Eur Biophys J (2013) 42:199–207
DOI 10.1007/s00249-012-0813-9

ORIGINAL PAPER

EBSA
Biophysics in Europe

Structure-based statistical analysis of transmembrane helices

Carlos Baeza-Delgado · Marc A. Marti-Renom · Ismael Mingarro

Received: 16 January 2012 / Revised: 21 March 2012 / Accepted: 14 April 2012 / Published online: 16 May 2012

Abstract Recent advances in determination of the high-resolution structure of membrane proteins now enable analysis of the main features of amino acids in transmembrane (TM) segments in comparison with amino acids in water-soluble helices. In this work, we conducted a large-scale analysis of the prevalent locations of amino acids by using a data set of 170 structures of integral membrane proteins obtained from the MPtopo database and 930 structures of water-soluble helical proteins obtained from the protein data bank. Large hydrophobic amino acids (Leu, Val, Ile, and Phe) plus Gly were clearly prevalent in TM helices whereas polar amino acids (Glu, Lys, Asp, Arg, and Gln) were less frequent in this type of helix. The distribution of amino acids along TM helices was also examined. As expected, hydrophobic and slightly polar amino acids are commonly found in the hydrophobic core of the membrane whereas aromatic (Trp and Tyr), Pro, and the hydrophilic amino acids (Asn, His, and Gln) occur more frequently in the interface regions. Charged amino acids are also statistically prevalent outside the hydrophobic core of the membrane, and whereas acidic amino acids are frequently found at both cytoplasmic and extra-cytoplasmic interfaces, basic amino acids cluster at the cytoplasmic interface. These results strongly support the experimentally demonstrated biased distribution of positively charged amino acids (that is, the so-called the positive-inside rule) with structural data.

Keywords Membrane protein · Transmembrane helices · Amino acid distribution · Statistical analysis

Introduction

Although helical membrane proteins constitute approximately one quarter of all proteins in living organisms (Wallin and von Heijne 1998), the rules governing their folding are still not completely established. The hydrophobic effect is a dominant force driving folding of water-soluble proteins, but its contribution to the folding of membrane proteins is more complex, given that these proteins "live" in a biophysical environment—the membrane—which is clearly different from aqueous media. The cell membrane is a very heterogeneous medium, composed mainly of phospholipids that are self-organized into two leaflets giving rise to the formation of a bilayer. The hydrocarbon core, of dimension approximately 30 Å, is the hydrophobic part of the membrane. The polar head groups of the phospholipids define the lipid/water interface and add approximately 15 Å to the thickness of each leaflet (White and Wimley 1999). It is in this complex environment that membrane proteins must fold into their native conformations.

The hydrocarbon core of biological membranes and the interior of folded water-soluble proteins are hydrophobic.

Special issue: Structure, function, folding and assembly of membrane proteins—Insight from Biophysics.

C. Baeza-Delgado · I. Mingarro (✉)
Departament de Bioquímica i Biologia Molecular,
Universitat de València, Burjassot, Spain
e-mail: Ismael.Mingarro@uv.es

M. A. Marti-Renom
Structural Genomics Team, Genome Biology Group,
National Center for Genomic Analysis (CNAG),
Barcelona, Spain

M. A. Marti-Renom (✉)
Structural Genomics Group, Center for Genomic Regulation
(CRG), Barcelona, Spain
e-mail: mmarti@cpb.ub.cat

In such a hydrophobic environment, the polarity of the polypeptide backbone is energetically unfavorable. Thus, in protein structures, nearly all the polar groups of the peptide bond (carbonyl and amide groups) tend to hydrogen bond with one another, leading to secondary structure that stabilizes the folded state. Alpha-helices are the commonest secondary structures found in water-soluble and membrane protein structures. However, the distribution of the helices in these two groups of proteins is very different. Whereas helices in water-soluble proteins can be exposed to both the hydrophobic core and the water-accessible surface, transmembrane (TM) helices in membrane proteins are surrounded by a hydrophobic lipid phase in which water is essentially absent. Therefore, for structural stabilization of helical membrane proteins that reside in this apolar (low dielectric) environment, hydrogen bonding and van der Waals packing forces are highly important.

Although the vast majority of membrane proteins integrate into biological membranes through the translocon (recently reviewed by Martínez-Gil et al. 2011), our current biophysical understanding of its folding and function is hampered by the scarcity of structural information. Fortunately, the number of high-resolution structures of membrane proteins has increased exponentially in recent years (White 2004, 2009). Consequently, a new statistical survey of the properties of TM helices is timely.

In this paper, we revisit the differences between helices from water-soluble proteins and TM helices in terms of length and amino acid composition. In addition, we analyze the distribution of amino acids in TM segments, which are energetically accommodated in the highly heterogeneous media of biological membranes as a result of favorable interaction with the local environment. This study involved 170 helical membrane proteins with known three-dimensional structure and topology, containing a total of 792 TM segments, which were compared with 7,348 helices from 930 water-soluble protein structures. Approximately half of all amino acids are randomly distributed when allocated to the membrane, but the others correlate strongly with amino acid positions along the TM regions.

Methods

Helix data sets

Two data sets for water-soluble and TM helices were obtained from the protein data bank (PDB) (Berman et al. 2000) and the MPtopo database (Jayasinghe et al. 2001b), respectively.

