THE STRUCTURE OF A TRANSMEMBRANE PROTEIN SORTING COMPLEX

Thesis by

Justin William Chartron

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

2013

Defended September 20, 2012

Acknowledgements

My career as a scientist, and in particular a structural biologist, began with a high school internship at The Scripps Research Institute. Professor C. David Stout went far beyond what was required and allowed me to work in the lab for five years, while I was an undergraduate at UCSD. He personally taught me X-ray crystallography, and I thank him immensely. Other lab members, Dr. Vidyasanker Sundaresan, Professor Laura Hunsicker-Wang, and Dr. Holly Heaslet also deserve thanks for their advice through the years. I thank Professors Carolyn Bertozzi and Kate Carroll for trusting an undergraduate to solve structures of their proteins.

At Caltech I found another supportive, yet challenging, environment. I thank my committee members Professors Shu-ou Shan, Douglas Rees and Jared Leadbetter, as well as Professors André Hoelz and Thomas Miller, for providing guidance on my research and encouragement when the pressures of competition bore down. I am grateful for the opportunity to work with my advisor, Professor William Clemons. I believe our personalities brought different strengths together that resulted in six productive, insightful years. He was patient with me as I stumbled my way through techniques that were new to both of us, yet provided a pressure necessary to maintain focus. If I am a more astute scientist after graduate school, it is largely due to Bil.

For their aid in my research and instruction I thank members of the Clemons lab, beginning with my partners in the GET pathway, Dr. Christian Suloway, Harry Gristick, and Jee-Young Mock. I also thank the members of the lab who were present while most of my work was accomplished, Dr. Axel Müller, Dr. Suresh Ramasamy, Dr. Shiho

Tanaka, Dr. Kyoung-Soon Jang, Matt Thornton and Ma'ayan Zaslaver. I also am grateful for the help of the undergraduate students I have mentored, Tom Lampo, Grecia Gonzalez, Carolyn Nguyen, Katie Brugman, and Doreen Chan.

When we began our NMR experiments there was a dearth of practical knowledge required for protein structure determination. Dr. David VanderVelde provided a deep, invaluable understanding of NMR upon which my work was built on. I thank him for the opportunities his instruction has opened.

Other members of the Caltech community deserve thanks: Dr. Troy Walton, Dr. Chris Ghandi, Dr. Jens Kaiser, Dr. Julie Hoy, Dr. Jost Vielmetter, Pavle Nikolovski, Mike Rome, Meera Rao, and Sowmya Chandrasaker.

I've made a lot of friends in graduate school to whom I owe tremendous gratitude, but two deserve recognition for providing driving forces in my research. I thank Dr. Ian Tonks, who most closely shared my scientific passions and goals, and Taylor Lenton, whose emotional support always accentuated the high points and softened the low ones.

Finally I thank my parents, Brenda and Fred Chartron, to whom I dedicate this thesis. Through their love, I never felt a single limit to what I could accomplish, and I never felt a single fear of disappointing them. Only in hindsight do I understand how special it was that providing me the opportunity to achieve my goals was their own most paramount. Although my father was only able to see the beginning of my career 11 years ago, he made sure those around him knew he was proud.

Abstract

The biogenesis of membrane proteins is an essential process in biology. It requires the protection of hydrophobic transmembrane domains from aggregation in the cytosol as well as targeting to the proper membrane. Tail-anchored (TA) proteins have a single transmembrane helix near their carboxyl termini and require a post-translational mechanism for targeting and insertion. In yeast, the Guided Entry of Tail-anchored proteins (GET) pathway delivers TA proteins to the endoplasmic reticulum (ER). A sorting complex comprising Get4, Get5, and Sgt2 load ER destined TA proteins onto the targeting factor Get3. X-ray crystallography, solution NMR, and small angle X-ray scattering were used to characterize this assembly. Get4 and Get5 form an extended adapter complex. Get4 maintains Get3 in a state competent to receive TA proteins. The N-terminus of Get5 tightly binds Get4, while the C-terminus of Get5 is a homodimerization domain, resulting in a heterotetrameric assembly. A ubiquitin-like domain within Get5 binds the heat-shock protein (HSP) co-chaperone Sgt2, providing a physical link between ER destined TA protein targeting and protein folding pathways. Sgt2 is also an extended homodimeric complex, and can directly bind four major classes of HSPs. The Get4/Get5/Sgt2 sorting complex is multivalent, flexible and the binding of individual components is transient. These results build a model for post-translational protein targeting in eukaryotes that is distinct from other pathways.

Table of Contents

Chapter 4

List of Figures and Tables

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Appendix

Introduction

The right place at the right time

The living cell is complex enough that its internal workings effectively mirror human civilization. The cell needs to acquire scarce, raw materials from a harsh environment, transform them into useful building blocks, direct these products to specific destinations, protect itself from external and internal threats and deal with the waste that is generated during all of these events. This is only the surface, and the analogy can be carried much further. The issues we squabble over are the same issues the cell has been dealing with for billions of years. Admittedly, putting an appropriate perspective on the scope of biology requires some melodrama.

However, treating the cell in this regard overlooks the remarkable fact that these struggles exist at all levels of organization in biology, from within individual cells to populations of cells, from tissues to organs to organisms, to communities, ecosystems, political states and nations. The reason the cell presents a unique area of study, though, is while higher levels can be explained by breaking them into their smaller biological constituents, the cell is a collection of chemicals guided by the most basic thermodynamic principles. It is the foundation, a battleground in understanding how the fundamental physical properties of the universe dictate life. Arguably the goal of biochemistry and molecular biophysics is to understand how actions occurring at the level of individual atoms and chemical bonds translate into biological responses. Aside from satisfying curiosity, this knowledge often has practical value in medicine and technology.

If the cell only obeys physical laws, what distinguishes the chemical reactions occurring within it from chemistry occurring outside of it? The most direct answer is control, and this can be further elaborated on. One aspect of control is temporal. The cell ultimately decides what chemical reactions can occur at any given time through its genetic programming. RNA and proteins are the encoded tools that perform these programs. The reactions that do occur, whether driven by environmental cues such as resource availability or physical stress, in turn dictate what genetic program to follow through transcriptional regulation. Coupled with intricate systems of protein modification and degradation, the cell can rapidly switch between the reactions it can perform.

A second form of control is spatial. Certain sets of reactions require or produce compounds that are otherwise incompatible with the materials of other sets. In the violent example of phagocytosis, the indiscriminate proteases required to digest a consumed cell must be separated from the rest of the cellular environment of the predator. The cell satisfies these requirements through compartmentalization. It can be as deceptively simple as providing a membrane separating the inside the cell and the space immediately outside of it, as in bacteria. Or as in the case of eukaryotes, it consists of a suite of membrane-enclosed substructures called organelles, each performing drastically different chemistry. Membranes themselves additionally provide level of spatial control by restricting some chemistry to a two-dimensional surface.

Two requirements arise from this simplistic model of biochemistry. Biomolecules must be synthesized with high fidelity at the proper time, and they must be directed to the proper location. For a major class of biomolecules, the proteins, the first process is

translation and folding, and the second process is known as targeting. For many eukaryotic proteins, the processes are streamlined by the secretory system. Proteins are synthesized in the cytosol by ribosomes, and then a single, generic marker directs them to the endoplasmic reticulum (ER). Once in the ER, additional specific markers further direct proteins to the various organelles through vesicular trafficking. The initial targeting event to the ER is performed concurrently with synthesis for the majority of proteins by the signal recognition particle pathway (reviewed in (Shan and Walter, 2005)).

A co-translational pathway targets the majority of secretory proteins

Proteins are synthesized by the ribosome reading a messenger RNA transcript. It is a linear process, with the nascent protein emerging into the environment one amino acid at a time in the encoded order. Relatively early in synthesis of secretory proteins, a stretch of residues called a signal peptide will become exposed (Blobel and Dobberstein, 1975). This sequence is recognized by the signal recognition particle (SRP), which binds to both the nascent protein and the surface of the translating ribosome (Walter and Blobel, 1981). The ribosome/SRP/nascent protein complex is captured by the ER-membrane associated SRP receptor (Gilmore et al., 1982). The ribosome and nascent protein are then transferred to a protein-conducting pore called the translocon (Simon and Blobel, 1991). As synthesis continues, the new protein passes through the translocon, unfolded, into the ER lumen. This process is called co-translational translocation. The system also can incorporate transmembrane proteins through a lateral gating mechanism, allowing the translocon to open into the plane of the membrane (Van den Berg et al., 2004). In this case, the first transmembrane helix can dually serve as a signal sequence and is

designated a signal anchor sequence. Once incorporated in the membrane, these proteins are trafficked through the same vesicular system as the soluble, luminal proteins.

Transmembrane proteins serve innumerous roles in the cell, ranging from structural and mechanical functions to enzymes to environmental sensors. Unlike soluble proteins, a transmembrane protein is exposed to three distinct environments: two aqueous environments on either side of the membrane, and the lipidic membrane itself. This is reflected in the diversity of structure observed. A transmembrane protein may consist of a single membrane-spanning helix, or it may thread back and forth across the membrane with multiple membrane-spanning helices. The positions of soluble termini of the protein chain may be in the same or opposite membrane face. There is apparently limitless variation on the lengths of soluble terminal domains and the loops connecting transmembrane helices.

The co-translational translocation pathway is remarkable because it can target and insert nearly all varieties of transmembrane proteins. There is one topologically restricted exception, however. If a single membrane-spanning helix is located near the carboxyl terminus of the nascent protein, the signal anchor sequence cannot bind the SRP because it sequestered within the ribosomal exit tunnel (Kutay et al., 1993). These proteins are called tail-anchored (TA) proteins and require post-translational targeting.

The biogenesis of TA proteins utilizes a distinct pathway

Post-translational targeting itself is not a rare process in biology. In the bacterial version of the secretory system, most secreted soluble proteins cross the cell membrane post-translationally (reviewed in (Driessen and Nouwen, 2008)). A cytosolic motor protein

called SecA binds to the signal sequence, and using energy generated from ATP hydrolysis, pushes unfolded protein through the translocon. Although the SecA system is absent in eukaryotes, mitochondrial and chloroplast proteins are targeted through other post-translational systems.

A post-translational targeting system specific to the ER that could accommodate TA proteins, however, proved more elusive to identify. Early investigation on synaptobrevin, a model TA protein involved in vesicle fusion (a SNARE protein), targeting to the ER required ATP hydrolysis, as well as unidentified ER membrane proteins distinct than the SRP/translocon machinery (Kutay et al., 1995). Some experimental evidence suggested that the SRP (Abell et al., 2004) or the chaperones Hsp40 and Hsp70 (Abell et al., 2007) can deliver TA proteins to the ER, although it is unclear how membrane insertion is performed by these routes, and how widely utilized they are as the data conflict with the earlier report (Kutay et al., 1995).

A breakthrough in our understanding of TA proteins occurred in 2007. Using sucrose density gradient fractionation and chemical crosslinking of *in vitro* translation systems, a 40 kDa ATPase was discovered to be the major protein associated with TA proteins in the soluble fraction of rabbit reticulocyte lysate (Favaloro et al., 2008; Stefanovic and Hegde, 2007). Designated TRC40, for Transmembrane domain Recognition Complex, this protein satisfies the postulated requirements for a targeting factor, including reversible binding to the ER membrane, and it can insert TA proteins when allowed to hydrolyze ATP. Prior to this discovery, the yeast homolog of TRC40, named Get3, was genetically linked to two membrane proteins, Get1 and Get2, and together appeared

significant for the recovery of ER resident proteins from the Golgi (Schuldiner et al., 2005). After the characterization of TRC40, Get3 proved to perform an identical role in yeast, with Get1 and Get2 forming the proteinaceous ER components required for insertion (Schuldiner et al., 2008). Knockouts of Get1 or Get2 lead to cytosolic aggregates of TA protein and Get3; a Get3 knockout leads to diffuse cytosolic localization of TA proteins. In any of these knockouts, ER TA proteins begin to appear in the mitochondria.

Get1, Get2 and Get3 were reassigned as TA protein targeting machinery, and the GET pathway was renamed for Guided Entry of Tail-anchored proteins (a full model of the pathway is presented in appendix Figure A.1). Shortly after, a genetic investigation of the unfolded protein response in the ER uncovered two more GET proteins (Jonikas et al., 2009). Little was known about Get4, other than that it is a conserved eukaryotic protein (Fernandes et al., 2008). More information was available for Get5 (also annotated as Mdy2). Mdy2 deletion leads to mating deficiency in yeast, with reduction in shmoo formation (Hu et al., 2006). Get5 was also known to interact with Sgt2, a heat shock protein (HSP) co-chaperone (Liou et al., 2007). There was some data suggesting Get5 also associates with the ribosome; however this interaction was destabilized by relatively low salt concentration (Fleischer et al., 2006). In our own investigations, we found no convincing evidence that Get4 or Get5 are ribosome-associated proteins. Regardless, Get4 and Get5 were postulated as acting upstream of Get3. Confirmation of a role for Sgt2 in TA protein targeting came through further genetic analysis, where it was shown that deletion of the gene resulted in a similar phenotype to other GET pathway members (Costanzo et al., 2010). A quantitative analysis of existing double-knockout phenotype

data proposed a linear "activity pathway" of Sgt2 followed by Get5, Get4 and finally Get3 (Battle et al., 2010).

By this point in time several structures of Get4 were available, including one from our laboratory (Bozkurt et al., 2010; Chang et al., 2010; Chartron et al., 2010). We characterized the Get4/Get5 complex in solution and showed that Get4 could bind Get3. These data led us to propose that Get4 somehow regulated Get3 prior to TA protein recognition, and Sgt2 and its associated chaperones either aided or branched to an alternate pathway. Concurrent to our structural work, Vladimir Denic's group provided direct experimental evidence for functions for Sgt2, Get4 and Get5 in TA protein targeting (Wang et al., 2010). Both Get4 and Get5 are required for efficient loading of TA proteins to Get3. Moreover, the C-terminal domain of Sgt2 is shown to directly bind the transmembrane domain of TA proteins. Amazingly, Sgt2 only seems to bind ER destined TA proteins as opposed to mitochondrial TA proteins, which can be bound to HSPs. ER proteins typically have shorter, more hydrophobic transmembrane domains and a mechanism for this discrimination remains an exciting question to be answered. From all this data, it appears that Get4 and Get5 form an adaptor complex that links Sgt2 and Get3, a broad HSP co-chaperone to a dedicated ER targeting factor.

Meanwhile, the story became more complicated in the mammalian system. A complex acting upstream of TRC40 was discovered, and it does contain homologs of Get4 and Get5, named TRC35 and Ubl4A respectively (Leznicki et al., 2010; Mariappan et al., 2010). However, it additionally contains a large, ~1100 amino acid protein named Bag-6 (alternatively named Bat-3 or Scythe). This complex has an additional role in selecting

protein substrates for ubiquitin-mediated degradation. These substrates may either be misfolded or mislocalized proteins as well as ER associated degradation (ERAD) retrotranslocated substrates (Hessa et al., 2011; Minami et al., 2010; Wang et al., 2011). SGTA, the human homolog of Sgt2, appears to have a direct role in handling substrates for both pathways and it will be exciting to determine the full interplay between targeting and quality control (Xu et al., 2012). Differences between fungal and mammalian homologs will be discussed throughout.

The GET pathway has proven fertile for structural biologists. Its attractiveness stemmed from the novel and important functions it performed, and that many components had no apparent homology to known proteins structures. It was also practical that many pathway members were soluble proteins. Christian Suloway spearheaded the efforts within our laboratory, determining the X-ray crystal structure of Get3 from *S. cerevisiae* and the filamentus fungus *Aspergillus fumigatus* (Suloway et al., 2009). I became involved by aiding the refinement and analysis of these structures. Within five months, four other reports of the Get3 structure were published (Bozkurt et al., 2009; Hu et al., 2009; Mateja et al., 2009; Yamagata et al., 2010). The structure and function of Get3 are the topics of Christian's thesis. A summary of Get3, models for its function and phylogenetic insights are provided in the appendix. The work presented as chapters here focuses on the complex composed by Get4, Get5, Sgt2 and various cellular chaperones.

Chapter 1 describes the initial structural analysis of the Get4 and Get5 proteins. A crystal structure of the majority of Get4 and the N-terminal domain of Get5 is presented. We identified the C-terminal domain of Get5 as a homodimerization domain and provided the

first characterization of the binding interface between Get4 and Get3. The Get5 homodimerization domain becomes the topic of chapter 2, where we determined a crystal structure and a solution NMR structure of the domain as well as a NMR structure of a homolog. Chapter 3 shifts focus to Sgt2, where a crystal structure of the HSP-binding TPR domain is presented from a fungal homolog. The structure gives insight into the relatively promiscuous binding of the domain. We further characterize Sgt2 using small-angle X-ray scattering. Chapter 4 ties the earlier chapters together and investigates the interaction between Sgt2 and the Get4/Get5 complex. Solution NMR and crystal structures are presented for the homodimerization domain of Sgt2 and the human homolog SGTA, as well as for the ubiquitin-like domain of Get5. NMR is further used to determine the structure of the complex between Sgt2 and Get5. Chapter 4 closes this thesis with a functional model of a transmembrane protein sorting complex.

Chapter 1

STRUCTURAL CHARACTERIZATION OF THE GET4/GET5 COMPLEX AND ITS INTERACTION WITH GET3

Abstract

The recently elucidated Get proteins are responsible for the targeted delivery of the majority of tail-anchored (TA) proteins to the endoplasmic reticulum. Get4 and Get5 have been identified in the early steps of the pathway mediating TA substrate delivery to the cytoplasmic targeting factor Get3. Here we report a crystal structure of Get4 and an N-terminal fragment of Get5 from *Saccharomyces cerevisae*. We show Get4 and Get5 (Get4/5) form an intimate complex that exists as a dimer (two copies of Get4/5) mediated by the C-terminus of Get5. We further demonstrate that Get3 specifically binds to a conserved surface on Get4 in a nucleotide dependent manner. This work provides further evidence for a model in which Get4/5 operates upstream of Get3 and mediates the specific delivery of a TA substrate.

Adapted from

Chartron, J.W., Suloway, C.J., Zaslaver, M., and Clemons, W.M., Jr. (2010). Structural characterization of the Get4/Get5 complex and its interaction with Get3. Proc. Natl. Acad. Sci. USA *107*, 12127-12132.

Introduction

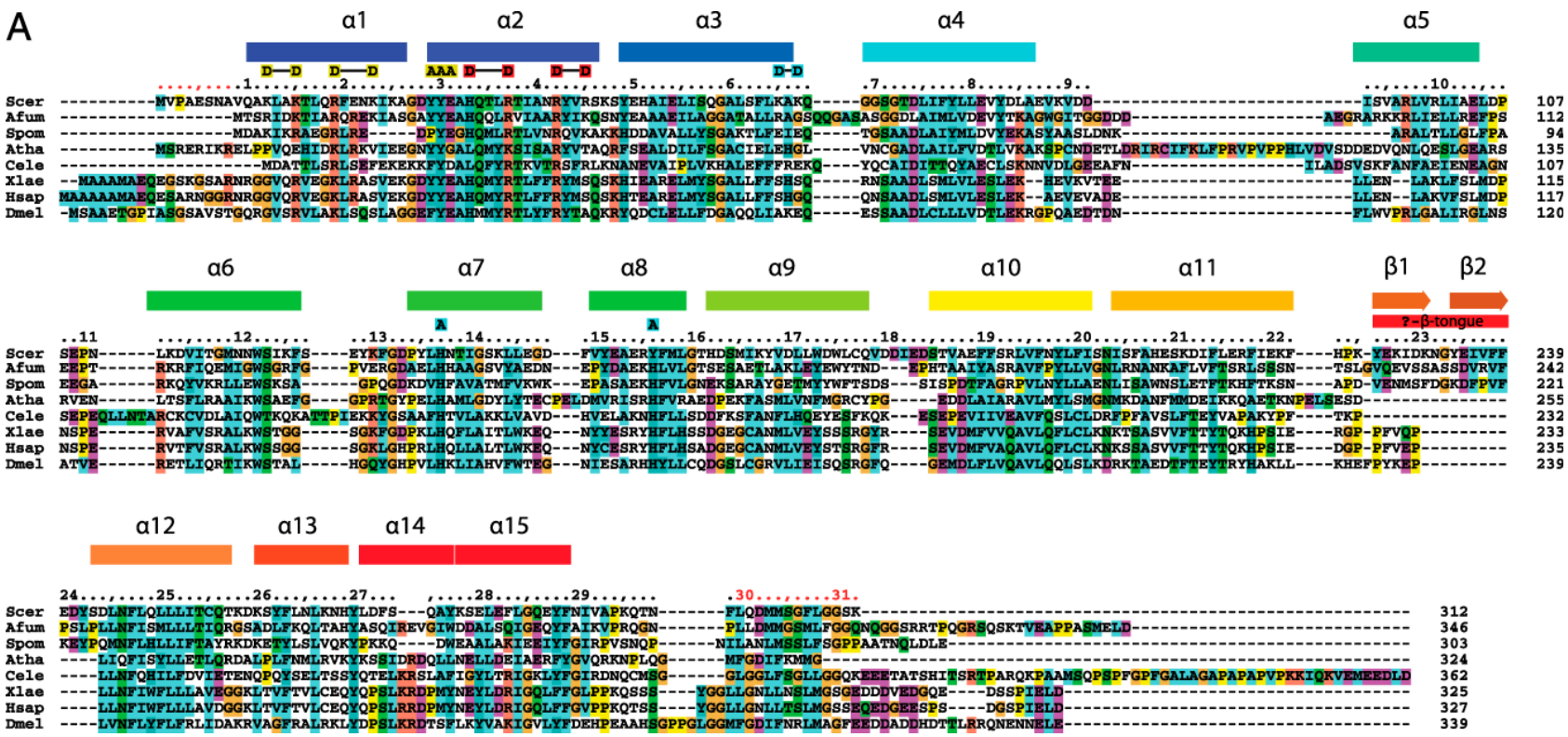
Targeted delivery of membrane proteins is a critical process. For a special class of membrane proteins, tail-anchored (TA) proteins, the targeting pathways have only recently begun to be understood. These proteins are a large and diverse class of integral membrane proteins found in all organisms. Examples include SNAREs, apoptosis factors, and protein translocation components. TA proteins are characterized by having a single transmembrane helix (TM) at their extreme C-terminus. Due to this topological constraint, these proteins are not able to follow the signal recognition particle (SRP) dependent cotranslational pathway that typifies most integral membrane proteins. Instead, these proteins must find their correct membrane for insertion post-translationally (Borgese et al., 2007; Kutay et al., 1995; Kutay et al., 1993; Rabu et al., 2009).

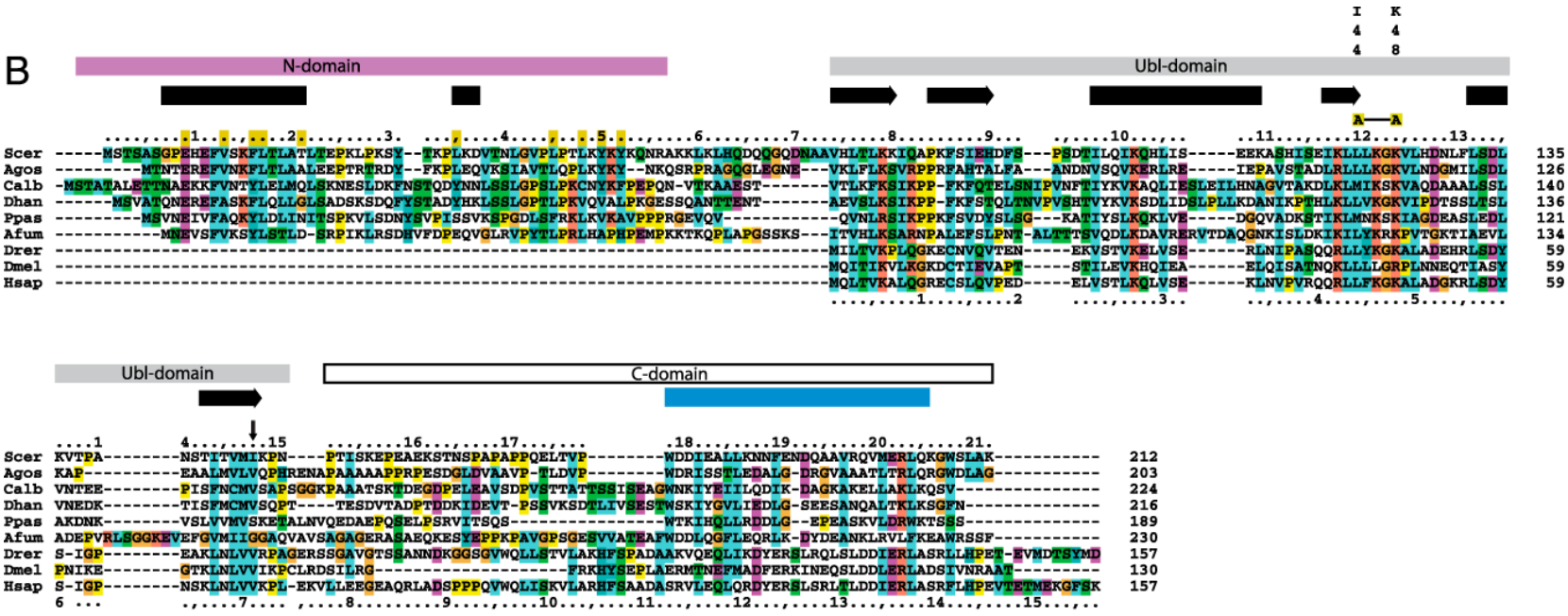
The newly characterized GET pathway (Guided Entry of Tail-anchored proteins) is the major targeting pathway for TA proteins in yeast. The first protein identified to specifically recognize a TA protein substrate is the ATPase Get3, which protects TA proteins in the cytoplasm and targets them to the endoplasmic reticulum (Favaloro et al., 2008; Jonikas et al., 2009; Schuldiner et al., 2005; Schuldiner et al., 2008; Stefanovic and Hegde, 2007). Deletions of this protein lead to mistargeting of TA proteins and growth sensitivity in a variety of conditions (Hillenmeyer et al., 2008; Jonikas et al., 2009). Recently, a number of structural studies of Get3 have led to the model where Get3 undergoes a dramatic conformational change upon nucleotide binding shifting from an open to a closed form and generating a TM binding pocket (Bozkurt et al., 2009; Hu et al., 2009; Mateja et al., 2009; Suloway et al., 2009; Yamagata et al., 2010). Despite the

many structures, the precise mechanism of how Get3 binds and releases substrate is not fully understood. At the ER, the Get3 TA protein complex binds two integral membrane proteins, Get1 and Get2, that are thought to act as receptors for the release of the protein substrate (Schuldiner et al., 2008). Upstream of Get3 are two proteins, Get4 and Get5 that are the subjects of this study.

Get4 [yeast locus Yor164c, human locus C7orf20 and *cee* in fish (Fernandes et al., 2008)] is a highly conserved protein that is estimated to have arisen early in evolution (Figure 1.1A) (Fernandes et al., 2008). Its high homology, 26% identity from yeast to humans, belies the fact that until recently very little was known about its biological role. It contains no known motifs and has only been annotated based on a series of genome-wide screens. Get4 localizes to the cytoplasm (Huh et al., 2003) and, although not essential, knockouts in yeast lead to sensitivity in a number of growth conditions (Giaever et al., 2002) whereas disruption of the homologue in *Caenorhabditis elegans* retards growth (Kamath et al., 2003; Simmer et al., 2003). Multiple protein interaction studies in yeast have implicated Get4 in binding to Get5 and Get3 and associating with Sgt2 and Hsp90-like proteins (Costanzo et al., 2010; Fleischer et al., 2006; Ito et al., 2001; Krogan et al., 2006; McClellan et al., 2007).

Get5/Mdy2 [yeast locus Yol111c, known as *GdX*/Ubl4a in mammals (Hu et al., 2006)] is a multidomain protein (Figure 1.1B). The N-terminal domain (Get5-N) is found only in fungi where it is conserved. Following that is a ubiquitin-like domain (Get5-Ubl) (Toniolo et al., 1988) and a C-terminal domain (Get5-C). Get5 was originally annotated based on a decreased mating phenotype (Hu et al., 2006; Iwanejko et al., 1999). Unlike


A
α1
α2
α3
α4
α5
α6
α7
α8
α9
α10
α11
β1
β2
?-β-tongue
α12
α13
α14
α15
Scer
Afum
Spom
Atha
Cele
Xlae
Hsap
Dmel


B
N-domain
Ubl-domain
I 4 4
K 4 8
A—A
Scer -----MSTSASGPEHEFVSKFLTLATLTEPKLPKSY--TKPLKDVTNLGVPLPTLKYKYKQNRAKKLKLHQDQGQDNAAVHLTLKKIQAPKFSIEHDFS---PSDTILQIKQHLIS-----EEKASHISEIKLLLKGKVLHDNLFLSDL 135
Agos ---------MTNTEREFVNKFLTLAALEEPTRTRDY--FKPLEQVKSIAVTLQPLKYKY--NKQSRPRAGQGLEGNE---VKLFLKSVRPPRFAHTALFA---ANDNVSQVKERLRE-----IEPAVSTADLRLLLKGKVLNDGMILSDL 126
Calb -MSTATALETTNAEKKFVNTYLELMQLSKNESLDKFNSTQDYNNLSSLGPSLPKCNYKFPEPQN-VTKAAEST-------VTLKFKSIKPP-FKFQTELSNIPVNFTIYKVKAQLIESLEILHNAGVTAKDLKLMIKSKVAQDAAALSSL 140
Dhan ------MSVATQNEREFASKFLQLLGLSADSKSDQFYSTADYHKLSSLGPTLPKVQVALPKGESSQANTTENT-------AEVSLKSIKPP-FKFSTQLTNVPVSHTVYKVKSDLIDSLPLLKDANIKPTHLKLLVKGKVIPDTSSLTSL 136
Ppas ---------MSVNEIVFAQKYLDLINITSPKVLSDNYSVPISSVKSPGDLSFRKLKVKAVPPPRGEVQV-----------QVNLRSIKPPKFSVDYSLSG---KATIYSLKQKLVE-----DGQVADKSTIKLMNKSKIAGDEASLEDL 121
Afum ----------MNEVSFVKSYLSTLD-ERPIKLRSDHVFDPEQVGLRVPYTLPRLHAPHPEMPKKTKQPLAPGSSKS---ITVHLKSARNPALEFSLPNT-ALTTTSVQDLKDAVRERVTDAQGNKISLDKIKILYKRKPVTGKTIAEVL 134
Drer ----------MILTVKPLQGKECNVQVTEN-----EKVSTVKELVSE------RLNIPASQQRLLYKGKALADEHRLSDY 59
Dmel ----------MQITIKVLKGKDCTIEVAPT-----STILEVKHQIEA------ELQISATNQKLLLLGRPLNNEQTIASY 59
Hsap ----------MQLTVKALQGRECSLQVPED-----ELVSTLKQLVSE------KLNVPVRQQRLLFKGKALADGKRLSDY 59
...,....1...,....2 ...,....3.. ..,....4....,....5....,....
Ubl-domain
C-domain
....1 4...,...15 ...,....16...,....17...,... .18...,....19...,....20...,....21.
Scer KVTPA--------NSTITVMIKPN----PTISKEPEAEKSTNSPAPAPPQELTVP--------WDDIEALLKNNFENDQAAVRQVMERLQKGWSLAK---------- 212
Agos KAP----------EAALMVLVQPHRENAPAAAAAPPRPESDGLDVAAVP-TLDVP--------WDRISSTLEDALG-DRGVAAATLTRLQRGWDLAG---------- 203
Calb VNTEE--------PISFNCMVSAPSGGKPAAATSKTDEGDPELEAVSDPVSTTATTSSISEAGWNKIYEIILQDIK-DAGKAKELLAKLKQSV---------------- 224
Dhan VNEDK--------TISFMCMVSQPT----TESDVTADPTDDKIDEVT-PSSVKSDTLIVSESTWSKIYGVLIEDLG-SEESANQALTKLKSGFN---------------- 216
Ppas AKDNK--------VSLVVMVSKETALNVQEDAEPQSELPSRVITSQS----------------WTKIHQLLRDDLG--EPEASKVLDRWKTSSS---------------- 189
Afum ADEPVRLSGGKEVEFGVMIIGGAQVAVSAGAGERASAEQKESYEPPKPAVGPSGESVVATEAFWDDLQGFLEQRLK-DYDEANKLRVLFKEAWRSSF---------- 230
Drer S-IGP--------EAKLNLVVRPAGERSSGAVGTSSANNDKGGSGVWQLLSTVLAKHFSPADAAKVQEQLIKDYERSLRQLSLDDIERLASRLLHPET-EVMDTSYMD 157
Dmel PNIKE--------GTKLNLVVIKPCLRDSILRG----------------------FRKHYSEPLAERMTNEFMADFERKINEQSLDDLERLADSIVNRAAT--------- 130
Hsap S-IGP--------NSKLNLVVKPL-EKVLLEEGEAQRLADSPPPQVWQLISKVLARHFSAADASRVLEQLQRDYERSLSRLTLDDIERLASRFLHPEVTETMEKGFSK 157
6 ... .,....7.... ,....8.....9....,....10...,....11...,....12...,....13...,....14...,....15...,..

most yeast proteins that contain a Ubl, Get5 does not interact with polyubiquinated proteins nor does it bind the 26S proteasome (Saeki et al., 2002). A biochemical and genetic study linked Get5 to both Sgt2, a tetratricopeptide repeat (TPR) containing protein (D'Andrea and Regan, 2003), and to Ydj1, a J-domain containing Hsp40 homologue that interacts with the cytosolic Hsp70 homologues Ssa1p/Ssa2p (Liou et al., 2007).

Genomic screens suggest that Get4 and Get5 form a stable complex (referred to here as Get4/5). Using epistatic arrays and biochemistry, Jonikas et al. showed that these two proteins were also involved in the GET pathway (Jonikas et al., 2009). These proteins operated upstream of Get3 and, based on a study where Get5 was found bound the ribosome (Fleischer et al., 2006), it was suggested that these proteins acted as the

---

Figure 1.1. Sequence alignments of Get4 and Get5. (A) Sequences of Get4 were aligned using ClustalX (Larkin et al., 2007). Residue coloring is based on the program output (amino acid type). Species in order are Scer (*Saccharomyces cerevisiae*), Afum (*Aspergillus fumigatus*), Spom (*Schizosaccharomyces pombe*), Atha (*Arabidopsis thal*iana), Cele (*Caenorhabditis elegans*), Xlae (*Xenopus laevis*), Hsap (*Homo sapiens*), and Dmel (*Drosophila melanogaster*). Numbering above the sequences is based on *S. cerevisiae*, with residues observed in the crystal structure in black and residues absent in red. Mutants generated for this study are indicated above the numbering, with double mutants connected with bars, and are highlighted based on wild type (cyan), weak (yellow) or weakest (red) interactions with Get3. Secondary structure is shown on top with helices indicated by boxes and β-strands by arrows color ramped to match Figure 1B. (B) Alignment of Get5. Species in order are Scer (*Saccharomyces cerevisiae*), Agos (*Ashbya gossypii*), Calb (*Candida albicans*), Dhan (*Debaryomyces hansenii*), Ppas (*Pichia pastoris*), Afum (*Aspergillus fumigatus*), Drer (*Danio rerio*), Dmel (*Drosophila melanogaster*), and Hsap (*Homo sapiens*). Secondary structure in the N-domain is from the crystal structure, whereas secondary structure of the Ubl-domain is based on the NMR structure of human Ubl4A (PDB ID code 2DZI). Residues 179–205 are predicted to be helical and are colored cyan. The arrow above residue 148 indicates the start of the C-domain peptide observed during purications. I44 and K48 labeled vertically above the Ubl-domain designation refer to residues in ubiquitin specically addressed in the text. The L120A/K120A mutant is indicated in yellow representing a moderate growth defect.

ribosome receptor for Get3. Kar2p, a resident ER protein, is secreted in mutants defective in TA targeting. A screen for these mutants confirmed Get4 and Get5 as part of the GET pathway (Copic et al., 2009).

During the preparation of this manuscript the first structure of yeast Get4 and a fragment of Get5, generated from unintended proteolysis, was published (Chang et al., 2010). The authors used two-hybrid screens to explore general interactions of Get4/5 implicating Sgt2 and Ydj1 binding to Get5 and Get3 binding to the N-terminal half of Get4. A second structure of Get4 alone from *Chaetomium thermophilum*has also been published (Bozkurt et al., 2010).

Here we report an independent structure of Get4 with an N-terminal fragment of Get5 (Get4/5-N) in a unique crystal form. Using the structure as a guide, we show that the full-length Get4/5 complex exists as a dimer (two copies of Get4/5) and identify important functional residues and the binding interface with Get3. Our results further define the structural elements of Get4/5 and provide strong evidence for the current model that has Get4 and Get5 acting as upstream factors of Get3 in the Get targeting pathway.

Results

Purification and structure determination of Get4/5-N

Full-length Get4/5 was expressed in *Escherichia coli* and purified using affinity chromatography. We were able to express Get4 alone; however, all of the protein went into inclusion bodies in all tested expression conditions. Further purification of Get4/5 by anion exchange chromatography resulted in two separate peaks. These peaks were stable and were injected onto a size exclusion column where they both ran separately

and larger than expected based on molecular weight (see below). Both failed to crystallize.

To address the possibility that disorder might have prevented crystallization, we performed *in situ* proteolysis by including protease in the crystallization trials. Crystals grew quickly using chymotrypsin and contained nearly full-length Get4 and the N-terminal third of Get5 (Get4/5-N). To improve crystallization, we performed limited proteolysis (Figure 1.2A) and then purified Get4/5-N by ion exchange chromatography (Figure 1.2B,C). Initial crystals were hexameric rods with reproducible twists halfway down their length that did not diffract. We found that several additives containing amines generated trigonal crystals. The final crystals grew using L-proline as an additive and diffracted to 2.8 Å (Figure 1.2D).

The structure was solved using seleno-methionine, single wavelength anomalous dispersion (SAD) phasing and 3-fold noncrystallographic symmetry. The final structure contained three almost identical Get4/5-N in the asymmetric unit, main-chain r.m.s.d. of approximately 0.6 Å (Figure 1.3C). The most complete Get4/5-N model contains nearly all of Get4, residues 9-299, and the N-terminus of Get5, residues 3–56 (Figure 1.3A). The structure refined to an R_{factor} of 18.2% and an R_{free} of 22.4%. Crystallographic statistics are presented in Table 1.1.

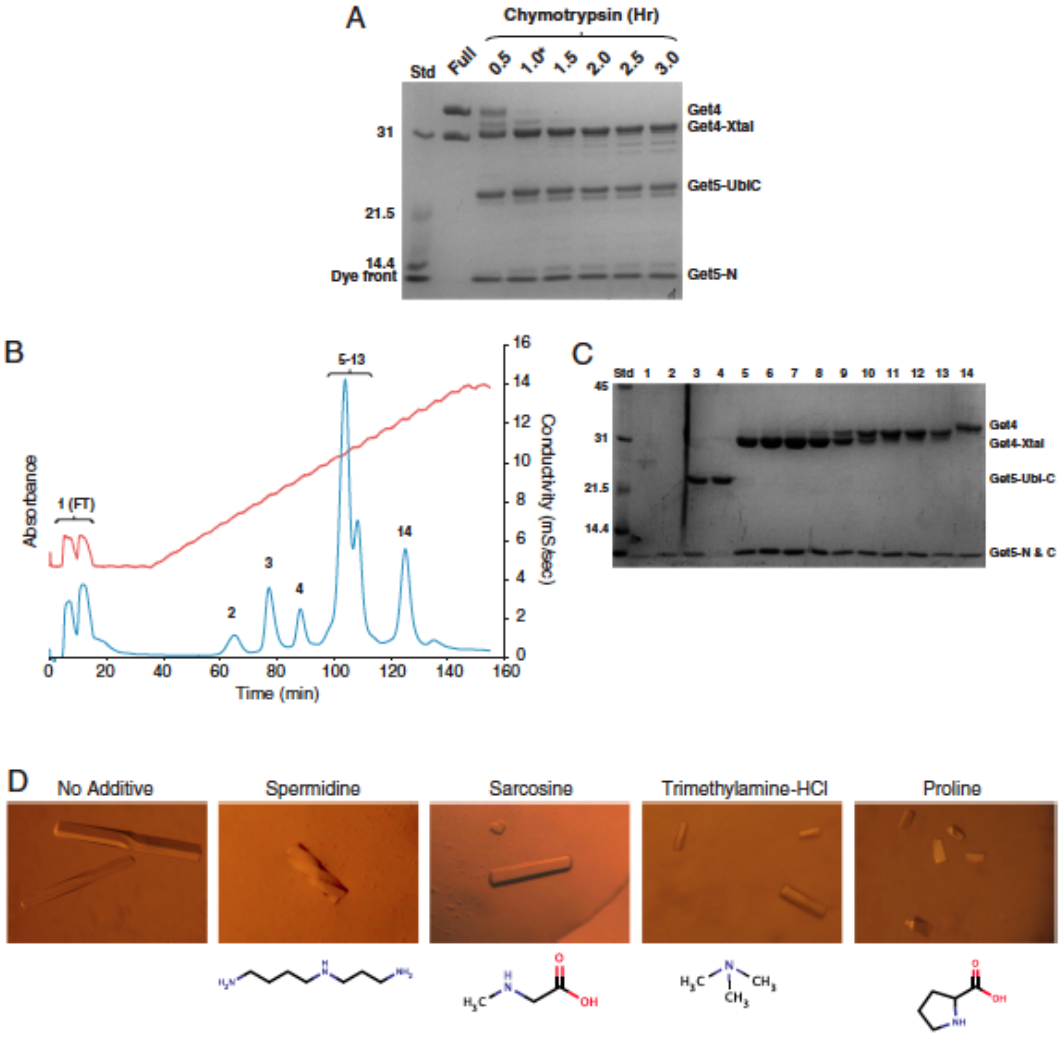


Figure 1.2. Proteolysis, purication, and crystallization. (A) Chymotrypsin is added to Ni-NTA anity puried Get4/5 on ice with aliquots removed at the indicated time points in hours and separated by SDS-PAGE. "Full" represents dimeric Get4/5 prior to the addition of enzyme. The identities of the Get4-Xtal, Get5-Ubl-C and Get5-N bands were confirmed by tryptic digest followed by LC/MS. The 1 h time point, indicated with an asterisk, represents the degree of digestion used for further purication (B) Anion exchange chromatogram of Get4/5 after digest by chymotrypsin using a 30–130 mM NaCl gradient. Conductivity is plotted as the red trace and A280 as the blue trace. The numbers above peaks represent lanes in C. (D) Representative crystals of Get4/5-N. Crystals are grown in 17% PEG 6000, 0.14 M ammonium sulfate, 0.1 M Bis-tris pH 5.5 with the following additives: None, 10 mM spermidine, 10 mM sarcosine, 10 mM Trimethylamine-HCL, or 10 mM L-proline.

Description of the Structure of the Get4/5-N Complex

Get4, which has no predicted sequence motifs, is essentially two rectangular blocks formed by right-handed α-helical coils that can be divided into N-terminal and C-terminal domains (N-domain and C-domain) (Figure 1.3B). The N-domain consists of the first 7 helices that are similar in length and are reminiscent of the TPR motif (D'Andrea and Regan, 2003). Unlike this motif, they do not contain an obvious internal consensus nor exhibit any curvature, a common feature of helical repeats. The C-domain continues with right-handed helical coils; however, the helix and loop lengths are more diverse. The loop between helices α11 and α12 is formed by two β-strands (β-tongue). The helix α13 makes a sharp turn into helix α14 that then bends to form helix α15 generating a U shape.

Get5-N forms an extended peptide that wraps tightly around the C-domain of Get4 (Figure 1.4). It begins with a helix that docks in a groove formed by helices α12, α13, and the β-tongue of Get4 (Figure 1.3A). Outside of the conserved hydrophobic interactions of the helix, the rest of Get5-N forms relatively few specific interactions to Get4 (Figure 1.4). The helix is followed by an extended, highly ordered loop that follows a groove contacting the loops formed by α12/α13, α10/α11, and α8/α9 in Get4 via backbone contacts. Comparing the structure to sequence conservation, the lengths of these Get4 loops are highly conserved implying that this platform is important in higher eukaryotes as well. The rest of Get5-N follows a groove formed by α8, α10, and α15 and then finally contacting the loop between α7/α8 of Get4. α8 is bookended by Get5-N and the short length of this helix appears to be conserved.

Table 1. 1 Crystallographic statistics

Data collection	Get4/5 Se peak
Beamline	SSRL 9-2
Wavelength (Å)	0.97862
Space group	P 3_1 2 1
Cell dimensions	
a, b, c (Å)	108.7, 108.7 169.8
α, β, γ (°)	90, 90, 120
Resolution range (Å)	30-2.8 (2.95−2.80)*
R_{merge}	0.104 (0.500)
I/σI	20.3 (4.7)
Completeness (%)	99.9 (100.0)
Multiplicity	12.1 (11.7)
Anom. compl. (%)	99.9 (99.9)
Anom. multiplicity	6.3 (6.1)
Phasing	
Se sites	9
PHASER figure of merit (8)	0.29
RESOLVE figure of merit (9)	0.65
Refinement	
No. reflections, working	27722
No. reflections, free	1464
No. atoms	
Protein	8439
Water	84
Ligands	16
Avg. B-factor (Å ^{2})	52.3
Working R-factor	0.18152
Free R-factor	0.22437
rmsd bonds (Å)	0.011
rmsd angles (°)	1.2

*Values in parentheses represent highest resolution shell.

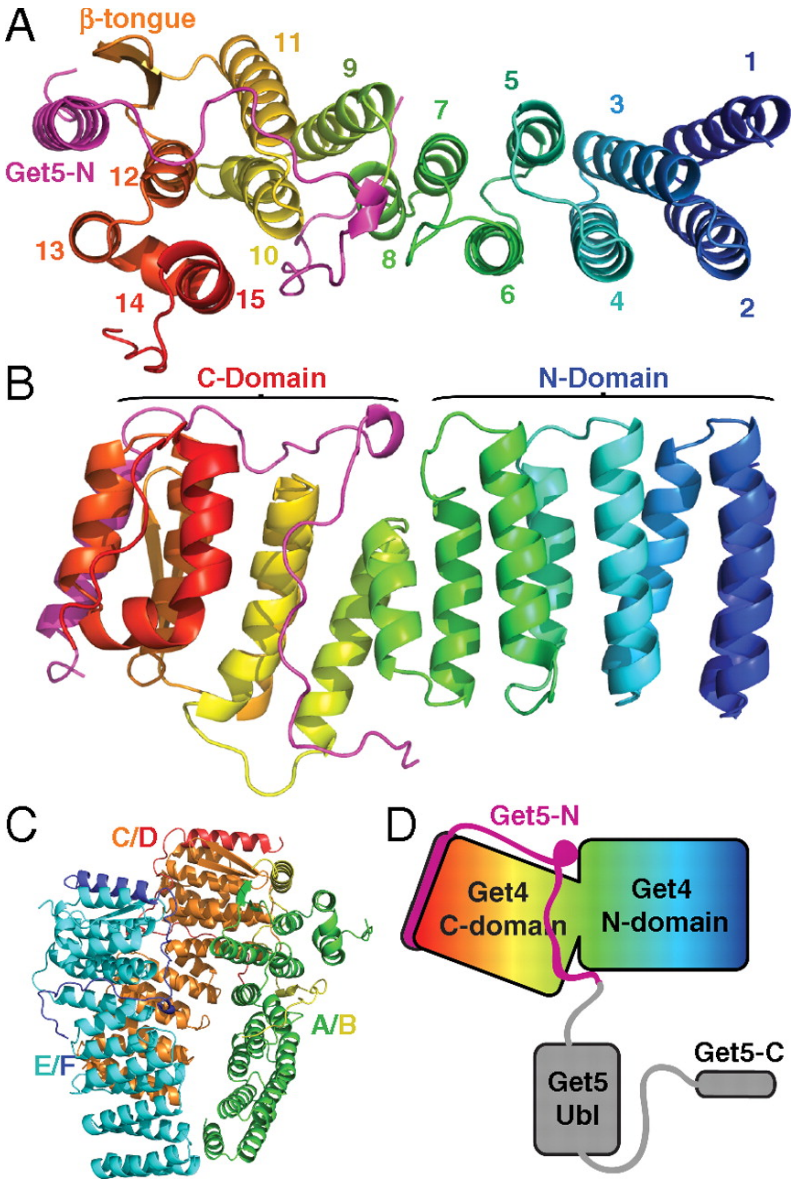


Figure 1.3. The structure of Get4/5-N. (A) Top view of Get4/5-N with Get4 color-ramped from N- (blue) to C-terminus (red) and Get5-N shown in magenta. Secondary structure elements are labeled as in Figure 1.1. (B) Side view relative to (A) with N- and C-domains indicated. (C) The asymmetric unit with each chain colored individually and labeled as in the deposited coordinates. (D) A cartoon of the structure based on (B). The missing Ubl- and C-domains of Get5 are shown in gray.

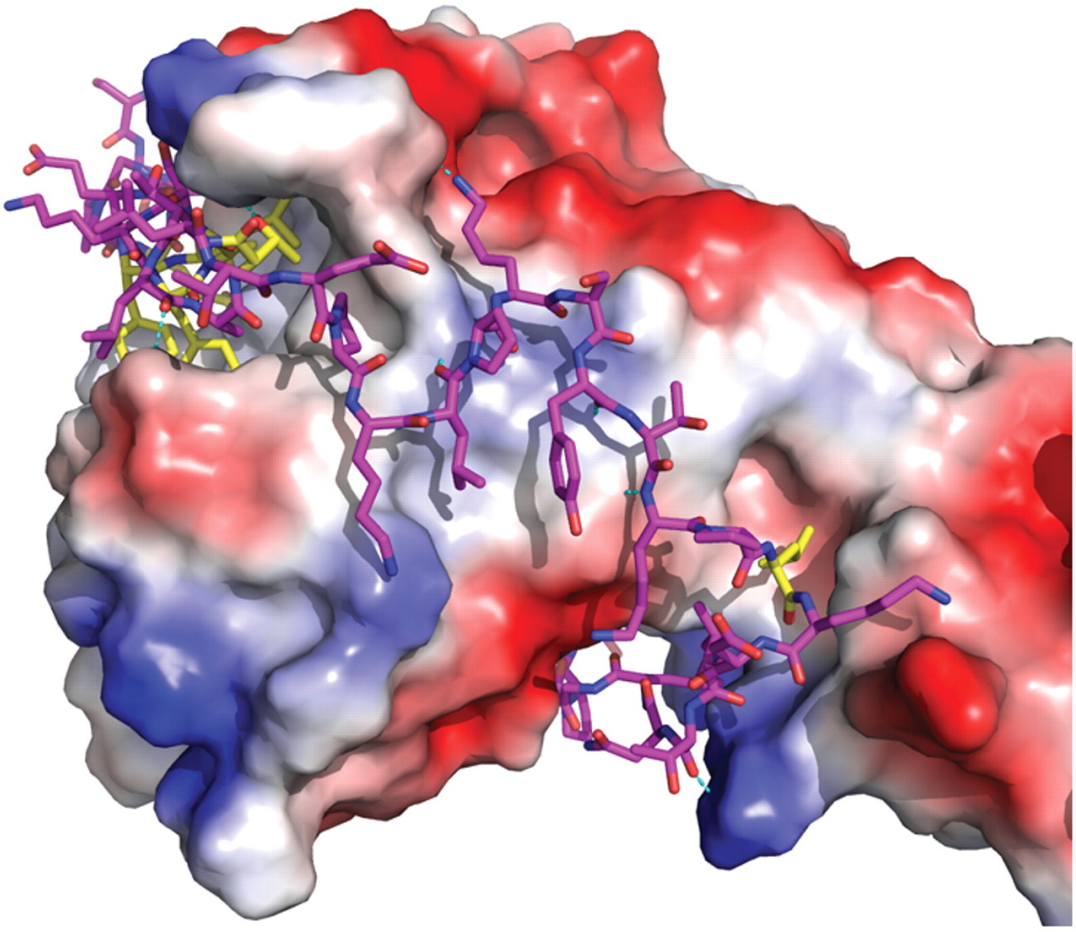


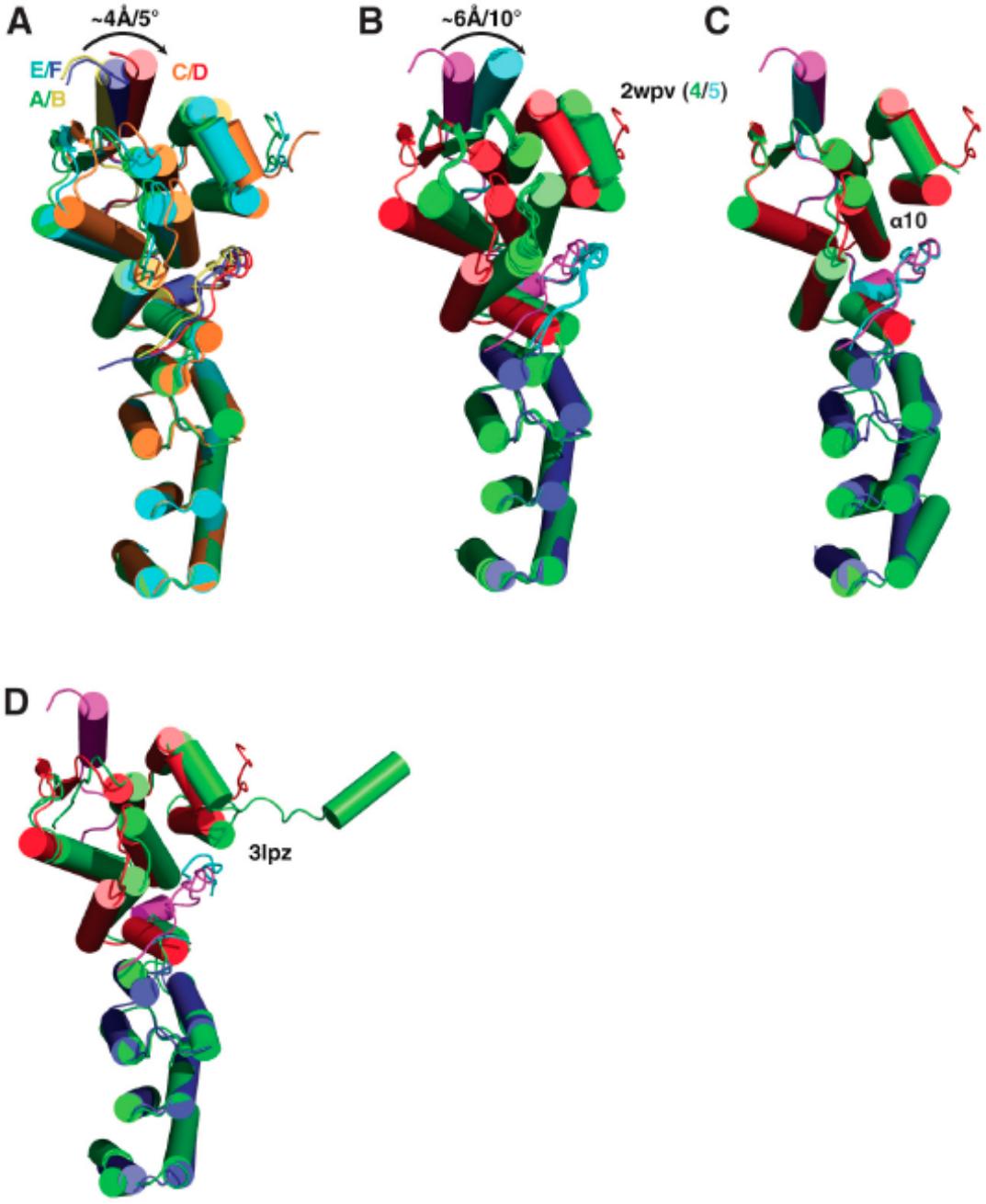
Figure 1.4. Binding of Get5-N. Get4 is shown as an accessible surface colored from positive (blue) to negative (red) Coulombic charge in an orientation similar to Figure 1.1A. Get5-N is shown as sticks in magenta with residues that are conserved making specific contacts in yellow.form helix α15 generating a U shape. The C-terminus forms an extended peptide that docks against a neighboring molecule.

Comparison to Other Get4 and Get5-N Structures

When the three molecules in our asymmetric unit are aligned based on their N-domain there is a clear twist in Get4 at α7/α8 that results in a relative bend in the C-domain. The greatest difference is between the A molecule and the C molecule with a relative rotation of about 5° and a maximal shift of about 4 Å (Figure 1.5A). In the Get4/5-N structure by Chang et al., there is very little difference between the four molecules in their asymmetric unit [Protein Data Bank (PDB) ID code 2 wpv] (Figure 1.5B) (Chang et al., 2010). When aligned individually relative to our structure, there are no major differences in the overall fold of the N-domain (Figure 1.5B) or the C-domain (Figure 1.5C); however, there is a larger twist between the two domains that results in a 10° rotation and a 6 Å shift (Figure 1.5B). Part of this rotation is taken up by a shift in α10 (Figure 1.5C). Other differences are extensions of most of the termini in the structure reported here. The structure of Get4 reported by Bozkurt et al. shows the conservation of the overall fold of Get4 and appears to be in a conformation similar to our A molecule; however, there are some significant distortions, presumably the result of the missing Get5-N (PDB ID code 3lpz) (Figure 1.5D) (Bozkurt et al., 2010).

Surface Features of Get4

As noted, based on sequence the overall fold of Get4 appears to be conserved across eukaryotes excluding the β-tongue (Figure 1.1A). The internal fold accounts for the majority of the highly conserved residues. Only two surfaces at the ends of the molecule have a high level of conservation (Figure 1.6A). The C-domain conserved surface contributes to the binding of Get5-N despite the lack of Get5-N in higher eukaryotes.


