12-2016

Interactions in the cpSRP Dependent Targeting of Light Harvesting Chlorophyll Binding Protein to the Thylakoid Membrane

Rory Henderson
University of Arkansas, Fayetteville

Interactions in the cpSRP Dependent Targeting of Light Harvesting Chlorophyll Binding Protein to the Thylakoid Membrane

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Chemistry

by

Rory Henderson
University of Arkansas
Bachelor of Science in Chemistry, 2011

December 2016
University of Arkansas

This dissertation is approved for recommendation to the Graduate Council.

_____
Dr. Suresh Thallapuranam
Dissertation Director

_____
Dr. Paul Adams
Committee Member

_____
Dr. Ralph Henry
Committee Member

_____
Dr. Colin Heyes
Committee Member

_____
Dr. Frank Millett
Committee Member

_____
Dr. Julie Stenken
Committee Member

Abstract

Targeting of proteins is a critical component of cellular function. A universally conserved targeting system of the cytosol utilizes a signal recognition particle (SRP) to target many proteins contranslationally to the endoplasmic reticulum in eukaryotes or the inner membrane in prokaryotes. A homologous SRP system exists in the chloroplast that delivers light harvesting chlorophyll binding proteins (LHCP) to they thylakoid membrane. The chloroplast SRP (cpSRP) is a heterodimer composed of a novel 43 kDa subunit and a 54 kDa subunit homologous to a component of the SRP system, SRP54. Many details regarding the interactions between the proteins of the cpSRP system have been determined. However, the three-dimensional arrangement of the cpSRP43 and cpSRP54 domains as well as their influence on one another has not been determined. The results of this study demonstrate both cpSRP43 and cpSRP54 are characterized by a significant amount of structural flexibility. Specifically, the domains of cpSRP43 and cpSRP54 are flexibly linked allowing for rapid conformational sampling of the domains. This flexibility allows cpSRP43 to sense the presence of cpSRP54 and subsequently alter its affinity for LHCP. Conversely, cpSRP54 domain flexibility allows it to scan cpSRP43 for the third transmembrane segment of LHCP in a manner surprisingly similar to SRP54 scanning for signal sequences at the ribosome. Together, the results of this structural investigation of the free and bound proteins has lead to the speculation of a cpSRP-LHCP transit complex structure capable of rationalizing the steps leading to the integration of LHCP.

Table of Contents

Portions of chapters II and III have been previously published:

II. Gao, F. *et al*. Regulation of Structural Dynamics within a Signal Recognition Particle Promotes Binding of Protein Targeting Substrates. *Journal of Biological Chemistry*, doi:10.1074/jbc.M114.624346 (2015).

III. Henderson, Rory C. *et al*. Domain Organization in the 54-kDa Subunit of the Chloroplast Signal Recognition Particle. *Biophysical Journal* **111**, 1151-1162, doi:10.1016/j.bpj.2016.08.004.

I. Introduction

Co-translational Cellular Protein Targeting

Targeting of proteins to their proper cellular location is a critical aspect of proper cellular function. Many targeting systems exist to ensure proper localization of proteins in the cell. Of the 30% of proteins destined for the endoplasmic reticulum (ER), most are targeted co-translationally by the signal recognition particle (SRP) pathway. This pathway delivers proteins to the endoplasmic reticulum in eukaryotes or the inner membrane in prokaryotes.[1] This delivery utilizes an N-terminal sequence recognized by a signal recognition particle (SRP). The SRP binds to the signal sequence during synthesis, halting the process while the m-RNA-ribosome-peptide-SRP complex reaches the appropriate SRP receptor on the membrane of the endoplasmic reticulum (ER).[2] Binding and hydrolysis of GTP results in the release of the SRP from the ribosome/peptide chain complex. Thereafter, protein synthesis continues through a translocation pore, co-translationally placing the protein into the ER lumen.[1,3] Once in the lumen the N-terminal signal sequence is cleaved. Additional sequences on the protein then allow for its direction to remain in the ER or move on to the Golgi apparatus, lysosome or exterior of the cell (Figure 1A. a.).[2]

The signal recognition particle is composed of a 54kDa GTPase protein, SRP54 in eukaryotes and archaea or Ffh in prokaryotes, and an SRP RNA. SRP54/Ffh is contains a four helix bundle N-terminal N-domain, a GTPase G-domain, and a methionine rich M-domain. The M-domain contains a hydrophobic pocket responsible for binding hydrophobic signal sequences as they emerge from the ribosome. The M-domain is linked to the G-domain via a flexible linking segment that allows the M-domain to change position relative to the N and G-domains.[1]

In fact, the M-domain positioning undergoes significant changes in position throughout the signaling process.[1] The SRP RNA is a universally conserved element in the cytosolic SRP pathway, acting as a scaffold for changes in the M-domain positioning and as a tether to accelerate NG-domain interaction with its receptor protein.[1] The SRP receptor (SR) protein in prokaryotes, FtsY, is homologous to the NG-domain of SRP54, containing an N-domain and GTPase G-domain. It contains an additional membrane binding segment termed the A-domain at its N-terminus.[1,4,5] The eukaryotic receptor is a heterodimer composed of an SRP54 binding α-subunit homologous to FtsY and a membrane associating GTPase β-subunit. The β-subunit GTPase element shares significant homology with the Arf family of GTPases. The GTPase active sites of both the SRP54/Ffh and the FtsY/α-subunit face one another in the bound configuration. However, association kinetics between the domains in the absence of SRP RNA a quite slow, as significant domain reorganization is required for efficient binding. The SRP RNA has therefore been proposed to act as a transient tether for SR to ensure these transition may occur. Upon binding of the receptor the RNC-SRP-SR complex is delivered to a translocation pore for delivery across or into the membrane.[1] This translocation pore is a transmembrane protein termed Sec61p in eukaryotes or Sec-YEG in prokaryotes. Upon delivery of the RNC to the translocation pore GTP hydrolysis signals the handover of cargo, destabilizing the SRP-SR interaction and resulting in the recycling of the SRP. The protein is then synthesized through the pore or into the membrane via a lateral gate in the translocon (Figure 1A & B).[1]

Post-translational Cellular Protein Targeting

Differing from the typical SRP-mediated targeting of cellular proteins, proteins destined for the nucleus, mitochondrion and chloroplasts follow an SRP-independent, post-translational path. Protein-specific chaperones direct the targeting process, which is generally more

complicated than the SRP mediated targeting due to the many different locations for the proteins in these organelles. These chaperones bind the completely synthesized protein in a conformation conducive to import and translocation across the organelle membrane. As in the ER lumen, the N-terminal sequence is cleaved, after which sequences on the native protein may be used for further targeting in the organelle. This is the case for thylakoid proteins, for instance, which must undergo further chaperone directed targeting to reach their intended location within the chloroplast.[2] Targeting to other organelles follows a similar scheme, utilizing signal sequence/chaperone specific receptors to capture proteins on the membrane surface.[6]

Two SRP-independent paths exist for targeting of bacterial proteins to the inner membrane as well. SRP-independent transport of cytosolic bacterial proteins occurs via the SecB-dependent and twin-arginine translocation (Tat), ΔpH pathways.[7,8] The SecB path uses a tetramer of acidic, 17kDa subunits. SecB binds the mature domain of the pre-protein maintaining it in a loosely folded state. SecB then transfers the pre-protein to Sec for for transport across the membrane.[8,9] The Sec-independent Tat transport requires a pH-gradient across the membrane. The Tat pathway depends on a twinned arginine signal sequence at the N-terminus containing an SRR-X-FLK consensus motif and is homologous to a chloroplastic pH-gradient dependent targeting system.[7,8]

Nuclear Protein Localization

Transport across the nuclear envelope involves a complex of nucleoporin proteins termed nuclear pore complexes (NPCs). While some small proteins passively diffuse through the NPC, larger proteins require chaperones called karyopherins (importins and exportins). Transport is controlled via binding of karyopherins to a small GTPase termed Ran. Ran's activity is controlled by exchange between GTP and GDP bound states. GTPase-activating proteins (GAPs)

activate the GTPase activity of Ran while guanine nucleotide-exchange factor proteins (GEFs) facilitated exchange of GDP for GTP. After traversing the pore for nuclear import, the protein-karyopherin complex is dissociated via Ran-GTP binding. For protein export, Ran-GTP bound karyopherins bind the export cargo and traverse the NPC. Once in the cytosol GAPs facilitate GTP hydrolysis and thereby the release of the karyopherin and target protein. The protein NTF2 may then facilitate Ran transition back into the nucleus where GEFs exchange Ran GDP for GTP to allow for another cycle of transport. Karyopherins typically recognize cargo based upon unstructured loops or tails in the targeted protein. However, some nuclear proteins are targeted based upon fold or by piggyback binding partners containing nuclear localization signals. For instance, proteins such as NTF2 and mRNA export protein PAP:p15 bypass the need for karyopherins for transport across the nuclear pore.[6]

Mitochondrial Protein Localization

The majority of mitochondrial proteins are encoded in the nucleus. A small number of tRNAs, rRNAs and respiratory complexe proteins are encoded in the mitochondrial genome. Transport of nuclear encoded proteins synthesized in the cytosol is controlled by the translocase of the outer membrane (TOM). Mitochondrial matrix proteins are directed via an N-terminal amphipathic sequence. The polar portion of the sequence carries a net positive charge recognized by the translocase at in inner membrane (TIM) during transport across TOM to the inner membrane space. This ATP driven process continues until the protein is transported into the matrix where the signal sequence is cleaved and degraded. Transport to the intermembrane space utilizes TOM and TIM as well. However, upon recognition of a specific hydrophobic sequence, TIM stops transport, releasing the protein to the intermembrane space where the signal is cleaved. An additional mechanism for intermembrane protein transport involves TOM and inner

membrane-anchored mitochondrial intermembrane space assembly machinery. This mechanism is utilized by certain cysteine rich proteins which utilize disulfide bond formation for transport. Inner membrane proteins utilize a path similar to the TOM/TIM intermembrane space proteins or, after transport in the cytosol by chaperones such as Hsp70 to TOC, are recruited by special chaperones in the intermembrane space to facilitate TIM insertion. Finally, outer membrane proteins are imported as in inner membrane protein involving TOM or by TOM-independent means.[6]

Chloroplast Protein Localization

Plastids are a ubiquitous, heterogeneous family of organelles in plants and algea that include chloroplast and others such as starch storing amyloplasts and carotenoid collecting chromoplasts.[10] Chloroplasts are a type of plastid containing 6 distinct compartments (Figure 3). In addition to photosynthesis, the chloroplast performs functions such as fatty acid and amino acid biosynthesis.[11] The chloroplasts of eukaryotes contain ~3,000 proteins of which 350 are thylakoid membrane proteins. Most chloroplast proteins are nuclear encoded and therefore follow a post translational targeting scheme.[12] Targeting for chloroplast proteins involves the general steps mentioned above: direction to the chloroplast, movement across the two envelope membranes and localization within the compartment. Typically, an N-terminal transit peptide directs proteins through both membrane structures within the chloroplast and is subsequently cleaved. In some cases, this transit peptide is bipartite, with one transit sequence responsible for transport to the stroma and the other to direct the protein further once in the stroma. Each is cleaved as the protein reaches its destination. In addition to, or in place of, a transit peptide, some proteins contain additional targeting sites within their structures (Figure 2).[11]

While the signal peptide provides information regarding where the proteins should go, the primary method for helping them reach their destination involves molecular chaperones such as Hsp70 and various import machinery.[10,11] The proteins first reach the outer envelope membrane of the chloroplast where they become reversibly associated with the membrane through protein-protein interactions and signal peptide interaction with the membrane. Translocation thereafter requires the input of ATP.[11] The proteins TOC and TIC (Translocon at the Outer and Inner envelop membrane of Chloroplasts respectively) are responsible for translocation of most cytosol synthesized proteins bound for the chloroplast. After the first membrane interaction step the TOC system begins the translocation of the protein followed by the simultaneous transfer of the protein through the TIC system into the stroma where the transit peptide is cleaved, leaving the mature protein in the stroma.[12,13] Thereafter, proteins destined for the thylakoids may use one of four different localization systems: the twin-arginine translocation (Tat) ΔpH pathway, a spontaneous insertion mechanism, a chloroplast Sec-dependent (cpSec) path and the cpSRP path.[3,11,12,14,15]

The cpSec-dependent pathway is a bacterial Sec-pathway homologue responsible for the delivery of certain photosynthetic proteins, acting as a translocon for the thylakoid. The cpSec pathway is composed of bacterial homologues of the ATP-dependent SecA translocation motor and SecYE translocation pore. Interestingly, typical bacterial Sec proteins including SecB, SecG, SecD, SecF and YajC are not found in the Arabidopsis genome suggesting the cpSec system operates on a minimum of translocation machinery.[10,11] The cpSec pathway delivers both luminal proteins, such as plastocyanin and the 33kDa oxygen-evolving protein, and thylakoid membrane proteins such as D1 (Figure 2A).[10]

The chloroplast Tat ΔpH pathway is homologous to the bacterial Tat pathway, utilizing a twin arginine motif in its signal sequence to facilitate targeting by Tat proteins.[16] The chloroplast Tat targeting system is composed of Hcf106, Tha4, and cpTatC, all orthologs of the *E. coli*. Tat system. cpTatC is an integral membrane protein that binds Hcf 106 to form a ~700kDa complex. The complex binds precursor proteins with cpTatC recognizing the twin-arginine motif. [17] Upon recognition of cargo Tha4 oligomerizes with the complex resulting in transport of the protein across the membrane. The primary driver of translocation is the pH gradient across the thylakoid membrane.[18] Interestingly, it was shown that the chloroplast Tat system uses the transmembrane electric potential as well as the proton gradient to facilitate protein transfer.[19] Two models for translocation have been proposed, one involving the formation of a translocation pore and another involving temporary weakening of the bilayer to allow the folded protein to transition through the membrane (Figure 2B).[18]

Several proteins destined for the thylakoid insert spontaneously without a requirement for any of the typical targeting and translocation machinery present in the chloroplast. These include photosystem proteins PsbS, PsbW, PsbX, PsbY, and PSI-G.[20,21] The significant hydrophobicity of the transmembrane helices is proposed to play a central role in the spontaneous insertion mechanism. However, loop regions have been shown to play an important role as well (Figure 2D).[21]

Light Harvesting Chlorophyll Binding Protein

The photosystems to which light harvesting chlorophyll binding proteins (LHCPs) belong are located in the thylakoid membrane. LHCP, a nuclear encoded protein, follows the cpSRP path for its integration. These three transmembrane helix, light harvesting complexes act as antennae, carrying chlorophylls *a* and *b* as well as carotenoid molecules all with specific

absorption bands in the visible spectrum (Figure 5D). The molecules absorb light and facilitate chemical energy production in plants.[22,23] LHCPs are the most abundant membrane proteins on earth and comprise roughly one third of the thylakoid membrane proteins, binding half of all the chlorophyll molecules.[15,22] It is synthesized in the cytosol where a transit peptide directs its movement across both membranes of the chloroplast through the TOC and TIC systems. The transit peptide is subsequently cleaved in the stroma where LHCP becomes involved with another transit complex, cpSRP.[24,25]

LHCP is targeted to the thylakoid membrane in a GTP dependent manner by the cpSRP system located in the stroma and consisting of a 54kDa protein homologue of the cytosolic SRP system and a unique 43kDa protein, each lacking the RNA present in the cytosolic system (figure 4).[15,26] LHCP has a large hydrophobic surface leaving it prone to aggregation, hence the need to form a soluble complex in the stroma.[27,28] It was shown that cpSRP54, cpFtsY and C-term Alb3, all interaction partners in the cpSRP system, could not prevent LHCP aggregation while cpSRP43 could. Further, LHCP copurifies with cpSRP43 in a manner that is not affected by salt concentrations. A K_d value of 97nM was obtained for this interaction demonstrating a high affinity of LHCP for cpSRP43.[27] This evidence indicates cpSRP43 is critical for the interaction with LHCP in this pathway and that the interaction is predominately hydrophobic in nature.[28]

The Chloroplast Signal Recognition Particle Pathway

Using knock-out mutants known as *ffc* and *chaos*, indicating the removal of coding regions for cpSRP54 and cpSRP43 respectively, early investigations demonstrated that plants grown without either of these two proteins produce pale green leaves.[6, 7] Removal of cpSRP54 through imunoprecipitation *in vitro* results in a reduction in LHCP integration. However, it was shown LHCP insertion did not increase upon reintroduction of cpSRP54 suggesting another

protein is involved in the process. This protein was identified as cpSRP43. Plant mutants lacking the genes coding for cpSRP43, CAO mutant, were deficient in chlorophyll b, only present in LHCP proteins, and certain carotenoids demonstrating the importance of cpSRP43 in the targeting of LHCPs to the thylakoid membrane. This deletion reduces oxygen production, and therefore light harvesting, without effecting electron transport in the photosystems.[15] It was shown that seven LHCP's were deficient in these plants. Removal of the gene coding for the protein ClpC in addition to cpSRP54 resulted in plant death, suggesting ClpC compensates for the lack of cpSRP54. Considering this and that about half of the stromal cpSRP54 is bound to ribosomes, cpSRP54 is likely involved in cotranslational targeting of many chloroplast proteins.[14,15] In fact, the *ffc* mutant results in several photosystem reaction center proteins failing to reach their intended location.[29] Important for the LHCP targeting path, cpSRP43 has been shown to operate absent cpSRP54 and cpFtsY. This was shown using cpSRP43ΔCD2 that is unable to bind cpSRP54.[12] The deletion of cpSRP54 resulted in similar deficiencies as well as a reduction in several chloroplast encoded thylakoid proteins.[14,30]

cpSRP Structural Arrangement and Homology

The domains of cpSRP43 include three chromodomains and four ankyrin repeats. Chromodomain 1 (CD1) at the N-terminus is fused to ankyrin 1 (Ank1) with chromodomain 2 (CD2) following from ankyrin 4 (Ank4) and followed thereafter by chromodomain 3 (CD3) (Figure 5A). Pairs of chromodomains are usually flexibly linked, however, this does not appear to be the case for cpSRP43.[31] The second protein of cpSRP, cpSRP54, is homologous to SRP54 in the cytosol, containing a four helix bundle N-domain, five conserved GTPase elements that make up a G-domain and a C-terminal M-domain (Figure 5B). The M-domain of SRP54, named for its methionine rich composition, is responsible for binding RNA as well as containing a

hydrophobic interaction site.[12,32,33] The M-domain of cpSRP54 may bind the transmembrane domains of LHCP, although some investigations are conflicting.[26,34] Structures are available in the protein data bank for eukaryotic SRP54 and prokaryotic SRP54. The eukaryotic G-domain of SRP54 is 62% similar to cpSRP54 while it is 73% similar to the prokaryotic SRP54. The M-domain's homology for the eukaryotic and prokaryotic proteins to cpSRP54 is 52% and 51%, respectively.[33]

An early study utilizing gel filtration and crosslinking demonstrated the possibility that cpSRP43 exists as a dimer. A subsequent investigation identified the location of the dimerization to be at the third and fourth ankyrin repeats.[31] The entire cpSRP complex was determined to be a heterotrimer consisting of two cpSRP43 molecules and one cpSRP54 molecule. However, conflicting reports suggest the cpSRP43 dimer is not a part of the functioning cpSRP system. It was also demonstrated that cpFtsY exists as a monomer.[35]

cpSRP43 Interactions

The 43kDa protein, cpSRP43, binds an 18 reside sequence on LHCP termed L18 (Figure 5D).[3,34] L-18 occurs in LHCP between the second and third transmembrane domains and binds cpSRP43 between ankyrin two to four. Furthermore, a deletion mutant of cpSRP43 containing only the N-terminal region (CD1-Ank1) still interacts with L18, albeit weaker than with full length cpSRP43.[31] However, other data has shown that deletion of any of the ankyrin repeats results in a loss of binding.[27] LHCP is bound in an extended form except for a DPLG type I turn. Without this turn it was shown in one study that L18 would not bind cpSRP43 while another investigation showed it was not necessary.[34] Hydrophobic interactions appear to play an important role as well since mutation of key hydrophobic residues result in a reduction in binding strength. Mutation of the tyrosine on the "hook" region of ankyrin 4 appeared to

eliminate binding altogether. However, another study could not confirm this, demonstrating that the L18 interaction was stabilized by complex formation with cpSRP54.[28,34] It was also demonstrated that positive charge around this site is important for binding. The L-18 region bound to cpSRP43 is very similar in appearance to unbound LHCP in the membrane, forming a complex around the tyrosine hook that is similar to the way LHCP interacts with a carotenoid molecule.[36] Some studies have also demonstrated the need for transmembrane segment 3 (TM3) for efficient interaction between LHCP and cpSRP43.[3,34] In fact, full length LHCP was shown to bind cpSRP43 with a 60-fold great affinity that the L18 peptide.[28]

cpSRP54 Interactions

Binding studies investigating the interaction between cpSRP54 and cpSRP43 have demonstrated that the C-terminal chromodomains are necessary for the interaction. Using deletion mutants it appeared that residues 259-282, at the N-terminus of chromodomain 2, were necessary for cpSRP54 interaction with cpSRP43 in addition to amino acids 259-339 of the third chromodomain.[31] It was demonstrated that CD2 alone could interact with cpSRP54, although weaker than full length cpSRP43.[32,37] The M-region of cpSRP54 appears to be responsible for this interaction.[31] In another study it was shown that binding would occur in the absence of CD3 and that CD2 interacts with cpSRP54 using an RRKR sequence on CD2.[36] An x-ray crystal structure of cpSRP43ΔCD3 complexed with the C-terminal tail of the M-region of cpSRP54 (M-peptide) showed the interaction takes place between Ank4 and CD2 on cpSRP43. A twinned cage-like interaction between two of M-peptide's arginines with both aromatic and one negatively charged amino acid in cpSRP43 mediates the interaction, forming a beta-completed structure. It was shown through mutation, however, that the negative charge was not the primary contributor to this interaction. Small angle x-ray scattering data demonstrated that CD2 is

flexible and is brought closer to Ank4 in the presence of the M-peptide (56). Another study demonstrated the importance of the negatively charged α-helix of CD2 for the interaction.[32] In fact, CD2 appears to undergo significant structural changes upon binding the cpSRP54 C-terminal tail region.[37] Other investigations have demonstrated the ability of cpSRP54 to interact with cpSRP43 with the C-terminal 26 residues containing the M-peptide removed. Chromodomain 1 of cpSRP43 has been shown to regulate the GTPase activity of cpSRP54. In vitro integration studies investigating LHCP integration into whole thylakoids demonstrated a marked reduction for integration when a cpSRP43ΔCD1 mutant was used in place of cpSRP43. However, this deletion did not affect transit complex formation suggesting CD1 is involved in function downstream. This same investigation demonstrated negative modulation of GTPase activity of cpSRP by CD1.[38] Removal of CD3 has been shown in several cases to have little effect on the interaction of cpSRP54 with cpSRP43.[32,36,38]

cpFtsY Interactions

In order to complete targeting of LHCP to the thylakoid membrane, the protein cpFtsY interacts with cpSRP54. This is similar to the bacterial SRP system which also contains a homologue to cpFtsY. The cpFtsY receptor is composed of domains homologous to the cpSRP54 NG-domains preceded by an A-domain (Figure 5C). This protein exists in both a membrane bound and stromal pool where an N-terminal, amphipathic helix, termed the A-domain, mediates membrane association. GTP hydrolysis by both cpSRP54 and cpFtsY is required to effectively transfer the complex to the membrane insertion apparatuses.[24] Utilizing non-hydrolyzable GTP it was shown that the cpSRP-cpFtsY complex is stable; however, the complex is will not disassemble [39]. Upon arrival to the thylakoid membrane, the LHCP-cpSRP complex then interacts with the membrane bound protein Albino 3. The interaction takes place between

cpSRP43 and the C-terminus as well as a transmembrane helix in Albino 3.[26] To determine the necessity of the three proteins for LHCP integration, all were placed in a thylakoid containing solution in addition to GTP. Together, efficient integration of LHCP was achieved but the removal of any component, including GTP, eliminated integration.[35] Interestingly, while no ATPase was found in this system, the presence of ATP stimulates LHCP integration.[12]

YidC/Oxa1/Alb3 Family of Translocases

Albino 3 (Alb3) belongs to a family of homologous proteins involved in nucleotide triphosphate hydrolysis independent membrane protein insertion and translocation pathways. These proteins are YidC, in the inner membrane of bacteria, Oxa1, in mitochondrial inner membrane, and Alb3 in chloroplast's thylakoid membrane. Usually more than one Oxa1 and/or Alb3 homologs are found in eukaryotes [40,41]. For example, in *Arabidopsis thaliana* there are six homologs, three in the mitochondria and three in the chloroplasts. Both Alb3 and Oxa1 contain five transmembrane helices in an $N_{in}C_{out}$ orientation in the thylakoid and mitochondrial inner membranes repectively (Figure 5E).[41,42] YidC contains six transmembrane domains with the C-terminus exposed to the cytosol.[43] Between the first and second transmebrane segments of YidC is an additional periplasmic, soluble protein domain. YidC function in a Sec-dependent and Sec-indepdenent fashion for insertion of innermembrane proteins. It has also been shown to act as a chaperone for some innermembrane proteins helping them to fold properly. In particular, it plays a critical role in respiratory complex delivery. Interestingly, YidC can be replaced by both Oxa1 and Alb3 in *E. coli* for membrane protein insertion, demonstrating the significant similarity between the homologues. Cryoelectron microscopy has demonstrated the ability of YidC and Oxa1 to form dimers while associated with the ribosome leading to the speculation that they may form translocation pores. While the periplasmic domain was observed to facilitate the interaction

between YidC and SecF, its deletion did not interfere with Sec-dependent translocase activity. However, residues in TMs 2, 3 and 6 (TMs 1,2 and 5 in Oxa1) are critical for its activity.[42] A recent YidC structure provides a Sec-independent mechanism by which transmembrane segments may be inserted. Polar residues lining the transmembrane helices were proposed to draw in negatively charged terminal segments in the TM-proteins, lowering the transition energy needed for translocation. However, this mechanism is unable to explain the insertion of all YidC translocase Sec-independent proteins.[43] Oxa1 is observed to play a similar role in the mitochondrial inner membrane, playing an important role in the transport of many energy-transducing respiratory complexes. Interestingly, Oxa1, in addition to its translocase activity, was shown to act as a ribosome receptor for insertion of Cox II. This receptor activity required the C-terminus of Oxa1. Indeed, the C-terminus in the YidC/Oxa1/Alb3 familily plays a universally, albeit non-identical, role in their translocase activity. As in YidC, Oxa has been observed to ensure proper folding/assembly of proteins as well.[42] Alb3 is involved in both post and co-translational thylakoid localization of proteins including the cpSRP path and the Sec-dependent paths.[22,39,40,42] Furthermore, Alb3 is observed to play a role in folding and assembly of thylakoid membrane proteins, similar to the activities of YidC and Oxa1.[42] An albino phenotype is observed for plants that do not produce Albino 3 and is seedling lethal.[14,44]

Albino 3 Interactions

The C-terminus of Alb3 (C-term Alb3) has been shown to be disordered due to it's larger than predicted Stokes radius and by the results of circular dichroism (CD) experiments. C-term Alb3 binds cpSRP43, forming alpha helices as monitored by CD.[22,39] This binding results in an increased thermal stability for cpSRP43 (~8ºC). The interaction occurs at the CD2CD3 domains of cpSRP43 and the ankyrin region.[22,45,46] The removal of CD3 reduces the binding affinity by

one fifth. However, it is believed CD2 plays a critical role in binding. This interaction appears to be mediated by aromatic residues on cpSRP43 and an SKRS sequence at the extreme C-terminus of Alb3 as well as an AKRS sequence near the center of the C-term Alb3 protein.[22,45] Demonstrating the importance of positive charge for this interaction, mutation of lysine or arginine to alanine in the C-terminal most site results in a tenfold decrease in binding affinity and mutating both abolishes binding. C-term Alb3 is able to bind cpSRP43 in the presence of cpSRP54 although the binding strength appears to be reduced.[22] An interaction between cpSRP43 and the fifth transmembrane helix of Alb3 has also been detected in addition to two stromal loops.[45] An interaction between Alb3 and LHCP in the complex has also been observed through co-purification of C-term Alb3 with LHCP.[47] Interestingly, full length Alb3 was demonstrated by blue native gel chromatography to exist in a monomeric and dimeric state.[45] C-term Alb3 has also been shown to stimulate GTP hydrolysis by cpSRP54/cpFtsY upon binding cpSRP43 which is an important step in regulating the release of LHCP.[46] It has been demonstrated that cpSRP43 is able to co-localize with Alb3 and continue LHCP integration in the absence of cpSRP54. However, it is still possible that other cellular machinery is necessary for this cpSRP54 absent integration.

Research Plan

Few structural details for the proteins in the cpSRP dependent targeting of light harvesting proteins have been revealed. Crystal structures for cpSRP43 have determined the LHCP L18 interaction site on cpSRP43 as well as the interaction site for a short cpSRP54 derived peptide. However, these structures lack the N-terminal tail and third chromodomain of cpSRP43. Additionally, a nearly complete crystal structure of cpFtsY has been determined. However, no three-dimensional structure for cpSRP54 or its domains has been determined.

Indeed, little is known regarding the structural details of the LHCP targeting and integration mechanism owing to lack of information regarding the spatial arrangement of the domains along the signaling path. Here, we set out to understand the structural significance of the interactions along the LHCP targeting pathway. In particular, we are interested in developing and supporting a structural model of the targeting system through to the first LHCP integration step. Because of the domain arrangement in cpSRP43, it is anticipated to be quite flexible in solution. Further, as major movement of the M-domain in SRP54/Ffh is observed along its signaling path, cpSRP54 is expected to be quite flexible as well. Using this flexibility, cpSRP54 and cpSRP43 together may be adapted to facilitating LHCP loading and unloading, all of which are monitored and signaled by position changes with each interaction. Using a recently determined crystal structure and translocase mechanism for YidC, together with the cpSRP-cpSR model, we predict the first step in the mechanism for integration of LHCP to involve insertion of TM3. The first step in the process of determining this model involves the development of a three-dimensional model of free cpSRP54 and monitoring the domain arrangement(s) of both free cpSRP54 and free cpSRP43. After this has been determine the influence of cpSRP43 on cpSRP54 and cpSRP54 on cpSRP43 would provide information on the cpSRP structure. Finally, the functional significance of these interactions can be determined to provide an overall picture of the signaling process (Figure 6).

References

1 Akopian, D., Shen, K., Zhang, X. & Shan, S.-o. Signal Recognition Particle: An Essential Protein-Targeting Machine. *Annual Review of Biochemistry* **82**, 693-721, doi:10.1146/annurev-biochem-072711-164732 (2013).

2 Whitford, D. *Proteins: Structure and Function*. (John Wiley & Sons Ltd, 2005).

3 DeLille, J. *et al*. A novel precursor recognition element facilitates posttranslational binding to the signal recognition particle in chloroplasts. *Proceedings of the National Academy of Sciences* **97**, 1926-1931, doi:10.1073/pnas.030395197 (2000).

4 Stengel, K. F., Holdermann, I., Wild, K. & Sinning, I. The structure of the chloroplast signal recognition particle (SRP) receptor reveals mechanistic details of SRP GTPase activation and a conserved membrane targeting site. *FEBS Letters* **581**, 5671-5676, doi:http://dx.doi.org/10.1016/j.febslet.2007.11.024 (2007).

5 Träger, C. *et al*. Evolution from the Prokaryotic to the Higher Plant Chloroplast Signal Recognition Particle: The Signal Recognition Particle RNA Is Conserved in Plastids of a Wide Range of Photosynthetic Organisms. *The Plant Cell Online* **24**, 4819-4836, doi:10.1105/tpc.112.102996 (2012).

6 Bauer, N. C., Doetsch, P. W. & Corbett, A. H. Mechanisms Regulating Protein Localization. *Traffic* **16**, 1039-1061, doi:10.1111/tra.12310 (2015).

7 Berks, B. C., Palmer, T. & Sargent, F. Protein targeting by the bacterial twin-arginine translocation (Tat) pathway. *Current Opinion in Microbiology* **8**, 174-181, doi:http://dx.doi.org/10.1016/j.mib.2005.02.010 (2005).

8 Fekkes, P. & Driessen, A. J. M. Protein Targeting to the Bacterial Cytoplasmic Membrane. *Microbiology and Molecular Biology Reviews* **63**, 161-173 (1999).

9 Driessen, A. J. M. & Nouwen, N. Protein Translocation Across the Bacterial Cytoplasmic Membrane. *Annual Review of Biochemistry* **77**, 643-667, doi:10.1146/annurev.biochem.77.061606.160747 (2008).

10 Jarvis, P. & Robinson, C. Mechanisms of Protein Import and Routing in Chloroplasts. *Current Biology* **14**, R1064-R1077, doi:http://dx.doi.org/10.1016/j.cub.2004.11.049 (2004).

11 Keegstra, K. & Cline, K. Protein Import and Routing Systems of Chloroplasts. *The Plant Cell Online* **11**, 557-570, doi:10.1105/tpc.11.4.557 (1999).

12 Richter, C. V., Bals, T. & Schünemann, D. Component interactions, regulation and mechanisms of chloroplast signal recognition particle-dependent protein transport. *European Journal of Cell Biology* **89**, 965-973, doi:http://dx.doi.org/10.1016/j.ejcb.2010.06.020 (2010).

13 Andrès, C., Agne, B. & Kessler, F. The TOC complex: Preprotein gateway to the chloroplast. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1803**, 715-723, doi:http://dx.doi.org/10.1016/j.bbamcr.2010.03.004 (2010).

14 Hutin, C. *et al*. Double mutation cpSRP43^{-}/cpSRP54^{-} is necessary to abolish the cpSRP pathway required for the thylakoid targeting of the light-harvesting chlorophyll proteins. *The Plant Journal* **29**, 531-543 (2001).

15 Klimyuk, V. I. *et al*. A Chromodomain Protein Encoded by the Arabidopsis CAO Gene Is a Plant-Specific Component of the Chloroplast Signal Recognition Particle Pathway That Is Involved in LHCP Targeting. *The Plant Cell Online* **11**, 87-99, doi:10.1105/tpc.11.1.87 (1999).

16 Motohashi, R. *et al*. An essential role of a TatC homologue of a ΔpH- dependent protein transporter in thylakoid membrane formation during chloroplast development in Arabidopsis thaliana. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 10499-10504, doi:10.1073/pnas.181304598 (2001).

17 Mori, H. & Cline, K. Post-translational protein translocation into thylakoids by the Sec and ΔpH-dependent pathways. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1541**, 80-90, doi:http://dx.doi.org/10.1016/S0167-4889(01)00150-1 (2001).

18 Berks, B. C. The Twin-Arginine Protein Translocation Pathway. *Annual Review of Biochemistry* **84**, 843-864, doi:10.1146/annurev-biochem-060614-034251 (2015).

19 Braun, N. A., Davis, A. W. & Theg, S. M. The Chloroplast Tat Pathway Utilizes the Transmembrane Electric Potential as an Energy Source. *Biophysical Journal* **93**, 1993-1998, doi:http://dx.doi.org/10.1529/biophysj.106.098731 (2007).

20 Woolhead, C. A. *et al*. Distinct Albino3-dependent and -independent Pathways for Thylakoid Membrane Protein Insertion. *Journal of Biological Chemistry* **276**, 40841-40846 (2001).

21 Zygadlo, A., Robinson, C., Scheller, H. V., Mant, A. & Jensen, P. E. The Properties of the Positively Charged Loop Region in PSI-G Are Essential for Its "Spontaneous" Insertion into Thylakoids and Rapid Assembly into the Photosystem I Complex. *Journal of Biological Chemistry* **281**, 10548-10554 (2006).

22 Falk, S., Ravaud, S., Koch, J. & Sinning, I. The C Terminus of the Alb3 Membrane Insertase Recruits cpSRP43 to the Thylakoid Membrane. *Journal of Biological Chemistry* **285**, 5954-5962, doi:10.1074/jbc.M109.084996 (2010).

23 Berg, J., Tymoczko, J. & Stryer, L. *Biochemistry*. 5th edn, (W. H. Freeman and Company, 2004).

24 Payan, L. A. & Cline, K. A stromal protein factor maintains the solubility and insertion competence of an imported thylakoid membrane protein. *The Journal of Cell Biology* **112**, 603-613, doi:10.1083/jcb.112.4.603 (1991).

25 Henry, R., Goforth, R. & Schunemann, D. Chloroplast SRP/FtsY and Alb3 in Protein Integration into the Thylakoid Membrane. *The Enzymes* **XXV**, 493-521 (2007).

26 Celedon, J. M. & Cline, K. Intra-plastid protein trafficking: How plant cells adapted prokaryotic mechanisms to the eukaryotic condition. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1833**, 341-351, doi:http://dx.doi.org/10.1016/j.bbamcr.2012.06.028 (2013).

27 Jaru-Ampornpan, P. *et al*. ATP-independent reversal of a membrane protein aggregate by a chloroplast SRP subunit. *Nat. Struct. Mol. Biol.* **17**, 696-702, doi:10.1038/nsmb.1836 (2010).

28 Falk, S. & Sinning, I. cpSRP43 Is a Novel Chaperone Specific for Light-harvesting Chlorophyll a,b-binding Proteins. *Journal of Biological Chemistry* **285**, 21655-21661, doi:10.1074/jbc.C110.132746 (2010).

29 Tzvetkova-Chevolleau, T. *et al*. Canonical signal recognition particle components can be bypassed for posttranslational protein targeting in chloroplasts. *Plant Cell* **19**, 1635-1648, doi:10.1105/tpc.106.048959 (2007).

30 Amin, P. *et al*. Arabidopsis mutants lacking the 43- and 54-kilodalton subunits of the chloroplast signal recognition particle have distinct phenotypes. *Plant Physiol* **121**, 61-70 (1999).

31 Jonas-Straube, E., Hutin, C., Hoffman, N. E. & Schünemann, D. Functional Analysis of the Protein-interacting Domains of Chloroplast SRP43. *Journal of Biological Chemistry* **276**, 24654-24660, doi:10.1074/jbc.M100153200 (2001).

32 Hermkes, R., Funke, S., Richter, C., Kuhlmann, J. & Schünemann, D. The α-helix of the second chromodomain of the 43 kDa subunit of the chloroplast signal recognition particle facilitates binding to the 54 kDa subunit. *FEBS Letters* **580**, 3107-3111, doi:http://dx.doi.org/10.1016/j.febslet.2006.04.055 (2006).

33 Franklin, A. E. & Hoffman, N. E. Characterization of a chloroplast homologue of the 54-kDa subunit of the signal recognition particle. *Journal of Biological Chemistry* **268**, 22175-22180 (1993).

34 Cain, P., Holdermann, I., Sinning, I., Johnson, A. & Robinson, C. Binding of chloroplast signal recognition particle to a thylakoid membrane protein substrate in aqueous solution and delineation of the cpSRP43–substrate interaction domain. *Biochem J.* **437**, 149-155 (2011).

35 Tu, C.-J., Schuenemann, D. & Hoffman, N. E. Chloroplast FtsY, Chloroplast Signal Recognition Particle, and GTP Are Required to Reconstitute the Soluble Phase of Light-harvesting Chlorophyll Protein Transport into Thylakoid Membranes. *Journal of Biological Chemistry* **274**, 27219-27224, doi:10.1074/jbc.274.38.27219 (1999).

36 Stengel, K. F. *et al*. Structural Basis for Specific Substrate Recognition by the Chloroplast Signal Recognition Particle Protein cpSRP43. *Science* **321**, 253-256, doi:10.1126/science.1158640 (2008).

37 Kathir, K. M. *et al*. Assembly of Chloroplast Signal Recognition Particle Involves Structural Rearrangement in cpSRP43. *Journal of Molecular Biology* **381**, 49-60, doi:http://dx.doi.org/10.1016/j.jmb.2008.05.065 (2008).

38 Goforth, R. L. *et al*. Regulation of the GTPase Cycle in Post-translational Signal Recognition Particle-based Protein Targeting Involves cpSRP43. *Journal of Biological Chemistry* **279**, 43077-43084, doi:10.1074/jbc.M401600200 (2004).

39 Moore, M., Goforth, R. L., Mori, H. & Henry, R. Functional interaction of chloroplast SRP/FtsY with the ALB3 translocase in thylakoids: substrate not required. *The Journal of Cell Biology* **162**, 1245-1254, doi:10.1083/jcb.200307067 (2003).

40 Stuart, R. A. Insertion of proteins into the inner membrane of mitochondria: the role of the Oxa1 complex. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1592**, 79-87, doi:http://dx.doi.org/10.1016/S0167-4889(02)00266-5 (2002).

41 Luirink, J., Samuelsson, T. & de Gier, J.-W. YidC/Oxa1p/Alb3: evolutionarily conserved mediators of membrane protein assembly. *FEBS Letters* **501**, 1-5, doi:http://dx.doi.org/10.1016/S0014-5793(01)02616-3 (2001).

42 Wang, P. & Dalbey, R. E. Inserting membrane proteins: The YidC/Oxa1/Alb3 machinery in bacteria, mitochondria, and chloroplasts. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1808**, 866-875, doi:http://dx.doi.org/10.1016/j.bbamem.2010.08.014 (2011).

43 Kumazaki, K. *et al*. Structural basis of Sec-independent membrane protein insertion by YidC. *Nature* **509**, 516-520, doi:10.1038/nature13167http://www.nature.com/nature/journal/v509/n7501/abs/nature13167.html - supplementary-information (2014).

