Packing of apolar amino acids is not a strong stabilizing force in transmembrane helix dimerization.

Gilbert J. Loiseau[1,2] and Alessandro Senes[1,@]

[1]Department of Biochemistry, University of Wisconsin-Madison, Madison, WI 53706

[2]Integrated Program in Biochemistry, University of Wisconsin-Madison, Madison, WI 53706

[@]Corresponding author: senes@wisc.edu

ORCID:

GJL: 0009-0003-8186-5275

AS: 0000-0002-3807-2275

Abstract

The factors that stabilize the folding and oligomerization of membrane proteins are still not well understood. In particular, it remains unclear how the tight and complementary packing between apolar side chains observed in the core of membrane proteins contributes to their stability. Complementary packing is a necessary feature since packing defects are generally destabilizing for membrane proteins. The question is the extent by which packing of apolar side chains – and the resulting van der Waals interactions – are a sufficient driving force for stabilizing the interaction between transmembrane helices in the absence of hydrogen bonding and polar interactions. We addressed this question with an approach based on high-throughout protein design and the homodimerization of single-pass helices as the model system. We designed hundreds of transmembrane helix dimers mediated by apolar packing in the backbone configurations that are most commonly found in membrane proteins. We assessed the association propensity of the designs in the membrane of *Escherichia coli* and found that they were most often monomeric or, at best, weakly dimeric. Conversely, a set of controls designed in the backbone configuration of the GAS_{right} motif, which is mediated by weak hydrogen bonds, displayed significantly higher dimerization propensity. The data suggest that packing of apolar side chains and van der Waals interactions may be a relatively weak force in driving transmembrane helix dimerization, unless highly optimized. It also confirms that GAS_{right} is a special configuration for achieving stability in membrane proteins.

Statement of Significance

The proteins that insert into the cellular membranes provide the cell with essential functionalities. However, how nature designed these proteins to fold into a precise conformation and become active is still poorly understood. In this article we investigate the role of one of the most important factors for protein folding, the van der Waals forces that produce favorable interactions when atoms are in close contact. We know that the formation of extensive van der Waals interactions through contacts are critical for proteins to be stable but we do not know if they are a strong driving force for folding a membrane protein. We addressed this question by designing hundreds of new membrane proteins design based on the optimization of these interatomic contacts and measuring if the resulting proteins come together to form stable pairs. The data suggest that the packing of apolar side chains and the resulting van der Waals interactions are not a strong force for supporting folding in the membrane.

Introduction

Like all proteins, helical integral membrane proteins need to fold into a specific tridimensional structure to support signaling, transport, catalysis, organization and plasticity and the many other essential functions they perform in the cell membranes. Folding in the membrane is a complex process that requires that proteins find their lowest energy state in an anisotropic environment that spans the hydrophobic core of the membrane, the polar lipid headgroup region and bulk water. Because of this complexity, it has been helpful to subdivide the folding process conceptually into two independent stages, the insertion of the transmembrane helices in the membrane (stage 1) and their association (stage 2) (1, 2). Even though membrane proteins occasionally incorporate within their fold helices that are too short or hydrophilic to be independently stable (3–11), the Two-Stage Model recognizes that most hydrophobic helices of sufficient length to span the membrane are thermodynamically stable in the membrane and can be considered independent folding domains.

The insertion of these helical domains in the membrane (stage 1) is primarily driven by the hydrophobic effect (12–16), i.e., the major driver of the folding of soluble proteins (3, 17, 18). Since this force is spent during insertion and is no longer available for folding, the question becomes what are the factors that promote the association of the helices in the second stage? Polar interactions such as electrostatics, cation→π and, especially, hydrogen bonding can in principle be energetically favorable in the apolar environment of the membrane (19, 20). Indeed, early studies established that individual polar side chains within an otherwise hydrophobic helix are sufficient to drive the self-association of TM helices (21–23). Further experiments suggest that the contributions of individual hydrogen bonds in membrane proteins are relatively modest but not insignificant (between 0.4 and ~2 kcal/mol) (24–30). Weak hydrogen bonds can also contribute to helix-helix interaction stability, such as those observed in the GAS_{right} interaction motif (31–33), involving activated Cα–H carbon donors (34–37) and backbone carbonyl acceptors on the opposing helix. However, in many cases, the interaction interfaces of TM helical pairs

do not display polar interactions. Polar amino acids are relatively rare in TM helices. Furthermore, the most common of them, the mildly polar Ser and Thr, tend to satisfy the hydrogen bonding potential of their hydroxyl group within their local backbone instead of participating in inter-helical hydrogen bonds (38, 39). Therefore, the binding interfaces of helical pairs are often mediated primarily by packing of apolar surfaces and van der Waals (vdW) interactions, raising the question of to what extent these forces can stabilize association.

There is no doubt that packing is a necessary feature of stable membrane proteins. The core of membrane proteins is tightly packed (40, 41) and disruption of packing, either by the creation of voids or the introduction of steric clashes, is generally destabilizing (24, 27, 42–45). Experimental measurements of the cost of creating packing defects within the core of bacteriorhodopsin suggest that the energetic contribution of vdW interactions in membrane proteins (approximately 20 cal mol^{-1} A^2) is similar to the values obtained for the water-soluble lysozyme (18 cal mol^{-1} A^2) (46). These studies indicate that packing defects are expensive and that tight packing is necessary for stability but do not demonstrate the extent to which packing is an actual driving force for helix-helix association, beyond being a mere necessity. A library screening analysis showed that TM helices composed of purely hydrophobic amino acids can activate the single-pass PDGFβ tyrosine receptor kinase, suggesting that apolar packing can be sufficient to mediate effective interactions leading to specific biological activity (47). However, the study was based on cellular response and did not confirm if the library elements formed a stable complex with the receptor. To date, the evidence that best supports the case that vdW interactions between apolar side chains can be a significant stabilizing force in membrane protein is a protein design study based on the scaffold of phospholamban, a regulator of the sarcoplasmic reticulum Ca^{2+} pump (48). The pentameric TM complex of phospholamban is mediated by a tightly packed hydrophobic C-terminal region, whereas the N-terminal region is dynamic and contains polar residues. When these polar residues were replaced with apolar side chains, the complex retained its stable pentameric assembly. The authors recognized a steric model based on a repeated LxxIxxx heptad motif (featuring β-branched amino acids at *e* positions and non-branched amino

acids at *g* positions) that favored tight packing in the pentameric configuration. These results provide compelling evidence that apolar side-chain packing can be sufficient as a strong driving force for oligomerization in the membrane. However, it was noted that the requirement for complementarity at the interface of this complex was very stringent, to the point that the pentamer can be completely destabilized by minor steric changes. This raises the question of whether the phospholamban model is a highly optimized case and to what extent packing can be a general stabilizing force across a variety of folds.

To answer this question, here we test the propensity of packing to promote association of TM helix dimers. Helical dimers are a simple and tractable system for studying the principles of TM helix-helix association. They are also extremely important biologically. For example, approximately 6% of the human proteome is composed of single-pass receptors, which often activate through ligand-induced stabilization of dimeric conformations (49–51). To understand the role of apolar packing across the variety of dimeric configurations observed in nature, here we use a high-throughput protein design strategy using a large number of starting backbones. We hypothesize is that if steric complementarity and vdW forces are sufficient for driving TM association, it should be possible to use this feature as a principle to design stable dimers.

We compare computationally predicted stability with the experimental dimerization propensity across a series of designed dimers, an approach similar to a previous study in which we explored the energetics of association of the common GAS_{right} dimerization motifs (52). That analysis indicated that vdW interactions and weak Cα–H hydrogen bonds combine to determine the stability of GAS_{right} dimers (52). In the model, the vdW interactions contribute a significant fraction of the predicted binding energetics of these dimers. This suggests that even if the hydrogen bonding component was somehow turned off, the vdW interactions should be sufficient for producing relatively strong association. However, here we found that designs based on packing alone are only weakly dimeric or monomeric. We conclude that, while it may be possible for vdW interactions to

produce highly stable structures in some cases, the contributions of packing are either too weak or their requirements are too stringent to be a general strong driving force in the dimerization in TM dimers.

Results and Discussion

Design strategy

To assess the general role of packing and vdW interactions in TM helix dimerization, we designed *de novo* a large number of dimers mediated primarily by packing and assessed their propensity to self-associate experimentally. To obtain the most favorable designs, they were based on backbones selected from the most common geometries observed in membrane proteins. The backbone geometries were identified by analyzing all pairs of interacting helices found in membrane proteins of known structure from the Orientations of Proteins in Membranes database, filtered by sequence similarity from the PDB (53, 54). Any two helices in close contact in the structures were considered as an individual helical pair. Since our designs are based on canonical α-helix conformation, unlike previous similar analyses (33, 55), we restricted the pairs to the subset that assumes strictly regular α-helix conformation (excluding all segments that contained kinks, curvature, 3_{10} helix and any other deviations from a standard α-helix) before the conformational parameters of each helical pair were computed (i.e., the crossing angle θ, the inter-helical distance *d*, the axial rotation of each helix ω, and the displacement in along the helical axis or *Z*-shift, Fig. 1a).

Fig. 1b illustrates the resulting helix-helix interaction landscape for parallel helical pairs plotted as a function of inter-helical distance *d* and the crossing angle θ. Two major high-density regions are present. In the left-handed region (positive crossing angle θ), helical pairs are found to interact most frequently with an inter-helical distance between 8.5 to 10 Å of and a crossing angle of 20 to 40°. We call this region the "Left" region (highlighted as a box in Fig. 1b). The right-handed region presents a broader range of populated inter-helical distances, from 6.5 to over 10 Å and crossing angles between -30 and -60°. We selected the maximum density

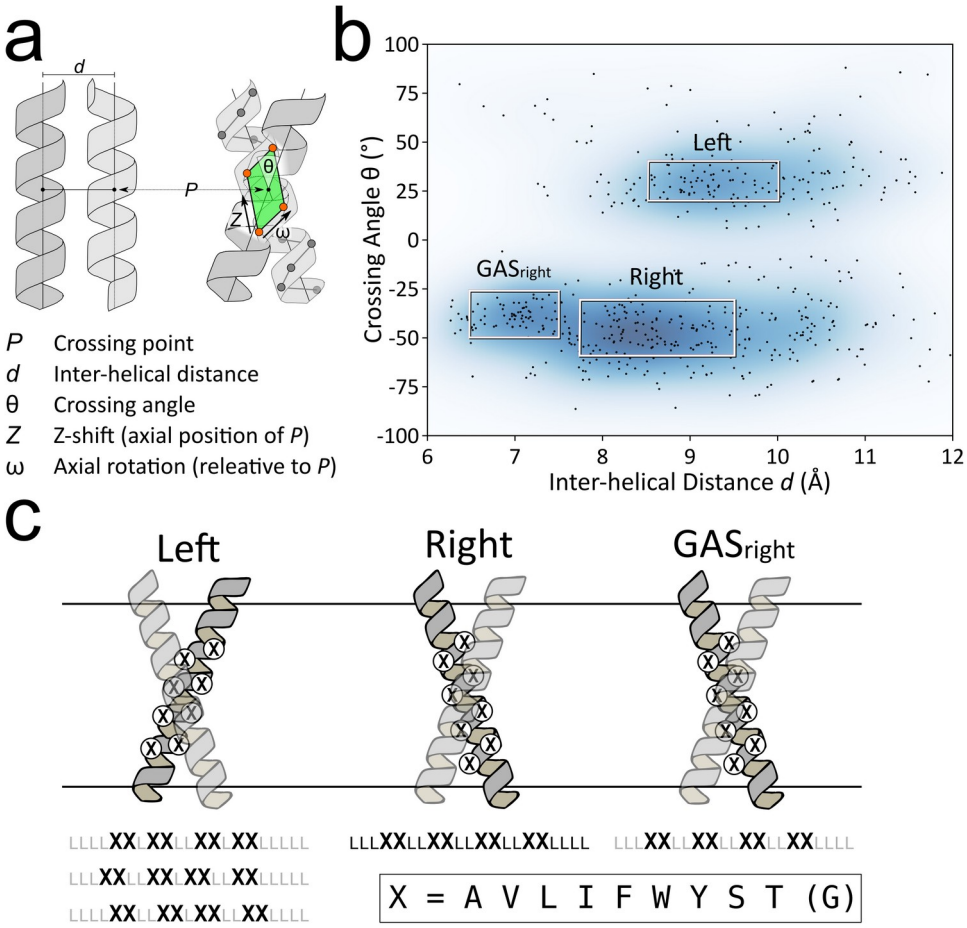


Fig. 1. Protein design strategy. a) Inter-helical parameters defining the geometry of a dimer: the inter-helical distance *d*; the crossing angle θ; the rotation of the helix around its axis ω; and the vertical position *Z* of the point of closest approach *P*. The ω and *Z* coordinates are relative to the green parallelogram (the unit cell in the helical lattice) formed by the four Cα of closest approach. b) Distribution of inter-helical distance *d* and crossing angle θ found in pairs of canonical α-helices in membrane proteins of known structure extracted from the Orientations of Proteins in membranes (OPM). The coordinates are overlaid over their kernel density estimation. c) All designs were based on a 21 amino acid backbone, with the crossing point placed in the middle of the membrane, 8 variable interfacial positions ("X") and all non-interfacial positions set to leucine ("L"). The placement of the interfacial positions was dependent on the ω and *Z* coordinates for designs in the Left region (see supplementary Fig. S1).

region with inter-helical distance between 7.75 and 9.5 Å and a -30 to -60° crossing angle (named "Right") as the second source of helical pair backbones.

The section of the right-handed region with the closest inter-helical distance (*d* = 6.5 to 7.5 Å, θ = -25 to -55°) corresponds to the GAS_{right} motif, the well-known helix-helix interaction motif (32, 56). GAS_{right} has a sequence

signature of small amino acids (Gly, Ala, Ser) at the dimer interface, forming its characteristic GxxxG and similar sequence motifs (GxxxA, SxxxG, etc.). These small amino acids allow for the short inter-helical distance that brings the helical backbones in contact, which results in the formation of characteristic networks of inter-helical weak hydrogen bonds between activated Cα–H carbon donors and backbone carbonyl acceptor groups on the opposing helix (Cα–H···O=C hydrogen bonds) (33). Since these weak hydrogen bonds contribute to stability in GAS_{right} (52) this region is not considered in our analysis of packing as a driving force. Instead, these designs were used as a positive control to validate the methodology applied to designs in the "Left" and "Right" regions.

Similarly to our previous GAS_{right} study (52), we based our designs on a 21-amino acid poly-Leu backbone with the crossing point between the two monomers placed near the center of the membrane, as illustrated in Fig. 1c. For each dimer, 8 positions at the helix-helix interface were set as variable during the design, whereas all non-interfacial positions were standardized to Leucine, a practice that reduces variability in overall hydrophobicity and helps to obtain relatively consistent expression levels and insertion into the membrane during the experimental phase (supplementary Fig. S2). The right-handed designs ("Right" and "GAS_{right}") follow a typical pattern alternating two variable interfacial positions (**x**) with two fixed positions (L), resulting in a LLL**xx**LL**xx**LL**xx**LL**xx**LLLL sequence (32, 52, 56). In the Left region, the interfacial positions follow the LxxLLxL heptad repeat typical of knobs-into-holes patterns found in leucine zippers (57–59). However, the specific registry of the interface positions within the 21-amino acid helix is dependent on the helical pair coordinates (*Z*-shift and axial rotation, supplemental Fig. S1). For this reason, three different interfacial patterns, all consisting of 8 variable positions, were adopted for designs in this region, as illustrated in Fig. 1c.

The selection of amino acids used at the variable positions at the interaction interfaces was limited to the most commonly found amino acids in membrane proteins. These included the aliphatic Ala, Val, Leu and Ile, which together represent 43% of the composition of TM helices (60). Also included were the aromatic residues

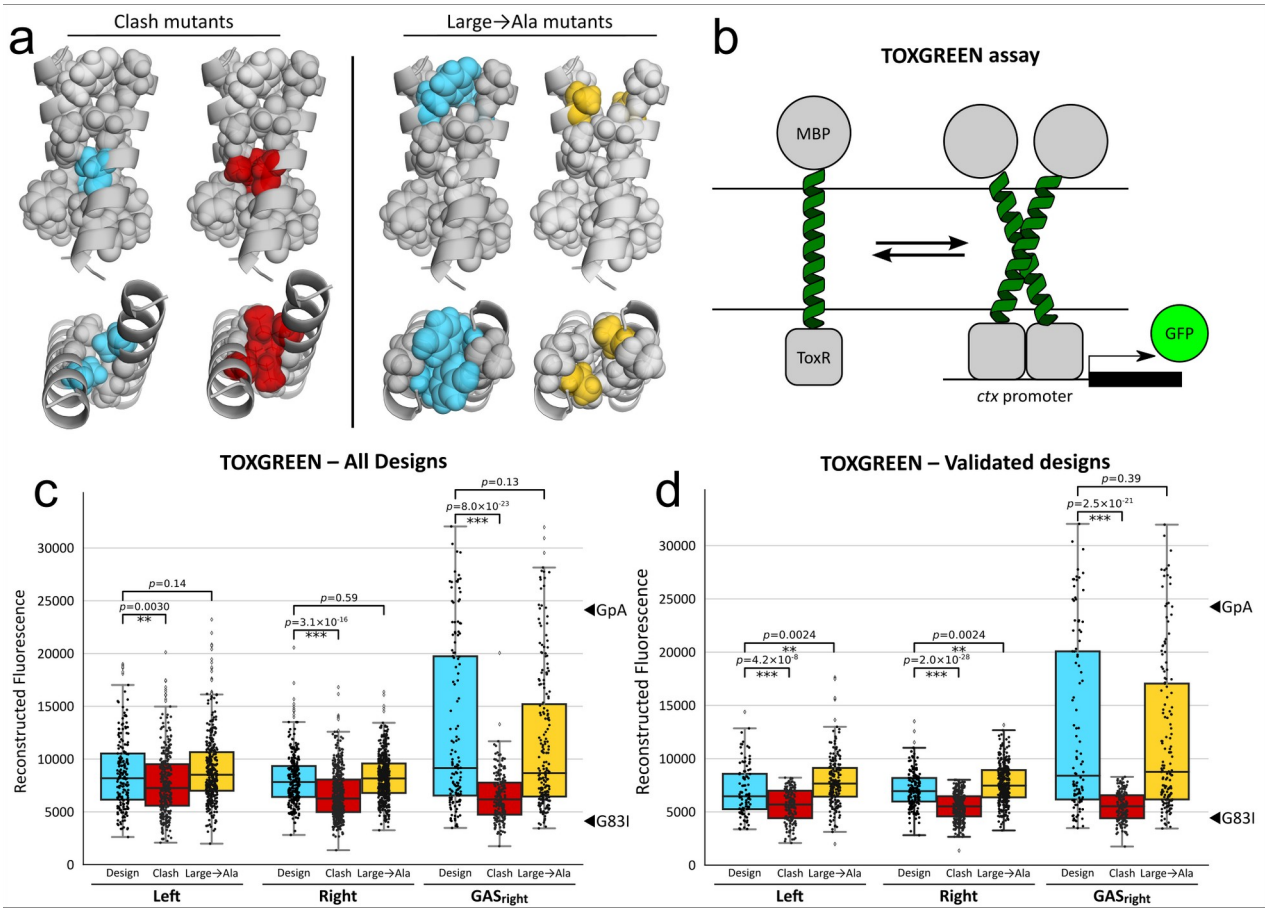


Fig. 2. Experimental validation of designed dimers. a) A set of validation mutations was designed for each design. "Clash mutants" are changes to interfacial small amino acids to isoleucine created to introduce significant steric clashes within the designed structure. "Large→Ala" mutants were created to reduce packing at the helix-helix interface by changing large amino acid to Ala. b) TOXGREEN is an *in vivo* genetic reporter assay that measures oligomerization propensities of TM helices relative to known standards. In the assay, the TM domain under investigation is fused to the ToxR transcriptional activator. TM helix-helix association results in ToxR's binding to the *ctx* promoter and the consequent in the expression of GFP. In this work, the TOXGREEN assay was used in high-throughput using its sort-seq variant. c) Measurement of dimerization of the designs in the membrane of *E. coli*. The figure displays the reconstructed TOXGREEN sort-seq fluorescence of Left, Right and GAS_{right} designs (blue) and their Clash (red) and Large→Ala mutants. The reconstructed fluorescence of the standards GpA (known strong dimer) and G83I (known monomer) are indicated on the side. The vdW-based designs in the Left and Right regions display a low propensity to dimerize. The GAS_{right} designs include constructs with TOXGREEN signals similar to or higher than the stable GpA standards. The introduction of clash mutations produces a significant reduction of dimerization. d) Same representation as panel c for designs validated by their class mutants (i.e., those designs whose clash mutants are in the monomeric range).

