

12-2019

## Toward Understanding the Mechanism of Protein Targeting in the Chloroplast Signal Recognition Particle Pathway

Mercede Furr  
*University of Arkansas, Fayetteville*

Follow this and additional works at: <https://scholarworks.uark.edu/etd>



Part of the [Amino Acids, Peptides, and Proteins Commons](#), [Biochemistry Commons](#), [Biology Commons](#), [Cell Biology Commons](#), and the [Molecular Biology Commons](#)

---

### Citation

Furr, M. (2019). Toward Understanding the Mechanism of Protein Targeting in the Chloroplast Signal Recognition Particle Pathway. *Theses and Dissertations* Retrieved from <https://scholarworks.uark.edu/etd/3468>

This Dissertation is brought to you for free and open access by ScholarWorks@UARK. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of ScholarWorks@UARK. For more information, please contact [ccmiddle@uark.edu](mailto:ccmiddle@uark.edu).

Toward Understanding the Mechanism of Protein Targeting in the Chloroplast Signal  
Recognition Particle Pathway

A dissertation submitted in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy in Cell and Molecular Biology

by

Mercede Furr  
Bachelor of Science in Biology, 2012

December 2019  
University of Arkansas

This dissertation is approved for recommendation to the Graduate Council.

---

Suresh Thallapuranam, Ph.D.  
Dissertation Director

---

David McNabb, Ph.D.  
Committee Member

---

Ralph Henry, Ph.D.  
Committee Member

---

Colin Heyes, Ph.D.  
Committee Member

## **Abstract**

Protein targeting is a vital cellular function. The signal recognition particle (SRP) pathway is a universally conserved targeting system present in the cytosol and used to co-translationally target many proteins to the inner membrane of prokaryotes and the endoplasmic reticulum of eukaryotes. The chloroplast has a homologous SRP system which post-translationally targets light harvesting chlorophyll binding proteins (LHCPs) to the thylakoid membrane for integration. The chloroplast SRP (cpSRP) is a heterodimer with a 54 kDa subunit equivalent to SRP54 in the canonical pathway. In addition, cpSRP contains a novel 43 kDa subunit which is a unique and irreplaceable component. cpSRP43 is central to targeting the highly hydrophobic LHCPs to the Albino3 translocase at the thylakoid membrane by operating as a chaperone capable even of disaggregating LHCPs without any external energy input. cpSRP43 has multiple binding partners in the cpSRP pathway. Many details about these binding interactions have been discovered however exact residual details regarding these interactions still requires elucidation. A structure for cpSRP43 bound to cpSRP54 was determined by X-Ray crystallography although cpSRP43 also functions in free form in the chloroplast. The results of this study demonstrate the significant amount of structural flexibility and thermal stability cpSRP43 as well as its potential use as a generic chaperone for proteins outside of the chloroplast. The structure of the c-terminal end of Albino3 (cAlb) was also investigated here since the exact binding interface between cpSRP43 and cAlb is still under debate. The results here reveal a region in cAlb which has a high propensity toward structure. This region may prove to be important in binding to cpSRP43 and could lead toward a better understanding of the process of integration for LHCP.

©2019 by Mercede Furr  
All Rights Reserved

## **Acknowledgements**

I would like to express my appreciation for Dr. Kumar as well as my committee members for their guidance and patience with me as I grew as a researcher through the years. I would like to thank my friends and loved ones for believing in me and caring about me. I am very thankful for the wonderful people I have meet in Dr. Kumar's group and throughout the Chemistry department. The help and friendship provided to me from my fellow lab mates has been is invaluable. I would like to thank my Mom and Dad and my sister, Amanda, for their unconditional love and support and my niece, Jenna, for making me smile when I needed it. I want to give a very special thanks to my grandmother, Onie, for loving me and supporting me. I would like to thank God for strength and refuge and for each day.

## Table of Contents

<b>I.</b>	<b>Chapter-1: Introduction.....</b>	<b>1</b>
	Signal Recognition Particles.....	1
	Light Harvesting Chlorophyll Binding Proteins.....	4
	The Chloroplast Signal Recognition Particle Pathway.....	6
	cpSRP43 Interactions and Chaperone Activity.....	8
	Alb3/Oxa1/YidC Translocase Family.....	15
	Intrinsically Disordered Proteins.....	18
	The c-terminal domain of Albino3.....	22
	Scope of Study.....	25
	References.....	28
<b>II.</b>	<b>Chapter-2: Structure and Stability of cpSRP43.....</b>	<b>35</b>
	Abstract.....	35
	Introduction.....	37
	Materials and Methods.....	41
	Results and Discussion.....	46
	Summary and Conclusions.....	66
	References.....	68
<b>III.</b>	<b>Chapter 3: Generic Chaperone Activity of cpSRP43.....</b>	<b>72</b>
	Abstract.....	72
	Introduction.....	73
	Materials and Methods.....	76
	Results and Discussion.....	80
	Summary and Conclusions.....	91
	References.....	98
<b>IV.</b>	<b>Chapter 4: Structural propensity in the C-terminal Domain of</b>	
	<b>Albino3.....</b>	<b>103</b>
	Abstract.....	103
	Introduction.....	104
	Materials and Methods.....	107
	Results and Discussion.....	119
	Summary and Conclusions.....	138
	Supplemental Material.....	142
	References.....	149
<b>V.</b>	<b>Conclusion.....</b>	<b>153</b>
	References.....	158
<b>VI.</b>	<b>Appendix A: Simplified Purification of GST-cpSRP43.....</b>	<b>160</b>
	Abstract.....	160
	Introduction.....	161
	Materials and Methods.....	162
	Results and Discussion.....	164
	Summary and Conclusions.....	172
	References.....	175
	Biosafety Committee Approval.....	176

## List of Abbreviations

**SRP** Signal Recognition Particle

**cpSRP** Chloroplast Signal Recognition Receptor

**ER** Endoplasmic Reticulum

**GTP/ATP** Guanosine Triphosphate/Adenosine Triphosphate

**SR** Eukaryotic SRP membrane receptor

**FtsY** Bacterial SRP membrane receptor

**cpFtsY** Chloroplast SRP membrane receptor

**RNA** Ribonucleic acid

**Ffh** Bacterial homolog of the 54kDa subunit of SRP

**N-G domain** N-terminal and GTPase activity domains on the 54kDa subunit of cpSRP

**M-domain** Methionine-rich domain of the 54kDa subunit of cpSRP

**LHCP** Light Harvesting Chlorophyll Binding Protein

**LHC** Light harvesting complex

**TM** Trans-membrane domain

**TIC/TOC** translocon at the inner/outer envelope membrane of chloroplasts

**Oxa1** Mitochondrial inner membrane insertase

**YidC** Membrane protein insertase in bacteria and archaea

**Alb3** Membrane protein insertase in chloroplast (Albino3)

**cAlb** the c-terminal of Albino3

**CD** Chromodomain

**ANK** Ankyrin repeat motif in proteins

**DPLG** a motif made up of the amino acids: Aspartic acid-Proline-Leucine-Glycine

**Kd** Dissociation constant

**ITC** Isothermal titration calorimetry

**Hsp104/ClpB** Heat shock protein 104/ Caseinolytic peptidase B protein (chaperone proteins)

**TF** Trigger factor (ribosome-associated chaperone in bacteria)

**NMR** Nuclear magnetic resonance

**smFRET** Single molecule fluorescence resonance energy transfer

**DnaK** bacterial molecular chaperone protein

**IMP** Inter membrane protein

**Sec pathway** general secretion route in bacteria

**SecYEG** bacterial protein translocation complex

**TOM/TIM** Mitochondrial protein import machinery at the outer and inner mitochondrial membranes

**CoxII** Cytochrome C oxidase subunit II (mitochondrial)

**IDPs/IDRs** Intrinsically disordered proteins/regions

**SLiMs** short linear motifs

**GST** Glutathione S-transferase

**SEC** Size exclusion chromatography

**CA** Carbonic anhydrase

**Con A** Concanavalin A

**sHSPs** Small heat shock proteins

**SBD** Substrate binding domain

**NBD** Nucleotide binding domain



## **I. Introduction**

### **Signal Recognition Particles**

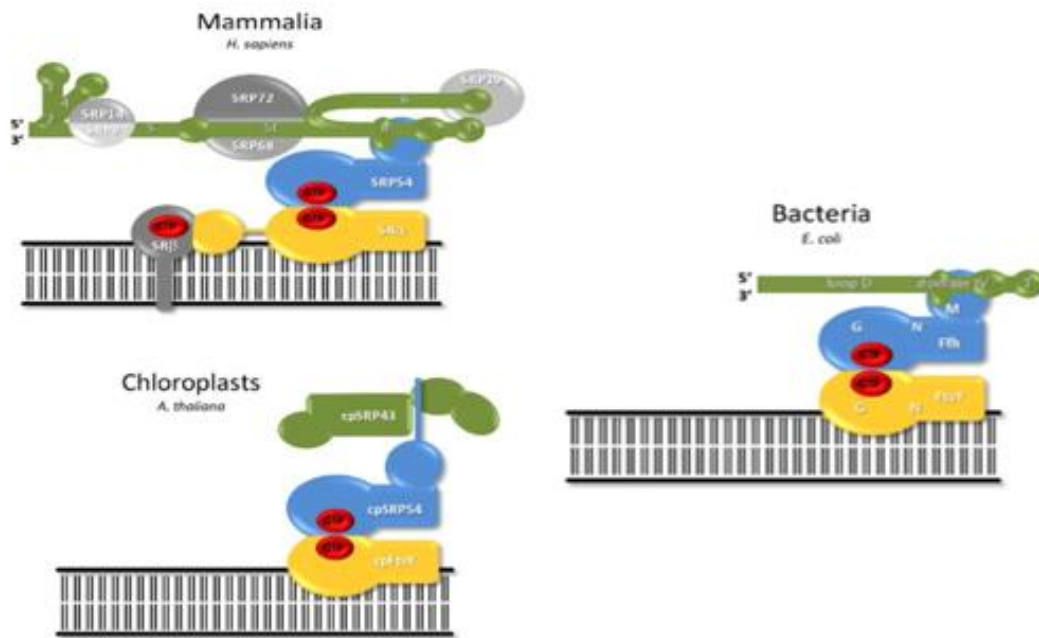
Signal recognition particles (SRP) are ubiquitous and conserved systems which enables the cell to target membrane and cytosolic proteins to their proper locations (1). SRPs are existent in all kingdoms of life. In eukaryotes and prokaryotes, the SRP co-translationally targets secreted and integral membrane proteins to the endoplasmic reticulum (ER) and the plasma membrane, respectively. However, it is fair to mention that certain specific substrates make use of other SRP-independent mechanisms for their translocation (2). To achieve this exceptionally important function, first the SRP recognizes and binds to the N-terminal hydrophobic signal sequence of the nascent chain emerging from the ribosome (3). This association time decreases the amount of protein synthesis occurring which allows for the time needed for the complex to reach the membrane (4). Elongation of the nascent chain is paused as the SRP-nascent chain complex targets the membrane receptor (SR) in a GTP-dependent interaction at the SR. The SRP then releases the nascent chain at the SR, the N-terminal signal sequence is cleaved, and the ribosome resumes translation with the protein directed in the translocon channel. Other sequences present in the proteins give direction for them to move to the exterior of the cell for transport out of the cell, the lysosome for digestion, the Golgi apparatus for further modification or to remain in current placement (5). GTP hydrolysis between the SRP and SR allows for the release of the SRP so that SRPs are available to continue targeting (6). SRP, in this convention, is a critical link between protein synthesis and translocation.

Mammalian SRP is made up of six polypeptides, which are referred to based on their molecular weights (SRP9, SRP14, SRP19, SRP54, SRP68 and SRP72), and one RNA

component termed 7SL (Fig.1). SRP9 and SRP14 function as a heterodimer required for elongation arrest before receptor docking by binding to the SRP RNA (4, 7). The 7SL RNA component plays a role in elongation arrest, operates as a scaffold for fixing the location of the methionine-rich domain of the SRP54 and expedites binding of SRP54 to the receptor (8). The Alu and S domains of SRP RNA are highly conserved which demonstrates the archaic nature of this translational system (9, 10). SRP19 assists in the binding of SRP54 to the SRP RNA acting as a scaffold and a chaperone (4, 11). SRP68 and SRP72 also form a heterodimer which facilitates docking at the SR by interaction with SR $\alpha$  subunit (7, 12). The eukaryotic SR is a membrane-anchored heterodimer made up of two GTPase subunits (SR $\alpha$  and SR $\beta$ ) (13). The 54kDa component (SRP54) is a GTPase responsible for signal sequence recognition and binding the RNA component although, its primary role is targeting the ribosome-nascent chain complexes (RNCs) to the ER membrane (14). SRP54 consists of a GTPase (G-domain), a C-terminal methionine-rich (M-domain) and an N-terminal domain which contains a four helix bundle (8, 11, 15). The N and G domains come together to form the catalytic core the M-domain is responsible for binding the signal sequences at the ribosome via a hydrophobic pocket (16, 17). Once the RNCs are delivered to the translocation pore at the membrane (Sec61p in eukaryotes), GTP hydrolysis leads to the unloading of the RNCs followed by subsequent instability of the SRP-SR association (8). The SRP machinery is recycled and protein synthesis can resume. Together, SRP19, SRP54 and 7SL RNA make up the conserved ribonucleic core which promotes the translocation of nascent chains (11, 18).

The bacterial SRP contains the homolog of SRP54, called Ffh, as well as an RNA component (termed 4.5S in *Escherichia Coli*) (Fig.1). SRP54/Ffh is composed of the methionine-rich, M domain, which binds the RNA and contains the hydrophobic pocket that

recognizes and binds the signal sequence at the N-terminal of proteins coming from the ribosome. The GTPase, or N-G domain, interacts with the receptor (8). FtsY is the membrane receptor of the bacterial SRP pathway homologous to the SR $\alpha$  subunit in the mammalian SRP pathway. The RNA component is conserved in cytosolic SRP systems which provides support for the proper positioning of the M-domain and expedites the interaction between the N-G domain and the receptor (8). Proper assembly of SRPs are in some measure dependent on the correct folding of the RNA component (19, 20). After GTP hydrolysis, the RNCs are unloaded at the membrane for further synthesis through the Sec-YEG translocation pore in prokaryotes (8). Although the bacterial SRP is simplified in comparison to the mammalian SRP, they both exhibit similar properties and characteristics. In fact, the Ffh and FtsY can substitute for SRP54 and SR $\alpha$  of the mammalian pathway to achieve protein targeting *in vitro* (21, 22). Post-translational protein targeting in cells is carried out by different and in some cases more intricate processes which are SRP-independent (23). Post-translational targeting requires protein-specific chaperones to protect fully synthesized proteins during translocation to nucleuses, mitochondria and chloroplasts (23-25). The SRP system is conserved in chloroplasts, albeit not identical, for the purpose of post-translational translocation of the Light Harvesting Chlorophyll Binding proteins (LHCPs) to the thylakoid membrane which is outlined in Figure 1 below.

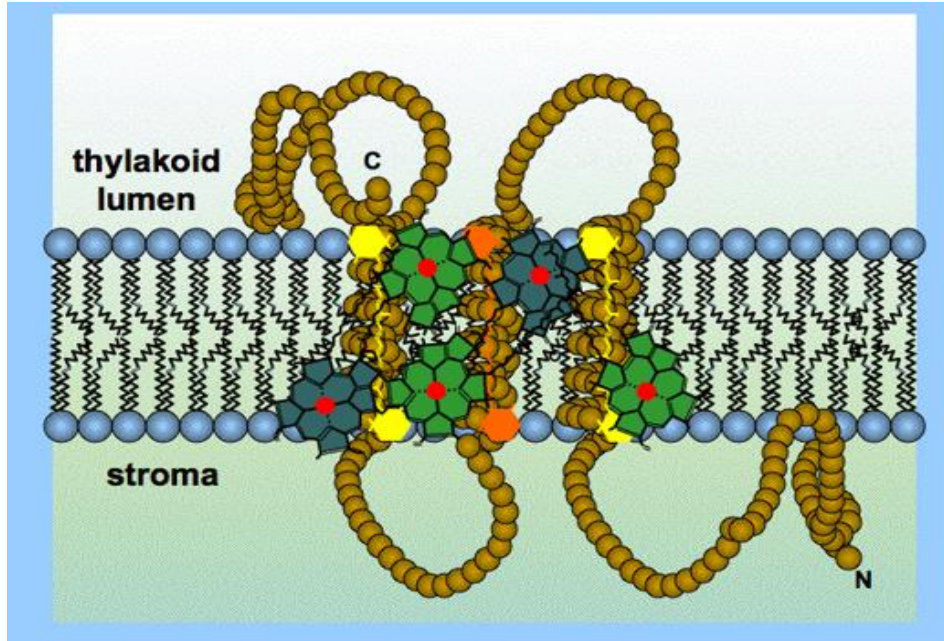


**Fig.1: Overview of the SRP pathways in eukaryotes, prokaryotes and chloroplasts. The cpSRP54 subunit is conserved in all SRP pathways. Mammals and bacteria contain the RNA component however it is absent in chloroplasts. The chloroplast pathway contains the novel cpSRP43 subunit. This figure is adapted from Wild, Klemens et al. J Mol Biol (2016) 428, 2880-2897.**

## Light Harvesting Chlorophyll Binding Proteins

Light harvesting chlorophyll *a/b* binding proteins (LHCPs) are the most plentiful membrane protein in nature (26). LHCPs are nuclear-encoded thylakoid proteins critical for the initial acquisition of light energy for the process of photosynthesis in chloroplasts. The photosystems for LHCPs are positioned in the thylakoid membranes. LHCP is comprised of three membrane spanning helices (TM1-3) and a hydrophilic region, L18, located on the N-terminal side of TM3 (26-28). Chlorophyll *a/b* and carotenoid cofactors, responsible for absorbing light in the visible spectrum, bind to TM1-3 and form light harvesting complexes (LHCs) consequent to the

insertion of LHCPs into the thylakoid membrane (Figure 2).



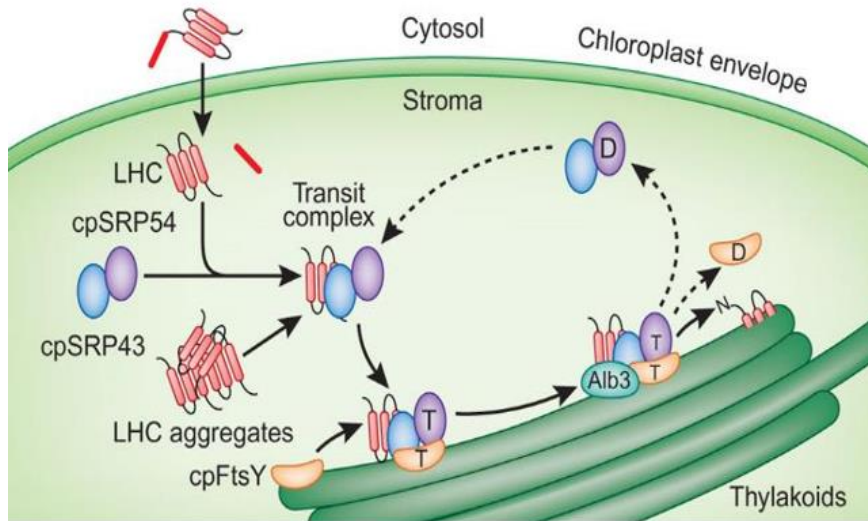
**Fig. 2: LHCPs are the largest membrane protein family, representing over 50% of the proteins in the thylakoid membrane. They bind to carotenoids and chlorophylls and have 3 trans-membrane segments. L18, the cpSRP43 binding unit is located between TM2 and TM3. LHCPs are post-translationally inserted and since they are highly hydrophobic, they require cpSRP43 as their molecular chaperone. Picture from: [http://plantphys.info/plant\\_physiology/lightrxn.shtml](http://plantphys.info/plant_physiology/lightrxn.shtml).**

The chloroplast organelle is of endosymbiotic origin and thereby it contains diverse structures and utilizes various transport pathways (29). Ancient pathways as well as evolutionarily young pathways make up the mixture of its transport machineries. Proper and efficient transport and targeting of LHCPs is a substantial requirement of the plant cell due to their abundance and significance. LHCPs are synthesized in the cytoplasm as precursors which possess N- terminal transit peptides (29). The transit peptides guide the precursors to the chloroplast then engage them with the TOC (translocon at the outer envelope membrane of chloroplasts) and TIC (translocon at the inner envelope membrane of chloroplasts) import system (30). The movement of the precursor LHCPs from their emergence at the ribosome to the TOC/TIC import system involves assistance from guidance complexes (29). After moving

through the import machinery into the stroma, the N-terminal peptide is cleaved and the LHCP is ready to be targeted to and inserted into the thylakoid membrane by the chloroplast signal recognition particle (cpSRP). LHCP is highly hydrophobic and thereby requires the support of a chaperone to prevent its aggregation upon translocation.

### **The Chloroplast Signal Recognition Particle Pathway**

Although there are other pathways to the thylakoid membrane once a nuclear-encoded chloroplast protein enters the stroma, the SRP-dependent pathway is the leading pathway for the integration of integral membrane proteins into the thylakoid membrane (Figure 3). The chloroplast signal recognition particle (cpSRP) is part of the major pathway for the post-translational integration of LHCP (29). LHCP forms a transit complex with cpSRP and is post-translationally targeted to the thylakoid membrane. There are many distinctions between the canonical SRP pathway and the cpSRP pathway. cpSRP is comprised of a 54 kDa homolog (cpSRP54) to the other SRPs and a novel 43kDa protein (cpSRP43). SRP-dependent integration pathways in the chloroplasts are capable of post-translational and co-translational targeting due to the presence of the cpSRP and a co-translationally active cpSRP54, respectively (27, 31). The presence of cpSRP43 allows for post-translational targeting while in the absence of cpSRP43, cpSRP54 targets proteins co-translationally.



**Fig.3: The post-translational targeting pathway of light harvesting chlorophyll binding proteins (LHCP) to the thylakoid membrane. LHCPs are synthesized in the cytosol then transported to the chloroplast stroma by way of a signal sequence. The signal sequence is cleaved after transport across the inner and outer membranes. The chloroplast signal recognition particle (cpSRP) heterodimer which consists of cpSRP43 and cpSRP54 binds LHCP to form the soluble “ transit complex”. Aggregates of LHCPs form in the absence of cpSRP43. The transit complex binds to the cpSRP receptor, cpFtsY, consequent to interaction with the Alb3 translocase. After LHCP is integrated into the thylakoid membrane, the GTPase activity of cpSRP54/cpFtsY allows recycling of the targeting machinery. This figure was adapted from Henry et. al. Nature 2012 Nature Structural & Molecular Biology 17,676– 677(2010).**

Before the identification of the transit complex components which transport LHCP to its location at the thylakoid membrane, the LHCPs were realized to be highly hydrophobic, however, they become soluble when they form the transit complex. This presented the question as to how such a highly hydrophobic substrate is made able to traverse the stroma. LHCP has a molecular weight of approximately 25kDa. LHCP was found to be part of a soluble and stable intermediate species having a molecular weight of 120kDa in the stroma (32). Initially termed as stromal factor these unknown components were discovered to play a role in preserving LHCP in a soluble state as well as maintaining it in a form conducive to successful integration (33). cpSRP54 was found to be present in the transit complex however, alone, it was insufficient for

complex formation. The cpSRP is particular in that it does not require an RNA molecule. In addition to the lack of the RNA molecule, the cpSRP54 of cpSRP was found to be unable to bind 4.5S RNAs from bacterial systems (34). Although homologous to each other, cytosolic SRP54 and cpSRP54 were found to be incapable of substituting for one another suggesting some significant differences between the two proteins (34).

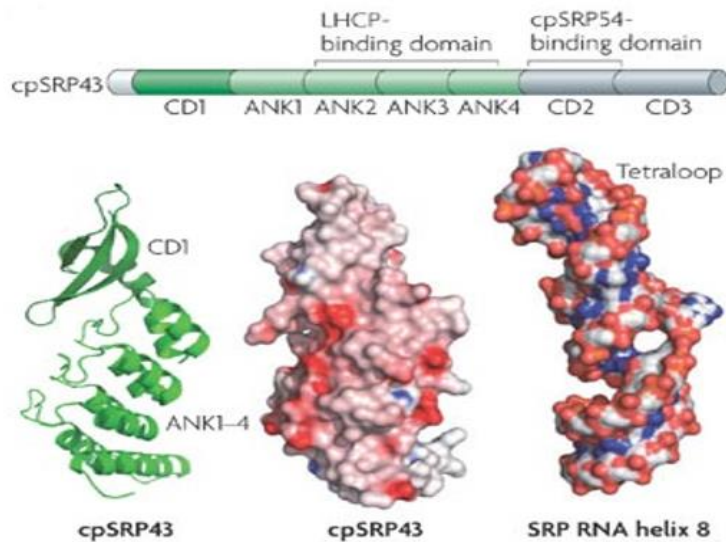
Later, the novel cpSRP43 was discovered, a component unique to the cpSRP, required for transit complex formation (35, 36). TM3 of LHCP maintains a hydrophobic interaction with cpSRP54 as in the other SRP pathways. L18, a hydrophilic loop region of LHCP mentioned above, interacts with cpSRP43. Once the transit complex is formed, it locates LHCP to the membrane receptor, cpFtsY, which is homologous to SR $\alpha$  and the bacterial FtsY (37). At the thylakoid membrane, the transit complex binds to cpFtsY and expedites the targeting of LHCP to Albino3 (Alb3), an integral thylakoid protein which mediates the insertion of LHCP into the thylakoid membrane. An interaction between the stromal c-terminal of Alb3 and cpSRP43 at the thylakoid membrane is necessary for proper integration. The discovery of cpSRP43 and subsequent studies elucidating its involvement in the cpSRP pathway have revealed it to be central and critical to the cpSRP system.

### **cpSRP43 Interactions and Chaperone activity**

cpSRP43 is structurally identified as having three chromodomains (CD1-3) and four ankyrin repeats (Ank1-4) located between CD1 and CD2 (2, 38) and it is not present in any other SRP systems (Fig.4). The crystal structure of cpSRP43 reveals two hydrophobic grooves which are disjointed by a positive ridge on one side and a highly negatively charged surface on the other side (39). Examination into the crystal structure of cpSRP43 proposes that the charge



distribution of cpSRP43 is reminiscent of the canonical SRP RNA component which, as previously mentioned, acts as a scaffold for positioning the M-domain of cpSRP54 in the formation of cpSRP (39).



**Fig. 4: The structure of cpSRP43 is composed of CD1 attached to the ankyrin repeat regions (ANK 1-4). ANK4 is fused by a short chain to CD2 which is linked to CD3 by another short chain. cpSRP43 resembles SRP RNA. This picture was adapted from Sinning et al. Science 2008.**

The chromodomains (CD2 and CD3) of cpSRP43 are closely spaced whereas in other chromoproteins these domains are typically separated by a linker region (40). Chromodomains of other proteins are known for regulating the structure of chromatin due to the ability of these domains to promote protein-protein interactions although, it is curious to find that cpSRP43 is the first non-nuclear chromoprotein (40). CD1 has the typical fold of a chromodomain as is found in Polycomb which is described as a three-stranded antiparallel beta sheet having an alpha helix across the indented portion of the beta sheet (39). CD1 and ANK1 are fused by the helix of CD1 and the N-terminal helix of ANK1 making the distinction between the end of CD1 and beginning of ANK1 difficult to distinguish (39). Ankyrin repeat regions are found in a

multitude of proteins with varied functions in biological processes which demonstrates the versatility of these regions to invoke different types of protein-protein interactions (40). ANK2 and 3 of cpSRP43 display the typical fold of an ankyrin region by having a helix-turn-helix motif. However, the flanking ANK regions, 1 and 4, contain elongated helices (39). The helices in ANK4 are extended by 16 residues, a noteworthy divergence from typical ANK structure, proposing a role in protein-protein interactions (39). The unique architecture and combination of the multi-domains which make up cpSRP43 make it an excellent framework upon which protein interactions can occur.

Experiments involving crosslinking and gel filtration revealed that the third and fourth ANK regions comprise a binding interface for cpSRP43 to exist as a dimer which led to the assertion that cpSRP exists as a heterotrimer with one cpSRP54 subunit and two cpSRP43 subunits (37, 38). However, that assertion was refuted by studies in which the functional cpSRP, for successful insertion of LHCP, was found to be a heterodimer comprised of one subunit of cpSRP43 and cpSRP54 each (27, 41). cpSRP54 by itself does not bind a significant quantity of LHCP, although cpSRP54 together with cpSRP43 in the cpSRP heterodimer creates the local area necessary for cpSRP54 to bind TM3 of LHCP which is essential for transit complex formation (37, 42). Studies using deletion mutants of cpSRP43 determined that the c-terminal chromodomains (CD2 and CD3) of cpSRP43 are the regions which bind to cpSRP54 (40). However, in different studies, a CD3 deletion mutant of cpSRP43 was found to have an insignificant effect on the interaction between cpSRP43 and cpSRP54 (39, 43, 44). CD2 alone can interact with cpSRP54 however the binding is not as strong as compared to binding between full-length cpSRP43 and cpSRP54 (42, 43). In a deletion mutant of cpSRP43 which did not contain CD2, there was no binding between the two subunits (45). This evidence revealed the

importance of CD2 in forming the cpSRP heterodimer. The M-domain of cpSRP54 mediates binding to cpSRP43 (39, 40). In integration studies using a CD1 deletion mutant of cpSRP43, the removal of CD1 did not affect the formation of transit complex but it did lead to a reduction in integration thereby suggesting that CD1 plays a role in regulating the GTPase activity of cpSRP54 at the receptor (44).

L18, an 18-amino acid, hydrophilic region between the second and third transmembrane domains of LHCP, binds to the ANK repeat region of cpSRP43, which greatly contributes to the formation of transit complex (42). Upon deletion of the ANK regions of cpSRP43, it was proven unable to provide chaperoning for LHCP (46). The chromodomains were also shown to be crucial for the chaperoning effect of cpSRP43 on LHCPs as a CD2 and CD3 deletion mutant of cpSRP43 was not able to inhibit aggregation of LHCPs (46). The L18 portion of LHCP has a lengthened structure except for its DPLG (Asp-Pro-Leu-Gly) motif which is characteristic of type I turns (39). Interaction between LHCP and cpSRP43 is primarily hydrophobic which has been displayed by co-purification at salt concentrations up to 1M as well as interactions between TM3 of LHCP and cpSRP43 (28, 41). In addition to hydrophobic interactions, the recognition of the DPLG motif by cpSRP43 is required for the formation of a transit complex (28, 39, 47). Binding between L18 and cpSRP43 has a  $K_d$  around  $1\mu\text{M}$  (39). The DPLG motif associates around the tyrosine<sub>204</sub> hook region on ANK4 of cpSRP43 and resembles the way in which this motif associates with the hydroxyls of carotenoids (41). A Y204A mutant of cpSRP43 presented a substantial reduction in binding affinity to the L18 peptide, proving this residue to be essential for binding (41).

cpSRP and its receptor complexes with the translocase, Alb3, which is necessary for the integration of LHCP into the thylakoid membrane. Dissociation of the cpSRP-LHCP complex

by the c-terminal of Alb3 (cAlb) at the thylakoid membrane takes place upon GTP hydrolysis by cpSRP54 and cpFtsY and it is dependent on an interaction between cAlb and cpSRP43 (48). Alb3 stimulates GTP hydrolysis upon binding to cpSRP43 which is essential for the release of LHCP (48). This interaction occurs between the ANK regions and CD2CD3 domains on cpSRP43 and cAlb3 (48-51). Binding studies between CD2 and CD3 domain constructs and cAlb show a dissociation constant of 20.6 $\mu$ M between CD2CD3 and cAlb and a slight reduction upon the removal of CD3 (51). CD2 has been proposed to be central to the binding event between the two proteins (49, 51). Dissociation constants between full length cpSRP43 and cAlb have been reported at 5.1 and 9.7 $\mu$ M (49, 51). In another study the binding affinity has been reported in the nanomolar range (48). Although there are discrepancies in ITC data reports, possibly attributed to differences in buffer compositions, the binding affinities do indicate an important role for the ANK regions of cpSRP43 within this interaction (48, 49, 51). Positively charged motifs at the center and c-terminus of cAlb have been implicated in the binding interaction with cpSRP43, which has been demonstrated by single mutations of lysine and arginine to alanine in motif IV at the extreme c-terminal resulting in lowered binding affinity and mutations of both simultaneously resulting in total loss of binding (49, 51). An interaction between the fifth transmembrane domain of Alb3, located near the luminal side of the thylakoid membrane, and cpSRP43 has been identified for contributing to the binding event between cAlb and cpSRP43 (50). In competition pull-down assays using cpSRP43, cAlb and the M-domain of cpSRP54 revealed that the cpSRP43 and the M-domain did not bind non-specifically to the resin used and M-domain did not bind to cAlb indicating that cpSRP43 can bind to cAlb and cpSRP54 simultaneously (49). Given the multiple interactions between cpSRP43 and the cpSRP pathway

components, cpSRP43 has proven to be a versatile binding partner central to the success of this pathway.

In addition to the pivotal role of cpSRP43 in the translocation and integration of LHCPs to the thylakoid membrane, cpSRP43 functions as a chaperone and disaggregase for the highly hydrophobic LHCPs to traverse the stroma. It has been established that cpSRP43 is highly specific for chaperoning LHCPs (28). In studies which monitored the aggregation of LHCPs in the presence of each component from the cpSRP pathway, cpSRP43 was shown to be the essential component for preventing the aggregation of LHCPs (28). cpSRP43 has unique disaggregase activity for the fact it does not require ATPase activity as is the case for the Hsp104/ClpB family of chaperones, however its ability to dissolve aggregates of LHCPs is comparable to that family of disaggregases (28, 46, 52, 53). Hsp70, Hsp60 and trigger factor (TF), a bacterial chaperone involved with hydrophobic regions of proteins, were shown not to be able to substitute for the disaggregase and chaperoning activity of cpSRP43, indicating the specificity of cpSRP43 to LHCPs (28, 33). Hydrophobic interactions between the trans-membrane regions of LHCP and cpSRP43 are also considered to contribute largely to the ability of cpSRP43 to successfully deliver the LHCPs to their destination. Hydrophobic interactions were revealed based on studies in which binding was not lost under high salt conditions, a 60-fold higher binding affinity between the full length LHCP versus the L18 peptide alone and the involvement of TM3 of LHCP for efficient binding (28, 41, 47). Previous studies have shown cpSRP43 to be vital component of the cpSRP, the molecular machine dedicated to targeting LHCPs to the thylakoid membrane. Reports have also indicated that cpSRP43 exists in free form as a dimer in the stroma, as a potential intermediate, functioning in the formation of cpSRP as well as a disaggregase for adrift LHCP molecules (28, 40).

Molecular chaperones are part of a diversified classification due to their scope of specific functions (54). Proteins in this group maintain critical roles in folding and assembly, refolding, translocation and preventing aggregation (54). Chaperones also assist under stress and disease conditions. The biogenesis of a large quantity of membrane proteins typically takes place in conjunction with their insertion into the membrane, however, in posttranslational translocation of membrane proteins, highly effective chaperones are required. Molecular chaperones which can prevent, and/or reverse protein aggregation provide a significant contribution in the maintenance of cellular stability, proper function and prevention of disease states. Molecular chaperones known as "disaggregases" are capable of reversing protein aggregation with the aid of external energy provided by ATP hydrolysis as well as other co-chaperones (55, 56). cpSRP43 is the first proteins to be described as having ATP-independent disaggregase activity. cpSRP43 provides protection to the most abundant protein on earth, giving it a highly specific and critical function. To investigate how cpSRP43 can recognize and reconstruct LHCP aggregates, the characteristics and structure of LHCP aggregates were studied (52, 57). LHCP aggregates are highly stable, insoluble, resistant to detergents or reversal by dilution (52). The L18 peptide and TM3 are shown to be solvent exposed in the aggregate formations which presents as the recognition site for cpSRP43 to first interact with the LHCP aggregate and begin the disassembly and reorganization of each LHCP protein (52). Recent studies using NMR and smFRET have also shown cpSRP43 to have high conformational flexibility (58, 59).

cpSRP43 is an indispensable component of the cpSRP pathway, involved in interactions with multiple partners leading to the successful translocation, disaggregation, refolding and integration of LHCPs. The chaperoning function of cpSRP43 in the cpSRP pathway is likened to that of the Hsp104/ClpB family of proteins. Hsp104 and ClpB work in

combination with Hsp70 or DnaK to disaggregate and remodel proteins under severe stress conditions and requires energy from ATP to perform (60). Although this family of proteins are impressive disaggregating machines, it is noteworthy to consider that cpSRP43 has similar thermostability and, in addition, can disaggregate and remodel the LHCP molecules without the input of external energy. The ability of cpSRP43 to disassemble and remodel LHCP aggregates relies on its binding interactions with its substrate in place of any external energy source and its flexibility demonstrates its effectiveness and potential adaptability as a generic chaperone. As cpSRP43 is such a unique and versatile protein, it seems unlikely that it would have only one job. This leads to a probing interest into the potential roles of free cpSRP43.

### **Alb3/Oxa1/YidC Translocase Family**

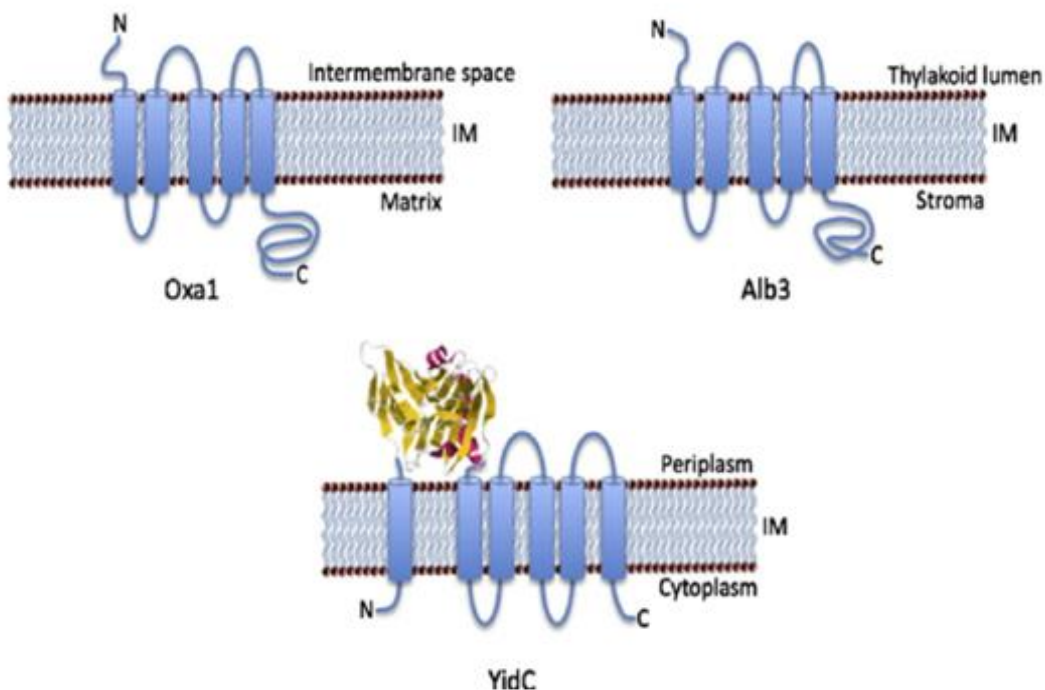
Alb3, Oxa1 and YidC are part of an evolutionarily conserved family of proteins involved in the folding, assembly and insertion of membrane proteins into thylakoid membranes within chloroplasts, inner mitochondrial and inner bacterial membranes, respectively (Fig. 5). It is maintained that the mitochondria and chloroplast organelles are of prokaryotic origins and evolved according to the theory of endosymbiosis. The thylakoid membrane of chloroplasts and the inner membrane of mitochondria are similar to the inner membrane of bacteria and therefore the process of protein localization for each follow similar mechanisms (61). These insertases share a conserved hydrophobic region comprised of five trans membrane domains in Alb3 and Oxa1 (61-63). YidC, belonging to gram positive bacteria, contains five trans membrane regions, however, YidC in gram negative bacteria, spans the bacterial inner membrane with six regions. YidC is a 60kDa protein responsible for the insertion of inner membranes proteins (IMPs) of bacterial membranes. YidC is involved in Sec-dependent and Sec-independent assembly of IMPs. Many bacterial IMPs are targeted to the Sec-YEG translocon pore by the SRP pathway

for export into the periplasm. Oxa1 is responsible for the insertion of nucleus and mitochondrial genome encoded proteins into the inner mitochondrial membrane. Oxa1 has a molecular weight of 36 kDa and spans the inner mitochondrial membrane with five helical trans membrane regions having an orientation of N (in at the intermembrane space) to C (out at the matrix) (64)

Investigation of Oxa1 mutants revealed that it plays a vital role in the assembly and insertion of respiratory complexes (62). Once Oxa1 was identified, homologs in chloroplasts and bacteria were soon discovered. Nuclear encoded proteins are targeted to the mitochondria by targeting sequences on the n-terminal of the polypeptides which is removed during or after import (64). In some cases, the target sequence is internal and therefore not cleaved. First, they are translocated across the outer and inner membrane by the outer membrane complex (TOM) and inner membrane complex (TIM) machinery. Movement across TIM requires the mt-Hsp70 chaperone which is ATP-driven. Most mitochondrial proteins are nuclear encoded however there are several respiratory complex proteins encoded by the mitochondrial genome which must be inserted into the inner membrane. Oxa1 interacts directly with nascent chain polypeptides undergoing synthesis as well as imported proteins destined for insertion. Interaction of the matrix exposed c-terminal of Oxa1 with the ribosome was determined to be necessary for coupling the translation and insertion of Cox11 (65). Removal of the shorter c-terminal present in YidC lead to reduction of nascent chain-YidC complexes observed *in vitro* (66). Both Oxa1 and YidC have been shown to form dimers above the exit of the ribosome tunnel of translating bacterial ribosomes. These dimers are believed to form translocation pores similar to the Sec translocon in bacteria (66). In the case of Sec-independent proteins, YidC operates on its own with the ability to stabilize the substrate and achieve insertion, although the mechanism has not been fully explored (62). There is a proposed theory that an electrostatic interaction between



polar residues of the trans membrane segments and negatively charged regions of the proteins to be inserted begets insertion by decreasing the transition energy although this theory warrants further investigation (63). Mutational studies have revealed key residues in TMDs 1, 2 and 6 in YidC and TMDs 1, 2 and 5 in Oxa1 are critical for activity (63). Regions specific to Alb3/Oxa1/YidC are present within the conserved hydrophobic core to allow for the association of other components which contribute to each system and this core domain can be substituted between different members of the family and retain functionality (67, 68). Alb3 and Oxa1 can substitute YidC in *E. Coli*. resulting in successful membrane protein insertion (69, 70). YidC of *E. Coli*. can replace Oxa1 in mitochondria if the c-terminal domain is affixed with that of YidC (71). The Alb3 homolog is present in co and post-translational thylakoid localization of proteins in the cpSRP pathway and the Sec-dependent pathways (49, 62, 64, 72).



**Fig. 5: Diagram of the Alb3/YidC/Oxa1 family of translocases. The hydrophobic core and c-terminal portion is conserved in each translocon. The c-terminal region is responsible for important interactions in each system. This picture is adapted from Wang, P and Ross E. Dabley, *Biochimica et Biophysica Acta*. (2011).**

Each member of this family has conserved roles in translocase activity although not identical. YidC and Oxa1 have been shown to play an important part in the proper folding and assembly of proteins as well. Alb3 is also involved in folding and assembly of thylakoid membrane proteins, much like the functions of YidC and Oxa1(63). In plants that do not express Albino3, an albino phenotype is observed, and the resultant seedlings are lethal (73, 74). Oxa1, Alb3 and YidC of gram-positive bacteria each have hydrophilic c-terminal extensions important for interactions with members of each system. The c-terminal domains of these three insertases are notably variable (67). Although the function of these insertases are conserved, the primary sequence similarity is low. The C-terminal extension of Oxa1 interacts with the ribosome for coupling translation and insertion (64). The C-terminal of YidC is important for associating the nascent-chain YidC complexes at the membrane while the C-terminal of Albino3 facilitates association with cpSRP at the thylakoid membrane through binding to cpSRP43 (62).

### **Intrinsically Disordered Proteins**

Proteins attain unique, ordered, equilibrium structures due to fixed positions within their backbone angles and atoms which is implicit to the structure-function paradigm (75). Intrinsically disordered proteins or intrinsically disordered regions of proteins (IDPs/IDRs), however, are abundant in the proteome of all living things and carry out important functions despite the absence of a specific fold (76). IDPs are being investigated to a greater extent as of recently to understand their functional and structural mechanisms by which they bind to partners. The stable conformations of protein structures are dictated by the primary amino acid sequence. In the case of IDPs, there is a tendency for the primary sequence to contain a low distribution of aromatic and hydrophobic residues thereby containing higher proportions of polar and charged residues. Elevated overall charge in combination with a low hydropathy index makes them

inadequate to form compact structures or become scaffolding/membrane proteins (77). Since IDPs do not need to obtain structure spontaneously, the area along their sequences is lengthened. The low sequence complexity of IDPs results in higher structural complexity (77). Another trademark of IDP function is that their main utility for mediating interactions is through short linear motifs (SLiMs). IDPs are well characterized for their roles in regulatory and signaling processes because of the ability of disordered portions to interact with multiple partners (78). This attribute allows the same protein to be used in several different pathways (78). The expression of IDPs in the cell must be strongly regulated to ensure accuracy and balance in signaling pathways. Divergence from the appropriate availability of IDPs can result in their binding to off-target partners or binding each other, giving rise to aggregation. IDPs have been implicated as the culprit in diseases such as cancer and neurodegeneration (78). Upon findings of important roles for IDPS within the cell, focus has shifted toward a better understanding of inherently unstructured proteins. Whereas the structure-function paradigm has already been well established for structured proteins the disorder-structure paradigm is emerging for IDPs (79). Computational methods in conjunction with experimental results, expressly from NMR, SAXS, CD and small-molecule techniques have been successful in elucidating distinct ensembles for these obscure proteins (80). Since a vast majority of unstructured proteins are uncharacterized, the development of new methods for this challenging task is anticipated to lead to exciting new insights into known biological mechanisms (79).

IDPs tend to lack heavy hydrophobic amino acids in their sequence which excludes them from having the ability to form stable, organized hydrophobic cores that give rise to structured domains (81, 82). IDRs were initially considered to be linker segments within structured proteins however it is now apparent that these regions participate in various functions carried out

by proteins and in some cases disordered regions/proteins are directly responsible for biological functions (79). These disordered regions often undergo post-translational modifications which augment to functional properties of proteins within the cell (83, 84). The many different properties and characteristics of IDPs along with their growing prevalence have induced the need for proper classification of the proteins which is still a work in progress. Currently IDPS can fall under the classifications of “intrinsically disordered” for incompletely folded regions and proteins, "intrinsically unstructured or natively unfolded" which denotes random-coil-like or globular forms and, and recently polyglutamine has been described as a collapsed random coil however it could also be referred to as natively unfolded (85, 86). It was previously described as being intrinsically disordered and now with the utilization of fluorescence correlation spectroscopy the ensemble of polyglutamine is revealed to be made up of a collection of collapsed structures (86). Further advancement in the studies of IDPs is leading to expanded distinction between them and in the case of polyglutamine a better understanding of the kinetics and thermodynamics involved in the aggregation process which is resultant in disease. In addition to low hydrophobicity, high net charge is a typical property of IDPs (85, 87). Characterization of conformational ensembles of arginine-rich IDPs (protamines), in aqueous surroundings, using a combination of molecular simulations and fluorescence measurements were used to distinguish which behavioral phase is predicted to be preferred. It was revealed that the net charge per residue was responsible for determining the behavior of the protein and which phase was preferred (87). In this case, protamines favored a globule- to- coil transition phase (87). This theory concluded that net charge can modulate the inclination of polypeptide backbones towards disorder. Crowding inside of cells has also been proposed as a condition in which IDPS may gain 3D structure. Experiments conducted to test this theory involve inducing

molecular crowding. *In vitro* NMR data of the disordered protein hallmark of Parkinson's disease,  $\alpha$ -synuclein, upon subjection to high glucose conditions showed that disorder remained for this protein (88). Preliminary results suggest that any fold or lack of fold in IDPs is not significantly dictated by environment however additional experiments involving other types of IDPs would be necessary to make any concrete conclusion. In the amino acid sequences of IDPs there are residues labeled as being order-promoting (cysteine, tryptophan, tyrosine, isoleucine, phenylalanine, valine, leucine, histidine, threonine, and asparagine) residues which are typically depleted in structured proteins or disorder-promoting (aspartic acid, methionine, lysine, arginine, serine, glutamine, proline, and glutamic acid) which are naturally enriched in disordered proteins (89, 90). Order predictions can thereby be made by profiling the amino acid composition of proteins which is used in a number of disorder prediction software commensurate with DisProt and PONDR (91, 92). This type of amino acid sequence analysis may be used as a starting point for comprehensive experimentation.

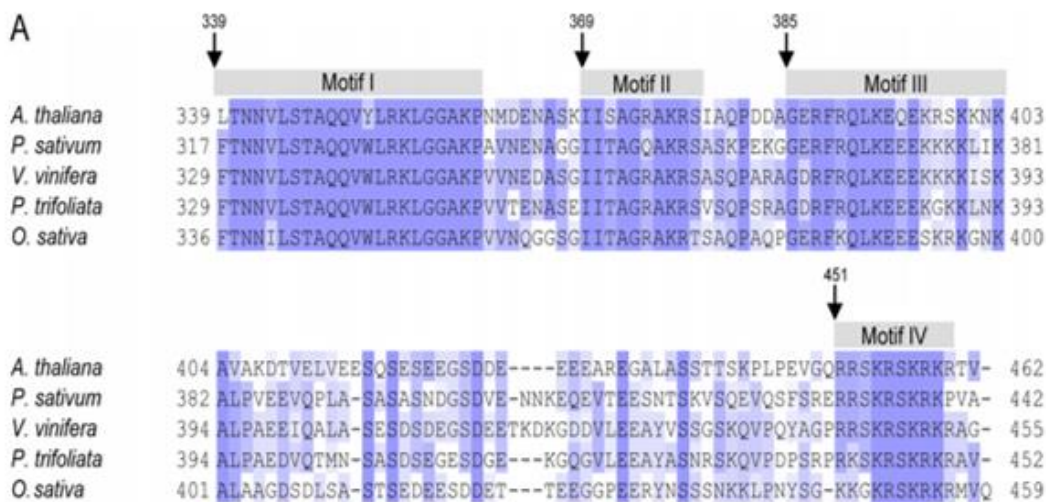
NMR is a particularly powerful tool for studying IDPs and partially folded proteins due to the structural observations which can be made during protein folding (93). Specifically, residual dipolar couplings (RDCs), SAXS and conformational sampling have proven to be a sensitive and effective technique for characterizing structural behavior in IDPs (94). In a study done by Bernardo et al., a structural model for a disordered 57 amino acid stretch within the nucleocapsid-binding domain of the Sendai virus phosphoprotein was determined combining these three methodologies. First conformational ensembles were generated using an algorithm called flexible-meccano then the structural ensembles were used to simulate small-angle x-ray scattering and averaged RDCs data. Simulated and experimental data from RDCs and small-angle x-ray scattering corroborated well with each other and provided a clear-cut characterization

of the dynamics, average dimensions and local structure of the ensemble of sampled structures (95). Data from RDCs can provide orientational constraints and dynamics on the microsecond to millisecond timescale (96). Data derived from small-angle scattering analysis can provide comprehensive information on the shape, size and average dimensions of molecules in solution (95). smFRET is a superiorly sensitive technique for studying the structure and unfolding processes of proteins (97). In smFRET experiments, there is a Donor and an Acceptor dye which form a FRET pair that is installed at specific sites in the protein (98). Analysis of the data reveals information regarding the dynamics and of the protein region which is interceding the two dyes. Combination of multiple biophysical techniques, computational algorithms and the information that can initially be derived from the amino acid sequence present a compelling system for studying natively unfolded proteins.

### **The c-terminal domain of Albino3**

Albino3 is the insertase present in thylakoid membranes of chloroplasts, critical for the post-translational integration of LHCPs and co-translational integration of other chloroplast membrane proteins. The c-terminal of Albino3 (cAlb) is an intrinsically disordered region of Albino3 which protrudes out into the stroma and facilitates association of the cpSRP at the thylakoid membrane through recognition and association with the cpSRP43 subunit. cAlb, upon binding to cpSRP43, gains a more compact structure as demonstrated in circular dichroism experiments (49). Association between cAlb and cpSRP43 also activates GTP hydrolysis by cpSRP54 and cpFtsY which is required for the release of LHCP (48). *In vitro* binding experiments have revealed an interaction between cAlb and LHCP (99). The cpSRP-LHCP complex was also found to bind more efficiently to cAlb than cpSRP alone. This finding indicates that cpSRP loaded with LHCP is critical for the interaction with cAlb and may also

prevent Albino3 from being blocked by cpSRP which is not part of the transit complex (99). There is no binding reported between cpSRP54 and cAlb which steers toward the binding between cpSRP43 and cAlb as the critical interaction taking place at the thylakoid membrane (50). Many investigations have been launched to determine the exact binding interactions between cAlb and cpSRP43. Within cpSRP43, the ankyrin repeat regions and CD2CD3 have been implicated as the important regions for binding with cAlb (48-50). Amino acid sequence alignments of Albino3 from different species have revealed four conserved, positively charged motifs (I-IV) in the stromal c-terminal region ((49), Fig. 6).



