Cell

Resource

A Comprehensive Comparison of Transmembrane Domains Reveals Organelle-Specific Properties

Hayley J. Sharpe,[1] Tim J. Stevens,[2] and Sean Munro[1,*]
[1]MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 0QH, UK
[2]Department of Biochemistry, University of Cambridge, Cambridge CB2 1GA, UK
*Correspondence: sean@mrc-lmb.cam.ac.uk
DOI 10.1016/j.cell.2010.05.037

SUMMARY

The various membranes of eukaryotic cells differ in composition, but it is at present unclear if this results in differences in physical properties. The sequences of transmembrane domains (TMDs) of integral membrane proteins should reflect the physical properties of the bilayers in which they reside. We used large datasets from both fungi and vertebrates to perform a comprehensive comparison of the TMDs of proteins from different organelles. We find that TMDs are not generic but have organelle-specific properties with a dichotomy in TMD length between the early and late parts of the secretory pathway. In addition, TMDs from post-ER organelles show striking asymmetries in amino acid compositions across the bilayer that is linked to residue size and varies between organelles. The pervasive presence of organelle-specific features among the TMDs of a particular organelle has implications for TMD prediction, regulation of protein activity by location, and sorting of proteins and lipids in the secretory pathway.

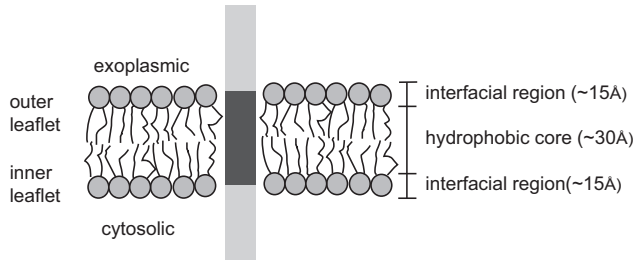
INTRODUCTION

Integral membrane proteins are encoded by ~30% of the genes in most genomes and perform numerous biological processes from signaling to transport (Almén et al., 2009; Stevens and Arkin, 2000). There are many indications that the activity of such proteins can be affected by physical properties of the lipid bilayer such as lipid order and hydrophobic thickness (Andersen and Koeppe, 2007; Bondar et al., 2009; Nyholm et al., 2007; Phillips et al., 2009). There is also considerable interest in the possibility that local differences in the physical properties of membranes could contribute to the lateral segregation of proteins during sorting or signaling (Bretscher and Munro, 1993; Dukhovny et al., 2009; Patterson et al., 2008; Ronchi et al., 2008; Simons and Ikonen, 1997). Determining the biological significance of such processes in eukaryotes is contingent on understanding the properties of the different bilayers of the cell. Organelle membranes vary in both their protein and lipid content, and even within one membrane the lipid composition of the two leaflets of the bilayer can be very different (van Meer et al., 2008). For instance, sterols and sphingolipids are scarce in the ER but abundant and asymmetrically distributed in the plasma membrane. These lipids differ from typical phospholipids in that sphingolipids are characterized by saturated acyl chains, and sterols by an inflexible core formed by four fused rings. In artificial liposomes the degree of acyl chain saturation and the levels of sterols affect such physical properties of the bilayer as thickness, order and viscosity (Brown and London, 1998). However, what effect they have at physiological levels in heterogeneous, protein-containing biological membranes is unclear.

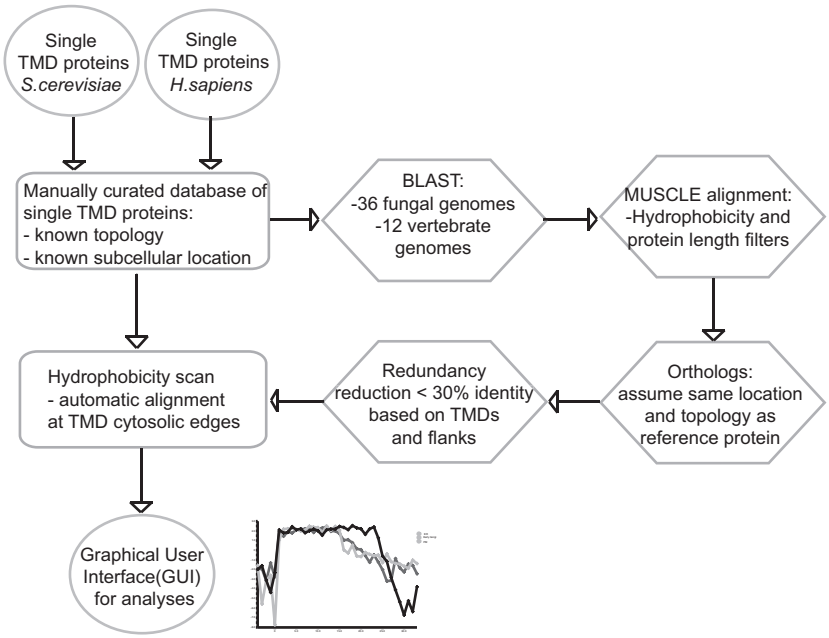
Most integral membrane proteins contain α-helical transmembrane domains (TMDs) that span the hydrophobic core of the lipid bilayer (Killian and von Heijne, 2000; White and Wimley, 1999). The primary constraint on all TMDs that enter the secretory pathway is that they must partition out of the Sec61 translocon into the membrane of the ER during synthesis. TMDs are greatly enriched in aliphatic hydrophobic residues, and these residues promote partitioning out of the translocon (Hessa et al., 2005, 2007; Killian and von Heijne, 2000). However, the physical properties of the bilayer in which a protein will eventually reside should also impose constraints upon the sequence of its TMD. Previous studies comparing the TMDs of Golgi and plasma membrane proteins have suggested a difference in TMD length and hence bilayer thickness (Bretscher and Munro, 1993; Levine et al., 2000). However, the full significance of this finding for cellular organization is unclear as the analysis was based on only a small number of proteins and did not include other organelles. Indeed the conclusions have been called into question by attempts to measure bilayer thickness of different compartments (Mitra et al., 2004).

To obtain a clear picture of organelle-specific constraints on TMDs, we have made use of the recent increase in available genome sequences to perform a comprehensive comparison of a large number of membrane proteins with a single TMD from the major secretory organelles from both fungi and vertebrates. Our findings validate the previous suggestions of a difference in TMD length between Golgi and plasma membrane and extend this to reveal an apparent step-change in bilayer thickness that occurs in the secretory pathway at the trans side of the Golgi. We also find that the TMDs of proteins from post-ER organelles show striking variations in amino acid composition across the bilayer. This results in an asymmetry in residue

A



B



C

Fungi	ER	Golgi	TGN/endo	vacuole	PM	Total
Reference proteins	32	30	8	5	17	92
Orthologs	767	463	163	110	145	1,648
Final (<30% identity)	502	309	99	94	115	1,119
Vertebrates	**ER**	**Golgi**	**TGN/endo**	**lyso**	**PM**	**Total**
Reference proteins	54	70	13	11	211	359
Orthologs	440	603	122	101	1,855	3,121
Final (<30% identity)	115	258	33	30	756	1,192

Figure 1. Overview of the Methodology for TMD Analysis

(A) Schematic of a typical single-pass or bitopic protein embedded in a lipid bilayer.

(B) Bitopic proteins of known topology and location from *S. cerevisiae* and *H. sapiens* were identified by literature and database searches. Orthologous proteins were identified using BLAST and aligned with the reference proteins. The starts of the TMDs were identified by a hydrophobicity scanning algorithm and used to align the TMDs at their cytosolic edges.

(C) The number of proteins from the indicated organelles that were used in the analyses of TMDs (PM, plasma membrane). Redundancy reduction was such that TMDs and flanking sequences have <30% identity. Reference proteins are listed in Table S1 and Table S2.

See also Figure S1.

composition that is linked to residue volume and correlates with changes in lipid asymmetry. Thus, eukaryotic TMDs are not a single type of entity but vary in a manner that implies that there are clear differences in the physical properties of the bilayers of the secretory pathway.

