High-throughput discovery of transmembrane helix dimers from human single-pass membrane proteins with TOXGREEN sort-seq

Samantha M. Anderson*[1,2], Joshua Choi*[1,2], Emma M. Cushman[1], Megan Leander[1,2], Srivatsan Raman[1,3,4] and Alessandro Senes[1,@]

[1]Department of Biochemistry

[2]Integrated Program in Biochemistry

[3]Department of Bacteriology

[4]Department of Chemical & Biological Engineering

University of Wisconsin-Madison, Madison, WI 53706

*These authors contributed equally

[@]Corresponding author: senes@wisc.edu

ORCID:

SMA: 0000-0001-6114-0427

JC: 0009-0002-3095-9987

EMC: 0009-0003-9415-4608

ML: 0000-0001-5257-2498

SR: 0000-0002-5673-6944

AS: 0000-0002-3807-2275

Abbreviations

AF3: AlphaFold 3

BSC: backscatter

BTNL3: Butyrophilin-like protein 3

CAT: Chloramphenicol AcetylTransferase

EVA1B: Protein eva-1 homolog B

FACS: Fluorescence-Activated Cell Sorting

FSC: forward scatter

FSC-A: forward scatter area

FSC-H: forward scatter height

GFP: Green Fluorescent Protein

GpA: glycophorin A

HAVCR2: Hepatitis A virus cellular receptor 2

IL-22R1: Interleukin-22 receptor subunit alpha-1

MBP: Maltose Binding Protein

NGS: Next-Generation Sequencing

RFU: Relative Fluorescence Units

SEMA: semaphorin

SSC-A: side scatter area

TfR1: Transferrin receptor protein 1

TM: transmembrane

TMEM81: Transmembrane protein 81

TSR: thrombospondin type-1 repeats

WT: wild-type

Abstract

The oligomerization of the transmembrane (TM) helices of single-pass membrane proteins is crucial to numerous biological functions and its misregulation can lead to cancer among many other diseases. The study of TM helix oligomerization is facilitated by the availability of genetic reporter assays, which have become essential tools for understanding the organization and biology of single-pass systems. In particular, reporter assays are crucial for mapping the oligomerization interfaces of TM helices through scanning mutagenesis. Their application, however, has thus far been limited by the need to clone and measure each construct individually. Here, we present "TOXGREEN sort-seq", a high-throughput version of the TOXGREEN assay that enables the direct measurement of TM helix oligomerization in large libraries using a combination of fluorescence-activated cell sorting and next-generation sequencing. We show that TOXGREEN sort-seq is robust, as it reproduces the direct measurements of individual constructs with good accuracy and sensitivity. The method produced high-quality mutational profiles from a library of 17,400 constructs designed to probe the interface of 100 potential GAS_{right} dimers predicted from the sequence of human single-pass membrane proteins. We report the validated structural model of twelve transmembrane helix dimers involved in a variety of biological functions, including immune response (interleukin-22 receptor subunit alpha-1, butyrophilin-like protein 3, hepatitis A virus cellular receptor 2), transport (transferrin receptor protein 1), and cell-surface signaling and proliferation (syndecan-3; semaphorins 5A, 6B and 6D). Remarkably, all three semaphorins in the dataset formed strong dimers and produced mutational profiles consistent with the computational structure. These findings open the possibility that GAS_{right}-mediated dimerization may be relevant to these proteins' activity and provide a validated interface for assessing their biological role.

Introduction

Proteins containing a single transmembrane (TM) helix are the most common type of membrane protein, accounting for over 20% of the total in humans and most organisms[1]. The TM helices of these "single-pass" membrane proteins often play active roles in biological function through the formation of oligomeric complexes[2]. One example is the large number of single-pass receptors whose activity often occurs through ligand-induced stabilization of dimeric conformations[3–5]. Several biophysical techniques have been developed to study the oligomerization of TM helices *in vitro*, using physical methods such as Förster Resonance Energy Transfer[6–13], sedimentation equilibrium analytical ultracentrifugation[14,15], disulfide exchange equilibrium[16,17] and steric trapping[18,19]. Genetic reporter systems are a popular complementary approach to measure the oligomerization propensity of TM helices *in vivo* in bacteria, by linking TM association to gene expression. While less quantitative than the biophysical methods, genetic assays enable faster experimentation since they do not require over-expression, purification, reconstitution, or labeling of proteins. They also have the advantage of measuring TM helix interactions in a natural membrane.

The first genetic reporter assay for measuring TM helix oligomerization was developed by Langosh and colleagues[20]. It was based on the expression in *Escherichia coli* of a fusion protein containing the TM helix of interest connected to the intracellular domain of ToxR, a membrane-bound transcriptional activator from *Vibrio cholera*[21]. In this system, TM helix oligomerization leads to the quantifiable expression of β-galactosidase. Derivatives of the system include versions suitable for characterizing hetero-oligomeric and antiparallel TM helix interactions[22–24] and a variant in which ToxR can act as either an activator or a repressor by binding to alternative promoter sequences[25,26]. Another ToxR-based system, TOXCAT[27], expresses an antibiotic resistance gene, enabling the selection of strongly oligomerizing TM helices in large libraries[28–30]. Other variants include systems in which the reporter gene was replaced by luciferase[31] or Green Fluorescent Protein (GFP)[32] for improved detection, along with a dominant-negative ToxR system for detecting heterologous association[33]. Other genetic assays use different transcriptional regulators, including GALLEX, which

utilizes both a wild-type (WT) and mutated DNA-binding domain of the LexA transcriptional repressor to report heterologous association[34], and AraTM, which promotes Type I orientation and allows for the assessment of both transmembrane and juxtamembrane regions in oligomerization[35,36].

These genetic reporter assays are particularly useful for scanning mutagenesis, which is a practical and effective strategy for experimentally mapping the interaction interface of TM helical dimers. Mutagenesis data can validate or inform the development of accurate computational models, which are particularly useful given that single-pass proteins remain severely underrepresented in the structural database[37]. Engelman and coworkers pioneered this strategy using SDS-PAGE as the experimental method for measuring oligomerization. The pattern of disruption in SDS-resistant self-association identified the association interface and supported the computation of early models of the glycophorin A (GpA) dimer[38,39] and the phospholamban pentamer[40,41]. The development of genetic reporter assays has since greatly streamlined the scanning mutagenesis strategy and it also expanded its applicability to dimers of moderate stability, which are unstable in SDS detergent. These genetic assays also enhanced the phenotype of certain mutants, particularly those of mildly polar residues (G, S, T and Y) that are common in membrane proteins and tolerated in a membrane environment[27] but can be disruptive irrespective of their position in helices solubilized in a detergent[38,40].

Just as genetic assays made it easier to map the interface of individual TM oligomers through scanning mutagenesis, the development of high-throughput methods opens the opportunity to extend scanning mutagenesis to large libraries of helices, enabling their simultaneous analysis with comparable effort. TOXCAT is already suited for analyzing large sequence libraries, although indirectly through the use of selection. The expression of the Chloramphenicol AcetylTransferase (CAT) as a reporter gene can be inferred through survival in media containing the antibiotic chloramphenicol, coupled with Next-Generation Sequencing (NGS) to determine the constructs that survive at increasing antibiotic concentrations. Using this strategy, Fleishman and colleagues

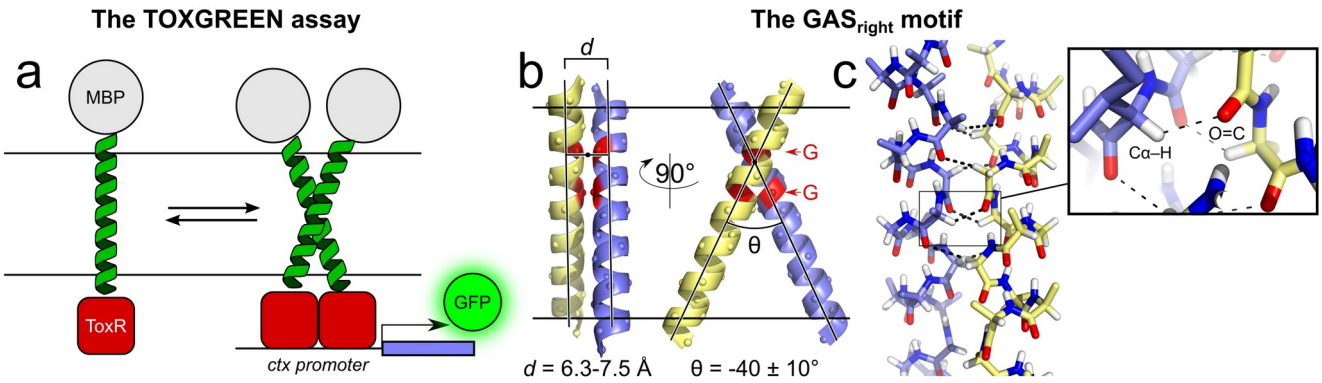

Fig. 1. a) The TOXGREEN assay is a derivative of TOXCAT, a widely used genetic assay to study interactions of transmembrane helices within the inner membrane of the bacterium *Escherichia coli*. The assay is based on a fusion construct that links a transmembrane domain of interest with a cytoplasmic DNA-binding domain from the Vibrio cholerae ToxR protein. Interaction driven by the transmembrane domain results in dimerization of the ToxR domain, which, in turn, activates the expression of the reporter gene superfolded GFP. The resulting fluorescence is used as a measure of the ability of the transmembrane domain to self-associate.

performed full mutational scanning of three single-pass model systems, exploring the energetics of membrane insertion and dimerization[30].

Although CAT resistance provides an indirect measure of oligomerization propensity, it would be useful to precisely measure reporter gene expression in high-throughput. Here we present TOXGREEN sort-seq, a high-throughput method that measures reporter gene expression directly. The method is based on the GFP-expressing variant of TOXCAT, TOXGREEN[32]. In traditional TOXGREEN, GFP expression is measured in a fluorescence microplate reader for each construct individually (Fig. 1a). TOXGREEN sort-seq significantly improves the throughput by using Fluorescence-Activated Cell Sorting (FACS) to read all individual cells in a large library and sort them into bins based on their fluorescence. NGS then quantifies the occurrence of each construct within these bins, and the fluorescence of each individual construct in the library is reconstructed using statistical methods.

We applied TOXGREEN sort-seq to validate the association interface of 100 computationally predicted human TM helix dimers in high-throughput using scanning mutagenesis. The opportunity for a large-scale validation of structural models of TM dimers was provided by a large dataset of potential TM dimers that were previously predicted from the sequences of all single-pass proteins in the human proteome using the program CATM[42,43]. These dimers were predicted to associate through the

formation of GAS_{right} structures, a common dimerization motif for single-pass proteins[44,45] illustrated in Fig. 1b. GAS_{right} dimers are characterized by a right-handed interaction angle and signature GxxxG and GxxxG-like sequence motifs (GxxxG, GxxxA, SxxxG, etc.)[46–48]. Their helical backbones are in close contact and the interactions are stabilized by van der Waals packing and networks of inter-helical weak hydrogen bonds between Cα–H carbon donors and carbonyl acceptors (Cα–H···O=C) (Fig. 1c)[42,43]. The strength of association of GAS_{right} dimers is modulated by its geometry and sequence[42], forming both strong structural dimers[49,50] as well as weak dynamic systems[4,51–56]. The dataset contains a range of dimers with different estimated energy scores, allowing for testing the scanning mutagenesis approach across a range of association strengths. Further, the selected proteins include many that are biologically and biomedically relevant, but are not yet known to form dimers or if they dimerize specifically through their TM helices.

Our analysis demonstrates that the statistically reconstructed fluorescence signals obtained with TOXGREEN sort-seq reproduce the direct fluorescence measurements of individual constructs with good accuracy and sensitivity. We show that the method is effective for measuring thousands of constructs consisting of WT sequences and point mutations simultaneously, generating high-quality, reproducible mutational profiles. Finally, we present the structural model of twelve GAS_{right} dimers of previously unknown structure that have been validated by our mutational analysis. These proteins participate in a variety of important processes, including transport, the modulation of immune response and cell-surface signaling, proliferation and adhesion. In particular, we discover that the TM helix of three different semaphorins is a strong dimerization domain. Semaphorins are a family of cellular surface proteins that mediate cell-to-cell communication and are critical for regulating cell morphology, cell differentiation and cell-cell interactions both in neuronal and non-neuronal cell types. They are known to dimerize through interactions in their extracellular domains. Our data indicates that the TM domain of these semaphorins is an additional dimerization domain and suggests it may be important for the assembly and signaling of these proteins.

Results and Discussion

The TOXGREEN sort-seq method

The TOXGREEN sort-seq method is schematically illustrated in Fig. 2. The experiment starts with a synthetic DNA oligo pool library in which each oligonucleotide encodes a unique TM helix sequence (although the method is applicable to other approaches for generating libraries, such as randomized mutagenesis). The pool of DNA inserts is cloned into the standard TOXGREEEN plasmid (pccGFPKan) and transformed into MM39 *E. coli* cells, resulting in a cell library expressing each oligo pool sequence. The cells are then sorted into multiple bins using FACS based on the expression

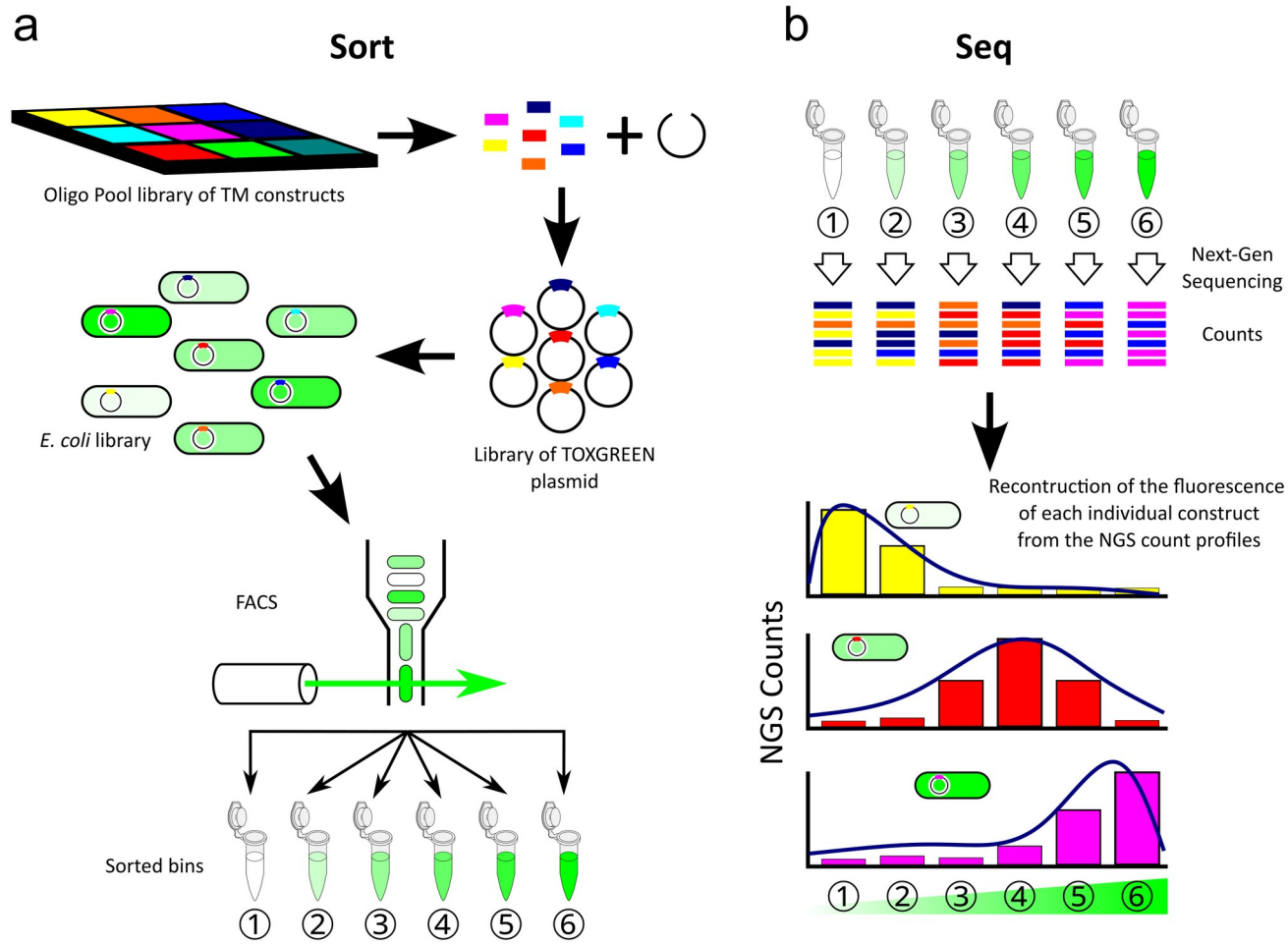

Fig. 2. Schematic representation of the TOXGREEN sort-seq method. a) The procedure starts with a library of oligonucleotides encoding the TM helices of interest. These are cloned into a TOXGREEN plasmid and transformed into *E. coli*. The cells are sorted into bins based on their GFP expression using FACS. b) The population of each construct in the bins is determined using Next-Generation Sequencing. The count distribution of each construct in each bin is used to reconstruct the fluorescence of each individual construct.

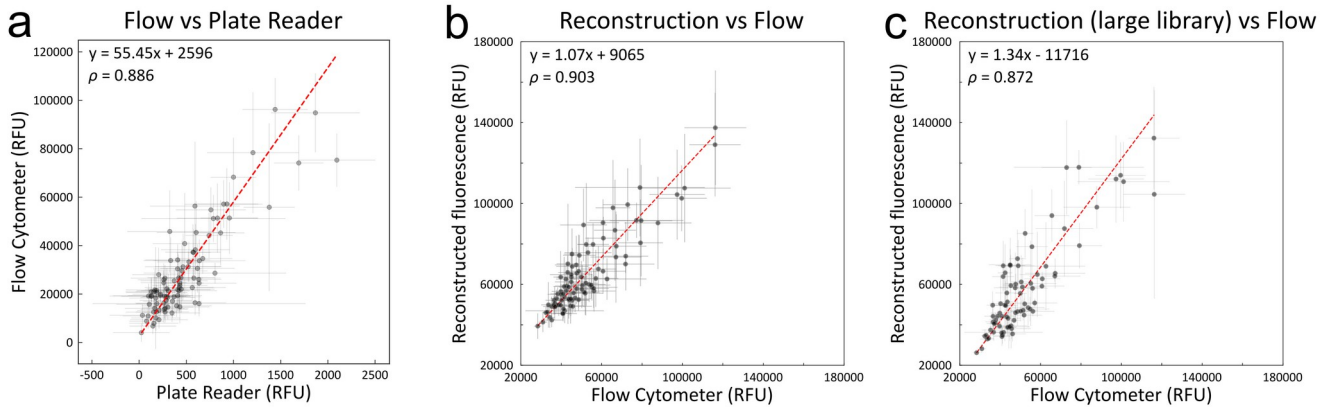

Fig. 3. Validation of the TOXGREEN sort-seq method. Validation of the TOXGREEN sort-seq method. A small library of 100 pre-characterized constructs (the validation set) was created for testing the method (Table S1). a) Comparison between direct measurement of GFP fluorescence in a plate reader vs direct measurement of individual constructs in a flow cytometer. b) Direct measurement in flow cytometer vs statistical reconstruction using sort-seq. c) Same as panel b, except that the validation library was mixed with a large library of over 17,000 unrelated constructs. The observed linear relationships establishes that TOXGREEN sort-seq can reconstruct the fluorescence of individual constructs with good accuracy and sensitivity in large pools of sequences. RFU = Relative Fluorescence Units. ρ = Pearson correlation coefficient. Linear fits performed using Deming regression. Error bars represent standard deviation of three replicates for the plate reader values and at least five replicates for the flow cytometer values and for the reconstructed fluorescence values.

levels of GFP. The DNA for each of these sorted populations is then extracted and analyzed by NGS to obtain a count for each library construct within each bin. As described in Methods, the distribution of the counts of each construct in the bins is then used to mathematically reconstruct their estimated GFP fluorescence, using a statistical inference approach based on a weighted average adapted from Kosuri et al.[57].

High-throughput data accurately recapitulates individual measurements

To assess the performance of TOXGREEN sort-seq, we performed three comparisons of directly measured fluorescence vs its reconstructed values across conditions of increased complexity. For this step, we used a small library of 100 TM helices (called the validation library, supplementary Table S1) that had been individually characterized and displayed varying degrees of GFP expression.

First, to establish the feasibility of using a cell sorter for measuring GFP expression in our systems, we compared the median fluorescence distribution of individual cells measured through flow cytometry with the fluorescence measured traditionally in bulk using a plate reader. As illustrated in

Fig. 3a, the comparison resulted in a strong linear correlation between the two instrument measurements (Deming regression, ρ = 0.886) with standard deviations (error bars) of similar relative magnitude. The experiment demonstrates that a fluorescence-activated cell sorter has the dynamic range and sensitivity to accurately measure the GFP fluorescence of TOXGREEN constructs.

Next, we assessed TOXGREEN sort-seq's statistical approach by comparing the fluorescence reconstructed from NGS data with the median fluorescence of each construct directly measured in a cell sorter. We first reconstructed the GFP fluorescence on a small scale by sorting a library containing exclusively the 100 members of the validation library (Fig. 3b). We then proceeded to reconstruct the fluorescence of the validation library when these were mixed with a large library of 17,305 additional sequences (Fig. 3c). Both measurements of the validation set resulted in excellent linear relationships between reconstructed and measured fluorescence (ρ = 0.903 and 0.872, respectively). These experiments establish that the TOXGREEN sort-seq method can reliably evaluate the association of TM helices on a massive scale.

