12-2019

Single Molecule Fluorescence Studies of protein structure and dynamics underlying the chloroplast signal recognition particle targeting pathway

Dustin R. Baucom
University of Arkansas, Fayetteville

Single Molecule Fluorescence Studies of protein structure and dynamics underlying the chloroplast signal recognition particle targeting pathway

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

by

Dustin R. Baucom
Northeastern State University
Bachelor of Science in Chemistry, 2013

December 2019
University of Arkansas

This dissertation is approved for recommendation to the Graduate Council.

_____
Colin Heyes, Ph.D.
Dissertation Director

_____
Paul Adams, Ph.D.
Committee Member

_____
Jingyi Chen, Ph.D.
Committee Member

_____
Ralph Henry, Ph.D.
Committee Member

_____
Suresh Thallapuranam, Ph.D
Committee Member

Abstract

The work presented in this dissertation explores the structural dynamics in the chloroplast signal recognition particle pathway. Findings include cpSRP shows scanning functionality similar to that in the cytosolic SRP with the ribosome. The intrinsically disordered C-terminal tail of the Albino3 protein has some transient secondary structure. Upon binding to cpSRP43 in solution, separate secondary structure formation was identified in the C-terminal tail of Albino3. Finally, to increase efficiency of analyzing fluorescence time traces for this work, a modular software was produced.

Table of Contents

Chapter 1: Introduction

1.1 Ubiquitous Protein Targeting via Signal Recognition Particle (SRP) pathways

The signal recognition particle (SRP) pathway is a ubiquitous protein targeting pathway whereby variants have been described in eukarya, archaea, bacteria and plants.[1] Some of these pathways are co-translational and some are post-translational. In co-translational targeting, SRP consists of a complex of six polypeptides and a bound RNA molecule originally described in mammals.[2, 3] The mammalian SRP complex was shown to perform three functions: signal recognition, elongation, and translocation. Each of these activities were shown to be the responsibility of separate portions of the complex.[4] For signal recognition, the complex interacts with nascent chains of secretory proteins, but does not interact with the nascent chains of cytosolic proteins. This differentiation in interaction arises from the presence of a signal sequence possessed on secretory proteins.[5] The subunit responsible for the recognition of the signal sequence of the nascent polypeptide chains on the ribosome was found to be the 54kDa subunit, termed SRP54.[6] The arresting of elongation pauses the translation of the nascent polypeptide to allow for translocation. The 9kDa and 14kDa subunits, SRP9 and SRP14 respectively, exist as a heterodimer and were found to be responsible for this arrest in elongation.[7, 8] For targeting of the ribosome-nascent polypeptide chain complex, the portion of the SRP implicated was the heterodimer consisting of 68kDa and 72kDa subunits, SRP68 and SRP72 respectively.[7, 8] The final protein subunit in the SRP complex SRP19 was shown to be critical for the interaction of SRP54 with the SRP.[8]

The primary role of the SRP pathway is to target the ribosome-nascent chain complex to a membrane. In the mammalian SRP pathway, the target of the SRP complex

is the SRP receptor.[9, 10] Upon interaction formation of a new complex that includes the SRP bound to the ribosome nascent chain complex and SRP receptor, the complex interacts with a translocon in a GTP driven manner to transfer the ribosome-nascent chain complex to the translocon and recycle the SRP complex and SRP receptor.[11, 12] The translocon is the protein machinery that transports the protein into or through the membrane. This machinery is the Sec61 complex in eukaryotes that allows protein transport across and into the endoplasmic reticulum.[13]

SRP pathways in archaea and bacteria are similar to the pathway in eukaryotes, but differ in the composition of the SRP, the SRP receptors, and can contain alternate pathways.[1] Bacteria only possess a homologue for the signal recognition subunit SRP54. This lone, conserved subunit is termed fifty-four homologue (Ffh).[14] Along with a structural variation of the SRP RNA, Ffh makes up the bacterial SRP.[15] For archaea, the SRP contains an SRP54 homolog, a SRP19 homolog, and SRP RNA. The differences in the SRP composition is shown in Figure 1.1.[16] Additionally, the SRP receptor is not present in bacteria and archaea. Instead, FtsY, a homologue of one of the two SRP receptor subunits performs the role of the receptor.[17] FtsY is not anchored to the membrane being targeted and has not been shown to have a specific membrane receptor.[18] However, studies have shown the capacity of FtsY to interact directly with the membrane and the SecYEG, the bacterial translocon.[19-21] One or both of these interactions may play a crucial role in the transient association of FtsY to the membrane.[22]

The discussion of the SRP systems has focused on the transport into and across membranes using the Sec machinery. It was discovered in bacteria and archaea that some

Sec independent membrane proteins are inserted by the YidC insertase.[23] This insertase has been shown to be able to act independently of and in conjunction with the Sec complex.[24-26] This variation of the pathway increases the complexity of the SRP targeting.

1.2 cpSRP pathway

A better understanding of the role of the different SRP targeting pathways may be gleaned by studying and comparing to the chloroplast SRP targeting pathway, which has some similarities and some differences to the eukaryotic and prokaryotic systems. Light harvesting chlorophyll-binding protein (LHCP) is produced in the cytosol and imported into the stroma of the chloroplast via an attached carrier peptide. This carrier peptide is then cleaved and the LHCP is targeted to the thylakoid membrane for insertion.[27] The exact process and molecular interactions to facilitate this posttranslational targeting once in the stroma were, until recently, poorly understood. Although progress to has been made to study the many interactions in this pathway, there are still several mysteries. The current understanding of the pathway is summarized/illustrated in Figure 1.2.[28]

A homologue of SRP54 was found in the chloroplast of *Arabidopsis thanliana*.[29] This homologue would later be name chloroplast SRP54 (cpSRP54).[30] The discovery of this homologue suggested the presence of an SRP targeting pathway in the chloroplast. Further experiments showed that LHCP required a stromal factor to maintain solubility and to be integrated into the thylakoid membrane of the chloroplast.[31, 32] To maintain solubility in the stroma, LHCP formed an approximately 120kDa complex.[33] Since RNAase showed no effect on this soluble factor and the targeting was posttranslational, an SRP mediated pathway was not initially suspected. The factor was strongly affected

by protease indicating that it was likely a protein.[32] Subsequently immunoprecipitation and crosslinking studies found that cpSRP54 plays a role in the integration of LHCP into the thylakoid membrane and was present in the LHCP complex.[34] These finding suggested that cpSRP54 could operate without bound RNA, in contrast with known examples from other SRP pathways. The composition of the cpSRP complex was also found to contain a novel 43kDa subunit, cpSRP43. This unique subunit of the chloroplast SRP was shown to form a transit complex with cpSRP54 and LHCP. This transit complex was also shown to be inactive in LHCP integration without the stroma present. This indicated that a second stromal factor was required for the integration of LHCP. It was also shown that approximately half of the cpSRP54 was bound to ribosomes and that cpSRP43 does not interact with ribosomes, strongly suggesting that cpSRP54 is involved in co-translational targeting that does not involve cpSRP43 as well as the post-translational targeting pathway that does involve cpSRP43.[30]

In vivo experiments have explored the result of mutations to lower or eliminate the population for various proteins involved in the cpSRP pathway. Reduction of cpSRP54 showed juvenile plants with yellow leaves. As the plants matured, the phenotypes started to resemble the wild-type.[35] A complete knockout mutant of cpSRP54 showed similar results to the reduction experiment. Further, knockout of cpSRP43 showed a different phenotype than the cpSRP54 knockout. The difference in the phenotypes suggest that the subunits of cpSRP may have roles that are separate from the cpSRP complex. In general, knockout of cpSRP54 was more severe and is suspected to interact in other pathways.[36, 37] Mutants lacking both cpSRP43 and cpSRP54 presented additive negative effects on the phenotype.[38]

The components of the stroma required for LHCP integration into the thylakoid membrane were found to be GTP and a homologue of the bacterial FtsY, cpFtsY.[39, 40] Similar to the mechanism in bacteria, cpFtsY interacts with the cpSRP and targets the complex to the thylakoid membrane.[39] Knockouts experiments of cpFtsY were performed to examine the phenotypes of plants missing the membrane targeting component of the pathway. It was found that mutants of *Arabidopsis thanliana* lacking cpFtsY were yellow and isolated thylakoids contained 10-33% as much LHCP when compared to the wild-type. These mutants were seedling-lethal. When given an external carbon source, the plants were yellow and stayed yellow throughout maturity. This evidence indicates that a cpFtsY-null mutant is more lethal than the double knockout of the whole cpSRP complex. One possibility to explain this result is that a large fraction of LHCP becomes trapped by cpSRP, but never targets to the membrane due to the lack of cpFtsY. This would suggest that some LHCP can bypass the transit complex formation when it is not present.[41] Interestingly, a knockout of cpFtsY in conjunction with a cpSRP54 knockout showed increased integration of LHCP as compared to the cpFtsY knockout alone. This would suggest that cpSRP43 may have some way of directly targeting and inserting LHCP into the thylakoid, albeit less efficiently.[42]

Although the chloroplast possesses a Sec pathway, cpSec, it was found that the cpSRP pathway for the integration of LHCP was unaffected when antibodies raised against cpSec were used.[43] This strongly suggested that the cpSec machinery was unnecessary in and independent of the cpSRP pathway. The cpSRP pathway was then found to involve the insertase Albino3, a homologue of YidC found in bacteria and Oxa1p found in mitochondria.[44] This insertase was shown to be required for both in vitro

and in vivo LHCP insertion. The phenotypes for the *Arabidopsis thanliana* knockout of the albino3 protein were white-yellow plants that were seedling-lethal. It was found that the leaves contained less than 10% of the LHCP present in the wild-type. This result, however, may be attributed to additional roles of Alb3. For example, Alb3 is found in significant concentrations in the roots of plants. This suggests there is a role of Alb3 beyond LHCP integration in the thylakoid membrane.[41]

1.3 Structural information

In order to understand specific molecular binding events and interactions in the cpSRP pathway, a general overview of the structures of proteins involved follows. cpSRP43, the novel subunit of cpSRP is composed of three chromatin organization modifier domains (chromodomains) and four ankyrin repeats.[36, 45] Interestingly, cpSRP43 was the first non-nuclear protein identified to possess chromodomains.[46] The ankyrin repeats are very common and typically serve as protein-protein interaction motifs.[47] A crystal structure of cpSRP43 revealed a negatively-charged backbone similar to that found in SRP RNA. Given this, cpSRP43 has been proposed to act as a direct replacement for SRP RNA.[48]

cpSRP54 consists of a conserved GTPase domain (G-domain), a conserved four helix bundle (N-domain), and a methionine-rich domain (M-domain). The N and G domains are consistent with Ffh and cytosolic SRP variants. The known crystal structures for Ffh and mammalian SRP54 show the M-domain in different positions. To solve for the relative positioning of the M-domain in cpSRP54, single molecule Förster resonance energy transfer was used to determine the position relative to the N-domain.[49]

For the membrane associated proteins, cpFtsY contains a conserved amphipathic helix near its N-terminus (N-domain) that mediates binding to the thylakoid membrane.[50] Based on crystal structure evidence, the conserved GTPase domain (G-domain) in cpFtsY was shown to have a more extensive interaction with the N-domain than in bacterial FtsY. This domain interaction in the free cpFtsY leads to a conformation that is similar to that of the cpFtsY-cpSRP54 bound state.[51] Albino3 is a membrane protein embedded in the thylakoid of chloroplasts. It contains 6 transmembrane helices, an N-terminal tail that is exposed to the lumen, and a long C-terminal tail that is exposed to the stroma. That contains a C-terminal soluble tail. This C-terminal tail has been proposed to be intrinsically disordered,[52, 53] although results in this dissertation will question this hypothesis.

LHCP, the substrate of the SRP targeting pathway, is synthesized in the cytosol as a preprotein with a carrier peptide. Upon entering the stroma, this carrier peptide is cleaved.[54] Once integrated into the thylakoid membrane, the structure of LHCP consists of three transmembrane helices.[48]

1.4 Molecular interactions

Despite the similarities between the cytosolic and chloroplast SRP54 subunit, it has been shown that the cytosolic forms such as bacterial Ffh cannot bind cpSRP43.[55] The unique interaction site on cpSRP54 that facilitates binding to cpSRP43 was identified as a 10-amino acid motif near the C-terminal. Sequence alignments confirmed the conservation of this motif amongst the six cpSRP homologues tested and the its absence in E. coli SRP54 homologue.[56] Later, this C-terminal binding extension motif was found to be predominantly disordered in nature.[49] The second chromodomain of cpSRP43

(CD2) was implicated as the primary binding region to cpSRP54.[57] To establish this binding interface, it was shown that CD2 had to undergo structural changes.[58]

A specific sequence in LHCP was found to be required for transit complex formation. cpSRP was shown to require an L18 motif and an appropriate hydrophobic domain for binding LHCP. Other proteins with a hydrophobic domain fused with L18 could also make transit complex, which suggests the L18 motif regulates the binding of cpSRP to hydrophobic proteins.[59] It was found that cpSRP43 binds this L18 domain and that cpSRP54 is primarily responsible for binding the hydrophobic domain. It was found that L18 and cpSRP54 are not competitors for binding of cpSRP43, highlighting different binding sites between the 3 proteins.[60] In fact, it was later discovered that cpSRP43 enhances L18 binding to cpSRP43.[61]

The membrane/stroma partitioned cpFtsY has been shown to be involved in targeting the transit complex to the thylakoid membrane for LHCP insertion.[39, 40] Three interactions for cpFtsY have been described. First, it has been shown that the N-terminal region of cpFtsY partially inserts into the thylakoid membrane. Deletion of this region results in the inability for the protein to bind the membrane. The nature of this membrane interaction seems to be generically non-specific to lipids and not only the thylakoid membrane.[50] The specificity of the interaction with the thylakoid membrane must therefore come from interactions with thylakoid-bound proteins or other molecular partners. This interaction likely comes from cpFtsY's association with Alb3 due to their close connection in the SRP targeting pathway. It was shown that cpFtsY and cpSRP both individually interact with Alb3, but the complex of cpSRP and cpFtsY is more efficient in this interaction.[62] This interaction was shown to only require cpSRP54 and

cpFtsY. Further, it was shown that using a non-hydrolyzable analogue of GTP, GMP-PNP, that the complex of cpSRP and cpFtsY were locked into the interaction with Alb3.[62] This indicates that the hydrolysis of GTP is necessary for the recycling of the cpSRP-cpFtsY complex back into the soluble form.[62] Lastly, the interaction of cpFtsY with cpSRP or the transit complex has been shown to be dependent only on cpSRP54. One key difference in their binding versus bacterial homologues is that the chloroplast versions bind 200 times faster than their bacterial counterparts.[63]

cpSRP43 has been shown to have an interaction with the C-terminal tail of Alb3.[52] This interaction is thought to have some effect on the targeting of cpSRP to the membrane via the interaction with cpSRP43, although the details are still unclear. The nature of this cpSRP43-Alb3 interaction has been conflictingly reported as taking place in the chromodomains of cpSRP43 and the ankyrin repeats. Lewis *et al.* report that the primary interaction occurs in the ankyrin repeats. In addition to this, they showed that CD3 can be deleted without any functional consequences.[64] However, at the same time Falk *et al.* reported that CD2-CD3 are responsible for the interaction between cpSRP43 and the C-terminal tail of Alb3.[52] This group subsequently provided detailed structural evidence to support their claim,[65] but they did not address the earlier functional studies of CD3 deletion or the claims of ankyrin binding.[64] As of yet, this debate has not been resolved.

1.5 Intrinsically disordered proteins

Intrinsically disordered proteins (IDPs) were first classified based on circular dichroism data. The disordered description may refer to a whole protein or region of a protein.[66] The term “intrinsically disordered” refers to a protein or part of a protein that

does not have well defined secondary structure based on its intrinsic sequence.[67] Although IDPs were first described over a half century ago, investigations into their structure-function relationship have increased dramatically over recent years.[68] Figure 1.3 shows the results of published articles containing synonyms for IDPs by year. The search is updates a previously published search of the query: "(inherently OR natively OR intrinsically) AND (disordered OR unfolded OR unstructured) AND protein."[68]

For many years in early structural biology, the paradigm was that the function of a protein was strictly related to its 3-dimensional, folded structure.[69] With the advent of increased genome sequencing and protein structural prediction software, this idea has been challenged.[70] By analyzing protein sequences from large databases, neural networks trained to identify structure in proteins revealed that DNA encodes for a large portion of amino acid sequences with little propensity towards structure.[71, 72] Another study analyzed the amount of proteins containing long, intrinsically disordered segments and found that the percentage of such proteins varied by domain of life. Interestingly, 33% of eukaryotic proteins possess such segments. Archaean and bacterial domains of life showed only 2% and 4.3% respectively.[73] In recent research, IDPs have been implicated in many functional roles. Some of the many examples of these roles include regulation, signaling, and chaperones.[68]

Since there is no fixed 3-D structure, an ensemble of many transient conformations exists for IDPs. The relative populations of each conformation depend on the distinct energetics of the protein.[74] This presents an experimental problem for any technique that measures an ensemble value. This ensemble value will not give

information about a relevant, physical structure, but will only provide information about the hypothetical average structure.

1.6 smFRET in the study of protein structure, protein dynamics and IDPs

Currently the most popular techniques to study IDPs use nuclear magnetic resonance (NMR), small-angle x-ray scattering (SAXS), and molecular dynamics simulations (MD).[68] Although NMR and SAXS have proved highly useful in identifying intrinsically disordered regions and providing constraints on the possible structures, there are some problems with obtaining detail conformational information since the results are reflective of the highly heterogeneous ensemble.[68] Molecular dynamics simulations can provide individual conformations that make up the ensemble and is therefore very useful in characterizing IDPs. However, the computational complexity in simulating disordered regions in apoproteins and the IDP upon binding to substrate is quite high and this computational cost can be prohibitive.[75] Single-molecule Förster resonance energy transfer (smFRET) can help to solve these problems by looking at individual conformations instead of an ensemble and, when used in conjunction with computational techniques, help to reduce the computational complexity.

FRET is a distance dependent phenomenon that involves the transfer of energy from an excited fluorophore (donor) to a second ground state fluorophore (acceptor).[76] This process is illustrated in Figure 1.4. The relationship between this energy transfer and the distance is shown in Equation 1.1:

$$\text{Equation 1.1} \quad E_{FRET} = \frac{R_0^6}{R_0^6 + r^6}$$

where E_{FRET} is the efficiency of the FRET, R_0 is the Förster distance, and r is the distance between the fluorophores. The Förster distance is an empirical value that depends on the

spectral overlap of the emission of the donor and the excitation of the acceptor. Since the FRET efficiency is inversely proportional to the sixth-power of the distance between the two fluorophores, measured FRET efficiency is extremely sensitive to small changes in distance.

Single-molecule techniques confer the advantage of examining individual molecules instead of the ensemble. This allows the examination of individual values for each measurement instead of an average, ensemble value. This can allow for the extraction of subpopulations possessing distinct values for the measurement that are not represented by the ensemble average value. For example, Kuzmenkina et al. observed distinct fold and unfolded populations of Rnase H under denaturing conditions.[77] Observing these distinct protein conformations and changes in conformations in response to stimuli would be impossible using ensemble techniques.

When using FRET at the single molecule level, distances between desired regions in proteins can be determined for individual conformations. By examining many of these conformations, subpopulations of predominant conformations can be built up. If many distance pairs are determined, 3D structure constraints can be compiled. These intramolecular distances can be used to confer conformational changes or to help reduce the computational complexity for MD simulations. By limiting the vast search spaces and possible conformations for IDPs, using known FRET distances, the simulations of IDPs becomes viable. This technique is a version of constrained MD.[78, 79]

These benefits of smFRET have been tracked in the recent increase in use with IDPs in publications. This data from a PubMed search that adds FRET to the previous IDP search illustrates this in Figure 1.5. Coupled with multiple reviews about the

significance of using smFRET to study IDPs, the consensus indicates the unique advantages of this technique.[78, 80, 81]

1.7 Figures

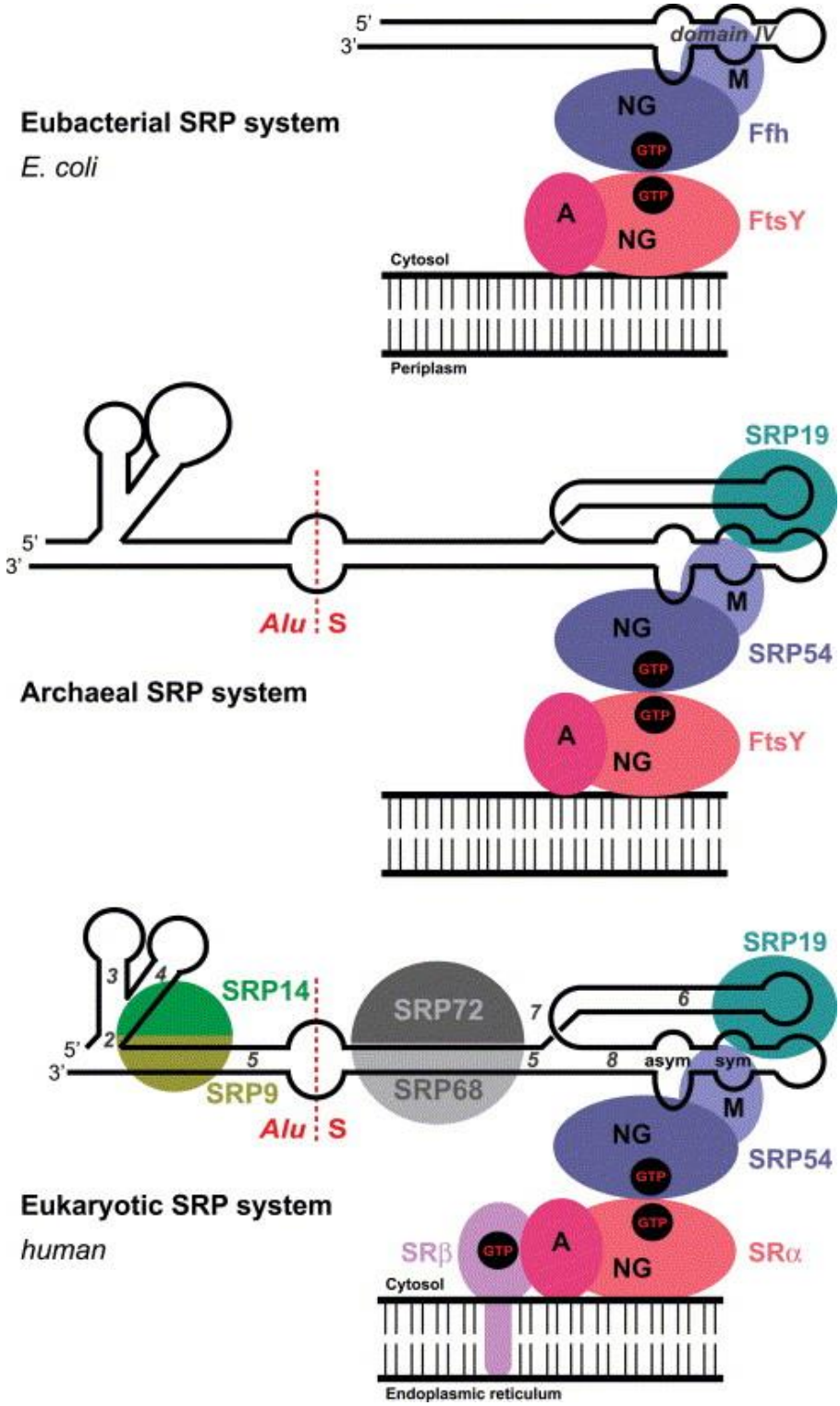


Figure 1.1 Shows the differences in cytosolic SRP across the three domains of life. Figure from SRP-mediated protein targeting: structure and function revisited.[16]

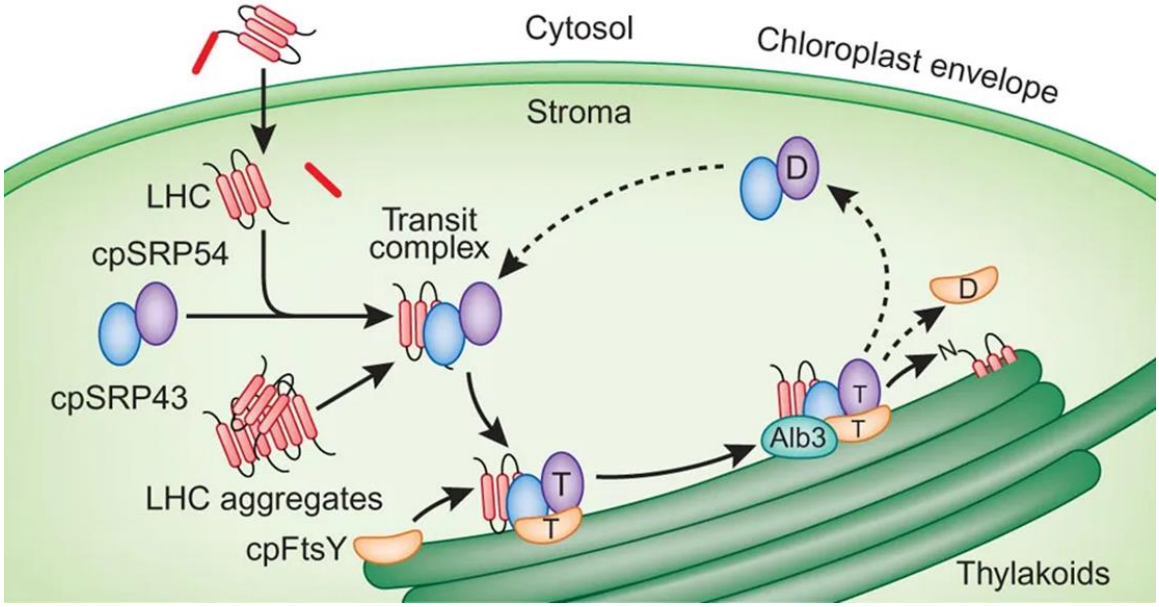


Figure 1.2 Shows the cpSRP mediated pathway for the targeting of LHCP to the thylakoid membrane. Figure from SRP: adapting to life in the chloroplast.[28]

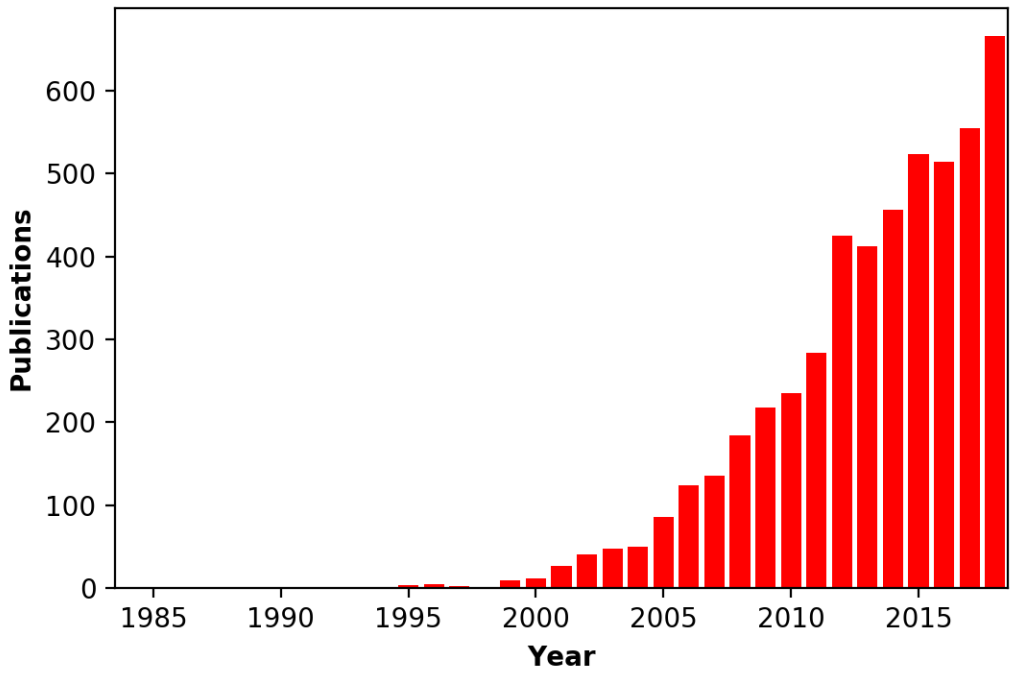


Figure 1.3 Publications by year from a search of the PubMed database using synonyms for IDPs.[68]

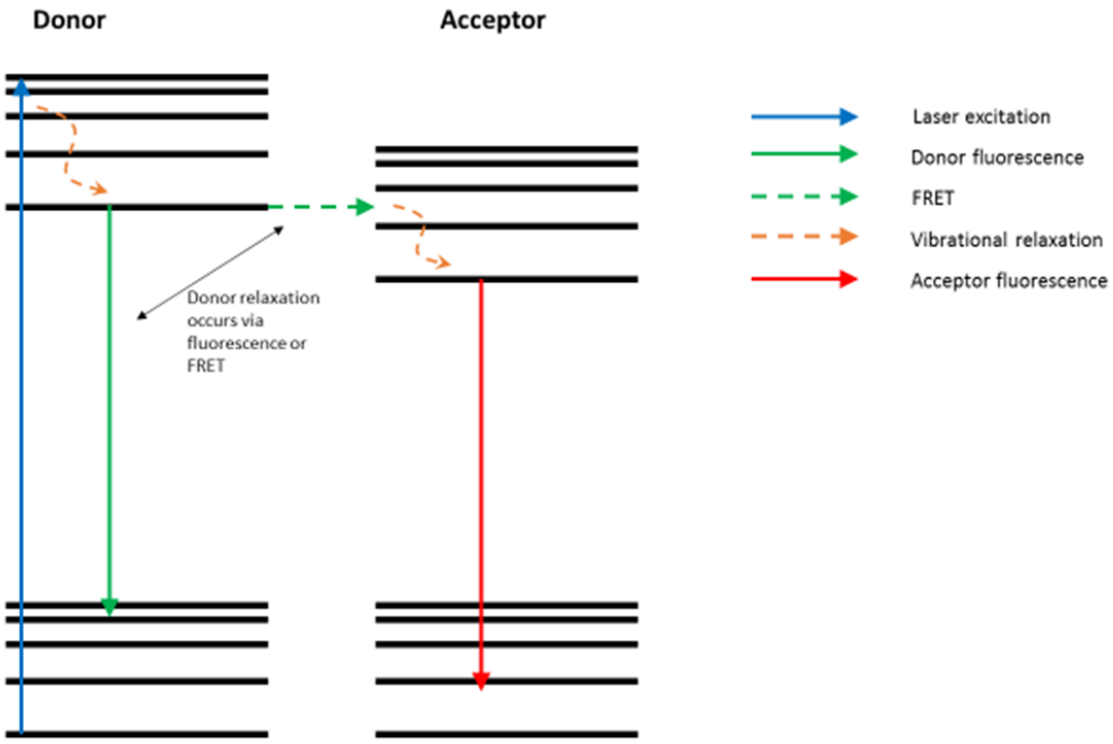


Figure 1.4 A Jablonksi diagram illustrating the process of FRET with two fluorophores.

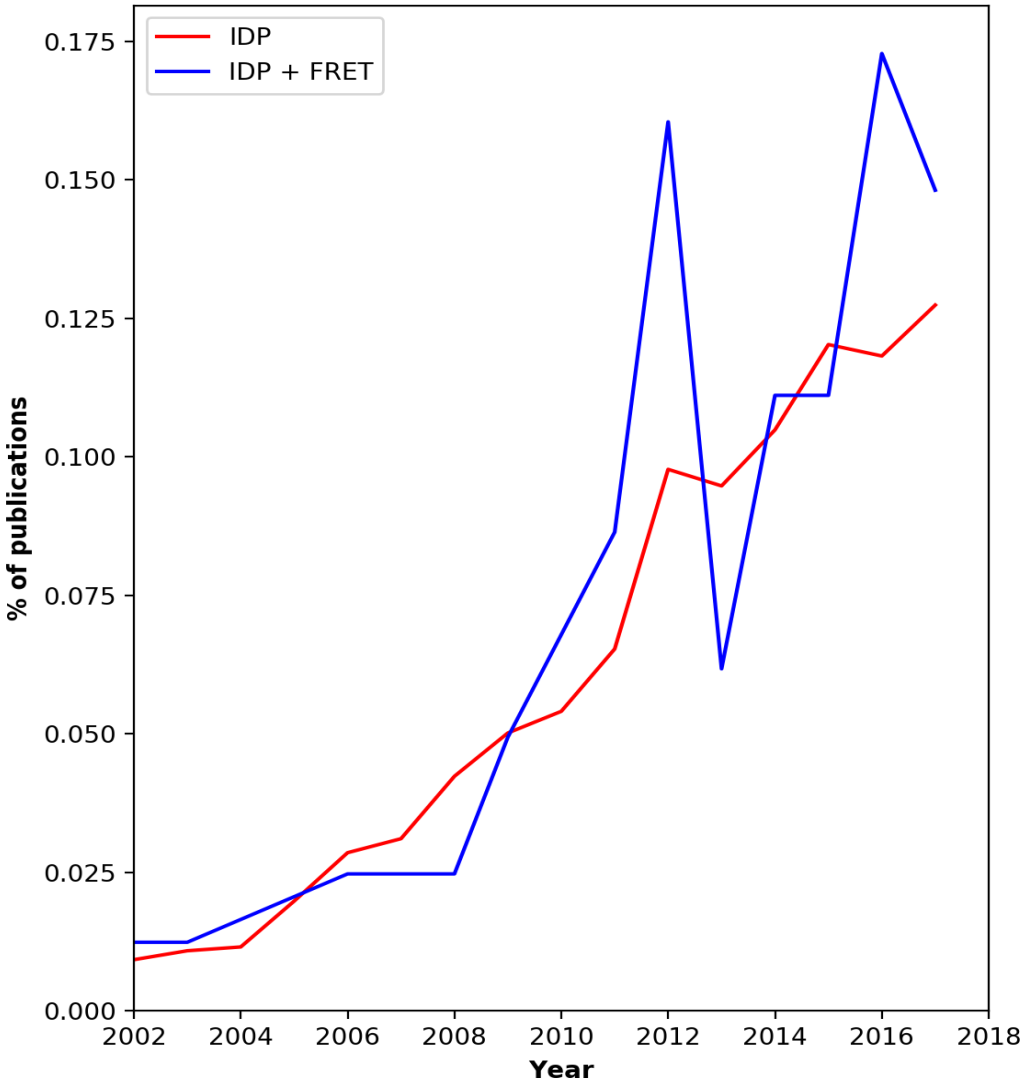


Figure 1.5 Publications containing synonyms for IDPs and synonyms for IDPs and FRET. This figure shows the relative popularity increase by year for this area of research.

Chapter 2: Mimicking a Ribosome Sampling Mechanism in SRP-Based Post-Translational Targeting

This chaptered is a collaborative work that is presented in full.

Rory C. Henderson[#], Dustin R. Baucom[#], Priyanka Sharma, Mercede Furr, Srinivas Jayanthi, Feng Gao, Alicia Kight, Robyn L. Goforth., Ralph L. Henry, Thallapuranam Krishnaswamy Suresh Kumar[*] and Colin D. Heyes[*].
[#]These authors contributed equally

2.1 Abstract

In co-translational targeting, the ribosome-bound SRP samples the exit tunnel for the emerging nascent chain to bind the substrate. The chloroplast SRP (cpSRP) is an example of a post-translational targeting system, in which the cpSRP is a heterodimer consisting of a conserved 54kDa analog and a unique 43kDa subunit. Here, we examine the structure of the cpSRP heterodimer and its targeting mechanism. We find that structural sampling of the substrate-binding region leads to interactions between both cpSRP subunits and the substrate, mimicking the ribosome sampling mechanism that underlies co-translational targeting. From these structures, we identify key residues in cpSRP54 that stabilizes substrate binding and a key region in cpSRP43 that facilitates substrate release into the membrane, verifying their importance using functional integration assays. These results highlight a possible common structural sampling mechanism underlying both co- and post-translational targeting.

2.2 Introduction

The chloroplast signal recognition particle (cpSRP) is a signal recognition particle (SRP) homologue that is responsible for the post-translational delivery of light harvesting complex proteins (LHCs) to the thylakoid membrane, where they bind chlorophylls as part of the photosynthetic machinery in higher plants.[82] The 54kDa subunit of cpSRP, cpSRP54, bears significant sequence and functional similarity to SRP54/Ffh found in the co-translational, ribosome-bound SRP pathway of the cytosol.[82, 83] Importantly, rather than the substantial scaffold that the ribosome and SRP RNA provide, cpSRP54 operates with a novel 43kDa subunit, cpSRP43, that is unique to chloroplasts.[83] cpSRP43 provides the majority of the interaction interface with LHCs, leaving a small segment of its third transmembrane segment (TM3) available for interaction with cpSRP54's M-domain.[48, 49, 84]) In addition to providing this interface, cpSRP43 has been observed to replace SRP RNA, stimulating the GTPase activity of cpSRP54 and cpFtsY.[83] However, it remains uncertain how cpSRP54 has been retrained to apply its co-translational signaling functionality in the absence of the typical ribosome machinery.

A recent surge in information regarding the molecular details of cpSRP signaling has revealed much information regarding the impact of interactions on cpSRP43's structure and signaling. cpSRP43 was first revealed to be rather flexible, with significant interdomain structural dynamics.[61] Furthermore, interaction with a cpSRP54 derived peptide significantly reduced some of this flexibility, particularly in the LHC-binding Ankyrin-repeat region, while also increasing the affinity of a critical cpSRP43-binding segment of LHC, L18.[61] This result was further substantiated, revealing greater detail regarding the specific changes occurring upon this interaction.[85] Additionally, it was shown that Albino 3 (M2 and M4) derived peptides could reverse this increased L18 affinity suggesting a mechanism for release of LHC to the thylakoid membrane.

The M4 region of Alb3 was recently shown to tether cpSRP43 at its second and third chromodomains, CD2 and CD3, respectively.[65]

While there has been a significant increase in our understanding of the mechanistic role of cpSRP43 in LHC targeting, details of the cpSRP54 three-dimensional structure, as well as the full cpSRP heterodimeric structure, have remained elusive. Recent work from our lab revealed that cpSRP54 is also a highly dynamic protein.[49] Specifically, the cpSRP54 N-terminal, N-domain and GTPase, G-domain cycle through interactions with the methionine rich, M-domain, regulated by two flexible segments of the protein. The results of this study indicated that cpSRP43 causes a transition in the conformational composition of cpSRP54 so that, opposite to the case of cpSRP43, it becomes more flexible when bound to cpSRP43. Further, it was found that a reduction in the hydrophobicity of the M-domain resulted in difficulty in forming transit complex. This result suggested that the M-domain of cpSRP54 could form an interaction with the TM3 region of LHC.[49]

Nevertheless, the molecular mechanism by which the M-domain facilitates loading of LHC to form the transit complex (TC) remains uncertain. In the ribosome-bound co-translational targeting system, SRP54/Ffh samples the ribosomal exit tunnel scanning for signal sequences.[86] Using small angle x-ray scattering (SAXS), combined with computational modeling of cpSRP54 structural dynamics based on fitting smFRET distributions of FRET-labeled cpSRP54 (BJ paper), we propose a structural model by which cpSRP54 ‘scans’ the substrate-binding region of cpSRP43. This proposal was tested using intermolecular smFRET experiments and functional assays. smFRET confirmed that the cpSRP heterodimer is characterized by a large amount of flexibility and adopts a more rigid, compact structure upon binding the L18-TM3 region of LHC, suggesting interactions between cpSRP54 and LHC. Functional assays were used to identify 2 sets of interactions underlying this sampling mechanism that underlies the targeting mechanism. First,

two specific residues in the M domain of cpSRP54 were identified to interact with the substrate to facilitate loading into the cpSRP. Second, transient interactions of a disordered N-terminal region preceding the CD1 domain of cpSRP43 facilitates release of the substrate into the thylakoid membrane via a proposed interaction with the C-terminal region of the N-domain of cpSRP54. These results suggest the possibility of a common mechanism between post-translational targeting and co-translational targeting by using a SRP heterodimer to mimick the "exit tunnel" sampling mechanism, thus allowing the ribosomal and SRP RNA scaffolding to be replaced.

2.3 Methods

Cloning, Expression and Purification of cpSRP proteins

Expression and purification of cpSRP43, cpSRP43ΔCD3, cpSRP54 M-domain, and cpSRP54 were performed as previously described.[45] Commercially synthesized genes (Life Technologies Corporation) for mature sized mutant cpSRP54 R66E K69E and mutant cpSRP43 with 23 amino acid residues deleted from the N-terminus were cloned into *pGEM-4Z* plasmid vector using KpnI and HindIII sites. *HindIII* linearized DNA was used to produce in vitro tanscribed/translated mutant cpSRP proteins in a wheat germ system. In vitro transcribed/translated wild type proteins: pLHC, cpSRP43, cpSRP54 and recombinant wild type cpFtsy were produced as described previously (Goforth et al., 2004 & Lewis et al., 2010).

Transit complex formation

Mutant cpSRP43 or mutant cpSRP54 described above were examined for their ability to support transit complex formation as described in Goforth et al., 2004 with the following changes: 20µl of each in vitro translated wild type cpSRP or mutant cpSRP proteins were incubated with 20µl of 35S-labeled in vitro translated pLHC, Mg-ATP (10mM final) in a 80µl assay. For each assay, 15µl of the sample was analyzed by native-PAGE and phosphorimaging.