**Fig 6: Multiple sequence alignments of cAlb domains from different plants highlighting the four conserved positively charged motifs. This picture is adapted from Falk, S JBC (2010).**

ITC binding experiments revealed that full length cAlb bound to cpSRP43 had a  $K_d$  of  $9.7\mu\text{M}$  and cAlb lacking motif I had a  $K_d$  of  $11.4\mu\text{M}$  when titrated into cpSRP43, with the difference in  $K_d$  being insignificant it was suggested that motif I was not critical to the binding interaction (49). In another study, a synthetic motif III peptide was unable to bind to cpSRP43 and further

truncations of cAlb from both the c and n-terminus showed lowered binding affinities (49). The conclusion that motifs II and IV are required for the interaction with cpSRP43 was made (49). Comparison of motifs II and IV reveal AKRS and SKRS sequences which are like the ARK(S/T) sequences of histone H3 tails (49, 51). This finding suggests that there is a likely affinity between these motifs and the chromodomains of cpSRP43 (49, 51). In studies in which lysine and arginine residues were mutated to alanine in in these motifs led to a reduction in affinity to cpSRP43 and binding did not occur with double mutations in both regions (49). Binding affinities between the ANK (1-4) regions and cAlb were found to be lower than what was observed for full length cpSRP43 and cAlb however it was in the nanomolar range implying that the ANK regions are essential to binding cAlb or at least for stabilizing the interaction between the two proteins (48). In an *in planta* study, a cAlb which was lacking motifs III and IV was compared to the wild type under normal light growth conditions which resulted in an 80% reduction of functionality when compared to the wild type (100). It is possible that the instability, faster degradation, and inability to interact with other proteins was a characteristic of the truncated form of cAlb. Truncated cAlb, grown under continuous low light conditions was able to accumulate and integrate LHCPs, with only slight reduced functionality compared to wild type (100). With the consensus of information, the study concluded that motifs I, II and TM5 were the areas of Albino3 important for sufficient interaction with cpSRP43 and insertion of LHCP into the membrane (100). Conflicting reports on which regions of cAlb are critical for binding cpSRP43 and initiating integration of LHCP have prompted further investigation of the cAlb structure and the regions which bind to its protein partners during the integration of LHCPs. This study combines biophysical techniques with computational methods to provide a thorough



examination of the c-terminal domain of Albino3 which will prove to be beneficial in the overall elucidation of the cpSRP pathway.

### **Scope of Study**

The comprehensive aim of the research project was to gain further insight into the cpSRP pathway by examining key players and their imperative roles. The cpSRP pathway is essential for the acquisition of light energy by plants through the proper translocation and integration of LHCPs into the thylakoid membrane of chloroplasts. cpSRP43 is central to the cpSRP pathway and this study characterizes the structure and stability of cpSRP43. The study also designates the ability of cpSRP43 to function as a generic chaperone. Given the distinctive domains and flexibility of cpSRP43, it can interact with diverse partners in the pathway to achieve proper translocation, integration, disaggregation and even refolding of LHCPs. cpSRP43 can disassemble and remodel LHCP aggregates based solely on the binding interactions between it and its substrate, granting it unique chaperoning and disaggregating capability in comparison to other molecular chaperones which require an external energy source and additional protein partners for activity. The unique structure and flexibility of cpSRP43 permits its adaptability and effectiveness as a generic chaperone. Although it is well established that LHCPs specifically require cpSRP43 as a chaperone, it stands to reason that cpSRP43 may also serve as a general chaperone which prevents aggregation and assists in the refolding of other proteins within and outside of chloroplasts. In the first part of this study we examine the structural characteristics and heat stability of cpSRP43 then continue forward to evaluate its ability to assist in refolding and prevention of heat-induced aggregation of proteins outside of the chloroplast. The significance of affirming cpSRP43 as a generic chaperone is substantial as further investigation will lead to the utilization of cpSRP43 as a protein protectant in other systems.

In the second part of this study, we examine the structure of cAlb. The role of cAlb in the cpSRP pathway is fundamental to the proper integration of LHCPs by way of its interaction with cpSRP43 at the thylakoid membrane. cAlb is considered an intrinsically disordered region of Albino3 which extends into the stroma and adopts an alpha helical structure upon binding to cpSRP43. Here we investigate whether cAlb contains any inherent structure as it floats in the stroma before binding to cpSRP43. The cAlb sequence from *Pisum sativum* was analyzed by disorder and secondary structure predicting software and in each case the output data indicated that the same stretch of amino acids in the N-terminal region of cAlb exhibited a propensity towards having an alpha helical structure. Based on the predictions, four amino acids in that region were selected for site-directed mutagenesis to glycine. Single amino acid mutations of cAlb were made and structural data was collected for comparison with the wild type protein. Double cysteine mutants were also analyzed using smFRET. By deciphering the distance between two labelled residues around the area predicted to contain a helix, the length in angstroms will reveal if a helix could be present or if that area of the protein is likely to be extended. Determining structural information at the residual level will lead to a better understanding of how cAlb is capable of locating and binding cpSRP43 to properly position the transit complex at the thylakoid membrane. In addition, binding details between cAlb and cpSRP43 are examined as well as the stability of the complex. Binding studies between cAlb and cpSRP43 have revealed areas on both proteins which may be crucial for the interaction however exact residual details of this interaction have yet to be elucidated. Here we utilized different biophysical techniques in combination with computational methods to determine binding affinities, stability of the complex and the proximity of different regions of the two

molecules upon the binding interaction. The data acquired here in comparison with previously published data will provide additional details to this crucial step in the cpSRP pathway.

## References

1. Gorodkin J, Knudsen B, Zwieb C, Samuelsson T. SRPDB (signal recognition particle database). *Nucleic acids research*. 2001;29(1):169-70.
2. Pool MR. Signal recognition particles in chloroplasts, bacteria, yeast and mammals (Review). *Molecular Membrane Biology*. 2005;22(1-2):3-15.
3. Walter P, Gilmore R, Blobel G. Protein translocation across the endoplasmic reticulum. *Cell*. 1984;38(1):5-8.
4. Luirink J, Dobberstein B. Mammalian and Escherichia coli signal recognition particles. *Molecular microbiology*. 1994;11(1):9-13.
5. Whitford D. *Proteins: structure and function*: John Wiley & Sons; 2013.
6. Keenan RJ, Freymann DM, Stroud RM, Walter P. The signal recognition particle. *Annual review of biochemistry*. 2001;70(1):755-75.
7. Bovia F, Bui N, Strub K. The heterodimeric subunit SRP9/14 of the signal recognition particle functions as permuted single polypeptide chain. *Nucleic acids research*. 1994;22(11):2028-35.
8. Akopian D, Shen K, Zhang X, Shan S-o. Signal Recognition Particle: An Essential Protein-Targeting Machine. *Annual Review of Biochemistry*. 2013;82(1):693-721.
9. Gundelfinger ED, Krause E, Melli M, Dobberstein B. The organization of the 7SL RNA in the signal recognition particle. *Nucleic acids research*. 1983;11(21):7363-74.
10. Ullu E, Tschudi C. Alu sequences are processed 7SL RNA genes. *Nature*. 1984;312(5990):171.
11. Egea PF, Napetschnig J, Walter P, Stroud RM. Structures of SRP54 and SRP19, the two proteins that organize the ribonucleic core of the signal recognition particle from *Pyrococcus furiosus*. *PLoS One*. 2008;3(10):e3528.
12. Siegel V, Walter P. Each of the activities of signal recognition particle (SRP) is contained within a distinct domain: analysis of biochemical mutants of SRP. *Cell*. 1988;52(1):39-49.
13. Schlenker O, Hendricks A, Sinning I, Wild K. The structure of the mammalian signal recognition particle (SRP) receptor as prototype for the interaction of small GTPases with Longin domains. *Journal of Biological Chemistry*. 2006;281(13):8898-906.
14. Rapiejko PJ, Gilmore R. Empty site forms of the SRP54 and SR $\alpha$  GTPases mediate targeting of ribosome–nascent chain complexes to the endoplasmic reticulum. *Cell*. 1997;89(5):703-13.
15. Nagai K, Oubridge C, Kuglstatter A, Menichelli E, Isel C, Jovine L. Structure, function and evolution of the signal recognition particle. *The EMBO journal*. 2003;22(14):3479-85.

16. Zheng N, Gierasch LM. Domain interactions in E. coli SRP: stabilization of M domain by RNA is required for effective signal sequence modulation of NG domain. *Molecular cell*. 1997;1(1):79-87.
17. Samuelsson T. A Mycoplasma protein homologous to mammalian SRP54 recognizes a highly conserved domain of SRP RNA. *Nucleic acids research*. 1992;20(21):5763-70.
18. Hauser S, Bacher G, Dobberstein B, Lütcke H. A complex of the signal sequence binding protein and the SRP RNA promotes translocation of nascent proteins. *The EMBO journal*. 1995;14(22):5485-93.
19. Wild K, Bange G, Bozkurt G, Segnitz B, Hendricks A, Sinning I. Structural insights into the assembly of the human and archaeal signal recognition particles. *Acta Crystallographica Section D: Biological Crystallography*. 2010;66(3):295-303.
20. Wild K, Weichenrieder O, Strub K, Sinning I, Cusack S. Towards the structure of the mammalian signal recognition particle. *Current opinion in structural biology*. 2002;12(1):72-81.
21. Powers T, Walter P. Co-translational protein targeting catalyzed by the Escherichia coli signal recognition particle and its receptor. *The EMBO Journal*. 1997;16(16):4880-6.
22. Bernstein HD, Zopf D, Freymann DM, Walter P. Functional substitution of the signal recognition particle 54-kDa subunit by its Escherichia coli homolog. *Proceedings of the National Academy of Sciences*. 1993;90(11):5229-33.
23. Bauer NC, Doetsch PW, Corbett AH. Mechanisms regulating protein localization. *Traffic*. 2015;16(10):1039-61.
24. Berks BC, Palmer T, Sargent F. Protein targeting by the bacterial twin-arginine translocation (Tat) pathway. *Current opinion in microbiology*. 2005;8(2):174-81.
25. Jarvis P, Robinson C. Mechanisms of protein import and routing in chloroplasts. *Current Biology*. 2004;14(24):R1064-R77.
26. Jansson S. A guide to the Lhc genes and their relatives in Arabidopsis. *Trends in plant science*. 1999;4(6):236-40.
27. Groves MR, Mant A, Kuhn A, Koch J, Dübel S, Robinson C, et al. Functional characterization of recombinant chloroplast signal recognition particle. *Journal of Biological Chemistry*. 2001.
28. Falk S, Sinning I. cpSRP43 is a novel chaperone specific for light-harvesting chlorophyll a, b-binding proteins. *Journal of Biological Chemistry*. 2010;285(28):21655-61.
29. Gutensohn M, Fan E, Frielingsdorf S, Hanner P, Hou B, Hust B, et al. Toc, Tic, Tat et al.: structure and function of protein transport machineries in chloroplasts. *Journal of Plant Physiology*. 2006;163(3):333-47.
30. Jarvis P, Soll J. Erratum to: "Toc, tic, and chloroplast protein import" [*Biochim. Biophys. Acta* 1541 (2001) 64–79]. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*. 2002;1590(1-3):177-89.

31. Nilsson R, Brunner J, Hoffman NE, van Wijk KJ. Interactions of ribosome nascent chain complexes of the chloroplast-encoded D1 thylakoid membrane protein with cpSRP54. *The EMBO Journal*. 1999;18(3):733-42.
32. Payan LA, Cline K. A stromal protein factor maintains the solubility and insertion competence of an imported thylakoid membrane protein. *The Journal of cell biology*. 1991;112(4):603-13.
33. 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. *Proceedings of the National Academy of Sciences*. 1993;90(18):8552-6.
34. Schuenemann D, Amin P, Hoffman NE. Functional divergence of the plastid and cytosolic forms of the 54-kDa subunit of signal recognition particle. *Biochemical and biophysical research communications*. 1999;254(1):253-8.
35. Li X, Henry R, Yuan J, Cline K, Hoffman NE. A chloroplast homologue of the signal recognition particle subunit SRP54 is involved in the posttranslational integration of a protein into thylakoid membranes. *Proceedings of the National Academy of Sciences*. 1995;92(9):3789-93.
36. Schuenemann D, Gupta S, Persello-Cartieaux F, Klimyuk VI, Jones JDG, Nussaume L, et al. A novel signal recognition particle targets light-harvesting proteins to the thylakoid membranes. *Proceedings of the National Academy of Sciences*. 1998;95(17):10312-6.
37. Tu C-J, Schuenemann D, Hoffman NE. 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*. 1999;274(38):27219-24.
38. Tu CJ, Peterson EC, Henry R, Hoffman NE. The L18 Domain of Light-harvesting Chlorophyll Proteins Binds to Chloroplast Signal Recognition Particle 43. *Journal of Biological Chemistry*. 2000;275(18):13187-90.
39. Stengel KF, Holdermann I, Cain P, Robinson C, Wild K, Sinning I. Structural basis for specific substrate recognition by the chloroplast signal recognition particle protein cpSRP43. *Science*. 2008;321(5886):253-6.
40. Jonas-Straube E, Hutin C, Hoffman NE, Schünemann D. Functional Analysis of the Protein-interacting Domains of Chloroplast SRP43. *Journal of Biological Chemistry*. 2001;276(27):24654-60.
41. Cain P, Holdermann I, Sinning I, Johnson AE, 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. *Biochemical Journal*. 2011;437(1):149-55.
42. Kathir KM, Rajalingam D, Sivaraja V, Kight A, Goforth RL, Yu C, et al. Assembly of Chloroplast Signal Recognition Particle Involves Structural Rearrangement in cpSRP43. *Journal of Molecular Biology*. 2008;381(1):49-60.

43. Hermkes R, Funke S, Richter C, Kuhlmann J, Schünemann D. The  $\alpha$ -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*. 2006;580(13):3107-11.
44. Goforth RL, Peterson EC, Yuan J, Moore MJ, Kight AD, Lohse MB, et al. Regulation of the GTPase cycle in post-translational signal recognition particle-based protein targeting involves cpSRP43. *Journal of Biological Chemistry*. 2004;279(41):43077-84.
45. Richter CV, Bals T, Schünemann D. Component interactions, regulation and mechanisms of chloroplast signal recognition particle-dependent protein transport. *European journal of cell biology*. 2010;89(12):965-73.
46. Jaru-Ampornpan P, Shen K, Lam VQ, Ali M, Doniach S, Jia TZ, et al. ATP-independent reversal of a membrane protein aggregate by a chloroplast SRP subunit. *Nature Structural & Molecular Biology*. 2010;17:696.
47. DeLille J, Peterson EC, Johnson T, Moore M, Kight A, Henry R. A novel precursor recognition element facilitates posttranslational binding to the signal recognition particle in chloroplasts. *Proceedings of the National Academy of Sciences*. 2000;97(4):1926-31.
48. Lewis NE, Marty NJ, Kathir KM, Rajalingam D, Kight AD, Daily A, et al. A dynamic cpSRP43-alb3 interaction mediates translocase regulation of cpSRP targeting components. *Journal of Biological Chemistry*. 2010.
49. 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*. 2010;285(8):5954-62.
50. 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*. 2011;286(40):35187-95.
51. Horn A, Hennig J, Ahmed YL, Stier G, Wild K, Sattler M, et al. Structural basis for cpSRP43 chromodomain selectivity and dynamics in Alb3 insertase interaction. *Nature Communications*. 2015;6:11.
52. Nguyen TX, Jaru-Ampornpan P, Lam VQ, Cao P, Piszkiwicz S, Hess S, et al. Mechanism of an ATP-independent protein disaggregase. I. Structure of a membrane protein aggregate reveals a mechanism of recognition by its chaperone. *Journal of Biological Chemistry*. 2013.
53. Jaru-Ampornpan P, Liang F-C, Nisthal A, Nguyen TX, Wang P, Shen K, et al. Mechanism of an ATP-independent protein disaggregase. II. Distinct molecular interactions drive multiple steps during aggregate disassembly. *Journal of Biological Chemistry*. 2013.
54. Hristozova N, Tompa P, Kovacs D. A novel method for assessing the chaperone activity of Proteins. *PloS one*. 2016;11(8):e0161970.

55. Doyle SM, Shorter J, Zolkiewski M, Hoskins JR, Lindquist S, Wickner S. Asymmetric deceleration of ClpB or Hsp104 ATPase activity unleashes protein-remodeling activity. *Nature Structural and Molecular Biology*. 2007;14(2):114.
56. Glover JR, Lindquist S. Hsp104, Hsp70, and Hsp40: a novel chaperone system that rescues previously aggregated proteins. *Cell*. 1998;94(1):73-82.
57. Jaru-Ampornpan P, Liang F-C, Nisthal A, Nguyen TX, Wang P, Shen K, et al. Mechanism of an ATP-independent protein disaggregase. II. Distinct molecular interactions drive multiple steps during aggregate disassembly. *Journal of Biological Chemistry*. 2013:jbc.M113.462861.
58. Gao F, Kight AD, Henderson R, Jayanthi S, Patel P, Murchison M, et al. Regulation of Structural Dynamics within a Signal Recognition Particle Promotes Binding of Protein Targeting Substrates. *Journal of Biological Chemistry*. 2015.
59. Liang F-C, Kroon G, McAvoy CZ, Chi C, Wright PE, Shan S-o. Conformational dynamics of a membrane protein chaperone enables spatially regulated substrate capture and release. *Proceedings of the National Academy of Sciences*. 2016:201524777.
60. Doyle SM, Wickner S. Hsp104 and ClpB: protein disaggregating machines. *Trends in biochemical sciences*. 2009;34(1):40-8.
61. Luirink J, Samuelsson T, de Gier J-W. YidC/Oxa1p/Alb3: evolutionarily conserved mediators of membrane protein assembly. *FEBS letters*. 2001;501(1):1-5.
62. Wang P, Dalbey RE. Inserting membrane proteins: the YidC/Oxa1/Alb3 machinery in bacteria, mitochondria, and chloroplasts. *Biochimica et Biophysica Acta (BBA)-Biomembranes*. 2011;1808(3):866-75.
63. Kumazaki K, Chiba S, Takemoto M, Furukawa A, Nishiyama K-i, Sugano Y, et al. Structural basis of Sec-independent membrane protein insertion by YidC. *Nature*. 2014;509(7501):516.
64. Stuart RA. Insertion of proteins into the inner membrane of mitochondria: the role of the Oxa1 complex. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*. 2002;1592(1):79-87.
65. Jia L, Dienhart M, Schrapf M, McCauley M, Hell K, Stuart RA. Yeast Oxa1 interacts with mitochondrial ribosomes: the importance of the C-terminal region of Oxa1. *The EMBO journal*. 2003;22(24):6438-47.
66. Kohler R, Boehringer D, Greber B, Bingel-Erlenmeyer R, Collinson I, Schaffitzel C, et al. YidC and Oxa1 form dimeric insertion pores on the translating ribosome. *Molecular cell*. 2009;34(3):344-53.
67. Funes S, Kauff F, van der Sluis EO, Ott M, Herrmann JM. Evolution of YidC/Oxa1/Alb3 insertases: three independent gene duplications followed by functional specialization in bacteria, mitochondria and chloroplasts. *Biological chemistry*. 2011;392(1-2):13-9.



68. Yuan J, Phillips GJ, Dalbey RE. Isolation of cold-sensitive yidC mutants provides insights into the substrate profile of the YidC insertase and the importance of transmembrane 3 in YidC function. *Journal of bacteriology*. 2007;189(24):8961-72.
69. van Bloois E, Nagamori S, Koningstein G, Ullers RS, Preuss M, Oudega B, et al. The Sec-independent function of *Escherichia coli* YidC is evolutionary-conserved and essential. *Journal of Biological Chemistry*. 2005;280(13):12996-3003.
70. Jiang F, Yi L, Moore M, Chen M, Rohl T, van Wijk K-J, et al. Chloroplast YidC homolog Albino3 can functionally complement the bacterial YidC depletion strain and promote membrane insertion of both bacterial and chloroplast thylakoid proteins. *Journal of Biological Chemistry*. 2002;277(22):19281-8.
71. Preuss M, Ott M, Funes S, Luirink J, Herrmann JM. Evolution of mitochondrial oxa proteins from bacterial YidC Inherited and acquired functions of a conserved protein insertion machinery. *Journal of Biological Chemistry*. 2005;280(13):13004-11.
72. Moore M, Goforth RL, Mori H, Henry R. Functional interaction of chloroplast SRP/FtsY with the ALB3 translocase in thylakoids: substrate not required. *The Journal of cell biology*. 2003;162(7):1245-54.
73. Hutin C, Havaux M, Carde JP, Kloppstech K, Meierhoff K, Hoffman N, et al. Double mutation cpSRP43–/cpSRP54–is necessary to abolish the cpSRP pathway required for thylakoid targeting of the light-harvesting chlorophyll proteins. *The Plant Journal*. 2002;29(5):531-43.
74. Moore M, Harrison MS, Peterson EC, 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*. 2000;275(3):1529-32.
75. Daughdrill GW, Pielak GJ, Uversky VN, Cortese MS, Dunker AK. Natively disordered proteins. *Protein folding handbook*. 2005:275-357.
76. Dunker AK, Brown CJ, Lawson JD, Iakoucheva LM, Obradović Z. Intrinsic disorder and protein function. *Biochemistry*. 2002;41(21):6573-82.
77. Uversky VN. Dancing protein clouds: the strange biology and chaotic physics of intrinsically disordered proteins. *Journal of Biological Chemistry*. 2016:jbc. R115. 685859.
78. Babu MM, van der Lee R, de Groot NS, Gsponer J. Intrinsically disordered proteins: regulation and disease. *Current opinion in structural biology*. 2011;21(3):432-40.
79. Van Der Lee R, Buljan M, Lang B, Weatheritt RJ, Daughdrill GW, Dunker AK, et al. Classification of intrinsically disordered regions and proteins. *Chemical reviews*. 2014;114(13):6589-631.
80. Rezaei-Ghaleh N, Blackledge M, Zweckstetter M. Intrinsically disordered proteins: from sequence and conformational properties toward drug discovery. *ChemBioChem*. 2012;13(7):930-50.

81. Uversky VN. Natively unfolded proteins: a point where biology waits for physics. *Protein science*. 2002;11(4):739-56.
82. Uversky VN, Gillespie JR, Fink AL. Why are “natively unfolded” proteins unstructured under physiologic conditions? *Proteins: structure, function, and bioinformatics*. 2000;41(3):415-27.
83. Iakoucheva LM, Radivojac P, Brown CJ, O’Connor TR, Sikes JG, Obradovic Z, et al. The importance of intrinsic disorder for protein phosphorylation. *Nucleic acids research*. 2004;32(3):1037-49.
84. Collins MO, Yu L, Campuzano I, Grant SG, Choudhary JS. Phosphoproteomic analysis of the mouse brain cytosol reveals a predominance of protein phosphorylation in regions of intrinsic sequence disorder. *Molecular & Cellular Proteomics*. 2008;7(7):1331-48.
85. Dunker AK, Oldfield CJ, Meng J, Romero P, Yang JY, Chen JW, et al. The unfoldomics decade: an update on intrinsically disordered proteins. *BMC genomics*. 2008;9(2):S1.
86. Crick SL, Jayaraman M, Frieden C, Wetzel R, Pappu RV. Fluorescence correlation spectroscopy shows that monomeric polyglutamine molecules form collapsed structures in aqueous solutions. *Proceedings of the National Academy of Sciences*. 2006;103(45):16764-9.
87. Mao AH, Crick SL, Vitalis A, Chicoine CL, Pappu RV. Net charge per residue modulates conformational ensembles of intrinsically disordered proteins. *Proceedings of the National Academy of Sciences*. 2010;107(18):8183-8.
88. McNulty BC, Young GB, Pielak GJ. Macromolecular crowding in the Escherichia coli periplasm maintains  $\alpha$ -synuclein disorder. *Journal of molecular biology*. 2006;355(5):893-7.
89. Radivojac P, Iakoucheva LM, Oldfield CJ, Obradovic Z, Uversky VN, Dunker AK. Intrinsic disorder and functional proteomics. *Biophysical journal*. 2007;92(5):1439-56.
90. Williams R, Obradovic Z, Mathura V, Braun W, Garner E, Young J, et al. The protein non-folding problem: amino acid determinants of intrinsic order and disorder. *Biocomputing 2001: World Scientific; 2000*. p. 89-100.
91. Xue B, Dunbrack RL, Williams RW, Dunker AK, Uversky VN. PONDR-FIT: A meta-predictor of intrinsically disordered amino acids. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*. 2010;1804(4):996-1010.
92. Sickmeier M, Hamilton JA, LeGall T, Vacic V, Cortese MS, Tantos A, et al. DisProt: the database of disordered proteins. *Nucleic acids research*. 2006;35(suppl\_1):D786-D93.
93. Dyson HJ, Wright PE. Unfolded proteins and protein folding studied by NMR. *Chemical reviews*. 2004;104(8):3607-22.
94. Jensen MR, Markwick PR, Meier S, Griesinger C, Zweckstetter M, Grzesiek S, et al. Quantitative determination of the conformational properties of partially folded and intrinsically disordered proteins using NMR dipolar couplings. *Structure*. 2009;17(9):1169-85.

95. Bernado P, Blanchard L, Timmins P, Marion D, Ruigrok RW, Blackledge M. A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering. *Proceedings of the National Academy of Sciences*. 2005;102(47):17002-7.
96. Dyson HJ, Wright PE. Elucidation of the protein folding landscape by NMR. *Methods in enzymology*. 394: Elsevier; 2005. p. 299-321.
97. Tompa P. Intrinsically disordered proteins: a 10-year recap. *Trends in biochemical sciences*. 2012;37(12):509-16.
98. Berger CL. Breaking the millisecond barrier: single molecule motors wobble to find their next binding sites. *Biophysical journal*. 2013;104(6):1219.
99. 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*. 2010;584(19):4138-44.
100. Urbischek M, Nick von Braun S, Brylok T, Gügel IL, Richter A, Koskela M, et al. The extreme Albino3 (Alb3) C terminus is required for Alb3 stability and function in *Arabidopsis thaliana*. *Planta*. 2015;242(3):733-46.

## **II. Structure and Stability of cpSRP43**

### **Abstract**

The novel multidomain protein, cpSRP43, is a unique subunit of the post-translational cpSRP targeting pathway in higher plants. The cpSRP pathway is responsible for targeting and insertion of light-harvesting chlorophyll a/b binding proteins (LHCPs) to the thylakoid membrane. Nuclear-encoded LHCPs are synthesized in the cytoplasm then imported into the chloroplast. Upon emergence into the stroma, LHCPs, form a soluble transit complex with the cpSRP heterodimer, which is composed of cpSRP43 and cpSRP54, a 54 kDa subunit homologous to the universally conserved GTPase in cytosolic SRP pathways. cpSRP43 is irreplaceable as a chaperone to LHCPs in their translocation to the thylakoid membrane and remarkable in its ability to dissolve aggregates of LHCPs without the need for external energy input. In previous studies, cpSRP43 has demonstrated significant flexibility and interdomain dynamics. However, the high flexibility and structural dynamics of cpSRP43 is yet unexplained by current crystal structures of cpSRP43. This is due, in part, to the fact that free full length cpSRP43 is so flexible that it is unable to crystalize. In this study, we explore the structural stability of cpSRP43 under different conditions using various biophysical techniques and find that this protein is concurrently highly stable and flexible. This conclusion is interesting considering that stable proteins typically possess a non-dynamic structure. Molecular dynamics simulations which correlated with data from biophysical experimentation were used to explain the basis of the extraordinary stability of cpSRP43. The data in agreement with the computational models allow us to obtain a highly detailed perspective on the conformational landscape of these proteins.

## Introduction

The Signal Recognition Particle (SRP), SRP receptor and translocase are components of a conserved and ubiquitous protein targeting system. The SRP pathway, mediated by the ribosome, co-translationally targets proteins to the endoplasmic reticulum (ER) of eukaryotes and the inner membrane of prokaryotes. The chloroplast signal recognition particle (cpSRP), in conjunction with its receptor (cpFtsY), can target proteins co- and post-translationally to the thylakoid membrane. cpSRP is composed of a 54kDa GTPase (cpSRP54) subunit, homologous to the 54kDa subunit in the canonical pathway, and a 43kDa subunit (cpSRP43) unique to the chloroplast. In the chloroplast, there are two pools of cpSRP54 which give rise to dual functionality. One pool of cpSRP54 associates with the ribosome and co-translationally targets proteins to their destination, while the other pool of cpSRP54 binds to cpSRP43, forming the cpSRP heterodimeric complex, and post-translationally targets light-harvesting chlorophyll-binding proteins (LHCs) to the thylakoid membrane (1, 2). LHCs function as an antenna complex to focus photons to produce high energy electrons (3-5). LHCs account for almost half of the membrane proteins in the thylakoids and are the most abundant protein on earth (6-8). cpSRP43 evolved in the chloroplast to aid in the translocation of LHCs without requiring the use of the ribosome or an RNA component. cpSRP binds to LHCs to form a soluble “transit complex” thereby enabling the LHCs to traverse the stroma without becoming involved in inappropriate interactions (9). cpSRP43 is required to prevent aggregation of the hydrophobic LHCs during their posttranslational targeting to the thylakoid membrane (10). The aggregation of LHCs was monitored in the presence of each cpSRP pathway component and cpSRP43 was shown to be the exclusive component for preventing the aggregation of LHCs (10). cpSRP43 does not require ATPase activity as is the case for the Hsp104/CipB family of chaperones, however its ability to

dissolve aggregates of LHCPs is comparable to that family of disaggregases (10-13). Studies have also revealed that Hsp70, Hsp60, and trigger factor (TF), a bacterial chaperone involved with hydrophobic regions of proteins, cannot substitute for cpSRP43 in the cpSRP pathway (11, 14). cpSRP43 interacts with the L18 region on LHCP, which is a hydrophilic peptide connecting the second and third trans-membrane domains of LHCP (TM2, TM3) (15, 16). Hydrophobic interactions between the trans-membrane regions of LHCP and cpSRP43 are also considered to contribute largely to the ability of cpSRP43 to successfully chaperone the LHCPs during translocation. The contribution of hydrophobic interactions is based on studies in which binding was maintained under high salt conditions. In addition, 60-fold higher binding affinity between the full length LHCP *versus* the L18 peptide alone was found; revealing the involvement of TM3 of LHCP for efficient binding (11, 17, 18).

cpSRP43 is a multidomain protein containing three chromodomains (CD1-3) and four ankyrin repeats (Ank1-4) located between CD1 and CD2 (9, 15). Chromodomains typically have a flexible linker between them however in cpSRP43 only a short chain connects CD2 and CD3 (Fig.1) (1,2). Chromodomains of other proteins are known for regulating the structure of chromatin due to the ability of these domains to promote protein-protein interactions, although, it is unique that cpSRP43 is the only non-nuclear chromoprotein (1). The chromodomains, located on the c-terminal of cpSRP43 interact with the methionine-rich domain of cpSRP54 and results in formation of heterodimeric cpSRP (15). The crystal structure of cpSRP43 reveals two hydrophobic grooves which are disjointed by a positive ridge on one side and a highly negatively charged surface on the other side (19). Examination into the crystal structure of cpSRP43 suggests that the charge distribution of cpSRP43 is reminiscent of the canonical SRP RNA component which acts as a scaffold for positioning the M-domain of cpSRP54 in the formation

of cpSRP. Ankyrin repeat regions are found in several different classes of proteins with varied functions suggesting that the versatility of ankyrin repeat structural units mediate different types of protein-protein interactions (1). ANK2 and ANK3 of cpSRP43 display the typical fold of an ankyrin region by having a helix-turn-helix motif. However, the flanking ANK regions, 1 and 4, contain elongated helices (3). The helices in ANK4 are extended by 16 residues, a noteworthy divergence from typical ANK structure, proposing an extended role in protein-protein interactions (3). The ANK regions of cpSRP43 bind the L18 domain of LHCP. The association of cpSRP43 with cpSRP54 and LHCP constitutes the soluble “transit complex” necessary for the translocation and integration of LHCP. In addition, cpSRP43 binds to the c-terminal portion of the integral membrane protein Albino3 (Alb3) to initiate and expedite docking at the receptor for integration (20). Reconstruction of the global structure of cpSRP43, based on small-angle X-ray scattering (SAXS) data combined with molecular dynamics (MD) simulations, reveals an elongated shape (~120 Angstroms in length) which supports the proposition that this protein requires extensive surface area for binding substrate and the SRP partner, cpSRP54 (11, 21). cpSRP43 has since displayed significant interdomain dynamics in studies utilizing smFRET, MD simulations and isothermal calorimetry (ITC) (22, 23). Based on these studies, cpSRP43 is speculated to have closed, open and extended conformational states which advocates for a high degree of flexibility across the entire protein (8). Decreased flexibility upon binding to cpSRP54 was detected in conjunction with an increased affinity for the L18 region of the LHCP binding substrate which demonstrates the assertion that cpSRP43 changes in flexibility to execute its role in the cpSRP pathway and is central to coordinating the proper sequence of targeting events of this pathway (8).

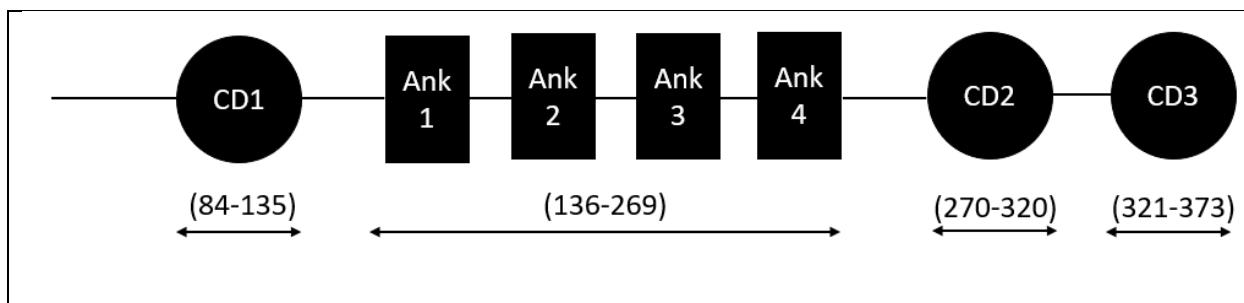
Computational analyses of chloroplast Sec pathway proteins have been implemented, such as the all-atom molecular dynamics of LHC and stromal proteins as well as cpFtsY, LHC proteins, however an inferior amount of molecular dynamics (MD) simulations have been directed towards examining certain conformations of cpSRP43 or the cpSRP43/cpSRP54 complex (24-26). Existing studies have concentrated on all atom simulations of the systems (27). In this study, we set out to reexamine their findings using all atom simulations carried out on the nanosecond level. First, we generated a docked model which attached the CD3 domain to the cpSRP43 structure. We then compared our models to a Small Angle X-Ray Scattering (SAXS) generated model of the cpSRP43 developed by the Kumar Lab (unpublished). SAXS models use data from X-Ray scattering to obtain several different conformations of proteins in solution which are then used to create atomic resolution models (28). These models can further be used to guide molecular dynamics simulations (29, 30). BILBOMD was used to fit the SAXS data into a molecular model (31). The SAXS and the generated docked model were both exposed to 940 nanoseconds of total simulation time to determine which displayed higher structural stability and the most stable model was investigated further with a 2-microsecond equilibrium MD simulation.

cpSRP43 has the capacity to bind multiple partners and the unique architecture and multi-domains which make up this protein comprise an excellent framework upon which protein-protein interactions can occur. The high degree of flexibility and role as a hub protein lead to our investigation into the extent of the structural stability of cpSRP43. In appreciating the information regarding the stability and dynamic nature of cpSRP43, an investigation into discovering an accurate conformation of free cpSRP43 was generated in the present study.

Our findings conclude with a precise depiction of the structure of monomeric cpSRP43, from *Arabidopsis thaliana*, based on the correlation between biophysical experimentation and



molecular modeling. Circular dichroism and fluorescence techniques were utilized to monitor the structural stability upon subjection to heat and chemical denaturants as well as tracking the unfolding and subsequent refolding nature of cpSRP43. Results of ANS binding and limited trypsin digestion of cpSRP43 show a high degree of flexibility in the backbone structure of cpSRP43 in comparison with a commensurate protein. Parameters obtained from intrinsic quenching experiments of cpSRP43 in agreement with *in silico* accessibility plots of tryptophan units confirm the microenvironment and degree of exposure of tryptophan units leading to the most explicit conformational model of cpSRP43. The simulation data revealed a stable conformation for the cpSRP43 monomer with a globular structure, as opposed to the mostly linear crystal structure.



**Fig. 1: Schematic diagram of the structural domains in cpSRP43. The residues spanning each domain are indicated with the double header arrows.**

## Materials and Methods

**Expression and Purification of Recombinant cpSRP43-** For expression of cpSRP43, BL-21 Star cells containing pGEX-6P-2-cpSRP43 were grown to an Optical Density of 0.6–8.0 at Abs<sub>600</sub> and incubated with 1 mM isopropyl β-D-thiogalactoside for 3.5 hours. Cells were harvested and resuspended in lysis buffer. Cells were sonicated for 25 cycles with 10 second on/off pulses. Supernatant was separated from the cell debris using centrifugation at 19,000 rpm for 30 minutes and subsequently passed over glutathione Sepharose that was then washed extensively with equilibration buffer (2.7mM KCl, 1.8mM KH<sub>2</sub>PO<sub>4</sub>, 15mM Na<sub>2</sub>HPO<sub>4</sub>, 137mM

NaCl, pH 7.2). GST-cpSRP43 was eluted using 10 mM L-Glutathione then exchanged into cleavage buffer (50 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA, 1 mM dithiothreitol, pH 7.0). Overnight cleavage in solution is setup at 4°C on a rocker with 10 units of PreScission Protease per liter of original cells for 16 hours. The cleavage product was passed back onto glutathione Sepharose to separate out the GST tag and the cleaved cpSRP43 was further purified by gel filtration chromatography. Purity of proteins were visualized using 15% Sodium Dodecyl Sulfate Poly Acrylamide Gel Electrophoresis (SDS-PAGE) followed by staining with brilliant blue.

**Ovalbumin** (A5503) was purchased from Sigma Co., St. Louis.

**Circular Dichroism-** Circular dichroism (CD) measurements were performed on a Jasco J-1500 CD spectrometer equipped with a variable temperature cell holder. Conformational changes in the secondary structure of cpSRP43 were monitored in the Far-UV region between 190 to 250 nm using a protein concentration of 8.5  $\mu$ M in 0.27mM KCl, 0.18mM KH<sub>2</sub>PO<sub>4</sub>, 1.5mM Na<sub>2</sub>HPO<sub>4</sub>, 13.7mM NaCl ( pH 7.2) in a 1 mm pathlength quartz cuvette. The scan speed, bandwidth and data pitch were set to 50nm/min, 1.00nm and 0.1nm, respectively. Three scans were collected (within a 1000 HT voltage range) and averaged to obtain the CD spectra. The thermal denaturation/renaturation scans were recorded from 25 to 80°C at 5°C increments.

**Fluorescence spectrometry and 8-Anilino-1-naphthalenesulfonic acid (ANS) binding assays-**

Fluorescence spectra were collected using a Fluorescence Spectrophotometer F-2500 (Hitachi). Initial Intrinsic fluorescence measurements on cpSRP43 were performed using an excitation wavelength of 280nm and emission intensity was recorded within a range from 300 nm to 450 nm. ANS binding assay measurements were made using a Fluorescence Spectrophotometer F-2500 (Hitachi) with a slit width set to 2.5nm. The concentration of cpSRP43 was quantified

using the Bradford method (32, 33). Protein concentrations of 8.5  $\mu\text{M}$  in phosphate buffer containing (2.7mM KCl, 1.8mM  $\text{KH}_2\text{PO}_4$ , 15mM  $\text{Na}_2\text{HPO}_4$ , 137mM NaCl, pH 7.2) were placed in a quartz cuvette. Titrations using an ANS stock were made by addition of 10  $\mu\text{M}$  increments of ANS followed with mixing and incubation for 2 minutes preceding each reading at 25°C. Fluorescence intensity was determined with an excitation at 380nm and emission intensity was recorded at 510nm.

**Equilibrium unfolding-** Thermal and chemical unfolding of cpSRP43 as probed by circular dichroism was performed on a Jasco-1500 spectrophotometer using a protein concentration of 8.5  $\mu\text{M}$  in 0.27mM KCl, 0.18mM  $\text{KH}_2\text{PO}_4$ , 1.5mM  $\text{Na}_2\text{HPO}_4$ , 13.7mM NaCl ( pH 7.2). Spectra were collected at 5-degree increments from 25 to 90 °C. Molar ellipticity values were recorded and plotted as a function of temperature. Urea was titrated into the sample up to a concentration of 6 M. Chemical denaturation as probed by fluorescence was carried out in titration mode using stock solutions of ultra-pure grade Urea (8M). Fraction unfolded was calculated from the recorded emission maximums upon chemical denaturant titration. Thermal denaturation as probed by fluorescence was carried by manual heating of samples at 5-degree increments for 5 minutes.

**Limited Trypsin Digestion-** Limited trypsin digestion of cpSRP43 and Ovalbumin was performed in phosphate buffer containing (2.7mM KCl, 1.8mM  $\text{KH}_2\text{PO}_4$ , 15mM  $\text{Na}_2\text{HPO}_4$ , 137mM NaCl, pH 7.2). The initial reaction tube contained 8.5  $\mu\text{M}$  of protein and 0.5  $\mu\text{g}$  of enzyme. The trypsin-containing samples were incubated at room temperature (25 °C). Digested samples were removed every 2 minutes for up to 16 minutes and then the reaction was stopped by the addition of 10% trichloroacetic acid at the end of 20 minutes and the samples were resolved on a 15% sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS– PAGE)

gel and subsequently stained using Coomassie Blue. UN-ScanIT software (Silk Scientific Inc.) was applied to identify the percent of cpSRP43 digested at each time point.

**Fluorescence quenching** -All measurements were made using Fluorescence Spectrophotometer F-2500 (Hitachi). The excitation wavelength was set at 295 nm and data was collected in the range of 315-450 nm. Fluorescence emission intensities were recorded for cpSRP43 at 340 nm upon titration with stock solutions of acrylamide, cesium chloride, potassium iodide and succinimide. Analysis of the quenching data was made using the Stern Volmer equation [1] and the modified Stern-Volmer equation [2] (34).

$$(1) F_0/F = 1 + K_{sv}[Q]$$

$$(2) F_0/\Delta F = f_a^{-1} + 1/(K_d f_a [Q])$$

Where  $F_0$  and  $F$  are the relative fluorescence intensities in the absence and presence of each quencher, respectively and  $[Q]$  is the concentration of each quencher.  $K_{sv}$  is the Stern-Volmer quenching constant for each quencher,  $f_a$  is the fraction of fluorescence accessible to each quencher and  $K_d$  is the corresponding quenching constant.

**MD Equilibrium Simulations of cpSRP43 and cpSRP43/cpSRP54-** To determine the stability of cpSRP43 as compared to the cpSRP43/cpSRP54 complex, VMD (35) and NAMD 2.10-12 (36) with CHARMM (37) force field parameters were used to perform MD simulations based on crystal structures from the RCSB Protein Data Bank (38). The PDB 3UI2 (39), which is the structure of the cpSRP54 tail bound to cpSRP43, was used for MD simulations of both the cpSRP43 and cpSRP43/cpSRP54 in complex. The cpSRP54 sequence was first removed from the PDB file. Then both systems were solvated with a 150 Å X 150 Å X 150 Å water box. cpSRP43 and cpSRP43/cpSRP54 complex systems were minimized for 10,000 steps with harmonic

restraints. A force constant of 1 Kcal/molÅ<sup>2</sup> in the C<sub>α</sub> atoms in both proteins was applied. The systems were heated to 300 K at 1 atm. Both the cpSRP43 and cpSRP54 systems were approximately 96,358 and 96,390 atoms, respectively.

**Molecular Docking of CD3-** VMD tcl scripting was used in order to create an entire cpSRP43 structure from the combination of PBD 3UI2 which included domains CD1 through CD2 and 2N88 (40), which consists of the CD3 domain. The Sequence Alignment tool on VMD was used to find the common amino acid to dock the two PDBs. GLY 316 was used as the anchor point for the docking procedure. The docked model was generated, and the final structure had the amino acid sequence of CD1(85-135) Ank1-Ank4(136-269) CD2 (270-318) CD3 (319-369) cpSRP54(528-540).

**Equilibrium MD Simulations of Docked Model-**The docked model was solvated in a 170 Å X 170 Å X 170 Å water box. The system was neutralized and minimized using the same procedure as described in the first set of simulations. The atoms for both cpSRP43 and the cpSRP43/cpSRP54 complex were approximately 130,800 and 130,945 atoms, respectively. The docked models of both cpSRP43 and the cpSRP43/cpSRP54 complex were run for 100ns of equilibrium simulations as displayed in Figure 1 of the supporting information.

**Determination of target SAXS structure for the SMD models based on Structural RMSD-** After the docked model was equilibrated for 100ns, we used the collective variable module of NAMD program to perform non-equilibrium simulations based on the SAXS structure obtained from the Kumar Lab (unpublished). The Sequence Alignment tool was used to determine the common residues between our SMD model and the SAXS structure. The residues from the Ank1-4, CD2, and CD3 were used in a script to identify similar structure based on RMSD values. 40 frames derived from the SAXS structures were based on variations of 4 different conformations.

The docked model was run through a structural RMSD-script to determine which of the 40 frames were most like the docked model based on structural RMSD. Once the most similar SAXS structure was determined, the centers of the Ank1-4, CD2, and CD3 domains in the SAXS model was determined in VMD using a script.

### **MD Equilibrium and Non-Equilibrium (Colvar) Simulations of SAXS Models and SMD**

**Generated Models-**Once the SMD models 1-4 were generated, they were run on 100 ns of equilibrium simulations. These models were not minimized. SAXS Models 1-4 were solvated in a 170 Å X 170 Å X 170 Å water box. The target SAXS models were minimized and neutralized using the same procedure outlined in the first set of simulations. The SAXS structures were run for 100 ns of equilibrium simulations.

### **Microsecond-level All-Atom Equilibrium MD Simulation of SMD Model 4-**

The equilibrium MD production run for the 4<sup>th</sup> SMD model was extended for 2 microseconds on the Anton2 supercomputer (Pittsburgh Supercomputing Center, D.E. Shaw Research), with a timestep of 2.5 fs. Conformations were collected every 240 picoseconds. Initial processing of the trajectories was carried out using VMD on the Kollman cluster (PSC, DEShaw).

## **Results and Discussion**

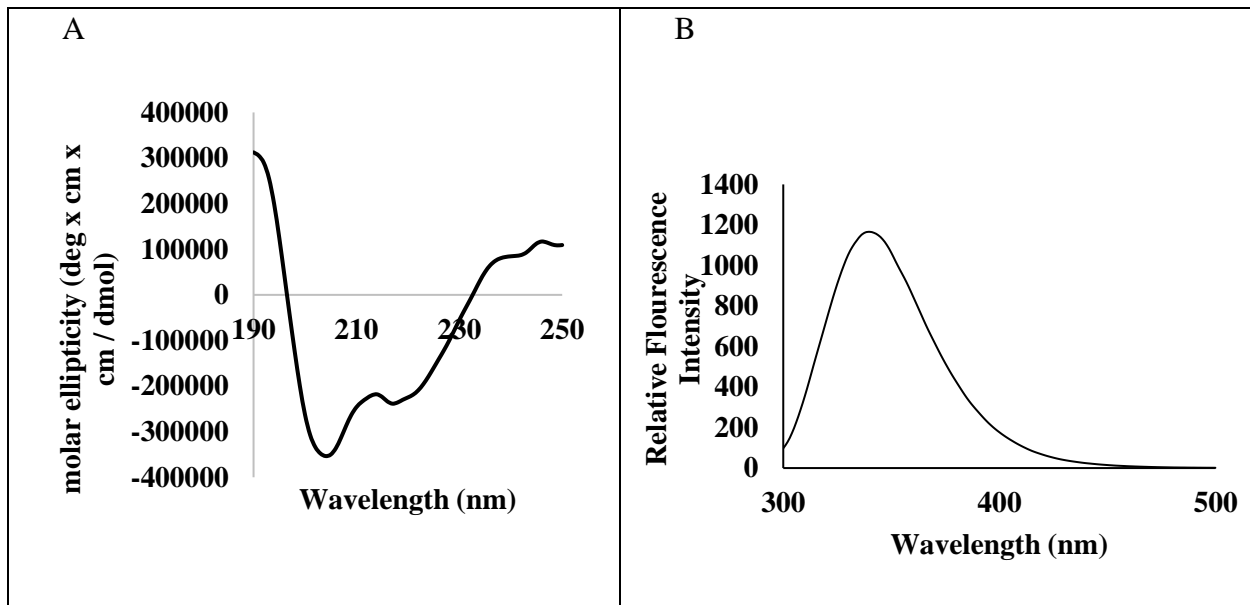
### ***Monitoring of secondary structure by Far-UV-CD Spectroscopy***

Far UV-circular dichroism [190-250nm] is a useful tool to probe the secondary structure of proteins in native conformation(s) and to evaluate any changes in conformation upon alterations made to the microenvironment of the protein. The far UV CD spectrum of cpSRP43 shows a double negative minimum at 208nm and 222nm suggesting that the backbone of the protein is predominately folded into a helix conformation. This aspect is consistent with the 3D structure obtained of cpSRP43 (Fig.2A),(19, 39-42). Circular dichroism spectra of the

chromodomain regions of cpSRP43 show that CD1 is characteristic of beta-barrel secondary structure and NMR studies reveal that CD1 does not contain any helical segments (41). Circular dichroism spectra of CD2 and CD3 reflect that there are areas of the backbone of these domains which have helical conformations. Through resonance assignments, CD2 and CD3 are each found to contain a triple-stranded-beta-sheet and an alpha helix (17). The ANK repeat segments within cpSRP43 contribute greatly to the helical signal of the CD spectra, and CD1 and CD2 contribute partially which leads to the overall predominate helical signal of the CD spectra of full length native cpSRP43 shown in this study.

### ***Monitoring of tertiary structural change by fluorescence***

Intrinsic protein fluorescence is sensitive to the local environment surrounding the indole ring of tryptophan units and is useful in determining the stability of a protein by tracking the folding/unfolding rates under different conditions. Intrinsic fluorescence spectrum of cpSRP43 in its native conformation shows an emission maximum at 340nm (Fig.2B). cpSRP43 contains 8 tryptophan residues, four are located within CD1, two are in CD2 and the remaining two are found in CD3. The wavelength of maximum emission indicates that the tryptophan units, on an average, are partially buried in the protein.