RESULTS

Computational Analysis of Fungal and Vertebrate Transmembrane Sequences from Distinct Subcellular Locations

To reliably compare TMDs that span different membranes, we curated a dataset of proteins with an experimentally determined topology and location and a single TMD (bitopic proteins, Figure 1A). Bitopic proteins represent ~40% of all membrane proteins in eukaryotic genomes, and their TMDs are those likely to have the most residues exposed to the lipid bilayer (Almén et al., 2009; Krogh et al., 2001). We assembled datasets of all single TMD proteins from what are probably the best characterized eukaryotic genomes, *Saccharomyces cerevisiae* and *Homo sapiens*. We then used literature searches and cross-referencing between databases to identify those proteins with a known organelle of residence and topology (Table S1 and Table S2). For the Golgi apparatus we pooled all the proteins from the various cisternae of the Golgi stack into a single "Golgi" set, with a separate set for those proteins that cycle between the trans-Golgi network (TGN) and endosomes. Only a few mammalian Golgi proteins have been accurately located within the Golgi stack, but for yeast, where this is more easily done, we found that the proteins of the early part of the stack were strikingly similar in TMD properties to those from the later part of the stack (see below), indicating that this pooling probably does not mask significant complexity.

Selecting only those proteins with a known location and topology inevitably reduced the size of the datasets, and so to expand the number of sequences available for analysis, we used BLAST searches to collect the orthologous proteins from all other complete fungal and vertebrate genomes. The topology and subcellular location of orthologs were assumed to be the same as for the reference protein. Many of their functions are highly organelle specific, and a global comparison of protein localization in the distantly related yeasts *S. cerevisiae* and

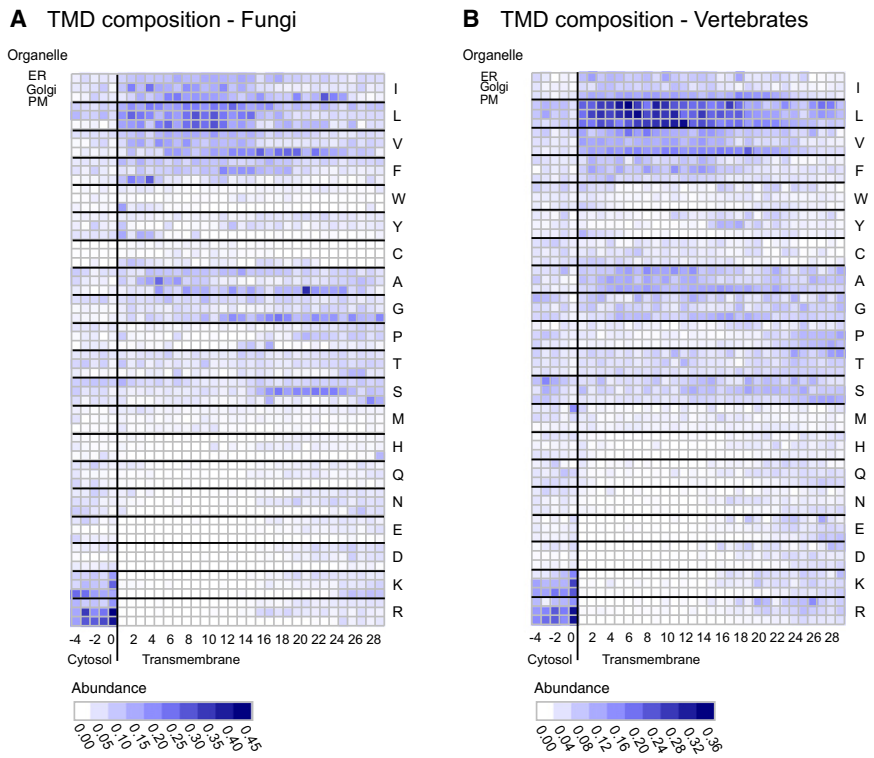


Figure 2. Positional Analysis of Amino Acid Composition of TMDs from Different Organelles in Fungi and Vertebrates

(A and B) The position relative to the cytosolic edge of the TMDs is on the horizontal axes, and the amino acids and organelles are on the vertical axes. Amino acids are listed in order of decreasing hydrophobicity (Goldman-Engelman-Steitz [GES] scale [Engelman et al., 1986]). Normalized residue abundance is color-coded such that white represents zero and dark blue a maximum of one. The abundance of serines in the region following the lumenal end of Golgi TMDs probably reflects the fact that this part of many Golgi enzymes forms a flexible linker that tethers the catalytic domain to the membrane (Paulson and Colley, 1989). Graphical plots for individual residues can be generated at http://www.tmdsonline.org. See Table S3 and Table S4 for numerical values.

Schizosaccharomyces pombe found the subcellular distributions of orthologs to be very similar (Matsuyama et al., 2006). The inclusion of orthologs significantly expanded our datasets, but this would be of little value if the proteins were very similar to the reference sequence. Thus the proteins from each organelle set were redundancy reduced by using BLASTClust to cluster them based on sequence similarity in their TMD and flanking sequences, and then we removed any with greater than 30% identity over this region (Altschul et al., 1997). Figure 1B summarizes the strategy used, and the numbers of proteins used for the analysis are provided in Figure 1C.

Alignment of TMDs Based on Their Cytosolic Ends

To compare the TMDs from different organelles, their sequences were aligned using the cytosolic ends of their hydrophobic cores. Initially, TMDs were located in the reference proteins using the TMHMM prediction algorithm (Krogh et al., 2001), and the orthologs were then aligned with the reference protein in order to assign their TMD positions. There is no established computational method for defining the ends of the part of a protein that spans the bilayer. Thus we implemented a scanning algorithm, which uses a sliding window and a threshold based on hydrophobicity. For this and subsequent analyses we used the hydrophobicity scale of Goldman, Engelman, and Steitz (GES) as it is designed for single-pass transmembrane helices and outperforms other scales in TMD prediction (Engelman et al., 1986; Koehler et al., 2009). However, to ensure that our findings were not dependent on this choice we also performed parallel analyses with the Wimley-White scale and the recently reported Biological scale from Hessa and coworkers, which is based on a completely different method (Hessa et al., 2007; Wimley and White, 1996). There is of course some flexibility in how charged residues are positioned at a bilayer interface, but by applying the same objective method to all organelles we should avoid bias in how TMD ends are assigned for the different datasets.

The scanning algorithm enabled us to align proteins from an organelle set at the position where a sharp change in hydropathy occurred, and the cytosolic end of the hydrophobic region was defined as position one. For all our analyses the hydrophobic spans were aligned with respect to their bilayer orientation, i.e. from the cytosolic side to the exoplasmic side (Figure 1A), rather than from N terminus to C terminus. We wanted to determine if residue preferences were influenced by position in the bilayer, which would be missed if all proteins (type I/III and type II) were simply analyzed from N to C terminus. In addition, the "positive-inside rule" indicates that the cytosolic flanking regions of TMDs are generally enriched in positively charged residues, thus allowing a clear definition of the cytosolic edges of hydrophobic spans (Nilsson et al., 2005).

TMDs from Different Organelles Exhibit Compositional Differences

Using the aligned sets of proteins, the frequency of each amino acid at each position through the hydrophobic region was calculated and plotted as matrices for fungi and vertebrates (Figures 2A and 2B, numerical values in Table S3 and Table S4). The residue preferences typically show a cluster of basic residues on the cytosolic side, followed as expected by the run of mostly aliphatic hydrophobic residues that spans the hydrophobic core of the bilayer. However, the matrices also reveal striking compositional differences between, and along, the TMDs. For both fungi and vertebrates, the regions enriched in hydrophobic residues are shorter for the ER and Golgi proteins than for plasma membrane proteins, indicating a difference in TMD length. In addition, the different hydrophobic residues were not uniformly

distributed through the hydrophobic TMD core. For example valine shows a clear enrichment in the exoplasmic side of the plasma membrane set in both vertebrates and fungi (Figure 2). To determine the extent and significance of such trends, we analyzed in more detail the changes in residue property and type through the bilayer.

Hydrophobic Lengths of TMDs Differ along the Secretory Pathway in Fungi and Vertebrates

To quantify trends in hydropathy, the mean hydrophobicity over all the sequences in each dataset was plotted relative to residue position. As noted above, the hydropathy plots for the fungal proteins from the early Golgi and late Golgi were found to be very similar, and so the datasets were combined to form a "Golgi" set (Figure S1 available online). For both fungi and vertebrates, the plasma membrane TMDs were on average hydrophobic for a greater length than those of the ER and Golgi (Figures 3A and 3B). For fungi the hydrophobicity values of the Golgi and plasma membrane TMDs were highly significantly different between positions 16 and 24 ($p < 1 \times 10^{-10}$ from two-sample independent t test, Figure S2A). For vertebrates, the difference between Golgi and plasma membrane TMDs was highly significant for positions 17 to 23 (Figure S2A).