Integration Assay

Mutant cpSRP proteins were examined for their ability to support LHC integration as described previously in Marty et al., 2009 with the following changes: 50µl of the buffer washed thylakoids were incubated with 4µg of recombinant cpFtsY, GTP (1mM final), 20µl of each in vitro translated wild type cpSRP or mutant cpSRP proteins and 20µl of 35S-labeled in vitro translated pLHC in a 150µl integration assay. After incubation, integrated LHC was recovered as a degradation product by digesting the thylakoids with thermolysin. 15µl of the sample from each assay was analyzed by SDS-PAGE and phosphorimaging. Buffer washed thylakoids may contain residual cpFtsY.

Ensemble FRET

Ensemble FRET titration experiments were performed titrating cpSRP43 into cpSRP54. cpSRP54 was labeled with the donor Alexa488 and cpSRP43 with the acceptor Alexa594, with a dye to protein ratio of 10 to minimize the nonspecific fluorescence labeling. Excess free dye was removed from labeled proteins via size exclusion chromatography. Fluorescence Correlation Spectroscopy (FCS) experiments for free Alexa594 and cpSRP43 protein labeled with Alexa594 were performed in order to confirm that size exclusion chromatography effectively separated the free dye from the labeled protein. Calculations determined less than 4% free dye was present in the cpSRP43-Alexa594 solution using a 2-component diffusion model. The Alexa488 labeled cpSRP54 protein concentration for the FRET experiments was fixed at 50 nM. with titration of Alexa594 labeled cpSRP43 or unlabeled cpSRP43 from 0 nM to 1000 nM. Titration data for unlabeled cpSRP43 was used to correct for quenching by the unlabeled protein. The fluorescence spectra were measured using Photon Technology International (PTI) Quantamaster 4 fluorometer.

The FRET efficiency is determined using

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

where I_D is the intensity of the donor in the presence of the acceptor labeled proteins or unlabeled proteins, and I_{D0} is the initial intensity of the donor. To remove any quenching caused by the unlabeled protein, the unlabeled cpSRP43 FRET values were subtracted from the labeled cpSRP43 FRET values. Each point was then divided by the labeling efficiency of cpSRP43-Alexa594. The data was fit using the Hill equation, which is $(V_{max} * x) / (k + x)$, where x is the concentration of Alexa594 labeled cpSRP43 protein. The fitted graphs provide a maximum FRET value (the FRET value of the fully-bound complex) and a k_D value. The maximum FRET is used to calculate the corresponding average interdomain distances of the cpSRP43 and cpSRP54 heterodimer. The FRET distance equation is $E = R_0^6 / R_0^6 + r^6$, where E is the maximum FRET value, and r is the FRET distance between the dyes. R_0 is corrected based on the Alexa488 quantum yields of Alexa488 labeled cpSRP54 proteins involving FRET, with the given Förster distance for the particular dye pair based on the quantum yield of the free dye (60 Angstroms or 6nm). The obtained k_D values are similar to those obtained by ITC experiments on unlabeled proteins, supporting that the fluorescent dye does not affect the binding of cpSRP43 and cpSRP54.

SAXS Data Collection & Fitting

SAXS data for cpSRP were collected at the Cornell High Energy Synchrotron Source (CHESS) as described at the G1 station.[87] A concentration series from 0.25 mg/ml to 7 mg/ml was collected using ten one-second exposures each to a *q* range of 0.8 $Å^{-1}$. Data was processed in the RAW software and visualized in the ATSAS package's Primus which was used for extrapolation to zero.[88] SAXS data for cpSRP43, cpSRP43 bound to cpSRP54 M-domain, and cpSRP43 bound to cpSRP54 M-domain were collected at the SIBYLS beam line as described.[89, 90] A concentration series from 1 mg/ml to 3 mg/ml was collected for each with a 6 second exposure. SIBYLS data

was processed in Primus. The Pored volume, distance distributions and plots were obtained using Primus. The volume of correlation (V_c) was determined using Scatter.[91]

Flexible fitting of the data was accomplished using BILBOMD for conformational ensemble generation while FoXs and MES were used for SAXS profile calculation and fitting.[92-94] Briefly, BILBOMD uses accelerated all-atom molecular dynamics simulation to sample configurational space according to a user selected radii of gyration (R_g) distribution. The full cpSRP43 structure was built from the folded segments comprised of a CD1-CD2 crystal structure (PDB ID 3UI2) and an NMR CD3 structure (PDB ID 1X3P).[95, 96] Residues 1 to 29 and 258-267 were assigned as disordered chains as 1 to 29 were predicted to be disordered according to MobiDB and previous FRET experiments indicate CD3 remains flexible when cpSRP54 is bound.[61, 97] For fitting of cpSRP43 the segment between the first and second helices of Ank4, residues 190 to 197, were permitted to be flexible, leaving the possibility of displacement of the second Ank4 helix. The previously developed homology model for cpSRP54 was used in this analysis.[49] The M-domain construct used in fitting of the cpSRP43 and cpSRP43ΔCD3 bound data was as described.[49] The NG-domain and M-domain core were kept folded while the G-domain to M-domain linker (residues 296 to 330), M-domain finger-loop (residues 345 to 365), and the C-terminal extension of the M-domain (residues 458 to 471) were regarded as intrinsically disordered as was determined using SAXS. The cpSRP structure was built from these using the previously determined crystal structure in which a short peptide of the C-terminal extension was bound to the linking segment between cpSRP43's Ank4 and CD2. This allowed the cpSRP54 C-terminal extension to remain bound to cpsRP43 in the simulations.

For cpSRP43 and cpSRP43 bound to cpSRP54 M-domain, an R_g range of 20 Å to 65 Å was used initially to find the optimum R_g space for sampling. This was determined for each

construct to be 35 Å to 55 Å. Final runs for the analysis of the SAXS for cpSRP43 M-domain bound and unbound used this range. For cpSRP, an R_g range of 10 Å to 100 Å was used initially to find the optimum R_g for sampling. This was determined to be 35 Å to 65 Å. Final runs for the analysis of SAXS for cpSRP used this range. A total of 600 configurations per Rg were sampled for all constructs. After each run the theoretical scattering patterns of each conformation were calculated using FoXs and a minimum ensemble of structure capable of fitting the SAXS data were calculated using a genetic algorithm, MES.

Intramolecular smFRET PDA Fitting

smFRET histograms from previously-published FRET-labeled cpSRP54[49] were fit using the probability distribution analysis (PDA) method as previously described.[98] Briefly, this method determines theoretical rates of interconversion between two or more states when the rate(s) is on the timescale of the molecule's diffusion through the confocal volume. FRET efficiencies for the states and an initial guess for the rate constants are provided to the algorithm which then minimizes the sum of squared error (SSE) between the theoretical and experimental histogram by changing the rate constants. The algorithm was implemented in C++ with the calculated FRET efficiencies of the modeled structures used as inputs (source code will be made available via github). Specifically, FRET efficiencies of 0.91 (+/- 0.06), 0.70 (+/- 0.06) and 0.25 (+/- 0.12) were used. A larger standard deviation was used for the 0.25 FRET state as it corresponds to a decoupled, flexible state. One-hundred iterations using random initial rate constants between $10s^{-1}$ and $1000s^{-1}$ were used to examine the minimization landscape and calculate the uncertainty in the rate constants and equilibrium distributions. The lowest five SSE for each were used for the final calculations.

Intermolecular smFRET

To compare the spatial relationship of the N-domain of cpSRP54 and CD1 domain of cpSRP43, smFRET was used. mcpSRP43 (C179A, C301A, E89C) was incubated with a 10x molar excess of AlexaFluor594-C5 maleimide in HKM with 100mM KCl for 30 minutes at room temperature. Cys- mcpSRP54his E20C was labeled with a 10x molar excess of AlexaFluor594-C5 maleimide in HKM with 100mM KCl and 50% glycerol for 30 minutes at room temperature. Labeled proteins were then separated from the free dye using a Bio-Gel P-6DG desalting gel filtration column. After elution, glycerol was added to the cpSRP54 to make a 50% solution. cpSRP was prepared by incubating the labeled cpSRP43 and cpSRP54 on ice for and the passing them over a sizing column. Transit complex was then prepared by incubating the prepared cpSRP with L18TM3 and purifying using a strep tag located on the L18TM3. Both the cpSRP and transit complex were analyzed via smFRET. Data collection and processing for the smFRET experiments were as described.[49]

2.4 Results

cpSRP54 structurally 'samples' the LHC-binding region of cpSRP43

A full small angle x-ray scattering (SAXS) study of the full cpSRP heterodimer is extremely difficult to achieve due to the size of the proteins and the flexibility in both cpSRP43 and cpSRP54 that previously identified using smFRET.[49, 61]As a first step towards elucidating the structure(s) of the cpSRP heterodimer, we examined free cpSRP43 and cpSRP43 bound to just the M-domain of cpSRP54 (54M), which is known to be the domain that primarily interacts with cpSRP43, via SAXS. Details of the SAXS data and fitting are given in the supporting information. The SAXS calculated density of 0.82g/ml from a V_c based molecular weight of ~101 kDa and a Porod volume of 203,418 $Å^3$.[61] All-atom molecular ensembles for flexible fitting of the SAXS data

were generated using BILBOMD. Fitting required the previously uncharacterized 27 amino acid N-terminal segment of cpSRP43 to be intrinsically disordered, which was also predicted using MobiDB, as well as flexibility in the linker connecting CD2 and CD3. To Fit the Kratky plot, 4 separate structures of cpSRP43 were necessary, supporting the previous conclusion of cpSRP43 flexibility.[61]

We also examined the cpSRP43-54M complex via SAXS (see supporting information for details). The best fit was obtained using 4 structures to represent an ensemble with relatively large RMSD values (up to 47.1) and variations in R_g (39.2-58.7 Å) indicating significant variation in the particle shape (Table 1). Fitting the cpSRP43-54M complex was obtained without flexibility in Ank4 and between Ank4 and CD2. However, it was found that the M-domain samples configurational space relative to cpSRP43 via an intrinsically disordered C-terminal extension. Our previous cpSRP54 smFRET results revealed that the flexibility between the M-domain and NG-domain of cpSRP54 increases upon binding to cpSRP43.[49]) Taken together, this data suggests that the flexibility initially present in the Ank4-CD2 region of cpSRP43 is transferred to the M-domain of cpSRP54 upon the formation of the cpSRP heterodimer. The is also in agreement with our previous smFRET results on cpSRP43 showing reduced flexibility between Ank2-Ank4 and between Ank2-CD2 upon binding to cpSRP54.[61]

In order to further examine the mechanism by which cpSRP43 transfers flexibility to cpSRP54, we employed the probability distribution analysis method using our previously collected smFRET data for cpSRP54.[49]Using results from the structural models of cpSRP54 obtained via homology modeling, we estimated the transition rates between the three previously hypothesizes states, two M-domain to NG-domain interacting "closed" states, termed states **1** and **2**, and a non M-NG interacting "open" state, termed state **3** (see supporting information). The results of the

PDA fitting of the smFRET histograms to the of the 3 structures show that the "closed" state **1** has a higher occupancy than the "closed" state **2**, thus indicating that state **1** is more stable than state **2**. The PDA fitting results are used to extract relative transition rates between these three states, and suggest that the rate of transition from the "open" state **3** toward the "closed" state **1** is strongly reduced while the reverse rate from **1** to **3** increases. The rate of transition from "open" state **3** to the "closed" state **2** increases, but so does the reverse rate, leading to no change in the relative occupancy of the more minor "closed" state **2**. These results suggest a weakening of the interactions in "closed" state **1** towards the "open" state **3** upon binding to cpSRP43, possibly stabilized by interactions between cpSRP54 with cpSRP43. Since the SAXS data indicates the cpSRP54 M-domain remains flexible in the presence of cpSRP43, such an interaction would likely exist between cpSRP43 and the NG-domain of cpSRP54.

When the NG domain is added to the structures derived from the SAXS and PDA results of the cpSRP43 and cpSRP54, structural model(s) of the cpSRP heterodimer are obtained and shown in figure 2.1. This reveals that the M (orange) and NG-domains (green) of cpSRP54 should sample cpSRP43 (blue) in the region near the binding site of the LHC (Ank 1-4). Furthermore, rather than a single NG-domain orientation, the NG-domain orientation changes relative to the M-domain and cpSRP43. The results support a structural model of the cpSRP heterodimer in which the NG-domain and M-domain are mobile relative to one another and that both are mobile relative to cpSRP43.

Using these proposed models, we determined the optimal place to place fluorescent labels on cpSRP43 and cpSRP54 respectively and performed intermolecular smFRET to probe these various structures. The multiple structures are overlaid onto each other are shown in figure 2.2a which highlights that if dyes are placed at the residues shown, a high FRET and a low FRET state

should be obvious in the smFRET histograms. This was indeed determined to be the case, as shown in figure 2.2(b). The black curve shows that the cpSRP heterodimer in the absence of the subsrtate samples multiple conformations leading to both low FRET and high FRET. More importnantly, when the substate is bound (red curve), there is a clear decrease in the low FRET state and an increase in the high FRET state. This is shown more clearly as a difference histogram in figure 2.2(c). These results highlight that the presence of the L18-TM3 portion (the major cpSRP43-binding motif) of the substrate stabilizes the state in which the NG domain (green) is brought into close proximity to the N-terminal tail of cpSRP43.

Specific residues in the M-domain of cpSRP54 stabilizes LHC binding.

The structural sampling model discussed above leads to the ability for the M-domain of cpSRP54 to interact with the LHC substrate that binds to cpSRP43. It is known that the L18 motif of LHC binds to the Ank-repeat region of cpSRP43[46], leaving the TM regions of LHC protruding from the cpSRP. These hydrophobic domains may interact with hydrophobic domains on cpSRP43 and/or cpSRP54 before the transit complex reaches the thylakoid membrane. The structural model allowed us to predict two key hydrophobic residues in the M domain of cpSRP54 as possible candidates for this interaction – V339 and L370 (Figure 2.3(a). When these residues were mutated to the Asn hydrophilic residue, transit complex formation was significantly reduced (Figure 2.3(b)). There was also a decrease in the integration of LHC, presumably a result of decreased TC formation efficiency.

The N-terminal region of cpSRP43 plays a role in LHC integration but not LHC loading

Interestingly, a combined crystal structure, electron density map of SRP at the ribosome shows that the N-domain of SRP54 forms a close interaction with ribosomal proteins L35 and L23 which bear structural and sequence similarity to the CD1-Ank1 regions of cpSRP43.[99] In

fact, electrostatic potential maps of cpSRP54 and cpSRP43 suggest a possible interaction between the C-terminal exposed loops of the cpSRP54 N-domain the CD1-Ank1 region of cpSRP43. The positively charged patch on cpSRP54 could complement the negatively charged portion of the disordered cpSRP43 N-terminal tail. Based on the structural sampling model discussed above, the flexibility in the cpSRP54 relative to cpSRP43 would allow for the NG domains of cpSRP54 to come into close proximity to the N-terminal tail of cpSRP43 (Figures 2.2(a) and 2.3(a)). Binding of LHC to the cpSRP locks the NG domain of cpSRP54 toward the N-term-CD1 domain of cpSRP43 (Figure 2.2), which can be further stabilized by the N-terminus amino acid residues on cpSRP43.

In order to test whether this interaction between cpSRP43 CD1 and the cpSRP54 N-domain is functionally important in the structural sampling mechanism described here, we tested a cpSRP43 N-terminal deletion mutant for its ability to bind LHC and support integration (Figure 2.3(b) and 2.3(c)). The 43 mutant was able to bind LHC to form the transit complex with no problems, but had a reduced integration efficiency (~60%) compared to the wild type 43 and was additive to the reduction in integration efficiency due to mutation V339 and L370 in cpSRP54. This suggests that interactions of the N-terminal region of cpSRP43 plays a role in preparing LHC for release at the membrane. Based on comparisons to ribosomal SRP54 together with our structural models, a strong candidate for this interaction is the C-terminal region of cpSRP54.

2.5 Discussion

Together, the structural and functional results obtained from SAXS, homology modeling, smFRET, PDA analysis, binding and integration assays reveal a remarkably complex series of interactions preparing cpSRP for interaction with LHC followed by preparing it for release at the membrane. Previously, the cpSRP54 C-terminal extension responsible for binding cpSRP43 was

demonstrated to be intrinsically disordered.[49] However, it was unclear at that point what advantage this disorder provided. The results here demonstrate the disorder in this extension is likely to be used by cpSRP54 to allow the M-domain to scan the region where TM3 is expected to bind cpSRP43 based upon a CD1-Ank4 crystal structure bound to L18 and cross-linking studies.[48, 84] This is noteworthy in that it is known that the SRP54 and Ffh scan the ribosomal "exit tunnel" apparently searching for signal sequences in a similar manner.[86] The missing link in understanding cpSRP43's remarkable ability to replace both SRP RNA and the ribosome to perform post-translational targeting was this segment of cpSRP54 which effectively allows for the transitioning of the M-domain between scanning states while bound to cpSRP43. It was recently shown that LHC can be fully integrated prior to GTPase activity and recycling of cpSRP for another round of LHC targeting.[100] Structurally, it was uncertain how this could be accomplished with the crowding that would be present with all the targeting machinery still in place. The flexibility of the C-terminal extension allows the M-domain to not only scan for TM3, but also for cpSRP54 and cpFtsY to provide cpSRP43 with the conformational freedom to carry out this final integration step. Furthermore, the M-domain alone is effectively able to regulate the positioning of the NG-domains, as no contact between the NG-domain and cpSRP43 is observed.

The cpSRP54 M-domain was observed to increase the GTPase activity of cpSRP54 and cpFtsY in the absence of SRP RNA and cpSRP43.[63] In another study it was observed that CD1 deletion increased GTPase activity.[45] In the locked configuration proposed here, the M-domain could be prevented from occupying the GTPase activating configuration. This would allow the cpSRP43 CD1 domain to effectively regulate M-domain activation of twinned GTPase activity. Further, as release of TM3 would destabilize this configuration, M-domain activation of GTPase activity would become possible once again since the M-domain would be permitted to sample

configurational space. This could act as a signal to the GTPase domains that cargo has been delivered. The model presented here for the transit complex suggests a mechanism by which cpSRP could provide a signal to cpFtsY that cargo has been loaded. Loading of cargo would also provide a specific orientation for the cpSRP54 GTPase domain relative to the remainder of the cpSRP structure. The proposed orientation would place the LHC TM3 tail toward the membrane upon interaction with cpFtsY. Following release of TM3, the M-domain would once again be available for activation of GTPase hydrolysis, signaling the successful delivery of cargo and the need for recycling of the complex. Ultimately, this model provides a framework for understanding the mechanism by which LHC is integrated at the membrane.

2.6 Conclusions

The present work demonstrates the remarkable versatility of a SRP54 homologue, cpSRP54. At some point during its evolution cpSRP54 was adapted to provide signaling and regulatory activity in the post-translational targeting of LHCs using cpSRP43 as a replacement for SRP RNA and the ribosome. Previous studies have mostly been limited in their discussion of cpSRP43 as being predominantly a replacement for specific binding partners in the co-translational path. cpSRP54 is involved in both co- and post-translational targeting, but cpSRP43 has only so far been identified to be involved in post-translational targeting. How it has adapted to this process is a key question. While cpSRP43 functions in a capacity similar to SRP RNA, it also provides a ribosome like scaffold for cpSRP54 interactions. This study demonstrates that it is possible that a common sampling mechanism underlies both co- and post-translational targeting, which may explain how the same cpSRP54 is so versatile to perform both functions. The most striking adaptation of cpSRP54 is its C-terminal extension which is proposed to binds cpSRP43. Through its marked disorder it allows the cpSRP54 M-domain to sample the region near where the TM3

domain of LHC exits cpSRP43, thereby scanning for the substrate in a manner similar to SRP54 scanning for signal sequences exiting the ribosome. In addition to cpSRP54 and cpSRP43 cooperative adaptation, cpFtsY itself is adapted to the absence of SRP RNA, prefolding into a state ready for interaction with cpSRP54.[101] We speculate here that cpSRP54's equilibrium of states acts as a regulatory element, disfavoring receptor interactions in the absence of cargo. As this final state would define a receptor orientation relative to the transit complex, we expect that this final state is critical for orienting the system at the membrane such that LHC integration is as efficient as possible. Upon release of TM3, this configuration would be lost allowing the M-domain of cpSRP54 to activate GTP hydrolysis. The structure of this complex and its mechanistic significance remain to be demonstrated experimentally. This work nevertheless provides a solid framework for dissecting this intricate process, finally exposing the structural framework for LHC targeting.

2.7 Acknowledgements

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

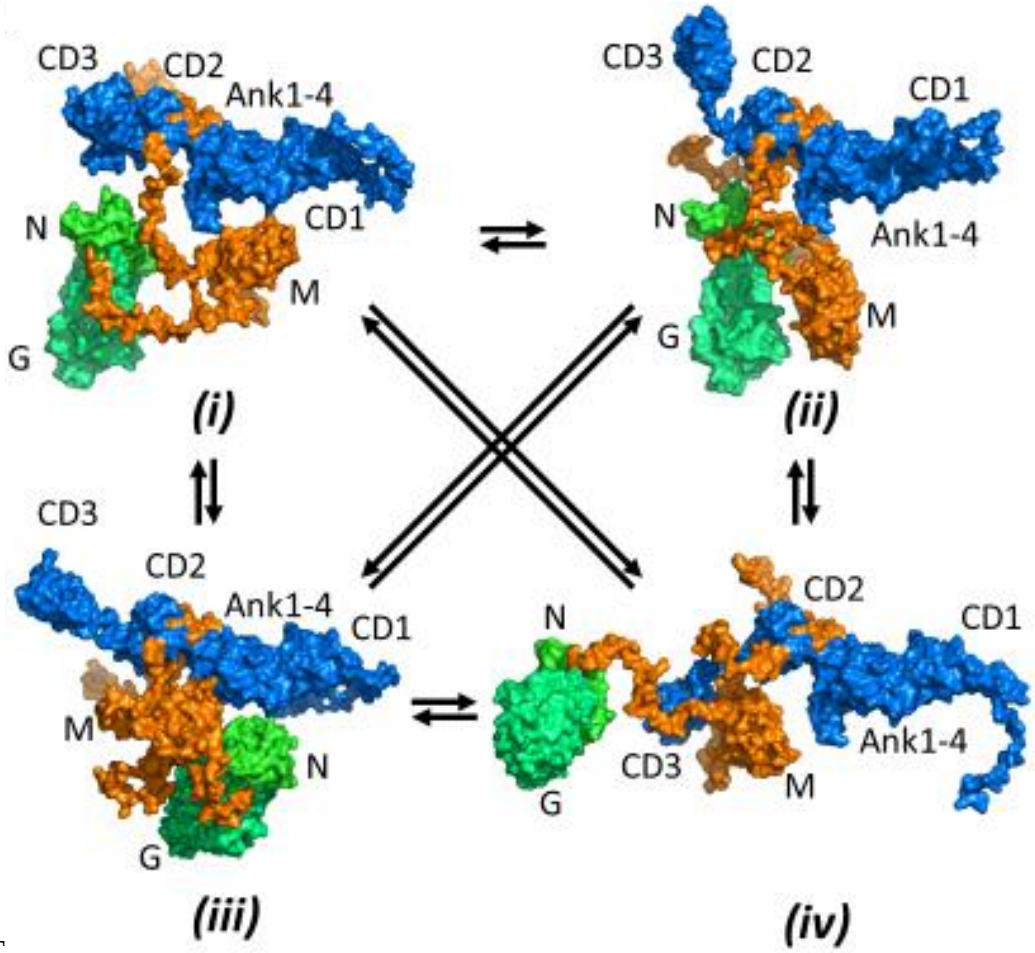
on behalf of the Department of Energy, Office of Basic Energy Sciences, through the Integrated Diffraction Analysis Technologies (IDAT) program, supported by DOE Office of Biological and Environmental Research. Additional ALS support comes from the National Institute of Health project MINOS (R01GM105404).

2.8 Tables and Figures

Table 2.1 shows the SAXS fitting parameters.

Ensemble Size	Chi^2	Structure Number (4 ensemble fit)	Weight (%)	RMSD	Rg (w. avg = 46.8)	Max Distance (w. avg = 176.9)	Experimental Rg (Guinier)
1	2.78	1	18.2	0.0	58.7	192.9	44.9
2	2.43	2	9.6	47.1	39.2	165.1	
3	2.32	3	44.7	32.8	44.7	174.2	Experimental Rg (Distance Distribution)
4	2.18	4	27.5	36.9	44.9	174.8	45.6

w. Avg -> weighted average



T

Figure 2.1 – Multiple structures of cpSRP heterodimer determined by combining SAXS data of cpSRP43 and 54M-cpSRP43, homology modeling of cpSRP54 and FRET PDA analysis. Figure by Henderson

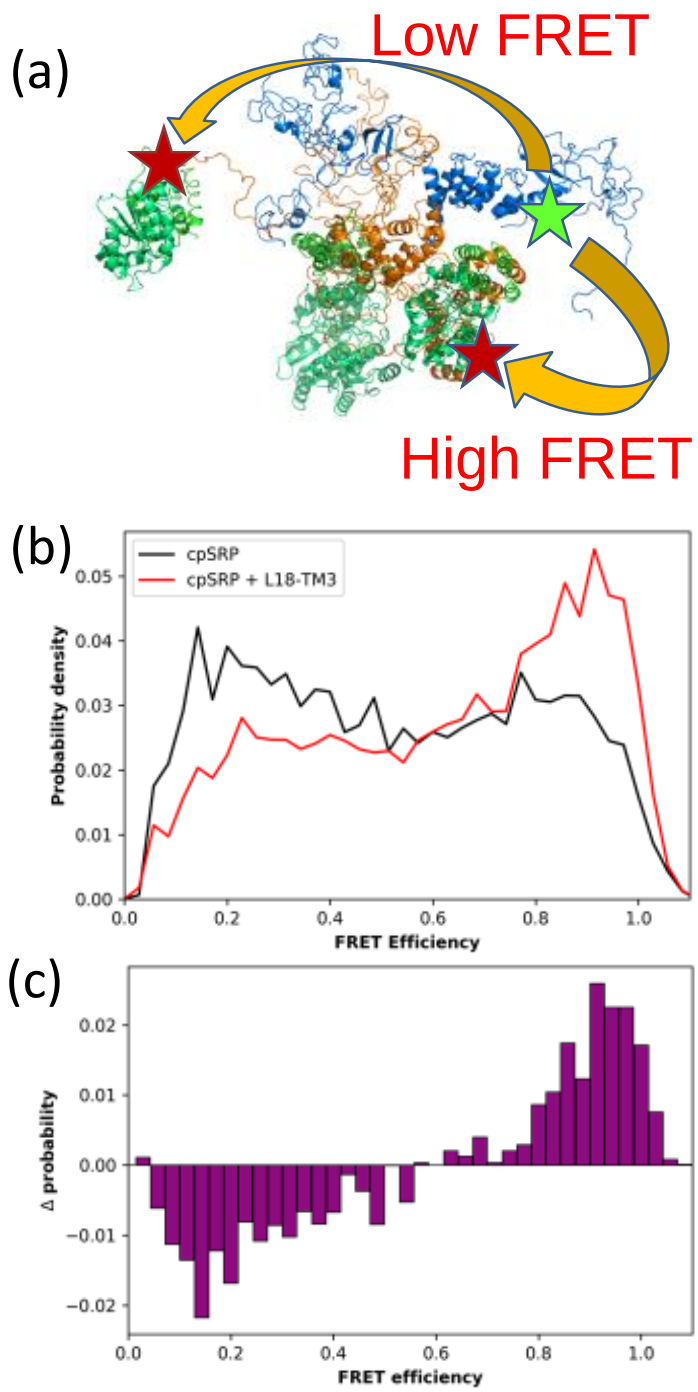


Figure 2.2 – (a) Overlay of multiple cpSRP structures with dye labels highlighted on cpSRP43 and cpSRP54. Multiple FRET states are predicted from these labeling positions (b) intermolecular smFRET histograms obtained from isolated cpSRP heterodimers with (black) and without (red) L18-TM3 substrate. (c) Difference histogram between L18-TM3-loaded cpSRP and unloaded cpSRP. Figure by Baucom and Henderson

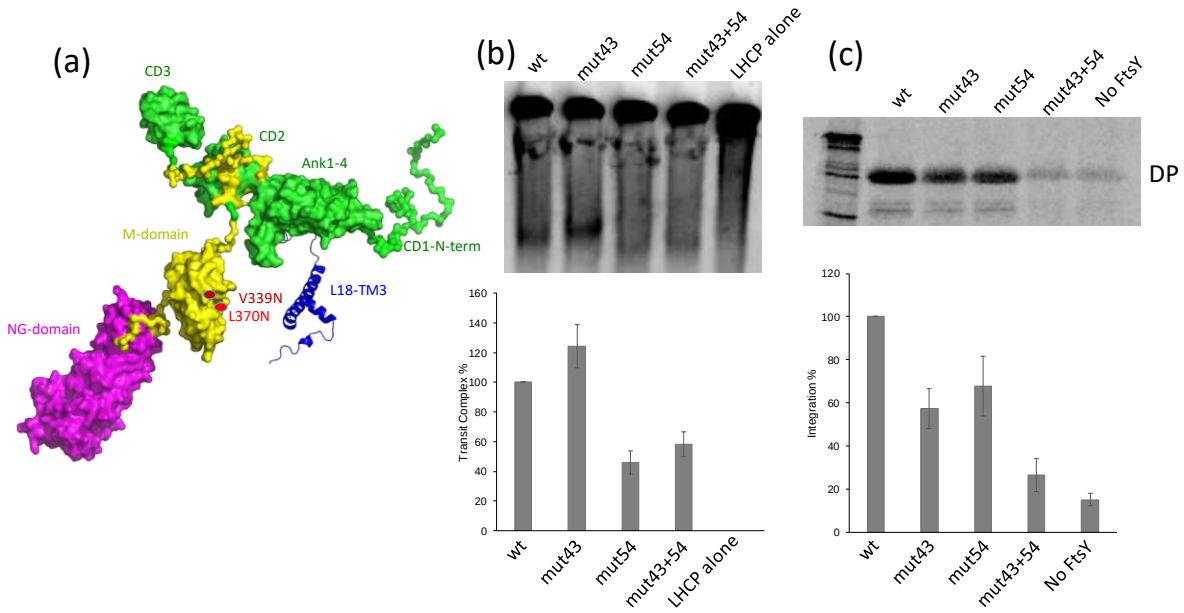


Figure 2.3 – (a) A representative structure of cpSRP heterodimer. cpSRP43 is shown in green. cpSRP54 NG-domain is shown in purple. cpSRP54 M-domain is shown in yellow. The position of a cpSRP54 double point mutant (V339N/L370N) predicted to eliminate LHC binding is indicated in the figure (red dots). The long N-term chain preceding CD1 of cpSRP43 is predicted to be important for LHC integration. (b) Transit complex assay. In vitro translated wild type (wt) cpSRP or mutant proteins and radiolabeled pLHCP were used. Proteins were incubated together at 25 degrees C for 30 minutes and the samples were analyzed on a denaturing gel. Bar graphs were generated by repeating the experiment four times. Error bars represent standard error of the mean. (c) Integration assay. In vitro translated wild type (wt) cpSRP or mutant proteins were checked for their ability to integrate pLHCP into buffer-washed thylakoids. Integration samples were protease treated and solubilized before loading on SDS PAGE gel. DP represents protease treated integrated degradation product. Bar graphs were generated by repeating the experiment seven times. Error bars represent standard error of the mean. Figure by Henderson and Sharma

2.9 Supplemental Material

Analysis of SAXS data

Previous SAXS fitting of cpSRP43 suggested the protein was a largely rigid, extended structure.[24] However, our recent smFRET data demonstrated cpSRP43 to be quite flexible.[5] In light of the smFRET data we decided to reinvestigate the cpSRP43 structure via SAXS. The Kratky plot for the data comprises a large peak followed by a rise at higher *q* values indicative of a protein with both folded and flexible segments (Figure 2.S1). Further, calculation of the protein density

from the Porod volume and V_c base molecular weight calculation reveal a density of 1.01g/ml, far lower than the average folded protein density of 1.37g/ml.[25,26] The Kratky plot and lower calculated protein density are both strong indicators of flexibility and indicate modeling via ensemble fitting is necessary. We therefore fit the data using BILBOMD/MES which combines accelerated molecular dynamics simulations with a genetic algorithm to sample configurational space and determine a set of structures capable of fitting the SAXS data. Flexibility in cpSRP43 was modeled using disorder for the N-terminal tail just before the residues comprising CD1 as is predicted by MobiDB as well as flexibility in the second helix of Ank4, between Ank4 and CD2, and between CD2 and CD3 (Figure 2.S1). The best fit was obtained using 4 structures although the improvement over the 3-structure fit is negligible. The fitting is nevertheless indicative of flexibility in cpSRP43. The distribution in the contribution of each structure to the fitted curve was relatively flat with the composition revealing significant variability in the positioning of the domains and disorder in the cpSRP43 N-terminal tail. Although the fitting method selected four structures to represent the ensemble and since the states selected produce no significant contact between the domains, the cpSRP43 structure likely samples a continuum of states.

As a next step, we examined the effect of the M-domain on the flexibility of cpSRP43 and whether the M-domain adopts a particular conformation relative to cpSRP43 or samples multiple conformations. Previous M-domain SAXS data revealed the M-domain C-terminal extension which binds cpSP43 between Ank4 and CD2 is quite flexible.[49] In both cases, the Kratky plots were indicative of flexibility in the protein complexes (Figure 2.S2). The results of fitting for cpSRP43 bound M-domain indicate that the flexibility in Ank4 and between Ank4 and CD2 is not required for fitting. Further, the M-domain remains quite flexible when bound to cpSRP43, sampling conformational space to one side of cpSRP43. The C-terminal peptide of cpSRP54 that

binds to cpSRP43 between Ank4 and CD2 was retained from its crystal-structure with the residues between the cpSRP54 M-domain and the binding peptide motif, as well as the residues following this peptide motif, modeled as flexible segments. Additionally, the residues between the G-domain and M-domain of cpSRP54 was modeled as a flexible linker. The best fit for cpSRP43 bound M-domain was obtained using 4 structures. The ensemble of structures reveals the M-domain C-terminal extension is disordered except in the cpSRP43 binding region. This allows the M-domain to sample configurational space to one side of cpSRP43 near ankyrin repeat 4 helix-2.

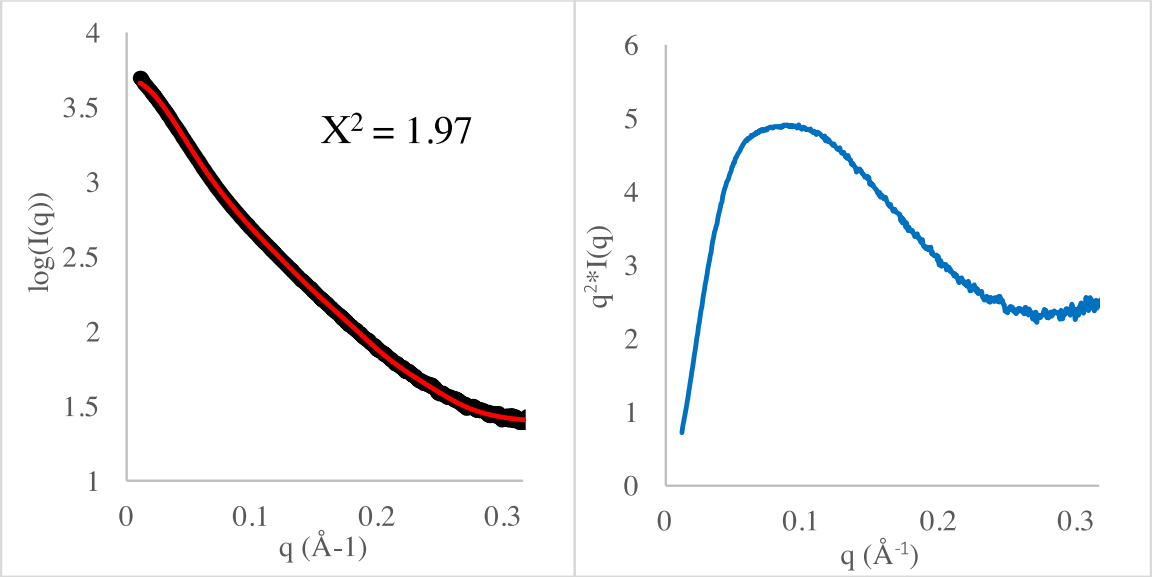


Figure 2.S1: Data fitting for free cpSRP43. (a) Ensemble fit (red) to the data (Black) of I(q) vs q. (b) Kratky plot for free cpSRP43. The rise in data at high q is indicative of high flexibility. Figure by Henderson

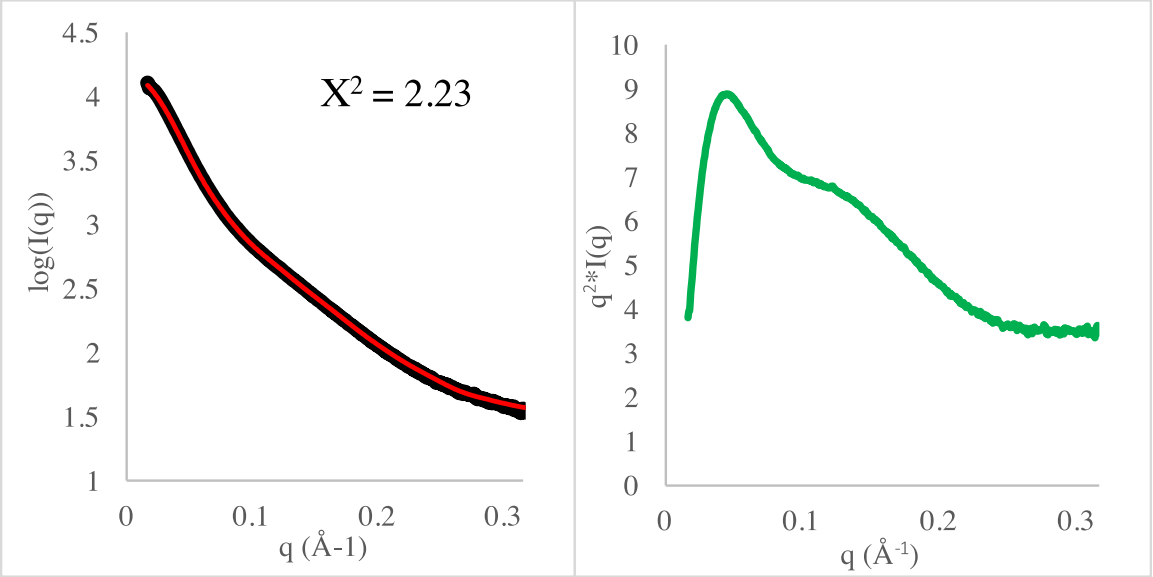


Figure 2.S2: Data fitting for cpSRP43-54M. (a) Ensemble fit (red) to the data (Black) of I(q) vs q. (b) Kratky plot for free cpSRP43. The rise in data at high q is lower than for free cpSRP43 indicative that bound cpSRP43-54M is less flexible than free cpSRP43, although it still does not approach zero, suggesting that it is still relatively flexibile. Figure by Henderson

PDA Analysis of smFRET distributions

Our previous smFRET experiments demonstrated the ability of cpSRP43 to influence the conformational distribution of cpSRP54 states, biasing cpSRP54 toward the open, flexible state. In order to gain a mechanistic understanding of the means by which cpSRP43 accomplishes this biasing, we utilized the FRET probability distribution analysis method to determine the kinetic rates of transitions between the previously postulated states. We first examined the ability of the three hypothesized structural states to describe the data.[49] Distances of 45Å, 54Å, 72Å determined from those structures, equivalent to FRET efficiencies of 0.91, 0.54, and 0.25, respectively, were used in the fitting process. These states were termed **1**, **2**, and **3**, respectively. State S1 corresponds to a state similar to that observed in a crystal structure for *T. aquaticus* (PDB ID 2FFH) in which the M-domain sits atop the NG-domain.[22] State **2** corresponds to a state similar to that occupied in an *S. solfataricus* crystal structure (1QZX) whose M-domain is stabilized by the G-domain to M-

domain linker near the N-domain.[23] The final state, S3, corresponds to a state in which the M-domain is fully decoupled from the NG domain and is therefore able to sample configurational space limited only by the length of the G-domain to M-domain linker. Interconversion between states **1** and **2** were not allowed as these would require a relatively complex mechanism of transition and would likely be quite slow compared to the diffusion timescale (~1.4ms) (Figure 2.1). Additionally, a greater FRET efficiency distribution used for state S3 as it corresponds to a highly flexible, decoupled state.

The results for free cpSRP54 indicate rapid exchange between states **1** and **3** with a rate of 830 s^{-1}(+/-94 s^{-1}) for k_1 and 985 s^{-1}(+/-160 s^{-1}) for k_{-1}. The rate of transition between states **3** and **2** is markedly slower with a rate of 101 s^{-1}(+/-38 s^{-1}) for k_2 and 396 s^{-1}(+/-120 s^{-1}) for k_{-2}. The rates result in a relatively low equilibrium population of state **2**, at 10 percent, with higher populations for states **1** and **3**, at 51 percent and 39 percent, respectively. Distinct changes in the rates of interconversion between states occur upon addition of cpSRP43. The forward rate for transitions from **1** to **3**, k_1, increases from 830 s^{-1}(+/-94 s^{-1}) to 926 s^{-1}(+/-145 s^{-1}) while the reverse rate, k_{-1}, decreases from 985 s^{-1}(+/-160 s^{-1}) to 552 s^{-1}(+/-203 s^{-1}). The rate of conversion from state **3** to **2**, k_2, does not change significantly, 101 s^{-1}(+/-38 s^{-1}) to 192 s^{-1}(+/-91 s^{-1}), while the reverse rate increases from 396 s^{-1}(+/-120 s^{-1}) to 743 s^{-1}(+/-163 s^{-1}). The result of the change in interconversion rates upon addition of cpSRP43 is that the equilibrium occupancy of state **1** is reduced from 51 percent to 39 percent with an increase in the occupancy of state **3** from 39 percent to 48 percent. State **2** is little changed with an increase from 10 percent to 13 percent (Figure 2.S4). It is important to note that the minimization algorithm may find rather shallow minima with multiple solutions of similar fit to the data. Indeed, the results here produced a relatively wide distribution in the minima.

