure 8). The transmembrane helices of the Mitochondrial Carrier protein are interspersed with a greater number of non-hydrophobic residues, especially positively charged residues, which correlates with the misleading membrane topology predictions for this type of protein. Other proteins with a lower than expected number of transmembrane helices, zero in some cases, that contain a high content of charged or polar residues include an organic anion transporter of the Major Facilitator Superfamily (protein 247: NP_696961, SLC22A7), all three members of the Mitochondrial Tricarboxylate Carrier family (protein 257: NP_112233, SFxn3; protein 258: NP_653180, SFxn5; protein 259: NP_073591, SFxn1), and two sulphate/anion exchangers of the Sulphate Permease family (protein 331: NP_597996, SLC26A10 and protein 333: NP_602297, SLC26A1) (Figure 6(E)). It appears that charged or polar residues located in the middle of a transmembrane helix can interfere with their identification by membrane topology tools resulting in missed helices in the prediction. This observation is not limited to the TOPCONS tool that was used for membrane topology predictions in this work. For example, analysis of sequences using the tool TMHMM predicts 12 transmembrane helices for GLUT1, but missing transmembrane helices in the bovine mitochondrial ADP:ATP translocase (Figure 8). This should be taken into account when performing topology predictions of membrane proteins and it may be useful to use multiple prediction tools for a more comprehensive analysis. The graphs of membrane topology results for the 336 proteins and 36 transporter families (Figures 6(E) and 7) also reveal highly consistent numbers of eleven putative transmembrane helices in the Anion Exchanger family (proteins 14-24), 12 helices in the Amino Acid-Polyamine-Organocation family (proteins 25-35), 6 helices in the Cation Diffusion Facilitator family (proteins 59-68), 11 helices in the Equilibrative Nucleoside Transporter family (proteins 122-125), 4 helices in the Multidrug Endosomal Transporter family (proteins 172-174), 12 helices in the Neurotransmitter: Sodium Symporter and Organo Anion Transporter families (proteins 266-282 and 283-294) and 8 helices in the
Zinc–Iron Permease family (proteins 335-336). General trends are revealed for 12 or 13 helices in the CCC family (proteins 50-58), 12 helices in the Major Facilitator Superfamily (proteins 175-254) and 13 or 14 helices in the Solute:Sodium Symporter family (proteins 309-318). Other families showed a more variable number of putative transmembrane helices.

A recent topological investigation of the entire human transmembrane proteome identified some of the reasons for inaccuracies in predicting the number of...
transmembrane helices in human proteins using existing methods (Dobson et al., 2015a). These included the possibility that the existing topology prediction algorithms were trained and tested on benchmark sets containing mostly prokaryotic transmembrane helices, whose properties can differ from eukaryotic transmembrane proteins. Another reason was a potential inability of existing algorithms to predict the presence of N-terminal signal peptides, which share similar physicochemical properties with transmembrane helices (Bendtsen et al., 2004; Käll et al., 2004; Bendtsen et al., 2007; Petersen et al., 2011). The TOPCONS tool that we used for analysis of the human secondary transporters has been updated to efficiently separate any N-terminal signal peptides from transmembrane regions, however (Tsirigos et al., 2015). Our analysis of human secondary transporters has demonstrated how AAC can influence the reliability of topology predictions for membrane proteins, especially due to the presence of charged and/or polar residues in the middle of transmembrane helices. Unless the topology prediction methods can handle such residues, an element of manual inspection is required along with use of other information such as a known number of experimentally determined transmembrane helices in homologous proteins, for example, from crystal structures. Otherwise, human membrane proteins with opportunities for experimental investigation or as drug targets may be missed. A new membrane topology prediction tool called Consensus Constrained TOPology (CCTOP, http://cctop.enzim.ttk.mta.hu), which uses ten different prediction methods and incorporates topology information from existing experimental and computational sources (Dobson et al., 2015a; Dobson et al., 2015b), appears to be more capable of handling the presence of charged and/or polar residues in the middle of transmembrane helices. For example, analysis of sequences for the human Mitochondrial Carrier family (proteins 131-171) using CCTOP, predicts an expected number of up to six transmembrane helices in the large majority of cases.