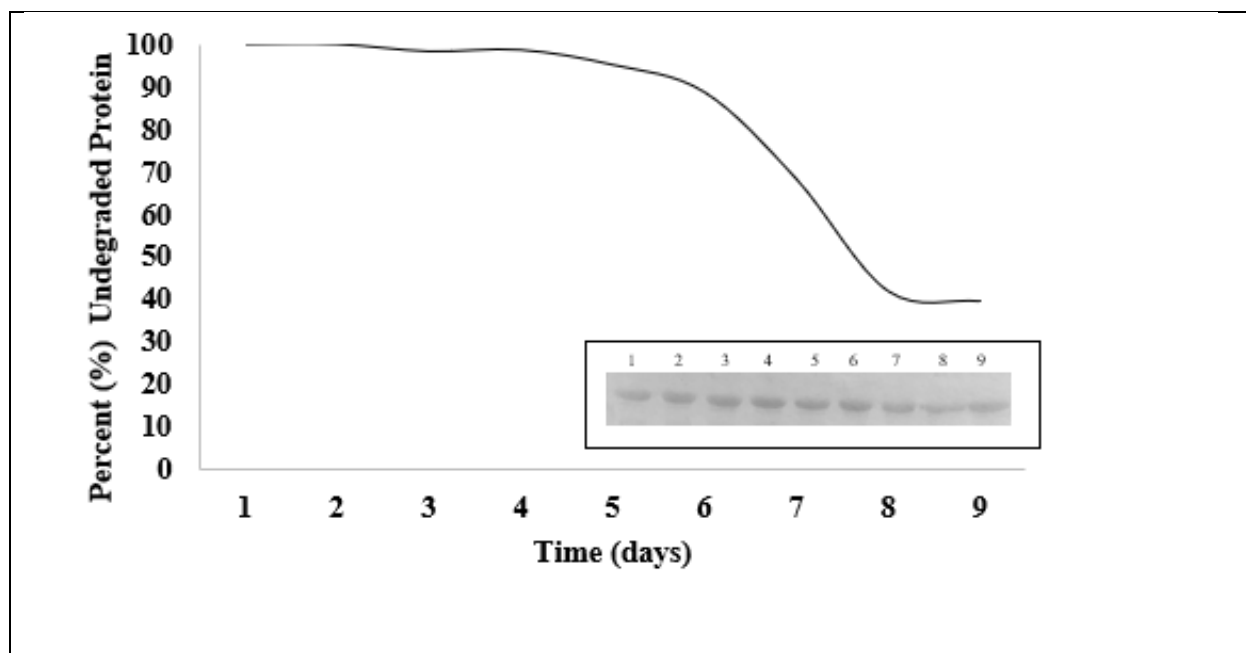


**Fig. 2: Far-UV CD (Panel-A) and Intrinsic Fluorescence (Panel-B) spectra of cpSRP43.**

### *Protein stability and degradation at room temperature*

Storage of protein left at room temperature often leads to degradation. In many cases, protein cannot withstand being left at room temperature for more than a day whereas some proteins may degrade directly. Protein stability is governed by a balance of different forces. The rate of degradation of cpSRP43, when left at 25 degrees C, was observed over a nine-day period (Fig. 3). cpSRP43 showed resistance to degradation up to four days. Significant degradation was not observed until seven days at room temperature. The average pixels recorded for the SDS-PAGE gel, days 1-9 samples using the UN-SCAN It densitometric software are as follows: 8.9,8.9,8.75,8.78,8.47,7.9,6.09,3.75 and 3.54, respectively.



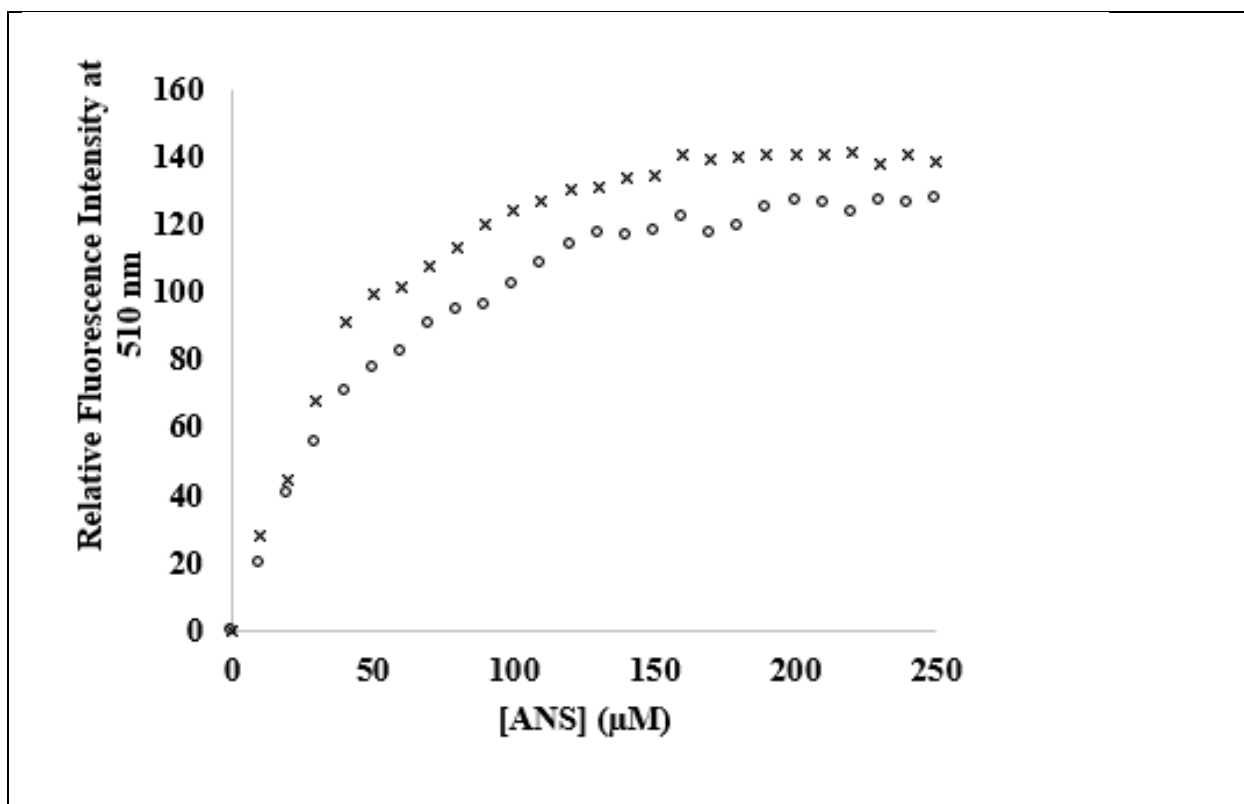


**Fig. 3: Densitometric plot depicting the rate of degradation of cpSRP43, incubated at 25°C, over a nine-day period. The insert shows the 15% SDS-PAGE with lanes 1-9 representing days 1-9.**

#### *Hydrophobic surface accessibility of cpSRP43*

Anilino naphthalene 8-sulfonate (ANS) is a non-polar dye designated to evaluate the extent of solvent-exposed hydrophobic surfaces in proteins. ANS fluoresces upon binding to hydrophobic regions within a protein. Hydrophobic properties of proteins can be crucial as a driving force toward proper folding as well as protein-protein interactions. The ANS binding curve of cpSRP43 is compared to that of Ovalbumin (Fig.4). cpSRP43 and Ovalbumin are comparable in size and overall charge. Ovalbumin is an acidic, globular protein with a soluble serpin structure. It consists of 385 amino acids and has a molecular weight of 42.7kDa, making it slightly larger than cpSRP43 which consists of 319 amino acids and has a molecular weight of 35.2kDa. Upon heating, the serpin structure turns into a beta-sheet structure and, with all hydrophobic surfaces exposed, is easily aggregated (43, 44). ANS bound to cpSRP43 shows an emission maximum at 510 nm and the surface hydrophobic regions are saturated with ANS at a concentration of 150  $\mu$ M ANS. Comparison of the ANS saturation curves of both reveals

insignificant changes in the relative fluorescence intensities which suggests that the two proteins contain a similar extent of solvent-exposed hydrophobic surface(s) in native conformations.

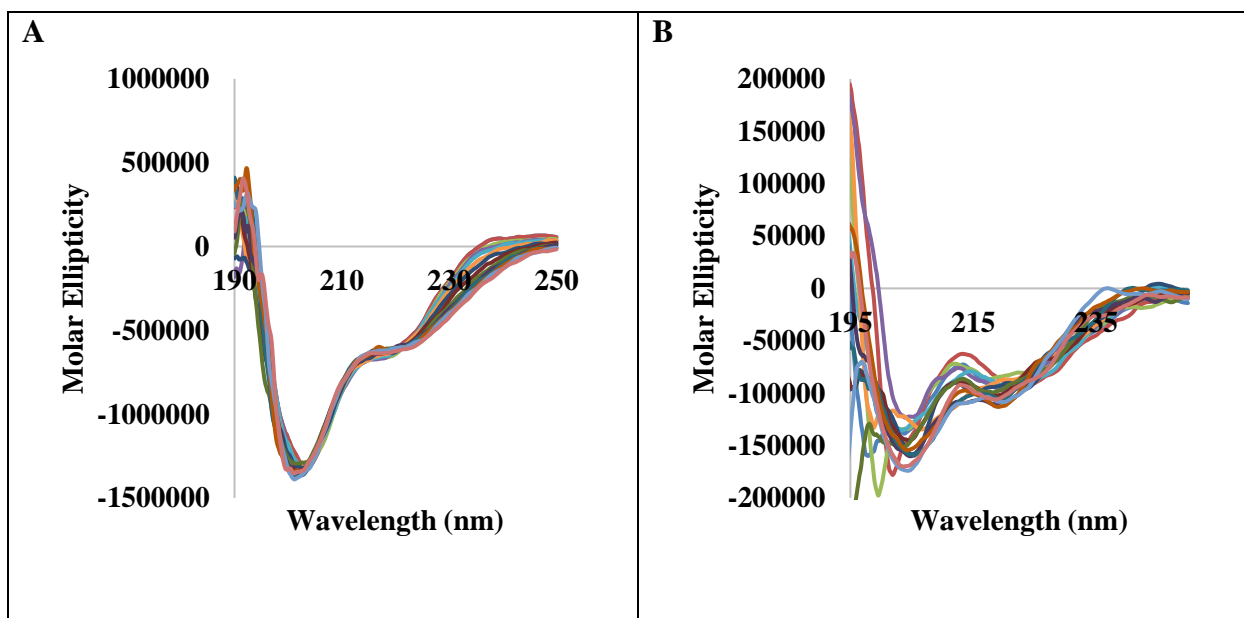


**Fig. 4:** ANS binding curves of cpSRP43 (o) and Ovalbumin (x).

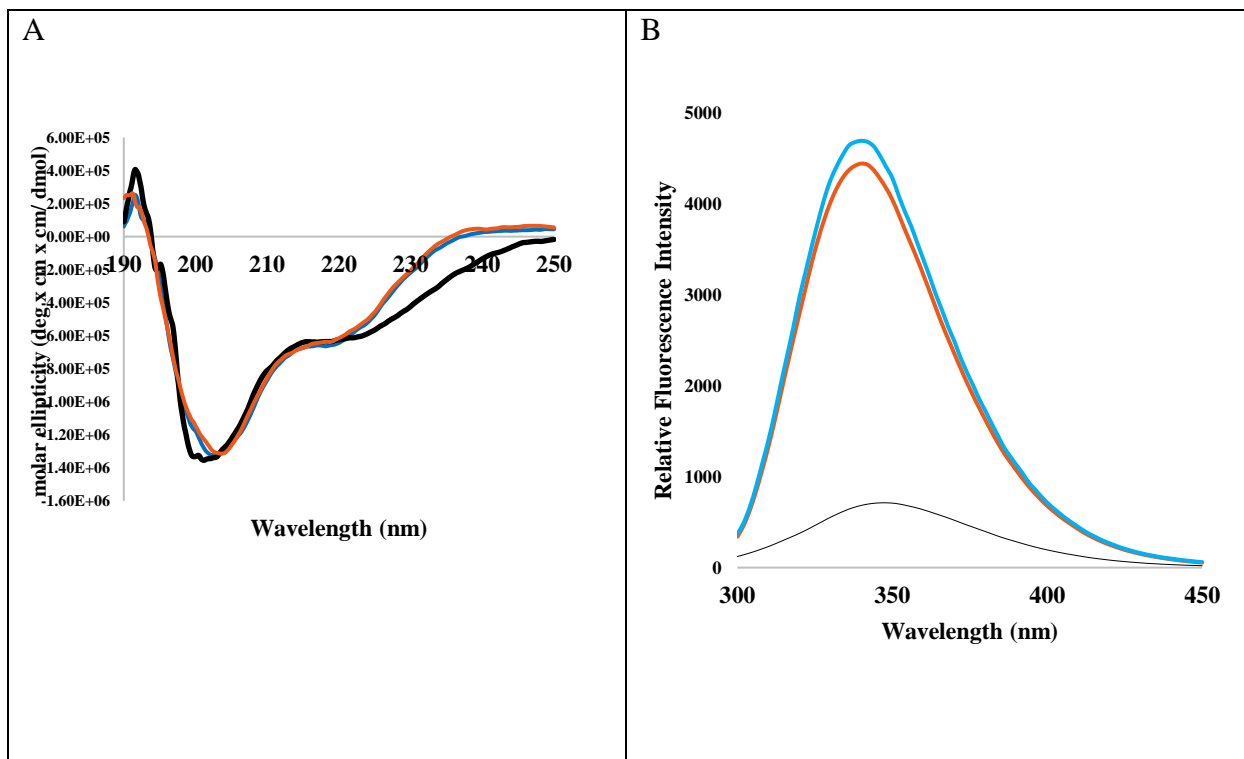
### *Thermal Stability of cpSRP43*

Thermal denaturation and renaturation of cpSRP43 was probed by circular dichroism from 25-90°C at 5°C increments. cpSRP43 retained its secondary structure upon heating and cooling cycles which proves to be an interesting feature for such a structurally dynamic protein. The individual circular dichroism spectra taken at each temperature increment show negligible change in comparison to one another upon heating and cooling cycles (Fig.5). The spectra of cpSRP43 before heating (25°C), upon heating to 90°C and after cooling back down to 25°C appear nearly identical when superimposed (Fig.6A). The conformity between spectra is also illustrated by the ratio of the double negative minima at each temperature increment (Fig.7A).

Upon heating, cpSRP43 shows a steady decrease in intrinsic fluorescence intensity (Fig.6B). Intrinsic fluorescence readings taken of cpSRP43 before heating show that the protein is in the native conformation with an emission maximum at 340 nm. The samples were then heated to 90°C at which point the emission maxima reached 348 nm and the relative intensities diminished, indicating that the protein was in an unfolded state. The fluorescence readings were taken again after the protein cooled down to room temperature and the emission maxima returned to 340nm suggesting that, upon cooling, the protein was able to regain its native fold (Fig.6B). Thermal stability data reveals that cpSRP43 is highly stable at extreme temperatures. The structure of cpSRP43 contributed by the backbone arrangement is not perturbed under high heat conditions and although the tertiary structure may undergo some changes upon heat exposure, any shift in the folded state of cpSRP43 is reversible upon cooling.



**Fig. 5: Thermal denaturation (Panel-A) and renaturation (Panel-B) of cpSRP43 as probed by circular dichroism at five-degree increments from 25 -90 degrees C.**

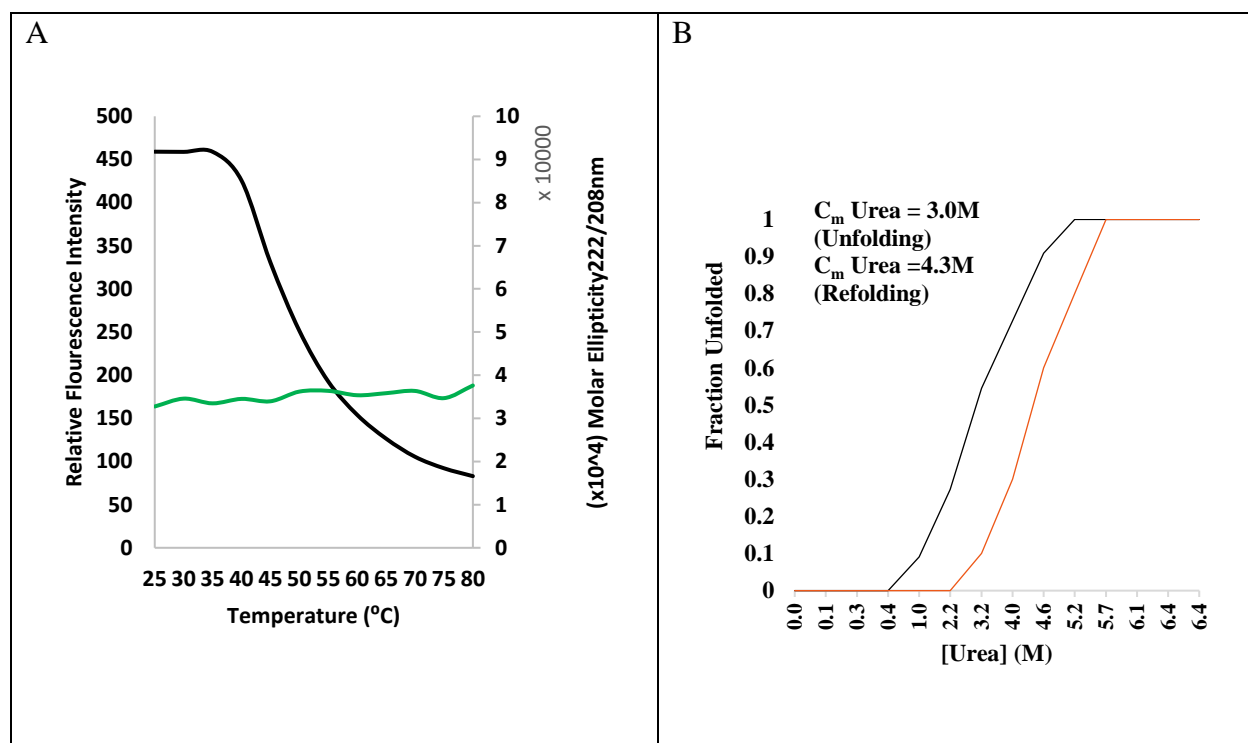


**Fig. 6: Thermal stability of cpSRP43 monitored by far UV-CD (Panel-A) and intrinsic fluorescence (Panel-B) spectroscopy. Native cpSRP43 at 25°C (blue), cpSRP43 heated to 90°C (black) and cpSRP43 upon cooling back down to 25°C (red).**

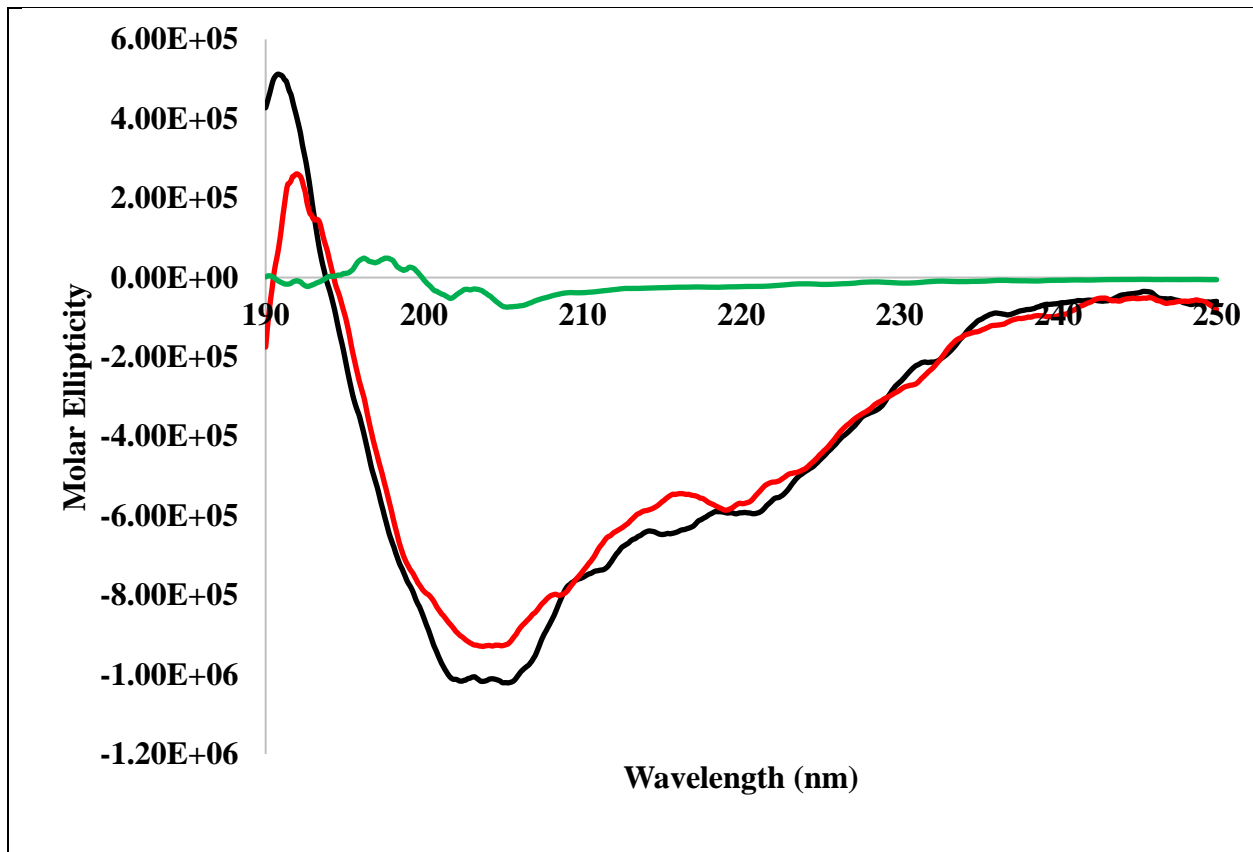
### *Stability of cpSRP43 upon chemical denaturation*

The stability of cpSRP43 was monitored by chemical-induced equilibration unfolding upon titration with urea (Fig. 7B). At zero concentration of denaturant, cpSRP43 is in the native conformation. At the saturation concentration for urea, being 5M, all cpSRP43 molecules exist in the denatured state. The  $C_m$  values which represent the concentration of denaturant at which 50% of the protein is in the denatured state(s), were calculated from the chemical unfolding curve. The  $C_m$  values for cpSRP43 is 3.1M in urea. Urea, being a neutral denaturant, can form hydrogen bonds with cpSRP43 thereby weakening intermolecular of the protein. To test if chemical-induced denaturation of cpSRP43 was reversible, dilution of cpSRP43 out of urea was carried out and monitored by circular dichroism. Initial circular dichroism spectra were recorded

for native cpSRP43. Next, the protein was denatured at a concentration of 6M Urea and again the spectra were recorded showing the protein to be in a denatured state. Circular dichroism spectrum was taken during dilution and the final spectra at the end of dilution revealed that cpSRP43 was able to refold (Fig.8).



**Fig.7: (Panel-A)-Relative fluorescence intensity spectra of cpSRP43 probed at 340nm (black) and the far-UV CD spectra of the 222 to 208nm ratio (green) upon heating. (Panel-B)-Urea-induced equilibrium unfolding (black) and refolding (red) curves for cpSRP43 were monitored by changes in the ratio of the intrinsic fluorescence intensity at 340 nm/350 nm.**

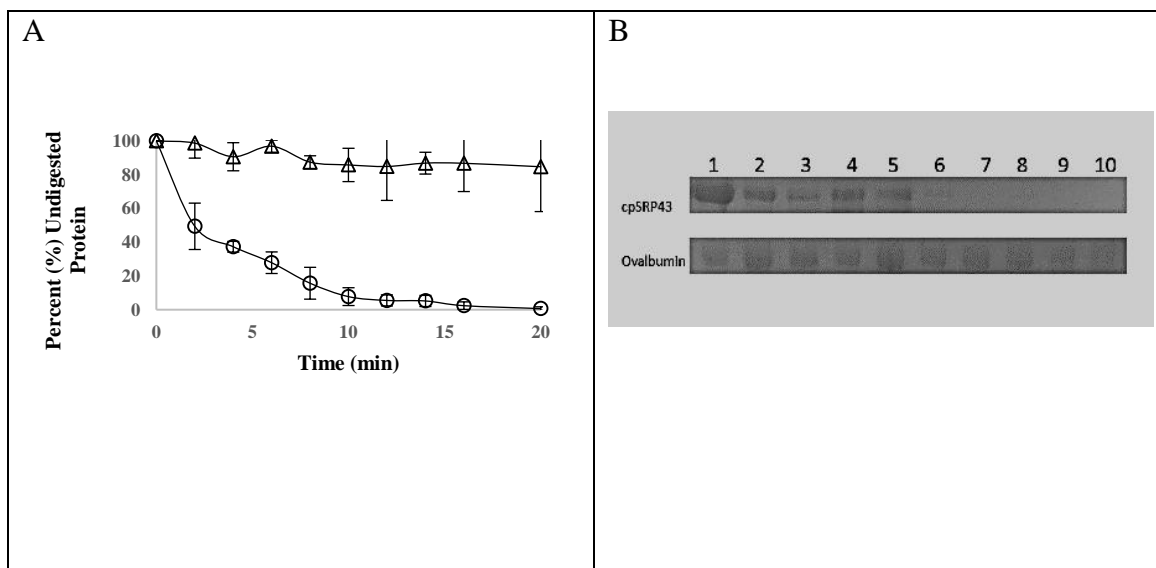


**Fig 8: Circular dichroism spectra of cpSRP43 before denaturation by urea (black), upon denaturation by urea (green) and dilution out of urea (red).**

#### *Susceptibility of cpSRP43 to trypsin digestion*

Limited trypsin digestion was used to evaluate the conformational flexibility of cpSRP43. Time-dependent trypsin digestion was carried out at 25°C and was observed on SDS-PAGE followed by analysis of the band intensities measured by densitometry (Fig. 9). cpSRP43 was highly susceptible to proteolytic cleavage which is reflected by the need to carry out the experiment at 25°C in the presence of a small quantity of trypsin. In initial trials in which cpSRP43 was subjected to digestion at varying concentrations of trypsin at 37°C, the degradation occurred almost immediately which lead to optimization of the experiment parameters. For comparison, limited trypsin digestion of Ovalbumin was performed under the same conditions. Trypsin cleaves at the carboxyl end of the amino acids lysine and arginine unless either are

followed by proline. cpSRP43 was easily digested within twenty minutes time whereas Ovalbumin was effectively unchanged. cpSRP43 contains 16 arginine residues and 16 Lysine residues of which three arginine residues are followed by proline. cpSRP43 is therefore having 29 potential cleavage sites which are contributing to the high susceptibility to trypsin digestion. The results here indicate that the backbone of cpSRP43 is more flexible than that of Ovalbumin which exhibited a lower susceptibility to trypsin digestion although it contains 33 cleavage sites for trypsin. The unusually high flexibility and dynamics of cpSRP43 previously reported in the acquisition of SAXS, smFRET and MD simulations data is consistent with conclusions drawn from the trypsin digestion data (22, 23, 45).



**Fig. 9: Densitometric analysis of the time dependent limited trypsin digestion of cpSRP43 (o) and Ovalbumin (Δ) as monitored by SDS-PAGE (Panel-A). Coomassie blue stained bands of cpSRP43 and Ovalbumin (Panel-B). Lane 1= No Trypsin, Lane 2= 2 minutes, Lane 3= 4 minutes, Lane 4= 6 minutes, Lane 5= 8 minutes, Lane 6= 10 minutes, Lane 7= 12 minutes, Lane 8= 14 minutes, Lane 9= 16 minutes, Lane 10= 20 minutes**

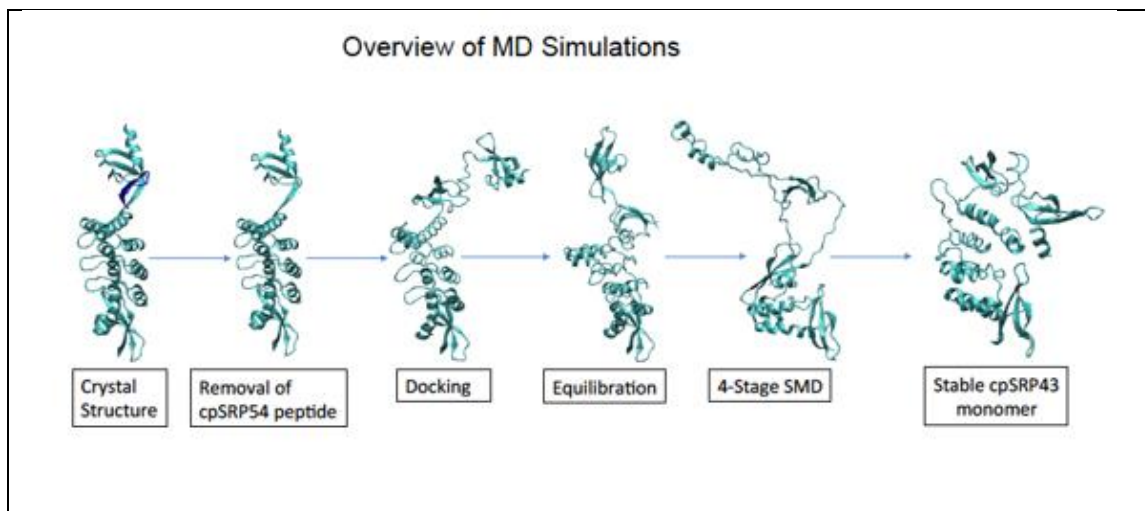
cpSRP43 is highly stable and undergoes reversible unfolding. In addition, the tryptophan residues in cpSRP43 are partially or superficially buried. These properties are characteristic of heat shock chaperone proteins. The coexisting high extent of stability and flexibility of cpSRP43

is not explained by the current linear structure derived from X-ray crystallography data. Molecular dynamics studies were then sought to revisit the structure of cpSRP43 in order to derive a model which can substantiate the occurrence of high stability and flexibility.

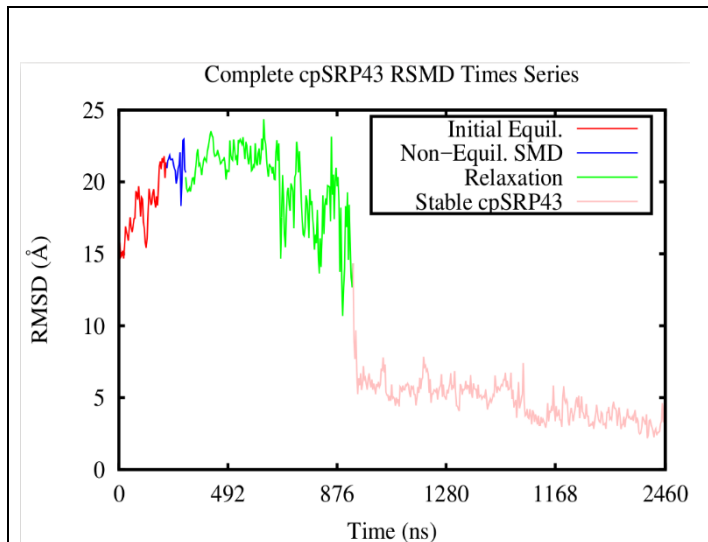
### ***Overview of MD Simulations***

A conformation of monomeric cpSRP43 was isolated using microsecond-level equilibrium MD simulations and found to be stable. The PDB crystal structure, 3UI2, was used after removal of the cpSRP54 tail and subsequent docking of the CD3 domain from PDB 2N88. Upon docking of the CD3 domain, equilibrium MD simulations were executed for 100 nanoseconds. SMD was achieved in four stages based on different SAXS structures. Four SMD models were developed and the 4<sup>th</sup> model was selected for its stability and subjected to a 2-microsecond equilibrium-MD simulation. An overview of the MD simulations process is illustrated below (Fig.10). An SMD model was produced from a conformation of monomeric cpSRP43 which was stable for approximately 1.6 microseconds according to RMSD values (Fig 11).





**Fig.10: Overview of molecular dynamics simulations process.**



**Fig. 11: The RMSD plot shows that the cpSRP43 monomer adopts a stable conformation at 0.4 nanoseconds and remains stable for 1.6 nanoseconds.**

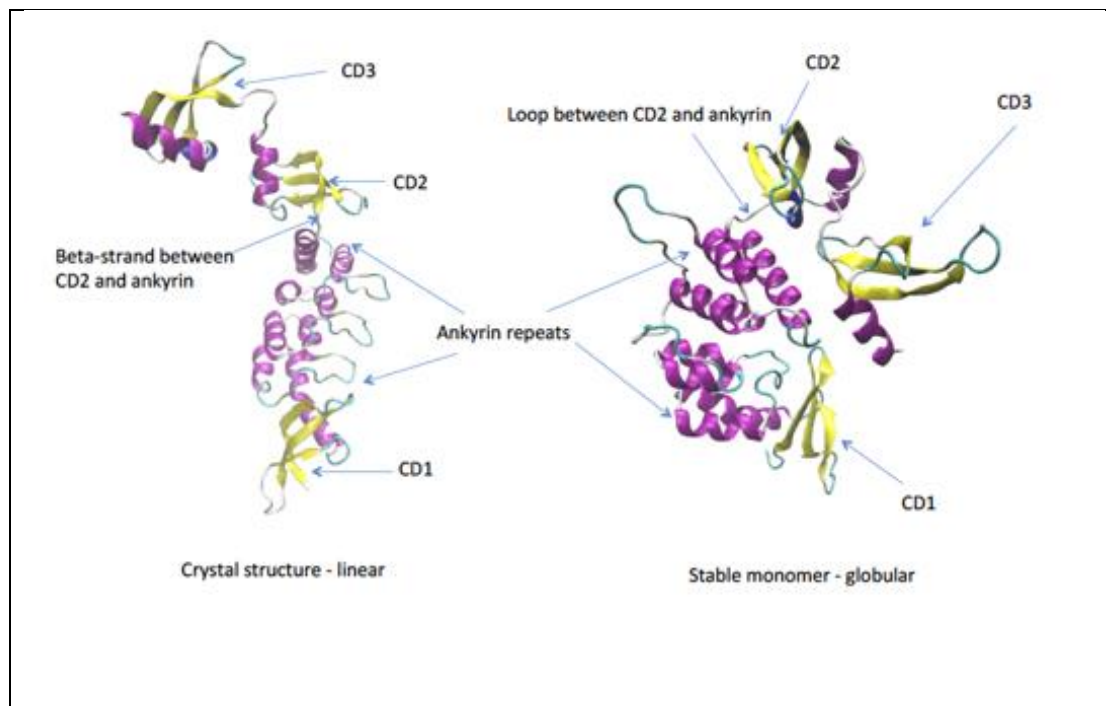
***Comparative analysis of the stability of an SMD generated cpSRP43 model versus a similar SAXS cpSRP43 model.***

SMD generated models are resultant from first creating an atomistic model based on a SAXS model. Using the CD3 domain docked cpSRP43 model, the system was guided towards the selected SAXS model. Equilibrium simulations were performed based on SAXS models and the corresponding SMD models. This allowed for the determination of stability as well as the presence of salt bridges. Overall stability and the newly found salt bridges were representative of the structural dynamics of both SAXS and SMD models. Eight equilibrium simulations were analyzed. This included four SAXS models which complemented four SMD models. The SAXS models showed a high degree of structural fluctuation based on RMSD and RMSF values. The 4<sup>th</sup> SMD model revealed a stable conformation upon analysis utilizing reverse RMSD. Reverse RMSD is determined by determining the RMSD values based on the final equilibrated conformation of the MD simulation.

***Comparative analysis of a stable conformation generated from the 4<sup>th</sup> SMD model and the crystal structure.***

The 4<sup>th</sup> SMD initially showed a stable conformation and therefore the equilibrium production run was subsequently continued for 2 microseconds. This conformation showed several fluctuations during the first 0.4 microseconds. The system then resolved itself to remain stable for 1.6 microseconds. Chromodomains 2 and 3 appeared to be the most flexible domains of this stable conformation (Fig.12). This result agrees with the data from the docking process of CD3, which indicated that the docking of CD3 to the crystal structure lead to a decrease in overall system stability. Trajectories of the atoms, upon visual inspection, revealed that ankyrin repeats 2 and 3 become rotated by 180 degrees in the final conformation, when compared to the crystal structure

(Fig. 12). According to the RMSD values, the generated monomeric form of cpSRP43 was more stable than the 3UI2 crystal structure.



**Fig.12: A comparison of the structural changes occurring in the stable monomer as compared to the crystal structure. A 180-degree rotation of the middle ankyrin repeats coupled with the beta strand to loop transformation (between CD2 and ANK1) causes different domains to interact. This results in the stable monomer adopting a globular structure.**

***Solvent accessible surface area of tryptophan residues is increased in the stable monomeric cpSRP43***

To order to assess the structural change between the model and the crystal structure, we calculated the combined solvent accessible surface area (SASA) of individual tryptophan residues as well as for all tryptophan residues (Table.1). Five out of eight tryptophan residues (106,132,133,291 and 299) in the stable monomer show an increase in average SASA. A significant increase in the average SASA for all tryptophan residues was observed in the stable conformation with an average SASA of all tryptophan residues at  $565 \text{ \AA}^2$  as compared to  $445 \text{ \AA}^2$

for the crystal structure. The net increase in SASA for tryptophan residues in the stable model supports the flexibility seen in the limited trypsin digestion. Fluorescence quenching studies were subsequently carried out to reinforce the validity of the selected SMD model.

**Table 1: Solvent accessible surface area (SASA) of individual tryptophan residues in cpSRP43 in the crystal structure and the stable monomer.**

Tryptophan Residue Number	Average SASA for the Crystal Structure ( $\text{\AA}^2$ )	Average SASA for the Stable Monomer ( $\text{\AA}^2$ )
Trp 106	62	95
Trp 114	112	107
Trp 132	2	27
Trp 133	13	55
Trp 291	16	47
Trp 299	83	88
Trp 343	52	43
Trp 351	104	101

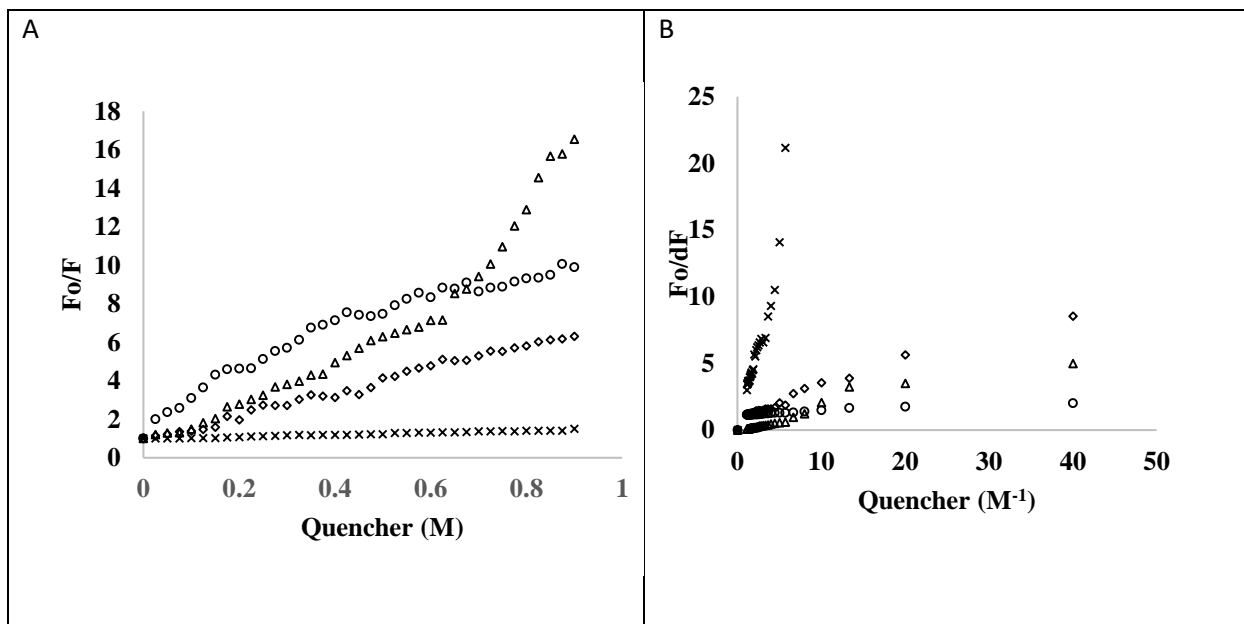
### *Fluorescence quenching of cpSRP43*

The aromatic side chains of tryptophan, tyrosine and phenylalanine may be exploited by intrinsic fluorescence methods to reveal protein structure and conformational state(s) (46). Tryptophan residues are selectively excited at 295 nm. Intrinsic fluorescence properties of tryptophan units are sensitive to changes in their microenvironment. The accessibility of tryptophan units within a protein can be evaluated using fluorescence quenching molecules (46-49). In this study, acrylamide, succinimide, potassium iodide and cesium chloride, were each titrated into cpSRP43. Acrylamide and succinimide are both neutral quenchers with the ability to quench all accessible tryptophan residues, however succinimide has a bulkier structure and is known to be a less efficient quencher (46). These quenchers provide information concerning the extent to which the tryptophan residues are buried within the protein matrix (46). An anionic quencher ( $I^-$ ) and a cationic quencher ( $Cs^+$ ) were used to provide details concerning the microenvironment of the tryptophan residues. Quenching accessibility of ionic quenchers depend on the charges proximally close to the tryptophan residues (46). Fluorescence quenching parameters were obtained from Stern-Volmer and modified Stern-Volmer plots (Fig. 14). Stern-Volmer plots showed a linear upward curvature except for the cesium chloride plot which was a flat line in comparison to the other quenchers. The  $K_{sv}$  values of acrylamide, potassium iodide and succinimide are significantly larger than that of cesium chloride which recommends that they are more efficient quenchers for cpSRP43 (Table 2). Acrylamide and succinimide displayed high fluorophore accessibility suggesting that the tryptophan residues are largely exposed to solvent or superficially buried throughout the protein. The accessibility of tryptophan units was higher for potassium iodide over cesium chloride which theorizes that the local charged environment of tryptophan residues within cpSRP43 are predominately positively

charged around tryptophan units which are partially buried. Given the overall high accessibility of these tryptophan residues to a range of different fluorophores and the calculated combined solvent accessible surface area (SASA) of all tryptophan residues as well as the SASA for individual tryptophan residues (Table 1), these collective results are in agreement with the high degree of flexibility shown throughout this study as well as in previous studies. Acrylamide and succinimide showed high fluorophore accessibility which agrees well with the SASA data indicating that the tryptophan residues are more exposed to solvent in the stable monomer. The cationic quencher had less accessibility to the tryptophan units over the anionic quencher which reveals that some of the tryptophan residues are partially buried in a charged microenvironment. Overall, fluorescence studies showed that the tryptophan residues are partially, albeit mostly superficially buried in a charged microenvironment; a result consistent with the fluorescence emission maximum at 340 nm and the SASA values derived from the stable monomer.

**Table 2: Fluorescence quenching on cpSRP43 using different quenchers.  $K_{sv}$ : Stern-Volmer quenching constant,  $K_d$ : quenching constant,  $f_a$ : fraction of the total fluorescence accessible to a given quencher.**

Quencher	$K_{sv} (M^{-1})$	$K_d (M^{-1})$	$f_a$
Acrylamide	10.24	0.82	0.91
Succinimide	6.0	0.73	0.89
Potassium Iodide	13.23	0.42	0.90
Cesium Chloride	0.50	0.99	0.69



**Fig. 14: Stern-Volmer plots (Panel-A) and Modified Stern-Volmer plots (Panel-B) of fluorescence quenching on cpSRP43. Acrylamide ( $\Delta$ ), Cesium Chloride (x), Succinimide ( $\diamond$ ) and Potassium Iodide (o).**

***Inter-domain distances and the presence of novel salt bridges support a globular structure for cpSRP43***

In the crystal structure of cpSRP43 there is a beta strand linking CD2 to ANK 4, however this is replaced by a collapsible loop in the stable monomer. This structural modulation enables the CD1 domain to better interact with ANK4 and CD3. The loop allows the different domains within cpSRP43 to come in closer proximity to each other as compared to the crystal structure. The average distances between the domain pairs CD1-CD2, CD1-ANK and CD2-ANK were assessed for the crystal structure and the stable monomer (Table 3). Our data indicates that the average distance between CD1-CD2 and CD2-Ankyrin repeat region is significantly lowered suggesting that these domain pairs can make closer associations with each other and this result also supports the high flexibility of cpSRP43 shown previously. Given the closer distances

between domain pairs we propose that cpSRP43 has adopted a more globular structure in the stable model as opposed to the mostly linear structure derived from crystallography.

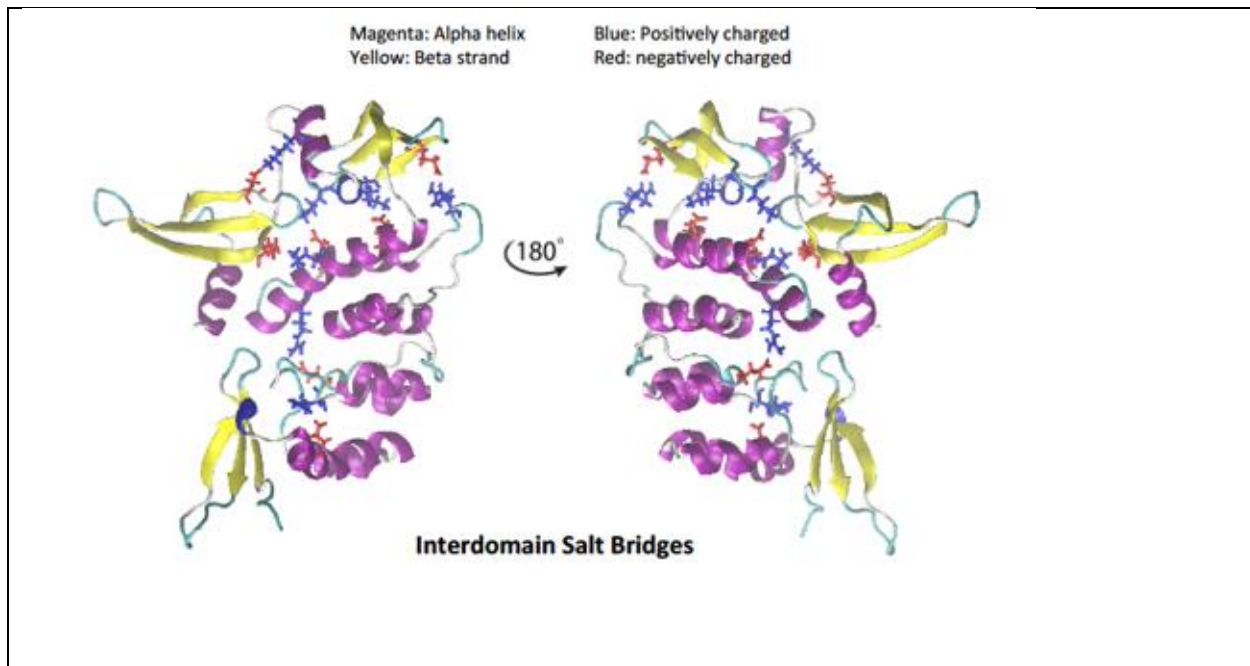
**Table 2: A comparison of the average distance between domains for the crystal structure and the stable monomer. There is a significant decrease in inter-domain distance for CD1-CD2 and CD2-ANK in the stable monomer and a much smaller one for CD1-ANK.**

Domain pairs	Average distance in crystal structure (Å <sup>2</sup> )	Average distance in stable monomer (Å <sup>2</sup> )
CD1-CD2	62.74	49.54
CD1-ANK	29.46	27.74
CD2-ANK	35.99	29.80

***Novel salt bridges found in the model give stability to the globular structure of cpSRP43***

Seven inter-domain salt bridges were found in the stable monomeric cpSRP43 model which were not present in the crystal structure. 1) GLU298-ARG226 (Ank1-4 and CD2); 2) GLU263-LYS301 (Ank1-4 and CD2); 3) GLU323-LYS257 (Ank1-4 and CD3); 4) ASP348-LYS311 (CD2 and CD3); 5) GLU129-ARG161 (CD1 and Ank1-4); 6) GLU256-LYS283 (Ank1-4 and CD2); 7) GLU181-ARG251 (CD1 and Ank1-4) (Fig.15). In order to compare the salt bridges, present in the stable model and the crystal structure, a program was used which isolated the salt bridges based on their length and longevity, under 4 Å° and greater than or equal to 40% of the simulation time, respectively. Seven unique inter-domain salt bridges were found in the stable monomeric cpSRP43. There were no inter-domain salt bridges present in the crystal structure.





**Fig. 15. Seven unique inter-domain salt-bridges occur in the stable monomer which are not present in the crystal structure.**

Two were formed between CD1 and Ank1-4, three were formed between CD2 and Ank1-4, one formed between CD2 and CD3 and one formed between CD3 and Ank1-4. The salt bridges which were significant according to the percentage that they were present during the simulation are as follows: GLU129-ARG161 (78% - CD1 to Ank1-4), GLU 256-LYS283 (62%- CD2 to Ank1-4) and GLU298-ARG226 (54%- CD2 to Ank1-4). Overall twenty-one salt bridges were found in the stable cpSRP43 monomer. Seven of these were novel inter-domain salt bridges, nine were found to be novel intra-domain salt bridges and five were common with the crystal structure. In the crystal structure there were a total of fifteen salt-bridges with ten of those being new. The presence of novel salt bridges in the stable model is likely formed due to the closer proximity between domain pairs and gives stability to the globular structure of cpSRP43.

## Summary and Conclusions

We have investigated the structure and stability of cpSRP43 under different conditions by utilizing various biophysical and *in silico* techniques. cpSRP43 is found to be a highly heat stable protein capable of maintaining its secondary structure upon subjection to significant increases in temperature. In addition, cpSRP43 was able to resist degradation when left for several days at room temperature. This protein also can refold itself upon cooling cycles and dilution out of a chemical denaturant. The degree of thermal stability displayed by cpSRP43 would be expected of a highly compact or rigid protein however the data derived from ANS binding and limited trypsin digestion are indicative of a highly flexible and dynamic structure. The extraordinary stability of cpSRP43 is unusual considering the parallel degree of high flexibility and structural dynamics within this protein. These aspects were previously unexplained by the existing crystal structure(s) of cpSRP43. Using stability tracking data analysis such as thermal and chemical equilibrium unfolding and refolding in alignment with suppositions made from SMD data, we have generated a structural model of monomeric cpSRP43. The SMD data clearly shows that the 180-degree rotation of the ankyrin repeats in combination with other secondary structural changes accounts for the presence of unique inter-domain salt-bridges in the stable cpSRP43 monomer. This change in conformation allows various domains to interact with each other thereby forming stable salt-bridges which, in combination reinforced the overall effects of the conformational change supporting a stable globular structure. This combination of secondary structural changes and interactions results in the globular structure of the stable monomer in contrast to the linear crystal structure. Solvent accessible surface area distances predicted for the 8 tryptophan residues of the stable model and Intrinsic fluorescence and fluorescence quenching data for cpSRP43 agree that the Tryptophan

residues are partially buried or superficially exposed in a charged microenvironment which all support the highly dynamic and flexible nature of cpSRP43. Experimental and computational data presented here provides a stable monomeric form of cpSRP43 which has a globular structure which is a structure that supports the significant degree of stability held within cpSRP43. There is likely a stable intermediate between bound and unbound states of cpSRP43 to cpSRP54. Equilibrium shifts to a more compact conformation to accommodate binding partners. The stability and new structural insights revealed information which may prove to be biologically relevant with respect to the function of cpSRP43 as a chaperone.

**Author Contributions:** Modeling and molecular dynamics simulations were performed and analyzed by Mitchell Benton and Dr. Mahmoud Moradi.

**Acknowledgements:** This work was supported by the DOE grant DE-02-01ER15161.

## References

1. Gutensohn M, Fan E, Frielingsdorf S, Hanner P, Hou B, Hust B, et al. Toc, Tic, Tat et al.: structure and function of protein transport machineries in chloroplasts. *Journal of Plant Physiology*. 2006;163(3):333-47.
2. Li X, Henry R, Yuan J, Cline K, Hoffman NE. A chloroplast homologue of the signal recognition particle subunit SRP54 is involved in the posttranslational integration of a protein into thylakoid membranes. *Proceedings of the National Academy of Sciences*. 1995;92(9):3789-93.
3. Büchel C. Evolution and function of light harvesting proteins. *Journal of Plant Physiology*. 2015;172:62-75.
4. Pietrzykowska M, Suorsa M, Semchonok DA, Tikkanen M, Boekema EJ, Aro E-M, et al. The Light-Harvesting Chlorophyll a/b Binding Proteins Lhcb1 and Lhcb2 Play Complementary Roles during State Transitions in *Arabidopsis*. *The Plant Cell Online*. 2014.
5. Kong F, Zhou Y, Sun P, Cao M, Li H, Mao Y. Identification of light-harvesting chlorophyll a/b-binding protein genes of *Zostera marina* L. and their expression under different environmental conditions. *Journal of Ocean University of China*. 2016;15(1):152-62.
6. Henry RL. SRP: adapting to life in the chloroplast. *Nature Structural and Molecular Biology*. 2010;17(6):676.
7. Xu Y-H, Liu R, Yan L, Liu Z-Q, Jiang S-C, Shen Y-Y, et al. Light-harvesting chlorophyll a/b-binding proteins are required for stomatal response to abscisic acid in *Arabidopsis*. *Journal of experimental botany*. 2011;63(3):1095-106.
8. Xia Y, Ning Z, Bai G, Li R, Yan G, Siddique KH, et al. Allelic variations of a light harvesting chlorophyll a/b-binding protein gene (Lhcb1) associated with agronomic traits in barley. *PLoS One*. 2012;7(5):e37573.
9. Schuenemann D, Gupta S, Persello-Cartieaux F, Klimyuk VI, Jones JDG, Nussaume L, et al. A novel signal recognition particle targets light-harvesting proteins to the thylakoid membranes. *Proceedings of the National Academy of Sciences*. 1998;95(17):10312-6.
10. Falk S, Sinning I. cpSRP43 Is a Novel Chaperone Specific for Light-harvesting Chlorophyll a,b-binding Proteins. *Journal of Biological Chemistry*. 2010;285(28):21655-61.
11. Jaru-Ampornpan P, Shen K, Lam VQ, Ali M, Doniach S, Jia TZ, et al. ATP-independent reversal of a membrane protein aggregate by a chloroplast SRP subunit. *Nature Structural & Molecular Biology*. 2010;17:696.
12. Nguyen TX, Jaru-Ampornpan P, Lam VQ, Cao P, Piszkiwicz S, Hess S, et al. Mechanism of an ATP-independent protein disaggregase. I. Structure of a membrane protein aggregate reveals a mechanism of recognition by its chaperone. *Journal of Biological Chemistry*. 2013.