Nevertheless, while the actual values vary somewhat about the average, the trends in each individual fitting yield similar results.

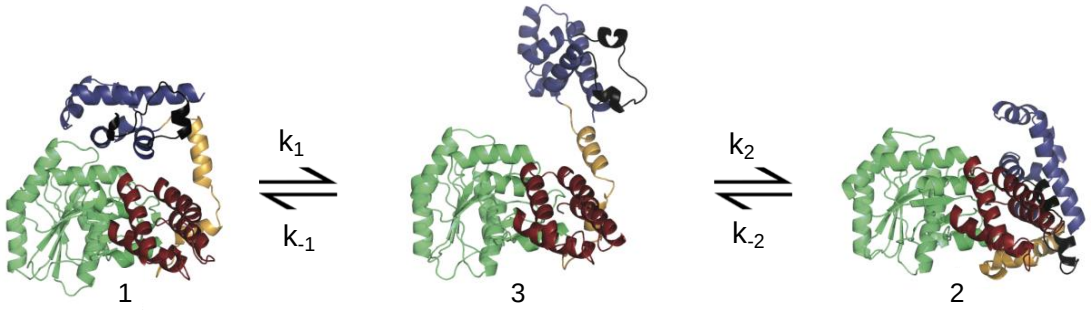


Figure 2.S3: Structural models obtained via homology modeling showing 2 "closed" states **1** and **2**, and an "open" state, **3**. The proposed kinetic model allows us to extract rate constant from the smFRET distributions reported in Henderson et al. Figure by Henderson

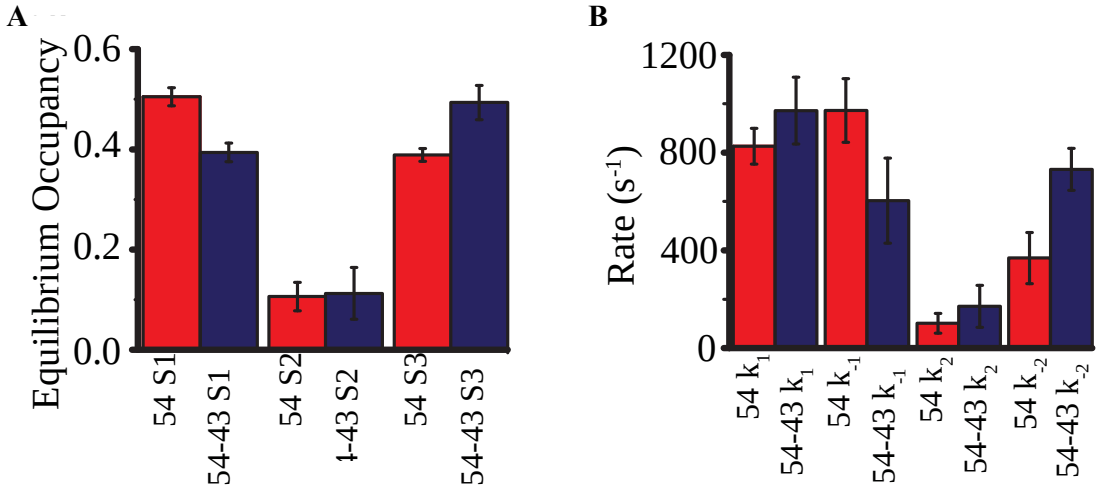


Figure 2.S4: Results of the PDA fitting to smFRET histograms and the kinetic model described above. (a) equilibrium occupancy of each state for free cpSRP54 (red) and cpSRP54 bound to cpSRP43 (blue). (b) transition rates between the different states for free cpSRP54 (red) and cpSRP54 bound to cpSRP43 (blue). Figure by Henderson

Chapter 3: Transient Structural Propensity in the C-terminal Domain of the Albino3 Translocase in Thylakoids

This chaptered is a collaborative work that is presented in full.

M. Furr, D. Baucom, V. Govind Kumar, P. Okoto, M. Moradi, C.D. Heyes, and T. K. S. Kumar

3.1 Abstract

Albino3 (Alb3) translocase in chloroplasts is fundamental to the integration of LHCPs into the thylakoid membrane, working in cooperation with the chloroplast signal recognition protein (cpSRP) and its receptor cpFtsY. Alb3 is an integral membrane protein containing five transmembrane helices in an N-in, C-out orientation in the thylakoid membrane. The c-terminus of Alb3 (cAlb3) is responsible for binding to the cpSRP43 subunit of cpSRP once it is bound to cpFtsY at the thylakoid membrane for successful integration of LHCPs. This region of Alb3 has previously been described as intrinsically disordered. In more recent developments, intrinsically disordered proteins have been shown to carry out vital functions within cells. While our findings show that cAlb is predominantly disordered, we have discovered a region in this protein which may have a relatively significant propensity towards an ordered structure. This region was first isolated by secondary structural sequence analysis using multiple prediction software. Single point mutations which are expected to contribute to the displacement of this structure were made within this region, followed by structural characterization. Results derived from Far-UV-Circular Dichroism, intrinsic fluorescence, and trypsin digestion reveal that the single-point mutations do indeed show a decrease in overall structure. Inter-residue distances were acquired from single-molecule fluorescence resonance energy transfer (smFRET) analysis of this region under native conditions and upon adding a denaturant. Changes in the smFRET histograms show that this region is able to be unfolded with increasing denaturant. Additionally, all-atom molecular dynamics

simulations conducted in conjunction with novel enhanced sampling techniques not only confirm the relatively high helical propensity of the identified region but also agree very well with the smFRET results at a quantitative level. These simulations thus provide a unique perspective on the structural dynamics of cAlb at an atomic resolution. The local structure of cAlb found in this study may prove to be important for binding of this protein to the cpSRP43 subunit of cpSRP, and may be useful to understanding the structure-function relationships of other translocases.

3.2 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.[102] 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.[102] Each homolog shares a conserved hydrophobic region comprised of five transmembrane domains . However, YidC, present in gram-negative bacteria, contains six trans membrane segments and a large periplasmic domain between TMs 1 and 2. Alb3 is required for the post-translational integration of LHCPs into the thylakoid membrane.[44] Nuclear-encoded LHCPs are imported from the cytoplasm into the stromal compartment of the chloroplast via a chloroplast targeting sequence.[33] Upon entrance into the stroma, the targeting sequence is cleaved and LHCP joins with cpSRP to form the transit complex.[59] Alb3 functions as an insertase and is critical for the post-translational integration of LHCPs and co-translational integration of other chloroplast membrane proteins. The c-terminal of Alb3 (cAlb) is an intrinsically disordered region of Alb3 which protrudes out into the stroma and facilitates association of the cpSRP at the thylakoid membrane through recognition and association with the cpSRP43 subunit.[52, 64] cAlb, upon binding to cpSRP43, gains a more compact

structure as demonstrated in circular dichroism experiments.[52] In addition, cAlb binding to cpSRP43 activates GTP hydrolysis by cpSRP54 and cpFtsY which is critical for the release of LHCP.[103] An interaction between LHCP and cAlb was observed in *in vitro* binding experiments.[104] It was also concluded that a cpSRP-LHCP complex binds more efficiently to cAlb than cpSRP alone which suggests that a cpSRP loaded with the LHCP cargo is important for the interaction with cAlb and may prevent the Alb3 protein from being blocked by a cargoless cpSRP.[104] 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. Many studies have investigated the binding interaction between cAlb and cpSRP43, but there is disagreement for which regions in both proteins are important for binding.[52, 64]

Sequence alignments of Alb3 have revealed four conserved, positively charged motifs (I-IV) in the stromal c-terminal region.[52] ITC binding experiments between cpSRP43 and full length cAlb revealed a K_d of 9.7μM and in a binding experiment where cAlb was lacking motif I the K_d was 11.4 μM. With a small change in dissociation constants, these researchers suggest that motif I is not important in the binding interaction.[52] 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 hypotheses that motifs II and IV are required for the interaction with cpSRP43.[52] 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% of its functionality as compared to the wild type.[105] This was possibly due to either 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 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.[105]

Within cpSRP43, the ankyrin repeat regions and CD2CD3 of cpSRP43 have been implicated as the important areas for association with cAlb.[48, 64] 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 cAlb or stabilizing the interaction between the two proteins.[64] However, examination of motifs II and IV of cAlb reveal AKRS and SKRS sequences which are similar to ARK(S/T) sequences present in histone H3 tails, suggesting a likely affinity between these motifs and the chromodomains of cpSRP43.[48] Lysine and arginine were independently mutated to alanine in both motifs and this led to reduced affinity to cpSRP43. When the mutations were applied simultaneously, binding with cpSRP43 was lost as observed by ITC. This result suggests that the positively charged side chains in motifs II and IV are important for binding cpSRP43.[48]

To attempt to resolve the debate over the interaction sites, examining the secondary structure and interactions between these proteins in a systematic manner is valuable. Four selected amino acids in a region predicted to contain transient helical structure, 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.[106, 107] Characteristic analysis of these designed mutations showed a decrease in residual structure within the proposed helical region. smFRET experiments, in which cysteine residues encompassing the region of interest were labeled were performed and found to display an inter-residue distance indicative of the presence of more ordered structure that that of random coil. Furthermore, using a denaturant, the smFRET data shows a decrease in the FRET between these same residues, indicating that this local structure can

be denatured and lending support for secondary structure in the native state in this region of cAlb. Computational modeling, using extensive all-atom molecular dynamics (MD) simulations of cAlb in explicit water in conjunction with enhanced sampling techniques, also provides further independent evidence for local structure propensity in the N-terminal region of cAlb and quantitatively agrees with the experimental smFRET distance distribution. 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 that leads to LHCP integration.

3.3 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 the pQE-80L vector and transformed in BL21 star cells (Obtained from Dr. Ralph L. Henry's lab). Protein production was induced with 1 mM isopropyl-1-thio-D-galactopyranoside at an Abs_{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 ultra centrifugation at 19,000 rpm. The supernatant was applied onto a Ni^{2+} 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. The molecular weight and purity were confirmed by mass spectroscopy (M.W.=15.17 kDa). The eluted protein was further purified to homogeneity using a Mono S ion exchange column.

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 DH5α 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^{2+} column and subsequent purification using a Mono S. Protein bands were visualized by staining the gels with Coomassie brilliant blue R-250 dye. 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_{600} and incubated with 1 mM isopropyl b -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_2PO_4, 15 mM Na_2HPO_4, 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. 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.

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 µM 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-napthalenesulfonic 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 µM 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 µls, mixing and incubation for 2 minutes preceding each reading at 25ºC. 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 µg of protein and 0.5 µg 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 µl 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}H-^{15}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 $^{15}NH_4Cl$ resulting in ^{15}N labeling. The experiment was acquired at 25 °C at a protein concentration of 300 µM. The sample was prepared in 90% H2O 10% D2O solution containing 50 mM Hepes buffer with150 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). 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 µM 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 Å-1. Buffer scattering data was collected before and after exposures of the protein. Data was analyzed using the RAW software then visualized using ATSAS Primus

. 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 DH5α 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.

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.[108] 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.[109, 110] PSIPRED secondary structure predictions use the same inputs as DISOPRED to provide further structural information. 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.[111] PrDOS2 output is a measure of residual disorder probability . 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.[112] Foldunfold is a method which locates the mean packing density of residues in order to predict disordered regions.[113] 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 15-38) 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.[114, 115] 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.

Fluorescent Labeling and Single-molecule FRET

The double cysteine cAlb mutants and wild type were both labeled under the same conditions. cAlb was incubated in pH 7.5 100 mM PBS, with 50 μM 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. The labeling efficiency was determined using UV-Vis absorption spectroscopy on a Hitachi U-3900H.

Labeled protein preparations were diluted to ~300 pM for smFRET. The dilution buffer for the denaturation experiments contained the appropriate concentration of guanidinium hydrochloride in pH 7.5 PBS buffer. Using a MicroTime 200 microscope (Picoquant GmbH Berlin, Germany) based on an Olympus IX71 microscope, the protein solutions were 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 μm pinhole was used to block out-of-focus fluorescence. Collected photons were then passed to a FF562-Di03 dichroic mirror (Chroma) to split the fluorescence into donor and acceptor channels based on wavelength. Further, each channel had a band-pass filters (Chroma) 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. Single photon avalanche diodes (SPAD) detectors (PDM, Microphotonic Devices,

Bolzano, Italy) were used to collect donor and acceptor signals in time-tagged time-resolved (tttr) format).

Single molecule traces were binned into 1 ms intervals and analyzed using a home-written analysis Python program. First, the donor channel was corrected using the γ-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 analyzed. 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 the smFRET histograms.

Enhanced sampling molecular dynamics (MD) simulations of cAlb

The LEAP (Loop prediction by Energy-Assisted Protocol) algorithm[116] was used to generate an all-atom model of cAlb in a fully extended conformation. Modeller[117] was then used to generate 16 different cAlb conformations using a Monte Carlo algorithm to minimize the Modeller objective function with different number of iterations. The 16 Modeller conformations were then used to set up 16 independent well-tempered metadynamics simulations[118] of cAlb in both implicit and explicit water. Explicit water systems were solvated in a box of TIP3P waters[119] and 0.15M NaCl. Implicit water systems used the generalized Born implicit solvent model (GBIS).[120] System size for the explicit water systems ranged from 23,745 atoms to 557,583 atoms due to different box sizes, while the implicit water systems had 2,138 atoms. Simulations were performed using the NAMD 2.13 simulation package[121] with the CHARMM36 all-atom additive force field[122]. Initially, we energy-minimized each system for 1000 steps using the conjugate gradient algorithm[123]. Then, the systems were equilibrated for 1 ns using a 2-fs time step at 310 K using a Langevin integrator with a damping coefficient of $\gamma = 1$ ps^{-1}. For the implicit water

simulations, the smoothed cutoff distance for nonbonded interactions was set to 256 Å. For the explicit water simulations, the smoothed cutoff distance for nonbonded interactions was set to 10−12 Å and long-range electrostatic interactions were computed with the particle mesh Ewald (PME) method[124].

After equilibration, well-tempered metadynamics simulations were run for 10 ns and 320 ns for the implicit and explicit water systems respectively. "alpha" (α) was used as the collective variable in all metadynamics simulations. This collective variable quantifies the α-helical propensity of a protein or peptide of length N:

$$\alpha = \frac{1}{2}\left(\frac{1}{N-2}\sum_{n=1}^{N-2} f(\theta_n) + \frac{1}{N-4}\sum_{n=1}^{N-4} g(d_n)\right),$$

where $\theta_n$is the angle formed by $C_\alpha^{(n)} - C_\alpha^{(n+1)} - C_\alpha^{(n+2)}$ and $d_n$is distance between $O^{(n)}$ and $N^{(n+4)}$. $f(\theta)$ and $g(d)$ are score functions (both ranging from 0 to 1) quantifying the likelihood of θ and d being associated with an α-helix. More specifically:

$$f(\theta) = \frac{1-\left(\frac{\theta-\theta_0}{\delta\theta}\right)^2}{1-\left(\frac{\theta-\theta_0}{\delta\theta}\right)^4}, \qquad g(d) = \frac{1-\left(\frac{d}{d_0}\right)^6}{1-\left(\frac{d}{d_0}\right)^8},$$

where θ_0 and $\delta\theta$ are 88 and 15 degrees, respectively and d_0 is 3.3 Å. The metadynamics algorithm used a width of 0.01, an initial Hill weight of 0.1, and an update frequency of 1000. The pseudo temperature used for the well-tempered feature was 1200 K[125]. All simulations were run on the Blue Waters supercomputer. Conformations were collected every 2 ps for the implicit simulations and every 4 ps for the explicit simulations.

Secondary structure determination from MD trajectories

VMD[126] was used to visualize the cAlb conformations and to generate molecular images. The STRIDE algorithm[127] was used to calculate the helical propensity of individual amino acids

based on the MD trajectories. All 16 MD trajectories in either explicit or implicit solvent simulations were combined to generate per-residue helical propensity profiles. Overall helical propensity (h) in each metadynamics trajectory (e.g., replica i) was calculated using the weighted average of the "alpha" collective variable using the metadynamics based free energy profile from trajectory i ($F_i(h)$):

$$\langle h \rangle_i = \frac{\int_0^1 e^{-\frac{F_i(h)}{RT}} h \, dh}{\int_0^1 e^{-\frac{F_i(h)}{RT}} dh}$$

Computational prediction of smFRET distance distributions

Atomic models of Alexa488 and Alexa594 dyes were attached to a cysteine amino acid and simulated in implicit water for 10 ns. The system size was 90 atoms for Alexa488 and 126 atoms for Alexa594. Conformations were collected every picosecond. Parameter and topology files were generated using CGenFF[128]. Simulations were performed using NAMD 2.13 with simulation parameters similar to those described above for cAlb. These simulations generated a library of 10,000 conformations of the two dyes. These models replaced the cAlb residues of interest (C14 and S52) in 5000 and 80,000 conformations of cAlb from implicit and explicit simulations, respectively. For the replacement, S52 was first mutated to C52 and the sampled conformations of both dyes were aligned with both C14 and C52. For each conformation of cAlb and the donor and acceptor dyes, two alternative models were generated (donor and acceptor attaching to C15 and C52, respectively, or vice versa). Mass center fluorophore-fluorophore distance distributions were then calculated for every conformation and compared with distance distributions derived from smFRET data (with an R_0 of 42 Å, determined from quantum yield measurements).

3.4 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.[129] The different structural motifs of proteins have characteristic CD spectra . Proteins containing predominate alpha-helices with have negative bands at 222 and 208 nm and will typically have a positive band at approximately 194 nm. Typical beta-sheet proteins show negative bands at 218 nm and positive bands at 195 nm.[130] Far UV-CD spectra of IDPs show a negative band around 195 nm indicating a random coil.[131] However the CD spectra of IDPs can fluctuate depending on the chemical conditions under which the protein is analyzed . 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.1A).

Intrinsic fluorescence measurements of proteins are performed by exciting the protein sample using a wavelength at 280 nm and observing the emission within a range of 300 nm to 400 nm . The emission wavelength peak depends on the polarity of the microenvironment surrounding the tryptophan residue(s) . 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, suggesting that the tryptophan residue is exposed to the solvent and cAlb is predominately unstructured in its native state (Fig.1B).

Heteronuclear single quantum coherence (HSQC) of cAlb is characteristic of an IDP

Two dimensional ^{1}H ^{15}N HSQC is a commonly-used NMR technique for determining structural details at the residual level for ^{15}N isotope-enriched proteins.[132] The acquired HSCQ

spectra displays the backbone conformation of proteins. ^{1}H ^{15}N HSQC shows the hydrogens that are directly attached to the nitrogen atoms . 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 . 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 . Crosspeaks for structured proteins occur over a range of ~7.5 to ~10 ppm in the proton dimension . However, HSQC spectra of IDPs are characterized by narrow spectra ranges. Crosspeaks for disordered proteins typically occur from ~8.0 to ~8.6 ppm which coincide with the range for amide protons in the central residue of unstructured tripeptides. ^{1}H ^{15}N HSQC spectra for hFGF1 and cAlb is compared below (Fig. 3.2). The structure of wt-hFGF1 has been well characterized complete with assigned resonances . hFGF1 is a 15 kDa protein comprised of twelve beta-strands arranged in an antiparallel fashion into a beta-barrel structure . The ^{1}H ^{15}N HSQC of hFGF1 shows well-dispersed crosspeaks along a spectral range of 6 to 11 ppm (Fig. 2A). 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 . The ^{1}H ^{15}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.2B).

Small Angle X-ray Scattering

Small Angle X-ray scattering (SAXS) is a powerful technique for providing global structural information for proteins and other molecules in solution.[133] SAXS profiles are a determination of the scattering intensity within a protein as a function of spatial frequency . The scattered light is a result of the difference between electron densities of the protein and the solvent, and therefore the solvent is subtracted out of the scattering data. The result thus represents the

overall shape of the protein and, when compared to corresponding X-ray structures and modeling data, can also provide conformational information . SAXS data is useful for indicating flexibility in proteins, and the shape of the Kratky plot ($I(q)*q^2$ vs. q) provides an assessment of disorder within a protein . 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 . Multi-domain proteins may show additional peaks at low q values but will still converge at the q axis at a high q value . 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 is calculated from the slope and may vary greatly depending on the level of expertise and interpretation by the user . The Guinier plot can also provide additional information about how the proteins behave upon exposure to the x-ray beam . A reliable Guinier plot should be linear meaning that the sample is monodisperse . 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 Å-1. Buffer scattering data was collected before and after exposure. 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.3.3).

Sequence analysis provides an estimation of secondary structural propensity in cAlb

Secondary structural elements present in the amino acid sequence of cAlb were determined using the consensus of five different web servers. 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, highlighted in bold in the sequence below (Fig.3.4). Sequence analysis data for cAlb from all the above listed servers is included in the supplementary material.

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

Four single point mutations were made and analyzed 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 A21, V24, L26 and K28, which were carefully selected for having helix promoting characteristics and mutated to glycine, a known helix breaker . According to an experimentally based helix propensity scale using 11 different systems developed by Pace et. al, alanine has the highest helix propensity, followed by leucine. Lysine was shown to have the 5^{th} highest helix propensity while glycine has the lowest helix propensity . 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 . 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. 3.5A.) 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 implies the presence of local secondary structure is receded in the A21G, V24G and L26G mutants when compared to the wild type cAlb (Fig.5A), indicating loss of secondary structure.

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

Intrinsic fluorescence can be used to monitor 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, as 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. 3.5B). There is a single tryptophan residue located within the region of predicted structure which appears to be unaffected by these mutations.

Estimation of secondary structural content from circular dichroism spectra of wt-cAlb and its mutants were analyzed using the online server algorithms in the CDPro program (Fig. 3.6). 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 , fits the CD spectrum of interest as a linear combination of the reference CD spectra by using a ridge regression procedure . CDSSTR, developed by Johnson , 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 . 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 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 . 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 .

Upon comparing the fractions for each secondary structural element in cAlb and each mutant, the values from CONTINLL and SELCON3 were very similar. 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.

The unordered fraction is consistently the largest fraction for wt-cAlb and its 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 disordered secondary structure. The percentages of regular and disordered sheet content were significantly low (less than 0.1) in the wild type and this result remained relatively consistent with all the mutants. The calculated percentages of turns were also comparable for wt-cAlb and the mutants. 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, as shown later.

Limited trypsin digestion reveals an overall decrease in stability for the cAlb mutants compared to wt-cAlb

Limited trypsin digestion provides information on the backbone flexibility of proteins. Arginine and lysine residues located in flexible, solvent-accessible regions of proteins are more vulnerable to trypsin cleavage. wt-cAlb contains 46 trypsin cleavage sites which, in conjunction with being predominately unstructured, results in it being 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. 3.7). 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.

smFRET of cAlb supports the presence of residual secondary structure

In an attempt to determine if the prediction software was correct in predicting an ordered, helical region in the n-terminal region of cAlb, smFRET was employed to determine if the inter-residue distances were consistent with those expected from structured or disordered polypeptides. By placing a FRET pair of fluorescent dyes at two residues encapsulating the proposed helical region, smFRET on freely-diffusing cAlb was used to estimate the distance between the dye positions in the protein. The first panel of Figure 8A shows the smFRET histogram of the cAlb under non-denaturing conditions. There are two populations of FRET efficiencies. The low FRET efficiency peak centered around zero results from statistically-labeled donor-only cAlb and provides no structural information. The high FRET peak is from the dual-labeled population. Using a measured R_0 of 42 angstroms (measured using polyproline standards as previously shown (cite Feng's JBC paper)), the distance between the two dyes is calculated to be ~37 angstroms. This is quite compact and is inconsistent with this region being a random coil. Also, of interest is the broadness of the peak. This could indicate that, as well as a preferred compact conformation, there is also some flexibility and/or disorder in this region.

To further investigate the possibility of secondary structure in this region, a denaturation titration was performed using guanidinium hydrochloride. As shown in Figure. 8A, as denaturant

is titrated in, the high FRET peak from the original conformation begins to decrease and a new peak at intermediate FRET efficiencies increased, continuing to increase in intensity and shift toward lower FRET efficiencies up to a concentration of 3M denaturant. In the smFRET histograms, three populations were fit. The peak at 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 separately measured donor-only labeled mutant. The high FRET peak was also fit to a lognormal function (reversed in its asymmetry), with parameters associated with the FRET labeled protein under non-denaturing conditions. A Gaussian peak was used to fit the subpopulation of denatured molecules at intermediate FRET. Using these fits, the donor-only peak can be removed to highlight the changes in smFRET, as shown in Figure. 8B in which the data is overlaid and the histogram bars removed for clarity. The diasppearance of the high FRET peak to be replaced with an intermediate FRET peak is strongly indicative that some structure was present, and that the protein becomes more disordered in this region as denaturant concentration increases. Additionally, an experiment was done to attempt to dialyze the denaturant from the labeled cAlb after the experiment to determine if the denaturation was reversible. This dialyzed sample was then measured and was shown to recover the high FRET peak of the native structure, although a significant proportion was lost, presumably due to aggregation and/or proteolytic degradation.

By comparing the integrated area of the high FRET peak and intermediate FRET peak fits, the free energy barrier associated with the unfolding process of the native state can be determined**.** For any given concentration of denaturant, the free energy difference between folded and unfolded states, ΔG_{FU}, is given by Equation 1 .

Equation 1: $\Delta G_{FU} = -RT\ln\frac{U}{F}$

Where U and F are the fractions of the unfolded and folded states respectively, R is the ideal gas constant, and T is the temperature. Since the free energy barrier between the two states decreases linearly as denaturant increases, this can be used to find the free energy difference in the cAlb under native conditions, ΔG_0, according to Equation 2.

Equation 2: $\Delta G_{FU} = \Delta G_0 - m[GdnHCl]$

By plotting the ΔG_{FU} versus denaturant concentration, the y-intercept yields ΔG_0 and the slope yields the cooperativity value, m, for the unfolding process. Figure 8C shows this plot and yields a free energy difference between the folded and unfolded states of 1.28kcal/mol and a cooperativity of 0.674 kcal/mol/M. This low energy difference is consistent with a low stability of secondary structure in the region. The low cooperativity value also indicates that the secondary structure is stabilized only slowly with removal of denaturant, indicating a low propensity for secondary structure and perhaps only transient formation.

Experimental smFRET and MD simulations of dye-dye distance distributions are in agreement

A lifferent CAlb conformations , B generated using Modeller[117], which were then used to set up 16 independent well-tempered metadynamics simulations using the alpha collective variable in explicit water. Alpha collective variable quantifies the helical propensity of the protein and metadynamics speeds up sampling along this collective variable. The metadynamics simulations were run for 320 ns per replica (~5μs in aggregate). Cysteine attached Alexa488 and Alexa594 dyes were separately simulated to generate a library of conformations and replace C14 and S52 to estimate the dye-dye distance C butions from the metadynamics simulations. The results were found to be in very good agreement with the distance distributions derived from experimental smFRET data for the double cysteine mutant (Fig. 3.9). The MD based distance distributions shown in Fig. 3.9 are generated using all 16 simulations (see Supplemental Fig. 3.12

for all 16 individual distributions). All 16 simulations were repeated in implicit water and produced similar results but more skewed towards the larger distances (or disordered conformations) (Supplemental Fig. 3.13). Given that the MD simulations did not use any information from the smFRET experiments above, one may argue that the similarity of the experimental and computational distance distributions validates the use of our novel approach for generating an ensemble of cAlb conformations, which can be used to provide a detailed atomic-level description of cAlb structural dynamics as discussed below.

Molecular dynamics simulations and helical propensity predictions corroborate the existence of a structured region near the cAlb N-terminus

Conformations derived from converged metadynamics simulations revealed that cAlb contains helical regions (Fig. 3.9) within the larger region of interest predicted by the software (amino acid residues 15-38) and experimentally tested by site-directed mutagenesis and smFRET. The STRIDE algorithm was used to calculate the helical propensity of individual cAlb amino acid residues (Fig. 3.10; also see Supplemental Fig. 3.14 for implicit simulation results). Structured regions located near the C-terminal end of the sequence were also reported by the simulations, which correlates with some of the predictions made initially. The helical propensity data obtained from STRIDE shows that the region of interest (residues 15-38) has a peak helical propensity of around 15% (Fig. 3.10). 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. 3.6). In addition to the STRIDE algorithm, overall helical propensity was also determined by weighted averaging of the free energies calculated along the alpha collective variable during the metadynamics simulations (Fig. 3.11, Supplemental Fig. 3.15). This was found to be 26±4% for the explicit water systems (Fig. 3.11) and 33±6% for the implicit water systems (Supplemental

Fig. 3.15). Overall, the computational dye-dye distance distributions and secondary structural predictions are in agreement with the experimental data and provide a complementary picture, where the cAlb exists as an intrinsically disordered domain that contains one or more helical segments at a time. However, these helical segments form and get disrupted continuously.

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 μM 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_2PO_4, 15mM Na_2HPO_4, 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.

3.5 Summary and Conclusions

In this study we have examined in detail the structure of cAlb to investigate the existence

of an ordered region in this predominately disordered protein. We combined structural predictions, biophysical approaches, spectroscopic techniques and computational methods to reveal a region of local structure in the n-terminal region of cAlb, approximately 30 amino acids long. While the NMR and SAXS data supports the conclusion that cAlb is intrinsically disordered, CD data shows an additional broad negative peak between 220 nm and 230 nm, implying the existence of some residual secondary structure, likely □-helical. Using site-directed mutants to reduce helical propensity in this region, our results show an overall decrease in structural integrity. All the mutants also showed lower structural stability, supporting the conclusion of decreased structural integrity. smFRET of cAlb under native conditions and upon increasing concentrations of denaturant further supports the presence of structure in the same region. In addition, the dye-dye distance distributions derived from experimental smFRET data agree with the dye-dye distance distributions computed from high accuracy all-atom MD simulations. Converged conformations from 16 independent well-tempered metadynamics simulations in explicit and implicit water confirm that cAlb has alpha-helical regions within a larger region of interest (residues 15-38) identified by the prediction software (Fig. 3.9-3.10, Supplemental Fig. 3.14) and that overall helical propensities (Fig. 3.11, Supplemental Fig. 3.15) are in agreement with the estimation of protein secondary structural content from the CD analysis using CDPro software (Fig. 3.6). 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.

3.6 Figures

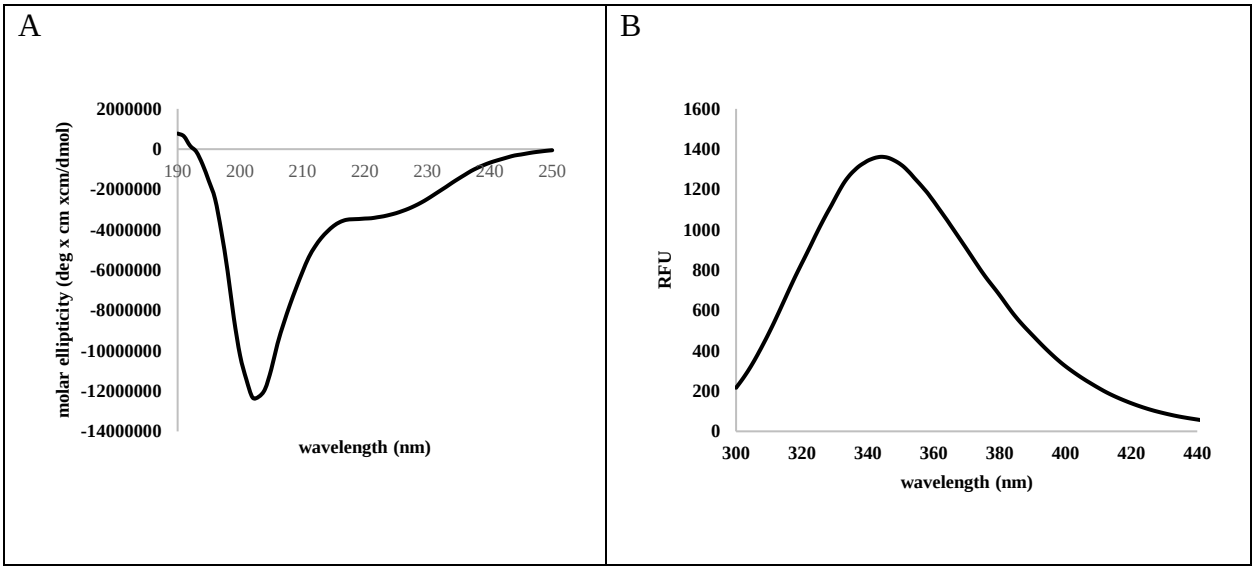


Fig. 3.1: (Panel-A) Far UV-CD spectra of cAlb. (Panel-B) Intrinsic fluorescence spectra of cAlb. Figure by Furr

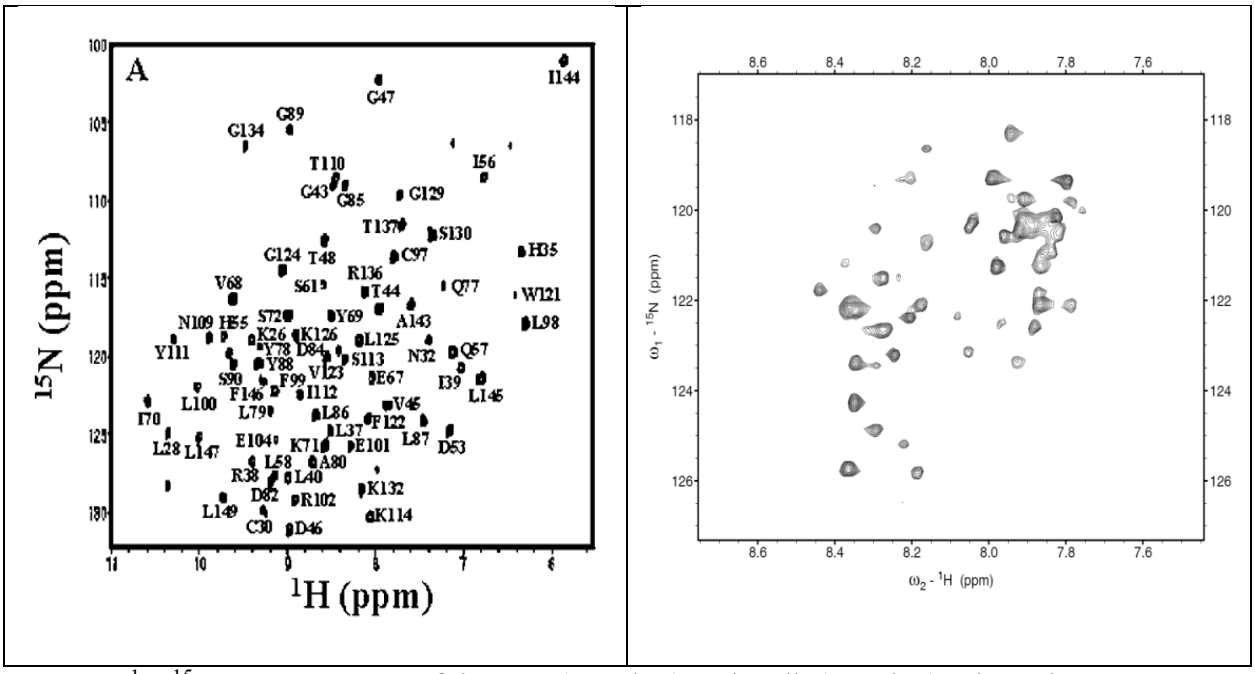


Fig. 3.2: ^{1}H ^{15}N HSQC spectra of hFGF1 (Panel-A) and cAlb (Panel-B). Figure by Furr

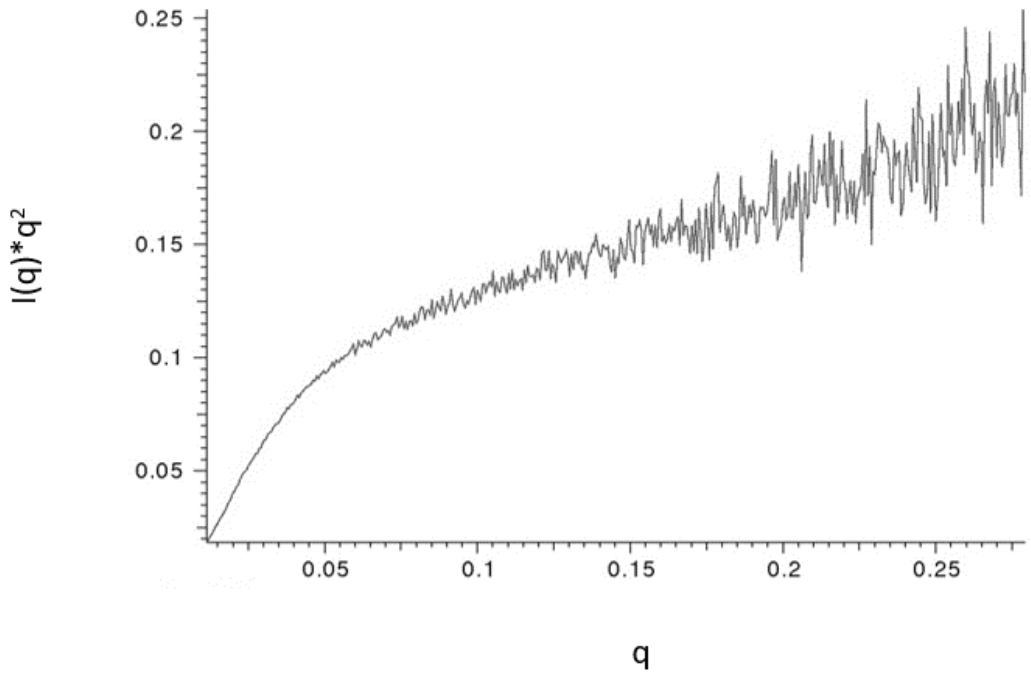


Fig. 3.3: Kratky plot of cAlb. Figure by Furr

```
        10         20         30         40         50         60
MRGSHHHHHH GSACNNVLST AQQVWLRKLG GAKPAVNENA GGIITAGQAK RSASKPEKGG

        70         80         90        100        110        120
ERFRQLKEEE KKKKLIKALP VEEVQPLASA SASNDGSDVE NNKEQEVTEE SNTSKVSQEV

       130
QSFSRERRSK RSKRKPVA
```

Fig. 3.4: Amino acid sequence of cAlb. The amino acids highlighted in bold depict the common region of predicted structure. Figure by Furr

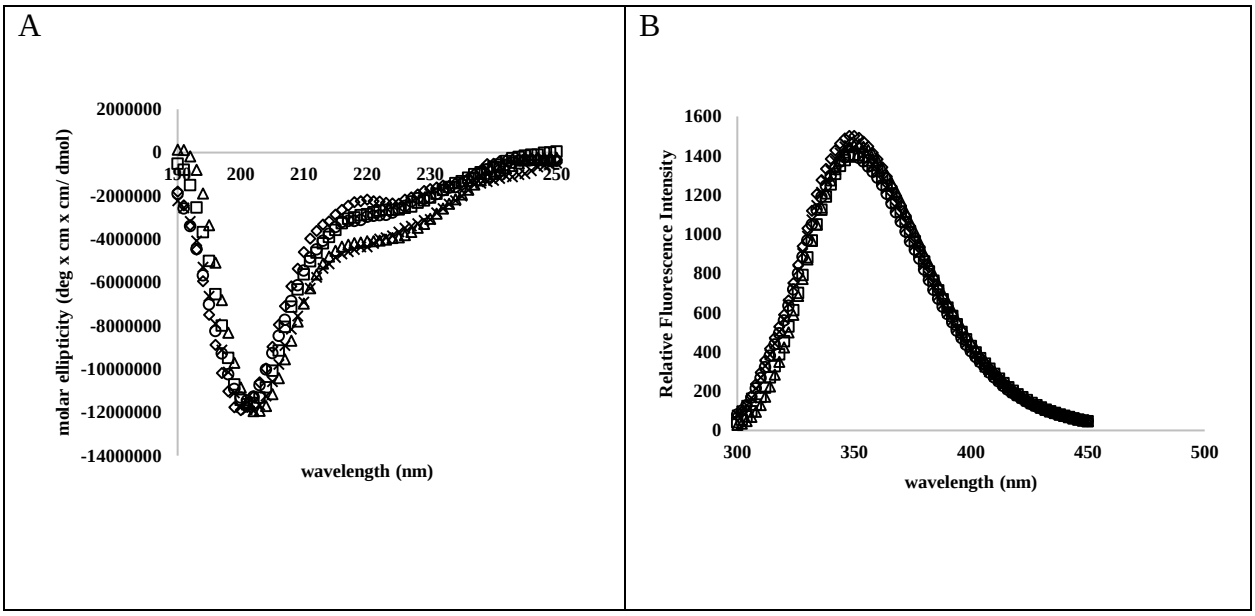


Fig. 3.5: The Far-UV CD spectra overlay (Panel-A) and intrinsic fluorescence (Panel-B) of wt-cAlb and the engineered mutations. wt-cAlb (Δ), A21G (◊), V24G (○), L26G (☐), K28G (x). Figure by Furr

Table. 3.6: 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. Figure by Furr

