Title
Molecular Design Principles of Membrane Protein Structure

Permalink
https://escholarship.org/uc/item/3j67h0hf

Author
Mravic, Marco

Publication Date
2020

Peer reviewed|Thesis/dissertation

Molecular Design Principles of Membrane Protein Structure

by
Marco Mravic

DISSERTATION
Submitted in partial satisfaction of the requirements for degree of
DOCTOR OF PHILOSOPHY

in

Biophysics

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

Bill DeGrado
_____
Chair

Tanja Kortemme
_____

Michael Grabe
_____

_____

_____
Committee Members

Dedication and Acknowledgements

This thesis is dedicated to my abuelita Emma 'Tita' Figueroa

Thank you to my parents L.M. and B.K.M. for the gift of unwavering confidence tempered with humility throughout my life and schooling. Thank you to my crazy Puerto Rican family for showing me the value of strong female role models and supporting my journey to being the first Figueroa PhD. Thank you to J.T.B and R.J. for making sure I always make time and be present for adventures, love, and happiness – no matter what is looming.

Thanks to my advisor Bill DeGrado, an ever curious and caring man who knows how to throw a party. You and Susan always made San Francisco and the lab feel like a home and family. Some say graduate students are like stem cells and the thesis advisor's job is to differentiate them. It is a testament to your endless creativity and passion for science that I feel more multi-potent than when I started, yet armed with the theoretical and practical foundation to tackle anything.

I want to acknowledge the scientific mentors who have shaped my perspective and purpose across my career journey: relevant to this work, my thesis committee members Tanja Kortemme and Michael Grabe along with thoughtful biophysics instructors Robert Stroud and James Fraser.

Molecular Design Principles of Membrane Protein Structure

Marco Mravic

Abstract

The forces and energetic principles that guide molecular structure and dynamics in cellular lipid bilayers are poorly understood, largely due to limited familiarity with physical chemistry outside of water. We strive to enhance our fundamental chemical intuition concerning membrane proteins because their biochemical behavior within lipid membranes governs critical cellular processes and dictates health or disease. Likewise, this intuition paired with the ability manipulate membrane protein structure and function using computational modeling, theoretical calculations, and molecular design are instrumental to the engineering of new therapeutics. The work described here focuses on using protein design and integrative computational approaches to dissect the fundamental biophysical principles underlying protein interactions and folding within the lipid environment, spanning studies of both natural and totally synthetic proteins. Chapter 1 describes the design and experimental characterization of a family of simple synthetic model membrane proteins that rely entirely on the steric packing of lipid-embedded apolar sidechains for their stabilization and folding. This work reveals new fundamental principles governing van der Waals forces for proteins in lipid which should dictate protein structure, function, and evolution. Chapter 2 reports discovery of an evolutionarily conserved interaction motif within the transmembrane (TM) domains of integrin family proteins regulating their biochemical function. This discovery was driven by molecular modeling and all-atom dynamics simulations paired with co-workers' optical tweezer force spectroscopy experiments. Chapter 3 details the development and application of software to design lipid-soluble peptides that target the transmembrane domains of

proteins. We report several molecules that activate integrin $\alpha_5\beta_1$ in human endothelial cell culture, and one with binding specificity for the β_1 subunit. Chapter 4 outlines a second campaign with similar software to engineer small synthetic transmembrane proteins that bind and inhibit signal-amplifying erythropoietin receptor (EpoR), a canonical cytokine family receptor. Here, we demonstrate computational design can be used to target membrane-soluble molecules to specifically adopt custom, non-native binding modes to engage targeted regions of protein embedded in lipid. Thus, the efforts described represent leaps in our capacity to compute designer tool molecules entirely from scratch useful for manipulating membrane-spanning regions. These works together define progress in our fundamental biophysical understanding of protein interactions and folding in lipid, as well as our ability to dissect membrane protein structure and function with *in silico* molecular modeling algorithms and novel chemical biology tools.

Table of Contents

List of Figures

List of Tables

Chapter 4

Chapter 1:

Packing of apolar side-chains enables accurate design of highly stable membrane proteins[†]

Packing of apolar side-chains enables accurate design of highly stable membrane proteins

Abstract

The features stabilizing the structures of membrane proteins remain poorly understood. Polar interactions contribute modestly, and the hydrophobic effect contributes little to the energetics of apolar side-chain packing in membranes. Disruption of steric packing can destabilize the native folds of membrane proteins, but is packing alone sufficient to drive folding in lipid? If so, membrane proteins stabilized by this feature should be readily designed and structurally characterized – yet this has not been achieved. Through simulation of the natural protein phospholamban and re-design of variants, we define a steric packing code underlying its assembly. Synthetic membrane proteins designed utilizing this code, stabilized entirely by apolar side-chains, conform to the intended fold. Although exceptionally stable, the steric complementarity required for their folding is surprisingly stringent. Structural informatics show that the designed packing motif recurs across the proteome, emphasizing a prominent role for precise apolar packing in membrane protein folding, stabilization, and evolution.

Introduction

As membrane proteins (MPs) exit the translocon, they complete folding and assembly in the lipid milieu. However, the forces that stabilize proteins in membranes are less well understood for MPs than for water-soluble proteins. The hydrophobic effect provided by packing apolar side-chains in the protein interior represents the predominant driving force for protein folding in water, yet it is negligible in lipid membranes. In MPs, it is unclear whether analogous side-chain packing in the native state provides significant structural stabilization. On the one hand, those same apolar moieties pack similarly with lipid tails in the exposed unfolded state. On the other hand, structural informatics suggests that side-chains pack more efficiently in membrane proteins (*1-4*), and hence stabilize folding via favorable van der Waals (vdW) interactions and possibly also lipid-specific effects such as "solvophobic" exclusion (*5-9*). Mutations to MPs that strongly disrupt vdW packing in the protein interior, either by introducing voids or steric clashes, have been shown to destabilize their native state to various degrees (*8, 10-12*). However, it has proven difficult to determine whether apolar packing can play a dominant role in MP folding, or whether this feature is secondary to other more stabilizing interactions, namely hydrogen bonding (*13-17*), topology (*18*), and weakly polar interactions (*19, 20*) (Fig. 1.1). If tight sterically-compatible apolar packing is strongly stabilizing, then it should be possible to design and structurally characterize folded MPs stabilized by this feature alone. However, this has not been accomplished despite many attempts (*14, 17, 21-24*), and all successfully designed MPs have relied on hydrogen bonding (*17, 25*), metal-ligand interactions (*26, 27*), small residue motifs (*28, 29*), or templating of their folds with extra-membrane loops and water-soluble domains (*30*). Here, we use a multi-prong approach to design MPs that show that apolar side-chain packing plays a major role throughout MP folding.

To isolate the role of packing in MP stabilization, we focus on the self-association of transmembrane (TM) α-helices from single-spanning MPs, chosen for their pervasive biological importance and technical advantages. Over 50% of all membrane proteins are single-spanning, yet they are the least structurally characterized class of MPs. Lateral interactions between single-span TM helices play vital roles from signaling to ion conduction (*31, 32*), and their aberrant assembly is central to devastating diseases from cancer to Alzheimer's Disease (*33, 34*). Additionally, unlike complex multi-pass proteins whose folds are subject to constraining loops and extra-membrane domains, single-span TM bundles allow investigation of unconstrained inter-helical interactions with a clear unfolded state – a monomeric α-helix – where conformational specificity and thermodynamics can be simply evaluated by the oligomeric distribution.

We began by analyzing the homo-pentameric assembly of the small single-span MP phospholamban (PLN), a regulator of the sarcoplasmic reticulum Ca^{2+} pump. PLN's TM helix contains a 'LxxIxxx' sequence repeat (Fig. 1.1), which has been extensively studied in water-soluble coiled-coils (*35, 36*). Given the canonical labeling for α-helical 7-residue repeats, "*abcdefg*", '*a*' and '*d*' side-chains (Leu and Ile of LxxIxxx, respectively) project towards the bundle core whereas '*e*' and '*g*' residues face the inter-subunit interface (*37*). Recent water-soluble peptide studies demonstrated that it was possible to design 4 to 8-stranded bundles by manipulating the physicochemical properties of the residues at the '*e*' and '*g*' positions (*35*). Previous strategies to design MP assemblies by apolar interactions focused primarily on '*a*' and '*d*' positions. Thus, we explored the role of both core and interfacial packing interactions, focusing on both PLN and rationally designed peptides. We show that when apolar packing is used as the sole stabilizing principle, folding requires side-chain geometric complementarity that is surprisingly stringent. Nevertheless, the optimization of steric packing alone can determine the thermodynamics and the

4

3-D architecture of these highly-stable TM protein bundles. These synthetic MPs provide comprehensible and robust model systems for future investigations of apolar packing in MP folding, including the relative contributions of dispersion forces and entropy.

Results

MD simulations of PLN identify a rigid pentameric domain stabilized by vdW packing.

The membrane-spanning segment (residues 24-52) of PLN forms a homo-pentameric helical bundle. The N-terminal third houses several strongly polar residues, while the C-terminal two thirds is highly hydrophobic – containing 2.5 copies of the critical LxxIxxx motif (Fig. 1.2)(*38*). The interactions stabilizing the pentameric structure remain unclear, particularly the role of the polar sector. Two early nuclear magnetic resonance (NMR) structures of the PLN helical bundle derived from relatively sparse experimental restraints differ substantially (backbone RMSD = 4.8 Å, residues 24-52) – most significantly at the polar TM sector (*39, 40*). To obtain a refined model of PLN's TM bundle we performed atomistic molecular dynamics (MD) simulations of the full-length protein in a fluid bilayer (1-palmitoyl-*2*-oleoyl-sn-glycero-3-phosphocholine, POPC), beginning with the more recent NMR model (PDB, 2KYV).

The initial model rapidly rearranged to a stable conformational ensemble (Figs. 1.2, 1.3) similar to a previous 15 ns simulation using the alternative NMR structure (PDB, 1ZLL) (*41*). The apolar C-terminal LxxIxxx-domain was tightly packed and remarkably rigid (backbone RMSF to medoid = 0.53 Å, residues 33-51) indicating it plays an important role in stabilizing the structure.

By contrast, the polar N-terminal TM sector was more dynamic (1.53 Å) and became widely splayed (Fig. 1.3) – thus, likely less critical for stabilization. Water rushed into this end of the bundle, accompanied by rapidly fluctuating hydration and inter-helical side-chain hydrogen

bonds (Fig. 1.3). These findings are consistent with previous mutational studies of PLN (*38, 42, 43*) and with other MPs in which polar TM residues were found to modulate functional dynamics rather than stability (*44*). In PLN, tuning of stability and dynamics is likely important for functional interactions with cyclic-AMP-dependent protein kinase and the Ca^{2+} ATPase.

Design and structural stabilization of apolar variants of the PLN pentamer

We next asked if the well-packed apolar region of PLN – when free of strongly polar residues – can drive pentameric assembly of a PLN-like peptide, designated PL5. PL5 was designed using an idealized model built *de novo* by mathematical parameterization of PLN's rigid C-terminal domain from simulation (Table 1.1). For the C-terminal half of PL5, PLN's native sequence was retained, changing only lipid-facing β-branched residues to Leu or Phe to facilitate synthesis. By contrast, the N-terminal TM portion of PLN was redesigned. Asparagines in the splayed polar sector of the bundle were converted to less polar residues more sterically compatible with the idealized bundle (Fig. 1.4). The remaining lipid-facing and terminal residues were chosen to facilitate synthesis, purification, and positioning within the bilayer. To determine whether the weak polarity of the interfacial Ser9, Cys11, and Cys16 of PL5 contributed to assembly, they were converted to the isosteric and fully hydrophobic S-C_α-ethyl-Gly (EtGly) in $PL5_{EtG}$ and $PL5_{EtG3}$ (Fig. 1.4) (*45*).

PL5 was shown to form stable pentamers by MD and solution measurements (Fig. 1.4). In all-atom MD, PL5's initial conformation was stable over 1.0 μs (0.71 Å mean RMSD versus medoid). Analytical ultracentrifugation of PL5 in myristyl dimethyl-3-ammonio-1-propanesulfonate micelles gave an apparent molecular weight of 18 kDa (Fig. 1.4), within the experimental error of the pentamer (19 kDa ± 1.5 kDa, based on propagation of a 3% error for the

computed partial specific volume). PL5, PL5$_{EtG}$, and PL5$_{EtG3}$ migrate as a single oligomeric state by SDS-PAGE, confirmed to be pentameric (Fig. 1.5). As well, the PL5 pentamer is highly stable even after heating (95° C) in lithium dodecyl sulfate (LDS) and 8 M urea (Fig. 1.4).

The 3.17 Å-resolution crystallographic structure of micelle-solubilized PL5 was in outstanding agreement with the guiding MD model (1.1 Å backbone RMSD) (Figs. 1.4, 1.6, 1.7; Table 1.2). This structure can be mathematically described to within 0.6 Å as either a parallel left-handed coiled-coil or a bundle of straight, tilted α-helices (Table 1.1). The bundle appears to maximize inter-helical packing at the expense of packing the central core, resulting in a small pore (Fig. 1.8). Throughout most of the bundle, the pore is too small to accommodate even a single water molecule and was dry during MD. However, diffuse density observed near the ends of the bundle might represent portions of a detergent tail (Fig. 1.4). The radius of the bundle expands by up to 1 Å as the steric bulk of the interfacial '*e*' and '*g*' residues progressively increases: S9, C11, C16, L18, I23, M25 (Fig. 1.9). As expected from previous experiments with PLN (*46*) and our simulations (Fig. 1.10), the side-chains of Ser9, Cys11, and Cys16 donate intra-helical hydrogen bonds to preceding carbonyls (residue *i-4*), rather than forming inter-helical hydrogen bonds.

A steric code for specific assembly and design of TM 5-helix bundles.

Inspection of PL5's X-ray structure revealed knobs-into-holes packing (Fig. 1.8), as in multi-stranded coiled-coils (*36, 37*). Side-chains at '*d*' and '*e*' positions of opposing helices interact intimately along a '*d/e*' interface; side-chains at '*a*' and '*g*' form the '*a/g*' interface (Fig. 1.8). At the '*a/g*' interface, the C$_\alpha$-C$_\beta$ bond vector of each side-chain is directed inward towards its helical neighbor, while the C$_\beta$-Cγ bond is directed outward to avoid a steric clash. The close approach of opposing C$_\beta$ atoms argues against placement of β-branched amino acids at '*g*'

positions in the context of PL5's sequence (Leu, '*a*'). The reverse is seen at the '*d/e*' interface; the C_α-C_β bond vector is directed outward and the C_β-C_γ bond points inward. Here, β-branched amino acids are sterically poised for tight packing. Although not β-branched, Ser and Cys also pack well at '*e*', because they can position a heavy atom (O, S) in place of C_γ when in a gauche$^+$ rotamer (i.e. -60°) that is further stabilized by intra-helical side-chain-main-chain hydrogen bonding. Interestingly, in contrast with the aforementioned packing code, water-soluble LxxIxxx-repeating peptides containing non-β-branched Glu at '*e*' and β-branched Ile at '*g*' specifically assembled into pentameric bundles (*35*), perhaps reflecting differences in forces between water-soluble versus membrane proteins.

To test our steric model, we first prepared a single-site variant of PL5, in which the '*g*' position Leu18 was replaced by its β-branched isomer Ile, resulting in a steric clash between this Ile18 C_γ methyl and Leu19 ('a') C_β methylene in models. In water-soluble coiled-coils, similar conservative single-site mutations at non-core positions are generally tolerated or, at worst, alter the stoichiometry of association (*35*). However, this replacement entirely destabilized the PL5 pentamer (Fig. 1.11, Table 1.3). Thus, the requirement for steric complementarity is so stringent that misplacement of a single methyl group per helix eliminates detectable pentamer formation.

Next, we designed peptides to test the proposed rules for steric packing at the '*e*' position in pentamer assembly (Fig. 1.12). The peptide sequences repeated a minimalist heptad $L_a xxI_d X_e xxL_g$, systematically varying the '*e*' position ('X_e') to either one of the β-branched amino acids Ile, Val, and Thr or Cys (designated '*e*'-Ile, '*e*'-Val, '*e*'-Thr, '*e*'-Cys). As stringent controls, we also prepared variants with Ala or Leu at '*e*', whose side-chains would not pack as favorably, but not necessarily clash.

SDS-PAGE showed that the designed peptides '*e*'-Ile, '*e*'-Val, '*e*'-Thr and '*e*'-Cys formed pentamers, while the negative controls ('*e*'-Ala, '*e*'-Leu) were entirely monomeric in both mild and harsh detergents, octyl glucoside (OG) and LDS respectively (Fig. 1.12). Thus, small modifications in the amino acid structure magnify to large changes in stability. Importantly, only very low intensity bands were observed between the monomer and pentamer, showing that the assembly is highly cooperative and specific.

The thermal stability of the assemblies was assessed after 95° C incubation and again after returning to 25° C (24 hours). In OG, the pentamers are essentially unperturbed by heating. In LDS micelles, heating partially dissociates the pentamers, with '*e*'-Thr being the most stable than '*e*'-Val and '*e*'-Cys. After cooling, the original distribution was largely recovered, demonstrating that assembly is enthalpically favorable and slowly reversible. Densitometric tracing (Tables 1.4, 1.5) allowed extraction of apparent equilibrium constants and associated energetics. The apparent free energy differences between '*e*'-Ile and '*e*'-Val relative to '*e*'-Thr are 3.4 and 2.8 kcal per mol pentamer in LDS (25° C), respectively. Furthermore, the apparent free energy difference between '*e*'-Ile and '*e*'-Leu isomers is at least 9.8 kcal per mol pentamer (OG, 25° C), given that '*e*'-Leu formed less than 0.5% pentamer under conditions where '*e*'-Ile was predominantly pentameric.

To confirm the structural basis for association, we determined X-ray structures of a slightly shortened variant of '*e*'-Val in three crystal forms between 1.90 and 2.50 Å resolution (Fig. 1.12, 1.13; Tables 1.2, 1.6). The pentameric structure, detailed side-chain packing, and rotamer distribution of the core residues were exactly as designed (0.63 Å backbone RMSD) (Fig. 1.12; Table 1.7). Each residue had clear density and all inter-helical contacts were between apolar sidechains. Thus, accurate design of stable TM domain architectures mediated entirely by apolar side-chains is indeed feasible.

Packing motifs in designed assemblies are observed across the membrane proteome.

The constitutive helix-helix packing interactions in coiled-coils such as PL5 and '*e*'-Val are often structurally analogous to those found at diverse tertiary contexts in numerous globular protein folds. We therefore asked whether the inter-subunit helix-helix geometries seen in '*e*'-Val and PL5 occur frequently within the structures of natural MPs. Successive pairs of 9-residue stretches (18 total, Fig. 1.14) from '*e*'-Val and PL5 were compared to the 3D structures of helical pairs from a non-redundant database of experimental MP structures at the main-chain level (Table 1.8)(*47*). For each stretch of '*e*'-Val, a large number of matches (RMSD < 0.85 Å) were found (50 to 91, mean = 79), demonstrating that structurally similar helix-helix geometries occur frequently in many MP architectures (Fig. 1.14). This geometry encompasses the most common parallel left-handed inter-helical packing motifs used by MPs (*3, 48*). Variations of this geometry were much less frequent (Fig. 1.15).

The degree of enrichment of specific amino acid types (Fig. 1.14, Table 1.9) in the aligned sequences was also in agreement with the packing analysis (Fig. 1.8). A perfect correspondence was not expected, since the sequence and structural contexts vary in different natural proteins. Nevertheless, β-branched amino acids were highly enriched at the '*d/e*' layer, with a clear preference for Ile at '*d*', precisely matching Ile6, Ile13, Ile20 in the LxxIxxx motif. Furthermore, the observed preference for Thr at '*e*' – rather than Val – matches the thermal stability seen for '*e*'-Thr, and is likely explained by the additional stabilization of its gauche^{+} rotamer by Thr's intra-helical main-chain hydrogen bonding (*49*). In the final heptad where the bundle radius has widened (Val21), the larger Met is instead enriched. Similar results were obtained probing PL5, displayed in Fig. 1.16.

Using '*e*'-Val as the search template, the amino acid enrichments observed at the '*a/g*' interface are also consistent with the steric pattern (Fig. 1.8). No β-branched amino acids were found enriched at the '*a/g*' interface, although they are also not strictly forbidden (Table 1.9). Cys was strongly enriched at '*g*' (p-value = 0.0001) for '*e*'-Val, matching evolutionary conservation seen in PLN (Cys36, '*g*'). For '*g*' at the equivalent PL5 helix-helix interface, shown in Figure 1.15, Cys, Cys, and Leu were highly enriched (p-values = 0.0001, 0.013, 0.004) – corresponding to Cys11, Leu18, and Met25. The enriched amino acids also progressively increase in size as the bundle widens (Fig. 1.14). At '*a*,' the leucine side-chains project towards the bundle center where their terminal atoms appear to be important for stabilizing the pentamer's core, rather than only maximizing pairwise inter-helical interactions. In our analysis, a number of non-β-branched residues are accommodated at '*a*' as expected, including the enriched Ser. Although slightly polar residues are commonly enriched in this and other positions, our battery of data argues that their steric properties rather than inter-helical hydrogen bonding capability are responsible. Moreover, in those natural examples containing Ser, Thr, or Cys, such hydrogen bonding was rare (7.3% of cases).

In summary, position-dependent propensities within the database of natural MPs are in concordance with our steric model (Fig. 1.8). Using this approach, we separate the trends general to the constitutive pairwise helix-helix packing motif – recurring in diverse contexts of natural MPs – apart from features distinct to pentameric bundles and LxxIxxx repeat, such as core-packing.

Finally, we asked how strict packing restraints are in MP versus water-soluble proteins by instead searching a non-redundant database of water-soluble proteins. Interestingly, bulky hydrophobic Leu and Ile residues are enriched at essentially all interfacial positions (Fig 1.17).

11

The distinct amino acid distributions seen in the two classes of protein likely reflect differences in the stabilizing forces at play. A strong hydrophobic driving force dominates folding in water, so natural proteins need not achieve stringent packing to fold; although, further optimization of packing is advantageous and can produce stabilities not seen in natural proteins (*50, 51*). Without a hydrophobic force in bilayers, it appears geometric complementarity must be more strictly optimized to achieve folding in MPs.

Conclusions

This work places on firm experimental ground what had been long hypothesized: apolar side-chain packing can provide a sufficient driving force to determine a membrane protein's fold. The design of highly stable TM assemblies can be encoded solely by steric side-chain complementarity. Furthermore, the inter-helical packing motifs of these pentamers are frequently found in diverse membrane protein families including transporters, ion channels, enzymes, and G-protein coupled receptors. Previous studies suggest that only a small number of inter-helical motifs, such as this one, constitute most tertiary building blocks used within MPs of distinct architectures (*3, 52*). The helix-helix interfaces of these commonly-occurring motifs are also rich in well-packed apolar side-chains (*3*). Thus, analogous steric principles likely encode these different motif geometries, and are used to direct a wide diversity of folds and stoichiometries throughout nature.

The sequence-specific stringency observed indicates cooperativity of the distributed vdW interactions across a large interface is crucial, particularly in the absence of alternative stabilizing interactions such as hydrogen-bonds. Local defects propagate over a large area, disrupting the precise geometric complementarity required for TM helix association. Thus, small changes in

protein-protein interactions can impart significant specificity. Notably, others also recently observed single Leu-to-Ile substitutions abolishing select TM helix-helix interactions (*53*). Furthermore, in the very common GxxxG motif, steric properties of apolar side-chains distant from the motif sequence strongly modulate stability and specificity, consistent with our findings (*7, 11*).

Our studies together with other recent work (*26, 30, 54*) places demonstrate that the principles for MP design has progressed significantly, and the engineering of more complex structures and functions within lipid bilayers is on the horizon.

Materials and Methods

Peptide synthesis and purification.

All peptides were synthesized by microwave-assisted solid-state synthesis using Fmoc chemistry, cleaved from the resin, and purified by reverse phase HPLC as previously described, resulting in a C-terminal carboxamide with a free amine N-terminus (*26*). The peptide was purified by HPLC on a C4 prep column (Vydac) using a linear gradient of solvents A and B at 15 mL/min (solvent B: 60/30/9.9/0.1 isoproanol/acetonitrile/water/trifluoroacetic acid (TFA); solvent A: water, 0.1% TFA). All peptides were confirmed to be >95% purity by analytical HPLC using a C4 column and to have the correct mass by MALDI mass spectrometry using α-cyano-4-hydroxycinnamic acid (Sigma) as the matrix.

SDS-PAGE analysis and thermodynamic analysis.

Peptide stocks of known concentration were made from dissolving in micelle solutions from lyophilized power as described previously (*14*). Samples were prepared in 2% sodium dodecyl sulfate (SDS) in 8M urea 2% lithium dodecyl sulfate (LDS), or 40 mM octyl-glucoside (OG) and

loaded in equal concentrations at 2 µg. The running buffer was 0.1% SDS in pH 7.3 2-ethanesulfonic acid (MES) buffer (ThermoFisher). Gels used were 12% NuPage Tris-Glycine (ThermoFisher) run at 200 V for 35 minutes under reducing conditions with greater than 10 mol equivalents tris(2-carboxyethyl)phosphine TCEP (sigma) with PageRuler protein standards (ThermoFisher), and then stained and visualized with SafeStain coomassie blue (ThermoFisher). Samples were heated on a thermal block for 30 minutes, followed by addition of additional reducing agent after room temperature was reached.

The difference in the free energy of pentamerization for a given pair of peptides can be obtained by measuring the concentration of the monomer and pentamer of each peptide according to equation 1:

$\Delta\Delta G^{app}$ = -RT ($\ln(K_{diss1})$ - $\ln(K_{diss2})$)   eqn. 1a

$\Delta\Delta G^{app}$ = -RT $\ln$ ($([mon_1]/[det])^5/[pent_1]/[det])$ - $\ln([mon_2]^5/[det])$ /$[pent_2]/[det])$ eqn. 1b

$\Delta\Delta G^{app}$ = -RT $\ln([mon_1]^5/[pent_1])$ - $\ln([mon_2]^5/[pent_2])$ eqn. 1c

In which, $\Delta\Delta G^{app}$ is the apparent free energy difference, K_{diss} is the dissociation constant, [mon], [pent] and [det] the molar concentrations of the monomer, pentamer, and detergent, respectively, and the subscripts reflect the identity of the two peptides.

The relevant concentration depends the choice of a standard state, which is either the molar concentration of the peptide, the peptide/detergent molar ratio (chosen in equation 1b), or the peptide/micelle molar ratio. Thus, different apparent free energies will be obtained for each

peptide depending on the choice of the standard state. However, if experiments for each peptide are carried out under identical (i.e., same peptide and detergent concentrations) and used to compare the free energy differences between two peptides the contributions from the choice of a standard cancels out (as in eqn. 1c), allowing a rigorous comparison of the free energies between the peptides. We refer to the obtained values of $\Delta\Delta G^{app}$ because the equilibria are very slow for the most stable peptides in OG. On the other hand, equilibrium appears to be reached in the harsher detergent, LDS, after overnight incubation (Fig. 1.12). We also do not report values below for the '*e*'-Cys, because of possible complications from irreversible Cys oxidation. The peptides were pre-incubated at 1 mg/mL (approximately 0.26 mM for all the peptides, all between 3.70-3.84 Da), giving a detergent/peptide mol ratio of 154 for 40 mM OG and 282 for 2% LDS.

To obtain a banding pattern in gel electrophoresis that reveals all oligomeric states, as observed for native full-length PLN, followed the protocol described in ref. (*58*). Detergent solubilized samples were heated to 95° C for 30 minutes, then allowed to cool for 1 hour without adding additional reducing agent. This assumes the heating decomposes most of the reducing agent in the solubilization buffer followed by formation of covalent species between numerous cysteine amino acids in the TM domain. Six total bands, likely monomer through hexamer, are observed at varying intensities after SDS-PAGE of the designed TM pentamers (Fig. 1.5) described in Figure 1.12, which typically migrate very similarly to PL5 under reducing conditions.

Circular Dichroism (CD) Spectroscopy.

CD spectra were recorded on a Jasco J810 CD spectrometer using 1 mm cuvettes with the following parameters: 1 nm data pitch, 20 nm/min scanning speed, 4 s response, 2 nm bandwidth,

and 3 accumulation*s*. Peptide samples of known concentration were made in 10 mM Tris HCl buffer pH 7.4, 2 mM TCEP, and 13.8 mM or 188 mM DPC.

Analytical ultracentrifugation sedimentation equilibrium and curve fitting

Samples were prepared by co-dissolving peptide and doceylphosphocholine in ethanol, evaporating the ethanol under a stream of N_2 then under vacuum for 1 hour, then re-hydrating the film with a 13% D_2O solution containing 50 mM sodium phosphate buffer, 200 mM potassium chloride, and 2.5 mM TCEP for a final concentration of 30 mM 3-(*N,N*-Dimethyltetradecylammonio)propanesulfonate (Sigma). All measurement were performed using a Optima XL-I Beckman Coulter ultracentrifuge. We first ensure correct density matching as described by Tanford and Reynolds (*60*). Theoretically, the protein-detergent particle will have an observed buoyant molecular weigh, M_b, related to the protein molecular weight, M_p, by:

$$M_b = M_p \left[1 - \bar{v}_p \rho + \delta_d (1 - \bar{v}_d \rho) \right]$$ Equation 2

where ρ is the solvent density, δ_d is the protein to detergent weight fraction, and $\bar{v}_p$ and $\bar{v}_d$ are the partial specific volume of the protein and detergent respectively. Under the solvent density condition $\rho = 1/\bar{v}_d$, the bound detergent's contribution to sedimentation is negligible. To ensure this condition was met, the sample cell was filled with a buffer containing myristyl-sulfobetaine, with a detergent-free reference buffer. Spins were conducted over 10 hours at 48 kRPM, taking radial interference measurements every hour to confirm equilibrium had been reached. The slope of the radial fringe displacement profile was confirmed to be negligible for all detergent concentrations tested (5-33 mM), indicating correct density matching and uniform distribution of micelles in the sample cell.

Then, sedimentation equilibrium experiments for PL5 in C-14 betaine solutions were performed at two concentrations and 3 rotor speeds. Peptide concentration was measured by UV absorbance at 280 nm before loading into the sample cell. The concentrations of PL5 used for AUC were 58 and 79 μM, giving a detergent to peptide molar ratio of 570 and, 420 respectively. Spin equilibration was conducted at each rotor speed 30,000, 35,000, and 40,000 rpm at 25° C for 12 hours before radial UV scans were taken at 280 nm.

The sedimentation equilibrium profiles were globally fit to a single species model using a non-linear least squares curve fitting procedure in Igor, as previously described (*14*). The protein partial specific volume was estimated as 0.78, calculated as the mass weighted average of amino-acid specific partial specific volumes of the Cohn-Edsall scale (*61*). The solvent density was estimated as 1.028 g/mL, calculated from the buffer components using SEDNTERP. The estimations of these constants are assumed to be correct with 3% accuracy, propagating to approximately 5-8% error in the experimentally determined molecular weight measurement.

Crystallization

Crystallization trials were performed by robotic-assisted hanging drop vapor diffusion at room temperature. MemGold, MemGold II, MemSys, MemStart (Molecular Dimensions) and PEG/Ion I & II (Hampton) sparse matrices were screen using 5 mg/mL solutions of PL5 in 30 mM C_8E_5 (Anatrace) at a drop to mother liquor ratio of 1:1 with a final drop volume of 200 nL. Crystals of 30-60 μm size and many different morphologies were observed overnight for ~10% of drops in the sparse matrix screen. No crystals were observed with trials set up in detergents OG or n-Dodecyl-N,N-Dimethylamine-N-Oxide (LDAO). Diffraction data used for structure determination

of the PL5 structure, PDB entry 6MQU, came from a single crystal (ca. 60 μm) grown for 3 weeks with a 1:1 volume ratio of protein solution and a well solution of 0.04 M magnesium chloride hexahydrate, 0.05 M NaCl, 0.1 M (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid), 32% (v/v) PEG 400, pH 7.5. Crystals were flash frozen without additional cryoprotectant and stored in liquid nitrogen.

For mini 'e'-V, sparse matrix screening was conducted similarly. Crystals were also observed in a large variety of conditions at similar concentrations. Data sets were collected and solved for crystals grown in 3 conditions, noted by crystal form in Table 1.2 and 1.6. Crystal form 1 (pdb, 6MCT), 5 mg/mL protein in 45 mM C_8E_4 added in a 1:1 volume ratio to a well solution of 2.5 M ammonium sulfate and 50 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) grown at 4° C – cyroprotected by briefly transferring the looped crystal to a solution of 30:60 2-Methyl-2,4-pentanediol (MPD) / mother liquor. Crystal form 2 (pdb, 6MQ2), 5 mg/mL protein in 45 mM C_8E_4 added in a 1:1 volume ratio to a well solution of 2.5 M ammonium sulfate, 0.1 M Na Acetate pH 4.6, 28 %(v/v) PEG MME 550 grown at 20° C; no additional cryo-protection. Crystal form 3 (pdb, 6MPW), 5 mg/mL protein in 45 mM C_8E_4 added in a 1:1 volume ratio to a well solution of 50 mM $MgCl_2$, 0.01 %(w/v) NaN_3, 0.1 M Na Cacodylate pH 6.7, 6.6 (%w/v) PEG 3350; cyro-protection by briefly transferring the looped crystal to a solution of 30/60 2-Methyl-2,4-pentanediol (MPD) / mother liquor

X-ray diffraction data collection and analysis of PL5

Crystallographic data were collected at the Advanced Light Source on beam 8.3.1. The beam size was 100 μm. The detector used was a Dectris Pilatus 3 6M. For PL5, data were collected at <100

K using a beam energy of 11.0 keV and detector-to-sample distance of 500 mm. The crystal was exposed to the beam for 1 second per frame, oscillated 2° during data collection. Diffraction images were processed using MOSFLM (*62*). Data were indexed then scaled and merged in CCP4 using Pointless (*63*) and Aimless (*64*), respectively. All 120 frames were used for a dataset indexed to the I 4_1 space group, with a resolution limit of 3.17 Å. As well, the data was indexed to the I 4_1 2 2 space group using the first 60 frames with a resolution limit of 3.07 Å with similar data collection statistics as shown in Table 1.2 (i.e. completeness, R_{merge}, CC1/2).

PHASER in the Phenix suite was used for molecular replacement, initially with the I 4_1 2 2 data set (*65*). The Matthews coefficient suggested 5 and 10 monomers per asymmetric unit for I 4_1 2 2 and I 4_1, respectively. An input search model of a *de novo* 29 poly-alanine C5-symmetric left-handed helical coiled-coil built using parametric equations: super-helical radius: 8.5 Å, pitch angle: -11°/residue. Side-chains and additional residue positions were built manually in Coot (*66*) and refinement was conducted using PHENIX Refine (*67*). An anti-parallel 5-helix bundle was also used for input phases, but gave poor R_{free} upon addition of side-chains and initial refinement, suggesting a parallel orientation of PL5.

The data set indexed to I 4_1 2 2 was refined to 0.25/0.28 R_{work}/R_{free}. However, the electron density map was missing strong density for many aromatic side-chains and had a larger than average R_{free} value given the resolution range. Thus, we repeated molecular replacement using the I 4_1 dataset searching two copies of the refined pentameric model derived from the I 4_1 2 2 dataset, including side-chains. A molecular replacement solution was found with two pentamers per asymmetric unit, and was refined to 0.2275/0.2420 with acceptable bond geometries. Upon symmetry expansion,

the overall crystal lattice was shown to be generally equivalent, suggesting the two pentamers in the asymmetric unit modeled from indexing to I 4_1 are related by pseudo-symmetry rather than true crystallographic symmetry.

Terminal residues of the helices were removed if their addition to the model decreased R_{free} or the residue could not be fit into the density with good geometries. Data collection and refinement statistics for the final model, the dataset indexed to I 4_1, are shown in Table 1.2. For the 249 residues in the model, 0 residues were modeled with dihedral angles that were Ramachandran outliers and 95.98% (n=239) had dihedral angles in the favored values range. Zero rotamer outliers or C-β deviations were modeled. The model's clash score is 9.48 and the MolProbity score is 1.78. The atomic coordinates were deposited to the protein data bank under the accession code 6MQU.

X-ray diffraction data collection and analysis of mini 'e'-Val

Crystallographic data were also collected at the Advanced Light Source on beam 8.3.1. The beam size was 100 μm. The detector used was a Dectris Pilatus3 6M. Data was collected at a temperature of <100 K using a beam energy of 11.1 keV.

For mini '*e*'-Val crystal form 1 (pdb, 6MCT), data was collected with a detector-to-sample distance of 250 mm, beam exposure of 0.6 second per frame, and oscillation 0.3° per frame for 1200 frames. Data was processed with XDS (*68*) using P 2_1 2_1 2 as the space group. Data statistics for data collection and refinement are shown in Table 1.2

The data showed apparent 5-fold NCS symmetry, and the designed pentameric helical bundle was used as search model for molecular replacement. Phaser was used for the molecular replacement. The native Patterson map showed a strong off-origin peak at (0.500, 0.1139, 0.000), suggestive of translational NCS (tNCS). Phaser was used for the molecular replacement. Two tNCS-related pentameric helical bundles were located first, and another bundle was added based on the map phased with this partial model. In total there were three helical bundles present in the asymmetric unit.

Structure refinement was done with Phenix, and manual model building was performed with Coot. Rigid body refinement was done using domains consisting of a single helix. Simulated annealing was used to remove bias from models. TLS refinement were performed with TLS domains as a single helix too. For crystal form 1, due to relatively small solvent content (calculated to be 38.2%), the solvent mask was fine-tuned with Phenix during refinement. Solvent molecules as well as detergent molecules were incorporated in late stage of refinement.

For mini 'e'-Val crystal form 2 (pdb, 6MQ2), data was collected with a detector-to-sample distance of 450 mm, beam exposure of 1 second, and oscillation 1° for 360 frames. Data was processed with XDS using C 2 2 2_1 as the space group. For mini 'e'-Val crystal form 3 (pdb, 6MPW), data was collected with a detector-to-sample distance of 350 mm, beam exposure of 0.7 second, and oscillation 0.5° for 720 frames. Data was processed with XDS using C 2 2 2_1 as the space group, differing by crystal form 2 by a 6.9 Å increase in the unit cell size in the b dimension. Data statistics for data collection and refinement are shown in Table 1.6, obtained using the same methods described.

Molecular Dynamics Simulation, Modeling, and Analysis

All simulations were performed using Gromacs 5.1.4. Atomistic simulation was performed using the CHARMM36 force field (*69*) A 2 fs time step was used throughout, and in conjunction with the LINCS constraint algorithm during the equilibration stage. The Parrinello-Rahman barostat was used with semi-isotropic coupling. The Nose-Hoover thermostat was used, with protein, lipid, and solvent independently coupled to a 300 K bath. Electrostatics were treated with Particle Mesh Ewald, and the cutoff for both Coulomb and van der Waals interactions (Lennard Jones potential) was 1.2 nm. A 4 ns equilibration simulation was performed with 1000 kJ/mol/nm^3 restraints on all protein heavy atoms. Subsequently, the restraints were removed for the production simulation.

The phospholamban structure was taken from the PDB (2KYV). The protein was placed in a POPC bilayer containing 493 lipids using the membrane builder tool in CHARMM-GUI (*70*) The system was then equilibrated for 4 ns as described above, and production MD was run for 380 ns.

For analysis of PLN simulation, a structural clustering of simulation frames was performed by hierarchical clustering with backbone atom RMSD (residues 32-52) using complete linkage (1.5 Å cutoff). To determine the medoid frame of the ensemble, a k-medoids clustering with one cluster was performed on the largest cluster. RMSF was calculated for backbone atoms of residues 24-52 by superimposing the coiled-coil region (residues 32-52)

To analyze hydration and inter-helical side-chain hydrogen-bonding between Asn and Gln residues in the PLN simulation (Fig. 1.3), simulation frames ever 200 ps were extracted. A hydrogen-bond

was considered if the distance between the acceptor and donor heavy atoms (O, N) was less than 3.3 Å and the angle between the acceptor oxygen, the donor proton, and the donor heavy atom is between 120º and 180°. For time-dependent analysis of the Asn34—Gln29 hydrogen-bonds, inter-protomer pairs were analyzed independently and given a binary classification as having a hydrogen-bond or not. A Pearson autocorrelation function was calculated by comparing the binary time series data to itself at different time delays (0.2 ns intervals for the first 70 ns), and averaging the time-dependent auto-correlation value calculated by repeating the analysis at all possible starting frames for the first 70 ns. Seventy nanoseconds was chosen by qualitatively observing the autocorrelation function decay to baseline. Each autocorrelation function for the inter-protomer hydrogen-bond data was fit to a simple biexponential decay function to calculate the components of the relaxation time for this interaction. For all 5 Asn34—Gln29 pairs in the PLN simulation, the smaller relaxation time was the lowest possible value we allowed in the fit, 0.2 ns, the interval at which frames were sampled. The other relaxation time component was between 2 and 12 ns. The time series data for each inter-protomer pair, frequency for bond breaking and bond forming events, and the autocorrelation function fits, are shown in Figure 1.3.

For the simulation of PL5, a protein model was build de novo based on a generalized set of parametric equations to describe coiled-coils (*56*) modeling residues 1-29. The parameters used to build the PL5 model were derived from the best fit to the apolar region of PLN from our MD simulation, shown in Table 1.1. This model was inserted in a membrane via a coarse-grained (CG) Martini self-assembly simulation as follows. The protein was converted to a CG representation using version 2.6 of the Martinize tool (*71*). The CG model was placed in a 7.5 nm cubic box with 1000 kJ/mol/nm^3 restraints placed on all particles. 120 POPC lipids were randomly placed, and

the system was solvated with SPC water and 100 mM sodium chloride. A 30 ns self-assembly simulation was performed with a 30 fs time step, using a velocity rescale thermostat, and the Berendsen barostat. Reaction field electrostatics were used, with a cutoff at 1.1 nm. The Lennard Jones potential was similarly cutoff at 1.1 nm. The assembled bilayer was then converted to an atomistic representation using the Backward script (*72*). After a constrained equilibration, a 1000 ns production simulation was performed, both as described above.

For mini '*e*'-Val, the input model was derived *de novo* from a parametrically generated a C5-symmetric left-handed helical coiled-coil poly-alanine model: super-helical radius: 8.7 Å, pitch angle: -16°/residue, α-helical phase: -160°. Rotamers in the intended gauche+ conformation were build manually. Simulations were conducted in an 80 x 80 x 9.5 nm box with 189 POPC lipids with a 30 ns equilibration period with restraints placed on the backbone atoms of the initial model, 1000 kJ/mol/nm^3 followed by a unrestrained 200 ns production run.

For each MD simulation, NMR structure, and x-ray structure, the fit to parametric coiled-coil equations were calculated (Table 1.1)(*56*). The channel width was calculated using HOLES2 (*73*). One pentamer of the PL5 X-ray structure was also fit to a C5-symmetric helical bundle of straight α-helices by first performing the superposition of an ideal α-helix onto residues 5-26 of PL5's chain A, then applying a C5-symmetry operation to the ideal α-helix. Then, the backbone RMSD between the 5 symmetrically positioned α-helices and PL5 was calculated to be 0.54 Å.

Membrane protein database construction

The database was constructed similarly to Zhang et al. with some modifications (*3*). First, all entries in the Orientations of Membrane Proteins database (*57*) were queried for structures from X-ray, neutron, or electron scattering experiments with a resolution less than 3.2 Å, or from electron microscopy with a resolution less than 3.0 Å. Then, the transmembrane regions were extracted by the residue ranges given for TM segments listed by OPM. Redundancy in the database was used by performing a comparison of total sequence identity between each pair of eligible PDB entries, considering only TM segments. Total sequence identity was calculated by performing each pairwise comparison between each TM segment from each of the compared PDB entries. Global sequence alignment between two TM segments was performed using the BLOSOM62 substitution matrix with gap penalties of -10 for opening and -2 for extension. The sequence identity for each TM segment was taken as the maximum sequence identity found after comparing it to each the opposing entry's TM segments. The total sequence identity between two PDB entries was calculated by averaging the sequence identities for each TM segment, weighted by each TM segments' residue length. PDB entries were then culled at 40% sequence identity; if two entries were >40% identical, the larger of the two entries were kept. This cut-off was chosen since the average sequence identity between TM segments in the database was approximately 20% with a standard deviation of 5%, suggesting 40% would remove a significant portion of database redundancy. This resulted in 203 entries, ranging widely in biological function and structural topology, consisting primarily of mostly multi-pass proteins.

Finally, all unique protein chains were identified in each pdb entry and extracted into new coordinate files to query for close structural matches. Additionally, residues from alternate chains

were included in this file, as well as residues 4 flanking residues. Thus, for each unique chain in an oligomeric protein, the chain itself as well as all of its unique protein-protein interactions are extracted; this minimizes double counting symmetrically repeated interfaces, though some intra-protein redundancy occurs at this stage that is later filtered. Each of these coordinate files was then converted to a binary query structural file, using the program MASTER (v1.3) and the CreatePDS executable (*47*).

Fragment construction and structural searches

Next, the helix-helix interface between adjacent chains A and B from the X-ray structure was divided into every possible fragment containing 9 consecutive residue stretches from both of the helices, i.e. with both chain A and B, 18 residues total. For PL5, residues Leu5 to Arg29 were analyzed. For mini '*e*'-Val, residues Leu3-Val21 were analyzed.

Each fragment was converted into a searchable binary structure file using MASTER (v1.3) and the CreatePDS executable. Then, each helical pair fragment was queried for close structural matched in the non-redundant set of diverse TM proteins, described above. A close match was counted if the backbone atom RMSD upon optimal superposition is less than 0.85 Å. This resulted in a large enough number of matches for every fragment for reliable sequence statistics to be derived (*74*). For PL5, the number of matches per fragment ranged between 48 to 85 (mean = 71). For mini '*e*'-Val, between 50 and 91 matches below the cutoff were found (mean = 79). For each residue, we analyzed the sequence compositions of the aligned structural matches, giving the amino acids found at each of the residue positions in that helix-helix interface within the natural TM proteins' context. When applicable, a residue's sequence alignment came from the query fragment in which

that residue was the central residue of the fragment (i.e. sequence analysis for Val14 is taken from the set of matches aligned to the query fragment Leu10-Leu18). At terminal residues, the sequence alignment comes from the corresponding terminal fragment.

To understand how large a given number of matches is relative to different helix-helix geometries and known motifs, we compared these data here to our previous work clustering all interacting helix pairs within membrane proteins (*3*). This work identified 1290 helical pairs that fell into clusters with at least 25 members (using a greedy clustering algorithm and 1.25 Å RMSD cutoff). Of these, about one-third (416) were parallel helical pairs, which were distributed between 7 clusters with member sizes ranging from 34 to 109. Using the same RMSD cut-off (1.25 Å) and fragment size (12 residue stretches), we queried a central segment of mini '*e*'-Val against that same database, resulting in 54 structural matches. This value represents 13% (54/416) of the total number of clustered parallel pairs, and it is larger than the median cluster size for parallel pairs.

The dimers from the natural MP database that match mini '*e*'-Val belong to one of two overlapping clusters identified in our previous database. These two clusters were originally grouped into a single cluster in an earlier survey (*48*) with a slightly less stringent RMSD cutoff. Together, they comprise 144 members, exceeding even the common GxxxG motif (109 members):

- '#3' with 94 members, inter-helical distance = 11.1 Å, crossing-angle = 14.0°;
- '#11' with 50 members, inter-helical distance = 9.8 Å, crossing angle = 34.4 °.

Of those 54 matches to 'e'-Val, 53 are shared with these two clusters (28 in #3 and 25 in #11).

We then assessed the how sensitive more subtly changes to the helix-helix geometry affect the number of structural matches we observed, summarized in Figure 1.15. We were concerned with how the pairwise inter-helical geometry varied as the pentameric bundle geometry varied. To investigate this, we used parametric equations to generate 800 models of C5 symmetric helical bundles, subtly varying in geometry by performing a grid search over the following parameters and physically relevant ranges: Super-helical radius (8.25 Å $< R_o < 9.75$ Å), Super-helical pitch angle (-24 ° $< \alpha < -8$ °), α-helical phase angle (1 ° $< \varphi_1 < 91$ °). From these bundle models, we extracted a 12-residue helical pair segment and searched our database from this study for the number of close geometric matches (1.1 Å). This cut-off was chosen because the big spread between the minimum and maximum number of matches found for the library of different helical pairs (min=8, max=111). The 'best' geometry giving the most matches forms a single peak at R_o ~ 8.5 Å, $\alpha \sim -16$ °, and $\varphi_1 \sim 91$ °. This corresponds to roughly the geometry for PL5 and '*e*'-Val (Table 1.1). Similarly, many structures of engineered and naturally occurring water-soluble coiled coils fall roughly in this geometry as well, particularly those with repeating motifs of hydrophobic residues at '*a*' and '*g*' positions (Fig. 1.15, Table 1.1, 1.7).

Analysis of sequence data from structural matches

The local environment surrounding each residue at the pairwise interfaces in the X-ray structures were analyzed by examining the sequences of all the database matching fragments. For each residue, a sequence profile is constructed from the sequences of close geometric matches (0.85 Å) extracted from the culled database of natural protein structures. For a specific residue in the protomer, the sequence distribution for that residue is built from the distribution of amino acids at the corresponding position in the set of matches to the query fragment. Though most residues are

included in several fragments, the sequence profile of a residue of interest is the amino acid distribution observed in sequences from close matches to the single fragment where that residue is the central residue of the query fragment. For terminal residues, the amino acid distribution is derived from matches to the terminal fragment. For example, the sequence profile for PL5 residue 9 from chain A is derived from close matches to a fragment spanning 5-14 (from both chain A and B), using the distribution of amino acids at just the corresponding position (the 5th position of helix A, residue 9 chain A) in the set of sequences of these close matches; the profile for residue 10 chain A would come from matches to the helix pair fragment spanning 6-15.

For each residue in the parent protein, the distribution of amino acids at the corresponding positions in a set of matches to the appropriate query fragment were analyzed by calculating the relative entropy (Kullback-Leibler divergence) versus the frequency of each amino acids in the TM protein structural database (Table 1.8). For residue x, the total relative entropy, D_x, is the sum of absolute values of the relative entropies calculated for each amino acid observed in the set of matches:

$$D_{x,i} = F_{x,i} \cdot \ln \left(F_{x,i} / Q_{x,i} \right) \qquad \text{Equation 3}$$

where $D_{x,i}$ is the relative entropy of residue x for amino acid i, $F_{x,i}$ is the observed frequency of amino acid i at the equivalent residue position x in the set of closely matching fragments, and $Q_{x,i}$ is the background frequency of that amino acid in our TM domain database. Sequence profile plots in WebLogo format were made with the total height of each letter is $D_{x,i}$.

The level of significance of the enrichment observed for a particular amino acid is calculated by a one-tailed binomial test. The p-value is calculated for each amino acid at each position individually, determining how significantly the frequencies we observe deviate from the expected

frequency: the observed distribution of amino acids found in the entirety of our membrane protein database (Table 1.8). For example, in the database matches at the position corresponding to Leu9 in '*e*'-Val, Cys is observed in 7 of 77 matching database sequences (9.09%) where its expected rate is (1.4%) giving a p-value for this 6.4-fold enrichment of 0.0001.

Given enrichments of slightly polar residues (Cys, Ser, Thr at the '*e/d*' or '*a/g*' interfaces), we examined the whether those side-chains formed inter-helical hydrogen bonds in the corresponding natural proteins structures. One caveat is side-chain modeling may not always be exactly accurate for models nearing the lower resolution range of our database, 3.2 Å. For all the '*a*,' '*d*,' '*e*,' and '*g*' positions in the middle heptad of '*e*'-Val, we observed 5 likely side-chain hydrogen bonds out of 68 observed Ser, Cys, or Thr residues (7.3%), all from models. For the '*g*' position, no hydrogen bonds were observed for those side-chains (Cys enriched). At '*e*' where Thr is highly enriched (13 instances), we observed two likely hydrogen bonds – one in a 2.2 Å resolution structure and another at 3.2 Å. Inter-helical hydrogen-bonds are not especially frequent in the enriched positions at the helix-helix interface, and should not impair our overall conclusions or statistical enrichments.

Authors contributions

MM and WFD conceived of the work and designed the experiments. MM and MT performed simulations. MM and PS produced and purified the proteins, and characterized them by analytical ultracentrifugation, gel electrophoresis, CD spectroscopy. MM crystallized the proteins. MM, JLT, and LL conducted X-ray diffraction, analyzed X-ray data, and solved the structures. MM conducted the bioinformatic analysis. All authors contributed to analysis of the data and writing of the manuscript.

References

1.   A. Oberai, N. H. Joh, F. K. Pettit, J. U. Bowie, Structural imperatives impose diverse evolutionary constraints on helical membrane proteins. *Proc Natl Acad Sci U S A* **106**, 17747-17750 (2009).

2.   M. Eilers, S. C. Shekar, T. Shieh, S. O. Smith, P. J. Fleming, Internal packing of helical membrane proteins. *Proc Natl Acad Sci U S A* **97**, 5796-5801 (2000).

3.   S. Q. Zhang *et al.*, The membrane- and soluble-protein helix-helix interactome: similar geometry via different interactions. *Structure* **23**, 527-541 (2015).

4.   L. Adamian, J. Liang, Helix-helix packing and interfacial pairwise interactions of residues in membrane proteins. *J Mol Biol* **311**, 891-907 (2001).

5.   D. Langosch, J. Heringa, Interaction of transmembrane helices by a knobs-into-holes packing characteristic of soluble coiled coils. *Proteins* **31**, 150-159 (1998).

6.   H. Hong, Toward understanding driving forces in membrane protein folding. *Arch Biochem Biophys* **564**, 297-313 (2014).

7.   S. M. Anderson, B. K. Mueller, E. J. Lange, A. Senes, Combination of Calpha-H Hydrogen Bonds and van der Waals Packing Modulates the Stability of GxxxG-Mediated Dimers in Membranes. *J Am Chem Soc* **139**, 15774-15783 (2017).

8.   N. H. Joh, A. Oberai, D. Yang, J. P. Whitelegge, J. U. Bowie, Similar energetic contributions of packing in the core of membrane and water-soluble proteins. *J Am Chem Soc* **131**, 10846-10847 (2009).

9.   Y. Yano, K. Kondo, R. Kitani, A. Yamamoto, K. Matsuzaki, Cholesterol-Induced Lipophobic Interaction between Transmembrane Helices Using Ensemble and Single-Molecule Fluorescence Resonance Energy Transfer. *Biochemistry* **54**, 1371-1379 (2015).

10.   R. P. Baker, S. Urban, Architectural and thermodynamic principles underlying intramembrane protease function. *Nature Chemical Biology* **8**, 759 (2012).

11.   A. K. Doura, F. J. Kobus, L. Dubrovsky, E. Hibbard, K. G. Fleming, Sequence context modulates the stability of a GxxxG-mediated transmembrane helix-helix dimer. *J Mol Biol* **341**, 991-998 (2004).

12.   R. Guo *et al.*, Steric trapping reveals a cooperativity network in the intramembrane protease GlpG. *Nat Chem Biol* **12**, 353-360 (2016).

13.   N. H. Joh *et al.*, Modest stabilization by most hydrogen-bonded side-chain interactions in membrane proteins. *Nature* **453**, 1266-1270 (2008).

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

15.   M. G. Teese, D. Langosch, Role of GxxxG Motifs in Transmembrane Domain Interactions. *Biochemistry* **54**, 5125-5135 (2015).

16.   A. Senes, I. Ubarretxena-Belandia, D. M. Engelman, The Calpha ---H...O hydrogen bond: a determinant of stability and specificity in transmembrane helix interactions. *Proc Natl Acad Sci U S A* **98**, 9056-9061 (2001).

17.   F. X. Zhou, H. J. Merianos, A. T. Brunger, D. M. Engelman, Polar residues drive association of polyleucine transmembrane helices. *Proc Natl Acad Sci U S A* **98**, 2250-2255 (2001).

18.   F. Cymer, G. von Heijne, S. H. White, Mechanisms of integral membrane protein insertion and folding. *J Mol Biol* **427**, 999-1022 (2015).

19.   R. M. Johnson, K. Hecht, C. M. Deber, Aromatic and cation-pi interactions enhance helix-helix association in a membrane environment. *Biochemistry* **46**, 9208-9214 (2007).

20.   D. Langosch, I. T. Arkin, Interaction and conformational dynamics of membrane-spanning protein helices. *Protein Sci* **18**, 1343-1358 (2009).

21.   R. Gurezka, R. Laage, B. Brosig, D. Langosch, A heptad motif of leucine residues found in membrane proteins can drive self-assembly of artificial transmembrane segments. *J Biol Chem* **274**, 9265-9270 (1999).

22.   Y. Yano *et al.*, Topological stability and self-association of a completely hydrophobic model transmembrane helix in lipid bilayers. *Biochemistry* **41**, 3073-3080 (2002).

23.   R. M. Johnson, C. L. Heslop, C. M. Deber, Hydrophobic helical hairpins: design and packing interactions in membrane environments. *Biochemistry* **43**, 14361-14369 (2004).

24.   P. Whitley, I. Nilsson, G. von Heijne, De novo design of integral membrane proteins. *Nat Struct Biol* **1**, 858-862 (1994).

25.   C. D. Tatko, V. Nanda, J. D. Lear, W. F. Degrado, Polar networks control oligomeric assembly in membranes. *J Am Chem Soc* **128**, 4170-4171 (2006).

26.   N. H. Joh *et al.*, De novo design of a transmembrane Zn(2)(+)-transporting four-helix bundle. *Science* **346**, 1520-1524 (2014).

27.   I. V. Korendovych *et al.*, De novo design and molecular assembly of a transmembrane diporphyrin-binding protein complex. *J Am Chem Soc* **132**, 15516-15518 (2010).

28.   H. Yin *et al.*, Computational design of peptides that target transmembrane helices. *Science* **315**, 1817-1822 (2007).

29.   M. Mravic *et al.*, De novo designed transmembrane peptides activating the alpha5beta1 integrin. *Protein engineering, design & selection : PEDS* **31**, 181-190 (2018).

30.   P. Lu *et al.*, Accurate computational design of multipass transmembrane proteins. *Science* **359**, 1042-1046 (2018).

31. A. L. Lomize, M. A. Lomize, S. R. Krolicki, I. D. Pogozheva, Membranome: a database for proteome-wide analysis of single-pass membrane proteins. *Nucleic Acids Res* **45**, D250-d255 (2017).

32. J. Kirrbach *et al.*, Self-interaction of transmembrane helices representing pre-clusters from the human single-span membrane proteins. *Bioinformatics* **29**, 1623-1630 (2013).