First, a total of 4,405 structural chains deposited in the PDB (as of November 17th, 2011) that passed the following criteria were selected:

1 their total secondary structure had more than 60 % α-helices and no β-strands;
2 their crystallographic resolution was 2.0 Å or higher; and
3 the word *MEMBRANE* did not appear in either the "TITLE" or "DESCRIPTION" fields of the PDB file.

Furthermore, to remove redundancy, the 4,405 chain sequences were compared with each other by use of *cd-hit* software (Huang et al. 2010) and pairs resulting in sequence alignments with 80 % or higher identity were discarded. The final set of 930 non-redundant PDB chains was parsed to identify a total of 7,348 helices from the "HELIX" fields of each PDB chain entry. Thus, the data set of water-soluble helices contained 930 non-redundant and high-resolution protein structures, 7,348 α-helices, and 108,277 amino acids.

Second, all α-helical membrane proteins deposited in the MPtopo database (last updated on January 19th, 2010) (Jayasinghe et al. 2001b), and thus with known membrane insertion topology, were selected. The initial set was filtered by:

1 removing any entry of unknown structure as based on the MPtopo entry classification (i.e., keeping only entries described as "3D_helix" and "1D_helix"); and
2 removing redundant pairs at 80 % sequence identity by use of *cd-hit* software (Huang et al. 2010).

The final data set of TM helices contained 170 non-redundant structures, 837 TM helices, and 20,079 amino acids. Furthermore, to properly analyze the prevalent locations of amino acids in single membrane-spanning TM helices, we discarded any helix shorter than 17 amino acids or larger than 38 amino acids. The resulting TM data subset contained 792 TM helices, and 19,356 amino acids.

Measurement of the prevalent locations of amino acids

We calculated three different measures:

1 probability and percent,
2 Odds, and
3 LogOdds.

The probability (p_i) of an amino acid i is defined as:

$$P_i = \frac{n_i}{N}$$

where i is the amino acid type (one of the 20 amino acids), n_i is the observation count of amino acid i, and N is all amino acids in the data set. Similarly, the percentage of a given amino acid i is defined as its probability multiplied by 100. The Odds (O_i) of an amino acid i is defined as:

$$O_i = \frac{p_{i,c}}{(1-p_{i,c})} \Big/ \frac{p_{i,r}}{(1-p_{i,r})}$$

where $p_{i,c}$ is the probability of amino acid i in class c (for example, TM helix) and $p_{i,r}$ is the probability of the amino acid i in class r (for example, water-soluble helix).

Similarly, the LogOdds for a given amino acid *i* is defined as the logarithm to the base 10 of its Odds. Briefly, Odds higher than 1 (or positive LogOdds) indicate over-occurrence of the amino acid type in the class. Odds smaller than 1 (or negative LogOdds) indicate under-representation of the amino acid type in the class.

Results and discussion

Helix length in membrane and water-soluble proteins

Length distributions for helices found in high-resolution structures deposited in the PDB (Berman et al. 2000) are very different for TM and water-soluble proteins (Fig. 1).

Helices in TM proteins are, on average, 24.0 (±5.6) amino acids long; this result differs slightly from previous data obtained by using databases with 45 (Bowie 1997) and 129 (Ulmschneider and Sansom 2001) TM helices, for which average helix length was 26.4 and 27.1 amino acids, respectively. Because the translation per amino acid in a canonical helix is 1.5 Å, a stretch of approximately 20 consecutive hydrophobic amino acids can span the 30 Å of the hydrocarbon core of a biological membrane. Indeed, the most prevalent (∼12 %) length of TM helices in our data set was 21 amino acids (Fig. 1). Longer helices can span the bilayer with concomitant tilting of the helix axis relative to the membrane plane. Other options are also feasible, ranging from lipid accommodation to polypeptide backbone deformation (Holt and Killian 2009).

Helices from water-soluble proteins have an average length of 14.7 (±8.7) amino acids, which agrees with previous studies in which the most prevalent helix length was 10–11 amino acids (Engel and DeGrado 2004; Pal et al. 2003). The shorter length of helices in water-soluble proteins is because of the absence of the restrictions imposed by the low dielectric constant at the hydrocarbon core of biological membranes, which forces the polypeptide backbone to adopt, on average, larger secondary structures.

Amino acid composition of α-helices

Amino acid composition was examined for both TM and water-soluble helices (Fig. 2). TM helices of lengths

