Molecular basis of tail-anchored integral membrane protein recognition by the cochaperone Sgt2

Ku-Feng Lin, Michelle Y. Fry‡ , Shyam M. Saladi‡ , and William M. Clemons Jr*
From the Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California, USA

Edited by Ursula Jakob

The targeting and insertion of tail-anchored (TA) integral membrane proteins (IMPs) into the correct membrane is critical for cellular homeostasis. The fungal protein Sgt2, and its human homolog SGTA, is the entry point for clients to the guided entry of tail-anchored protein (GET) pathway, which targets endoplasmic reticulum-bound TA IMPs. Consisting of three structurally independent domains, the C terminus of Sgt2 binds to the hydrophobic transmembrane domain (TMD) of clients. However, the exact binding interface within Sgt2 and molecular details that underlie its binding mechanism and client preference are not known. Here, we reveal the mechanism of Sgt2 binding to hydrophobic clients, including TA IMPs. Through sequence analysis, biophysical characterization, and a series of capture assays, we establish that the Sgt2 C-terminal domain is flexible but conserved and sufficient for client binding. A molecular model for this domain reveals a helical hand forming a hydrophobic groove approximately 15 Å long that is consistent with our observed higher affinity for client TMDs with a hydrophobic face and a minimal length of 11 residues. This work places Sgt2 into a broader family of TPR-containing cochaperone proteins, demonstrating structural and sequence-based similarities to the DP domains in the yeast Hsp90 and Hsp70 coordinating protein, Sti1.

An inherently complicated problem of cellular homeostasis is the biogenesis of hydrophobic integral membrane proteins (IMPs), which are synthesized in the cytoplasm and must be targeted and inserted into a lipid bilayer. Accounting for ~25% of transcribed genes [1], IMPs are primarily targeted by cellular signal-binding factors that recognize a diverse set of hydrophobic α-helical signals as they emerge from the ribosome [2–4]. One important class of IMPs are tail-anchored (TA) proteins whose hydrophobic signals are their single helical transmembrane domain (TMD) located near the C terminus and are primarily targeted posttranslationally to either the endoplasmic reticulum (ER) or mitochondria [5–9]. In the case of the canonical pathway for ER-destined TA IMPs, each is first recognized by homologs of mammalian SGTA (small glutamine tetratricopeptide repeat protein alpha) [4, 6, 10, 11]. Common to all signal-binding factors is the need to recognize, bind, and then hand off a hydrophobic helix. How such factors can maintain specificity to a diverse set of hydrophobic clients that must subsequently be released remains an important question.

Homologs of *Saccharomyces cerevisiae* Sgt2 (small glutamine-rich tetratricopeptide repeat-containing protein 2, refered to here as ySgt2) and *Homo sapiens* SGTA (referred to here as hSgt2 and collectively Sgt2 for simplicity) are involved in a variety of cellular processes regarding the homeostasis of membrane proteins including the targeting of TA IMPs [9, 12–14], retrograde transport of membrane proteins for ubiquitination and subsequent proteasomal degradation [15], and regulation of mislocalized membrane proteins (MLPs) [16, 17]. Among these, the role of Sgt2 in the primary pathways responsible for targeting TA clients to the ER is best characterized, *i.e.*, the fungal Guided Entry of Tail-anchored proteins (GET) or the mammalian Transmembrane Recognition Complex (TRC) pathway. In the GET pathway, Sgt2 functions by binding a cytosolic TA client, then transferring the TA client to the ATPase chaperone Get3 (human homolog is also Get3) with the aid of the heteromeric Get4/Get5 complex (human Get4/Get5/Bag6 complex) [13, 18–20]. In this process, TA client binding to Sgt2, after hand-off from Hsp70, is proposed as the first committed step to ensure that ER TA clients are delivered to the ER membrane while mitochondrial TA clients are excluded [3, 13, 21]. Subsequent transfer of the TA client from Sgt2 to the ATP-bound Get3 induces conformational changes in Get3 that trigger ATP hydrolysis, releasing Get3 from Get4 and favoring binding of the Get3-TA client complex to the Get1/2 receptor at the ER leading to release of the TA client into the membrane [22–26]. Deletions of yeast GET genes (*i.e.*, *get1Δ*, *get2Δ*, or *get3Δ*) cause cytosolic aggregation of TA clients dependent on Sgt2 [26, 27].

In addition to targeting TA IMPs, there is evidence that hSgt2 promotes degradation of IMPs through the proteasome by cooperating with the Bag6 complex, a heterotrimer containing Bag6, hGet4, and hGet5, which acts as a central hub for a diverse physiological network related to protein targeting and quality control [19, 28–30]. The Bag6 complex can associate with ER membrane-embedded ubiquitin regulatory protein UbxD8, transmembrane protein gp78, proteasomal component Rpn10c, and an E3 ubiquitin protein ligase RNF126, thereby connecting hSgt2 to ER-associated degradation (ERAD) and proteasomal activity. Depletion of hSgt2 significantly inhibits turnover of ERAD IMP clients and elicits the unfolded protein response

‡ These authors contributed equally to this work.
* For correspondence: William M. Clemons Jr., clemons@caltech.edu.

Characterization of the client-binding domain of Sgt2

(16). Furthermore, the cellular level of MLPs in the cytoplasm could be maintained by coexpression with hSgt2, which possibly antagonizes ubiquitination of MLPs to prevent proteasomal degradation (15, 17). These studies demonstrate an active role of hSgt2 in triaging IMPs in the cytoplasm and the breadth of hSgt2 clients including TA IMPs, ERAD, and MLPs all harboring one or more TMDs. Roles for hSgt2 in disease include polyomavirus infection (31), neurodegenerative disease (27, 32), hormone-regulated carcinogenesis (33, 34), and myogenesis (35), although the underlying molecular mechanisms are still unclear.

The architecture of Sgt2 includes three structurally independent domains that define the three different interactions of Sgt2 (Fig. 1A) (12, 36–39). The N-terminal domain forms a homodimer composed of a four-helix bundle with twofold symmetry that primarily binds to the ubiquitin-like domain (UBL) of Get5/Ubl4A for TA IMP targeting (36, 40) or interacts with the UBL on

the N-terminal region of Bag6 (41) where it is thought to initiate downstream degradation processes (15, 28, 29). The central region comprises a cochaperone domain with three repeated TPR motifs arranged in a right-handed superhelix forming a "carboxylate clamp" for binding the C terminus of heat-shock proteins (HSP) (12, 42). The highly conserved TPR domain was demonstrated to be critical in modulating propagation of yeast prions by recruiting HSP70 (27) and may associate with the proteasomal factor Rpn13 to regulate MLPs (43). More recently, it was demonstrated that mutations to residues in the TPR domain, which prevent Hsp70 binding, impair the loading of TA clients onto ySgt2 (21), consistent with a direct role of Hsp70 in TA IMP targeting *via* the TPR domain. The C-terminal methionine-rich domain of Sgt2 is responsible for binding to hydrophobic clients such as TA IMPs (11, 37, 44). Other hydrophobic segments have been demonstrated to interact with

Figure 1. Structural characteristics of free Sgt2 C domain. *A, top,* schematic of the domain organization of Sgt2. *Below,* representative sequences from a large-scale multiple sequence alignment of the C domain: fungal Sgt2 from *S. cerevisiae*, *S. pombe*, and *C. thermophilum* and metazoan Sgt2 from *C. savignyi*, *X. laevis*, and *H. sapiens*. Protease susceptible sites on ySgt2-C identified by mass spectrometry are indicated by *red arrowheads*. Predicted helices of ySgt2 (*blue*) and hSgt2 (*orange*) by Jpred (83) and/or structure prediction are shown. *Blue/orange* color scheme for ySgt2/hSgt2 is used throughout the text. Residues noted in the text are highlighted by an *asterisk*. *B,* overlay of size-exclusion chromatography traces of ySgt2-C (*blue line*), hSgt2-C (*orange line*), ySgt2-TPR (*blue dash*), and hSgt2-TPR (*orange dash*). Traces are measured at 214 nm, baseline-corrected, and normalized to the same peak height. *C,* Far UV CD spectrum of 10 μM of purified ySgt2-C (*blue*) and hSgt2-C (*orange*) at RT with secondary structure decomposition from BestSel (68). *D,* ^{1}H-^{15}N HSQC spectrum of ySgt2-C at 25 °C. The displayed chemical shift window encompasses all N-H resonances from both backbone and side chains. The range of backbone amide protons, excluding possible side-chain NH$_2$ of Asn/Gln, is indicated by pairs of *red dashed lines*. *E,* as in *D* for hSgt2-C at 25 °C.

this domain such as the membrane protein Vpu (viral protein U) from human immunodeficiency virus type-1 (HIV-1), the TMD of tetherin (44), the signal peptide of myostatin (35), and the N domain of the yeast prion forming protein Sup35 (27). All of these studies suggest that the C terminus of Sgt2 binds broadly to hydrophobic stretches, yet structural and mechanistic information for client recognition is lacking.

In this study, we provide the first structural characterization of the C domains from Sgt2 (Sgt2-C) and show that, in the absence of substrate, it is relatively unstructured. We demonstrate that a conserved region of the C domain, defined here as C_{cons}, is sufficient for client binding. Analysis of the C_{cons} sequence identifies six amphipathic helices whose hydrophobic residues are required for client binding. Based on this, we computationally generate an *ab initio* structural model that is validated by point mutants and disulfide cross-linking. Artificial clients are then used to define the properties within clients critical for binding to Sgt2-C. The results show that Sgt2-C falls into a larger STI1 family of TPR-containing cochaperones and allow us to propose a mechanism for client binding.

Results

The flexible Sgt2-C domain

Based on sequence alignment (Fig. 1A), the Sgt2-C contains a conserved core of six predicted helices flanked by unstructured loops that vary in length and sequence. Previous experimental work suggested that this region is particularly flexible, as this domain in the *Aspergillus fumigatus* homolog is sensitive to proteolysis (12). Similarly, for ySgt2-TPR-C, the sites sensitive to limited proteolysis primarily occur within the loops flanking the conserved helices (Fig. 1A, *red arrows* and Fig. S1B). This flexible nature of the C-domain likely contributes to its anomalous passage through a gel-filtration column where Sgt2-C elutes much earlier than the similarly sized, but well-folded, Sgt2 TPR domain (Fig. 1B), as is typical for unstructured proteins [REF]. The larger hydrodynamic radius matches previous small-angle X-ray scattering measurement of the ySgt2 TPR-C domain that indicated a partial unfolded characteristic in a Kratky plot analysis. The circular dichroism (CD) spectra for both homologs suggest that the C domain and a predicted six α-helical methionine-rich region of Sgt2-C (Fig. 1A), hereafter referred to as Sgt2-C_{cons}, largely assume a random-coil conformation, with 40 to 45% not assignable to a defined secondary structure category (Fig. 1C, Fig. S1A) (45). The well-resolved, sharp, but narrowly dispersed chemical shifts of the backbone amide protons in ^1H-^{15}N HSQC spectra of Sgt2-C (Fig. 1, D and E) and Sgt2-C_{cons} (Fig. S1, B and C) indicate a significant degree of backbone mobility, similar to natively unfolded proteins (46) and consistent with results seen by others (47), further highlighting the lack of stable tertiary structure (12). Taken all together, Sgt2-C appears to be a flexible domain.

The conserved region of the C domain is sufficient for substrate binding

We then asked if the flexible Sgt2-C is the site of client binding in the cochaperone and if so, where within this domain is the binding region. During purification Sgt2-C is susceptible to proteolytic activity being cut at several specific sites (Fig. 1A). Proteolysis occurred primarily at Leu_{327} and in the poorly conserved N-terminal region (between Asp_{235}-Gly_{258}). Given the intervening region (ySgt2 G258-L327) is conserved (Fig. 1A), it and the corresponding region in hSgt2 may mediate client binding (Fig. 2A, *gray*). To test this, we established a set of His-tagged Sgt2 constructs of various lengths (Fig. 2C, Fig. S2A). These Sgt2-C truncations were coexpressed with an MBP-tagged client, Sbh1, and binding was detected by the presence of captured TA clients in nickel elution fractions (Fig. 2B). The TA protein Sbh1 is the yeast homolog of the mammalian Sec61-gamma, a component of the ER-resident Sec translocon. While the relative efficiency of MBP-Sbh1 capture cannot be assessed in this assay due to differences in total protein levels (Fig. S2B), we can demonstrate the ability of a given construct to bind to the client. As previously seen (13), we confirm that Sgt2-TPR-C alone is sufficient for capturing a client (Fig. 2C). As one might expect, the C domain was also sufficient for binding the client. Interestingly, Sgt2-C_{cons} is sufficient for binding to Sbh1. Even a minimal region of the last five helices (referred to as ΔH0) also captures Sbh1 (Fig. 2C). The predicted helices in Sgt2-C_{cons} are amphipathic and their hydrophobic faces may be used for client binding (Fig. 2D).