A
~4Å/5°
E/F
A/B
C/D
B
~6Å/10°
2wpv (4/5)
C
α10
D
3lpz

The largest conserved surface is the N-domain face (Figure 1.6A). The overall surface of Get4/5-N is acidic with a single basic patch directly correlating to the conserved face of the Get4 N-domain (Figure 1.6B).

Dimerization

As stated above, Get4/5 eluted as two peaks by ion exchange chromatography. Each peak ran anomalously large on a size exclusion column in the expected size range of 4–8 copies relative to typical globular proteins (Figure 1.7A, blue trace). We confirmed that the lower molecular weight peak corresponded to a single copy of Get4/5 and the higher molecular weight peak to two copies of Get4/5 by multiangle light scattering (MALS) (Figure 1.8A). We refer to these two forms as the "monomer" and dimer in the rest of the text. We noted that the dimer peak was stable while the monomer peak partially converted to the dimer peak over time. These results suggest that the proteins have exaggerated hydrodynamic properties relative to what would be predicted based on molecular weight.

Figure 1.5. Comparison to other crystal structures. All views are from the bottom of Get4/5-N relative to figure 1.1A. (A) Aligning the three molecules in the asymmetric unit relative to the first six helices in molecule A. The molecules are colored the same as in figure 1.1C. The rotation relative to the domains is indicated. (B) Similar to (A) with the A/B molecule colored N-domain (blue), C-domain (red), and Get5-N (magenta). The four molecules in the asymmetric unit of Chang et al. (1 wpv) are aligned to the N-domain as in (A) colored by Get4 (green) and Get5-N (cyan). (C) Similar to (B) but molecules 1 wpv-A/B are aligned to the C-domain. The shifted α10 is indicated. (D) The Bozkurt et al. (3lpz) Get4 (green) aligned to the A molecule N-domain. Included are a C-terminal peptide seen in that crystal structure that docks into the Get5-N binding groove (cyan).

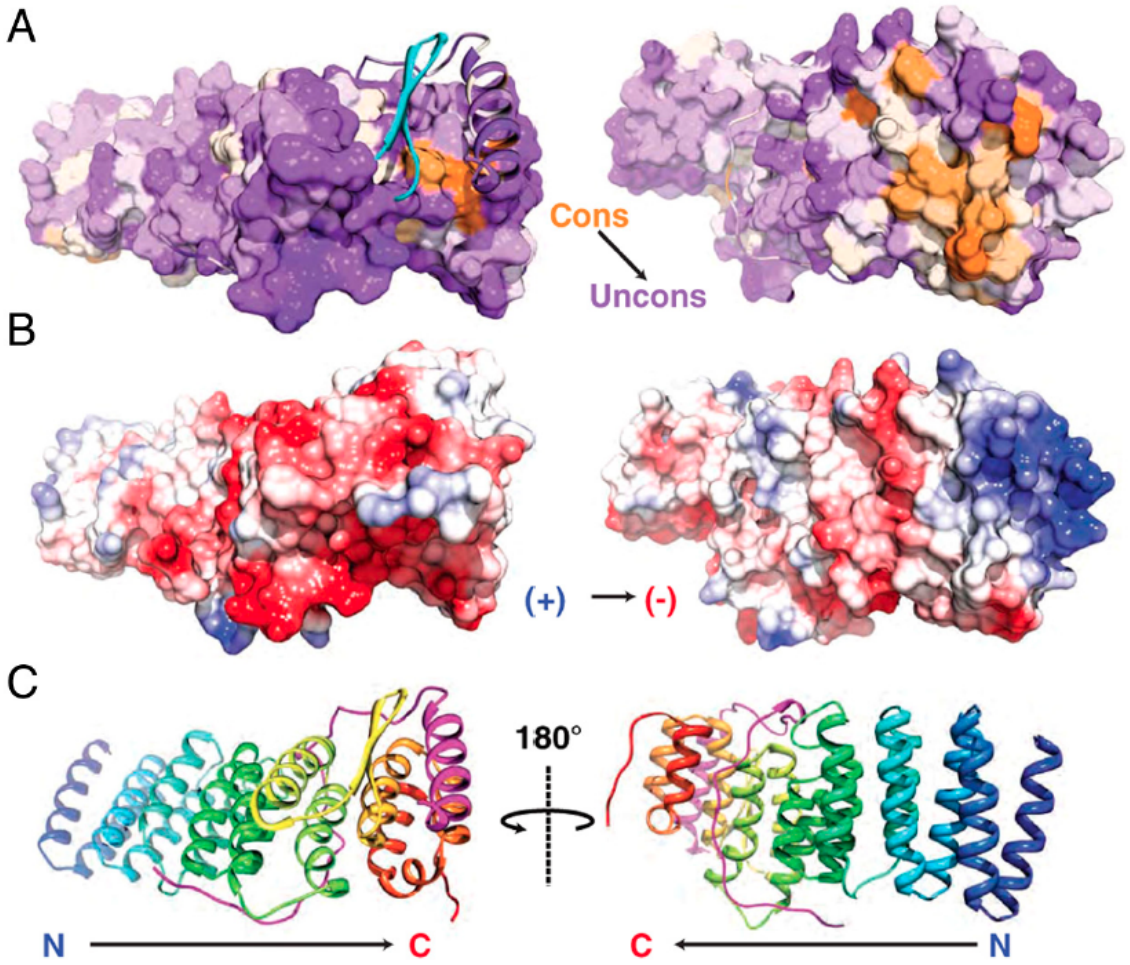
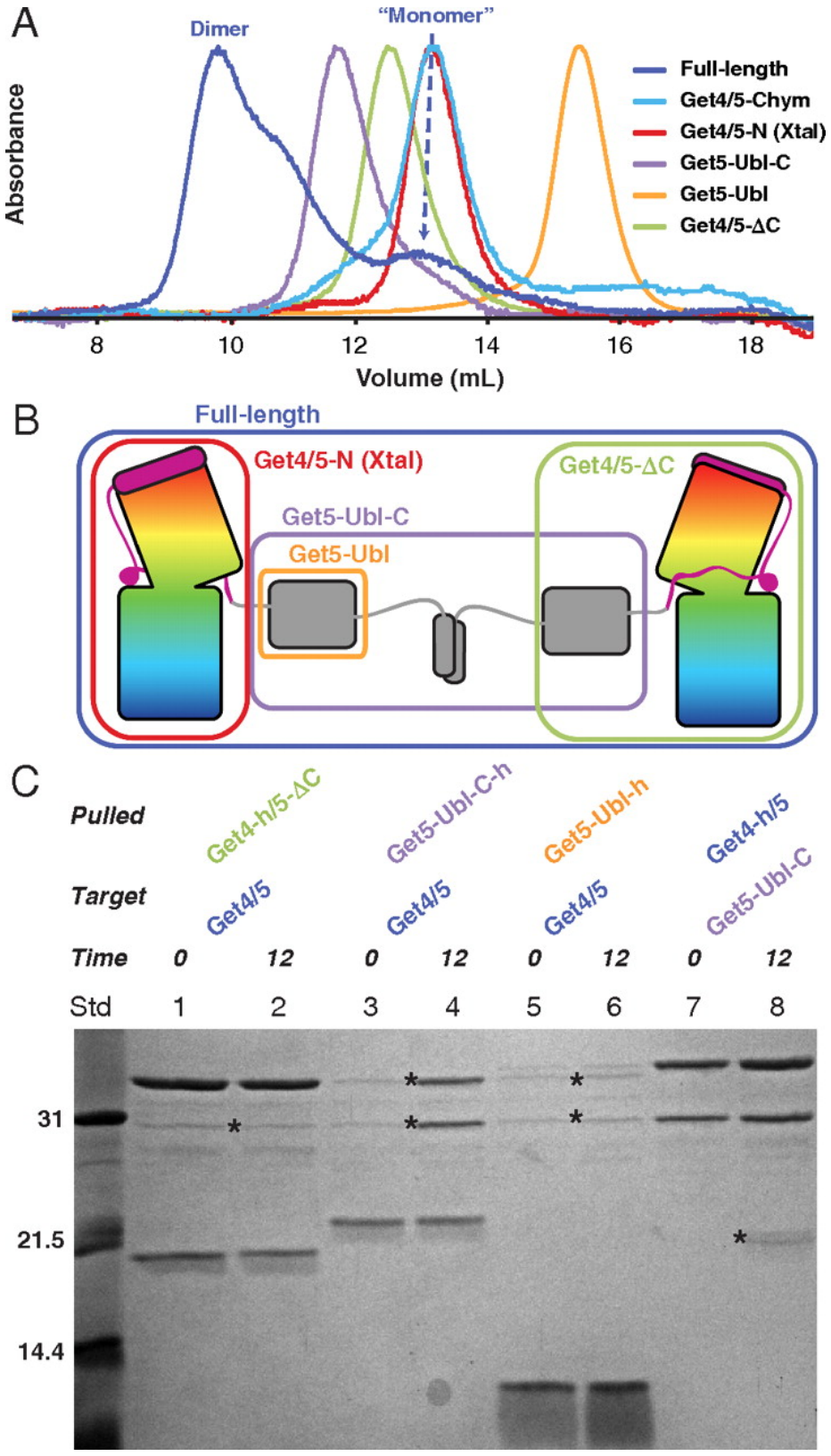


Figure 1.6. Surface conservation and charge. (A) Get4 shown as an accessible surface and Get5-N as a ribbon colored from most (orange) to least (purple) conserved. The fungal specific β-tongue is shown as a cyan ribbon. (B) As in A with the total accessible surface colored by Coulombic charge. (C) Ribbons diagram in the same orientations for reference.


A
Dimer
"Monomer"
Absorbance
Full-length
Get4/5-Chym
Get4/5-N (Xtal)
Get5-Ubl-C
Get5-Ubl
Get4/5-ΔC
8
10
12
14
16
18
Volume (mL)
B
Full-length
Get4/5-N (Xtal)
Get4/5-ΔC
Get5-Ubl-C
Get5-Ubl
C
Pulled
Target
Time
Std
Get4-h/5-ΔC
Get5-Ubl-C-h
Get5-Ubl-h
Get4-h/5
Get5-Ubl-C
Get4/5
0
12
1
2
3
4
5
6
7
8
31
21.5
14.4

Analyzing the sequence features of Get5, there are two probable flexible loops connecting the Ubl to the N- and C-domains that could account for the larger radius. Addition of chymotrypsin to either monomer or dimer immediately cleaved the loop between Get5-N and the Get5-Ubl-C fragment; however, the loop connecting to the C-domain was resistant to cleavage (Figure 1.2A). When this proteolyzed mix was run on the size exclusion column the peak shifted dramatically to a much smaller size with a small shoulder that eluted earlier (Figure 1.7A, cyan trace). The two dominant proteolysis fragments, Get4/5-N and Get5-Ubl-C, could be purified by ion exchange chromatography (Figure 1.2B,C). When these fragments were run on the size exclusion column Get4/5-N (Figure 1.7A, red trace) corresponded to the bulk of the proteolysis peak and Get5-Ubl-C was the leading shoulder (purple trace). These peaks were clearly resolved; therefore, Get4/5-N does not form a stable complex with Get5-Ubl-C. The estimated extinction coefficient is much lower for Get5-Ubl-C and the smaller leading shoulder in the protease fragment reflects this (cyan trace). Get5-Ubl-C (15.9 kDa), a dimer by MALS, ran anomalously large on the sizing column even ahead of the larger Get4/5-N (40 kDa) (Figure 1.7A) This implies a model that involves dimerization of Get4/5 by the Get5-Ubl-C fragment (Figure 1.7B).

Figure 1.7. Dimerization by Get5-C. (A) Size exclusion chromatograms from various constructs of Get4/5 after affinity purification. “Chym” represents Get4/5 after a 15-minute chymotrypsin digest. The absorbances are normalized to the highest recorded value. The dimer and monomer peaks for full-length Get4/5 are indicated. (B) Cartoon of the Get4/5 dimerization by Get5-C. Constructs used in A and C are indicated by colored boxes. (C) Domain swapping by dimerization. His-tagged proteins, indicated by “h”, with and without Get5-C are mixed with untagged protein containing Get5-C for 0 and 12 h.

The mass of the Get4/5 monomer peak by MALS was approximately 8 kDa larger than that predicted for Get4/5 and contained a small peptide in the range of 7 kDa (Figure 1.8). We tested the peptide by mass spectrometry and N-terminal sequencing and determined that it corresponded to an internal start site at the single methionine in our Get5 construct; i.e., it was an expressed Get5-C domain. The extra C-domain in the monomer prevented dimer formation with full-length Get4/5 leading to the conclusion that this domain was involved in dimerization. To further explore this we expressed and purified the additional constructs of Get5-Ubl-C, Get4/5-ΔC and Get5-Ubl (Figure 1.7B). The constructs without C-domains both ran as monomers as predicted (green and orange traces, respectively). Get5-Ubl-C behaved similarly to the cleaved Get5-Ubl-C fragment and we again detected a fraction of the protein bound to a Get5-C peptide similar to the full-length monomer (Figure 1.8B). Get5-C, including its disordered loop, is found in all eukaryotes and in yeast the sequence of its C-terminal helical region is consistent with a possible coiled-coil, as predicted by COILS (Lupas et al., 1991).We hypothesized that two different constructs of Get4/5 may exchange their C-domains to form mixed Get4/5 dimers. To test this, various constructs of Get4/5 (Figure 1.9A) were mixed with full-length Get4/5 and the products were analyzed after a 12 h incubation (Figure 1.7C). An affinity tagged Get4/5-ΔC was unable to capture a full-length Get4/5 after 12 h (lane 2 compared to lane 1); however, Get5-Ubl-C was able to capture Get4/5 (lane 4) and Get4/5 could capture Get5-Ubl-C (lane 8). Get5-Ubl alone was unable to capture Get4/5 (lane 6). These results verify that the C-domain was responsible for Get4/5 dimerization and rule out a role for the Ubl-domain.

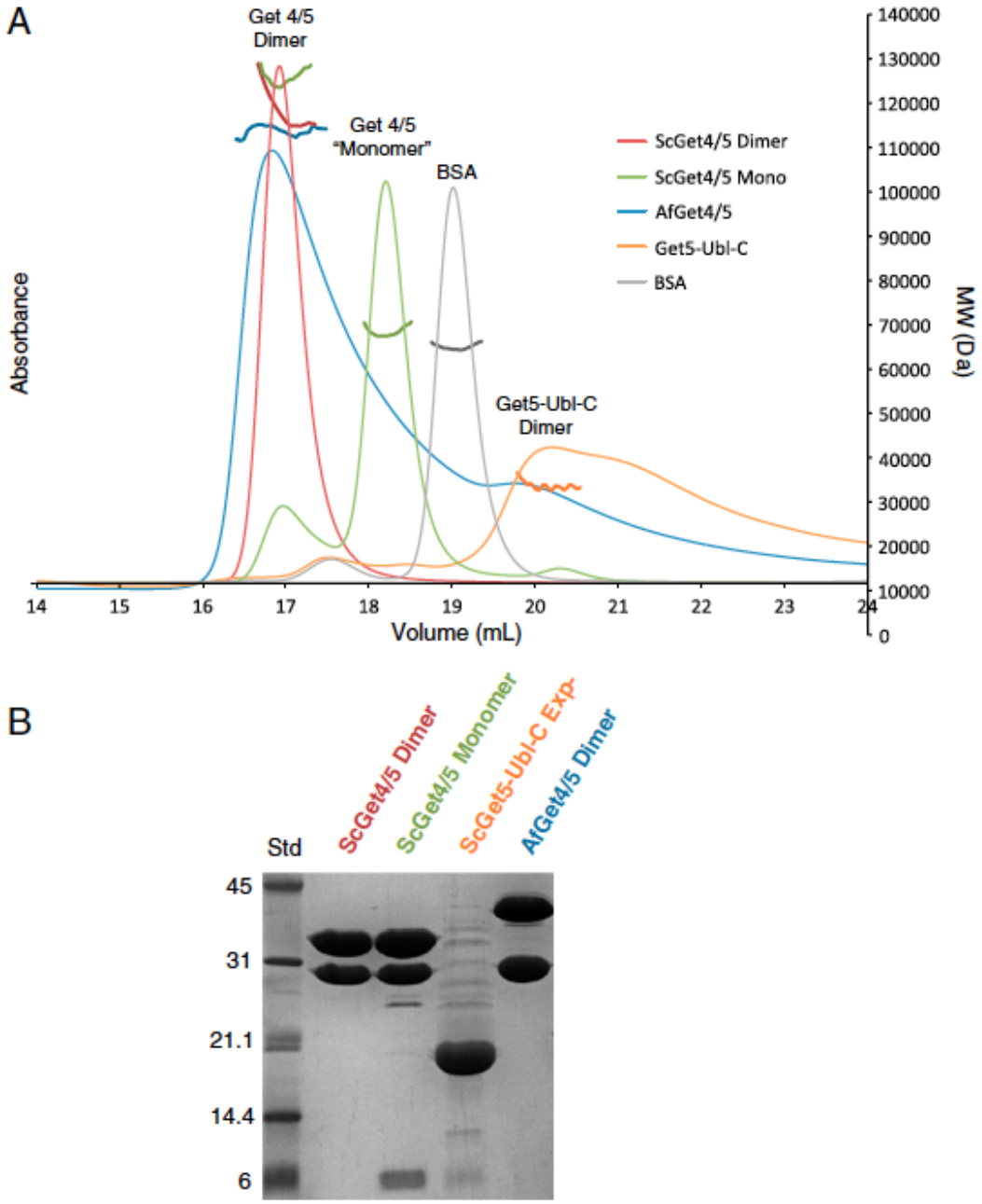


Figure 1.8. Multiangle light scattering. (A) Overlay of size exclusion chromatograms using a Shodex KW 803 column. Traces represent UV absorbance (left axis). For each peak, molecular weight as determined using multiangle light scattering and refractive index is plotted (right axis). (B) Coomassie-stained SDS-PAGE of samples used in (A).

A coiled-coil prediction for the Get5 C-terminal region was found in higher eukaryotes but was not seen in all fungi. An example is *Aspergillus fumigatus* Get4/5, and we wanted to see if Get4/5 formed a dimer in this species. By MALS, the purified full-length *Af*Get4/5 corresponded to a stable dimer (Figure 1.8A). We never detected a monomer fraction for this species nor did we ever see the equivalent of the Get5-C fragment (Figure 1.8B). To confirm dimerization by the C-domain, we generated *Af*Get4/5 constructs with the C-domain removed (*Af*4/5-ΔC) and the Get5 Ubl-C-domain alone (*Af*5-Ubl-C). All three *Af* constructs behaved the same as their *Sc* equivalents on a size exclusion column (Figure 1.10A,B). Additionally, the *Af* constructs were able to exchange based on the presence of the C-domain (Figure 1.10C lanes 1-10). Not surprisingly, the *Sc*5-Ubl-C was not able to swap with the *Af*5-Ubl-C confirming that dimerization is conserved independent of sequence (Figure 1.10C, lanes 13 and 14).

In Vitro Interaction of Get4/5 and Get3

Previous studies have demonstrated that Get4/5 can form a complex with Get3; however, these studies did not address the specifics of the interaction nor did they address the role of nucleotide in binding (Jonikas et al., 2009; Yu et al., 2008). We further explored this interaction using purified components. Initially, an affinity-his-tagged Get4/5 was used to test the binding of Get3 (Figure 1.11A) Using Ni-affinity beads, very little Get3 could be captured by Get4/5 when mixed in the absence of nucleotide (lane 2). The addition of ADP (lane 3) or ATP (lane 4) dramatically increased the amount of bound Get3. The structures of the apo form of Get3 were always in an

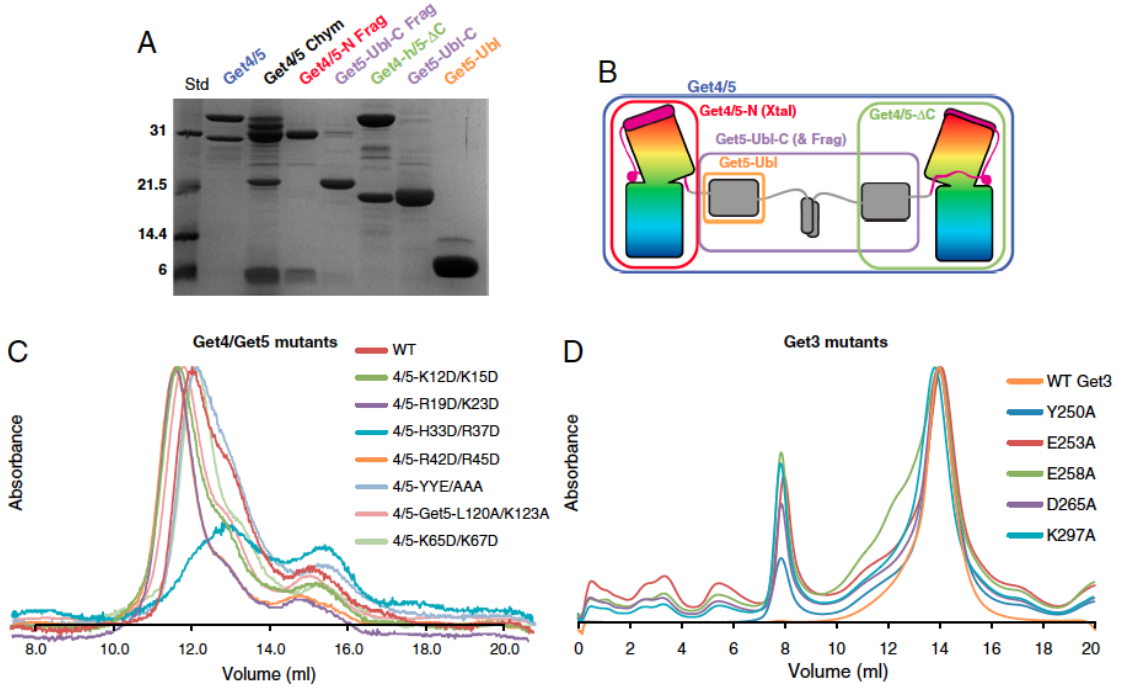


Figure 1.9. Mutant purification. (A) Coomassie-stained SDS-PAGE of various puried Get4/5 constructs. Get4/5 Chym is the result of a 15 minute room temperature chymotrypsin digest. Get4/5-N Frag and Get5-UblC Frag are purified from the chymotrypsin digest using anion exchange chromatography as in Figure 1.2B. Get4-h/ 5-C, Get5-UblC, and Get5-Ubl are generated at the genetic level. (B) Cartoon diagram of Get4/5 dimer with constructs boxed and colored as in A. (C) Size exclusion chromatograms of mutants of Get4/5 using a Superdex 200 10/300 column. The absorbences are normalized to the highest recorded value. (D) As in C, for mutants of Get3.

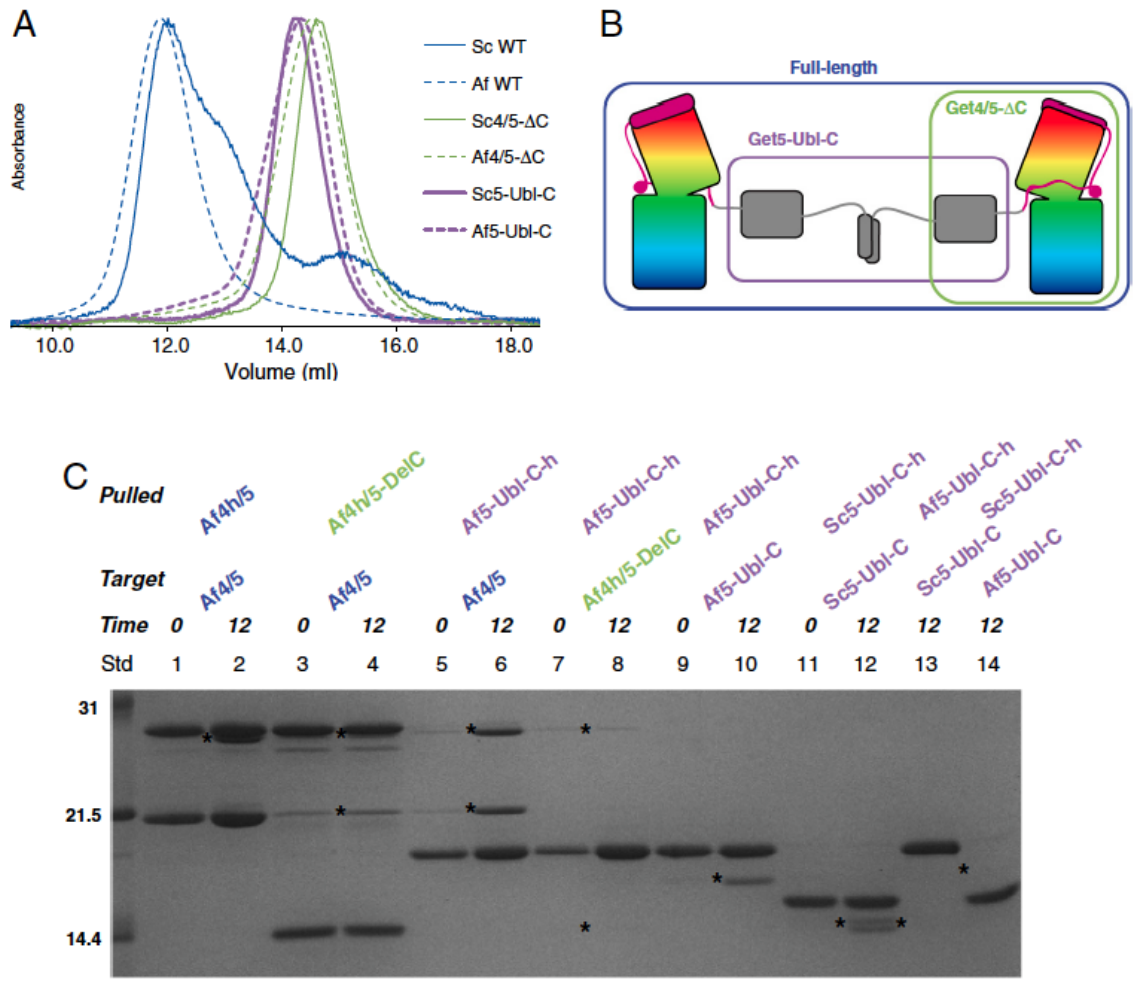
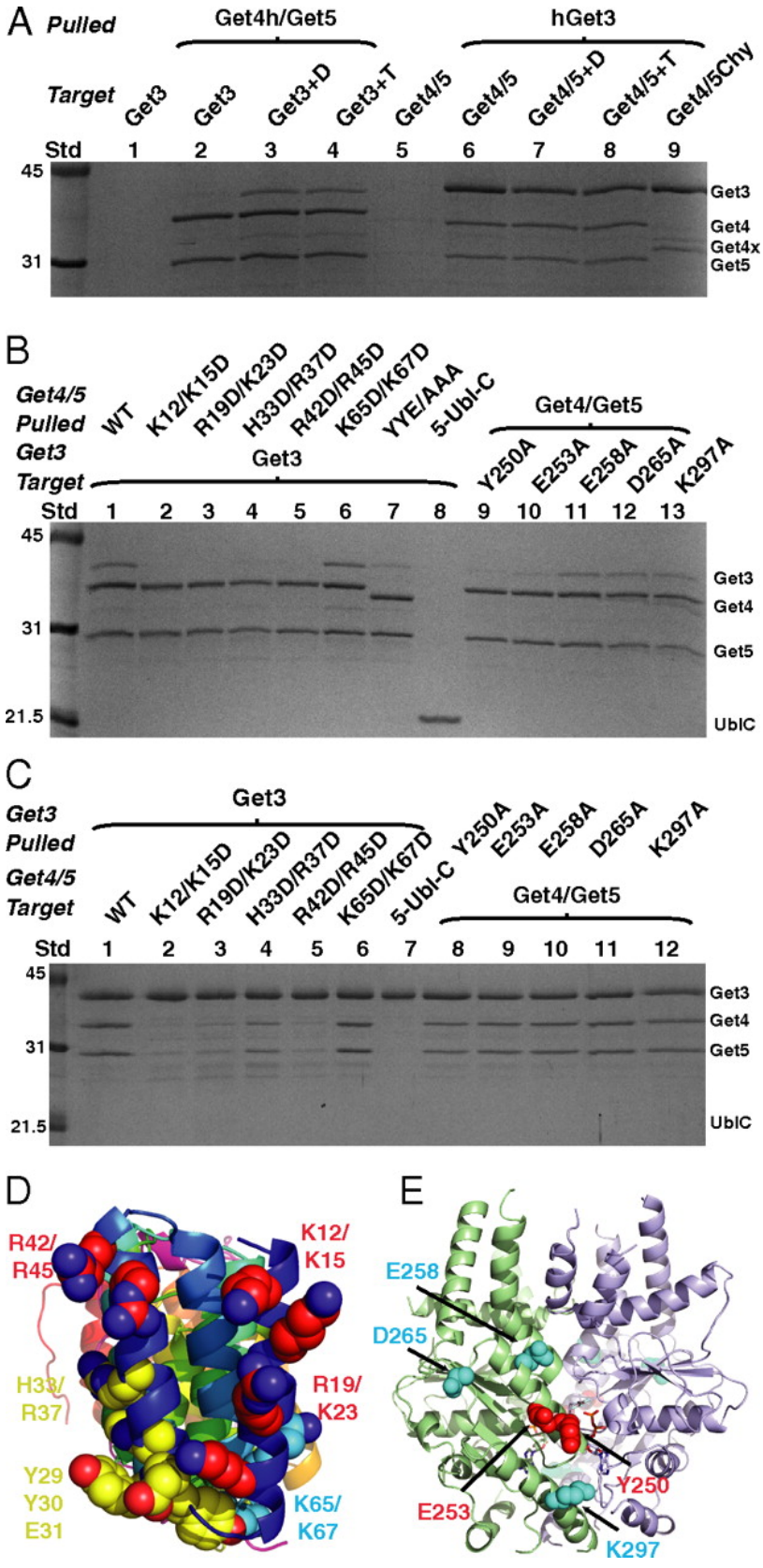


Figure 1.10. Dimerization of AfGet4/5 by Get5-C. (A) Size exclusion chromatograms of constructs of Get4/5 from Sc, as in Figure 1.9A, and Af (dashed lines). The absorbances are normalized to the highest recorded value. (B) Cartoon of constructs used in (A) indicated by colored boxes. (C) Domain swapping by dimerization. His-tagged proteins, indicated by "h," with and without Get5-C are mixed with untagged protein containing Get5-C for 0 and 12 h. Respective Af and Sc constructs are indicated. Asterisks (*) represent expected bands for successful capture.

open form (Hu et al., 2009; Mateja et al., 2009; Suloway et al., 2009); however, structures of Get3 adopt both open and closed forms in the presence of nucleotide (Bozkurt et al., 2009; Mateja et al., 2009; Suloway et al., 2009). The fact that binding is enhanced by nucleotide would imply that nucleotide has shifted the equilibrium from the open to the closed form, and it is the closed form that is recognized by Get4/5. Get3 was able to capture Get4/5 (lane 6) but this was not enhanced by nucleotide (lanes 7 and 8). Our Get3 has an N-terminal affinity tag, and binding this to beads may also shift the equilibrium to the closed form. Get3 captured only Get4/5-N from our proteolyzed pool and not Get5-Ubl-C (lane 9); therefore, Get3 binds specifically to Get4/5-N.

Based on these results, we searched for mutants that would affect Get4/5 binding to Get3. The most conserved surface of Get4 is on the N-terminus where there is a patch of highly conserved positive residues (Figure 1.6). To evaluate potential effects of these charged residues in the Get3 interaction, we mutated pairs of positive charges to aspartates to fully disrupt possible interfaces (K12D/K15D, R19D/K23D, H33D/R37D, and R42D/R45D). In addition, to rule out the general effects of charge swapping, we generated another pair of mutants of unconserved residues (K65D/K67D) as a control. Finally, we mutated the highly conserved Tyr 29, Tyr 30, and Glu 31 to alanines (YYE/AAA) (Fernandes et al., 2008). All of these mutants were purified and behaved similar to wild type on a size exclusion column (Figure 1.9C).

The purified mutants were mixed with Get3 and captured using Ni-affinity beads. All of the conserved mutants showed a significant loss in the ability to capture Get3 (Figure 1.11B, lanes 1–6). Two of the mutants, H33D/R37D (lane 3) and YYE/AAA (lane 7),


A
Pulled
Target
Get4h/Get5
hGet3
Get3
Get3
Get3+D
Get3+T
Get4/5
Get4/5
Get4/5+D
Get4/5+T
Get4/5Chy
Std 1 2 3 4 5 6 7 8 9
45
31
Get3
Get4
Get4x
Get5
B
Get4/5 Pulled
Get3 Target
WT
K12/K15D
R19D/K23D
H33D/R37D
R42D/R45D
K65D/K67D
YYE/AAA
5-Ubl-C
Get3
Get4/Get5
Y250A
E253A
E258A
D265A
K297A
Std 1 2 3 4 5 6 7 8 9 10 11 12 13
45
31
21.5
Get3
Get4
Get5
UblC
C
Get3 Pulled
Get4/5 Target
Get3
WT
K12/K15D
R19D/K23D
H33D/R37D
R42D/R45D
K65D/K67D
5-Ubl-C
Y250A
E253A
E258A
D265A
K297A
Get4/Get5
Std 1 2 3 4 5 6 7 8 9 10 11 12
45
31
21.5
Get3
Get4
Get5
UblC
D
R42/
R45
K12/
K15
H33/
R37
R19/
K23
Y29
Y30
E31
K65/
K67
E
E258
D265
Y250
E253
K297

captured Get3 at a markedly reduced level. As expected, Get5-Ubl-C was unable to capture Get3 (lane 8). By reversing the experiment we saw the same pattern of mutant Get4/5 capture by Get3 (Figure 1.11C, lanes 2–7). These results center the binding of Get3 to the positive N-terminal face of Get4 (Figure 1.11D and Figure 1.12A,B).

In our previous study of Get3, we had mutated a number of conserved Get3 surface residues that were unable to rescue a *Δget3* knockout (Suloway et al., 2009). These residues could not be explained by contacts in the Get3 dimer in either the open or closed state and many localized to a negative surface (Figure 1.12C–E). We tested several for Get4 binding by alanine mutation (Y250, E253, E258, D265, and K297). All of the mutants expressed well and behaved similar to wild type on the size exclusion column (Figure 1.9D). When wild type Get4/5 was used to capture the Get3 mutants only two (Y250A and E253A) showed a significant decrease in binding (Figure 1.11B, lanes 9–13 compared to 1). These two mutants are near the interface of the Get3 dimer and presumably destabilize the closed state (Figure 1.11E). All of the Get3 mutants were able to capture Get4/5 (Figure 1.11C, lanes 8–12). The bead bound Get3 is less sensitive to nucleotide; therefore, the closed form may be favored despite the mutants.

Figure 1.11. Binding of Get4/5 to Get3. (A) Tagged wild type Get4/5 and Get3 are incubated for 2 h at room temperature then bound to Ni-beads. "D" represents incubation in the presence of 2 mM ADP-Mg^{2+}, and "T" in the presence of 2 mM ATP-Mg^{2+}. (B) His-tagged Get4/5 and Get3 mutants are incubated in the presence of 2 mM ADP-Mg^{2+} as in (A). The YYE/AAA Get4 mutant lacks a TEV cleavage site and is slightly smaller than the other tagged mutants. (C) As in (B), using his-tagged Get3. (D) View from the N-terminal face of Get4. Mutated residues are displayed as spheres with carbons and labels colored based on wild type level (cyan), weak (yellow), or weakest (red) interactions with Get3. (E) The closed state of Get3 (PDB ID code 2WOJ) with mutated residues displayed and colored as in (D).

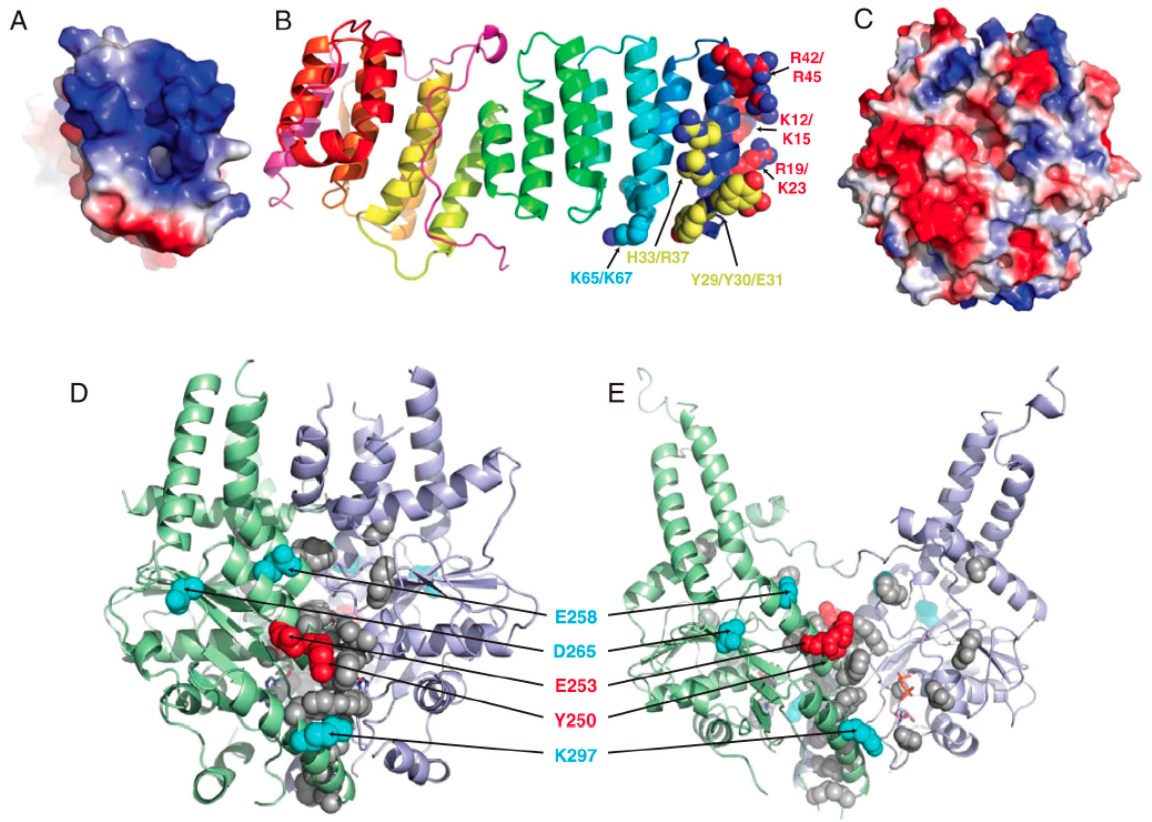


Figure 1.12. Mutants on Get4 and Get3. (A) View from the N-terminal face of Get4 as an accessible surface colored from positive (blue) to negative (red) Coulombic charge. (B) Side view of Get4/5 as in Figure 1.3B. The mutants used in Figure 1.11B,C are displayed as spheres as colored based on wild type (cyan), weak (yellow), or weakest (red) binding to Get3 as in Figure 1.11D. (C) Get3 similar to (A) in the closed state (PDB ID code 2WOJ) (D) Ribbon diagram of the closed state of Get3 with mutants that had previously been shown to have strong phenotypes as spheres. Tested mutants are colored as in Figure 1.11E. (E) Ribbon diagram of the open state of Get3 colored as in (D) (PDB ID code 3IBG).

In Vivo Effects of Get4 and Get5 Mutants

To assess the mutants *in vivo*, we utilized the fact that deletion of either Get4 or Get5 showed growth phenotypes under stress conditions and tested for rescue by our constructs (Hillenmeyer et al., 2008). We rescued the Δ*get4* sensitivity to temperature and copper sulfate by expression of wild type Get4 using the native promoter on a plasmid (Figure 1.13A). With this construct, we generated all of the mutants used in the binding experiment. All of these mutants were unable to completely rescue the phenotype. The poorest rescue was by H33/R37, R42/R45, and YYE/AAA, which are on the highly conserved helix α2 (Figure 1.13B). We generated two mutants at the interface of the Get4 N- and C-domain (H137A and Y156A) that showed no phenotype (Figure 1.13).

The structure of the Ubl domain from the human homologue of Get5, Ubl4a/GDX, has been solved by NMR and deposited to the PDB (ID code 2dzi). Using this as a template we generated a homology model of Get5-Ubl using SWISS-MODEL and our alignment (Figure 1.14A) (Arnold et al., 2006). In ubiquitin isoleucine 44 is highly conserved and always involved in protein binding interactions (Hicke et al., 2005). In Get5 homologues, this residue is highly conserved as a leucine or methionine (L120) (Figure 1.1B). Based on the homology model, we noted that L120 was part of a conserved interface that included a number of positively charged residues including the conserved K124, which corresponds to the commonly conjugated K48 in ubiquitin (Figure 1.14A–C). Get5 deletion mutants are viable in rich media but show phenotypes under stress conditions (Hillenmeyer et al., 2008). Mutation of these residues to alanines (L120A/K124A) was

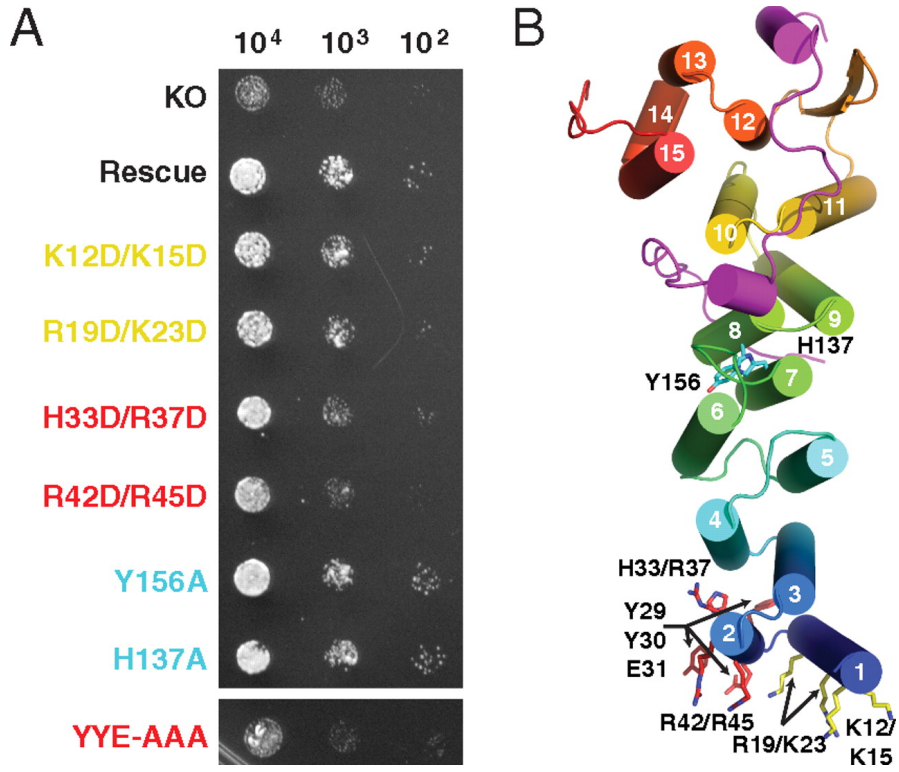


Figure. 1.13. Get4 rescue. (A) Spot plate growth assays of YEp-352 derived rescue plasmids under control of genomic promoters in the BY4741 Get4::KanMX background. Plates consisted of Sc-Ura supplemented with 2 mM $CuSO_4$ and were incubated at 37 °C. The panel is generated from a single plate. "KO" represents transformations with empty YEp-352 vector and "Rescue" represents wild type Get4. Rescue mutants labels are colored based on no (cyan), moderate (yellow), and strong (red) phenotypes. (B) Get4/5-N with helices shown as cylinders. Sets of mutated residues are shown as sticks with carbons colored based according to phenotype.

not able to fully rescue the *Δget5* at 37 °C (Figure 1.14D) despite the protein being stable in solution (Figure 1.9C)

Discussion

Get4 and Get5 are two highly conserved proteins whose functions have only recently begun to be understood. The growing consensus is a direct role for these proteins operating upstream of Get3 in the TA protein targeting pathway. Here we have presented a structure of the yeast Get4/5-N complex along with a model for the structural elements of Get5-Ubl-C. We have shown that the purified Get4/5 complex dimerizes mediated by the Get5 C-domain. The Get4 N-terminal face forms part of the recognition interface with Get3, apparently preferring to bind to the closed form, and this surface is important *in vivo*.

Get4 and the N-domain of Get5 form a stable and intimate complex. The fact that Get5-N appears to be important in yeast is somewhat surprising considering the absence of this domain in Get5 homologues in higher eukaryotes. One might speculate that in higher eukaryotes another protein can perform a similar interaction with Get4. An attractive option would be another protein that could bridge between Ubl4a and the mammalian Get4 homologue. This theoretical protein would contain a Get5-N-like region for binding to Get4, a dimerization domain for binding to Ubl4a and possibly a second Ubl. This complex would retain many of the features of Get4/5.

Get3 is a soluble protein that transiently interacts with Get4. In the current model, Get4/5 facilitates binding of TA proteins to Get3. The preference for a closed state of Get3 fits

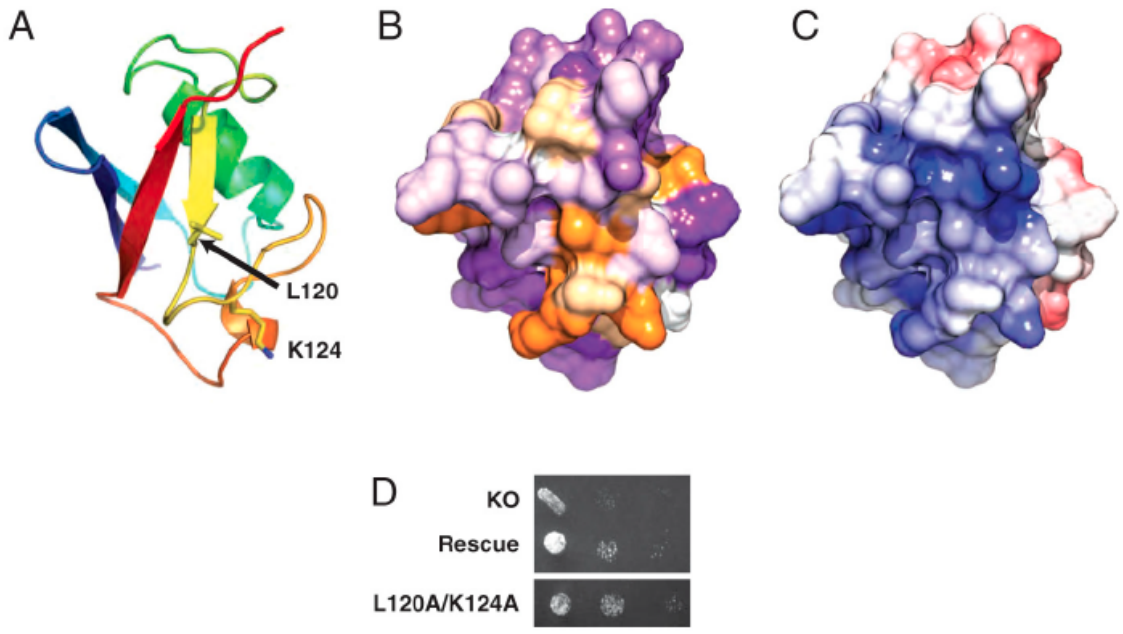


Figure 1.14. Get5 model with phenotypic rescue. (A) Ribbons diagram of a homology model of the Get5-Ubl-domain generated by sequence threading through the NMR structure of human UBL4A (PDB ID code 2DZI) using SWISS-MODEL, color ramped from N- (blue) to C-terminus (red). Residues L120 and K124 are displayed as sticks. (B) Get5-Ubl oriented as in A and shown as an accessible surface colored from least (purple) to most (orange) conserved. (C) As in (B), colored from positive (blue) to negative (red) Coulombic charge. (D) Spot plate growth assays in the BY4741 Get5::KanMX background, performed as in Figure 1.13.

nicely into the model where Get4/5 act as mediators to the ribosome and bind Get3 in a state competent for TA protein binding. The dimerization of Get4/5 adds the additional possibility that the 2-fold symmetrical Get3 dimer presents a binding site for each of the Get4 binding sites in the Get4/5 dimer. A mechanism such as this would lead to cooperative binding with a much higher affinity for Get3 in the correct state.

The results presented here allow us to clarify the role of Get4 and Get5 as intermediaries in TA targeting (Figure 1.15). Get4/5 are able to recognize the nucleotide state of Get3 and localize the closed form of Get3 to the ribosome dependent on an emerging TA substrate. Binding of the TA protein to Get3 leads to a conformational change that releases the Get3/TA complex from Get4/5 and the ribosome. Sgt2 and cellular chaperones either facilitate this transfer of substrate or act as parts of an alternate pathway.

The steps in the Get targeting pathway continue to become clear but there are many outstanding questions that remain. The yeast Get4/5 appears to interact with the ribosome via Get5 (Fleischer et al., 2006); however, it remains to be demonstrated that this is a direct interaction. The role of nucleotide hydrolysis in the targeting pathway is not clear and may be involved in fidelity of substrate selection at the ribosome or in release of the TA protein at the ER. Finally, although the precise role of dimerization remains to be elucidated, the fact that Get4/5 forms dimers is a provocative result in light of the symmetry of the Get3 dimer.

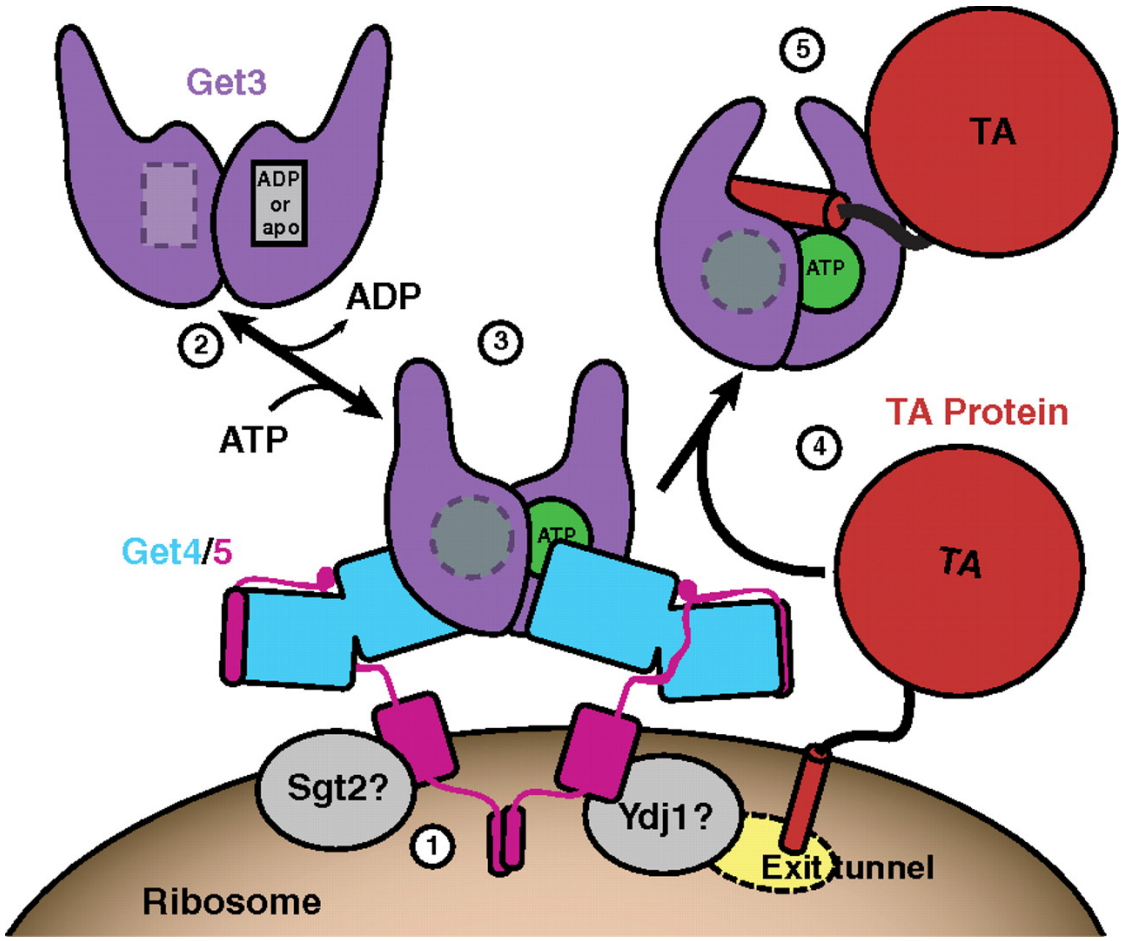


Figure 1.15. A model for the role of Get4/5. (1) Dimeric Get4/5 presumably binds the ribosome near the exit tunnel. Sgt2 and Ydj1 associate with the Get5-Ubl-C-domains. (2) Get3 in the open state is free in the cytoplasm. (3) Get3, in a closed state, is recruited to the Get4/5 complex upon ATP binding. (4) Tail-anchored proteins emerge from the ribosome and become associated with the Get3/Get4/5 complex. (5) The soluble Get3/TA protein complex is released to the cytoplasm for targeting to the ER.

Methods

Cloning, expression, and purification of proteins

The sequences of Get4 (YOR164c) and Get5 (YOL111c) were obtained from the *Saccharomyces* Genome Database, codon optimized for expression in *Escherichia coli* using DNAWorks and synthesized by PCR (Hoover and Lubkowski, 2002). Residues 1–9 of Get4 were truncated to Met-Gly due to lack of conservation and predicted disorder (Figure S1A). The genes were inserted sequentially into a pET33b(+)-derived vector (Novagen) and separated by an internal ribosome binding site. The C-terminus of Get4 was fused to a tobacco etch virus TEV protease cleavage site followed by a hexa-histidine tag. The proteins were overexpressed in BL21(DE3) (Novagen) grown in 2xYT media for 3 h at 37 °C after induction with 0.3 mM IPTG. Cells were lysed using an M-110L pneumatic microfluidizer (Microfluidics) and purified as a complex by Ni-affinity chromatography (Qiagen). The affinity tag was removed by a 3 h TEV protease digest at room temperature while dialyzing against 20 mM Tris pH 8.0, 30 mM NaCl and 12 mM β-mercaptoethanol. The sample separated into two peaks using a 6 mL Resource Q anion exchange column (GE Healthcare). Each peak was further purified using a Superdex 200 16/60 size exclusion column (GE Healthcare) equilibrated with 5 mM Tris pH 8.0, 300 mM NaCl and 12 mM β-mercaptoethanol and concentrated to 10 mg⁄mL.

The *Aspergillus fumigatus* homologs of Get4 (NCBI sequence XP_747572) and Get5 (NCBI sequence XP_748165) were cloned, expressed, and purified in the same way as the *Saccharomyces cerevisiae* proteins. Get5 started from the second methionine in the Genbank sequence, omitting the first 42 annotated residues, as they were not consistent

with our multiple sequence alignments. The sequence used in this study is shown in Figure 1.1B. Expression and purification of Get3 were performed as previously described (Suloway et al., 2009).

Mutations and truncations of Get3, Get4, or Get5 were introduced using the Quikchange method (Stratagene). All mutants were expressed under identical conditions as the wild type Get4/5, purified by Ni-affinity chromatography and cleaved by TEV protease where indicated. Analytical size exclusion chromatography was performed using either a Superdex 200 10/300 column with a QuadTec UV detector (Biorad) or a Shodex KW 803 column with a Dawn Heleos MALS detector (Wyatt Technology).

Formation and detection of complexes between Get3 and Get4/5

Complexes were formed by incubating 800 pmol Get3 and 400 pmol Get4/5 for 2 h at room temperature in 200 μL of 50 mM HEPES pH 7.3, 100 mM NaCl, 2 mM MgCl2, 30 mM imidazole and, where indicated, 2 mM ADP or 2 mM ATP. After incubation, the reaction was added to 10 μL of gravity settled Ni-NTA agarose beads and mixed. The beads were washed twice with 100 μL of the incubation buffer within 1 minute. Bound proteins were eluted with 25 μL of incubation buffer containing 300 mM imidazole.

Limited proteolysis and crystallization of Get4/5

Crystallization screening was performed using a TTP LabTech Mosquito robot and commercially purchased kits (Hampton Research, Qiagen, Molecular Dimensions Limited). Adding TLCK treated α-chymotrypsin (Sigma) at a 1:125 mass ratio to either oligomeric state immediately prior to screening, as described (Dong et al., 2007), resulted in crystals growing as hexaganol rods within a day at room temperature. Limited

proteolysis was performed at room temperature for 15 or 60 minutes on ice and stopped by the addition of 0.1 mM PMSF. Six products were separated by a Resource Q column, and only the major product, Get4/5-N, crystallized after concentration to 5 mg⁄mL. Sequences of the proteolytic fragments were determined using LC/MS with tryptic digestion and/or N-terminal sequencing. After refinement of the condition, crystals approximately 150 μm in length grew after 3 days at room temperature, although the majority had a twist approximately halfway down their length and were unable to diffract (Figure 1.2D).

A small molecule additive screen (Hampton Research) revealed that the addition of spermidine aggravated twisting while 10 mM L-proline, trimethylamine-HCL or sarcosine yielded three-sided single crystals (Figure 1.2D). Diffracting crystals were grown using the hanging drop vapor diffusion method where 1 μL of Get4/5-N at 5 mg⁄mL in 20 mM Tris pH 8.0, 90 mM NaCl, and 12 mM β-mercaptoethanol was mixed with 1 μL of a reservoir of 17% PEG 6000, 0.14 M ammonium sulfate, 0.1 M Bis-tris pH 5.5, and 10 mM L-proline. Crystals were transferred into 20 μL of reservoir, cryoprotected by repeatedly removing and adding reservoir solution supplemented with 1% increments of glycerol to 10%, and flash frozen in liquid nitrogen. Selenomethionine-derived protein was expressed following established methods (Van Duyne et al., 1993) and purified, crystallized and frozen as described for the native protein.