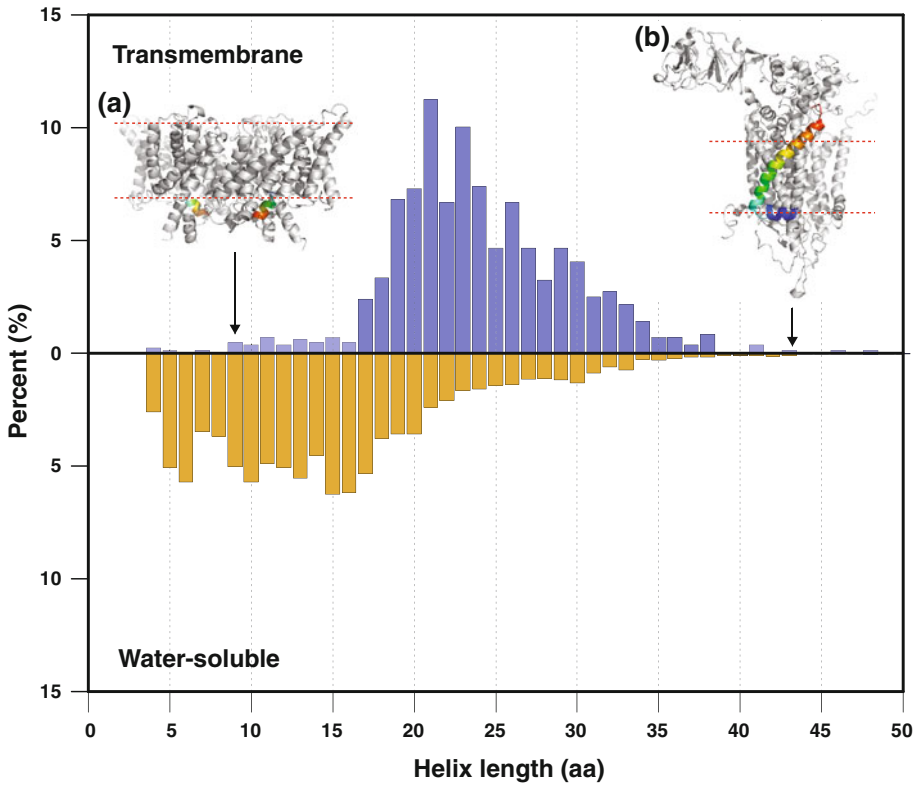

Fig. 1 Length distributions for 837 TM and 7,348 water-soluble helices from a set of non-redundant proteins of known structure (see the "Methods" section). Transmembrane helices are shown in *blue* (*pale blue* corresponds to discarded lengths) and water-soluble helices are shown in *orange*. **a** Example of a short nine-amino-acid-length helix in the ClC chloride channel from *E. coli* (1KPK entry in PDB). Membrane boundaries were obtained from the PPM server (Lomize et al. 2012). The selected membrane is shown in *rainbow coloring* from the N-terminal (*blue*) end to the C-terminal (*red*) end. **b** Example of a large 43-amino-acid-length helix in the chicken cytochrome BC1 complex (1BCC entry in the PDB); the N-terminus of the helix (*blue*) lies at the membrane/water interface. Representation as in inset **a**

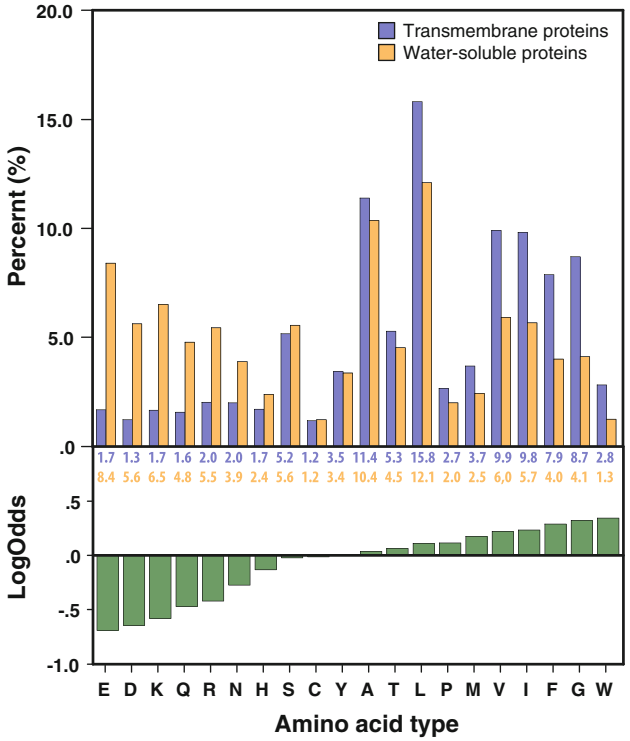

Fig. 2 Amino acid type distribution from 792 TM and 7,348 water-soluble helices from a set of non-redundant proteins of known structure (see the "Methods" section). (*Upper plot*) Amino acid type distribution for TM helices in *blue* and for water-soluble helices in *orange*. (*Lower plot*) LogOdds values for comparison of the relative abundance of each amino acid type in TM and water-soluble helices. Amino acid types are ordered by LogOdds values

between 17 and 38 amino acids were selected from the MPtopo database (Jayasinghe et al. 2001b); these included helical segments that do completely span the hydrophobic core of the membrane. TM helices shorter than 17 amino acids and larger than 38 amino acids were excluded, because they may not cross the membrane entirely (Fig. 1, inset a) or may contain segments parallel to the membrane (Fig. 1, inset b). Note that for water-soluble helices all lengths were included in our analysis because no restrictions in terms of length can be assumed for water-soluble proteins in an aqueous environment.

As expected, hydrophobic amino acids Leu, Ala, Val, and Ile constitute the bulk of the amino acids in the TM region accounting for almost half (47.0 %) of all amino acids. Similarly, these amino acids are also frequently found in the helices of water-soluble proteins (34.1 %). However, there are, as noted previously using smaller datasets (Bywater et al. 2001), differences between the composition of the two types of helix. Despite sharing the same structural features, the differences between the two types of helix are reflected by their preferential occurrences, as measured by the logarithm of the Odds of finding a given amino acid in a TM helix compared with its frequency in a water-soluble helix (Fig. 2 bottom panel). For example, whereas charged and polar amino acids are much more frequently found in helices of water-soluble proteins, Trp, Gly, and Phe are more likely to occur in TM helices. Interestingly, in contrast with their prevalent conformations in water, the likelihood of amino acids such as Val, Ile, Phe, and Met occurring in a helical structure are notably increased in the membrane environment, and it has been suggested that their prevalence in helices depends primarily on their side chain hydrophobicity and on the hydropathy of the local polypeptide region in which the amino acids reside spanning the membrane (Li and Deber 1994). Significantly, Gly and Pro are more frequent in TM helices than in water-soluble helices. Although commonly regarded as "helix breakers" it has been reported that Gly occurs frequently in TM helix–helix interactions, especially in association with β-branched residues at neighboring positions (Senes et al. 2000), and that Pro, in addition to its function in signal transduction and gating across the membrane, may also be significantly involved in these processes (Orzáez et al. 2004).