To determine the prevalence of this difference in length within the datasets, we used the scanning algorithm described above to also define the exoplasmic ends of the TMDs and thus obtain a measure of the hydrophobic length for all of the TMDs. Distribution plots of TMD length show clearly distinguishable profiles for Golgi versus plasma membrane proteins (Figures 3C and 3D), with mean values that are highly significantly different (Figure 3E). As described above, the definition of TMD ends and the analyses of hydrophobic lengths were also performed using the Wimley-White and Biological hydrophobicity scales to avoid bias arising from using one particular hydrophobicity scale (Hessa et al., 2007; Wimley and White, 1996). The plots of TMD hydropathy and TMD length distribution determined using these differently derived scales show very similar trends to those obtained with the GES scale (Figures S2D–S2G). We also examined the distribution of TMD lengths predicted for the proteins by TMD prediction program Zpred2 (Papaloukas et al., 2008) and again found similar trends (Figure S2H).

Overall, plasma membrane TMDs have a greater hydrophobic length than those TMDs that span Golgi membranes, and this difference is conserved between fungi and vertebrates. However, the plots also reveal some differences between fungi and vertebrates. In fungi, the TGN/endosomal TMDs have a mean hydrophobic length intermediate between those of the Golgi and plasma membrane TMDs. However, in vertebrates, the TGN/endosomal TMDs appear to be of similar lengths to those of the plasma membrane. Indeed, the mean hydrophobicity values of the two sets of vertebrate proteins are not very significantly different in the region where the TMDs come to their end (residues 16–24), whereas for the fungal proteins there is a highly significant difference between the plasma membrane and TGN/endosomal TMDs at positions 21–24 ($p < 1 \times 10^{-10}$, Figure S2C).

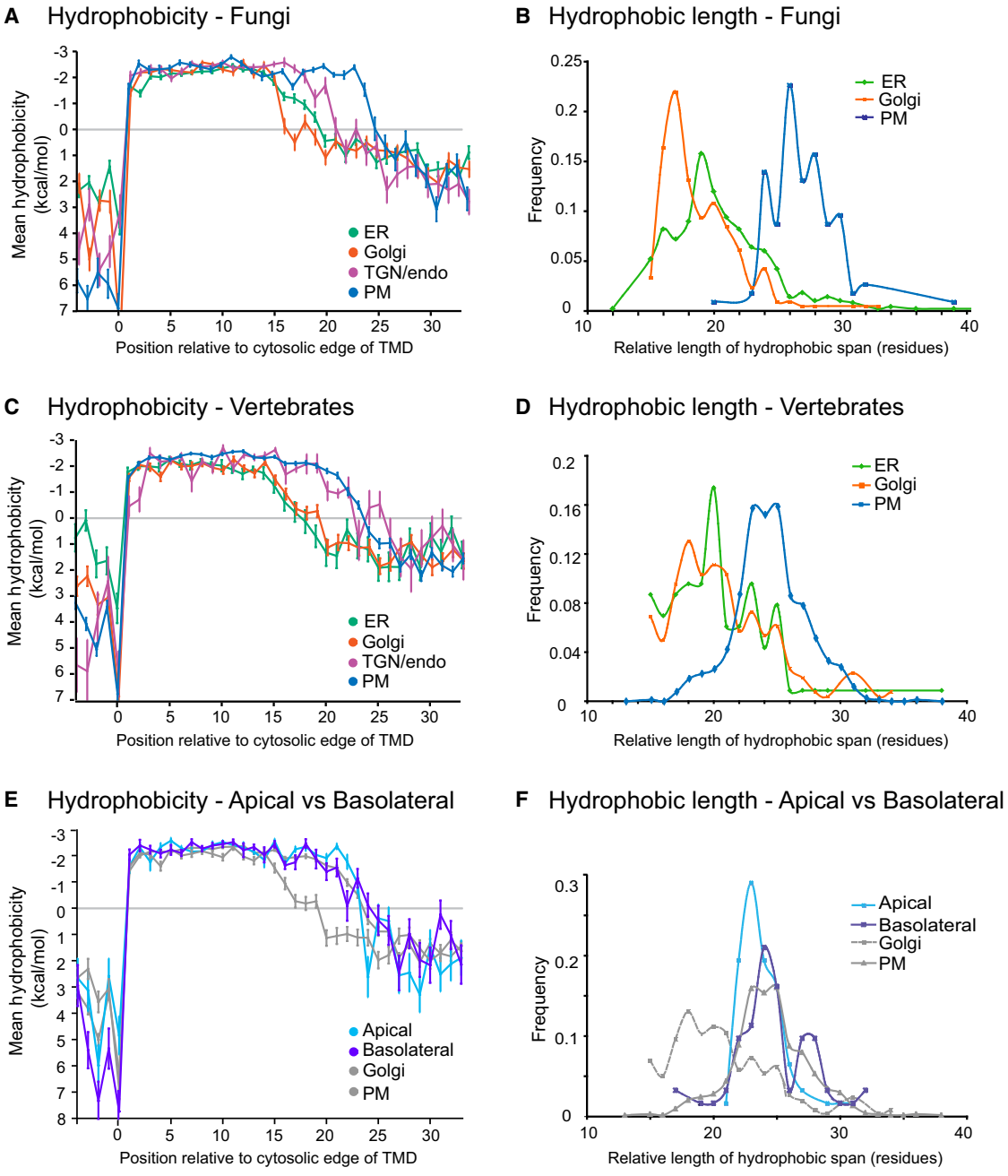
There have been suggestions that interactions with lipids could contribute to the sorting of membrane proteins to the apical surface in polarized epithelia (Simons and van Meer, 1988). We identified 15 apical and 12 basolateral reference proteins with a single TMD that were expanded to sets of 62 of each after adding orthologs and redundancy reduction to <30% identity (Table S2). However, the hydrophobic plots of the two sets are similar to each other and to the total plasma membrane set (Figure 3E), and the apical TMDs appear no longer than those of the basolateral surface (Figure 3F).

TMD Lengths Vary along the Secretory Pathway Irrespective of Type I versus Type II Topology

We noted that the Golgi datasets from both species groups consist only of proteins with a type II topology (N terminus in the cytosol). Conversely, the plasma membrane proteins from the fungal set all have a type I topology. It has been reported that topology has little influence on the sequence of TMDs in terms of partitioning out of the translocon (Lundin et al., 2008). Nonetheless, it seemed important to address the possibility that the trends observed here relate to differences in topology rather than location. Thus, we divided the organelle sets on the basis of topology. In fungi, there are type I proteins in the ER and TGN/endosomes sets in addition to the plasma membrane. The hydropathy plot in Figure S2I demonstrates that the TMDs of type I proteins from the plasma membrane are longer than those of type I proteins from the TGN/endosomes and ER sets, as observed for the combined topologies. For vertebrates, both type I and II proteins are present in the plasma membrane and ER datasets. The hydrophobicity plot in Figure S2J demonstrates that the plasma membrane proteins have longer average hydrophobic regions than ER proteins irrespective of whether type I or type II proteins are compared. Thus the trends we observe in hydrophobic length appear related to subcellular location rather than topology.

Hydrophobic Residues Are Distributed Asymmetrically in Plasma Membrane and Golgi TMDs

Although the hydrophobic cores of the TMDs from the various organelles differ in length, they all have similar hydropathy values that do not vary greatly along the length of this core. However, the residue frequency plots above suggest that the abundance of individual hydrophobic residues changes along the length of the TMDs (Figure 2). The residues valine, glycine, and leucine are uniformly distributed through the fungal Golgi TMDs, but all are asymmetrically distributed in plasma membrane TMDs, with valine and glycine being favored in more exoplasmic positions, whereas leucine shows the opposite trend (Figures 4A and 4B). To quantify further the degree of residue asymmetry, the relative lengths of each TMD in an organelle set were calculated as above and used to define the halves of the TMD corresponding to the inner and outer leaflets of the membrane. The abundance of each amino acid in the "inner" leaflet was subtracted from the abundance in the "outer" leaflet and divided by the total abundance to give a ratio for the leaflet preference. The mean ratios for each hydrophobic residue in each organelle set are shown in Figures 4C (fungi) and 4D (vertebrates). The values are plotted against the volume of each amino acid residue (Pontius et al., 1996). For fungal plasma membrane proteins, the overall trend is for the outer leaflet half



G Mean hydrophobic lengths (residues)

	ER	Golgi	TGN/endo	PM	apical	basolat.
Fungi	20.6	19.2	23.6	27.0		
Vertebrates	20.3	20.6	24.8	24.4	24.0	24.8

Figure 3. Positional Analysis of TMD Hydropathy from Different Organelles in Fungi and Vertebrates

(A) The mean hydrophobicity (GES scale) of the residues at each position along the aligned TMDs relative to the cytosolic edge was plotted for the indicated protein sets from fungi. The hydrophobicity values represent the free energy for partitioning from water into a hydrophobic environment, and therefore negative values indicate a preference for the interior of a lipid bilayer. Bars show standard error of mean.