4. Conclusions

Human secondary transport proteins are responsible for transporting a wide range of ions, metals, organic and inorganic solutes involved in energy transduction, control of membrane potential and osmotic balance, metabolic processes and in the absorption or efflux of drugs and xenobiotics. They are also emerging as important potential targets for development of new drugs as well as being target sites for drug delivery to specific organs or tissues, especially for allowing drugs to cross the blood–brain barrier. AAC and phylogenetic analyses of 336 human secondary transport proteins have confirmed family classifications and allowed an inspection for trends in AAC and correlations with structural domains and specific substrates and/or function. In this respect, some proteins showed statistically high contents of individual amino acids or of groups of amino acids with similar physicochemical properties. One recurring trend identified by the analysis was a correlation between high contents of charged and/or polar residues with misleading results in predictions of membrane topology, which was especially prevalent in Mitochondrial Carrier family proteins. We demonstrated how charged and/or polar residues located in the middle of transmembrane helices can interfere with their identification by membrane topology tools resulting in missed helices in the prediction. When predicting the number of transmembrane helices in membrane proteins it is therefore useful to use a number of membrane topology tools, manual inspection of protein sequences and other information such as a known number of experimentally determined transmembrane helices in homologous proteins, for example, from crystal structures. Otherwise, the human transmembrane proteome will not be assigned completely and accurately and human membrane proteins with opportunities for experimental investigation or as drug targets may be missed. Comparison of AAC in the human proteins with that in 235 secondary transport proteins from E. coli revealed similar overall trends along with differences in average contents for some individual amino acids and groups of similar amino acids. These included higher average contents of charged residues in the human proteins compared with the bacterial proteins. The differences are presumed to result from a greater number of functions and complexity in the higher organism. The bioinformatics methods of AAC analysis, phylogenetic analysis and membrane topology prediction, which are all based on protein sequence information, are important tools to use alongside experimental techniques for achieving a complete and accurate picture and understanding of the human transmembrane proteome leading to structural and functional characterisation. This includes full identification and classification, determination of structural fold and of specific substrates and function, elucidation of molecular mechanism and characterisation of ligand interactions, dynamics and multidimensional relationships. Such information has potential to lead pathways for the discovery of new drugs and for the tissue-specific delivery of drugs in the treatment of human diseases.

Supplemental material

The supplemental material for this paper is available at http://dx.doi.org/10.1080/07391102.2016.1167622.
Johansson, A. C., & Lindahl, E. (2009). Protein contents in biological membranes can explain abnormal solvation of charged and polar residues. Proceedings of the National Academy of Sciences the United States of America, 106, 15684–15689.

Käll, L., Krogh, A., & Sonnhammer, E. L. L. (2004). A combined transmembrane topology and signal peptide prediction method. Journal of Molecular Biology, 338, 1027–1036.

Käll, L., Krogh, A., & Sonnhammer, E. L. L. (2007). Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server. Nucleic Acids Research, 35, W429–W432.

Kirschke, C. P., & Huang, L. (2015). The potential clinical impact of the yeast mitochondrial ADP/ATP translocator. Nature Reviews Drug Discovery, 5, 993–1002.

Klepper, J., & Voit, T. (2002). Facilitated glucose transporter protein type 1 (GLUT1) deficiency syndrome: Impaired glucose transport into brain - a review. European Journal of Pediatrics, 161, 295–304.

Krogh, A., Larsson, B., von Heijne, G., & Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden markov model: Application to complete genomes. Journal of Molecular Biology, 305, 567–580.

Kuan, J., & Saier, M. H. Jr. (1993). The mitochondrial carrier family of transport proteins: Structural, functional, and evolutionary relationships. Critical Reviews in Biochemistry and Molecular Biology, 28, 209–233.

Kurz, A. K., Gallicchio, G., Heine, C., Mole, S. E., Quitsch, A., & Braulke, T. (2010). Pathogenic mutations cause rapid degradation of lysosomal storage disease-related membrane protein CLN6. Human Mutation, 31, E1163–E1174.

Lee, S. M., Grass, G., Haney, C. J., Fan, B., Rosen, B. P., Anton, A., ... Rensing, C. (2002). Functional analysis of the Escherichia coli zinc transporter ZitB. FEMS Microbiology Letters, 215, 273–278.

Letunic, I., & Bork, P. (2007). Interactive Tree Of Life (iTOL): An online tool for phylogenetic tree display and annotation. Bioinformatics, 23, 127–128.

Letunic, I., & Bork, P. (2011). Interactive Tree Of Life v2: Online annotation and display of phylogenetic trees made easy. Nucleic Acids Research, 39, W475–W478.

Liu, H., & Ding, H. (2011). Predicting ion channels and their types by the dipeptide mode of pseudo amino acid composition. Journal of Theoretical Biology, 269, 64–69.

Lin, L., Yee, S. W., Kim, R. B., & Giacomini, K. M. (2015). SLC transporters as therapeutic targets: Emerging opportunities. Nature Reviews Drug Discovery, 14, 543–560.