Comparison of amino acid frequency between TM and water-soluble helices confirmed that strongly polar amino acids (Glu, Lys, Asp, Arg, and Gln) are more prevalent in water-soluble helices (Fig. 3). These amino acids constitute only 8.2 % of the amino acids within TM helices compared with 30.9 % of those in water-soluble helices. Despite their lower occurrence, polar amino acids are evolutionary

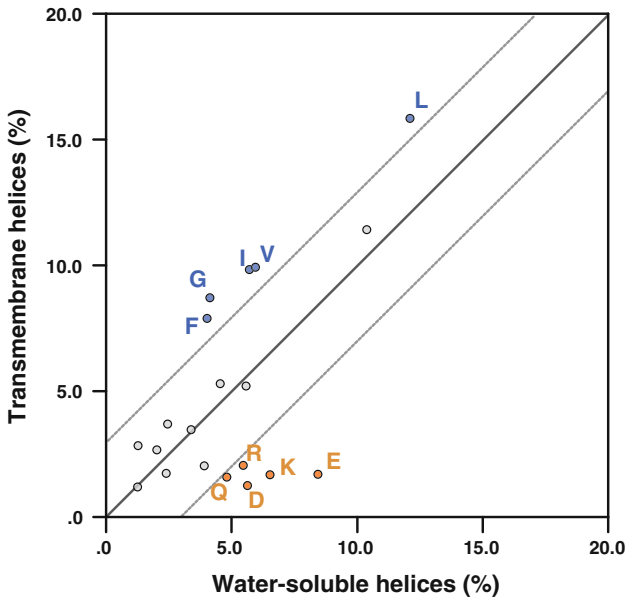

Fig. 3 Amino acid type percentage comparison between TM and water-soluble helices. *Blue colored* amino acids are over-represented (difference >3 % points) in TM helices compared with water-soluble helices. *Orange colored* amino acids are over-represented (difference >3 % points) in water-soluble helices compared with TM helices. *Dashed grey lines* indicate a cut-off of 3 % difference points

conserved in TM proteins, which has been partially explained by their tendency to be buried in the protein interior and, in many cases, because of their direct involvement in the function of the protein (Illergård et al. 2011). Conversely, hydrophobic amino acids (Leu, Val, Ile, Gly, and Phe) are over-represented in TM helices (Fig. 3). Interestingly, Ala, although the second most abundant amino acid in TM helices (Fig. 2), it is not over-represented in this type of helix; this is probably because its greater tendency to participate in a helical structure in aqueous environments (Blaber et al. 1993) than in membrane-mimetic environments (Li and Deber 1994). In fact, both biological (Nilsson et al. 2003; Hessa et al. 2005) and biophysical (Jayasinghe et al. 2001a) measurements have placed Ala at the threshold between those amino acids that promote membrane integration of TM helices and those that preclude membrane insertion.

Position-dependent distribution of amino acids in TM helices

Comparison of amino acid frequency at different positions in a TM segment, taking as reference the TM center, confirmed that approximately half of the natural amino acids have similar distributions in positive positions (toward the inside of the cell) than at negative positions (toward the outside of the cell) (Fig. 4). It was found that not only the strongly hydrophobic amino acids but also Gly and the hydroxylated amino acids Ser and Thr are equally distributed along the hydrophobic core of the membrane. It is important to note that Gly is normally regarded as being conducive to turn (Williams et al. 1987), yet it is a common amino acid in TM helices (Fig. 2). There are important folding reasons for incorporating Gly into TM helices. The absence of a side-chain from Gly enables bulkier groups to

