The GxxxG motif and its relation to the dimerization of α-helical transmembrane proteins

Bachelorthesis

Bauke Gaastra S1864726, supervisor: Dirk Jan Slotboom

26-7-2012

Contents

Abstract

Membrane proteins are broadly represented in the genomes of all organisms and serve as the gates between the inside and outside of the cells. Because only a limited amount of high resolution structures are available, studying these proteins is more challenging. This paper presents an overview of the research on the GxxxG motif that is found on the interface of helix-helix interactions inside the membrane. Most of this research has been performed with glycophorin A as a model peptide. Through multiple mutational studies more data became available on the importance of the GxxxG motif. Later studies also confirmed that this motif is also present on interfaces of helix-helix interactions in water soluble proteins. The GxxxG motif proves to be a powerful motif to predict helix-helix interactions, though it is not sufficient or necessary for dimerization itself.

Introduction

One of the largest groups of proteins encoded in the genomes of organisms are the membrane proteins. Estimates range from 20 to 30%[1]. Membrane proteins form the gate between the inside and the outside of a cell and despite their abundance, the available high resolution X-ray structures are very limited, in contrast to structures of water soluble proteins. As a result of the low number of available structures, the knowledge about the folding of these proteins has been lagging behind.

Overall, membrane proteins can be divided into two classes. The first class folds as β-barrels inside the outer cell membrane of gram-negative bacteria. The second class folds into bundles of α-helices. Membrane proteins can be monotopic, bitopic or polytopic. Monotopic proteins do not span the entire membrane, but interact with one of the membrane surfaces or have a lipid anchor. Bitopic and polytopic proteins are transmembrane proteins which span the membrane once or twice, respectively. Glycophorin A, the protein that is mainly discussed in this paper, is a bitopic protein. It crosses the membrane once with an α-helix and forms homodimers. The proteins discussed at the end of the paper are polytopic, their structures are generally known as α-bundles.

Knowledge about the folding of membrane proteins has been lagging behind their water soluble counterparts. It is important to gain more knowledge on these proteins, as they comprise a large part of all the proteins. Also, mutations inside the transmembrane parts of specific proteins has been connected to some diseases, for example cystic fibrosis[2]. In cystic fibrosis a neutral to charged mutation inside a transmembrane helix of the cystic fibrosis transmembrane conductance regulator (CFTR) introduces a hydrogen bond within the protein, which causes a loss in function. Increased knowledge on the folding and oligomerization of membrane proteins could lead to new insights on how to deal with these diseases.

A lot of research about protein-protein interactions in membrane proteins has been done using Glycophorin A and its embedded GxxxG motif as model. But how are the results of these researches related to natural proteins? And how well do the dimerization motifs really predict interfaces between α-helices inside the membrane? In this paper, an overview of the work on Glycophorin A and other peptides is presented after which their relation to other natural proteins is discussed.

Glycophorin A

Glycophorin A (GpA) is one of the most used proteins in research on membrane protein-protein interactions. GpA is an integral transmembrane protein which is located in human erythrocyte (red blood cell) membranes. This protein was used in early studies on interactions between membrane proteins because it had been studied extensively in other research fields and was known to form stable dimers in the presence of SDS[3]. At the time, it was known that from a tryptic digest of GpA, only the trans-membrane part was able to dimerize with an intact GpA molecule, indicating that the interactions between the monomers have to be inside of the membrane[4]. Later research revealed that the interactions between the GpA monomers were reversible and specific[3].

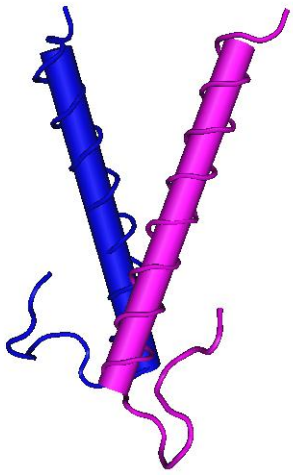

Figure 1: A cartoon of the GpATM structure (PDB ID: 1AFO). Two transmembrane segments of Glycophorin A are visible. Rendering was done by the program Cn3D.

More information became available with the determination of the GpA structure by solution

NMR. The structure allowed the researches to describe the physical basis of helix-helix association with the help of the two-stage model. This two-stage model states that α-helical membrane proteins are first inserted into the membranes as separate α-helices, after which these proteins can interact with other molecules to form complexes[5]. The NMR-structure (see figure 1) revealed that the two trans-membrane helices present in the dimer interact at a -40° (right handed) crossing angle and that the interactions between the helices mainly consist of favourable van der Waals interactions (later on also hydrogen bonds between the helices were proposed as stabilizing interactions for dimerization). It was also noted that two glycines, separated by three other residues, provided a surface for backbone to backbone interactions[6].

Prior to the determination of the Glycophorin A structure, it was already known that the interaction between the two monomers was sequence specific. A study performed by Lemmon et al.[7] in 1992 provided the first indications that the dimerization was sequence dependent. In this study, only the trans-membrane helix of GpA (GpATM) was used and a library was made with single mutants of this helix. These mutant GpATM helices were assayed for their ability to form dimers. Although no motif was identified in this work, the researchers did observe that the sensitive positions are separated on average by 3,9 residues, consistent with one helix turn. Especially the glycine residue at position 83, and to a somewhat lesser extend the glycine residue at position 79, were found to be very sensitive towards mutations. The researchers noted that these two amino acids may be needed at these positions to allow close contact between the two helices.