33. A. W. Partridge, A. G. Therien, C. M. Deber, Missense mutations in transmembrane domains of proteins: phenotypic propensity of polar residues for human disease. *Proteins* **54**, 648-656 (2004).

34. J. P. Schlebach, C. R. Sanders, The safety dance: biophysics of membrane protein folding and misfolding in a cellular context. *Quarterly reviews of biophysics* **48**, 1-34 (2015).

35. A. R. Thomson *et al.*, Computational design of water-soluble alpha-helical barrels. *Science* **346**, 485-488 (2014).

36. P. Harbury, T. Zhang, P. Kim, T. Alber, A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants. *Science* **262**, 1401-1407 (1993).

37. F. H. Crick, The packing of-helices: simple coiled-coils. *Acta crystallographica* **6**, 689-697 (1953).

38. I. T. Arkin *et al.*, Structural organization of the pentameric transmembrane alpha-helices of phospholamban, a cardiac ion channel. *Embo j* **13**, 4757-4764 (1994).

39. R. Verardi, L. Shi, N. J. Traaseth, N. Walsh, G. Veglia, Structural topology of phospholamban pentamer in lipid bilayers by a hybrid solution and solid-state NMR method. *Proc Natl Acad Sci U S A* **108**, 9101-9106 (2011).

40. K. Oxenoid, J. J. Chou, The structure of phospholamban pentamer reveals a channel-like architecture in membranes. *Proc Natl Acad Sci U S A* **102**, 10870-10875 (2005).

41.   T. Kim, J. Lee, W. Im, Molecular dynamics studies on structure and dynamics of phospholamban monomer and pentamer in membranes. *Proteins: Structure, Function, and Bioinformatics* **76**, 86-98 (2009).

42.   J. Fujii, K. Maruyama, M. Tada, D. H. MacLennan, Expression and site-specific mutagenesis of phospholamban. Studies of residues involved in phosphorylation and pentamer formation. *J Biol Chem* **264**, 12950-12955. (1989).

43.   H. K. Simmerman, Y. M. Kobayashi, J. M. Autry, L. R. Jones, A leucine zipper stabilizes the pentameric membrane domain of phospholamban and forms a coiled-coil pore structure. *J Biol Chem* **271**, 5941-5946 (1996).

44.   J. U. Bowie, Membrane protein folding: how important are hydrogen bonds? *Curr Opin Struct Biol* **21**, 42-49 (2011).

45.   C. B. Karim *et al.*, Role of cysteine residues in structural stability and function of a transmembrane helix bundle. *J Biol Chem* **276**, 38814-38819 (2001).

46.   I. T. Arkin *et al.*, Structure of the transmembrane cysteine residues in phospholamban. *J Membr Biol* **155**, 199-206 (1997).

47.   J. Zhou, G. Grigoryan, Rapid search for tertiary fragments reveals protein sequence-structure relationships. *Protein Sci* **24**, 508-524 (2015).

48.   R. F. Walters, W. F. DeGrado, Helix-packing motifs in membrane proteins. *Proc Natl Acad Sci U S A* **103**, 13658-13663 (2006).

49.   K. R. MacKenzie, J. H. Prestegard, D. M. Engelman, A transmembrane helix dimer: structure and implications. *Science* **276**, 131-133 (1997).

50.   P. S. Huang *et al.*, High thermodynamic stability of parametrically designed helical bundles. *Science* **346**, 481-485 (2014).

51.   J. R. Moll, S. B. Ruvinov, I. Pastan, C. Vinson, Designed heterodimerizing leucine zippers with a ranger of pIs and stabilities up to 10(-15) M. *Protein Sci* **10**, 649-655 (2001).

52.   X. Feng, P. Barth, A topological and conformational stability alphabet for multipass membrane proteins. *Nat Chem Biol* **12**, 167-173 (2016).

53.   L. He *et al.*, Single methyl groups can act as toggle switches to specify transmembrane Protein-protein interactions. *Elife* **6**, (2017).

54.   M. Young *et al.*, Computational design of orthogonal membrane receptor-effector switches for rewiring signaling pathways. *Proceedings of the National Academy of Sciences* **115**, 7051-7056 (2018).

55.   G. E. Crooks, G. Hon, J. M. Chandonia, S. E. Brenner, WebLogo: a sequence logo generator. *Genome Res* **14**, 1188-1190 (2004).

56.   G. Grigoryan, W. F. Degrado, Probing designability via a generalized model of helical bundle geometry. *J Mol Biol* **405**, 1079-1100 (2011).

57.   M. A. Lomize, A. L. Lomize, I. D. Pogozheva, H. I. Mosberg, OPM: orientations of proteins in membranes database. *Bioinformatics* **22**, 623-625 (2006).

58.   L. G. Reddy *et al.*, Functional reconstitution of recombinant phospholamban with rabbit skeletal Ca(2+)-ATPase. *J Biol Chem* **270**, 9390-9397 (1995).

59.   N. W. Schmidt, G. Grigoryan, W. F. DeGrado, The accommodation index measures the perturbation associated with insertions and deletions in coiled-coils: Application to understand signaling in histidine kinases. *Protein Sci* **26**, 414-435 (2017).

60. C. Tanford, J. A. Reynolds, Characterization of membrane proteins in detergent solutions. *Biochimica et Biophysica Acta (BBA) - Reviews on Biomembranes* **457**, 133-170 (1976).

61. C. L. A. Schmidt, Proteins, Amino Acids and Peptides as Ions and Dipolar Ions (Cohn, Edwin J.; Edsall, John T.). *Journal of Chemical Education* **20**, 415 (1943).

62. T. G. Battye, L. Kontogiannis, O. Johnson, H. R. Powell, A. G. Leslie, iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr D Biol Crystallogr* **67**, 271-281 (2011).

63. P. Evans, Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr* **62**, 72-82 (2006).

64. P. R. Evans, G. N. Murshudov, How good are my data and what is the resolution? *Acta Crystallographica Section D: Biological Crystallography* **69**, 1204-1214 (2013).

65. A. J. McCoy *et al.*, Phaser crystallographic software. *Journal of Applied Crystallography* **40**, 658-674 (2007).

66. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501 (2010).

67. P. D. Adams *et al.*, PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-221 (2010).

68. W. Kabsch, Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr D Biol Crystallogr* **66**, 133-144 (2010).

69. J. Huang, A. D. MacKerell, Jr., CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. *J Comput Chem* **34**, 2135-2145 (2013).

70.   S. Jo *et al.*, CHARMM-GUI 10 years for biomolecular modeling and simulation. *J Comput Chem* **38**, 1114-1124 (2017).

71.   D. H. de Jong *et al.*, Improved Parameters for the Martini Coarse-Grained Protein Force Field. *J Chem Theory Comput* **9**, 687-697 (2013).

72.   T. A. Wassenaar, H. I. Ingolfsson, M. Priess, S. J. Marrink, L. V. Schafer, Mixing MARTINI: electrostatic coupling in hybrid atomistic-coarse-grained biomolecular simulations. *J Phys Chem B* **117**, 3516-3530 (2013).

73.   O. S. Smart, J. M. Goodfellow, B. A. Wallace, The pore dimensions of gramicidin A. *Biophys J* **65**, 2455-2460 (1993).

74.   T. D. Schneider, R. M. Stephens, Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res* **18**, 6097-6100 (1990).
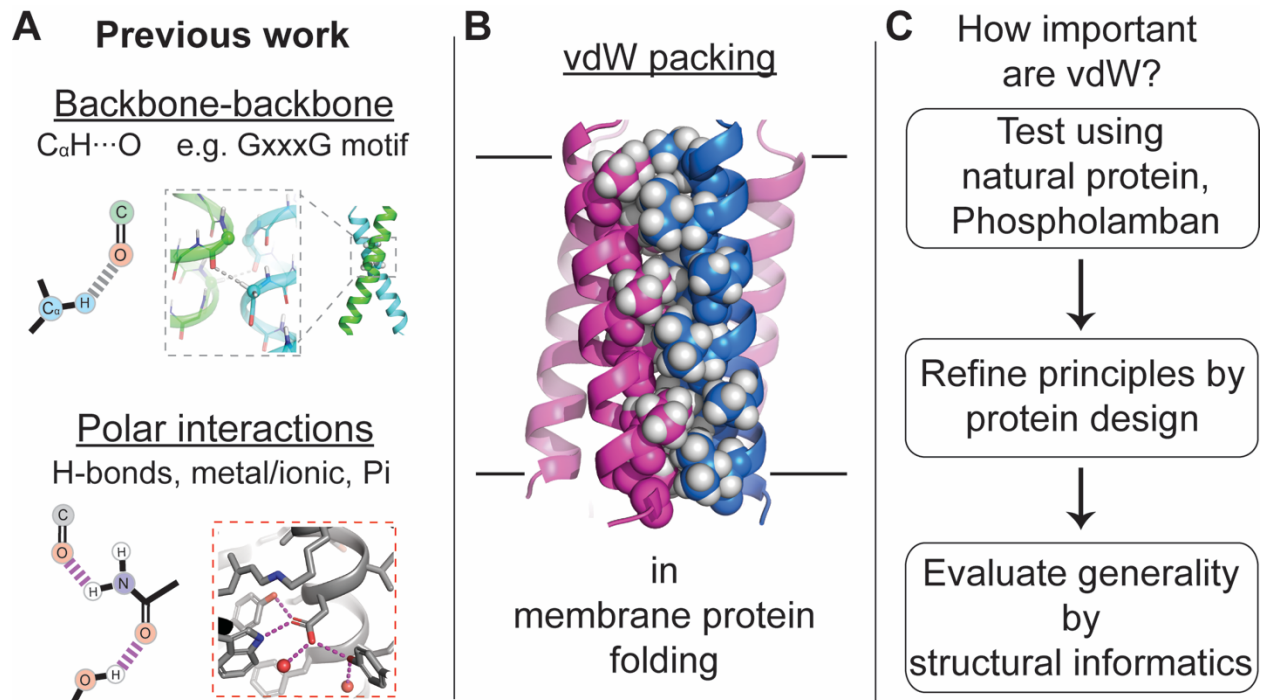
Figure 1.1. Non-covalent forces in membrane protein folding. (A) Polar interactions are known to stabilize membrane protein structures. **(B)** vdW packing is abundant in the folded state, but similar interactions with membrane lipids occur in the unfolded state; it is unknown whether packing alone can drive membrane protein folding. **(C)** Overview of the multi-prong approach used.
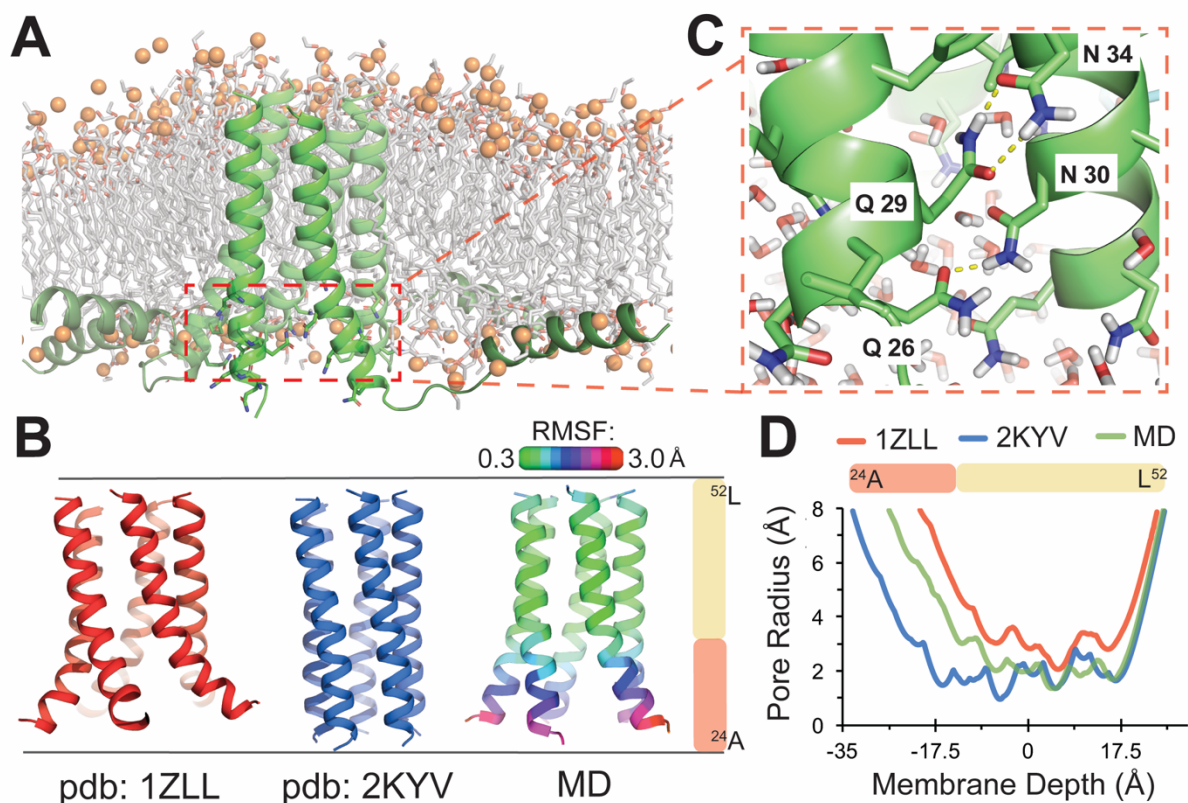
Figure 1.2. All-atom molecular dynamics simulation of phospholamban (PLN), a pentameric transmembrane α-helical bundle, illuminates critical apolar packing interactions. **(A)** Final simulation frame for full-length PLN in a POPC bilayer; water, hidden. **(B)** Comparison of the TM domains of published NMR structures of PLN (PDB: 1ZLL, red; 2KYV, blue), compared to our MD simulation. RMSF of backbone atoms versus the simulation medoid displayed as ribbon color. **(C)** Snapshot showing the water-filled central cavity and rapidly fluctuation polar side-chain interactions at the splayed N-terminal third of the TM-spanning α-helices. **(D)** Central cavity size within each helical bundle versus membrane depth.
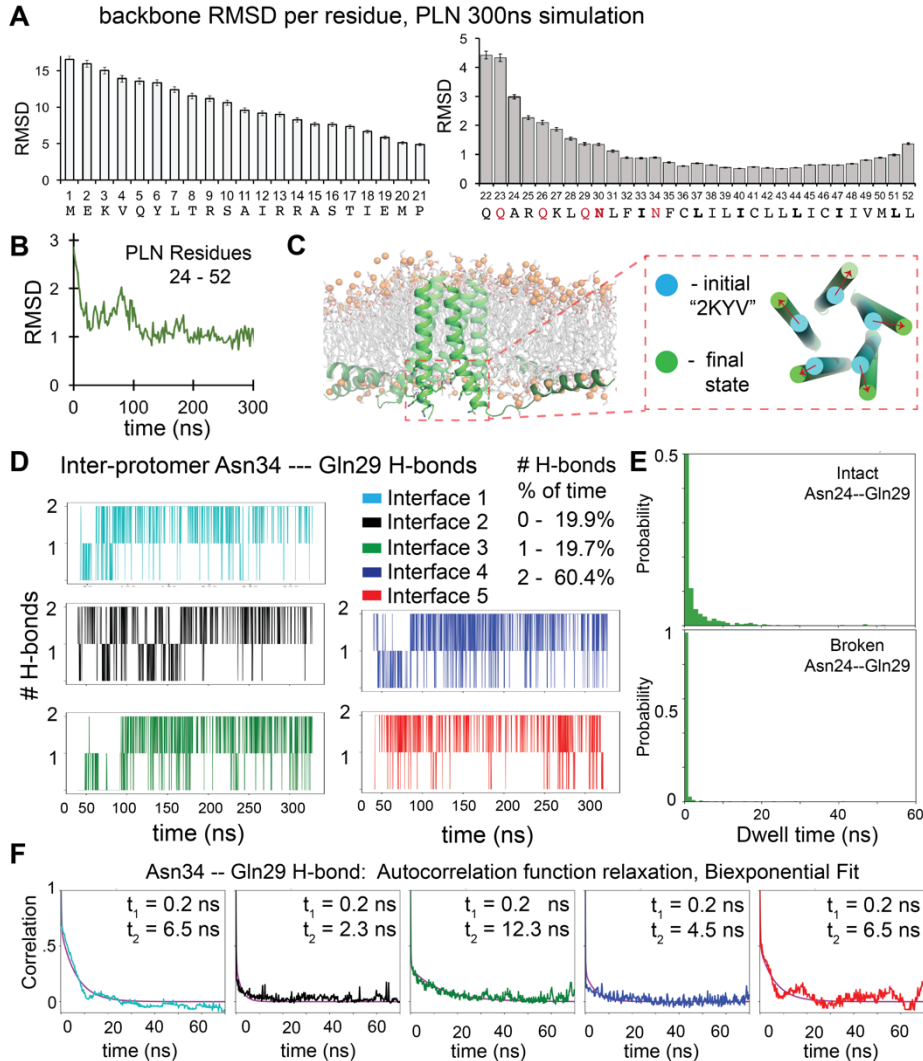
Figure 1.3. Analysis of full-length phospholamban (PLN) molecular dynamics simulation
(A) Mean RMSD of backbone atoms of the full-length PLN throughout the simulation (300 ns) after aligning the coiled-coil region (residues 32-52) of each frame to those atoms in the medoid frame. **(B)** Backbone RMSD over time for residues 24-52 **(C)** Snapshot and cartoon depiction showing the splaying of the N-terminal portion of the PLN transmembrane helical bundle in the simulation (green) versus the reported NMR model which was the starting point (blue, PDB: 2KYV). **(D)** Tracking the number of hydrogen-bonds between the Asn34 to Gln29 side-chains across each of the 5 equivalent helix-helix interfaces in the pentameric helical bundle throughout the simulation. **(E)** Frequency of observed dwell time between events of breaking hydrogen-bonds and making one or more hydrogen-bonds, for data in **(C)**. **(F)** The autocorrelation function of each trace in **(C)** was calculated and fit to a biexponential decay function, giving two time components corresponding. Each function had a fast component the lowest allowed value of the fit, 0.2 ns, and a longer component (6.5, 2.3, 12.3, 4.5, 6.5 ns). Curve fit, purple; trace colors match legend in **(C)**.
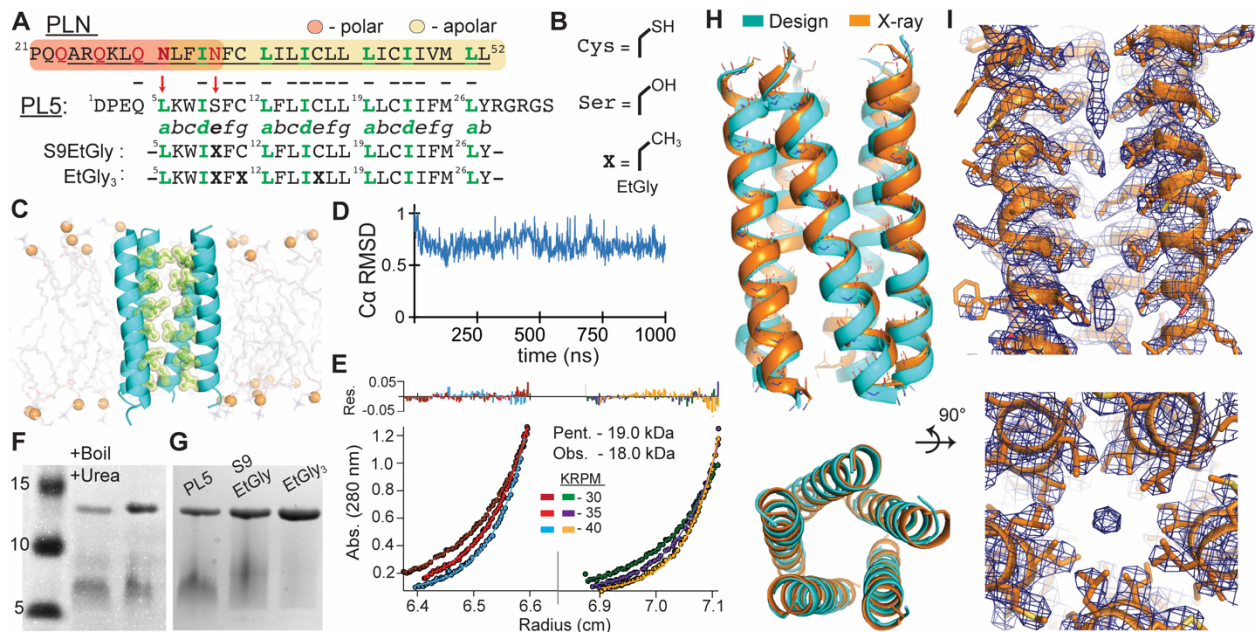
Figure 1.4. Design and structural characterization of phospholamban-like pentameric protein PL5.

(**A**) Sequence of PLN and PL5. PLN's polar region (red) and apolar region (yellow) highlighted; membrane-spanning α-helix underlined (residues 24-52). The 7-residue α-helical repeat is labeled "*abcdefg*"; LxxIxxx motif, green. Polar to apolar mutations, red arrows. Bolded 'X' denotes S-C_α-ethyl-Glycine (EtGly) residues. (**B**) EtGly is approximately isosteric to Ser and Cys. (**C**) Snapshot from a 1.0 microsecond MD simulation of PL5 in a POPC bilayer. (**D**) PL5 shows conformational rigidity by C_α RMSD versus medoid frame (residues 5-29). (**E**) Equilibrium analytical ultracentrifugation, PL5 at 58 μM and 79 μM in 33 mM myristyl sulfobetaine micelles, globally fit to a single species model: apparent molecular weight = 4.74 monomers. (**F**) SDS-PAGE of PL5 in reducing conditions shows a single oligomeric state, resistant to heating (95° C, 30 minutes) in 2% LDS, 8 M urea. Similar to PLN, PL5 exhibits aberrant gel migration. Figure 1.5 confirms the slower band is pentameric. (**G**) PL5, PL5$_{EtG}$, and PL5$_{EtG3}$ have similar oligomeric distributions. (**H**) The pentameric x-ray structure of PL5 (pdb: 6MQU) closely matches the MD-refined design model (cyan). (**I**) Well-packed side-chains are well-resolved in the $2F_o$-F_c electron density map (σ = 1.0). Elongated density is present only at divergent ends of the bundle.

43
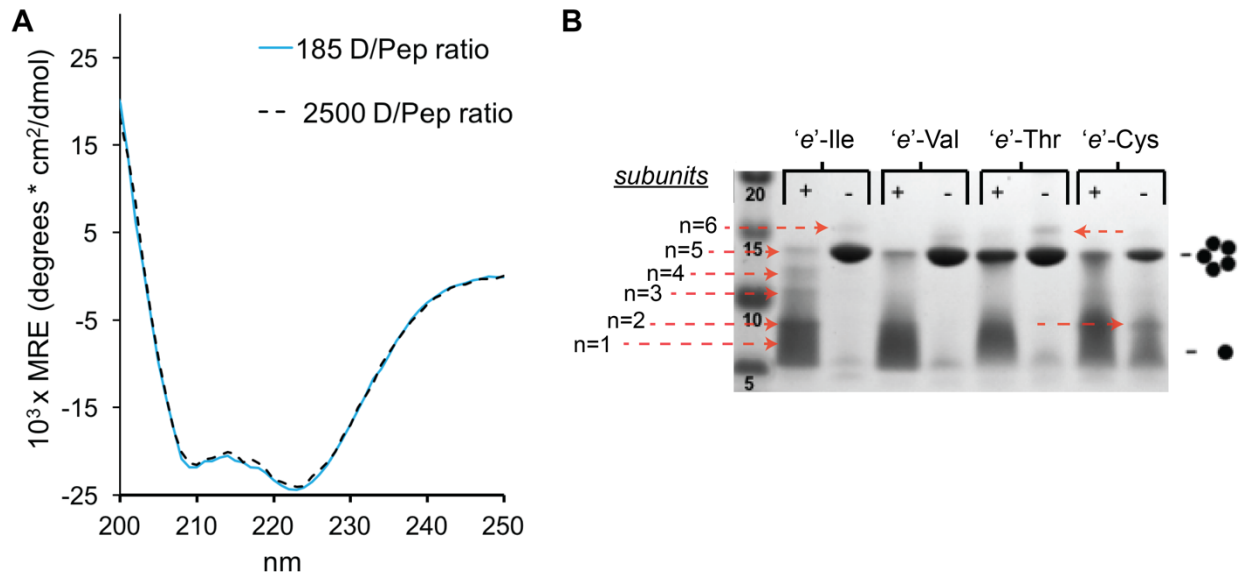
Figure 1.5. Circular dichroism spectra of PL5 peptide and gel migration in detergent micelles
(A) Synthetic PL5 full sequence : DPEQLKWISFCLFLICLLLLCIIFMLYRGrGS-carboxamide:
r, D-Arg. Far UV circular dichroism spectra of PL5 at 75 μM in 10 mM Tris HCl buffer pH 7.4,
2 mM TCEP, solubilized by dodecylphoshocholine (DPC) micelles either 185 or 2500 molar
excess. 'D/Pep,' detergent to peptide ratio. The spectra are characteristic of α-helices.
(B) SDS-PAGE of designed TM peptides described in Figure 1.12, under conditions previous
used to count the number of monomers in PLN. Oligomeric bands, ranging from n=1 to n=5, and
possibly n=6, are labelled. Under these conditions, the peptide was solubilized in SDS and
reducing agent, then an aliquot was boiled, similar to previously described for PLN (*58*).
Presumably due to degradation of the reducing agent and covalent disulfide bond formation,
additional oligomeric bands at intermediate molecular weight are observed for the Ile peptide.
This banding pattern is not observed if fresh reducing agent is added after boiling, as in Figures
1.4 and 1.12. This migration pattern allows us to assign the pentameric oligomer (major band)
and confirm that alternative oligomeric species seen in this gel are not observed under the
reducing conditions of our experiments.

44
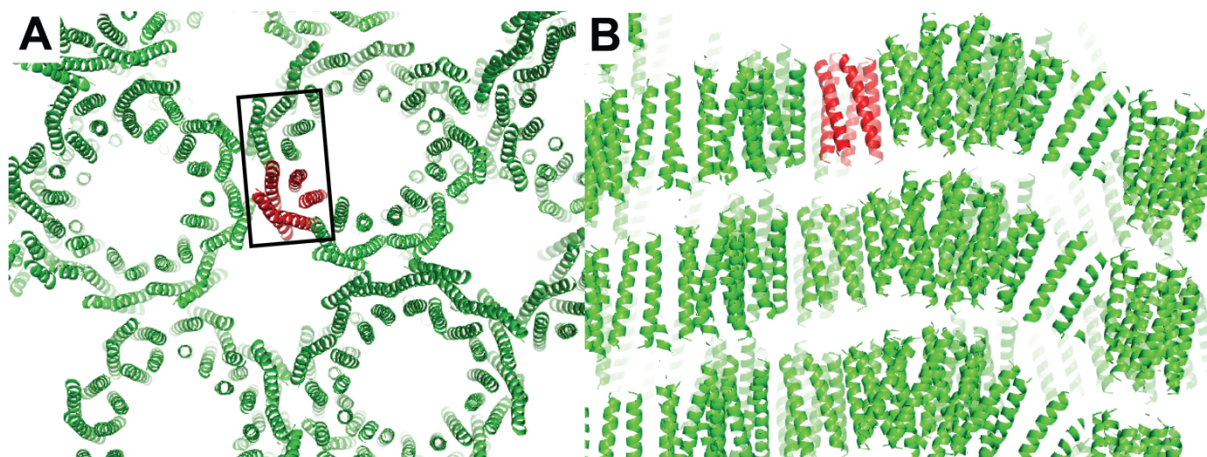
Figure 1.6. Crystal lattice of PL5.
(A) The asymmetric unit contains two structurally similar PL5 pentamers (backbone RMSD = 0.18 Å) matching the designed model (black outline), that are anti-parallel to one another. Eight pentamers (40 PL5 peptides) form large (ca. 45 Å) pores with diffuse electron density, suggesting a disordered solvent-filled area. **(B)** These 2D porous layers stack in the 3rd dimension, approximately parallel to the PL5 coiled-coil axis. A biologically relevant pentamer whose structure we used for analysis is shown in red in both **(A)** and **(B)**.
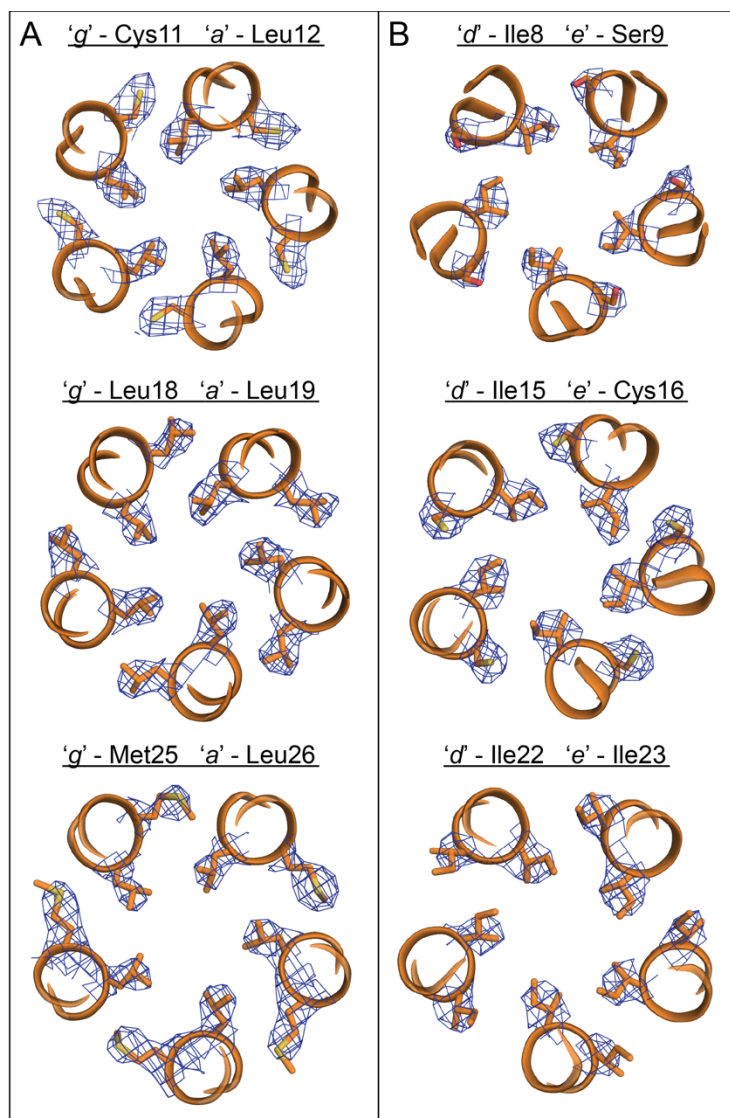
Figure 1.7. **2F$_0$-F$_c$ maps showing the definitions of PL5 side-chain conformations**.
(A) Electron density from a 2F$_0$ − F$_C$ map is shown in blue contoured at σ=1 for the side-chains at each 'layer' of interacting 'a' and 'g' residues within the repeating PL5 helical bundle.
(B) The 'd' and 'e' layer.
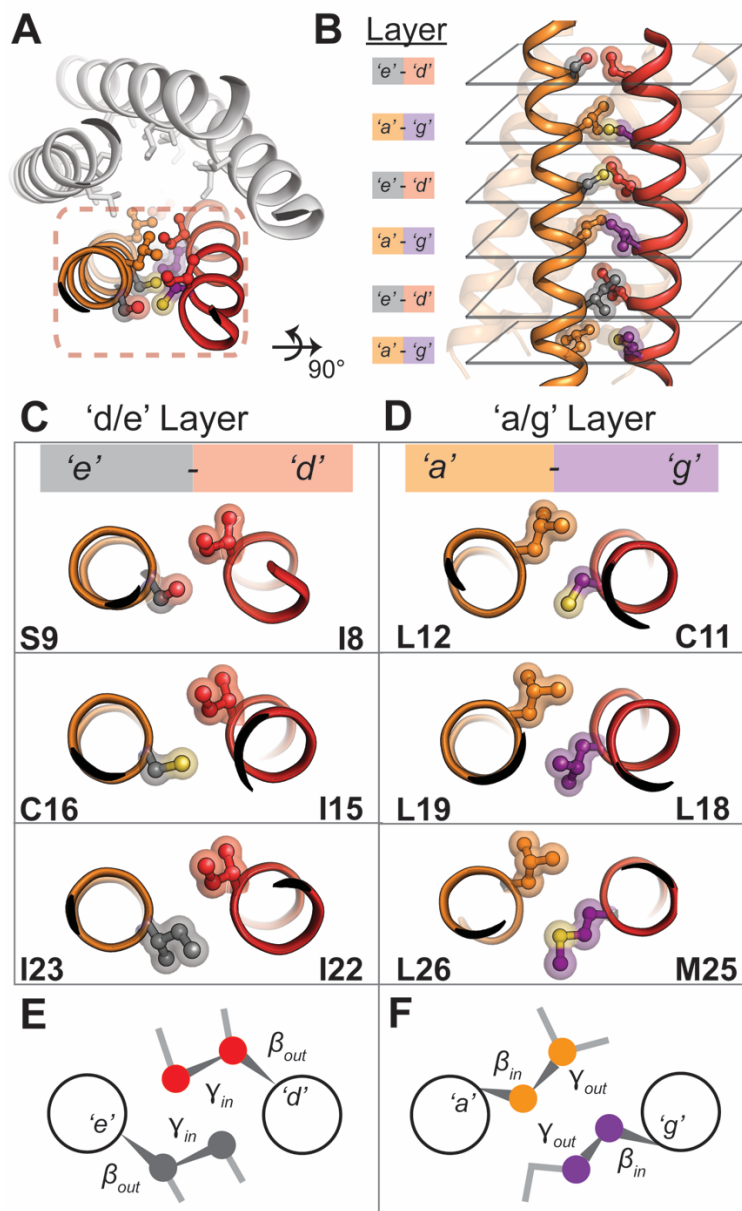
Figure 1.8. Side-chain steric packing at PL5's symmetric helix-helix interface. (A) The pairwise interaction of helices, symmetrically repeated, provides the primary stabilization for PL5. **(B)** High geometric complementarity of interacting residues across the helix-helix interface, roughly in layers: the '*a/g*' and '*e/g*' layers. **(C-D)** Axial view of side-chain packing of individual layers. **(E)** A potential stereo-chemical code required for pentameric assembly. At the '*e/d*' layer, the C_α-C_β bond vector of each amino acid points outward from the helix-helix interface (β_{out}), while the C_β-$C\gamma$ bond vector faces inward (γ_{in}). This suggests a heavy atom (i.e. N, C, O, S) at the gauche$^+$ position is required for tight inter-helical packing. **(F)** In the '*a/g*' layer, the opposite is true; the C_α-C_β bond vector points inward (β_{in}) while the C_β-Cγ bond vector places outward (γ_{out}).

A Super-helical Radius B Super-Coiling Twist C Pitch Angle
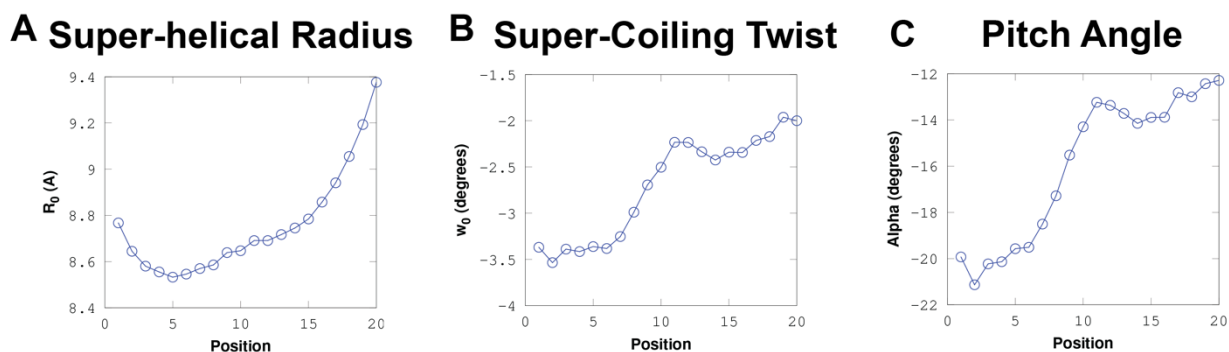
Figure 1.9. Coiled-coil parameters of PL5 fit in 7-residue spans.
Seven residue segments of PL5, starting with 4-11 (window position 1) and ending with 22-29 (window position 20), were fit independently to a generalized version of Crick's coiled-coil equations, imposing C5 symmetry, using the accommodation index web-server (*59*). The equation parameters for successive 7 residue spans are plotted by position: the **(A)** super-helical radius (R_0), **(B)** the super-helical frequency or super-coiling twist (w_0), and **(A)** the super-helical pitch angle or interface angle (α). As the bundle moves from the N- to the C-termini, R_0 increases gradually, while the values of α and w_0 increase suddenly at the C-terminal half of the bundle, where the 'e' and 'g' residues are much larger (Leu18, Ile23, Met25) versus the N-terminal half (Ser9, Cys11, Cys16). In summary, at the helix-helix interface the helices diverge and straighten, and the coiled-coil unwinds at its end.
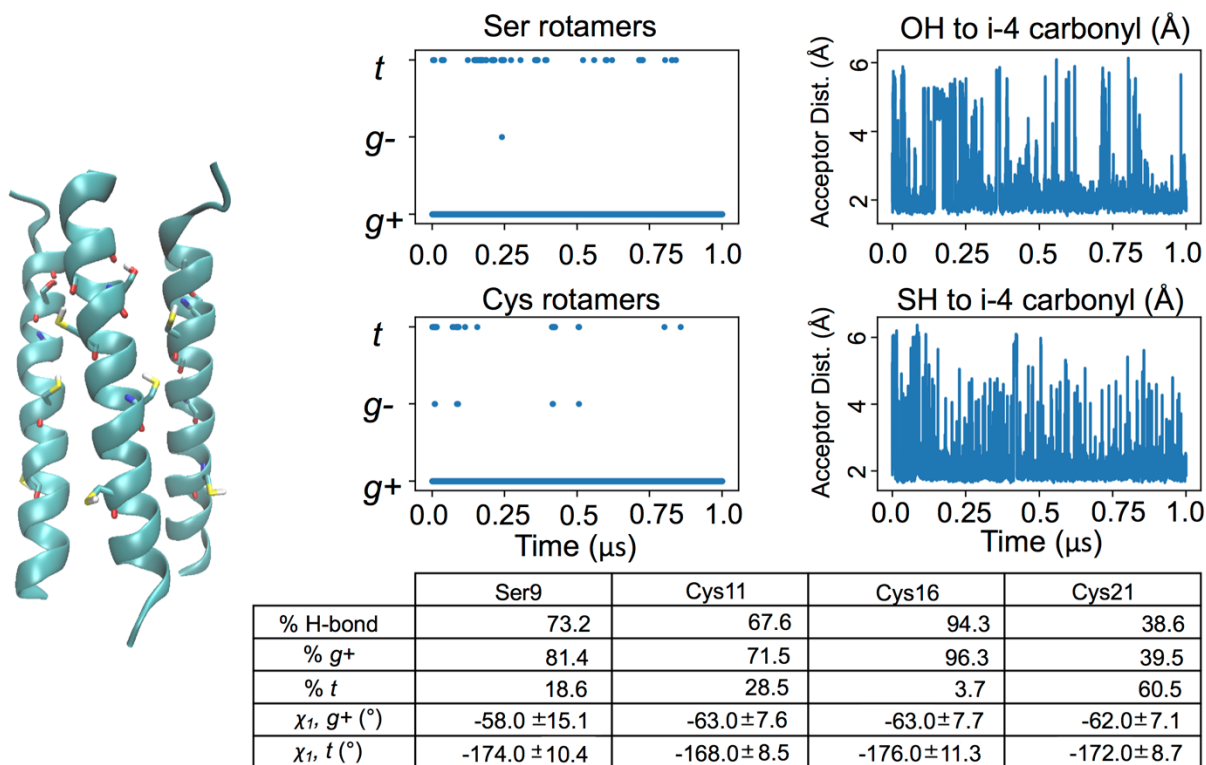
Figure 1.10. Cys and Ser residues in PL5 predominantly adopt an intra-helical hydrogen-bond and a single predominant rotameric state.

Left, three of five helices of PL5 from the final simulation frame with the Ser and Cys side-chains shown in sticks. Right, the plots show traces of rotamer distribution and hydrogen-bond distance for a chosen representative side-chain during the PL5 simulation. Values in the table below represent trajectory averages across all monomers. Interface facing residues Ser9, Cys11, and Cys16 maintain mostly a single rotameric state, gauche$^+$, stabilized by intra-helical hydrogen-bonds to a preceding main-chain carbonyl atom (at position i-4), while the much more solvent exposed Cys22 shows less of a preference.

The table below the plots:

	Ser9	Cys11	Cys16	Cys21
% H-bond	73.2	67.6	94.3	38.6
% $g+$	81.4	71.5	96.3	39.5
% t	18.6	28.5	3.7	60.5
χ_1, $g+$ (°)	-58.0 ±15.1	-63.0±7.6	-63.0±7.7	-62.0±7.1
χ_1, t (°)	-174.0 ±10.4	-168.0±8.5	-176.0±11.3	-172.0±8.7
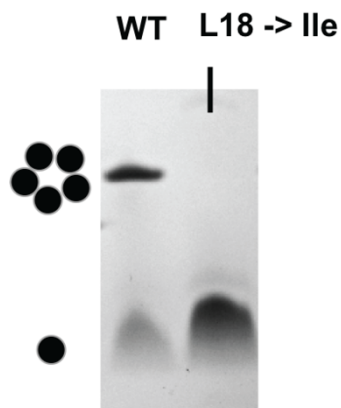
Figure 1.11. Leu to Ile substitution at interfacial '*g*' position of recombinant PL5 disrupts side-chain packing and self-assembly.

SDS-PAGE of recombinant PL5, left lane, and the Leu18Ile mutant, right. For PL5, a single oligomeric species is observed, the pentamer. However, only the monomeric band is observed for the mutant. This isomeric Leu-to-Ile substitution of Leu to a β-branched amino acid disrupts side-chain packing at the inter-subunit helix-helix interface and completely impairs pentameric self-assembly. This finding supports a sequence guideline that the introduction of a β-branched residue at the '*g*' position of the 7-residue repeat ('*abcdefg*') in PL5's sequence disrupts side-chain steric complementarity required for pentamerization.
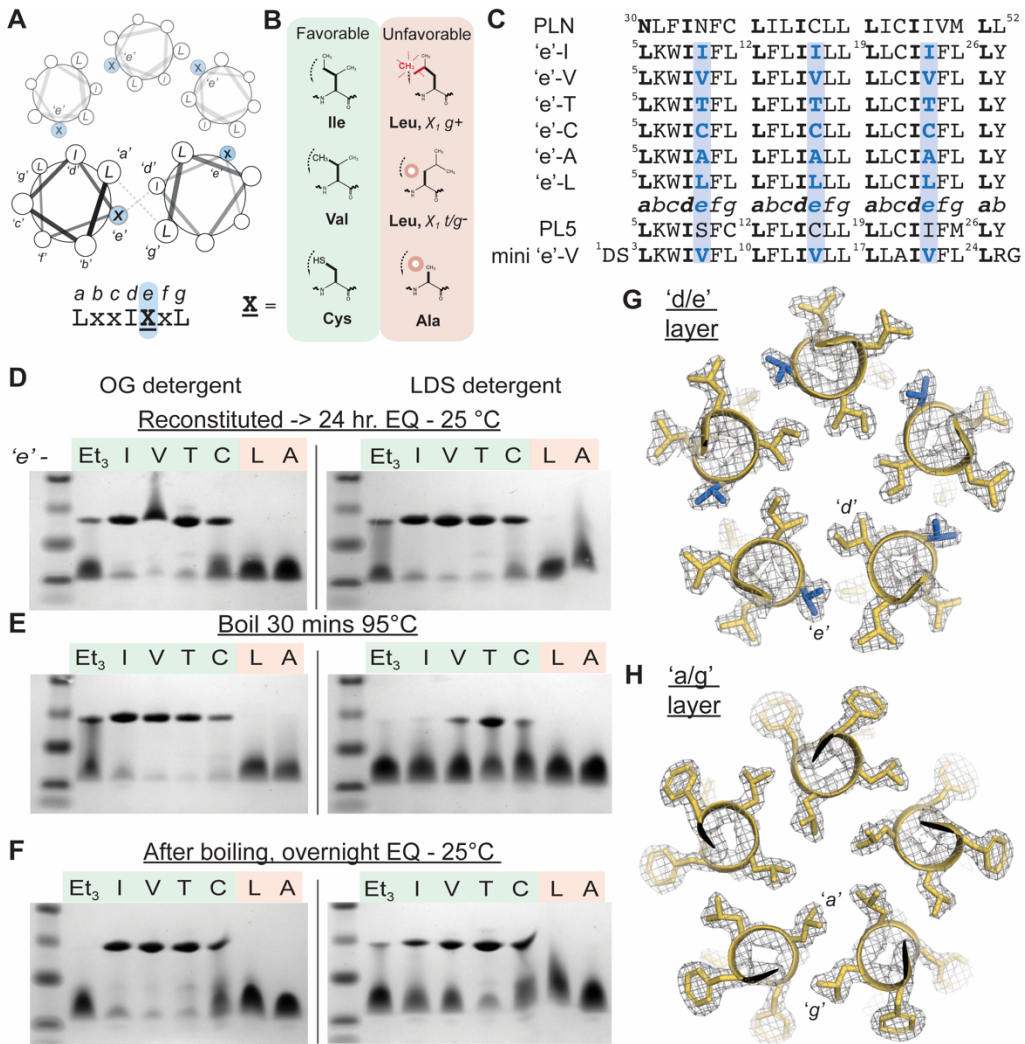
Figure 1.12. Design of pentameric membrane proteins from first principles of steric packing

(A) Helical wheel diagram of a pentameric bundle, the identity of the '*e*' position at the inter-subunit interface. **(B)** Given the proposed steric code in Fig. 1.8, Ile, Val, and Cys should facilitate favorable packing at position '*e*', while Leu and Ala are expected to be unfavorable. **(C)** Aligned sequences of '*e*'-series peptides, PLN, and PL5. Leu is fixed at each '*g*' while '*e*' is systematically varied. **(D-F)** SDS-PAGE of peptides under different detergent, temperature, and incubation conditions; the PL5$_{EtG3}$ peptide is abbreviated as Et$_3$. After incubating in the indicated micelle and temperature in the presence of 4 mM of reducing agent Tris(2-carboxyethyl)phosphine (TCEP), 2 µg loaded was loaded. After heating, fresh TCEP was added prior to electrophoresis to 4 mM final. **(G-H)** The pentameric X-ray crystal structure of the mini '*e*'-Val peptide (1.9 Å resolution, gold). The 2F$_o$-F$_c$ electron density map ($\sigma = 1.5$) highlights layers of '*e/g*' **(G)** and '*a/g*' **(H)** packing layers in the central LxxIVxL repeat; Val at '*e*,' blue.
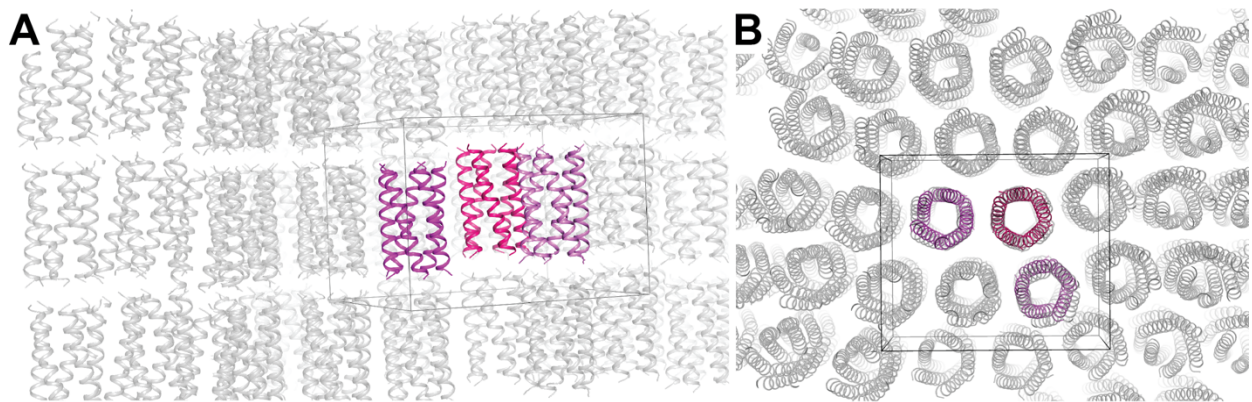
Figure 1.13. Crystal lattice of mini 'e'-Val membrane-soluble peptide, crystal form 1.
(a) P 2_1 2_1 2 space group. Asymmetric unit and unit cell, with three pentamers colored in purple, pink, and magenta, packing in layers. **(b)** Top down packing of the pentamers in the crystal lattice, highlighting the low solvent content.
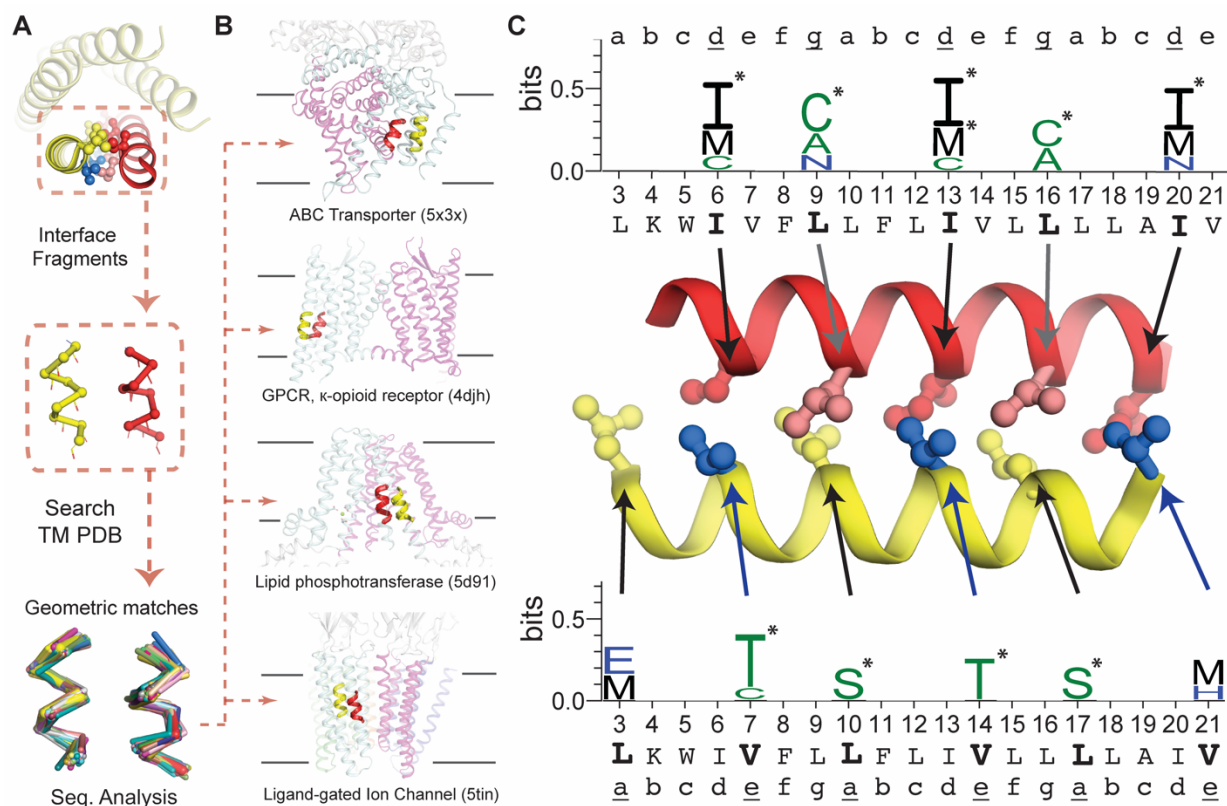
Figure 1.14. Helix-helix packing motifs stabilizing the designed TM bundles are common across the membrane proteome. **(A)** Adjacent helices comprising the repeated helix-helix interface in mini '*e*'-Val were decomposed into successive 18 residue fragments. Backbone atoms were used to search for close structural matches (<0.85 Å RMSD) within a non-redundant database of membrane protein experimental structures. **(B)** The helix-helix geometries in mini '*e*'-Val are found frequently in nature, within membrane proteins of diverse architectures and functions. **(C)** Amino acids enriched in the fragments (p < 0.05) at equivalent position in '*e*'-Val are plotted in Weblogo format (*55*) with asterisks denoting amino acids enriched at >3 standard deviations (p-value < 0.003). Steric bulk of enriched amino acids at '*d*,' '*e*,' and '*g*' are similar to those in '*e*'-Val, and all are consistent with our proposed steric code. A total of nine positions have at least one amino acid enriched with p-values less than 0.003. The binomial probability associated with finding even one position with this level of enrichment is 0.03, while the probability for finding nine positions thusly enriched is $1.1 \cdot 10^{-21}$.
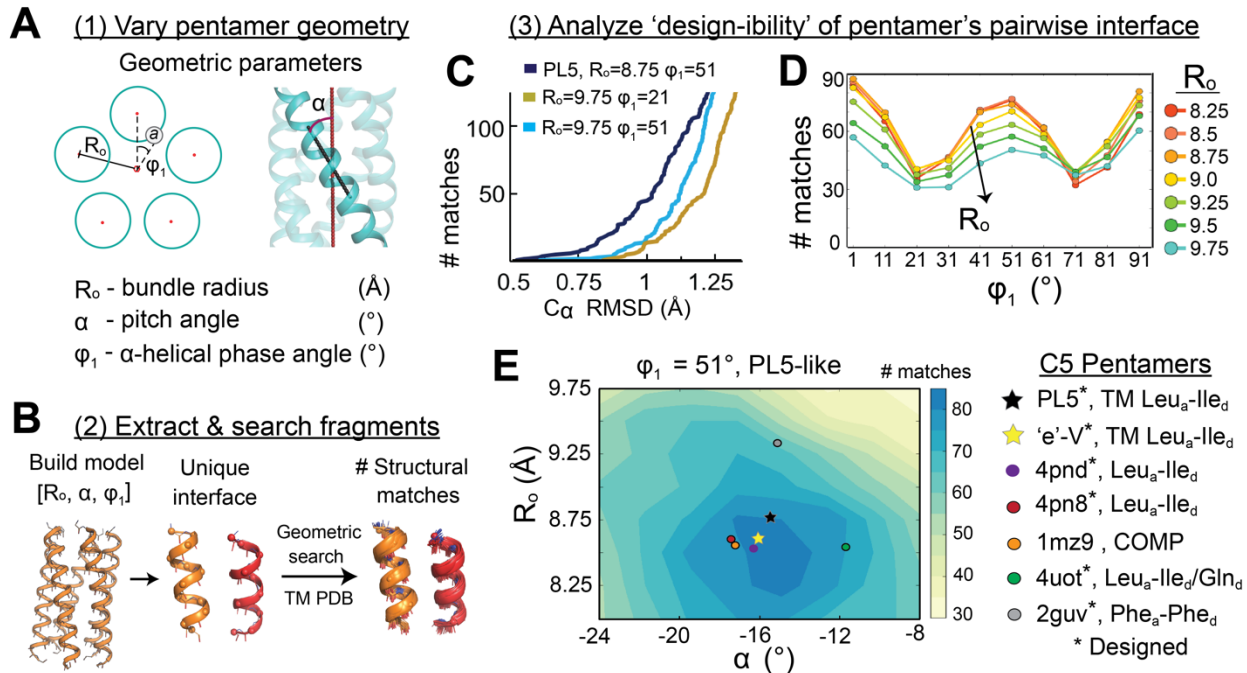
Figure 1.15. Helix-helix geometry observed in the PL5 and *'e'*-Val X-ray structures are near optimal in terms of structural frequency of similar helical pairs seen in membrane proteins in the PDB.

We wished to determine whether the inter-helical geometries in PL5 and *'e'*-Val are similar to those seen in natural MP proteins, and the extent to which systematic changes to the geometry of PL5 and *'e'*-Val would result in fewer matches to the database proteins. To accomplish this, we first fit the structures of PL5 and *'e'*-Val to ideal coiled coils using Crick's parametric equations as in Table 1.1 (panel **A**). We then systematically changed the values of the super-helical radius (R_o), super-helical pitch angle (α), and α-helical phase angle (φ_1) to create a grid of structures with slightly varying inter-subunit helix-helix geometries. Next, we excised a 9-residue helical dimer from each pentamer model (panel **B**) and determined the number of structural matches in the database of natural MPs to this "probe dimer" at a given RMSD cutoff (shown overlain at far right).

The numbers of matches are plotted as a function of the RMSD cutoff in panel **C** for probe dimers that were generated from the native structure of PL5; other models in which the radius or minor helical phase was varied slightly were shown alongside PL5. For example, at 1.1 Å, the maximum and minimum structural frequencies for all the pairs are 111 and 8, respectively. Highly "designable" structures have more hits at each RMSD cutoff. Note that the curves shift to the right as the probe structure deviates from the values seen for PL5. Similarly, panel **D** shows the variation of structural frequency with respect to changes in the super-helical phase and radius, taking the average frequency across the different values of α. The curve shows minima spaced by $360°/7 = 51°$, corresponding to the conversion of a change in register of the coiled coil (e.g., the rotational component required for conversion of an *'a'* to *'d'* (*59*)).