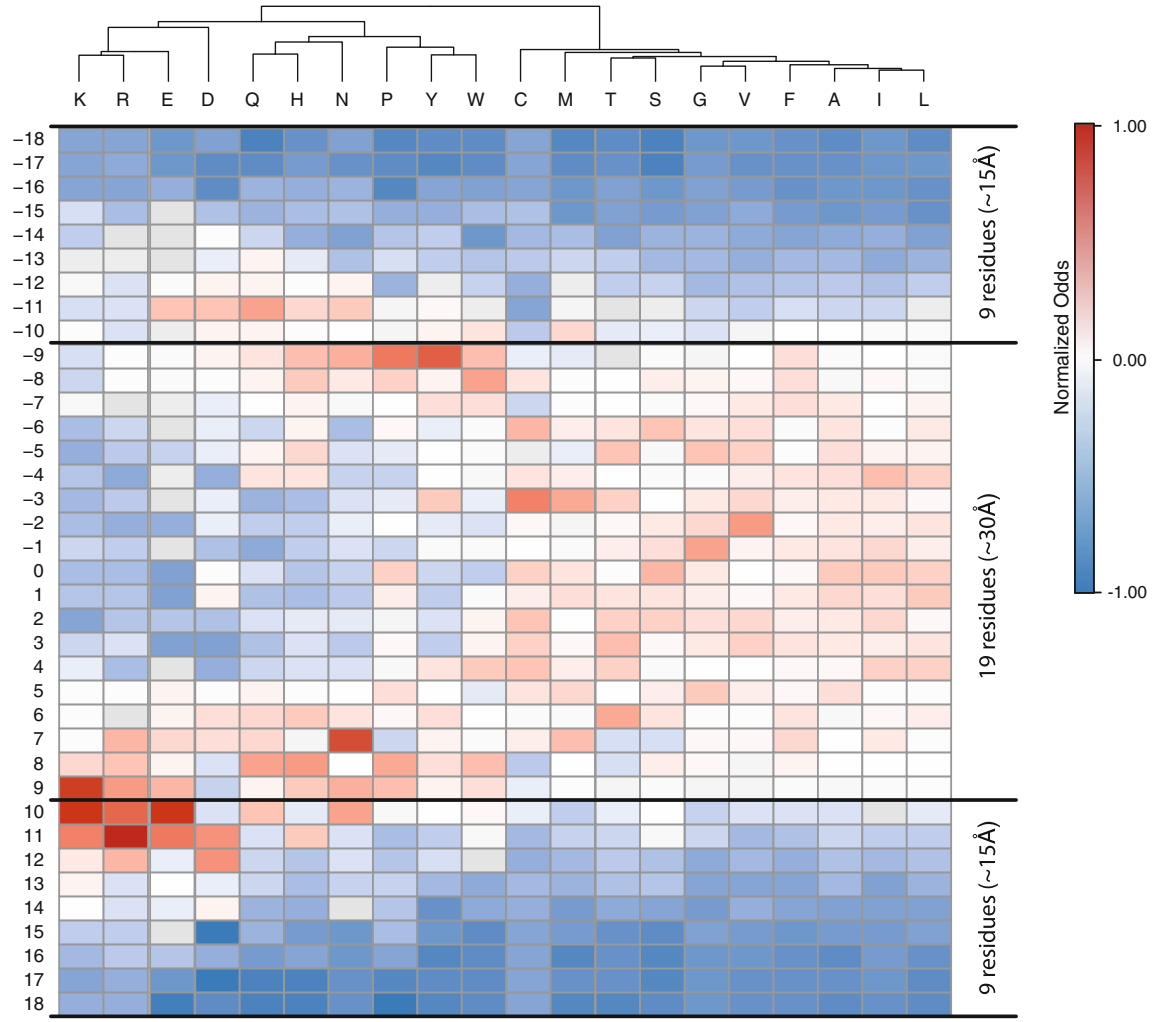

Fig. 4 Amino acid type and position distribution in TM helices. Each amino acid type and its positioning in the TM helix is represented by its position-normalized Odds (that is, for each column the Odds are normalized to an average of zero and a standard deviation of unity). The amino acids are clustered on the basis of their positional normalized Odds within the helices. Positively labeled positions indicate the cytoplasmic side of the membrane and its flanking region whereas negatively labeled positions are indicative of extra-cytoplasmic regions

be accommodated close to the polypeptide backbone of the TM helices. This might be important for intramolecular helix–helix packing, for homo-oligomerization, or for recognition of other membrane proteins, among other factors. Indeed, it has been observed that Gly has the highest overall packing value in membrane proteins (Eilers et al. 2002). Ser or Thr within TM helices participate in hydrogen-bonding networks by hydrogen bonding of the side chain oxygen atom to the acceptor side chain or peptide bond groups. These effects, intimate packing (Gly) and hydrogen bonding (Ser and Thr), can be relevant at any position along the TM region, which could explain the absence of position prevalence of these amino acids in TM helices. Met or Cys are also frequent at different locations within the hydrophobic core, but relative prevalence can be observed in a region that would correspond to the initial portion of the polar headgroups of the phospholipids, consistent with the slightly amphipathic nature of these amino acids and in agreement with its distribution in the lipid bilayer recently obtained from molecular dynamics simulation (MacCallum et al. 2008).

Whereas Phe has a flat distribution in TM helices, behaving as a hydrophobic amino acid, distribution of Trp, Tyr, and Pro is biased—they are most likely to be found at the ends of the bilayer (i.e. at the interface between the hydrophobic core of the bilayer and the bulk water). At this location, aromatic amino acids may serve as anchors for the TM helices in the membrane. In fact, Trp and Tyr positioned 7–9 amino acids away from the center of a TM segment result in reduction of the free energy (Hessa et al. 2007), which correlates well with our statistical distribution from three-dimensional structures (Fig. 4). The biophysical reason for the observed distribution of Trp and Tyr could rely on the relatively amphipathic nature of their side chains, which can form hydrogen bonds and also have hydrophobic character. Actually, this prevalent location has previously been observed not only for α-helical but also β-barrel membrane proteins (Ulmschneider and Sansom 2001). A similar distribution is observed for Pro, although increased prevalence is detectable toward the center of the bilayer, which might be associated with the fundamental and subtle function of Pro in the dynamics, structure, and function of many membrane proteins of inducing the formation of molecular hinges (Cordes et al. 2002). Indeed, thirteen TM helices with known structure have Pro at the 0 position, which in all cases results in a kink in the helix. Nevertheless, it should be noted that the interfacial prevalence of these three amino acids is somehow more pronounced at the non-cytoplasmic interface. This was also observed for the aromatic amino acids (Trp and Tyr) in a membrane protein prediction analysis using sequence information from 107 genomes (Nilsson et al. 2005).