Each of the six helices in Sgt2-C_{cons} was mutated to replace the larger hydrophobic residues with alanines, dramatically reducing the overall hydrophobicity. For all of the helices, alanine replacement of the hydrophobic residues significantly reduces binding of Sbh1 to Sgt2-C (Fig. 2, E and F). While these mutants expressed at similar levels to the wild-type (WT) sequence, one cannot rule out that some of these changes may affect the tertiary structure of this domain. In general, these results imply that these amphipathic helices are necessary for client binding since removal of the hydrophobic faces disrupts binding. The overall effect on binding by each helix is different, with mutations in helices 1 to 3 having the most dramatic reduction in binding suggesting that these are more crucial for Sgt2–client complex formation. It is also worth noting, as this is a general trend, that hSgt2 is more resistant to mutations that affect binding (Fig. 2F) than ySgt2, which likely reflects different thresholds for binding.

Molecular modeling of Sgt2-C domain

Despite the need for a molecular model, the C domain has resisted structural studies, likely due to the demonstrated inherent flexibility. Based on the six conserved α-helical amphipathic segments (Fig. 1A) that contain hydrophobic residues critical for client binding (Fig. 2, D and E), we expect some folded structure to exist. Therefore, we performed *ab initio* molecular modeling of Sgt2-C using a variety of prediction methods resulting in a diversity of putative structures (48–52). As expected, all models showed buried hydrophobic residues as this is a major criterion for *in silico* protein folding. Residues outside the ySgt2-C_{cons} region adopted varied conformations consistent with their expected higher flexibility.
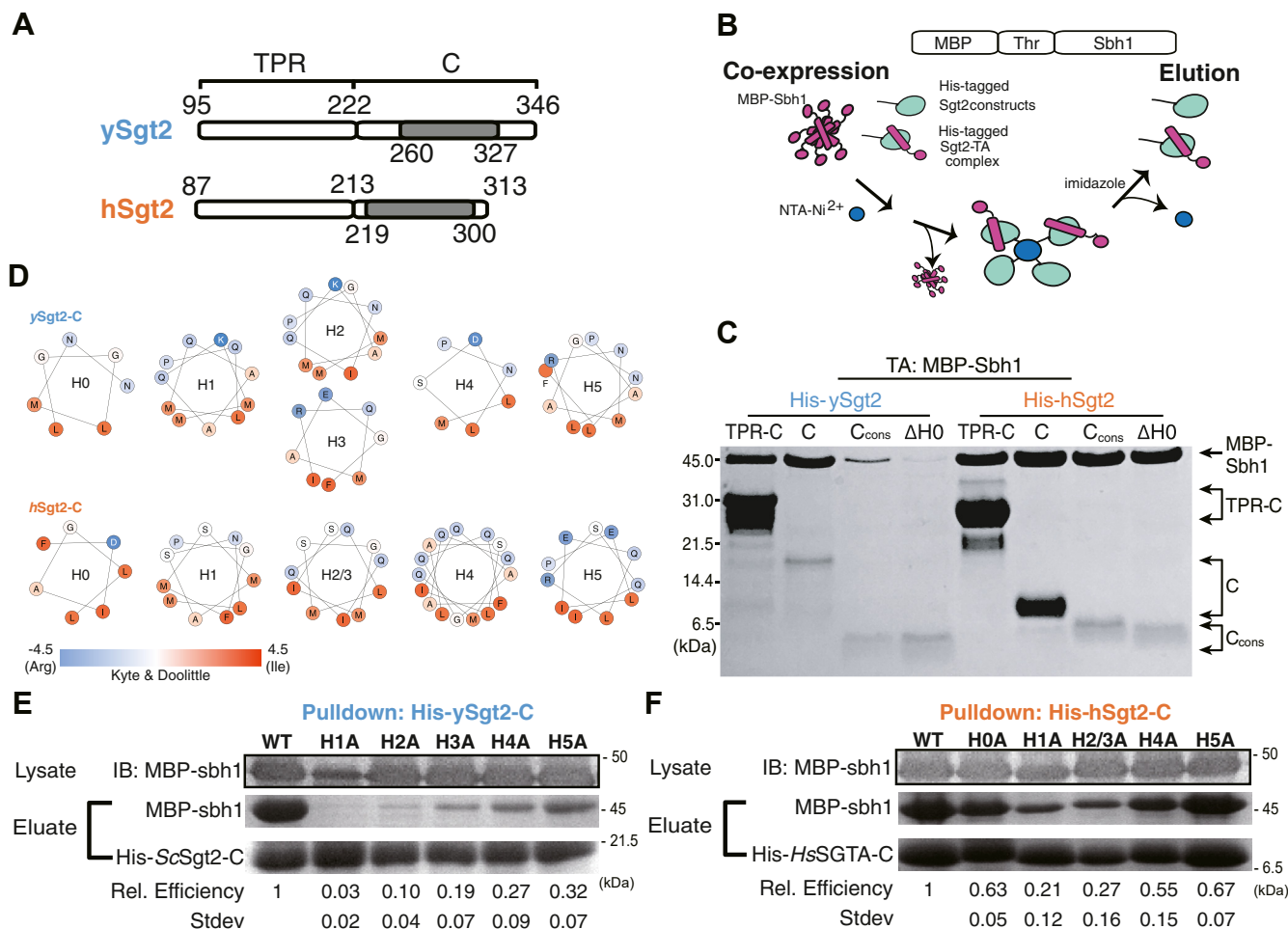
Figure 2. The minimal binding region of Sgt2 for client binding. *A*, diagram of the protein truncations tested for client binding that include the TPR-C domain, C-domain (C), C$_{cons}$, and C$_{cons}$ ΔH0 (ΔH0) from ySgt2 and hSgt2. The residues corresponding to each domain are indicated, and *gray blocks* highlight the C$_{cons}$ region. *B*, schematic of capture experiments of MBP-tagged Sbh1 separated by a thrombin (Thr) cleavage site (MBP-Sbh1) by Sgt2 variants. After coexpression, cell pellets are lysed and NTA-Ni^{2+} is used to capture His-tagged Sgt2-TPR-C. *C*, Tris-Tricine-SDS-PAGE gel (84) of coexpressed and purified MBP-Sbh1 and His-tagged Sgt2 truncations visualized with Coomassie Blue staining. *D*, helical wheel diagrams of predicted helices (see Fig. 1A) in the C$_{cons}$ domain of ySgt2 and hSgt2. Residues are colored by the Kyte and Doolittle hydrophobicity scale (85). *E*, all of the hydrophobic residues (L, I, F, and M) in a predicted helix (H0, H1, etc.) are replaced with alanines and tested for the ability to capture MBP-Sbh1. Protein levels were quantified by Coomassie staining. Relative binding efficiency of MBP-Sbh1 by ySgt2 C domain (ySgt2-C) variants was calculated relative to total amount of ySgt2-C captured (MBP-Sbh1/Sgt2-C), then normalized to the wild-type ySgt2-C. Experiments were performed 3 to 4 times and the standard deviations are presented. Total expression levels of the MBP-Sbh1 were similar across experiments as visualized by immunoblotting (IB) of the cell lysate. *F*, as in *E* but for hSgt2.

Pruning these N- and C-terminal regions to focus on the ySgt2-C$_{cons}$ region (Fig. S3A) revealed a potential binding interface for a hydrophobic substrate. Examples are seen in Quark models (1, 4, and 6 shown), Robetta 1 and 2, and I-TASSER 2 and 3, whereas other models had no clearly distinguishable groove. Given the intrinsic flexibility of the Sgt2-C domain, it is possible that models without a groove are found in the non-TMD-bound structural ensemble.

For a working model of TMD-bound ySgt2-C, we chose the highest scoring Quark structures where a general consistent architecture is seen (Fig. 3A) (48). The overall model contained a potential client-binding site, a hydrophobic groove formed by the amphipathic helices. The groove is approximately 15 Å long, 12 Å wide, and 10 Å deep, which is sufficient to accommodate three helical turns of an α-helix, ~11 amino acids (Fig. 3B).

To validate the model, we interrogated the accuracy of the predicted structural arrangement by determining distance

constraints from cross-linking experiments. We selected four pairs of residues in close spatial proximity and one pair far apart based on the Quark models (Fig. 4A). Calculating a C$_β$–C$_β$ distance between residue pairs for each model (Fig. 4F), the Quark models 1 and 3 were the most consistent with an expected distance of 9 Å or less for the close pairs. In all alternative models, the overall distances are much larger and should not be expected to form disulfide bonds *in vitro* if they represent a TMD-bound state. For Robetta, a number of the models have pairs of residues within 9 Å, and Robetta's per-residue error estimate suggests relatively high confidence in the C$_{cons}$ region (Fig. S3B).

As a control, we first confirmed that the cysteine–mutant pairs do not affect the function of ySgt2. We utilized an *in vitro* capture assay where a yeast Hsp70 homolog Ssa1 loaded with a TA client, Bos1, delivers the client to ySgt2 (21, 49, 50) (Fig. 4C). Purified Ssa1 is mixed with detergent-solubilized strep-tagged Bos1-TMD (a model ER TA client)
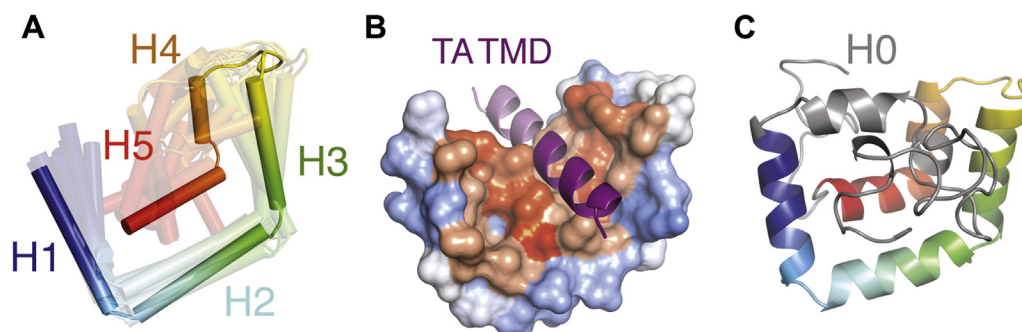
Figure 3. A structural model for Sgt2-C$_{cons}$. *A*, the top ten models of the ySgt2-C$_{cons}$ generated by the template-free algorithm Quark (48) are overlaid with the highest scoring model in solid. Models are color-ramped from N- (*blue*) to C terminus (*red*). *B*, a model of ySgt2-C$_{cons}$ (surface colored by Kyte–Doolittle hydrophobicity) bound to a TMD (*purple helix*) generated by rigid-body docking through Zdock (80). The *darker purple* corresponds to an 11-residue stretch. *C*, the entire ySgt2-C from the highest scoring model from Quark (C$_{cons}$ in rainbow with the rest in *gray*) highlighting H0 and the rest of the flexible termini that vary considerably across models.

that contained a p-benzoyl-l-phenylalanine (BPA)-labeled residue, Bos1$_{BPA,}$ and diluted to below the critical micelle concentration resulting in soluble complexes of Bos1$_{BPA}$/Ssa1. Full-length ySgt2 variants were each tested for the ability to capture Bos1$_{BPA}$ from Ssa1. After the transfer reaction, each was UV-treated to generate Bos1 cross-links. Successful capture of the TA clients by ySgt2 was detected for all cysteine variants using an anti-strep western blot and the appearance of a Bos1$_{BPA}$/ySgt2 cross-link band, suggesting that the mutations do not affect the structure or function of ySgt2 (Fig. 4*C*).

