The client-binding domain of the cochaperone Sgt2 has a helical-hand structure that binds a short hydrophobic helix

(120 of 120)

Ku-Feng Lin, Michelle Y. Fry*, Shyam M. Saladi*, and William M. Clemons, Jr.

* contributed equally to this work

Correspondence: W.M.C, clemons@caltech.edu
Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California 91125

Keywords: Tail-anchored proteins, Protein targeting, co-chaperones, Sti1, Hop

Abstract (150 words – currently 149)

The targeting and insertion of tail-anchored (TA) integral membrane proteins (IMP) into the correct membrane is critical for cellular homeostasis. The fungal protein Sgt2, and its human homolog SGTA, binds hydrophobic clients and is the entry point for targeting of ER-bound TA IMPs. Here we reveal molecular details that underlie the mechanism of Sgt2 binding to TA IMP clients. We establish that the Sgt2 C-terminal region is flexible but conserved and sufficient for client binding. A molecular model for this domain reveals a helical hand forming a hydrophobic groove, consistent with a higher affinity for TA IMP clients with hydrophobic faces and a minimal length of 11 residues. This work places Sgt2 into a broader family of TPR-containing co-chaperone proteins.

Introduction

An inherently complicated problem of cellular homeostasis is the biogenesis of hydrophobic 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 targeted post-translationally to either the 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) [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 (ySgt2) and *Homo sapiens* SGTA (referred to here as hSgt2 for simplicity), collectively Sgt2, 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 endoplasmic reticulum (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 (human Get1/Get2) receptor at the ER leading to release of the TA client into the membrane [22-26]. Deletions of yeast GET genes (*i.e. get1*Δ, *get2*Δ, or *get3*Δ) cause cytosolic aggregation of TA clients dependent on Sgt2 [26,27].

61 In addition to targeting TA IMPs, there is evidence hSgt2 promotes degradation of IMPs
62 through the proteasome by cooperating with the Bag6 complex, a heterotrimer containing Bag6,
63 hGet4, and hGet5, which acts as a central hub for a diverse physiological network related to protein
64 targeting and quality control [19,28-30]. The Bag6 complex can associate with ER membrane-
65 embedded ubiquitin regulatory protein UbxD8, transmembrane protein gp78, proteasomal
66 component Rpn10c, and an E3 ubiquitin protein ligase RNF126 thereby connecting hSgt2 to ER
67 associated degradation (ERAD) and proteasomal activity. Depletion of hSgt2 significantly inhibits
68 turnover of ERAD IMP clients and elicits the unfolded protein response[16]. Furthermore, the
69 cellular level of MLPs in the cytoplasm could be maintained by co-expression with hSgt2, which
70 possibly antagonize ubiquitination of MLPs to prevent proteasomal degradation [15,17]. These
71 studies demonstrate an active role of hSgt2 in triaging IMPs in the cytoplasm and the breadth of
72 hSgt2 clients including TA IMPs, ERAD, and MLPs all harboring one or more TMD. Roles for
73 hSgt2 in disease include polyomavirus infection [31], neurodegenerative disease [27,32],
74 hormone-regulated carcinogenesis [33,34], and myogenesis [35], although the underlying
75 molecular mechanisms are still unclear.
76 The architecture of Sgt2 includes three structurally independent domains that define the three
77 different interactions of Sgt2 (Fig. 1*A*) [12,36-39]. The N-terminal domain forms a homo-dimer
78 composed of a four-helix bundle with 2-fold symmetry that primarily binds to the ubiquitin-like
79 domain (UBL) of Get5/Ubl4A for TA IMP targeting [36,40] or interacts with the UBL on the N-
80 terminal region of BAG6 [41] where it is thought to initiate downstream degradation processes
81 [15,28,29]. The central region comprises a co-chaperone domain with three repeated TPR motifs
82 arranged in a right handed-superhelix forming a 'carboxylate clamp' for binding the C-terminus
83 of heat-shock proteins (HSP) [12,42]. The highly conserved TPR domain was demonstrated to be
84 critical in modulating propagation of yeast prions by recruiting HSP70 [27] and may associate
85 with the proteasomal factor Rpn13 to regulate MLPs [43]. More recently, it was demonstrated that
86 mutations to residues in the TPR domain which prevent Hsp70 binding impair the loading of TA
87 clients onto ySgt2 [21], consistent with a direct role of Hsp70 in TA IMP targeting via the TPR
88 domain. The C-terminal methionine-rich domain of Sgt2 is responsible for binding to hydrophobic
89 clients such as TA IMPs [11,37,44]. Other hydrophobic segments have been demonstrated to
90 interact with this domain such as the membrane protein Vpu (viral protein U) from human
91 immunodeficiency virus type-1 (HIV1), the TMD of tetherin [44], the signal peptide of myostatin

92 [35], and the N-domain of the yeast prion forming protein Sup35 [27]. All of these studies suggest
93 that the C-terminus of Sgt2 binds broadly to hydrophobic stretches, yet structural and mechanistic
94 information for client recognition is lacking.

95 In this study, we provide the first structural characterization of the C-domains from Sgt2
96 (Sgt2-C) and show that, in the absence of substrate, it is relatively unstructured. We demonstrate
97 that a conserved region of the C-domain, defined here as C_{cons}, is sufficient for client binding.
98 Analysis of the C_{cons} sequence identifies six amphipathic helices whose hydrophobic residues are
99 required for client binding. Based on this, we computationally generate an *ab initio* structural
100 model that is validated by point mutants and disulfide crosslinking. Artificial TA clients are then
101 used to define the properties within clients critical for binding to Sgt2-C. The results show that
102 Sgt2-C falls into a larger STI1 family of TPR-containing co-chaperones and allow us to propose a
103 mechanism for TA client binding.
104

105 ## Results

106 ### The flexible Sgt2-C domain

107 Based on sequence alignment (Fig. 1*A*), the Sgt2-C contains a conserved core of six predicted
108 helices flanked by unstructured loops that vary in length and sequence. Previous experimental
109 work suggested that this region is particularly flexible, as this domain in the *Aspergillus fumigatus*
110 is sensitive to proteolysis [12]. Similarly, for ySgt2-TPR-C, the sites sensitive to limited
111 proteolysis primarily occur within the loops flanking the conserved helices (Fig. 1*A, red arrows*
112 and S1*B*). This flexible nature of the C-domain likely contributes to its anomalous passage through
113 a gel-filtration column where Sgt2-C elutes much earlier than the similarly-sized, but well-folded
114 Sgt2 TPR-domain (Fig. 1*B*). The larger hydrodynamic radius matches previous small-angle X-ray
115 scattering measurement of the ySgt2 TPR-C domain that indicated a partial unfolded characteristic
116 in a Kratky plot analysis. The circular dichroism (CD) spectra for both homologs suggests that the
117 C-domain largely assumes a random-coil conformation, with 40-45% not assignable to a defined
118 secondary structure category (Fig. 1*C*) [45]. The well-resolved, sharp, but narrowly dispersed
119 chemical shifts of the backbone amide protons in ^{1}H-^{15}N HSQC spectra of Sgt2-C (Fig. 1*D,E*),
120 indicate a significant degree of backbone mobility, similar to natively unfolded proteins [46] and
121 consistent with results seen by others [47], further highlighting the lack of stable tertiary structure.
122 [12]. Taken all together, Sgt2-C appears to be a flexible domain.

123

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

125 We then asked if the flexible Sgt2-C is the site of client binding in the co-chaperone and if so,
126 where within this domain is the binding region. During purification Sgt2-C is susceptible to
127 proteolytic activity being cut at several specific sites (Fig. 1*A*). Proteolysis occurred primarily at
128 Leu_{327} and in the poorly conserved N-terminal region (between Asp_{235}-Gly_{258}). Given the
129 intervening region, between Gly_{258} and Leu_{327} on ySgt2, is conserved (Fig 1*A*), it and the
130 corresponding region on hSgt2, may mediate TA client binding (Fig. 2*A, grey*). To test this, we
131 established a set of his-tagged Sgt2 constructs of various lengths (Fig. 2*C*). These Sgt2-C mutants
132 were co-expressed with an MBP-tagged TA client, Sbh1, and binding was detected by the presence
133 of captured TA clients in nickel elution fractions (Fig. 2*B*). As previously seen [13], we confirm
134 that Sgt2-TPR-C alone is sufficient for capturing a TA client (Fig. 2*C*). As one might expect, the
135 C-domain was also sufficient for binding the TA client. A predicted six α-helical methionine-rich
136 region of Sgt2-C (Fig. 1*A*), hereafter referred to as Sgt2-C_{cons}, is sufficient for binding to Sbh1.
137 For ySgt2, a minimal region H1-H5 (ΔH0) poorly captures Sbh1, while for hSgt2 the equivalent
138 minimal region is sufficient for capturing the client at a similar level as the longer C_{cons} domain
139 (Fig. 2*C*). The predicted helices in Sgt2-C_{cons} are amphipathic and their hydrophobic patches could
140 be used for client binding (Fig. 2*D*).

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

152

153 **Molecular modeling of Sgt2-C domain**

154 Despite the need for a molecular model, the C-domain has resisted structural studies, likely
155 due to the demonstrated inherent flexibility. Based on the six conserved α-helical amphipathic
156 segments (Fig. 1*A*) that contain hydrophobic residues critical for TA client binding (Fig. 2*D-E),*
157 we expect some folded structure to exist. Therefore, we performed *ab initio* molecular modeling
158 of Sgt2-C using a variety of prediction methods resulting in a diversity of putative structures [48-
159 52]. As expected, all models showed buried hydrophobic residues as this is a major criterion for *in*
160 *silico* protein folding. Residues outside the ySgt2-C_{cons} region adopted varied conformations
161 consistent with their expected higher flexibility. Pruning these N- and C-terminal regions to focus
162 on the ySgt2-C_{cons} region (Fig. S2*A*) revealed a potential binding interface for a hydrophobic
163 substrate, examples are seen in Quark models (1, 4, & 6 shown), Robetta 1 & 2, and I-TASSER 2
164 & 3, whereas others models had no clearly distinguishable groove. Given the intrinsic flexibility
165 of the Sgt2-C domain, it is possible that models without a groove are found in the non-TMD bound
166 structural ensemble.
167 For a working model of TMD-bound ySgt2-C, we chose the highest scored Quark structures
168 where a general consistent architecture was seen (Fig. 3*A*) [48]. The overall model contained a
169 potential TA client binding site, a hydrophobic groove formed by the amphipathic helices. The
170 groove is approximately 15 Å long, 12 Å wide, and 10 Å deep, which is sufficient to accommodate
171 three helical turns of an α-helix, ~11 amino acids (Fig. 3*B*).
172 To validate the model, we interrogated the accuracy of the predicted structural arrangement
173 by determining distance constraints from crosslinking experiments. We selected four pairs of
174 residues in close spatial proximity and one pair far apart based on the Quark models (Fig. 4*A*).
175 Calculating a C_β-C_β distance between residue pairs for each model (Fig. 4*E*), the Quark models 2
176 and 3 were the most consistent with an expected distance of 9Å or less for the close pairs. In all
177 alternative models, the overall distances are much larger and should not be expected to form
178 disulfide bonds *in vitro* if they represent a TMD-bound state. For Robetta, a number of the models
179 have pairs of residues within 9Å and Robetta's per-residue error estimate suggests relatively high
180 confidence in the C_{cons} region (Fig. S2*B*).
181 As a control, we first confirmed that the cysteine-mutant pairs do not affect the function of
182 ySgt2. We utilized an *in vitro* transfer assay where a yeast Hsp70 homolog Ssa1 loaded with a TA
183 client delivers the client to ySgt2 [21,49,50] (Fig. 4*C*). Purified Ssa1 is mixed with detergent
184 solubilized strep-tagged Bos1-TMD (a model ER TA client) that contained a p-benzoyl-l-

185 phenylalanine (BPA) labeled residue, Bos1$_{BPA,}$ are diluted to below the critical micelle
186 concentration resulting in soluble complexes of Bos1$_{BPA}$/Ssa1. Full-length ySgt2 variants were
187 each tested for the ability to capture Bos1$_{BPA}$ from Ssa1. After the transfer reaction, each was UV-
188 treated to generate Bos1 crosslinks. Successful capture of the TA clients by ySgt2 was detected
189 using an anti-strep Western blot and the appearance of a Bos1$_{BPA}$/ySgt2 crosslink band (Fig. 4*D*).
190 All of the cysteine variants of ySgt2 successfully captured Bos1$_{BPA}$ from Ssa1 similar to wild-type
191 suggesting that the cysteine mutations did not affect the structure or function of ySgt2.

192 For the distance experiment, each of the cysteine-mutant pairs was made in ySgt2-TPR-C
193 which lacks the dimerization domain. Each variant was coexpressed with an artificial TA client, a
194 cMyc-tagged BRIL (small, 4-helix bundle protein [51]) with a C-terminal TMD consisting of eight
195 leucines and three alanines, denoted 11[L8], and purified via nickel-affinity chromatography in
196 reducing buffer (Fig. S3*A*). All of the ySgt2 mutants bound the TA client and behaved similar to
197 the wild-type (cysteine-free) further suggesting the mutants did not perturb the native structure
198 (Fig. S3*B*). For disulfide crosslink formation, each eluate was oxidized and crosslinks were
199 identified by the visualization of a reducing-agent sensitive ~7.7kDa fragment in gel
200 electrophoresis (Fig. 3*B*). For both the wild-type construct and in N285C/G329C, where the pairs
201 are predicted from the Quark models to be too distant for disulfide bond formation, no higher
202 molecular weight band was observed. For the remaining pairs that are predicted to be close enough
203 for bond formation, the 7.7kDa fragment was observed in each case and is labile in reducing
204 conditions. Again, these results support the C_{cons} model derived from Quark.

205 With the four crosslinked pairs as distance constraints, new models were generated using
206 Robetta with a restraint on the corresponding pairs of C_{β} atoms less than 9Å (Fig. S4*A*). The
207 Robetta models from these runs are similar to the top scoring models from Quark (Fig. 3).
208 Satisfyingly, the pair of residues that do not form disulfide crosslinks are generally consistent (Fig.
209 S4*B*).