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

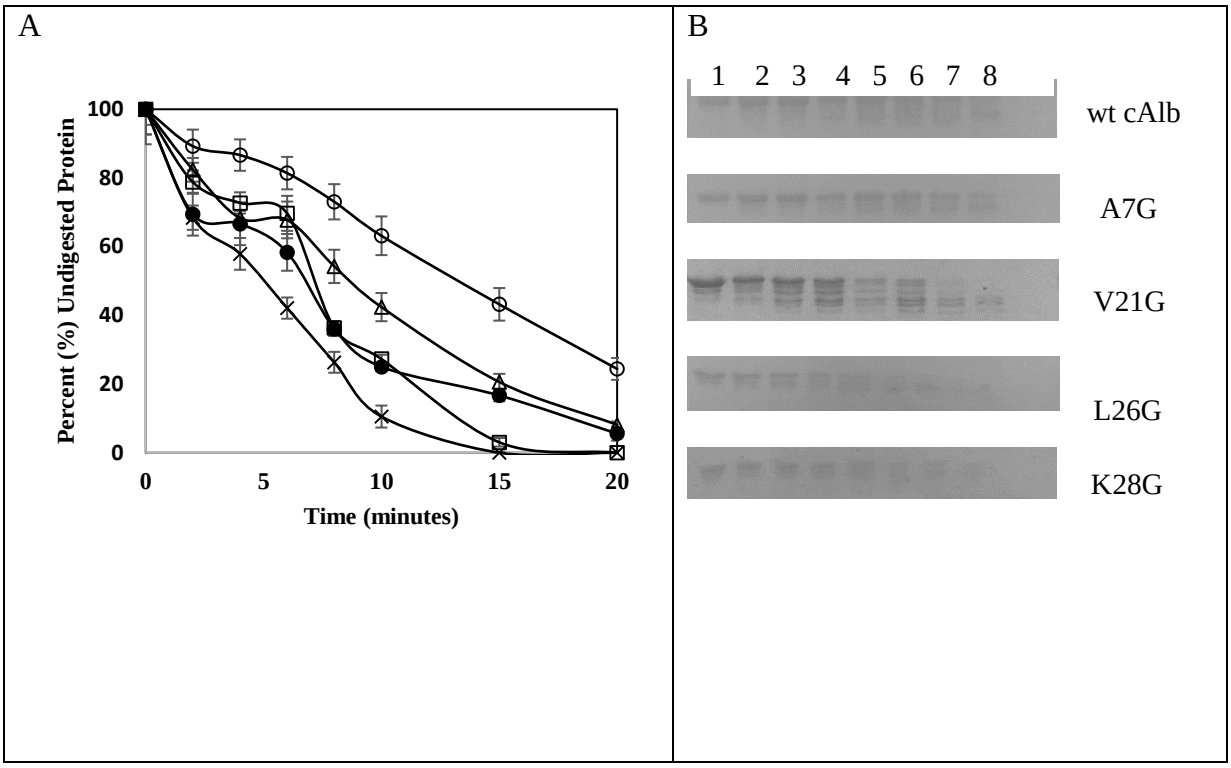


Fig. 3.7: Densitometric analysis of the SDS-PAGE gels depicting the resistance of wt-cAlb (0), A21G (Δ), V24G (□), L26G (x) and K28G (●) to limited trypsin digestion (Panel-A). SDS-PAGE gels of limited trypsin digestion for cAlb and mutants (Panel-B). Lane 1= Zero Trypsin, Lane 2= 2 minutes, Lane 3= 4 minutes, Lane 4= 6 minutes, Lane 5= 8 minutes, Lane 6= 10 minutes, Lane 7= 15 minutes, Lane 8= 20 minutes. Figure by Furr

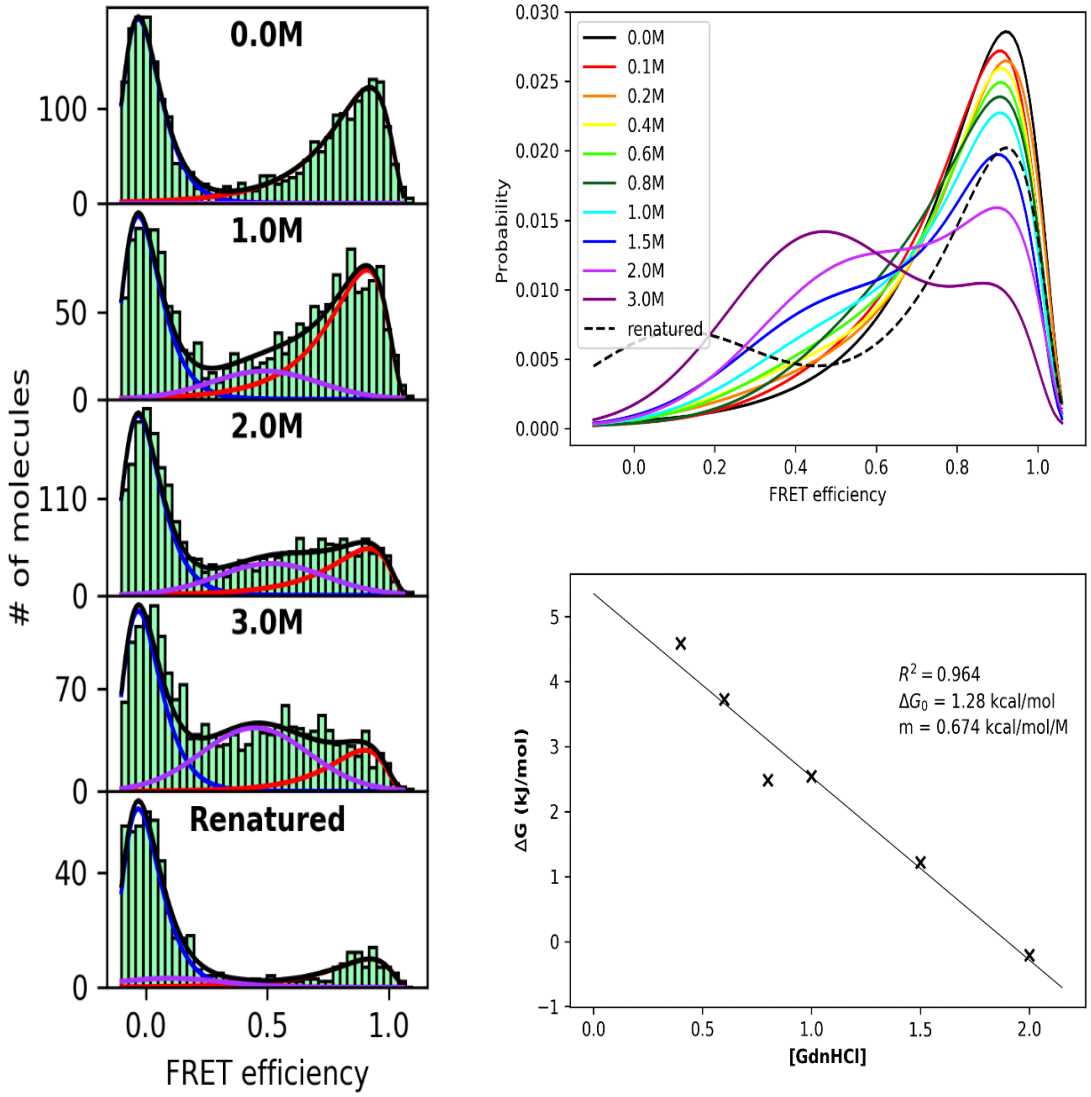


Fig. 3.8: Selected smFRET histograms of S52C cAlb in increasing guanidinium hydrochloride (Panel-A). Fit values with the donor only peak removed of the titration smFRET histograms (Panel-B). Plot of the unfolding energetics of S52C (Panel-C). By plotting the ΔG_{FU} versus denaturant concentration, the y-intercept yields ΔG_0 and the slope corresponds to the negative cooperativity value, m, for the unfolding. Figure by Baucom

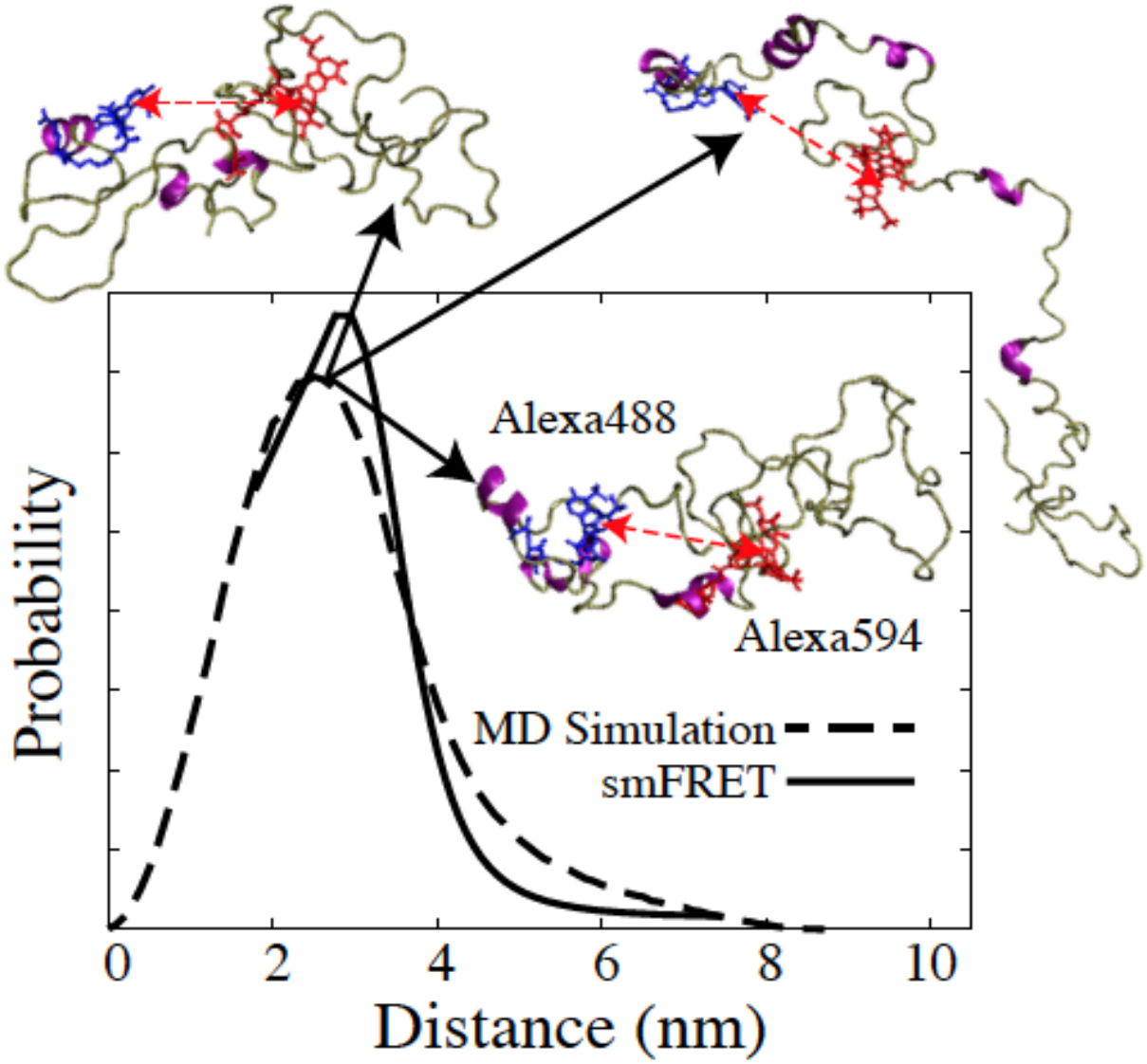


Fig. 3.9: Comparison of experimental smFRET and computational dye-dye distance distributions. Three representative converged conformations from the explicit water simulations are depicted above the distribution plot. Helical regions within the larger region of interest (residues 15-38) are shown in purple. Figure by Baucom and V. Kumar

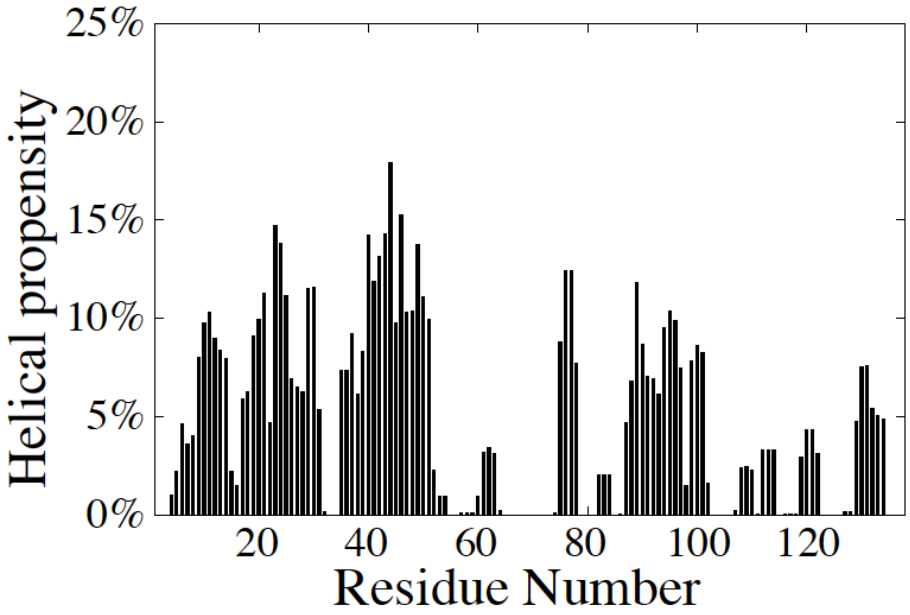


Fig. 3.10: Helical propensity per residue from MD simulations. STRIDE algorithm was used to calculate the helical propensity of individual amino acid residues. The region of interest (residues 15-38) has a peak helical propensity of around 15%. Figure by V. Kumar

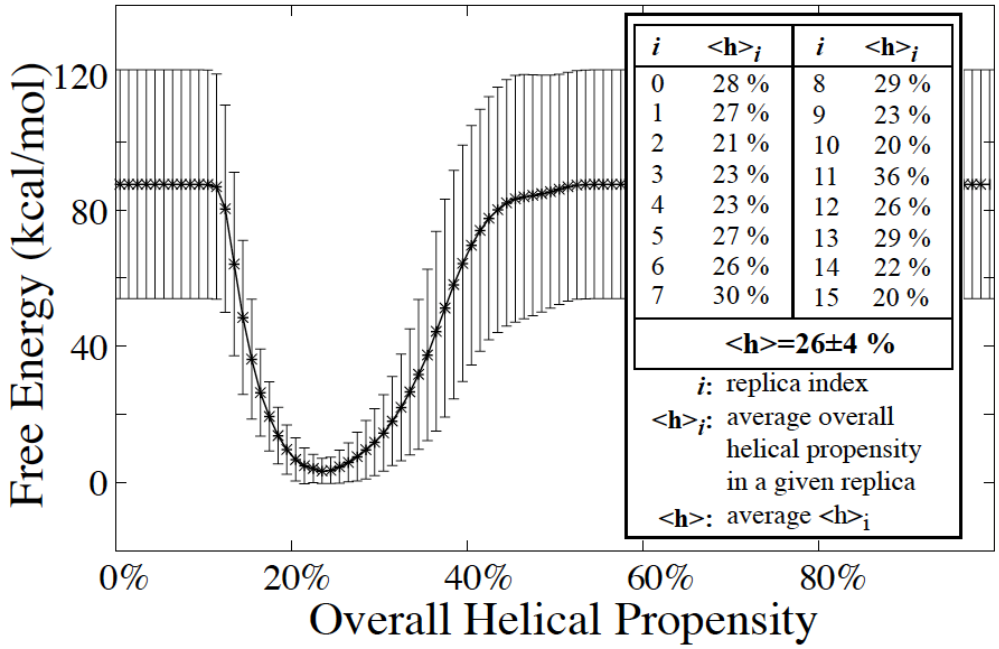
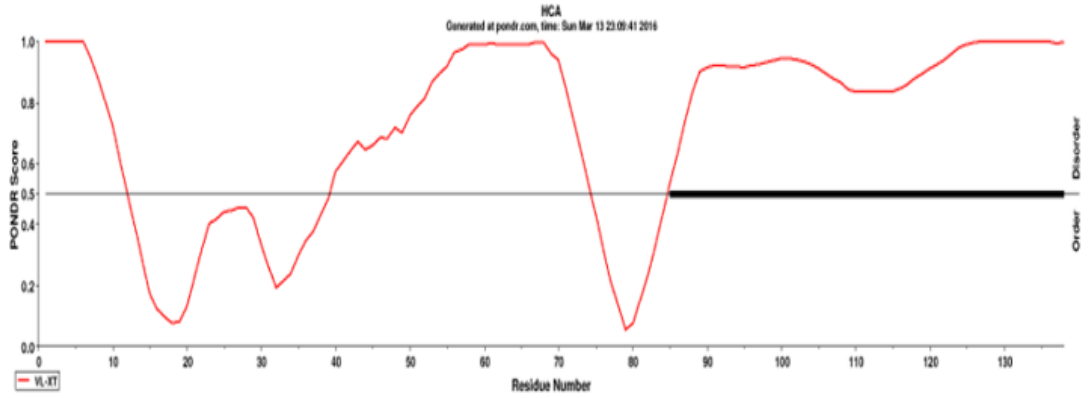


Fig. 3.11: Overall helical propensity from MD simulations. Weighted averaging of the free energies calculated along the alpha collective variable during the well-tempered metadynamics simulations was used to calculate the overall helical propensity. Average helical propensities from each of the 16 independent cAlb simulations are shown inset. Figure by V. Kumar

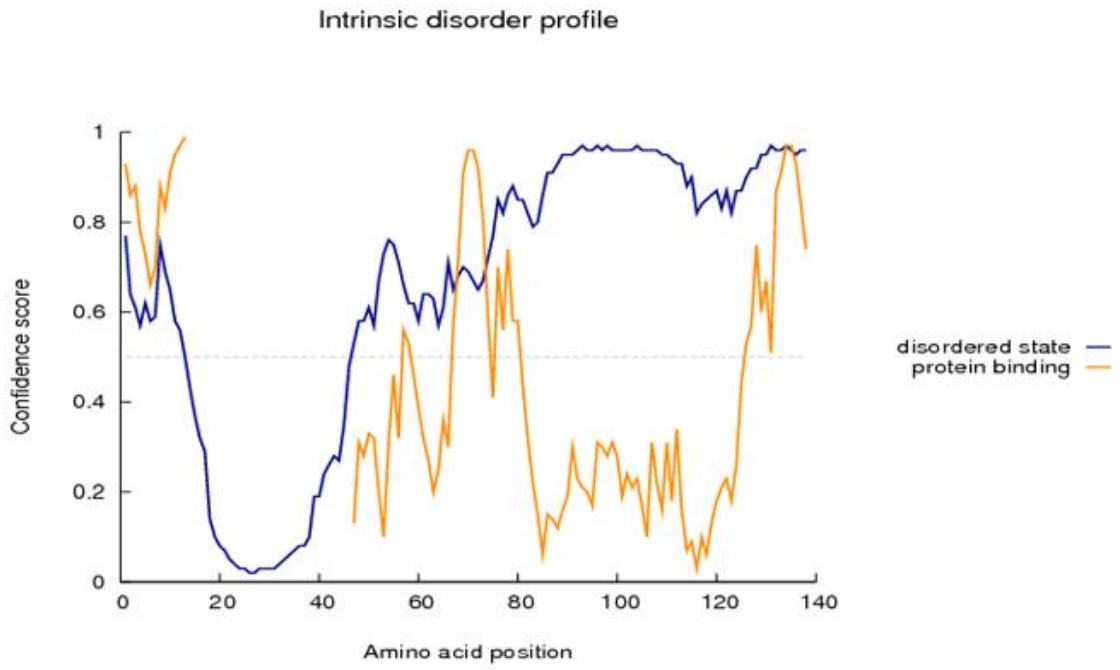
Supplemental material

```
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
```

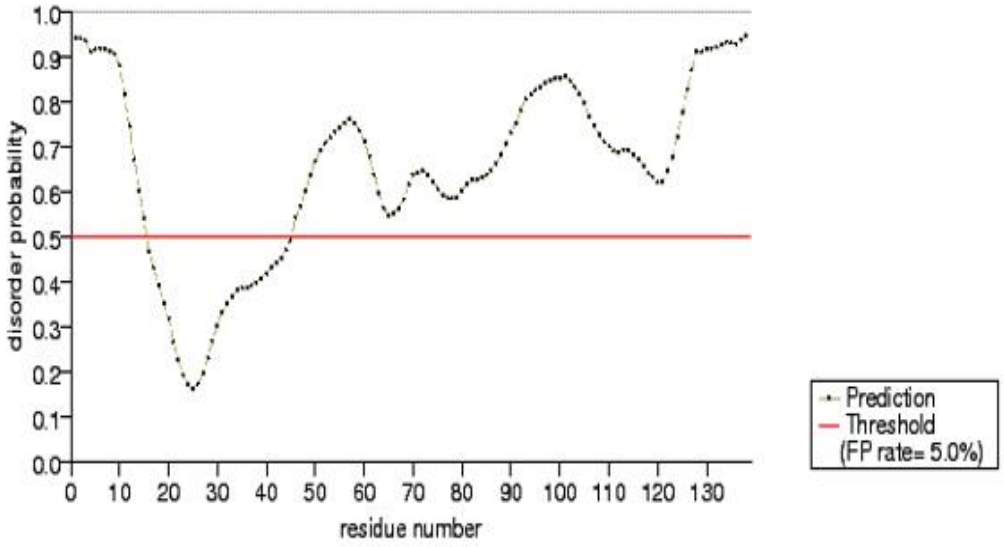


Supplemental Fig. 3.1: Primary sequence of cAlb analyzed with PONDR indicates that it is mostly disordered with a 28 amino acid ordered region at the N-terminus (residues 12-40) and a small proportion of ordered structure from residues 75-84. Figure by Furr

Supplemental Fig. 3.2: 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. Figure by Furr



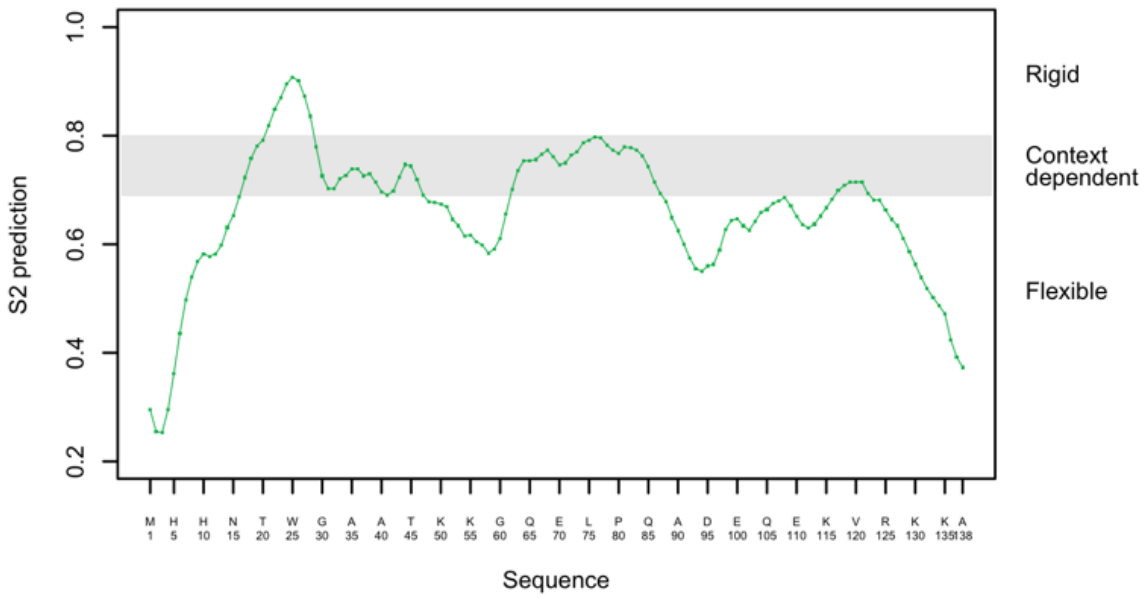
Supplemental Fig. 3.3: Primary sequence of cAlb analyzed by DISOPRED. Amino acids in the input sequence are considered disordered when the blue line is above the grey dashed line, that is the confidence score is higher than 0.5. The orange line shows the confidence of disordered protein binding residue predictions. Residues from ~15-45 were predicted to contain an ordered structure. Figure by Furr



Red: Disordered residues Black: Ordered residues)

1	MRGSHHHHHH	GSACNNVLST	AQQVWLRKLG	GAKPAVNENA	GGIITAGQAK	50
51	RSASKPEKGG	ERFRQLKEEE	KKKKLIKALP	VEEVQPLASA	SASNDGSDVE	100
101	NNKEQEVTEE	SNTSKVSQEV	QSFSRERRSK	RSKRKPVA		150

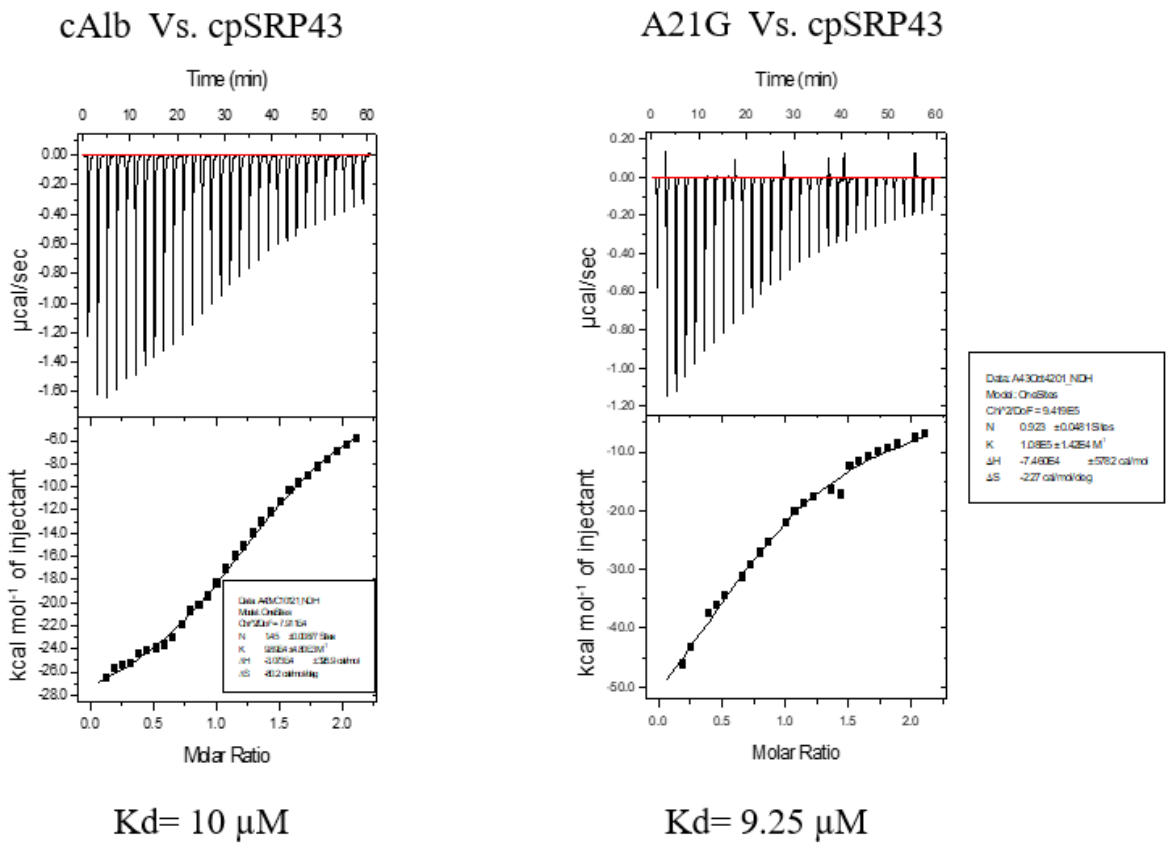
Supplemental Fig. 3.4: Primary sequence of cAlb analyzed by PrDOS2 indicates that residues 16-45 form an ordered structure. Figure by Furr



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

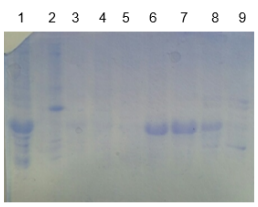
Predictor of protein backbone dynamics. A value of 1 means complete order (stable conformation), while a value of 0 means fully random bond vector movement (highly dynamic).

Supplemental Fig. 3.5: Primary sequence of cAlb analyzed by DynaMine indicates that residues 24, 25 and 26 have a high probability to contribute to a stable conformation. Figure by Furr

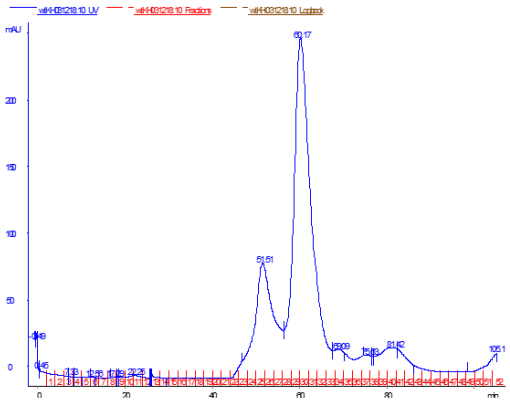


Supplemental Fig.6: ITC of cAlb titrated into cpSRP43 (left) and ITC of A21G titrated into cpSRP43 (right).

SEC elution of cAlb at 60 mls

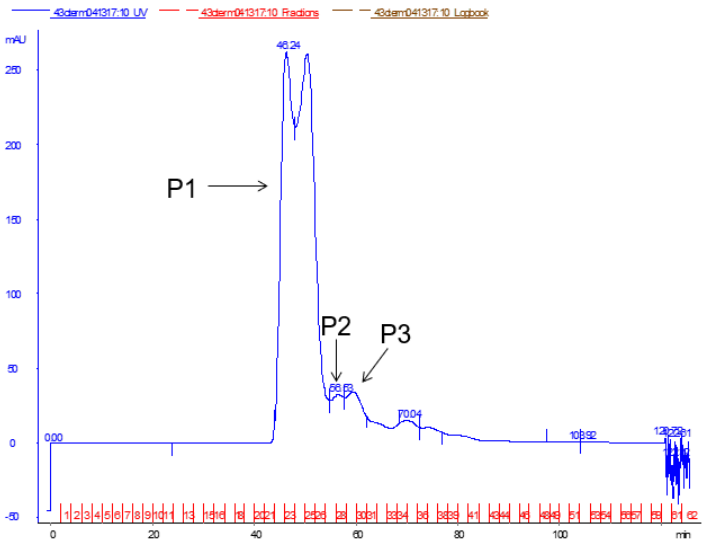


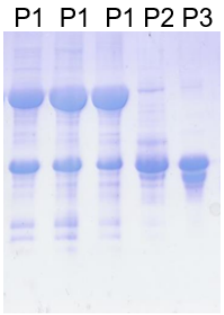
1= Preload
2= Fraction 25
3= Fraction 26
4= Fraction 27
5= Fraction 28
6= Fraction 29
7= Fraction 30
8= Fraction 31
9= Fraction 32



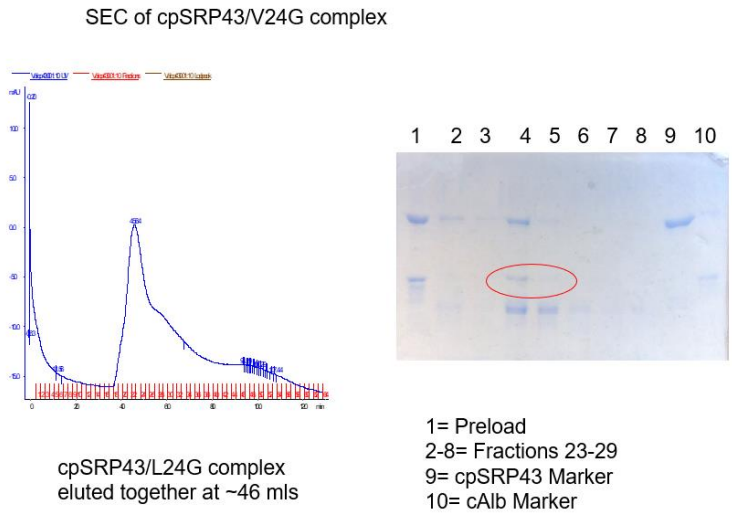
Supplemental Fig.7: Pure cAlb elutes at 60 mls.

cpSRP43/cAlb complex elutes at 46.24 mls

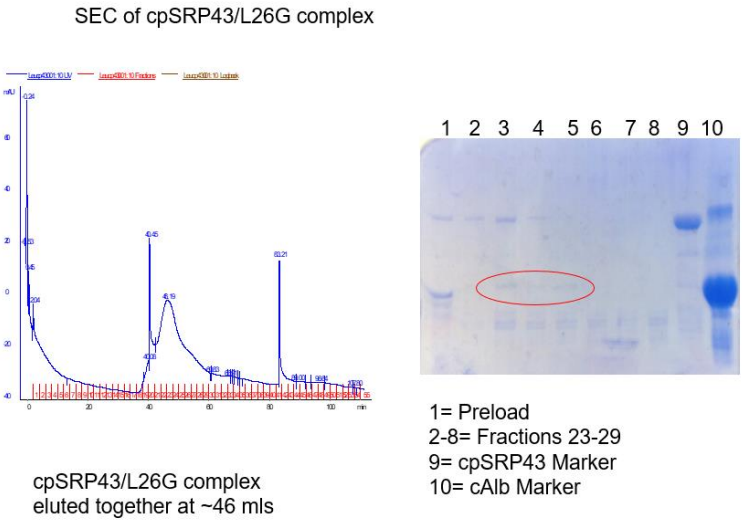




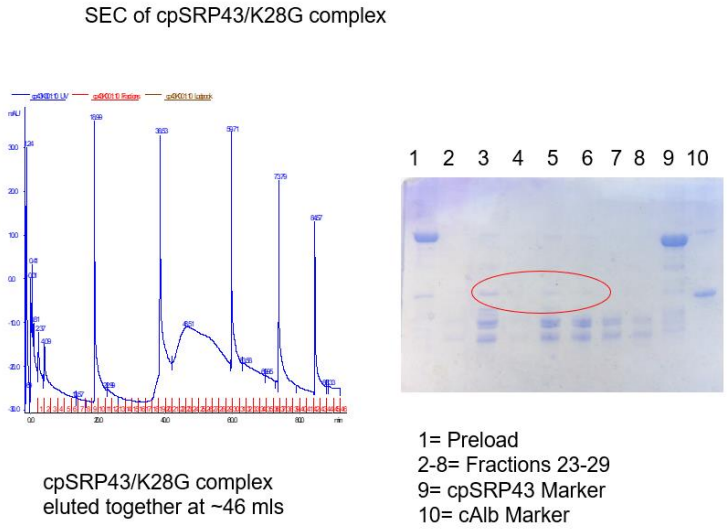
Supplemental Fig. 8: The complex of cpSRP43/cAlb elutes at 46-52 mls of buffer volume.



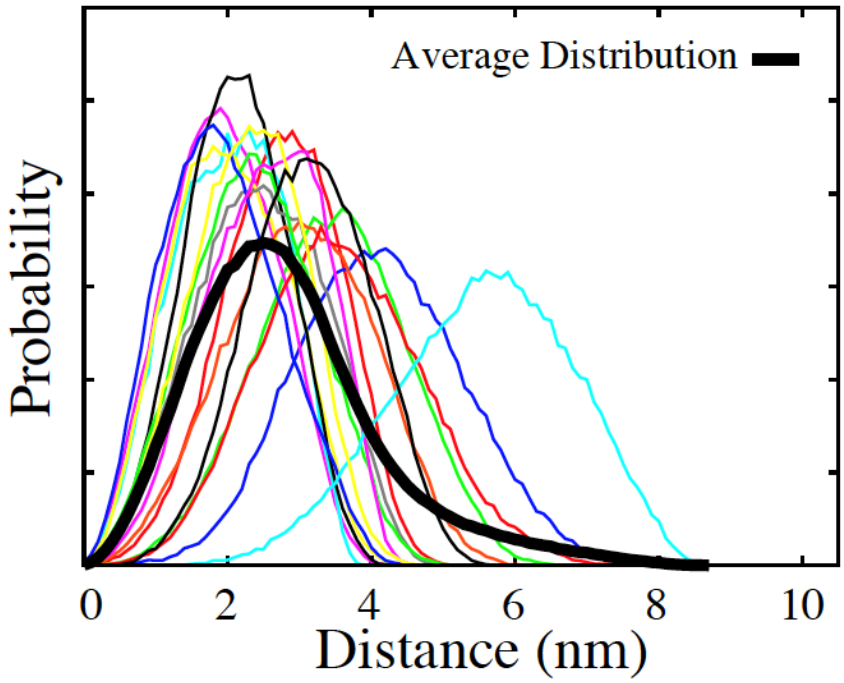
Supplemental Fig. 9: The complex of cpSRP43/V24G elutes together at 46-52 mls of buffer volume. The V24G band is circled in red.



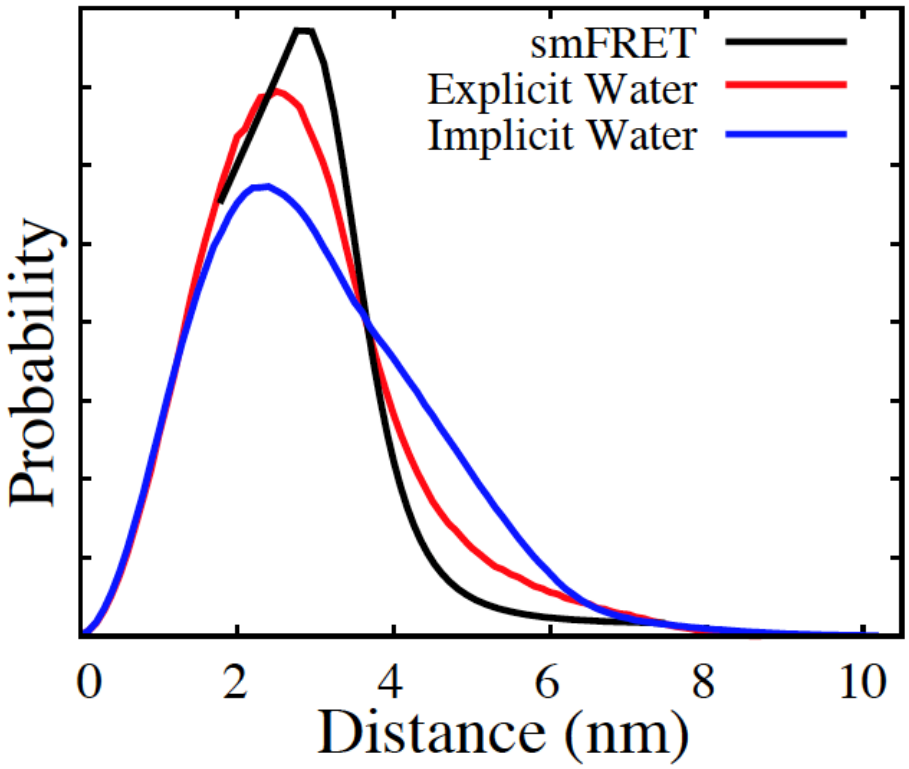
Supplemental Fig. 10: The complex of cpSRP43/L26G elutes together at 46-52 mls of buffer volume. The L26G band is circled in red.



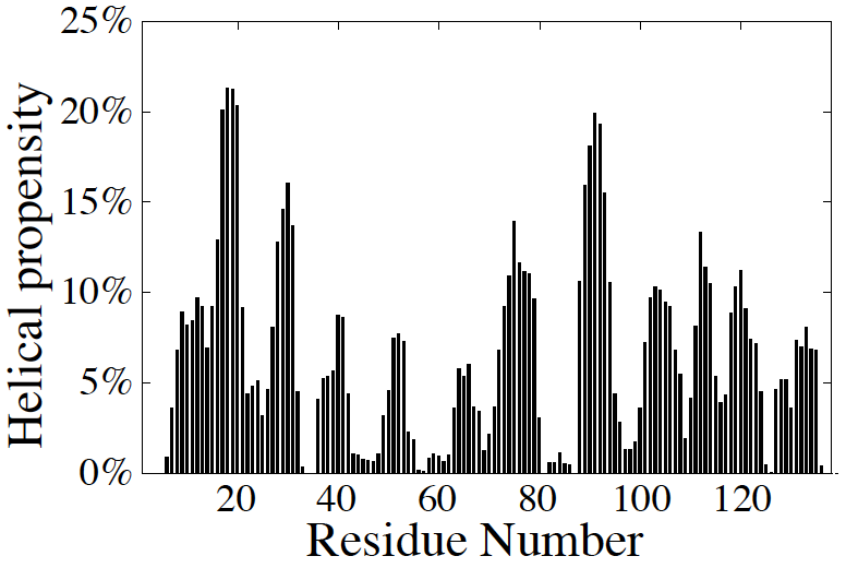
Supplemental Fig. 11: The complex of cpSRP43/K28G elutes together at 46-52 mls of buffer volume. The K28G band is circled in red.



