FULL-LENGTH PAPERS

The STI1-domain is a flexible alpha-helical fold with a hydrophobic groove

Michelle Y. Fry | **Shyam M. Saladi** | **William M. Clemons Jr** [ID]

Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California

Correspondence
William M. Clemons, Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA, USA
Email: clemons@caltech.eduz

Funding information
National Science Foundation Graduate Research, Grant/Award Number: 1144469; National Research Service Award Training, Grant/Award Number: T32 GM07616; National Institutes of Health, Grant/Award Numbers: GM097572, GM105385

Abstract

STI1-domains are present in a variety of co-chaperone proteins and are required for the transfer of hydrophobic clients in various cellular processes. The domains were first identified in the yeast Sti1 protein where they were referred to as DP1 and DP2. Based on hidden Markov model searches, this domain had previously been found in other proteins including the mammalian co-chaperone SGTA, the DNA damage response protein Rad23, and the chloroplast import protein Tic40. Here, we refine the domain definition and carry out structure-based sequence alignment of STI1-domains showing conservation of five amphipathic helices. Upon examinations of these identified domains, we identify a preceding helix 0 and unifying sequence properties, determine new molecular models, and recognize that STI1-domains nearly always occur in pairs. The similarity at the sequence, structure, and molecular levels likely supports a unified functional role.

KEYWORDS
co-chaperones, HIP, hop, protein targeting, SGTA, Sti1, ubiquilins, UBL-UBA

1 | INTRODUCTION

Hydrophobic stretches that are exposed during protein biosynthesis can aggregate which poses a risk to cellular homeostasis. To avoid this, cells have evolved proteins that bind to and protect hydrophobic segments. Several protein domains exist to assist these proteins and are found in protein families that occur broadly in eukaryotes, here we focus on the STI1-domain. Named for the yeast protein Sti1 (STress Inducible 1) where they were first identified, STI1-domains are referred to as heat-shock chaperonin-binding domains in databases.[1] Solved structures from Sti1 revealed an alpha-helical domain with five amphipathic helices that present a hydrophobic groove, an likely binding site for hydrophobic segments of a client[2,3] (Figure 1a,b). In addition to Sti1 homologs, a number of protein families were bioinformatically

identified to contain this domain including the co-chaperones HIP (HSP interacting protein) and SGTA (Small Glutamine-rich TPR-containing protein A), the DNA damage response protein Rad23 (RADiation sensitive 23), yeast UBL-UBA family member Dsk2 (Dominant Suppressor of Kar1 2), human KPC2 (Kip1 ubiquitylation-Promoting Complex 2), human ubiquilins (UBQLNs) 1–4, and the plant chloroplast import protein Tic40 (Translocon at the of the Inner envelope membrane of Chloroplasts 40).[4,5] The identified STI1-domain containing proteins can be broadly classified into two categories: either co-chaperones (homologs of Sti1, HIP, and SGTA and the unique plant Tic40) or adaptors of the ubiquitin proteasome system (AUPS) (homologs of the mammalian Rad23, UBQLNs, KPC2, and yeast Dsk2).

STI1-domain containing proteins have been identified only in eukaryotes. HOP and Rad23 are found throughout eukaryotes including in some protists. SGTA homologs are similarly prevalent, although not readily

---

Michelle Y. Fry and Shyam M. Saladi contributed equally to this work.
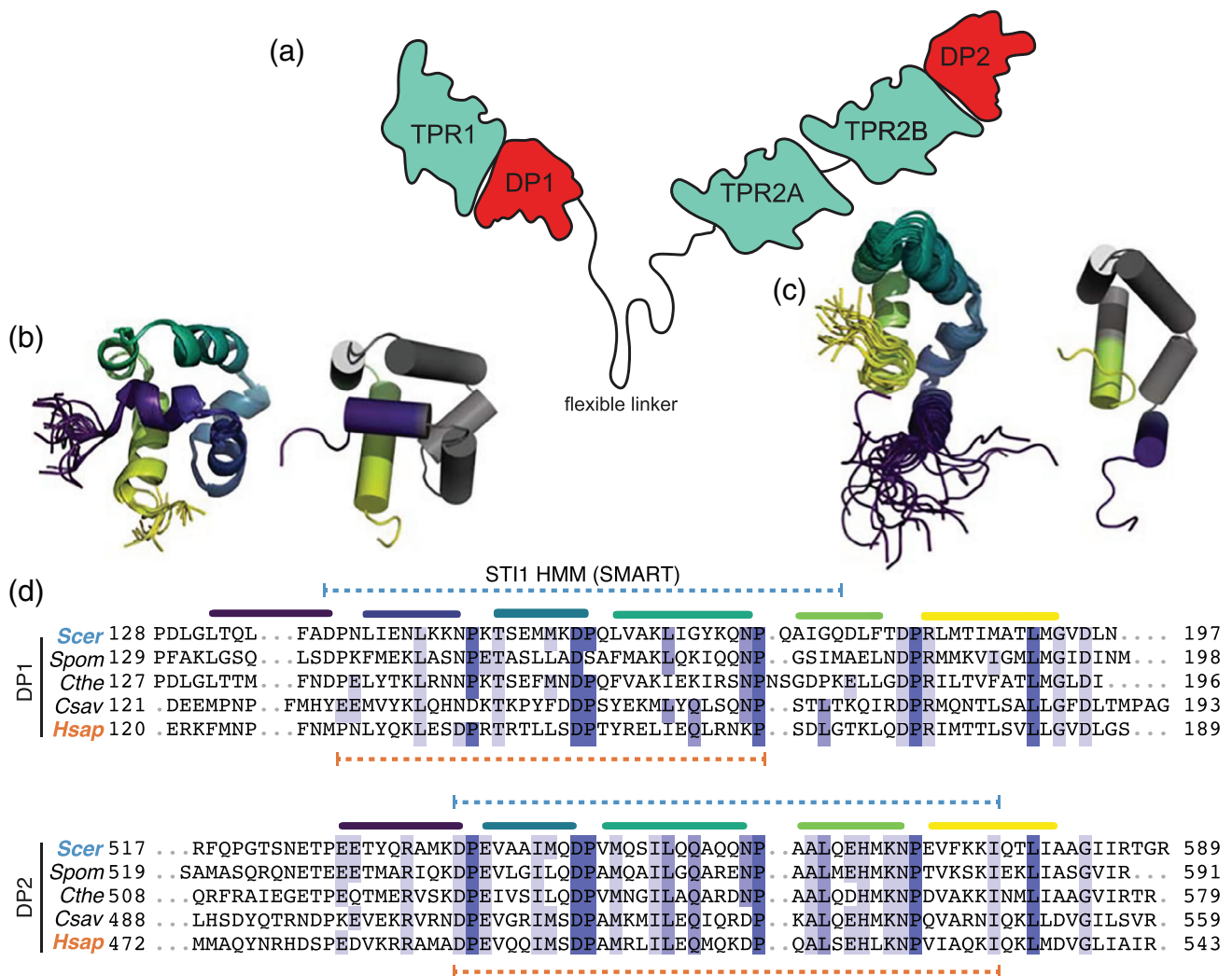
FIGURE 1 Sequence alignments and structural characterizations of the STI1-domains in HOP. (a) A cartoon representation of the HOP yeast homolog, Sti1, the two STI1-domains (DP2 and DP1) colored in red and three TPR domains in cyan. (b,c) Ribbon and cylinder cartoon depictions of the structures of (b) DP1 (PDBID:2LLV) and (c) DP2 (PDBID: 2LLW), colored using Viridis from N- to C-termini (purple to yellow). Grey cylinders indicate residues covered by the SMART STI1 HMM. (d) Sequence alignments of DP1 (top) and DP2 (bottom) from HOP homologs from *Saccharomyces cerevisiae* (Scer), *Schizosaccharomyces pombe* (Spom), *Chaetomium thermophilum* (Cthe), *Ciona savignyi* (Csav), and *Homo sapiens* (Hsap). Predicted helices are annotated above with the SMART STI1-HMM definition marked by dotted lines in blue (yeast) and orange (human). Names for this and subsequent figures are colored in blue for *S. cerevisiae* and orange for *H. sapiens*

identifiable in *Viridiplantae*.[4] UBQLNs, the closest mammalian relatives to yeast Dsk2, can be found across multicellular eukaryotes.[4,5] Tic40 is restricted to *Archaeplastida* (algae and land plants). It is likely that more distant homologs for each of these STI1-domain containing proteins exist in taxa that currently seem excluded.[6]

We consider the co-chaperones first. Broadly, co-chaperones are binding partners for Hsp90 or Hsp70 that enhance the function of these chaperones with a subset directly involved in binding to clients.[7] A first example is the mammalian Sti1-homolog HOP (Hsp70/Hsp90 organizing protein) that coordinates the essential transfer of clients between the chaperones Hsp70 and Hsp90. The abundant chaperone Hsp90 and its homologs are involved in multiple cellular pathways and many clients are first loaded from homologs of the chaperone Hsp70.[8] The domain organization of HOP homologs includes two STI1-domains, originally named DP1 and DP2 due to a repeated DP motif,[9,10] that are preceded by Hsp70/90-binding tetratricopeptide-repeat (TPR) domains[2,11–13] (Figure 1a). In yeast, in vivo deletion of the second STI1-domain in Sti1 (DP2) is detrimental, impairing native activity of the glucocorticoid receptor.[2] In vitro, removal of DP2 results in the loss of the transfer of the progesterone receptor to Hsp90.[12] These results implicate DP2 in client interaction. Besides simply bridging client transfer between the two HSPs, HOP has also been implicated in prion-protein binding.[14–16] Like

HOP, HIP also aids in the transfer of client from Hsp70 to Hsp90 through direct interaction with Hsp70.[17–19]

For the mammalian SGTA and its homologs, including yeast Sgt2, the STI1-domain directly binds to clients.[20] SGTA homologs play a number of roles, with the best characterized involving the targeting of tail-anchored (TA) proteins to the ER membrane as a member of the Guided Entry of TA protein (GET) pathway.[21–25] SGTA has been suggested to play a role in the degradation of mislocalized membrane proteins in conjunction with the protein Bag6.[26–31] Additionally, SGTA is involved with disease, including polyomavirus infection,[32] neurodegenerative disease,[33,34] hormone-regulated carcinogenesis,[35,36] and myogenesis,[37] although the underlying molecular mechanisms are still unclear.