High-throughput evaluation of proper membrane insertion (maltose growth assay)

To assess whether a TOXGREEN construct is expressed in the membrane and inserted with the correct topology, it is standard practice to perform a separate MalE complementation assay[27]. The MM39 *E. coli* strain used in TOXGREEN is Maltose Binding Protein (MBP)-deficient; it can grow in a M9 minimal media where maltose is the only available carbon source only if the chimeric TOXGREEN constructs are expressed and inserted into the inner membrane with the C-terminal MBP moiety oriented to the periplasmic side. Canonically, this assay is performed using M9 maltose selection plates with individually streaked samples, but this approach is not suitable for large libraries. For compatibility, we scaled up the assay by implementing a liquid culture variant where the entirety of the library is grown in liquid M9 maltose minimal media. The make-up of the surviving library members is identified at time points via next-generation sequencing until the non-inserting constructs are out-competed from the population due to poor uptake of the maltose, leading to a massive decrease in enrichment score.

To assess the feasibility, we compared the growth of the constructs in the validation library with the traditional solid M9 media and the new liquid media protocol. We found the same samples that failed to grow on the selection plates (supplementary Fig. S1a) also displayed greatly diminished presence in the population after a period of extended growth in liquid M9 maltose minimal media (supplementary Fig. S1b and c), suggesting combined library growth in liquid media recapitulates the original test. However, we found the necessary growth time and enrichment score to somewhat vary between library samples. Therefore, for calibration with different libraries and populations, we suggest the inclusion of a set of standard non-inserting constructs (such as 1E01, 2E11, 2H11, 2F12 and 1E11, listed in supplementary Table S1) to identify the most appropriate threshold enrichment score. Any samples with similar growth levels to these negative controls in M9 minimal media should be considered as likely poorly expressing or inserting.

High-throughput mutagenesis study with TOXGREEN sort-seq

We applied TOXGREEN sort-seq to a large-scale scanning mutagenesis analysis aiming to validate the helix-helix interaction interfaces of 100 computationally predicted GAS_{right} TM dimers[42,43]. These potential dimers were selected from a list of 1,141 GAS_{right} interfaces that we previously predicted from all single-pass proteins annotated in the human genome[42]. The 100 TM sequences were selected using the following criteria: i) they had a predicted TM helix of 21 amino acids (the most common length annotated in Uniprot, for consistency); ii) they did not contain more than one strongly polar residue (D, E, H, K, N, Q); iii) they did not contain a proline residue; and iv) their model did not display significant inter-helical interactions in the juxtamembrane polar region. From the remaining sequences, we selected the top 100 (listed in supplementary Table S2) for the mutational analysis based on their CATM energy score.

A wide variety of mutational strategies have been applied in the literature to oligomerizing systems TM helices. Some studies have used very large sets of mutations, up to thirteen[38] and even seventeen[58] changes at each position. A TOXCAT study identified the dimerization interface of BNIP3 using eight mutations per position[59]. Our group has typically substituted seven hydrophobic amino

acids at all positions (A, I, F, G, L, M and V)[60,61], but others have used more minimalistic approaches. For example, TOXCAT was used with a combination of alanine- and cysteine-scanning (the latter in combination with cross-linking) to identify the interfaces for the thrombopoietin receptor TM[62]. Another study used either Ala or Ile at each position (depending on size and polarity) to identify the interface of the human B-cell receptor in GALLEX[63]. Others have used scanning mutagenesis with a single amino acid, such as Ala[64] or Phe[65]. Aiming for an optimally sized library, we decided a middle-ground approach, introducing three mutations at each position, namely to Ala, Ile or Leu. If the WT identity at a position consisted of one of these three amino acids, we introduced an additional Phe mutation to maintain the total of three. As explained below, we also explored the effect of TM insert length in scanning mutagenesis. For this reason, each of the 100 constructs was assessed at its original 21 amino acid length (called the "21-AA" library) and with truncations to 19 ("19-AA") and 17 amino acids ("17-AA"). The final combination of the number of constructs, the helix sizes and the mutations yields a theoretical library of 17,400 constructs, consisting of 300 WT sequences (100 constructs each tested at three lengths) and 17,100 point mutants.

Comparison of TOXGREEN signal as a function of construct length

The length and hydrophobicity of the TM helix insert can influence the reporter output in TOXCAT/TOXGREEN[66], which is critical for sensitivity in scanning mutagenesis. To systematically evaluate how the length of the TM domain affects this outcome, we compared the reconstructed fluorescence output of the same constructs (including WT sequences and all point mutants) in the three libraries. We found that the longer 21-AA constructs displayed significantly higher reconstructed fluorescence (78,000 ± 32,000 RFU) compared to the 19-AA (58,000±22,000 RFU) and 17-AA libraries (52,000±26,000 RFU) (Fig. S2a). Linear regression analysis of the samples across the three lengths (Fig. S3) shows significant proportionality in the fluorescence, indicating that overall there is still correspondence in the relative TOXGREEN output of a sequence at varying lengths. It should be noted, however, that many of the constructs that display moderate to high signals in the 21-AA library are near the baseline when shortened to 17 residues (Fig. S3c). This data suggests that for improved

Outcomes	Cases
Confirmed GAS$_{right}$ interface	12
Disruption maps to alternative interface	13
Uninterpretable disruption	6
Sensitive single polar amino acid	3
No or minimal disruption	31
TOXGREEN signal too low for mutagenesis	35
Total	**100**

Fig. 4. Examples of mutagenesis outcomes. The mutagenesis analysis of 65 constructs produced outcomes that can be classified into 5 groups. a) Twelve constructs displayed a profile that is consistent with the predicted interface of their GAS$_{right}$ model. b) Thirteen constructs displayed a pattern of disruption that maps on one face of the helix but does not match the predicted GAS$_{right}$ model. c) Six constructs had a pattern of disruption that does not map to a clear interface. d) Three constructs displayed a pattern consisting of a single polar amino acid as the only critical position for oligomerization. e) 31 constructs displayed no disruption or a minimal level of disruption across all positions in all the available mutagenesis profiles at various lengths.

fluorescence output, the 21-amino acid TM helix is preferable. However, as noted below, the analysis of the mutagenesis profiles indicates that the longer length may not always be the optimal choice.

A summary of the mutagenesis analysis is provided in Table S3. The WT TOXGREEN signal of 35 constructs had a reconstructed fluorescence value close to the monomeric baseline in all three lengths and thus these constructs were excluded from the mutagenesis analysis. The mutagenesis profiles of the 65 remaining constructs are reported in supplementary Fig. S4, where phenotypes are classified according to five levels of disruptions, ranging from WT-like (white) to completely disruptive (red). For each position, we calculated a position-specific “disruption score” by averaging the disruption levels of all mutants of that residue. As illustrated in Fig. 4 we identified five types of profiles. Twelve constructs displayed a profile that is consistent with the predicted interface of their

GAS_{right} model (Fig. 4a). A comparable number (13) displayed a pattern of disruption that maps on one face of the helix but does not match the predicted GAS_{right} model (Fig. 4b), suggesting that these helices may possibly oligomerize through a different structure. Interestingly, three of these profiles match reasonably well with their AlphaFold 3 (AF3) prediction, suggesting that they are likely to highlight biological interfaces (P78380: Oxidized low-density lipoprotein receptor 1, P16410: Cytotoxic T-lymphocyte protein 4, and Q9ULG6: Cell cycle progression protein 1, illustrated in supplementary Fig. S5). Six constructs had a pattern of disruption that does not map to a clear interface (Fig. 4c). Three constructs displayed a single polar amino acid as the only stringent requirement for oligomerization (Fig. 4d), consistent with the tendency of these amino acids to drive promiscuous association of TM helices[15,67]. Surprisingly, the remaining 31 constructs displayed a minimal level of disruption across all positions (Fig. 4e).

When the mutational profiles at different TM lengths are compared, there is an overall notable tendency for mutations to have a stronger effect in the shorter constructs, which is useful for identifying critical positions. The mutational profile of the 21-AA constructs displayed minimal or no disruption in 60% of the cases, whereas it occurs only about 13% and 14% of the 19-AA and 17-AA constructs, respectively. This is evident in the examples in Fig. 4a and b, in which the mutations of the 21-AA constructs cause minimal disruption but the 19-AA and especially 17-AA mutagenesis reveals a clear pattern. This is also observed for constructs O75056, Q13591, Q6UXE8, Q8IYV9, Q9H3T3, Q9NVM1, Q9ULG6 in the profiles of all proteins, provided in supplementary Fig. S4. In some cases, however, the mutational profiles are essentially identical across the three lengths. Interestingly, this is not necessarily dependent on the reconstructed GFP fluorescence of the WT constructs, which can be very similar for the three lengths or differ wildly (e.g., Q8N6P7 vs Q8NFY4 and Q6UW88, supplementary Fig. S4). Remarkably, however, if the significant disruption patterns are observed in more than one TM length, these patterns are always consistent, as evident in Fig. 4a and also in all constructs in Fig. 5. This suggests that while different length constructs may behave

differently with respect to their tendency to display disruption, the underlying interactions between the helices tend to be consistent.

Mutagenesis supports the model of twelve predicted GAS_{right} homodimers

The twelve GAS_{right} dimers validated by the mutagenesis are illustrated in Fig. 5. They include proteins involved in important biological functions in processes, such as cell-surface signaling, proliferation and adhesion (semaphorins 5A, 6B and 6D, syndecan-3), transport (transferrin receptor protein 1) and the modulation of immune response (interleukin-22 receptor subunit alpha-1, butyrophilin-like protein 3, hepatitis A virus cellular receptor 2). The computationally predicted interfacial positions are highlighted in green in the structural representation and the sequence. A darker shade of green identifies the most central four amino acids that are involved in the tightest interaction at the point of closest approach between the helical pairs. These positions include GxxxG or GxxxG-like motif, with the Gly residues expected to be the most sensitive positions to mutation. Some of these proteins were previously known to form homodimers (all semaphorins, syndecan-3 and the transferrin receptor protein 1), but of these, the TM domain is known to mediate dimerization only for syndecan-3. There is no prior evidence of self-association for the other seven proteins (although some are known to form heterologous complexes with other proteins), therefore these findings may be helpful to shed light on their biological mechanisms.

To visualize the TM domain in the context of the entire protein, we also produced predictions using AF3. The TM domains of these predictions are compared to the CATM models in Fig. 6. In nine of the twelve cases, the models generated by the two programs are in excellent agreement with each other. In two cases, AF3 places the TM domains in proximity but not in contact or poorly packed. Interestingly, in one case (Butyrophilin-like protein 3), AF3 predicts with high confidence a well-packed helical dimer but its structure is different than the CATM model.

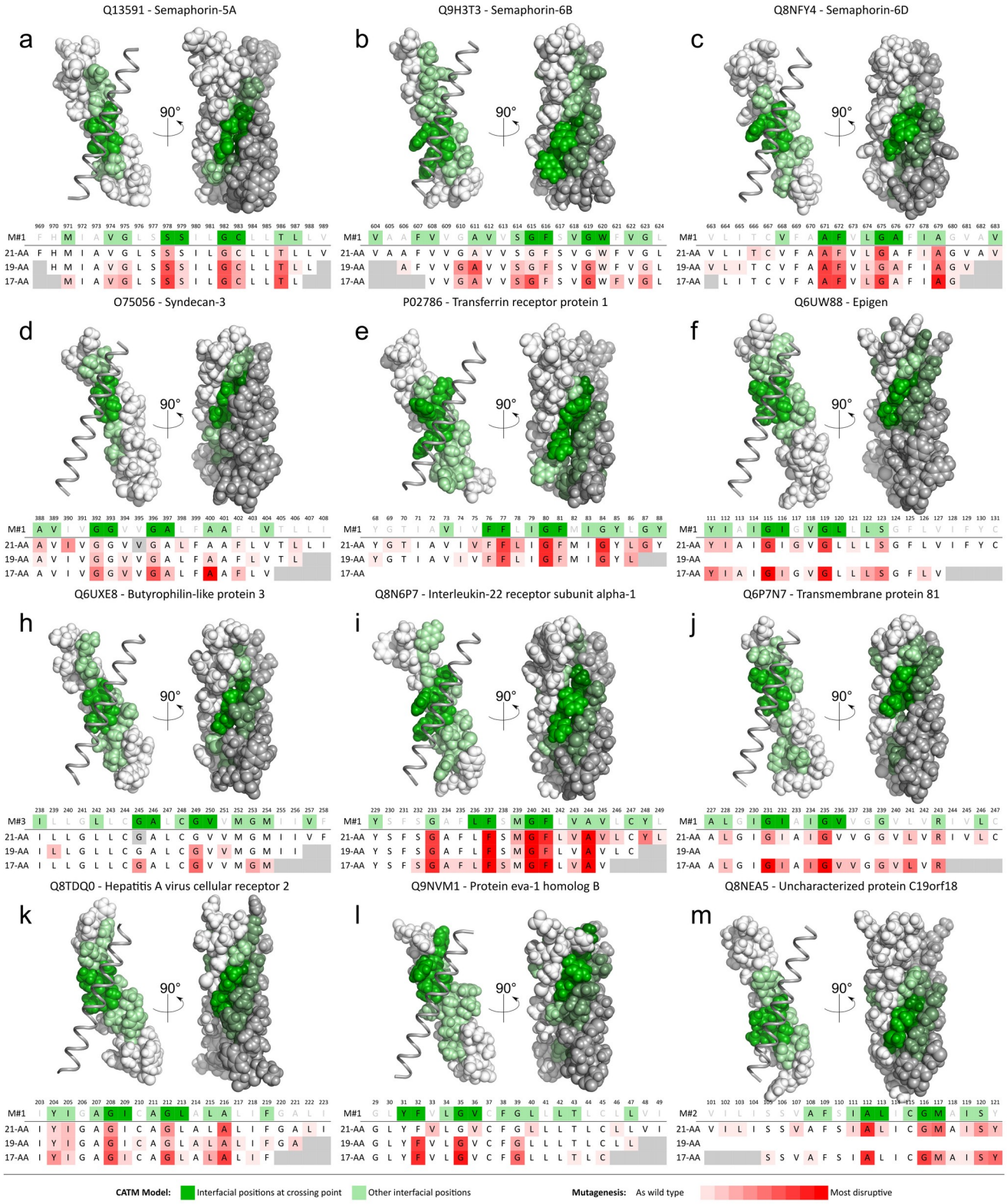

Fig. 5. The mutagenesis of twelve constructs is consistent with the predicted GAS_{right} interface. The figure illustrates the structural model, with the interfacial positions highlighted in green. The darker green identifies the four residues that mediate the tight interactions at the crossing point between the helices. The same color-coding is provided in the sequence. M#1, M#2 and M#3 refer to CATM's first, second and third top model, chosen for their matching with the mutational profile. The average profile of the three length constructs (21-AA, 19-AA and 17-AA) is show with the degree of disruption indicated in red.

A remarkable finding of the AF3/CATM comparison is that the level of correspondence between CATM and AF3 is very high only among this set of validated GAS_{right} constructs (9 out of 12, Table S3) whereas AF3 and CATM agree only in 3 of the remaining 88 constructs (3.4%). This disproportion is highly significant statistically (χ^2 test, $p = 4.4\times10^{-10}$). The converging outcomes between an AI-based method that uses co-evolutionary information and a physics-based method increase the confidence that these structural predictions correspond to biologically relevant structures.

Semaphorins 5A, 6B and 6D

The 100 TM helix library contained three members of the semaphorin (SEMA) family (SEMA 5A, 6B and 6D) and, remarkably, all three displayed strong TOXGREEN signals and mutagenesis profiles consistent with the predicted GAS_{right} model (Fig. 5a, b and c). Semaphorins are a diverse family of cell-surface signaling proteins originally discovered for their role in guiding axon development[68] but are critical for regulating cell morphology, differentiation and cell-cell interactions in both neuronal and non-neuronal cell types, including in angiogenesis and immune response. Semaphorins engage in short-range cell-to-cell interactions with plexin receptors in adjacent cells[69]. These *"trans"* (between cells) interactions are mediated by semaphorins' characteristic *sema* domain, which forms homodimers that bring together two plexin molecules, triggering their signaling[70–73]. The response to the signals typically induces significant cellular changes in the cytoskeletal and adhesive elements that specify cell morphology[74–76], affecting integrin function and actin dynamics[77,78], inducing actin filament disassembly[79], and modifying membrane dynamics such as endo- and exocytosis[80]. Semphorins/plexin interactions can also occur in *"cis"*, between molecules on the same cells, and these are important to modulate or suppress intercellular signaling[81].

SEMA 5A regulates neural circuit development in vertebrates[71,82,83]. In humans, it has been identified as an autism susceptibility gene[84,85], and, in mice, SEMA 5A mutants negatively display increased numbers of excitatory synapses and deficits in social interaction, a hallmark of autism-spectrum disorders[86]. Additionally, SEMA 5A promotes angiogenesis in endothelial cells[87]. Structurally, SEMA 5A forms disulfide-linked dimers through its extracellular thrombospondin type-1

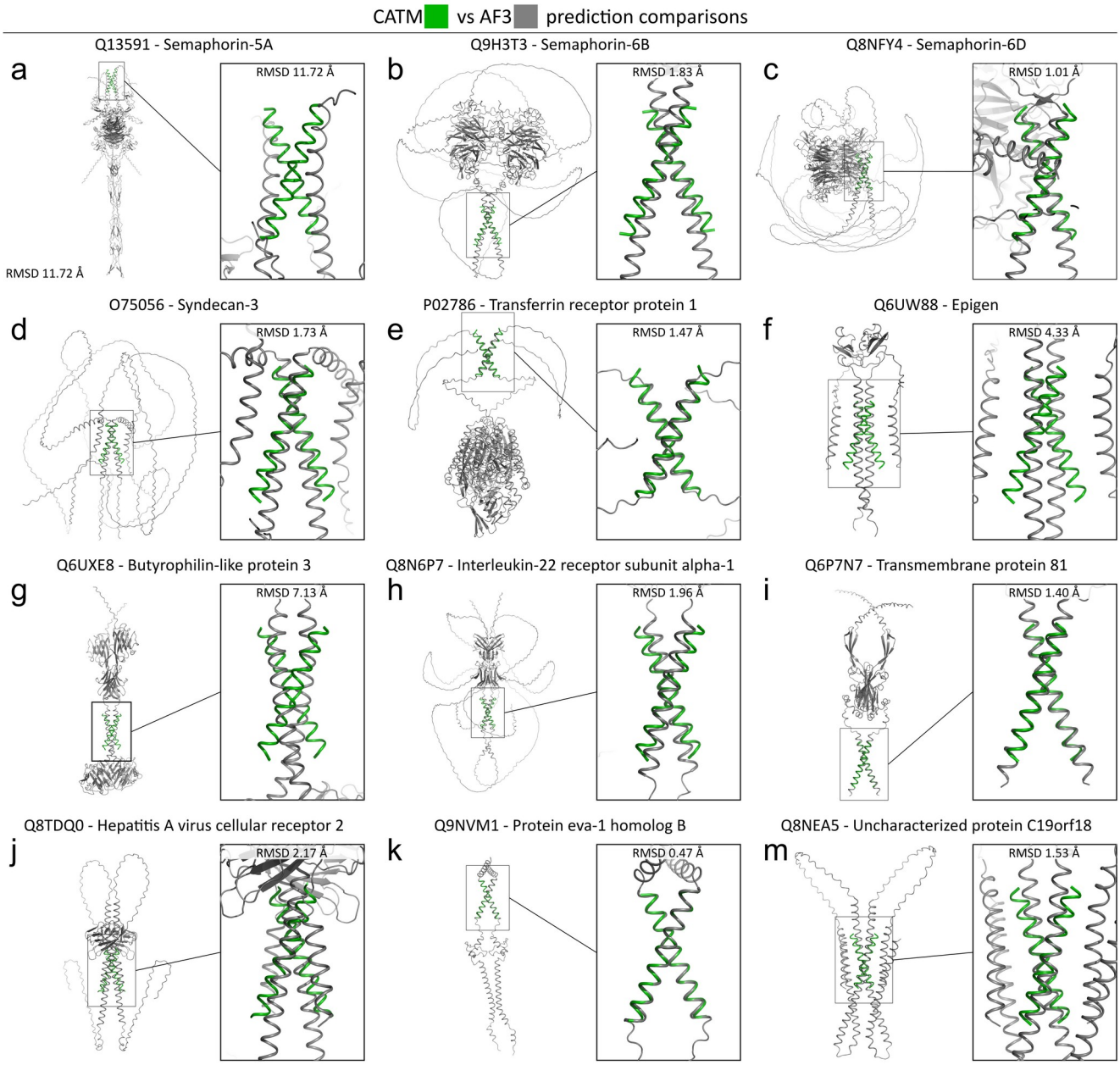

Fig. 6. Correspondence between the CATM and AlphaFold3 models for twelve validated GAS_{right} constructs. The prediction from CATM is represented in green. The AF3 predictions are represented in gray.

repeats (TSR)[88], a unique feature of the semaphorin class 5 family that enables their interactions with the glycosaminoglycan moieties proteoglycans[89]. The crystal structure of SEMA 5A shows that its third and fourth TSR repeats link the two chains through a domain-swapped fold[90]. Our mutagenesis confirms that the TM helix SEMA 5A forms stable GAS_{right} dimers through an interface that involves residues M_{971}xxVGxxSSxxGCxxT_{986} (Fig. 5a). The data therefore suggests that two SEMA 5A chains

form an elongated dimeric structure interacting at the N-terminus through the sema domain, through the TSR domains in the middle of the protein, and via a GAS_{right} motif in the C-terminal TM domain.

The other two semaphorins (SEMA 6B and 6D) are members of the type 6 family, which is involved in neuronal and non-neuronal morphogenesis, primarily through interactions with class A plexins[91,92]. SEMA 6B is known as the receptor of a toxin released by the clostridium Paeniclostridium sordellii, which causes an almost invariably lethal toxic shock syndrome associated with gynecological infections[93,94]. SEMA 6D is involved in cardiac morphogenesis[95], bone homeostasis, and immune response[96]. SEMA 6D is an interesting potential drug target since it can hamper cytotoxic T cell infiltration in head and neck cancer, therefore its modulation could increase the efficacy of immune checkpoint inhibitors for tumors resistant to immunotherapy[97].

A characteristic of the semaphorin 6 family, is their long cytoplasmic C-terminal region[98], which is critical for their ability to perform reverse signaling, i.e., transmitting external signals to their own cell through the recruitment and activation of kinases[99]. In SEMA 6B, mutations leading to truncation of this cytoplasmic region cause progressive myoclonic epilepsy in humans and produce defective development of brain neurons in zebrafish[100]. In SEMA 6D, reverse signaling regulates anti-inflammatory polarization and lipid metabolism in macrophages[101] and is also important in chickens for cardiac development[99].