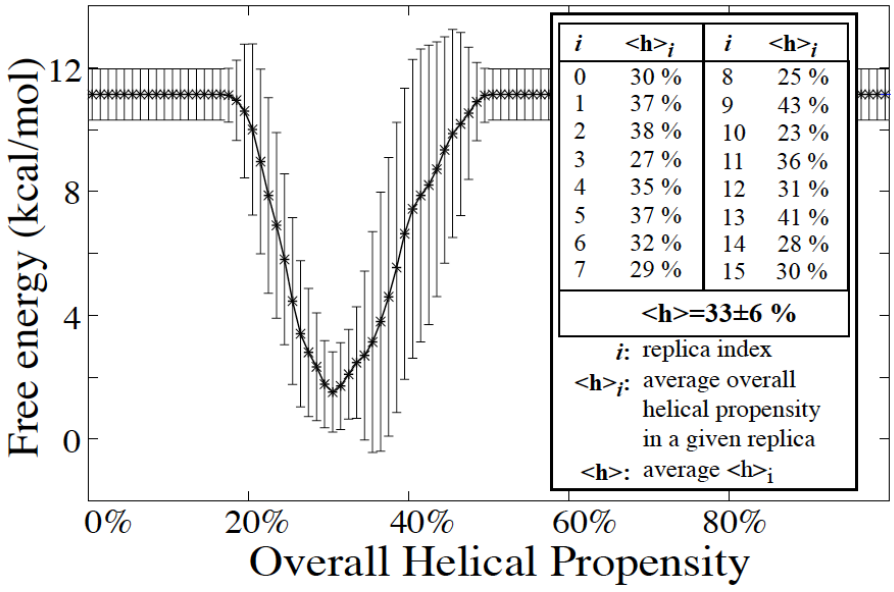
Supplemental Fig. 3.12: An average dye-dye distance distribution was calculated from 16 independent well-tempered metadynamics simulations of cAlb (explicit water). The average distribution is shown in black. Alexa 488 and Alexa 594 dyes were aligned with both C14 and S52C to calculate the computational distance distributions. Related to Figure 3.9. Figure by V. Kumar



Supplemental Fig. 3.13: Comparison of experimental smFRET and computational dye-dye distance distributions. Alexa 488 and Alexa 594 dyes were aligned with both C14 and S52C to calculate the computational distance distributions. The experimental smFRET (black) and explicit water (red) computational dye-dye distance distributions are in agreement. The implicit water distribution (blue) is similar but does not show a pronounced agreement with the experimental data (Related to Figure 3.9). Figure by Baucom and V. Kumar



Supplemental Fig. 3.14: Helical propensity per residue from implicit water simulations. STRIDE algorithm was used to calculate the helical propensity of individual amino acid residues. The region of interest (residues 15-38) has a peak helical propensity of around 20% in implicit water. Figure by V Kumar



Supplemental Fig. 3.15: Overall helical propensity from implicit water simulations. Weighted averaging of the free energies calculated along the alpha collective variable during the well-tempered metadynamics simulations was used to calculate the overall helical propensity. Average helical propensities from each of the 16 independent cAlb simulations are shown inset. Figure by V. Kumar

Chapter 4: Changes in the Structural Ensembles of the C-terminal domain of Albino3 Insertase upon Binding to cpSRP43

4.1 Abstract:

Light-harvesting chlorophyll-binding protein (LHCP) is post-translationally targeted for insertion into the thylakoid membrane of chloroplasts. The mechanism behind this targeting involves a unique chloroplast version of the signal recognition particle (SRP). This version of SRP contains a subunit unique to chloroplasts known as cpSRP43. This novel protein has been shown to interact with the stroma-soluble C-terminus of the thylakoid-bound Albino3 insertase (Alb3C-term). Although Alb3 is a member of a larger family of insertates, consisting of YidC in bacteria and Oxa1c in mitochondria, the fact that it interacts with the unique cpSRP43 subunit rather than the ubiquitous cpSRP54 is an important question that will hopefully shed light onto both common and unique aspects of protein targeting. The Alb3C-term has been described as intrinsically disordered, but chapter XX has shown that it contains transient secondary structure, resulting in an ensemble of states. In this study, we use a combination of single-molecule FRET and molecular dynamics simulations in a concerted fashion to examine how the structural ensembles of the Alb3C-term changes upon binding cpSRP43 to evaluate its role in post-translational targeting. Alb3C-term adopts a more compact structure upon binding cpSRP43 compared to the unbound form, suggesting that transient secondary structure in this intrinsically disordered region plays an important role in post-translational LHCP targeting.

4.2 Introduction

Signal recognition particle (SRP) is a multicomponent complex of protein subunits and RNA that was found to be required for the co-translational targeting of nascent membrane proteins to their destination.[2, 3, 10] SRP is a conserved mechanism for targeting nascent membrane

proteins to their destinations that has been found across all domains of life.[1] A specialized SRP has been found to exist in the chloroplasts of plants.[29] This chloroplast SRP (cpSRP) consists of a conserved 54kDa subunit (cpSRP54) and a novel 43kDa subunit (cpSRP43).[30] Interestingly, cpSRP does not contain a bound RNA molecule like all other known forms of SRP in the cytosol of eukaryotes, bacteria, and archaea.[3, 15, 134]

This cpSRP complex has been implicated in the post-translational targeting of light-harvesting chlorophyll-binding protein (LHCP) for insertion into the thylakoid membrane.[30, 34] In addition to cpSRP, the targeting to the thylakoid membrane was shown to require one other soluble factor, cpFtsY.[39, 40] After the LHCP is successfully targeted to the thylakoid membrane, membrane insertion requires interaction with Albino3 (Alb3), an Oxa1P homolog.[44] After insertion of LHCP, the cpSRP complex is recycled with the use of GTP.[64] Further, it was shown that although cpSRP43 is not a GTPase, it is required for the GTPase activity of cpFtsY and cpSRP54.[64] Truncated mutants showed that cpSRP43 required the ankyrin repeat domains and the second chromodomain were necessary for stimulation of GTP hydrolysis by Alb3.[64] This cpSRP43 interaction with Alb3-cterm is subject to some debate in the literature. It was originally proposed that chromodomains 2 and 3 of cpSRP4.[52, 135] However, this claim and its methodology were refuted by Lewis et al. who proposed the primary interaction site to be the ankyrin repeats of cpSRP43.[64] Further evidence that the chromodomains provide the primary interaction sites was subsequently provided using a combination of NMR, X-ray crystallography, and binding studies.[65]

To explore the intrinsically disordered structure of Alb3-cterm, single-molecule Förster resonance energy transfer (smFRET) provides the ability to examine inter-residue distances, a distribution of conformations, and the ability to observe conformational changes.[136] By creating

multiple mutants, distance distributions from smFRET can be used with molecular dynamics simulations to narrow the search space for the bound state. At the same time, molecular dynamics simulations can provide valuable insight in where in the protein to examine for structure.

4.3 Methods

Mutations

Mutations and purified Alb3-cterm mutants from Pisum sativum were provided by Patience Okoto from the Kumar group at the University of Arkansas. These mutations and purification were critical for this work but will not be included in this dissertation.

Fluorescent Labeling

Fluorescent labeling of cysteine residues was done in a statistical fashion using maleimide chemistry. Alb3-cterm mutants (1mg/mL) were labeled with Alexa488 C5-maleimide and Alexa594 C5-maleimide at molar ratios of 1:1:20, Alb3-cterm:Alexa488:Alexa594. Free dye was then separated from the labeled protein with the use of gravity gel filtration using Bio-Gel P6 polyacrylamide gel. Additionally, each mutant was labeled with donor-only as a control for smFRET. This labeling was performed in the same fashion, but a 1:10 protein to Alexa488 labeling ratio was used. All experiments were performed using a buffer consisting of 10mM HEPES, 10mM $MgCl_2$, and 100mM KCl with 10% glycerol at pH 7.5.

smFRET

The experimental FRET methodology is the same as that described in Gao et al. except that a γ factor of 0.6 was used and a new home-written software written in Python was used .[61]

Molecular dynamics

Molecular dynamics simulations were performed by Vivek Kumar from the Moradi group at the University of Arkansas. These simulations are heavily intertwined with the research presented in this chapter and will be presented accordingly.

4.4 Results & Discussion

Figure 1 is the computationally predicted propensity of each residue to form a helical secondary structure. Additionally, the span of three mutant dye locations is shown. The region examined in all of the mutants includes the helix-forming region from residue 77 to residue 100 to examine its significance.

Successful labeling of the Alb3-cterm mutants was verified using UV-vis absorption spectroscopy. Figure 2 illustrates the dyes pair positions for the three mutants used in this study, N15C, S124C (blue), S12C, S111C (green), and S54C, S117C (red).

After labeling, smFRET was performed on each FRET labeled and donor-only labeled mutant. Due to the statistical nature of the labeling process, each protein molecule is labeled with some combination of one of the dyes or no dye at each cysteine location. This result in several possibilities. Occurrences with no dyes, one acceptor dye, or two acceptor dyes are not of concern because those molecules will not be efficiently excited by the 485nm laser. However, molecules with only donor dyes (1 or 2) will be excited and give a signal which is indistinguishable from very low FRET using this setup. This artifactual zero-FRET peak can be measured by obtaining the smFRET curve of the donor-only labeled dye. A representative of a donor-only smFRET curve is shown in Figure 3. Further, this peak can be normalized to the zero-FRET peak for the FRET-labeled data and can be used to subtract this artifact out the FREt-labeled smFRET curve.

The first mutant N15C/S124C has a very large separation between the labeling sites. For an intrinsically disordered protein, this long distance should result in a very low FRET with a wide distribution. Shown in Figure 4 is the raw smFRET curve that was obtained for this mutant. The low FRET efficiency between the two dyes blends the peak with the donor-only artifact peak. Further characterization of this peak requires a donor-only subtraction.

Next, the FRET labeled mutant was incubated with 1μM wild-type cpSRP43 to force the interaction. Although there is some dispute in the literature about the K_d value for this interaction,[52, 64] this large excess of cpSRP43 will force binding. The smFRET histogram for this mixture was then obtained, Figure 4, and examined for differences caused by binding. Additionally, to ensure that cpSRP43 does not cause artifacts due to background fluorescence, the 1μM cpSRP43 was also examined for FRET signal. The minimum amount of background fluorescence introduced with the addition of such a high concentration of cpSRP43 results in a negligible FRET signal as shown in purple on Figure 4. The smFRET histogram in Figure 5 shows a noticeable increase in the high FRET state upon binding the cpSRP43 compared to unbound Alb3C-term. Also, a decrease in the low-FRET peak to donor-only peak ratio results in the ability to distinguish the two peaks more clearly.

Using the aformentioned donor-only subtraction from the smFRET histograms and conversion to probability space allows for direct comparison of Alb3-cterm N15C, S124C with and without binding of cpSRP43, as shown in Figure 6. This comparison shows that molecules from the low FRET peak are shifted to higher FRET conformations upon binding to cpSRP43. The broadness of this high FRET state, ranging from ~0.6 - 1.0, indicates that molecules in conformations corresponding to the binding of cpSRP43 are quite flexible in this region.

This same process was performed for the second mutant, S12C/S111C. This mutant covers a similar range in the protein and should give similar results. Although it was originally produced due to the fact that not all mutants purify and label, here it serves as a good control experiment. As such, Figure 7 shows a smFRET histogram that similar to the N15C, S124C mutant. The low-FRET peak combines with the donor-only artifact peak to be almost indistinguishable. Upon binding cpSRP43, the smFRET curve, Figure 8, shows same appearance of high FRET conformations. Comparison of the S12C, S111C smFRET curves with and without cpSRP43 after donor-subtraction and conversion to probability, Figure 9 shows similar results to the previous mutant. The peaks below zero FRET are the result of a poor donor-only under-subtraction. Despite this, the formation of the high FRET peak is evident.

For the final mutant in this study, S54C/S117C, a smaller region was probed. According to molecular dynamics simulations, a FRET efficiency of approximately 0.5 was predicted for this mutant. The smFRET histogram obtained for this mutant, Figure 10, strongly agrees with that predicted from molecular dynamics simulations. The smFRET histogram upon binding cpSRP43, shown in Figure 11, shows a reduction of this peak centered around 0.5 and a marked increase in a high FRET peak. As opposed to previous mutants, this peak is quite narrow.

In Figure 12, the subtracted FRET curve for the unbound state of S54C, S117C mutant of Alb3-cterm shows a wide distribution of conformations. This suggests that the protein is very flexible in this region. Upon binding cpSRP43, however, the Alb3-cterm seems to be locked in close-proximity, high-FRET state.

Based on the empirical smFRET histograms, the mutant dye pairs exhibit very high FRET efficiency peaks in the cpSRP43-bound state. By assuming that the bound state of the Alb3-cterm is one of many possible conformations from the molecular dynamics simulations of

the unbound Alb3-cterm, a search for the bound state using the empirical smFRET distance parameters can be conducted. Using clustering analysis of the generated Alb3-cterm conformations is underway to predict the bound structure. Two possible structures of the bound state based on the FRET data are shown in Figure 13.

All mutants show Alb3-cterm tested show a shift to a high FRET efficiency upon binding cpSRP43. This indicates the Alb3-cterm adopts a more compact structure in the regions being probed. All regions probed include the helical region around residues 77-100. Since the S54C, S117C mutant shows a dramatic shift to higher FRET, we hypothesize that this helical propensity is increased upon binding to cpSRP43, allowing the helix to form. Further, the shift in this region is responsible for the increases in FRET efficiency in the other two mutants tested. This formation of this helix may be critical to the cpSRP43-Alb3-cterm interaction.

4.5 Conclusions

In this work, the structure of the stroma-soluble C-terminal region of Alb3 was explored using smFRET and MD simulations. Further, examination of the structure upon interaction with cpSRP43 showed a preferred compact structure was adopted by Alb3-cterm. By using smFRET results to narrow down the MD conformational search space, potential bound conformations of Alb3-cterm are proposed. Future work in this area could include using the current structure bank to propose smFRET experiments to further narrow the possible favored conformations. One possible experiment would be to introduce a helix-breaking mutation in the region probed and examine the effect on the binding with cpSRP43. Additionally, exploration of the conformational dynamics of cpSRP43 upon binding with Alb3-cterm could shed light into the interaction between the two proteins.

4.6 Figure

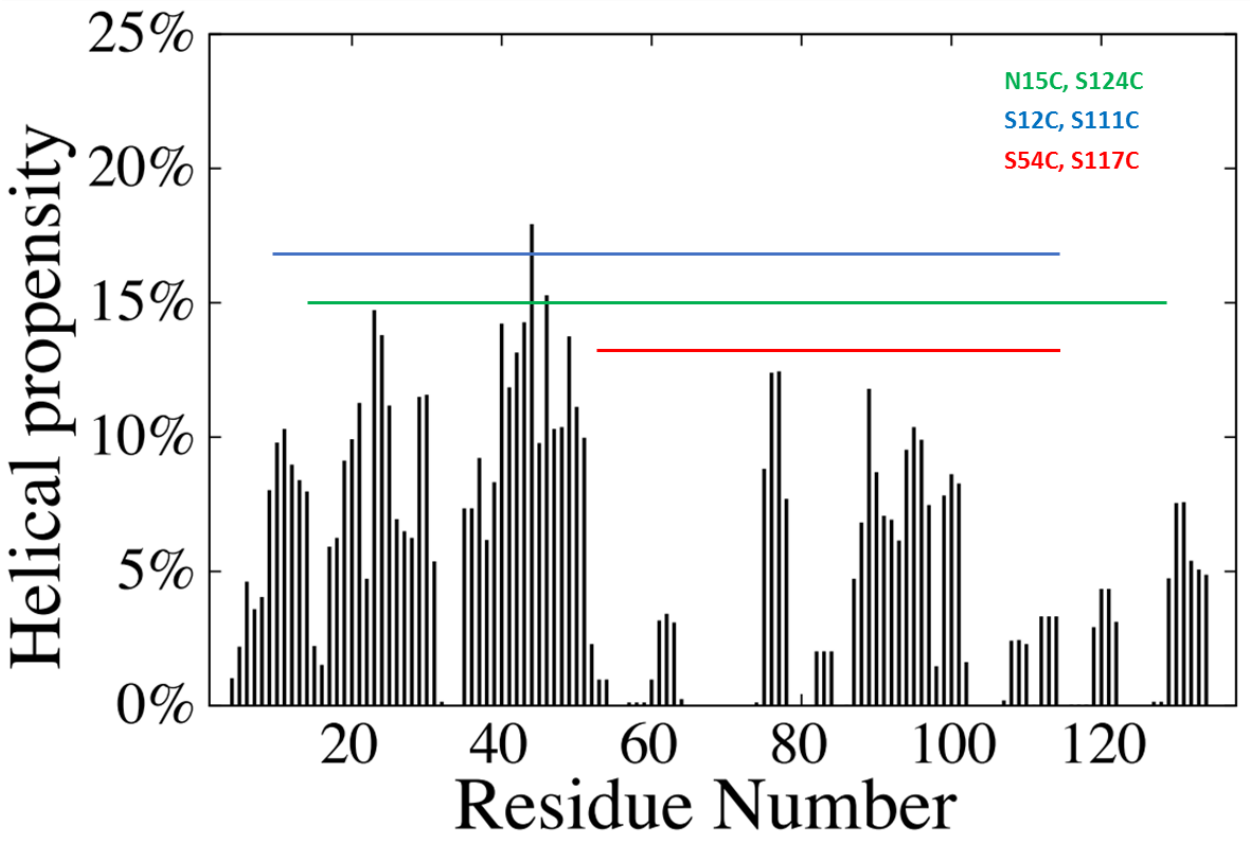


Figure 4.1 shows the predicted helical secondary structure propensity by residue for Alb3-cterm as calculated from molecular dynamics simulations.

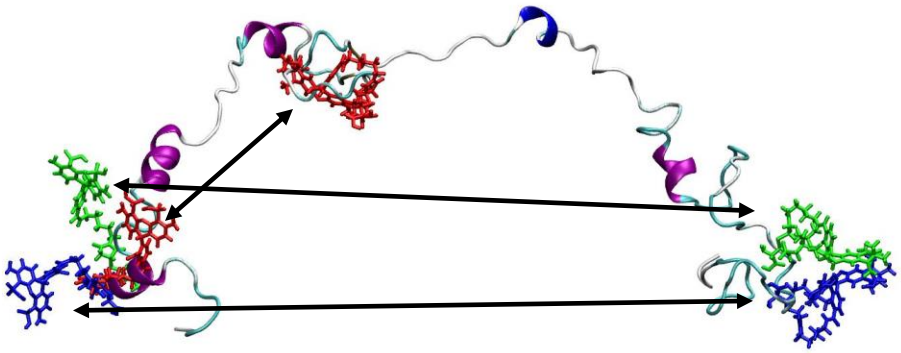


Figure 4.2: A representative conformation of the Alb3-cterm mutant showing the three dye pair positions tested. Alexa488 and Alexa594 fluorescent dyes are color coded in the structure to distinguish mutant labeling sites. Blue corresponds to the N15C, S124C mutant, green corresponds to the S12C, S111C mutant, and red corresponds to the S54C, S117C mutant.

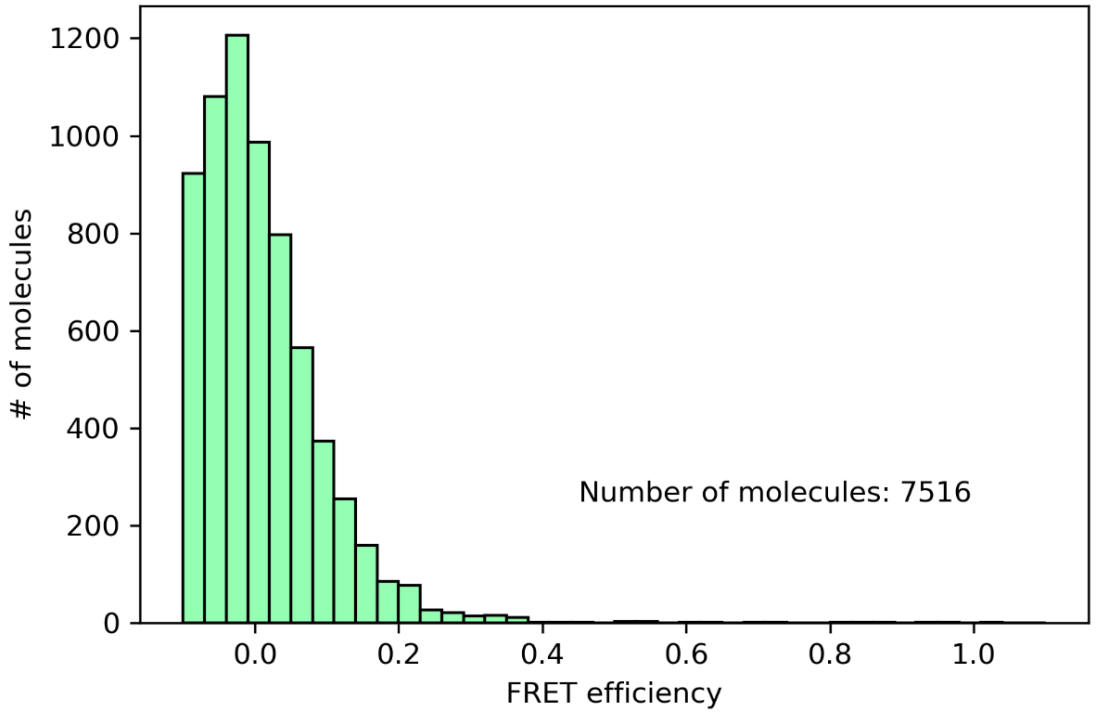


Figure 4.3: Representative smFRET histogram of donor-only labeled Alb3-cterm. This histogram is measured for every FRET mutant so that its contribution can be accurately subtracted.

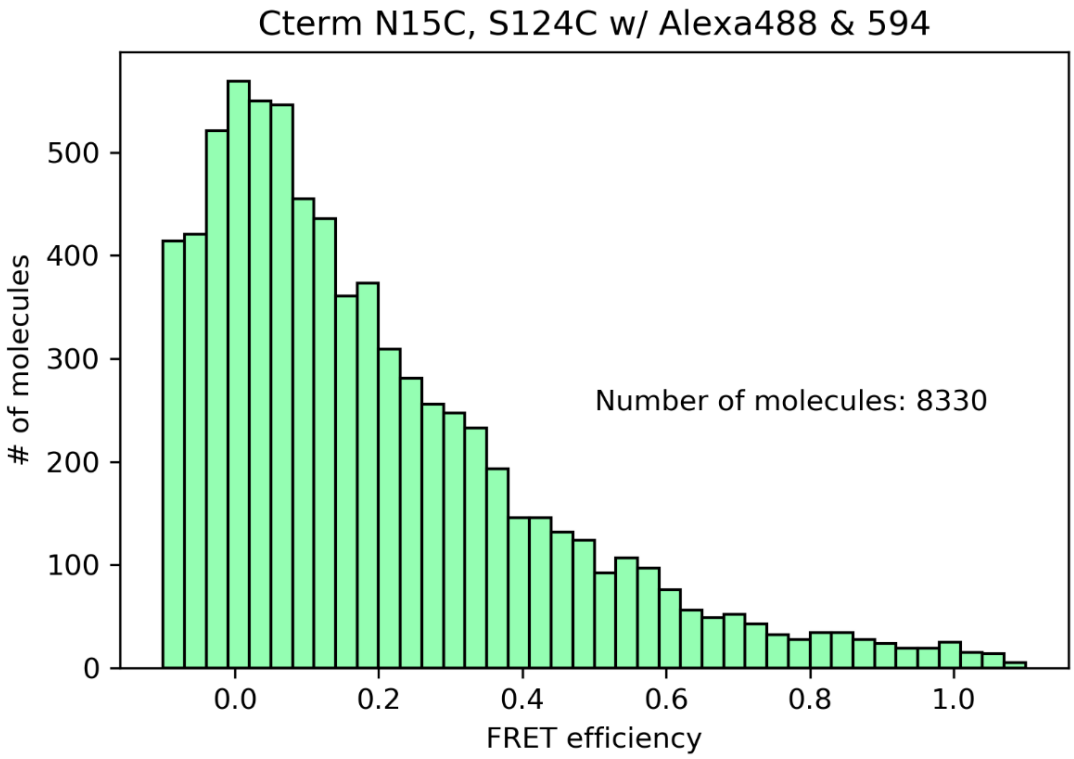


Figure 4.4: The raw smFRET histogram for the N15C, S124C Alb3-cterm mutant in the absence of cpSRP43. The peak centered around zero FRET primarily comes from molecules that are not labeled with an acceptor dye.

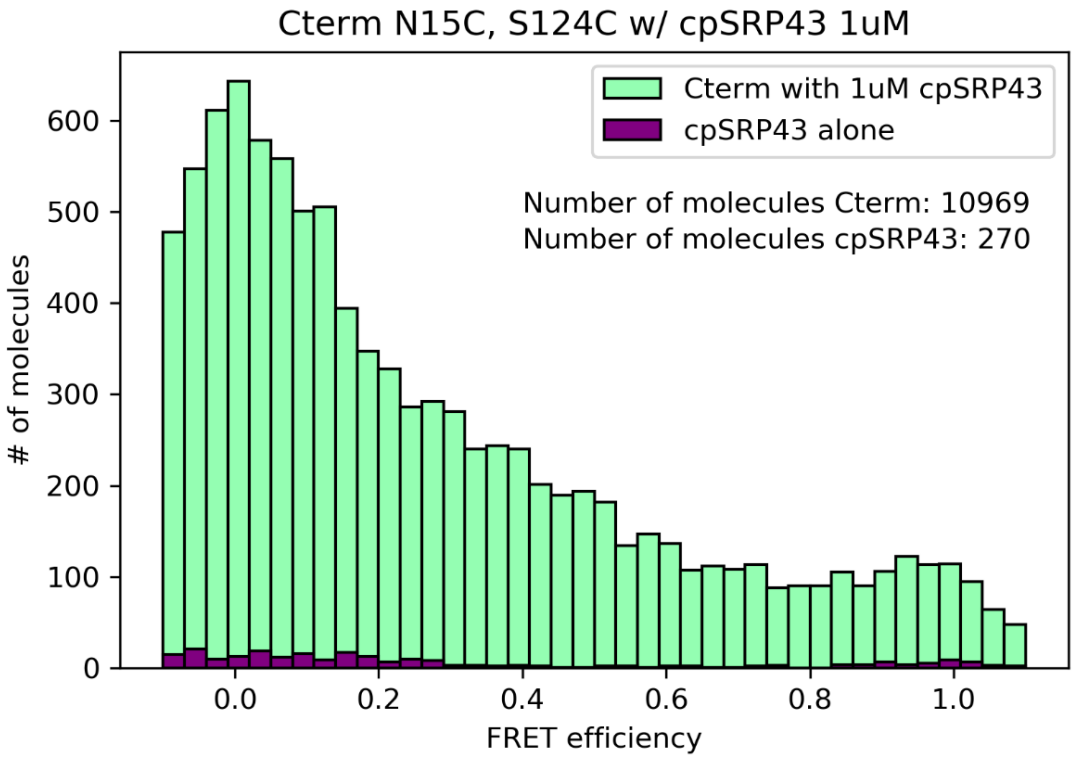


Figure 4.5: The smFRET histogram of FRET labeled Alb3-cterm mutant with the addition of WT-cpSRP43 (green) and the smFRET histogram created by background fluorescence from the cpSRP43 sample (purple).

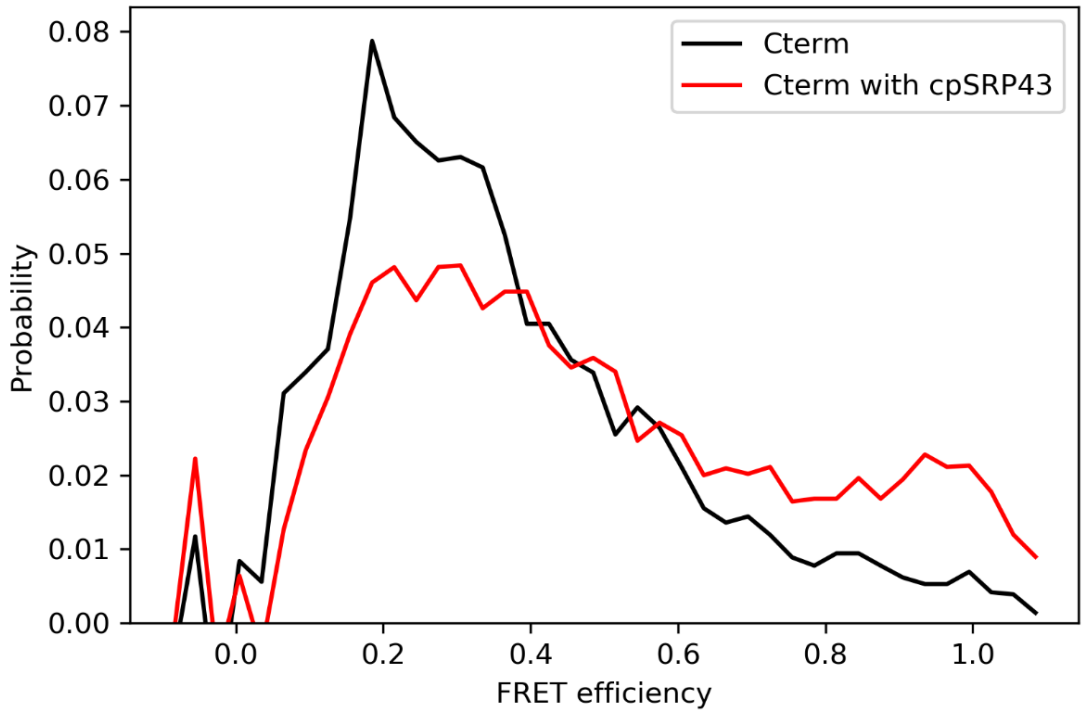


Figure 4.6: The smFRET curves for bound and unbound Alb3-cterm are shown here after donor-only subtraction for the N15C, S124C Alb3-cterm mutant.

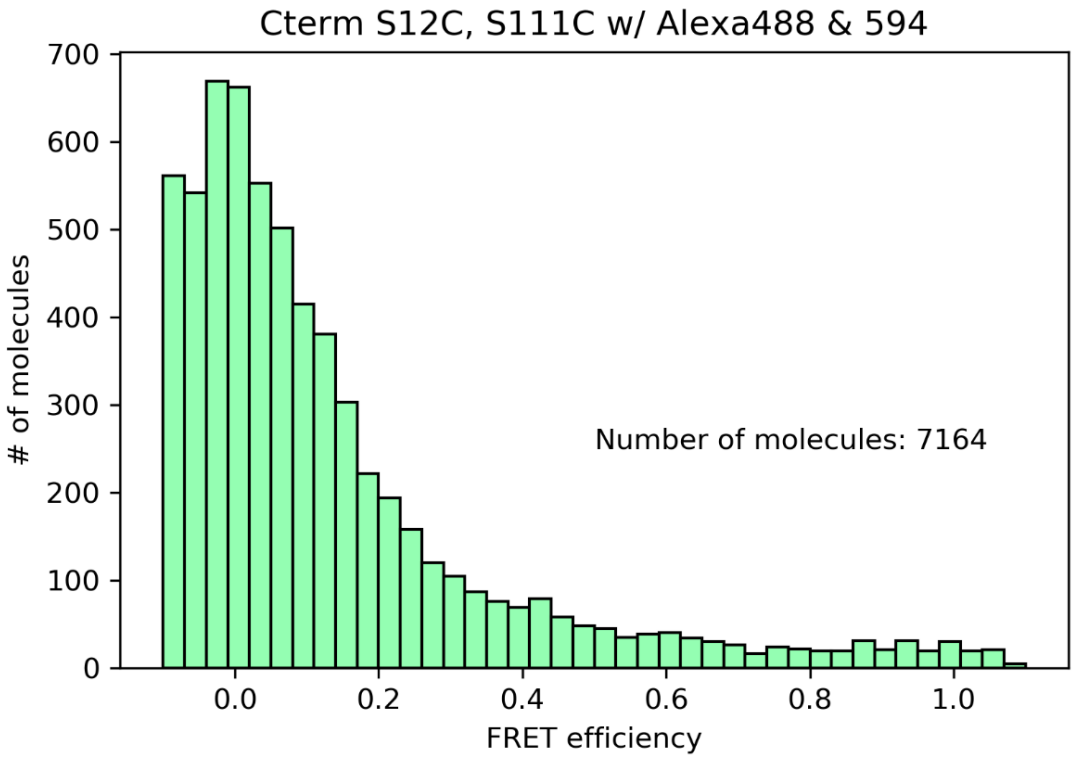


Figure 4.7: The raw smFRET histogram for the S12C, S111C Alb3-cterm mutant in the absence of cpSRP43. The peak centered around zero FRET primarily comes from molecules that are not labeled with an acceptor dye.

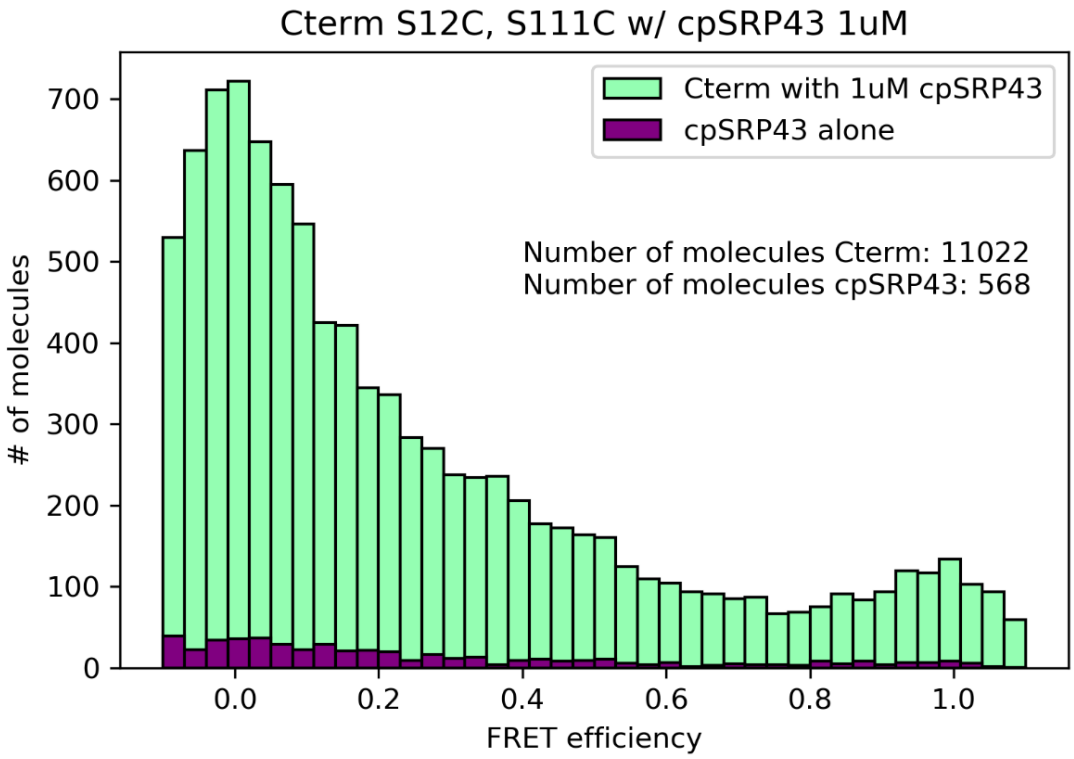


Figure 4.8: The smFRET histogram of FRET labeled Alb3-cterm mutant with the addition of WT-cpSRP43 (green) and the smFRET histogram created by background fluorescence from the cpSRP43 sample (purple).

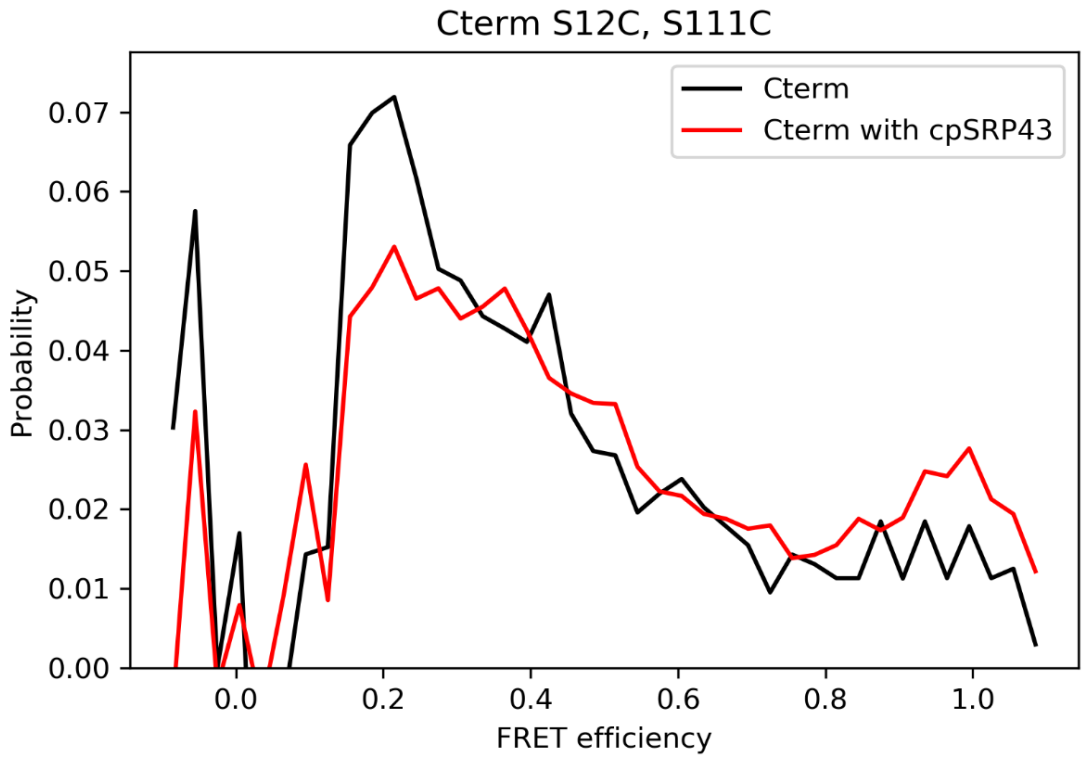


Figure 4.9: The smFRET curves for bound and unbound Alb3-cterm are shown here after donor-only subtraction for the S12C, S111C Alb3-cterm mutant.

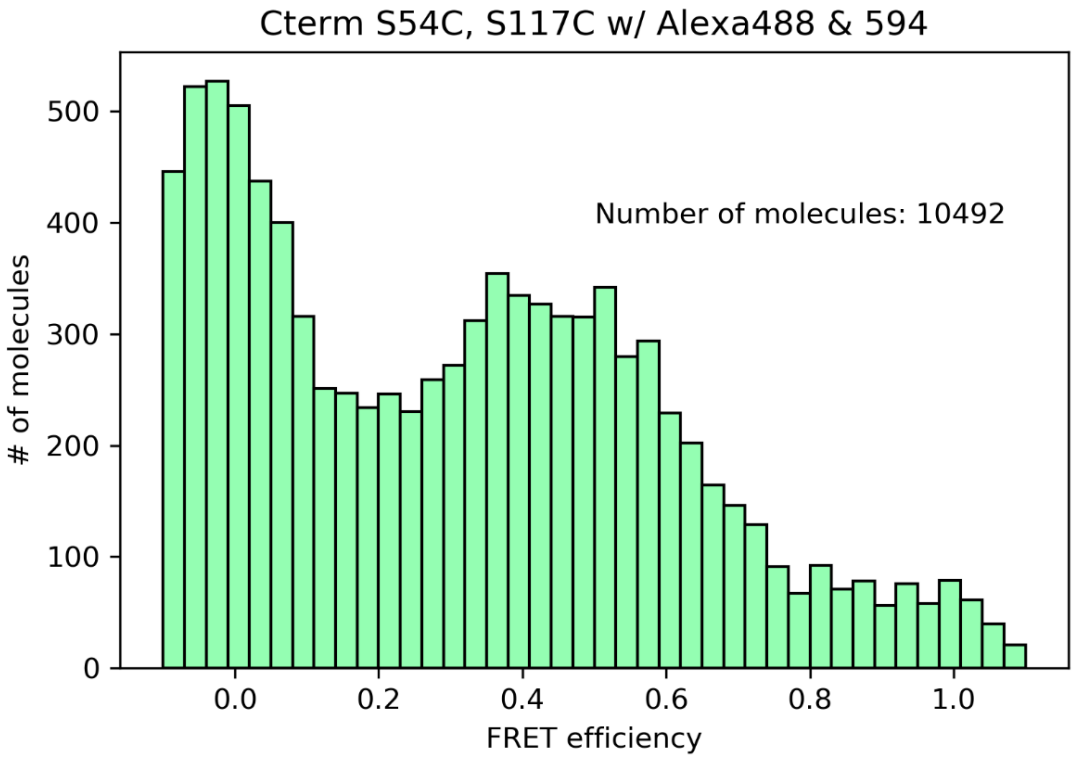


Figure 4.10: The raw smFRET histogram for the N15C, S124C Alb3-cterm mutant in the absence of cpSRP43. The peak centered around zero FRET primarily comes from molecules that are not labeled with an acceptor dye.