The final member of the co-chaperone family is the chloroplast protein Tic40. In *Arabidopsis thaliana*, Tic40 is found in the inner membrane of the chloroplast and has been suggested to be a co-chaperone for the stroma chaperone complex for protein transport across the inner membrane.[38,39] Deleting Tic40 leads to a decrease in the import of precursors into the chloroplast.[40] Where studied, STI1-domains in each co-chaperone interact with clients,[2,3,12,20–22,30,41–43] thus due to the role of the STI1 motif in other co-chaperones, the STI1-domain in Tic40 may also interact with clients being transported across the outer chloroplast membrane and into the stroma.[38–40,44]

The second group of STI1-domain containing proteins, the AUPS, primarily delivers clients to the proteasome. They contain an N-terminal ubiquitin-like (UBL) domain and a C-terminal ubiquitin-associated domain (UBA). One of the earliest identified UBL-containing protein in yeast was Rad23; this protein shuttles some proteins to the proteasome and also protects some clients from degradation by preventing ubiquitin elongation.[45–49] Rad23 has also been implicated in nucleotide excision repair as a complex with Rad4 that recognizes DNA damage.[5,50] Likewise the fungal Dsk2 acts as an adaptor to target ubiquitin-labeled proteins to the proteasome for degradation.[51] The UBA domain of Dsk2 recognizes the poly-ubiquitin tail on proteins and the UBL domain interacts with the proteasome regulatory subunit, Rpn1. Another UBL-UBA containing adapter protein is KPC2,[52] a subunit of the KPC E3 ligase complex where it acts as an adapter for p27 ubiquitination in the G1 phase of the cell cycle.[53]

The closest mammalian homologs to Dsk2 are the four ubiquilins, UBQLN-1 to 4.[5] While UBQLN-1 is universally expressed and UBQLN-2 & -4 are expressed in most tissues, UBQLN-3 is expressed only in the testes.[54,55] The best characterized of these, UBQLN-1, functions similar to Dsk2 and Rad23 by delivering poly-ubiquitinated proteins to the 26S proteasome. Other demonstrated roles for UBQLN-1 are an association with aggregates for delivery to the lysosome for degradation and in the ER-associated degradation (ERAD) pathway.[56–59] In UBQLN-1, STI1-domains have been shown to bind to TMDs of mitochondrial membrane proteins and target them to the proteasome for degradation.[60] The direct involvement of UBQLNs in client degradation suggests a broader role than simply being shuttling factors.[60]

Here, we inspect these identified STI1-domains and the proteins they reside in to clarify the criteria for this domain. Upon examination, there are clear similarities in the structural features of most STI1-domains, while some currently defined STI1-domains are likely misannotated. Based on structure-based sequence alignments and similarity in predicted secondary structure, we develop a new definition that has allowed the identification of other STI1-domains and clarification of previously misannotated domains. We employ structural prediction methods to model uncharacterized STI1-domains revealing a consistent alpha-helical hand architecture. When considering proteins that contain STI1-domains, we find similar functional roles and domain architecture. In total, this work provides a comprehensive definition of the STI1-domain.

2 | RESULTS

2.1 | Amalgamating and examining predicted STI1-domains

A search through protein databases shows a variety of entries with a name or protein domain annotation that includes "Sti1". Close homologs of HOP have "Sti1" or "Sti1-like" in their entry name. These are typically bidirectional BLAST best-hits, that is, for a new sequence, the top scoring hit in a reference database identifies the original sequence when searched against the set of new sequences.[61] Other entries also have "Sti1" domain annotation. This annotation originates from a hit to the STI1 Hidden Markov Model (HMM) created by the Simple Modular Architecture Research Tool (SMART) database (since 2001)[1] and more recently from an HMM in the Pfam v32 database[62] (Figure 1d). The STI1-domain is named as such due to the prevalence of Sti1 homologs in the seed sequences of the SMART and Pfam HMMs. HMM-based methods reliably identify homologs of lower sequence similarity and are much more sensitive than sequence-to-sequence searches such as BLAST. HMM-based searching achieves higher sensitivity by using an alignment of query sequences (a "seed") to search a target database. Typically, the set of protein sequence regions that comprise seeds are curated by hand, but the

creators of SMART developed an automated method to compile seed protein regions for the categorization of protein domains distinct from those identified by human curators, as in the case of Pfam.

The SMART HMM for the STI1 family provides a useful initial annotation, yet when comparing the HMM to known structures and sequence alignments of STI1-domains, several issues come to light. The first being a partial hit for each DP domain in Sti1 where only four of the five helices are recovered by the SMART HMM (Figure 1d). The last helix of DP1 and the first helix of DP2 are both not covered by the HMM hits in the Sti1 sequence. Accordingly, we sought to understand to what degree the protein regions identified by this HMM actually reflect a single homologous family.

We performed structural alignments[63] (Figure 2a) across putative members of this family with experimentally solved structures (Sti1-DP1, Sti1-DP2, Tic40-STI1-II, and Rad23). Sti1-DP1 and Sti1-DP2 have clear homology indicated both by structural similarity (Figure 1b,c) and structure-guided sequence alignment using PROMALS3D[64] (Figure 1d). Tic40-STI1-II closely resembles Sti1-DP2 (Figure 2b). For Rad23, the orientation and register of the helices differs from the other domains. This can be visualized with respect to the substrate binding position where Rad23-STI1 forms
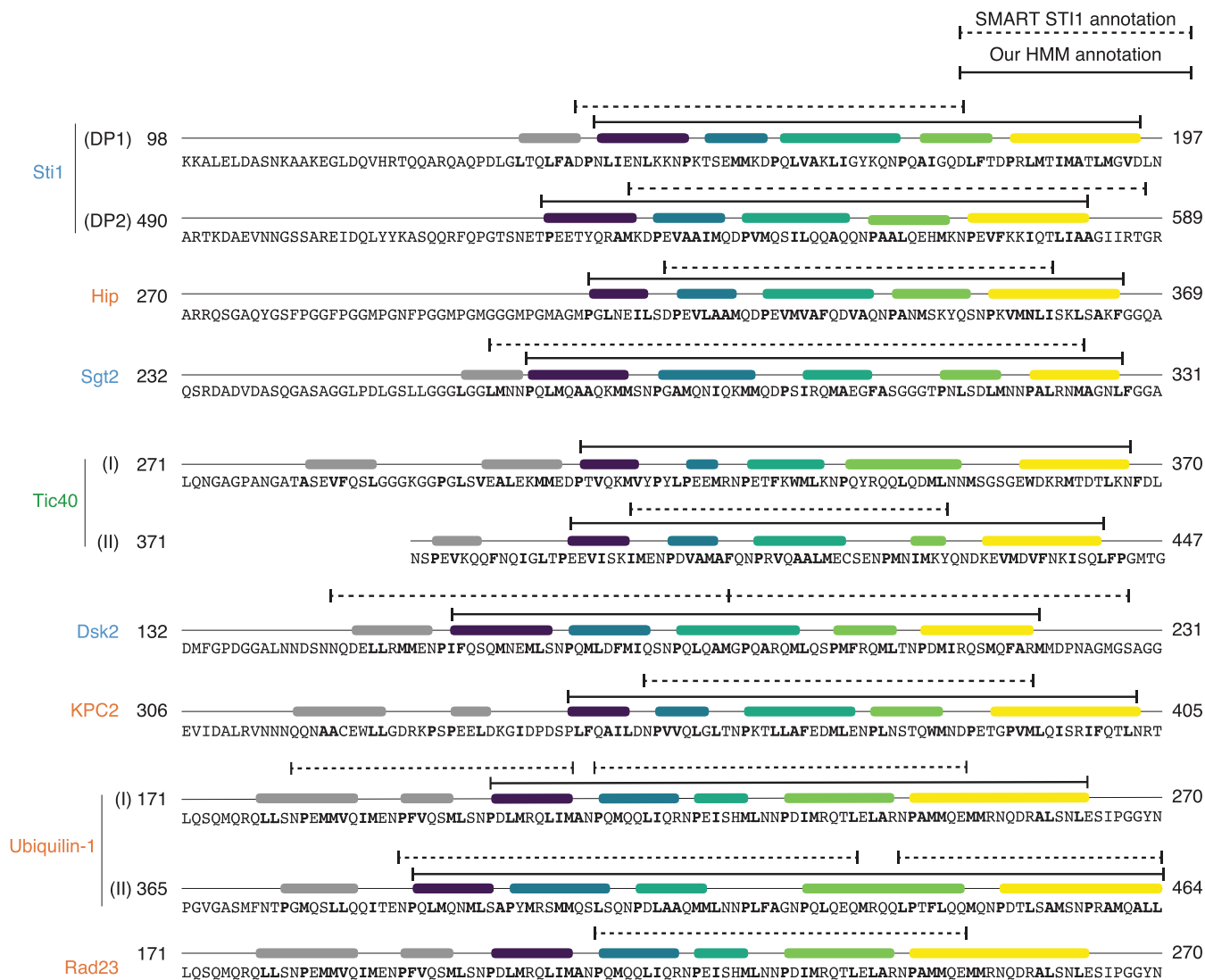
the binding-groove with a rotated helix organization compared to the other STI1-domains (Figure 2b). In this orientation, the first helix of Rad23-STI1 aligns to the third helix of Sti1-DP2 and the fourth (last) helix of Rad23-STI1 occupies a position similar to the first two helices of Sti1-DP2. Based on alignments and structures Tic40-STI1 is clearly a member of the STI1-domain family whereas Rad23-STI1 may be erroneously annotated (Figure 2a).

Despite its adequate utility, the SMART definition for STI1-domains can also lead to erroneous annotations of putative domains that have not been structurally characterized. Drawing the stretches of each protein with a hit from the SMART HMM alongside experimental or predicted helical regions, we can separate proper yet partial hits, for example, Sti1-DP1 and Sti1-DP2, from potentially erroneous ones (Figure 3). An erroneous hit might originate from more than one consecutive hit to the HMM. UBQLN-1, -2, & -4 are predicted by the SMART HMM as having two pairs of abutting STI1-domains, for a total of four (Figure 3, *Ubiquilin-1 I&II*).[65] In each case, the total length of each abutting hit was around 70 residues, not 100 which would be necessary for two STI1-domains considering the length of structurally determined STI1-domains. In addition, the secondary structure prediction within this region
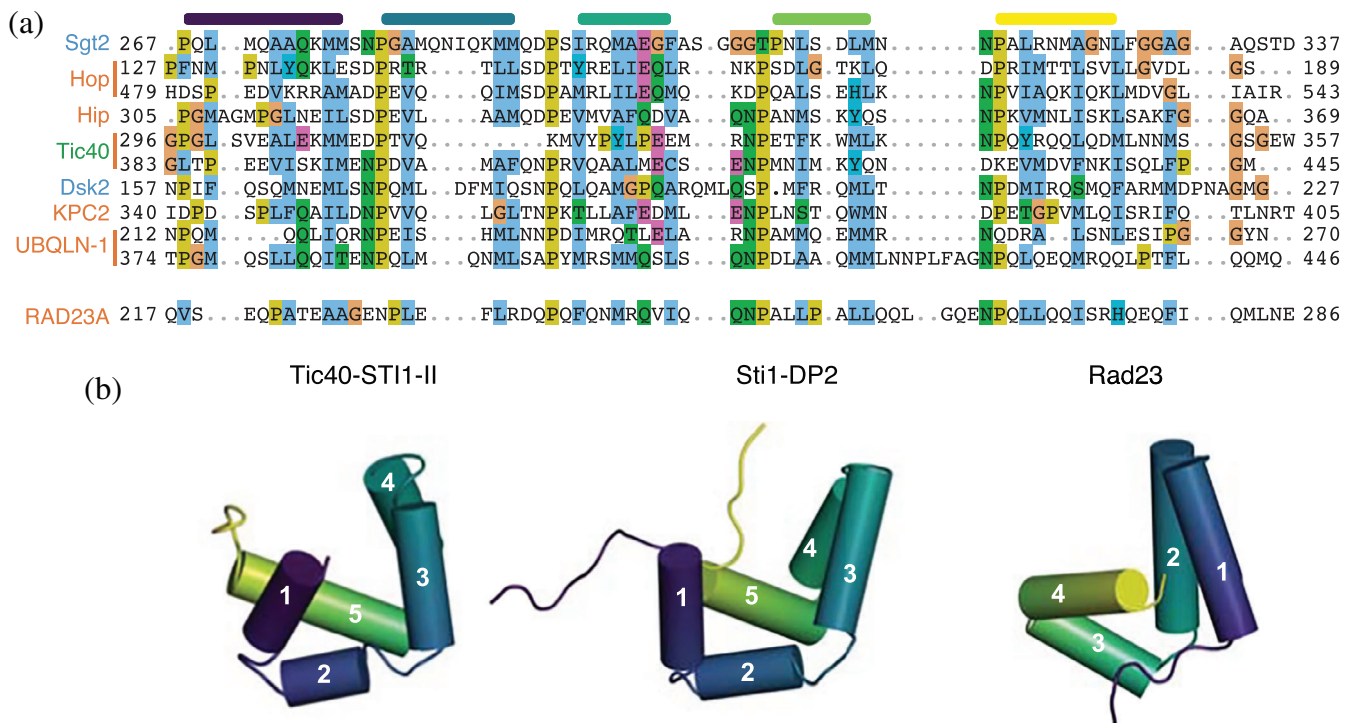
FIGURE 2 Structure-based sequence alignment of identified STI1-domains. (a) Structure-based alignment of identified STI1-domains colored using the ClustalX color scheme[100] with secondary structure elements indicated above the alignment. Names are as in Figure 1 with the addition of green for *A. thaliana*. (b) Cartoon representation of the structures of Tic40-STI1-II (PDBID: 2LNM), Sti1-DP2 (PDBID: 2LLW) and Rad23 (PDBID: 1X3W) colored using a viridis color scale from N-terminus (purple) to C-terminus (yellow). Helices are numbered in white

