Protein Science (2000), *9*:1246–1253. Cambridge University Press. Printed in the USA.

Influence of the C-terminus of the glycophorin A transmembrane fragment on the dimerization process

MAR ORZÁEZ, ENRIQUE PÉREZ-PAYÁ, AND ISMAEL MINGARRO

Departament de Bioquímica i Biologia Molecular, Universitat de València, E-46 100 Burjassot, Spain

(RECEIVED January 31, 2000; FINAL REVISION April 10, 2000; ACCEPTED April 13, 2000)

Abstract

The monomer–dimer equilibrium of the glycophorin A (GpA) transmembrane (TM) fragment has been used as a model system to investigate the amino acid sequence requirements that permit an appropriate helix–helix packing in a membrane-mimetic environment. In particular, we have focused on a region of the helix where no crucial residues for packing have been yet reported. Various deletion and replacement mutants in the C-terminal region of the TM fragment showed that the distance between the dimerization motif and the flanking charged residues from the cytoplasmic side of the protein is important for helix packing. Furthermore, selected GpA mutants have been used to illustrate the rearrangement of TM fragments that takes place when leucine repeats are introduced in such protein segments. We also show that secondary structure of GpA derivatives was independent from dimerization, in agreement with the two-stage model for membrane protein folding and oligomerization.

Keywords: flanking regions; glycophorin A; helix packing; transmembrane fragments

The structural motifs that have been observed in membrane proteins are the α-helices, composed mainly of hydrophobic amino acids flanked by clusters of polar residues, and the β-barrels, in which hydrophobic residues face outward, toward the lipid bilayer (von Heijne, 1996; Wallin & von Heijne, 1998; White & Wimley, 1999). The α-helix type proteins are the most abundant and can be made up by a single helix or by multiple helices packed together in bundles. These membrane-spanning α-helices, rather than serving merely as featureless hydrophobic anchors of the proteins into the membranes facilitating their insertion, may be also responsible in part for stabilizing the assembly of membrane protein complexes through helix–helix interactions (reviewed in Lemmon & Engelman, 1994).

In fact, although the sequence characteristics that control the membrane insertion step and the orientation (topology) of the helices in the membrane are quite well understood (von Heijne, 1994), less is known about the molecular interactions that drive the helix–helix packing. This lack of information is due both to the scarcity of informative assays for helix–helix interactions (both in vivo and in vitro), and to the well-known difficulty of obtaining high-resolution structural information for membrane proteins.

Reprint requests to: Ismael Mingarro, Departament de Bioquímica i Biologia Molecular, Universitat de València, E-46 100 Burjassot, Spain; e-mail: Ismael.Mingarro@uv.es.

Abbreviations: EDTA, ethylenediaminetetraacetic acid; GpA, glycophorin A; SDS-PAGE, sodium dodecylsulfate polyacrilamide-gel electrophoresis; TM, transmembrane; N-terminal, amino-terminal; C-terminal, carboxi-terminal.

Some of the few existing assay systems that allow this kind of questions to be addressed are based on the single transmembrane (TM) domain from the human erythrocyte sialoglycoprotein glycophorin A (GpA). This TM domain drives a SDS-stable, noncovalent homodimerization of the protein, rendering perhaps the best-characterized example of interaction between TM α-helices and providing a convenient model of membrane protein folding where the dimer forms a right-handed coiled-coil.

Several mutagenic studies in detergent micelles (Lemmon et al., 1992b; Mingarro et al., 1996) computational modeling (Adams et al., 1996; MacKenzie & Engelman, 1998), and mainly, solution NMR (MacKenzie et al., 1997), have shown that a homodimeric complex results from the association between the GpA TM monomers mediated by helix–helix contacts. These contacts involve chiefly a seven-residue motif ($L^{75}I^{76}xxG^{79}V^{80}xxG^{83}V^{84}xxT^{87}$), present on one face of each TM α-helix. The dimerization motif established from these studies has been successfully grafted onto polyleucine stretches (Lemmon et al., 1994), as well as different host sequences allowing in vivo genetic assay systems in biological membranes (Langosch et al., 1996; Leeds & Beckwith, 1998; Russ & Engelman, 1999).

Although it is now well accepted the key role of this seven-residue motif, the contribution of each single residue in this motif to the sequence-specific interaction has become lately controversial by several lines of evidence. Replacement of Gly83, previously found as the most crucial residue for dimerization in a saturation mutagenesis study (Lemmon et al., 1992b), by isoleucine has only a mild effect on dimer formation (Leeds & Beckwith, 1998). Individual mutation of residues L^{75}, I^{76}, V^{80}, and V^{84} within

this seven-residue motif to alanine did not result in significant reduction of dimer formation (Langosch et al., 1996). Alteration of the α-helical pathway by means of alanine insertion at different positions between Ile76 and Gly79 of the GpA wild-type sequence had no deleterious effect, compromising the importance of $L^{75}I^{76}$ in the dimerization process (Mingarro et al., 1996). Furthermore, following insertion of 3–5 contiguous residues into the middle of the seven-residue motif (between Met81 and Ala82), mutation of Gly79 did not abrogate dimer formation (Mingarro et al., 1997). All together, these results point toward a global contribution of the TM fragment, and probably the flanking region, in the dimer formation process more than a very specific and local interaction. Within this context, we have explored the C-terminal part of the TM fragment and its flanking region to identify relevant requirements for helix packing in membrane proteins. This region presents well-defined secondary structure (MacKenzie et al., 1997), and no key residues for dimer formation have been described so far.

Here, we present evidence that extend our knowledge on the contribution of the C-terminal part of the α-helix to the GpA homodimer formation. Experiments utilizing deletion and point mutation of the wild-type amino acid sequence shed light on new features that redefine the contribution of the C-terminal part of the TM fragment and the flanking cytoplasmic charged regions, toward GpA homodimer stability in a membrane-like environment. Thus, we have found residues that initially were not assigned as significant for dimer formation disclose a dramatic effect on the packing of this model membrane protein.