The distribution pattern for Asn, His, and Gln, corresponds to an interfacial preference close to the end of the TM regions, which is consistent with the amphipathic nature of these molecules. This pattern was previously reported for His (Ulmschneider and Sansom 2001), and is in good agreement with our results. Interestingly, in more recent studies using computer simulations, it has been noted that small molecule analogs of Asn (MacCallum et al. 2008) and Asn, His, and Gln (Johansson and Lindahl 2007) result in an energy minimum for partition into model lipid bilayers.

Because the energy cost of inserting an ionizable group in the hydrophobic environment of the membrane is very high (White and Wimley 1999), charged amino acids should generally be excluded from the hydrophobic core of the TM helices. Interestingly, nearly all membrane proteins with six or more predicted TM helices contain at least one ionizable amino acid (Arkin and Brunger 1998). However, charged amino acids consistently cluster at the TM flanking regions (Fig. 4). For example, increased distribution of acidic amino acids (Asp and Glu) occurs on both the cytoplasmic and extra-cytoplasmic sides of the membrane, although with some prevalence for the cytoplasmic region. Distribution of positively charged amino acids (Arg and Lys) is even more strongly asymmetric between opposite sides of the membrane, in good agreement with the positive-inside rule (von Heijne 1992). Moreover, it has been demonstrated experimentally that basic amino acids act as stronger topological signals than acidic amino acids (Nilsson and von Heijne 1990; Saurí et al. 2009), which is reflected by their different statistical occurrence on either end of the TM segments. Nevertheless, when considered globally, charged amino acids cluster predominantly near the cytoplasmic end of the TM segments (Fig. 5, orange line). This effect has already been noted in a previous structure-based analysis that included the fewer structures available at the time (Ulmschneider et al. 2005). In contrast, although polar amino acids (Gln, His, and Asn) mimic the distribution pattern of charged amino acids, avoiding the more hydrophobic region of the bilayer, they tend to occur in the extra-cytoplasmic region (Fig. 5). Trp, Tyr, and Pro are more abundant approximately eight or nine amino acid positions from the center of the membrane, that is, within the interface region, but with some bias toward the extra-cytoplasmic interface. The other natural amino acids are more abundant at the center of the bilayer, within seven amino acid positions on both sides of the membrane normal, but are also very frequently found beyond this boundary, as noted by their overall proximity to the Odd value of 1 for positions >10 on both sides of the center of the membrane (Fig. 5). Interestingly, the amino acid distribution patterns in both interface regions are slightly different. There is a sharper transition from mainly

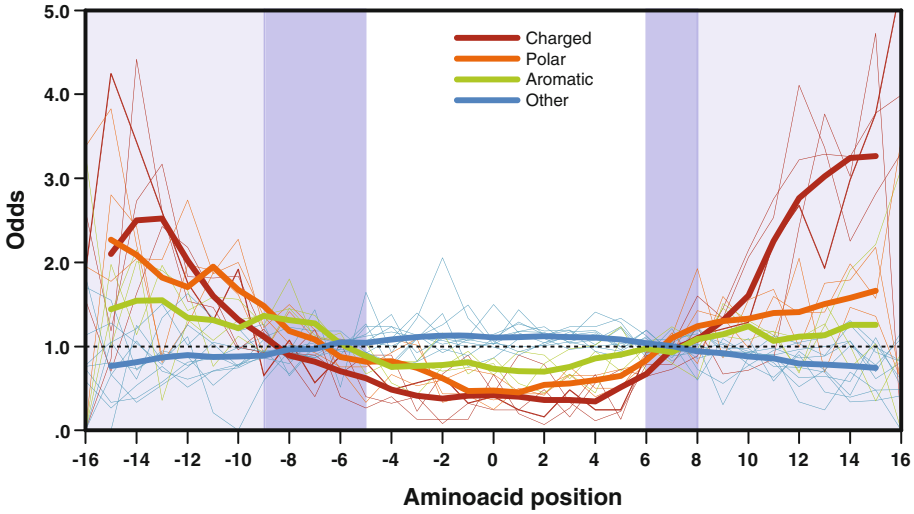

Fig. 5 Most likely positions of amino acid groups in a membrane. *Thin lines* represent the positional Odds for each amino acid individually, whereas *thick lines* represent the average positional Odds for each group of amino acids obtained from Fig. 4. Amino acid types are grouped as in the dendrogram in Fig. 4, i.e. charged amino acids (*red* KRED), polar amino acids (*orange* QHN), aromatic amino acids plus Pro (*green* PYW), and the other amino acids (*blue* CMTSGVFAIL)

hydrophobic to charged, polar, and aromatic amino acids on the cytoplasmic side of the membrane (positions 6–8) than on the extra-cytoplasmic side (positions −5 to −9). The different lipid composition between the two lipid leaflets in biological membranes and the strong electrochemical potential over the prokaryotic inner cell membranes can exert an important effect, which may be reflected by this difference. For instance, asymmetry in the distribution of amino acids within TM segments from plasma membrane proteins has recently been reported (Sharpe et al. 2010), and has been attributed to asymmetry in the state of lipid order in the membrane. Such asymmetry is likely to be because of enrichment of lipids, for example sterols and sphingolipids, in the extra-cytoplasmic leaflet, where more gradual amino acid distribution can be expected.