13. Jaru-Ampornpan P, Liang F-C, Nisthal A, Nguyen TX, Wang P, Shen K, et al. Mechanism of an ATP-independent protein disaggregase. II. Distinct molecular interactions drive multiple steps during aggregate disassembly. *Journal of Biological Chemistry*. 2013.
14. 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. *Proceedings of the National Academy of Sciences*. 1993;90(18):8552-6.
15. Jonas-Straube E, Hutin C, Hoffman NE, Schünemann D. Functional Analysis of the Protein-interacting Domains of Chloroplast SRP43. *Journal of Biological Chemistry*. 2001;276(27):24654-60.
16. Tu CJ, Peterson EC, Henry R, Hoffman NE. The L18 Domain of Light-harvesting Chlorophyll Proteins Binds to Chloroplast Signal Recognition Particle 43. *Journal of Biological Chemistry*. 2000;275(18):13187-90.
17. Cain P, Holdermann I, Sinning I, Johnson Arthur 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. *Biochemical Journal*. 2011;437(1):149-55.
18. DeLille J, Peterson EC, Johnson T, Moore M, Kight A, Henry R. A novel precursor recognition element facilitates posttranslational binding to the signal recognition particle in chloroplasts. *Proceedings of the National Academy of Sciences*. 2000;97(4):1926-31.
19. Stengel KF, Holdermann I, Cain P, Robinson C, Wild K, Sinning I. Structural Basis for Specific Substrate Recognition by the Chloroplast Signal Recognition Particle Protein cpSRP43. *Science*. 2008;321(5886):253-6.
20. 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*. 2009.
21. Henderson Rory C, Gao F, Jayanthi S, Kight A, Sharma P, Goforth Robyn L, et al. Domain Organization in the 54-kDa Subunit of the Chloroplast Signal Recognition Particle. *Biophysical Journal*. 2016;111(6):1151-62.
22. Gao F, Kight AD, Henderson RC, Jayanthi S, Patel P, Goforth RL, et al. Role of Structural Flexibility of cpSRP43 in Binding Substrates during Post-Translational Targeting. *Biophysical Journal*. 2015;108(2):61A-A.
23. Gao F, Kight AD, Henderson R, Jayanthi S, Patel P, Murchison M, et al. Regulation of Structural Dynamics within a Signal Recognition Particle Promotes Binding of Protein Targeting Substrates. *Journal of Biological Chemistry*. 2015.
24. Yang M-J, Pang X-Q, Zhang X, Han K-L. Molecular dynamics simulation reveals preorganization of the chloroplast FtsY towards complex formation induced by GTP binding. *Journal of Structural Biology*. 2011;173(1):57-66.
25. Robert B, Horton P, Pascal AA, Ruban AV. Insights into the molecular dynamics of plant light-harvesting proteins in vivo. *Trends in Plant Science*. 2004;9(8):385-90.

26. Zhang Y, Ding Y. Molecular dynamics simulation and bioinformatics study on chloroplast stromal ridge complex from rice (*Oryza sativa* L.). *BMC Bioinformatics*. 2016;17(1):28.
27. Gao F, Kight AD, Henderson R, Jayanthi S, Patel P, Murchison M, et al. Regulation of structural dynamics within a signal recognition particle promotes binding of protein targeting substrates. *Journal of Biological Chemistry*. 2015:jbc. M114. 624346.
28. Pelikan M, Hura GL, Hammel M. Structure and flexibility within proteins as identified through small angle X-ray scattering. *General physiology and biophysics*. 2009;28(2):174-89.
29. Kimanius D, Pettersson I, Schluckebier G, Lindahl E, Andersson M. SAXS-guided metadynamics. *Journal of chemical theory and computation*. 2015;11(7):3491-8.
30. Oroguchi T, Ikeguchi M. MD-SAXS method with nonspherical boundaries. *Chemical Physics Letters*. 2012;541:117-21.
31. Pelikan M, Hura GL, Hammel M. Structure and flexibility within proteins as identified through small angle X-ray scattering. *General physiology and biophysics*. 2009;28(2):174.
32. Hammond JB, Kruger NJ. The Bradford method for protein quantitation. *New Protein Techniques*: Springer; 1988. p. 25-32.
33. Kruger NJ. The Bradford method for protein quantitation. *The protein protocols handbook*: Springer; 2009. p. 17-24.
34. Eftink MR, Ghiron CA. Fluorescence quenching studies with proteins. *Analytical biochemistry*. 1981;114(2):199-227.
35. Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. *Journal of Molecular Graphics*. 1996;14(1):33-8.
36. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, et al. Scalable molecular dynamics with NAMD. *Journal of computational chemistry*. 2005;26(16):1781-802.
37. Brooks BR, Brooks III CL, Mackerell Jr AD, Nilsson L, Petrella RJ, Roux B, et al. CHARMM: the biomolecular simulation program. *Journal of computational chemistry*. 2009;30(10):1545-614.
38. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The protein data bank. *Nucleic acids research*. 2000;28(1):235-42.
39. Holdermann I, Meyer NH, Round A, Wild K, Sattler M, Sinning I. Chromodomains read the arginine code of post-translational targeting. *Nature Structural & Molecular Biology*. 2012;19:260.
40. Horn A, Hennig J, Ahmed YL, Stier G, Wild K, Sattler M, et al. Structural basis for cpSRP43 chromodomain selectivity and dynamics in Alb3 insertase interaction. *Nature Communications*. 2015;6:11.

41. Sivaraja V, Kumar TKS, Leena PST, Chang A-n, Vidya C, Goforth RL, et al. Three-Dimensional Solution Structures of the Chromodomains of cpSRP43. *Journal of Biological Chemistry*. 2005;280(50):41465-71.
42. Kathir KM, Rajalingam D, Sivaraja V, Kight A, Goforth RL, Yu C, et al. Assembly of Chloroplast Signal Recognition Particle Involves Structural Rearrangement in cpSRP43. *Journal of Molecular Biology*. 2008;381(1):49-60.
43. Huntington JA, Stein PE. Structure and properties of ovalbumin. *Journal of Chromatography B: Biomedical Sciences and Applications*. 2001;756(1):189-98.
44. Aoki T, Hiidome Y, Kitahata K, Sugimoto Y, Ibrahim HR, Kato Y. Improvement of heat stability and emulsifying activity of ovalbumin by conjugation with glucuronic acid through the Maillard reaction. *Food Research International*. 1999;32(2):129-33.
45. Liang FC, Kroon G, McAvoy CZ, Chi C, Wright PE, Shan SO. Conformational dynamics of a membrane protein chaperone enables spatially regulated substrate capture and release. *Proc Natl Acad Sci U S A*. 2016;113(12):E1615-24.
46. Di Giambattista M, Ide G, Engelborghs Y, Cocito C. Analysis of fluorescence quenching of ribosome-bound virginiamycin S. *Journal of Biological Chemistry*. 1984;259(10):6334-9.
47. Samworth CM, Degli ME, Lenaz G. Quenching of the intrinsic tryptophan fluorescence of mitochondrial ubiquinol--cytochrome-c reductase by the binding of ubiquinone. *European journal of biochemistry*. 1988;171(1-2):81-6.
48. Jamir K, Seshagirirao K. Fluorescence quenching, structural and unfolding studies of a purified cysteine protease, ZCPG from *Zingiber montanum* rhizome. *International journal of biological macromolecules*. 2018;106:277-83.
49. Papadopoulou A, Green RJ, Frazier RA. Interaction of flavonoids with bovine serum albumin: a fluorescence quenching study. *Journal of agricultural and food chemistry*. 2005;53(1):158-63.

### **III. Chaperone Activity of cpSRP43**

#### **Abstract**

Integration of light-harvesting chlorophyll binding proteins into the thylakoid membrane requires a specific chaperone, being the cpSRP43 subunit, of the signal recognition particle pathway in chloroplasts. cpSRP43, unique to the chloroplast, is responsible for transport of LHCPs through the stroma as well as assisting in the correct folding, assembly and disaggregation of these proteins for the acquisition of light energy. cpSRP43 is a highly flexible, multidomain protein capable of binding distinct partners in the cpSRP pathway. cpSRP43 is an irreplaceable component, necessary for the accurate and successful integration of LHCPs. It can act as a disaggregase without any input of external energy. Its action is based on the ability to associate with variable regions of different proteins owing to the domains and flexibility within its distinctive structure. Understanding the unique capabilities of cpSRP43 in the chloroplast begs the question of its usefulness outside of the plant cell, as well as its yet unknown roles still within the plant cell. Although the capabilities of cpSRP43 as a hub protein, adept to binding many unknown partners, has been alluded to in other works, it has yet to be thoroughly investigated. In this study we discover that cpSRP43 can act as a generic chaperone for proteins other than LHCP/not native to the chloroplast. The high thermal stability of cpSRP43 has been demonstrated in the previous chapter by its ability to retain its secondary structure as well as withstand aggregation upon heating and cooling cycles as confirmed by absorbance, intrinsic tryptophan fluorescence and far UV circular dichroism spectroscopy. This property gives cpSRP43 the basis to act as a generic chaperone and provide protection like that of typical heat shock proteins. Carbonic anhydrase, Concanavalin A and hFGF1 (acidic human fibroblast growth factor), were selected as candidates for chaperoning activity by cpSRP43. In all three cases, heat-induced aggregation of the candidate protein was either eliminated or significantly



reduced in the presence of cpSRP43. In the case of hFGF1, the bioactivity was preserved after heat-treatment in the presence of cpSRP43. We have proposed a mechanism by which cpSRP43 is able to execute this action however further investigation is warranted to determine the exact mechanism(s) which may vary dependent on the target protein.

## **Introduction**

The chloroplast signal recognition particle (cpSRP) executes the translocation of Light-harvesting Chlorophyll Binding Proteins (LHCPs) through the chloroplast stroma and assists, in coordination with the cpSRP receptor (FtsY) and the integral membrane protein ALB3, with their integration into the thylakoid membrane (1). The cpSRP consists of the conserved 54-kDa subunit (cpSRP54) and a unique 43-kDa subunit (cpSRP43) (2, 3). cpSRP and LHCP form a soluble "transit complex" thereby enabling the LHCPs to traverse the stroma without becoming involved in inappropriate interactions (1, 4). cpSRP43 is required to prevent aggregation of the hydrophobic LHCPs during their posttranslational targeting to the thylakoid membrane (5). cpSRP43 has been structural identified as having three chromodomains (CD1-3) and four ankyrin repeats (Ank1-4) located between CD1 and CD2 (4, 6). The crystal structure of cpSRP43 reveals two hydrophobic grooves which are disjointed by a positive ridge on one side and a highly negatively charged surface on the other side (7). The interaction between cpSRP and LHCP is achievable, in part, by the binding between the ANK regions of cpSRP43 and the L18 domain of LHCP, an 18-amino acid, hydrophilic peptide which connects the second and third trans membrane domains (TM2, TM3) (6, 8). The chromodomains located on the c-terminal of cpSRP43 allow for its ability to bind the methionine-rich domain of cpSRP54 lending to the conformation of the transit complex (6). In addition, cpSRP43 binds to portions of the integral membrane protein ALB3 to initiate and expedite docking at the receptor for integration. The

unique architecture and multi-domains which make up cpSRP43 make it an excellent framework upon which protein interactions can occur.

Investigation into the structure and chaperone cycle of cpSRP43 suggests that it is highly specific for chaperoning LHCPs (9). The aggregation of LHCPs was monitored in the presence of each cpSRP pathway component and cpSRP43 was shown to be the exclusive component for preventing the aggregation of LHCPs (5). cpSRP43 does not require ATPase activity as is the case for the Hsp104/ClpB family of chaperones, however its ability to dissolve aggregates of LHCPs is comparable to that family of disaggregases (9-12). Studies have also revealed that Hsp70, Hsp60 and trigger factor (TF), a bacterial chaperone involved with hydrophobic regions of proteins, cannot substitute for cpSRP43 indicating the specificity of cpSRP43 to LHCPs (5, 13, 14). Hydrophobic interactions between the trans-membrane regions of LHCP and cpSRP43 are also considered to contribute largely to the ability of cpSRP43 to successfully chaperone and convoy the LHCPs to their destination. The contribution of hydrophobic interactions has been asserted based on studies in which binding was not lost under high salt conditions, a 60-fold higher binding affinity between the full length LHCP versus the L18 peptide alone and the involvement of TM3 for efficient binding (5, 15, 16).

Proteins which act as molecular chaperones are part of a diversified classification due to their scope of specific functions (17). Proteins in this group maintain crucial roles in folding and assembly, refolding, translocation and preventing aggregation (17). In addition, chaperones assist under stress and disease conditions. The biogenesis of an abundant amount of membrane proteins typically takes place in conjunction with their insertion into the membrane, however, in posttranslational translocation of membrane proteins, the requirement of effective chaperones is imperative. Molecular chaperones which can prevent, and/or reverse protein aggregation

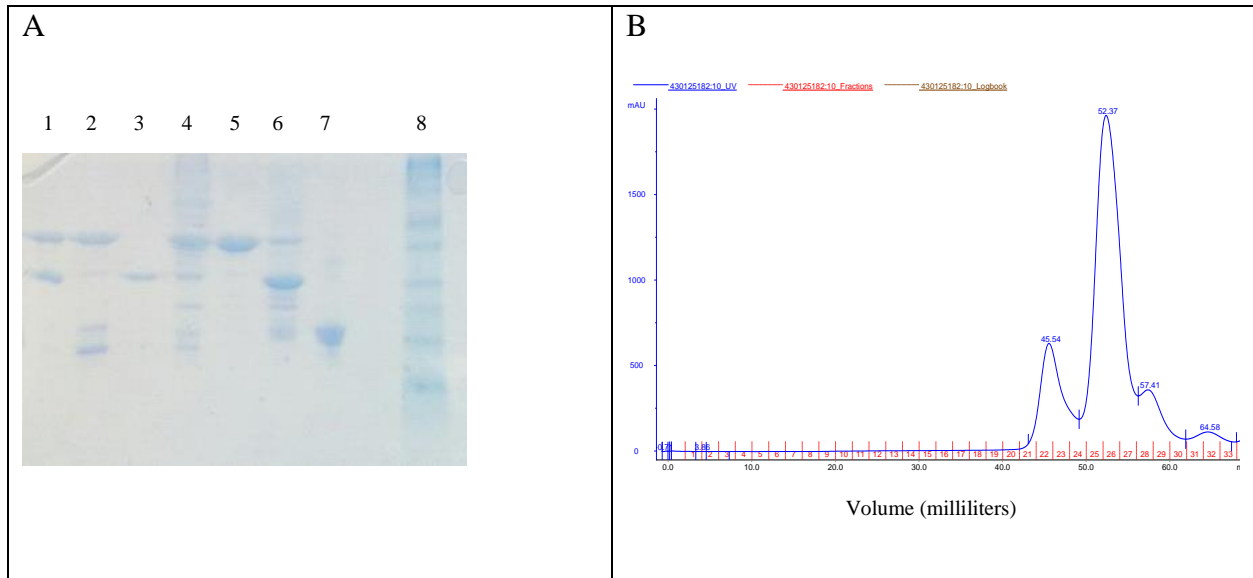
provide a significant contribution in the maintenance of cellular stability, proper function and prevention of disease states. Molecular chaperones known as "disaggregases" are capable of reversing protein aggregation with the aid of external energy provided by ATP hydrolysis as well as other co-chaperones (14, 18, 19). cpSRP43 is one of the first proteins to be described as an ATP-independent disaggregase. cpSRP43 provides protection to the most abundant protein on earth, giving it a highly specific and critical function. To investigate how cpSRP43 can recognize and reconstruct LHCP aggregates, the characteristics and structure of LHCP aggregates was elucidated (11, 20). LHCP aggregates are highly stable, insoluble, resistant to detergents or reversal by dilution (11). The L18 peptide and TM3 are shown to be solvent exposed in the aggregate formations which presents as the recognition site for cpSRP43 to first interact with the LHCP aggregate and begin the disassembly and reorganization of each LHCP protein (11). Recent studies using NMR and smFRET have also shown cpSRP43 to have high conformational flexibility (21, 22).

cpSRP43 is an indispensable component of the cpSRP pathway, involved in interactions with multiple partners leading to the successful translocation, disaggregation, refolding and integration of LHCPs. The ability of cpSRP43 to disassemble and remodel LHCP aggregates relies on its binding interactions with its substrate in place of any external energy source and its flexibility demonstrates its effectiveness and potential adaptability as a generic chaperone. Although LHCPs specifically require cpSRP43 as a chaperone, it stands to reason that cpSRP43 may also serve as a general chaperone which prevents aggregation and assists in the refolding of other proteins. The distinctive heat stability of cpSRP43 was outlined in the previous chapter. In this study, we demonstrate its ability to assist in refolding and prevention of heat-induced aggregation of three proteins which cpSRP43 from *Arabidopsis Thaliana* does not ordinarily

associate with by using Carbonic anhydrase from bovine erythrocytes, Concanavalin A from *Canavalia ensiformis* (Jack Bean) and, hFGF1 as substrates.

## **Materials and Methods**

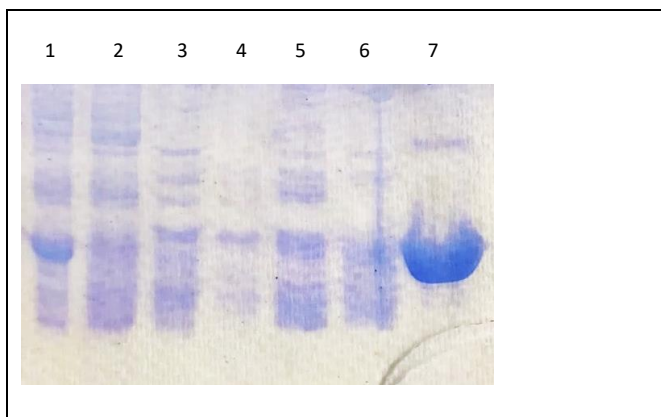
**Expression and Purification of Recombinant cpSRP43-** BL-21-star cells were transformed with the expression vector pGEX T2 encoding the sequence of cpSRP43 grown at 37°C lysogeny broth containing 10 µg/ml ampicillin. Protein expression was induced at an optical density of 0.8 measured at A600 with 1mM isopropyl- D-thiogalactoside. The cells were incubated for an additional 3.5 hours followed by harvesting by centrifugation. Cell pellets were resuspended and lysed by sonication for 25 cycles with 10 second on/off pulses using 10 W of output power. GST-cpSRP43 was bound to glutathione Sepharose then washed thoroughly with equilibration buffer (2.7mM KCl, 1.8mM KH<sub>2</sub>PO<sub>4</sub>, 15mM Na<sub>2</sub>HPO<sub>4</sub>, 137mM NaCl, pH 7.2). GST-cpSRP43 was eluted with 10 mM L-Glutathione and exchanged into cleavage buffer (50 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA, 1 mM dithiothreitol, pH 7.0). Overnight cleavage was setup, in solution, at 4°C on a rocker with 10 units of PreScission Protease per liter of original cells for 16 hours. The cleavage product was again put onto glutathione for separation of the GST tag and the cleaved cpSRP43 was further purified by size exclusion chromatography (SEC) as seen in Fig. 1 below.



**Fig.1: Purification of cpSRP43, SDS-PAGE (Panel-A) and the SEC profile (Panel-B) of cpSRP43 purification. Panel-B: Lane 1= Cleavage product containing cpSRP43 and GST, Lane 2= cleaved cpSRP43 eluted in the unbound fraction from GSH column, Lane 3= GST tag eluted with 10 mM L-Glutathione, Lane 4= Peak 1 from SEC eluted at 45 milliliters containing cpSRP43 and contaminants, Lane 5= Peak 2 from SEC eluted at 52 milliliters containing cpSRP43, Lane 6= Peak 3 from SEC eluted at 57 milliliters containing cpSRP43, GST and contaminants Lane 7= Peak 4 eluted from SEC containing low molecular weight contaminants.**

**Expression and Purification of hFGF1-** hFGF1 was overexpressed in BL-21(DE3) cells and grown to an Optical Density of 0.6–1.0 at Abs<sub>600</sub> and incubated with 1 mM isopropyl β-D-thiogalactoside for 3.5 hours. Cells were harvested, resuspended and sonicated. The supernatant was separated from the cell debris using ultra centrifugation. hFGF1 was purified on a heparin sepharose column using a stepwise salt gradient in 10mM sodium phosphate buffer containing 25mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> pH 7.2 and purity was analyzed using 15% SDS-PAGE (Fig.2). Pure hFGF1 elutes at a concentration of 1500 mM NaCl.

**Carbonic Anhydrase** from bovine erythrocytes and **Concanavalin A** from *Canavalia ensiformis* (Jack Bean) Type V were purchased commercially (Sigma).



**Fig.2: Purification of wtFGF1 on Heparin Sepharose column. Lane 1= Supernatant, Lane 2= Unbound fraction, Lane 3= 100mM NaCl elution, Lane 4= 300 mM NaCl elution, Lane 5= 500 mM NaCl elution, Lane 6= 800 mM NaCl elution, Lane 7= 1500 mM NaCl elution.**

**Circular Dichroism-** Circular dichroism (CD) measurements were performed on a Jasco J-1500 CD spectrometer equipped with a variable temperature cell holder. Conformational changes in the secondary structure of cpSRP43 were monitored in the far-UV region between 190 to 250 nm with a protein concentration of 0.3mg/ml in a quartz cuvette with a pathlength of 1 mm. The scanning speed, band width and data pitch were set to 50 nm/min, 1.00 nm and 0.1nm, respectively. Three scans were taken (within a 1000 HT voltage range) and averaged to obtain the CD spectra. The thermal denaturation/renaturation scans were recorded from 25°C to 90°C at 5°C increments.

**Size-Exclusion Chromatography (SEC)**— cpSRP43, hFGF1 and cpSRP43/hFGF1 proteins for the mixture samples were over-expressed and purified as described above. In the heat-treated and non-heat-treated samples, concentrations of hFGF1 and cpSRP43 used were 20 and 25  $\mu$ M, respectively. The heat-treated and non-heat-treated samples were loaded to a Superdex 75 16/600 column (GE Healthcare, Pittsburgh, PA) equilibrated in 2.7mM KCl, 1.8mM  $\text{KH}_2\text{PO}_4$ , 15mM  $\text{Na}_2\text{HPO}_4$ , 137mM NaCl, pH 7.2 on an AKTA FPLC and ran at a flow rate of 1 milliliter

per minute. The elution volume was monitored by absorbance at 280 nm. Fractions were collected and analyzed by SDS-PAGE.

**Refolding of Carbonic Anhydrase in the presence and absence of cpSRP43 as monitored by turbidity measurements-** All readings were obtained using an Agilent UV-VIS Spectrophotometer. Absorbance readings were recorded at 280 nm and 350 nm in order to track concentration and aggregation, respectively. Carbonic anhydrase was denatured and reduced in 8 M Urea with 1 mM BME and added to refolding buffer (10 mM phosphate, 100 mM NaCl, 0.1 mM GSH and 1 mM G-S-S-G at pH 7.2) starting at 0.1 mg/ml and increasing by 0.1 mg/ml for each reading. A total of ten readings were acquired for each experiment and the experiments were repeated three times each for denatured and reduced carbonic anhydrase increasing in refolding buffer in the presence and absence of cpSRP43. A constant concentration of 0.2 mg/ml (6  $\mu$ M) of cpSRP43 was present in the refolding experiments carried out in the presence of cpSRP43.

**Temperature-Dependent Aggregation Inhibition Experiments-** All absorbance readings were recorded at 280 and 350 nm wavelengths using an Agilent UV-VIS Spectrophotometer. The increasing temperatures were maintained using a variable temperature water-bath. Carbonic Anhydrase, Concanavalin A and hFGF1 were each heated at 5°C increments in the presence and absence of cpSRP43. All samples were held at each temperature increment for 5 minutes prior to measurement. The concentration of cpSRP43 was held constant for each substrate protein in the presence of cpSRP43 experiments. The concentrations of the substrate proteins were increasing for each reading in the presence and absence of cpSRP43.

**Bioactivity Assays-** 3T3 fibroblast cells obtained from ATCC (Manassas, VA) were cultured in media consisting of DMEM supplemented with 10% bovine calf serum. Cells were grown and

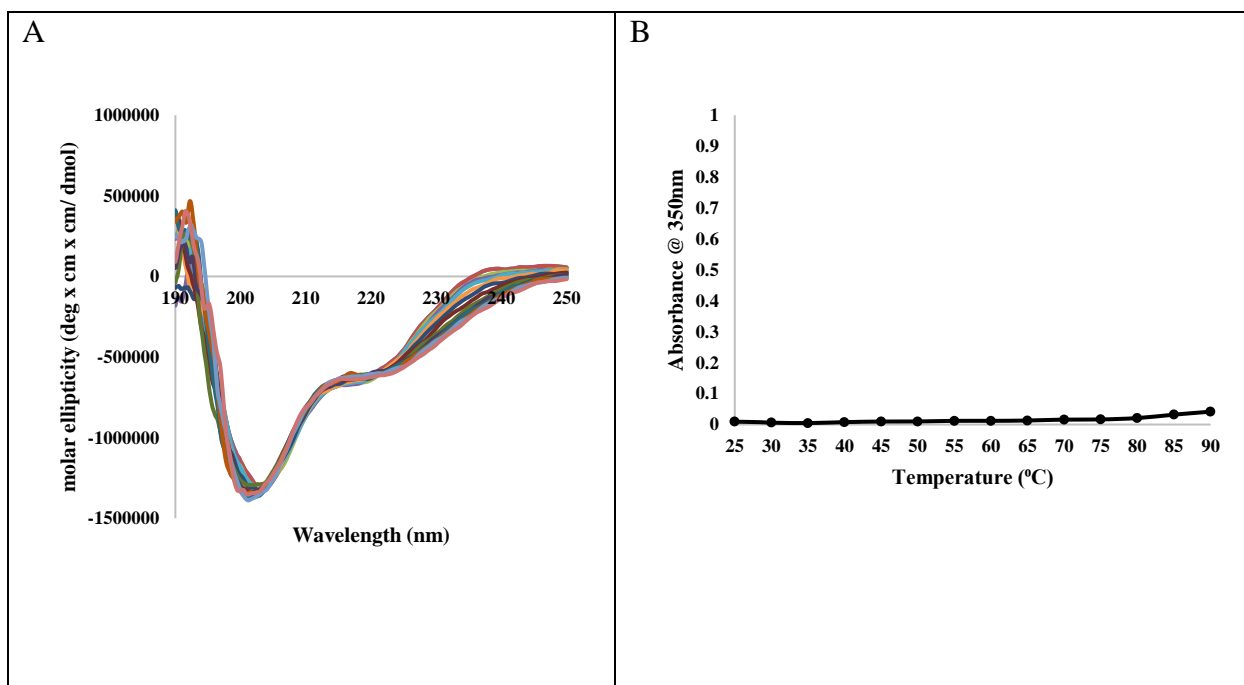
were incubated overnight at 37 °C. The bioactivity of FGF1 was determined by quantifying the cell number increase after being incubated with FGF1 post heat treatment at 60°C and FGF1 in the presence of cpSRP43 post heat treatment at 60°C. Starved 3T3 fibroblasts were collected and seeded in a well plate at a seeding density of 10,000 cells/well. The bioactivity assays were performed five times under the same condition. 3T3 cell proliferation was assessed by the Cell Titer-Glo (Promega, Madison, WI) cell proliferation assay after 24 hours.

## **Results and Discussion**

### ***cpSRP43 possesses high thermal stability***

cpSRP43 shows a circular dichroism spectrum with negative bands at 208 and 222 nm which indicate that it has a predominate, typical  $\alpha$  helical secondary structure. An intrinsic fluorescence emission at 341 nm indicates that the protein is in its native conformation. Thermal denaturation of cpSRP43 was probed by circular dichroism from 25-90°C at 5°C increments and retained its secondary structure both upon heating and cooling cycles as shown in the previous chapter. cpSRP43 is heat stable up to 90°C and can retain its secondary structure during heating. In addition, cpSRP43 was heated at five-degree increments from 25 to 90°C and the absorbance at 350 nm was recorded after the protein was held for five minutes at each temperature. cpSRP43 did not show visible aggregation or aggregation reflected by absorbance readings up to 90°C (Fig.3B). Subsequently, ultracentrifugation of the post heat treated samples did not indicate protein precipitation upon visual inspection.



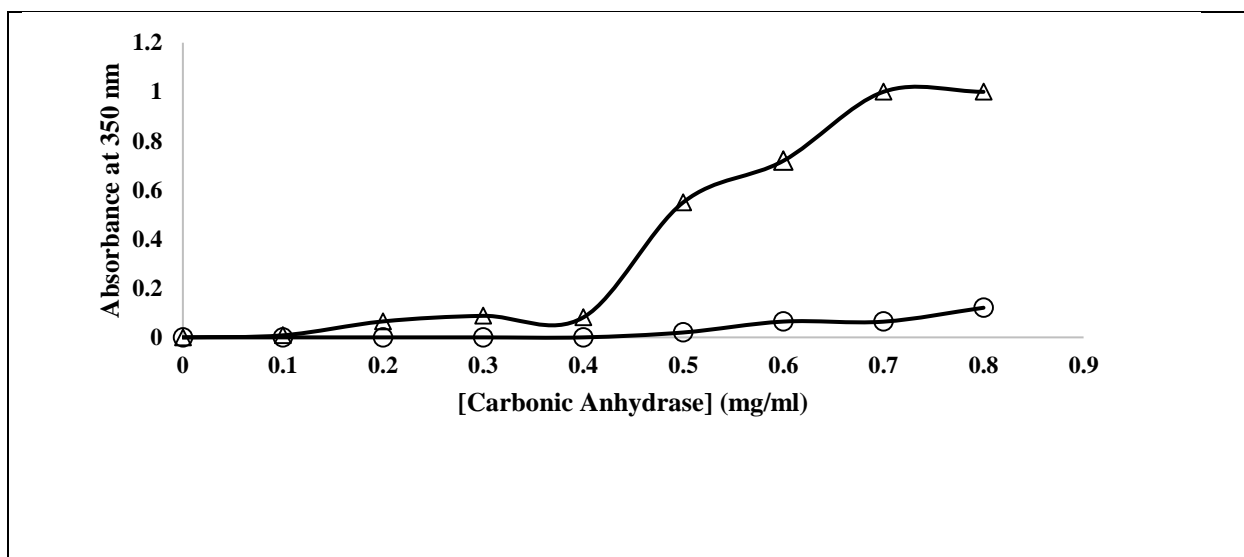


**Fig.3: Overlay of Far-UV CD spectra of cpSRP43 upon thermal denaturation (Panel-A). Aggregation profile of cpSRP43 (Panel-B) upon heat treatment as probed by absorbance at 350 nm.**

*cpSRP43 assists in the refolding of denatured and reduced Carbonic Anhydrase*

Carbonic anhydrase (CA), found in all kingdoms of life, catalyzes the reversible conversion of carbon dioxide and water to bicarbonate and also plays a role in ion transport, pH regulation and water and electrolyte balance (23, 24). CA II deficiency is responsible for osteopetrosis, an inherited syndrome in humans which is characterized with renal tubular acidosis and brain calcification (25-28). CA has been studied more closely over the past decade for its role in biological systems pertaining to bone resorption, renal acidification, and for its suggested importance in normal brain development (25, 28). CA was used in this study to illustrate the ability of cpSRP43 to aid in the refolding of a protein other than LHCP and although CA is naturally abundant in chloroplasts, the CA used here originated from bovine erythrocytes. A mixture of reduced and oxidized glutathione was incorporated in the refolding buffer for its usefulness in facilitating and/or accelerating disulfide bond formation once destroyed by

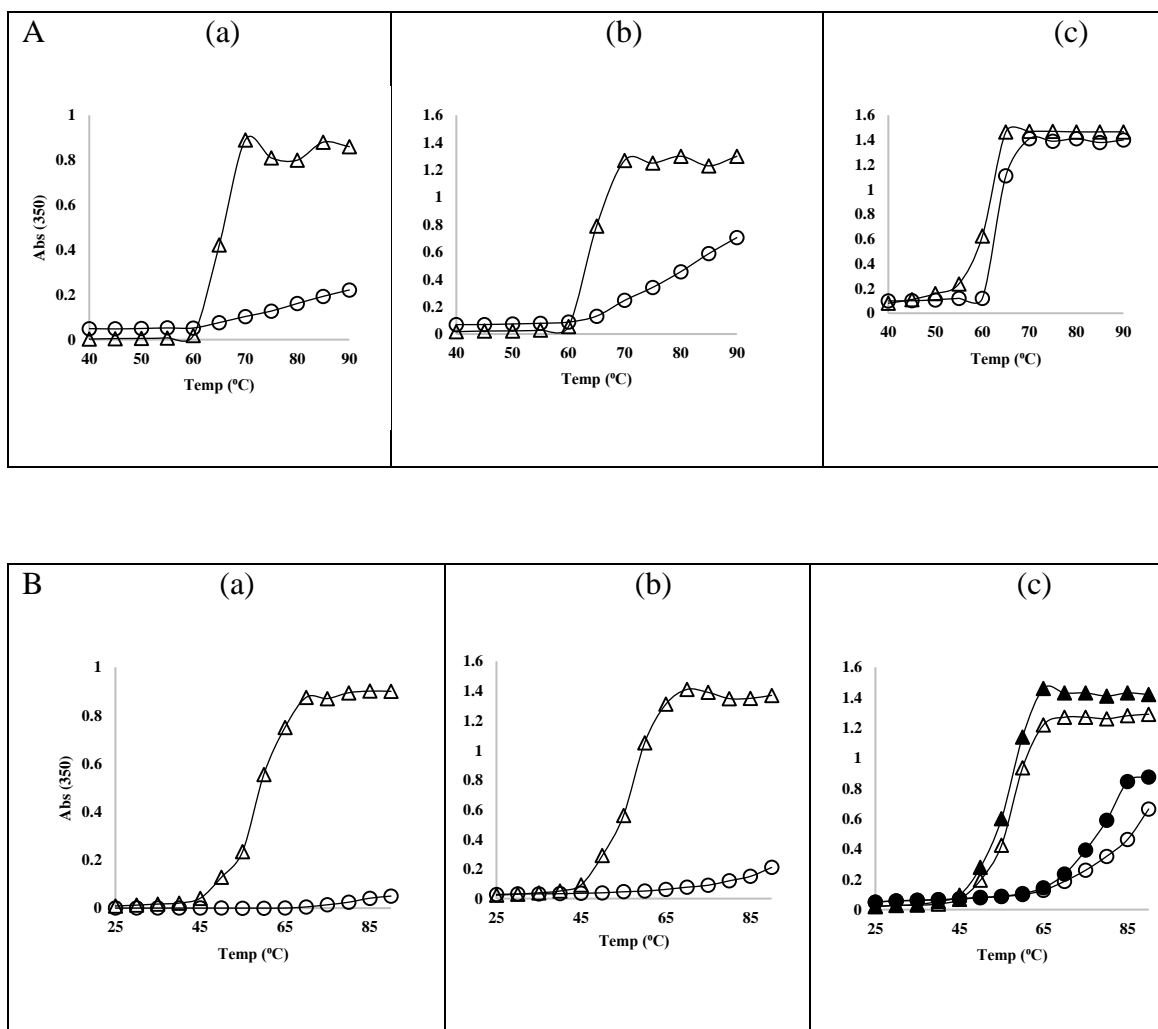
solubilization in a denaturant(29-31). A stock solution of CA was denatured and reduced in 8 M Urea containing 1 mM BME. A Urea buffer was prepared containing (10 mM phosphate, 100 mM NaCl, 8M Urea and 1 mM BME at pH 7.2) and a refolding buffer was prepared containing (10 mM phosphate, 100 mM NaCl, 0.1 mM GSH and 1 mM G-S-S-G at pH 7.2) Increasing concentrations of CA were added (0.1mg/ml increments) to refolding buffer in the presence and absence of cpSRP43 (Fig 4). In refolding buffer that did not contain cpSRP43, the CA began to aggregate at a concentration of 0.2 mg/ml. In refolding buffer containing cpSRP43 (constant concentration of 0.2 mg/ml), the CA did not show any aggregation as probed by absorbance readings taken at 350 nm until the CA reached a concentration of 0.5 mg/ml and did not show visible aggregation until the concentration reached 0.6 mg/ml (Fig.4). It appears that without the assistance of cpSRP43 the CA was aggregating almost immediately, and the aggregates were visible at a concentration of 0.3 mg/ml. The refolding buffer composition alone was not adequate to reduce the rate of aggregation of CA however the presence of cpSRP43 in the refolding buffer did postpone aggregate formation up to three times that of the refolding buffer alone. This result indicates that cpSRP43 can assist in the refolding of CA, being a concentration dependent action.



**Fig.4: Refolding of carbonic anhydrase in the presence (o) and absence (Δ) of cpSRP43.**

### ***cpSRP43 prevents heat-induced aggregation of Carbonic Anhydrase and Concanavalin A***

The chaperone ability of cpSRP43 was investigated by monitoring the heat-induced aggregation of different target proteins in the absence and presence of cpSRP43. A preliminary heat treatment experiment done using CA alone determined that CA was heat stable up to 60°C therefore the experiment was carried out from a starting temperature of 40°C. Increasing concentrations of carbonic anhydrase, in the presence and absence of cpSRP43, were heated from 40-90°C at 5°C increments and held for 5 minutes at each temperature before readings were taken at an absorbance of 350 nm. Each concentration of CA was added to eleven sample tubes (1 for each temperature) and made up with buffer in the control experiments and buffer containing 25 µM of cpSRP43 in the presence of cpSRP43 experiments prior to heating. The experiment was repeated at increasing concentrations of CA in the absence and presence of a constant concentration of cpSRP43. At the micromolar ratio in which cpSRP43 was in the most excess to CA (0.2: 1, Fig. 5A(a)), its presence was able to provide the best degree of protection. In the control experiment with CA alone it became completely aggregated by 70°C and at the same temperature in the presence of excess cpSRP43 there was still some aggregation present, but it was nine times less the amount of aggregation as in the control experiment. At equivalent ratios it appears that protection from heat-induced aggregation was not provided by cpSRP43 suggesting that the event in which cpSRP43 is acting as a chaperone for CA is concentration dependent (Fig.5A(c)). A similar trend was seen in the same experiments in which Concanavalin A (ConA) was used as the target protein for chaperoning activity provided by cpSRP43, however a better rate of protection was detected between Con A and cpSRP43 in comparison (Fig.5B).



**Fig. 5: (Panel-A) Heat-induced aggregation profile of Carbonic anhydrase in the presence (o) and absence (Δ) of cpSRP43. (Panel-B) Heat-induced aggregation profile of Concanavalin A. Concentrations are reported in micromolar ratios: Panel-A(a)(0.2:1), (b)(0.4:1), (c)(1:1). Panel-B(a)(0.13:1),(b)(0.27:1),(c)(Δ:o)(0.36:1), (c)(▲:●)(0.45:1).**

Con A is a lectin (carbohydrate-binding protein) which was initially isolated from the jack bean (*Canavalia ensiformis*) (32). It is a homotetramer at pH 7.0 containing 26.5 kDa subunits which bind to the non-reducing terminal  $\alpha$ -D-glucosyl and  $\alpha$ -D-manosyl groups (32, 33). Lectins are known to play physiological roles in animals but the roles of plant lectins are less understood (34). Although, Con A derived from plant proteins have proven to be especially useful as a probe in the study of cell surface membrane dynamics and mitosis in lymphocytes (33, 35, 36).

In chaperoning activity experimentation between Con A and cpSRP43, increasing concentrations of Con A, in the presence and absence of cpSRP43, were heated from 25-90°C at 5°C increments and held for 5 minutes at each temperature increment before readings were taken at an absorbance 350 nm, the same as described above for CA. Significant protection from heat-induced aggregation was observed in the mixture samples of Con A and cpSRP43 in which the ratio of Con A to cpSRP43 was lower i.e. cpSRP43 was in excess (Fig.5B). In all control experiments of Con A, aggregation began at 45°C and complete aggregation was measured and visually observed at 65°C. At 65°C, at Con A to cpSRP43 ratios of 0.13:1 and 0.27:1 micromolar, heat-induced aggregation of Con A was eliminated entirely and reduced by nearly 100%, respectively (Fig.3B (a) and (b)). In the heat-treatment experiments executed at higher ratios of Con A to cpSRP43 protection against aggregation was not significantly observed (Fig.5B (c)) which again supports the assertion that the chaperoning activity of cpSRP43 *in vitro* is concentration dependent.

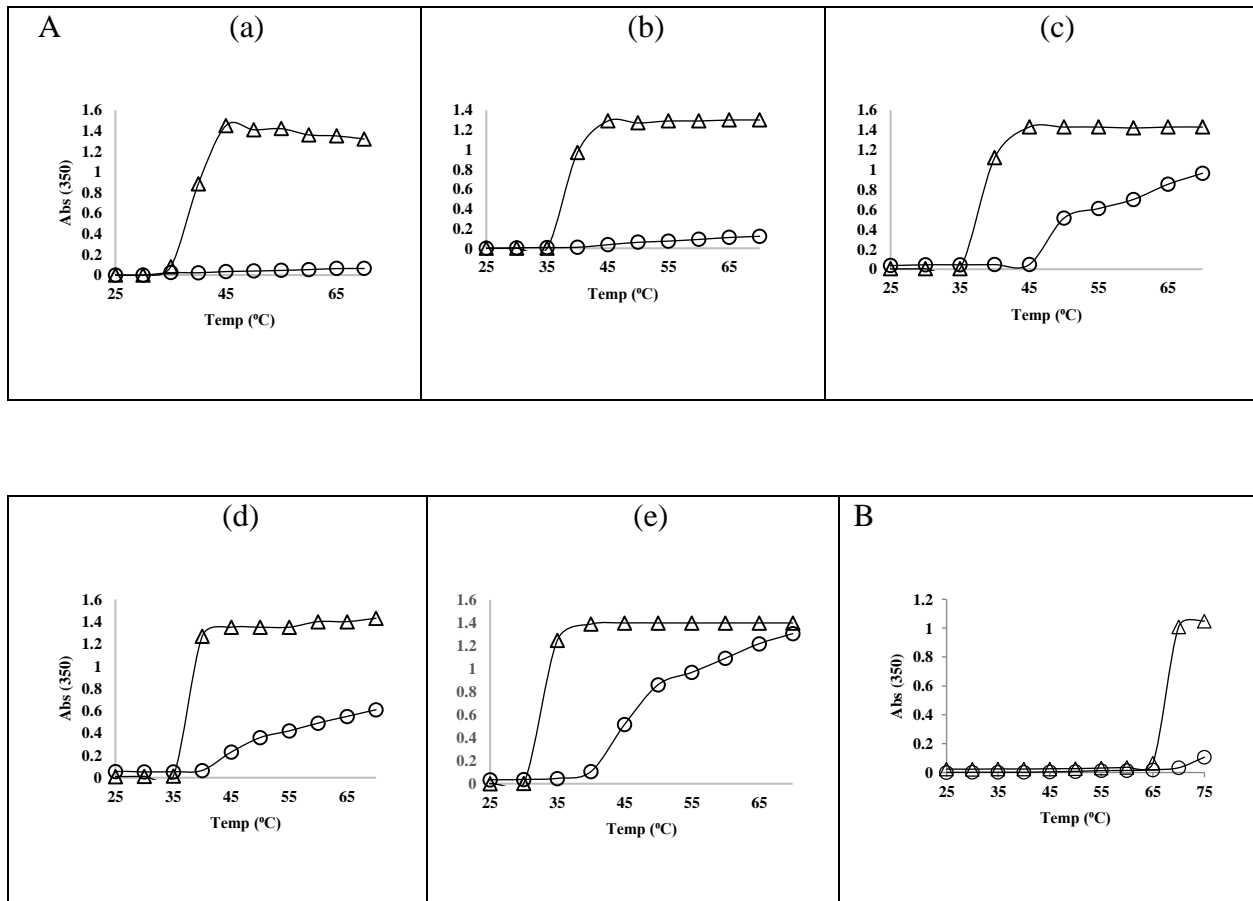
### ***cpSRP43 protects hFGF1 and lysozyme from heat-induced aggregation***

Human fibroblast growth factor (hFGF1) was selected to display the chaperoning effects of cpSRP43 on a protein which is not native to the chloroplast. Members of the FGF family are involved in signaling cell growth, differentiation and proliferation in addition to having roles in tissue repair, vascular stability and stress response (37-39). FGFs act by binding specific cell surface receptors (FGFRs) (38, 39). FGF1 is considered a heat-labile protein since it has a  $T_m$  of approximately 40°C which is close to physiological temperature (40, 41). The process by which the heat-induced aggregation inhibition experiments were executed are the same as previously described for CA and Con A. Absorbance readings of hFGF1 alone show that most of the protein has aggregated at 40°C and hFGF1 is completely aggregated at 45°C. Absorbance

readings of cpSRP43 alone at 350 nm do not show any turbidity upon heating from 25°C-90°C as shown previously. Heat-induced aggregation of hFGF1 is eliminated at a micromolar ratio of hFGF1 to cpSRP43 of 0.4:1 for up to 70°C (Fig.6A(a)) and up to 65°C at a ratio of 0.8:1 (Fig.6A(b)). Considering that in the absence of cpSRP43, hFGF1 was completely aggregated at 45°C the ability to delay this aggregation for 25 more degrees C is significant. In the previous experiments with CA and Con A, when the molar ratios were equivalent, cpSRP43 did not provide any protection against aggregation. However, in the case of hFGF1, at a ratio of 1.2:1 of hFGF1 to cpSRP43, hFGF1 is protected at 45°C, when it becomes completely aggregated otherwise (Fig.6A(c)). At higher molar ratios, protection is only seen at 40°C which indicates that when the quantity of hFGF1 exceeds that of cpSRP43 there is no reduction in the normal aggregation level of hFGF1 at its  $T_m$  (Fig.6A(d) and (e)). The protection which cpSRP43 provides to the heat induced aggregation of hFGF1 is concentration dependent with the best protection provided by cpSRP43 at an excess of cpSRP43. This result is consistent throughout experimentation with all three substrates.

In addition to the afore mentioned chaperone activity of cpSRP43 candidate proteins, lysozyme was also investigated. Lysozyme is an enzyme which catalyzes the breakdown of cell walls within certain bacteria as a component of the innate immune response in animals (42, 43). As a result of its antibacterial activity it is present in saliva, tears, mucus and human milk as well as being used in medicine and as a food preservative (44, 45). Lysozyme is already a highly heat-stable protein having a  $T_m$  of 72°C at pH 7.0 (46). In the control experiment, lysozyme aggregated completely at 70°C (Fig.4B). In the corresponding experiment in which the same concentration of lysozyme was heated in the presence of cpSRP43 at a micromolar ratio of lysozyme to cpSRP43 of 0.8:1, aggregation was eliminated at 70°C was decreased by 90% at

75°C (Fig.4B). Although cpSRP43 was not in great excess in this trial it still provided significant protection against heat-induced aggregation of lysozyme at temperatures at which this enzyme would have been completely aggregated. Testing the biological/enzymatic activity of each of these post heat treated chaperoning candidates would have been ideal. However, hFGF1 was chosen for investigation into its bioactivity post heat treatment in the presence of cpSRP43 and yielded positive results as reported in the following section.

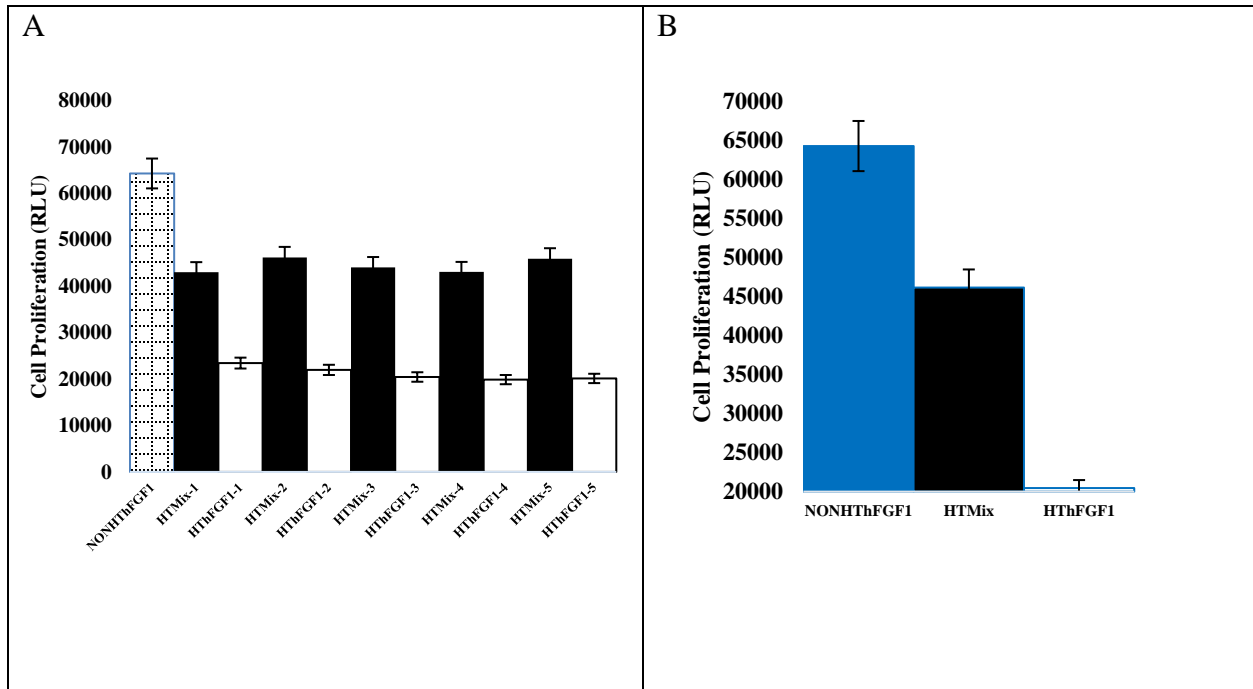


**Fig. 6: (Panel-A) Heat-induced aggregation profile of hFGF1 in the presence (o) and absence ( $\Delta$ ) of cpSRP43. (Panel-B) Heat-induced aggregation profile of lysozyme in the presence (o) and absence ( $\Delta$ ) of cpSRP43. Concentrations are reported in micromolar ratios: Panel-A(a)(0.4:1), (b)(0.8:1), (c)(1.2:1), (d)(1.6:1), (e)(2.4:1). Panel-B (0.8:1)**

***Post heat-treated hFGF1 (in the presence of cpSRP43) retains its cell proliferation activity***

The cell proliferation activity of hFGF1 on NIH 3T3 cells was quantified using post heat treated hFGF1 only and a post heat treated mixture of hFGF1 and cpSRP43 and hFGF1 only as a control. The same experiment was repeated five times. The concentrations of hFGF1 and cpSRP43 were the same in all trials. This was to examine whether or not hFGF1 retained its bioactivity after heating in the presence and absence of cpSRP43 and if so to what extent as compared to non-heat treated hFGF1. Bioactivity assays were performed on hFGF1 which did not undergoes heat treatment to represent the normal amount of cell proliferation expected. The same concentration of hFGF1 was heated at 50°C in the presence and absence of cpSRP43. Both samples were centrifuged at 13,000 rpm for 10 minutes. The supernatant of centrifuged, post heat-treated hFGF1 only sample was carefully removed from the precipitate and tested for bioactivity. The post heat treated mixture of hFGF1 to cpSRP43 used was at a micromolar concentration ratio of 0.4:1 since this concentration which showed the best result in the heat-induced aggregation inhibition experiments. Although the bioactivity of hFGF1 in the post heat-treated mixture did not reach the same levels as non-heat treated hFGF1, the presence of cpSRP43 during heating did lead to significantly better cell proliferation in comparison with the heated hFGF1 sample in the absence of cpSRP43 (Fig.7). Bioactivity of hFGF1 after heat treatment in the presence of cpSRP43 was doubled when compared to the cell proliferation activity of heat treated hFGF1 in the absence of cpSRP43. This is due to the fact that cpSRP43 was able to protect heated hFGF1 from aggregation. A mechanism by which cpSRP43 is able to provide this protection is explored in the discussion and conclusions section.



