4 April 2018

Influence of proline residues in transmembrane helix packing

Mar Orzáez, Jesús Salgado, Ana Giménez-Giner, Enrique Pérez-Payá and Ismael Mingarro*

Departament de Bioquímica i Biologia Molecular

Universitat de València

E-46 100 Burjassot, Spain

Running title: Prolines in transmembrane packing

Text: Microsoft Word 2001 for Mac (29 pages)

Figures included: 6

* Corresponding author: I. Mingarro
Phone: Int+34-96-354 3796
Fax: Int+34-96-354 4635
E-mail address of the corresponding author: Ismael.Mingarro@uv.es

Current address: E. Pérez-Payá, FVIB, Amadeo de Saboya 4, València. E-46010

Abstract

Integral membrane proteins often contain proline residues in their α-helical transmembrane (TM) fragments, which may strongly influence their folding and association. Pro-scanning mutagenesis of the helical domain of glycophorin A (GpA) showed that replacement of the residues located at the center abrogates helix packing while substitution of the residues forming the ending helical turns allows dimer formation. Synthetic TM peptides revealed that a point mutation of one of the residues of the dimerization motif (L75P) located at the N-terminal helical turn of GpA TM fragment, adopts secondary structure and oligomeric state similar to the wild type sequence in detergents. In addition, both glycosylation mapping in biological membranes and molecular dynamics showed that the presence of a proline residue at the lipid/water interface has as an effect the extension of the helical end. Thus, helix packing can be an important factor that determines appearance of proline in TM helices. Membrane proteins might accumulate proline residues at the two ends of their TM segments in order to modulate the exposition of key amino acids at the interface for molecular recognition events while allowing stable association and native folding.

Key words

Transmembrane helix, helix packing, proline, protein structure, glycophorin A

Abbreviations

DPC: dodecyl phosphocholine; ER endoplasmic reticulum; GpA: Glycophorin A; Lep: leader peptidase; MDG: minimal glycosylation distance; OST: oligosaccharyl transferase; SDS-PAGE: sodium dodecylsulfate polyacryamide-gel electrophoresis; TM: transmembrane; Wt: wild type;

Introduction

The transmembrane (TM) segments of integral membrane proteins are embedded in a phospholipid bilayer, where the environment strongly limits the range of possible structures [1]. All membrane protein structures solved to date show that TM segments fold either as α helices or β strands, due to the physical and chemical constrains imposed by the hydrophobic environment [2]. The α helical-type TM proteins are most abundant and can be made up of a single helix or of multiple helices packed in bundles.

The folding of constitutive α-helical membrane proteins has been conceptualized, in its simplest form, as a two-stage process, in which the helices are first independently formed across the membrane and then laterally assembled to form the native protein [3]. The formation of individual helices (the first stage) is mainly a consequence of main-chain hydrogen bonding and the hydrophobic effect of the lipid bilayer, and it has been explored by studying the membrane partitioning-folding coupling of fragments of membrane proteins, i.e. peptides (reviewed in [4]). Concerning the side-to-side helix association (the second stage), other interactions, like van der Waals forces, electrostatic effects, steric clashes, or differential lipid effects must dominate membrane protein assembly [2,5]. Although there are several studies addressing this latter feature, in some of them is difficult to separate the intrinsic helix-helix interactions from non-TM contributions, like those involving prosthetic groups [6,7] or the loops

connecting TM helices [8,9]. One of the few experimental models that allow overcoming these difficulties is the study of the homodimerization process of the single-helix TM protein glycophorin A (GpA), which has permitted a detailed examination of intrinsic helix-helix interactions (reviewed in [10,11]). The GpA homodimer defines a dimerization interface that has been extensively studied by diverse techniques, such as saturation mutagenesis [12], alanine-insertion scanning [13], computational modeling [14], solution NMR in dodecyl phosphocholine micelles [15] and solid-state NMR in lipid membranes [16]. The output of these studies describes a dimerization motif in the TM fragment composed of seven residues, $L^{75}I^{76}xxG^{79}V^{80}xxG^{83}V^{84}xxT^{87}$, that is responsible of the dimerization process. The study of this motif as a model for helix-helix packing turned to be especially meaningful since statistical analysis of amino acid paterns in TM helices highlighted the importance of the GxxxG motif (in association with β□□□□nched residues) in helix-helix interactions [17].

Knowledge of the principles of TM helix packing has grown in the last few decades, increasing our capacity to identify and test the role of potential TM associations (recently reviewed by [18]). However, there are still important aspects that remain to be addressed to fully understand both the insertion and packing of TM α-helices. One of these issues is the presence of proline residues in these idiosyncratic protein domains.

The cyclic structure of proline makes it unique among the 20 naturally occurring amino acid residues because its amide group lacks the proton necessary for hydrogen bond stabilization of regular secondary structures. Nonetheless, while not abundant, this residue has been observed to be frequently present in TM helices [19,20]. This contrasts with the long time accepted role of proline as a helix-breaker in water-soluble proteins [21]. Hence, a substantial number of proline residues are found in TM helices, where their most common effect is to create a bend in the helix axis [22,23], although the bend can be attenuated by local conformations of the main chain resulting in some cases in a straight helix [24].

Helicity of these proline-containing hydrophobic sequences has been evidenced mainly by the analysis of synthetic peptides in the presence of detergents or incorporated into liposomes [19,25,26], and by means of molecular approaches in natural membranes [27]. Taking this into account, the effect of proline residues in membrane protein folding should rely on its capacity to perturb helix packing more than the helicity of the region itself.

In the present study we addressed the consequences of replacing specific residues by proline all along the helical region [15] of the GpA TM fragment on the dimerization of this model membrane protein. Our results showed that the presence of proline at several positions of the TM segment still allows helix-helix association. In particular, the replacement of one of the previously defined

seven key residues for dimerization, rendered a chimeric protein that displayed dimer formation. An analysis by glycosylation mapping in endoplasmic reticulum membranes showed a partially extended conformation in this region of the dimerization domain. Finally, our experimental results were rationalized by means of molecular modeling.

Results

Proline-scanning mutagenesis

Pro-scanning mutagenesis was envisioned as a strategy to study the conformational effect of proline residues in the packing of TM helices. To this aim, every residue of the helical domain [15] of GpA was individually substituted by proline, and dimerization was determined for each mutant by SDS-PAGE. This domain drives a SDS-stable, noncovalent homodimerization of the protein [28], providing a convenient model for the analysis of membrane protein folding.

Fig. 1 shows that proline is easily tolerated in dimer formation when replacing residues located at both the N- and C-terminal ends of the helix, roughly at the first helical turn. A first interpretation of these results is that substitutions in the central part of the helix, where the dimerization motif is located, abrogates dimer formation by perturbing the described ridge-into-groove arrangement [15,29]. Conversely, mutations in the first helical turn at both ends of the TM segment would leave the dimerization motif in proper orientation for dimer formation.