Panel **E** shows the interpolated heat map for the landscape for the number of structural matches of bundles with smoothly varying R_o and α (the other Crick parameters were held fixed at the value obtained for native PL5). This plot reveals a clear peak near $R_o = 8.5$ Å and $\alpha = -16.5$°, matching the value for PL5. We also compare the best-fit Crick parameters for our TM

peptide structures (stars) and previously water-soluble C_5 symmetric bundle structures (circles; Table 1.7). The legend denotes the PDB code and the sequence motif used for those peptides (i.e. LxxIxxx = Leu$_a$-Ile$_d$); COMP - cartilage oligomeric matrix protein. All of the LxxIxxx-containing bundles lie 0.5 Å and 4 ° of the maximum, with the exception of the helical bundle, stabilized by hydrogen-bonds at different layers (green circle, mix of Ile and Gln at '*d*').
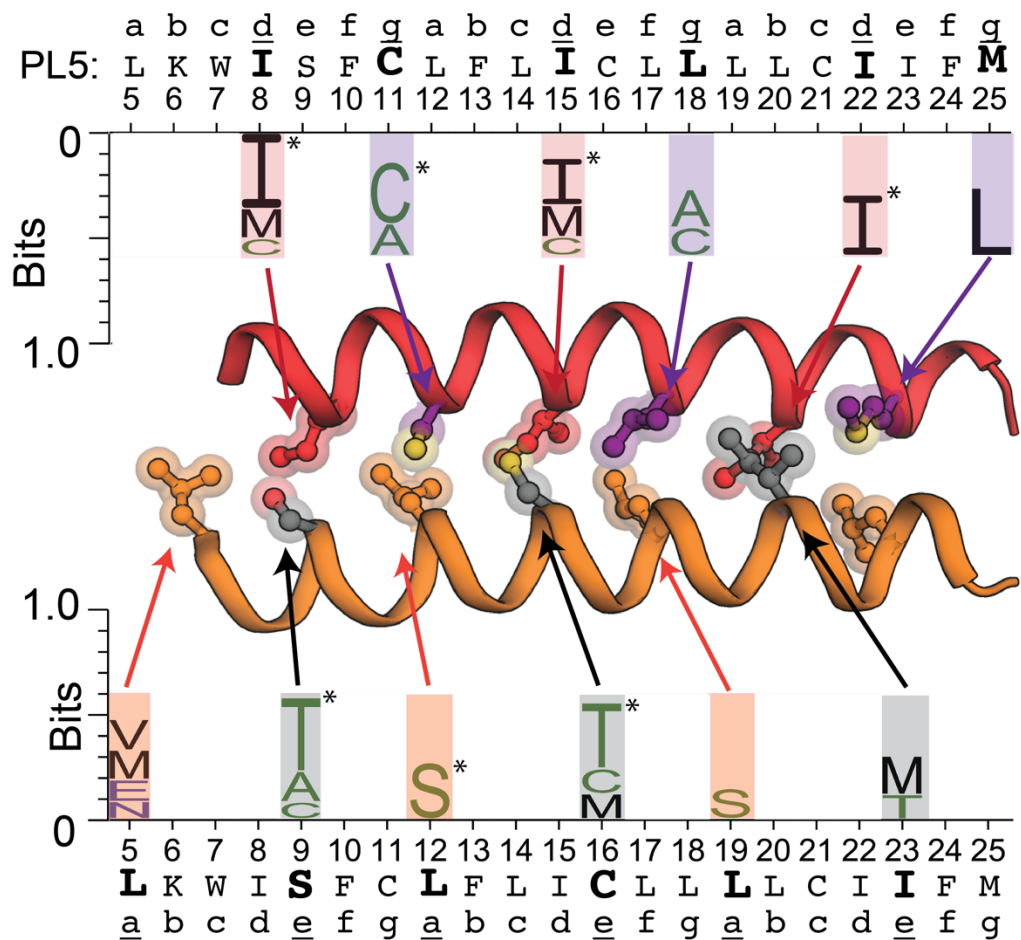
Figure 1.16. Membrane protein database search of helix pair fragments of PL5 interface shows enrichment of amino acids with similar steric patterns.
The analysis performed in Figure 1.14 repeated on the PL5 X-ray crystal structure. Again, Ile is consistently enriched at the '*d*' positions in the database matches. At '*g*', Cys was enriched at equivalent positions in '*e*'-Val, corresponding to Leu amino acids. In PL5, Cys and Ala are enriched at Cys11 and Leu18, while the larger Leu is enriched at subsequent '*g*' position corresponding to the also large Met25 in PL5. At '*e*' positions, Thr is enriched just as in '*e*'-Val, for PL5 residues Ser9 and Cys16. At the wider region of the helix-helix interface, Ile23 at '*e*', the distribution shifts to include the larger Met, retains β-branched Thr, but excludes the smaller Cys and Ala. At the '*a*' position, the amino acids enrichment shows less correlation with the Leu of PL5's LxxIxxx motif; this side-chain points more towards the bundle's central core than the helix-helix interface. Asterisks denote amino acids enriched at >3 standard deviations (p-value < 0.003)
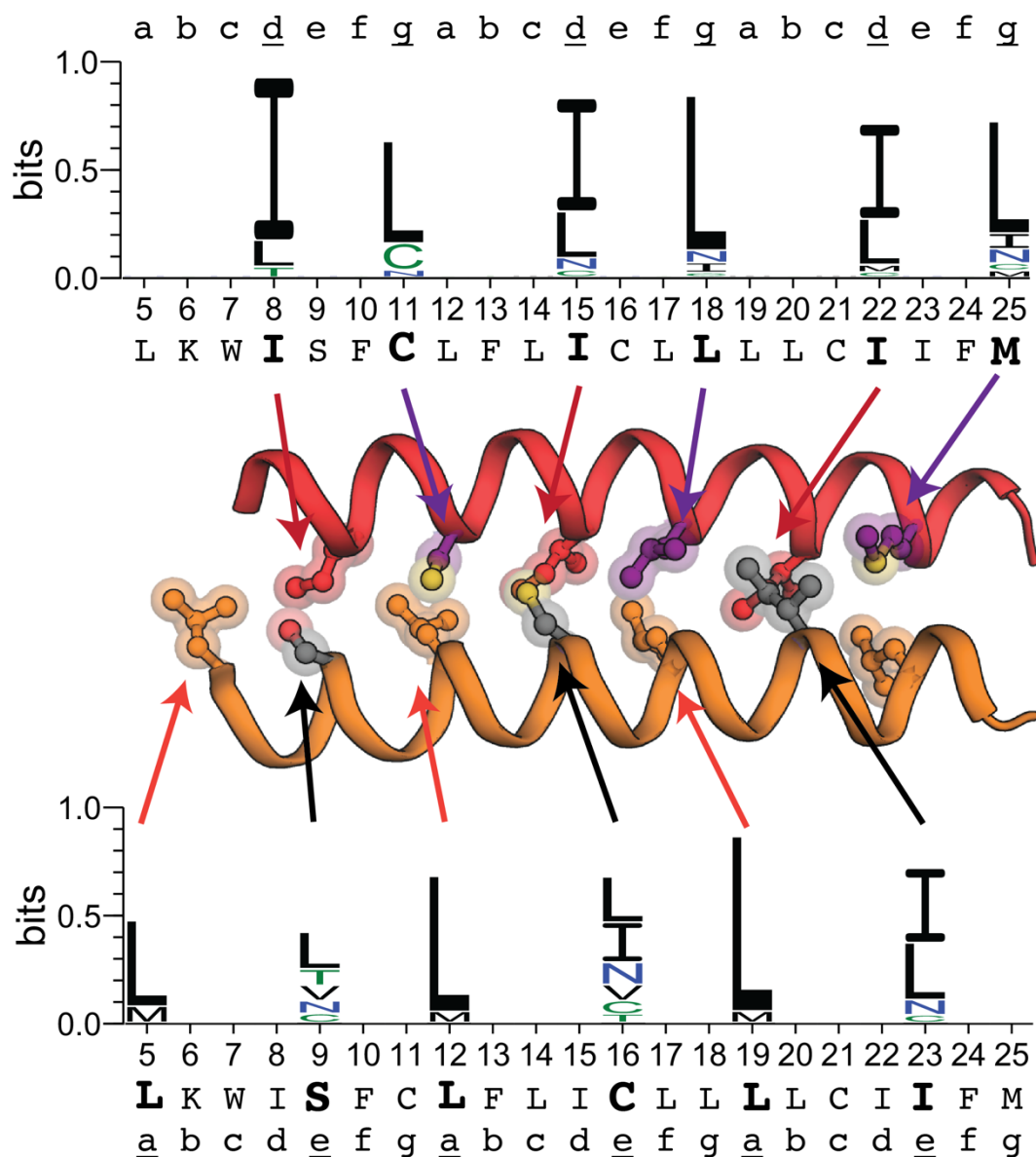
Figure 1.17. Sequence preferences for water-soluble proteins using the same helix-helix geometries as used in PL5 differ from those preferences observed in membrane proteins

The structural fragment based bioinformatic analysis shown in Figure 1.16 is repeated here except performed by searching a database of water-soluble protein structures, showing striking differences. The sequence profile shows interesting similarities and differences when compared to the one generated for membrane proteins (Fig. 1.16). Whereas the enriched amino acids at the 'a' position did not match the motif in the analysis of membrane-soluble proteins, Leu is predominant for water-soluble proteins at 'a'. Instead of small-medium sized residues preferred at 'e' and 'g' interfacial positions (Cys, Ser, Thr, Ala), large Leu and Ile residues are consistently enriched in the database sequences interface.

Table 1.1. Coiled-coil parameters for PL5 and PLN models.
Parameters derived from fitting the carbon-α coordinates of each model to a generalized version of Crick's coiled-coiled parametric equations (*56*). For the MD-derived models, the medoid frame of the major cluster in the simulation was used. The '*e*'-Val X-ray column corresponds to mini '*e*'-Val, crystal form 1. The fits are to a C_5-symmetrical model.

Coiled-coil parameters	PLN 1ZLL	PLN 2KYV	PLN MD	PL5 MD	PL5 X-ray	PL5 Top	PL5 Bottom	'*e*'-Val X-ray
Super-helical radius, R_0 (Å)	9.5	8.4	8.8	8.6	8.7	8.6	9.0	8.6
Super-helical frequency, w_o (°/residue)	-2.4	-1.5	-1.8	-1.9	-2.8	-3.4	-2.2	-2.8
Helical phase, ϕ_1 (°)	-158	-157	-151	-158	-153	-153	-3	-155
Pitch (Å)	226	345	292	272	188	151	241	189
Residue range	37-52	37-52	37-52	5-26	5-26	5-16	17-28	3-24
Fit RMSD (Å)	0.64	0.54	0.59	0.57	0.50	0.35	0.52	0.35
Atoms in fit	80	80	80	110	110	60	60	110
Heptad index of starting residue	*a*	*a*	*a*	*a*	*a*	*a*	f	*a*

Table 1.2. Data collection and refinement statistics

	PL5, C_8E_4 (PDB: 6MQU)	Mini 'e'-Val Form 1, C_8E_4 (PDB: 6MCT)
Data collection		
Space group	$I\,4_1$	$P\,2_1\,2_1\,2$
Cell dimensions		
a, b, c (Å)	124.20, 124.20, 51.94	52.89, 165.63, 41.07
α, β, γ (°)	90.00, 90.00, 90.00	90.00, 90.00, 90.00
Resolution (Å)	3.17 (3.17-3.39)[*]	44.58-1.89 (2.01-1.89)
R_{merge}	0.171 (0.549)	0.108 (1.384)
$I / \sigma I$	6.2 (2.7)	14.49 (1.75)
Completeness (%)	99.0 (98.2)	99.6 (99.8)
Redundancy	3.9 (3.9)	12.6 (12.3)
Refinement		
Resolution (Å)	3.17	1.90
No. reflections	6784	29360
R_{work} / R_{free}	0.2275 / 0.2420	0.1847/0.2281
No. atoms		
Protein	2271	3256
Ligand/ion	0	259
Water	6	23
B-factors		
Protein	46.4	38.2
Ligand/ion	N/A	70.2
Water	35	61.0
R.m.s. deviations		
Bond lengths (Å)	0.009	0.016
Bond angles (°)	1.12	1.24

For both proteins, 1 crystal was used to determine their structure. * The values in parentheses correspond to those for the outmost resolution shells.

Table 1.3. Sequence of recombinant PL5 fusion protein and mutant.
The PL5 transmembrane peptide was expressed in K12 E. coli as a c-terminal fusion to T4-Lysozyme with an N-terminal 6x His-tag, having a thrombin protease cleavage site between the lysozyme and PL5.

DNA sequence of 6xHis-T4-Lysozyme-PL5
ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGCATCTTCGAGATGCGCATATGGGT AACATCTTCGAGATGCTGCGTATCGACGAAGGTCTGCGTCTTAAAATCTACAAGGACACAGAAGGG TACTACACAATCGGAATTGGTCACCTTTTAACCAAGTCTCCCAGCCTTAATGCTGCGAAATCGGAGT TAGACAAGGCTATCGGACGCAACACGAACGGGGTCATCACAAAGGACGAGGCGGAGAAATTATTC AATCAGGATGTCGACGCGGCGGTGCGTGGCATCCTGCGCAACGCTAAGTTGAAGCCAGTATATGAC TCCTTAGACGCCGTACGTCGCGCGGCTCTGATTAACATGGTCTTCCAGATGGGCGAGACGGGTGTAG CTGGGTTTACAAATAGCTTACGTATGTTGCAACAAAAACGTTGGGATGAAGCGGCTGTCAACTTGG CGAAGAGCCGCTGGTACAATCAGACACCCAACCGCGCCAAACGTGTGATTACCACATTCCGTACGG GGACCTGGGATGCCTACGCGGCAGGGGGGCTCAGGGTCTTTGGTTCCACGCGGATCATCCGGAGGGT CCGACCCCGAACAACTTAAGTGGATTTCATTCTGTTTGTTTCTGATTTGTCTGTTGTTATTGTGCATT ATTTTTATGCTGTATCGCGGACGTCGCTAA

Protein sequence of rPL5
GSSGGSDPEQLKWISFCLFLICLLLLCIIFMLYRGRR

DNA sequence of 'Leu18Ile' mutant
- GAC CCC GAA CAA CTT AAG TGG ATT TCA TTC TGT TTG TTT CTG ATT TGT CTG ATC TTA TTG TGC ATT ATT TTTATG CTG TAT CGC GGA CGT CGC TAA

Protein sequence of 'Leu18Ile' mutant
GSSGGS DPEQLKWISFCLFLICL<u>I</u>LLCIIFMLYRGRR

Table 1.4. SDS-PAGE densitometry and apparent free energy of pentamerization in OG.
Densitometry analysis was applied to SDS-PAGE images, Figure 1.12. The mole fraction of monomer, pentamer, and the peptide concentration were used to determine the apparent monomer-pentamer dissociation constants (app. K_{diss}), and the apparent free energy differences of self-association between peptides ($\Delta\Delta G_{app}$) were calculated as described in the methods. Confidence intervals for $\Delta\Delta G_{app}$ were calculated by simulating and propagating 10% errors in gel densitometry integration, reported in parentheses next to each value to an additional significant figure. Pent., Pentamer; Mon., Monomer.

393 K in OG							
	PL5-Et$_3$	'e'-Ile	'e'-Val	'e'-Thr	'e'-Cys	'e'-Leu	'e'-Ala
Pent.	0.389	0.775	0.907	0.932	0.685	< 0.005	< 0.005
Dimer	0.000	0.000	0.000	0.000	0.000	0	0
Mon.	0.611	0.225	0.093	0.068	0.315	> 0.995	> 0.995
App. K_{diss}	2.4 E-10	7.2 E-13	7.6 E-15	1.6 E-15	4.5 E-12	>1.9 E-7	>1.9 E-7

$\Delta\Delta G_{app}$		State 1					
		'e'-Ile	'e'-Val	'e'-Thr		'e'-Leu	'e'-Ala
	LxxIIxL	0.0	3.5 (3.46, 3.62)	4.8 (4.67, 4.87)		<-9.8	<-9.8
	LxxIVxL	-3.5 (-3.54, -3.46)	0.0	1.2 (1.22, 1.25)		<-13.3	<-13.3
State 2	LxxITxL	-4.8 (-4.78, -4.68)	-1.2 (-1.23, -1.22)	0.0		<-14.5	<-14.5
	LxxLLxL	>9.8	>13.3	>14.5		0.0	0.0
	LxxIAxL	>9.8	>13.3	>14.5		0.0	0.0

298 K in OG							
	PL5-Et$_3$	'e'-Ile	'e'-Val	'e'-Thr	'e'-Cys	'e'-Leu	'e'-Ala
Pent.	0.004	0.783	0.882	0.882	0.284	< 0.005	< 0.005
Dimer	0.000	0.006	0.006	0.039	0.000	0.000	0.000
Mon.	0.996	0.211	0.112	0.080	0.716	> 0.995	> 0.995
App. K_{diss}	2.4 E-07	5.3 E-13	2.0 E-14	4.3 E-15	6.6 E-10	>1.90 E-7	>1.90 E-7

$\Delta\Delta G_{app}$		State 1					
		'e'-Ile	'e'-Val	'e'-Thr		'e'-Leu	'e'-Ala
	LxxIIxL	0.0	1.9 (1.88, 1.97)	2.8 (2.78, 2.97)		<-7.6	<-7.6
	LxxIVxL	-1.9 (-1.97, -1.88)	0.0	0.9 (0.90, 0.93)		<-9.5	<-9.5
State 2	LxxITxL	-2.8 (-2.90, -2.78)	-0.9 (-0.93, -0.90)	0.0		<-10.4	<-10.4
	LxxLLxL	>7.6	>9.5	>10.4		0.0	0.0
	LxxIAxL	>7.6	>9.5	>10.4		0.0	0.0

Table 1.5. SDS-PAGE densitometry and apparent free energies of pentamerization in LDS. Densitometry analysis was applied to SDS-PAGE images, Figure 1.12. The mole fraction of monomer, pentamer, and the peptide concentration were used to determine the apparent monomer-pentamer dissociation constants (app. K_{diss}), and the apparent free energy differences of self-association between peptides ($\Delta\Delta G_{app}$) were calculated as described in the methods. Confidence intervals for $\Delta\Delta G_{app}$ were calculated by simulating and propagating 10% errors in gel densitometry integration, reported as a range in parentheses next to each value with an additional significant figure. Pent., Pentamer; Mon., Monomer.

393 K in LDS							
	PL5-Et$_3$	'*e*'-Ile	'*e*'-Val	'*e*'-Thr	'*e*'-Cys	'*e*'-Leu	'*e*'-Ala
Pent.	0.007	0.010	0.153	0.404	0.160	< 0.005	< 0.005
Dimer	0.000	0.000	0.000	0.000	0.000	0	0
Mon.	0.993	0.990	0.847	0.596	0.840	> 0.995	> 0.995
App. K_{diss}	1.5 E-07	8.9 E-08	2.9 E-09	1.9 E-10	2.6 E-09	>1.9 E-7	>1.9 E-7

$\Delta\Delta G_{app}$		State 1				
		'*e*'-Ile	'*e*'-Val	'*e*'-Thr	'*e*'-Leu	'*e*'-Ala
State 2	LxxIIxL	0.0	2.7 (2.60, 2.78)	4.8 (4.58, 5.07)	<-0.6	<-0.6
	LxxIVxL	-2.7 (-2.78, -2.60)	0.0	2.1 (1.97, 2.29)	<-3.3	<-3.3
	LxxITxL	-4.8 (-5.07, -4.58)	-2.1 (-2.29, -1.97)	0.0	<-5.4	<-5.4
	LxxLLxL	>0.6	>3.3	>5.4	0.0	0.0
	LxxIAxL	>0.6	>3.3	>5.4	0.0	0.0

298 K in LDS							
	PL5-Et$_3$	'*e*'-Ile	'*e*'-Val	'*e*'-Thr	'*e*'-Cys	'*e*'-Leu	'*e*'-Ala
Pent.	0.122	0.276	0.369	0.721	0.366	< 0.005	< 0.005
Dimer	0.000	0.000	0.000	0.000	0.000	0	0
Mon.	0.878	0.724	0.631	0.279	0.634	> 0.995	> 0.995
App. K_{diss}	4.7 E-9	7.0 E-10	2.7 E-10	2.4 E-12	2.8 E-10	>1.9 E-7	>1.9 E-7

$\Delta\Delta G_{app}$		State 1				
		'*e*'-Ile	'*e*'-Val	'*e*'-Thr	'*e*'-Leu	'*e*'-Ala
State 2	LxxIIxL	0.0	0.6 (0.52, 0.61)	3.4 (3.16, 3.58)	<-3.3	<-3.3
	LxxIVxL	-0.6 (-0.61, -0.52)	0.0	2.8 (2.64, 2.97)	<-3.9	<-3.9
	LxxITxL	-3.4 (-3.58, -3.16)	-2.8 (-2.97, -2.64)	0.0	<-6.7	<-6.7
	LxxLLxL	>3.3	>3.9	>6.7	0.0	0.0
	LxxIAxL	>3.3	>3.9	>6.7	0.0	0.0

Table 1.6. Data collection and refinement statistics for alternative crystal forms of Mini '*e*'-Val

For both proteins, 1 crystal was used to determine their structure. * The values in parentheses correspond to those for the outmost resolution shells. Crystal form 1 for Mini '*e*'-Val is in Table 1.2.

	Mini '*e*'-Val Form 2, C_8E_4 (PDB: 6MQ2)	Mini '*e*'-Val Form 3, C_8E_4 (PDB: 6MPW)
Data collection		
Space group	C 2 2 2_1	C 2 2 2_1
Cell dimensions		
a, b, c (Å)	55.55, 81.89, 47.84	55.58, 88.802, 48.207
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Resolution (Å)	50.0 - 2.38 (2.53 - 2.38)*	50.0 - 2.48 (2.63 - 2.48)
R_{merge}	0.15 (1.731)	0.067 (0.58)
$I / \sigma I$	9.45 (0.84)	19.71 (4.84)
Completeness (%)	92.7 (53.5)	99.0 (94.1)
Redundancy	11.68 (3.15)	12.3 (11.9)
Refinement		
Resolution (Å)	2.50	2.50
No. reflections	3957	4311
R_{work} / R_{free}	0.2129 / 0.2341	0.2116 / 0.2324
No. atoms		
Protein	1031	1056
Ligand/ion	49	79
Water	6	0
B-factors		
Protein	66.15	87.38
Ligand/ion	78.51	105.09
Water	72.86	N/A
R.m.s. deviations		
Bond lengths (Å)	0.003	0.002
Bond angles (°)	0.49	0.39

Table 1.7. Structural comparisons for pentameric helical bundles.
The X-ray structures described here are structurally similar to previous solved LxxIxxx-containing water-soluble symmetric pentameric helical bundles ('cc-Pent1' and 'cc-Pent2' from ref (*35*)). When mathematically described as parametric coiled-coil structures, in Figure 1.15, the coiled-coil parameters for these structures are all similar. The backbone atom RMSD was calculated and reported below for the optimal superposition of mini 'e'-Val (crystal form 1) and PL5 onto the water-soluble 'cc-Pent1' and 'cc-Pent2,' as well as on each other. For mini 'e'-Val (form 1), 'P1,' 'P2,' and 'P3' denote each of the three pentamers in the crystallographic asymmetric unit.

ID (A)	Residues	ID (B)	Residues	RMSD
Mini 'e'-Val P1	3-24	cc-Pent1 (pdb, 4PND)	4-25	0.44
PL5	5-26	cc-Pent1 (pdb, 4PND)	4-25	0.66
Mini 'e'-Val P1	3-24	cc-Pent2 (pdb, 4PN8)	4-25	0.48
PL5	5-26	cc-Pent2 (pdb, 4PN8)	4-25	0.60
Mini 'e'-Val P1	3-24	Mini 'e'-Val P2	3-24	0.30
Mini 'e'-Val P1	3-24	Mini 'e'-Val P3	3-24	0.31
Mini 'e'-Val P2	3-24	Mini 'e'-Val P3	3-24	0.39
Mini 'e'-Val P2	5-26	PL5	5-26	0.60

Table 1.8. Transmembrane protein database used for fragment-based searching.
PDB IDs of structures culled at 40% sequence identity between all TM segments of compared structures, all having <3.2 Å resolution for X-ray, neutron, or electron diffraction structures and <3.0 Å cryo-EM structures.   Distribution of amino acids from this database is also reported. Downloaded Nov 2017. The TM positions are defined as in the Orientations of Membrane Proteins database (*57*).

2Z73	4X5M	4J05	4ZR1	5AX0	3RLB
4KPP	5SYT	5N6M	3UX4	4YL3	5U1X
4BBJ	5XAP	4WD8	1KQF	3VR8	4W6V
4JKV	5DSG	5EDL	5LWE	2XQ2	2A65
4PL0	5Y79	5UHQ	2WSW	2R9R	1E12
2QKS	5I32	5DIR	5HVX	5D91	3WMG
2BHW	4MYC	4DW1	3M73	4TQ4	3TIJ
1OTS	5XSZ	2F2B	4XTL	1U7G	4V1F
3V5U	5O0T	5CBG	1LDF	5OGL	4RDQ
5BZ3	3D31	4GC0	4K5Y	3V2Y	4D2E
5SVK	5WUF	3WXV	4CAD	5GLI	2CFQ
4Z3N	5WUD	4YMK	5UIW	5XPD	4XES
4XU4	1PW4	3LLQ	5VRG	4O6M	4FC4
5AZB	5MKK	3TLW	4R0C	4O6Y	3VW7
3C02	5KKZ	4YMU	5EE7	4YAY	5I20
3SYA	1U19	1KF6	4O9P	2HYD	4RYM
5JKI	5JJE	4U4W	1OKC	4UMW	3L1L
5TSA	4RI2	2J58	5A8E	3UG9	5V8K
4F4S	4K1C	4S0V	2XTV	4TPJ	5IWS
2UUH	4QND	4XNV	4XT1	4DVE	5KBW
4OO9	3RQW	1Q16	1ORS	5XLS	4Z7F
5SV0	5XJJ	3TUI	4JQ6	3N5K	5IWK
4ZW9	2BL2	5ER7	4RP9	5LEV	5TIN
1R3J	4PHU	2DYR	4H33	1M0L	3KLY
1YQ3	5EH4	3DDL	5EIK	4PGS	5U09
5XW6	5MLZ	2BS2	1XIO	4DJH	3GIA
4I9W	3WVF	5AJI	3LDC	4DJI	4UC1
2JLN	5CTG	3WDO	5AEZ	5DQQ	5DA0
2X2V	4EIY	5W3S	5JSI	3B9Y	
4LP8	5WIV	5JWY	2QTS	4TKR	
4AYT	5X3X	5T77	5JSZ	4HYJ	
5JNQ	4WIS	4ZP0	5IRX	5F15	
4QKM	4MND	4I0U	5AYN	3TX3	
4IKV	3H90	4DX5	4QUV	5GUF	
2W2E	2WCD	3OUF	2XQU	4PHZ	

AA	Count	Fraction
A	4195	0.1102
C	535	0.0141
E	677	0.0178
D	535	0.0141
G	2862	0.0752
F	3137	0.0824
I	3834	0.1007
H	390	0.0102
K	771	0.0203
M	1303	0.0342
L	5804	0.1525
N	857	0.0225
Q	640	0.0168
P	1133	0.0298
S	2069	0.0544
R	975	0.0256
T	1976	0.0519
W	985	0.0259
V	3823	0.1005
Y	1557	0.0409
total	38058	

Table 1.9. Observed amino acid frequencies and enrichment in helix-helix pairs with high structural similarity (RMSD ≤ 0.85 Å) to 'e'-Val

Successive 9-residue stretches of the 'e'-Val helix-helix interface were matched to structurally similar helical pairs within our MP database. For interfacial residues, we analyzed the amino acid distributions at the equivalent position in the matching database sequences as described in the methods and plotted in Figure 1.14. The counts ('#aa') for each amino acid ('aa') are listed for each position, alongside the propensity (number observed/expected) for those amino acids relative to the background amino acid distribution in MPs listed in Table 1.8. The p-value from a one-tailed binomial test is reported, expressing the statistical significance that observed amino acid frequency at that position is enriched (for propensities > 1) or under-represented (for propensities < 1) relative to the expected background frequency.

Residue 3 - LEU, 'a'				Residue 6 - ILE, 'd'				Residue 7 - VAL, 'e'			
aa	#aa	Propensity	p-value	aa	#aa	Propensity	p-value	aa	#aa	Propensity	p-value
A	2	0.36	0.0757	V	5	0.62	0.1732	G	1	0.15	0.0084
L	6	0.79	0.3432	A	6	0.68	0.2082	F	5	0.69	0.2583
G	2	0.53	0.2641	G	2	0.33	0.0546	P	1	0.38	0.2584
F	3	0.73	0.4012	F	4	0.61	0.2016	R	1	0.44	0.338
W	1	0.77	0.6272	L	10	0.82	0.308	N	1	0.51	0.4084
N	1	0.89	0.6894	Y	1	0.31	0.1562	M	3	1	0.6451
Y	2	0.98	0.6643	R	1	0.49	0.3896	E	0	0	0.2059
K	1	0.99	0.7302	K	1	0.62	0.5152	D	0	0	0.3024
C	0	0	0.4916	W	2	0.97	0.657	K	0	0	0.1645
D	0	0	0.5068	E	0	0	0.2377	Q	0	0	0.2252
H	0	0	0.5989	D	0	0	0.3371	W	0	0	0.0993
Q	0	0	0.4286	Q	0	0	0.2578	Y	0	0	0.0254
P	0	0	0.2203	P	0	0	0.0889	I	9	1.02	0.5321
R	0	0	0.2734	H	1	1.23	0.5597	S	5	1.04	0.5254
T	0	0	0.0696	N	3	1.67	0.2688	V	10	1.13	0.3908
V	8	1.59	0.1247	S	6	1.38	0.2689	H	2	2.23	0.2264
S	6	2.21	0.0532	T	7	1.69	0.1215	L	15	1.12	0.3629
I	9	1.79	0.06	C	4	3.55	0.0268	C	4	3.22	0.0363
M	5	2.92	0.0277	M	8	2.92	0.0061	A	15	1.55	0.0574
E	4	4.49	0.0121	I	19	2.36	0.0003	T	16	3.5	<0.0001

Residue 9 - LEU, 'g'				Residue 10 - LEU, 'a'				Residue 13 - ILE, 'd'			
aa	#aa	Propensity	p-value	aa	#aa	Propensity	p-value	aa	#aa	Propensity	p-value
V	5	0.65	0.2019	G	4	0.58	0.1771	V	5	0.6	0.1479
G	2	0.35	0.0647	F	5	0.67	0.2299	G	2	0.32	0.046
I	6	0.77	0.3322	L	12	0.86	0.3548	F	4	0.58	0.1764
Y	1	0.32	0.1719	P	1	0.37	0.2419	A	7	0.77	0.2927
M	2	0.76	0.5073	W	1	0.42	0.314	L	11	0.87	0.374
F	6	0.95	0.5488	T	4	0.85	0.4868	Y	1	0.29	0.1418
D	1	0.96	0.7211	Y	3	0.81	0.4865	P	1	0.4	0.2882
H	0	0	0.4541	D	0	0	0.2903	W	1	0.47	0.3632
K	0	0	0.2061	H	0	0	0.3934	K	1	0.59	0.4958
Q	0	0	0.2713	K	0	0	0.1547	E	0	0	0.2252
R	0	0	0.1358	N	0	0	0.1261	D	0	0	0.3236
W	0	0	0.1326	Q	0	0	0.214	Q	0	0	0.2451
T	4	1	0.6296	R	0	0	0.0944	R	0	0	0.1162

	Residue 9 - LEU, *'g'*				Residue 10 - LEU, *'a'*				Residue 13 - ILE, *'d'*		
aa	#aa	Propensity	p-value	aa	#aa	Propensity	p-value	aa	#aa	Propensity	p-value
E	2	1.46	0.3991	C	2	1.56	0.3679	H	1	1.18	0.573
S	5	1.19	0.409	E	3	1.85	0.2209	S	5	1.11	0.4735
L	13	1.11	0.3918	I	11	1.2	0.308	N	4	2.14	0.1176
P	4	1.74	0.1974	M	5	1.61	0.2012	T	7	1.62	0.1397
N	5	2.89	0.03	A	13	1.3	0.1998	C	4	3.42	0.0302
A	14	1.65	0.0408	V	14	1.53	0.0706	M	9	3.17	0.0021
C	7	6.45	0.0001	S	13	2.63	0.0013	I	20	2.39	0.0002

	Residue 14 - VAL, *'e'*				Residue 16 - LEU, *'g'*				Residue 17 - LEU, *'a'*		
aa	#aa	Propensity	p-value	aa	#aa	Propensity	p-value	aa	#aa	Propensity	p-value
F	5	0.74	0.3225	G	2	0.39	0.1005	F	4	0.58	0.1764
P	1	0.41	0.2944	F	4	0.7	0.3186	G	5	0.8	0.4002
R	1	0.48	0.3762	Y	1	0.35	0.221	P	1	0.4	0.2882
L	12	0.96	0.5137	V	6	0.87	0.452	W	1	0.47	0.3632
S	4	0.9	0.5371	M	1	0.42	0.312	Y	3	0.88	0.5577
V	8	0.97	0.5573	S	3	0.8	0.4787	V	8	0.96	0.5425
E	0	0	0.2293	D	0	0	0.3915	T	4	0.93	0.5674
D	0	0	0.3281	H	0	0	0.4929	D	0	0	0.3236
G	0	0	0.0016	K	0	0	0.2429	H	0	0	0.427
K	0	0	0.1861	Q	0	0	0.3107	K	0	0	0.1823
Q	0	0	0.2492	R	0	0	0.1671	N	0	0	0.1512
W	0	0	0.1163	W	0	0	0.1635	Q	0	0	0.2451
Y	0	0	0.0326	I	7	1.01	0.5502	R	0	0	0.1162
N	2	1.08	0.5532	E	2	1.63	0.3483	E	2	1.35	0.436
H	2	2.39	0.204	P	3	1.46	0.3387	I	9	1.08	0.4599
C	3	2.59	0.1099	T	5	1.4	0.2877	L	14	1.11	0.3859
I	11	1.33	0.2004	N	3	1.93	0.2031	A	11	1.2	0.3052
M	6	2.14	0.0621	L	14	1.33	0.1588	C	3	2.56	0.1129
A	14	1.55	0.064	A	13	1.71	0.037	M	6	2.11	0.0651
T	13	3.05	0.0003	C	5	5.14	0.003	S	12	2.66	0.0018

	Residue 20 - ILE, *'d'*				Residue 21 - VAL, *'e'*		
aa	#aa	Propensity	p-value	aa	#aa	Propensity	p-value
V	5	0.62	0.1732	G	1	0.19	0.03
G	1	0.17	0.0144	S	2	0.53	0.2684
A	7	0.79	0.3329	F	5	0.88	0.4924
S	3	0.69	0.3612	R	1	0.57	0.4699
Y	2	0.61	0.3597	E	0	0	0.2896
R	1	0.49	0.3896	D	0	0	0.3915
K	1	0.62	0.5152	K	0	0	0.2429
L	12	0.98	0.5514	Q	0	0	0.3107
D	1	0.93	0.7062	P	0	0	0.124
E	0	0	0.2377	W	0	0	0.1635
Q	0	0	0.2578	Y	0	0	0.0561
P	0	0	0.0889	I	7	1.01	0.5502
W	0	0	0.1225	V	7	1.01	0.548
H	1	1.23	0.5597	C	2	2.06	0.2542
F	7	1.06	0.4919	N	3	1.93	0.2031
C	2	1.77	0.3115	L	13	1.24	0.247

Residue 20 - ILE, *'d'*				Residue 21 - VAL, *'e'*			
aa	#aa	Propensity	p-value	aa	#aa	Propensity	p-value
V	5	0.62	0.1732	G	1	0.19	0.03
G	1	0.17	0.0144	S	2	0.53	0.2684
A	7	0.79	0.3329	F	5	0.88	0.4924
S	3	0.69	0.3612	R	1	0.57	0.4699
Y	2	0.61	0.3597	E	0	0	0.2896
R	1	0.49	0.3896	D	0	0	0.3915
K	1	0.62	0.5152	K	0	0	0.2429
L	12	0.98	0.5514	Q	0	0	0.3107
D	1	0.93	0.7062	P	0	0	0.124
E	0	0	0.2377	W	0	0	0.1635
Q	0	0	0.2578	Y	0	0	0.0561
P	0	0	0.0889	I	7	1.01	0.5502
W	0	0	0.1225	V	7	1.01	0.548
H	1	1.23	0.5597	C	2	2.06	0.2542
F	7	1.06	0.4919	N	3	1.93	0.2031
C	2	1.77	0.3115	L	13	1.24	0.247

Chapter 2

Unique Transmembrane Domain Interactions Differentially Modulate Integrin αvβ3 and αIIbβ3 Function[†]

Unique Transmembrane Domain Interactions Differentially Modulate Integrin αvβ3 and αIIbβ3 Function

Abstract

Lateral transmembrane (TM) helix-helix interactions between single-span membrane proteins play an important role in the assembly and signaling of many cell-surface receptors. Often these helices contain two highly conserved yet distinct interaction motifs, arranged such that the motifs cannot be engaged simultaneously. However, there is sparse experimental evidence that dual-engagement mechanisms play a role in biological signaling. Here, we investigate the function of the two conserved interaction motifs in the TM domain of the integrin β3 subunit. The first motif uses reciprocating "large-large-small" amino acid packing to mediate the interaction of the β3 and αIIb TM domains and maintain the inactive resting conformation of the platelet integrin αIIbβ3. The second motif, $S-X_3-A-X_3-I$, is a variant of the classical "$G-X_3-G$" motif. Using site-directed mutagenesis, optical trap-based force spectroscopy, and molecular modeling, we show that $S-X_3-A-X_3-I$ does not engage αIIb, but rather mediates the interaction of the β3 TM domain with the TM domain of the αv subunit of the integrin αvβ3. Like αIIbβ3, αvβ3 on circulating platelets is inactive and in the absence of platelet stimulation is unable to interact with components of the subendothelial matrix. However, disrupting any residue in the β3 $S-X_3-A-X_3-I$ motif by site-directed mutations is sufficient to induce αvβ3 binding to the αvβ3 ligand osteopontin and to the monoclonal antibody WOW-1. Thus, the β3 integrin TM domain is able to engage in two mutually-exclusive interactions that produces alternate α-subunit pairing, creating two integrins with distinct biological functions.

Introduction

Lateral transmembrane (TM) helix-helix interactions are central to the regulation of many proteins that transmit signals across membranes including integrins (*1*). As a relevant example, a heterodimer composed by the TM helices of its αIIb and β3 subunits stabilizes the integrin αIIbβ3 on circulating platelets in its resting inactive conformation, thereby preventing the spontaneous formation of platelet aggregates in the circulation and insuring that αIIbβ3 is only activated at sites of vascular injury (*2*). Molecular modeling has provided insight into the structure of the αIIbβ3 TM heterodimer (*3*). Initial models were built by combining rigid body sampling and physics-based energy minimization subject to numerous restraints derived from biochemical experiments (*4, 5*). The resulting structural models were in remarkably good agreement with subsequent NMR structures of the TM domain heterodimer, highlighting the power of data-driven computer models of TM complexes in the absence of high-resolution structural information (*6, 7*).

The interface of the αIIbβ3 TM domain heterodimer consists of a tightly packed structure in which small and large residues interdigitate by efficient van der Waals packing along the heterodimer interface. Thus, a sequence motif in the αIIb TM domain, $G\text{-}x_3\text{-}G\text{-}x_3\text{-}L$, packs in a reciprocal manner with the β3 TM domain sequence $V\text{-}x_3\text{-}I\text{-}x_3\text{-}G$ such that bulky residues from one TM helix contact a hole formed by a small Gly residue on the neighboring helix (*8*) (Fig. 2.1).

Platelets contain a second β3 integrin, αvβ3. Like αIIbβ3, αvβ3 on resting platelets is inactive until platelets are stimulated, after which it is able to bind to immobilized ligands such as the extravascular matrix proteins osteopontin (OPN) and vitronectin (*9*). Although αv and αIIb are homologous proteins, they share only 36.1% sequence identity (*10*). Nonetheless, compared to the αIIb TM domain motif interfacing with β3, the homologous motif on αv contains only one

conservative Gly-to-Ala substitution (A-x$_3$-G-x$_3$-L) (*8*). On the other hand, most integrin β subunits, including β3, also contain a highly conserved small residue-x$_3$-small residue motif (*8*) (Fig. 2.1, Table 2.1) – a motif in which close interactions between small residues (Gly, Ala, and Ser) between two TM domains stabilize dimeric TM complexes (*11*). Yet, the small residue-x$_3$-small residue motif, S-x$_3$-A, on the β3 TM helix is located on the face of the helix opposite to the interface of the αIIbβ3 TM heterodimer (*12*) (Fig. 2.1).

Platelets circulate in a milieu containing a high concentration of the principal αIIbβ3 ligand fibrinogen (*13*). Thus, αIIbβ3 function is tightly regulated (*14*). However, given the specialized nature of αIIbβ3 function, we asked whether the motifs mediating the interaction of the β3 TM domain with αv and αIIb are necessarily the same. To address this question, we scanned the TM helix of intact β3 with leucine and alanine replacements and used optical trap-based force spectroscopy to identify replacements that caused constitutive αvβ3 binding to immobilized OPN, a physiological αvβ3 ligand (*15*) and to the activation-dependent monovalent αvβ3-specific monoclonal antibody WOW-1 (*16*). In contrast to αIIbβ3, we found a clear shift in the helical register and periodicity of the αvβ3 TM heterodimer such that the αv and β3 helices both interacted via small-x$_3$-small residue motifs. Subsequent unrestrained molecular dynamics simulations revealed that this helical interface was stable and in striking agreement with our functional results. Thus, the β3 integrin TM domain employs two distinct sequences that mediate mutually-exclusive interactions producing alternate α-subunit pairing, creating two integrins with distinct biological functions.

Results

Effect of β3 TM domain mutations on OPN binding to αvβ3

For integrin αIIbβ3, we previously showed that perturbing side-chain van der Waals packing at the TM domain interface shifts the protein from its resting inactive conformation to its active conformation, the latter having high ligand binding affinity (*17, 18*). Scanning mutagenesis identified the β3 TM domain V-x_3-I-x_3-G motif responsible for stabilizing αIIbβ3 in its inactive state (*19*). Here, we similarly interrogate the second platelet integrin αvβ3 via mutagenesis of the β3 TM domain in conjunction with optical trap-based force spectroscopy to probe the activation state of the αvβ3 mutants (*17, 20*).

Figure 2.2 illustrates optical trap-based force spectroscopy of OPN-coated latex beads with CHO cells expressing wild-type (WT) αvβ3 in the resting state. The cumulative probability of detecting rupture forces >20 pN was calculated, normalized for the level of αvβ3 expression. Under these conditions, the vast majority of WT αvβ3 expressed on the CHO cell surface was in its resting state, given a cumulative probability of 1.7±0.8% (Fig. 2.2). However, there was a small peak of rupture force at 55-70 pN, indicating that some of the WT αvβ3 was able to constitutively interact with the OPN-coated beads. Next, we repeated the measurement after addition of 1 mM Mn^{2+} to the suspension medium, which is known to shift the equilibrium to favor the high-affinity conformation. There was a striking increase in the force spectrum, with a peak in the range of 55-65 pN and with detectable rupture forces as large as 100 pN (Fig. 2.2). Moreover, the cumulative probability of rupture forces >20 pN significantly increased to 9.1 ± 4.3% when compared to WT αvβ3 in the absence of Mn^{2+} (p<0.001). Furthermore, EDTA is known to eliminate the adhesion of proteins to extracellular ligands, as shown previously for αvβ3 binding to OPN (*21*). Indeed the

force spectrum in the presence of EDTA is similar to that of WT $\alpha v\beta 3$ alone, with the vast majority of the measured rupture forces again in the range of 0-40 pN and no statistical difference in the cumulative binding probability for forces >20 pN (p>0.05). In summary, these measurements demonstrate that the vast majority of the WT $\alpha v\beta 3$ expressed on the surface of transfected CHO cells is in its resting inactive conformation, and that under appropriate conditions the inactive $\alpha v\beta 3$ can be shifted to its active ligand binding conformation.

We next performed these force spectroscopy measurements on TM domain mutants of $\beta 3$. We first highlight the $\beta 3$ mutant S669L, in which the first residue in the small residue-x_3-small residue motif is replaced by a bulky leucine residue, expected to disrupt the $\alpha v\beta 3$ TM complex. The S699L mutant displayed strong constitutive OPN binding, having a rupture force spectrum essentially identical to that of WT $\alpha v\beta 3$ activated by 1 mM Mn^{2+} (Fig. 2.2). The range of rupture forces for S699L extended to 95 pN with a peak at 60-75 pN and a cumulative probability of rupture forces > 20 pN of 9.9 ± 4.2 % (Fig. 2.3). Addition of 1 mM Mn^{2+} to S699L induced only a modest increase in the cumulative binding probability for forces > 20 pN, 11.3 ± 3.8 %, while addition of 5 mM EDTA caused substantial inhibition of rupture forces in the range of 30-55 pN and complete abrogation of those >55 pN (Fig 2.3). Thus, increases in the rupture force spectrum can be attributed to specific binding between mutant $\alpha v\beta 3$ and OPN. Moreover, S699L causes near-maximal constitutive activation of $\alpha v\beta 3$.

The results of scanning the entire $\beta 3$ TM domain with alanine or leucine replacements on $\alpha v\beta 3$ binding to OPN are shown in Figure 2.4 and Table 2.2. They reveal that mutating either S699 or A703 comprising the small residue-x_3-small residue motif in the $\beta 3$ TM domain causes a high degree of constitutive activation and OPN binding. By contrast, mutating these same residues

did not cause constitutive fibrinogen binding to αIIbβ3 (*19*). To directly compare the differences in how β3 mutants disrupt the TM domain interfaces of αvβ3 and αIIbβ3, we plotted the fractional activation of αvβ3 (OPN binding) versus the fractional activation of αIIbβ3 (fibrinogen binding) caused by mutating β3 TM domain residues 697-705 in Figure 2.5. The curves for constitutive αvβ3 and αIIbβ3 activation are completely out of phase, indicating that the residues critical for the interaction of αIIb versus αv with β3 reside on opposites sides of the β3 helix (Fig. 2.5). Furthermore, besides the high degree of αvβ3 activation caused by S699L and A703L, G702L and I707L are also highly activating. G702 and I707 are both located on same face of the β3 helix as the ^{699}S-x$_3$-A^{703} motif, with G702 adjacent to A703 and I707 one full α-helical turn down (Fig. 2.4). The β3 residues whose mutations cause strong constitutive αvβ3 activation (S699, G702, A703, I707) are clearly on a different helical register from those residues at the interface of the αIIbβ3 NMR structures (V700, M701, I704, G708) (Fig. 2.5). Thus, these mutants define a contiguous interaction interface on β3 with αv, distinct from the αIIbβ3 interface.

To confirm these results, we repeated the optical trap-based force spectroscopy measurements using the activation-dependent monovalent monoclonal antibody WOW-1 as a second αvβ3 ligand (*16*). As shown in Figure 2.4 and Table 2.2, force spectroscopy measurements using WOW-1 as the αvβ3 ligand were generally in agreement with those using OPN. As with OPN, large effects were seen for the S699L, A703L, I707L mutants. Also, the significant effect of subtle isomeric replacement of Ile707 to Leu resulted in activation of αvβ3, demonstrating that the native van der Waals packing interactions are highly stereochemically-specific. One distinction of WOW-1 binding was that mutating the residues V700, I704, and G708 that neighbor S699, A703, and I707, respectively, was more activating than detected for OPN. This might reflect a higher affinity of

WOW-1 for αvβ3 such that marginally activated mutants were detected more easily. By contrast, replacing G702 with leucine was less activating when WOW-1 was the αvβ3 ligand. Thus, it is possible that WOW-1 binds to a different range of activated αvβ3 conformations than the physiological αvβ3 ligand OPN. Despite these differences, the overall trend is similar between mutants binding to OPN and WOW-1. These results confirm that the conserved β3 S-X$_3$-A-X$_3$-I motif is critical for gating αvβ3 activation, and that the TM domain surface on β3 used to bind αv is quite distinct from that used in αIIbβ3.

Modeling and molecular dynamics simulation of αvβ3

We next sought to gain a better structural understanding of the αvβ3 TM domain interface. We began by comparing the helical register implied from our scanning mutagenesis of αvβ3 to the interfaces calculated for two independently published solution NMR structures of aIIbB3 (*6, 7*). First, we analyzed the per-residue inter-helical distances between each residue in β3 and the closest residues in αIIb by Cα atom for the lowest energy model reported in each NMR study. To make comparisons to the per-residue change associated with αvβ3 mutant binding to OPN, the inter-helical distances were converted to a Normalized Interhelical Closeness (NIC) value, as described in the Materials and Methods, where NIC = 1 for the closest β3 residue to the αIIb backbone and NIC = 0 for the furthest β3 residue (Fig. 2.6). When the NIC values for the two αIIbβ3 NMR structures were plotted alongside the Fractional Activation values for αvβ3 binding to OPN shown in Figure 2.4, it is clear that mutating residues in the β3 TM helix that are closest to the αIIb helix would have no effect on the activation state of avβ3 (Fig. 2.6). Further, this

analysis indicates that any attempt to model the avβ3 TM dimer from a αIIbβ3 NMR structure would require large-scale conformational rearrangements.

Next, we sought to build a model for all-atom molecular dynamics simulations to sample the energetic landscape of the αvβ3 TM and cytoplasmic domains. An initial model was built by threading the sequences of αv and β3 onto the coordinates of the first conformation of the reported αIIbβ3 NMR model (pdb: 2knc). This model was first structurally relaxed in an implicit lipid bilayer using the Rosetta modeling suite, applying rounds of side-chain rotamer sampling, cartesian minimization, and rigid-body reorientation. The resultant lowest energy model of was inserted into a POPC lipid bilayer, solvated, and equilibrated with constraints on the initial Cα atomic positions. Constraints were then released for 400 ns of unbiased simulation time. Thereafter, each frame was structurally clustered using a 2.6 Å cutoff to identify unique conformations. Additionally, for each simulation frame, a Pearson correlation was calculated between the per-residue inter-helical distances across the αvβ3 TM domain (i.e., NIC of β3 residues) and the per-residue Fractional Activation datasets for both OPN and WOW-1 binding (Fig. 2.7).

Within the first 20 ns of unrestrained simulation, the αv and β3 TM helices rapidly re-oriented to a new geometry with S699, A703, I707 at the interface, resulting in a marked increase in the correlation between the model and the scanning mutagenesis results (Fig. 2.7). After remaining in this associated state for 75 ns, the TM helices dissociated in a conformation that persisted for an additional 70 ns. Meanwhile, the cytoplasmic domains remained in contact, including the αv R995 – β3 E723 salt bridge (Fig. 2.8). The αvβ3 TM heterodimer was

reconstituted with an inter-helical geometry similar to that prior to dissociation, persisting in this conformation for the remaining 245 ns.

The major TM domain conformation (cluster medoid, 80.4% of 400 ns) had a consistently low structural deviation within the cluster (mean Cα RMSD = 1.20 ± 0.01) suggesting that it is energetically stable. Further, there was a statistically significant correlation between frames in this cluster and the rupture force spectroscopy results ($p < 0.05$). An overlay of the rupture force spectroscopy results for each β3 mutation and the inter-helical distance for each β3 residue in the αvβ3 TM domain, as measured from a representative model of the major conformation, is shown in Figure 2.7. It is noteworthy that the modeled interface agrees well with both data sets. Further, while the average correlation with simulation frames is greater for the WOW-1 data than for the OPN data, the change in correlation coefficient between the initial to the medoid frame is greater for the OPN data than for the WOW-1 data.

A representative simulation snapshot depicting the αvβ3 TM dimer interface is shown in Figure 2.7. The helices adopt a geometry that is similar to the canonical GAS_{Right} motif (*1*) in its parallel right-handed crossing (-23.9 ± 0.3 °, SEM), but differs in its larger inter-helical distance of 10.2 ± 0.1 Å. We find αv side chains packing against the β3 residues most sensitive to mutation. Overall, the agreement of the major conformation from the simulation with the rupture force spectroscopy results is remarkable, given that no additional forces were applied to the system to bias helix association or geometry.

Discussion

TM domain interactions, mediated by specific sequence motifs, stabilize integrins in their inactive conformations (*2, 22, 23*). In the best characterized example, a cannoical G-x$_3$-G motif in the αIIb TM helix packs against a reciprocating large residue-small residue motif in the β3 TM helix to stabilize the integrin αIIbβ3 in its inactive state (Fig. 2.1). Given the high degree of sequence conservation in both α and β chains, it seemed likely that this TM domain complex might be common to all integrins. This notion was further supported by our previous examination of the heterodimeric interaction strength between the TM domain segments of integrins in *E. coli* using the dominant negative TOXCAT assay (*8*). For the interaction of isolated α and β TM segments in bacterial membranes, mutating residues homologous to the αIIbβ3-like reciprocating large-small motif within β1-, β2-, and β3-containing integrin pairings was highly disruptive to dimerization. In striking contrast however, the results presented here using the full-length protein in mammlian membranes indicates that the second platelet intergin αvβ3 is instead regulated by a distinct TM domain interface, which is also highly conserved (Table 2.1).

When expressed in bacterial cell membranes, the β3 residues most sensitive to mutations were residues at the αIIbβ3 interface (*8*). Thus, the geometry of the isolated αvβ3 TM domain heterodimer assembled in bacterial membranes appears to differ from the geometry of TM domain heterodimer when it is present in the context of the full-length receptor and expressed in mammalian membranes. This suggests that either the lipid environment or contiguous extracellular or intracellular sequences influence the conformation of the TM domain (*24, 25*).

Crystal structures of the αvβ3 and αIIbβ3 ectodomains and negatively-stained electron microscope images of the full-length proteins are similar (*26-29*). However, in the more membrane-distal regions of the β3 ectodomain, the interactions with either αv or αIIb that help maintain the inactive conformations of αvβ3 and αIIbβ3 differ (*25*). Consequently, the orientation of the distal β3 stalk and the contiguous β3 TM helix might be expected to be shifted with respect to complementary portions of αv and αIIb as well. While the composition of the αIIbβ3 TM domain interface heterodimer is well-established, there has been until now a paucity of information about the structure of the αvβ3 TM domain. Although each contain small residue-x_3-small residue motifs in approximately the same position, the van der Waal surfaces of the αv and αIIb TM helices are certainly distinct and distinguishable. Computationally designed TM domain peptides were able to interact exclusively with either αv and αIIb (*30*). Then, it is not suprising that β3 can interact with αv and αIIb via distinct interfaces, varying in affinity (*17, 31*).

The distinct TM domain interface between αvβ3 than αIIbβ3 might be rationalized by their varying requirements for stringent regulation of the high-affininty ligand binding conformation given their different physiological roles. αvβ3 is widely expressed, mediating cell adhesion and migration on a variety of immobilized ligands containing an RGD motif (*32*), while αIIbβ3 expression is limited to platelets. Moreover, αIIbβ3 readily interacts with soluble ligands, whereas αvβ3 does not. Previously, using optical trap-based force spectroscopy, we found that under essentially identical conditions, the average adhesion strength for OPN-αvβ3 interactions was 47 ± 7 pN, whereas the average adhesion strength of fibrinogen binding to αIIbβ3 was nearly 2-fold greater at 80-90 pN (*20, 33*). This likely reflects that the physiologic requirement that fibrinogen binding to αIIbβ3 on aggregated platelets must be sufficiently strong to resist the shear forces

present in arterial blood, which would be not the case for the adherence and spreading of single platelets mediated by ligand binding to αvβ3. Further, we found that a substantial fraction of unstimulated platelets exhibited a rupture force distribution for interactions with OPN similar to ADP-stimulated platelets (*20*), whereas the fraction of unstimulated platelets that interacted spontaneously with fibrinogen was negligible (*33*). This implies that αIIbβ3 is less prone to activation than αvβ3, perhaps a consequence of the differences in the interfaces of their TM domains.

From a biophysical perspective, the structural consequences of mutation observed here are in accordance with the literature but also offer new insights. Employing small-to-large or large-to-small mutations successfully induced major disruptions in van der Waals packing at the TM heterodimer interface. However, deviation of this pattern to much more subtle mutants, namely Leu-Ile or Ile-Leu isomers, also had dramatic effects. For example, Ile707Leu had a marked effect on both OPN and WOW1 binding, despite only one methyl group being stereochemically repositioning in the entire protein complex. This result illustrates a strict steric requirement for van der Waals packing interactions between TM domain side-chains that dictates the stabilization and regulation of the αvβ3 resting state. Similar Ile-Leu mutations have been observed to abrogate select TM domain interactions for shorter engineered TM peptides (*34*), yet it is noteworthy to find a similar result within a large protein complex interacting simultaneously through several extra-membrane domains. The mutational consequences might also be rationalized in the context of previous reports for small-x_3-small motifs. Replacement of the neighboring β-branched residues V700 and I704 with alanine and leucine, respectively, caused constitutive WOW-1 binding to αvβ3 as well. In a statistical analysis of amino acid patterns in TM helices, Senes et al found that not only were small-x_3-small residue motifs were frequent, but these motifs were commonly

associated with β-branched aliphatic residues in neighboring $i \pm 1$ positions (*11*). Further, they posited that because the side chains of valine and isoleucine are constrained to only one rotamer when they are present in a helix, these residues contribute to the stability of transmembrane helix dimers by minimizing entropy loss upon helical packing. Thus, it is possible that by mutating either V700 to alanine and I704 to leucine, the αvβ3 TM domain is destabilized sufficiently to permit WOW-1 binding to different range of activated αvβ3 conformations, but not the more stringent binding of OPN. In summary, the structural consequences of these αvβ3 mutations provide insight into the strong evolutionary pressure that dictates both the sequence conservation and deviation at TM domains across the integrin family. We find that steric compatibility between the apolar van der Waals surfaces in the TM region likely sets the threshold for differential subunit pairing and functional regulation in integrins.

Our MD simulations are also consistent with numerous experimentally-reported features of the αvβ3 cytosolic and juxtamembrane regions (*24, 35*). The homology model began with the entire cytoplasmic domains of both αv and β3 as α-helical, yet we observed that only αv V993-V1024 and β3 D718-N769 have persistent α-helical structure, with the remaining regions in extended conformations. This is consistent with NMR measurements of a β3 TM-cytoplasmic construct in DHPC:DMPC bicelles where β3 was helical through residue A737, well beyond the TM domain and into the cytoplasm (*24*). Also consistent with this NMR study, K713 (numbered K716 in ref (*24*)) snorkeled its amine moiety to the cytoplasmic lipid head group region in the majority of simulations. Another feature is the reported R995 to E723 salt bridge, previously suggested to stabilize the low-affinity state of αvβ3 (*35*). This salt bridge was formed during the initial structural relaxation of the αvβ3 homology model and maintained through the first 260 ns of simulation, even during the TM helix dissociation event (Figure 2.8). Yet, in the last the last 60

ns we observe a switch to a K994 – E723 salt bridge instead. This simulation does not completely sample the orientational and conformational energy landscapes available to αvβ3, yet in this short time scale, we observed behavior largely consistent with previous structural and functional experiments of αvβ3, implying that the simulated conformation may be physiologically relevant.

In concert with the optical-trap force spectroscopy, our MD simulations are consistent with a unique interface for the αvβ3 TM heterodimer. The interhelical distances observed in the major dimeric conformation in simulation match those we expect, given the disruption of the resting state measured experimentally. Our final model, derived entirely from unbiased simulations, is shown in Figure 2.7, and provides atomic level understanding of the TM domain interactions that constrain αvβ3 in its inactive state, as well as insight into how homologous integrins uniquely regulate cellular signal transduction.