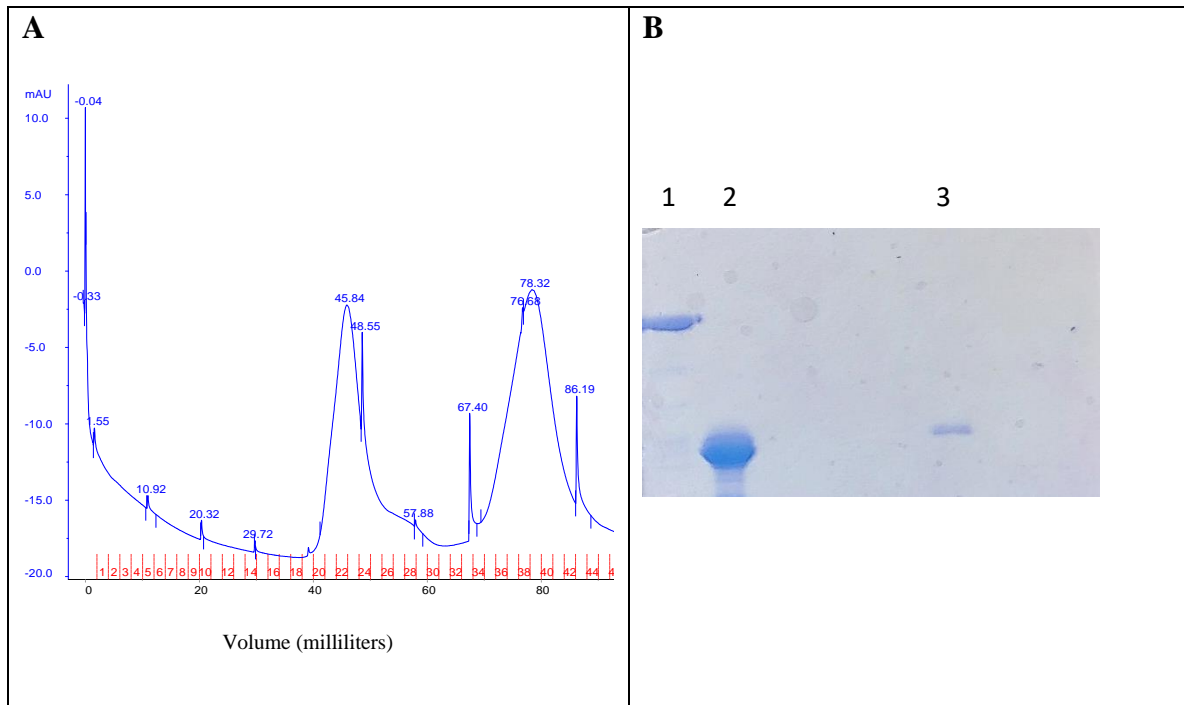


**Fig. 7: (Panel-A)-Proliferation of NIH 3T3 cells by heat-treated hFGF1 (HThFGF1-5) only and the heat- treated hFGF1/cpSRP43 mixture (HTMix1-5), repeated five times, samples 1-5. Cell proliferation of non-heat-treated hFGF1 is shown in the far left column of the bar graph (NONHThFGF1). The standard errors were calculated from triplicate measurements. (Panel-B)-One set of data selected from Panel-A. The base amount of starting cells is 20,000.**

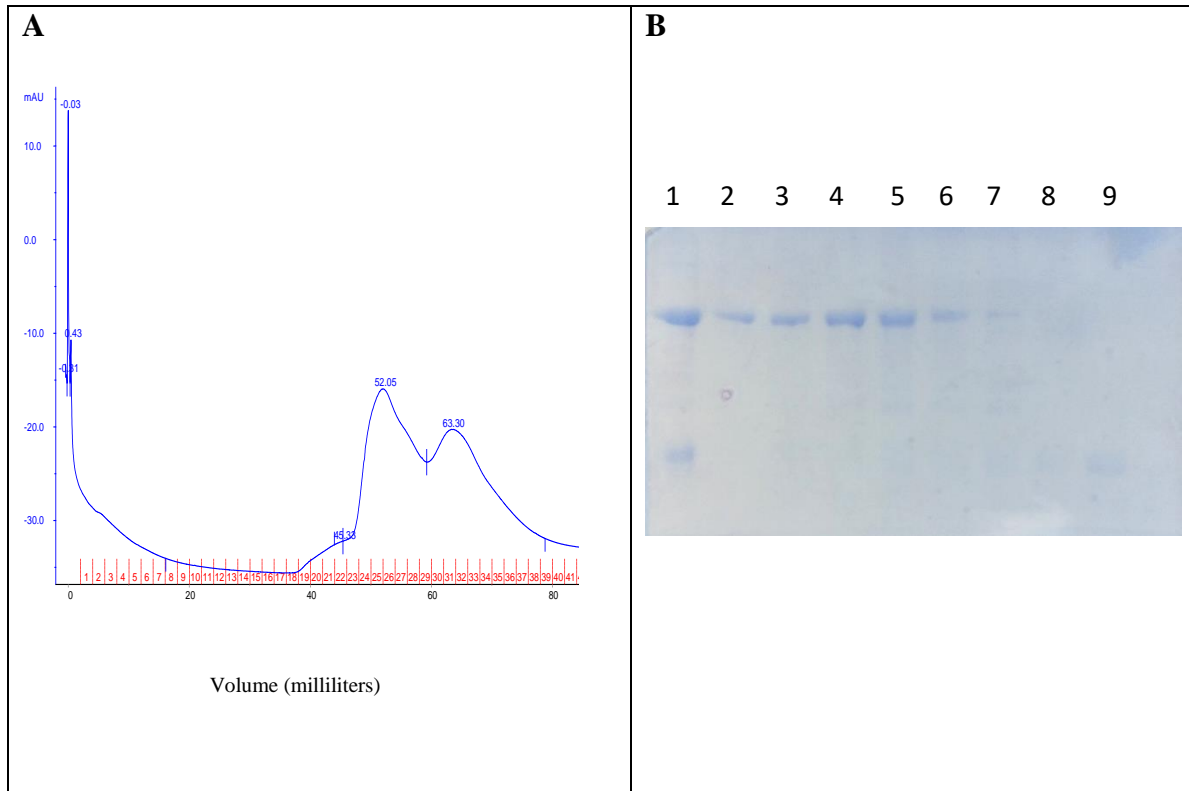
***Size exclusion chromatography reveals an interaction between cpSRP43 and hFGF1 under heated conditions which is not present under non-heated conditions***

Size exclusion chromatography was performed on heat-treated and non-heat-treated samples of the cpSRP43/hFGF1 mixtures (Fig.8 and 9). The chromatograph of the non-heat-treated mixture contained two well separated peaks at the appropriate elution volumes for cpSRP43 and hFGF1, at 45-52 milliliters and 76 milliliters, respectively (Fig.8A). The corresponding SDS-PAGE revealed a protein band of cpSRP43 and hFGF1 in the separate peaks (Fig.8B). This result suggests that cpSRP43 and hFGF1 do not bind or form associations with each other when combined and incubated at room temperature (20-25°C). The heat-treated SEC profile shows less resolution between the peaks with cpSRP43 eluting at its normal volume (45-52 milliliters) and hFGF1 eluting slightly earlier at 68 milliliters (Fig.9A). The results suggest that, upon

heating, cpSRP43 and hFGF1 undergo transient associations although it is not indicative of tight binding.



**Fig. 8: Size-exclusion Chromatography (Panel-A) and SDS-PAGE analysis (Panel-B) of the non-heat-treated mixture of hFGF1 and cpSRP43. Panel-B: Lane 1= Peak 1 collected containing cpSRP43, Lane 2= hFGF1 marker, Lane 3= Peak 2 collected containing hFGF1.**



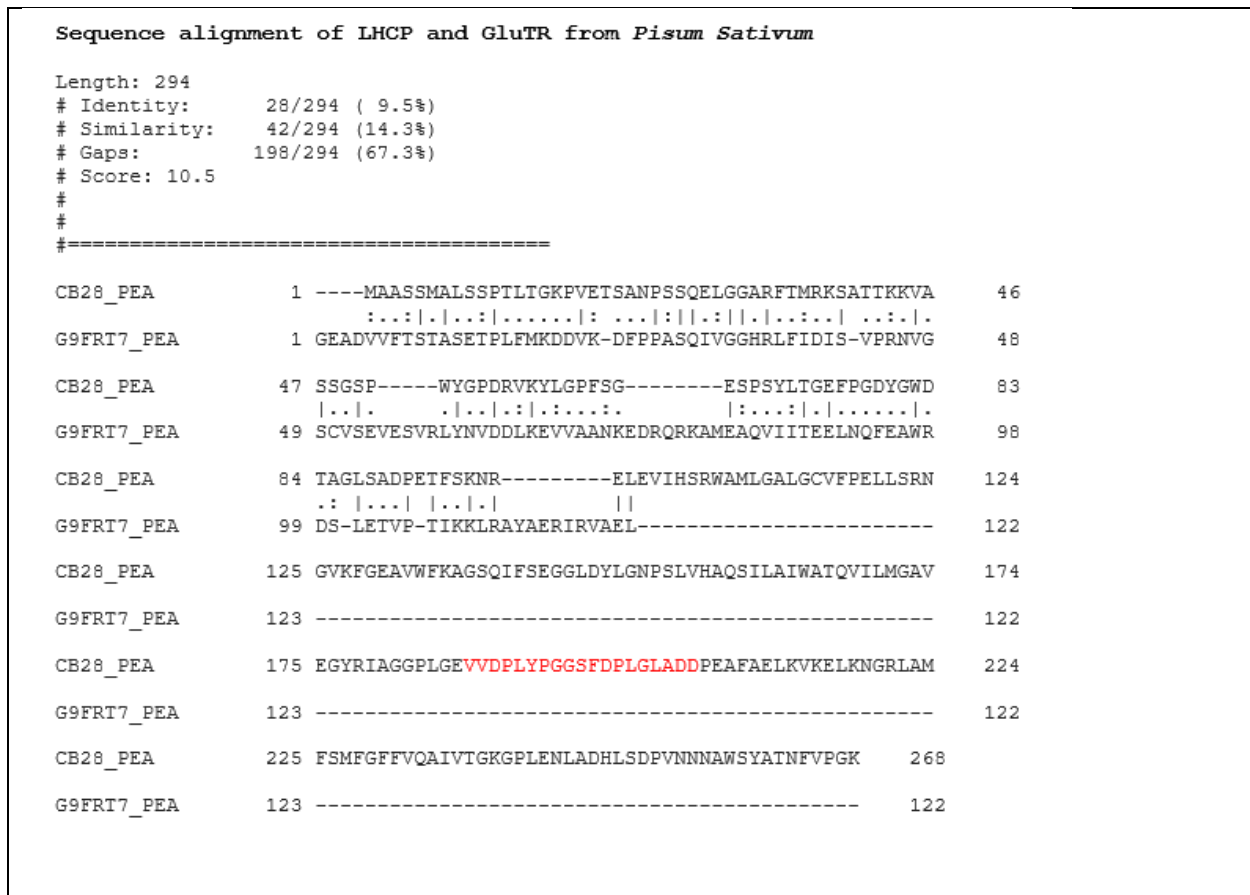
**Fig. 9: Size-exclusion Chromatography (Panel-A) and SDS-PAGE analysis (Panel-B) post heat-treated mixture of hFGF1 and cpSRP43. Panel-B: Lane 1= Preload to SEC containing cpSRP43 and hFGF1, Lanes 2-7= Peak 1 (fractions containing cpSRP43), Lanes 8 and 9= Peak 2 (fractions containing hFGF1).**

## Summary and Discussion

cpSRP43 is a vital component of the cpSRP pathway for its ability to interact with key players in this system and chaperone LHCPs through the stroma to the thylakoid membrane in an ATP-independent manner. LHCPs are apt to misfold and aggregate in aqueous environments due to their high hydrophobic content consequent to being integral membrane proteins. Certain reports have indicated that cpSRP43 exists in free form as a dimer in the stroma, as a potential intermediate, functioning in the formation of cpSRP as well as a disaggregase for adrift LHCP molecules (5,6). cpSRP43 has been shown to recognize and bind the L18 sequence, between TM2 and TM3 of LHCP (8). The conserved Tyrosine at position 204 in ANK3 of cpSRP43

recognizes the FDPLGL sequence in L18 (7, 8, 15, 20, 22). However, this interaction alone would not be adequate to guard the trans-membrane domains (TMDs) from aggregation (47). Investigation into how cpSRP43 can protect the TMDs of LHCP has only recently been communicated. Mapping of interaction sites between cpSRP43 and LHCb5 was done by alkylation efficiency of LHCb5 cysteine mutants using N-ethylmaleimide followed by site-specific cross-linking and the results of this combination of experiments revealed that cpSRP43 makes contacts and protects all three TMDs of the LHCb5 substrate (47). In site-directed mutagenesis studies, residues were categorized into two classes being mutations which disrupted cpSRP43's chaperoning of LHCP but not the binding of L18 and mutations in which both were disrupted (47). These experiments were the basis for designing a model of the interaction surface on cpSRP43 for the protection of the TMDs of LHCP which is comprised of hydrophobic surfaces on ANK4, the bridging helix at the N-terminal of the ankyrin regions and, the beta- hairpins along the ankyrin regions (47). Since cpSRP43 is a known chaperone of LHCPs, it is conceivable that the binding interface between these two proteins can be elucidated in detail upon adequate investigation. The binding interface between cpSRP43 and other substrates is likely similar. However, considering the composite structure of cpSRP43 and the intricate rearrangements which it undergoes to bind its known partners, defining the exact binding interface for cpSRP43 and the substrates in this study would require extensive examination. The binding interface proposed by McAvoy et al. is primarily made up of hydrophobic residues suggesting that the protection mode of cpSRP43 is driven by the formation of hydrophobic interactions. This theory is upheld in another recent study in which cpSRP43 has been identified as a chaperone for glutamyl-tRNA reductase (GluTR), the rate-limiting enzyme in tetrapyrrole biogenesis (48). Tetrapyrrole is at the core of chlorophyll and the biogenesis of

chlorophyll is coordinated with that of LHCPs in order for the proper assembly of light-harvesting complexes to proceed (48). Here they found that a deficit of cpSRP43 as well as GluTR-binding protein (GBP) significantly reduced production of GluTR and lead to a reduction in chlorophyll biogenesis which suggests that both cpSRP43 and GBP play distinct roles in the stabilization of GluTR (35). This group was able to form a ternary complex between cpSRP43, L18 and GluTR allowing them to speculate that both substrates are not competing for the same binding surface on cpSRP43 (48). Upon further examination, cpSRP43 was found to interact with the N-terminal of GluTR (48). The N- and C-terminal domains of GluTR can bind to different regulators to issue various control mechanisms for the enzyme (49) GBP was already known to interact with a heme-binding domain on the N-terminal of GluTR and this group conducted bimolecular fluorescence complementation assays (BiFC) and HIS pull-down assays in order to determine which GluTR domain interacts with cpSRP43 (48). cpSRP43 was found to bind to GluTR-N64–163 which contains the heme-binding domain as well as part of a catalytic domain, adjacent to the heme-binding domain, in the N-terminal (48). GluTR, in excess, is prone to aggregation and by using an aggregation-predicting algorithm, TANGO (50), the catalytic domain of GluTR was predicted to be an aggregation-prone region (48). Although this is a newly discovered role for cpSRP43, it is not surprising that cpSRP43 is providing a chaperoning function for another protein within the chloroplast. The L18 sequence in the LHCP substrate is available for specific recognition by cpSRP43. Upon comparison between the sequence of GluTR and LHCP from *Pisum Sativum*, the L18 recognition sequence is not present in GluTR however there is approximately 14.3% sequence similarity and 9.5% identity along the N-terminal region of each protein. Since there is no L18 like sequence or otherwise known recognition sequence in GluTR the GluTR protein, the question presents itself



**Fig.10: Sequence alignment of GluTR and LHCP from *Pisum Sativum*. Sequences were derived from uniprot.org. Sequence alignment was carried out using EMBOSS needle pairwise alignment.**

as to how cpSRP43 is targeting this enzyme. Proper folding and assembly of LHCPs require the incorporation of chlorophylls and therefore chlorophyll biosynthesis needs to be coordinated with the insertion of LHCPs into the thylakoid membrane (48). How this coordination is taking place as well as how cpSRP43 is able to identify the GluTR as a substrate requires further investigation. We have seen that cpSRP43 undergoes differing conformational rearrangements upon binding to partners along the cpSRP pathway (51). In addition, the exceptional ability of cpSRP43 to solubilize a substrate is key for its chaperoning function and can be achieved through making hydrophobic and electrostatic contacts (1, 4, 6, 8, 47, 52, 53). These exceptional properties of cpSRP43 may make it challenging to define the binding interface between

cpSRP43 and its substrates. As mentioned earlier, cpSRP43 is extremely sensitive to point mutations wherein the binding of L18 and chaperoning effect of cpSRP43 on LHCP was lost (47). These results suggest that cpSRP43 has many key residues which make it able to specifically chaperone LHCP, however the data collected in this chapter has demonstrated that cpSRP43 can act as a generic chaperone. There are many factors which can trigger protein aggregation including changes in ionic strength, pH, temperature and any other factor which can lead to interfacial instabilities (54). In this study we focused primarily on the role of cpSRP43 in the inhibition of heat-induced aggregation. The chaperoning function of cpSRP43 in the cpSRP pathway resembles that of the Hsp104/ClpB family of proteins. Hsp104 and ClpB work in combination with Hsp70 or DnaK to disaggregate and remodel proteins under severe stress conditions and requires energy from ATP to perform (20). Although this family of proteins are impressive disaggregating machines, it is noteworthy to consider that cpSRP43 has the same thermoresistance and, in addition, can disaggregate and remodel the LHCP proteins without the input of external energy. The structure of Hsp70 proteins consists of an N-terminal nucleotide binding domain (NBD) and a C-terminal polypeptide substrate binding domain (SBD) which are connected by a conserved, flexible linker (55). The SBD is divided into two subdomains, a  $\beta$ -sheet subdomain containing a peptide binding pocket and an  $\alpha$ -helical subdomain which acts as a lid (56). The NBD contains four subdomains which are separated into two sections that comprise an ATP binding site (57). The function of Hsp70s depends on interactions between the SBD and a short region within the substrate molecule(s) however the affinity of Hsp70s to their substrates is regulated by the NBD (58). Hsp70s are approximately 40% larger than cpSRP43 and have from 20-30% sequence similarity on average. Structural similarities between Hsp70s and cpSRP43 are low however they both contain regions which make critical binding interactions

with substrates/binding partners. Small heat shock proteins (sHSPs) in plants have structures similar to conserved sHSPs which consists of a variable N-terminal region, an  $\alpha$ -crystallin domain consisting of conserved  $\beta$ -sheets and a variable C-terminal extension (59). In comparison of cpSRP43 with Hsp21, a small heat shock protein found in *Arabidopsis Thaliana*, the sequence and structural similarities appear minimal, however the functional comparison between cpSRP43 and sHSPs shows many similarities. Upon unfolding, the murine Hsp25 loses its tertiary and quaternary structure while its secondary structure remains intact and unfolding is completely reversible (60). These are the same remarkable physical properties which were observed for cpSRP43 and reported in the previous chapter. Here we have shown that cpSRP43 has the capability to delay or inhibit the heat-induced aggregation of proteins of diverse sizes and characteristics. CA, a 30 kDa enzyme, was able to resist aggregation in the presence of cpSRP43 upon the application of heat and subsequent to denaturation and reduction. Con A, a homotetramer with each subunit having a molecular weight of 26.5 kDa showed substantial reduction in its propensity to aggregate in the presence of cpSRP43 as compared to the samples of Con A which were incrementally heated in the absence of cpSRP43. The rate of aggregation of hFGF1 was significantly reduced while the bioactivity was retained after being heated in the presence of cpSRP43 as compared to in the absence of cpSRP43 wherein most of the hFGF1 became aggregated. The best protection in each scenario was found at ratios of substrate to cpSRP43 in which cpSRP43 was in higher abundance as well as providing the best protection to smaller target proteins supporting the theory that this is a size and concentration-dependent function. Given the information known about the interaction between cpSRP43 and LHCP as well as the more recent report in which cpSRP43 is serving as a chaperone for GluTR, it is feasible to conclude that the mechanism of protection is mainly interchangeable between



substrate. cpSRP43 can predominately form hydrophobic interactions with its substrate thereby shielding it from aggregation or aiding in protein refolding. Although we know that cpSRP43 is more successful when present in excess to its substrate, the question remains as to how many molecules of cpSRP43 are interacting along each target protein. This information is expected to vary depending on the substrate. In this study, the chaperoning effect of cpSRP43 has been extended to and substantiated by proteins which are not native to the chloroplast. In the case of all three candidate proteins requiring chaperone facilitation, either the existence of concentration/heat-induced aggregation was not detected, or the rate of aggregate accumulation was significantly lowered in the presence of cpSRP43. The exact mechanism by which cpSRP43 can successfully provide protection to these proteins requires further investigation. It is highly anticipated that many more roles for cpSRP43 in the chloroplast will be revealed and now that we have displayed the ability of cpSRP43 to act as a generic chaperone, further investigation using a variety of substrates may lead to the use of cpSRP43 in applied biochemical and medical science as well as the identification of additional roles within its native chloroplast.

**Author Contributions:** Bioactivity Assays were performed by Dr. Ravi Kumar.

**Acknowledgements:** This work was supported by the DOE grant DE-02-01ER15161.

## References

1. Li X, Henry R, Yuan J, Cline K, Hoffman NE. A chloroplast homologue of the signal recognition particle subunit SRP54 is involved in the posttranslational integration of a protein into thylakoid membranes. *Proceedings of the National Academy of Sciences*. 1995;92(9):3789-93.
2. Klimyuk VI, Persello-Cartieaux F, Havaux M, Contard-David P, Schuenemann D, Meierhoff K, 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*. 1999;11(1):87-99.
3. Schünemann D. Structure and function of the chloroplast signal recognition particle. *Current genetics*. 2004;44(6):295-304.
4. Schuenemann D, Gupta S, Persello-Cartieaux F, Klimyuk VI, Jones JDG, Nussaume L, et al. A novel signal recognition particle targets light-harvesting proteins to the thylakoid membranes. *Proceedings of the National Academy of Sciences*. 1998;95(17):10312-6.
5. Falk S, Sinning I. cpSRP43 is a novel chaperone specific for light-harvesting chlorophyll a, b-binding proteins. *Journal of Biological Chemistry*. 2010;285(28):21655-61.
6. Jonas-Straube E, Hutin C, Hoffman NE, Schünemann D. Functional Analysis of the Protein-interacting Domains of Chloroplast SRP43. *Journal of Biological Chemistry*. 2001;276(27):24654-60.
7. Stengel KF, Holdermann I, Cain P, Robinson C, Wild K, Sinning I. Structural basis for specific substrate recognition by the chloroplast signal recognition particle protein cpSRP43. *Science*. 2008;321(5886):253-6.
8. Tu CJ, Peterson EC, Henry R, Hoffman NE. The L18 Domain of Light-harvesting Chlorophyll Proteins Binds to Chloroplast Signal Recognition Particle 43. *Journal of Biological Chemistry*. 2000;275(18):13187-90.
9. Falk S, Sinning I. cpSRP43 Is a Novel Chaperone Specific for Light-harvesting Chlorophyll a,b-binding Proteins. *Journal of Biological Chemistry*. 2010;285(28):21655-61.
10. Jaru-Ampornpan P, Shen K, Lam VQ, Ali M, Doniach S, Jia TZ, et al. ATP-independent reversal of a membrane protein aggregate by a chloroplast SRP subunit. *Nature Structural & Molecular Biology*. 2010;17:696.
11. Nguyen TX, Jaru-Ampornpan P, Lam VQ, Cao P, Piszkiwicz S, Hess S, et al. Mechanism of an ATP-independent protein disaggregase. I. Structure of a membrane protein aggregate reveals a mechanism of recognition by its chaperone. *Journal of Biological Chemistry*. 2013.
12. Jaru-Ampornpan P, Liang F-C, Nisthal A, Nguyen TX, Wang P, Shen K, et al. Mechanism of an ATP-independent protein disaggregase. II. Distinct molecular

- interactions drive multiple steps during aggregate disassembly. *Journal of Biological Chemistry*. 2013.
13. 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. *Proceedings of the National Academy of Sciences*. 1993;90(18):8552-6.
  14. Doyle SM, Wickner S. Hsp104 and ClpB: protein disaggregating machines. *Trends in biochemical sciences*. 2009;34(1):40-8.
  15. Cain P, Holdermann I, Sinning I, Johnson Arthur 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. *Biochemical Journal*. 2011;437(1):149-55.
  16. DeLille J, Peterson EC, Johnson T, Moore M, Kight A, Henry R. A novel precursor recognition element facilitates posttranslational binding to the signal recognition particle in chloroplasts. *Proceedings of the National Academy of Sciences*. 2000;97(4):1926-31.
  17. Hristozova N, Tompa P, Kovacs D. A novel method for assessing the chaperone activity of Proteins. *PloS one*. 2016;11(8):e0161970.
  18. Doyle SM, Shorter J, Zolkiewski M, Hoskins JR, Lindquist S, Wickner S. Asymmetric deceleration of ClpB or Hsp104 ATPase activity unleashes protein-remodeling activity. *Nature Structural and Molecular Biology*. 2007;14(2):114.
  19. Glover JR, Lindquist S. Hsp104, Hsp70, and Hsp40: a novel chaperone system that rescues previously aggregated proteins. *Cell*. 1998;94(1):73-82.
  20. Jaru-Ampornpan P, Liang F-C, Nisthal A, Nguyen TX, Wang P, Shen K, et al. Mechanism of an ATP-independent protein disaggregase. II. Distinct molecular interactions drive multiple steps during aggregate disassembly. *Journal of Biological Chemistry*. 2013;jbc. M113. 462861.
  21. Gao F, Kight AD, Henderson R, Jayanthi S, Patel P, Murchison M, et al. Regulation of Structural Dynamics within a Signal Recognition Particle Promotes Binding of Protein Targeting Substrates. *Journal of Biological Chemistry*. 2015.
  22. Liang F-C, Kroon G, McAvoy CZ, Chi C, Wright PE, Shan S-o. Conformational dynamics of a membrane protein chaperone enables spatially regulated substrate capture and release. *Proceedings of the National Academy of Sciences*. 2016:201524777.
  23. Lindskog S. Structure and mechanism of carbonic anhydrase. *Pharmacology & therapeutics*. 1997;74(1):1-20.
  24. DiMario RJ, Clayton H, Mukherjee A, Ludwig M, Moroney JV. Plant carbonic anhydrases: structures, locations, evolution, and physiological roles. *Molecular plant*. 2017;10(1):30-46.

25. Sly WS, Hu PY. Human carbonic anhydrases and carbonic anhydrase deficiencies. *Annual review of biochemistry*. 1995;64(1):375-401.
26. Hassan MI, Shajee B, Waheed A, Ahmad F, Sly WS. Structure, function and applications of carbonic anhydrase isozymes. *Bioorganic & medicinal chemistry*. 2013;21(6):1570-82.
27. Sly WS, Hewett-Emmett D, Whyte MP, Yu Y-S, Tashian RE. Carbonic anhydrase II deficiency identified as the primary defect in the autosomal recessive syndrome of osteopetrosis with renal tubular acidosis and cerebral calcification. *Proceedings of the National Academy of Sciences*. 1983;80(9):2752-6.
28. Bhatt A, Mondal UK, Supuran CT, Ilies MA, McKenna R. Crystal structure of carbonic anhydrase II in complex with an activating ligand: implications in neuronal function. *Molecular neurobiology*. 2018;55(9):7431-7.
29. Cumming RC, Andon NL, Haynes PA, Park M, Fischer WH, Schubert D. Protein disulfide bond formation in the cytoplasm during oxidative stress. *Journal of biological chemistry*. 2004;279(21):21749-58.
30. Frand AR, Cuozzo JW, Kaiser CA. Pathways for protein disulphide bond formation. *Trends in cell biology*. 2000;10(5):203-10.
31. Sies H. Glutathione and its role in cellular functions. *Free Radical Biology and Medicine*. 1999;27(9-10):916-21.
32. Kiernan J. Localization of  $\alpha$ -D-glucosyl and  $\alpha$ -D-mannosyl groups of mucosubstances with concanavalin A and horseradish peroxidase. *Histochemistry*. 1975;44(1):39-45.
33. Saleemuddin M, Husain Q. Concanavalin A: a useful ligand for glycoenzyme immobilization—a review. *Enzyme and microbial technology*. 1991;13(4):290-5.
34. Gunther GR, Wang JL, Yahara I, Cunningham BA, Edelman GM. Concanavalin A derivatives with altered biological activities. *Proceedings of the National Academy of Sciences*. 1973;70(4):1012-6.
35. Wang JL, Edelman GM. Binding and functional properties of concanavalin A and its derivatives. II. A proteolytic product with saccharide-binding activity. *Journal of Biological Chemistry*. 1978;253(9):3008-15.
36. Dwyer J, Johnson C. The use of concanavalin A to study the immunoregulation of human T cells. *Clinical and experimental immunology*. 1981;46(2):237.
37. Prudovsky I, Kacer D, Davis J, Shah V, Jayanthi S, Huber I, et al. Folding of fibroblast growth factor 1 is critical for its nonclassical release. *Biochemistry*. 2016;55(7):1159-67.
38. Lin X, Buff EM, Perrimon N, Michelson AM. Heparan sulfate proteoglycans are essential for FGF receptor signaling during *Drosophila* embryonic development. *Development*. 1999;126(17):3715-23.

39. Ornitz DM, Itoh N. Fibroblast growth factors. *Genome biology*. 2001;2(3):reviews3005. 1.
40. Zakrzewska M, Krowarsch D, Wiedlocha A, Otlewski J. Design of fully active FGF-1 variants with increased stability. *Protein Engineering Design and Selection*. 2004;17(8):603-11.
41. Blaber SI, Culajay JF, Khurana A, Blaber M. Reversible thermal denaturation of human FGF-1 induced by low concentrations of guanidine hydrochloride. *Biophysical journal*. 1999;77(1):470-7.
42. Ibrahim HR, Matsuzaki T, Aoki T. Genetic evidence that antibacterial activity of lysozyme is independent of its catalytic function. *FEBS letters*. 2001;506(1):27-32.
43. Jollès P, Jollès J. What's new in lysozyme research? *Molecular and cellular biochemistry*. 1984;63(2):165-89.
44. Cegielska-Radziejewska R, Lesnierowski G, Kijowski J. Properties and application of egg white lysozyme and its modified preparations-a review. *Polish Journal of food and nutrition sciences*. 2008;58(1).
45. Cunningham F, Proctor V, Goetsch S. Egg-white lysozyme as a food preservative: an overview. *World's Poultry Science Journal*. 1991;47(2):141-63.
46. Venkataramani S, Truntzer J, Coleman DR. Thermal stability of high concentration lysozyme across varying pH: A Fourier Transform Infrared study. *Journal of pharmacy & bioallied sciences*. 2013;5(2):148.
47. McAvoy CZ, Siegel A, Piszkiwicz S, Miaou E, Yu M, Nguyen T, et al. Two distinct sites of client protein interaction with the chaperone cpSRP43. *Journal of Biological Chemistry*. 2018;293(23):8861-73.
48. Wang P, Liang F-C, Wittmann D, Siegel A, Shan S-o, Grimm B. Chloroplast SRP43 acts as a chaperone for glutamyl-tRNA reductase, the rate-limiting enzyme in tetrapyrrole biosynthesis. *Proceedings of the National Academy of Sciences*. 2018;115(15):E3588-E96.
49. Brzezowski P, Richter AS, Grimm B. Regulation and function of tetrapyrrole biosynthesis in plants and algae. *Biochimica Et Biophysica Acta (BBA)-Bioenergetics*. 2015;1847(9):968-85.
50. Fernandez-Escamilla A-M, Rousseau F, Schymkowitz J, Serrano L. Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nature biotechnology*. 2004;22(10):1302.
51. Gao F, Kight AD, Henderson R, Jayanthi S, Patel P, Murchison M, et al. Regulation of structural dynamics within a signal recognition particle promotes binding of protein targeting substrates. *Journal of Biological Chemistry*. 2015;jbc. M114. 624346.

52. Tu C-J, Schuenemann D, Hoffman NE. 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*. 1999;274(38):27219-24.
53. Eichacker L, Henry R. Function of a chloroplast SRP in thylakoid protein export. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*. 2001;1541(1-2):120-34.
54. Van Der Walle C. *Peptide and protein delivery*: Academic Press; 2011.
55. Kityk R, Kopp J, Sinning I, Mayer MP. Structure and dynamics of the ATP-bound open conformation of Hsp70 chaperones. *Molecular cell*. 2012;48(6):863-74.
56. Zhu X, Zhao X, Burkholder WF, Gragerov A, Ogata CM, Gottesman ME, et al. Structural analysis of substrate binding by the molecular chaperone DnaK. *Science*. 1996;272(5268):1606-14.
57. Flaherty KM, DeLuca-Flaherty C, McKay DB. Three-dimensional structure of the ATPase fragment of a 70K heat-shock cognate protein. *Nature*. 1990;346(6285):623.
58. Mayer M, Bukau B. Hsp70 chaperones: cellular functions and molecular mechanism. *Cellular and molecular life sciences*. 2005;62(6):670.
59. Waters ER. The evolution, function, structure, and expression of the plant sHSPs. *Journal of experimental botany*. 2012;64(2):391-403.
60. Ehrnsperger M, Lilie H, Gaestel M, Buchner J. The dynamics of Hsp25 quaternary structure structure and function of different oligomeric species. *Journal of Biological Chemistry*. 1999;274(21):14867-74.

#### **IV. Structural Propensity in the C-terminal Domain of Albino3**

##### **Abstract**

The c terminus of the Albino3 (cAlb) translocase in chloroplasts is a region fundamental to the integration of LHCPs into the thylakoid membrane in cooperation with the cpSRP.

Albino3 (Alb3) is an integral membrane protein containing five transmembrane helices in an N-in, C-out orientation in the thylakoid membrane. cAlb is responsible for recruiting the cpSRP43 subunit of the transit complex to the thylakoid membrane for successful integration of LHCPs.

cAlb has previously been described as intrinsically disordered. In more recent developments, intrinsically disordered proteins have been shown to serve as hub proteins to regulate key cellular functions. While our findings show that cAlb is predominately disordered, we have discovered a region in this protein which has a high propensity to form local ordered structures. This region may be crucial to its functionality in the chloroplast. Our results show a possible helix in the N-terminal region of cAlb. It was first identified by secondary structure sequence analysis using multiple prediction software. Single point mutations which would contribute to the disruption of structure were designed within this region, followed by characterization of the residual helix structure. Results obtained from Far-UV-Circular Dichroism, intrinsic fluorescence, and trypsin digestion reveal that the designed mutations disrupt the residual helix structure in the c-terminal domain of Alb3. Inter-residue distances and energetics were acquired from smFRET analysis of native and denatured cAlb. The results show that cAlb is unfolding upon titration with guanidinium hydrochloride. Additionally, computational analysis corroborates the presence of the local helix structure in the N-terminal region of cAlb. This region may prove to be important for the recognition and binding of cAlb to the cpSRP43 subunit of cpSRP.

## Introduction

The evolutionarily conserved YidC/Oxa1/Alb3 protein family mediates assembly and insertion of membrane proteins vital to the processes of energy production in bacteria, mitochondria and chloroplasts, respectively. Located in the inner membrane of bacteria, mitochondria, and the thylakoid membrane of chloroplasts, these insertases have similar functions with the exception of some species-specific differences (1). Each homolog shares a conserved hydrophobic region comprised of five transmembrane domains (1, 2). However, YidC, present in gram-negative bacteria, contains six trans membrane segments and a large periplasmic domain between TMs 1 and 2 (2). Alb3 is required for the post-translational integration of LHCPs into the thylakoid membrane. Nuclear-encoded LHCPs are imported from the cytoplasm into the stromal compartment of the chloroplast via a chloroplast targeting sequence. Upon entrance into the stroma, the targeting sequence is cleaved and LHCP becomes part of a soluble transit complex with the cpSRP subunits. Post-translational integration of LHCPs and co-translational integration of other chloroplast membrane proteins is carried out by the Alb3 insertase once they reach cAlb which is the final receptor in the cpSRP pathway. cAlb is an intrinsically disordered region of Alb3 which protrudes out into the stroma and facilitates association with the cpSRP at the thylakoid membrane through recognition and association with the cpSRP43 subunit. cAlb, upon binding to cpSRP43, gains a more compact structure as demonstrated in circular dichroism experiments (3). In addition, cAlb binding to cpSRP43 activates GTP hydrolysis by cpSRP54 and cpFtsY which is critical for the release of LHCP (4). An interaction between LHCP and cAlb has been observed in *in vitro* binding experiments (5). It was also concluded that a cpSRP-LHCP complex binds more efficiently to cAlb than cpSRP alone which suggests that a cpSRP loaded with LHCP is important for the interaction with cAlb



and may prevent the Alb3 protein from being blocked by a cpSRP which is not containing cargo (5). cpSRP43 binds cAlb whereas cpSRP54 does not, again confirming that binding between cpSRP43 and cAlb is critical for interaction of the complex at the thylakoid membrane (6). Many studies have investigated the binding interaction between cAlb and cpSRP43. The ankyrin repeat regions and CD2CD3 of cpSRP43 have been implicated as important areas for association with cAlb (4, 7). Sequence alignments of Albino3 from different plant species have revealed four conserved, positively charged motifs (I-IV) in the stromal c-terminal region which have been investigated for their contribution to the binding event between cpSRP43 and cAlb (3). Key regions in both proteins have been identified as critical or necessary for the interaction between the two proteins however when the data is compiled there appears to be some conflicting reports. ITC binding experiments showed that full length cAlb had a  $K_d$  of  $9.7\mu\text{M}$  and cAlb lacking motif I had a  $K_d$  of  $11.4\mu\text{M}$  upon binding cpSRP43, with a small change in dissociation constants these researchers suggest that motif I is not important in the binding interaction (3). A synthetic motif III peptide did not bind to cpSRP43 and further truncations of cAlb from both the c and n-terminus showed lowered binding affinities leading to the recommendation that motifs II and IV are required for the interaction with cpSRP43 (3). Examination of motifs II and IV reveal AKRS and SKRS sequences which are similar to ARK(S/T) sequences present in histone H3 tails suggesting that there would likely be affinity between these motifs and the chromodomains of cpSRP43 (3, 8). As lysine and arginine were independently mutated to alanine in motifs II and IV and bound to cpSRP43 this led to reduced affinity (3). When the same mutations were applied concurrently, binding was lost as observed by ITC suggesting that the positively charged side chains in these areas of cAlb are important for binding to cpSRP43 (3). Binding affinities between the ankyrin (ANK) (1-4) regions and cAlb

were lower than observed for full length cpSRP43 and cAlb although it was still in the nanomolar range suggesting that the ANK regions are important for binding or stabilizing the interaction between the two proteins (7). In an *in planta* study, a cAlb which was lacking motifs III and IV was compared to the wild type under normal light growth conditions and the result was that the truncated cAlb lost 80% functionality of that of the wild type (9). This was possibly due to the instability and faster degradation of the truncated cAlb or its inability to interact with other proteins. When grown under continuous low light conditions, the truncated cAlb (lacking motifs II and IV) was able to accumulate and integrate LHCPs, with only a minor reduction compared to wild type, therefore the researchers concluded that motifs I, II and TM5 were the areas of Alb3 important for sufficient interaction with cpSRP43 and insertion of LHCP into the membrane (9). There still exists some debate on which domains (ANK(1-4) or chromodomains) of cpSRP43 are critical for binding to cAlb, there also seems to be further debate over the precise area within cAlb that is essential to the binding event. Given conflicting reports on the exact regions of cAlb which are critical for binding to cpSRP43 and a lack of binding details at the residual level, ongoing investigation into this interaction is important for a better understanding of targeting proficiency in the cpSRP pathway.

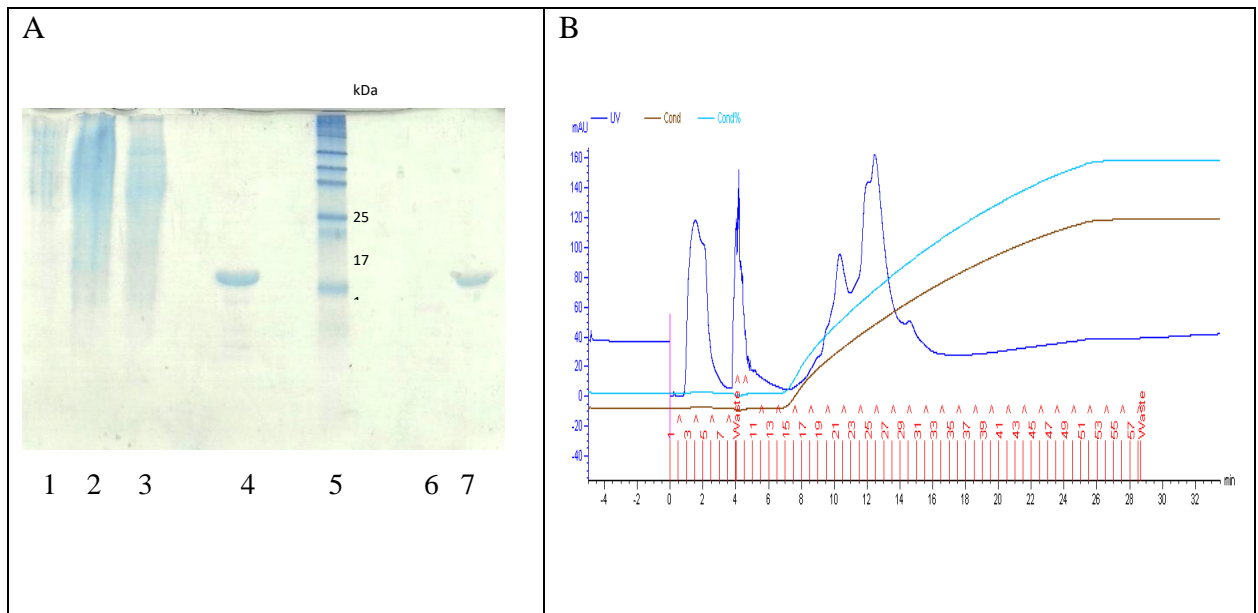
Four selected amino acids, in the region predicted to contain a residual helix, were mutated to glycine in order to disrupt the predicted structure. While glycine is common in transmembrane proteins it is also known to destabilize helices in water- soluble proteins (10-13). Characteristic analysis of these designed mutations showed a decrease in residual structure within the proposed helical region. In the smFRET experiments, cysteine residues which encompassed the region of interest were labelled and analyzed to display an inter-residue distance indicative of the presence of an ordered structure. The computational studies again agree

with a high propensity of structure in the N-terminal of cAlb. This study combines biophysical techniques with computational methods to provide a thorough examination of cAlb to reveal a region of residual structure within the c-terminal domain of Alb3 which may prove to be beneficial in the overall elucidation of the cpSRP pathway.

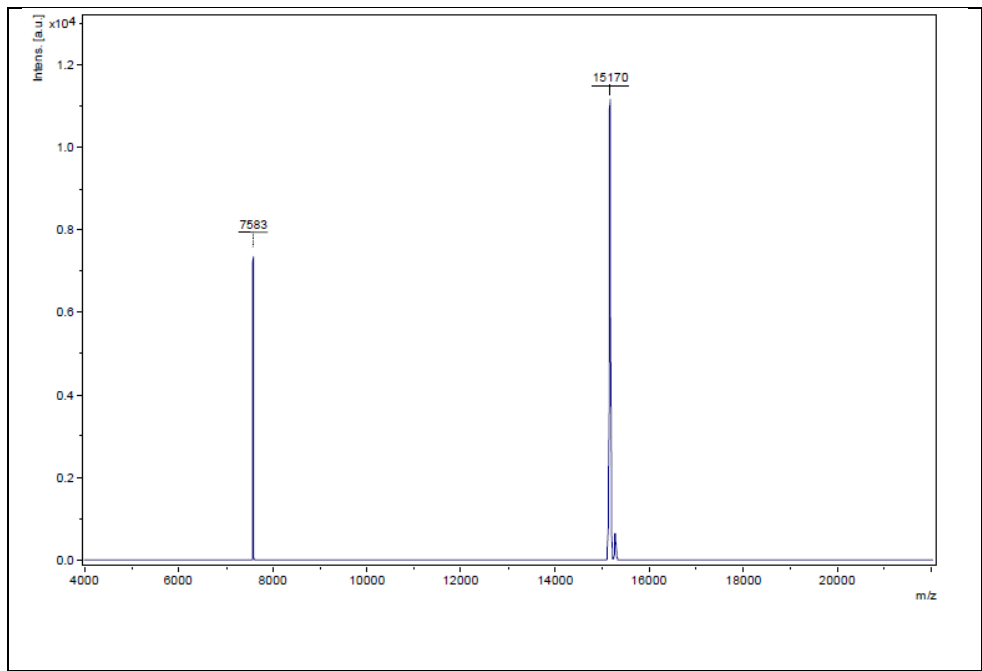
## **Materials and Methods**

### **Cloning, Expression and Purification of His-cAlb**

N-terminally His-tagged c-term of PPF1 (Alb3 in *Pisum sativum*) was cloned into pQE-80L vector and transformed in BL21 star cells (Obtained from Dr. Henry's lab). Protein production was induced with 1 mM isopropyl-1-thio-D-galactopyranoside at an  $A_{600}$  of 0.6, the cells were harvested after 3.5 hours at 37 °C. For the purification of His-tagged cAlb, the cells were resuspended in buffer (50mM Hepes, 150mM NaCl, 10% (w/v) glycerol, 0.02% 1-thioglycerol, pH 7.5). The cells were lysed using ultrasonication and the cell lysate was separated using ultracentrifugation at 19,000 rpm. The supernatant was applied onto a Ni<sup>2+</sup> Sepharose column (GE Healthcare), washed with resuspension buffer and eluted at 300 mM imidazole (IMZ) using a stepwise IMZ gradient of 10, 20, 50, 100, 300 and 500 mM IMZ concentrations. cAlb migrates slightly higher than expected; between the 25 and 17 kDa (Fig.1A) The molecular weight and purity was confirmed by mass spectroscopy (M.W.=15.17 kDa) (Fig. 2). The eluted protein was further purified to homogeneity using a Mono S ion exchange column (Fig. 1A and B).



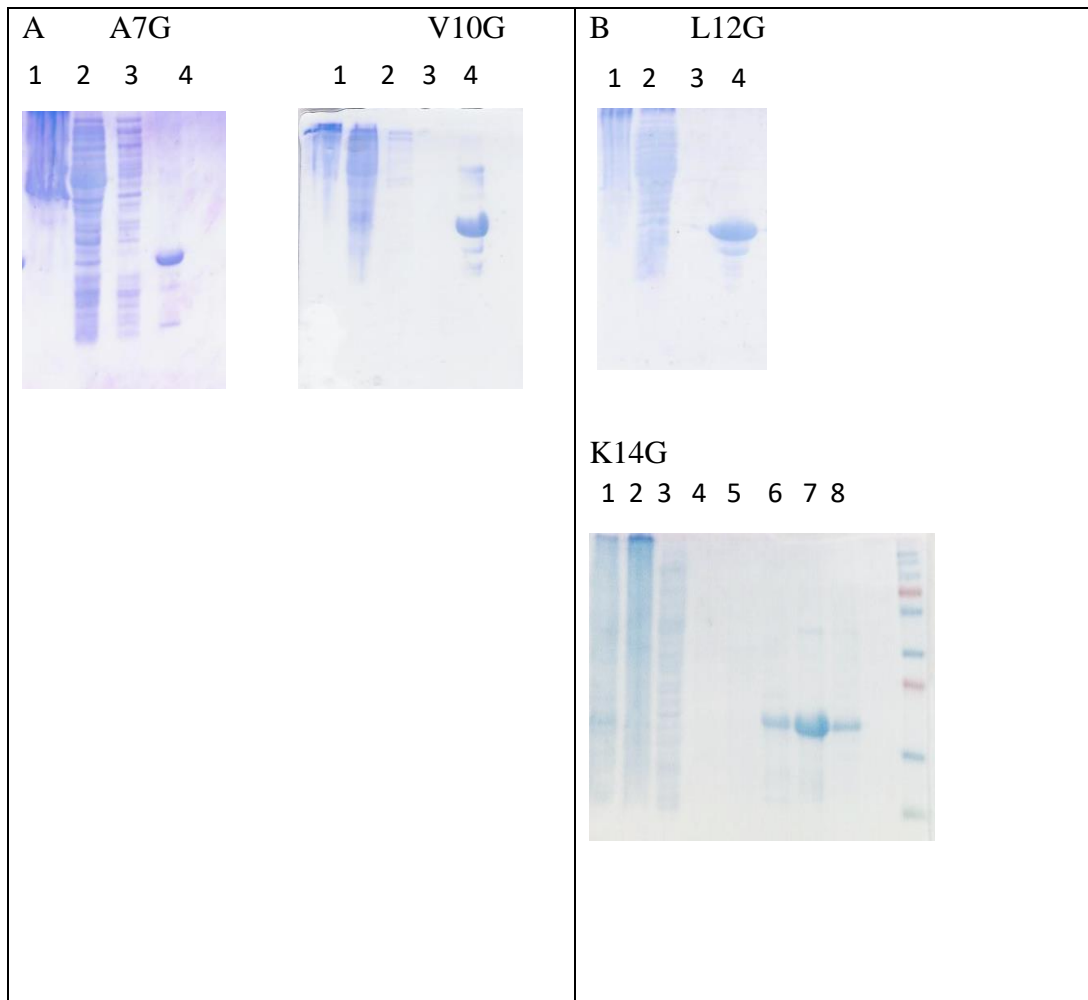
**Fig.1: SDS-PAGE depicting the purification of wt-cAlb (Panel-A) and the Mono S profile of wt-cAlb (Panel-B), Lane 1= Pellet, Lane 2= Supernatant, Lane 3= Unbound, Lane 4= 300 mM IMZ, Lane 5= Protein Marker, Lane 6= Fractions 19-22 from Mono S, Lane 7= Fractions 24-27 from Mono S**



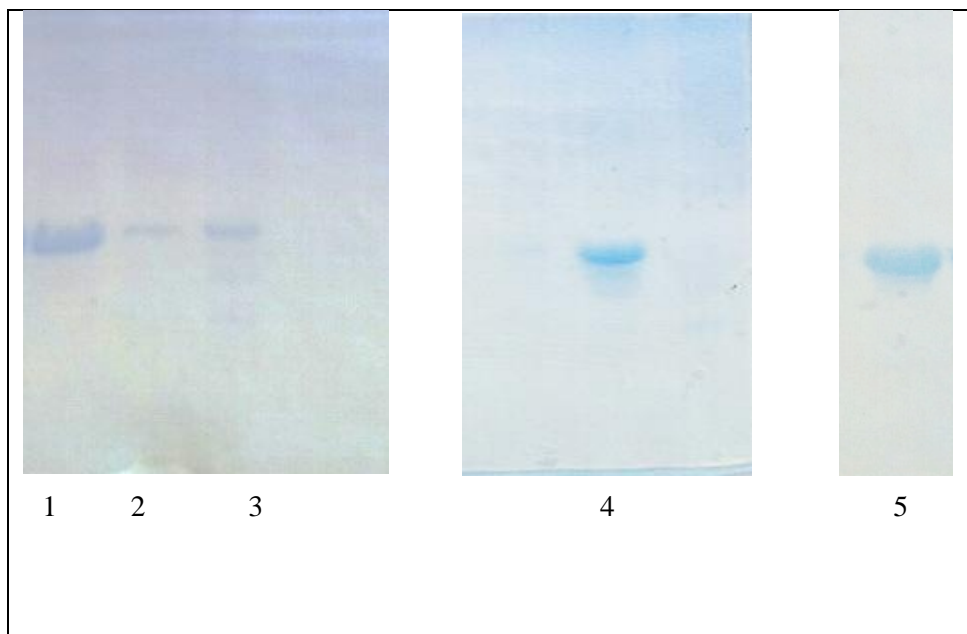
**Fig. 2: Mass spectroscopy of cAlb.**

### **Construction and purification of His-cAlb mutants**

Based on predictions made with secondary structure and disorder predicting software, single point mutants of cAlb (A21G, V24G, L26G and K28G) were cloned into pQE-80L bacterial expression vector which was used as a template for site-directed mutagenesis. Primers were designed using an Agilent primer design program (IDT DNA Inc.). Site-directed mutagenesis was performed using a QuikChange II XL kit followed by polymerase chain reaction (PCR). The plasmid was then transformed into DH5 $\alpha$  competent cells. Upon confirmation of the correct plasmid sequences, each mutant was overexpressed in BL-21-Star Escherichia coli cells cultured in lysogeny broth (LB) at 37 °C with agitation at 250 rpm. Overexpressed cells were lysed using ultrasonication and separated from the cell debris using ultra centrifugation at 19,000 rpm. cAlb mutants were purified as described above for the wild type cAlb. The purity of the cAlb mutant proteins were assessed by 15% sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) from Ni<sup>2+</sup> column (Fig. 3A and B) and subsequent purification from Mono S (Fig.4). Protein bands were visualized by staining the gels with Coomassie brilliant blue R-250 dye.