(B) The distribution of TMD lengths for fungal organelles. The exoplasmic ends of the TMD were defined using the hydrophobicity scanning algorithm as for the cytosolic ends.

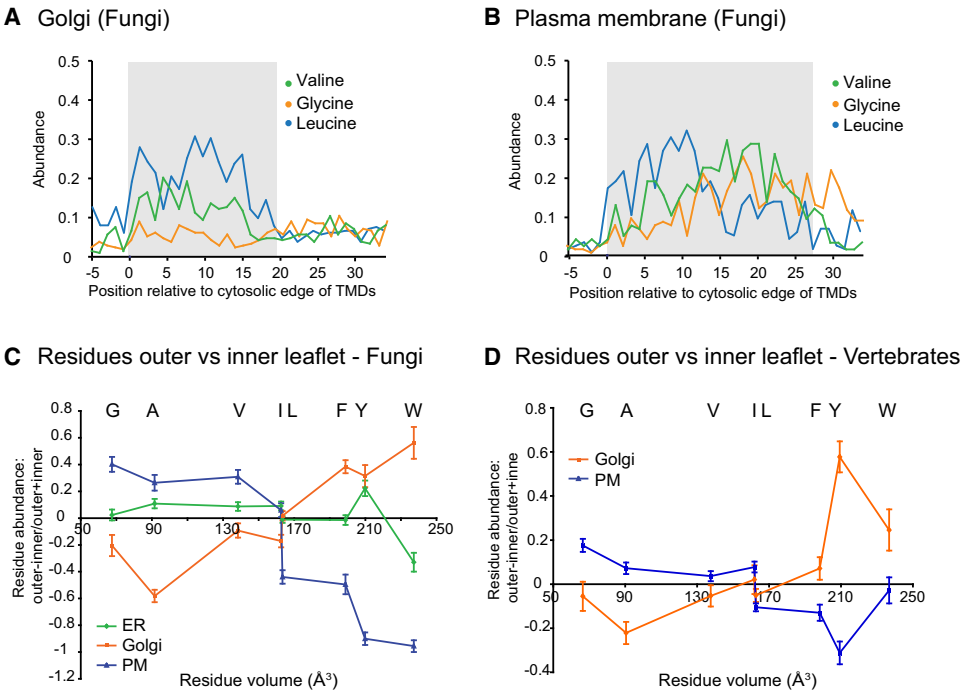


Figure 4. Analysis of the Compositional Asymmetry of TMDs from Different Organelles of Fungi and Vertebrates

(A and B) Analysis of the abundance of valine, glycine, and leucine along the TMDs from Golgi and plasma membrane proteins of fungi. Shaded areas represent the mean length of the hydrophobic regions for each protein set (Figure 3G).

(C and D) Analysis of amino acid asymmetry in ER, Golgi, and plasma membrane (PM) TMDs from fungi and in Golgi and plasma membrane TMDs from vertebrates. The abundance of each residue in the "inner" leaflet was subtracted from the abundance in the "outer" leaflet and divided by the total abundance to give a ratio of leaflet preference (0 = no preference). Leaflet position was defined by dividing the mean hydrophobic length for each organelle into two equal parts, and values for the different residues are plotted along the x axis according to residue volume. Error bars represent the standard error of the mean.

of the TMD to have an increase in smaller residues and decrease in larger residues, with the opposite trend for the Golgi proteins, whereas ER TMDs show no difference in relative abundance of hydrophobic residues between the leaflets. For vertebrates, a comparison of Golgi and plasma membrane asymmetry shows a similar trend to that seen in fungi, albeit smaller in scale. Overall, these results suggest that the constraints on amino acid composition of TMDs differ between the two leaflets of the bilayer.

TMD Compositions Appear Constrained by Residue Volume

The asymmetry described above appears to correlate to residue volume, and so we calculated the mean residue volume at positions along the TMDs. Figure 5A shows that Golgi and plasma membrane TMDs from fungi have similar profiles of residue volume in the halves of their TMD closest to the cytosol (positions 1–11 from the cytosolic side). However, in the exoplasmic portion of the TMDs, there is a bifurcation after residue 11; the plasma membrane TMDs have smaller mean volumes and the Golgi larger ones, with these differences being seen over most of the exoplasmic part of the TMDs. A similar trend is seen for the vertebrate proteins, with mean residue volumes similar in the cytosolic half and then splitting after position 12, with the larger amino acids for the Golgi TMDs and smaller amino acids in the plasma membrane TMDs (Figure 5B). The differences in amino acid volume are highly statistically significant for positions 12–19 in fungi and positions 14–19 in vertebrates ($p < 1 \times 10^{-10}$ for both, Figure 5C). In addition, very similar differences are observed if the TMD ends are defined with the Biological scale instead of the GES scale (Figure 5D). This indicates that there is an increase in average residue volume in the outer leaflet portion of TMDs from Golgi proteins and a reduction in volume for this part of plasma membrane TMDs.

(C and D) As for (A) and (B), but for vertebrate proteins.

(E and F) As for (C) and (D), but for vertebrate proteins of the apical and basolateral domains of the plasma membrane. The Golgi and total plasma membrane plots from (C) and (D) are included for comparison. The 15 apical and 12 basolateral reference proteins are listed in Table S2.

(G) The mean values for the TMD hydrophobic lengths of the indicated organelles shown in (B) and (D). For fungi, the differences between Golgi and TGN, Golgi and plasma membrane (PM), and TGN and PM are statistically significant ($p < 10^{-12}$, two sample t tests), whereas for vertebrates this was the case for Golgi and TGN and Golgi and PM ($p < 10^{-10}$) but not TGN and PM.

See also Figure S2 for tests of robustness and significance of data.

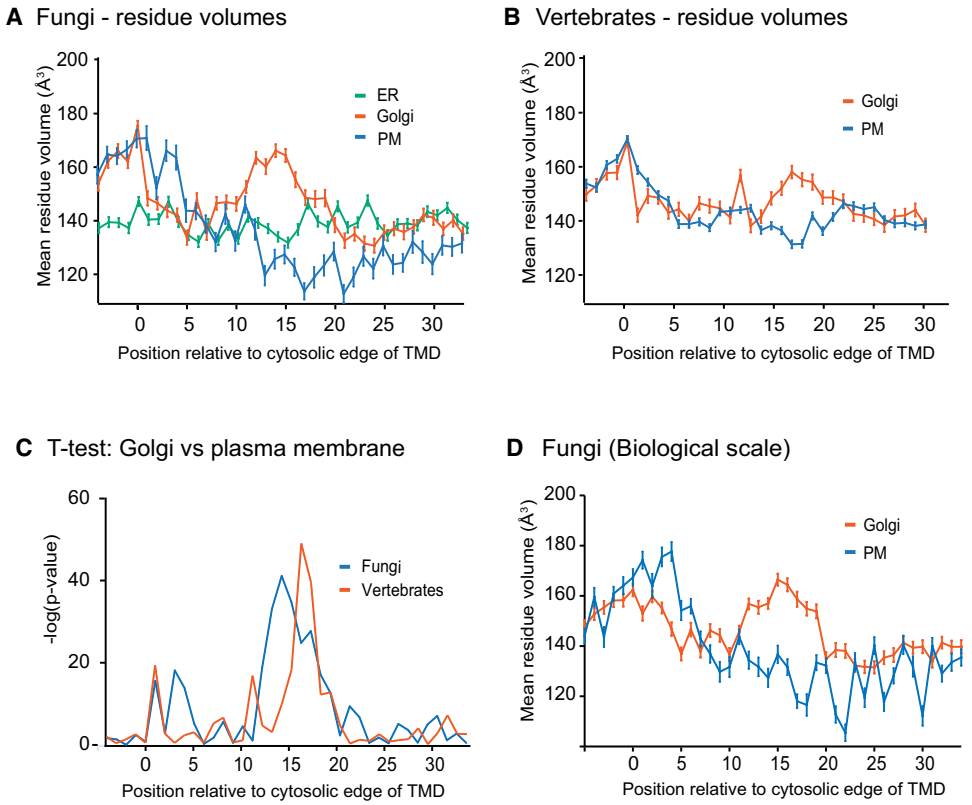


Figure 5. Positional Analysis of Amino Acid Volume from Different Organelles in Fungi and Vertebrates

(A and B) The mean values for residue volume (Pontius et al., 1996), at each position along the TMDs from fungi and vertebrates. Error bars indicate standard error of the mean.