Results

It was suggested early that the interaction to form a dimer occurred through the transmembrane domain of GpA (Bormann et al., 1989). Using dimerization in the presence of SDS as an in vitro assay, replacement mutagenesis was performed to identify the TM residues that constitute the GpA dimerization interface (Lemmon et al., 1992b). This study suggested that a seven-residue motif ($L^{75}I^{76}xxG^{79}V^{80}xxG^{83}V^{84}xxT^{87}$, where "x" represents a noncrucial residue for dimer formation) could define the dimerization interface. In early studies, this motif appeared to be averse to amino acid substitution; therefore, even conservative substitutions in one of these residues diminish or abolish GpA dimerization (Lemmon et al., 1992b). More recently, the solution NMR structure of the dimeric GpA TM fragment confirmed that actually, the seven-residue motif constitute the dimer interface (MacKenzie et al., 1997).

Nevertheless, it seems that not all seven residues contribute equally to the dimerization process. In fact, starting by the N-terminal end of the motif, our previous results suggested that Leu75 and Ile76 are less critical than the other interface residues (Mingarro et al., 1996). Concerning the central GV pairs, in an in vivo system based on the ToxR transcription activator (where the natural TM fragment of ToxR protein is substituted by sequences derived from GpA TM fragment), the central glycines ($G^{79}xxxG^{83}$) appeared to be the most crucial residues within the seven-residue motif, depending on the context (Brosig & Langosch, 1998). However, these two glycine residues in the natural GpA context have been shown previously to have a different susceptibility toward alanine replacement, while mutant G79A shown significant dimer, mutant G83A was absolutely disruptive (Lemmon et al., 1992b). To reinforce the idea of a different relevance of these two GV pairs, we designed the double mutant G79A,V80A. These two residues are closely packed in the solved structure (MacKenzie et al., 1997), and it was already known that their independent replacement by alanine residues provided significant dimer in both point mutants (Lemmon et al., 1992b). As shown in Figure 2 (Lane 2), this double mutant forms significant dimer, which means that residues 79 and 80 are permissive to conservative amino acid replacement, pointing toward the second GV pair ($G^{83}V^{84}$) as more important for an efficient helix–helix packing.

Along the same lines, our previous results indicated that sequences with major deviations from the wild-type GpA TM sequence (including replacements and insertions of several residues) could still support efficient dimerization. It appeared that the right-handed helix–helix crossing found for wild-type GpA was retained in these new dimers, and that Gly83 (and probably also Val84 and Thr87) was critical in all cases (Mingarro et al., 1997). All in all, because it seems that the C-terminal part of the motif could have a weighty contribution to the dimerization process, we focused our studies to this segment as a relevant region for the helix packing.

Dimerization equilibrium of C-terminal deleted constructs

In an attempt to elucidate the relevant residues of the C-terminus of the GpA TM fragment for helix packing, a residue deletion strategy was designed to analyze this region (Fig. 1A). It should be mentioned here that none of these residues was previously defined as a sensitive position (Lemmon et al., 1992b). We found that deletion of segment L^{89}–Y^{93} (Δ89–93) (Fig. 2, Lane 3), as well as subsequent partial deletions of this segment (Δ89–91 and Δ91–93) (Fig. 2, Lanes 4 and 5, respectively) abrogated dimer formation. These three constructs all contain deletion of the amino acid residue Ile91, located at the contact interface between the two TM fragments (Fig. 1B). However, a mutant protein with a point deletion of this residue (Δ91) dimerized as efficiently as the wild-type (Fig. 2, Lane 6), indicating that Ile91 is probably not a specifically relevant residue and bringing up the hypothesis of an important role for helix length in the dimezation process.

Leucine replacement of C-terminal residues

Next, we evaluated whether an amino acid–dependent specificity event is by itself responsible for the lack of protein dimerization in these regions. Leucine stretches have been long used as artificial TM fragments (Kuroiwa et al., 1991). Because leucine has a high α-helical propensity in membrane environments (Li & Deber, 1994; Blondelle et al., 1997) and because the GpA-TM fragment is entirely α-helical (MacKenzie et al., 1997), it seemed unlikely that leucine mutations would cause gross secondary structure perturbation. Thus, a mutant where residues I^{91}–Y^{93} were replaced by leucines (91–93L) showed a dimer formation indistinguishable from the wild-type (Fig. 2, Lane 7), suggesting that the specific amino acid sequence $I^{91}SY^{93}$, is not involved in key residue interactions when the dimerization of the TM fragment takes place. Furthermore, extension of leucine replacement until residue Ile95 (91–95L) gave a similar result (Fig. 2, Lane 8), accordingly to previous replacement data in this region (Lemmon et al., 1994).

Conversely, as mentioned above, when this sequence ($I^{91}SY^{93}$) is deleted dimer formation is precluded (Fig. 2, Lane 5). Deletion of three or five residues from an α-helical segment implies elimination of about one helical turn and this would place residues in unfavorable location for GpA dimer formation. In addition, when such a deletion is made at the C-terminal end of a TM fragment, it