SMART STI1 annotation

Our HMM annotation

Sti1

(DP1) 98 — 197

KKALELDASNKAAKEGLDQVHRTQQARQAQPDLGLTQ**LFAD**P**NLIE**NLKKNPKTSE**MMKD**P**QLVAK**L**IG**YKQN**PQAIG**QD**LFTD**P**RLMT**I**MATLMG**V**D**LN

(DP2) 490 — 589

ARTKDAEVNNGSSAREIDQLYYKASQQRFQPGTSNET**PEETYQR**A**MKD**P**EVAAIM**QD**PVM**QSILQQAQQN**PAAL**QEHMKN**PEVF**KKIQT**LIAA**GIIRTGR

Hip 270 — 369

ARRQSGAQYGSFPGGFPGGMPGNFPGGMPGMGGGMPGMAGM**PGLNEI**LSD**PEVLAAM**QD**PEVMVAF**QD**VA**QN**PANM**SKYQSN**PKVMNLI**SKL**SAKF**GGQA

Sgt2 232 — 331

QSRDADVDASQGASAGGLPDLGSLLGGGLGG**LMNN**P**QLMQAA**QKMMSN**PGAM**QNIQ**KMMQD**P**SI**RQ**MAEGFA**SGGGT**PNLSD**L**MNN**PALR**NMAG**NLF**GGA

Tic40

(I) 271 — 370

LQNGAGPANGATA**SEVF**QS**L**GGGKGG**PGLSVEAL**EKMMED**PTVQKM**VYP**YLPEEMRNPET**FKWMLKNPQYRQQLQDMLNN**MSGSGEWDKRMTDT**LKNF**DL

(II) 371 — 447

NS**PEVKQQF**NQIG**LTPEEVISKIM**EN**PDVAMAF**QN**PRVQAALM**ECSEN**PMNIM**KYQNDKE**VMDVF**NKISQ**LFP**GMTG

Dsk2 132 — 231

DMFGPDGGALNNDSNNQDE**LLRMM**EN**PIF**QSQMNE**MLSN**P**QMLDFMI**QSN**PQLQAMG**P**QARQMLQS**PMF**RQM**LTN**PDMIRQS**MQFARM**MDPNAGMGSAGG

KPC2 306 — 405

EVIDALRVNNNQQNA**ACEWLLGDRKPSPEELDKGIDPDSPLFQAILDNPVVQLGLTNPKTLLAFEDMLENPLNSTQWMNDPETGPVMLQISRIFQTLNRT

Ubiquilin-1

(I) 171 — 270

LQSQMQRQ**LLSN**P**EMMVQIM**EN**PFVQSML**SN**PDLMRQLIMAN**P**QMQQLIQRN**PEISHMLNN**PDIMRQTLELARN**PAMM**QE**MMRNQDRALSN**LESIPGGYN

(II) 365 — 464

PGVGASMFNT**PGMQSLLQQITEN**P**QLMQNMLSAP**YMRS**MMQSLSQN**PDLAAQMMLNN**PLFAGN**PQLQEQMRQQLPTFLQQMQN**PDTLSAMSN**PRAMQALL

Rad23 171 — 270

LQSQMQRQ**LLSN**P**EMMVQIM**EN**PFVQSML**SN**PDLMRQLIMAN**P**QMQQLIQRN**PEISHMLNN**PDIMRQTLELARN**PAMM**QE**MMRNQDRALSN**LESIPGGYN

FIGURE 3 Redefining the STI1-domain model to properly account for number of helices. Secondary structure prediction is depicted for each identified STI1-domain. Helices residing within the STI1-domain are colored in viridis (as in Figure 1) and N-terminal helices, H0, are colored in grey. The HMMs are indicated by lines over the secondary structure for each model showing the SMART HMM (dashed line) and our new HMM (solid line). Hydrophobic residues within the STI1-domains and H0 are indicated in bold

showed only seven helices (Figures 2a and 3), sufficient for a single domain only. Thus, it is unlikely that there are two pairs of abutting domains.

As discussed earlier, the structure of the SMART identified STI1-domain in Rad23 differs from that of Sti1 DP domains and Tic40-STI1-II, again making it a possible erroneous annotation. The SMART HMM covers the same region as the Pfam XPC-binding HMM, but with a slightly lower score (27 vs. 43); both align over a similar number of residues (40 vs. 44). Given that the SMART HMM identifies the clear structural homolog Tic40-STI1-II with a similar score of 35, we cannot rule out Rad23 as a member of the STI1 containing family solely based on the SMART HMM score. Since the XPC-binding HMM uses Rad23-STI1 as part of the seed

sequences, it could also be possible that the XPC-binding domain is a subfamily of STI1-domains. Although the Rad23 structure in this region appears distinct from other STI1-domain structures, it could be a member of this family based on its score and alignment to the SMART HMM. An HMM with higher specificity could more clearly delineate the difference between Rad23-STI1 and other STI1-domains, which we now aim to define.

2.2 | A new definition for the STI1-domain

In light of these issues with the SMART definition for STI1-domains, we sought to generate an HMM that better

defines the full-length of the STI1-domain and thereby more sensitively captures the full breadth of the STI1 family. First, we created an alignment of protein sequences that correspond to both structurally characterized STI1-domains and close homologs then aligned others with constraints from molecular models (see below) and/or secondary structure predictions. As expected, this multiple sequence alignment reveals a strong conservation of structural features (Figure 2a). The predicted helical regions, helices 1–5 (H1–H5), in several proteins align directly with the structurally determined helices of DP2 from Sti1 and Tic40-STI1-II. This includes complete conservation of helix breaking prolines and close alignment of hydrophobic residues defining amphipathic helices. The amphipathic nature of these helices appear important for client binding; experiments mutating the hydrophobic faces in these helices in SGTA to less hydrophobic alanine affected binding to tail-anchored protein substrates.[20]

The resulting HMM clears up a number of the issues with the SMART HMM defining STI1-domains. Most identified STI1-domains align well between the two lists with the exception being Rad23 (Figure 2a). Along with other factors discussed later, this suggests Rad23 belongs to a class of STI1-like domains, which could include proteins like Ddi1.[66] In the case of UBQLNs, the resulting HMM identifies the annotated abutting STI1-domains as a single STI1-domain (Figure 3).

Opposite to what was revealed in the UBQLNs, where a reduction in the number of identified STI1-domains was observed, a second N-terminal STI1-domain is identified by our HMM in Tic40 (Figure 3). Previously this region of Tic40 was suggested to be a TPR domain primarily on the basis of a binding to an anti-TPR1 antibody by western blot.[38] While it is reasonable to suspect that Tic40 possesses a TPR domain since TPR domains precede STI1-domains in HOP, HIP, and SGTA, the TPR domain HMMs[62] do not suggest a hit in this region. However, this region does produce a hit by our STI1 HMM. The anti-TPR1 antibody was generated against full-length rat TPR1,[67] and it is possible it lacks specificity for this plant TPR domain. We were unable to identify identical peptides longer than five residues between Tic40 and rat STI1 that could easily explain the cross reactivity. Due to the bioinformatic support for a STI1-domain in this region, we refer to it as Tic40-STI1-I and the structurally solved STI1-domain as Tic40-STI1-II.

2.3 | STI1-domains are preceded by an N-terminal helix

Along with a curated set of STI1-domains, this new HMM reveals a conserved N-terminal sixth helix,

hereafter referred to as Helix 0 (H0), which like H1-H5 is also amphipathic (Figure 4). As already noted, UBLQNs were mistakenly characterized as having four STI1-domains. The incorrect annotation was likely because three helices N-terminal to the domain were combined with the five helices of the STI1-domain, which resulted in eight contiguous helices recognized by SMART as two adjacent STI1-domains. The presence of N-terminal helices appears to be a general feature of STI1-domains. Based on secondary structure prediction and structures, it is clear, in most cases, that at least one helix precedes the STI1-domain, the exceptions being Sti1 DP2 and HIP (Figure 3). While the roles of the additional helices are not clear, H0 is well conserved within each protein and are also amphipathic in nature as the other helices (Figures 3 and 4).
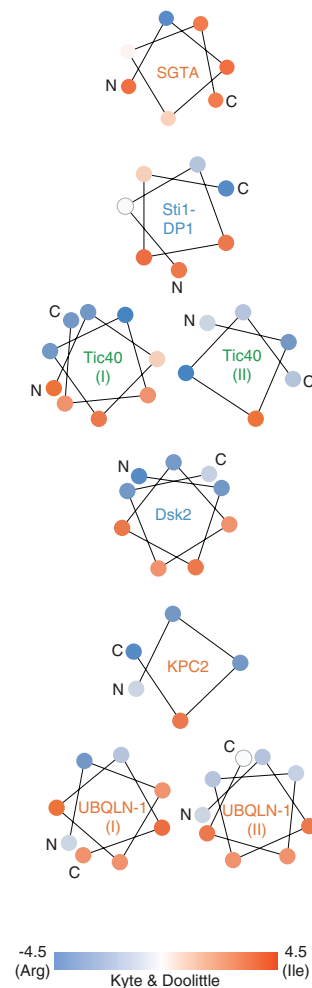