The mutagenesis profile of both SEMA 6B-TM and 6D-TM are in good agreement with the CATM model. The profile of SEMA 6B identifies the positions most sensitive to mutation as the A_{611}xxxGxxxG_{619} motif at the center of the predicted interface (Fig. 5b). The profile of SEMA 6D is also in excellent agreement with a predicted interface where A_{671}FxLGAxxA_{679} are the most important residues (Fig. 5c).

The CATM models of SEMA 6B and 6D are confirmed by AF3 (1.73 and 1.01 Å RMSD, respectively), as illustrated in Fig. 6b and c. Interestingly, SEMA 5A is one of the two cases among the twelve validated GAS_{right} dimers for which AF3 does not predict an interaction between the TM helices (Fig. 6a). Overall, the computational and mutational data provide strong evidence that the TM

helix of these semaphorins is a dimerization domain. Considering the frequent involvement of TM helices in mediating receptor signals in and out of the cell[3–5], the sensitive positions identified by the validated interfaces provide the opportunity for assessing whether the TM dimers act as functional elements in these semaphorins.

Syndecan-3

Syndecans are a family of four proteins that are responsible for cell signaling, proliferation, and adhesion in mammals. Their extracellular N-terminal region is covalently linked to glycosaminoglycan chains, which mediate interactions with a multitude of ligands, including growth factors, chemokines, cytokines, proteinases, adhesion receptors, and collagens[102,103]. The interactions between the cytoplasmic domain of syndecans and intracellular kinases and the actin cytoskeleton regulate cell adhesion and migration, with implications in both tumor progression and suppression[104]. All four syndecans homodimerize and form heterologous complexes among each other through interactions mediated by the GxxxG motif in their transmembrane domain[105,106]. Our mutational analysis included syndecan-3, which is expressed in neural tissues and developing musculoskeletal tissues. The mutagenesis confirms that syndecan-3 associates through a GAS_{right} dimer centered around its G_{392}xxxGxxxA_{400} motif (Fig. 5d). The AF3 model of syndecan-3 is overall very low confidence but the program identifies a GAS_{right} motif similar to that predicted by CATM (1.73 Å RMSD, Fig. 6d).

Transferrin receptor protein 1

The Transferrin receptor protein 1 (TfR1, Fig. 5e) mediates the cellular uptake of iron. Insoluble iron is distributed through the organism as Fe^{3+} bound to transferrin. Endocytosis of the TfR1/transferrin complex leads to the release of iron in the endosomal compartment through acidification, followed by the recycling of the apotransferrin-receptor complex to the cell surface[107]. The receptor is necessary for the development of erythrocytes, and the immune and nervous systems and patients with TfR1 mutations can have immunodeficiency characterized by impaired T and B cell function[108]. The extracellular domain of TfR1 is a homodimer in solution[109,110] therefore our data

indicate that the TM domain is an additional dimerization domain for TfR1. The CATM model of the TfR1 TM dimer is in good agreement with the AF3 dimeric prediction (1.47 Å RMSD), which also models the transferrin binding domain connected to the TM domain by an extended low confidence region (supplementary Fig. 6e).

Butyrophilin-like protein 3 (BTNL3)

Butyrophilin-like protein 3 (BTNL3, Fig. 5g) is one of five members of the human butyrophilin-like protein family. They are poorly understood proteins with co-stimulatory and inhibitory roles in the immune response[111,112]. BTNL3-TM displayed limited disruption in its mutagenesis profile, with a pattern emerging only in the 17-AA construct, highlighting the sensitivity of the G_{245}xxxGxxxG_{253} motif. Interestingly, AF3 produces a model in which the interaction interface is diametrically opposite to this GxxxGxxxG motif (Fig. 6g). The AF3 model is high-confidence and well packed, however, its interfacial positions (L_{240}xxxCxxLxxxVxxM_{254}) are not susceptible to mutation in our TOXGREEN analysis. The presence of two potential interaction interfaces, however, may explain the limited sensitivity to mutation. If these two configurations are physiological, it suggests that BTNL3's function somehow involves a conformational switch in its TM region, corresponding to a nearly 180° rotation around its TM helices.

Interleukin-22 receptor subunit alpha-1 (IL-22R1)

IL-22R1 and IL-22R2 form a heterodimeric transmembrane receptor complex that binds to the interleukin 22 cytokine and activates a number of immune cell types[113]. There are currently no reports in the literature indicating that IL-22R1 can form homodimeric complexes. The mutagenesis of IL-22R1-TM (Fig. 5h) displays similar profiles across the 21-, 19- and 17-AA constructs, with strong disruption at positions G_{233}xxxFxxGFxxA_{244}, in excellent agreement with CATM's molecular model. The AF3 prediction is also in agreement with CATM (RMDS 1.96 Å, Fig. 6x).

Transmembrane protein 81 (TMEM81)

TMEM81 was recently discovered to be part of a trimeric complex formed with two other single-pass fertilization factors, Izumo1 and Spaca6, which are essential in the binding and fusion of sperm and egg[114]. In this recent study, the authors produced an AlphaFold multimer model of the Izumo1/Spaca6/TMEM81 complex. In that model, the TM domains of Izumo1 and TMEM81 form a GAS_{right} heterodimer using the same positions (A_{227}LxxGIxxGVxxG_{239}) that are involved in the homodimer predicted by CATM (Fig. 5i). The AF3 homodimeric model of TMEM81 is in excellent agreement with CATM's model (RMDS 1.40 Å, Fig. 6i). If both the homodimeric and heterodimeric forms of TMEM81 are physiological, they would have to be alternative states occurring at different functional or assembly stages of these proteins.

Hepatitis A virus cellular receptor 2 (HAVCR2)

HAVCR2, also known as TIM-3, is a cell-surface receptor implicated in modulating innate and adaptive immune responses in T-cells and other immune cells. TIM-3 is recruited to the immunological synapse[115,116]. It is known to bind to several ligands, which induce the phosphorylation of tyrosine residues in its cytoplasmic tail[117]. This triggers the dissociation of BAT3 adaptor protein and association with the FYN kinase, leading to the inhibitory activity on T cell receptor signaling[118]. Our mutagenesis indicates that TIM-3-TM homodimerizes through an interface formed by Y_{204}IxxGIxxxGLxxLA_{216} (Fig. 5j). The prediction is confirmed by AF3, which is relatively similar and utilizes the same interface (RMDS 2.17 Å, Fig. 6j).

Other proteins

Protein eva-1 homolog B (EVA1B) is a poorly studied member of the EVA1 family. Its better known paralog EVA1A localized in the endoplasmic reticulum and lysosome and it is involved in regulating autophagy and apoptosis[119]. EVA1B is upregulated in glioma and is considered a poor prognostic biomarker in this disease[120]. We found that EVA1B-TM oligomerizes strongly in TOXGREEN in a manner that is highly sensitive to mutation in its F_{32}xxGxxxG_{39} motif (Fig. 5k). The

AF3 model is nearly identical to the CATM model (RMDS 0.47 Å, Fig. 6k). **Epigen** is a small single-pass membrane protein. Following proteolysis, its extracellular domain acts as an Epidermal growth factor receptor ligand, promoting cell differentiation[121,122]. We found no reports indicating epigen homodimerization. CATM's prediction of Epigen-TM (Fig. 5f) is not replicated by AF3, which produces two poorly interacting parallel helices (Fig. 6f). **C19orf18** is a protein of unknown function (Fig. 5m) that is up-regulated in pancreatic cancer[123]. CATM and AF3 predict similar models (RMDS 1.53 Å, Fig. 6m).

Finally, three of the mutagenesis profiles identified here do not match the CATM predictions but are consistent with their AF3 models (illustrated in supplementary Fig. S5). These are **Oxidized low-density lipoprotein receptor 1**, which is implicated in organ damage caused by dyslipidemia and that mediates the internalization and degradation of oxidatively modified low density lipoprotein by endothelial cells[124]. The second is **Cytotoxic T-lymphocyte protein 4**, a receptor that acts as a negative regulator of T-cell responses[125]. Lastly, **Cell cycle progression protein 1** is ER-resident protein that is transcriptionally upregulated upon ER stress and is important for stimulating ER selective autophagy[126].

Conclusions

TOXCAT is a widely employed assay for the determination of the oligomeric properties of TM helices. It has contributed to our understanding of the biology and the folding of numerous single-pass membrane protein systems. Here we demonstrate that its GFP-expressing variant, TOXGREEN, is applicable to the simultaneous analysis of oligomerization in large libraries of TM helices. This rapid, large-scale evaluation opens the door to new experiments for understanding the interactions, energetics and biology of single-pass protein systems as well as new protein engineering approaches.

We have shown that the reconstructed fluorescence values obtained statistically from FACS and NGS data reproduce the direct measurements of individual constructs with good accuracy and sensitivity. We also established a maltose growth test compatible with large libraries, which is necessary for identifying constructs that may not be inserted correctly or efficiently in the membrane. One aspect of the workflow that cannot be performed yet on a large scale is the systematic assessment of protein expression, which is typically performed using western blots. To address this issue, Fleishman and coworkers previously replaced the MBP moiety of the TOXCAT construct with the antibiotic resistance enzyme β-lactamase and used selection in ampicillin as a proxy for expression and insertion[30]. Since the TOXGREEN and TOXCAT constructs are identical (the only difference being the nature of the reporter gene), the same strategy could be adopted in TOXGREEN sort-seq. It should be noted that controlling for expression is not necessary for obtaining quality mutagenesis profiles because closely related constructs, such as point mutants, tend to have similar levels of expression[27,42,127].

TOXGREEN sort-seq produced high-quality, reproducible mutational profiles from a large library of 17,400 constructs. We found that the longer TM inserts, and particularly the 21-AA library, produced stronger fluorescence signals but were less susceptible to disruption. The shorter constructs tend to fluoresce more weakly, but they have a higher propensity to display disruption patterns when they produce sufficient signal. Based on these observations, we did not identify an optimal insert length

and suggest that testing each TM helix at multiple lengths may be an effective way to address this issue. Our large-scale analysis also confirmed that the "limited" strategy of three mutations at each position can be effective in determining an interaction interface, although a larger selection of mutants (such as A, I, F, G, L, M and V) is likely to be more informative when affordable.

TOXGREEN sort-seq can be used to rapidly evaluate the dimerization propensity of entire groups of single-pass proteins, to study their energetics or to engineer TM dimers. The present analysis was used as an opportunity to further assess CATM's performance in predicting GAS_{right} homodimers from primary sequence. The algorithm performed well when tested against the small set of known GAS_{right} dimers[43], but its propensity for producing false positives was not previously assessed because of the limited availability of single-pass proteins in the structural database. In the present analysis, 35 constructs produced a weak TOXGREEN signal and 31 did not display an appreciable disruption pattern. Of the remaining 34 constructs, CATM's GAS_{right} predictions were supported in 12 cases (35%). Remarkably, this subset has a high rate of correspondence with the AF3 predictions, supporting the hypothesis that the models correspond to biologically relevant structures. The high-throughout mutagenesis data provided by TOXGREEN sort-seq could be used in the future to improve the algorithm's precision. It should also be noted that 13 constructs produced a pattern that mapped to a possible alternative interface. Three of these mutagenesis profiles are corroborated by the AF3 prediction, suggesting that these interfaces are also biologically relevant.

Methods

Computational prediction of dimeric GAS_{right} structures

The structural models of the 100 GAS_{right} dimers were predicted in previous work from all human protein sequences annotated as single-pass membrane proteins in Uniprot using CATM[42]. These models are available at https://catm.biochem.wisc.edu/CATM. As summarized in supplementary Table S4, the sequences were selected among models with a CATM energy score of at least -10 kcal/mol. For consistency, we limited the selection to sequences with a predicted length of 21 amino acids, which is the most common length annotated for transmembrane domains in Uniprot. Sequences containing proline were excluded because their tendency to create kinks in α-helices. Sequences containing more than one strongly polar amino acid (D, E, H, K, N, or Q) were also removed because these amino acids have a tendency to cause non-specific association in TM helices driven by hydrogen bonding[67]. Any sequences for which part of the predicted interface included any residues outside of the TM domain were also removed. The final library consisted of 100 biological TM helices expressed at three different lengths, a 21 amino acid full-length construct and truncations to 19 and 17 amino acids operated at the N- and/or C-terminal ends of the helices, eliminating positions predicted not to be involved at the dimerization interface. The final library thus consisted of 300 WT sequences, plus three mutants for each position at each length, for a theoretical total of 17,400 constructs (100 constructs ×(3 WT + 3×21 + 3×19 + 3×17 mutants)).

The full-length dimeric structures of each candidate GAS_{right} protein were also predicted using AlphaFold 3[128] via the AlphaFold Server available at https://alphafoldserver.com/. The full sequence of every protein was used as input with two copies of every sequence for multimer prediction. All parameters were set as default.

Cloning and Expression of Chimeric Protein Libraries

All sequences encoding for TM domains of interest were ordered as oligo pools from Twist Bioscience. Distinct predefined subsets (segments) of the oligo pools were individually amplified using

unique primer pairs specific to each segment by quantitative PCR (KAPA SYBR Fast qPCR master mix, Roche) based on the protocol outlined in Kosuri et al.[129] followed by another traditional PCR reaction using the qPCR reaction as a template. These amplified segments were digested with NheI and DpnII and cloned into the pccGFPKan vector[32] at the NHeI-BamHI restriction sites. The DNA sequences of the inserts were designed such that the last two nucleotides are "TT" before the DpnII cut site "G" resulting in an in-frame "L" once ligated into the corresponding pccGFPKan restriction sites. The resulting chimeric protein sequences are: RAS–TM Sequence of Interest–LILI (e.g., RAS–LLAAGILGAGALIAGMCFIII–LILI for LLAAGILGAGALIAGMCFIII). TOXGREEN construct library segments were transformed into *malE* deficient *E. coli* strain MM39. Each transformation outgrowth was grown overnight at 37 °C in 3 mL LB containing 100 µg/mL ampicillin to store segments or libraries as 25% glycerol stocks. Segments were determined to be successfully cloned if the calculated colony-forming units on LB agar plates supplemented with 100 µg/mL ampicillin were at least tenfold the number of sequences in the segment. To create composite libraries from individual segments or colonies, each library component was inoculated into 3 mL LB containing 100 µg/mL ampicillin and grown overnight at 37 °C. 30 uL of each overnight culture was used to inoculate 3 mL fresh LB containing 100 µg/mL ampicillin and incubated at 37 °C until an optical density of approximately 0.1 at 600 nm was reached. The optical density at 600 nm of each culture was normalized by theoretical unique sequence count and then combined in equal proportion for every component to a final volume of 1 mL. 25% glycerol stocks were prepared for all combined libraries for storage at -80 °C.

Selection of individual constructs for the validation set

A "validation set" of approximately 100 constructs was created for the purpose of developing the TOXGREEN sort-seq method (supplementary Table S1). The mutation length libraries were sorted using a SONY LE-SH800 into three populations: low, medium, and high fluorescence. The low bin was defined with any fluorescence lower than the median fluorescence of the monomeric mutant of GpA, G83I. The medium bin was drawn with the lower bound set to the upper bound of the low bin

and the upper bound set to the median fluorescence value of GpA. The high gate was set to any cells with fluorescence above the median fluorescence level of GpA. The sorted cells were plated on LB supplemented with 100 μg/mL ampicillin and incubated at 37°C overnight. A total of 109 colonies were chosen between each of these pools and sent for sequencing. This resulted in 98 confirmed variants and the standard three traditional controls of GpA, GpA-G83I, and the base No-TM pccGFPKan plasmid. TOXGREEN fluorescence was measured individually for each construct in the set.

MalE Complementation Assay

To confirm proper membrane insertion and orientation of individual TOXGREEN constructs, overnight cultures were plated on M9 minimal medium plates containing 0.4% maltose as the only carbon source and grown at 37 °C for 72 h[27].

MalE complementation assay for the libraries was performed using liquid M9 maltose minimal media. Overnight cultures of each library were used to separately inoculate 500 mL M9 maltose media to a final concentration of 0.00125 optical density at 600 nm. Samples were taken following 30-36 h of growth at 37°C. DNA was extracted from M9 media samples along with the overnight LB cultures used for initial inoculation and sent for NGS.

The negative (non-inserting) subset of constructs from the validation set (highlighted in supplementary Table S1) was included in any liquid MalE complementation assay sample as a threshold growth level for any failing samples. For all protein sequences retrieved from the NGS, the fractional change in population percentage was calculated between the 30-36 h M9 time point and the LB overnight culture sample. Any sample with a fractional change equal to or worse than the best-performing negative sample was removed from further analysis.

The sequences for the negative control set as well as two positive controls have been deposited on the AddGene repository with the following accession numbers: 1E01: AddGene: 239013; 1E11: AddGene: 239083; 2E11: AddGene: 239084; 2F12: AddGene: 239085; 2H11: AddGene: 239086; 1F02: AddGene: 239087; 2H07: AddGene: 239088.

Traditional TOXGREEN

Standard plate reader-based sfGFP expression was quantified in stationary phase as described prior[32]. Briefly, a freshly streaked colony was inoculated into 3 mL of LB broth containing 100 µg/mL ampicillin and grown overnight at 37 °C for all samples. To reduce background, 1.5 mL of cells were collected by centrifugation at 17,000 g and concentrated by re-suspension in 0.5 mL in PBS solution (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 2 mM KH_2PO_4, pH 7.4). 300 µL of each sample was transferred to a 96-well black-walled, clear bottom plate (Fisher Scientific) for fluorescence measurement. Fluorescence measurements were performed using an Infinite M1000 Pro plate reader (Tecan), using an excitation wavelength of 485 nm and recording emission from 500 to 600 nm. The relative sfGFP expression was calculated by normalizing the fluorescence emission at 512 nm to the optical density of the sample at 600 nm measured on an Agilent Technologies Cary 60 UV-Vis.

Flow cytometry reading of individual TOXGREEN clones

Flow cytometry-based sfGFP expression was measured using a SONY LE-SH800 with a SONY 70 µM chip. The forward scatter (FSC) threshold was set at 0.05%. The sensor gains for FL1 (GFP) and backscatter (BSC) were set at 90% and 30%, respectively. The first gate was drawn to select for cells by eliminating cell debris on a side scatter area (SSC-A) vs. forward scatter area (FSC-A) plot. The second gate was drawn to eliminate doublet cells on a forward scatter height (FSC-H) vs. FSC-A plot. A freshly streaked colony was inoculated into 3 mL of LB broth containing 100 µg/mL ampicillin and grown overnight at 37 °C for all samples. To reduce background, overnight cultures were added to a PBS solution at a dilution level such that the event rate was below a maximum of 10,000 events per second. Median GFP fluorescence was calculated from gated cells for each variant. The FCS files were analyzed using FlowJo.

Immunoblotting for individual constructs

Expression level of the length variants of several constructs was assessed by Western blot analysis (Fig. S-WESTERN_BLOTS). 2 mL of overnight culture were pelleted, resuspended, and

lysed through chemical lysis (SoluLyse, AMSBIO). For each sample, 45 µL of lysate was added to 15 µL 4X LDS loading buffer (Novex, Life Technologies) with β-mercaptoethanol and heated to 70 °C for 3 min. Cell lysate protein concentrations were determined by bicinchoninic acid assay (Pierce). 10 µg of each LDS protein sample was loaded onto a NuPAGE 4-12% Bis-Tris SDS-PAGE gel (Invitrogen) to separate by SDS-PAGE and transferred to polyvinylidene difluoride membrane (VWR). Peroxidase-conjugated monoclonal anti-maltose binding protein antibodies (Sigma-Alridch) were used for immunoblotting analysis. Blots were developed with ECL Western Blotting Substrate Kit (Pierce) and imaged using an iBright FL1500 (Invitrogen).

Fluorescence-activated cell sorting

50 µL of any library glycerol stock was used to inoculate 3 mL of LB broth supplemented with 100 µg/mL ampicillin and grown overnight at 37 °C. FACS was performed using a SONY LE-SH800 cell sorter with a 70-µm microfluidic sorting chip. The FSC threshold was set at 0.05%. The sensor gains for FL1 (GFP) and BSC were set at 90% and 30%, respectively. The first gate was drawn to select for cells by eliminating cell debris on a SSC-A vs. FSC-A plot. The second gate was drawn to eliminate doublet cells on a FSC-H vs. FSC-A plot. Every library was appropriately diluted in PBS buffer with a flow rate set such that the event rate was at most 10,000 events per second. A suitable number of gates (4-7) were drawn to span the fluorescence distribution of the entire library ensuring every bin was distinguishable from its adjacent bins when sorted. 500,000 events were sorted into every bin. All bins were sorted into 2 mL LB broth containing 100 µg/mL ampicillin. Sorted populations were grown in a total of 5 mL LB broth containing 100 µg/mL ampicillin until an optical density at 600 nm of approximately 0.1 and plasmid DNA was extracted.

Library preparation and deep sequencing

For each sample for NGS, the sequences corresponding to the TM helices were amplified through PCR from the extracted TOXGREEN plasmid DNA followed up by another PCR reaction in order to add the necessary Illumina TruSeq adapter, unique barcode, and “stem” sequences. The products

were submitted to the University of Wisconsin Biotechnology Center DNA Sequencing Facility for Index PCR (TruSeq) and Illumina NovaSeq6000 (2×150 shared) Sequencing services.

Sequence Analysis and TOXGREEN fluorescence reconstruction

A minimum cutoff of 10 copies of a particular sequence had to be present in the NGS for it to be retained. After filtering, each DNA sequence was translated to amino acids and matched to its appropriate TM in length, protein ID, and mutation. All of the above analysis was performed with *ad hoc* Perl scripts. Statistical reconstruction calculations of sfGFP fluorescence output were performed in R and calculated as a weighted average. The distribution of each individual library construct in the bins is then used to reconstruct their estimated GFP fluorescence value with a statistical inference approach based on a weighted average, adapted from Kosuri et al.[57]. Briefly, for each construct *i* within each sorted bin *j,* we first calculate the normalized fractional contribution a_{ij} as follows:

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

where c_{ij} is the sequencing count from NGS for construct *i* in bin *j*, and f_j is the fraction of the sorted sample population in bin *j*.