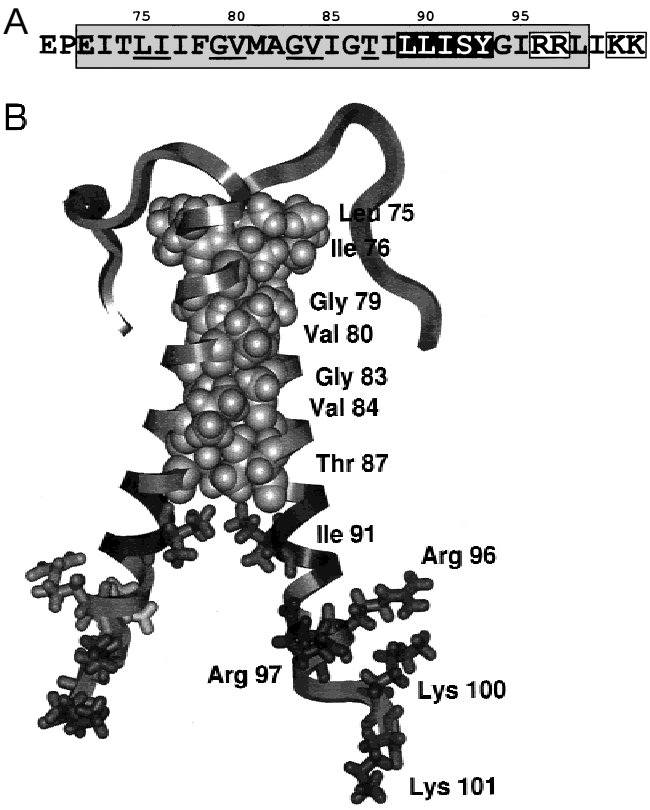

Fig. 1. A: Sequence of the wild-type GpA transmembrane fragment. The seven critical interface residues defined by replacement mutagenesis (Lemmon et al., 1992b) and NMR spectroscopy (MacKenzie et al., 1997) are underlined. The helical domain in the solved structure (MacKenzie et al., 1997) is enclosed in a shaded box. Residues at which deletion or replacement mutations were made are white in a black box. White boxes emphasize positively charged residues at the C-terminus of the TM sequence. **B:** Ribbon drawing derived from the transmembrane helix of the experimental NMR structure (PDB file 1AFO). The seven critical interface residues are shown in space-filling mode, deleted or replaced residues in dark grey and Ile91, Arg96, Arg97, Lys100, and Lys101 in sticks mode. The drawing was generated using a Silicon Graphics Indigo workstation and the program Insight II (Molecular Simulations, San Diego, California).

may force to place amino acids that in the wild-type protein are part of extra-membrane (interface) domains, in a chemically unfriendly environment. The NMR solution structure of the GpA-TM fragment in a micellar medium (MacKenzie et al., 1997) reflects the formation of an α-helix that extents from Ile73 to Leu98 (Fig. 1A). Although paying attention to the regular membrane thickness, the segment I^{73}–S^{92} (20 amino acids) would be enough to expand the hydrocarbon core of the membrane (Fig. 3). Taking into account that characteristically tryptophan and tyrosine side chains are located at the membrane interfaces for all known membrane proteins (White & Wimley, 1999), residues $I^{91}SYGI^{95}$ would be placed at the membrane-water interface. It is also important to mention that Tyr93 is the only of these two residues present in the GpA TM fragment (Fig. 1A). Furthermore, it is noteworthy that the presence of a glycine residue in this region could be significant for conferring, to the polypeptide chain, enough flexibility to accommodate the highly positively charged extra-membrane residues $R^{96}RLIKK^{101}$, brought in close proximity when the protein

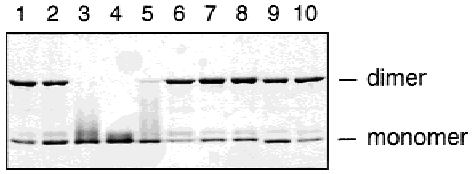

Fig. 2. SDS-PAGE analysis of GpA mutants. Lane 1: nuclease-GpA wild-type fusion protein. Lane 2: double mutant G79A,V80A. Lane 3: Δ89–93. Lane 4: Δ89–91. Lane 5: Δ91–93. Lane 6: Δ91. Lane 7: replacement of residues 91–93 by leucines (91–93L). Lane 8: replacement of residues 91–95 by leucines (91–95L). Lane 9: deletion of arginines 96 and 97 (Δ96/97). Lane 10: replacement of arginines 96 and 97 by leucine residues (96/97L).

dimer is formed. Taking this picture into account, those mutants with deletions of three (Δ91–93) or five (Δ89–93) amino acid residues would tend to incorporate the end of the helix too close to the central dimerization motif (i.e., the seven-residue motif $L^{75}I^{76}xxG^{79}V^{80}xxG^{83}V^{84}xxT^{87}$). As a result, the charged extra-membrane segment is forced to be near the seven-residue motif and in close contact with the membrane, making the energetic balance unfavorable for dimer formation.

Influence of the C-terminal charged residues on hydrophobic helix–helix interactions

Since deletion of three residues imply roughly elimination of one helical turn, we started to consider the contribution of the charged residues located at the end of the helix in the solved structure (MacKenzie et al., 1997) and flanking the hydrophobic residues putatively embedded in a cellular membrane. In this context, with one helical turn deleted, arginines 96 and 97 could end up too close to the core of the contact interface. Deletion (Δ96/97) or replacement by leucines (96/97L) of these arginines in the wild-type sequence had little effect on the dimerization process (Fig. 2, Lanes 9 and 10, respectively). In the same direction, one would expect that if arginines have a pernicious effect in deleted mutants dimerization, after removal the system should dimerize. Surprisingly, deletion (Δ96/97) or replacement by leucines (96/97L) of these arginines on a shortened version of the TM fragment (i.e., in mutant Δ89–91) did not restore dimer formation (Table 1), probably because in these constructs lysines 100 and 101 abolish, somehow, the correct packing of the helices. Significantly, in this sense, replacement (by leucines) but not deletion of arginines 96 and 97 on mutant Δ91 allow dimer formation (Table 1), stressing the importance of a minimal distance between the core of the dimerization (seven-residue motif) and the lysine residues when arginines are not present. Actually, it should be noted that mutants Δ89–91,96/97L and Δ91,Δ96/97 showed the same incapacity to form dimers (Table 1). Interestingly, in these two mutants the lysine residues have to play the same role since in both, three residues are deleted and no other charged residues precede them. To confirm this hypothesis, we have replaced these lysines by leucines in mutant Δ89–91,96/97L, and found a significant amount of dimer (mutant Δ89–91,96/97L,100/101L in Table 1). These combined results suggest that the TM fragment driven dimerization of GpA depend on both the previously described seven-residue dimerization motif and the distance to the membrane and/or relative orientation of the charged extra-membrane domain. In