FIGURE 4 The amphipathic nature of the N-terminal helix preceding STI1-domains. Helical wheel plots of the H0 helix immediately preceding identified STI1-domains. Residues making up H0 are represented as circles and colored based on their hydrophobicity using the Kyte & Doolittle scale.[101] The N-terminus (N) and C-terminus (C) of each helix is annotated

2.4 | Structural similarity between STI1-domains

With this new list of STI1-domains we inspected their predicted and structurally determined secondary structures. Broadly, these domains share several features including four to five amphipathic helices, as annotated (Figure 2a). For STI1-domains that have been structurally characterized (DP1, DP2, Tic40-STI1-II), the helices assemble into a tertiary structure that resembles a helical-hand forming a hydrophobic groove (Figure 5a–c) and are characterized by structural flexibility.[20] Though no structures of a STI1-domain from a co-chaperone exist with a client occupying the hydrophobic groove, it
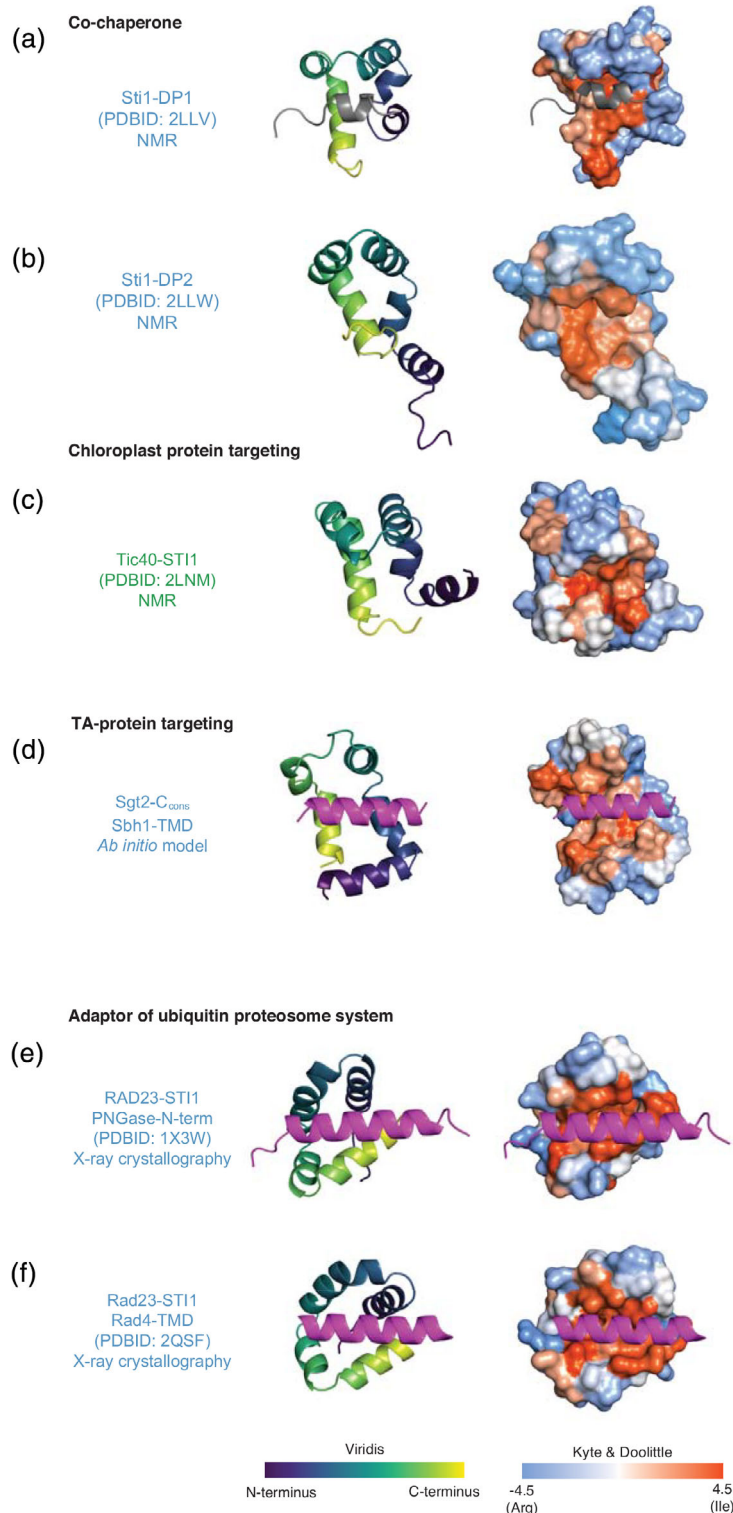
FIGURE 5 Published structures and models of STI1-domains reveal an alpha-helical hand that forms a hydrophobic groove. Structural models of STI1-domains as cartoon and surface hydrophobicity representations: DP1 and DP2 from yeast Sti1 (Sti1-DP1 and Sti1-DP2) (PDBID:2LLV and 2LLW), Tic40 from *Arabidopsis thaliana* (Tic40-STI1) (PDBID:2LNM), a computational model of the C-domain from the yeast co-chaperone Sgt2, yeast Rad23 (Rad23-STI1) bound to the N-terminus of PNGase (PNGase-N-term) (PDBID: 1X3W) and yeast Rad23 bound to the TMD of yeast Rad4 (Rad4-TMD) (PDBID: 2QSF). Cartoons are colored as in Figure 1b with bound helices in magenta. The surface representation is colored based on hydrophobicity

presents an appealing pocket for the binding site of hydrophobic segments. The flexibility may contribute to the ability of these domains to specifically bind and then release substrates as part of their functional role.

DP1 and DP2 were the first STI1-domains to be structurally characterized. The two structures have five amphipathic alpha helices arranged like a cupped hand presenting a hydrophobic groove. In the structure of DP1 the hydrophobic hand is occupied by H0, possibly mimicking client binding (grey helix Figure 5a). Both structures of DP1 and DP2 were solved using NMR. When comparing the states from the models for each domain, DP2 appears to be more flexible than DP1, with the N-terminal H0 in DP1 likely stabilizing its core region.[2]

Tic40 contains another structurally characterized STI1-domain in the absence of a client occupying the groove. Like DP1 and DP2, Tic40-STI1-II consists of five alpha helices that arrange into a similar helical hand with a hydrophobic groove (Figure 5c). Tic40 is predicted to have an H0 that was not included in the determined structure. Other STI1-domains have remained resistant to structure determination.

Alternative structural methods have been used to characterize other STI1-domains. One domain in particular is the C-terminal domain of fungal homolog of SGTA, Sgt2, which remains recalcitrant to experimental structural determination. Ab initio molecular modeling of Sgt2-C followed by experimental validation of residue-pair distances further suggests that the domain is part of the STI1 family[20] (Figure 5d). Residues of a conserved region resolve a potential binding interface for a helical hydrophobic substrate. Outside this region they adopt varied conformations consistent with expected high flexibility. The working model contains a potential TA 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 alpha-helix, ~11 amino acids. Like the NMR structures of other STI1-domains found in co-chaperones, the ab initio model of Sgt2-C resembles the general STI1-domain structure.

For the STI1-domain containing AUPS proteins, no experimental or ab initio structures currently exist. To predict structures of these STI1-domains including those in the UBQLNs we employed the Robetta transform-restrained (TR) tool, a state-of-the-art structure prediction method[68] where a deep neural network predicts pairwise residue distances and angles followed by energy minimization. Distinct from template- or fragment-based approaches, Robetta TR generates de novo structures from restraints where structures are not explicitly used. As validation for our domains, we first compare the prediction of multiple STI1-domains by Robetta TR versus the experimentally derived structure. Providing the full-length sequences from ScSti1 and AtTic40, Robetta TR provides a model of the full-length protein (Figure S1A). We isolated the STI1-domains from each and compare them to the structures solved by NMR (Figure 6a, Figure S1). For the top predicted models, DP1, DP2, and Tic40-STI1-II have five helices that assemble into a helical hand (Figure 6a, Figure S1B,C). These predictions are in close agreement with the NMR derived structures, with the last five helices of the prediction overlaying with the five alpha helices in the NMR structure—supporting that the prediction method can provide data broadly on STI1-domains (Figure 6a, Figure S1B,C).

We proceeded to predict the structures of uncharacterized STI1-domains in other co-chaperones. The predicted structure for Tic40-STI1-I has a similar fold to other STI1-domains, the five alpha-helical hand, supporting the new domain definition (Figure 6b). This model further reduces the likelihood that this region contains a TPR domain that is structurally distinct alpha-solenoids. The STI1-domain from human HIP (Figure 6c), the last of the uncharacterized co-chaperone STI1-domains, is similar to the experimentally determined Tic40-STI1-II (Figure 6a), DP2, and DP1 (Figure 5a,b). Across all co-chaperones we observe five helices coming together to form an alpha helical hand.

We next predicted the structures of the uncharacterized STI1-domains in the AUPS family (Dsk2, KPC2, and UBQLNs) (Figure 6d–g). Like HIP, the predicted structure of KPC2 is consistent with the experimentally determined structures of STI1-domains from the co-chaperones. The predicted structure of the Dsk2 and UBQLN-1 also display a variation of the STI1 helical hand observed in the solved structures (Figure 6d,f,g). A helical groove compatible for binding a hydrophobic alpha helix is formed by five helices in each of these cases. Dsk2-STI1 (Figure 6d) and the second STI1-domain of UBQLN-1 (UBQLN-1-STI1-II) differs from the co-chaperone STI1-domains as their helices form the groove in the reverse order (Figure 6g). Unlike the prediction of UBQLN-1-STI1-II, the first STI1-domain (UBQLN-1-STI1-I) forms a nearly enclosed groove, which could accommodate client binding with some rearrangement (Figure 6f). These predictions are models for what these STI1-domains could look like and in both the co-chaperones and AUPS these models suggest that these STI1-domains can form helical hands to accommodate client binding. UBQLNs (UBQLN 1–4) are a particularly important focus of research and experimental work to test these predictions will be broadly useful.[60,69–72]