We and others have demonstrated that a monomeric Sgt2 is sufficient for binding to clients (13). For the distance experiment, each of the cysteine–mutant pairs was made in the more stable monomeric variant ySgt2-TPR-C. Each variant was coexpressed with an artificial client—a cMyc-tagged BRIL (small, four-helix bundle protein used in previous work to aid in the crystallization of GPCRs (51)) with a C-terminal TMD consisting of eight leucines and three alanines, denoted as 11 [L8], and purified *via* nickel-affinity chromatography in reducing buffer (Fig. S4*A*). All of the ySgt2 mutants bound to the client and behaved similar to the WT (cysteine-free) further suggesting that the mutants did not perturb the native structure (Fig. S4*B*). For disulfide cross-link formation, each eluate was oxidized, digested using the protease Glu-C, and cross-links were identified by the visualization of a reducing-agent sensitive ~7.7 kDa fragment in gel electrophoresis (Fig. 4*D*). For both the WT construct and in N285C/G329C, where the pairs are predicted from the Quark models to be too distant for disulfide bond formation, no higher-molecular-weight band was observed. For the remaining pairs that are predicted to be close enough for bond formation, the 7.7 kDa fragment was observed in each case and is labile in reducing conditions. Again, these results support the C$_{cons}$ model derived from Quark.

With the four cross-linked pairs as distance constraints, new models were generated using Robetta with a restraint on the corresponding pairs of C$_\beta$ atoms less than 9 Å (Fig. S5*A*). The Robetta models from these runs are similar to the top scoring models from Quark (Fig. 3). Satisfyingly, the pair of residues that do not form disulfide cross-links are generally consistent (Fig. S5*B*).

The improvement of the ySgt2 models predicted by Robetta with restraints included encouraged us to generate models for hSgt2-C with constraints. For this, pairs were defined based on sequence alignments of Sgt2 (Fig. 1*A*) and used as restraints. The resulting predictions had architectures consistent with the equivalent regions predicted for ySgt2-C$_{cons}$, for example, Robetta 4 (Fig. S5*C*, top). Although in general the predicted hSgt2 model is similar to that for ySgt2, the region that corresponds to H2 occupies a position that precludes a clear hydrophobic groove. For ySgt2, the longer N-terminal loop occupies the groove preventing the exposure of hydrophobics to solvent (Fig. 3*C*, *gray*). For hSgt2, the shorter N-terminal loop may not be sufficient to similarly occupy the groove and allow for the clear hydrophobic hand seen for the ySgt2-C. To correct for this, we replaced the sequence of the N-terminal loop of hSgt2-C with the ySgt2-C loop and ran structure prediction with the pairwise distance restraints. This resulted in a model where the loop occupies the groove and, when pruned away, suggests the hydrophobic hand seen in yeast (Fig. S5*C*, middle boxed). Of note, we also generated models of hSgt2-C using the most recent Robetta method (transform-restrained), which produces new structures with a groove and similar helical-hand architecture across the board (Fig. S5*C*, bottom).

We sought to further test the robustness of our model considering the intrinsic flexibility of Sgt2-C by probing for disulfide bond formation with neighboring residues of one of our cross-linking pairs. While the C$_\beta$–C$_\beta$ distance puts these adjacent pairs at farther than 9 Å, mutating residues to cystines and measuring S–S distances across all possible pairs of rotamers provide a wider interval on possible distances and, therefore, the likelihood that a disulfide bond will form (Fig. 4*E*). Cysteine mutants were introduced to the residues adjacent to M289 and A319 in ySgt2-TPR-C resulting in four additional pairs: K288C/A319C, M290C/A319C, M289C/P318C, and M289C/L320C. As described previously, these mutants were coexpressed with a substrate, in this case the cMyc-tag was replaced with an MBP-tag. The MBP-tag on the artificial client allows for tandem amylose- and nickel-affinity chromatography to ensure eluates contained only Sgt2-TPR-C bound to substrate. Disulfide bond formation was conducted
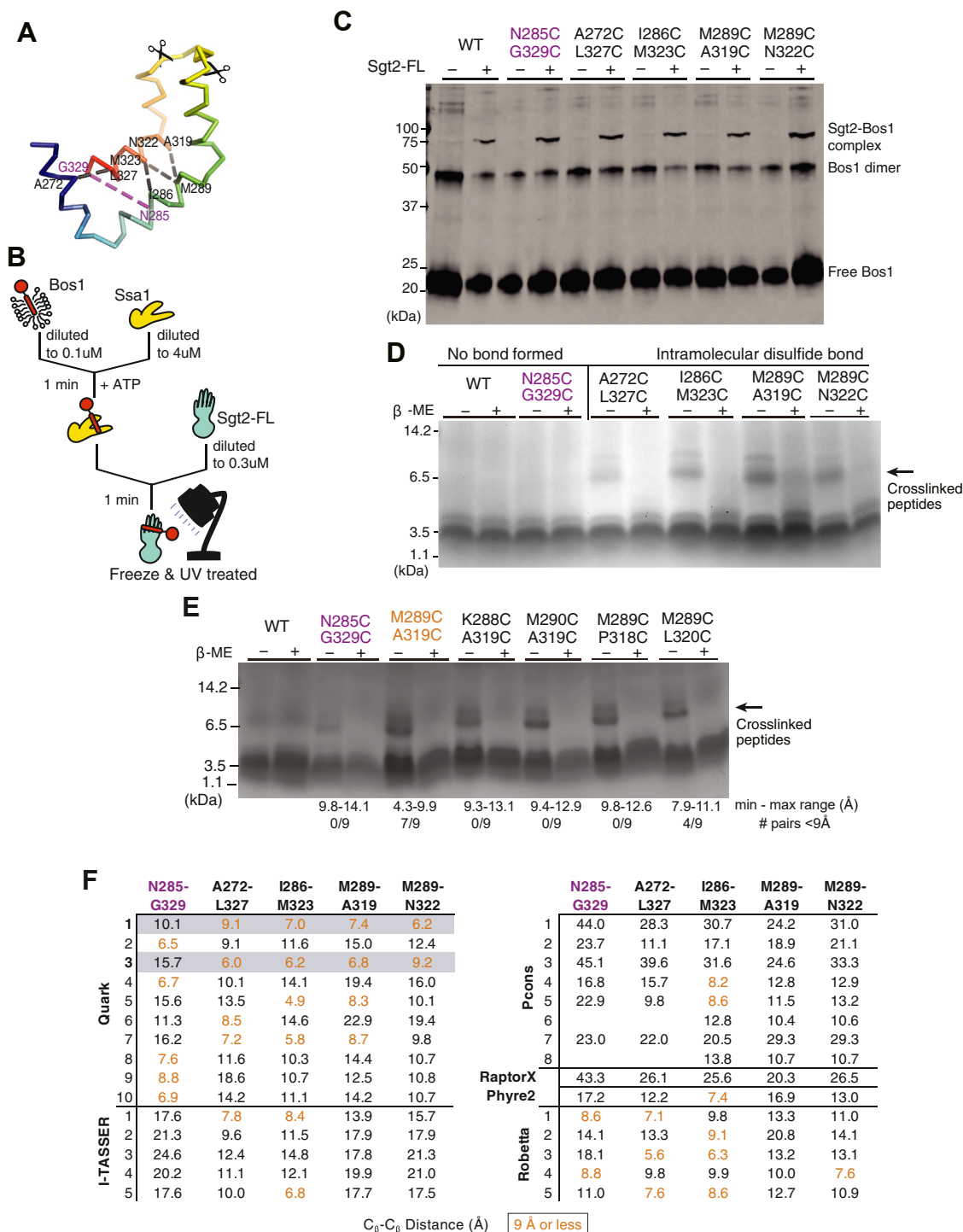
Figure 4. Validating the structural model with disulfide bond formation. Variants of His-ySgt2-TPR-C (WT or cysteine double mutants) were coexpressed with the artificial client, cMyc-BRIL-11[L8]. After lysis, ySgt2-TPR-C proteins were purified, oxidized, then digested by Glu-C protease, and analyzed by gel in either nonreducing or reducing buffer. *A*, Cα ribbon of ySgt2-C_{cons} color-ramped with various pairs of cysteines highlighted. *Scissors* indicate protease cleavage sites resulting in fragments less than 3 kDa in size. *B*, a schematic of the transfer of Bos1$_{BPA}$ from Ssa1 to full-length ySgt2 to demonstrate that the double cysteine mutants are still functional. *C*, a western blot visualizing cross-linked ySgt2-Bos1 complexes. All samples tested, WT, N285C/G329C, A272C/L327C, I286C/M323C, M289C/A319C, and M289C/N322C, had a higher molecular weight appearing after the addition of ySgt2, which corresponds to the size of the cross-linked complex (86). *D*, for the WT (cys-free), no significant difference was found between samples in nonreducing *versus* reducing conditions. All close residue pairs (A272/L327, I286/M323, M289/A319, and M289/N322) show peptide fragments (higher MW) sensitive to the reducing agent and indicate disulfide bond formation (indicated by *arrow*). A cysteine pair (N285/G329) predicted to be far apart by the model does not result in the higher MW species. *E*, tris-Glycine SDS-PAGE gel probing the flexibility of ySgt2-C_{cons}. All new pairs (K288/A319, M290/A319, M289/P318, M289/L320) show peptide fragments sensitive to the reducing agent (indicated by *arrow*). The range of distances of the eight closest possible rotamer pairs is annotated below. The cysteine pair (N285/G329) shown to be far apart by the model does have a faint higher-molecular-weight band. *F*, $C_β$–$C_β$ distances between the residues mutated to cysteines based on various models predicted by the Quark, I-TASSER, PCONS, and Robetta. Cysteine pairs that are 9 Å or less colored in orange and are expected to be close enough to form disulfide bonds. Where all five pair distances are consistent with the experiment (four near and one far), the row is shaded in *gray*.

as before, and a reductant sensitive band at 7.7 kDa is observed for each of these adjacent pairs. While the geometry of each of these C–C pairs might suggest against disulfide bond formation, given the intrinsic flexibility of Sgt2-C, it is not surprising that each of these pairs is able to form disulfide bonds. As before, disulfide bond formation was detected for the M289C/A319C pair. In this new construct, we now see a small amount of disulfide bond formation in the distant N285C/G329C pair, likely an effect of switching to the MBP tag.

Structural similarity of Sgt2-C domain to STI1 domains

Attempts to glean functional insight for Sgt2-C from Basic Local Alignment Search Tool searches did not reliably return other families or non-Sgt2 homologs making functional comparisons difficult. A more extensive profile-based search using hidden Markov models from the SMART database (52) identified a similarity to domains in the yeast cochaperone Sti1 (HOP in mammals). First called DP1 and DP2, due to their prevalence of aspartates (D) and prolines (P), these domains have been shown to be required for client binding by Sti1 (53, 54) and are termed "STI1" domains in bioinformatics databases (52). In yeast Sti1 and its human homolog HOP (combination will be referred to here as Sti1), each of the two STI1 domains (DP1 and DP2) is preceded by Hsp70/90-binding TPR domains, similar to the domain architecture of Sgt2. Deletion of the second, C-terminal STI1-domain (DP2) from Sti1 *in vivo* is detrimental, impairing native activity of the glucocorticoid receptor (53). *In vitro*, removal of the DP2 domain from Sti1 results in the loss of recruitment of the progesterone receptor to Hsp90 without interfering in Sti1-Hsp90 binding (55). These results implicate DP2 in binding of Sti1 clients. In addition, others have noted that, broadly, STI1 domains may present a hydrophobic groove for binding the hydrophobic segments of a client (53, 54). Furthermore, the similar domain organizations (*i.e.*, Sgt2 TPR-C, Sti1 TPR-STI1) and molecular roles could imply an evolutionary relationship between these cochaperones. Indeed, a multiple sequence alignment of the Sgt2-C$_{cons}$ with several yeast STI1 domains (Fig. 5A) reveals strong conservation of structural features. H1–H5 of the predicted helical regions in C$_{cons}$ align directly with the structurally determined helices in the DP2 domain of Sti1; this includes complete conservation of helix-breaking prolines and close alignment of hydrophobic residues in the amphipathic helices (53).