(C) Independent (two sample) t tests were applied at positions along the TMDs to assess the significance of differences between the mean values of amino acid volumes for Golgi and plasma membrane proteins shown in (A) and (B).

(D) The Biological scale of Hessa and coworkers was used to define cytosolic TMD edges and thus align the TMDs from different organelles at their cytosolic ends (Hessa et al., 2005). This alignment was then used for analysis of amino acid volume along the fungal Golgi and plasma membrane TMDs. Error bars indicate standard error of the mean.

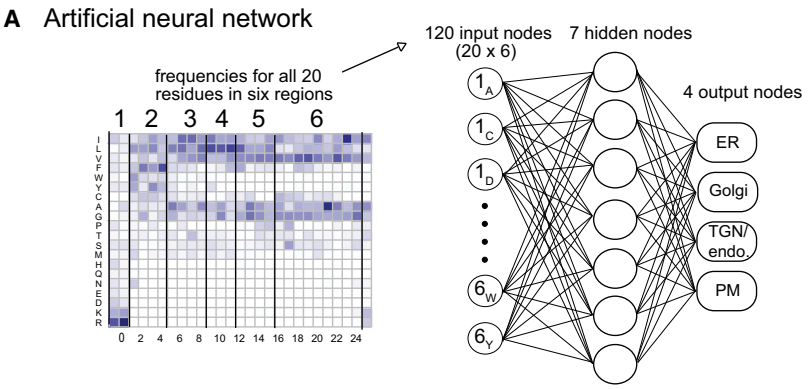
See also Figure S3.

Plasma Membrane TMDs Do Not Display a "Size Moment"

One possible explanation for the reduction in average residue volume in the exoplasmic side of the plasma membrane TMDs is that we were detecting a relative enrichment of GXXXG-like oligomerization motifs. This motif aligns two glycines or other small residues on one face of the helix, and these allow the TMDs to pack closely and dimerize via their backbones (Russ and Engelman, 2000). In order to test if this was the case, we quantified the helical size bias of the TMDs in the different datasets. Residue volume was treated as a vector from the helix, and the values summed for two turns (i.e., seven successive residues) to give a "size moment." If one side of the helix is flattened, i.e., has more small residues than the opposing side, then the size moment will be higher over that region. Glycophorin A has a GXXXG motif within its TMD and was used as a positive control (Russ and Engelman, 2000). The plasma membrane TMD sets do not show a large peak in size moment in their exoplasmic positions such as that seen for glycophorin A (Figures S3A and S3B). This implies that the exoplasmic parts of the plasma membrane datasets are not substantially enriched in flat dimerization motifs and indicates that the increased proportion of smaller residues instead reflects a difference in overall residue composition all round the transmembrane helix.

An Artificial Neural Network Can Classify Subcellular Location Based on TMD Sequence

To evaluate further the scale of organelle-specific heterogeneity among TMDs, we tested whether the differences are sufficiently great to have predictive value. An artificial neural network was implemented with the aim of classifying proteins into organelles using the amino acid composition of delineated regions through the TMDs (Figure 6A). To avoid bias arising from differences in dataset size, proteins were randomly removed from the redundancy-reduced sets of fungal proteins such that each organelle was represented by the same number of proteins (n = 99). The neural network was then trained on these fungal proteins from the ER, Golgi, TGN/endosomes, and plasma membrane. Networks were tested by cross-validation in which groups of proteins were removed from the training set and then used to



B ANN - fivefold cross validation

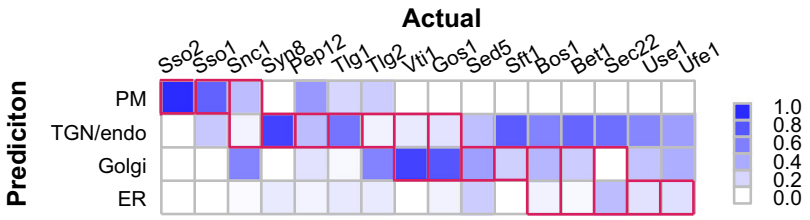
Actual \ Prediction	ER	Golgi	TGN/ endo	PM
ER	0.70	0.10	0.09	0.11
Golgi	0.11	0.82	0.05	0.01
TGN/ endo	0.12	0.03	0.71	0.14
PM	0.06	0.06	0.06	0.81

Mean accuracy: 76.1%

C Localisation predictors compared

Organelle	Prediction accuracy ANN	Euk-mPloc	WoLF PSORT	SherLoc
ER	0.70	0.39	0.06	0.30
Golgi + TGN/endo	0.77	0.30	0.07	0.54
PM	0.82	0.47	0.11	0.05

D SNARE prediction - normal topology



E SNARE prediction - reverse topology

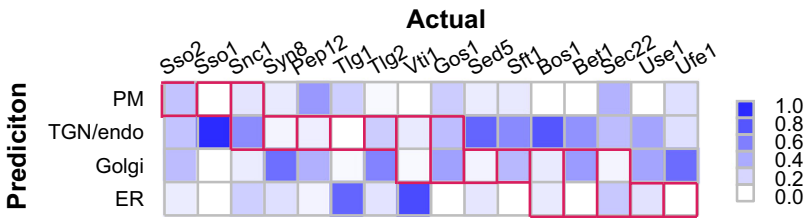


Figure 6. An Artificial Neural Network Classifier of Subcellular Location Based on TMD Sequence

(A) Overview of the neural network used for classifying proteins. The compositions of six regions along the TMDs from each fungal organelle set were encoded into input vectors to train the network.

(B) Test of the accuracy of the ability of the neural network to predict localization. Performance was assessed using a 5-fold "leave-one-out" cross-validation in which groups of proteins were removed from the training set and then used to test the network trained with the remaining proteins. The predicted location was that with the highest score, with a mean accuracy calculated over all proteins in each set.

(C) A comparison of predictive accuracy of the network (ANN) to that of existing subcellular localization prediction methods when applied to the *S. cerevisiae* reference proteins.

(D and E) Prediction of SNARE localization using the neural network trained on TMD regions. The SNAREs from *S. cerevisiae* and 36 other fungi were examined with the network trained on the datasets that do not include the SNAREs, and the frequencies of predictions were normalized and plotted in a matrix against subcellular locations. Red boxes indicate the experimentally determined localizations of the SNAREs. SNARE TMD sequences were reversed prior to analysis in (E).

test the network trained with the remaining proteins. Using a 5-fold cross-validation, the network correctly classified over 70% of proteins from ER, early Golgi, TGN/endosomes, and plasma membrane (Figure 6B).

We compared this outcome to that obtained with widely used algorithms for predicting subcellular localization. The most recent methods for location prediction are based on a combination of text-mining and ab initio sequence-based methods. We thus challenged three of the major location predictors (WoLF PSORT, SherLoc, Euk-mPLoc) with the complete sequences of the *S. cerevisiae* proteins from our datasets (Chou and Shen, 2007; Horton et al., 2007; Shatkay et al., 2007). The neural network, using only the sequence of the TMDs, outperformed all three predictors using the complete protein sequences when classifying bitopic proteins between ER, Golgi, and plasma membrane (Figure 6C). This suggests that incorporation of analysis of TMD sequences could improve the accuracy of current methods for predicting subcellular localization.

SNARE TMDs Exhibit Organelle-Specific Trends in Composition

To seek further evidence for organelle-specific constraints on TMDs, we used the neural network to examine the proteins of SNARE family that mediate fusion between vesicles and organelles (Jahn and Scheller, 2006). Most SNAREs have a single TMD, and these all have the same type II topology. Individual SNAREs all perform the same general fusogenic role but are located to different organelles within the exocytic and endocytic pathways. They were not included in the datasets analyzed above, and so we tested whether the neural network could detect organelle-specific

differences in the TMDs of the SNARE proteins. Predictions were performed on the fungal orthologs of all of the SNAREs from *S. cerevisiae* (Figure 6D). Overall, the outcome was far from random with SNAREs from the early secretory pathway predicted to be ER or Golgi, and plasma membrane and endosomal SNAREs predicted to be TGN/endosomes or plasma membrane. The accuracy of prediction is less than that obtained with the datasets examined above (50% correct rather than >70%), which may reflect the multitude of factors involved in the recycling and localization of SNAREs, and the TMDs potentially having a role in SNARE function (Stein et al., 2009). However, when the sequences of the SNAREs were reversed, and hence the orientation of their TMDs with respect to the bilayer, there was no particular trend or accuracy in the prediction (22% correct, Figure 6E). Thus, despite the SNAREs all sharing a common general function, there are constraints imposed on the sequences of SNARE TMDs that are shared with the TMDs of unrelated proteins from the same organelle, and the asymmetry of these constraints is a major feature detected by the neural network.