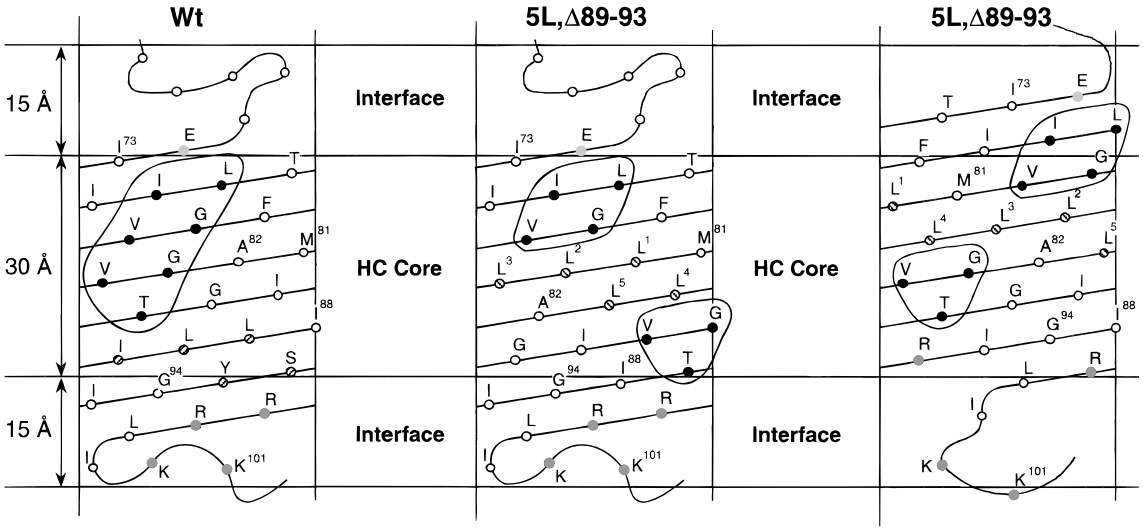

Fig. 3. Helical net plots of the wild-type GpA TM fragment (left) and mutant 5L, Δ89–93. Cartoons show how the terminal segments of the helix are pulled-away of the membrane, either the C-terminal (central) or the N-terminal (right). The membrane is comprised of interfacial regions, each about 15 Å thick, and the hydrocarbon (HC) core, which has a thickness of about 30 Å. Helical area is represented according to the solved structure (Fig. 1B). Black balls represent the seven-residue motif. Grey balls indicate charged residues. Striped balls represent either deleted or inserted residues.

fact, the presence of charged residues at the cytoplasmic side of the GpA TM fragment seems important, since replacement of both arginine and lysine residues, exhibited a diminished dimer formation in the wild-type sequence (mutant 96/97L,100/101L in Table 1).

Effect of deletions and replacements in a different packing interface

The relevance of these results can be helpful in the definition of packing interfaces when, as an example, leucine stretches are in-

Table 1. *Fraction of dimer for GpA mutants*[a]

Mutant	Sequence	% Dimer
Wild-type	itLIifGVmaGVigT^{87}illisygirrlikk	84
G79A,V80A	itLIif**AA**maGVigT^{87}illisygirrlikk	54
Δ89/93	itLIifGVmaGVigT^{87}i-----girrlikk	0
Δ89-91	itLIifGVmaGVigT^{87}i---sygirrlikk	0
Δ91-93	itLIifGVmaGVigT^{87}ill---girrlikk	9
Δ91	itLIifGVmaGVigT^{87}ill-sygirrlikk	86
91-93L	itLIifGVmaGVigT^{87}ill**lll**girrlikk	78
91-95L	itLIifGVmaGVigT^{87}ill**lllll**rrlikk	77
Δ89-91,Δ96/97	itLIifGVmaGVigT^{87}i---sygi--likk	2
Δ89-91,96/97L	itLIifGVmaGVigT^{87}i---sygi**ll**likk	2
Δ89-91,96/97L,100/101L	itLIifGVmaGVigT^{87}i---sygi**ll**li**ll**	73
Δ91,Δ96/97	itLIifGVmaGVigT^{87}ill-sygi--likk	0
Δ91,96/97L	itLIifGVmaGVigT^{87}ill-sygi**ll**likk	45
Δ96/97	itLIifGVmaGVigT^{87}illisygi--likk	69
96/97L	itLIifGVmaGVigT^{87}illisygi**ll**likk	74
96/97L,100/101L	itLIifGVmaGVigT^{87}illisygi**ll**li**ll**	55
5L	itLIifGVm<u>llllll</u>aGVigT^{87}illisygirrlikk	34
5L,Δ89-93	itLIifGVm<u>llllll</u>aGVigT^{87}i-----girrlikk	5
5L,Δ89-91	itLIifGVm<u>llllll</u>aGVigT^{87}i---sygirrlikk	1
5L,Δ91-93	itLIifGVm<u>llllll</u>aGVigT^{87}ill---girrlikk	8
5L,91-93L	itLIifGVm<u>llllll</u>aGVigT^{87}ill**lll**girrlikk	48
5L,91-95L	itLIifGVm<u>llllll</u>aGVigT^{87}ill**lllll**rrlikk	50

[a]Seven residues corresponding to the wild-type interface are shown in capital letters and other residues of GpA sequence are in small letters. Mutated residues are in bold and inserted residues are underlined. Dashed position symbols deleted residues.

