doi:10.1016/j.jmb.2006.09.020 J. Mol. Biol. (2007) 366, 436–448

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

E_z, a Depth-dependent Potential for Assessing the Energies of Insertion of Amino Acid Side-chains into Membranes: Derivation and Applications to Determining the Orientation of Transmembrane and Interfacial Helices

Alessandro Senes†, Deborah C. Chadi†, Peter B. Law Robin F. S. Walters, Vikas Nanda and William F. DeGrado*

Deparment of Biochemistry and Molecular Biophysics School of Medicine University of Pennsylvania Philadelphia, PA 19104-6059 USA

**Corresponding author*

We have developed an empirical residue-based potential (E_z potential) for protein insertion in lipid membranes. Propensities for occurrence as a function of depth in the bilayer were calculated for the individual amino acid types from their distribution in known structures of helical membrane proteins. The propensities were then fit to continuous curves and converted to a potential using a reverse-Boltzman relationship. The E_z potential demonstrated a good correlation with experimental data such as amino acid transfer free energy scales (water to membrane center and water to interface), and it incorporates transmembrane helices of varying composition in the membrane with trends similar to those obtained with translocon-mediated insertion experiments. The potential has a variety of applications in the analysis of natural membrane proteins as well as in the design of new ones. It can help in calculating the propensity of single helices to insert in the bilayer and estimate their tilt angle with respect to the bilayer normal. It can be utilized to discriminate amphiphilic helices that assume a parallel orientation at the membrane interface, such as those of membrane-active peptides. In membrane protein design applications, the potential allows an environment-dependent selection of amino acid identities.

Keywords: membrane proteins; protential; hydrophobicity; membrane insertion; transmembrane helices

Introduction

The primary amino acid sequence of membrane proteins directs the proper positioning of the helices in the lipid bilayer following insertion *via* the translocon apparatus. Thus, considerable effort has been made to determine the thermodynamics of transfer of the amino acid side-chains from water to various regions of the membrane.[1] Schematically, the bilayer is generally considered to consist of distinct sectors representing the hydrophobic hydrocarbon core, a polar headgroup region, and the extra-membrane aqueous region. The free energies of transfer of amino acid derivatives from water to the headgroup or hydrocarbon regions of a bilayer have been determined using a variety of experimental techniques.[1–3] Also, amino acid sequence data for single-span and multi-span helical membrane proteins have been evaluated to determine the propensity of various amino acids to form helices at distinct sectors of the bilayer.[4–7]

Although the multi-sector approach provides a good first approximation for the bilayer environment, it would be advantageous to determine the free energy profile for each residue type as a continuous function of its depth of insertion within a membrane. Here, we determine the frequency of occurrence of various amino acid types as a function of their position in a bilayer. Using the reverse-Boltzmann statistical approach, these propensities are converted to pseudo-energies, which are shown

† A.S. and D.C.C. contributed equally to this work.

E-mail address of the corresponding author: wdegrado@mail.med.upenn.edu

0022-2836/$ - see front matter © 2006 Published by Elsevier Ltd.

to correlate well with experimental free energies of transfer of side-chains from water to distinct regions of the bilayer. The resulting method is used to predict the locations and orientations of helices in membrane proteins and has a variety of applications in the analysis of natural membrane proteins as well as in the design of novel structures. For example, there is keen interest in determining the extent to which polar side-chains, particularly Arg, can be accommodated stably within the bilayer, particularly as this relates to the mechanism of voltage-sensing in potassium channels,[8–10] as well as the mechanism of membrane disruption by cationic membrane-permeabilizing and antimicrobial peptides.[11–14] The energy function can help to define the orientation of a protein in the membrane, which cannot be inferred directly from a crystal structure because the positions of the phospholipids are generally not defined, although isolated lipids are visible in the electron density in some cases. Given the large amount of available sequence data, another important application is the estimation of the propensity for insertion of predicted transmembrane domains and their preferred angle with respect to the bilayer normal. Finally, a computationally efficient empirical energy function is useful in membrane protein design to guide the selection of specific side-chains at various positions of the peptide chain as a function of their environment.[7]