44 Moore, M., Harrison, M. S., Peterson, E. C. & Henry, R. Chloroplast Oxa1p Homolog Albino3 Is Required for Post-translational Integration of the Light Harvesting Chlorophyll-binding Protein into Thylakoid Membranes. *Journal of Biological Chemistry* **275**, 1529-1532, doi:10.1074/jbc.275.3.1529 (2000).

45 Dünschede, B., Bals, T., Funke, S. & Schünemann, D. Interaction Studies between the Chloroplast Signal Recognition Particle Subunit cpSRP43 and the Full-length Translocase Alb3 Reveal a Membrane-embedded Binding Region in Alb3 Protein. *Journal of Biological Chemistry* **286**, 35187-35195, doi:10.1074/jbc.M111.250746 (2011).

46 Lewis, N. E. *et al*. A Dynamic cpSRP43-Albino3 Interaction Mediates Translocase Regulation of Chloroplast Signal Recognition Particle (cpSRP)-targeting Components.

Journal of Biological Chemistry **285**, 34220-34230, doi:10.1074/jbc.M110.160093 (2010).

47 Bals, T., Dünschede, B., Funke, S. & Schünemann, D. Interplay between the cpSRP pathway components, the substrate LHCP and the translocase Alb3: An in vivo and in vitro study. *FEBS Letters* **584**, 4138-4144, doi:http://dx.doi.org/10.1016/j.febslet.2010.08.053 (2010).

Figures

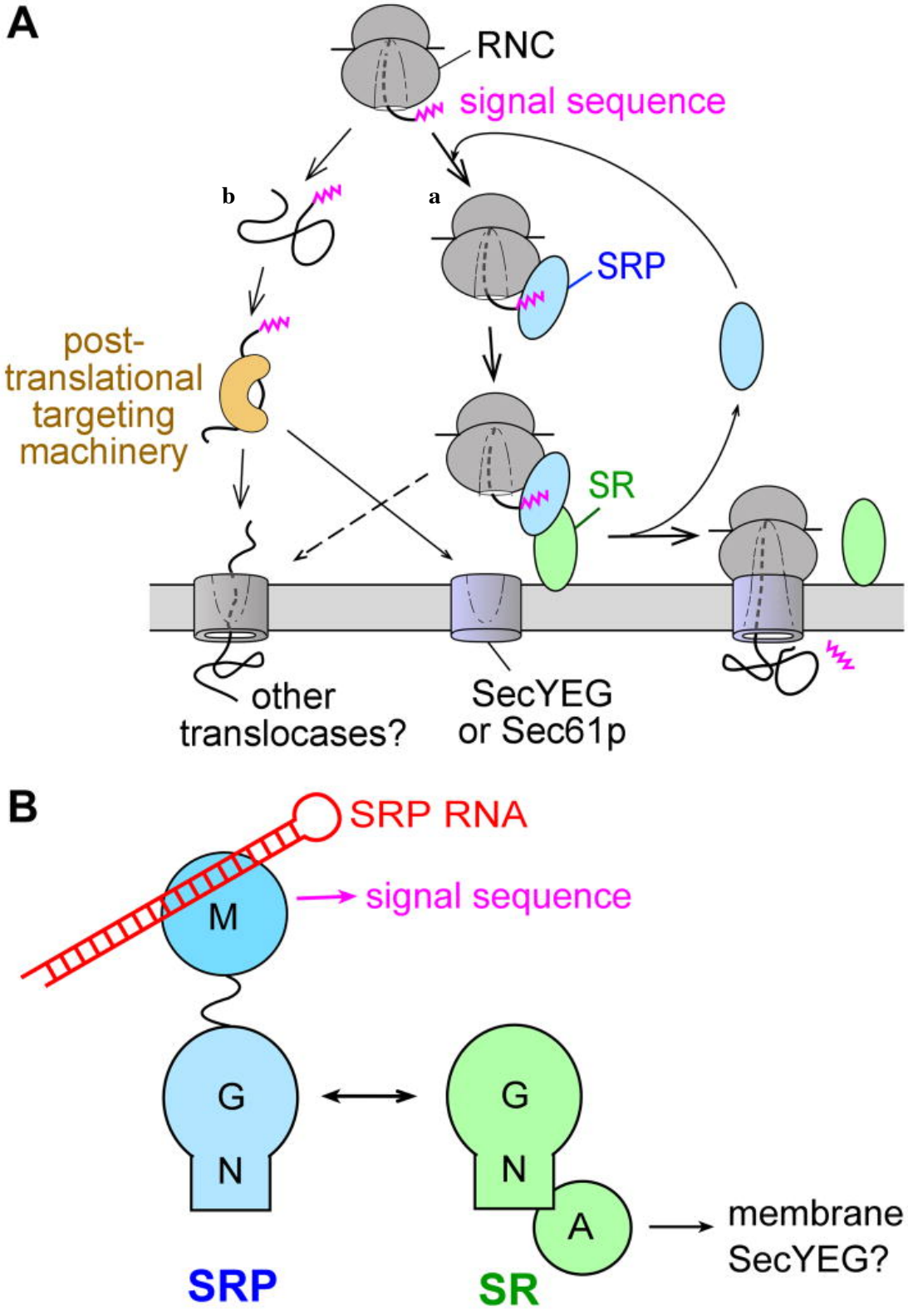
A
RNC
signal sequence
b
a
SRP
post-
translational
targeting
machinery
SR
other
translocases?
SecYEG
or Sec61p
B
SRP RNA
M
signal sequence
G
N
G
N
A
membrane
SecYEG?
SRP
SR

Figure 1: An overview of the co-translational and post-translational targeting of proteins. *(A)* Targeting begins with the synthesis of the first twenty to seventy amino acids *(a)* Proteins with a signal sequences specific for the co-translational targeting system are bound by SRP which halts protein synthesis. The RNC-SRP complex is then transported the SRP receptor and finally the translocon. GTP hydrolysis recycles the SRP-SR after which protein synthesis resumes, transporting the protein across the membrane. *(b)* Proteins without co-translational signal sequences are fully synthesized in the cytosol and transported to their proper location via specific chaperones and targeting machinery. *(B)* The SRP is composed of a 54kDa SRP54 which contains an N-terminal N-domain and a GTPase G-domain flexibly linked to a methionine rich M-domain. It is the M-domain that binds the universally conserved SRP RNA. The SR in prokaryotes is composed of a membrane binding A-domain and an SRP NG-domain homologous NG-domain. In eukaryotes the membrane binding and GTPase elements are split into a heterodimer. The SRP and SR form a twinned GTPase interaction which is transported to the translocon, SecYEG in bacteria. This figure was adapted from *Akopain et. al., Annual review of biochemistry. 2013;82:693-721*

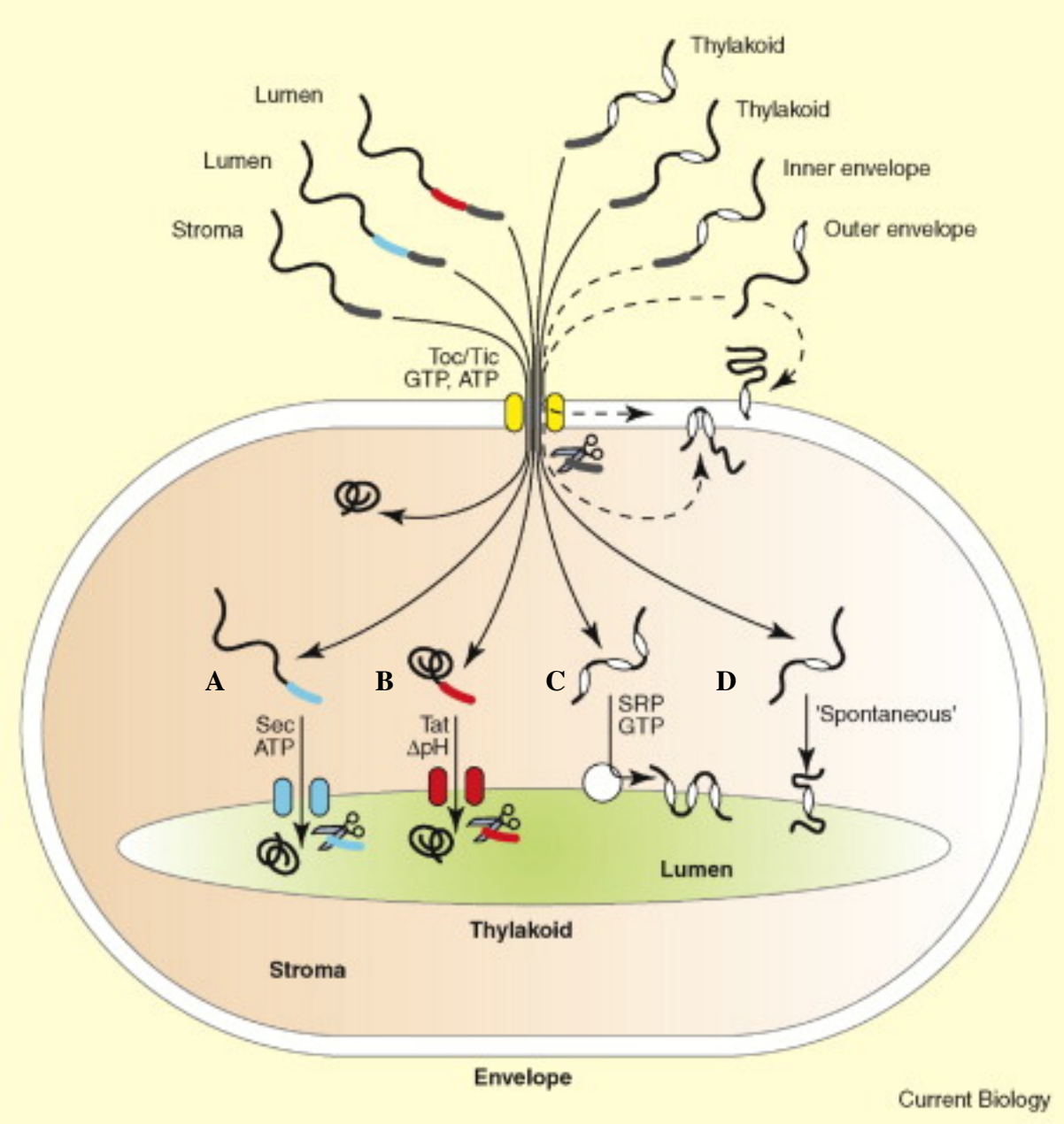
Thylakoid
Lumen
Thylakoid
Lumen
Inner envelope
Stroma
Outer envelope
Toc/Tic
GTP, ATP
A
B
C
D
Sec
ATP
Tat
ΔpH
SRP
GTP
'Spontaneous'
Lumen
Thylakoid
Stroma
Envelope
Current Biology

Figure 2: Targeting of proteins to and in the chloroplast. Nuclear encoded chloroplast proteins are synthesized in the cytosol and directed to the Toc/Tic translocation machinery by specific signal sequence or, for outer membrane proteins, may be directed by sequence specific attributes. Once in the stroma the signal sequence is cleaved. For proteins destined for the thylakoid lumen or membrane, additional cleavable signal sequences or sequence attributes direct them to their proper location. Four different targeting mechanisms exist for this. *(A)* The Sec-dependent pathway is a bacterial SecA/Sec-YEG translocation homologue which utilizes ATP to drive proteins across the membrane. *(B)* An additional pathway exists which does not use nucleotide triphosphates, but rather the pH gradient across the membrane to drive translocation. *(C)* A special SRP in the chloroplast delivers light harvesting chlorophyll binding proteins to the thylakoid membrane via a heterodimer composed of an SRP54 homologue, cpSRP54, and a novel 43kDa protein. Additionally, cpSRP54 is found to bind stromal ribosomes a facilitate the delivery of proteins via the traditional co-translational path. *(D)* A final mechanism for insertion of some thylakoid proteins occurs in the absence of translocation machinery. Targeting elements in the membrane helices and protein loops allow for spontaneous insertion of the proteins into the thylakoid membrane. This figure was adapted from *Jarvis et. al. Current Biology, 2004;14:R1064–R1077*

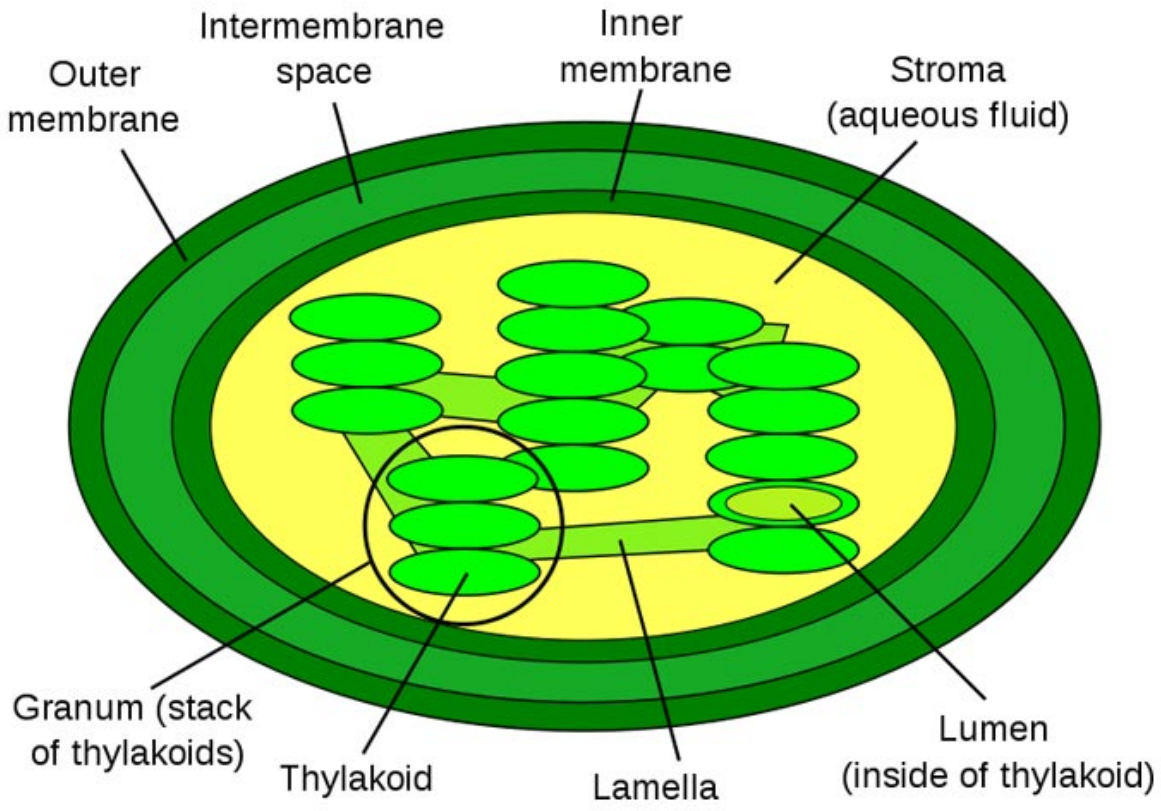

Figure 3: The chloroplast. The chloroplast is a specialized plastid responsible primarily for the process of photosynthesis. It is enclosed by an outer and an inner membrane between which is an intermembrane space. Inside the inner membrane is the stroma which contains stacks of thylakoids referred to as grana which are connected via lamella. Inside the thylakoids is an acidic space called the thylakoid lumen. This figure was adapted from *http://byjus.com/biology/chloroplasts//*.

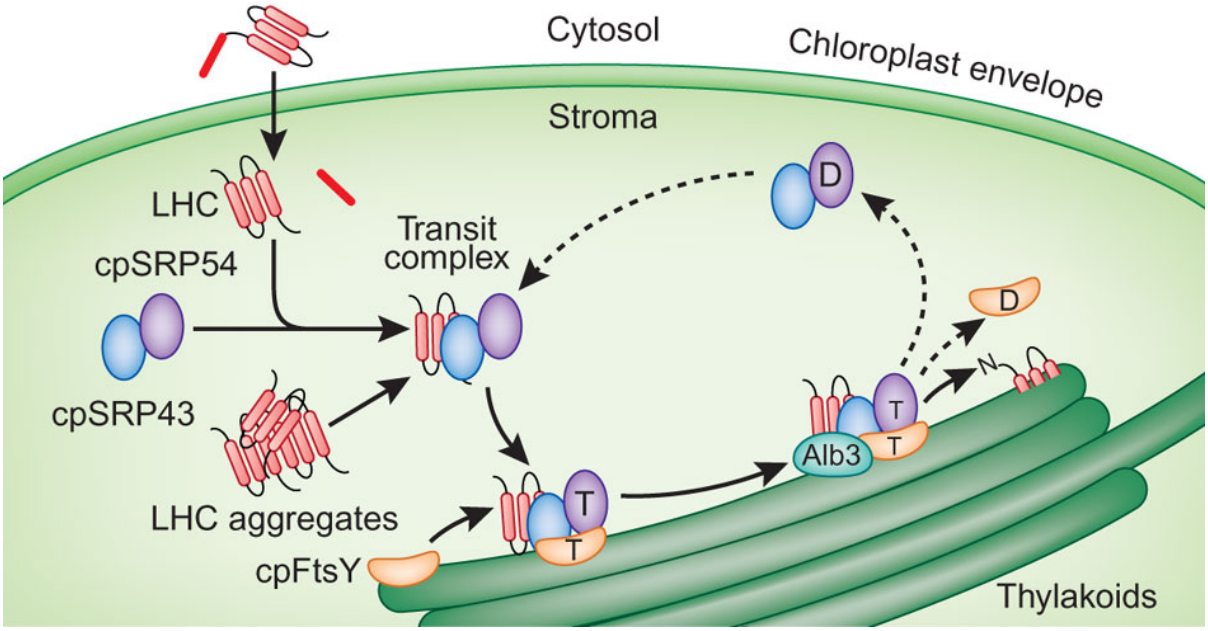

Figure 4: Targeting of light harvesting chlorophyll binding proteins (LHCP) to the thylakoid membrane. The LHCP protein is first synthesized in the cytosol and transported to the chloroplast stroma via a signal sequence. After transport across the outer and inner membranes of the chloroplast the signal sequence is cleaved. In the absence of cpSRP LHCP forms aggregates. The chloroplast signal recognition particle (cpSRP) heterodimer, composed of cpSRP43 and cpSRP54, binds LHCP to form transit complex. The transit complex then binds the cpSRP receptor cpFtsY, which transports the complex to the Alb3 translocase. LHCP is then integrated into the thylakoid membrane followed by twinned cpSRP54/cpFtsY GTPase ativity which recycles the targeting machinery. This figure was adapted from *Henry et. al. Nature 2012 Nature Structural & Molecular Biology 17, 676–677 (2010)*

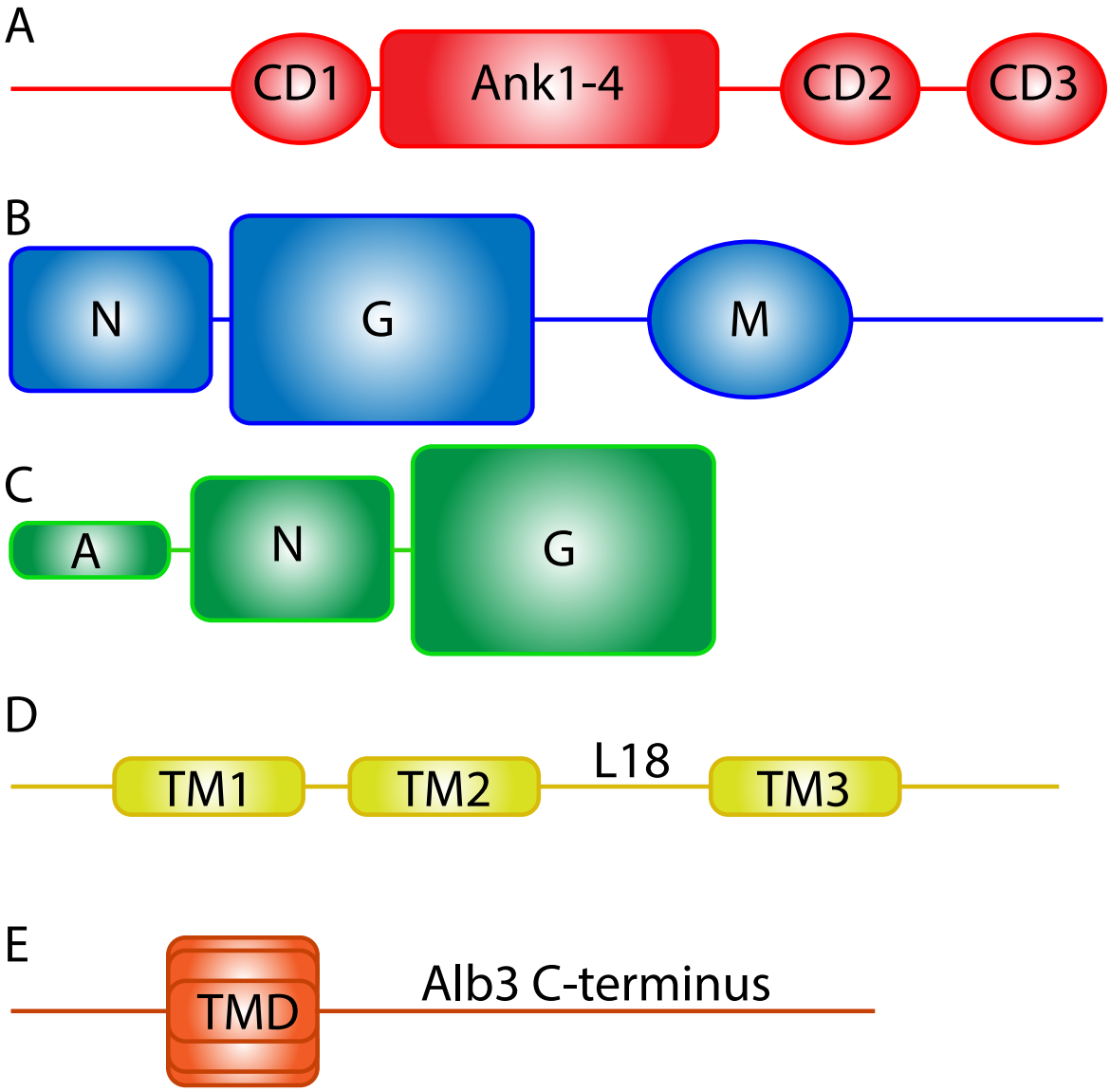
A
CD1
Ank1-4
CD2
CD3
B
N
G
M
C
A
N
G
D
TM1
TM2
L18
TM3
E
TMD
Alb3 C-terminus

Figure 5: The domain structure of the chloroplast signal recognition particle (cpSRP) proteins. *(A)* cpSRP43; cpSRP43 contains an N-terminal tail followed by the first chromodomain (CD1) which is fused to the first and second ankyrin repeats (Ank1 and Ank2). These are followed by Ank3 and Ank4. The ankyrin repeats are responsible the binding of LHCP. Ank4 is linked by a short chain to chromodomain 2 (CD2) responsible for binding cpSRP54 CD2 is followed by chromodomain 3 (CD3) which is linked via another short chain. *(B)* cpSRP54; cpSRP54 is composed of an N-terminal N-domain closely associated to a GTPase G-domain. The G-domain is connected to the methionine rich M-domain by a long linking segment. The chloroplast SRP54 contains an additional C-terminal extension from the M-domain responsible for binding cpSRP43. *(C)* cpFtsY; cpFtsY is composed of a membrane binding, N-terminal A-domain followed by a cpSRP54 NG-domain homologue NG-domain *(D)* Light harvesting chlorophyll binding protein (LHCP); the N-terminal stromal segment of LHCP is followed by transmembrane segments TM1 and TM2. Between TM2 and TM3 is the cpSRP43 binding element L18. The remaining portion after TM3 resides in the luminal portion of the thylakoid. *(E)* Albino3; The translocase for the cpSRP system, Alb3 contains a luminal N-terminus followed by five transmembane helices. The stromal C-terminus is disordered and is responsible for binding of cpSRP43

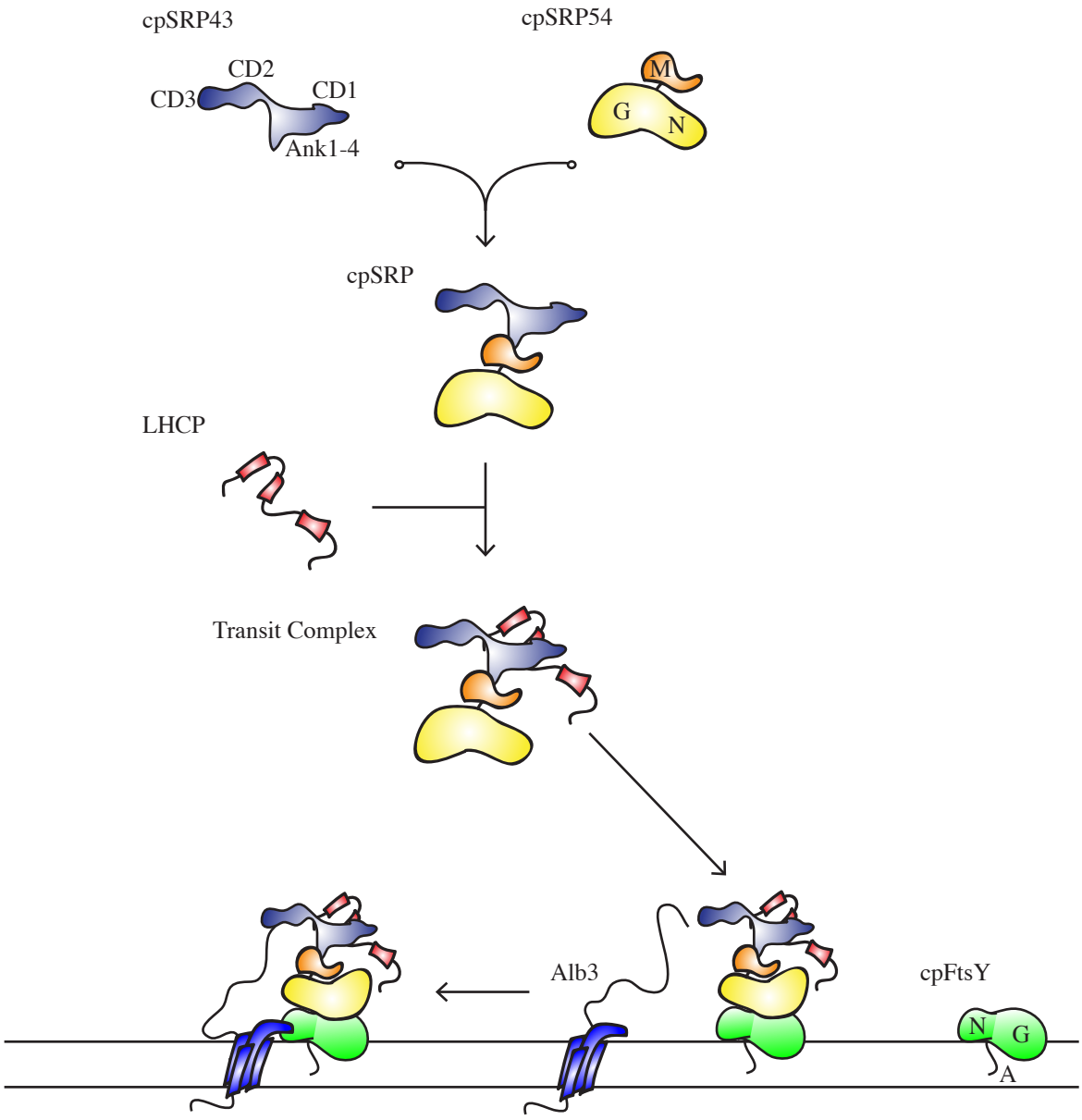
cpSRP43
CD2
CD3
CD1
Ank1-4
cpSRP54
M
G
N
cpSRP
LHCP
Transit Complex
Alb3
cpFtsY
N
G
A

Figure 6: The current structural understanding of the chloroplast signal recognition particle (cpSRP) during the targeting of light harvesting chlorophyll binding protein (LHCP) to the thylakoid membrane. *(A)* Free cpSRP43. Based upon the connectivity of the domains in cpSRP43 a large degree of flexibility between Ank4-CD2 and CD2-CD3 may be present. *(B)* Free cpSRP54. The conformational states in the cytosolic SRP system are quite variable and dependent upon SRP RNA. Due to the absence of SRP RNA in the cpSRP system it is uncertain what orientation or orientations the cpSRP54 M-domain takes relative to the NG-domain. *(C)* The cpSRP heterodimer. Little information is available for the full cpSRP. A small peptide derived from the C-terminus of cpSRP54 is bound to cpSRP43 between Ank4 and CD2 in a previously published crystal structure which provides some clues as to the positioning of the M-domain. *(D)* Light-harvesting chlorophyll binding protein. The three transmembrane helices bind chlorophylls and carotenoids present in the thylakoid membrane, holding them in a specific orientation. *(E)* Transit Complex. Binding of LHCP to cpSRP forms the transit complex. The M-domain may interact with TM3 of LHCP although this has yet to be confirmed in the cpSRP system. *(F)* cpSRP54 receptor, cpFtsY. The A-domain of cpFtsY helps to anchor it in the thylakoid membrane with the cpSRP54 binding GTPase region facing away from the membrane. *(G)* No information is available to support a hypothetical configuration for the receptor bound transit complex relative to the thylakoid membrane. However, as the cpSRP54/cpFtsY homolouge Ffh/FtsY interaction orientation is known, and since cpFtsY retains the FtsY NG-domain fold, the interaction with cpSRP54 and cpFtsY is likely identical. This limits the potential orientations of the transit complex/cpFtsY structure at the membrane. *(H)* Alb3 is known to interact with cpSRP43. However, little is known about the structural impact of this interaction.

II. Regulation of Structural Dynamics within a Signal Recognition Particle Promotes Binding of Protein Targeting Substrates

Feng Gao[1], Alicia D. Kight[2], Rory Henderson[1], Srinivas Jayanthi[1], Parth Patel[1], Marissa Murchison[1], Priyanka Sharma[2], Robyn L. Goforth[2], Thallapuranam Krishnaswamy Suresh Kumar[1*], Ralph L. Henry[2*], and Colin D. Heyes[1*]

[1]Department of Chemistry and Biochemistry, University of Arkansas, Fayetteville, AR 72701

[2]Department of Biological Sciences, University of Arkansas, Fayetteville, AR 72701

Abstract

Protein targeting is critical in all living organisms and involves a signal recognition particle (SRP), an SRP receptor and a translocase. In co-translational targeting, interactions between these proteins is mediated by the ribosome. In chloroplasts, the light-harvesting chlorophyll-binding protein (LHCP) in the thylakoid membrane is targeted post-translationally without a ribosome. A multi-domain chloroplast-specific subunit of the SRP, cpSRP43, is proposed to take on the role of coordinating the sequence of targeting events. Here, we model the observed inter-domain dynamics of cpSRP43. We show that the affinity of cpSRP43 for the binding motif of LHCP (L18) increases when cpSRP43 is complexed to the binding motif of cpSRP54 ($cpSRP54_{pep}$). These results support the conclusion that substrate binding to the chloroplast SRP is modulated by protein structural dynamics, in which a major role of cpSRP54 is to improve substrate-binding efficiency to the cpSRP.

Introduction

Protein-targeting in cells occurs via a wide variety of sorting pathways, depending on both the synthesis site and target location of the protein. Some proteins are targeted co-translationally, with central involvement of the ribosome, while other proteins are targeted post-translationally. For example, in chloroplasts, nuclear-encoded thylakoid proteins are routed from the stroma to the thylakoid membrane following their import from the cytosol via multiple pathways [1,2]. The multiple pathways accommodate the various transport requirements of a broad range of proteins that function in the thylakoid lumen or in the lipid bilayer. One of the most important of these nuclear-encoded proteins is the light harvesting chlorophyll-binding proteins (the LHCPs). A unique chloroplast signal recognition particle (cpSRP) pathway has evolved in

the stroma to target this protein without the need for the ribosome to coordinate the sequence of targeting events. Similar to SRPs in the cytosol of pro- and eukaryotes, cpSRP in the stroma contains a conserved 54 kDa GTPase (cpSRP54), although an RNA moiety is conspicuously absent [3]. Fractionation of chloroplast stroma proteins demonstrated that cpSRP54 is present in two pools [4]. One pool is associated with ribosomes and appears to function in co-translational targeting of proteins to the chloroplast. A second cpSRP54 pool is present in a heterodimeric complex bound to cpSRP43, a 43 kDa protein that is unique to chloroplasts [5-7]. The cpSRP heterodimer functions post-translationally to bind and target full-length LHCP targeting substrates to the thylakoid membrane by a ribosome-independent mechanism [3]. cpSRP43 therefore appears to take on the role of the ribosome in coordinating the sequence of protein targeting events.

The only known targeting substrates of the post-translational cpSRP localization pathway are LHCPs, a protein family encoded by lhca and lhcb genes in Arabidopsis. LHCPs coordinate pigment binding in peripheral photosystem 1 and 2 light harvesting complexes. The importance of the post-translational cpSRP-based targeting mechanism is apparent when considering that LHCPs can account for up to 50% of the bulk protein in thylakoids, making them one of the most abundant membrane proteins on earth [8]. Like other nuclear-encoded thylakoid proteins, LHCPs are synthesized in the cytosol as full-length precursor proteins and imported into the chloroplast stroma as directed by a cleavable N-terminal chloroplast-targeting domain [1,2]. Processing of the chloroplast targeting peptide yields a mature-sized LHCP in the stroma that is targeted to thylakoid membranes as a soluble cpSRP-LHCP complex termed ‘transit complex’ [4,7,9].

Binding of LHCP to cpSRP serves to maintain the solubility and insertion competence of LHCPs [10]. *In vitro* studies demonstrate that addition of LHCPs to heterodimeric cpSRP, but not to cpSRP54 or cpSRP43 alone, reconstitutes transit complex formation and supports LHCP insertion into isolated thylakoids by a mechanism that requires GTP, the protein insertase Albino3 (Alb3), and cpFtsY, a GTPase homologue of the SRP receptor in bacteria [3,11]. Cross linking studies indicate that cpSRP54 interacts with LHCPs in transit complex [6], but cpSRP54 alone shows no binding to LHCPs. In contrast, cpSRP43 alone and in transit complex binds LHCPs owing to its interaction with a conserved LHCP motif, termed L18 [12,13]. This interaction is required for transit complex formation and for ATP-independent chaperon activity exhibited by cpSRP43 alone [14,15]. The cpSRP43 chaperon activity is likely important for efficient transit complex formation that leads to additional interaction of LHCP transmembrane domains 2 and 3 with cpSRP43 [16,17]. Vectorial LHC targeting to the thylakoid membrane relies on both cpSRP subunits. cpSRP54 exhibits affinity for lipid-bound cpFtsY [18] and cpSRP43 interacts with Alb3 [19-21]. Both the timing of LHCP release from cpSRP and release of cpSRP from its membrane-bound receptor are triggered by interaction between cpSRP43 with Alb3 [19]. The central importance of cpSRP43 is underscored by its range of interaction partners and the abundance of interaction domains that dominate its structure; three chromatin-binding domains (CD) and four Ankyrin-repeats (Ank) [3]. We hypothesize that the order of these interactions will be controlled through changes in cpSRP43 structural dynamics that alter the relative position of these domains in response to association of cpSRP43 with each of its binding partners. It is this hypothesis that forms the basis of the present study.

Single molecule fluorescence has proven to be extremely powerful for studying the biophysics of complex biological systems, such as folding of proteins and RNA [22-26],

biomolecular diffusion on membranes and cells [27-30], protein-protein interactions [31] and protein conformational dynamics [32-34]. We employed single molecule FRET (smFRET) [35,36] to probe the inter-domain conformational dynamics of cpSRP43 and how they are affected by binding cpSRP54. We used the crystal [17,37] and NMR structures [38] as well as SAXS data [15] of cpSRP43 to design mutants for site-specific labeling within its various domains for the smFRET studies.

Specifically, constructs containing two cysteines, one in each domain to be studied, were either retained for the two native cysteines in Ank2 and CD2, or modified such that the native cyteines were removed and new ones placed in the appropriate domains. The positions probed were Ank2-CD3, Ank2-Ank4, An2-CD2, CD2-CD3, CD1-Ank1 and a CD2-CD3 construct in which CD1-Ank4 was deleted. These were termed P1, P2, P3, P4, P5, and P4Δ, respectively (Table 1 and Figure 1) [39].

In order to ensure the observed smFRET corresponded specifically to changes in a native like cpSRP43, several control experiments were performed. First, the proteins were over-labeled in order to quantify the integration activity of the label cpSRP43. This was done to ensure the labels did not interfere with integration so that the experiments could be reliably related to the function of cpSRP43. The results indicated labeled cpSRP43 retained its ability to support integration. Next, the affinity between labeled cpSRP43 and cpSRP54 was demonstrated to be comparable to the already determined native affinity [40]. The remaining two controls were specific to the smFRET experiment itself. First, the zero FRET peak, generated due to the presence of donor only and donor-donor labeling, in polyproline peptides was compared to the zero FRET peak in the cpSRP43 data. The similarity between the two allowed for the subtraction of the zero FRET portion of the histograms from cpSRP43 data, improving the clarity of the results. Finally, as changes in κ^2 can affect the FRET histogram generation, the dyes on cpSRP43 were checked

for their ability to rotate freely. The results of polarization anisotropy experiments revealed to be ~2/3, consistent for freely rotating dyes. Together, these control experiments allowed for the interpretation of the results as being consistent with the properties of an unlabeled, native cpSRP43 [39].

The results showed that cpSRP43 is characterized by a large amount of structural heterogeneity. Specifically, histograms for each constructs displayed a wide distribution in the FRET efficiency. Further, the experimental distributions (red and blue) were markedly wider than the FRET peaks calculated from the available SAXS envelope (grey) (Figure 2) [41]. Fluorescence correlation spectroscopy for cpSRP43 affixed to a PEG surface showed that the distributions represent a flexible particle rather than multiple static states. The various constructs compared cpSRP43 alone (red) and cpSRP54 bound cpSRP43 (blue) to determine the effect of cpSRP54 on cpSRP43 (Figure 2). The Ank2-CD3 (P4) histogram showed significant flexibility that was largely unchanged by the presence of cpSRP54. In contrast, both the Ank2-Ank4 and Ank2-CD2, P2 and P3, respectively, showed significant flexibility with a significant reduction in flexibility upon interaction with cpSRP54. The CD2-CD3 position (P4) gave similar results, *albeit*, with a less pronounced affect of cpSRP54 on the cpSRP43 flexibility. The CD2-CD3 only construct (P4Δ) was quite similar to the full length cpSRP43 CD2-CD3 probe suggesting CD2 and CD3 form no interaction with the other domain capable of significantly affecting the flexibility. This is consistent with the fact that CD2 interactions strongly with cpSRP54 [40]. The final domains probed were between CD1 and Ank2 (P5). These domains showed the least flexibility and were unaffected by the presence of cpSRP54. Together, the data demonstrated cpSRP43 was characterized by a large amount of flexibility and that this flexibility is

dramatically reduced in regions near the LHCP L18 binding site [17] via interaction with cpSRP54 [39].

All-atom and course-grained molecular dynamics simulations are used here to support the smFRET data and provide more detailed structural models. These results show that the dynamics observed in the smFRET experiments for cpSRP43 alone exhibits wide inter-domain conformational dynamic heterogeneity (i.e. high flexibility), and upon binding to cpSRP54 shows reduced regional flexibility. We hypothesized that this reduced flexibility favors binding of LHCP targeting substrates, which was then tested using isothermal titration calorimetry (ITC). We found that the affinity of cpSRP43 for L18 increased by ~3-fold when cpSRP43 was complexed with a cpSRP54 peptide that corresponds to the cpSRP43 binding site on cpSRP54. Together, these findings support a model in which cpSRP54 modulates the structural dynamics of cpSRP43, which is critical for the promotion of efficient binding and localization of LHCP targeting substrates.

Methods

All-Atom Molecular Dynamics Simulations

The crystal structure for cpSRP43 CD1-CD2 (PDB ID 3UI2) was used for the 100 ns simulations [37]. Two different structures were prepared, one containing the cpSRP54 peptide and one with the peptide removed. Both were equilibrated separately. Another construct in which the NMR CD3 structure (PDB ID 1X3P) was attached to the C-terminus of the 3UI2 structure was produced [38]. For this, structure eighteen of 1X3P was submitted to the ModRefine server followed by energy minimization in SYBYL-X using the AMBER-FF99 forcefield. Next, the CD1-CD2, 3UI2 structure and the refined CD3 structure were aligned with the cpSRP43 SAXS density produced

from the dummy atom model (provided by the Shan group, Caltech) with SITUS using a combination of manual and automated fitting in UCSF Chimera [15,42,43]. Since fitting required adjusting the linking region between Ank4 and CD2 such that CD2 is distant from Ank4, to keep the structure modifications to a minimum, this fitting was used as a guide to attach CD3 to CD2 in the unaltered 3UI2 structure. This structure was used in the simulations needed to prepare the shape based coarse grained (SBCG) model.