However, the case of the L75P mutant deserves a special attention. It has been demonstrated that the dimerization motif of GpA when grafted on a polyleucine stretch is $L^{75}I^{76}xxG^{79}V^{80}xxG^{83}V^{84}xxT^{87}$ [30]. In a seminal exhaustive saturation mutagenesis study [12], it was shown that the replacement of Leu75 by relatively hydrophobic residues like cysteine, methionine or phenylalanine still

tolerates significant dimerization of the chimeric protein, whereas the presence of tryptophan or valine at this position completely abolishes helix-helix association. In this respect, since the proline side chain has a marked hydrophobic character, it is not that surprising that our L75P mutant showed a significant amount of dimer (Fig. 1).

Secondary structure and oligomeric state of TM peptides in detergents

Although the presence of proline residues normally compromises regular secondary structure formation, the dimerization degree observed for the L75P mutant should be compatible with α-helical structures. In order to test this hypothesis peptides containing wild type (Wt) and mutant (L75P) GpA TM sequences were chemically synthesized using previously described protocols (see Materials and Methods).

The secondary structure adopted by both synthetic peptides was evaluated by circular dichroism (CD) spectroscopy in micellar SDS, a solvent system that mimics the natural environment of TM sequences [31], and dodecyl phosphocholine (DPC), the micellar phase previously used to obtain the NMR structure of the GpA TM domain [15]. As seen in Fig. 2(a), only small differences were found between the Wt and L75P peptides in each membrane mimetic environment. The lowered helical conformation adopted by the two synthetic TM peptides in DPC micelles when compared to SDS micelles, could be attributed to a frayed effect of the positively charged Lys-tags in the zwitterionic

detergent (DPC) with respect to the negatively charged SDS micelles. Commonly used deconvolution algorithms of CD spectra [32] indicate over 80% helical content for both the Wt and L75P peptides in SDS micellar media.

The $\Theta_{220\ nm}/\Theta_{208\ nm}$ ratio has been used in CD spectroscopy to assess whether α-helices are implicated in coiled-coil motifs. In our case, both peptides showed a ratio close to 1 for the two micellar media, which is the value proposed for a two stranded α-helical coiled-coil [33,34], in agreement with a dimeric conformation. In addition, the very high degree of helical secondary structure in this media was maintained between 15 and 90 ºC for both peptides (Fig. 2(a), inset), indicating that they are similarly stable. However, as pointed out previously [35], it is possible that temperature denaturation of membrane proteins in a micelle or lipid bilayer could perturb helix packing without significantly affecting the secondary structure. The inexistence of any isodichroic point in our thermal denaturation experiments and the maintenance of a $\Theta_{220\ nm}/\ \Theta_{208\ nm}$ ratio of ~1 for both peptides within the studied temperature range, suggest subsistence of the homodimer.

The ability of these peptides to self-associate in micelles was assessed by tricine SDS-PAGE. This method has been recently used to examine the association of TM peptides [36-39] since SDS detergent micelles mimic the membrane environment. As shown in Fig. 2(b), both the Wt (as previously reported [37]) and L75P peptides migrate as a single band with a mobility similar to that expected for a dimer in a concentration independent manner (not

shown). The small electrophoretical mobility alteration observed between the two dimers should not be adscribed to differences in the specificity of their interhelical packing, since an equimolecular mixture of both peptides migrates as a single band with an intermediate mobility (Fig. 2(b), right lane).

To determine whether dimerization of L75P results from molecular interactions similar to those that mediate dimerization of natural GpA, competition experiments, where synthetic Wt or L75P peptides are mixed with wild type or mutant proteins, were performed. For all combinations, synthetic TM peptides disrupted protein association to a significant extent, generating the concomitant peptide–protein heterodimers, and indicating that interactions between the wild type and the mutant sequences are specific and closely related (Fig. 3). A quantitative analysis provided apparent dissociation constants of (12±2) $x10^{-6}$ M and (10±4) x 10^{-6} M for the wild type protein/L75P peptide and mutant L75P protein/L75P peptide complexes, respectively. When both the Wt and L75P mutant protein were incubated with the Wt peptide at very high peptide concentrations, an additional (less abundant) band appeared above the one corresponding to the heterodimer, interpreted as heterotrimers composed of one protein and two Wt peptides (Fig. 3(c),(d)). It should be mentioned that a homotrimer of the Wt peptide was not observed in the tricine SDS-PAGE analysis (Fig. 2(b) and [37]) and that similar heterotrimers have been observed using a hydrophobic synthetic peptide and chimeric proteins, closely related to ours, comprising designed TM helices [35]. Although this apparent tendency to

form heterotrimers complicates the quantitative analysis, we were able to estimate the apparent dissociation constants for the formation of heterodimers at concentration conditions where the heterotrimers accounted for less than 10% of the total mass amount. In those particular conditions we obtained values of (5±1) x 10^{-6} M and (11±1) x 10^{-6} M for wild type protein/Wt peptide and mutant L75P protein/ Wt peptide complexes, respectively. The estimated apparent dissociation constants showed affinities that are in all cases within the same order of magnitude, although higher for the wild type sequences, as expected.

Helix-length modulation effects of proline residues in the endoplasmic reticulum membrane

In order to test the molecular effect of the proline residue in the L75P mutant in biological membranes we have used a glycosylation mapping technique [40]. The basic idea behind this approach is that the endoplasmic reticulum (ER) enzyme oligosaccharyl transferase (OST) can transfer a glycosyl moiety to an acceptor Asn residue in a nascent membrane protein only when the Asn-Xxx-Thr/Ser (being Xxx any amino acid but Pro) acceptor site is placed a minimum number of residues away from the lumenal end of a TM fragment, the so called "minimum glycosylation distance" (MGD). This technique has been successfully used previously to study the helix-breaking effects of proline residues in TM helices of different lengths and orientations [27]. To assess the effect of proline

replacement on position 75 of the GpA TM fragment we inserted this hydrophobic sequence in place of the second TM fragment of the well-characterized *Escherichia coli* inner membrane protein leader peptidase (Lep). Although of bacterial origin, Lep integrates efficiently into dog pancreas microsomes with the same topology (Fig. 4(a)) as in *E. coli*[40,41].

As shown in Fig. 4(b) the MGD for the wild type sequence of GpA was found to be 11 residues, while replacement of Leu75 by proline allowed efficient glycosylation at a smaller number of residues. These results indicate that the presence of the proline in the upper most turn of the helical segment enlarges the distance to the membrane of the acceptor site, probably by changing this turn of the TM α-helix into a more flexible, extended conformation. In a fully extended chain, the amino acid residues are staggered, so the linear dimension of a polypeptide stretch with *n* residues can be considered as *n* times ~3.3Å, while this distance for the same number of residues in a canonical α-helix would be *n* times ~1.5Å (being 3.3Å and 1.5Å the rise per residue for extended and α-helical conformations, respectively) (Fig. 4(a)). These structural differences account for the shorter MGD found for the glycosylation of the L75P mutant respect to the wild type sequence (Fig. 4(b)). Thus, the proline residue in this construct is probably changing the structure on top of Pro75, inducing a more extended conformation (Fig. 4(c)). In this regard, it is worth noting that a helix-breaking effect of proline residues located at an analogous position has been found in Lep constructs harboring, in place of its second TM fragment, either

artificial polyleucine stretches or a TM helix of the *Rhadobacter sphaeroides* photosynthetic reaction center [27].