Phe, Trp and Tyr (14% of the composition, in total). Trp and Tyr have a tendency to be enriched in the membrane head-group region and Tyr has a tendency to "snorkel", and thus expose its η-hydroxyl group to water in the

head-group region (61). Finally, we included Ser and Thr, since they are also common in membrane proteins (7% each). These two amino acids contain a hydroxyl group and could potentially form inter-helical hydrogen bonds. However, their side chains tend to satisfy the potential by forming intra-helical hydrogen bonds with carbonyl groups at *i*-3 and *i*-4 (38, 39) and are thus compatible with a helix-helix interaction interface mediated primarily by non-polar interactions. Gly is common in membrane proteins (8%) but it was excluded from the Left and Right designs because it plays a critical role in the interface of the GAS_{right} dimerization motif, to exclude the accidental formation of GAS_{right} structures (Fig. 1b).

The design algorithm started with selecting an initial backbone with crossing angle and inter-helical distance values randomly assigned within the boundaries of each region, as well as random *Z*-shift and axial rotation coordinates. The side chains were selected and optimized with a Monte Carlo procedure, producing the final protein sequence. The structure was further refined by iterative cycles of consisting local backbone moves followed by side-chain optimization (19, 62–65). As in our previous GAS_{right} prediction studies (32, 52), each design was evaluated using a computational energy score composed of a vdW function (66), hydrogen bonding (67), and the IMM1 implicit membrane solvation (68). The procedure resulted in 1045 designed sequences divided across the "Left", "Right" and "GAS_{right}" regions, covering a range of both structure and stability.

To control for the possibility that constructs may associate through different modes than their designed structures, we designed point mutations expected to disrupt association (Fig. 2a). Small interfacial amino acids were changed to the larger Isoleucine residues to create significant clashes within the designed structure ("Clash mutants"). In addition, large amino acids were changed to Ala ("Large→Ala") to reduce packing at the helix-helix interface. Both Clash and Large→Ala mutants were expected to potentially decrease association, with Clash mutants expected to result in drastic decreases as observed in previous studies (52, 69–73).

Experimental strategy

The propensity for dimerization of the designs was experimentally assessed using the high-throughput "TOXGREEN sort-seq" membrane oligomerization assay (74). TOXGREEN (75, 76) is an *in vivo* assay that reports the self-association of TM helices in the biological membrane of *Escherichia coli* (Fig. 2b). The assay is based on a chimeric protein in which the TM domain of interest is fused to the ToxR transcriptional activator. Dimerization of these constructs in the inner membrane induces ToxR binding to a specific promoter, resulting in the expression of GFP. Quantification of fluorescence provides an indication of a TM domain's propensity for oligomerization. We recently demonstrated that the TOXGREEN signal of a series of GAS_{right} dimers based on a standardized poly-Leu backbone correlates with their thermodynamic stability measured *in vitro* (77). The recently developed high-throughput version of the assay, TOXGREEN *sort-seq,* is suitable for evaluating large libraries of TM helices (74). It is based on Fluorescence-Activated Cell Sorting (FACS), which separates cells in different bins based on GFP expression. The relative frequency of each construct in the various bins is then measured using Next-Generation Sequencing (NGS) and the distribution profile is used to reconstruct the GFP fluorescence of each individual construct.

The direct comparison of the relative dimerization propensity of the various constructs assumes that these are expressed at similar levels in the bacterial membrane. As mentioned in the previous section, the adoption of a standardized poly-Leu backbone substantially reduces expression variability in TOXGREEN (52). The high-throughput workflow of the present analysis does not allow for screening each construct individually and eliminating those that display anomalous expression levels. We verified by Western blot a random subset of 21 constructs, illustrated in supplementary Fig. S2. The majority of the constructs display relatively similar levels of expression. Some variability is expected (observable most notably in constructs R2, R4 and L10, whose bands display reduced density), deriving from inhomogeneity in the rates of protein synthesis, membrane insertion or

degradation. While expression variability introduces noise in the apparent dimerization propensities, the large scale of the analysis mitigates this issue, enabling the assessment of statistical trends as long as the majority of the constructs are relatively homogeneous.

The Left and Right designs are weak in comparison to GAS_{right}

The TOXGREEN *sort-seq* analysis recovered 938 of the 1045 (89%) designed sequences. A total of 234 constructs did not grow in maltose minimal media, suggesting that these constructs were inserted poorly in the membrane (76). For 91 designs no validation mutants were found in the reconstruction and were also discarded, resulting in a final set of 613 designs, consisting of 198 Left, 277 Right and 138 GAS_{right} designs (supplemental Fig. S3). The distribution of the computational energy scores of these designs is illustrated in supplemental Fig. S4. The computational energy score of the Right designs was overall less favorable (average -18 kcal/mol) compared to the Left and GAS_{right} (both around -28 kcal/mol), suggesting that the latter backbone configurations are more designable. The structural models of the Right designs also displayed a smaller interface average surface area (median 773 Å), compared to the Left (917 Å) and GAS_{right} designs (858 Å, supplemental Fig. S5a).

Fig. 2c shows the distribution of the reconstructed GFP fluorescence of the designs (blue). The GFP fluorescence values of the standard TOXGREEN controls (the strong dimer glycophorin A and its monomeric G83I variants) are indicated as a reference. We found that the majority of the designs in the Left and Right regions are in a range that indicates that they are monomeric or have a weak propensity to oligomerize. Notably, the GAS_{right} designs span a wider range of fluorescence, including constructs with signals comparable to or higher than the strong GpA dimer. To provide a measure, approximately 40% of the GAS_{right} designs have a reconstructed fluorescence above 60% of GpA (14,500 fluorescent units in the figure), compared to only 8% and 4% of the Left and Right designs. The data confirms the strong propensity of GAS_{right} for self-association and suggests that vdW packing alone is a weak driver for dimerization.

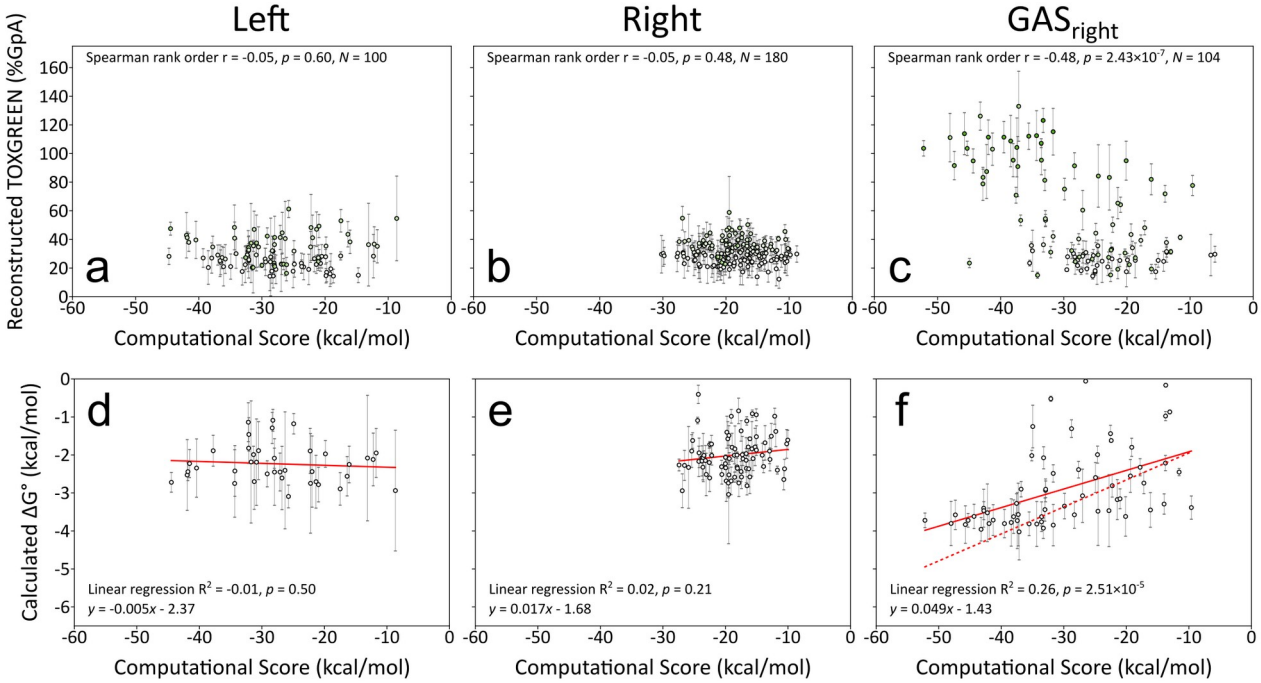


Fig. 3. Comparison between computational and extrapolated experimental energetics of the designs. a-c) Comparison between TOXGREEN signal (normalized to the GpA standard) and computational energy score. No trend is observed for the low stability Left and Right designs, whereas a statistically significant relationship is present for the designs with GAS$_{right}$ configuration. d-f) Computational score plotted against an empirical conversion of TOXGREEN signal to ΔG° of association. As in panels a and b, no significant relationship is observed for the Left and Right constructs, which have estimated low dimerization stabilities. A significant linear relationship is observed for the GAS$_{right}$ constructs. The linear relationship (solid line) is relatively similar to the original model of Díaz Vázquez (dashed line).

When the "Clash" mutants are considered (red bars in Fig. 2c), a reduction in the distribution of reconstructed fluorescence is observed across all three regions. The reduction is small but statistically significant for the Left and Right designs. A dramatic decrease is observed for the GAS$_{right}$ designs. The data support the hypothesis that a majority of the constructs associate according to their designed interface. The Clash mutants were used to validate a subset of designs that behaved consistently with their predicted interface. This dataset of validated designs included only those "wild type" sequences whose Clash mutants were in the monomeric range using a strict threshold as the criterion (below 35% of the GpA TOXGREEN signal). This selection resulted in the retention of 384 (63%) of the original designs (listed in supplemental Table S1). The distribution of reconstructed

TOXGREEN signal of the validated dataset is shown in Fig. 2d. We observed a further reduction of the overall dimerization propensity of the Left and Right designs, indicating that a substantial number of the more stable designs in these regions associated through a spurious interface. The data further confirms a low propensity for the apolar packing designs to self-associate. The distribution of reconstructed fluorescence does not change for the GAS_{right} designs, which retains the entire range of reconstructed fluorescence, including many constructs with fluorescent signals consistent with a strong dimer.

Fig. 2c also shows the reconstructed fluorescence of the sets of Large→Ala mutants (orange) designed to reduce the amount or quality of packing at the interface. Because of the narrow interface of a helix dimer, these mutations do not generally create cavities, which can be energetically expensive (46). Rather, they reduce the interface contact area by approximately 10% on average (supplemental Fig. S5b) and can produce small packing defects. In natural TM dimers, large-to-small mutations can either be well tolerated or reduce dimerization (25, 78–81). For this reason, we hypothesized that we would observe a moderate downward shift in fluorescence distribution. We found that the fluorescence distribution is statistically indistinguishable from WT for the GAS_{right} designs. In the case of the Left and Right designs, we even observed a small fluorescence increase. The data suggest that these mutations designed to reduce packing do not produce an effect measurable with TOXGREEN sort-seq. With regards to the small increase observed in the Left and Right mutants, we speculate that Ala residues may slightly favor non-specific weak association over larger amino acids.

The energetics of the GAS_{right} designs are consistent with a previous model of association

Comparison of the experimental dimerization propensities with the predicted computational energies can help obtain insights into the energetics of dimerization. Fig. 3a-c show scatter plots of the reconstructed TOXGREEN fluorescence of the designs against the energy scores of their structural models. No notable trend was found for the Left (Fig. 3a) and Right (Fig. 3b) designs. These designs populate only the bottom of the

association range of TOXGREEN signal (monomeric or weakly dimeric), therefore it is not possible to assess whether there is any correlation between their association propensity and their computational energetics. Conversely, a statistically significant correlation between association propensity and energy scores is observed for the GAS_{right} designs, a trend that is similar to our previous analysis of association of this motif (52).

To further understand the factors that are important for stabilizing the designs, we applied an empirical conversion formula that estimates the free energy of association of constructs from their TOXGREEN signal. This conversion is based on our recent study that found good agreement between the ΔG° of association of a series of poly-Leu GAS_{right} dimers measured *in vitro* in decyl-β-maltopyranoside detergent and their TOXGREEN dimerization propensities (77). Here, we found that there is a statistically significant linear correlation between the computational energy score of the GAS_{right} designs and their calculated $\Delta G°$ of association (Fig. 3f). The linear regression parameters obtained here ($y = 0.049x - 1.43$, solid line, where y is the calculated $\Delta G°$ and x is the computational score) are relatively close to the original model of Díaz Vázquez ($y = 0.076x - 1.33$, shown in dashed line), indicating that the present set of GAS_{right} designs is consistent with the previous model. As before, no correlation can be found for the weakly associating Left and Right designs (Fig. 3d and e).

Conclusion

We have presented a large-scale analysis of transmembrane helix association that addressed the question if packing of apolar residues and vdW interactions are an effective driving force for TM helix dimerization. We found that designs based on this feature are monomeric or weakly dimeric in the membrane of *E. coli*. A set of GAS_{right} dimers designed as a control with the same methodology displayed significant propensities to dimerize.

The finding that GAS_{right} dimers are overall more stable than the vdW-based Left and Right designs is not surprising since GAS_{right} and its GxxxG sequence are well-known TM helix interaction motifs that can produce strong dimers. Weak hydrogen bonds contribute significantly to the stability of GAS_{right}. However, in our previous energetic model, we estimated that the vdW component represented the largest contributor to association, by approximately a factor of two compared to hydrogen bonding (52). This suggested that even if the inter-helical hydrogen bonding was somehow turned off, vdW packing alone would be sufficient to produce rather stable GAS_{right} dimers. The present data suggest that this is not a general case.

The contributions of vdW interactions in GAS_{right} association may be over-estimated by the model. It is also possible that our high-throughput procedure – which produced thousands of designs across a range of stabilities – may not have had the level of optimization needed for satisfying packing complementarity if these are very stringent. Such stringency was suggested by Mravic and colleagues, who noted that their stable apolar pentameric designs had very strict requirements for complementarity, to the point that the relocation of a single methyl group was sufficient to completely destabilize the oligomer (48). It may be possible that a more intensive computational procedure might identify highly optimized dimers that are stabilized primarily by apolar side-chain packing.

It should also be noted that, even if apolar packing appears much weaker in driving dimerization than interactions involving weak hydrogen bonding, vdW remains a fundamental force in membrane protein folding.

First, the association of helical pairs is less costly when they are connected via short loops within the same protein chain; therefore, an interface that produces marginally stable dimers may be sufficiently strong for stabilizing the interaction between two helices in a polytopic membrane protein. Similarly, the association of single-pass membrane proteins often involves extensive interactions between extracellular and intracellular domains, in addition to the TM helices, as, for example, in receptor protein kinases (82); cooperativity with other interacting domains can reduce the energetic requirements for stabilizing TM helix pairs. Additionally, weak interaction can also be important in dynamic biological processes that involve conformational changes, such as the activation and deactivation of signaling molecules (83). Thus, the present study should be viewed as a stringent test for comparing the stability of different interaction motifs and reaffirms the critical role of vdW forces in membrane protein folding.

Methods

Structural database analysis of the geometry of ideal helical pairs

A database of non-redundant membrane proteins was obtained from OPM (53) and curated by removing homologous sequences with less than 30% sequence similarity using the program MMseq2 (54). The TM segments were extracted as annotated in OPM and then assessed for their helical nature with a program implemented using the MSL C++ libraries (64). Segments consisting of quadruplets of Cα atoms were assessed with the HELANAL algorithm (84) and were flagged as being helical if their geometry fell within the following parameters: rise per residue: 1.25-1.75Å; rotation per residue: 90-110°; and radius: 2.12-2.42Å. Helical segments composed of at least 13 consecutive helical amino acids were extracted as individual helices. Any two helices with at least 3 Cα carbons within 9 Å of distance from each other were extracted as an individual helical pair. The analysis resulted in a curated database of 835 unique parallel pairs of canonical α-helices extracted from 1541 membrane protein structures. The inter-helical parameters (schematically illustrated in Fig._1a) were then calculated as in Mueller et al. (32), namely the crossing angle θ, the inter-axial helical distance *d*, the axial rotation of each helix ω_1 and ω_2, and the position of the crossing point with respect of each helical axis, Z_1 and Z_2. The distribution of the resulting crossing angles and inter-helical distances of the helical pairs are plotted in Fig._1b.