Conclusion

TM helix-helix interactions are known play important roles in the function of many type 1 membrane proteins, including integrins and receptor-linked tyrosine kinases. Often the TM helices of these multi-domain proteins contain two TM-helix interaction motifs, arranged such that both motifs cannot simultaneously engage a single target without large structural distortions. At the outset of our investigation, we expected that the different helix-helix interaction motifs might be alternatively engaged during the dynamics of signaling, as has been proposed previously for members of the EGF receptor family (*36, 37*). Therefore, we were somewhat surprised to find that the two motifs found in the β3 TM helix mediate interactions with two different α subunits. Given the combinatorial assembly of integrin subunits into heterodimers with distinct biologic functions, the ability for a single TM domain to use two different faces of its helix for different interaction

partners is noteworthy. This phenomenon might contribute to both the specificity of interaction and the different activation energies required for different receptors. Moreover, the ability to achieve binding on two faces of a single helix provides an attractive mechanism to evolve new partners in the integrin family as well as other membrane proteins.

Materials and Methods

Stable Expression of Wild-Type and Mutant αvβ3 in CHO cells

Full-length cDNAs for human αv and β3 were inserted into pcDNA3.1(+) Neo and pcDNA3.1(+) Zeo (Invitrogen), respectively. Single amino acid substitutions in the β3 TM domain were introduced into the β3 cDNA using the QuickChange Site-Directed Mutagenesis Kit (Stratagene) as described previously (*25*). All mutated sequences were confirmed by DNA sequence analysis.

CHO cells were cultured in Ham's F-12 media (Gibco/BRL) supplemented with 10% fetal bovine serum (Hyclone Laboratories). Plasmids containing the cDNAs for human αv and β3 were introduced into the CHO cells using FuGENE 6 according to the manufacturer's instructions (Roche Molecular Biochemicals). Two days after transfection, cells were transferred to a selection medium containing 500 μg/mL G418 (Gibco/BRL) and 300 μg/mL Zeocin (Invitrogen). After three weeks of selection, αvβ3 expression was assessed by flow cytometry using the αvβ3-specific monoclonal antibody LM609 (Millipore).

Optical Trap-Based Measurements of Single Molecule αvβ3 Interactions with OPN and the
Monoclonal Antibody WOW-1

To quantify αvβ3 binding to OPN and WOW-1, we used optical trap-based force spectroscopy, a method to measure single-molecule nanomechanics that we used previously to study OPN binding to αvβ3 on platelets (*20*), as well as the interaction of αIIbβ3 with fibrinogen and fibrin (*33, 38, 39*). OPN is an acidic phosphorylated subendothelial matrix glycoprotein whose Arg-Gly-Asp (RGD) motif interacts with a variety of αv- and β1-containing intergrins, including the activated conformation of αvβ3 (*21*). Because OPN does not interact with αIIbβ3 (*30*), it identifies β3 mutations that specifically activate αvβ3. WOW-1 is an engineered αvβ3 ligand in which the 19 amino acid H-CDR3 of the αIIbβ3-specific activation-dependent PAC-1 Fab was replaced by the 50 amino acid αv binding domain of the adenovirus type 2 penton base, thereby switching the specificity of the Fab from αIIbβ3 to the αv-containing integrins αvβ3 and αvβ5 (*16*). Similar to OPN, WOW-1 binds preferentially to activated αvβ3, but unlike OPN, WOW-1 also binds to inactive αvβ3, albeit at lower levels (*16*). The strength and probability of OPN and WOW-1 binding to CHO cells expressing WT and mutant αvβ3 were used as a measure of the αvβ3 activation state.

To prepare the OPN- and WOW-1-coated latex bead for these experiments, human OPN (R&D Systems, Minneapolis, MN) and the WOW-1 monoclonal antibody (a gift of Dr. S. J. Shattil) were covalently bound to 1.87-μm carboxylate-modified latex beads (Bangs Laboratories, Inc., Fishers, IN) using N-(3-dimethylaminopropyl)-N′-ethylcarbodiimide hydrochloride (Sigma, St. Louis, MO) as a cross-linking agent in a two-step procedure. After activation of the bead surface in 0.1 M MES, pH 5.2, by mixing 1 ml of the buffer, 10 μl of a 10% bead suspension, 1.5 mg of dry

carbodiimide, followed by constant shaking for 15 min at 4ºC, the beads were sedimented (4,000g, 2 min), washed once with binding buffer (0.055 M borate buffer pH 8.5), re-sedimented, and suspended in 20 μg/ml OPN or WOW-1 solution in binding buffer. After 15 min incubation with stirring at 4ºC, the beads were sedimented and re-suspended in BSA solution (2 mg/ml in 0.055 M borate buffer pH 8.5) or 1 M ethanolamine solution in the same buffer. OPN- and WOW-1-coated beads were freshly prepared before each experiment.

The experimental model system for these measurements have been previously described in detail (*40*). Briefly, a mixture of the OPN- or WOW-1-coated latex beads (~10^6/ml each) and transfected CHO cells were suspended in culture medium containing 0.3 mM Ca^{2+} and 0.6 mM Mg^{2+} and placed in a flow chamber. After the chamber was placed on a microscope stage, the CHO cells were allowed to settle onto the polylysine-coated bottom of the chamber for 5-10 minutes. Using a manual stage, an adherent CHO cell was selected and brought to the center of the microscopic field. An OPN- or WOW-1-coated bead in the vicinity of the cell was then trapped by the laser beam and the microscope focus was adjusted so that the bead and cell centers were approximately the same distance from the bottom surface of the chamber. The position of the optical trap, and hence the position of the bead, was oscillated in the X-direction in a triangular waveform at a frequency of 10 Hz and at constant peak-to-peak amplitude of 0.8 μm. The trap stiffness was 0.2 pN/nm and loading rate 3,200 pN/s. Separation of the cell and bead was progressively reduced in 10-nm steps using a piezostage until repeated contacts between the bead and cell were observed. Data collection was initiated at first contact between the bead and cell and the force exerted by the trap to separate the bead and cell was recorded. If the force was sufficient to detach the bead from the CHO cell surface, it was registered as the rupture force and reflected the strength of binding of the bead to the cell.

Data were recorded and analyzed with custom LabVIEW (National Instruments, Austin, TX) software. The position of a trapped bead and its displacement from the laser beam focus were recorded with 0.5 ms time resolution as a digitalized voltage curve consisting of many individual voltage (force) signals, each reflecting a binding/rupture event (40). The software recognized the peak, calculated the amplitude of each signal, and presented each peak in force units based on a calibration coefficient. Rupture forces were sorted into a histogram with 5 pN-wide bins and normalized by the total number of contacts. The percentage of events in a bin represents the probability of a bond rupture in that force range. As inferred from numerous controls, forces in the range of 0 to 10 pN represent background noise or optical artifacts and were neglected from data analysis (20, 38, 40). Accordingly, the first bin was neglected without affecting the results since meaningful events only occur within the force range of 10-100 pN. In order to obtain reasonable statistics, rupture force measurements were made at a high repetition rate. The results of experiments for multiple (50-100) cell-bead pairs under similar experimental conditions were averaged so that each force histogram represented about 10^4 to 10^5 contacts between similar reacting surfaces. To preclude the inclusion of non-specific surface interactions of CHO cells in the data analysis, the cumulative binding probability of forces >20 pN was used as a quantitative measure of the ligand binding activity of a particular αvβ3 mutant (17). For each β3 TM domain mutant, we calculated a Fractional Activation (FA) value, a measure of the ability of a mutation to cause αvβ3 activation, using the following equation:

$$FA = P(mut)/P(WT\text{-}Mn^{2+}), \hspace{2cm} \text{(Equation 1)}$$

where P(mut) is the cumulative probability of αvβ3 mutant binding to OPN or WOW-1 in the absence of Mn^{2+} and $P(WT\text{-}Mn^{2+})$ is the cumulative probability for WT αvβ3 binding to OPN or

WOW-1 binding in the presence of 1 mM Mn^{2+}, representing maximal ligand binding to unmutated activated αvβ3.

Modeling of αvβ3 using Rosetta

A model of the TM and cytoplasmic (CYTO) domains of αvβ3 was built by threading their sequences, aligned to αIIb, onto the coordinates from the first conformation of the reported αIIbβ3 NMR model (pdb: 2knc). The Uniprot identifiers corresponding protein sequence ranges are as follows: Human Integrin αv (990 – 1035), P06756; Uniprot Human Integrin αIIb (993 – 1039), P08514; Human Integrin β3 (715 – 788), P05106. These sequences, as well as the sequence alignment of the α subunits are shown in Figure 2.9 and Table 2.1. The Rosetta molecular modeling suite was then used to add hydrogens and to minimize and relax the model. The implicit membrane RosettaMP framework was used as described by Alford et al (*41*) and implemented as a RosettaScripts (*42*). The αvβ3 model was oriented in the membrane based on the Orientations of Membrane Protein server prediction (*43*) for the parent αIIbβ3 NMR model. Fifty trajectories of the FastRelax protocol were conducted, with each trajectory including eight rounds of Markov chain Monte Carlo sampling with simulated annealing, including stages of membrane rigid body re-orientation, protein rotamer sampling, and protein Cartesian minimization subject to evaluation by the membrane all-atom high resolution score function (weights: "mpframework_smooth_fa_2012").

Agreement between each of the OPN and WOW-1 binding experiments and the TM domain interface for each structural model of αvβ3 was assessed by measuring the Pearson correlation between the disruption index for each β3 TM domain residue and the distance of the Cα atom of

88

that residue to the closest αv C$_\alpha$ atom. Prior to this statistical analysis, the inter-helical distance

for each residue was transformed into a Normalized Interhelical Closeness (NIC) value:

$$NIC_i = (D_{max} - D_i) / (D_{max} - D_{min}) \qquad \text{(Equation 2)}$$

where NIC_i is the normalized inter-helical distance of residue i, D_i is the inter-helical position at

residue i, and D_{min} and D_{max} are minimum and maximum inter-helical distances equal to 5 Å and

10.5 Å, respectively. Thus, $NIC_i = 1$ when residue i on $\beta3$ is $\leq$5 Å from αv (i.e., likely interacting)

and $NIC_i = 0$ when residue i is $\geq$ 10.5 Å from αv (i.e., likely not interacting). The NIC value was

created to normalize the interhelical-distance data to a value ranging from 0 to 1 so that it can be

visually compared and correlated to the D_i for each residue. Each Pearson correlation value was

considered to be statistically significant at a p value <0.05

Molecular dynamics simulation and analysis

The model obtained from RosettaMP was used as the input conformation for all-atom

Langevin dynamics simulations under NPT conditions in a 2-Oleoyl-1-palmitoyl-sn-glycero-3-

phosphocholine (POPC) bilayer using the CHARMM36 force field and the ACEMD engine (*44,

45*). The system was built in VMD (*46*) as a 60 Å by 60 Å bilayer with a total height of 135 Å,

having a water box extending 75 Å above the bilayer and 15 Å below the bilayer to solvate the

extended CYTO domains and the N-termini, respectively. The N-termini were modeled as charged

amines to ensure that the flexible N-terminal residues remain outside the membrane, mimicking

the large soluble extracellular domains of integrins. The final system was neutralized, brought to

0.15 mM salt with NaCl, and contained 44211 atoms. A Berendsen barostat (1.01325 bar) and

Langevin thermostat (310 K) with a 4 fs time step alongside a hydrogen mass repartitioning

scheme with a re-scaling factor of 4 were used throughout (*47*). Electrostatics were calculated using a Particle Mesh Ewald grid with spacing of 1 Å, a cutoff of 9 Å, a switching function beginning at 7.5 Å, and exclusion non-bonded calculations under 4 bonded atoms away, typical for this MD engine. van der Waals interactions were calculated with this switching distance as well. The initial model was minimized, then an equilibration simulation was conducted for 15 ns with simple harmonic positional constraints of backbone Cα atoms to their initial conditions with a Langevin damping constant of 1.0 ps^{-1}. Constraints were then lifted for a 400 ns production simulation with a Langevin damping constant of 0.1 ps^{-1}.

For each simulation frame (0.98 ns/frame), the protein was clustered by the RMSD of its TM domain using a hierarchical clustering algorithm with complete linkage using a 2.6 Å cutoff. The TM domain span used for this analysis is underlined in Figure 2.1. The cluster centroid was identified from the major structural cluster using a *k*-medoids clustering algorithm. The atomic RMSD for each Cα atom within the TM domain (40 atoms) from this cluster centroid, after superposition, was calculated for each frame. Additionally, the inter-helical crossing angle and distance were calculated. For this geometric analysis, each of the TM helices was independently fit to a cylindrical surface to approximate the central axis unit vector of the helix. The inter-helical crossing angle was then calculated as the dihedral angle between the two helical axis vectors. The inter-helical distance was found by approximating the closest point between the two helical axis vectors by least-squares minimization, where the length of each vector was found by projecting each terminal residue onto the helical axis unit vector.

Acknowledgments

Contributions

R.L., M.M., J.W., W.D., J.B. designed the experiments. R.L. performed the force spectroscopy. H.Z. performed the molecular biology. M.M. performed the molecular modeling and simulations. R.L., M.M., W.D., and J.B. analyzed the data and wrote the manuscript.

References

1. R. F. Walters, W. F. DeGrado, Helix-packing motifs in membrane proteins. *Proc Natl Acad Sci U S A* **103**, 13658-13663 (2006).

2. R. Li *et al.*, Activation of integrin aIIbb3 by modulation of transmembrane helix associations. *Science* **300**, 795-798 (2003).

3. K. E. Gottschalk, P. D. Adams, A. T. Brunger, H. Kessler, Transmembrane signal transduction of the alpha(IIb)beta(3) integrin. *Protein Sci* **11**, 1800-1812 (2002).

4. J. Zhu *et al.*, The structure of a receptor with two associating transmembrane domains on the cell surface: integrin alphaIIbbeta3. *Mol Cell* **34**, 234-249 (2009).

5. D. G. Metcalf, D. W. Kulp, J. S. Bennett, W. F. DeGrado, Multiple approaches converge on the structure of the integrin alphaIIb/beta3 transmembrane heterodimer. *J Mol Biol* **392**, 1087-1101 (2009).

6. J. Yang *et al.*, Structure of an integrin alphaIIb beta3 transmembrane-cytoplasmic heterocomplex provides insight into integrin activation. *Proc Natl Acad Sci U S A* **106**, 17729-17734 (2009).

7. T. L. Lau, C. Kim, M. H. Ginsberg, T. S. Ulmer, The structure of the integrin alphaIIbbeta3 transmembrane complex explains integrin transmembrane signalling. *Embo J* **113**, 4747-4753 (2009).

8. B. W. Berger *et al.*, Consensus motif for integrin transmembrane helix association. *Proc Natl Acad Sci U S A* **107**, 703-708 (2010).

9. J. S. Bennett, C. Chan, G. Vilaire, S. A. Mousa, W. F. DeGrado, Agonist-activated avb3 on platelets and lymphocytes binds to the matrix protein osteopontin. *J. Biol. Chem.* **272**, 8137-8140 (1997).

10.   L. A. Fitzgerald *et al.*, Comparison of cDNA-derived protein sequences of the human fibronectin and vitronectin receptor a-subunits and platelet glycoprotein IIb. *Biochemistry* **26**, 8158-8165 (1987).

11.   A. Senes, M. Gerstein, D. M. Engelman, Statistical analysis of amino acid patterns in transmembrane helices: the GxxxG motif occurs frequently and in association with beta-branched residues at neighboring positions. *J Mol Biol* **296**, 921-936 (2000).

12.   M. Hoefling, H. Kessler, K. E. Gottschalk, The transmembrane structure of integrin alphaIIbbeta3: significance for signal transduction. *Angew Chem Int Ed Engl* **48**, 6590-6593 (2009).

13.   J. S. Bennett, Platelet-fibrinogen interactions. *Ann N Y Acad Sci* **936**, 340-354 (2001).

14.   J. S. Bennett, Structure and function of the platelet integrin alphaIIbbeta3. *J Clin Invest* **115**, 3363-3369 (2005).

15.   C. Giachelli, L. Liaw, C. Murry, S. Schwartz, M. Almeida, Osteopontin expression in cardiovascular diseases. *Annals.N.Y.Acad.Sci.* **760**, 109-126 (1995).

16.   N. Pampori *et al.*, Mechanisms and consequences of affinity modulation of integrin alpha(V)beta(3) detected with a novel patch-engineered monovalent ligand. *J Biol Chem* **274**, 21609-21616 (1999).

17.   R. I. Litvinov *et al.*, Activation of individual alphaIIbbeta3 integrin molecules by disruption of transmembrane domain interactions in the absence of clustering. *Biochemistry* **45**, 4957-4964 (2006).

18.   W. Li *et al.*, A push-pull mechanism for regulating integrin function. *Proc Natl Acad Sci U S A* **102**, 1424-1429 (2005).

19.	H. Zhu *et al*., Specificity for homooligomer versus heterooligomer formation in integrin transmembrane helices. *J Mol Biol* **401**, 882-891 (2010).

20.	R. I. Litvinov, G. Vilaire, H. Shuman, J. S. Bennett, J. W. Weisel, Quantitative analysis of platelet alpha v beta 3 binding to osteopontin using laser tweezers. *J Biol Chem* **278**, 51285-51290 (2003).

21.	O. Helluin *et al*., The activation state of avb3 regulates platelet and lymphocyte adhesion to intact and thrombin-cleaved osteopontin. *J Biol Chem* **275**, 18337-18343 (2000).

22.	B. H. Luo, T. A. Springer, J. Takagi, A specific interface between integrin transmembrane helices and affinity for ligand. *PLoS Biol* **2**, e153 (2004).

23.	D. T. Moore, B. W. Berger, W. F. DeGrado, Protein-protein interactions in the membrane: sequence, structural, and biological motifs. *Structure* **16**, 991-1001 (2008).

24.	Z. Lu *et al*., Implications of the differing roles of the beta1 and beta3 transmembrane and cytoplasmic domains for integrin function. *Elife* **5**, (2016).

25.	J. E. Donald, H. Zhu, R. I. Litvinov, W. F. DeGrado, J. S. Bennett, Identification of interacting hot spots in the beta3 integrin stalk using comprehensive interface design. *J Biol Chem* **285**, 38658-38665 (2010).

26.	J. P. Xiong *et al*., Crystal structure of the extracellular segment of integrin aVb3. *Science* **294**, 339-345. (2001).

27.	B. D. Adair *et al*., Three-dimensional EM structure of the ectodomain of integrin {alpha}V{beta}3 in a complex with fibronectin. *J Cell Biol* **168**, 1109-1118 (2005).

28.	J. Zhu *et al*., Structure of a complete integrin ectodomain in a physiologic resting state and activation and deactivation by applied forces. *Mol Cell* **32**, 849-861 (2008).

29. E. T. Eng, B. J. Smagghe, T. Walz, T. A. Springer, Intact alphaIIbbeta3 integrin is extended after activation as measured by solution X-ray scattering and electron microscopy. *J Biol Chem* **286**, 35218-35226 (2011).

30. H. Yin *et al.*, Computational design of peptides that target transmembrane helices. *Science* **315**, 1817-1822 (2007).

31. G. Pagani, H. Gohlke, On the contributing role of the transmembrane domain for subunit-specific sensitivity of integrin activation. *Sci Rep* **8**, 5733 (2018).

32. R. O. Hynes, Integrins: bidirectional, allosteric signaling machines. *Cell* **110**, 673-687 (2002).

33. R. I. Litvinov, D. H. Farrell, J. W. Weisel, J. S. Bennett, The Platelet Integrin alphaIIbbeta3 Differentially Interacts with Fibrin Versus Fibrinogen. *J Biol Chem* **291**, 7858-7867 (2016).

34. L. He *et al.*, Single methyl groups can act as toggle switches to specify transmembrane Protein-protein interactions. *Elife* **6**, (2017).

35. M. A. Muller *et al.*, Cytoplasmic salt bridge formation in integrin alphavss3 stabilizes its inactive state affecting integrin-mediated cell biological effects. *Cell Signal* **26**, 2493-2503 (2014).

36. M. Lelimousin, V. Limongelli, M. S. Sansom, Conformational Changes in the Epidermal Growth Factor Receptor: Role of the Transmembrane Domain Investigated by Coarse-Grained MetaDynamics Free Energy Calculations. *J Am Chem Soc* **138**, 10611-10622 (2016).

37. J. K. L. Sinclair, A. S. Walker, A. E. Doerner, A. Schepartz, Mechanism of Allosteric Coupling into and through the Plasma Membrane by EGFR. *Cell Chem Biol* **25**, 857-870 e857 (2018).

38. R. I. Litvinov, H. Shuman, J. S. Bennett, J. W. Weisel, Binding strength and activation state of single fibrinogen-integrin pairs on living cells. *Proc Natl Acad Sci U S A* **99**, 7426-7431 (2002).

39. P. Hook *et al.*, Strong Binding of Platelet Integrin alphaIIbbeta3 to Fibrin Clots: Potential Target to Destabilize Thrombi. *Scientific reports* **7**, 13001 (2017).

40. R. I. Litvinov, J. S. Bennett, J. W. Weisel, H. Shuman, Multi-step fibrinogen binding to the integrin (alpha)IIb(beta)3 detected using force spectroscopy. *Biophys J* **89**, 2824-2834 (2005).

41. R. F. Alford *et al.*, An Integrated Framework Advancing Membrane Protein Modeling and Design. *PLoS Comput Biol* **11**, e1004398 (2015).

42. S. J. Fleishman *et al.*, RosettaScripts: a scripting language interface to the Rosetta macromolecular modeling suite. *PLoS One* **6**, e20161 (2011).

43. M. A. Lomize, A. L. Lomize, I. D. Pogozheva, H. I. Mosberg, OPM: orientations of proteins in membranes database. *Bioinformatics* **22**, 623-625 (2006).

44. M. J. Harvey, G. Giupponi, G. D. Fabritiis, ACEMD: Accelerating Biomolecular Dynamics in the Microsecond Time Scale. *J Chem Theory Comput* **5**, 1632-1639 (2009).

45. J. Huang, A. D. MacKerell, Jr., CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. *J Comput Chem* **34**, 2135-2145 (2013).

46. W. Humphrey, A. Dalke, K. Schulten, VMD: visual molecular dynamics. *J Mol Graph* **14**, 33-38, 27-38 (1996).

47. H. B. Feenstra KA, and Berendsen HJC, Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. *J Computational Chem* **20**, 786-798 (1999).
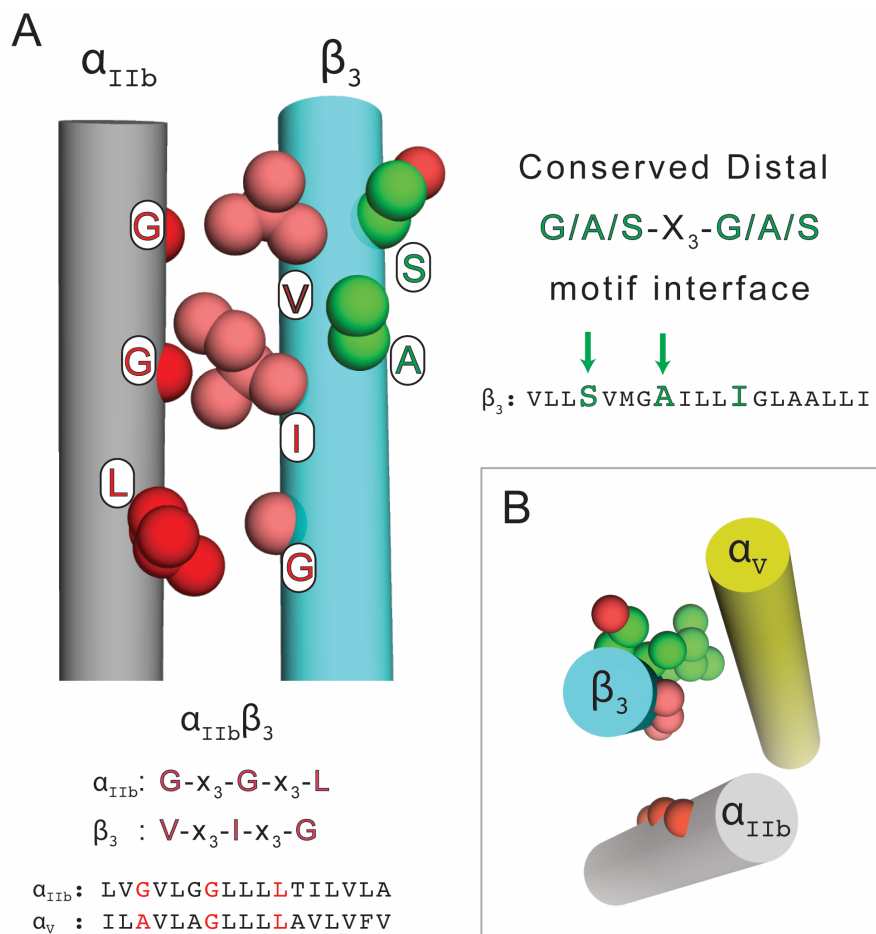
Figure 2.1. Model of the αIIbβ3 TM domain heterodimer highlighting two conserved β3 interaction motifs. **A**. The location of a conserved reciprocating large-small residue interfacial motif responsible for stabilizing the αIIbβ3 TM domain heterodimer is shown as red space-filling balls and its sequence is denoted by red text. The location of a conserved G-x$_3$-G-like motif is shown as green space-filling balls and it sequence denoted by green text and arrows. A presence of a conserved distal isoleucine residue is also denoted in the text. **B**. Top down view of the αIIbβ3 helices with only αIIbβ3 Cα atoms shown. Also displayed is a potential interaction partner for the G-x$_3$-G-like motif, αv (yellow). Given the offset of their helical registers, the two β3 interaction motifs constitute two distinct interaction interfaces.

Figure 2.2. Rupture force histograms of the interaction of OPN-coated latex beads with CHO cells expressing wild-type (WT) human αvβ3. Optical trap-based force spectroscopy was performed as described in the "Materials and Methods". Individual rupture force signals were collected into 5-pN bins and plotted as the percentage of total bead-cell contact-detachment cycles in a particular bin. Rupture forces of <20 pN representing background noise, optical artifacts, or non-specific interactions were neglected. **A**. Rupture force histograms resulting from OPN binding to unstimulated CHO cells expressing WT αvβ3. **B**. Rupture force histograms resulting from OPN binding to CHO cells expressing WT αvβ3 in the presence of 5 mM EDTA. **C**. Rupture force histograms resulting from OPN binding to CHO cells expressing WT αvβ3 in the presence of 1 mM Mn^{2+}. **D.** Bar graph derived from the histograms shown in panels A-C indicating the cumulative probability of detecting rupture forces >20 pN. The statistical significance of differences in cumulative binding probability was determined using the Mann-Whitney U test with $p<0.05$ considered to be statistically significant.
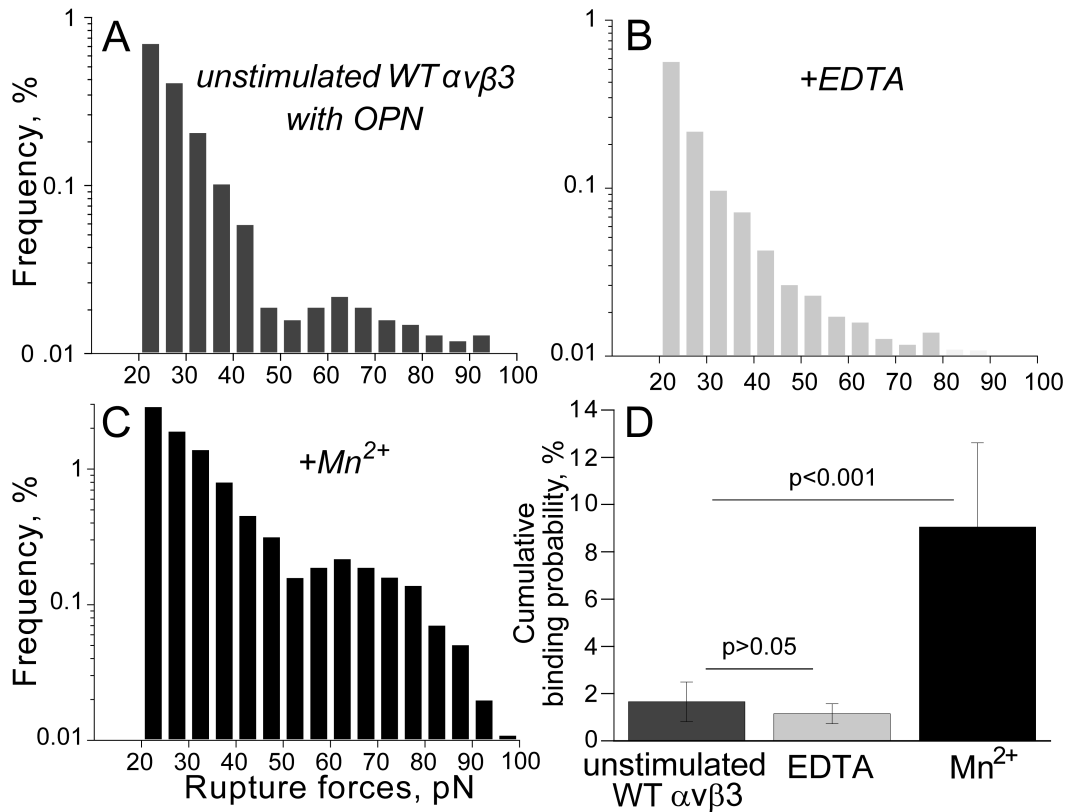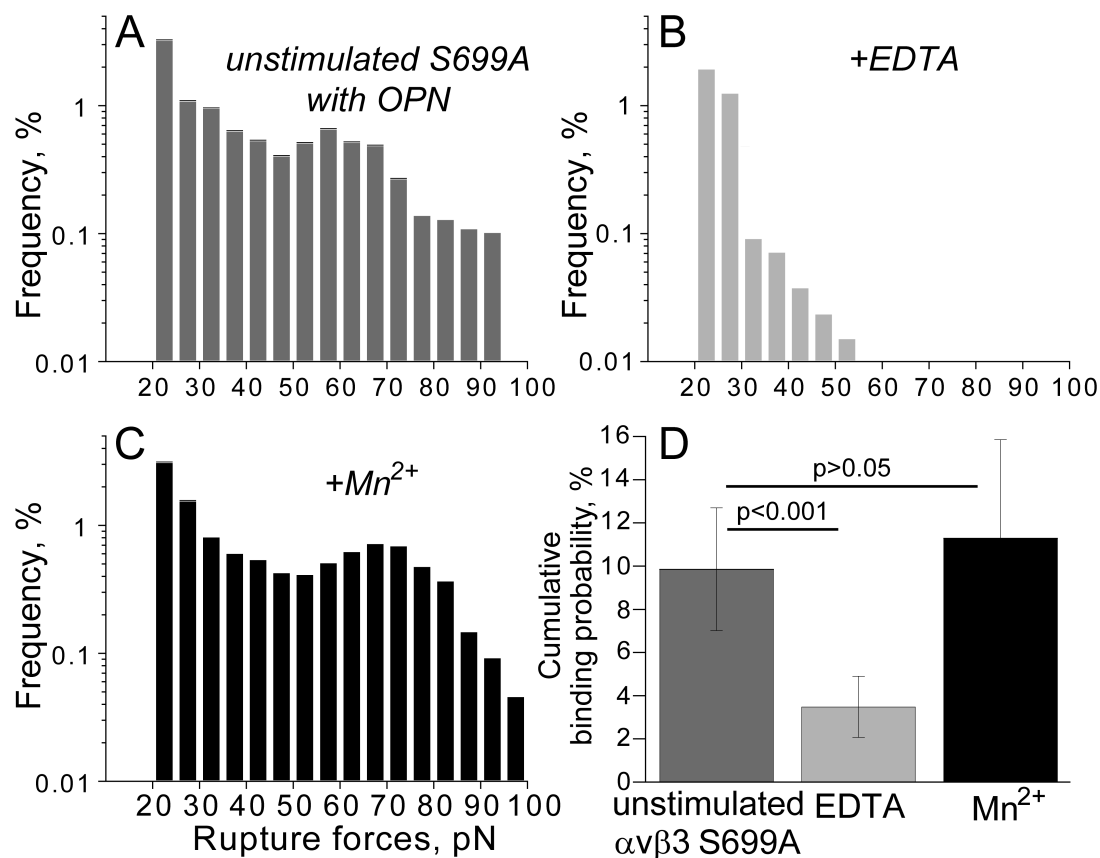
Figure 2.3. Rupture force histograms of the interaction of OPN-coated latex beads with CHO cells expressing human αvβ3 containing the β3 TM helix mutation S699L. Optical trap-based force spectroscopy was performed and analyzed as described in Figure 2.2. **A**. Rupture force histograms resulting from OPN binding to unstimulated CHO cells expressing the β3 S699L αvβ3 mutant. **B**. Rupture force histograms resulting from OPN binding to CHO cells expressing the β3 S699L αvβ3 mutant in the presence of 5 mM EDTA. **C**. Rupture force histograms resulting from OPN binding to CHO cells expressing the β3 S699L αvβ3 mutant in the presence of 1 mM Mn^{2+}. **D.** Bar graph derived from the histograms shown in panels A-C indicating the cumulative probability of detecting rupture forces >20 pN.
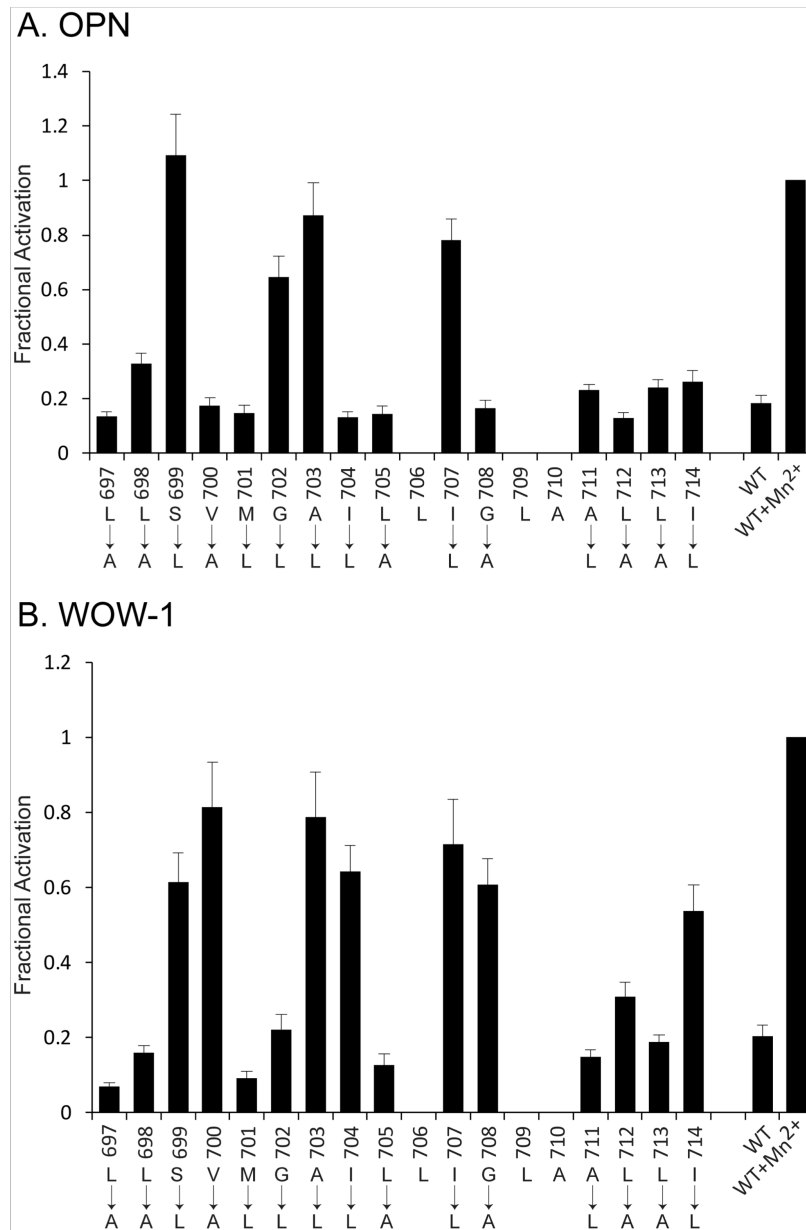
Figure 2.4. Effects of leucine or alanine replacements in the β3 TM helix on the cumulative probability of constitutive OPN (A) and WOW-1 (B) binding to αvβ3 expressed by CHO cells. Alanine or leucine mutations were introduced into the human β3 TM helix as described in the "Materials and Methods". The β3 mutants were then co-expressed with WT human αv in CHO cells and the cumulative probability of constitutive OPN and WOW-1 binding to αvβ3 was measured by optical trap-based force spectroscopy. The cumulative probability of OPN and WOW-1 binding to WT αvβ3 in the absence and presence of 1 mM Mn^{2+} used as negative and positive controls, respectively. The bars and error bars in the figure correspond to the mean and standard error generated from rupture force histograms representing 10^4-10^5 bead-cell contacts in at least 10 independent experiments.
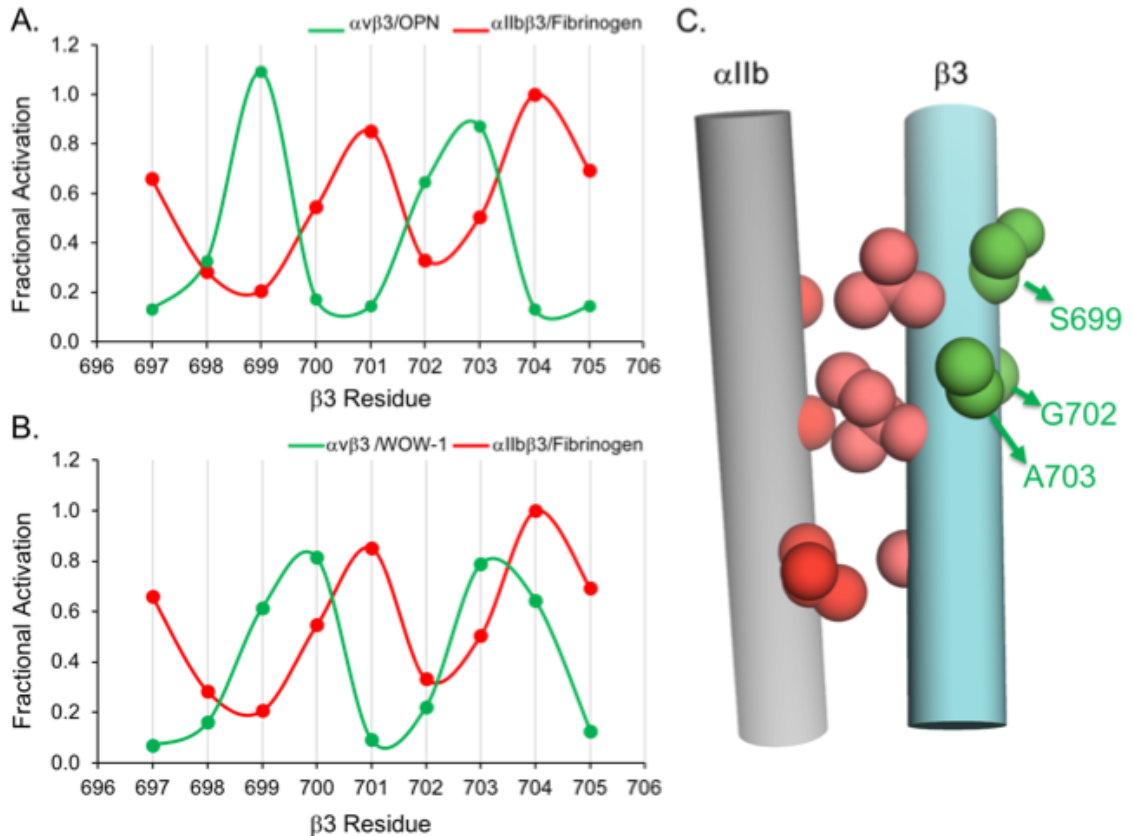
Figure 2.5. The motifs in the β3 TM helix that mediate its association with αv and αIIb are completely out of phase and are located on opposites sides of the β3 helix. The curves shown in **A.** compare the effect of replacing β3 residues 697-705 with leucine or alanine on constitutive αvβ3 binding to OPN using data from Figure 2.4 and the effect of the same replacements on constitutive αIIbβ3 to fibrinogen using data from reference (*4*). The curves shown in **B.** compare the effect of replacing β3 residues 697-705 on constitutive αvβ3 binding to WOW-1 using data from Figure 2.4 and the effect of these replacements on constitutive αIIbβ3 to fibrinogen using data from reference (*4*). The cylinders representing the αIIb and β3 TM helices shown in **C.** demonstrate that the residues responsible for the interaction of β3 with αIIb versus αv reside on opposites sides of the β3 helix. The β3 residues shown in green (S699, G702, and A703) are the residues whose mutation causes constitutive binding of αvβ3 to OPN.
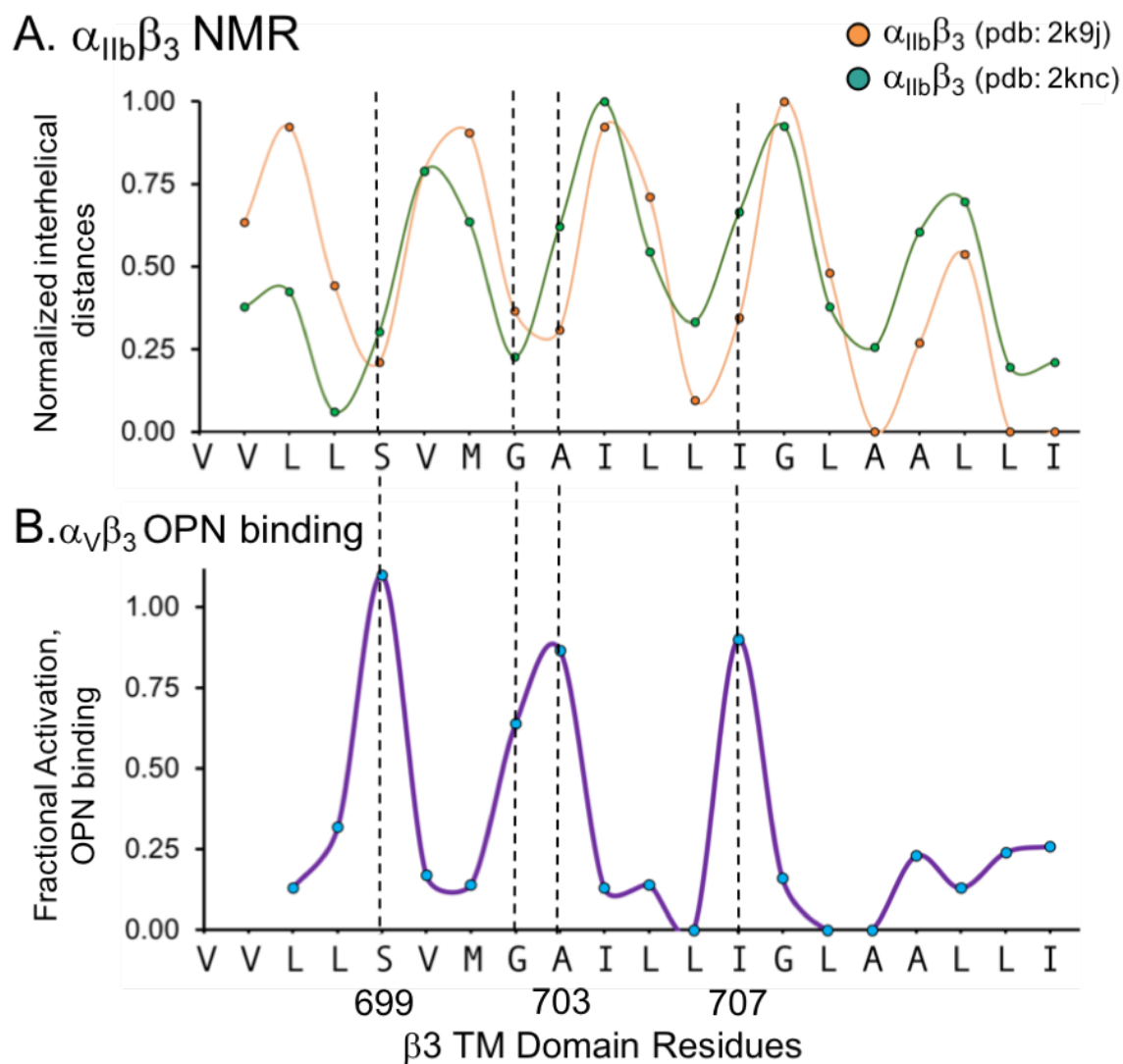
Figure 2.6. The effect of β3 TM domain mutations on αvβ3 binding to OPN and WOW-1 correlate poorly with inter-atomic distances derived from αIIbβ3 NMR structures. A. Plot of the normalized interhelical distances (Inter-helical Closeness) of residues in the β3 TM domain from residues in the αIIb TM domain derived from αIIbβ3 NMR structures (PDB ids: 2k9j, 2knc). **B.** Plot of the fractional activation of αvβ3 caused by β3 TM domain mutations using the OPN binding data from Figure 2.4. The vertical dotted lines correspond to the local maxima for αvβ3 activation.
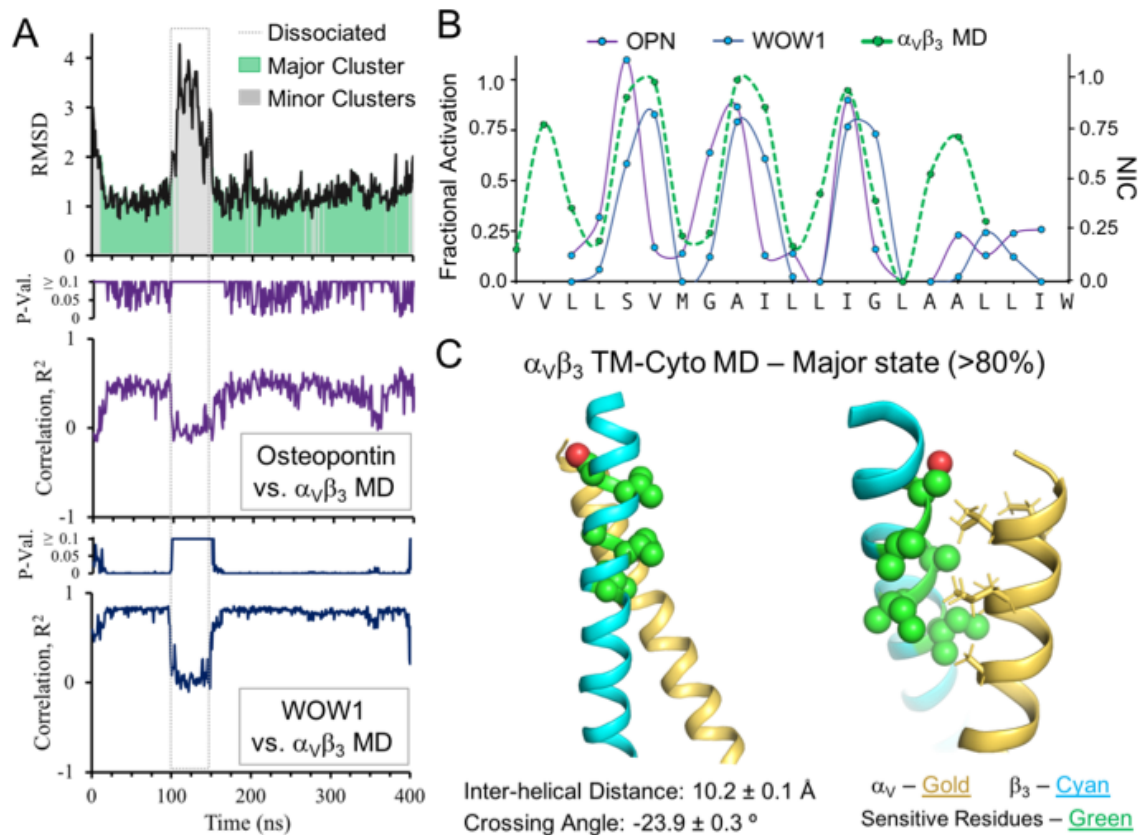
Figure 2.7. Molecular Dynamics (MD) simulations of αvβ3 TM and CYTO domain interactions during αvβ3 binding to OPN and WOW-1. A. Upper panel: RMSD of the αvβ3 TM domain to the centroid of the major structural cluster. The color of each frame indicates whether it is belongs to the major cluster (green) or a minor cluster (grey). Middle and Bottom panels: Pearson correlation coefficients and p-values of αvβ3 TM domain inter-helical distances for each β3 mutation and the cumulative probability of OPN (Middle panel) or WOW-1 (Bottom panel) binding. **B.** Overlay of the normalized inter-helical distance values (inter-helical closeness) for the major αvβ3 TM domain conformation observed in the MD simulations and the Fractional Activation of αvβ3 caused by β3 TM helix mutations using the results shown in Figure 2.4. **C.** Representative model of the major conformation and the mean geometric parameters ± SEM of the αvβ3 TM domain heterodimer observed during the MD simulations.
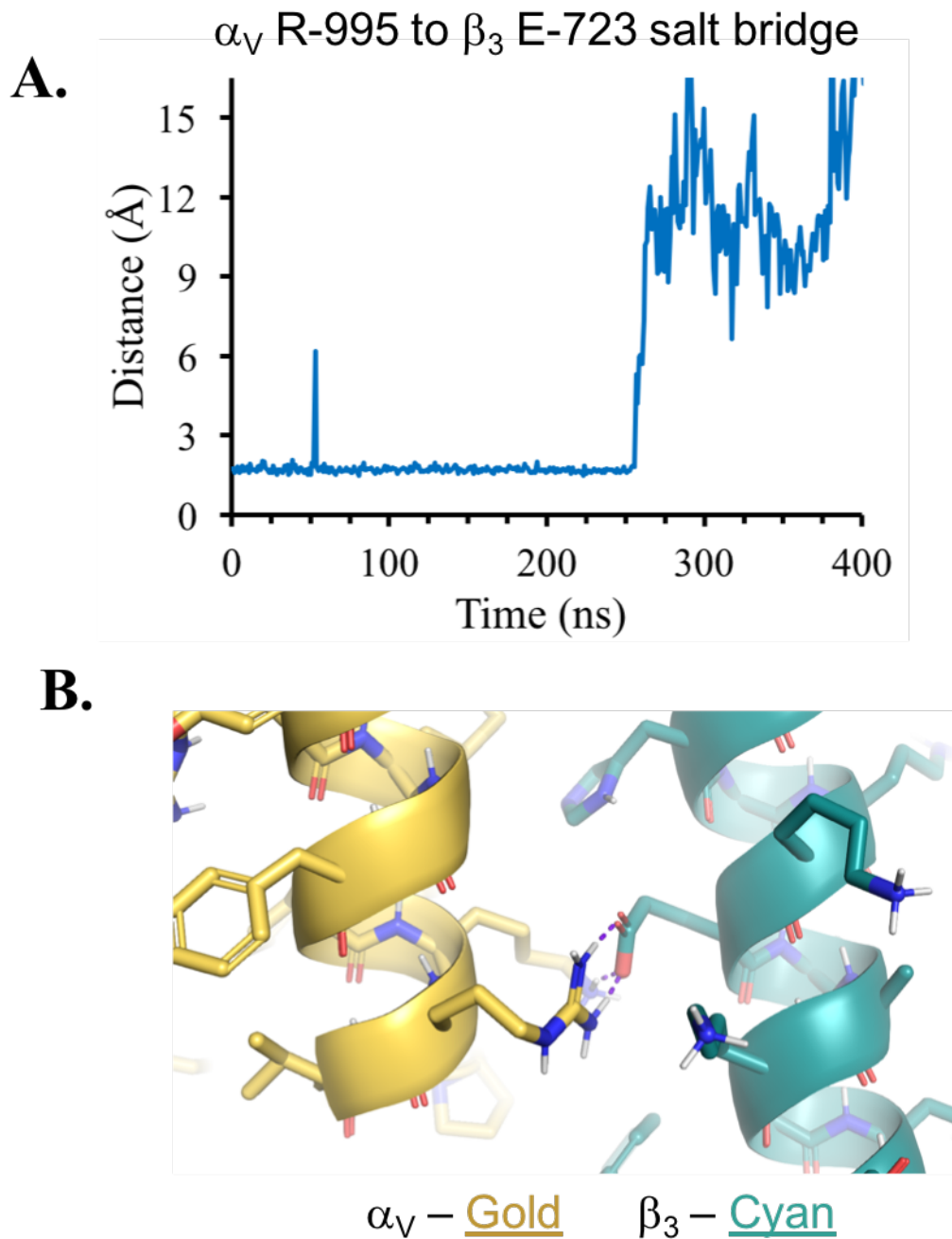
Figure 2.8. Demonstration of an αv R995-β3 E723 salt bridge during MD simulations. A.
The minimum distance observed between the hydrogen bond donor and acceptor atoms of αv R995
and β3 E723, respectively. These residues are in hydrogen bonding range for the first 260 ns of
the simulation, despite the helix dissociation event between the 95 ns and 145 ns. **B.** Model of the
αv R995-β3 E723 salt bridge derived from the centroid frame of the major TM domain structural
cluster.

A. Human β3 residues 715 - 788

```
KGPDIL VVLLSVMGAI LLIGLAALLI WKLLITIHDR KEFAKFEEER ARAKWDTANN PLYKEATSTF TNITYRGT
```

B. Comparison human αv residues 990-1035 with human αIIb residues 993-1039

```
 αv    MPVPVWVIIL AVLAGLLLLA VLVFVMYRMG FFKRVRPPQE EQEREQL
       ___*_*____ _**_*****_ _**__*___* ****_***_* *___*__
αIIb   RAIPIWWVLV GVLGGLLLLT ILVLAMWKVG FFKRNRPPLE EDDEEGE
```

Figure 2.9. β3, αV, and αIIb sequences used for molecular modeling. **A**. Sequence of β3 residues 715-788. The sequence of the β3T M domain is underlined. **B**. Sequence alignment of human αv residues 990-1035 with αIIb residues 993-1039. The αv and αIIb TM domain sequences are shown in bold type. Asterisks correspond to sequence identity.

Table 2.1. Conservation of interaction motifs in human integrin α and β subunit TM domains.
Red indicates the presence of G/A/S-x$_3$-G/A/S type (Small-Small-Large) motifs in both integrin α and β subunit TM domains. It is noteworthy that this motif is absent in β4 and β8. Green indicates the presence of a conserved Large-Large-Small motif in integrin β subunits. In the β subunits, the two interaction motifs are out of helical register.

αIIb:	LVGVLGGLLLLTILVLA	β3:	VLLSVMGAILLIGLAALLI
αV :	ILAVLAGLLLLAVLVFV	β1:	IVAGVVAGIVLIGLALLLI
α1 :	LLSAFAGLLLLMLLILA	β2:	IVGGTVAGIVLIGILLLVI
α2 :	IGSIIAGILLLLALVAI	β5:	ILLAVVGSILLVGLALLAI
α3 :	LVAVGAGLLLLGLIILL	β6:	IMLGVSLAILLIGVVLLCI
α4 :	SSSLLLGLLLLGLIILL	β7:	IVLGCVGGIVAVGLGLVLA
α5 :	ILAILFGLLLLGLLIYI	β8:	IFFIIFIVTFLIGLLKVLI
α6 :	LVAILAGILMLALLVFI	β4:	LIPLLLLLLPLLALLLLLC
α7 :	LLAVLAGLLVLALLVLL		
α8 :	ILAILLGLLVLAILTLA		
α9 :	AISLLVGILILLLLAVL		
α10 :	IGAVLGGLLLLALLVFC		
α11 :	VGSTLGGLLLLALLVLA		
αD :	MGSSVGALLLLALITAT		
αE :	IKGSVGGLLVLIVILVI		
αL :	VLSGIGGLLLLLLIFIV		
αM :	VGSSVGGLLLLALITAA		
αX :	VGSSIGGLLLLALITAV		

Table 2.2. Reactivity* of CHO cells expressing αvβ3 TM β3 mutants and WT αvβ3 with OPN and WOW-1 in the absence and presence of Mn^{2+}, an integrin activator

αvβ3 variant	Ligand	Cumulative binding probability, %	Fractional Activation (FA)**	Ligand	Cumulative binding probability, %	Fractional Activation (FA)**
β3 L697A	OPN	1.2±0.2	0.13±0.02	WOW-1	0.6±0.1	0.07±0.01
β3 L698A	OPN	2.9±0.4	0.32±0.04	WOW-1	1.5±0.2	0.16±0.02
β3 S699L	OPN	9.9±1.3	1.08±0.15	WOW-1	5.6±0.7	0.61±0.08
β3 S699L + Mn^{2+}	OPN	11.3±1.2	-	-	-	-
β3 V700A	OPN	1.6±0.3	0.17±0.03	WOW-1	7.5±1.0	0.81±0.12
β3 V700A + Mn^{2+}	OPN	8.3±1.0	-	-	-	-
β3 M701L	OPN	1.3±0.2	0.14±0.03	WOW-1	0.8±0.2	0.09±0.02
β3 G702L	OPN	5.8±0.7	0.64±0.08	WOW-1	2.0±0.3	0.22±0.4
β3 A703L	OPN	7.9±1.1	0.87±0.12	WOW-1	7.2±1.0	0.79±0.12
β3 I704L	OPN	1.2±0.2	0.13±0.02	WOW-1	5.9±0.7	0.64±0.07
β3 L705A	OPN	1.3±0.2	0.14±0.03	WOW-1	1.2±0.3	0.13±0.03
β3 L706	-	-	-	-	-	-
β3 I707L	OPN	7.1±0.8	0.78±0.08	WOW-1	6.5±1.1	0.71±0.12
β3 G708A	OPN	1.5±0.3	0.16±0.03	WOW-1	5.6±0.7	0.61±0.07
β3 L709	-	-	-	-	-	-
β3 A710	-	-	-	-	-	-
β3 A711L	OPN	2.1±0.2	0.23±0.02	WOW-1	1.4±0.2	0.15±0.02
β3 L712A	OPN	1.2±0.2	0.13±0.02	WOW-1	2.8±0.4	0.31±0.04
β3 L713A	OPN	2.2±0.3	0.24±0.03	WOW-1	1.7±0.2	0.19±0.02
β3 I714L	OPN	2.4±0.3	0.26±0.04	WOW-1	4.9±0.6	0.54±0.07
WT	OPN	1.8±0.3	0.17±0.03	WOW-1	1.9±0.2	0.20±0.02
WT + Mn^{2+}	OPN	9.1±1.4	1	WOW-1	9.2±1.4	1

*Reactivity is quantified as a cumulative probability of the receptor-ligand interactions with rupture forces >20 pN.
**Fractional Activation (FA) value for each mutant was calculated as FA = P(mut)/P(WT-Mn^{2+}), where P(mut) is the cumulative probability for the mutant in the absence of Mn^{2+} and P(WT-Mn^{2+}) is the average cumulative probability for the WT αvβ3 activated with Mn^{2+}. The results are presented as mean ± standard error of mean from at least 10 independent experiments.