Computational modeling of proline mutants

In order to better understand the effect of proline substitutions, some of the mutants were generated "in silico" and simulated by using a protocol of energy minimization and molecular dynamics. In general, the consequences of the substitution of a residue by proline in an α-helix are a constrained ϕ rotamer at the position of the proline, the loss of a H bond donor and the appearance of steric clashes between the proline cyclic side chain and the main chain backbone. All these effects may eventually produce a kink of the TM helix at a position about four residues N-terminal of the proline location [42]. If the kinked helix participates in packing interactions, an interference on TM association is to be expected.

Molecular modeling of proline substituted GpA dimers shows that mutations at the first turn of the N-terminus produce a destabilization of this turn, which may be viewed as an extension of the peptide chain (as inferred form Fig. 4). There is a subtle change in the case of the L75P mutant (Fig. 5(a)) because the corresponding residue is at the end of the first turn in the structural model of the TM peptide [15] used as a base for the simulations, but the extension is more evident in the case of the I77P (Fig. 5(b)) and I76P (not shown) mutants. With respect to the F78P mutant, proline substitution produces an

increase of the pitch of the turns around this residue (Fig. 5(c)), although the main features of the helix structure and the relative position of the residues involved in helix-helix interactions are conserved. These observations agree with the different effect on dimerization found for the reciprocal mutants (Fig. 1). Thus, proline substitution at position 75 permits dimerization and, in the correspondent model, we observe no significant perturbation in the structure of the helix at the level of the interaction surface (Fig. 6(a),(b)). In contrast, proline substitution at positions 76 or 77 clearly perturbs the interaction surface (see Fig. 6(c) for the I77P mutant), in agreement with the lack of dimer formation in these two mutants (Fig. 1). Interestingly, proline at position 78 produces only a minor effect in the helix interface (Fig. 6(d)), which again agrees with the large percentage of dimer formation found, even though this residue is placed immediately before the crucial Gly79.

When proline residues are present at the C-terminal side of the interaction motif, a kink of the helix is observed in most cases (see for example the I88P mutant, Fig. 5(d)). This kink changes the direction of the helices and may affect the relative orientation of the relevant residues for helix-helix interaction and so impair dimerization. For example, small variations on the relative position of the Thr87 β-hydroxyl group will affect inter-helix hydrogen bond formation [43]. When proline is more than four residues away from Thr87, as in S92P, it is again tolerated. Even at this region, important disturbing effects may still occur, as is demonstrated by the fact that the Y93P mutant does not

dimerize (Fig. 1). It is not clear why this may happen (Fig. 5(e)). However, the structural effect of the proline residue itself in this mutant should not be very strong because it was possible to observe the formation of heterodimers between the Y93P protein and a peptide corresponding to the wild type sequence (Fig. 5(e), inset), while no other heterodimers were found between the Wt peptide and any of the other non-dimerizing chimeric proteins (not shown).

Discussion

The GpA homodimeric complex results from the association between TM fragments through helix-helix contacts involving chiefly a motif of seven-residues ($L^{75}I^{76}xxG^{79}V^{80}xxG^{83}V^{84}xxT^{87}$). This motif is placed asymmetrically close to the N-terminal end of the TM α-helix. Thus, one might have expected that proline substitutions would have a more pronounced disrupting effect when amino acids close to this end were replaced. However, proline is more tolerated in terms of dimer formation at the N-terminus than at the C-terminal side. Especially significant in this sense are the L75P mutant, where a residue of the dimerization motif is substituted, and the F78P mutant, where substitution is made just before the crucial residue Gly79. These two mutant proteins register a high degree of dimer formation. On the contrary, at the C-terminal side, proline is permitted only when is placed far away from the interacting residues. Several factors could contribute to this observed pattern. First, the proline side chain projects towards the N-terminus, inducing a stronger distortion on residues placed in the turn immediately before the proline residue. Such an effect is due to sterical hindrance and the lack of canonical hydrogen bonding. In fact, this preferred anisotropic distortion is observed statistically if a number of proline-containing TM helices of known structure are superimposed [42]. Second, the asymmetry of the effect of proline substitution with respect to the interaction motif may be due to the asymmetry of the motif itself. Thus, it has been recognized that, among the seven residues that compose this motif, the pair

$G^{83}V^{84}$ exert the main contribution to the stability of the dimer [16]. Additionally, the Thr87 residues are responsible of the formation of interhelical hydrogen bonds [43], and these directional interactions could be easily affected by the structural distortions caused by proline substitution at the C-terminal side.

The behavior of L75P mutant, usually considered an important element of the dimerization motif of GpA, deserves a detailed discussion. Leu75 was initially included in the dimerization motif, but its contribution to dimer stabilization has become controversial. The seven-residue pattern was clearly defined as the first known dimerization motif for TM α-helices when, grafted onto polyleucine stretches, promoted specific dimerization [30]. Due to the nature of such a scaffold (Leu residues) the contribution of Leu75 cannot be evaluated in these constructs. Later on, alteration of the α-helical pathway by means of different insertions in the critical helix-helix interface compromised in several mutants the importance of Leu75 in the dimerization process [13,44]. Additionally, individual Leu75Ala [45,46], Leu75Met [47], Leu75Phe [46] and Leu75Val [48] mutations did not result in significant reduction of dimer formation in biological membranes in several genetic assay systems.

In our L75P mutant, the proline residue is placed at the third position of the helical domain in the solved GpA TM peptide structure [15]. Thus, its presence should only compromise in a limited manner the secondary structure of this region, as the synthetic TM peptides demonstrates (Fig. 2(a)). In fact, the helical propensity of proline residues in model peptides has been previously shown to

be greatly enhanced in membrane mimetic environments [19]. Both glycosylation mapping and molecular modeling (Figs. 4 and 5) showed that the presence of proline at position 75 has the effect of extending the N-terminal end of the helix, leaving the rest of the dimerization motif competent for dimer formation. Altogether these results point towards a less relevant role for Leu75 compared to the rest of the residues included in the dimerization motif of GpA.