Computational Dimer Design

Protein design was performed with a program implemented using the MSL C++ libraries on a helical pair consisting of two standard α-helices of 21 amino acids. The backbone coordinates of the pairs were selected by choosing random values for inter-helical distance *d* and a crossing angle θ within the defined regions, as for the following: "Left": *d* = 8.5 to 10 Å and θ = 20 to 40°; "Right": *d* = 7.75 to 9.5 Å and θ = -30 to -60°; GAS_{right}: *d* = 6.5

to 7.5 Å and θ = -25 to -55°. The axial rotation ω and the *Z* displacement were also assigned randomly in areas of high density. All helical pairs were placed in the middle of a virtual membrane, which was set parallel to the *XY* plane with the *Z* = 0 coordinate corresponding to the center of the bilayer. The *z* coordinate of inter-helical crossing point *P* (Fig. 1a) was placed near this center.

For right-handed designs (GAS_{right} and Right), the variable interfacial positions in the 21-amino acid helix were selected as previously as LLLxxLLxxLLxxLLxxLLLL (52), where x corresponds to a variable position and L corresponds to a non-interfacial position, which was set to Leucine. In order to identify the interfacial position of the left-handed backbones, we calculated the steric requirement as a function of the four geometric parameters *d*, θ, ω and *Z.* As shown in supplemental Fig. S1a, the analysis identified 3 distinct "stripes" of configurations compatible with knobs-into-holes packing. The first stripe ranged approximately from ω=0° and *Z*=1 Å to ω=40° and *Z*=6 Å, the second from ω=40° and *Z*=0 Å to ω=80° and *Z*=6 Å, and the third covered approximately from ω=80° and *Z*=0 Å to ω=100° and *Z*=1 Å. These three regions produced different sets of interfacial residues, corresponding to variable positions patterns LLLxxLLxxLxxLLxxL, LLLLxxLxxLLxxLxxL and LLLLxxLLxxLxxLLxxL, respectively, as illustrated in supplemental Fig. S1b.

Designs were performed using a Monte Carlo sequence optimization algorithm implemented with MSL, starting from 5000 backbones from each of the Left, Right and GAS_{right} regions. The interfacial positions were allowed to assume any of the amino acids most prevalently found in membrane protein sequences, Ala, Val, Leu, Ile, Phe, Tyr, Trp, Ser and Thr. Gly was also allowed for designs based on GAS_{right} geometry. The side chain mobility was modeled using the Energy-Based Conformer Library applied at the 95% level (85). Non-interfacial positions were restricted to Leu and sampled at the 60% level. The variable positions were initially set to a random sequence and then the sequence was optimized with Monte Carlo moves consisting of randomly changing the amino acid at one position, followed by side chain optimization performed with 3 cycles of a greedy

algorithm (86). The initial search consisted of 50 cycles from a temperature of 5000 K and cooled to 0.5 K an exponential cooling schedule. Once the end temperature is reached, another search consisting of 100 cycles at a temperature of 3649 K was used to further explore local sequence variation.

Energies were calculated using the CHARMM22 van der Waals function (66), the IMM1 membrane implicit solvation term (68), and the hydrogen bonding function of SCWRL4 (67) as implemented in MSL (64). For Cα–H hydrogen bonds the following parameters were used, as reported previously: B = 60.278, D_0 = 2.3 Å, σ_d = 1.202 Å, α_{max} = 74.0°, and β_{max} = 98.0° (32). The dimer energy was calculated as a binding energy by subtracting the energy of the monomer. During the search, the monomeric energy was approximated by a baseline monomer energy calculated from the identity of each amino acid at each position. Additionally, a sequence entropy term was added to penalize low-complexity sequences. This term is sufficiently strong to steer the overall composition of the pool of constructs to approach that of natural membrane proteins, while allowing each individual dimer to adopt the sequence most favored by the interaction energetics.

$$\Delta E_{binding} = E_{dimer(VDW,\,HBOND,\,IMM1)} - 2\,E_{monomer\,baseline} - E_{sequence\,entropy}$$

The baseline and sequence entropy terms were not used to compute the final binding energy of a design. Instead, the energy of a side chain-optimized monomer was subtracted from the energy of the dimer.

$$\Delta E_{binding} = E_{dimer(VDW,\,HBOND,\,IMM1)} - E_{monomer(VDW,\,HBOND,\,IMM1)}$$

Monomer energy approximation with baseline energies

A monomer energy approximation was developed by measuring the interactions (van der Waals, hydrogen bonding and solvation energies) for individual amino acids in a large set of monomeric helices of different amino acid compositions. The baseline energy includes a self term (interactions within the single amino acid) and pair term (interaction of the amino acid with other amino acids in the helix), which were computed for 10000

random sequences in a helical configuration and averaged. The pair terms are limited to 10 positions following an amino acid in the helix.

During design, for a given sequence, its monomeric baseline energy $E_{baseline}$ was calculated as:

$$E_{baseline} = \sum_{i=1}^{N} \left(SB_i + \sum_{j=i+1;\, j \le N}^{(i+10)} PB_{i,j} \right)$$

where *i* and *j* are the positions in a helix, *N* is the total number of amino acids in the helix, SB_i is the self baseline energy at position *i*, and $PB_{i,j}$ is the pair baseline energy between positions *i* and *j*. SB_i is dependent only on the identify of the amino acid at position *i* whereas $PB_{i,j}$ is dependent on the identities of positions *i* and *j* and their distance in the primary sequence *d=j-i*.

Supplemental Fig. S6 displays the correlation between computed monomer energy and their baseline monomer energy estimates. The table of self and pair baseline energies is provided as supplemental files Baselines.xlsx.

Sequence Entropy Term

The sequence entropy term was calculated from the probability of each amino acid in the transmembrane domains extracted from the OPM database.

$$E_{sequence\ entropy} = w \times -\log(P_{seq}) \times RT$$

where *w* is a weight factor of 10, P_{seq} is the probability of a given sequence, R is the gas constant and T is temperature defaulted to 298 K. P_{seq} was calculated as

$$P_{seq} = \prod_{AA} f_{AA}^{N_{AA}} \times \frac{n!}{\prod_{AA} N_{AA}!}$$

where N_{AA} is the number of amino acids of type AA found in the sequence and f_{AA} is the frequency of the amino acid in the OPM database, reported in supplemental Table S2.

Final structure refinement

After each design was completed, its structure was subjected to a final Monte Carlo procedure to optimize the backbone for the final sequence. The procedure consisted of 100 cycles, with an initial temperature of 1000K cooled to 0 K with an exponential schedule, where the four geometric parameters (*d*, θ, ω and *Z)* were locally varied and the side chains were optimized with 10 cycles of a greedy optimization algorithm (86). This relaxation was repeated 3 times for each protein. The final energy after this backbone relaxation step was subtracted from the energy of a side chain-optimized monomeric helix with the same sequence and used as the design computational energy score.

$$\Delta E_{binding} = E_{dimer} - 2\,E_{monomer}$$

Library cloning

DNA sequences for the designs were ordered as an oligo pool from Twist Biosciences. Individual segments were amplified using primer pairs specific to each segment by quantitative PCR (Roche KAPA SYBR Fast) based on the protocol outlined in Kosuri et al. (87) followed by traditional PCR reaction using the qPCR reaction as a template. Segments were digested using restriction enzymes NheI-HF and DpnII. The pccGFPKan (TOXGREEN) vector (75) was digested in preparation for ligation using restriction enzymes NheI-HF, BamHI-HF, and CIP. Plasmids were assembled by incubating segment DNA with vector in a 1:10 backbone:insert ratio with ElectroLigase for 2 hours at room temperature, followed by 10 minutes at 65°C to inactivate the enzyme. The inserts were designed such that the last two nucleotides were "TT" before the DpnII cut site "G" resulting in coding an in-frame Leu once ligated into the corresponding pccGFPKan restriction sites resulting in the following general insert sequence RAS**LLLxxLLxxLLxxLLxxLLLL**LILI. 5 µL of the ligation mixture were added to 50 µL of NEB 10-beta *E. coli* cells and cloned through electroporation (BIO-RAD MicroPulser Electroporator), grown for 1 hour in 950 µL NEB Stable Outgrowth Medium media in 37 °C shaker, and diluted 1 mL in 4 mL LB media containing

100 µg/mL ampicillin. 50 µL of a 1:10 and 1:100 dilution were plated on LB plates containing 100 µg/mL ampicillin. Plates were grown for 14-16 hours at 37°C and colony forming units (CFUs) were calculated for each segment (number of Colonies × dilution/number of sequences in the segment). A CFU > 5 was accepted. Overnight cultures corresponding to plates were spun down and the plasmid DNA was extracted (Qiagen). The plasmid library was transformed through electroporation into 50 µL malE deficient *E. coli* strain MM39, grown for 1 hour in 950 µL SOC media in a 37°C shaker, and 100 µL of 1:10 dilution plated on LB plates containing 100 µg/mL ampicillin with the rest being grown overnight in 3 mL LB media containing 100 µg/mL ampicillin. CFUs were again counted and any plates with values > 5 were accepted. Overnight cultures for each segment were stored in 25% glycerol stocks. 50 µL samples and controls from glycerol stocks were grown in 3 mL LB media containing 100 µg/mL ampicillin for 2-4 hours to reach an optical density of approximately 0.1 OD_{600} and the corresponding number of sequences per sample were used to calculate how much to add to a full library glycerol stock. All enzymes acquired from NEB.

TOXGREEN sort-seq

TOXGREEN sort-seq was performed according to Anderson et al. (74). 50 µL of library glycerol stock were grown in 3 mL LB containing 100 µg/mL ampicillin overnight for 14-16 hours in a 37 °C shaker. FACS was performed using a Sony MA900 cell sorter. The FSC threshold was set at 0.05%. The sensor gains for FL1 (GFP) and backscatter (BSC) were set at 90% and 30%, respectively. The first gate was drawn to select for cells by eliminating cell debris on a side scatter area (SSC-A) vs. forward scatter area (FSC-A) plot. The second gate was drawn to eliminate doublet cells on a forward scatter height (FSC-H) vs. FSC-A plot. 10 µL of sample culture were diluted in 50 mL PBS buffer and 10,000 events were recorded per second in the Fluorescence-Activated Cell Sorter. Control cells containing the GpA, G83I, and No-TM controls were flowed to calibrate the instrument and determine proper gating to remove dead cells. Individual sample libraries were first flowed through the

instrument and the fluorescence profile was subdivided into 4 bins. 100,000 events were sorted into bins 1-3, and 50,000 events were sorted into bin 4. Each library sample was sorted in triplicate from biological replicates of overnights into 2 mL LB broth containing 100 µg/mL ampicillin. Sorted populations were grown for 1-5 hours to reach an optical density of approximately 0.3 OD_{600} and plasmid DNA was extracted. Samples were prepared for NGS using PCR: each sorted population was amplified with unique stem primers, allowing us to distinguish between populations for each NGS run. These amplified samples were sent to the DNA Sequencing Center at the University of Wisconsin-Madison Biotech Center for library preparation and next-generation sequencing for 500 reads for each sequence in the library. Library Preparation Services: Index PCR (TruSeq). Sequencing Services: Illumina (NovaSeq) Sequencing [2x150 Shared (10M read increments)].

The GFP fluorescence of each individual construct in the library was reconstructed as in Anderson et al. (74) using a weighted average approach (88). This method normalizes the NGS reads per construct per bin with the fraction of the population found in that bin. To ensure NGS data quality, a cutoff of 1 incorrect base pair per sequence was implemented based on the Phred score (Q) of each base, measuring the probability that the base is incorrect where $Q = -10 \log_{10}P$ (89, 90). Further, a minimum cutoff of 10 copies of a particular sequence had to be present in the NGS for it to be retained. The normalized fractional contribution a_{ij} for each construct i in each bin j was calculated as

$$a_{ij} = \frac{\frac{f_j \cdot c_{ij}}{\sum_i c_{ij}}}{\sum_j \frac{f_j \cdot c_{ij}}{\sum_i c_{ij}}}$$

where c_{ij} is the NGS sequencing count for construct i in bin j, and f_j is the fraction of the sorted sample population in bin j. The reconstructed fluorescence (RFU_i) for each individual protein sequence, expressed as

Relative Fluorescence Units was then derived by summing the product of the normalized fractional contribution a_{ij} and the median fluorescence m_j over each sorted bin *j*.

$$RFU_i = \sum_j a_{ij} \cdot m_j$$

Normalized TOXGREEN conversion

TOXGREEN fluorescence is typically converted to a relative signal by normalizing its fluorescence to the fluorescence of the strong glycophorin A (GpA) dimer. To reduce the dependency of the normalization from the variability associated with an individual construct, we created a calibration set of approximately 10 constructs to be used for this normalization. Standard TOXGREEN was performed on a set of individual constructs from each design region, including GpA and its monomeric variant G83I, used as the negative control. The fluorescence values, the normalized (% GpA) values and the reconstructed fluorescence values of these constructs are reported in supplemental Table S3. As illustrated in Fig. S7, linear regression was used to determine the relationship between the individual TOXGREEN signal of each construct, expressed as % GpA, and their reconstructed fluorescence. The resulting linear regression equation was then applied to transform the reconstructed fluorescence values to % GpA normalized values.

Conversion of TOXGREEN signal to an estimated free energy of association

TOXGREEN data was converted to an estimated ΔG of association in detergent using a previously found relationship (77). First, a conversion factor was applied to account for the difference between TOXGREEN and TOXCAT signals.

$$TOXCAT = \frac{TOXGREEN - 3.8}{1.1}$$

Then a theoretical fraction dimer *FD* was calculated according to the following relationship (77):

$$FD = \frac{TOXCAT - 24.9}{332.2}$$

The fraction dimer was converted to a dissociation constant K_D, assuming a concentration of protein χ_T of 1.8×10^{-4} protein molecules per detergent molecule.

$$K_D = 2\chi_T \frac{(FD-1)^2}{FD}$$

Finally, an estimated free energy of association in detergent ΔG was calculated from the dissociation constant, where R is the gas constant and T is 4 °C.

$$\Delta G = RTln(K_D)$$

Immunoblotting

Protein expression was confirmed for an assortment of designs and mutants using immunoblotting (supplemental Fig. S2). Cell lysates were normalized by ThermoScientific Pierce Protein BCA Assay Kit and 10uL loaded onto NuPage 4-12% bis-tris SDS-PAGE gels (ThermoFisher) and then transferred to PVDF membranes (VWR) for 1 hour at 100 millivolts. Blots were blocked using 5% milk (Nestle Instant Nonfat Dry Milk) in TBS-Tween buffer (50 mM Tris, 150 mM NaCl, 0.05% Tween 20) for 2 hours at 4 °C and incubated overnight with peroxidase-conjugated monoclonal anti-maltose binding protein antibodies (Sigma-Aldrich). Blots were developed with Pierce ECL Western Blotting Substrate Kit; 1 mL of ECL solution was added to the blot and incubated for 90 seconds. Chemiluminescence was measured using an ImageQuant LAS 4000 (GE Healthsciences).

MalE Complementation Assay

To confirm proper membrane insertion and orientation of the TOXGREEN constructs, the library was tested with a MalE Complementation Assay performed in M9 media with maltose as the sole carbon source. The MM39

E. coli strain used in the TOXGREEN assay is Maltose Binding Protein (MBP)-deficient. TOXGREEN constructs can support growth in M9 maltose minimal media only if the chimeric protein is expressed and inserted into the inner membrane with the C-terminal MBP moiety correctly oriented to the periplasmic side (76).

The library version of the MalE Complementation Assay in liquid media was performed as described previously (74). Briefly, overnight cultures of each library were used to inoculate 500 mL M9 media containing 0.8% maltose (w/v) with a dilution to a final cell concentration of 0.00125 OD_{600}. Samples were taken at time points 30 and 36 hours in triplicate of growth at 37°C. DNA was extracted and submitted for NGS analysis. Samples from the overnight LB cultures used for initial inoculation were also analyzed by NGS for comparison with non-selective conditions. All sequences that had a fractional NGS count change worse than a set of known maltose-negative (non-inserting) controls (constructs "1E01", "1E11", "2E11", "2F12", "2H11", "1F02", "2H07") (74) were excluded from our analysis.

Programs and data availability

Programs used to determine the preferred interhelical geometries from the structural database and design the dimers are available at the GitHub repository https://github.com/TmdimerDesign/vdwDesign/, also available at https://doi.org/10.5281/zenodo.16876329. TOXGREEN plasmids for control constructs are available in the AddGene repository (https://www.addgene.org). These include the original TOXGREEN plasmid (pccGFPKAN: AddGene ID #73649), the TOXGREEN positive (GpA: #73651) and negative controls (G83I: #73650) (75) and an extended set of maltose-negative and maltose-positive control constructs (1E01: #239013; 1E11: #239083; 2E11: #239084; 2F12: #239085; 2H11: #239086; 1G01: #239087; 2H01 #239760; 2H07: #239088) (74). The PDB files of the designed dimers are available for download from the Dryad repository at https://doi.org/10.5061/dryad.5dv41nsjg. The repository also include raw sort-seq data used for the calculation

of TOXGREEN sort-seq values and for the MalE Complementation Assay. Scripts for the analysis of TOXGREEN sort-seq data are available at https://doi.org/10.5061/dryad.qnk98sfv8 (74).

Author Contributions

GJL: Conception and Design, Data Acquisition, Analysis, and Interpretation, Wrote Manuscript.

AS: Conception and Design, Data Analysis, and Interpretation, Wrote Manuscript.

Acknowledgments

This work was supported by National Institutes of Health grant R35-GM130339 to A.S. and National Science Foundation grant CHE-1710182 to A.S.. G.J.L. acknowledges the support of National Institutes of Health training grant T32GM008505 to the Chemistry-Biology Interface Training Program (CBI).