After an initial minimization of the structure, the protein was solvated, leaving a minimum of 10 Å distance between the protein and the solvent box edge. The system was neutralized with NaCl and the effective salt concentration was set to 0.150 M. To prepare the structure for production runs, the system was subjected to several refinement steps to relax the backbone, side chains, and surrounding water molecules and ions. Equilibration of the system was carried out in the NPT ensemble using NAMD 2.9 and the CHARMM27 forcefield [44,45]. A temperature of 300 K was maintained using Langevin dynamics, with a damping coefficient of 1 ps^{-1}. A constant pressure of 1 atm was maintained using the Langevin piston method with a period of 100 fs and a decay time of 50 fs. Long-range electrostatic interactions were computed using the particle mesh Ewald method with periodic boundary conditions. Electrostatic and Van der Waals interactions were cut-off beyond a distance of 12 Å using switching functions beginning at 10 Å. All hydrogen atoms were kept rigid. The simulation time step was 1 fs unless stated otherwise. In the first step of preparation the protein side-chains were minimized for 10,000 steps with the backbone atoms fixed in the absence of water and ions. The resulting structure was solvated as described above. To facilitate the relaxation of water around the protein, all protein atoms were fixed and the system was subjected to 1,500 steps of minimization and 50 ps of dynamics. For further relaxation of the solvent system around the

protein harmonic constraints were used on the protein atoms with a force constant of 1 kcal/(mol $Å^2$). In the final step before production runs a gradual temperature increase from 0 K to 300 K was implemented. Temperature reassignment by an increment of 10 K every 2 ps was used to reach 300 K followed by 250 ps of dynamics. Production run preparedness was determined based upon leveling of the RMSD and energy fluctuations.

In order to evaluate flexible regions in the 3UI2 proteins, simulations of 100 ns were performed. A 2 fs time step was used for these runs with all other parameters identical to the last 250 ps of the final equilibration run. Backbone RMSDs were evaluated using the RMSD trajectory tool and the simulations visualized and C-α distances measured in VMD [46]. The full length cpSRP43 structure was simulated similarly for a total of 5 ns with harmonic restraints on the backbone atoms of CD3 with a force constant of 1 kcal/(mol $Å^2$). Restraints were used due to instability of the CD3 domain observed in prior simulations.

Shape Based Coarse Grained (SBCG) Modeling

The minimized cpSRP43 from the all atom cpSRP43 CD1-CD3 run was used for the SBCG building. The coarse grained model building methods are as described [47]. The VMD CGTools plugin was used to build a 30 bead model using 25,000 iterations following the chain connectivity of the all-atom structure. The mass and charge of each bead is determined based upon the atoms assigned to them. Hydrophobic interactions are modeled via a Leonard Jones (LJ) potential with the parameters determined from the all atom structure setting the maximum well depth to 10 kcal/mol and adding an additional 1 Å to the LJ radius. Bonded interactions are modeled via a harmonic potential. For further refinement the appropriate values are extracted from the restrained all-atom simulation. Briefly, several iterations of short SBCG simulation

were run to determine the effective Force constants in the coarse model which were individually adjusted between iterations to improve the agreement with the values obtained from the all-atom simulation.

A temperature of 300K was maintained using Langevin dynamics, with a damping coefficient of 2 ps^{-1}. Periodic boundary conditions were used with a cut-off beyond a distance of 35 Å using switching functions beginning at 20 Å. Boundaries were selected such that the beads in the adjacent systems would be non-interacting. The simulation time step for simulations was 100 fs. In order to improve sampling of configurational space, each construct was simulated a total of 40-50 μs twelve times using different random number seeds in each simulation.

Isothermal Titration Calorimetry

All isothermal titration calorimetric experiments were performed at 25°C using a VP-ITC (MicroCal Inc. Northampton, MA). The concentration of protein samples used in the isothermal titration calorimetry was estimated using the Bradford micro-dye binding assay. The concentration of the peptides ($cpSRP54_{pep}$ and L18 peptide) was estimated by Scopes method using the 205 nm absorbance [48]. Protein and peptide samples were centrifuged at 4ºC at 13,000 rpm for 15 minutes prior to the ITC experiments to remove any aggregated or precipitated material(s). All sample solutions were degassed under vacuum and equilibrated for 30 min prior to titration. 0.025 mM of cpSRP43 [10 mM HEPES, 10 mM $MgCl_2$ buffer (pH 8.0)] was placed in a reaction cell and titrated with 0.25 mM of the L18 peptide dissolved in the same buffer. The contents of the syringe (L18 peptide) were added sequentially in 6 μl aliquots to the reaction cell (0.025 mM cpSRP43 alone or 5:1 mixture of $cpSRP54_{pep}$ and cpSRP43) at an interval of 120 seconds. The titration curves were corrected by performing appropriate blank titrations. The data

were analyzed using *Origin* (MicroCal Inc). The raw ITC data were individually fitted to the one-set-of-sites binding model.

Results and Discussion

Molecular dynamics (MD) simulations suggest support for multiple conformational states in cpSRP43

All-atom and shape-based coarse-grained (SBCG) simulations were run to lend support to the observations based on FRET data that cpSRP43 is a very flexible protein [39]. All-atom simulations were run using the crystal structure of the CD1-Ank-CD2 portion of the protein (PDB ID 3UI2) in the presence and absence of the cpSRP54 peptide for a total of 100 ns each to examine the changes in the motion(s) of CD2 relative to CD1-Ank4 upon binding to cpSRP54. Movies of these simulations are provided in the supporting information (Movie S1 and S2, respectively). From these movies, the inter-domain distance between the C_α carbons of the labeled residues in Ank2 and CD2 was extracted as a function of time (Figure 3a). Figure 3a shows that the flexibility of the protein in this region is significantly reduced when the $cpSRP54_{pep}$ is bound.

Shape based coarse graining (SBCG) of a protein structure allows for the assignment of clusters of atoms to a set of beads that inherit the aggregate properties of the atoms assigned to it helping to reduce the system size and achieve significantly increased simulation time scales. The SBCG simulations include CD3 and were run for 20 μs in order to reach timescales similar to those of the FRET experiments. The bead connectivity was adjusted to match the shape of cpSRP43. Based on the shape, the final selected construct included relatively flexible linkers between Ank4 and CD2 and between CD2 and CD3 (Figure 3b). An additional construct was

prepared via the addition of a restricting bond between the terminal Ank4 bead and a bead in CD2 to mimic the effect of the cpSRP54 peptide, which is known to bind to this region. Importantly, due to the close association of the Ankyrin repeats in the crystal structure, Ank4 beads included bonds to its adjacent Ank3 beads. This was necessary due to the nature of the model building. It was not reasonable to assign Ank4 atoms to a single bead based on this structure and underscores a major difference in the potential of Ank4 movement versus the flexibly linked domain architecture of CD2 and CD3. Example movies of the course grained modeling are provided in the supporting information (Movies S3-S5). Three major conformations, which are denoted as "closed", "open" and "extended" were identified (Figure 3c). The "open" conformation bears the closest resemblance to the published SAXS structure of cpSRP43 [15]. The "closed" structure is characterized by a strong interaction between CD3 and Ank4, and the "extended" structure depicts a particularly large inter-domain separation between the Ank domains and CD3. The "closed" conformation was observed in both simulations (i.e. whether the peptide constraint between Ank4 and CD2 was applied or not, movies S4 and S5), suggesting that binding of cpSRP54 does not hinder the ability for CD3 to interact with Ank4. However, the relative orientation of the CD2 domain to the CD3 is affected by cpSRP54 binding.

Distances between the beads representing the residues in which cpSRP43 was labeled in the smFRET experiments were measured and plotted as a function of time. Figure 3d shows example traces highlighting transitions between the three dominant conformations. P1, which probes the distance between the C_δ carbons on the labeled residues in Ank2 and CD3, shows states with distances at ~42 Å, 65 Å and 88 Å. These three distances between the labeled residues in Ank2 and CD3 describe the major conformations that we classified as "closed", "open" and "extended", respectively. It must be noted here that the orange trace in Figure 3 does

start at a distance of 65 Å (open), but there is a rapid decrease to 42 Å (closed) within the first few bins, which is not easily visible on this timescale. P3 probes the distance between residues in Ank2 and CD2 with the three major conformations at ~40 Å, 50 Å and 55 Å inter-residue separation. P4 shows major conformations with distances between residues in CD2 and CD3 at ~22 Å, 30 Å and 46 Å, respectively. The simulations did not identify significant distance changes between the residues labeled in Ank2-Ank4 (P2) or CD1-Ank2 (P5), due to the limitations in bead connectivity described above. Multiple transitions between the three major conformations were not observed in a single trace, indicating that these states were long lived compared to the 50-100 μs simulation timescales. Therefore, in this study, we limit our discussion only to the fact that the SBCG model supports the *existence* of various states that underlie the structural dynamics observed in smFRET [39]without consideration of their relative *stabilities*. While the 3 main conformations identified from the modeling may lend support, at least in part, to the fact that low, mid and high FRET efficiencies might be observed due to the "closed", "open" and "extended" conformations, respectively, more refinements to the model are needed to make more robust conclusions. Future studies will examine the relative stabilities of the states in more detail and further refine the SBCG model toward this end.

Reducing the flexibility of cpSRP43 upon binding to cpSRP54 increases its affinity for the LHCP substrate

LHCP transport and integration into the thylakoid membrane is a multi-step process involving both sequential and concurrent protein-protein interactions between cpSRP43, cpSRP54, LHCP, cpFtsY, and Alb3. The formation of the ternary cpSRP43-cpSRP54-LHCP transit complex in the chloroplast stroma is the first key step in this process [11]. Cross-linking

experiments have shown that the N-terminal portion of TM3 in LHCP binds to cpSRP43, independent of the presence of cpSRP54 [17]. Our smFRET results showed that the regional flexibility of cpSRP43 is reduced, particularly between Ank2 and Ank4 [39]. Two hydrophobic grooves were identified in the crystal structure of cpSRP43 [17], one groove between Ank2 and Ank4 (Groove 1) and the other between CD1 and Ank1 (Groove 2), as shown in Figure 4a. Residues that are known to interact with both the L18 domain of LHCP and $cpSRP54_{pep}$ are in the Ank2-Ank4 region of the protein. MD simulations showed that the root-mean square fluctuations of the C_α carbons at the interaction sites of both L18 and $cpSRP54_{pep}$ in cpSRP43 reduced upon binding $cpSRP54_{pep}$ (Figure 4b). smFRET experiments showed that $cpSRP54_{pep}$ reduces the flexibility of cpSRP43 in the Ank2-Ank4 region in a similar way to the full-length cpSRP54 (Figure 4c) [39]. Figure 4c shows that $cpSRP54_{pep}$ changes the smFRET similar to the full-length cpSRP54, albeit to a lower extent. The combination of smFRET data and MD simulations led us to hypothesize that the affinity between cpSRP43 and the L18 domain of LHCP is increased by the reduced flexibility caused by cpSRP54 binding. To test this hypothesis, isothermal titration calorimetry (ITC) experiments were performed between cpSRP43 and L18 in the absence and presence of $cpSRP54_{pep}$. The ITC data for the binding of L18 to cpSRP43 in the absence and presence of $cpSRP54_{pep}$ is shown in Figure 4d. There is a ~3-fold increase in the affinity between cpSRP43 and L18 after $cpSRP54_{pep}$ is bound. Since $cpSRP54_{pep}$ changes the smFRET of cpSRP43 to a lower extent than full-length cpSRP54 (Figure 4c), it is possible that other residues of cpSRP54 are involved in interactions with cpSRP43 and the 3-fold increase in affinity may actually be an underestimation of the effect of cpSRP54 binding on the LHCP affinity. However, it must also be noted that this change in affinity is only to the L18 loop region of LHCP and other regions of LHCP may be involved in the interaction

with cpSRP43 and/or cpSRP54 which may serve to either increase or decrease this effect. Clearly, more work on measuring the affinity changes of the other domains is needed.

Functional implications for LHCP targeting

The central role of cpSRP43 in LHCP post-translational targeting has been highlighted by a range of studies. cpSRP43 alone, but not cpSRP54, is able to solubilize LHCP aggregates *in vitro* [15,21]. While the CD1 domain of cpSRP43 is not required for binding LHCP in the water soluble transit complex, it is required for LHCP integration and is involved in regulating the GTP hydrolysis rates of cpSRP54/cpFtsY [49]. Clearly, cpSRP43 is absolutely necessary for both forming the water soluble LHCP transit complex and for membrane integration. In fact, cpSRP43's ability to interact with both LHCP and the Alb3 integrase supports the operation of a less efficient LHCP targeting pathway that functions in the absence of cpSRP54 [19,21]. It was found that plants lacking both cpSRP54 and its membrane receptor cpFtsY recover to a great extent from the chlorotic phenotype observed when cpFtsY alone is missing [50]. Hence, cpFtsY is only required when LHCP is targeted by a cpSRP43/cpSRP54 dimeric complex. Therefore, the role of the two GTPases, cpSRP54 and cpFtsY, is to increase the efficiency of LHCP routing conducted by cpSRP43 and Alb3. The data presented in this study supports that one role of cpSRP54 is to promote LHCP binding to cpSRP43 by reducing inter-domain protein dynamics in cpSRP43. Two grooves were identified in the crystal structure of cpSRP43 [17]. The P2 and P5 smFRET proteins were designed to separately probe these two grooves (Figure 4a) [39]. Based on the smFRET results for P2, altering the structure of groove 1 in the region of cpSRP43 is a strong candidate for the source of improved binding of the LHCP substrate [39].

A strong interaction of CD3 with the Ank4 region was found by SBCG MD simulations to form a "closed" conformation. It is interesting to speculate that this interaction helps to effect a conformational change in groove 1. The interaction of cpSRP54 with cpSRP43 still allows the "closed", "open" and "extended" structures to be sampled, as evidenced from the wide smFRET histogram observed for P1 [39]. It has been postulated that a portion of the TM3 helix in LHCP also binds to cpSRP54 [17]. Since LHCP binding to cpSRP fails to occur when the L18 region of LHCP is mutated, LHCP binding to cpSRP is undoubtedly a multistep process initiated by L18 binding into groove 1 in which cpSRP54 helps to promote. Other portions of LHCP with groove 2 on cpSRP43 and cpSRP54 would follow, although the order of these interactions is uncertain. The continued flexibility of cpSRP43 after binding cpSRP54 may be necessary to allow LHCP to form these favorable interactions with both cpSRP43 and cpSRP54, extending an earlier model proposed by the Sinning group [17] to include structural dynamics in the formation of the transit complex. Finally, it is speculated that downstream interactions of cpSRP43 in the transit complex with Alb3 may be connected to inter-domain structural dynamics that enable both efficient binding of cpSRP43 to Alb3 and release of LHCP from cpSRP to Alb3 at the thylakoid membrane. The fact that cpSRP43 alone binds LHCP, albeit with a lower affinity and subsequently integrates it into the thylakoid membrane by a cpSRP54-independent pathway could suggest that cpSRP54 may have evolved in chloroplasts after cpSRP43 as a way to improve LHCP-binding to the SRP and subsequent integration efficiency. A combination of smFRET, MD simulations and thermodynamic measurements is a powerful approach that will help further identify the roles of protein dynamics in protein-targeting mechanisms.

Summary and Conclusions

We have combined MD simulations and ITC with the previous smFRET to relate the inter-domain structural dynamics of cpSRP43 to its function in post-translational targeting of LHCP. 3 major conformations were speculated – "open", "closed" and "extended" – that leads to cpSRP43 showing a high degree of flexibility across the whole protein. Certain regions showed decreased flexibility upon binding to its cpSRP binding partner, cpSRP54, particularly in the Ank2-Ank4 and CD2-CD3 regions [39]. We demonstrated that this decreased flexibility upon binding cpSRP54 is connected with an increased affinity of cpSRP43 to the L18 motif of LHCP. These results support a model in which changes in the flexibility of cpSRP43 are related to its ability to bind multiple substrates and promote its post-translational targeting function and may provide a clue that the evolutionary development of cpSRP54 in the chloroplast protein-targeting pathway may have come after the cpSRP43 pathway as a way to improve targeting efficiency.

References

1 Cline, K. & Dabney-Smith, C. Plastid protein import and sorting: different paths to the same compartments. *Curr Opin Plant Biol* **11**, 585-592, doi:10.1016/j.pbi.2008.10.008 (2008).

2 Cline, K. in *Light-Harvesting Antennas in Photosynthesis* Vol. 13 *Advances in Photosynthesis and Respiration* (eds Beverley Green & W.W. Parson) Ch. 12, 353-372 (Kluwer Academic Publishers, 2003).

3 Henry, R., Goforth, R. L., Schunemann, D., Ross E. Dalbey, C. M. K. & Fuyuhiko, T. in *The Enzymes* Ch. Volume 25, 493-521 (Academic Press, 2007).

4 Franklin, A. E. & Hoffman, N. E. Characterization of a chloroplast homologue of the 54-kDa subunit of the signal recognition particle. *J Biol Chem* **268**, 22175-22180 (1993).

5 Payan, L. A. & Cline, K. A stromal protein factor maintains the solubility and insertion competence of an imported thylakoid membrane protein. *J Cell Biol* **112**, 603-613 (1991).

6 Li, X., Henry, R., Yuan, J., Cline, K. & Hoffman, N. E. A chloroplast homologue of the signal recognition particle subunit SRP54 is involved in the posttranslational integration of a protein into thylakoid membranes. *Proc Natl Acad Sci U S A* **92**, 3789-3793 (1995).

7 Schünemann, D. *et al*. A novel signal recognition particle targets light-harvesting proteins to the thylakoid membranes. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 10312-10316 (1998).

8 BR, G. & AH, S. in *Molecular Genetics of Photosynthesis* (eds Andersson B, Salter AH, & Barber J) 75–103 (IRL Press, 1996).

9 Groves, M. R. *et al*. Functional Characterization of Recombinant Chloroplast Signal Recognition Particle. *Journal of Biological Chemistry* **276**, 27778-27786 (2001).

10 Yuan, J., Henry, R. & Cline, K. Stromal factor plays an essential role in protein integration into thylakoids that cannot be replaced by unfolding or by heat shock protein Hsp70. *Proc Natl Acad Sci U S A* **90**, 8552-8556 (1993).

11 Henry, R. L. SRP: adapting to life in the chloroplast. *Nat Struct Mol Biol* **17**, 676-677, doi:10.1038/nsmb0610-676 (2010).

12 DeLille, J. *et al*. A novel precursor recognition element facilitates posttranslational binding to the signal recognition particle in chloroplasts. *Proc Natl Acad Sci U S A* **97**, 1926-1931, doi:10.1073/pnas.030395197 (2000).

13 Tu, C. J., Peterson, E. C., Henry, R. & Hoffman, N. E. The L18 Domain of Light-harvesting Chlorophyll Proteins Binds to Chloroplast Signal Recognition Particle 43. *Journal of Biological Chemistry* **275**, 13187-13190 (2000).

14 Falk, S. & Sinning, I. cpSRP43 Is a Novel Chaperone Specific for Light-harvesting Chlorophyll a,b-binding Proteins. *Journal of Biological Chemistry* **285**, 21655-21661 (2010).

15 Jaru-Ampornpan, P. *et al*. ATP-independent reversal of a membrane protein aggregate by a chloroplast SRP subunit. *Nat Struct Mol Biol* **17**, 696-702 (2010).

16 Cain, P., Holdermann, I., Sinning, I., Johnson, A. E. & Robinson, C. Binding of chloroplast signal recognition particle to a thylakoid membrane protein substrate in aqueous solution and delineation of the cpSRP43-substrate interaction domain. *Biochem J* **437**, 149-155, doi:10.1042/BJ20110270 (2011).

17 Stengel, K. F. *et al*. Structural Basis for Specific Substrate Recognition by the Chloroplast Signal Recognition Particle Protein cpSRP43. *Science* **321**, 253-256 (2008).

18 Marty, N. J. *et al*. The Membrane-binding Motif of the Chloroplast Signal Recognition Particle Receptor (cpFtsY) Regulates GTPase Activity. *Journal of Biological Chemistry* **284**, 14891-14903 (2009).

19 Lewis, N. E. *et al*. A Dynamic cpSRP43-Albino3 Interaction Mediates Translocase Regulation of Chloroplast Signal Recognition Particle (cpSRP)-targeting Components. *Journal of Biological Chemistry* **285**, 34220-34230 (2010).

20 Dünschede, B., Bals, T., Funke, S. & Schünemann, D. Interaction Studies between the Chloroplast Signal Recognition Particle Subunit cpSRP43 and the Full-length Translocase Alb3 Reveal a Membrane-embedded Binding Region in Alb3 Protein. *J. Biol. Chem.* **286**, 35187-35195 (2011).

21 Falk, S., Ravaud, S., Koch, J. & Sinning, I. The C Terminus of the Alb3 Membrane Insertase Recruits cpSRP43 to the Thylakoid Membrane. *Journal of Biological Chemistry* **285**, 5954-5962 (2010).

22 Jia, Y. *et al*. Folding dynamics of single GCN-4 peptides by fluorescence resonant energy transfer confocal microscopy. *Chemical Physics* **247**, 69-83 (1999).

23 Talaga, D. S. *et al*. Dynamics and folding of single two-stranded coiled-coil peptides studied by fluorescent energy transfer confocal microscopy. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 13021-13026 (2000).

24 Schuler, B., Lipman, E. A. & Eaton, W. A. Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. *Nature* **419**, 743-747 (2002).

25 Kuzmenkina, E. V., Heyes, C. D. & Nienhaus, G. U. Single-molecule Foerster resonance energy transfer study of protein dynamics under denaturing conditions. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15471-15476 (2005).

26 Zhuang, X. *et al.* A single-molecule study of RNA catalysis and folding. *Science* **288**, 2048-2051 (2000).

27 Nishimura, S. Y. *et al.* Diffusion of lipid-like single-molecule fluorophores in the cell membrane. *J Phys Chem B* **110**, 8151-8157, doi:10.1021/jp0574145 (2006).

28 Cui, B. *et al.* One at a time, live tracking of NGF axonal transport using quantum dots. *Proc Natl Acad Sci U S A* **104**, 13666-13671, doi:10.1073/pnas.0706192104 (2007).

29 Gao, F., Mei, E., Lim, M. & Hochstrasser, R. M. Probing Lipid Vesicles by Bimolecular Association and Dissociation Trajectories of Single Molecules. *Journal of the American Chemical Society* **128**, 4814-4822 (2006).

30 Sharonov, A. & Hochstrasser, R. Wide-field subdiffraction imaging by accumulated binding of diffusing probes. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 18911-18916, doi:10.1073/pnas.0609643104 (2006).

31 Chen, P. *et al.* Tackling metal regulation and transport at the single-molecule level. *Nat Prod Rep* **27**, 757-767, doi:10.1039/b906691h (2010).

32 Dickson, R. M., Cubitt, A. B., Tsien, R. Y. & Moerner, W. E. On/off blinking and switching behaviour of single molecules of green fluorescent protein. *Nature* **388**, 3558-3358 (1997).

33 Lu, H. P., Xun, L. & Xie, X. S. Single Molecule Enzymatic Dynamics. *Science* **282**, 1877-1882 (1998).

34 Yang, H. *et al.* Protein conformational dynamics probed by single-molecule electron transfer. *Science* **302**, 262-266 (2010).

35 Ha, T. *et al.* Probing the interaction between two single molecules: fluorescence resonance energy transfer between a single donor and a single acceptor. *Proc Natl Acad Sci U S A* **93**, 6264-6268 (1996).

36 Roy, R., Hohng, S. & Ha, T. A practical guide to single-molecule FRET. *Nat Methods* **5**, 507-516 (2010).

37 Holdermann, I. *et al.* Chromodomains read the arginine code of post-translational targeting. *Nat Struct Mol Biol* **19**, 260-263 (2012).

38 Sivaraja, V. *et al.* Three-Dimensional Solution Structures of the Chromodomains of cpSRP43. *Journal of Biological Chemistry* **280**, 41465-41471 (2005).

39 Gao, F. *et al*. Regulation of Structural Dynamics within a Signal Recognition Particle Promotes Binding of Protein Targeting Substrates. *Journal of Biological Chemistry*, doi:10.1074/jbc.M114.624346 (2015).

40 Sivaraja, V. *et al*. Three-Dimensional Solution Structures of the Chromodomains of cpSRP43. *J. Biol. Chem.* **280**, 41465-41471, doi:10.1074/jbc.M507077200 (2005).

41 Jaru-Ampornpan, P. *et al*. ATP-independent reversal of a membrane protein aggregate by a chloroplast SRP subunit. *Nat. Struct. Mol. Biol.* **17**, 696-702, doi:10.1038/nsmb.1836 (2010).

42 Wriggers, W. Using Situs for the integration of multi-resolution structures. *Biophys Rev* **2**, 21-27, doi:10.1007/s12551-009-0026-3 (2010).

43 Pettersen, E. F. *et al*. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-1612, doi:10.1002/jcc.20084 (2004).

44 Phillips, J. C. *et al*. Scalable molecular dynamics with NAMD. *J Comput Chem* **26**, 1781-1802, doi:10.1002/jcc.20289 (2005).

45 Brooks, B. R. *et al*. CHARMM: the biomolecular simulation program. *J Comput Chem* **30**, 1545-1614, doi:10.1002/jcc.21287 (2009).

46 Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J Mol Graph* **14**, 33-38, 27-38 (1996).

47 Arkhipov, A., Freddolino, P. L. & Schulten, K. Stability and dynamics of virus capsids described by coarse-grained modeling. *Structure* **14**, 1767-1777, doi:10.1016/j.str.2006.10.003 (2006).

48 Scopes, R. K. Measurement of protein by spectrophotometry at 205 nm. *Anal Biochem* **59**, 277-282 (1974).

49 Goforth, R. L. *et al*. Regulation of the GTPase Cycle in Post-translational Signal Recognition Particle-based Protein Targeting Involves cpSRP43. *Journal of Biological Chemistry* **279**, 43077-43084 (2004).

50 Tzvetkova-Chevolleau, T. *et al*. Canonical Signal Recognition Particle Components Can Be Bypassed for Posttranslational Protein Targeting in Chloroplasts. *Plant Cell* **19**, 1635-1648 (2007).

Tables and Figures

Table 1: cpSRP43 mutants used in this study

Designation	Site Descriptions	Domains Labeled
P1	mcpSRP43 C179A R181C C301A R333C	Ank2 / CD3
P2	mcpSRP43 C179 V265C C301A	Ank2 / Ank4
P3	mcpSRP43 wild-type C179 C301	Ank2 / CD2
P4	mcpSRP43 C179A C301 R333C	CD2 / CD3
P4Δ	CD2CD3 C301 R333C	CD2 / CD3
P5	mcpSRP43 E89C C179 C301A	CD1 / Ank2

Figure 1

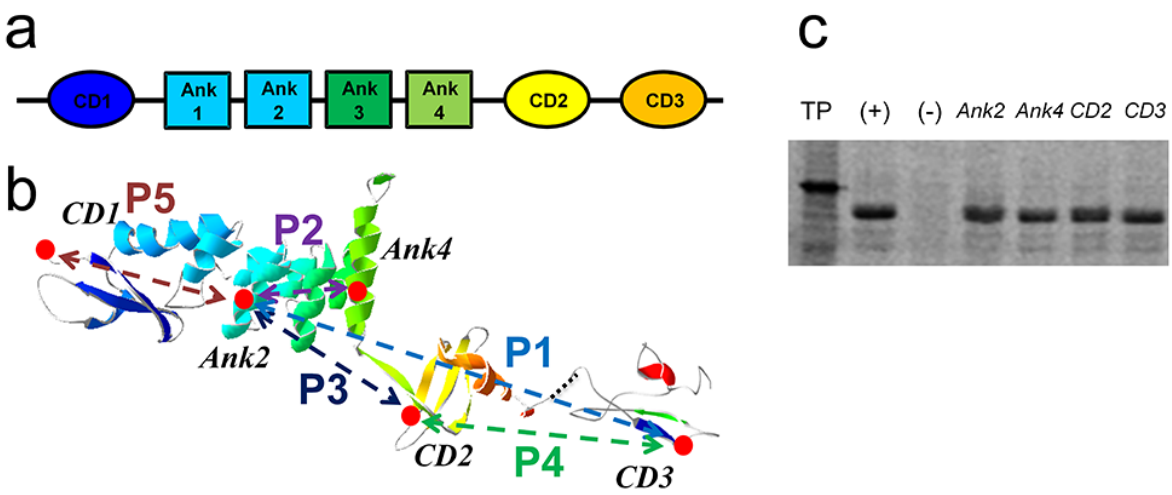

Figure 1. a) The arrangement of domains in cpSPR43 b) The amino acid positions used for fluorescence labeling highlighted in red for five double Cys cpSRP43 proteins, labeled P1-P5. The structure for the CD1-Ank1-4-CD2 region is taken from pdb code 3UI2, and the structure for CD3 is taken from pdb code 1X3P. c) LHCP integration assays confirming that fluorescently-labeled cpSRP43 are able to transport and integrate LHCP into thylakoid membranes. TP: translation product; (+) positive control with native cpSRP43/54; (-) negative control with no cpSPR43/54; (1) wt cpSRP54 + Ank2-labeled cpSRP43; (2) wt cpSRP54 + Ank4-labeled cpSRP43; (3) wt cpSRP54 + CD2 labeled cpSRP43; (4) wt cpSRP54 + CD3 labeled cpSRP43.

Figure 2

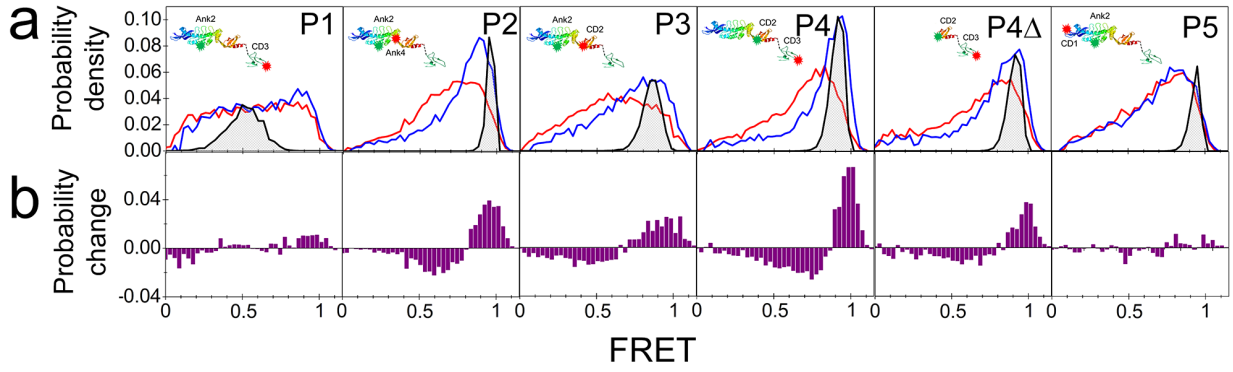

Figure 2. a) smFRET histograms of cpSRP43 proteins without (red) and with (blue) bound cpSRP54. The expected Poisson noise-limited smFRET histograms from the average FRET based on the SAXS structure are overlaid (gray). Insets are schematic representations of the labeled proteins. b) Calculated difference histograms of the cpSRP43 proteins after binding cpSRP54.

Figure 3

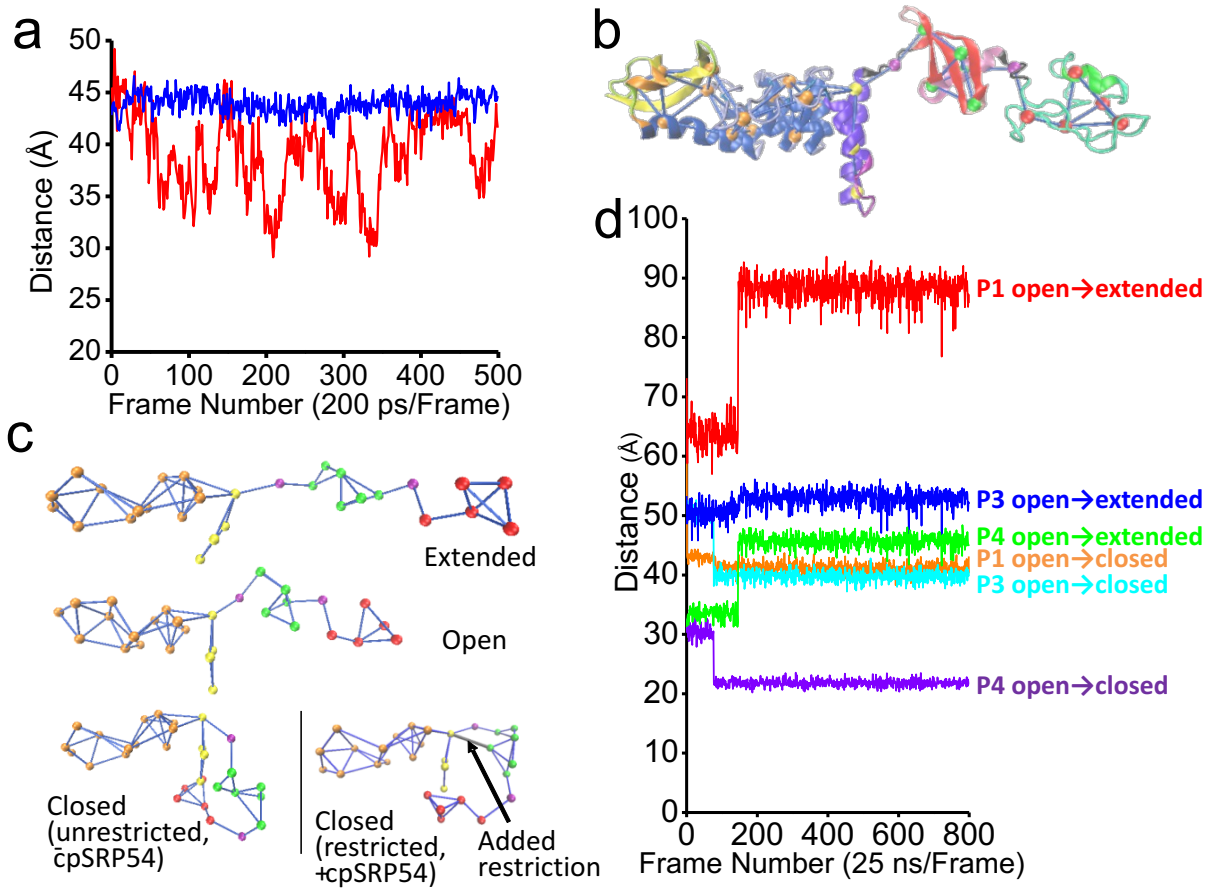

Figure 3. a) Distance between the labeled residues on the Ank2 and CD2 domains (P3) as a function of time from all-atom simulations in the absence (red) and presence (blue) of $cpSRP54_{pep}$. b) Overlay of the all-atom structure with the SBCG model (spheres). c) The 3 main conformations identified from the SBCG simulations, "extended", "open" and "closed. The "closed" structure for both the unrestricted (cpSRP54 unbound) and the restricted (cpSRP54 bound) simulations are shown. d) Distances between the labeled residues of cpSRP43 probed in P1 (Ank2-CD3), P3 (Ank2-CD2) and P4 (CD2-CD3) during open-to-extended or open-to-closed transitions.

Figure 4

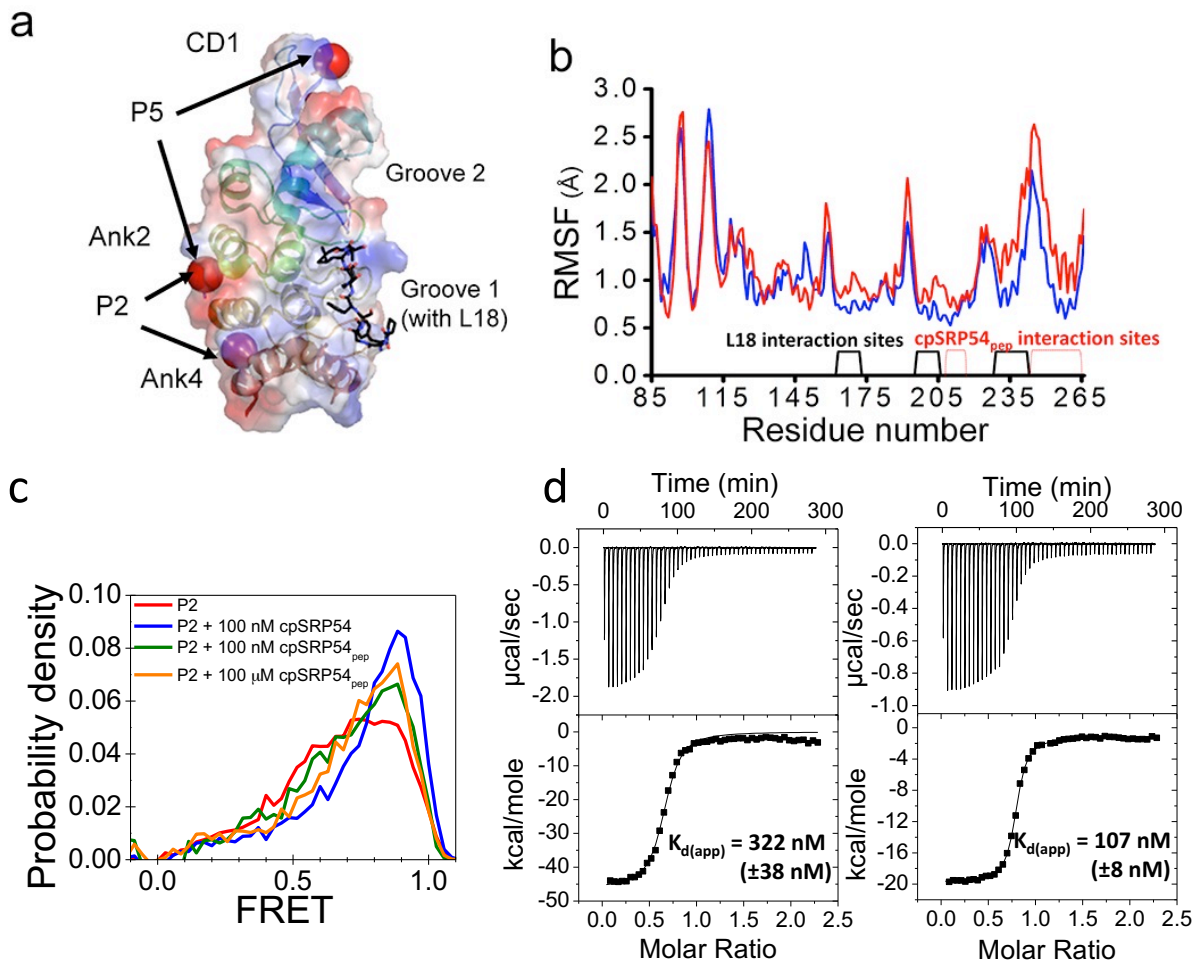

Figure 4. a) CD1-Ank4 region of cpSRP43 highlighting the relationship between the FRET labels in P2 and P5 with groove 1, to which the L18 motif binds, and groove 2. b) Root mean square fluctuations in the C_α carbons of residues in the Ank2-Ank4 region in the absence (red) and presence (blue) of $cpSRP54_{pep}$. The L18 and $cpSRP54_{pep}$ interaction residues on cpSRP43 are shown on the x-axis. c) smFRET histograms for P2 in the absence (red) and presence of different concentrations of $cpSRP54_{pep}$ (green, 100 nM; orange 100 μM). No significant difference between adding 100 nM or 100 μM $cpSRP54_{pep}$ was observed, suggesting that the change in the smFRET of P2 is already saturated at 100 nM $cpSRP54_{pep}$. For comparison, the smFRET of P2 + 100 nM cpSRP54 protein (blue) is shown. d) ITC data for the addition of L18 to cpSRP43 in the absence (left) and presence (right) of $cpSRP54_{pep}$.

III. Domain Organization in the 54kDa Subunit of the Chloroplast Signal Recognition Particle

Rory C. Henderson R.C.H[1], Feng Gao F.G.[1], Srinivas Jayanthi S.J.[1], Alicia Kight A.K.[2], Robyn L. Goforth R.L.G.[2], Colin D. Heyes C.D.H.[1], Ralph Henry R.H.[2], and Thallapuranam Krishnaswamy Suresh Kumar T.K.S.K.[1]

1. Department of Chemistry & Biochemistry 2. Department of Biological Sciences
University of Arkansas, Fayetteville, 72701

BJ, 2016 doi:10.1016/0005-2728(86)90197-0

Abstract

Chloroplast signal recognition particle (cpSRP) is a heterodimer composed of an evolutionarily conserved 54kDa GTPase (cpSRP54) and a unique 43kDa subunit (cpSRP43) responsible for delivering light harvesting chlorophyll binding proteins (LHCP) to the thylakoid membrane. While a nearly complete three-dimensional structure of cpSRP43 has been determined, no high-resolution structure is yet available for cpSRP54. In this study, we developed and examined an *in silico* three-dimensional model of the structure of cpSRP54 by homology modeling using cytosolic homologues. Model selection was guided by single molecule FRET experiments, which revealed the presence of at least two distinct conformations. Small angle x-ray scattering showed that the linking region between the GTPase (G-domain) and methionine rich (M-domain) domains, an M-domain loop, and the cpSRP43 binding c-terminal extension of cpSRP54 are predominantly disordered. Interestingly, the linker and loop segments were observed to play an important role in organizing the domain arrangement of cpSRP54. Further, deletion of the finger loop abolished loading of the cpSRP cargo, LHCP. These data highlight important structural dynamics relevant to cpSRP54's role in the post- and co-translational signaling processes.

