Biochimica et Biophysica Acta 1858 (2016) 2804–2817

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbamem

Helix perturbations in membrane proteins assist in inter-helical interactions and optimal helix positioning in the bilayer

Ashish Shelar, Manju Bansal *

Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560012, Karnataka, India

ARTICLE INFO

Article history:
Received 20 April 2016
Received in revised form 13 July 2016
Accepted 7 August 2016
Available online 10 August 2016

Keywords:
Helix kink
π-Helix
3_{10}-Helix
Hydrophobic mismatch
Membrane protein modelling
Helix design

ABSTRACT

Transmembrane (TM) helices in integral membrane proteins are primarily α-helical in structure. Here we analyze 1134 TM helices in 90 high resolution membrane proteins and find that apart from the widely prevalent α-helices, TM regions also contain stretches of 3_{10} (3 to 8 residues) and π-helices (5 to 19 residues) with distinct sequence signatures. The various helix perturbations in TM regions comprise of helices with kinked geometry, as well as those with an interspersed 3_{10}/π-helical fragment and show high occurrence in a few membrane proteins. Proline is frequently present at sites of these perturbations, but it is neither a necessary nor a sufficient requirement. Helix perturbations are also conserved within a family of membrane proteins despite low sequence identity in the perturbed region. Furthermore, a perturbation influences the geometry of the TM helix, mediates inter-helical interactions within and across protein chains and avoids hydrophobic mismatch of the helix termini with the bilayer. An analysis of π-helices in the TM regions of the heme copper oxidase superfamily shows that interspersed π-helices can vary in length from 6 to 19 amino acids or be entirely absent, depending upon the protein function. The results presented here would be helpful for prediction of 3_{10} and π-helices in TM regions and can assist the computational design of membrane proteins.

1. Introduction

Helical transmembrane (TM) proteins govern critical and diverse processes in cells [1] and thus form 60% of drug targets [2]. They also represent a large fraction of the available crystal structures that have been solved and are listed in the Mpstruc database (http://blanco.biomol.uci.edu/mpstruc). These proteins are packed in the form of α-helical bundles within the membrane, yet they show wide diversity in their folds, occurring in 46 of the 58 folds listed under the 'Membrane and cell surface proteins and peptides' in the SCOP database [3].

Hydrophobic amino acids constitute a major component of TM helices to minimize the free energy cost required for their insertion within the bilayer [4,5]. Another hallmark of TM helices is the over-representation of proline and glycine in the body of the helix [6,7] These amino acids, along with serine, threonine, asparagine and glutamine are known to alter the intra-helical backbone N—H…O hydrogen bond pattern and cause deviations from regular α-helical character [8,9]. The resulting perturbations are broadly termed as 'kinks' and they either cause a deviation from linearity in the helical path [10] or place a functionally important residue in the required position within the 3D structure [11,12]. Recent analysis by Deane and co-workers also suggests that helix kinks are equally prevalent in globular and membrane proteins [13]. Helix kinks are known to occur in functionally important regions of various helical membrane proteins like GPCRs [12,14] and ion channels [15,16]. Extensive experimental and computational analysis of TM helix kinks has been carried out to understand their amino acid preferences [7,8,12,17,18], helix bending and wobble angles [9,18,19] as also crowd sourcing experiments to identify geometries of helices [20]. The methodologies to characterize kinks using the helical path traced by three dimensional atomic coordinates of the C^{α} atoms of the protein backbone have been explored in depth over three decades [21–24]. Other means of characterization of helix geometries include defining smaller linear helical stretches within a larger fragment to trace the change in the helical path [17]. Recently, a bi-pronged sequence and structure based approach to characterize kinks using neural networks has also been developed [7]. Based on the above mentioned studies, a helix 'kink' can generally be defined as a sharp local bend in the helical body that causes a deviation in backbone torsion angles (φ-ψ), disrupts backbone intra-helical N—H…O hydrogen bonds and changes the helical path. However, helices in membrane proteins are not only 'kinked' but display a wide range of subtle variations in their conformations [18,25] as well as adjustments to accommodate extra amino acids to form α-bulges [26]. The role of

Abbreviations: MaxBA, maximum local bending angle; OPM, orientation of proteins in membrane; MM, main-chain main chain; SM, side-chain main chain; D_{Ori}, distance between local helix origins; Rad, radius.

* Corresponding author at: Lab-109 Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560012, India.
E-mail address: mb@mbu.iisc.ernet.in (M. Bansal).

http://dx.doi.org/10.1016/j.bbamem.2016.08.003
0005-2736/

proline in modulating various types of conformations and perturbed backbone hydrogen bond patterns of α-helices in globular proteins has already been highlighted [27]. However, a detailed and systematic analysis to understand the effect of different perturbations on the local helix parameters, the geometry of the overall TM region and the fold of the membrane protein has not been performed to the best of our knowledge. Due to the role of helix perturbations in maintaining structure [28] and functionality [29] of a membrane protein as well as their importance in evolution [12], an improved understanding would greatly assist computational tools used to model, predict and design structures of membrane proteins [30–32].