The study described above has given some valuable clues on the sequence specificity of the dimerization of GpA, but did not report a motif. The first motif was identified in later work by Lemmon et al.[8] In this paper, the motif LIxxGVxxGVxxT is presented. The question was asked whether this motif, when introduced in any α-helix, is able to induce dimerization of the helices. To answer this question, the proposed motif was fused to the carboxy terminus of the nuclease from *Staphallococcus aureus*, which normally is a monomeric protein. The added mass of the nuclease enable the researches to determine the oligomeric state of the peptides on an SDS-PAGE gel. All the residues between the residues that form the motif were changed into leucines, as well as the valine residue at position 80 (see Figure 2). These helices showed no reduction in dimerization, indicating that the motif is capable of inducing dimerization in helices that are made entirely out of leucine residues, apart from the residues of the motif itself. To validate the interactions between the helices are the same in these synthetic peptides, a number of mutations were introduced in the motif. The behaviour of these mutant helices was compared to the known characteristics of mutations in the GpA dimer. The results turned out to be very similar and strengthened the hypothesis that this motif is capable of inducing dimerization in membrane embedded α-helices.

To validate that the interactions between the helices were specific, competition experiments were carried out[8]. In these experiments, dimers of chimeric peptides (the SN/GpA, SN/$L_{23/7}$ and SN/L_{23-6} peptides as described above) are used. Transmembrane domains of several proteins (including the transmembrane part of SN/Gpa, SN/$L_{23/7}$ and SN/L_{23-6} (shaded boxes in figure 2)) are added to the dimers, which compete in the dimerization. Because these peptides are smaller than the chimeric proteins, the resulting dimers are smaller than the original chimeric dimers, which is clearly visible on an SDS-

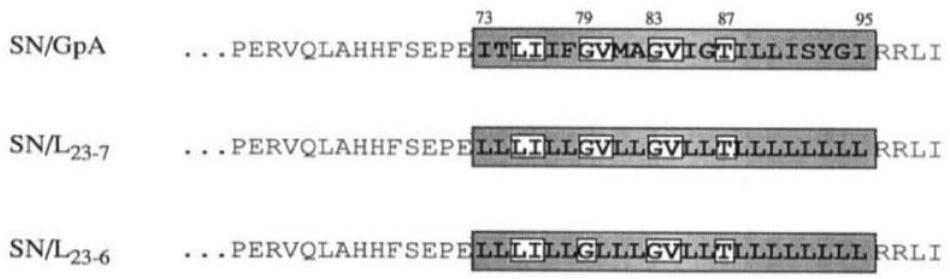

Figure 2. Sequences used in the paper of Lemmon et al. The transmembrane helix (shaded boxes) of GpA are fused to the C-terminus of the nuclease from *S.* aureus (SN/GpA). In SN/L_{23-7} all the residues in the transmembrane helix part that are not part of the motif were mutated to leucine. SN/L_{23-6}, in addition to SN/L_{23-7}, also has V80 mutated to leucine[8].

PAGE gel (see figure 3). Indeed, a significant drop in the molecular weight of the dimers was observed when GpATM was added to either the chimeric GpA or the leucine variant protein. However, when trans membrane parts of other membrane proteins were added, this was not observed, indicating that the interactions between the helices are specific.

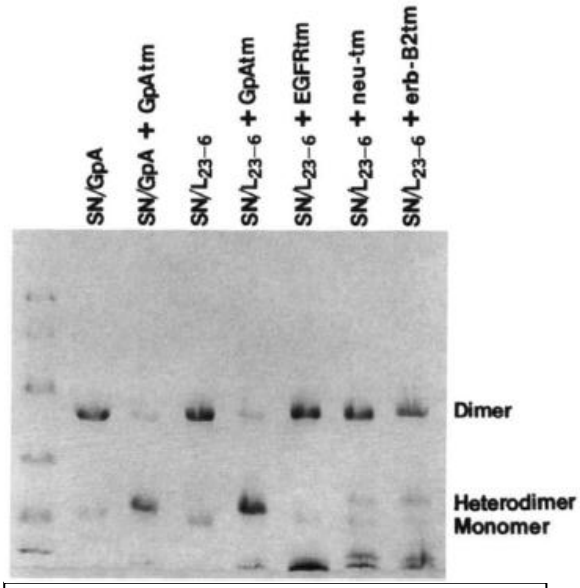

Figure 3. An SDS-PAGE gel of a competition experiment between SN/GpA or SN/L_{23-6} and the transmembrane domains of GpA, EGFR, neu and erb-B2. The transmembrane domains are added in tenfold concentration relative to SN/GpA or SN/L_{23-6}. Homo-dimerization was only blocked by adding the transmembrane domain of GpA (causing the band of the homodimer to disappear and appearance of the band corresponding to the heterodimer[8].

Another study, performed by Langosch et al.[9], made use of a different chimeric structure. In this study, the transmembrane segment of the ToxR protein from *Vibrio cholerae*, a transcription factor, was replaced by the transmembrane part of GpA which contains the motif described above. The transcription factor ToxR is normally an integral bacterial inner membrane protein, and it is only active in a dimeric state. When this transcription factor is incorporated into an indicator strain it is a useful tool to quantitatively monitor the dimerization of the selected helices. In this indicator strain (*Escherichia coli*), the ctx promotor, which is controlled by the ToxR transcription factor, is placed in front of the lacZ gene, which codes for the β-galactosidase enzyme. X-gal, a compound which can be supplemented to the medium, is cleaved by β-galactosidase. This produces a blue dye which can be used as a visible indicator of the activity of β-galactosidase[10], and thereby the dimerization of the ToxR-helix chimer (See figure 4). In contrast with previous studies, which examined the dimerization of GpA in SDS-micelles, this assay with the ToxR transcription factor allows the researches to study the dimerization in a biological membrane. This has some implications. First of all, in biological membranes, the helices are already preorientated, because of the planar shape of the membrane, which enhances the dimerization. Secondly, the (local) concentration is higher in biological membranes (the GpA molecules are confined to the membrane), than the concentration in an SDS solution. Taken together, this implicates that dimerization is more favourable in biological membranes than SDS-micelles. Therefore, mutations are expected to have a smaller effect on the dimerization in biological membranes than in SDS-micelles, only the most critical mutations will be identified[9].