Lindskog, C. (2015). The potential clinical impact of the tissue-based map of the human proteome. Expert Review of Proteomics, 12, 213–215.

Liu, J., & Rost, B. (2001). Comparing function and structure between entire proteomes. Protein Science, 10, 1970–1979.

Liu, Q., Zhu, Y., Wang, B., & Li, Y. (2003). Identification of β-barrel membrane proteins based on amino acid composition properties and predicted secondary structure. Computational Biology and Chemistry, 27, 355–361.

Lundstrom, K. (2006). Latest development in drug discovery on G protein-coupled receptors. Current Protein and Peptide Science, 7, 465–470.

McKerian, C. J., & Friedlander, M. (1999). The retinal rod Na+/Ca2+–K+ exchanger contains a noncleaved signal sequence required for translocation of the N Terminus. Journal of Biological Chemistry, 274, 38177–38182.

Mezei, M. (2011). Discriminatory power of stoichiometry-driven protein folding? Journal of Biomolecular Structure and Dynamics, 28, 625–626.

Mikitish, J. L., & Chacko, A. M. (2014). Pathways for small molecule delivery to the central nervous system across the blood-brain barrier. Perspectives in Medicinal Chemistry, 6, 11–24.

Miseta, A., & Csutora, P. (2000). Relationship between the occurrence of cysteine in proteins and the complexity of organisms. Molecular Biology and Evolution, 17, 1232–1239.

Mishra, S. (2011). ‘Chargaff’s rules’ for protein folding: Stoichiometric leitmotif made visible. Journal of Biomolecular Structure and Dynamics, 28, 649–652.

Mishra, N. K., Chang, J., & Zhao, P. X. (2014). Prediction of membrane transport proteins and their substrate specificities using primary sequence information. PLoS One, 9, e100278.

Mittal, A., & Jayaram, B. (2011). Backbones of folded proteins reveal novel invariant amino acid neighborhoods. Journal of Biomolecular Structure and Dynamics, 28, 443–454.

Mittal, A., Jayaram, B., Shenoy, S., & Bawa, T. S. (2010). A stoichiometry driven universal spatial organization of backbones of folded proteins: Are there Chargaff’s rules for protein folding? Journal of Biomolecular Structure and Dynamics, 28, 133–142.

Mohabatkar, H., Mohammad Beigi, M., & Esmaeili, A. (2011). Prediction of GABAA receptor proteins using the concept of Chou’s pseudo-amino acid composition and support vector machine. Journal of Theoretical Biology, 281, 18–23.

Monné, M., & Palmieri, F. (2014). Antiporters of the mitochondrial carrier family. Current Topics in Membranes, 73, 289–320.

Moreland, J. L., Gramada, A., Buzko, O. V., Zhang, Q., & Bourne, P. E. (2005). The Molecular Biology Toolkit (MBT): A modular platform for developing molecular visualization applications. BMC Bioinformatics, 6, 21.

Nelson, D. R., Lawson, J. E., Klingenberg, M., & Douglas, M. G. (1993). Site-directed mutagenesis of the yeast mitochondrial ADP/ATP translocator. Six arginines and one lysine are essential. Journal of Molecular Biology, 230, 1159–1170.

Nishikawa, K., & Ooi, T. (1982). Correlation of the amino acid composition of a protein to its structural and biological characteristics. Journal of Biochemistry, 91, 1821–1824.

Nury, H., Blesnac, I., Ravaud, S., & Pebbey-Peyroula, E. (2010). Structural approaches of the mitochondrial carrier family. Methods in Molecular Biology, 654, 105–117.

Nury, H., Dahlout-Gonzalez, C., Trézéguet, V., Lauquin, G., Brandolin, G., & Pebbey-Peyroula, E. (2005). Structural basis for lipid-mediated interactions between mitochondrial ADP/ATP carrier monomers. FEBS Letters, 579, 6031–6036.

Ou, Y. Y., Chen, S. A., & Gromiha, M. M. (2010). Classification of transporters using efficient radial basis function networks with position-specific scoring matrices and biochemical properties. Proteins, 78, 1789–1797.

Overington, J. P., Al-Lazikani, B., & Hopkins, A. L. (2006). How many drug targets are there? Nature Reviews Drug Discovery, 5, 993–996.

Palmieri, F. (1994). Mitochondrial carrier proteins. FEBS Letters, 346, 48–54.