210 The improvement of the ySgt2 models predicted by Robetta with restraints included
211 encouraged us to generate models for hSgt2-C with constraints. For this, pairs were defined based
212 on sequence alignments of Sgt2 (Fig. 1*A*) and used as restraints. The resulting predictions had
213 architectures consistent with the equivalent regions predicted for ySgt2-C_{cons}, for example Robetta
214 4 (Fig. S4*C*, top). Although in general the predicted hSgt2 model is similar to that for ySgt2, the
215 region that corresponds to H2 occupies a position that precludes a clear hydrophobic groove. For

216 ySgt2, the longer N-terminal loop occupies the groove preventing the exposure of hydrophobics
217 to solvent (Fig. 3*C*, grey). For hSgt2, the shorter N-terminal loop may not be sufficient to similarly
218 occupy the groove and allowing for the clear hydrophobic hand seen for the ySgt2-C. To correct
219 for this, we replaced the sequence of the N-terminal loop of hSgt2-C with the ySgt2-C loop and
220 ran structure prediction with the pairwise distance restraints. This resulted in a model where the
221 loop occupies the groove and, when pruned away suggests the hydrophobic hand seen in yeast
222 (Fig. S4*C*, middle boxed). Of note, we also generated models of hSgt2-C using the most recent
223 Robetta method (transform-restrained) which produces new structures with a groove and similar
224 helical-hand architecture across the board (Fig. S4*C*, bottom).
225 We sought to further test the robustness of our model considering the intrinsic flexibility of
226 Sgt2-C by probing for disulfide bond formation with neighboring residues of one of our
227 crosslinking pairs. While the C_β-C_β distance puts these adjacent pairs at farther than 9Å, mutating
228 residues to cystines and measuring S-S distances across all possible pairs of rotamers provides a
229 wider interval on possible distances and, therefore, the likelihood a disulfide bond will form (Fig.
230 4*F*). Cysteine mutants were introduced to the residues adjacent to M289 and A319 in ySgt2-TPR-
231 C for four additional pairs: K288C/A319C, M290C/A319C, M289C/P318C, and M289C/L320C.
232 As described previously, these mutants were coexpressed with a TA substrate, in this case the
233 artificial BRIL-11[L8] which has a MBP-tag instead of a cMyc-tag. Complexes were purified by
234 amylose and then nickel affinity chromatography to ensure eluates contained only Sgt2-TPR-C
235 bound to substrate. Eluates were incubated in oxidizing conditions, quenched with 50mM NEM,
236 and digested with Glu-C protease. Again, a reductant sensitive band at 7.7kDa is observed for
237 each of these adjacent pairs. While the geometry of each of these C-C pairs might suggest against
238 disulfide bond formation, given the intrinsic flexibility of Sgt2-C, it is not surprising that each of
239 these pairs are able to form disulfide bonds. As before, disulfide bond formation was detected for
240 the M289C/A319C pair. In this new construct, we now see a small amount of disulfide bond
241 formation in the distant N285C/G329C pair, likely an effect of switching to the MBP tag.
242
243 **Structural similarity of Sgt2-C domain to STI1 domains**
244 Attempts to glean functional insight for Sgt2-C from BLAST searches did not reliably return
245 other families or non-Sgt2 homologs making functional comparisons difficult. A more extensive
246 profile-based search using hidden Markov models from the SMART database [52] identified a

247 similarity to domains in the yeast co-chaperone Sti1 (HOP in mammals). First called DP1 and DP2,
248 due to their prevalence of aspartates (D) and prolines (P), these domains have been shown to be
249 required for client-binding by Sti1 [53,54] and are termed 'STI1'-domains in bioinformatics
250 databases [52]. In yeast Sti1 and its human homolog HOP (combined will be referred to here as
251 Sti1), each of the two STI1 domains (DP1 and DP2) are preceded by Hsp70/90-binding TPR
252 domains, similar to the domain architecture of Sgt2. Deletion of the second, C-terminal STI1-
253 domain (DP2) from Sti1 *in vivo* is detrimental, impairing native activity of the glucocorticoid
254 receptor [53]. *In vitro*, removal of the DP2 domain from Sti1 results in the loss of recruitment of
255 the progesterone receptor to Hsp90 without interfering in Sti1-Hsp90 binding [55]. These results
256 implicate DP2 in binding of Sti1 clients. In addition, others have noted that, broadly, STI1-domains
257 may present a hydrophobic groove for binding the hydrophobic segments of a client [53,54].
258 Furthermore, the similar domain organizations (*i.e.* Sgt2 TPR-C, Sti1 TPR-STI1) and molecular
259 roles could imply an evolutionary relationship between these co-chaperones. Indeed, a multiple
260 sequence alignment of the Sgt2-C_{cons} with several yeast STI1 domains (Fig. 5*A*) reveals strong
261 conservation of structural features. H1-H5 of the predicted helical regions in C_{cons} align directly
262 with the structurally determined helices in the DP2 domain of Sti1; this includes complete
263 conservation of helix breaking prolines and close alignment of hydrophobic residues in the
264 amphipathic helices [53].
265 Based on the domain architecture and homology, a direct comparison between the STI1
266 domain and Sgt2-C_{cons} can be made. A structure of DP2 solved by solution NMR reveals that the
267 five amphipathic helices assemble to form a flexible helical-hand with a hydrophobic groove [53].
268 The lengths of the α-helices in this structure concur with those inferred from the alignment in Fig.
269 4*A*. Our molecular model of Sgt2-C_{cons} is strikingly similar to this DP2 structure (Fig. 5*B,C*). An
270 overlay of the DP2 structure and our molecular model demonstrates both Sgt2-C_{cons} and DP2 have
271 similar lengths and arrangements of their amphipathic helices (Fig. 5*D*). Consistent with our
272 observations of flexibility in Sgt2-C_{cons}, Sti1-DP2 generates few long-range NOEs between its
273 helices indicating that Sti1-DP2 also has a flexible architecture [53]. We consider this flexibility a
274 feature of these helical-hands for reversible and specific binding of a variety of clients.
275
276 **Binding mode of TA clients to Sgt2**
277 We examined the Sgt2-C_{cons} surface that putatively interacts with TA clients by constructing

278 hydrophobic-to-charge residue mutations that are expected to disrupt capture of TA clients by Sgt2.
279 Similar to the helix mutations in Fig. 2, the capture assay was employed to establish the relative
280 effects of individual mutations. A baseline was established based on the amount of the TA client
281 Sbh1 captured by wild-type Sgt2-TPR-C. In each experiment, Sbh1 was expressed at the same
282 level; therefore, differences in binding should directly reflect the affinity of Sgt2 mutants for clients.
283 In all cases, groove mutations from hydrophobic to aspartate led to a reduction in TA client binding
284 (Fig. 6). The effects are most dramatic with ySgt2 where each mutant significantly reduced binding
285 by 60% or more (Fig. 6*A*). While all hSgt2 individual mutants saw a significant loss in binding,
286 the results were more subtle with the strongest a ~36% reduction (M233D, Fig. 6*B*). Double
287 mutants were stronger with a significant decrease in binding relative to the individual mutants,
288 more reflective of the individual mutants in ySgt2. As seen before (Fig. 2*E&F*), we observe that
289 mutations toward the N-terminus of Sgt2-C have a stronger effect on binding than those later in
290 the sequence.
291
292 **Sgt2-C domain binds clients with a hydrophobic segment ≥ 11 residues**
293 With a molecular model for ySgt2-C_{cons} and multiple lines of evidence for a hydrophobic
294 groove, we sought to better understand the specific requirements for TMD binding. We and others
295 have demonstrated that a monomeric C-domain from Sgt2 is sufficient for binding to TA clients
296 [13]. To study the minimal constraints on TA client binding, we chose to focus on a monomeric
297 construct of Sgt2 (Sgt2-TPR-C) binding to variable TMDs. TA clients were designed where the
298 overall (sum) and average (mean) hydrophobicity, length, and the distribution of hydrophobic
299 character were varied in the TMDs. These artificial TMDs, a Leu/Ala helical stretch followed by
300 a Trp, were constructed as C-terminal fusions to the soluble protein BRIL (Fig. 7*A*). The total and
301 mean hydrophobicity are controlled by varying the helix-length and the Leu/Ala ratio. For clarity,
302 we define a syntax for the various artificial TA clients to highlight the various properties under
303 consideration: hydrophobicity, length, and distribution. The generic notation is TMD-
304 length[number of leucines] which is represented, for example, as 18[L6] for a TMD of 18 amino
305 acids containing six leucines.
306 Our first goal with the artificial clients was to define the minimal length of a TMD to bind to
307 the C-domain. As described earlier, captures of his-tagged Sgt2-TPR-C with the various TA clients
308 were performed. We define a relative binding efficiency as the ratio of captured TA client by a

309 Sgt2-TPR-C normalized to the ratio of a captured wild-type TA client by Sgt2-TPR-C. In this case
310 we replaced the TMD in our artificial clients with the native TMD of Bos1 ($Bos1_{TMD}$). The
311 artificial client 18[L13] shows a comparable binding efficiency to Sgt2-TPR-C as that of $Bos1_{TMD}$
312 (Fig. 7*B*). From the helical wheel diagram of the TMD for Bos1, we noted that the hydrophobic
313 residues favored one face of the helix. We explored this 'hydrophobic face' by using model clients
314 that maintained this orientation while shortening the length and maintaining the average
315 hydrophobicity of 18[L13] (Fig. 7*B*). Shorter helices of 14 or 11 residues, 14[L10] and 11[L8],
316 also bound with similar affinity to Bos1. Helices shorter than 11 residues, 9[L6] and 7[L5], were
317 not able to bind Sgt2-TPR-C (Fig. 7*B*) establishing a minimal length of 11 residues for the helix
318 consistent with the dimensions of the groove predicted from the structural model (Fig. 3). In the
319 context of the full-length Sgt2, which exists as a dimer, an 11-residue cut-off suggests that two C-
320 domains could come together and bind to a single TA whose TMD lengths range from 18-24
321 residues.

322 Since a detected binding event occurs with TMDs of at least 11 amino acids, we decided to
323 probe this limitation further. The dependency of client hydrophobicity was tested by measuring
324 complex formation of Sgt2-TPR-C and artificial TA clients containing an 11 amino acid TMD with
325 increasing number of leucines (11[Lx]). As shown in Fig. 7*C*, increasing the number of leucines
326 monotonically enhances complex formation, echoing previous results [56]. hSgt2-TPR-C binds to
327 a wider spectrum of hydrophobic clients than ySgt2-TPR-C, which could mean it has a more
328 permissive hydrophobic binding groove, also reflected in the milder impact of Ala replacement
329 and Asp mutations in hSgt2-TPR-C to TA client binding (Fig. 2*C* and Fig. 6*A*).
330

331 **Sgt2-C preferentially binds to TMDs with a hydrophobic face**

332 Next, we address the properties within the TMD of TA clients responsible for Sgt2 binding.
333 In the case of ySgt2, it has been suggested that the co-chaperone binds to TMDs based on
334 hydrophobicity and helical propensity [56]. In our system, our artificial TMDs consist of only
335 alanines and leucines which have high helical propensities [57], and despite keeping the helical
336 propensity constant and in a range that favors Sgt2 binding, there is still variation in binding
337 efficiency. For the most part, varying the hydrophobicity of an artificial TA client acts as expected,
338 the more hydrophobic TMDs bind more efficiently to Sgt2 TPR-C (Fig. 7*C*). Our C_{cons} model
339 suggests the hydrophobic groove of ySgt2-C protects a TMD with highly hydrophobic residues

340 clustered to one side (see Fig. 3*B*). To test this, various TMD pairs with the same hydrophobicity,
341 but different distributions of hydrophobic residues demonstrates TA clients with clustered leucines
342 have a higher relative binding efficiency than those with a more uniform distribution (Fig. 7*D*).
343 Helical wheel diagrams demonstrate the distribution of hydrophobic residues along the helix (*e.g.*
344 bottom Fig. 7*D*). The clustered leucines in the TMDs create a hydrophobic face which potentially
345 interacts with the hydrophobic groove formed by the Sgt2-C_{cons} region, corresponding to the model
346 in Fig. 3*B*.
347

348 **Discussion**

349 Sgt2, the most upstream component of the GET pathway, plays a critical role in the targeting
350 of TA IMPs to their correct membranes. Its importance as the first confirmed selection step of ER
351 versus mitochondrial [56] destined TA IMPs necessitates a molecular model for TA client binding.
352 Previous work demonstrated a role for the C-domain of Sgt2 to bind to hydrophobic clients, yet
353 the exact binding domain remained to be determined. Through the combined use of biochemistry,
354 bioinformatics, and computational modeling, we conclusively identify the minimal client-binding
355 domain of Sgt2. Pulldown experiments revealed a minimal six α-helical region as the TMD binding
356 domain of Sgt2 that is also conserved in structure and sequence analyses. This domain
357 preferentially binds TMDs with hydrophobic residues organized onto one face of the helix. While
358 NMR spectroscopy displays an intrinsically disordered domain in the absence of substrate,
359 computational models of ySgt2-C consistently predict a structured helical hand in its conserved
360 region, while the rest of the domain remained flexible and varied between models. Structural
361 similarities between the ySgt2-C model and the STI1 domains DP1 and DP2 now place Sgt2
362 among a class of co-chaperones. Together, these results allow us to present a validated structural
363 model of the Sgt2 C-domain as a methionine-rich helical hand for grasping a hydrophobic helix
364 and to provide a mechanistic explanation for binding a TMD of at least 11 hydrophobic residues.

365 We confidently identify the C-domain of Sgt2 as containing a STI1 domain for client binding
366 through sequence alignments and structural homology. This places Sgt2 into a larger context of
367 conserved co-chaperones (Fig. 8*A*). In the co-chaperone family, the STI1 domains predominantly
368 follow HSP-binding TPR domains connected by a flexible linker, reminiscent of the domain
369 architecture of Sgt2. As noted above, for Sti1 these domains are critical for client-processing and
370 coordinated hand-off between Hsp70 and Hsp90 homologs [58] as well as coordinating the

371 simultaneous binding of two heat shock proteins. Both Sgt2 and the co-chaperone Hip coordinate
372 pairs of TPR and STI1 domains by forming stable dimers via their N-terminal dimerization
373 domains [59]. With evidence for a direct role of the carboxylate-clamp in the TPR domain of Sgt2
374 for client-binding now clear [21], one can speculate that the two TPR domains may facilitate TA
375 client entry into various pathways that use multiple heat shock proteins.

376 Computational modeling reveals a conserved region sufficient for TA client binding that
377 consists of a helical hand of five alpha-helices that form a hydrophobic groove to bind the client
378 TMD. The concept of TMD binding by a helical hand is reminiscent of other proteins involved in
379 membrane protein targeting. Like Sgt2, the signal recognition particle (SRP) contains a
380 methionine-rich domain that binds signal sequences and TMDs. While the helical order is inverted,
381 again five amphipathic helices form a hydrophobic groove that cradles the client signal peptide
382 [60] (Fig. 8*B*). Here once more, the domain has been observed to be flexible in the absence of
383 client [61,62] and, in the resting state, occupied by a region that includes a helix which must be
384 displaced [60]. Another helical-hand example recently shown to be involved in TA IMP targeting
385 is calmodulin where a crystal structure reveals two helical hands coordinating to clasp a TMD at
386 either end (Fig. 8*B*) [63]. Considering an average TMD of 18-20 amino acids (to span a ~40Å
387 bilayer), each half of calmodulin interacts with about 10 amino acids. This is in close
388 correspondence to the demonstrated minimal 11 amino acids for a TA client to bind to the
389 monomeric Sgt2-TPR-C. In the context of the full-length Sgt2, one can speculate that the Sgt2
390 dimer may utilize both C-domains to bind to a full TMD, similar to calmodulin. Cooperation of
391 the two Sgt2 C-domains in client-binding could elicit conformational changes in the complex that
392 could be recognized by downstream factors, such as additional interactions that increase the
393 affinity to Get5/Ubl4A.

394 Intriguingly, Sgt2-TPR-C preferentially binds to artificial clients with clustered leucines. If
395 the C-domain forms a hydrophobic groove as suggested by the computational model, it provides
396 an attractive explanation for this preference. In order to bind to the hydrophobic groove, a client
397 buries a portion of its TMD in the groove leaving the other face exposed. Clustering hydrophobic
398 residues contributes to the hydrophobic effect driving binding efficiency and protecting them from
399 the aqueous environment. Indeed, GET pathway substrates have been suggested to be more
400 hydrophobic TMDs than EMC substrates [64]. Of these, for the most hydrophobic substrates, like
401 Bos1, residues on both sides of the TMD could be protected by a pair of C-domains. Alternatively,

402 the unstructured N-terminal loop through H0 could act as a lid surrounding the circumference of
403 the client's TMD. Unstructured regions participating in substrate binding as well as capping a
404 hydrophobic groove have both suggested in the context of other domains, *e.g.* with Get3 [4]. The
405 role for this clustering of hydrophobic residues in TA client recognition and targeting merits further
406 investigation.

