

DATABASE

Open Access



TMalphaDB and TMbetaDB: web servers to study the structural role of sequence motifs in α -helix and β -barrel domains of membrane proteins

Marc Perea^{1†}, Ivar Lugtenburg^{1†}, Eduardo Mayol¹, Arnau Cordero¹, Xavier Deupi^{1,3}, Leonardo Pardo^{1*} and Mireia Olivella^{2*}

Abstract

Background: Membrane proteins represent over 25 % of human protein genes and account for more than 60 % of drug targets due to their accessibility from the extracellular environment. The increasing number of available crystal structures of these proteins in the Protein Data Bank permits an initial estimation of their structural properties.

Description: We have developed two web servers—TMalphaDB for α -helix bundles and TMbetaDB for β -barrels—to analyse the growing repertoire of available crystal structures of membrane proteins. TMalphaDB and TMbetaDB permit to search for these specific sequence motifs in a non-redundant structure database of transmembrane segments and quantify structural parameters such as ϕ and ψ backbone dihedral angles, χ_1 side chain torsion angle, unit bend and unit twist.

Conclusions: The structural information offered by TMalphaDB and TMbetaDB permits to quantify structural distortions induced by specific sequence motifs, and to elucidate their role in the 3D structure. This specific structural information has direct implications in homology modeling of the growing sequences of membrane proteins lacking experimental structure. TMalphaDB and TMbetaDB are freely available at <http://lmc.uab.cat/TMalphaDB> and <http://lmc.uab.cat/TMbetaDB>.

Keywords: Membrane proteins, Transmembrane segments, Sequence motifs, Structural distortion

Background

Membrane proteins represent over 25 % of all proteins in sequenced genomes and mediate the interaction of the cell with its surroundings, including selective molecular transport, signalling, respiration and motility [1]. Because of their accessibility from the extracellular environment, membrane proteins are targets of over 60 % of currently marketed drugs [2–4]. Due to the difficulty in over-expressing, purifying and crystallizing membrane proteins [5], only 2 % of the structures deposited in Protein Data Bank are membrane proteins [6, 7]. Membrane proteins

display specific features that differ from those of water-soluble ones, due to their different environment [8]. For instance, the number of folds that membrane proteins can adopt is limited to α -helix bundles and β -barrels due to the physical constraints imposed by the lipid bilayer. The lipid bilayer, where the transmembrane (TM) regions are located, is predominantly lipophilic, lacks hydrogen-bonding potential, and provides little screening of electrostatic interactions. Thus, α -helix and β -sheets secondary structure elements maximize the hydrogen bond interactions among backbone atoms, whereas hydrophobic side chains are preferentially oriented toward the membrane lipids. This results in significant differences in amino acid composition [9] and in the probabilities of amino acid substitutions during evolution [10, 11] relative to globular proteins.

* Correspondence: Leonardo.pardo@uab.es; mireia.olivella@uvic.cat

†Equal contributors

¹Laboratori de Medicina Computacional, Unitat de Bioestadística, Facultat de Medicina, Universitat Autònoma de Barcelona, Bellaterra, Spain

²Departament de Biologia de Sistemes, Universitat de Vic, Vic, Barcelona, Spain
Full list of author information is available at the end of the article

Biological function of membrane proteins involves conformational rearrangement of the TM regions. For example, activation of the G protein-coupled receptor family requires the binding of the C-terminal α -helix of the G protein to the intracellular cavity that is opened by the conformational rearrangement of TM6 [12]. Similarly, multidrug transporters are flexible proteins that switch from outward-open to inward-open conformations, facilitating the release of the substrate [13]. Such conformational changes require local flexibility or distortions in the TM regions, which can be provided by specific structural motifs. For instance, our laboratory has shown that serine or threonine, either alone [14] or in combination with proline [15], induces distinctive TM distortions to accommodate the structural needs of specific protein functions [16, 17]. To address this issue we have developed two non-redundant databases of 3D structures of TM segments consisting on α -helix bundles and β -barrels that are accessible through the TMAalphaDB and TMBetaDB web servers, respectively. The main advantage of these servers is their ability to systematically survey sequences of TM regions and provide to the users main structural parameters, such as backbone ϕ and ψ dihedral angles and side chain χ_1 angle, as well as helix bend and twist angles. This structural information allows to quantify distortions induced by residues or motifs and to elucidate their role in the structure and function of membrane proteins.

