1 Helix-Helix Packing Between Transmembrane Fragments

Mar Orzáez, Francisco J. Taberner, Enrique Pérez-Payá and Ismael Mingarro

Abstract

The rules that govern the folding of membrane proteins are still not completely understood when compared with the well-detailed set of principles described for the folding of soluble proteins. Although the molecular determinants of the folding mechanism should be basically the same for both types of proteins, the main difference, which in turn could also represent the main difficulty, is the media that surrounds the protein. In this sense, it would be useful to develop new molecular techniques that allow for the study of the folding mechanism of membrane proteins in their natural media, the membrane. In the present work we have collected a series of studies devoted to understanding the principles that govern the molecular mechanism of folding and packing of membrane proteins, an important group of the proteome. In particular, we have focused our attention on Glycophorin A (GpA), a single span membrane protein, which has been extensively used as a model system to study helix-helix transmembrane (TM) packing. Here, we report on the importance of the molecular distance between the critical oligomerisation motif and polar residues in TM fragments. We have also given an account of the influence of the length of the TM fragment as well as the effect of proline residues on the packing of TM alpha helices.

1.1 Introduction

It is estimated that membrane proteins constitute about 30 % of fully sequenced genomes– a major part of biological life (Wallin and von Heijne, 1998). These types of protein play an important role in cell function, acting as cell receptors, transporters, channels and as essential components of respiratory and photosynthetic complexes. Furthermore, membrane proteins are

Principles and Practice
Methods in Proteome and Protein Analysis
R.M. Kamp, J. J. Calvete, T. Choli-Papadopoulou (Eds.)

also implicated in essential processes inside the cell. They have attracted the interest not only of basic, but also of more applied fields like medical chemistry, due to their potential as targets for therapeutic intervention. However, due to their hydrophobic character, the fine characterisation of different aspects of the folding, membrane insertion and structure-function relationships of membrane proteins has been delayed more than expected. For example, very few membrane protein structures are known at atomic resolution (~20 out of over ~20,000 known structures), which is a major stumbling block in our understanding of how they function, how to correct their malfunctions and how to exploit them. In soluble proteins, the knowledge of protein structures paralleled with the definition of the molecular mechanisms that govern protein folding has permitted a better understanding of protein function. This has facilitated the rational design of new pharmaceuticals directed to selected target proteins. It is reasonable to believe that the same approach could be extended to membrane proteins. Thus, there is increasing interest in the definition of the molecular mechanism that drives the folding of proteins in the membrane environment, and in particular in the rules that permit TM segments to be energetically stabilised through packing interactions.

When studying membrane proteins one has to clearly realise that these proteins "*live*" in an environment completely different from the aqueous media, the membrane. The cell membrane is a highly heterogeneous media, composed mainly of phospholipids that are self-organised in two leaflets giving rise to the formation of a bilayer. The hydrocarbon core (HC) is the hydrophobic part of the membrane that is approximately 30 Å thick. The polar heads of the phospholipids define the lipid/water interphases (IF) and add 15 Å to the thickness of each leaflet. It is this complex environment, with physical and chemical properties different from aqueous media, into which membrane proteins are integrated.

All membrane protein structures solved to date show that TM domains fold as either α-helices or β-strands, due to the physical and chemical constraints imposed by the hydrophobic environment (White and Wimley, 1999). The α-helical-type proteins are most abundant and can be made up of a single helix or of multiple helices packed together in bundles. The folding of constitutive α-helical membrane proteins has been conceptualised, in its simplest form, as a two-stage process (Fig. 1.1), in which the helices are first independently formed across the membrane and then laterally assembled to form the native protein (Popot and Engelman, 1990). The formation of the individual helices is a consequence of main-chain hydrogen bonding, as the hydrophobic effect of the lipid bilayer has an influence, restraining other unordered peptide structures that would expose polar peptide bonds (reviewed in (White *et al.*, 1998)). The side-to-side helix association or protein assembly, the second stage, is driven by different interactions, like van der Waals forces, electrostatic effects, steric clashes, or differential effect of asymmetrically distributed lipids (Popot and Engelman, 2000; White and Wimley, 1999).