corporated in TM fragments. In this sense, we have previously shown that when a stretch of five leucines was incorporated in the middle of the GpA-TM fragment (mutant 5L; Mingarro et al., 1997), the protein dimerization was allowed, although the dimerization interface should be necessarily different from that of the wild-type TM fragment. Previous results showed that Gly79 was not a significant amino acid residue in the dimerization process of this mutant 5L, while it was highly significant in the wild-type sequence. However, amino acid residues Gly83, Val84, and Thr87 were significant for dimerization of both, the 5L mutant and the wild-type (Mingarro et al., 1997). These results point toward the interpretation that following the insertion of the five leucine stretch, the N-terminal end of the GpA-TM segment could be, to some extent, pushed-out of the membrane (or of the SDS micelles), reinforcing the role of the C-terminus for an appropriate helix–helix packing. To corroborate such an interpretation, we decided to explore the C-terminal end of the TM fragment in this mutant. The working hypothesis is based on the idea that the insertion of the leucine stretch leads to a major rearrangement of the whole TM segment resulting in one of two possible scenarios (see Fig. 3). Either the C-terminal part is pulled-away from the membrane (Fig. 3, center panel), or it is the N-terminal one (Fig. 3, right panel). In the first possibility, a further mutant with a deletion of five amino acids at the C-terminus ($L^{89}LISY^{93}$-mutant 5L,Δ89–93; Table 1) will render a TM fragment that would dimerize, because the location of the amino acid residues at the C-terminal end of the TM fragment will be very similar to that of the wild-type protein (compare left and center panel in Fig. 3). If the second possibility reflects the actual arrangement of the TM fragment in the 5L mutant, a further mutant, with the above-mentioned five amino acid residue deletion, will not dimerize based in our results with mutant Δ89–93 (Table 1). In fact, mutant 5L,Δ89–93 did not dimerize indicating that the insertion of the five leucine stretch could tend to pull-away from the membrane the N-terminal end of the GpA-TM segment, dragging the positively charged C-terminal end too close to the membrane and preventing protein dimerization (Fig. 3, right panel). It is important to note here that mutant 5L,Δ89–93 keeps the same hydrophobic helix length as the wild-type (23 residues), avoiding any contribution of this factor in the packing process of GpA derivatives. Hydrophobic length contribution has been demonstrated using synthetic peptides in a recent study where, within membranes, self-association of a polyleucine stretches of 19 leucines appears to be weaker than that of 23 leucines (Ren et al., 1999).

Parallel with our results with the C-terminal end of the wild-type GpA-TM fragment, deletion of only three amino acids on the 5L mutant (5L,Δ89–91 and 5L,Δ91–93) did not restore dimerization (Table 1). In contrast, when the positively charged C-terminal end is placed at its original position by means of additional leucine replacement (mutants 5L,91–93L and 5L,91–95L), the dimerization ability was restored (Table 1).

Secondary structure of peptides derived from the C-terminal TM fragment

To explore the α-helical contribution of this region to the global dimerization process, we have studied the secondary structure of a series of C-terminal derivatives. Peptide fragments of proteins and synthetic peptides have proved very useful in understanding helix formation and stability (see, for example, Baldwin, 1995), and helical propensities of peptide fragments of several proteins have been examined to understand the local folding of these proteins. The structure of the dimerization domain (shown in Fig. 1B), obtained from the NMR studies (MacKenzie et al., 1997), showed an extension of the helical structure beyond the hydrophobic residues of the TM fragment. To unravel the structure of this region out of the TM context, we have synthesized peptides covering the C-terminal part of the GpA sequence (Fig. 4A). Far-ultraviolet circular dichroism (CD) spectroscopy was used to detect whether dimerization propensity affected the secondary structure of the GpA derivatives. Figure 4B shows the CD spectra of the peptides in SDS at 25 °C. Two minima, one around 208 nm and the other at 222 nm characteristic of helices (Holzwarth & Doty, 1965; Rohl & Baldwin, 1998), are observed for all peptides (Fig. 4B). Interestingly, peptides 88–99,Δ89–91 and 88–102,Δ89–91, derived from mutants where no dimer formation was detected (Fig. 2), showed a clear helical content. It should be noted that neither in the first of these two peptides, which is a shortened version, nor in the second peptide, where three polar residues from the C-terminal part of the wild-type sequence of the protein were added (to keep peptide length), their secondary structure has been perturbed. These results are in the same lane of evidence than a recent study using fluo-

A

Mutant	Peptide sequence	θ_{222}
Wt	Ac-ILLISYGIRRLI-NH_2	-17,032
88-99,96/97K	Ac-ILLISYGI**KK**LI-NH_2	-14,277
88-99,Δ91	Ac-ILLSYGIRRLI-NH_2	-18,225
88-99,Δ89-91	Ac-ISYGIRRLI-NH_2	-17,901
88-102,Δ89-91	Ac-ISYGIRRLIKKS-NH_2	-19,206

B

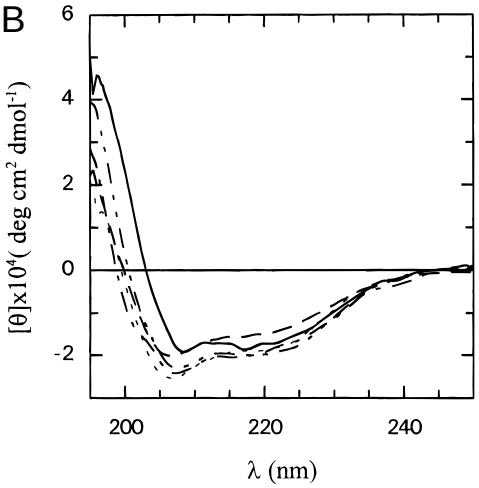

Fig. 4. A: Sequence of the synthetic peptides used in the CD studies with their molar ellipticity at 222 nm. **B:** CD spectra of the synthetic peptides in the presence of 3 mM SDS. Wild-type (——); 88–99,96/97K (– – –); 88–99,Δ91 (· · · ·); 88–99,Δ89–91 (—·—); 88–102,Δ89–91 (—··—).

rescent labeled GpA peptides (based on the TM sequence) where it has been confirmed a long-standing assumption about the GpA TM domain that helix formation is uncoupled from dimerization (Fisher et al., 1999).