Construction and content

TMAalphaDB and TMBetaDB are web-based servers that combine a MySQL database management system and Python programs with a dynamic web interface based on PHP.

Non-redundant databases of transmembrane segments structures of alpha and beta membrane proteins

TMAalphaDB and TMBetaDB currently contain 330 structures of α -helix bundles and 107 structures of β -barrels, respectively, with a resolution lower than 3.5 Å. To avoid redundancy, only one structure for each protein is selected (i.e. one structure per UniProt accession code). Among different structures with the same UniProt accession code, the one with best resolution and resemblance to the native state (i.e. without mutations, native pH) is selected. Additionally, for multimeric proteins, only one subunit is extracted. The complete list of structures, together with the unique subunit database, can be downloaded at <http://lmc.uab.cat/TMAalphaDB/info.php> and <http://lmc.uab.cat/TMBetaDB/info.php>. These databases are regularly and automatically updated, in order to include new solved proteins. Each structure is characterized by the Protein Data Bank identification code (PDBID) [18], protein name, UniProt accession code [19], family

name according to Orientations of Proteins in Membranes [20] and organism. Moreover, because the hydrophobic nature of the lipid bilayer conditions the structure and features of the membrane-embedded regions relative to the water-exposed ones [8, 11], we used PDBTM [7, 21] to download only the coordinates of the domain of the protein that is inserted in the lipid bilayer.