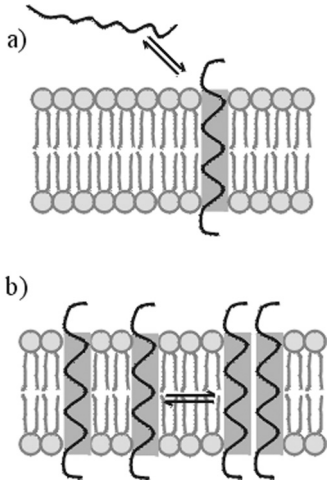

Fig. 1.1. The two-stage model. **a** The first stage, insertion of the prefolded helix in the lipid bilayer. **b** The second stage, lateral association between inserted *TM* helices to produce the folded structure

These theoretical principles have been, to some extent, demonstrated by means of membrane spanning synthetic peptides that become folded in the membrane media and by experimental systems that allow the study of membrane multispanning helices. However, the main problem is to obtain detailed information on protein structures. This is due, in part, to the loss of structural stability of the proteins outside their natural media, and on the extreme complexity that the membrane introduces when classical methods of protein structure analysis are contemplated (DeGrado et al., 2003). Thus, the development of new, or the improvement of "classical", strategies to study membrane proteins at the atomic level is one of the key areas in structural biology. In this sense, the use of well-characterised model proteins could allow the design of experimental techniques directed to the study of complex questions still to be solved regarding membrane proteins. One of the best-suited models of a membrane protein that oligomerises (more specifically, dimerises) through interactions of its TM α-helices is undoubtedly Glycophorin A (Arkin, 2002; Lemmon and Engelman, 1994).

1.2 Glycophorin A as a Model System

Glycophorin A (GpA) was the first membrane protein whose sequence was determined. A long stretch of hydrophobic amino acid residues was identified from these studies (Tomita and Marchesi, 1975). Furthermore, GpA has provided the first clear example of non-covalent membrane protein oligomerisation due to specific interactions of its TM α-helices (see Bormann and Engelman, 1992 for a review). The free energy decrease associated with the dimerisation process is enough to confer dimerisation of an unrelated water-

soluble protein when fused to the TM domain of GpA to form a chimeric protein (Lemmon et al., 1992a). Moreover, the addition of a His-tag at the extreme C-terminus of this fusion for purification purposes did not perturb TM association (Mingarro et al., 1996).

The wide use of this protein as a model membrane protein is also based on its intrinsic simplicity, since its single TM fragment drives a detergent resistant homodimerisation of the protein. Thus, the dimerisation process and those factors that could affect or modify it can be analysed using SDS-PAGE. The GpA homodimer, defines a dimerisation interface that has been extensively studied by diverse techniques such as saturation mutagenesis (Lemmon et al., 1992b), alanine-insertion scanning (Mingarro et al., 1996), computational modelling (Adams et al., 1996), solution NMR in dodecylphosphocholine micelles (MacKenzie et al., 1997) and solid-state NMR in lipid membranes (Smith *et al.*, 2001). The output of these studies describes a dimerisation motif in the TM fragment composed of seven residues, L^{75}IxxGVxxGVxxxT^{87}, which is responsible for the dimerisation process (Fig. 1.2).

In the present study, we have focused on three major factors that have an influence on the molecular mechanism of helix-helix packing in a membrane-like environment. First, we analysed how important it is to keep a minimum distance between the dimerisation motif and the flanking charged residues on the cytoplasmic side of the protein. Secondly, we described the minimisation of the GpA dimerisation motif that allows dimer formation in homogeneous poly-leucine stretches of different length. Finally, we would like to achieve an experimental rationalisation for the observed and unexpected over-representation of proline residues in TM fragments as well as their role in membrane protein assembly.