Discussion

We have explored the sequence requirements for helix–helix packing in TM segments of integral membrane proteins by taking advantage of the known SDS-resistance of GpA dimers. We have found both that hydrophobic residues from the C-terminus previously described as not significantly implicated in the helix–helix packing, and other residues located at the putative extramembrane end of the reported GpA helix, have an essential and specific role for dimer stability in membrane-like environments (Table 1). This role could be related to the necessity of maintaining an appropriate distance and orientation with respect to the core of the dimerization motif and the charged residues that are usually present at the flanking regions of the TM fragments (Wallin & von Heijne, 1998). Therefore, deletions in this part of the helix would bring the charged residues, too close to the core of the dimerization motif (i.e., mainly Gly83, Val84, and Thr87) impairing helix packing.

To generalize our findings and to gain a better understanding of how these amino acid residues (located at the water-membrane interface) influence the overall stability and final arrangement of membrane proteins, it would be useful to analyze how these additional amino acids (beyond the seven-residue motif) participate in the assembly of proteins where the TM domain of GpA has been grafted. Functional examples can be seen in recent studies where segments covering amino acids 73 to 89 (Brosig & Langosch, 1998) and 75 to 87 (Russ & Engelman, 1999) of the TM domain of GpA were enough to drive the dimerization of the ToxR transcription factor. The apparent discrepancy between the previous mentioned grafted systems and our approach could arise from the fact that those chimeric constructs did not include the charged residues flanking the TM fragment in the wild-type GpA sequence. In fact, the last C-terminal residue from GpA included in these systems is Leu89 (Miller et al., 1987; Langosch et al., 1996; Russ & Engelman, 1999). The same type of limitation has been observed in a genetic system based on the lambda repressor N-terminal DNA-binding domain (Leeds & Beckwith, 1998), where the last residue included was Ile95, just before the arginine residues located at positions 96 and 97, again not including charged residues at the flanking region.

Along the same lines, a recently developed algorithm proposed to predict the packing of TM helices has been applied to obtain the global energy minimum for the TM helix dimer of GpA (Pappu et al., 1999). The results from this study suggest that the $G^{79}V^{80}xA^{82}G^{83}$ sequence motif is the major determinant of the overall helix dimer structure. Once more, the sequence used in this calculation (T^{74}-I^{91}) included neither the entire C-terminal hydrophobic portion of the helix, nor the charged residues 96, 97, 100, and 101.

In the context of existing data, our results have raised new questions and provided new explanations for the necessity of a certain distance between the core of the dimerization motif (i.e., the seven-residue motif) and the extramembrane charged residues. Deletions in the hydrophobic area would render these charged residues too close to the membrane, and it is likely that this proximity is what diminishes or abolishes dimer formation in every construct tested, emphasizing the importance of these residues. In the mutant Δ91,96/97L, where arginines were replaced by leucines, dimerization was maintained, probably because deletion of a single residue did not sufficiently compromise the hydrophobic contribution of this region. However, this was not the case when the arginines were deleted (mutant Δ91,Δ96/97). Similarly, in the mutant Δ89–91,96/97L, which has the arginines replaced by leucines and also three residues deleted, dimerization was not allowed. The reason could be that in these two last mutants three residues (roughly one helical turn) were deleted, and lysines 100 and 101 end up located at the same position on both, preventing dimer formation in a similar way. These data are consistent with observations made for other membrane proteins, where the substitution of arginine residues (located at the flanking regions joining TM segments) by lysines revealed no absolute requirement for arginine side chains in two different cytoplasmic loops from a glucose transporter (Sato & Mueckler, 1999). In the same line of evidence, our data point toward a global contribution of the basic residues of this region instead of a specific contribution of the arginine or lysine residues independently.

It should also be mentioned that in constructs derived from mutant 5L, where the 5 leucine stretch was inserted between the GV pairs, the two GV pairs of the seven-residue motif need to be located in opposite helix interfaces. In this construct, it was demonstrated that the pair $G^{83}V^{84}$ (the C-terminal pair) is crucial for TM packing (Mingarro et al., 1997). Concerning the relative position of the previously adduced arginines in this 5L construct, deletion of five residues (5L,Δ89–93) will not only bring the charged residues too close to the $G^{83}V^{84}$ pair, but will also place them in the same helical interface, impairing dimer formation, as can be inferred from Figure 3. Interestingly, the electrostatic repulsion between charged residues at the interface could also occur in the mutant Δ89-93, derived from the wild-type sequence, where the same five residue deletion would leave the arginines facing each other in both helices of the dimer, once again impairing dimerization.

As stated earlier in this study, the α-helix is an ubiquitous structural feature found in TM segments of integral membrane proteins. This structural motif has been found to play a key role in the assembly of this particular class of proteins because it has to keep in an appropriate balance helix–helix and helix–phospholipid interactions. However, due to the complexity and compound variables involved in such interactions, the identities of the sequences that contribute to the folding and stability are not yet well understood. In this present work, the role of the hydrophilic balance at the flanking regions in relation to the GpA dimerization process was shown to be major determinant bearing on dimer formation. At the same time, we have shown that in the presence of membrane-like environments, peptides derived from mutant sequences that did not dimerize are able to fold into an α-helical conformation. This is in agreement with previous approaches where labeled peptides were used to show that GpA helix formation is independent of dimerization in different micellar environments (Fisher et al., 1999). Our results provide evidence showing that, in contrast to soluble coiled-coils in which helix formation is coupled to oligomerization, the requisite of adopting a defined structure in membrane proteins is generally not a sufficient requisite to allow the fine-tuned interactions at the amino acid level that would render an efficient helix–helix packing. And these findings are in accordance with the two-stage model for integral membrane protein folding and oligomerization (Popot & Engelman, 1990), where it was hypothesized that the formation of individual TM helices may be

energetically distinct from their assembly into helical bundles and oligomeric structures.