The structure of Rad23-STI1 bound to substrate supports the exclusion of Rad23 from the STI1-domain containing family of proteins (Figure 5d,e). While several
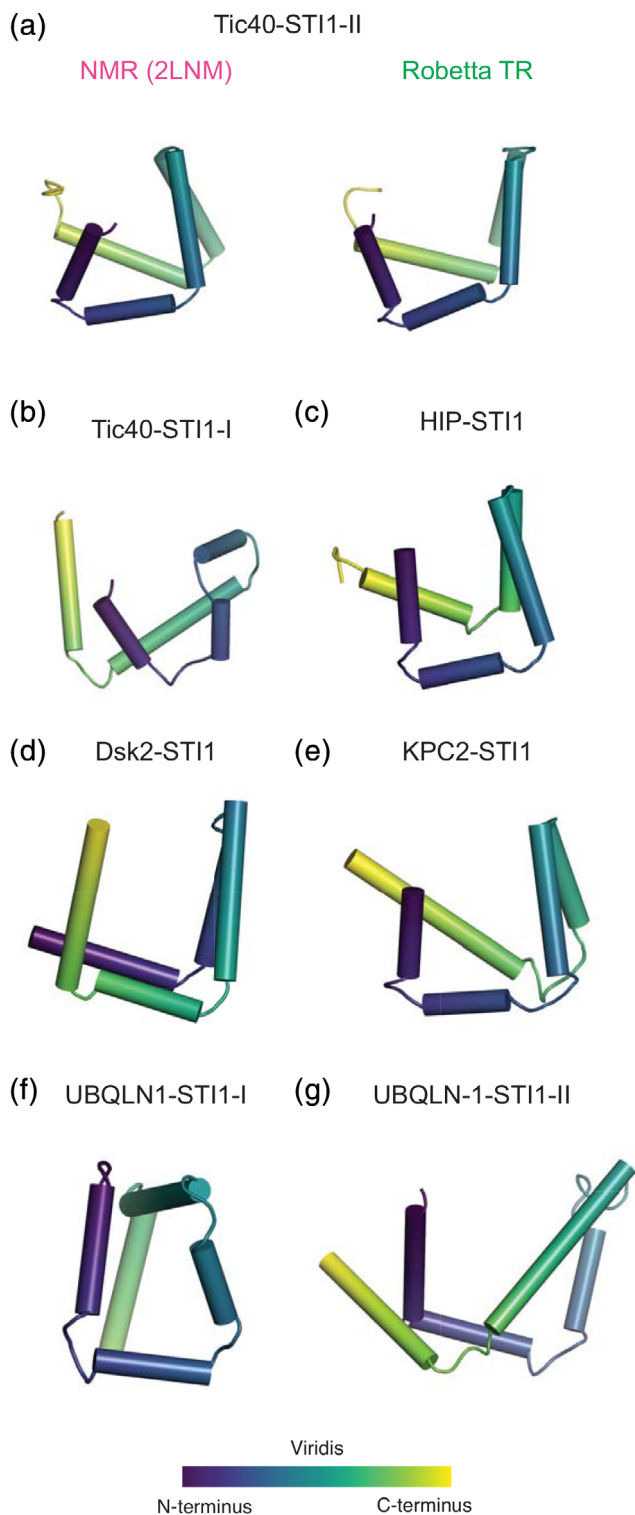
FIGURE 6 Predicted structures of uncharacterized STI1-domains reveal a hydrophobic groove as seen in the NMR solved structures. (a) A comparison of the Tic40-STI1-II structure model either determined by NMR (left) or predicted using Robetta-TR (right). The predicted models using Robetta-TR of the uncharacterized STI1-domain(s) from (b) Tic40 (Tic40-STI1-I), (c) HIP, (d) Dsk2, (e) KCP2, and (f,g) UBQLN-1. All structures are colored as in Figure 1b

structures of complexes of Rad23-STI1 bound to amphipathic clients show in each that the client-helix binds via a hydrophobic groove, the domain architecture differs from those determined in the co-chaperones and the predicted structures of other STI1-domain containing proteins (Figures 2c and 6b–g). Despite Rad23-STI1 being a helical bundle that binds clients similar to the co-chaperones, the absence of a fifth helix supports that Rad23 does not contain a STI1-domain, but is instead a STI1-like domain which also utilizes a hydrophobic groove. It has been observed that the first three helices of the Rad23 XCP domain are structurally similar to the first three helices of the N-terminal domain of the Helical Domain (HDDnt) from the DNA damage inducible 1 protein (Ddi1), but the fourth helix deviates and goes in a different direction.[66] Ddi1 differs from other shuttle proteins because of its proteolytic role and interacting partners.[5] Like the shuttle proteins described above, Ddi1 contains a UBA and UBL domain, but the UBA domain has been lost in mammalian homologs. Ddi1 also contains a retrovirus protease (RVP) fold domain. The UBL domain of Ddi1 has an unusual binding preference, unlike the domain in Rad23 or Dsk2, it does not interact with its UBA domain or Rpn10 and interacts weakly with Rpn1. It has been suggested that Ddi1 may assist Rad23 or Dsk2 instead of acting as a shuttle factor on its own.

Due to its homology with the XCP domain of Rad23, which has been implicated through protein–protein interactions, HDDnt may play a similar role.[66] It has also been observed that HDDnt has a similar structure to other DNA binding domains suggesting HDDnt may bind directly to DNA. With structural similarities to various domains, it is reasonable to think Rad23 and Ddi1 both contain STI1-like domains.

2.5 | Similarity in the domain structures of STI1 proteins

When examining the predicted secondary structure of the entirety of STI1-domain containing proteins several common characteristics became clear. Dual STI1-domains are present in both the co-chaperones and AUPS (Figure 7). Within the co-chaperones, two distinct groups emerge—ones that possess two STI1-domains (HOP, Tic40) and those that possess a dimerization domain (SGTA, HIP). As stated previously, HOP contains two STI1-domains separated by multiple TPR domains with both required for efficient client transfer.[2] It has been speculated that DP1 and the first TPR domain (TPR1) act as an intermediate in the shuttling of a client from Hsp70 to the TPR2A&B and DP2-bound Hsp90.[2,19,73–78]
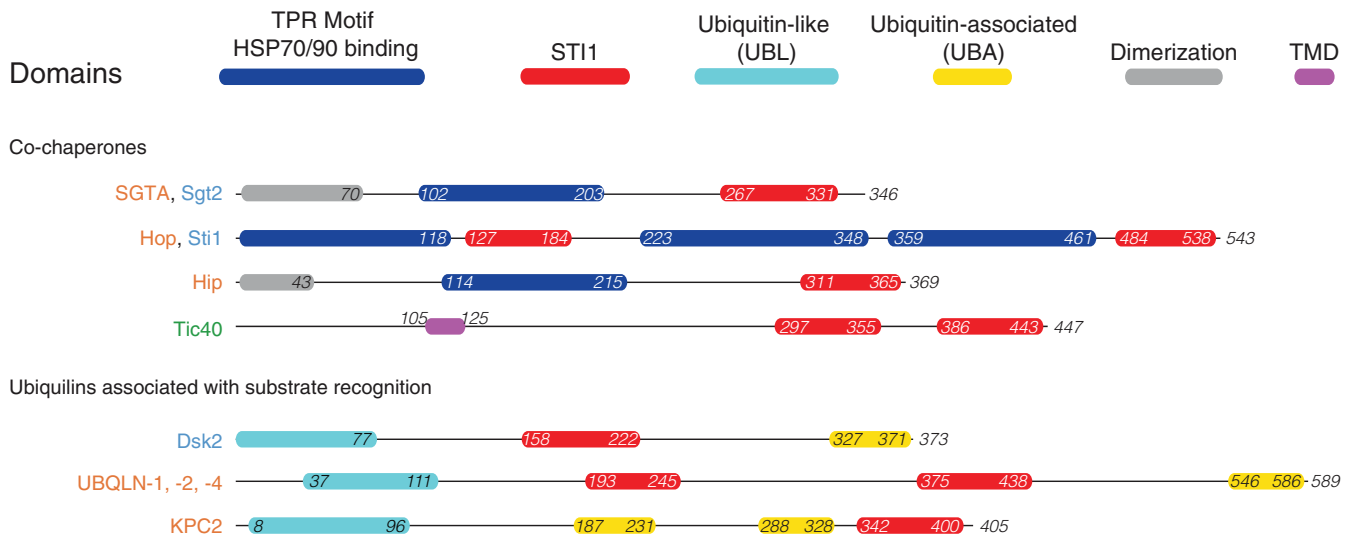
FIGURE 7   Various domain architectures of STI1-domain containing proteins. The domain definitions of proteins containing at least one STI1-domain were obtained initially from InterPro[102] and then adjusted as discussed in the text. Each domain within a protein is colored relative to the key. Numbering where it is not clear is in reference to the human protein. Names are colored as in Figures 1 and 2

Unlike HOP, Tic40 contains two abutting STI1-domains. Given our recent identification of the first STI1-domain, its function has yet to be determined. Found in the chloroplast inner membrane with the STI1-domains in the stroma, the C-terminal domain can be replaced with the STI1-domain from HIP without loss-of-function.[79] In HIP the STI1-domain interacts with the leucine-rich chemokine receptor,[43] the previously proposed TPR domain, now STI1-I, of Tic40 interacts with the leucine-rich mature region of Tic110.[39]

The homodimerization of HIP and SGTA, each containing a single STI1-domain in the monomer, results in a complex with two STI1-domains. Small angle X-ray scattering (SAXS) data has revealed that for both proteins in solution, the dimers[22] form elongated, flexible complexes.[3] The elongated form would put maximal distance between the two STI1-domains which are found opposite of the dimerization domains. The fact that the STI1-domain from HIP can functionally replace a STI1-domain from Tic40, both in pairs, but one through dimerization domain and the other encoded in the monomer, suggests the STI1-domains have similar overall functions, the significance of these pairs in the co-chaperones is still unclear.

The presence of a pair of STI1-domains is also observed in UBQLNs. As discussed earlier, UBLQNs contain two STI1-domains and not four as previously thought. This clarification relates UBQLNs to the co-chaperones HOP and Tic40, where each have a STI1-domain pair encoded in their monomer, while still separating them from other AUPS family members which contain a single STI1-domain (Figure 7). The observation of a pair of STI1-domains in UBLQNs and the previously identified active roles in protein targeting and degradation[60] set these proteins apart from the other AUPS family members. The previously defined M domain of UBQLN-1 contains both identified STI1-domains and is responsible for its ability to shield TMDs of mitochondrial membrane proteins from the cytosol and to deliver them for degradation.[60] Identifying pairs of STI1-domains in both co-chaperones and UBLQNs, proteins with a known role in protecting TMDs in the cytosol through their STI1-domains, suggests these pairs aid in this role.

While a pair of STI1-domains is a found in both AUPS and co-chaperone proteins, an HSP-binding TPR domain preceding the STI1-domain(s) connected by a flexible linker is exclusively observed in the co-chaperones (Figure 7). These TPR domains have been shown to aid in client hand-off in these proteins. The multiple TPR domains in HOP are used to coordinate simultaneous binding of Hsp70 and Hsp90, facilitating client transfer between the two chaperones.[2,80,81] In contrast, HIP contains a TPR domain that only interacts with Hsp70. Additionally in Sgt2, the TPR domain increases the efficiency of capture of TA clients by coordinating with a client bound Hsp70 homolog, Ssa1.[82] While HOP has two TPR domains within a monomer, both SGTA and HIP link two TPR and STI1-domains by forming stable dimers via N-terminal dimerization domains.[83] For the SGTA and HIP homodimers, a cooperative role between the two copies of each TPR- and STI1-domain remains a possibility.

Differing from the other co-chaperones, the relatively more distant chloroplast Tic40 has its own domain

architecture. Previously, the N-terminal STI1-domain was annotated as a TPR domain but, as discussed earlier, bioinformatics and computational models counter this claim. The rest of the protein lacks a clear TPR domain and has an N-terminal TMD. How Tic40 fits mechanistically into this group of co-chaperones is less clear due to it missing a TPR domain and being membrane bound.

2.6 | Amino acid distribution in STI1-domains

STI1-domains were initially described as DP domains due to two repeats of a DPEV motif in HIP[10] and a DPEV and DPAM motif in HOP.[9] From the solved structures (Figure 5) and predicted secondary structure (Figure 3), we see this motif localizing to the N-terminus of helices likely acting as a cap. Aspartate and threonine most frequently occur at the cap of a helix often followed by either a glutamine or proline.[84] The role for this motif where found is likely as a stabilizing N-cap accounting for its conservation. When analyzed broadly, a repeat DP motif is not observed in the majority of STI1-domains, it not even found in all HOP homologs.