Introduction

The signal recognition particle (SRP) protein sorting pathway is universally conserved and utilized in both eukaryotes and prokaryotes as a means to deliver newly synthesized polypeptides to their appropriate cellular locations.(1-3) This process is initiated in the cytosol via a co-translational mechanism whereby a conserved 54kDa SRP subunit (SRP54 in eukaryotes and Ffh in bacteria) binds to a nascent polypeptide chain as it emerges from the ribosome. In

eukaryotes, this binding event halts synthesis of the ribosome-associated nascent chain. The ribosome/nascent chain complex is then delivered to the endoplasmic reticulum membrane where SRP54 interacts with its receptor.(4) GTP hydrolysis by SRP54 and its receptor coordinate SRP release of the nascent polypeptide chain to a protein translocase. This is followed by recycling of SRP.(5-7) Protein synthesis is resumed and the nascent chain is either inserted into the membrane or transported across the lipid bilayer.(4)

Targeting to the thylakoid of a subset of integral membrane proteins is achieved via a novel post-translational targeting mechanism that relies on a heterodimeric signal recognition particle in chloroplasts (cpSRP).(8) Unlike cytosolic SRPs, the cpSRP dimer functions exclusively by a post-translational mechanism.(8-10) The only known substrates of the post-translational cpSRP targeting system are light harvesting chlorophyll-binding proteins (LHCPs), which bind chlorophyll-a/b and serve as the peripheral antennae for PS2 and PS1 in photosynthetic light capture.(4, 8)

The cpSRP heterodimer is composed of an SRP54 homolog, cpSRP54, and a unique 43kDa protein, cpSRP43.(8, 11) A cpSRP-LHCP targeting complex, termed transit complex, is formed in the stroma following LHCP import into the chloroplast.(11, 12) A critical binding site for LHCP occurs on cpSRP43 in a groove between its four ankyrin repeats that interact with an 18 amino acid segment (L18) of LHCP located between the second and third transmembrane (TM) helices (TM2 and TM3, respectively).(11-13) Based upon homology, cpSRP54 contains a four-helix bundle N-domain and five conserved GTPase elements that make up a G-domain, which together comprise the NG-domain. A c-terminal methionine rich M-domain lacks the RNA binding motif found in cytosolic SRPs but has retained a hydrophobic interaction site suitable for binding hydrophobic signal sequences. A flexible linker connects the G-domain and

M-domain. This segment is capable of forming a helix-turn-helix motif referred to as the greasy slide due to hydrophobic contacts between the helices. An additional flexible segment occurs between the first and second helices of the M-domain and is referred to as the finger loop. The linker and finger loop appear to facilitate signal sequence binding based upon crystal structure data.(14-16) The M-domain of cpSRP54 has been proposed to bind TM3 of LHCP via evidence of cross-linking between TM3 and cpSRP54 in elongation arrested ribosomes.(17) However, a cross-link was not observed between cpSRP54 and LHCP in cpSRP leaving some question as to whether cpSRP54 interacts with LHCP in the transit complex.(17, 18) Although cpSRP54 functions post-translationaly in the cpSRP system, approximately half of the cpSRP54 in stroma is bound to ribosomes and it is known that cpSRP54 is involved in co-translational targeting of some chloroplast synthesized proteins.(19-22) Interestingly, unlike its cytosolic counterpart, cpSRP54 does not interact with SRP RNA leaving some question as to how cpSRP54 retains these co-translational targeting abilities.(8)

Protein flexibility is known to play a major role in molecular recognition allowing single proteins to occupy multiple roles in a system.(23, 24) Flexibility via the G-domain to M-domain linker has been observed in the cytosolic SRP system using crystallography and FRET. These studies indicate large movements of the M-domain relative to the NG-domain of SRP54 throughout the signaling process.(4, 25, 26) Additionally, flexibility in the finger loop of the M-domain likely confers the ability to recognize various signal sequences at the ribosome.(4) Atomic level details of LHCP targeting by cpSRP, however, remain elusive. Recent work has demonstrated significant conformational dynamics in cpSRP43 that support downstream interactions.(27) However, little structural information exists to explain how cpSRP54 is able to function in the classical, co-translational signaling process as well as in the non-traditional, post-

translational targeting of LHCP. In this study, we examine the conformational states of cpSRP54 both before and after the formation of heterodimeric cpSRP. We also study the significance of these conformational states on the loading of LHCP to form the transit complex. Results from molecular modeling and FRET indicate that cpSRP54 exists in at least two distinct conformations. Additionally, mutational data and molecular dynamics simulations suggest cpSRP54 interacts with TM3 of LHCP and that, prior to LHCP binding, two intrinsically disordered segments of cpSRP54 coordinate to organize cpSRP into an LHCP binding competent state. Deletion of one of these segments, the M-domain finger loop, abolishes LHCP uptake by cpSRP.

Methods

Homology Modeling

The sequence for mature cpSRP54 (residues 76-564) was obtained from the UniProt database (accession number P37107; *Arabidopisis thaliana*).(28) As a first step in producing a three-dimensional model for cpSRP54, a gapped BLAST search of the Protein Data Bank (PDB) was performed for the full-length protein using the Swiss Model template identification tool and the NCBI BLAST tool using an e-value cut-off of $1x10^{-10}$ and the BLOSSUM64 matrix.(29-31) Since the M-domain shares less sequence similarity with cytosolic SRPs, a BLAST search against the PDB was performed for the M-domain residues 371-564 using an e-value cut-off of $1x10^{-10}$ and the BLOSSUM64 matrix. Of the sequences specified in the full-length search, four were identified in this search. These included cpSRP54/Ffh from *Canis familiaris* (PDB ID 2J37, chain W), *Sulfolobus solfataricus* (*S. sol. SRP54*) (PDB ID 1QZW, chain A), *Thermus aquaticus* (*T. aq.* Ffh) (PDB ID 2FFH, chain A), and *Escherichia coli* (E. col. Ffh) (PDB ID 2XXA, chain

A).(14, 16, 32, 33) A multiple sequence alignment of these four structures and cpSRP54 was performed using ClustalW2 and Clustal Omega.(34, 35) The ClustalW2 alignment was used in Modeller v9.12 to produce a multi-template 3D structure of cpSRP54.(36) In order to evaluate model quality and to determine the effect(s) of individual sequence similarity on the final model, I-TASSER was used to produce additional 3D structures.(37-39) Additionally, Ffh from *Methanocaldococcus jannaschii* (M. jan. SRP54) (PDB ID 3NDB, chain B) was submitted as a control template for cpSRP54 to I-TASSER to assess whether differences would appear with a reduction in sequence similarity to the M-domain.(15) I-TASSER was used to produce three-dimensional structures using each template individually without alignment.

Each structural model was evaluated visually using PyMOL and submitted to the Swiss-Model structure assessment tool for structure quality validation.(40) Specifically, Anolea, Gromos, QMEAN and PROCHECK were used in the validation.(41-44) Template models were also submitted for use in determining the ideal template for modeling. Structural alignments were performed using PyMOL and MultiProt.(45)

Molecular Dynamics Simulation

The *T. aq*. Ffh crystal structure 2FFH and the cpSRP54 structure modeled from 2FFH were selected for simulation. The highest scoring I-TASSER model for cpSRP54 (1.40 C-score) was used. The crystal structure for 2FFH chain A was prepared from the downloaded pdb file. The NG to M linker resides ^{308}G-K^{318} were missing in the structure due to weak electron density. These residues were added using the SYBYL-X suite.(46) Since the C-terminal residues of cpSRP54 were modeled using limited data, and since most cytosolic forms do not contain this C-terminal extension, the segment was removed for simulation. Additionally, the intrinsic disorder

predictor, PONDR, indicates this segment is likely disordered (Suppl. Fig. 3), a prediction supported by an instability index of 47 as calculated using the ProtParam module in Expasy and confirmed via small angle x-ray scattering (SAXS). (47-50)

To prepare the structures for production runs, the system was subjected to several refinement steps to relax the backbone, side chains, surrounding TIP3 water molecules, and Na^+ and Cl^- ions. The final system sizes were 52,586 atoms and 53,516 atoms for cpSRP54 and 2FFH respectively. Minimization and equilibration of the system was carried out in the NPT ensemble using NAMD 2.9 and the CHARMM27 or CHARMM36 forcefields.(51-53) The protein-solvent system was equilibrated as described previously.(27) In order to monitor the stability of the proteins and to evaluate structurally variable regions, simulations of 100ns using a 2fs time step were performed. Visual molecular dynamics (VMD) and tools therein were used for preliminary analysis and visualization.(54)

Steered molecular dynamics

SMD simulations were performed using constructs of 2FFH and the cpSRP54 model built using 2FFH as a template. The G-M linking regions (2FFH residues ^{293}G-L^{320} and cpSRP54 residues ^{296}G-F^{328}) were removed from each protein. The direction for pulling was determined as the vector connecting the NG and M-domain's centers of mass. The system was prepared as above leaving 25 Å space between the M-domain and the solvent box edge in the direction of pulling. The final system sizes were 61,460 atoms and 62,317 atoms for cpSRP54 and 2FFH respectively. The 2FFH L367E and cpSRP54 E379L mutant structures were prepared using the VMD Mutator plugin. Each protein was equilibrated for 50 nano second (ns) using a time-step of 2 femto second (fs) with the backbone of the NG-domains restrained with a force constant of 1

kcal/(mol $Å^2$). Constant velocity pulling was employed using a spring constant of 7 kcal/(mol $Å^2$) pulling from the center of mass of the M-domain while maintaining the constraints on the NG-domain's backbone atoms. A pulling speed of $1x10^{-6}$ Å /step for a total distance of 18 Å using a 1fs time-step. These conditions were selected after simulating several pulling conditions. Fifteen pulling simulations were performed and averaged for each construct with all other simulation conditions identical to those above. A pseudo-random number seed was used in each to ensure that no two simulations were identical.

Bioinformatics

The G-domain to M-domain interaction site residues from T. aq. Ffh were compared to a set of SRP54/Ffh/cpSRP54 proteins to examine the degree of conservation for this site. An additional, larger data set was used to examine the rate of mutation of the flexible G-domain to M-domain linker and the M-domain finger loop to determine whether these segments undergo an accelerated mutation rate compared to the rigid structural elements. The method used in these investigations is described in detail in the Supporting Material.

Cloning of cpSRP proteins

Primer-encoded Asn point mutations were introduced by polymerase chain reaction (PCR) into the mature cpSRP54 coding sequence (starting with amino acid sequence MAFGQL and including a c-terminal 6 HIS tag) at positions V339 (QTRAVAKMG) and L370 (AEKNLLVME) to make $cpSRP54_{V339N\ L370N}$. A loop deletion mutant, $cpSRP54_{\Delta 345\text{-}359}$, was made by splicing via overlap extension (SOE) using internal overlapping primers to remove 15 amino acids (MTRVLGMIPGMGKVS). The resulting PCR products were cloned into pGEM-4Z

(Promega) using KpnI and HindIII sites for *in vitro* transcription/translation. Cloning of a Cys replacement mutant for *E. coli* expression and labeling for single molecule FRET studies was produced by first using PCR to replace Cys residues in mature sized cpSRP54 with Ala. The resulting cysteine-less construct (cpSRP54his-cys) served as a template to introduce Cys residues at position 54 and 407 resulting in $cpSRP54_{S54C/K407C}$. The resulting construct was then cloned into pPROLar.A122 (Clontech) using KpnI and HindIII and transformed into BL21 Star (Life Technologies). Both $cpSRP54_{S54C/K407C}$ and cpSRP54 M-domain were produced as previously described.(55) All constructs were sequence verified by Molecular Resource Laboratory, University of Arkansas for Medical Sciences, Little Rock.

Co-precipitation assays

A protein binding assay was performed as described in Goforth et al. using 100 pmol of GST-cpSRP43 and 35S-labeled mcpSRP54 mutant produced by *in vitro* transcription/translation.(55)

Transit Complex Formation

Mature sized cpSRP54 or mutants of cpSRP54 described above were examined for their ability to support transit complex formation as in Goforth et al. with the following changes.(55) Twenty-five picomoles of recombinant cpSRP43 was added to *in vitro* translated cpSRP54 or cpSRP54 mutant (indicated in the figure legend) along with 35S-labeled *in vitro* translated pLHCP or pLHCP ΔTM3, both of which are described in Delille et al.(13) One third of the assay was examined by native-PAGE as described in Goforth et al.(55)

Single Molecule FRET

A cpSRP54 double Cys mutant ($cpSRP54_{S54C/K407C}$) was site-specifically labeled using Alexa488 C5 maleimide and Alexa594 C5 maleimide (Life Technologies), and the smFRET experiments were performed in HKMK buffer (10 mM HEPES, 10 mM $MgCl_2$, and 100 mM KCl, pH 8.0) and carried out with a MicroTime 200 microscope controlled by SymPhoTime software, collecting fluorescence photons on a photon-by-photon basis, as previously described.(27) Briefly, a water immersion objective with an N.A. value of 1.3 was used for efficient collection of single molecule fluorescence. A 100 μm pinhole in the fluorescence detection path served to suppress out of focus fluorescence. Excitation was initiated using a ~45 μw laser at 485 nm operating in continuous wave (cw) mode. A dichroic mirror FF562-Di03 served to split Alexa488 and Alexa594 fluorescence signal onto two Single Photon Avalanche Diode Detectors (SPAD, MPI, Microphotonic devices, Bolano, Italy). High quality fluorescence filters 520/35 and 620/60 were placed in front of each SPAD detector respectively for collecting the Alexa488 and Alexa594 fluorescence. The donor to acceptor leak-through was ~6% with ɣ of 0.5. ɣ is used to balance the donor and acceptor signals and was determined by measuring the fluorescence intensities of Alexa488 solution in the donor channel and of Alexa594 solution in the acceptor channel having the same optical density at 485 nm, using the same experimental set-up as for smFRET experiments. For smFRET experiments, the fluorescently labeled cpSRP54 protein concentration was ~150 pM. The femtoliter confocal laser focus was placed ~30 μm above a #1 glass coverslip surface and ~20 minute long traces showing bursts of diffusing single cpSRP54 proteins were collected. The donor and acceptor fluorescence traces were binned at 1-ms time-resolution and analyzed using a homemade Matlab program to generate smFRET histograms. The single protein bursts were selected from the traces only when the total burst intensity met a

threshold of 10 counts/ms, then the FRET values were calculated using $E=I_A/(I_A+\gamma I_D)$, with γ of 0.5. The smFRET histogram was then obtained by compiling the FRET values for all the detected bursts or molecules. Control experiments are described in the supplemental section (Figs. S6-S9).

Small Angle X-Ray Scattering

SAXS data for the M-domain of cpSRP54 was collected at the Cornell High Energy Synchrotron Source (CHESS) as described at the G1 station.(56) A concentration series from 1mg/ml to 5mg/ml was collected using ten one-second exposures each to a q range of 0.5 $Å^{-1}$. Buffer scattering data was collected before and after exposures of the protein concentration series. Data was processed using the RAW software and visualized in the ATSAS package's Primus. A slight concentration dependence was observed in addition to a shift at wider angles in the curve at concentrations above 3mg/ml. Only the data for 1-3mg/ml were used in Primus for extrapolation to zero concentration. This curve was used for analysis downstream. Structural fitting was performed using the Ensemble Optimization Method (EOM) as described.(57, 58) Briefly, randomly oriented c-α traces are generated to represent intrinsically disordered segments of the protein. A large pool of these structures was generated (10,000). This pool was then used to determine a minimum set of structures from the pool whose combined theoretical scattering curves were capable of fitting the experimental SAXS curve. This is accomplished via a genetic algorithm. Here, the sequence for the M-domain with the added histidine tag was used in the pool generation. The folded core was defined as α–M1 residues F328-K342 and α–M2 to α–M4 residues K369-M427 built from the 2FFH template model. Ten separate runs of the genetic

algorithm (100 iterations per run) were performed and compiled to determine the size distribution of the protein.

Results

Orientation of the M-domain varies in different SRP54s/Ffhs

In order to begin probing the structure of cpSRP54 and guide single molecule experiments a series of homology models were built. Crystal structures for cytosolic SRP54/Ffh proteins all feature a 4-helix bundle N-domain, an α/β GTPase G-domain that binds Mg^{2+} and GTP (collectively referred to together as the NG-domain), and a methionine rich M-domain (Fig. 1). The M-domain is connected to the G-domain via a flexible linker that allows the M-domain to move relative to the NG-domain. Indeed, no two crystal structures for SRP54/Ffh result in a similar positioning of the M-domain and therefore homology models of cpSRP54 position the M-domain in differing positions depending on the template used (Figure 1). Sequence similarity suggests the *T. aq* structure should most closely represent the free cpSRP54 structure (Table S1).

An important feature of the homology models is that, regardless of the starting template, the intra-domain backbone coordinates are quite similar as determined using PyMOL alignments, which is further supported by the high degree of similarity between templates observed from MultiProt data. Displacement of the first N-domain helix in 2XXA resulted in fewer aligned residues (Supplementary Table-2). Exclusion of the 2XXA structure from the N-domain alignment yields an RMSD of 1.40Å for 67 residues (77.0%). The discrepancy between the NG-domain's alignment and the individual N and G alignments is explained by the differences in orientation of the N-domain to the G-domain in each structure. The individual domain similarity is important since large differences in domain positioning exist in the templates. Ultimately, a

structural model(s) must be selected to represent the structure for cpSRP54 and internal domain differences would exacerbate the problem of model selection. The model quality statistics indicate that each model is of similar quality in terms of Ramachandran statistics and the scoring function QMEAN, as well as the overall z-scores for the models *versus* the Swiss Model database statistics for available crystal structures (Supplementary Table-1).

cpSRP54 exists in multiple conformational states

As there is significant variability in the M-domain positioning in the homology models we next used single molecule FRET experiments to investigate the positing of the M-domain relative the the NG-domain in cpSRP54. In *E. col*. Ffh, FRET experiments have revealed that in solution the M-domain occupies a position consistent with the *T. aq*. *Ffh* crystal structure.(14, 26) Since cpSRP54 shares significant homology with bacterial Ffh, we placed dyes in similar positions to those used in the *E. col*. FRET experiment, anticipating a similar positioning. The mutants produced were S54C in the N-domain and K407C in the M-domain of cpSRP54 ($cpSRP54_{S54C/K407C}$). The peak in the smFRET histogram is wider than expected from a single state as discussed previously(27), indicating at least two conformational states are present (Fig. 2).

Next, to learn more about the character of the interaction holding the M-domain in place, we used a KCl titration to determine whether the interaction(s) could be affected by ionic strength. A titration from 0mM to 1M additional KCl was performed. The results of this titration reveal that the high FRET peak observed at ~0.90 FRET, corresponding to a distance of ~42Å, decreases discernibly as the ionic strength is increased, indicating a significant electrostatic component to the interaction (Fig. 2A). The increase in FRET efficiency centered at ~0.54 FRET

becomes more predominant at high ionic strength, suggesting that the interaction holding the domains together for this state is predominantly hydrophobic. The ratio of the two FRET peaks shows a saturation with increasing KCl concentration, indicating that no further increase in ionic strength would alter the orientation of the domains (Fig. 2C).

The cytosolic SRP54/Ffh undergoes gross domain movement as various interaction partners such as SRP RNA and ribosomal proteins become involved. Therefore, we asked what affect cpSRP43 might have on the positioning of the M-domain. The results indicate cpSRP54 interaction with cpSRP43 alters the domain organization compared to cpSRP54 alone (Fig. 2D & 2E). A new low FRET value of ~0.35 is observed when cpSRP43 binds to cpSRP54 which indicates that the relative position of the M-domain is similar, but not identical to that observed at high ionic strength.

It is important to note that transitions occurring on the diffusion timescale in the smFRET experiments will result in broadening of the FRET peaks toward a weighted population average between the peaks. The rate of transition between the peaks can be determined using probability distribution analysis method.(59) Such fitting of the data is beyond the scope of this study, but will be explored further in a forthcoming publication.

Differences in the M-domain to NG-domain interaction exist between cpSRP54 and Thermus aquaticus Ffh

We next examined the residue specific differences between the *T. aq*. Ffh crystal structure and the cpSRP54 homology model M-domain to NG-domain interaction interface since smFRET experiment's 0.90 FRET efficiency gives a distance consistent with the *T. aq*. Ffh crystal structure. Simulations on the cpSRP54 model and *T. aq*. Ffh (PDB ID) 2FFH were

performed to examine the M-NG-domain binding interface. Critical M-domain to NG binding residues were selected from the *T. aq*. Ffh simulations since this interface has experimental support from *E. coli*. FRET experiments (Fig. S5). Specifically, cross-linkers were placed at positions on the G-domain and M-domain in positions determined using the T. aq Ffh crystal structure and were found to produce a cross-link. Further, FRET data indicated the distance between the N-domain and M-domain was consistent with this structure. Finally, a relatively strong interaction was observed when isolated M-domain was titrated into NG-domain. (60) In *T. aq*. Ffh, a large contact surface between the M-domain and NG-domains is characterized by electrostatic, hydrogen bonding and hydrophobic interactions. Details of the residues involved are provided in the supplemental section.

The previous *E. coli*. FRET experiments were used to determine the interaction strength via titration of isolated M-domain into isolated NG-domain. A Kd of ~62nM was obtained indicating relatively tight binding.(60) As the cpSRP54 smFRET experiments indicated rapid transitions between M-domain coupled states, we expected cpSRP54 M-domain to NG-domain interaction strength to be markedly weaker. Using the *T. aq*. Ffh crystal structure as a proxy for *E. coli* Ffh, we employed steered molecular dynamics (SMD) simulations to examine the relative strength of M-NG binding. The results indicate that the M-domain of the *T. aq*. Ffh is tightly bound to the NG domains, requiring ~45kcal/mol work to achieve decoupling (Fig. 4b). As expected, the binding of the M-domain of cpSRP54 is much weaker than that of 2FFH, requiring only ~17kcal/mol work.

We next used sequence analysis to determine first whether differences in the 23 important residues could explain differences in the free SRP54/Ffh crystal structures and second to determine whether sufficient conservation exists between cpSRP54 and *T. aq*. Ffh that the

0.90 FRET state could be reasonably expected to occupy the same position as *T. aq*. Ffh in cpSRP54. Sequence analysis indicates the 23 residues involved in the M-domain to NG-domain interaction site tend to be conserved in bacterial Ffh and cpSRP54. However, significantly less conservation is observed in archaeal and eukaryotic SRP54 (Fig. S2 in the supporting material). This seems reasonable in view of available crystal structures that show a large degree of variability in the positioning of the M-domain, even in structures that contain no other binding partners such as SRP RNA or the ribosome.(14-16, 32, 33, 61-65) Indeed, little conservation is observed overall for surface residues in SRP54s, aside from the GTPase region (Fig. S1 in the supporting material). The sequence data is consistent with both the crystal structure data for SRP54s/Ffhs as well as the smFRET data for cpSRP54. Further, conservation between the *T. aq*. Ffh 23 residues and the cpSRP54 23 residues suggest the interaction could be retained (Fig. S4 in the supporting material).

Interaction between the intrinsically disordered M-domain linker and finger loop in cpSRP54 is crucial for domain organization

In light of the apparent transitions observed in the smFRET experiments we attempted to determine a possible mechanism by which this could be accomplished. Several iterations of all-atom molecular dynamics simulation were run to examine interplay between the G-domain to M-domain linker and the M-domain finger loop as these play a significant role in the interaction observed in the *S. sol*. SRP54 crystal structure. Results of the simulations revealed interplay between the linker and finger loop (Fig. 4c). Indeed, interaction of the G-domain to M-domain Linker with the M-domain finger loop occurred in several simulations during model refinement. Herein, two particular interaction events are described in detail.

Firstly, the interaction between the G-domain to M-domain linker and α–M1, called the greasy slide (GL), forms in a manner similar to the interaction observed in the crystal structure of *M. jan*. SRP54 (3NDB).(15) This occurs in both the simulations of cpSRP54. Details of the residues involved are provided in the supplemental section. In general, it is known that intrinsically disordered segments tend to undergo an accelerated rate of mutation.(66) If the linker and finger loop were structural determinants of the heterogeneity in the M-domain positioning then they would be expected to undergo increased rates of mutation. Indeed, estimates of the rate of mutation in these two segments indicate the linker mutation rates are higher in archaea and eukaryota whereas mutations rates are high in all domains of life in the finger loop (Fig. S4). This is discussed further in the supplemental section.

The second simulation demonstrates the ability of the finger loop and G-domain to M-domain linker to coordinate shuttling of the M-domain between conformations. In the initial moments of the simulation, the M-domain becomes decoupled from the NG-domain due to poorly formed M-domain to G-domain contacts (Fig. 4c). The linker helix, however, quickly forms the GL with α-M1. Specifically, L319, I323 and F328 form close contacts in addition to K322 and D332 forming an electrostatic interaction. In this case, the core with the finger loop does not form. Instead the finger loop remains open, in a conformation reminiscent of that observed in a signal sequence bound crystal structure of SRP54 from *Sulfobus solfataricus* (PDB ID 3KL4).(15) While the new positioning of the M-domain relative to the NG-domains is not identical to the free *S. sol*. SRP54 structure (PDB ID 1QZX), the overall fold is quite similar.

In order for the G-domain and M-domain linker and the finger loop to play a significant role in the M-domain positioning it is necessary that they possess significant disorder. Indeed, the cpSRP54 M-domain was predicted to contain several disordered segments (Fig. S3). These

segments included the linker between the G-domain and M-domain, the finger loop and the C-terminal extension. We therefore used an isolated M-domain construct for SAXS experiments the contains the G-domain to M-domain linker thorugh the cpSRP54 M-domain C-terminal extension. The results reveal the linker, finger loop, and C-terminal extension are indeed disordered with a fit χ^2 of 1.09 (Fig. S10) to the extrapolated data. The size distribution is wide and supports an extended structure with large-scale disorder (Figure 4a). However, the distribution favors a more closed structure. A review of the structures indicates the primary driver of extension in the closed structure is the C-terminal extension. This is remarkably consistent with a closed state between the linker and the finger loop and is therefore supportive of an interaction between them.

Mutation in the M-domain and deletion of the finger Loop hinder substrate binding

Considering the observations made in the smFRET experiments and the observation from the SAXS experiment that the M-domain C-terminal extension is highly disordered we hypothesized that the M-domain may be positioned for interaction with TM3 of LHCP. TM3 exits cpSRP43 on the same side the M-domain is predicted to reside based upon the positioning of the M-domain C-terminal extension peptide in a cpSRP43 crystal structure (PDB ID 3UI2).(67) Previously, crosslinking experiments were unable to determine whether cpSRP54 interacts with TM3 when bound to cpSRP43 although it had been observed to cross-link with TM3 in elongation arrested ribosomes.(17, 18) However, cross-linkers placed at hydrophobic positions that the M-domain would need to interact with may have eliminated the ability of the M-domain to interact with TM3 and therefore prevent cross-linking. Although transit complex formation was nevertheless possible through cpSRP43 interaction, it remains possible cpSRP54

interacts with TM3 and that this interaction allows cpSRP54 to help regulate integration. As the M-domain in cytosolic SRP54/Ffh binds hydrophobic signal sequences and TM3 is significantly hydrophobic, we hypothesized a reduction in the hydrophobicity of the M-domain binding pocket would have a significant impact on M-domain interaction with TM3 if it occurs. If TM3 binding occurs, it would likely have a dramatic affect on transit complex formation. To verify this hypothesis cpSRP54 was produced in which two hydrophobic residues in the M-domain signal sequence binding groove, Val339 and Leu370, were substituted with Asn. This M-domain mutant ($cpSRP54_{V339N/L370N}$) was produced by *in vitro* transcription/translation and compared with similarly produced wild type cpSRP54 for the ability to bind cpSRP43 (Figure 5a) and the ability to promote transit complex formation when mixed with recombinant cpSRP43 and *in vitro* translated radiolabeled LCHP (Figure 5b). Co-precipitation assays with GST-43 and GST alone demonstrate that radiolabeled $cpSRP54_{V339N/L370N}$ is efficiently co-precipitated by glutathione sepharose only when GST is fused to cpSRP43, which demonstrates that the M-domain point mutations have no influence on cpSRP43 binding. In contrast, when transit complex formation was examined using native-PAGE, $cpSRP54_{V339N/L370N}$ exhibited a severe deficiency in the ability to form transit complex. Next, we reasoned if TM3 must form an interaction with cpSRP54 in the loading step, then deletion of TM3 would recover LHCP loading. Indeed, $cpSRP54_{V339N/L370N}$ exhibited a greater ability to support transit complex formation when LHCP was replaced with pLHCPΔTM3, which lacks TM3. This data suggests that an interaction of TM3 with the cpSRP54 M-domain has a significant impact on the formation of cpSRP-LHCP transit complex.

Since molecular dynamics simulations suggest a potential mechanism by which M-domain shuttling may occur through interaction between the G-domain to M-domain linker and the finger loop, we next removed the finger loop (residues 345-359; $cpSRP54_{\Delta 345\text{-}359}$) from the M-

domain. If the finger-loop plays a major role in transit complex formation beyond facilitating interaction of the M-domain with TM3 then the formation of transit complex would likely be affected to a greater degree than the double mutant cpSRP54$_{V339N/L370N}$. That is, if the finger loop is also critical for orienting the NG-domain and M-domain, these results would differ significantly. Figure 5a shows that while the loop deletion did not affect cpSRP43 binding, transit complex formation was completely abolished and no recovery of transit complex formation was observed when TM3 was removed from the pLHCP targeting substrate as in the cpSRP54$_{V339N/L370N}$ double mutant (Fig. 5b). These results suggest that the finger loop is likely responsible for more than hydrophobic peptide binding, consistent with a role for the linker-loop interaction.

Discussion

The current study, along with recent findings for cpSRP43, reveals the cpSRP system is characterized by a large amount of structural heterogeneity.(27) While cpSRP43 prepares an energetic funnel favoring LHCP interactions downstream from cpSRP formation, cpSRP54 appears to use various domain orientations to allow it to act in two different targeting systems. Indeed, there are significant differences between the traditional, co-translational signaling pathway of SRP54/Ffh *versus* the cpSRP54 post-translational targeting pathway. Our smFRET data indicates that at least two bound orientations exist for the M-domain relative to the NG-domains in cpSRP54. In each of the three free SRP54s/Ffh crystal structure templates available for cpSRP54 homology modeling, the M-domain adopts different positions relative to the NG-domains. (14, 33, 61) In a structure from *S. sol.* SRP54, the M-domain occupies a state stabilized by the "greasy slide" while a structure from *T. aq.* Ffh occupies a state in which the M-domain

interacts directly with the NG-domain.(14, 33) The *P. fur*. SRP54 structure contains no apparent contact between the M-domain and NG-domains that would likely render the molecule quite flexible in solution.

The 0.90 FRET peak corresponds well with the cpSRP54 model built from the *T. aq*. Ffh crystal structure after simulation with an M-N-domain distance of ~43Å. In the *S. sol*. SRP54 structure, the M-domain is positioned near the N-domain, forming a hydrophobic interaction between the G-M-domain linker and α-M1. The positions between the FRET α-carbons from cpSRP54 modeled from *S. sol*. SRP54 is ~52 Å (~0.70 FRET) which is which is inconsistent with the apparent 0.54 FRET peak (Fig. 2A). However, the broad shoulder at lower FRET is suggestive of an additional state or states. A change in the rate of exchange between a 0.70 or 0.90 FRET state with a lower FRET state centered at ~0.35 would lead to an increase in the 0.54 FRET region as increasing ionic strength shifted the equilibrium toward the 0.70 FRET state. The 0.35 FRET efficiency corresponds to a distance of ~66 Å, which is only possible for a decoupled M-domain based upon the G-domain to M-domain linker distance.

Using the results of the molecular dynamics simulations and sequence data, we propose that the G-domain to M-domain linker interacts with the finger loop to shuttle the M-domain between bound states through the decoupled state observed in the smFRET. Indeed, our results show that deletion of the finger loop has a dramatic affect on the loading of the LHCP cargo. Interestingly, a similar finger loop deletion in a cytosolic SRP54 efficiently bound targeting cargo, but was hindered in SRP RNA stimulation of GTP hydrolysis and cargo handover.(68) This difference suggests that complex assembly in the cpSRP system differs significantly from the signal peptide binding of the co-translational SRP system.

In conclusion, we hypothesize the significance of this orientational plasticity is to overcome the necessity of SRP RNA for separation of the M-domain from the NG-domain and to allow the cpSRP54 M-domain to scan for TM3. The mutational data support our notion that the M-domain interacts with TM3 and that the finger loop of the M-domain is involved in more than facilitating the interaction between cpSRP54 and LHCP. The loss of LHCP binding in the absence of the finger loop may be related to the heterogeneity in M-domain positioning observed in the smFRET experiments in which cpSRP54 is bound to cpSRP43. It may be necessary that the M-domain of cpSRP54 remains flexible when bound to cpSRP43 in order to facilitate LHCP binding, a process likely facilitated by the finger loop. It remains uncertain at this point whether binding of cpSRP43 to cpSRP54 or binding of cpSRP54 to the ribosome requires one of the cpSRP54 conformations observed here, or if the ability of cpSRP54 to transition between bound and unbound M-NG states is responsible for facilitating these interactions. Further, the cpSRP structure in its entirety remains to be experimentally determined, as does the structure of the transit complex. Nevertheless, it is clear that domain heterogeneity of both cpSRP proteins plays a major role in the targeting process.

Author Contributions

R.C.H. designed and performed computational research. R.C.H., S.J. and T.K.S.K. analyzed computational data. R.C.H and F.G. performed FRET experiments. R.C.H, F.G., C.D.H. and T.K.S.K. designed and analyzed FRET experiments. R.C.H. performed SAXS experiments. R.C.H. and T.K.S.K. designed and analyzed SAXS experiments. A.K. performed protein assay experiments. A.K., R.L.G and R.H. designed and analyzed protein assay experiments. R.C.H. prepared and wrote the manuscript. All authors edited and approved the manuscript.

Acknowledgements

This research was generously supported by the Department of Energy (grant number DE-FG02-01ER15161) (to R.L.H. and T.K.S.K.), the National Institutes of Health/National Cancer Institute (NIH/NCI) (1 RO1 CA 172631) and the NIH through the COBRE program (P30 GM103450), and the Arkansas Biosciences Institute. MD simulations are supported in part by the National Science Foundation through grants MRI #0722625 (Star of Arkansas), MRI-R2 #0959124 (Razor), ARI #0963249, #0918970 (CI-TRAIN), and a grant from the Arkansas Science and Technology Authority, with resources managed by the Arkansas High Performance Computing Center. Chess is supported by the NSF & NIH/NIGMS via NSF award DMR-1332208, and the MacCHESS resource is supported by NIH/NIGMS award GM-103485.

Supporting Citations

References (72-79) appear in the Supporting Material.

References

1. Cross, B. C. S., I. Sinning, J. Luirink, and S. High. 2009. Delivering proteins for export from the cytosol. Nat. Rev. Mol. Cell Biol. 10:255-264.

2. Driessen, A. J. M., and N. Nouwen. 2008. Protein translocation across the bacterial cytoplasmic membrane. Annu. Rev. Biochem. 77:643-667.

3. Pool, M. 2005. Signal recognition particles in chloroplasts, bacteria, yeast and mammals. Mol. Membr. Biol. 22:3-15.

4. Akopian, D., K. Shen, X. Zhang, and S.-o. Shan. 2013. Signal Recognition Particle: An Essential Protein-Targeting Machine. Annual Review of Biochemistry 82:693-721.

5. Connolly, T., P. J. Rapiejko, and R. Gilmore. 1991. Requirement of GTP hydrolysis for dissociation of the signal recognition particle from its receptor. Science (Washington, D. C., 1883-) 252:1171-1173.

6. Zhang, X., C. Schaffitzel, N. Ban, and S.-o. Shan. 2009. Multiple conformational switches in a GTPase complex control co-translational protein targeting. Proc. Natl. Acad. Sci. U. S. A. 106:1754-1759.

7. Shan, S.-o., S. Chandrasekar, and P. Walter. 2007. Conformational changes in the GTPase modules of the signal reception particle and its receptor drive initiation of protein translocation. J. Cell Biol. 178:611-620.

8. Richter, C. V., T. Bals, and D. Schünemann. 2010. Component interactions, regulation and mechanisms of chloroplast signal recognition particle-dependent protein transport. European Journal of Cell Biology 89:965-973.

9. Schünemann, D. 2004. Structure and function of the chloroplast signal recognition particle. Current Genetics 44:295-304.

10. Aldridge, C., P. Cain, and C. Robinson. 2009. Protein transport in organelles: Protein transport into and across the thylakoid membrane. FEBS Journal 276:1177-1186.

11. Groves, M. R., A. Mant, A. Kuhn, J. Koch, S. Dübel, C. Robinson, and I. Sinning. 2001. Functional Characterization of Recombinant Chloroplast Signal Recognition Particle. Journal of Biological Chemistry 276:27778-27786.

12. Tu, C. J., E. C. Peterson, R. Henry, and N. E. Hoffman. 2000. The L18 Domain of Light-harvesting Chlorophyll Proteins Binds to Chloroplast Signal Recognition Particle 43. Journal of Biological Chemistry 275:13187-13190.

13. DeLille, J., E. C. Peterson, T. Johnson, M. Moore, A. Kight, and R. Henry. 2000. A novel precursor recognition element facilitates posttranslational binding to the signal recognition particle in chloroplasts. Proceedings of the National Academy of Sciences 97:1926-1931.

14. Keenan, R. J., D. M. Freymann, P. Walter, and R. M. Stroud. 1998. Crystal structure of the signal sequence binding subunit of the signal recognition particle. Cell 94:181-191.

15. Hainzl, T., S. Huang, G. Meriläinen, K. Brännström, and A. E. Sauer-Eriksson. 2011. Structural basis of signal-sequence recognition by the signal recognition particle. Nat Struct Mol Biol 18:389-391.

16. Halic, M., M. Blau, T. Becker, T. Mielke, M. R. Pool, K. Wild, I. Sinning, and R. Beckmann. 2006. Following the signal sequence from ribosomal tunnel exit to signal recognition particle. Nature 444:507-511.

17. High, S., R. Henry, R. M. Mould, Q. Valent, S. Meacock, K. Cline, J. C. Gray, and J. Luirink. 1997. Chloroplast SRP54 Interacts with a Specific Subset of Thylakoid Precursor Proteins. Journal of Biological Chemistry 272:11622-11628.

18. Cain, P., I. Holdermann, I. Sinning, A. Johnson, and C. Robinson. 2011. Binding of chloroplast signal recognition particle to a thylakoid membrane protein substrate in aqueous solution and delineation of the cpSRP43–substrate interaction domain. In Biochem J. The Authors Journal compilation. 149-155.

19. Hutin, C., M. Havaux, J.-P. Carde, K. Kloppstech, K. Meiherhoff, N. Hoffman, and L. Nussaume. 2001. Double mutation cpSRP43^{-}/cpSRP54^{-} is necessary to abolish the cpSRP pathway required for the thylakoid targeting of the light-harvesting chlorophyll proteins. In The Plant Journal. Blackwell Science Ltd. 531-543.

20. Klimyuk, V. I., F. Persello-Cartieaux, M. Havaux, P. Contard-David, D. Schuenemann, K. Meiherhoff, P. Gouet, J. D. G. Jones, N. E. Hoffman, and L. Nussaume. 1999. A Chromodomain Protein Encoded by the Arabidopsis CAO Gene Is a Plant-Specific Component of the Chloroplast Signal Recognition Particle Pathway That Is Involved in LHCP Targeting. The Plant Cell Online 11:87-99.

21. Nilsson, R., J. Brunner, N. E. Hoffman, and K. J. Van Wijk. 1999. Interactions of ribosome nascent chain complexes of the chloroplast-encoded D1 thylakoid membrane protein with cpSRP54. EMBO J. 18:733-742.

22. Nilsson, R., and K. J. van Wijk. 2002. Transient interaction of cpSRP54 with elongating nascent chains of the chloroplast-encoded D1 protein; 'cpSRP54 caught in the act'. FEBS Lett. 524:127-133.

23. Dunker, A. K., J. D. Lawson, C. J. Brown, R. M. Williams, P. Romero, J. S. Oh, C. J. Oldfield, A. M. Campen, C. M. Ratliff, K. W. Hipps, J. Ausio, M. S. Nissen, R. Reeves,

C. Kang, C. R. Kissinger, R. W. Bailey, M. D. Griswold, W. Chiu, E. C. Garner, and Z. Obradovic. 2001. Intrinsically disordered protein. Journal of Molecular Graphics and Modelling 19:26-59.

24. Haynes, C., C. J. Oldfield, F. Ji, N. Klitgord, M. E. Cusick, P. Radivojac, V. N. Uversky, M. Vidal, and L. M. Iakoucheva. 2006. Intrinsic Disorder Is a Common Feature of Hub Proteins from Four Eukaryotic Interactomes. PLoS Comput Biol 2:e100.

25. Buskiewicz, I. A., J. Jöckel, M. V. Rodnina, and W. Wintermeyer. 2009. Conformation of the signal recognition particle in ribosomal targeting complexes. RNA 15:44-54.

26. Buskiewicz, I., F. Peske, H.-J. Wieden, I. Gryczynski, M. V. Rodnina, and W. Wintermeyer. 2005. Conformations of the Signal Recognition Particle Protein Ffh from Escherichia coli as Determined by FRET. Journal of Molecular Biology 351:417-430.

27. Gao, F., A. D. Kight, R. Henderson, S. Jayanthi, P. Patel, M. Murchison, P. Sharma, R. L. Goforth, T. K. S. Kumar, R. L. Henry, and C. D. Heyes. 2015. Regulation of Structural Dynamics within a Signal Recognition Particle Promotes Binding of Protein Targeting Substrates. Journal of Biological Chemistry.

28. Anon. 2014. Activities at the Universal Protein Resource (UniProt). Nucleic Acids Res. 42:D191-D198.

29. Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25:3389-3402.

30. Westbrook, J., Z. Feng, L. Chen, H. Yang, and H. M. Berman. 2003. The Protein Data Bank and structural genomics. Nucleic Acids Res. 31:489-491.

31. Schwede, T., J. Kopp, N. Guex, and M. C. Peitsch. 2003. SWISS-MODEL: an automated protein homology-modeling server. Nucleic Acids Res. 31:3381-3385.