Requirement of flanking regions in oligomerization processes has been described in other systems. In the self-association of heparan sulfate proteoglycan N-syndecan, it has been shown that the TM domain of the N-syndecan core protein was required but not sufficient for the formation of stable complexes (Asundi & Carey, 1995). The minimal amino acid sequence that conferred the ability to form protein complexes in this system, included four charged residues of the flanking region. Furthermore, point mutations that changed the basic residues to alanine residues within this region either partially or totally, abolished the ability of the N-syndecan core protein to form complexes (Asundi & Carey, 1995).

A variety of theoretical and experimental complications must be considered when attempting to characterize a system involving integral membrane protein folding or oligomerization. It is therefore necessary to explore simple model systems to wade through this difficult problem. In our approach, it is possible that for some constructs we do not detect dimerization in SDS, merely because they have a weaker affinity, but could associate in an actual membrane at certain concentrations. Despite this caveat, all together our results indicate that sequence context influences the previously established dimerization motif of GpA. Bearing this in mind, more attention has to be paid to the flanking regions of the TM fragments to fully understand helix packing of membrane proteins and to better use this knowledge to comprehend membrane protein–protein interactions, which are currently at an early stage.

Materials and methods

Plasmid constructs

Construction of the plasmids encoding the chimeric protein (SN/GpA) is described by Lemmon et al. (1992a). Shortly, the *Hind*III–*Bam*HI fragment from pSN/GpA was cloned in M13mp18. For purification purposes, a His6 tag was added by site-directed mutagenesis at the extreme C-terminus of the coding region using Kunkel's method (Kunkel, 1985) as modified by Geisselsoder et al. (1987). Modified *Hind*III–*Bam*HI fragment was cloned in the high level expression vector pT7SN/GpA (Lemmon et al., 1992a). Mutations at the TM fragment of GpA were obtained by site-directed mutagenesis using the QuickChange™ site-directed mutagenesis kit (Stratagene, La Jolla, California). All mutants were confirmed by DNA sequencing.

Expression, extraction, and purification of SN/GpA

For SN/GpA production, pT7SN/GpA was transformed into *Escherichia coli* BL21 (DE3) strain containing the plasmid pLYS-S (Novagen, Madison, Wisconsin). Colonies were picked and grown to logarithmic phase in LB at 37 °C. Cultures were diluted 1:100 into terrific-broth (TB) and grown to an A_{600} of 2.5. Isopropyl β-D-thiogalactopyranoside (IPTG) was then added to 0.8 mM and growth continued for a further 3 h. After harvesting by centrifugation, cells were resuspended 1:20 in 50 mM Tris-HCl (pH 8), 5 mM EDTA, 1 mM phenylmethylsulfonyl fluoride (PMSF), 0.025% (w/v) NaN_3. Cells were lysed by three rounds of freeze-thaw in this suspension, lysis being aided by the constitutive expression of T7 lysozyme directed by the pLYS-S plasmid. $CaCl_2$ was added to 10 mM to activate the nuclease moiety of SN/GpA. The resulting cellular DNA hydrolysis was complete after incubation for 15 min on ice. The lysate was clarified by centrifugation, and protein was extracted from the resulting pellet by sonication for 1 min at 4 °C (at 1:10 dilution) in a solution containing 20 mM Tris-HCl, 150 mM NaCl (pH 7.5; TBS), 1% (w/v) SDS, 1 mM PMSF. Nonsolubilized products were removed by centrifugation and chimeric proteins were purified using Ni-NTA agarose resin (Qiagen, Hilden, Germany). After loading, resins were washed with 10 mM imidazole, 0.5% SDS in TBS and eluted with 100 mM imidazole in the same solution. The presence of the His6 tail was found not to affect the dimerization efficiency and the chimeric product had the expected mobility on SDS-PAGE (Mingarro et al., 1996). Protein concentration was determined by the bicinchoninic acid protein assay (Pierce, Rockford, Illinois) using bovine serum albumin as a standard.

SDS-PAGE analysis

Two microliters of a 1 mg/mL (38 μM) solution of purified protein was loaded onto 12% SDS polyacrylamide mini-gels (BioRad, Hercules, California). The loading buffer contained 2% SDS, and samples were boiled for 5 min prior to electrophoresis. Gels were stained with Coomassie blue, and the percentage of monomer and dimer were estimated with an LKB Ultroscan 2202 laser densitometer with a 3390A Hewlett-Packard integrator.

Peptide synthesis

Peptides were prepared by simultaneous multiple peptide synthesis (SMPS) (Houghten, 1985) using Fmoc chemistry (Fields & Noble, 1990). After cleavage from the resin peptides were extracted with 95% acetic acid, and lyophilized. Individual peptides were analyzed by mass spectroscopy. Peptide purification was performed using a preparative reversed phase-high performance liquid chromatography (RP-HPLC) system to purity greater than 95% as determined by analytical RP-HPLC.

CD spectroscopy

All measurements were carried out on a Jobin Yvon CD6 CD spectropolarimeter (CD-Longjumeau, France) calibrated with isoandrosterone. The spectra were measured in a 1 mm pathlength cell. Data were taken with a 0.2 nm step size, 8 s average time, 20 nm/min speed, and the results of 10 scans were averaged. The CD spectra were taken at 25 °C, in 10 mM Tris-HCl buffer, pH 8 in the presence of 3 mM SDS. The peptide concentration was 25 μM as determined by UV spectroscopy using $\epsilon_{276} = 1{,}450\ M^{-1}\ cm^{-1}$ for tyrosine (Chakrabartty et al., 1993).

Acknowledgments

Thanks are due to Drs. Ramón Sendra and Cheryl Vaughan for critical reading of the manuscript, Desireé Camús and Ana Giménez for technical assistance, and Vicent Esteve for help in the use of computer graphics tools. We also thank Prof. Gunnar von Heijne (Stockholm University) for providing us with the original plasmids. This work was supported by the Spanish CICYT project no. SAF97-0067 and EU biotech grant BIO4-CT 97-2086. M.O. was recipient of a predoctoral fellowship from the Generalitat Valenciana (Spain).