DISCUSSION

The analysis described here is, to the best of our knowledge, the first report of a comprehensive comparison of TMDs from all the major compartments of the eukaryotic secretory pathway. We find overwhelming evidence that there is not a "generic" type of TMD shared by eukaryotic membrane proteins. There are, of course, protein-specific constraints on TMD sequences imposed by the interactions and function of a particular protein. However, it appears that TMDs also vary depending on their organelle of residence in both length and composition. The structural consequences of these compositional differences can be illustrated by modeling the "consensus" TMDs for the fungal Golgi and plasma membrane (Figure 7A). These organelle-specific trends have obvious implications for improving the prediction of TMD presence and topology, as TMD features recognized by prediction algorithms will, in part, reflect the localization of the membrane proteins used to train the algorithm.

Our observations also have implications for how and why the different bilayers of the cell vary in their physical properties. The TMDs from the plasma membrane proteins of both fungi and vertebrates are longer than those from the proteins of internal membranes, even though the two sets of plasma membrane proteins are otherwise unrelated by sequence or function. This length difference was suggested by previous analyses of much smaller datasets from the plasma membrane and the Golgi (Bretscher and Munro, 1993; Levine et al., 2000) but is unequivocally validated by these much larger datasets. In addition, the analysis has now been extended to all of the secretory pathways of both vertebrates and fungi, revealing that TMD lengths are similarly short in both the ER and Golgi and then increase in compartments beyond the Golgi stack. This difference could reflect a shared tendency for post-Golgi TMDs to tilt in the bilayer of their organelle of residence, but this seems highly implausible, especially as the increased levels of order-inducing lipids in post-Golgi membranes would be expected to discourage tilting (see below). Thus the simplest explanation of the difference in TMD length is that for both fungi and vertebrates the plasma membrane is thicker than the membranes of the ER and Golgi. The length of an α helix increases by 1.5 Å per residue, and so these differences in TMD length would equate to an increase in bilayer thickness of ~12 Å (42%) from Golgi to plasma membrane in fungi and ~6 Å for vertebrates.

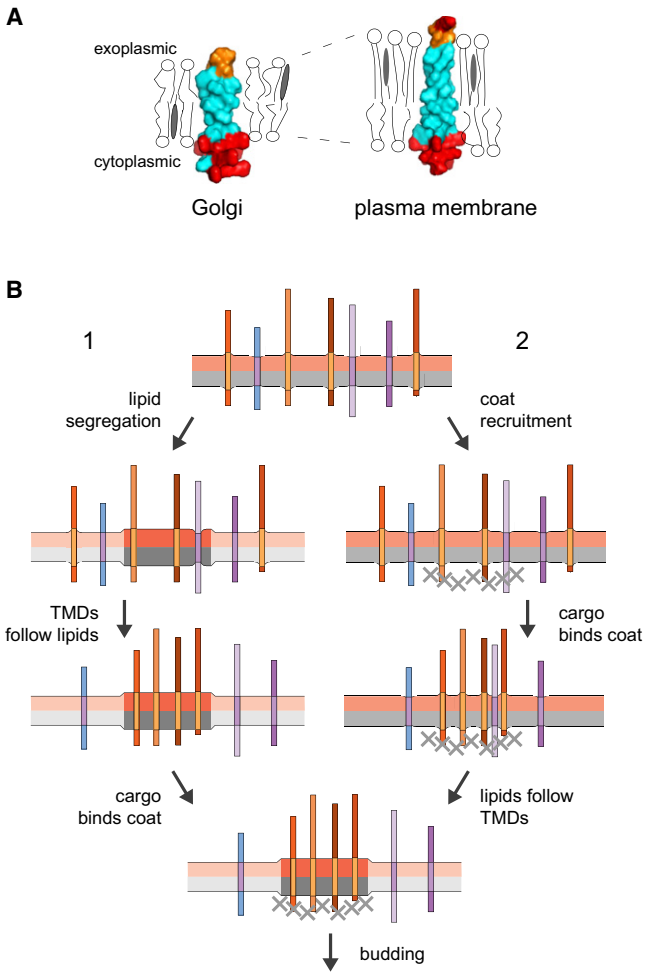


Figure 7. Organelle-Specific TMDs and Their Relationship with the Lipid Bilayer

(A) Consensus TMDs from the fungal Golgi and plasma membrane datasets based on the most abundant residue at each position. Residues were modeled on an α helix using PyMOL. Hydrophobic residues (AGILVFWY) are colored cyan, polar residues (HNQST) orange, and basic residues (KR) red. The representation of the bilayer assumes that the plasma membrane is thicker and has a higher content of saturated lipids and sterols in the outer leaflet than do Golgi membranes.

(B) Sorting of proteins sharing distinct TMD properties could either be driven by lipid sorting or could drive lipid sorting. For example, a domain of thicker, more ordered lipids could attract proteins with longer TMDs, and these could then attract a coat (1). Alternatively, if the cargo proteins for a particular class of vesicle have longer TMDs than the resident proteins, then their collection by coat into a forming transport carrier could affect the lipid composition around them, which would sort lipids and exclude residents with shorter TMDs (2). Either system could alternatively act on short TMDs if they were collected into vesicle by coats or segregated into thinner domains.

See also Figure S4.

Although the trend for longer TMDs in post-Golgi compartments is broadly similar in fungi and vertebrates, there also appear to be some differences. The TMD lengths imply that the fungal plasma membrane is even thicker than that of vertebrates, and also the membranes of the TGN/endosomal system are similar in thickness to the plasma membrane in vertebrates, but in fungi their thickness is intermediate between those of the Golgi and plasma membrane. The TGN/endosomal route is followed by proteins taken in from the plasma membrane or traveling from the Golgi to the vacuole or lysosome (Bonifacino and Traub, 2003; Bowers and Stevens, 2005). We did not include these lytic compartments in the analysis above because only a few bitopic proteins are known for each. However, when the methods used above are applied to these small datasets, the vertebrate lysosomal proteins appear similar to plasma membrane proteins, with longer TMDs and a preference for smaller residues in the exoplasmic half of the bilayer (Figure S4). In contrast, the fungal vacuolar proteins have shorter TMD lengths and an increased abundance of bulky aromatic residues compared to lysosomal TMDs (Figure S4). These differences cannot be viewed as definitive given the small numbers of reference proteins, but they are at least consistent with all post-Golgi membranes in vertebrates being equally thickened compared to the Golgi and ER, whereas in fungi the plasma membrane is particularly thick and the other post-Golgi membranes are intermediate in thickness compared to the Golgi.

The thickness of a fluid lipid bilayer has been shown to depend on acyl chain length and the presence of lipids such as sterols or sphingolipids (Brown and London, 1998; Lewis and Engelman, 1983). Sterols are rigid and sphingolipids have saturated acyl chains, and so both increase acyl chain order and thus thicken the bilayer and reduce permeability to solutes. The plasma membranes of fungi and mammals are enriched in sterols and sphingolipids compared to the ER and Golgi, which would be consistent with an increase in bilayer thickness (Holthuis et al., 2001). Sphingolipids are synthesized in the exoplasmic leaflet of the trans-Golgi from where they move, via mechanisms that are not understood, up a concentration gradient into post-Golgi compartments (Holthuis and Levine, 2005; Klemm et al., 2009; Tafesse et al., 2006; van Meer, 1989). The vacuole and endosomes of *S. cerevisiae* have relatively low levels of sterols and sphingolipids compared to the fungal plasma membrane or vertebrate lysosomes, which would be consistent with the apparent differences in the bilayer thickness between these organelles (Klemm et al., 2009; Schneiter et al., 1999).