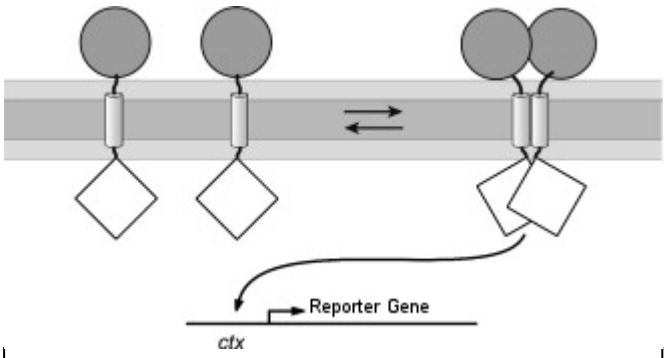

Figure 4. A schematic representation of the ToxR system. ToxR is visualised with the diamond shapes and fused to a transmembrane α-helix. The circles represent a maltose binding protein domain which aids the insertion of the construct into the membrane and can be used to quantify the amount of expression of the construct. On dimerization, it activates the transcription of genes with the ctx promoter[26].

All residues inside the transmembrane part of GpA were separately mutated to alanine to study the specificity of the motif for dimerization. Alanine was chosen because it had proven before to significantly reduce or completely abolish the dimerization of GpA in SDS-micelles. Only the (wildtype) alanine residue at position 82 was mutated to leucine. The strongest effect were again seen on the G79 and the G83 positions, dimerization was reduced with about 70% by both mutations. Additionally, the I76A and T87A mutations led to reductions of 35 and 43 % respectively. All the other mutations on the other residues gave reductions smaller than

25% compared to the wildtype, except for the A82L mutant. This mutant showed an increase of 22% compared to the wildtype, for which the researches had no explanation. The G79, G83, I76 and the T87 residues, which caused the largest reduction in dimerization when mutated to alanine, are all contained in the proposed motif. Other mutations in motif residues were found to be less critical in the biological membrane of the indicator strain used in this experiment, indicating that the G79, G82, I76 and the T87 residues are the most important residues within the motif.

In another study[11] a similar approach was taken to identify strong transmembrane helix packing motifs. The researches constructed a randomized sequence library based on the right-handed dimerization motif of Glycophorin A. In this library of transmembrane peptides, only the residues that were presumed to participate in the interaction between helices (positioned at one side of the helix) were mutated (see figure 5). The residues between the interface residues were either all alanine or all leucine, divided into two separate libraries (called ALAlib and LEUlib). The ability of these peptides to form dimers was assayed with the TOXCAT system, a system similar to the assay with the ToxR transcription factor described above. This TOXCAT system differs in the fact that the CAT gene is placed downstream of the ctx promotor. This CAT gene codes for the chloramphenicol acetyltransferase protein, which enables resistance to the antibacterial compound chloramphenicol (CAM). The synthetic peptides are again fused to the transcription factor ToxR, leaving only the strains with dimerizing peptides able to grow on the selective medium. Only the colonies with high resistance were used for further analysis. These colonies were obtained by using a high CAM concentration in the medium.

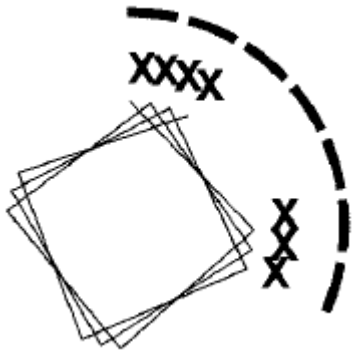

Figure 5. A schematic top view of an α-helix. The X's represent the positions that were mutated in this research. These positions are all located on the same side of the helix[11].

The results showed that in a large majority of the cases a motif of two glycines spaced by three other residues was found (in 96% of the sequences identified in the LEUlib, and in 79% of the sequences in the ALAlib). In the remaining sequences, one of those glycines was most commonly replaced by a serine. One striking difference between the two libraries was the fact that, although both showed a large dominance of the GxxxG motif in their sequences, the motif is shifted by one residue towards the carboxy terminal of the helix in the ALAlib (the glycine residues are located on position 5 and 9 for the Leulib and 6 and 10 for the ALAlib). The motifs that were found (the GxxxG and some poor-defined motifs) were validated for natural occurrence by searching the non-homology trans membrane domain database from SwissProt. The GxxxG motif had 1641 matches in this database, which is roughly 12 % of the whole database (13 606 trans membrane domains in total). This result suggest that these motifs, found with synthetic peptides, are also used in natural proteins.

The context of the GxxxG motif is important for dimerization of GpA

Not only the motif residues are important in the dimerization, the context of the residues also play an important role. In a study by Doura et al.[12] the motif flanking residues are examined for their ability to influence the stability of the dimerization of GpATM peptides. The researchers approach the problem in a systematic way, in which single point mutations to glycine, isoleucine, leucine and valine, were introduced at each position on the helix-helix interface. The effects of the mutations were determined by using sedimentation equilibrium analytical ultracentrifugation. This method allows to gather details on the free energy of association. In general, almost all of the mutations inside the dimer interface of GpA result in a lower stability of the dimers, but none of the peptides show no dimerization at all, indicating that the GxxxG motif is not necessary for dimerization. Only

the conservative I76V and V80L mutations show dimerization comparable to the wildtype. In peptides with an intact GxxxG motif, the stability of the dimerization can be altered by mutating the flanking residues, indicating that the context is important in this process. Also, the threonine residue at position 87 (this residue is positioned at the same side of the helix as the motif residues) seems to be very important for the dimerization. When this residue is mutated to either leucine or isoleucine, the effect is of the same proportion as an mutation to either one of the glycines in the motif. This suggests that this threonine residue may complement the strong interhelical interactions with the GxxxG motif.