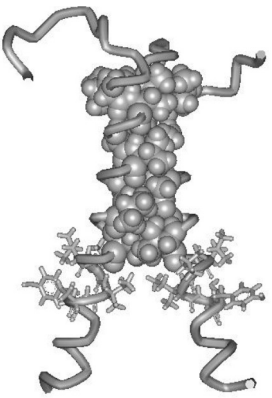

Fig. 1.2. Ribbon drawing derived from the *GpA TM* helix of the experimental *NMR* structure (PDB file 1AFO). The seven critical interface residues are shown in space-filling mode, residues Leu89-Tyr93 in ball-and-sticks mode. The drawing was generated using the program WebLab Viewer Pro 3.7 (Molecular Simulations, San Diego, California)

1.3 Influence of the Distance Between the Dimerisation Motif and the Flanking Charged Residues on the Packing Process Between TM Helices

The presence of positive charged residues at the cytoplasmic side of TM fragments is a feature of many membrane proteins. In most organisms the orientation of TM fragments in the membrane seems to be influenced by the charge distribution flanking the hydrophobic core of the TM segment. Positively charged amino acid residues often direct a charged domain to remain on the cytosolic side of the membrane (the side of protein synthesis). This experimental observation is formally known as the "*positive-inside* rule" (von Heijne, 1992).

Glycophorin A has at its C-terminus, adjacent to the TM fragment, arginine and lysine residues that corroborate the general rule. In particular, the TM fragment of GpA has a stretch of basic residues located eight residues downstream from the last amino acid of the dimerisation motif (Thr87). Although the residues included in the hydrophobic stretch (Ile88-Ile95) are not directly implicated in the net of van der Waals interactions that maintain the dimeric structure of this protein, deletions in this region (mutants Δ89/93, Δ89/91and Δ91/93, for example) abrogate dimer formation, (see Table 1.1). These last three constructs all contain a deletion of the amino acid residue Ile91, located at the contact interface of the two TM helices (Fig. 1.2). However, a mutant protein with a point deletion of this residue (Δ91) dimerised as efficiently as the wild type (Table 1.1), indicating that Ile91 is probably not specifically

Table 1. Dimerisation capacity of GpA C-terminal mutants

Mutant	Sequence	Dimer (%)
Wt	^{72}EITLIIFGVMAGVIGTILLISYGIRRLIKK^{101}	84
Δ89/93	^{72}EITLIIFGVMAGVIGTI---GIRRLIKK^{101}	0
Δ89/91	^{72}EITLIIFGVMAGVIGTI--SYGIRRLIKK^{101}	0
Δ91/93	^{72}EITLIIFGVMAGVIGTILL--GIRRLIKK^{101}	9
Δ91	^{72}EITLIIFGVMAGVIGTILL-SYGIRRLIKK^{101}	86
91–93L	^{72}EITLIIFGVMAGVIGTILLLLLGIRRLIKK^{101}	78
91–95L	^{72}EITLIIFGVMAGVIGTILLLLLLLRRLIKK^{101}	77
Δ89/91,Δ96/97	^{72}EITLIIFGVMAGVIGTI--SYGI–LIKK^{101}	2
Δ89/91,96/97L	^{72}EITLIIFGVMAGVIGTI--SYGILLLIKK^{101}	2
Δ89/91,96/97L,100/101L	^{72}EITLIIFGVMAGVIGTI--SYGILLLILL^{101}	73
Δ91,Δ96/97	^{72}EITLIIFGVMAGVIGTILL-SYGI–LIKK^{101}	0
Δ91,96/97L	^{72}EITLIIFGVMAGVIGTILL-SYGILLLIKK^{101}	45
Δ96/97	^{72}EITLIIFGVMAGVIGTILLISYGI–LIKK^{101}	69
96/97L	^{72}EITLIIFGVMAGVIGTILLISYGILLLIKK^{101}	74
96/97L,100/101L	^{72}EITLIIFGVMAGVIGTILLISYGILLLILL^{101}	55