References

1. Engelman, D.M., Y. Chen, C.-N. Chin, A.R. Curran, A.M. Dixon, A.D. Dupuy, A.S. Lee, U. Lehnert, E.E. Matthews, Y.K. Reshetnyak, A. Senes, and J.-L. Popot. 2003. Membrane protein folding: beyond the two stage model. *FEBS Lett*. 555:122–125.
2. Popot, J.L., and D.M. Engelman. 1990. Membrane protein folding and oligomerization: the two-stage model. *Biochemistry*. 29:4031–7.
3. Marinko, J.T., H. Huang, W.D. Penn, J.A. Capra, J.P. Schlebach, and C.R. Sanders. 2019. Folding and Misfolding of Human Membrane Proteins in Health and Disease: From Single Molecules to Cellular Proteostasis. *Chem Rev*. 119:5537–5606.
4. Monné, M., I. Nilsson, A. Elofsson, and G. von Heijne. 1999. Turns in transmembrane helices: determination of the minimal length of a "helical hairpin" and derivation of a fine-grained turn propensity scale. *J Mol Biol*. 293:807–814.
5. Pitonzo, D., and W.R. Skach. 2006. Molecular mechanisms of aquaporin biogenesis by the endoplasmic reticulum Sec61 translocon. *Biochimica et Biophysica Acta (BBA) - Biomembranes*. 1758:976–988.
6. Sadlish, H., and W.R. Skach. 2004. Biogenesis of CFTR and other Polytopic Membrane Proteins: New Rolesfor the Ribosome-Translocon Complex. *J Membrane Biol*. 202:115–126.
7. Sato, T., T. Kawakami, K. Akaji, H. Konishi, K. Mochizuki, T. Fujiwara, H. Akutsu, and S. Aimoto. 2002. Synthesis of a membrane protein with two transmembrane regions. *Journal of Peptide Science*. 8:172–180.
8. Vitrac, H., M. Bogdanov, and W. Dowhan. 2013. In vitro reconstitution of lipid-dependent dual topology and postassembly topological switching of a membrane protein. *Proc Natl Acad Sci U S A*. 110:9338–9343.
9. Vitrac, H., D.M. MacLean, V. Jayaraman, M. Bogdanov, and W. Dowhan. 2015. Dynamic membrane protein topological switching upon changes in phospholipid environment. *Proc Natl Acad Sci U S A*. 112:13874–13879.
10. Woodall, N.B., S. Hadley, Y. Yin, and J.U. Bowie. 2017. Complete topology inversion can be part of normal membrane protein biogenesis. *Protein Sci*. 26:824–833.
11. De Marothy, M.T., and A. Elofsson. 2015. Marginally hydrophobic transmembrane α-helices shaping membrane protein folding. *Protein Sci*. 24:1057–1074.
12. Cymer, F., G. von Heijne, and S.H. White. 2015. Mechanisms of Integral Membrane Protein Insertion and Folding. *Journal of Molecular Biology*. 427:999–1022.
13. Hessa, T., H. Kim, K. Bihlmaier, C. Lundin, J. Boekel, H. Andersson, I. Nilsson, S.H. White, and G. von Heijne. 2005. Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature*. 433:377–381.
14. Hessa, T., N.M. Meindl-Beinker, A. Bernsel, H. Kim, Y. Sato, M. Lerch-Bader, I. Nilsson, S.H. White, and G. von Heijne. 2007. Molecular code for transmembrane-helix recognition by the Sec61 translocon. *Nature*. 450:1026–1030.
15. Marx, D.C., and K.G. Fleming. 2021. Local Bilayer Hydrophobicity Modulates Membrane Protein Stability. *J. Am. Chem. Soc.* 143:764–772.
16. Moon, C.P., and K.G. Fleming. 2011. Side-chain hydrophobicity scale derived from transmembrane protein folding into lipid bilayers. *Proceedings of the National Academy of Sciences*. 108:10174–10177.
17. Tanford, C. 1980. The hydrophobic effect: formation of micelles and biological membranes 2d ed. J. Wiley.
18. Yang, A.-S., K.A. Sharp, and B. Honig. 1992. Analysis of the heat capacity dependence of protein folding. *Journal of Molecular Biology*. 227:889–900.
19. Senes, A. 2011. Computational design of membrane proteins. *Curr Opin Struct Biol*. 21:460–6.
20. Senes, A., D.E. Engel, and W.F. DeGrado. 2004. Folding of helical membrane proteins: the role of polar, GxxxG-like and proline motifs. *Curr. Opin. Struct. Biol*. 14:465–479.
21. Choma, C., H. Gratkowski, J.D. Lear, and W.F. DeGrado. 2000. Asparagine-mediated self-association of a model transmembrane helix. *Nat Struct Biol*. 7:161–6.
22. Gratkowski, H., J.D. Lear, and W.F. DeGrado. 2001. Polar side chains drive the association of model transmembrane peptides. *Proc. Natl. Acad. Sci. U.S.A.* 98:880–885.
23. Zhou, F.X., H.J. Merianos, A.T. Brunger, and D.M. Engelman. 2001. Polar residues drive association of polyleucine transmembrane helices. *Proc Natl Acad Sci U S A*. 98:2250–5.

24. Doura, A.K., F.J. Kobus, L. Dubrovsky, E. Hibbard, and K.G. Fleming. 2004. Sequence context modulates the stability of a GxxxG-mediated transmembrane helix-helix dimer. *J. Mol. Biol.* 341:991–998.
25. Doura, A.K., and K.G. Fleming. 2004. Complex Interactions at the Helix–Helix Interface Stabilize the Glycophorin A Transmembrane Dimer. *Journal of Molecular Biology*. 343:1487–1497.
26. Duong, M.T., T.M. Jaszewski, K.G. Fleming, and K.R. MacKenzie. 2007. Changes in apparent free energy of helix-helix dimerization in a biological membrane due to point mutations. *J. Mol. Biol*. 371:422–434.
27. Fleming, K.G., and D.M. Engelman. 2001. Specificity in transmembrane helix–helix interactions can define a hierarchy of stability for sequence variants. *Proceedings of the National Academy of Sciences*. 98:14340–14344.
28. Hong, H., T.M. Blois, Z. Cao, and J.U. Bowie. 2010. Method to measure strong protein–protein interactions in lipid bilayers using a steric trap. *Proc Natl Acad Sci U S A*. 107:19802–19807.
29. Joh, N.H., A. Min, S. Faham, J.P. Whitelegge, D. Yang, V.L. Woods, and J.U. Bowie. 2008. Modest stabilization by most hydrogen-bonded side-chain interactions in membrane proteins. *Nature*. 453:1266–1270.
30. Stanley, A.M., and K.G. Fleming. 2007. The Role of a Hydrogen Bonding Network in the Transmembrane β-Barrel OMPLA. *Journal of Molecular Biology*. 370:912–924.
31. Arbely, E., and I.T. Arkin. 2004. Experimental measurement of the strength of a C alpha-H...O bond in a lipid bilayer. *J. Am. Chem. Soc*. 126:5362–5363.
32. Mueller, B.K., S. Subramaniam, and A. Senes. 2014. A frequent, GxxxG-mediated, transmembrane association motif is optimized for the formation of interhelical Cα-H hydrogen bonds. *Proc Natl Acad Sci U S A*. 111:E888-95.
33. Senes, A., I. Ubarretxena-Belandia, and D.M. Engelman. 2001. The Cα—H···O hydrogen bond: A determinant of stability and specificity in transmembrane helix interactions. *Proceedings of the National Academy of Sciences*. 98:9056–9061.
34. Sandoval-Lira, J., L. Fuentes, L. Quintero, H. Höpfl, J.M. Hernández-Pérez, J.L. Terán, and F. Sartillo-Piscil. 2015. The Stabilizing Role of the Intramolecular C–H···O Hydrogen Bond in Cyclic Amides Derived From α-Methylbenzylamine. *J. Org. Chem.* 80:4481–4490.
35. Scheiner, S. 2005. Chapter 29 - The CH···O hydrogen bond: A historical account. In: Dykstra CE, G Frenking, KS Kim, GE Scuseria, editors. Theory and Applications of Computational Chemistry. Amsterdam: Elsevier. pp. 831–857.
36. Scheiner, S., T. Kar, and Y. Gu. 2001. Strength of the CαH··O Hydrogen Bond of Amino Acid Residues *. *Journal of Biological Chemistry*. 276:9832–9837.
37. Vargas, R., J. Garza, D.A. Dixon, and B.P. Hay. 2000. How Strong Is the Cα−H···OC Hydrogen Bond? *J. Am. Chem. Soc.* 122:4750–4755.
38. Bower, M.J., F.E. Cohen, and R.L. Dunbrack. 1997. Prediction of protein side-chain rotamers from a backbone-dependent rotamer library: a new homology modeling tool1. *Journal of Molecular Biology*. 267:1268–1282.
39. Chamberlain, A.K., Y. Lee, S. Kim, and J.U. Bowie. 2004. Snorkeling preferences foster an amino acid composition bias in transmembrane helices. *J. Mol. Biol.* 339:471–479.
40. Eilers, M., S.C. Shekar, T. Shieh, S.O. Smith, and P.J. Fleming. 2000. Internal packing of helical membrane proteins. *Proc. Natl. Acad. Sci. U.S.A*. 97:5796–5801.
41. Oberai, A., N.H. Joh, F.K. Pettit, and J.U. Bowie. 2009. Structural imperatives impose diverse evolutionary constraints on helical membrane proteins. *Proc Natl Acad Sci U S A*. 106:17747–17750.
42. Baker, R.P., and S. Urban. 2012. Architectural and thermodynamic principles underlying intramembrane protease function. *Nat Chem Biol*. 8:759–768.
43. Guo, R., K. Gaffney, Z. Yang, M. Kim, S. Sungsuwan, X. Huang, W.L. Hubbell, and H. Hong. 2016. Steric trapping reveals a cooperativity network in the intramembrane protease GlpG. *Nat Chem Biol*. 12:353–360.
44. Lemmon, M.A., J.M. Flanagan, J.F. Hunt, B.D. Adair, B.J. Bormann, C.E. Dempsey, and D.M. Engelman. 1992. Glycophorin A dimerization is driven by specific interactions between transmembrane alpha-helices. *J Biol Chem*. 267:7683–9.
45. Mersch, K., T.N. Ozturk, K. Park, H.-H. Lim, and J.L. Robertson. 2021. Altering CLC stoichiometry by reducing non-polar side-chains at the dimerization interface. *J Mol Biol*. 433:166886.
46. Joh, N.H., A. Oberai, D. Yang, J.P. Whitelegge, and J.U. Bowie. 2009. Similar energetic contributions of packing in the core of membrane and water-soluble proteins. *J Am Chem Soc*. 131:10846–7.
47. Ptacek, J.B., A.P.B. Edwards, L.L. Freeman-Cook, and D. DiMaio. 2007. Packing contacts can mediate highly specific interactions between artificial transmembrane proteins and the PDGFbeta receptor. *Proc Natl Acad Sci U S A*. 104:11945–11950.

48. Mravic, M., J.L. Thomaston, M. Tucker, P.E. Solomon, L. Liu, and W.F. DeGrado. 2019. Packing of apolar side chains enables accurate design of highly stable membrane proteins. *Science*. 363:1418–1423.
49. Bugge, K., K. Lindorff-Larsen, and B.B. Kragelund. 2016. Understanding single-pass transmembrane receptor signaling from a structural viewpoint-what are we missing? *FEBS J*. 283:4424–4451.
50. Endres, N.F., R. Das, A.W. Smith, A. Arkhipov, E. Kovacs, Y. Huang, J.G. Pelton, Y. Shan, D.E. Shaw, D.E. Wemmer, J.T. Groves, and J. Kuriyan. 2013. Conformational coupling across the plasma membrane in activation of the EGF receptor. *Cell*. 152:543–556.
51. Hubbard, S.R., and W.T. Miller. 2007. Receptor tyrosine kinases: mechanisms of activation and signaling. *Curr Opin Cell Biol*. 19:117–123.
52. Anderson, S.M., B.K. Mueller, E.J. Lange, and A. Senes. 2017. Combination of Cα-H Hydrogen Bonds and van der Waals Packing Modulates the Stability of GxxxG-Mediated Dimers in Membranes. *J Am Chem Soc*. 139:15774–15783.
53. Lomize, M.A., A.L. Lomize, I.D. Pogozheva, and H.I. Mosberg. 2006. OPM: orientations of proteins in membranes database. *Bioinformatics*. 22:623–625.
54. Steinegger, M., and J. Söding. 2017. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nature biotechnology*. 35:1026–1028.
55. Walters, R.F.S., and W.F. DeGrado. 2006. Helix-packing motifs in membrane proteins. *Proceedings of the National Academy of Sciences*. 103:13658–13663.
56. Russ, W.P., and D.M. Engelman. 2000. The GxxxG motif: a framework for transmembrane helix-helix association. *J Mol Biol*. 296:911–9.
57. Ash, W.L., T. Stockner, J.L. MacCallum, and D.P. Tieleman. 2004. Computer modeling of polyleucine-based coiled coil dimers in a realistic membrane environment: insight into helix-helix interactions in membrane proteins. *Biochemistry*. 43:9050–60.
58. Bornberg-Bauer, E., E. Rivals, and M. Vingron. 1998. Computational approaches to identify leucine zippers. *Nucleic Acids Research*. 26:2740–2746.
59. Walshaw, J., and D.N. Woolfson. 2003. Extended knobs-into-holes packing in classical and complex coiled-coil assemblies. *Journal of Structural Biology*. 144:349–361.
60. Liu, Y., D.M. Engelman, and M. Gerstein. 2002. Genomic analysis of membrane protein families: abundance and conserved motifs. *Genome Biol*. 3:research0054.
61. Liang, J., L. Adamian, and R. Jackups. 2005. The membrane–water interface region of membrane proteins: structural bias and the anti-snorkeling effect. *Trends in Biochemical Sciences*. 30:355–357.
62. Huang, B., Y. Xu, X. Hu, Y. Liu, S. Liao, J. Zhang, C. Huang, J. Hong, Q. Chen, and H. Liu. 2022. A backbone-centred energy function of neural networks for protein design. *Nature*. 602:523–528.
63. Kuhlman, B., G. Dantas, G.C. Ireton, G. Varani, B.L. Stoddard, and D. Baker. 2003. Design of a Novel Globular Protein Fold with Atomic-Level Accuracy. *Science*. 302:1364–1368.
64. Kulp, D.W., S. Subramaniam, J.E. Donald, B.T. Hannigan, B.K. Mueller, G. Grigoryan, and A. Senes. 2012. Structural informatics, modeling, and design with an open-source Molecular Software Library (MSL). *J Comput Chem*. 33:1645–61.
65. Nash, A., R. Notman, and A.M. Dixon. 2015. De novo design of transmembrane helix-helix interactions and measurement of stability in a biological membrane. *Biochim Biophys Acta*. 1848:1248–57.
66. MacKerell, A.D., D. Bashford, M. Bellott, R.L. Dunbrack, J.D. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F.T. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen, B. Prodhom, W.E. Reiher, B. Roux, M. Schlenkrich, J.C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, and M. Karplus. 1998. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B*. 102:3586–616.
67. Krivov, G.G., M.V. Shapovalov, and R.L. Dunbrack. 2009. Improved prediction of protein side-chain conformations with SCWRL4. *Proteins*. 77:778–95.
68. Lazaridis, T. 2003. Effective energy function for proteins in lipid membranes. *Proteins*. 52:176–92.
69. Khadria, A.S., B.K. Mueller, J.A. Stefely, C.H. Tan, D.J. Pagliarini, and A. Senes. 2014. A Gly-Zipper Motif Mediates Homodimerization of the Transmembrane Domain of the Mitochondrial Kinase ADCK3. *J. Am. Chem. Soc.* 136:14068–14077.
70. LaPointe, L.M., K.C. Taylor, S. Subramaniam, A. Khadria, I. Rayment, and A. Senes. 2013. Structural Organization of FtsB, a Transmembrane Protein of the Bacterial Divisome. *Biochemistry*. 52:2574–2585.

71. Lawrie, C.M., E.S. Sulistijo, and K.R. MacKenzie. 2010. Intermonomer Hydrogen Bonds Enhance GxxxG-Driven Dimerization of the BNIP3 Transmembrane Domain: Roles for Sequence Context in Helix–Helix Association in Membranes. *Journal of Molecular Biology*. 396:924–936.
72. Wei, P., X. Liu, M.-H. Hu, L.-M. Zuo, M. Kai, R. Wang, and S.-Z. Luo. 2011. The dimerization interface of the glycoprotein Ibβ transmembrane domain corresponds to polar residues within a leucine zipper motif. *Protein Science*. 20:1814–1823.
73. Wei, P., B.-K. Zheng, P.-R. Guo, T. Kawakami, and S.-Z. Luo. 2013. The Association of Polar Residues in the DAP12 homodimer: TOXCAT and Molecular Dynamics Simulation Studies. *Biophysical Journal*. 104:1435–1444.
74. Anderson, S.M., J. Choi, E.M. Cushman, M. Leander, S. Raman, and A. Senes. 2025. High-throughput discovery of transmembrane helix dimers from human single-pass membrane proteins with TOXGREEN sort-seq. *bioRxiv*. 2025.04.22.650048.
75. Armstrong, C.R., and A. Senes. 2016. Screening for transmembrane association in divisome proteins using TOXGREEN, a high-throughput variant of the TOXCAT assay. *Biochim Biophys Acta*. 1858:2573–2583.
76. Russ, W.P., and D.M. Engelman. 1999. TOXCAT: a measure of transmembrane helix association in a biological membrane. *Proc Natl Acad Sci U S A*. 96:863–8.
77. Díaz Vázquez, G., Q. Cui, and A. Senes. 2023. Thermodynamic analysis of the GASright transmembrane motif supports energetic model of dimerization. *Biophysical Journal*. 122:143–155.
78. Fleming, K.G., and D.M. Engelman. 2001. Computation and mutagenesis suggest a right-handed structure for the synaptobrevin transmembrane dimer. *Proteins: Structure, Function, and Bioinformatics*. 45:313–317.
79. Howitt, S.M., A.J.W. Rodgers, P.D. Jeffrey, and G.B. Cox. 1996. A Mutation in Which Alanine 128 Is Replaced by Aspartic Acid Abolishes Dimerization of the b-Subunit of the F0F1-ATPase from Escherichia coli(∗). *Journal of Biological Chemistry*. 271:7038–7042.
80. Metcalf, D.G., P.B. Law, and W.F. DeGrado. 2007. Mutagenesis data in the automated prediction of transmembrane helix dimers. *Proteins: Structure, Function, and Bioinformatics*. 67:375–384.
81. Mingarro, I., P. Whitley, G.V. Heijne, and M.A. Lemmon. 1996. Ala-insertion scanning mutagenesis of the glycophorin a transmembrane helix: A rapid way to map helix-helix interactions in integral membrane proteins. *Protein Science*. 5:1339–1341.
82. Bocharov, E.V., G.V. Sharonov, O.V. Bocharova, and K.V. Pavlov. 2017. Conformational transitions and interactions underlying the function of membrane embedded receptor protein kinases. *Biochim Biophys Acta Biomembr*. 1859:1417–1429.
83. Matthews, E.E., M. Zoonens, and D.M. Engelman. 2006. Dynamic helix interactions in transmembrane signaling. *Cell*. 127:447–450.
84. Bansal, M., S. Kumar, and R. Velavan. 2000. HELANAL: a program to characterize helix geometry in proteins. *J Biomol Struct Dyn*. 17:811–819.
85. Subramaniam, S., and A. Senes. 2012. An energy-based conformer library for side chain optimization: improved prediction and adjustable sampling. *Proteins*. 80:2218–34.
86. Xiang, Z., P.J. Steinbach, M.P. Jacobson, R.A. Friesner, and B. Honig. 2007. Prediction of side-chain conformations on protein surfaces. *Proteins*. 66:814–823.
87. Kosuri, S., N. Eroshenko, E.M. Leproust, M. Super, J. Way, J.B. Li, and G.M. Church. 2010. Scalable gene synthesis by selective amplification of DNA pools from high-fidelity microchips. *Nat Biotechnol*. 28:1295–1299.
88. Kosuri, S., D.B. Goodman, G. Cambray, V.K. Mutalik, Y. Gao, A.P. Arkin, D. Endy, and G.M. Church. 2013. Composability of regulatory sequences controlling transcription and translation in Escherichia coli. *Proc Natl Acad Sci U S A*. 110:14024–14029.
89. Ewing, B., L. Hillier, M.C. Wendl, and P. Green. 1998. Base-Calling of Automated Sequencer Traces UsingPhred. I. Accuracy Assessment. *Genome Res.* 8:175–185.
90. Ewing, B., and P. Green. 1998. Base-Calling of Automated Sequencer Traces Using Phred. II. Error Probabilities. *Genome Res.* 8:186–194.