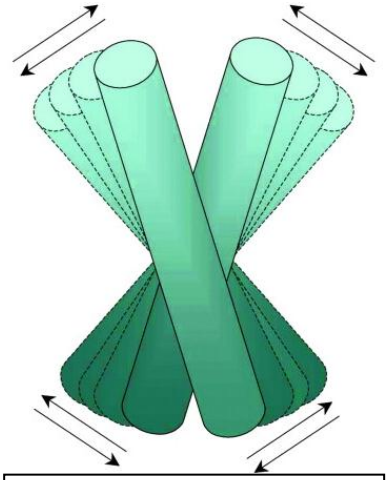

Figure 6. A representation of the proposed conformational changes caused by long ranged coupling. This coupling can cause the angle of the interaction between the helices to change[13].

In a follow-up study[13] more information became available about the important of the motif context. In this study, the researches performed a mutational analysis again, this time with double alanine mutants. The results are again gathered by sedimentation equilibrium analytical ultracentrifugation. A total of 21 double alanine mutants were characterized. Also, the results of the double mutants were compared to the expected result of the combination of the two separate single mutants. In general, all mutants appear to be destabilizing, with the exception of the L75A/V84A mutant. Dimerization occurred in each of the mutants, aside from the L75A/T87A mutant. The free energy of coupling (subtracting the experimental free energy of dimerization of the double mutants by the calculated free energy of dimerization of both single-point mutants) revealed that the majority of the mutants have a stabilizing coupling energy. This means that the double mutant is less destabilizing than the summed free energy of the two single mutants. The data suggest that by introducing a double mutation to alanine, some favourable interactions might be recovered as compared to both single mutations.

Surprisingly, when both of the glycines in the GxxxG motif are mutated to alanine residues, the GpATM helices continues to dimerize, with a stabilizing coupling energy. This provides evidence for an AxxxA motif capable of inducing dimerization of trans membrane helices. When leucine is introduced at the motif positions, dimerization does occur, but at a lower affinity than the alanine double mutant.

With all this data gathered, the researchers conclude that the GxxxG motif is neither sufficient or necessary for dimerization. They base this conclusion on the fact that some peptides with an intact GxxxG motif do not appear to dimerize, while peptides without the motif do form dimers. It is also surprising that the largest degree of coupling is observed between helices that are relatively far away from each other. This suggest that this long range coupling is caused by larger conformational changes of the helices, for example the crossing angle (see figure 6).

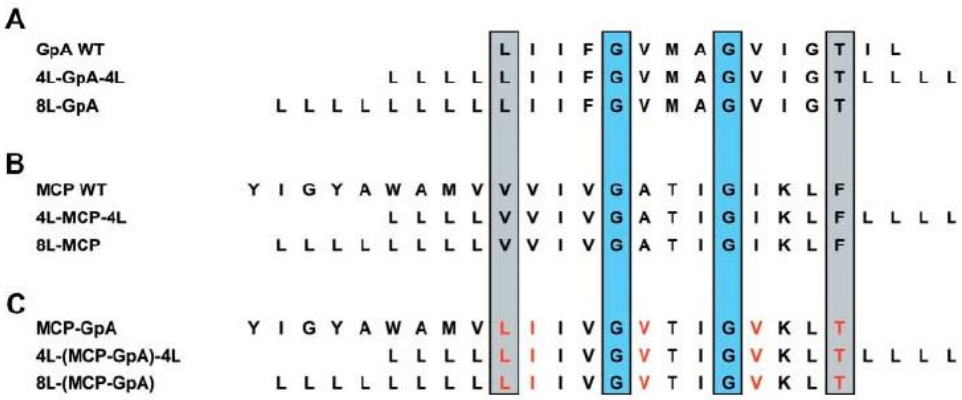

Figure 7. A multiple sequence alignment of the sequences used in this study. A. These sequences are derived from GpATM. The upper one is the wt sequence. The middle sequence has four leucine residues added at both the C- and N-terminal of the motif. The lower sequence has eight leucine residues added to the N-terminal of the motif. B. These sequences contain the GxxxG motif, but contain the context of the MCP protein. C. These are the chimera sequences. The main sequence is derived from the MCP protein, but the motif flanking residues (in red) are replaced by the residues that are found on these positions in the GpATM sequence[14].

Position of the GxxxG motif

One factor that may be involved in the stability of dimer formation has been largely ignored by most studies. Johnson et al.[14] post the question whether the position of the GxxxG motif within a transmembrane part of a protein influences the dimerization. To address this question, two different proteins were used as starting points, the GpATM peptide and the bacteriophage M13 coat protein (MCP). In glycophorin A, the motif is positioned close to the center of the trans membrane helix, whereas the motif is located near the C-terminal end of the trans membrane helix of MCP. The motifs within these trans membrane helixes were repositioned by adding sets of four leucines to the ends of the helix; either one set at both ends of the helix (4L-4L) or two sets (eight leucine residues) on the N-terminal (8L-)(see picture 7). Additionally, chimeric peptides were created, with the MCP sequence as a starting point, but with all of the interfacial residues of the GpATM peptide. The propensity of the helices to dimerize was again measured with the aid of the TOXCAT system.

First, the wildtype sequences of the GpATM, MCP and the chimera peptides are tested as positive controls. The wt MCP dimerizes with an affinity that is approximately 30% of the wt GpATM affinity, as expected. The chimer, which has the same interfacial residues as the wt GpATM sequence, dimerizes with an affinity that lies between that of the MCP and GpATM peptides. Secondly, the position of the motif in the GpATM sequence is evaluated. When the wt and the 4L-4L peptides are compared (in which the motif is similarly positioned), no significantly difference is observed. The 8L-sequence, however, shows a twofold lower affinity for dimerization compared to the wt and the 4L-4L sequences. A similar story applies to the MCP peptides. The peptides in which the motif is positioned in a comparable way (wt and the 8L- variant) show no big difference in dimerization. When the motif is shifted towards the center however, a threefold increase in dimer affinity was observed. Additionally, when the flanking residues of GxxxG motif are placed in this 4L-MCP-4L sequence (creating the chimer sequence with four leucine residues at each peptide end), the dimer affinity is restored to a level that is indistinguishable with the affinity of the 4L-GpATM-4L peptide.