Initial structural analysis on helix packing arrangements and interaction motifs within the membrane were performed either on smaller datasets [33] of proteins with modest resolution [34,35]. Later studies that reported structural examination of higher resolution (<3.0 Å) membrane proteins also had only a few structures [36]. However, over the past decade a large number of high resolution membrane protein structures have been deposited in the Protein Data Bank [37] due to the significant advancements in X-ray [38,39] and cryo-electron microscopy techniques [40–42]. Taking advantage of this wealth of membrane protein structures, we have created a sequence-unique high resolution dataset (<2.5 Å) of 90 membrane proteins and performed a systematic analysis of various helix types and perturbations using **A**ssignment of **S**econdary **S**tructure in **P**roteins (ASSP) [43]. TM helices have been shown to occasionally adopt 3_{10}- and π-helical conformations within the bilayer [29,44]. In the present analysis, we report for the first time that apart from the ubiquitous α-helices, TM regions frequently have stretches of 3_{10}- and π-helices, with distinct sequence preferences. We have also characterized the geometry of TM α-helical regions using HELANAL-Plus [21]. In another first, a detailed analysis of TM helices with different geometries and with short 3_{10} and π-helices enables us to group the various helix perturbations into 9 categories each of which show unique trends in twist and rise, upstream and downstream of the perturbation. These perturbations act as key players in influencing the geometry of the TM helix, mediating inter-helical interactions and avoiding the hydrophobic mismatch of the helix termini with the membrane. Hence, they are crucial determinants for oligomerization and proper folding of the membrane protein.

2. Materials and methods

2.1. X-ray crystal structure dataset

A non-homologous dataset of X-ray crystal structures of polytopic helical membrane proteins with resolution better than 2.5 Å and sequence identity <25% was created using the PISCES server [45]. The dataset comprised of 90 proteins with 199 chains. Coordinates for these 90 proteins were downloaded from the Orientation of Proteins in Membrane (OPM) database [46] which aligns the protein structure along the *Z*-axis and also provides the membrane (hydrophobic core) boundaries for the TM region of the protein.

2.2. Secondary structure assignment and helix position nomenclature

The Assignment of Secondary Structure in Proteins (ASSP) program [43] which uses the path traversed by the C^{α} atoms of the protein was used to identify secondary structures in the 90 membrane proteins. ASSP uses a window of four contiguous C^{α} atoms to compute the local step geometric parameters such as helical twist, rise per residue and virtual torsional angle. This window of four C^{α} atoms slides along the length of the helix one C^{α} atom at a time. Uniform stretches within the protein structure are then defined based on these geometric parameters and classified into different secondary structural elements based on the average values of the local parameters (Fig. 1 in [47]). Different types of helices identified by ASSP have been listed in Table 1 along with their median length and values for backbone torsion angles (φ-ψ), helical twist and rise per residue.

Amino acid propensities for 15 positions (9 helical and 6 near-helical) in α-helices belonging to membrane proteins have been studied in detail previously [48]. Since majority of 3_{10}- and π-helices helices are 4 or 5 amino acids in length, residue propensities have only been examined for five helical positions (N1, N2, Mid, C2, C1) and two near-helical positions (N′, Ncap, C′, Ccap) at each termini. N1 and C1 represent the first and last helical positions within the main helix body. The mid position represents all the residues, after excluding the two terminal positions at each end of the helix. Of the sequence propensities analyzed for 15 positions in α-helices previously [48], preferences have been shown only for the above mentioned 9 positions.

2.3. Programs used

HELANAL-Plus [21], a program that uses least square 3D line and sphere fitting to local helix origin points was used for assigning geometries to the helical segments. MolBridge [49] was used with default cut-off values for calculation of non-bonded interactions.

Identification of closely related sequences was carried out using BLAST [50] against a non-redundant protein database. Multiple sequence alignment of protein sequences was performed using ClustalΩ [51].

2.4. Distribution of amino acids and position-wise propensity

Distribution of amino acids was computed for helical and near-helical positions in 3_{10}- and π-helices. Positionwise propensity (P_{ij}) for amino acids to occur at the helical and near-helical positions was calculated using the following formulae:

$$P_{ij} = (n_{ij}/n_i)/(N_j/N)$$

where:

n_{ij}	number of amino acids 'i' at position 'j'
n_i	total number of amino acids of type 'i' in 90 membrane proteins
N_j	number of amino acids at position 'j'
N	total number of amino acids in 90 membrane proteins.

Table 1
Different types of helices defined by ASSP along with their parameters in 90 membrane proteins containing 199 chains.

			Backbone torsion angles		Helical parameters	
Type of helix	Number	Median length	φ (°)	ψ (°)	Twist (°)	Rise (Å)
α-Helix	1704 (68.2%)	13	−62.2 (15.6)	−41.2 (13.7)	99.1 (2.5)	1.5 (0.3)
3_{10}-Helix	523 (21%)	3	−69.5 (18.3)	−29 (11.4)	105.3 (5.4)	1.6 (0.2)
π-Helix	129 (5%)	6	−80.7 (21)	−33.4 (22.6)	85.3 (8.2)	1.1 (0.2)
Poly proline II	143 (5.8%)	3	−75 (16.2)	142 (11.5)	234.2 (3.7)	3.2 (0.1)

The standard deviation values are given in parentheses. Calculation of local step helical parameters and representative examples of each helix type have been depicted in Figs. 1 and 2 in [47] respectively.