Based on the domain architecture and homology, a direct comparison between the STI1 domain and Sgt2-C$_{cons}$ can be made. A structure of DP2 solved by solution NMR reveals that the five amphipathic helices assemble to form a flexible helical hand with a hydrophobic groove (53). The lengths of the α-helices in this structure concur with those inferred from the alignment in Figure 4A. Our molecular model of Sgt2-C$_{cons}$ is strikingly similar to this DP2 structure (Fig. 5, B and C). An overlay of the DP2 structure and our molecular model demonstrates both Sgt2-C$_{cons}$ and DP2 have similar lengths and
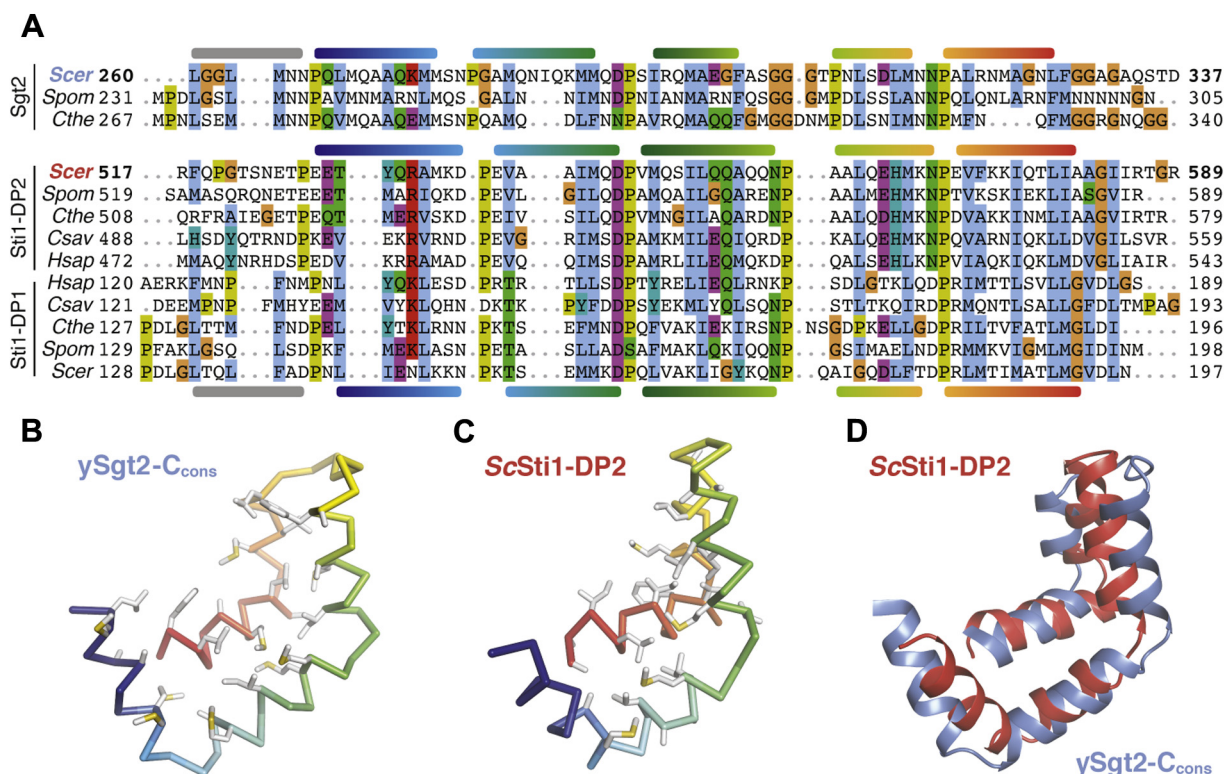

Figure 5. Comparison of Sti1 domains and the Sgt2-C$_{cons}$ model. *A*, multiple sequence alignment of Sgt2-C with STI1 domains (DP1, DP2) from STI1/Hop homologs. Helices are shown based on the Sgt2-C$_{cons}$ model and the *Sc*Sti1-DP1/2 structures. Species for representative sequences are from *S. cerevisiae* (*Scer*), *S. pombe* (*Spom*), *C. thermophilum* (*Cthe*), *C. savignyi* (*Csav*), and *H. sapiens* (*Hsap*). *B*, Cα ribbon of *Sc*Sgt2-C$_{cons}$ color-ramped with large hydrophobic sidechains shown as *gray sticks* (sulfurs in *yellow*). *C*, similar to *B* for the solution NMR structure of Sti1-DP2$_{526-582}$ (PDBID: 2LLW) (53). *D*, superposition of the Sgt2-C$_{cons}$ (*blue*) and Sti1-DP2$_{526-582}$ (*red*) drawn as cartoons.

arrangements of their amphipathic helices (Fig. 5*D*). Consistent with our observations of flexibility in Sgt2-C$_{cons}$, Sti1-DP2 generates few long-range Nuclear Overhauser effects between its helices indicating that Sti1-DP2 also has a flexible architecture (53). We consider this flexibility a feature of these helical hands for reversible and specific binding of a variety of clients.

Binding mode of clients to Sgt2

We examined the Sgt2-C$_{cons}$ surface that putatively interacts with clients by constructing hydrophobic-to-charge residue mutations that are expected to disrupt capture of clients by Sgt2. Similar to the helix mutations in Figure 2, *E* and *F*, the capture assay was employed to establish the relative effects of individual mutations. A baseline was established based on the amount of the TA client Sbh1 captured by WT Sgt2-TPR-C. In each experiment, Sbh1 was expressed at the same level; therefore, differences in binding should directly reflect the affinity of Sgt2 mutants for clients. In all cases, groove mutations from hydrophobic to aspartate led to a reduction in client binding (Fig. 6). The effects are most dramatic with ySgt2 where each mutant significantly reduced binding by 60% or more (Fig. 6*A*). While all hSgt2 individual mutants saw a significant loss in binding, the results were more subtle with the strongest, a ~36% reduction (M233D, Fig. 6*B*). Double mutants were stronger with a significant decrease in binding relative to the individual mutants, more reflective of the individual mutants in ySgt2. As seen before (Fig. 2, *E* and

F), we observe that mutations toward the N terminus of Sgt2-C have a stronger effect on binding than those later in the sequence, whether single-point mutants in the case of ySgt2 or double mutants for hSgt2.

Sgt2-C domain binds clients with a hydrophobic segment ≥11 residues

With a molecular model for ySgt2-C$_{cons}$ and multiple lines of evidence for a hydrophobic groove, we sought to better understand the specific requirements for TMD binding. TMD clients were designed where the overall (sum) and average (mean) hydrophobicity, length, and the distribution of hydrophobic character were varied in the TMDs. These artificial TMDs, a Leu/Ala helical stretch followed by a Trp, were constructed as C-terminal fusions to the soluble protein BRIL (Fig. 7*A*). The total and mean hydrophobicities are controlled by varying the helix length and the Leu/Ala ratio. For clarity, we define a syntax for the various artificial TMD clients to highlight the various properties under consideration: hydrophobicity, length, and distribution. The generic notation is TMD length [number of leucines], which is represented, for example, as 18[L6] for a TMD of 18 amino acids containing six leucines.

Our first goal with the artificial clients was to define the minimal length of a TMD to bind to the C domain. As described earlier in our single-point mutation capture assays, captures of His-tagged Sgt2-TPR-C with the various TMD clients were performed. We define a relative binding efficiency
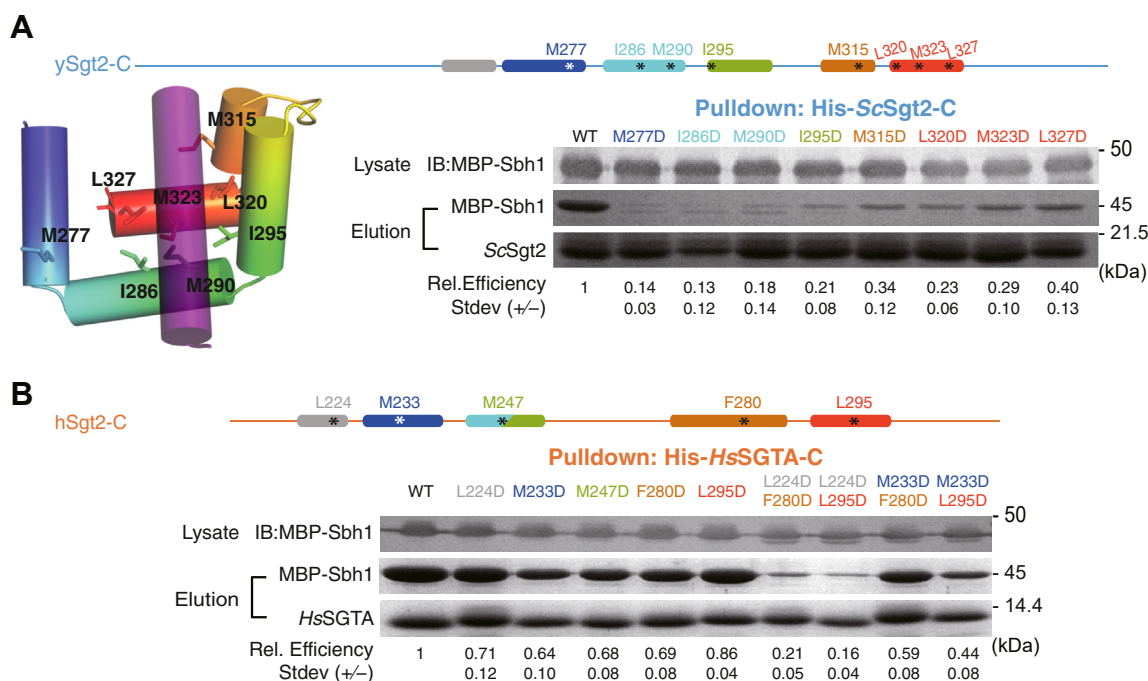

Figure 6. Effects on client binding of charge mutations to the putative hydrophobic groove of Sgt2-C$_{cons}$. For these experiments, individual point mutations are introduced into Sgt2-C and tested for their ability to capture Sbh1 quantified as in Figure 2*D*. *A*, for ySgt2-C, a schematic and cartoon model are provided highlighting the helices and sites of individual point mutants both color-ramped for direct comparison. For the cartoon, the docked TMD is shown in purple. Binding of MBP-Sbh1 to His-tagged ySgt2-C and mutants were examined as in Figure 2*E*. Lanes for mutated residues are labeled in the same color as the schematic. *B*, same analysis as in *A* for hSgt2-C. In addition, double point mutants are included. Each capture assay was repeated three times.
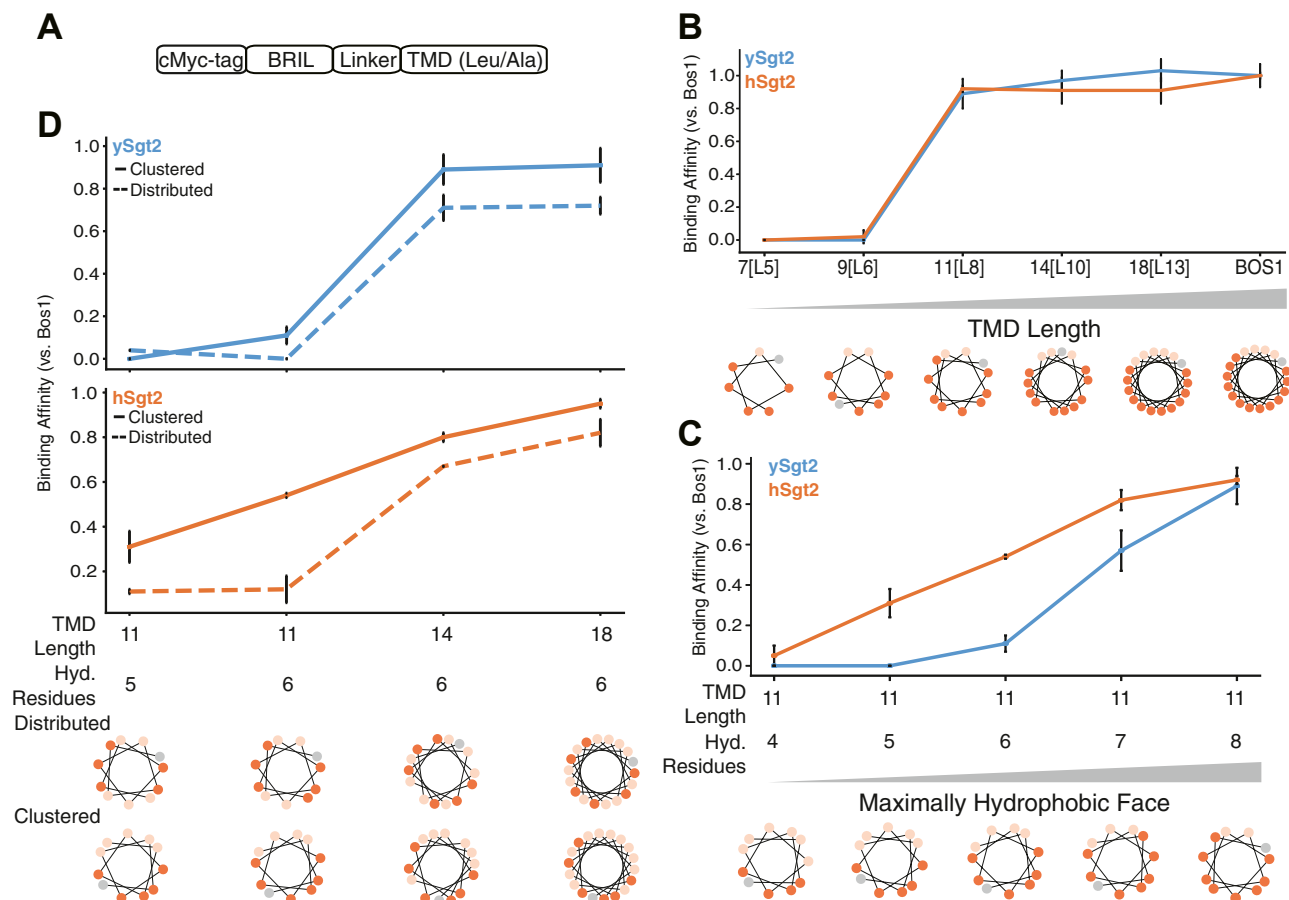
Figure 7. Minimal requirements for client recognition by Sgt2. *A*, schematic of model clients. From capture assays, quantification of complex formation in the eluate is calculated and normalized to that of complexes with Bos1$_{TMD}$, here defined as relative binding efficiency. *B*, complex formation of ySgt2 (*blue*) and hSgt2 (*orange*) with the TA client Bos1$_{TMD}$ and several artificial clients noted x[Ly], where x denotes the length of the TMD and y denotes the number of leucines in the TMD. The helical wheel diagrams of the TMD of clients here and for subsequent panels with leucines colored in *dark orange*, alanines colored in *pale orange*, and tryptophans colored in *gray*. Each assay was performed four times except for ySgt2-Bos1 and hSgt2-9[L6], which were performed three times. *C*, complex formation of ySgt2-TPR-C and hSgt2-TPR-C with artificial clients with TMDs of length 11 and increasing numbers of leucine. Capture assays were repeated either two or three times. *D*, comparison of complex formation of ySgt2-TPR-C and hSgt2-TPR-C with artificial clients of the same lengths and hydrophobicities but differences in the distribution of leucines, *i.e.*, clustered (*solid line*) versus distributed (*dotted line*). Each assay was performed either three or four times.