407 What is the benefit of the flexible helical-hand structure for hydrophobic helix binding? While
408 it remains an open question, it is notable that evolution has settled on similar simple solutions to
409 the complex problem of specific but temporary binding of hydrophobic helices. For all of the
410 domains mentioned above, the flexible helical-hands provide an extensive hydrophobic surface to
411 capture a client-helix—driven by the hydrophobic effect. Typically, such extensive interfaces are
412 between pairs of pre-ordered surfaces resulting in high affinities and slow off rates. These helical
413 hands are required to only engage temporarily, therefore the flexibility offsets the favorable free
414 energy of binding by charging an additional entropic cost for ordering a flexible structure in the
415 client-bound complex. The benefit for TA clients is a favorable transfer to downstream components
416 in the GET pathways as seen for ySgt2 [21] and hSgt2 [50]. The demonstration that TA transfer
417 from hSgt2 to Get3 is twice as fast as disassociation from hSgt2 into solution, perhaps interaction
418 with Get3 leads to conformational changes that further favor release [50].

419 While hSgt2 and ySgt2 share many properties, there are a number of differences between the
420 two homologs that may explain the different biochemical behavior. For the C_{cons}-domains, hSgt2
421 appears to be more ordered in the absence of client as the peaks in its NMR spectra are broader
422 (Fig. 1*E*). Comparing the domains at the sequence level, while the high glutamine content in the
423 C-domain is conserved it is higher in hSgt2 (8.8% versus 15.2%). The additional glutamines are
424 concentrated in the predicted longer H4 helix (Fig. 1*A*). The linker to the TPR domain is shorter
425 compared to ySgt2 while the loop between H3 and H4 is longer. Do these differences reflect
426 different roles? As noted, in every case the threshold for hydrophobicity of client-binding is lower
427 for hSgt2 than ySgt2 (Fig. 1*E*, 5, and 6) implying that the mammalian protein is more permissive
428 in client binding. The two C-domains have similar hydrophobicity, so this difference in binding
429 might be due to a lower entropic cost paid by having the hSgt2 C-domain more ordered in the
430 absence of client or the lack of an unstructured N-terminal loop.

431 The targeting of TA clients presents an intriguing and enigmatic problem for understanding
432 the biogenesis of IMPs. How subtle differences in each client modulates the interplay of hand-offs

433 that direct these proteins to the correct membrane remains to be understood. In this study, we focus
434 on a central player, Sgt2 and its client-binding domain. Through biochemistry and computational
435 analysis, we provide a structural model that adds more clarity to client discrimination.

436 # Material and Methods

437 **Plasmid constructs**

438 MBP-Sbh1 full length, *ySgt2*$_{95\text{-}346}$ (*ySgt2*-TPR-C), *ySgt2*$_{222\text{-}346}$ (*ySgt2*-C), *ySgt2*$_{260\text{-}327}$ (*ySgt2*-
439 C$_{cons}$), *ySgt2*$_{266\text{-}327}$ (*ySgt2*-ΔH0), *hSgt2*$_{87\text{-}313}$ (*hSgt2*-TPR-C), *hSgt2*$_{213\text{-}313}$ (*hSgt2*-C), *hSgt2*$_{219\text{-}300}$
440 (*hSgt2*-C$_{cons}$), and *hSgt2*$_{228\text{-}300}$ (*hSgt2*-ΔH0) were prepared as previously described [12,65]. Genes
441 of *ySgt2* or *hSgt2* variants were amplified from constructed plasmids and then ligated into an
442 pET33b-derived vector with a 17 residue N-terminal hexa-histidine tag and a tobacco etch virus
443 (TEV) protease site. Single or multiple mutations on Sgt2 were constructed by site-direct
444 mutagenesis. Artificial TA clients were constructed in a pACYC-Duet plasmid with a N-terminal
445 cMyc tag, BRIL fusion protein [66], GSS linker, and a hydrophobic C-terminal tail consisting of
446 leucines and alanines and ending with a tryptophan.

447 **Protein expression and purification**

448 All proteins were expressed in *Escherichia coli* NiCo21 (DE3) cells (New England BioLabs).
449 To co-express multiple proteins, constructed plasmids were co-transformed as described [65].
450 Protein expression was induced by 0.3 mM IPTG at OD_{600} ~ 0.7 and harvested after 3 hours at
451 37°C. For structural analysis, cells were lysed through an M-110L Microfludizer Processor
452 (Microfluidics) in lysis buffer (50 mM Tris, 300 mM NaCl, 25 mM imidazole supplemented with
453 benzamidine, phenylmethylsulfonyl fluoride (PMSF), and 10 mM β-mercaptoethanol (BME), pH
454 7.5). For capture assays, cells were lysed by freeze-thawing 3 times with 0.1 mg/mL lysozyme. To
455 generate endogenous proteolytic products of *ySgt2*-TPR-C for MS analysis, PMSF and
456 benzamidine were excluded from the lysis buffer. His-tagged Sgt2 and his-tagged Sgt2/TA
457 complexes were separated from the lysate by batch incubation with Ni-NTA resin at 4°C for 1hr.
458 The resin was washed with 20 mM Tris, 150 mM NaCl, 25 mM imidazole, 10 mM BME, pH 7.5.
459 The complexes of interest were eluted in 20 mM Tris, 150 mM NaCl, 300 mM imidazole, 10 mM
460 BME, pH 7.5.

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

464 Measurement of Sgt2 protein concentration was carried out using the bicinchoninic acid
465 (BCA) assay with bovine serum albumin (BSA) as standard (Pierce Chemical Co.). Samples for

466 NMR and CD analyses were concentrated to 10-15 mg/mL for storage at −80°C before
467 experiments.

468 For the *in vitro* transfer assay, plasmids encoding for the full-length ySgt2 cysteine mutants
469 were transformed into BL21 Star cells (Invitrogen). Cells were grown in 2x yeast-tryptone (2xYT)
470 media and induced with 0.1mM IPTG at an OD_{600} of 0.6 then harvested after 3 hours at 30°C by
471 centrifugation. Cells were lysed in 50mM Tris pH 8.0, 500mM NaCl, 10% glycerol, and 1x
472 BugBuster (Millipore Sigma), supplemented with protease inhibitors (4-(2-
473 aminoethyl)benzenesulfonyl fluoride hydrochloride (Roche), benzamidine, and BME). Full-length
474 his-tagged ySgt2 and cysteine mutants were separated from the lysate by batch incubation with
475 Ni-NTA resin (Qiagen) at 4°C for 1 hour. The resin was washed with 50mM Tris pH 8.0, 500mM
476 NaCl, 10% glycerol, and 25mM imidazole and then the protein was eluted in 50mM Tris pH 8.0,
477 500mM NaCl, 10% glycerol, and 300mM imidazole. For storage, protein was dialyzed in 25mM
478 K-HEPES pH 7.5, 150mM KOAc, and 20% glycerol at 4°C and then flash frozen in liquid nitrogen.
479 Purified Bos1 with p-benzoyl-l-phenylalanine (BPA) labeled at residue 230 ($Bos1_{BPA}$) and yeast
480 Ssa1 were gifts from the lab of Shu-ou Shan (Caltech).

481 **NMR Spectroscopy**

482 ^{15}N-labeled proteins were generated from cells grown in auto-induction minimal media as
483 described [67] and purified in 20 mM phosphate buffer, pH 6.0 (for ySgt2-C, 10mM Tris, 100mM
484 NaCl, pH 7.5). The NMR measurements of ^{15}N-labeled Sgt2-C proteins (~0.3-0.5 mM) were
485 collected using a Varian INOVA 600 MHz spectrometer at either 25°C (ySgt2-C) or 35°C (hSgt2-
486 C) with a triple resonance probe and processed with TopSpin™ 3.2 (Bruker Co.).

487 **CD Spectroscopy**

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

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

494 Complexes of the co-expressed wild type or double cysteine mutated His-*ySgt2*-TPR-C and

495 the artificial TA client, 11[L8], with either a cMyc or MBP tag were purified as the other His-Sgt2
496 complexes described above or initially purified via amylose affinity chromatography before nickel
497 chromatography explained earlier. The protein complexes were mixed with 0.2 mM $CuSO_4$ and
498 0.4 mM 1,10-phenanthroline at 24°C for 20 min followed by 50 mM N-ethyl maleimide for 15
499 min. Sequencing-grade Glu-C protease (Sigma) was mixed with the protein samples at an
500 approximate ratio of 1:30 and the digestion was conducted at 37°C for 22 hours. Digested samples
501 were mixed with either non-reducing or reducing SDS-sample buffer, resolved via SDS-PAGE
502 using Mini-Protean® Tris-Tricine Precast Gels (10-20%, Bio-Rad), and visualized using
503 Coomassie Blue staining.

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

505 The *in vitro* transfer assays were performed as in Chio et al. 2019 and Shao et al. 2017 [49,50].
506 Specifically, 39µM $Bos1^{BPA}$ (50mM HEPES, 300mM NaCl, 0.05% LDAO, 20% glycerol) was
507 diluted to a final concentration of 0.1µM and added to 4µM Ssa1 supplemented with 2mM ATP
508 (25mM HEPES pH7.5, 150mM KOAc). After one minute, 0.3µM of full-length ySgt2 or mutant
509 was added to the reaction. Samples were flash frozen after one minute and placed under a 365nm
510 UV lamp for 2 hours on dry ice to allow for BPA crosslinking.

511 **Protein immunoblotting and detection**

512 For western blots, protein samples were resolved via SDS-PAGE and then transferred onto
513 nitrocellulose membranes by the Trans-Blot® Turbo™ Transfer System (Bio-Rad). Membranes
514 were blocked in 5% non-fat dry milk and hybridized with antibodies in TBST buffer (50 mM Tris-
515 HCl pH 7.4, 150 mM NaCl, 0.1% Tween 20) for 1 hour of each step at 24°C. The primary
516 antibodies were used at the following dilutions: 1:1000 anti-penta-His mouse monoclonal (Qiagen),
517 1:5000 anti-cMyc mouse monoclonal (Sigma), and a 1:3000 anti-Strep II rabbit polyclonal
518 (Abcam). A secondary antibody conjugated either to alkaline phosphatase (Rockland, 1:8000) or
519 a 800nm fluorophore was employed, and the blotting signals were chemically visualized with
520 either the nitro-blue tetrazolium/5-bromo-4-chloro-3'-indolyphosphate (NBT/BCIP) chromogenic
521 assay (Sigma) or infrared scanner. All blots were photographed and quantified by image
522 densitometry using ImageJ [69] or ImageStudioLite (LI-COR Biosciences).

523 **Quantification of Sgt2-TA complex formation**

524 The densitometric analysis of MBP-Sbh1 capture by His-Sgt2-TPR-C quantified the intensity
525 of the corresponding protein bands on a Coomassie Blue G-250 stained gel. The quantified signal
526 ratios of MBP-Sbh1/His-Sgt2-TPR-C are normalized to the ratio obtained from the wild-type
527 (WT). Expression level of MBP-Sbh1 was confirmed by immunoblotting the MBP signal in cell
528 lysate. Average ratios and standard deviations were obtained from 3-4 independent experiments.
529 In artificial TA client experiments, both his-tagged Sgt2-TPR-C and cMyc-tagged artificial
530 TA clients were quantified via immunoblotting signals. The complex efficiency of Sgt2-TPR-C
531 with various TA clients was obtained by

532 $$E_{complex} = \frac{E_{TA}}{T_{TA}} \times \frac{1}{E_{capture}} \quad (1)$$

533 where E_{TA} is the signal intensity of an eluted TA client representing the amount of TA client co-
534 purified with Sgt2-TPR-C and T_{TA} is the signal intensity of a TA client in total lysate which
535 corresponds to the expression yield of that TA client. Identical volumes of elution and total lysate
536 from different TA clients experiments were analyzed and quantified. In order to correct for possible
537 variation the amount of Sgt2-TPR-C available for complex formation, $E_{capture}$ represents the
538 relative amount of Sgt2-TPR-C present in the elution (E_{Sgt2}) compared to a pure Sgt2-TPR-C
539 standard ($E_{purified,Sgt2}$).
540

541 $$E_{capture} = \frac{E_{Sgt2}}{E_{purified,Sgt2}}, \quad (2)$$

542

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

546 $$\% \text{ Complex} = \frac{E_{complex,TA}}{E_{complex,Bos1}} \times 100 \quad (3)$$

547 **Sequence alignments**

548 An alignment of Sgt2-C domains was carried out as follows: all sequences with an annotated
549 N-terminal Sgt2/A dimerization domain (PF16546 [70]), at least one TPR hit (PF00515.27,
550 PF13176.5, PF07719.16, PF13176.5, PF13181.5), and at least 50 residues following the TPR
551 domain were considered family members. Putative C-domains were inferred as all residues
552 following the TPR domain, filtered at 90% sequence identity using CD-HIT [71], and then aligned
553 using MAFFT G-INS-i [72]. Other attempts with a smaller set (therefore more divergent) of

554 sequences results in an ambiguity in the relative register of H0, H1, H2, and H3 when comparing
555 Sgt2 with SGTA.
556 Alignments of Sti1 (DP1/DP2) and STI1 domains were created by pulling all unique domain
557 structures with annotated STI1 domains from Uniprot. Where present, the human homolog was
558 selected and then aligned with PROMALS3D [73]. PROMALS3D provides a way of integrating
559 a variety of costs into the alignment procedure, including 3D structure, secondary structure
560 predictions, and known homologous positions.
561 All alignments were visualized using Jalview [74].

562 **Molecular modeling**

563 Putative models for ySgt2-C were generated with I-TASSER, PCONS, Quark, Robetta (*ab*
564 *initio* and transform-restrained modes), Phyre2, and RaptorX via their respective web servers
565 [48,75-78]. The highest scoring model from Quark was then chosen to identify putative TA client
566 binding sites by rigid-body docking of various transmembrane domains modelled as α-helices (3D-
567 HM [79]) into the ySgt2-C_{cons} through the Zdock web server [80]. Pairwise distances were
568 calculated between C_β atoms (the closer C_α proton on glycine) using mdtraj [81] . Based on our
569 disulfide crosslinks, new models were predicted using Robetta in *ab initio* mode specifying C_β-C_β
570 atom distance constraints bounded between 0 and 9 Å.
571 For hSgt2, using the same set of structure prediction servers above, we were only able to
572 produce a clear structural model using the Robetta transform-restrained mode. We were also unable
573 to generate a reliable model by directly using the ySgt2-C model as a template [82]. To crosslink
574 distance data from ySgt2 as restraints for hSgt2, pair positions were transferred from one protein
575 to the other via an alignment of Sgt2-C domains (excerpt in Fig. 1A) and ran Robetta *ab initio*.
576 Also, we grafted the N-terminal loop of ySgt2-C on hSgt2-C with the same set of restraints.
577 Images were rendered using PyMOL 2.3 (www.pymol.org).

578

579 ## Acknowledgements

580 We thank D. G. VanderVelde for assistance with NMR data collection; S. Mayo for providing
581 computing resources; S. Shan, H. J. Cho, Y. Liu, and members of the Clemons lab for support and
582 discussion. We thank J. Mock and A. M. Thinn for comments on the manuscript. This work was

583 supported by the National Institutes of Health (NIH) grants GM105385 and GM097572 (to WMC),
584 NIH/National Research Service Award Training Grant GM07616 (to SMS and MYF), and a
585 National Science Foundation Graduate Research fellowship Grant 1144469 (to SMS).