Finally, we analyzed and plotted the odds ratio for each amino acid in three regions in a membrane, that is, taking the hydrophobic TM region as the central 19 positions (∼30 Å) and nine amino acid positions (∼15 Å) on both sides as the extra-cytoplasmic (from −10 to −18 amino acids) and cytoplasmic (from 10 to 18) flanking regions (Fig. 6). Hydrophobic amino acids (blue colored) predominated in the hydrophobic center. However, this trend is not observed for the more prevalent amino acids in TM segments (for example Leu, Fig. 2), which are also frequently found in the flanking regions. A minor increase is observed for Trp, Tyr, and Pro (green) in the extra-cytoplasmic flanking region. The absence of larger differences for the distribution of these amino acids is probably because of their precise location at the interface between the hydrophobic core and the flanking hydrophilic environment. Polar (orange) amino acids (Gln, His, and Asn) predominate in both flanking regions, because their presence within the membrane core is energetically unfavorable. These amino acids do not ionize at physiological pH

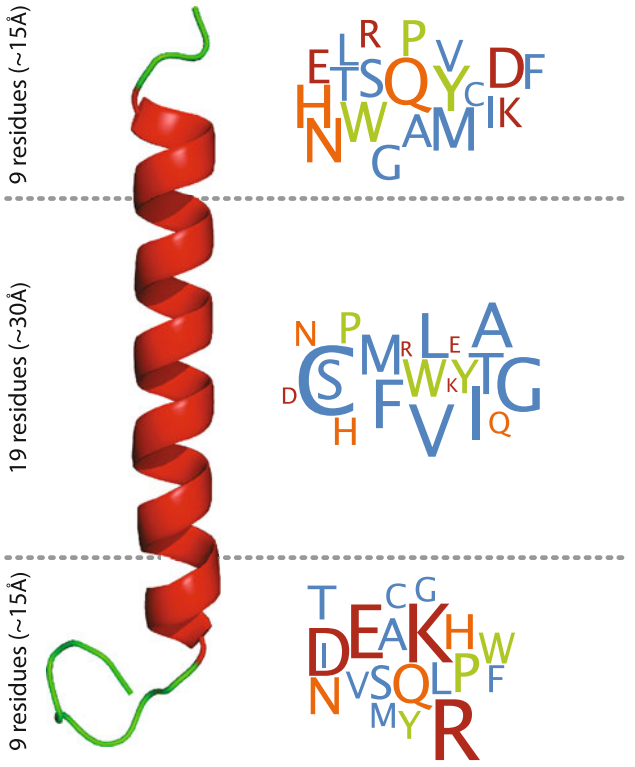

Fig. 6 Amino acid location prevalence in a membrane. *Letter size* is proportional to the odds (relative prevalence) of finding a given amino acid in the three regions in a membrane (i.e., from *top* to *bottom* outer, membrane, and inner regions). Amino acids are colored as in Fig. 5

and can donate and accept hydrogen bonds simultaneously. This effect is manifested as greater occurrence of Gln, His, and Asn in the rich hydrogen-bond network environment of the interface. Charged amino acids (red) were underrepresented in the hydrophobic core and tended to occur in the cytoplasmic flanking region, with acidic amino acids more prevalent in the extra-cytoplasmic flanking region. Furthermore, basic amino acids are strong topological

determinants that heavily populate the cytoplasmic flanking region. The effect of positively charged amino acids located near the cytoplasmic end of hydrophobic segments has been estimated to contribute approximately −0.5 kcal/mol to the apparent free energy of membrane insertion (Lerch-Bader et al. 2008). This energy contribution can be extremely relevant for precise anchoring of hydrophobic regions to biological membranes.

Concluding remarks

We have compared the length and amino acid composition of helices in TM and water-soluble proteins. Overall, significant differences are observed for these proteins; these may be attributed to the biophysical differences between the two environments in which they fold.

- First, TM helices adapt their length to the dimensions and constraints of biological membranes, whereas water-soluble helices are statistically shorter because they do not have to satisfy the demanding restrictions imposed by the complexity of the membrane environment.
- Second, the observed differences indicate that in the lipid bilayer, an environment which forces secondary structure formation, amino acid side chain hydrophobicity prevails over helicity. Accordingly, aliphatic amino acids with reduced tendency to form a helix (Val, Ile, Gly, and Phe) are abundant in TM helices, whereas polar amino acids (Glu, Lys, and Arg) with high tendency to form a helix are consistently less frequent in TM helices.
- Third, half of the natural amino acids are equally distributed along TM helices whereas aromatic, polar, and charged amino acids plus Pro are biased toward the ends of the TM helices.
- Fourth, as previously observed, the distribution of charged amino acids was asymmetric, occurring more frequently on the cytoplasmic side of the membrane, which causes net charge unevenness on both sides of the membrane. In addition to this asymmetry, Trp, Tyr, and Pro were found to be more frequent at the extracytoplasmic interface of the membrane and the polar amino acids (Gln, His, and Asn) at the extra-cytoplasmic flanking region of the TM helices.
- Fifth, transitions between the different types of amino acid at the ends of the hydrophobic core occur in a more defined region on the cytoplasmic side than at the extracytoplasmic face, probably reflecting the different lipid composition of both leaflets of biological membranes.

The conclusions on TM helix architecture described here should prove useful for constructing models of membrane proteins with desired properties, which could help filling in some of the many gaps in our knowledge in this field.

Acknowledgments This work was supported by grants BFU2009-08401 (to I.M.) and BFU2010-19310 (to M.A.M-R.) from the Spanish Ministry of Science and Innovation (MICINN, ERDF supported by the European Union), and by PROMETEO/2010/005 and ACOMP/2012/226 (to I.M.) and ACOMP/2011/048 (to M.A.M-R.) from the Generalitat Valenciana. C.B–D. was recipient of a predoctoral FPI fellowship from the MICINN.