32. Ataide, S. F., N. Schmitz, K. Shen, A. Ke, S.-o. Shan, J. A. Doudna, and N. Ban. 2011. The crystal structure of the signal recognition particle in complex with its receptor. Science 331:881-886.

33. Rosendal, K. R., K. Wild, G. Montoya, and I. Sinning. 2003. Crystal structure of the complete core of archaeal signal recognition particle and implications for interdomain communication. Proc Natl Acad Sci U S A 100:14701-14706.

34. Sievers, F., A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Soding, J. D. Thompson, and D. G. Higgins. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7:539.

35. Larkin, M. A., G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson, and D. G. Higgins. 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23:2947-2948.

36. Eswar, N., B. Webb, M. A. Marti-Renom, M. S. Madhusudhan, D. Eramian, M.-y. Shen, U. Pieper, and A. Sali. 2001. Comparative Protein Structure Modeling Using MODELLER. In Current Protocols in Protein Science. John Wiley & Sons, Inc.

37. Zhang, Y. 2008. I-TASSER server for protein 3D structure prediction. BMC Bioinf. 9:No pp. given.

38. Roy, A., A. Kucukural, and Y. Zhang. 2010. I-TASSER: a unified platform for automated protein structure and function prediction. Nat. Protocols 5:725-738.

39. Xu, D., and Y. Zhang. 2012. Ab initio protein structure assembly using continuous structure fragments and optimized knowledge-based force field. Proteins: Structure, Function, and Bioinformatics 80:1715-1735.

40. Schrödinger, L. 2010. The PyMOL Molecular Graphics System, Version 1.3.

41. Melo, F., and E. Feytmans. 1998. Assessing protein structures with a non-local atomic interaction energy. J. Mol. Biol. 277:1141-1152.

42. Scott, W. R. P., P. H. Huenenberger, I. G. Tironi, A. E. Mark, S. R. Billeter, J. Fennen, A. E. Torda, T. Huber, P. Krueger, and W. F. van Gunsteren. 1999. The GROMOS Biomolecular Simulation Program Package. J. Phys. Chem. A 103:3596-3607.

43. Benkert, P., S. C. E. Tosatto, and D. Schomburg. 2008. QMEAN: a comprehensive scoring function for model quality assessment. Proteins: Struct., Funct., Bioinf. 71:261-277.

44. Laskowski, R. A., M. W. MacArthur, D. S. Moss, and J. M. Thornton. 1993. PROCHECK: a program to check the stereochemical quality of protein structures. Journal of Applied Crystallography 26:283-291.

45. Shatsky, M., R. Nussinov, and H. J. Wolfson. 2004. A method for simultaneous alignment of multiple protein structures. Proteins: Structure, Function, and Bioinformatics 56:143-156.

46. SYBYL-X 2.1. Tripos International, 1699 South Hanley Rd., St. Louis, Missouri, 63144, USA.

47. Gasteiger, E., C. Hoogland, A. Gattiker, S. Duvaud, M. R. Wilkins, R. D. Appel, and A. Bairoch. 2005. Protein identification and analysis tools on the ExPASy server. Humana Press Inc. 571-607.

48. Li, X., P. Romero, M. Rani, A. K. Dunker, and Z. Obradovic. 1999. Predicting protein disorder for N-, C- and internal regions. Genome Inf. Ser. 10:30-40.

49. Romero, P., Z. Obradovic, X. Li, E. C. Garner, C. J. Brown, and A. K. Dunker. 2001. Sequence complexity of disordered protein. Proteins: Struct., Funct., Genet. 42:38-48.

50. Romero, P., Z. Obradovic, and A. K. Dunker. 1997. Sequence data analysis for long disordered regions prediction in the calcineurin family. Genome Inf. Ser. 8:110-124.

51. Brooks, B. R., C. L. Brooks, III, A. D. Mackerell, Jr., L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, and M. Karplus. 2009. CHARMM: The biomolecular simulation program. J. Comput. Chem. 30:1545-1614.

52. Phillips, J. C., R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, and K. Schulten. 2005. Scalable molecular dynamics with NAMD. J. Comput. Chem. 26:1781-1802.

53. Best, R. B., X. Zhu, J. Shim, P. E. M. Lopes, J. Mittal, M. Feig, and A. D. MacKerell. 2012. Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone φ, ψ and Side-Chain χ1 and χ2 Dihedral Angles. J. Chem. Theory Comput. 8:3257-3273.

54. Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD: visual molecular dynamics. J Mol Graph 14:33-38, 27-38.

55. Goforth, R. L., E. C. Peterson, J. Yuan, M. J. Moore, A. D. Kight, M. B. Lohse, J. Sakon, and R. L. Henry. 2004. Regulation of the GTPase Cycle in Post-translational Signal Recognition Particle-based Protein Targeting Involves cpSRP43. Journal of Biological Chemistry 279:43077-43084.

56. Acerbo, A. S., M. J. Cook, and R. E. Gillilan. 2015. Upgrade of MacCHESS facility for X-ray scattering of biological macromolecules in solution. Journal of Synchrotron Radiation 22:180-186.

57. Bernado, P., E. Mylonas, M. V. Petoukhov, M. Blackledge, and D. I. Svergun. 2007. Structural Characterization of Flexible Proteins Using Small-Angle X-ray Scattering. J. Am. Chem. Soc. 129:5656-5664.

58. Konarev, P. V., V. V. Volkov, A. V. Sokolova, M. H. J. Koch, and D. I. Svergun. 2003. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. Journal of Applied Crystallography 36:1277-1282.

59. Santoso, Y., J. P. Torella , and A. N. Kapanidis. 2010. Characterizing Single-Molecule FRET Dynamics with Probability Distribution Analysis. ChemPhysChem 11:2209-2219.

60. Buskiewicz, I., A. Kubarenko, F. Peske, M. V. Rodnina, and W. Wintermeyer. 2005. Domain rearrangement of SRP protein Ffh upon binding 4.5S RNA and the SRP receptor FtsY. RNA 11:947-957.

61. Egea, P. F., J. Napetschnig, P. Walter, and R. M. Stroud. 2008. Structures of SRP54 and SRP19, the Two Proteins that Organize the Ribonucleic Core of the Signal Recognition Particle from <italic>Pyrococcus furiosus</italic>. PLoS ONE 3:e3528.

62. Hainzl, T., S. Huang, and A. E. Sauer-Eriksson. 2007. Interaction of signal-recognition particle 54 GTPase domain and signal-recognition particle RNA in the free signal-recognition particle. Proceedings of the National Academy of Sciences 104:14911-14916.

63. Janda, C. Y., J. Li, C. Oubridge, H. Hernandez, C. V. Robinson, and K. Nagai. 2010. Recognition of a signal peptide by the signal recognition particle. Nature 465:507-510.

64. von Loeffelholz, O., K. Knoops, A. Ariosa, X. Zhang, M. Karuppasamy, K. Huard, G. Schoehn, I. Berger, S.-o. Shan, and C. Schaffitzel. 2013. Structural basis of signal sequence surveillance and selection by the SRP–FtsY complex. Nat Struct Mol Biol 20:604-610.

65. Schaffitzel, C., M. Oswald, I. Berger, T. Ishikawa, J. P. Abrahams, H. K. Koerten, R. I. Koning, and N. Ban. 2006. Structure of the E. coli signal recognition particle bound to a translating ribosome. Nature 444:503-506.

66. Brown, C. J., S. Takayama, A. M. Campen, P. Vise, T. W. Marshall, C. J. Oldfield, C. J. Williams, and A. K. Dunker. 2002. Evolutionary rate heterogeneity in proteins with long disordered regions. J. Mol. Evol. 55:104-110.

67. Holdermann, I., N. H. Meyer, A. Round, K. Wild, M. Sattler, and I. Sinning. 2012. Chromodomains read the arginine code of post-translational targeting. Nat. Struct. Mol. Biol. 19:260-263.

68. Ariosa, A. R., S. S. Duncan, I. Saraogi, X. Lu, A. Brown, G. J. Phillips, and S.-O. Shan. 2013. Fingerloop activates cargo delivery and unloading during cotranslational protein targeting. Molecular Biology of the Cell 24:63-73.

Figures

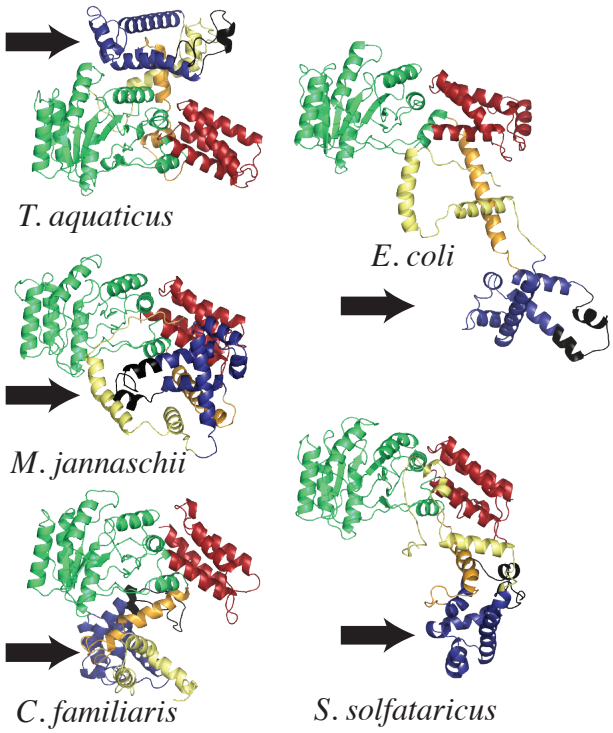

Figure 1: Homology models for cpSRP54. The organism name listed is that of the template used to build each model. The NG-domain's orientations are identical in each image to highlight the variability in the M-domain (black arrows) orientation in each template and therefore in each homology model. The PDB IDs for *T. aquaticus*, *E. coli*, *M. jannaschii, C. familiaris*, *and S. solfataricus* are 2FFH, 2XXA, 3NDB, 2J37, and 1QZX, respectively. The N-domain (red) and G-domain (green) are connected to the M-domain (blue) *via* a flexible linker (orange). The M-domain contains two additional flexible segments, the finger loop (black) and the c-terminal extension (yellow).

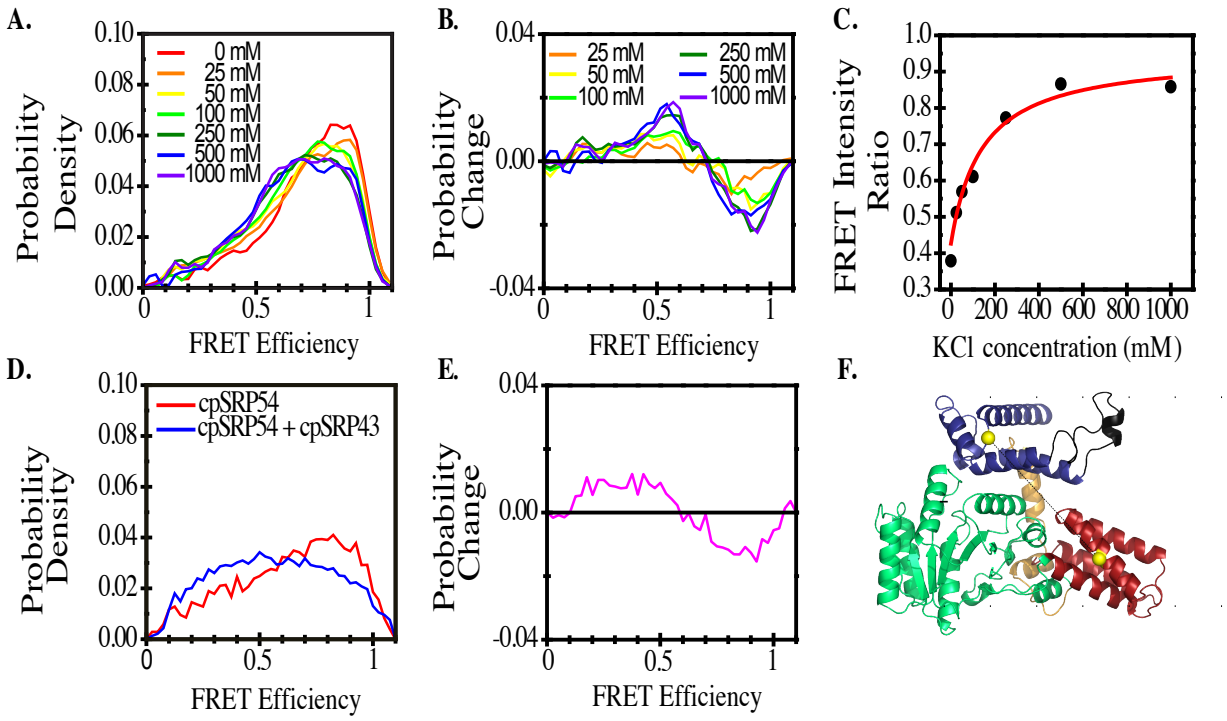

Figure 2: A- The smoothed FRET histograms of cpSRP54 at different concentrations of KCl (smoothed histograms are provided here for visual clarity; the unsmoothd curves are provided in the supplementary section). As the KCl concentration is increased the 0.90 FRET peak is diminished while a 0.70 FRET peak begins to emerge. This suggests the high FRET state is stabilized by salt-bridges while the middle FRET state is primarily hydrophobic in nature. **B-** Subtracted FRET data as a function of KCl concentration. The subtraction is the KCl concentration FRET efficiency minus the initial 0mM additional KCl FRET efficiency. **C-** Titration curve showing the change in the ratio of the FRET efficiency at 0.54 *versus* 0.86 as the KCl concentration is increased **D-** FRET of cpSRP54 in the presence and absence of cpSRP43. Similar to the KCl titration results, the presence of cpSRP43 results in a conformational change to a middle FRET peak. However, this state appears at lower FRET suggesting the state is not identical to the state observed to increase at higher KCl concentration. **E-** Subtraction of cpSRP54 alone from cpSRP43/cpSRP54 FRET. **F-** Structure of cpSRP54; beads indicate the location of the Dye molecules.

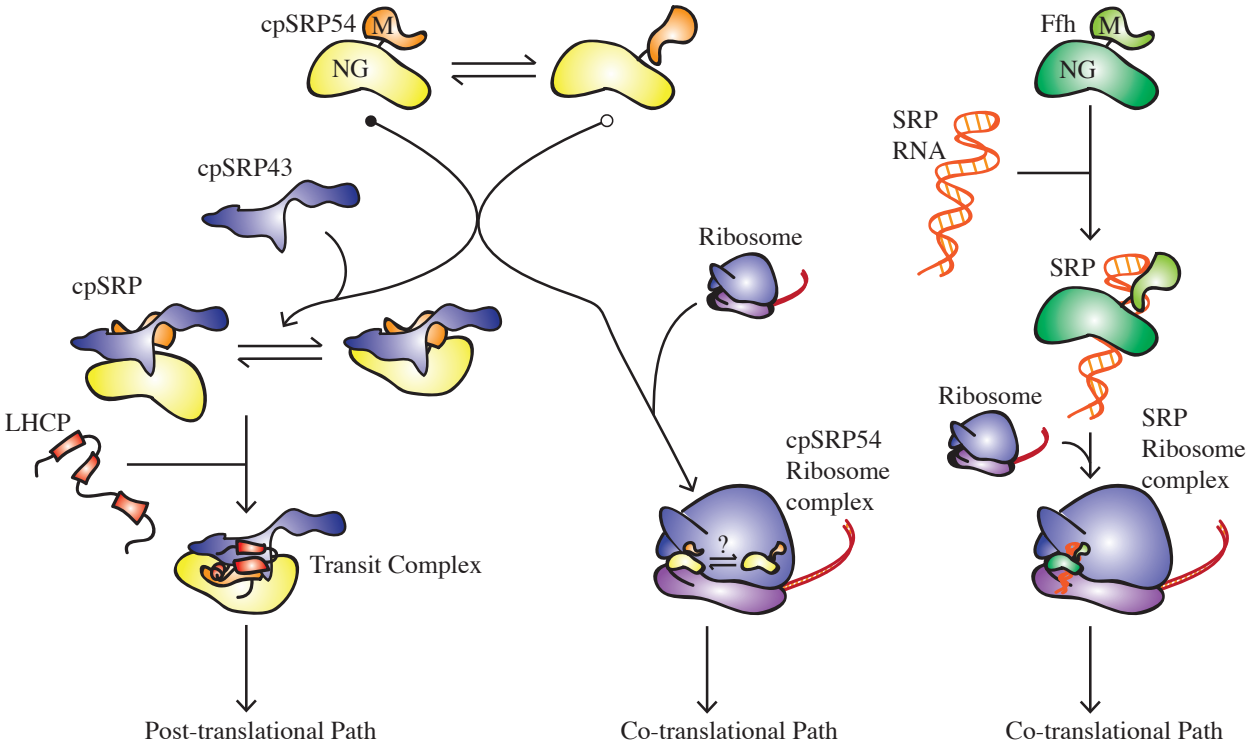

Figure 3*:* Left- cpSRP path depicting events leading to the post- and co-translational pathways. cpSRP54 exists in at least two different conformations; low ionic strength favors a state similar to that observed for *T. aq*. Ffh while high ionic strength favors a state similar to *S. sol* SRP54. The *S. sol* SRP54 like state is consistent with the state occupied by bacterial Ffh when bound to SRP RNA and possibly allows cpSRP54 to act in the absence of SRP RNA. When bound to cpSRP43, cpSRP54 retains its structural heterogeneity. Upon binding LHCP cpSRP54's M-domain possibly interacts with TM3. **Right-** The traditional SRP54/Ffh co-translational signaling pathway. This pathway differs starkly in that it requires SRP RNA and functions only in a co-translational pathway. Whether cpSRP54 continues to sample multiple conformations at the ribosome remains to be determined.

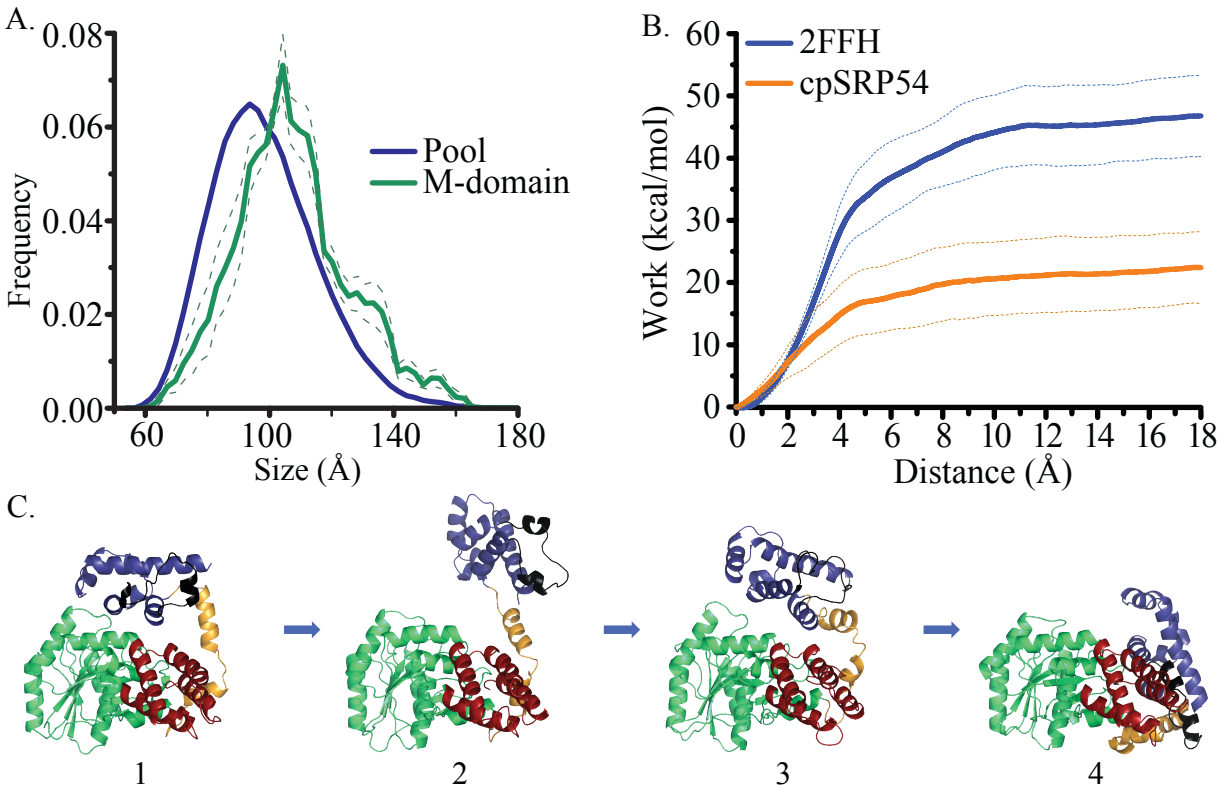

Figure 5: A- Size distribution determined using the ensemble optimization method for the M-domain. The wide M-domain distribution (green) shifted toward higher size compared to the pool (blue) is indicative of a large amount of disorder. **B-** Profiles of the averaged accumulated work (solid lines) from the steered molecular dynamics simulations of 2FFH (blue) and cpSRP54 (orange) with the standard deviations for each (dashed lines). The results indicate the 2FFH M-domain is tightly bound to the NG-domains and that the cpSRP54 M-NG binding is significantly weaker. **C-** The M-domain position shift. The initial position in (1) quickly transitions to the decoupled M-domain (blue) in state (2). Folding of the linker (green) helix with α-M1 to form the "greasy slide" along with folding of the finger loop onto the linker results in recoupling of the M-domain to the N–domain (red) and G-domain (orange) as seen in (3) and (4).

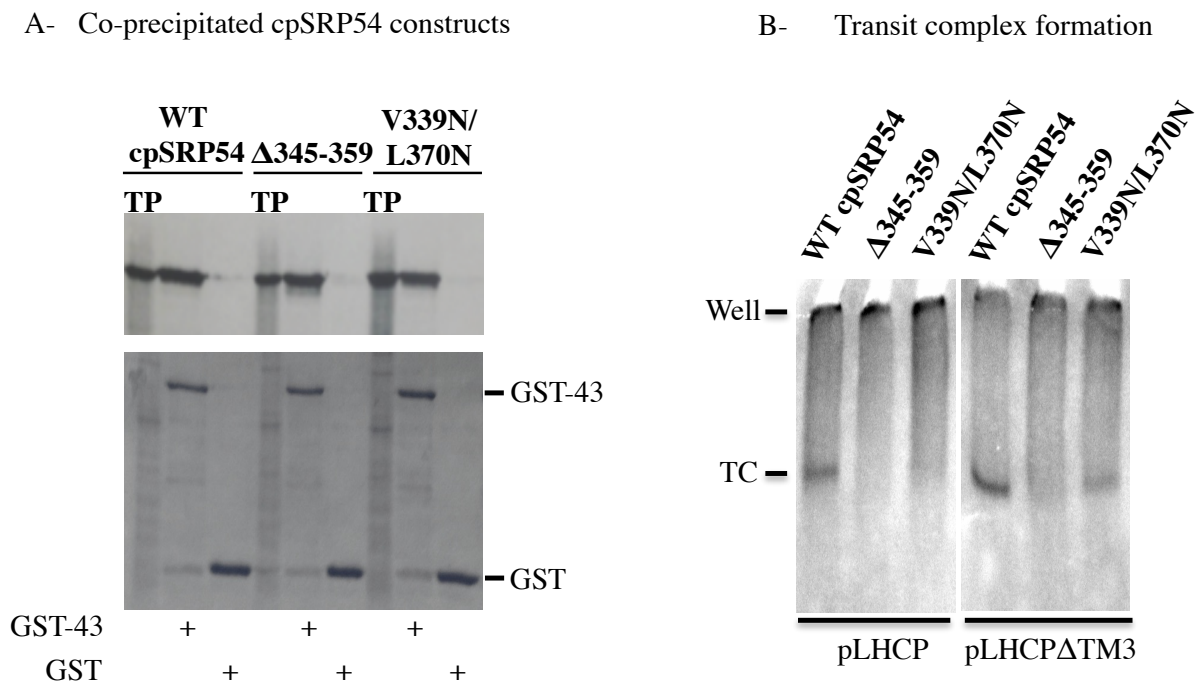

Figure 5: cpSRP54 M-domain is critical for interaction with TM3 domain of LHCP. **A-** *In vitro* translated radiolabeled cpSRP54 (WT cpSRP54) or cpSRP54 mutants ($cpSRP54_{V340N/L371N}$ and $cpSRP54_{\Delta 346-360}$) were incubated with GST or GST-cpSRP43 (GST-43) and described in methods. Proteins were recovered with glutathione sepharose and separated by SDS-PAGE. Radiolabeled translation product (TP) and co-precipitated cpSRP54 constructs were visualized by fluorography (top) or by protein staining (bottom). **B-** The indicated cpSRP54 construct was produced by *in vitro* translation and incubated with recombinant cpSRP43 and radiolabeled *in vitro* translated pLHCP or pLHCP ΔTM3 as described in methods. Transit complex (TC) containing the indicated radiolabeled LHCP construct was examined by native-PAGE/fluorography as described in methods. The position of aggregated LHCP constructs in the well is indicated in the figure.

Supporting Material Methods

Bioinformatics

The sequence for cpSRP54 used for homology modeling was used for conservation analysis. The sequence was submitted for a BLAST search using BLAST 2.2.29+ against the Swiss-Prot Database using an e-value cut-off of $1x10^{-10}$ using the BLOSSUM64 matrix. A total of one hundred and forty-one hits were obtained containing SRP54's as well as SRP receptor proteins. The receptor proteins were removed from the data set, leaving a total of ninety-two unique hits. Sequences were collected from UniProt and aligned using Clustal Omega. The residues identified as important for M-NG coupling in the *T. aq*. Ffh simulation were checked for conservation in the multiple sequence alignment from Clustal Omega. A simple conservation scheme was employed in which substitutions compared to 2FFH residues between Arg and Lys, Glu and Asp, Gln and Asn, as well as Leu, Val, and Ile were considered conserved. Note, aromatic residue substitutions of Leu, Val, and Ile were not allowed as the positioning of these residues in the binding interface suggests core packing would be hindered in the presence of bulky, aromatic side-chains.

Sequence conservation was calculated as the fraction of the total conserved residues versus the total searched. In order to determine whether the interaction site is more conserved than any specific set, twenty-three residues were selected at random among all surface residues of 2FFH since a total of twenty-three important residues were identified for the M-domain to NG-domain interaction site. Surface residues were selected as those with greater than $20Å^2$ of solvent-accessible surface area, as determined using VMD. These residues were evaluated for their conservation. This process was iterated 100,000 times to produce a dataset consisting of

9,200,000 protein site percent-conservation totals. These sets were then separated into sets corresponding to totals for archaeal, eukaryotic and bacterial proteins. These were compared to the interaction site data set. Additionally, since cpSRP54 is particularly similar to 2FFH, the random site distribution for cpSRP54 was evaluated in the manner above. All analysis was performed using custom Perl scripts.

The relative sequence conservation was also determined using ConSurf, an online server that weighs substitutions based on less stringent conservation rules as well as on the phylogenetic distribution in the data set.(2-5) The Bayesian method was used in this analysis and, since there are a greater number of residues covering more of all the proteins in the data set, the cpSRP54 structure and sequence was used for submission.

The rate of evolution for the G-domain to M-domain linker, the M-domain finger loop, and a set in which both the G-domain to M-domain linker and M-domain loop sequences were removed from the full SRP54/Ffh sequence (ordered set) was determined using the PHYLIP program protdist using the Dayhoff PAM matrix.(1, 6) The M-NG interaction site residues were evaluated similarly. The distance difference was determined by subtracting the disordered/interaction site distances from the ordered set distances.(1) For this, a new and larger dataset was generated. Rates were determined for bacteria, archaea and eukaryotes. BLAST searches were submitted for plants, eukaryotes, archaea and bacteria separately using the cpSRP54 against the TrEMBL database using an e-value cut-off of $1x10^{-4}$ and the BLOSSUM64 matrix with a maximum of 1,000 hits each. The sequences were then combined after removing duplicates. In order to ensure only Ffh/SRP54/cpSRP54 proteins were included in the dataset, all sequences with fewer than four hundred amino acids were removed. Subsequently, the sequences were aligned using Clustal Omega followed by trimming of the sequences using Trimal 1.4 using

a gap threshold of 0.6, a residue overlap of 0.40, and a sequence overlap of 90 percent. Any sequence missing the linker and/or loop residues was removed leaving a total of 1,538 proteins in the dataset. Data processing was performed using custom Perl scripts.

Supporting Material Results

Bioinformatics

A BLAST search of the SwissProt database for the cpSRP54 sequence was performed to examine the degree of conservation of the residues involved in the M-domain binding to the NG-domains. A set of sequences for the BLAST hits in which the 23 M-domain to NG-domain interaction site residues were removed was prepared for random sampling. For all proteins in this set, the distribution of percent conservation in the interaction site residues is shifted toward greater conservation than the randomly selected set. Evolutionary distances determined using the PHYLIP program protdist indicates the M-domain to NG-domain binding residues in bacterial and eukaryotic proteins tend to be more conserved. However, these same residues are less conserved in archaeal SRP54s; this is consistent with the fact that the M-domain to NG-domain interaction site observed in bacteria is not observed in archaea. The most conserved interaction site residues are found in the sequences for bacterial Ffh. Importantly, the interaction site residues show significantly higher similarity than the random selection (Fig. S2). The M-domain to NG-domain interaction site conservation in cpSRP54 compared to the T. aq. Ffh interaction site is ~70%. Comparing this conservation to the randomly selected set of surface residues for cpSRP54 indicates that only ~7% of the sets will have greater conservation than this site. Considering that these sets are not necessarily clustered in the same location and can include the highly conserved GTPase residues suggests that the site is particularly well conserved in

cpSRP54. Several proteins in the data set have significantly low conservation as compared to the random sets, suggesting this interaction site is not well conserved across all SRP54s.

The relative rate of mutation in the G-domain to M-domain linker and the M-domain finger loop was estimated for a large set of SRP54/Ffh sequences. The difference in the rate of mutation in the linker *versus* ordered segments of bacterial Ffh is quite small (-0.04), however mutation rates among archaeal and eukaryotic SRP54 (-0.32 and -0.27 respectively) is considerably higher (Fig. S4). In contrast, the finger loop has a greater rate of mutation relative to the ordered segments in all domains of life (-0.43, -0.33 and -0.29 for bacteria, archaea and eukaryote, respectively) (Fig. S4). These results indicate increased sampling of residues in these regions of SRP54/Ffh that may explain differences observed in the structures of free SRP54s/Ffhs.

M-domain to NG-domain Interaction Details

The 2FFH packed hydrophobic core of residues formed between Leu367 of the M-domain core and residues Leu106, Ile194, Val229 and Ala232 of the G-domain, remains in place throughout the simulations (Fig. S5). This region represents a significant contact surface between the M-domain and NG-domains and likely contributes to significant coupling of these domains. Backbone hydrogen bonding between Ile194(NH) of the G-domain and of Ser368(CO) of the M-domain and G-domain Glu196(NH) and M-domain Gly391(CO) stabilizes the contact at the M-domain core (Fig. S5). These contacts are further stabilized *via* side-chain interactions between *G*-domain Arg231 and *M*-domain Glu363, *G*-domain Gln224 and *M*-domain Lys327, *G*-domain Gln193 and *M*-domain Thr370, *G*-domain Gln193 and *M*-domain Ser368, as well as *G*-domain Glu225 and *M*-domain Arg374 (Fig. S5). The latter three are somewhat transient and may not likely contribute significantly to the binding free energy. Although the M-domain core near the

G-domain appears to be the strongest region of interaction, several important contacts occur between the M and N-domain interface, particularly near the C-terminus of α-M1 and the N-terminus of α-M2 (Fig. S5). Arg334 forms a close contact with the backbone of residue Val43(CO) and the Asp42 side-chain of the N-domain. This segment of the N-domain is relatively well conserved and has been implicated in mediating signal sequence binding.(7) Additional contacts were observed between N-domain Glu46 and M-domain Lys360, M-domain Glu356 and N-domain Arg49 in the simulation. In one simulation, the N-domain rotates to accommodate these interactions, consistent with previous SRP54 simulations in which the N-domain is capable of rotation about the N-G-domain interface.(8, 9)

M-domain Recouping Interaction Residues

Prior to the formation of the GL, the finger loop (residues M346-M352) folds over the peptide-binding site close to the C-terminus of α–M4. The G-domain to M-domain linker then forms the GL *via* interaction of Ile323, Ala326 and Phe328 with a finger loop helix (Thr347-M352) and Met427 in α–M4. The transition results in a packed core between the M-domain to NG-domain linker and finger loop residues V349 and M352, as well as a single electrostatic interaction between K322 and D332.

Control Experiments for smFRET

We performed several control experiments to ensure that changes in the smFRET histograms are due to structural changes in the protein and not due to environmental effects from change in the concentration of KCl. To examine for changes in the fluorescent properties of the dye-labeled proteins with KCl concentration at the ensemble and single-molecule level, we labeled the double-cysteine mutant cpSRP54 proteins with only Alexa488 or only Alexa594. As shown in figure S6, the fluorescence intensity of Alexa488 or Alexa594 labeled cpSRP54

proteins did not show a change with KCl concentration up to 1000 mM. Furthermore, Figure S7 also shows that the photon counting histograms for single Alexa488-labeled cpSRP54 did not show KCl concentration dependent effect(s).

Additionally, Figure S8 shows that the threshold chosen for detecting single molecule fluorescence bursts from the background does not affect the smFRET histograms at any of the KCl concentrations. The only difference is that the number of detected molecules decreases (from about ~7000 to ~5000 per trace) when a higher threshold is chosen (10 counts/ms vs 20 counts/ms), and thus the statistical noise increases slightly.

This data is further quantified in table S3, whereby the average intensity of the donor and acceptor signal is not affected by KCl concentration for a given threshold level and that the average dwell time in the laser focus is unaffected by either threshold or KCl. The numbers in parentheses are the standard deviations of the values.

Taken together, this control data strongly supports that the both the ensemble and single molecule fluorescence properties of the dyes are unaffected by KCl concentration under the experimental conditions used, thus validating that changes in the smFRET histograms with KCl are due to protein structural changes and not due to dye photophysics.

Figure S9 shows that neither smoothing of the smFRET data or choice of binning introduced artifacts. This is highlighted by the similar Kd values obtained in the titration; only the R^2 values and standard errors are different due to differing levels of statistical noise. The smoothed data allows for better viewing of the trends by reducing statistical noise but does not bias the data analysis.

Supporting References

1. Brown, C. J., S. Takayama, A. M. Campen, P. Vise, T. W. Marshall, C. J. Oldfield, C. J. Williams, and A. K. Dunker. 2002. Evolutionary rate heterogeneity in proteins with long disordered regions. J. Mol. Evol. 55:104-110.

2. Landau, M., I. Mayrose, Y. Rosenberg, F. Glaser, E. Martz, T. Pupko, and N. Ben-Tal. 2005. ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. Nucleic Acids Research 33:W299-W302.

3. Ashkenazy, H., E. Erez, E. Martz, T. Pupko, and N. Ben-Tal. 2010. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucleic Acids Research 38:W529-W533.

4. Glaser, F., T. Pupko, I. Paz, R. E. Bell, D. Bechor-Shental, E. Martz, and N. Ben-Tal. 2003. ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information. Bioinformatics 19:163-164.

5. Celniker, G., G. Nimrod, H. Ashkenazy, F. Glaser, E. Martz, I. Mayrose, T. Pupko, and N. Ben-Tal. 2013. ConSurf: Using Evolutionary Data to Raise Testable Hypotheses about Protein Function. Israel Journal of Chemistry 53:199-206.

6. Felsenstein, J. 2005. PHYLIP (Phylogeny Inference Package) version 3.6. *Distributed by the author*, *Department of Genome Sciences, University of Washington, Seattle*.

7. Newitt, J. A., and H. D. Bernstein. 1997. The N-domain of the signal recognition particle 54-kDa subunit promotes efficient signal sequence binding. Eur. J. Biochem. 245:720-729.

8. Yang, M., X. Zhang, and K. Han. 2010. Molecular dynamics simulation of SRP GTPases: Towards an understanding of the complex formation from equilibrium fluctuations. Proteins: Struct., Funct., Bioinf. 78:2222-2237.

9. Chang, S., H.-q. He, J.-p. Hu, X. Jiao, and X.-h. Tian. 2012. Network models reveal stability and structural rearrangement of signal recognition particle. J. Biomol. Struct. Dyn. 30:150-159.

Supporting Tables and Figures

Table S1.

Modeling Statistics						
cpSRP54 Structure (Template)	Core (%)	Allowed (%)	Generously Allowed (%)	Disallowed (%)	QMEAN	Z-Score
Modeller	88.1	8.4	2.6	1	0.575	-2.32
3NDB	92.1	6	1.4	0.5	0.606	-1.95
2XXA	21.4	6.2	1.4	1	0.599	-2.03
2J37	88.1	7.9	2.4	1.7	0.49	-3.34
1QZW	87.6	8.1	3.3	1	0.523	-2.95
2FFH	91.2	6	1.7	1.2	0.569	-2.4
cpSRP54 BLAST (Template)	Bits	e-value	Identities (ratio - %)	Positives (ratio - %)	Gaps (ratio - %)	Coverage (%)
3NDB	254	1.00E-67	164/411 - 39	245/411 – 59	29/411 – 7	83
2XXA	340	1.00E-93	200/427 - 46	268/427 -62	18/427 – 4	86
2J37	217	2.00E-56	133/438 – 30	235/438 – 53	11/438 – 2	87
1QZW	227	2.00E-59	143/391 – 36	226/391 – 57	16/391 – 4	77
2FFH	308	6.00E-84	276/429 – 64	276/429 – 64	24/429 – 5	87
cpSRP54 M-domain BLAST (Template)	Bits	e-value	Identities (ratio - %)	Positives (ratio - %)	Gaps (ratio - %)	Coverage (%)
2XXA	68	4.00E-12	50/136 - 36	17/136 -52	17/136 -12	66
2J37	66	2.00E-11	44/152 -28	82/152 -53	9/152 – 5	74
1QZW	51	5.00E-07	41/121 – 33	64/121 -52	14/121 – 11	57
2FFH	71	7.00E-13	44/124 – 35	74/124 – 59	6/124 – 4	69

Table S2.

MultiProt Multiple Structure Alignments		
Domain(s) (cpSRP54 models)	RMSD (Å)	Number of Residues (total aligned – percent aligned)
NG (E1-L295)	1.61	164 - 55.6%
N-All (E1-M87)	1.33	46 - 52.9%
N-Minus 2XXA (E1-M87)	1.40	67 -77.0%
G (P101-L295)	1.44	150 - 76.9%
M-core (K369-G432)	0.99	58 - 90.6%

Table S3.

	Average donor burst height (counts/ms)		Average acceptor burst height (counts/ms)		Average dwell time (ms)	
	Threshold 10	Threshold 20	Threshold 10	Threshold 20	Threshold 10	Threshold 20
0mM KCL	20	32	20	32	1.4 (0.8)	1.3 (0.6)
25mM KCL	20	31	18	30	1.4 (0.8)	1.3 (0.6)
50mM KCL	20	30	17	28	1.4 (0.7)	1.3 (0.6)
100mM KCL	20	31	17	28	1.4 (0.8)	1.2 (0.5)
250mM KCL	19	30	17	29	1.4 (0.8)	1.3 (0.6)
500mM KCL	19	30	18	30	1.4 (0.7)	1.2 (0.6)
1000mM KCL	18	29	17	27	1.4 (0.7)	1.2 (0.5)
average	19	30	18	29	1.4 (0.8)	1.3 (0.6)

Figure S1.

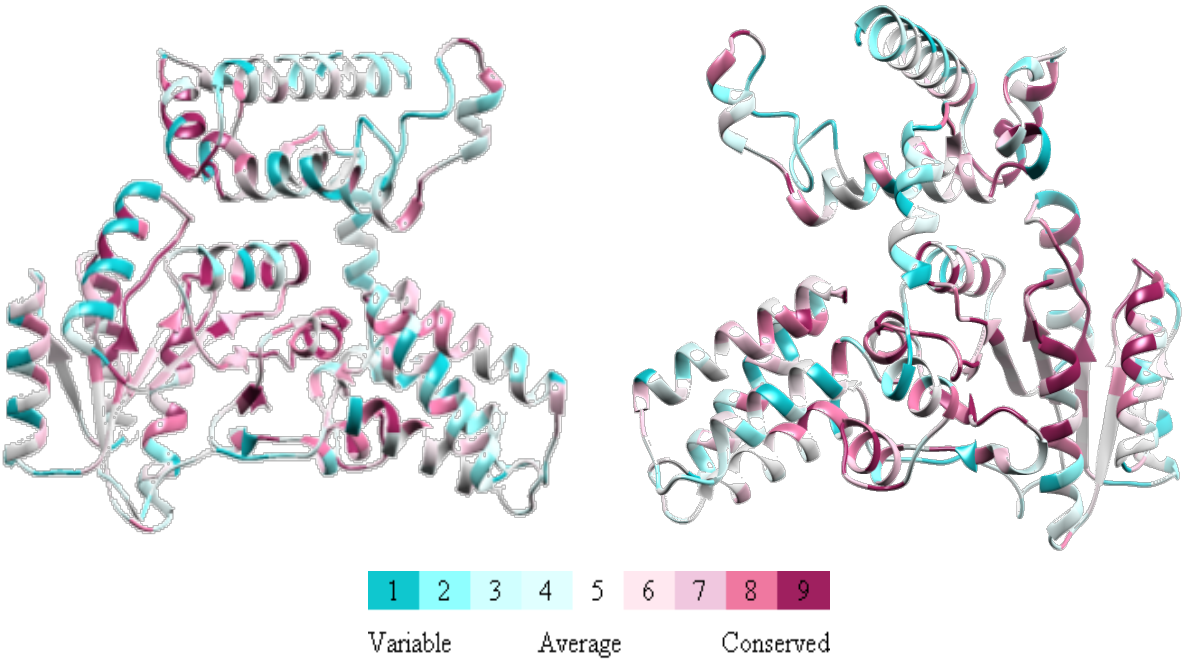

Figure S1: ConSurf conservation statistics depicted on the cpSRP54 structure. Conservation scores are represented as colors ranging from blue (least conserved) to white (moderately conserved) to red (highly conserved). As can be seen on the structure, surface residues tend to be poorly conserved for SRP54 proteins in this data set.