Pao, S. S., Paulsen, I. T., & Saier, M. H. Jr. (1998). Major facilitator superfamily. Microbiology and Molecular Biology Reviews, 62, 1–34.
Partridge, A. W., Therien, A. G., & Deber, C. M. (2002). Polar mutations in membrane proteins as a biophysical basis for disease. Biopolymers, 66, 350–358.

Pascal, G., Médigue, C., & Danchin, A. (2006). Persistent biases in the amino acid composition of prokaryotic proteins. BioEssays, 28, 726–738.

Patching, S. G. (2015). Roles of facilitative glucose transporter GLUT1 in [18F]FDG positron emission tomography (PET) imaging of human diseases. Journal of Diagnostic Imaging in Therapy, 2, 30–102.

Patching, S. G. (2016). Glucose transporters at the blood-brain barrier: Function, regulation and gateways for drug delivery. Molecular Neurobiology [Epub ahead of print], doi: 10.1007/s12053-015-9672-6.

Patel, M. M., Goyal, B. R., Bhadada, S. V., Bhatt, J. S., & Amin, A. F. (2009). Getting into the brain: Approaches to enhance brain drug delivery. CNS Drugs, 23, 35–58.

Pebay-Peyroula, E., Dahout-Gonzalez, C., Kahn, R., Trézaguet, V., Lauquin, G. J., & Brandolin, G. (2003). Structure of mitochondrial ADP/ATP carrier in complex with carbohydrate-lycoylase. Nature, 426, 39–44.

Petersen, T. N., Brunak, S., von Heijne, G., & Nielsen, H. (2011). SignalP 4.0: Discriminating signal peptides from transmembrane regions. Nature Methods, 8, 785–786.

Quadri, M., Federico, A., Zhao, T., Brevedveld, G. J., Battisti, C., Delnooz, C., … Bonifati, V. (2012). Mutations in SLC30A10 cause parkinsonism and dystonia with hypermanganesemia, polycthemia, and chronic liver disease. The American Journal of Human Genetics, 90, 467–477.

Ragona, F., Marcolin, S., Castellotti, B., Patrini, M., Ferri, E., Binelli, S., & Granata, T. (2014). Refractory absence epilepsy and glut1 deficiency syndrome: A new case report and literature review. Neuropediatrics, 45, 328–332.

Rahman, M., Patching, S. G., Ismat, F., Henderson, P. J., Herbert, R. B., Baldwin, S. A., & McPherson, M. J. (2008). Probing metal ion substrate-bind to the E. coli ZntB exporter in native membranes by solid state NMR. Molecular Membrane Biology, 25, 683–690.

Rask-Andersen, M., Almén, M. S., & Schiöth, H. B. (2011). Trends in the exploitation of novel drug targets. Nature Reviews Drug Discovery, 10, 579–590.

Reddy, V. S., Shlykov, M. A., Castillo, R., Sun, E. I., & Saier, M. H. Jr. (2012). The major facilitator superfamily (MFPS) revisited. FEBS Journal, 279, 2022–2035.

Regeer, R. R., Lee, A., & Markovich, D. (2003). Characterization of the human sulfate anion transporter (hsat-1) protein and gene (SAT1; SLC26A1). Journal of Molecular Structure, 66, 143–158.

Ren, Q., Chen, K., & Paulsen, I. T. (2007). TransportDB: A comprehensive database resource for cytoplasmic membrane transport systems and outer membrane channels. Nucleic Acids Research, 35, D274–D279.

Rosenbaum, D. M., Rasmussen, S. G., & Kobikla, B. K. (2009). The structure and function of G-protein-coupled receptors. Nature, 459, 356–363.

Ruprecht, J. J., Hellawell, A. M., Harding, M. C., Crichton, P. G., McCoy, A. J., & Kunji, E. R. (2014). Structures of yeast mitochondrial ADP/ATP carriers support a domain-based alternating-access transport mechanism. Proceedings of the National Academy of Sciences the United States of America, 111, E426–E434.

Saidijam, M., & Patching, S. G. (2015). Amino acid composition analysis of secondary transport proteins from Escherichia coli with relation to functional classification, ligand specificity and structure. Journal of Biomolecular Structure and Dynamics, 33, 2205–2220.

Saier, M. H., Jr,Beaty, J. T., Goffeau, A., Harley, K. T., Heijne, W. H., Huang, S. C., … Virk, P. S. (1999). The major facilitator superfamily. Journal of Molecular Microbiology and Biotechnology, 1, 257–279.

Sanchez-Covarrubias, L., Slosky, L. M., Thompson, B. J., Davis, T. P., & Ronaldson, P. T. (2014). Transporters at CNS barrier sites: Obstacles or opportunities for Drug Delivery? Current Pharmaceutical Design, 20, 1422–1449.