Data collection, structure solution, and refinement

A single wavelength anomalous dataset to 2.8 Å resolution was collected on beam line 9-2 at the Stanford Synchrotron Radiation Laboratory (SSRL) (Table 1.1). Data were

integrated and scaled using XDS (Kabsch, 2010), the space group was determined by POINTLESS and the data were merged and converted to amplitudes with SCALA/CTRUNCATE (CCP4, 1994). The selenium substructure was determined using SHELXD (Sheldrick, 2008) and initial phases, density modification, and model building were performed by PHASER (McCoy et al., 2007) and RESOLVE (Terwilliger, 2000) as implemented automatically by PHENIX (Adams et al., 2010). The asymmetric unit consisted of three copies of Get4/5-N arranged along a distorted 3-fold axis. Chain E was the most complete copy of Get4, with continuous electron density observed for residues 9–299. Similarly, chain F was the most complete copy of Get5-N consisting of residues 3–56. Manual rebuilding was performed using COOT (Emsley et al., 2010). Difference densities in two copies near the N-terminal poles of α3 of Get4 (residues 49–52) that were too large for ordered waters were modeled as L-proline. Reciprocal space refinement with local NCS restraints was performed using REFMAC v5.6 (Winn et al., 2003) using the direct single wavelength anomalous dispersion (SAD) target (Skubak et al., 2004). A native dataset of comparable quality was collected on beam line 11-1 at SSRL, however refinement against the SAD data resulted in a higher quality model (determined by R-factors and model geometry) and thus the selenomethionine structure was used for deposition into the PDB (ID code 3LKU). The final model had an R_{crys} of 18.2% and an R_{free} of 22.4% with residues in the Ramachandran plot in 97.9% preferred, 2.1% allowed, and 0.0% in the disallowed and restricted regions [COOT (Emsley et al., 2010)]. Structure figures were prepared using PyMol (Delano, 1998) or UCSF Chimera (Pettersen et al., 2004; Sanner et al., 1996).

Yeast growth assays

Knockout strains BY4741 YOR164C::KanMX and BY4741 YOL111c::KanMX were purchased from America Type Culture Collection (ATCC) (Winzeler et al., 1999). The Get4 rescue plasmid was constructed by PCR amplifying the open reading frame with 339 bp upstream and 87 bp downstream flanking regions from BY4741 genomic DNA and inserting into the YEp-352 shuttle vector (Hill et al., 1986). Similarly, the Get5 open reading frame was amplified with 506 bp upstream and 69 bp downstream flanking sequence. Mutations were then introduced with the Quikchange method. The parent BY4741 and deletion strains were transformed using the LiAc/single-stranded carrier DNA/PEG method (Gietz and Schiestl, 2007). Phenotypic rescue was determined by growing each transformant in SC-Ura media at 30 °C to an OD600 nm between 1 and 2, diluting to 3.85×10^6 cells⁄mL and spotting 4 µL of serial dilutions onto SC -Ura agar plates in the presence or absence of 2 mM CuSO4. Plates were then incubated at 37 °C for 20 h and photographed.

Acknowledgements

We thank J. Howard, A. Müller, and A. Palazzo for discussion and critical comments on the manuscript, and T. Walton for help with MALS. We thank Gordon and Betty Moore for support of the Molecular Observatory at Caltech. All data collection was performed at beamline 9-2 and 11-1 at SSRL. Operations at Stanford Synchrotron Radiation Laboratory are supported by the US Department of Energy and the National Institutes of Health. W. M. C. is supported by the Searle Scholar program and a Burroughs–Wellcome Fund Career Award for the Biological Sciences.

Chapter 2

GET5 CARBOXYL-TERMINAL DOMAIN IS A NOVEL DIMERIZATION MOTIF THAT TETHERS AN EXTENDED GET4/GET5 COMPLEX

Abstract

Tail-anchored transmembrane proteins are targeted to membranes post-translationally. The proteins Get4 and Get5 form an obligate complex that catalyzes the transfer of tail-anchored proteins destined to the endoplasmic reticulum from Sgt2 to the cytosolic targeting factor Get3. Get5 forms a homodimer mediated by its carboxyl domain. We show here that a conserved motif exists within the carboxyl domain. A high-resolution crystal structure and solution NMR structures of this motif reveal a novel and stable helical dimerization domain. We additionally determined a solution NMR structure of a divergent fungal homolog, and comparison of these structures allows annotation of specific stabilizing interactions. Using solution X-ray scattering and the structures of all folded domains, we present a model of the full-length Get4/Get5 complex.

Adapted from

Chartron, J.W., VanderVelde, D.G., Rao, M., and Clemons, W.M., Jr. (2012). Get5 carboxyl-terminal domain is a novel dimerization motif that tethers an extended Get4/Get5 complex. J. Biol. Chem. *287*, 8310-8317.

Introduction

The targeted delivery of transmembrane proteins to the proper membrane is a critical cellular process. The signal recognition particle pathway delivers the majority of transmembrane proteins co-translationally in all organisms. Tail-anchored (TA) proteins are important exceptions to this pathway (Shan and Walter, 2005). TA proteins contain a single transmembrane helix within 30 residues of the carboxyl terminus, and this feature necessitates post-translational targeting. After insertion, the N terminus remains in the cytoplasm. TA proteins are found in all membranes exposed to the cytoplasm and have a wide variety of roles, such as vesicle fusion, regulating apoptosis, and protein translocation (Borgese et al., 2007; Kutay et al., 1993).

Eukaryotic pathways for TA protein delivery to the ER have been elucidated and are best described for yeast [for review, see (Rabu et al., 2009; Simpson et al., 2010)]. The majority of TA proteins are targeted via the GET (guided entry of TA proteins) pathway. Targeting progresses from the Get4/Get5/Sgt2 sorting complex that loads ER destined TA protein onto the ATPase Get3 (Battle et al., 2010; Chartron et al., 2010; Wang et al., 2010). Mitochondrial TA proteins appear to be initially retained on Sgt2-bound heat-shock cognate protein chaperones. Get3 then targets the TA protein to the ER membrane via Get1/Get2 (Schuldiner et al., 2008).

Get4 and Get5 form an obligate heterodimer mediated by the amino domain of Get5 (Get5-N) and the carboxyl domain of Get4 (Chang et al., 2010; Chartron et al., 2010). Get4 is an α-helical repeat protein that binds Get3 through a conserved basic face. Following the Get5 amino domain in sequence is a ubiquitin-like domain (Get5-Ubl) that

mediates interaction with Sgt2 (Chang et al., 2010; Chartron et al., 2011). The carboxyl domain of Get5 (Get5-C) is a homodimerization domain, resulting in a heterotetrameric Get4/Get5 complex (Chartron et al., 2010).

A similar pathway for TA targeting exists in mammals. TRC35 and Ubl4A, homologs of Get4 and Get5, respectively, form a stable complex with the protein Bag-6/Bat-3/Scythe (Mariappan et al., 2010). This Bag-6 complex is required for efficient TA protein targeting by transferring them to TRC40, the Get3 homolog, after synthesis is complete (Leznicki et al., 2010; Mariappan et al., 2010). It is also involved in the degradation of defective nascent polypeptides and the stabilization of hydrophobic segments of proteins retrotranslocated from the ER prior to degradation by the proteasome (Minami et al., 2010; Wang et al., 2011). The human homolog of Sgt2, SGTA, may also interact with this complex, suggesting a shared mechanism of TA sorting with yeast (Hegde and Keenan, 2011; Winnefeld et al., 2006).

Structural studies of GET pathway members continue to provide details on the molecular series of events that occur in TA targeting. To understand the basis for specificity of homodimerization by Get5, we determined both a crystal and solution structure of the carboxyl domain of Get5 along with the solution structure of a fungal homolog. These structures reveal the nature of the conserved dimerization motif. We characterize the oligomeric state of human Ubl4A, which is amendable to the alternate architecture of the mammalian complex. Moreover, we use solution small angle X-ray scattering (SAXS) to define the overall structure of the full Get4/Get5 heterotetramer, providing the first molecular framework of this complex.

Results

Conservation of the Get5 Carboxyl Domain

Get5 contains an amino domain, a ubiquitin-like domain, and a carboxyl domain that mediates homodimerization (Figure 2.1A) (Chartron et al., 2010). Get5-C, 152–212, contains the entire sequence from the end of the Get5-Ubl domain to the carboxyl terminus. Residues 152–176 have poor overall conservation in sequence identity or length. Residues 177–212 form an ~35-residue conserved motif that is found in Get5 homologs from two of the three subphyla of Ascomycota, the largest described fungal phylum (Figure 2.2) (James et al., 2006). The motif has greater variability within Saccharomycotina than in Pezizomycotina. Taphrinomycotina, the third subphylum that includes *Schizosaccharomyces pombe*, does not appear to have homologs of Get5 that contain a carboxyl-terminal domain.

Ubl4A lacks the amino Get4 binding domain of Get5, with the Ubl domain as the amino terminus (Chartron et al., 2010). Ubl4A is well conserved in vertebrates. The carboxyl domain of Ubl4A contains the 35-residue dimerization motif seen in Get5 (Figure 2.2). The linker between the dimerization motif and the Ubl domain is generally shorter than in fungal homologs, and Ubl4A has an additional 30 conserved residues following the motif.

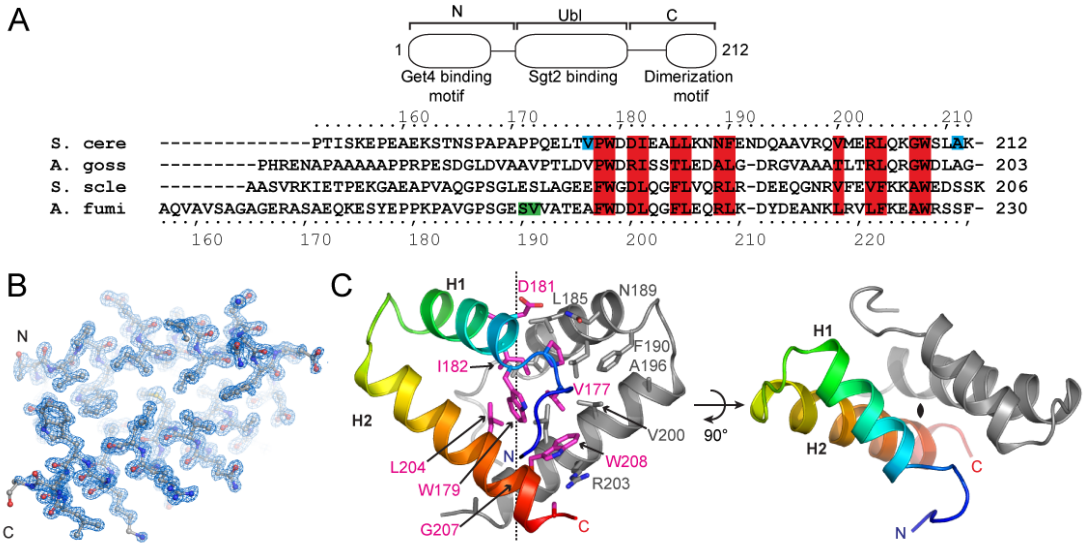
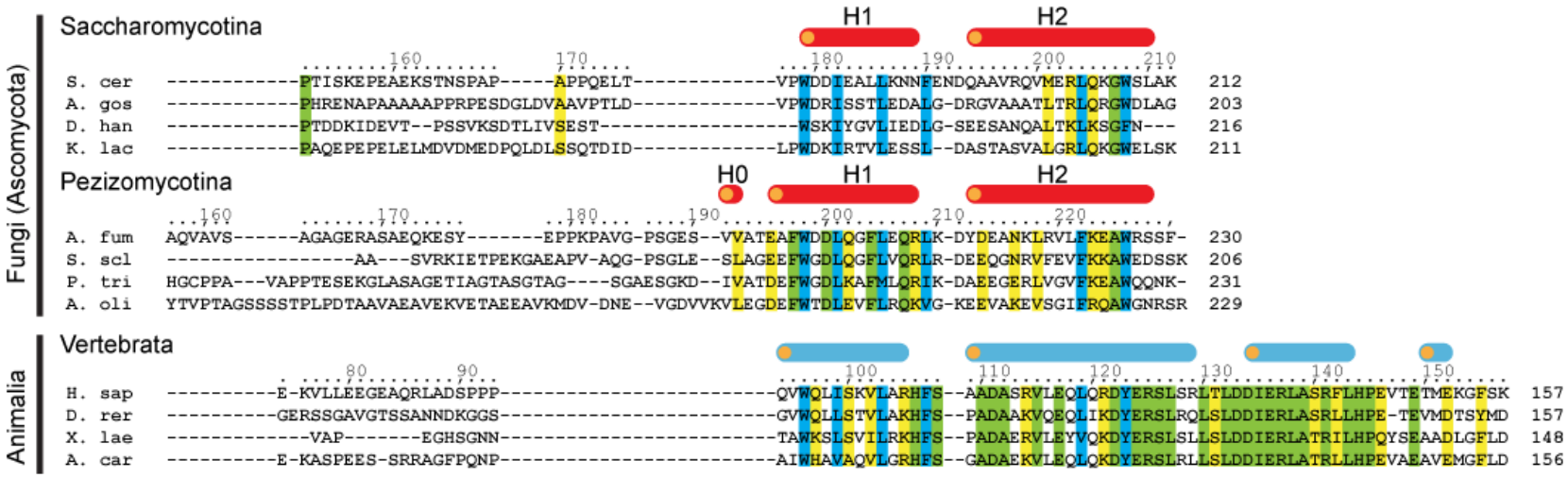


Figure 2.1. Crystal structure of Get5-C. (A) schematic of the domain organization of Get5 and sequence alignment of the carboxyl domain in fungi. Lines indicate flexible regions. Sequences are: S. cere, *S. cerevisiae*; A. goss, *Ashbya gossypii*; S. scle, *Sclerotina sclerotiorum*; A. fumi, *A. fumigatus*. Residues that mediate intermolecular contacts in both Get5 and AfGet5 dimers are highlighted in red, and residues specific to Get5 or AfGet5 are highlighted in blue or green, respectively. (B) Dimerization interface of a monomer of Get5-C with σ_a-weighted $2|F_o|-|F_c|$ electron density contoured at 1.5σ. (C) asymmetric unit of Get5-C. One monomer is color ramped from amino (blue) to carboxyl (red) terminus. The 2-fold axis is indicated with a dotted line. Side chains that make intermolecular contacts are shown (left).


Fungi (Ascomycota)
Saccharomycotina
H1
H2
S. cer PTISKEPEAEKSTNSPAP-----APPQELT----------VPWDDIEALLKNNFENDQAAVRQVMERLQKGWSLAK 212
A. gos PHRENAPAAAAAPPRPESDGLDVAAVPTLD----------VPWDRISSTLEDALG-DRGVAAATLTRLQRGWDLAG 203
D. han PTDDKIDEVT--PSSVKSDTLIVSEST----------WSKIYGVLIEDLG-SEESANQALTKLKSGFN--- 216
K. lac PAQEPEPELELMDVDMEDPQLDLSSQTDID----------LPWDKIRTVLESSL--DASTASVALGRLQKGWELSK 211
Pezizomycotina
H0
H1
H2
A. fum AQVAVS------AGAGERASAEQKESY-------EPPKPAVG-PSGES--VVATEAFWDDLQGFLEQRLK-DYDEANKLRVLFKEAWRSSF- 230
S. scl ----------AA---SVRKIETPEKGAEAPV-AQG-PSGLE--SLAGEEFWGDLQGFLVQRLR-DEEQGNRVFEVFKKAWEDSSK 206
P. tri HGCPPA---VAPPTESEKGLASAGETIAGTASGTAG----SGAESGKD--IVATDEFWGDLKAFMLQRIK-DAEEGERLVGVFKEAWQQNK- 231
A. oli YTVPTAGSSSSTPLPDTAAVAEAVEKVETAEEAVKMDV-DNE--VGDVVKVLEGDEFWTDLEVFLRQKVG-KEEVAKEVSGIFRQAWGNRSR 229
Animalia
Vertebrata
H. sap ----------E-KVLLEEGEAQRLADSPPP----------QVWQLISKVLARHFS--AADASRVLEQLQRDYERSLSRLTLDDIERLASRFLHPEVTETMEKGFSK 157
D. rer ----------GERSSGAVGTSSANNDKGGS----------GVWQLLSTVLAKHFS--PADAAKVQEQLIKDYERSLRQLSLDDIERLASRLLHPE-TEVMDTSYMD 157
X. lae ----------VAP-------EGHSGNN----------TAWKSLSVILRKHFS--PADAERVLEYVQKDYERSLSLLSLDDIERLATRILHPQYSEAADLGFLD 148
A. car ----------E-KASPEES-SRRAGFPQNP----------AIWHAVAQVLGRHFS--GADAEKVLEQLQKDYERSLRLLSLDDIERLATRLLHPEVAEAVEMGFLD 156

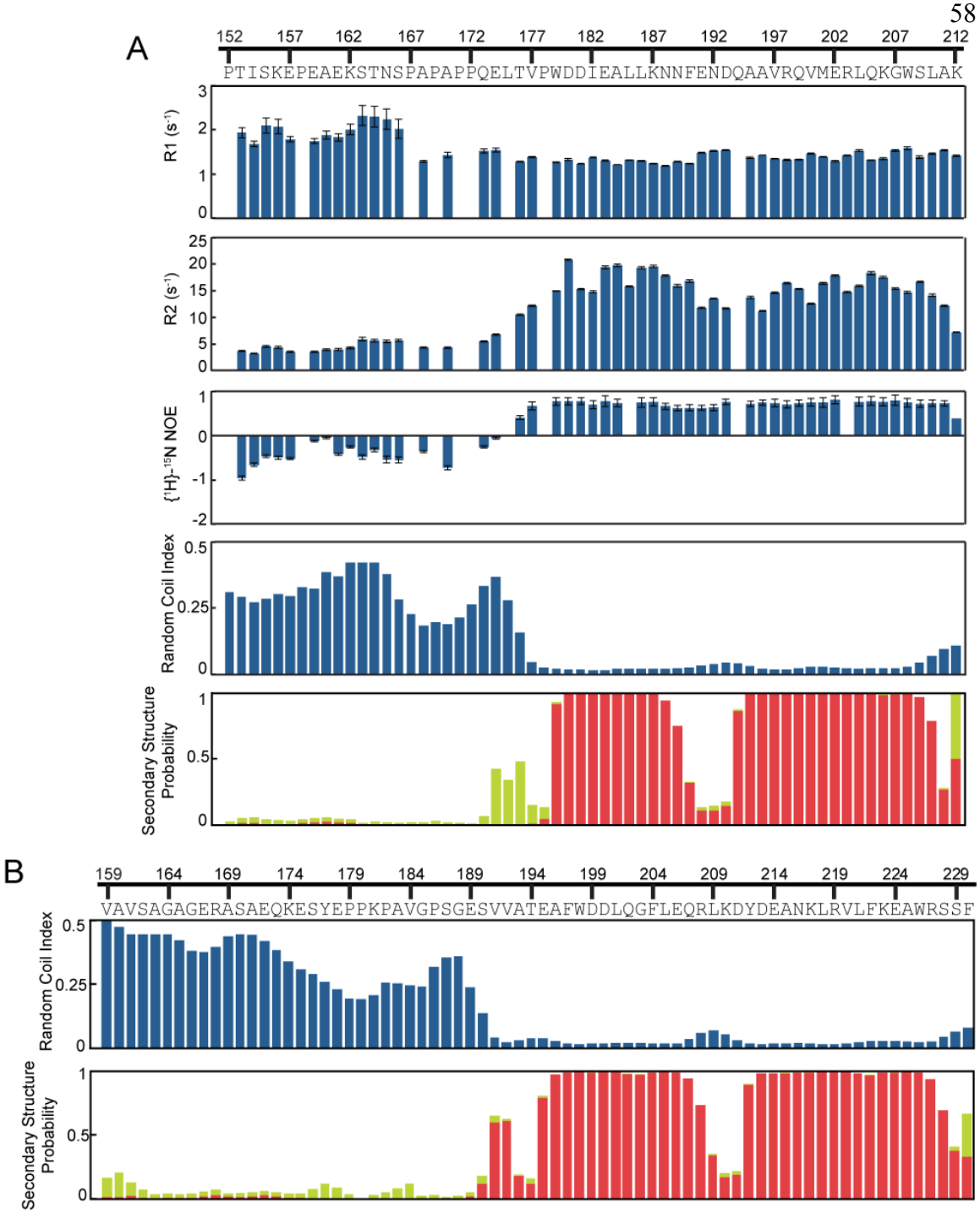
Structure of Get5 Dimerization Motif

We cloned, expressed, and purified Get5-C for NMR investigation. The main chain chemical shifts ($^{1}H^{N}$, $^{15}N^{N}$, $^{13}C^{CO}$, $^{13}C^{\alpha}$, $^{13}C^{\beta}$, and $^{1}H^{\alpha}$) of residues 152–175 have random coil character, whereas those of 179–190 and 194–210 are characteristic of helices (Figure 2.3). These helices are designated H1 and H2, respectively. Relaxation rates and heteronuclear NOE values also indicate rapid motions for residues 152–175. Consistent with this secondary structure, $^{1}H^{N}$ of residues 182–187 and 200–208 were the most protected from solvent deuterium exchange whereas residues 152–177 readily exchanged. Therefore, we concluded that residues 152–175 were unstructured and not significantly contributing to the stability of the folded dimerization domain.

Based on these results a truncated version of the carboxyl domain consisting of residues 175–212 was generated for crystallographic studies. This variant crystallized in space group $P2_1$ and diffracted to 1.23 Å resolution. Experimental phases were determined by the single-wavelength anomalous diffraction technique using an iodide-soaked crystal

---

Figure 2.2. Conservation of the dimerization motif within Get5. The sequences shown are from the end of the ubiquitin-like domain to the carboxyl terminus of each homolog. Secondary structures of *S. cerevisiae* Get5 and the *A. fumigatus* homolog are shown as red helices above their respective sequences. Secondary structure for H. sapiens Ubl4A is predicted using JPRED-3 (Cole et al., 2008) and is shown as blue helices above the sequence. Residue numbers above the sequences are from *S. cerevisiae*, *A. fumigatus* or *H. sapiens*. Conserved or identical residues within each subphylum are highlighted in yellow or green, respectively. Residues that are conserved across all sequences are highlighted in blue. Sequences are: S. cer, *Saccharomyces cerevisiae*; A. gos, *Ashbya gossypii*; D. han, *Debaryomyces hansenii*; K. lac, *Kluyveromyces lactis*; A. fum, *Aspergillus fumigatus*; S. scl, *Sclerotinia sclerotiorum*; P. tri, *Pyrenophora triticirepentis*; A. oli, *Arthrobotrys oligospora*; H. sap, *Homo sapiens*; D. rer, *Danio rerio*; X. lae, *Xenopus laevis*; A. car, *Anolis carolinensis*.


A
152 157 162 167 172 177 182 187 192 197 202 207 212
PTISKEPEAEKSTNSPAPAPPQELTVPWDDIEALLKNNFENDQAAVRQVMERLQKGWSLAK
R1 (s-1)
R2 (s-1)
{1H}-15N NOE
Random Coil Index
Secondary Structure Probability
B
159 164 169 174 179 184 189 194 199 204 209 214 219 224 229
VAVSAGAGERASAEQKESYEPPKPAVGPSGESVVATEAFWDDLQGFLEQRLKDYDEANKLRVLFKEAWRSSF
Random Coil Index
Secondary Structure Probability

(Dauter et al., 2000). Two copies of Get5-C are present in the asymmetric unit of the native crystal. Unambiguous electron density allowed modeling of residues 175–212 of one copy and 175–211 of the second (Figure 2.1B). The structure was refined to an R_{free} of 0.204. Crystallographic statistics are presented in Table 2.1.

The two copies of Get5 in the asymmetric unit interact extensively, burying 870 $Å^2$ of solvent-accessible surface area (~25% of total) per monomer and sequester the majority of hydrophobic residues (Figure 2.1C). In contrast, the most extensive crystallographic contact buries 410 $Å^2$; therefore, the asymmetric unit contains the physiologically relevant dimer.

A 2-fold axis relates one monomer onto the other (Figure 2.1C). The H1 helices pack antiparallel to one another. A four-residue linker connects to H2, and the H2 helices cross at conserved Gly-207 forming an ~95° angle. The association between the two subunits in the dimer is almost entirely by hydrophobic interactions. The dimer axis is lined by Ile-182, Leu-185, Leu-204, and Gly-207. Leu-185 and Pro-178 extend from H1 of one copy into hydrophobic pockets formed by additional H1 residues the second copy. Asp-181 and Asn-189 form the only intermolecular hydrogen bond across the dimer.

Figure 2.3. Solution NMR analysis of Get5-C. ^{15}N R1 and R2 relaxation rates, {1H}-15N heteronuclear NOE, random coil index (Berjanskii and Wishart, 2007) and secondary structure probability determined from main-chain chemical shifts using TALOS+ (Shen et al., 2009). plotted against amino acid sequence for Get5-C (A) or AfGet5-C (B). The greater R1 and lower R2 values of residues 152-175 of Get5-C mean this region is tumbling more rapidly than 176-212. Negative {1H}-15N heteronuclear NOE values for residues 152-175 indicate rapid (picosecond to nanosecond) motions.

Table 2.1. Crystallographic data and model refinement statistics

Dataset	Native	Iodide
Data collection	SSRL BL12-2	Micromax-007 HF
Wavelength (Å)	1.0000	1.5418
Resolution range (Å)	28.08–1.23 (1.26–1.23)[a]	26.52–1.60 (1.69–1.60)
Space Group	P 2_1	P $2\ 2_1\ 2$
Cell Parameters		
a, b, c (Å)	25.02, 46.34, 28.27	25.10, 26.52, 46.17
α, β, γ (°)	90.00, 96.90, 90.00	90.00, 90.00, 90.00
Unique reflections	18288	4395
Completeness (%)	97.9(94.6)	99.6(97.5)
Redundancy	4.2(3.9)	12.2(11.2)
R_{merge}[b]	0.058(0.609)	0.088(0.605)
Mean I/σ(I)	11.6(2.1)	19.5(3.9)
Refinement		
Reflections: work/free[c]	17185/1089	
R_{work}/R_{free}[d]	0.177/0.204	
Number of atoms		
Protein	714	
Water	86	
Phosphate	5	
Average B-factors (Å^2)		
Protein	13.6	
Water	24.8	
Phosphate	33.9	
r.m.s.d.		
Bond lengths (Å)	0.010	
Bond angles (°)	1.082	
Ramachandran favored[e]	100%	

[a]Values in parentheses are for the highest resolution shell

[b] $R_{merge} = \Sigma_{hkl} \Sigma_i |I_i(hkl) - \langle I(hkl) \rangle| / \Sigma_{hkl} \Sigma_i I_i(hkl)$, where $I_i(hkl)$ is the *i*th observation of reflection hkl and $\langle I(hkl) \rangle$ is the weighted average intensity for all observations *i* of reflection hkl.

[c]The free set represents a random 6% of reflections not included in refinement

[d] $R = \Sigma_{hkl}(||F_{obs}| - |F_{calc}||)/\Sigma_{hkl}|F_{obs}|$, where $|F_{obs}|$ and $|F_{calc}|$ are the observed and calculated structure factor amplitudes, respectively.

[e]Percentage of residues in Ramachandran plot regions were determined using PROCHECK (Laskowski et al., 1993).

There are several interactions that restrict the conformation of H1 relative to H2 within each monomer. Phe-190, which begins the connecting loop, fits in a pocket formed by Leu-186, Ala-196, and Val-200 of the same copy (Figure 2.1C). The indole ring of Trp-179 is normal to the ring of Trp-208 of the same copy, and the side chain of Val-177 fits into the resulting pocket. Trp-208 points toward the dimer interface, where the side chain of Arg-203 extends from the opposite copy to form a cation–pi interaction. Arg-203 is conserved as an arginine or lysine in Saccharomycotina, and Trp-208 is conserved across all eukaryotes (Figure 2.2). The side chain of Trp-208 also contacts Val-200 and Leu-204 of the opposite copy.

Solution NMR Structure of Get5-C

We determined the solution structure of Get5-C using NOE-derived and hydrogen bond distance restraints, ϕ and ψ dihedral angle restraints, and residual dipolar couplings restraints. Residues 152–172 did not converge to a consistent structure due to a lack of interresidue NOE-derived distance restraints; therefore, they were omitted in the final model calculations. Statistics for the structure and restraints are summarized in Table 2.2, and the ensembles of the 10 lowest energy structures are shown in Figure 2.4A.

The solution structure is very similar to the crystal structure, with an average backbone root mean square deviation of 0.94 ± 0.06 Å across residues 177–212 (Figure 2.5A). The interaction between Arg-203 and Trp-208 and the interactions between the methyl groups of Val-177 with Trp-179 and Trp-208 are maintained in solution (Figure 2.5B). The unique environment created by the two aromatic residues is reflected in unusual upfield chemical shifts of interacting residues, with Val-177 methyl protons at −0.994 and

Table 2.2. NMR structural constraints and structure statistics for the 10 lowest energy structures

	Get5-C[a]	*Af*Get5-C[a]
No. of restraints		
NOE-based distance restraints[b]		
Intraresidue ($\|i - j\| = 0$)	644	598
Sequential ($\|i - j\| = 1$)	396	334
Medium range ($2 \leq \|i - j\| < 5$)	562	454
Long range ($\|i - j\| \geq 5$)	178	158
Intermolecular	276	140
Ambiguous	824	460
Total	2880	2144
Dihedral angle restraints (φ/ψ)	132 (66/66)	152 (76/76)
Hydrogen bond restraints	96	96
Residual dipolar coupling restraints	64	70
Restraints statistics		
r.m.s.d. of distance violations		
NOE restraints	0.11 ± 0.02 Å	0.07 ± 0.01 Å
H-bond restraints	0.009 ± 0.001Å	0.009 ± 0.002 Å
r.m.s.d. of dihedral violations	1.18 ± 0.06°	0.61 ± 0.05°
RDC Q-factors	0.0715 ± 0.004	0.05 ± 0.02
Coordinate precision r.m.s.d.		
Backbone	0.24 ± 0.06 Å	0.37 ± 0.07 Å
Heavy atom	0.60 ± 0.06 Å	0.69 ± 0.04 Å
Structural quality		
Ramachandran statistics[c]		
Most favored regions	84.70%	93.90%
Allowed regions	9.30%	4.64%
Generously allowed regions	6.00%	0.96%
Disallowed regions	0.00%	0.48%
WHAT-IF Z-score[d]		
Backbone conformation	0.775 ± 0.250	0.537 ± 0.580
2nd generation packing quality	7.644 ± 2.380	6.273 ± 2.062
Ramachandran plot appearance	−2.170 ± 0.359	−0.388 ± 0.339
χ1/χ2 rotamer normality	0.514 ± 0.488	−1.883 ± 0.528

[a]Statistics are reported for dimers of Get5-C residues 173–212 and *Af*Get5-C residues *186–230*.
[b]A single NOE generates two restraints for atom pairs from each monomer.
[c]Percentage of residues in Ramachandran plot regions were determined using PROCHECK (Laskowski et al., 1993).
[d]Z-scores were calculated by WHAT-IF(Vriend and Sander, 1993).

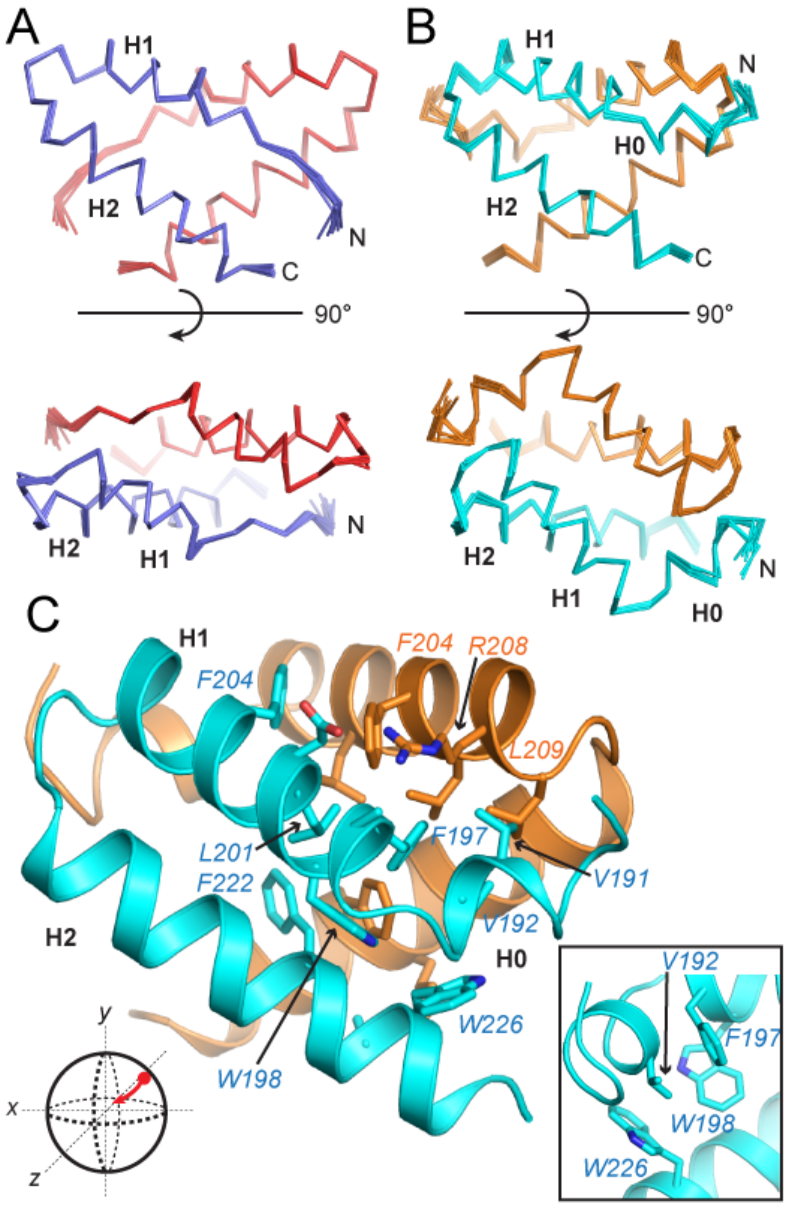
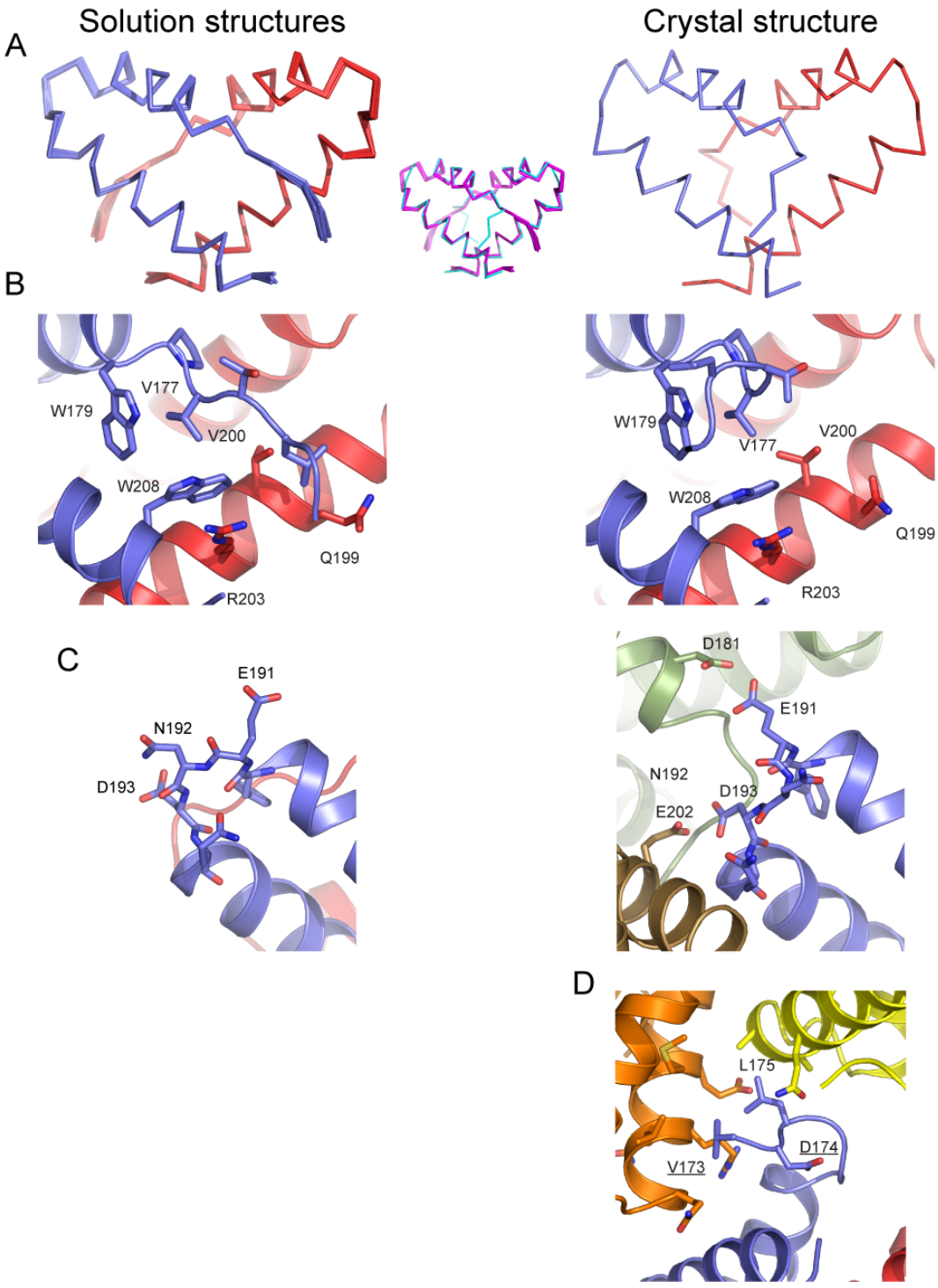


Figure 2.4. Solution structures of Get5-C and AfGet5-C. (A) and (B), ensembles of the 10 lowest energy solution structures of the ordered regions of Get5-C (A) and *Af*Get5-C (B). (C) ribbon diagram of *Af*Get5-C with side chains making intermolecular contacts. *Af*Get5-C is rotated from the orientation in the top of (B) as indicated by the sphere in the corner. *Inset*, view of the environment around Val-192.


Solution structures
Crystal structure
A
B
W179
V177
V200
W208
Q199
R203
W179
V177
V200
W208
Q199
R203
C
E191
N192
D193
D181
E191
N192
D193
E202
D
L175
D174
V173

−0.613 ppm, Arg-205 $^{1}H^{\delta}$ at 1.573 ppm and Gln-205 $^{1}H^{\gamma}$, and a side chain amide proton at −0.056 ppm and 4.656 ppm, respectively.

Two significant differences exist between the solution and crystal structures, and both can be explained by the crystal lattice. First, the loop connecting H1 and H2 is rearranged (Figure 2.5C). For the X-ray structure, crystals grew in pH 5.0, and this pH allows Glu-191 and Asp-193 to form hydrogen bonds with Asp-181 and Glu-202, respectively, from two different crystallographic copies. This results in an alternate conformation of the loop, predominantly by a rotation of the peptide bond plane between Glu-191 and Gln-192. The second difference is at the N terminus of the crystal structure, which has a three-residue cloning artifact prior to residue 175 of the crystallization construct. Leu-175 and nonnative Val-173 form hydrophobic contacts with two different crystallographic copies (Figure 2.5D), resulting in the terminus turning outward from the helices. In solution, residues 173–176 are extended (Figure 2.5A). This full domain results in 1090 $Å^2$ of solvent-accessible surface area buried per monomer (~30% of total).

Figure 2.5. Comparison of solution and crystal structures of Get5-C (A) *Left,* the ensemble of the 10 lowest energy solution structures of the ordered region of Get5-C. *Right,* the crystal structure of Get5-C. *Middle,* overlay of solution (magenta) and crystal (cyan) structures. (B) The arrangement of Val-177, Trp-179, Trp-208 and Arg-203 in the solution (left) and crystal (right) structures. For the solution structure a representative model is displayed. (C) The loop connecting H1 and H2 in the solution (left) and crystal (right) structures. Two crystallographic copies of the Get5 dimer are colored brown and light green. (D) The amino terminus of the Get5-C crystal structure with two crystallographic copies of the dimer colored orange and yellow. Val-173 and Asp-174, underlined, are cloning artifacts and are Gln-173 and Glu-174 in the wild type protein.

Structure of Get5-C Domain Homolog

We additionally investigated the carboxyl domain of the Get5 homolog from the filamentous fungi *A. fumigatus* (*Af*Get5-C). Several structures of other GET pathway members have been determined for this organism (Chartron et al., 2011; Suloway et al., 2009). *A. fumigatus* is in the subphylum Pezizomycotina, and Get5 homologs from this group have conserved features distinct from Saccharomycotina (Figure 2.2). There is more conservation in the residues immediately preceding the tryptophan at the amino terminus of H1, and there are three additional conserved phenylalanines. We cloned, expressed, and purified *Af*Get5-C residues *159–230*, comprising the full sequence of the carboxyl domain for NMR investigation.

Main chain chemical shifts indicated that residues *159–186* are likely random coil whereas *212–228* and *195–208* formed helices corresponding to H1 and H2 (Figure 2.3B). An additional short coil is predicted at residues *Ser-190* and *Val-191*, designated H0. We determined the solution structure of *Af*Get5-C. Residues *159–185* did not converge to a consistent structure due to a lack of interresidue NOE-derived distance restraints and were omitted from the final structure calculation. Structure and restraint statistics are provided in Table 2.2, and the ensembles of the 10 lowest energy structures are shown in Figure 2.4B.

The overall structure of *Af*Get5-C is similar to the Get5-C (Figure 2.4B,C). However, *Af*Get5-C has an additional turn at the amino terminus of H1, comprising residues *Glu-195*, *Ala-196*, and Pezizomycotina-specific *Phe-197* (Figure 2.2). Coil H0 caps the

helical bundle on either end, and *Val-191* and *Val-192* extend into the core. The two other conserved phenylalanines, *Phe-204*, and *Phe-222*, line the symmetry axis within the dimer along with *Leu-201* and *Ala-225*. The dimer interface is homologous to Get5-C, with *Phe-204* forming interactions that are equivalent to those of Leu-185. *Val-191* and *Phe-197* together occupy a position similar to Pro-178. There is a single intermolecular electrostatic interaction made by *Arg-208* and *Asp-200*, similar to the hydrogen bond between Asp-181 and Asn-189. The H2 helices cross at an ~110° angle and cross between *Phe-222* and *Ala-225*. This reconfiguration accommodates the aromatic rings of *Phe-204* and *Phe-222* and the side chains extending from H0 in the core.

*Af*Get5-C has an arrangement of tryptophans that is distinct from the Val-177/Trp-179/Trp-208 interaction of Get5 (Figure 2.4C, inset). The extra turn in H1 positions the aromatic ring of *Phe-197* normal to the ring of *Trp-198* and results in upfield shifts of 5.799 and 5.919 ppm for $^{1}H^{\delta}$ and $^{1}H^{\epsilon}$, respectively, of *Phe-197*. The side chain of *Val-192* occupies a pocket formed by *Phe-197*, *Trp-198*, and *Trp-226*. *Af*Get5-C does not have an intermolecular cation-pi interaction, and the side chain of *Trp-226* is flipped relative to the equivalently positioned Trp-208; the increased distance of the main chains between the two copies prevents any extended side chain interaction to this position. Arg-203 is not conserved in Pezizomycotina.

Get5 Carboxyl Domain Is Highly Stable Dimer

The thermal stabilities of Get5-C and *Af*Get5-C were determined by circular dichroism (CD) spectroscopy (Figure 2.6). Get5-C has a thermal melting curve with a single transition from folded to unfolded states. The midpoint of this transition, T_m, is ~74°C.

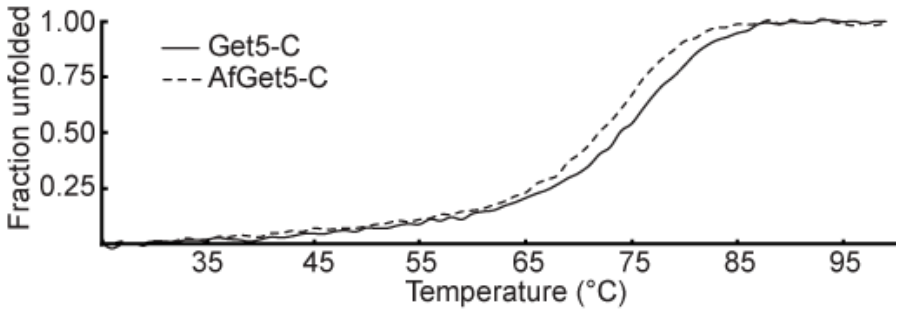


Figure 2.6. Get5-C is a stable dimer. Thermal melting curves of Get5-C (solid line) and AfGet5-C (dashed line) measured by CD spectroscopy.

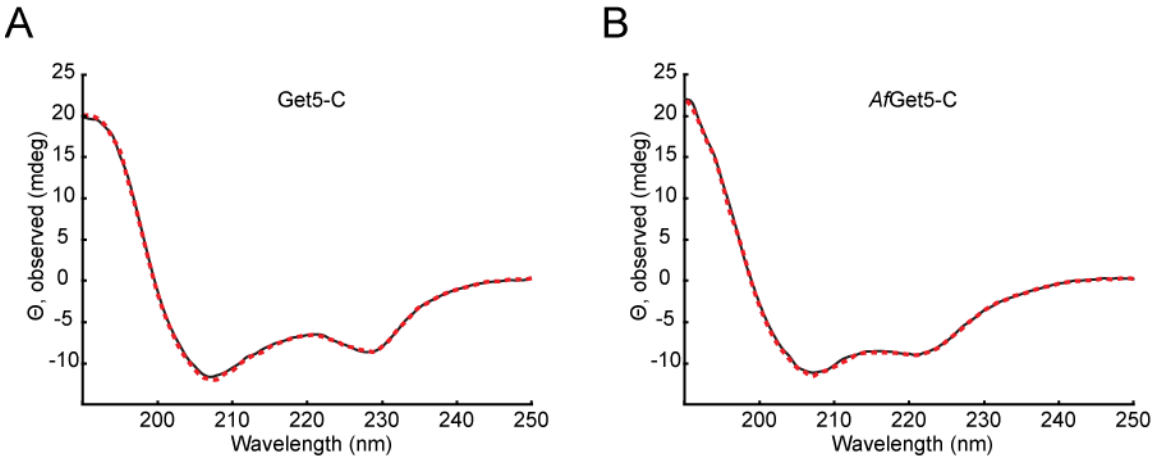


Figure 2.7. Refolding of Get5-C and AfGet5-C. (A) and (B). Circular dichroism spectra recorded at 25°C before (solid, black) and after (dashed, red) complete denaturation at 99°C for Get5-C (A) and AfGet5-C (B). In both proteins, the spectra are nearly completely overlapping indicating complete refolding.

*Af*Get5-C has a surprisingly similar thermal melting curve and a T_m of 72 °C. After heating to 99 °C, both proteins can refold completely (Figure 2.7).

Mammalian Get5 Homolog Ubl4A Does Not Homodimerize

We expressed and purified recombinant human Ubl4A from *E. coli*. Ubl4A, with a molecular mass of 20.0 kDa, elutes later from a size exclusion column than a dimeric Get5-Ubl-C construct that is lacking the amino-terminal Get4 binding domain (15.9 kDa), and slightly earlier than the monomeric Get5-Ubl domain alone (10.0 kDa) (Figure 2.8A). Surprisingly, these data are most consistent with Ubl4A existing as a monomer in solution. Alone, Ubl4A is unstable and precipitates over several days when stored at 4 °C.

It has been demonstrated previously that Get5 dimers exchange monomers within hours at room temperature (Chartron et al., 2010). In that experiment, we were unable to demonstrate the formation of a heterodimer between Get5 and *Af*Get5. We tested whether Ubl4A can form heterodimeric complexes with either homolog (Figure 2.8B). A heterodimer could be detected with Get5 but not *Af*Get5, suggesting that the Ubl4A dimerization domain is most similar to Get5-C.

Get4/Get5 Complex Is Extended in Solution

SAXS can be used to generate low-resolution models of protein complexes in solution when high-resolution structures of individual domains are available. We collected SAXS data on the full-length Get4/Get5 heterotetramer, which has a molecular mass of 120 kDa. The pair–distance distribution function obtained by an indirect Fourier

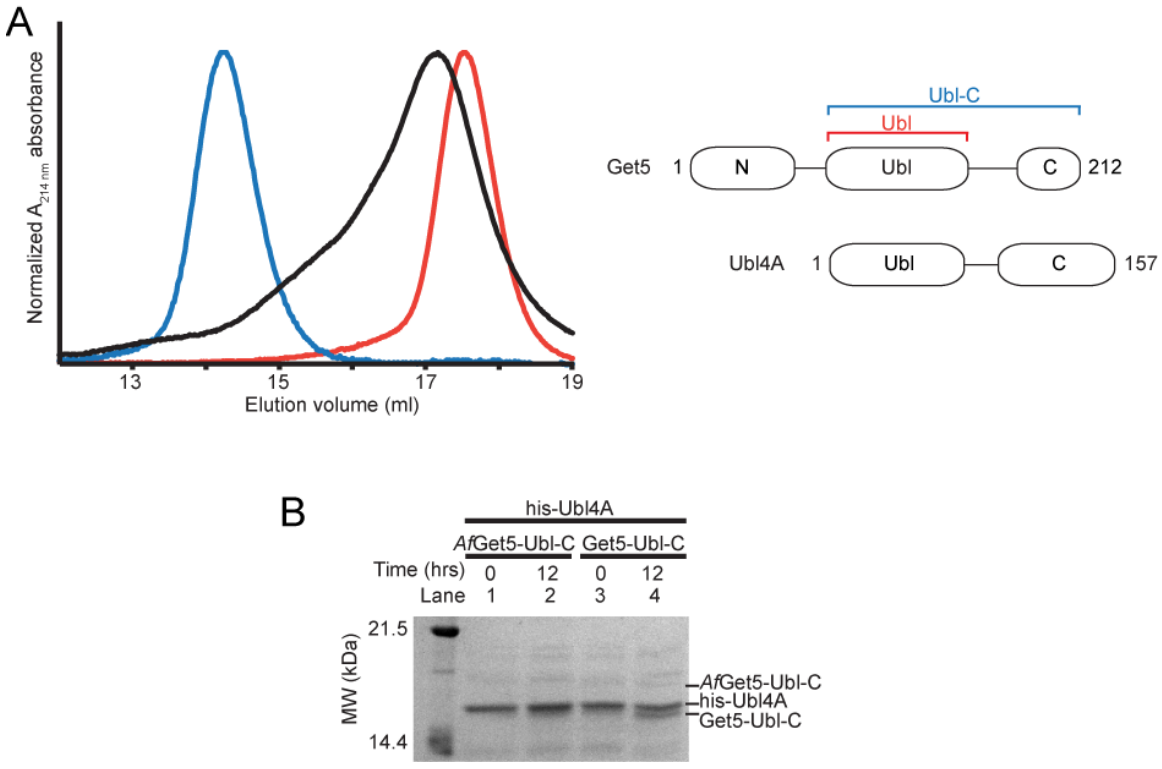


Figure 2.8. Ubl4A can form heterodimers. (A) Size exclusion chromatography of Ubl4A (black), Get5- Ubl (red) and Get5-Ubl-C (blue) using a Superdex 200 10/300 column. Absorbances are normalized against peak values. Cartoons of each construct are shown as in Figure 2.1A. (B) Polyhistidine tagged Ubl4A incubated with either AfGet5-Ubl-C or Get5-Ubl-C for the indicated time, precipitated with Ni-NTA agarose beads, eluted and separated by SDS-PAGE.

transform, $P(r)$, is characteristic of an elongated particle (Mertens and Svergun, 2010) as there is an asymmetric peak that decreases to a large distance in a linear fashion (Figure 2.9A). Compact particles have a more symmetric, parabolic peak. The radius of gyration and maximum particle diameter obtained from this analysis are 68.1 ± 0.1 Å and 240 Å, respectively. These values are consistent with the observation that Get4/Get5 appears atypically large by size exclusion chromatography (Chartron et al., 2010). An *ab initio* reconstruction using the indirect Fourier transform generates the surface shown in Figure 2.9B. This surface is the average of 10 individual models and represents the most probable volume.

With high-resolution structures now available for each ordered domain of the Get4/Get5 complex, rigid body fitting against the SAXS data can generate unbiased models of the heterotetramer in solution. The crystal structure of Get4 and Get5-N (Protein Data Base ID 3LKU), a homology model of Get5-Ubl generated from the Ubl domain of Ubl4A (Protein Data Base ID 2DZI) (Chartron et al., 2010), and the solution structure of Get5-C were used for fitting by the program CORAL. This program models flexible termini and interdomain linkers as chains of dummy residues (Petoukhov and Svergun, 2005), restraining the possible orientations of the rigid domains fit using simulated annealing. We performed the calculation eight times. The model with the best fit to the data (Figure 2.9C, *top*) is in agreement with the dimensions of the averaged *ab initio* model (Figure 2.9B). In all trials there is a similar extended spatial arrangement of the ordered domains, although there is variation in the overall rotations of the Get5-Ubl domain and Get4/Get5-N relative to Get5-C as well as in the angle that Get4 makes with the long

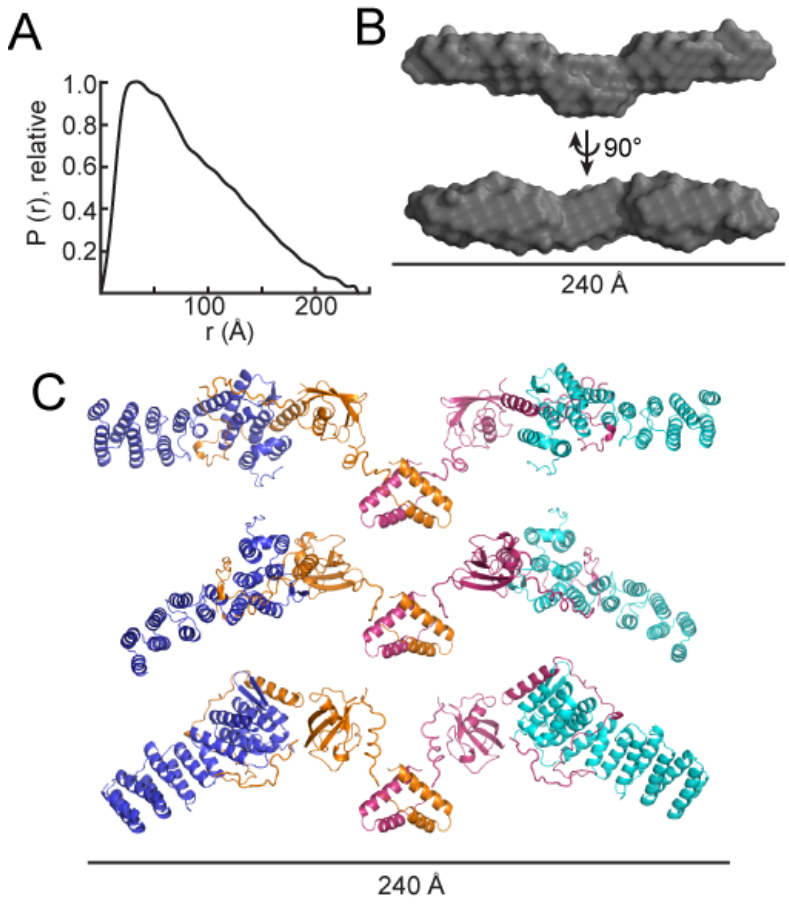


Figure 2.9. SAXS of Get4/Get5. (A) Pair–distance distribution function of the Get4/Get5 heterotetramer derived from SAXS. (B) Averaged *ab initio* reconstruction of Get4/Get5 shown as a gray surface in two orientations. (C) Rigid body models generated by independent simulated annealing calculations using known structures against SAXS data. In each model, the two copies of Get5 are colored magenta and orange, and the two copies of Get4 are colored cyan and blue. The three top models are shown ranked according to χ^2 fit to the experimental SAXS curve.

axis of the particle (Figure 2.9C). The *ab initio* model likely reflects an average of these orientations rather than a single species.

Discussion

Get5 homodimerization is mediated by an ~35-residue motif within its carboxyl domain. Two copies of this motif form a domain with a novel arrangement of four α-helices. The sequence of the motif has some similarity to coiled-coils (Chartron et al., 2010); however, rather than arranged in a zipper like fashion, hydrophobic residues are buried within a small core. Elements of the motif are conserved across the eukaryotic kingdom, such as the positions of hydrophobic residues that mediate the dimer interface and the aromatic residues toward the beginning and end of the motif (Figure 2.2). Of 13 positions that are at the dimer interface, 6 are conserved across eukaryotes (Figure 2.1A and Figure 2.2). There is extensive divergence of the remaining sequence, especially at surface positions, yet an overall conservation of structure as well as thermal stability.

The overall solution data of the Get4/Get5 complex points to an extended conformation in solution. With this architecture, the two copies of the amino-terminal face of Get4, which binds Get3, point outward at opposite ends of the particle. Get5-C presumably plays a structural role in this by orienting the Ubl domain that in turn orients Get4.