We then obtain the reconstructed fluorescence, expressed as Relative Fluorescence Units (RFU_i) for each construct *i* by summing the product of the normalized fractional contribution a_{ij} and the median fluorescence m_j over each sorted bin *j*.

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

Mutagenesis Analysis

The disruption level for every mutant was set based on its relative disruption compared to its WT reconstructed fluorescence. A minimum threshold of 50,000 RFU was set for performing the mutagenesis analysis to guarantee sufficient dynamic range. This value was established empirically

and corresponds to approximately the 50% percentile in the distribution of RFU of all WT and mutant constructs (supplementary Fig. S6). For each mutant, a disruption value from 1 to 5 was calculated, with 1 being the most disruptive and 5 being similar or better than the WT. First, a quintile RFU_Q was calculated as one-fifth of the difference between the reconstructed fluorescence of the WT RFU_{WT} minus a base fluorescence value in the monomeric range RFU_B.

$$RFU_Q = \frac{RFU_{WT} - RFU_B}{5}$$

RFU_B was chosen as 35,000, corresponding to approximately the 20% percentile in the distribution of all constructs (supplementary Fig. S6). Then the disruption index $DI_{m,i}$ of a mutant *m* at position *i* was calculated from its reconstructed fluorescence $RFU_{m,i}$ as the following:

$$\begin{aligned}
DI_{m,i} &= 1 \; for \; RFU_{m,i} < RFU_B + 0.5\, RFU_Q \\
DI_{m,i} &= 2 \; for \; RFU_{m,i} < RFU_B + 1.5\, RFU_Q \\
DI_{m,i} &= 3 \; for \; RFU_{m,i} < RFU_B + 2.5\, RFU_Q \\
DI_{m,i} &= 4 \; for \; RFU_{m,i} < RFU_B + 3.5\, RFU_Q \\
DI_{m,i} &= 5 \; for \; RFU_{m,i} \geq RFU_B + 3.5\, RFU_Q
\end{aligned}$$

Finally, for each position *i*, an overall position-based disruption index DI_i was calculated by averaging the disruption value of all *M* mutants (with typically *M* = 3) at that particular position:

$$DI_i = \frac{\sum_m DI_{m,i}}{M}$$

Acknowledgments

This work was supported by National Institutes of Health grant R35 GM130339 to A.S. and National Science Foundation grant CHE-1710182 to A.S.. S.M.A. acknowledges the support of National Library of Medicine training grant T15 LM007359 to the CIBM Training Program and the Dr. James Chieh-Hsia Mai Wisconsin Distinguished Graduate Fellowship in Biochemistry. J.C. acknowledges the support of a Denis R.A. and Martha Washburn Wharton Fellowship in Biochemistry. M.L. acknowledges the support of National Institutes of Health training grant T32 GM08293 to the Molecular Biophysics Training Program. We are grateful to Dr. Kyle Nishikawa for helpful suggestions and discussion.

Conflicts of Interest

The authors declare that they have no conflicts of interest with the contents of this article.

References

(1) Krogh, A.; Larsson, B.; von Heijne, G.; Sonnhammer, E. L. Predicting Transmembrane Protein Topology with a Hidden Markov Model: Application to Complete Genomes. *J. Mol. Biol.* **2001**, *305* (3), 567–580. https://doi.org/10.1006/jmbi.2000.4315.

(2) Moore, D. T.; Berger, B. W.; DeGrado, W. F. Protein-Protein Interactions in the Membrane: Sequence, Structural, and Biological Motifs. *Structure* **2008**, *16* (7), 991–1001. https://doi.org/10.1016/j.str.2008.05.007.

(3) Bugge, K.; Lindorff-Larsen, K.; Kragelund, B. B. Understanding Single-Pass Transmembrane Receptor Signaling from a Structural Viewpoint-What Are We Missing? *FEBS J.* **2016**, *283* (24), 4424–4451. https://doi.org/10.1111/febs.13793.

(4) Endres, N. F.; Das, R.; Smith, A. W.; Arkhipov, A.; Kovacs, E.; Huang, Y.; Pelton, J. G.; Shan, Y.; Shaw, D. E.; Wemmer, D. E.; Groves, J. T.; Kuriyan, J. Conformational Coupling across the Plasma Membrane in Activation of the EGF Receptor. *Cell* **2013**, *152* (3), 543–556. https://doi.org/10.1016/j.cell.2012.12.032.

(5) Hubbard, S. R.; Miller, W. T. Receptor Tyrosine Kinases: Mechanisms of Activation and Signaling. *Curr. Opin. Cell Biol.* **2007**, *19* (2), 117–123. https://doi.org/10.1016/j.ceb.2007.02.010.

(6) Fisher, L. E.; Engelman, D. M.; Sturgis, J. N. Detergents Modulate Dimerization, but Not Helicity, of the Glycophorin A Transmembrane Domain. *J. Mol. Biol.* **1999**, *293* (3), 639–651. https://doi.org/10.1006/jmbi.1999.3126.

(7) Khadria, A.; Senes, A. Measurement of Transmembrane Peptide Interactions in Liposomes Using Förster Resonance Energy Transfer (FRET). *Methods Mol. Biol. Clifton NJ* **2013**, *1063*, 19–36. https://doi.org/10.1007/978-1-62703-583-5_2.

(8) Khadria, A. S.; Senes, A. The Transmembrane Domains of the Bacterial Cell Division Proteins FtsB and FtsL Form a Stable High-Order Oligomer. *Biochemistry* **2013**, *52* (43), 7542–7550. https://doi.org/10.1021/bi4009837.

(9) Khadria, A. S.; Senes, A. Fluorophores, Environments, and Quantification Techniques in the Analysis of Transmembrane Helix Interaction Using FRET. *Pept. Sci.* **2015**, *104* (4), 247–264. https://doi.org/10.1002/bip.22667.

(10) Merzlyakov, M.; You, M.; Li, E.; Hristova, K. Transmembrane Helix Heterodimerization in Lipid Bilayers: Probing the Energetics behind Autosomal Dominant Growth Disorders. *J. Mol. Biol.* **2006**, *358* (1), 1–7. https://doi.org/10.1016/j.jmb.2006.01.086.

(11) You, M.; Li, E.; Wimley, W. C.; Hristova, K. Förster Resonance Energy Transfer in Liposomes: Measurements of Transmembrane Helix Dimerization in the Native Bilayer Environment. *Anal. Biochem.* **2005**, *340* (1), 154–164. https://doi.org/10.1016/j.ab.2005.01.035.

(12) Díaz Vázquez, G.; Cui, Q.; Senes, A. Thermodynamic Analysis of the GASright Transmembrane Motif Supports Energetic Model of Dimerization. *Biophys. J.* **2023**, *122* (1), 143–155. https://doi.org/10.1016/j.bpj.2022.11.018.

(13) Li, E.; Placone, J.; Merzlyakov, M.; Hristova, K. Quantitative Measurements of Protein Interactions in a Crowded Cellular Environment. *Anal. Chem.* **2008**, *80* (15), 5976–5985. https://doi.org/10.1021/ac800616u.

(14) Fleming, K. G. Standardizing the Free Energy Change of Transmembrane Helix-Helix Interactions. *J. Mol. Biol.* **2002**, *323* (3), 563–571. https://doi.org/10.1016/S0022-2836(02)00920-8.

(15) Choma, C.; Gratkowski, H.; Lear, J. D.; DeGrado, W. F. Asparagine-Mediated Self-Association of a Model Transmembrane Helix. *Nat. Struct. Mol. Biol.* **2000**, *7* (2), 161–166. https://doi.org/10.1038/72440.

(16) Cristian, L.; Lear, J. D.; DeGrado, W. F. Determination of Membrane Protein Stability via Thermodynamic Coupling of Folding to Thiol-Disulfide Interchange. *Protein Sci. Publ. Protein Soc.* **2003**, *12* (8), 1732–1740. https://doi.org/10.1110/ps.0378603.

(17) Cristian, L.; Lear, J. D.; DeGrado, W. F. Use of Thiol-Disulfide Equilibria to Measure the Energetics of Assembly of Transmembrane Helices in Phospholipid Bilayers. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100* (25), 14772–14777. https://doi.org/10.1073/pnas.2536751100.

(18) Hong, H.; Blois, T. M.; Cao, Z.; Bowie, J. U. Method to Measure Strong Protein–Protein Interactions in Lipid Bilayers Using a Steric Trap. *Proc. Natl. Acad. Sci.* **2010**, *107* (46), 19802–19807. https://doi.org/10.1073/pnas.1010348107.

(19) Hong, H.; Chang, Y.-C.; Bowie, J. U. Measuring Transmembrane Helix Interaction Strengths in Lipid Bilayers Using Steric Trapping. *Methods Mol. Biol. Clifton NJ* **2013**, *1063*, 37–56. https://doi.org/10.1007/978-1-62703-583-5_3.

(20) Langosch, D.; Brosig, B.; Kolmar, H.; Fritz, H. J. Dimerisation of the Glycophorin A Transmembrane Segment in Membranes Probed with the ToxR Transcription Activator. *J. Mol. Biol.* **1996**, *263* (4), 525–530. https://doi.org/10.1006/jmbi.1996.0595.

(21) Kolmar, H.; Hennecke, F.; Götze, K.; Janzer, B.; Vogt, B.; Mayer, F.; Fritz, H. J. Membrane Insertion of the Bacterial Signal Transduction Protein ToxR and Requirements of Transcription Activation Studied by Modular Replacement of Different Protein Substructures. *EMBO J.* **1995**, *14* (16), 3895–3904.

(22) Schanzenbach, C.; Schmidt, F. C.; Breckner, P.; Teese, M. G.; Langosch, D. Identifying Ionic Interactions within a Membrane Using BLaTM, a Genetic Tool to Measure Homo- and Heterotypic Transmembrane Helix-Helix Interactions. *Sci. Rep.* **2017**, *7*, 43476. https://doi.org/10.1038/srep43476.

(23) Julius, A.; Laur, L.; Schanzenbach, C.; Langosch, D. BLaTM 2.0, a Genetic Tool Revealing Preferred Antiparallel Interaction of Transmembrane Helix 4 of the Dual-Topology Protein EmrE. *J. Mol. Biol.* **2017**, *429* (11), 1630–1637. https://doi.org/10.1016/j.jmb.2017.04.003.

(24) Lindner, E.; Langosch, D. A ToxR-Based Dominant-Negative System to Investigate Heterotypic Transmembrane Domain Interactions. *Proteins* **2006**, *65* (4), 803–807. https://doi.org/10.1002/prot.21226.

(25) Gurezka, R.; Langosch, D. In Vitro Selection of Membrane-Spanning Leucine Zipper Protein-Protein Interaction Motifs Using POSSYCCAT. *J. Biol. Chem.* **2001**, *276* (49), 45580–45587. https://doi.org/10.1074/jbc.M105362200.

(26) Lindner, E.; Unterreitmeier, S.; Ridder, A. N. J. A.; Langosch, D. An Extended ToxR POSSYCCAT System for Positive and Negative Selection of Self-Interacting Transmembrane Domains. *J. Microbiol. Methods* **2007**, *69* (2), 298–305. https://doi.org/10.1016/j.mimet.2007.01.015.

(27) Russ, W. P.; Engelman, D. M. TOXCAT: A Measure of Transmembrane Helix Association in a Biological Membrane. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96* (3), 863–868.

(28) Dawson, J. P.; Weinger, J. S.; Engelman, D. M. Motifs of Serine and Threonine Can Drive Association of Transmembrane Helices. *J. Mol. Biol.* **2002**, *316* (3), 799–805. https://doi.org/10.1006/jmbi.2001.5353.

(29) Russ, W. P.; Engelman, D. M. The GxxxG Motif: A Framework for Transmembrane Helix-Helix Association. *J. Mol. Biol.* **2000**, *296* (3), 911–919. https://doi.org/10.1006/jmbi.1999.3489.

(30) Elazar, A.; Weinstein, J.; Biran, I.; Fridman, Y.; Bibi, E.; Fleishman, S. J. Mutational Scanning Reveals the Determinants of Protein Insertion and Association Energetics in the Plasma Membrane. *eLife* **2016**, *5*. https://doi.org/10.7554/eLife.12125.

(31) Bennasroune, A.; Gardin, A.; Auzan, C.; Clauser, E.; Dirrig-Grosch, S.; Meira, M.; Appert-Collin, A.; Aunis, D.; Crémel, G.; Hubert, P. Inhibition by Transmembrane Peptides of Chimeric Insulin Receptors. *Cell. Mol. Life Sci. CMLS* **2005**, *62* (18), 2124–2131. https://doi.org/10.1007/s00018-005-5226-9.

(32) Armstrong, C. R.; Senes, A. Screening for Transmembrane Association in Divisome Proteins Using TOXGREEN, a High-Throughput Variant of the TOXCAT Assay. *Biochim. Biophys. Acta* **2016**, *1858* (11), 2573–2583. https://doi.org/10.1016/j.bbamem.2016.07.008.

(33) Berger, B. W.; Kulp, D. W.; Span, L. M.; DeGrado, J. L.; Billings, P. C.; Senes, A.; Bennett, J. S.; DeGrado, W. F. Consensus Motif for Integrin Transmembrane Helix Association. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107* (2), 703–708. https://doi.org/10.1073/pnas.0910873107.

(34) Schneider, D.; Engelman, D. M. GALLEX, a Measurement of Heterologous Association of Transmembrane Helices in a Biological Membrane. *J. Biol. Chem.* **2003**, *278* (5), 3105–3111. https://doi.org/10.1074/jbc.M206287200.

(35) Su, P.-C.; Berger, B. W. Identifying Key Juxtamembrane Interactions in Cell Membranes Using AraC-Based Transcriptional Reporter Assay (AraTM). *J. Biol. Chem.* **2012**, *287* (37), 31515–31526. https://doi.org/10.1074/jbc.M112.396895.

(36) Su, P.-C.; Berger, B. W. A Novel Assay for Assessing Juxtamembrane and Transmembrane Domain Interactions Important for Receptor Heterodimerization. *J. Mol. Biol.* **2013**, *425* (22), 4652–4658. https://doi.org/10.1016/j.jmb.2013.07.022.

(37) Lomize, A. L.; Schnitzer, K. A.; Todd, S. C.; Cherepanov, S.; Outeiral, C.; Deane, C. M.; Pogozheva, I. D. Membranome 3.0: Database of Single-Pass Membrane Proteins with AlphaFold Models. *Protein Sci. Publ. Protein Soc.* **2022**, *31* (5), e4318. https://doi.org/10.1002/pro.4318.

(38) Lemmon, M. A.; Flanagan, J. M.; Treutlein, H. R.; Zhang, J.; Engelman, D. M. Sequence Specificity in the Dimerization of Transmembrane Alpha-Helices. *Biochemistry* **1992**, *31* (51), 12719–12725.

(39) Adams, P. D.; Engelman, D. M.; Brünger, A. T. Improved Prediction for the Structure of the Dimeric Transmembrane Domain of Glycophorin A Obtained through Global Searching. *Proteins* **1996**, *26* (3), 257–261. https://doi.org/10.1002/(SICI)1097-0134(199611)26:3<257::AID-PROT2>3.0.CO;2-B.

(40) Arkin, I. T.; Adams, P. D.; MacKenzie, K. R.; Lemmon, M. A.; Brünger, A. T.; Engelman, D. M. Structural Organization of the Pentameric Transmembrane Alpha-Helices of Phospholamban, a Cardiac Ion Channel. *EMBO J.* **1994**, *13* (20), 4757–4764. https://doi.org/10.1002/j.1460-2075.1994.tb06801.x.

(41) Adams, P. D.; Arkin, I. T.; Engelman, D. M.; Brünger, A. T. Computational Searching and Mutagenesis Suggest a Structure for the Pentameric Transmembrane Domain of Phospholamban. *Nat. Struct. Biol.* **1995**, *2* (2), 154–162.

(42) Anderson, S. M.; Mueller, B. K.; Lange, E. J.; Senes, A. Combination of Cα-H Hydrogen Bonds and van Der Waals Packing Modulates the Stability of GxxxG-Mediated Dimers in Membranes. *J. Am. Chem. Soc.* **2017**, *139* (44), 15774–15783. https://doi.org/10.1021/jacs.7b07505.

(43) Mueller, B. K.; Subramaniam, S.; Senes, A. A Frequent, GxxxG-Mediated, Transmembrane Association Motif Is Optimized for the Formation of Interhelical Cα-H Hydrogen Bonds. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (10), E888-895. https://doi.org/10.1073/pnas.1319944111.

(44) Walters, R. F. S.; DeGrado, W. F. Helix-Packing Motifs in Membrane Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103* (37), 13658–13663. https://doi.org/10.1073/pnas.0605878103.

(45) Zhang, S.-Q.; Kulp, D. W.; Schramm, C. A.; Mravic, M.; Samish, I.; DeGrado, W. F. The Membrane- and Soluble-Protein Helix-Helix Interactome: Similar Geometry via Different Interactions. *Struct. Lond. Engl. 1993* **2015**, *23* (3), 527–541. https://doi.org/10.1016/j.str.2015.01.009.

(46) Senes, A.; Gerstein, M.; Engelman, D. M. 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.* **2000**, *296* (3), 921–936. https://doi.org/10.1006/jmbi.1999.3488.

(47) Russ, W. P.; Engelman, D. M. The GxxxG Motif: A Framework for Transmembrane Helix-Helix Association. *J. Mol. Biol.* **2000**, *296* (3), 911–919. https://doi.org/10.1006/jmbi.1999.3489.

(48) Brosig, B.; Langosch, D. The Dimerization Motif of the Glycophorin A Transmembrane Segment in Membranes: Importance of Glycine Residues. *Protein Sci. Publ. Protein Soc.* **1998**, *7* (4), 1052–1056. https://doi.org/10.1002/pro.5560070423.

(49) MacKenzie, K. R.; Prestegard, J. H.; Engelman, D. M. A Transmembrane Helix Dimer: Structure and Implications. *Science* **1997**, *276* (5309), 131–133.

(50) Sulistijo, E. S.; MacKenzie, K. R. Sequence Dependence of BNIP3 Transmembrane Domain Dimerization Implicates Side-Chain Hydrogen Bonding and a Tandem GxxxG Motif in Specific Helix-Helix Interactions. *J. Mol. Biol.* **2006**, *364* (5), 974–990. https://doi.org/10.1016/j.jmb.2006.09.065.

(51) Bragin, P. E.; Mineev, K. S.; Bocharova, O. V.; Volynsky, P. E.; Bocharov, E. V.; Arseniev, A. S. HER2 Transmembrane Domain Dimerization Coupled with Self-Association of Membrane-Embedded Cytoplasmic Juxtamembrane Regions. *J. Mol. Biol.* **2016**, *428* (1), 52–61. https://doi.org/10.1016/j.jmb.2015.11.007.

(52) Mineev, K. S.; Bocharov, E. V.; Pustovalova, Y. E.; Bocharova, O. V.; Chupin, V. V.; Arseniev, A. S. Spatial Structure of the Transmembrane Domain Heterodimer of ErbB1 and ErbB2 Receptor Tyrosine Kinases. *J. Mol. Biol.* **2010**, *400* (2), 231–243. https://doi.org/10.1016/j.jmb.2010.05.016.

(53) Bocharov, E. V.; Mineev, K. S.; Goncharuk, M. V.; Arseniev, A. S. Structural and Thermodynamic Insight into the Process of "Weak" Dimerization of the ErbB4 Transmembrane Domain by Solution NMR. *Biochim. Biophys. Acta* **2012**, *1818* (9), 2158–2170. https://doi.org/10.1016/j.bbamem.2012.05.001.

(54) Bocharov, E. V.; Mineev, K. S.; Volynsky, P. E.; Ermolyuk, Y. S.; Tkach, E. N.; Sobol, A. G.; Chupin, V. V.; Kirpichnikov, M. P.; Efremov, R. G.; Arseniev, A. S. Spatial Structure of the Dimeric Transmembrane Domain of the Growth Factor Receptor ErbB2 Presumably Corresponding to the Receptor Active State. *J. Biol. Chem.* **2008**, *283* (11), 6950–6956. https://doi.org/10.1074/jbc.M709202200.

(55) Bocharov, E. V.; Mayzel, M. L.; Volynsky, P. E.; Goncharuk, M. V.; Ermolyuk, Y. S.; Schulga, A. A.; Artemenko, E. O.; Efremov, R. G.; Arseniev, A. S. Spatial Structure and pH-Dependent Conformational Diversity of Dimeric Transmembrane Domain of the Receptor Tyrosine Kinase EphA1. *J. Biol. Chem.* **2008**, *283* (43), 29385–29395. https://doi.org/10.1074/jbc.M803089200.

(56) Lau, T.-L.; Kim, C.; Ginsberg, M. H.; Ulmer, T. S. The Structure of the Integrin alphaIIbbeta3 Transmembrane Complex Explains Integrin Transmembrane Signalling. *EMBO J.* **2009**, *28* (9), 1351–1361. https://doi.org/10.1038/emboj.2009.63.

(57) Kosuri, S.; Goodman, D. B.; Cambray, G.; Mutalik, V. K.; Gao, Y.; Arkin, A. P.; Endy, D.; Church, G. M. Composability of Regulatory Sequences Controlling Transcription and Translation in Escherichia Coli. *Proc. Natl. Acad. Sci.* **2013**, *110* (34), 14024–14029. https://doi.org/10.1073/pnas.1301301110.

(58) Arkin, I. T.; Adams, P. D.; MacKenzie, K. R.; Lemmon, M. A.; Brunger, A. T.; Engelman, D. M. Structural Organization of the Pentameric Transmembrane Cx-Helices of Phospholamban, a Cardiac Ion Channel.

(59) Lawrie, C. M.; Sulistijo, E. S.; MacKenzie, K. R. Intermonomer Hydrogen Bonds Enhance GxxxG-Driven Dimerization of the BNIP3 Transmembrane Domain: Roles for Sequence Context in Helix-Helix Association in Membranes. *J. Mol. Biol.* **2010**, *396* (4), 924–936. https://doi.org/10.1016/j.jmb.2009.12.023.