In contrast, when we compared the TMDs of proteins that reside in the apical or basolateral domains of epithelial cells, we did not find a clear difference in hydrophobic length or trends in residue volume (Figure 3 and data not shown). There have been suggestions that TMD:lipid interactions could contribute to sorting of proteins to the apical surface (Simons and van Meer, 1988), but we are not aware of any previous report of a comparison of the TMDs from the two sets of proteins. The lack of apparent difference in TMD length may reflect the relatively small number of reference proteins, and indeed Mitra and coworkers have used low-angle X-ray scattering to measure the thickness of membranes of polarized hepatocytes and reported that the apical membrane was 3–5 Å thicker than the Golgi and ER, but the basolateral membrane was, if anything, thinner (Mitra et al., 2004). However, it should be noted that although X-ray scattering is an interesting approach, the method requires that organelles are isolated from cells, treated with carbonate to rupture them, and then treated for several hours with protease. This could perturb aspects of the bilayers and so may not have provided a definitive measure of in vivo properties. Moreover, the protocol used to isolate basolateral membranes removes apical membranes but not all others, with inner mitochondrial membranes alone constituting 22% of the basolateral fraction (Meier et al., 1984). It should also be noted that whereas glycolipids are ~2-fold more concentrated on the apical surface of many epithelia, the other order-inducing lipids cholesterol and sphingomyelin can be equally distributed, and sphingomyelin even concentrated at the basolateral surface in some cell types (Brasitus and Schachter, 1980; Simons and van Meer, 1988; van IJzendoorn et al., 1997). Further work is clearly needed to understand the different properties of the apical and basolateral surfaces, but at present it seems possible that the major difference in bilayer thickness in epithelial cells could occur between pre- and post-Golgi compartments rather than between apical and basolateral domains.

In addition to variations in TMD length, we also found an asymmetry in the distribution of residue volume in the plasma membrane TMDs. Extrapolating from studies of bilayer permeability, small and more compact side chains would be expected to fit better into a bilayer that has well-ordered lipid acyl chains (Mathai et al., 2008; Mitragotri et al., 1999). This implies that there is an asymmetry in the state of lipid order in the plasma membrane. Such an asymmetry is more easily accounted for by lipids such as sterols and sphingolipids, which are enriched in one leaflet, than by proteins that span both leaflets. This suggests that lipids contribute, at least in part, to differences in bilayer order between organelles or subdomains. Indeed TMD asymmetry may explain why plasma membrane proteins show a surprising exclusion from "plasma membrane-like" lipid domains in liposomes (Bacia et al., 2004), as liposomes are symmetric and so the residues of the TMD adapted to the cytosolic leaflet would be exposed to a lipid organization that is only experienced in vivo by the outer leaflet residues.

The results of our analysis strongly imply that the different bilayers of eukaryotic cells have different physical properties, and these differences seem likely to be, at least in part, imposed by differences in lipid composition. Changes in membrane properties would provide an indication of location that could be used to control the activity of proteins such as channels and transporters as they move through the secretory pathway. However, a striking aspect of the data is how pervasive the differences between TMDs are in the large datasets that we have examined, implying that the TMDs of many of the proteins in a particular compartment share organelle-specific properties. This is perhaps clearest for TMD length in the case of fungi (Figure 3C), but even for vertebrates 92% of the plasma membrane TMDs are longer than the mean length for the Golgi and ER. Previous theoretical and experimental work has suggested that integral membrane proteins can influence the organization of the lipids that surround them (Andersen and Koeppe, 2007; Mitra et al., 2004; Mouritsen and Bloom, 1993). In addition, a quantitative analysis of the

composition of synaptic vesicles revealed that TMDs account for ~20% of the area of the membrane, indicating that most lipids are close to proteins, and this very high protein density is unlikely to be unique to this particular membrane (Frick et al., 2007; Takamori et al., 2006). If many of the proteins in the same compartment or forming vesicle share TMD shapes then they could contribute to bilayer properties, and in particular to thickness, if they are at a high enough concentration. Protein clustering in forming vesicles could thus cause local changes in bilayer physical properties, which could result in lipid sorting, especially at the late Golgi where sphingolipids are synthesized and a major transition in bilayer thickness seems to occur (Figure 7B). This means that the answer to the long-standing question of how cells sort lipids to different destinations could be that it is an emergent property of the traffic of membrane proteins that are at a high density and share organelle-specific TMD properties. This need not exclude the resulting protein/lipid microdomains attracting further cargo or excluding residents based on physical properties alone. Determining the relative contributions of proteins and lipids to each other's sorting is likely to be a key issue for future studies of the biogenesis of eukaryotic membranes.

Further work will be required to investigate these issues in detail, but irrespective of the outcome of such studies, our analysis clearly shows that eukaryotic TMDs are not a generic entity that is varied solely for protein-specific functions. Rather, TMD sequences are optimized for insertion, function, and also the variable and asymmetric physical properties of their bilayers of residence.

EXPERIMENTAL PROCEDURES

Full methods and associated references are in the Extended Experimental Procedures online. In summary, proteins with a single TMD from *S. cerevisiae* and *H. sapiens* were collated from databases. Those with a known location and topology were identified from the literature (Table S1 and Table S2), and their TMDs located with the prediction program TMHMM (Krogh et al., 2001). Orthologs from a further 36 fungi or 12 vertebrates were identified by BLAST searching of RefSeq genomes, and the TMDs in the orthologs identified by aligning them to the references sequences.

The cytosolic and exoplasmic edges of the TMDs were defined as the point at which the residue hydropathy in a small window sliding out from the middle of the TMD fell below a fixed threshold. For analysis of residue properties all the TMDs were aligned at their cytosolic edges. For type II proteins, residues were thus analyzed starting from the N-terminal end of their TMDs, and for type I and III from the C-terminal end. The resulting datasets were analyzed using custom software with a graphical user interface, and plots of residue properties or abundance can be generated at http://www.tmdsonline.org.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, four figures, and four tables and can be found with this article online at doi:10.1016/j.cell.2010.05.037.

ACKNOWLEDGMENTS

We would like to thank Arne Elofsson (Stockholm University) for help with Zpred2 and Madan Babu, Chantal Christis, Matthew Freeman, Alison Gillingham, Tobias Kloeper, Katja Röper, and Isabel Torres for comments on the manuscript. Funding was from the Medical Research Council (UK).

Received: February 19, 2010
Revised: April 15, 2010
Accepted: May 10, 2010
Published: July 8, 2010

REFERENCES

Almén, M.S., Nordström, K.J.V., Fredriksson, R., and Schiöth, H.B. (2009). Mapping the human membrane proteome: a majority of the human membrane proteins can be classified according to function and evolutionary origin. BMC Biol. *7*, 50.

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.

Andersen, O.S., and Koeppe, R.E. (2007). Bilayer thickness and membrane protein function: an energetic perspective. Annu. Rev. Biophys. Biomol. Struct. *36*, 107–130.

Bacia, K., Schuette, C.G., Kahya, N., Jahn, R., and Schwille, P. (2004). SNAREs prefer liquid-disordered over "raft" (liquid-ordered) domains when reconstituted into giant unilamellar vesicles. J. Biol. Chem. *279*, 37951–37955.

Bondar, A., del Val, C., and White, S.H. (2009). Rhomboid protease dynamics and lipid interactions. Structure *17*, 395–405.

Bonifacino, J.S., and Traub, L.M. (2003). Signals for sorting of transmembrane proteins to endosomes and lysosomes. Annu. Rev. Biochem. *72*, 395–447.

Bowers, K., and Stevens, T.H. (2005). Protein transport from the late Golgi to the vacuole in the yeast Saccharomyces cerevisiae. Biochim. Biophys. Acta *1744*, 438–454.

Brasitus, T.A., and Schachter, D. (1980). Lipid dynamics and lipid-protein interactions in rat enterocyte basolateral and microvillus membranes. Biochemistry *19*, 2763–2769.

Bretscher, M.S., and Munro, S. (1993). Cholesterol and the Golgi apparatus. Science *261*, 1280–1281.

Brown, D.A., and London, E. (1998). Structure and origin of ordered lipid domains in biological membranes. J. Membr. Biol. *164*, 103–114.

Chou, K.-C., and Shen, H.-B. (2007). Euk-mPLoc: a fusion classifier for large-scale eukaryotic protein subcellular location prediction by incorporating multiple sites. J. Proteome Res. *6*, 1728–1734.

Dukhovny, A., Yaffe, Y., Shepshelovitch, J., and Hirschberg, K. (2009). The length of cargo-protein transmembrane segments drives secretory transport by facilitating cargo concentration in export domains. J. Cell Sci. *122*, 1759–1767.

Engelman, D.M., Steitz, T.A., and Goldman, A. (1986). Identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins. Annu. Rev. Biophys. Biophys. Chem. *15*, 321–353.

Frick, M., Schmidt, K., and Nichols, B.J. (2007). Modulation of lateral diffusion in the plasma membrane by protein density. Curr. Biol. *17*, 462–467.

Hessa, T., Kim, H., Bihlmaier, K., Lundin, C., Boekel, J., Andersson, H., Nilsson, I., White, S.H., and von Heijne, G. (2005). Recognition of transmembrane helices by the endoplasmic reticulum translocon. Nature *433*, 377–381.