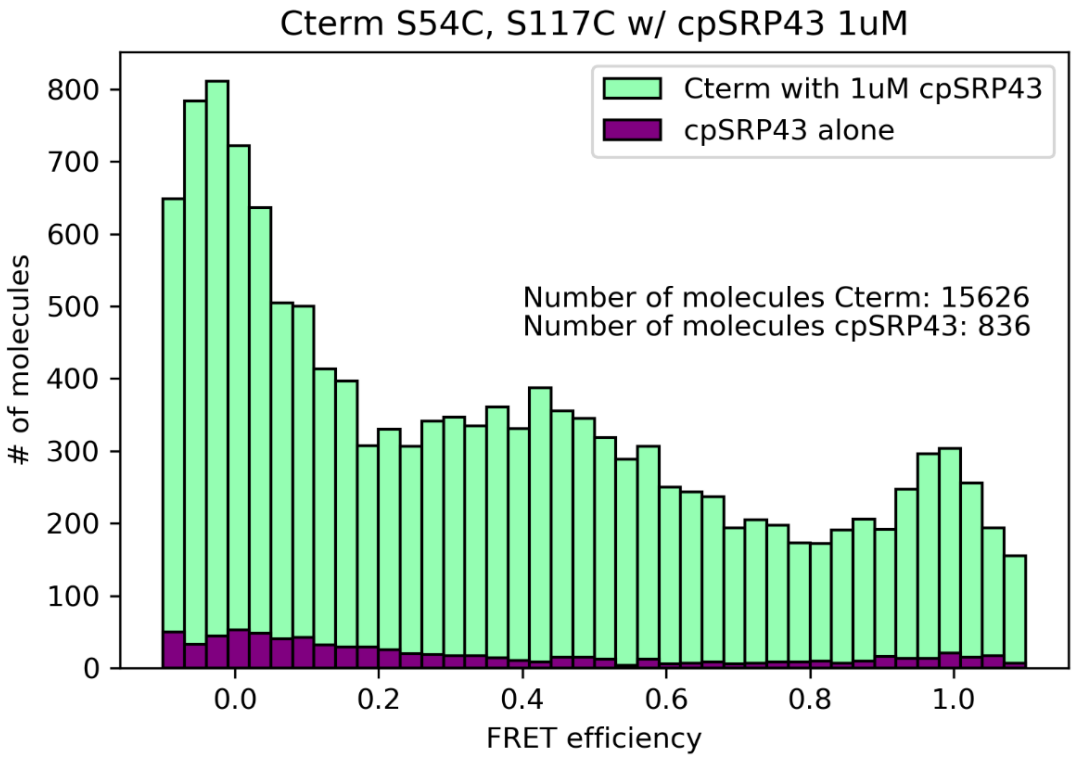


Figure 4.11: The smFRET histogram of FRET labeled Alb3-cterm mutant with the addition of WT-cpSRP43 (green) and the smFRET histogram created by background fluorescence from the cpSRP43 sample (purple).

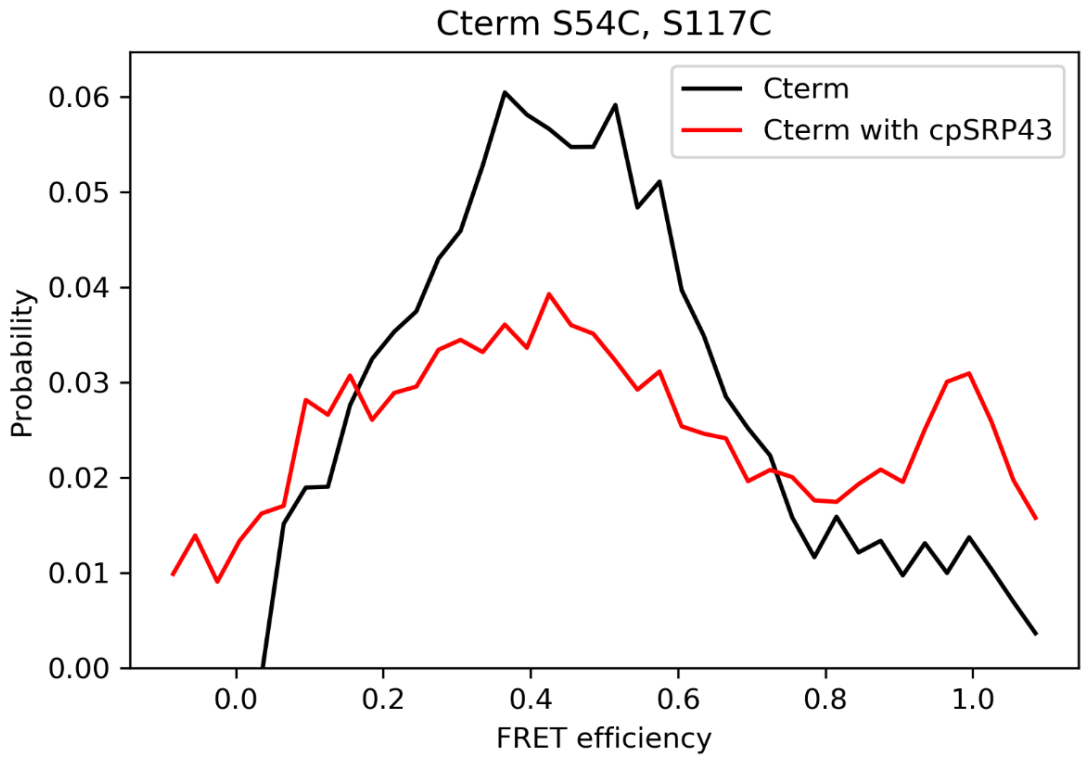


Figure 4.12: The smFRET curves for bound and unbound Alb3-cterm are shown here after donor-only subtraction for the S54C, S117C Alb3-cterm mutant.

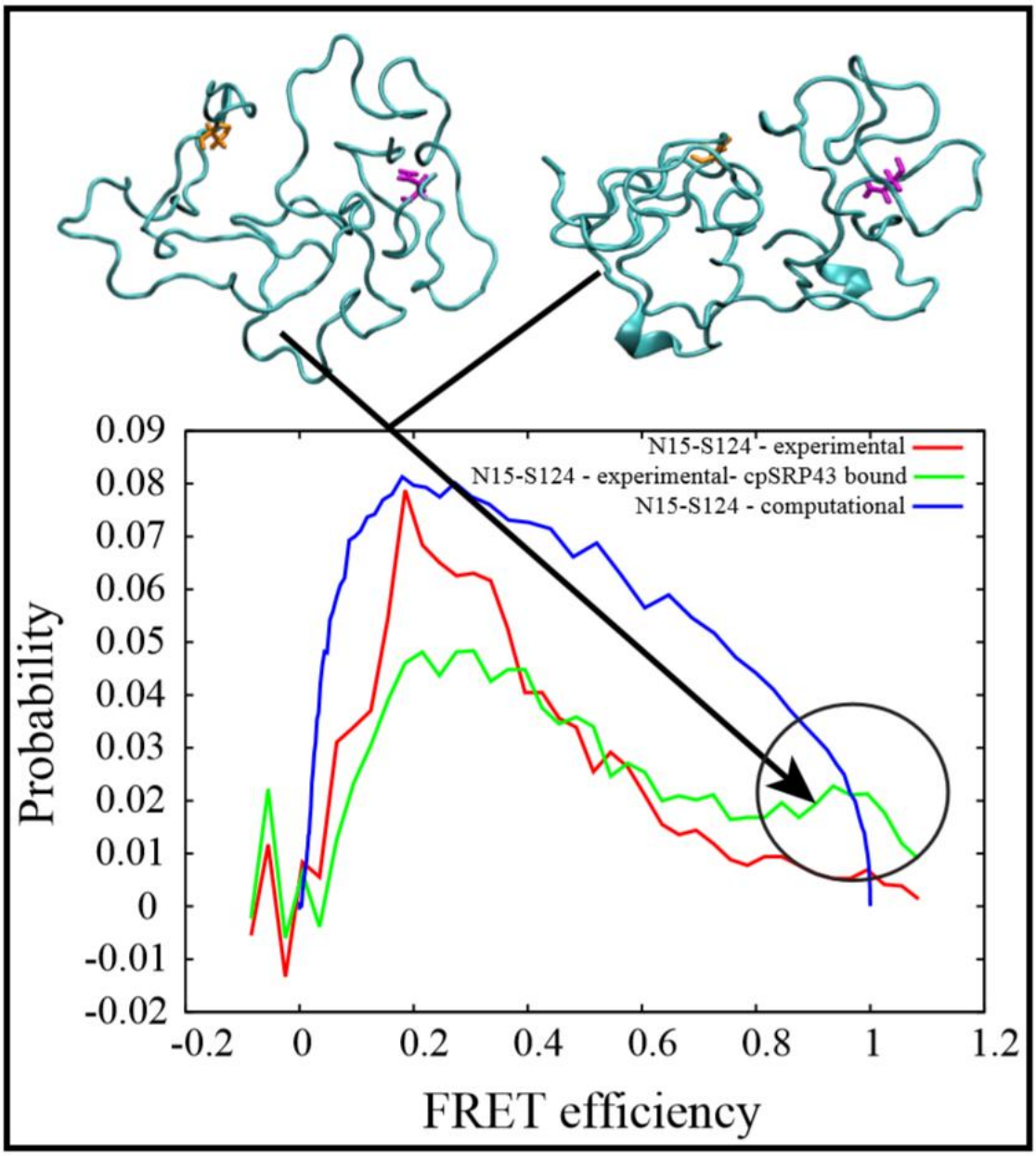


Figure 4.13: The search for the bound conformation of Alb3-cterm is performed by matching subpopulations from the unbound simulation to the differences in the FRET curves.

Chapter 5: Modular program for multi-parametric analysis of single-molecule fluorescence traces

5.1 Abstract:

Quantum dots (QDs) are fluorescent semiconductor nanocrystals that have great promise as fluorescent reporters for smFRET applications. This promise is due to the ability to tune the fluorescent properties of QDs to fit the particular application. To select for the desired, application-specific properties, appropriate synthesis methods must be developed. In the course of discovering the synthesis preparations that lead to the desired properties, detailed analysis of single QDs is required. One property that is of particular importance to smFRET applications is blinking. Blinking, a typically undesired property, is the tendency of the fluorescent reporter to stop fluorescing for up to many seconds at a time. Because blinking is a single-molecule property, it must be analyzed by observing many QDs from a sample for an extended period of time. This fact and the difficulty of the analysis and aggregation of population statistics makes this process a bottleneck for QD characterization. To streamline blinking analysis of QDs enough to make it a routine endeavor, an all-in-one program for the analysis of single QD fluorescence and population statistics was developed using the Python programming language. The program analyzes wide-field videos of many QDs simultaneously fluorescing and automatically calculates desired population statistics for direct comparison of samples produced via differing synthesis methods. Additionally, the software's ability to vastly improve sample populations led to the ability to highlight previously unrecognized intrasample heterogeneity. Finally, with the addition of changepoint analysis to the codebase, the program's utility was extended to examine other classes of fluorescent traces such as photobleaching and labeling stoichiometry traces.

5.2 Introduction

The physical processes of protein dynamics cover a wide range of timescales ranging from picoseconds to hundreds of seconds.[137] The focus of this work involves the dynamics of the interdomain flexibility of proteins and the conformational changes upon binding. These particular processes occur on longer timescales, typically ranging from milliseconds to hundreds of seconds.[137] The long timescales and potential presence of other protein binding partners makes these interactions difficult to characterize using molecular dynamics simulations.[138] One way to study these interactions in real time is to monitor distance changes by using smFRET. In smFRET experiments, distances between two specific locations on the protein can be monitored as a function of time.[78] Using multiple FRET pair locations on the protein, detailed information about the dynamics of the protein's flexibility and/or conformational changes upon binding a protein partner or substrate can be obtained. Additionally, the information obtained via smFRET can be used to bias the search space in molecular modeling to help produce a converged, accurate model in a reasonable period of time.[139]

In order to monitor conformational changes of proteins in real time, the smFRET-labeled proteins must be immobilized on a surface or restricted to some known observation volume.[140, 141] Further, the smFRET-labeled protein must be exposed to excitation laser light for periods of time longer than the timescales of the interactions of interest. One problem that arises with this type of experiment is that organic dyes, the most common fluorescent reporters used for smFRET, are susceptible to photobleaching on these timescales.[142] Photobleaching is the process by which a fluorophore undergoes some reaction that permanently disables its ability to fluoresce.[143] In organic dyes, photobleaching typically involves breaking the highly conjugated bonds via oxidation.[143]

Several strategies to mitigate photobleaching have been developed. First, since the has been shown to be excitation power dependent, using lower laser powers extends the average lifespan of a fluorophore before photobleaching. However, using this strategy has the large trade-off of reducing the signal-to-background and signal-to-noise ratios. Overall, this strategy leads to a loss of fluorescent information. This trade-off is usually not properly compensated with a justifiable increase in the lifespan of the fluorophore.

A second strategy to lengthen the lifespan of a fluorophore involves the use of a cocktail containing systems to scavenge oxygen from the buffer being used and to prevent the organic dyes from residing into the triplet electronic state.[144] Although this method shows significant improvements in the lifespan of the organic dyes, most of the molecules will still photobleach on timescales that are not useful for monitoring real-time protein dynamics. Figure 5.1 shows the bleaching of a common organic dye in the presence of multiple solutions designed to extend the lifespan of the fluorophore.[144] Additionally, the introduction of foreign species to the buffer at relatively high concentrations could affect the protein dynamics if some interaction with the protein of interest occurs. For example, one common oxygen-scavenging system involves the catalytic oxidation of glucose by using the glucose-glucose oxidase-catalase system.[145] This system results in the temporary production of peroxide in solution and the final result of lowering the pH of the buffer.

A third strategy of using a fluorophore that is much less susceptible to oxidative photobleaching is the focus of this chapter. Quantum dots (QDs) are a candidate for this role.[142] QDs are semiconductor nanocrystals that do not rely on a complex network of conjugated bonds to produce fluorescence.[146] Because of this fact, their fluorescence is much more resilient to the

prospect of oxidation. QDs can last several orders of magnitude longer than organic dyes before photobleaching occurs even if the best photobleaching preventative cocktails are used for the organic dyes.[142]

Although the use of QDs almost completely solves the photobleaching problem, the QDs are not without their own problems. One such problem, known as blinking, is the tendency of a fluorophore to enter into a long-lived excited state that is not conducive to fluorescence.[147] This so-called "off state" can last for many seconds at a time.[148] While this fluorophore is turned off, the results from the smFRET signal would be impossible to determine. To avoid this dark state, it is possible to design synthesis methods that produce QDs that have less of a tendency to populate it.[148] In QDs the origin of this dark state is often the trapping of the excited state in a surface defect.[149] This defect can come from a mismatch in the materials used to create the QDs. Strategies to mitigate surface defects and to reduce blinking have been a consistent area of research effort.[150] One such strategy is to make QDs that have many shell layers of low lattice mismatch.[151] This can isolate the fluorescent core of the QD from the true surface enough to minimize the probability of entering into the dark state. This strategy has been shown to produce almost non-blinking QDs. However, this particular strategy has a negative impact on the prospects for the use of QDs in smFRET applications. The shelling process makes the QDs larger with each shell. This results in very large particles at the end of the process. These large particles, when conjugated to a protein of interest, can hinder the natural dynamics and conformational changes of the protein.

As an alternative to shelling, using a variety material for shelling to minimize surface defects and create smaller QDs has also shown a tendency to create QD with different sizes and tendencies to blink. This strategy may produce QDs that are both relatively small in size and are

suitably non-blinking for smFRET or single particle tracking applications. In attempting to synthesize non-blinking QDs through various strategies, the analysis of blinking must occur for every sample. This analysis involves observing the fluorescence of immobilized QDs in the presence of continuous excitation laser light. Since QDs are heterogeneous, observation of many QDs is required to build population statistics that are indicative of the sample. Manually conducting blinking analysis from a wide-field fluorescence video can make it a time-intensive task. Often, only samples that have shown to maximize other properties such as quantum yield, size, or desired spectra are analyzed for their blinking properties. This bias can lead to a false impression of sample utility.

To decrease this time-intensive bottleneck that is prohibitive to routine blinking analysis, two strategies can be employed. The first is the use of wide-field methods that collect fluorescence from a large area at a time. This allows for the collection of fluorescence information from many quantum dots simultaneously. One common approach is to use a technique called total internal reflectance fluorescence microscopy (TIRF-M) in conjunction with an charge coupled display (CCD) chip to collect fluorescent intensity information as a function of location. The TIRF is a strategy of exciting the surface of a sample. By illuminating and collecting the fluorescent traces of many quantum dots simultaneously, vast amounts of individual information can be obtained from each quantum dot. By processing the desired properties for each QD, population statistics can be acquired. This data processing of TIRF-M movies to extract out relevant QD information then becomes the bottleneck of the process. It is intractable to manually extract individual traces for QD samples as a routine analysis. To overcome this problem, automated methods of analyzing TIRF-M data for immobilized QDs

must be developed. This development of custom software to automate the data analysis and collect blinking statistics from TIRF-M movies is the subject of this chapter.

Once the need for automating the blinking analysis was clear, the decision of which framework to use was examined. First, although the popular image analysis program ImageJ is extensible through adding modules, it was determined that multiple problems with such an extension existed. First, ImageJ primarily processes and produces images. Since the functionality is limited in scope, it would be very difficult to add a system by which to manage the individual traces in a convenient fashion. Once this decision not to use ImageJ was made, Python and MATLAB, being the premier of scientific computing languages, were considered. Python was chosen due to the free nature and extensive open-source architecture. MATLAB would have required an expensive license to distribute to any computer for which future code was to be run.

5.3 Methods

For a TIRF-M experiment on the blinking of quantum dots, the software MicroManager is typically used to obtain videos. The videos consist of hundreds to thousands of individual images called frames that are compiled into the ome.tif file format. This file format was designed specifically to contain the data and metadata for microscopy experiments. Fortunately, this specialty file is an extension of the much more popular file format .tif. A library called pillow, a fork of the Python Imaging Library, contains a ready-made solution for extracting the individual frames for the .tif format.

After accessing the frames of the TIRF-M video, the question of how to locate and identify the blinking QDs needs to be addressed. This problem poses a major challenge as the QDs may or may not be in the fluorescent on-state in any given frame. Figure 5.2 **s**hows the

difference in the first and last frames of a video of blinking QDs. The QDs that are on in the last frame are different from those in the first frame because of blinking and possibly, over a long period of time, photobleaching. One potential strategy to identify the QDs would be to identify every QD in every frame and collect their locations. At the end of the movie, these locations could be compared and spots that are identified within some local region can be classified as the same QD. In addition to be computationally expensive, this would introduce problems differentiating unique QDs from frame to frame and two QDs that share some local region. The solution introduced here is to create a composite image from the frames in the movie based on some criteria. This composite image would then be searched for the QDs. Two composite image types that have proven to be useful in the search for QDs, the max intensity image, Figure 5.3, and the mean intensity image, shown in Figure 5.4.

The mean intensity image constructs the image by taking the average value for each pixel throughout the course of the movie and representing the proper pixel location with that mean. Alternatively, the max composite image represents each pixel as the maximum intensity that each pixel obtained throughout the whole movie. The use of either of these search images introduces a clear bias in the selection of quantum dots. The mean search image rewards QDs that were both on for a long period of time and are bright. The max search image soley recognizes QDs that are bright for any moment, regardless of how fluorescent it is throughout the rest of the movie. With proper QD density and low background signal, the max search image can identify all QDs that were ever on in the region of interest. This strategy is ideal for identifying all QDs and saving them for further analysis.

One of the main problems with the production of the search image arises from the technical aspects of an objective-based TIRF experiment, uneven illumination.[152] Since the

center of an objective based TIRF experiment will always be brighter than the edges, this creates a bias towards selecting only QDs near the center of the video. Comparison of brighter QDs in the center of the laser excitation to the background is possible, but comparing darker QDs on the fringes to this same background is not favorable. It is difficult to compare the darker QDs when pixels of the background in the center on the same image may be brighter than the fringe QD pixels. The naive option in this case would be to compare every pixel to its local background level when determining its potential to be a QD. This algorithm would be exhaustive and be very expensive computationally. To combat this problem, a way to subtract an uneven background from the search image would be ideal as it would directly eliminate this problem. One promising method for this subtraction that became obvious was the white tophat transform. The tophat transform is a morphological image manipulation that subtracts local background from an image, but leaves small features. The size of the features left after subtraction depend on the structuring element size chosen. This can be illustrated in Figure 5.5. Based on this figure, the optimal structuring element size is one that is larger than the particles of interest, but as small as possible to minimize the background.

Fortunately, the background is much larger and more diffuse than the QDs. The size of the QDs in the video do not depend on the QD size, but rather the point-spread-function based on the diffraction limit. Since the fluorescence of the QDs is of known size and the fluorescent intensity of the QDs is much greater than that of the background, a structuring element of a certain pixel size can be chosen. The difference in the before and after results of the background subtraction are shown in Figure 5.6.

In Figure 5.6, it is clear that without the background subtraction, the center of the image is overrun with background signal. This colormap of the before and after images is set to

highlight this fact. Although the images are at different scales, no scale would have the “after” image displaying an increased background in the center relative to the edges.

Next, the actual search algorithm can be divided into two parts. First, a user selected threshold is applied to the search image. Every pixel below this threshold is set to zero and the search image is updated to reflect this. Every pixel that is non-zero, “bright” pixels, are assumed to come from QD fluorescence. Since the fluorescence of QDs spans a range of pixels according to the point-spread-function of the fluorophore, it is important to identify the center of each QD. To do this, all of the bright pixels are checked to see if they are the brightest in their local area. Once a pixel is determined to be the brightest in its local area, that pixel location is then tagged as the center of a QD. Figure 5.7 shows all of the centers that are identified using this method. The identification of the bright spots is well represented by the algorithmically tagged QD centers, shown in red. It is apparent that there are some artifacts in two vertical lines near the center. It is belived that these bright pixels arise from some problems associated with the optics. To remove these artifacts, a new criteria can be applied that limits QDs in the same row or column of the search image to a maximum number. If this number is exceeded, the whole column or row is excluded. The result of applying this algorithm to Figure 5.7 is shown in Figure 5.8.

Once the user is satisfied with the threshold selection and QD identification as shown in Figure 5.8, the pixel of interest and all adjacent pixels for every QD are extracted from the video. The extraction area for a single QD is shown in Figure 5.9. Although not all pixels are extracted, this is on purpose to ensure the average pixel intensity is much higher than the background. The intensity values and location of the QD are then collected and used in the creation of a Trace object. This object represents a QD and further properties will be calculated and attached to it.

Creating an object for each QD means that the manipulation of individual QDs becomes much simpler programmatically. This is especially useful for subsequent sorting of QDs by their individual fluorescent properties.

For every individual trace that is extracted, a threshold that differentiates between the on state and off state must be determined. One way to do this is to use Otsu's method. Otsu's method assumes a bimodal distribution and divides the fluorescent intensity histogram into an on and off state with a threshold that maximizes the weighted sum of the variances between the two distributions. Two differently blinking traces and the Otsu determination of their thresholds is shown in Figures 5.10-5.13.From Figures 5.10-5.13, it is clear that Otsu's method is able to set an accurate, automatic threshold despite vast differences in blinking statistics and populations of each state amongst QDs. Alternatively, the user may select a custom threshold that qualitatively fits the QD trace better. Once the threshold is set, lengths of on and off times and the frequencies of the transitions can be calculated. Also, the average intensity of the QD in the on state can be found. Once all of the desired statistics are calculated for each QD in a sample, statistics about the sample population can be compiled. For instance, the average brightness of a sample of QDs in the on state can be determined. Also, with many blinking events, the length a QD is in the on or off state can be collected and expressed as a probability distribution. This distribution for the on and off states are key blinking statistics called P_{on} and P_{off}. P_{on} and P_{off} for two synthesis methods involving differing amounts of zinc are shown in Figures 5.14 and 5.15 respectively.

Figures 5.14 and 5.15 show the program's power to do complicated analysis very quickly and routinely. Based on the figures, it is clear that the sample with more zinc is more likely to stay in the on state, while the sample with less zinc is more likely to go into and stay in the off state.

One of the main benefits of this work is the ability to not only compare the populations of two different synthesis methods, but to compare intrasample heterogeneity. Since QDs are not molecules and their specific optical properties depend on unique surface properties, there is variation between each QD. This can lead to the formation of sub-populations in a sample if multiple formation paths of the QDs are likely during the reaction. This can allow for the blinking statistics between two samples to be similar, but some fraction of the QDs possess remarkably different properties. Recognizing these intrasample sub-populations is only possible in single molecule experiments. By recognizing these subpopulations, it may be possible to tune a synthesis method to maximize a subpopulation with desired properties. Figures 5.16 and 5.17 show that both samples, less and more Zn have a large population of QDs that spend only a small fraction of time in the on state. However, there is some small subpopulation the more Zn sample that is relatively non-blinking. It may be possible upon identification of these subpopulations to tune the synthesis and examine the data for a change in this subpopulation.

Finally, a graphical user interface (GUI) to make the program as accessible as possible was built using the QT framework. This work was done using the Python ported library called PyQT. This step is crucially important for the uptake of this software.

Recent work has seen the extension of the QD analysis program via the addition of changepoint analysis. Changepoint analysis is a collection of methods that allow for the detection of more than two states and the changes between them. Since changepoint analysis is a general method and does not have to apply to QDs, this, coupled with the trace extraction method allows the program to garner useful information from a variety of fluorescence traces. The main contribution of this portion program is the detailed labeling of the states in between the changepoints in a fluorescence trace.

For the standard QD analysis as described above, Otsu's method is used to distinguish between the on and the off states of QDs. However, this method only works for QDs in which the intensity trace is bimodal. For certain QD preparations, multiple on states have been reported. Not only does this increase the complexity of the analysis, it can also provide more mechanistic information. Transitions between states are not likely to have the same probability. Changepoint analysis can be used to find the transition points in a QD blinking trace.[153] Figure 5.18 and 5.19 show the application of changepoint analysis and state labeling on a simulated QD trace and a known grey-state trace from Gao et al.[154] With this automated labeling of the states, transition probabilities can be calculated across a sample. Figures 5.20 shows two different transition heat maps for trace populations with different statistics. The first heatmap is a simulated QD sample with equal transition probabilities between off, gray, and on states. For the sample corresponding to the second sample, the probability of transitioning from gray to on was reduced one-hundred times.

The simulated samples correspond to one 5 minute widefield movies. This small sample size shows the power of this type of analysis.

Moving past QDs, the methods described above can be applied to different types of fluorescent time traces. Figure 5.21 shows the stepwise photobleaching of multiple cyanine3 dyes conjugated to a streptavidin protein. By detecting the changepoints, the states were labeled. Subsequent state criteria-based searching was performed to identify the number of dyes attached to each streptavidin molecule. The specific search criteria involved molecules that had the appropriate number of subsequent downward transitions. By collecting population statistics for streptavidin labeled with two different dyes, Cy3 and Alexa488, the dyes per protein population statistics were found at the single molecule level. This is illustrated in Figure 5.22. Although

Figure 5.22 does not give any information about the unlabeled portion of streptavidin molecules, it does give valuable insight into the relative percentages of signal-producing multi-dye systems. By analyzing the single-bleach molecules, the bleaching decay rates of a dye can be determined in a similar fashion to Figure 5.1. Figure 5.23 an analysis of the bleaching decay curves of Alexa488 and Cy3 under different power and buffer conditions.

5.4 Conclusions

Here I have demonstrated a new software to streamline the process of analyzing blinking data from QDs immobilized on a surface. Further, this software was able to identify previously unobserved properties within a population of QDs from a particular synthesis method. Hopefully the adoption of this software within our lab will allow for the routine analysis of blinking for every synthesized sample. Further, the addition of changepoint analysis extends the utility of this program to the analysis of other fluorescent time traces. Future improvements to the software should involve the introduction statistical clustering of states, super-resolution mapping, and single particle tracking.

5.5 Figures

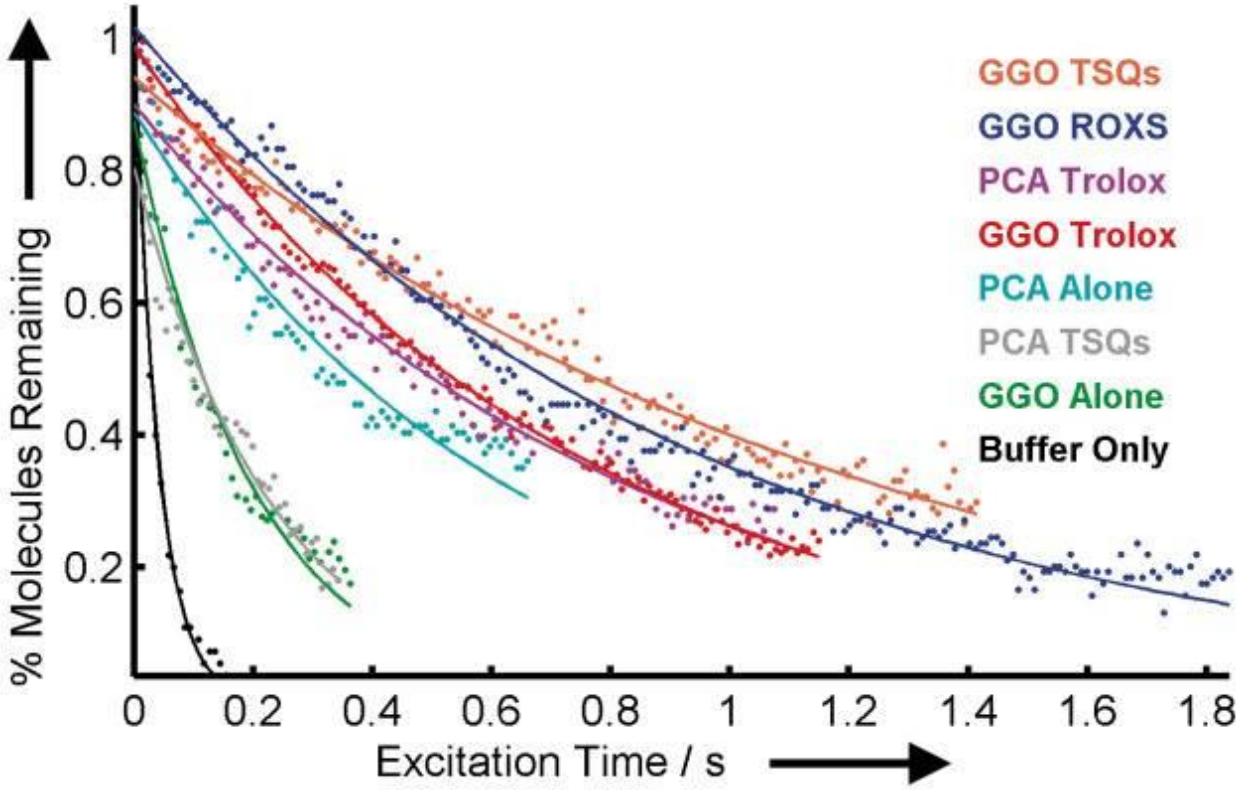


Figure 5.1 From Cooper et. Al, 2013[144] Shows the percentage of cyanine dyes still fluorescing after a certain amount of time for multiple lifespan extending solutions.

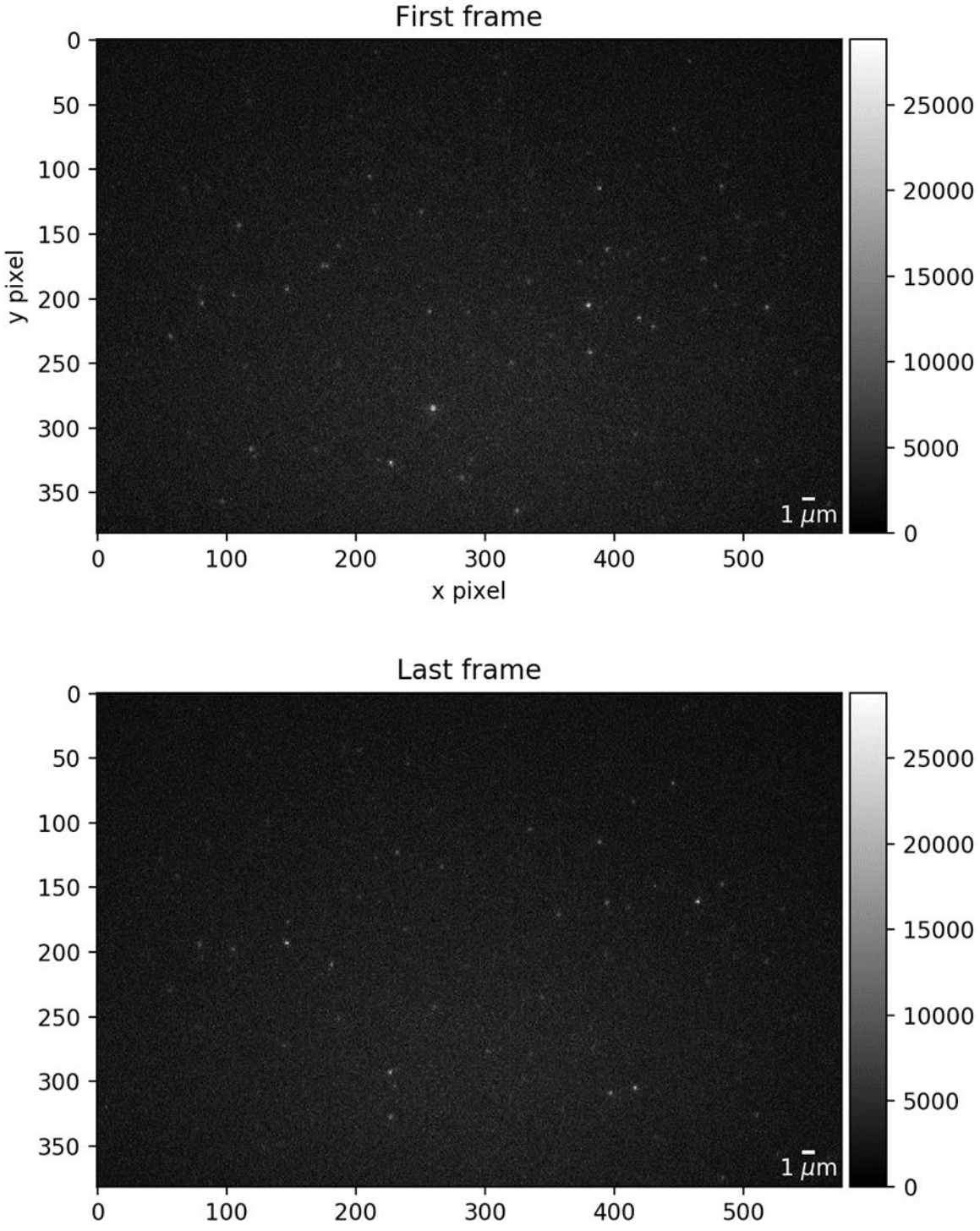


Figure 5.2 shows that the QDs that are fluorescing in the first frame of a video are not necessarily going to be the QDs that are on in the last frame of the video.

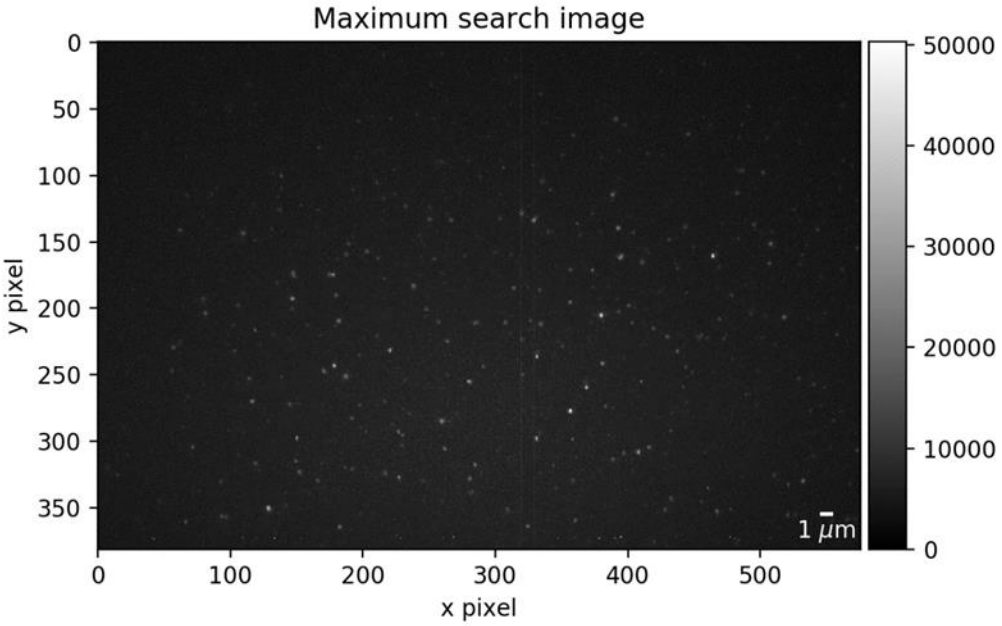


Figure 5.3 The maximum intensity image is useful for identifying all QDs in a sample that are on at any given time in the movie. Each pixel is the maximum brightness of that pixel throughout the entire movie.

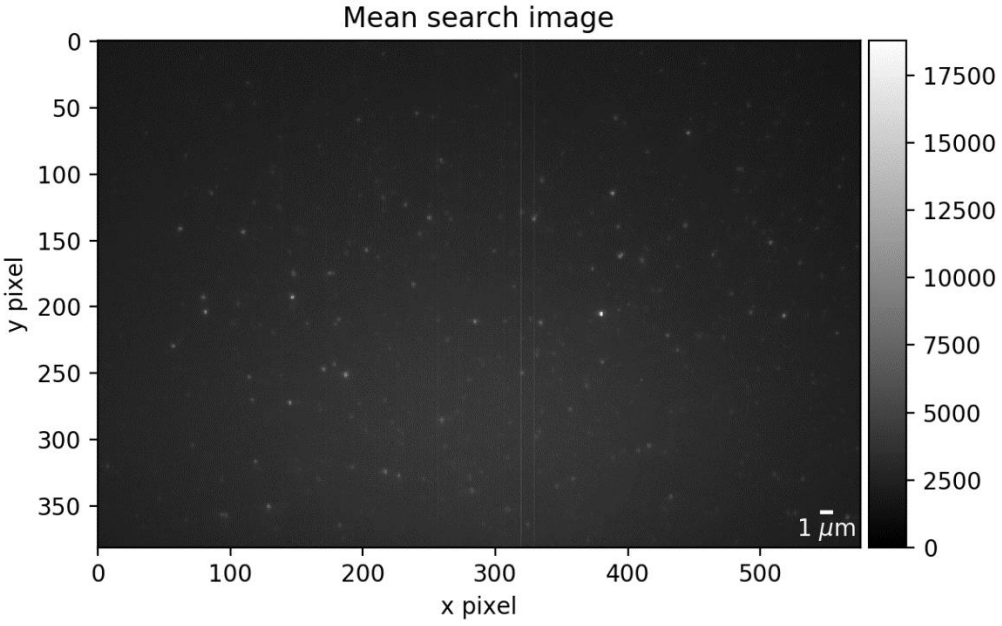


Figure 5.4 The mean intensity image is useful for identifying QDs that have a significant combination of brightness and on times. It is composed of the average value of each pixel throughout the course of the movie.

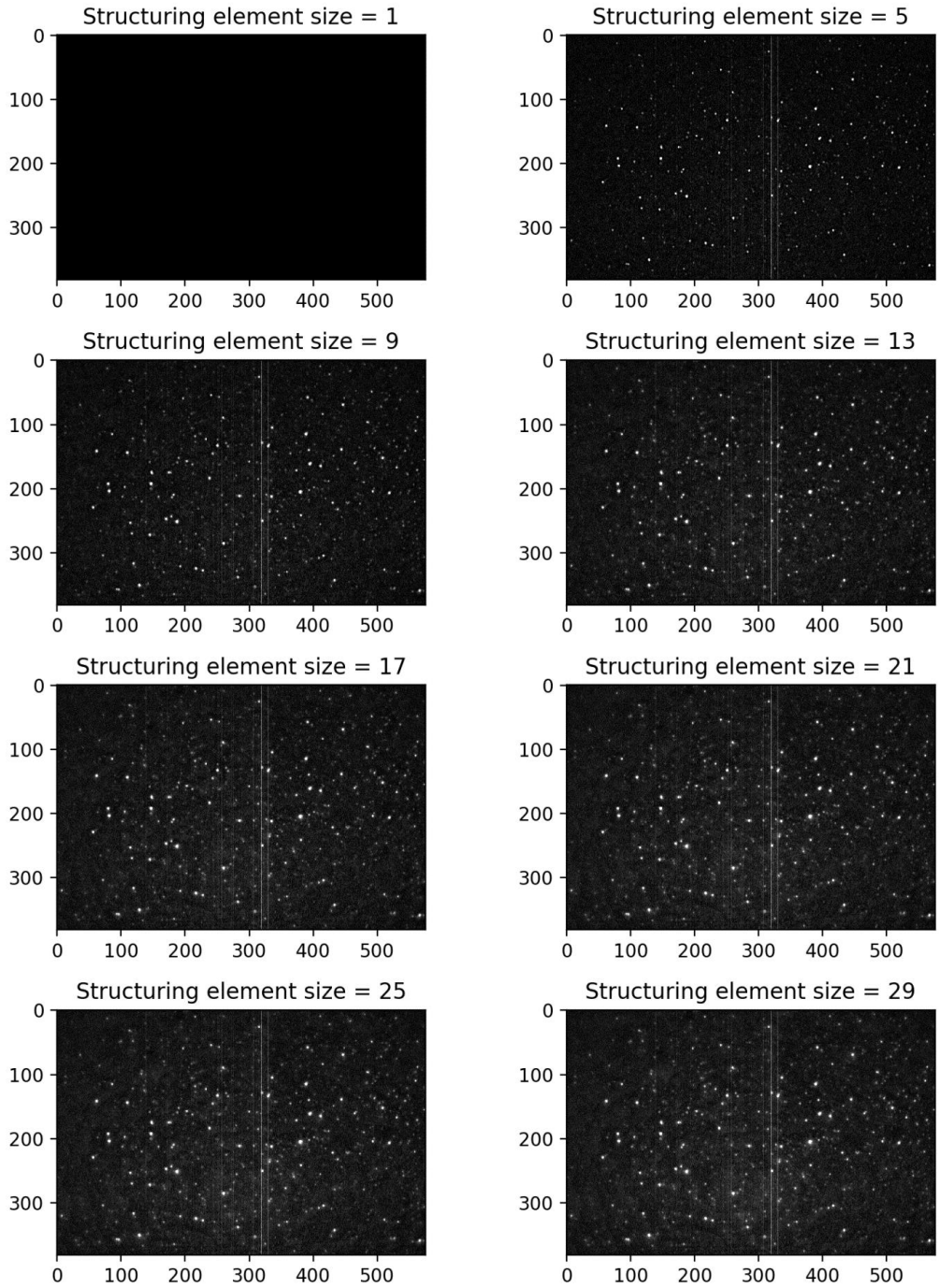


Figure 5.5 shows the white tophat transform as a function of the size of the structuring element.