Chapter 3

De Novo Designed Transmembrane Peptides Activating the α5β1 Integrin

De Novo Designed Transmembrane Peptides Activating the α₅β₁ Integrin

Abstract

Computationally designed transmembrane α-helical peptides (CHAMP) have been used to compete for helix-helix interactions within the membrane, enabling the ability to probe the activation of the integrins $\alpha_{IIb}\beta_3$ and $\alpha_v\beta_3$. Here, this method is extended towards the design of CHAMP peptides that inhibit the association of the $\alpha_5\beta_1$ transmembrane domains, targeting the Ala-X$_3$-Gly motif within α_5. Our previous design algorithm was performed alongside a new workflow implemented within the widely used Rosetta molecular modeling suite. Peptides from each computational approach activated integrin $\alpha_5\beta_1$ but not $\alpha_V\beta_3$ in human endothelial cells. Two CHAMP peptides were shown to directly associate with an α_5 transmembrane domain peptide in detergent micelles to a similar degree as a β_1 transmembrane peptide does. By solution-state NMR, one of these CHAMP peptides was shown to bind primarily the integrin β_1 transmembrane domain, which itself has a Gly-X$_3$-Gly motif. The second peptide associated modestly with both α_5 and β_1 constructs, with slight preference for α_5. Although the design goal was not fully realized, this work characterizes novel CHAMP peptides activating $\alpha_5\beta_1$ that can serve as useful reagents for probing integrin biology.

Introduction

Membrane proteins play crucial roles in many fundamental cellular processes including signal transduction, cell adhesion, and membrane trafficking. As a result, a battery of molecular tools has been developed to modulate the structural and functional states of membrane proteins, primarily through their water-soluble domains. However, oligomerization and induced conformational changes of their transmembrane (TM) α-helices often regulate these essential functions (*1*). Thus, accessing critical functional states and signaling outputs of membrane proteins requires the development of agents that act by directly targeting TM domains as well (*2, 3*). Previous efforts have relied on natural sequences (*4-6*) as dominant negative competitors or on large scale mutational scanning (*7-9*) of natural or artificial TM helices to generate agents for this purpose. However, recent work suggests that direct computational design of inhibitors of TM protein-protein interactions represents a promising strategy.

Our group has developed a method to computationally designed peptides known as CHAMPs (<u>C</u>omputed <u>H</u>elical <u>A</u>nti-<u>M</u>embrane <u>P</u>eptides), and used this approach to prepare peptides that were specifically directed against the TM domains of α_{IIb} and α_V, the α subunits of Integrins $\alpha_{IIb}\beta_3$ and $\alpha_V\beta_3$ *(10, 11)*. In the resting state, the TM helices of $\alpha_{IIb}\beta_3$ and $\alpha_V\beta_3$ interact, but they separate when the integrins are activated by pharmacological agonists. Accordingly we designed CHAMP peptides to determine the degree to which TM helix separation contributes to the process of integrin activation. Each CHAMP peptide was found to bind to its target TM domain with reasonable affinity in micelles by a FRET assay, in bacterial cell membranes via the TOXCAT assay, and in mammalian cell membranes by integrin activation assays. Importantly, the CHAMP peptides specifically enabled intact integrins to bind their cognate adhesion proteins when they

were expressed in mammalian cells. More recently, the anti-α_{IIb} CHAMP peptide was used to study the role of TM interactions in outside-in platelet signaling, and it was found to activate β_3-associated Src and Syk, likely by inducing $\alpha_{IIb}\beta_3$ clustering (*12*).

Here we sought to design a CHAMP peptide that selectively binds to the TM domain of α_5, thereby causing the activation of the integrin $\alpha_5\beta_1$. This would provide a reagent useful to deciphering the significantly overlapping biochemical signaling responses of integrins $\alpha_5\beta_1$ versus $\alpha_v\beta_3$ in endothelial cell activation (*13-15*). Because we expected it would be difficult to design CHAMPs with specificity for α_5, given the sequence similarity within the integrin α-subunit family (Figure 3.1), we explored modifications on the original method that introduce computational advances that were made since the original development of the CHAMPs procedure.

The original CHAMP protocol proceeded in several discrete steps. First, potential interaction motifs were targeted in the sequence of the target TM domain. Here, we targeted the G/A/S-X_3-G/A/S sequence motif of α_5, a motif common to dimeric TM domain complexes interacting in a parallel right-handed geometry (GAS$_{right}$) (*16, 17*). Next, examples of three-dimensional structures of these motifs were chosen from the Protein Data Bank (PDB) to guide modeling of a CHAMP peptide (*18*). The sequence of the target protein, here the α_5 TM domain, was threaded in the appropriate register onto one of the two helices in the extracted helical pair, and the sequence of the neighboring CHAMP peptide was computationally designed using a sidechain packing algorithm. The exterior-facing residues of the CHAMP peptide were chosen based on the preference of specific sidechains to lie in distinct regions of a bilayer (*17*). Previously, we used this protocol to successfully design two distinct CHAMPs that were able to differentiate between the distinct, but homologous G/A/S-X_3-G/A/S-X_3-Leu motifs of the α_{IIb} and α_v TM helices.

111

Despite its simplicity, to our knowledge, similar *de novo* TM peptide design methods have not been reported to date. To increase the accessibility of this method to the broader scientific community, we revised the protocol to make use of the ROSETTA molecular modeling suite (*19*) and to automate the scoring and ranking of modeled CHAMP peptides. To compare the original and revised methods, we designed anti-α_5 peptides using both methods in parallel and assessed the ability of the resulting CHAMPs to cause integrin activation in a mammalian cell system. The ability of the peptides to associate with either α_5 or β_1 TM domain fragments was then tested by thiol-disulfide exchange equilibrium crosslinking (*20, 21*) and NMR spectroscopy. This work highlights the promise and limitations of current computational approaches to engineer specific α-helical protein-protein interactions within the membrane in the background of a promiscuous small-X_3-small motif. Furthermore, the designed peptides expand the set of reagents available to dissect the mechanisms and outcomes of integrin signaling through $\alpha_5\beta_1$ and potentially other β_1 integrin.

Results

CHAMP Design

We performed structure-based protein design to generate sequences we expected to bind specifically to the TM domain of the integrin α_5. The designs were achieved using two slightly different protocols, as summarized in Figure 3.1: the first protocol was previously described and used to generate sequences that bind specifically to the α_{IIb} and α_v TM domains (*22*) and the second relies on Rosetta's membrane protein framework, RosettaMP (*19*). Both protocols began by selecting pairs of α-helices interacting in the GAS$_{Right}$ geometry from our clustered database of helix pairs within natural TM proteins (*23*), using their backbone coordinates to model the dimeric

CHAMP - α_5 TM complex. For protocol 1, we selected one pair of helices and the lowest energy sequence was synthesized (CHAMP #1). For protocol 2, we identified two backbone models from the 109 pairs in the GAS_{Right} cluster that contained a sequence motif similar to the α_5 TM domain and also a distance restraint. After using RosettaMP for 800 sequence design trials, 159 and 26 unique sequences were generated for backbones 1 and 2, respectively. Sequences were clustered hierarchically and evaluated using an interfacial packing score and a relative association energy. The sequences in each cluster with the most favorable packing score were taken from the three most common clusters generated using backbone 1 (CHAMPs 2 to 4). Only one cluster from backbone 2 had a low Rosetta energy and packing score and we chose this sequence for experimental evaluation (CHAMP #5). The sequences for peptides generated using Protocol 2 are shown in Table.1.

CHAMP peptides activate $\alpha_5\beta_1$ and not $\alpha_V\beta_3$

As a primary test, HAEC were incubated with increasing concentrations of either CHAMP #1 and the anti-α_v CHAMP peptide, after which $\alpha_5\beta_1$ activation was assessed using the $\alpha_5\beta_1$ ligand mimetic GST-FNIII$_{9-11}$, as previously described (*33, 39*). Although GST-FNIII$_{9-11}$ (a recombinant, GST-tagged fusion protein containing then 9th through 11th type III repeat of fibronectin) can interact with several integrins (*40*), the soluble recombinant protein interacts specifically with activated $\alpha_5\beta_1$ integrins, similar to endogenous fibronectin (*33, 41*). CHAMP peptide #1 robustly induced GST-FNIII$_{9-11}$ binding to HAEC at both 4 μM and 8 μM concentrations as assessed by Western blotting for the GST-tag (Figure 3.2). By contrast, anti-α_v CHAMP peptide (8 μM) did not activate $\alpha_5\beta_1$. Similarly, CHAMPs #4, generated using Protocol 2, activated $\alpha_5\beta_1$, although the magnitude of GST-FNIII$_{9-11}$ induced by CHAMP #1 was consistently greater (Figure 3.2). Finally,

113

CHAMP #2, CHAMP #3, and CHAMP #5 caused only a mild increase in GST-FNIII$_{9-11}$ binding, with only that by CHAMP #2 reaching statistical significance (Figure 3.3).

To verify the specificity of anti-α_5 CHAMP-induced $\alpha_5\beta_1$ activation, HAEC were incubated with increasing concentrations of CHAMP #1, CHAMP #4, and anti-α_v CHAMP, and $\alpha_v\beta_3$ activated was measured using the ligand-mimetic antibody WOW-1. WOW-1 is an engineered $\alpha_v\beta_3$ ligand in which the 19 amino acid H-CDR3 of the $\alpha_{IIb}\beta_3$-specific activation-dependent PAC-1 Fab were replaced by the 50 amino acid α_v binding domain of the adenovirus type 2 penton base, switching the specificity of the Fab from $\alpha_{IIb}\beta_3$ to the α_v-containing integrins, $\alpha_v\beta_3$ and $\alpha_v\beta_5$ (*42*). WOW-1 binding to the HAEC was only observed when the cells were incubated with anti-α_v CHAMP peptides and then only at high peptide concentrations (Figure 3.3). Taken together, these data indicate that the CHAMP #1 and CHAMP #4 peptides strongly induce $\alpha_5\beta_1$ activation in HAEC and do so without activating $\alpha_v\beta_3$.

CHAMP peptides #1 and #4 associate with the α₅ TM domain in micelles

To measure the association of CHAMP peptides #1 and #4 with a peptide fragment of the α_5 TM in vitro, we conducted thiol-disulfide exchange equilibrium. First, peptides were synthesized, including β_1 and α_5 TM fragments, with a N-terminal cysteine followed by a flexible triple glycine linker (Figure 3.4 and Figure 3.5). As expected, the peptides were pre-dominantly helical in DPC micelles at a detergent to peptide ratio of 100:1 (Figure 3.6). Mixtures of peptides in DPC micelles were brought to an equilibrium of monomeric and dimeric species in redox buffer in a two-step process of helix association followed by thiol oxidation (Figure 3.5). Covalent species were then separated and measured by analytical HPLC (Figure 3.7) and the relative ratios of the monomers, as well as the homo-dimeric and hetero-dimeric species were quantified as shown previously (*20,*

21) (Fig. 3.4, 3.5). Under these conditions, if the peptides are primarily monomeric then the primary products are the free peptide thiol and a mixed disulfide between glutathione and the cys-containing peptide. On the other hand, if the peptides dimerize in a geometry conducive to forming a disulfide, the primary product is a disulfide-bonded dimer. Additionally, if two different peptides are present in equimolar amounts, then the mole fraction of the homo- versus hetero-dimer provides an indication of the preference for homo- versus hetero-dimer formation. In the limit where the association constant for the homo- and hetero-dimers are the same, the ratio of the peaks would be 1:2:1. Deviations from this ratio indicate differences in the affinities of the individual homo- and hetero-dimeric species.

The β_1 TM peptide had a slight propensity to form homo-dimers, in agreement with previous TOXCAT bacterial reporter assay experiments (*27*). Both CHAMP peptides had strong propensities for homo-dimerization as well. In particular, CHAMP #1 demonstrated a marked tendency to self-associate. Although the CHAMPs and β_1 self-associated, they also formed a hetero-dimers with the α_5 TM peptide. Moreover, the association of CHAMP #4 with the α_5 TM hetero-dimers was similar to the association of α_5 with β_1, while all CHAMPs decreased the equilibrium concentration of the α_5 to the same degree as β_1. This behavior is consistent with a modest to weak association of either β_1 or the CHAMPs to the α_5 TM fragment.

Solution NMR Titration of CHAMPs and α_5 and β_1 TM constructs in bicelles

To determine the how the CHAMP #1 and #4 peptides, caused $\alpha_5\beta_1$ activation, we titrated either ^{15}N-labeled α_5 or β_1 TM and cytoplasmic domain (TM-CT) constructs in 20% (w/v) DHPC/DMPC bicelles with unlabeled CHAMP peptides and measured changes in previously characterized ^{1}H-^{15}N TROSY-HSQC spectra of the integrin(*36*). Figure 3.8 shows an overlay of

the ^{1}H-^{15}N HSQC spectra of α_5 in isolation and after the addition of 8 molar equivalents of unlabeled CHAMP #1. Even at the highest equivalent of CHAMP #1, only minor changes in either chemical shift or peak intensity can be observed for all resonances in the α_5 TM-CT spectra, suggesting negligible interaction. The full titration series, including 0.5, 1, 2, 4, 6 and 8 molar equivalents of CHAMP #1, is shown in Figure 3.9. However, drastic changes are observed in the β_1 TM-CT ^{1}H-^{15}N HSQC spectra upon addition of 1 molar equivalent of CHAMP #1 (Fig. 3.9) which become more pronounced at 3 molar equivalent (Fig. 3.8), indicative of a tight association. In the β_1 spectra, the peak intensities of many resonances previously assigned to TM domain residues are markedly decreased, consistent with association-dissociation events between the β_1 TM-CT and CHAMP #1 within the slow-exchange regime on the NMR time scale. This same behavior was observed upon mixing of integrin α_5 TM-CT with β_1 TM-CT (*36*). Thus, since CHAMP #1 binds more strongly to the integrin β_1 TM domain than to integrin α_5 TM-CT, it is likely that the CHAMP #1 causes $\alpha_5\beta_1$ activation in cells by binding the β_1 TM domain, thereby the inhibiting the resting $\alpha_5\beta_1$ TM hetero-dimer.

The similar behavior of CHAMP #4 and #1 in the thiol-disulfide equilibrium assay suggests CHAMP #4 might also activate $\alpha_5\beta_1$ in a similar manner. As shown in Figure 3.10, there were slight decreases in peak intensity in the ^{1}H-^{15}N TROSY-HSQC of both α_5 and β_1 TM-CT with the addition of CHAMP #4 at 0.5, 1, 2, 3, 4 molar ratios, suggesting a modest association of the peptide to both α_5 and β_1 TM-CT constructs. For the β_1 TM-CT, decay of peak intensities for the TM domain residues were plotted and globally fit to a simple equimolar binding model. This is shown in Figure 3.10 for the W751 indole proton (peak 1) and resonances of L749, G744 and V736 amide protons. Similarly, eight resonances in the α_5 spectra thought to be from TM domain residues were tracked and fit (Figure 3.10), although these peaks have not yet been assigned. The peak decay fits

give equilibrium dissociation constants of 344 ± 33 µM (peptide/detergent 1.6 mol%) and 435 ± 53 µM (2.03 mol%) for α_5 TM-CT and β_1 TM-CT respectively. This is an order of magnitude weaker than the dissociation constant of 0.17 ± 0.1 mol% similarly measured for the integrin α_5 TM-CT and β_1 TM-CT complex previously (*36*).

Discussion

Reagents that specifically bind to one integrin TM helix provide a way to interrogate the paradigm of integrin activation that is heavily based on the behavior of the platelet integrin $\alpha_{IIb}\beta_3$. However, the behavior of $\alpha_{IIb}\beta_3$ may not be typical of most integrins. For example, $\alpha_1\beta_1$ and $\alpha_2\beta_1$ activation may be less regulated and less dependent on a TM domain hetero-dimer to constrain their activity (*36*). CHAMP peptides that specifically bind to one integrin helix in situ provide a tool for studying the general role of TM domain interactions in regulating integrin function.

Our goal here was to specifically target the integrin $\alpha_5\beta_1$ by designing CHAMP peptides that bind to the α_5 TM domain. To this end, our design strategy was successful in that we generated and characterized two novel CHAMP peptides that specifically activate the $\alpha_5\beta_1$ expressed by human aortic endothelial cells and not the co-expressed $\alpha_v\beta_3$. Nonetheless, subsequent biophysical characterization of the two most active CHAMP peptides indicated that they do not bind specifically to α_5, but instead bind either to the β_1 TM domain specifically or both α_5 and β_1 TM domains. CHAMP #1 was found by solution NMR in bicelles to interact strongly with the β_1 TM domain but had negligible interaction with α_5 and it preferentially self-associated in the presence of α_5 in DPC micelles. Thus, by binding the β_1 TM domain, CHAMP #1 likely activates $\alpha_5\beta_1$ by competitively dissociates the TM domain hetero-dimer of inactive $\alpha_5\beta_1$. CHAMP #4 interacted modestly with both the α_5 and β_1 TM domains in the NMR titration assay with a slight preference

for α_5 in bicelles. In DPC micelles, it bound equally well to both the α_5 and β_1 TM domains. Thus, the molecular mechanism by which CHAMP #4 activates $\alpha_5\beta_1$ is less clear. It is possible that at high concentrations, the peptide accumulates in the cell membrane relative to $\alpha_5\beta_1$ and competitively inhibits the $\alpha_5\beta_1$ TM hetero-dimer given its modest affinity. Nonetheless, because neither CHAMP peptide activates $\alpha_V\beta_3$, even at high concentrations, it is clear that their interaction with the α_5 and β_1 G/A/S-X$_3$-G/A/S TM helix motifs is specific. However, the link between biological potency and binding affinity is convoluted and difficult to interpret given the many factors at play: peptide delivery and insertion into the membrane, peptide topology upon insertion, and promiscuous binding to other membrane proteins. Regardless, the CHAMPs designed here should be useful reagents to probe the biological role of TM domain interactions in integrin biology. In particular, the use of CHAMP #1 to specifically target β_1-containing integrins across distinct cell types holds the most promise.

The cellular activity and the biophysical properties of our peptides emphasize key sequence-structure relationships within the G/A/S-X$_3$-G/A/S motif critical to CHAMP design outcome. Principally, it is thought that the composition and sequence context of a G/A/S-X$_3$-G/A/S motif provides limits on how strongly or weakly it interacts with complementary sequences. Motifs with strong homo- and hetero-dimerization propensities most commonly contain two Gly residues (G-X$_3$-G) surrounded by either beta-branched or additional small residues (*16, 24, 43-45*). It is also critical for the second small residue to have the potential to donate a side chain or Cα proton hydrogen bond across the helix (small-X$_3$-G or G-X$_3$-S) (*46*). The α_5 TM domain has a Leu-rich A-X$_3$-G motif containing a bulky Phe residue unique within the α-subunit family (L-<u>A</u>-I-L-<u>F</u>-<u>G</u>-L, Fig. 3.1). Accordingly, we would expect close packing at this interface to be difficult, making it a difficult sequence to target. Consistent with this expectation, we found that the association

between the α_5 TM and CHAMP constructs to be weak. For example, CHAMP #5, which contained a single A-X$_3$-A motif that could not donate a hydrogen bond in the GAS$_{Right}$ geometry, was the only peptide that did not to induce detectable $\alpha_5\beta_1$ activation (Table 3.1). On the other hand, the β_1 TM domain has features consistent with strong dimerization and potential promiscuity: a G-X$_3$-G motif, preceded by additional small residues and flanked on each side by beta-branched amino acids (V-A-G-V-V-A-G-I). Likewise, CHAMP #1 has 4 small-X$_3$-small motifs, including adjacent strong G-X$_3$-G and G-X$_3$-S motifs (L-G-G-L-I-G-S-L,). Thus, CHAMP #1 had a strong propensity to self-associate in micelles and bound to the β_1 G-X$_3$-G motif at near-stoichiometry. The unanticipated binding of CHAMP #1 and #4 to β_1 suggest that much more rigorous modeling and energetic analysis will be required to overcome the inherent dimerization potential of strong small-X$_3$-small motifs to achieve specificity for difficult targets like α_5.

In addition to designing functional anti-α_5 CHAMP peptides, our goal here was to test whether modifications to the de novo design protocol provided advantages over the original algorithm. Given that no specific anti-α_5 peptide was identified, it is difficult to attribute changes in the computational algorithm to positive or negative biochemical outcomes. In the future, it will be important to implement changes in the potential function and introduce multi-state negative design. Here, we included the RosettaMP potential function, which was derived and trained on structure prediction a decade ago, for the design of CHAMPs #2-5 (*28*). Subsequent to designing our sequences, there was a report detailing the poor performance of RosettaMP and other potential functions on benchmark tests for TM proteins, basic tasks typically passed by standard Rosetta scoring functions for water-soluble proteins (*47*). However, others have supplemented RosettaMP with orientation restraints to achieve unprecedented success in TM homo-dimer structure prediction, utilizing protein family sequence analysis, residue entropy, and co-evolution (*48*).

Fortunately, efforts to revise and reparametrize the potential function used by RosettaMP has been recently reinvigorated. Alongside the imperative published by Kroncke *et al.* (*47*), the Fleishman group recently focused on the depth-dependent insertion energy of amino acids in TM helices by deriving a statistical function from deep mutational scanning experiments with Glycophorin A and ErbB2 (*49*), which was subsequently benchmarked for topology prediction (*50*). Meanwhile, simpler potential functions, such as PREDIMER (*51*) and CATM (*46*), capture critical structural features and achieve reasonable prediction of structure and energetics of TM homo-dimers utilizing just a polar surface complementation term (PREDIMER) or a two term function comprised of van der Waals and hydrogen bonds, including Cα-H donors (CATM). Despite limitations in modeling and energetic analysis of TM protein structure, exploring and benchmarking alternative potential functions holds great promise, especially for design applications such as CHAMP.

Given the current promiscuity of CHAMP binding, we hope to prevent unexpected interactions by implementing a multi-state modeling scheme to enable negative design. Specifically, we plan to explicitly model and evaluate of the energy gap between ensembles of the desired state (e.g., CHAMP with α_5) and alternative undesired states (i.e., CHAMP with β_1 and CHAMP homo-dimers). Thereafter, mutations can be introduced to destabilize the undesired models without compromising the desired CHAMP hetero-dimer. The utility of this extra modeling stage can be directly evaluated by experiment. However, the accuracy and relevance of energy gaps calculated between models depends on the quality of the potential function, again highlighting the importance of potential functions that perform well for future TM protein modeling and design. Nonetheless, we anticipate that inclusion of these steps will advance the

CHAMP algorithm towards routinely achieving specificity among difficult targets from closely related protein families such as integrins.

In summary, executing the CHAMP design protocol has provided structural insights into engineering protein-protein interaction in the membrane and the computational methodology to do so. Since pairwise interactions of α-helices serve as tertiary building blocks of TM proteins (*18, 23*), this type of *de novo* TM dimer design provides a forum for testing principles of TM protein folding. Likewise, CHAMP design, being the simplest TM protein design application, provides a stepping-stone towards the design of complex and functional multi-pass and higher-order oligomeric TM proteins.

Materials and Methods

Computational CHAMP Design

A schematic showing the general steps of the CHAMP design workflow is shown in Figure 3.1. Two approaches were taken in this study: Protocol 1 mimics Yin et al. (*22*); Protocol 2 has pre-determined criteria to enable automated computer-based decision-making. In both cases, we first defined the approximate preferred position of the target integrin sequence in a bilayer. An atomic model for the α_5 target was built by threading the α_5 TM domain sequence onto an idealized alpha-helix and building side chain coordinates from a rotamer library. In protocol 1, side chains were built manually and oriented by the EZ-potential (*17*). In protocol 2, side chains were built through rounds of side chain repacking, rigid body motions, and Cartesian minimization, all relative to an implicit bilayer with a depth-dependent dielectric solvation model using the Rosetta membrane protein framework (RosettaMP) (*19*).

Next, a dimer model was built by extracting coordinates from a geometrically clustered database of pairs of α-helices from a non-redundant set of natural TM protein x-ray structures (*23*). The CHAMP was designed to bind the target integrin TM domain at its conserved G/A/S-X$_3$-G/A/S-X$_3$-Leu motif. These sequence motifs commonly associate in a "GAS$_{right}$" geometry with parallel right-handed crossing angle (-35º) and close inter-helical distance (8.3 Å) (*23-25*).

In protocol 1, the helix dimer backbone coordinates were selected by choosing a subset of the helical pairs from the PDB with this geometry. Specifically, a subset of the parallel right-handed dimers from cluster #4 was chosen based on the criterion that they should have inter-helical distances less than one standard deviation from the mean inter-helical distance of all the members within the cluster (n=109, mean=8.3 Å) in order to facilitate close backbone packing and interaction within the helix dimer. All candidates were then culled to eliminate those with insufficient helix length, steric clashes with integrin α$_5$ side chains, unusual conformations, and non-α helical backbone hydrogen bonding patterns.

The protocol 1 sequence design was conducted by first fixing residues of non-interfacial positions to Val to simplify rotamer selection and interface positions to Gly as a starting sequence. Only residues at the dimer's interface were sampled. Given that the dielectric environment within protein cores and the lipid bilayer are similar, rotamer trials were conducted using Rosetta with the Talaris2013 score function modified to utilize only the Lennard-Jones and rotamer energies, with equal weights (*26*). Lowest energy designs were parsed and selected by visualization for inter-helical packing. Finally, the non-interfacial residues were selected semi-randomly: with a 0.6 chance of Leu and 0.1 chance of each Ala, Val, Phe, and Ile. Terminal Lys residues were added or mutated to aid solubility, synthesis, and purification. In protocol 2, the dimer model was selected by a series of automated criteria. First, the cluster (n=109 members) was searched for helices with

a sequence matching the pattern: [G/A/S]-X_2-[L/F/Y/W/M]-[G/A/S]-X_3-[L/F/V/I/M] where X is variable. This pattern was used to capture the sterics of the unique Phe1008 of the integrin α_5 TM domain, which is usually a small residue within the α-subunit family (Fig. 3.1) (27). Furthermore, it was required that template backbones were in van der Waals contact (Cα distances of <5.0 Å) at the positions of the small Ala and Gly residues. Also, the signature Phe residue's Cα was required to be within 7 Å of a non-Ala/Gly Cα atom on the designed helix. Next, the coordinates of helical pairs matching the motif were superimposed onto the modeled TM domain of the target so that the query and target sequence motifs overlap. The helical pairs were extended by superimposing 3-residue fragments in an idealized alpha-helical conformation until both helices spanned the implicit bilayer (length = 34 Å). Helices that were angled such that they required >32 residues to span the membrane were discarded.

The protocol 2 sequence design was conducted by only selecting interfacial residues. However, here we applied RosettaMP and the standard TM full-atom energy function (28). Interface residues were selected through the FastRelax application implemented in RosettaScripts, (29) incorporating rounds of simulated annealing with rigid body re-orientation in the bilayer, rotamer trials, and Cartesian minimization. For each dimer scaffold, 800 design trajectories were conducted. In addition to the calculated energy score, each model was evaluated for void volume in the helix interface by the Rosetta PackStat score as described previously (30). Additionally, the individual helices of each model were separated by 40 Å and re-scored after FastRelax trajectories to allow calculation of $\Delta E_{association}$ in Rosetta energy units:

$$\Delta E_{Association} = E_{Dimer} - (E_{CHAMP} + E_{\alpha5}) \qquad \text{Equation 1}$$

in which E_{dimer} is the Rosetta energy of the helices modeled as a dimer, and E_{CHAMP} and $E_{\alpha5}$ are the Rosetta energies of those monomeric α-helical peptides separated by 40 Å. Finally, unique

sequences for each backbone were independently clustered hierarchically with complete linkage using a normalized BLOSOM sequence similarity score as the distance function with a matrix corresponding to the average pairwise sequence identity of all designs, ca. 80% (e.g. BLOSOM80), using BioPython (*31, 32*). Of the four most populated clusters, the sequence with the highest PackStat score was selected for synthesis (additionally, sequences with a tryptophan in the apolar membrane region were not prepared as they might not insert properly into bilayers). Non-interfacial residues were randomized between A, L, I, V, or F as in protocol 1. Flanking lysine residues were added to aid synthesis, purification, and solubility, and one tryptophan was added for spectroscopic detection.

Peptide Synthesis and Purification

Peptides were synthesized at 0.1 mmol scale using a Biotage Initiator+ Alstra automated microwave peptide synthesizer on preloaded Rink-amide resin (Chem-Impex). Standard Fmoc deprotection was performed twice at 70°C for 5 min. Coupling reactions were performed twice for 5 minutes at 75°C using Fmoc-protected amino acids (5 equivalent, 0.5 M, Chem-Impex), HCTU (4.95 equivalent, 0.5 M) and DIEA (10 equivalent, 0.5 M) in DMF. N-terminal acetylation was carried out at room temperature in the presence of 10 equivalent of acetic anhydride and 20 equivalents of DIEA in DMF. Peptide cleavage was performed on the dried resin using TFA-TIPS-H$_2$O (95:2.5:2.5) for 2 hr at 22°C ; in the presence of Cys or Met, the cleavage cocktail was changed to TFA-TIPS-H$_2$O-EDT (92.5:2.5:2.5:2.5). The crude peptide mixture was dried using N$_2$ gas, precipitated with cold ether, dried again under N$_2$ gas, and dissolved in a 1:1 mixture of solvents A and B (Solvent A: 0.1% TFA in H$_2$O; solvent B: 0.1% TFA in 60% isopropanol, 30% CH$_3$CN and 10% H$_2$O). The mixture was purified by RP-HPLC (Higgens 300 Å C4 column, 10 µm, 10 x

50 mm) with a flow rate of 5 mL/min over a linear gradient of 60% to 100% of buffer B over 40 min. Peptide mass was analyzed using MALDI (Shimadzu Axima) and electrospray ionization mass spectrometry (Qtrap 3200, ABSCI EX). Purity was determined by analytical HPLC. All peptides used were at least 95% pure. Organic solvents were purchased from Sigma.

Cell Culture

Human aortic endothelial cells (HAEC) were purchased at passage 3 (Lonza) and cultured in MCDB131 media supplemented with 10% fetal bovine serum (FBS), 60 µg/ml heparin sodium, 25 µg/ml bovine brain extract (isolated from bovine hypothalamus, Pel-Freeze), 10 U/ml penicillin, and 100 µg/ml streptomycin. Cells were used between passage 6 and 10, and experiments performed in low serum (0.5% FBS) media.

Integrin Activation Assay

A fusion protein composed of the 9th to 11th type III repeat of fibronectin (GST-FNIII$_{9-11}$) and a glutathione S-transferase (GST) tag employed as a ligand mimetic for $\alpha_5\beta_1$ as previously described (*33*). The activation state-sensitive anti-$\alpha_v\beta_3$ antibody His-WOW-1 (a gift of Sanford Shattil, University of California, San Diego) was used to measure $\alpha_v\beta_3$ activation. Briefly, cells were stimulated with CHAMP peptides in media contain either 20 µg/ml GST-FNIII$_{9-11}$ or 20 µg/ml His-WOW1 at 37°C for 30 minutes. The cells were then washed to remove unbound ligand, lysed in SDS sample buffer, and integrin activation assessed by Western blotting for GST (GST-FNIII$_{9-11}$) or His (His-WOW-1).

Immunoblotting

Cell lysis and immunoblotting was performed as previously described(*34*). Antibodies included mouse anti-GST (Santa Cruz Biotechnology, Cat# sc-138, 1:1000 dilution), mouse anti-His (Thermo Fisher, Cat# R930-25, 1:1000 dilution), mouse anti-β actin (Cat# sc-47778, 1:2500 dilution), and rabbit anti-GAPDH (Cell Signaling Technology, Cat# 5174, 1:5000 dilution).

Disulfide Exchange Equilibrium in Micelles

Sample preparation and thiol disulfide exchange were performed as previously described (*21, 35*). For these experiments, peptides were synthesized with an N-terminal Cys residue followed by a three residue Gly linker (e.g. CGGG-R). Thiol-containing peptides and detergent dodecylphosphocholine (DPC) were co-dissolved in ethanol with a peptide to detergent/lipid ratio of 1:100, dried under N_2 gas, and left under vacuum overnight. Samples were re-hydrated with thiol exchange buffer (100mM Tris–HCl pH 8.6, 100 mM KCl and 1 mM EDTA) and incubated for 4 hours under reversible redox conditions in a glutathione buffer containing 0.45 mM oxidized (GSSG) and 1.05 mM reduced (GSH) glutathione before quenching with HCl (final 0.12 M). The mixture was separated by analytical reverse phase HPLC (Vydac™ 214TP C4 Column). Homo-dimer and target-CHAMP hetero-dimer peaks were identified by comparing the HPLC spectrum with samples containing only one peptide monomer. Overlapping peaks were analyzed with OriginPro 8 (OriginLab, Northampton, MA) multipeak fitting.

NMR Spectroscopy

Construction of the α_5 and β_1 TM and cytoplasmic (TM-CT) domain fragment plasmid, expression, purification, and NMR sample preparation were performed as previously described

(*36*). The final NMR samples contained 20% (w/v) DHPC/DMPC q=0.3 bicelles, 1 mM EDTA and 10% v/v D_2O in 50 mM $NaPO_4$ at pH 6.5, in which the integrin concentration was 150 uM. The appropriate amount of 15 mM CHAMP peptide stock solution in the same NMR buffer was added to the NMR tube to make a series of mixtures of either integrin α_5 or β_1 with the CHAMP peptide in CHAMP:integrin molar ratios ranging from 0 to 8. An $^1H^{15}N$-TROSY-HSQC spectrum was measured for each titration point at 45° C on a BRUKER 800 MHz spectrometer equipped with cryogenic triple-resonance probes with z-axis pulsed field gradients. The data were processed using NMRPIPE (*37*) and analyzed using NMRVIEWJ (*38*). The observed decay in peak intensities for several resonances over the titration series were fit to a simple equimolar binding model as previously described (*36*).

Contributions

M.M., H.H., J.S.B., A.W.O., and W.F.D. conceived of and designed the project. M.M. and H.H. performed the computational design and synthesized the designed peptides. H.H. performed the circular dichroism and thiol-disulfide exchange experiments. A.W.O. performed the cell culture

integrin activation experiments. Z. L. performed NMR experiments. Z. L. and C.S.S. analyzed the

NMR data. All authors contributed to the writing of the manuscript.

References

1.   D. M. Engelman *et al.*, Membrane protein folding: beyond the two stage model. *Febs Letters* **555**, 122-125 (2003).

2.   H. Yin, A. D. Flynn, Drugging membrane protein interactions. *Annual review of biomedical engineering* **18**, 51-76 (2016).

3.   T. A. Stone, C. M. Deber, Therapeutic design of peptide modulators of protein-protein interactions in membranes. *Biochim Biophys Acta* **1859**, 577-585 (2017).

4.   A. Bennasroune *et al.*, Transmembrane peptides as inhibitors of ErbB receptor signaling. *Molecular biology of the cell* **15**, 3464-3474 (2004).

5.   B. E. Poulsen, C. M. Deber, Drug efflux by a small multidrug resistance protein is inhibited by a transmembrane peptide. *Antimicrobial agents and chemotherapy* **56**, 3911-3916 (2012).

6.   A. Fink *et al.*, Assembly of the TLR2/6 transmembrane domains is essential for activation and is a target for prevention of sepsis. *The Journal of Immunology* **190**, 6410-6422 (2013).

7.   T. J. Cammett *et al.*, Construction and genetic selection of small transmembrane proteins that activate the human erythropoietin receptor. *Proceedings of the National Academy of Sciences* **107**, 3447-3452 (2010).

8.   L. L. Freeman-Cook *et al.*, Selection and characterization of small random transmembrane proteins that bind and activate the platelet-derived growth factor beta receptor. *J Mol Biol* **338**, 907-920 (2004).

9.   E. N. Heim *et al.*, Biologically active LIL proteins built with minimal chemical diversity. *Proc Natl Acad Sci U S A* **112**, E4717-4725 (2015).

10.   G. A. Caputo *et al.*, Computationally Designed Peptide Inhibitors of Protein− Protein Interactions in Membranes†. *Biochemistry* **47**, 8600-8606 (2008).

11.   H. Yin *et al.*, Computational design of peptides that target transmembrane helices. *Science* **315**, 1817-1822 (2007).

12.   K. P. Fong *et al.*, Directly Activating αIIbβ3 Initiates Outside-In Signaling by Causing αIIbβ3 Clustering. *Journal of Biological Chemistry*, jbc. M116. 716613 (2016).

13.   A. Yurdagul, Jr. *et al.*, alpha5beta1 integrin signaling mediates oxidized low-density lipoprotein-induced inflammation and early atherosclerosis. *Arterioscler Thromb Vasc Biol* **34**, 1362-1373 (2014).

14.   J. Chen *et al.*, alphavbeta3 Integrins Mediate Flow-Induced NF-kappaB Activation, Proinflammatory Gene Expression, and Early Atherogenic Inflammation. *Am J Pathol* **185**, 2575-2589 (2015).

15.   S. Yun *et al.*, Interaction between integrin alpha5 and PDE4D regulates endothelial inflammatory signalling. *Nat Cell Biol* **18**, 1043-1053 (2016).

16.   M. G. Teese, D. Langosch, Role of GxxxG Motifs in Transmembrane Domain Interactions. *Biochemistry* **54**, 5125-5135 (2015).

17.   A. Senes *et al.*, E(z), a depth-dependent potential for assessing the energies of insertion of amino acid side-chains into membranes: derivation and applications to determining the orientation of transmembrane and interfacial helices. *J Mol Biol* **366**, 436-448 (2007).

18.   R. F. Walters, W. F. DeGrado, Helix-packing motifs in membrane proteins. *Proc Natl Acad Sci U S A* **103**, 13658-13663 (2006).

19.   R. F. Alford *et al.*, An integrated framework advancing membrane protein modeling and design. *PLOS Comput Biol* **11**, e1004398 (2015).

20.   B. North *et al.*, Characterization of a membrane protein folding motif, the Ser zipper, using designed peptides. *Journal of molecular biology* **359**, 930-939 (2006).

21.   Y. Zhang, D. W. Kulp, J. D. Lear, W. F. DeGrado, Experimental and computational evaluation of forces directing the association of transmembrane helices. *Journal of the American Chemical Society* **131**, 11341-11343 (2009).

22.   H. Yin *et al.*, Computational design of peptides that target transmembrane helices. *Science* **315**, 1817-1822 (2007).

23.   S. Q. Zhang *et al.*, The membrane- and soluble-protein helix-helix interactome: similar geometry via different interactions. *Structure* **23**, 527-541 (2015).

24.   A. Senes, M. Gerstein, D. M. Engelman, Statistical analysis of amino acid patterns in transmembrane helices: the GxxxG motif occurs frequently and in association with beta-branched residues at neighboring positions. *J Mol Biol* **296**, 921-936 (2000).

25.   D. T. Moore, B. W. Berger, W. F. DeGrado, Protein-protein interactions in the membrane: sequence, structural, and biological motifs. *Structure* **16**, 991-1001 (2008).

26.   A. Leaver-Fay *et al.*, ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods in enzymology* **487**, 545 (2011).

27.   B. W. Berger *et al.*, Consensus motif for integrin transmembrane helix association. *Proceedings of the National Academy of Sciences* **107**, 703-708 (2010).

28.   P. Barth, J. Schonbrun, D. Baker, Toward high-resolution prediction and design of transmembrane helical protein structures. *Proc Natl Acad Sci U S A* **104**, 15682-15687 (2007).

29.   S. J. Fleishman *et al.*, RosettaScripts: a scripting language interface to the Rosetta macromolecular modeling suite. *PLoS One* **6**, e20161 (2011).

30. W. Sheffler, D. Baker, RosettaHoles2: a volumetric packing measure for protein structure refinement and validation. *Protein Science* **19**, 1991-1995 (2010).

31. S. Henikoff, J. G. Henikoff, Amino acid substitution matrices from protein blocks. *Proceedings of the National Academy of Sciences* **89**, 10915-10919 (1992).

32. P. J. Cock *et al.*, Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**, 1422-1423 (2009).

33. A. W. Orr, M. H. Ginsberg, S. J. Shattil, H. Deckmyn, M. A. Schwartz, Matrix-specific suppression of integrin activation in shear stress signaling. *Mol Biol Cell* **17**, 4686-4697 (2006).

34. A. W. Orr *et al.*, The subendothelial extracellular matrix modulates NF-kappaB activation by flow: a potential role in atherosclerosis. *J Cell Biol* **169**, 191-202 (2005).

35. L. Cristian, J. D. Lear, W. F. DeGrado, Use of thiol-disulfide equilibria to measure the energetics of assembly of transmembrane helices in phospholipid bilayers. *Proceedings of the National Academy of Sciences* **100**, 14772-14777 (2003).

36. Z. Lu *et al.*, Implications of the differing roles of the beta1 and beta3 transmembrane and cytoplasmic domains for integrin function. *Elife* **5**, (2016).

37. F. Delaglio *et al.*, NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *Journal of biomolecular NMR* **6**, 277-293 (1995).

38. B. A. Johnson, From Raw Data to Protein Backbone Chemical Shifts Using NMRFx Processing and NMRViewJ Analysis. *Methods Mol Biol* **1688**, 257-310 (2018).

39. P. E. Hughes *et al.*, Suppression of integrin activation: a novel function of a Ras/Raf-initiated MAP kinase pathway. *Cell* **88**, 521-530 (1997).

40. M. K. Magnusson, D. F. Mosher, Fibronectin: structure, assembly, and cardiovascular implications. *Arterioscler Thromb Vasc Biol* **18**, 1363-1370 (1998).

41. S. Huveneers, H. Truong, R. Fassler, A. Sonnenberg, E. H. Danen, Binding of soluble fibronectin to integrin alpha5 beta1 - link to focal adhesion redistribution and contractile shape. *J Cell Sci* **121**, 2452-2462 (2008).

42. N. Pampori *et al.*, Mechanisms and consequences of affinity modulation of integrin alpha(V)beta(3) detected with a novel patch-engineered monovalent ligand. *J Biol Chem* **274**, 21609-21616 (1999).

43. W. P. Russ, D. M. Engelman, The GxxxG motif: a framework for transmembrane helix-helix association. *J Mol Biol* **296**, 911-919 (2000).

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

45. S. M. Anderson, B. K. Mueller, E. J. Lange, A. Senes, Combination of Calpha-H Hydrogen Bonds and van der Waals Packing Modulates the Stability of GxxxG-Mediated Dimers in Membranes. *J Am Chem Soc* **139**, 15774-15783 (2017).

46. B. K. Mueller, S. Subramaniam, A. Senes, A frequent, GxxxG-mediated, transmembrane association motif is optimized for the formation of interhelical Cα–H hydrogen bonds. *Proceedings of the National Academy of Sciences* **111**, E888-E895 (2014).

47. B. M. Kroncke *et al.*, Documentation of an Imperative To Improve Methods for Predicting Membrane Protein Stability. *Biochemistry* **55**, 5002-5009 (2016).

48. Y. Wang, P. Barth, Evolutionary-guided de novo structure prediction of self-associated transmembrane helical proteins with near-atomic accuracy. *Nat Commun* **6**, 7196 (2015).

49.   A. Elazar *et al.*, Mutational scanning reveals the determinants of protein insertion and association energetics in the plasma membrane. *Elife* **5**, (2016).

50.   A. Elazar, J. J. Weinstein, J. Prilusky, S. J. Fleishman, Interplay between hydrophobicity and the positive-inside rule in determining membrane-protein topology. *Proceedings of the National Academy of Sciences* **113**, 10340-10345 (2016).

51.   A. A. Polyansky, P. E. Volynsky, R. G. Efremov, Multistate organization of transmembrane helical protein dimers governed by the host membrane. *J Am Chem Soc* **134**, 14390-14400 (2012).

a Integrin α-subunit family, G/A/S-X$_3$-G/A/S-X$_3$-L

Target: α$_5$ – ^{1003}ILAILFGLLLLGLLIYI^{1019}

α$_V$ – ^{998}ILAVLAGLLLLAVLVFV^{1014}

α$_{IIb}$– ^{1001}LVGVLGGLLLLTILVLA^{1017}

b

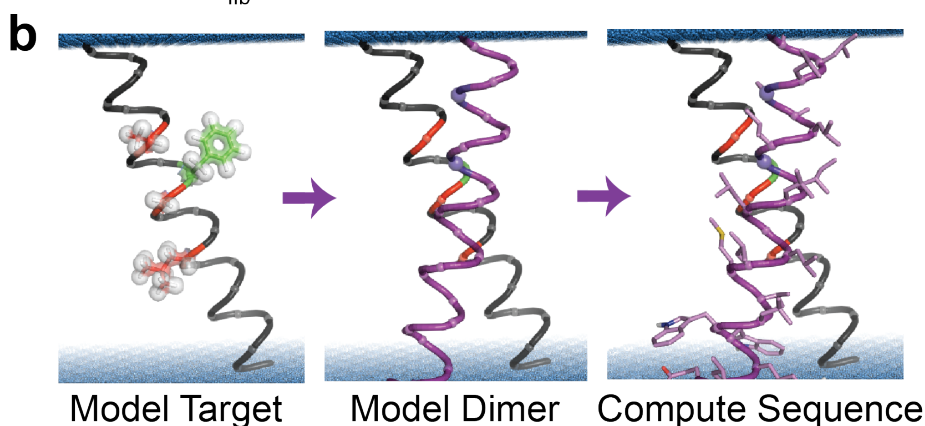Model Target   Model Dimer   Compute Sequence

Figure 3.1. CHAMP peptide design strategy. (**a**) First the intended interface of the target is defined, here the conserved A-xxx-G-xxx-L motif of the integrin α-subunit TM domain, colored red in the sequences and the weblogo for human isoforms. Integrin α$_5$ has a unique large residue (Phe, green) in a residue position typically having a small residue. (**b**) Next an atomic model of the monomer is built, a second helix is selected from a database of pairs of associated helices extracted from natural membrane protein X-ray structures and modeled. Finally, the sequence of the CHAMP is sampled and selected by rotamer trials given a potential energy function.
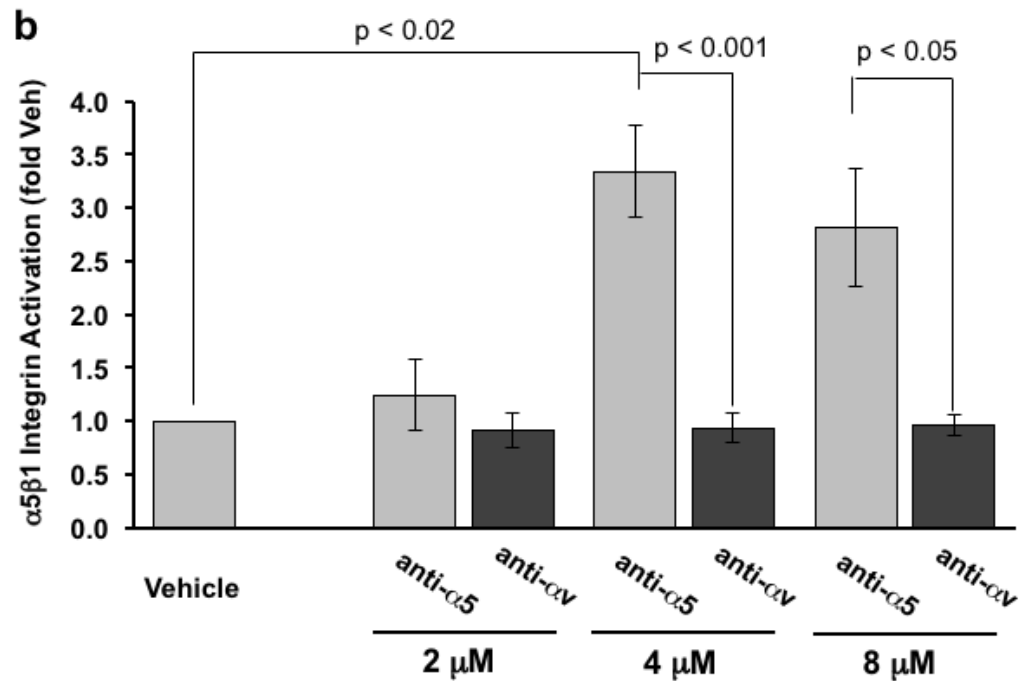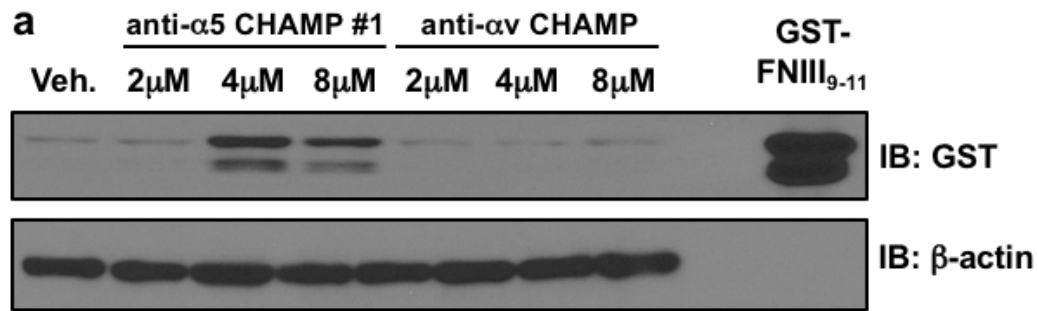
Figure 3.2. CHAMP #1 induces activation of $\alpha_5\beta_1$, increasing cell adhesion to GST-Fibronectin ligand mimetic in endothelial cells. Human aortic endothelial cells were treated for 30 minutes with the indicated doses of anti-α_5 CHAMP peptides, anti-α_v CHAMP peptides, or vehicle (ethanol) control in the presence of the $\alpha_5\beta_1$ ligand mimetic GST-FNIII$_{9-11}$. Retention of GST-FNIII$_{9-11}$ was assessed by lysing cells, Western blotting for the GST-tag, and normalizing to β-actin. (**a**) Representative immunoblots are shown with recombinant GST-FNIII$_{9-11}$ included as a positive control. (**b**) Quantification of the GST-FNIII9-11 retention. Results shown are mean ± S.E.M. n = 3-4.
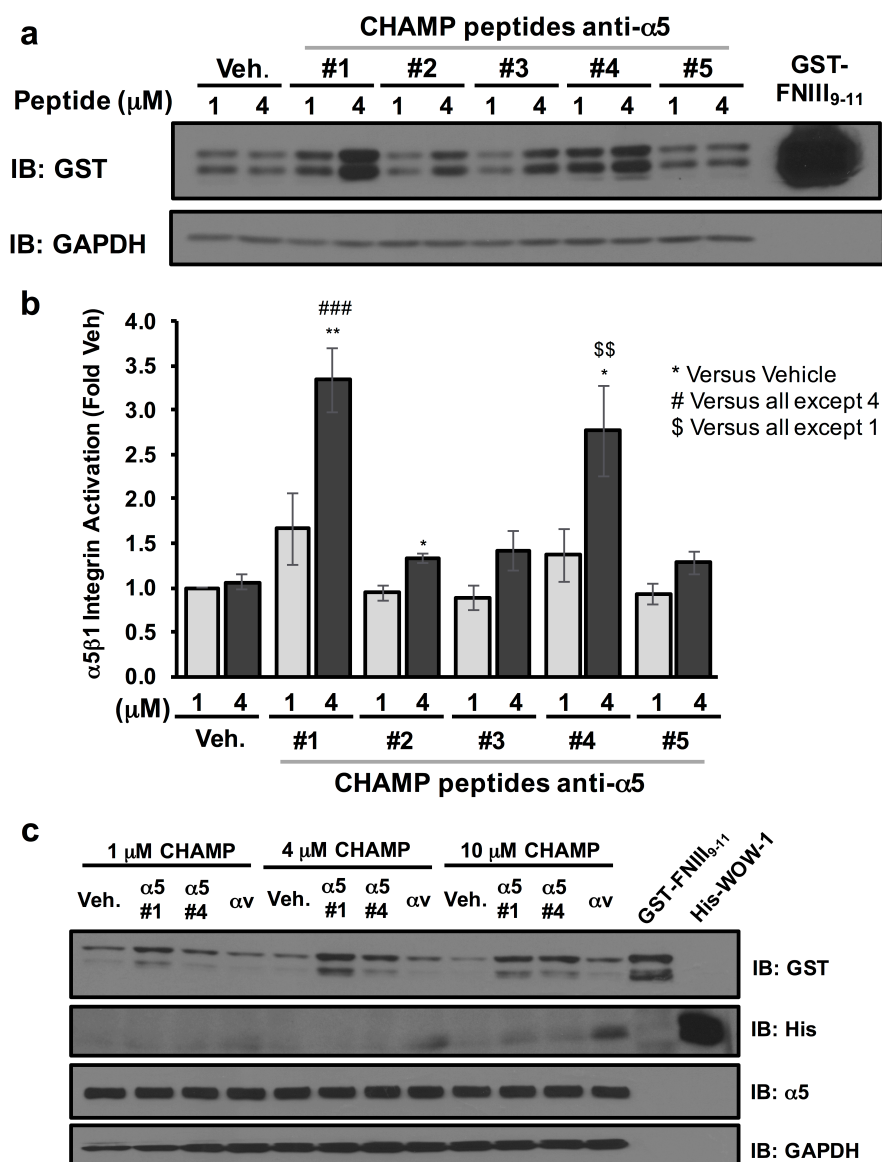
Figure 3.3. CHAMP peptides selectively activate $\alpha_5\beta_1$, but not $\alpha_v\beta_3$ in endothelial cells. Human aortic endothelial cells were treated for 30 minutes with the indicated doses of anti-α_5 CHAMP peptides or vehicle (ethanol) control in the presence of the $\alpha_5\beta_1$ ligand mimetic GST-FNIII$_{9\text{-}11}$. Retention of GST-FNIII$_{9\text{-}11}$ was assessed by lysing cells, Western blotting for the GST-tag, and normalizing to GAPDH. (**a**) Representative immunoblots are shown with recombinant GST-FNIII$_{9\text{-}11}$ included as a positive control. (**b**) Quantification of the GST-FNIII$_{9\text{-}11}$ retention. Results shown are mean ± S.E.M. n = 4. (**c**) Endothelial cells were treated with the indicated doses of CHAMP #1, CHAMP #4, anti-α_v CHAMP, or vehicle control for 30 minutes in the presence of both the $\alpha_5\beta_1$ ligand mimetic GST-FNIII$_{9\text{-}11}$ and His-WOW1 antibody specific for the $\alpha_v\beta_3$ activated state. Western blotting for GST-tag (GST-FNIII$_{9\text{-}11}$) or His-tag (His-WOW1) was used to assess $\alpha_5\beta_1$ and $\alpha_v\beta_3$ activation respectively. Representative immunoblots are shown with recombinant GST-FNIII$_{9\text{-}11}$ and recombinant His-WOW1 included as positive controls. n = 3.

a

```
Integrin α5 TM  : CGGGKKPLWIIILAILFGLLLLGLLIYILYKKKKKK
CHAMP #1        : CGGGKKAWSLVLGGLIGSLIAFAVFLLLWKK
CHAMP #4        : CGGGKKVLMVLLLLLAALLGILLGLIMVLWVKK
Integrin β1 TM  : CGGGKKIIPIVAGVVAGIVLIGLALLLIWKK
```


b Integrin α5 TM Homodimer HPLC Peak Area (AU)
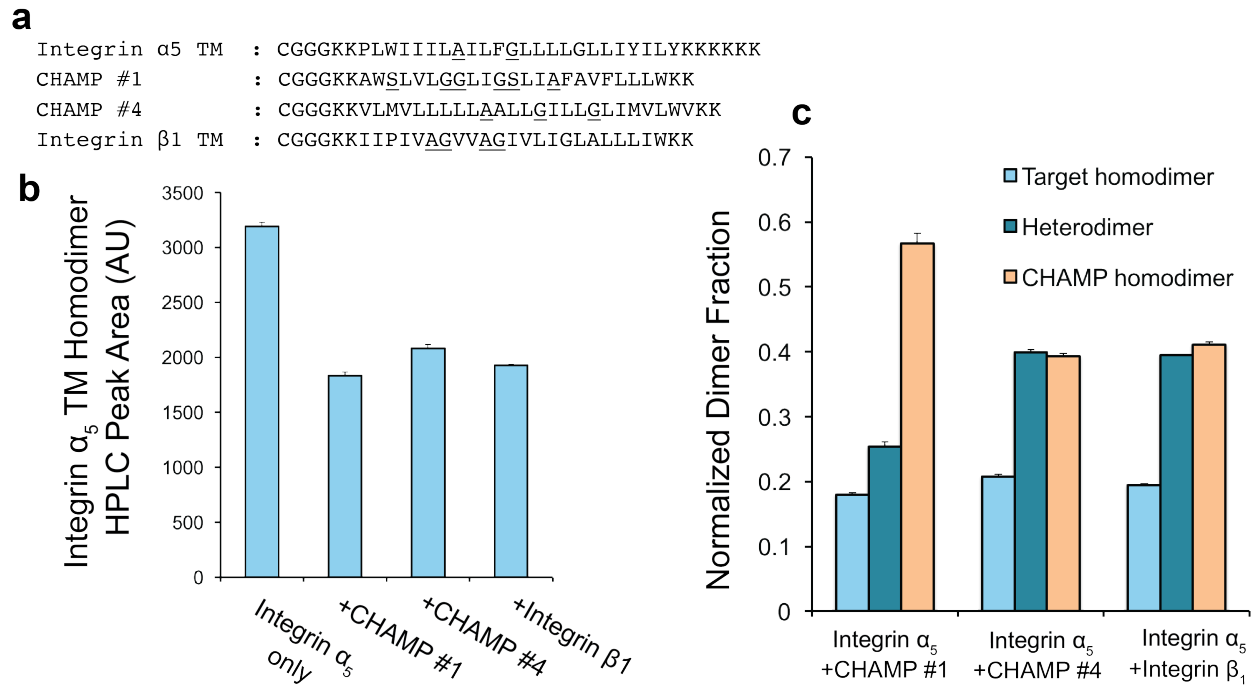
c Normalized Dimer Fraction

Figure 3.4. Disulfide Exchange Equilibrium Assay indicated different dimerization potential of CHAMP peptides. (a) Sequences of Cys labeled peptides used; a cysteine was added to the N terminus of each peptide followed by a triple-glycine flexible linker to disulfide bond formation. **(b)** Integrin α_5 TM homo-dimerization decrease due to a competitive binder. **(c)** Fraction of different dimers in equilibrium, indicating different preference of hetero-dimer formation over homo-dimer. Results shown are mean ± S.E.M. n = 3.
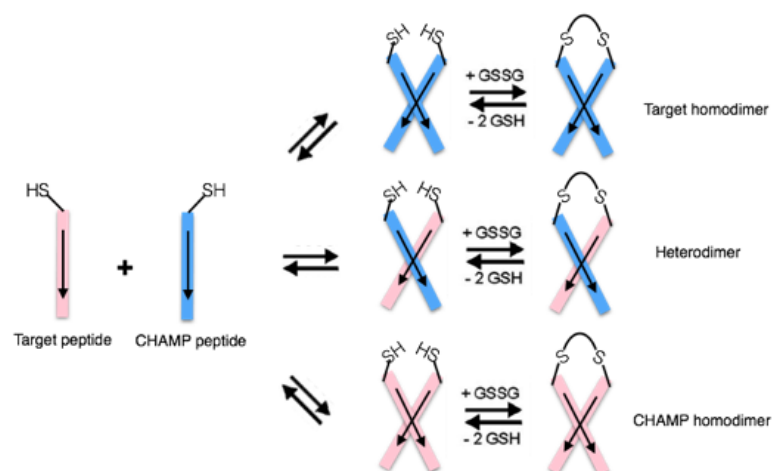
Figure 3.5. Potential monomeric and dimeric species formed and observed in the thiol-disulfide exchange assay.
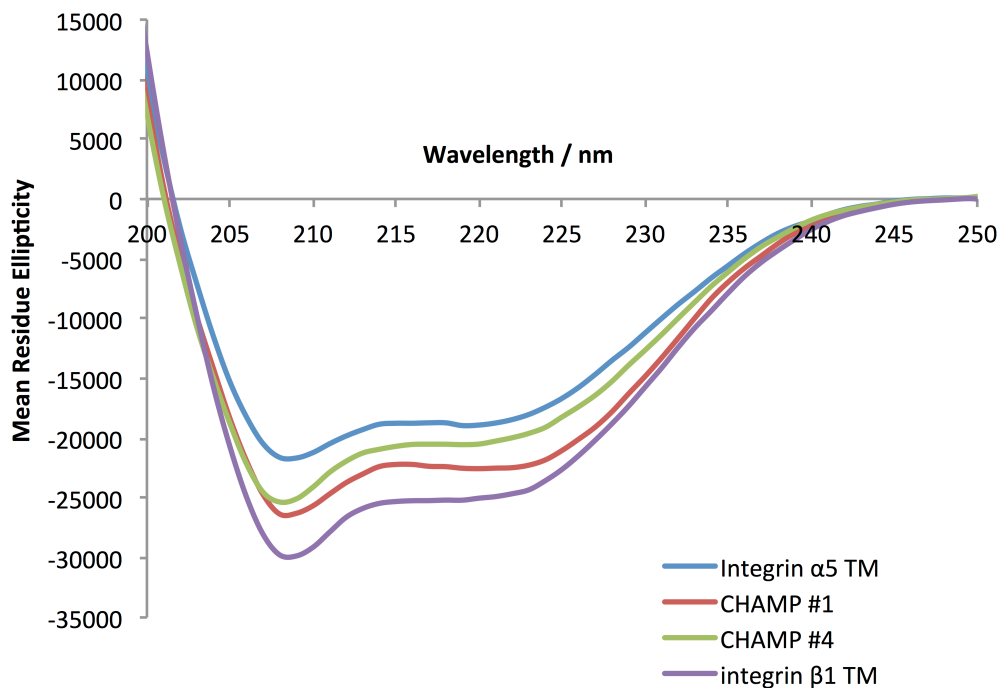
Figure 3.6. **CD Spectra of Peptides in Disulfide Exchange Equilibrium Assay**
(horizontal units: deg cm^2 dmol^{-1} res^{-1}). All four peptides are predominately α-helical at the thiol-disulfide exchange reaction condition (peptide/detergent ratio of 1/100). The experiments were performed in aqueous buffer containing 2.5 mM Tris–HCl (pH 8.6), 2.5 mM KCl, 2.5μM EDTA and 1 mM TCEPs with a peptide concentration of 100 μM. The spectra were acquired with a Jasco J-810 spectropolarimeter and a 0.1 cm quartz cell at 25℃.