We were interested if there were common residues overrepresented in STI1-domains, considering that in some cases these domains have been referred to as methionine-rich.[20,60] To quantify this, we began with compiling a list of homologs for HOP, SGTA, HIP, UBQLN-1, and KPC2 using EnsemblGenomes[85] and then calculated the distribution of amino acids within the STI1-domain. As for Dsk2 homologs, only two hits were found after searching EnsemblGenomes, therefore we omitted Dsk2 from this analysis. Only an overrepresentation of methionine, asparagine, and leucine is observed across all five protein groups (Figure 8). An overrepresentation of methionine has also been observed in other hydrophobic segment binding domains such as the M domain of the signal recognition
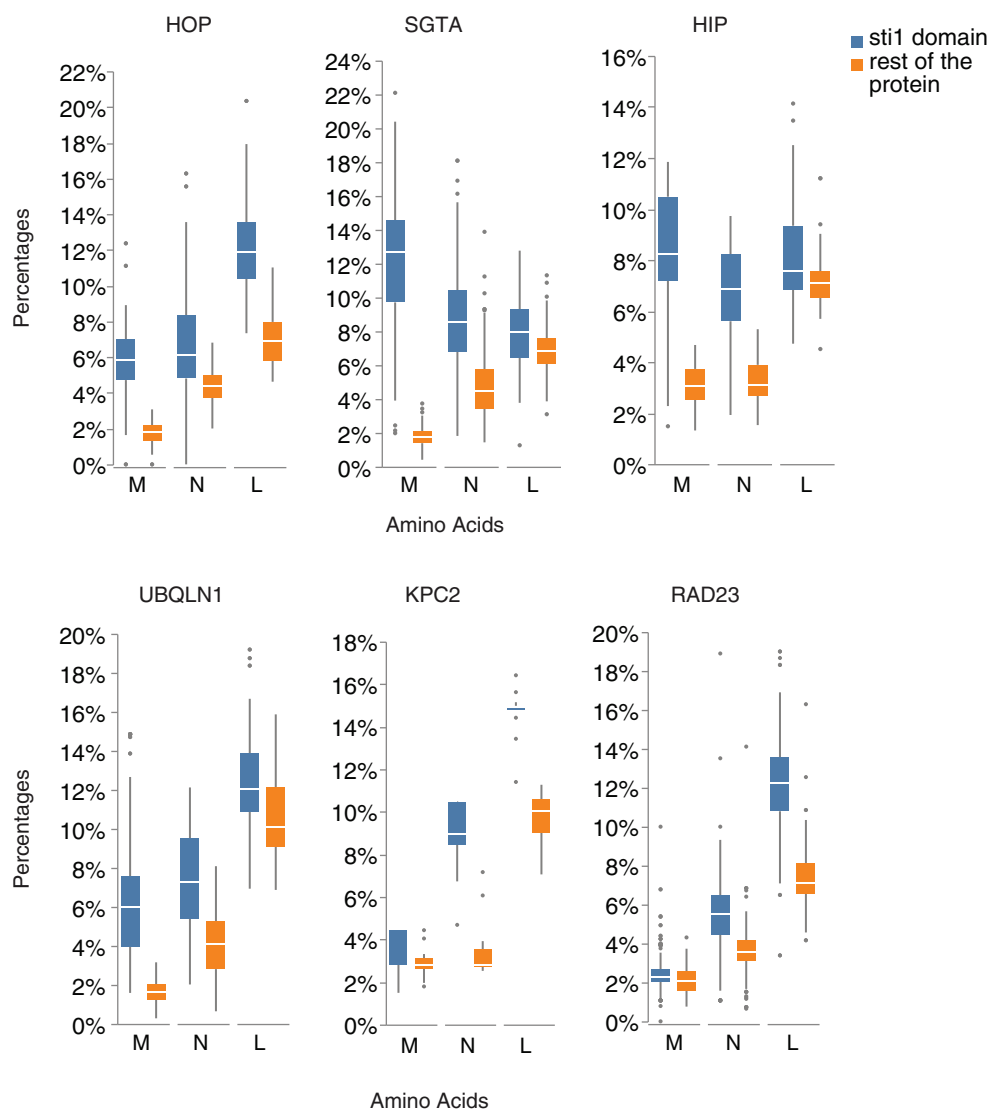
FIGURE 8 Methionine, asparagine, and leucine are overrepresented in STI1-domains. (a) Box plots of the prevalence of methionine (M), asparagine (N), and leucine (L) in homologs of HOP (176 sequences), SGTA (155 sequences), HIP (39 sequences), UBQLN-1 (95 sequences), KPC2 (11 sequences), and Rad23 (169 sequences) within the STI1-domains (blue) and the rest of the protein (orange). Amino acids that were overrepresented across all five STI1-domain containing protein groups are shown

particle (SRP)[86,87] and Get3.[88,89] As discussed previously, Rad23 likely does not contain a STI1-domain. We applied the same analysis of the amino acid distribution in the previous annotated STI1-domain of Rad23 and found that, unlike verified STI1-domains, methionine is not overrepresented (Figure 8). Overall, this analysis reveals that over representation of methionine, asparagine, and leucine is a feature of STI1-domains.

3 | DISCUSSION

STI1-domains have been predicted in a number of proteins essential for protein biogenesis. Here we explicate a definition for STI1-domains and curate a list of STI1-domain containing proteins through structure-based sequence alignments, validating some previously predicted domains as well as identifying new ones. Solved structures and computational models reveal STI1-domains consist of five to six helices organized into a helical hand with a hydrophobic groove. Upon close inspection, STI1-domain containing proteins can be classified into two families—co-chaperones or AUPS—with several common features noted. Overall, this work presents the first in-depth examination of STI1-domains and the essential proteins for where they are found.

Previously, STI1-domains were identified by the SMART database, which lead to the identification of erroneous hits and omissions for the family. Our revised HMM that encompasses a minimal five helix region helps uncover and clarify the full breadth of STI1-domains. As a result, this new definition revealed new STI1-domains and corrections to previous identifications. For example, the annotations of four abutting STI1-domains in UBQLNs are actually a set of two STI1-domains and the annotated TPR domain in the co-chaperone Tic40 is more likely a STI1-domain. Furthermore, it is now clear that Rad23 does not contain a STI1-domain, but has a distinct helical hand formed of only four helices.

This revised list was evaluated to determine common structural features. The overall five amphipathic helices forming a flexible helical hand are seen in the structures of DP2 and Tic40-STI1-II.[2] Differing from the structure of the DP2, the DP1 structure contains a sixth helix (H0) that resides in the groove (Figures 1b and 5). With our new HMM, an H0 was identified preceding most STI1-domains and its role is yet to be determined (Figure 3). Due to the flexibility of STI1-domains, one possibility is that H0 fills the groove as seen in the DP1 structure.[90] In this model, the hydrophobic residues in H0 would dock in the groove of the helical hand stabilizing the domain in the absence of client. H0 would then be displaced by an appropriate client.

Outside of the STI1-domains, several common features surface in these STI1-domain containing proteins. A distinct feature of the co-chaperone family is the TPR domains. We previously discussed the various TPR domains identified in Sgt2, HOP, and HIP, noting sequence features that define specificity in interacting partners.[22] TPR domains consist of multiple repeats of 34 amino acids in a helix-turn-helix fold, with antiparallel alpha-helices. Differences in the binding pocket of TPR domains allow for selectivity of a diverse set of chaperone partners. In the case of SGTA, the TPR-domain works together with the STI1-domain to coordinate client binding. Ssa1 carrying STI1-domain clients interact directly with the TPR domains allowing for client capture by the STI1-domains. Here we demonstrate that the suggested TPR domain of Tic40 is more likely a STI1-domain based on its higher score against the STI1 HMM versus the TPR HMM. This adjustment to the domain structure within Tic40 may suggest that Tic40 does not interact with HSPs to capture clients as seen for the other co-chaperones.

Pairs of STI1-domains are found in both co-chaperones and the AUPS family, either encoded in a single monomer or joined through a dimerization region. While the co-chaperones identified in this paper contain STI1-domains in pairs, for the AUPS family this is only true for UBQLNs. Of the AUPS family members, UBQLNs are the only ones so far shown to play a direct role in preventing client aggregation in the cytosol and facilitating the degradation of mitochondrial membrane proteins that fail to insert into the mitochondrial membrane.[60] These roles in both protein targeting and degradation are similar to those of SGTA—handing off TA clients to chaperones in the GET pathway for insertion and handing off mislocalized proteins to Bag6 for degradation.[91] The pair of STI1-domains lies in the identified client binding domain of UBQLN-1. Perhaps due to similar roles, these domains in UBQLNs function similarly to the STI1-domains in SGTA. The details of how STI1-domain pairs affect function and if they interact with one-another are important areas for future study.

How might pairs of STI1-domains cooperate for client specificity and selection? Conceptually, a pair of STI1-domains may simultaneously bind the same client TMD in the case of UBQLNs and SGTA. We have shown previously that a single STI1-domain from Sgt2 can bind to a minimum of 11 amino acids in a client.[20] Conceivably, one model is that the two STI1-domains in the Sgt2 dimer simultaneously bind a single client—binding side-by-side on a client TMD that averages 20aa. This would require that these domains come close together altering the overall architecture. A related model is that the pair of STI1-domains could cooperate

to increase the apparent affinity for a client TMD by increasing the local concentration of the binding domain. The simplest model is that each STI1-domain binds a separate client either to increase client load (two clients per monomer/dimer instead of one) or there is a necessity to bind two clients at once. On the other hand, it is also possible that two STI1-domains are necessary for different functions, as proposed previously in the case of HOP by Schmid and colleagues where glucocorticoid receptor activation cannot by rescued by replacing DP2 with DP1. It is worth noting that low resolution structural studies have suggested that dimeric Sgt2 and HIP position their STI1-domains on opposite ends of a dimer molecule in the absence of client.[22,83] Still, given the noted flexibility in these proteins,[2,20] the possibility of cooperation remains, with the molecular details an open question.

Flexibility is a common motif seen within STI1-domains and the proteins where they are identified. NMR studies of STI1-domains have suggested that these domains are flexible. We consider this flexibility a feature of these helical-hands for reversible and specific binding of a variety of clients. But 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 with experimentally determined structures, the flexible helical-hands provide an extensive hydrophobic surface to capture the client-helix. Required to only engage temporarily, the flexibility of the helical hand could offset the favorability of the domain to bind a hydrophobic client, allowing the client to be released. This would account for the favorable transfer seen from Sgt2[82] and SGTA[92] to downstream components.

This work provides a comprehensive HMM to define and identify STI1-domains in proteins and recognizes common features observed within the domains themselves and across STI1-domain containing proteins. These patterns leave open questions that have yet to be determined. What is the role of flexibility within STI1-domains and STI1-domain containing proteins? Does helix zero act as a stand in for clients in their absence or does it have a role as a lid? What is the benefit of having two STI1-domains in the co-chaperones? Do UBQLNs have a larger role than other AUPS family members and do their two STI1-domains contribute to this different role? There is much still to be understood about the underlying mechanisms that result in specificity and client handoff of STI1-domains. This comprehensive list of STI1-domains provides a coherent starting point.