as the ratio of captured TMD client by a Sgt2-TPR-C normalized to the ratio of a captured WT TA client by Sgt2-TPR-C. In this case we replaced the TMD in our artificial clients with the native TMD of Bos1 (Bos1$_{TMD}$). The artificial client 18[L13] shows a comparable binding efficiency to Sgt2-TPR-C as that of Bos1$_{TMD}$ (Fig. 7B). From the helical wheel diagram of the TMD for Bos1, we noted that the hydrophobic residues favored one face of the helix. We explored this "hydrophobic face" by using model clients that maintained this orientation while shortening the length and maintaining the average hydrophobicity of 18[L13] (Fig. 7B). Shorter helices of 14 or 11 residues, 14[L10] and 11[L8], also bound with similar affinity to Bos1. Helices shorter than 11 residues, 9[L6] and 7[L5], were not able to bind Sgt2-TPR-C (Fig. 7B), establishing a minimal length of 11 residues for the helix, consistent with the dimensions of the groove predicted from the structural model (Fig. 3).

Since a detected binding event occurs with TMDs of at least 11 amino acids, we decided to probe this limitation further. The dependency of client hydrophobicity was tested by

measuring complex formation of Sgt2-TPR-C and artificial TMD clients containing an 11 amino acid TMD with increasing number of leucines (11[Lx]). As shown in Figure 7C, increasing the number of leucines monotonically enhances complex formation, echoing previous results (56). hSgt2-TPR-C binds to a wider spectrum of hydrophobic clients than ySgt2-TPR-C, which could mean it has a more permissive hydrophobic binding groove, also reflected in the milder impact of Ala replacement and Asp mutations in hSgt2-TPR-C to TMD client binding (Figs. 2F and 6B).

Sgt2-C preferentially binds to TMDs with a hydrophobic face

Next, we address the properties within the TMD of clients responsible for Sgt2 binding. In the case of ySgt2, it has been suggested that the cochaperone binds to TMDs based on hydrophobicity and helical propensity (56). In our system, our artificial TMDs consist of only alanines and leucines, which have high helical propensities (57), and despite keeping the helical propensity constant and in a range that favors Sgt2 binding, there is still variation in binding efficiency. For the

most part, varying the hydrophobicity of an artificial TMD client acts as expected, the more hydrophobic TMDs bind more efficiently to Sgt2-TPR-C (Fig. 7*C*). Our C$_{cons}$ model suggests that the hydrophobic groove of ySgt2-C protects a TMD with highly hydrophobic residues clustered to one side (see Fig. 3*B*). To test this, various TMD pairs with the same hydrophobicity, but different distributions of hydrophobic residues, demonstrate that TMD clients with clustered leucines have a higher relative binding efficiency than those with a more uniform distribution (Fig. 7*D*). Helical wheel diagrams demonstrate the distribution of hydrophobic residues along the helix (*e.g.*, bottom Fig. 7*D*). The clustered leucines in the TMDs create a hydrophobic face, which potentially interacts with the hydrophobic groove formed by the Sgt2-C$_{cons}$ region, corresponding to the model in Figure 3*B*.

Discussion

Sgt2, the most upstream component of the GET pathway, plays a critical role in the targeting of TA IMPs to their correct membranes along with other roles in maintaining cellular homeostasis. Its importance as the first confirmed selection step of ER *versus* mitochondrial (56) destined TA IMPs necessitates a molecular model for client binding. Previous work demonstrated a role for the C domain of Sgt2 to bind to hydrophobic clients, yet the exact binding domain remained to be determined. Through the combined use of biochemistry, bioinformatics, and computational modeling, we conclusively identify the minimal client-binding domain of Sgt2 and preferences in client binding. Here we present a validated structural model of the Sgt2 C domain as a methionine-rich helical hand for grasping a hydrophobic helix and to provide a mechanistic explanation for binding a TMD of at least 11 hydrophobic residues.

Identifying the C domain of Sgt2 as containing an STI1 places Sgt2 into a larger context of conserved cochaperones (Fig. 8*A*). In the cochaperone family, the STI1 domains predominantly follow HSP-binding TPR domains connected by a flexible linker, reminiscent of the domain architecture of Sgt2. As noted above, for STI1 these domains are critical for coordinated hand-off between Hsp70 and Hsp90 homologs (58) as well as coordinating the simultaneous binding of two HSPs. Both Sgt2 and the cochaperone Hip coordinate pairs of TPR and STI1 domains by forming stable dimers *via* their N-terminal dimerization domains (59). With evidence for a direct role of the carboxylate clamp in the TPR domain of Sgt2 for TA client binding now clear (21), one can speculate that the two TPR domains may facilitate TA client entry into various pathways that use multiple HSPs.

Computational modeling reveals that a conserved region, sufficient for client binding, forms a five-alpha-helical hand, which is reminiscent of other proteins involved in membrane protein targeting. Like Sgt2, the signal recognition particle (SRP) contains a methionine-rich domain that binds signal sequences and TMDs. While the helical order is inverted, again five amphipathic helices form a hydrophobic groove that cradles the client signal peptide (60) (Fig. 8*B*). Here once more,

the domain has been observed to be flexible in the absence of client (61, 62) and, in the resting state, occupied by a region that includes a helix, which must be displaced (60). Another helical-hand example recently shown to be involved in TA IMP targeting is calmodulin where a crystal structure reveals two helical hands coordinating to clasp a TMD at either end (Fig. 8*B*) (63). Considering an average TMD of 18 to 20 amino acids (to span a ∼40 Å bilayer), each half of calmodulin interacts with about ten amino acids. This is in close correspondence to the demonstrated minimal 11 amino acids for a TMD client to bind to the monomeric Sgt2-TPR-C. In the context of the full-length Sgt2, one can speculate that the Sgt2 dimer may utilize both C domains to bind to a full TMD, similar to calmodulin. Cooperation of the two Sgt2 C domains in client binding could elicit conformational changes in the complex that could be recognized by downstream factors, such as additional interactions that increase the affinity to Get5/Ubl4A.

Intriguingly, Sgt2-TPR-C preferentially binds to artificial clients with clustered leucines. The hydrophobic groove presented in the computational model provides an attractive explanation for this preference. In order to bind to the hydrophobic groove, a client buries a portion of its TMD in the groove leaving the other face exposed. Clustering the most hydrophobic residues contributes to the hydrophobic effect driving binding efficiency and protecting them from the aqueous environment. Indeed, when focusing on Sgt2's role in TA IMP targeting, GET pathway substrates have been suggested to be more hydrophobic TMDs than Endoplasmic Reticulum Membrane Complex substrates (64). Of these, for the most hydrophobic substrates, such as Bos1, residues on both sides of the TMD could be protected by a pair of C domains. Alternatively, the unstructured N-terminal loop through H0 could act as a lid surrounding the circumference of the client's TMD. Unstructured regions participating in substrate binding as well as capping a hydrophobic groove have been suggested in the context of other domains, *e.g.*, with Get3 (4). The role for this clustering of hydrophobic residues in client recognition and targeting merits further investigation.

What is the benefit of the flexible helical-hand structure for hydrophobic helix binding? While it remains an open question, it is notable that evolution has settled on similar simple solutions to the complex problem of specific but temporary binding of hydrophobic helices. For all of the domains mentioned above, the flexible helical hands provide an extensive hydrophobic surface to capture a client-helix— driven by the hydrophobic effect. Typically, such extensive interfaces are between pairs of preordered surfaces resulting in high affinities and slow off rates. These helical hands are required to only engage temporarily, therefore the flexibility offsets the favorable free energy of binding by charging an additional entropic cost for ordering a flexible structure in the client-bound complex. The benefit for clients is a favorable transfer to downstream components in the GET pathway as seen for ySgt2 (21) and hSgt2 (50). The demonstration that TA client transfer from hSgt2 to Get3 is twice as fast as disassociation from hSgt2 into solution, perhaps interaction
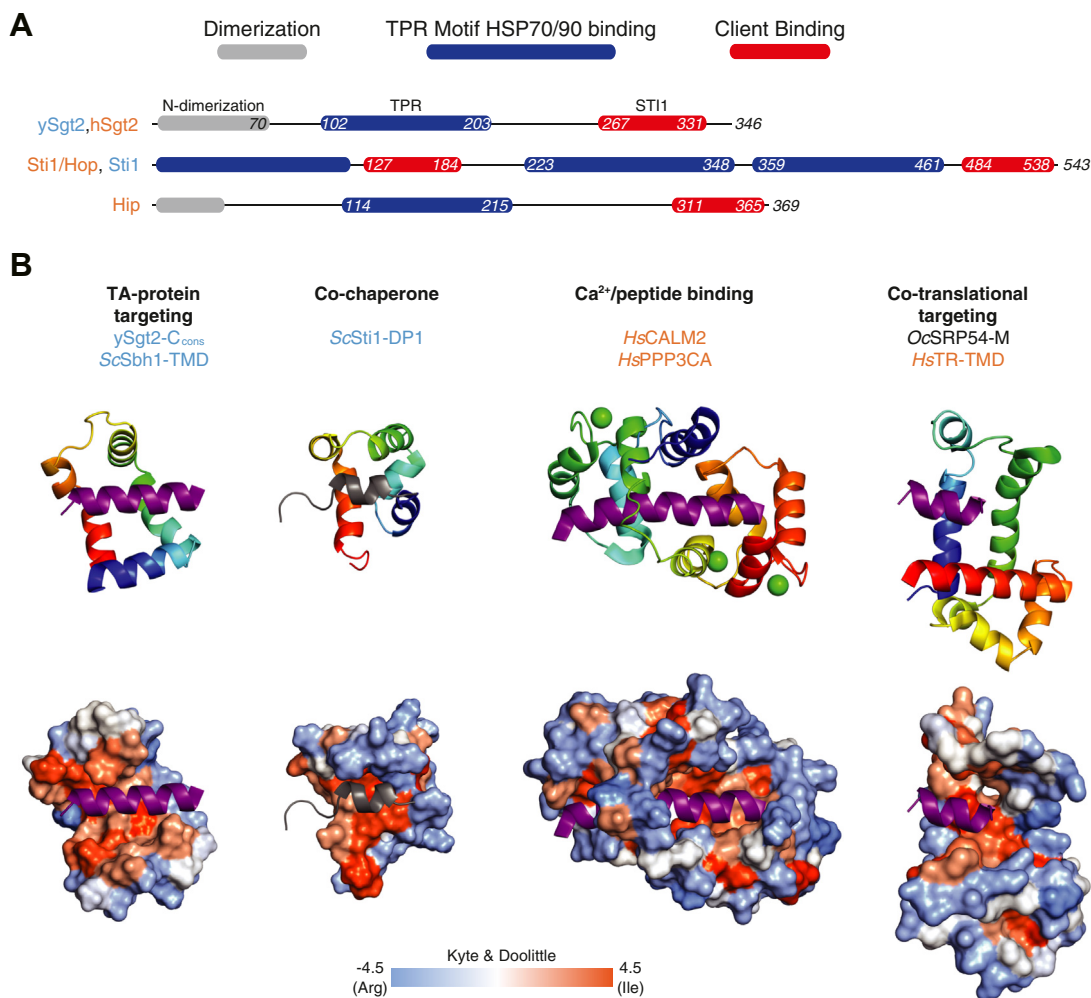
Figure 8. Various domain structures of STI1 and other helical-hand-containing proteins. *A*, the domain architectures of proteins with an STI1 domain were obtained initially from InterPro (87) and then adjusted as discussed in the text. Each domain within a protein is colored relative to the key. *B*, structural comparison of various hydrophobic-binding helical-hand protein complexes. For each figure only relevant domains are included. *Upper row*, color-ramped cartoon representation with bound helices in *purple*. *Lower row*, accessible surface of each protein colored by hydrophobicity again with docked helical clients in *purple*. In order, the predicted complex of ySgt2-C_{cons} and *Sc*Sbh2-TMD, DP1 domain from yeast Sti1 with N terminus containing H0 in *gray* (*Sc*Sti1-DP1)(PDBID: 2LLV), human calmodulin (*Hs*CALM2) bound to a hydrophobic domain of calcineurin (*Hs*PPP3CA) (PDBID: 2JZI), and M domain of SRP54 from *Oryctolagus cuniculus* (*Oc*SRP54-M) and the signal sequence of human transferrin receptor (*Hs*TR-TMD) (PDBID: 3JAJ).