586 **References**

587

588 1. Pieper U, Schlessinger A, Kloppmann E, Chang GA, Chou JJ, Dumont ME, Fox BG,
589 Fromme P, Hendrickson WA, Malkowski MG, et al. (2013) Coordinating the impact of
590 structural genomics on the human α-helical transmembrane proteome. *Nat Struct Mol Biol*
591 **20**: 135–138.
592 2. Aviram N, Schuldiner M (2017) Targeting and translocation of proteins to the
593 endoplasmic reticulum at a glance. *J Cell Sci* **130**: 4079–4085.
594 3. Shao S, Hegde RS (2011) Membrane Protein Insertion at the Endoplasmic Reticulum.
595 *Annu Rev Cell Dev Biol* **27**: 25–56.
596 4. Guna A, Hegde RS (2018) Transmembrane Domain Recognition during Membrane
597 Protein Biogenesis and Quality Control. *Curr Biol* **28**: R498–R511.
598 5. Kutay U, Hartmann E, Rapoport TA (1993) A class of membrane proteins with a C-
599 terminal anchor. *Trends Cell Biol* **3**: 72–75.
600 6. Hegde RS, Keenan RJ (2011) Tail-anchored membrane protein insertion into the
601 endoplasmic reticulum. *Nat Rev Mol Cell Biol* **12**: 787–798.
602 7. Denic V (2012) A portrait of the GET pathway as a surprisingly complicated young man.
603 *Trends Biochem Sci* **37**: 411–417.
604 8. Wattenberg B, Lithgow T (2001) Targeting of C-Terminal (Tail)-Anchored Proteins:
605 Understanding how Cytoplasmic Activities are Anchored to Intracellular Membranes.
606 *Traffic* **2**: 66–71.
607 9. Chartron JW, Clemons WM, Suloway CJM (2012) The complex process of GETting tail-
608 anchored membrane proteins to the ER. *Curr Opin Struct Biol* **22**: 217–224.
609 10. Chio US, Cho H, Shan S-O (2017) Mechanisms of Tail-Anchored Membrane Protein
610 Targeting and Insertion. *Annu Rev Cell Dev Biol* **33**: 417–438.
611 11. Shao S, Hegde RS (2011) A Calmodulin-Dependent Translocation Pathway for Small
612 Secretory Proteins. *Cell* **147**: 1576–1588.
613 12. Chartron JW, Gonzalez GM, Clemons WM Jr. (2011) A Structural Model of the Sgt2
614 Protein and Its Interactions with Chaperones and the Get4/Get5 Complex. *J Biol Chem*
615 **286**: 34325–34334.
616 13. Wang F, Brown EC, Mak G, Zhuang J, Denic V (2010) A Chaperone Cascade Sorts
617 Proteins for Posttranslational Membrane Insertion into the Endoplasmic Reticulum. *Mol*
618 *Cell* **40**: 159–171.
619 14. Simon AC, Simpson PJ, Goldstone RM, Krysztofinska EM, Murray JW, High S, Isaacson
620 RL (2013) Structure of the Sgt2/Get5 complex provides insights into GET-mediated
621 targeting of tail-anchored membrane proteins. *Proc Natl Acad Sci USA* **110**: 1327–1332.
622 15. Xu Y, Cai M, Yang Y, Huang L, Ye Y (2012) SGTA Recognizes a Noncanonical
623 Ubiquitin-like Domain in the Bag6-Ubl4A-Trc35 Complex to Promote Endoplasmic
624 Reticulum-Associated Degradation. *Cell Rep* **2**: 1633–1644.
625 16. Wunderley L, Leznicki P, Payapilly A, High S (2014) SGTA regulates the cytosolic
626 quality control of hydrophobic substrates. *J Cell Sci* **127**: 4728–4739.
627 17. Leznicki P, High S (2012) SGTA antagonizes BAG6-mediated protein triage. *Proc Natl*
628 *Acad Sci USA* **109**: 19214–19219.

629 18. Gristick HB, Rao M, Chartron JW, Rome ME, Shan S-O, Clemons WM (2014) Crystal
630 structure of ATP-bound Get3–Get4–Get5 complex reveals regulation of Get3 by Get4.
631 *Nat Struct Mol Biol* **21**: 437–442.
632 19. Mock J-Y, Chartron JW, Zaslaver M, Xu Y, Ye Y, Clemons WM Jr. (2015) Bag6
633 complex contains a minimal tail-anchor–targeting module and a mock BAG domain. *Proc*
634 *Natl Acad Sci USA* **112**: 106–111.
635 20. Wang F, Whynot A, Tung M, Denic V (2011) The Mechanism of Tail-Anchored Protein
636 Insertion into the ER Membrane. *Mol Cell* **43**: 738–750.
637 21. Cho H, Shan S-O (2018) Substrate relay in an Hsp70-cochaperone cascade safeguards
638 tail-anchored membrane protein targeting. *EMBO J* **37**:.
639 22. Stefer S, Reitz S, Wang F, Wild K, Pang Y-Y, Schwarz D, Bomke J, Hein C, Löhr F,
640 Bernhard F, et al. (2011) Structural basis for tail-anchored membrane protein biogenesis
641 by the Get3-receptor complex. *Science* **333**: 758–762.
642 23. Rome ME, Chio US, Rao M, Gristick H, Shan S-O (2014) Differential gradients of
643 interaction affinities drive efficient targeting and recycling in the GET pathway. *Proc Natl*
644 *Acad Sci USA* **111**: E4929–E4935.
645 24. Vilardi F, Lorenz H, Dobberstein B (2011) WRB is the receptor for TRC40/Asna1-
646 mediated insertion of tail-anchored proteins into the ER membrane. *J Cell Sci* **124**: 1301–
647 1307.
648 25. Yamamoto Y, Sakisaka T (2012) Molecular Machinery for Insertion of Tail-Anchored
649 Membrane Proteins into the Endoplasmic Reticulum Membrane in Mammalian Cells. *Mol*
650 *Cell* **48**: 387–397.
651 26. Schuldiner M, Metz J, Schmid V, Denic V, Rakwalska M, Schmitt HD, Schwappach B,
652 Weissman JS (2008) The GET Complex Mediates Insertion of Tail-Anchored Proteins
653 into the ER Membrane. *Cell* **134**: 634–645.
654 27. Kiktev DA, Patterson JC, Müller S, Bariar B, Pan T, Chernoff YO (2012) Regulation of
655 chaperone effects on a yeast prion by cochaperone Sgt2. *Mol Cell Biol* **32**: 4960–4970.
656 28. Xu Y, Liu Y, Lee J-G, Ye Y (2013) A Ubiquitin-like Domain Recruits an Oligomeric
657 Chaperone to a Retrotranslocation Complex in Endoplasmic Reticulum-associated
658 Degradation. *J Biol Chem* **288**: 18068–18076.
659 29. Rodrigo-Brenni MC, Gutierrez E, Hegde RS (2014) Cytosolic Quality Control of
660 Mislocalized Proteins Requires RNF126 Recruitment to Bag6. *Mol Cell* **55**: 227–237.
661 30. Hessa T, Sharma A, Mariappan M, Eshleman HD, Gutierrez E, Hegde RS (2011) Protein
662 targeting and degradation are coupled for elimination of mislocalized proteins. *Nature*
663 **475**: 394–397.
664 31. Dupzyk A, Williams JM, Bagchi P, Inoue T, Tsai B (2017) SGTA-Dependent Regulation
665 of Hsc70 Promotes Cytosol Entry of Simian Virus 40 from the Endoplasmic Reticulum. *J*
666 *Virol* **91**: 23.
667 32. Long P, Samnakay P, Jenner P, Rose S (2012) A yeast two-hybrid screen reveals that
668 osteopontin associates with MAP1A and MAP1B in addition to other proteins linked to
669 microtubule stability, apoptosis and protein degradation in the human brain. *Eur J*
670 *Neurosci* **36**: 2733–2742.
671 33. Trotta AP, Need EF, Selth LA, Chopra S, Pinnock CB, Leach DA, Coetzee GA, Butler
672 LM, Tilley WD, Buchanan G (2013) Knockdown of the cochaperone SGTA results in the
673 suppression of androgen and PI3K/Akt signaling and inhibition of prostate cancer cell
674 proliferation. *Int J Cancer* **133**: 2812–2823.

675 34. Buchanan G, Ricciardelli C, Harris JM, Prescott J, Yu ZCL, Jia L, Butler LM, Marshall
676 VR, Scher HI, Gerald WL, et al. (2007) Control of Androgen Receptor Signaling in
677 Prostate Cancer by the Cochaperone Small Glutamine Rich Tetratricopeptide Repeat
678 Containing Protein. *Cancer Res* **67**: 10087–10096.
679 35. Wang H, Zhang Q, Zhu D (2003) hSGT interacts with the N-terminal region of myostatin.
680 *Biochem Biophys Res Commun* **311**: 877–883.
681 36. Chartron JW, VanderVelde DG, Clemons WM Jr. (2012) Structures of the Sgt2/SGTA
682 Dimerization Domain with the Get5/UBL4A UBL Domain Reveal an Interaction that
683 Forms a Conserved Dynamic Interface. *Cell Rep* **2**: 1620–1632.
684 37. Liou S-T, Wang C (2005) Small glutamine-rich tetratricopeptide repeat-containing protein
685 is composed of three structural units with distinct functions. *Arch Biochem Biophys* **435**:
686 253–263.
687 38. Cziepluch C, Kordes E, Poirey R, Grewenig A, Rommelaere J, Jauniaux JC (1998)
688 Identification of a novel cellular TPR-containing protein, SGT, that interacts with the
689 nonstructural protein NS1 of parvovirus H-1. *J Virol* **72**: 4149–4156.
690 39. Callahan MA, Handley MA, Lee YH, Talbot KJ, Harper JW, Panganiban AT (1998)
691 Functional interaction of human immunodeficiency virus type 1 Vpu and Gag with a
692 novel member of the tetratricopeptide repeat protein family. *J Virol* **72**: 5189–5197.
693 40. Winnefeld M, Grewenig A, Schn lzer M, Spring H, Knoch TA, Gan EC, Rommelaere J,
694 Cziepluch C (2006) Human SGT interacts with Bag-6/Bat-3/Scythe and cells with reduced
695 levels of either protein display persistence of few misaligned chromosomes and mitotic
696 arrest. *Exp Cell Res* **312**: 2500–2514.
697 41. Darby JF, Krysztofinska EM, Simpson PJ, Simon AC, Leznicki P, Sriskandarajah N,
698 Bishop DS, Hale LR, Alfano C, Conte MR, et al. (2014) Solution Structure of the SGTA
699 Dimerisation Domain and Investigation of Its Interactions with the Ubiquitin-Like
700 Domains of BAG6 and UBL4A. *PLoS ONE* **9**: e113281–19.
701 42. Dutta S, Tan Y-J (2008) Structural and functional characterization of human SGT and its
702 interaction with Vpu of the human immunodeficiency virus type 1. *Biochemistry* **47**:
703 10123–10131.
704 43. Leznicki P, Korac-Prlic J, Kliza K, Husnjak K, Nyathi Y, Dikic I, High S (2015) Binding
705 of SGTA to Rpn13 selectively modulates protein quality control. *J Cell Sci* **128**: 3187–
706 3196.
707 44. Waheed AA, MacDonald S, Khan M, Mounts M, Swiderski M, Xu Y, Ye Y, Freed EO
708 (2016) The Vpu-interacting Protein SGTA Regulates Expression of a Non- glycosylated
709 Tetherin Species. *Sci Rep* **6**: 1–14.
710 45. Luo P, Baldwin RL (1997) Mechanism of helix induction by trifluoroethanol: a
711 framework for extrapolating the helix-forming properties of peptides from
712 trifluoroethanol/water mixtures back to water. *Biochemistry* **36**: 8413–8421.
713 46. Dyson HJ, Wright PE (2004) Unfolded Proteins and Protein Folding Studied by NMR.
714 *Chem Rev* **104**: 3607–3622.
715 47. Martínez-Lumbreras S, Krysztofinska EM, Thapaliya A, Spilotros A, Matak-Vinkovic D,
716 Salvadori E, Roboti P, Nyathi Y, Muench JH, Roessler MM, et al. (2018) Structural
717 complexity of the co-chaperone SGTA: a conserved C-terminal region is implicated in
718 dimerization and substrate quality control. *BMC Biol* **16**: 76.
719 48. Xu D, Zhang Y (2012) Ab initio protein structure assembly using continuous structure
720 fragments and optimized knowledge-based force field. *Proteins* **80**: 1715–1735.

721 49. Chio US, Chung S, Weiss S, Shan S-O (2019) A Chaperone Lid Ensures Efficient and
722 Privileged Client Transfer during Tail-Anchored Protein Targeting. *Cell Rep* **26**: 37–
723 44.e37.
724 50. Shao S, Rodrigo-Brenni MC, Kivlen MH, Hegde RS (2017) Mechanistic basis for a
725 molecular triage reaction. *Science* **355**: 298–302.
726 51. Chun E, Thompson AA, Liu W, Roth CB, Griffith MT, Katritch V, Kunken J, Xu F,
727 Cherezov V, Hanson MA, et al. (2012) Fusion Partner Toolchest for the Stabilization and
728 Crystallization of G Protein-Coupled Receptors. *Structure* **20**: 967–976.
729 52. Letunic I, Doerks T, Bork P (2015) SMART: recent updates, new developments and status
730 in 2015. *Nucleic Acids Res* **43**: D257–D260.
731 53. Schmid AB, Lagleder S, Gräwert MA, Röhl A, Hagn F, Wandinger SK, Cox MB,
732 Demmer O, Richter K, Groll M, et al. (2012) The architecture of functional modules in the
733 Hsp90 co-chaperone Sti1/Hop. *EMBO J* **31**: 1506–1517.
734 54. Li Z, Hartl FU, Bracher A (2013) Structure and function of Hip, an attenuator of the
735 Hsp70 chaperone cycle. *Nat Struct Mol Biol* **20**: 929–935.
736 55. Nelson GM, Huffman H, Smith DF (2003) Comparison of the carboxy-terminal DP-repeat
737 region in teh co-chaperones Hop and Hip. *EMBO J* **8**: 125–133.
738 56. Rao M, Okreglak V, Chio US, Cho H, Walter P, Shan S-O (2016) Multiple selection
739 filters ensure accurate tail-anchored membrane protein targeting. *eLife* **5**: e21301–e21324.
740 57. Pace CN, Scholtz JM (1998) A Helix Propensity Scale Based on Experimental Studies of
741 Peptides and Proteins. *Biophys J* **75**: 422–427.
742 58. Röhl A, Wengler D, Madl T, Lagleder S, Tippel F, Herrmann M, Hendrix J, Richter K,
743 Hack G, Schmid AB, et al. (2015) Hsp90 regulates the dynamics of its cochaperone Sti1
744 and the transfer of Hsp70 between modules. *Nat Commun* **6**: 6655.
745 59. Coto ALS, Seraphim TV, Batista FAH, Dores-Silva PR, Barranco ABF, Teixeira FR,
746 Gava LM, Borges JC (2018) Structural and functional studies of the Leishmania
747 braziliensis SGT co-chaperone indicate that it shares structural features with HIP and can
748 interact with both Hsp90 and Hsp70 with similar affinities. *Int J Biol Macromol* **118**: 693–
749 706.
750 60. Voorhees RM, Hegde RS (2015) Structures of the scanning and engaged states of the
751 mammalian SRP-ribosome complex. *eLife* **4**: 1485–21.
752 61. Keenan RJ, Freymann DM, Walter P, Stroud RM (1998) Crystal Structure of the Signal
753 Sequence Binding Subunit of the Signal Recognition Particle. *Cell* **94**: 181–191.
754 62. Clemons WM, Gowda K, Black SD, Zwieb C, Ramakrishnan V (1999) Crystal structure
755 of the conserved subdomain of human protein SRP54M at 2.1 A resolution: evidence for
756 the mechanism of signal peptide binding. *J Mol Biol* **292**: 697–705.
757 63. Tidow H, Nissen P (2013) Structural diversity of calmodulin binding to its target sites.
758 *FEBS J* **280**: 5551–5565.
759 64. Guna A, Volkmar N, Christianson JC, Hegde RS (2018) The ER membrane protein
760 complex is a transmembrane domain insertase. *Science* **359**: 470–473.
761 65. Suloway CJ, Rome ME, Clemons WM (2011) Tail-anchor targeting by a Get3 tetramer:
762 the structure of an archaeal homologue. *EMBO J* **31**: 707–719.
763 66. Chu R, Takei J, Knowlton JR, Andrykovitch M, Pei W, Kajava AV, Steinbach PJ, Ji X,
764 Bai Y (2002) Redesign of a Four-helix Bundle Protein by Phage Display Coupled with
765 Proteolysis and Structural Characterization by NMR and X-ray Crystallography. *J Mol*
766 *Biol* **323**: 253–262.