The presence of proline residues, at least at the N-terminal turn of a TM helix, could also afford an alternative and attractive role for this residue based on its incapability to form standard β-sheets [39]. In the context of membrane protein biogenesis, translation and folding occurs in the complex environment of the translocon. In this membrane channel, TM fragments are temporarily exposed to the aqueous pore prior to membrane insertion. In a survey of amino acid preferences for specific locations in TM segments, it has been evidenced an enrichment in residues that have high β-sheet propensities in water at the N-terminal half of type I single spanning membrane proteins like GpA [49]. In this scenario, a Pro-dependent preferential destabilization of β-conformations could allow the polypeptide chain to sample other conformations, including α-helices, while being in the aqueous channel of the translocon. These TM helices can then laterally assemble to form the native protein (second-stage of membrane protein folding). Conversely, in this second-stage, proline residues located at the central positions of the interacting TM segments will be pernicious in cases where extensive helix packing is needed, like in the example studied here.

Finally, the influence of proline residues on GpA dimer formation, as depicted in Fig. 1, can be closely related to the occurrence of this residue at different positions in TM α-helices in general, as studied by Sansom and co-workers [50,51]. According to these authors the frequency of appearance of proline residues in TM □-helices diminishes, in an almost regular manner, as we move from the lipid/water interface to the center of the membrane, in good agreement with the results presented here (compare Fig. 1 with Fig. 13 from reference [51]). There is then a nice correlation between this reported data survey compilation and our experimental observations, suggesting that proline residues can be tolerated at the end of the TM helix, close to the lipid/water interface. However, a more recent study from the same group [42] shows a relatively higher frequency of proline in the center of TM helices.

All in all, the general conclusion that can be extracted from the present study is that helix packing of TM helices may be one of the factors that determine the frequency of appearance of proline residues in TM proteins. Thus, in addition to the functional role suggested for prolines in signal transduction and in the gating mechanism of ion channels, significant effects on helix-helix interactions are found, that should be taken into account as far as membrane protein folding is concerned.

Materials and Methods

Plasmid constructs

Construction of the plasmids encoding the His-tagged chimeric proteins (SN/GpA) are described by [13,28]. Mutations at the TM fragment of GpA were obtained by site directed mutagenesis using the QuickChange™ site directed mutagenesis kit (Stratagene, La Jolla, California). Introduction of the TM fragment from GpA into the Lep sequence was carried out by replacing H2 segment of Lep vectors with designed glycosylation acceptor sites at different positions [27] by PCR amplification of the GpA sequence with forward and backward primers containing appropriate restriction sites as previously described [52]. All mutants were confirmed by DNA sequencing.

Protein expression and purification

Overexpression and purification of His-tagged SN/GpA was performed as described [53]. *In vitro* transcription of Lep-derived constructs was done as previously [40]. The reactions were incubated at 37°C for 2h. The mRNAs were purified using Qiagen RNeasy clean up kit and verified on a 1% agarose gel. *In vitro* translation of the mRNA synthesized from the *in vitro* transcription was done in the presence of reticulocyte lysate and [^{35}S]-Met.

Peptide synthesis

Peptides containing wild type (Wt) and mutant (L75P) GpA TM sequences and non-native N- and C-terminal lysine residues (Fig. 2, top) were chemically synthesized using previously reported protocols [37]. The lysine residues were appended to confer water solubility to these hydrophobic peptides, which facilitates peptide purification and handling while keeping secondary structure and dimeric state of TM GpA as demonstrated by Melnyk *et al.* [37]. Solid phase synthesis of the peptides was performed using Fmoc chemistry on an Applied Biosystems 433A Peptide synthesizer. The low-load polyethylene glycol (PAL-PEG-PS) resin was from Applied Biosystems. Extended coupling conditions with HATU/DIEA activation pair were used with an 8-fold molar excess amino acids (Senn Chemicals). Double and triple couplings were applied on difficult residues [54]. Deprotection and cleavage reactions were carried out in 88% trifloroacetic acid (TFA)/5% phenol/5% water/2% triisobutylsilane (TIBS) (v/v) (adapted from [37]). Cleaved peptides were precipitated with ice-cold diethyl ether. Centrifugated pellets were dried, redissolved in water, and lyophilized. Peptide purification was performed using a C18 preparative reversed phase-high performance liquid chromatography (RP-HPLC) system to purity larger than 95% as determined by analytical RP-HPLC. Individual peptides were analyzed by mass spectroscopy to confirm their molecular weights.

SDS-PAGE analysis

Purified proteins were loaded onto 12% SDS polyacrylamide mini-gels (BioRad). Peptide samples were subjected to SDS-PAGE using 10-20% tricine precast gels (Novex Corporation). The loading buffer contained 2% SDS, and samples were boiled for 5 minutes prior to electrophoresis. Gels were stained with Coomassie blue, and the percentages of monomer, homodimer and heterodimer were estimated with an LKB Ultroscan 2202 laser densitometer with a 3390A Hewlett-Packard integrator. Gels with radioactive samples were dried at 80º C and scanned using a Fuji FLA-3000 phosphorimager using the Image Reader 1.0 software.

CD spectroscopy

All measurements were carried out on a Jasco J-810 CD spectropolarimeter, equiped with a Neslab RTE 110 water bath and temperature controller calibrated with isoandrosterone [52]. The spectra were measured in a 1 mm path length cell. Data were taken with a 0.2 nm step size, 8 s average time, 20 nm/min speed, and the results of 20 scans were averaged. Thermal melts were performed by collecting data at 222 nm every 0.2 ºC, in 10 mM Tris/HCl buffer, pH 7 in the presence of 5 mM SDS. The peptide concentration was 30 µM as determined by UV spectroscopy using ε_{276}= 1450 $M^{-1}cm^{-1}$ for tyrosine [55].

Molecular modeling

Structural models of dimeric proline mutants of GpA were created *in silico* starting from the coordinates of a dimeric α-helix fragment (residues Ser69 to Lys101) corresponding to the structure of the TM domain of the wild type protein as determined by NMR in detergent micelles (PDB ID 1AFO) [15]. For each model, the side chain of the residue to be mutated was replaced by the side chain of proline, while maintaining the coordinates of the backbone atoms, using the program Swiss-PdbViewer [56]. The structures obtained were then subjected to a protocol of energy minimization and molecular dynamics (MD) in vacuum using the GROMACS [57,58] package with double precision. MD simulations were run for 100 pico seconds without restraints. The trajectories obtained in this way were analyzed with the help of the program VMD [59] and characteristic structures were selected and energy minimized.

Figure legends

Figure 1. Pro-scanning mutagenesis of the helical residues in the GpA TM fragment. The seven residues associated with dimer formation are shown in green. The yellow band corresponds to the approximate location of the hydrophobic core of the bilayer (width 30Å). The blue bar denotes wild type GpA standard dimerization and the orange one emphasize dimerization of L75P. Each data point represents the average of at least three independent protein expressions and purification experiments ± standard deviation (SD).