Summarized, the affinity for dimerization is the highest when the GxxxG motif is located in the center of the transmembrane helix. Additionally, flanking interfacial residues to the GxxxG motif seem to influence the affinity for dimerization as well. A possible explanation for the preferred position of the GxxxG motif inside the helices is the fact that the dielectric constant is the lowest inside the middle of the membrane. Hydrogen bonds are the strongest when the environment has a low dielectric constant, this this might increase the affinity for dimerization.

In addition to the GxxxG motif, several other motifs were also identified to be inducing the oligomerization of transmembrane α-helices. In a study of Dawson et al.[15] a randomized library of peptides was created. In these peptides, seven residues were varied with the most frequent occurring residues in transmembrane helices, but glycine was omitted in these variable residues to prevent the identification of the GxxxG motif (see figure 8). The results were again gathered with the TOXCAT system.

The most notable results were the strong selection for a threonine residue at position 13 and serine residues at position 9 and 10 of the peptides. Additional mutational and modelling experiments suggested that the driving force of these peptides are interhelical hydrogen bonds. Threonine at position 13 is able to accept a hydrogen bond from a serine residue at position 9 of the opposite helix. Proline was also found to be conserved in the library, at position 5 or 6.

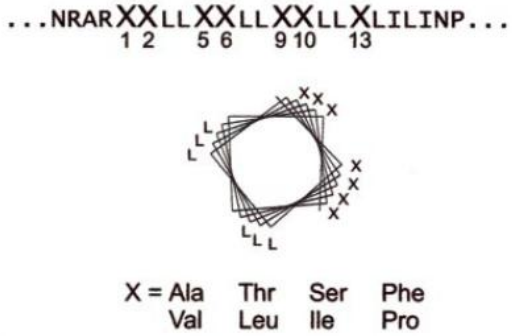

Figure 8. Residues on the positions tagged with an X were mutated into the eight most common residues in transmembrane helices. All the mutated residues are on the same side of the helix[15].

All sequences including a proline at this position also included a serine or threonine residue spaced 4 residues from the proline residue. When proline is inserted into a helix, the carbonyl oxygen atom at position *i*-4 is freed. This atom is thought to be able to form interhelical hydrogen bonds with a threonine on the opposing helix which stabilizes the dimer.

The GxxxG motif is also present in water soluble proteins

The prevalence of the GxxxG motif is not confined to only transmembrane proteins. A study by Kleiger et al.[16] shows that the motif is also present in α-helices of water soluble proteins. In this study, the researches searched for the GxxxG motif in α-helices inside the nonreduntant Protein DataBank. In total 152 proteins were found containing the helix motif, which is almost 10% of all the proteins searched. When the significance is calculated for these results, the motif is found 41 ± 9% more often than expected (if glycine residues were spaced randomly inside the helices). Intriguingly, in addition to the GxxxG motif, the motif AxxxA is also widespread among the α-helices of soluble proteins. This motif is found in 1533 of the proteins, 30 ± 3% more often than expected. To get a better understanding of the function of these motifs, the structures of the proteins containing the motifs were analysed. In 97 (56 %) of the proteins containing the GxxxG motif an interaction is observed between the helix containing the motif and another helix and in 26 of these cases the interaction is highly similar to that of Glycophorin A. When the distance between the helices is evaluated, only the distances between helices with the GxxxG motif appear to be significantly shorter than the average distances. The distance between helices with the AxxxA motif is not different from average values. This is in accordance to previously statements that the GxxxG motif allows the close packing of α-helices.

The researchers also performed a search in the predicted α-helices of 24 fully sequenced genomes. The results indicate that the motifs occur more frequently than expected in most of these genomes. The enrichment of AxxxA is especially high in the genomes of thermophiles, whereas there does not seem to be such a correlation with the GxxxG motif. The researchers state that the AxxxA motif may be important to provide stability to the proteins, but are unable to give a good explanation about the mechanisms of this stabilising interaction.

Prompted by this discovery of the AxxxA motif in soluble protein helices, another group of researchers posed the question whether the GxxxG and AxxxA motifs could be more generally defined as two small residues separated by three other residues[17]. To address this question, the researchers substituted either of the two glycine residues of the motif in the transmembrane part of Glycophorin A to the small residues alanine and serine, or to the bulky amino acid Isoleucine. To measure the dimerization, the GALLEX system was used. This system is similar to the TOXCAT system described earlier, but the transcription factor LexA (wildtype and a mutated version) is used instead of ToxR, which enables the researchers to measure both homo- and heterodimerization. Substitution of one of both glycines by isoleucine was found to be highly disruptive to the homodimerization of GpATM, which was expected. Substitutions by alanine were only slightly less disruptive than isoleucine substitutions. When one of the glycines is replaced by a serine, the dimerization is also reduced, but to a lower extend than with the alanine mutants. This result is in consensus with the results obtained by Russ et al. described earlier. Serine is larger than alanine and should disrupt the tight packing of the two α-helices stronger. Therefore, there must be another interaction that compensates for the loss in tight packing. Serine possesses a hydroxyl group, which might interact with the hydroxyl group of the other serine residue or with backbone atoms on the other helix. Additional experiments in which sets of two different GpATM helices were allowed to heterodimerize, pointed towards the conclusion that serine residues are mainly involved in hydrogen bond interactions with the carbonyl oxygen atoms on the adjacent helix. In soluble proteins, this might lead to breaking of the α-helix because the backbone carbonyl oxygen atoms are usually involved in hydrogen bonds stabilizing the helix, but as the secondary structure of α-helical

membrane proteins is quite stable (the two stage model)[18], this may not be the case for these proteins.