SUPPLEMENTARY INFORMATION

Packing of apolar amino acids is not a strong stabilizing force in transmembrane helix dimerization.

Gilbert J. Loiseau[1,2] and Alessandro Senes[1,@]

[1]Department of Biochemistry, University of Wisconsin-Madison, Madison, WI 53706

[2]Integrated Program in Biochemistry, University of Wisconsin-Madison, Madison, WI 53706

@Corresponding author: senes@wisc.edu

ORCID:

GJL: 0009-0003-8186-5275

AS: 0000-0002-3807-2275

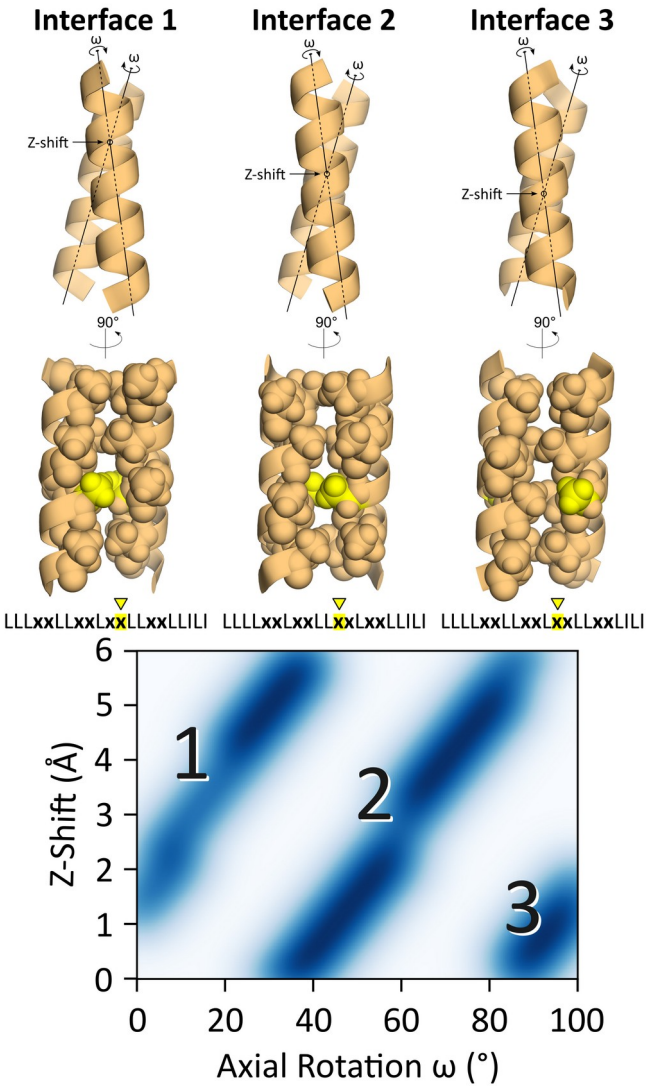
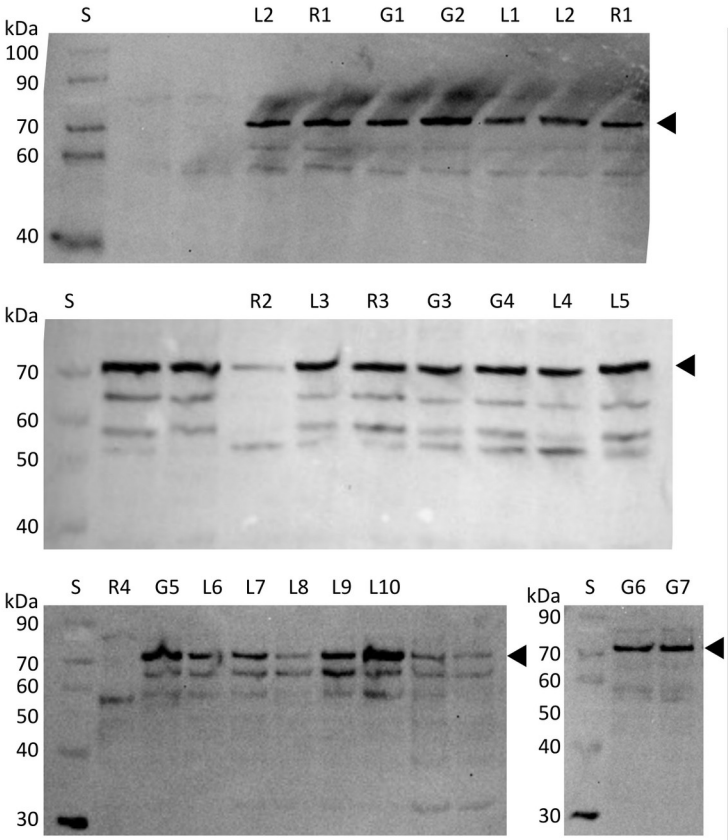


Fig. S1. Selection of the interfacial position for the "Left" designs. The interaction interface in the 21-amino acid TOXGREEN construct for the "Left" designs depends on the ω (axial rotation) and *Z*-shift (the position of the crossing point relative to the helical axis. These two coordinates generate three different regions of know-into-hole packing, corresponding to three different arrangements of the variable (**x**) interfacial position during the design process.



Western ID	Sequence	Design ID
G1	LLL**AV**LL**LT**LL**GG**LL**FA**LILI	
G2	LLL**YV**LL**GA**LL**GI**LL**TL**LILI	G_016
L1	LLLLL**FL**L**AT**LL**IL**L**SV**LILI	L_021
L2	LLL**FS**LL**LL**L**VA**LL**LT**LLILI	
R1	LLL**YI**LL**TA**LL**VA**LL**SL**LILI	R_051
R2	LLL**FY**LL**VA**LL**TA**LL**SL**LILI	R_140
L3	LLLL**AV**LL**FL**L**AL**L**TS**LLILI	L_031
R3	LLL**TV**LL**LA**LL**FA**LL**SI**LILI	R_090
G3	LLL**TA**LL**LA**LL**FG**LL**FS**LILI	G_060
G4	LLL**TA**LL**IG**LL**FG**LL**VL**LILI	
L4	LLLL**YL**LL**AV**L**TA**LL**FS**LILI	L_022
L5	LLLL**LA**LL**LI**L**FV**LL**ST**LILI	
R4	LLL**TS**LL**LA**LL**FV**LL**IL**LILI	
G5	LLL**VA**LL**AL**LL**GT**LL**SF**LILI	G_052
L6	LLLL**FA**LL**LA**LL**YV**L**TS**LLILI	
L7	LLLL**AI**LL**AV**L**FT**LL**SL**LILI	
L8	LLL**YA**LL**FV**L**TA**LL**LS**LLILI	L_006
L9	LLLL**AV**LL**LF**LL**AL**L**TS**LLILI	L_031
L10	LLL**FA**LL**LV**L**IT**LL**YS**LLILI	
G6	LLL**TV**LL**AL**LL**GF**LL**GS**LILI	G_022
G7	LLL**VA**LL**AL**LL**GT**LL**SF**LILI	G_052

Fig. S2. Protein expression of designs. Protein expression was confirmed by randomly sampling a number of constructs by immunoblotting with anti-maltose binding protein antibodies. L: "Left" designs. R: "Right" designs. G: GAS_{right} designs, S: molecular standards. The triangle indicates the main band of interest. Although some variability is observed (most notably R2, R4 and L10), the expression level is overall consistent across the constructs.

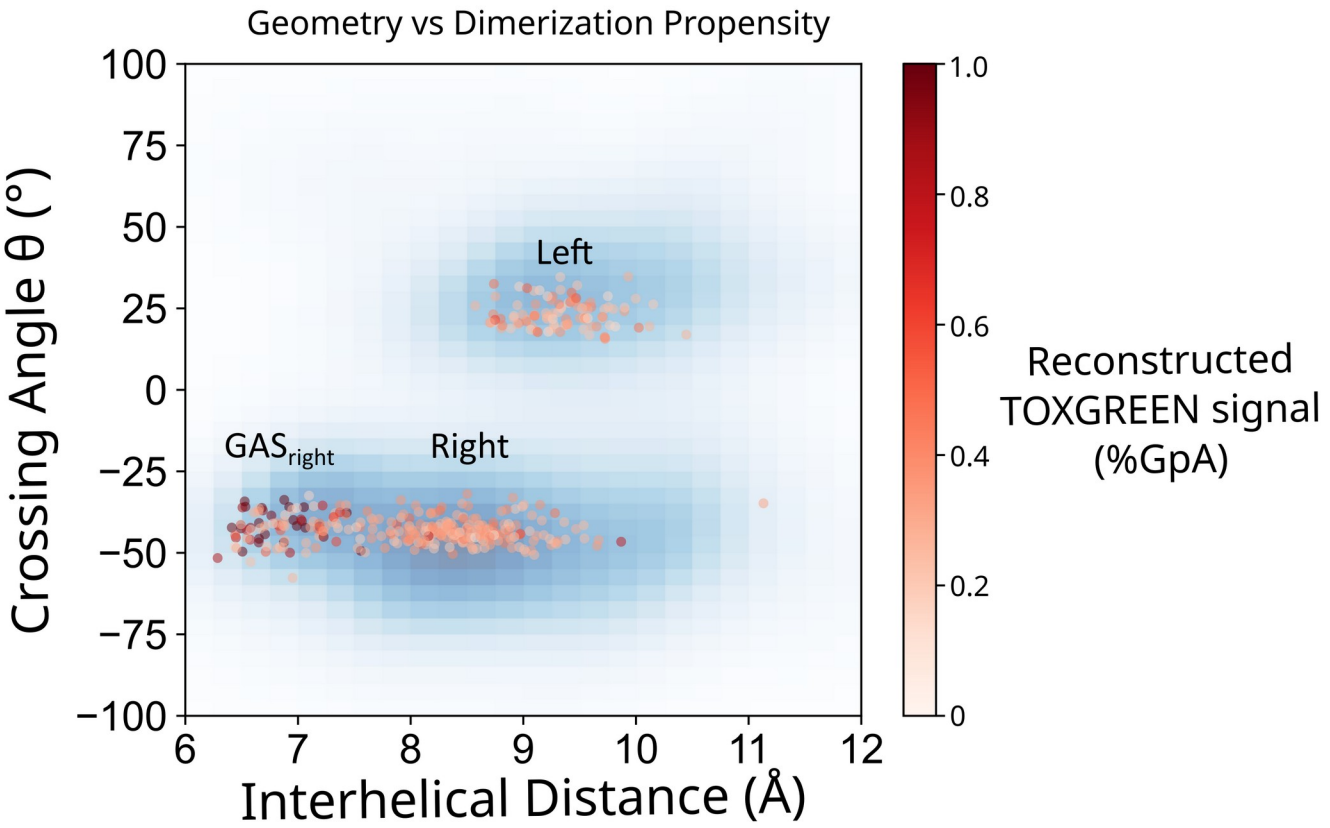


Fig. S3. Mapping of the final set of 613 as a function of crossing angle θ and interhelical distance *d*. The color coding shows the reconstructed TOXGREEN dimerization propensity of each design.

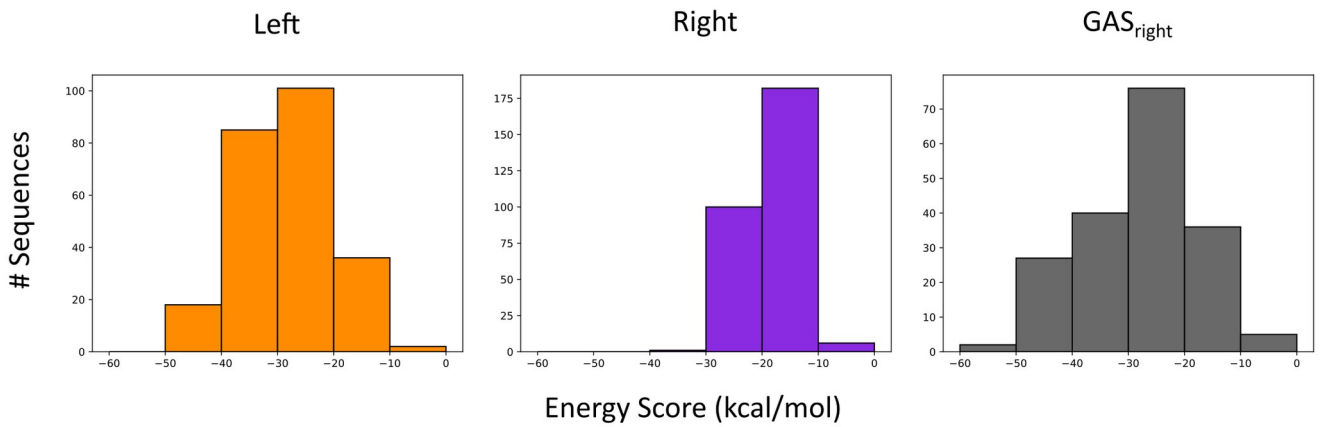


Fig. S4. Distribution of the computational energy score of the final set of 613 designs in the three design regions.

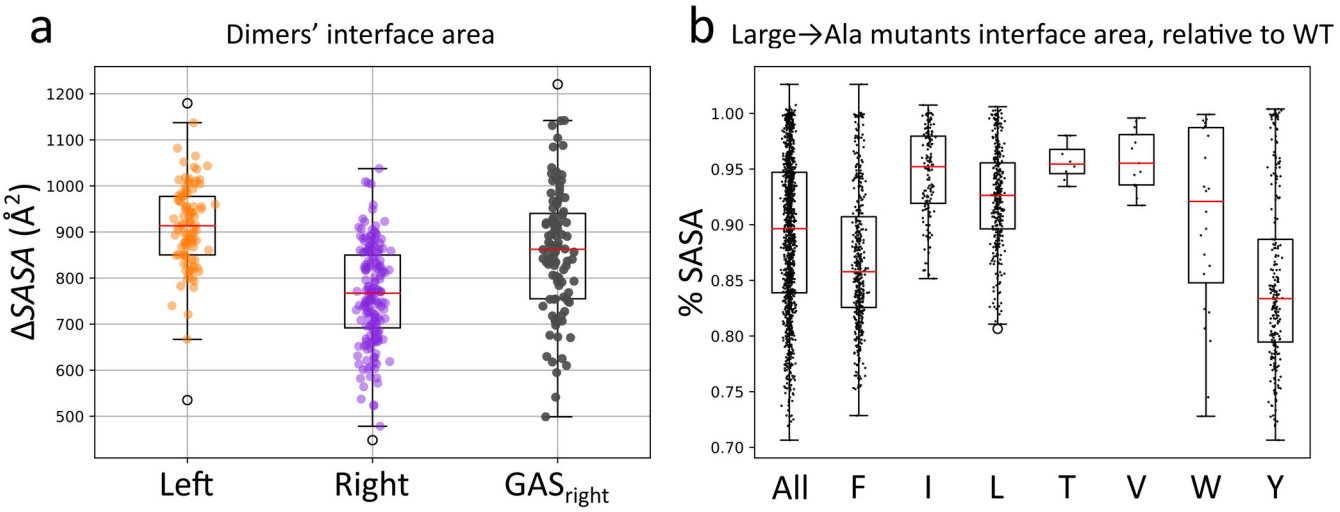


Fig. S5. Surface area of the design's dimer interaction interface and effect of mutation. a) Distribution of the surface area of the interaction interface in the computational models of the final set of 613 Left, Right and GAS_{right} designs. The interaction surface areas were calculated by subtracting the Solvent Accessible Surface Area (SASA) of the helix in a monomeric configuration from the SASA of the dimer. b) Change in interaction interface area of various types of mutants normalized to the area of their respective wild-type sequences.

Table S1. Dataset of validated designs. This dataset includes all constructs whose whose Clash mutants were in the monomeric range using >35% of the GpA TOXGREEN signal as the criterion.