Ubl4A is a component of the Bag-6 complex along with Bag-6/Bat-3/Scythe and TRC35. This complex is involved in directing TA protein biogenesis (Leznicki et al., 2010; Mariappan et al., 2010) and has more recently been implicated in the stabilization of proteins retrotranslocated from the ER prior to degradation (Wang et al., 2011). TRC35 is

believed to mask a nuclear localization signal within Bag-6 and modulates the population of cytoplasmic and nuclear resident protein (Wang et al., 2011); however, nothing is known regarding the physical interactions of components of the complex. Purified Ubl4A is an unstable monomer but can form an artificial heterodimer with Get5-Ubl-C. This suggests that the carboxyl domain of Ubl4A has the conserved, exposed protein-protein interaction face and may mediate its interaction with the Bag-6 complex. We were unable to identify a feature in Bag-6 that contained the Get5-C motif; therefore, one might expect a unique heterodimeric interaction. The architecture of the Bag-6 complex is distinct from the Get4/Get5/Sgt2 complex. Certain steps, such as the recognition of TA proteins by SGTA, the binding of SGTA to the Ubl domain of Ubl4A, and the subsequent loading of TRC40 are likely mechanistically analogous with yeast. The rearrangement to include Bag-6/Bat-3/Scythe could allow interplay between TA targeting pathways, protein degradation and apoptosis.

Oligomerization in biology plays an important role, and there are many examples of dimerization domains; however, there are few examples of small (less than 70 residues) globular dimerization motifs. We searched for other examples of helix-turn-helix homodimers similar to Get5-C in that they are neither intercalated nor have extensive coiled-coils. There are four proteins in this category with high-resolution structures available. They are the Qua1 domain from STAR proteins (Beuck et al., 2010; Meyer et al., 2010) , the protein kinase A (PKA) type Iα and type IIα regulatory subunits (Banky et al., 2003; Newlon et al., 1999), and the Siah-interacting protein (SIP) (Santelli et al., 2005). Of these structures, Get5-C has the shortest sequence and the highest ratio of buried to exposed surface area (30% versus 28% for the next most, the crystal structure of

the PKA type IIα regulatory subunit) (Gold et al., 2006). It is significantly more thermostable than the only other dimer where this was measured, the Qua1 domain, with a T_m of 74 °C compared with 63 °C (Beuck et al., 2010). Therefore, this domain is currently unique in biology.

Engineering dimerization into proteins is an important goal (Bolon et al., 2005; Kuhlman et al., 2001). Much of this type of work focuses on the use of coiled-coils in the form of leucine zippers (Apostolovic et al., 2010). The motif we describe here may provide a useful alternative. Recombinant Get5-C expresses very well and is easily purified. The 35-residue sequence is also amenable to chemical synthesis. The Get4/Get5 complex proves that Get5-C is sufficient in maintaining dimers of significantly larger protein assemblies. The requirement of only a few key residues suggests that there is ample opportunity for protein design to alter stability, specificity, and surface properties.

It has recently been proposed that TA protein binding to Get3 causes two dimers of Get3 to assemble into a tetrameric complex (Suloway et al., 2012). This is supported by crystal structures of a tetrameric archael Get3 homolog and by SAXS analysis of TA protein-bound Get3 complexes that show similar overall size and shape. This tetramer has a longest dimension of ~150 Å, which makes it possible for the two copies of Get4 within the Get4/Get5 complex to interact with the two different Get3 dimers prior to loading of TA proteins. Sgt2, bound to at least one Get5 ubiquitin-like domain (Chartron et al., 2011), would then be positioned near the hydrophobic cavity created by Get3 (Figure 2.10). Alternatively, the two copies of Get4 may interact with the two Get3 subunits within a single lower order dimer. This necessitates a dramatic conformational change

within Get4/Get5 to bring the Get4 N terminus into proximity. Flexibility within Sgt2 may be amendable with this model. The exact mechanism of TA protein transfer from Sgt2 to Get3 awaits further characterization.

Get5 plays a central role in localizing the various components involved in TA targeting forming the nexus of the so-called sorting complex. The unusually stable dimerization domain of Get5 is likely critical to the sorting function. It remains unclear whether this domain plays a broader role beyond simple dimerization. The novelty of the architecture provides ample opportunity for future studies.

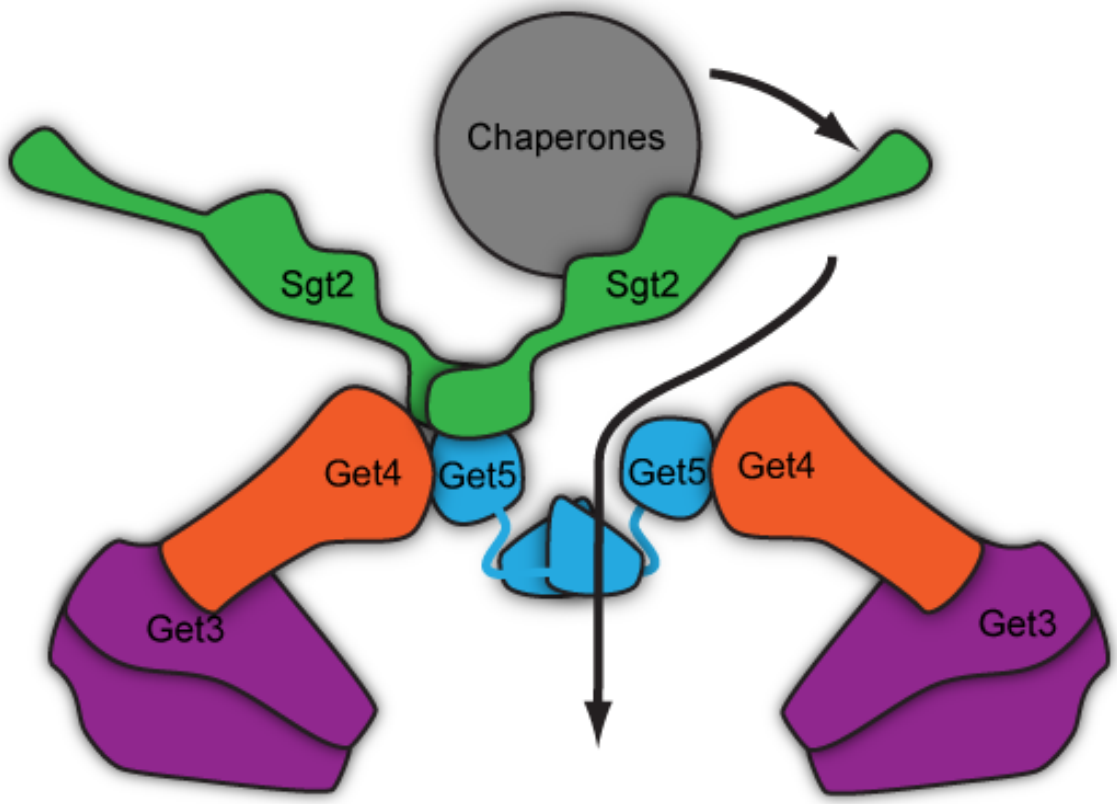


Figure 2.10. Model of the Get4/Get5/Sgt2 sorting complex and its interaction with Get3. Schemes are drawn to scale, with the exception of chaperones bound to Sgt2. The arrow indicates the path of the TA protein from chaperones to Sgt2 to Get3, which is held by Get4.

Methods

Cloning, Expression, and Purification

Get5-C residues 152–212 or 175–212 from *Saccharomyces cerevisiae* S288C or 159–230 from *Aspergillus fumigatus* 118 were amplified from vectors described previously (Chartron et al., 2010). The coding sequences were then inserted into a pET33b-derived plasmid. Unlabeled proteins were expressed in *Escherichia coli* BL21(DE3)Star (Invitrogen) for 3 h at 37 °C after induction with 250 μM isopropyl-β-D-thiogalactopyranoside (Affymetrix). Uniformly ^{15}N-labeled proteins were produced using N-5052 autoinduction media (Studier, 2005), and uniformly ^{13}C/^{15}N-labeled proteins were produced using the method of Marley et al. (2001). Cells were lysed using an S-4000 sonicator (Misonix) and purified by immobilized metal affinity chromatography (Qiagen). The full-length Get4/Get5 complex, Get5-Ubl-C (residues 74–212), and *Af*Get5-Ubl-C (residues *66–230*) were prepared as described previously (Chartron et al., 2010). (For clarity, residues that are italicized will refer specifically to the *A. fumigatus* homolog.)

Human Ubl4A was amplified from a cDNA-containing plasmid (ATCC) and inserted into a pET33b-derived plasmid. Expression and purification were as for Get5-C. Domain swap experiments and pulldown assays were performed as described previously (Chartron et al., 2010). Briefly, Get5-Ubl-C or *Af*Get5-Ubl-C was incubated in 2-fold stoichiometric excess with polyhistidine-tagged Ubl4A in 50 mM HEPES, 100 mM NaCl and 30 mM imidazole, pH 7.3. Ubl4A was precipitated with nickel-nitrilotriacetic acid-

agarose beads (Qiagen). The beads were washed twice with incubation buffer and proteins eluted with 20 mM EDTA.

Crystallization, Data Collection, and Structure Determination

Get5-C residues 175–212 were concentrated to 30 mg/ml in 20 mM sodium phosphate, pH 6.1. Crystallization screening was performed using the sitting-drop vapor diffusion method with commercially available screens (Qiagen) and a Mosquito robot (TTP Labtech). Clusters of plate like crystals grew after 1 week in a 1:1 ratio drop of protein solution to a reservoir of 3.4 M ammonium sulfate and 0.1 M sodium citrate, pH 5.0, at 22 °C. The clusters were broken apart, and individual crystals were transferred to reservoir solution supplemented with 10% glycerol for 5 min and then cryopreserved in liquid nitrogen. Iodide derivatives were generated by soaking crystals in freshly prepared 2.9 M ammonium sulfate, 0.5 M ammonium iodide, 0.1 M sodium citrate, and 10% glycerol, pH 5.0, for 5 min prior to cryopreservation.

X-ray diffraction data from a single native crystal were collected on beam line 12-2 at the Stanford Synchrotron Radiation Lightsource (SSRL) at 100 K using a Pilatus 6M detector and a microbeam. Positions on the crystal were screened for diffraction using an automated raster scanning protocol. The best position allowed for collection of a near complete dataset to maximum 1.23 Å resolution. Diffraction data from an iodide derivative were collected using a Micromax-007 HF rotating copper-anode generator and an R-axis IV++ image plate detector (Rigaku) to a maximum 1.6 Å resolution and ~12-fold redundancy. Data were integrated, scaled, and merged using XDS (Kabsch, 2010), iodide substructure determination, phasing, and initial model building and refinement

were performed in PHENIX (Adams et al., 2010), and manual building and model refinement were performed using COOT (Emsley et al., 2010). Structure figures were prepared using PyMOL (Schrödinger).

NMR Spectroscopy

All NMR measurements were collected using a Varian INOVA 600 MHz spectrometer at 25 °C with a triple resonance probe. Uniformly $^{13}C/^{15}N$-labeled Get5-C residues 152–212 or *Af*Get5-C residues *159–230* were concentrated to a monomer concentration of 3.5 mM in 20 mM sodium phosphate, pH 6.1. Chemical shift assignments were determined using standard triple-resonance experiments (Sattler et al., 1999). Data were processed using either TopSpin (Bruker) or NMRPipe (Delaglio et al., 1995) and analyzed using CCPN (Vranken et al., 2005). The PINE web server aided initial assignments (Bahrami et al., 2009). Distance restraints were derived from ^{13}C- and ^{15}N-edited NOESY spectra. An asymmetrically labeled dimer of *Af*Get5-C was prepared by mixing 4 mM unlabeled protein with 2 mM uniformly $^{13}C/^{15}N$-labeled protein and incubating at room temperature for 3 days. Intermolecular distance restraints were then determined using a $^{13}C/^{15}N$-filtered ^{13}C-edited NOESY spectrum.

H^N-N residual dipolar couplings were measured using the IPAP-HSQC experiment. A 2.5 mM solution of ^{15}N-labeled Get5-C was aligned in a 4% strained polyacrylamide gel, and a 2 mM solution of ^{15}N-labeled *Af*Get5-C was aligned in a 5% strained polyacrylamide gel (Chou et al., 2001).

Solution Structure Determination

ARIA2.3 was used for automated NOE cross peak assignment and structure calculation with hydrogen bond, dihedral and residual dipolar coupling-derived restraints (Rieping et al., 2007). Every NOE cross peak was treated ambiguously as an inter- or intramolecular contact, with initial sets of unambiguous intermolecular contacts determined by constraints imposed by secondary structure (Bardiaux et al., 2009). For *Af*Get5-C, experimentally determined intermolecular restraints were also utilized. Restraints for ϕ/ψ dihedral angles were predicted from chemical shifts using TALOS+ (Shen et al., 2009). An energy term is included to maintain symmetry within each model. Hydrogen bond restraints were initially assigned to the amide protons most resistant to deuterium exchange. After structure calculation, additional hydrogen bonds that were supported by the ensemble of models were added within the helical segments and the calculations repeated. The initial structures were also used to determine the axial and rhombic components of the alignment tensors with the program REDCAT (Valafar and Prestegard, 2004). Residual dipolar couplings were added as restraints during subsequent calculations. A log-harmonic energy potential was used during the second Cartesian cooling phase of the simulated anneal protocol with automatic determination of weights for NOE-derived and hydrogen bond distance restraints (Nilges et al., 2008). One hundred models were generated in the final iteration of ARIA, and the 10 lowest energy models were selected for refinement in explicit water.

SAXS

Multiple concentrations of *S. cerevisiae* Get4/Get5 were prepared as described previously (Chartron et al., 2011), with dialysis against 50 mM Tris, 300 mM NaCl, and 5 mM 2-

mercaptoethanol, pH 8.0. Data were collected at SSRL beam line 4-2 using a Rayonix MX225-HE detector, 1.13 Å wavelength X-rays, and a detector distance of 2.5 m for a momentum transfer range of 0.0055–0.3709 $Å^{-1}$. Data were processed using MARPARSE (Smolsky et al., 2007), and Guiner analysis was performed with PRIMUS (Konarev et al., 2003). The distance distribution function was determined with GNOM (Svergun, 1992). Ten independent *ab initio* models were generated using DAMMIF (Franke and Svergun, 2009) and real space data. These were superposed, averaged, and filtered using DAMAVER (Volkov and Svergun, 2003). High-resolution structures of Get4 and Get5 components were used as input into the program CORAL (Petoukhov and Svergun, 2005) for rigid body fitting with reciprocal space data. All calculations used imposed 2-fold symmetry.

Circular Dichroism Spectroscopy

Thermal denaturation measurements were collected using an Aviv 62A DS circular dichroism spectrometer. Get5-C or *Af*Get5-C at 10 μM in 20 mM sodium phosphate, pH 7.5, was heated from 25 to 99°C in 1 °C increments, and ellipticity was measured at 227 or 221 nm, respectively. Fractions of unfolded and folded protein were approximated using plateau values at low and high temperature.

Acknowledgements

We thank S. O. Shan, M. E. Rome, C. J. M. Suloway, and H. B. Gristick for critical reading of the manuscript; members of the laboratory for support and useful discussions; Julie Hoy for help in data collection on our home source; Graeme Card, Ana Gonzalez, and Michael Soltice for help with data collection at SSRL BL12-2; and Tsutomu Matsui and Hiro Tsuruta (1962–August 2011) for help with bioSAXS data collection and processing at SSRL BL4-2.

Chapter 3

A STRUCTURAL MODEL OF THE SGT2 PROTEIN AND ITS INTERACTIONS WITH CHAPERONES AND THE GET4/GET5 COMPLEX

Abstract

The insertion of tail-anchored transmembrane (TA) proteins into the appropriate membrane is a post-translational event that requires stabilization of the transmembrane domain and targeting to the proper destination. Sgt2 is a heat-shock protein cognate (HSC) co-chaperone that preferentially binds endoplasmic reticulum–destined TA proteins and directs them to the GET pathway via Get4 and Get5. Here, we present the crystal structure from a fungal Sgt2 homolog of the tetratricopeptide repeat (TPR) domain and part of the linker that connects to the C-terminal domain. The linker extends into the two-carboxylate clamp of the TPR domain from a symmetry-related molecule mimicking the binding to HSCs. Based on this structure, we provide biochemical evidence that the Sgt2 TPR domain has the ability to directly bind multiple HSC family members. The structure allows us to propose features involved in this lower specificity relative to other TPR containing co-chaperones. We further show that a dimer of Sgt2 binds a single Get5 and use small angle X-ray scattering to characterize the domain arrangement of Sgt2 in solution. These results allow us to present a structural model of the Sgt2-Get4/Get5-HSC complex.

Adapted from

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

Introduction

The eukaryotic cell is a complex environment of multiple membrane-bound organelles, each with a unique set of resident integral membrane proteins. Biogenesis of these proteins requires mechanisms for targeting and insertion into the correct membrane. For the majority destined to the ER membrane, this is accomplished via the signal recognition particle pathway (Walter and Johnson, 1994). Major exceptions to this rule are tail-anchor transmembrane (TA) proteins that are defined topologically by a single transmembrane helix within 30 residues of the C-terminus. Examples are found in all membranes exposed to the cytoplasm (Borgese et al., 2007; Kutay et al., 1993). A dedicated targeting pathway for ER destined TA proteins has been elucidated and is called the GET pathway (Guided Entry of TA proteins) in yeast (Schuldiner et al., 2008). The central player is Get3, a cytosolic ATPase that sequesters the transmembrane segment of a newly synthesized TA protein for targeting. A multiprotein complex consisting of Get4/Get5 and Sgt2 loads the TA protein onto Get3 (Wang et al., 2010). Get4 is an α-helical repeat protein that forms an obligate dimer with the N-terminal domain of Get5 (Bozkurt et al., 2010; Chang et al., 2010; Chartron et al., 2010), which also contains a ubiquitin-like domain and a C-terminal dimerization domain (Chartron et al., 2010).

Sgt2, the small glutamine-rich tetratricopeptide repeat (TPR) containing protein (SGT in mammals), is a 38-kDa protein highly conserved across eukaryotes (Kordes et al., 1998). It consists of an N-terminal homodimerization domain, a TPR domain composed of three TPR repeats, and a C-terminal domain that is rich in glutamine and methionine (Liou and Wang, 2005; Tobaben et al., 2003; Worrall et al., 2008). The Sgt2 dimer has a larger hydrodynamic radius than expected for a globular protein suggesting an extended conformation (Liou and Wang, 2005; Worrall et al., 2008). SGT interacts with a variety of proteins, notably heat-shock proteins and their cognates (referred to here in general as HSC) such as Hsc70 and Hsp90, which bind directly to the TPR domain (Angeletti et al., 2002; Liou et al., 2007; Liou and Wang, 2005; Worrall et al., 2008). SGT binding to HSCs appears to modulate the chaperone ATPase activity and folding rates dependent on other HSC co-chaperones. Binding to Hsc70 decreases ATPase activity and protein folding rates, whereas a neuronal Hsc70 complex, including SGT and cysteine string protein, stimulates ATPase activity (Angeletti et al., 2002; Tobaben et al., 2001; Wu et al., 2001). SGT also binds a number of viral proteins, and, in one case, the interaction was mediated by the TPR domain (Callahan et al., 1998; Cziepluch et al., 1998; Fielding et al., 2006). The C-terminal domain of SGT is capable of binding hydrophobic regions of protein, such as the N-terminal signal sequence of myostatin and *in vitro* translated type 1 glucose transporter (Liou and Wang, 2005; Wang et al., 2003).

The initial links between Sgt2 and Get5 were from proteome-wide yeast two-hybrid and tandem-affinity purification assays (Krogan et al., 2006; Uetz et al., 2000). Additional yeast two-hybrid and pulldown assays demonstrated that an N-terminal construct of Sgt2 was necessary and sufficient for binding to Get5, and consequently Get4, in an

interaction also dependent on the Ubl domain of Get5 (Chang et al., 2010; Liou et al., 2007). A direct role for Sgt2 in the TA targeting pathway was shown by two independent genetic interaction analyses indicating a strong functional connection with other GET pathway members (Battle et al., 2010; Costanzo et al., 2010). Both demonstrated TA protein mislocalization in Sgt2 deletion strains, and Battle et al. (2010) proposed that Sgt2 acts functionally upstream of Get5 and the other GET members. Most recently, it has been shown using an *in vitro* translation system that Sgt2 can bind to ER destined TA proteins directly through the C-terminal hydrophobic-binding domain. With the aid of Get4/Get5, the TA protein is then transferred to Get3 (Wang et al., 2010). TA proteins destined to the mitochondria do not bind Sgt2 directly but are bound to TPR domain-associated chaperones.

TPR domains are defined by a variable number of two-helix, 34-residue motifs and frequently mediate protein-protein interaction (D'Andrea and Regan, 2003). A subclass acts as co-chaperones of HSCs by regulating nucleotide hydrolysis cycles, physically linking multiple HSC families, and/or connecting protein-folding pathways with alternate pathways such as ubiquitination and degradation (Meacham et al., 2001). Examples are found in protein phosphatase 5, Hsp organizing protein (HOP in human, Sti1 in yeast) and C-terminus of Hsc70 interacting protein (CHIP) (Das et al., 1998; Scheufler et al., 2000; Zhang et al., 2005). This TPR subclass is composed of three repeats followed in sequence by a C-terminal capping helix and recognizes the C-terminal residues of the HSC, exemplified by IEEVD in Hsc70 and MEEVD in Hsp90. Five conserved residues from the TPR domain mediate this interaction forming a motif named the two-

carboxylate clamp, based on the recognition of the terminal acidic residue and the main-chain carboxylate (Scheufler et al., 2000).

Here we report the crystal structure of the Sgt2 TPR domain from the filamentous fungus *Aspergillus fumigatus* (*Af*Sgt2). A crystallographic contact generates a serendipitous interaction that mimics the carboxyl sequence of many HSCs binding to a TPR co-chaperone. Based on a structural analysis, we demonstrate biochemically that the Sgt2 TPR domain in *Saccharomyces cerevisiae* (Sgt2) can directly bind to at least four different HSC families. We also test the potential stoichiometry between Sgt2 and Get5 and characterize the structure of Sgt2 in solution, allowing us to present a structural model for the higher order complex between chaperones, Sgt2 and Get4/Get5.

Results

Crystal Structure of a Fungal Sgt2 TPR Domain

We expressed the TPR and C-terminal domains of the *A. fumigatus* homolog of Sgt2 (*Af*Sgt2), corresponding to residues 109–341, in *E. coli* and purified it using nickel-nitrilotriacetic acid affinity chromatography. There was an extensive range of proteolysis, and two major species could be resolved by SEC (Figure 3.1). The smaller protein is a proteolytic product with a molecular mass of 17,547 Da, consistent with the predicted

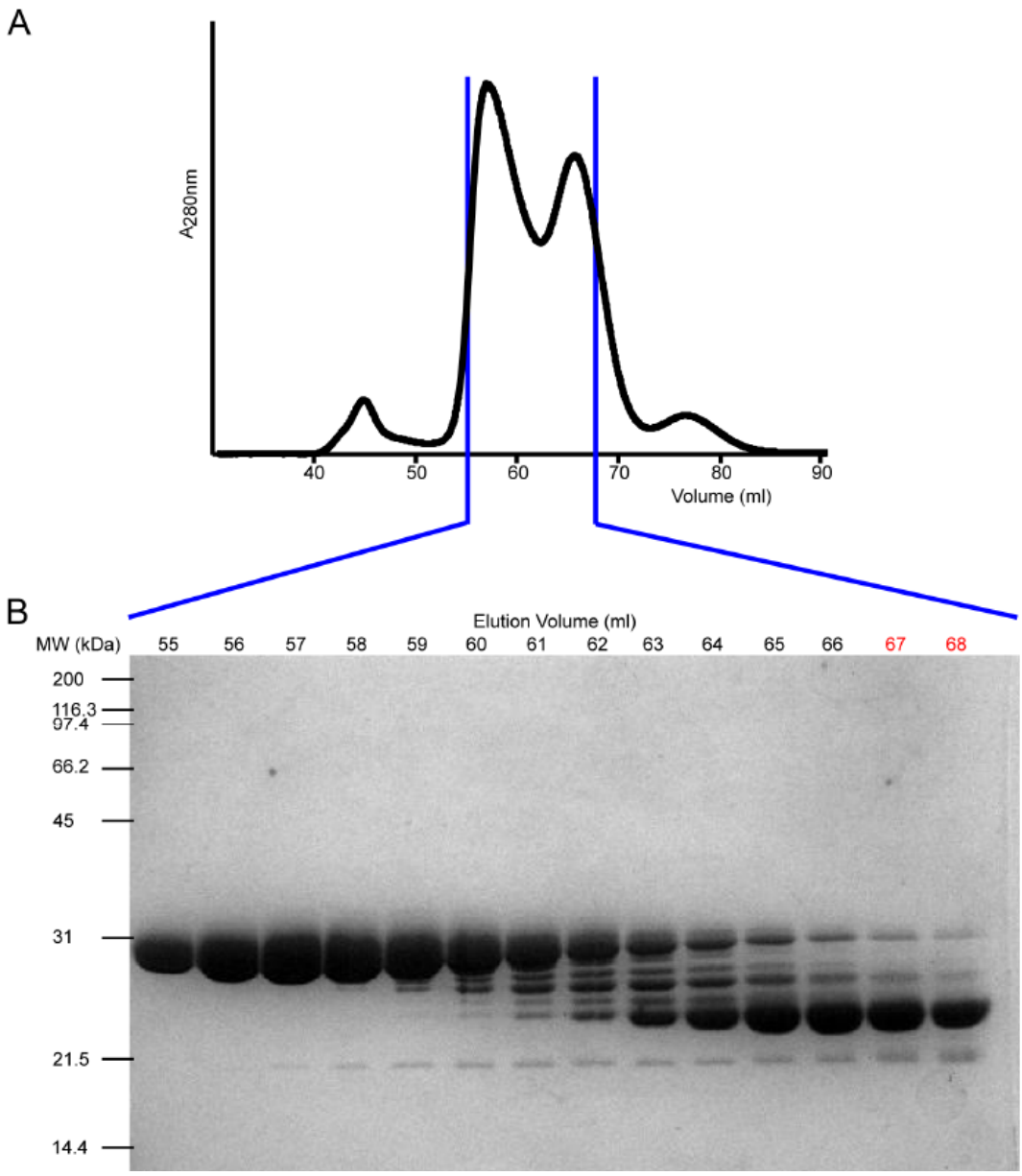


Figure 3.1. Purification of *Af*Sgt2-TPR-C. (A) Size exclusion chromatography of *Af*Sgt2-TPR-C after Ni-affinity chromatography using a Superdex 75 16/60 column. (B) SDS-PAGE of the region indicated in (A) with elution volume indicated per lane. Fractions labeled in red were used for crystallization trials without further purification. The major band has a molecular weight of 17,547 Da by mass spectrometry and migrates slower than the molecular weight standards on the gel.

molecular mass of residues 109–267. This protein yielded orthorhombic crystals in space group *F*222 that diffracted to a maximum of 1.72 Å resolution. The crystal structure was solved by molecular replacement using the human SGT α-isoform TPR domain (SGTA) (Dutta and Tan, 2008) as a search model. A single copy of the TPR domain (*Af*Sgt2-TPR) was located in the asymmetric unit (Figure 3.2A). Electron density extending from the C-terminal helix of the TPR domain could be modeled as 21 residues of the linker connecting to the C-terminal domain. Overall, unambiguous electron density throughout the entire chain allowed residues 109–254 to be modeled and refined to an R_{factor} of 0.168 and an R_{free} of 0.208. Complete crystallographic statistics are shown in Table 3.1.

The *Af*Sgt2-TPR has 38% sequence identity and 60% similarity to the SGTA TPR domain and shares an architecture consisting of three TPR repeats comprised by helices α1–α6 with a "capping" helix, α7 (Figure 3.2A,E). Like other domains composed of three TPR repeats, these seven helices are arranged in a right-handed supercoil (D'Andrea and Regan, 2003) resulting in a concave surface lined by α1, α3, α5, and α7. The C_α root mean square deviation between the TPR domains of the SGTA and *Af*Sgt2 is 1.2 Å. Relative to the truncated SGTA construct, α7 is extended by five residues, and the angle formed between α6 and α7 is increased by ~10° (Figure 3.2B). Unlike the SGTA crystal structure, which lacks extensive intermolecular contacts, *Af*Sgt2-TPR forms a crystallographic dimer with a symmetry-related copy, burying 2108 $Å^2$ of solvent-accessible surface area (Figure 3.2A, inset). This interface is mediated by the α7 helices, which pack head to tail against one another, and residues 240–254 of the linker to the C-terminal domain, which bind into the TPR groove of the opposite copy in an extended

Table 3.1. Data collection and refinement statistics for *Af* Sgt2-TPR

Data collection	
Space Group	F 2 2 2
Unit Cell	
a, b, c	59.6, 86.77, 116.54 Å
α, β, γ	90°, 90°, 90°
Resolution	26.54-1.72 Å (1.81-1.72 Å)
R_{merge}	0.060 (0.502)
I/σ(I)	14 (4.2)
Completeness	97.2% (98.8%)
Redundancy	4.0 (4.1)
Refinement	
Resolution	26.54-1.72 Å
R_{work}/R_{free}	0.169/0.208
No. atoms	
Protein	1124
Water	109
B-factors	
Protein	25.9
Water	39.2
Root mean square deviations	
Bond lengths	0.004 Å
Bond angles	0.74°
Ramachandran statistics	
Favored	100%
Allowed	0%
Outlier	0%

[a] Values in parentheses are for highest resolution shell

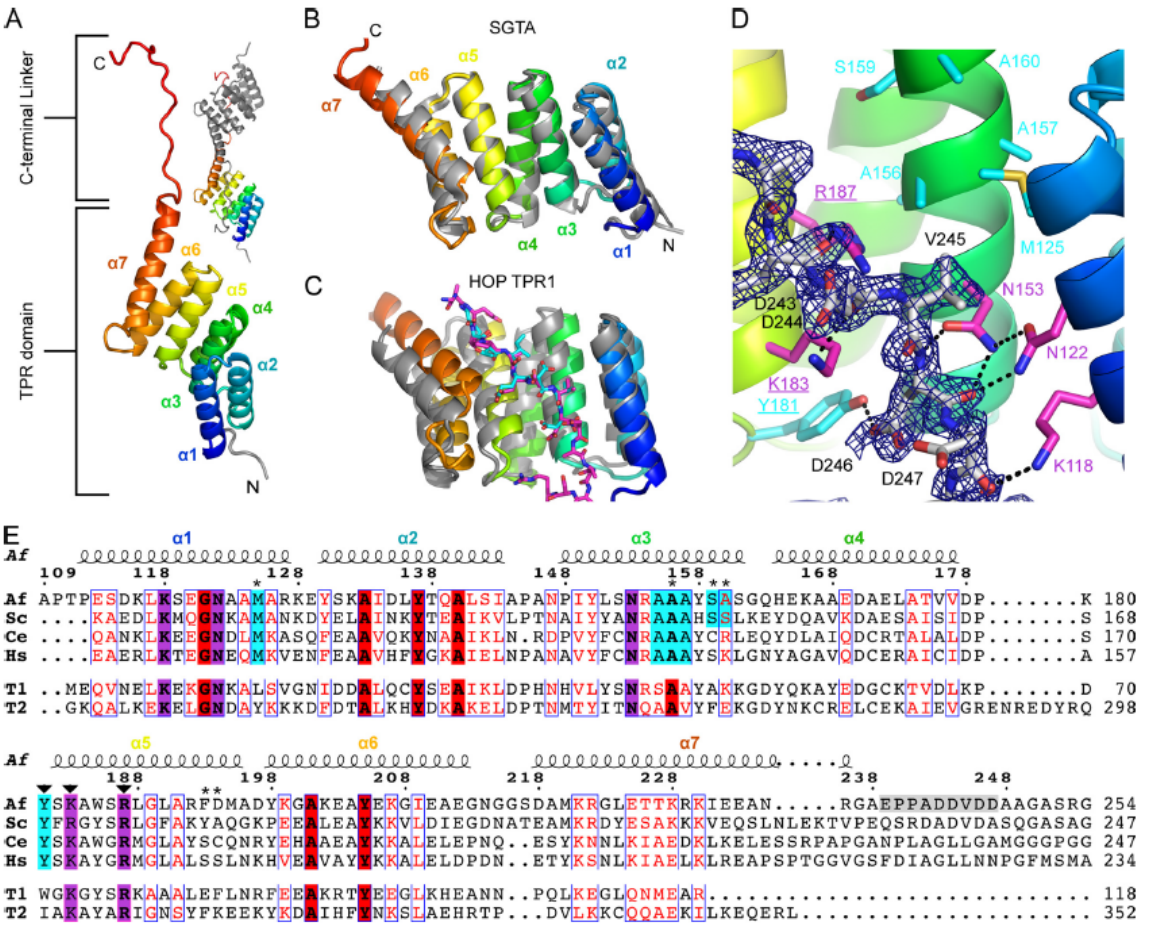


Figure 3.2. Crystal structure of AfSgt2 TPR domain. (A) The asymmetric unit shown as a ribbons diagram with color ramped from the N (blue) to C (red) termini. The region in gray is a cloning artifact. The crystallographic dimer is inset. (B) *Af*Sgt2-TPR as in (A) superposed on human SGTA TPR domain (gray, PDBID 2VYI). (C) Superposition of *Af*Sgt2 and bound C-terminal linker (rainbow, magenta carbons) onto Hsc70 peptide-bound HOP TPR1A (gray, cyan carbons, PDBID 1ELW). (D) *Af*Sgt2 TPR groove with a 1σ $2|F_o|-|F_c|$ simulated annealing omit map of the C-terminal linker of *Af*Sgt2 (gray carbons). Hydrogen bonds to conserved two-carboxylate clamp residues (magenta carbons) are indicated as black dashes. Residues are labeled based on the *Af*Sgt2 sequence. Sgt2-specific residues are shown as cyan carbons. The FAA mutant positions are underlined. (E) Alignment of the sequence observed in the crystal structure. Sequences are: Af, *A. fumigatus*; Sc, *S. cerevisiae*; Ce, *C. elegans*; Hs, *Homo sapiens*; T1, *H. sapiens* HOP TPR1; T2, *H. sapiens* HOP TPR2a. The numbering above is from *A. fumigatus*. Residues predicted to be involved in substrate specificity are marked with asterisks. The locations of the FAA mutations are indicated with arrowheads. Two-carboxylate clamp residues and conserved Sgt2 residues are highlighted in magenta and cyan, respectively. Further TPR conserved residues are highlighted in red. The C-terminal linker bound in the TPR groove is highlighted in gray.

conformation. This includes a well-ordered interaction between two acidic side chains (Glu-239 to Asp-194) outside the groove that is unlikely to occur *in vivo* and may be stabilized by the low pH of crystallization.

Fortuitously, the sequence of residues 240–247 in *Af*Sgt2, PPADDVDD, resembles the C-terminal residues of Hsp70 and Hsp90 homologs, exemplified respectively in yeast by PTVEEVD in Ssa1 and TEMEEVD in Hsc82. These sequences are recognized by a variety of TPR domains containing a two-carboxylate clamp that anchors the EEVD motif (Scheufler et al., 2000). Hsp70/Hsp90 Organizing Protein (HOP) contains three TPR domains designated TPR1, TPR2a, and TPR2b. TPR1 specifically binds Hsc70, whereas TPR2a binds Hsp90, and the crystal structures of these interactions have allowed an understanding of determinants of substrate specificity in TPR domains (Odunuga et al., 2003; Scheufler et al., 2000). The main chain conformation of the C-terminal linker region of *Af*Sgt2 in our crystal structure is identical to that of the GPTIEEVD peptide bound to HOP TPR1 (Figure 3.2C). The two-carboxylate clamp is composed of five highly conserved residues that make hydrogen bonds to the main chain of the EEVD motif. In *Af*Sgt2, Arg-187 (Arg-175; for clarity when referencing the structure *A. fumigatus* numbering will be used and *S. cerevisiae* sequence numbering will be provided in parenthesis) and Lys-183 (Arg-171) interact with the carbonyl of Asp-244 and Arg-187 (Arg-175) makes an additional contact to the carbonyl of Asp-243 (Figure 3.2D). Asn-122 (Asn-110) extends from α3 to form a hydrogen bond with Asn-153 (Asn-141) from α5, and these two asparagines hydrogen bond with the amide and carbonyl of Asp-246. Lys-118 (Lys-106) hydrogen bonds with the carbonyl of Asp-247, the equivalent position of the terminal carboxyl of Hsp70/90. In addition to these five canonical

residues, Tyr-181 (Tyr-169), conserved across eukaryotes, forms a hydrogen bond with the side chain of Asp-246 (Figure 3.2D,E).

The TPR Domain of Sgt2 Is a General HSC Binding Interface

Sgt2 is physically linked with several families of heat-shock proteins. Hsp70 homologs Ssa1 and Ssa2, Hsp110 homologs Sse1 and Sse2, the Hsp90 homolog Hsc82, and the Hsp100 homolog Hsp104 co-purify with Sgt2 from yeast lysate, and all of these interactions are abolished by mutation of residues in the two-carboxylate clamp (Costanzo et al., 2010; Wang et al., 2010). TPR domain containing co-chaperones often have specificity for different HSCs; therefore, we asked if Sgt2 can either bind each of these chaperone families directly or if co-immunoprecipitation of certain families were mediated by sub-complexes between chaperones. For example, Sse1 can form a stable complex with Ssa1 (Shaner et al., 2005). SGT and a homolog from *C. elegans* can bind directly to either Hsp70 or Hsp90 (Angeletti et al., 2002; Liou and Wang, 2005; Worrall et al., 2008). Using a nickel affinity pulldown assay, purified Ssa1 and Hsc82 bind directly to a polyhistidine-tagged TPR domain of Sgt2 (Figure 3.3A lanes 1-2 and 4-5). A triple mutant of the Sgt2-specific Y169F and the two-carboxylate clamp residues R171A and R175A (designated "FAA" corresponding to *Af*Sgt2 numbering Y181/K183/R187 (Figure 3.2D,E), reduced binding to background levels (Figure 3.3A, lanes 3 and 6). Hsp104 has a C-terminal sequence of MEIDDDLD and binds to the two-carboxylate clamp in Cpr7 and the TPR1 domain of Sti1 (Abbas-Terki et al., 2001). The C-terminal sequence of Sse1 is more divergent with the sequence EGDVDMD. Both

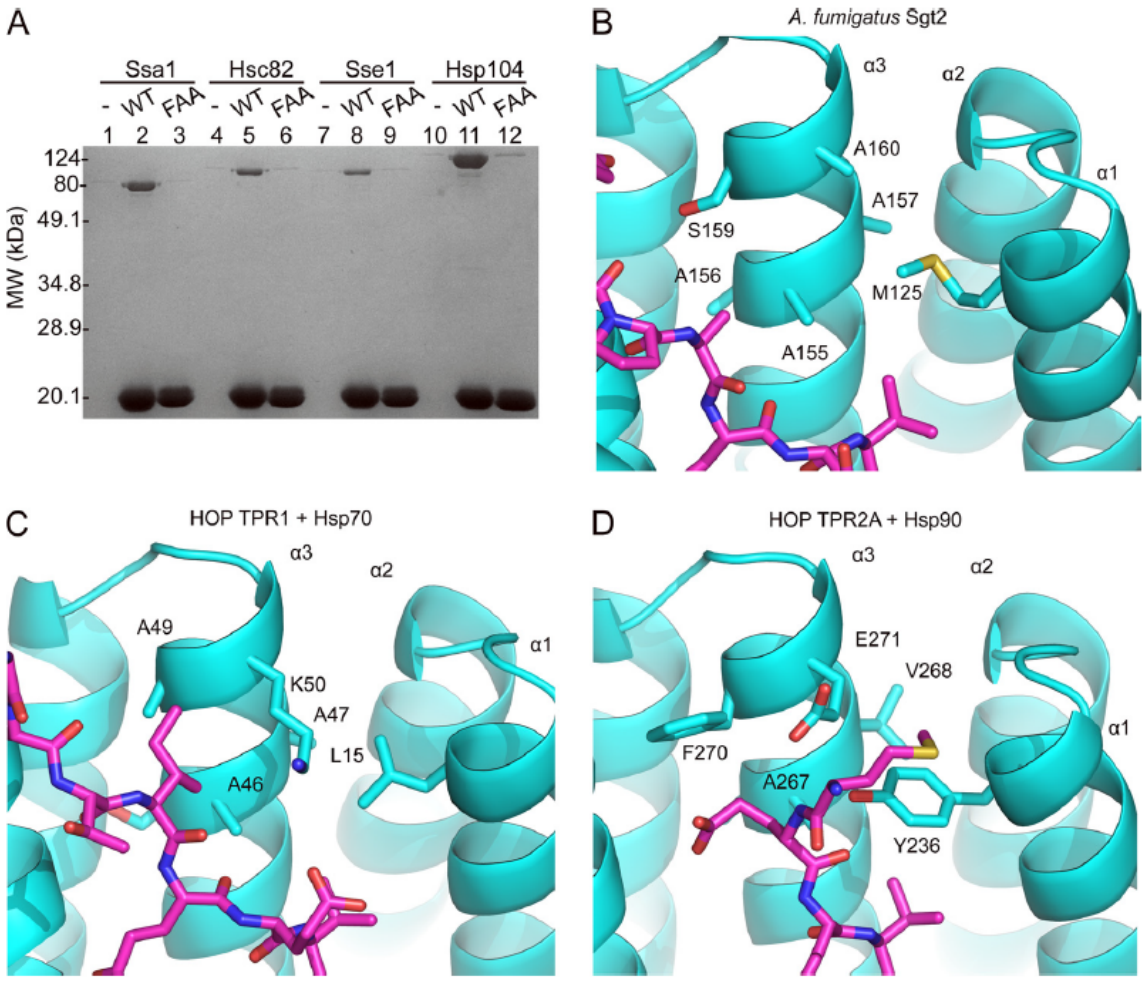


Figure 3.3. Sgt2 binds multiple chaperone families. (A) SDS-PAGE of nickel affinity pulldown assays of Ssa1, Hsc82, Sse1, or Hsp104 in the absence of Sgt2 (−), with wild type polyhistidine-tagged Sgt2-TPR (WT) or polyhistidine-tagged Sgt2-TPR-FAA (FAA). (B–D) Hydrophobic binding pocket on the TPR groove (cyan) with bound peptides (magenta) of *Af*Sgt2-TPR with symmetry molecule (B), *H. sapiens* HOP TPR1 with Hsp70-derived peptide (PDB ID 1ELW) (C), and *H. sapiens* HOP TPR2A with Hsp90-derived peptide (PDBID 1ELR) (D).

Sse1 and Hsp104 bind to the TPR domain of Sgt2 dependent on the two-carboxylate clamp (Figure 3.3A, lanes 7–12).

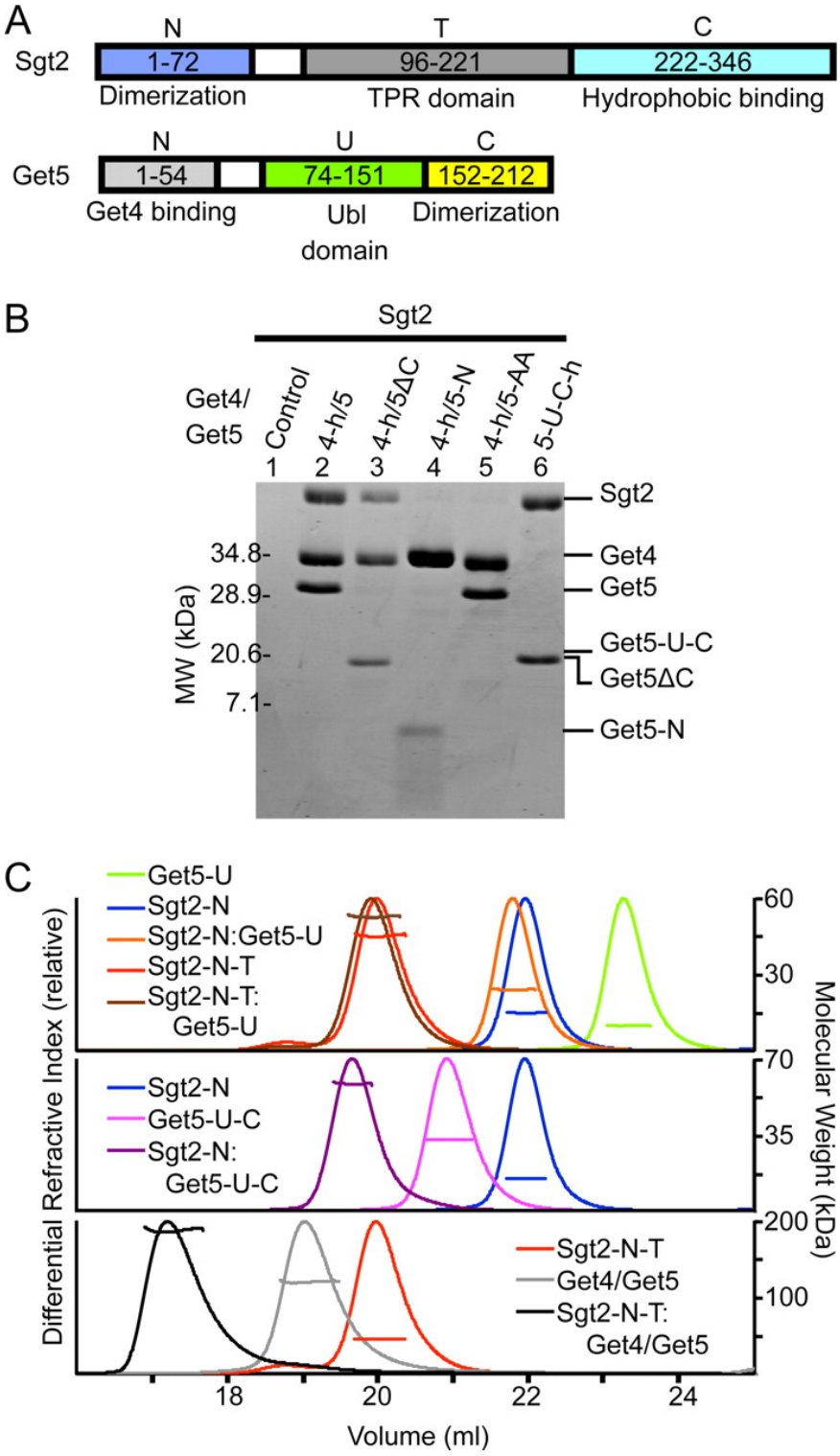
The structural basis for TPR domain specificity has been well characterized in HOP/Sti1. The TPR1 domain of HOP/Sti1 binds Hsp70 and Hsp100, whereas the TPR2A domain binds Hsp90 (Abbas-Terki et al., 2001; Scheufler et al., 2000). The identity of the hydrophobic residues preceding the EEVD terminus of Hsp70 and Hsp90 is critical for specific binding (Brinker et al., 2002). Structure-guided mutational analysis determined a set of TPR residues that impact specificity, in particular TPR residues equivalent to positions Met-125 (Met-113), Ser-159 (Ser-147), and Ala-160 (Ser-148) in *Af*Sgt2 (Figures 3.2E and 3.3B) (Odunuga et al., 2003). In HOP-TPR1, these positions are occupied by Leu-15, Ala-49, and Lys-50, and the isoleucine of the Hsp70 GPTIEEVD sequence fits between the alanine and lysine side chains on α3 (Figure 3.3C). Alternatively, in HOP-TPR2a, the respective positions are occupied by Tyr-236, Phe-270, and Glu-271, and the methionine of the Hsp90 MEEVD sequence is found between α3 and α1 interacting with the tyrosine and glutamate (**Figure 3.3D**). A structure of HOP-TPR2a bound with non-cognate Hsp70-derived peptide shows that the isoleucine is not accommodated and becomes solvent exposed (Kajander et al., 2009). In contrast with the HOP/Sti1 TPR domains, these positions in Sgt2 create an open pocket that we predict accommodates a wider range of substrates (Figure 3.3B). Met-125 (Met-113) is conserved across the eukaryotic kingdom; Ser-159 (Ser-147) and Ala-160 (Ser-148) are highly conserved in fungi (Figure 3.2E). Higher eukaryotes have a conserved basic residue at position 160, perhaps indicating variation in specificity. Additionally, Ala-155

(Ala-143), Ala-156 (Ala-144), and Ala-157 (Ala-145) are strictly conserved in Sgt2 and complete the substrate interacting face of α3 (Figures 3.2E and 3.3B).

Another characterized example of binding to multiple heat-shock protein families is the C-terminal TPR domain of Hsp70 interacting protein, CHIP, which can bind either Hsp70 or Hsp90. Structures with peptides derived from either chaperone show that, rather than binding in an extended conformation, the bound peptide kinks immediately prior to the EEVD motif to position the upstream isoleucine or methionine into a large hydrophobic pocket lined by α5, α6, and α7 that is unique to CHIP (Wang et al., 2011; Zhang et al., 2005).

The Sgt2 Dimer Binds a Single Copy of Get5 at a Canonical Ubl Interface

Yeast two-hybrid assays (Chang et al., 2010; Liou et al., 2007) and analytical SEC (Wang et al., 2010) indicated that the N terminus of Sgt2 and the Ubl domain of Get5 are necessary for binding between these two proteins; however, molecular details of this interaction remain to be defined. We used a nickel affinity pulldown assay to further characterize the complex. Residue ranges of the domains of Sgt2 and Get5 are defined in Figure 3.4A. Polyhistidine-tagged Get4/Get5 can bind Sgt2 (Figure 3.4B, lane 2), as well as Get4/Get5ΔC (Figure 3.4B, lane 3), indicating that dimerization of Get4/Get5 is not essential for the interaction. Further deletion of the Ubl domain abolishes binding to Sgt2 (Figure 3.4B, lane 4). We previously identified a double mutant of Get5, L120A/K124A, that resulted in incomplete rescue of a Get5 deletion strain and is predicted to disrupt a conserved binding interface in ubiquitin-like domains (Chartron et al., 2010; Hicke et al., 2005). Get4/Get5-L120A/K124A is unable to bind Sgt2 (Figure 3.4B, lane 5).


A
N
T
C
Sgt2
1-72
96-221
222-346
Dimerization
TPR domain
Hydrophobic binding
N
U
C
Get5
1-54
74-151
152-212
Get4 binding
Ubl domain
Dimerization
B
Sgt2
Get4/ Get5
Control
4-h/5
4-h/5ΔC
4-h/5-N
4-h/5-AA
5-U-C-h
1
2
3
4
5
6
MW (kDa)
34.8-
28.9-
20.6-
7.1-
Sgt2
Get4
Get5
Get5-U-C
Get5ΔC
Get5-N
C
Get5-U
Sgt2-N
Sgt2-N:Get5-U
Sgt2-N-T
Sgt2-N-T: Get5-U
Sgt2-N
Get5-U-C
Sgt2-N: Get5-U-C
Sgt2-N-T
Get4/Get5
Sgt2-N-T: Get4/Get5
Differential Refractive Index (relative)
Molecular Weight (kDa)
60
30
70
35
200
100
18
20
22
24
Volume (ml)

In the absence of Get4, recombinant Get5 is unstable and prone to forming inclusion bodies or susceptible to proteolysis (Chartron et al., 2010; Liou et al., 2007; Wang et al., 2010). Removing the N-terminal Get4 binding domain results in a stable Get5-Ubl-C (Chartron et al., 2010), which is capable of binding to Sgt2 (Figure 3.4B, lane 6). We generated a construct of the N-terminal 72 residues of Sgt2. This minimal domain alone was sufficient to bind to Get4/Get5 (Figure 3.5A). From this assay, we conclude that the interaction between Get4/Get5 and Sgt2 is predominantly between the Ubl domain of Get5, and the N-terminal domain of Sgt2 and is mediated by a canonical binding interface.

We next investigated the stoichiometry of the interaction between Sgt2 and Get5. After formation, complexes of Sgt2 and Get5 are stable and can be purified from unbound protein by SEC (Figure 3.6). Purified complexes are stable upon further SEC and were subjected to SEC coupled with MALLS for molecular weight determination (Figure 3.4C and Table 2). The Sgt2-N dimer possibly has two unique binding interfaces; therefore, up to two copies of Get5-Ubl might be expected to bind. However, only a single higher weight peak was observed after incubation with a 3-fold stoichiometric excess of Get5-Ubl (Figure 3.6). The MALLS molecular weight of this complex was

---

Figure 3.4. The Sgt2-Get4/Get5 complex. (A) Schematics indicating residue ranges of domains described in text with corresponding letter abbreviations. (B) SDS-PAGE of nickel affinity pulldown assays. Get4 and Get5 are abbreviated as 4 and 5, respectively, with "h" indicating the polyhistidine-tagged protein. (C) SEC-MALLS of Sgt2 and Get4/Get5 complexes. Traces are normalized to maximum differential refractive index. Horizontal lines are experimentally measured molecular weight (right axis). Proteins added in 3-fold stoichiometric excess to generate complexes are Get5-Ubl (top), Sgt2-N (middle), Sgt2-N- TPR (bottom).

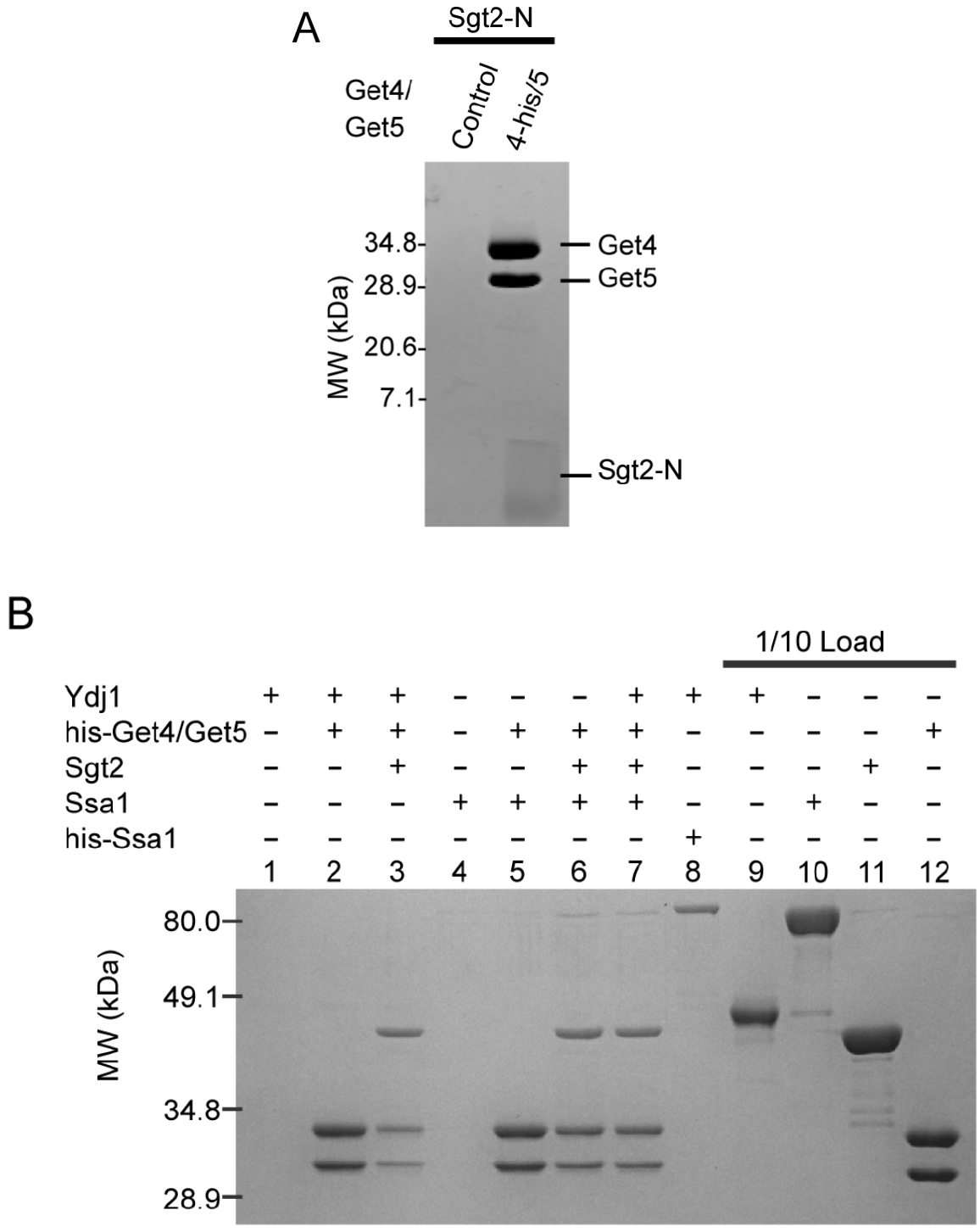


Figure 3.5. Sgt2 interactions. (A) Nickel affinity capture of Sgt2-N (residues 1-72) by poly-histidine tagged Get4/Get5. (B) Nickel affinity capture of indicated protein mixtures. The lanes indicated with the horizontal bar are 1/10 the total amount of protein added to each mixture.