Prevalence of the GxxxG motif in membrane proteins of which the structure is determined

A lot of research has been done on the GxxxG motif, mainly in the context of the Glycophorin A protein. These researches have given valuable information about the mechanisms involved in the dimerization of α-helices and role of the GxxxG motif in the process. But exactly how does all of this research relate to natural proteins of which the high resolution structures have been determined?

In a paper by Kim et al.[19] which focusses on glycine zippers (extended GxxxG motifs, e.g. GxxxGxxxG) a number of structures of unrelated channel proteins (Potassium channel pore-lining channel KcsA, mechanosensitive channel of large conductance MscL, vacuolating toxin A anion selective channel VacA and mechanosensitive channel of small conductance MscS) are discussed. In all these proteins a zipper motif is present, either with glycine on the main positions or one of the other small residues, alanine or threonine. All these channel proteins form homo-oligomers and the glycine zippers seem to provide a surface for the helices to get in close contact with each other.

The GxxxG motif is also present in the erbB2 receptor, in which is plays a rather unique role[20]. This receptor has one transmembrane helix, and forms homo- or heterodimers with homologue proteins when substrate is bound. Within this transmembrane helix, two dimerization motifs are present. The first and most stable dimerization motif is a GxxxG motif. The conformation of dimers using this motif is related to an inactive state. When substrate is bound, the protein switches to another conformation in which the other dimerization motif is utilized. This state is thought to be the active state. This second dimerization motif is similar to the GxxxG motif, but is far less strictly defined, as the glycines are often replaced by serine, threonine or alanine. The researchers suspect that the motifs form a relatively flat surface which facilitates close packing of the helices, as seen in previous work on the GxxxG motif.

In ECF-transporters an AxxxA motif is thought to be mediating the interaction between the energizing module and the substrate specific component. These transporters consist of an energizing module, which is a complex of one transmembrane protein and two identical or homologue periplastic nucleotide binding domains, and a substrate specific component (S-component). Several S-components can compete for the same energizing module. When the AxxxA motif (positioned on the S-component, ThiT in this study) is mutated to either WxxxA or AxxxW, the complex falls apart and the transport of the substrate is completely abolished, indicating that this motif is involved in the formation of the complex[21].

With all the data gathered from the studies discussed above, it is tempting to scan some structures of α-helical membrane-spanning proteins for the presence of the GxxxG motif. Although the resources available to scan the structures are very limited, it has yielded in some surprising results. To find GxxxG motifs in the structure of membrane proteins, the program Cn3D was used and a pattern search for 'GXXXG' was performed. A first attempt was made on the structure of the ClC chloride channel protein from *Escherichia coli*[22](PDB ID: 1KPK). This channel protein catalyses the selective flow of Cl^- ions across the cell membrane. The protein forms a homodimer inside the membrane (see figure 9). The first question asked was whether GxxxG motifs are present inside the dimer interface. The dimer interface consists of four helices. Neither of these helices appear to have a GxxxG motif or a (small)xxx(small) motif. However, this protein is polytopic; its structure consists of multiple membrane-spanning α-helices, which interact with each other to form the tertiary structure. When the whole structure is scanned for the presence of the GxxxG motif, multiple of these motifs are found on the interfaces between these helices. In total, the motif is found nine times and seven of the motifs seem to be involved into helix-helix interactions within the monomer. This indicates that the GxxxG motif is not only able to assist in interfaces between peptides but is also able to aid in the formation of the tertiary structure of proteins.

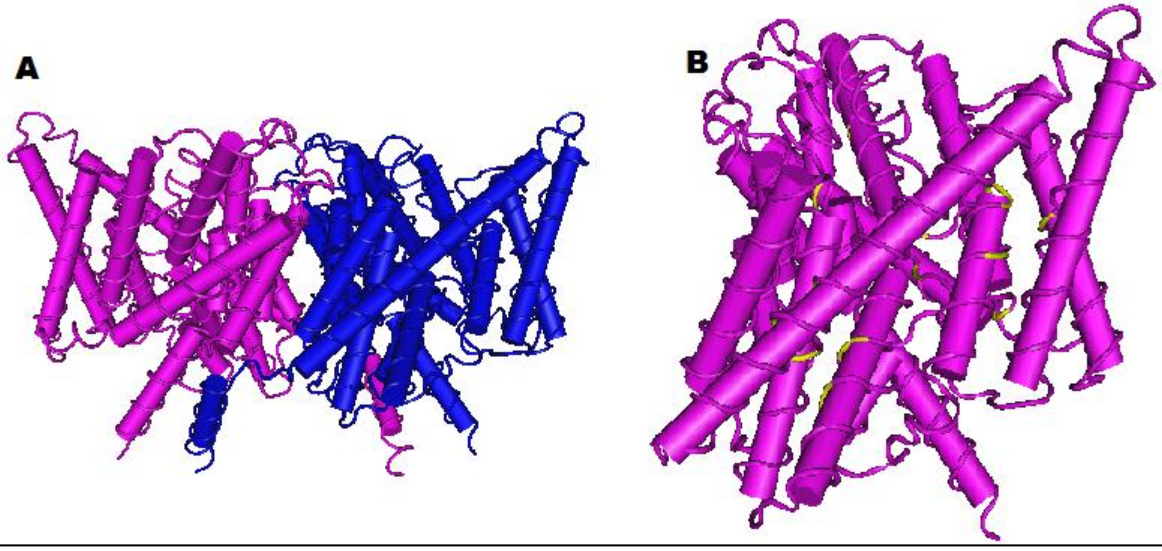

Figure 9. A. The complete ClC transporter. The dimer is located inside the membrane. The upper side is located at the extracellular side of the membrane, the lower part on the intracellular side. B. One monomer of the ClC transpoter. The positions of the GxxxG motifs are indicated in yellow. Pictures are rendered by Cn3D.