(60) LaPointe, L. M.; Taylor, K. C.; Subramaniam, S.; Khadria, A.; Rayment, I.; Senes, A. Structural Organization of FtsB, a Transmembrane Protein of the Bacterial Divisome. *Biochemistry* **2013**, *52* (15), 2574–2585. https://doi.org/10.1021/bi400222r.

(61) Khadria, A. S.; Mueller, B. K.; Stefely, J. A.; Tan, C. H.; Pagliarini, D. J.; Senes, A. A Gly-Zipper Motif Mediates Homodimerization of the Transmembrane Domain of the Mitochondrial Kinase ADCK3. *J. Am. Chem. Soc.* **2014**, *136* (40), 14068–14077. https://doi.org/10.1021/ja505017f.

(62) Matthews, E. E.; Thévenin, D.; Rogers, J. M.; Gotow, L.; Lira, P. D.; Reiter, L. A.; Brissette, W. H.; Engelman, D. M. Thrombopoietin Receptor Activation: Transmembrane Helix Dimerization, Rotation, and Allosteric Modulation. *FASEB J.* **2011**, *25* (7), 2234. https://doi.org/10.1096/fj.10-178673.

(63) Lockey, C.; Young, H.; Brown, J.; Dixon, A. M. Characterization of Interactions within the Igα/Igβ Transmembrane Domains of the Human B-Cell Receptor Provides Insights into Receptor Assembly. *J. Biol. Chem.* **2022**, *298* (5). https://doi.org/10.1016/j.jbc.2022.101843.

(64) Chai, M.; Liu, B.; Sun, F.; Zhu, Z.; Xu, L.; Luo, S.-Z. The Dimerization of PSGL-1 Is Driven by the Transmembrane Domain via a Leucine Zipper Motif. *Proteins Struct. Funct. Bioinforma.* **2018**, *86* (8), 844–852. https://doi.org/10.1002/prot.25498.

(65) Wang, H.; Barreyro, L.; Provasi, D.; Djemil, I.; Torres-Arancivia, C.; Filizola, M.; Ubarretxena-Belandia, I. Molecular Determinants and Thermodynamics of the Amyloid Precursor Protein Transmembrane Domain Implicated in Alzheimer's Disease. *J. Mol. Biol.* **2011**, *408* (5), 879–895. https://doi.org/10.1016/j.jmb.2011.03.028.

(66) Hellmann, N.; Schneider, D. Hydrophobic Mismatch and Sequence Specificity Compete When Transmembrane Helix-Helix Interactions Are Measured with the TOXCAT Assay. *Front. Chem.* **2022**, *10*, 1049310. https://doi.org/10.3389/fchem.2022.1049310.

(67) Zhou, F. X.; Merianos, H. J.; Brunger, A. T.; Engelman, D. M. Polar Residues Drive Association of Polyleucine Transmembrane Helices. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98* (5), 2250–2255. https://doi.org/10.1073/pnas.041593698.

(68) Kolodkin, A. L.; Matthes, D. J.; Goodman, C. S. The Semaphorin Genes Encode a Family of Transmembrane and Secreted Growth Cone Guidance Molecules. *Cell* **1993**, *75* (7), 1389–1399. https://doi.org/10.1016/0092-8674(93)90625-z.

(69) Junqueira Alves, C.; Yotoko, K.; Zou, H.; Friedel, R. H. Origin and Evolution of Plexins, Semaphorins, and Met Receptor Tyrosine Kinases. *Sci. Rep.* **2019**, *9* (1), 1970. https://doi.org/10.1038/s41598-019-38512-y.

(70) Janssen, B. J. C.; Robinson, R. A.; Pérez-Brangulí, F.; Bell, C. H.; Mitchell, K. J.; Siebold, C.; Jones, E. Y. Structural Basis of Semaphorin-Plexin Signalling. *Nature* **2010**, *467* (7319), 1118–1122. https://doi.org/10.1038/nature09468.

(71) Nogi, T.; Yasui, N.; Mihara, E.; Matsunaga, Y.; Noda, M.; Yamashita, N.; Toyofuku, T.; Uchiyama, S.; Goshima, Y.; Kumanogoh, A.; Takagi, J. Structural Basis for Semaphorin Signalling through the Plexin Receptor. *Nature* **2010**, *467* (7319), 1123–1127. https://doi.org/10.1038/nature09473.

(72) Liu, H.; Juo, Z. S.; Shim, A. H.-R.; Focia, P. J.; Chen, X.; Garcia, K. C.; He, X. Structural Basis of Semaphorin-Plexin Recognition and Viral Mimicry from Sema7A and A39R Complexes with PlexinC1. *Cell* **2010**, *142* (5), 749–761. https://doi.org/10.1016/j.cell.2010.07.040.

(73) Lu, D.; Shang, G.; He, X.; Bai, X.-C.; Zhang, X. Architecture of the Sema3A/PlexinA4/Neuropilin Tripartite Complex. *Nat. Commun.* **2021**, *12* (1), 3172. https://doi.org/10.1038/s41467-021-23541-x.

(74) Kruger, R. P.; Aurandt, J.; Guan, K.-L. Semaphorins Command Cells to Move. *Nat. Rev. Mol. Cell Biol.* **2005**, *6* (10), 789–800. https://doi.org/10.1038/nrm1740.

(75) Neufeld, G.; Sabag, A. D.; Rabinovicz, N.; Kessler, O. Semaphorins in Angiogenesis and Tumor Progression. *Cold Spring Harb. Perspect. Med.* **2012**, *2* (1), a006718. https://doi.org/10.1101/cshperspect.a006718.

(76) Kanth, S. M.; Gairhe, S.; Torabi-Parizi, P. The Role of Semaphorins and Their Receptors in Innate Immune Responses and Clinical Diseases of Acute Inflammation. *Front. Immunol.* **2021**, *12*, 672441. https://doi.org/10.3389/fimmu.2021.672441.

(77) Bos, J. L.; Rehmann, H.; Wittinghofer, A. GEFs and GAPs: Critical Elements in the Control of Small G Proteins. *Cell* **2007**, *129* (5), 865–877. https://doi.org/10.1016/j.cell.2007.05.018.

(78) Etienne-Manneville, S.; Hall, A. Rho GTPases in Cell Biology. *Nature* **2002**, *420* (6916), 629–635. https://doi.org/10.1038/nature01148.

(79) Hung, R.-J.; Terman, J. R. Extracellular Inhibitors, Repellents, and Semaphorin/Plexin/MICAL-Mediated Actin Filament Disassembly. *Cytoskelet. Hoboken NJ* **2011**, *68* (8), 415–433. https://doi.org/10.1002/cm.20527.

(80) Tojima, T.; Hines, J. H.; Henley, J. R.; Kamiguchi, H. Second Messengers and Membrane Trafficking Direct and Organize Growth Cone Steering. *Nat. Rev. Neurosci.* **2011**, *12* (4), 191–203. https://doi.org/10.1038/nrn2996.

(81) Haklai-Topper, L.; Mlechkovich, G.; Savariego, D.; Gokhman, I.; Yaron, A. Cis Interaction between Semaphorin6A and Plexin-A4 Modulates the Repulsive Response to Sema6A. *EMBO J.* **2010**, *29* (15), 2635–2645. https://doi.org/10.1038/emboj.2010.147.

(82) Matsuoka, R. L.; Chivatakarn, O.; Badea, T. C.; Samuels, I. S.; Cahill, H.; Katayama, K.-I.; Kumar, S. R.; Suto, F.; Chédotal, A.; Peachey, N. S.; Nathans, J.; Yoshida, Y.; Giger, R. J.; Kolodkin, A. L. Class 5 Transmembrane Semaphorins Control Selective Mammalian Retinal Lamination and Function. *Neuron* **2011**, *71* (3), 460–473. https://doi.org/10.1016/j.neuron.2011.06.009.

(83) Lett, R. L. M.; Wang, W.; O'Connor, T. P. Semaphorin 5B Is a Novel Inhibitory Cue for Corticofugal Axons. *Cereb. Cortex N. Y. N 1991* **2009**, *19* (6), 1408–1421. https://doi.org/10.1093/cercor/bhn179.

(84) Kim, S. A.; Kim, B.-N.; Kim, J.-W.; Shin, M.-S.; Park, T.-W.; Son, J.-W.; Chung, U.-S.; Park, M. Polymorphism in the Promoter Region of SEMA5A Is Associated with Sociality Traits in Korean Subjects with Autism Spectrum Disorders. *Psychiatry Investig.* **2017**, *14* (6), 876–878. https://doi.org/10.4306/pi.2017.14.6.876.

(85) Mosca-Boidron, A.-L.; Gueneau, L.; Huguet, G.; Goldenberg, A.; Henry, C.; Gigot, N.; Pallesi-Pocachard, E.; Falace, A.; Duplomb, L.; Thevenon, J.; Duffourd, Y.; St-Onge, J.; Chambon, P.; Rivière, J.-B.; Thauvin-Robinet, C.; Callier, P.; Marle, N.; Payet, M.; Ragon, C.; Goubran Botros, H.; Buratti, J.; Calderari, S.; Dumas, G.; Delorme, R.; Lagarde, N.; Pinoit, J.-M.; Rosier, A.; Masurel-Paulet, A.; Cardoso, C.; Mugneret, F.; Saugier-Veber, P.; Campion, D.; Faivre, L.; Bourgeron, T. A de Novo Microdeletion of SEMA5A in a Boy with Autism Spectrum Disorder and Intellectual Disability. *Eur. J. Hum. Genet. EJHG* **2016**, *24* (6), 838–843. https://doi.org/10.1038/ejhg.2015.211.

(86) Duan, Y.; Wang, S.-H.; Song, J.; Mironova, Y.; Ming, G.; Kolodkin, A. L.; Giger, R. J. Semaphorin 5A Inhibits Synaptogenesis in Early Postnatal- and Adult-Born Hippocampal Dentate Granule Cells. *eLife* **2014**, *3*, e04390. https://doi.org/10.7554/eLife.04390.

(87) Sadanandam, A.; Rosenbaugh, E. G.; Singh, S.; Varney, M.; Singh, R. K. Semaphorin 5A Promotes Angiogenesis by Increasing Endothelial Cell Proliferation, Migration, and Decreasing Apoptosis. *Microvasc. Res.* **2010**, *79* (1), 1–9. https://doi.org/10.1016/j.mvr.2009.10.005.

(88) Adams, R. H.; Betz, H.; Püschel, A. W. A Novel Class of Murine Semaphorins with Homology to Thrombospondin Is Differentially Expressed during Early Embryogenesis. *Mech. Dev.* **1996**, *57* (1), 33–45. https://doi.org/10.1016/0925-4773(96)00525-4.

(89) Kantor, D. B.; Chivatakarn, O.; Peer, K. L.; Oster, S. F.; Inatani, M.; Hansen, M. J.; Flanagan, J. G.; Yamaguchi, Y.; Sretavan, D. W.; Giger, R. J.; Kolodkin, A. L. Semaphorin 5A Is a Bifunctional Axon Guidance Cue Regulated by Heparan and Chondroitin Sulfate Proteoglycans. *Neuron* **2004**, *44* (6), 961–975. https://doi.org/10.1016/j.neuron.2004.12.002.

(90) Nagy, G. N.; Zhao, X.-F.; Karlsson, R.; Wang, K.; Duman, R.; Harlos, K.; El Omari, K.; Wagner, A.; Clausen, H.; Miller, R. L.; Giger, R. J.; Jones, E. Y. Structure and Function of Semaphorin-5A Glycosaminoglycan Interactions. *Nat. Commun.* **2024**, *15* (1), 2723. https://doi.org/10.1038/s41467-024-46725-7.

(91) Tanaka, T.; Ekimoto, T.; Nagatomo, M.; Neyazaki, M.; Shimoji, E.; Yamane, T.; Kanagawa, S.; Oi, R.; Mihara, E.; Takagi, J.; Akashi, S.; Ikeguchi, M.; Nogi, T. Hybrid in Vitro/in Silico Analysis of Low-Affinity Protein-Protein Interactions That Regulate Signal Transduction by Sema6D. *Protein Sci. Publ. Protein Soc.* **2022**, *31* (11), e4452. https://doi.org/10.1002/pro.4452.

(92) Tran, T. S.; Kolodkin, A. L.; Bharadwaj, R. Semaphorin Regulation of Cellular Morphology. *Annu. Rev. Cell Dev. Biol.* **2007**, *23*, 263–292. https://doi.org/10.1146/annurev.cellbio.22.010605.093554.

(93) Lee, H.; Beilhartz, G. L.; Kucharska, I.; Raman, S.; Cui, H.; Lam, M. H. Y.; Liang, H.; Rubinstein, J. L.; Schramek, D.; Julien, J.-P.; Melnyk, R. A.; Taipale, M. Recognition of Semaphorin Proteins by P. Sordellii Lethal Toxin Reveals Principles of Receptor Specificity in Clostridial Toxins. *Cell* **2020**, *182* (2), 345-356.e16. https://doi.org/10.1016/j.cell.2020.06.005.

(94) Tian, S.; Liu, Y.; Wu, H.; Liu, H.; Zeng, J.; Choi, M. Y.; Chen, H.; Gerhard, R.; Dong, M. Genome-Wide CRISPR Screen Identifies Semaphorin 6A and 6B as Receptors for Paeniclostridium Sordellii Toxin TcsL. *Cell Host Microbe* **2020**, *27* (5), 782-792.e7. https://doi.org/10.1016/j.chom.2020.03.007.

(95) Toyofuku, T.; Zhang, H.; Kumanogoh, A.; Takegahara, N.; Suto, F.; Kamei, J.; Aoki, K.; Yabuki, M.; Hori, M.; Fujisawa, H.; Kikutani, H. Dual Roles of Sema6D in Cardiac Morphogenesis through Region-Specific Association of Its Receptor, Plexin-A1, with off-Track and Vascular Endothelial Growth Factor Receptor Type 2. *Genes Dev.* **2004**, *18* (4), 435–447. https://doi.org/10.1101/gad.1167304.

(96) Takegahara, N.; Takamatsu, H.; Toyofuku, T.; Tsujimura, T.; Okuno, T.; Yukawa, K.; Mizui, M.; Yamamoto, M.; Prasad, D. V. R.; Suzuki, K.; Ishii, M.; Terai, K.; Moriya, M.; Nakatsuji, Y.; Sakoda, S.; Sato, S.; Akira, S.; Takeda, K.; Inui, M.; Takai, T.; Ikawa, M.; Okabe, M.; Kumanogoh, A.; Kikutani, H. Plexin-A1 and Its Interaction with DAP12 in Immune Responses and Bone Homeostasis. *Nat. Cell Biol.* **2006**, *8* (6), 615–622. https://doi.org/10.1038/ncb1416.

(97) Fernández-Nogueira, P.; Linzoain-Agos, P.; Cueto-Remacha, M.; De la Guia-Lopez, I.; Recalde-Percaz, L.; Parcerisas, A.; Gascon, P.; Carbó, N.; Gutierrez-Uzquiza, A.; Fuster, G.; Bragado, P. Role of Semaphorins, Neuropilins and Plexins in Cancer Progression. *Cancer Lett.* **2024**, *606*, 217308. https://doi.org/10.1016/j.canlet.2024.217308.

(98) Wagner, W.; Ochman, B.; Wagner, W. Semaphorin 6 Family-An Important Yet Overlooked Group of Signaling Proteins Involved in Cancerogenesis. *Cancers* **2023**, *15* (23), 5536. https://doi.org/10.3390/cancers15235536.

(99) Toyofuku, T.; Zhang, H.; Kumanogoh, A.; Takegahara, N.; Yabuki, M.; Harada, K.; Hori, M.; Kikutani, H. Guidance of Myocardial Patterning in Cardiac Development by Sema6D Reverse Signalling. *Nat. Cell Biol.* **2004**, *6* (12), 1204–1211. https://doi.org/10.1038/ncb1193.

(100) Hamanaka, K.; Imagawa, E.; Koshimizu, E.; Miyatake, S.; Tohyama, J.; Yamagata, T.; Miyauchi, A.; Ekhilevitch, N.; Nakamura, F.; Kawashima, T.; Goshima, Y.; Mohamed, A. R.; Ch'ng, G.-S.; Fujita, A.; Azuma, Y.; Yasuda, K.; Imamura, S.; Nakashima, M.; Saitsu, H.; Mitsuhashi, S.; Mizuguchi, T.; Takata, A.; Miyake, N.; Matsumoto, N. De Novo Truncating Variants in the Last Exon of SEMA6B Cause Progressive Myoclonic Epilepsy. *Am. J. Hum. Genet.* **2020**, *106* (4), 549–558. https://doi.org/10.1016/j.ajhg.2020.02.011.

(101) Kang, S.; Nakanishi, Y.; Kioi, Y.; Okuzaki, D.; Kimura, T.; Takamatsu, H.; Koyama, S.; Nojima, S.; Nishide, M.; Hayama, Y.; Kinehara, Y.; Kato, Y.; Nakatani, T.; Shimogori, T.; Takagi, J.; Toyofuku, T.; Kumanogoh, A. Semaphorin 6D Reverse Signaling Controls Macrophage Lipid Metabolism and Anti-Inflammatory Polarization. *Nat. Immunol.* **2018**, *19* (6), 561–570. https://doi.org/10.1038/s41590-018-0108-0.

(102) Czarnowski, D. Syndecans in Cancer: A Review of Function, Expression, Prognostic Value, and Therapeutic Significance. *Cancer Treat. Res. Commun.* **2021**, *27*, 100312. https://doi.org/10.1016/j.ctarc.2021.100312.

(103) Gopal, S. Syndecans in Inflammation at a Glance. *Front. Immunol.* **2020**, *11*, 227. https://doi.org/10.3389/fimmu.2020.00227.

(104) Afratis, N. A.; Nikitovic, D.; Multhaupt, H. A. B.; Theocharis, A. D.; Couchman, J. R.; Karamanos, N. K. Syndecans - Key Regulators of Cell Signaling and Biological Functions. *FEBS J.* **2017**, *284* (1), 27–41. https://doi.org/10.1111/febs.13940.

(105) Dews, I. C.; Mackenzie, K. R. Transmembrane Domains of the Syndecan Family of Growth Factor Coreceptors Display a Hierarchy of Homotypic and Heterotypic Interactions. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (52), 20782–20787.

(106) Choi, S.; Lee, E.; Kwon, S.; Park, H.; Yi, J. Y.; Kim, S.; Han, I.-O.; Yun, Y.; Oh, E.-S. Transmembrane Domain-Induced Oligomerization Is Crucial for the Functions of Syndecan-2 and Syndecan-4. *J. Biol. Chem.* **2005**, *280* (52), 42573–42579. https://doi.org/10.1074/jbc.M509238200.

(107) Senyilmaz, D.; Virtue, S.; Xu, X.; Tan, C. Y.; Griffin, J. L.; Miller, A. K.; Vidal-Puig, A.; Teleman, A. A. Regulation of Mitochondrial Morphology and Function by Stearoylation of TFR1. *Nature* **2015**, *525* (7567), 124–128. https://doi.org/10.1038/nature14601.

(108) Jabara, H. H.; Boyden, S. E.; Chou, J.; Ramesh, N.; Massaad, M. J.; Benson, H.; Bainter, W.; Fraulino, D.; Rahimov, F.; Sieff, C.; Liu, Z.-J.; Alshemmari, S. H.; Al-Ramadi, B. K.; Al-Dhekri, H.; Arnaout, R.; Abu-Shukair, M.; Vatsayan, A.; Silver, E.; Ahuja, S.; Davies, E. G.; Sola-Visner, M.; Ohsumi, T. K.; Andrews, N. C.; Notarangelo, L. D.; Fleming, M. D.; Al-Herz, W.; Kunkel, L. M.; Geha, R. S. A Missense Mutation in TFRC, Encoding Transferrin Receptor 1, Causes Combined Immunodeficiency. *Nat. Genet.* **2016**, *48* (1), 74–78. https://doi.org/10.1038/ng.3465.

(109) Lebrón, J. A.; Bennett, M. J.; Vaughn, D. E.; Chirino, A. J.; Snow, P. M.; Mintier, G. A.; Feder, J. N.; Bjorkman, P. J. Crystal Structure of the Hemochromatosis Protein HFE and Characterization of Its Interaction with Transferrin Receptor. *Cell* **1998**, *93* (1), 111–123. https://doi.org/10.1016/s0092-8674(00)81151-4.

(110) Cheng, Y.; Zak, O.; Aisen, P.; Harrison, S. C.; Walz, T. Structure of the Human Transferrin Receptor-Transferrin Complex. *Cell* **2004**, *116* (4), 565–576. https://doi.org/10.1016/s0092-8674(04)00130-8.

(111) Nussbaumer, O.; Thurnher, M. Functional Phenotypes of Human Vγ9Vδ2 T Cells in Lymphoid Stress Surveillance. *Cells* **2020**, *9* (3), 772. https://doi.org/10.3390/cells9030772.

(112) Arnett, H. A.; Escobar, S. S.; Viney, J. L. Regulation of Costimulation in the Era of Butyrophilins. *Cytokine* **2009**, *46* (3), 370–375. https://doi.org/10.1016/j.cyto.2009.03.009.

(113) Wolk, K.; Witte, E.; Witte, K.; Warszawska, K.; Sabat, R. Biology of Interleukin-22. *Semin. Immunopathol.* **2010**, *32* (1), 17–31. https://doi.org/10.1007/s00281-009-0188-x.

(114) Deneke, V. E.; Blaha, A.; Lu, Y.; Suwita, J. P.; Draper, J. M.; Phan, C. S.; Panser, K.; Schleiffer, A.; Jacob, L.; Humer, T.; Stejskal, K.; Krssakova, G.; Roitinger, E.; Handler, D.; Kamoshita, M.; Vance, T. D. R.; Wang, X.; Surm, J. M.; Moran, Y.; Lee, J. E.; Ikawa, M.; Pauli, A. A Conserved Fertilization Complex Bridges Sperm and Egg in Vertebrates. *Cell* **2024**, *187* (25), 7066-7078.e22. https://doi.org/10.1016/j.cell.2024.09.035.