ID	Sequence	TOXGREEN (GpA % ± S.D.)	Reconstructed fluorescence (RFU ± S.D.)	Calculated ΔG (kcal/mol ± S.D.)	SASA (Å)
Left Library					
L_001	RAS LLLLYVLLATLLALLSF LILI	28 ± 8%	6627 ± 1603		881
L_003	RAS LLLLAILLLLLTVLFSL LILI	18 ± 8%	4193 ± 1659		898
L_004	RAS LLLLFVLTILLALLYSL LILI	19 ± 10%	4488 ± 1962		721
L_005	RAS LLLVALLLILTFLLASL LILI	35 ± 6%	8314 ± 1235	-1.97 ± 0.35	923
L_007	RAS LLLLWTLIALLLVLFSL LILI	32 ± 8%	7451 ± 1641	-1.09 ± 0.29	1044
L_008	RAS LLLLFVLLALLTILLSL LILI	15 ± 5%	3450 ± 991		667
L_009	RAS LLLYALLLLLVTLLFSL LILI	22 ± 8%	5172 ± 1497		878
L_011	RAS LLLYALLFVLLLLLTSL LILI	17 ± 7%	4017 ± 1305		804
L_013	RAS LLLLLALLLILVTLLSF LILI	32 ± 10%	7549 ± 1995	-1.29 ± 0.41	831
L_014	RAS LLLLYVLIALLTLLFSL LILI	28 ± 5%	6551 ± 1007		823
L_015	RAS LLLVALLFLLTSLLYLL LILI	21 ± 7%	4962 ± 1302		1065
L_018	RAS LLLVALLTILLALLFSL LILI	41 ± 15%	9726 ± 2906	-2.45 ± 0.88	819
L_019	RAS LLLYALLFLLTSLLFSL LILI	14 ± 8%	3401 ± 1643		1011
L_020	RAS LLLLFALLTILVALLSL LILI	36 ± 5%	8556 ± 907	-2.09 ± 0.26	899
L_022	RAS LLLLLLLLAVLYTLLSF LILI	23 ± 4%	5512 ± 820		850
L_024	RAS LLLFSLLVLLTALLYLL LILI	23 ± 5%	5471 ± 1070		943
L_025	RAS LLLYFLLAILVTLLLSL LILI	19 ± 11%	4483 ± 2079		977
L_029	RAS LLLLAILLTLLAVLFSL LILI	61 ± 6%	14387 ± 1162	-3.09 ± 0.3	840
L_033	RAS LLLLYLLLFILAVLLST LILI	21 ± 9%	4887 ± 1756		940
L_036	RAS LLLLYVLLALLFTLISL LILI	18 ± 12%	4184 ± 2264		882
L_038	RAS LLLLFLLLAILLVLLST LILI	21 ± 5%	4950 ± 906		867
L_039	RAS LLLLYVLLALLFTLLSI LILI	23 ± 7%	5493 ± 1378		872
L_043	RAS LLLLAYLLALLFLLTSL LILI	32 ± 7%	7491 ± 1423	-1.18 ± 0.27	1137
L_045	RAS LLLLLLLLTALLIFLVSL LILI	23 ± 9%	5343 ± 1783		905
L_050	RAS LLLLAVLLFLLALLTSL LILI	41 ± 8%	9723 ± 1654	-2.45 ± 0.5	978
L_051	RAS LLLLTILLLVLALLLSF LILI	23 ± 19%	5368 ± 3764		848
L_052	RAS LLLVTLLLALLLLLFSL LILI	27 ± 6%	6242 ± 1250		955
L_054	RAS LLLYALLFVLTALLLSL LILI	49 ± 8%	11588 ± 1580	-2.78 ± 0.45	906
L_055	RAS LLLLYLLLAVLTALLFS LILI	27 ± 9%	6399 ± 1734		947
L_056	RAS LLLLLFLLATLLILLSV LILI	25 ± 5%	5776 ± 958		793
L_057	RAS LLLLLLLLFVLATLLSI LILI	23 ± 7%	5316 ± 1405		921
L_058	RAS LLLLYALLFLLTALLFS LILI	45 ± 14%	10490 ± 2794	-2.61 ± 0.83	1008
L_060	RAS LLLLTLLLALIVLLFSL LILI	47 ± 12%	11067 ± 2371	-2.7 ± 0.69	849
L_061	RAS LLLYLLLTVLALLLFSL LILI	23 ± 6%	5287 ± 1228		867
L_062	RAS LLLLVLLLTILFLLLSA LILI	23 ± 4%	5347 ± 771		800
L_063	RAS LLLLYFLLALLAVLTSL LILI	27 ± 9%	6246 ± 1803		929
L_064	RAS LLLLTLLVALLLFLISL LILI	41 ± 26%	9542 ± 5102	-2.4 ± 1.54	969
L_065	RAS LLLLYVLLATLLFLLSL LILI	41 ± 16%	9677 ± 3106	-2.44 ± 0.94	983
L_067	RAS LLLVALLLLLFLLLTSL LILI	48 ± 23%	11353 ± 4520	-2.75 ± 1.31	884
L_068	RAS LLLLLLLLFILAVLLST LILI	16 ± 5%	3845 ± 1032		825
L_069	RAS LLLLAVLLFLLATLISL LILI	22 ± 5%	5252 ± 997		916
L_070	RAS LLLVSLLILLTALLFLL LILI	35 ± 11%	8174 ± 2132	-1.89 ± 0.59	992
L_072	RAS LLLLLFLTALLYVLISL LILI	28 ± 27%	6518 ± 5327		888
L_074	RAS LLLLALLLLFLTVLLSI LILI	24 ± 7%	5724 ± 1430		831
L_079	RAS LLLFALLLLLVTLLYSL LILI	29 ± 7%	6835 ± 1375		999
L_080	RAS LLLLAILLAVLFLLLST LILI	25 ± 11%	5900 ± 2165		913
L_081	RAS LLLVLLLAILFLLLTSL LILI	24 ± 9%	5648 ± 1697		850
L_082	RAS LLLYTLLAVLLLLLASL LILI	15 ± 5%	3526 ± 1009		882
L_084	RAS LLLLFVLLLLLATLISL LILI	27 ± 8%	6418 ± 1512		955
L_086	RAS LLLLAVLTALLLLLFSL LILI	21 ± 13%	4940 ± 2649		954
L_087	RAS LLLLYVLTALLFLLISL LILI	26 ± 10%	6181 ± 1973		1019
L_095	RAS LLLLAILLALLFVLTSL LILI	30 ± 10%	7064 ± 2005		948
L_097	RAS LLLLAYLTVLLSILFLL LILI	38 ± 8%	8998 ± 1606	-2.25 ± 0.48	740
L_099	RAS LLLLVSLLILLTALLFL LILI	41 ± 11%	9612 ± 2213	-2.42 ± 0.67	917
L_104	RAS LLLLVLLLTILLFLLSA LILI	21 ± 9%	4946 ± 1726		950
L_106	RAS LLLLWYLVSLLILLTFL LILI	48 ± 16%	11370 ± 3068	-2.75 ± 0.89	867
L_108	RAS LLLLFVLLAILYLLLST LILI	35 ± 8%	8165 ± 1473	-1.89 ± 0.41	994
L_110	RAS LLLYALLLTLIVLLFSL LILI	55 ± 30%	12846 ± 5802	-2.94 ± 1.59	815
L_113	RAS LLLLLLLLAVLTALLFI LILI	28 ± 6%	6623 ± 1111		895
L_114	RAS LLLLVALLLLLTSLLFL LILI	48 ± 5%	11172 ± 889	-2.72 ± 0.26	1037
L_115	RAS LLLYALLLVLITLLFSL LILI	35 ± 12%	8269 ± 2276	-1.95 ± 0.64	849
L_118	RAS LLLLVALLLLLTALLFI LILI	43 ± 16%	10084 ± 3102	-2.53 ± 0.93	1004
L_119	RAS LLLFALLLILTALLYVL LILI	37 ± 12%	8635 ± 2364	-2.12 ± 0.69	942
L_120	RAS LLLLFYLTALLILLVSL LILI	28 ± 15%	6652 ± 2973		816
L_122	RAS LLLLLALLFVLTSLLYL LILI	41 ± 4%	9688 ± 783	-2.44 ± 0.24	1016
L_124	RAS LLLLFWLLAVLTALLYL LILI	38 ± 6%	8926 ± 1236	-2.23 ± 0.37	1179
L_130	RAS LLLLTVLLALLFLLISL LILI	20 ± 12%	4790 ± 2306		934
L_133	RAS LLLLVALLLLLTSLIFL LILI	27 ± 11%	6347 ± 2074		1041
L_136	RAS LLLLYSLTALLFVLLLL LILI	27 ± 6%	6356 ± 1270		987
L_138	RAS LLLLYVLIALLLLLTSL LILI	36 ± 29%	8543 ± 5667	-2.08 ± 1.65	1009
L_139	RAS LLLLYWLLTILAFLLSL LILI	40 ± 13%	9315 ± 2541	-2.34 ± 0.77	1013
L_145	RAS LLLLYVLLLALTILLFS LILI	19 ± 11%	4574 ± 2067		896
L_146	RAS LLLLSYLVALLLTLFLL LILI	37 ± 28%	8806 ± 5416	-2.18 ± 1.61	967
L_147	RAS LLLAALLLLLVTLLFSL LILI	53 ± 8%	12461 ± 1539	-2.89 ± 0.43	934
L_148	RAS LLLYWLLTVLFLLLASL LILI	28 ± 3%	6695 ± 553		921

L_149	RAS LLLLAVLLALLLILLST LILI	20 ± 2%	4785 ± 409		851
L_150	RAS LLLYSLLLLLTFLLAVL LILI	20 ± 18%	4791 ± 3482		978
L_152	RAS LLLLALLFLLTSLLYVL LILI	36 ± 4%	8350 ± 766	-1.99 ± 0.22	1052
L_156	RAS LLLLYFLLATLVALLLS LILI	47 ± 13%	11046 ± 2494	-2.7 ± 0.73	956
L_160	RAS LLLLTVLLAILFLLLSL LILI	29 ± 9%	6855 ± 1700		915
L_161	RAS LLLLAVLLLILTLLLSF LILI	38 ± 20%	8821 ± 3830	-2.19 ± 1.14	796
L_162	RAS LLLLYWLLTLLVALLSF LILI	42 ± 6%	9941 ± 1139	-2.5 ± 0.34	780
L_163	RAS LLLLILLLALLFVLTSL LILI	35 ± 14%	8161 ± 2756	-1.89 ± 0.76	861
L_164	RAS LLLLYILLALLFVLLST LILI	22 ± 2%	5156 ± 303		535
L_165	RAS LLLLLVLLAILLFLLST LILI	24 ± 10%	5712 ± 1882		816
L_168	RAS LLLYALLFVLLILLFSL LILI	14 ± 5%	3390 ± 932		783
L_173	RAS LLLLYLLLAVLLSLLFT LILI	26 ± 8%	6180 ± 1489		1082
L_174	RAS LLLYFLLAVLLTLLFSL LILI	16 ± 9%	3757 ± 1725		876
L_175	RAS LLLLAVLLLTLALLLSF LILI	26 ± 3%	6076 ± 643		875
L_176	RAS LLLYTLLAVLIALLLSL LILI	44 ± 9%	10245 ± 1730	-2.56 ± 0.52	893
L_182	RAS LLLYFLLAVLITLLLSL LILI	18 ± 9%	4137 ± 1857		934
L_184	RAS LLLLFILLAVLLTLLLS LILI	22 ± 13%	5256 ± 2459		814
L_185	RAS LLLYLLLAVLTSLLFLL LILI	14 ± 10%	3367 ± 1988		886
L_186	RAS LLLLVLLLAILFTLLSL LILI	29 ± 7%	6858 ± 1309		839
L_188	RAS LLLVALLLFLLSLLYTL LILI	27 ± 8%	6418 ± 1636		1005
L_190	RAS LLLLFLLLLALTVLLSI LILI	33 ± 10%	7664 ± 1928	-1.46 ± 0.44	960
L_192	RAS LLLLTYLLAILFVLLLS LILI	25 ± 12%	5884 ± 2351		1014
L_195	RAS LLLLFLLLAVLYLLLST LILI	32 ± 14%	7472 ± 2811	-1.14 ± 0.51	965
L_197	RAS LLLLVALLLILFLLLST LILI	28 ± 8%	6551 ± 1561		904
L_198	RAS LLLLSYLLAVLFTLLLI LILI	34 ± 8%	8059 ± 1490	-1.82 ± 0.4	983
Right Library					
R_002	RAS LLLYTLLVSLLALLLAF LILI	31 ± 7%	7071 ± 1374	0.43 ± 0.1	667
R_004	RAS LLLYLLLFTLLAVLLSI LILI	30 ± 6%	6843 ± 1116		661
R_006	RAS LLLYALLVALLFTLLSL LILI	26 ± 5%	5872 ± 1025		747
R_009	RAS LLLTYLLFVLLAILLSL LILI	29 ± 10%	6627 ± 1965		860
R_010	RAS LLLYTLLVLLLAFLLSI LILI	30 ± 7%	6931 ± 1288		662
R_011	RAS LLLLVLLIALLFTLLSY LILI	33 ± 4%	7623 ± 827	-1.62 ± 0.21	745
R_012	RAS LLLYTLLIALLFVLLSL LILI	28 ± 4%	6437 ± 780		1004
R_013	RAS LLLYSLLVALLFTLLFL LILI	23 ± 10%	5298 ± 1984		928
R_014	RAS LLLYTLLISLLFVLLAL LILI	31 ± 8%	7077 ± 1616	0.25 ± 0.07	862
R_015	RAS LLLSYLLFVLLALLLST LILI	39 ± 7%	8861 ± 1317	-2.28 ± 0.4	841
R_016	RAS LLLSYLLTLLLAFLLSI LILI	55 ± 8%	12574 ± 1628	-2.94 ± 0.45	861
R_019	RAS LLLYFLLVALLTLLLSI LILI	27 ± 9%	6149 ± 1743		775
R_020	RAS LLLSYLLFTLLALLLAV LILI	38 ± 3%	8827 ± 515	-2.26 ± 0.15	891
R_021	RAS LLLYTLLISLLVALLFL LILI	25 ± 10%	5759 ± 2025		1038
R_024	RAS LLLYSLLLALLFVLLTI LILI	30 ± 6%	6982 ± 1110		860
R_025	RAS LLLYSLLIALLFVLLTL LILI	28 ± 10%	6395 ± 2044		959
R_027	RAS LLLSYLLVLLLAFLLTI LILI	35 ± 8%	7984 ± 1557	-1.9 ± 0.43	855
R_028	RAS LLLYSLLLTLLFVLLAL LILI	30 ± 13%	6799 ± 2568		770
R_029	RAS LLLSYLLFTLLAILLTL LILI	30 ± 9%	6945 ± 1773		898
R_030	RAS LLLWTLLVFLLAILLSL LILI	29 ± 6%	6546 ± 1170		888
R_033	RAS LLLLYLLVALLISLLFT LILI	39 ± 18%	9033 ± 3567	-2.33 ± 1.07	770
R_034	RAS LLLYALLLTLLVALLSI LILI	33 ± 5%	7610 ± 1038	-1.61 ± 0.26	748
R_036	RAS LLLYALLIALLFTLLVL LILI	30 ± 3%	6866 ± 517		645
R_038	RAS LLLYALLTTLLSVLLFL LILI	44 ± 7%	10095 ± 1455	-2.58 ± 0.43	752
R_039	RAS LLLTALLIALLFVLLSL LILI	33 ± 1%	7578 ± 165	-1.57 ± 0.04	818
R_040	RAS LLLYSLLLTLLVALLIF LILI	24 ± 11%	5465 ± 2086		668
R_041	RAS LLLYSLLLALLATLLVF LILI	24 ± 7%	5592 ± 1288		729
R_043	RAS LLLYVLLTILLALLLSF LILI	34 ± 13%	7894 ± 2632	-1.84 ± 0.72	478
R_045	RAS LLLYLLLATLLFVLLSI LILI	31 ± 3%	7216 ± 646	-0.91 ± 0.1	838
R_049	RAS LLLYTLLIALLAVLLSL LILI	37 ± 5%	8376 ± 1069	-2.1 ± 0.31	864
R_050	RAS LLLFYLLITLLVALLSL LILI	34 ± 9%	7823 ± 1841	-1.79 ± 0.49	659
R_052	RAS LLLYLLLATLLAILLFS LILI	14 ± 6%	3184 ± 1124		823
R_053	RAS LLLYILLTALLVALLSL LILI	31 ± 5%	7224 ± 1007	-0.94 ± 0.15	696
R_055	RAS LLLYALLVALLSLLLFT LILI	35 ± 3%	7946 ± 504	-1.87 ± 0.14	824
R_056	RAS LLLYVLLIALLATLLSL LILI	33 ± 6%	7529 ± 1219	-1.52 ± 0.29	601
R_057	RAS LLLYSLLTALLVALLFL LILI	29 ± 8%	6560 ± 1564		822
R_058	RAS LLLYALLLTLLSILLVF LILI	22 ± 4%	5017 ± 773		727
R_059	RAS LLLYALLLALLFSLLVT LILI	30 ± 10%	6796 ± 2019		832
R_062	RAS LLLYALLVALLITLLFL LILI	30 ± 7%	6898 ± 1379		652
R_063	RAS LLLLYLLTALLIALLFV LILI	27 ± 6%	6187 ± 1201		914
R_065	RAS LLLYILLATLLAFLLSL LILI	27 ± 12%	6142 ± 2359		873
R_067	RAS LLLYSLLVALLATLLFL LILI	26 ± 7%	6071 ± 1329		870
R_068	RAS LLLAILLVTLLFTLLSL LILI	36 ± 9%	8158 ± 1693	-1.99 ± 0.48	754
R_069	RAS LLLYALLTALLSVLLFL LILI	37 ± 3%	8390 ± 611	-2.1 ± 0.18	702
R_071	RAS LLLFSLLLTLLVALLSY LILI	26 ± 8%	5910 ± 1491		855
R_072	RAS LLLYILLFALLTVLLSL LILI	26 ± 6%	5859 ± 1267		647
R_075	RAS LLLVYLLTALLIALLSL LILI	37 ± 5%	8481 ± 1049	-2.14 ± 0.31	875
R_076	RAS LLLYSLLLTLLVALLSF LILI	29 ± 10%	6730 ± 2053		695
R_077	RAS LLLYALLVALLTALLSL LILI	35 ± 8%	8103 ± 1477	-1.96 ± 0.42	805
R_078	RAS LLLFYLLVALLLTLLSL LILI	40 ± 8%	9163 ± 1575	-2.36 ± 0.48	769
R_079	RAS LLLTALLLALLAVLLSF LILI	40 ± 6%	9119 ± 1252	-2.35 ± 0.38	772
R_080	RAS LLLYILLLTLLAVLLSF LILI	24 ± 4%	5609 ± 826		614
R_081	RAS LLLYFLLVALLLALLST LILI	32 ± 3%	7281 ± 617	-1.1 ± 0.11	820
R_083	RAS LLLTALLVALLAYLLSF LILI	50 ± 4%	11529 ± 843	-2.81 ± 0.24	774
R_084	RAS LLLVSLLLTLLFALLAY LILI	29 ± 3%	6761 ± 644		824
R_085	RAS LLLYSLLLTLLATLLAF LILI	21 ± 8%	4808 ± 1512		680
R_086	RAS LLLYLLLATLLAVLLSF LILI	25 ± 7%	5760 ± 1403		655
R_089	RAS LLLFYLLVILLALLLST LILI	29 ± 14%	6579 ± 2662		768
R_090	RAS LLLFYLLVALLISLLTL LILI	34 ± 5%	7824 ± 893	-1.79 ± 0.24	896