The structure of a glutamate transporter homologue from *Pyrococcus horikoshii*[23](PDB ID: 1XFH) was also evaluated for the presence of the GxxxG motif (see figure 10). This transporter catalyses the uptake of glutamate into the cell. The protein is an integral membrane protein and forms homo-trimers inside the membrane. Again, no GxxxG motifs are present on the interfaces between the subunits. However, three GxxxG were found within the monomers. Two of these motifs are positioned inside an α-helix, and could be contributing to the tertiary structure. Surprisingly, one of these motifs consists of an extended GxxxG motif: GxxxGxxxGxxxG. This glycine zipper might provide extra stability to the protein.

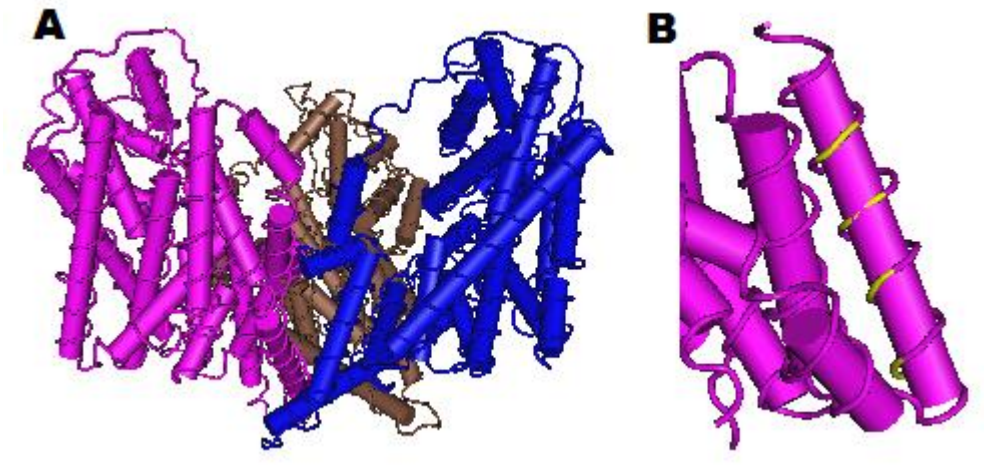

Figure 10. A. The complete structure of the glutamate transporter homologue in trimeric form (PDB ID: 1XFH). B. A highlight of the found GxxxGxxxGxxxG motif. The backbone atoms of the glycines are colored yellow. Pictures rendered by Cn3D.

Finally, aquaporin 4[24] (PDB ID: 3GD8) was examined for the presence of the GxxxG motif. This human protein is expressed in the brain and has a function in the maintenance of the cerebral water balance. This protein forms homotetramers inside the membrane (see figure 11). When the sequence is scanned for the GxxxG motif three hits are found. Two of these hits are inside an α-helix. Again, none of the found GxxxG motifs are on the interface between the subunits. The two motifs that were found seem to be stabilizing the structure of the monomer itself, as they were found at the location where two helices cross each other. When the aquaporin 5 protein from *Homo sapiens*[25] (PDB ID: 3D9S) is scanned for the GxxxG motif, it is found six times. In three of these occasions it appears to be involved in a helix-helix interaction, but again only between helices within a monomer.

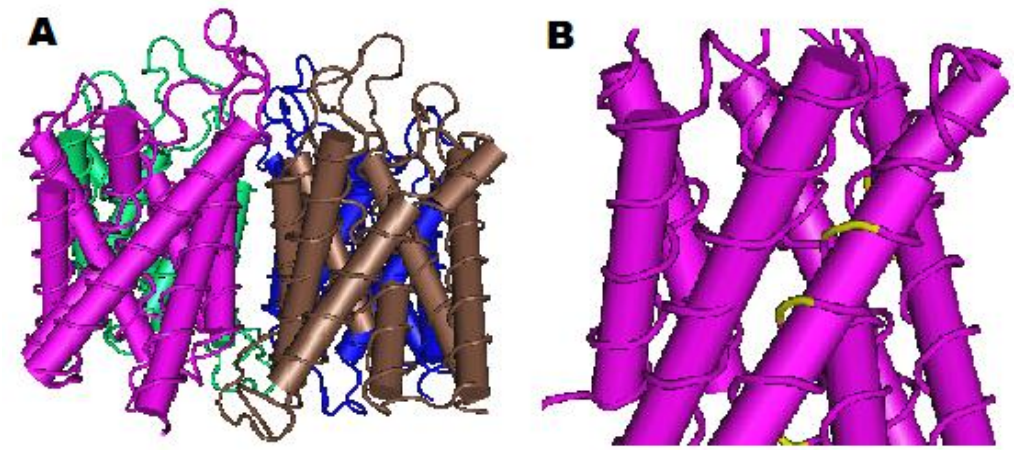

Figure 11. A. The complete structure of the aquaporin 4 protein in tetrameric form. B. A zoomed in picture of one of the GxxxG motifs found in the transporter. The GxxxG motif is indicated by the yellow backbone ribbon structure. Pictures are rendered by Cn3D.

Conclusions

On the basis of all the work that has been performed on Glycophorin A and several other peptides, it is not possible to ignore the importance of the GxxxG motif. Although the motif is not capable of inducing dimerization by itself, it proves to be a powerful sequence motif to predict helix-helix interactions. The prevalence of the GxxxG motif in solved structures of membrane proteins only

seems to strengthen this conclusion. Though the motif in the cases studied is not involved in the oligomerization of the proteins, it does seem to be involved in the interactions between helices inside the monomer. The results from the analyses performed on the ClC transporter, the glutamate transporter homologue and the aquaporin 4 and 5 proteins within this paper should however not be overestimated, as they do not represent a significant sample of all the membrane proteins.