**Fig. 3: SDS-PAGE gels of A7G: Lane 1= Pellet, Lane 2= Supernatant, Lane 3= Unbound, Lane 4= 300 mM IMZ, V10G: Lane 1= Pellet, Lane 2= Supernatant, Lane 3= Unbound, Lane 4= 300 mM IMZ (Panel-A). SDS-PAGE gels of L12G: Lane 1= Pellet, Lane 2= Supernatant, Lane 3= Unbound, Lane 4= 300 mM IMZ, K14G: Lane 1= Pellet, Lane 2= Supernatant, Lane 3= Unbound, Lane 4= 10 mM IMZ, Lane 5= 50 mM IMZ, Lane 6 = 100 mM IMZ, Lane 7= 300 mM IMZ, Lane 8= 500 mM IMZ (Panel-B).**



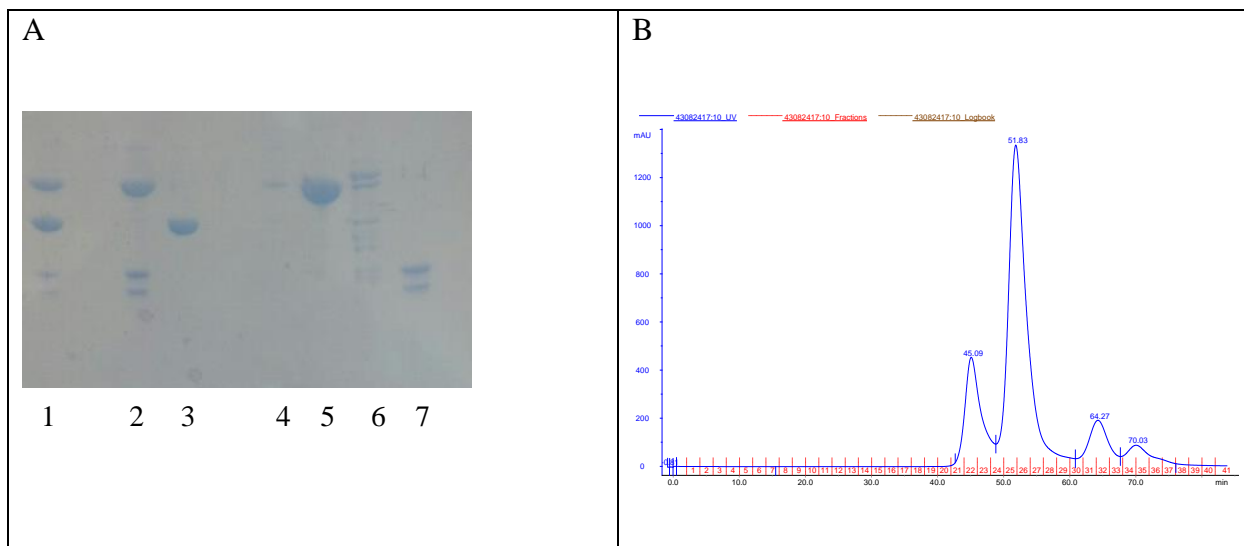
**Fig. 4: SDS-PAGE gels of pure cAlb and mutants. Lane 1= cAlb, Lane 2= L12G, Lane 3= K14G, Lane 4= A7G, Lane 5= V10G.**

**Ovalbumin** (A5503) was purchased from Sigma (St. Louis, MO).

### **Expression and Purification of Recombinant cpSRP43**

For expression of cpSRP43, BL-21 Star cells containing pGEX-6P-2-cpSRP43 were grown to an Optical Density of 0.6–0.8 at Abs<sub>600</sub> and incubated with 1 mM isopropyl β-D-thiogalactoside for 3.5 hours. Cells were harvested and sonicated for 30 on/off cycles (20 seconds) using 10 W output. GST-cpSRP43 fusion proteins were bound to glutathione Sepharose then washed extensively with equilibration buffer (2.7 mM KCl, 1.8 mM KH<sub>2</sub>PO<sub>4</sub>, 15 mM Na<sub>2</sub>HPO<sub>4</sub>, 137 mM NaCl, pH 7.2). Elution of fusion protein was done using 10 mM L-Glutathione with subsequent exchange into cleavage buffer (50 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA, 1 mM dithiothreitol, pH 7.0). Overnight cleavage in solution was setup at 4°C on a rocker with 10 units of PreScission Protease per liter of cells for 16 hours. The cleavage product was passed back onto glutathione Sepharose to separate out the GST tag and the cleaved cpSRP43 was

further purified by gel filtration chromatography (Fig.5B). Purity of proteins were visualized using 15% sodium dodecyl sulfate poly acrylamide gel electrophoresis (SDS-PAGE) followed by staining with Coomassie brilliant blue G-250 dye (Fig. 5A).



**Fig. 5: SDS-PAGE of purification of cpSRP43 (Panel-A), Lane 1= Cleavage product, Lane 2=Cleaved cpSRP43, preload to SEC, Lane 3= GST tag eluted in 10 mM L-Glutathione, Lanes 4-7= Peaks 1-4 of the SEC profile, Pure cpSRP43 is eluted in Peak 2 (Lane 5). SEC profile (Panel-B).**

### Circular Dichroism

Circular dichroism (CD) measurements were performed on a Jasco J-1500 CD spectrometer equipped with a variable temperature cell holder. Conformational changes in the secondary structure of cAlb were monitored in the Far-UV region between 190 to 250 nm with a protein concentration of 20  $\mu$ M in a quartz cuvette with a path length of 1 mm. The scanning speed, band width and data pitch were set to 50 nm/min, 1.00 nm and 0.1 nm, respectively. Three scans were taken (within a 1000 HT voltage range) and averaged to obtain the CD spectra.



### **Fluorescence spectrometry and 8-Anilino-1-naphthalenesulfonic acid (ANS) binding assay**

Fluorescence spectra were collected using a Fluorescence Spectrophotometer F-2500 (Hitachi). Initial intrinsic fluorescence readings of cAlb were taken at an excitation of 280nm and the emission intensity was recorded within a range from 300 nm to 450 nm. ANS binding assay measurements were made using a Fluorescence Spectrophotometer F-2500 (Hitachi) with a slit width set to 2.5 nm. A 20 mM ANS stock was prepared and then diluted to a 5 mM stock. The cAlb sample was quantified using Bradford estimation. Protein concentrations of 20  $\mu$ M in buffer (50 mM Hepes, 150 mM NaCl, pH 7.2) was placed in a quartz cuvette. Titrations using the 5 mM ANS stock were made by adding 2  $\mu$ ls, mixing and incubation for 2 minutes preceding each reading at 25°C. Fluorescence intensity was determined with an excitation at 380 nm and emission intensity was recorded at 510 nm.

### **Limited Trypsin Digestion**

Limited trypsin digestion of cAlb and cAlb mutants was performed in Hepes buffer . The initial reaction tube contained 500  $\mu$ g of protein and 0.5  $\mu$ g of enzyme. The trypsin-containing samples were incubated at room temperature (25 °C). Digested samples were removed every 2 minutes for up to 15 minutes and then the final sample was taken at 20 minutes. The reaction was stopped via the addition of 10% trichloroacetic acid to each 100  $\mu$ l aliquot. TCA precipitated samples were resolved on a 15% sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) gel and subsequently stained using Coomassie Blue. UN-ScanIT software (Silk Scientific Inc.) was applied to identify the percent of digestion based on the size of the parent band remaining at each time point.

## **Nuclear magnetic spectroscopy**

Heteronuclear (2D  $^1\text{H}$ - $^{15}\text{N}$ ) single quantum coherence (HSQC) spectroscopy was carried out using a Bruker 500 MHz NMR. Expression of the cAlb sample was done in M9 minimal media containing  $^{15}\text{NH}_4\text{Cl}$  resulting in  $^{15}\text{N}$  labeling. The experiment was acquired at 25 °C at a protein concentration of 300  $\mu\text{M}$ . The sample was prepared in 90% H<sub>2</sub>O 10% D<sub>2</sub>O solution containing 50 mM Hepes buffer with 150 mM NaCl (pH 7.5). Experiments were conducted using a Bruker Avance 700 MHz NMR spectrometer set at 298 K. NMR data was analyzed using Sparky 3.114 software (T.D. Goddard and D.G. Kneller, SPARKY 3.114, University of California, San Francisco).

## **Small angle x-ray scattering**

SAXS data for cAlb was acquired at the Cornell High Energy Synchrotron Source (CHESS)(14). cAlb was purified to homogeneity using two chromatographic steps. The supernatant from centrifugation of the cell lysate was first applied to a nickel sepharose column and the purest, largest fraction of cAlb was eluted at 250 mM IMZ. This fraction was subsequently concentrated to 1 ml and applied to a Mono S column using FPLC with a salt gradient ranging from 150 mM NaCl to 500 mM NaCl. The pure fraction of cAlb was eluted at approximately 300 mM NaCl. This purification process was repeated six times to obtain a 300  $\mu\text{M}$  concentrated sample of cAlb. This sample was shipped on dry ice to Cornell University. This entire process was repeated three times because the first two attempts did not yield reliable data. It is possible that the cAlb preparations degraded during delivery or degraded during the SAXS experiment due to radiation damage. A series at different concentrations (1 mg/ml to 5 mg/ml) was collected using ten one-second exposures to a  $q$  range of 0.5  $\text{\AA}^{-1}$ . Buffer scattering data was collected

before and after exposures of the protein. Data was analyzed using the RAW (15) software then visualized using ATSAS Primus (16).

### **Isothermal titration calorimetry**

The binding affinity of cpSRP43 to cAlb and the A21G mutant was measured using isothermal titration calorimetry (ITC). Protein samples were prepared with 2.7mM KCl, 1.8mM KH<sub>2</sub>PO<sub>4</sub>, 15mM Na<sub>2</sub>HPO<sub>4</sub>, 137mM NaCl, pH 7.2 in 10% glycerol and degassed prior to loading. A 1:10 ratio of cpSRP43 to cAlb and A21G (30μM cpSRP43: 300μM cAlb and A21G) was applied on a MicroCal iTC200 (Malvern Inc.) in which cAlb and A21G were each titrated into cpSRP43 samples. A series of 30 titrations were carried out at 25 °C with a stir speed of 1000 rpm and excess heats which could have come from the buffer were appropriately subtracted out and binding affinities were calculated.

### **Size-Exclusion Chromatography (SEC)**

cpSRP43, cAlb and the single point mutants were over-expressed and purified as described above. The samples of cpSRP43 bound to cAlb and mutants used were at 30 and 15 μM, respectively. The samples were incubated on a rocker for an hour at 4°C prior to being loaded to a Superdex 75 16/600 column (GE Healthcare, Pittsburgh, PA) equilibrated in 2.7mM KCl, 1.8mM KH<sub>2</sub>PO<sub>4</sub>, 15mM Na<sub>2</sub>HPO<sub>4</sub>, 137mM NaCl, pH 7.2 on an AKTA FPLC and ran at a flow rate of 1 milliliter per minute. The elution volumes were monitored by absorbance at 280 nm. Fractions were collected and analyzed by SDS-PAGE.

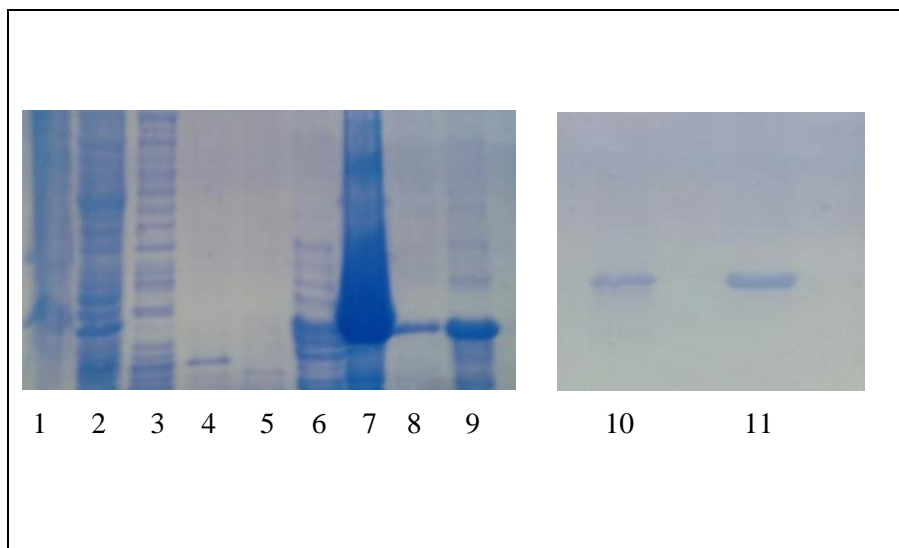
## **Predictions of structured regions within cAlb**

Structural predictions from the amino acid sequence of cAlb were made using six different web servers: PONDR, DISOPRED, PrDOS2, PSIPred, Dynamine and Foldunfold. PONDR is a feedforward neural network which takes sequence attributes and averages them and these values are used to make predictions (17). DISOPRED uses a knowledge-based method by initially running a PSI-BLAST search to predict dynamically disordered regions from the primary sequence and the output is a residual probability estimate of disorder (18, 19). PSIPRED secondary structure predictions use the same inputs as DISOPRED to provide further structural information (20). PrDOS2 is a server which employs an SVM (support vector machine) algorithm which combines a template-based predictor with sequence profiles to predict natively disordered regions of a protein from the primary sequence input (21). PrDOS2 output is a measure of residual disorder probability (21). DynaMine is a server used to quantify backbone dynamics movements on the residual level by using only the protein sequence of interest as input and returns predictions on different structural regions and also identifies disordered regions (22, 23). Foldunfold is a method which locates the mean packing density of residues in order to predict disordered regions (24). The primary sequence of cAlb was uploaded to each server and the output was collected and analyzed to find a region of consensus between all algorithms used. An ordered region was predicted in the area ranging from residues 12-45 for all the software. A helix was predicted by PSIPred within residues 15-29. The region of predicted structure common to all algorithms (residues 15-39) was used for biophysical and computational studies. Estimation of protein secondary structure derived from circular dichroism spectra of cAlb and the single point mutants was analyzed using CONTIN, CDSSTR, and SELCON available with the CDPro program package online. The circular dichroism data points from 190-250 nm for

cAlb and each mutant was input into each of the CDPro algorithms and the output gave fractional estimations of alpha-helical content (regular and distorted), beta-sheet content (regular and distorted), turns and unordered regions.

### **Cloning, expression and purification of double cysteine cAlb mutant**

An endogenous cysteine residue was kept at position 14 in the original sequence and an additional cysteine residue was introduced at position 52 (S52C). Primers were designed using an Agilent design program and ordered from IDT DNA Inc. Site-directed mutagenesis was performed using a QuikChange II XL kit. The primer-encoded point mutation was introduced into the coding sequence of HIS-cAlb by polymerase chain reaction (PCR). The plasmid was then transformed into DH5 $\alpha$  competent cells and the construct was sequence verified by Molecular Resource Laboratory, University of Arkansas for Medical Sciences, Little Rock. Upon confirmation of the plasmid sequence, the mutant protein was overexpressed in BL-21 Star Escherichia coli cells cultured in lysogeny broth (LB) at 37 °C with agitation at 250 rpm. Overexpression, purification and quantification methods were the same as described above for HIS-cAlb except for the second chromatographic step was carried out on a size exclusion column. Purity of S52C is depicted by SDS-PAGE gels stained with Coomassie brilliant blue G-250 dye (Fig.6).



**Fig. 6: SDS-PAGE of S52C double cAlb mutant. Ni<sup>2+</sup> column: Lane 1= Pellet, Lane 2= Supernatant, Lane 3= Unbound, Lane 4= 10 mM IMZ, Lane 5= 50 mM IMZ, Lane 6= 100 mM IMZ, Lane 7= 300 mM IMZ, Lane 8= 500 mM IMZ, Lane 9= cAlb marker (300 mM IMZ). Lanes 10= S52C collected from SEC, Lane 11= cAlb marker.**

### Fluorescent Labeling

The double cysteine cAlb mutants and wild type were both labeled under the same conditions. In pH 7.5 PBS, 100  $\mu$ M cAlb was incubated with 50  $\mu$ M Alexa488 C5 maleimide and 1 mM Alexa594 C5 maleimide for 30 minutes at room temperature. Tris (2-carboxyethyl) phosphine hydrochloride (TCEP) treatment is used to inhibit the formation of disulfide bonds. To remove the excess free dye, the labeling mixture was separated on a Biogel P-6DG gel filtration column.

### Single-molecule FRET

Labeled protein preparations were diluted to  $\sim$ 300 pM for the smFRET experiment. The dilution buffer for titration experiments contained the appropriate concentration of guanidinium hydrochloride in pH 7.5 PBS buffer. Using a MicroTime 200 microscope, the protein solutions were then exposed to 485 nm continuous wave excitation for 3-minute intervals, after which the sample aliquot was changed. Fluorescence was collected using a 60X water immersion objective with a numerical aperture of 1.3. A 100  $\mu$ m pinhole was used to block out of focus fluorescence.

Collected photons were then passed to a FF562-Di03 dichroic mirror to split the fluorescence into donor and acceptor channels based on wavelength. Further, each channel had a band-pass filter to restrict photons to the respective fluorescent ranges of the donor and acceptor fluorophores. These filters were 520/35 and 620/60 for the donor and acceptor channel respectively.

### **smFRET analysis**

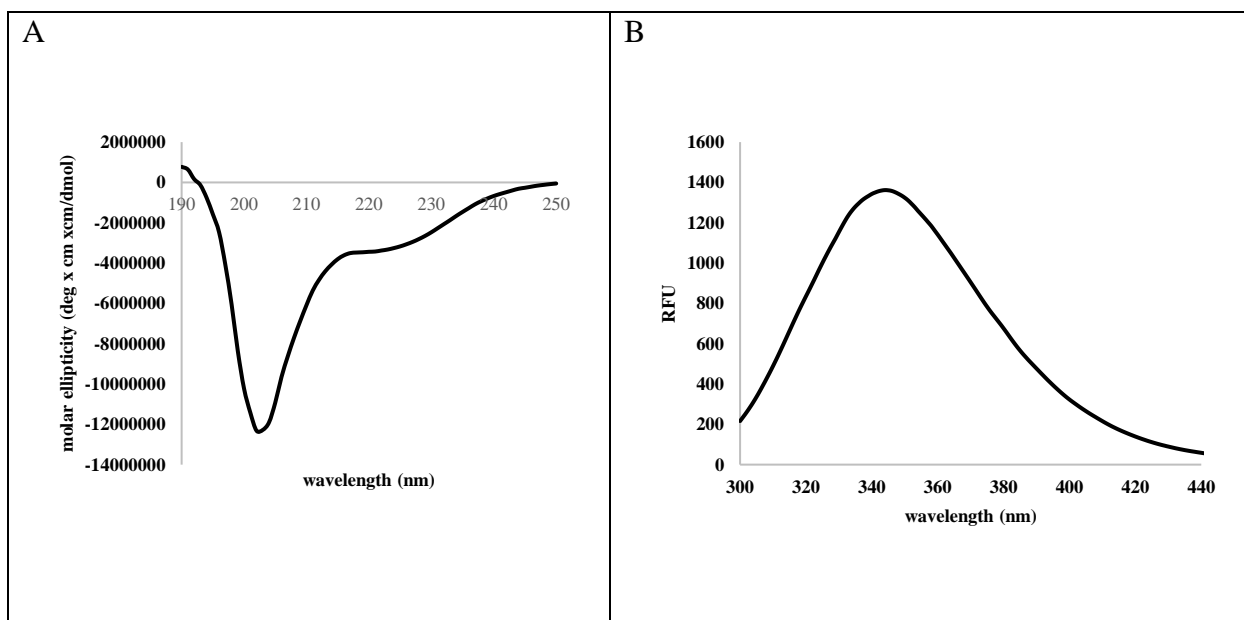
Single molecule traces were binned into 1 ms intervals and analyzed using a home-written analysis program. First, the donor channel was corrected using the  $\gamma$ -factor that was determined experimentally to be 0.6. This gamma-corrected donor value was added to the acceptor channel value and any burst meeting the threshold of 20 counts/ms was collected. Ratiometric FRET efficiencies were calculated for all bursts using the gamma factor and the donor fluorescence leak into the acceptor channel which was found to be 7%. All FRET efficiencies for a sample were binned to produce a smFRET histogram.

## **Results and Discussion**

### ***cAlb secondary structure and native conformation***

Far UV-CD analysis is a commonly used technique to determine the secondary structure within a protein of interest by investigating the absorption bands of optically active chiral molecules (25, 26). The different structural properties of proteins have characteristic CD spectra (27). Proteins containing predominate alpha-helices will have negative bands at 222 and 208 nm and will typically have a positive band at approximately 194 nm (27). Typical beta-sheet proteins show negative bands at 218 nm and positive bands at 195 nm (27). Far UV-CD spectra of IDPs show a negative band around 195 nm however the CD spectra of IDPs can fluctuate depending on the

chemical conditions under which the protein is analyzed (28, 29). Far UV-CD of cAlb reveals a negative peak at 200 nm, which is characteristic of a random coil. There is an additional broad negative peak at approximately 225 nm, which indicates there may be some local structure present (Fig.7A). Intrinsic fluorescence measurements of proteins are made by exciting the protein sample using an ultraviolet light wavelength at 280 nm observing the maximum emission value within a range of 300 nm to 400 nm (30). The emission wavelength depends on the polarity of the microenvironment surrounding the tryptophan residue(s) (30). cAlb has one endogenous tryptophan residue located at the N-terminal region. The intrinsic fluorescence spectra of native cAlb reveals an emission maximum at 350 nm which suggests that the tryptophan residue is exposed to the solvent and cAlb is predominately unfolded in its native state (Fig.7B).

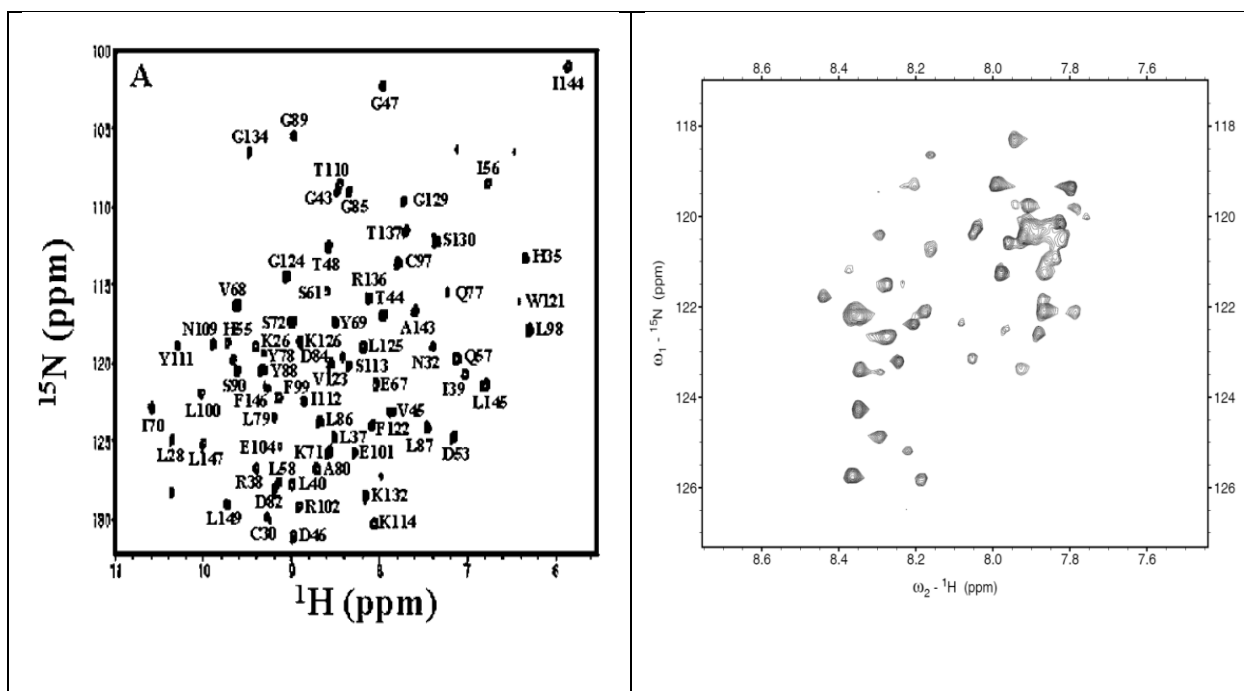


**Fig. 7: (Panel-A) Far UV-CD spectra of cAlb. (Panel-B) Intrinsic fluorescence spectra of cAlb.**



### ***Heteronuclear single quantum coherence of cAlb is characteristic of an IDP***

Two dimensional  $^1\text{H}$   $^{15}\text{N}$  HSQC is a commonly used NMR technique for determining structural details at the residual level for  $^{15}\text{N}$  isotope enriched proteins. Acquired HSCQ spectra displays the backbone conformation of proteins.  $^1\text{H}$   $^{15}\text{N}$  HSQC works by showing the hydrogens that are directly attached to the nitrogen atoms (31, 32). Each crosspeak in a spectrum is representative of the conformational position of each amino acid in the protein by giving the chemical shift related to the proton and nitrogen couple (31, 32). The proton attached to the nitrogen is probed because it has a greater equilibrium magnetization, meaning that it is better aligned to the static magnetic field and therefore can create a stronger signal (32). Crosspeaks for structured proteins occur over a range of  $\sim 7.5$  to  $\sim 10$  ppm in the proton dimension (32). However, HSQC spectra of IDPs are characterized by narrow spectra ranges. Crosspeaks for disordered proteins typically occur from  $\sim 8.0$  to  $\sim 8.6$  ppm which coincide with the range for amide protons in the central residue of unstructured tripeptides.  $^1\text{H}$   $^{15}\text{N}$  HSQC spectra for hFGF1 and cAlb is compared below (Fig. 8). The structure of wt-hFGF1 has been well characterized complete with assigned resonances (33, 34). hFGF1 is a 15 kDa protein comprised of twelve beta-strands arranged in an antiparallel fashion into a beta-barrel structure (35). The  $^1\text{H}$   $^{15}\text{N}$  HSQC of hFGF1 shows well-dispersed crosspeaks along a spectral range of 6 to 11 ppm (Fig. 8A). The chemical shift dispersion is much higher in folded proteins which indicates that the nuclei in these proteins are having a stable secondary and/or tertiary structure in their spatiotemporal environment (36). The  $^1\text{H}$   $^{15}\text{N}$  HSQC of cAlb displays clustered peaks in a smaller spectral window of 7.8 to 8.4 ppm which is characteristic of unfolded proteins (Fig.8B).

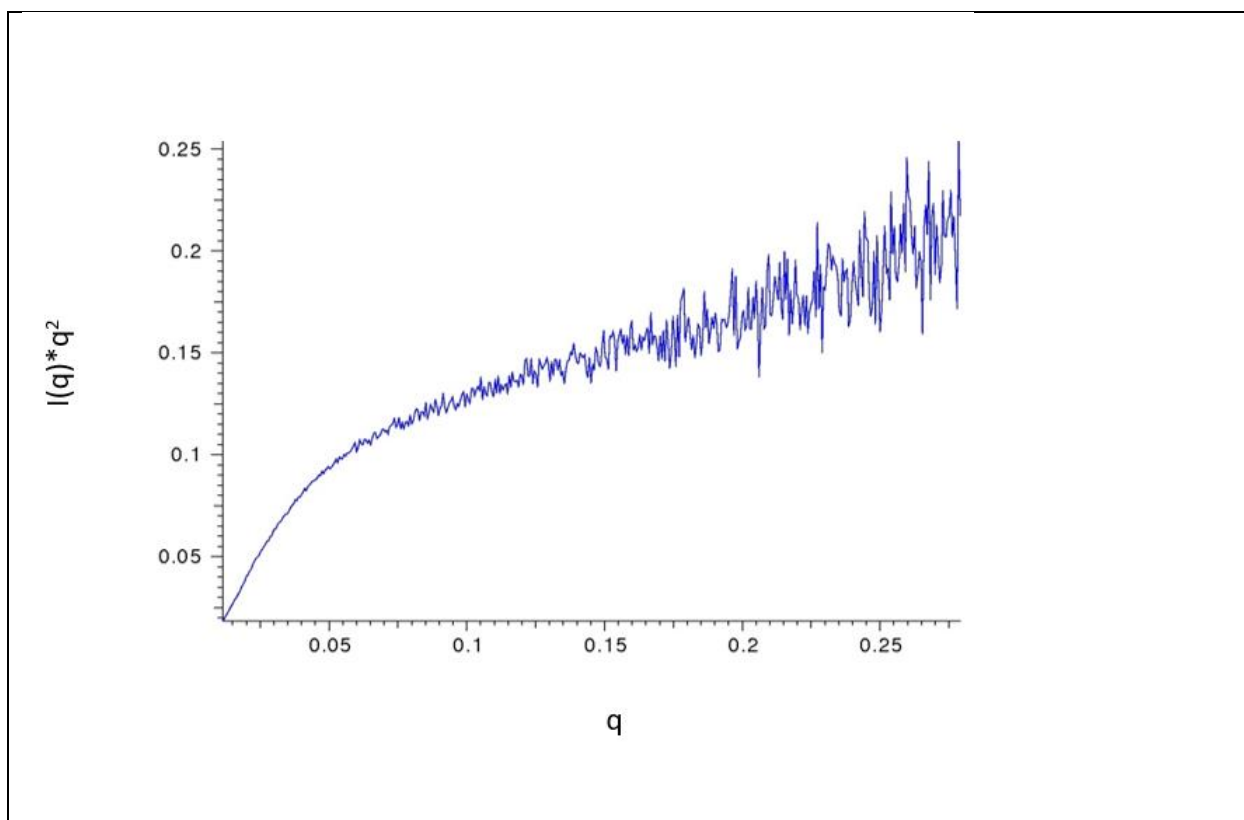


**Fig. 8:**  $^1\text{H}$   $^{15}\text{N}$  HSQC spectra of hFGF1 (Panel-A) and cAlb (Panel-B).

### *Small Angle X-ray Scattering*

Small Angle X-ray scattering (SAXS) is a useful technique for providing low resolution structural information for proteins and other molecules in solution (37). SAXS profiles are a determination of the scattering intensity within a protein as a measure of spatial frequency (37). The scattered light is a result of the difference between electron densities of the protein sample and the solvent, therefore the solvent is subtracted out of the scattering data and what is left represents the overall shape of the protein (38). SAXS data is useful for indicating flexibility in proteins (39). Profiles derived from SAXS experiments can also provide conformational information when compared to corresponding X-ray structures and modeling data (37). The Kratky plot ( $I(q) \cdot q^2$  vs.  $q$ ) is a qualitative check for flexibility and globularity and gives an assessment of disorder within a protein (37). The Kratky plot of a well-structured, globular protein will show a bell-shaped peak at a low  $q$  value and will converge with the  $q$  axis at a high

q value (40). Multi-domain proteins may show additional peaks at low q values but will still converge at the q axis at a high q value (40). However, the Kratky plot of a protein with evident flexibility will not converge at the q axis. A Guinier plot ( $\ln I(q)$  vs.  $q^2$ ) will provide an estimate of the radius of gyration ( $R_g$ ) which can give an idea of the overall size of the protein (41). The calculated  $R_g$  value comes from the slope of the line in the Guinier region and may vary greatly depending on the level of expertise and interpretation by the user (42). The Guinier plot can also provide additional information about how the proteins are behaving upon exposure to the x-ray beam (38). A reliable Guinier plot should be linear which means that the sample is monodisperse (41). SAXS data for the cAlb was collected at the Cornell High Energy Synchrotron Source (CHESS) as described above (14). A series at different concentrations (1 mg/ml to 5 mg/ml) were collected using ten one-second exposures to a q range of  $0.5 \text{ \AA}^{-1}$ . Buffer scattering data was collected before and after exposures of the protein. The data was analyzed using RAW software then visualized by ATSAS Primus as described in the methods. The Guinier plot of cAlb showed a downturned shape at a low angle which is indicative of possible inter-particle repulsion within the sample however the Kratky plot of cAlb shows a slightly increasing upward trajectory which is typical of flexibly extended proteins (Fig.9).



**Fig. 9: Kratky plot of cAlb.**

*Sequence analysis gives a prediction of secondary structural content in cAlb*

Secondary structural elements present in the amino acid sequence of cAlb were determined using the consensus of five different web servers (S1-5). PONDR, PSIPRED, DISOPRED, PrDOS2, Dynamine and FoldUnfold are free web server algorithms used for the prediction of disordered and ordered regions extrapolated from the primary sequences of proteins and peptides as described in the methods. The prediction software collectively suggested an area of local order located near the N-terminal of cAlb. Sequence analysis data for cAlb from the above listed servers is included in the supplementary material. Four single point mutations were engineered in this region for the purpose of decreasing or eliminating any structural propensity held within. The region of interest is highlighted in bold in the sequence below (Fig.10). This region was also the focus of smFRET experiments and computational studies.

10	20	30	40	50	60
MRGSHHHHHH	GSACNNVLST	<b>AQQVWLRKLG</b>	<b>GAKPAVNENA</b>	GGIITAGQAK	RSASKPEKGG
70	80	90	100	110	120
ERFRQLKEEE	KKKLIKALP	VEEVQPLASA	SASNDGSDVE	NNKEQEVTEE	SNTSKVSQEV
130					
QSFSRERRSK	RSKRKPVA				

**Fig. 10: Amino acid sequence of cAlb. The amino acids highlighted in bold depict the common region of predicted structure.**

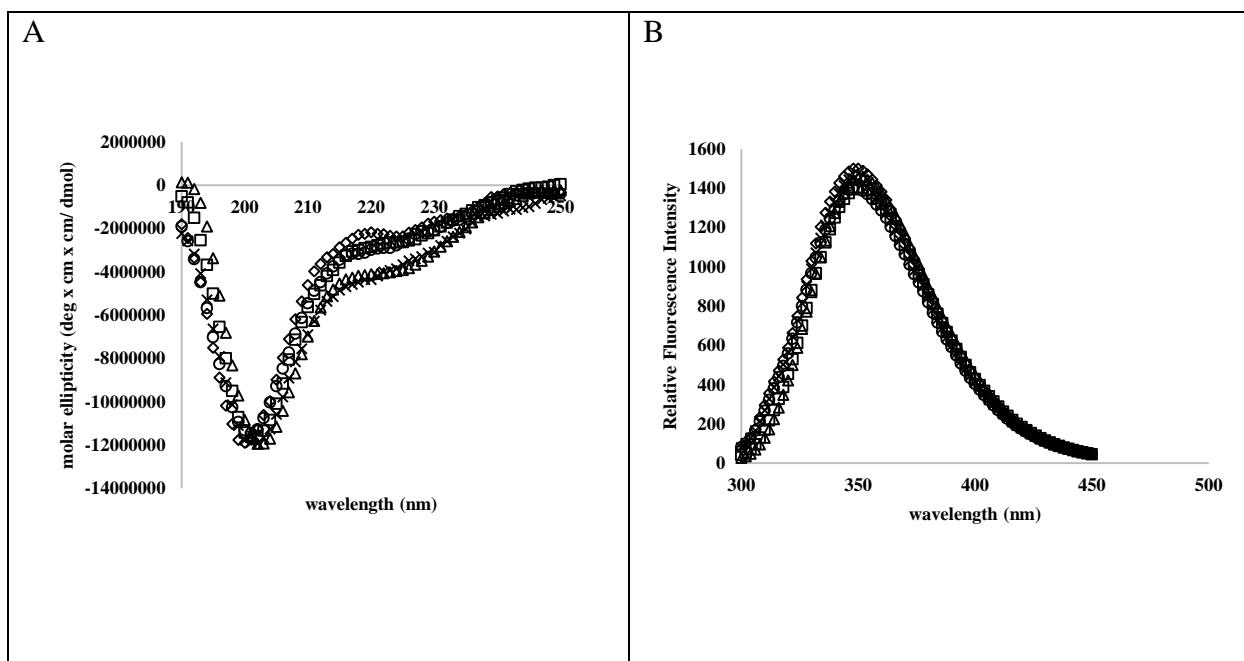
### *A21G, V24G and L26G mutations cause a loss of local secondary structure*

Four single point mutations were made and analyzed one at a time by utilizing far-UV-CD, intrinsic fluorescence and limited trypsin digestion in order to probe for any changes in structure and stability. The four amino acids selected were A21G, V24G, L26G and K28G. All of which are located in the N-terminal region of cAlb. The amino acids were carefully selected for having helix promoting characteristics and mutated to glycine, a known helix breaker (12). According to an experimentally based helix propensity scale using 11 different systems in proteins and peptides developed by Pace et. al, Alanine has the highest helix propensity, the second highest is Leucine, Lysine showed to be the 5<sup>th</sup> highest while Glycine has the lowest helix propensity (12). This is due to the extent of conformational entropy provided by each residue making them have a more or less favorable contribution to the stabilization of the protein (43, 44). The far UV circular dichroism technique is often used to collect reliable information regarding the secondary structure of proteins (45). The far-UV CD spectra of the K28G mutant superimposes well with the wild type cAlb which suggests that this mutation does not cause significant perturbation in the secondary structure of cAlb (Fig.11A.) Comparison of the wild type cAlb with the A21G, V24G, and L26G mutants reveals that they superimpose well with the negative minima at 200

nm which is characteristic of a disordered protein CD spectrum. However, the additional negative shoulder at approximately 225 nm which implies the presence of local secondary structure is receded in the A21G, V24G and L26G mutants when compared to the wild type cAlb (Fig.11A). This indicates a loss of local structure.

***A21G, V24G, L26G, and K28G mutations do not significantly perturb the tertiary structure***

Intrinsic fluorescence is a spectral technique which can be used for monitoring changes in the tertiary structure of proteins. An emission maximum at 350 nm in the wild type cAlb indicates that the native conformation is in a predominately unfolded state which is expected of an intrinsically disordered protein. All four of the single point mutations introduced into this region of the wild type sequence did not appear to perturb the tertiary structure of cAlb as revealed by an insignificant shift in the wavelengths of maximum emission (Fig. 11B). There is a single tryptophan residue located within the region of predicted structure which appears to be unaffected by these mutations.



**Fig. 11: The Far-UV CD spectra overlay (Panel-A) and intrinsic fluorescence (Panel-B) of wt-cAlb and the engineered mutations. wt-cAlb ( $\Delta$ ), A21G ( $\diamond$ ), V24G ( $\circ$ ), L26G ( $\square$ ), K28G ( $\times$ ).**

*Secondary structural estimates made by CD spectra shows a correlation between the mutations to glycine and an increase in disorder*

Estimation of secondary structural content from circular dichroism spectra of wt-cAlb and mutants were analyzed using the online server algorithms in the CDPro program (46, 47) (Fig.12). CONTIN, CDSSTR, and SELCON available with the CDPro software provides an estimation of protein secondary structure derived from circular dichroism spectra. CONTIN, developed by Provencher and Glockner (48), fits the CD spectrum of interest as a linear combination of the reference CD spectra by using a ridge regression procedure (46). CDSSTR, developed by Johnson (49), uses a singular value decomposition (SVD) algorithm to evaluate the content of secondary structure from a CD spectrum. The SELCON method first makes an initial approximation of the structural content of the spectrum of interest by using a basis set then solves the resulting equation using the SVD algorithm and repeats this process until self-

consistency is achieved (50). Data points from a spectral range of 190-250 nm for cAlb and each mutant was input into each of the CDPro algorithms and the output gave fractional estimations of alpha-helical content (regular and distorted), beta-sheet content (regular and distorted), turns and unordered regions. These programs use multiple sets of reference proteins and the reference sets with the largest number of proteins were used for each analysis. According to a comparison of CDPro software done by Sreerama et al., CONTINLL provided the best performance for the unordered fractions as well as providing the best overall performance in comparison with other methods (46). They reported that the reason for this may be, in part, due to larger errors in the estimations of unordered fractions by the other methods. In addition, they concluded that SELCON3 performed slightly better than CDSSTR (46). Upon comparing the fractions for each secondary structural element in cAlb and each mutant, the values from CONTINLL and SELCON3 were very similar overall. In cases where the fractions were significantly different from one another (greater or less than 0.2), the three fractions were averaged and would typically be equivalent to the values reported by CONTINLL. CONTINLL values were thereby used in comparison with other data.



cAlb	H (regular)	H (distorted)	S (regular)	S (distorted)	Turn	Unordered
CONTINLL	.149	.244	.001	.072	.270	.265
CDSSTR	.193	.244	.076	.067	.181	.254
SELCON3	.138	.209	.013	.063	.259	.329

A21G	H (regular)	H (distorted)	S (regular)	S (distorted)	Turn	Unordered
CONTINLL	.037	.133	.113	.093	.237	.388
CDSSTR	.010	.057	.165	.105	.242	.414
SELCON3	.063	.111	.132	.080	.224	.409

V24G	H (regular)	H (distorted)	S (regular)	S (distorted)	Turn	Unordered
CONTINLL	.038	.129	.076	.079	.257	.421
CDSSTR	.039	.132	.126	.089	.260	.356
SELCON3	.065	.123	.099	.092	.242	.409

L26G	H (regular)	H (distorted)	S (regular)	S (distorted)	Turn	Unordered
CONTINLL	.065	.152	.140	.083	.224	.336
CDSSTR	.055	.152	.110	.093	.238	.353
SELCON3	.085	.142	.119	.074	.229	.347

K28G	H (regular)	H (distorted)	S (regular)	S (distorted)	Turn	Unordered
CONTINLL	.046	.162	.043	.082	.274	.393
CDSSTR	.074	.171	.117	.086	.226	.324
SELCON3	.076	.177	.020	.073	.255	.387

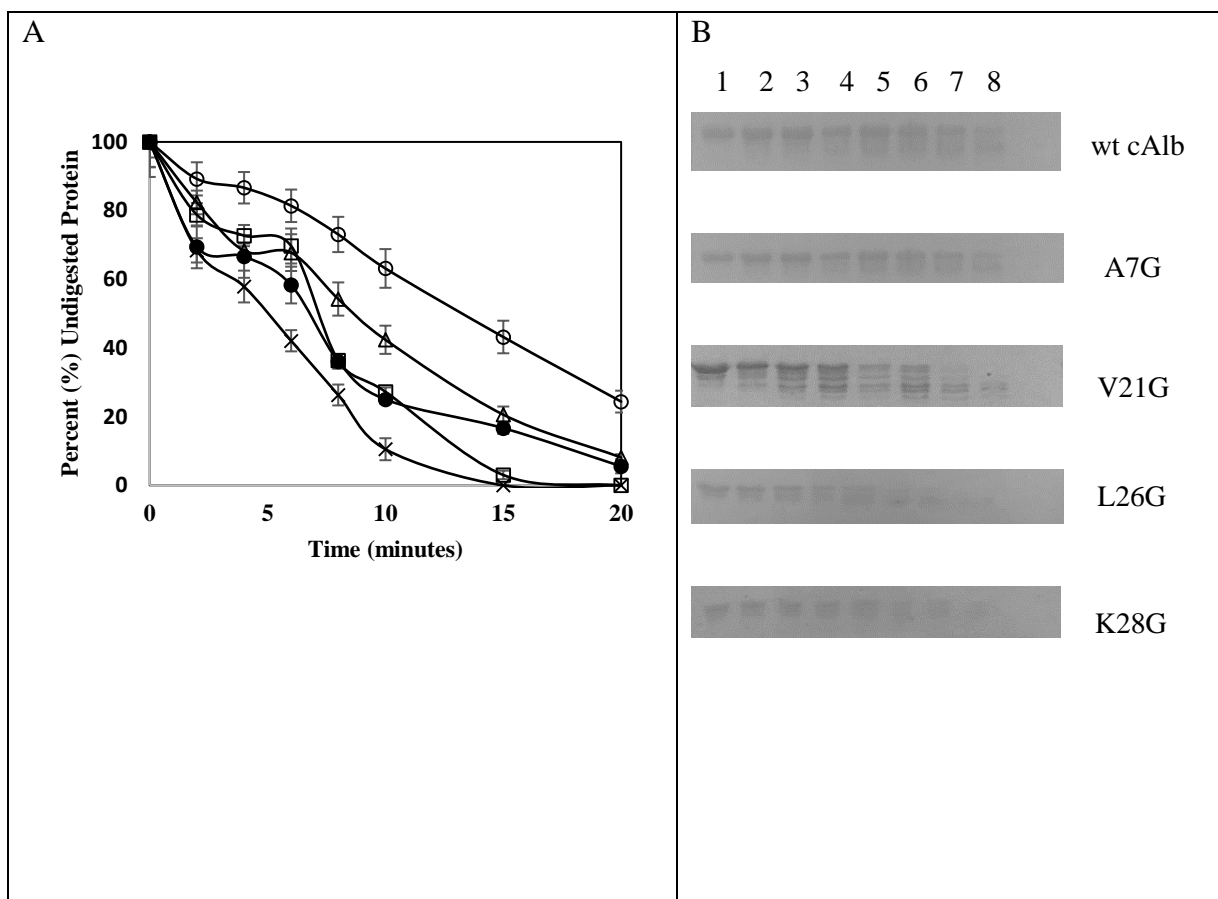
**Fig. 12: Estimation of protein secondary structure from far-UV CD spectra of wt-cAlb and the mutants, the percent of structured and unordered content using was calculated using CONTINLL, CDSSTR and SELCON3.**

The unordered fraction is consistently the largest fraction for cAlb and the mutants which is consistent with the CD spectra and NMR data revealing the predominance of disorder. The combined percentages of regular and disordered helices were reduced by at least 0.2 in all the mutants as compared to the wild type. The percentages for unordered content were consequently raised for all mutants with the V24G mutant showing the highest percentage of unordered secondary structure. The percentages of regular and disordered sheet content were significantly low (less than 0.1) in the wild type and this result stayed relatively consistent with all the mutants. The calculated percentages of turns were comparable for cAlb and the single point mutants to glycine. The fraction of helical structure estimated from the CD spectrum of wild type cAlb matches closely with the percent of helical content predicted by the secondary

prediction software and the propensity towards structure calculated by computational methods shown later in this chapter. To conclude, the single point mutants showed an increase in unordered percentage as compared to the wild type cAlb thereby supporting a reduction in overall order for these mutants.

### ***Limited trypsin digestion reveals an overall decrease in stability in the cAlb mutants***

Limited trypsin digestion provides information about the backbone flexibility of proteins. Arginine and lysine residues located in flexible, solvent-accessible regions of proteins are more vulnerable to trypsin cleavage. Wild type cAlb contains 46 trypsin cleavage sites in conjunction with being predominately unstructured makes it highly susceptible to enzymatic cleavage. After 20 minutes of incubation with trypsin at 25°C the cAlb parent band, migrating at 15 kDa, is reduced to 25% of its original quantity. At the same time point, A21G, V24G, L26G and K28G are reduced to 8%, 0%, 0% and 5 %, respectively (Fig.13). Comparison of wt-cAlb with the mutants reveals an increase in the susceptibility of the cAlb mutants to trypsin digestion. The results here suggest that the introduction of these single point mutations led to a significant increase in the backbone flexibility of cAlb and therefore an overall loss of structural integrity. The observations here are consistent with the assertion of a decrease in local structure made by the Far-UV-CD spectra analyses.



**Fig. 13: Densitometric analysis of the SDS-PAGE gels depicting the resistance of wt-cAlb (○), A21G (△), V24G (□), L26G (x) and K28G (●) to limited trypsin digestion (Panel-A). SDS-PAGE gels of limited trypsin digestion for cAlb and mutants (Panel-B). Lane 1= Zero Trypsin, Lane 2= 2 minutes, Lane 3= 4 minutes, Lane 4= 6 minutes, Lane 5= 8 minutes, Lane 6= 10 minutes, Lane 7= 15 minutes, Lane 8= 20 minutes.**

***ITC and SEC analysis reveals that the single point mutants do not significantly affect the binding to cpSRP43***

cAlb and A21G were each titrated into cpSRP43 at a 1:10 ratio. The buffer was subtracted from each profile and the binding affinities were measured and reported by the equilibrium dissociation constant (Kd). The Kd values for cpSRP43 Vs. cAlb and cpSRP43 Vs. A21G were comparable with only a 0.75  $\mu$ M difference with the mutant displaying the slightly lower affinity (S6). Co-elution via size exclusion chromatography between cpSRP43 and the remaining mutants (V24G, L26G and K28G) was determined to assess the presence of a binding event.

cpSRP43 elutes at approximately 46 to 52 mls with a molecular weight of 35 kDa while cAlb elutes at 60 mls with a molecular weight of 15 kDa (S7). The complex of cpSRP43/cAlb elutes from 46 to 52 mls (S8). cpSRP43 and V24G, L26G and K28G were each incubated on a rocker at 4°C for an hour at a concentration of 30 μM cpSRP43 to 15 μM mutant. Samples were then passed over a Superdex 75 column which was pre-equilibrated in the same buffer used in the incubation of both proteins ( 2.7mM KCl, 1.8mM KH<sub>2</sub>PO<sub>4</sub>, 15mM Na<sub>2</sub>HPO<sub>4</sub>, 137mM NaCl, pH 7.2 ). In each complex, the co-elution was observed by absorbance at 280 nm and the elution volumes matched that of the cpSRP43/cAlb complex (S9-11). This suggests that the single point mutations did not significantly affect the binding of cAlb to cpSRP43.

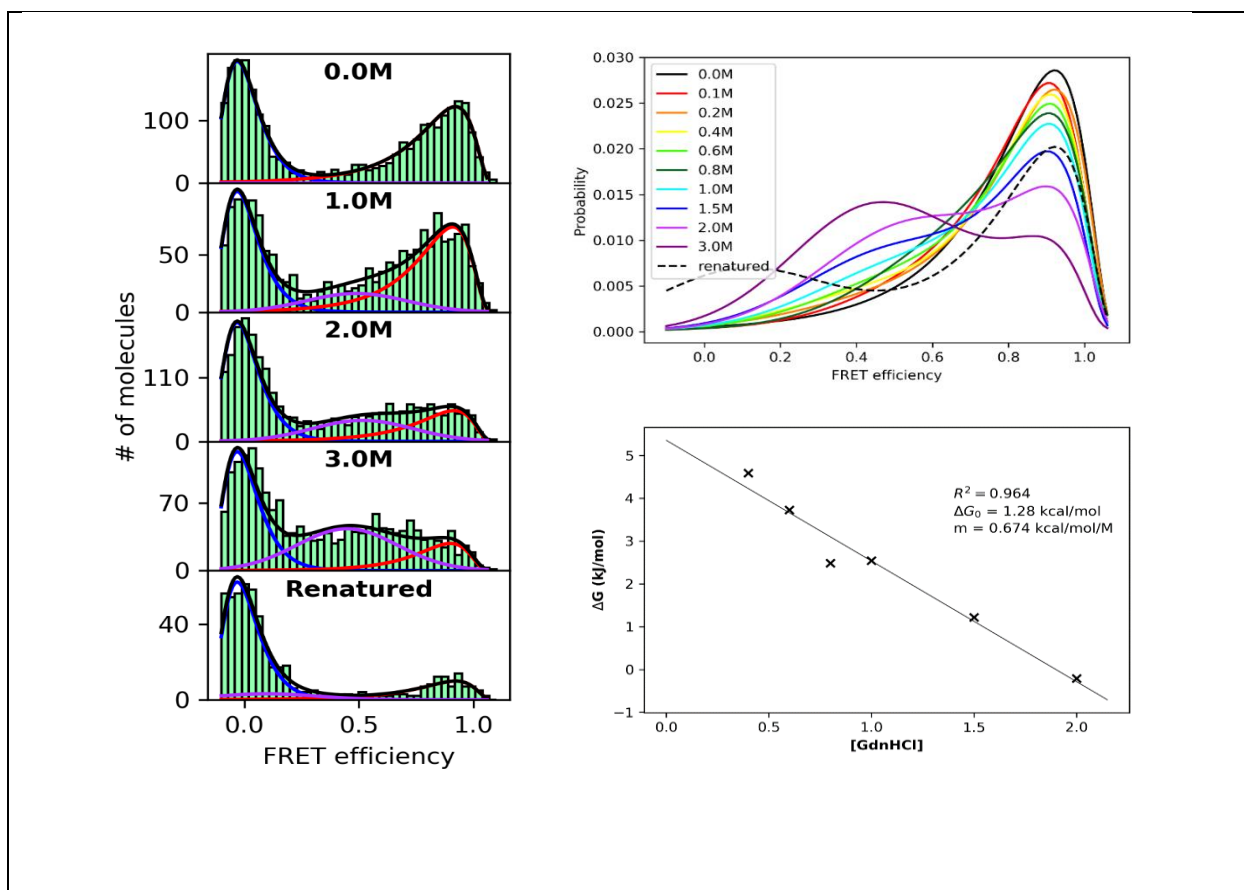
***smFRET of cAlb supports the presence of residual secondary structure***

Single-molecule Förster resonance energy transfer spectroscopy is a useful technique for measuring distances in single molecules and for the study of proteins under the denaturing conditions (51, 52). smFRET was used to examine the same region of cAlb to determine if any local structure was present as it was predicted by secondary structure software. A double cysteine mutant was designed with the first cysteine on the far N-terminal at position 14 and the second cysteine was located at position 52. The placement of the cysteine residues was engineered to be flanking the region of interest with enough distance to probe for the existence of possible structure. Cysteine-maleimide chemistry and site-specific dye labeling was applied in smFRET to probe the distances between the dye positions of freely diffusing cAlb under native conditions and during titration with guanidinium hydrochloride. As a control to ensure that the wild type cAlb protein was properly labeled, it was treated under the same experimental conditions as the mutant. Although the wild type did show some high FRET molecules, the non-specific labeling showed a broad histogram of FRET molecules and the labeling was much less

efficient. Polarization anisotropy was performed on donor-only labeled mutant cAlb. It was found that there was complete rotational averaging of the Alexa488 dye and that  $\kappa^2$  may be assumed to be 2/3. Under standard buffer conditions the mutant displays a single, high FRET efficiency peak (Fig.14A, top panel). The high FRET peak is the dual-labeled population. Using a measured  $R_0$  of 57 angstroms, the distance between the two dyes is  $\sim 37$  angstroms. This distance suggests that the region is compact and would be contrary to that region being analogous to a random coil. Additionally, the peak is broad which may indicate that although a more compact conformation is preferred in this region, there is also some degree flexibility. To determine if the prediction software was correct in predicting an ordered region, between the two cysteine residues, titration using increasing amounts of guanidinium hydrochloride was applied. smFRET which monitored the inter-dye-dye distance showed unfolding upon titration of guanidine hydrochloride.