with Get3 leads to conformational changes that further favor release (50).

While hSgt2 and ySgt2 share many properties, there are a number of differences between the two homologs that may explain the different biochemical behavior. For the C_{cons} domains, hSgt2 appears to be more ordered in the absence of client as the peaks in its NMR spectra are broader (Fig. S1, *B* and *C*). Comparing the domains at the sequence level, while the high glutamine content in the C domain is conserved, it is higher in hSgt2 (8.8% *versus* 15.2%). The additional glutamines are concentrated in the predicted longer H4 helix (Fig. 1*A*). The linker to the TPR domain is shorter compared with ySgt2 while the loop between H3 and H4 is longer. Do these differences reflect different roles? As noted, in every case the threshold for hydrophobicity of client binding is lower for hSgt2 than ySgt2 (Figs. 2*E*, 6, and 7) implying that the mammalian protein is more permissive in client binding. The two C domains have similar hydrophobicity, so this difference in binding might be due to a lower entropic cost paid by

having the hSgt2 C domain more ordered in the absence of client or the lack of an unstructured N-terminal loop.

The targeting of TA clients presents an intriguing and enigmatic problem for understanding the biogenesis of IMPs. How subtle differences in each client modulate the interplay of hand-offs that direct these proteins to the correct membrane remains to be understood. In this study, we focus on a central player, Sgt2, and its client-binding domain. Through biochemistry and computational analysis, we provide a structural model that adds more clarity to client binding both within and outside of the GET pathway.

Experimental procedures

Plasmid constructs

MBP-Sbh1 full length, $ySgt2_{95-346}$ (*ySgt2-TPR-C*), $ySgt2_{222-346}$ (*ySgt2-C*), $ySgt2_{260-327}$ (*ySgt2-C_{cons}*), $ySgt2_{266-327}$ (*ySgt2-ΔH0*), $hSgt2_{87-313}$ (*hSgt2-TPR-C*), $hSgt2_{213-313}$ (*hSgt2-C*), $hSgt2_{219-300}$ (*hSgt2-C_{cons}*), and $hSgt2_{228-300}$ (*hSgt2-ΔH0*)

were prepared as previously described (12, 65). Genes of *ySgt2* or *hSgt2* variants were amplified from constructed plasmids and then ligated into a pET33b-derived vector with a 17-residue N-terminal hexa-histidine tag and a tobacco etch virus (TEV) protease site. Single or multiple mutations on Sgt2 were constructed by site-direct mutagenesis. Artificial clients were constructed in a pACYC-Duet plasmid with an N-terminal cMyc tag, BRIL fusion protein (66), GSS linker, and a hydrophobic C-terminal tail consisting of leucines and alanines and ending with a tryptophan.

Protein expression and purification

All proteins were expressed in *Escherichia coli* NiCo21 (DE3) cells (New England BioLabs). To coexpress multiple proteins, constructed plasmids were cotransformed as described (65). Protein expression was induced by 0.3 mM IPTG at $OD_{600} \sim 0.7$ and harvested after 3 h at 37 °C. For structural analysis, cells were lysed through an M-110L Microfluidizer Processor (Microfluidics) in lysis buffer (50 mM Tris, 300 mM NaCl, 25 mM imidazole supplemented with benzamidine, phenylmethylsulfonyl fluoride (PMSF), and 10 mM β-mercaptoethanol (BME), pH 7.5). For capture assays, cells were lysed by freeze-thawing three times with 0.1 mg/ml lysozyme. To generate endogenous proteolytic products of *ySgt2*-TPR-C for MS analysis, PMSF and benzamidine were excluded from the lysis buffer. His-tagged Sgt2 and His-tagged Sgt2/TA complexes were separated from the lysate by batch incubation with Ni-NTA resin at 4 °C for 1 h. The resin was washed with 20 mM Tris, 150 mM NaCl, 25 mM imidazole, 10 mM BME, pH 7.5. The complexes of interest were eluted in 20 mM Tris, 150 mM NaCl, 300 mM imidazole, 10 mM BME, pH 7.5.

For structural analysis, the affinity tag was removed from complexes collected after the nickel elution by an overnight TEV digestion against lysis buffer followed by size-exclusion chromatography using a HiLoad 16/60 Superdex 75 prep grade column (GE Healthcare).

Measurement of Sgt2 protein concentration was carried out using the bicinchoninic acid (BCA) assay with bovine serum albumin (BSA) as standard (Pierce Chemical Co). Samples for NMR and CD analyses were concentrated to 10 to 15 mg/ml for storage at −80 °C before experiments.

For the *in vitro* transfer assay, plasmids encoding for the full-length ySgt2 cysteine mutants were transformed into BL21 Star cells (Invitrogen). Cells were grown in 2x yeast-tryptone (2xYT) media and induced with 0.1 mM IPTG at an OD_{600} of 0.6, then harvested after 3 h at 30 °C by centrifugation. Cells were lysed in 50 mM Tris pH 8.0, 500 mM NaCl, 10% glycerol, and 1x Bug-Buster (Millipore Sigma), supplemented with protease inhibitors (4-(2-aminoethyl)benzenesulfonyl fluoride hydrochloride (Roche), benzamidine, and BME). Full-length His-tagged ySgt2 and cysteine mutants were separated from the lysate by batch incubation with Ni-NTA resin (Qiagen) at 4 °C for 1 h. The resin was washed with 50 mM Tris pH 8.0, 500 mM NaCl, 10% glycerol, and 25 mM imidazole, and then the protein was eluted in 50 mM Tris pH 8.0, 500 mM NaCl, 10% glycerol, and 300 mM imidazole. For storage, protein was dialyzed in 25 mM K-HEPES pH 7.5, 150 mM KOAc, and 20% glycerol at 4 °C and then flash-frozen in

liquid nitrogen. Purified Bos1 with p-benzoyl-l-phenylalanine (BPA) labeled at residue 230 ($Bos1_{BPA}$) and yeast Ssa1 were gifts from the lab of Shu-ou Shan (Caltech).

NMR spectroscopy

^{15}N-labeled proteins were generated from cells grown in autoinduction minimal media as described (67) and purified in 20 mM phosphate buffer, pH 6.0 (for ySgt2-C, 10 mM Tris, 100 mM NaCl, pH 7.5). The NMR measurements of ^{15}N-labeled Sgt2-C proteins (∼0.3–0.5 mM) were collected using a Varian INOVA 600 MHz spectrometer at either 25 °C (ySgt2-C) or 35 °C (hSgt2-C) with a triple-resonance probe and processed with TopSpin 3.2 (Bruker Co).

CD spectroscopy

The CD spectra were recorded at 24 °C with an Aviv 202 spectropolarimeter using a 1 mm path length cuvette with 10 μM protein in 20 mM phosphate buffer, pH 7.0. The CD spectrum of each sample was recorded as the average over three scans from 190/195 to 250 nm in 1 nm steps. Each spectrum was then decomposed into its most probable secondary structure elements using BeStSel (68).

Glu-C digestion of the double cysteine mutants on ySgt2-C

Complexes of the coexpressed WT or double cysteine mutated His-*ySgt2*-TPR-C and the artificial client, 11[L8], with either a cMyc or an MBP tag, were purified as the other His-Sgt2 complexes described above or initially purified *via* amylose affinity chromatography before nickel chromatography explained earlier. The protein complexes were mixed with 0.2 mM CuSO_4 and 0.4 mM 1,10-phenanthroline at 24 °C for 20 min followed by 50 mM N-ethyl maleimide for 15 min. Sequencing-grade Glu-C protease (Sigma) was mixed with the protein samples at an approximate ratio of 1:30, and the digestion was conducted at 37 °C for 22 h. Digested samples were mixed with either nonreducing or reducing SDS-sample buffer, resolved *via* SDS-PAGE using Mini-Protean Tris-Tricine Precast Gels (10–20%, Bio-Rad), and visualized using Coomassie Blue staining.

In vitro *transfer assay of Bos1 from Ssa1 to ySgt2*

The *in vitro* transfer assays were performed as in Chio *et al.* 2019 and Shao *et al.* 2017 (49, 50). Specifically, 39 μM $Bos1_{BPA}$ (50 mM HEPES, 300 mM NaCl, 0.05% LDAO, 20% glycerol) was diluted to a final concentration of 0.1 μM and added to 4 μM Ssa1 supplemented with 2 mM ATP (25 mM HEPES pH7.5, 150 mM KOAc). After 1 min, 0.3 μM of full-length ySgt2 or mutant was added to the reaction. Samples were flash-frozen after 1 min and placed under a 365 nm UV lamp for 2 h on dry ice to allow for BPA cross-linking.

Protein immunoblotting and detection

For western blots, protein samples were resolved *via* SDS-PAGE and then transferred onto nitrocellulose membranes by the Trans-Blot Turbo Transfer System (Bio-Rad). Membranes were blocked in 5% nonfat dry milk and hybridized with antibodies in TBST buffer (50 mM Tris-HCl pH 7.4, 150 mM

NaCl, 0.1% Tween 20) for 1 h of each step at 24 °C. The primary antibodies were used at the following dilutions: 1:1000 anti-penta-His mouse monoclonal (Qiagen), 1:5000 anti-cMyc mouse monoclonal (Sigma), and a 1:3000 anti-Strep II rabbit polyclonal (Abcam). A secondary antibody conjugated to either alkaline phosphatase (Rockland, 1:8000) or a 800 nm fluorophore was employed, and the blotting signals were chemically visualized with either the nitro-blue tetrazolium/5-bromo-4-chloro-3'-indolyphosphate (NBT/BCIP) chromogenic assay (Sigma) or infrared scanner. All blots were photographed and quantified by image densitometry using ImageJ (69) or ImageStudioLite (LI-COR Biosciences).

Quantification of Sgt2-TA complex formation

The densitometric analysis of MBP-Sbh1 capture by His-Sgt2-TPR-C quantified the intensity of the corresponding protein bands on a Coomassie Blue G-250 stained gel. The quantified signal ratios of MBP-Sbh1/His-Sgt2-TPR-C are normalized to the ratio obtained from the WT. Expression level of MBP-Sbh1 was confirmed by immunoblotting the MBP signal in cell lysate. Average ratios and standard deviations were obtained from 3 to 4 independent experiments.