Figure 2. Secondary structure and oligomeric state of synthetic TM peptides in detergent. **(a)** The secondary structure and thermal stability. CD of the synthetic TM peptides Wt (black) and L75P (gray) at 30 μM in aqueous buffer (doted lines) and in detergent buffers containing 10 mM SDS (solid lines) and DPC (dashed lines) micelles. The inset shows the mean residue ellipticity (MRE) at 222 nm with increasing temperature. **(b)** Dimeric state of synthetic TM peptides probed by SDS-PAGE analysis. Peptides (25 μM) dissolved in SDS-containing sample buffer were boiled for 5 min prior to electrophoresis using 10-20% tricine gels. Each lane is labeled on top except for molecular weight standards (third lane), the sizes of which are indicated at the right side. Fourth lane shows an equimolecular mixture of both peptides. Synthetic peptide sequences are shown on top with Lys-taggs underlined.

Figure 3. Competition experiments with Wt and L75P synthetic peptides. Purified SN/GpA chimeric wild type ((**a**) and (**c**)) and L75P mutant ((**b**) and (**d**)) proteins were mixed with Wt ((**a**) and (**b**)) and L75P ((**c**) and (**d**)) synthetic TM peptides at different molar ratios. Samples were tested for disruption of chimera homodimer by the peptides in SDS-PAGE. Positions of the monomer and homodimer of the chimeras, heterodimer and heterotrimer of the chimera and peptide are marked on the right.

Figure 4. Glycosylation mapping for the GpA TM helix. **(a)** Model of Lep chimeras showing the relation between glycosylation and the length of an extended chain from a TM helix. The second TM fragment of Lep was replaced by the GpA TM amino acid sequence (orange box). **(b)** Glycosylation efficiency for acceptor sites located at different distances *d* from the GpA helix and for the indicated L75P mutant. **(c)** Model of the TM chimeras for *d*=9 (counting from Glu72). GpA residues are shown in upper case, those resulting from the cloning of GpA TM and Lep sequence are shown in lower case (acceptor Asn, in bold, is included).

Figure 5. Structural models of dimeric proline mutants of the TM fragment of GpA compared with the structure of the wild type protein. Models were generated by *in silico* site-directed mutagenesis made on the structure of wild type GpA (NMR structure, PDB ID 1AFO [15]), followed by short runs of MD simulations and energy minimization (see Materials and Methods). Only a cartoon of the backbone (green color for the wild type protein, purple color for

the mutants) and the substituting Pro side chain (in yellow) is represented. **(a)**, **(b)** and **(c)** are three mutants at the N-terminal side of the dimerization motif. **(d)** and **(e)** are mutants at the C-terminal side. The inset in **(e)** shows the formation of hetero-oligomers between the Y93P GpA mutant and a peptide corresponding to the wild type TM sequence. Pictures made using the program VMD [59].

Figure 6. Top view of the interface of wild type GpA and various proline mutants of residues at the N-terminal end of the TM α-helix. The backbone of the proteins is represented as a tube. Heavy atoms of side chains from the substituting proline and residues that conform the dimer interface are shown as balls with the van der Waals radius (yellow for Pro, pink for Leu, iceblue for Ile, green for Gly and orange for Val). The rest of the structure is omitted for clarity. For comments on the generation of the models see Fig. 5 and Materials and Methods. Pictures made using the program VMD [59].

Acknowledgments

We thank Prof. Gunnar von Heijne (Stockholm University) who kindly provided us with the Lep vectors. This work was supported by grants BMC2000-1448 from the Spanish MCyT, GV00-040-5 from Generalitat Valenciana and Fundación Ramón Areces (Spain).

References

1. von Heijne, G. (1996). Principles of membrane protein assembly and structure. *Progr Biophys Mol Biol* **66**(2), 113-139.
2. White, S. H., and Wimley, W. C. (1999). Membrane protein folding and stability: physical principles. *Annu Rev Biophys Biomol Struct* **28**, 319-65.
3. Popot, J. L., and Engelman, D. M. (1990). Membrane protein folding and oligomerization - The 2-stage model. *Biochemistry* 29(17), 4031-4037.
4. White, S. H., Wimley, W. C., Ladokhin, A. S., and Hristova, K. (1998). Protein folding in membranes: determining energetics of peptide-bilayer interactions. *Methods Enzymol* **295**, 62-87.
5. Popot, J. L., and Engelman, D. M. (2000). Helical membrane protein folding, stability, and evolution. *Annu Rev Biochem* **69**, 881-922.
6. Booth, P. J. (1997). Folding alpha-helical membrane proteins: kinetic studies on bacteriorhodopsin. *Fold Des* **2**(6), R85-92.
7. Lu, H., Marti, T., and Booth, P. J. (2001). Proline residues in transmembrane alpha helices affect the folding of bacteriorhodopsin. *J Mol Biol* **308**(2), 437-46.
8. Kahn, T. W., Sturtevant, J. M., and Engelman, D. M. (1992). Thermodynamic measurements of the contributions of helix-connecting loops and of retinal to the stability of bacteriorhodopsin. *Biochemistry* **31**(37), 8829-39.
9. Therien, A. G., and Deber, C. M. (2002). Interhelical Packing in Detergent Micelles. Folding of a cystic fibrosis transmembrane conductance regulator construct. *J. Biol. Chem.* **277**(8), 6067-6072.
10. Bormann, B. J., and Engelman, D. M. (1992). Intramembrane Helix-Helix Association in Oligomerization and Transmembrane Signaling. *Annu Rev Biophys Biomol Struc* **21**, 223-242.
11. Lemmon, M. A., and Engelman, D. M. (1994). Specificity and promiscuity in membrane helix interactions. *Q Rev Biophys* **27**(2), 157-218.

12. Lemmon, M. A., Flanagan, J. M., Treutlein, H. R., Zhang, J., and Engelman, D. M. (1992). Sequence specificity in the dimerization of transmembrane α-helices. *Biochemistry* **31**(51), 12719-12725.
13. Mingarro, I., Whitley, P., Lemmon, M. A., and von Heijne, G. (1996). Ala-insertion scanning mutagenesis of the glycophorin A transmembrane helix. A rapid way to map helix-helix interactions in integral membrane proteins. *Protein Sci.* **5**, 1339-1341.
14. Adams, P., Engelman, D., and Brünger, A. (1996). Improved prediction for the structure of a dimeric transmembrane domain of glycophorin A obtained through global searching. *PROTEINS - Struct Funct Genet* **26**, 257-261.
15. MacKenzie, K. R., Prestegard, J. H., and Engelman, D. M. (1997). A transmembrane helix dimer: Structure and implications. *Science* **276**, 131-133.
16. Smith, S. O., Song, D., Shekar, S., Groesbeek, M., Ziliox, M., and Aimoto, S. (2001). Structure of the transmembrane dimer interface of glycophorin A in membrane bilayers. *Biochemistry* **40**(22), 6553-6558.
17. Senes, A., Gerstein, M., and Engelman, D. M. (2000). Statistical analysis of amino acid patterns in transmembrane helices: the GxxxG motif occurs frequently and in association with beta-branched residues at neighboring positions. *J Mol Biol* **296**(3), 921-36.
18. DeGrado, W. F., Gratkowski, H., and Lear, J. D. (2003). How do helix-helix interactions help determine the folds of membrane proteins? Perspectives from the study of homo-oligomeric helical bundles. *Protein Sci* **12**(4), 647-665.
19. Li, S.-C., Goto, N. K., Williams, K. A., and Deber, C. M. (1996). α-helical but not ß-sheet, propensity of proline is determined by peptide environment. *Proc Natl Acad Sci USA* **93**, 6676-6681.
20. Bywater, R. P., Thomas, D., and Vriend, G. (2001). A sequence and structural study of transmembrane helices. *J Comput Aided Mol Des* **15**(6), 533-52.