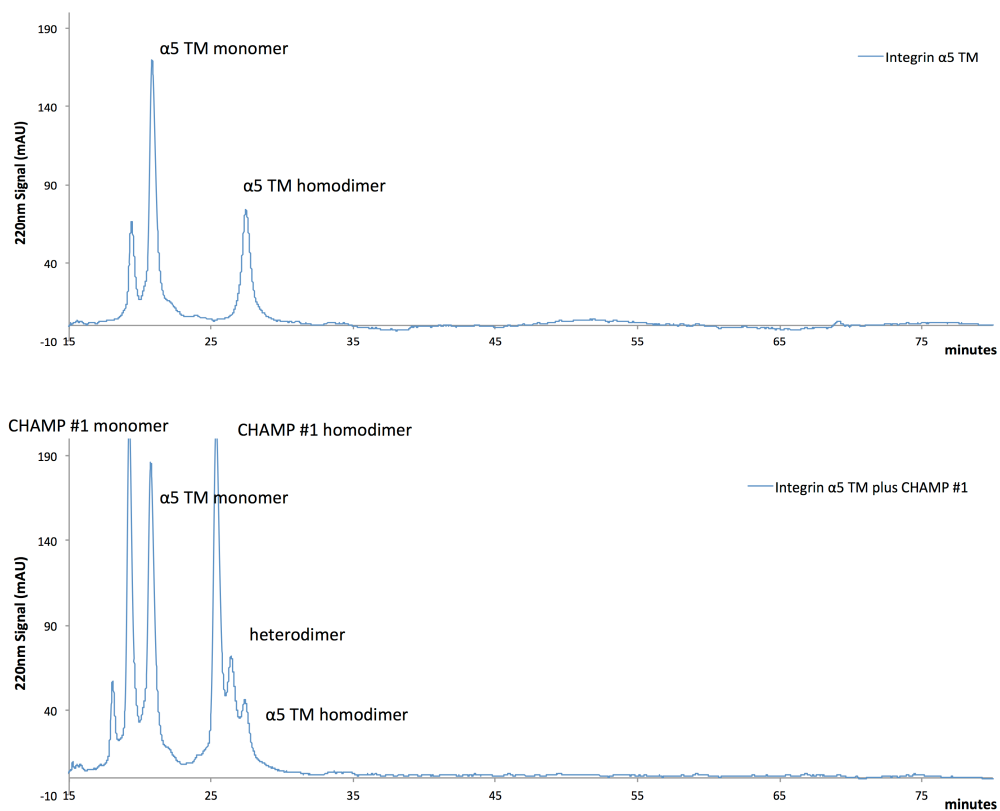**Figure 3.7. HPLC absorbance (220 nm) chromatograms of disulfide exchange equilibrium assay in DPC micelle**.
Peaks were labeled on the top; peaks with no labels were GSH adducts. The identities of elutant's were determined by ESI or MALDI mass spectrometry as well as compared HPLC chromatograms for individual monomeric peptides and monomer/dimer mixtures for each CHAMP formed by air-oxidation (not shown).
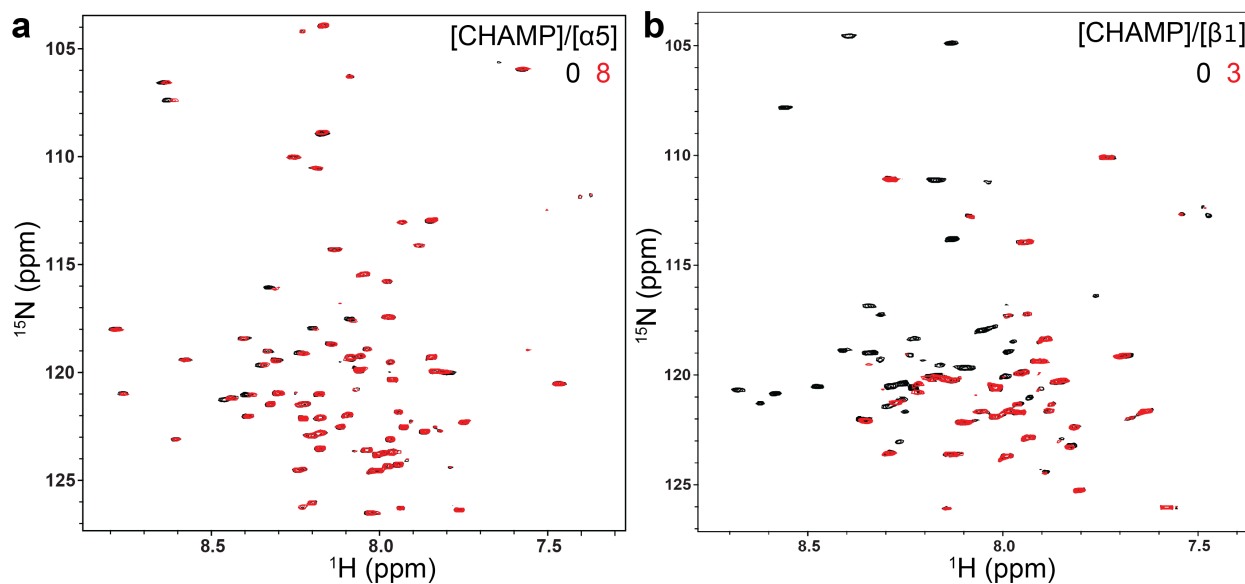
Figure 3.8. NMR Titration experiments of integrin α_5 and β_1 TM-CT with CHAMP #1.
(**a**) and (**b**), superimposed 800 MHz ^{1}H-^{15}N-TROSY-HSQC spectra from titrations of unlabeled CHAMP #1 into ^{15}N-integrin α_5 TM-CT or β_1 TM-CT in bicelles, respectively. The concentrations of the integrin α_5 and β_1 TM-CT were fixed at 150 uM for all samples. Little change in the α_5 TM-CT were observed upon addition of even 8 mol eqv. CHAMP#1, while significant peak intensity decay was observed for many resonances within the β_1 TM domain with addition of CHAMP #1, 3 mol eqv. shown here.
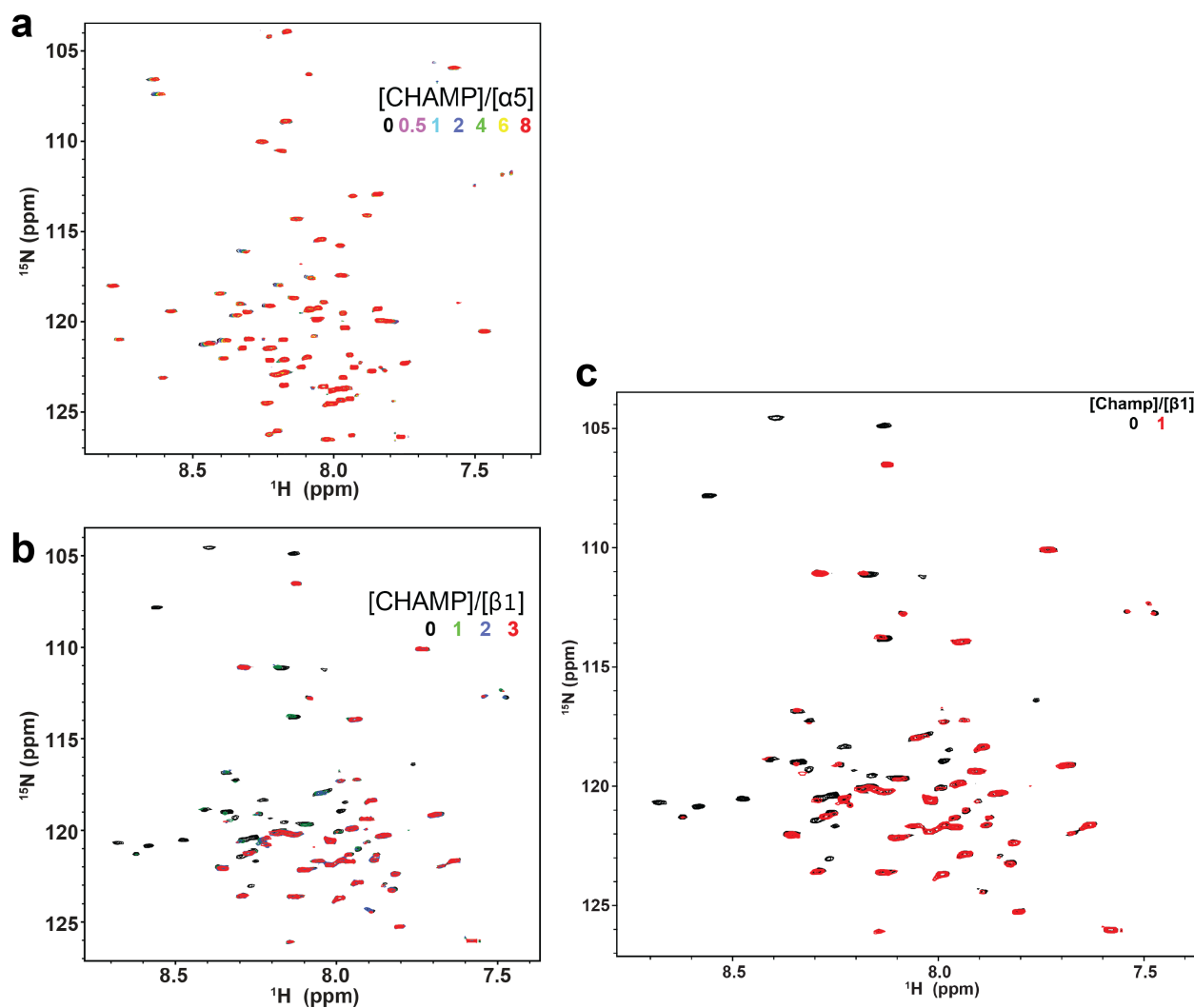
Figure 3.9. Full NMR titration series of integrin α_5 and β_1 TM-CT with CHAMP #1. (**a**) and (**b**), superimposed 800 MHz ^{1}H-^{15}N-TROSY-HSQC spectra from full titration series of unlabeled CHAMP #1 into ^{15}N-integrin α_5 TM-CT or β_1 TM-CT in bicelles, respectively. The molar ratio of CHAMP #1 peptide to the corresponding integrin subunit is listed in the top right in a color that matches the spectra displayed for that titration. (**c**) Significant changes in peak intensity for many TM domain residue resonances are seen upon addition of just 1 mol eqv. CHAMP#1 to β_1 TM-CT.
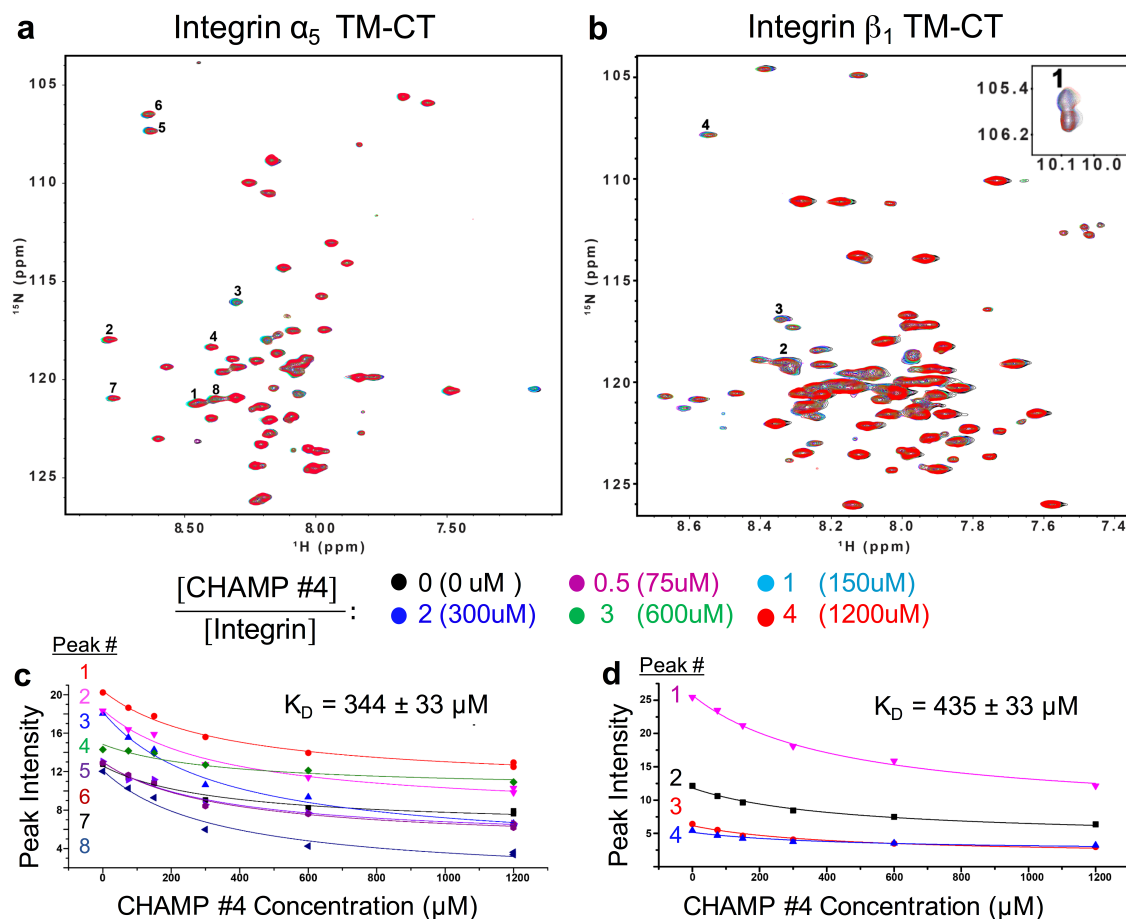
Figure 3.10. NMR Titration experiments of integrin α_5 and β_1 TM-CT with CHAMP #4.
(**a**) and (**b**), superimposed 800 MHz ^{1}H-^{15}N-TROSY-HSQC spectra from titration of ^{15}N-integrin α_5 transmembrane and cytoplasmic domain (TM-CT) construct or β_1 TM-CT, respectively, in bicelles with unlabeled CHAMP #4 peptide. The concentrations of the integrin α_5 and β_1 TM-CT were fixed at 150 uM for all samples. (**c**) and (**d**), the intensity decay of several peaks are tracked for both spectra upon addition of CHAMP #4 at increased molar ratios, numbered consistently in the spectra (**a**) and (**b**) and the plots (**c**) and (**d**), respectively. The trace indicates the global fit of this data to an equimolar binding model, with listed dissociation constants: CHAMP#4-integrin α_5, 344 ± 33 uM (1.6 mol%); CHAMP#4-integrin β_1, 435 ± 53 uM (2.03 mol%). Trp 751 indol proton is shown in the inset of (**b**).

Table 3.1. Sequences of CHAMP peptides designed in this study

Name	Sequence	Design pipeline	Backbone source
CHAMP #1	KKAWSLVLGGLIGSLIAFAVFLLLWKK	Original protocol	PDB: 2w2e
CHAMP #2	KKLIMLVFLLFAALLGLILGLLLVFLWKK	RosettaMembrane	PDB: 1kpl
CHAMP #3	KKLVMVLLLLIAAILGIALGLVAVWLLKK	RosettaMembrane	PDB: 1kpl
CHAMP #4	KKVLMVLLLLLAALLGILLGLIMVLWVKK	RosettaMembrane	PDB: 1kpl
CHAMP #5	KKWFAFVLILAVMVFIALLVALLLMLFAILKK	RosettaMembrane	PDB: 1r3j

Chapter 4

Computational Design of Transmembrane Peptides Inhibiting the Erythropoietin Receptor in a Custom Binding Mode

Computational Design of Transmembrane Peptides Inhibiting the Erythropoietin Receptor in a Custom Binding Mode

Abstract

Transmembrane (TM) domains as simple as a single span perform complex biological functions using entirely lipid-embedded chemical features. Computational design provides a platform for generating custom polypeptide tool molecules directly targeting and perturbing functionally active TM regions. Thus far, *de novo* designed TM targeting agents have been limited to mimicking binding modes and motifs of natural TM interaction partners. Here, we demonstrate design of synthetic TM proteins targeting the erythropoietin receptor (EpoR) TM domain in a custom binding topology competitive with its homodimerization. Expressed designer TM domains complex with EpoR in mammalian cell membranes and inhibit erythropoietin-induced cell proliferation through the JAK/STAT pathway. Structural characterization *in vitro* shows the TM domain complex has a 1:1 stoichiometry with antiparallel helices, outcompetes EpoR's native parallel homodimerization, and involves the intended amino acids, agreeing with our guiding computational model. Thus, membrane protein TM regions can now be targeted in custom designed binding modes.

Introduction

Cells rely on membrane proteins as simple as a single transmembrane (TM) span for many complex biochemical functions: assembling multi-protein complexes, sensing mechanical or electrical cues, and signal transduction across lipid barriers. The molecular features executing these tasks are often completely embedded in lipid. Thus, it is important to develop chemical probe molecules that target and perturb proteins directly at their functional membrane-spanning regions. The ability to monitor and manipulate TM domains similar to how antibodies and small molecules allow control over globular domains would transformative for studying the active roles these domains play in biological function and disease(*1, 2*).

Approaches for targeting TM domains exist, but remain limited. The earliest and most simple approaches mimic natural lipid-inserting sequences as peptides. Delivered to cell membranes, TM mimic peptides obstruct TM domain assembly by competing with TM protein-protein interactions or intramolecular folding of multi-spanning targets(*3-6*). However, simple TM mimics have limited scope, tending to denature multi-spanning protein targets rather than binding to intact folds. These molecules also often have sub-optimal affinity, off-target promiscuity, and toxicity that are difficult to improve. Alternative approaches tailoring synthetic membrane-soluble molecules overcome these boundaries to different extents, including rational derivatization(*3*), computational design(*7*), or screening expressed libraries of randomized TM protein variants(*8-10*). Cellular phenotype-based selections of TM protein libraries have a history of success(*11*) and provide the advantage that the structure, functional conformation, or sensitive epitope of an intended membrane protein target need not be known. However, synthetic TM proteins are often selected that do not form stable complexes with the intended target, including those acting by alternative signaling complexes(*12*) or weak, transient, indirect, or lipid-specific

148

assemblies that can be difficult to biochemically and biophysically characterize*(13)*. Computational protein design has the potential to target specific regions of folded multi-spanning proteins, including customizing interaction geometry, recognizing distinct conformational sub-states, and excluding unwanted binding events. However, membrane protein design has major boundaries and sparse precedent due to the requirements of a confident atomic model for the membrane protein target and a reliable potential function for estimating protein interaction energetics in lipid(*14*). Although the latter has been historically lacking, recent force-field developments are noted(*15, 16*).

Here, we focused the capability of using protein design to target specific TM domain region in a custom interaction geometry, e.g. parallel or anti-parallel helix orientation. Achieving this milestone is requisite for the more challenging objective of binding complex surface sites of multi-spanning protein folds to perturb their conformational equilibria (e.g. ion channels, G-protein coupled receptors) without inducing unfolding or aggregation. Designed TM peptides so far have only mimicked binding modes of natural TM interaction partners known *a priori* and have relied on the field's best-known sequence motif for driving TM domain interaction in a parallel orientation, GxxxG (*17, 18*). Initial proof-of-concept succeeded in designing GxxxG-containing *de novo* TM peptides that could bind and selectively distinguish the highly similar single-pass TM domains of integrins α_{IIb} and α_V, each also having GxxxG motifs(*7*). Recent work targeting integrin α_5 again relied on GxxxG-like motifs for affinity, and yielded an $\alpha_5\beta_1$-activating peptide which instead unexpectedly bound the β_1 subunit(*19*). Thus, 13 years after this algorithm's proof-of-concept, *de novo* design of TM targeting polypeptides is far from routine. Likewise, the binding geometries and spectrum of membrane protein targets accessible have remained extremely limited.

We present three advancements to our Computed Helical Anti-Membrane Proteins

(CHAMP) technology, using similar theoretical principles with different implementation. First, we expand the algorithm beyond the field's best-understood TM interaction motif (GxxxG). Second, we tailor the interaction geometry by which our *de novo* CHAMP binds the target region of a receptor's TM domain, showing that mimicry or knowledge of native interaction partners are no longer pre-requisites for design. Third, we target and effectively inhibit a signal-amplifying receptor, imposing a strict functional requirement. The membrane protein of interest is the highly signal amplifying single-span mouse Erythropoetin (EPO) receptor (mEpoR). EPO binding to the extracellular domain of only 6% of cell surface mEpoRs yields ~50% maximal cellular response(*20, 21*). We designed synthetic TM domains to compete with the homodimerization epitope of mEpoR's TM α-helix by binding in non-native anti-parallel TM helix orientation, impairing cross-membrane signal conduction (Fig. 4.1). While most design expeditions have success rates under 10%, our algorithm on first attempt yielded *de novo* TM sequences that when expressed assemble with mEpoR in mammalian cells and inhibit EPO-EpoR signal conduction. The TM domain complexes formed *in vivo* and *in vitro* are consistent with the topology and sidechain interactions predicted from our design model. This adaptation of the CHAMP algorithm executed in RosettaMembrane(*22*) is effective in designing TM targeting peptides adopting custom binding modes and should be promising for more complex application to multi-spanning proteins.

Results

Design of anti-mEpoR TM peptides to bind in a non-native anti-parallel complex

EpoR is a model cytokine receptor whose transmembrane domain contributes to both receptor homodimerization(*23-25*) and efficiency of cross-membrane conformational coupling upon ligand binding(*26, 27*), activating the cytoplasmic associated JAK kinase to drive cell

proliferation and differentiation. Likewise, known natural (gp55(*28*)) and library engineered(*11, 13*) TM domain interaction partners exist for mEpoR; these target EpoR's dimeric state to activate EPO-independent signaling. Interest in instead inhibiting mEpoR with CHAMP peptides stemmed in part from the two coexisting 7-residue sequence motifs lining its expected homodimerization interface (Fig. 1c): a small-X_6-small motif (small = Gly, Ser, Ala)(*29, 30*) and a Ser-Leu zipper(*31*). Model TM peptides repeating this Ser-Leu motif form stable parallel dimers in the absence of extra-membrane loops(*32*). Yet, Ser-Leu zipper sequences can be found mediating both parallel and antiparallel TM helices packing within multi-spanning proteins(*31*). Thus, mEpoR's parallel homodimerization motif provides a molecular surface and energy landscape that is a tractable target for a CHAMP peptide to instead engage in a non-native anti-parallel binding mode, posing a clear challenge for our algorithm. Given the presence of the small-X_6-small repeat motif (S230-S237-A244) and this motif's correlation with an antiparallel TM interaction geometry in our previous informatics studies(*33*), we adapted a protocol that models and stabilizes a dimeric CHAMP/mEpoR complex (Fig. 4.1) mediated by mutual small-X_6-small repeats.

In the CHAMP algorithm, the target TM domain (mEpoR) is first modeled in an implicit membrane at an energy-optimized insertion depth and orientation(*34*). Second, a molecular model is built positioning a putative CHAMP binding partner relative to the embedded TM domain target region. This step circumvents traditional random rigid-body docking of TM helices in mostly unproductive conformations. Instead, we select from a curated set of favorable TM helix-helix interaction geometries discovered in our previous analysis of the most common tertitary motifs stabilizing membrane protein structures(*33*). Here, we idealized coordinates from >100 structural examples of a prominent anti-parallel interaction motif having close-packing left-handed helices whose consensus sequence is small-X_6-small repeated, selecting strictly those examples whose

sequences contains three consecutive small-X_6-small motifs (Fig. 4.2). The resulting interaction geometry (inter-helical distance 8.1 Å; crossing-angle, -175 degrees; Z-offset, 2 Å) was used to precisely position a poly-alanine CHAMP helix relative to mEpoR's small-X_6-small in the anti-parallel CHAMP/mEpoR binding mode we aimed to stabilize. Next, we implemented a sidechain repacking algorithm in RosettaMP to design the CHAMP sequence(*35*). Lipid-facing positions were selected semi-randomly based on their depth in the membrane (see methods). Of the remaining 13 of 24 residues, the 4 small-X_6-small positions were allowed to independent sample Gly, Ser, and Ala amino acids and the remaining 9 positions sampled a limited lipid-friendly alphabet (GATSVLIFM) using a routine of rotamer repacking trials followed by a combined step of minimization and re-orientation in the membrane (RosettaMP FastRelax)(*22*), repeated twice each. The critical final step was ranking and identifying sequences having the most stable modeled complex for experimental testing. Due to the documented poor accuracy of predicted energies by RosettaMembrane(*14*), we instead as a proxy ranked models primarily on a distinct molecular feature: the lack of sidechain packing voids, a model quality metric with promising predictive power in both protein folding simulations and *de novo* design(*36*) (Fig. 4.2). Sequence clustering of the top 10% most well-packed models generated two groups differing only by a pair of residues, defining the final anti-mEpoR CHAMP-1 and -2 sequences (Fig. 4.1). We also selected a random small-X_6-small-containing TM sequence as a "No Design" negative control to explore selectivity requirements for motif to bind mEpoR's TM domain.

While previous work drastically narrowed *de novo* sequence exploration by following known principles for encoding parallel right-handed TM dimers with GxxxG-mediated motifs(*7, 19*), we instead focus on stabilizing an anti-parallel inter-helical geometry whose sequence-structure principles are not well defined. It is unclear how diverse mixtures of Ser/Ala/Gly at

"small" positions and different sidechain chemistries at intermediate helix turns ('*d*' or '*e*' residues, '*abcdefg*' 7-residue repeats; i.e. Leu of Ser-Leu zipper) interact to define stability and topology in the context of the small-X$_6$-small repeat. Thus, we actively test our design algorithm's ability to effectively sample and recognize sidechain packing geometries stabilizing this desired TM binding geometry *de novo* regardless. Specific adjustments made to the CHAMP algorithm here included (A) use of RosettaMembrane to conduct sidechain packing, increasing user accessibility, (B) ranking designs on the interface packing metric (not relying on predicted interaction energy), and (C) idealizing a single coordinate set for the CHAMP/target model rather than conducting independent design trials for 5-10 examples of similar backbones. Finally, previous work selected the final CHAMP sequences by human visual evaluation of models(*7*), a step that limits reproducibility and introduces user disparities. The adaptations described here enable complete automation of model building and sequence design.

CHAMP peptides inhibit EPO-induced growth by binding the mEpoR TM domain

To test the biological activity of these designed sequences, FLAG-tagged synthetic TM proteins were expressed (Fig. 4.1) by retroviral transduction in mouse BaF3 cells stably expressing mEpoR (BaF3/mEpoR cells). Proliferation of these cells can be stimulated by IL-3 (EpoR-independent) or EPO (EpoR-dependent) growth factors. When BaF3/mEpoR cells were transduced with synthetic TM domains supplemented with IL-3 in the absence of EPO, robust proliferation was observed showing these TM proteins are not cytotoxic (Fig. 4.3). Likewise, proliferation is not observed in the absence of IL-3 and EPO (Fig. 4.3), ruling out the possibility these proteins induce EPO-independent mEpoR activation.

We next tested whether the expressed TM proteins impair EPO-induced cell proliferation by interacting with EpoR. Over an 8-day period, these cells showed markedly reduced proliferation compared to cells lacking CHAMP expression (Fig. 4.1), whereas the No Design Control TM protein did not impair growth. At 8 days, cells expressing CHAMP-1 and CHAMP-2 showed cell counts at $38 \pm 5\%$ (n=6) and $40 \pm 6\%$ (n=6) of the expected vector control, respectively (p-value, both $p < 0.005$) (Fig. 4.1). When hEpoR was expressed instead of mEpoR (BaF3/hEpoR), cells co-expressing the TM proteins in the presence of EPO instead proliferated similarly to vector-only cells (Fig. 4.3). Similarly, the CHAMP proteins did not inhibit cells expressing a previously published construct "mhm-EpoR", consisting of mEpoR's water-soluble domains and hEpoR's TM domain (Fig. 4.3). Thus, the inhibition of EPO-stimulated growth by CHAMP-1 and CHAMP-2 is highly specific to mEpoR's TM domain sequence, given that hEpoR and mEpoR TM domains only differ by three amino acids. Finally, expressing constructs containing the native TM domain sequence of mEpoR or the unrelated PDGFβR in BaF3/mEpoR cells do not inhibit proliferation (Fig. 4.1, 4.3). Thus, these TM mimic proteins are non-functional, whereas *de novo* designed CHAMP proteins inhibit EPO-stimulated growth likely via mEpoR's TM domain.

We also conducted co-immunoprecipitation experiments to assess whether the CHAMP proteins physically associated with the mEpoR. Detergent lysate of BaF3/mEpoR cells expressing a CHAMP protein or the No Design control protein were immunoprecipitated with antibody recognizing the N-terminal FLAG tag on the designed proteins, then immunoblotted with antibody recognizing HA on the EpoR. Of the FLAG-tagged TM proteins expressed, only CHAMP-1 and CHAMP-2 co-immunoprecipitated with HA-tagged mEpoR (Fig. 4.1).

Next, we studied the relative orientation and association of synthetic peptides consisting of the designed sequences and mEpoR's TM domain *in vitro* using an established equilibrium thiol-disulfide exchange assay (Fig. 4.4; sequences, Table 4.1)(*37, 38*).   In this experiment, the preference for formation of specific covalent dimeric species over others provides insight into the underlying population distribution of non-covalently TM assemblies and their interaction topologies, given altered cysteine proximities. The mEpoR TM domain peptide containing an N-terminal cysteine and a designed peptide containing a C-terminal cysteine were reconstituted in dodecylphosphocholine (DPC) detergent micelles or in 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) small unilamellar vesicles (SUVs) at high molar ratio to promote association (100:1 detergent/lipid: peptide) and subject to glutathione-assisted reversible oxidation.   The disulfide-bonded dimer species fractions quantified by reverse-phase high-performance liquid chromatography (RP-HPLC) for all three CHAMP-1, CHAMP-2, and the No Design Control peptides show a strong deviation from random in their preference for heterodimers with mEpoR (55-66%) in both DPC micelles or POPC SUVs (Fig. 4.4), ranging from 2.5 to 16-fold enrichment relative to the native parallel mEpoR homodimer.   Thus, the parallel self-interaction of mEpoR's TM domain is less stable and outcompeted by its anti-parallel heteromeric assembly with the synthetic small-X_6-small-containing peptides (i.e. N-to-C disulfide bond) in detergent micelles and SUVs, consistent with competitive inhibition observed in cell membranes.

CHAMP-1 competitively inhibits the EpoR pathway in sequence-dependent manner

Next for one design, CHAMP-1, we investigated the mechanism and sequence features driving its inhibition of EPO-mEpoR-dependent cell proliferation.   First, we expressed the CHAMP-1 construct under a titratable doxycycline-repressible promoter in BaF3/mEpoR cells

which revealed CHAMP-1's inhibition is dependent on its expression level (Fig. 4.4). Likewise, the inhibitory effect was negatively correlated with the concentration of EPO between 0 and 0.24 U/mL added to stimulate mEpoR signaling and BaF3 cell growth as expected (Fig. 4.4). Likewise, expression of CHAMP-1 in BaF3/mEpoR cells drastically reduced the EPO-stimulated tyrosine phosphorylation of JAK2 and STAT5 effector proteins downstream of mEpoR activation (Fig. 4.4). Thus, CHAMP-1 competes with EPO-mEpoR cross-membrane signal conduction and effectively dampens downstream catalytic JAK/STAT phosphorylation and signaling amplification events driving cell proliferation.

Next, we used mutagenesis to identify amino acids in the mEpoR TM domain and CHAMP-1 responsible for mEpoR-specific BaF3 cell growth inhibition. First, CHAMP-1's ability to inhibit proliferation (% cell count versus empty vector transfection, day 6) when expressing hEpoR-like single and double mutants of mEpoR at the 3 dissimilar TM positions at 0.06 U/mL EPO (Fig. 4.5). Compared to expected $42 \pm 5\%$ inhibition of wildtype (WT) mEpoR (day 6, n=10), CHAMP-1 showed similar potency towards L235V, $46 \pm 15\%$, and modestly reduced inhibition with S237L, $57 \pm 8\%$, neither reaching statistical significance versus for WT mEpoR. However, L238V showed significantly reduced inhibition by CHAMP-1 at $80 \pm 11\%$, versus WT mEpoR ($p < 0.003$). For all double mutants, inhibition was significantly dampened ($p < 0.01$). S237L-L238V was completely uninhibited at ($101 \pm 13\%$), while L235V-S237 and L235V-L238V maintained partial CHAMP-1 inhibition at $73 \pm 4\%$ and $74 \pm 8\%$, respectively. Due to L235V's individual lack of responsiveness, S237L's and L238V's partially responsiveness, and cooperativity between S237L and L238V to completely abrogate CHAMP-1's effect, S237 and L238V could possibly lie on the CHAMP-1-interacting face of mEpoR's TM helix, whereas L235 is lipid-facing. This interpretation differs slightly from the helical phase and registry of our

156

designed model by a slight helix rotation (Fig. 4.5), yet is consistent with target residue S237 participating in the complex.

Next, a panel of both polar and apolar mutants of CHAMP-1 at predicted mEpoR-facing positions were similarly tested for inhibition of EPO-stimulated WT BaF3/mEpoR cells (Fig. 4.5). S8Q, S8N, S8D, or S8E mutants lose all inhibitory potency, cells proliferating at ≥ 100 % of vector transfected cells. T19Q maintains modestly inhibition ($70 \pm 8\%$, p = 0.01). However, S1Q was completely tolerated, inducing significant inhibition ($42 \pm 7\%$, p < 0.001) within error of WT CHAMP-1. Interestingly, CHAMP-1 S8Q was found to still co-immunoprecipitate mEpoR (Fig. 4.3). We attribute the loss of potency to the propensity of strongly polar membrane-embedded sidechains to drive TM domain association in a depth-dependent manner(*39*) thus limiting the monomeric pool of CHAMP-1, maybe also partially disrupting the mEpoR/CHAMP-1 interface. We additionally mutated the "S1-S8-G15-G22" small-X_6-small motif of CHAMP-1 prioritized during the design process to large Ile or Leu amino acids, mutants "I-I-I-I" ("4I") and "L-L-L-L" ("4L") (Fig. 4.4). "I-I-I-I" exhibited significantly reduced inhibitory potency versus CHAMP-1, $76 \pm 28\%$ (p < 0.05), but interestingly did not completely abolish activity. By contrast, "L-L-L-L" lost inhibitory potency and instead exhibited growth-factor independent proliferation (Fig. 4.3, 4.5), thus clearly altering the TM domain interaction equilibrium.

CHAMP-1 forms an anti-parallel dimeric complex with the EpoR TM domain in vitro and in mammalian cells

We next characterized the topology of freely associating mEpoR and CHAMP-1 TM peptides in detergent micelles solutions. First, we used an established FRET-based fluorescence quenching assay to determine the stoichiometry of the interacting complex(*40, 41*) (sequences,

Table 4.1).   When increasing molar ratios of Fluorescein-5-maleimide-labeled CHAMP-1 (acceptor) to (diethylamino-4-methylcoumarin-3-yl)maleimide-labeled (donor) mEpoR peptides were reconstituted in myristyl sulfobetaine (C14-B) micelles (fixed 175:1 detergent to total peptide ratio), donor fluorescence emission was observed to decay linearly to half intensity at 1:1 acceptor:donor ratio (Fig. 4.6). Given FRET-based quenching expected from theoretical equations (see methods)(*41*), the data are best explained by a dimeric complex of CHAMP-1/mEpoR predominantly associated at these peptide/detergent mol fractions (0.2 – 0.5 % mol).   Decreasing the detergent ratio concentrating the peptides (100:1) did not cause additional quenching, but further diluting the TM domain complex in detergent (250:1) somewhat reduced fluorescence decay linearly (Fig. 4.7).   This indicates a modest affinity interaction, yet stronger than previously measured for mEpoR's TM domain homo-dimerzation in C14-B by AUC ([monomer] ≈ [dimer] at 0.3% mol)(*23*).

        We next tested whether the non-covalent CHAMP-1/mEpoR complex preferentially adopts a parallel or anti-parallel orientation by measuring which of an N-terminal or C-terminal cysteine-containing CHAMP-1 peptide (nCys-CHAMP-1, cCys-CHAMP-1) more readily disulfide bonds with a biotinylated N-terminal cysteine-containing mEpoR TM peptide (biotin-nCys-mEpoR) in C14B micelle solutions as a consequence of cysteine proximity (Fig. 4.6).   After reconstituting and reversibly oxidizing the three peptide mixture at 40:1 C14B to total peptide mole ratio and 2:2:1 nCys-CHAMP-1:cCys-CHAMP-1:nCys-mEpoR, mEpoR-containing species were captured by streptavidin beads at pH 4.5, quenching oxidation and washing off unbound and non-covalent species (Fig. 4.8).   The pool of monomeric TM peptides disulfide bonded to biotin-nCys-mEpoR were eluted off the beads with excess reducing agent and C14B, finally quantifying species fractions by RP-HPLC (Fig. 4.6).   This workflow revealed a strong preference (62% of total) for

cCys-CHAMP-1 to disulfide bond with biotin-nCys-mEpoR, 3-fold more than either nCys-CHAMP-1 or biotin-nCys-mEpoR, indicating TM helices of the non-covalent CHAMP-1/mEpoR complex are predominantly anti-parallel.

We also used a flow cytometry-based split GFP protein complementation assay in living BaF3 cells to examine CHAMP-1's TM orientation and association with mEpoR in mammalian plasma membranes (Fig. 4.6, 4.9). Well-characterized constructs of GFP's beta-strands 1-10 (GFP1-10) or the 11th strand alone (GFP11) were used(*42*). GFP1-10 was fused via a short flexible linker to mEpoR or hEpoR replacing EpoR's C-terminal cytoplasmic domain while GFP11 was fused to putative interaction partners and co-expressed (Table 4.2). Control experiments with hEpoR-GFP1-10 show low cellular mean fluorescence intensity (MFI) (1.5×10^3) when expressed alone and only a minor (2.5×10^3, < 2-fold) MFI increase when hEpoR-GFP11 is co-expressed suggested negligible hEpoR self-interaction. However, co-expressing a short TM protein known to induce hEpoR EPO-independent activation showed dramatic 10-fold increase in MFI consistent with GFP complement (Fig. 4.9). We co-expressed CHAMP-1 constructs as either an N- or C-terminal fusion to GFP11 to measure their complementation of hEpoR-GFP1-10 and mEpoR-GFP1-10 (sequences, Table 4.2). The N-terminal GFP-11-CHAMP-1 fusion resulted in cells with 5-fold higher MFI (1×10^4) than that of uncomplemented mEpoR-GFP1-10 or parental BaF3 cells. Co-expression of GFP-11-CHAMP-1 with hEpoR-GFP1-10 yielded a smaller 2-fold increase consistent with background complementation (Fig. 4.6). The C-terminal CHAMP-1-GFP-11 fusion was not present in cell lysates at a detectable level and did not increase fluorescence, suggesting poor membrane insertion or instability. The ability of N-terminal GFP11-CHAMP-1 fusion to strongly complement mEpoR-GFP1-10 is indicative of its anti-parallel TM helix interaction with mEpoR in mammalian cell membranes as designed.

159

Solution NMR shows a direct sidechain-mediated complex at the targeted mEpoR epitope

We next measured spectral changes for U-^{15}N or U-^{15}N,^{13}C mEpoR TM peptides in the presence of unlabeled CHAMP-1 peptides by solution NMR in membrane mimics (sequences, Tables 4.1, 4.3). As rigorous controls, detergent- or bicelle-only titrations were performed in parallel to probe the monomer-homodimer equilibrium, and distinguish mEpoR monomer or homodimer resonances from those of the heterodimer mEpoR/CHAMP-1 complex. First, conditions in C14B mimicking our FRET experiment (180:1 detergent:peptide) were used for ^{1}H-^{15}N HSQC spectra (800 MHz, 45° C, pH 5.2) of expressed U-^{15}N mEpoR TM construct titrated with 1.3 molar equivalents of a CHAMP-1 peptide (Fig. 4.10). A set of slow-exchanging peaks emerge, distinct from mEpoR's homodimer resonances, indicative of this minor population in stable complex with CHAMP-1. Altering the sequences of their flanking water-exposed regions, a second mEpoR construct (mEpoR-TM2) and different CHAMP-1 peptide were similarly assayed (Table 4.1). CHAMP-1 titration to U-^{15}N mEpoR-TM2 in C14B again induced chemical shift perturbations instead in fast chemical exchange, whereas mEpoR-TM2's monomer-homodimer equilibrium was in slow-exchange. (Fig. 4.11). In 1,2-dimyristoyl-sn-glycero-3-phosphocholine (DPMC)/1,2-dihexanoyl-sn-glycero-3-phosphocholine (DHPC) q=0.3 bicelles, fast-exchanging chemical shift perturbations were observed for both bicelle and CHAMP-1 titrations yet differed in directionality (Fig. 4.12), allowing assignment of distinct monomer, homodimeric, and heterodimeric mEpoR-TM2 states. A dissociation constant of 2.7 ± 0.6 mol % (versus DMPC) was fit for CHAMP-1. In DPC micelles, the well-dispersed monomeric ^{1}H-^{15}N resonances (Fig. 4.13) underwent distinct slow-exchanging behavior for both DPC (Fig. 4.14) or CHAMP-1 titration (Fig. 4.15) allowing unambiguous assignment of homodimer and heterodimer states (Fig.

4.16). mEpoR and CHAMP-1 seemingly interact in solution for all membrane mimics and peptide constructs tested, albeit varying in chemical exchange. Surprisingly, these constructs' association affinities inferred from population fractions were much weaker than expected from our FRET experiments, possibly attributed to sequence changes at mEpoR's membrane-proximal regions (e.g. histidines) and lowered pH (5.2) used for NMR. Nonetheless, the *de novo* designed CHAMP-1/mEpoR-TM2 complex could be uniquely isolated as the primary mEpoR population, with best intensities, linewidths, and dispersion in DPC using both excess detergent (>400:1 DPC:mEpoR) and CHAMP-1 (>1.5 mol %, or 6-8 equivalents) suitable for resonance assignment.

In ^{2}H-DPC, we next characterized how CHAMP-1 perturbs mEpoR's sidechain environment. First, we assigned the α-helical monomeric state in ^{2}H-DPC (600:1 detergent:mEpoR-TM2 U-^{1}H,^{15}N,^{13}C) using triple resonance HNCA and HNCB spectra followed by (H)CC(CO)NH and H(CC)(CO)NH backbone-sidechain TOCSY spectra to separate heavily overlapping ^{1}H-^{13}C spectral regions (Fig. 4.13, 4.17, 4.18). We next performed independent sequential backbone assignment and partial sidechain assignment of the mEpoR-TM2(U-^{1}H,^{15}N,^{13}C)/CHAMP-1 heterodimer resonances (Fig. 4.19). Comparison of ^{1}H-^{13}C-HSQC spectra (800 MHz, 45° C, pH 5.2) of mEpoR's monomeric and CHAMP-1 bound states showed new peaks, widespread chemical shift perturbations, and select resonance broadening across diverse sidechain chemical groups (Fig. 4.15, 4.20), indicating a direct complex with an extensive sidechain-mediated interface. Numerous mEpoR side-chains in close interaction (< 4 Å) with CHAMP-1 in our design model experienced significant changes, including V17 CG2 (shift), I19 CG1 (broadening) and CG2 (shift), T24 CB (shift), Ser13 and 20 CA (overlapped, broadening), and Ser 20 CB (shift). Comparing ^{1}H-^{13}C spectra of a largely homodimeric ^{13}C mEpoR sample (Fig. S15) showed that about half of observed peak shifts (assigned or unassigned) are in similar

directions to those induced by CHAMP-1 binding (e.g. V17, A27, S30, possibly H31, R32). However, numerous distinct shift behaviors for CHAMP-1-bound versus homodimer resonances are also observed in many cases, including membrane proximal residues (H3, W4, N5, P7) and many of those residues specifically targeted by the designed CHAMP (I9, I19, S20, T24).

Interestingly, the per-residue shift perturbations ($\Delta\delta$) to mEpoR Cα atoms upon CHAMP-1 binding exhibited a clear pattern of 3-4 residue periodicity (Fig. 5d), also mirrored by ^{1}H-^{15}N $\Delta\delta$ (Fig. 4.21). This periodicity could indicate mEpoR's helical register in its interaction with CHAMP-1, given that the phase of this periodic $\Delta\delta$ is roughly in-phase with the inter-helical Cα-Cα distances between mEpoR and CHAMP-1 are the TM core region of our design model; mEpoR backbone atoms with the largest $\Delta\delta$ perturbations are closest to CHAMP-1 and residues least perturbed are lipid-facing. However, this correlation was not exact and became out of phase towards both helix termini, hinting CHAMP-1 might possibly have a right-handed crossing angle relative to mEpoR's α-helix (Fig. 4.15, middle). Incomplete assignment and overlaps, particularly for leucines, precluded analogous thorough comparison of sidechain resonance behaviors on opposing faces of its TM helix for monomeric versus CHAMP-1-bound mEpoR. However, the few unambiguously assigned resonances having notable changes are plotted on mEpoR's structure in Figure 4.15 (*right*). Although shift perturbations occur at many distinct regions of mEpoR's TM helix, slightly more highly perturbed sidechain atoms occur at the face of mEpoR's helix expected to bind with CHAMP-1 in our design model and in phase with the periodic Cα $\Delta\delta$ (Fig. 4.15).

Finally, we measured direct inter-atomic interactions between ^{1}H,^{13}C,^{15}N-mEpoR-TM2 and unlabeled CHAMP-1 in ^{2}H-DPC through a F1-^{13}C-edited/F3-^{13}C,^{15}N-filtered HSQC-NOESY experiment. Intermolecular transfer nuclear Overhauser effect (NOE) resonances were observed between numerous ^{13}C-attached protons (mEpoR) and ^{12}C or ^{14}N-attached protons (CHAMP-1) in

the 2D projection spectrum (Fig. 4.15). One such resonance involving a CHAMP-1 amide proton (8.14 ppm) could be unambiguously attributed to interaction with the ^{13}C resonance of mEpoR (58.4 ppm) corresponding to several overlapping Leu Cα atoms (Fig. 4.20). Thus, we directly observe close backbone-backbone packing of CHAMP-1 with mEpoR, corroborating the intimate sidechain interaction network suggested from the broad ^{1}H-^{13}C spectral changes for the complex.

Discussion

Here, we follow the guiding principles of the original CHAMP algorithm in a distinct, automated software implementation to design *de novo* TM domains that robustly inhibit the target mEpoR via a custom binding topology in mammalian cell membranes. *In vitro* characterization reveals CHAMP-1 forms a complex with mEpoR's TM domain that is stable in diverse membrane mimics, at elevated temperature (45° C), and varying pH (5-8.5) by forming a direct and extensive sidechain-mediated complex. This robust binding of CHAMP-1 to mEpoR strongly contrasts that observed with previously engineered anti-EpoR synthetic TM domains selected from randomized protein libraries; they co-immunoprecipitated with EpoR and activated EPO-independent signaling in BaF3 cells, but induced negligible sidechain chemical shift changes in analogous solution NMR studies of TM domain fragments as described here(*12, 13*). In addition to selectively and stably binding its target, CHAMP-1 adopts a heterodimeric and antiparallel TM domain complex with mEpoR as designed, achieving our primary engineering goal.

However, the finer structure and interacting amino acids stabilizing the complex are not fully clear from our cellular or biophysical data, precluding assessment of our designed model's accuracy in lieu of high-resolution structural information. Interestingly, the structure and participating amino acids seem to differ substantially between the distinct chemical environments

of the cell membrane and in detergent micelles for solution NMR at high concentrations. mEpoR residues L238 followed by S237 are most sensitive to hEpoR-like mutations, likely both located at the interface of the CHAMP-1 inhibitory complex within cellular lipid membranes. NMR studies in DPC imply strongly perturbed residues I236 (I19 in NMR construct) and S237 (S20) lie at the CHAMP-1/mEpoR interaction interface as well as V234 (V17) and T241 (T24), as expected from our design model. Thus, we conclude the targeted central serine (S237) of the small-X_6-small motif is likely engaged by CHAMP-1 in both cell membranes and detergent micelles, and mEpoR's homodimerization motif is bound by CHAMP-1 in the lowest energy state observed via NMR. However, I236 lies on a distinct face of mEpoR's TM domain than L238 (an α-helix rotates ~102° per residue), suggesting major structural differences in how CHAMP-1 may engage mEpoR in the two distinct apolar environments. On the CHAMP-1 TM domain, bulky isoleucines substituted at small-X_6-small positions, intended to be critical for stabilization of the designed complex, caused significant yet incomplete loss of potency in inhibiting EPO-stimulated BaF3/mEpoR cell proliferation. This behavior may be rationalized by a slightly shifted CHAMP-1/mEpoR complex versus the design model or a second unintended binding mode. Thus, the exact structural basis of these synthetic TM domains is difficult to parse out.

Nonetheless, we demonstrate effective design of small synthetic TM targeting proteins with robust binding affinity, target specificity and control of interaction topology completely from scratch to modulate membrane protein targets of interest. The CHAMP-1 sequence should serve as a novel tool to complement the existing suite of water-soluble engineered polypetides for studying the EPO receptor's signal conduction mechanism(*43-46*) and contribution the complex process of erythropoiesis(*47*). This milestone is instrumental for advancing protein design to generate custom chemical biology tools targeting diverse surfaces of complex membrane proteins.

Methods

Peptide synthesis and purification

TM peptides were synthesized by solid-state synthesis using fmoc chemistry on rink amide resin in microwave-heated reaction vessels by robotic liquid handling (Biotage Initiatior+ Alstra), cleaved using a trifluoroacetic acid (TFA) cocktail (Sigma) from solid-phase resin, and purified by reverse phase HPLC as previously described(*19*). All peptides were produced as C-terminal carboxamides and with free amino N-termini, except for biotin-nCys-mEpoR. Precusor nCys-mEpoR peptide was labeled at its free amino N-terminus as a protected peptide on resin with NHS-Biotin (Sigma) by swelling the resin with DMF having 10 equivalents N,N-Diisopropylethylamine, then adding 1.5 mol equivalents of NHS-Biotin dissolved in minimal DMF to be stirred at room temperature for 45 minutes – performed twice. Peptides were purified by RP-HPLC using a C4 prep column (10 μm 214TP Vydac) using a linear gradient of solvent A (water, 0.1% TFA) and solvent B (60/30/9.9/0.1 isoproanol/acetonitrile/water/TFA). Peptide purity of >95% was achieved in all cases, confirmed using analytical HPLC (C4 Vydac). Correct product masses were confirmed by MALDI mass spectrometry using the matrix α-cyano-4-hydroxycinnamic acid (Sigma).

Peptides for fluorescence quenching cCys CHAMP-1 and nCys mEpoR were derivitized in solution with fluorescein-5-maleimide and (diethylamino-4-methylcoumarin-3-yl)maleimide, respectively. Four mg lyophilized peptide (TFA salts) was dissolved with 10 molar equivalents maleimido-fluorphore in 1 mL DMF, and 0.2 mL water with pH 7 HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid) buffer (final, 25 mM) was added followed by incubating the reaction overnight at room temperature under nitrogen gas on a rotating shaker. Product

fluorescent labelled peptides were re-purified by HPLC as described above.

FRET-based Fluorescence Quenching of TM peptides for Stoichiometry Determination

Following previous published protocols (*40*), we reconstituted donor (diethylamino-4-methylcoumarin-3-yl)maleimido) labeled mEpoR TM peptide at different donor to CHAMP-1 acceptor molar ratios yet at a fixed total peptide concentration across the titration, fixed peptide:detergent ratio, and total equimolar mEpoR to CHAMP-1 ratio. This was done by co-dissolving mEpoR (from a stock solution in trifluoroethanol, TFE, stock solution) with a mixture of unlabeled CHAMP-1 and fluorescein-5-maleimido CHAMP-1 (separately as ethanol stock stock solutions) and C14-B (TFE stock solution) at final concentrations of 1.5 μM donor-labeled mEpoR and 1.5 μM CHAMP-1 (combined unlabeled and acceptor-labelled) with C14-B at 0.5 mM (~100:1 detergent:peptide), 0.725 mM (~175:1), or 0.95 mM (~250:1); assumed C14-B, CMC = 0.2 mM and 90 detergent monomers per micelle. The peptide/detergent mixtures were evaporated, then reconstituted in 50 mM Tris-HCl pH 8, 100 mM NaCl, 0.5 mM EDTA, 5 mM TCEP, subject to bath sonication (30 minutes), vortexed (3 mintutes), equilibrated overnight, then aliquoted in triplicate into 96-well black round-bottom plates and read in a SpectraMax H5 fluorescence monochromator (Molecular Devices). Fluorescence emission scans were recorded after excitation at 410 nm (435 nm cut-off) and relative intensity at 460 nm was plotted for samples of increasing acceptor-labelled CHAMP-1 mole ratios ([donor]/[acceptor]). Emission intensity decay was compared to theoretical quenching expected for different donor/acceptor molecular complexes different stoichiometry using classic theoretical equations (see ref (*41, 48*)). To these theoretical equations, we also added a probabilistic micelle crowding factor using Poisson statistics to calculate the mole fraction of donor additionally quenched due to the probability of non-specific

co-occupation of the same micelle as a acceptor peptide, given the concentration of C14-B in the micelle phase, C14-B's aggregation number, and the detergent:peptide ratio.

Thiol-disulfide Equilibrium Exchange

For the two peptide mixtures (Fig. 4.4) samples were prepared by mixing TFE stock solution of nCys mEpoR, ethanol stock solution of cCys designed TM peptide, and methanol stock solution of either DPC or POPC to a final total peptide to detergent ratio of 1:100 and equimolar peptide ratio. Solutions were dried under nitrogen gas and stored under vacuum overnight, then reconstituted at 100 μM concentration of each peptide in 100mM Tris–HCl pH 8.6, 100 mM KCl and 1 mM EDTA with 0.45 mM oxidized (GSSG) and 1.05 mM reduced (GSH) glutathione to initiate reversible redox conditions. After equilibrium was reached overnight, samples were quenched by acidification by adding HCl to 0.1 M final concentration. Each reaction mixture was separate by analytical reverse phase HPLC using a C4 column (Vydac 214TP, 5 μm) and eluted peaks were collected and identified by mass spectrometry. Peaks from the UV chromatogram were integrated to quantify the relative species fractions of disulfide-bonded dimer species: homodimers and heterodimers.