Tools to analyse sequence and structure of membrane proteins

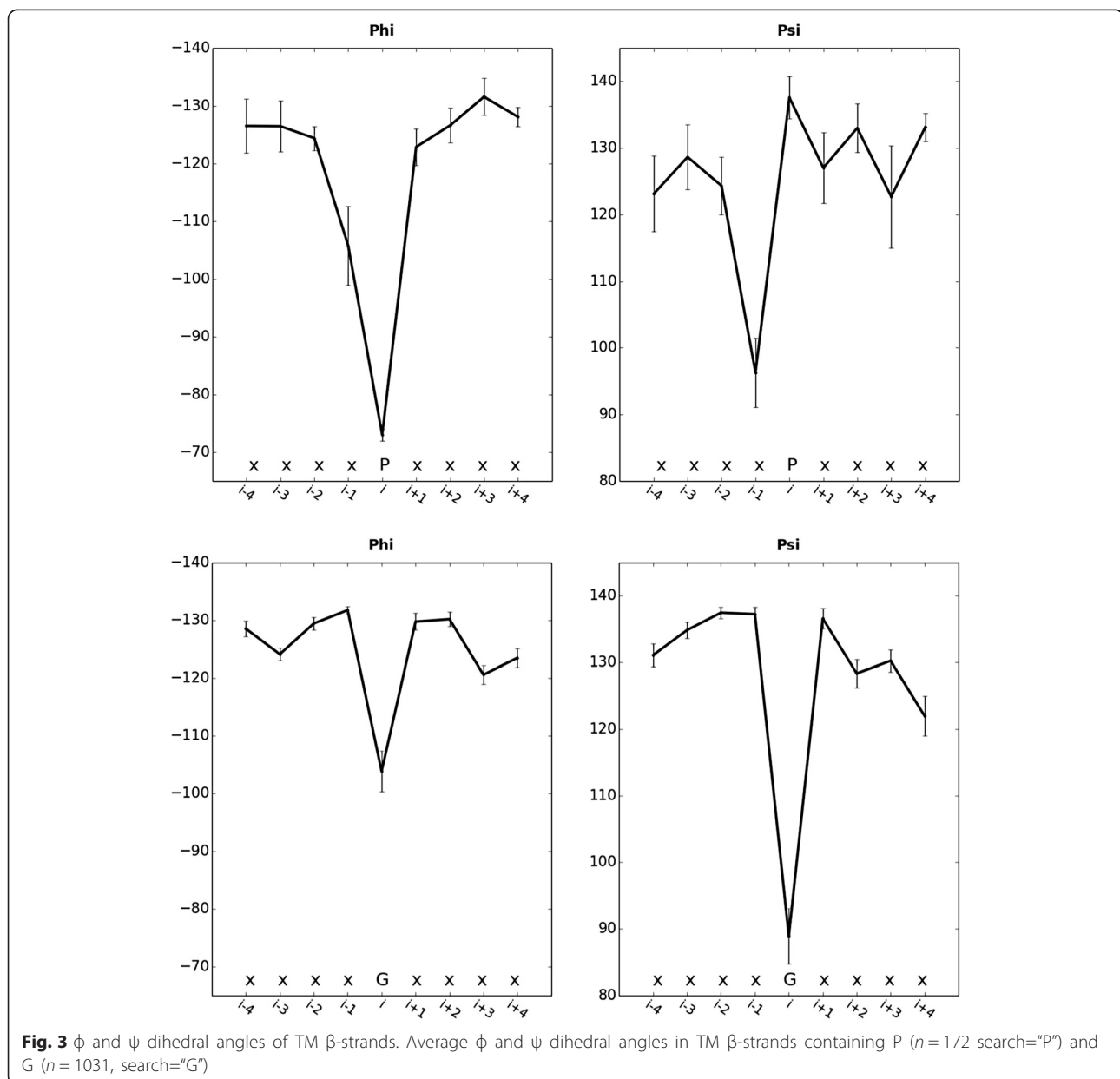
The importance of TMAalphaDB and TMBetaDB is their ability to search and analyse specific residues or sequence motifs in TM segments of membrane proteins. The search is performed using the single-letter amino acid code or/and in combination with wildcard characters such as '+' (positively charged, K/R), '-' (negatively charged, D/E), '*' (charged, K/R/H/D/E), '@' (aromatic, F/W/Y/H), '~' (hydrophobic, I/L/V/M/F/A/P), '^' (polar, D/E/N/Q/K/R/H/S/T/C/W/Y), '%' (aliphatic, L/V/I/M), '#' (distorting, P/G), '?' (hydroxylic, S/T/Y), '\$' (sulphur-containing, C/M), '.' (tiny, G/A), '!' (aromatic amphipathic, W/Y/H), or 'x' (any amino acid). Moreover, the search (advanced options) can be filtered by the proximity of the sequence motif to the beginning/end of the TM domain (as the structural parameters can be highly influenced by the loops) or the presence of certain amino acids within the sequence motif (as, for instance, Pro or/and Gly can distort the secondary structure conformation). The output consists of a list of proteins, identified by the PDBID, UniProt accession code, the name and identifier of the first residue in the motif, the sequence of the TM segment with the requested motif highlighted and the family name of the protein. The coordinates of the entire protein, the TM segments and/or a unique subunit of the protein can be downloaded for each entry. The user can select all TM segments, unique TM segments (i.e. only one TM segments is selected for repeated subunits), or a manually selection can be performed. Average backbone ϕ and ψ angles and side chain χ_1 angle for the selected TM segments can be downloaded or/and displayed in a plot. When all the analysed sequences feature the same type of residue (according to the wildcards previously defined), for a specific position, the plot uses this representation. In TMAalphaDB, bend and twist angles, two relevant parameters to measure local distortions of TM helices, are also calculated and plotted for the identified/selected TM segments using HELANAL [22]. Local bend angles are calculated as the angle between the axes of the cylinders formed by the $C\alpha$ atoms of the residues preceding ($i-3$, i) and following (i , $i+3$) a given amino acid i . Unit twist angles are calculated for sets of four consecutive $C\alpha$ atoms, i.e. one turn, to analyze helical uniformity. An ideal α -helix, with approximately 3.6

flexible point and assisting in helix movements or stabilizing local regions of structural relevance [24]. In order to illustrate the use of TAlphaDB and TMbetaDB, we have surveyed and quantified structural distortions induced by P and PP motifs in TM α -helices and P and G residues in TM β -strands.

P and PP motifs in TM α -helices

Although Pro presents the smallest helix-forming tendency among naturally occurring amino acids [25], Pro residues are often observed in TM helices [26] where they induce a significant distortion. This is produced to avoid a steric clash between the pyrrolidine ring of Pro and the carbonyl oxygen of the

residue in the preceding turn, leading to a bending of the helical structure [27]. Moreover, two consecutive Pro residues are also observed in sequences of membrane proteins. In order to study the distortion induced by the PP motif, relative to P, we scanned TAlphaDB. The search resulted in 349 unique TM helices containing P (search="P") and 8 TM helices containing PP (search="PP"). Figure 1 shows a snapshot of the obtained output for the "P" search, plots for phi and psi dihedrals and unit twist/bend and a Pymol session showing the residues located near "P". The obtained average bend angle plots for P and PP are shown to compare the structural distortion induced by each sequence motif (Fig. 2a). Clearly, the



distortion in bend induced by PP is lower than for P and is not the sum of individual Pro distortions, suggesting a modulating effect.