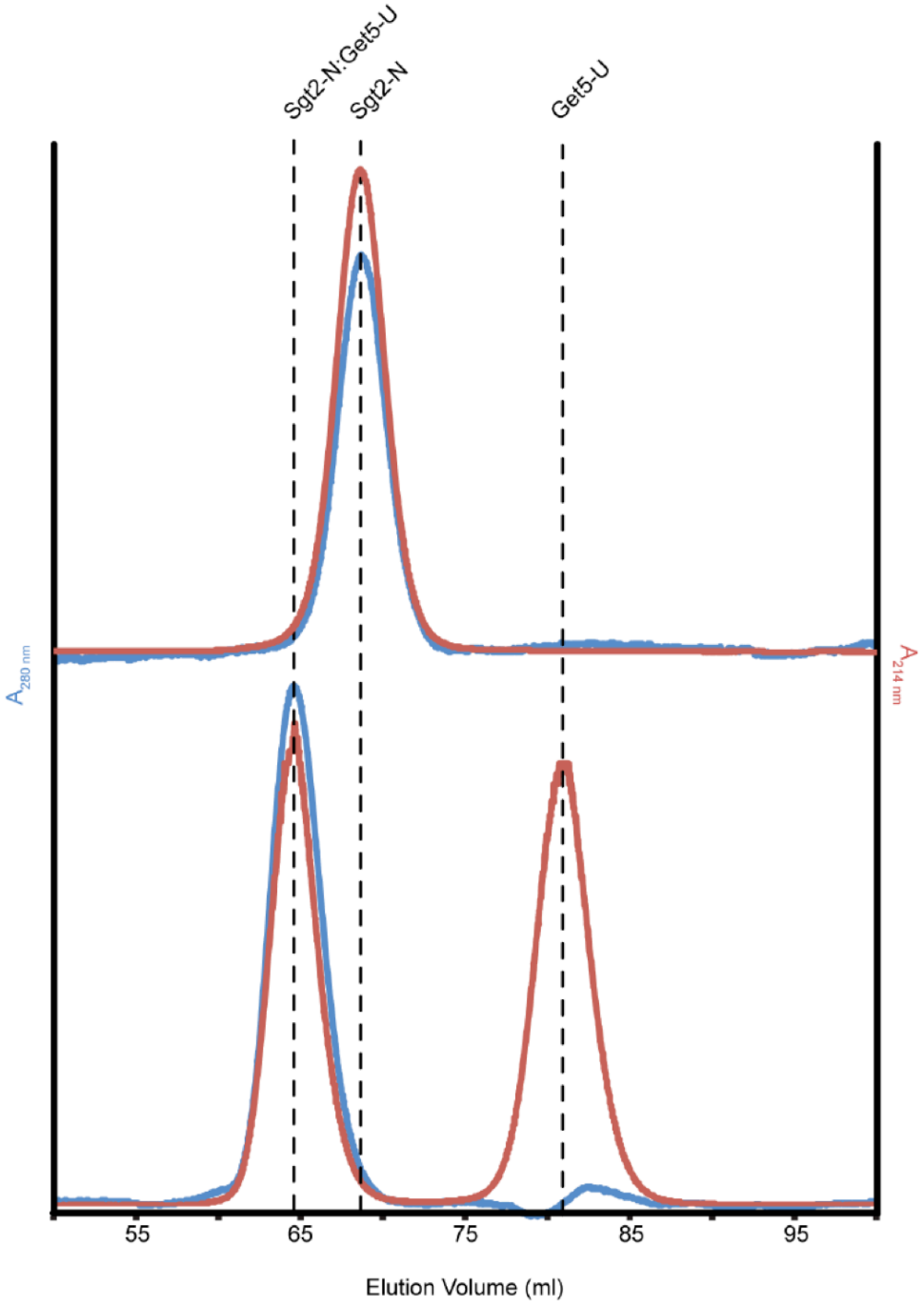


Figure 3.6. Formation of the Sgt2-N/Get5-Ubl complex. Size exclusion chromatography was performed with a Superdex 75 16/60 column. The top chromatogram is Sgt2-N alone and the bottom is Sgt2-N incubated with a three-fold stoichiometric excess of Get5-Ubl. Get5-Ubl does not contain tryptophan or tyrosine but absorbs at 214 nm.

most consistent with an Sgt2-N dimer and a single copy of Get5-Ubl (Figure 3.4C, top, and Table 3.2). Similarly, when Sgt2-N-TPR was incubated with excess Get5-Ubl the determined molecular weight of the resulting complex was in closer agreement with a single bound copy of Get5-Ubl than two copies. Under the conditions tested here, we did not detect Sgt2 binding to more than one copy of Get5.

TABLE 3 . 2
SEC-MALLS- and SAXS-derived parameters

	Sequence molecular mass	MALLS molecular mass	SAXS	
			R_g	D_{max}
	kDa		Å	Å
$Sgt2\text{-}N_{Dimer}$	15.8	14.66	16.9	65
$Sgt2\text{-}N\text{-}TPR_{Dimer}$	49.1	44.8	42.5	155
$Sgt2\text{-}TPR\text{-}C_{Monomer}$			38.4[a]	122[a]
$Get4/Get5_{Dimer}$	120.0	119.4		
$Get5\text{-}Ubl\text{-}C_{Dimer}$	33.6	32.31		
$Get5\text{-}Ubl_{Monomer}$	10.0	9.68		
Complexes				
Get5-Ubl/Sgt2-N	25.7, 35.7 [b]	23.78	20.7	70
Get5-Ubl/Sgt2-N-TPR	59.1, 69.1	52.19		
Sgt2-N/Get5-Ubl-C	49.4, 65.2	57.88		
Sgt2-N-TPR/Get4/Get5	169.1, 218.2	186.2		

[a] Values are the average for a 50-model ensemble selected by EOM .
[b] The molecular mass was estimated from sequences for 1:1 and 2:1 ratios.

When Get5-Ubl-C was incubated with excess Sgt2-N, the resulting complex had a determined molecular mass of 57.88 kDa (Figure 3.4C, middle). A complex of a Get5-Ubl-C dimer with a single Sgt2-N dimer has an expected molecular mass of 49.4 kDa and also with two Sgt2-N dimers of 65.2 kDa. When Get4/Get5 was incubated with excess Sgt2-N-TPR, the resulting complex had a molecular mass of 186 kDa, a value closer to a single copy of the Get4/Get5 dimer with a single copy of the Sgt2-N-TPR dimer (Figure 3.4C, bottom). Although full-length Sgt2 can interact with Get4/Get5 (Figure 3.4B), the same sample by SEC showed no higher peaks relative to each protein run individually.

This could suggest a conformational change resulting in a similar hydrodynamic radius for the complex relative to the individual proteins; however, it is more likely that the complex simply was not stable by this method. It is possible that steric clashes between Get4 and the Sgt2-C domain, or increased entropic costs, reduce the binding affinity.

Ydj1 Does Not Directly Bind the Sgt2-Get4/Get5 Complex

The yeast DnaJ homolog Ydj1 is reported to bind Get5, which mediates a genetic interaction between Ydj1 and Sgt2 (Liou et al., 2007). In that study, recombinant Get5, in the absence of Get4, formed an *in vitro* complex with Ydj1. We tested whether Ydj1 binds to purified Get4/Get5 but did not see any enrichment in a pulldown assay, nor could we detect binding with the addition of Sgt2 or Ssa1 (Figure 3.5B). The N-terminal domain of Get5 causes the protein to aggregate in the absence of Get4, and we propose that Ydj1 binds in response to this aggregation. Rather than directly interacting with Sgt2, Get4, or Get5, the genetic linkage is likely due to the role of Ydj1 in regulating Ssa1. It is noteworthy that Sis1, the other DnaJ homolog in yeast, co-immunoprecipitates with Sgt2 dependent upon two-carboxylate clamp (Wang et al., 2010).

SAXS of Sgt2

Sgt2 is a multidomain dimeric protein with an extended conformation whose domain arrangement is unknown. Small angle X-ray scattering (SAXS) allows determination of particle size, analysis of flexibility, and *ab initio* determination of low-resolution structure. This allows modeling of higher order assemblies when coupled with high-resolution structures of individual domains. SAXS was performed on Sgt2 using the constructs indicated in Table 3.2. Values for radius of gyration (R_g) and the maximum

particle distance (D_{max}) were obtained from the indirect Fourier transform processed in GNOM. We generated *ab initio* models of Sgt2-N and Sgt2-N-TPR using GASBOR (Svergun et al., 2001), which uses dummy residues restrained with simulated chain connectivity to fit experimental data. A total of 20 independent models was generated for each construct, and these were superposed, averaged, and filtered using DAMAVER (Volkov and Svergun, 2003). The resulting model represents the most probable volume shared by the individual models. Sgt2-N was reconstructed as a somewhat spherical particle, in agreement with the p(r) function, which has a single peak that smoothly approaches zero (Figure 3.7A,B). The p(r) function of Sgt2-N-TPR is indicative of multiple folded domains as it has more than one peak (Figure 3.7C). Sgt2-N-TPR reconstructed as a curved tubular shape, with two volumes, appropriately sized for TPR domains, extending out from the dimerization domain in the same plane (Figure 3.7D). We additionally modeled Sgt2-N-TPR with the program BUNCH, which allows fitting to multiple SAXS curves in cases where scattering from truncations are available, and it also fits high resolution structures as rigid bodies (Petoukhov and Svergun, 2005). In this case, the fit utilized the TPR crystal structure and Sgt2-N curve in addition to the Sgt2-N-TPR curve. Unknown regions were generated *ab initio* as dummy residues. The resulting models are in agreement with the averaged GASBOR model (Figure 3.7E). The orientation of the TPR domain groove is not resolved by SAXS; it may be resolution-limited or averaged due to flexibility in the inter-domain linker. Importantly, the angle formed between the TPR domains and the N-domain, as well as the end-to-end distance of the TPR domains is consistent between models from multiple methods.

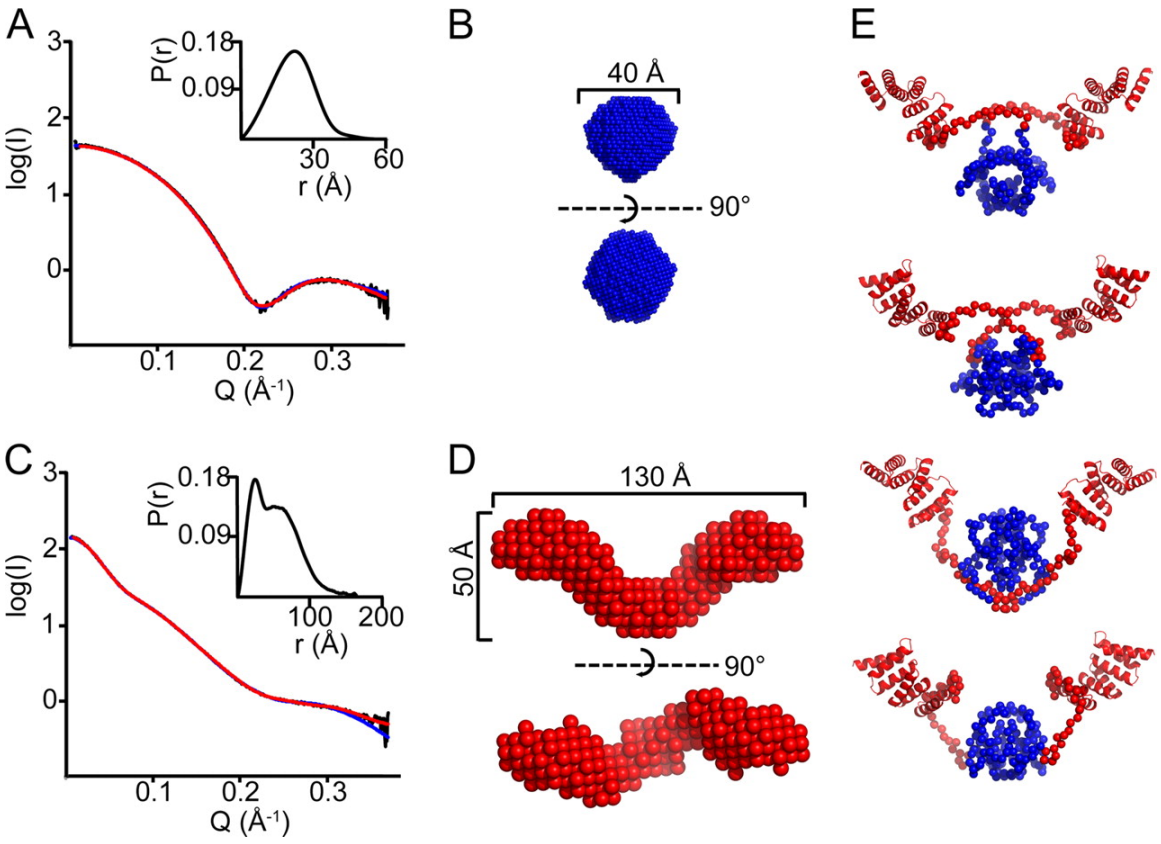


Figure 3.7. SAXS of Sgt2-N and Sgt2-N-TPR. (A) and (C) Experimental SAXS curve (black), fits of the best GASBOR (red) and BUNCH (blue) models and pair-probability functions (inset) for Sgt2-N (A) and Sgt2-N-TPR (C). (B) and (D) *ab initio* reconstructions of 20 averaged and filtered GASBOR models of Sgt2-N (B) and Sgt2-N-TPR (D). Bottom images are rotated relative to top images. (E) Four example BUNCH models. Blue regions were fit to both Sgt2-N and Sgt2-N-TPR data, while red regions were fit to only Sgt2-N-TPR.

From primary sequence, Sgt2-TPR-C is expected to have high flexibility, and the shape of the Kratky plot is characteristic of a partially folded protein as intensity plateaus to a nonzero value as Q increases (Doniach, 2001) (Figure 3.8A). The program EOM interprets SAXS data of flexible proteins by selecting ensembles of structures to fit scattering from a large, diverse pool of structures (Bernado et al., 2007). The *Af*Sgt2-TPR crystal structure was linked with 10,000 random C_α models of the C-terminal domain. An ensemble of 50 structures was fit to the data (Figure 3.8B). The R_g and D_{max} distributions (Figure 3.8C,D) of the fitted structures have similar center and shape to the entire random pool of structures, indicating unrestricted flexibility between the TPR and C-terminal domains.

We further used SAXS to confirm the stoichiometry between Sgt2 and Get5. Using data for the purified Sgt2-N/Get5-Ubl complex, the molecular envelope was reconstructed with the program DAMMIF (Figure 3.9A,B). DAMMIF uses dummy atoms to fill a volume that satisfies the SAXS curve and, unlike GASBOR, does not require total residue number as an input, reducing bias of the complex stoichiometry. The averaged model is ellipsoidal, with a long dimension of ~60 Å. This can only be fit with the 40 Å diameter reconstruction of the Sgt2-N dimer (Figure 3.7B) and a single Get5-Ubl of ~20 Å diameter, previously modeled from an NMR structure (Figure 3.9C) (Chartron et al., 2010).

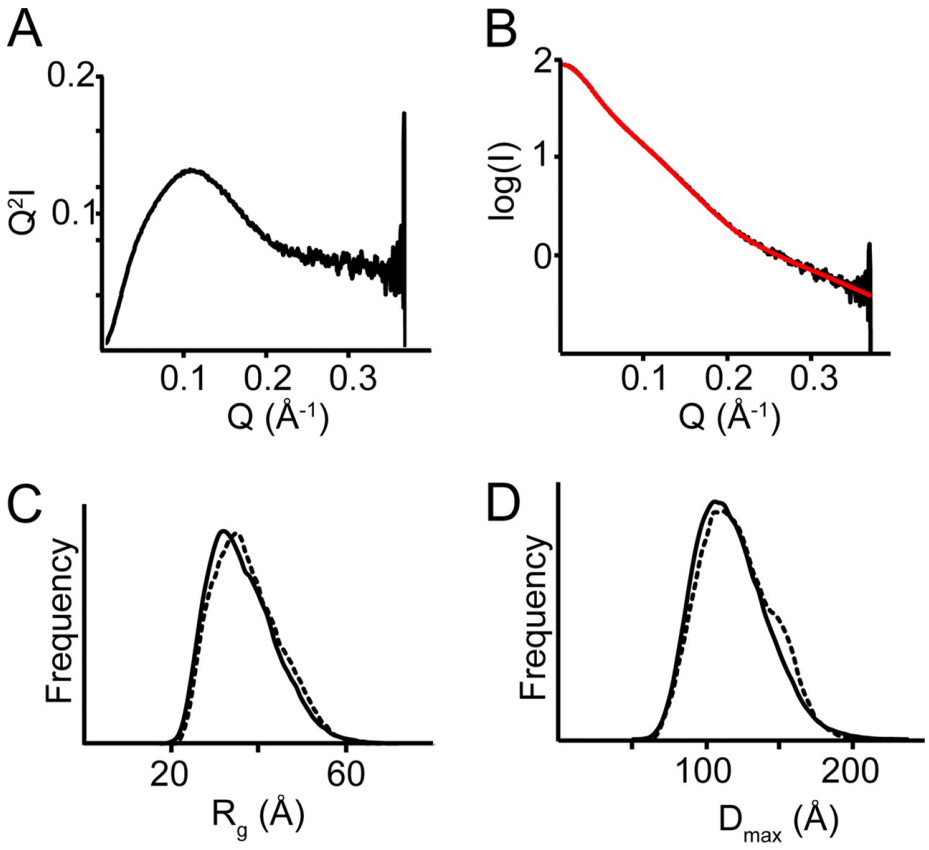


Figure 3.8. SAXS of the flexible Sgt2-TPR-C. (A) Kratky plot; (B) experimental SAXS curve of Sgt2-TPR-C (black) with EOM fit (red); (C) and (D) Rg and Dmax distributions, respectively, for 10,000 random C-domain models relative to the TPR domain (dashed lines) versus 50 models fit to experimental data (solid lines). Correlation suggests unrestricted motion of the C-domain relative to the TPR domain.

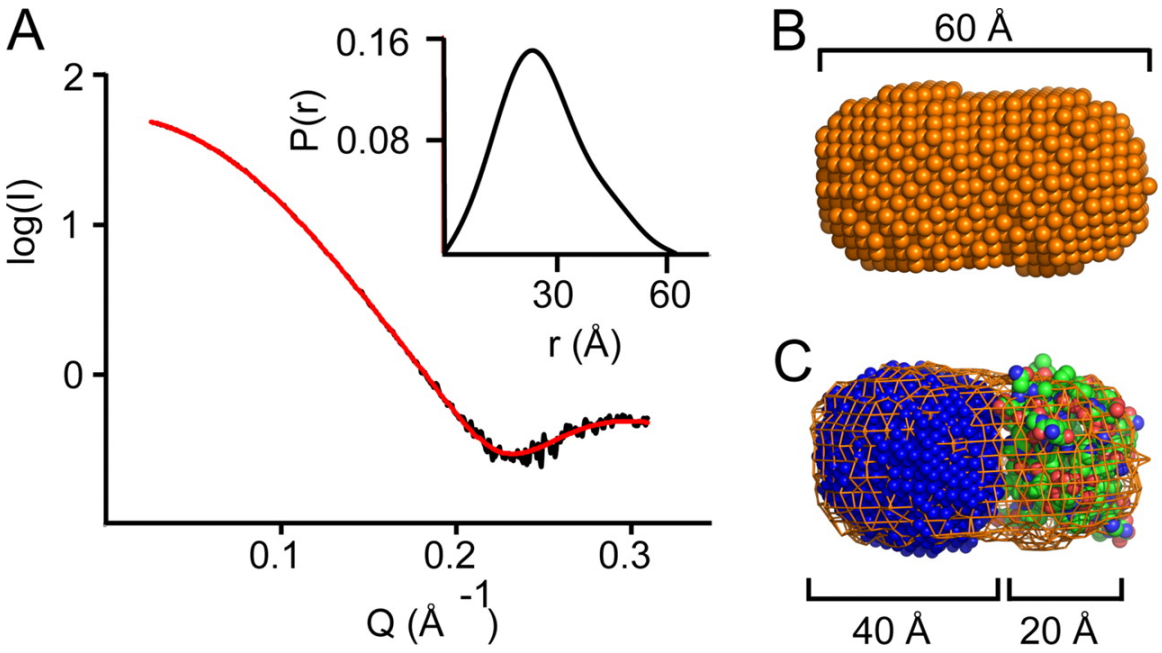


Figure 3.9. SAXS of an Sgt2/Get5 complex. (A) Experimental SAXS curve (black) versus fit of the best DAMMIF model (red). The pair-probability function of Sgt2-N/Get5-U is inset. (B) Filtered average of 20 DAMMIF models. (C) Filtered average GASBOR model of Sgt2-N (blue spheres) and space filling model of a homology model of Get5-Ubl (green) fit into the filtered average model of Sgt2-N/Get5-Ubl (orange mesh).

Discussion

The mechanism for the sorting of TA proteins among the variety of target membranes is only beginning to be understood. Sgt2 appears to selectively bind ER destined TA proteins transferring these substrates with the aid of Get4/Get5 to Get3. Sgt2 contains a TPR domain that physically links the GET pathway to other chaperone pathways. Although other characterized TPR domain-containing co-chaperones are limited in their binding to one or two heat-shock protein families, we demonstrate here that the Sgt2 TPR domain can bind directly to members of at least four families. This promiscuity is explained by the structure of the Sgt2 TPR domain. A phylogenetic analysis of TPR domain sequences indicated that the Sgt2 TPR domain is most similar to the HOP domains (Schlegel et al., 2007) and, indeed, in our structure, bound peptide adopts an identical conformation. The determinants of substrate specificity include the conserved binding pocket formed by residues Met-125 (Met-113), Ala-155 (Ala-143), Ala-156 (Ala-144), Ala-157 (Ala-145), Ser-159 (Ser-147), and Ala-160 (Ser-148). This wider pocket is sterically less restrictive than the HOP TPR domains presumably allowing for the binding of non-canonical two-carboxylate clamp substrates such as the EGDVDMD of Sse1. Moreover, human SGTA TPR can interact with an internal stretch in the androgen receptor that does not contain a clear binding motif suggesting an even broader specificity (Buchanan et al., 2007). In addition to the five two-carboxylate clamp residues, a conserved tyrosine contributes a hydrogen bond to the side chain of the residue corresponding to the terminal residue in Hsp70 or Hsp90. It is possible that this further stabilizes binding to longer sequences.

Is it simply fortuitous that the linker following the *Af*Sgt2 TPR contains the sequence PPADDVDD, resembling an HSC termini? Evidence for a possible functional role is that the sequence is conserved in the Eurotiomycetes family, of which *A. fumigatus* is a member, and *S. cerevisiae* Sgt2 contains the related sequence SRDADVDA (Figure 3.2E). Unfortunately, similar regions are not found more broadly in other fungal homologs, and there are no comparable sequences found in higher eukaryotes. This does not rule out family-specific specialization, perhaps binding other co-chaperones; however, it makes a general conserved role for this sequence unlikely.

Sgt2 forms a direct complex with Get4/Get5 mediated by the dimerization domain of Sgt2 and the Ubl domain of Get5. Despite possible symmetry of Sgt2-N, only a single copy of Get5-Ubl can bind with high affinity. The Vps9-CUE domain is a homodimeric α-helical domain that binds ubiquitin at its symmetry axis and undergoes a conformational change that breaks symmetry to resemble a ubiquitin-associating domain (Shih et al., 2003). The Sgt2/Get5 complex may undergo similar rearrangements. Alternatively, binding of one copy may simply occlude a second binding site. The inability of Sgt2 to bind a second Ubl domain creates a potential for the Get4/Get5 dimer to bind two dimers of Sgt2. *In vitro*, complexes between the minimal binding domains are stable indefinitely; however, as additional domains are included only a single dimer of Sgt2 can bind Get4/Get5, and the full-length proteins do not form a complex stable over SEC. This leaves the overall *in vivo* stoichiometry ambiguous. It is likely that other factors, such as the Get4 to Get3 interaction, may result in a dynamic complex.

Based on the data we present here, we can suggest an updated model for the role of Sgt2 in TA targeting and an overall structure of the Sgt2/Get4/Get5/HSC complex (Figure 3.10). The SAXS analysis of Sgt2-N-TPR suggests that the TPR domains have independent motion. Coupled with the observations that N-domain deletions still bind both HSC and TA proteins (Wang et al., 2010), we conclude that the two Sgt2-TPR-C domains act independently. This would allow for the Sgt2 dimer to bind multiple HSCs simultaneously. The dimerization domain of Sgt2 would bind one Ubl of the Get4/Get5 heterotetramer. Sgt2 will sequester an ER-bound TA and deliver it specifically to Get3.

This work allows us to speculate on the interplay of HSCs and the GET pathway. The low specificity for HSC families suggests that Sgt2 can interact with unfolded proteins distributed among the majority of protein folding pathways in the cell. This may allow Sgt2 to act as a general recovery pathway for TA proteins through its TPR domain, consistent with a recent observation that the TPR domain is not essential for targeting of certain substrates by Get3 under low stress conditions (Kohl et al., 2011). Perhaps more prominently, given the genetic role of Sgt2 in the GET pathway, Sgt2 uses the TPR domain to couple multiple folding pathways with TA targeting. This would allow TA substrates with a variety of folding needs to enter the GET pathway. Combined, this suggests two possible routes for TA proteins through Sgt2: either HSCs bind the TA protein first and deliver them to Sgt2, or Sgt2 binds the TA first and chaperones are

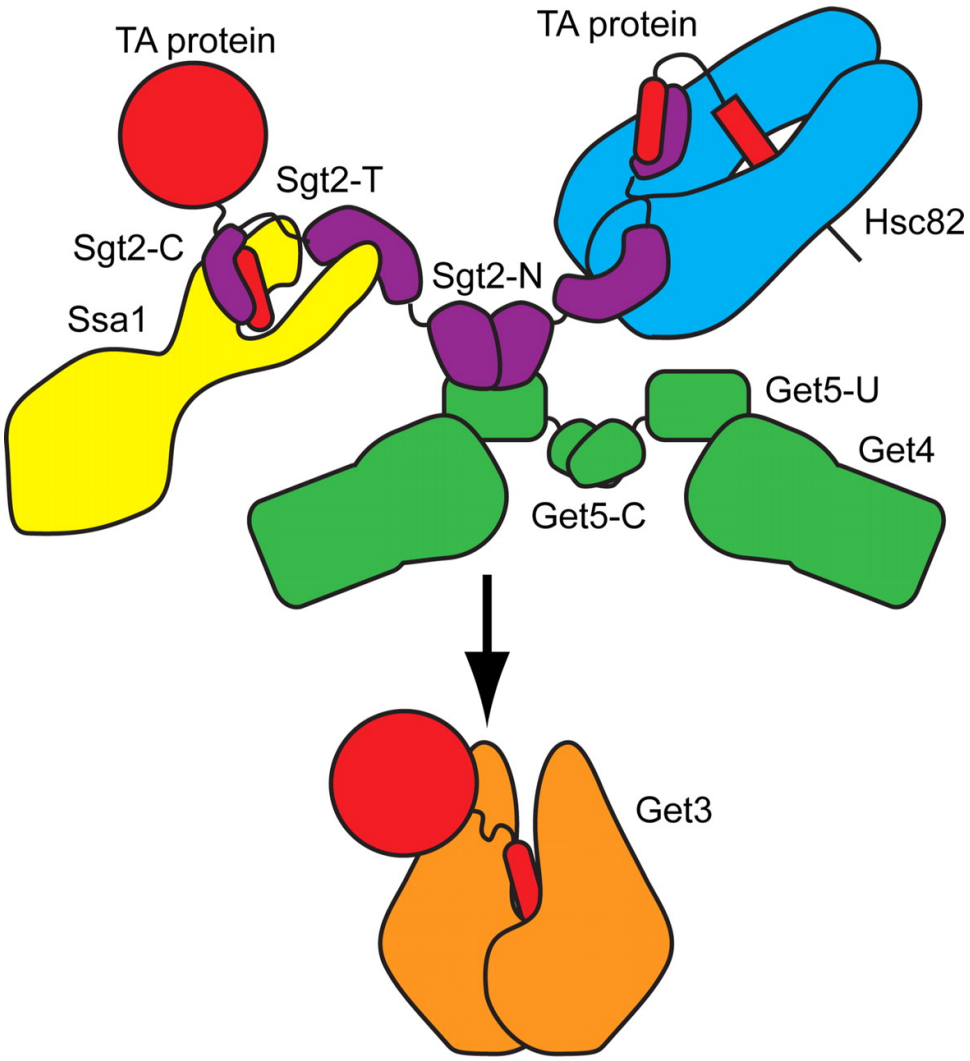


Figure 3.10. Model for the Sgt2/Get4/Get5/HSC complex. Sgt2 (purple) is bound to Get4/Get5 via the Ubl domain (green). TA protein (red) transmembrane domain may initially be in complex with an HSC (e.g., with Ssa1, yellow) with subsequent binding to the Sgt2-TPR and transfer to Sgt2-C. Alternatively, the transmembrane domain initially binds to Sgt2-C and chaperones required for substrate stabilization are bound to the Sgt2-TPR. The complex then releases the TA protein to Get3.

recruited to either aid in folding or to act as acceptors if the TA protein is not ER destined.

TA proteins are diverse with the majority specifically targeted to either the ER or mitochondria. For those destined to the ER, both the GET pathway and Hsp70 operate post-translational targeting pathways, the former being critical as TA hydrophobicity increases (Rabu et al., 2008). Targeting of TA proteins to the mitochondria is less characterized. It has been suggested that this may only depend on Hsp70/90-mediated targeting to the co-chaperone TPR receptors on the mitochondria (Abell and Mullen, 2011; Borgese and Fasana, 2011). This type of targeting could be more general as co-chaperone TPR receptors exist on every organelle (Kriechbaumer et al., 2012; Schlegel et al., 2007). These overlapping or alternative pathways may allow delivery independent of the GET pathway, possibly explaining why none of the individual GET components are essential under optimal conditions. All of this would suggest an important role for co-chaperone TPR proteins in delivery of TA proteins. Sgt2 would be the central co-chaperone acting as a sortase to optimize correct delivery to the GET pathway. If so, TA targeting by Sgt2 may be more akin to panning for gold, retaining substrate proteins for entry to the GET pathway from different points in protein biogenesis.

Methods

Cloning, Expression, and Purification

Get4/Get5 was prepared as previously described (Chartron et al., 2010). Sgt2, Ssa1, Sse1, Hsc82, and Hsp104 were amplified from genomic DNA isolated from *S. cerevisiae* strain S288C (*Af*Sgt2 from *A. fumigatus* strain 118, ATCC) and ligated into pET33b-derived vectors that added N-terminal hexahistidine tags separated by a tobacco etch virus protease site. Truncations were prepared by using QuikChange mutagenesis (Stratagene). Except for the crystallography, all experiments used Sgt2 and its variants from the *S. cerevisiae* homolog.

Sgt2 variants were expressed in *Escherichia coli* BL21(DE3)Star (Invitrogen). Cells were lysed by sonication, and proteins were purified by nickel affinity chromatography (Qiagen). Proteins were digested with tobacco etch virus protease for 3 h at room temperature, and uncut protein and protease were removed by incubation with nickel-nitrilotriacetic acid-agarose beads. Proteins were further purified by size-exclusion chromatography (SEC) using Superdex 200 or Superdex 75 (GE Healthcare) and concentrated to 10–20 mg/ml in 20 mM Tris, 100 mM NaCl, 5 mM 2-mercaptoethanol, pH 8.0. The molecular weight of the *Af*Sgt2 truncation product was determined by LC/MS at the Protein/Peptide MicroAnalytical Laboratory at Caltech.

Ssa1, Sse1, Hsc82, and Hsp104 were expressed in Rosetta(DE3) (Novagen), lysed by sonication in 50 mM K-HEPES, 300 mM KCl, 20 mM imidazole, 5 mM 2-mercaptoethanol, and 0.1% Triton X-100, pH 8.0 and purified by nickel affinity chromatography. Proteins were dialyzed against a buffer containing 0.1 mM EDTA

during cleavage by tobacco etch virus protease and further purified by anion-exchange chromatography (ResourceQ, GE Healthcare) and SEC. Proteins were concentrated to 10–20 mg/ml in 20 mM K-HEPES, 100 mM KCl, 5 mM 2-mercaptoethanol, pH 8.0.

Pulldown Assays

Binding reactions between Get4/Get5 and Sgt2 were performed in 50 µl of a 10% slurry of nickel-nitrilotriacetic acid-agarose in a binding buffer of 20 mM Tris, 100 mM NaCl, 40 mM imidazole, pH 8.0. Per reaction, 8 µM of each protein was mixed for 10 min at room temperature. The resin was washed three times with 100 µl of binding buffer and eluted with binding buffer supplemented with 20 mM EDTA. Reactions between Ssa1, Sse1, Hsc82, and Hsp104 with the Sgt2 TPR domain were performed in 50 µl of 20 mM K-HEPES, 100 mM KCl, 20 mMimidazole, 5 mM $MgCl_2$, 5 mM ADP, 5 mM 2-mercaptoethanol, pH 7.5. 50 µM His-TPR domain were incubated with 5 µM chaperone on ice for 2 h and then added to 5 µl of nickel-nitrilotriacetic acid-agarose. The resin was washed twice with 100 µl of binding buffer and eluted with binding buffer with 300 mM imidazole.

SEC with MALLS

Complexes between Sgt2 and Get4/Get5 were formed in 20 mM Tris, 100 mMNaCl, 5 mM 2-mercaptoethanol, pH 7.0, and resolved by SEC using Superdex 200 or Superdex 75. To generate saturated complexes, 3-fold stoichiometric excesses of the smaller protein were used. Complex peaks were confirmed by SDS-PAGE, concentrated to 10 mg/ml, and separated on a Shodex KW-804 column with multiangle laser light scattering

(MALLS) data collected by a DAWN HELEOS and Optilab rEX detector. Data were processed with ASTRA (Wyatt).

Crystallization

Crystallization screening was performed using the sitting drop vapor-diffusion method with commercially available screens (Hampton Research, Qiagen, Molecular Dimensions) set up by a Mosquito robot (TTP Labtech) then incubated at room temperature. A proteolytic product of *Af*Sgt2 crystallized after 1 week as rectangular prisms against a reservoir of 25% PEG 1500 and 0.1 MMES/malic acid/Tris buffer, pH 4.0 (PACT Premier condition 37), with dimensions of ~50 × 50 × 25 μm. They were soaked in reservoir solution with glycerol added to 10% for 15 min, surrounded with perfluoropolyether PFO-X175/08 (Hampton Research) and flash frozen in liquid nitrogen.

Data Collection, Structure Solution, and Refinement

X-ray diffraction data were collected on beam line 12-2 at the Stanford Synchrotron Radiation Laboratory (SSRL) using a Pilatus 6M pixel array detector at 100 K. A complete dataset was collected from a single crystal to 1.72 Å resolution. Data were integrated, scaled, and merged using MOSFLM (Battye et al., 2011) and SCALA (Evans, 2006). Phases were determined by molecular replacement using the crystal structure of human SGT (PDB ID 2VYI) as a search model by PHASER (McCoy et al., 2007), and the model was rebuilt using RESOLVE (Terwilliger, 2004). The model was refined using COOT (Emsley et al., 2010) and PHENIX (Adams et al., 2010). Secondary structure-matching root mean square deviation values were obtained using COOT. Solvent-

accessible surface area between copies was calculated using PISA (Krissinel and Henrick, 2007). Structure figures were prepared using PyMOL (Delano, 1998).

Small Angle X-ray Scattering

Samples were concentrated to 25 mg/ml and filtered through 0.22 μm membranes. Overnight dialysis was performed against 50 mM Tris, 100 mM NaCl, 5 mM 2-mercaptoethanol, pH 8.0. Dilutions to 1, 2, 5, and 10 mg/ml were made using the dialysate, and concentrations were determined using an ND-1000 spectrophotometer (Nanodrop Technologies) and theoretical extinction coefficients derived from protein sequences.

Data were collected at SSRL beam line 4-2 using a Rayonix MX225-HE detector, 1.13 Å wavelength X-rays, and a detector distance of 2.5 m. For each concentration, 20 exposures of 1 s were collected covering a momentum transfer range of 0.0055–0.3709 $Å^{-1}$. Data were reduced, averaged, and buffer-subtracted using MARPARSE (Smolsky et al., 2007). Extrapolation to infinite dilution and merging were performed with PRIMUS (Konarev et al., 2003). Guiner analysis was performed using AutoRG, and distance distribution functions were determined with GNOM (Svergun, 1992). *Ab initio* reconstructions were performed using software available in the ATSAS package (Konarev et al., 2006). Sgt2-N and Sgt2-N-TPR were reconstructed with imposed 2-fold symmetry as an additional constraint on the data. Sgt2-N/Get5-Ubl was reconstructed without imposed symmetry.

Acknowledgments

We thank D. C. Rees and S. O. Shan for critical reading of the manuscript. We thank members of the laboratory for support and useful discussions. We thank Graeme Card, Ana Gonzalez, and Michael Soltice for help with data collection at SSRL BL12-2, Tsutomu Matsui and Hiro Tsuruta for help with data collection and processing at the bioSAXS SSRL BL4-2, and Troy Walton for help with MALLS. We are grateful to Gordon and Betty Moore for support of the Molecular Observatory at Caltech.

Chapter 4

STRUCTURES OF THE SGT2/SGTA DIMERIZATION DOMAIN WITH THE GET5/UBL4A UBL DOMAIN REVEAL A NOVEL INTERACTION THAT FORMS A CONSERVED DYNAMIC INTERFACE

Abstract

In the cytoplasm, the correct delivery of membrane proteins is an essential and highly regulated process. The post-translational targeting of the important tail-anchor membrane (TA) proteins has recently been under intense investigation. A specialized pathway, called the GET pathway in yeast and the TRC pathway in vertebrates, recognizes ER targeted TA proteins and delivers them through a complex series of handoffs. An early step is the formation of a complex between Sgt2/SGTA, a co-chaperone with a presumed ubiquitin-like-binding domain (UBD), and Get5/Ubl4a, a ubiquitin-like domain (UBL) containing protein. We structurally characterize this novel UBD/UBL interaction for both the yeast and human proteins. This is supported by biophysical studies that demonstrate that complex formation is mediated by electrostatics generating an interface that has high affinity with rapid kinetics. In total, this work provides a refined model of the interplay of Sgt2 homologs in TA targeting.

Adapted from

Chartron, J.W., VanderVelde, D.G., and Clemons, W.M., Jr. (2012). Structures of the Sgt2/SGTA dimerization domain with the Get5/Ubl4a UBL domain reveal a novel interaction that forms a conserved dynamic interface.

Introduction

Two homologous pathways have been elucidated for the targeting of TA proteins to the ER (recently reviewed in Chartron et al., 2012a; Hegde and Keenan, 2011). The best characterized is the fungal GET pathway (Guided Entry of TA proteins). In its simplest form, a TA sorting complex comprising Sgt2, Get4, and Get5 (alternatively name Mdy2) facilitates the loading of a TA substrate onto the Get3 ATPase. The Get3/TA complex is targeted to the ER where the TA is released for insertion by the membrane proteins Get1 and Get2. Vertebrates have a related system, referred to as the TRC pathway (Transmembrane domain Recognition Complex). In this case, the sorting complex similarly contains the proteins TRC35 and Ubl4a/Gdx, homologs of Get4 and Get5 respectively; however, they form a three-component complex with the protein Bag6. From here, the TA is handed to a Get3 homolog, TRC40, that delivers the protein to WRB, a homolog of Get1.

Sgt2, a heat-shock protein (HSP) co-chaperone, facilitates the first committed step in TA protein targeting. It recruits a variety of HSP families via an internal tetratricopeptide repeat (TPR) domain (Chartron et al., 2011; Wang et al., 2010). The Sgt2 C-terminal domain can bind to sequences of six or more hydrophobic residues (Liou and Wang, 2005) including ER destined TA proteins, which are then handed to Get3 (Wang et al., 2010). Mitochondrial TA proteins, which can also co-purify with the Sgt2/Get4/Get5 complex, are associated with bound HSPs and are not transferred to Get3. The role of Sgt2 in sorting between target organelles is supported by the mislocalization of ER resident TA proteins to the mitochondria in *Δsgt2* cells (Costanzo et al., 2010). The C-

terminal domain, rich in asparagine, glutamine and methionine, contains only a short conserved sequence and is weakly predicted as helical. Sgt2 is flexible, especially between the TPR and C-terminal domains (Chartron et al., 2011; Liou and Wang, 2005). A small N-terminal homodimerization domain (Sgt2-N) mediates the association with Get5, where a dimer of Sgt2 binds a single copy of Get5 (Chang et al., 2010; Chartron et al., 2011; Liou et al., 2007).

Get4 and Get5 form an adaptor required for the transfer of TA proteins from Sgt2 to Get3 (Wang et al., 2010). Get4 forms a complex via the N-terminal domain of Get5 and sequesters a nucleotide bound Get3 (Bozkurt et al., 2010; Chang et al., 2010; Chartron et al., 2010). Get5 has a central ubiquitin-like domain (Get5-UBL) that binds Sgt2-N and a C-terminal homodimerization domain, resulting in an extended Get4/Get5 heterotetrameric complex (Chartron et al., 2010; Chartron et al., 2012b). Disruption of the Get5-UBL/Sgt2-N interaction leads to incomplete rescue of *Δget5* growth defects under stress conditions (Chartron et al., 2010; Liou et al., 2007). Moreover, the TA protein transfer reaction is competed by excess Get5, which cannot alone form a productive complex with Get3 (Wang et al., 2010). These results underscore the importance of the physical interaction between Sgt2 and the Get4/Get5 complex. The molecular details of the transfer of TA proteins, including energetic requirements and the *in vivo* stoichiometry of Sgt2, Get4/Get5 and Get3 over the course of the hand off, remain to be established.

The yeast Get4/Get5/Sgt2/HSP complex has also been dubbed the TRC (Wang et al., 2010). An analogous complex to Get4/Get5/Sgt2 exists in vertebrates; however, with

different organization and additional functions. The homologs of Get4 and Get5, named TRC35 and Ubl4a, along with Bag6 (alternatively named Bat-3 or Scythe) form the Bag6 complex (Mariappan et al., 2010). Bag6 is a large, ~1100-residue protein with an N-terminal UBL domain, a C-terminal BAG domain, and internal proline rich regions. SGTA, the homolog of Sgt2, associates with Bag6 through its N-terminus (Winnefeld et al., 2006) and Ubl4A is postulated to bridge this interaction (Chartron et al., 2012a; Hegde and Keenan, 2011). The Bag6 complex has the ability to load TA proteins onto TRC40 (Leznicki et al., 2010; Mariappan et al., 2010). In addition, it is involved in the degradation of mislocalized membrane and other defective proteins (Hessa et al., 2011; Minami et al., 2010; Wang et al., 2011).

Here we use a combination of structural biology and biochemistry to define the interaction between Sgt2 and Get5 homologs. We present the first structures of the Sgt2-N and SGTA-N homodimerization domains and characterize them as a new class of UBDs. Further, we solve the structure of the Get5-UBL demonstrating it as a novel UBL. Finally, we determined the structure of a complex between the Sgt2-N dimer and the Get5-UBL revealing an interaction strongly influenced by electrostatics. We use a variety of methods to demonstrate that this interaction has high affinity with rapid binding kinetics. This provides critical context for understanding the Sgt2/Get4/Get5 complex. Similarly, this work provides a mechanism for flux into the seemingly more complicated mammalian system.

Results

The structures of Sgt2 and SgtA dimerization domains

The Sgt2-N domain is both a homodimerization domain and a binding platform for the Get5-UBL domain. This provides the physical link between the Sgt2 chaperone complex and the GET pathway. Sgt2-N does not have sequence homology to known structures or other characterized UBDs. To understand how these dual functions are accomplished, we first determined the structure of *S. cerevisiae* Sgt2-N homodimer using solution NMR. Statistics for NMR structure calculations are in Table 4.1. A monomer of Sgt2-N consists of three helices (Figure 4.1A–C). The first two helices are of similar length and mediate homodimerization, forming a four-helix bundle with 2-fold symmetry, consistent with the postulated coiled-coil (Tobaben et al., 2003). A third, shorter helix packs against either side of the bundle away from the dimer interface. The residues C-terminal to this helix give weaker signals than the helical region in all NMR experiments and only partial chemical shift assignments were possible. This is due to different rates of motion relative to the rest of the protein rather than proteolysis, as signals for the terminal residues are observed. The few NOE derived contacts observed in this sequence restrain it in partially folded conformations (Figure 4.1C).

The symmetry axis places the equivalent helices from each subunit head-to-tail resulting in two unique surfaces. We designate the surface comprised of the α2 helices the Get5 binding surface, and the opposite comprised of the α1 helices the α1 surface (Figures 4.1A and 4.2A,B). The dimer contacts in α2 are made by conserved small residues (Ser32, Ala36, Cys39, and Ala43), resulting in close packing between the main chains of

Table 4.1. NMR structural constraints and structure statistics

	Sgt2-N	Get5-UBL
No. of restraints[a,b]		
NOE-based distance restraints		
Intra-residue ($\|i - j\| = 0$)	1008	663
Sequential ($\|i - j\| = 1$)	580	271
Medium range ($2 \leq \|i - j\| < 5$)	730	183
Long range ($\|i - j\| \geq 5$)	348	334
Inter-molecular (homodimer)	120	-
Ambiguous	716	321
Total	3502	1772
$\phi + \psi$ dihedral angle restraints	240	124
Hydrogen bond restraints	100	52
Residual dipolar coupling restraints	88	-
Magnitude (D_a)	7.0	-
Rhombicity (R)	0.54	-
Restraints statistics		
r.m.s.d. from experimental distance restraints	0.008 ± 0.001 Å	0.018 ± 0.002 Å
r.m.s.d. from experimental dihedral restraints	0.3 ± 0.1°	0.5 ± 0.1°
Cross-validated RDC Q-factors[c], (No. used for validation)	0.11 ± 0.03 (18)	-
Model Statistics		
Residue Range[d]	5-56 (x2)	73-149
Coordinate precision r.m.s.d.		
Backbone	0.33 ± 0.07 Å	0.34 ± 0.09 Å
Heavy atom	0.73 ± 0.08 Å	0.74 ± 0.09 Å
Structural quality		
Ramachandran statistics[e]		
Most favored regions	99.00%	83.10%
Allowed regions	1.00%	15.60 %
Generously allowed regions	0.00%	0.00%
Disallowed regions	0.00%	1.70%[g]
WHAT-IF Z-score[f]		
Backbone conformation	1.306 ± 0.417	–1.428 ± 0.327
2nd generation packing quality	5.440 ± 1.865	5.730 ± 2.413
Ramachandran plot appearance	0.169 ± 0.418	–3.309 ± 0.319
$\chi1/\chi2$ rotamer normality	–1.610 ± 0.588	–3.189 ± 0.537

[a]For Sgt2-N, a single NOE generates two restraints for atom pairs from each monomer.

[b]Structures were calculated using the full sequence of the protein.

[c]Q-factors were calculated with PALES using RDCs omitted from structure calculations (Zweckstetter and Bax, 2000)

[d]Validation was performed over the ordered sequence of the protein.

[e]Ramachandran statistics calculated by PROCHECK (Laskowski et al., 1993)

[f]Z-scores were calculated by WHAT-IF (Vriend and Sander, 1993). Positive scores are better than average and negative scores are worse

[g]K85, which follows *cis* P84, is the only residue in the disallowed region. TALOS dihedral angle restraints were omitted for P84 and K85.

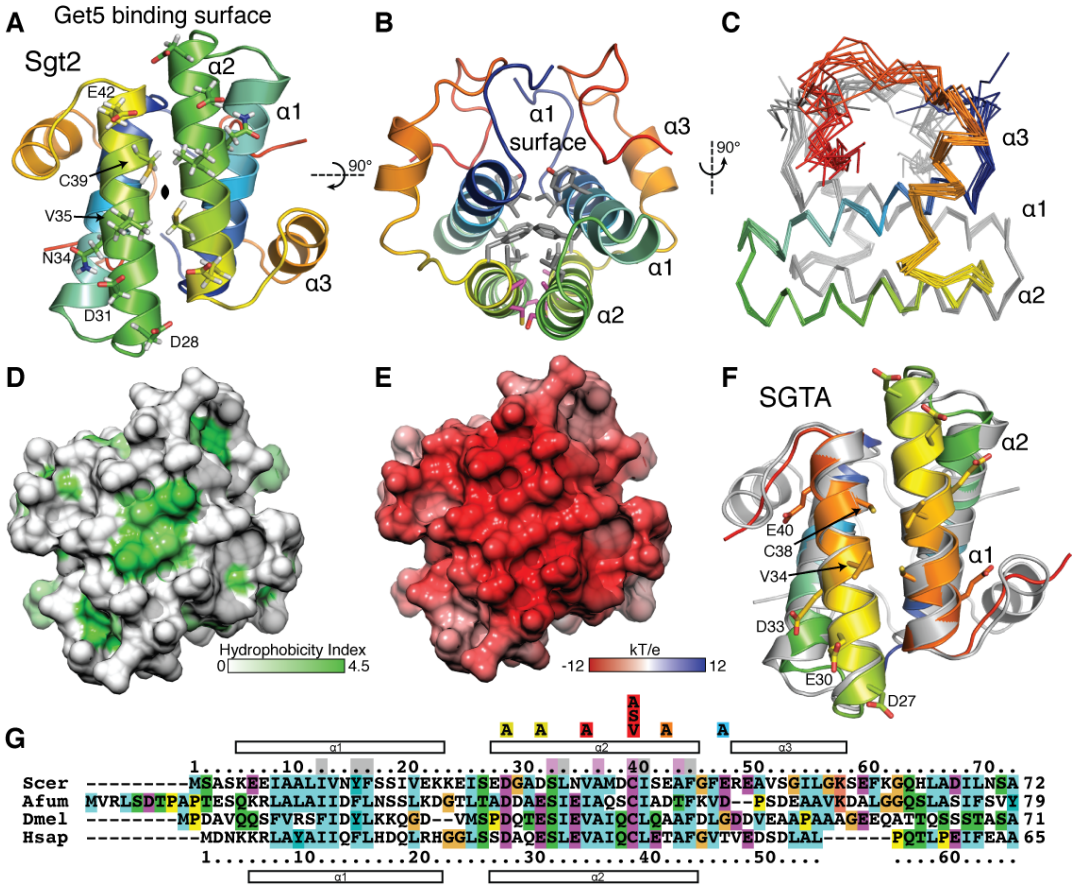


Figure 4.1. Atomic structure of the Sgt2-N/SGTA-N dimerization domain. (A) Solution NMR structure of *S. cerevisiae* Sgt2-N orientated to show the conserved binding face. The cartoon diagram of a representative structure is color ramped from N- to C-terminus (blue to red) for each subunit of the dimer. Conserved residues are shown as sticks with noncarbon atoms colored. (B) Similar to (A) rotated 90° forward looking down the bundle axis. Interior small (magenta) and bulky (gray) residue side chains are highlighted as sticks. (C) Ribbon diagram of the overlaid ensemble of the ten lowest-energy NMR structures rotated 90° to the right relative to (B). One subunit is color ramped and one is in gray. (D) Surface representation of the binding face highlighting exposed hydrophobicity. The surface is color ramped from 0 (gray) to 4.5 (green) on the Kyte-Doolittle hydrophobicity scale (Kyte and Doolittle, 1982). (E) Surface representation of the binding face showing surface charge. The surface is color ramped based on electrostatic potential colored from negative (red) to positive (blue). (F) X-ray crystal structure of human SGTA-N aligned to Sgt2-N (gray) similar to (A). (G) Sequence alignment of Sgt2-N homologs. The species are Scer (*S. cerevisiae*), Afum (*Aspergillus fumigatus*), Dmel (*D. melongaster*) and Hsap (*H. sapiens*). Alignment and residue coloring are based on ClustalX output (Larkin et al., 2007). Numbering and secondary structure (rectangles for helices) are indicated above or below the corresponding sequence. Mutations tested in Figure 4.5 are highlighted in the numbers colored based on effect.

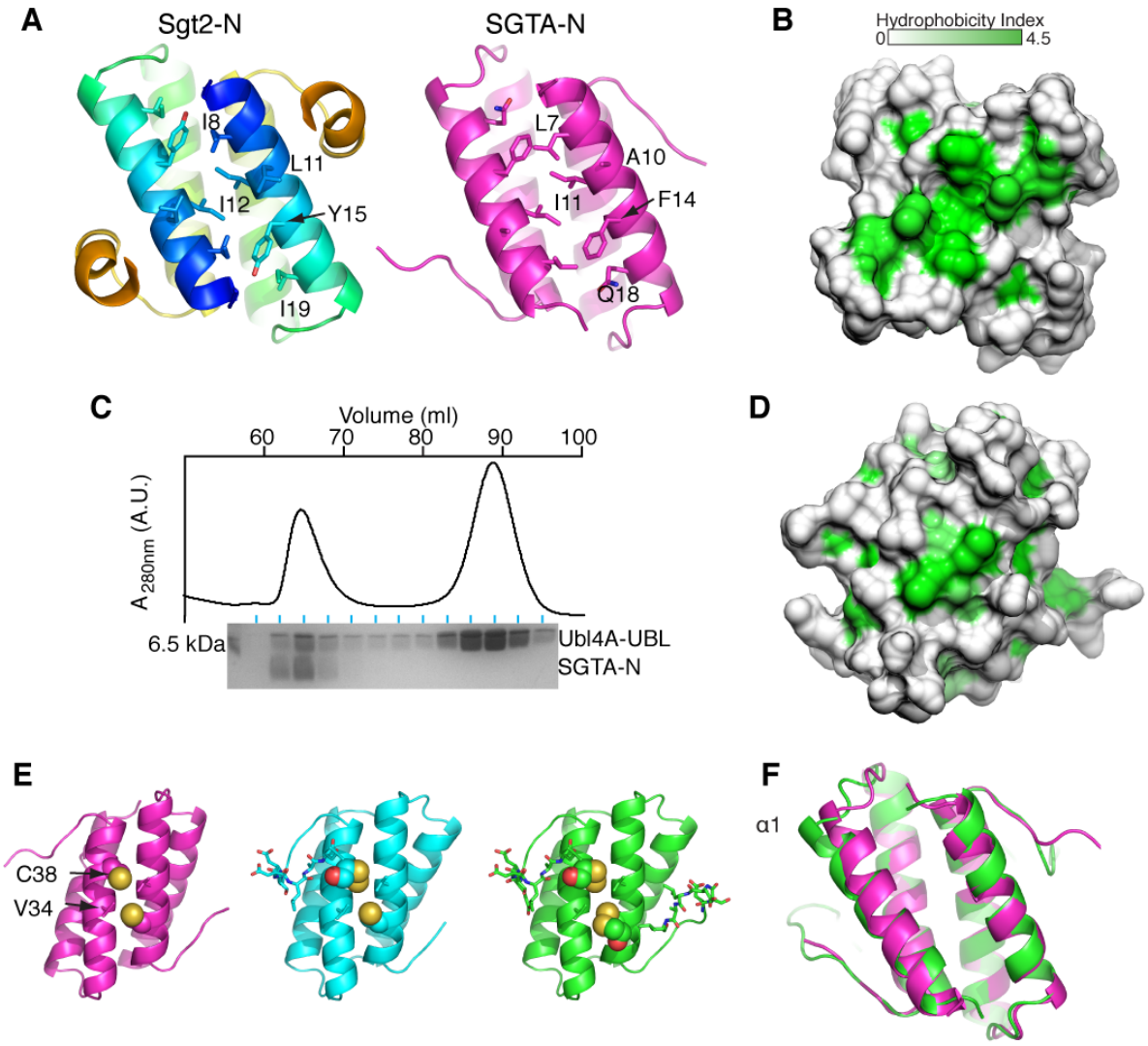


Figure 4.2. Comparison of the structures of Sgt2-N and SGTA-N. (A) View of the α1 surface. Sgt2-N is color ramped from N- (blue) to C- (red) termini, as in Figure 4.1A. (B) Surface representation of the α1 surface of Sgt2-N. The surface is color ramped from 0 (gray) to 4.5 (green) on the Kyte-Doolittle hydrophobicity scale. (C) SgtA-N was co-expressed and co-purified with affinity tagged Ubl4A-UBL. Shown here is the size exclusion chromatogram of the separation of excess Ubl4A-UBL from Ubl4A-UBL in complex with SGTA-N. Coomassie stained SDS-PAGE of indicated 1 ml fraction ranges is shown below. The complex peak was used for crystallization trials, although only crystals of the co-purified SgtA were obtained. (D) Hydrophobic representation of the Ubl4A-UBL binding surface of SGTA-N, colored as (B). (E) Comparison of the crystal structures of SGTA-N with fully reduced C38 (magenta), single (cyan) or double (green) β-mercaptoethanol covalent adducts. (F) Overlay of the reduced (magenta) and fully oxidized (green) SGTA-N structures. The view is of the α1 surface. Along with the reconfiguration of the C-terminal residues, the packing of α1 is altered in the covalently modified structures.

the two subunits (Figure 4.1B). Cys39 and Val35 make a small hydrophobic surface at the center of the Get5 binding face (Figure 4.1A,D). The partially exposed Cys39 is strictly conserved across eukaryotes; Val35 is fully solvent exposed and is conserved as either valine or isoleucine. Despite the close proximity of the Cys39 sulfhydryl groups (5.3 ± 0.4 Å sulfur-sulfur distance), the C_β shifts of 27.819 ppm argue against disulfide bond formation, even after incubation for several months (Sharma and Rajarathnam, 2000). The conserved acidic residues Asp28 and Glu31 and the conserved positioning of Glu42, Glu47 and Glu49 result in a negatively charged ring surrounding the hydrophobic patch (Figure 4.1A and E).

The helices of the α1 surface are held further apart than the α2 interface by the interlocking of large hydrophobic side chains (Figures 4.1B and 4.2A). This results in an exposed hydrophobic surface that is protected, to some extent, from solvent by the partially folded carboxyl terminal linker (Figure 4.1C and 4.2B). When an Sgt2-N variant with this linker deleted is purified, it forms a higher order oligomer, likely due to aggregation at this face (data not shown).

We also investigated the N-terminal domain of human SGTA (SGTA-N). The domain was co-expressed and purified with an affinity tagged UBL domain of Ubl4A (Ubl4a-UBL), demonstrating a stable complex that could be separated from excess Ubl4A (Figure 4.2C). We attempted to crystallize this complex; however, in the three resulting distinct crystal structures, only density for SGTA-N was observed. Structures were determined by molecular replacement using Sgt2-N at 1.35-1.45 Å resolutions. The structures differ predominantly by a post purification oxidation of the conserved cysteine

with buffer components (Figure 4.2E,F). Crystallographic data collection and refinement statistics are provided in Table 4.2. SGTA-N shares the four-helix bundle topology with Sgt2-N (Figure 4.1F). As expected from the solution data, the cysteine sulfhydryl groups do not form a disulfide bond between subunits.