4 | MATERIAL AND METHODS

4.1 | Molecular visualization

All STI1-domains with experimentally determined structures were retrieved from the RCSB. Ab initio structure prediction was employed for other STI1-domains using RobettaTR (transform restrained)[93] with the full-length protein sequences of each protein, with the specific STI1 region of interest visualized. Images were rendered using PyMOL 2.4 (www.pymol.org) with a viridis coloring scheme.[94] Helical wheel diagrams were rendered in R using a fork (https://github.com/smsaladi/heliquest) of the HELIQUEST source code.[95]

4.2 | Sequence analyses

Alignments of Sti1 (DP1/DP2) and STI1-domains were created by pulling all unique domain structures with annotated STI1-domains from Uniprot. Sequences were clustered at 50% similarity present, with the human, yeast, and *A. thaliana* preferred and then aligned with PROMALS3D[64] along with all experimentally determined structures of STI1-domains. PROMALS3D provides a way of integrating a variety of costs into the alignment procedure, including 3D structure, secondary structure predictions, and known homologous positions. The human, yeast, and *A. thaliana* homologs were selected from this alignment for display. An HMM for the STI1 domain was generated using HMMER v3.3.1. Alignments were visualized using Jalview.[96] Secondary structure where indicated is calculated using DSSP[97] on experimentally determined or predicted structures.[93]

4.3 | Amino acid composition of STI1-domains

Homologs of *Sc*Sti1, *Hs*HOP, *Sc*Sgt2, *Hs*SGTA, *Hs*HIP, *Sc*Dsk2, *Sc*KPC2, *Hs*UBQLN-1, and *Hs*Rad23A were compiled from EnsemblGenomes[4] and filtered for redundancy at 70% sequence identity using the CD-HIT Suite.[98] Sequences were then aligned using MAFTT.[99] The STI1-domain(s) in each sequence were identified by alignment to the known STI1-domains of *Hs*HOP and *Hs*SGTA to yield two segments: the STI1-domain and the non-STI1-domain region, that is, the "rest" of the protein. For each segment of each protein, the percentage of all individual amino acids was calculated. Significance was determined by permutation testing, comparing the difference between the first quartile of an amino acid's percentage between each

segment, that is, within versus outside of the STI1-domain.

AUTHOR CONTRIBUTIONS

Michelle Fry: Conceptualization; data curation; formal analysis; investigation; methodology; validation; visualization; writing-original draft; writing-review & editing. **Shyam Saladi:** Conceptualization; data curation; formal analysis; investigation; methodology; software; validation; visualization; writing-original draft; writing-review & editing. **William Clemons:** Conceptualization; project administration; supervision; writing-original draft; writing-review & editing.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

DATA AVAILABILITY STATEMENT

Our refined HMM for the STI1-domain (HMMER format), alignment shown in Figure 2 (FASTA format), and each of the structural models (PDB format) are included as supplementary data. Structural models should be used with caution.

ORCID

William M. Clemons Jr https://orcid.org/0000-0002-0021-889X

REFERENCES

1. Letunic I, Bork P. 20 years of the SMART protein domain annotation resource. Nucleic Acids Res. 2017;46:D493–D496.
2. Schmid AB, Lagleder S, Gräwert MA, et al. The architecture of functional modules in the Hsp90 co-chaperone Sti1/Hop. EMBO J. 2012;31:1506–1517.
3. Li Z, Hartl FU, Bracher A. Structure and function of Hip, an attenuator of the Hsp70 chaperone cycle. Nat Struct Mol Biol. 2013;20:929–935.
4. Howe KL, Contreras-Moreira B, De Silva N, et al. Ensembl genomes 2020—Enabling non-vertebrate genomic research. Nucleic Acids Res. 2019;48:D689–D695.
5. Zientara-Rytter K, Subramani S. The roles of ubiquitin-binding protein shuttles in the degradative fate of ubiquitinated proteins in the ubiquitin-proteasome system and autophagy. Cell. 2019;8:40–32.
6. Weisman CM, Murray AW, Eddy SR. Many, but not all, lineage-specific genes can be explained by homology detection failure. PLoS Biol. 2020;18:e3000862.
7. Caplan AJ. What is a co-chaperone? Cell Stress Chaperones. 2003;8:105–107.
8. Prodromou C. Mechanisms of Hsp90 regulation. Biochem J. 2016;473:2439–2452.
9. Chen S, Smith DF. Hop as an adaptor in the heat shock protein 70 (Hsp70) and Hsp90 chaperone machinery. J Biol Chem. 1998;273:35194–35200.
10. Prapapanich V, Chen S, Smith DF. Mutation of hip's carboxy-terminal region inhibits a transitional stage of progesterone receptor assembly. Mol Cell Biol. 1998;18:944–952.
11. Onuoha SC, Coulstock ET, Grossmann JG, Jackson SE. Structural studies on the co-chaperone hop and its complexes with Hsp90. J Mol Biol. 2008;379:732–744.
12. Nelson GM, Huffman H, Smith DF. Comparison of the carboxy-terminal DP-repeat region in the co-chaperones Hop and Hip. EMBO J. 2003;8:125–133.
13. Kajander T, Sachs JN, Goldman A, Regan L. Electrostatic interactions of Hsp-organizing protein tetratricopeptide domains with Hsp70 and Hsp90: Computational analysis and protein engineering. J Biol Chem. 2009;284:25364–25374.
14. Martins VR, Graner E, Garcia-Abreu J, et al. Complementary hydropathy identifies a cellular prion protein receptor. Nat Med. 1997;3:1376–1382.
15. Zanata SM, Lopes MH, Mercadante AF, et al. Stress-inducible protein 1 is a cell surface ligand for cellular prion that triggers neuroprotection. EMBO J. 2002;21:3307–3316.
16. da Fonseca ACC, Matias D, Geraldo LHM, et al. The multiple functions of the co-chaperone stress inducible protein 1. Cytokine Growth Factor Rev. 2021;57:73–84.
17. Prapapanich V, Chen S, Toran EJ, Rimerman RA, Smith DF. Mutational analysis of the hsp70-interacting protein Hip. Mol Cell Biol. 1996;16:6200–6207.
18. Lässle M, Blatch GL, Kundra V, Takatori T, Zetter BR. Stress-inducible, murine protein mSTI1. Characterization of binding domains for heat shock proteins and in vitro phosphorylation by different kinases. J Biol Chem. 1997;272:1876–1884.
19. Reidy M, Kumar S, Anderson DE, Masison DC. Dual roles for yeast Sti1/Hop in regulating the Hsp90 chaperone cycle. Genetics. 2018;209:1139–1154.
20. Lin K-F, Fry MY, Saladi SM, Clemons WM Jr. The client-binding domain of the cochaperone SGTA/Sgt2 has a helical-hand structure that binds a short hydrophobic helix. J Biol Chem. 2020; in press. https://doi.org/10.1016/j.jbc.2021.100441.
21. Wang F, Brown EC, Mak G, Zhuang J, Denic V. A chaperone cascade sorts proteins for posttranslational membrane insertion into the endoplasmic reticulum. Mol Cell. 2010;40:159–171.
22. Chartron JW, Gonzalez GM, Clemons WM Jr. A structural model of the Sgt2 protein and its interactions with chaperones and the Get4/Get5 complex. J Biol Chem. 2011;286:34325–34334.
23. Chartron JW, VanderVelde DG, Clemons WM Jr. Structures of the Sgt2/SGTA dimerization domain with the Get5/UBL4A UBL domain reveal an interaction that forms a conserved dynamic interface. Cell Rep. 2012;2:1620–1632.
24. Rao M, Okreglak V, Chio US, Cho H, Walter P, Shan S-O. Multiple selection filters ensure accurate tail-anchored membrane protein targeting. Elife. 2016;5:e21301–e21324.

25. Simon AC, Simpson PJ, Goldstone RM, et al. Structure of the Sgt2/Get5 complex provides insights into GET-mediated targeting of tail-anchored membrane proteins. Proc Natl Acad Sci U S A. 2013;110:1327–1332.

26. Hessa T, Sharma A, Mariappan M, Eshleman HD, Gutierrez E, Hegde RS. Protein targeting and degradation are coupled for elimination of mislocalized proteins. Nature. 2011;475:394–397.

27. Xu Y, Cai M, Yang Y, Huang L, Ye Y. SGTA recognizes a non-canonical ubiquitin-like domain in the Bag6-Ubl4A-Trc35 complex to promote endoplasmic reticulum-associated degradation. Cell Rep. 2012;2:1633–1644.

28. Wunderley L, Leznicki P, Payapilly A, High S. SGTA regulates the cytosolic quality control of hydrophobic substrates. J Cell Sci. 2014;127:4728–4739.

29. Leznicki P, Korac-Prlic J, Kliza K, et al. Binding of SGTA to Rpn13 selectively modulates protein quality control. J Cell Sci. 2015;128:3187–3196.

30. Mock J-Y, Chartron JW, Zaslaver M, Xu Y, Ye Y, Clemons WM Jr. Bag6 complex contains a minimal tail-anchor–targeting module and a mock BAG domain. Proc Natl Acad Sci U S A. 2015;112:106–111.

31. Rodrigo-Brenni MC, Gutierrez E, Hegde RS. Cytosolic quality control of mislocalized proteins requires RNF126 recruitment to Bag6. Mol Cell. 2014;55:227–237.

32. Dupzyk A, Williams JM, Bagchi P, Inoue T, Tsai B. SGTA-dependent regulation of Hsc70 promotes cytosol entry of simian virus 40 from the endoplasmic reticulum. J Virol. 2017;91:23.

33. Kiktev DA, Patterson JC, Müller S, Bariar B, Pan T, Chernoff YO. Regulation of chaperone effects on a yeast prion by cochaperone Sgt2. Mol Cell Biol. 2012;32:4960–4970.

34. Long P, Samnakay P, Jenner P, Rose S. 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. 2012;36:2733–2742.

35. Trotta AP, Need EF, Selth LA, et al. 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. 2013;133:2812–2823.

36. Buchanan G, Ricciardelli C, Harris JM, et al. Control of androgen receptor signaling in prostate cancer by the cochaperone small glutamine rich tetratricopeptide repeat containing protein. Cancer Res. 2007;67:10087–10096.

37. Wang H, Zhang Q, Zhu D. hSGT interacts with the N-terminal region of myostatin. Biochem Biophys Res Commun. 2003;311:877–883.

38. Chou M-L, Fitzpatrick LM, Tu S-L, et al. Tic40, a membrane-anchored co-chaperone homolog in the chloroplast protein translocon. EMBO J. 2003;22:2970–2980.

39. Bédard J, Kubis S, Bimanadham S, Jarvis P. Functional similarity between the chloroplast translocon component, Tic40, and the human co-chaperone, Hsp70-interacting protein (Hip). J Biol Chem. 2007;282:21404–21414.

40. Kovacheva S, Bédard J, Patel R, et al. In vivo studies on the roles of Tic110, Tic40 and Hsp93 during chloroplast protein import. Plant J. 2004;41:412–428.