Hessa, T., Meindl-Beinker, N.M., Bernsel, A., Kim, H., Sato, Y., Lerch-Bader, M., Nilsson, I., White, S.H., and von Heijne, G. (2007). Molecular code for transmembrane-helix recognition by the Sec61 translocon. Nature *450*, 1026–1030.

Holthuis, J.C., and Levine, T.P. (2005). Lipid traffic: floppy drives and a superhighway. Nat. Rev. Mol. Cell Biol. *6*, 209–220.

Holthuis, J.C., Pomorski, T., Raggers, R.J., Sprong, H., and Van Meer, G. (2001). The organizing potential of sphingolipids in intracellular membrane transport. Physiol. Rev. *81*, 1689–1723.

Horton, P., Park, K.-J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C.J., and Nakai, K. (2007). WoLF PSORT: protein localization predictor. Nucleic Acids Res. *35*, W585–W587.

Jahn, R., and Scheller, R.H. (2006). SNAREs - engines for membrane fusion. Nat. Rev. Mol. Cell Biol. *7*, 631–643.

Killian, J.A., and von Heijne, G. (2000). How proteins adapt to a membrane-water interface. Trends Biochem. Sci. *25*, 429–434.

Klemm, R.W., Ejsing, C.S., Surma, M.A., Kaiser, H.J., Gerl, M.J., Sampaio, J.L., de Robillard, Q., Ferguson, C., Proszynski, T.J., Shevchenko, A., et al. (2009). Segregation of sphingolipids and sterols during formation of secretory vesicles at the trans-Golgi network. J. Cell Biol. *185*, 601–612.

Koehler, J., Woetzel, N., Staritzbichler, R., Sanders, C.R., and Meiler, J. (2009). A unified hydrophobicity scale for multispan membrane proteins. Proteins *76*, 13–29.

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

Levine, T.P., Wiggins, C.A., and Munro, S. (2000). Inositol phosphorylceramide synthase is located in the Golgi apparatus of *Saccharomyces cerevisiae*. Mol. Biol. Cell *11*, 2267–2281.

Lewis, B.A., and Engelman, D.M. (1983). Lipid bilayer thickness varies linearly with acyl chain length in fluid phosphatidylcholine vesicles. J. Mol. Biol. *166*, 211–217.

Lundin, C., Kim, H., Nilsson, I., White, S.H., and von Heijne, G. (2008). Molecular code for protein insertion in the endoplasmic reticulum membrane is similar for N(in)-C(out) and N(out)-C(in) transmembrane helices. Proc. Natl. Acad. Sci. USA *105*, 15702–15707.

Mathai, J.C., Tristram-Nagle, S., Nagle, J.F., and Zeidel, M.L. (2008). Structural determinants of water permeability through the lipid membrane. J. Gen. Physiol. *131*, 69–76.

Matsuyama, A., Arai, R., Yashiroda, Y., Shirai, A., Kamata, A., Sekido, S., Kobayashi, Y., Hashimoto, A., Hamamoto, M., Hiraoka, Y., et al. (2006). ORFeome cloning and global analysis of protein localization in the fission yeast Schizosaccharomyces pombe. Nat. Biotechnol. *24*, 841–847.

Meier, P.J., Sztul, E.S., Reuben, A., and Boyer, J.L. (1984). Structural and functional polarity of canalicular and basolateral plasma membrane vesicles isolated in high yield from rat liver. J. Cell Biol. *98*, 991–1000.

Mitra, K., Ubarretxena-Belandia, I., Taguchi, T., Warren, G., and Engelman, D.M. (2004). Modulation of the bilayer thickness of exocytic pathway membranes by membrane proteins rather than cholesterol. Proc. Natl. Acad. Sci. USA *101*, 4083–4088.

Mitragotri, S., Johnson, M.E., Blankschtein, D., and Langer, R. (1999). An analysis of the size selectivity of solute partitioning, diffusion, and permeation across lipid bilayers. Biophys. J. *77*, 1268–1283.

Mouritsen, O.G., and Bloom, M. (1993). Models of lipid-protein interactions in membranes. Annu. Rev. Biophys. Biomol. Struct. *22*, 145–171.

Nilsson, J., Persson, B., and von Heijne, G. (2005). Comparative analysis of amino acid distributions in integral membrane proteins from 107 genomes. Proteins *60*, 606–616.

Nyholm, T.K., Ozdirekcan, S., and Killian, J.A. (2007). How protein transmembrane segments sense the lipid environment. Biochemistry *46*, 1457–1465.

Papaloukas, C., Granseth, E., Viklund, H., and Elofsson, A. (2008). Estimating the length of transmembrane helices using Z-coordinate predictions. Protein Sci. *17*, 271–278.

Patterson, G.H., Hirschberg, K., Polishchuk, R.S., Gerlich, D., Phair, R.D., and Lippincott-Schwartz, J. (2008). Transport through the Golgi apparatus by rapid partitioning within a two-phase membrane system. Cell *133*, 1055–1067.

Paulson, J.C., and Colley, K.J. (1989). Glycosyltransferases. Structure, localization, and control of cell type-specific glycosylation. J. Biol. Chem. *264*, 17615–17618.

Phillips, R., Ursell, T., Wiggins, P., and Sens, P. (2009). Emerging roles for lipids in shaping membrane-protein function. Nature *459*, 379–385.

Pontius, J., Richelle, J., and Wodak, S.J. (1996). Deviations from standard atomic volumes as a quality measure for protein crystal structures. J. Mol. Biol. *264*, 121–136.

Ronchi, P., Colombo, S., Francolini, M., and Borgese, N. (2008). Transmembrane domain-dependent partitioning of membrane proteins within the endoplasmic reticulum. J. Cell Biol. *181*, 105–118.

Russ, W.P., and Engelman, D.M. (2000). The GxxxG motif: a framework for transmembrane helix-helix association. J. Mol. Biol. *296*, 911–919.

Schneiter, R., Brügger, B., Sandhoff, R., Zellnig, G., Leber, A., Lampl, M., Athenstaedt, K., Hrastnik, C., Eder, S., Daum, G., et al. (1999). Electrospray ionization tandem mass spectrometry (ESI-MS/MS) analysis of the lipid molecular species composition of yeast subcellular membranes reveals acyl chain-based sorting/remodeling of distinct molecular species en route to the plasma membrane. J. Cell Biol. *146*, 741–754.

Shatkay, H., Höglund, A., Brady, S., Blum, T., Dönnes, P., and Kohlbacher, O. (2007). SherLoc: high-accuracy prediction of protein subcellular localization by integrating text and protein sequence data. Bioinformatics *23*, 1410–1417.

Simons, K., and Ikonen, E. (1997). Functional rafts in cell membranes. Nature *387*, 569–572.

Simons, K., and van Meer, G. (1988). Lipid sorting in epithelial cells. Biochemistry *27*, 6197–6202.

Stein, A., Weber, G., Wahl, M.C., and Jahn, R. (2009). Helical extension of the neuronal SNARE complex into the membrane. Nature *460*, 525–528.

Stevens, T.J., and Arkin, I.T. (2000). Do more complex organisms have a greater proportion of membrane proteins in their genomes? Proteins *39*, 417–420.

Tafesse, F.G., Ternes, P., and Holthuis, J.C. (2006). The multigenic sphingomyelin synthase family. J. Biol. Chem. *281*, 29421–29425.

Takamori, S., Holt, M., Stenius, K., Lemke, E.A., Grønborg, M., Riedel, D., Urlaub, H., Schenck, S., Brügger, B., Ringler, P., et al. (2006). Molecular anatomy of a trafficking organelle. Cell *127*, 831–846.

van IJzendoorn, S.C., Zegers, M.M., Kok, J.W., and Hoekstra, D. (1997). Segregation of glucosylceramide and sphingomyelin occurs in the apical to basolateral transcytotic route in HepG2 cells. J. Cell Biol. *137*, 347–357.

van Meer, G. (1989). Lipid traffic in animal cells. Annu. Rev. Cell Biol. *5*, 247–275.

van Meer, G., Voelker, D.R., and Feigenson, G.W. (2008). Membrane lipids: where they are and how they behave. Nat. Rev. Mol. Cell Biol. *9*, 112–124.

White, S.H., and Wimley, W.C. (1999). Membrane protein folding and stability: physical principles. Annu. Rev. Biophys. Biomol. Struct. *28*, 319–365.

Wimley, W.C., and White, S.H. (1996). Experimentally determined hydrophobicity scale for proteins at membrane interfaces. Nat. Struct. Biol. *3*, 842–848.