Overall the GxxxG motif does seem to be a powerful motif to predict helix-helix interactions. However, the absence of this motif does not imply in any way the absence of an interaction between helices. Research has shown that there are also other motifs capable of stimulating interhelical interactions. More and more high-resolution structures of membrane proteins are becoming available. This will strongly benefit the research on membrane proteins and the interactions of these proteins.

References

1. Wallin E, von Heijne G. Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Sci*. 1998;7(4):1029-1038.

2. Therien A, Grant F, Deber C. Interhelical hydrogen bonds in the CFTR membrane domain. *Nat Struct Biol*. 2001;8(7):597-601.

3. BORMANN B, KNOWLES W, MARCHESI V. Synthetic peptides mimic the assembly of transmembrane glycoproteins. *J Biol Chem*. 1989;264(7):4033-4037.

4. FURTHMAYR H, MARCHESI V. Subunit structure of human erythrocyte glycophorin-a. *Biochemistry (N Y)*. 1976;15(5):1137-1144.

5. Popot JL, Engelman DM. Membrane protein folding and oligomerization: The two-stage model. *Biochemistry*. 1990;29(17):4031-4037.

6. MacKenzie K, Prestegard J, Engelman D. A transmembrane helix dimer: Structure and implications. *Science*. 1997;276(5309):131-133.

7. Lemmon Ma, Flanagan JM, Treutlein HR, Zhang J, Engelman DM. Sequence specificity in the dimerization of transmembrane alpha-helices. *Biochemistry (N Y)*. 1992;31(51):12719-12725.

8. LEMMON M, TREUTLEIN H, ADAMS P, BRUNGER A, ENGELMAN D. A dimerization motif for transmembrane alpha-helices. *Nat Struct Biol*. 1994;1(3):157-163.

9. Langosch D, Brosig B, Kolmar H, Fritz H. Dimerisation of the glycophorin a transmembrane segment in membranes probed with the ToxR transcription activator. *J Mol Biol*. 1996;263(4):525-530.

10. KOLMAR H, FRISCH C, KLEEMANN G, GOTZE K, STEVENS F, FRITZ H. Dimerization of bence-jones proteins - linking the rate of transcription from an escherichia-coli promoter to the association constant of rei(v). *Biol Chem Hoppe-Seyler*. 1994;375(1):61-70.

11. Russ WP, Engelman DM. The GxxxG motif: A framework for transmembrane helix-helix association. *J Mol Biol*. 2000;296(3):911-919.

12. Doura AK, Kobus FJ, Dubrovsky L, Hibbard E, Fleming KG. Sequence context modulates the stability of a GxxxG-mediated transmembrane helix-helix dimer. *J Mol Biol*. 2004;341(4):991-998.

13. Doura AK, Fleming KG. Complex interactions at the helix-helix interface stabilize the glycophorin A transmembrane dimer. *J Mol Biol*. 2004;343(5):1487-1497.

14. Johnson R, Rath A. The position of the gly-xxx-gly motif in transmembrane segments modulates dimer affinity. *Biochemistry and cell*. 2006;1012:1006-1012.

15. Dawson JP, Weinger JS, Engelman DM. Motifs of serine and threonine can drive association of transmembrane helices. *J Mol Biol*. 2002;316(3):799-805.

16. Kleiger G, Grothe R, Mallick P, Eisenberg D. GXXXG and AXXXA: Common alpha-helical interaction motifs in proteins, particularly in extremophiles. *Biochemistry (N Y)*. 2002;41(19):5990-5997.

17. Schneider D, Engelman DM. Motifs of two small residues can assist but are not sufficient to mediate transmembrane helix interactions. *J Mol Biol*. 2004;343(4):799-804.

18. POPOT J, ENGELMAN D. Membrane-protein folding and oligomerization - the 2-stage model. *Biochemistry (N Y)*. 1990;29(17):4031-4037.

19. Kim S, Jeon T, Oberai A, Yang D, Schmidt JJ, Bowie JU. Transmembrane glycine zippers: Physiological and pathological roles in membrane proteins. *Proc Natl Acad Sci U S A*. 2005;102(40):14278-14283.

20. Fleishman S, Schlessinger J, Ben-Tal N. A putative molecular-activation switch in the transmembrane domain of erbB2. *Proc Natl Acad Sci U S A*. 2002;99(25):15937-15940.

21. Erkens GB, Berntsson RP, Fulyani F, et al. The structural basis of modularity in ECF-type ABC transporters. *Nat Struct Mol Biol*. 2011;18(7):755-U23.

22. Dutzler R, Campbell E, Cadene M, Chait B, MacKinnon R. X-ray structure of a CIC chloride channel at 3.0 angstrom reveals the molecular basis of anion selectivity. *Nature*. 2002;415(6869):287-294.

23. Yernool D, Boudker O, Jin Y, Gouaux E. Structure of a glutamate transporter homologue from pyrococcus horikoshii. *Nature*. 2004;431(7010):811-818.

24. Ho JD, Yeh R, Sandstrom A, et al. Crystal structure of human aquaporin 4 at 1.8 angstrom and its mechanism of conductance. *Proc Natl Acad Sci U S A*. 2009;106(18):7437-7442.

25. Horsefield R, Norden K, Fellert M, et al. High-resolution x-ray structure of human aquaporin 5. *Proc Natl Acad Sci U S A*. 2008;105(36):13327-13332.

26. Duong MT, Jaszewski TM, Fleming KG, MacKenzie KR. Changes in apparent free energy of helix-helix dimerization in a biological membrane due to point mutations. *Journal of molecular biology*. 2007;371(2):422-34.