As was the case for structures of the Sgt2/SGTA-TPR domains (Chartron et al., 2011; Dutta and Tan, 2008), Sgt2-N and SGTA-N have very similar architecture with an r.m.s.d. of 1.24 ± 0.07 Å over equivalent C_α atoms (Figure 4.1F). The sequences have high homology (Figure 4.1G) and all of the general features are conserved including the hydrophobic patch surrounded by charge at the binding face (Figures 4.1F and 4.2D). One notable difference is that SGTA-N does not have a third α-helix. The residues that would correspond to α3 are disordered or involved in non-physiological crystallographic contacts.

The structure of the Get5-UBL domain

The UBL of Get5 and Ubl4A have a number of features that make them unique compared to other UBLs with high homology to ubiquitin. Get5 binds Sgt2 via its UBL and this interaction includes Leu120, the equivalent of Ile44 of ubiquitin (Chang et al., 2010; Chartron et al., 2011). The solution NMR structure of the Ubl4A-UBL domain showed that it has the expected ubiquitin fold (Figure 4.3A) (PDBID:2DZI, RIKEN Structural Genomics Initiative). Get5-UBL has several small sequence insertions suggesting some structural differences (Figure 4.3F); therefore, to fully characterize the yeast system, we determined the structure of the Get5-UBL domain. Initially, we solved a structure by solution NMR (Figure 4.4A). Simultaneously, we obtained crystals of Get5-UBL that

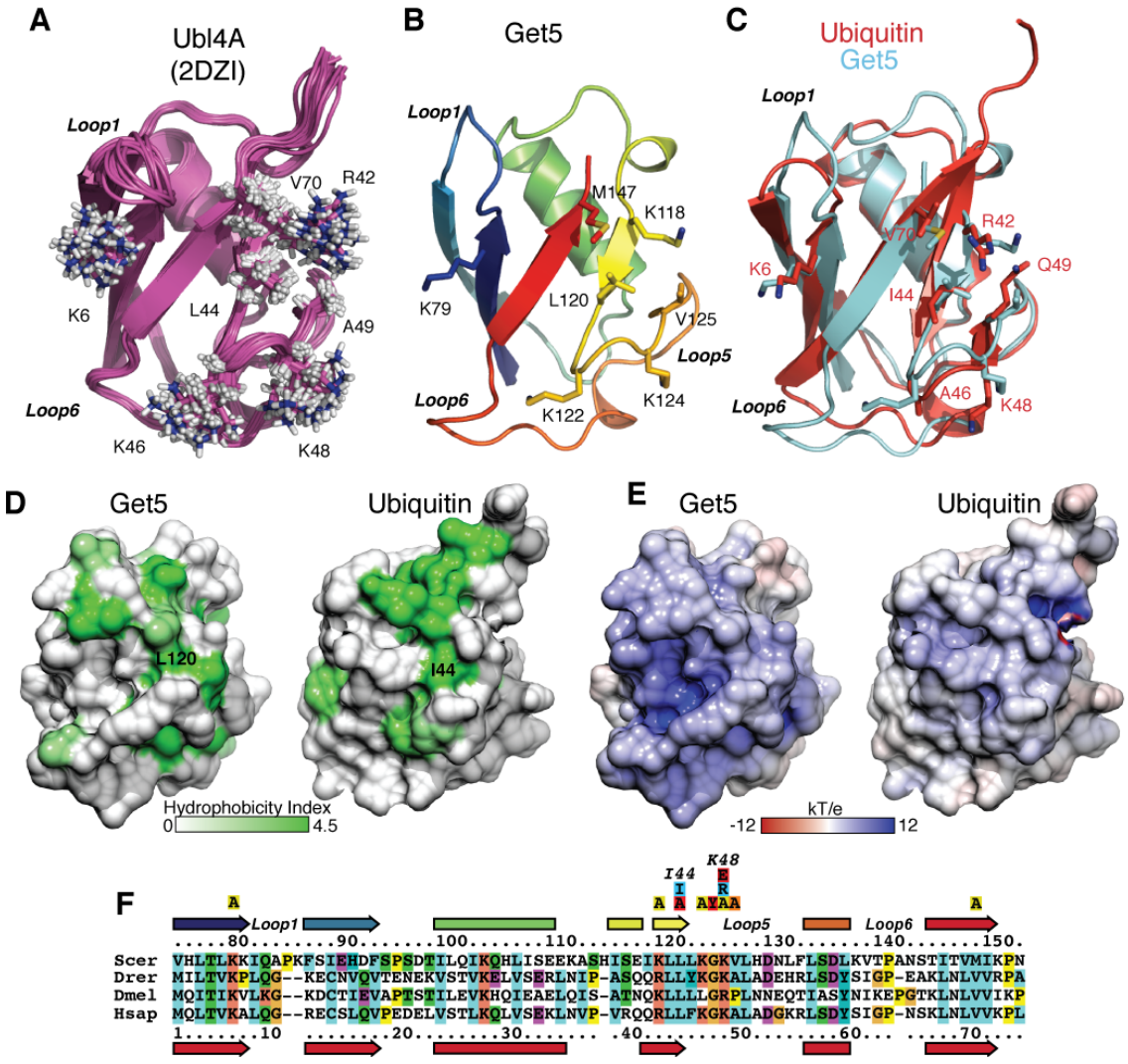


Figure 4.3. Atomic structure of the Get5-UBL/Ubl4A-UBL domain. (A) Ribbon diagram of the human Ubl4A-UBL domain solution structure (PDBID: 1DZI). Residues equivalent to those tested in Figure 4.5 are highlighted as sticks in (A)-(C). (B) X-ray crystal structure of the *S. cerevisiae* Get5-UBL domain. The cartoon of chain A is color ramped between termini. (C) Overlay of a representative structure of ubiquitin (red) is shown (PDBID:1UBQ) as a ribbon diagram with Get5-UBL (light blue). (D) Surface representation highlighting exposed hydrophobicity of Get5 and ubiquitin as in Figure 4.1D. (E) Surface representation showing surface electrostatic potential of Get5 and ubiquitin as in Figure 4.1E. (F) Sequence alignment of Get5-UBL homologs. Sequences are displayed similar to Figure 4.1G. In addition, beta-sheets are shown as arrows, Drer is *Danio rerio*, and the residues equivalent to the important ubiquitin residues Ile44 and Lys48 are labeled.

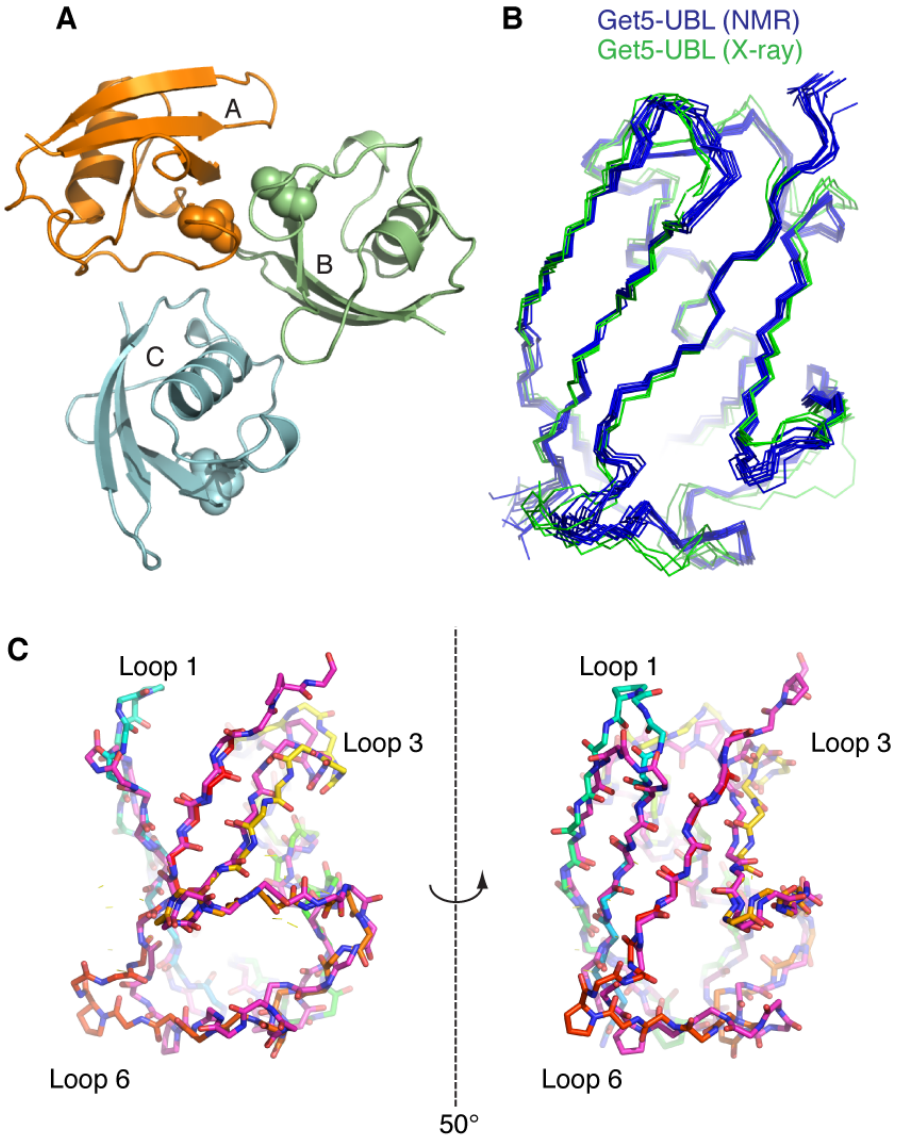


Figure 4.4. Comparison of the structures of Get5-UBL and Ubl4A-UBL. (A) Overlay of the backbone nitrogen and carbon atoms of the ensemble of the 10 lowest energy NMR structures of Get5-UBL (blue) and the three copies of the Get5-UBL in the asymmetric unit of the crystal structure (green). Chain A is used in figures and in discussion of the Get5-UBL structure. (B) The three molecules found in the asymmetric unit of the Get5-UBL crystal structure with Leu120 shown as spheres for reference. (C) Overlay of the main-chain atoms of the crystal structure of Get5-UBL (carbons color ramped blue to red from N- to C-terminus) and a representative structure of Ubl4A-UBL (magenta carbons). Insertions in Loops 1, 3, and 6 of Get5-UBL cause most of the structural differences between the homologs.

diffracted to 2.4 Å. We were able to obtain phases by molecular replacement using the Get5-UBL solution structure. Three copies of Get5-UBL were present in the asymmetric unit (Figures 4.3B and 4.4B). They have an average r.m.s.d. of 0.75 Å over main chain atoms. Statistics for the solution structure calculations and crystallographic data collection and refinement are presented in Tables 4.1 and 4.2, respectively. The solution and crystal structures are very similar with an average main chain r.m.s.d. of 1.23 Å (Figure 4.4A), with most variation in loops 1 and 5.

While overall the structures from Get5 and Ubl4A are similar, there are a few differences due to insertions (Figure 4.4C). Based on the structures and sequences of other animal homologs, Get5-UBL has a two-residue insertion around Pro84 in Loop 1. Pro84 is *cis* allowing for a tight turn, and there is no detectable *cis* to *trans* isomerization by NMR. There are two additional insertions in Get5-UBL, His113 in Loop 3, and Ala141 that causes a short coil-like turn to extend at the end of Loop 6. Despite these differences, the surface elements on the face of the β-sheet are conserved between Get5 and Ubl4A (Figure 4.3A,B).

Most UBLs share some functional roles with ubiquitin. They occur either as independent units known as Type I UBLs that can be conjugated onto other proteins or as Type II UBLs that are domains in larger proteins that frequently mediate binding to the proteasome (Jentsch and Pyrowolakis, 2000). Get5 falls into the latter class but does not associate with the 20S proteasome or polyubiquitin chains (Hu et al., 2006; Saeki et al., 2002). Therefore, one would expect that there are significant features that distinguish

Table 4.2. Crystallographic statistics

	SGTA Residues 4-54			Get5-UBL Residues 74-150
Dataset	**Reduced Cys-38**	**Mercaptoethanol adduct**	**2x Mercaptoethanol adduct**	**Native**
Wavelength (Å)	1.0000	1.0000	1.0000	1.0000
Resolution range (Å)	28.22-1.40 (1.42-1.40)[a]	29.73-1.45 (1.47-1.45)	19.39-1.35 (1.37-1.35)	29.07-2.40 (2.46-2.40)
Space group	P 2 2_1 2_1	P 2 2_1 2_1	P 2 2_1 2_1	P 2_1 2_1 2_1
Cell parameters: *a*, *b*, *c* (Å)	23.60, 57.05, 64.95	29.73, 44.13, 63.98	29.61, 43.61, 63.47	49.36, 63.62, 71.58
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
Unique reflections	17693 (851)	15303 (767)	18421 (876)	9092 (666)
Completeness (%)	98.7 (98.9)	98.8 (99.9)	98.9 (99.6)	98.2 (98.7)
Redundancy	3.1 (3.2)	3.6 (3.6)	3.9 (3.6)	3.3 (3.3)
R_{merge}[b]	0.078 (0.640)	0.093 (0.728)	0.058 (0.692)	0.133 (0.613)
Mean I/σ(I)	6.7 (1.6)	7.8 (1.8)	10.1 (1.7)	6.6 (1.8)
Refinement				
Reflections: work/free	16385/1281	14183/1093	17298/1090	8177/907
R_{work}/R_{free}[c]	0.1965/0.2083	0.1808/0.2030	0.1806/0.1984	0.1737/0.2453
No. protein atoms	726	755	811	1803
No. water atoms	110	107	104	102
No. ligand atoms	16	1	1	0
Protein B-factors (Å2)	16.4	14.5	16.3	22.5
Water/ligand B-factors (Å2)	29.1	26.1	25.1	25.5
r.m.s.d. of bond lengths (Å)	0.006	0.006	0.006	0.008
r.m.s.d. of bond angles (°)	1.02	0.98	1.04	1.15
Ramachandran favored	100%	98%	96%	98%
Ramachandran outliers	0%	0%	0%	0%

[a]Values in parentheses are for the highest resolution shell

[b]Rmerge = $\Sigma_{hkl} \Sigma_i |I_i(hkl) - \langle I(hkl)\rangle| / \Sigma hkl \Sigma_i I_i(hkl)$, where $I_i(hkl)$ is the ith observation of reflection hkl and $\langle I(hkl)\rangle$ is the weighted average intensity for all observations i of reflection hkl.

[c] R = Σ(| Fobs - Fcalc |)/Σ| Fobs |, where Fobs and Fcalc are observed and calculated structure factors amplitudes, respectively

Get5 homologs from ubiquitin. Comparing Get5-UBL to ubiquitin, the two most significant structural differences are the conformations of Loop1 and Loop6 (Figure 4.3C). Ubiquitin is noted for the hydrophobic "I44 patch." In Get5, the conformations of Loop1 and Loop6 significantly reduce the size of the equivalent patch (Figure 4.3D). Moreover, the surface charge of both proteins around the I44 patch has a positive charge; however, it is significantly more pronounced in Get5 (Figure 4.3E).

Characterization of the putative interface between Sgt2-N and Get5-UBL

We previously demonstrated that only a single copy of Get5-UBL binds to dimeric Sgt2-N and the double mutation L120A/K122A prevents complex formation (Chartron et al., 2011). We decided to probe the interaction further using isothermal titration calorimetry (ITC) of wild type and mutant proteins (Figure 4.5 and 4.6). All of the variants behaved similarly to wild type during purification (data not shown). Sgt2-N and Get5-UBL interact with 10^{-8} M affinity independent of which protein is used as a titrant. The interaction is slightly exothermic, releasing approximately 1.4 kcal/mol. The Get5-UBL L120A mutation has a thousand fold lower binding affinity consistent with the significance of this position in UBLs. The Get5-UBL L120I mutant bound Sgt2-N with similar affinity to wild type. This is surprising considering that a leucine at this position is completely conserved, in contrast to the isoleucine in most UBLs. Mutations of three other nearby hydrophobic residues that compose part of the hydrophobic patch, G123Y, V125A, and M147A also significantly lowered the binding affinity.

The complementing surface charge of the two proteins suggests that a significant component of the interaction involves electrostatics. As expected for this type of binding,

Sgt2	Get5	[NaCl] (mM)	K_d (nM)
wt[a]	wt	100	34
wt	wt	150	89
wt	wt	200	158
wt	wt	300	1317
wt	K79A	100	136
wt	K118A	100	191
wt	L120A	100	17100
wt	L120I	100	20
wt	K122A	100	463
wt	G123Y	100	17200
wt	K124A	100	181
wt	V125A	100	916
wt	M147A	100	497
wt	*wt*	100	15
D28A	*wt*	100	183
D31A	*wt*	100	175
V35A	*wt*	100	2060
C39A	*wt*	100	4320
C39S	*wt*	100	10600
C39V	*wt*	100	1470
E42A	*wt*	100	720
E47A	*wt*	100	16

[a]Italicized proteins were titrants

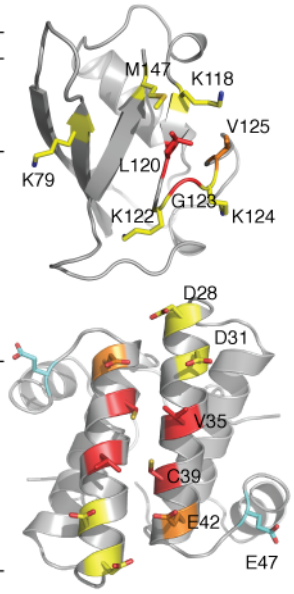


Figure 4.5. Importance of conserved residues of both Sgt2 and Get5 at the putative binding interface by ITC. The table is a summary of the data obtained by ITC where mutants were tested for binding affinity. Wild type is represented by "wt." Mutations tested are colored based on the loss of binding affinity (red; strong effect, orange; moderate effect, yellow; mild effect, blue; no significant effect). On the right, mutations are shown colored based on effect as sticks on a ribbon diagram of Get5-UBL (top) and Sgt2-N (bottom).

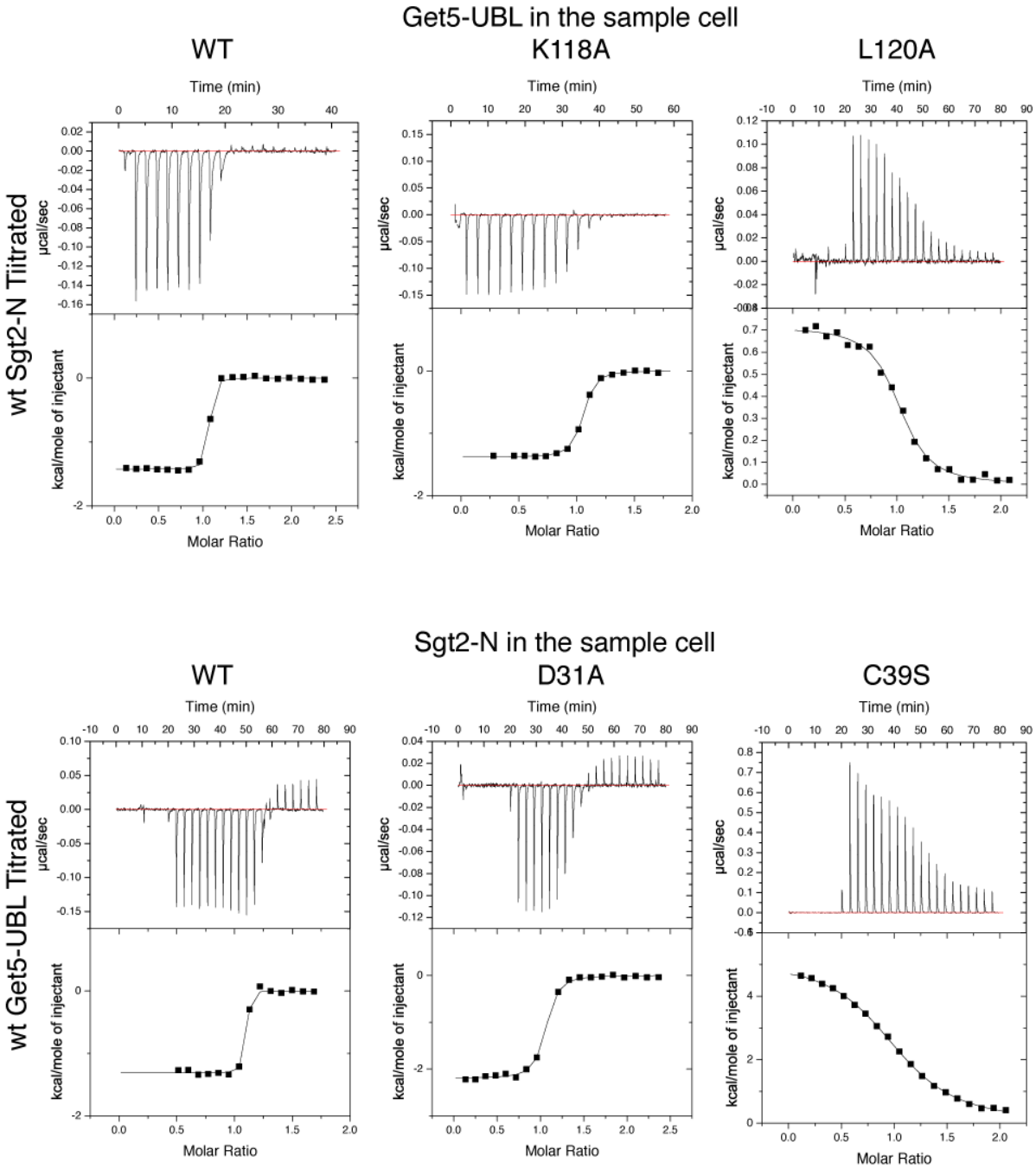


Figure 4.6. Isothermal titration calorimetry. Representative ITC isotherms for the interaction between Sgt2-N and Get5-UBL. Raw data are shown in the top panel of each trial as the power input to the sample cell over time. Integrated data are shown in the bottom panels in terms of total energy required for equilibration as a function of molar ratio. Data using Get5-UBL as a titrant were corrected for heat of dilution of the protein. Note that trials are individually scaled to demonstrate the quality of fits. The top row utilized Get5-UBL variants in the sample cell and wild type Sgt2-N was titrated. The bottom row used Sgt2-N variants in the sample cell and wild type Get5-UBL was titrated.

affinity decreases rapidly as salt concentration increases. Moreover, mutations of any of the lysines on this face of Get5-UBL to alanine (79, 118, 122, and 124) reduce binding affinity 5–10-fold. Interestingly, ubiquitin, which contains most of the residues tested, binds Sgt2-N with negligible affinity (data not shown).

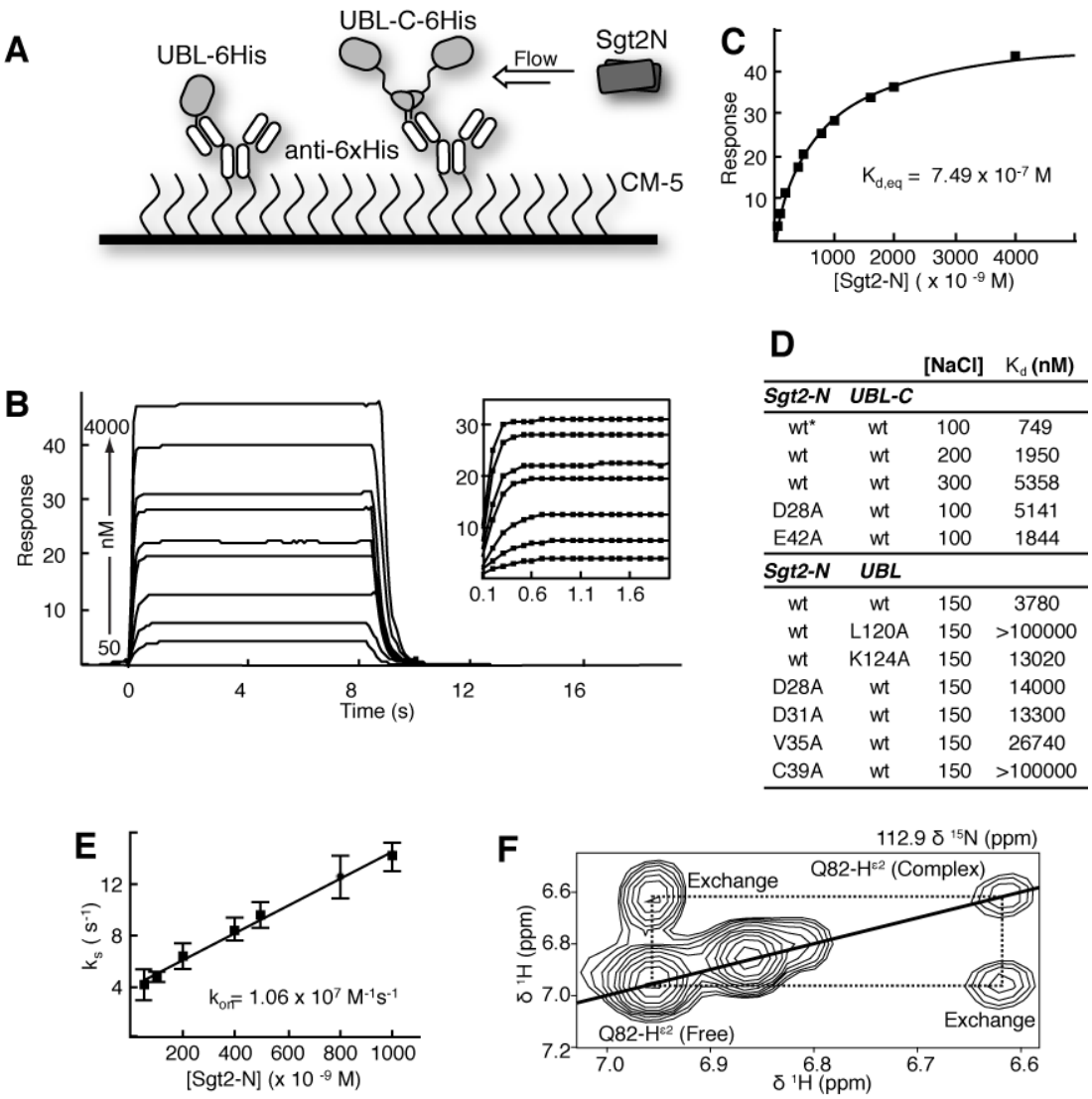
We made reciprocal mutations to the conserved face of Sgt2-N and tested their binding to Get5-UBL (Figure 4.5). Similar to Get5, mutation of the two exposed hydrophobics to alanine had the strongest effect with over 100-fold lower affinity (V35A and C39A). The presence of the completely conserved Cys was curious. We decided to do a series of typically minimal changes at this position to test for effect on complex formation. A slightly bulkier hydrophobic side chain, C39V, resulted in a 100-fold lower affinity while removing the sulfhydryl to a smaller amino acid, C39A, resulted in a nearly 300-fold lower affinity. The strongest effect was conversion of the sulfhydryl to the more polar hydroxyl, C39S, resulting in a ~700-fold lower affinity. Mutations of the acidic residues that comprise the charged face have a similar effect on affinity as the basic residues of Get5-UBL. One exception is that the peripheral Glu47, located at the beginning of α3, had no effect on affinity.

Characterization of the binding kinetics of Sgt2-N and Get5-UBL

The complex between the minimal Get5-UBL and Sgt2-N domains appears homogenous and stable over multiple rounds of size exclusion chromatography. Multiple lines of evidence demonstrate complex formation between Get5 and Sgt2 (Chang et al., 2010; Chartron et al., 2011; Liou et al., 2007; Wang et al., 2010). Based on ITC, the complex has high affinity; however, investigations using co-immunoprecipitation find variable

amounts of Sgt2 or SGTA associated with the Get5/Ubl4A partners. Although there are many possible reasons why the *in vivo* stoichiometry is variable, we hypothesized that fast binding kinetics could explain how co-purification could be dependent on experimental conditions. To measure the association and disassociation rate constants we turned to surface plasmon resonance (SPR). In this experiment, polyhistidine-tagged Get5-UBL or Get5-UBL-C was immobilized for SPR analysis and Sgt2-N was used as the analyte (Figure 4.7A). Get5-UBL-C includes the C-terminal dimerization domain of Get5 that is separated from the UBL domain by a flexible linker (Chartron et al., 2010). We previously demonstrated that the less restrictive Get5-UBL-C dimer can bind two Sgt2-N domains (Chartron et al., 2011); therefore, we expect that each will act independently. The proteins were well behaved on the chip giving stable concentration dependent saturation (Figure 4.7B). The rapid saturation of response units after injection (<1 s) and then rapid reduction after the analyte injection is stopped are indicative of fast on and off rates consistent with our hypothesis.

The equilibrium-binding constant can be calculated by plotting response units as a function of Sgt2-N concentration after response units reach equilibrium (Figure 4.7B–D). For Get5-UBL-C, the plot fit to a K_d of 7.49×10^{-7} M. Compared to ITC, this is nearly two orders of magnitude lower affinity. Although the different techniques are not expected to give identical results due to experimental conditions, the relative results are consistent. For example, mutation of charged residues, Sgt2-N D28A and D42A, resulted in a significant loss in affinity. We suspect that steric constraints based on interactions with the antibody affect the measured rates. This is seen when we used immobilized



		[NaCl]	K_d (nM)
Sgt2-N	***UBL-C***		
wt*	wt	100	749
wt	wt	200	1950
wt	wt	300	5358
D28A	wt	100	5141
E42A	wt	100	1844
Sgt2-N	***UBL***		
wt	wt	150	3780
wt	L120A	150	>100000
wt	K124A	150	13020
D28A	wt	150	14000
D31A	wt	150	13300
V35A	wt	150	26740
C39A	wt	150	>100000

Figure 4.7. Binding is characterized by rapid on and off rates mediated by electrostatics as analyzed by SPR. (A) Scheme used for SPR analysis. Either his-tagged Get5-UBL (UBL-6His) or Get5-UBL-C (UBL-C-6His) were attached to an anti-6xHis antibody sparsely immobilized on a CM-5 chip. Sgt2-N was flowed over the chip at varying concentrations. (B) Representative SPR experiment flowing varying concentrations of Sgt2-N over immobilized Get5-UBL-C. Inset, the first two seconds of the experiment, used to derive k_s for (E). (C) Plot of equilibrium response units versus concentration of Sgt2-N for calculation of equilibrium K_d. (D) Table of the effects of mutants and salt concentration on binding. An asterisk signifies that the experiment was performed in triplicate. (E) Plot of the coefficient that fits k_s versus Sgt2-N concentration. The slope gives the association constant (k_a) while the Y-intercept gives the dissociation rate (k_{off}). (F) Plot of a ^{15}N plane from an EXSY-HSQC spectrum showing a side chain amide proton of Gln82 in free or complex Get5-UBL. The cross peaks indicate exchange between the two states within the timescale of the experiment.

Get5-UBL, which is expected to bring the Sgt2 binding face into close proximity to the immobilizing antibody. This set up had consistently lower binding affinities, for example wild type binding is reduced approximately 5-fold to a K_d of 3.78 x 10^{-6} M. Mutants in Sgt2-N or Get5-UBL show similar changes in binding affinity compared to ITC. Charged mutants Get5-UBL K124A, and Sgt2-N D28A and D31A all had similar affinities on the order of 3–4-fold weaker than the wild type protein. Also, residues in the hydrophobic interface had the strongest effect, reducing the affinity of Get5-UBL L120A or Sgt2-N C39A to below what could be accurately determined in this experimental set up.

The importance of charge complementarity is consistent with two preformed interfaces that are electrostatically steered toward complex formation. These types of protein interactions are known to form quickly and then rapidly dissociate (Sheinerman et al., 2000). Here, the association phase of the SPR data were used to determine a k_{on} of 1.06 x 10^7 M^{-1} s^{-1} and a dissociation rate constant of ~4 s^{-1} (Figure 4.7E). The dissociation rate constant could be independently estimated from the equilibrium K_d and k_{on} values, which gives an approximate k_{off} of 7.9 s^{-1}. Consistent with both the electrostatic mechanism and the ITC data, increasing salt concentration from 100 to 300 mM salt resulted in nearly an order of magnitude change in K_d (Figure 4.7D). All of this data points to a highly specific interface that has rapid dynamics.

For an independent verification of the kinetics, we measured exchange rates by NMR using EXSY spectra (Perrin and Dwyer, 1990). A 2:1 ratio of ^{15}N-labeled Get5-UBL to unlabeled Sgt2-N homodimer was prepared, allowing detection of approximately equimolar free and complex forms of Get5-UBL. The complex dissociates and reforms

within the timescale of the experiment (mixing times of 60 or 120 ms) and exchange cross peaks are observed between the two states of specific Get5-UBL protons (Figure 4.7F). The side chain amide protons of Gln82 and the main chain amide proton of Asn129 gave well-resolved exchange peaks, the magnitudes of which were used to calculate a $k_{off,NMR}$ of 8.3 ± 1.4 s^{-1} and an apparent on rate of 6.5 ± 0.4 s^{-1}. Under these conditions the forward and reverse rates are expected to be equal, with apparent on rate as the product of the rate constant $k_{on,NMR}$ and K_d. Using the ITC-derived K_d of 34 nM, $k_{on,NMR}$ is 1.9 × 10^8 M^{-1} s^{-1}, roughly 20-fold faster than measured by SPR. Because the solution and immobilized Get5 off rates are similar, the lower binding affinity seen by Get5 immobilization in the SPR experiment is from a slower k_{on}. A possible explanation could be that restricting the rotational freedom of Get5 reduces the rate acceleration caused by an electrostatic steering mechanism.

Structure of the Sgt2-N and Get5-UBL complex

Compared to the proteins alone, the NMR spectra of the Sgt2-N/Get5-UBL complex have a dramatic reduction in resolution. This is a result of peak broadening due to both slower tumbling and the fast kinetics of complex formation and dissociation. Therefore, rather than generating a uniformly $^{13}C/^{15}N$ labeled sample of 235 residues, we opted to investigate two asymmetrically labeled complexes to reduce the amount of chemical shift overlap. Chemical shift perturbations (CSP) on the 2D ^{1}H-15N-HSQC spectra were examined for both proteins (Figure 4.8A). On Sgt2, the most drastic CSPs occur at the Get5 binding surface, with very little change occurring in the rest of the protein (Figure 4.8B). Binding of a single Get5-UBL is anticipated to break the symmetry of Sgt2-N,

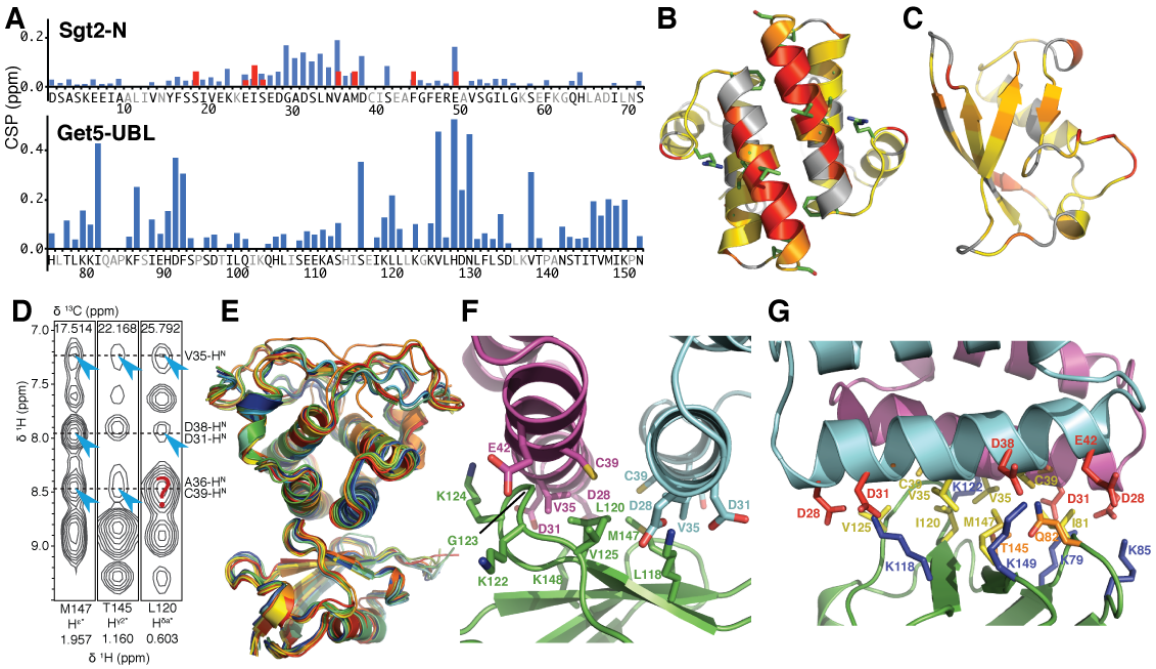


Figure 4.8. Solution structure of the Sgt2-N/Get5-UBL complex. (A) Plots of the change in chemical shifts of the ^{1}H-^{15}N HSQC as a combined chemical shift perturbation (CSP) upon complex formation for each residue at 25°C. Changes for residues in gray could not be accurately determined due to cross peak overlap. For Sgt2-N, several residues were split into two peaks, breaking the symmetry. These are indicated in red. (B) Cartoon representation of Sgt2-N illustrating regions of chemical shift change between the free protein and the complex (cyan). Coloring is based on values from (A). Residues are color ramped from smallest (yellow) to largest (red) CSP. Residues that could not be measured are shown in gray. Residues with split chemical shifts upon complex formation are shown as green sticks. (C) Cartoon representation of Get5-UBL similar to (B). (D) Representative intermolecular NOEs from a ^{13}C-edited NOESY-HSQC spectrum at 37°C constrained to be less than 6Å away during docking. Peaks identified as intermolecular are highlighted by blue arrowhead and labeled. The red "?" represents a peak that is unresolvable from an expected intramolecular NOE. (E) Overlay of the 10 best-scored models obtained after NOE, AIR and RDC driven docking. (F and G) Two views of the binding interface. In (F), residues determined to interact by ITC are highlighted as sticks. In (G), residues at the interface are drawn as sticks with the following color scheme positive (blue), negative (red), polar (orange) and hydrophobic (yellow).

indeed several residues on the Get5 binding surface face split into two cross peaks with reduced peak height (Figures 4.8A and 4.9A). Residues without perturbation maintain a single cross peak, indicating that symmetry remains away from the binding site. The Get5-UBL domain is more broadly affected by binding to Sgt2-N, but the most intense CSPs occur at Ile81 and the loop consisting of residues 123-129, including Gly123 and Val125.

Inspection of NOESY spectra failed to conclusively identify enough new cross peaks resulting from intermolecular contacts to directly determine the structure of the complex. This is not surprising, based on the presumed interface. Electrostatic interactions between a glutamate or aspartate and a lysine yield weak proton NOE cross peaks. Additionally, the few expected hydrophobic interactions occur in crowded regions of the spectra. We proceeded to determine the individual solution structures of Sgt2-N and Get5-UBL as they are in complex. Data collection for the full assignments of the proteins in complex was performed at 37°C, which reduced the severity of line broadening effects. This had the additional effect of averaging the two states of Sgt2-N. We therefore treated Sgt2-N as symmetric in these calculations.

Overall, the structures of Sgt2-N and Get5-UBL while in complex are not significantly different from the free solution structures (Figure 4.9B,C). A few intermolecular NOEs could be identified and were used in the calculation of the complex structure (Figure 4.8D). In the absence of substantial numbers of NOE derived distance restraints, the structures of complexes can be determined by molecular docking driven by other

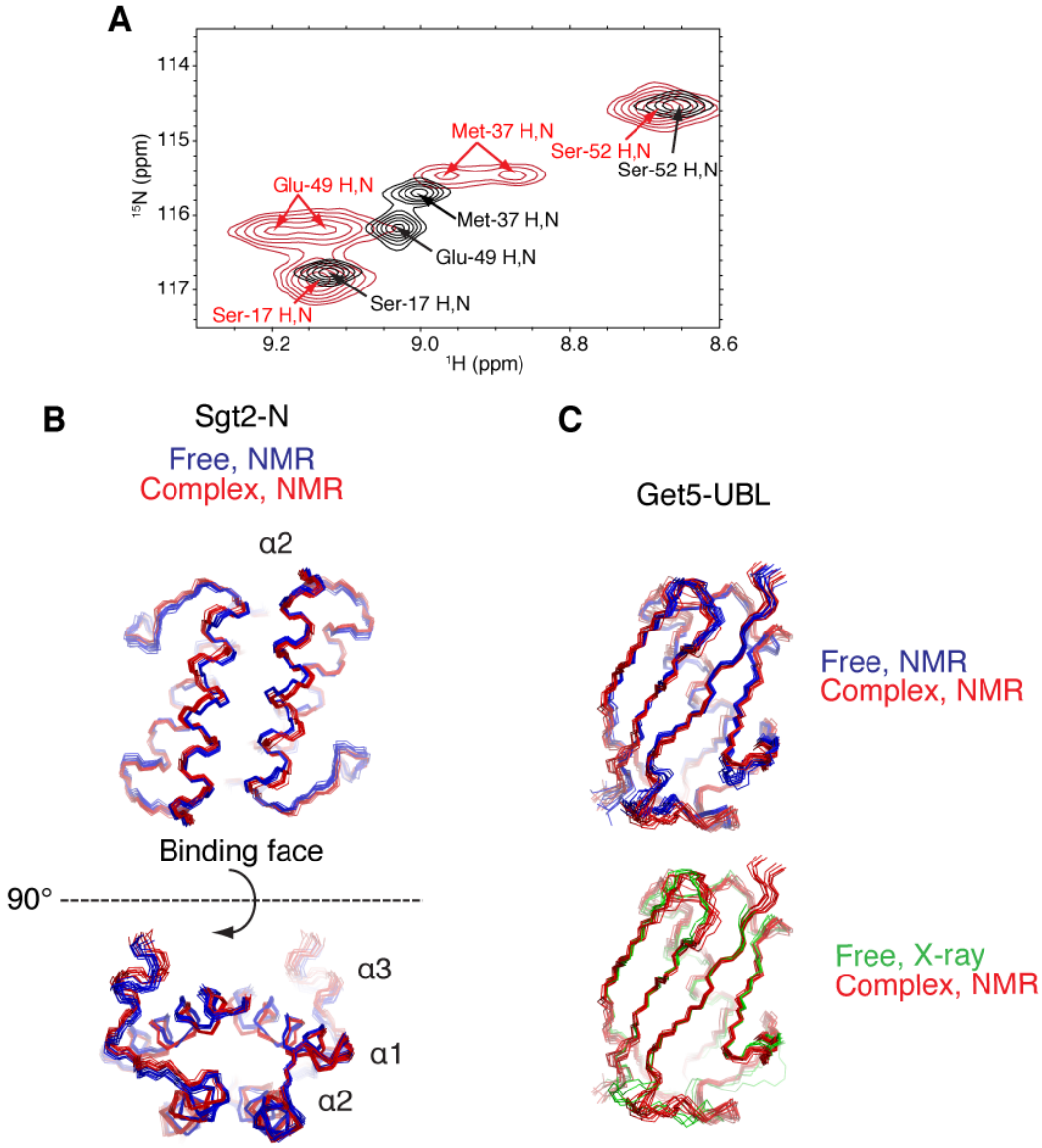


Figure 4.9. Comparison of the structures of free and bound proteins. (A) A section of overlaid ^{1}H-^{15}N HSQC spectra of uniformly $^{13}C/^{15}N$ labeled Sgt2-N (black) or labeled Sgt2-N in a saturated complex with natural abundance Get5-UBL (red). Some cross peaks split into two indicating asymmetric binding. (B) Overlay views of the backbone nitrogen and carbon atoms of the ensembles of Sgt2-N (blue) and Sgt2-N while in complex with Get5-UBL (red). For clarity only residues 5-56 are shown. (C). Overlay views of ensembles of Get5-UBL structures. In the top panel, the NMR structures of free (blue) Get5-UBL and Get5-UBL in complex with Sgt2-N (red). The bottom panel is an overlay of the complex NMR structure (red) and the crystal structure (green).

experimental data introduced as ambiguous interaction restraints (AIRs) (Dominguez et al., 2003). We defined 14 AIRs using the CSP and mutagenesis data, the details of which are provided in the methods section of this chapter. Moreover, we collected residual dipolar couplings that restrict rotational freedom between the models during docking. We performed docking between the two separate complex structures to generate a full model of Sgt2-N and Get5-UBL (Figure 4.8E). Statistics for the structure calculation of each component as well as the complex are provided in Tables 4.3 and 4.4.

The experimentally restrained docking returns a well-converged structure where the predicted binding faces interact (Figure 4.8E). All of the residues that were identified to be involved experimentally are found at the interface (Figure 4.8F). The interface is comprised of the hydrophobic patch on Get5-UBL (Ile81, Leu120, Val125, Gly143 and Met147) that docks against the reciprocal patch that contains two each of Cys39 and Val35 from the Sgt2-N dimer (Figure 4.8G). Additionally, Thr145 packs against one copy of Val35. The conserved lysines 79, 85, 118, 122 and 149 all make electrostatic contacts to the charged face of Sgt2-N (two each of Asp 28, 31, 38 and Glu42). The interface has a surface area of ~700 $Å^2$.

The complex is a unique UBD/UBL interaction

Ubiquitin and UBL domains are abundant in cells and all share common features (Winget and Mayor, 2010). Perhaps more abundant are the different UBD motifs found in a wide variety of frameworks (Hicke et al., 2005; Husnjak and Dikic, 2012). The combinatorial use of UBLs and UBDs results in a wide diversity of interactions that are utilized in many different contexts. The complex between Sgt2 and Get5 introduces a novel interaction.

Table 4.3. NMR structural constraints and structure statistics for proteins in complex

$^{15}N/^{13}C$ labeled protein	Sgt2-N	Get5-UBL
No. of restraints[a,b]		
NOE-based distance restraints		
Intra-residue ($\|i - j\| = 0$)	770	597
Sequential ($\|i - j\| = 1$)	336	378
Medium range ($2 \leq \|i - j\| < 5$)	276	198
Long range ($\|i - j\| \geq 5$)	208	258
Inter-molecular	70	-
Ambiguous	668	419
Total	2328	1850
$\phi + \psi$ dihedral angle restraints	204	126
Hydrogen bond restraints	100	52
Residual dipolar coupling restraints	52	45
Magnitude (D_a)	6.4	
Rhombicity (R)	0.61	
Restraints statistics		
r.m.s.d. from experimental distance restraints	0.010 ± 0.003 Å	0.014 ± 0.002 Å
r.m.s.d. from experimental dihedral restraints	2.2 ± 0.2°	0.37 ± 0.08°
Cross-validated RDC Q-factors[c], (No. used for validation)	0.29 ± 0.10 (12)	0.35 ± 0.07 (10)
Model Statistics		
Residue Range[d]	5-56 (×2)	73-149
Coordinate precision r.m.s.d.		
Backbone	0.36 ± 0.04 Å	0.49 ± 0.09 Å
Heavy atom	0.76 ± 0.05 Å	0.88 ± 0.07 Å
Structural quality		
Ramachandran statistics[e]		
Most favored regions	98.10%	86.10%
Allowed regions	1.90%	13.20%
Generously allowed regions	0.00%	0.30%
Disallowed regions	0.00%	0.40%
WHAT-IF Z-score[f]		
Backbone conformation	0.939 ± 0.578	–0.546 ± 0.500
2nd generation packing quality	4.137 ± 1.486	4.471 ± 1.997
Ramachandran plot appearance	–1.227 ± 0.499	–2.464 ± 0.505
$\chi 1/\chi 2$ rotamer normality	–2.358 ± 0.811	–3.074 ± 0.386

[a]For Sgt2-N, a single NOE generates two restraints for atom pairs from each monomer.

[b]Structures were calculated using the full sequence of the protein.

[c]Q-factors were calculated with PALES using RDCs omitted from structure calculations (Zweckstetter and Bax, 2000).

[d]Validation was performed over the ordered sequence of the protein.

[e]Ramachandran statistics calculated by PROCHECK (Laskowski et al., 1993).

[f]Z-scores were calculated by WHAT-IF (Vriend and Sander, 1993).

Table 4.4. Calculation of the Get5-UBL and Sgt2-N complex

No. of restraints					
Intermolecular NOE-based distance restraints			14		
Ambiguous Interaction Restraints			10		
Residual dipolar coupling restraints			97		
$\phi + \psi$ dihedral angle restraints			326		
Hydrogen bond restraints			152		
Calculation statistics					
Statistics for top 5 clusters[a]	1	2	3	4	5
No. of structures	48	20	16	6	8
HADDOCK score of 10 best structures[b]	-63 ± 7	-43 ± 7	-43 ± 15	-40 ± 15	-39 ± 15
Statistics for 10 best structures of best cluster					
Buried surface area (Å^2)			1125 ± 42		
Backbone r.m.s.d.			0.58 ± 0.22 Å		
Heavy atom r.m.s.d.			0.92 ± 0.22 Å		
RDC Q-factors[c]					
Working + cross-validated (No.)			0.176 ± 0.014 (120)		
Cross-validated (No.)			0.311 ± 0.036 (23)		
Structural quality[d]					
Ramachandran statistics[e]					
Most favored regions			92.80%		
Allowed regions			6.80%		
Generously allowed regions			0.40%		
Disallowed regions			0.00%		
WHAT-IF Z-score[f]					
Backbone conformation			0.036 ± 0.242		
2nd generation packing quality			4.258 ± 1.911		
Ramachandran plot appearance			–1.758 ± 0.275		
$\chi 1/\chi 2$ rotamer normality			–5.498 ± 0.412		

[a]Clusters generated using a 2 Å r.m.s.d. interface similarity cutoff (Daura et al., 1999).

[b]HADDOCK score is a sum of electrostatic, van der Waals, desolvation and AIR energy terms.

[c]Q-factors were calculated with PALES (Zweckstetter and Bax, 2000). Cross-validated RDCs were omitted from docking, while working RDCs were included used in calculations.

[d]Validation was performed over Get5-UBL residues 73-149 and Sgt2-N residues 5-56.

[e]Ramachandran statistics calculated by PROCHECK (Laskowski et al., 1993).

[f]Z-scores were calculated by WHAT-IF (Vriend and Sander, 1993).

As is typical for the most commonly observed UBD interactions, Sgt2-N binds at the face that contains the I44 patch (Figure 4.10A,B). Different from other UBDs, Sgt2 uses a symmetrical dimer interface to interact with its UBL. The only other example of a UBD dimer is the swapped CUE motif found in Vps9 that binds similar to other CUE domains (Figure 4.10B, PDBID:2P3Q) (Prag et al., 2003). Most other of the characterized UBDs bind primarily with a single polypeptide. Comparing UBD binding interfaces, Sgt2 buries an atypically large surface area (682.8 $Å^2$) (Figure 4.10A,B). The next closest interface for a yeast UBD is that of Ufd2 bound to the UBL of Rad23, which has an interface of 614.3$Å^2$ (Hanzelmann et al., 2010).

Most of the I44 patch-binding UBDs use α-helical motifs, as does Sgt2. These proteins interact with similar groups of residues on their respective UBLs (Figure 4.10C,D). UBLs have a number of conserved residues around the I44 patch. For ubiquitin, these include Leu8, Arg42, Ile44, His68 and Val70. These interactions are conserved on Get5 in the interface with Sgt2 (Ile81, K118, Leu120, Gly123, Thr145 and Met147). For these residues, most are similar in nature to the canonical residues except for Thr145 that is a histidine in most UBLs.

Of the remaining Sgt2/Get5 interactions, the most interesting are the residues that are unique in the interface relative to all other UBLs (Figure 4.10C,D). Four residues fit this description. The most provocative, the highly conserved Ile44 of ubiquitin is a leucine in all Get5 homologs. Surprisingly, mutating this to isoleucine had no affect on binding affinity (Figure 4.5). The second is Gln82 in Get5 that forms a conserved H-bond network with negative charges on Sgt2. In ubiquitin, this residue is a threonine that likely

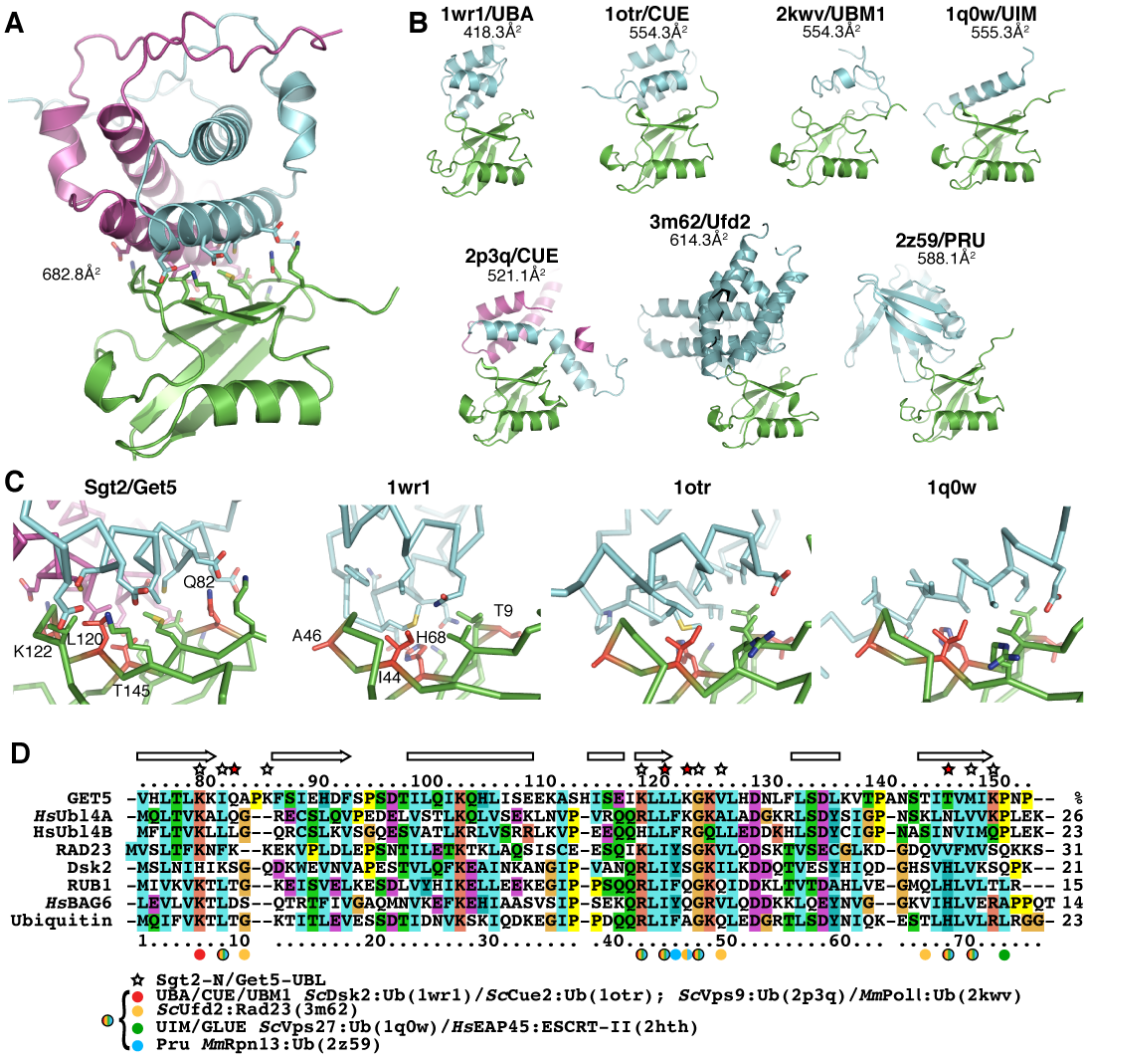


Figure 4.10. The Sgt2-N/Get5-UBL complex is a unique UBD/UBL interface. (A) A view of the complex shown as a cartoon diagram. Sgt2-N in cyan and purple, Get5-UBL in green. Residues involved in the interface are drawn as sticks. The interface surface area, determined by PISA, is indicated (Krissinel and Henrick, 2007). (B) Representative structures of various UBD (cyan)/UBL (green) complexes that interact with the "I44" face shown as cartoon aligned to the Get5-UBL in (A). Each structure is labeled with its PDBID and the UBD group that is represented (Reviewed in Hicke et al., 2005; Husnjak and Dikic, 2012). Proteins are named in (D). Interface surface areas are indicated. (C) View of the binding interface of representative helical UBDs. Residues at the interface are shown as sticks (D). Get5-UBL residues Q83, L120, K122 and T145 are colored red. Equivalent residues in other structures are also highlighted. (D) Sequence alignment of *S. cerevisiae* UBL domains with human Ubl4A, Ubl4B and BAG6 UBLs included. Alignment is similar to Figure 4.1G. Numbers at the end of the sequence are percent identity to Get5. Residues that interact with various UBDs are highlighted based on the legend. Stars filled with red are highlighted in (C).

cannot contribute to a similar network and, in fact, is pointed away from the I44 patch. Next is Lys122, a positive charge that is conserved in Get5 homologs adjacent to the I44 patch yet is missing in other UBLs. This lysine forms salt bridges with the conserved Asp28 and Asp31 on Sgt2. In ubiquitin, this position is an alanine that points its side chain away from the interface. The final residue is Thr145 whose equivalent in ubiquitin is a histidine (His68). In Ubl4A, the position is an asparagine, a more polar residue. For ubiquitin, His68 is typically described as a component of the hydrophobic pocket lining the edge of the I44 patch. For the Get5/Sgt2 complex, the smaller threonine is likely required to accommodate the tight interface.