21. Chou, P. Y., and Fasman, G. D. (1974). Prediction of protein conformation. *Biochemistry* **13**, 222-245.

22. von Heijne, G. (1991). Proline kinks in transmembrane α-helices. *J. Mol. Biol.* **218**(3), 499-503.

23. Ostermeier, C., Harrenga, A., Ermler, U., and Michel, H. (1997). Structure at 2.7 A resolution of the Paracoccus denitrificans two-subunit cytochrome c oxidase complexed with an antibody FV fragment. *Proc Natl Acad Sci U S A* **94**(20), 10547-53.

24. Deisenhofer, J., Epp, O., Sinning, I., and Michel, H. (1995). Crystallographic refinement at 2.3 A resolution and refined model of the photosynthetic reaction centre from Rhodopseudomonas viridis. *J Mol Biol* **246**(3), 429-57.

25. Li, S.-C., and Deber, C. M. (1994). A measure of helical propensity for amino acids in membrane environments. *Nat Struct Biol* **1**, 368-373.

26. Kim, M. K., and Kang, Y. K. (1999). Positional preference of proline in alpha-helices. *Protein Sci* **8**(7), 1492-9.

27. Nilsson, I., Sääf, A., Whitley, P., Gafvelin, G., Waller, C., and von Heijne, G. (1998). Proline-induced disruption of a transmembrane α-helix in its natural environment. *J Mol Biol* **284**, 1165-1175.

28. Lemmon, M. A., Flanagan, J. M., Hunt, J. F., Adair, B. D., Bormann, B.-J., Dempsey, C. E., and Engelman, D. M. (1992). Glycophorin A dimerization is driven by specific interactions between transmembrane α-helices. *J Biol Chem* **267**, 7683-7689.

29. Smith, S. O., and Bormann, B. J. (1995). Determination of helix-helix interactions in membranes by rotational resonance NMR. *Proc Natl Acad Sci USA* **92**(2), 488-491.

30. Lemmon, M. A., Treutlein, H. R., Adams, P. D., Brünger, A. T., and Engelman, D. M. (1994). A dimerization motif for transmembrane α-helices. *Nat Struct Biol* **1**, 157-163.

31. Gierasch, L. M., Lacy, J. E., Thompson, K. F., Rockwell, A. L., and Watnick, P. I. (1982). Conformations of model peptides in membrane-mimetic environments. *Biophys J* **37**(1), 275-84.

32. Yang, J. T., Wu, C. S., and Martinez, H. M. (1986). Calculation of protein conformation from circular dichroism. *Methods Enzymol* **130**, 208-69.

33. Lau, S. Y., Taneja, A. K., and Hodges, R. S. (1984). Synthesis of a model protein of defined secondary and quaternary structure. Effect of chain length on the stabilization and formation of two-stranded alpha-helical coiled-coils. *J Biol Chem* **259**(21), 13253-61.

34. Goetz, M., Carlotti, C., Bontems, F., and Dufourc, E. J. (2001). Evidence for an alpha-helix --> pi-bulge helicity modulation for the neu/erbB-2 membrane-spanning segment. A 1H NMR and circular dichroism study. *Biochemistry* **40**(21), 6534-40.

35. Zhou, F. X., Cocco, M. J., Russ, W. P., Brunger, A. T., and Engelman, D. M. (2000). Interhelical hydrogen bonding drives strong interactions in membrane proteins. *Nat Struct Biol* **7**(2), 154-60.

36. Choma, C., Gratkowski, H., Lear, J. D., and DeGrado, W. F. (2000). Asparagine-mediated self-association of model transmembrane helix. *Nat Struct Biol* **7**(2), 161-166.

37. Melnyk, R. A., Partridge, A. W., and Deber, C. M. (2001). Retention of native-like oligomerization states in transmembrane segment peptides: application to the Escherichia coli aspartate receptor. *Biochemistry* **40**(37), 11106-13.

38. Partridge, A. W., Melnyk, R. A., and Deber, C. M. (2002). Polar residues in membrane domains of proteins: molecular basis for helix-helix association in a mutant CFTR transmembrane segment. *Biochemistry* **41**(11), 3647-53.

39. Wigley, W. C., Corboy, M. J., Cutler, T. D., Thibodeau, P. H., Oldan, J., Lee, M. G., Rizo, J., Hunt, J. F., and Thomas, P. J. (2002). A protein sequence that

can encode native structure by disfavoring alternate conformations. *Nat Struct Biol* **9**(5), 381-8.

40. Nilsson, I., and von Heijne, G. (1993). Determination of the distance between the oligosaccharyltransferase active site and the endoplasmic reticulum membrane. *J Biol Chem* **268**(8), 5798-5801.
41. Johansson, M., Nilsson, I., and von Heijne, G. (1993). Positively charged amino acids placed next to a signal sequence block protein translocation more efficiently in *Escherichia coli* than in mammalian microsomes. *Mol Gen Genet* **239**(1-2), 251-256.
42. Cordes, F. S., Bright, J. N., and Sansom, M. S. (2002). Proline-induced distortions of transmembrane helices. *J Mol Biol* **323**(5), 951-60.
43. Smith, S. O., Eilers, M., Song, D., Crocker, E., Ying, W., Groesbeek, M., Metz, G., Ziliox, M., and Aimoto, S. (2002). Implications of threonine hydrogen bonding in the glycophorin A transmembrane helix dimer. *Biophys J* **82**(5), 2476-86.
44. Mingarro, I., Elofsson, A., and von Heijne, G. (1997). Helix-helix packing in a membrane-like environment. *J Mol Biol* **272**(4), 633-641.
45. Langosch, D., Brosig, B., Kolmar, H., and Fritz, H. J. (1996). Dimerization of the glycophorin A transmembrane segment in membranes probed with the ToxR transcription activator. *J Mol Biol* **263**(4), 525-530.
46. Russ, W. P., and Engelman, D. M. (1999). TOXCAT: a mesure of transmembrane helix association in a biological membrane. *Proc Natl Acad Sci USA* **96**, 863-868.
47. Brosig, B., and Langosch, D. (1998). The dimerization motif of the glycophorin A transmembrane segment in membranes: Importance of glycine residues. *Protein Sci* **7**, 1052-1056.
48. Schneider, D., and Engelman, D. M. (2003). GALLEX, a Measurement of Heterologous Association of Transmembrane Helices in a Biological Membrane. *J. Biol. Chem.* **278**(5), 3105-3111.