Figure S2.

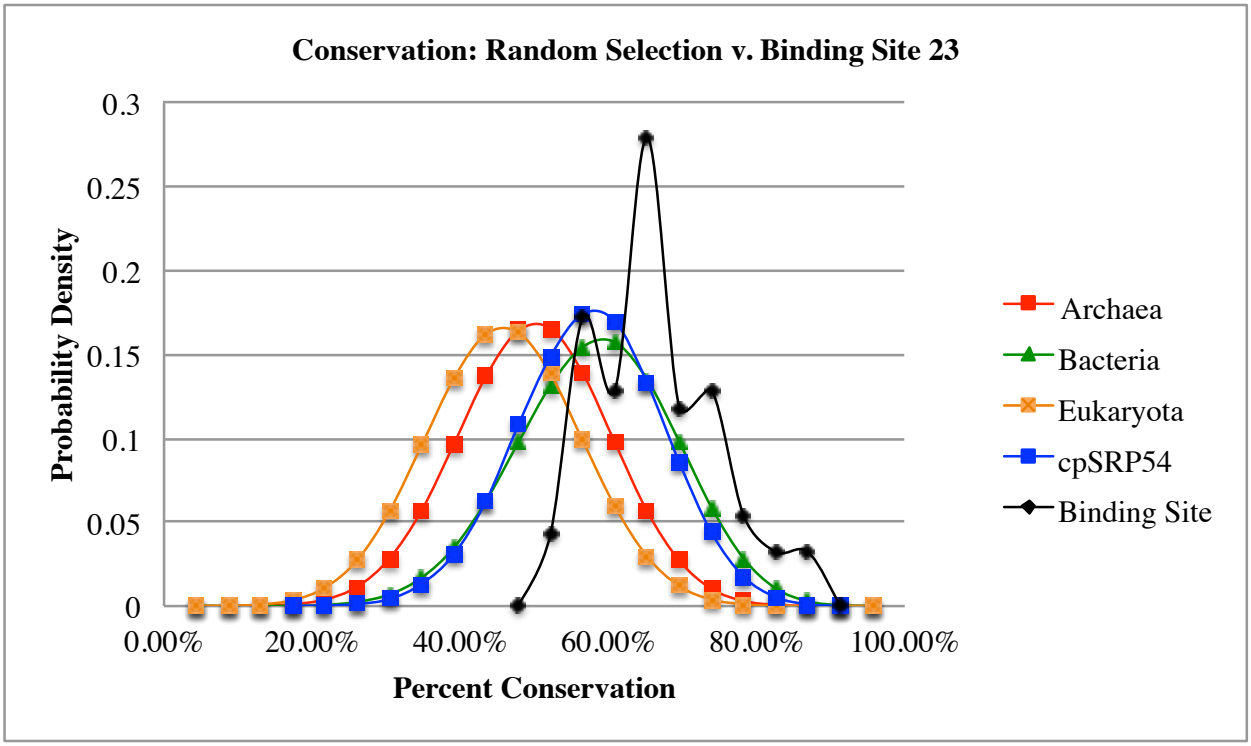

Figure S2: Conservation statistics for the 23 selected residues from the 2FFH simulation compared to a random selection of 100,000 sets of 23 residues for each protein compiled for eukaryotic, bacterial and archaeal proteins compared to 2FFH. The results show that the interaction site 23 tend toward greater conservation as compared to the random selection. In particular, the random selection leaves only ~6.5% of sets greater than the percent conservation for the 2FFH interaction site in cpSRP54. Smooth lines are drawn for visual clarity.

Figure S3.

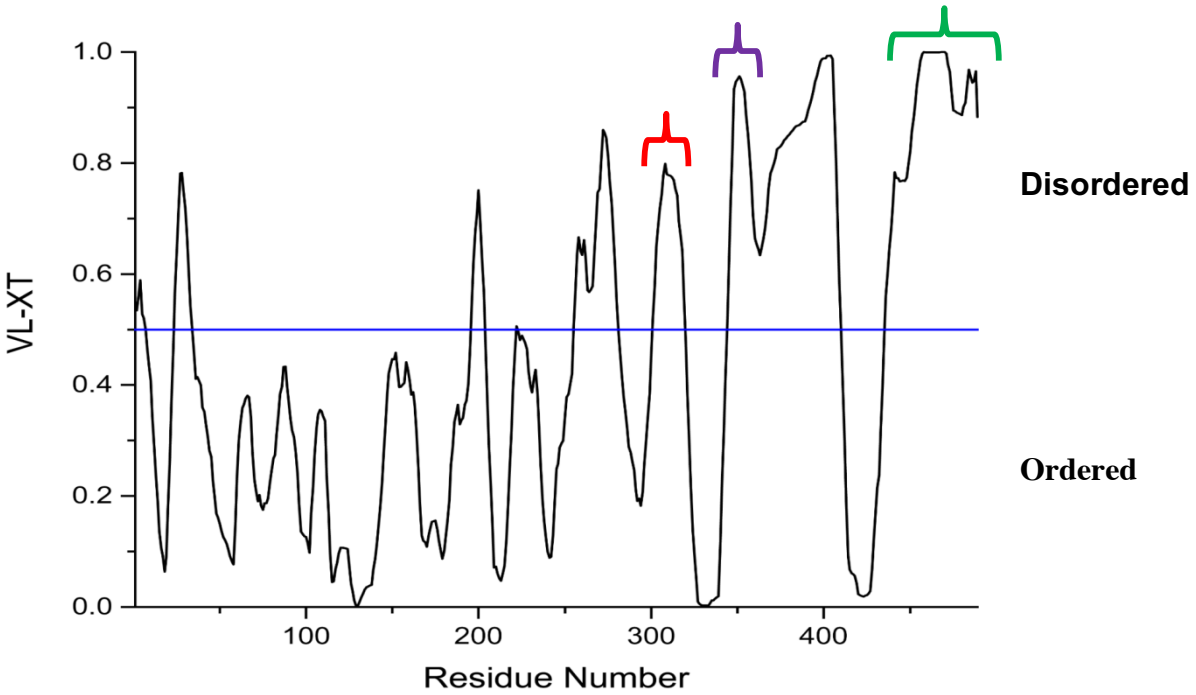

Figure S3: PONDR disorder prediction indicating disorder in the G-domain to M-domain linker region (red), the finger loop (purple), and the C-terminal extension (green).

Figure S4.

```
N-domain 1          11         21         31         41         51
         |          |          |          |          |          |
2FFH     -FQ-QLSARL QEAIGRLRGR GRITE-EDLK ATLREIRRAL MDADVNLEVA \RDFVERVRE
cpSRP54  M.G...TGG. EA.WSK.K.E EVL.K.DNIA EPM.D..... LE...S.P.V ..R..QS.SD
1QZX     .--.M.ENI- RD.VRKFLTG STPY..KAVD EFIKDLQKS. ISS...VKLV .FSLTAKIK.
3DM5     VLD.N.GKA. ANTLKKIARA SSVD..ALI. ELV.D.Q... IQ....VRLV .LQLTREIQR

G-domain 181        191        201        211        221        231
         |          |          |          |          |          |
2FFH     VEEKARLEAR DLILVDTAGR LQI\DEPLMG ELARLKEVLG PDEVLLVLDA MTGQEALSVA
cpSRP54  GLKE.KKNNV .VVIM..... ...KGM.D ..KDV.KF.N .T.....V.. .....AALV
1QZX     GVDIFVKNKM .I.I..... HGY.ETK.LE .MKEMYD..K ..D.I..I.. SI..K.YDL.
3DM5     GVDYFKSKGV .I.I..... HKE..KA.IE .MKQISN.IH .H..I..I.G TI..Q.YNQ.

M-domain 361        371        381        391        401        411
         |          |          |          |          |          |
2FFH     GLLPGVPQ-- ---GLKVD-E KAIKRLEAIV LSMTPEERKD PRIL-GS\RR KRIAKGSGTS
cpSRP54  .MI..MGK.S P.AQIREA.. .NLLVM...MI EV.....ER .EL..PE... .....D..KT
1QZX     QHI..LGIMP SEDQ..IG.. EK.R.WL.AL N...YK.LEN .N.I.K...M R...E...LE
3DM5     RMI..LGYS. PDDVISIG.. ERL.KFKV.M D...E..LLN .E.I.Y...I ....R.....
```

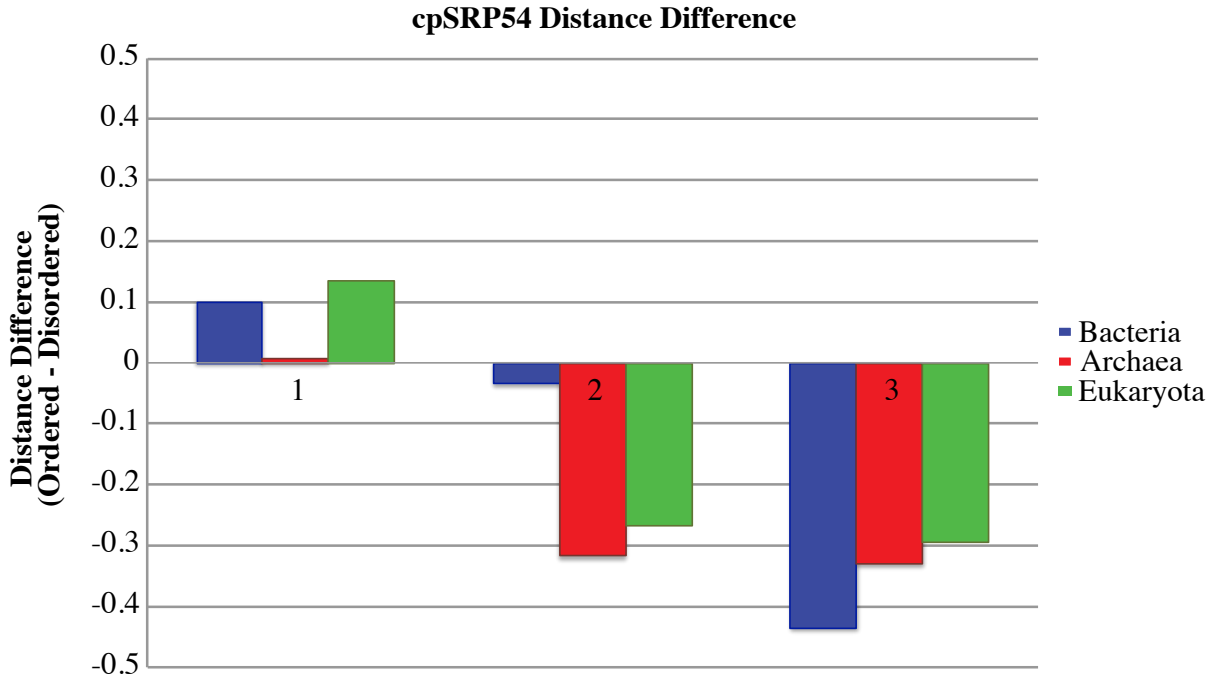

Figure S4: A- Multiple sequence alignment of *T. aq*. Ffh, cpSRP54, *S. sol*. SRP54 *& P. fur*. SRP54 (PDB ID 3DM5) sequences depicting the residues at the interaction site (2FFH residues highlighted in black). Comparison of the sequences reveals 5 sites which are different in *S. sol*. SRP54 and *Pyrococcus furiosus* (red) but not in cpSRP54 (green). Changes in these residues likely contribute to the difference in M-domain positioning observed in crystal structures. In blue is the critical residue, L367, for the formation of a buried hydrophobic pocket between the M-domain and G-domain in 2FFH. The alignment indicates this residue is hydrophilic in all three except *T. aq*. Ffh. **B-** Distance difference for SRP54/Ffh: (a) M-domain to NG-domain interaction site (b) G-domain to M-domain linker (c) M-domain finger loop.

Figure S5.

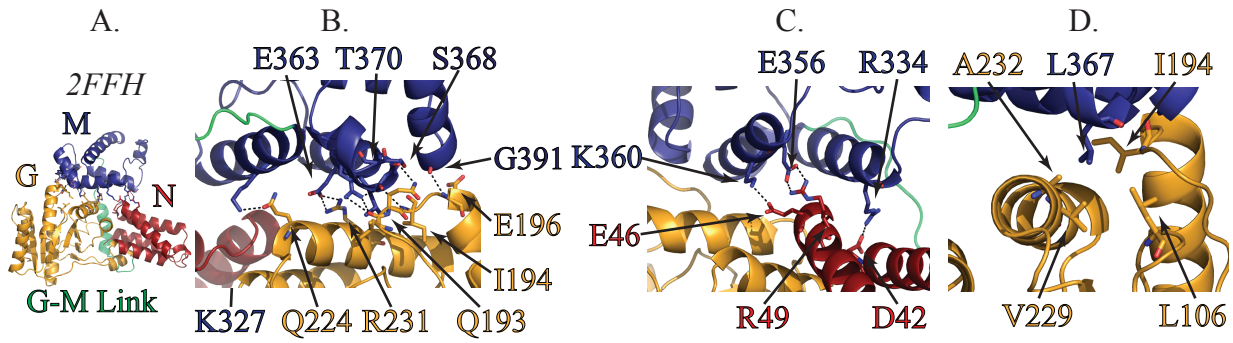

Figure S5: A- *T. aq*. Ffh (2FFH) M-domain to NG-domain interaction site from molecular dynamics simulation: **B-** the hydrophobic core between the G & M-domains, **C-** the G & M-domain polar contacts, **D-** and the N & M-domain polar contacts.

Figure S6.

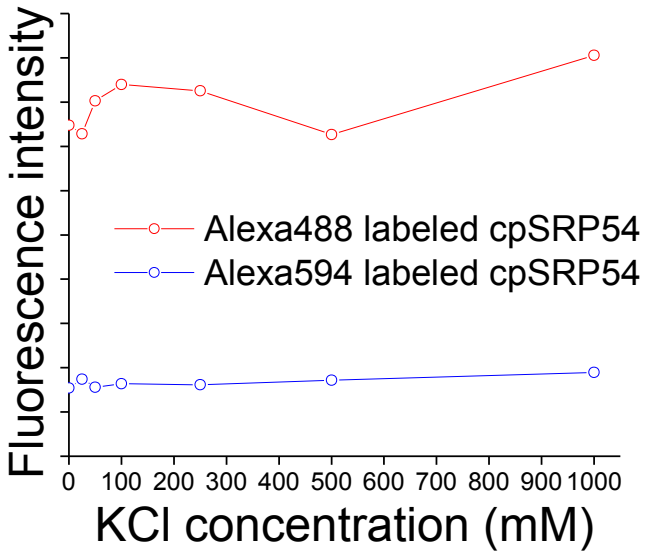

Figure S6: Ensemble fluorescence intensity at different concentrations of KCl for Alexa 488 labeled cpSRP54 and Alexa594 labeled cpSRP54. The results indicate increasing concentrations of KCl do not significantly affect the fluorescence intensity of the dyes.

Figure S7.

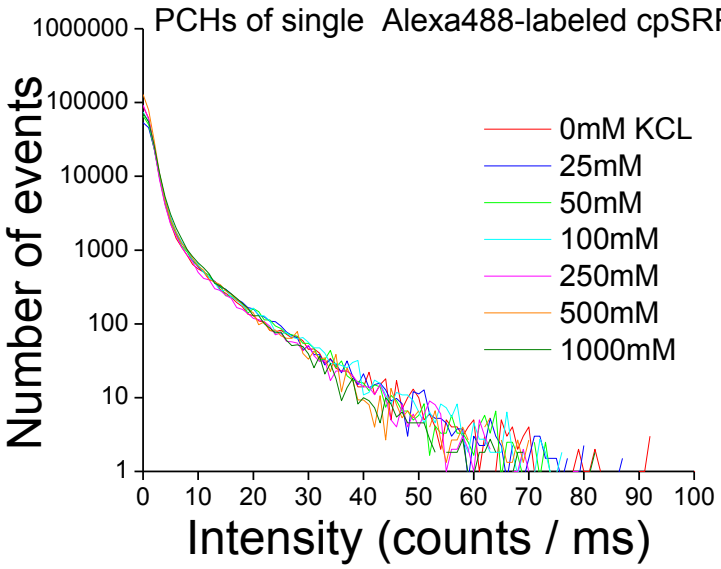

Figure S7: Single photon-counting histograms at different concentrations of KCl. The results indicate increasing concentrations of KCl do not significantly affect the photon counts.

Figure S8.

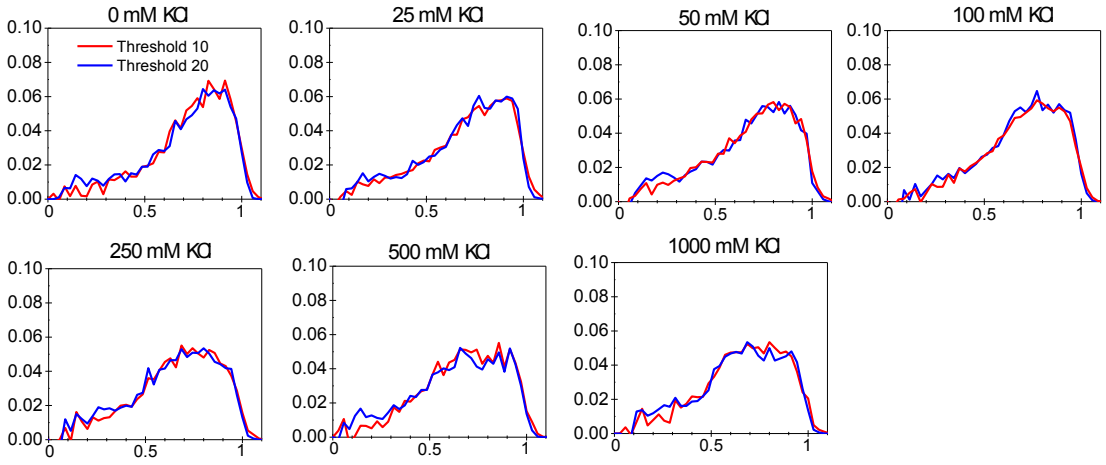

Figure S8: FRET histograms resulting from using 10 (red) and 20 (blue) counts/ms thresholds. The results indicate the threshold has little effect on the shape of the histogram at the various KCl concentrations used.

Figure S9.

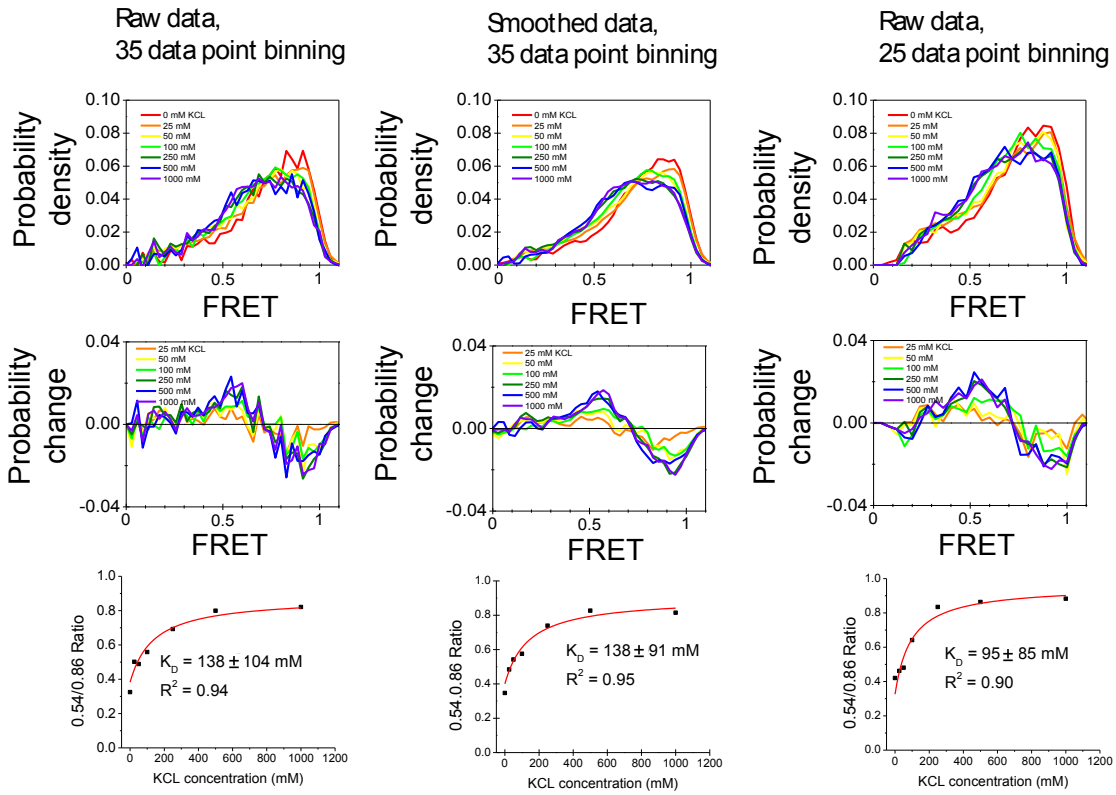

Figure S9: Binning and smoothing of the KCl smFRET data. The results indicate bin size has little influence on the overall shape of the histograms. Further, smoothing of the data does not affect the overall shape and, therefore, the visual interpretation of the results.

Figure S10

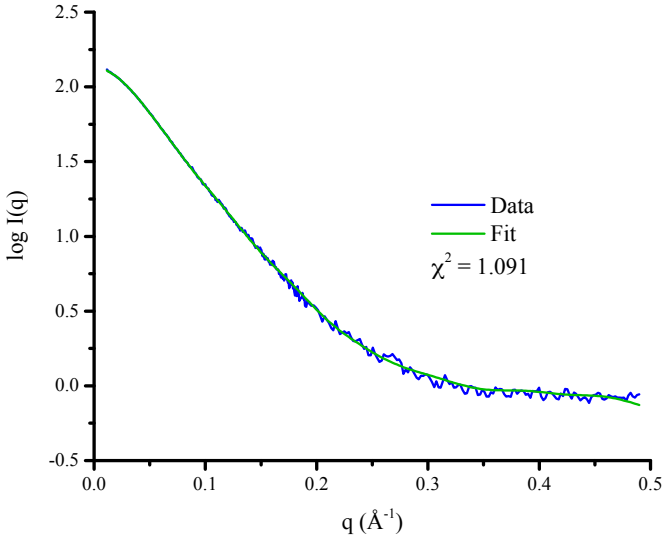

Figure S10: M-domain SAXS fit. The blue line represents the data while the green line represents the fit to the data.

IV. Ribosome Mimicry: cpSRP54 Scans the cpSRP43 "Exit Tunnel" for TM3

Rory C. Henderson R.C.H, Feng Gao F.G., Srinivas Jayanthi S.J., Alicia Kight A.K., Robyn L. Goforth R.L.G., Colin D. Heyes C.D.H., Ralph Henry R.H., and Thallapuranam Krishnaswamy Suresh Kumar T.K.S.K.
1. Department of Chemistry & Biochemistry 2. Department of Biological Sciences
University of Arkansas, Fayetteville, 72701
* Corresponding Authors

Abstract

The chloroplast signal recognition pathway is responsible for delivering light harvesting chlorophyll binding proteins post-translationally to the thylakoid membrane. The central binding machinery, cpSRP, is composed of an SRP54 homologue cpSRP54 and a novel 43kDa subunit cpSRP43. Previous FRET data demonstrated that both proteins were quite flexible even upon interaction with one another. Here, we examine the domain orientations of the cpSRP subunits using small angle x-ray scattering and FRET probability distribution analysis, revealing a large amount of flexibility between the domains of cpSRP54. Specifically, a C-terminal extension of the cpSRP54 M-domain allows the M-domain to sample the region of cpSRP43 where the third transmembrane segment of LHCP is expected to reside. These results help to rationalize the mechanism by which LHCP is loaded onto cpSRP to form transit complex and to predict a possible transit complex structure.

Introduction

The chloroplast signal recognition particle (cpSRP) is a signal recognition particle (SRP) homologue responsible for the post-translational the delivery of light harvesting-chlorophyll binding proteins (LHCPs) to the thylakoid membrane.[1] The chloroplast 54kDa subunit, cpSRP54, bears significant sequence and functional similarity to SRP54 found in the co-translational, ribosome bound SRP pathway of the cytosol.[1,2] Significantly, rather than the massive scaffold the ribosome and SRP RNA provide\, cpSRP54 operates with a novel 43kDa subunit, cpSRP43.[2] cpSRP43 provides the majority of the interaction interface with LHCP leaving a small segment of its third transmembrane segment (TM3) available for interaction with cpSRP54's M-domain.[3,4] In addition to providing this interface, cpSRP43 has been observed to

replace SRP RNA, stimulating the GTPase activity of cpSRP54 and cpFtsY.[2] However, it remains uncertain how cpSRP54 has been retrained to apply its co-translational signaling functionality in the absence of the typical ribosome machinery.

A recent surge in information regarding the molecular details of cpSRP signaling have revealed much information regarding the impact of interactions on cpSRP43's structure and signaling influence. cpSRP43 was first revealed to be quite flexible. Further, interaction with a cpSRP54 derived peptide significantly reduce some of this flexibility while increasing the affinity of a critical cpSRP43 binding segment of LHCP, L18.[5] This result was further substantiated, revealing greater detail regarding the specific changes occurring upon this interaction.[6] Additionally, it was shown that Albino 3 (M2 and M4) derived peptides could reverse this increased L18 affinity suggesting a mechanism for release of LHCP to the membrane. The M4 region of Alb3 was recently shown to tether cpSRP43 at the second and third chromodomains (CD2 and CD3, respectively).[7]

While there has been a significant increase in our understanding of cpSRP43's mechanistic role in LHCP targeting, details of the cpSRP54 three-dimensional structure as well as the full cpSRP structure have remained elusive. Recent work has revealed cpSRP54 is highly dynamic. (Ch. III.) Specifically, the cpSRP54 N-terminal, N-domain and GTPase, G-domain cycle through interactions with the methionine rich, M-domain, regulated by two flexible segments of the protein. The results of this study indicated that cpSRP43 causes a transition in the conformational composition of cpSRP54. Further, it was found that a reduction in the hydrophobicity of the M-domain resulted in difficulty in forming transit complex. The results suggested that M-domain forms an interaction with TM3 of LHCP.(Ch. III.)

Nevertheless, the molecular mechanism by which the M-domain facilitates loading of LHCP remains uncertain. In the ribosome bound co-translational system, SRP54/Ffh samples the ribosomal exit tunnel scanning for signal sequences.[8] We show here using small angle x-ray scattering (SAXS) that an intrinsically disordered, C-terminal segment of cpSRP54 allows the cpSRP54 M-domain to scan the region of cpSRP43 in which TM3 is expected to be exposed. Results of ensemble FRET experiments confirm cpSRP is characterized by a large amount of flexibility while probability distribution analysis (PDA) of the previous single molecule FRET experiments for cpSRP54 reveal that cpSRP43 alters the organizational equilibrium to favor a decoupled M-domain from the NG-domain. The results provide a mechanistic picture of the LHCP loading process and reveal that cpSRP43 and cpSRP54's C-terminal extension act functionally to replace the ribosomal scaffolding. Further, these results lead to the speculation of a transit complex structure.

Methods

Expression and Purification

Expression and purification of cpSRP43, cpSRP43ΔCD3, cpSRP54 M-domain, and cpSRP54 were performed as described.[9]

Ensemble FRET

Ensemble FRET titration experiments were performed titrating cpSRP43 into cpSRP54. cpSRP54 was labeled with the donor Alexa488 and cpSRP43 with the acceptor Alexa594, with a dye to protein ratio of 10 to minimize the nonspecific fluorescence labeling. Excess free dye was removed from labeled proteins via size exclusion chromatography. Fluorescence Correlation

Spectroscopy (FCS) experiments for free Alexa594 and cpSRP43 protein labeled with Alexa594 were performed in order to confirm that size exclusion chromatography effectively separated the free dye from the labeled protein. Calculations determined less than 4% free dye was present in the cpSRP43-Alexa594 solution using a 2-component diffusion model. The Alexa488 labeled cpSRP54 protein concentration for the FRET experiments was fixed at 50 nM. with titration of Alexa594 labeled cpSRP43 or unlabeled cpSRP43 from 0 nM to 1000 nM. Titration data for unlabeled cpSRP43 was used to correct for quenching by the unlabeled protein. The fluorescence spectra were measured using Photon Technology International (PTI) Quantamaster 4 fluorometer.

The FRET efficiency is determined using

$E = 1 - (I_D / I_{D0})$

where I_D is the intensity of the donor in the presence of the acceptor labeled proteins or unlabeled proteins, and I_{D0} is the initial intensity of the donor. To remove any quenching caused by the unlabeled protein, the unlabeled cpSRP43 FRET values were subtracted from the labeled cpSRP43 FRET values. Each point was then divided by the labeling efficiency of cpSRP43-Alexa594. The data was fit using the Hill equation, which is $(V_{max} * x) / (k + x)$, where x is the concentration of Alexa594 labeled cpSRP43 protein. The fitted graphs provide a maximum FRET value (the FRET value of the fully-bound complex) and a k_D value. The maximum FRET is used to calculate the corresponding average interdomain distances of the cpSRP43 and cpSRP54 heterodimer. The FRET distance equation is $E = R_0^6 / R_0^6 + r^6$, where E is the maximum FRET value, and r is the FRET distance between the dyes. R_0 is corrected based on the Alexa488 quantum yields of Alexa488 labeled cpSRP54 proteins involving FRET, with the given Förster distance for the particular dye pair based on the quantum yield of the free dye (60

Angstroms or 6nm). The obtained k_D values are similar to those obtained by ITC experiments on unlabeled proteins, supporting that the fluorescent dye does not affect the binding of cpSRP43 and cpSRP54.

Single Molecule FRET

Data collection and processing for the smFRET experiments were as described.(Ch. III.)

SAXS Data Collection & Fitting

SAXS data for cpSRP were collected at the Cornell High Energy Synchrotron Source (CHESS) as described at the G1 station.[10] A concentration series from 0.25 mg/ml to 7 mg/ml was collected using ten one-second exposures each to a *q* range of 0.8 $Å^{-1}$. Data was processed in the RAW software and visualized in the ATSAS package's Primus which was used for extrapolation to zero.[11] SAXS data for cpSRP43, cpSRP43ΔCD3 bound to cpSRP54 M-domain, and cpSRP43 bound to cpSRP54 M-domain were collected at the SIBYLS beam line as described.[12,13] A concentration series from 1 mg/ml to 3 mg/ml was collected for each with a 6 second exposure. SIBYLS data was processed in Primus. The Pored volume, distance distributions and plots were obtained using Primus. The volume of correlation (V_c) was determined using Scatter.[14]

Flexible fitting of the data was accomplished using BILBOMD for conformational ensemble generation while FoXs and MES were used for SAXS profile calculation and fitting.[15-17] Briefly, BILBOMD uses accelerated all-atom molecular dynamics simulation to sample configurational space according to a user selected radii of gyration (R_g) distribution. The full cpSRP43 structure was built from the folded segments comprised of a CD1-CD2 crystal

structure (PDB ID 3UI2) and an NMR CD3 structure (PDB ID 1X3P).[18,19] Residues 1 to 29 and 258-267 were assigned as disordered chains as 1 to 29 were predicted to be disordered according to MobiDB and previous FRET experiments indicate CD3 remains flexible when cpSRP54 is bound.[5,20] For fitting of cpSRP43 the segment between the first and second helices of Ank4, residues 190 to 197, were permitted to be flexible, leaving the possibility of displacement of the second Ank4 helix. CD3 was not included in the cpSRP43ΔCD3 construct. The previously developed homology model for cpSRP54 was used in this analysis.(Ch. III.) The M-domain construct used in fitting of the cpSRP43 and cpSRP43ΔCD3 bound data was as described.(Ch. III.) The NG-domain and M-domain core were kept folded while the G-domain to M-domain linker (residues 296 to 330), M-domain finger-loop (residues 345 to 365), and the C-terminal extension of the M-domain (residues 458 to 471) were regarded as intrinsically disordered as was determined using SAXS. The cpSRP structure was built from these using the previously determined crystal structure in which a short peptide of the C-terminal extension was bound to the linking segment between cpSRP43's Ank4 and CD2. This allowed the cpSRP54 C-terminal extension to remain bound to cpsRP43 in the simulations.

For cpSRP43, cpSRP43ΔCD3 bound to cpSRP54 M-domain, and cpSRP43 bound to cpSRP54 M-domain, a R_g range of 20 Å to 65 Å was used initially to find the optimum R_g space for sampling. This was determined for each construct to be 35 Å to 55 Å. Final runs for the analysis of the SAXS for cpSRP43 M-domain bound and unbound used this range. For cpSRP, an R_g range of 10 Å to 100 Å was used initially to find the optimum R_g for sampling. This was determined to be 35 Å to 65 Å. Final runs for the analysis of SAXS for cpSRP used this range. A total of 10 runs were used to gather statistics on the region sampled by both the M-domain and NG-domains. A total of 800 configurations were sampled in each iteration for all constructs.

After each run the theoretical scattering patterns of each conformation were calculated using FoXs and a minimum ensemble of structure capable of fitting the SAXS data were calculated using a genetic algorithm, MES. The distances were then calculated and related to the ensemble FRET experiments.

FRET PDA

smFRET histograms were fit using the probability distribution analysis (PDA) method as described.[21] Briefly, this method determines theoretical rates of interconversion between two or more states when this rate is on the timescale of the molecules diffusion through the confocal volume. FRET efficiencies for the states and an initial guess for the rate constants are provided to the algorithm which then minimizes the sum of squared error (SSE) between the theoretical and experimental histogram by changing the rate constants. The algorithm was implemented in C++ with the calculated FRET efficiencies of the modeled structures used as inputs (source code will be made available via the corresponding author's webpage). Specifically, FRET efficiencies of 0.91 (+/- 0.06), 0.70 (+/- 0.06) and 0.25 (+/- 0.12) were used. A larger standard deviation was used for the 0.25 FRET state as it corresponds to a decoupled, flexible state. One-hundred iterations using random initial rate constants between $10s^{-1}$ and $1000s^{-1}$ were used to examine the minimization landscape and calculate the uncertainty in the rate constants and equilibrium distributions. The lowest five SSE for each were used for the final calculations.

Results & Discussion

cpSRP54 cycles rapidly between three states

We first examined the ability of the three hypothesized structural states to describe the data. (Ch. III.) Distances of 45Å, 54Å, 72Å determined from those structures, equivalent to FRET efficiencies of 0.91, 0.54, and 0.25, respectively, were used in the fitting process. These states were termed S1, S2, and S3, respectively. State S1 corresponds to a state similar to that observed in a crystal structure for *T. aquaticus* (PDB ID 2FFH) in which the M-domain sits atop the NG-domain.[22] State S2 corresponds to a state similar to that occupied in an *S. solfataricus* crystal structure (1QZX) whose M-domain is stabilized by the G-domain to M-domain linker near the N-domain.[23] The final state, S3, corresponds to a state in which the M-domain is fully decoupled from the NG domain and is therefore able to sample configurational space limited only by the length of the G-domain to M-domain linker. Interconversion between states S1 and S2 were not allowed as these would require a relatively complex mechanism of transition and would likely be quite slow compared to the diffusion timescale (~1.4ms) (Figure 1). Additionally, a greater FRET efficiency distribution used for state S3 as it corresponds to a highly flexible, decoupled state.

The results for free cpSRP54 indicate rapid exchange between states S1 and S3 with a rate of 830 s^{-1}(+/-94 s^{-1}) for k_1 and 985s^{-1}(+/-160 s^{-1}) for k_{-1}. The rate of transition between states S3 and S2 is markedly slower with a rate of 101 s^{-1}(+/-38 s^{-1}) for k_2 and 396 s^{-1}(+/-120 s^{-1}) for k_{-2}. The rates result in a relatively low equilibrium population of state S2, at 10 percent, with higher populations for states S1 and S3, at 51 percent and 39 percent. Distinct changes in the rates of interconversion between states occur upon addition of cpSRP43. The forward rate for transitions from S1 to S3, k_1, increases from 830 s^{-1}(+/-94 s^{-1}) to 926 s^{-1}(+/-145 s^{-1}) while the reverse rate, k_{-}

$_1$, decreases from 985 s^{-1}(+/-160 s^{-1}) to 552 s^{-1}(+/-203 s^{-1}). The rate of conversion from state S3 to S2, k_2, does not change significantly, 101 s^{-1}(+/-38 s^{-1}) to 192 s^{-1}(+/-91 s^{-1}), while the reverse rate increases from 396 s^{-1}(+/-120 s^{-1}) to 743 s^{-1}(+/-163 s^{-1}). The result of the change in interconversion rates upon addition of cpSRP43 is that the equilibrium occupancy of state S1 is reduced from 51 percent to 39 percent with an increase in the occupancy of state S3 from 39 percent to 48 percent. State S2 is little changed with an increase from 10 percent to 13 percent (Figure 2). It is important to note that the minimization algorithm may find rather shallow minima with multiple solutions of similar fit to the data. Indeed, the results here produced a relatively wide distribution in the minima. Nevertheless, while the actual values vary somewhat about the average, the trends in each individual fitting yield similar results providing greater confidence in the overall picture of the kinetic scheme employed in the sampling.

The results indicate the decoupled state is favored in the presence of cpSRP43. The most significant changes in the rates show the transition out of S3 to S1 is slowed with an increase in the rate of transition from S2 to S3. The increased population of the S3 state without a reduction in the S3 to S2 rate allows for a steady S2 population *albeit* with an apparent weakening of the interaction since transitions out of S2 occur more readily. As these changes in cpSRP represent the particle just prior to loading of LHCP, it is apparent a highly mobile structure is need for proper cargo loading.

The cpSRP54 M-domain remains flexible when bound to cpSRP43

From the distances, transition rates, and equilibrium occupancies alone it is difficult to discern the significance of the structural states being sampled. Therefore, we next used small angle x-ray scattering to investigate the positioning of the domains of each protein relative to one

another further. In order to dissect changes in the domains of cpSRP43 and cpSRP54, we investigated free cpSRP43, cpSRP43ΔCD3 and each bound to the M-domain prior to investigating the full cpSRP structure.

Previous SAXS fitting of cpSRP43 suggested the protein was a largely rigid, extended structure.[24] However, recent smFRET data demonstrated cpSRP43 to be quite flexible.[5] In light of the smFRET data we decided to reinvestigate the cpSRP43 structure via SAXS. The Kratky plot for the data comprises a large peak followed by a rise at higher *q* values indicative of a protein with both folded and flexible segments (Figure 3). Further, calculation of the protein density from the Porod volume and V_c base molecular weight calculation reveal a density of 1.01g/ml, far lower than the average folded protein density of 1.37g/ml.[25,26] The Kratky plot and lower protein density are both strong indicators of flexibility and signify modeling via ensemble fitting is necessary. We therefore fit the data using BILBOMD/MES which combines accelerated molecular dynamics simulations with a genetic algorithm to sample configurational space and determine a set of structures capable of fitting the SAXS data. The best fit to the data was obtained using flexibility in the N-terminal tail before CD1, as is predicted by MobiDB, as well as in the second helix of Ank4, between Ank4 and CD2, and between CD2 and CD3 (Figure 3). These are all consistent with the previously obtained smFRET data. The best fit was obtained using 4 structures with RMSDs of 24.7 Å, 21.0 Å and 18.8 Å relative to the first arbitrarily selected structure and R_gs of 30.4 Å, 38.0 Å, 38.2 Å, and 37.3 Å for the first and remaining three, respectively. The distribution in the contribution of each structure to the scattering was relatively flat with 26.2 percent, 19.8 percent, 29.6 percent, and 24.5 percent for each, respectively. The overall composition of the structures reveals significant variability in the positioning of the domains and disorder in the N-terminal tail. Although the fitting method selected four structures

to represent the ensemble and since the states selected produce no significant contact between the domains, the cpSRP43 structure likely samples a continuum of states.

As a next step, we examined the affect of the M-domain on cpSRP43's flexibility and whether the M-domain adopts a particular conformation relative to cpSRP43 or samples multiple conformations. Further, we examined what affect CD3 might have on the positioning of the M-domain. Previous M-domain SAXS data revealed the M-domain C-terminal extension, which binds cpSP43 between Ank4 and CD2, is quite flexible.(Ch. III.) In both cases, the Kratky plots were indicative of flexibility in the protein complexes (Figure 4). The results of fitting for both cpSRP43 and cpSRP43ΔCD3 bound M-domain indicate that the flexibility in Ank4 and between Ank4 and CD2 is dramatically decreased, requiring no flexibility in the fitting. Further, the M-domain remains quite flexible when bound to cpSRP43 and cpSRP43ΔCD3, sampling conformational space to one side of cpSRP43. The best fit for cpSRP43 bound M-domain was obtained using 4 structures with RMSDs of 24.4 Å, 12.6 Å, and 29.5 Å relative to the first arbitrarily selected structure and R_gs of 41.0 Å, 47.8 Å, 41.5 Å, and 43.2 Å for the first and remaining three, respectively (Figure 4). The largest occupied state was the 41.5 Å Rg structure at 38.9 percent with the first second and fourth structures at 24.3 percent, 23.2 percent, and 13.7 percent, respectively. The ensemble of structures reveals the M-domain C-terminal extension is disordered except in the cpSRP43 binding region. This allows the M-domain to sample configurational space to one side of cpSRP43 near ankyrin repeat 4 helix-2.