In artificial client experiments, both His-tagged Sgt2-TPR-C and cMyc-tagged artificial clients were quantified *via* immunoblotting signals. The complex efficiency of Sgt2-TPR-C with various clients was obtained by

$$E_{complex} = \frac{E_{TMD}}{T_{TMD}} \times \frac{1}{E_{capture}} \qquad (1)$$

where E_{TMD} is the signal intensity of an eluted client representing the amount of client copurified with Sgt2-TPR-C, and T_{TMD} is the signal intensity of a client in total lysate, which corresponds to the expression yield of that client. Identical volumes of elution and total lysate from different client experiments were analyzed and quantified. In order to correct for possible variation in the amount of Sgt2-TPR-C available for complex formation, $E_{capture}$ represents the relative amount of Sgt2-TPR-C present in the elution (E_{Sgt2}) compared with a pure Sgt2-TPR-C standard ($E_{purified,Sgt2}$).

$$E_{capture} = \frac{E_{Sgt2}}{E_{purified,Sgt2}}, \qquad (2)$$

Each E_{TMD} and T_{TMD} value was obtained by blotting both simultaneously, *i.e.*, adjacently on the same blotting paper. To facilitate comparison between clients, the Sgt2-TPR-C/TA client complex efficiency $E_{complex,TMD}$ is normalized by Sgt2-TPR-C/Bos1 complex efficiency $E_{complex,Bos1}$.

$$\% \text{ Complex} = \frac{E_{complex,TMD}}{E_{complex,Bos1}} \times 100 \qquad (3)$$

Sequence alignments

An alignment of Sgt2-C domains was carried out as follows: all sequences with an annotated N-terminal Sgt2/A dimerization domain (PF16546 (70)), at least one TPR hit (PF00515.27, PF13176.5, PF07719.16, PF13176.5, PF13181.5), and at least 50 residues following the TPR domain were considered family members. Putative C domains were inferred as all residues following the TPR domain, filtered at 90% sequence identity using CD-HIT (71), and then aligned using MAFFT G-INS-i (72). Other attempts with a smaller set (therefore more divergent) of sequences result in an ambiguity in the relative register of H0, H1, H2, and H3 when comparing Sgt2 with SGTA.

Alignments of Sti1 (DP1/DP2) and STI1 domains were created by pulling all unique domain structures with annotated STI1 domains from Uniprot. Where present, the human homolog was selected and then aligned with PROMALS3D (73). PROMALS3D provides a way of integrating a variety of costs into the alignment procedure, including 3D structure, secondary structure predictions, and known homologous positions.

All alignments were visualized using Jalview (74).

Molecular modeling

Putative models for ySgt2-C were generated with I-TASSER, PCONS, Quark, Robetta (*ab initio* and transform-restrained modes), Phyre2, and RaptorX *via* their respective web servers (48, 75–78). The highest scoring model from Quark was then chosen to identify putative TA client-binding sites by rigid-body docking of various transmembrane domains modeled as α-helices (3D-HM (79)) into the ySgt2-C_{cons} through the Zdock web server (80). Pairwise distances were calculated between C_β atoms (the closer C_α proton on glycine) using mdtraj (81). Based on our disulfide crosslinks, new models were predicted using Robetta in *ab initio* mode specifying C_β–C_β atom distance constraints bounded between 0 and 9 Å.

For hSgt2, using the same set of structure prediction servers above, we were only able to produce a clear structural model using the Robetta transform-restrained mode. We were also unable to generate a reliable model by directly using the ySgt2-C model as a template (82). To cross-link distance data from ySgt2 as restraints for hSgt2, pair positions were transferred from one protein to the other *via* an alignment of Sgt2-C domains (excerpt in Fig. 1*A*) and ran Robetta *ab initio*. Also, we grafted the N-terminal loop of ySgt2-C on hSgt2-C with the same set of restraints.

Images were rendered using PyMOL 2.3 (www.pymol.org).

Data availability

All data is provided in the article.

Abbreviations—The abbreviations used are: BPA, p-benzoyl-l-phenylalanine; ER, endoplasmic reticulum; ERAD, ER-associated degradation; GET, guided entry of tail-anchored protein; IMP, integral membrane protein; MLP, mislocalized membrane protein; PMSF, phenylmethylsulfonyl fluoride; SGTA, small glutamine tetratricopeptide repeat protein alpha; SRP, signal recognition particle; TA, tail-anchored; TMD, transmembrane domain; TRC, transmembrane recognition complex; UBL, ubiquitin-like domain.

References

1. Pieper, U., Schlessinger, A., Kloppmann, E., Chang, G. A., Chou, J. J., Dumont, M. E., Fox, B. G., Fromme, P., Hendrickson, W. A., Malkowski, M. G., Rees, D. C., Stokes, D. L., Stowell, M. H. B., Wiener, M. C., Rost, B., *et al.* (2013) Coordinating the impact of structural genomics on the human α-helical transmembrane proteome. *Nat. Struct. Mol. Biol.* **20**, 135–138

2. Aviram, N., and Schuldiner, M. (2017) Targeting and translocation of proteins to the endoplasmic reticulum at a glance. *J. Cell Sci.* **130**, 4079–4085

3. Shao, S., and Hegde, R. S. (2011) Membrane protein insertion at the endoplasmic reticulum. *Annu. Rev. Cell Dev. Biol.* **27**, 25–56

4. Guna, A., and Hegde, R. S. (2018) Transmembrane domain recognition during membrane protein biogenesis and quality control. *Curr. Biol.* **28**, R498–R511

5. Kutay, U., Hartmann, E., and Rapoport, T. A. (1993) A class of membrane proteins with a C-terminal anchor. *Trends Cell Biol.* **3**, 72–75

6. Hegde, R. S., and Keenan, R. J. (2011) Tail-anchored membrane protein insertion into the endoplasmic reticulum. *Nat. Rev. Mol. Cell Biol.* **12**, 787–798

7. Denic, V. (2012) A portrait of the GET pathway as a surprisingly complicated young man. *Trends Biochem. Sci.* **37**, 411–417

8. Wattenberg, B., and Lithgow, T. (2001) Targeting of C-terminal (tail)-anchored proteins: Understanding how cytoplasmic activities are anchored to intracellular membranes. *Traffic* **2**, 66–71

9. Chartron, J. W., Clemons, W. M., and Suloway, C. J. M. (2012) The complex process of GETting tail-anchored membrane proteins to the ER. *Curr. Opin. Struct. Biol.* **22**, 217–224

10. Chio, U. S., Cho, H., and Shan, S.-O. (2017) Mechanisms of tail-anchored membrane protein targeting and insertion. *Annu. Rev. Cell Dev. Biol.* **33**, 417–438

11. Shao, S., and Hegde, R. S. (2011) A calmodulin-dependent translocation pathway for small secretory proteins. *Cell* **147**, 1576–1588

12. Chartron, J. W., Gonzalez, G. M., and Clemons, W. M., Jr. (2011) A structural model of the Sgt2 protein and its interactions with chaperones and the Get4/Get5 complex. *J. Biol. Chem.* **286**, 34325–34334

13. Wang, F., Brown, E. C., Mak, G., Zhuang, J., and Denic, V. (2010) A chaperone cascade sorts proteins for posttranslational membrane insertion into the endoplasmic reticulum. *Mol. Cell* **40**, 159–171

14. Simon, A. C., Simpson, P. J., Goldstone, R. M., Krysztofinska, E. M., Murray, J. W., High, S., and Isaacson, R. L. (2013) Structure of the Sgt2/Get5 complex provides insights into GET-mediated targeting of tail-anchored membrane proteins. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 1327–1332

15. Xu, Y., Cai, M., Yang, Y., Huang, L., and Ye, Y. (2012) SGTA recognizes a noncanonical ubiquitin-like domain in the Bag6-Ubl4A-Trc35 complex to promote endoplasmic reticulum-associated degradation. *Cell Rep.* **2**, 1633–1644

16. Wunderley, L., Leznicki, P., Payapilly, A., and High, S. (2014) SGTA regulates the cytosolic quality control of hydrophobic substrates. *J. Cell Sci.* **127**, 4728–4739

17. Leznicki, P., and High, S. (2012) SGTA antagonizes BAG6-mediated protein triage. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 19214–19219

18. Gristick, H. B., Rao, M., Chartron, J. W., Rome, M. E., Shan, S.-O., and Clemons, W. M. (2014) Crystal structure of ATP-bound Get3–Get4–Get5 complex reveals regulation of get3 by Get4. *Nat. Struct. Mol. Biol.* **21**, 437–442

19. Mock, J.-Y., Chartron, J. W., Zaslaver, M., Xu, Y., Ye, Y., and Clemons, W. M., Jr. (2015) Bag6 complex contains a minimal tail-anchor–targeting module and a mock BAG domain. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 106–111

20. Wang, F., Whynot, A., Tung, M., and Denic, V. (2011) The mechanism of tail-anchored protein insertion into the ER membrane. *Mol. Cell* **43**, 738–750

21. Cho, H., and Shan, S.-O. (2018) Substrate relay in an Hsp70-cochaperone cascade safeguards tail-anchored membrane protein targeting. *EMBO J.* **37**, e99264

22. Stefer, S., Reitz, S., Wang, F., Wild, K., Pang, Y.-Y., Schwarz, D., Bomke, J., Hein, C., Löhr, F., Bernhard, F., Denic, V., Dötsch, V., and Sinning, I. (2011) Structural basis for tail-anchored membrane protein biogenesis by the Get3-receptor complex. *Science* **333**, 758–762

23. Rome, M. E., Chio, U. S., Rao, M., Gristick, H., and Shan, S.-O. (2014) Differential gradients of interaction affinities drive efficient targeting and recycling in the GET pathway. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E4929–E4935

24. Vilardi, F., Lorenz, H., and Dobberstein, B. (2011) WRB is the receptor for TRC40/Asna1-mediated insertion of tail-anchored proteins into the ER membrane. *J. Cell Sci.* **124**, 1301–1307

25. Yamamoto, Y., and Sakisaka, T. (2012) Molecular machinery for insertion of tail-anchored membrane proteins into the endoplasmic reticulum membrane in mammalian cells. *Mol. Cell* **48**, 387–397

26. Schuldiner, M., Metz, J., Schmid, V., Denic, V., Rakwalska, M., Schmitt, H. D., Schwappach, B., and Weissman, J. S. (2008) The GET complex mediates insertion of tail-anchored proteins into the ER membrane. *Cell* **134**, 634–645

27. Kiktev, D. A., Patterson, J. C., Müller, S., Bariar, B., Pan, T., and Chernoff, Y. O. (2012) Regulation of chaperone effects on a yeast prion by cochaperone Sgt2. *Mol. Cell. Biol.* **32**, 4960–4970

28. Xu, Y., Liu, Y., Lee, J.-G., and Ye, Y. (2013) A ubiquitin-like domain recruits an oligomeric chaperone to a retrotranslocation complex in endoplasmic reticulum-associated degradation. *J. Biol. Chem.* **288**, 18068–18076

29. Rodrigo-Brenni, M. C., Gutierrez, E., and Hegde, R. S. (2014) Cytosolic quality control of mislocalized proteins requires RNF126 recruitment to Bag6. *Mol. Cell* **55**, 227–237

30. Hessa, T., Sharma, A., Mariappan, M., Eshleman, H. D., Gutierrez, E., and Hegde, R. S. (2011) Protein targeting and degradation are coupled for elimination of mislocalized proteins. *Nature* **475**, 394–397

31. Dupzyk, A., Williams, J. M., Bagchi, P., Inoue, T., and Tsai, B. (2017) SGTA-dependent regulation of Hsc70 promotes cytosol entry of simian virus 40 from the endoplasmic reticulum. *J. Virol.* **91**, 23

32. Long, P., Samnakay, P., Jenner, P., and Rose, S. (2012) A yeast two-hybrid screen reveals that osteopontin associates with MAP1A and MAP1B in addition to other proteins linked to microtubule stability, apoptosis and protein degradation in the human brain. *Eur. J. Neurosci.* **36**, 2733–2742

33. Trotta, A. P., Need, E. F., Selth, L. A., Chopra, S., Pinnock, C. B., Leach, D. A., Coetzee, G. A., Butler, L. M., Tilley, W. D., and Buchanan, G.