References

Adams P, Engelman D, 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.

Asundi VK, Carey DJ. 1995. Self-association of N-syndecan (syndecan-3) core protein is mediated by a novel structural motif in the transmembrane domain and ectodomain flanking region. *J Biol Chem 270*:26404–26410.

Baldwin RL. 1995. α-Helix formation by peptides of defined sequence. *Biophys Chem 55*:127–135.

Blondelle SE, Forood B, Houghten RA, PerezPaya E. 1997. Secondary structure induction in aqueous vs membrane-like environments. *Biopolymers 42*:489–498.

Bormann B-J, Knowles WJ, Marchesi VT. 1989. Synthetic peptides mimic the assembly of transmembrane glycoproteins. *J Biol Chem 264*:4033–4037.

Brosig B, Langosch D. 1998. The dimerization motif of the glycophorin A transmembrane segment in membranes: Importance of glycine residues. *Protein Sci 7*:1052–1056.

Chakrabartty A, Kortemme T, Padmanabhan S, Baldwin RL. 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.

Fields GB, Noble RL. 1990. Solid phase peptide synthesis utilizing 9-fluorenylmethoxycarbonyl amino acids. *Int J Pept Prot Res 35*:161–214.

Fisher LE, Engelman DM, Sturgis JN. 1999. Detergents modulate dimerization, but not helicity of the glycophorin A transmembrane domain. *J Mol Biol 293*:639–651.

Geisselsoder J, Witney F, Yuckenberg P. 1987. Efficient site-directed in vitro mutagenesis. *BioTechniques 5*:786–791.

Holzwarth G, Doty P. 1965. The ultraviolet circular dichroism of polypeptides. *J Am Chem Soc 87*:218–228.

Houghten RA. 1985. General method for the rapid solid-phase synthesis of large number of peptides: Specificity of antigen-antibody interaction at the level of individual amino acids. *Proc Natl Acad Sci USA 82*:5131–5135.

Kunkel TA. 1985. Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proc Natl Acad Sci USA 82*:488–492.

Kuroiwa T, Sakaguchi M, Mihara K, Omura T. 1991. Systematic analysis of stop-transfer sequence for microsomal membrane. *J Biol Chem 266*:9251–9255.

Langosch D, Brosig B, Kolmar H, Fritz HJ. 1996. Dimerization of the glycophorin A transmembrane segment in membranes probed with the ToxR transcription activator. *J Mol Biol 263*:525–530.

Leeds JA, Beckwith J. 1998. Lambda repressor N-terminal DNA-binding domain as an assay for protein transmembrane segment interactions *in vivo*. *J Mol Biol 280*:799–810.

Lemmon MA, Engelman DM. 1994. Specificity and promiscuity in membrane helix interactions. *Q Rev Biophys 27*:157–218.

Lemmon MA, Flanagan JM, Hunt JF, Adair BD, Bormann B-J, Dempsey CE, Engelman DM. 1992a. Glycophorin A dimerization is driven by specific interactions between transmembrane α-helices. *J Biol Chem 267*:7683–7689.

Lemmon MA, Flanagan JM, Treutlein HR, Zhang J, Engelman DM. 1992b. Sequence specificity in the dimerization of transmembrane α-helices. *Biochemistry 31*:12719–12725.

Lemmon MA, Treutlein HR, Adams PD, Brünger AT, Engelman DM. 1994. A dimerization motif for transmembrane α-helices. *Nat Struct Biol 1*:157–163.

Li S-C, Deber CM. 1994. A measure of helical propensity for amino acids in membrane environments. *Nat Struct Biol 1*:368–373.

MacKenzie KR, Engelman DM. 1998. Structure-based prediction of the stability of transmembrane helix–helix interactions: The sequence dependence of glycophorin A dimerization. *Proc Natl Acad Sci USA 95*:3583–3590.

MacKenzie KR, Prestegard JH, Engelman DM. 1997. A transmembrane helix dimer: Structure and implications. *Science 276*:131–133.

Miller VL, Taylor RK, Mekalanos JJ. 1987. Cholera toxin transcriptional activator ToxR is a transmembrane DNA binding protein. *Cell 48*:271–279.

Mingarro I, Elofsson A, von Heijne G. 1997. Helix-helix packing in a membrane-like environment. *J Mol Biol 272*:633–641.

Mingarro I, Whitley P, Lemmon MA, 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.

Pappu RV, Marshall GR, Ponder JW. 1999. A potential smoothing algorithm accurately predicts transmembrane helix packing. *Nat Struct Biol 6*:50–55.

Popot JL, Engelman DM. 1990. Membrane protein folding and oligomerization—The 2-stage model. *Biochemistry 29*:4031–4037.

Ren J, Lew S, Wang J, London E. 1999. Control of the transmembrane orientation and interhelical interactions within membranes by hydrophobic helix length. *Biochemistry 38*:5905–5912.

Rohl CA, Baldwin RL. 1998. Deciphering rules of helix stability in peptides. *Methods Enzymol 295*:1–26.

Russ WP, Engelman DM. 1999. TOXCAT: A mesure of transmembrane helix association in a biological membrane. *Proc Natl Acad Sci USA 96*:863–868.

Sato M, Mueckler M. 1999. A conserved amino acid motif (R-X-G-R-R) in the Glut1 glucose transporter is an important determinant of membrane topology. *J Biol Chem 274*:24721–24725.

von Heijne G. 1994. Membrane proteins: From sequence to structure. *Annu Rev Biophys Biomol Struct 23*:167–192.

von Heijne G. 1996. Principles of membrane protein assembly and structure. *Progr Biophys Mol Biol 66*:113–139.

Wallin E, von Heijne G. 1998. Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Sci 7*:1029–1038.

White SH, Wimley WC. 1999. Membrane protein folding and stability: Physical principles. *Annu Rev Biophys Biomol Struct 28*:319–365.