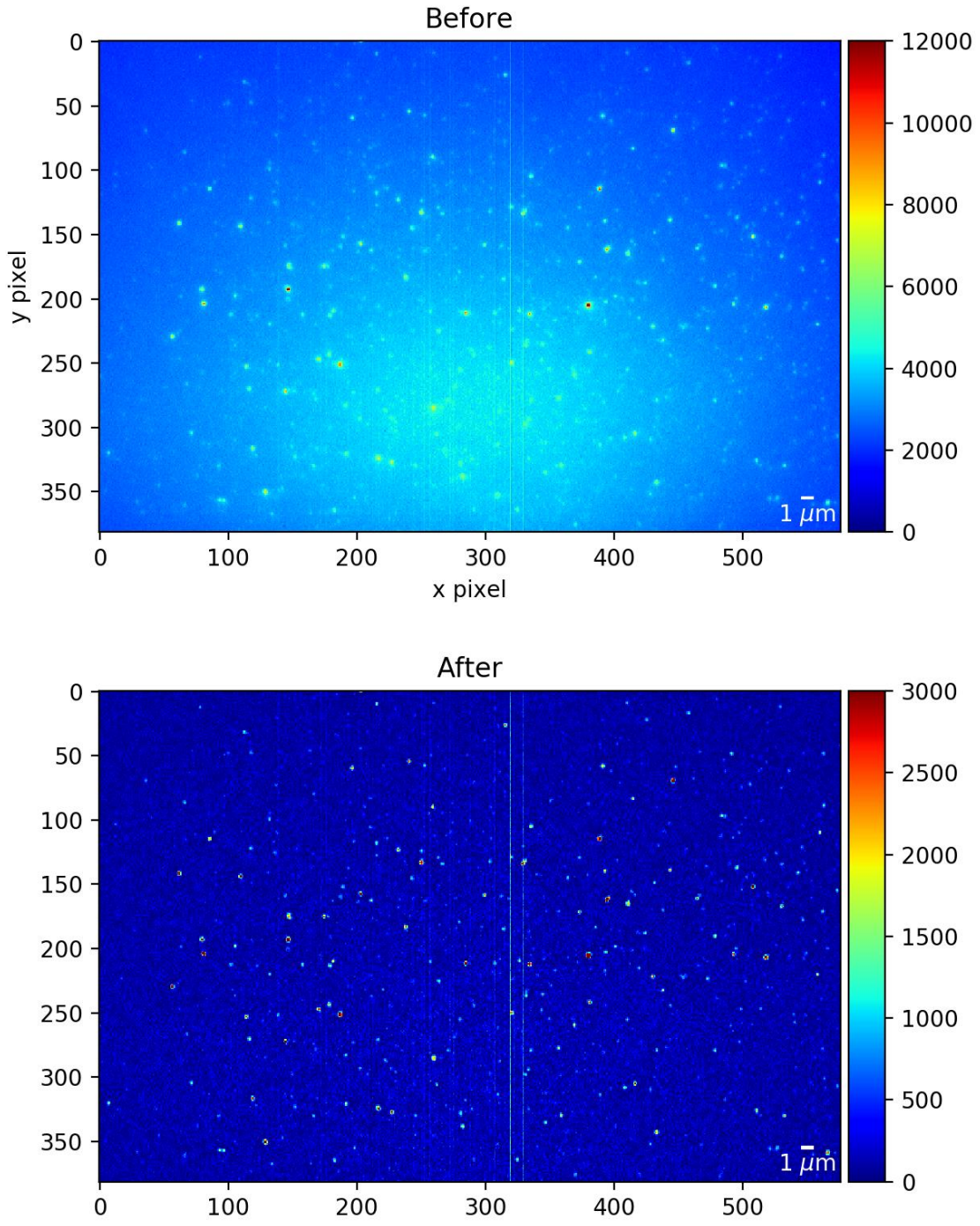


Figure 5.6 shows the before and after images of the background subtraction via the white tophat transform.

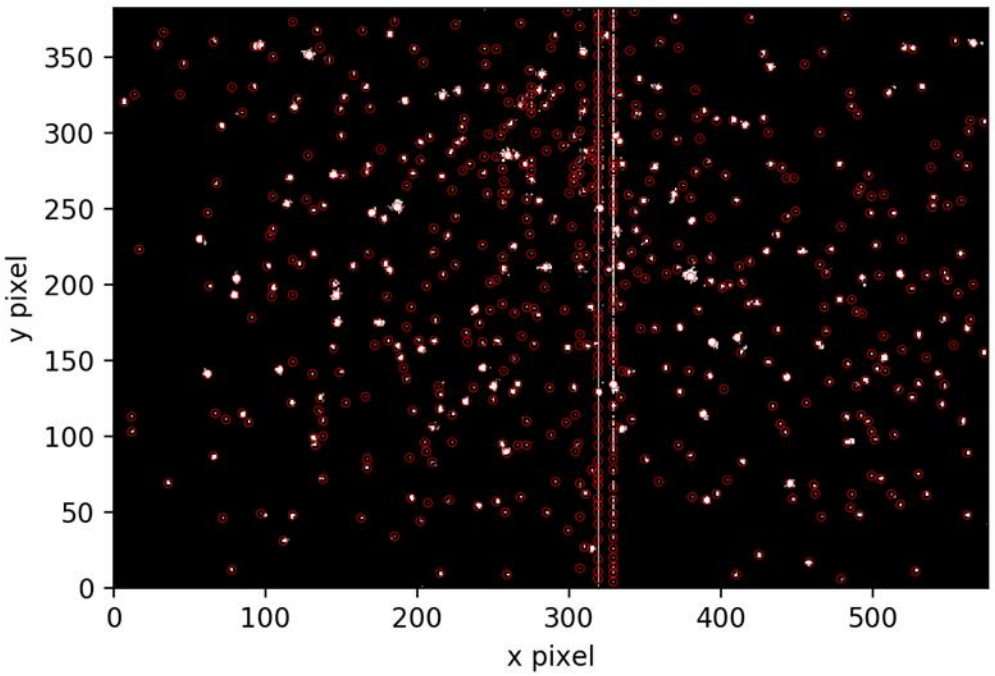


Figure 5.7 shows all of the potential QDs identified as the local brightness centers above the user chosen threshold.

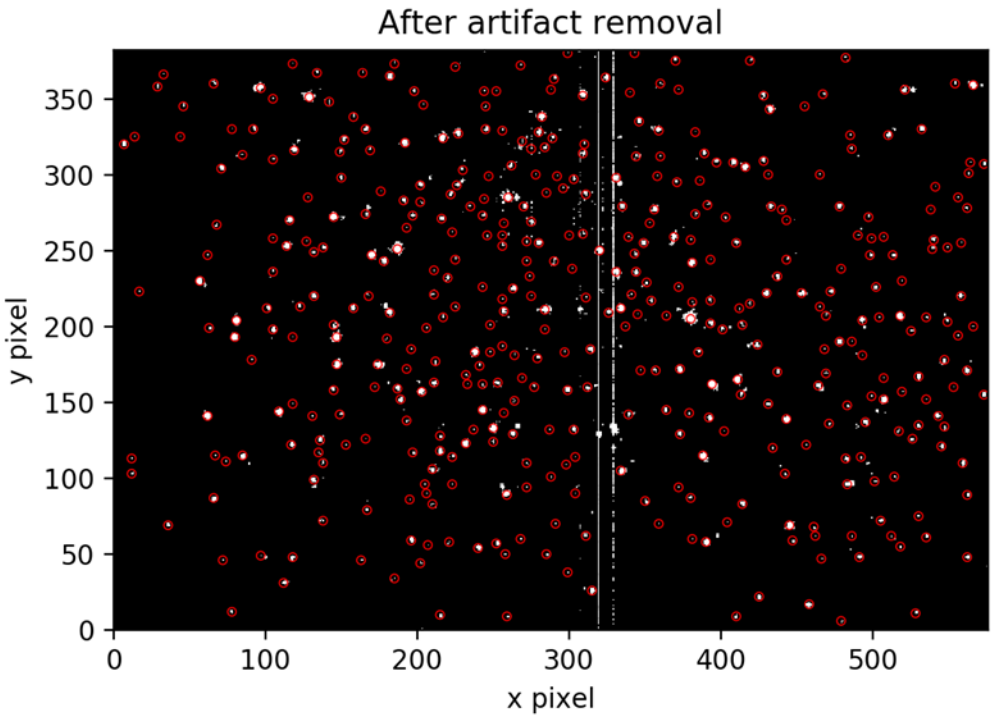


Figure 5.8 shows the search image after application of the criteria that limits the QDs in a row or column.

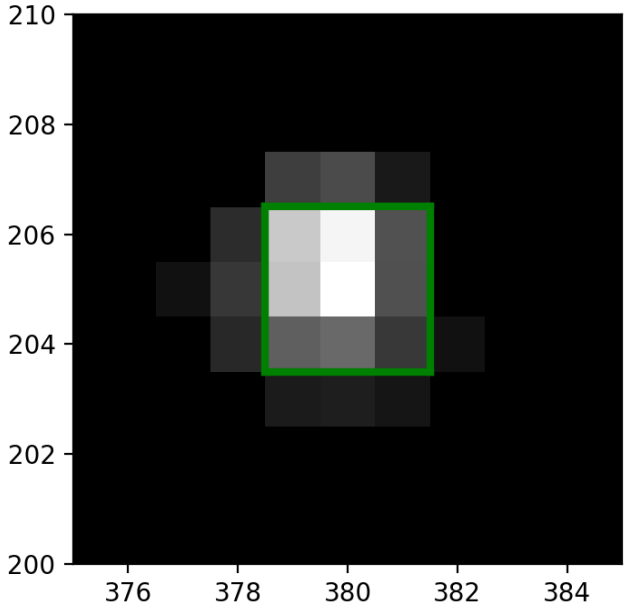


Figure 5.9 shows the chosen pixels of an individual QD to be extracted into a Trace object.

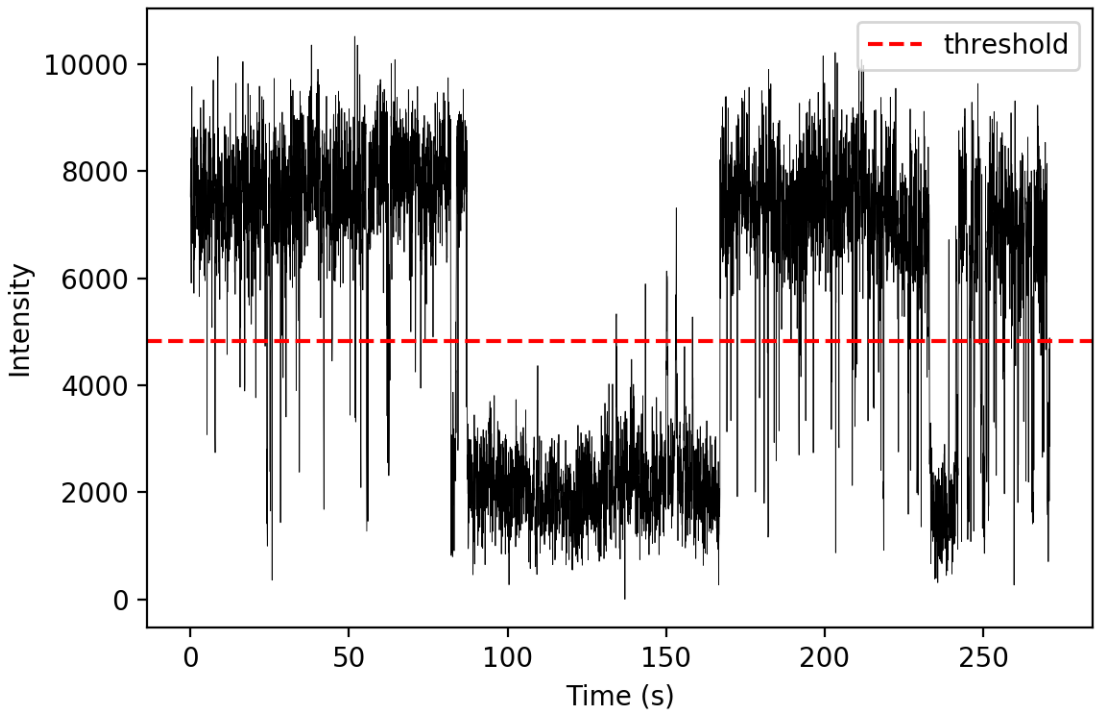


Figure 5.10 shows a trace with well defined on and off states with approximately equal time spent in each.

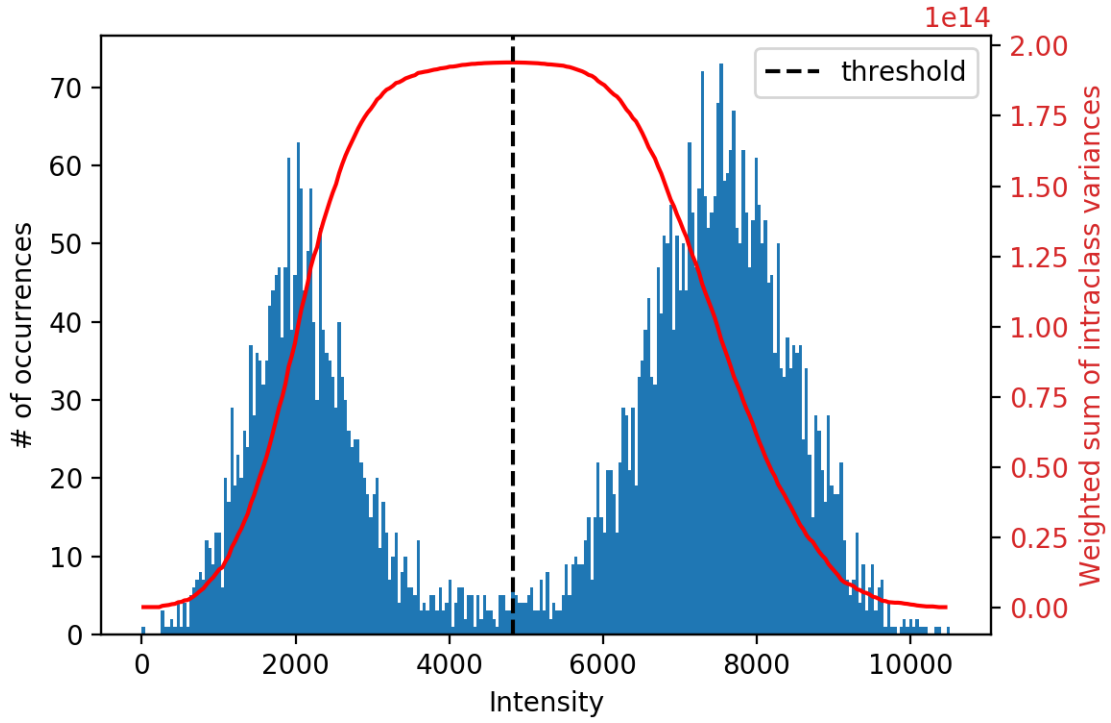


Figure 5.11 shows the Otsu method of calculating the optimal threshold to distinguish between signal and background.

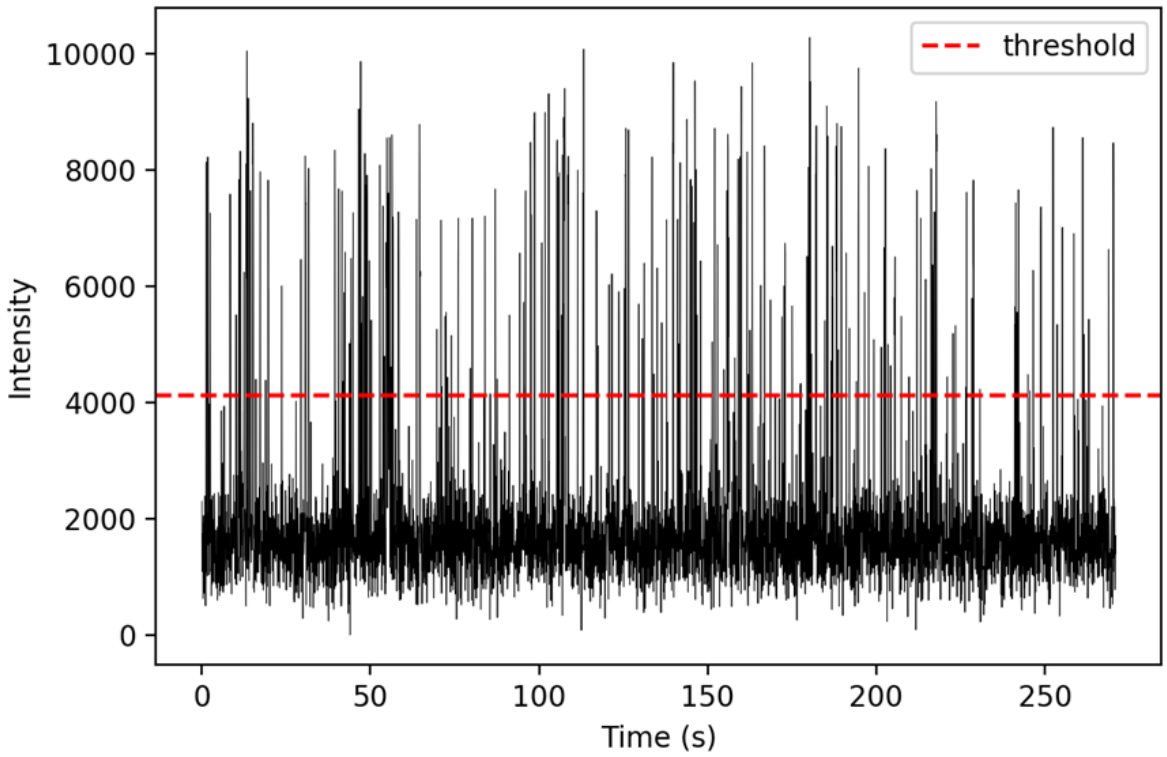


Figure 5.12 shows a rapidly blinking QD that spends the vast majority of its time in the off state.

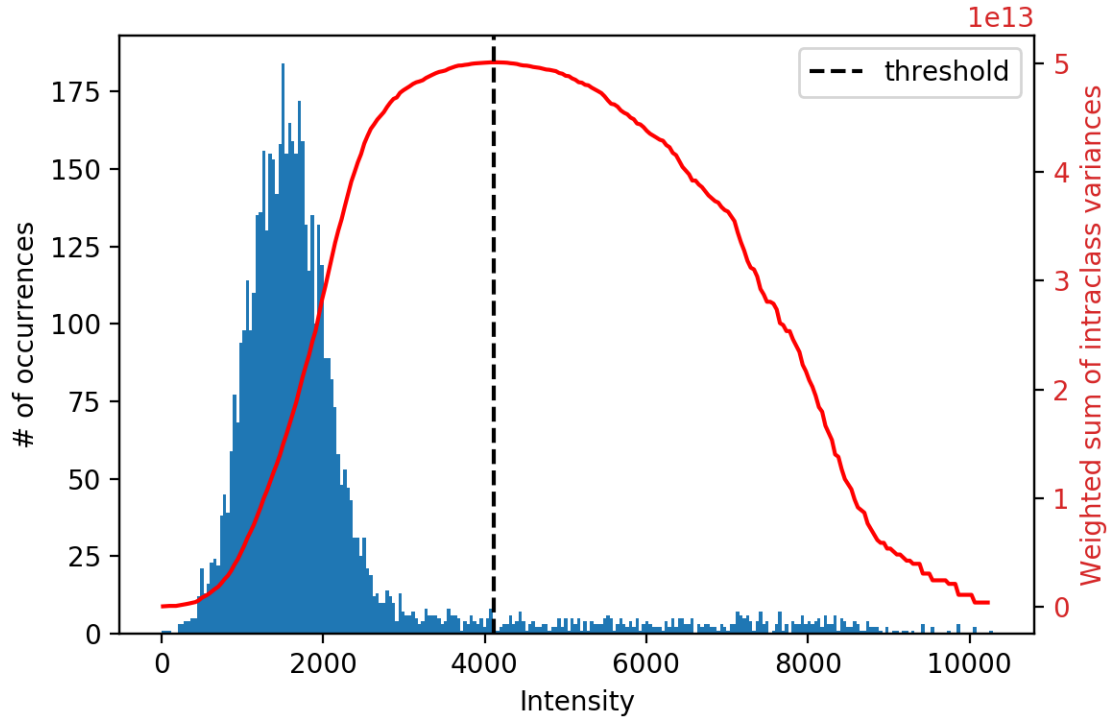


Figure 5.13 shows the Otsu method of calculating the optimal threshold to distinguish between signal and background when the two states are very imbalanced.

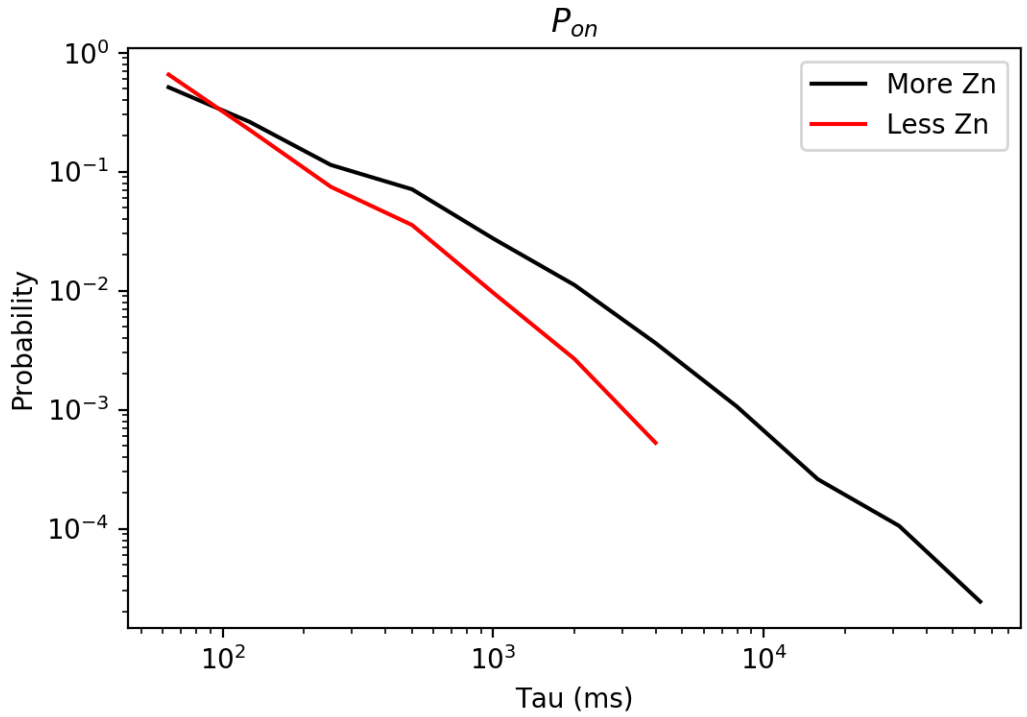


Figure 5.14 shows a log-log plot of the probability of an on state lasting a certain amount of time, tau for two different samples.

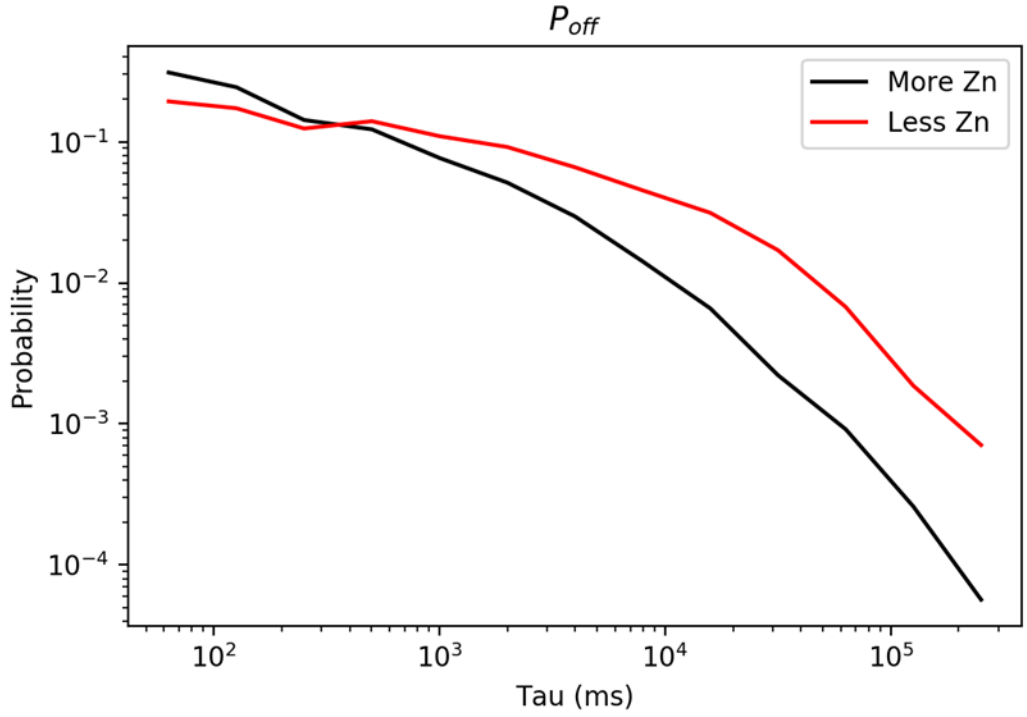


Figure 5.15 shows a log-log plot of the probability of an off state lasting a certain amount of time, tau for two different samples.

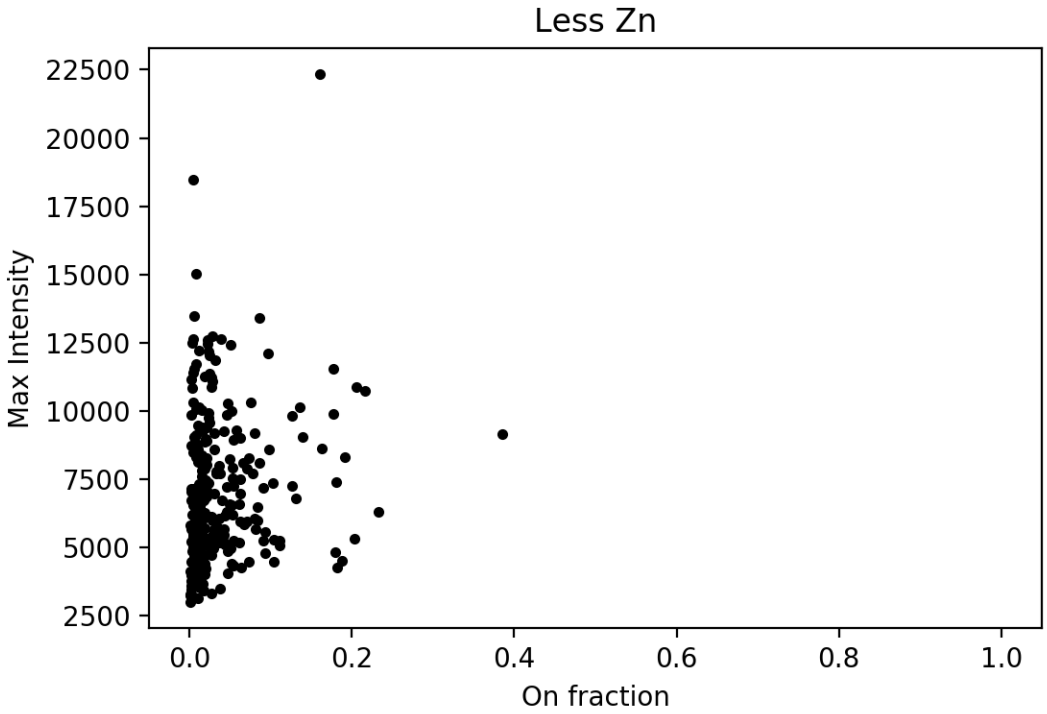


Figure 5.16 shows that the sample prepared with less Zn has a tight distribution of QDs that have a low on fraction.

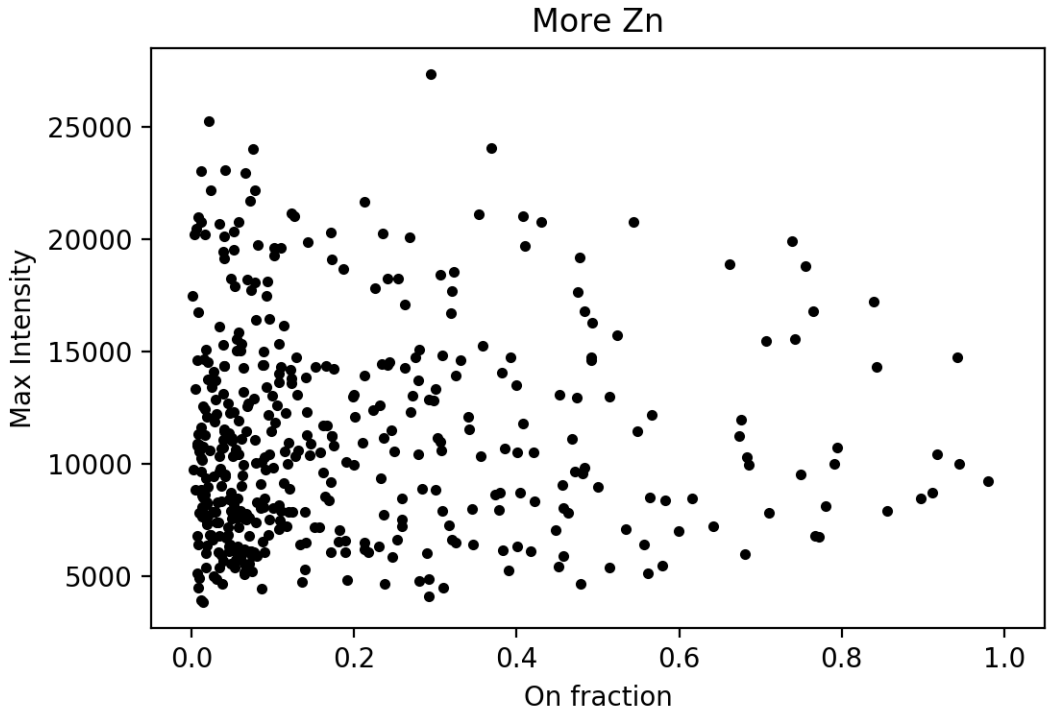


Figure 5.17 shows a dense subpopulation of mostly off QDs and a diffuse distribution of some relatively non-blinking QDs.

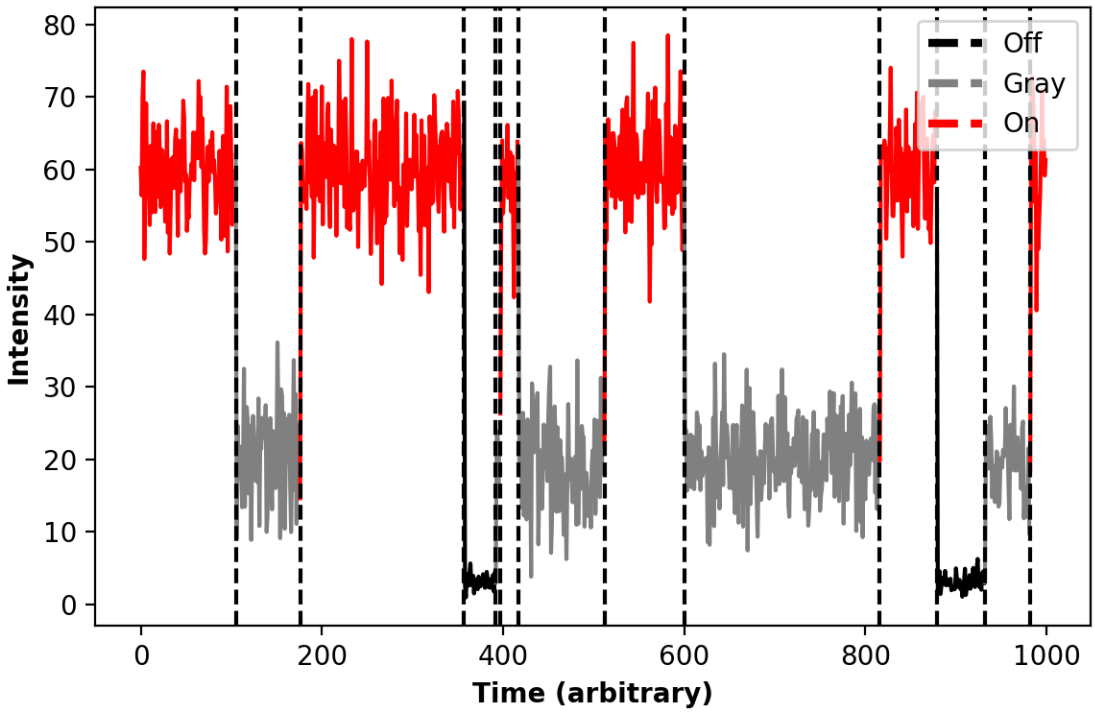


Figure 5.18 shows a simulated fluorescence trace automatically labeled with states using changepoint analysis.

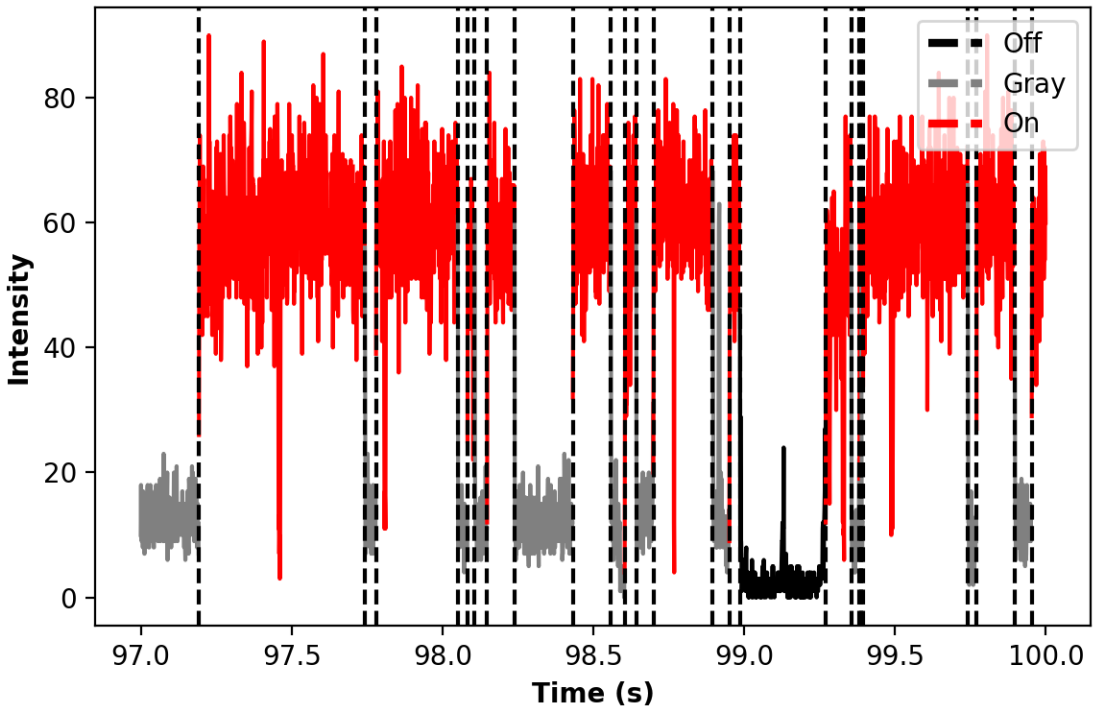


Figure 5.19 shows state identification using changepoint analysis on a published, three-state QD blinking trace from Gao et al.[154]

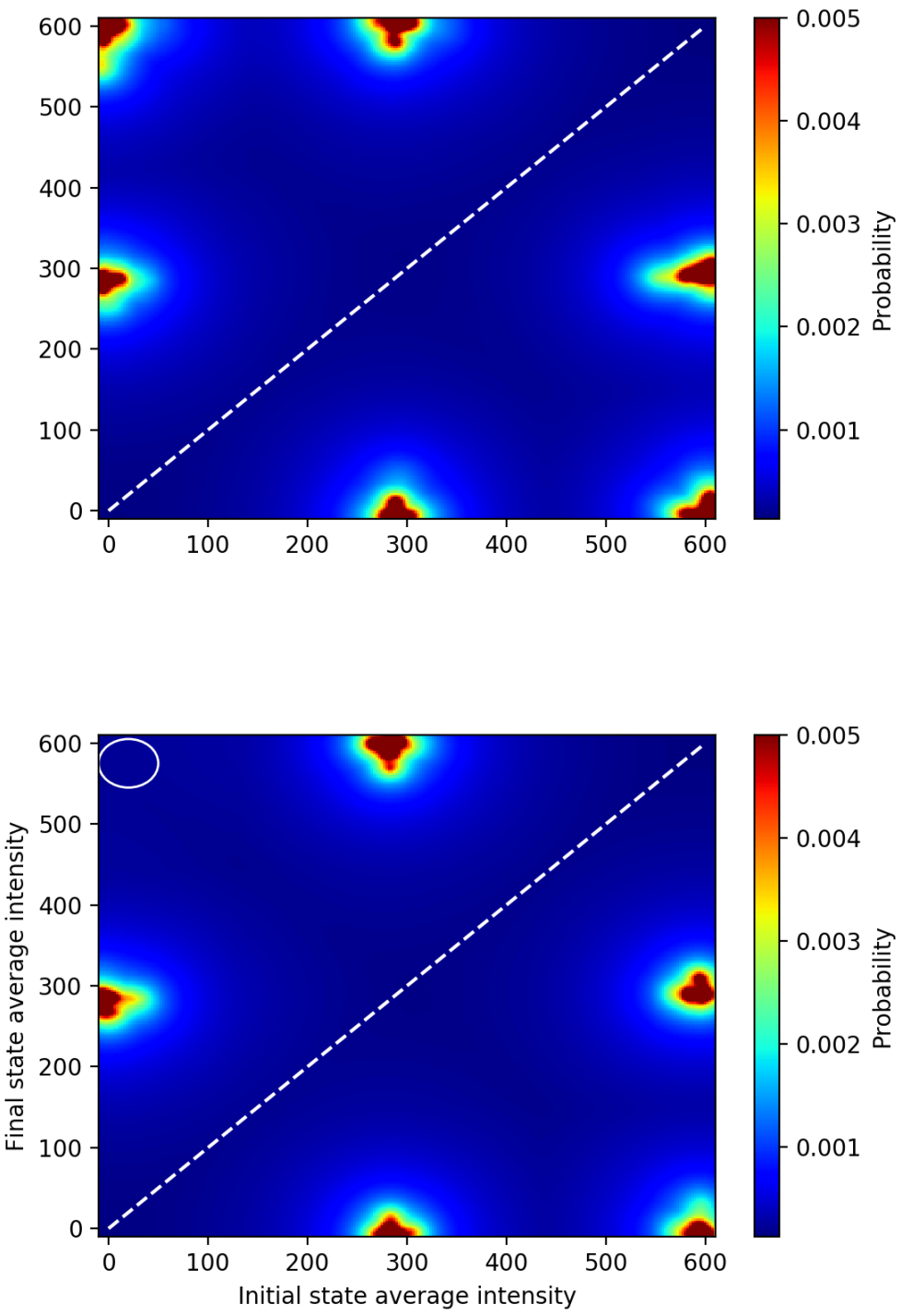


Figure 5.20 a) Shows the transition probability for simulated traces with symmetric transition probabilities. b) Shows the transition probability for simulated traces where the transition from the off state to the on state has been made significantly less probable.

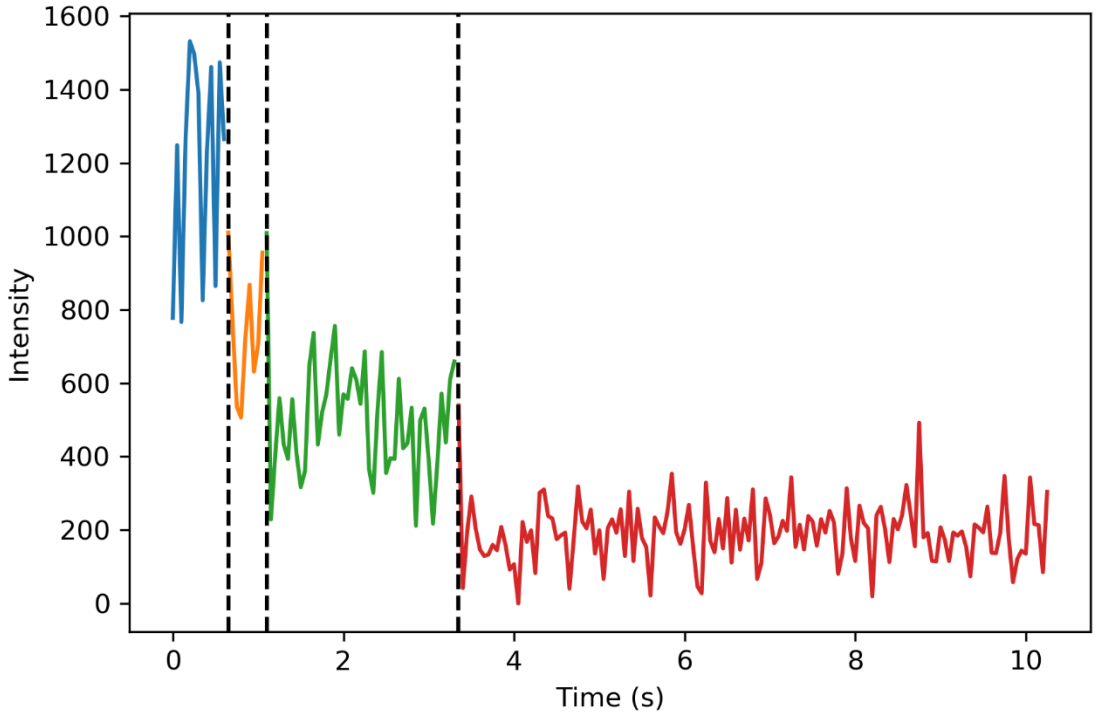


Figure 5.21 shows the stepwise photobleaching of three cyanine3 dyes.