Sanders, C. R., & Myers, J. K. (2004). Disease-related misassembly of membrane proteins. Annual Review of Biophysics and Biomolecular Structure, 33, 25–51.

Schaadt, N. S., & Helms, V. (2012). Functional classification of membrane transporters and channels based on filtered TM/non-TM amino acid composition. Biopolymers, 97, 558–567.

Sharpe, H. J., Stevens, T. J., & Munro, S. (2010). A comprehensive comparison of transmembrane domains reveals organelle-specific properties. Cell, 142, 158–169.

Shi, Y. (2013). Common folds and transport mechanisms of secondary active transporters. Annual Review of Biophysics, 42, 51–72.

Shukla, H. D., Vaitiekunas, P., & Cotter, R. J. (2012). Advances in membrane proteomics and cancer biomarker discovery: Current status and future perspective. Proteomics, 12, 3085–3104.

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., … Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular Systems Biology, 7, 539.

Stenbjheim, D. D., Hartz, A. M., Bauer, B., & Anderson, G. W. (2009). Novel and emerging strategies in drug delivery for overcoming the blood-brain barrier. Future Medicinal Chemistry, 1, 1623–1641.

Striano, P., Weber, Y. G., Toliat, M. R., Schubert, J., Leu, C., Chaimana, R., … Marini, C. (2012). GLUT1 mutations are a rare cause of familial idiopathic generalized epilepsy. Neurology, 78, 557–562.

Suls, A., Mullen, S. A., Weber, Y. G., Verhaert, K., Ceulemans, B., Guerrini, R., … Scheffer, I. E. (2009). Early-onset absence epilepsy caused by mutations in the glucose transporter GLUT1. Annals of Neurology, 66, 415–419.

Tashima, T. (2015). Intriguing possibilities and beneficial aspects of transporter-conscious drug design. Bioorganic and Medicinal Chemistry, 23, 4119–4131.

Tsiligros, K. D., Peters, C., Khu, N., Käll, L., & Elofsson, A. (2015). The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides. Nucleic Acids Research, 43, W401–W407.

Tusnády, G. E., & Simon, I. (2010). Topology prediction of helical transmembrane proteins: How far have we reached? Current Protein and Peptide Science, 11, 550–561.

Uhlen, M., Fagerberg, L., Hallström, B. M., Lindskog, C., Oksvold, P., … Pontén, F. (2015). Proteomics. Tissue-based map of the human proteome. Science, 347, 1260419.

von Heijne, G. (2007). The membrane protein universe: What’s out there and why bother? Journal of Internal Medicine, 261, 543–557.

Wallin, E., & von Heijne, G. (1998). Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. Protein Science, 7, 1029–1038.

Wang, H., Huang, W., Sugawara, M., Devoe, L. D., Leibach, F. H., Prasad, P. D., & Ganapathy, V. (2000). Cloning and functional expression of ATA1, a subtype of amino acid transporter A, from human placenta. Biochemical and Biophysical Research Communications, 273, 1175–1179.
Wang, J., Li, Y., Wang, Q., You, X., Man, J., Wang, C., & Gao, X. (2012). ProClusEnsem: Predicting membrane protein types by fusing different modes of pseudo amino acid composition. Computers in Biology and Medicine, 42, 564–574.

Watanabe, H., Koopmann, T. T., Le Scouarnec, S., Yang, T., Ingram, C. R., Schott, J. J., … Bezzina, C. R. (2008). Sodium channel β1 subunit mutations associated with Brugada syndrome and cardiac conduction disease in humans. Journal of Clinical Investigation, 118, 2260–2268.

Weber, Y. G., Storch, A., Wuttke, T. V., Brockmann, K., Kempfle, J., Maljevic, S., … Lerche, H. (2008). GLUT1 mutations are a cause of paroxysmal exertion-induced dyskinesias and induce hemolytic anemia by a cation leak. Journal of Clinical Investigation 118, 2157–2168.

Yan, N. (2013). Structural advances for the major facilitator superfamily (MFS) transporters. Trends in Biochemical Sciences, 38, 151–159.

Yan, N. (2015). Structural biology of the major facilitator superfamily transporters. Annual Review of Biophysics, 44, 257–283.

Zia-Ur-Rehman, & Khan, A. (2012). Identifying GPCRs and their types with Chou’s pseudo amino acid composition: An approach from multi-scale energy representation and position specific scoring matrix. Protein and Peptide Letters, 19, 890–903.