767 67. Studier FW (2005) Protein production by auto-induction in high-density shaking cultures.
768 *Protein Express Purif* **41**: 207–234.
769 68. Micsonai A, Wien F, Kernya L, Lee Y-H, Goto Y, Réfrégiers M, Kardos J (2015)
770 Accurate secondary structure prediction and fold recognition for circular dichroism
771 spectroscopy. *Proc Natl Acad Sci USA* **112**: E3095–E3103.
772 69. Schneider CA, Rasband WS, Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image
773 analysis. *Nat Methods* **9**: 671–675.
774 70. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M,
775 Qureshi M, Sangrador-Vegas A, et al. (2016) The Pfam protein families database: towards
776 a more sustainable future. *Nucleic Acids Res* **44**: D279–D285.
777 71. Li W, Godzik A (2006) Cd-hit: a fast program for clustering and comparing large sets of
778 protein or nucleotide sequences. *Bioinformatics* **22**: 1658–1659.
779 72. Katoh K, Standley DM (2013) MAFFT Multiple Sequence Alignment Software Version
780 7: Improvements in Performance and Usability. *Mol Biol Evol* **30**: 772–780.
781 73. Pei J, Kim B-H, Grishin NV (2008) PROMALS3D: a tool for multiple protein sequence
782 and structure alignments. *Nucleic Acids Res* **36**: 2295–2300.
783 74. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ (2009) Jalview Version
784 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**:
785 1189–1191.
786 75. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y (2015) The I-TASSER Suite: protein
787 structure and function prediction. *Nat Methods* **12**: 7–8.
788 76. Wallner B (2006) Identification of correct regions in protein models using structural,
789 alignment, and consensus information. *Prot Sci* **15**: 900–913.
790 77. Bradley P, Misura KM, Baker D (2005) Towards high-resolution de novo structure
791 prediction for small proteins. *Science* **309**: 1868–1871.
792 78. Yang J, Anishchenko I, Park H, Peng Z, Ovchinnikov S, Baker D (2019) Improved
793 protein structure prediction using predicted inter-residue orientations. *bioRxiv* **13**:
794 e1005324–e1005324.
795 79. Reißer S, Strandberg E, Steinbrecher T, Ulrich AS (2014) 3D Hydrophobic Moment
796 Vectors as a Tool to Characterize the Surface Polarity of Amphiphilic Peptides. *Biophys J*
797 **106**: 2385–2394.
798 80. Pierce BG, Wiehe K, Hwang H, Kim BH, Vreven T, Weng Z (2014) ZDOCK server:
799 interactive docking prediction of protein-protein complexes and symmetric multimers.
800 *Bioinformatics* **30**: 1771–1773.
801 81. McGibbon RT, Beauchamp KA, Harrigan MP, Klein C, Swails JM, Hernández CX,
802 Schwantes CR, Wang L-P, Lane TJ, Pande VS (2015) MDTraj: A Modern Open Library
803 for the Analysis of Molecular Dynamics Trajectories. *Biophys J* **109**: 1528–1532.
804 82. Webb B, Sali A (2002) *Comparative Protein Structure Modeling Using MODELLER*.
805 John Wiley & Sons, Inc., Hoboken, NJ, USA.
806 83. Drozdetskiy A, Cole C, Procter J, Barton GJ (2015) JPred4: a protein secondary structure
807 prediction server. *Nucleic Acids Res* **43**: W389–W394.
808 84. Schägger H (2006) Tricine–SDS-PAGE. *Nat Protoc* **1**: 16–22.
809 85. Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a
810 protein. *J Mol Biol* **157**: 105–132.
811 86. Fernandez-Patron C, Hardy E, Sosa A, Seoane J, Castellanos L (1995) Double staining of
812 coomassie blue-stained polyacrylamide gels by imidazole-sodium dodecyl sulfate-zinc

813 reverse staining: sensitive detection of coomassie blue-undetected proteins. *Anal Biochem*
814 **224**: 263–269.
815 87. Bateman A, Martin MJ, O'Donovan C, Magrane M, Alpi E, Antunes R, Bely B, Bingley
816 M, Bonilla C, Britto R, et al. (2017) UniProt: the universal protein knowledgebase.
817 *Nucleic Acids Res* **45**: D158–D169.
818

819 ## Figures

820 **Fig. 1. Structural characteristics of free Sgt2 C-domain.** (*A*) *Top*, Schematic of the domain
821 organization of Sgt2. *Below,* representative sequences from a large-scale multiple sequence
822 alignment of the C domain: fungal Sgt2 from *S. cerevisiae*, *S. pombe*, and *C. thermophilum* and
823 metazoan Sgt2 from *C. savignyi*, *X. laevis*, and *H. sapiens*. Protease susceptible sites on ySgt2-C
824 identified by mass spectrometry are indicated by red arrowheads. Predicted helices of ySgt2 (blue)
825 and hSgt2 (orange) by Jpred [83] and/or structure prediction are shown. Blue/orange color scheme
826 for ySgt2/hSgt2 is used throughout the text. Residues noted in the text are highlighted by an
827 asterisk. (*B*) Overlay of size-exclusion chromatography traces of ySgt2-C (blue line), hSgt2-C
828 (orange line), ySgt2-TPR (blue dash), and hSgt2-TPR (orange dash). Traces are measured at 214
829 nm, baseline-corrected and normalized to the same peak height. (*C*) Far UV CD spectrum of 10
830 μM of purified ySgt2-C (blue) and hSgt2-C (orange) at RT with secondary structure decomposition
831 from BestSel [68]. (*D*) ^{1}H-^{15}N HSQC spectrum of ySgt2-C at 25°C. The displayed chemical shift
832 window encompasses all N-H resonances from both backbone and side chains. The range of
833 backbone amide protons, excluding possible side-chain NH_2 of Asn/Gln, is indicated by pairs of
834 red dashed lines. (*E*) As in *D* for hSgt2-C at 25°C.

835

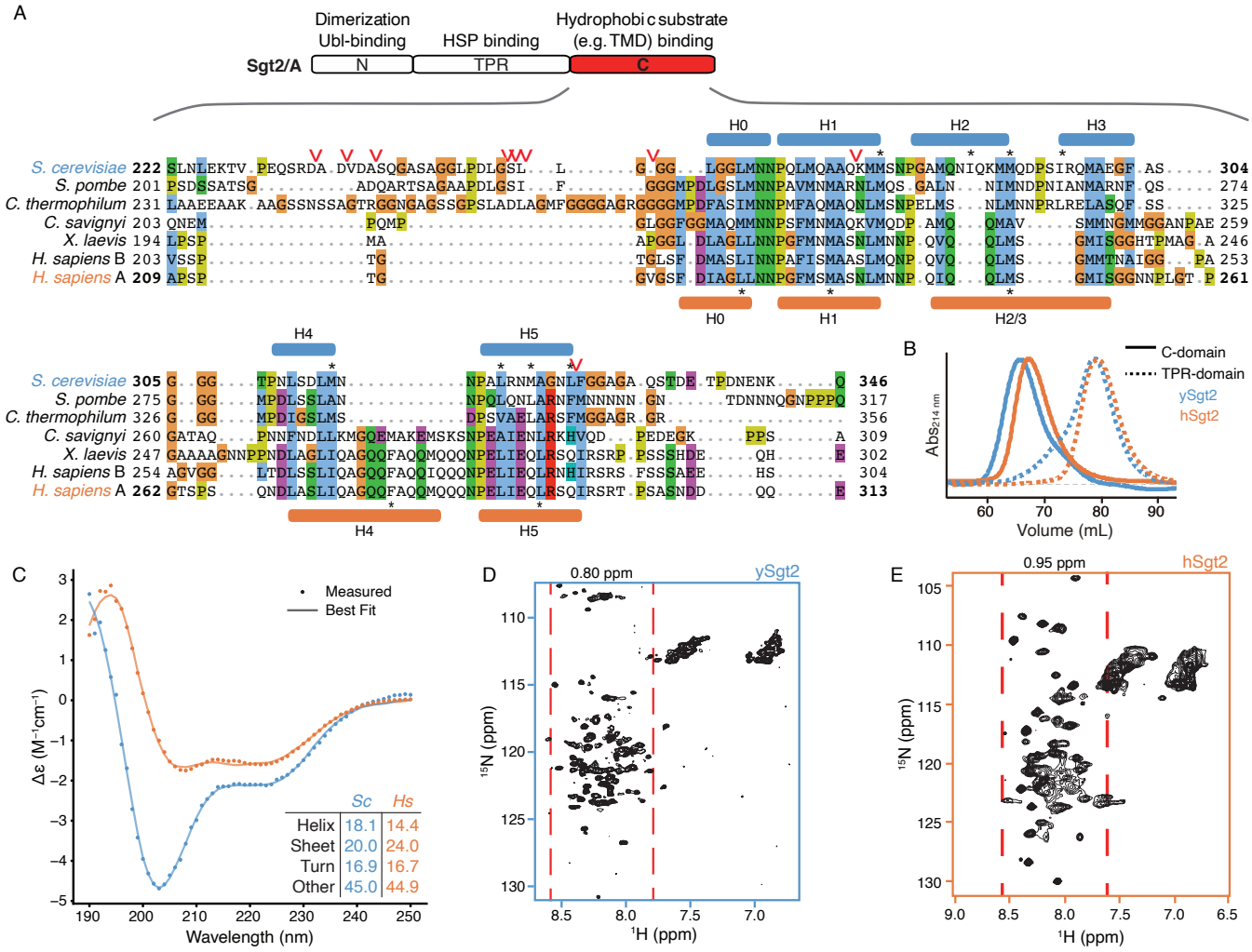

A
Dimerization
Ubl-binding
HSP binding
Hydrophobic substrate
(e.g. TMD) binding
Sgt2/A
N
TPR
C
H0
H1
H2
H3
S. cerevisiae
S. pombe
C. thermophilum
C. savignyi
X. laevis
H. sapiens B
H. sapiens A
H2/3
H4
H5
B
C-domain
TPR-domain
ySgt2
hSgt2
Abs214 nm
Volume (mL)
C
Measured
Best Fit
Δε (M⁻¹cm⁻¹)
Wavelength (nm)
Sc
Hs
Helix 18.1 14.4
Sheet 20.0 24.0
Turn 16.9 16.7
Other 45.0 44.9
D
0.80 ppm
ySgt2
15N (ppm)
1H (ppm)
E
0.95 ppm
hSgt2
15N (ppm)
1H (ppm)

Figure 1

837 **Fig. 2. The minimal binding region of Sgt2 for TA client binding.** (*A*) Diagram of the protein
838 truncations tested for TA client binding that include the TPR-C domain, C-domain (C), C_{cons}, and
839 C_{cons} ΔH0 (ΔH0) from ySgt2 and hSgt2. The residues corresponding to each domain are indicated,
840 and grey blocks highlight the C_{cons} region. *(B)* Schematic of capture experiments of MBP-tagged
841 Sbh1 (MBP-Sbh1) by Sgt2 variants. After co-expression, cell pellets are lysed and NTA-Ni^{2+} is
842 used to capture his-tagged Sgt2-TPR-C. (*C*) Tris-Tricine-SDS-PAGE gel [84] of co-expressed and
843 purified MBP-Sbh1 and his-tagged Sgt2 truncations visualized with Coomassie Blue staining. (*D*)
844 Helical wheel diagrams of predicted helices (see Fig. 1*A*) in the C_{cons} domain of ySgt2 and hSgt2.
845 Residues are colored by the Kyte and Doolittle hydrophobicity scale [85]. (*E*) All of the
846 hydrophobic residues (L, I, F, and M) in a predicted helix (H0, H1, etc.) are replaced with alanines
847 and tested for the ability to capture MBP-Sbh1. Protein levels were quantified by Coomassie
848 staining. Relative binding efficiency of MBP-Sbh1 by ySgt2 C-domain (ySgt2-C) variants was
849 calculated relative to total amount of ySgt2-C captured (MBP-Sbh1/Sgt2-C) then normalized to
850 the wild-type ySgt2-C. Experiments were performed 3-4 times and the standard deviations are
851 presented. Total expression levels of the MBP-Sbh1 were similar across experiments as visualized
852 by immunoblotting (IB) of the cell lysate. (*F)* As in *E* but for hSgt2.

853

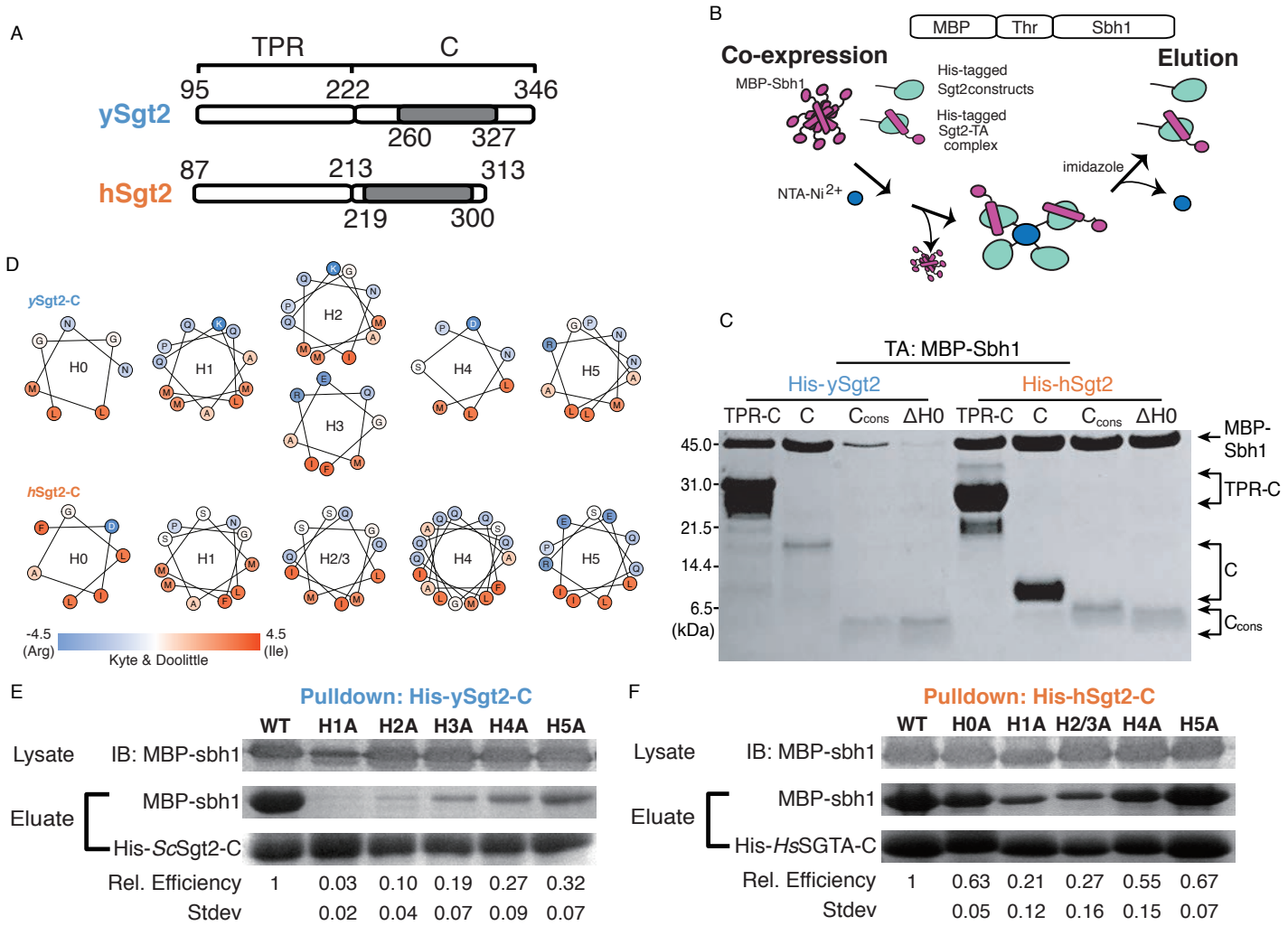

A
TPR
C
95
222
346
ySgt2
260
327
87
213
313
hSgt2
219
300
B
MBP
Thr
Sbh1
Co-expression
Elution
MBP-Sbh1
His-tagged Sgt2constructs
His-tagged Sgt2-TA complex
imidazole
NTA-Ni2+
C
TA: MBP-Sbh1
His-ySgt2
His-hSgt2
TPR-C
C
Ccons
ΔH0
TPR-C
C
Ccons
ΔH0
45.0
31.0
21.5
14.4
6.5
(kDa)
MBP-Sbh1
TPR-C
C
Ccons
D
ySgt2-C
H0
H1
H2
H3
H4
H5
hSgt2-C
H0
H1
H2/3
H4
H5
-4.5 (Arg)
4.5 (Ile)
Kyte & Doolittle
E
Pulldown: His-ySgt2-C
WT
H1A
H2A
H3A
H4A
H5A
Lysate
IB: MBP-sbh1
Eluate
MBP-sbh1
His-ScSgt2-C
Rel. Efficiency
1
0.03
0.10
0.19
0.27
0.32
Stdev
0.02
0.04
0.07
0.09
0.07
F
Pulldown: His-hSgt2-C
WT
H0A
H1A
H2/3A
H4A
H5A
Lysate
IB: MBP-sbh1
Eluate
MBP-sbh1
His-HsSGTA-C
Rel. Efficiency
1
0.63
0.21
0.27
0.55
0.67
Stdev
0.05
0.12
0.16
0.15
0.07

