STRUCTURE-BASED STATISTICAL AND EXPERIMENTAL ANALYSIS OF TRANSMEMBRANE HELICES

Carlos Baeza[1,2], Marc A. Marti-Renom[1] and Ismael Mingarro[2]

[1] *Structural Genomics Group, National Center for Genomic Analysis (CNAG), Valencia, Spain*

[2] *Membrane Proteins Laboratory, Department of Biochemistry and Molecular Biology, Universitat de València, Burjassot, Spain*

SUMMARY

The thickness of the hydrocarbon core of the membrane bilayer is ~30 Å, which leads to the expectation that transmembrane (TM) helices of helix-bundle membrane proteins (MPs) should be at least ~20 amino acids long to span this region. Although some studies have previously shown that 10-residue polyleucine segments can be inserted into lipid bilayers [1, 2], there are no experimental evidences on the minimum length for a TM segment to be inserted into the membrane by the translocon machinery. Recent advances in high-resolution structure determination of membrane proteins enable now the analysis of the main features of amino acids in transmembrane (TM) segments. Using structures of integral membrane proteins and structures of water-soluble proteins, we have performed a statistical analysis of length and amino acid composition in TM and compared it with that of water-soluble helices to emphasize the differences between them. We have also calculated the position distribution of amino acids in TM domains and, using these data, we have designed a set of putative TM segments with different lengths and amino acid composition. We finally have calculated the probability of insertion for the designed sequences and validated our predictions using an *in vitro* experimental system based on the *Escherichia Coli* inner-membrane protein Lep as a model protein [3, 4].

COMPUTATIONAL RESULTS

Two data sets of water-soluble and TM helices were obtained from the Protein Data Bank (PDB) [5] and the MPTOPO databases [6]. The data set of water-soluble helices contained 930 non-redundant and high-resolution protein structures, 7,348 a-helices and 108,277 amino acids. The data set of TM helices contained 170 non-redundant structures, 837 TM helices, and 20,079 amino acids.

LENGTH DISTRIBUTIONS FOR TM AND WATER-SOLUBLE HELICES

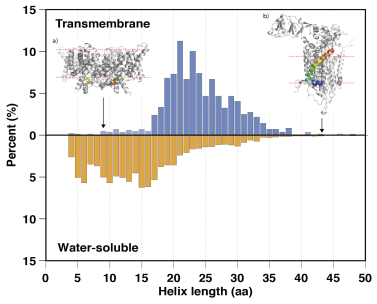

Figure 1. Length distributions for 837 TM and 7,348 water-soluble helices from a set of non-redundant proteins of known structure. Transmembrane helices in blue and water-soluble helices in orange. (a) Example of a short 9 amino acid length helix from 1KPK entry in PDB. Membrane boundaries were obtained from the OPM database [7]. The selected membrane is shown in rainbow coloring from N- (blue) to C-terminal (red) ends. (b) Example of a large 43 amino acid length helix from 1BCC entry in PDB, which N-terminus of the helix (blue) lies at the membrane/water interface. Representation as in inset (a).

AMINO ACID TYPE DISTRIBUTION

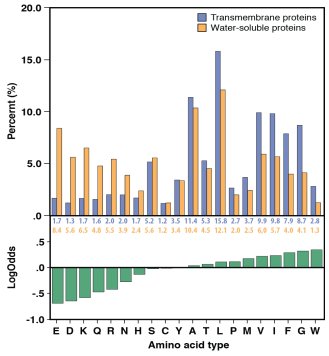

Figure 2. Amino acid type distribution from 792 TM (any helix shorter than 17 amino acids or larger than 38 amino acids were discarded from the 837 TM helices in the data set) and 7,348 water-soluble helices from a set of non-redundant proteins of known structure. (Upper plot) Amino acid type distribution for TM helices in blue and for water-soluble helices in orange. (Lower plot) LogOdds values for comparing the relative abundance of each amino acid type in TM and water-soluble helices. Amino acid types are ordered by its LogOdds.

AMINO ACID POSITION DISTRIBUTION IN TM SEGMENTS

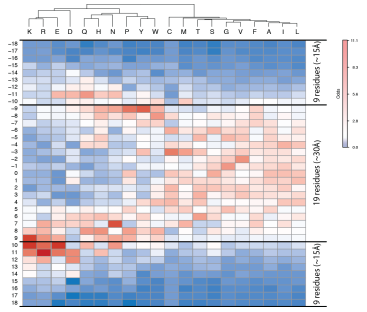

Figure 3. Amino acid type and position distribution in TM helices. Each amino acid type and their positioning in the TM helix is represented by their positional Odds. The amino acids are clustered based on their positional Odds within the helices. Positively labeled positions refer to the cytoplasmic side of the membrane and its flanking region whilst negatively labeled positions refer to extra-cytoplasmic regions