The novel biotin capture thiol-disulfide exchange was performed to isolate only the mEpoR-interacting species from the 9 possible monomeric or disulfide-bonded species when 3 cysteine-containing peptides are mixed together for competitive reversible oxidation. Samples were prepared by co-dissolving peptides and detergent in organic solvent, evaporating solvent, and reconstitution in aqueous solution. Biotinylated nCys-mEpoR was reconstituted at 100 μM with 4-fold molar excess of CHAMP, 200 μM nCys CHAMP-1 and 200 μM cCys CHAMP-1, in 20 mM C14-B (40:1 detergent to total peptide ratio) in 20 mM Tris pH 8, 50 mM NaCl as well as 3

mM glutathione (5:1, [GSH] / [GSSG]), allowed to undergo reversible oxidation overnight. SDS-PAGE confirmed TM peptide oxidation (Fig. 4.8). The mixture was quenched by reducing the pH to 4.5 by addition of concentrated sodium acetate and bound in batch to streptavidin-conjugated biotin beads overnight – capturing a fraction of the biotin-nCys-mEpoR monomeric and disulfide bonded species. Non-covalently associations were diluted and excess unbound TM peptides were removed by washing the beads (5 bead volumes, 5 times) with 200 mM C14-B, 100 mM sodium acetate, 50 mM NaCl. Peptides disulfide bonded to mEpoR were eluted from the beads by washing with 10 mM TCEP added to the same high C14-B content buffer, reducing disulfide bonds and also diluting non-covalent TM peptide interactions. The eluted fractions were pooled, concentrated and separated by analytical RP-HPLC using linear solvent gradient with C4 column to quantify specie fractions as described above.

Recombinant Expression of Isotopically Enriched mEpoR TM Domain Fragments

Constructs of mEpoR TM domain fused to His-tagged cysteine-free T4 lysozyme or SUMO having a thrombin cleavage site or a sequence-specific nickel-assisted cleavage (SNAC) site, respectively, were cloned into a pET28a(+) vector (Table 4.3). Proteins were expressed in BL21(DE3) or C43 cells in M9 minimal media supplemented with 0.5 g ^{15}N NH$_4$Cl$_2$ (Cambridge isotopes) and/or 3 g ^{13}C glucose (Sigma) per liter as well as appropriately isotope-enriched ISOGROW supplements (Sigma) and induced with 0.4 mM IPTG at an optical density (600 nm) of 0.8 followed by overnight growth at 30 °C or 37 °C. Cells were pelleted, subject to freeze thaw and reconstituted in a lysis buffer effective at solubilizing inclusion bodies: 8 M urea, 0.5 mM EDTA, 50 mM Na Phosphate pH 7.5, 2% (w/v) sodium dodecyl sulfate (SDS). Cycles of tip sonication (10 minutes) and rotary shaking (30 minutes) were repeated until a clear homogeneous

(not viscous or clumpy) solution was achieved. After centrifugation to remove any insoluble debris (30 minutes, 35,000 g), the lysate was poured over a gravity column of Ni-NTA agarose resin (HisPur, Thermo Fisher). The Ni-NTA column was washed with excess detergent-free lysis buffer (10 column volumes), then washed with detergent-free imidazole-containing lysis buffer (20 mM, 4 column volumes), before elution in lysis buffer containing 1% SDS and 250 mM imidazole. For thrombin cleavage, the eluted fraction was concentrated and buffer exchanged into 50 mM Tris pH 8, 100 mM NaCl with 0.1% n-Dodecyl-B-D-Maltoside to remove excess SDS using an 30 kDa amicon centrifugal filter (1/10 dilution, 3 times; EMD Millipore) before a final overnight dialysis (1/500 dilution) using a 20 kDa membrane. For SNAC cleavage, imidazole was removed (<0.5 mM) using an amicon 10 kDa centrifugal filter, buffer exchanged to 100 mM N-Cyclohexyl-2-aminoethanesulfonic acid (CHES) pH 8.5, 100 mM NaCl buffer for the SNAC peptide self-cleavage reaction, followed by addition of $NiCl_2$ to 2 mM final to initiate cleavage. Due to the high residual SDS content that reduces the rate of the SNAC reaction (typically >95% complete in 16 hours), the SUMO-SNAC-mEpoR-TM2 construct underwent SNAC cleavage at 42 °C, typically reaching >95% completion in 24-36 hours. Cleaved isotope-enriched mEpoR TM domain peptide was then purified by RP-HPLC and lyophilized, typically yielding ~10 mg TM peptide per 1 L culture.

Solution NMR in membrane mimics

Samples were prepared from ethanol or methanol stocks of isotopically-enriched recombinant mEpoR TM domain fragments and synthetic CHAMP-1 peptides with stocks of lipid or detergent, dried under nitrogen gas stream, and further dried under vacuum. Samples were reconstituted in our NMR buffer used throughout this study, 40 mM sodium acetate pH 5.2, 20

mM NaCl, 0.5 mM EDTA, 10 mM DTT, 5 % (v/v) D_2O, filtered (0.2 μm), and transferred to a 3 mm Shigemi tube. ^1H-^{15}N and ^1H-^{13}C HSQC spectra were recorded at 45° C for detergent or CHAMP-1 titrations at a fixed mEpoR fragment concentration on a Bruker 800 MHz spectrometer with cryogenic triple-resonance probes. Peptide and detergent concentrations of each sample are detailed in each figure legend. For the monomeric mEpoRTM-2 samples, HNCA and HNCB as well as (H)CC(CO)NH and H(CC)(CO)NH backbone-sidechain TOCSY were recorded on the same Bruker 800 MHz spectrometer for a sample of 400 mM ^2H-DPC and 1 mM $^1H,^{15}N,^{13}C$ mEpoR-TM2. For the CHAMP-1 bound mEpoR sample, HNCA, HNCB, ^{13}C-edited NOESY-HSQC, and ^{13}C-edited/$^{13}C,^{15}N$-filtered HSQC-NOSEY were recorded for a sample of 400 mM ^2H-DPC, 1 mM $^1H,^{15}N,^{13}C$ mEpoR-TM2, and 7 mM unlabeled CHAMP-1 on a Bruker 800 MHz spectrometer. As well, HNCA and HN(CO)CA were recorded for a sample comprised of 800 mM ^2H-DPC, 2 mM $^1H,^{15}N,^{13}C$ mEpoR-TM2, and 12 mM unlabeled CHAMP-1 on a Bruker 900 MHz spectrometer with a triple resonance cryogenic probe.

For mEpoR in 10% (w/v) DMPC/DHPC q=0.3 bicelles titrated with CHAMP-1, the nearly saturating chemical shift perturbations were fit by non-linear least squares to a biomolecular interaction scheme using each peptide's concentration as its mole % (100*mole fraction) relative to DMPC:

$$\Delta\delta_i = \Delta\delta_{max}\left[([A] + [B] + K_D) - \sqrt{([A] - [B] + K_D)^2 - 4[A][B]}\,\right]/2[A] \qquad \text{Equation 1.}$$

where [A] is mEpoR's concentration and [B] is the CHAMP-1 concentration expressed in mole %, while K_D is the bimolecular dissociation constant, $\Delta\delta_{max}$ is the maximum chemical shift perturbation observed, and $\Delta\delta_i$ is the chemical shift at the particular CHAMP-1 concentration measured. K_D was globally fit across all titration series fit while each $\Delta\delta_{max}$ was locally fit for each individual resonance titration series.

Chemical shift perturbation at CA atoms were normalized to the largest induced shift value and plotted on the monomeric TM helix of mEpoR (Fig. 4.15), thusly scaling the relative sphere size and color (green, least perturbed; red, most perturbed) of each CA atom. For sidechain atoms, atom's ^{1}H or ^{13}C having measured shift perturbation were split into three groups by the magnitude of this perturbation (green, least perturbed; yellow, modestly perturbed; red, most perturbed), not include resonances significantly broadened (S13, T11).

Cloning and Vectors for Mammalian Cell Expression

The HA-tagged hEPOR and HA-tagged mEPOR genes were originally obtained from S. Constantinescu, Ludwig Institute, and subcloned into pMSCV-neo (Clontech) using EcoRI and HpaI restriction sites. The chimeric mhmEPOR and mEPOR mutants containing point mutations in the mEPOR TMD were constructed using double-stranded DNA gBlock™ Gene Fragments (Integrated DNA Technologies), as previously described [ref]. The hEPOR-GFP1-10 fusion protein was constructed by replacing the C-terminus of hEPOR downstream of residue 258 with a 10-amino-acid flexible linker (GGSGGGGSGG) followed by the sequences encoding the GFP 1-10 fragment (residue 1 – 215 [need citation]) using DNA gBlock™ Gene Fragments and BglII restriction sites. The mEPOR-GFP1-10 fusion protein was constructed by replacing the sequence encoding hEPOR1-258 in hEPOR-GFP1-10 with the sequence encoding mEPOR1-257 using DNA gBlock™ Gene Fragments and HpaI/BstBI restriction sites.

CHAMP proteins and their mutants were constructed by cloning DNA gBlock™ Gene Fragments into pMSCV-puro by using XhoI and EcoRI restriction sites.

Cells, retrovirus infections, and growth inhibition assays

Human embryonic kidney (HEK) 293T cells were maintained in DMEM-10 medium: Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% fetal bovine serum (FBS) (Gemini Bioproducts), 4 mM L-glutamine, 20 mM HEPES (pH 7.3), and 1X penicillin/streptomycin (P-S). To produce retrovirus stocks, 2 μg pantropic pVSV-G (Clontech), 3 μg pCL- (Imgenex), and 5 μg of the retroviral expression plasmid of interest were mixed with 250 μl of 2x HEBS. 250 μl 0.25 M calcium chloride was then bubbled into each mixture. The mixture (~500 μl) was incubated for 20 minutes at room temperature and then added drop-wise into 2.0 x 10^6 293T cells plated the day before in 100 mm tissue culture dishes in DMEM-10. The cells were incubated with the transfection mixture for 6-8 hours at 37°C, and the medium was replaced with 5 mL fresh DMEM-10 medium. The cells were incubated for another 48 hours at 37°C, then the viral supernatant was harvested, filtered through a 0.45 μm filter (Millipore), and either used immediately or stored at -80°C.

Murine interleukin-3 (IL-3)-dependent BaF3 and derivative cells were maintained in RPMI-10 media: RPMI-1640 supplemented with 10% heat-inactivated FBS, 5% WEHI-3B cell-conditioned medium (as the source of IL-3), 4 mM L-glutamine, 0.06 mM β-mercaptoethanol, and 1X P-S. BaF3 cells expressing mEPOR, hEPOR, and all EPOR mutants and chimeras were generated by infecting BaF3 cells with pMSCV-neo vector containing the desired HA-tagged EPOR gene. 5x10^5 BaF3 cells were washed with phosphate buffered saline (PBS) and then re-suspended in 500 μl RPMI-10 medium with 4 μg/mL polybrene. 500 μl retroviral supernatant or 500 μl DMEM-10 for mock-infection was added to the re-suspended cells that were then incubated for eight hours at 37°C. After incubation, 9 ml RPMI-10 was added and the cells were incubated overnight at 37°C prior to selection in 1 mg/mL G418. Wild-type and mutant CHAMP proteins

cloned in MSCV-puro were introduced into cells by infection, followed by selection in 1 μg/ml puromycin.

For growth inhibition assays, 2×10^5 BaF3 and derivative cells expressing the appropriate genes were washed in PBS three times to remove IL-3. Cell pellets were resuspended in 10 mL RPMI-10 medium lacking WEHI-3B cell-conditioned medium but supplemented with 0.06 U/ml human erythropoietin (Epoetin Alfa, Amgen). Viable cells were counted six to eight days after IL-3 removal. All growth inhibition assays were performed in three independent biological replicates (*i.e.,* independent infections to express TM proteins). All reported experiments included positive and negative controls that performed as expected, and no outliers in these experiments were excluded. All graphs show average values for IL-3 tests +/- standard error of the mean (SEM). Statistical significance of differences between control and experimental samples was evaluated by either one-tailed or two-tailed Student's t-tests with unequal variance, performed using T.TEST function in Microsoft Excel (2013).

Construction and analysis of inducible cell lines

BaF3 cells were transduced to express an engineered version of the tetracycline-controlled transactivator protein, tTA-Advance (tTA), via retroviral infection with the pRetroX-Tet-Off Advanced (Clontech) vector and selection with 1 mg/ml G418. CHAMP-1 cloned in the expression vector pRetroX-TIGHT-puro (Clontech) was introduced into cells expressing tTA by retroviral infection and selection with 1 μg/ml puromycin. HA-mEPOR was retrovirally transduced with pMSCVneo (Clontech) and selected with 0.6 U/ml human erythropoietin (Epoetin Alfa, Amgen) in the absence of IL-3.

To assess expression levels of the CHAMP proteins, BaF3/mEPOR/tTA cells expressing a CHAMP protein were grown in 10 ml cultures in RPMI-10/IL-3 medium in the absence of doxycycline (DOX) or supplemented with 100 or 200 pg/ml DOX for 48 hours. Cells were pelleted on ice in the presence of 1mM phenylmethylsulfonyl fluoride (PMSF) for 10 min at 1,500 RPM at 4°C. Cell extracts were prepared and electrophoresed using 20-30 μg of total protein. After transfer to 0.2 micron PVDF membranes and blocking in 5% milk in TBST (20 mM Tris, 150 mM NaCl, 0.1% Tween 20), blots were incubated overnight 4°C with 1:1000 Anti-FLAG-hrp (Sigma-Aldrich) in 5% Milk in TBST. Blots were then washed and visualized using enhanced chemiluminescence.

For growth inhibition assays, BaF3/mEPOR/tTA cells expressing CHAMP proteins were first cultured in 10 mL RPMI-10/IL-3 medium in the absence of DOX or supplemented with 100 or 200 pg/ml DOX for 48 hours. $2x10^5$ BaF3 cells expressing the appropriate genes were then washed in PBS, resuspended in IL-3-free medium supplemented with the same concentration of DOX and 0.06 U/ml human erythropoietin and counted as described above.

Immunoprecipitation and Immunoblotting

To assess protein phosphorylation, BaF3 cells and their derivatives were first starved in RPMI-10 IL-3-free media for 3 hrs at 37°C, and were then acutely stimulated with 1 U/mL EPO for 10 min at 37°C. Cells were then washed twice with ice-cold PBS containing 1 mM phenylmethylsulfonyl fluoride (PMSF). For phosphotyrosine and phospho-protein blots, 1X HALT Protease and Phosphatase Inhibitor Cocktail (Thermo Scientific) and 500 μM hydrogen peroxide-activated sodium metavanadate were also added to the wash solution. Cells were lysed in FLAG-lysis buffer (50 mM Tris pH 7.4, 150 mM NaCl, 1 mM EDTA, 1% Triton-100)

supplemented with protease and phosphatase inhibitors as described above. All lysates were incubated on ice for 20 minutes, followed by centrifugation at 14,000 rpm for 30 minutes at 4°C. The total protein concentration of the supernatants was determined using a bicinchoninic acid (BCA) protein assay kit (Pierce).

To immunoprecipitate FLAG-tagged CHAMP peptides, 50 μl of anti-FLAG M2 matrix gel (Sigma-Aldrich) was added to 0.5 mg of total protein and rotated overnight at 4°C. Immunoprecipitated samples were washed four times with 1 mL NET-N buffer (100 mM NaCl, 0.1 mM EDTA, 20 mM Tris-HCl pH 8.0, 0.1% Nonidet P-40) supplemented with protease inhibitors as above, pelleted and re-suspended in 2x Laemmli sample buffer (2x SB) supplemented with 200 mM dithiothreitol (DTT) and 5% β-mercaptoethanol (β-ME). Precipitated proteins and whole cell lysates were heated at 95°C for 5 min and then resolved by SDS-PAGE on either 7.5%, 10%, or 20% polyacrylamide gels according to the size of the protein of interest. The resolving gel was then transferred by electrophoresis to a 0.2 μm nitrocellulose or PVDF membrane. 0.09% SDS was added to the transfer buffer for membranes used to detect phosphorylated proteins.

Membranes were blocked with gentle agitation for two hours at room temperature in 5% nonfat dry milk/TBST. To detect the phosphorylated forms of JAK2 and STAT5, anti-phospho-JAK2 (Tyr1008) (clone D4A8, Cell Signaling) and anti-phospho-STAT5 (Y694) #9351 (Cell Signaling) were used. To detect the total JAK2 and STAT5, anti-JAK2 (clone D2E12, Cell Signaling) and anti-STAT5 #9363 (Cell Signaling) were used. An HRP-conjugated mouse anti-HA (clone 6E2, Cell Signaling) was used to detect the HA-tagged EPOR and all EPOR mutants. All antibodies were used at 1:1000 dilution. Membranes were incubated overnight with gentle agitation in primary antibody at 4°C, washed five times in TBST, and then incubated with gentle agitation for one hour at room temperature in donkey anti-mouse or donkey anti-rabbit HRP

175

(Jackson Immunoresearch), as appropriate, at a 1:10,000 dilution. To re-probe membranes, they were stripped in Restore Western Stripping Buffer (Thermo Scientific) for 15 min at room temperature with gentle agitation, washed five times in TBST, blocked in 5% milk/TBST for one hour at room temperature, and incubated overnight at 4°C with antibody, as described above. Membranes were incubated with Super Signal West Pico or Femto Chemiluminescent Substrates (Pierce) to detect protein bands.

Split GFP Complementation Assay

BaF3 cells were transduced to express GFP1-10 fragment fused to EPOR, GFP11 fragment fused to CHAMP-1, or both. The EPOR-GFP1-10 fusion protein consists of (from N-terminus to C-terminus) the residue 1-258 from hEPOR or residue 1-257 from mEPOR, a 10-amino-acid flexible linker (GGSGGGGSGG), and the GFP1-10 segment. The GFP11-N1 fusion proteins consists of (from N-terminus to C-terminus) a FLAG tag, the GFP11 segment (residue 216-231), a GGG linker, and the CHAMP-1 sequence. For flow cytometry, 5×10^5 cells were collected by centrifugation at 1,000 rpm for 10 minutes at 4°C, washed in cold PBS, and re-suspended in 300ul cold PBS. Cells were then analyzed by Cytoflex using green laser and plotted in FlowJo.

Author Contributions

M.M. and H.H. performed the computational design. L.H., W.B., and A.E. performed all cell biology experiments. M.M., H.H., and H.J. synthesized and purified the peptides. M.M. and H.H. executed thiol-disulfide exchange experiments. M.M. and S.E.N. conducted the fluorescence quenching experiments. M.M., H.T.K., and Y.W. performed the NMR experiments. M.M., L.H., D.M., and W.F.D. designed the experiments, analyzed the data, and wrote the manuscript.

References

1.   J. M. Westerfield, F. N. Barrera, Membrane receptor activation mechanisms and transmembrane peptide tools to elucidate them. *Journal of Biological Chemistry* **295**, 1792-1814 (2020).

2.   T. A. Stone, C. M. Deber, Therapeutic design of peptide modulators of protein-protein interactions in membranes. *Biochim Biophys Acta* **1859**, 577-585 (2017).

3.   V. Ovchinnikov, T. A. Stone, C. M. Deber, M. Karplus, Structure of the EmrE multidrug transporter and its use for inhibitor peptide design. *Proceedings of the National Academy of Sciences* **115**, E7932-E7941 (2018).

4.   D. S. Alves *et al.*, A novel pH-dependent membrane peptide that binds to EphA2 and inhibits cell migration. *Elife* **7**, (2018).

5.   A. W. Partridge, R. A. Melnyk, D. Yang, J. U. Bowie, C. M. Deber, A transmembrane segment mimic derived from Escherichia coli diacylglycerol kinase inhibits protein activity. *J Biol Chem* **278**, 22056-22060 (2003).

6.   L. He, N. Shobnam, K. Hristova, Specific inhibition of a pathogenic receptor tyrosine kinase by its transmembrane domain. *Biochim Biophys Acta* **1808**, 253-259 (2011).

7.   H. Yin *et al.*, Computational design of peptides that target transmembrane helices. *Science* **315**, 1817-1822 (2007).

8.   E. N. Heim *et al.*, Biologically active LIL proteins built with minimal chemical diversity. *Proc Natl Acad Sci U S A* **112**, E4717-4725 (2015).

9.   L. He, A. R. Hoffmann, C. Serrano, K. Hristova, W. C. Wimley, High-throughput selection of transmembrane sequences that enhance receptor tyrosine kinase activation. *J Mol Biol* **412**, 43-54 (2011).

10.   E. Bloch *et al.*, Disrupting the transmembrane domain-mediated oligomerization of protein tyrosine phosphatase receptor J inhibits EGFR-driven cancer cell phenotypes. *J Biol Chem* **294**, 18796-18806 (2019).

11.   T. J. Cammett *et al.*, Construction and genetic selection of small transmembrane proteins that activate the human erythropoietin receptor. *Proceedings of the National Academy of Sciences* **107**, 3447-3452 (2010).

12.   L. He *et al.*, Transmembrane Protein Aptamer Induces Cooperative Signaling by the EPO Receptor and the Cytokine Receptor beta-Common Subunit. *iScience* **17**, 167-181 (2019).

13.   L. He *et al.*, Single methyl groups can act as toggle switches to specify transmembrane Protein-protein interactions. *Elife* **6**, (2017).

14.   B. M. Kroncke *et al.*, Documentation of an Imperative To Improve Methods for Predicting Membrane Protein Stability. *Biochemistry* **55**, 5002-5009 (2016).

15.   J. Y. Weinstein, A. Elazar, S. J. Fleishman, A lipophilicity-based energy function for membrane-protein modelling and design. *PLoS Comput Biol* **15**, e1007318 (2019).

16.   R. F. Alford, P. J. Fleming, K. G. Fleming, J. J. Gray, Protein Structure Prediction and Design in a Biologically Realistic Implicit Membrane. *Biophys J* **118**, 2042-2055 (2020).

17.   M. G. Teese, D. Langosch, Role of GxxxG Motifs in Transmembrane Domain Interactions. *Biochemistry* **54**, 5125-5135 (2015).

18.   W. P. Russ, D. M. Engelman, The GxxxG motif: a framework for transmembrane helix-helix association. *J Mol Biol* **296**, 911-919 (2000).

19.   M. Mravic *et al.*, De novo designed transmembrane peptides activating the alpha5beta1 integrin. *Protein engineering, design & selection : PEDS* **31**, 181-190 (2018).

20. D. L. Johnson *et al.*, Identification of a 13 amino acid peptide mimetic of erythropoietin and description of amino acids critical for the mimetic activity of EMP1. *Biochemistry* **37**, 3699-3710 (1998).

21. J. S. Philo, K. H. Aoki, T. Arakawa, L. O. Narhi, J. Wen, Dimerization of the extracellular domain of the erythropoietin (EPO) receptor by EPO: one high-affinity and one low-affinity interaction. *Biochemistry* **35**, 1681-1691 (1996).

22. R. F. Alford *et al.*, An Integrated Framework Advancing Membrane Protein Modeling and Design. *PLoS Comput Biol* **11**, e1004398 (2015).

23. A. Z. Ebie, K. G. Fleming, Dimerization of the erythropoietin receptor transmembrane domain in micelles. *J Mol Biol* **366**, 517-524 (2007).

24. W. Ruan, V. Becker, U. Klingmüller, D. Langosch, The Interface between Self-assembling Erythropoietin Receptor Transmembrane Segments Corresponds to a Membrane-spanning Leucine Zipper. *Journal of Biological Chemistry* **279**, 3273-3279 (2004).

25. S. N. Constantinescu *et al.*, Ligand-independent oligomerization of cell-surface erythropoietin receptor is mediated by the transmembrane domain. *Proceedings of the National Academy of Sciences* **98**, 4379-4384 (2001).

26. K. F. Kubatzky *et al.*, Self assembly of the transmembrane domain promotes signal transduction through the erythropoietin receptor. *Curr Biol* **11**, 110-115 (2001).

27. N. Seubert *et al.*, Active and inactive orientations of the transmembrane and cytosolic domains of the erythropoietin receptor dimer. *Molecular cell* **12**, 1239-1250 (2003).

28. S. N. Constantinescu *et al.*, Activation of the erythropoietin receptor by the gp55-P viral envelope protein is determined by a single amino acid in its transmembrane domain. *Embo j* **18**, 3334-3347 (1999).

29.   Y. Liu, D. M. Engelman, M. Gerstein, Genomic analysis of membrane protein families: abundance and conserved motifs. *Genome Biol* **3**, research0054 (2002).

30.   Y. Zhang, D. W. Kulp, J. D. Lear, W. F. DeGrado, Experimental and computational evaluation of forces directing the association of transmembrane helices. *J Am Chem Soc* **131**, 11341-11343 (2009).

31.   L. Adamian, J. Liang, Interhelical hydrogen bonds and spatial motifs in membrane proteins: polar clamps and serine zippers. *Proteins* **47**, 209-218 (2002).

32.   B. North *et al.*, Characterization of a membrane protein folding motif, the Ser zipper, using designed peptides. *J Mol Biol* **359**, 930-939 (2006).

33.   S. Q. Zhang *et al.*, The membrane- and soluble-protein helix-helix interactome: similar geometry via different interactions. *Structure* **23**, 527-541 (2015).

34.   A. Senes *et al.*, E(z), a depth-dependent potential for assessing the energies of insertion of amino acid side-chains into membranes: derivation and applications to determining the orientation of transmembrane and interfacial helices. *J Mol Biol* **366**, 436-448 (2007).

35.   S. J. Fleishman *et al.*, RosettaScripts: a scripting language interface to the Rosetta macromolecular modeling suite. *PLoS One* **6**, e20161 (2011).

36.   W. Sheffler, D. Baker, RosettaHoles: rapid assessment of protein core packing for structure prediction, refinement, design, and validation. *Protein Sci* **18**, 229-239 (2009).

37.   L. Cristian, J. D. Lear, W. F. DeGrado, Determination of membrane protein stability via thermodynamic coupling of folding to thiol-disulfide interchange. *Protein Sci* **12**, 1732-1740 (2003).

38.   L. Cristian, J. D. Lear, W. F. DeGrado, Use of thiol-disulfide equilibria to measure the energetics of assembly of transmembrane helices in phospholipid bilayers. *Proceedings of the National Academy of Sciences* **100**, 14772-14777 (2003).

39.   J. D. Lear, H. Gratkowski, L. Adamian, J. Liang, W. F. DeGrado, Position-dependence of stabilizing polar interactions of asparagine in transmembrane helical bundles. *Biochemistry* **42**, 6400-6407 (2003).

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

41.   M. Li *et al.*, A fluorescence energy transfer method for analyzing protein oligomeric structure: application to phospholamban. *Biophys J* **76**, 2587-2599 (1999).

42.   S. Cabantous, T. C. Terwilliger, G. S. Waldo, Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. *Nat Biotechnol* **23**, 102-107 (2005).

43.   O. Livnah *et al.*, An antagonist peptide–EPO receptor complex suggests that receptor dimerization is not sufficient for activation. *Nature Structural Biology* **5**, 993-1004 (1998).

44.   O. Livnah *et al.*, Functional mimicry of a protein hormone by a peptide agonist: the EPO receptor complex at 2.8 A. *Science* **273**, 464-471 (1996).

45.   R. S. Syed *et al.*, Efficiency of signalling through cytokine receptors depends critically on receptor orientation. *Nature* **395**, 511-516 (1998).

46.   H. Zhang, I. A. Wilson, R. A. Lerner, Selection of antibodies that regulate phenotype from intracellular combinatorial antibody libraries. *Proc Natl Acad Sci U S A* **109**, 15728-15733 (2012).

47.   K. Mohan *et al.*, Topological control of cytokine receptor signaling induces differential effects in hematopoiesis. *Science (New York, N.Y.)* **364**, eaav7532 (2019).

48.   J. D. Lear, A. L. Stouffer, H. Gratkowski, V. Nanda, W. F. Degrado, Association of a model transmembrane peptide containing gly in a heptad sequence motif. *Biophys J* **87**, 3421-3429 (2004).
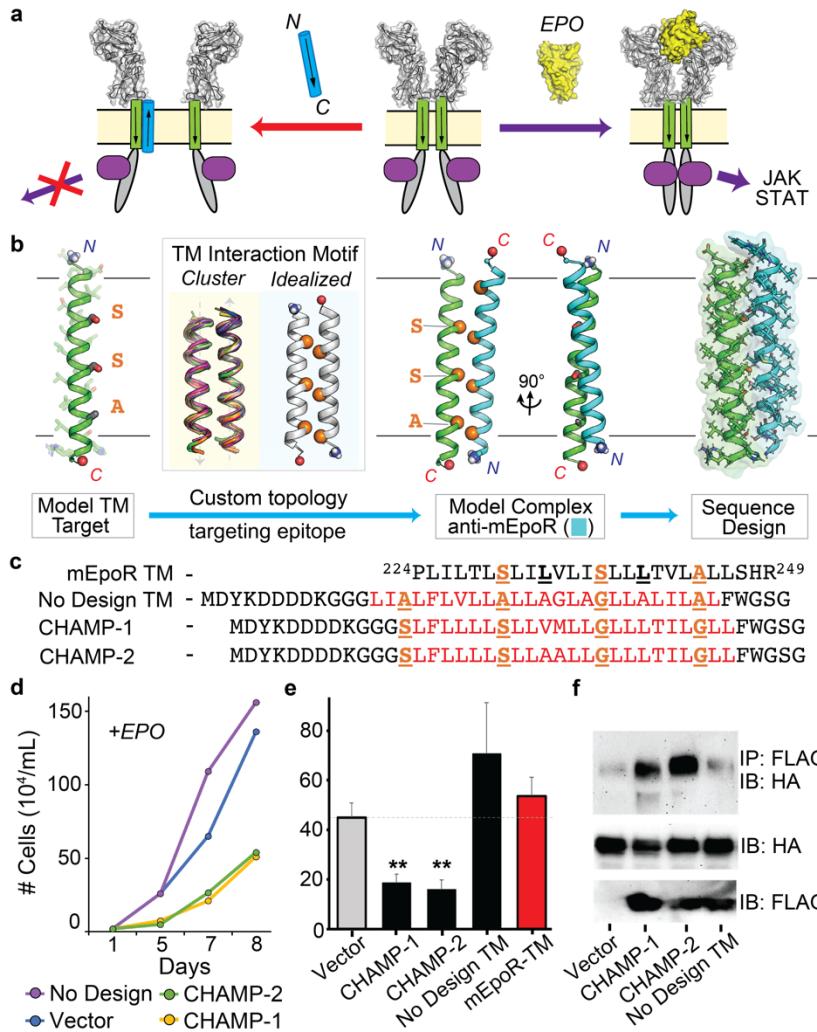
Figure 4.1. Computational design of anti-mouse erythropoietin receptor (mEpoR) TM domain peptides **(a)** EpoR ligand-independent monomer-homodimer equilibrium is mediated in part by transmembrane (TM) domains. Erythropoietin (EPO, green) binding stabilizes a homodimer conformation activating JAK/STAT signaling. Computed Helical Anti-Membrane Proteins (CHAMPs, red) targeting mEpoR's TM homodimerization epitope in a non-native anti-parallel geometry to competitively inhibit signaling. **(c)** mEpoR TM domain, Ser-X_6-Ser-X_6-Ala motif (orange), overlapping Ser-Leu (underlined) repeat. Expressed FLAG-tagged constructs: synthetic TM regions (red), small-X_6-small motifs (orange, underlined). **(d)** Representative EPO-induced cell proliferation curve of mouse BaF3 cells stably expressing mEpoR transfected with an empty vector or the synthetic TM domains in IL-3-free media with 0.06 U/mL EPO. Proliferation is significantly impaired by CHAMP-1 and CHAMP-2 expression. **(e)** Cell counts (10^4/mL) at day 6 summarizing at least 3 independent experiment with standard error denoting significant differences (p-values < 0.05) versus vector control cells by asterisks. No Design control TM and a minimal mEpoR TM construct (mEpoR-TM) are not inhibitory. **(f)** Co-immunoprecipitation (co-IP) of FLAG-tagged CHAMP-1 and CHAMP-2, but not the No Design Control TM with HA-tagged mEpoR (IB, western immunoblot; IP: immunoprecipitation) from co-expressing BaF3 cells; representative of 3 experiments.

184
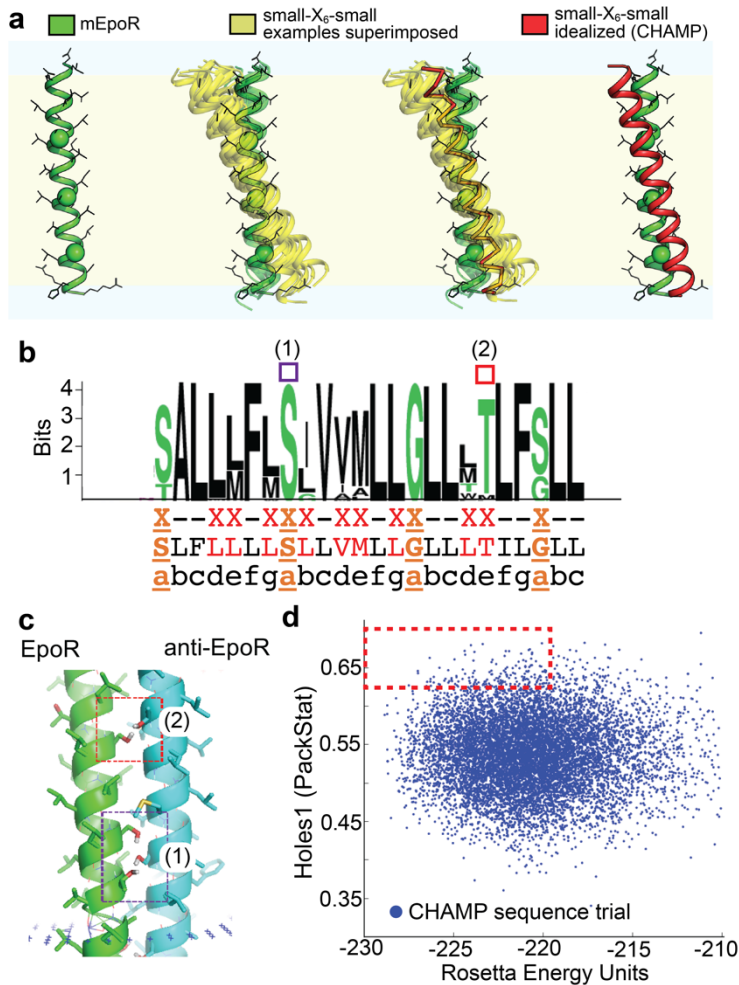
Figure 4.2. Protein design and ranking of output anti-mEpoR CHAMP sequences.
(a) Given the position of mEpoR's TM domain (green, cartoon) in an implicit membrane and its small-X_6-small sequence (spheres), natural examples of the dimeric antiparallel structural cluster enriched in mutual small-X_6-small repeats at interface positions can be superimposed by one helix onto mEpoR's helix, matching their small-X_6-small registry. The second interacting helix of these many examples (yellow, partially transparent cartoon) are roughly overlaid – potential CHAMP backbone coordinated positioned to bind mEpoR's small-X_6-small residues in an antiparallel orientation. From the coordinate sets an idealized model can be built of the interacting pair of helices, where the idealized mEpoR interaction partner helix (red) is shown both as CA ribbon with natural examples (yellow) overlaid or alone (*right*). **(b)** Sequence Logo of all sequence output from CHAMP design, showing residues that were fixed (black hyphen), those allowed to sample any residue from our lipid-friendly library (red X; GATSVLIFM), and designated small-X_6-small positions (orange X). **(c)** Output CHAMP/mEpoR design model showing tight interaction and some inter-helical hydrogen bonds, with colored boxes corresponding to CHAMP residues denoted in **(b)**. **(d)** Scatter plot of each sequence design trial's model Rosetta energy and packing voids score (RosettaHoles1(*36*) via the "packstat" filter in RosettaScripts(*35*)). Only sequences with Rosetta energy greater than the mean energy of all designs were considered, and of those, only those having Holes scores in the top 10% were clustered by sequence similarity to give the final CHAMP-1 and CHAMP-1 sequences.

185
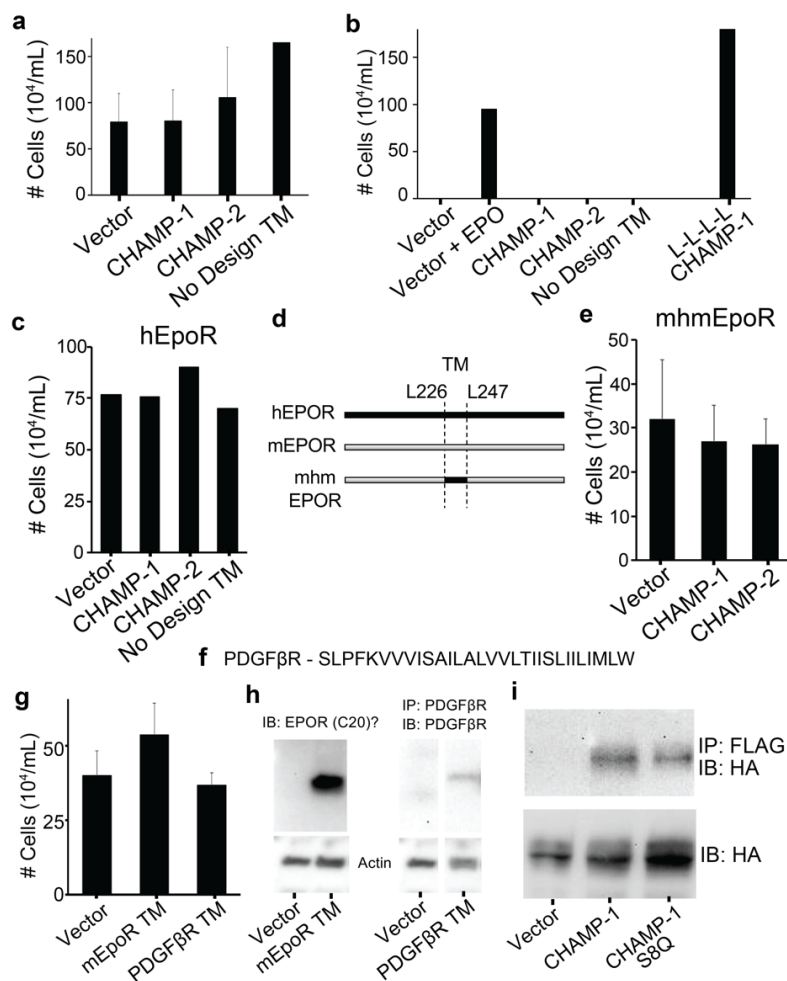
Figure 4.3. Effects of expressed exogenous TM domain on EpoR expressing BaF3 cells.
(a) BaF3/mEpoR cells expressing empty vector, CHAMP-1, CHAMP-1, or No Design TM were incubated in medium supplemented with 5% WEHI-conditioned medium as source of IL-3. The number of live cells at day 4 is shown. Three replicates are shown with standard deviations as error bars for all except for No Design TM (n=1). **(b)** Artificial TM domains do not stimulate cell growth of BaF3/mEpoR cells in the absence of IL-3 and EPO, except for CHAMP-1 mutant L-L-L-L replacing the small-X_6-small motif residues with Leu. **(c)** BaF3/hEPOR cells expressing empty vector, CHAMP-1, CHAMP-1, or No Design TM were incubated in IL-3-free medium supplemented with 0.06 U/mL EPO. The number of live cells at day 4 is shown. **(d)** construct 'mhm' EpoR where human EpoR TM domain is used in place of the mEpoR TM domain. **(e)** BaF3/mhmEpoR cells expressing empty vector, CHAMP-1, CHAMP-1, or No Design TM were incubated in IL-3-free medium supplemented with 0.06 U/mL EPO. The average number of live cells at day 8 is shown with standard deviation of three independent experiments. **(f)** TM domain sequence of unrelated PDGFβR receptor. (**g**) BaF3/mEpoR cells expressing empty vector, a short mEpoR TM domain construct plus signal peptide, and a PDGFβR TM domain construct in IL-3-free medium supplemented with 0.06 U/mL EPO, showing the average number of live cells at day 8 is shown with standard deviation of three independent experiments. (**h**) Western blots; the mEpoR TM construct (C20 cytoplasmic epitope) and PDGFβR TM construct are expressed.

186
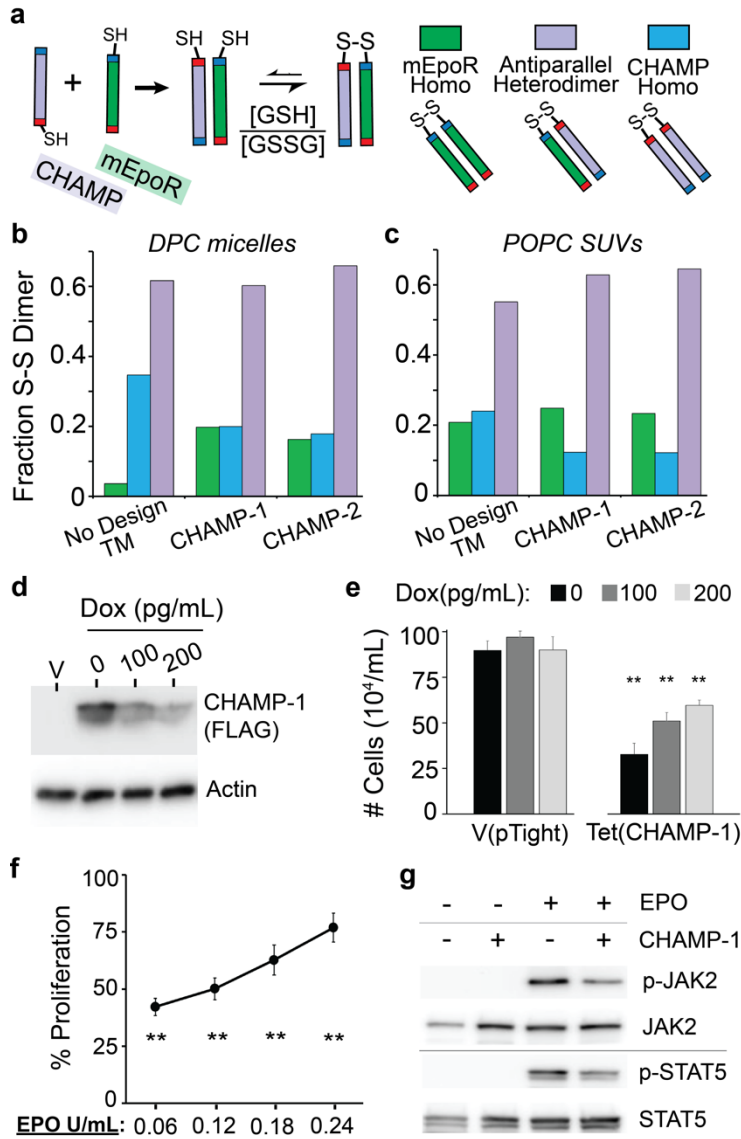
Figure 4.4. Designed TM peptides target mEpoR TM domain to inhibit the EPO-mEpoR signaling cascade.
(a) Equilibrium thiol-disulfide exchange measures the population of non-covalent complexes by the distribution of disulfide-bonded species formed via altered thiol proximity. Association and overnight reversible oxidation of 100 μM of each N-terminal cysteine mEpoR TM peptide (green) and a C-terminal cysteine designed TM peptide (purple) (sequences, Table S1) in 50 mM Tris-HCl, 200 mM KCl, 0.5 mM ethylenediaminetetraacetic acid (EDTA) buffer pH 8.6 with 10 equivalents glutathione (3:1 reduced:oxidized) solubilized at 1:50 molar ratio to **(b)** dodecylphosphocholine (DPC) detergent micelles or **(c)** 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) small unilamerlar vesicles (SUVs) giving a mixture of covalent mEpoR TM homodimer (green), designed heterodimer (purple), designed peptide homodimer (cyan). Species fractions measured by reverse phase liquid chromatography show a strong non-random (>2-fold) preference for the antiparallel mEpoR/designed peptide complex in all cases. **(d)** Doxycycline (Dox) repression of CHAMP-1 expression levels from tetracycline-responsive

promoter Tet(CHAMP-1) in BaF3/mEpoR cells expressing the tTA tetracycline transactivator: IB of lysates; V, empty vector. **(e)** BaF3/mEpoR cell proliferation in IL-3-free media with 0.06 U/mL EPO and either 0, 100, or 200 pg/mL Dox shows Tet(CHAMP-1) at day 6 (n=3; bars, standard error) CHAMP-1-concentration dependent growth inhibition versus vector only, V(pTight). **(f)** CHAMP-1 growth inhibition in (n=4) is decreased at increasing EPO concentration suggesting competition with EPO-mEpoR signal conduction. **(g)** IB detecting reduced phosphorylation of JAK2 and STAT5 after 10 minutes of 1 U/mL EPO stimulation when CHAMP-1 is co-expressed, BaF3/mEpoR cells. Total JAK2 and STAT5 levels do not change.

Significance (p<0.05) versus vector control denoted by asterisks.
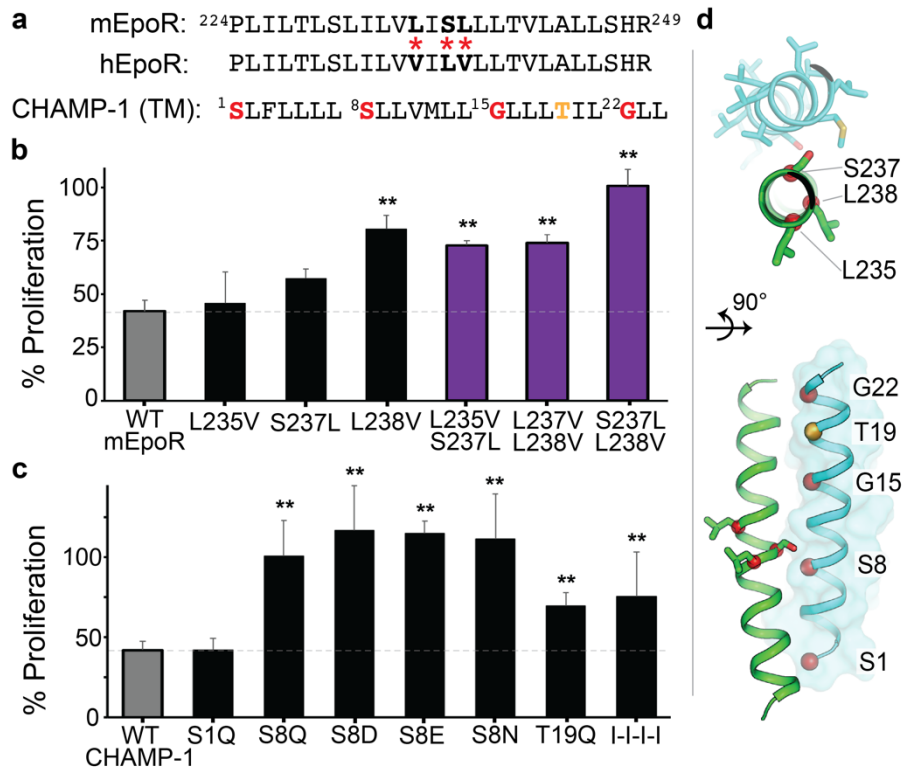
Figure 4.5. Sequence-specific interaction between mEpoR and CHAMP-1 (a) Wild-type (WT) sequences for core TM regions of mouse EpoR (mEpoR), human EpoR (hEpoR), and CHAMP-1 (peptide or expressed protein constructs). Dissimilar residues between mEpoR and hEpoR, bold and red astericks, mutated in **(b)**. CHAMP-1 interface residues mutated in **(c)** bolded and colored: small-X_6-small repeat Ser1-Ser8-Gly15-Gly22, red; Thr19, orange. **(b)** BaF3 stably expressing single and double mutants of mEpoR to hEpoR-like amino acids expressed with WT CHAMP-1 cultured in IL3-free media supplemented with 0.06 U/mL EPO. Cell counts at day 6 are shown normalized to vector control for 3 independent experiments (bars, standard error) as percentage growth inhibition by CHAMP-1. Of single mutants, CHAMP-1 potency is mitigated most by L238V and is negligibly perturbed by L235V. All double mutants have significantly reduced CHAMP-1 inhibition, with S237L-L237V completely reversing CHAMP-1 effects. **(c)** CHAMP-1 mutants co-expressed in BaF3 cells expressing WT mEpoR as percentage growth inhibition at day 6 (n=3), numbering from **(a)**. Strongly polar mutants to lipid-embedded positions S8 and T19 are significantly less inhibitory, with E, D, Q, and N completely reversing CHAMP-1. Quadruple mutant I-I-I-I, Ile1-Ile8-Ile15-Ile22, targeting the small-X_6-small motif is significantly less potent, yet unexpectedly maintains some inhibition to EPO-induced cell proliferation. **(d)** Design model of mEpoR (green) and CHAMP-1 (cyan) TM complex with anticipated interface residues subject to mutation labeled: mEpoR, sticks and red Cα atom spheres; CHAMP-1, Cα atom spheres, colored as **(a)**.
Significance (p<0.05) in difference versus WT mEpoR with WT CHAMP-1 denoted by asterisks.
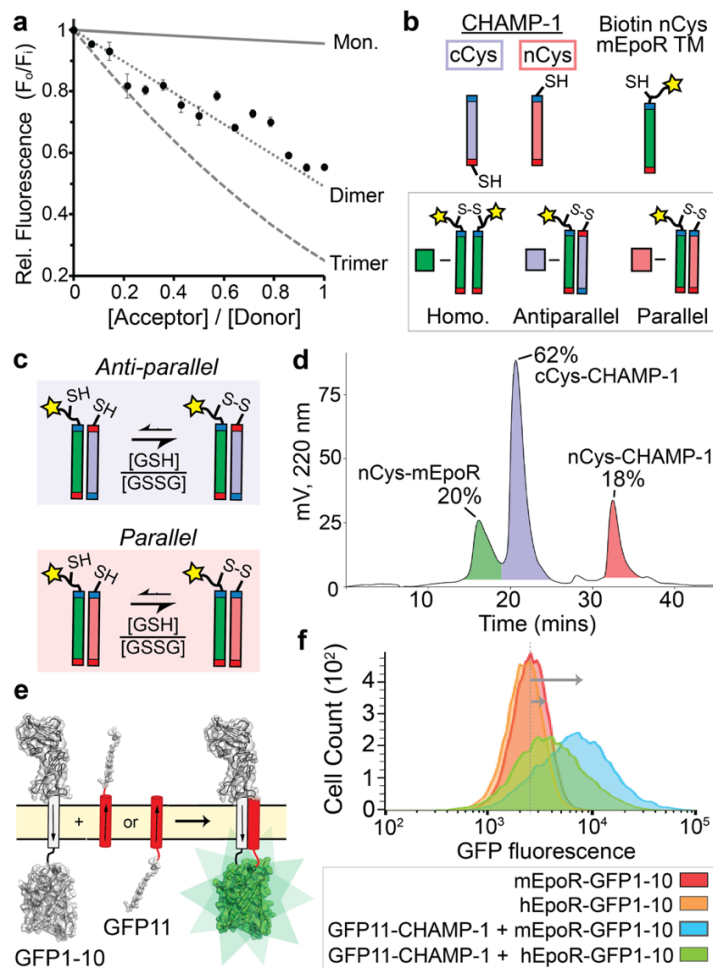
Figure 4.6. **Dimeric anti-parallel topology of the mEpoR/CHAMP-1 complex**
(a) Relative (Rel.) 460 nm fluorescence emission of 1.5 µM 7-diethylamino-4-methylcoumerin-labeled mEpoR TM peptide shows FRET quenching as fluorescein-labeled CHAMP-1 is titrated in 0.6 mM C14-betaine, 50 mM Tris-HCl pH 8, 100 mM NaCl, 0.5 mM EDTA, 5 mM TCEP buffer at fixed equimolar total peptide concentration (n=3; bars, standard error). Theoretical FRET curves for monomeric (Mon.), dimeric, and trimeric complexes including a micelle crowding factor (gray). **(b)** N-terminal (nCys) and C-terminal cysteine (cCys) CHAMP-1 variants with a biotin-labeled N-terminal cysteine (biotin nCys) mEpoR TM peptide (sequence, Table S1) reconstituted (2:2:1 ratio, 0.5 mM total) in 20 mM C14-betaine. Reversible oxidation and streptadvidin-bead capture yields mEpoR homodimer (green) and heterodimers with nCys (red) or cCys (purple) CHAMP-1. **(c)** The non-covalent topology (parallel or antiparallel) of the CHAMP-1/mEpoR complex determines which of nCys- or cCys-CHAMP-1 preferentially forms disulfide bonds with nCys-mEpoR. **(d)** Representative reverse phase liquid chromatogram in an isopropanol/acetonitrile/water/TFA (60:30:9.9:0.1) gradient separating eluted monomeric TM peptides after capture. **(e)** Split green fluorescent protein (GFP) complemention assay in BaF3 cells. mEpoR and hEpoR constructs (gray) have GFP fragment with β-strands 1-10 (GFP1-10) replacing the cytoplasmic domain. Constructs of CHAMP-1 fused to GFP's 11th β-strand (GFP11) at its N or C terminus. Sequences, Table 4.3. **(f)** Cellular fluorescence when GFP11-CHAMP-1 is co-expressed with mEpoR-GFP1-10 (representative plot, n=3).