Figure 2

855 **Fig 3. A structural model for Sgt2-C_{cons}.** *(A)* The top 10 models of the ySgt2-C_{cons} generated by
856 the template-free algorithm Quark [48] are overlaid with the highest scoring model in solid.
857 Models are color-ramped from N- (blue) to C-terminus (red). *(B)* A model of ySgt2-C_{cons} (surface
858 colored by Kyte-Doolittle hydrophobicity) bound to a TMD (purple helix) generated by rigid-body
859 docking through Zdock [80]. The darker purple corresponds to an 11 residue stretch. *(C)* The entire
860 ySgt2-C from the highest scoring model from Quark (C_{cons} in rainbow with the rest in grey)
861 highlighting H0 and the rest of the flexible termini that vary considerably across models.

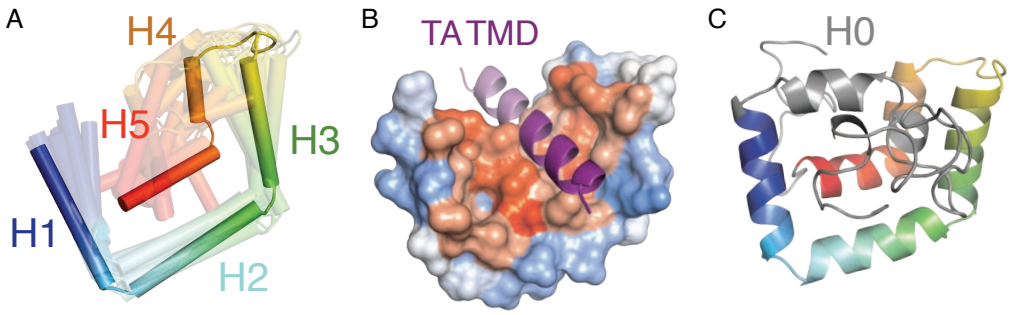
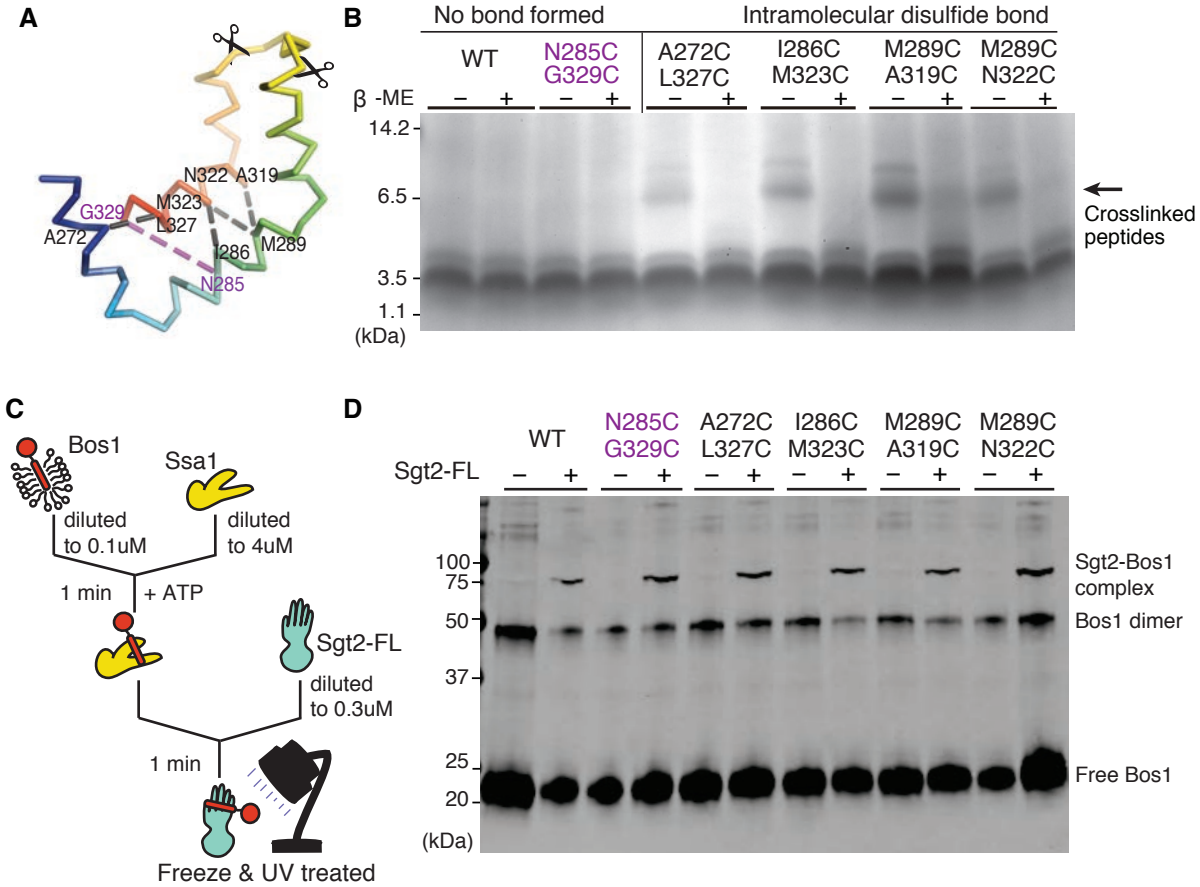


Figure 3

863 **Fig 4. Validating the structural model with disulfide bond formation.** Variants of His-ySgt2-
864 TPR-C (WT or cysteine double mutants) were co-expressed with the artificial TA client, cMyc-
865 BRIL-11[L8]. After lysis, ySgt2-TPR-C proteins were purified, oxidized, then digested by Glu-C
866 protease and analyzed by gel either in non-reducing or reducing buffer. (*A*) Cα ribbon of ySgt2-
867 C_{cons} color-ramped with various pairs of cysteines highlighted. Scissors indicate protease cleavage
868 sites resulting in fragments less than 3 kDa in size. (*B*) Tris-Glycine-SDS-PAGE gel visualized by
869 imidazole-SDS-zinc stain [86]. For the WT (cys-free) no significant difference was found between
870 samples in non-reducing vs. reducing conditions. All close residue pairs (A272/L327, I286/M323,
871 M289/A319, and M289/N322) show peptide fragments (higher MW) sensitive to the reducing
872 agent and indicate disulfide bond formation (indicated by arrow). A cysteine pair (N285/G329)
873 predicted to be far apart by the model does not result in the higher MW species. (*C*) A schematic
874 of the transfer of $Bos1^{BPA}$ from Ssa1 to full-length ySgt2 to demonstrate the double cysteine
875 mutants are still functional. (*D*) A western blot visualizing cross-linked ySgt2-$Bos1^{BPA}$ complexes.
876 All samples tested, WT, N285C/G329C, A272C/L327C, I286C/M323C, M289C/A319C, and
877 M289C/N322C, had a higher molecular weight appear after the addition of ySgt2 which
878 corresponds to the size of the cross-linked complex. As described in Cho et al., a band for cross-
879 linked Ssa1-$Bos1^{BPA}$ complexes was not observed. (*E*) C_β-C_β distances between the residues
880 mutated to cysteines based on various models predicted by the Quark, I-TASSER, PCONS, and
881 Robetta. Cysteine pairs that are 9Å or less apart are colored in orange and are expected to be close
882 enough to form disulfide bonds. Where all five pair distances are consistent with the experiment
883 (4 near and 1 far), the row is shaded in grey. (*F*) Tris-Glycine SDS-PAGE gel probing the flexibility
884 of ySgt2-C_{cons}. All new pairs (K288/A319, M290/A319, M289/P318, M289/L320) show peptide
885 fragments sensitive to the reducing agent (indicated by arrow). The range of distances of the eight
886 closest possible rotamer pairs is annotated below. The cysteine pair (N285/G329) shown to be far
887 apart by the model does have a faint higher molecular weight band.



E

		N285-G329	A272-L327	I286-M323	M289-A319	M289-N322
Quark	1	10.1	9.1	7.0	7.4	6.2
	2	6.5	9.1	11.6	15.0	12.4
	3	15.7	6.0	6.2	6.8	9.2
	4	6.7	10.1	14.1	19.4	16.0
	5	15.6	13.5	4.9	8.3	10.1
	6	11.3	8.5	14.6	22.9	19.4
	7	16.2	7.2	5.8	8.7	9.8
	8	7.6	11.6	10.3	14.4	10.7
	9	8.8	18.6	10.7	12.5	10.8
	10	6.9	14.2	11.1	14.2	10.7
I-TASSER	1	17.6	7.8	8.4	13.9	15.7
	2	21.3	9.6	11.5	17.9	17.9
	3	24.6	12.4	14.8	17.8	21.3
	4	20.2	11.1	12.1	19.9	21.0
	5	17.6	10.0	6.8	17.7	17.5

		N285-G329	A272-L327	I286-M323	M289-A319	M289-N322
Pcons	1	44.0	28.3	30.7	24.2	31.0
	2	23.7	11.1	17.1	18.9	21.1
	3	45.1	39.6	31.6	24.6	33.3
	4	16.8	15.7	8.2	12.8	12.9
	5	22.9	9.8	8.6	11.5	13.2
	6			12.8	10.4	10.6
	7	23.0	22.0	20.5	29.3	29.3
	8			13.8	10.7	10.7
RaptorX		43.3	26.1	25.6	20.3	26.5
Phyre2		17.2	12.2	7.4	16.9	13.0
Rosetta	1	8.6	7.1	9.8	13.3	11.0
	2	14.1	13.3	9.1	20.8	14.1
	3	18.1	5.6	6.3	13.2	13.1
	4	8.8	9.8	9.9	10.0	7.6
	5	11.0	7.6	8.6	12.7	10.9

C_β-C_β Distance (Å) 9 Å or less

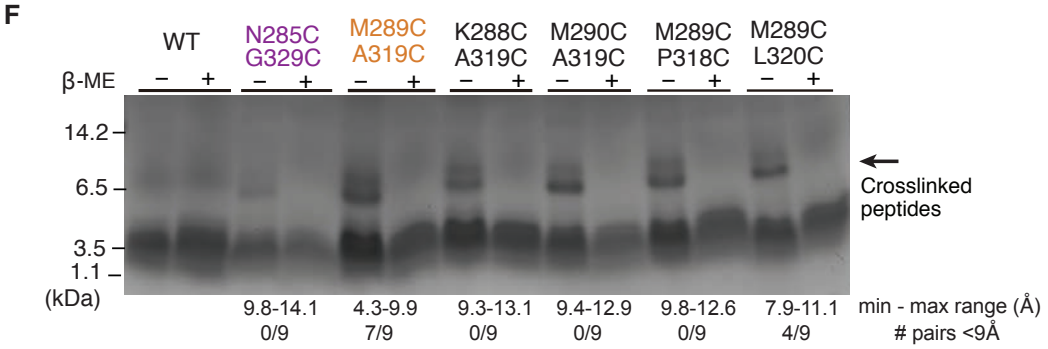


Figure 4

889 **Fig. 5. Comparison of Sti1 domains and the Sgt2-C_{cons} model.** (*A*) Multiple sequence alignment
890 of Sgt2-C with STI1 domains (DP1, DP2) from STI1/Hop homologs. Helices are shown based on
891 the Sgt2-C_{cons} model and the *Sc*Sti1-DP1/2 structures. Species for representative sequences are
892 from *S. cerevisiae (Scer)*, *S. pombe* (*Spom*), *C. thermophilum* (*Cthe*), *C. savignyi* (*Csav*), and *H.*
893 *sapiens* (*Hsap*). (*B*) Cα ribbon of *Sc*Sgt2-C_{cons} color-ramped with large hydrophobic sidechains
894 shown as grey sticks (sulfurs in yellow). (*C*) Similar to *B* for the solution NMR structure of Sti1-
895 $DP2_{526-582}$ (PDBID: 2LLW) [53]. (*D*) Superposition of the Sgt2-C_{cons} (blue) and Sti1-$DP2_{526-582}$
896 (red) drawn as cartoons.