R_092	RAS LLLAVLLLTLLFALLSY LILI	26 ± 6%	5881 ± 1110		736
R_093	RAS LLLLILLSTLLATLLVF LILI	23 ± 5%	5331 ± 1018		709
R_094	RAS LLLYILLLTLLVALLSF LILI	28 ± 1%	6363 ± 212		687
R_095	RAS LLLLALLFALLVSLLTW LILI	30 ± 0%	6958 ± 93		771
R_097	RAS LLLTYLLLVLLAILLSF LILI	19 ± 12%	4471 ± 2338		668
R_099	RAS LLLYVLLIALLFTLLSL LILI	26 ± 7%	5914 ± 1446		738
R_100	RAS LLLYVLLTILLALLLAF LILI	32 ± 9%	7425 ± 1850	-1.38 ± 0.4	604
R_102	RAS LLLYLLLSTLLAFLLAI LILI	28 ± 4%	6316 ± 837		836
R_105	RAS LLLTYLLIVLLALLLSF LILI	41 ± 2%	9290 ± 338	-2.4 ± 0.1	752
R_106	RAS LLLYLLLFTLLAALLTS LILI	12 ± 6%	2813 ± 1238		753
R_108	RAS LLLTYLLAVLLAILLFL LILI	32 ± 7%	7236 ± 1323	-0.98 ± 0.21	886
R_110	RAS LLLYALLFALLLSLLVT LILI	27 ± 2%	6097 ± 397		767
R_111	RAS LLLYVLLTTLLFALLSL LILI	28 ± 8%	6513 ± 1529		739
R_114	RAS LLLYVLLAILLTLLLSF LILI	27 ± 10%	6117 ± 1926		685
R_115	RAS LLLAALLLSLLFTLLVF LILI	46 ± 5%	10458 ± 917	-2.65 ± 0.27	775
R_116	RAS LLLYLLLATLLAVLLSI LILI	24 ± 7%	5435 ± 1416		721
R_117	RAS LLLYALLLSLLTALLVF LILI	28 ± 9%	6367 ± 1834		572
R_118	RAS LLLTALLIALLFLLLSV LILI	40 ± 4%	9163 ± 826	-2.36 ± 0.25	816
R_119	RAS LLLYVLLFALLTVLLSL LILI	30 ± 2%	6859 ± 423		608
R_120	RAS LLLYSLLFVLLAILLTL LILI	25 ± 10%	5675 ± 1942		664
R_121	RAS LLLAYLLVLLLATLLSF LILI	34 ± 7%	7719 ± 1462	-1.71 ± 0.38	770
R_124	RAS LLLYVLLATLLALLLSF LILI	26 ± 3%	5984 ± 622		653
R_127	RAS LLLTVLLLALLFALLSI LILI	36 ± 6%	8183 ± 1102	-2.01 ± 0.32	723
R_129	RAS LLLYVLLFTLLAALLSL LILI	28 ± 6%	6450 ± 1178		696
R_131	RAS LLLYVLLLTLLFALLSL LILI	30 ± 11%	6962 ± 2187		758
R_132	RAS LLLVYLLLALLTSLLFI LILI	39 ± 3%	8952 ± 542	-2.3 ± 0.16	708
R_134	RAS LLLYSLLLILLAVLLTF LILI	24 ± 7%	5579 ± 1292		700
R_138	RAS LLLTYLLVFLLALLLSI LILI	34 ± 9%	7687 ± 1769	-1.68 ± 0.45	905
R_139	RAS LLLYALLLTLLSVLLFI LILI	30 ± 7%	6970 ± 1350		773
R_143	RAS LLLALLLVTLLAILLSF LILI	17 ± 7%	3955 ± 1386		586
R_147	RAS LLLTVLLALLLAFLLSI LILI	35 ± 6%	7985 ± 1194	-1.9 ± 0.33	662
R_148	RAS LLLTYLLVALLIALLFL LILI	37 ± 5%	8424 ± 1026	-2.12 ± 0.3	949
R_149	RAS LLLYALLLTLLVALLSF LILI	25 ± 8%	5767 ± 1554		525
R_153	RAS LLLYLLLVTLLAILLSF LILI	17 ± 9%	3970 ± 1779		537
R_155	RAS LLLYLLLTTLLVALLSF LILI	44 ± 5%	10102 ± 1022	-2.58 ± 0.31	673
R_156	RAS LLLLVLLITLLFALLSY LILI	32 ± 6%	7416 ± 1256	-1.37 ± 0.27	869
R_160	RAS LLLTYLLVALLAILLSL LILI	28 ± 2%	6362 ± 481		881
R_161	RAS LLLYSLLLTLLAVLLFS LILI	21 ± 10%	4824 ± 1906		791
R_162	RAS LLLLYLLITLLFSLLVA LILI	30 ± 5%	6890 ± 1051		627
R_163	RAS LLLYSLLLALLFALLVT LILI	27 ± 5%	6159 ± 1071		817
R_164	RAS LLLYSLLVALLFTLLIL LILI	28 ± 12%	6389 ± 2275		889
R_167	RAS LLLFSLLLTLLAALLSY LILI	21 ± 7%	4871 ± 1454		619
R_168	RAS LLLYILLSTLLALLLAF LILI	21 ± 4%	4878 ± 727		631
R_172	RAS LLLYTLLATLLFLLLSV LILI	27 ± 11%	6293 ± 2125		705
R_173	RAS LLLYVLLFILLALLLST LILI	22 ± 10%	4956 ± 1918		564
R_174	RAS LLLYILLSVLLALLLTF LILI	19 ± 8%	4415 ± 1504		722
R_176	RAS LLLYTLLVALLFALLSL LILI	30 ± 7%	6902 ± 1290		923
R_178	RAS LLLYLLLAVLLAFLLTS LILI	24 ± 9%	5502 ± 1814		823
R_180	RAS LLLFYLLVALLTALLSL LILI	39 ± 3%	8975 ± 562	-2.31 ± 0.17	713
R_182	RAS LLLVYLLITLLFALLSL LILI	43 ± 4%	9780 ± 841	-2.51 ± 0.25	816
R_183	RAS LLLYSLLITLLFVLLAL LILI	25 ± 9%	5674 ± 1793		746
R_184	RAS LLLYVLLLTLLATLLSF LILI	22 ± 12%	5011 ± 2302		603
R_185	RAS LLLYWLLTALLLSLLFV LILI	21 ± 5%	4820 ± 973		997
R_186	RAS LLLTYLLFTLLAVLLSL LILI	36 ± 4%	8197 ± 877	-2.01 ± 0.25	882
R_187	RAS LLLYWLLVALLISLLTL LILI	32 ± 3%	7276 ± 492	-1.09 ± 0.09	1007
R_188	RAS LLLYTLLIALLVLLLFS LILI	31 ± 18%	7120 ± 3620	-0.41 ± 0.24	820
R_189	RAS LLLYTLLVSLLFALLIL LILI	23 ± 6%	5206 ± 1144		865
R_190	RAS LLLVYLLTALLISLLFL LILI	37 ± 7%	8556 ± 1351	-2.17 ± 0.4	1009
R_191	RAS LLLWYLLLVLLATLLSF LILI	35 ± 7%	8067 ± 1380	-1.95 ± 0.39	732
R_192	RAS LLLYSLLLTLLATLLVF LILI	21 ± 8%	4807 ± 1605		662
R_193	RAS LLLYILLATLLTLLLSF LILI	21 ± 11%	4907 ± 2208		617
R_194	RAS LLLYSLLVALLFALLTL LILI	37 ± 7%	8525 ± 1339	-2.16 ± 0.4	892
R_195	RAS LLLFVLLATLLAYLLSL LILI	36 ± 3%	8318 ± 496	-2.07 ± 0.14	910
R_196	RAS LLLSYLLVFLLATLLSL LILI	39 ± 6%	9011 ± 1113	-2.32 ± 0.34	885
R_197	RAS LLLVYLLLALLTALLSI LILI	36 ± 11%	8313 ± 2182	-2.07 ± 0.63	829
R_198	RAS LLLYSLLLTLLALLLAF LILI	30 ± 8%	6992 ± 1553		667
R_199	RAS LLLYSLLLTLLFALLVI LILI	36 ± 2%	8193 ± 418	-2.01 ± 0.12	782
R_200	RAS LLLYVLLATLLAFLLSL LILI	29 ± 6%	6646 ± 1275		740
R_202	RAS LLLLYLLVTLLFALLSI LILI	37 ± 7%	8503 ± 1350	-2.15 ± 0.4	775
R_204	RAS LLLATLLISLLFTLLVL LILI	43 ± 9%	9772 ± 1857	-2.51 ± 0.56	902
R_206	RAS LLLYLLLATLLFVLLSL LILI	21 ± 12%	4819 ± 2289		812
R_207	RAS LLLYALLLALLVSLLFT LILI	38 ± 6%	8637 ± 1268	-2.2 ± 0.38	827
R_208	RAS LLLYSLLLALLFALLTI LILI	45 ± 6%	10229 ± 1202	-2.6 ± 0.36	827
R_209	RAS LLLYALLTSLLVALLFL LILI	34 ± 2%	7740 ± 302	-1.73 ± 0.08	798
R_210	RAS LLLYVLLATLLTFLLSL LILI	29 ± 2%	6707 ± 435		753
R_211	RAS LLLYALLTILLTFLLSL LILI	28 ± 15%	6477 ± 2881		731
R_213	RAS LLLVYLLIALLFSLLTL LILI	34 ± 4%	7724 ± 877	-1.71 ± 0.23	857
R_215	RAS LLLLSLLFALLIALLVT LILI	29 ± 3%	6663 ± 672		749
R_216	RAS LLLYILLTVLLALLLSF LILI	26 ± 5%	5952 ± 962		522
R_217	RAS LLLYSLLVALLFLLLAT LILI	28 ± 10%	6329 ± 2000		613
R_219	RAS LLLYFLLAILLAVLLTL LILI	24 ± 6%	5448 ± 1080		835
R_221	RAS LLLVALLLALLFALLST LILI	36 ± 4%	8188 ± 787	-2.01 ± 0.23	862
R_222	RAS LLLLALLFALLTVLLSI LILI	44 ± 2%	10127 ± 442	-2.59 ± 0.13	738
R_223	RAS LLLVALLLTLLAILLSF LILI	28 ± 5%	6414 ± 999		680
R_224	RAS LLLYILLVALLATLLTL LILI	34 ± 11%	7898 ± 2224	-1.84 ± 0.61	621

R_225	RAS LLLYLLLTVLLAFLLSI LILI	27 ± 5%	6209 ± 988		582
R_230	RAS LLLFVLLTTLLAILLSL LILI	22 ± 2%	4945 ± 381		636
R_232	RAS LLLLALLVALLSTLLIF LILI	39 ± 6%	8988 ± 1223	-2.31 ± 0.37	703
R_233	RAS LLLTYLLIALLLALLFV LILI	40 ± 1%	9114 ± 182	-2.35 ± 0.05	871
R_237	RAS LLLYSLLVALLTLLLFI LILI	31 ± 13%	7196 ± 2459	-0.84 ± 0.34	812
R_238	RAS LLLFYLLLTLLVALLSI LILI	48 ± 6%	11016 ± 1227	-2.74 ± 0.36	782
R_239	RAS LLLYSLLLALLAVLLTI LILI	31 ± 22%	7074 ± 4413	0.32 ± 0.23	675
R_240	RAS LLLYLLLVALLATLLSF LILI	22 ± 8%	5059 ± 1512		583
R_241	RAS LLLYTLLATLLVLLLSF LILI	25 ± 7%	5792 ± 1369		771
R_243	RAS LLLTYLLTILLAFLLSL LILI	41 ± 5%	9444 ± 886	-2.44 ± 0.27	857
R_244	RAS LLLTYLLVALLIALLSL LILI	36 ± 2%	8177 ± 298	-2 ± 0.09	796
R_245	RAS LLLLTLLIALLVALLSF LILI	34 ± 7%	7835 ± 1374	-1.8 ± 0.37	712
R_246	RAS LLLYVLLTALLFALLSL LILI	32 ± 6%	7236 ± 1130	-0.98 ± 0.18	792
R_247	RAS LLLVYLLLALLFTLLSL LILI	47 ± 6%	10676 ± 1239	-2.68 ± 0.36	778
R_249	RAS LLLYSLLLTLLAALLVF LILI	23 ± 8%	5230 ± 1489		715
R_250	RAS LLLLYLLIALLVALLST LILI	30 ± 5%	6822 ± 973		741
R_251	RAS LLLYTLLVALLFLLLSL LILI	27 ± 10%	6256 ± 1949		859
R_252	RAS LLLFYLLTALLIALLSL LILI	38 ± 3%	8723 ± 502	-2.23 ± 0.15	879
R_253	RAS LLLYWLLVALLLALLST LILI	36 ± 1%	8302 ± 135	-2.06 ± 0.04	923
R_254	RAS LLLFALLTILLAYLLSL LILI	40 ± 6%	9073 ± 1138	-2.34 ± 0.34	892
R_256	RAS LLLTYLLLALLVSLLFI LILI	47 ± 9%	10681 ± 1835	-2.68 ± 0.54	802
R_257	RAS LLLYALLVALLLTLLSW LILI	27 ± 4%	6195 ± 880		747
R_258	RAS LLLYWLLLALLVALLST LILI	30 ± 4%	6956 ± 710		752
R_259	RAS LLLYSLLFTLLAVLLAL LILI	32 ± 10%	7439 ± 1907	-1.4 ± 0.42	751
R_260	RAS LLLLALLIALLFVLLST LILI	27 ± 5%	6138 ± 1022		743
R_263	RAS LLLYTLLVSLLAILLTL LILI	33 ± 9%	7579 ± 1857	-1.58 ± 0.45	763
R_264	RAS LLLFYLLTALLITLLVL LILI	31 ± 9%	7028 ± 1736		689
R_265	RAS LLLSYLLVFLLALLLAT LILI	48 ± 7%	11020 ± 1341	-2.74 ± 0.39	908
R_267	RAS LLLTYLLVALLTFLLSL LILI	59 ± 25%	13502 ± 4920	-3.04 ± 1.29	855
R_268	RAS LLLYILLFTLLVALLSL LILI	19 ± 6%	4360 ± 1162		628
R_269	RAS LLLLTLLIALLFVLLYS LILI	43 ± 6%	9826 ± 1145	-2.52 ± 0.34	845
R_270	RAS LLLAILLLTLLFVLLSL LILI	33 ± 6%	7503 ± 1109	-1.49 ± 0.26	713
R_271	RAS LLLYTLLVALLSTLLFL LILI	23 ± 9%	5236 ± 1793		857
R_272	RAS LLLLTLLATLLAVLLSF LILI	37 ± 3%	8373 ± 537	-2.09 ± 0.16	727
R_275	RAS LLLAVLLLTLLFTLLSY LILI	21 ± 11%	4727 ± 2206		715
GAS_{right} library					
G_001	RAS LLLTSLLLGLLFALLAY LILI	22 ± 13%	5270 ± 2604		871
G_003	RAS LLLVTLLLGLLFALLST LILI	25 ± 5%	5995 ± 896		841
G_004	RAS LLLYALLFLLLGTLLSI LILI	17 ± 5%	4209 ± 994		853
G_005	RAS LLLYFLLATLLGLLLGV LILI	36 ± 2%	8621 ± 478	-2.02 ± 0.14	1104
G_006	RAS LLLYVLLGLLLGTLLAF LILI	32 ± 14%	7721 ± 2817	-1.25 ± 0.56	944
G_007	RAS LLLYLLLGALLGILLST LILI	75 ± 8%	18111 ± 1480	-3.35 ± 0.34	862
G_008	RAS LLLYFLLALLLGTLLGS LILI	23 ± 2%	5657 ± 393		1088
G_009	RAS LLLYVLLATLLGLLLGF LILI	30 ± 6%	7260 ± 1187		618
G_010	RAS LLLLVLLATLLGGLLSI LILI	26 ± 5%	6184 ± 912		610
G_012	RAS LLLFALLASLLGILLTL LILI	19 ± 4%	4696 ± 689		920
G_014	RAS LLLYLLLGALLGALLTF LILI	112 ± 17%	27099 ± 3413	-3.82 ± 0.59	806
G_017	RAS LLLYLLLASLLGILLAF LILI	15 ± 2%	3617 ± 413		718
G_020	RAS LLLTLLLGLLLGVLLGS LILI	91 ± 21%	21903 ± 4125	-3.57 ± 0.83	892
G_021	RAS LLLAFLLGTLLGVLLGL LILI	133 ± 24%	32043 ± 4794	-4.02 ± 0.74	1011
G_022	RAS LLLYSLLALLLGTLLGF LILI	15 ± 5%	3664 ± 902		991
G_023	RAS LLLLFLLGTLLGALLAY LILI	115 ± 16%	27769 ± 3197	-3.85 ± 0.54	857
G_024	RAS LLLYLLLGILLGTLLAF LILI	53 ± 4%	12849 ± 730	-2.9 ± 0.2	871
G_025	RAS LLLYTLLVALLALLLGL LILI	33 ± 5%	7843 ± 970	-1.44 ± 0.22	877
G_026	RAS LLLTVLLLGLLFGLLFA LILI	50 ± 5%	12113 ± 1044	-2.81 ± 0.3	801
G_030	RAS LLLAGLLFLLLGVLLAT LILI	33 ± 1%	8009 ± 113	-1.62 ± 0.03	908
G_031	RAS LLLYVLLTLLLGALLGF LILI	26 ± 8%	6361 ± 1638		707
G_032	RAS LLLFVLLTILLGALLSL LILI	25 ± 5%	6109 ± 927		630
G_033	RAS LLLAYLLGVLLGLLLAF LILI	112 ± 9%	27007 ± 1800	-3.81 ± 0.31	975
G_034	RAS LLLYGLLFALLGTLLSL LILI	19 ± 1%	4539 ± 235		754
G_035	RAS LLLLVLLTGLLFGLLYA LILI	31 ± 8%	7449 ± 1550	-0.06 ± 0.01	859
G_036	RAS LLLYALLALLLGTLLGF LILI	19 ± 3%	4664 ± 618		915
G_037	RAS LLLSFLLVLLLGTLLAY LILI	19 ± 7%	4619 ± 1301		804
G_038	RAS LLLAALLALLLGVLLGT LILI	39 ± 4%	9480 ± 745	-2.33 ± 0.22	759
G_040	RAS LLLYVLLFTLLGALLGL LILI	22 ± 3%	5243 ± 592		707
G_041	RAS LLLLTLLVALLAGLLFS LILI	31 ± 3%	7359 ± 537		827
G_042	RAS LLLVALLALLLGTLLSF LILI	28 ± 5%	6828 ± 947		895
G_043	RAS LLLTFLLAVLLGLLLGS LILI	60 ± 14%	14569 ± 2705	-3.08 ± 0.7	872
G_044	RAS LLLTFLLAILLGLLLGV LILI	64 ± 5%	15482 ± 1044	-3.15 ± 0.26	1040
G_045	RAS LLLYVLLALLLGTLLGF LILI	18 ± 3%	4379 ± 497		984
G_046	RAS LLLVTLLFSLLGALLSL LILI	48 ± 5%	11598 ± 955	-2.74 ± 0.28	921
G_047	RAS LLLYILLGLLLGTLLAS LILI	31 ± 4%	7496 ± 768	-0.52 ± 0.07	755
G_048	RAS LLLVTLLLTLLGALLSA LILI	28 ± 3%	6685 ± 584		705
G_049	RAS LLLYGLLFLLLGVLLAT LILI	40 ± 4%	9716 ± 723	-2.39 ± 0.22	920
G_050	RAS LLLLFLLATLLGYLLSI LILI	42 ± 4%	10144 ± 757	-2.49 ± 0.23	981
G_054	RAS LLLYLLLGALLGTLLSF LILI	53 ± 4%	12873 ± 789	-2.9 ± 0.22	739
G_055	RAS LLLTVLLGLLLGLLLGF LILI	54 ± 6%	13131 ± 1096	-2.93 ± 0.3	830
G_056	RAS LLLTALLFLLLGVLLAS LILI	82 ± 11%	19751 ± 2118	-3.45 ± 0.45	916
G_057	RAS LLLTFLLAVLLGLLLGI LILI	65 ± 11%	15745 ± 2129	-3.18 ± 0.53	1141
G_060	RAS LLLYFLLGVLLGTLLGL LILI	107 ± 3%	25829 ± 608	-3.76 ± 0.11	1221
G_061	RAS LLLYLLLTTLLAGLLFS LILI	17 ± 10%	4203 ± 2055		837
G_063	RAS LLLYVLLGALLGILLTL LILI	95 ± 10%	22981 ± 2005	-3.62 ± 0.39	863
G_064	RAS LLLLTLLVGLLFALLYG LILI	25 ± 2%	6130 ± 414		939
G_065	RAS LLLAVLLTSLLGGLLFL LILI	25 ± 4%	6099 ± 840		755
G_066	RAS LLLAALLFSLLGLLLGT LILI	30 ± 9%	7118 ± 1861		784