(115) Clayton, K. L.; Haaland, M. S.; Douglas-Vail, M. B.; Mujib, S.; Chew, G. M.; Ndhlovu, L. C.; Ostrowski, M. A. T Cell Ig and Mucin Domain-Containing Protein 3 Is Recruited to the Immune Synapse, Disrupts Stable Synapse Formation, and Associates with Receptor Phosphatases. *J. Immunol. Baltim. Md 1950* **2014**, *192* (2), 782–791. https://doi.org/10.4049/jimmunol.1302663.

(116) Kataoka, S.; Manandhar, P.; Lee, J.; Workman, C. J.; Banerjee, H.; Szymczak-Workman, A. L.; Kvorjak, M.; Lohmueller, J.; Kane, L. P. The Costimulatory Activity of Tim-3 Requires Akt and MAPK Signaling and Its

Recruitment to the Immune Synapse. *Sci. Signal.* **2021**, *14* (687), eaba0717. https://doi.org/10.1126/scisignal.aba0717.

(117) Joller, N.; Anderson, A. C.; Kuchroo, V. K. LAG-3, TIM-3, and TIGIT: Distinct Functions in Immune Regulation. *Immunity* **2024**, *57* (2), 206–222. https://doi.org/10.1016/j.immuni.2024.01.010.

(118) Rangachari, M.; Zhu, C.; Sakuishi, K.; Xiao, S.; Karman, J.; Chen, A.; Angin, M.; Wakeham, A.; Greenfield, E. A.; Sobel, R. A.; Okada, H.; McKinnon, P. J.; Mak, T. W.; Addo, M. M.; Anderson, A. C.; Kuchroo, V. K. Bat3 Promotes T Cell Responses and Autoimmunity by Repressing Tim-3–Mediated Cell Death and Exhaustion. *Nat. Med.* **2012**, *18* (9), 1394–1400. https://doi.org/10.1038/nm.2871.

(119) Wang, L.; Yu, C.; Lu, Y.; He, P.; Guo, J.; Zhang, C.; Song, Q.; Ma, D.; Shi, T.; Chen, Y. TMEM166, a Novel Transmembrane Protein, Regulates Cell Autophagy and Apoptosis. *Apoptosis Int. J. Program. Cell Death* **2007**, *12* (8), 1489–1502. https://doi.org/10.1007/s10495-007-0073-9.

(120) Qu, S.; Liu, J.; Wang, H. EVA1B to Evaluate the Tumor Immune Microenvironment and Clinical Prognosis in Glioma. *Front. Immunol.* **2021**, *12*, 648416. https://doi.org/10.3389/fimmu.2021.648416.

(121) Dm, F.; Nj, B.; A, K.; E, S.-C.; Po, B.; Jo, M.; Cc, V.; Km, F.; Dj, L.; Ds, L.; Ma, L. EGFR Ligands Differentially Stabilize Receptor Dimers to Specify Signaling Kinetics. *Cell* **2017**, *171* (3). https://doi.org/10.1016/j.cell.2017.09.017.

(122) Bs, K.; D, H.; A, D.-S.; G, M.-R.; L, L.; G, G.; G, K.; A, C.; S, L.; R, E.; V, C.-C.; Z, E.; E, P.; R, P.-K.; Ss, B.; Y, Y. Epigen, the Last Ligand of ErbB Receptors, Reveals Intricate Relationships between Affinity and Mitogenicity. *J. Biol. Chem.* **2005**, *280* (9). https://doi.org/10.1074/jbc.M413919200.

(123) Makler, A.; Narayanan, R. Mining Exosomal Genes for Pancreatic Cancer Targets. *Cancer Genomics Proteomics* **2017**, *14* (3), 161–172. https://doi.org/10.21873/cgp.20028.

(124) Nagase, M.; Kaname, S.; Nagase, T.; Wang, G.; Ando, K.; Sawamura, T.; Fujita, T. Expression of LOX-1, an Oxidized Low-Density Lipoprotein Receptor, in Experimental Hypertensive Glomerulosclerosis. *J. Am. Soc. Nephrol. JASN* **2000**, *11* (10), 1826–1836. https://doi.org/10.1681/ASN.V11101826.

(125) Teft, W. A.; Kirchhof, M. G.; Madrenas, J. A Molecular Perspective of CTLA-4 Function. *Annu. Rev. Immunol.* **2006**, *24*, 65–97. https://doi.org/10.1146/annurev.immunol.24.021605.090535.

(126) Smith, M. D.; Harley, M. E.; Kemp, A. J.; Wills, J.; Lee, M.; Arends, M.; von Kriegsheim, A.; Behrends, C.; Wilkinson, S. CCPG1 Is a Non-Canonical Autophagy Cargo Receptor Essential for ER-Phagy and Pancreatic ER Proteostasis. *Dev. Cell* **2018**, *44* (2), 217-232.e11. https://doi.org/10.1016/j.devcel.2017.11.024.

(127) Duong, M. T.; Jaszewski, T. M.; Fleming, K. G.; MacKenzie, K. R. Changes in Apparent Free Energy of Helix-Helix Dimerization in a Biological Membrane Due to Point Mutations. *J. Mol. Biol.* **2007**, *371* (2), 422–434. https://doi.org/10.1016/j.jmb.2007.05.026.

(128) Abramson, J.; Adler, J.; Dunger, J.; Evans, R.; Green, T.; Pritzel, A.; Ronneberger, O.; Willmore, L.; Ballard, A. J.; Bambrick, J.; Bodenstein, S. W.; Evans, D. A.; Hung, C.-C.; O'Neill, M.; Reiman, D.; Tunyasuvunakool, K.; Wu, Z.; Žemgulytė, A.; Arvaniti, E.; Beattie, C.; Bertolli, O.; Bridgland, A.; Cherepanov, A.; Congreve, M.; Cowen-Rivers, A. I.; Cowie, A.; Figurnov, M.; Fuchs, F. B.; Gladman, H.; Jain, R.; Khan, Y. A.; Low, C. M. R.; Perlin, K.; Potapenko, A.; Savy, P.; Singh, S.; Stecula, A.; Thillaisundaram, A.; Tong, C.; Yakneen, S.; Zhong, E. D.; Zielinski, M.; Žídek, A.; Bapst, V.; Kohli, P.; Jaderberg, M.; Hassabis, D.; Jumper, J. M. Accurate Structure Prediction of Biomolecular Interactions with AlphaFold 3. *Nature* **2024**, *630* (8016), 493–500. https://doi.org/10.1038/s41586-024-07487-w.

(129) Kosuri, S.; Eroshenko, N.; LeProust, E. M.; Super, M.; Way, J.; Li, J. B.; Church, G. M. Scalable Gene Synthesis by Selective Amplification of DNA Pools from High-Fidelity Microchips. *Nat. Biotechnol.* **2010**, *28* (12), 1295–1299. https://doi.org/10.1038/nbt.1716.

SUPPLEMENTARY INFORMATION

High-throughput discovery of transmembrane helix dimers from human single-pass membrane proteins with TOXGREEN sort-seq

Samantha M. Anderson*[1,2], Joshua Choi*[1,2], Megan Leander[1,2] , Srivatsan Raman[1,3,4] and Alessandro Senes[1,@]

[1]Department of Biochemistry

[2]Integrated Program in Biochemistry

[3]Department of Bacteriology

[4]Department of Chemical & Biological Engineering

University of Wisconsin-Madison, Madison, WI 53706

*These authors contributed equally

[@]Corresponding author: senes@wisc.edu

Table S1. Set of 100 constructs used for the validation of the TOXGREEN sort-seq method (validation set).

Maltose Negative Constructs			
Sample ID	**Sequence**	**TOXGREEN (RFU)**	**Addgene ID***
1E01	RAS FALGLGFCLPAGTSLSV LILI	42283 ± 4549	239013
1E11	RAS ILFVIAVASELGYFLCI LILI	40500 ± 4786	239083
2E11	RAS VVIIAVVCCVVGTSLVWIVII LILI	58301 ± 6984	239084
2F12	RAS VVIIAIVCCVVGTSLVWVVII LILI	56432 ± 9758	239085
2H11	RAS GIYFVLGVCFGLLLTLCLLVI LILI	103789 ± 15214	239086
Maltose Positive Constructs			
Sample ID	**Sequence**	**TOXGREEN (RFU)**	**Addgene ID***
1A01 (No TM)	N/A	28564 ± 4134	
1A02 (GpA)	RAS LIIFGMAGVIGT LILI	84488 ± 30200	
1A03 (G83I)	RAS LIIFGMAIVIGT LILI	37388 ± 6366	
1A04	RAS ILTVIIIVVVLLMGIVGAI LILI	47881 ± 10226	
1A05	RAS VMFFTLFILLAGTAVMIIA LILI	41911 ± 5143	
1A06	RAS YSFSGAFLFSMGFLVAVFC LILI	59261 ± 8950	
1A07	RAS LIFIIALGSIAGILIVTMI LILI	44979 ± 5647	
1A09	RAS IFIFLIVAAILGVTIGLLV LILI	52553 ± 9835	
1A11	RAS ILTIVGTIAGIVILSIIIA LILI	46559 ± 6129	
1A12	RAS LILLAVLLLLLCGITAGCV LILI	55531 ± 12767	
1B01	RAS LILLAALLLLLCGVTAGCV LILI	51806 ± 6552	
1B02	RAS GLYFVLGVCFGILLTLCLL LILI	120520 ± 16825	
1B04	RAS FFTAFMVIALLGVWTSVAV LILI	45356 ± 6120	
1B05	RAS ILLIMIFYACLAGGLILAI LILI	42958 ± 6812	
1B06	RAS YLIVVLFGISITGILFNTI LILI	64040 ± 9537	
1B07	RAS FLLWILAAVSLGLFFYSFL LILI	46906 ± 6975	
1B08	RAS IALTIVSLACILGVLLASG LILI	48792 ± 8963	
1B09	RAS GLLICGSLILITGLTFAIF LILI	49588 ± 11334	
1B10	RAS LILLAVFLLLLCGVTAGCV LILI	48667 ± 8701	
1B11	RAS VLLVVVMSLFVGGTLLVFS LILI	41015 ± 4257	
1B12	RAS IFIFLGVAAILGVTIILLL LILI	57145 ± 13770	
1C02	RAS YVTVLAFTLLVLTGGFTWL LILI	38451 ± 5143	
1C03	RAS VLGAAGTALLCAGLLLSLF LILI	37121 ± 4548	
1D03	RAS IMIIVVVCCVVGTSLIWVI LILI	42172 ± 7965	
1D07	RAS MIAVGLSSSIFGCLLTL LILI	77352 ± 14577	
1D08	RAS LILSSLLLLLLGLLVAI LILI	50397 ± 13091	
1D09	RAS YSFSGAFLFIMGFLVAV LILI	50540 ± 6365	
1D10	RAS GVISFALAVGVGAKVAL LILI	63168 ± 8851	
1D11	RAS LITCVFAAFVLGAFIFG LILI	38517 ± 4651	
1D12	RAS LLLSLILVSVGFVLTVF LILI	39523 ± 3141	
1E03	RAS IMIIICCVVLGVVLASS LILI	39178 ± 3812	
1E04	RAS MFVALLGLILGQVVCSV LILI	70148 ± 9480	
1E05	RAS FSLSVMAALTFGCFITA LILI	44838 ± 6918	
1E07	RAS LLLSSLLLLLLGLLVII LILI	51863 ± 4893	
1E08	RAS VVIIVILAILTGAGLAA LILI	57268 ± 6078	
1E09	RAS SLVAFSIALICGMAISY LILI	47237 ± 6035	
1E10	RAS IIAVGLSSSALGCLLTL LILI	71144 ± 12816	
1E12	RAS IWVYGVSGGIFLIMIFL LILI	44502 ± 3130	
1F01	RAS LIMIFYACLAGGLFLAY LILI	38511 ± 3847	
1F03	RAS ITVVIAAAGGALLLILG LILI	33812 ± 1869	
1F04	RAS LFLGTFFISSGLILSVA LILI	35292 ± 2934	
1F05	RAS TVIIIVVVLLMIFVGAV LILI	47629 ± 6864	
1F08	RAS VLISGAVLLLFGALGAF LILI	43995 ± 20663	
1F09	RAS IILSLALAGILGICLVV LILI	41715 ± 5375	

*denotes constructs that have been deposited to AddGene

Table S1. (continued)

Maltose Positive Constructs (continued)			
Sample ID	**Sequence**	**TOXGREEN (RFU)**	**Addgene ID***
1F10	RAS VVIIAVVCCVVGTSLIW LILI	29919 ± 2572	
1F11	RAS LAPATLGVLCLALVVTIL LILI	38072 ± 3532	
1F12	RAS VILYLMVMIGMFAFIIV LILI	45873 ± 7411	
1G01	RAS VALWLAGLLGAGGTVSV LILI	36018 ± 4355	
1F02	RAS CAVVVGVGLIVGFAVGL LILI	40114 ± 4593	239087
1G03	RAS IVVAMTAVGGSIIVMLV LILI	32679 ± 2995	
1G04	RAS IMIIICCVVLGVVLAAS LILI	43265 ± 5954	
1G06	RAS GVISFAAAVGVGAKVAF LILI	52482 ± 4095	
1G07	RAS LGIAMKVLGGVALFWII LILI	36832 ± 3814	
2B09	RAS RLCWLVSYSFAVLLLGCLLFL LILI	58829 ± 13176	
2C01	RAS LALGLGFCLLAGTSLSLLWVY LILI	60075 ± 10503	
2C02	RAS FFWLLLLLTAACSGLLFALHL LILI	80849 ± 13413	
2C06	RAS LGFILSLALAGILGICIVVVV LILI	48972 ± 4340	
2C09	RAS VMFFTLFALLAGTAVIIIAYH LILI	51149 ± 4559	
2C12	RAS ALILLILCLGMVVGLVALGIW LILI	71089 ± 24462	
2D01	RAS GVVLLYILLGTIGTLVALLAA LILI	75511 ± 20110	
2D02	RAS IFYVTVLAATLIVLTGGFTWL LILI	54322 ± 7017	
2D04	RAS LALGLGFCLLAGTSISVLWVY LILI	44782 ± 5595	
2D05	RAS FAFSLSVMAILTFGCFITTAF LILI	63521 ± 12087	
2D06	RAS LFLGTFAISSGLILSVAGFLY LILI	51327 ± 11776	
2D07	RAS LIALVTSGALLAVLGITGYFA LILI	45688 ± 6551	
2D08	RAS YVVCGVISFALAVGVGAKLAF LILI	50581 ± 8857	
2D10	RAS LIALVTSGALLAVLLITGYFL LILI	48688 ± 8081	
2E01	RAS VILIAAVGGGVLLLSALGLII LILI	57675 ± 12561	
2E03	RAS VMFFTLFALPAGTALMIIAAH LILI	48856 ± 8238	
2E04	RAS VLSISCIIFGIVIVGAFCAAF LILI	51104 ± 11471	
2E05	RAS LIILVTSGALLAVLGITGYFL LILI	44356 ± 9136	
2E06	RAS LALGLGACLLAGTSLSVLWVY LILI	60257 ± 12889	
2F02	RAS IFYVTVLFFTLIVLTGGFTWL LILI	55726 ± 6721	
2F03	RAS YIFSGVFLFSMGFLVAVLCYL LILI	Not determined	
2F04	RAS LLAAGILGAGALIALMCFIII LILI	86820 ± 13734	
2F07	RAS GLIFVLGVCFGLLLTLCLLVI LILI	41698 ± 4236	
2F08	RAS VFLLAILGGMAFIFLVLLCLL LILI	104964 ± 14524	
2F11	RAS VIIFFAAVLLLSGALAYCLAL LILI	85508 ± 35055	
2G01	RAS YSFLGAFLFSMGFLVAVLRYL LILI	51438 ± 7071	
2G03	RAS LLAGGILGAGALIALMCFIII LILI	81016 ± 18436	
2G04	RAS AVGALLTTLLVIGIILAVVFI LILI	84434 ± 13010	
2G05	RAS VVLLSGAVLLLFGAIGAFYFW LILI	52724 ± 6885	
2G06	RAS VAAFVIGAVVSGFSVGWFVGL LILI	109036 ± 27813	
2G07	RAS VILLAAVGAGCVIFLLIIIFL LILI	86855 ± 18649	
2G08	RAS WLLGAAMVGAVLTALLAGLVI LILI	97937 ± 20237	
2G09	RAS LGIILSLALAGILGICIVVLV LILI	64310 ± 10115	
2G10	RAS IALTLISLACILGVLLASGLI LILI	67935 ± 15230	
2H01	RAS FHMIAVGLSSSILGCLITLLV LILI	125015 ± 15284	
2H02	RAS VVLISGAVLLLFGAIGAFYFW LILI	48506 ± 5557	
2H03	RAS IFQTLLLLTVVFGFLYGFMLY LILI	75071 ± 15475	
2H05	RAS LIQSLIASGIAGSMIGVITFY LILI	63770 ± 8922	
2H07	RAS IIVAMTAVGGSICVMLVVICL LILI	63416 ± 12577	239088
2H08	RAS ALIITMSVVGGTLLLGIAFCC LILI	49071 ± 12152	
2H09	RAS VILIAAVGGGVLLLSFLGLII LILI	56248 ± 6879	
2H10	RAS VAAFVVGAVVSGFAVGWFVGL LILI	Not determined	

*denotes constructs that have been deposited to AddGene

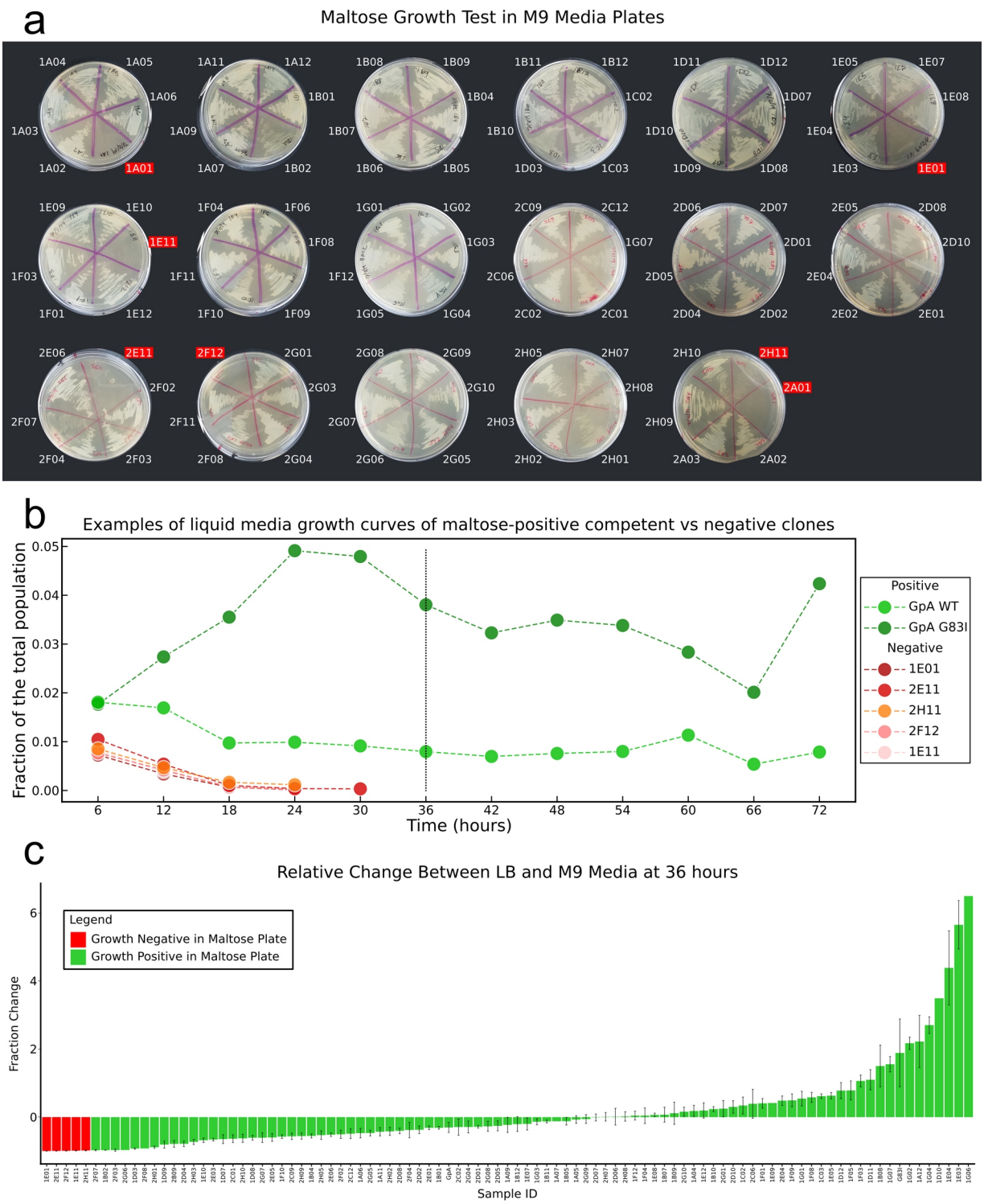

Fig. S1. Adaptation of the maltose growth test for libraries in liquid minimal media. a) Traditional maltose growth test performed in M9 minimal media with maltose as the sole carbon source on each individual constructs of the validation library. The maltose-negative constructs are highlighted in red. b) Liquid media competition maltose growth assay. The figure illustrates the relative abundance of two maltose-positive constructs (shades of green) five maltose-negative constructs (shades of red). By 36 hours the maltose-negative constructs have disappeared from the colture. c) Fraction change between LB and M9 minimal maltose media for all constructs of the validation library. Error bars represent standard deviation of three replicates.

Table S2. Library of 100 computationally predicted GAS_{right} dimers.