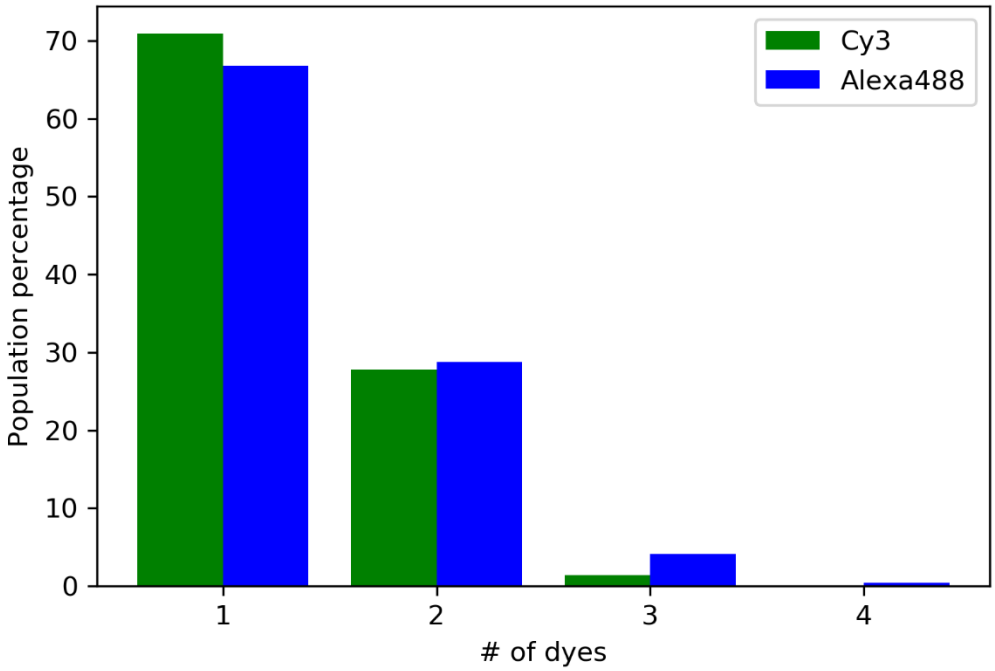


Figure 5.22 shows the distribution of dye labeling for streptavidin using Cy3 or Alexa488.

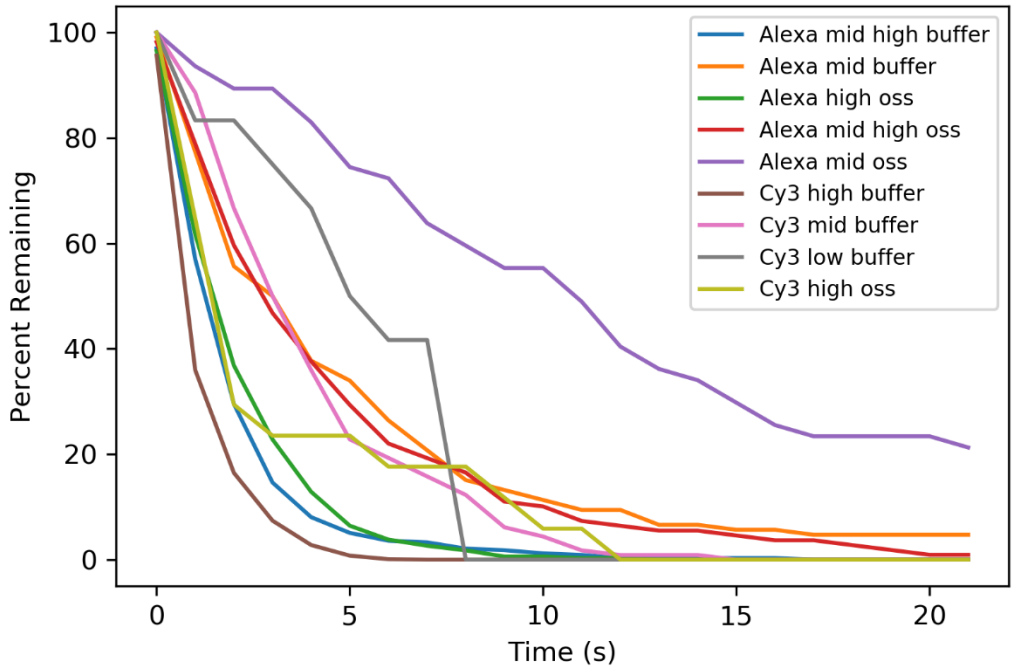


Figure 5.23 shows the photobleaching decay curve for dyes under various powers and photoprotective situations.

Conclusions

In this work, I have presented supporting evidence for a molecular model of the cpSRP and transit complexes, evidence for a previously unidentified secondary structure motif in the intrinsically disordered C-terminus of the Albino3 protein, shown evidence for the formation of secondary structure of the C-terminus of the Albino3 protein upon binding cpSRP43, and developed software that will lay the foundation of time-resolved single molecule experiments in the future.

Acknowledgments

I would like to thank my committee. I would like to thank the cpSRP collaboration. Many members contributed greatly to this work. I would like to thank my advisor, Colin Heyes, for many things. First, his research knowledge and guidance were invaluable to me over the years. Second, I would like to thank him for allowing me to develop at my own pace. And finally, I would like to thank him for his kind demeanor, from which I take great inspiration. I would like to thank all of my friends who have helped me along the way. All of my teachers that helped get me here.

References

1. Pohlschröder, M.; Prinz, W. A.; Hartmann, E.; Beckwith, J., Protein translocation in the three domains of life: variations on a theme. *Cell* **1997,** *91* (5), 563-6.

2. Walter, P.; Blobel, G., Purification of a membrane-associated protein complex required for protein translocation across the endoplasmic reticulum. *Proc Natl Acad Sci U S A* **1980,** *77* (12), 7112-6.

3. Walter, P.; Blobel, G., Signal recognition particle contains a 7S RNA essential for protein translocation across the endoplasmic reticulum. *Nature* **1982,** *299* (5885), 691-8.

4. 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.

5. Walter, P.; Ibrahimi, I.; Blobel, G., Translocation of proteins across the endoplasmic reticulum. I. Signal recognition protein (SRP) binds to in-vitro-assembled polysomes synthesizing secretory protein. *J Cell Biol* **1981,** *91* (2 Pt 1), 545-50.

6. Kurzchalia, T. V.; Wiedmann, M.; Girshovich, A. S.; Bochkareva, E. S.; Bielka, H.; Rapoport, T. A., The signal sequence of nascent preprolactin interacts with the 54K polypeptide of the signal recognition particle. *Nature* **1986,** *320* (6063), 634-6.

7. Siegel, V.; Walter, P., Elongation arrest is not a prerequisite for secretory protein translocation across the microsomal membrane. *The Journal of cell biology* **1985,** *100* (6), 1913-1921.

8. Walter, P.; Blobel, G., Disassembly and reconstitution of signal recognition particle. *Cell* **1983,** *34* (2), 525-33.

9. Gilmore, R.; Blobel, G.; Walter, P., Protein translocation across the endoplasmic reticulum. I. Detection in the microsomal membrane of a receptor for the signal recognition particle. *J Cell Biol* **1982,** *95* (2 Pt 1), 463-9.

10. Gilmore, R.; Walter, P.; Blobel, G., Protein translocation across the endoplasmic reticulum. II. Isolation and characterization of the signal recognition particle receptor. *J Cell Biol* **1982,** *95* (2 Pt 1), 470-7.

11. Görlich, D.; Prehn, S.; Hartmann, E.; Kalies, K. U.; Rapoport, T. A., A mammalian homolog of SEC61p and SECYp is associated with ribosomes and nascent polypeptides during translocation. *Cell* **1992,** *71* (3), 489-503.

12. Rapiejko, P. J.; 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.

13. Hanein, D.; Matlack, K. E.; Jungnickel, B.; Plath, K.; Kalies, K. U.; Miller, K. R.; Rapoport, T. A.; Akey, C. W., Oligomeric rings of the Sec61p complex induced by ligands required for protein translocation. *Cell* **1996,** *87* (4), 721-32.

14. Phillips, G. J.; Silhavy, T. J., The E. coli ffh gene is necessary for viability and efficient protein export. *Nature* **1992,** *359* (6397), 744-6.

15. Poritz, M. A.; Bernstein, H. D.; Strub, K.; Zopf, D.; Wilhelm, H.; Walter, P., An E. coli ribonucleoprotein containing 4.5S RNA resembles mammalian signal recognition particle. *Science* **1990,** *250* (4984), 1111-7.

16. Luirink, J.; Sinning, I., SRP-mediated protein targeting: structure and function revisited. *Biochim Biophys Acta* **2004,** *1694* (1-3), 17-35.

17. Seluanov, A.; Bibi, E., FtsY, the prokaryotic signal recognition particle receptor homologue, is essential for biogenesis of membrane proteins. *J Biol Chem* **1997,** *272* (4), 2053-5.

18. Miller, J. D.; Bernstein, H. D.; Walter, P., Interaction of E. coli Ffh/4.5S ribonucleoprotein and FtsY mimics that of mammalian signal recognition particle and its receptor. *Nature* **1994,** *367* (6464), 657-9.

19. de Leeuw, E.; Poland, D.; Mol, O.; Sinning, I.; ten Hagen-Jongman, C. M.; Oudega, B.; Luirink, J., Membrane association of FtsY, the E. coli SRP receptor. *FEBS Lett* **1997,** *416* (3), 225-9.

20. Angelini, S.; Deitermann, S.; Koch, H. G., FtsY, the bacterial signal-recognition particle receptor, interacts functionally and physically with the SecYEG translocon. *EMBO Rep* **2005,** *6* (5), 476-81.

21. Akimaru, J.; Matsuyama, S.; Tokuda, H.; Mizushima, S., Reconstitution of a protein translocation system containing purified SecY, SecE, and SecA from Escherichia coli. *Proc Natl Acad Sci U S A* **1991,** *88* (15), 6545-9.

22. Bhuiyan, S. H.; Gowda, K.; Hotokezaka, H.; Zwieb, C., Assembly of archaeal signal recognition particle from recombinant components. *Nucleic acids research* **2000,** *28* (6), 1365-1373.

23. Samuelson, J. C.; Chen, M.; Jiang, F.; Möller, I.; Wiedmann, M.; Kuhn, A.; Phillips, G. J.; Dalbey, R. E., YidC mediates membrane protein insertion in bacteria. *Nature* **2000,** *406* (6796), 637-41.

24. Serek, J.; Bauer-Manz, G.; Struhalla, G.; van den Berg, L.; Kiefer, D.; Dalbey, R.; Kuhn, A., Escherichia coli YidC is a membrane insertase for Sec-independent proteins. *EMBO J* **2004,** *23* (2), 294-301.

25. Scotti, P. A.; Urbanus, M. L.; Brunner, J.; de Gier, J. W.; von Heijne, G.; van der Does, C.; Driessen, A. J.; Oudega, B.; Luirink, J., YidC, the Escherichia coli homologue of mitochondrial Oxa1p, is a component of the Sec translocase. *EMBO J* **2000,** *19* (4), 542-9.

26. Samuelson, J. C.; Jiang, F.; Yi, L.; Chen, M.; de Gier, J. W.; Kuhn, A.; Dalbey, R. E., Function of YidC for the insertion of M13 procoat protein in Escherichia coli: translocation of mutants that show differences in their membrane potential dependence and Sec requirement. *J Biol Chem* **2001,** *276* (37), 34847-52.

27. Kohorn, B. D.; Tobin, E. M., Chloroplast Import of Light-Harvesting Chlorophyll a/b-Proteins with Different Amino Termini and Transit Peptides. *Plant Physiol* **1986,** *82* (4), 1172-4.

28. Henry, R. L., SRP: adapting to life in the chloroplast. *Nat Struct Mol Biol* **2010,** *17* (6), 676-7.

29. Franklin, A. E.; Hoffman, N. E., Characterization of a chloroplast homologue of the 54-kDa subunit of the signal recognition particle. *J Biol Chem* **1993,** *268* (29), 22175-80.

30. Schuenemann, D.; Gupta, S.; Persello-Cartieaux, F.; Klimyuk, V. I.; Jones, J. D.; Nussaume, L.; Hoffman, N. E., A novel signal recognition particle targets light-harvesting proteins to the thylakoid membranes. *Proc Natl Acad Sci U S A* **1998,** *95* (17), 10312-6.

31. Yuan, J.; Henry, R.; Cline, K., Stromal factor plays an essential role in protein integration into thylakoids that cannot be replaced by unfolding or by heat shock protein Hsp70. *Proc Natl Acad Sci U S A* **1993,** *90* (18), 8552-6.

32. Fulson, D. R.; Cline, K., A Soluble Protein Factor is Required in Vitro for Membrane Insertion of the Thylakoid Precursor Protein, pLHCP. *Plant Physiol* **1988,** *88* (4), 1146-53.

33. Payan, L. A.; Cline, K., A stromal protein factor maintains the solubility and insertion competence of an imported thylakoid membrane protein. *J Cell Biol* **1991,** *112* (4), 603-13.

34. Li, X.; Henry, R.; Yuan, J.; Cline, K.; Hoffman, N. E., A chloroplast homologue of the signal recognition particle subunit SRP54 is involved in the posttranslational integration of a protein into thylakoid membranes. *Proc Natl Acad Sci U S A* **1995,** *92* (9), 3789-93.

35. Pilgrim, M. L.; van Wijk, K. J.; Parry, D. H.; Sy, D. A.; Hoffman, N. E., Expression of a dominant negative form of cpSRP54 inhibits chloroplast biogenesis in Arabidopsis. *Plant J* **1998,** *13* (2), 177-86.

36. Klimyuk, V. I.; Persello-Cartieaux, F.; Havaux, M.; Contard-David, P.; Schuenemann, D.; Meiherhoff, K.; Gouet, P.; Jones, J. D.; Hoffman, N. E.; Nussaume, L., 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. *Plant Cell* **1999,** *11* (1), 87-99.

37. Amin, P.; Sy, D. A.; Pilgrim, M. L.; Parry, D. H.; Nussaume, L.; Hoffman, N. E., Arabidopsis mutants lacking the 43- and 54-kilodalton subunits of the chloroplast signal recognition particle have distinct phenotypes. *Plant Physiol* **1999,** *121* (1), 61-70.

38. Hutin, C.; Havaux, M.; Carde, J. P.; Kloppstech, K.; Meiherhoff, K.; Hoffman, N.; Nussaume, L., Double mutation cpSRP43--/cpSRP54-- is necessary to abolish the cpSRP

pathway required for thylakoid targeting of the light-harvesting chlorophyll proteins. *Plant J* **2002,** *29* (5), 531-43.

39. Kogata, N.; Nishio, K.; Hirohashi, T.; Kikuchi, S.; Nakai, M., Involvement of a chloroplast homologue of the signal recognition particle receptor protein, FtsY, in protein targeting to thylakoids. *FEBS Lett* **1999,** *447* (2-3), 329-33.

40. Tu, C. J.; Schuenemann, D.; Hoffman, N. E., Chloroplast FtsY, chloroplast signal recognition particle, and GTP are required to reconstitute the soluble phase of light-harvesting chlorophyll protein transport into thylakoid membranes. *J Biol Chem* **1999,** *274* (38), 27219-24.

41. Asakura, Y.; Kikuchi, S.; Nakai, M., Non-identical contributions of two membrane-bound cpSRP components, cpFtsY and Alb3, to thylakoid biogenesis. *Plant J* **2008,** *56* (6), 1007-17.

42. Tzvetkova-Chevolleau, T.; Hutin, C.; Noël, L. D.; Goforth, R.; Carde, J. P.; Caffarri, S.; Sinning, I.; Groves, M.; Teulon, J. M.; Hoffman, N. E.; Henry, R.; Havaux, M.; Nussaume, L., Canonical signal recognition particle components can be bypassed for posttranslational protein targeting in chloroplasts. *Plant Cell* **2007,** *19* (5), 1635-48.

43. Mori, H.; Summer, E. J.; Ma, X.; Cline, K., Component specificity for the thylakoidal Sec and Delta pH-dependent protein transport pathways. *J Cell Biol* **1999,** *146* (1), 45-56.

44. Moore, M.; Harrison, M. S.; Peterson, E. C.; Henry, R., Chloroplast Oxa1p homolog albino3 is required for post-translational integration of the light harvesting chlorophyll-binding protein into thylakoid membranes. *J Biol Chem* **2000,** *275* (3), 1529-32.

45. Goforth, R. L.; Peterson, E. C.; Yuan, J.; Moore, M. J.; Kight, A. D.; Lohse, M. B.; Sakon, J.; Henry, R. L., Regulation of the GTPase cycle in post-translational signal recognition particle-based protein targeting involves cpSRP43. *J Biol Chem* **2004,** *279* (41), 43077-84.

46. Jonas-Straube, E.; Hutin, C.; Hoffman, N. E.; Schünemann, D., Functional analysis of the protein-interacting domains of chloroplast SRP43. *J Biol Chem* **2001,** *276* (27), 24654-60.

47. Mosavi, L. K.; Cammett, T. J.; Desrosiers, D. C.; Peng, Z. Y., The ankyrin repeat as molecular architecture for protein recognition. *Protein Sci* **2004,** *13* (6), 1435-48.

48. Stengel, K. F.; 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.

49. Henderson, R. C.; Gao, F.; Jayanthi, S.; Kight, A.; Sharma, P.; Goforth, R. L.; Heyes, C. D.; Henry, R. L.; Suresh Kumar, T. K., Domain Organization in the 54-kDa Subunit of the Chloroplast Signal Recognition Particle. *Biophys J* **2016,** *111* (6), 1151-1162.

50. Marty, N. J.; Rajalingam, D.; Kight, A. D.; Lewis, N. E.; Fologea, D.; Kumar, T. K.; Henry, R. L.; Goforth, R. L., The membrane-binding motif of the chloroplast signal recognition particle receptor (cpFtsY) regulates GTPase activity. *J Biol Chem* **2009,** *284* (22), 14891-903.

51. Chandrasekar, S.; Chartron, J.; Jaru-Ampornpan, P.; Shan, S. O., Structure of the chloroplast signal recognition particle (SRP) receptor: domain arrangement modulates SRP-receptor interaction. *J Mol Biol* **2008,** *375* (2), 425-36.

52. Falk, S.; Ravaud, S.; Koch, J.; Sinning, I., The C terminus of the Alb3 membrane insertase recruits cpSRP43 to the thylakoid membrane. *J Biol Chem* **2010,** *285* (8), 5954-62.

53. Barros, T.; Kühlbrandt, W., Crystallisation, structure and function of plant light-harvesting Complex II. *Biochim Biophys Acta* **2009,** *1787* (6), 753-72.

54. Abad, M. S.; Oblong, J. E.; Lamppa, G. K., Soluble Chloroplast Enzyme Cleaves preLHCP Made in Escherichia coli to a Mature Form Lacking a Basic N-Terminal Domain. *Plant Physiol* **1991,** *96* (4), 1220-7.

55. Schuenemann, D.; Amin, P.; Hoffman, N. E., Functional divergence of the plastid and cytosolic forms of the 54-kDa subunit of signal recognition particle. *Biochem Biophys Res Commun* **1999,** *254* (1), 253-8.

56. Funke, S.; Knechten, T.; Ollesch, J.; Schünemann, D., A unique sequence motif in the 54-kDa subunit of the chloroplast signal recognition particle mediates binding to the 43-kDa subunit. *J Biol Chem* **2005,** *280* (10), 8912-7.

57. 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 Lett* **2006,** *580* (13), 3107-11.

58. Kathir, K. M.; Rajalingam, D.; Sivaraja, V.; Kight, A.; Goforth, R. L.; Yu, C.; Henry, R.; Kumar, T. K., Assembly of chloroplast signal recognition particle involves structural rearrangement in cpSRP43. *J Mol Biol* **2008,** *381* (1), 49-60.

59. DeLille, J.; Peterson, E. C.; Johnson, T.; Moore, M.; Kight, A.; Henry, R., A novel precursor recognition element facilitates posttranslational binding to the signal recognition particle in chloroplasts. *Proc Natl Acad Sci U S A* **2000,** *97* (4), 1926-31.

60. Tu, C. J.; Peterson, E. C.; Henry, R.; Hoffman, N. E., The L18 domain of light-harvesting chlorophyll proteins binds to chloroplast signal recognition particle 43. *J Biol Chem* **2000,** *275* (18), 13187-90.

61. Gao, F.; Kight, A. D.; Henderson, R.; Jayanthi, S.; Patel, P.; Murchison, M.; Sharma, P.; Goforth, R. L.; Kumar, T. K.; Henry, R. L.; Heyes, C. D., Regulation of Structural Dynamics within a Signal Recognition Particle Promotes Binding of Protein Targeting Substrates. *J Biol Chem* **2015,** *290* (25), 15462-74.

62. Moore, M.; Goforth, R. L.; Mori, H.; Henry, R., Functional interaction of chloroplast SRP/FtsY with the ALB3 translocase in thylakoids: substrate not required. *J Cell Biol* **2003,** *162* (7), 1245-54.

63. Jaru-Ampornpan, P.; Nguyen, T. X.; Shan, S. O., A distinct mechanism to achieve efficient signal recognition particle (SRP)-SRP receptor interaction by the chloroplast srp pathway. *Mol Biol Cell* **2009,** *20* (17), 3965-73.

64. Lewis, N. E.; Marty, N. J.; Kathir, K. M.; Rajalingam, D.; Kight, A. D.; Daily, A.; Kumar, T. K.; Henry, R. L.; Goforth, R. L., A dynamic cpSRP43-Albino3 interaction mediates translocase regulation of chloroplast signal recognition particle (cpSRP)-targeting components. *J Biol Chem* **2010,** *285* (44), 34220-30.

65. Horn, A.; Hennig, J.; Ahmed, Y. L.; Stier, G.; Wild, K.; Sattler, M.; Sinning, I., Structural basis for cpSRP43 chromodomain selectivity and dynamics in Alb3 insertase interaction. *Nat Commun* **2015,** *6*, 8875.

66. Jirgensons, B., Classification of proteins according to conformation. *Die Makromolekulare Chemie* **1966,** *91* (1), 74-86.

67. Dunker, A. K.; Babu, M. M.; Barbar, E.; Blackledge, M.; Bondos, S. E.; Dosztányi, Z.; Dyson, H. J.; Forman-Kay, J.; Fuxreiter, M.; Gsponer, J.; Han, K. H.; Jones, D. T.; Longhi, S.; Metallo, S. J.; Nishikawa, K.; Nussinov, R.; Obradovic, Z.; Pappu, R. V.; Rost, B.; Selenko, P.; Subramaniam, V.; Sussman, J. L.; Tompa, P.; Uversky, V. N., What's in a name? Why these proteins are intrinsically disordered: Why these proteins are intrinsically disordered. *Intrinsically Disord Proteins* **2013,** *1* (1), e24157.

68. Oldfield, C. J.; Dunker, A. K., Intrinsically disordered proteins and intrinsically disordered protein regions. *Annu Rev Biochem* **2014,** *83*, 553-84.

69. Wright, P. E.; Dyson, H. J., Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J Mol Biol* **1999,** *293* (2), 321-31.

70. Lesk, A. M.; Tramontano, A., The Computational Analysis of Protein Structures: Sources, Methods, Systems and Results. *J Res Natl Inst Stand Technol* **1989,** *94* (1), 85-92.

71. Dyson, H. J.; Wright, P. E., Coupling of folding and binding for unstructured proteins. *Curr Opin Struct Biol* **2002,** *12* (1), 54-60.

72. Romero, P.; Obradovic, Z.; Kissinger, C. R.; Villafranca, J. E.; Garner, E.; Guilliot, S.; Dunker, A. K., Thousands of proteins likely to have long disordered regions. *Pac Symp Biocomput* **1998**, 437-48.

73. Ward, J. J.; Sodhi, J. S.; McGuffin, L. J.; Buxton, B. F.; Jones, D. T., Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J Mol Biol* **2004,** *337* (3), 635-45.

74. Baker, C. M.; Best, R. B., Insights into the binding of intrinsically disordered proteins from molecular dynamics simulation. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2014,** *4* (3), 182-198.

75. Bhattacharya, S.; Lin, X., Recent Advances in Computational Protocols Addressing Intrinsically Disordered Proteins. *Biomolecules* **2019,** *9* (4).

76. Day, R. N.; Davidson, M. W., Fluorescent proteins for FRET microscopy: monitoring protein interactions in living cells. *Bioessays* **2012,** *34* (5), 341-50.

77. Kuzmenkina, E. V.; Heyes, C. D.; Nienhaus, G. U., Single-molecule Forster resonance energy transfer study of protein dynamics under denaturing conditions. *Proc Natl Acad Sci U S A* **2005,** *102* (43), 15471-6.

78. LeBlanc, S. J.; Kulkarni, P.; Weninger, K. R., Single Molecule FRET: A Powerful Tool to Study Intrinsically Disordered Proteins. *Biomolecules* **2018,** *8* (4).

79. Best, R. B.; Hofmann, H.; Nettels, D.; Schuler, B., Quantitative interpretation of FRET experiments via molecular simulation: force field and validation. *Biophys J* **2015,** *108* (11), 2721-31.

80. Gomes, G. N.; Gradinaru, C. C., Insights into the conformations and dynamics of intrinsically disordered proteins using single-molecule fluorescence. *Biochim Biophys Acta Proteins Proteom* **2017,** *1865* (11 Pt B), 1696-1706.

81. Ferreon, A. C.; Moran, C. R.; Gambin, Y.; Deniz, A. A., Single-molecule fluorescence studies of intrinsically disordered proteins. *Methods Enzymol* **2010,** *472*, 179-204.

82. Richter, C. V.; Bals, T.; Schünemann, D., Component interactions, regulation and mechanisms of chloroplast signal recognition particle-dependent protein transport. *European Journal of Cell Biology* **2010,** *89* (12), 965-973.

83. 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.

84. Cain, P.; Holdermann, I.; Sinning, I.; Johnson, A. E.; Robinson, C., Binding of chloroplast signal recognition particle to a thylakoid membrane protein substrate in aqueous solution and delineation of the cpSRP43-substrate interaction domain. *Biochem J* **2011,** *437* (1), 149-55.

85. Liang, F.-C.; Kroon, G.; McAvoy, C. Z.; Chi, C.; Wright, P. E.; 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**.

86. Holtkamp, W.; Lee, S.; Bornemann, T.; Senyushkina, T.; Rodnina, M. V.; Wintermeyer, W., Dynamic switch of the signal recognition particle from scanning to targeting. **2012,** *19* (12), 1332-1337.

87. Acerbo, A. S.; Cook, M. J.; Gillilan, R. E., Upgrade of MacCHESS facility for X-ray scattering of biological macromolecules in solution. *Journal of Synchrotron Radiation* **2015,** *22* (1), 180-186.

88. Konarev, P. V.; Volkov, V. V.; Sokolova, A. V.; Koch, M. H. J.; Svergun, D. I., PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *Journal of Applied Crystallography* **2003,** *36* (5), 1277-1282.

89. Hura, G. L.; Menon, A. L.; Hammel, M.; Rambo, R. P.; Poole Ii, F. L.; Tsutakawa, S. E.; Jenney Jr, F. E.; Classen, S.; Frankel, K. A.; Hopkins, R. C.; Yang, S.-j.; Scott, J. W.; Dillard, B. D.; Adams, M. W. W.; Tainer, J. A., Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS). *Nat Meth* **2009,** *6* (8), 606-612.

90. Putnam, C. D.; Hammel, M.; Hura, G. L.; Tainer, J. A., X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Quarterly Reviews of Biophysics* **2007,** *40* (03), 191-285.

91. Forster, S.; Apostol, L.; Bras, W., Scatter: software for the analysis of nano- and mesoscale small-angle scattering. *Journal of Applied Crystallography* **2010,** *43* (3), 639-646.

92. 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-W544.

93. Schneidman-Duhovny, D.; Hammel, M.; Tainer, John A.; Sali, A., Accurate SAXS Profile Computation and its Assessment by Contrast Variation Experiments. *Biophysical Journal* **2013,** *105* (4), 962-974.

94. Pelikan, M.; Hura, G. L.; Hammel, M., Structure and flexibility within proteins as identified through small angle X-ray scattering. *Gen. Physiol. Biophys.* **2009,** *28* (2), 174-189.

95. Sivaraja, V.; Kumar, T. K.; Leena, P. S.; Chang, A. N.; Vidya, C.; Goforth, R. L.; Rajalingam, D.; Arvind, K.; Ye, J. L.; Chou, J.; Henry, R.; Yu, C., Three-dimensional solution structures of the chromodomains of cpSRP43. *J Biol Chem* **2005,** *280* (50), 41465-71.

96. Holdermann, I.; Meyer, N. H.; Round, A.; Wild, K.; Sattler, M.; Sinning, I., Chromodomains read the arginine code of post-translational targeting. *Nat Struct Mol Biol* **2012,** *19* (2), 260-3.

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

98. Santoso, Y.; Torella , J. P.; Kapanidis, A. N., Characterizing Single-Molecule FRET Dynamics with Probability Distribution Analysis. *ChemPhysChem* **2010,** *11* (10), 2209-2219.

99. Rosendal, K. R.; Wild, K.; Montoya, G.; Sinning, I., Crystal structure of the complete core of archaeal signal recognition particle and implications for interdomain communication. *Proc Natl Acad Sci U S A* **2003,** *100* (25), 14701-6.

100. Sharma, P.; Kight, A.; Goforth, R.; Henry, R., GTP Hydrolysis is Not Required for Post-translational Protein Insertion by a Chloroplast Signal Recognition Particle. *The FASEB Journal* **2015,** *29* (1 Supplement).

101. Stengel, K. F.; Holdermann, I.; Wild, K.; Sinning, I., The structure of the chloroplast signal recognition particle (SRP) receptor reveals mechanistic details of SRP GTPase activation and a conserved membrane targeting site. *FEBS Letters* **2007,** *581* (29), 5671-5676.

102. Wang, P.; Dalbey, R. E., Inserting membrane proteins: The YidC/Oxa1/Alb3 machinery in bacteria, mitochondria, and chloroplasts. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **2011,** *1808* (3), 866-875.

103. Chandrasekar, S.; Shan, S. O., Anionic Phospholipids and the Albino3 Translocase Activate Signal Recognition Particle-Receptor Interaction during Light-harvesting Chlorophyll a/b-binding Protein Targeting. *J Biol Chem* **2017,** *292* (1), 397-406.

104. 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 Lett* **2010,** *584* (19), 4138-44.

105. Urbischek, M.; von Braun, S. N.; Brylok, T.; xfc; gel, I. L.; Richter, A.; Koskela, M.; Grimm, B.; Mulo, P.; xf; lter, B.; Soll, J.; rgen; Ankele, E.; Schwenkert, S., The extreme Albino3 (Alb3) C terminus is required for Alb3 stability and function in Arabidopsis thaliana. *Planta* **2015,** *242* (3), 733-746.

106. Dong, H.; Sharma, M.; Zhou, H. X.; Cross, T. A., Glycines: role in α-helical membrane protein structures and a potential indicator of native conformation. *Biochemistry* **2012,** *51* (24), 4779-89.

107. Chakrabartty, A.; Schellman, J. A.; Baldwin, R. L., Large differences in the helix propensities of alanine and glycine. *Nature* **1991,** *351* (6327), 586-8.

108. Xue, B.; Dunbrack, R. L.; Williams, R. W.; Dunker, A. K.; Uversky, V. N., PONDR-FIT: A meta-predictor of intrinsically disordered amino acids. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2010,** *1804* (4), 996-1010.

109. Ward, J. J.; McGuffin, L. J.; Bryson, K.; Buxton, B. F.; Jones, D. T., The DISOPRED server for the prediction of protein disorder. *Bioinformatics* **2004,** *20* (13), 2138-2139.

110. Altschul, S. F.; Madden, T. L.; Schäffer, A. A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D. J., Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research* **1997,** *25* (17), 3389-3402.

111. Ishida, T.; Kinoshita, K., PrDOS: prediction of disordered protein regions from amino acid sequence. *Nucleic Acids Research* **2007,** *35* (suppl_2), W460-W464.

112. Cilia, E.; Pancsa, R.; Tompa, P.; Lenaerts, T.; Vranken, W. F., The DynaMine webserver: predicting protein dynamics from sequence. *Nucleic acids research* **2014,** *42* (W1), W264-W270.

113. Galzitskaya, O. V.; Garbuzynskiy, S. O.; Lobanov, M. Y., FoldUnfold: web server for the prediction of disordered regions in protein chain. *Bioinformatics* **2006,** *22* (23), 2948-2949.

114. Provencher, S. W., CONTIN: a general purpose constrained regularization program for inverting noisy linear algebraic and integral equations. *Computer Physics Communications* **1982,** *27* (3), 229-242.

115. Sreerama, N.; Woody, R. W., 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), 252-260.

116. Liang, S.; Zhang, C.; Zhou, Y., LEAP: highly accurate prediction of protein loop conformations by integrating coarse-grained sampling and optimized energy scores with all-atom refinement of backbone and side chains. *J Comput Chem* **2014,** *35* (4), 335-41.

117. Eswar, N.; Eramian, D.; Webb, B.; Shen, M.-Y.; Sali, A., Protein Structure Modeling with MODELLER. In *Structural Proteomics: High-Throughput Methods*, Humana Press: Totowa, NJ, 2008; pp 145-159.

118. Barducci, A.; Bussi, G.; Parrinello, M., Well-tempered metadynamics: a smoothly converging and tunable free-energy method. *Phys Rev Lett* **2008,** *100* (2), 020603.

119. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of simple potential functions for simulating liquid water. *Journal of Chemical Physics* **1983,** *79*, 926-935.

120. Tanner, D. E.; Chan, K. Y.; Phillips, J. C.; Schulten, K., Parallel Generalized Born Implicit Solvent Calculations with NAMD. *J Chem Theory Comput* **2011,** *7* (11), 3635-3642.

121. Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K., Scalable molecular dynamics with NAMD. *J Comput Chem* **2005,** *26* (16), 1781-802.

122. Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E.; Mittal, J.; Feig, M.; Mackerell, A. D., Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone φ, ψ and side-chain χ(1) and χ(2) dihedral angles. *J Chem Theory Comput* **2012,** *8* (9), 3257-3273.

123. Reid, J., Large Sparse Sets of Linear Equations. Academic Press: London and New York, 1971.

124. Darden, T.; York, D.; Pedersen, L., Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics* **1993,** *98* (12), 10089-10092.

125. Bernardin; A; JR, C.; G, F.; H, F.; J, H.; A, K.; F, M.; JV, V.; AD, W., COLLECTIVE VARIABLES MODULE - Reference manual for NAMD. Version -

October 2019.

126. Humphrey, W.; Dalke, A.; Schulten, K., VMD: visual molecular dynamics. *J Mol Graph* **1996,** *14* (1), 33-8, 27-8.

127. Andersen, C. A.; Palmer, A. G.; Brunak, S.; Rost, B., Continuum secondary structure captures protein flexibility. *Structure* **2002,** *10* (2), 175-84.

128. Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D., CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J Comput Chem* **2010,** *31* (4), 671-90.

129. Greenfield, N. J., Using circular dichroism spectra to estimate protein secondary structure. *Nature protocols* **2006,** *1* (6), 2876.

130. Johnson, W. C., Analyzing protein circular dichroism spectra for accurate secondary structures. *Proteins: Structure, Function, and Bioinformatics* **1999,** *35* (3), 307-312.

131. Woody, R. W., *Circular dichroism of intrinsically disordered proteins*. John Wiley & Sons, Hoboken, NJ: 2010.

132. Keeler, J., *Understanding NMR spectroscopy*. John Wiley & Sons: 2011.

133. Kikhney, A. G.; Svergun, D. I., A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. *FEBS letters* **2015,** *589* (19PartA), 2570-2577.

134. Ronimus, R. S.; Musgrave, D. R., Identification of a gene in the euryarchaeal Thermococcus species AN1 encoding a protein homologous to the alpha subunit of the eukaryal signal recognition particle (SRP) receptor. *Biochim Biophys Acta* **1997,** *1351* (1-2), 1-8.

135. Falk, S.; Sinning, I., The C terminus of Alb3 interacts with the chromodomains 2 and 3 of cpSRP43. *J Biol Chem* **2010,** *285* (53), le25-6; author reply le26-8.

136. Schuler, B.; Lipman, E. A.; Steinbach, P. J.; Kumke, M.; Eaton, W. A., Polyproline and the "spectroscopic ruler" revisited with single-molecule fluorescence. *Proc Natl Acad Sci U S A* **2005,** *102* (8), 2754-9.

137. Kandt, C.; Monticelli, L., Membrane protein dynamics from femtoseconds to seconds. *Methods Mol Biol* **2010,** *654*, 423-40.

138. Klepeis, J. L.; Lindorff-Larsen, K.; Dror, R. O.; Shaw, D. E., Long-timescale molecular dynamics simulations of protein structure and function. *Curr Opin Struct Biol* **2009,** *19* (2), 120-7.

139. Matsunaga, Y.; Sugita, Y., Linking *time-series* of single-molecule experiments with molecular dynamics simulations by machine learning. *Elife* **2018,** *7*.

140. Choi, U. B.; Weninger, K. R.; Bowen, M. E., Immobilization of proteins for single-molecule fluorescence resonance energy transfer measurements of conformation and dynamics. *Methods Mol Biol* **2012,** *896*, 3-20.

141. Rhoades, E.; Gussakovsky, E.; Haran, G., Watching proteins fold one molecule at a time. *Proc Natl Acad Sci U S A* **2003,** *100* (6), 3197-202.

142. Resch-Genger, U.; Grabolle, M.; Cavaliere-Jaricot, S.; Nitschke, R.; Nann, T., Quantum dots versus organic dyes as fluorescent labels. *Nat Methods* **2008,** *5* (9), 763-75.

143. Eggeling, C.; Volkmer, A.; Seidel, C. A. M., Molecular Photobleaching Kinetics of Rhodamine 6G by One- and Two-Photon Induced Confocal Fluorescence Microscopy. *ChemPhysChem* **2005,** *6* (5), 791-804.

144. Cooper, D.; Uhm, H.; Tauzin, L. J.; Poddar, N.; Landes, C. F., Photobleaching lifetimes of cyanine fluorophores used for single-molecule Förster resonance energy transfer in the presence of various photoprotection systems. *Chembiochem* **2013,** *14* (9), 1075-80.

145. Shi, X.; Lim, J.; Ha, T., Acidification of the oxygen scavenging system in single-molecule fluorescence studies: in situ sensing with a ratiometric dual-emission probe. *Anal Chem* **2010,** *82* (14), 6132-8.

146. Bera, D.; Qian, L.; Tseng, T.-K.; Holloway, P. H., Quantum Dots and Their Multimodal Applications: A Review. *Materials* **2010,** *3* (4), 2260-2345.

147. Efros, A. L.; Nesbitt, D. J., Origin and control of blinking in quantum dots. *Nat Nanotechnol* **2016,** *11* (8), 661-71.

148. Mahler, B.; Spinicelli, P.; Buil, S.; Quelin, X.; Hermier, J. P.; Dubertret, B., Towards non-blinking colloidal quantum dots. *Nat Mater* **2008,** *7* (8), 659-64.

149. Yuan, G.; Gómez, D. E.; Kirkwood, N.; Boldt, K.; Mulvaney, P., Two Mechanisms Determine Quantum Dot Blinking. *ACS Nano* **2018,** *12* (4), 3397-3405.

150. Huang, B.; Yang, H.; Zhang, L.; Yuan, Y.; Cui, Y.; Zhang, J., Effect of surface/interfacial defects on photo-stability of thick-shell CdZnSeS/ZnS quantum dots. *Nanoscale* **2018,** *10* (38), 18331-18340.

151. Vela, J.; Htoon, H.; Chen, Y.; Park, Y. S.; Ghosh, Y.; Goodwin, P. M.; Werner, J. H.; Wells, N. P.; Casson, J. L.; Hollingsworth, J. A., Effect of shell thickness and composition on blinking suppression and the blinking mechanism in 'giant' CdSe/CdS nanocrystal quantum dots. *J Biophotonics* **2010,** *3* (10-11), 706-17.

152. Fiolka, R.; Belyaev, Y.; Ewers, H.; Stemmer, A., Even illumination in total internal reflection fluorescence microscopy using laser light. *Microscopy Research and Technique* **2008,** *71* (1), 45-50.

153. Shuang, B.; Cooper, D.; Taylor, J. N.; Kisley, L.; Chen, J.; Wang, W.; Li, C. B.; Komatsuzaki, T.; Landes, C. F., Fast Step Transition and State Identification (STaSI) for Discrete Single-Molecule Data Analysis. *J Phys Chem Lett* **2014,** *5* (18), 3157-3161.

154. Gao, F.; Bajwa, P.; Nguyen, A.; Heyes, C. D., Shell-Dependent Photoluminescence Studies Provide Mechanistic Insights into the Off-Grey-On Transitions of Blinking Quantum Dots. *ACS Nano* **2017,** *11* (3), 2905-2916.