Results and Discussion

Depth-dependent propensity profiles for amino acid side-chains

To determine the propensities of the amino acid side-chains across the bilayer depth, we analyzed previously studied helical membrane proteins positioned in an implicit bilayer. Ideally, we would consider the cytoplasmic and extracellular ends of the helix differently, and would additionally discriminate the N-terminal ends from the C-terminal ends of the helices (because the residues have different rotamer distributions on either end of the helix).[15,16] However, to maximize the signal-to-noise ratio, these features were not differentiated, and the distance of the residues C^{β} (C^{α} for Gly) from the bilayer center was measured.

Figure 1 illustrates the propensity for the side-chains to occupy consecutive 2 Å regions beginning at the center of the bilayer, and extending to 30 Å along a line that is normal to the bilayer plane. The z coordinate defines the distance from the center of the bilayer. We define the position-dependent propensity (P_z) for a given residue as the number of residues of this type in each 2 Å increment, divided by the mean value for all 2 Å increments for the entire 30 Å region considered. At a qualitative level, the resulting distributions are consistent with conventional wisdom concerning the partitioning of substances between apolar and polar environments.[1] The hydrophobic residues (Figure 1(a)) have the highest propensity to occur near the center of the bilayer, and the polar residues (Figure 1(b)) show the highest propensities to occur on the exterior of the bilayer ($z>20$ Å). Pro, which behaved similarly to the polar residues, has a hydrophobic side-chain but in a transmembrane helix it often causes one or more main chain hydrogen bonding groups to be unsatisfied, which are thermodynamically unfavorable in the non-polar core of the membranes.[17] While it can be accommodated easily in a helix when embedded deep within a membrane,[8,18] Pro is also the amino acid with the highest propensity to promote helical hairpin formation when present near the headgroup region.[19,20]

Aromatic residues with a single polar group (Tyr and Trp, Figure 1(c)) have a high tendency to locate at the headgroup region, whereas, as expected, Phe behaves similarly to the other hydrophobic amino acids.[21,22] Finally, relatively flat profiles are observed for the small residues Gly, Ser, Thr and Cys (Figure 1(d)), which are mildly polar but can be accommodated favorably also in apolar environments.[23]

A continuous z-dependent potential can be derived from the amino acid propensities, which allows the calculation of pseudo-energies of insertion of a protein as a function of conformation, orientation, and depth in the membrane bilayer. The relationship of the amino acid propensities with energy can be expressed as:

$$P_z = P_{\mathrm{aq}} \mathrm{e}^{-\frac{\Delta E_z}{RT}} \tag{1}$$

where P_z is the propensity as a function of the depth coordinate, z; P_{aq} is the propensity in pure water when z is extrapolated to infinity, and ΔE_z is the energy difference between the residue in water and at a given depth within the bilayer as defined by its z coordinate; R is the gas constant and T is the absolute temperature.

It is important to choose a functional form of ΔE_z to account for the differences in the free energy of solvation of the various residues at different portions of the bilayer. With the exception of Trp and Tyr, the P_z *versus* z distributions tend to be sigmoidal, with varying degrees of steepness. This behavior is captured well by the equation:

$$\Delta E_z = \frac{\Delta E_0}{1 + \left(\frac{z}{Z_{\mathrm{mid}}}\right)^n} \tag{2}$$

where ΔE_0 is the pseudo-energy difference between center of the bilayer ($z=0$) and water (the standard state, $z=\infty$) for a residue; Z_{mid} is the z coordinate at which the energy is half-maximal; and n is a parameter that defines the steepness of the transition.

Equation (2) was inserted into equation (1) and non-linear least-squares analysis was used to obtain the values of P_{aq}, ΔE_0, Z_{mid}, and n (Table 1), leading