Uniprot Entries					Reconstructed Fluorescence		
#	AC	Uniprot ID	Protein Name	TM helix sequence	21 AA	19 AA	17 AA
1	A1L1A6	IGS23_HUMAN	Immunoglobulin superfamily member 23	LLAAGILGAGALIAGMCFIII	60574		37306
2	A6NL88	SHSA7_HUMAN	Protein shisa-7	YVVCGVISFALAVGVGAKVAF	33044	35726	54589
3	B2RUZ4	SMIM1_HUMAN	Small integral membrane protein 1	LGIAMKVLGGVALFWIIFILG	38718		33313
4	F2Z333	FND10_HUMAN	Fibronectin type III domain-containing protein 10	IVVAMTAVGGSICVMLVVICL	45422		24675
5	O00461	GOLI4_HUMAN	Golgi integral membrane protein 4	IFQTLLLLTVVFGFLYGAMLY	64129	40904	38401
6	O14788	TNF11_HUMAN	Tumor necrosis factor ligand superfamily member 11	MFVALLGLGLGQVVCSVALFF	120811		103405
7	O15533	TPSN_HUMAN	Tapasin	GLFLSAFLLLGLFKALGWAAV	94182		43772
8	O43557	TNF14_HUMAN	Tumor necrosis factor ligand superfamily member 14	VGLGLLLLLMGAGLAVQGWFL	93118		40891
9	O43570	CAH12_HUMAN	Carbonic anhydrase 12	LGIILSLALAGILGICIVVVV	41038	38238	42740
10	O43736	ITM2A_HUMAN	Integral membrane protein 2A	MLTLLGLSFILAGLIVGGACI	49116	41160	34750
11	O60344	ECE2_HUMAN	Endothelin-converting enzyme 2	LVLAGASLLLAALLLGCLVAL	130118	62187	48085
12	O75056	SDC3_HUMAN	Syndecan-3	AVIVGGVVGALFAAFLVTLLI	114656	69984	74318
13	O75829	LECT1_HUMAN	Leukocyte cell-derived chemotaxin 1	VVLISGAVLLLFGAIGAFYFW	38753		38741
14	O94898	LRIG2_HUMAN	Leucine-rich repeats and immunoglobulin-like domains protein 2	IVIIVVVCCVVGTSLIWVIVI		30019	26496
15	O95256	I18RA_HUMAN	Interleukin-18 receptor accessory protein	GVVLLYILLGTIGTLVAVLAA	62811		39918
16	O95976	IGSF6_HUMAN	Immunoglobulin superfamily member 6	FLTALVSLLSVYVTGVCVAFI		64674	55502
17	P02786	TFR1_HUMAN	Transferrin receptor protein 1	YGTIAVIVFFLIGFMIGYLGY	81018	88605	46030
18	P11117	PPAL_HUMAN	Lysosomal acid phosphatase	VIVALAVCGSILFLLIVLLLT	108645		48600
19	P11362	FGFR1_HUMAN	Fibroblast growth factor receptor 1	IIIYCTGAFLISCMVGSVIVY	33392	32176	31624
20	P14679	TYRO_HUMAN	Tyrosinase	WLLGAAMVGAVLTALLAGLVS	70105		51214
21	P16410	CTLA4_HUMAN	Cytotoxic T-lymphocyte protein 4	FLLWILAAVSSGLFFYSFLLT		51547	38137
22	P22897	MRC1_HUMAN	Macrophage mannose receptor 1	GVVIIVILLILTGAGLAAYFF	38760		47652
23	P28906	CD34_HUMAN	Hematopoietic progenitor cell antigen CD34	LIALVTSGALLAVLGITGYFL	36343	37037	36137
24	P53801	PTTG_HUMAN	Pituitary tumor-transforming gene 1 protein-interacting protein	ALIITMSVVGGTLLLGIAICC	39370	45271	32269
25	P56975	NRG3_HUMAN	Pro-neuregulin-3, membrane-bound isoform	VLSISCIIFGIVIVGMFCAAF	40623		35201
26	P61009	SPCS3_HUMAN	Signal peptidase complex subunit 3	FAFSLSVMAALTFGCFITTAF	63518	48899	53700
27	P61266	STX1B_HUMAN	Syntaxin-1B	IMIIICCVVLGVVLASSIGGT	71428		36364
28	P78357	CNTP1_HUMAN	Contactin-associated protein 1	ILLGFLVAFLLLGLVGMLVLF	99675	56243	47215
29	P78380	OLR1_HUMAN	Oxidized low-density lipoprotein receptor 1	LAAATLGVLCLGLVVTIMVLG		50214	46938
30	P78562	PHEX_HUMAN	Phosphate-regulating neutral endopeptidase	IALVVFVGGTLVLGTILFLVS		64041	34042
31	P98172	EFNB1_HUMAN	Ephrin-B1	VALFAAVGAGCVIFLLIIIFL	71098		35047
32	Q07075	AMPE_HUMAN	Glutamyl aminopeptidase	VAILCAVVVGVGLIVGLAVGL	68741	68017	36794
33	Q09013	DMPK_HUMAN	Myotonin-protein kinase	LLLFAVVLSRAAALGCIGLVA		34964	37669
34	Q12797	ASPH_HUMAN	Aspartyl/asparaginyl beta-hydroxylase	FFTWFMVIALLGVWTSVAVVW	64574	46580	44647
35	Q12836	ZP4_HUMAN	Zona pellucida sperm-binding protein 4	VLWVAGLSGTLILGALLVSYL			33567
36	Q12864	CAD17_HUMAN	Cadherin-17	AVGILLTTLLVIGIILAVVFI	53019		40056
37	Q13308	PTK7_HUMAN	Inactive tyrosine-protein kinase 7	IGLSVGAAVAYIIAVLGLMFY	58752	39132	41709
38	Q13444	ADA15_HUMAN	Disintegrin and metalloproteinase domain-containing protein 15	GLLLSLLVLLVLVMLGASYWY	41370	37667	35274
39	Q13591	SEM5A_HUMAN	Semaphorin-5A	FHMIAVGLSSSILGCLLTLLV	130913	91379	119744
40	Q13651	I10R1_HUMAN	Interleukin-10 receptor subunit alpha	VIIFFAFVLLLSGALAYCLAL	88755		39982
41	Q14626	I11RA_HUMAN	Interleukin-11 receptor subunit alpha	SLGILSFLGLVAGALALGLWL	37477		36345
42	Q3ZCQ8	TIM50_HUMAN	Mitochondrial import inner membrane translocase subunit TIM50	VALWLAGLLGAGGTVSVVYIF	33041		32996
43	Q5VUB5	F1711_HUMAN	Protein FAM171A1	VFLLAILGGMAFILLVLLCLL	113435		54175
44	Q5VV43	K0319_HUMAN	Dyslexia-associated protein KIAA0319	IFYVTVLAFTLIVLTGGFTWL	49573	36947	35645
45	Q6P7N7	TMM81_HUMAN	Transmembrane protein 81	ALGIGIAIGVVGGVLVRIVLC	93654		57356
46	Q6UW88	EPGN_HUMAN	Epigen	YIAIGIGVGLLLSGFLVIFYC	129492		84371

Table S2. (continued)

Uniprot Entries					Reconstructed Fluorescence		
#	AC	Uniprot ID	Protein Name	TM helix sequence	21 AA	19 AA	17 AA
47	Q6UWL2	SUSD1_HUMAN	Sushi domain-containing protein 1	LLQMAGVGLGSLAVVIILTFL	34870	35238	34019
48	Q6UXE8	BTNL3_HUMAN	Butyrophilin-like protein 3	ILLGLLCGALCGVVMGMIIVF	124359	113444	93910
49	Q6UXM1	LRIG3_HUMAN	Leucine-rich repeats and immunoglobulin-like domains protein 3	VVIIAVVCCVVGTSLVWVVII	40066	28851	25697
50	Q6UY09	CEA20_HUMAN	Carcinoembryonic antigen-related cell adhesion molecule 20	IVIGILAVIAVASELGYFLCI	98263	61792	42022
51	Q6V0L0	CP26C_HUMAN	Cytochrome P450 26C1	LSVLGAAGTALLCAGLLLSLA		38471	34878
52	Q6ZWK6	TM11F_HUMAN	Transmembrane protease serine 11F	FTLAIVAIIGIAIGIVTHFVV	80649	64201	42804
53	Q7Z410	TMPS9_HUMAN	Transmembrane protease serine 9	SIGVVATSLVVLTLGVLLAFL	48900		40460
54	Q86T26	TM11B_HUMAN	Transmembrane protease serine 11B	IFIFLGVAAILGVTIGLLVHF	40916	47705	37695
55	Q86YA3	ZGRF1_HUMAN	Protein ZGRF1	LIQSLIASGIAGSMIGVITLY	45976		46662
56	Q8IUS5	EPHX4_HUMAN	Epoxide hydrolase 4	SLLFWSLVYCYCGLCASIHLL	122391		70754
57	Q8IWT1	SCN4B_HUMAN	Sodium channel subunit beta-4	TLIILAVVGGVIGLLILILLI	80023		45116
58	Q8IYV9	IZUM1_HUMAN	Izumo sperm-egg fusion protein 1	LLGLLICGSLALITGLTFAIF	100806	50946	37658
59	Q8N350	CBARP_HUMAN	Voltage-dependent calcium channel beta subunit-associated regulatory protein	VLLVVVMSLFVGGTLVVLSGV		39732	25062
60	Q8N5D6	GBGT1_HUMAN	Globoside alpha-1,3-N-acetylgalactosaminyltransferase 1	LALGLGFCLLAGTSLSVLWVY	34581	36053	39058
61	Q8N6P7	I22R1_HUMAN	Interleukin-22 receptor subunit alpha-1	YSFSGAFLFSMGFLVAVLCYL	88164	87050	86547
62	Q8N7C0	LRC52_HUMAN	Leucine-rich repeat-containing protein 52	YIFLLLIGFCIFAAGTVAAWL	80486		34535
63	Q8N7S6	ARI2O_HUMAN	Uncharacterized protein ARIH2OS	CILTALLAVSFHSIGVVIMTS	101889	85757	34266
64	Q8N967	LRTM2_HUMAN	Leucine-rich repeat and transmembrane domain-containing protein 2	VIIAGVVCGVVCIMMVVAAAY	71877		38653
65	Q8NDY8	TMM52_HUMAN	Transmembrane protein 52	VGLILLAVLLLLLCGVTAGCV		63260	48591
66	Q8NEA5	CS018_HUMAN	Uncharacterized protein C19orf18	VILISSVAFSIALICGMAISY	95006	48027	51554
67	Q8NFY4	SEM6D_HUMAN	Semaphorin-6D	VLITCVFAAFVLGAFIAGVAV	140712	71364	87855
68	Q8TAB3	PCD19_HUMAN	Protocadherin-19	LIFIIALGSIAGILFVTMIFV	86280	46857	31971
69	Q8TDQ0	HAVR2_HUMAN	Hepatitis A virus cellular receptor 2	IYIGAGICAGLALALIFGALI	119051	50444	56484
70	Q8TEM1	PO210_HUMAN	Nuclear pore membrane glycoprotein 210	VMFFTLFALLAGTAVMIIAYH	45133	39096	36835
71	Q92932	PTPR2_HUMAN	Receptor-type tyrosine-protein phosphatase N2	IALTLVSLACILGVLLASGLI	51485	42639	40627
72	Q96AW1	VOPP1_HUMAN	Vesicular, overexpressed in cancer, prosurvival protein 1	WYFWFLLMMGVLFCCGAGFFI			38218
73	Q96HA4	CA159_HUMAN	Uncharacterized protein C1orf159	LFLGTFFISSGLILSVAGFFY	40711		34816
74	Q9BVV7	TIM21_HUMAN	Mitochondrial import inner membrane translocase subunit Tim21	FTYLIVVLFGISITGGLFYTI	30684	30201	24211
75	Q9BY79	MFRP_HUMAN	Membrane frizzled-related protein	VLLLSSLLLLLLGLLVAIILA	120415		77327
76	Q9C0I4	THS7B_HUMAN	Thrombospondin type-1 domain-containing protein 7B	IWVYGVSGGAFLIMIFLIFTS	35420		50863
77	Q9C0J1	B3GN4_HUMAN	N-acetyllactosaminide beta-1,3-N-acetylglucosaminyltransferase 4	RLCWLVSYSLAVLLLGCLLFL	78006	52249	42528
78	Q9H158	PCDC1_HUMAN	Protocadherin alpha-C1	LVIALACISFLFLGCLLFFVC	91886		37248
79	Q9H1E5	TMX4_HUMAN	Thioredoxin-related transmembrane protein 4	VFFVIATLVFGLFMGLVLVVI	92525		44398
80	Q9H1U4	MEGF9_HUMAN	Multiple epidermal growth factor-like domains protein 9	IIILTVIIIVVVLLMGFVGAV	77286	54805	45208
81	Q9H2X3	CLC4M_HUMAN	C-type lectin domain family 4 member M	ALVLQLLSFMLLAGVLVAILV	41328	67467	36584
82	Q9H3R2	MUC13_HUMAN	Mucin-13	ILTIVGTIAGIVILSMIIALI		45044	35317
83	Q9H3S3	TMPS5_HUMAN	Transmembrane protease serine 5	AVLGALGLLAGAGVGSWLLVL	63246	38649	41333
84	Q9H3T3	SEM6B_HUMAN	Semaphorin-6B	VAAFVVGAVVSGFSVGWFVGL	97476	87596	116181
85	Q9H5V8	CDCP1_HUMAN	CUB domain-containing protein 1	VILIAAVGGGVLLLSALGLII	44791		31139
86	Q9NPU4	CN132_HUMAN	Uncharacterized protein C14orf132 {ECO:0000305}	ILFAVIILTGASFAHLFVALF	122932		44417
87	Q9NVM1	EVA1B_HUMAN	Protein eva-1 homolog B	GLYFVLGVCFGLLLTLCLLVI	111769	114334	142204
88	Q9NY15	STAB1_HUMAN	Stabilin-1	GVGAVLAAGALLGLVAGALYL	38986	58238	49853
89	Q9P126	CLC1B_HUMAN	C-type lectin domain family 1 member B	ALILLILCVGMVVGLVALGIW	54301		47598
90	Q9UHP7	CLC2D_HUMAN	C-type lectin domain family 2 member D	FFLIMFLTIIVCGMVAALSAI	62777	43182	60868
91	Q9UJ14	GGT7_HUMAN	Gamma-glutamyltransferase 7	LTVIVTACLTFATGVTVALVM	62433		28854
92	Q9UJ37	SIA7B_HUMAN	Alpha-N-acetylgalactosaminide alpha-2,6-sialyltransferase 2	FFWLLLLLTAACSGLLFALYF	74841		62143

Table S2. (continued)

Uniprot Entries					Reconstructed Fluorescence		
#	**AC**	**Uniprot ID**	**Protein Name**	**TM helix sequence**	**21 AA**	**19 AA**	**17 AA**
93	Q9UJ90	KCNE5_HUMAN	Potassium voltage-gated channel subfamily E regulatory beta subunit 5	LYILLIMIFYACLAGGLILAY		48028	36338
94	Q9ULG6	CCPG1_HUMAN	Cell cycle progression protein 1	CVILALVIAISMGFGHFYGTI	72405	68118	72848
95	Q9ULI3	HEG1_HUMAN	Protein HEG homolog 1	ITVVIAAAGGGLLLILGIALI	48180		25698
96	Q9UN70	PCDGK_HUMAN	Protocadherin gamma-C3	LLLSLILVSVGFVVTVFGVII	76105	43692	39749
97	Q9Y287	ITM2B_HUMAN	Integral membrane protein 2B	WCMCFGLAFMLAGVILGGAYL	59148		46588
98	Q9Y3D6	FIS1_HUMAN	Mitochondrial fission 1 protein	LVGMAIVGGMALGVAGLAGLI	44801		45404
99	Q9Y5F7	PCDGL_HUMAN	Protocadherin gamma-C4	LAVSLVAICFVSFGSFVALLS	65006	45907	26795
100	Q9Y6J6	KCNE2_HUMAN	Potassium voltage-gated channel subfamily E member 2	VILYLMVMIGMFSFIIVAILV			42183

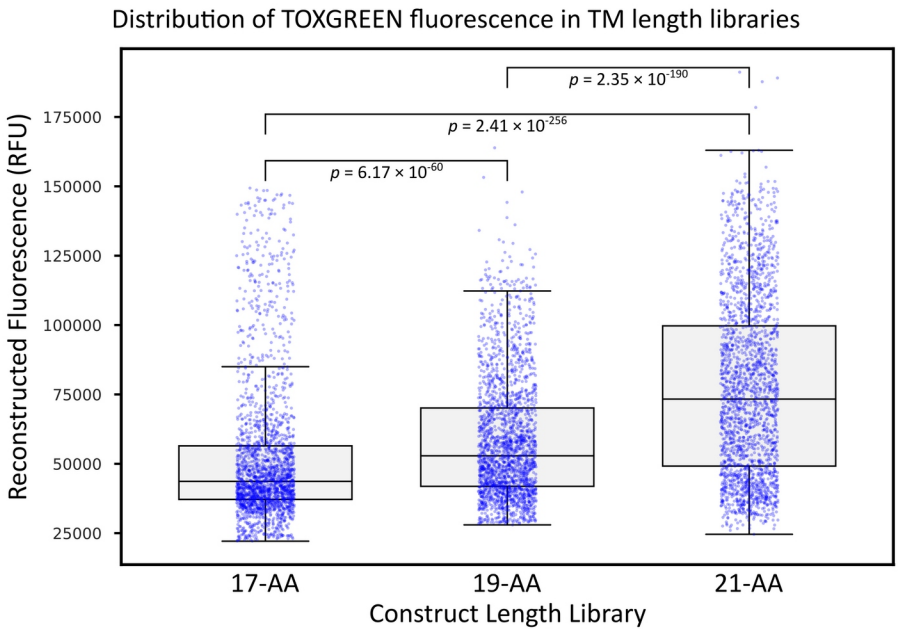

Fig. S2. Distribution of reconstructed fluorescent signal for all constructs from libraries with different TM helix length. Statistical significance calculated with the Wilcoxon Signed-Rank Test.

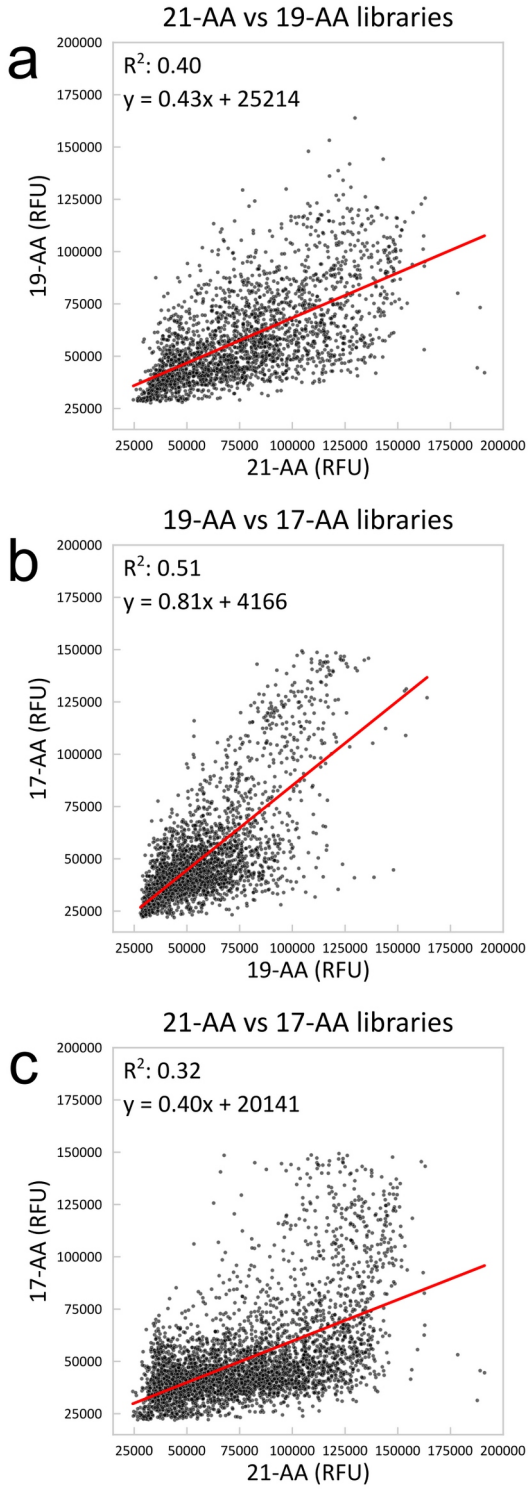

Fig. S3. Correlation of the reconstructed fluorescence in the three length libraries. a) 21-AA vs 19-AA libraries. b) 19-AA vs 17-AA libraries. c) 21-AA vs 17-AA libraries.