897

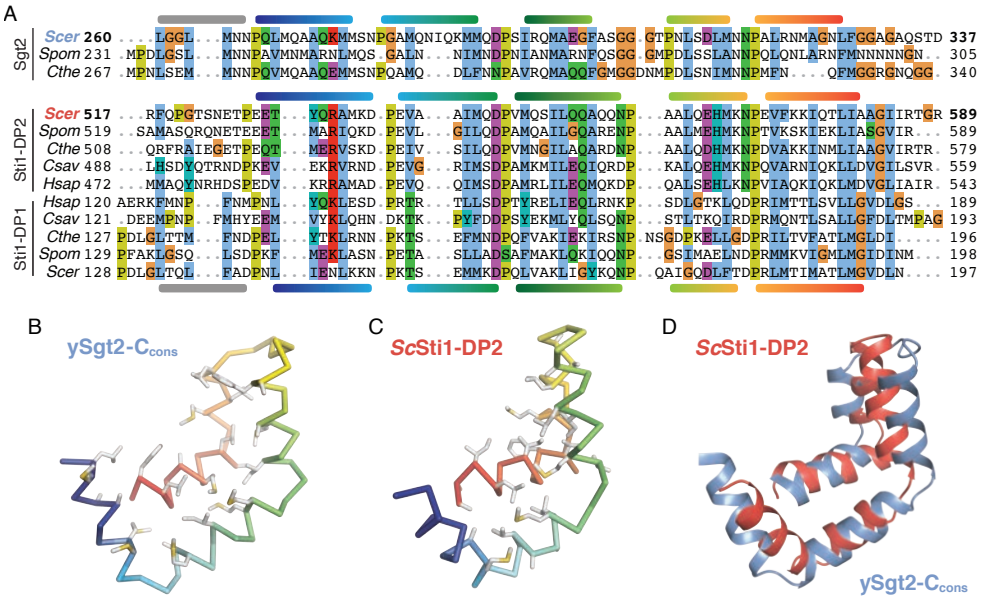

A
Sgt2
Scer 260 ...LGGL...MNNPQLMQAAQKMMSNPGAMQNIQKMMQDPSIRQMAEGFASGG.GTPNLSDLMNNPALRNMAGNLFGGAGAQSTD 337
Spom 231 .MPDLGSL...MNNPAVMNMARNLMQS.GALN...NIMNDPNIANMARNFQSGG.GMPDLSSLANNPQLQNLARNFMNNNNNGN.. 305
Cthe 267 .MPNLSEM...MNNPQVMQAAQEMMSNPQAMQ...DLFNNPAVRQMAQQFGMGGDNMPDLSNIMNNPMFN....QFMGGRGNQGG. 340
Sti1-DP2
Scer 517 ...RFQPGTSNETPEET...YQRAMKD.PEVA...AIMQDPVMQSILQQAQQNP...AALQEHMKNPEVFKKIQTLIAAGIIRTGR 589
Spom 519 ..SAMASQRQNETEEET...MARIQKD.PEVL...GILQDPAMQAILGQARENP...AALMEHMKNPTVKSKIEKLIASGVIR... 589
Cthe 508 ...QRFRAIEGETPEQT...MERVSKD.PEIV...SILQDPVMNGILAQARDNP...AALQDHMKNPDVAKKINMLIAAGVIRTR. 579
Csav 488 ...LHSDYQTRNDPKEV...EKRVRND.PEVG...RIMSDPAMKMILEQIQRDP...KALQEHMKNPQVARNIQKLLDVGILSVR. 559
Hsap 472 ...MMAQYNRHDSPEDV...KRRAMAD.PEVQ...QIMSDPAMRLILEQMQKDP...QALSEHLKNPVIAQKIQKLMDVGLIAIR. 543
Sti1-DP1
Hsap 120 AERKFMNP...FNMPNL...YQKLESD.PRTR...TLLSDPTYRELIEQLRNKP...SDLGTKLQDPRIMTTLSVLLGVDLGS... 189
Csav 121 .DEEMPNP..FMHYEEM...VYKLQHN.DKTK...PYFDDPSYEKMLYQLSQNP...STLTKQIRDPRMQNTLSALLGFDLTMPAG 193
Cthe 127 PDLGLTTM...FNDPEL...YTKLRNN.PKTS...EFMNDPQFVAKIEKIRSNP.NSGDPKELLGDPRILTVFATLMGLDI..... 196
Spom 129 PFAKLGSQ...LSDPKF...MEKLASN.PETA...SLLADSAFMAKLQKIQQNP...GSIMAELNDPRMMKVIGMLMGIDINM... 198
Scer 128 PDLGLTQL...FADPNL...IENLKKN.PKTS...EMMKDPQLVAKLIGYKQNP..QAIGQDLFTDPRLMTIMATLMGVDLN.... 197
B
ySgt2-Ccons
C
ScSti1-DP2
D
ScSti1-DP2
ySgt2-Ccons

Figure 5

899 **Fig. 6. Effects on TA client binding of charge mutations to the putative hydrophobic groove**
900 **of Sgt2-C$_{cons}$.** For these experiments, individual point mutations are introduced into Sgt2-C and
901 tested for their ability to capture Sbh1 quantified as in Figure 2*D*. (*A*) For ySgt2-C, a schematic
902 and cartoon model are provided highlighting the helices and sites of individual point mutants both
903 color-ramped for direct comparison. For the cartoon, the docked TMD is shown in purple. Binding
904 of MBP-Sbh1 to his-tagged ySgt2-C and mutants were examined as in Figure 2*D*. Lanes for
905 mutated residues are labeled in the same color as the schematic *(B)* Same analysis as in *A* for
906 hSgt2-C. In addition, double point mutants are included.

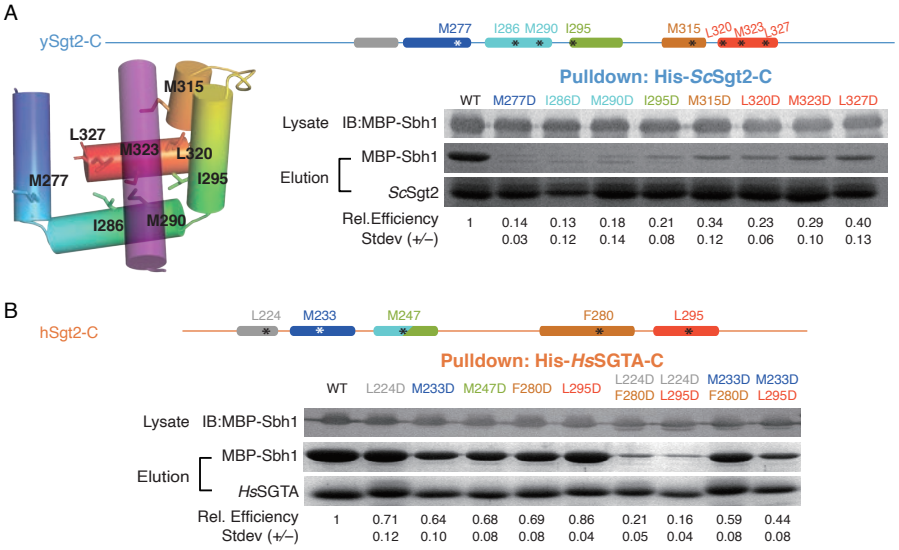

A
ySgt2-C
M277
I286
M290
I295
M315
L320
M323
L327
M315
L327
M323
L320
M277
I295
I286
M290
Pulldown: His-ScSgt2-C
WT M277D I286D M290D I295D M315D L320D M323D L327D
Lysate IB:MBP-Sbh1
Elution
MBP-Sbh1
ScSgt2
Rel.Efficiency 1 0.14 0.13 0.18 0.21 0.34 0.23 0.29 0.40
Stdev (+/–) 0.03 0.12 0.14 0.08 0.12 0.06 0.10 0.13
B
hSgt2-C
L224
M233
M247
F280
L295
Pulldown: His-HsSGTA-C
WT L224D M233D M247D F280D L295D L224D F280D L224D L295D M233D F280D M233D L295D
Lysate IB:MBP-Sbh1
Elution
MBP-Sbh1
HsSGTA
Rel. Efficiency 1 0.71 0.64 0.68 0.69 0.86 0.21 0.16 0.59 0.44
Stdev (+/–) 0.12 0.10 0.08 0.08 0.04 0.05 0.04 0.08 0.08

Figure 6

908 **Fig. 7. Minimal requirements for client recognition by Sgt2.** (*A*) Schematic of model TA clients.
909 Quantification of complex formation is calculated and normalized to that of complexes with Bos1-
910 TMD, here defined as relative binding efficiency. (*B*) Complex formation of ySgt2 (blue) and hSgt2
911 (orange) with the TA client Bos1TMD and several artificial TA clients noted x[Ly], where x denotes
912 the length of the TMD and y denotes the number of leucines in the TMD. The helical wheel
913 diagrams of the TMD of TA clients here and for subsequent panels with leucines colored in dark
914 orange, alanines colored in pale orange, and tryptophans colored in grey. (*C*) Complex formation
915 of ySgt2-TPR-C and hSgt2-TPR-C with artificial TA clients with TMDs of length 11 and
916 increasing numbers of leucine. (*D*) Comparison of complex formation of ySgt2-TPR-C and hSgt2-
917 TPR-C with artificial TA clients of the same lengths and hydrophobicities but differences in the
918 distribution of leucines, i.e. clustered (solid line) vs distributed (dotted line).

919

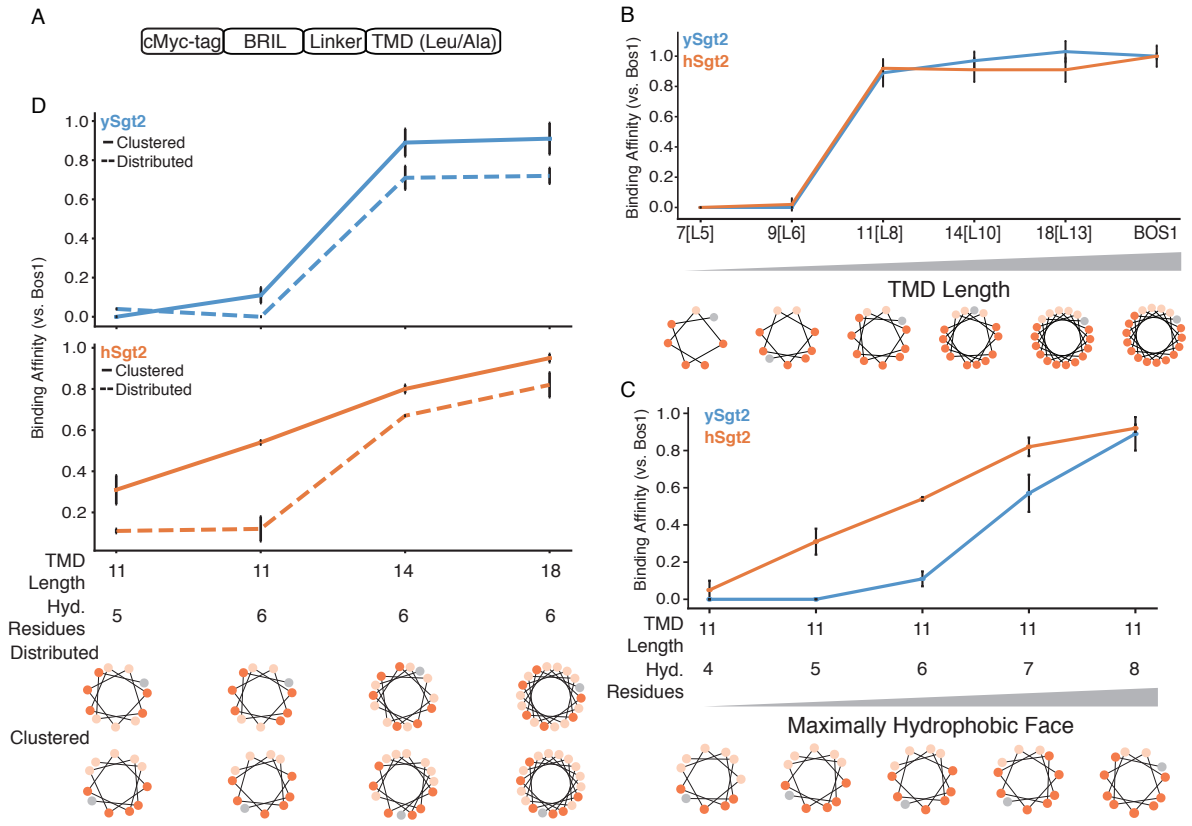

A
cMyc-tag
BRIL
Linker
TMD (Leu/Ala)
D
ySgt2
Clustered
Distributed
hSgt2
Clustered
Distributed
Binding Affinity (vs. Bos1)
TMD Length
Hyd. Residues
Distributed
Clustered
B
ySgt2
hSgt2
Binding Affinity (vs. Bos1)
7[L5]
9[L6]
11[L8]
14[L10]
18[L13]
BOS1
TMD Length
C
ySgt2
hSgt2
Binding Affinity (vs. Bos1)
TMD Length
Hyd. Residues
Maximally Hydrophobic Face

Figure 7

921 **Fig. 8. Various domain structures of STI1 and other helical-hand containing proteins.** (*A*)
922 The domain architectures of proteins with a STI1 domain were obtained initially from InterPro
923 [87] and then adjusted as discussed in the text. Each domain within a protein is colored relative to
924 the key. (*B*) Structural comparison of various hydrophobic-binding helical-hand protein complexes.
925 For each figure only relevant domains are included. Upper row, color-ramped cartoon
926 representation with bound helices in purple. Lower row, accessible surface of each protein colored
927 by hydrophobicity again with docked helical clients in purple. In order, the predicted complex of
928 ySgt2-C_{cons} and *Sc*Sbh2-TMD, DP1 domain from yeast Sti1 with N-terminus containing H0 in
929 grey (*Sc*Sti1-DP1)(PDBID: 2LLV), human calmodulin (*Hs*CALM2) bound to a hydrophobic
930 domain of calcineurin (*Hs*PPP3CA) (PDBID: 2JZI), and M domain of SRP54 from *Oryctolagus*
931 *cuniculus* (*Oc*SRP54-M) and the signal sequence of human transferrin receptor (*Hs*TR-TMD)
932 (PDBID: 3JAJ).
933

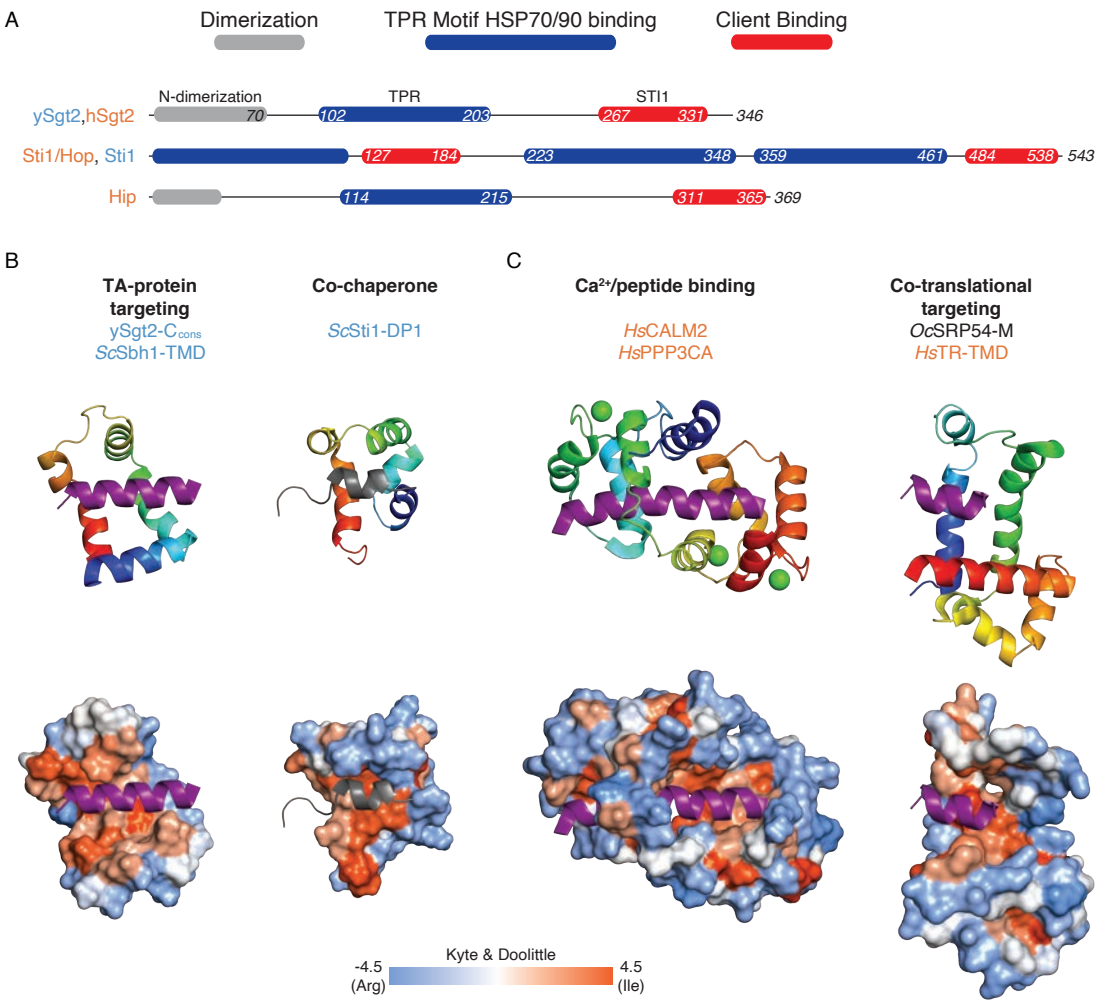

A
Dimerization
TPR Motif HSP70/90 binding
Client Binding
N-dimerization
TPR
STI1
ySgt2,hSgt2
70
102
203
267
331
346
Sti1/Hop, Sti1
127
184
223
348
359
461
484
538
543
Hip
114
215
311
365
369
B
TA-protein targeting
ySgt2-Ccons
ScSbh1-TMD
Co-chaperone
ScSti1-DP1
C
Ca2+/peptide binding
HsCALM2
HsPPP3CA
Co-translational targeting
OcSRP54-M
HsTR-TMD
Kyte & Doolittle
-4.5 (Arg)
4.5 (Ile)

Figure 8

935 **Fig. S1. Biophysical characterization of the Sgt2-C_{cons} domain.** (*A*) CD spectra as in Fig. 1*C*
936 for the conserved C-terminal domains of ySgt2 (blue) and hSgt2 (orange). NMR spectra as in Fig.
937 1*D* & *E* for ySgt2-C_{cons} (*B*, blue) and hSgt2-C_{cons} (*C*, orange).