190
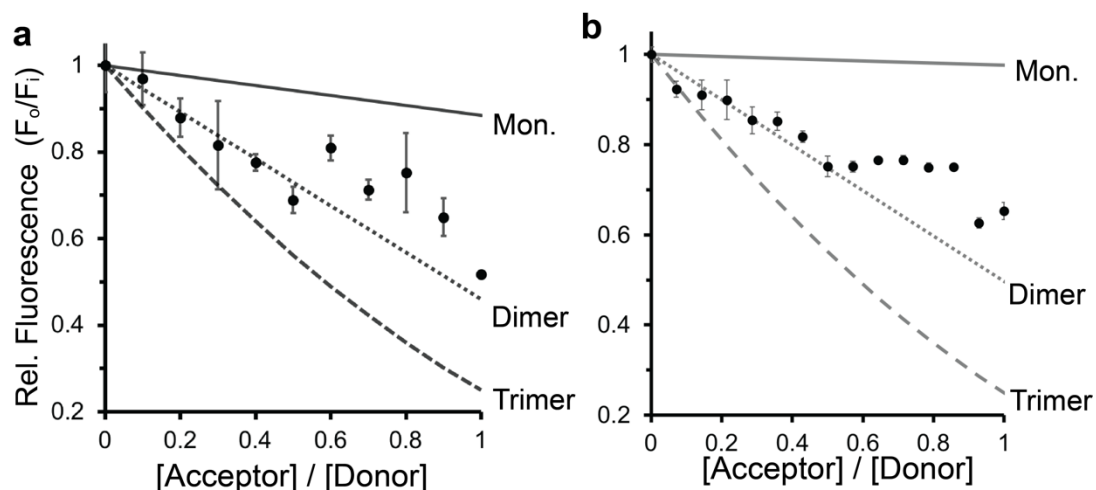
Figure 4.7. Fluorescence quenching of labeled-mEpoR by CHAMP-1 in C14B micelles at different peptide:detergent ratios.
Relative (Rel.) 460 nm fluorescence emission of 1.5 µM 7-diethylamino-4-methylcoumerin-labeled mEpoR TM peptide shows FRET quenching as fluorescein-labeled CHAMP-1 is titrated in C14-betaine, 50 mM Tris-HCl pH 8, 100 mM NaCl, 0.5 mM EDTA, 5 mM TCEP buffer at fixed equimolar total peptide concentration (n=3-6; bars, standard error). Theoretical FRET curves for monomeric (Mon.), dimeric, and trimeric complexes including a micelle crowding factor (gray). **(a)** Using a detergent to total peptide ratio of 100:1, or 1.1 micelles per total peptide. **(b)** Using a detergent to total peptide ratio of 250:1, or 2.8 micelles per total peptide.
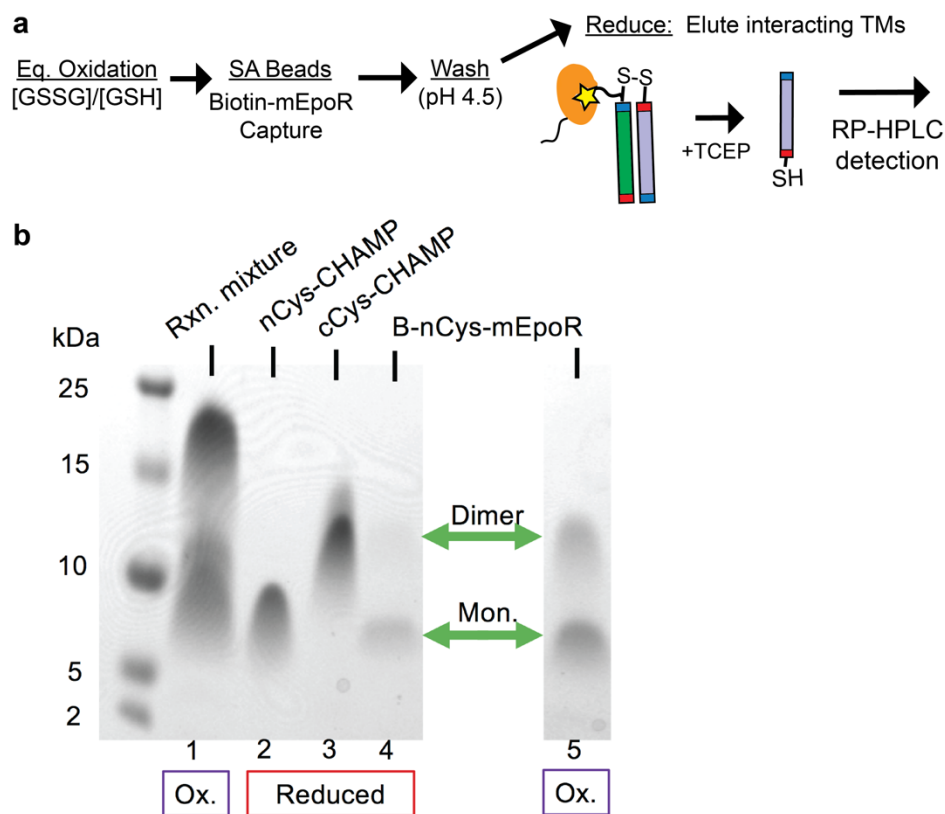
Figure 4.8. Schematic and SDS-PAGE of the three peptide thiol-disulfide exchange with biotin capture

(a) Workflow for equilibrium thiol-disulfide exchange modified for biotin capture and isolation of only mEpoR-containing disulfide bonded TM dimers, followed by reduction and quantification of the interacting TM helices by reverse phase high-performance liquid chromatography (RP-HPLC). SA, Streptavidin; GSSG, GSH, oxidized and reduced glutathione, respectively; Eq., equilibrium. **(b)** Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) of equilibrated oxidized TM peptides in 20 mM myristyl sulfobetaine (C14B) micelles (lanes 1, 5), or purified peptides (lanes 2, 3, 4) reduced in 10 molar equivalents of tris(2-carboxyethyl)phosphine (TCEP). *Lane 1*, mixture of 100 µM Biotinylated nCys-mEpoR with 4-fold molar excess of CHAMP – 2 equivalents of each nCys-CHAMP1 and cCys-CHAMP1 V2 oxidized overnight at a molar ratio of oxidized to reduced glutathione of 0.2. *Lane 5*, 100 µM Biotinylated nCys-mEpoR alone oxidized under the same conditions. Green arrows denote mEpoR monomer and dimer bands. Ox., oxidizing conditions; Mon., Monomer; kDa, kilodaltons; Rxn., Reaction.
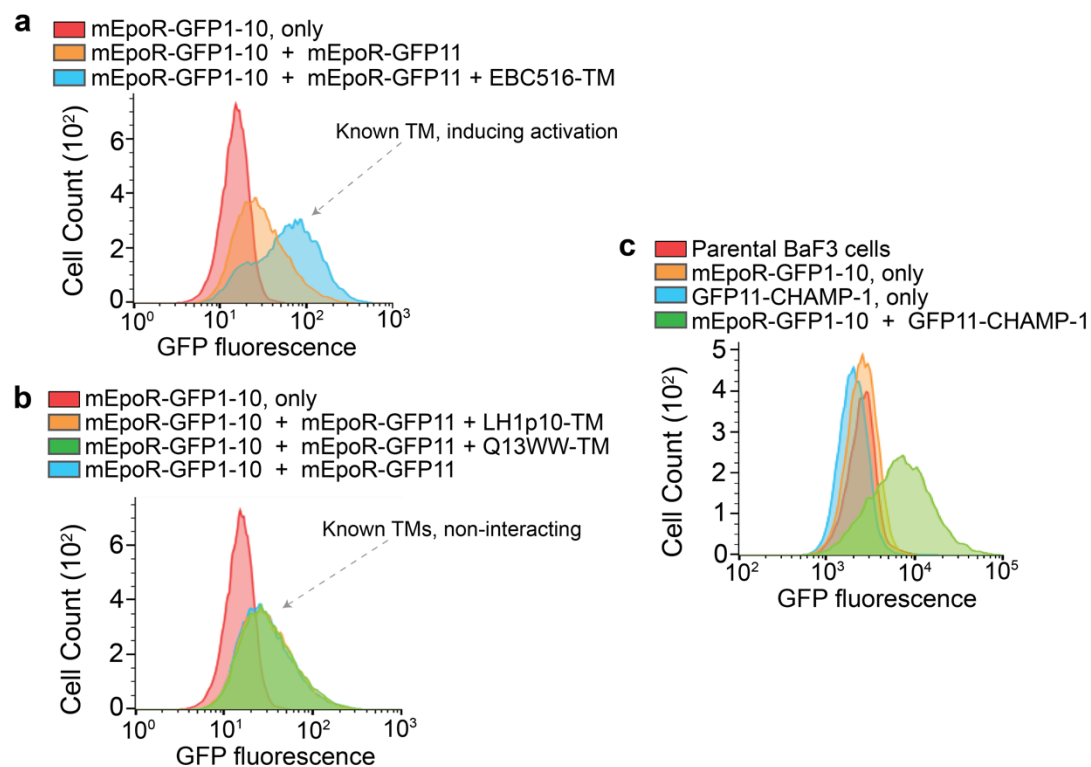
Figure 4.9. Control hEpoR and mEpoR GFP protein complement experiments
(a) Flow cytometry of hEpoR-GFP1-10 expressed alone, hEpoR-GFP1-10 and hEpoR-GFP11 co-expressed, and hEpoR-GFP1-10 and hEpoR-GFP11 co-expressed with known activating synthetic TM protein EBC516-TM, inducing hEpoR dimerization and GFP complement. **(b)** Flow cytometry of hEpoR-GFP1-10 expressed alone, hEpoR-GFP1-10 and hEpoR-GFP11 co-expressed, and hEpoR-GFP1-10 and hEpoR-GFP11 co-expressed with known non-interacting TM proteins (LH1p10-TM, Q13WW-TM), representing the background complementation of non-interacting species. **(c)** Flow cytometry of parental BaF3 cells, BaF3/mEpoR-GFP1-10 cells, BaF3/GFP11-CHAMP-1 cells, and BaF3 cells co-expressing mEpoR-GFP1-10 cells and GFP11-CHAMP-1.
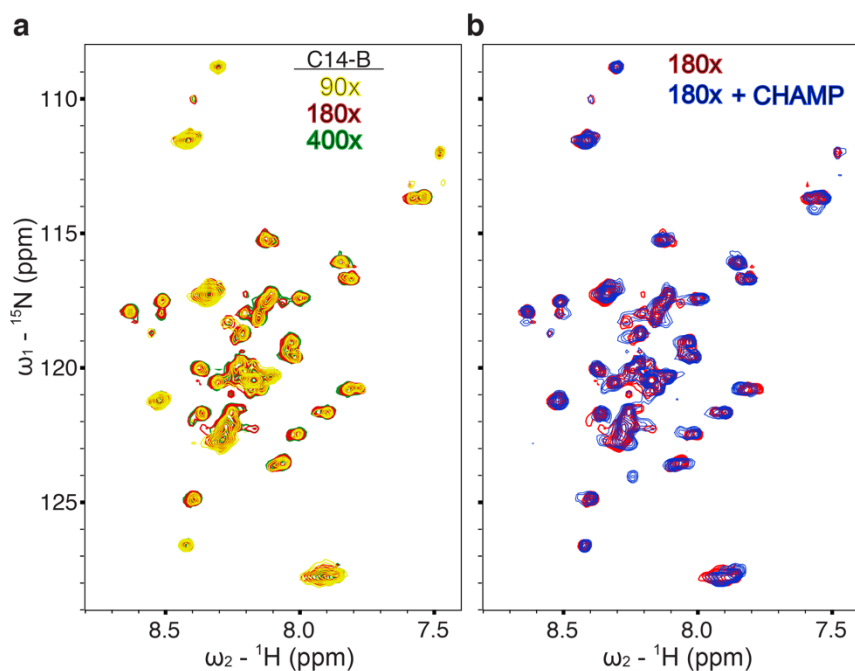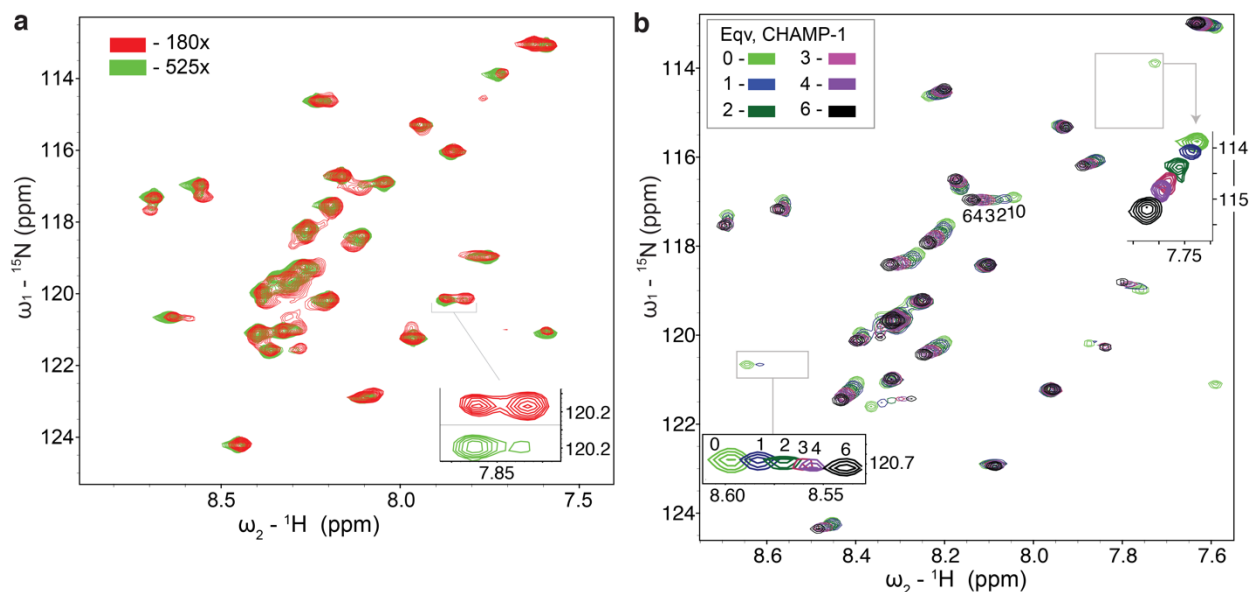
Figure 4.10. ^{15}N mEpoR-TM1 monomer-homodimer behavior in C14-Betaine micelles and binding to CHAMP-1

(a) ^{1}H-^{15}N HSQC spectra of 0.3 μM ^{15}N mEpoR-TM1 when diluted in myristyl-sulfobetaine (C14-B) micelles at 90, 180, and 400 molar equivalents of detergent to peptide (27 mM, yellow; 54 mM, red; 120 mM, blue) in 40 mM sodium acetate buffer pH 5.2, 20 mM NaCl, 0.5 mM EDTA, 5 mM DTT at 45° C and 800 MHz. C14-B ~90 detergent molecules per micelle. Multiple populations are observed which show resonance-dependent peak broadening over the titration, but negligible chemical shift perturbations. **(b)** Maintaining both the same ^{15}N mEpoR-TM1 concentration (0.3 μM) and the total detergent:protein molar ratio (180), 1.35 molar equivalents of unlabeled CHAMP-1 were added and the ^{1}H-^{15}N HSQC spectra is shown overlaid (Blue, 0.4 μM CHAMP, 108 mM C14-B; 360x C14-B per mole mEpoR-TM1). Many new peaks are observed and some existing peaks show large decays in intensity, indicative of mEpoR-TM1's interaction with CHAMP-1 slow on the NMR timescale.

Figure 4.11. ^{15}N mEpoR-TM2 monomer-homodimer behavior in C14-Betaine micelles and binding to CHAMP-1

(a) The ^{1}H-^{15}N HSQC spectra of 0.3 μM ^{15}N mEpoR-TM2 at 45° C, pH 5.2, and 800 MHz when diluted in myristyl-sulfobetaine (C14-B) micelles at 180 and 520 molar equivalents of detergent (54 mM, red; 156 mM, green). The more concentrated sample (180x, red) shows a strong second set of peaks emerging in exchange that is slow on the NMR time scale, likely representing the mEpoR-TM2 in the homodimeric state. *Right inset*, example pair of related, interconverting peaks dependent on the detergent concentration. **(b)** Titration of 1, 2, 3, 4, and 6 molar equivalents (eqv.) of CHAMP (navy, dark green, red, purple, black, respectively) to 0.3 μM ^{15}N mEpoR-TM2 where additional C14-B is added alongside each eqv. CHAMP to maintain a constant molar ratio of C14-B to total peptide of 180:1. Reference spectra (no CHAMP, green) in (**b**) is predominantly the monomeric species at 520:1 ratio C14-B:mEpoR-TM2. Insets track shifting resonances at lower contour.
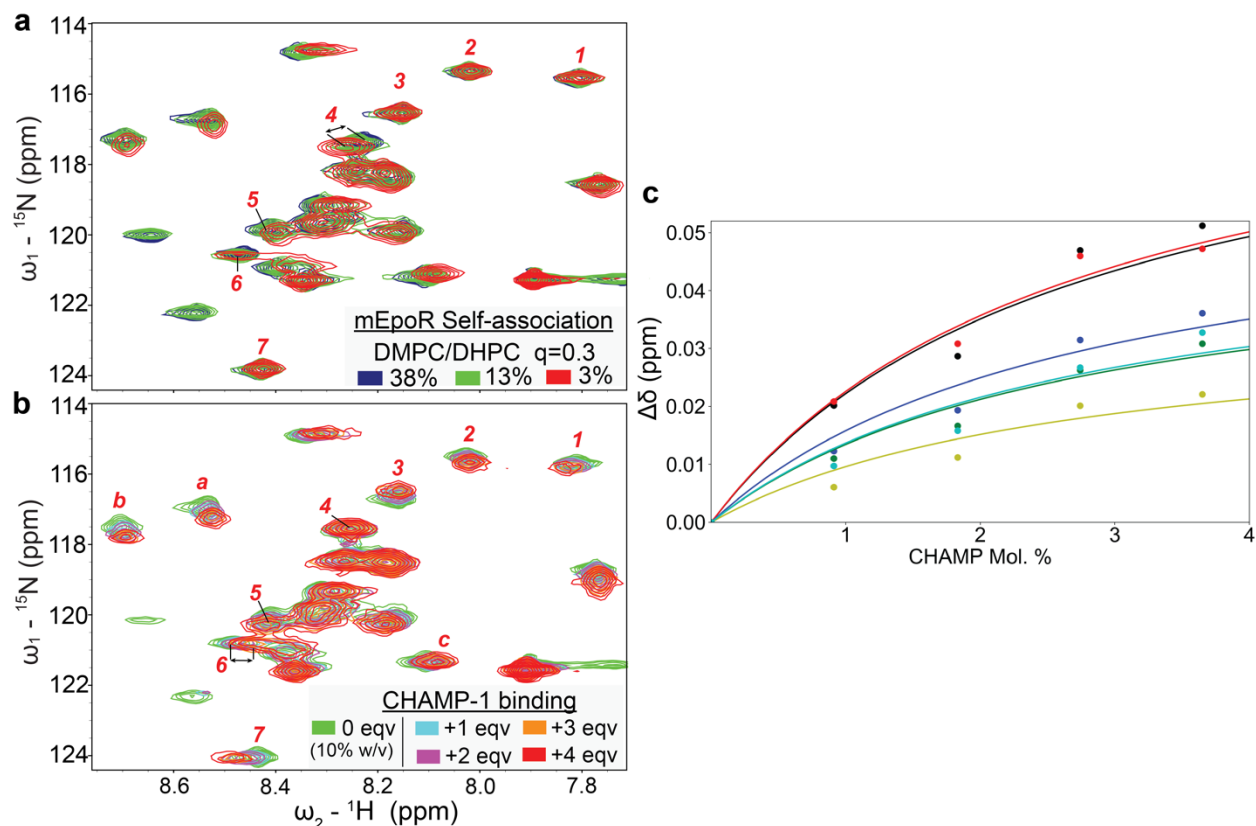
Figure 4.12. mEpoR-TM2 monomer-homodimer equilibrium and binding to CHAMP-1 in DHPC/DMPC q=0.3 bicelles.

^{1}H-^{15}N HQSC spectral changes from 0.5 mM ^{15}N-mEpoR-TM2 at 45° C, pH 5.2, and 800 MHz upon (**a**) titration of DMPC/DMPC q=0.3 bicelles at 38%, 13%, and 3% (w/v) at constant mEpoR-TM2 (navy, green, red, respectively) or (**b**) titration of 1, 2, 3, and 4 molar equivalent (eqv.) unlabeled CHAMP (cyan, orange, maroon, red, respectively) at a constant bicelle concentration of 10%, overlaid on a 10% bicelle (no CHAMP, green) reference spectrum. Both (**a**) the mEpoR homo-oligomer (likely homodimer) and (**b**) mEpoR-CHAMP heterodimer complex exhibit chemical shift perturbation relative to monomer resonances indicating a 'fast' chemical exchange. The numbered resonances denote induced chemical shift changes that differ significantly between the two titrations, suggesting two distinct states (**a**, mEpoR homodimer; **b** mEpoR-CHAMP heterodimer). Peaks 1, 2, 3, 6 and 7 do not shift as mEpoR converts from monomer to homodimer, while these peaks shift significantly upon CHAMP binding; peak 4 shifts upon homodimerization but not upon CHAMP binding; peak 5 shifts in different directions. Lowercase letters denote resonances experiencing peak shift similar between **a** and **b**. (**c**) Global fit to induced chemical shifts of resonances *a, b, c, 1, 2, 5* upon CHAMP titration to calculate a binding affinity (K_d) of 2.7 ± 0.6 mol %, in mol % peptide relative to DMPC.
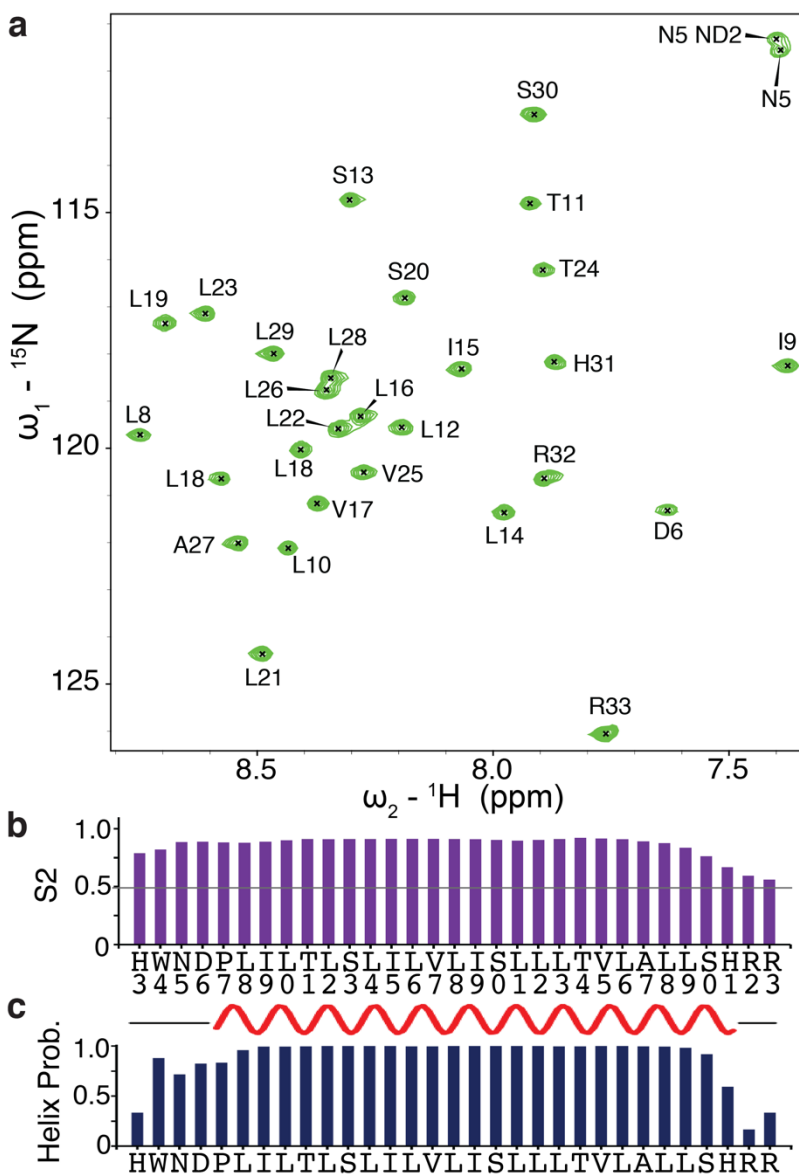
Figure 4.13. ^{1}H-^{15}N HSQC spectra and per-residue structural features of monomeric mEpoR-TM2 in excess DPC.
(a) Assigned ^{1}H-^{15}N HSQC spectra of 200 uM ^{15}N mEpoR at 45° C, 800 MHz, pH 5.2, and high [DPC] : [protein] molar ratio: 600 or 12 micelles per protein.   **(b)** predicted order parameter (S2) from TALOS+.
(c) Predicted helicity from chemical shift index software. *Top*, predicted helical (red) and disordered (black) residues from CSI-3. *Bottom*, probability of helical secondary structure predicted by TALOS+. Prob., probability.
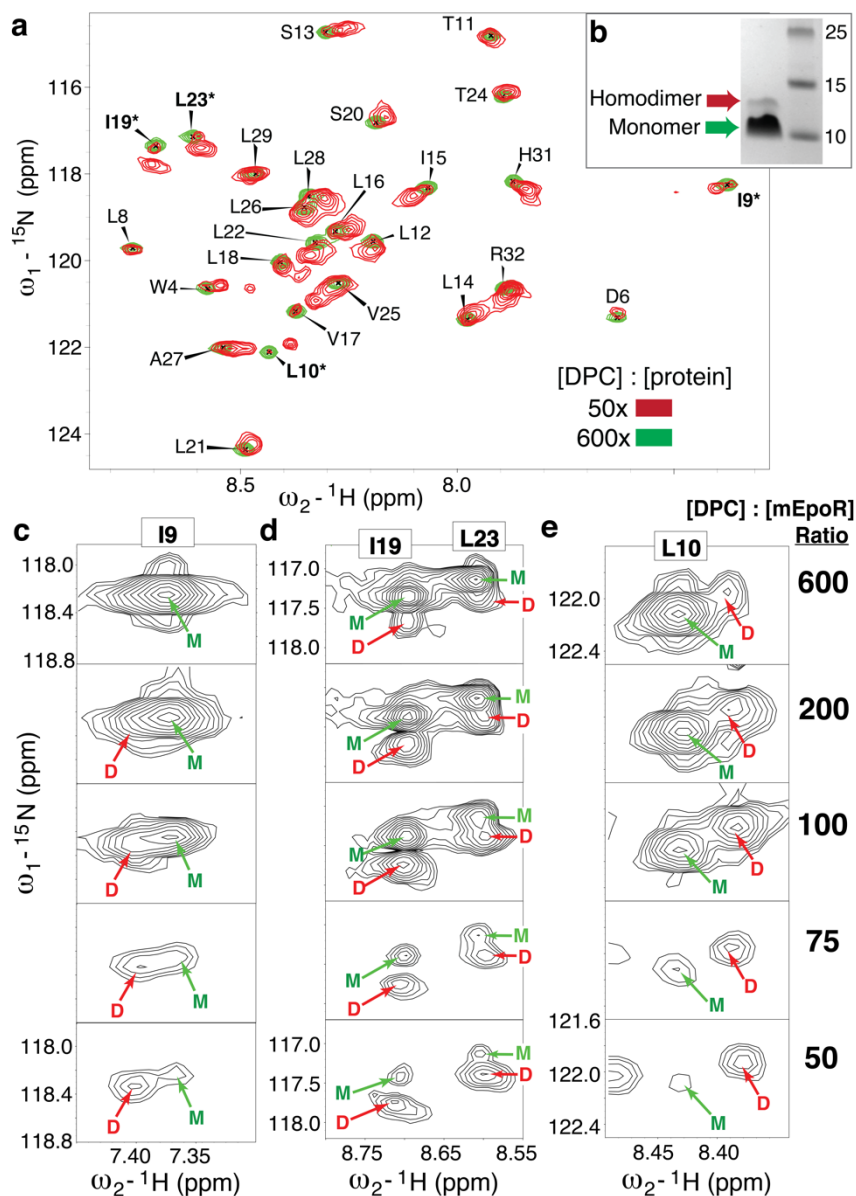
Figure 4.14. mEpoR Monomer-Homodimer equilibrium upon DPC titration
(a). Assigned ^{1}H-^{15}N HSQC spectrum of 150 uM ^{15}N mEpoR at high DPC molar ratio (600x, green) with the spectrum at low DPC concentration (50x, red) overlaid; the latter [DPC] concentration corresponds to approximately 1 TM peptide per 1 micelle. Recorded at 45° C, 800 MHz, pH 5.2, with the predominantly monomer spectra (green) contoured 10-fold higher than the largely homodimer spectra (red). **(b)** SDS-PAGE of mEpoR showing non-covalent monomer-dimer equilibrium. Mon, monomer. **(c-e)** Examples of newly emerging peaks (starred, bolded residues in **a**) of the mEpoR TM homodimer species upon reduction of DPC detergent concentration, titrated at 600, 200, 100, 75, and 50 molar ratios of detergent to TM peptide, showing the mEpoR TM peptide monomer-homodimer equilibria in the absence of CHAMP. Further concentrating mEpoR-TM2 to only 25x DPC molar ratio results in a sparse spectrum suggestive on peptide aggregation, where only 2 peaks (R32 and R33) are observed.
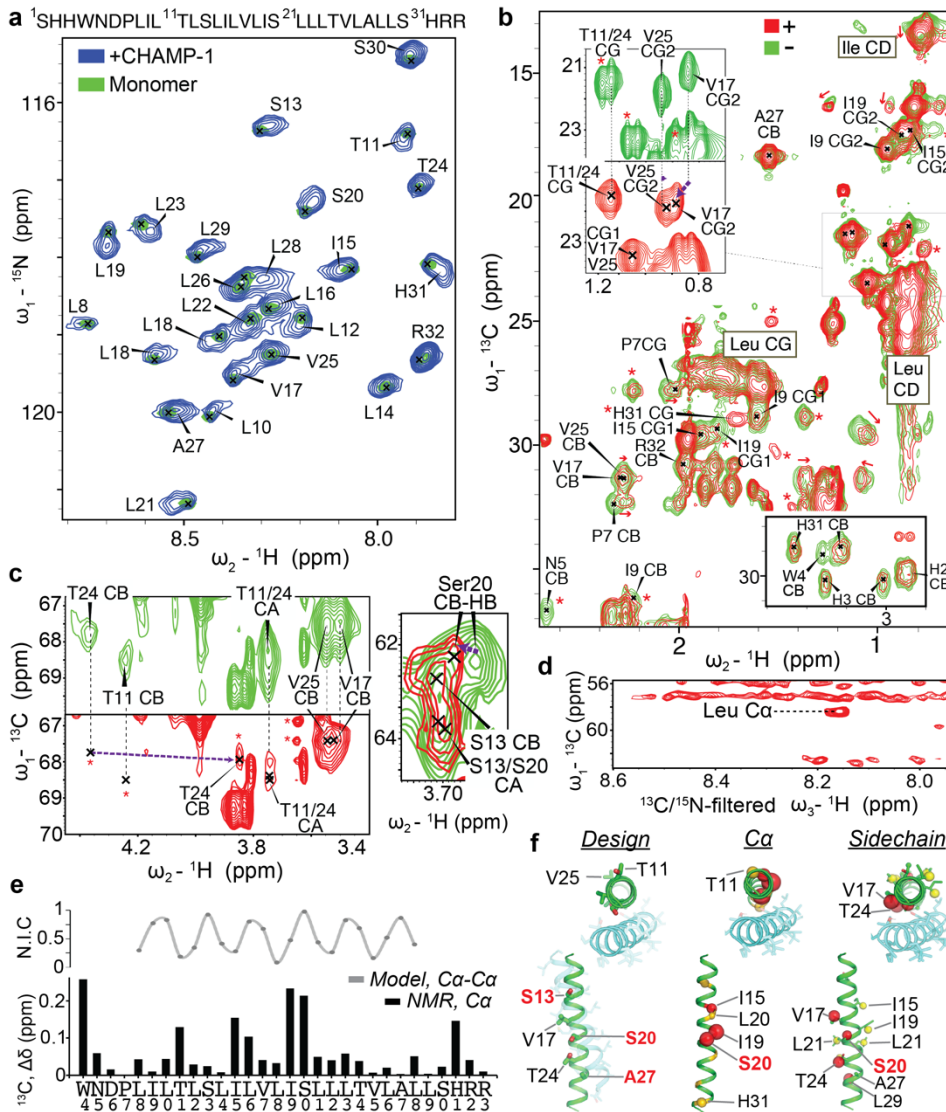
Figure 4.15. Solution NMR of the side-chain mediated CHAMP-1/mEpoR complex in DPC micelles

(a) mEpoR-TM2 sequence and ^{1}H-^{15}N HSQC spectra at 200 µM (U-^{15}N^{13}C^{1}H) in 120 mM ^{2}H-DPC in 40 mM sodium acetate pH 5.2, 20 mM NaCl, 0.5 mM EDTA, 5 mM dithiothritol (45° C, 800 MHz) in monomeric (green; peaks, X's) and CHAMP-1-bound states (blue, 1 mol %, unlabeled) independently assigned. (b) ^{1}H-^{13}C HSQC spectra of mEpoR-TM2 monomer (green) and CHAMP-1-bound (red) states from (a) have widespread differences: chemical shift perturbations (red arrows); new or broadened peaks (red asterisks). *Top inset* (5x contour), V17 CG2-HG2 peak shift. (c) Target epitope residues. *Left*, shift perturbation of T24 CB (purple arrow), new unassigned peaks (red asterisks), broadening of T11 Cα/CB and T24 CA resonances. *Right*, shift perturbation of S20 CB-HB resonance, alongside broadening of S13 Cα, S13 CB, and S20 Cα. (d) 2D F1-^{13}C-edited / F3-^{13}C,^{15}N-filtered HSQC-NOESY spectrum. Transferred NOE cross-peaks indicate direct contact between mEpoR-TM2 ^{13}C atoms, e.g. Leu Cα, and CHAMP-1 ^{14}N/^{12}C-attached protons, e.g. backbone amide proton(s). (e) Helical periodicity in mEpoR-TM2 Cα shift perturbation upon CHAMP-1 binding and its agreement with expected CHAMP-1/mEpoR

199

Cα-Cα distances from the design model, plotted as normalized inter-helical closeness (N.I.C., see methods). **(f)** *Left*, CHAMP-1 (cyan) mEpoR (green) design model noting protein- and lipid-facing sidechains (sticks) and targeted small-X_6-small repeat (red). *Middle*, perturbed Cα atoms (scaled spheres, green-red color scale) lie on one face of mEpoR's TM α-helix, overlapping the targeted epitope and CHAMP-1 interface; minimally perturbed Cα are lipid-facing in the design model. *Right*, perturbed sidechain atoms enriched at one helix face, including V17, S20, T24.
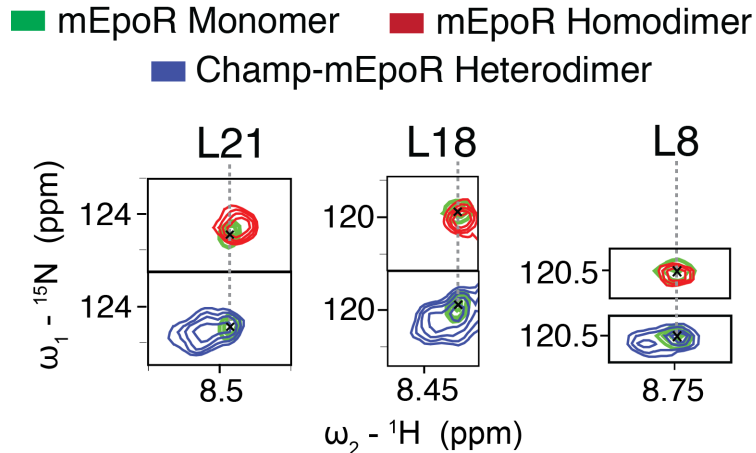
Figure 4.16. Characteristic ^{1}H-^{15}N HSQC peaks of mEpoR in the monomeric, homodimeric, and heterodimer (complex with CHAMP-1) states

^{1}H-^{15}N HSQC spectrum of mEpoR-TM2 in its monomeric state (green, diluted in 600x mol excess DPC) is shown as a reference. Either increasing the mEpoR-TM2 concentration relative to detergent or adding CHAMP-1 peptide induce a second set of peaks corresponding to significant populations of slow-exchanging mEpoR dimeric species; full spectra are shown in Figures Xa and Sx. Respectively, these correspond to the homodimeric assembly (red, more concentrated 50x mol excess DPC, i.e. about 1 micelle per mEpoR-TM2 peptide) and the heterodimeric complex with the CHAMP-1 peptide (blue, 7-fold molar excess CHAMP). Most of the ^{1}H-^{15}N induced chemical shift changes between mEpoR homodimer and heterodimer similar and nearly overlap. However, several notable resonances such as for Leu21 and Leu18 show distinct characteristic peak shifts (i.e. different directions) that distinguish the mEpoR homodimer and heterodimer state. All spectrum at 800 MHz, pH 5.2, and 45° C.
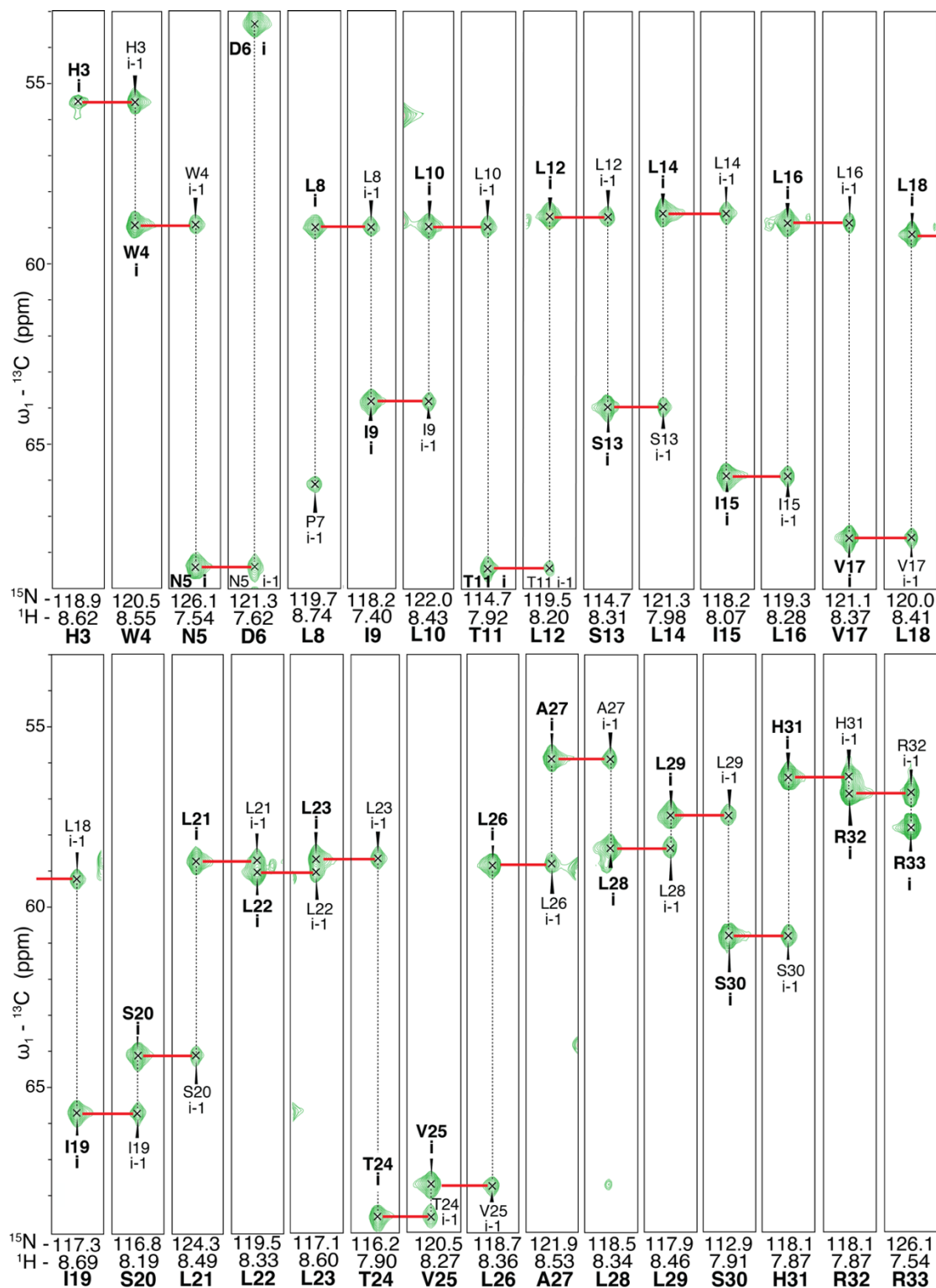
Figure 4.17. HNCA Assignment of mEpoR-TM2 monomer. ^{1}H^{13}C^{15}N mEpoR at 1 mM in 400 mM ^{2}H-DPC at 45° C, 800 MHz, pH 5.2. Red lines denote *i* and *i-1* resonance connectivity; X axis, ^{15}N. Pro 7 assigned by L8 i-1 and lack of ^{15}N-^{1}H i peak or D6 i-1 peaks.
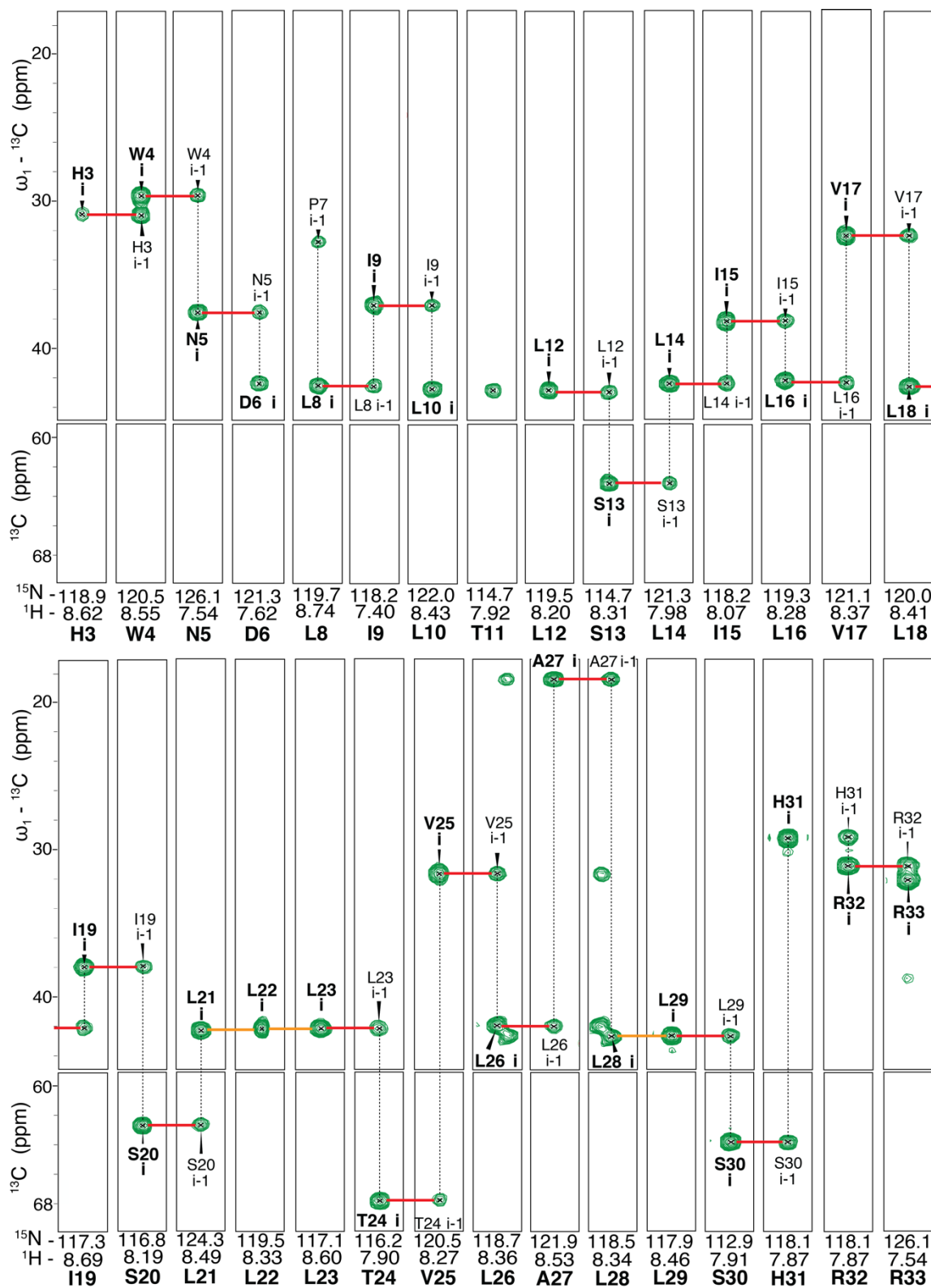
Fig 4.18. HNBA Assignment of mEpoR-TM2 monomer. $^1H^{13}C^{15}N$ mEpoR at 1 mM in 400 mM ^{2}H-DPC at 45° C, 800 MHz, pH 5.2. Orange lines denote ambiguous connectivity due to overlap of *i* and *i-1* resonances. Red lines denote *i* to *i-1* links. X axis, ^{15}N.
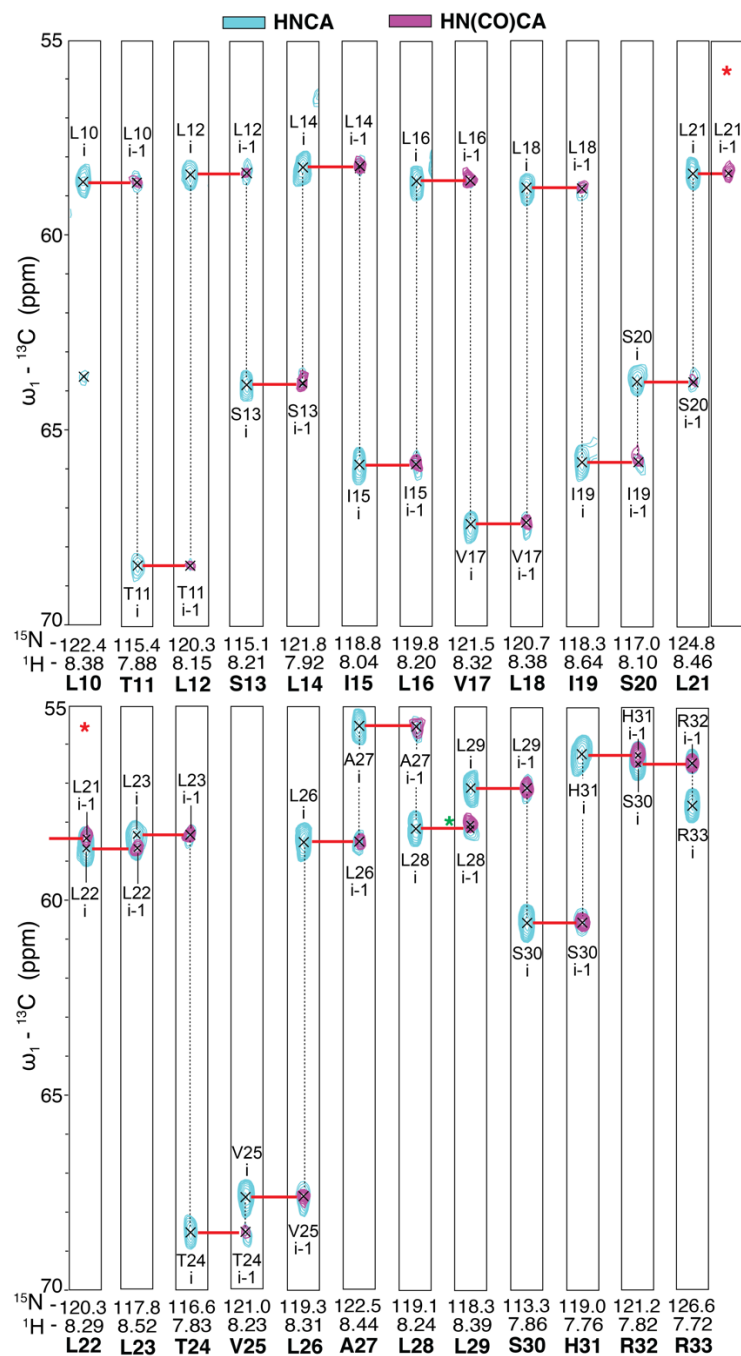
Figure 4.19. Paired HNCA and HNcoCA spectra of $^1H^{13}C^{15}N$ mEpoR in complex with CHAMP-1.
$^1H^{13}C^{15}N$ mEpoR at 2 mM with 10 mM CHAMP-1 in 800 mM ^2H-DPC at 45° C, 900 MHz, pH 5.2; HNCA, cyan; HNcoCA, magenta. Red lines denote *i* and *i-1* resonance connectivity; X axis, ^{15}N. Red asterisk, repeated L22 strip of HCcoCA to show connectivity to L21 across rows of strips. Green asterisk, discrepancy in HNCA and HNcoCA peak position of L28 CA by 0.15 ppm (^{13}C). H31, R32, and R33 resonances contoured 5, 10, 20 times higher, respectively.
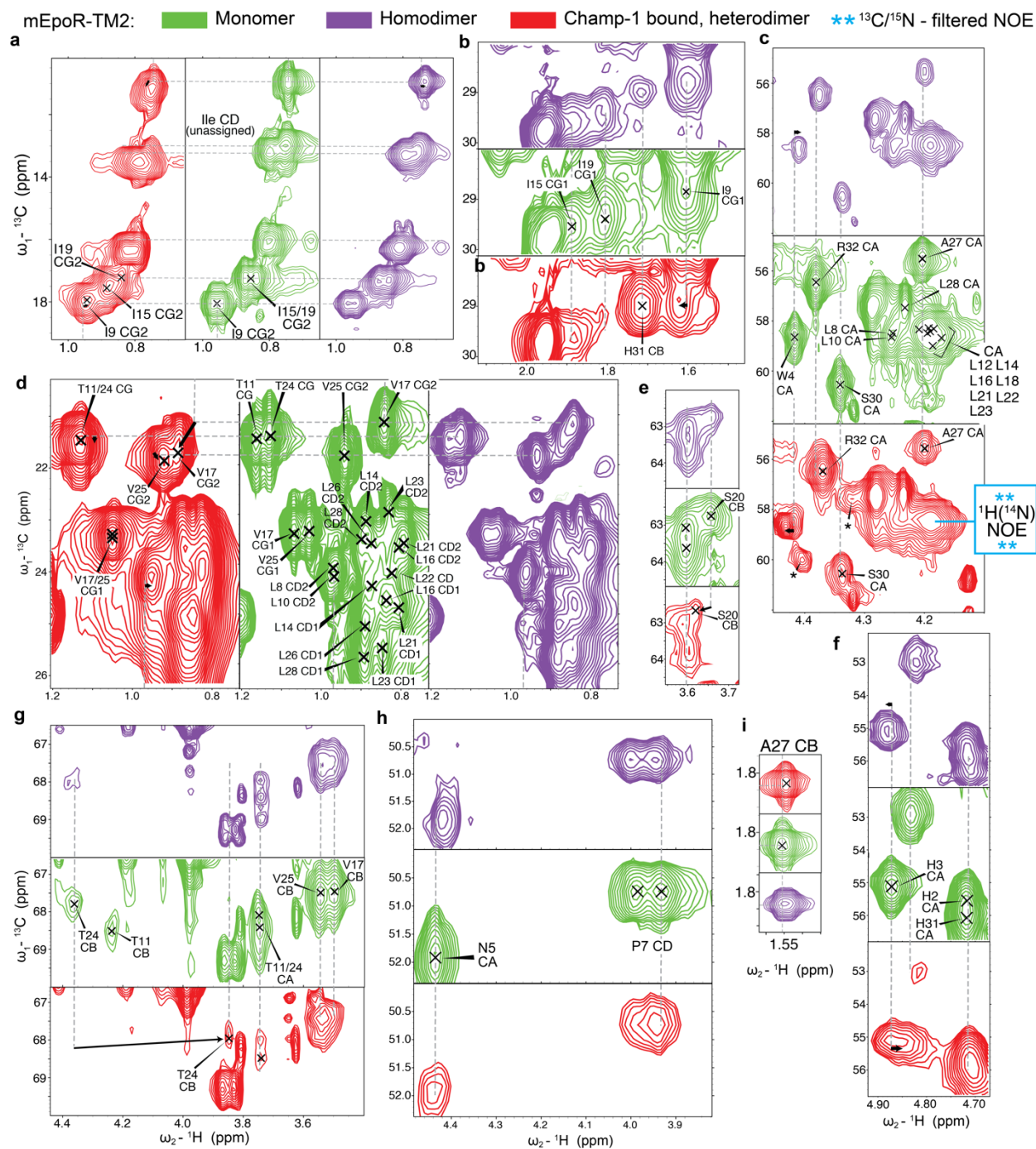
Figure 4.20. Comparison of ¹H-¹³C HSQC spectra of mEpoR-TM2 from monomeric (green), CHAMP-1-bound (red), and partially homodimeric (purple) samples.

(a-i) Different spectral regions perturbed. More monomeric resonances assigned from backbone-backbone and backbone sidechain experiments. Noted heterodimer assignments were derived independently by backbone-backbone, backbone-sidechain, and ¹⁵N-edited NOESY. *Cyan inset*, location of the isolated ¹³C resonance having a transfer NOE to a ¹⁴N-attached amide ¹H peaks in the 13C-edited/¹³C,¹⁵N-filtered HSQC-NOESY (Fig. 4.18).
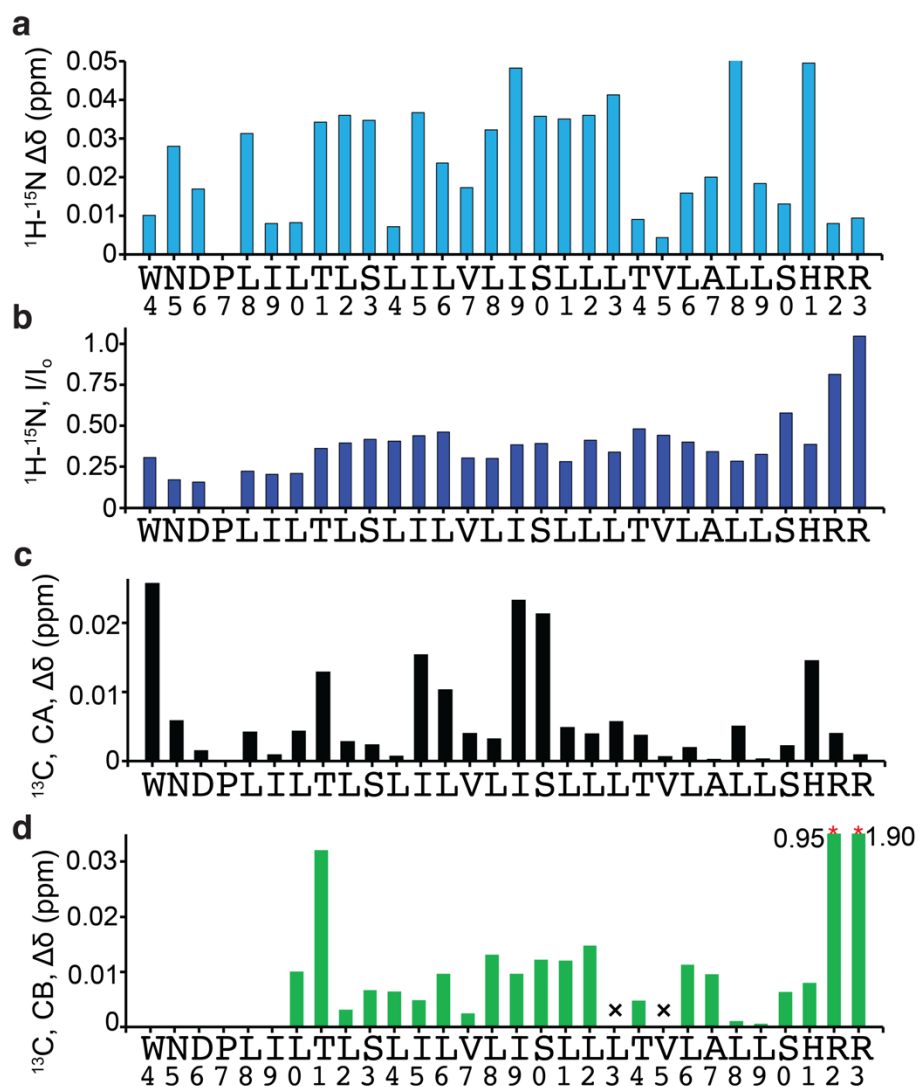
Figure 4.21. Per-residue chemical shift and intensity perturbations of mEpoR CHAMP-1 bound state versus monomeric state.
(a) Hybrid ^{1}H-^{15}N chemical shift perturbation of monomeric and CHAMP-1 bound mEpoR-TM2. **(b)** Relative intensity (I) between ^{1}H-^{15}N resonances in HSQC spectra of monomeric and CHAMP-1 bound mEpoR-TM2. **(c)** CA chemical shift perturbation of monomeric and CHAMP-1 bound mEpoR-TM2. **(d)** CB chemical shift perturbation of monomeric and CHAMP-1 bound mEpoR-TM2. Missing residues notes by 'x'. R32 and R33 are off-scale, and their induced shift values listed on top.

Table 4.1. TM peptide sequences used in this work

Synthesized peptides were produced by solid-phase fmoc synthesis as C-terminal carboxamides and N-terminal free amines. "Biotin nCys mEpoR" was produced with biotin conjugated by peptide bond to the amino terminus. Fluorophore labeled TM peptides for FRET were produced as cysteine-maleimido conjugates. mEpoR-TM constructs for NMR were recombinantly expressed as fusion proteins: mEpoR-TM1 fused to T4 Lysozyme and cleaved by thrombin enzyme; mEpoR-TM2 fused to SUMO and cleaved chemically by Ni^{2+}-induced SNAC self-cleavage.

Thiol-disulfide	Fig 2
nCys mEpoR	NH3–WKK**C**GGGPLILTL**S**LILVLI**S**LLLTVL**A**LLSHRR–NH2
cCys CHAMP-1	NH3–SWGR**S**LFLLLL**S**LLVMLL**G**LLLTIL**G**LLKSGGG**C**–NH2
cCys CHAMP-2	NH3–SWGR**S**LFLLLL**S**LLAALL**G**LLLTIL**G**LLKSGGG**C**–NH2
cCys no Design TM	NH3–SWGRLI**A**LFLVLL**A**LLAGLA**G**LLALIL**A**LKSGGG**C**–NH2
FRET, Cys-maleimido	Fig 3, S3
nCys mEpoR	NH3–WKK**C***GGGPLILTL**S**LILVLI**S**LLLTVL**A**LLSHRR–NH2
cCys CHAMP-1	NH3–SWGR**S**LFLLLL**S**LLVMLL**G**LLLTIL**G**LLKSGGG**C***–NH2
Thiol-disulfide biotin capture	Fig 3, S4
Biotin nCys mEpoR	Biotin–GWK**C**GGGPLILTL**S**LILVLI**S**LLLTVL**A**LLSHRR
cCys CHAMP-1 v2	NH3–KKQNRR**S**LWLLLL**S**LLVMLL**G**LLLTIL**G**LLKSGG**C**–NH2
nCys CHAMP-1	NH3–**C**GGSR**S**LWLLLL**S**LLVMLL**G**LLLTIL**G**LLK–NH2
NMR, construct pair 1	Fig S6
mEpoR-TM1	GSCGQSDNDPLILTLSLILVLISLLLTVLALLSHRR
cCys CHAMP-1	NH3–SWGR**S**LFLLLL**S**LLVMLL**G**LLLTIL**G**LLKSGGG**C***–NH2
NMR, construct pair 1	Fig 5; S7-16
mEpoR-TM2	SHHWNDPLILTLSLILVLISLLLTVLALLSHRR
CHAMP-1 v3	NH3–β–RSLFLLLLSLLVMLLGLLLTILGLLKSGGC–NH2
	β = β-alanine

Table 4.2. GPF complement assay constructs in BaF3 cells.
GFP11 highlighted in yellow and the designed CHAMP-1 TM domain underlined.

---

mEPOR257-10-GFP1-10

MDKLRVPLWPRVGPLCLLLAGAAWAPSPSLYPYDVPDYAPDPKFESKAALLASRGSEELLCFTQRLED
LVCFWEEAASSGMDFNYSFSYQLEGESRKSCSLHQAPTVRGSVRFWCSLPTADTSSFVPLELQVTEA
SGSPRYHRIIHINEVVLLDAPAGLLARRAEEGSHVVLRWLPPPGAPMTTHIRYEVDVSAGNRAGGTQR
VEVLEGRTECVLSNLRGGTRYTFAVRARMAEPSFSGFWSAWSEPASLLTASDLDPLILTLSLILVLISLL
LTVLALLSHRRTLQQKIWGGSGGGGSGGMVSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDAT
IGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGKYK
TRAVVKFEGDTLVNRIELKGTDFKEDGNILGHKLEYNFNSHNVYITADKQKNGIKANFTVRHNVEDGSV
QLADHYQQNTPIGDGPVLLPDNHYLSTQTVLSKDPNEK

---

hEPOR258-10-GFP1-10

MDHLGASLWPQVGSLCLLLAGAAWAPPPNLYPYDVPDYAPDPKFESKAALLAARGPEELLCFTERLE
DLVCFWEEAASAGVGPGNYSFSYQLEDEPWKLCRLHQAPTARGAVRFWCSLPTADTSSFVPLELRVT
AASGAPRYHRVIHINEVVLLDAPVGLVARLADESGHVVLRWLPPPETPMTSHIRYEVDVSAGNGAGSV
QRVEILEGRTECVLSNLRGRTRYTFAVRARMAEPSFGGFWSAWSEPVSLLTPSDLDPLILTLSLILVVIL
VLLTVLALLSHRRALKQKIWGGSGGGGSGGMVSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGD
ATIGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGK
YKTRAVVKFEGDTLVNRIELKGTDFKEDGNILGHKLEYNFNSHNVYITADKQKNGIKANFTVRHNVEDG
SVQLADHYQQNTPIGDGPVLLPDNHYLSTQTVLSKDPNEK

GPF11-CHAMP-1

MDYKDDDDKGGG<mark>RDHMVLHEYVNAAGIT</mark>GGGSLFLLLLSLLVMLLGLLLTILGLLFWGSG

Table 4.3. Fusion protein constructs for NMR

Constructs for isotopically enriched recombinant expression in *E. coli*. Final TM constructs after thrombin protease cleavage or sequence-specific nickel-assisted cleavage (SNAC) used for NMR are underlined.

7x His T4 Lysozyme (thrombin) mEpoR TM-1
MGSSHHHHHHHGGSGRGSHMGNIFEMLRIDEGLRLKIYKDTEGYYTIGIGHLLTKSPSL NAAKSELDKAIGRNTNGVITKDEAEKLFNQDVDAAVRGILRNAKLKPVYDSLDAVRRAALINMVFQM GETGVAGFTNSLRMLQQKRWDEAAVNLAKSRWYNQTPNRAKRVITTFRTGTWDAYAAGGSGSTE NGLVPRGSCGQSDNDPLILTLSLILVLISLLLTVLALLSHRR

7x His SUMO (SNAC) mEpoR TM-2
MHHHHHHGSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRL MEAFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHREQIGGAGG SHHWNDPLILTLSLILVLISLLLTVLALLSHRR

209

Publishing Agreement

It is the policy of the University to encourage open access and broad distribution of all theses, dissertations, and manuscripts. The Graduate Division will facilitate the distribution of UCSF theses, dissertations, and manuscripts to the UCSF Library for open access and distribution. UCSF will make such theses, dissertations, and manuscripts accessible to the public and will take reasonable steps to preserve these works in perpetuity.

I hereby grant the non-exclusive, perpetual right to The Regents of the University of California to reproduce, publicly display, distribute, preserve, and publish copies of my thesis, dissertation, or manuscript in any form or media, now existing or later derived, including access online for teaching, research, and public service purposes.

DocuSigned by:

Marco Mravic

D55723CC8EDD4C3...   Author Signature

9/4/2020

Date