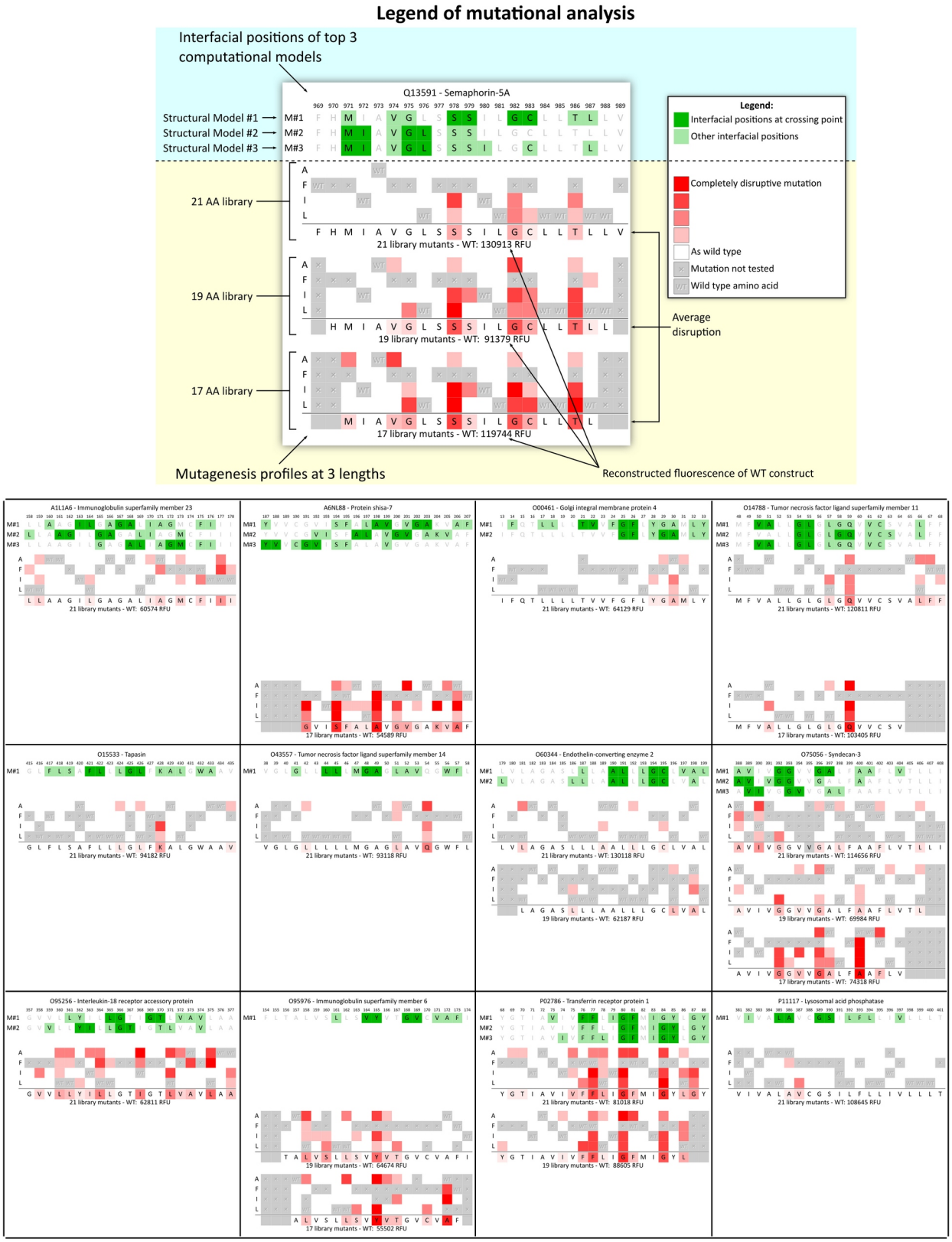

Fig. S4. Mutational profiles of the predicted GAS_{right} constructs compared with the pattern of interfacial position predicted by CATM top three structural models. The mutational profiles were not computed for any constructs and lengths if the wild-type reconstructed fluorescence fell under a minimum threshold of 50,000 RFU.

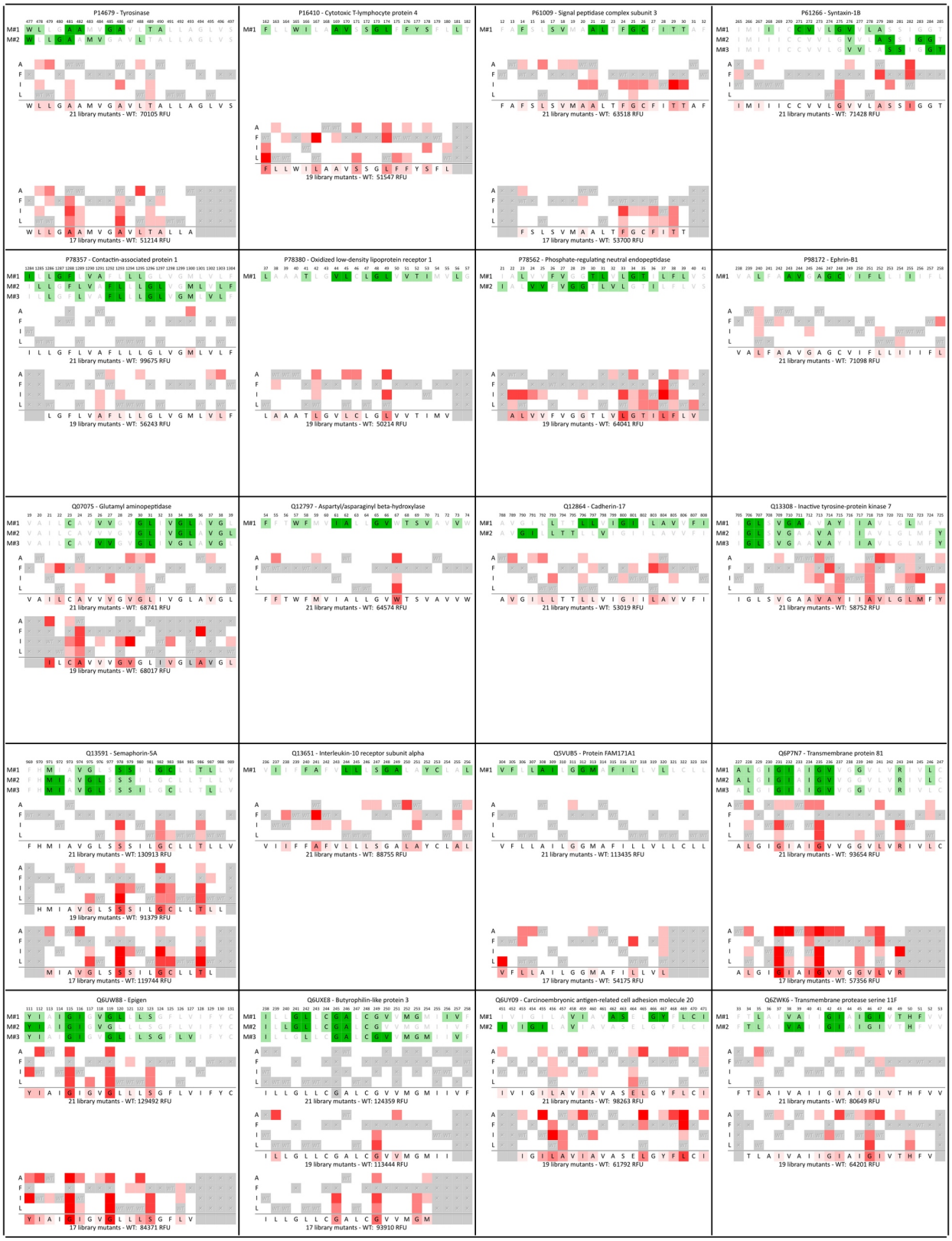

Fig. S4. (continued)

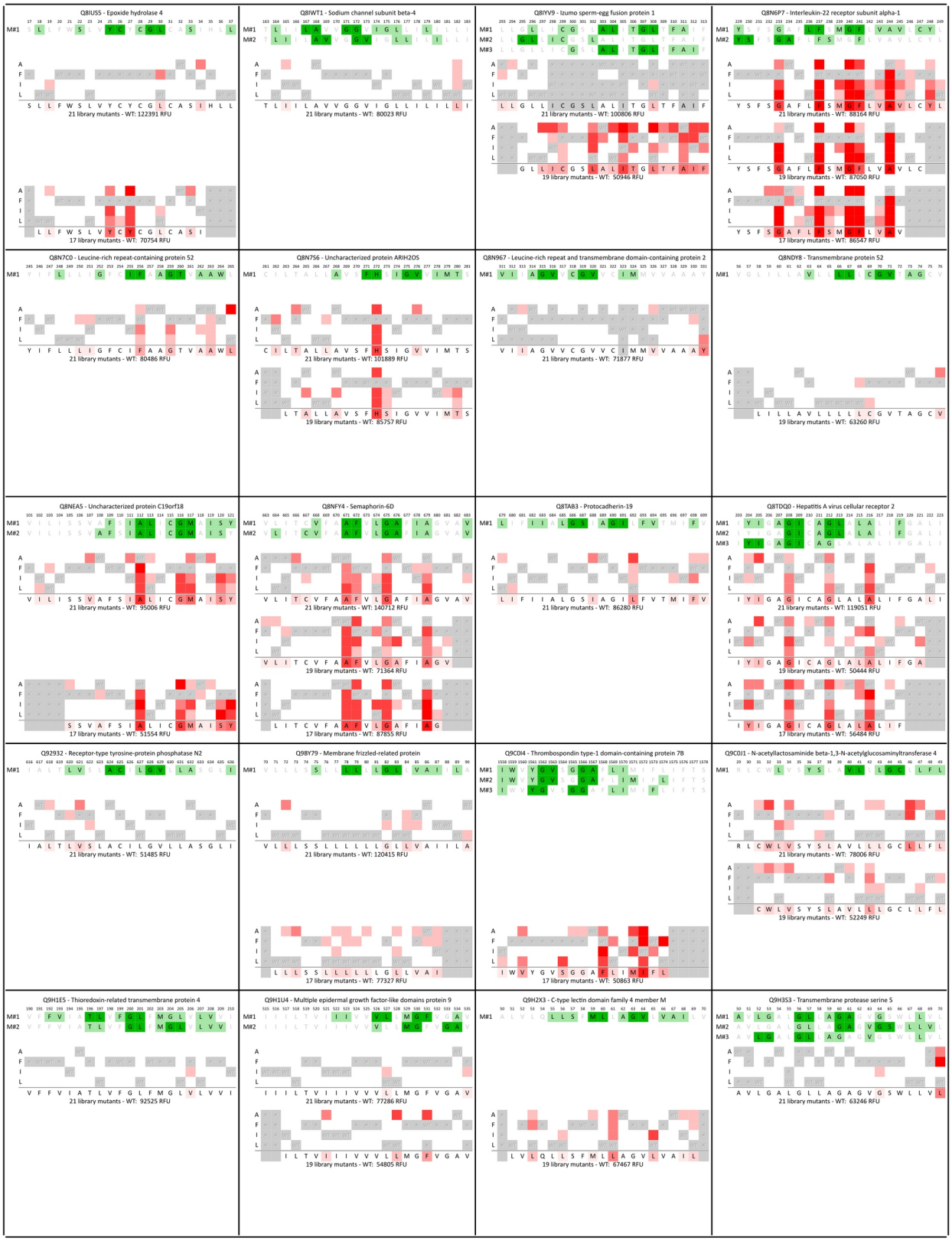

Fig. S4. (continued)

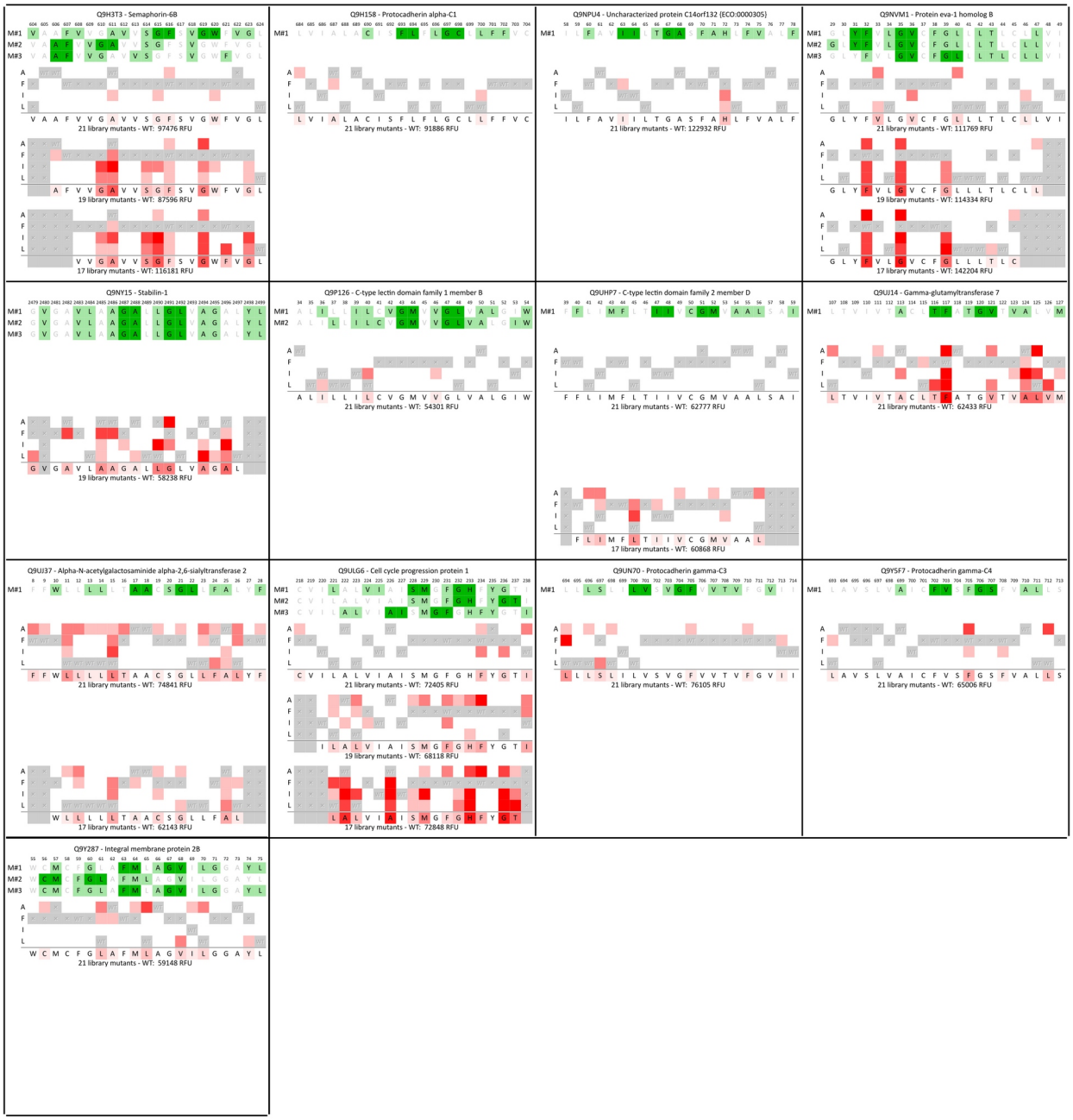

Fig. S4. (continued)

Table S3. Summary of the classes of mutagenesis profiles. The number of predictions that are similar between CATM and AlphaFold3 are also reported.

Mutagenesis result	**Total**	**AF3 and CATM match[1]**	
Disruption matches CATM predicted interface	12	9	(75%)
Disruption maps to an alternative interface	13	0	(0%)
Disruption does not map to a clear interface	6	1	(17%)
Single polar disruptive position	3	0	(0%)
No or minimal disruption observed	31	1	(3%)
TOXGREEN signal too low for mutagenesis	35	1	(3%)

[1]AF3 and CATM models are considered matching if their RMSD is under 2.2 Å

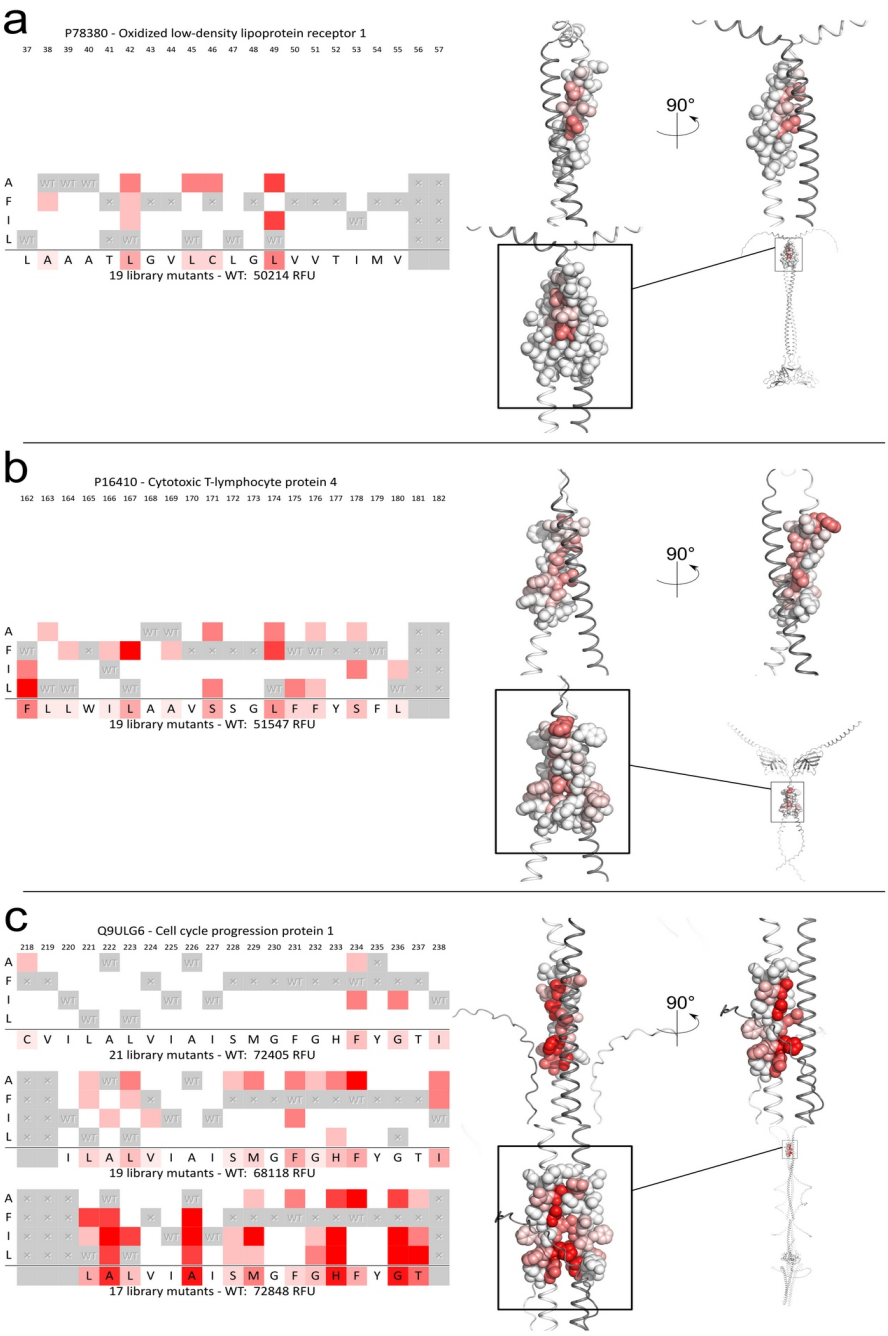

Fig. S5. Mutagenesis profiles that match with an AlphaFold3 model. The three mutational profiles did not match the GAS_{right} configuration predicted by CATM but the interface highlighted by the mutagenesis is compatible with the dimeric AF3 structure.

Table S4. Series of steps that led to the selection of 100 predicted GAS_{right} constructs for the mutational analysis.

Step	Constructs left
Human Single Pass Proteins Annotated in Uniprot	2,383
Potential GAS_{right} structures predicted by CATM	1141
TM sequences that do not contain proline residues	609
TM sequences with annotated sequence length of 21 amino acids	454
CATM energy score below -10 kcal/mol	363
Structural models without significant interactions in juxtamembrane region	122
Sequences with not more than one polar residue	110
Sequences selected for experimental analysis (top 100 energy scorer)	100

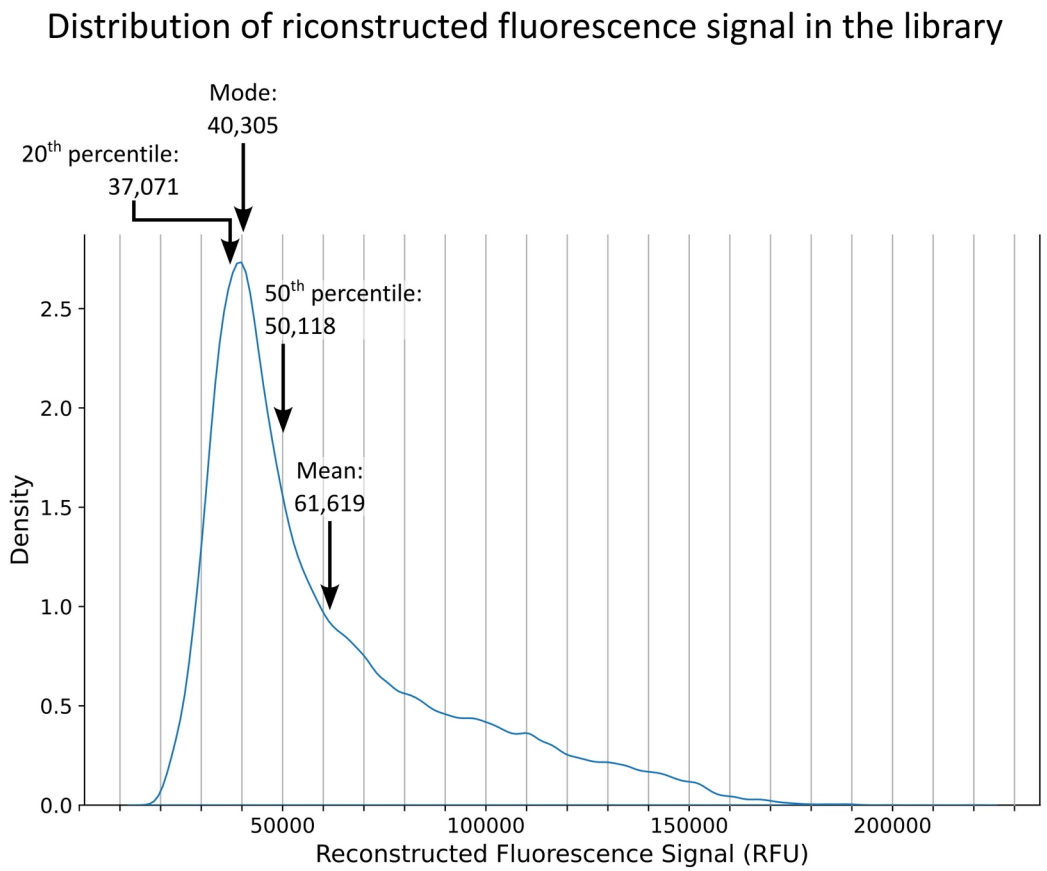

Fig. S6. Distribution of reconstructed fluorescence in all constructs of the library. The distribution informed the choices for thresholds for the mutational analysis. A minimum threshold of 50,000 RFU (50th percentile) was set for performing the mutagenesis analysis to guarantee sufficient dynamic range. A threshold of 35,000 RFU (~ 20th percentile) was chosen as the monomeric baseline threshold for the classification in the mutational analysis (see Methods).