Discussion

The biogenesis of TA membrane proteins requires sorting in the cytoplasm, targeting to the membrane followed by insertion into the bilayer. For TAs destined for the ER, the conserved GET pathway governs this process. Sgt2 and Get5 are members of the so-called yeast TRC that includes Get4 and HSPs. This complex is responsible for binding and then sorting of ER destined TAs to the targeting factor Get3 (Wang et al., 2010). Sgt2 contains three domains (Figure 4.11A). The C-terminal domain binds hydrophobic peptides with varying affinities. Mitochondrial TA proteins are typically less polar compared to ER destined substrates and it is thought that this feature allows Sgt2 to selectively bind the latter with higher affinity (Wang et al., 2010). The central domain contains three TPR repeats that bind multiple classes of HSP proteins whose structure was recently solved (Chartron et al., 2011). At the N-terminus is a homodimerization domain. We report the first structure of this domain here. In solution, Sgt2 forms an

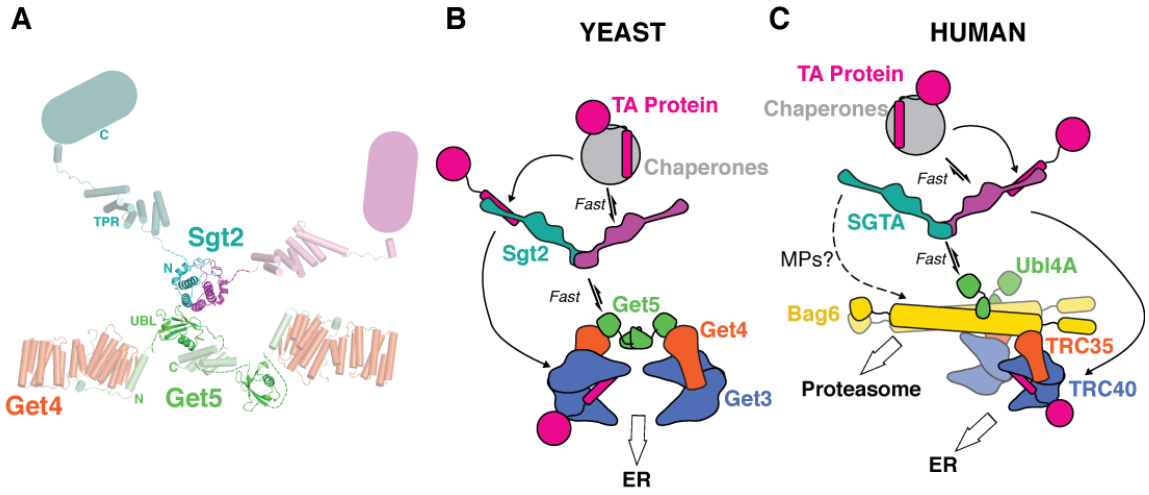


Figure 4.11. A model for the role of the Sgt2/Get5 complex. (A) Composite model of the Sgt2/Get4/Get5 complex based on all available structural data. Central is the complex reported here. Sgt2 is an extended dimer (cyan and magenta). The TPR domains extend away from each other (PDBID:3SZ7). The C-terminal domains, represented as rounded rectangles, are flexible. The Get4/Get5 complex is an extended homodimer mediated by the Get5-C domain (PDBID:3VEJ). Get4 forms a complex with the Get5-N domain (PDBID:3LKU). (B) Model of the TRC complex in the GET pathway. Chaperones binding hydrophobic proteins rapidly bind and dissociate from the TPR domain of Sgt2. The C-terminal domain of Sgt2 binds to ER destined TA proteins. Sgt2 is rapidly binding and dissociating from Get4/Get5, which binds to Get3. The TA protein is then transferred from Sgt2 to Get3. (C) Model of the mammalian TRC pathway. Chaperones and SGTA act analogously to the yeast system in (B). SGTA rapidly binds to and dissociates from Ubl4A. The Bag6 complex then sorts substrates between the TA targeting pathway, mediated by TRC35 and TRC40 and the proteasome.

extended dimeric complex with the C-terminal domain moving freely at the end. Get5 forms an obligate heterotetramer with Get4 (Figure 4.11A). Get5 contains three domains. An N-terminal domain that wraps around Get4 forms the heterodimer interface. A C-terminal domain forming a small, stable dimerization motif whose structure was also recently solved (Chartron et al., 2012b). The central domain is a novel UBL domain whose structure we report for the first time here. The Get4/Get5 complex is also extended in solution (Chartron et al., 2011; Chartron et al., 2010).

The initial identification of Get4 and Get5 as bona fide members of the GET pathway did not reveal the connection to Sgt2 (Jonikas et al., 2009). In hindsight, this is surprising as previous biochemical and genetic links had been reported (Liou et al., 2007). Subsequent studies clearly linked the N-terminal dimerization domain of Sgt2 to the UBL domain of Get5 solidifying the role of Sgt2 in TA targeting (Battle et al., 2010; Chang et al., 2010; Chartron et al., 2011; Wang et al., 2010). We previously demonstrated that the interface formed was stable to purification, yet sensitive to mutation (Chartron et al., 2011; Chartron et al., 2012b). The structures reveal a strong electrostatic component to the interface between Get5 and Sgt2. This results in a complex with fast on and off rates.

The affinity of Sgt2 to Get5 is remarkably high (Figure 4.5). In the cell, every Get4/Get5 heterotetramer will, on average, be bound by an Sgt2, as one falls off another quickly replaces it. Our previous work demonstrated that only a single Sgt2 dimer could bind to the Get4/Get5 heterotetramer at one time (Chartron et al., 2011); therefore, Get5 has two potential binding sites for Sgt2 that are rapidly sampling. This makes sense in a model for TA targeting (Figure 4.11B), as a Get3/Get4/Get5 complex would be stable in the

cytoplasm. This complex would screen, over multiple rounds, Sgt2 proteins to find one that stably bound an ER destined TA protein. This interaction would lead to a handoff of the TA to Get3 that would be released from Get4 to find its ER receptors. How Sgt2 finds TA proteins is a matter of conjecture. The simplest model is that it captures free TAs in the cytoplasm; however, it remains seductive to imagine that the highly abundant HSPs provide at least one route into the pathway. In that context, Sgt2 would bind HSPs transiently allowing for multiple rounds of binding to find appropriate substrates. Once a TA protein was bound, it would be a stable complex that could be found by Get4/Get5.

In metazoans the picture becomes more complicated (Figure 4.11C). In addition to homologs for all the yeast GET proteins, TA targeting includes a large multidomain protein called Bag6. Bag6 contains an N-terminal UBL that has the typical features to ubiquitin (Figure 4.10D) and a C-terminal BAG domain. It is linked to TA targeting by forming a stable complex with TRC35 and Ubl4A (Mariappan et al., 2010). TRC35 and Ubl4A lack the features necessary for direct complex formation in yeast (Chartron et al., 2010; Chartron et al., 2012b) and it is likely that they both bind Bag6 directly. The first structure solved from this complex is the UBL domain from Ubl4A, which has not been described in the literature (PDBID:2DZI). Although it has yet to be experimentally demonstrated, it seems very likely that SGTA performs a similar role in TA targeting to its yeast counterpart (Figure 4.11C). The structure of the dimerization motif of SGTA supports this, as it is a highly conserved domain, like Ubl4a-UBL, with all of the features that are important for dimer formation. Bag6 is demonstrated to be a dimer (Yihong Ye, personal communication) and forms a complex with SGTA via the UBL domain of

Ubl4a. This then would mirror all of the components of the yeast system including two each of SGTA, Ubl4A and TRC35 (Figure 4.11C).

In mammalian cells, the Bag6 complex is linked to the degradation of mislocalized membrane proteins and dislocated ER products (Hessa et al., 2011; Wang et al., 2011). TA proteins can be redirected to the degradation pathway suggesting that the two pathways intersect (Hessa et al., 2011). SGTA has now been shown to be an important component for targeted degradation of ERAD substrates transiently binding hydrophobic membrane proteins prior to handoff to Bag6 (Figure 4.11C) (Xu et al., 2012). This additional role would also benefit from rapid sampling of SGTA to the Bag6 complex. This raises the possibility that, in yeast, Sgt2 could have multiple roles as well.

Where the proteins are linked to degradation pathways, it seems plausible that the additional role for UBLs may have co-evolved. This is clear for the Bag6-UBL, which has all of the features of UBLs involved in degradation and is critical for targeting to the proteasome (Hessa et al., 2011; Wang et al., 2011; Xu et al., 2012). The Get5/Ubl4A-UBL represents a unique class with features that clearly distinguish them from other UBLs (Figure 4.10). In fact, when the conserved histidine of the Bag6-UBL is converted to an asparagine, it completely loses its ability to bind to standard UBDs (Xu et al., 2012). A similar effect is not seen when the conserved Leu120 is replaced by an isoleucine, although this could be a change to prevent other UBDs from binding Get5.

Sgt2 and SGTA are novel UBDs that have very specific binding partners. An interface dominated by electrostatics is unique amongst UBL/UBD complexes. This presumably allows high-affinity, rapid binding while strongly rejecting unfavorable interactions with

other UBLs and UBDs. One interesting side note is that in mammals, both Ubl4A and SGTA have tissue specific isoforms, Ubl4B and SGTB (Tobaben et al., 2003; Yang et al., 2007). While SGTA and SGTB have similar conserved sequence elements, Ubl4B is missing a number of the residues that likely form the SGTA/Ubl4A interface (Figure 4.10D). This suggests it may have lower affinity or perhaps an unknown UBD.

In this chapter we demonstrate a novel, conserved UBD/UBL interaction critical for TA targeting. The complex is in a dynamic equilibrium that allows for rapid sampling of the various components. This attribute is likely essential for the various roles that Sgt2 homologs must play. This work opens the door to understanding the steps of target selection and discrimination that are required for a regulated process. The finer details of the process of TA targeting continue to be resolved at a rapid pace; however, each new insight leads to unexpected elaborations. With the recent link to regulated proteolysis, TA targeting is becoming part of the greater picture of homeostasis in the cell.

Methods

Expression and purification of samples for NMR

We previously described the cloning of *S. cerevisiae* Sgt2 residues 1-72 and Get5 residues 74-151 (Chartron et al., 2011; Chartron et al., 2010). A hexahistidine tag and TEV protease cut site were fused to the N-terminus of Sgt2-N, and a hexahistidine tag was fused to the C-terminus of Get5-UBL. Uniformly ^{15}N or $^{13}C/^{15}N$ labeled proteins were produced using the media and protocol described by Marley et al. (2001). Proteins were expressed in NiCo21(DE3) cells (New England Biolabs) at 37°C for 5 hours after induction with 0.3 mM IPTG. Cell pellets were resuspended in 20 mM Tris, 300 mM NaCl and 20 mM imidazole pH 7.5 and lysed by sonication. Clarified lysates were passed over Ni-NTA agarose beads (Qiagen) and washed with 20 column volumes of lysis buffer. Proteins were eluted using 20 mM Tris, 100 mM NaCl and 300 mM imidazole, pH 7.5. For Sgt2-N, the affinity tag was removed by digestion with hexahistidine tagged TEV protease while dialyzing against buffer containing 20 mM imidazole. A second Ni-NTA agarose column removed any remaining tagged Sgt2-N and protease. Samples were further purified by size exclusion chromatography (SEC) using a Superdex 75 16/60 column (GE healthcare) equilibrated with 20 mM sodium phosphate pH 6.1 for Sgt2-N, or 20 mM sodium phosphate, 100 mM NaCl pH 6.1 for Get5-UBL. Proteins were then concentrated to 1 mM (2 mM of Sgt2-N monomers) and D_2O was added to 10% for data collection.

Preparation of labeled complexes between Sgt2-N and Get5-UBL

Two asymmetrically labeled complex samples were prepared for NMR investigation. Uniformly $^{13}C/^{15}N$ labeled Sgt2-N was incubated with 2 equivalents of natural abundance Get5-UBL. Excess Get5-UBL was then removed by using a Superdex 75 16/60 column equilibrated with low-conductivity buffer RE (10 mM Bis-Tris, 50 mM L-arginine and 50 mM L-glutamate titrated to pH 6.1 with MES) (Hautbergue and Golovanov, 2008; Kelly et al., 2002). The saturated complex between Sgt2-N and Get5-UBL is stable over multiple rounds of SEC (Chartron et al., 2011). For the other complex sample, uniformly $^{13}C/^{15}N$ labeled Get5-UBL was incubated with a slight excess of natural abundance Sgt2-N and separated over a Superdex 75 16/60 column equilibrated with buffer RE. Free Sgt2-N cannot be completely resolved from the complex by SEC, and so only enough Sgt2-N was added for a peak of free protein to be observed within the shoulder of the complex peak. Complex samples were concentrated to 1 mM and D_2O was added to 10% for data collection.

Expression and purification of the complex between SGTA-N and Ubl4A-UBL

Human SGTA residues 1–54 and Ubl4A residues 1–71 were amplified from cDNA clones (ATCC, Mammalian Gene Collection numbers MGC:4672 and MGC:49894) and inserted into a pET33b-derived plasmid to yield a polycistronic mRNA upon expression. A hexahistidine tag and a TEV protease cleavage site were fused to the N-terminus of Ubl4A. The proteins were co-expressed using NiCo21(DE3) cells in LB media at 37°C for 3 hours after induction with 0.3 mM IPTG. Cells were lysed by sonication in 20 mM Tris, 100 mM NaCl, 20 mM imidazole and 5 mM β-mercaptoethanol, pH 7.5. A Ni-NTA agarose column was used to capture all Ubl4A from lysate and was washed with 20

column volumes lysis buffer. Proteins were eluted in 20 mM Tris, 100 mM NaCl, 300 mM imidazole and 5 mM β-mercaptoethanol, pH 7.5. The affinity tag on Ubl4A was removed by digestion with TEV protease during dialysis against a buffer containing 20 mM imidazole. A second Ni-NTA agarose column removed protease and any remaining tagged protein. The complex between Ubl4A-UBL and co-purified SGTA was separated from free Ubl4A-UBL by a Superdex 75 16/60 column equilibrated with 20 mM Tris, 100 mM NaCl and 5 mM β-mercaptoethanol, pH 7.5 (Figure 4.2C). The complex was concentrated to approximately 5 mg/ml for use in crystallization trials.

Expression and purification of Get5-UBL for crystallization

Get5 residues 74–148 were amplified from the expression vector used for NMR studies and inserted into a pET33b-derived plasmid, fusing a hexahistidine tag and a TEV protease cleavage site to the N-terminus. The protein was expressed using NiCo(DE3) cells grown in LB media at 30°C for 5 hours after induction with 0.3 mM IPTG. The majority of the protein localized in inclusion bodies. Get5-UBL was purified from the soluble fraction using the protocol described for the NMR samples. The hexahistidine tag was removed by digestion with TEV protease, and the final Superdex 75 16/60 column run was equilibrated with 20 mM Tris, 100 mM NaCl, pH 7.5. The protein was concentrated to approximately 10 mg/ml for crystallization trials.

Expression and purification of proteins for ITC and SPR

Mutants were introduced into the expression vectors used for NMR study by the Quikchange method (Stratagene). We previously described the expression plasmid for Get5-UBL-C residues 74–212 (Chartron et al., 2010). Proteins were expressed using

NiCo(DE3) cells grown in LB media at 37°C for 3 hours after induction with 0.3 mM IPTG and purified using the protocol described for the NMR samples. All proteins were eluted from a Superdex 75 16/60 column using the same batch of 20 mM Tris and 100 mM NaCl pH 7.5. Proteins used in the sample cell of the ITC experiments were not concentrated following SEC and had concentrations of 30–100 μM. Proteins used in the injection syringe of the ITC experiments were concentrated to 800–1000 μM with conical centrifugal filter units (Millipore) that were extensively washed with the SEC buffer prior to use. Concentrations were determined using either absorbance at 280 nm with extinction coefficients calculated from sequence and the Bradford assay using hen egg white lysozyme and bovine ubiquitin as standards. Denaturation of proteins in 8 M urea prior to the UV assay did not affect absorbance for any of the proteins tested. Proteins used as the analyte in the SPR experiments were concentrated to 1 mM and diluted into the sample buffers, which consisted of 20 mM Tris pH 7.5, 0.05% Tween-20, 3 mM EDTA and the NaCl concentrations indicated in Figure 4.7. The following protein concentrations were used in trials: 4000, 2000, 1600, 1000, 800, 500, 400, 200, 100, and 50 nM.

Isothermal titration calorimetry

Data were collected using a MicroCal iTC-200 calorimeter. Proteins in the sample cell were at 30–100 μM and proteins in the injection syringe were concentrated to 800–1000 μM. For each trial, an initial injection of 0.4 μl was followed by 19 injections of 2 μl each. The cell was allowed to equilibrate for 120 seconds in between each titration. Data were processed using Origin v7.0 (OriginLab) software using a single binding site model.

Surface plasmon resonance

Data were collected using a Biacore T-100 system upgraded to T-200 sensitivity (GE healthcare). Mouse anti-pentahistidine antibody (Qiagen) was covalently linked to a CM5 dextran chip using standard amide coupling chemistry. Hexahistidine tagged Get5-UBL (wild type or mutants) or Get5-UBL-C were then immobilized. Equilibrium binding analysis was performed using BIAevaluation software (GE healthcare). Kinetic analysis between wild type Sgt2-N and Get5-UBL was performed using Kaleidagraph (Synergy Software). The slope values k_s at each concentration were determined by linear regression fitting of the association phase to the integrated 1st-order rate equation. Values of k_s were plotted against concentration and the association rate k_{on} was determined from the linear fit to equation (1).

$$k_s = k_{on} \times [\text{Sgt2-N}] + k_{off} \qquad (1)$$

This fit results in a k_{off} of 3.9 s^{-1}, but since this may not accurately be determined by this method (Karlsson et al., 1991), we additionally estimate k_{off} using the equilibrium dissociation constant as the ratio of k_{off} to k_{on}.

NMR data collection

All data for Sgt2-N, Get5-UBL and two complex samples were collected using a Varian INOVA 600 MHz spectrometer with a triple resonance probe with the following exceptions: HNCACB spectra of Sgt2-N and the labeled Get5-UBL complex, ^{15}N-edited NOESY-HSQC spectra of Sgt2-N and the labeled Get5-UBL complex, and a ^{13}C-edited NOESY-HSQC spectrum of the labeled Get5-UBL complex were collected using a Bruker AVANCE 800 MHz spectrometer with a TCI cryoprobe. Data for the free

proteins were collected at 25°C, and data for the complexes were collected at either 25°C or 37°C. Chemical shift assignments were made using standard triple resonance experiments (Sattler et al., 1999), including ^{1}H-^{15}N HSQC, HNCACB, HNCO, (H)C(CO)NH-TOCSY, CBCA(CO)NH, (H)CCH-TOCSY, H(C)CH-COSY, H(C)CH-TOCSY, ^{15}N-edited TOCSY-HSQC, and aliphatic or aromatic ^{1}H-^{13}C HSQC. Distance constraints were derived from ^{15}N- or ^{13}C-edited NOESY-HSQC spectra. Data were processed using NMRPipe (Delaglio et al., 1995) and analyzed using CCPNMR (Vranken et al., 2005). Residual dipolar couplings (RDCs) $^{1}D_{HN-N}$ were measured by comparing ^{15}N-edited IPAP-HSQC spectra of unaligned samples to samples partially aligned in either 7% (Sgt2-N alone) or 3.5% (complexes) strained polyacrylamide gel (Chou et al., 2001). To make the alignment within the two complex samples as similar as possible, the same stock of acrylamide solution was used to cast both gels. The gels were compressed using the same NMR tube, and data collection began within an hour of either sample preparation. Main chain dihedral angle restraints were predicted from chemical shifts by the program TALOS+ (Shen et al., 2009). The proline conformations were analyzed from chemical shifts by the program PROMEGA (Shen and Bax, 2010) Pro84 is the only proline with chemical shifts indicating a *cis* conformation.

Exchange rates between free Get5-UBL and protein in complex with Sgt2-N were determined using three-dimensional EXSY spectra, collected using the standard ^{15}N-edited NOESY-HSQC pulse sequence. The sample consisted of 400 μM ^{15}N-labeled Get5-UBL and 200 μM unlabeled Sgt2-N homodimer in 10 mM phosphate, 100 mM NaCl pH 6.1. Mixing times were either 60 or 120 ms for the exchange experiments and 6 ms mixing time for the reference experiment. Rates were then determined using

EXSYCALC (Mestrelab Research) using a side chain amide proton of Gln82 and the main chain amide proton of Asn129. The numbers provided in the main text are the average of four values obtained from either proton at both 60 ms and 120 ms mixing times. The binding reaction and the measured rates are described by equation 2.

$$\text{Sgt2-N} + \text{Get5-UBL} \underset{k_{off,NMR}}{\overset{k_{on,apparent}}{\rightleftharpoons}} \text{Sgt2-N/Get5-UBL} \quad (2)$$

The measured association rate determined by EXSYCALC, $k_{on,apparent}$, is related to the rate constant $k_{on,NMR}$ according to equation 3. The dissociation constant K_d is defined by equation 4.

$$k_{on,apparent} = k_{on,NMR} \times [\text{Sgt2-N}] \quad (3)$$

$$K_d = \frac{[\text{Sgt2-N}] \times [\text{Get5-UBL}]}{[\text{Sgt2-N/Get5-Ubl}]} = \frac{k_{off,NMR}}{k_{on,NMR}} \quad (4)$$

Under the conditions used in the EXSY experiment, [Get5-UBL] is approximately equal to [Sgt2-N/Get5-UBL] and can be removed from equation 3. Thus $k_{on,apparent}$ becomes the product of K_d and $k_{on,NMR}$, and is expected to be equal to $k_{off,NMR}$ (equation 5).

$$k_{on,apparent} = k_{on,NMR} \times K_d = k_{off,NMR} \quad (5)$$

Structure calculation of Sgt2-N and Get5-UBL

The assignments of the NOESY spectra were performed automatically during structure calculation using the ambiguous restraints with iterative assignment method in ARIA2.3 (Rieping et al., 2007). The initial rounds of structure calculation for Sgt2-N or Get5-UBL used only NOE-derived distance restraints and main-chain dihedral angle restraints. For

Sgt2-N, every NOE cross peak was initially treated ambiguously as inter- or intramolecular between the two symmetric subunits. Secondary structure was predicted from backbone dihedral angles, and constraints on possible NOE assignment imposed by helices were used to generate an initial set of unambiguous intermolecular contacts (Bardiaux et al., 2009). A ^{15}N-filtered, ^{13}C-edit NOESY-HSQC spectra of Sgt2-N in the presence of 2 equivalents of natural abundance protein was collected to observe unambiguous intermolecular contacts. While these validated the structures, they were not required for calculation convergence. The initial structures were used to assign main-chain hydrogen bond restraints between amide protons and carbonyl oxygens that were within 3.3 Å and satisfied expected helix or sheet geometry. The magnitudes (D_a) and rhombicities (R) of the gel alignment tensors were determined from initial structural models and $^{1}D_{HN-N}$ values using PALES (Zweckstetter and Bax, 2000). Hydrogen bond and RDCs were incorporated in subsequent structure calculations. In the final calculations 100 structures were generated, and the lowest 10 energy structures were selected for further refinement in an explicit water environment (Linge et al., 2003). Statistics for solution structure calculations are provided in Table 4.1.

Calculation of the complex between Sgt2-N and Get5-UBL

Chemical shift changes between free Sgt2-N or Get5-UBL and the proteins while in complex were quantified using the method described by Ayed et al. (2001). For each main chain cross peak that could be clearly resolved on ^{1}H-^{15}N HSQC spectra collected at 25°C, changes in the ^{1}H and ^{15}N dimensions were recorded. The combined chemical shift perturbation (CSP) was calculated according to equation 6, scaling the nitrogen shifts by 0.15.

$$\mathrm{CSP} = \sqrt{\Delta\delta_{\mathrm{H}}^2 + 0.15 * \Delta\delta_{\mathrm{N}}^2} \quad (6)$$

Symmetry within the Sgt2-N dimer is disrupted while in complex with Get5-UBL, and for a subset of residues two chemical shift distances were determined, as indicated in Figure 4.8 and Figure 4.9A.

Although we could resolve this in the two-dimensional HSQC experiments, we were unable perform complete chemical assignment at 25°C due to signal degradation from line broadening. Instead, we collected data at 37°C, significantly improving data quality. This also acted to average two states of Sgt2-N for all residues. We therefore artificially treated the protein as symmetric during assignment. The 2D ^{15}N-edited HSQC, ^{15}N- or ^{13}C-edited NOESY-HSQC and ^{15}N-edited TOCSY-HSQC spectra were sufficient to reassign main chain and side chain atoms of the helical regions using the data for the free protein as a reference. TALOS+ predicted secondary structure and dihedral angles were similar to free protein. The structure of Sgt2-N while in complex with Get5-UBL was determined using the same protocol as for the free protein, again imposing symmetry. We later rely on the Get5-UBL structure to break the symmetry in complex. Statistics for solution structure calculations are provided in Table 4.3. Compared to the free structure of Sgt2-N, the second helices are slightly arched away from the 2-fold symmetry axis (Figure 4.9B).

We used ^{1}H-^{15}N and ^{1}H-^{13}C HSQC, HNCACB, HNCO, H(C)CH-TOCSY and (H)CCH-TOCSY spectra to assign chemical shifts for labeled Get5-UBL in complex in natural abundance Sgt2-N. We calculated a structure of Get5-UBL in complex using the same protocol as described for the free protein. Statistics for solution structure calculations are

provided in Table 4.3. There are very little structural differences between the structure of Get5-UBL in complex and the crystal and solution structures of free Get5-UBL (Figure 4.9C).

We were unable to obtain usable signal on isotope filtered and edited NOESY-HSQC to use to assign unambiguous intermolecular NOEs. We instead identified 14 intermolecular NOE cross peaks from the ^{15}N- and ^{13}C-edited NOESY-HSQC spectra of the labeled Get5-UBL complex. These were cross peaks that could not be explained by the structure of Get5-UBL and had chemical shifts matching residues on helix 2 of Sgt2-N. They were all between residues that showed chemical shift perturbation or loss of binding affinity upon mutation. Only contacts between the main chain or side chain protons of Get5 and the main chain amide proton of Sgt2-N were considered, as the amide proton regions of the spectra were less crowded than the methylene or methyl proton regions. These NOE cross peaks were converted to distance constraints with upper bounds of 6 Å and were ambiguous between the two subunits of Sgt2-N.

We used these constraints as well as residual dipolar couplings and Ambiguous Interaction Restraints (AIRs) as input into HADDOCK to calculate a structure of the complex (De Vries et al., 2007; Dominguez et al., 2003). AIRs are defined on the level of residues based upon chemical shift perturbation, mutagenesis and solvent accessibility. Interacting residues are designated as active or passive. A restraint is satisfied if any atom from an active residue of one protein falls is within 2 Å of any atom of an active or passive restraint on the second protein. Rigid body docking generates an initial set of 1000 structures, and the 200 lowest energy structures are selected for semi-flexible

refinement followed by refinement in explicit solvent. One-half of AIRs are randomly deactivated in each structure generated. Since we artificially treated Sgt2-N as symmetric while interpreting data, we relied on Get5-UBL to explain asymmetry. We implemented this by designating interacting Sgt2-N residues as passive only (Asp31, Val35, and Cys39 on either subunit). Conversely, all interacting Get5-UBL residues were designated active (Lys79, Ile81, Lys118, Leu120, Lys122, Gly123, Lys124, Val125, Thr145, Met147). Dihedral and hydrogen bond restraints were also included in the docking calculations to preserve geometry during the semi-flexible refinement. Results of the docking calculation are presented in Table 4.4.

Crystallization, data collection and structure determination of SGTA-N and Get5-UBL

Sitting drop vapor diffusion crystallization trials were set up using commercially available screens (Hampton Research, Qiagen) and a Mosquito liquid handling robot (TTP Labtech). In each condition, 100 nl of protein solution were mixed with 100 nl of reservoir solution. Crystals of SGTA-N appeared in 45% 2-methyl-2,4-pentanediol, 0.2M ammonium acetate and 0.1 M Tris pH 8.5 at 4°C after two days. Additional crystals, corresponding to either single or double covalent adducts between Cys38 and β-mercaptoethanol, grew after 3 weeks at 4°C in 10% 2-propanol, 0.1 M sodium citrate and 26% PEG 400 pH 5.0. Crystals from either condition were plates with dimensions of approximately 10 × 20 microns. Clusters of crystals of Get5-UBL were initially observed overnight in 2.4 M sodium malonate pH 7.0 at 37°C. To slow down the vapor diffusion, a layer of 300 µl 1:1 paraffin oil to silicon oil was layered over the 200 µl of reservoir

solution. After 5 days, single microcrystals appeared that were approximately 1–2 microns across.

All crystals were flash frozen in liquid nitrogen directly from crystallization drops. Diffraction data were collected on beam line 12-2 at the Stanford Synchrotron Radiation Lightsource (SSRL) at 100 K using a Pilatus 6M detector and a microbeam. Crystals were located within cryoloops using an automated X-ray raster protocol. Datasets were collected from single crystals, integrated with XDS, and merged and scaled using AIMLESS (Evans, 2006; Kabsch, 2010). Phase information was recovered for crystals of SGTA or Get5-UBL by molecular replacement using the respective solution structures as a search models. Molecular replacement was performed using PHASER as implemented in PHENIX (Adams et al., 2010; McCoy et al., 2007). The Get5-UBL crystals contained three copies of the protein in the asymmetric unit. The complete sequence of the Get5-UBL protein could be modeled in each copy. All SGTA-N crystals contained a single homodimer in the asymmetric units. The 2.4 Å resolution structure has both Cys38 sulfhydryl groups reduced and is used for structural comparisons in the text. Residues 4–49 and residues 4–48 could be modeled for either subunit, with the last 5 or 6 residues of the C-terminus disordered. Two molecules of MPD are associated with the hydrophobic face of the four-helix bundle.

The other two crystals show a covalent modification at one or both copies of Cys38 (Figure 4.2E). We modeled this modification as a covalent adduct with β-mercaptoethanol present in the buffer. It is notable that these crystals appeared after the sample aged for several weeks, while the reduced crystals appeared within days of

sample preparation. The modification results in the most C-terminal residues, which are disordered in the reduced structure, folding over the conserved binding face and interacting with the hydroxyl group of the β-mercaptothanol adduct. The entire sequence could be modeled for subunits with the modification. Reciprocal space refinement was performed using PHENIX and manual rebuilding using COOT (Emsley et al., 2010). Statistics for data collection and model refinement are presented in Table 4.2.

Structure analysis and figures

Cartoon representations of protein structures were prepared using PyMol (Schrodinger, LLC), while surface representations were prepared using UCSF Chimera (Pettersen et al., 2004). Surface hydrophobicity was determined in Chimera using the scale of Kyte and Doolittle (1982) for individual residues. Electrostatic surface potentials were calculated using APBS with default values as implemented in the PDB2PQR webserver (Baker et al., 2001; Dolinsky et al., 2004).

Acknowledgements

We thank Yihong Ye, NIH, for providing unpublished data and discussion. We thank S. O. Shan, D. C. Rees, Y. Ye, A. Müller, M. Rao, C. J. M. Suloway and M. E. Rome for critical reading of the manuscript. We thank members of the laboratory for support and useful discussions. We thank Graeme Card, Ana Gonzalez and Michael Soltice for help with data collection at SSRL BL12-2. We thank Robert Peterson and Juli Feigon for use of the 800 MHz NMR spectrometer at UCLA. We thank Jost Vielmetter, Harry Gristick, Claude Rogers and Abigail Pulsipher for help with Biacore and ITC experiments. We are grateful to Gordon and Betty Moore for support of the Molecular Observatory at Caltech. Operations at SSRL are supported by the US DOE and NIH. W. M. C. is supported by NIH grant R01GM097572. The atomic coordinates, structure factors and/or NMR restraints have been deposited in the RCSB Protein Data Bank, www.pdb.org (PDB ID codes 2LXB, 4GOD, 4GOE, 4GOC, 2LXA, 4GOC and 2LXC for the Sgt2-N (NMR), three SGTA-N (X-ray), Get5-UBL (NMR), Get5-UBL (X-ray) and Get5-N/Sgt2 structures respectively).

Appendix A

THE COMPLEX PROCESS OF GETTING TAIL-ANCHORED MEMBRANE PROTEINS TO THE ER

Abstract

Biosynthesis of membrane proteins requires that hydrophobic transmembrane (TM) regions be shielded from the cytoplasm while being directed to the correct membrane. Tail-anchored (TA) membrane proteins, characterized by a single C-terminal TM, pose an additional level of complexity because they must be post-translationally targeted. In eukaryotes, the GET pathway shuttles TA proteins to the endoplasmic reticulum. The key proteins required in yeast (Sgt2 and Get1–5) have been under extensive structural and biochemical investigation during recent years. The central protein Get3 utilizes nucleotide linked conformational changes to facilitate substrate loading and targeting. Here we analyze this complex process from a structural perspective, as understood in yeast, and further postulate on similar pathways in other domains of life.

Adapted from

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

Delivery of membrane proteins to the proper membrane is a critical process that is highly regulated. The field of protein targeting has recently added a new pathway involved in targeting the special class of tail-anchored (TA) membrane proteins to the ER. After the initial discovery of proteins involved in this pathway a wealth of genetics, biochemistry and structural information has rapidly elucidated a complex process of handoffs of the TA substrate. The pathway starts by transfer of the substrate to the Get4/Get5/Sgt2 sorting complex that then loads the TA onto the targeting chaperone Get3, which is subsequently released at the membrane by the Get1/Get2 receptor complex. While recent reviews describe these initial characterizations (Hegde and Keenan, 2011; Shao and Hegde, 2011), in this review we will summarize the current state of the field from a structural perspective and more broadly comment on the possibility of similar pathways beyond fungi. Figure A.1 is an overall model of TA protein targeting, including the various ambiguities in certain steps, which will be referred to throughout the text.

Structural changes in the Get3 molecular machine

The general structure and mechanism of Get3, which we briefly summarize here, has been recently reviewed (Hegde and Keenan, 2011; Simpson et al., 2010). Get3 is a nucleotide hydrolase that modulates its conformation through nucleotide state. Similar to other SIMIBI class NTPases (named for representative members *si*gnal recognition particle, *Mi*nD and*Bi*oD), Get3 forms a homodimer through interactions between the nucleotide hydrolase domains (NHD) (Leipe et al., 2002). The subunits rotate relative to each other from an open state in apo form or bound to ADP to a more compact closed state bound to $Mg^{2+}ADP{\cdot}AlF_4^-$, Mg^{2+}AMPPNP or Mg^{2+}ADP, forming an intersubunit

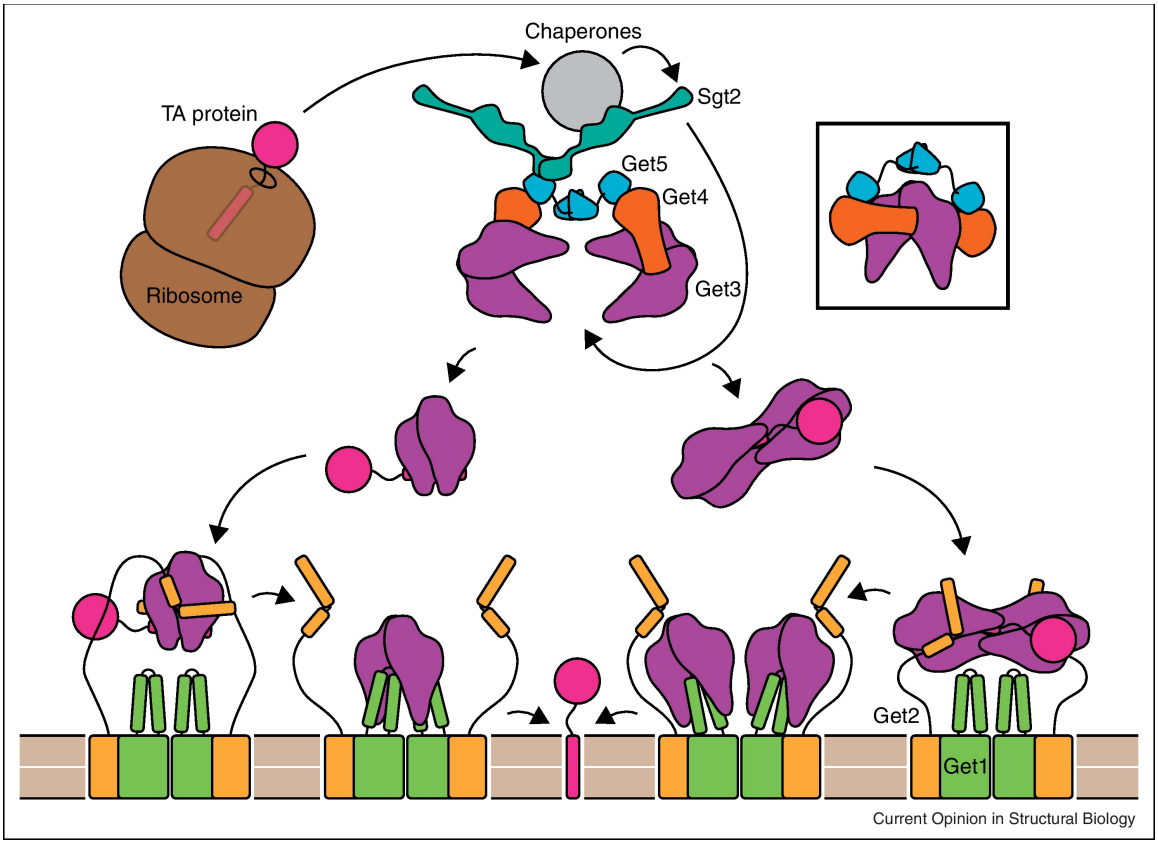


Figure A.1. A model for TA targeting by the GET pathway. After protein synthesis is complete, a complex consisting of cellular chaperones, two copies each of Get4 and Get5 and at least one dimer of Sgt2 binds the TM helix of TA proteins. The Get4/Get5 complex recruits Get3, and Sgt2 transfers ER destined TA proteins to Get3. Each Get4 may bind a separate dimer of Get3, or the same dimer (boxed). The Get3/TA protein complex may contain two or four copies of Get3. The stoichiometry of the Get1/Get2 complex within the ER membrane is unknown but is shown here as a dimer of 1:1 Get1 to Get2 dimers. Initially Get2 binds the Get3/TA protein complex and is then displaced by Get1. Get1 binding is coupled to the opening of Get3, leading to release and integration of the TA protein. In the case of a dimeric Get3/TA complex, Get1 and Get2 could bind the Get3 subunits symmetrically (left pathway). Alternatively, Get2 and Get1 could bind a tetramer of Get3 asymmetrically (right pathway).

hydrophobic groove from α-helices connected to the NHD (Bozkurt et al., 2009; Hu et al., 2009; Suloway et al., 2009; Yamagata et al., 2010).

There are three models for how Get3 sequesters a TA protein predominantly based on structures (Figure A.2A,B). The first two models are based on a dimer of Get3, which is the dominant form of purified fungal Get3. In the prevailing model, an α-helical TA is captured by the hydrophobic groove formed by the closed Get3 dimer (Mateja et al., 2009) (Figure A.2D). The alternative dimer model posits that pairs of amphipathic α-helices extending out from either side of the groove bind the TA through a hydrophobic patch (Yamagata et al., 2010). The third model correlates the observation that heterologously purified Get3/TA complexes contain four copies of Get3 with the structure of a tetrameric archaeal homolog. In this model, the TA would be sequestered in a hydrophobic chamber formed by the grooves of opposing dimers (Figure A.2E and Figure A.3) (Suloway et al., 2012). In all models, similar residues mediate potential TA interactions. The importance of these regions in TA binding has been shown through various mutagenesis experiments (Mateja et al., 2009; Suloway et al., 2009; Yamagata et al., 2010). Additionally, changes in these regions upon TA binding have been demonstrated through hydrogen exchange mass spectrometry of Get3 complexed with TA protein (Bozkurt et al., 2009). In general, results from biochemical experiments are compatible with both dimer and tetramer models. The simplicity of the dimer model is consistent with structures seen in complex with partners (see below) and binding of functionalized TA substrates (Leznicki et al., 2011). The tetramer model allows for complete shielding of the hydrophobic domain but necessitates additional assembly and

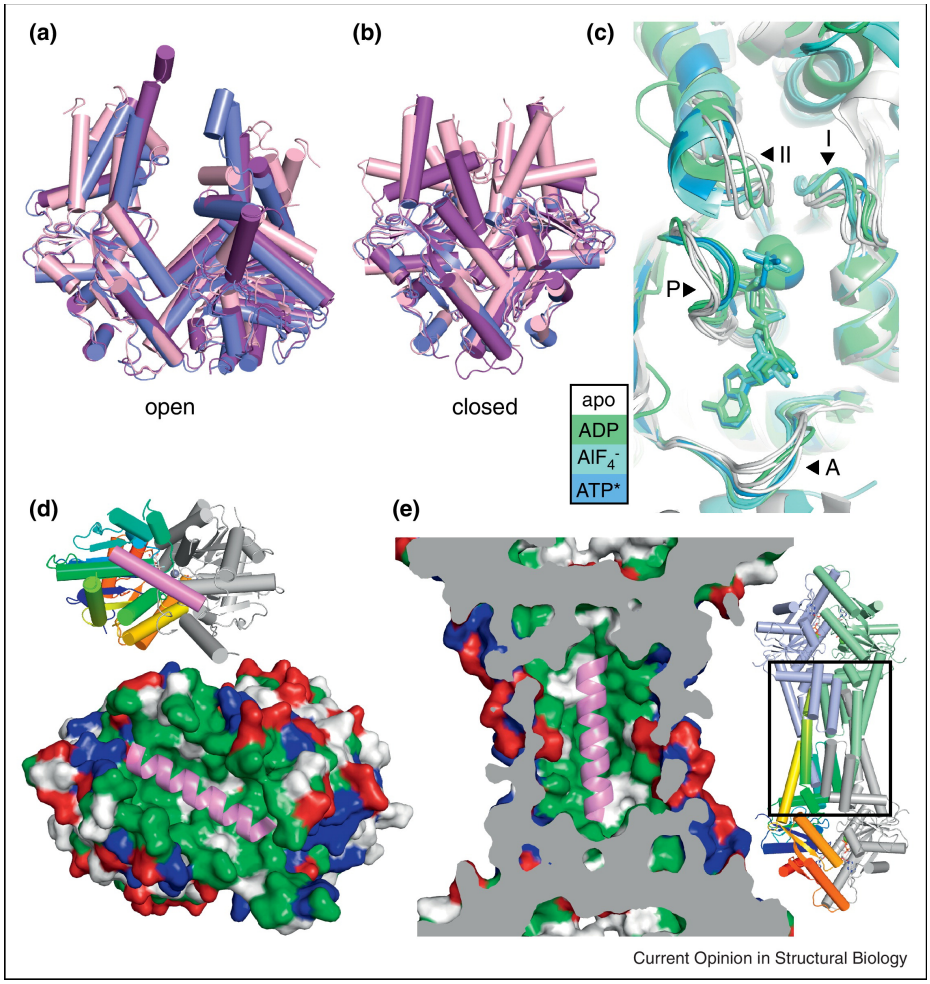


Figure A.2. Conformations of Get3 and models for TA binding. (A) Open form structures of Get3 in a cartoon representation aligned by the right subunit, using PDB IDs 3A36 (purple), 3H84 (blue) and 2WOO (pink). (B) Closed form structures shown as in (A), using PDB IDs 3IQW (purple), 3IO3 (blue) and 2WOJ (pink). (C) The NHD of aligned Get3 structures from one subunit. Motifs are indicated with arrowheads and labeled A (A-loop), P (P-loop), I (switch I) and II (switch II). Structures are colored according to the nucleotide bound with white for no nucleotide (PDB ID: 3A36, PDB ID: 2WOO, PDB ID: 3SJA and PDB ID: 3SJC), green for ADP (PDB ID: 3IQX and PDB ID: 3SJD), cyan for ADP·AlF_4^- (PDB ID: 2WOJ, PDB ID: 3ZQ6 and PDB ID: 3ZS9) and blue for AMPPNP (PDB ID: 3IQW). (D) A model for TA binding by the Get3 dimer. Get3 (PDB ID: 2WOJ) is shown in a surface representation colored hydrophobic (green), positively (blue) and negatively charged (red). A cartoon representation of a TA from Secβ (PDB ID: 1RHZ), colored pink, is shown in the groove formed in the closed structure of Get3. The inset shows a cartoon representation of Get3 with the left subunit colored ramped and the right subunit in gray. (E) A TA binding model for a tetramer of Get3 (Suloway et al., 2012) represented as in (D). The structure of tetrameric Get3 (PDB ID: 3UG6) is cut away to show the central cavity with a TA modeled inside, similar to (D). The area shown is indicated with a box on the overall structure in the inset.

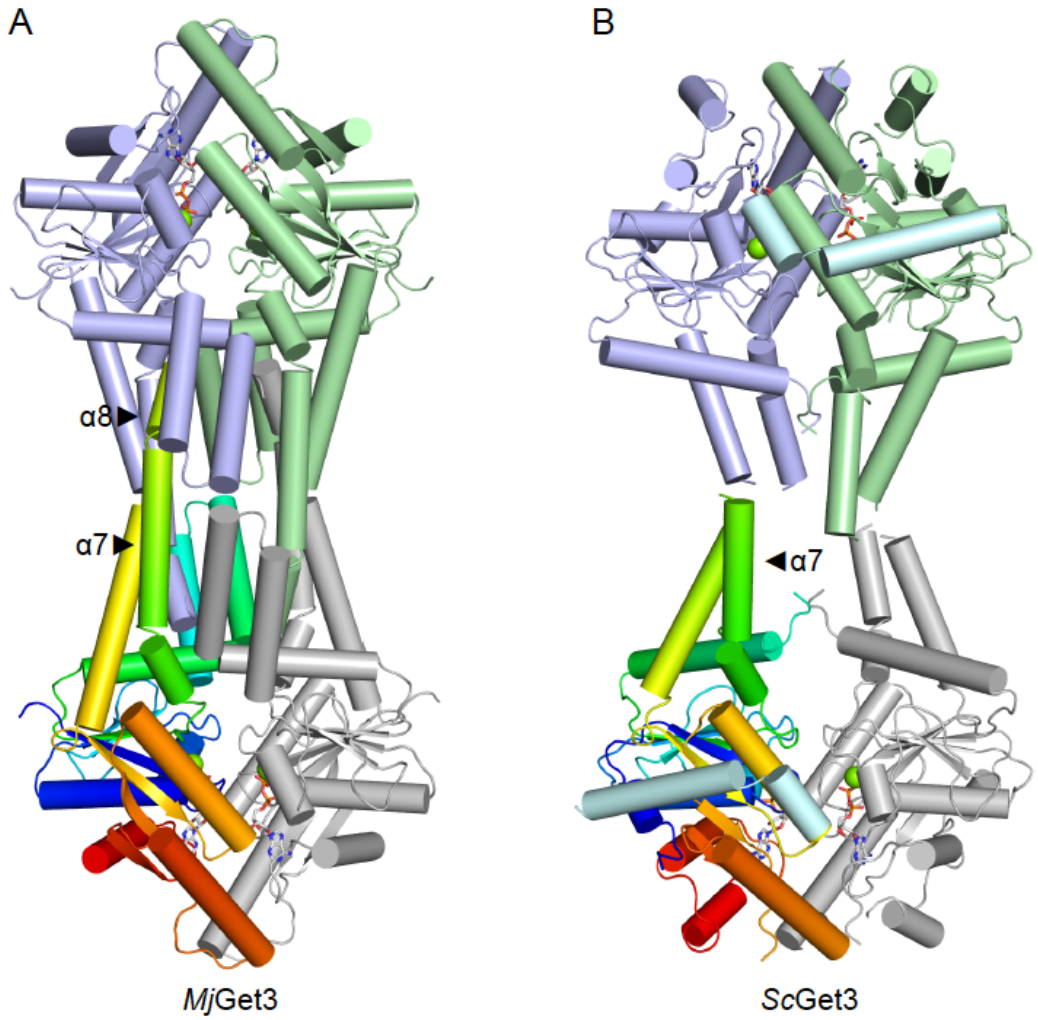


Figure A.3. Structures of Get3 tetramers. (A) Tetrameric *Mj*Get3 (PDB ID: 3UG6) shown in a cartoon representation with one subunit color ramped from N to C-terminus and additional subunits colored gray, blue and green. (B) A crystallographic tetramer of *Sc*Get3 (PDB ID: 3SJD, shown as in (A)) in complex with the soluble domain of Get2 (cyan).

disassembly steps, yet agrees with complexes analyzed in solution (Bozkurt et al., 2009; Favaloro et al., 2010; Suloway et al., 2012). The oligomeric state of Get3 in the native TA complex remains a subject for future study.

Many nucleotide hydrolases are molecular machines that act as switches, utilizing nucleotide state and the energy of hydrolysis to alter their conformation with a broad spectrum of functionality from signaling interactions to actively driving processes. The Get3 homodimer uses ATP hydrolysis to switch between open and closed conformations during its targeting cycle (Figure A.2A,B). Related SIMIBI class NTPases in the MinD/Mrp family (Leipe et al., 2002) switch by changes in dimer conformation like Get3 (NifH and ArsA pseudodimer) or convert between monomeric and dimeric states (MinD and Soj) (Lutkenhaus and Sundaramoorthy, 2003). Changes from open and closed dimer states were originally seen for NifH (Schindelin et al., 1997) and have been modeled for ArsA (Ajees et al., 2011).

In SIMIBI proteins, nucleotide induced conformational changes control interactions with protein factors. For Get3, this involves ferrying the TA substrate from the sorting complex (Sgt2/Get4/5) to the membrane receptor (Get1/2) (see below). This is analogous to ArsA that receives arsenite from ArsD and then exports through the membrane protein ArsB (Lin et al., 2007). NifH switches between a MoFe bound state ferrying electrons along the nitrogenase cycle (Burgess and Lowe, 1996). In the best characterized of these cases, in addition to rotation across the dimer interface, the switch II loop alters its conformation among different nucleotide states (Figure A.2C and Figure A.4). This loop connects to structural features mediating partner interactions. In the case of Get3,

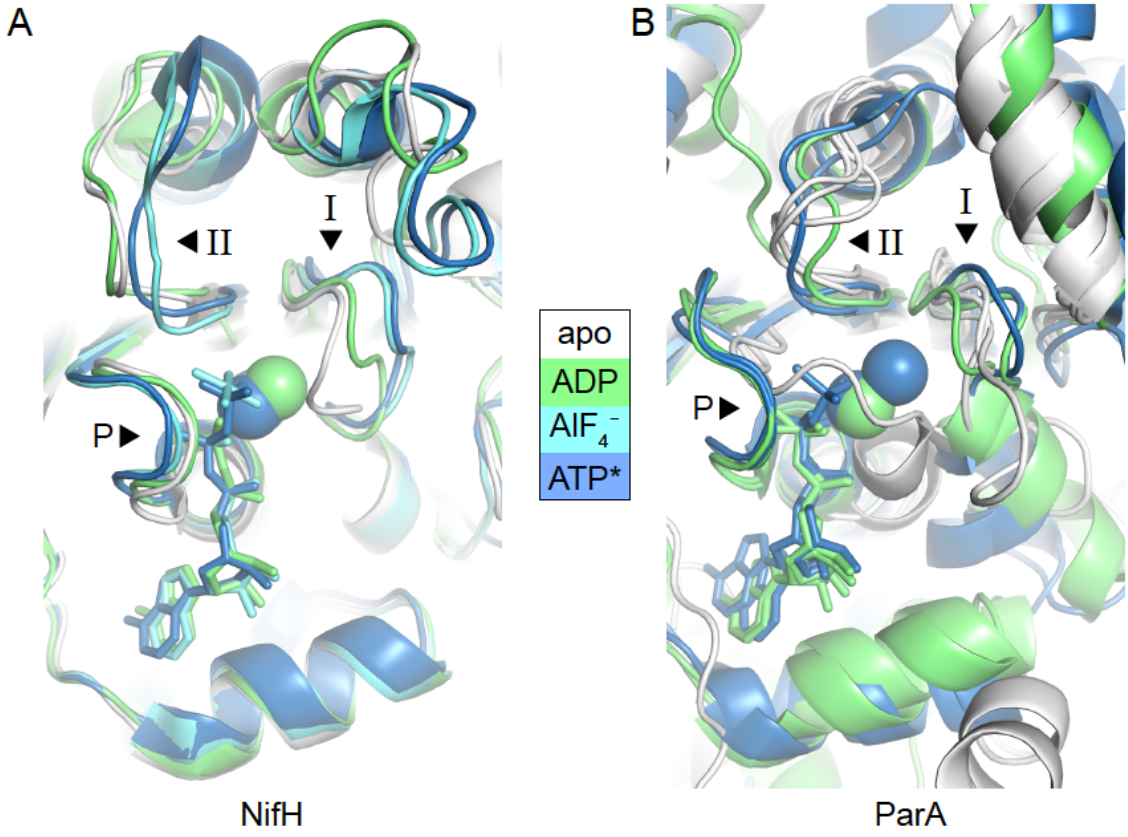


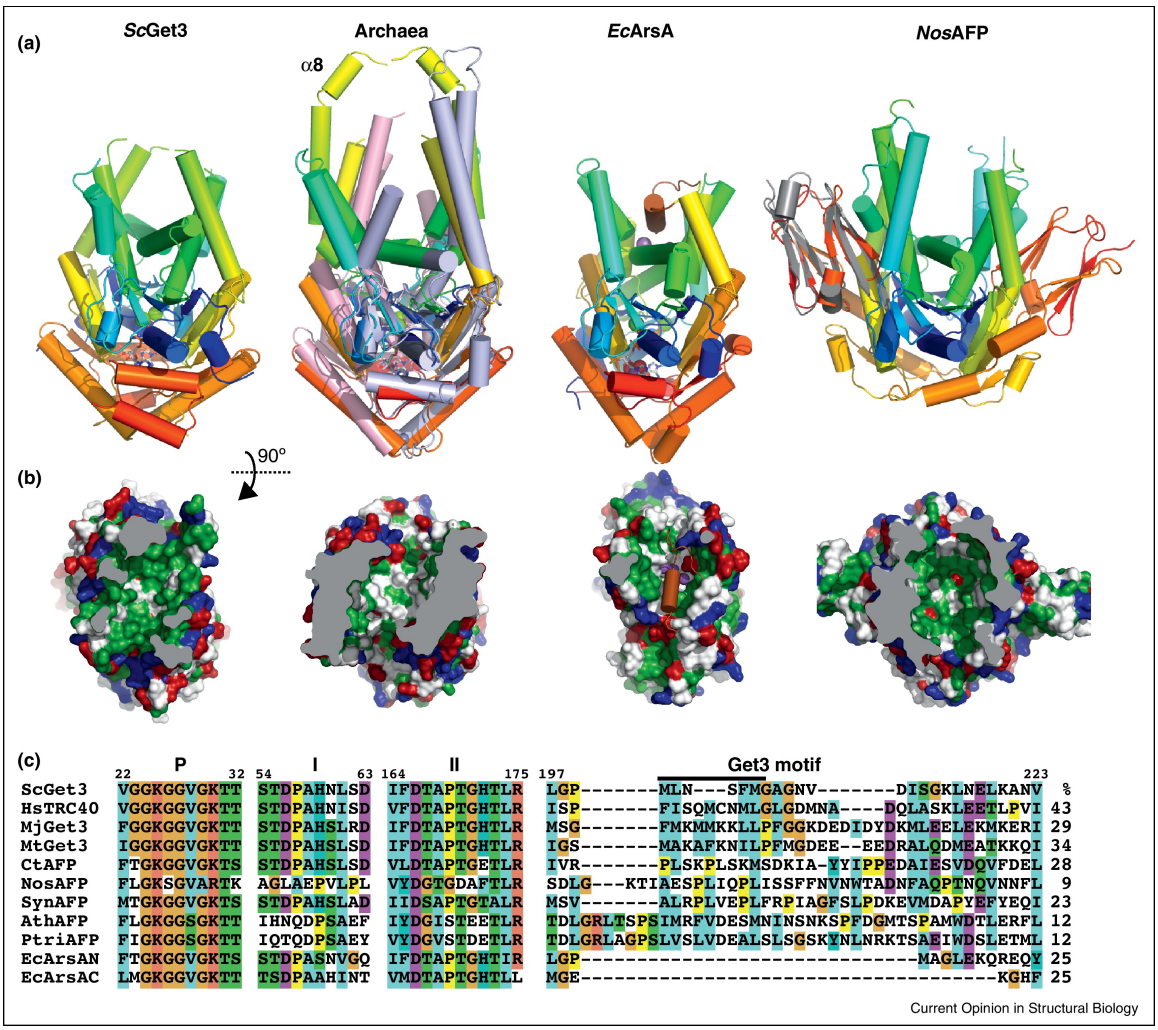
Figure A.4. Conformations of the NHDs of NifH and ParA in different nucleotide states. (A) A structural alignment of NifH NHDs represented as in Figure A.2C. Apo (PDB ID: 2AFH), ADP bound (PDB ID: 2AFI), ADP·AlF_4^- bound (PDB ID: 1M34) and AMPPCP bound (PDB ID: 2AFK) forms are shown. (B) As in (A) for ParA structures. Apo (PDB ID:3EZ7, PDB ID: 3EZ9, PDB ID: 3EZF), ADP bound (PDB ID: 3CWQ, PDB ID: 3EZ2) and ATP/ATPγS bound (PDB ID: 3EA0, PDB ID: 2OZE) forms are shown.

switch II is linked to α6, which is connected to the putative substrate-binding groove (Figure A.2C). Therefore, the nucleotide state can modulate transitions associated with TA-binding. Similar switch II conformational changes occur during the NifH and ParA nucleotide cycle (Figure A.3).

The ArsA fold

Get3 belongs to a subset of the SIMIBI family of proteins characterized by the first member ArsA. ArsA contains two NHDs linked together forming a pseudodimer (Figure A.5A) (Zhou et al., 2000). In the first structure, substrate antimony atoms bind in a groove formed by two loops that extended from the NHD, linking the nucleotide-binding pocket to liganding cysteines and histidines (Figure A.5B). ATP hydrolysis is proposed to induce conformational changes that coordinate the release of the toxic metals to the ArsB exporter. Interestingly, the two halves have only 26% identity, and this asymmetry appears to play a functional role (Fu et al., 2010). The substrate-binding groove appears to be a unique identifier for this family (Figure A.5A,B).