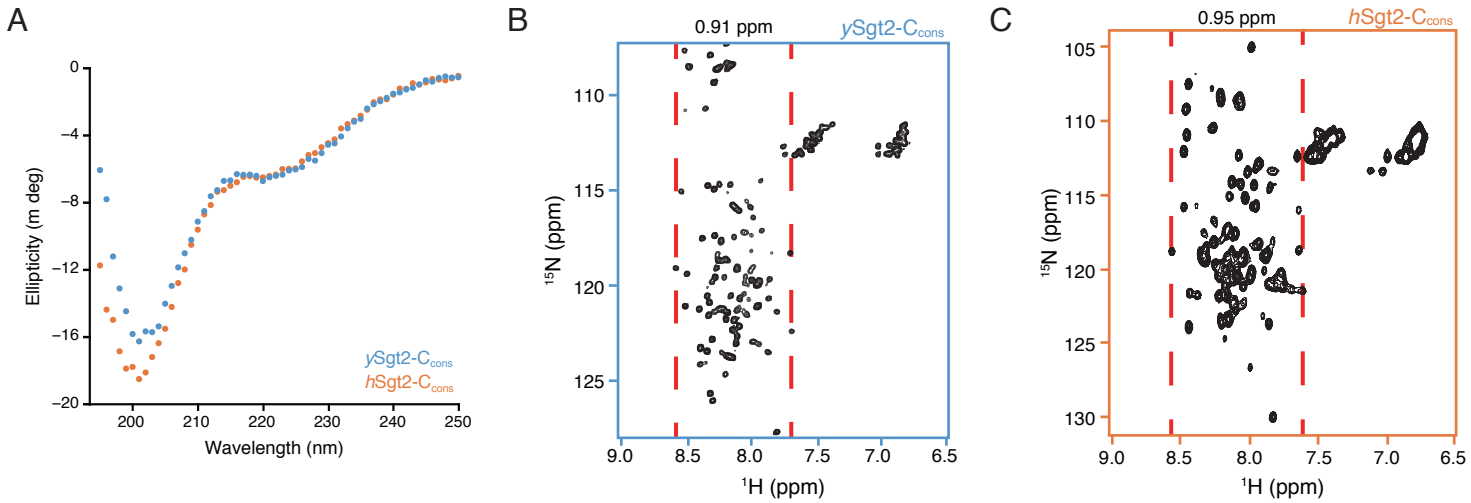

A
Ellipticity (m deg)
Wavelength (nm)
ySgt2-Ccons
hSgt2-Ccons
B
0.91 ppm
ySgt2-Ccons
15N (ppm)
1H (ppm)
C
0.95 ppm
hSgt2-Ccons
15N (ppm)
1H (ppm)

Figure S1

939 **Fig. S2. Structural models across prediction methods.** (*A*) Predictions from Quark, I-TASSER,
940 Pcons, Phyre2, RaptorX, and Robetta. Methods produce between 5 and 10 models. (*B*) Robetta
941 provides a residue-wise estimated error in Angstroms; this is shown below the corresponding
942 models with a grey bar indicating the C_{cons} region.

943

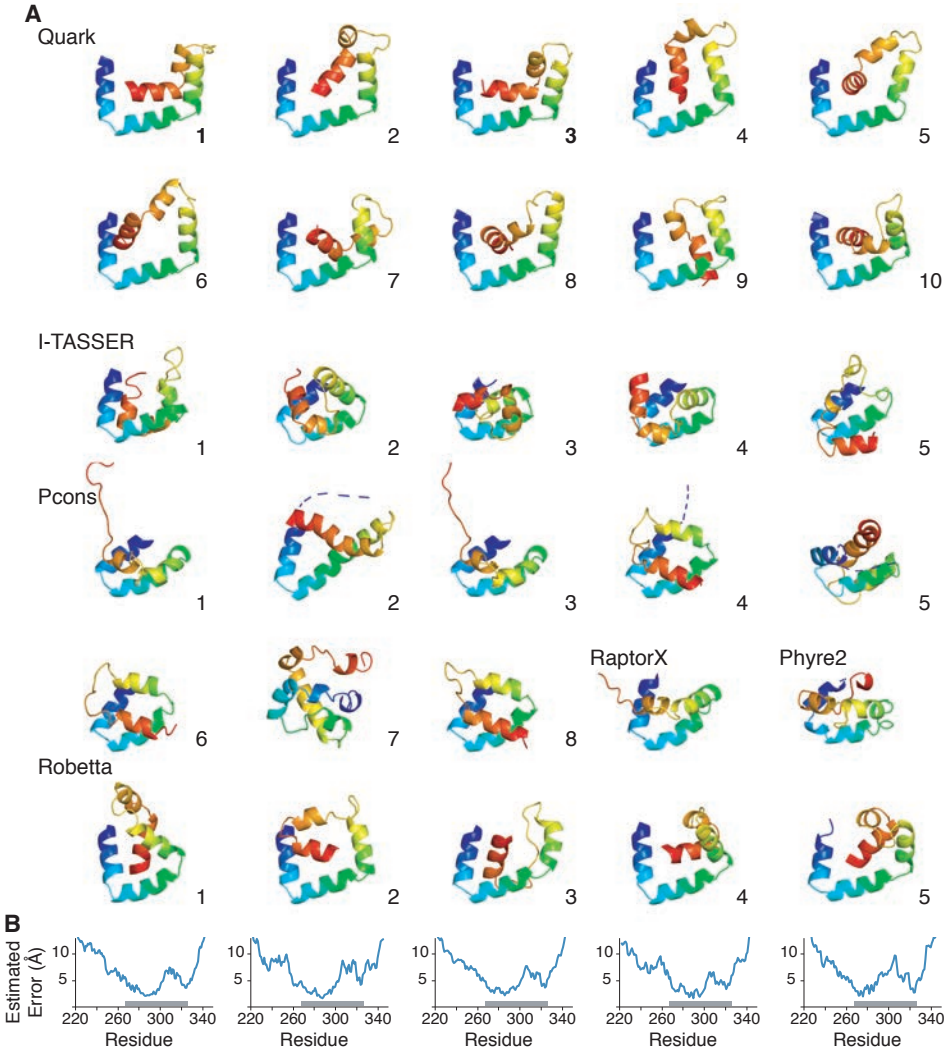

A
Quark
1
2
3
4
5
6
7
8
9
10
I-TASSER
1
2
3
4
5
Pcons
1
2
3
4
5
6
7
8
RaptorX
Phyre2
Robetta
1
2
3
4
5
B
Estimated Error (Å)
10
5
220
260
300
340
Residue

Figure S2

945 **Fig. S3. Cysteine mutants are capable of binding to TA clients.** (*A*) Schematic showing how
946 his-tagged ySgt2-TPR-C and double cysteine mutant constructs were coexpressed with the TA
947 client 11[L8], and complexes were purified by nickel affinity chromatography. (*B*) A SDS-PAGE
948 gel of the elution fractions demonstrates that 11[L8] was present in the elution suggesting double
949 cysteine mutations do not affect client binding. (*C*) An anti-cMyc western blot of the fractions
950 represented in the SDS-PAGE gel also demonstrates that 11[L8] was present in all elutions.
951

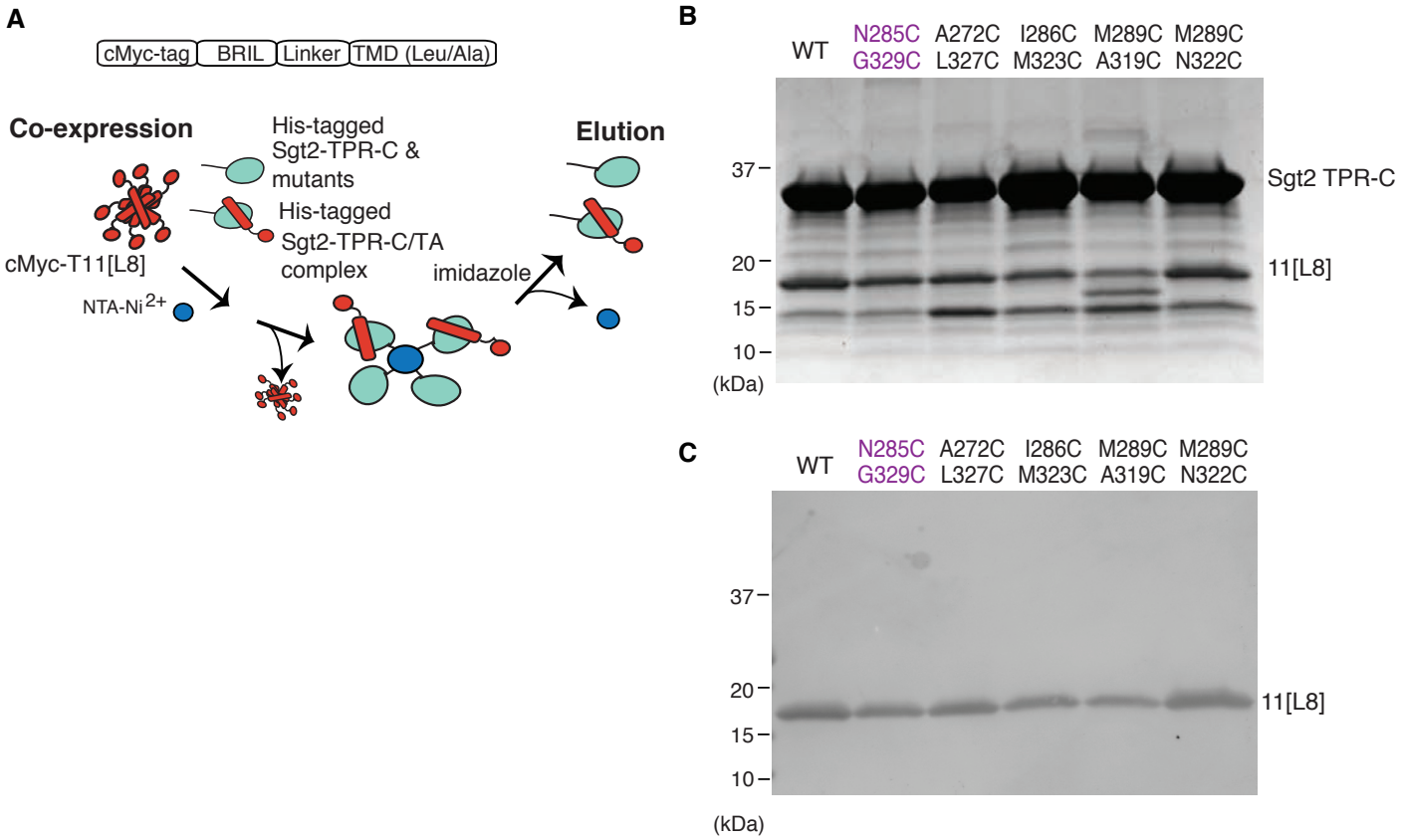
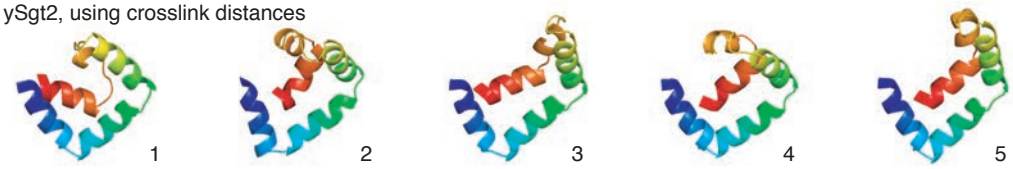

A
cMyc-tag
BRIL
Linker
TMD (Leu/Ala)
Co-expression
His-tagged Sgt2-TPR-C & mutants
His-tagged Sgt2-TPR-C/TA complex
cMyc-T11[L8]
NTA-Ni2+
imidazole
Elution
B
WT
N285C G329C
A272C L327C
I286C M323C
M289C A319C
M289C N322C
37
20
15
10
(kDa)
Sgt2 TPR-C
11[L8]
C
WT
N285C G329C
A272C L327C
I286C M323C
M289C A319C
M289C N322C
37
20
15
10
(kDa)
11[L8]

Figure S3

953 **Fig. S4. Distance restraints lead to improved ySgt2-C and suggestive hSgt2-C models.** (*A*)
954 Prediction of ySgt2-C using distances from in vitro crosslinking. (*B*) Cβ-Cβ distances between
955 residues probed by *in vitro* disulfide crosslinking for each ySgt2 model. Distances 9 Å or less are
956 colored orange. For models where all distances correspond (4 near and 1 far), the row is shaded
957 grey. (*C*) Models for hSgt2-C using restraints, adding a N-terminal loop, and via the new Robetta
958 TR method.

A ySgt2, using crosslink distances



B

	N285-G329	A272-L327	I286-M323	M289-A319	M289-N322
1	18.6	4.6	6.8	7.3	12.4
2	10.3	5.6	5.5	9.0	7.4
3	12.7	6.7	4.8	5.0	8.0
4	9.6	8.2	7.7	9.0	6.7
5	8.5	8.0	7.2	7.2	5.1

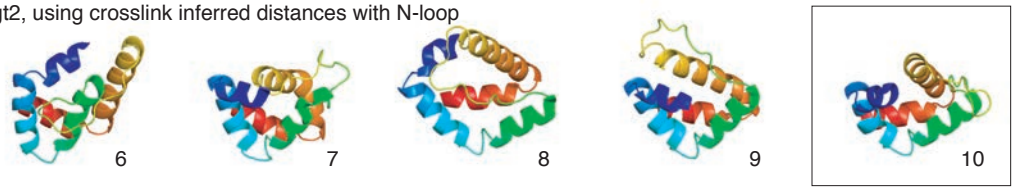
C_β-C_β Distance (Å)

9 Å or less

C

hSgt2, using crosslink inferred distances

hSgt2, using crosslink inferred distances with N-loop



hSgt2, Robetta TR

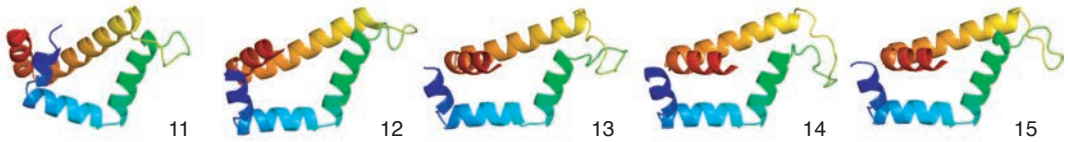


Figure S4