The best fit for the cpSRP43ΔCD3 was obtained using 4 structures with RMSDs of 19.9 Å, 28.9 Å, 40.5 Å relative to the first arbitrarily selected structure and R_gs of 44.2 Å, 44.4 Å, 32.6 Å, and 31.9 Å for the first and remaining three, respectively (Figure 5). A fairly even distribution in the occupancy for the first three states was obtained, yielding 25.5 percent, 26.0

percent, and 32.9 percent for each, respectively. The final state was slightly lower in population at 15.7 percent. Interestingly, the M-domain samples regions of configurational space around cpSRP43ΔCD3 not sampled in the cpSRP43 bound M-domain suggesting CD3 may modulate the positioning of the M-domain. This region is specifically around the L18 binding site suggesting CD3 might prevent L18 competition with the M-domain. Together, these data confirm the flexibility observed in the smFRET experiments for cpSRP43 and indicates that the cpSRP43 binding M-domain peptide represents the majority of M-domain contact with cpSRP43, leaving the M-domain core quite flexible.

The cpSRP54 scans the TM3 exit site of cpSRP43

We next examined what influence the NG-domain might have on the cpSRP structure. The cpSRP54 smFRET experiments showed significant flexibility exists between the M-domain and NG-domain even when bound to cpSRP43. As the M-domain is quite flexible in the absence of the NG-domain, it is possible that the NG-domain remains flexible with the M-domain or that it adopts a specific position on cpSRP43. Indeed, small SAXS data reveals cpSRP to be flexible as determined using the Kratky plot and the SAXS calculated density of 0.74g/ml. This is consistent with the previously obtained smFRET data for both cpSRP43 and cpSRP54.[5] (Ch. III.) As in the previous results, the best fit was obtained using 4 structures (Figures 6 and 7). The RMSDs for the structures were 15.5 Å, 31.7 Å, and 37.1 Å relative to the first arbitrarily selected structure. The R_gs for these structures were 44.9 Å, 44.4 Å, 40.9 Å, and 58.8 Å for the first and remaining three, respectively. A relatively even distribution in the occupancy of each state was obtained, yielding 34.6 percent, 27.3 percent, 17.0 percent, and 21.1 percent for each, respectively. The results reveal the M-domain and NG-domain of cpSRP54 samples the region

near the theoretical "exit" site of LHCP TM3 (Figure 6). Rather than a single NG-domain orientation, the NG-domain position simply changes relative to the M-domain. Furthermore, the NG-domain possesses no apparent contact with cpSRP43 in the complex. Importantly, in none of the ensembles did the cpSRP54 domains sample the theoretical LHCP TM1 and TM2 interaction site near the L18 binding site.

In order to support the SAXS results, ensemble FRET was used with dyes on cpSRP54 and cpSRP43 in several pairs of domains. The FRET donor dye was placed in either the cpSRP54 N-domain or the cpSRP54 G-domain. Each labeled cpSRP54 construct was then titrated with acceptor labeled cpSRP43 with labels at Ank2 or CD3. The distances obtained were 6.66 nm, 8.84nm, 7.39nm, and 6.86 nm for N-domain to Ank2, G-domain to Ank2, N-domain to CD3, and G-domain to CD3, respectively. The population weighted distances for the cpSRP SAXS structures were 6.4 nm, 8.5 nm, 5.7 nm, and 7.7 nm for N-domain to Ank2, G-domain to Ank2, N-domain to CD3, and G-domain to CD3, respectively. The FRET distances are quite consistent with the cpSRP SAXS distances. There is some disagreement with regard to the positioning of the N-domain relative to CD3. This could be attributed to the uncertainty in the weights for each conformer in solution as compared to the FRET averaging in solution.

Together, the smFRET PDA data for cpSRP54 combined with the ensemble FRET and SAXS data as well as the previous cpSRP54 data reveal a remarkably complex series of interactions preparing cpSRP for interaction with LHCP. Previously, the cpSRP54 C-terminal extension responsible for binding cpSRP43 was demonstrated to be intrinsically disordered. (Ch. III.) However, it was unclear what advantage this disorder provided. The results here demonstrate the extension is used by cpSRP54 to allow the M-domain to scan the region where TM3 is expected to exit cpSRP43 based upon a CD1-Ank4 crystal structure bound to L18 and

cross-linking studies.[3,4] This is noteworthy in that it is known that the SRP54 and Ffh scan the ribosomal exit tunnel apparently searching for signal sequences.[8] The missing link in understanding cpSRP43's remarkable ability to replace SRP RNA and apparently the ribosome was this segment of cpSRP54 which effectively allows for the transitioning of the M-domain between scanning states while bound to cpSRP43. It was recently shown that LHCP can be fully integrated prior to GTPase activity and recycling of cpSRP for another round of LHCP targeting.[27] Structurally, it was uncertain how this could be accomplished with the crowding that would be present with all the targeting machinery still in place. The flexibility of the C-terminal extension allows the M-domain to not only scan for TM3, but also for cpSRP54 and cpFtsY to provide cpSRP43 with the conformational freedom to carry out this final integration step. Furthermore, the M-domain is effectively able to regulate the positioning of the NG-domains, as no contact between the NG-domain and cpSRP43 is observed.

It is interesting to note that the N-terminal tail of cpSRP43 is predicted to be disordered and was necessarily disordered for fitting of the SAXS data. It is known that deletion of the N-terminal tail through CD1 is capable of forming cpSRP and transit complex but is unable to facilitate integration.[9] The current model allows us to propose that, upon loading of LHCP, the cpSRP54 M-domain would force the NG-domains into close proximity to CD1 and the cpSRP43 N-terminal tail. A combined crystal structure, electron density map of SRP at the ribosome shows that the N-domain of SRP54 forms a close interaction with ribosomal proteins L35 and L23 which bear structural and sequence similarity to CD1-Ank1. Electrostatic potential maps of cpSRP54 and cpSRP43 strongly suggest an interaction between the N-domain loops opposite the N-domain and G-domain interface. A positively charged patch on cpSRP54 would complement the electrostatic potential of CD1-Ank1 and the negatively charged cpSRP43 tail.

The cpSRP54 M-domain was observed to increase the GTPase activity of cpSRP54 and cpFtsY in the absence of SRP RNA and cpSRP43.[28] In another study it was observed that CD1 deletion increased GTPase activity.[9] In this locked configuration the M-domain may be prevented from occupying the GTPase activating configuration. This would allow the cpSRP43 CD1 domain to effectively regulate M-domain activation of twinned GTPase activity. Further, as release of TM3 would destabilize this configuration, M-domain activation of GTPase activation would become possible once again since the M-domain would be permitted to sample configurational space. This would act as a signal to the GTPase domains that cargo has been delivered.

This is a provocative model for the transit complex as it could act as a signal to cpFtsY that cargo has been loaded and force the cpSRP54 GTPase domain in to a particular orientation relative to the remainder of the cpSRP structure. The proposed orientation would place the LHCP TM3 tail toward the membrane upon interaction with cpFtsY. Following release of TM3, the M-domain would once again be available for activation of GTPase hydrolysis, signaling the successful delivery of cargo and the need for recycling of the complex. Ultimately, this would provide a solid framework for understanding the mechanism by which LHCP is integrated at the membrane.

Conclusions

The present work demonstrates the remarkable versatility of an SRP54 homologue, cpSRP54. At some point during its evolution cpSRP54 was adapted to provide signaling and regulatory activity in the delivery of LHCPs with cpSRP43. Previous studies have restricted the discussion of cpSRP43 to being predominantly a replacement for specific binding partners in the

co-translational path. However, this study clearly demonstrates a combined adaptation of both cpSRP54 and cpSRP43 to one another. While cpSRP43 functions in a capacity similar to SRP RNA, it also provides a ribosome like scaffold for cpSRP54 interactions. The most striking adaptation of cpSRP54 is its C-terminal extension which binds cpSRP43. Through its marked disorder it allows the cpSRP54 M-domain to sample the region near where TM3 exits cpSRP43, thereby scanning for the TM in a manner similar to SRP54 scanning for signal sequences exiting the ribosome. In addition to cpSRP54 and cpSRP43 cooperative adaptation, cpFtsY itself is adapted to the absence of SRP RNA, prefolding into a state ready for interaction with cpSRP54.[29] We speculate here that cpSRP54's equilibrium of states acts as a regulatory element, disfavoring receptor interactions in the absence of cargo. We anticipate an additional transition once cargo is bound that would allow cpSRP54's NG-domain to be no longer occluded by the M-domain, thereby exposing the cpFtsY interaction site. As this final state would define a receptor orientation relative to the transit complex, we expect that this final state is critical for orienting the system at the membrane such that LHCP integration is as efficient as possible. Upon release of TM3, this configuration would be lost allowing the M-domain of cpSRP54 to activate GTP hydrolysis. The structure of this complex and its mechanistic significance remain to be demonstrated experimentally. This work nevertheless provides a solid framework for dissecting this intricate process, finally exposing the structural framework for LHCP targeting.

Acknowledgements

This research was generously supported by the Department of Energy (grant number DE-FG02-01ER15161) (to R.L.H. and T.K.S.K.), the National Institutes of Health/National Cancer

Institute (NIH/NCI) (1 RO1 CA 172631) and the NIH through the COBRE program (P30 GM103450), and the Arkansas Biosciences Institute. MD simulations are supported in part by the National Science Foundation through grants MRI #0722625 (Star of Arkansas), MRI-R2 #0959124 (Razor), ARI #0963249, #0918970 (CI-TRAIN), and a grant from the Arkansas Science and Technology Authority, with resources managed by the Arkansas High Performance Computing Center. Chess is supported by the NSF & NIH/NIGMS via NSF award DMR-1332208, and the MacCHESS resource is supported by NIH/NIGMS award GM-103485. The Advanced Light Source (ALS), a national user facility operated by Lawrence Berkeley National Laboratory on behalf of the Department of Energy, Office of Basic Energy Sciences, through the Integrated Diffraction Analysis Technologies (IDAT) program, supported by DOE Office of Biological and Environmental Research. Additional ALS support comes from the National Institute of Health project MINOS (R01GM105404).

References

1 Richter, C. V., Bals, T. & Schünemann, D. Component interactions, regulation and mechanisms of chloroplast signal recognition particle-dependent protein transport. *European Journal of Cell Biology* **89**, 965-973, doi:http://dx.doi.org/10.1016/j.ejcb.2010.06.020 (2010).

2 Akopian, D., Shen, K., Zhang, X. & Shan, S.-o. Signal Recognition Particle: An Essential Protein-Targeting Machine. *Annual Review of Biochemistry* **82**, 693-721, doi:10.1146/annurev-biochem-072711-164732 (2013).

3 Cain, P., Holdermann, I., Sinning, I., Johnson, A. & Robinson, C. Binding of chloroplast signal recognition particle to a thylakoid membrane protein substrate in aqueous solution and delineation of the cpSRP43–substrate interaction domain. *Biochem J.* **437**, 149-155 (2011).

4 Stengel, K. F. *et al*. Structural Basis for Specific Substrate Recognition by the Chloroplast Signal Recognition Particle Protein cpSRP43. *Science* **321**, 253-256, doi:10.1126/science.1158640 (2008).

5 Gao, F. *et al*. Regulation of Structural Dynamics within a Signal Recognition Particle Promotes Binding of Protein Targeting Substrates. *Journal of Biological Chemistry*, doi:10.1074/jbc.M114.624346 (2015).

6 Liang, F.-C. *et al*. Conformational dynamics of a membrane protein chaperone enables spatially regulated substrate capture and release. *Proceedings of the National Academy of Sciences* (2016).

7 Horn, A. *et al*. Structural basis for cpSRP43 chromodomain selectivity and dynamics in Alb3 insertase interaction. *Nature Communications* **6**, 8875, doi:10.1038/ncomms9875 (2015).

8 Holtkamp, W. *et al*. Dynamic switch of the signal recognition particle from scanning to targeting. **19**, 1332-1337 (2012).

9 Goforth, R. L. *et al*. Regulation of the GTPase Cycle in Post-translational Signal Recognition Particle-based Protein Targeting Involves cpSRP43. *Journal of Biological Chemistry* **279**, 43077-43084, doi:10.1074/jbc.M401600200 (2004).

10 Acerbo, A. S., Cook, M. J. & Gillilan, R. E. Upgrade of MacCHESS facility for X-ray scattering of biological macromolecules in solution. *Journal of Synchrotron Radiation* **22**, 180-186, doi:10.1107/S1600577514020360 (2015).

11 Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J. & Svergun, D. I. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *Journal of Applied Crystallography* **36**, 1277-1282, doi:10.1107/S0021889803012779 (2003).

12 Hura, G. L. *et al*. Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS). *Nat Meth* **6**, 606-612, doi:http://www.nature.com/nmeth/journal/v6/n8/suppinfo/nmeth.1353_S1.html (2009).

13 Putnam, C. D., Hammel, M., Hura, G. L. & Tainer, J. A. X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Quarterly Reviews of Biophysics* **40**, 191-285 (2007).

14 Forster, S., Apostol, L. & Bras, W. Scatter: software for the analysis of nano- and mesoscale small-angle scattering. *Journal of Applied Crystallography* **43**, 639-646, doi:10.1107/S0021889810008289 (2010).

15 Schneidman-Duhovny, D., Hammel, M. & Sali, A. FoXS: a web server for rapid computation and fitting of SAXS profiles. *Nucleic Acids Research* **38**, W540-W544, doi:10.1093/nar/gkq461 (2010).

16 Schneidman-Duhovny, D., Hammel, M., Tainer, John A. & Sali, A. Accurate SAXS Profile Computation and its Assessment by Contrast Variation Experiments. *Biophysical Journal* **105**, 962-974, doi:10.1016/j.bpj.2013.07.020 (2013).

17 Pelikan, M., Hura, G. L. & Hammel, M. Structure and flexibility within proteins as identified through small angle X-ray scattering. *Gen. Physiol. Biophys.* **28**, 174-189, doi:10.4149/gpb_2009_02_174 (2009).

18 Sivaraja, V. *et al*. Three-Dimensional Solution Structures of the Chromodomains of cpSRP43. *J. Biol. Chem.* **280**, 41465-41471, doi:10.1074/jbc.M507077200 (2005).

19 Holdermann, I. *et al*. Chromodomains read the arginine code of post-translational targeting. *Nat. Struct. Mol. Biol.* **19**, 260-263, doi:10.1038/nsmb.2196 (2012).

20 Potenza, E., Domenico, T. D., Walsh, I. & Tosatto, S. C. E. MobiDB 2.0: an improved database of intrinsically disordered and mobile proteins. *Nucleic Acids Research* (2014).

21 Santoso, Y., Torella , J. P. & Kapanidis, A. N. Characterizing Single-Molecule FRET Dynamics with Probability Distribution Analysis. *ChemPhysChem* **11**, 2209-2219, doi:10.1002/cphc.201000129 (2010).

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

23 Rosendal, K. R., Wild, K., Montoya, G. & Sinning, I. Crystal structure of the complete core of archaeal signal recognition particle and implications for interdomain communication. *Proc Natl Acad Sci U S A* **100**, 14701-14706 (2003).

24 Jaru-Ampornpan, P. *et al*. ATP-independent reversal of a membrane protein aggregate by a chloroplast SRP subunit. *Nat. Struct. Mol. Biol.* **17**, 696-702, doi:10.1038/nsmb.1836 (2010).

25 Rambo, R. P. & Tainer, J. A. Accurate assessment of mass, models and resolution by small-angle scattering. *Nature* **496**, 477-481, doi:10.1038/nature12070 (2013).

26 Rambo, R. P. & Tainer, J. A. Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod-Debye law. *Biopolymers* **95**, 559-571, doi:10.1002/bip.21638 (2011).

27 Sharma, P., Kight, A., Goforth, R. & Henry, R. GTP Hydrolysis is Not Required for Post-translational Protein Insertion by a Chloroplast Signal Recognition Particle. *The FASEB Journal* **29** (2015).

28 Jaru-Ampornpan, P., Nguyen, T. X. & Shan, S.-o. A Distinct Mechanism to Achieve Efficient Signal Recognition Particle (SRP)–SRP Receptor Interaction by the Chloroplast SRP Pathway. *Molecular Biology of the Cell* **20**, 3965-3973, doi:10.1091/mbc.E08-10-0989 (2009).

29 Stengel, K. F., Holdermann, I., Wild, K. & Sinning, I. The structure of the chloroplast signal recognition particle (SRP) receptor reveals mechanistic details of SRP GTPase activation and a conserved membrane targeting site. *FEBS Letters* **581**, 5671-5676, doi:http://dx.doi.org/10.1016/j.febslet.2007.11.024 (2007).

Figures

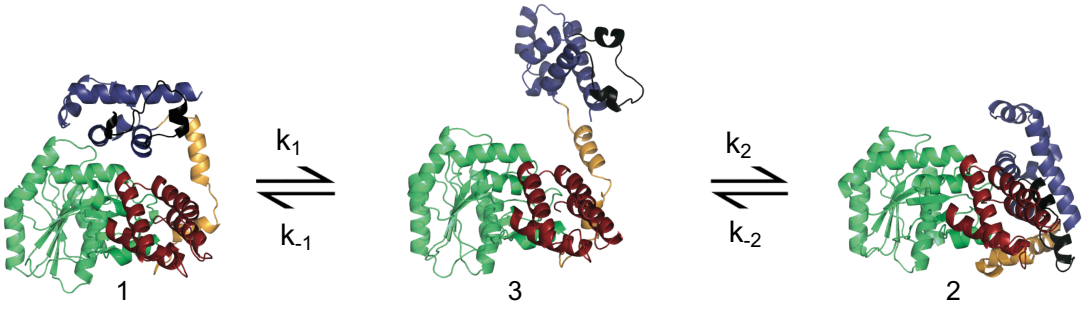

Figure 1: *(A)* The transition mechanism used in the cpSRP54 FRET PDA fitting. The *T. aq.* like state (left) transitions between a *S. sol.* like structure (right) through a decoupled intermediate (center).

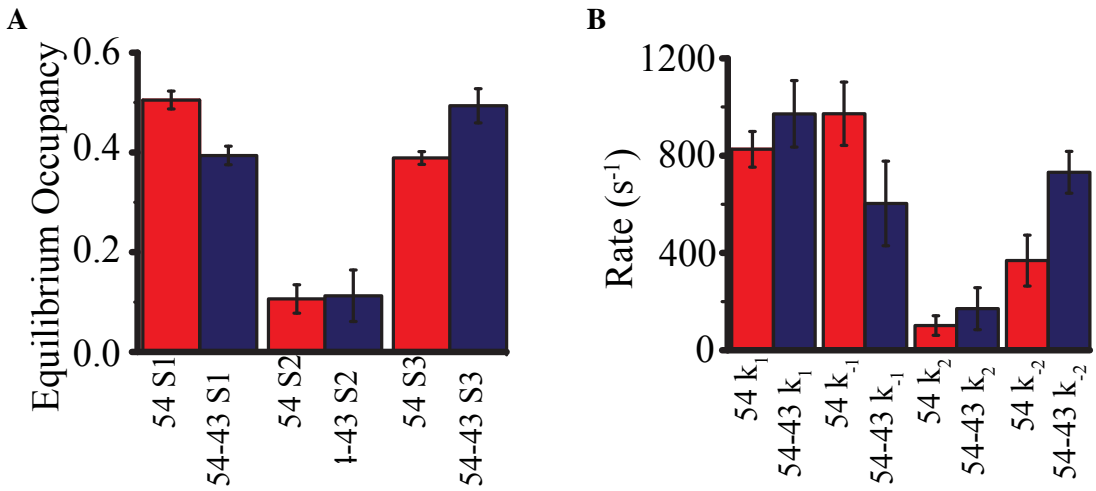

Figure 2: Results of the FRET PDA fitting for free cpSRP54 and cpSRP43 bound cpSRP54. *(A)* The equilibrium occupancy of each state used in the FRET PDA fitting for the cpSRP54 v. cpSRP54-43 data. The results indicate that the occupancy of state S1 is reduced upon interaction with cpSRP43 with a concomitant increase in the occupancy of state S3. State S2 remains unchanged. *(C)* Transition rates determined using the FRET PDA algorithm for free cpSRP54 and cpSRP54 bound to cpSRP43. Transition rates between states S1 and S3 are similar and quite rapid while the rates between S3 and S2 are relatively slow. Upon interaction with cpSRP43, the rate of transition from S3 to S1 is decreased with an increase in the transition rates between S2 and S3.

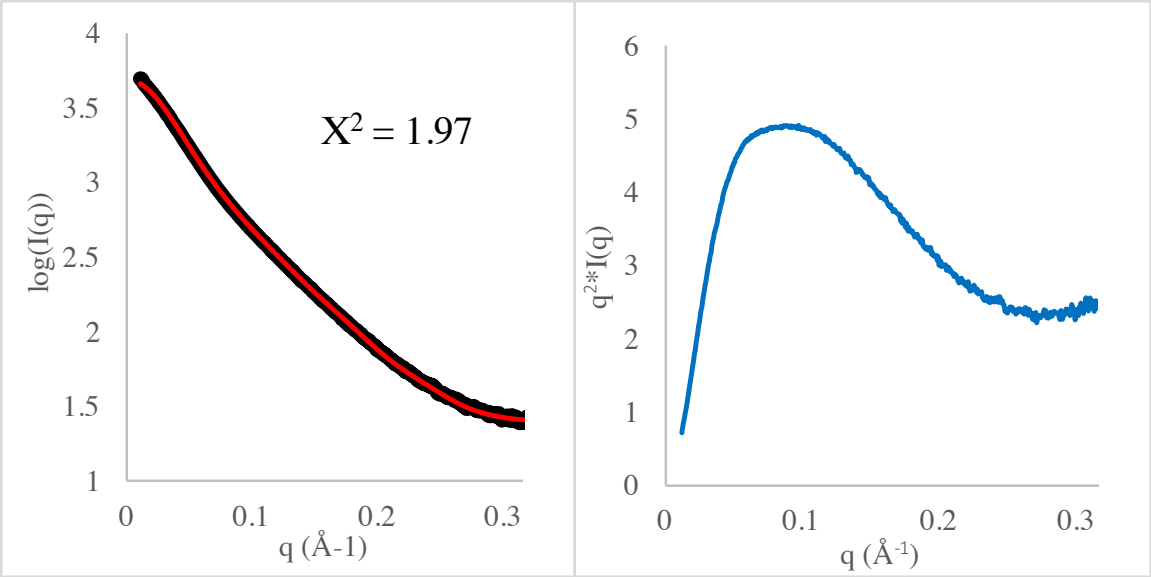

Figure 3: Kratky Plots and data fitting for cpSRP43. *(A)* Ensemble fit (red) to the data (black) yielding a X^2 of 1.97. *(B)* Kratky plot for the cpSRP43 data. The rise in the data at higher q is indicative of flexibility in the particle.

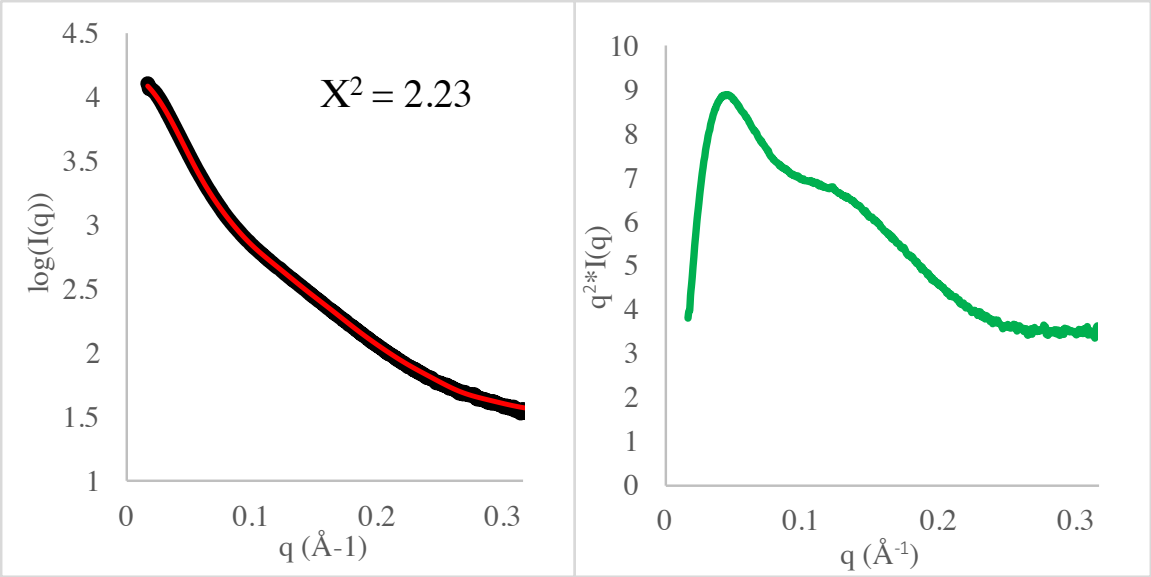

Figure 4: Kratky Plots and data fitting for cpSRP43 bound to the M-domain. *(C)* Ensemble fit (red) to the data (black) yielding a X^2 of 2.23. *(D)* Kratky plot for the cpSRP43 bound to the M-domain data. As the curve does not approach zero at higher q the particle contains some flexibility. However, comparison of this curve to that of cpSRP43 indicates a stronger peak suggestive of increased order in the system.

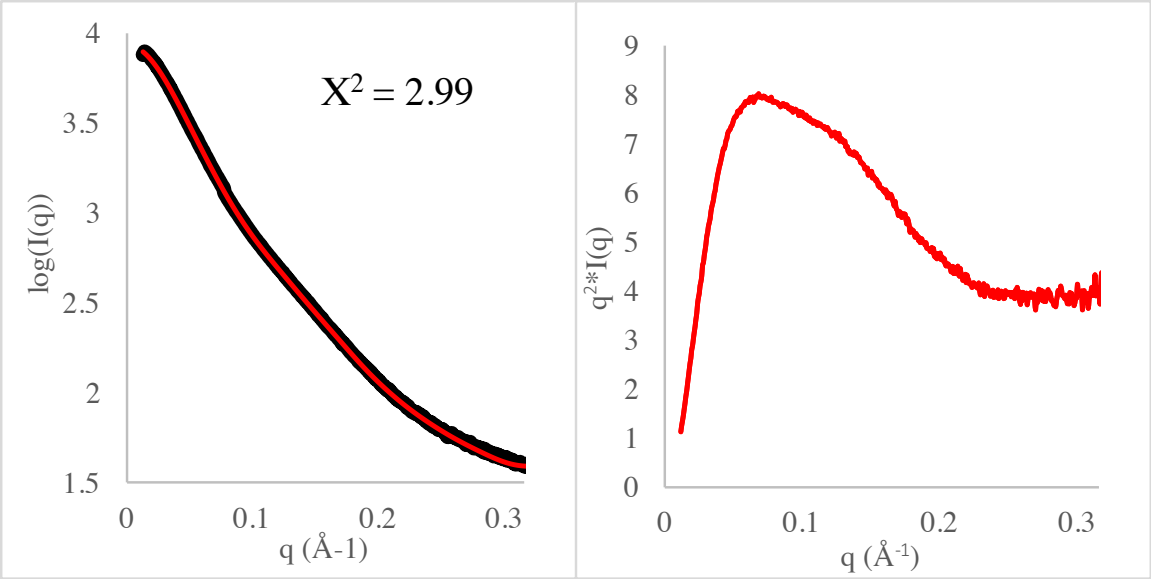

Figure 5: Kratky Plots and data fitting cpSRP43ΔCD3 bound to the M-domain. *(E)* Ensemble fit (red) to the data (black) yielding a X^2 of 2.99. *(F)* Kratky plot for the cpSRP43ΔCD3 bound to the M-domain data. The curve does not approach zero suggesting of disorder. Interestingly, comparison of this curve to that of cpSRP43 bound to the M-domain indicates a weaker peak suggestive of increased disorder in this system vs. the full length cpSRP43 bound to the M-domain.

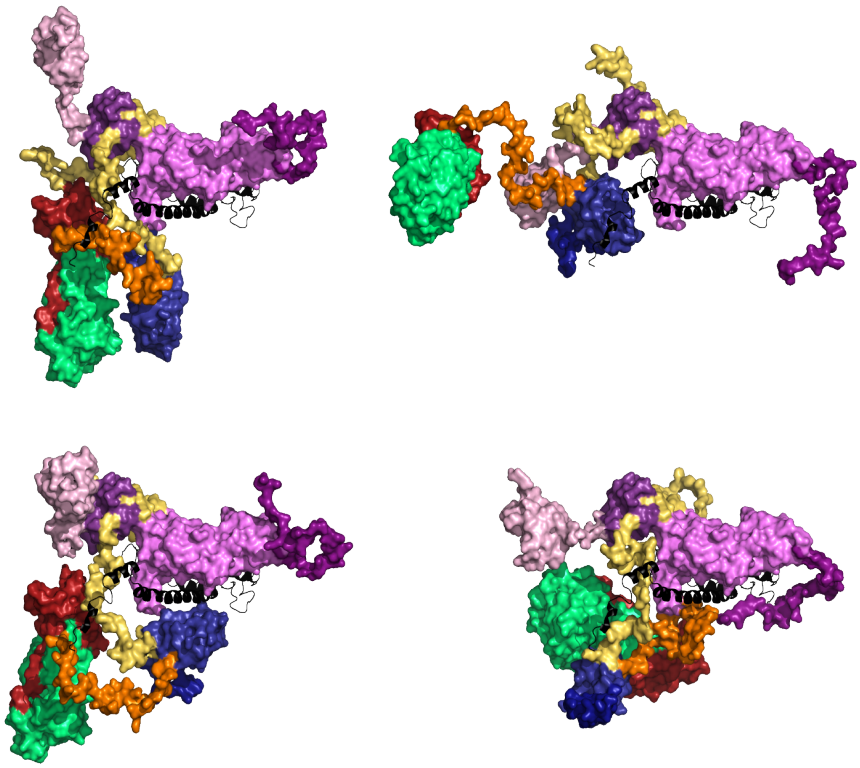

Figure 6: Structural states selected using BILBOMD-MES. The results show a strong preference for the M-domain (blue) to remain near Ank4 (pink) of cpSRP43 where TM3 (black) theoretically exits cpSRP43 (TM3 is shown for illustrative purposes only). Further, the results are consistent with the smFRET experiments, showing the NG-domain (red and green) of cpSRP54 and CD3 (light pink) of cpSRP43 are quite mobile in the complex.

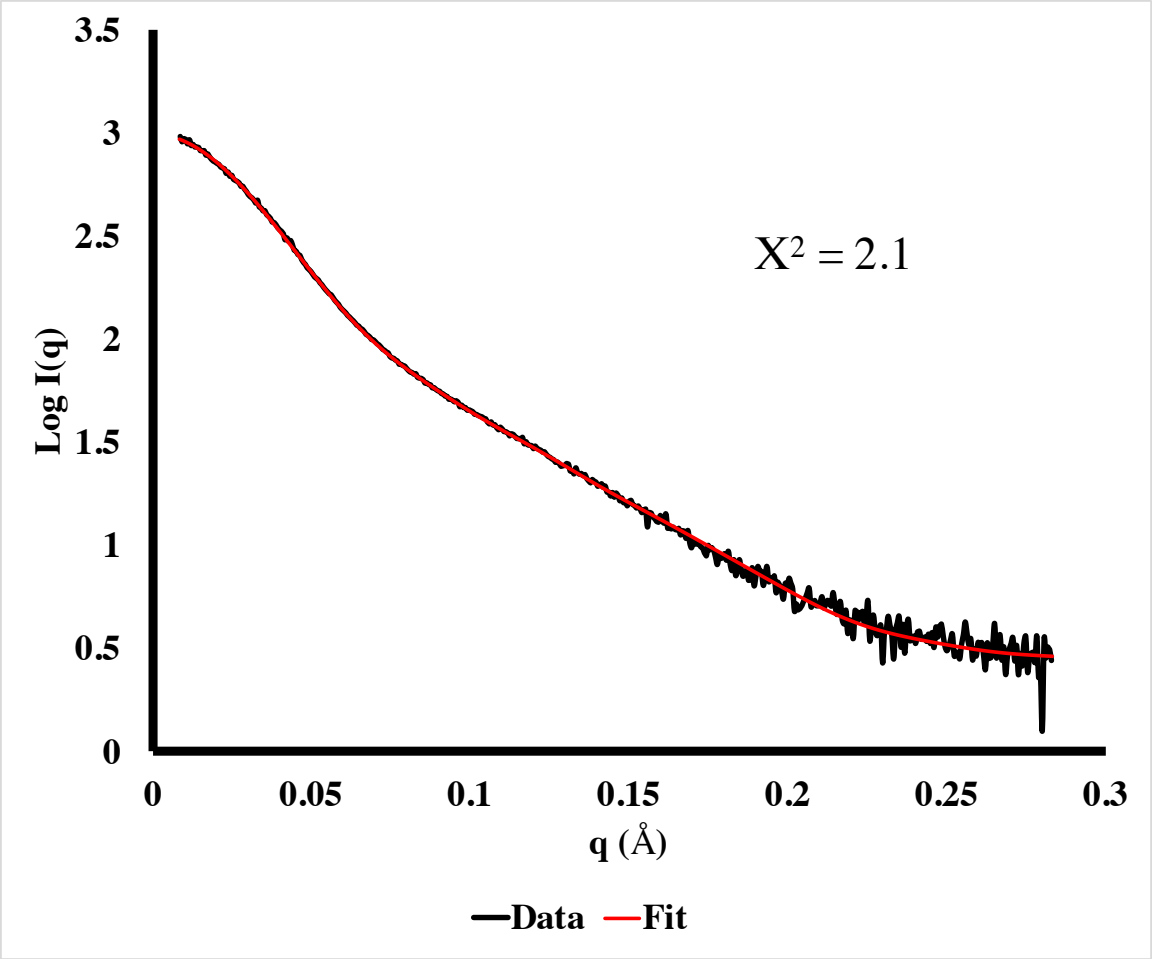

Figure 7: The fit (red) for the MES selected structure to the experimental data (black; $X^2 = 2.1$) indicates an excellent fit to the data.

V. An Efficient Algorithm for the Generation of Conformational Ensembles for Fitting of Small and Wide Angle X-ray Scattering Data

Rory Henderson R.H. Colin Heyes (CH), Ralph Henry (RH), and Thallapuranam Krishnaswamy Suresh Kumar T.K.S.K.

Abstract

Fitting of small angle x-ray scattering (SAXS) data for flexible proteins is accomplished *via* selection from large pools of possible conformations. However, data in the SAXS regime generally limits analysis to a low resolution, quantitative description of the ensembles. Recent interest in the wide angle scattering regime (WAXS) has lead to improvements in scattering profile determinations that may allow for studies to increase resolution of flexible systems. This requires the ability to generate all-atom ensembles of protein structures. Here we introduce an algorithm for the simple and rapid generation of all-atom protein ensembles of any degree of complexity. The results of test runs demonstrate that the structures produced by the algorithm reliably fit SAXS data as accurately as the most popular structure generating algorithm currently available. Furthermore, our implementation provides a simple to use method for the generation of all-atom structures of arbitrary complexity.

Introduction

Small angle x-ray scattering (SAXS) has become increasingly popular over the last decade as a method to probe the structure of biological macromolecules.(1, 2) SAXS is a solution technique in which the protein sample is illuminated by a highly collimated x-ray beam. The light is scattered due to differences in the electron densities of the solvent and protein. Subtraction of the solvent scattering from the protein solution scattering results in a scattering pattern related to the protein shape.(3) While the use of early experiments was limited to measures of broad physical properties of the protein(s), such and the radius of gyration, many computational tools have been developed to produce low resolution models of proteins shapes.(4) Additionally, rapid computation of accurate theoretical scattering patterns from

atomic models has allowed for testing of structures against the scattering data.(5, 6) SAXS is particularly well suited for highly flexible systems which often present significant challenges for study by traditional methods such as x-ray crystallography (7, 8). Fitting methods for flexible systems take advantage of the ability to calculate scattering patterns from models and algorithms for selecting from a large set of possible structures.(9, 10) A recent surge in interest in wide angle x-ray scattering (WAXS), scattering beyond $0.5Å^{-1}$, has brought solution x-ray scattering to significantly higher resolution.(2, 11) Although the information content of such solution scattering techniques is insufficient for high resolution structure determination, the full range of scattering data allows for the determination of whether a particular structure is consistent with the data.(1) When combined with additional biophysical techniques, structures of predictive value can be reliably produced.(12)

Several methods have been developed for fitting of SAXS data for flexible systems.(9, 10, 13-20) As WAXS data can probe down to the 3-4 Å level, it is necessary to produce all-atom structures. Methods for reliable prediction of WAXS curves from high resolution models have recently be implemented.(2, 5, 21-25). However, atomic level structures are required for these calculations. Currently available methods typically use a coarse grained approach for at least part of the system.(10, 15, 16, 18) Nevertheless, there exist atomic level methods for ensemble generation.(9, 14, 17, 20) These methods, however, are limited in the size and complexity of the system and can require significant computational resources to run efficiently. Here, we report the development of an efficient algorithm for rapid generation of all-atom protein structural ensembles. Specifically, our approach makes use of rigid body movement for folded domains, which can include RNA and DNA, followed by the addition of disordered connecting chains and termini via sampling of torsion angle space. Additionally, proteins and domains can be fixed in

place and user constraints can be used to ensure movable domains sample a specific region relative to either movable or fixed segments. Runtime is fast, producing ensembles usable (typically 10,000 structures) in hours on a single processor, while open source availability and compact design using a popular programming language allows users to easily modify the code to suit their needs without advanced molecular simulation knowledge.

Results and Discussion

The methionine rich M-domain of the 54 kDa subunit of the chloroplast signal recognition particle is predicted to contain three disordered segments. Further, the Kratky plot of the SAXS data shows a clear increase in $q^2*I(q)$ as q increases indicative of large scale flexibility. Therefore, as a test of the Conformational Ensemble Generator (CEG) algorithm, data for the M-domain out to a q of 0.5 $Å^{-1}$ was fitted using GAJOE (Genetic Algorithm Judging Optimization of Ensembles) with ensembles built using the ATSAS (All That Small Angle Scattering) RANdom Chain generator (RANCH; a part of the Ensemble Optimization Method [EOM]) or CEG. Both fit the data quite well with X^2 values of 1.091 and 1.046 respectively. (Figure 1:left) Notably, the fit using the RANCH C-α trace begins to diverge from the experimental curve above a q of 0.45 $Å^{-1}$ suggesting the C-α trace is problematic for fitting at wider angles. The overall size-distributions are quite similar despite the differences in the pools from which the structures were selected. (Figure 1: middle and right) This is to be expected because the overall size distribution is reflected in the SAXS region while folding details are contained in the WAXS region. Therefore, a good fit in the SAXS region, regardless of the level of detail in the structures, should produce reliable size distributions.

As a further test of the algorithm, glutathione-S transferase (GST) fusion proteins with GST-linked to chromodomain 2 of cpSRP43 (GST-CD2) or linked to chromodomains 2 and 3 (GST-CD2CD3) were tested using both the RANCH and CEG. Nuclear magnetic resonance experiments have demonstrated the chromodomains are flexibly linked to GST via the thrombin cleavage site. Further, single molecule Forster resonance energy transfer (smFRET) experiments have demonstrated CD2 and CD3 are flexibly linked.(26) The SAXS (to 0.28 $Å^{-1}$) fitting results indicate both algorithms produce a similar fit to the data (GST-CD2 $X^2 = 1.105$ and 1.164 for the EOM and CEG respectively; GST-CD2CD3 $X^2 = 1.560$ and 1.612 for EOM and CEG respectively). (Figures 2 and 3:left) The size distributions are again quite similar with a bimodal distribution produced using both ensemble generators for GST-CD2CD3. (Figures 2 and 3: middle and right) The fit at the lowest angles for GST-CD2CD3 is slightly lower than the experimental data, giving rise to higher X^2 values. This is likely due to differences between the crystal structure and solution structures of GST. While the distributions are similar for all constructs the frequency of the selected sizes differ slightly between the conformer generating algorithms. This may be due to the differences in the frequency with which the required conformers are generated in each algorithm or may be a fact of the degeneracy of solutions SAXS models produce. Nevertheless, both algorithms provide data of similar value in terms of size and distribution width.

Conclusion

Here, we have developed an efficient, user-friendly algorithm for the production of protein conformational ensembles of arbitrary complexity. We show that fitting of experimental data at small angles is as effective as the EOM method. Nevertheless, challenges remain to bring

the algorithm to full functionality. The current implementation does not provide the necessary restrictions to ensure backbone bond and torsional angles are physically realistic. Furthermore, the distribution in the distances sampled must be adjusted such that smaller distances are more heavily sampled. Future work will bring additional acceptable conditions for possible backbone conformations and bond angles. Further, current algorithms for conformer selection are restricted to the SAXS up to 0.5$Å^{-1}$ limiting the application of the conformational ensembles to the SAXS regime. Nevertheless, this algorithm provides the basis for structure prediction of flexible systems down to the atomic level using WAXS.