Fungal Get3 has approximately 25% identity to each half of ArsA; therefore, based on homology alone, distinguishing an ArsA from a Get3 is difficult. In most genomic annotations members of this fold family are identified as ArsA homologs. Several key differences to distinguish Get3 from ArsA have been identified. The simplest is that Get3 is a homodimer. Instead of containing the ArsA metal ligand residues (Boskovic et al., 1996; Stefanovic and Hegde, 2007), Get3 homologs contain a unique "Get3 motif" insertion (Mateja et al., 2009) (Figure A.5C). Moreover, Get3 homologs typically contain a pair of cysteines, a CXXC motif, at their dimer interface that coordinate zinc


(a)
ScGet3
Archaea
EcArsA
NosAFP
α8
(b)
90°
(c)
P
I
II
Get3 motif
ScGet3 VGGKGGVGKTT STDPAHNLSD IFDTAPTGHTLR LGP-------MLN---SFMGAGNV-------DISGKLNELKANV %
HsTRC40 VGGKGGVGKTT STDPAHNISD VFDTAPTGHTLR ISP-------FISQMCNMLGLGDMNA----DQLASKLEETLPVI 43
MjGet3 FGGKGGVGKTT STDPAHSLRD IFDTAPTGHTLR MSG-------FMKMMKKLLPFGGKDEDIDYDKMLEELEKMKERI 29
MtGet3 IGGKGGVGKTT STDPAHSLSD IFDTAPTGHTLR IGS-------MAKAFKNILPFMGDEE--EEDRALQDMEATKKQI 34
CtAFP FTGKGGVGKTS STDPAHSLSD VLDTAPTGETLR IVR-------PLSKPLSKMSDKIA-YYIPPEDAIESVDQVFDEL 28
NosAFP FLGKSGVARTK AGLAEPVLPL VYDGTGDAFTLR SDLG---KTIAESPLIQPLISSFFNVNWTADNFAQPINQVNNFL 9
SynAFP MTGKGGVGKTS STDPAHSLAD IIDSAPTGTALR MSV-------ALRPLVEPLFRPIAGFSLPDKEVMDAPYEFYEQI 23
AthAFP FLGKGGSGKTT IHNQDPSAEF IYDGISTEETLR TDLGRLTSPSIMRFVDESMNINSNKSPFDGMTSPAMWDTLERFL 12
PtriAFP FIGKGGSGKTT IQTQDPSAEY VYDGVSTDETLR TDLGRLAGPSLVSLVDEALSLSGSKYNLNRKTSAEIWDSLETML 12
EcArsAN FTGKGGVGKTS STDPASNVGQ IFDTAPTGHTIR LGP----------MAGLEKQREQY 25
EcArsAC LMGKGGVGKTT TSDPAAHINT VMDTAPTGHTLL MGE----------KGHF 25
Current Opinion in Structural Biology

presumably to stabilize the dimer, analogous to the ArsA linker (Metz et al., 2006; Stefanovic and Hegde, 2007).

Archaea contain both ArsA and Get3 homologs (Borgese and Righi, 2010). Only about half of the sequenced genomes contain the Get3 homolog and half of these lack the CXXC motif (Suloway et al., 2012). While the role of the ArsA homolog is likely similar to that in eubacteria, the presence of a Get3 homolog suggests membrane protein targeting occurs. Indeed, in heterologous systems archaeal Get3 can form complexes with a variety of TA proteins and in one case facilitates insertion (Sherrill et al., 2011; Suloway et al., 2012). This implies a unique pathway, as other GET pathway members

---

Figure A.5. Members of the Get3/ArsA fold family. (A) Cartoon representations of dimers of members of the ArsA fold viewed down the putative TA binding groove. The structures are the transition state *S. cerevisiae* Get3 (PDB ID:2WOJ), the archaeal *M. jannaschii* (PDB ID: 3UG6, monomers color ramped) and *M. thermoautotropicum* Get3s (PDB ID: 3ZQ6, monomers light blue/pink), *E. coli* ArsA (PDB ID: 1F48, each pseudodimer color ramped with the helix filling the groove in brown), and the cyanobacterial homolog *Nostoc* sp. all4481 (PDB ID: 3IGF, monomers color ramped). The human α-crystallin structure (PDB ID: 2WJ7) has been aligned in gray to the equivalent cyanobacterial domain. Metals are shown as spheres and nucleotides as sticks. (B) Surface representations of the structures in (A) rotated 90° with residues colored as in figure A.2D. Some foreground residues have been removed for clarity and the truncation is colored gray. *Mj*Get3 is shown for the archaeal homologs and the loop filling the groove has been removed in the ArsA surface calculation to show the comparable groove. (C) An alignment of notable regions of the ArsA family proteins from ClustalW. Numbering and region names are based on the *S. cerevisiae* protein similar to figure A.2C. A black bar identifies residues corresponding to α8. The sequences are ScGet3 (*S. cerevisiae* UniProt ID: Q12154), HsTRC40 (*Homo sapiens* UniProt ID: O43681), MjGet3 (*Methanocaldococcus jannaschii* UniProt ID: Q58542), MtGet3 (*Methanobacter thermoautotropicum* UniProt ID: O27555) CtAFP (*Chlorobium tepidum* UniProt ID: Q46366), NosAFP (*Nostoc* sp. *PCC 7120* UniProt ID: Q8YNT0), SynAFP (*Synechocystis* sp. *PCC 6803* UniProt ID: F7UTP7), AthAFP (*Arabidopsis thaliana* UniProt ID: Q6DYE4), PtriAFP (*Populus trichocarpa* or California poplar UniProt ID: B9HWM7) and EcArsA (*E. coli* plasmid R773 N-terminal and C-terminal domains UniProt ID: P08690). Numbers to the right are percentage identity to *Sc*Get3 excluding crystallin domains.

have not been identified. Recent structural studies of the homologs from *Methanobacter thermoautotripicum* (*Mt*Get3) and *Methanocaldococcus jannaschii* (*Mj*Get3) confirm the structural homology of archaeal Get3 to fungal Get3 (Figure A.5A,B) (Sherrill et al., 2011; Suloway et al., 2012). Both archaeal structures are in the closed form; however, unlike fungal Get3 homologs the substrate binding loops are ordered and extended. In *Mj*Get3 this results in a tetramer, a dimer of dimers, with a closed hydrophobic chamber stabilized by α8 (Figure A.2E and Figure A.3) (Suloway et al., 2012). Moreover, of the archaeal homologs where oligomeric state was analyzed, all are capable of forming stable tetramers in solution including the *Thermococcus kodakaerensis* homolog, which lacks the CXXC motif and is competent for TA binding (Suloway et al., 2012).

Looking more broadly at sequence homology a third class of this fold family becomes apparent. Here, the fold is found in photosynthetic organisms. For clarity, we will refer to members of this protein class as the "ArsA family Fold associated with Photosynthesis" or AFP. They can be found in green sulfur bacteria, cyanobacteria and are nuclear encoded with chloroplast-targeting signals in plants. Each of these organisms is characterized by photosynthetic machinery in specialized membranes, either the chlorosome or thylakoid (Hohmann-Marriott and Blankenship, 2011). A recent structure from this class, all4481 from the cyanobacteria *Nostoc* sp. *PCC 7120* (*Nos*AFP), was deposited by the Northeast Structural Genomics Consortium. The structure reveals a closed Get3-like fold containing the hydrophobic groove but lacking the nucleotide-binding pocket (Figure A.5A,B). A small heat shock protein (HSP)/crystallin domain is appended to the C-terminus that contacts the Get3 surface used in yeast for partner recognition (see below). The remarkable structural similarity of *Nos*AFP is not evident

from the low sequence identity (9%) to *Sc*Get3; however, similar homologs such as that from the green sulfur bacteria *Chlorobium tepidum* (*Ct*AFP) have higher identity (28%) and appear to include the nucleotide-binding pocket (Figure A.5C). The conservation of the AFP class from ancient photosynthetic organisms to modern plants suggests a critical role in biosynthesis of the photosynthetic membranes. It will be exciting to see if there is a role for this protein in protein targeting to membranes.

Get1 and Get2 regulate membrane insertion steps

The integral membrane proteins Get1 and Get2 form the ER membrane bound complex required for TA-insertion by Get3 (Schuldiner et al., 2008); however, only a homolog for Get1 has been found in higher eukaryotes (Vilardi et al., 2011). Both proteins have single cytoplasmic domains that can bind Get3 in the absence of the membrane components. The Get1 cytoplasmic domain, which connects the first and second transmembrane helices, is a coiled-coil motif that extends between the subunits of a Get3 dimer (Mariappan et al., 2011; Stefer et al., 2011) (Figure A.6A). A portion of Get1 binds at a groove on the surface of Get3 formed by α10 and α11. At the other Get3 subunit, Get1 extends into the nucleotide-binding pocket, preventing ATP or ADP from binding (Figure A.6A,C,D). Get3 is in an open or semi-open conformation, with Get1 acting as a wedge that would block a complete transition to the closed state. Nearly the entire cytoplasmic loop is modeled in these structures; therefore, bound Get3 must be in close proximity to the ER membrane.

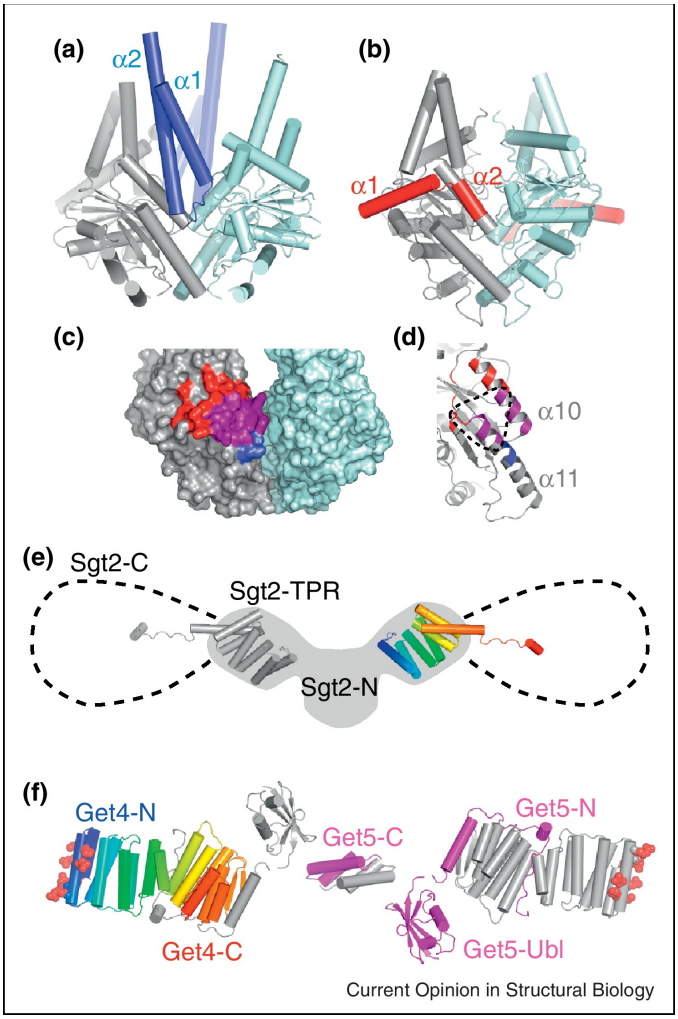


Figure A.6. Interactions between Get3 and other GET pathway members. (A) Crystal structure of Get3 in complex with the Get1 cytoplasmic domain (blue) (PDB ID 3ZS8). One Get3 subunit is gray and the other pale blue. (B) Crystal structure of Get3 in complex with the Get2 cytoplasmic domain (red) (PDB ID: 3SJD). (C) Surface rendering of open Get3 (PDB ID: 3A36). Regions where Get1 (blue) and Get2 (red) interact with the gray subunit are highlighted, with the overlapping region in magenta. (D) Close-up of Get3 α10 and α11, colored as in (C). The expected Get4 binding interface is indicated with a dotted line. (E) Model of the structure of Sgt2. Two copies of the TPR domain (PDB ID: 3SZ7) are displayed over a cartoon of a SAXS model of the N-terminal dimerization and TPR domains (gray, adapted from Chartron et al., 2011). The additional 107 residues of the C-terminal domains rotate freely from the TPR and a potential range of motion is indicated with dotted lines. (F) Structure of the Get4/Get5 heterotetramer. Residues shown to be critical for Get3 interaction are shown as spheres (Chartron et al., 2010). Subunits are arranged based upon SAXS models (Chartron et al., 2012b). One copy of Get4 is color ramped from N-termini (blue) to C-termini (red). One copy of Get5 is colored magenta and the remaining Get4 and Get5 subunits are gray.

The cytoplasmic domain of Get2 consists of the N-terminus of the protein. Overall sequence conservation is poor, with the exception of the first 35 amino acids that are sufficient to bind Get3 (Mariappan et al., 2011; Stefer et al., 2011). This region forms two helices that wrap along the outer surface of Get3, including the groove where Get1 interacts (Figure A.6B–D). Unlike Get1, the ordered portions of Get2 do not contact both subunits of Get3. With bound $Mg^{2+}ADP{\cdot}AlF_4^-$, Get3 is nearly identical to previous transition state structures (Bozkurt et al., 2009; Mariappan et al., 2011; Mateja et al., 2009). A second structure containing $Mg^{2+}ADP$ also is a closed Get3 dimer (Stefer et al., 2011). Despite this, helices that line the TA protein-binding site are in conformations similar to open structures, although resolution limited the extent of modeling and refinement (Figure A.3). Intriguingly, in both structures the surface of Get3 that interacts with Get2 is unperturbed relative to Get3 alone. Since Get2 does not contact the TA protein binding loops or the nucleotide-binding pocket, it is unclear why a closed conformation of Get3 would be selected. It is also noteworthy that in the $Mg^{2+}ADP$ bound structure a crystallographic axis relates a second Get3 dimer into an arrangement similar to the *M. jannaschii* tetramer (Suloway et al., 2012) (Figure A.3).

The structures of Get1 and Get2 and other biochemical data led to a general model for events at the membrane (Mariappan et al., 2011; Stefer et al., 2011; Wang et al., 2010) (Figure A.1). The flexible cytoplasmic domain of Get2 initially captures the Get3/TA protein complex. As Get3 approaches the membrane, Get1 first displaces Get2 and then facilitates a transition from closed to open Get3. This corresponds with integration of the TA protein into the membrane and release of bound nucleotide. ATP then displaces Get1, releasing Get3 from the membrane complex. The exact nature, however, of the

Get1/Get2/Get3 complex at the membrane remains to be determined and is crucial for understanding the mechanism of insertion. Without a clear structural explanation for state selectivity for Get2, it is not obvious how the Get3/TA protein complex is favored from free Get3 or how immediate rebinding after an insertion cycle is prevented. Uncertainty about the stoichiometry of Get1 and Get2 results in alternative models of Get1 and Get2 binding at different subunits of dimeric or tetrameric Get3. Finally, the biophysical mechanism of TA protein insertion into the membrane and the requirements of Get3 or full-length Get1 and Get2, in this process are largely unknown.

Sgt2, Get4 and Get5 load Get3 with TA protein

The HSP co-chaperone Sgt2 mediates the committed step in TA protein targeting. It recruits a variety of HSP families via an internal tetratricopeptide repeat (TPR) domain (Chartron et al., 2011; Wang et al., 2010). The Sgt2 C-terminal domain binds the transmembrane helix of ER destined TA proteins, which are then handed to Get3 (Wang et al., 2010; Wang et al., 2011). In contrast, mitochondrial TA proteins remain associated with bound HSPs. This domain, rich in glutamine, methionine and asparagine, contains only a short conserved sequence and is weakly predicted as helical. The mechanism of TA protein selection and handoff to Get3 are, therefore, intriguing problems. Sgt2 is a homodimer mediated by its small N-terminal domain, and the TPR and C-terminal domains of the two subunits extend away from each other (Chartron et al., 2011) (Figure A.6E). The N-terminal domain also mediates the association with Get5, where a dimer of Sgt2 binds a single copy of Get5 (Chang et al., 2010; Chartron et al., 2011).

Get4 and Get5 act as an adaptor complex linking Sgt2 to Get3. Get4 is an alpha helical repeat protein that tightly binds to the N-terminal domain of Get5 (Bozkurt et al., 2010; Chang et al., 2010; Chartron et al., 2010). The Get5 sequence is followed by a ubiquitin-like domain (Ubl) and a C-terminal homodimerization domain, resulting in an extended heterotetrameric complex (Chartron et al., 2010; Chartron et al., 2012b) (Figure A.6F). The N-terminal face of Get4 is a conserved basic surface that mediates interaction with Get3 (Chang et al., 2010; Chang et al., 2012; Chartron et al., 2010). Although a high-resolution structure of this complex remains to be determined, mutagenesis and molecular modeling indicate that, like Get1 and Get2, Get4 binds Get3 at α10 and α11 (Figure A.6D). The *in vivo* stoichiometry, however, has not been established. In order for a single Get4/Get5 heterotetramer to bind both subunits of a Get3 dimer, as it has been proposed (Chang et al., 2012; Chartron et al., 2010), the Ubl and C-terminal domains of Get5 would either thread through the TA binding groove of Get3 or wrap around the NHDs, possibly positioning Sgt2 away from the TA protein binding groove. If the two Get4 copies bind independent Get3 dimers, the Sgt2/TA protein complex could be more accessible to Get3. This model also provides a framework for Get3 to tetramerize upon TA protein capture (Figure A.1) (Chartron et al., 2012b).

A complex analogous to Get4/Get5/Sgt2 appears to exist in vertebrates although with some differences in architecture and function. The homologs of Get4 and Get5, named TRC35 and Ubl4a, bind to Bag6 (alternatively named Bat-3 or Scythe) forming the Bag6 complex (Mariappan et al., 2010). Bag6 is an approximately 1000-residue protein with an N-terminal Ubl domain and a C-terminal Bag domain, with internal proline rich regions. SGTA, the homolog of Sgt2, uses its N-terminal domain to associate with Bag6, and

although it has not yet been demonstrated experimentally, Ubl4A is expected to bridge these proteins (Hegde and Keenan, 2011; Winnefeld et al., 2006). The Bag6 complex not only has the ability to load TA proteins onto the Get3 homolog, TRC40 (Leznicki et al., 2010; Mariappan et al., 2010), but also mediates protein degradation pathways for mislocalized membrane proteins, retrotranslocated ER proteins and other defective proteins (Hessa et al., 2011; Minami et al., 2010; Wang et al., 2011). Determining how the interplay between these functions makes or breaks membrane proteins will provide fundamental insight into how TM segments are manipulated by the cell.

Concluding remarks

We are rapidly acquiring structural information for all of the components of the TA protein-targeting pathway. Careful analysis of analogous systems and components of the large and dynamic complexes generate testable models. These are tied together by low-resolution and biochemical methodologies. Despite recent leaps in understanding, many crucial questions remain to be answered. How are TA proteins initially directed towards the Get4/Get5/Sgt2 sorting complex? How does this complex recruit Get3 and what is the mechanism for handoff? What is the structure of the physiological Get3/TA protein complex? How is the TA protein integrated into the membrane? Further afield are questions about the functional roles of structurally homologous proteins in other domains of life. Clearly, understanding the detailed biosynthesis of this important class of proteins will continue to excite for some time to come.

Bibliography

Abbas-Terki, T., Donze, O., Briand, P.A., and Picard, D. (2001). Hsp104 interacts with Hsp90 cochaperones in respiring yeast. Mol. Cell. Biol. *21*, 7569–7575.

Abell, B.M., and Mullen, R.T. (2011). Tail-anchored membrane proteins: exploring the complex diversity of tail-anchored-protein targeting in plant cells. Plant Cell Rep. *30*, 137–151.

Abell, B.M., Pool, M.R., Schlenker, O., Sinning, I., and High, S. (2004). Signal recognition particle mediates post-translational targeting in eukaryotes. EMBO J. *23*, 2755–2764.

Abell, B.M., Rabu, C., Leznicki, P., Young, J.C., and High, S. (2007). Post-translational integration of tail-anchored proteins is facilitated by defined molecular chaperones. J. Cell Sci. *120*, 1743–1751.

Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2010). PHENIX: A comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. Sect. D. Biol. Crystallogr. *66*, 213–221.

Ajees, A.A., Yang, J., and Rosen, B.P. (2011). The ArsD As(III) metallochaperone. BioMetals *24*, 391–399.

Angeletti, P.C., Walker, D., and Panganiban, A.T. (2002). Small glutamine-rich protein/viral protein U-binding protein is a novel cochaperone that affects heat shock protein 70 activity. Cell Stress Chaperones *7*, 258–268.

Apostolovic, B., Danial, M., and Klok, H.A. (2010). Coiled coils: Attractive protein folding motifs for the fabrication of self-assembled, responsive and bioactive materials. Chem. Soc. Rev. *39*, 3541–3575.

Arnold, K., Bordoli, L., Kopp, J., and Schwede, T. (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics *22*, 195–201.

Ayed, A., Mulder, F.A.A., Yi, G.S., Lu, Y., Kay, L.E., and Arrowsmith, C.H. (2001). Latent and active p53 are identical in conformation. Nat. Struct. Biol. *8*, 756–760.

Baker, N.A., Sept, D., Joseph, S., Holst, M.J., and McCammon, J.A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. Proc. Natl. Acad. Sci. USA *98*, 10037–10041.

Bahrami, A., Assadi, A.H., Markley, J.L., and Eghbalnia, H.R. (2009). Probabilistic interaction network of evidence algorithm and its application to complete labeling of peak lists from protein NMR spectroscopy. PLoS Comput. Biol. *5*, e1000307.

Banky, P., Roy, M., Newlon, M.G., Morikis, D., Haste, N.M., Taylor, S.S., and Jennings, P.A. (2003). Related protein-protein interaction modules present drastically different surface topographies despite a conserved helical platform. J. Mol. Biol. *330*, 1117-1129.

Bardiaux, B., Bernard, A., Rieping, W., Habeck, M., Malliavin, T.E., and Nilges, M. (2009). Influence of different assignment conditions on the determination of symmetric homodimeric structures with ARIA. Proteins *75*, 569–585.

Battle, A., Jonikas, M.C., Walter, P., Weissman, J.S., and Koller, D. (2010). Automated identification of pathways from quantitative genetic interaction data. Mol. Syst. Biol. *6*, 379.

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

Berjanskii, M., and Wishart, D. (2007). The RCI server: rapid and accurate calculation of protein flexibility using chemical shifts. Nucleic Acids Res. *35*, W531–W537.

Bernado, P., Mylonas, E., Petoukhov, M.V., Blackledge, M., and Svergun, D.I. (2007). Structural characterization of flexible proteins using small-angle X-ray scattering. J. Am. Chem. Soc. *129*, 5656–5664.

Beuck, C., Szymczyna, B.R., Kerkow, D.E., Carmel, A.B., Columbus, L., Stanfield, R.L., and Williamson, J.R. (2010). Structure of the GLD-1 homodimerization domain: insights into STAR protein-mediated translational regulation. Structure *18*, 377–389.

Blobel, G., and Dobberstein, B. (1975). Transfer of proteins across membranes. I. Presence of proteolytically processed and unprocessed nascent immunoglobulin light chains on membrane-bound ribosomes of murine myeloma. J. Cell Biol. *67*, 835–851.

Bolon, D.N., Grant, R.A., Baker, T.A., and Sauer, R.T. (2005). Specificity versus stability in computational protein design. Proc. Natl. Acad. Sci. USA *102*, 12724–12729.

Borgese, N., Brambillasca, S., and Colombo, S. (2007). How tails guide tail-anchored proteins to their destinations. Curr. Opin. Cell Biol. *19*, 368–375.

Borgese, N., and Fasana, E. (2011). Targeting pathways of C-tail-anchored proteins. Biochim. Biophys. Acta *1808*, 937–946.

Borgese, N., and Righi, M. (2010). Remote origins of tail-anchored proteins. Traffic *11*, 877–885.

Boskovic, J., Soler-Mira, A., Garcia-Cantalejo, J.M., Ballesta, J.P., Jimenez, A., and Remacha, M. (1996). The sequence of a 16,691 bp segment of *Saccharomyces cerevisiae* chromosome IV identifies the DUN1, PMT1, PMT5, SRP14 and DPR1 genes, and five new open reading frames. Yeast *12*, 1377–1384.

Bozkurt, G., Stjepanovic, G., Vilardi, F., Amlacher, S., Wild, K., Bange, G., Favaloro, V., Rippe, K., Hurt, E., Dobberstein, B.*, et al.* (2009). Structural insights into tail-anchored

protein binding and membrane insertion by Get3. Proc. Natl. Acad. Sci. USA *106*, 21131–21136.

Bozkurt, G., Wild, K., Amlacher, S., Hurt, E., Dobberstein, B., and Sinning, I. (2010). The structure of Get4 reveals an alpha-solenoid fold adapted for multiple interactions in tail-anchored protein biogenesis. FEBS Lett. *584*, 1509–1514.

Brinker, A., Scheufler, C., Von Der Mulbe, F., Fleckenstein, B., Herrmann, C., Jung, G., Moarefi, I., and Hartl, F.U. (2002). Ligand discrimination by TPR domains. Relevance and selectivity of EEVD-recognition in Hsp70 x Hop x Hsp90 complexes. J. Biol. Chem. *277*, 19265–19275.

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

Burgess, B.K., and Lowe, D.J. (1996). Mechanism of molybdenum nitrogenase. Chem. Rev. *96*, 2983–3012.

Callahan, M.A., Handley, M.A., Lee, Y.H., Talbot, K.J., Harper, J.W., and Panganiban, A.T. (1998). Functional interaction of human immunodeficiency virus type 1 Vpu and Gag with a novel member of the tetratricopeptide repeat protein family. J. Virol. *72*, 8461.

CCP4 (1994). The CCP4 suite: programs for protein crystallography. Acta Crystallogr. Sect. D. Biol. Crystallogr. *50*, 760–763.

Chang, Y.W., Chuang, Y.C., Ho, Y.C., Cheng, M.Y., Sun, Y.J., Hsiao, C.D., and Wang, C. (2010). Crystal structure of Get4-Get5 complex and its interactions with Sgt2, Get3, and Ydj1. J. Biol. Chem. *285*, 9962–9970.

Chang, Y.W., Lin, T.W., Li, Y.C., Huang, Y.S., Sun, Y.J., and Hsiao, C.D. (2012). Interaction surface and topology of Get3-Get4-Get5 protein complex, involved in targeting tail-anchored proteins to endoplasmic reticulum. J. Biol. Chem. *287*, 4783–4789.

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

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

Chartron, J.W., Suloway, C.J., Zaslaver, M., and Clemons, W.M., Jr. (2010). Structural characterization of the Get4/Get5 complex and its interaction with Get3. Proc. Natl. Acad. Sci. USA *107*, 12127–12132.

Chartron, J.W., VanderVelde, D.G., Rao, M., and Clemons, W.M., Jr. (2012b). Get5 carboxyl-terminal domain is a novel dimerization motif that tethers an extended Get4/Get5 complex. J. Biol. Chem. *287*, 8310–8317.

Chou, J.J., Gaemers, S., Howder, B., Louis, J.M., and Bax, A. (2001). A simple apparatus for generating stretched polyacrylamide gels, yielding uniform alignment of proteins and detergent micelles. J. Biomol. NMR *21*, 377–382.

Cole, C., Barber, J., and Barton, G. (2008) The JPRED 3 secondary structure prediction server. Nucleic Acids Res. *36*, W197–W201.

Copic, A., Dorrington, M., Pagant, S., Barry, J., Lee, M.C., Singh, I., Hartman, J.L.t., and Miller, E.A. (2009). Genomewide analysis reveals novel pathways affecting endoplasmic reticulum homeostasis, protein modification and quality control. Genetics *182*, 757–769.

Costanzo, M., Baryshnikova, A., Bellay, J., Kim, Y., Spear, E.D., Sevier, C.S., Ding, H., Koh, J.L., Toufighi, K., Mostafavi, S., et al. (2010). The genetic landscape of a cell. Science *327*, 425–431.

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

D'Andrea, L.D., and Regan, L. (2003). TPR proteins: the versatile helix. Trends Biochem. Sci. *28*, 655–662.

Das, A.K., Cohen, P.W., and Barford, D. (1998). The structure of the tetratricopeptide repeats of protein phosphatase 5: implications for TPR-mediated protein-protein interactions. EMBO J. *17*, 1192–1199.

Daura, X., Gademann, K., Jaun, B., Seebach, D., van Gunsteren, W.F., and Mark, A.E. (1999). Peptide folding: When simulation meets experiment. Angew. Chem. Int. Ed. *38*, 236–240.

Dauter, Z., Dauter, M., and Rajashankar, K.R. (2000). Novel approach to phasing proteins: derivatization by short cryo-soaking with halides. Acta Crystallogr. Sect. D. Biol. Crystallogr. *56*, 232–237.

De Vries, S.J., van Dijk, A.D.J., Krzeminski, M., van Dijk, M., Thureau, A., Hsu, V., Wassenaar, T., and Bonvin, A.M.J.J. (2007). HADDOCK versus HADDOCK: New features and performance of HADDOCK2.0 on the CAPRI targets. Proteins *69*, 726–733.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J. Biomol. NMR *6*, 277–293.

Delano, W.L. (1998). The PyMol Molecular Graphics System. (Delano Scientific, Palo Alto, CA).

Dominguez, C., Boelens, R., and Bonvin, A.M. (2003). HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. J. Am. Chem. Soc. *125*, 1731–1737.

Dong, A., Xu, X., Edwards, A.M., Chang, C., Chruszcz, M., Cuff, M., Cymborowski, M., Di Leo, R., Egorova, O., Evdokimova, E., et al. (2007). In situ proteolysis for protein crystallization and structure determination. Nat. Methods *4*, 1019–1021.

Doniach, S. (2001). Changes in biomolecular conformation seen by small angle X-ray scattering. Chem. Rev. *101*, 1763–1778.

Driessen, A.J., and Nouwen, N. (2008). Protein translocation across the bacterial cytoplasmic membrane. Annu. Rev. Biochem. *77*, 643–667.

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

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

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

Favaloro, V., Spasic, M., Schwappach, B., and Dobberstein, B. (2008). Distinct targeting pathways for the membrane insertion of tail-anchored (TA) proteins. J. Cell Sci. *121*, 1832–1840.

Favaloro, V., Vilardi, F., Schlecht, R., Mayer, M.P., and Dobberstein, B. (2010). Asna1/TRC40-mediated membrane insertion of tail-anchored proteins. J. Cell Sci. *123*, 1522–1530.

Fernandes, J.M., Macqueen, D.J., Lee, H.T., and Johnston, I.A. (2008). Genomic, evolutionary, and expression analyses of cee, an ancient gene involved in normal growth and development. Genomics *91*, 315–325.

Fielding, B.C., Gunalan, V., Tan, T.H., Chou, C.F., Shen, S., Khan, S., Lim, S.G., Hong, W., and Tan, Y.J. (2006). Severe acute respiratory syndrome coronavirus protein 7a interacts with hSGT. Biochem. Biophys. Res. Commun. *343*, 1201–1208.

Fleischer, T.C., Weaver, C.M., McAfee, K.J., Jennings, J.L., and Link, A.J. (2006). Systematic identification and functional screens of uncharacterized proteins associated with eukaryotic ribosomal complexes. Genes Dev. *20*, 1294–1307.

Franke, D., and Svergun, D.I. (2009). DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. J. Appl. Crystallogr. *42*, 342–346.

Fu, H.L., Ajees, A.A., Rosen, B.P., and Bhattacharjee, H. (2010). Role of signature lysines in the deviant walker a motifs of the ArsA ATPase. Biochemistry *49*, 356–364.

Giaever, G., Chu, A.M., Ni, L., Connelly, C., Riles, L., Veronneau, S., Dow, S., Lucau-Danila, A., Anderson, K., Andre, B., et al. (2002). Functional profiling of the Saccharomyces cerevisiae genome. Nature *418*, 387–391.

Gietz, R.D., and Schiestl, R.H. (2007). High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. Nature Protocols *2*, 31–34.

Gilmore, R., Blobel, G., and Walter, P. (1982). Protein translocation across the endoplasmic reticulum. I. Detection in the microsomal membrane of a receptor for the signal recognition particle. J. Cell Biol. *95*, 463–469.

Gold, M.G., Lygren, B., Dokurno, P., Hoshi, N., McConnachie, G., Tasken, K., Carlson, C.R., Scott, J.D., and Barford, D. (2006). Molecular basis of AKAP specificity for PKA regulatory subunits. Mol. Cell *24*, 383–395.

Hanzelmann, P., Stingele, J., Hofmann, K., Schindelin, H., and Raasi, S. (2010). The yeast E4 ubiquitin ligase Ufd2 interacts with the ubiquitin-like domains of Rad23 and Dsk2 via a novel and distinct ubiquitin-like binding domain. J. Biol. Chem. *285*, 20390–20398.

Hautbergue, G.M., and Golovanov, A.P. (2008). Increasing the sensitivity of cryoprobe protein NMR experiments by using the sole low-conductivity arginine glutamate salt. J. Magn. Reson. *191*, 335–339.

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

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

Hicke, L., Schubert, H.L., and Hill, C.P. (2005). Ubiquitin-binding domains. Nat. Rev. Mol. Cell Biol. *6*, 610–621.

Hill, J.E., Myers, A.M., Koerner, T.J., and Tzagoloff, A. (1986). Yeast/E. coli shuttle vectors with multiple unique restriction sites. Yeast *2*, 163–167.

Hillenmeyer, M.E., Fung, E., Wildenhain, J., Pierce, S.E., Hoon, S., Lee, W., Proctor, M., St Onge, R.P., Tyers, M., Koller, D*.,* et al. (2008). The chemical genomic portrait of yeast: uncovering a phenotype for all genes. Science *320*, 362–365.

Hohmann-Marriott, M.F., and Blankenship, R.E. (2011). Evolution of photosynthesis. Annu. Rev. Plant Biol. *62*, 515–548.

Hoover, D.M., and Lubkowski, J. (2002). DNAWorks: an automated method for designing oligonucleotides for PCR-based gene synthesis. Nucleic Acids Res. *30*, e43.

Hu, J., Li, J., Qian, X., Denic, V., and Sha, B. (2009). The crystal structures of yeast Get3 suggest a mechanism for tail-anchored protein membrane insertion. PloS one *4*, e8061.

Hu, Z., Potthoff, B., Hollenberg, C.P., and Ramezani-Rad, M. (2006). Mdy2, a ubiquitin-like (UBL)-domain protein, is required for efficient mating in Saccharomyces cerevisiae. J. Cell Sci. *119*, 326–338.

Huh, W.K., Falvo, J.V., Gerke, L.C., Carroll, A.S., Howson, R.W., Weissman, J.S., and O'Shea, E.K. (2003). Global analysis of protein localization in budding yeast. Nature *425*, 686–691.

Husnjak, K., and Dikic, I. (2012). Ubiquitin-binding proteins: decoders of ubiquitin-mediated cellular functions. Annu. Rev. Biochem. *81*, 291–322.

Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M., and Sakaki, Y. (2001). A comprehensive two-hybrid analysis to explore the yeast protein interactome. Proc. Natl. Acad. Sci. USA *98*, 4569–4574.

Iwanejko, L., Smith, K.N., Loeillet, S., Nicolas, A., and Fabre, F. (1999). Disruption and functional analysis of six ORFs on chromosome XV: YOL117w, YOL115w (TRF4), YOL114c, YOL112w (MSB4), YOL111c and YOL072w. Yeast *15*, 1529–1539.

James, T.Y., Kauff, F., Schoch, C.L., Matheny, P.B., Hofstetter, V., Cox, C.J., Celio, G., Gueidan, C., Fraker, E., Miadlikowska, J.*,* et al. (2006). Reconstructing the early evolution of Fungi using a six-gene phylogeny. Nature *443*, 818–822.

Jentsch, S., and Pyrowolakis, G. (2000). Ubiquitin and its kin: how close are the family ties? Trends Cell Biol. *10*, 335–342.

Jonikas, M.C., Collins, S.R., Denic, V., Oh, E., Quan, E.M., Schmid, V., Weibezahn, J., Schwappach, B., Walter, P., Weissman, J.S.*,* et al. (2009). Comprehensive characterization of genes required for protein folding in the endoplasmic reticulum. Science *323*, 1693–1697.

Kabsch, W. (2010). Xds. Acta Crystallogr. Sect. D. Biol. Crystallogr. *66*, 125–132.

Kajander, T., Sachs, J.N., Goldman, A., and Regan, L. (2009). Electrostatic interactions of Hsp-organizing protein tetratricopeptide domains with Hsp70 and Hsp90: computational analysis and protein engineering. J. Biol. Chem. *284*, 25364–25374.

Kamath, R.S., Fraser, A.G., Dong, Y., Poulin, G., Durbin, R., Gotta, M., Kanapin, A., Le Bot, N., Moreno, S., Sohrmann, M.*,* et al. (2003). Systematic functional analysis of the Caenorhabditis elegans genome using RNAi. Nature *421*, 231–237.

Karlsson, R., Michaelsson, A., and Mattsson, L. (1991). Kinetic analysis of monoclonal antibody-antigen interactions with a new biosensor based analytical system. J. Immunol. Methods *145*, 229–240.

Kelly, A.E., Ou, H.D., Withers, R., and Dotsch, V. (2002). Low-conductivity buffers for high-sensitivity NMR measurements. J. Am. Chem. Soc. *124*, 12013–12019.

Kohl, C., Tessarz, P., von der Malsburg, K., Zahn, R., Bukau, B., and Mogk, A. (2011). Cooperative and independent activities of Sgt2 and Get5 in the targeting of tail-anchored proteins. Biol. Chem. *392*, 601–608.

Konarev, P.V., Petoukhov, M.V., Volkov, V.V., and Svergun, D.I. (2006). ATSAS 2.1, a program package for small-angle scattering data analysis. J. Appl. Crystallogr. *39*, 277–286.

Konarev, P.V., Volkov, V.V., Sokolova, A.V., Koch, M.H.J., and Svergun, D.I. (2003). PRIMUS: a Windows PC-based system for small-angle scattering data analysis. J. Appl. Crystallogr. *36*, 1277–1282.

Kordes, E., Savelyeva, L., Schwab, M., Rommelaere, J., Jauniaux, J.C., and Cziepluch, C. (1998). Isolation and characterization of human SGT and identification of homologues in Saccharomyces cerevisiae and Caenorhabditis elegans. Genomics *52*, 90–94.

Kriechbaumer, V., von Loffelholz, O., and Abell, B.M. (2012). Chaperone receptors: guiding proteins to intracellular compartments. Protoplasma *249*, 21–30.

Krissinel, E., and Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. J. Mol. Biol. *372*, 774–797.

Krogan, N.J., Cagney, G., Yu, H., Zhong, G., Guo, X., Ignatchenko, A., Li, J., Pu, S., Datta, N., Tikuisis, A.P., et al. (2006). Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. Nature *440*, 637–643.

Kuhlman, B., O'Neill, J.W., Kim, D.E., Zhang, K.Y., and Baker, D. (2001). Conversion of monomeric protein L to an obligate dimer by computational protein design. Proc. Natl. Acad. Sci. USA *98*, 10687–10691.

Kutay, U., Ahnert-Hilger, G., Hartmann, E., Wiedenmann, B., and Rapoport, T.A. (1995). Transport route for synaptobrevin via a novel pathway of insertion into the endoplasmic reticulum membrane. EMBO J. *14*, 217–223.

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

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

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., et al. (2007). Clustal W and Clustal X version 2.0. Bioinformatics *23*, 2947–2948.

Laskowski, R.A., Macarthur, M.W., Moss, D.S., and Thornton, J.M. (1993). Procheck—a Program to Check the Stereochemical Quality of Protein Structures. J. Appl. Crystallogr. *26*, 283–291.

Leipe, D.D., Wolf, Y.I., Koonin, E.V., and Aravind, L. (2002). Classification and evolution of P-loop GTPases and related ATPases. J. Mol. Biol. *317*, 41–72.

Leznicki, P., Clancy, A., Schwappach, B., and High, S. (2010). Bat3 promotes the membrane integration of tail-anchored proteins. J. Cell Sci. *123*, 2170–2178.

Leznicki, P., Warwicker, J., and High, S. (2011). A biochemical analysis of the constraints of tail-anchored protein biogenesis. Biochem. J. *436*, 719–727.

Lin, Y.F., Yang, J., and Rosen, B.P. (2007). ArsD: an As(III) metallochaperone for the ArsAB As(III)-translocating ATPase. J. Bioenerg. Biomembr. *39*, 453–458.

Linge, J.P., Williams, M.A., Spronk, C.A.E.M., Bonvin, A.M.J.J., and Nilges, M. (2003). Refinement of protein structures in explicit solvent. Proteins *50*, 496–506.

Liou, S.T., Cheng, M.Y., and Wang, C. (2007). SGT2 and MDY2 interact with molecular chaperone YDJ1 in Saccharomyces cerevisiae. Cell Stress Chaperones *12*, 59–70.

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

Lupas, A., Van Dyke, M., and Stock, J. (1991). Predicting coiled coils from protein sequences. Science *252*, 1162–1164.

Lutkenhaus, J., and Sundaramoorthy, M. (2003). MinD and role of the deviant Walker A motif, dimerization and membrane binding in oscillation. Mol. Microbiol. *48*, 295–303.

Mariappan, M., Li, X., Stefanovic, S., Sharma, A., Mateja, A., Keenan, R.J., and Hegde, R.S. (2010). A ribosome-associating factor chaperones tail-anchored membrane proteins. Nature *466*, 1120–1124.

Mariappan, M., Mateja, A., Dobosz, M., Bove, E., Hegde, R.S., and Keenan, R.J. (2011). The mechanism of membrane-associated steps in tail-anchored protein insertion. Nature *477*, 61–66.

Marley, J., Lu, M., and Bracken, C. (2001). A method for efficient isotopic labeling of recombinant proteins. J. Biomol. NMR *20*, 71–75.

Mateja, A., Szlachcic, A., Downing, M.E., Dobosz, M., Mariappan, M., Hegde, R.S., and Keenan, R.J. (2009). The structural basis of tail-anchored membrane protein recognition by Get3. Nature *461*, 361–366.

McClellan, A.J., Xia, Y., Deutschbauer, A.M., Davis, R.W., Gerstein, M., and Frydman, J. (2007). Diverse cellular functions of the Hsp90 molecular chaperone uncovered using systems approaches. Cell *131*, 121–135.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. J. Appl. Crystallogr. *40*, 658–674.

Meacham, G.C., Patterson, C., Zhang, W., Younger, J.M., and Cyr, D.M. (2001). The Hsc70 co-chaperone CHIP targets immature CFTR for proteasomal degradation. Nat. Cell Biol. *3*, 100–105.

Mertens, H.D., and Svergun, D.I. (2010). Structural characterization of proteins and complexes using small-angle X-ray solution scattering. J. Struct. Biol. *172*, 128–141.

Metz, J., Wachter, A., Schmidt, B., Bujnicki, J.M., and Schwappach, B. (2006). The yeast Arr4p ATPase binds the chloride transporter Gef1p when copper is available in the cytosol. J. Biol. Chem. *281*, 410–417.

Meyer, N.H., Tripsianes, K., Vincendeau, M., Madl, T., Kateb, F., Brack-Werner, R., and Sattler, M. (2010). Structural basis for homodimerization of the Src-associated during mitosis, 68-kDa protein (Sam68) Qua1 domain. J. Biol. Chem. *285*, 28893–28901.

Minami, R., Hayakawa, A., Kagawa, H., Yanagi, Y., Yokosawa, H., and Kawahara, H. (2010). BAG-6 is essential for selective elimination of defective proteasomal substrates. J. Cell Biol. *190*, 637–650.

Newlon, M.G., Roy, M., Morikis, D., Hausken, Z.E., Coghlan, V., Scott, J.D., and Jennings, P.A. (1999). The molecular basis for protein kinase A anchoring revealed by solution NMR. Nat. Struct. Biol. *6*, 222–227.

Nilges, M., Bernard, A., Bardiaux, B., Malliavin, T., Habeck, M., and Rieping, W. (2008). Accurate NMR structures through minimization of an extended hybrid energy. Structure *16*, 1305–1312.

Odunuga, O.O., Hornby, J.A., Bies, C., Zimmermann, R., Pugh, D.J., and Blatch, G.L. (2003). Tetratricopeptide repeat motif-mediated Hsc70-mSTI1 interaction. Molecular characterization of the critical contacts for successful binding and specificity. J. Biol. Chem. *278*, 6896–6904.

Perrin, C.L., and Dwyer, T.J. (1990). Application of 2-Dimensional Nmr to Kinetics of Chemical-Exchange. Chem. Rev. *90*, 935–967.

Petoukhov, M.V., and Svergun, D.I. (2005). Global rigid body modeling of macromolecular complexes against small-angle scattering data. Biophys. J. *89*, 1237–1250.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. J. Comput. Chem. *25*, 1605–1612.

Prag, G., Misra, S., Jones, E.A., Ghirlando, R., Davies, B.A., Horazdovsky, B.F., and Hurley, J.H. (2003). Mechanism of ubiquitin recognition by the CUE domain of Vps9p. Cell *113*, 609–620.

Rabu, C., Schmid, V., Schwappach, B., and High, S. (2009). Biogenesis of tail-anchored proteins: the beginning for the end? J. Cell Sci. *122*, 3605–3612.

Rabu, C., Wipf, P., Brodsky, J.L., and High, S. (2008). A precursor-specific role for Hsp40/Hsc70 during tail-anchored protein integration at the endoplasmic reticulum. J. Biol. Chem. *283*, 27504–27513.

Rieping, W., Habeck, M., Bardiaux, B., Bernard, A., Malliavin, T.E., and Nilges, M. (2007). ARIA2: automated NOE assignment and data integration in NMR structure calculation. Bioinformatics *23*, 381–382.

Saeki, Y., Saitoh, A., Toh-e, A., and Yokosawa, H. (2002). Ubiquitin-like proteins and Rpn10 play cooperative roles in ubiquitin-dependent proteolysis. Biochem. Biophys. Res. Commun. *293*, 986–992.

Sanner, M.F., Olson, A.J., and Spehner, J.C. (1996). Reduced surface: an efficient way to compute molecular surfaces. Biopolymers *38*, 305–320.

Santelli, E., Leone, M., Li, C., Fukushima, T., Preece, N.E., Olson, A.J., Ely, K.R., Reed, J.C., Pellecchia, M., Liddington, R.C., et al. (2005). Structural analysis of Siah1-Siah-interacting protein interactions and insights into the assembly of an E3 ligase multiprotein complex. J. Biol. Chem. *280*, 34278–34287.

Sattler, M., Schleucher, J., and Griesinger, C. (1999). Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. Prog. Nucl. Magn. Reson. Spectrosc. *34*, 93–158.

Scheufler, C., Brinker, A., Bourenkov, G., Pegoraro, S., Moroder, L., Bartunik, H., Hartl, F.U., and Moarefi, I. (2000). Structure of TPR domain-peptide complexes: critical elements in the assembly of the Hsp70-Hsp90 multichaperone machine. Cell *101*, 199–210.

Schindelin, H., Kisker, C., Schlessman, J.L., Howard, J.B., and Rees, D.C. (1997). Structure of ADP x AIF4(-)-stabilized nitrogenase complex and its implications for signal transduction. Nature *387*, 370–376.

Schlegel, T., Mirus, O., von Haeseler, A., and Schleiff, E. (2007). The tetratricopeptide repeats of receptors involved in protein translocation across membranes. Mol. Biol. Evol. *24*, 2763–2774.

Schuldiner, M., Collins, S.R., Thompson, N.J., Denic, V., Bhamidipati, A., Punna, T., Ihmels, J., Andrews, B., Boone, C., Greenblatt, J.F., et al. (2005). Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile. Cell *123*, 507–519.

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

Shan, S.O., and Walter, P. (2005). Co-translational protein targeting by the signal recognition particle. FEBS Lett. *579*, 921–926.

Shaner, L., Wegele, H., Buchner, J., and Morano, K.A. (2005). The yeast Hsp110 Sse1 functionally interacts with the Hsp70 chaperones Ssa and Ssb. J. Biol. Chem. *280*, 41262–41269.

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

Sharma, D., and Rajarathnam, K. (2000). C-13 NMR chemical shifts can predict disulfide bond formation. J. Biomol. NMR *18*, 165–171.

Sheinerman, F.B., Norel, R., and Honig, B. (2000). Electrostatic aspects of protein-protein interactions. Curr. Opin. Struct. Biol. *10*, 153–159.

Sheldrick, G.M. (2008). A short history of SHELX. Acta Crystallogr. Sect. A: Found. Crystallogr. *64*, 112122.

Shen, Y., and Bax, A. (2010). Prediction of Xaa-Pro peptide bond conformation from sequence and chemical shifts. J. Biomol. NMR *46*, 199–204.

Shen, Y., Delaglio, F., Cornilescu, G., and Bax, A. (2009). TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. J. Biomol. NMR *44*, 213–223.

Sherrill, J., Mariappan, M., Dominik, P., Hegde, R.S., and Keenan, R.J. (2011). A conserved archaeal pathway for tail-anchored membrane protein insertion. Traffic *12*, 1119–1123.

Shih, S.C., Prag, G., Francis, S.A., Sutanto, M.A., Hurley, J.H., and Hicke, L. (2003). A ubiquitin-binding motif required for intramolecular monoubiquitylation, the CUE domain. EMBO J. *22*, 1273–1281.

Simmer, F., Moorman, C., van der Linden, A.M., Kuijk, E., van den Berghe, P.V., Kamath, R.S., Fraser, A.G., Ahringer, J., and Plasterk, R.H. (2003). Genome-wide RNAi of C. elegans using the hypersensitive rrf-3 strain reveals novel gene functions. PLoS Biol. *1*, E12.

Simon, S.M., and Blobel, G. (1991). A protein-conducting channel in the endoplasmic reticulum. Cell *65*, 371–380.

Simpson, P.J., Schwappach, B., Dohlman, H.G., and Isaacson, R.L. (2010). Structures of Get3, Get4, and Get5 provide new models for TA membrane protein targeting. Structure *18*, 897–902.

Skubak, P., Murshudov, G.N., and Pannu, N.S. (2004). Direct incorporation of experimental phase information in model refinement. Acta Crystallogr. Sect. D. Biol. Crystallogr. *60*, 2196–2201.

Smolsky, I.L., Liu, P., Niebuhr, M., Ito, K., Weiss, T.M., and Tsuruta, H. (2007). Biological small-angle X-ray scattering facility at the Stanford synchrotron radiation laboratory. J. Appl. Crystallogr. *40*, S453–S458.

Stefanovic, S., and Hegde, R.S. (2007). Identification of a targeting factor for posttranslational membrane protein insertion into the ER. Cell *128*, 1147–1159.

Stefer, S., Reitz, S., Wang, F., Wild, K., Pang, Y.Y., Schwarz, D., Bomke, J., Hein, C., Lohr, F., Bernhard, F., et al. (2011). Structural basis for tail-anchored membrane protein biogenesis by the Get3-receptor complex. Science *333*, 758–762.

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

Suloway, C.J., Chartron, J.W., Zaslaver, M., and Clemons, W.M., Jr. (2009). Model for eukaryotic tail-anchored protein binding based on the structure of Get3. Proc. Natl. Acad. Sci. USA *106*, 14849–14854.

Suloway, C.J., Rome, M.E., and Clemons, W.M., Jr. (2012). Tail-anchor targeting by a Get3 tetramer: the structure of an archaeal homologue. EMBO J. *31*, 707–719.

Svergun, D.I. (1992). Determination of the regularization parameter in indirect-transform methods using perceptual criteria. J. Appl. Crystallogr. *25*, 495–503.

Svergun, D.I., Petoukhov, M.V., and Koch, M.H. (2001). Determination of domain structure of proteins from X-ray solution scattering. Biophys. J. *80*, 2946–2953.

Terwilliger, T.C. (2004). SOLVE and RESOLVE: automated structure solution, density modification and model building. J. Synchrotron Radiat. *11*, 49–52.

Terwilliger, T.C. (2000). Maximum-likelihood density modification. Acta Crystallogr. Sect. D. Biol. Crystallogr. *56*, 965–972.

Tobaben, S., Thakur, P., Fernandez-Chacon, R., Sudhof, T.C., Rettig, J., and Stahl, B. (2001). A trimeric protein complex functions as a synaptic chaperone machine. Neuron *31*, 987–999.

Tobaben, S., Varoqueaux, F., Brose, N., Stahl, B., and Meyer, G. (2003). A brain-specific isoform of small glutamine-rich tetratricopeptide repeat-containing protein binds to Hsc70 and the cysteine string protein. J. Biol. Chem. *278*, 38376–38383.

Toniolo, D., Persico, M., and Alcalay, M. (1988). A "housekeeping" gene on the X chromosome encodes a protein similar to ubiquitin. Proc. Natl. Acad. Sci. USA *85*, 851–855.

Uetz, P., Giot, L., Cagney, G., Mansfield, T.A., Judson, R.S., Knight, J.R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P., et al. (2000). A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. Nature *403*, 623–627.

Valafar, H., and Prestegard, J.H. (2004). REDCAT: a residual dipolar coupling analysis tool. J. Magn. Reson. *167*, 228–241.

Van den Berg, B., Clemons, W.M., Jr., Collinson, I., Modis, Y., Hartmann, E., Harrison, S.C., and Rapoport, T.A. (2004). X-ray structure of a protein-conducting channel. Nature *427*, 36–44.

Van Duyne, G.D., Standaert, R.F., Karplus, P.A., Schreiber, S.L., and Clardy, J. (1993). Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. J. Mol. Biol. *229*, 105–124.

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

Volkov, V.V., and Svergun, D.I. (2003). Uniqueness of *ab initio* shape determination in small-angle scattering. J. Appl. Crystallogr. *36*, 860–864.

Vranken, W.F., Boucher, W., Stevens, T.J., Fogh, R.H., Pajon, A., Llinas, M., Ulrich, E.L., Markley, J.L., Ionides, J., and Laue, E.D. (2005). The CCPN data model for NMR spectroscopy: development of a software pipeline. Proteins *59*, 687–696.

Vriend, G., and Sander, C. (1993). Quality-control of protein models—directional atomic contact analysis. J. Appl. Crystallogr. *26*, 47–60.

Walter, P., and Blobel, G. (1981). Translocation of proteins across the endoplasmic reticulum. II. Signal recognition protein (SRP) mediates the selective binding to microsomal membranes of in-vitro assembled polysomes synthesizing secretory protein. J. Cell Biol. *91*, 551–556.

Walter, P., and Johnson, A.E. (1994). Signal sequence recognition and protein targeting to the endoplasmic reticulum membrane. Annu. Rev. Cell Biol. *10*, 87–119.

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

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

Wang, Q., Liu, Y., Soetandyo, N., Baek, K., Hegde, R., and Ye, Y. (2011). A ubiquitin ligase-associated chaperone holdase maintains polypeptides in soluble states for proteasome degradation. Mol. Cell *42*, 758–770.

Winget, J.M., and Mayor, T. (2010). The diversity of ubiquitin recognition: hot spots and varied specificity. Mol. Cell *38*, 627–635.

Winn, M.D., Murshudov, G.N., and Papiz, M.Z. (2003). Macromolecular TLS refinement in REFMAC at moderate resolutions. Methods Enzymol. *374*, 300–321.

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

Winzeler, E.A., Shoemaker, D.D., Astromoff, A., Liang, H., Anderson, K., Andre, B., Bangham, R., Benito, R., Boeke, J.D., Bussey, H., et al. (1999). Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis. Science *285*, 901–906.

Worrall, L.J., Wear, M.A., Page, A.P., and Walkinshaw, M.D. (2008). Cloning, purification and characterization of the Caenorhabditis elegans small glutamine-rich tetratricopeptide repeat-containing protein. Biochim. Biophys. Acta *1784*, 496–503.

Wu, S.J., Liu, F.H., Hu, S.M., and Wang, C. (2001). Different combinations of the heat-shock cognate protein 70 (hsc70) C-terminal functional groups are utilized to interact with distinct tetratricopeptide repeat-containing proteins. Biochem. J. *359*, 419–426.

Xu, Y., Cai, M., Liu, Y., Lee, J.-g., Huang, L., Clore, M., and Ye, Y. (2012). A UBL dependent protein interaction network regulates ER-associated protein degradation.

Yamagata, A., Mimura, H., Sato, Y., Yamashita, M., Yoshikawa, A., and Fukai, S. (2010). Structural insight into the membrane insertion of tail-anchored proteins by Get3. Genes Cells *15*, 29–41.

Yang, F., Skaletsky, H., and Wang, P.J. (2007). Ubl4b, an X-derived retrogene, is specifically expressed in post-meiotic germ cells in mammals. Gene Expr. Patterns *7*, 131–136.

Yu, H., Braun, P., Yildirim, M.A., Lemmens, I., Venkatesan, K., Sahalie, J., Hirozane-Kishikawa, T., Gebreab, F., Li, N., Simonis, N., et al. (2008). High-quality binary protein interaction map of the yeast interactome network. Science *322*, 104–110.

Zhang, M., Windheim, M., Roe, S.M., Peggie, M., Cohen, P., Prodromou, C., and Pearl, L.H. (2005). Chaperoned ubiquitylation—crystal structures of the CHIP U box E3 ubiquitin ligase and a CHIP-Ubc13-Uev1a complex. Mol. Cell *20*, 525–538.

Zhou, T., Radaev, S., Rosen, B.P., and Gatti, D.L. (2000). Structure of the ArsA ATPase: the catalytic subunit of a heavy metal resistance pump. EMBO J. *19*, 4838–4845.

Zweckstetter, M., and Bax, A. (2000). Prediction of sterically induced alignment in a dilute liquid crystalline phase: Aid to protein structure determination by NMR. J. Am. Chem. Soc. *122*, 3791–3792.