References

Arkin IT, Brunger AT (1998) Statistical analysis of predicted transmembrane alpha-helices. Biochim Biophys Acta 1429:113–128

Berman HM, Berman HM, Westbrook J, Westbrook J, Feng Z, Feng Z, Gilliland G, Gilliland G, Bhat TN, Bhat TN, Weissig H, Weissig H, Shindyalov IN, Shindyalov IN, Bourne PE, Bourne PE (2000) The protein data bank. Nucleic Acids Res 28:235–242

Blaber M, Zhang XJ, Matthews BW (1993) Structural basis of amino acid alpha helix propensity. Science 260:1637–1640

Bowie JU (1997) Helix packing in membrane proteins. J Mol Biol 272:780–789

Bywater RP, Thomas D, Vriend G (2001) A sequence and structural study of transmembrane helices. J Comput Aided Mol Des 15:533–552

Cordes FS, Bright JN, Sansom MSP (2002) Proline-induced distortions of transmembrane helices. J Mol Biol 323:951–960

Eilers M, Patel AB, Liu W, Smith SO (2002) Comparison of helix interactions in membrane and soluble alpha-bundle proteins. Biophys J 82:2720–2736

Engel DE, DeGrado WF (2004) Amino acid propensities are position-dependent throughout the length of alpha-helices. J Mol Biol 337:1195–1205

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

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

Holt A, Killian JA (2009) Orientation and dynamics of transmembrane peptides: the power of simple models. Eur Biophys J 39:609–621

Huang Y, Huang Y, Niu B, Niu B, Gao Y, Gao Y, Fu L, Fu L, Li W, Li W (2010) CD-HIT Suite: a web server for clustering and comparing biological sequences. Bioinformatics 26:680–682. Available at: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmedandid=20053844andretmode=refandcmd=prlinks

Illergård K, Kauko A, Elofsson A (2011) Why are polar residues within the membrane core evolutionary conserved? Proteins 79:79–91

Jayasinghe S, Hristova K, White SH (2001a) Energetics, stability, and prediction of transmembrane helices. J Mol Biol 312:927–934

Jayasinghe S, Jayasinghe S, Hristova K, Hristova K, White SH, White SH (2001b) MPtopo: a database of membrane protein topology. Protein Sci 10:455–458

Johansson ACV, Lindahl E (2007) Position-resolved free energy of solvation for amino acids in lipid membranes from molecular dynamics simulations. Proteins 70:1332–1344

Lerch-Bader M, Lundin C, Kim H, Nilsson I, von Heijne G (2008) Contribution of positively charged flanking residues to the insertion of transmembrane helices into the endoplasmic reticulum. Proc Natl Acad Sci USA 105:4127–4132

Li SC, Deber CM (1994) A measure of helical propensity for amino acids in membrane environments. Nat Struct Biol 1:558

Lomize MA, Pogozheva ID, Joo H, Mosberg HI, Lomize AL (2012) OPM database and PPM web server: resources for positioning of proteins in membranes. Nucleic Acids Res 40:370–376

MacCallum JL, Bennett WFD, Tieleman DP (2008) Distribution of amino acids in a lipid bilayer from computer simulations. Biophys J 94:3393–3404

Martínez-Gil L, Saurí A, Marti-Renom MA, Mingarro I (2011) Membrane protein integration into the endoplasmic reticulum. FEBS J 278:3846–3858

Nilsson I, von Heijne G (1990) Fine-tuning the topology of a polytopic membrane protein: role of positively and negatively charged amino acids. Cell 62:1135–1141

Nilsson I, Johnson AE, von Heijne G (2003) How hydrophobic is alanine? J Biol Chem 278:29389–29393

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

Orzáez M, Salgado J, Giménez-Giner A, Pérez-Payá E, Mingarro I (2004) Influence of proline residues in transmembrane helix packing. J Mol Biol 335:631–640

Pal L, Chakrabarti P, Basu G (2003) Sequence and structure patterns in proteins from an analysis of the shortest helices: implications for helix nucleation. J Mol Biol 326:273–291

Saurí A, Tamborero S, Martínez-Gil L, Johnson AE, Mingarro I (2009) Viral membrane protein topology is dictated by multiple determinants in its sequence. J Mol Biol 387:113–128

Senes A, Gerstein M, Engelman DM (2000) Statistical analysis of amino acid patterns in transmembrane helices: the GxxxG motif occurs frequently and in association with beta-branched residues at neighboring positions. J Mol Biol 296:921–936

Sharpe HJ, Stevens TJ, Munro S (2010) A comprehensive comparison of transmembrane domains reveals organelle-specific properties. Cell 142:158–169

Ulmschneider MB, Sansom MS (2001) Amino acid distributions in integral membrane protein structures. Biochim Biophys Acta 1512:1–14

Ulmschneider MB, Sansom MSP, Di Nola A (2005) Properties of integral membrane protein structures: derivation of an implicit membrane potential. Proteins 59:252–265

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

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

White SH (2004) The progress of membrane protein structure determination. Protein Sci 13:1948–1949

White SH (2009) Biophysical dissection of membrane proteins. Nature 459:344–346

White SH, Wimley WC (1999) Membrane protein folding and stability: physical principles. Annu Rev Biophys Biomol Struct 28:319–365

Williams RW, Chang A, Juretić D, Loughran S (1987) Secondary structure predictions and medium range interactions. Biochim Biophys Acta 916:200–204