49. Landolt-Marticorena, C., Williams, K. A., Deber, C. M., and Reithmeier, R. A. F. (1993). Non-Random distribution of amino acids in the transmembrane segments of human Type-I single span membrane proteins. *J Mol Biol* **229**(3), 602-608.

50. Sansom, M. S., and Weinstein, H. (2000). Hinges, swivels and switches: the role of prolines in signalling via transmembrane alpha-helices. *Trends Pharmacol Sci* **21**(11), 445-51.

51. Ulmschneider, M. B., and Sansom, M. S. (2001). Amino acid distributions in integral membrane protein structures. *Biochim Biophys Acta* **2**(1), 1-14.

52. Vilar, M., Sauri, A., Monne, M., Marcos, J. F., von Heijne, G., Perez-Paya, E., and Mingarro, I. (2002). Insertion and Topology of a Plant Viral Movement Protein in the Endoplasmic Reticulum Membrane. *J. Biol. Chem.* **277**(26), 23447-23452.

53. Orzaez, M., Perez-Paya, E., and Mingarro, I. (2000). Influence of the C-terminus of the glycophorin A transmembrane fragment on the dimerization process. *Protein Sci* **9**(6), 1246-53.

54. Fisher, L. E., and Engelman, D. M. (2001). High-Yield Synthesis and Purification of an [alpha]-Helical Transmembrane Domain. *Analytical Biochemistry* **293**(1), 102-108.

55. Chakrabartty, A., Kortemme, T., Padmanabhan, S., and Baldwin, R. L. (1993). Aromatic side-chain contribution to far-ultraviolet Circular Dichroism of helical peptides and its effect on measurement of helix propensities. *Biochemistry* **32**, 5560-5565.

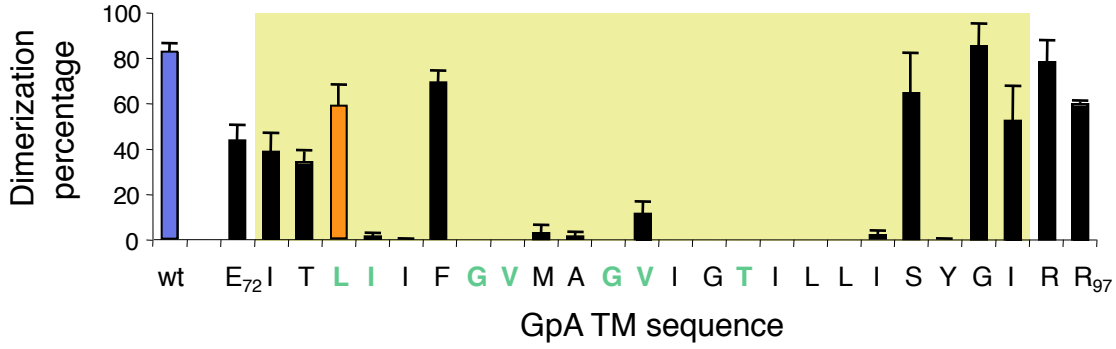
56. Guex, N., and Peitsch, M. C. (1997). SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* **18**(15), 2714-23.

57. Berendsen, H. J. C., van der Spoel, D., and van Drunen, R. (1995). GROMACS: A message-passing parallel molecular dynamics implementation. *Comp Phys Comm* 91, 43-56.

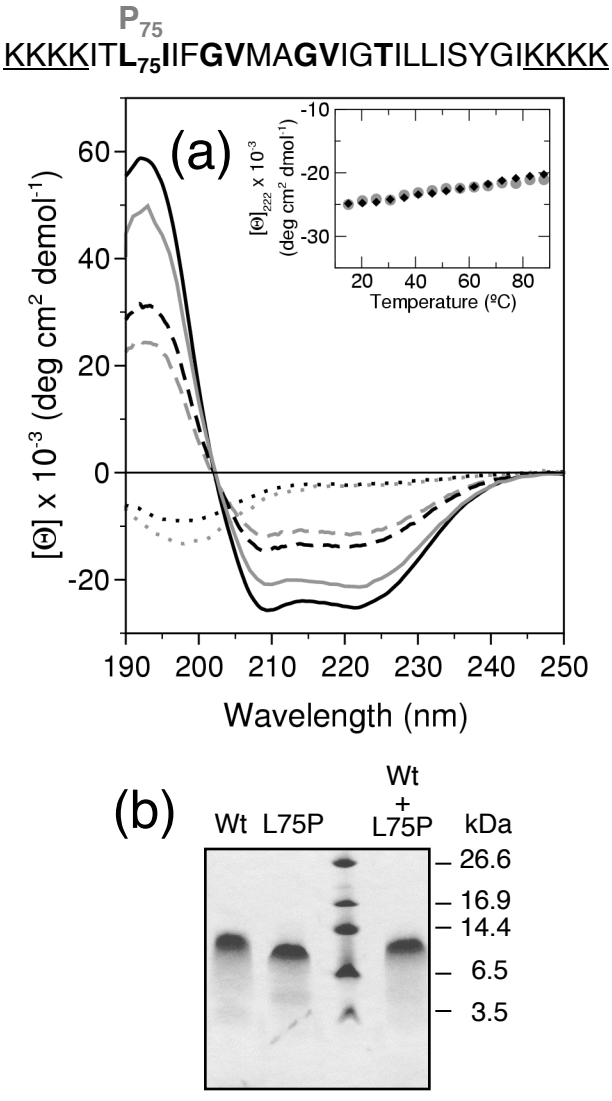
58. Lindahl, E., Hess, B., and van der Spoel, D. (2001). GROMACS 3.0: A package for molecular simulation and trajectory analysis. *J Mol Model* **7**(8), 306-317.

59. Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. *J Mol Graph* **14**(1), 33-8.