INSERTION PROBABILITY OF GENERATED SEQUENCES

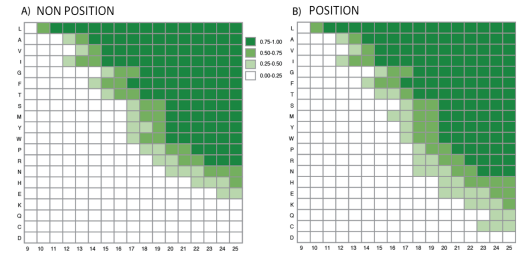

Figure 3. TM segment insertion probability. Each square shows the average of the probability insertion of a subset of 1,000 sequences generated using the distribution of aa in TM (A) and using the distribution per position (B). The probability of insertion were calculated using the *ΔG Prediction Server* [3, 4].

EXPERIMENTAL RESULTS

EXPERIMENTAL SYSTEM: THE LEP MODEL PROTEIN

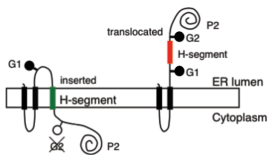

Figure 4. The Lep model protein. H-segments (green/red) are engineered into the P2 domain with two flanking glycosylation acceptor sites (G1, G2). Constructs for which the H-segment is integrated into the endoplasmic reticulum membrane as a TM helix are glycosylated only on the G1 site (left), whereas those for which the H-segment is translocated across the membrane are glycosylated on both the G1 and G2 sites (right). Adapted from Hessa et al, 2005 [3].

EXPRESSION *IN VIVO* AND GLYCOSYLATION ASSAY

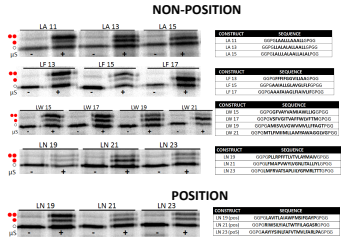

Figure 5. Glycosilation assay, SDS-PAGE and autoradiography. Constructs were transcribed and translated in the *TnT Quick System* (Promega) with ^{35}S-Met and in the presence (+) and absence (-) of dog pancreas rough microsomes (μS). Translation products were analysed by SDS–PAGE and the gels were quantified using a Fuji FLA-3000 phosphorimager. Bands of unglycosylated protein are indicated by a white dot; singly and doubly glycosylated proteins are indicated by one and two red dots, respectively.

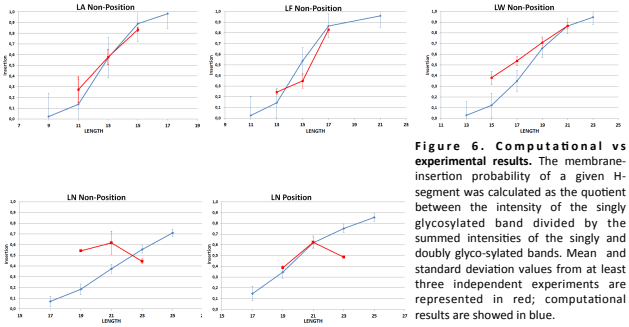

Figure 6. Computational vs experimental results. The membrane-insertion probability of a given H-segment was calculated as the quotient between the intensity of the singly glycosylated band divided by the summed intensities of the singly and doubly glyco-sylated bands. Mean and standard deviation values from at least three independent experiments are represented in red; computational results are showed in blue.

REFERENCES

[1] Chen HF and Kendall DA *J Biol Chem* **270(23)** (1995) 14115.
[2] Jaud S. et al. *Proc Natl Acad Sci U S A* **106(28)** (2009) 11588.
[3] Hessa T. et al. *Nature* **433(7024)** (2005) 377.
[4] Hessa T. et al. *Nature* **450(7172)** (2007) 1026.
[5] Berman HM et al. *Nucleic Acids Res.* **28(1)** (2000) 235.
[6] Jayasinghe S. et al. *Protein Sci.* **10** (2001) 455.
[7] Lomize MA et al. *Nucleic Acid Res.* (2012) in process.
[8] Von Heijne G. *J Mol Biol* **225(2)** (1992) 487.

CONCLUSIONS

- TM helices adapt their length to the dimensions and constraints of biological membranes, while water-soluble helices are statistically shorter.
- The observed differences highlight that in the lipid bilayer, which environment forces secondary structure formation, amino acid side chain hydrophobicity prevails to helicity.
- The position within α-helical TMs is strongly restricted for interfacial and polar charged residues. Positively charged (Arg and Lys) residues distribution is even more strongly asymmetric between opposite sides of the membrane, in good agreement with the positive-inside rule [8].
- The designed sequences maintaining the structure-based distribution of amino acid residues can theoretically insert into membranes at shorter lengths.
- TM helix architecture described here should prove useful for constructing models of membrane proteins with desired properties.

CONTACT: carlos.baeza@uv.es