P and G in TM β -barrels

The cyclic structure of the side chain of Pro locks the ϕ dihedral angle at approximately -60° , which is incompatible with ϕ values near -130° observed in β -strands. We scanned TMbetaDB in order to calculate the backbone ϕ and ψ dihedral angles of Pro when located in β -barrel domains of membrane proteins. The TMbetaDB search resulted in 172 TM segments containing P whose average ϕ and ψ dihedral angles are plotted in Fig. 3. Relative to the energetically preferred ϕ and ψ dihedral angles near -130° and 130° of β -strands, Pro increases ϕ and triggers a decrease in ψ at position $i-1$. In contrast to Pro, the absence of a side chain in Gly allows high flexibility in the polypeptide chain as well as dihedral angles. In order to calculate the observed ϕ and ψ dihedral angles of Gly in β -barrel domains TMbetaDB was scanned. The search resulted in 1031 TM segments containing Gly. Figure 3 shows that, on average, Gly increases ϕ and decreases ψ dihedral angles. These results indicate that both Pro and Gly induce a distortion in the conformation of main polypeptide chain in TM β -strands.

Conclusions

The structural data provided by TAlphaDB and TMbetaDB quantify structural distortions induced by specific amino acids or motifs, and elucidate their role in the structure of membrane proteins. This specific structural information can be, for instance, incorporated in the homology modelling of membrane proteins lacking experimental structure. Thus, these servers emerge as valuable tools to fill the growing gap between the pool of known sequences of membrane proteins and the number of experimentally determined structures.

Availability and requirements

TAlphaDB and TMbetaDB are freely available at <http://lmc.uab.cat/TAlphaDB> and <http://lmc.uab.cat/TMbetaDB>.

Abbreviation

TM: Transmembrane segments; PDBID: Protein Data Bank identification code.

Competing interests

The authors declare that they have no competing interests.

Author's contributions

LP, XD and MO participated in the research design. IL, MP, EM, AC and MO wrote the required computer software. LP, AC, EM and MO contributed to the writing of the manuscript. All authors read and approved the final manuscript.

Acknowledgements

This work is supported by Ministerio de Ciencia e Innovación (SAF2013-48271-C2-2-R) to LP and Instituto de Salud Carlos III (CD09/00150) to AC. We thank Víctor Urrea for his help.

Author details

¹Laboratori de Medicina Computacional, Unitat de Bioestadística, Facultat de Medicina, Universitat Autònoma de Barcelona, Bellaterra, Spain. ²Department de Biologia de Sistemes, Universitat de Vic, Vic, Barcelona, Spain. ³Present address: Condensed Matter Theory Group and Laboratory of Biomolecular Research, Paul Scherrer Institut, Villigen PSI, Switzerland.