41. Ko K, Banerjee S, Innes J, Taylor D, Ko Z. The Tic40 translocon components exhibit preferential interactions with

42. Chang HC, Nathan DF, Lindquist S. In vivo analysis of the Hsp90 cochaperone Sti1 (p60). Mol Cell Biol. 1997;17:318–325.

43. Fan G-H, Yang W, Sai J, Richmond A. Hsc/Hsp70 interacting protein (Hip) associates with CXCR2 and regulates the receptor signaling and trafficking. J Biol Chem. 2002;277:6590–6597.

44. Chou M-L, Chu C-C, Chen L-J, Akita M, Li H-M. Stimulation of transit-peptide release and ATP hydrolysis by a cochaperone during protein import into chloroplasts. J Cell Biol. 2006;175:893–900.

45. Watkins JF, Sung P, Prakash L, Prakash S. The Saccharomyces cerevisiae DNA repair gene RAD23 encodes a nuclear protein containing a ubiquitin-like domain required for biological function. Mol Cell Biol. 1993;13:7757–7765.

46. Wade SL, Auble DT. The Rad23 ubiquitin receptor, the proteasome and functional specificity in transcriptional control. Transcription. 2014;1:22–26.

47. Fishbain S, Prakash S, Herrig A, Elsasser S, Matouschek A. Rad23 escapes degradation because it lacks a proteasome initiation region. Nat Commun. 2011;2:192.

48. Elsasser S, Gali RR, Schwickart M, et al. Proteasome subunit Rpn1 binds ubiquitin-like protein domains. Nat Cell Biol. 2002;4:725–730.

49. Saeki Y, Sone T, Toh-e A, Yokosawa H. Identification of ubiquitin-like protein-binding subunits of the 26S proteasome. Biochem Biophys Res Commun. 2002;296:813–819.

50. Dantuma NP, Heinen C, Hoogstraten D. The ubiquitin receptor Rad23: At the crossroads of nucleotide excision repair and proteasomal degradation. DNA Repair. 2009;8:449–460.

51. Lowe ED, Hasan N, Trempe JF, et al. Structures of the Dsk2 UBL and UBA domains and their complex. Acta Cryst D. 2006;62:177–188.

52. Hara T, Kamura T, Kotoshiba S, et al. Role of the UBL-UBA protein KPC2 in degradation of p27 at G1 phase of the cell cycle. Mol Cell Biol. 2005;25:9292–9303.

53. Kotoshiba S, Kamura T, Hara T, Ishida N, Nakayama KI. Molecular dissection of the interaction between p27 and Kip1 ubiquitylation-promoting complex, the ubiquitin ligase that regulates proteolysis of p27 in G1 phase. J Biol Chem. 2005;280:17694–17700.

54. Conklin D, Holderman S, Whitmore TE, Maurer M, Feldhaus AL. Molecular cloning, chromosome mapping and characterization of UBQLN3 a testis-specific gene that contains an ubiquitin-like domain. Gene. 2000;249:91–98.

55. Yuan S, Qin W, Riordan CR, McSwiggin H, Zheng H, Yan W. Ubqln3, a testis-specific gene, is dispensable for embryonic development and spermatogenesis in mice. Mol Reprod Dev. 2015;82:266–267.

56. Zhang D, Raasi S, Fushman D. Affinity makes the difference: Nonselective interaction of the UBA domain of ubiquilin-1 with monomeric ubiquitin and polyubiquitin chains. J Mol Biol. 2008;377:162–180.

57. Seok Ko H, Uehara T, Tsuruma K, Nomura Y. Ubiquilin interacts with ubiquitylated proteins and proteasome through its ubiquitin-associated and ubiquitin-like domains. FEBS Lett. 2004;566:110–114.

different forms of the Oee1 plastid protein precursor. Funct Plant Biol. 2004;31:285–294.

58. Ayadi El A, Stieren ES, Barral JM, Boehning D. Ubiquilin-1 and protein quality control in Alzheimer disease. Prion. 2013; 7:164–169.

59. Lim PJ, Danner R, Liang J, et al. Ubiquilin and p97/VCP bind erasin, forming a complex involved in ERAD. J Cell Biol. 2009;187:201–217.

60. Itakura E, Zavodszky E, Shao S, Wohlever ML, Keenan RJ, Hegde RS. Ubiquilins chaperone and triage mitochondrial membrane proteins for degradation. Mol Cell. 2016;63:21–33.

61. Ward N, Moreno-Hagelsieb G. Quickly finding orthologs as reciprocal best hits with BLAT, LAST, and UBLAST: How much do we miss? PLoS ONE. 2014;9:e101850–e101856.

62. El-Gebali S, Mistry J, Bateman A, et al. The Pfam protein families database in 2019. Nucleic Acids Res. 2018;47:D427–D432.

63. Shindyalov IN, Bourne PE. Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. Protein Eng. 1998;11:739–747.

64. Pei J, Kim B-H, Grishin NV. PROMALS3D: A tool for multiple protein sequence and structure alignments. Nucleic Acids Res. 2008;36:2295–2300.

65. Letunic I, Doerks T, Bork P. SMART: Recent updates, new developments and status in 2015. Nucleic Acids Res. 2015;43: D257–D260.

66. Trempe J-F, Šašková KG, Sivá M, et al. Structural studies of the yeast DNA damage-inducible protein Ddi1 reveal domain architecture of this eukaryotic protein family. Sci Rep. 2016;6: 33671.

67. Liu FH, Wu SJ, Hu SM, Hsiao CD, Wang C. Specific interaction of the 70-kDa heat shock cognate protein with the tetratricopeptide repeats. J Biol Chem. 1999;274:34425–34432.

68. Yang J, Anishchenko I, Park H, Peng Z, Ovchinnikov S, Baker D. Improved protein structure prediction using predicted interresidue orientations. Proc Natl Acad Sci U S A. 2020;117:1496–1503.

69. Deng H-X, Chen W, Hong S-T, et al. Mutations in UBQLN2 cause dominant X-linked juvenile and adult-onset ALS and ALS/dementia. Nature. 2011;2:192.

70. Şentürk M, Lin G, Zuo Z, et al. Ubiquilins regulate autophagic flux through mTOR signalling and lysosomal acidification. Nat Cell Biol. 2019;21:384–396.

71. Subudhi I, Shorter J. Ubiquilin 2: Shuttling clients out of phase? Mol Cell. 2018;69:919–921.

72. Marín I. The ubiquilin gene family: Evolutionary patterns and functional insights. BMC Evol Biol. 2014;14:63–22.

73. Röhl A, Wengler D, Madl T, et al. Hsp90 regulates the dynamics of its cochaperone Sti1 and the transfer of Hsp70 between modules. Nat Commun. 2015;6:6655.

74. Biebl MM, Buchner J. Structure, function, and regulation of the Hsp90 machinery. Cold Spring Harb Perspect Biol. 2019; 11:a034017–a034033.

75. Lott A, Oroz J, Zweckstetter M. Molecular basis of the interaction of Hsp90 with its co-chaperone Hop. Protein Sci. 2020;29: 2422–2432.

76. Kirschke E, Goswami D, Southworth D, Griffin PR, Agard DA. Glucocorticoid receptor function regulated by coordinated action of the Hsp90 and Hsp70 chaperone cycles. Cell. 2014;157:1685–1697.

77. Schopf FH, Biebl MM, Buchner J. The HSP90 chaperone machinery. Nat Rev Mol Cell Biol. 2017;18:345–360.

78. Röhl A, Rohrberg J, Buchner J. The chaperone Hsp90: Changing partners for demanding clients. Trends Biochem Sci. 2013; 38:253–262.

79. Bédard J, Trösch R, Wu F, et al. Suppressors of the chloroplast protein import mutant tic40 reveal a genetic link between protein import and thylakoid biogenesis. Plant Cell. 2017;29: 1726–1747.

80. Scheufler C, Brinker A, Bourenkov G, et al. Structure of TPR domain–peptide complexes. Cell. 2000;101:199–210.

81. Zeytuni N, Zarivach R. Structural and functional discussion of the tetra-trico-peptide repeat, a protein interaction module. Cell Struct Funct. 2012;20:397–405.

82. Cho H, Shan S-O. Substrate relay in an Hsp70-cochaperone cascade safeguards tail-anchored membrane protein targeting. EMBO J. 2018;37:e99264.

83. Coto ALS, Seraphim TV, Batista FAH, et al. Structural and functional studies of the *Leishmania braziliensis* SGT cochaperone indicate that it shares structural features with HIP and can interact with both Hsp90 and Hsp70 with similar affinities. Int J Biol Macromol. 2018;118:693–706.

84. Aurora R, Rose GD. Helix capping. Protein Sci. 1998;7:21–38.

85. Burri L, Lithgow T. A complete set of SNAREs in yeast. Traffic. 2003;5:45–52.

86. Bernstein HD, Poritz MA, Strub K, Hoben PJ, Brenner S, Walter P. Model for signal sequence recognition from amino-acid sequence of 54K subunit of signal recognition particle. Nature. 1989;340:482–486.

87. Zopf D, Bernstein HD, Johnson AE, Walter P. The methionine-rich domain of the 54 kd protein subunit of the signal recognition particle contains an RNA binding site and can be crosslinked to a signal sequence. EMBO J. 1990;9: 4511–4517.

88. Mateja A, Szlachcic A, Downing ME, et al. The structural basis of tail-anchored membrane protein recognition by Get3. Nature. 2009;461:361–366.

89. Suloway CJM, Chartron JW, Zaslaver M, Clemons WM. Model for eukaryotic tail-anchored protein binding based on the structure of Get3. Proc Natl Acad Sci U S A. 2009;106: 14849–14854.

90. Guna A, Hegde RS. Transmembrane domain recognition during membrane protein biogenesis and quality control. Curr Biol. 2018;28:R498–R511.

91. Leznicki P, High S. SGTA antagonizes BAG6-mediated protein triage. Proc Natl Acad Sci U S A. 2012;109:19214–19219.

92. Shao S, Rodrigo-Brenni MC, Kivlen MH, Hegde RS. Mechanistic basis for a molecular triage reaction. Science. 2017;355: 298–302.

93. Yang J, Anishchenko I, Park H, Peng Z, Ovchinnikov S, Baker D. Improved protein structure prediction using predicted inter-residue orientations. bioRxiv. 2019;13: e1005324–e1005324.

94. Saladi SM, Maggiolo AO, Radford K, Clemons WM Jr. Structural biologists, let's mind our colors. bioRxiv. 2020;27: 14–16.

95. Gautier R, Douguet D, Antonny B, Drin G. HELIQUEST: A web server to screen sequences with specific α-helical properties. Bioinformatics. 2008;24:2101–2102.

96. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2—A multiple sequence alignment

editor and analysis workbench. Bioinformatics. 2009;25: 1189–1191.

97. Kabsch W, Sander C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. Biopolymers. 1983;22:2577–2637.

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

99. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol Biol Evol. 2013;30:772–780.

100. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The CLUSTAL_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 1997;25:4876–4882.

101. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. J Mol Biol. 1982;157: 105–132.

102. Bateman A, Martin MJ, O'Donovan C, et al. UniProt: The universal protein knowledgebase. Nucleic Acids Res. 2017;45: D158–D169.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.