Upon titration, the high FRET peak begins to disappear and a shift to lower FRET efficiencies continue through a concentration of 3M guanidinium hydrochloride. Three peaks were fit in the smFRET histograms. The peak having very low FRET efficiencies represents the donor-only population of molecules and was fit with a lognormal function with parameters corresponding to the smFRET signal of a donor-only label mutant. The high FRET peak was also fit to a lognormal function with parameters associated with the FRET labeled protein under non-denaturing conditions. Lastly, a third, free Gaussian peak was added to show the subpopulation of denatured molecules. Using these fits, the donor-only peak was removed (Fig.14B). The denaturation of the high FRET peak indicates that some local structure was present. The region moves more toward disorder as the denaturant concentration is increased. This would indicate a loss of structure in this region. Additionally, the denaturant was removed, by dialysis, from the

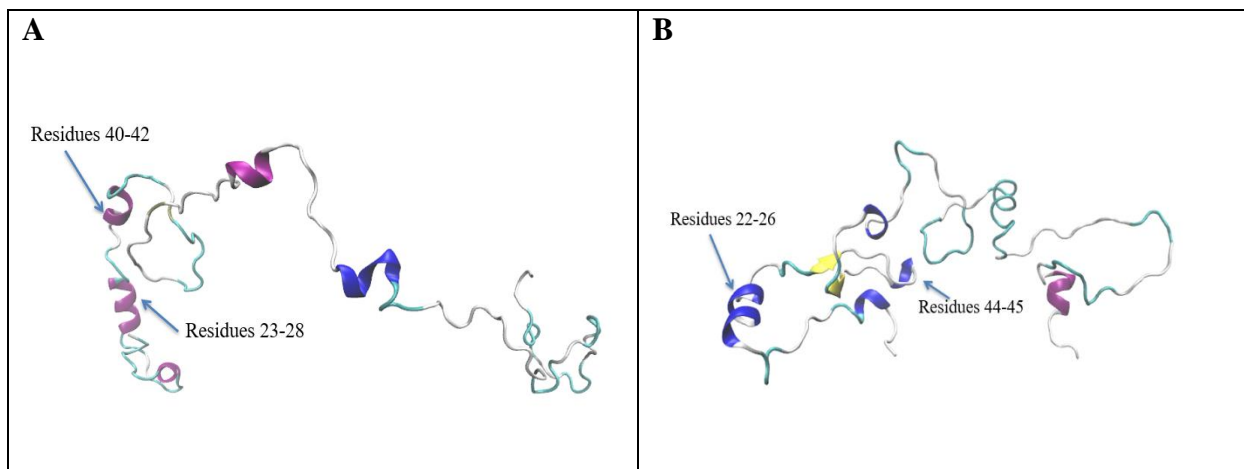
labeled cAlb after the experiment. The dialyzed sample was then measured and was able to recover the high FRET peak. Next the free energy barrier associated with the unfolding process of the undenatured state was determined by comparing the free energy difference between folded and unfolded populations. Free energy difference between the folded and unfolded states is 1.28 kcal/mol with a cooperativity of 0.674 kcal/mol/M. This low energy difference is indicative of low stability in the region of interest.



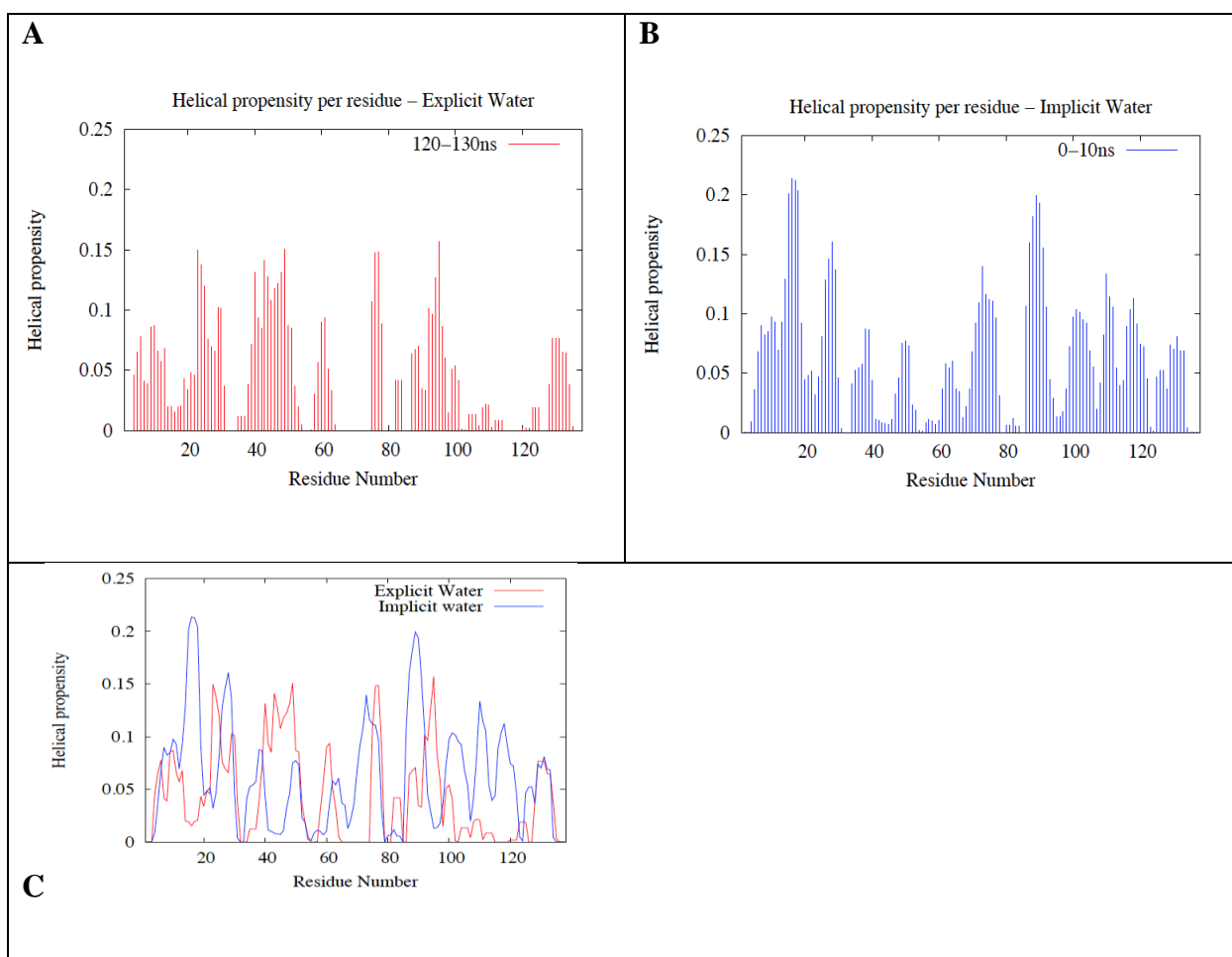
**Fig. 14:** Selected smFRET histograms of S52C cAlb in increasing guanidinium hydrochloride (Panel-A: Left). Fit values with the donor only peak removed of the titration smFRET histograms (Panel-B: Right top). Plot of the unfolding energetics of S52C (Panel-C: Right bottom ). By plotting the  $\Delta G_{FU}$  versus denaturant concentration, the y-intercept yields  $\Delta G_0$  and the slope corresponds to the negative cooperativity value,  $m$ , for the unfolding.

***Molecular simulations and helical propensity predictions support the same region of local structure found in experimentation***

Application of sixteen independent metadynamics simulations of the cAlb were performed in both implicit and explicit water (Fig.15). Conformations derived from these simulations revealed that cAlb contained helical formations within the region predicted by the software (amino acid residues 15-38) and experimentally tested by site-directed mutagenesis and smFRET. Regions of structure located near the C-terminal end of the sequence were also reported by the simulations which correlates with some of the predictions made initially however our focus was on the amino acid stretch near the N-terminal because this region was reported to contain structure by all of the different predicting algorithms that were applied at the start. The helical propensities purported by the simulations showed that the region of interest (residues 15-45) has a peak helical propensity of around 15% in explicit water and 20% in implicit water (Fig.16). This data agrees with the helical content fractions extrapolated from the circular dichroism of the single-point mutations to glycine in the same region (Fig. 12). Regions of structural propensity determined by the computational method also fall within the same region (Fig.17 and 18). Overall, the conformations, helical/structural propensities and smFRET calculations constructed by computational methods closely matched the experimental data.

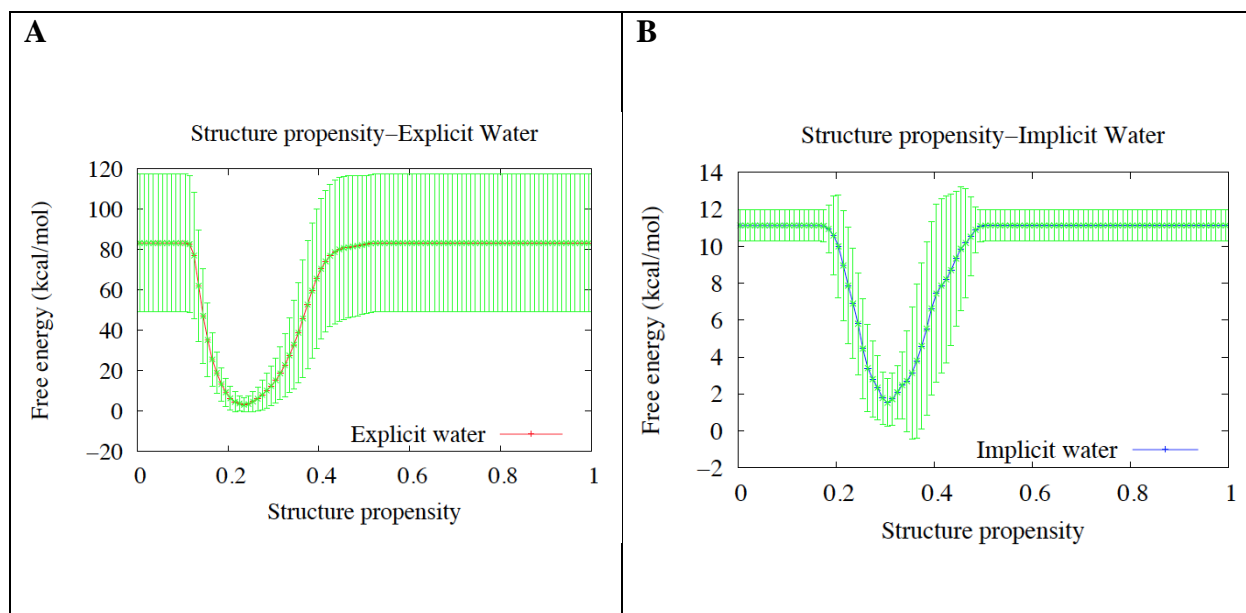


**Fig. 15: A representative of converged conformation in explicit (Panel-A) and implicit (Panel-B) water.**



**Fig. 16: Helical propensity (per residue) in explicit (Panel-A) and implicit (Panel-B) water. STRIDE algorithm was used to calculate the helical propensity of individual amino acid residues. Comparison of helical propensity in explicit and implicit water (Panel-C).**





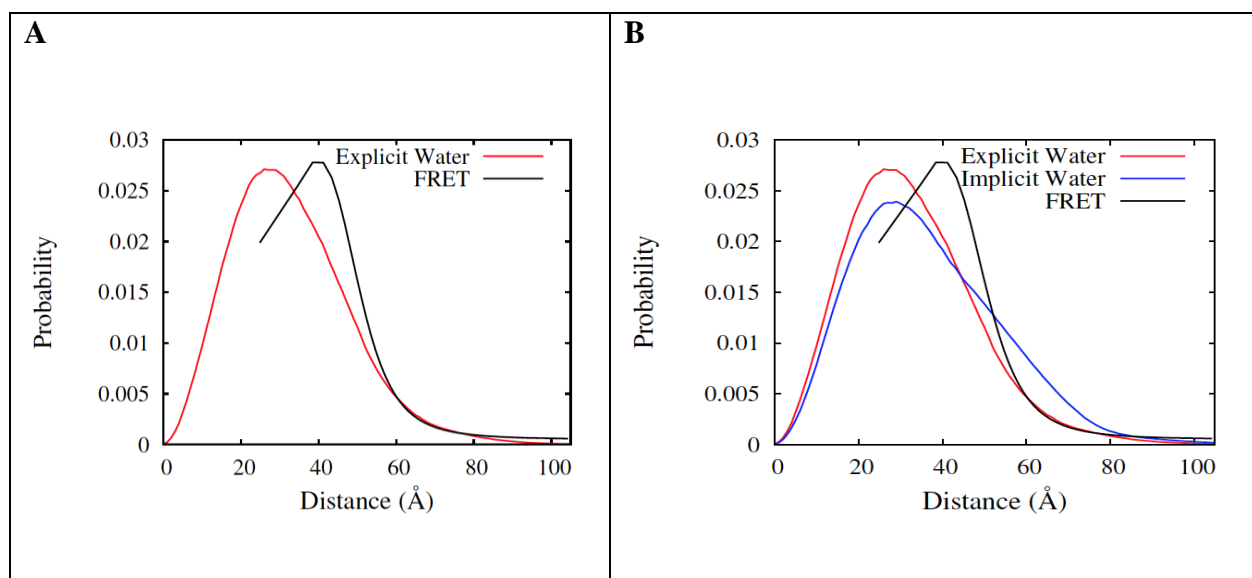
**Fig. 17: Structure propensity in explicit (Panel-A) and implicit (Panel-B) water.**

<b>Explicit Water</b>		<b>Implicit Water</b>	
0	26 %	0	30 %
1	27 %	1	37 %
2	23 %	2	38 %
3	23 %	3	27 %
4	21 %	4	35 %
5	27 %	5	37 %
6	27 %	6	32 %
7	30 %	7	29 %
8	27 %	8	25 %
9	25 %	9	43 %
10	20 %	10	23 %
11	36 %	11	36 %
12	24 %	12	31 %
13	29 %	13	41 %
14	22 %	14	28 %
15	20 %	15	30 %
<b>Avg</b>	<b>26 %</b>	<b>Avg</b>	<b>33 %</b>
<b>SD</b>	<b>4 %</b>	<b>SD</b>	<b>6 %</b>

**Fig. 18: Overall structure propensity was around 26% in explicit water and 33% in implicit water.**

### *Experimental smFRET and computational smFRET distances reported similar dye-dye distances*

Comparison of the experimental smFRET distances done with the double cysteine mutant and the smFRET predictions made upon simulation proved to have similar distances between the dye pairs (Fig.19). Similar smFRET distances derived computationally and experimentally provides added confidence to the theory that cAlb contains an area of local structure.



**Fig. 19: Comparison of experimental and computational dye-dye distance distributions.**

### **Summary and Conclusions**

The C-terminus of Alb3 is required for the proper functionality of Alb3 in the integration of LHCPs into the thylakoid membrane (53). cAlb interacts with multiple partners at the thylakoid membrane and serves as the last regulator in the cpSRP pathway by first interacting with the cpSRP43 subunit of cpSRP for the subsequent docking of LHCP. The association between cAlb and cpSRP43 initiates GTP hydrolysis by cpSRP54/cpFtsY and the release of LHCP which is followed by disassociation of cpSRP and the receptor from Alb3 (4). Although

there have been extensive studies to investigate the binding interface between cAlb and cpSRP43 it is still indeterminate as to the precise interaction which proves to be functionally dominate. Based on CD spectrum analysis of a cAlb and cpSRP43 mixture, cAlb is theorized to adopt a more compact structure involving the formation of helices upon binding to cpSRP43 (3). The classification of cAlb as an intrinsically disordered protein supports a folding upon binding model of interaction between cAlb and cpSRP43. The binding event between cAlb and cpSRP43 is consider to be a critical interaction for providing stability to the transit complex at the docking site (3). In this study we have taken a closer look at the structure of cAlb to investigate thoroughly the possible existence of an ordered region in this predominately disordered protein. In this study, we have combined structural predictions, biophysical approaches, spectroscopic techniques and computational methods to reveal a region of local structure in cAlb. Further examination is required to verify if this region, located in the N-terminal of cAlb, is functionally important in the interaction with cpSRP43. The structural nature of cAlb is initially identified in its primary sequence, which shows a low percentage of hydrophobic residues (27%) and a high content of polar or charged residues (67%) with the remainder being composed of non-polar residues. This type of amino acid composition presupposes that cAlb will be lacking the driving force necessary to fold into a compact structure. The secondary structure and disorder predictions illustrate that cAlb is a predominately disordered protein, with a stretch of residues, approximately 30 amino acids, at the N-terminal, which contain helical secondary structure. While the NMR and SAXS data supports the conclusion that cAlb is intrinsically disordered, the CD data having an additional broad negative peak between 220 nm and 230 nm, implies the existence of some residual secondary structure (Fig.7A). Using site-directed mutagenesis, our results show an overall

decrease in structural integrity when helix-promoting residues in the N-terminal region of cAlb were replaced with glycine, known to be a helix breaker. CD spectrum analysis of these mutants revealed that three out of the four mutations lead to a reduction in the local structure present when compared with the CD spectra of wild type cAlb (Fig.11A). All the mutants showed lowered structural stability during limited trypsin digestion experiments however the introduction of glycine in each position did not have a strong influence on binding to cpSRP43 (Fig.13, S6-11). smFRET of cAlb under native conditions and upon increasing concentrations of denaturant affirms the presence of structure in the same region which was focused on for mutagenesis (Fig.14). In addition, the dye-dye distances reported from experimental smFRET matches well with the computational smFRET distances (Fig.19). Molecular modeling depicts a structure of cAlb having a helix located within in the same 30 amino acid stretch identified by the prediction software (Fig.15). Molecular dynamics simulations calculated helical/structural propensities in the region of interest of cAlb which agrees with the estimation of protein secondary structural content from the CD analysis using CDPro software (Fig.12 and 16-18). The structural characteristics of cAlb exhibit general disorder and high flexibility with a relatively higher ordered structure propensity in the N-terminal region. In this study, multiple techniques were employed to converge at the same conclusion. The region in cAlb having a propensity toward structure may perhaps serve as a nucleation site for binding to the cpSRP43 subunit in the cpSRP pathway, an interaction critical to the integration of LHCPs in the thylakoid membrane. This region may contain a transient structure which is moving between helical and disorder as cAlb is sampling the cpSRP43. Considering that the single point mutations revealed a decrease in order yet did not affect binding to cpSRP43, these amino acids may not be the key residues important for binding to cpSRP43 or a helix could be forming in another region which is

important to binding. In the future, multiple mutations in this region ,and other regions identified to have a high propensity toward structure, followed by binding studies to cpSRP43 and the L18 peptide in LHCPs will provide information on which region(s) on cAlb are required binding event with cpSRP43 and L18. This investigation will lead to a model of the structure of cAlb which is structurally critical for the binding events between cpSRP43 and L18 for proper integration of LHCPs.

**Author Contributions:** smFRET was performed by Duston Baucom and Dr. Colin Heyes. Molecular simulations and experimental smFRET values were performed by Vivek Kumar and Dr. Mahmoud Moradi.

**Acknowledgements:** This work was supported by the DOE grant DE-02-01ER15161.

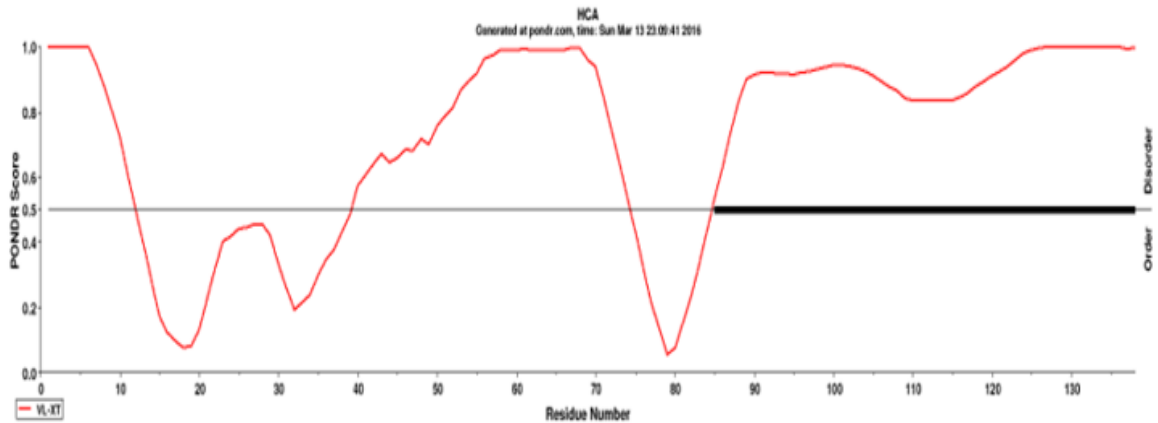
Supplemental material

```

1      MRGSHHHHHH GSACNNVLST AQQVWLRKLG GAKPAVNENA GGIITAGQAK
VLXT   DDDDDDDDDD D                                     D DDDDDDDDDD

51     RSASKPEKGG ERFRQLKEEE KKKKLIKALP VEEVQPLASA SASNDGSDVE
VLXT   DDDDDDDDDD DDDDDDDDDD DDDD          DDDDDD DDDDDDDDDD

101    NNKEQEVTEE SNTSKVSQEV QSFSRERRSK RSKRKPVA
VLXT   DDDDDDDDDD DDDDDDDDDD DDDDDDDDDD DDDDDDDD
  
```



**Supplemental Fig. 1:** Primary sequence of cAlb analyzed with PONDR indicates that it is mostly disordered with a 28 amino acid ordered region at the N-terminus (residues 12-40) and a small proportion of ordered structure from residues 75-84.

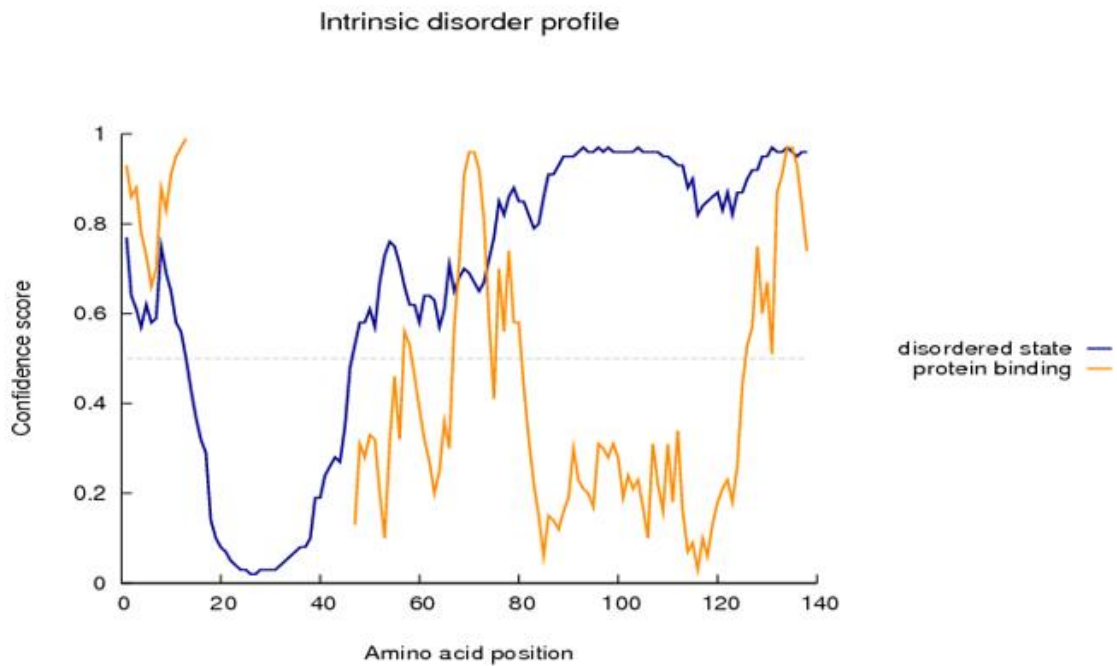
Secondary Structure Map

Feature predictions are colour coded onto the sequence according to the sequence feature key shown below.

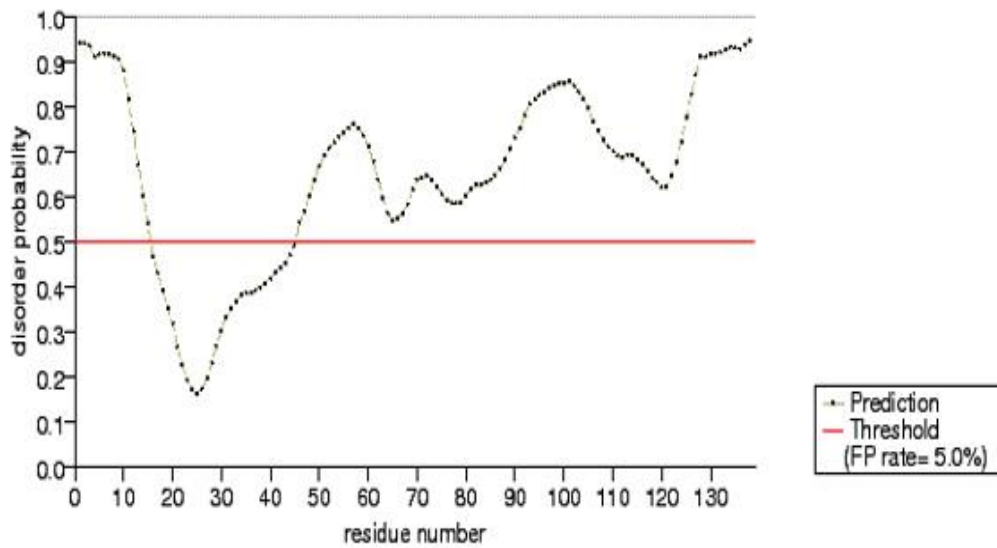


KEY	Helix	Sheet	Disordered	Disordered protein binding	Dompred Boundary	DomSSEA Boundary
Annotations	H	S	D	E	A	B

**Supplemental Fig. 2:** Primary sequence of cAlb analyzed by PSIPRED indicates that residues 15-29 form a helical structure. Residues 63-77 and 6 more residues in the C-terminal are indicated to form a helical structure along with disordered/disordered protein binding structure.



**Supplemental Fig. 3:** Primary sequence of cAlb analyzed by DISOPRED. Amino acids in the input sequence are considered disordered when the blue line is above the grey dashed line, that is the confidence score is higher than 0.5. The orange line shows the confidence of disordered protein binding residue predictions. Residues from ~15-45 were predicted to contain an ordered structure.

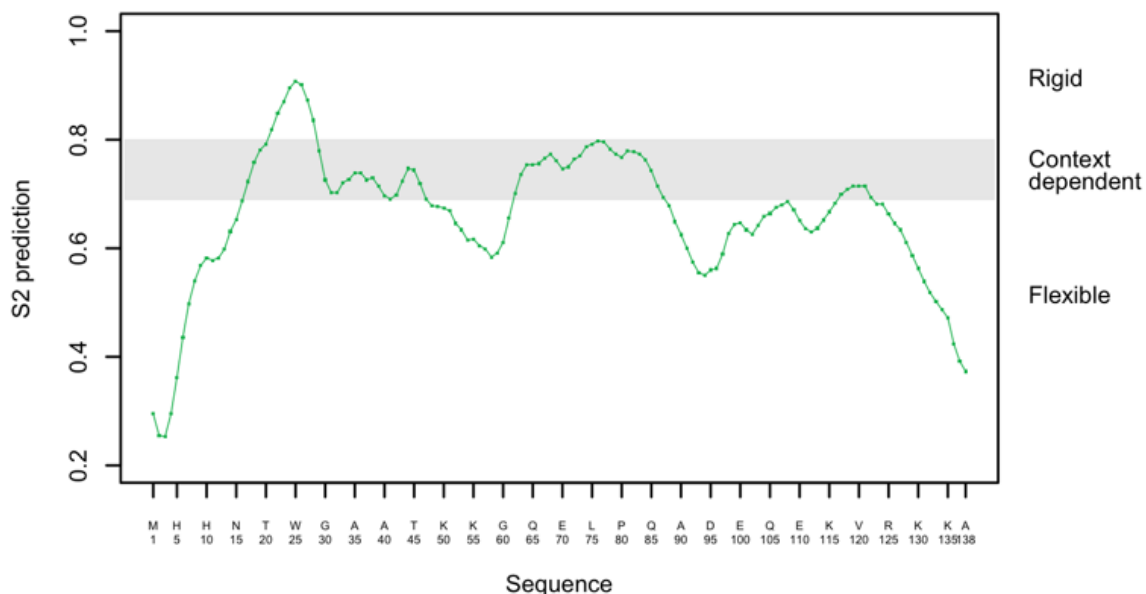


Red: Disordered residues Black: Ordered residues)

1	<b>M</b> R <b>G</b> S <b>H</b> H <b>H</b> H <b>H</b> H	<b>G</b> S <b>A</b> C <b>N</b> N <b>V</b> L <b>S</b> T	<b>A</b> Q <b>Q</b> V <b>W</b> L <b>R</b> K <b>L</b> G	<b>G</b> A <b>K</b> P <b>A</b> V <b>N</b> E <b>N</b> A	<b>G</b> G <b>I</b> I <b>T</b> A <b>G</b> Q <b>A</b> K	60
51	<b>R</b> S <b>A</b> S <b>K</b> P <b>E</b> K <b>G</b> G	<b>E</b> R <b>F</b> R <b>Q</b> L <b>K</b> E <b>E</b> E	<b>K</b> K <b>K</b> K <b>L</b> I <b>K</b> A <b>L</b> P	<b>V</b> E <b>E</b> V <b>Q</b> P <b>L</b> A <b>S</b> A	<b>S</b> A <b>S</b> N <b>D</b> G <b>S</b> D <b>V</b> E	100
101	<b>N</b> N <b>K</b> E <b>Q</b> E <b>V</b> T <b>E</b> E	<b>S</b> N <b>T</b> S <b>K</b> V <b>S</b> Q <b>E</b> V	<b>Q</b> S <b>F</b> S <b>R</b> E <b>R</b> R <b>S</b> K	<b>R</b> S <b>K</b> R <b>K</b> P <b>V</b> A		150

**Supplemental Fig. 4:** Primary sequence of cAlb analyzed by PrDOS2 indicates that residues 16-45 form an ordered structure.



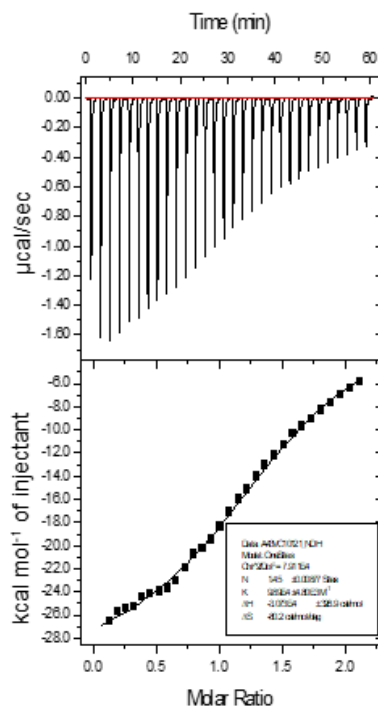


Sequence num.	Residue	Probability
24	VAL	0.90
25	TRP	0.91
26	LEU	0.90

Predictor of protein backbone dynamics. A value of 1 means complete order (stable conformation), while a value of 0 means fully random bond vector movement (highly dynamic).

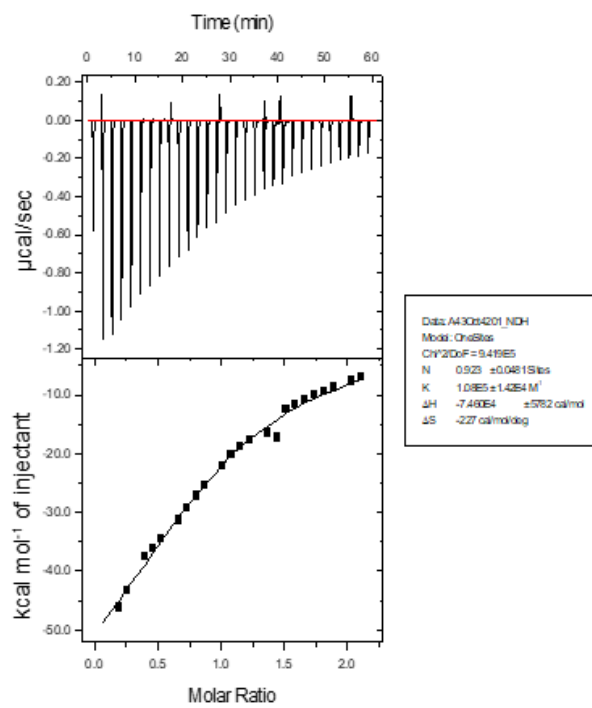
**Supplemental Fig. 5:** Primary sequence of cAlb analyzed by DynaMine indicates that residues 24, 25 and 26 have a high probability to contribute to a stable conformation.

### cAlb Vs. cpSRP43



$K_d = 10 \mu\text{M}$

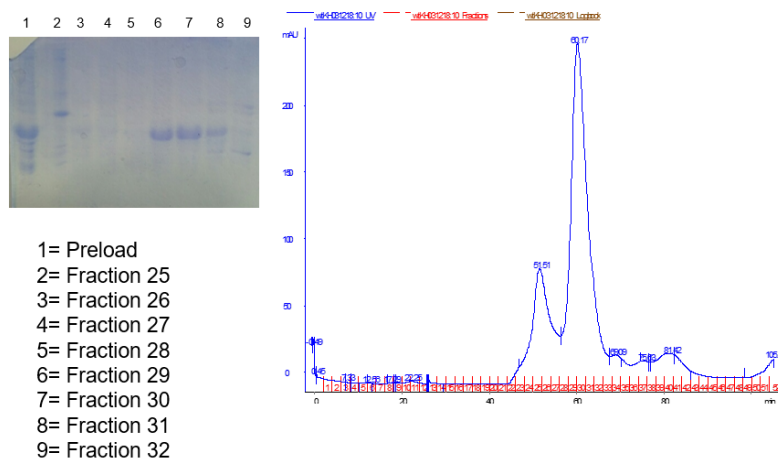
### A21G Vs. cpSRP43



$K_d = 9.25 \mu\text{M}$

**Supplemental Fig.6:** ITC of cAlb titrated into cpSRP43 (left) and ITC of A21G titrated into cpSRP43 (right).

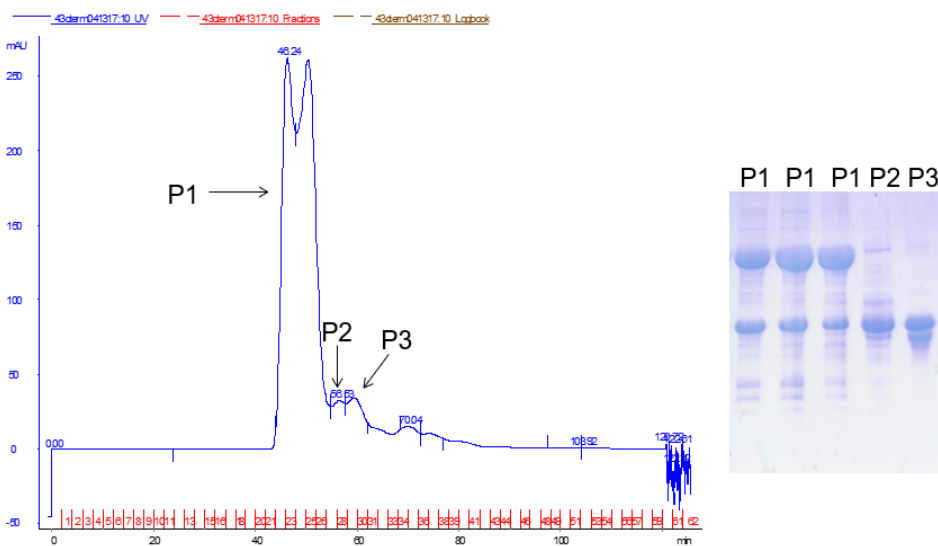
### SEC elution of cAlb at 60 mls



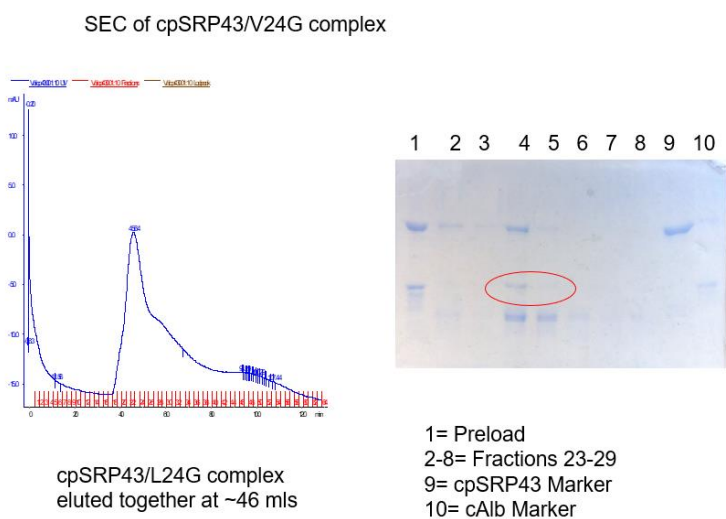
- 1= Preload
- 2= Fraction 25
- 3= Fraction 26
- 4= Fraction 27
- 5= Fraction 28
- 6= Fraction 29
- 7= Fraction 30
- 8= Fraction 31
- 9= Fraction 32

**Supplemental Fig.7:** Pure cAlb elutes at 60 mls.

cpSRP43/cAlb complex elutes at 46.24 mls

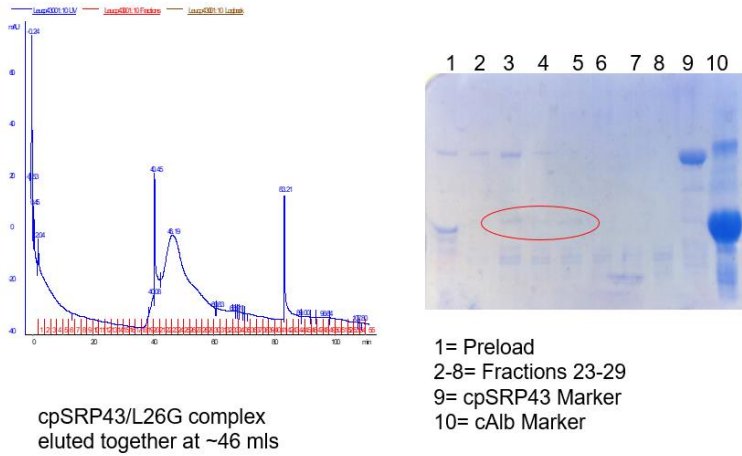


**Supplemental Fig. 8:** The complex of cpSRP43/cAlb elutes at 46-52 mls of buffer volume.



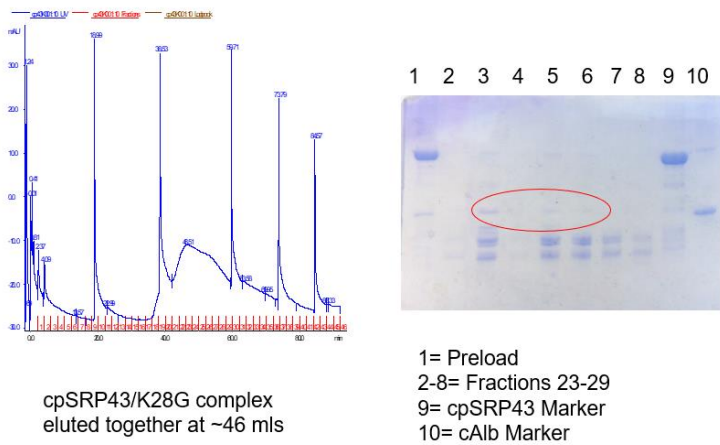
**Supplemental Fig. 9:** The complex of cpSRP43/V24G elutes together at 46-52 mls of buffer volume. The V24G band is circled in red.

SEC of cpSRP43/L26G complex



**Supplemental Fig. 10:** The complex of cpSRP43/L26G elutes together at 46-52 mls of buffer volume. The L26G band is circled in red.

SEC of cpSRP43/K28G complex



**Supplemental Fig. 11:** The complex of cpSRP43/K28G elutes together at 46-52 mls of buffer volume. The K28G band is circled in red.

## References

1. Wang P, Dalbey RE. Inserting membrane proteins: The YidC/Oxa1/Alb3 machinery in bacteria, mitochondria, and chloroplasts. *Biochimica et Biophysica Acta (BBA) - Biomembranes*. 2011;1808(3):866-75.
2. Kumazaki K, Chiba S, Takemoto M, Furukawa A, Nishiyama K-i, Sugano Y, et al. Structural basis of Sec-independent membrane protein insertion by YidC. *Nature*. 2014;509:516.
3. 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*. 2010;285(8):5954-62.
4. Lewis NE, Marty NJ, Kathir KM, Rajalingam D, Kight AD, Daily A, et al. A dynamic cpSRP43-albino3 interaction mediates translocase regulation of cpSRP targeting components. *Journal of Biological Chemistry*. 2010.
5. 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*. 2010;584(19):4138-44.
6. 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*. 2011;286(40):35187-95.
7. Stengel KF, Holdermann I, Cain P, Robinson C, Wild K, Sinning I. Structural Basis for Specific Substrate Recognition by the Chloroplast Signal Recognition Particle Protein cpSRP43. *Science*. 2008;321(5886):253-6.
8. Horn A, Hennig J, Ahmed YL, Stier G, Wild K, Sattler M, et al. Structural basis for cpSRP43 chromodomain selectivity and dynamics in Alb3 insertase interaction. *Nature Communications*. 2015;6:11.
9. Urbischek M, Nick von Braun S, Brylok T, Gügel IL, Richter A, Koskela M, et al. The extreme Albino3 (Alb3) C terminus is required for Alb3 stability and function in *Arabidopsis thaliana*. *Planta*. 2015;242(3):733-46.
10. O'Neil KT, DeGrado WF. A thermodynamic scale for the helix-forming tendencies of the commonly occurring amino acids. *Science*. 1990;250(4981):646-51.
11. Lyu PC, Liff MI, Marky LA, Kallenbach NR. Side chain contributions to the stability of alpha-helical structure in peptides. *Science*. 1990;250(4981):669-73.
12. Pace CN, Scholtz JM. A helix propensity scale based on experimental studies of peptides and proteins. *Biophysical journal*. 1998;75(1):422-7.
13. Chou PY, Fasman GD. Conformational parameters for amino acids in helical,  $\beta$ -sheet, and random coil regions calculated from proteins. *Biochemistry*. 1974;13(2):211-22.

14. Acerbo AS, Cook MJ, Gillilan RE. Upgrade of MacCHESS facility for X-ray scattering of biological macromolecules in solution. *Journal of Synchrotron Radiation*. 2015;22(1):180-6.
15. Nielsen SS, Møller M, Gillilan RE. High-throughput biological small-angle X-ray scattering with a robotically loaded capillary cell. *Journal of Applied Crystallography*. 2012;45(2):213-23.
16. Konarev PV, Volkov VV, Sokolova AV, Koch MH, Svergun DI. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *Journal of applied crystallography*. 2003;36(5):1277-82.
17. Xue B, Dunbrack RL, Williams RW, Dunker AK, Uversky VN. PONDR-FIT: A meta-predictor of intrinsically disordered amino acids. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*. 2010;1804(4):996-1010.
18. Ward JJ, McGuffin LJ, Bryson K, Buxton BF, Jones DT. The DISOPRED server for the prediction of protein disorder. *Bioinformatics*. 2004;20(13):2138-9.
19. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research*. 1997;25(17):3389-402.
20. Jones DT, Ward JJ. Prediction of disordered regions in proteins from position specific score matrices. *Proteins: Structure, Function, and Bioinformatics*. 2003;53(S6):573-8.
21. Ishida T, Kinoshita K. PrDOS: prediction of disordered protein regions from amino acid sequence. *Nucleic Acids Research*. 2007;35(suppl\_2):W460-W4.
22. Cilia E, Pancsa R, Tompa P, Lenaerts T, Vranken WF. From protein sequence to dynamics and disorder with DynaMine. *Nature communications*. 2013;4:2741.
23. Cilia E, Pancsa R, Tompa P, Lenaerts T, Vranken WF. The DynaMine webserver: predicting protein dynamics from sequence. *Nucleic acids research*. 2014;42(W1):W264-W70.
24. Galzitskaya OV, Garbuzynskiy SO, Lobanov MY. FoldUnfold: web server for the prediction of disordered regions in protein chain. *Bioinformatics*. 2006;22(23):2948-9.
25. Whitmore L, Wallace BA. Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases. *Biopolymers: Original Research on Biomolecules*. 2008;89(5):392-400.
26. Fasman GD. *Circular dichroism and the conformational analysis of biomolecules*: Springer Science & Business Media; 2013.
27. Greenfield NJ. Using circular dichroism spectra to estimate protein secondary structure. *Nature protocols*. 2006;1(6):2876.
28. Chemes LB, Alonso LG, Noval MG, de Prat-Gay G. Circular dichroism techniques for the analysis of intrinsically disordered proteins and domains. *Intrinsically disordered protein analysis*: Springer; 2012. p. 387-404.

29. Woody RW. Circular dichroism of intrinsically disordered proteins: John Wiley & Sons, Hoboken, NJ; 2010.
30. Eftink MR. Intrinsic fluorescence of proteins. Topics in fluorescence spectroscopy: Springer; 2002. p. 1-15.
31. Keeler J. Understanding NMR Spectroscopy (2004). 2016.
32. Keeler J. Understanding NMR spectroscopy: John Wiley & Sons; 2011.
33. Chi Y-H, Kumar TKS, Kathir KM, Lin D-H, Zhu G, Chiu I-M, et al. Investigation of the structural stability of the human acidic fibroblast growth factor by hydrogen– deuterium exchange. *Biochemistry*. 2002;41(51):15350-9.
34. Canales A, Lozano R, López-Méndez B, Angulo J, Ojeda R, Nieto PM, et al. Solution NMR structure of a human FGF-1 monomer, activated by a hexasaccharide heparin-analogue. *The FEBS journal*. 2006;273(20):4716-27.
35. Arunkumar AI, Srisailam S, Kumar TKS, Kathir KM, Chi Y-H, Wang H-M, et al. Structure and stability of an acidic fibroblast growth factor from *Notophthalmus viridescens*. *Journal of Biological Chemistry*. 2002;277(48):46424-32.
36. Yao J, Dyson HJ, Wright PE. Chemical shift dispersion and secondary structure prediction in unfolded and partly folded proteins. *FEBS letters*. 1997;419(2-3):285-9.
37. Schneidman-Duhovny D, Hammel M, Sali A. FoXS: a web server for rapid computation and fitting of SAXS profiles. *Nucleic acids research*. 2010;38(suppl\_2):W540-W4.
38. Boldon L, Laliberte F, Liu L. Review of the fundamental theories behind small angle X-ray scattering, molecular dynamics simulations, and relevant integrated application. *Nano reviews*. 2015;6(1):25661.
39. Putnam CD, Hammel M, Hura GL, Tainer JA. X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Quarterly reviews of biophysics*. 2007;40(3):191-285.
40. Semisotnov GV, Timchenko AA, Melnik BS, Kimura K, Kihara H. Kratky plot as a tool to evaluate the molecular mass of globular proteins. *Photon Factory Activity Report 2002*. 2002;20.
41. Kikhney AG, Svergun DI. A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. *FEBS letters*. 2015;589(19PartA):2570-7.
42. Grant TD, Luft JR, Carter LG, Matsui T, Weiss TM, Martel A, et al. The accurate assessment of small-angle X-ray scattering data. *Acta Crystallographica Section D: Biological Crystallography*. 2015;71(1):45-56.
43. Hermans J, Anderson AG, Yun R. Differential helix propensity of small apolar side chains studied by molecular dynamics simulations. *Biochemistry*. 1992;31(24):5646-53.
44. Creamer TP, Rose GD.  $\alpha$ -Helix-forming propensities in peptides and proteins. *Proteins: Structure, Function, and Bioinformatics*. 1994;19(2):85-97.

45. Micsonai A, Wien F, Kernya L, Lee Y-H, Goto Y, Réfrégiers M, et al. Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy. *Proceedings of the National Academy of Sciences*. 2015;201500851.
46. Sreerama N, Woody RW. Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Analytical biochemistry*. 2000;287(2):252-60.
47. Sreerama N, Venyaminov SY, Woody RW. Estimation of protein secondary structure from circular dichroism spectra: inclusion of denatured proteins with native proteins in the analysis. *Analytical biochemistry*. 2000;287(2):243-51.
48. Provencher SW. CONTIN: a general purpose constrained regularization program for inverting noisy linear algebraic and integral equations. *Computer Physics Communications*. 1982;27(3):229-42.
49. Johnson WC. Analyzing protein circular dichroism spectra for accurate secondary structures. *Proteins: Structure, Function, and Bioinformatics*. 1999;35(3):307-12.
50. Sreerama N, Woody RW. A self-consistent method for the analysis of protein secondary structure from circular dichroism. *Analytical biochemistry*. 1993;209(1):32-44.
51. Kuzmenkina EV, Heyes CD, Nienhaus GU. Single-molecule Förster resonance energy transfer study of protein dynamics under denaturing conditions. *Proceedings of the National Academy of Sciences*. 2005;102(43):15471-6.
52. Kuzmenkina EV, Heyes CD, Nienhaus GU. Single-molecule FRET study of denaturant induced unfolding of RNase H. *Journal of molecular biology*. 2006;357(1):313-24.
53. Urbischek M, von Braun SN, Brylok T, Gügel IL, Richter A, Koskela M, et al. The extreme Albino3 (Alb3) C terminus is required for Alb3 stability and function in *Arabidopsis thaliana*. *Planta*. 2015;242(3):733-46.