Experimental Methods

Expression and Purification

The cpSRP54 M-domain was produced as previously described.(27) The GST fusion proteins were produced as described previously.(28) The final buffer for the M-domain was 100mM KCl, 10mM HEPES and 10mM Mg2+ at a pH of 8.8. The final buffer for GST-CD2 and GST-CD2CD3 was 137mM NaCl and 10mM phosphate at a pH of 7.2.

SAXS Data Collection and Processing

Small angle X-ray scattering data was obtained at the Cornell High Energy Synchrotron Source (CHESS) as described at the G1 station.(29) For the M-domain, a concentration series from 1 mg/ml to 5 mg/ml was obtained out to a q of 0.5 $Å^{-1}$. The GST fusion proteins were collected at 1 mg/ml, 3 mg/ml and 5 mg/ml concentrations. Ten one second exposures were averaged for each protein and the buffer. Buffer data were collected before and after the concentration series. The RAW software was used to process the data which was visualized in

the ATSAS package's Primus.(30, 31) A slight concentration dependence was observed for the M-domain in addition to a shift at wider angles with data for concentrations above 3 mg/ml. Therefore, only the data for 1-3 mg/ml were used in Primus for extrapolation to zero concentration. No concentration dependence was observed for the GST fusion proteins. Fitting of the SAXS data was carried out using the the ATSAS package's GAJOE. The ATSAS package's ensemble generator RANCH was used for the generation of C-α flexible segment ensembles.(10) The M-domain was built using I-TASSER using *Thermus aquaticus* (PDB ID 2FFH) as a template.(32, 33) The flexible segments were defined as 296 - 327, 342 - 368, and 428 to 489 with numbering beginning at the first N-terminal residue of mature cpSRP54. The *Schistosoma haematobium* GST crystal structure (PDB ID 2c80) with *Arabidopsis thaliana* CD2 (PDB ID 1X3Q) or CD2CD3 (PDB ID 1X3Q and 1X3P) were used for the fusion proteins with the sequence ATPFLELVPRGS between GST and CD2 and the sequence GLEYAV between CD2CD3 assigned flexibility allowing the backbones to sample torsion angle space.(34, 35)

Conformational Ensemble Generation Method

The conformations are generated in a fashion similar to that employed in the first part of the ATSAS EOM method (RANCH) employing rigid body movement of the folded domains with addition of connecting chains. In the conformational ensemble generator (CEG) algorithm the user provides a configuration file that specifies the protein PDB file names, connectivity, whether the protein should remain fixed in position or move, and any atom-atom distance constraints for the ensemble generation. An additional file contains the flexible linking sequences and/or flexible N or C-terminal segments. After reading in the data from these files the coordinates from the PDB files are transferred to the origin to allow for rotation of the

molecules. For each iteration the program cycles through several steps to prepare each conformation. The steps are iterated until the user specified number of PDB files have been generated. The steps used are as follows:

1. Generation of translation and rotation matrices that do not violate distance constraints.
2. Movement of the specified protein domains.
3. Overlap check (defined as any two C-α's within 3Å).
4. If no overlap has occurred, flexible chain addition begins. The steps are as follows:
    a. Determination of a set of distance pairs, one along the vector connecting the first and last chain C-α's and one orthogonal to this vector.
    b. Consecutive placement of C-α atoms at these distances
    c. Addition of backbone atoms
    d. Checking of torsion angles
    e. Overlap check
    f. Side-chain addition after building of the entire segment backbone
5. Printing of the PDB file.

Protein backbone atoms are added in a fashion similar to that used in the Flexible Mecanno approach making use of the peptide planar geometry.(16) Torsion angles are used from the Top500 database.(36) Residue side-chains were obtained from SwissSideChain.(37) The equations used in the generation of the conformers are as follows:

A random translation matrix for the movable folded domains is selected according to:

EQ 1-3:

$$x = d \cdot \cos(\theta) \cdot \sin(\phi)$$
$$y = d \cdot \sin(\theta) \cdot \cos(\phi)$$
$$z = d \cdot \cos(\phi)$$

where *d* is a distance randomly selected from a uniform distribution between 0 and the maximum distance possible, θ and ϕ are randomly selected according to:

EQ 4-5:

$$\theta = 2\pi \cdot u$$
$$\phi = cos^{-1}(2v - 1)$$

where *u* and *v* are values randomly selected from a uniform distribution (a distribution in which all values in the range are selected with equal probability) between 0 and 1. The selection of ϕ ensures there is no bias in the rotation selection. Such bias is illustrated by the problem of point picking on a sphere where, since the area of a particular region of the sphere is defined by $d\Omega = \sin(\phi)d\theta d\phi$ is a function of , the points are more concentrated at the poles.

A random rotation matrix is selected using Euler angels according to:

EQ 6:

$$\begin{bmatrix} a_{11} & a_{12} & a_{31} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

where:

EQ 7-15:

$$a_{11} = \cos(\psi) \cdot \cos(\phi) - \cos(\theta)\sin(\phi)\sin(\psi)$$
$$a_{12} = \cos(\psi) \cdot \sin(\phi) + \cos(\theta)\cos(\phi)\sin(\psi)$$
$$a_{13} = \sin(\psi) \cdot \sin(\theta)$$
$$a_{21} = -\sin(\psi) \cdot \cos(\phi) - \cos(\theta)\sin(\phi)\cos(\psi)$$
$$a_{22} = -\sin(\psi) \cdot \sin(\phi) + \cos(\theta)\cos(\phi)\cos(\psi)$$
$$a_{23} = \cos(\psi) \cdot \sin(\theta)$$
$$a_{31} = \sin(\theta) \cdot \sin(\phi)$$
$$a_{32} = \sin(\theta) \cdot \sin(\phi)$$
$$a_{33} = \cos(\theta)$$

where θ and ϕ is a randomly selected according to:

EQ 16-17:

$$\theta = 2\pi \cdot k$$
$$\phi = cos^{-1}(2l\ - 1)$$
$$\psi = 2\pi \cdot m$$

where *k, l and m* are values randomly selected from a uniform distribution between 0 and 1. The selection of ϕ ensures there is no bias in the rotation selection.

A set of n_t distances to travel along the vector connecting the chain C-α's, where n_t equals the number of residues, is selected prior to chain building. It is designed such that:

EQ 18:

$$\sum_{i=1}^{n_t} \mu_a \approx d_t$$

where μ_a are the mean acceptance values used in the Gaussian distribution for distance selection, n_t is the number of residues, and d_t is the total distance to be covered. Distances are selected along the vector according to:

EQ 19:

$$\mu_a = (d_r \cdot n_r) - d_c$$

where d_r is defined as the distance to cover divided by n_t, and d_c is the distance covered. The distance is then sampled from the Gaussian distribution with acceptance probability:

EQ 20:

$$a = e^{-\frac{(d_a - \mu_a)^2}{\sigma_a}}$$

where d_a is the distance under consideration where d_a is randomly selected from a uniform distribution between -3.8 and 3.8 whose absolute values are the maximum distance a C-α can travel.(38) The variance, σ_a, is defined as:

EQ 21:

$$\sigma_a = e^{-\frac{d_t \cdot C}{n_t}}$$

where C is a constant. For each d_c an orthogonal distance to travel from the vector connecting the start and end points of the chain is determined such that:

EQ 22:

$$\sum_{i=1}^{n_t} \mu_b \approx b_t$$

where μ_b are the mean acceptance values used in distance selection and b_t is the maximum distance to be covered in the orthogonal direction defined as average protein C-α to C-α distance, 3.8Å. Distances are selected according to:

EQ 23:

$$\mu_b = (b_r \cdot n_r) - b_c$$

where b_r is the distance per residue, which is defined as half the distance to cover divided by n_r which equals 1.9Å, and d_c is the distance covered. The distance is then sampled from the Gaussian distribution with acceptance probability:

EQ 24:

$$b = e^{-\frac{(b_a - \mu_b)^2}{\sigma_b}}$$

where b_a is the distance under consideration and σ_a is defined as:

EQ 25:

$$\sigma_b = {c_r}^2$$

where c_r is the defined as:

EQ 26:

$$c_r = ({b_t}^2 - {b_a}^2)^{\frac{1}{2}}$$

The distance pairs are then reordered such that the chain is more likely to continue in its current direction, toward or away, from the vector connecting the C-α's with probability 0.75. The chain is first built along the z-axis beginning from the origin. The z-coordinate for each C-α is defined as the distance to travel along the C-α vector. The x and y-coordinates are calculated from orthogonal distance according to:

EQ 27:

$$x = c_r \cdot \cos(\theta_i)$$
$$y = c_r \cdot \sin(\theta_i)$$

where θ_i is defined using the law of cosines as:

EQ 28:

$$\pm\theta_i = cos^{-1}\left(\frac{{b_c}^2 + {b_a}^2 - {c_r}^2}{2 \cdot b_c \cdot b_a}\right)$$

This is the angle from the x-axis for the first C-α. Subsequent C-α use the prior angle as a reference, adding or subtracting to it to obtain the new angle.

After checking that the new point does not overlap (<3 Å) with any of the previous, the backbone atoms are added excluding hydrogen. These are added using the peptide planar

geometry.(16) The previous C-α_1 and current C-α_2 are used in the calculation, adding the CO of C-α_1 and N of C-α_2 in each iteration. The coordinates for each calculation are as follows:

EQ 29:

$$A_x = A_r \cdot \cos(\theta_c) \cdot \vec{u}_x + A_r \cdot \sin(\theta_c) \cdot \vec{v}_x$$
$$A_y = A_r \cdot \cos(\theta_c) \cdot \vec{u}_y + A_r \cdot \sin(\theta_c) \cdot \vec{v}_y$$
$$A_z = A_r \cdot \cos(\theta_c) \cdot \vec{u}_z + A_r \cdot \sin(\theta_c) \cdot \vec{v}_z$$

where A_x, A_y, and A_z are the x, y, and z coordinates for the atom respectively, A_r is the radius of the atom, 0.53Å for C, 1.75Å for O, 0.39Å for N, θ_c is a randomly angle selected from a uniform distribution between 0 to 2π, and $\vec{u}$ is the unit vector orthogonal to the vector connecting the beginning and end points and $\vec{v}$ is a vector orthogonal to the vector connecting C-α_1 and C-α_2 . The angle θ_c for N is θ_c - π. The phi, ϕ, and psi, ψ, angles are calculated as:

EQ 30-32:

$$\overrightarrow{s_1} = \widehat{p_1} \times \widehat{p_2}$$
$$\overrightarrow{s_2} = \widehat{p_2} \times \widehat{p_3}$$
$$\vec{t} = \overrightarrow{s_1} \times \widehat{p_2}$$

where $\widehat{p_1}$ is the unit vector pointing from the first and second atom, C_1 to N_1 for ϕ and N_1 to C-α for ψ, $\widehat{p_2}$ is the unit vector pointing from the second and third atom, N_1 to C-α for ϕ and C-α to C_2 for ψ, and $\widehat{p_3}$ is the unit vector pointing from the third and fourth atom, C-α to C_2 for ϕ and C_2 to N_2 for ψ. The dot products $S = \overrightarrow{s_1} \cdot \overrightarrow{s_2}$, and $T = \vec{t} \cdot \overrightarrow{s_2}$ are then used to calculate the dihedral angle according to:

EQ 33:

$$D = cos^{-1}(T, S)$$

where *D* is the dihedral angle ϕ or ψ. If *T* is negative the angles sign is negative. The angle is then converted from radians to degrees since the angles in the Top500 database are in degrees.

Once the chain is complete it is transferred from the z-axis to its final location in space using the vector pointing from the origin to final position along the z-axis, $\vec{j}$, and the vector pointing from the first position to the last position determined earlier in the chain building, $\vec{k}$. This is accompished using a rotation matrix determined by the following:

EQ 34:

$$R = I + v_x + {v_x}^2 \cdot \frac{1 - c}{s^2}$$

where R is the rotation matrix, I is the identity matrix, c is the dot product of $\vec{j}$ and $\vec{k}$, and s is the magnitude of $\vec{v}$ which equals the cross product between $\vec{j}$ and $\vec{k}$, and v_x is the skew-symmetric cross-product matrix of $\vec{v}$ which is:

EQ 35:

$$v_x = \begin{bmatrix} 0 & -v_3 & v_2 \\ v_3 & 0 & -v_1 \\ -v_2 & v_1 & 0 \end{bmatrix}$$

the chain is then translated to its correct location in space adding to each coordinate the first position vector.

The residue atoms are added using coordinates for the backbone and side-chains atoms obtained from SwissProt and with the C-α centered at the origin. In order to obtain the correct side-chain geometry the C-α to N vector, $\vec{a}$, of the generated coordinates are aligned with the C-

α to N vector, $\vec{b}$, of the side-chain. This is accomplished using a rotation matrix determined using EQs 35 and 36 with c as the dot product of $\vec{a}$ and $\vec{b}$, and s is the magnitude of $\vec{v}$ which equals the cross product between $\vec{a}$ and $\vec{b}$. The residue coordinate for the side-chain are transferred to the correct position using the C-α coordinates.

References

1. Graewert, M. A., and D. I. Svergun. 2013. Impact and progress in small and wide angle X-ray scattering (SAXS and WAXS). Current Opinion in Structural Biology 23:748-754.

2. Knight, C. J., and J. S. Hub. 2015. WAXSiS: a web server for the calculation of SAXS/WAXS curves based on explicit-solvent molecular dynamics. Nucleic Acids Research.

3. Boldon, L., F. Laliberte, and L. Liu. 2015. Review of the fundamental theories behind small angle X-ray scattering, molecular dynamics simulations, and relevant integrated application. Nano Reviews 6:10.3402/nano.v3406.25661.

4. Petoukhov, M. V., D. Franke, A. V. Shkumatov, G. Tria, A. G. Kikhney, M. Gajda, C. Gorba, H. D. T. Mertens, P. V. Konarev, and D. I. Svergun. 2012. New developments in the ATSAS program package for small-angle scattering data analysis. Journal of Applied Crystallography 45:342-350.

5. Svergun, D., C. Barberato, and M. H. J. Koch. 1995. CRYSOL - a program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates. J. Appl. Crystallogr. 28:768-773.

6. Schneidman-Duhovny, D., M. Hammel, John A. Tainer, and A. Sali. 2013. Accurate SAXS Profile Computation and its Assessment by Contrast Variation Experiments. Biophysical Journal 105:962-974.

7. Receveur-Bréchot, V., and D. Durand. 2012. How Random are Intrinsically Disordered Proteins? A Small Angle Scattering Perspective. Current Protein & Peptide Science 13:55-75.

8. Kikhney, A. G., and D. I. Svergun. 2015. A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. FEBS Letters 589:2570-2577.

9. Pelikan, M., G. L. Hura, and M. Hammel. 2009. Structure and flexibility within proteins as identified through small angle X-ray scattering. Gen. Physiol. Biophys. 28:174-189.

10. Tria, G., H. D. T. Mertens, M. Kachala, and D. I. Svergun. 2015. Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. IUCrJ 2:207-217.

11. Makowski, L. 2010. Characterization of Proteins with Wide-angle X-ray Solution Scattering (WAXS). Journal of structural and functional genomics 11:9-19.

12. Venditti, V., T. K. Egner, and G. M. Clore. 2016. Hybrid Approaches to Structural Characterization of Conformational Ensembles of Complex Macromolecular Systems

Combining NMR Residual Dipolar Couplings and Solution X-ray Scattering. Chemical Reviews.

13. Bernado, P., E. Mylonas, M. V. Petoukhov, M. Blackledge, and D. I. Svergun. 2007. Structural Characterization of Flexible Proteins Using Small-Angle X-ray Scattering. J. Am. Chem. Soc. 129:5656-5664.

14. Curtis, J. E., S. Raghunandan, H. Nanda, and S. Krueger. 2012. SASSIE: A program to study intrinsically disordered biological molecules and macromolecular ensembles using experimental scattering restraints. Comput. Phys. Commun. 183:382-389.

15. Yang, S., L. Blachowicz, L. Makowski, and B. Roux. 2010. Multidomain assembled states of Hck tyrosine kinase in solution. Proceedings of the National Academy of Sciences of the United States of America 107:15757-15762.

16. Ozenne, V., F. Bauer, L. Salmon, J.-r. Huang, M. R. Jensen, S. Segard, P. Bernadó, C. Charavay, and M. Blackledge. 2012. Flexible-meccano: a tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables. Bioinformatics 28:1463-1470.

17. Antonov, L. D., S. Olsson, W. Boomsma, and T. Hamelryck. 2016. Bayesian inference of protein ensembles from SAXS data. Physical Chemistry Chemical Physics 18:5832-5838.

18. Różycki, B., Y. C. Kim, and G. Hummer. 2011. SAXS ensemble refinement of ESCRT-III CHMP3 conformational transitions. Structure (London, England : 1993) 19:109-116.

19. Kashtanov, S., and F. M. Ytreberg. Characterizing Structural Ensembles of Intrinsically Disordered Proteins Determined via a Broad Ensemble Generator with Reweighting. Biophysical Journal 102:631a.

20. Daughdrill, G. W., S. Kashtanov, A. Stancik, S. E. Hill, G. Helms, M. Muschol, V. Receveur-Bréchot, and F. M. Ytreberg. 2012. Understanding the Structural Ensembles of a Highly Extended Disordered Protein(). Molecular bioSystems 8:308-319.

21. Grishaev, A., L. Guo, T. Irving, and A. Bax. 2010. Improved Fitting of Solution X-ray Scattering Data to Macromolecular Structures and Structural Ensembles by Explicit Water Modeling. J. Am. Chem. Soc. 132:15484-15486.

22. Schneidman-Duhovny, D., M. Hammel, and A. Sali. 2010. FoXS: a web server for rapid computation and fitting of SAXS profiles. Nucleic Acids Research 38:W540-W544.

23. Poitevin, F., H. Orland, S. Doniach, P. Koehl, and M. Delarue. 2011. AquaSAXS: a web server for computation and fitting of SAXS profiles with non-uniformally hydrated atomic models. Nucleic Acids Research 39:W184-W189.

24. Merzel, F., and J. C. Smith. 2002. SASSIM: a method for calculating small-angle X-ray and neutron scattering and the associated molecular envelope from explicit-atom models of solvated proteins. Acta Crystallographica Section D 58:242-249.

25. Liu, H., A. Hexemer, and P. H. Zwart. 2012. The Small Angle Scattering ToolBox (SASTBX): an open-source software for biomolecular small-angle scattering. Journal of Applied Crystallography 45:587-593.

26. Gao, F., A. D. Kight, R. Henderson, S. Jayanthi, P. Patel, M. Murchison, P. Sharma, R. L. Goforth, T. K. S. Kumar, R. L. Henry, and C. D. Heyes. 2015. Regulation of Structural Dynamics within a Signal Recognition Particle Promotes Binding of Protein Targeting Substrates. Journal of Biological Chemistry.

27. Goforth, R. L., E. C. Peterson, J. Yuan, M. J. Moore, A. D. Kight, M. B. Lohse, J. Sakon, and R. L. Henry. 2004. Regulation of the GTPase Cycle in Post-translational Signal Recognition Particle-based Protein Targeting Involves cpSRP43. Journal of Biological Chemistry 279:43077-43084.

28. Kathir, K. M., D. Rajalingam, V. Sivaraja, A. Kight, R. L. Goforth, C. Yu, R. Henry, and T. K. S. Kumar. 2008. Assembly of Chloroplast Signal Recognition Particle Involves Structural Rearrangement in cpSRP43. Journal of Molecular Biology 381:49-60.

29. Acerbo, A. S., M. J. Cook, and R. E. Gillilan. 2015. Upgrade of MacCHESS facility for X-ray scattering of biological macromolecules in solution. Journal of Synchrotron Radiation 22:180-186.

30. Nielsen, S. S., K. N. Toft, D. Snakenborg, M. G. Jeppesen, J. K. Jacobsen, B. Vestergaard, J. P. Kutter, and L. Arleth. 2009. BioXTAS RAW, a software program for high-throughput automated small-angle X-ray scattering data reduction and preliminary analysis. Journal of Applied Crystallography 42:959-964.

31. Konarev, P. V., V. V. Volkov, A. V. Sokolova, M. H. J. Koch, and D. I. Svergun. 2003. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. Journal of Applied Crystallography 36:1277-1282.

32. Roy, A., A. Kucukural, and Y. Zhang. 2010. I-TASSER: a unified platform for automated protein structure and function prediction. Nat. Protocols 5:725-738.

33. Keenan, R. J., D. M. Freymann, P. Walter, and R. M. Stroud. 1998. Crystal structure of the signal sequence binding subunit of the signal recognition particle. Cell 94:181-191.

34. Baiocco, P., L. J. Gourlay, F. Angelucci, J. Fontaine, M. Hervé, A. E. Miele, F. Trottein, M. Brunori, and A. Bellelli. 2006. Probing the Mechanism of GSH Activation in Schistosoma haematobium Glutathione-S-transferase by Site-directed Mutagenesis and X-ray Crystallography. Journal of Molecular Biology 360:678-689.

35. Sivaraja, V., T. K. S. Kumar, P. S. T. Leena, A.-n. Chang, C. Vidya, R. L. Goforth, D. Rajalingam, K. Arvind, J.-L. Ye, J. Chou, R. Henry, and C. Yu. 2005. Three-Dimensional Solution Structures of the Chromodomains of cpSRP43. J. Biol. Chem. 280:41465-41471.

36. Lovell, S. C., I. W. Davis, W. B. Arendall, P. I. W. de Bakker, J. M. Word, M. G. Prisant, J. S. Richardson, and D. C. Richardson. 2003. Structure validation by Cα geometry: φ,ψ and Cβ deviation. Proteins: Structure, Function, and Bioinformatics 50:437-450.

37. Gfeller, D., O. Michielin, and V. Zoete. 2013. SwissSidechain: a molecular and structural database of non-natural sidechains. Nucleic Acids Research 41:D327-D332.

38. Kleywegt, G. J. 1997. Validation of protein models from Cα coordinates alone1. Journal of Molecular Biology 273:371-376.

Figures

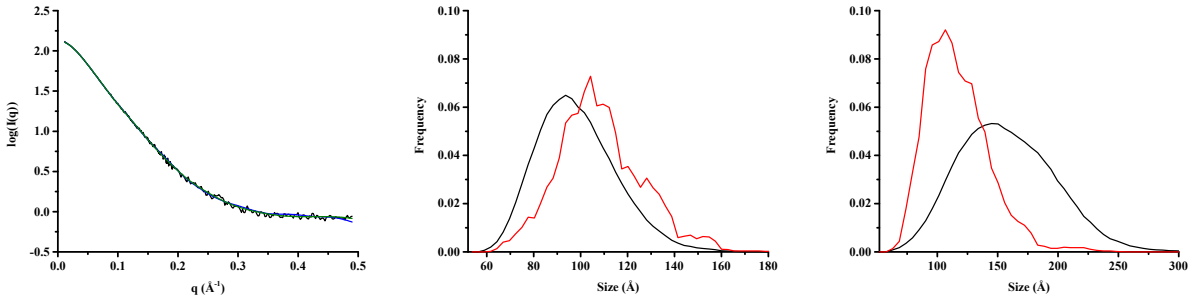

Figure 1: Fitting results for the cpSRP54 M-domain. *Left:* Fit for RANCH (blue) and CEG (green) generated ensembles to the experimental data (black). The results indicated both sets of ensembles fit the data quite well with chi squared values of 1.091 and 1.046 for RANCH and CEG respectively although the RANCH produced ensembles begin to deviate above q = 0.45 $Å^{-1}$. *Middle:* Size distribution for the fit using RANCH generated ensembles. A peak maximum at ~110 Å is observed. *Right:* Size distribution for the fit using CEG generated ensembles. A peak at ~110 Å is observed consistent with that observed for the RANCH generated ensembles. The frequency with which this size occurs is slightly higher for the CEG produced ensembles. However, both produce a relatively similar profile.

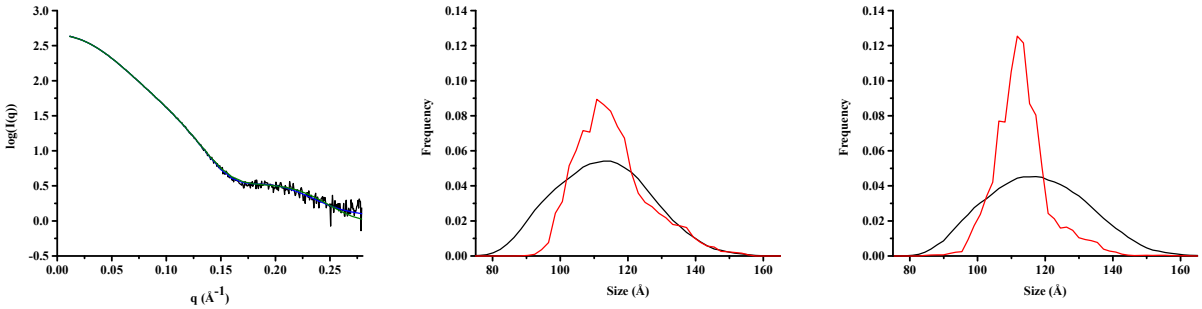

Figure 2: Fitting results for GST-CD2. *Left:* Fit for RANCH (blue) and CEG (green) generated ensembles to the experimental data (black). The results indicated both sets of ensembles fit the data quite well with chi squared values of 1.105 and 1.164 for RANCH and CEG respectively. *Middle:* Size distribution for the fit using RANCH generated ensembles. A peak maximum at ~112 Å is observed. *Right:* Size distribution for the fit using CEG generated ensembles. A peak at ~112 Å is observed consistent with that observed for the RANCH generated ensembles. The frequency with which this size occurs is slightly higher for the CEG produced ensembles. However, both produce a relatively similar profile.

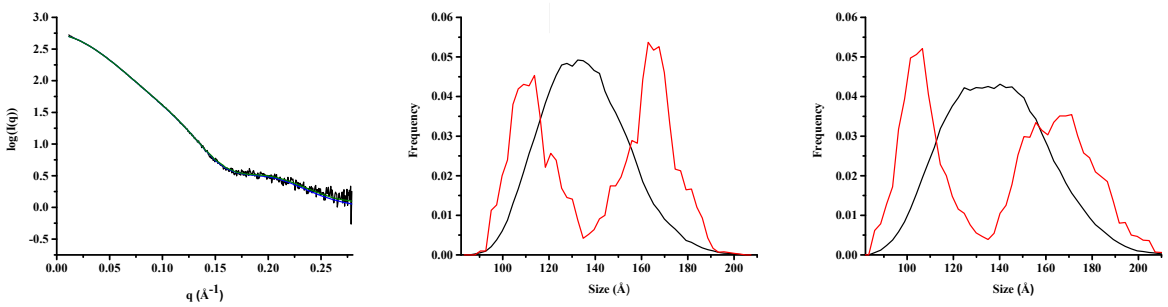

Figure 3: Fitting results for GST-CD2CD3 *Left:* Fitting results for GST-CD2 Left: Fit for RANCH (blue) and CEG (green) generated ensembles to the experimental data (black). The results indicated both sets of ensembles fit the data quite well with chi squared values of 1.560 and 1.612 for RANCH and CEG respectively. *Middle:* Size distribution for the fit using RANCH generated ensembles. A bimodal distribution is observed with peak maximum at ~110 Å and ~170 Å. This suggest two conformations, one relatively closely coupled with another very extended. *Right:* Size distribution for the fit using CEG generated ensembles. A bimodal distribution is observed as in the RANCH fitting with peak maximum at ~110 Å and ~170 Å. ensembles. The frequency with which the sizes occur differs for the CEG produced ensembles with the higher peak presenting a flatter distribution. Nevertheless, both produce a relatively similar profile.

IV. Conclusion

Several decades of study have revealed the chloroplast signal recognition particle pathway to be quite unique. Not only does it contain a novel chaperone with ATP independent disaggregase activity for its substrate, but it has also adapted a cytosolic-like 54 kDa signal recognition particle protein to a seemingly orthogonal system.[1] The results of this study reveal the cpSRP heterodimer and its components to be quite unusual (Figure 1). While early studies focused on cpSRP43 as an RNA replacement and scaffold for LHCP with cpSRP54 acting in a relatively passive role, simply aiding the delivery of LHCP via interaction with cpFtsY, this investigation has revealed the proteins as fully adapting themselves to one another.[2] Specifically, cpSRP43's ability to sample configurational space is tuned such that interaction with cpSRP54 increases its ability to function as a chaperone. In a follow up study to our investigation, another group found that cpSRP43's affinity for substrate is again modified by an interaction with the Alb3 translocase.[3] In contrast to the view of cpSRP54 as an interaction partner with a small, *albeit* important role in LHCP targeting, these results have demonstrated not only that cpSRP54 is directly involved in LHCP binding, but that cpSRP54 itself has been adapted to cpSRP43. This is accomplished *via* intrinsic disorder in an extension at its C-terminus. This allows cpSRP54 to sample configurational space relative to cpSRP43, apparently scanning for the third transmembrane segment of LHCP. This is surprisingly similar to the scanning observed to occur at the ribosome in the co-translational cytosolic SRP.[4] This is remarkable due to the major differences between the cytosolic system and the cpSRP system. The ribosome and SRP RNA provide a large scaffold responsible for dictating the positioning the SRP54/Ffh domains, especially in the scanning mode searching for signal sequences. The cpSRP54 results show that

the protein, in conjunction with changes in cpSRP43, manages to undergo similar fluctuations despite the absence of those scaffolds.

The predictive ability of the protein model developed in this work continues beyond identifying LHCP loading preparatory steps (Figure 1). While the cpSRP system is quite dissimilar in its mechanism, the cytosolic SRP system has nevertheless proved essential in guiding our understanding of what general structural states should be anticipated. A combined crystal structure and electron microscope image of the signal sequence bound ribosome/SRP provides an attractive option for predicting the structural outcome of LHCP loading for cpSRP.[5] In the SRP/ribosome structure the N-domain of SRP54 forms a close contact with ribosomal proteins L23 and L35 (Figure 1). What is interesting about this structure is that it defines an orientation for the NG-domain receptor binding site of SRP54. Supposing a specific orientation for cpSRP54 relative to cpSRP43 and LHCP is defined upon LHCP binding, what might be the coarse shape of this complex? Further, what specific sequence and structural features could be critical in this positioning? Published data for deletion constructs of cpSRP43 provide the strongest evidence for a specific interaction that may be of structural and functional relevance. It was found that deletion of the first chromodomain, which includes what is now known as the N-terminal tail of cpSRP43, supports LHCP loading but is deficient in integration. Considering the structural and sequence similarity between L23/L35 and CD1, it seems reasonable to predict a similar interaction between cpSRP54's N-domain and CD1. Indeed, as in the SRP54 structure, the M-domain and CD1 together could stabilize the NG-domain in a particular orientation. What is interesting about this model is that, since the cpFtsY binding site is conserved, it would define an orientation at the membrane that would place the TM3 C-terminal tail pointing toward the membrane (Figure 2). This could explain why, while LHCP loading is unaffected by CD1

deletion, integration is hindered. Indeed, nothing in the model proposed here would suggest LHCP binding would be hindered.

It remains uncertain from this information alone, however, why a specific orientation would be so critical. A recent YidC, an Alb3 homologe, structure may help to explain this.[6] YidC's transmembrane domains contain a surprisingly large number of polar residues. Among these is a critical arginine residue. The proposed mechanism for insertion of some transmembrane helices involves the arginine attracting the negatively charged tail of the TM segment and pulling it across the membrane using those polar residues and perhaps the proton gradient across the membrane (Figure 3). While *Arabidopsis* LHCP's C-terminus contains two charged residues, only one is conserved among confirmed LHCP sequences (Figure 4). However, there are several glutamine residues. Of the TM helices found to integrate *via* the mechanism put forth using the YidC structure, a significant stretch of negatively charged residues was present. As an Alb3 homology model shows that the polar residues in the TM segments are retained, with a lysine replacing the arginine, it seems reasonable that a similar mechanism of insertion might take place for TM3. Since there is less electrostatic attraction from a negatively charged LHCP C-terminus to the Alb3 buried lysine, cpSRP54 may ensure the tail is oriented toward these polar residues and is therefore more likely to form and maintain an initial interaction. The subsequent step would be the partitioning of LHCP TM3 from the M-domain into the membrane. The following steps would likely involve collection of chlorophylls and carotenoids along with integration of TMs 1 and 2.

Once TM3 has been released from the M-domain, since the M-domain is responsible for stabilizing the N-domain contact, the cpSRP54 NG-domain interaction with cpSRP43 would be disrupted. Interestingly, the M-domain has been observed to stimulate GTP hydrolysis in the

cpSRP54-cpFtsY complex. Since CD1 is known to inhibit GTP hydrolysis by cpSRP54-cpFtsY, this release could act as a signal to indicate that the cargo has been delivered and consequently allow the M-domain to activate GTP hydrolysis for recycling of the complex. If this is correct, the transit complex model with the N-domain at CD1 must orient the M-domain and NG-domains such that M-domain stimulated GTP hydrolysis is blocked.

Ultimately these predictions must be tested. The first, and most critical, is to determine whether CD1 and the cpSRP54 N-domain actually form an interaction and whether this interaction is responsible for the observed loss in integration in the absence of CD1. Since the ΔCD1 mutant also removed the N-terminal tail it is uncertain whether CD1 itself of CD1 in combination with this tail caused the reduction in integration. Since the N-terminal tail contains a significant string of negatively charged residues that would complement positive charge on the N-domain, we have chosen first to delete the tail from cpSRP43. Further, a twinned positive and negative charge on CD1 and Ank1 that has a reverse complement on the N-domain is being tested for another possible linkage between the domains. Specifically, the loops between the first and second and third and forth helices of the N-domain contain positive and negative charge (Figure 6). As the overall charge at CD1-Ank1 is negative, we have chosen to convert the N-domain loops to be completely negatively charged as this would create significant repulsion. The results, if this hypothetical interaction and its functional significance is indeed valid, should be identical to those of the CD1 deletion. That is, LHCP loading would be unaffected while integration would be significantly hindered. If the experiments yield promising results, the negative charges on CD1-Ank1 will be transformed to positive charges. If these results are similar to the cpSRP54 N-domain mutations, they will together provide strong support for the proposed model. Indeed, if these results are positive, another set of mutations with a charge

reversal on CD1-Ank1 and The N-domain together could further suggest a specific interaction and a specific orientation in that interaction. The significance of the LHCP C-terminal tail is being investigated as well. It will be deleted to determine whether integration is possible in its absence. If not, it would be supportive of the above mechanism for integration and indicate further study is warranted. It is interesting to hypothesize that addition of a stretch of negative charges to the tail might recover the ability of cpSRP43 to support integration in the absence of cpSRP54.

The result of this body of work provides a significant increase in our understanding of the molecular details of LHCP targeting. Furthermore, the final model offers ample opportunity for further investigation and may lead to work in understanding not only the targeting of LHCP, but also to understanding the mechanism by which integration takes place. Together with folding investigations of LHCP this work provides a markedly clearer picture of the workings of this system.

References

1 Jaru-Ampornpan, P. *et al*. ATP-independent reversal of a membrane protein aggregate by a chloroplast SRP subunit. *Nat. Struct. Mol. Biol.* **17**, 696-702, doi:10.1038/nsmb.1836 (2010).

2 Stengel, K. F. *et al*. Structural Basis for Specific Substrate Recognition by the Chloroplast Signal Recognition Particle Protein cpSRP43. *Science* **321**, 253-256, doi:10.1126/science.1158640 (2008).

3 Liang, F.-C. *et al*. Conformational dynamics of a membrane protein chaperone enables spatially regulated substrate capture and release. *Proceedings of the National Academy of Sciences* (2016).

4 Holtkamp, W. *et al*. Dynamic switch of the signal recognition particle from scanning to targeting. **19**, 1332-1337 (2012).

5 Halic, M. *et al*. Following the signal sequence from ribosomal tunnel exit to signal recognition particle. *Nature* **444**, 507-511 (2006).

6 Kumazaki, K. *et al*. Structural basis of Sec-independent membrane protein insertion by YidC. *Nature* **509**, 516-520, doi:10.1038/nature13167, http://www.nature.com/nature/journal/v509/n7501/abs/nature13167.html - supplementary-information (2014).

Figures

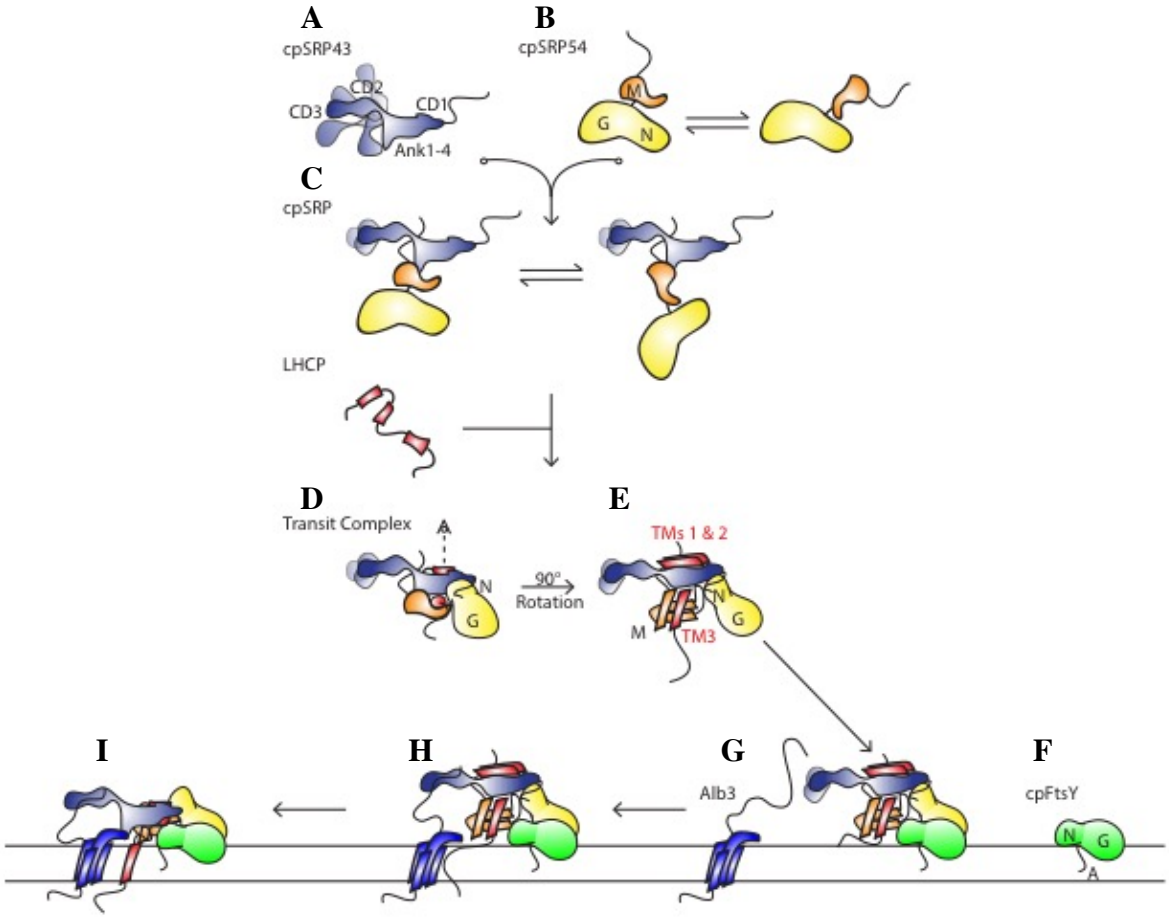
A
cpSRP43
CD2
CD3
CD1
Ank1-4
B
cpSRP54
M
G
N
C
cpSRP
LHCP
D
Transit Complex
A
N
G
90°
Rotation
E
TMs 1 & 2
N
G
M
TM3
I
H
G
Alb3
F
cpFtsY
N
G
A

Figure 6: The predicted structural states of the chloroplast signal recognition particle (cpSRP) during the targeting of light harvesting chlorophyll binding protein (LHCP) to the thylakoid membrane. *(A)* Free cpSRP43. Based upon the connectivity of the domains in cpSRP43 a large degree of flexibility in the protein is expected. *(B)* Free cpSRP54. The conformational states in the cytosolic SRP system are quite variable and dependent upon SRP RNA. Due to the absence of SRP RNA in the cpSRP system, cpSRP54 might exist in multiple states in its free form. *(C)* The cpSRP heterodimer. Although crystal structures have indicated CD2 of cpSRP43 loses its flexibility in the presence of the cpSRP54-cpSRP43 binding peptide, no apparent interaction site that would stabilize the remainder of cpSRP43 or cpSRP54 has been identified. For this reason it is likely that flexibility would be retained. *(D)* Transit Complex. Binding of LHCP to cpSRP forms the transit complex. The M-domain likely interacts with TM3 of LHCP and, based upon crystal structures of SRP54 at the ribosome, *(E)* lock the NG-domain of cpSRP54 in a conformation ready for receptor binding. *(F)* cpSRP54 receptor, cpFtsY. The A-domain of cpFtsY helps to anchor it in the thylakoid membrane with the cpSRP54 binding GTPase region facing away from the membrane. *(G)* The locked configuration for cpSRP54's NG domain would provide a specific orientation for all domains and proteins in the transit complex relative to the thylakoid membrane. The orientation would place the TM3 C-terminal tail toward the membrane. *(H)* If Alb3 LHCP integration functions in a manner similar to YidC's insertion mechanism for single pass membrane proteins, the C-terminus of LHCP is pull across the membrane by a special hydrophilic groove. *(I)* After the C-terminal tail is been pulled across the membrane, TM3 could then partition from the M-domain hydrophobic groove into the membrane.