(2013) Knockdown of the cochaperone SGTA results in the suppression of androgen and PI3K/Akt signaling and inhibition of prostate cancer cell proliferation. *Int. J. Cancer* **133**, 2812–2823

34. Buchanan, G., Ricciardelli, C., Harris, J. M., Prescott, J., Yu, Z. C. L., Jia, L., Butler, L. M., Marshall, V. R., Scher, H. I., Gerald, W. L., Coetzee, G. A., and Tilley, W. D. (2007) Control of androgen receptor signaling in prostate cancer by the cochaperone small glutamine rich tetratricopeptide repeat containing protein. *Cancer Res.* **67**, 10087–10096

35. Wang, H., Zhang, Q., and Zhu, D. (2003) hSGT interacts with the N-terminal region of myostatin. *Biochem. Biophys. Res. Commun.* **311**, 877–883

36. Chartron, J. W., VanderVelde, D. G., and Clemons, W. M., Jr. (2012) Structures of the Sgt2/SGTA dimerization domain with the Get5/UBL4A UBL domain reveal an interaction that forms a conserved dynamic interface. *Cell Rep.* **2**, 1620–1632

37. Liou, S.-T., and Wang, C. (2005) Small glutamine-rich tetratricopeptide repeat-containing protein is composed of three structural units with distinct functions. *Arch. Biochem. Biophys.* **435**, 253–263

38. Cziepluch, C., Kordes, E., Poirey, R., Grewenig, A., Rommelaere, J., and Jauniaux, J. C. (1998) Identification of a novel cellular TPR-containing protein, SGT, that interacts with the nonstructural protein NS1 of parvovirus H-1. *J. Virol.* **72**, 4149–4156

39. Callahan, M. A., Handley, M. A., Lee, Y. H., Talbot, K. J., Harper, J. W., and Panganiban, A. T. (1998) Functional interaction of human immunodeficiency virus type 1 Vpu and Gag with a novel member of the tetratricopeptide repeat protein family. *J. Virol.* **72**, 5189–5197

40. Winnefeld, M., Grewenig, A., Schn lzer, M., Spring, H., Knoch, T. A., Gan, E. C., Rommelaere, J., and Cziepluch, C. (2006) Human SGT interacts with Bag-6/Bat-3/Scythe and cells with reduced levels of either protein display persistence of few misaligned chromosomes and mitotic arrest. *Exp. Cell Res.* **312**, 2500–2514

41. Darby, J. F., Krysztofinska, E. M., Simpson, P. J., Simon, A. C., Leznicki, P., Sriskandarajah, N., Bishop, D. S., Hale, L. R., Alfano, C., Conte, M. R., Martnez-Lumbreras, S., Thapaliya, A., High, S., and Isaacson, R. L. (2014) Solution structure of the SGTA dimerisation domain and investigation of its interactions with the ubiquitin-like domains of BAG6 and UBL4A. *PLoS One* **9**, e113281–19

42. Dutta, S., and Tan, Y.-J. (2008) Structural and functional characterization of human SGT and its interaction with Vpu of the human immunodeficiency virus type 1. *Biochemistry* **47**, 10123–10131

43. Leznicki, P., Korac-Prlic, J., Kliza, K., Husnjak, K., Nyathi, Y., Dikic, I., and High, S. (2015) Binding of SGTA to Rpn13 selectively modulates protein quality control. *J. Cell Sci.* **128**, 3187–3196

44. Waheed, A. A., MacDonald, S., Khan, M., Mounts, M., Swiderski, M., Xu, Y., Ye, Y., and Freed, E. O. (2016) The Vpu-interacting protein SGTA regulates expression of a non- glycosylated tetherin species. *Sci. Rep.* **6**, 1–14

45. Luo, P., and Baldwin, R. L. (1997) Mechanism of helix induction by trifluoroethanol: A framework for extrapolating the helix-forming properties of peptides from trifluoroethanol/water mixtures back to water. *Biochemistry* **36**, 8413–8421

46. Dyson, H. J., and Wright, P. E. (2004) Unfolded proteins and protein folding studied by NMR. *Chem. Rev.* **104**, 3607–3622

47. Martínez-Lumbreras, S., Krysztofinska, E. M., Thapaliya, A., Spilotros, A., Matak-Vinkovic, D., Salvadori, E., Roboti, P., Nyathi, Y., Muench, J. H., Roessler, M. M., Svergun, D. I., High, S., and Isaacson, R. L. (2018) Structural complexity of the co-chaperone SGTA: A conserved C-terminal region is implicated in dimerization and substrate quality control. *BMC Biol.* **16**, 76

48. Xu, D., and Zhang, Y. (2012) *Ab initio* protein structure assembly using continuous structure fragments and optimized knowledge-based force field. *Proteins* **80**, 1715–1735

49. Chio, U. S., Chung, S., Weiss, S., and Shan, S.-O. (2019) A chaperone lid ensures efficient and privileged client transfer during tail-anchored protein targeting. *Cell Rep.* **26**, 37–44.e7

50. Shao, S., Rodrigo-Brenni, M. C., Kivlen, M. H., and Hegde, R. S. (2017) Mechanistic basis for a molecular triage reaction. *Science* **355**, 298–302

51. Chun, E., Thompson, A. A., Liu, W., Roth, C. B., Griffith, M. T., Katritch, V., Kunken, J., Xu, F., Cherezov, V., Hanson, M. A., and Stevens, R. C.

(2012) Fusion partner toolchest for the stabilization and crystallization of G protein-coupled receptors. *Structure* **20**, 967–976

52. Letunic, I., Doerks, T., and Bork, P. (2015) Smart: Recent updates, new developments and status in 2015. *Nucleic Acids Res.* **43**, D257–D260

53. Schmid, A. B., Lagleder, S., Gräwert, M. A., Röhl, A., Hagn, F., Wandinger, S. K., Cox, M. B., Demmer, O., Richter, K., Groll, M., Kessler, H., and Buchner, J. (2012) The architecture of functional modules in the Hsp90 co-chaperone Sti1/Hop. *EMBO J.* **31**, 1506–1517

54. Li, Z., Hartl, F. U., and Bracher, A. (2013) Structure and function of Hip, an attenuator of the Hsp70 chaperone cycle. *Nat. Struct. Mol. Biol.* **20**, 929–935

55. Nelson, G. M., Huffman, H., and Smith, D. F. (2003) Comparison of the carboxy-terminal DP-repeat region in teh co-chaperones Hop and Hip. *EMBO J.* **8**, 125–133

56. Rao, M., Okreglak, V., Chio, U. S., Cho, H., Walter, P., and Shan, S.-O. (2016) Multiple selection filters ensure accurate tail-anchored membrane protein targeting. *Elife* **5**, e21301–e21324

57. Pace, C. N., and Scholtz, J. M. (1998) A helix propensity scale based on experimental studies of peptides and proteins. *Biophys. J.* **75**, 422–427

58. Röhl, A., Wengler, D., Madl, T., Lagleder, S., Tippel, F., Herrmann, M., Hendrix, J., Richter, K., Hack, G., Schmid, A. B., Kessler, H., Lamb, D. C., and Buchner, J. (2015) Hsp90 regulates the dynamics of its cochaperone Sti1 and the transfer of Hsp70 between modules. *Nat. Commun.* **6**, 6655

59. Coto, A. L. S., Seraphim, T. V., Batista, F. A. H., Dores-Silva, P. R., Barranco, A. B. F., Teixeira, F. R., Gava, L. M., and Borges, J. C. (2018) Structural and functional studies of the Leishmania braziliensis SGT co-chaperone indicate that it shares structural features with HIP and can interact with both Hsp90 and Hsp70 with similar affinities. *Int. J. Biol. Macromol.* **118**, 693–706

60. Voorhees, R. M., and Hegde, R. S. (2015) Structures of the scanning and engaged states of the mammalian SRP-ribosome complex. *Elife* **4**, e07975

61. Keenan, R. J., Freymann, D. M., Walter, P., and Stroud, R. M. (1998) Crystal structure of the signal sequence binding subunit of the signal recognition particle. *Cell* **94**, 181–191

62. Clemons, W. M., Gowda, K., Black, S. D., Zwieb, C., and Ramakrishnan, V. (1999) Crystal structure of the conserved subdomain of human protein SRP54M at 2.1 A resolution: Evidence for the mechanism of signal peptide binding. *J. Mol. Biol.* **292**, 697–705

63. Tidow, H., and Nissen, P. (2013) Structural diversity of calmodulin binding to its target sites. *FEBS J.* **280**, 5551–5565

64. Guna, A., Volkmar, N., Christianson, J. C., and Hegde, R. S. (2018) The ER membrane protein complex is a transmembrane domain insertase. *Science* **359**, 470–473

65. Suloway, C. J., Rome, M. E., and Clemons, W. M. (2011) Tail-anchor targeting by a Get3 tetramer: The structure of an archaeal homologue. *EMBO J.* **31**, 707–719

66. Chu, R., Takei, J., Knowlton, J. R., Andrykovitch, M., Pei, W., Kajava, A. V., Steinbach, P. J., Ji, X., and Bai, Y. (2002) Redesign of a four-helix bundle protein by phage display coupled with proteolysis and structural characterization by NMR and X-ray crystallography. *J. Mol. Biol.* **323**, 253–262

67. Studier, F. W. (2005) Protein production by auto-induction in high-density shaking cultures. *Protein Expr. Purif.* **41**, 207–234

68. Micsonai, A., Wien, F., Kernya, L., Lee, Y.-H., Goto, Y., Réfrégiers, M., and Kardos, J. (2015) Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E3095–E3103

69. Schneider, C. A., Rasband, W. S., and Eliceiri, K. W. (2012) NIH image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675

70. Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., Potter, S. C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G. A., Tate, J., and Bateman, A. (2016) The Pfam protein families database: Towards a more sustainable future. *Nucleic Acids Res.* **44**, D279–D285

71. Li, W., and Godzik, A. (2006) Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659

72. Katoh, K., and Standley, D. M. (2013) MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780

73. Pei, J., Kim, B.-H., and Grishin, N. V. (2008) PROMALS3D: A tool for multiple protein sequence and structure alignments. *Nucleic Acids Res.* **36**, 2295–2300

74. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., and Barton, G. J. (2009) Jalview version 2–a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191

75. Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., and Zhang, Y. (2015) The I-TASSER suite: Protein structure and function prediction. *Nat. Methods* **12**, 7–8

76. Wallner, B. (2006) Identification of correct regions in protein models using structural, alignment, and consensus information. *Protein Sci.* **15**, 900–913

77. Bradley, P., Misura, K. M., and Baker, D. (2005) Towards high-resolution de novo structure prediction for small proteins. *Science* **309**, 1868–1871

78. Yang, J., Anishchenko, I., Park, H., Peng, Z., Ovchinnikov, S., and Baker, D. (2019) Improved protein structure prediction using predicted inter-residue orientations. *bioRxiv* **13**, e1005324–24

79. Reißer, S., Strandberg, E., Steinbrecher, T., and Ulrich, A. S. (2014) 3D hydrophobic moment vectors as a tool to characterize the surface polarity of amphiphilic peptides. *Biophys. J.* **106**, 2385–2394

80. Pierce, B. G., Wiehe, K., Hwang, H., Kim, B. H., Vreven, T., and Weng, Z. (2014) ZDOCK server: Interactive docking prediction of protein-protein complexes and symmetric multimers. *Bioinformatics* **30**, 1771–1773

81. McGibbon, R. T., Beauchamp, K. A., Harrigan, M. P., Klein, C., Swails, J. M., Hernández, C. X., Schwantes, C. R., Wang, L.-P., Lane, T. J., and Pande, V. S. (2015) MDTraj: A modern open library for the analysis of molecular dynamics trajectories. *Biophys. J.* **109**, 1528–1532

82. Webb, B., and Sali, A. (2002) *Comparative Protein Structure Modeling Using MODELLER* (Vol 235,. John Wiley & Sons, Inc, Hoboken, NJ: 5.6.1–5.6.37

83. Drozdetskiy, A., Cole, C., Procter, J., and Barton, G. J. (2015) JPred4: A protein secondary structure prediction server. *Nucleic Acids Res.* **43**, W389–W394

84. Schägger, H. (2006) Tricine–SDS-PAGE. *Nat. Protoc.* **1**, 16–22

85. Kyte, J., and Doolittle, R. F. (1982) A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105–132

86. Fernandez-Patron, C., Hardy, E., Sosa, A., Seoane, J., and Castellanos, L. (1995) Double staining of coomassie blue-stained polyacrylamide gels by imidazole-sodium dodecyl sulfate-zinc reverse staining: Sensitive detection of coomassie blue-undetected proteins. *Anal. Biochem.* **224**, 263–269

87. Bateman, A., Martin, M. J., O'Donovan, C., Magrane, M., Alpi, E., Antunes, R., Bely, B., Bingley, M., Bonilla, C., Britto, R., Bursteinas, B., Bye-A-Jee, H., Cowley, A., Silva, A. D., Giorgi, M. D., *et al.* (2017) Uni-Prot: The universal protein knowledgebase. *Nucleic Acids Res.* **45**, D158–D169

ASBMB