G_067	RAS LLLTFLLGLLLGYLLAS LILI	123 ± 8%	29664 ± 1648	-3.93 ± 0.27	1000
G_068	RAS LLLVGLLFALLGILLTL LILI	36 ± 5%	8750 ± 995	-2.08 ± 0.29	967
G_069	RAS LLLTFLLALLLGYLLGS LILI	95 ± 14%	22883 ± 2667	-3.62 ± 0.52	890
G_070	RAS LLLYFLLGTLLGALLGL LILI	81 ± 7%	19595 ± 1407	-3.44 ± 0.3	754
G_071	RAS LLLVLLLTGLLFALLAY LILI	25 ± 7%	5947 ± 1359		826
G_074	RAS LLLYLLLGALLAFLLST LILI	18 ± 1%	4355 ± 288		834
G_075	RAS LLLYVLLFLLLGALLGT LILI	17 ± 2%	4138 ± 433		710
G_078	RAS LLLALLLGTLLGFLLSI LILI	126 ± 10%	30386 ± 1928	-3.96 ± 0.31	856
G_079	RAS LLLSYLLGVLLGTLLGF LILI	79 ± 12%	18987 ± 2262	-3.4 ± 0.5	906
G_080	RAS LLLSYLLGLLLGTLLGF LILI	83 ± 6%	20084 ± 1117	-3.47 ± 0.24	952
G_081	RAS LLLTYLLGVLLGLLLGF LILI	87 ± 19%	21057 ± 3692	-3.52 ± 0.76	1015
G_082	RAS LLLSTLLGALLGLLLAV LILI	91 ± 9%	22029 ± 1790	-3.58 ± 0.36	895
G_083	RAS LLLYFLLAVLLGLLLGS LILI	25 ± 15%	5980 ± 2916		1142
G_084	RAS LLLYFLLATLLGVLLGL LILI	34 ± 5%	8232 ± 888	-1.8 ± 0.24	878
G_086	RAS LLLAYLLGVLLGLLLGS LILI	111 ± 12%	26845 ± 2359	-3.81 ± 0.41	1027
G_087	RAS LLLALLLGLLLGVLLGS LILI	104 ± 21%	25123 ± 4037	-3.73 ± 0.74	936
G_088	RAS LLLYGLLFALLGILLTL LILI	28 ± 1%	6712 ± 124		914
G_090	RAS LLLSYLLGVLLGLLLGT LILI	95 ± 8%	22852 ± 1647	-3.62 ± 0.32	1012
G_091	RAS LLLYGLLFLLLGVLLAS LILI	32 ± 6%	7753 ± 1117	-1.31 ± 0.23	834
G_092	RAS LLLLVLLGALLGTLLSF LILI	44 ± 7%	10699 ± 1319	-2.6 ± 0.39	672
G_094	RAS LLLTVLLALLLGFLLGS LILI	78 ± 7%	18720 ± 1373	-3.39 ± 0.3	499
G_095	RAS LLLYSLLFALLGALLTL LILI	36 ± 4%	8567 ± 784	-1.99 ± 0.22	671
G_096	RAS LLLTALLVGLLGGLLFL LILI	29 ± 5%	7053 ± 979		747
G_097	RAS LLLYLLLGALLGVLLST LILI	84 ± 22%	20309 ± 4272	-3.48 ± 0.9	793
G_098	RAS LLLYILLTLLLGALLAF LILI	29 ± 14%	7006 ± 2824		541
G_099	RAS LLLSSLLFLLLGVLLGT LILI	30 ± 6%	7150 ± 1082		892
G_101	RAS LLLYVLLGLLLGFLLST LILI	104 ± 5%	24963 ± 1048	-3.72 ± 0.19	964
G_102	RAS LLLYLLLGALLGFLLST LILI	111 ± 17%	26779 ± 3318	-3.8 ± 0.58	959
G_103	RAS LLLFYLLGVLLGLLLAS LILI	92 ± 10%	22070 ± 1936	-3.58 ± 0.39	1084
G_104	RAS LLLAYLLGTLLGLLLGF LILI	114 ± 15%	27435 ± 2873	-3.83 ± 0.49	1024
G_105	RAS LLLALLLGLLLGTLLSF LILI	104 ± 5%	24963 ± 983	-3.72 ± 0.18	842
G_106	RAS LLLTSLLVGLLAYLLGL LILI	41 ± 2%	9983 ± 312	-2.45 ± 0.09	768
G_107	RAS LLLYTLLAILLGLLLGF LILI	24 ± 3%	5662 ± 632		595
G_108	RAS LLLYLLLGTLLGILLAF LILI	43 ± 7%	10470 ± 1447	-2.55 ± 0.43	676
G_110	RAS LLLSYLLGVLLGLLLAT LILI	103 ± 11%	24834 ± 2211	-3.71 ± 0.41	968
G_112	RAS LLLVLLLATLLGSLLGT LILI	31 ± 7%	7457 ± 1340	-0.17 ± 0.04	727
G_113	RAS LLLTALLLALLFGLLFS LILI	38 ± 3%	9117 ± 672	-2.21 ± 0.2	842
G_114	RAS LLLLALLFALLGILLST LILI	25 ± 12%	6118 ± 2389		718
G_115	RAS LLLYVLLTLLLGGLLFA LILI	32 ± 4%	7600 ± 760	-0.98 ± 0.12	791
G_116	RAS LLLSYLLGLLLGFLLST LILI	109 ± 18%	26195 ± 3519	-3.78 ± 0.62	934
G_117	RAS LLLALLLGVLLGTLLGF LILI	95 ± 12%	22974 ± 2273	-3.62 ± 0.44	912
G_119	RAS LLLYVLLATLLGGLLFL LILI	21 ± 6%	5019 ± 1176		711
G_120	RAS LLLYILLGLLLGTLLAF LILI	25 ± 9%	5904 ± 1715		846
G_121	RAS LLLVTLLLGLLFGLLFA LILI	83 ± 23%	20072 ± 4462	-3.47 ± 0.95	967
G_122	RAS LLLALLLTVLLGGLLFS LILI	26 ± 4%	6337 ± 876		793
G_123	RAS LLLAGLLFLLLGTLLSY LILI	14 ± 2%	3477 ± 416		927
G_124	RAS LLLYALLTALLGGLLFL LILI	22 ± 6%	5399 ± 1205		698
G_126	RAS LLLAVLLFSLLGLLLGT LILI	27 ± 4%	6618 ± 744		706
G_127	RAS LLLYFLLAVLLGLLLGT LILI	28 ± 6%	6728 ± 1098		1131
G_128	RAS LLLLVLLTTLLGGLLFA LILI	29 ± 4%	6926 ± 849		750
G_129	RAS LLLTALLVGLLGLLLGF LILI	27 ± 2%	6534 ± 318		625
G_132	RAS LLLTVLLGILLGLLLAF LILI	71 ± 6%	17082 ± 1082	-3.28 ± 0.25	818
G_133	RAS LLLAYLLGLLLGFLLGT LILI	111 ± 11%	26844 ± 2140	-3.81 ± 0.37	1030
G_135	RAS LLLTALLFLLLGVLLGI LILI	72 ± 6%	17311 ± 1142	-3.29 ± 0.27	868
G_137	RAS LLLTALLAGLLFLLLSV LILI	31 ± 2%	7567 ± 331	-0.87 ± 0.05	920

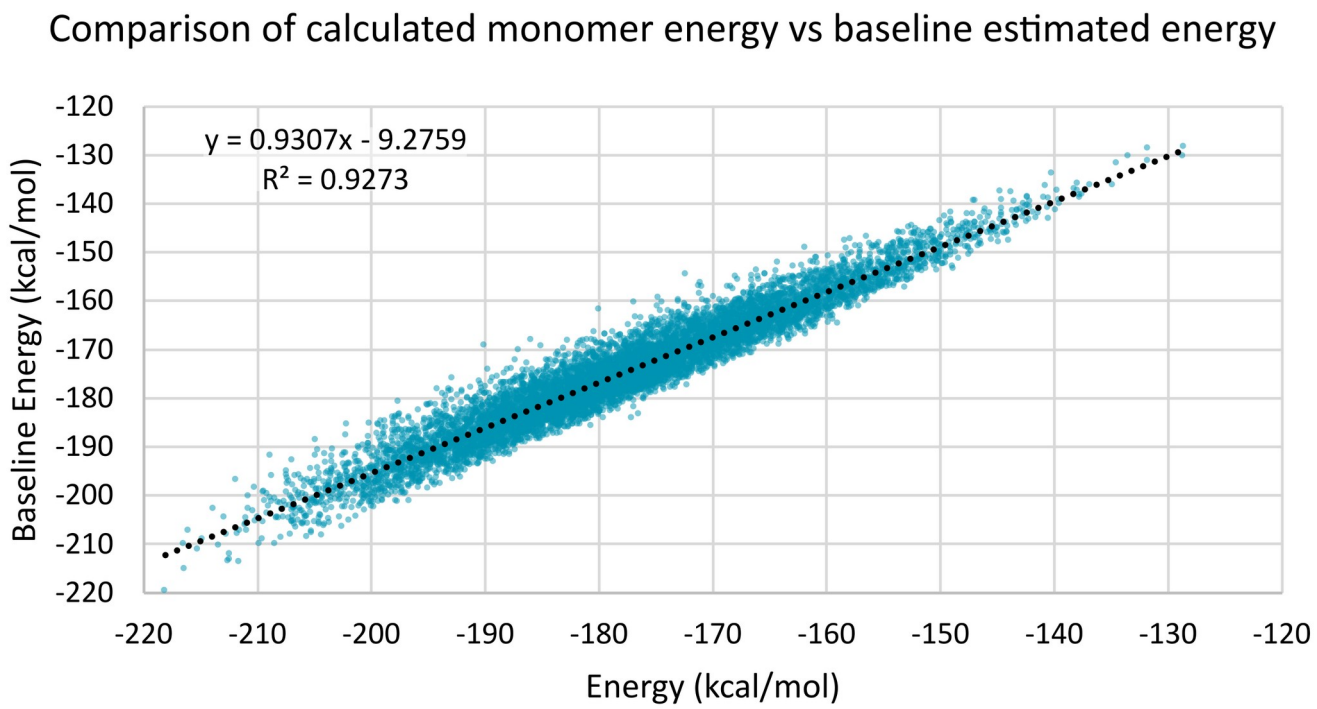


Fig. S6. Calculation of the monomer baseline energies. A monomer energy approximation (baseline term) for approximating the energy of the monomeric helices during design. The term was developed by averaging the interactions of each individual amino acid in a large set of monomeric helices of different amino acid compositions. The baseline energies include a self term that corresponds to the interactions that occur within the single amino acid, and a pair term that corresponds to the interaction of an amino acid with other amino acids in the helix. The figure shows the comparison between the baseline energy and the actual energy computed for the corresponding monomer in 10000 random sequences in a helical configuration.

Table S2. Frequency of the hydrophobic amino acids within the transmembrane region of membrane proteins in the OPM database.

Amino acid	Frequency
ALA	0.143
PHE	0.098
GLY	0.085
ILE	0.121
LEU	0.211
SER	0.073
THR	0.066
VAL	0.126
TRP	0.028
TYR	0.049

Table S3. Standard constructs used for fluorescence reconstruction normalization. The table reports the individually measured GFP fluorescence (TOXGREEN) and their normalized value to the customary GpA standard (% GpA). The table also reports the reconstructed fluorescence values for the same constructs as obtained for the Left, Right and GAS_{right} libraries. Linear fitting was performed for these individual measurement and their library reconstruction vaues, as illustrated in supplementary Fig. S6. The resulting linear relationships were used used to calibrate the conversion of the reconstructed fluorescence values for three libraries to normalized % GpA values.

TM Helix Sequence	TOXGREEN (RFU)	% GpA	Left Library (RFU)	Right Library (RFU)	GAS_{right} Library (RFU)
RAS LIIFGVMAGVIGT ILI	19627 ± 3749	100%	32343 ± 1317	29104 ± 2087	25479 ± 1368
RAS LIIFGVMAIVIGT ILI	5236 ± 2054	27%	27077 ± 1246	24008 ± 1517	18152 ± 1495
RAS FHMIAVGLSSSILGCLITLLVL ILI	31104 ± 2695	158%	34162 ± 1703	31287 ± 2330	27636 ± 1888
RAS VVIIAIVCCVVGTSLVWVVIIL ILI	6355 ± 3823	32%	29941 ± 1410	26540 ± 2108	20793 ± 1382
RAS CAVVVGVGLIVGFAVGLL ILI	7496 ± 615	38%	27836 ± 1125	24509 ± 1625	18670 ± 1389
RAS LLLLAYLLLTLAVLLFSL ILI	15374 ± 7141	78%	29139 ± 1148		
RAS LLLLFALLALLYVLTSLL ILI	12303 ± 916	63%	29650 ± 1264		
RAS LLLYALLFVLTALLLSLL ILI	10838 ± 1766	55%	29414 ± 1400		
RAS LLLLYLLLAVLTALLFSL ILI	8879 ± 3335	45%	28122 ± 712		
RAS LLLLLALLLILFVLLSTL ILI	7647 ± 1822	39%	28730 ± 1225		
RAS LLLFYLLVALLLTLLSTL ILI	9200 ± 1889	47%		25217 ± 1500	
RAS LLLTVLLLALLFALLSIL ILI	7881 ± 1712	40%		25005 ± 1314	
RAS LLLTVLLALLLAFLLSIL ILI	10974 ± 733	56%		24960 ± 1802	
RAS LLLYALLTALLFVLLSLL ILI	10364 ± 616	53%		24771 ± 1597	
RAS LLLVYLLIALLFSLLTLL ILI	9835 ± 1358	50%		24900 ± 1450	
RAS LLLTVLLALLLGFLLGSL ILI	19807 ± 3298	101%			23535 ± 1000
RAS LLLVALLALLLGTLLSFL ILI	11079 ± 1331	56%			19271 ± 1426
RAS LLLTALLIGLLFGLLVLL ILI	8308 ± 5363	42%			20184 ± 824

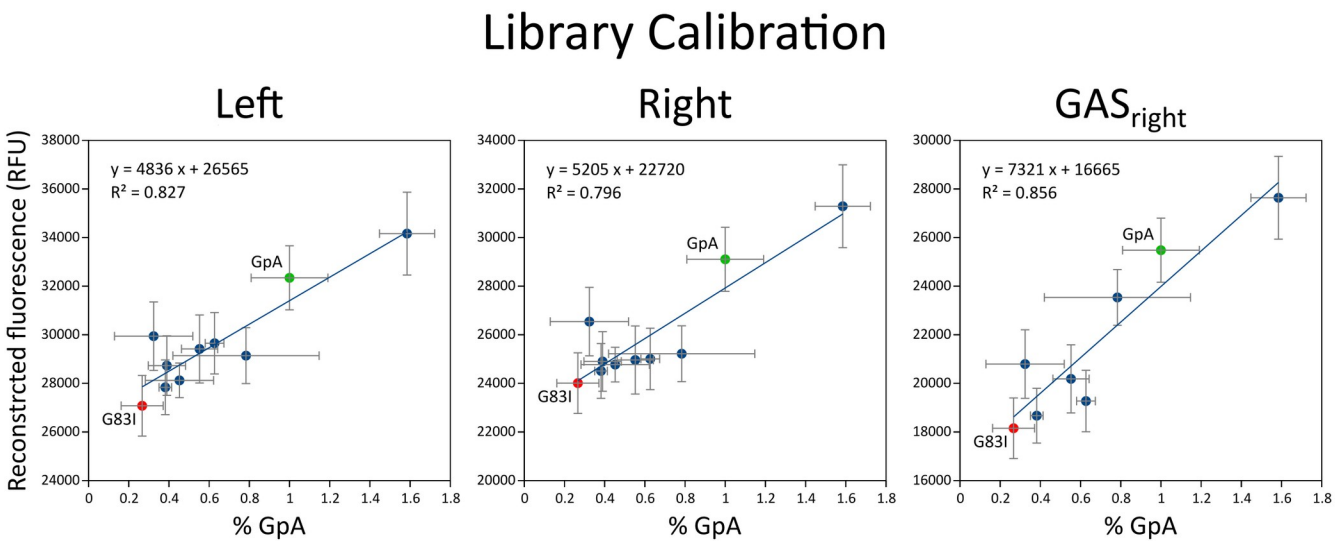


Fig. S7. Normalization of reconstructed fluorescence to standard. The TOXGREEN signal of a construct is typically reported normalized to the signal of the strong dimer glycophorin A (GpA, green dot). The signal is also compared to the to the monomeric variant of GpA, G83I (red dot). To avoid the dependency of the conversion on the variability of the single GpA construct, we calculated a calibration curve for each library using a set of 8-10 constructs of known TOXGREEN signal. Normalization was performed according to the resulting linear relationship. X-axis: normalized TOXGREEN signal measured individually in a fluorescence plate reader. Y-axis: reconstructed fluorescence of the same constructs in the libraries.