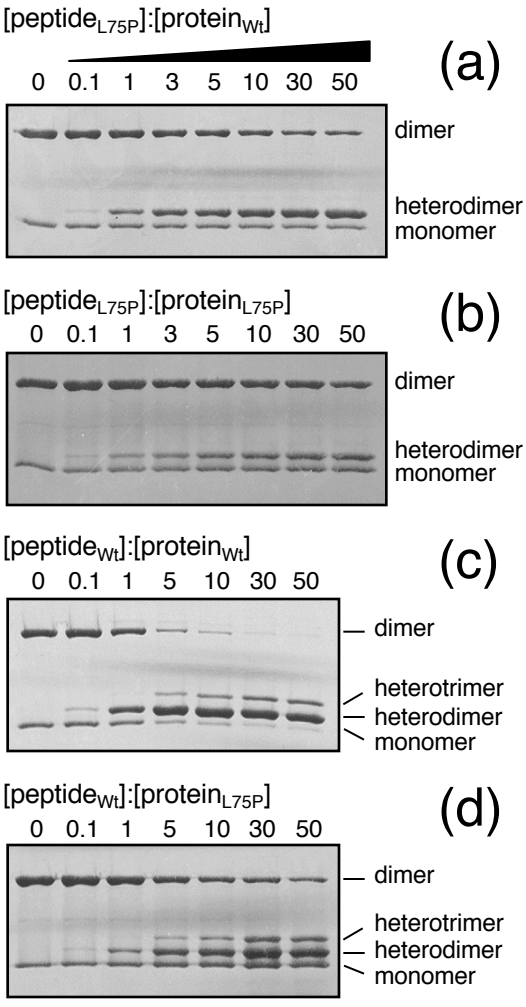
Orzáez *et al.* Fig. 1



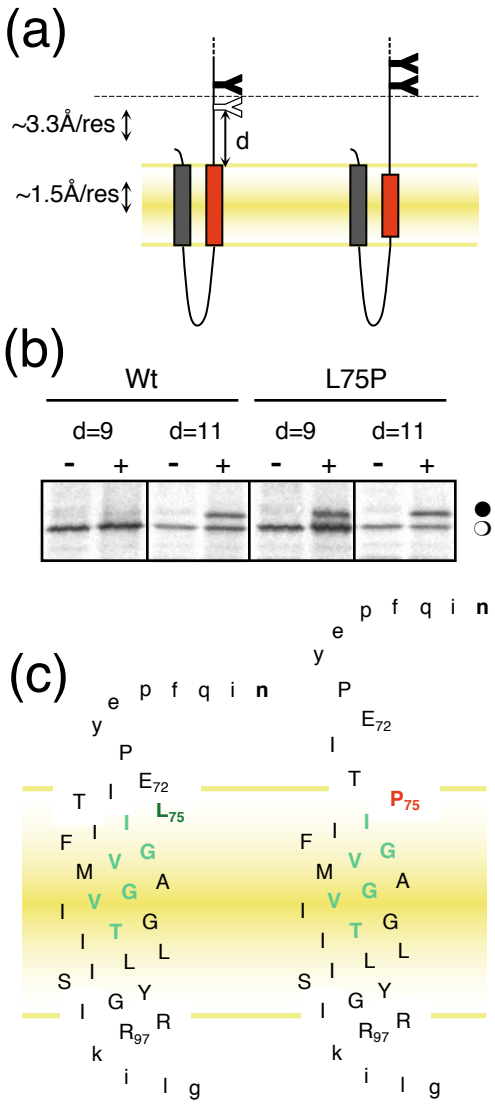
Orzáez *et al.* Fig. 2



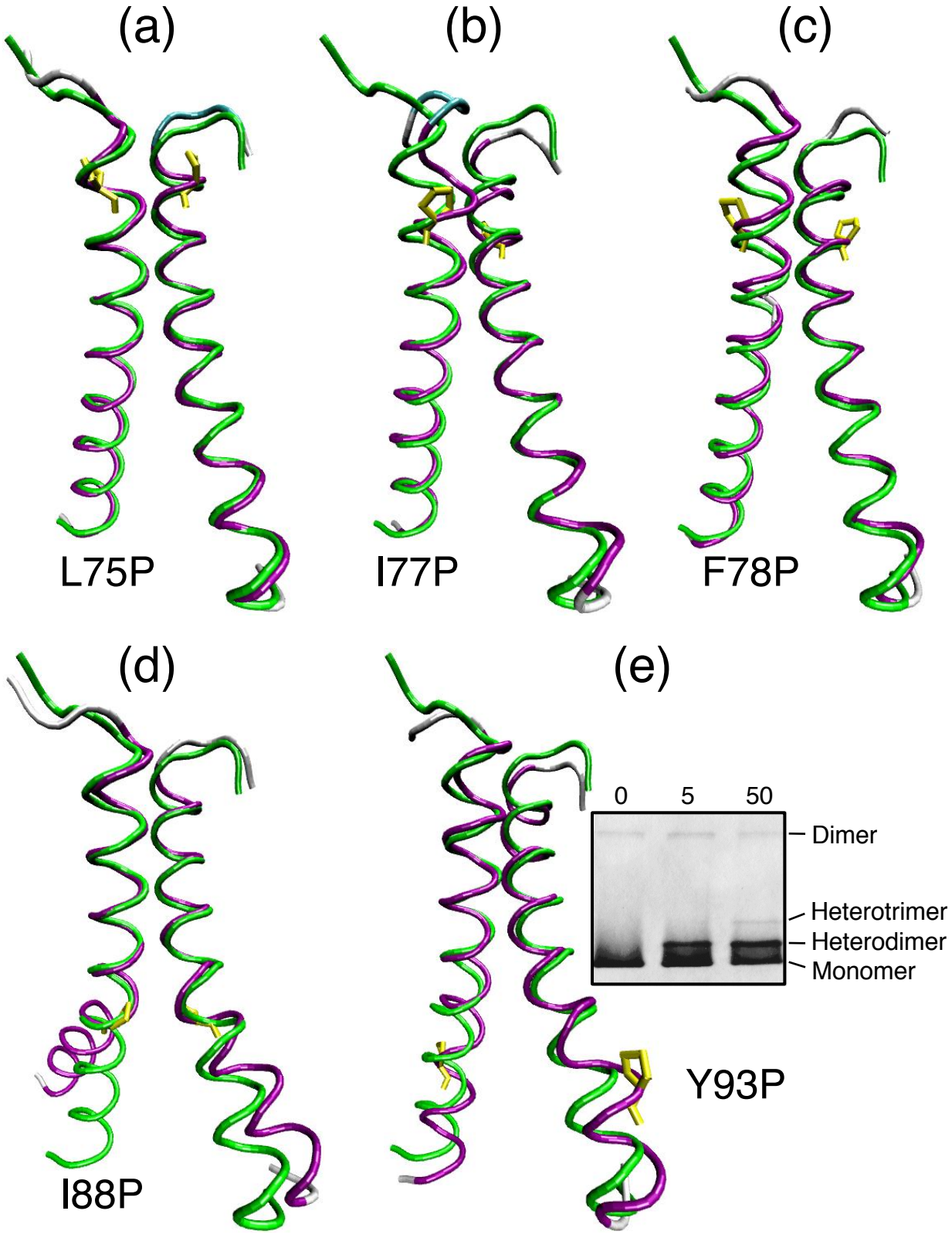
Orzáez *et al.* Fig. 3



Orzáez *et al.* Fig. 4



Orzáez *et al.* Fig. 5



Orzáez *et al.* Fig. 6