## **V. Conclusion**

The information detailed in this thesis further elucidated the unique structure and remarkable stability of cpSRP43 as well as describing its role as a chaperone for proteins external to the chloroplast. In addition, an area having residual structure was found in the c-terminal portion of the Albino 3 protein. The structure of cpSRP43 is unique for its blend of distinctive domains capable of an array of various protein-protein interactions. The different domains along with its structural flexibility have been outlined in previous studies (1-3). cpSRP43 is a highly stable yet highly dynamic protein which is not explained by the current crystal structure (4). This study highlights the thermal stability of cpSRP43 despite its highly flexible character and develops a structural model which justifies the co-existence of flexibility and stability. cpSRP43 can retain its secondary structure upon heating and cooling cycles and undergoes reversible unfolding when subjected to high temperatures and chemical denaturants. These properties are characteristic of heat shock chaperone proteins. The coexisting high extent of stability and flexibility of cpSRP43 is not explained by the current linear structure derived from X-ray crystallography data and therefore molecular dynamics studies were sought to revisit the structure of cpSRP43. A stable conformation of monomeric cpSRP43 was isolated using microsecond-level equilibrium MD simulations and solvent accessible surface area (SASA) distances were predicted for the 8 tryptophan residues of the stable model. Intrinsic fluorescence and fluorescence quenching data acquired for cpSRP43 revealed that the tryptophan residues are partially buried or superficially exposed in a charged microenvironment which supports the highly dynamic and flexible nature of cpSRP43 and agrees with the predicted SASA distances. Experimental and computational data combined here provides a stable monomeric form of cpSRP43 which has a globular structure which is a structure that supports the significant degree

of stability held within cpSRP43. There is likely a stable intermediate between bound and unbound states of cpSRP43 to cpSRP54 with the equilibrium shifting from a more compact, stable conformation to an extended, flexible conformation in order to accommodate binding partners. Although this protein has been proven to be indispensable for its roles in the chloroplast, ongoing investigation into the structure of cpSRP43 will provide clues for its uses still within the plant cell and for its potential applications outside of the chloroplast as a generic chaperone.

cpSRP43's capabilities as a hub protein, adept to binding many partners, has been alluded to in other works yet still requires thorough characterization (5-7). The structural domains of cpSRP43 consist of a combination of ankyrin, chromodomain and flexible regions, making it a specific chaperone for LHCPs (8, 9). It also functions as a unique disaggregase for the fact that an external input of energy is not required for its action. Here we have shown that cpSRP43 can protect proteins outside of the chloroplast from heat-induced aggregation by using carbonic anhydrase, Concanavalin A and hFGF1 as target proteins. In the case of FGF1, not only was heat-induced aggregation arrested but the ability of this protein to assist in cell proliferation was preserved upon heat treatment in the presence of cpSRP43. The protection which cpSRP43 provides to the heat induced aggregation of target proteins is dependent on the molecular weight and concentration of the target protein. The exact mechanism by which cpSRP43 can successfully provide protection to these proteins requires further investigation and is expected to vary depending on the substrate. A substrate binding domain (SBD) comprised of ankyrin repeat motifs 1-4, capped at the N terminus by a CD1 and at the C-terminus by a bridging helix with a binding interface primarily made up of hydrophobic residues has been established in previous literature (10-13). The full extent of cpSRP43's chaperoning capacity is yet to be concretely

defined. Future work emerging from these studies stand to illuminate the role of cpSRP43 as a generic chaperone. Although we know its action is based on the ability to associate with variable regions of different proteins owing to the divergent domains and flexibility within its unique structure, a mechanism of action needs to be determined. Continued heat-induced aggregation and refolding studies of cpSRP43 with a wide variety of candidate proteins will help to define both the extent and the limitations of its action for specific partners. Size exclusion chromatography and isothermal calorimetry can be utilized to determine the strength of cpSRP43 associations with its partners. Binding studies carried out under different concentrations of salt and denaturants will give insight into the nature of the associations and whether the protection provided by cpSRP43 is steered by electrostatic or hydrophobic forces. Information deduced from the combination of these techniques will provide evidence of a mechanism of action which may indeed vary from protein to protein. Further investigation into the potential chaperoning capabilities of cpSRP43 will lead to the application of the unique protein in other systems. Chaperone activity of cpSRP43 could be extrapolated to its use as a therapeutic agent, thus providing protection for aggregation prone proteins.

The c-terminal domain of the Albino 3 protein (cAlb) in chloroplasts has an important job recruiting and docking the LHCP substrate for subsequent integration by the integral membrane portion of Albino 3 through its interaction with cpSRP43. cAlb in chloroplasts has previously been described as intrinsically disordered (14). We observed that in the Far UV circular dichroism spectra of cAlb there exists an additional negative shoulder at ~225 nm. While not indicative of a typical helix, this suggests that there may be a small region of structure within cAlb which is also supported by the NMR spectra. It is understood that cAlb gains helical structure upon binding to cpSRP43. However, our findings indicate that cAlb does contain a

more ordered state, localized at N-terminal end of cAlb, as opposed to a completely random coil before it binds to cpSRP43. Using site-directed mutagenesis, CD analysis and Limited Trypsin Digestion our results show an overall decrease in structural integrity for mutations to glycine within the region of predicted order. However, binding studies between the cAlb mutants and cpSRP43 show that these glycine mutants ,individually, do not have a strong influence on binding. smFRET of cAlb under native conditions and upon increasing concentrations of denaturant confirms the presence of local structure in the same region which was focused on for mutagenesis. Molecular modeling depicts a structure of cAlb having a helix located within in the same 30 amino acid stretch identified by the prediction software. Molecular dynamics simulations calculated helical/structural propensities in the region of interest of cAlb which agrees with the estimation of protein secondary structural content from the CD analysis using CDPro software. The structural characteristics of cAlb exhibit general disorder and high flexibility along with a region having a propensity toward structure in the N-terminal region. Further examination is required to verify if this region, located in the N-terminal of cAlb, is functionally important in the interaction with cpSRP43. The exact binding interface between cpSRP43 and cAlb has not been described at a residual level. Assessment of any predicted structure or perhaps undulating state(s) in cAlb will lead us towards a better understanding of this interaction. Future work to ascertain this information may include double, triple and quadruple point mutations in the region of predicted structure in this study as well as other regions followed by characterization using various biophysical techniques to assess the stability and structure of the mutants in comparison with the wild type sequence. Additional FRET experiments with the cysteine conjugated dyes placed further apart may reveal more precise information about the length of the structured region(s). Characterization and functional studies

done on mutant and truncated forms of cAlb may narrow down the specific amino acids which contribute to binding and perhaps reveal the key residues which are lining up to form the binding interface between cpSRP43 and cAlb. The knowledge gained from these studies will provide residual binding information between cpSRP43 and cAlb and ultimately give insight on the structural impact of this interaction.

## References

1. Gao F, Kight AD, Henderson R, Jayanthi S, Patel P, Murchison M, et al. Regulation of Structural Dynamics within a Signal Recognition Particle Promotes Binding of Protein Targeting Substrates. *Journal of Biological Chemistry*. 2015.
2. Gao F, Kight AD, Henderson RC, Jayanthi S, Patel P, Goforth RL, et al. Role of Structural Flexibility of cpSRP43 in Binding Substrates during Post-Translational Targeting. *Biophysical Journal*. 2015;108(2):61A-A.
3. Liang F-C, Kroon G, McAvoy CZ, Chi C, Wright PE, Shan S-o. Conformational dynamics of a membrane protein chaperone enables spatially regulated substrate capture and release. *Proceedings of the National Academy of Sciences*. 2016:201524777.
4. Holdermann I, Meyer NH, Round A, Wild K, Sattler M, Sinning I. Chromodomains read the arginine code of post-translational targeting. *Nature Structural & Molecular Biology*. 2012;19:260.
5. Falk S, Sinning I. cpSRP43 is a novel chaperone specific for light-harvesting chlorophyll a, b-binding proteins. *Journal of Biological Chemistry*. 2010;285(28):21655-61.
6. Wang P, Liang F-C, Wittmann D, Siegel A, Shan S-o, Grimm B. Chloroplast SRP43 acts as a chaperone for glutamyl-tRNA reductase, the rate-limiting enzyme in tetrapyrrole biosynthesis. *Proceedings of the National Academy of Sciences*. 2018;115(15):E3588-E96.
7. Falk S, Sinning I. cpSRP43 Is a Novel Chaperone Specific for Light-harvesting Chlorophyll a,b-binding Proteins. *Journal of Biological Chemistry*. 2010;285(28):21655-61.
8. Gutensohn M, Fan E, Frielingsdorf S, Hanner P, Hou B, Hust B, et al. Toc, Tic, Tat et al.: structure and function of protein transport machineries in chloroplasts. *Journal of Plant Physiology*. 2006;163(3):333-47.
9. Li X, Henry R, Yuan J, Cline K, Hoffman NE. A chloroplast homologue of the signal recognition particle subunit SRP54 is involved in the posttranslational integration of a protein into thylakoid membranes. *Proceedings of the National Academy of Sciences*. 1995;92(9):3789-93.
10. McAvoy CZ, Siegel A, Piskiewicz S, Miaou E, Yu M, Nguyen T, et al. Two distinct sites of client protein interaction with the chaperone cpSRP43. *Journal of Biological Chemistry*. 2018;293(23):8861-73.
11. Jaru-Ampornpan P, Shen K, Lam VQ, Ali M, Doniach S, Jia TZ, et al. ATP-independent reversal of a membrane protein aggregate by a chloroplast SRP subunit. *Nature Structural & Molecular Biology*. 2010;17:696.

12. Liang FC, Kroon G, McAvoy CZ, Chi C, Wright PE, Shan SO. Conformational dynamics of a membrane protein chaperone enables spatially regulated substrate capture and release. *Proc Natl Acad Sci U S A*. 2016;113(12):E1615-24.
13. Stengel KF, Holdermann I, Cain P, Robinson C, Wild K, Sinning I. Structural Basis for Specific Substrate Recognition by the Chloroplast Signal Recognition Particle Protein cpSRP43. *Science*. 2008;321(5886):253-6.
14. 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*. 2010;285(8):5954-62.

## **VI. Appendix: Simplified Purification of GST-cpSRP43**

### **Abstract**

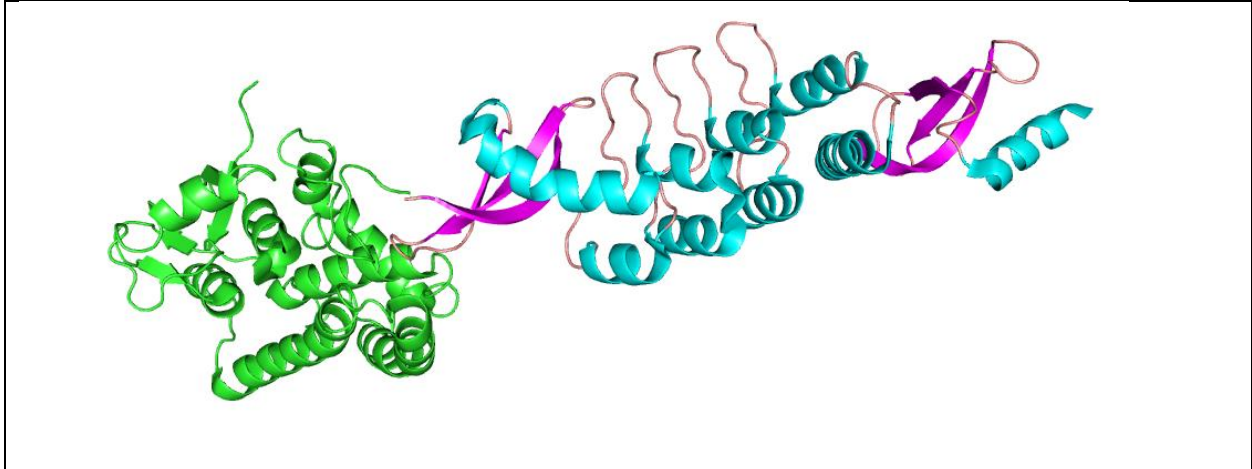
Protein purification for research and medical uses is costly and time consuming in any cases. As an approach to improve the cost and time investment in protein purification we examined the use of a heat treatment step in place of a secondary column for the purification of cpSRP43. Given that cpSRP43 is heat stable up to 90°C and GST is known to aggregate at 45°C, the post cleavage sample of GST and cpSRP43 can be subjected to high heat in an effort to remove the GST tag by aggregating it out of the cleavage product mixture and thereby leaving pure cpSRP43 in the soluble layer. This method effectively removes the GST tag and eliminates the use of a second chromatographic step. This purification method was explored and optimized for cpSRP43, yet it may be applied to the purification of small unstructured proteins and larger heat stable proteins. These results are anticipated to lower the cost and time investment for purification of ideal candidate proteins.



## **Introduction**

A challenging stage of protein research is in the process of obtaining pure proteins in high yields. Protein chromatography requires expensive materials that must typically be purchased commercially which can exhaust funding. Many proteins require the use of several chromatographic steps to acquire purity. Large quantities of protein can be lost during this process thereby reducing the net yield (1). The monetary and time expenditure can make the challenging task of protein purification for researchers and those in the commercial industry even more difficult. This study focused on a new method for the purification of heat stable proteins aimed at eliminating extra chromatographic steps and lowering costs.

Many different types of affinity tags are fused to protein partners to facilitate the purification of proteins in affinity chromatography (2). Glutathione S-transferase (GST) is a commonly used affinity tag for its ability to bind to various metal based resins like nickel and cobalt and can also be expressed in prokaryotic and eukaryotic systems (3-5). GST begins aggregating at 45°C and is completely aggregated at 55°C. GST is also known to stabilize its fusion partner by increasing the overall solubility (6,7). cpSRP43 is stable at much higher temperatures as illustrated in the previous chapters. cpSRP43 has evolved in chloroplasts to chaperone and prevent aggregation of the light-harvesting chlorophyll binding substrate (8,9). cpSRP43 is stable up to over 90°C and the ability to withstand much higher temperatures than GST made the fusion protein an excellent choice for this heat treatment method of purification. This method appears to be applicable for the purification of proteins and peptides with a  $T_m$  is greater than 65 °C.



**Fig.1: Cartoon representation of the fusion protein, GST-cpSRP43.**

## **Materials and Methods**

### **Expression, purification and enzymatic cleavage of GST-cpSRP43**

Lysogeny broth containing 100 µg/ml ampicillin was inoculated under aseptic conditions with 5% (v/v) of the bacterial culture. Incubation was carried out at 37°C with 250 rpm applied. Once the OD<sub>600</sub> reached 0.6, cells were induced with 1mM IPTG and incubation was continued at 37°C for 3.5 to 4 hours. The cells were harvested by centrifugation at 6,000 rpm. Cell pellets were resuspended and lysed by sonication. GST-cpSRP43 was bound to glutathione Sepharose then washed thoroughly with equilibration buffer (2.7mM KCl, 1.8mM KH<sub>2</sub>PO<sub>4</sub>, 15mM Na<sub>2</sub>HPO<sub>4</sub>, 137mM NaCl, pH 7.2). GST-cpSRP43 was eluted with 10mM L-Glutathione and exchanged into cleavage buffer (50 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA, 1 mM dithiothreitol, pH 7.0). Overnight cleavage was setup, in solution, at 4°C on a rocker with 10 units of PreScission Protease per liter of original cells for 16 hours.

### **Expression and purification of GST**

An empty pGEX-2T vector was transformed into BL-21-star cells and overexpressed by the same method described above for GST-cpSRP43. The harvested cells were resuspended and lysed by sonication. Supernatant was separated from cell debris was removed by centrifugation at

19,000 rpm for 30 minutes. The soluble GST protein was loaded onto a pre-equilibrated GSH-Sepharose column at a flow rate of 1 milliliter per minute. Contaminants were removed by washing the column extensively with equilibration buffer (2.7mM KCl, 1.8mM KH<sub>2</sub>PO<sub>4</sub>, 15mM Na<sub>2</sub>HPO<sub>4</sub>, 137mM NaCl, pH 7.2). The GST protein was eluted using 10 mM reduced glutathione (Sigma, St. Louis, MO). Glutathione was subsequently removed by concentration and buffer exchanging. Protein concentration was quantified by Bradford estimation and purity was monitored on 15% SDS PAGE.

### **Aggregation profile of GST**

To acquire an aggregation profile of GST, pure GST was incubated for 20 minutes at 5°C increments from 25-80°C. The samples were centrifuged at each temperature increment at 13,000 rpm for 10 minutes for complete separation of the aggregated protein from the supernatant. The aggregated protein was dissolved in 8 M Urea. Loading dye was added and the samples were resolved on a 15% SDS-PAGE gel.

### **Heat Treatment of GST-cpSRP43 cleavage products**

Once the cleavage was complete, the cleavage products were heated for 20 minutes at 5°C increments from 25-80°C. The samples were centrifuged at each temperature increment at 13,000 rpm for 10 minutes for complete separation of the aggregated protein from the supernatant. The aggregated protein was dissolved in 8 M Urea. Loading dye was added and the samples were resolved on a 15% SDS-PAGE gel.

### **Heat Treatment of Independently purified GST and cpSRP43 proteins**

Pure GST and cpSRP43 were placed in separate centrifuge tubes. The two proteins were then combined in a 1:1 ratio in a third centrifuge tube. These samples were then heat treated as described above. At each temperature increment, the samples were centrifuged, and the insoluble

material and supernatant were separated. The supernatant was immediately separated from any aggregated protein that appeared after centrifugation. These samples were resolved on 15% SDS-PAGE gels.

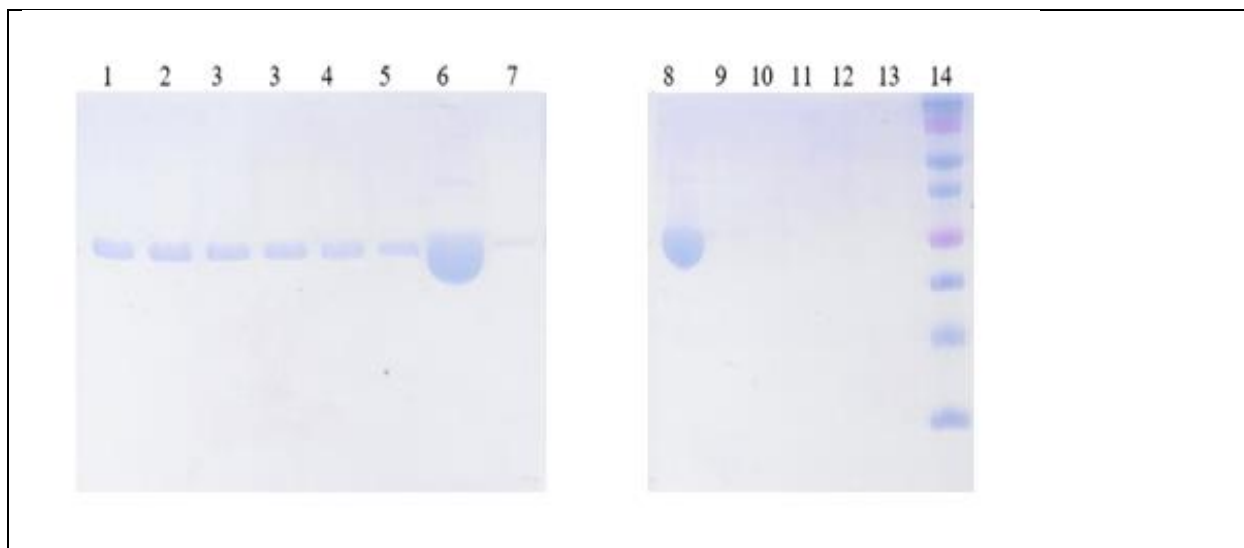
### **Heat Treatment of GST-cpSRP43 cleavage products in NaCl and Urea**

The cleaved sample was divided into two centrifuge tubes. NaCl was added to one tube to make the concentration 0.5 M NaCl. Urea was added to the other tube in order to reach the concentration 0.5 M Urea. Both tubes were heat treated at 5°C increments for 20 minutes at each temperature from 25-80°C. Separation of supernatant and insoluble material was acquired by centrifugation and samples were resolved on 15% SDS-PAGE gels as above. Gels were analyzed using UN-ScanIT software (Silk Scientific Inc.).

### **Results and Discussion**

#### ***GST is completely aggregated at 55°C***

GST begins to aggregate at 50°C and is completely aggregated at 55°C as seen in lanes 6 and 8 which corresponds to the pellets obtained at these temperatures, respectively. This is consistent with the findings from Kaplan et al. where they reported that GST lost enzymatic activity and had a melting temperature at 51°C. The GST tag can be completely removed from the supernatant at 65°C as visualized in lane 10 (Fig.2).

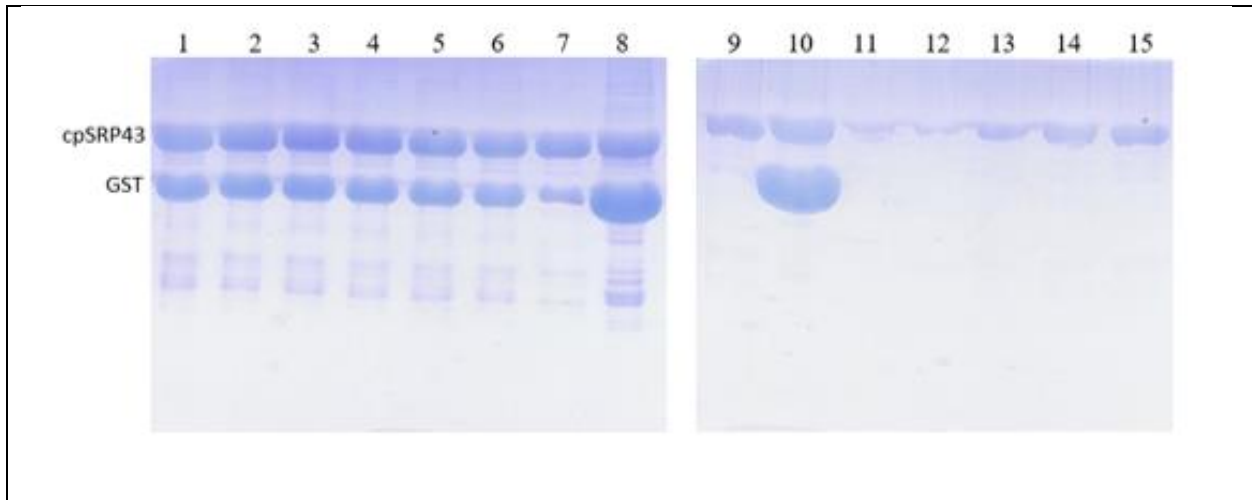


**Fig.2: SDS-PAGE of GST heat stability. Lane 1= 25°C Supernatant, Lane 2= 30°C Supernatant, Lane 3= 35°C Supernatant, Lane 4= 40°C Supernatant, Lane 5= 45°C Supernatant, Lane 6= 50°C Pellet, Lane 7= 50°C Supernatant, Lane 8= 55°C Pellet, Lane 9= 55°C Supernatant, Lane 10= 65°C Supernatant, Lane 11= 70°C Supernatant, Lane 12= 75°C Supernatant, Lane 13= 80°C Supernatant, Lane 14= Prestained Protein Marker**

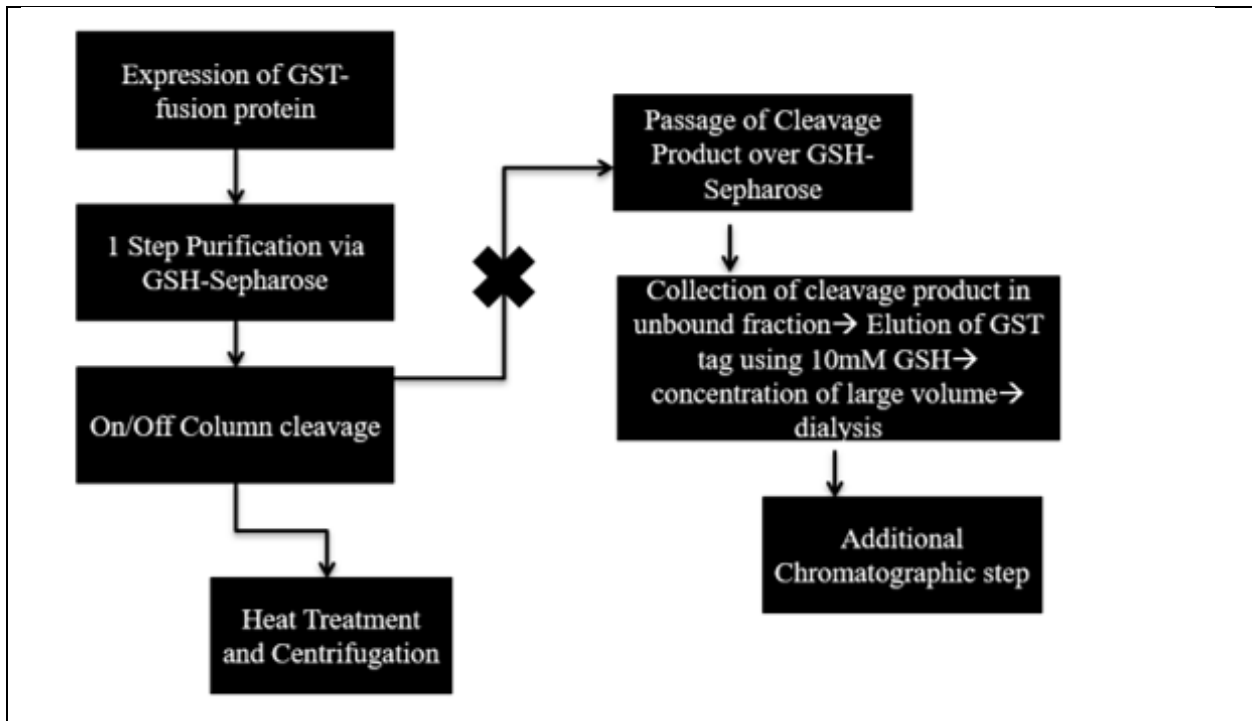
***Purification of the cleaved fusion product using the heat treatment procedure yields pure cpSRP43***

GST and cpSRP43 were both present in the sample from which was heated from 25-50°C. At 50 and 55°C the GST began to aggregate and a pellet was visible. When visualized on the gel it is apparent that a portion of the cpSRP43 was also present in the pellet which is possibly due to an association between GST and cpSRP43 even after separation by enzymatic cleavage causing cpSRP43 to be lost in the pellet along with the aggregation of GST. It is unclear at this point whether the association between these two proteins is driven by electrostatic or hydrophobic forces. A significant amount of cpSRP43 was lost along with the GST however protein left after 60°C is pure cpSRP43 as seen in lanes 11-15 (Fig.3). The procedure was then developed for the rapid heat treatment of the sample upon enzymatic cleavage. A comparison of this procedure compared to the conventional purification method is outlined in the flow chart

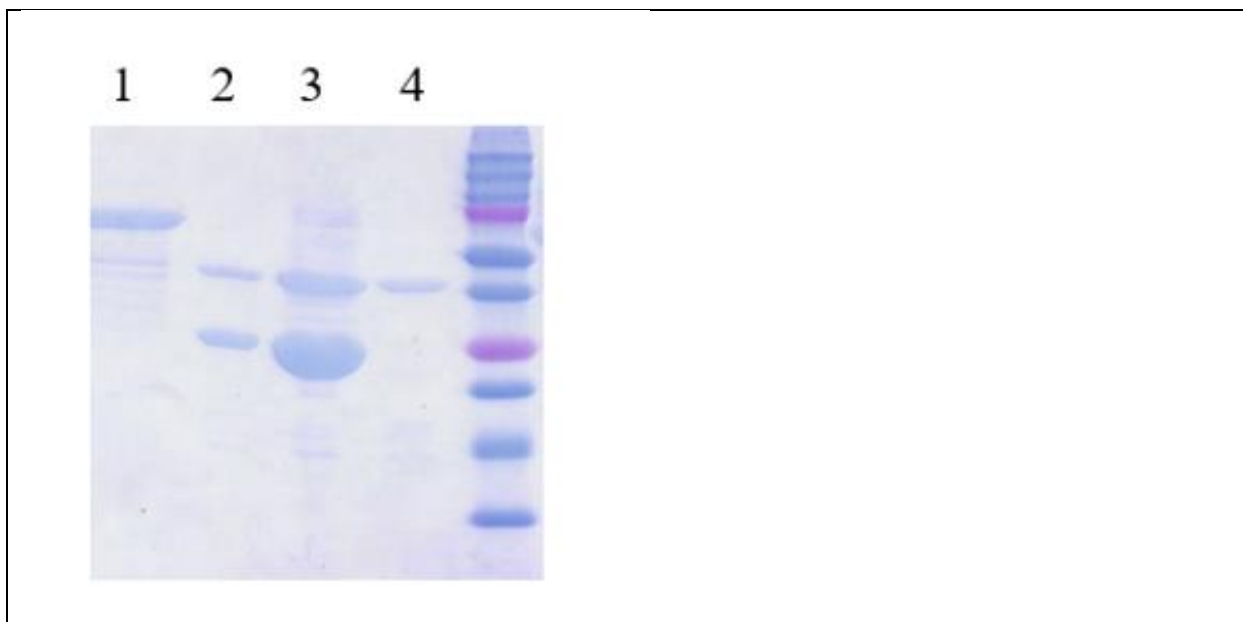
(Fig.4). The heat treatment at 65°C was subsequently carried out directly on the cleavage product (Fig. 5).



**Fig. 3: SDS-PAGE gels of cleaved GST-cpSRP43 subjected to incremental heat treatments**  
 Lane 1= cleavage product, Lane 2= 25°C Supernatant, Lane 3= 30°C Supernatant, Lane 4= 35°C Supernatant, Lane 5= 40°C Supernatant, Lane 6= 45°C Supernatant, Lane 7= 50°C Supernatant, Lane 8= 50°C Pellet, Lane 9= 55°C Supernatant, Lane 10= 55°C Pellet, Lane 11= 60°C Supernatant, Lane 12= 65°C Supernatant, Lane 13= 70°C Supernatant, Lane 14= 75°C Supernatant, Lane 15= 80°C Supernatant



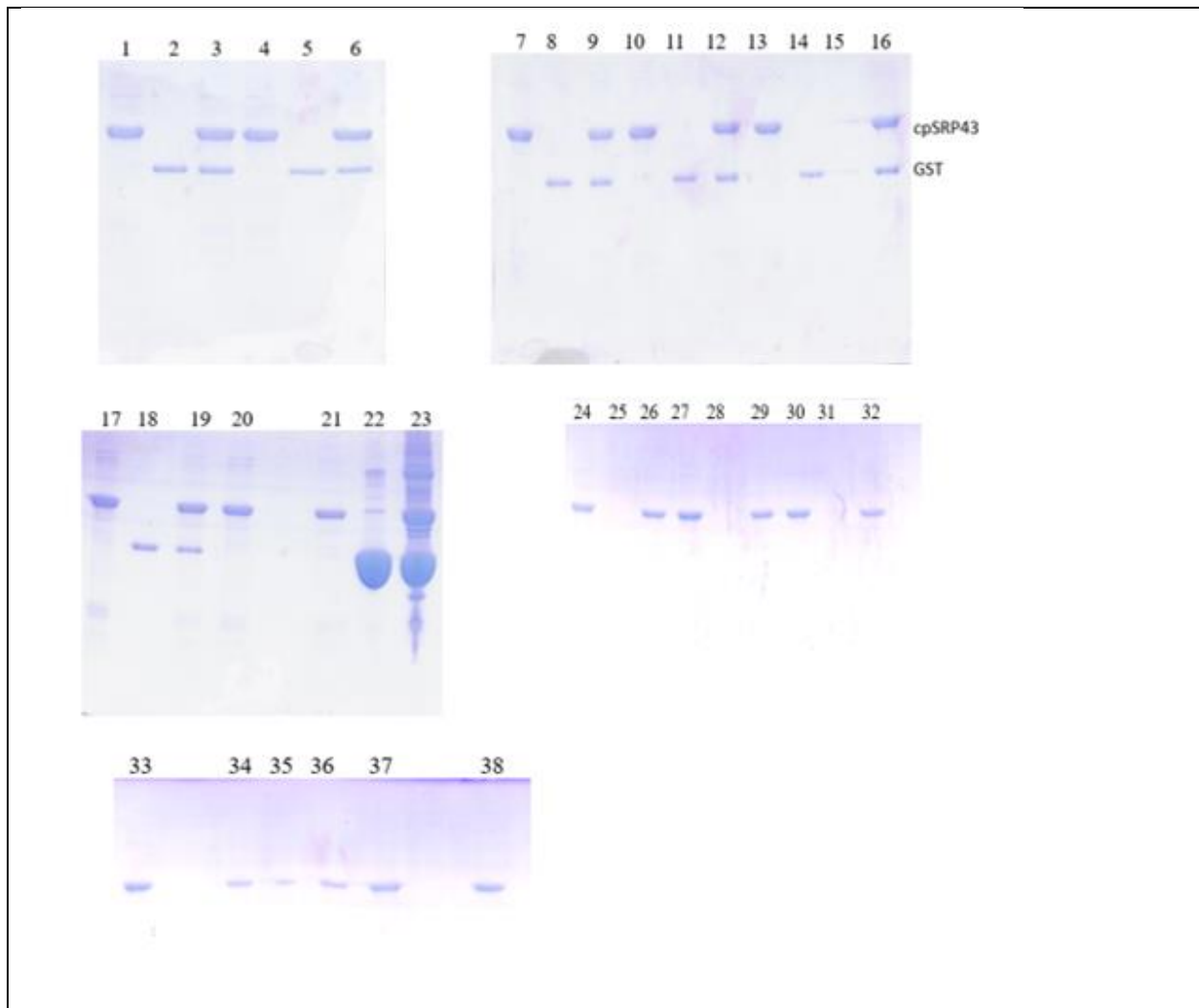
**Fig. 4: Flow chart comparing conventional purification method and the heat treatment method.**



**Fig.5: Direct heat treatment of cleavage product. Lane 1= GST-cpSRP43 eluted from GSH Sepharose, Lane 2= Cleavage product, Lane 3= pellet of post 70°C HT, Lane 4= supernatant of post 70°C HT.**

***Heat Treatment applied to combined, independently purified GST and cpSRP43***

GST and cpSRP43 were purified independently of one another and GST alone, cpSRP43 alone and the proteins in combination were subjected to the same heat treatment procedure done on the cleavage product (Fig.6). GST aggregated alone and in combination with cpSRP43 at 55°C as seen in lanes 22 and 23 below, respectively. A significant amount of cpSRP43 is still pulled down into the insoluble layer at 55°C as in the heat treatment procedure of the cleavage product. cpSRP43 and GST appear to make the same associations with each other whether they are purified independently and combined or together in the same tube as a result of cleavage. Further investigation into the forces driving this association is examined next.

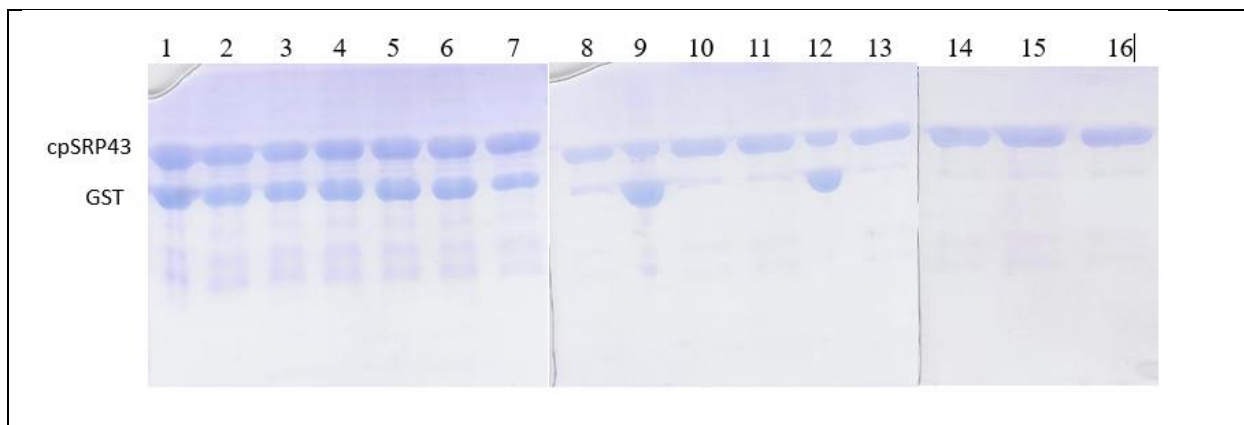


**Fig.6: SDS-PAGE gels of the Heat Treatment of purified GST and cpSRP43 proteins. Lane 1= cpSRP43 (43) 25°C Supernatant, Lane 2= GST 25°C Supernatant, Lane 3= 43 + GST 25°C Supernatant, Lane 4= 43 30°C Supernatant, Lane 5= GST 30°C Supernatant, Lane 6= 43 + GST 30°C Supernatant, Lane 7= 43 35°C Supernatant, Lane 8= GST 35°C Supernatant, Lane 9= 43 + GST 35°C Supernatant, Lane 10= 43 40°C Supernatant, Lane 11= GST 40°C Supernatant, Lane 12= 43 + GST 40°C Supernatant, Lane 13= 43 45°C Supernatant, Lane 14= GST 45°C Supernatant, Lane 15= 43 + GST 45°C Supernatant, Lane 16= 43 50°C Supernatant, Lane 17= GST 50°C Supernatant, Lane 18= 43 + GST 50°C Supernatant, Lane 19= 43 55°C Supernatant, Lane 20= GST 55°C Supernatant, Lane 21= 43 + GST 55°C Supernatant, Lane 22= GST 55°C Pellet, Lane 23= 43 + GST 55°C Pellet, Lane 24= 43 60°C Supernatant, Lane 25= GST 60°C Supernatant, Lane 26= 43 + GST 60°C Supernatant, Lane 27= 43 65°C Supernatant, Lane 28= GST 65°C Supernatant, Lane 29= 43 + GST 65°C Supernatant, Lane 30= 43 70°C Supernatant, Lane 31= GST 70°C Supernatant, Lane 32= 43 + GST 70°C Supernatant, Lane 33= 43 75°C Supernatant, Lane 34= GST 75°C Supernatant, Lane 35= 43 + GST 75°C Supernatant, Lane 36= 43 80°C Supernatant, Lane 37= GST 80°C Supernatant, Lane 38= 43 + GST 80°C Supernatant.**

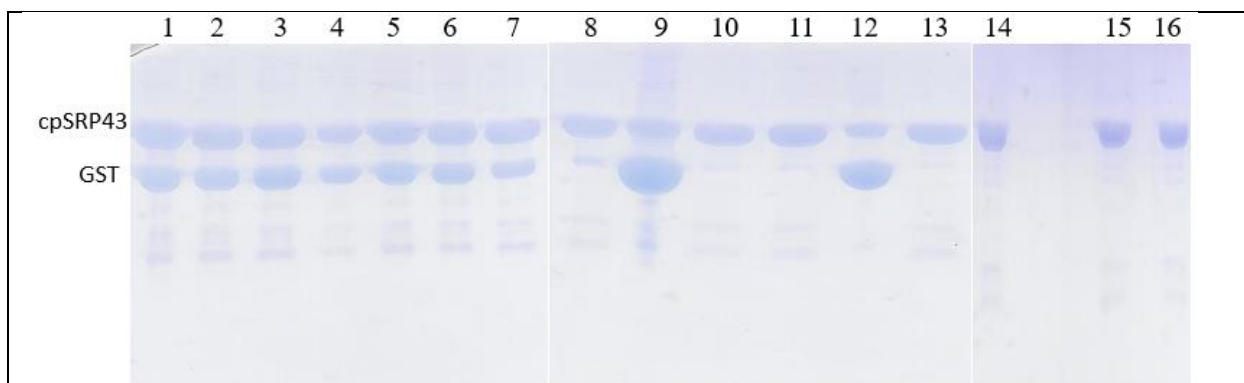


### ***Heat Treatment procedure of Cleaved GST-cpSRP43 in NaCl and Urea***

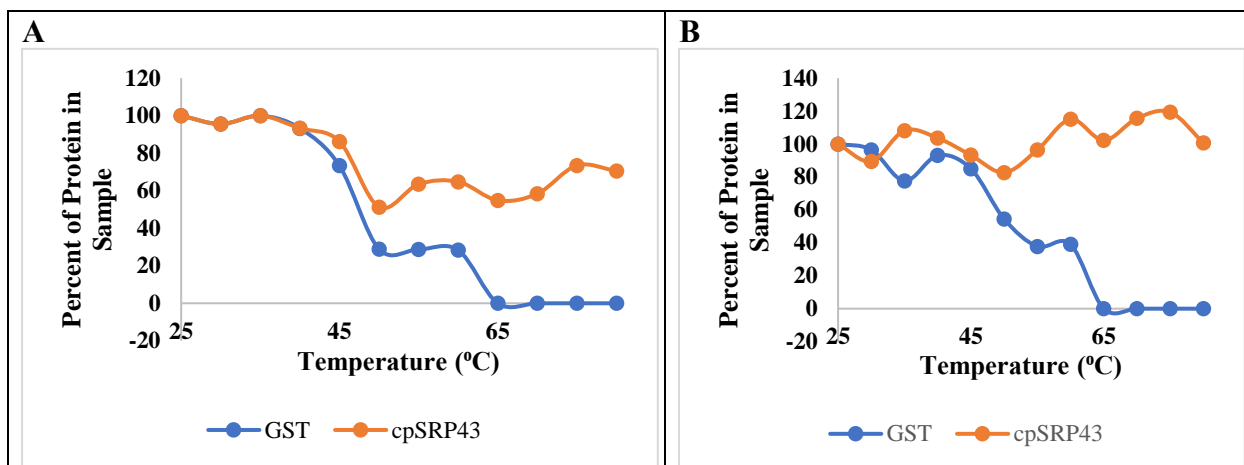
To determine if the association between cpSRP43 and GST was being driven by electrostatic or hydrophobic forces the incorporation of salt and a denaturant was added separately and in combination to the samples during the heat treatment procedures. cpSRP43 was successfully purified when subjected to a gradual heat treatment in the presence of 0.5 M Urea and 0.5 M NaCl. The GST aggregated between 50-60°C however, some cpSRP43 was still also being aggregated along with the GST between 50 and 60°C. The amount of cpSRP43 that was lost under the new conditions although was less than the amount that was lost in the heat treatment performed in the regular buffer. Figure 9 shows the densitometric scans of the supernatant bands on the SDS-PAGE gels from Figures 7 and 8. The presence of a GST band disappears at higher temperatures as the cpSRP43 is still present. A decent quantity of cpSRP43 can be retrieved under both conditions however it appears that there is a higher yield of cpSRP43 in the heat treated sample which contained 0.5M Urea which perhaps suggest that the association occurring between GST and cpSRP43 is predominately caused by the formation of hydrophobic bonds between the two proteins. In an attempt to retrieve the maximum yield of cpSRP43 post heat treatment we decided to use a constant Urea concentration (2M) and apply an increasing concentration of NaCl (1-3.5M). The gels from this experiment were analyzed by densitometric scanning and a comparison was made between the cpSRP43 protein which was co-aggregated with GST after subjection to 50°C and the cpSRP43 which was remaining in the soluble layer after subjection to 50°C (Fig.10). It appears that the optimal condition for the highest yield of cpSRP43 in the soluble layer was under the conditions of 2M Urea and 1M NaCl and increasing the NaCl concentration did not appear to improve yield. This conclusion supports the claim that the association between cpSRP43 and GST is primarily hydrophobic.



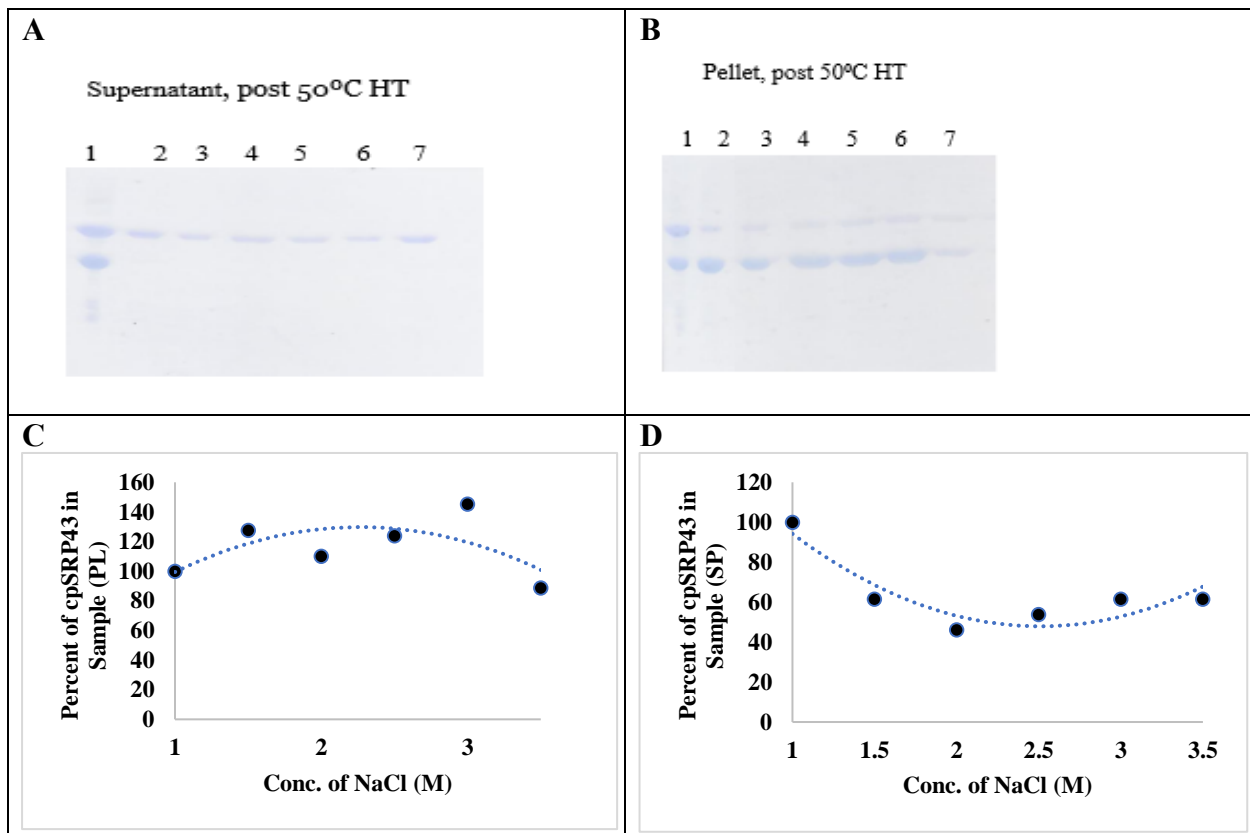
**Fig. 7: SDS-PAGE gels of cleaved GST-cpSRP43 subjected to incremental heat treatments in 0.5 M NaCl. Lane 1= cleavage product, Lane 2= post incubation in 0.5 M NaCl, Lane 3= 25°C Supernatant, Lane 4= 30°C Supernatant, Lane 5= 35°C Supernatant, Lane 6= 40°C Supernatant, Lane 7= 45°C Supernatant, Lane 8= 50°C Supernatant, Lane 9= 50°C Pellet, Lane 10= 55°C Supernatant, Lane 11= 60°C Supernatant, Lane 12= 60°C Pellet, Lane 13= 65°C Supernatant, Lane 14= 70°C Supernatant, Lane 15= 75°C Supernatant, Lane 16= 80°C Supernatant.**



**Fig.8: SDS-PAGE gels of cleaved GST-cpSRP43 subjected to incremental heat treatments in 0.5 M Urea. Lane 1= cleavage product, Lane 2= post incubation in 0.5 M Urea, Lane 3= 25°C Supernatant, Lane 4= 30°C Supernatant, Lane 5= 35°C Supernatant, Lane 6= 40°C Supernatant, Lane 7= 45°C Supernatant, Lane 8= 50°C Supernatant, Lane 9= 50°C Pellet, Lane 10= 55°C Supernatant, Lane 11= 60°C Supernatant, Lane 12= 60°C Pellet, Lane 13= 65°C Supernatant, Lane 14= 70°C Supernatant, Lane 15= 75°C Supernatant, Lane 16= 80°C Supernatant**



**Fig. 9: Densitometric analysis of SDS-Page gels for the cleavage product in 0.5M NaCl (Panel-A) and in 0.5M Urea (Panel-B).**



**Fig. 10: SDS-PAGE of post heat-treated supernatant (Panel-A) and pellet (Panel-B)3, of the cleavage product in a combination of constant Urea concentration and increasing NaCl concentrations. Lane 1= Cleavage Product, Lane 2= 2M Urea, 1M NaCl, Lane 3= 2M Urea, 1.5M NaCl, Lane 4= 2M Urea, 2M NaCl, Lane 5= 2MUrea, 2.5M NaCl, Lane 6= 2M Urea, 3M NaCl, Lane 7= 2M Urea, 3.5M NaCl. Densitometric analysis of the cpSRP43 band which was co-aggregated along with GST (Panel-C) and the band of cpSRP43 which was remaining in the supernatant (Panel-D) at the different concentrations of Urea and NaCl post heat-treatment at 50°C.**

### Summary and Conclusions

GST is a commonly used affinity tag which is known for its ability to help solubilize its fusion partner (4,10,11). GST can be selectively precipitated out of a sample by applying a temperature of 65° C. This purification method was able to work for the large cpSRP43 protein due to its innate heat stability. Upon heat treatment of the GST-cpSRP43 cleavage products, the GST and other small contaminants became precipitated while pure cpSRP43 was left in the supernatant. This method can potentially be used for other large proteins if they can withstand

temperatures up to 70°C. Protein purification by chromatography can be costly and time consuming and, in many cases, require multiple chromatographic steps. As additional chromatographic steps are needed for purification the overall protein yield decreases significantly. The typical method of purifying cpSRP43 involves three chromatographic steps and a sixteen-hour cleavage in order to obtain pure cpSRP43 for biophysical analysis as illustrated by the flow chart (Fig.4). This method eliminates the need for any additional chromatographic steps and speeds up the purification process significantly. The overall yield of cpSRP43 by the conventional method is typically 1 mg/ml which is approximately the same for the heat treatment method. Thereby making the heat treatment method less time consuming and more cost effective. Protein is lost during the conventional method due to the multiple column passes and possible degradation during enzymatic cleavage. During heat treatment purification, cpSRP43 is lost in the precipitate due to an association between GST and cpSRP43. To improve the protein yield from the heat treatment method, we considered the potential driving forces behind the association between GST and cpSRP43 by examining whether the addition of salt and/or denaturant would improve net protein yields. Adjustments in the concentrations of Urea and NaCl were made in order to disrupt any hydrophobic or electrostatic interactions which may be occurring between the two protein during heat treatment. However, there was not a significant yield increase found with the addition of salt or denaturant independently or with both added to the heat-treated samples. Optimization in this manner may be dependent on the candidate protein for this purification method. In cases with peptides co-aggregation did not pose an issue. cpSRP43 is a larger structured protein and given these properties, there is a higher probability of interaction between the protein and the tag. In this study, protein yields by the conventional method versus the heat-treatment method were comparable thus a reduction in time consumption for purification of cpSRP43 was gained as well

as lowered cost due to the elimination of additional chromatographic steps. This method, when applied to proper candidate proteins, could lower cost and potentially give higher protein yields.

**Acknowledgements:** This work was supported by the DOE grant DE-02-01ER15161

## References

1. Freitag R, Horváth C. Chromatography in the downstream processing of biotechnological products. *Downstream Processing Biosurfactants Carotenoids*: Springer; 1995. p. 17-59.
2. Terpe K. Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Applied microbiology and biotechnology*. 2003;60(5):523-33.
3. Zhao X, Li G, Liang S. Several affinity tags commonly used in chromatographic purification. *Journal of analytical methods in chemistry*. 2013;2013.
4. Harper S, Speicher DW. Purification of proteins fused to glutathione S-transferase. *Protein Chromatography*: Springer; 2011. p. 259-80.
5. Mitchell DA, Marshall TK, Deschenes RJ. Vectors for the inducible overexpression of glutathione S-transferase fusion proteins in yeast. *Yeast*. 1993;9(7):715-22.
6. Arnau J, Lauritzen C, Petersen GE, Pedersen J. Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins. *Protein expression and purification*. 2006;48(1):1-13.
7. Dyson MR, Shadbolt SP, Vincent KJ, Perera RL, McCafferty J. Production of soluble mammalian proteins in *Escherichia coli*: identification of protein features that correlate with successful expression. *BMC biotechnology*. 2004;4(1):32.
8. Tu CJ, Peterson EC, Henry R, Hoffman NE. The L18 Domain of Light-harvesting Chlorophyll Proteins Binds to Chloroplast Signal Recognition Particle 43. *Journal of Biological Chemistry*. 2000;275(18):13187-90.
9. Henry RL. SRP: adapting to life in the chloroplast. *Nature structural & molecular biology*. 2010;17(6):676.
10. Waugh DS. Making the most of affinity tags. *Trends in biotechnology*. 2005;23(6):316-20.
11. Lichty JJ, Malecki JL, Agnew HD, Michelson-Horowitz DJ, Tan S. Comparison of affinity tags for protein purification. *Protein expression and purification*. 2005;41(1):98-105.

## Biosafety Committee Approval

Biosafety protocol number:10044




UNIVERSITY OF  
ARKANSAS

Office of Research Compliance

January 30, 2014

MEMORANDUM

TO: Dr. Ralph Henry 

FROM: W. Roy Penney  
Institutional BioSafety Committee

RE: IBC Protocol Approval

IBC Protocol #: 10044

Protocol Title: "Protein Targeting to the Chloroplast Thylakoid Membrane"

Approved Project Period: Start Date: January 26, 2014  
Expiration Date: January 25, 2017

The Institutional Biosafety Committee (IBC) has approved the Renewal of Protocol 10044, "Protein Targeting to the Chloroplast Thylakoid Membrane". You may begin your study.

If further modifications are made to the protocol during the study, please submit a written request to the IBC for review and approval before initiating any changes.

The IBC appreciates your assistance and cooperation in complying with University and Federal guidelines for research involving hazardous biological materials.