Received: 2 March 2015 Accepted: 11 August 2015

Published online: 20 August 2015

References

- Fagerberg L, Jonasson K, von Heijne G, Uhlen M, Berglund L. Prediction of the human membrane proteome. *Proteomics*. 2010;10:1141–9.
- Overington JP, Al-Lazikani B, Hopkins AL. How many drug targets are there? *Nat Rev Drug Discov*. 2006;5:993–6.
- Arinaminpathy Y, Khurana E, Engelman DM, Gerstein MB. Computational analysis of membrane proteins: the largest class of drug targets. *Drug Discov Today*. 2009;14:1130–5.
- Bakheet TM, Doig AJ. Properties and identification of human protein drug targets. *Bioinformatics*. 2009;25:451–7.
- Bill RM, Henderson PJ, Iwata S, Kunji ER, Michel H, Neutze H, et al. Overcoming barriers to membrane protein structure determination. *Nat Biotechnol*. 2011;29:335–40.
- Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissing H, et al. The Protein Data Bank. *Nucleic Acids Res*. 2000;28:235–42.
- Kozma D, Simon I, Tusnady GE. PDBTM: Protein Data Bank of transmembrane proteins after 8 years. *Nucleic Acids Res*. 2013;41:D524–529.
- Olivella M, Gonzalez A, Pardo L, Deupi X. Relation between sequence and structure in membrane proteins. *Bioinformatics*. 2013;29:1589–92.
- Donnelly D, Overington JP, Ruffe SV, Nugent JH, Blundell TL. Modeling alpha-helical transmembrane domains: the calculation and use of substitution tables for lipid-facing residues. *Protein Sci*. 1993;2:55–70.
- Jones DT. De novo protein design using pairwise potentials and a genetic algorithm. *Protein Sci*. 1994;3:567–74.
- Li SC, Deber CM. A measure of helical propensity for amino acids in membrane environments. *Nat Struct Biol*. 1994;1:368–73.
- Rasmussen SG, DeVree BT, Zou Y, Kruse AC, Chung KY, Kobilka TS, et al. Crystal structure of the beta2 adrenergic receptor-Gs protein complex. *Nature*. 2011;477:549–55.
- Masureel M, Martens C, Stein RA, Mishra S, Ruyschaert JM, Mchaourab HS, et al. Protonation drives the conformational switch in the multidrug transporter LmrP. *Nat Chem Biol*. 2014;10:149–55.
- Deupi X, Olivella M, Sanz A, Dolker N, Campillo M, Pardo L. Influence of the g- conformation of Ser and Thr on the structure of transmembrane helices. *J Struct Biol*. 2010;169:116–23.
- Deupi X, Olivella M, Govaerts C, Ballesteros JA, Campillo M, Pardo L. Ser and Thr residues modulate the conformation of pro-kinked transmembrane alpha-helices. *Biophys J*. 2004;86:105–15.
- Sansuk K, Deupi X, Torrecillas IR, Jongejan A, Nijmeijer S, Bakker RA, et al. A structural insight into the reorientation of transmembrane domains 3 and 5 during family A G protein-coupled receptor activation. *Mol Pharmacol*. 2011;79:262–9.
- Boiteux C, Vorobyov I, Allen TW. Ion conduction and conformational flexibility of a bacterial voltage-gated sodium channel. *Proc Natl Acad Sci U S A*. 2014;111:3454–9.
- Bernstein FC, Koetzle TF, Williams GJ, Meyer Jr EF, Brice MD, Rodgers JR, et al. The Protein Data Bank: a computer-based archival file for macromolecular structures. *J Mol Biol*. 1977;112:535–42.
- UniProt. Activities at the Universal Protein Resource (UniProt). *Nucleic Acids Res*. 2014;42:D191–198.
- Lomize MA, Pogozheva ID, Joo H, Mosberg HI, Lomize AL. OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic Acids Res*. 2012;40:D370–376.
- Tusnady GE, Dosztanyi Z, Simon I. PDB_TM: selection and membrane localization of transmembrane proteins in the protein data bank. *Nucleic Acids Res*. 2005;33:D275–278.

22. Bansal M, Kumar S, Velavan R. HELANAL: a program to characterize helix geometry in proteins. *J Biomol Struct Dyn*. 2000;17:811–9.
23. Hanson RM, Prilusky J, Renjian Z, Nakane T, Sussman JL. JSmol and the next-generation Web-based representation of 3D molecular structure as applied to proteopedia. *Isr J Chem*. 2013;53:207–16.
24. Gonzalez A, Cordoní A, Caltabiano G, Campillo M, Pardo L. Impact of helix irregularities on sequence alignment and homology modelling of G protein-coupled receptors. *Chembiochem*. 2012;13:1393–9.
25. O'Neil KT, DeGrado WF. A thermodynamic scale for the helix-forming tendencies of the commonly occurring amino acids. *Science*. 1990;250:646–51.
26. Senes A, Gerstein M, Engelman DM. Statistical analysis of amino acid patterns in transmembrane helices: the GxxxG motif occurs frequently and in association with b-branched residues at neighboring positions. *J Mol Biol*. 2000;296:921–36.
27. Rey J, Deville J, Chabbert M. Structural determinants stabilizing helical distortions related to proline. *J Struct Biol*. 2010;171:266–76.

**Submit your next manuscript to BioMed Central
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

