Rubisco folding and oligomeric assembly: Detailed analysis of an assembly intermediate

Amanda Windhof
(geb. Starling)

aus

Delaware, Ohio, USA

2011

Acknowledgments

Contents

1   Summary

To become biologically active, a protein must fold into a distinct three-dimensional structure. Many non-native proteins require molecular chaperones to support folding and assembly. These molecular chaperones are important for *de novo* protein folding as well as refolding of denatured proteins under stress conditions. A certain subset of chaperones, the chaperonins, are required for the folding of the enzyme ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco); furthermore, correct folding of Rubisco is also aided by the Hsp70 chaperone system.

Rubisco catalyzes the initial step of CO_2 assimilation in the Calvin-Benson-Bassham (CBB) cycle. Unfortunately, this enzyme is extremely inefficient, not only does it exhibit a slow catalytic rate (three CO_2 molecules fixed per second per Rubisco) but it also discriminates poorly between the assimilation of CO_2 and O_2 to its sugar-phosphate substrate ribulose-1,5-bisphosphate (RuBP), the latter resulting in loss of photosynthetic efficiency. Due to these inefficiencies, carbon fixation by Rubisco is the rate limiting step of the CBB cycle. Photosynthetic organisms must produce tremendous amounts of Rubisco to alleviate these shortcomings; therefore significant quantities of nitrogen stores are invested in the production of Rubisco making Rubisco the most abundant protein on earth. These drawbacks of Rubisco have important implications in increasing CO_2 concentrations and temperatures in the context of global warming. The ability to engineer a more efficient Rubisco could potentially reduce photosynthetic water usage, increase plant growth yield, and reduce nitrogen usage is plants. However, eukaryotic Rubisco cannot fold and assemble outside of the chloroplast, hindering advancements in creating a more efficient Rubisco.

Form I Rubisco, found in higher plants, algae, and cyanobacteria, is a hexadecameric complex consisting of a core of eight ~50 kDa large subunits (RbcL), which is capped by four ~15 kDa small subunits (RbcS) on each end. The discovery of a Rubisco-specific assembly chaperone, RbcX, has lead to a better understanding of the components necessary for the form I Rubisco assembly process. RbcX is a homodimer of ~15 kDa subunits consisting of four α-helices aligned in an anti-parallel fashion along the α4 helix. $RbcX_2$ functions as a stabilizer of folded RbcL by recognizing a highly conserved C-terminal sequence of RbcL: EIKFEFD, termed the C-terminal recognition motif. As has been demonstrated by studies of cyanobacterial Rubisco, *de novo* synthesized RbcL is folded by the chaperonins, whereupon $RbcX_2$ stabilizes the folded RbcL monomer upon release from the folding cavity and then assists in the formation of the $RbcL_8$ core. $RbcX_2$ forms a dynamic complex with $RbcL_8$ and as a result, $RbcX_2$ is readily displaced by RbcS docking in an ATP-independent manner,

thereby creating the functional holoenzyme. However, the exact mechanism by which RbcS binding displaces $RbcX_2$ from the $RbcL_8$ core is still unknown. Furthermore, though much advancement has been made in the understanding of form I Rubisco folding and assembly, an exact and detailed mechanism of form I Rubisco assembly is still lacking. The highly dynamic complex of RbcL/RbcX is critical for the formation of the holoenzyme; however it has hindered attempts to characterize critical regions of RbcL that interact with the peripheral regions of $RbcX_2$. An important observation arose when heterologous RbcL and $RbcX_2$ components interacted; a stable complex could form enabling in depth characterization of the RbcL/$RbcX_2$ interaction.

In the present study, the detailed structural mechanism of $RbcX_2$-mediated cyanobacterial form I Rubisco assembly is elucidated. To obtain molecular insight into the $RbcX_2$-mediated assembly process of cyanobacterial form I Rubisco, cryo-EM and crystallographic studies in concert with mutational analysis were employed by taking advantage of the high affinity interaction between RbcL and $RbcX_2$ in the heterologous system (*Synechococcus* sp. PCC6301 RbcL and *Anabaena* sp. CA $RbcX_2$). Structure guided mutational analysis based on the 3.2 Å crystal structure of the $RbcL_8/(RbcX_2)_8$ assembly intermediate were utilized to determine the precise interaction site between the body of RbcL and the peripheral region of $RbcX_2$. From these studies a critical salt bridge could be identified that functions as a guide-point for correct dimer formation, and it was observed that $RbcX_2$ exclusively mediates Rubisco dimer assembly. Furthermore, the mechanism of $RbcX_2$ displacement from the $RbcL_8$ core by RbcS binding was elucidated as well as indications of how RbcS docking on the $RbcL_8$ core is imperative for full form I Rubisco catalytic function by stabilizing the enzymatically competent conformation of an N-terminal loop of Rubisco termed the '60ies loop'. Finally, initial attempts in *in vitro* reconstitution of eukaryotic Rubisco are reported along with the characterization of *Arabidopsis thaliana* $RbcX_2$ binding to the C-terminal recognition motif of the Rubisco large subunit from various species.

2   Introduction

Proteins are polymers of amino acids linked together by peptide bonds formed between adjacent carboxyl and amino groups. They are necessary for the correct functioning of biological processes in cells from all domains of life. For example these macromolecular molecules are involved in processes such as cellular structure, metabolism, signaling, enzymatic catalysis, immune response, storage, transportation, movement, and regulation.

2.1   Protein folding

In order to become biologically functional, proteins must fold into the proper three-dimensional structure. The physiological function of a protein is dependent on its unique structure as established by the primary, secondary, tertiary, and quaternary structures. The primary structure of a protein is determined by the sequence of amino acids along the polypeptide chain as encoded by the respective gene upon translation. Amino acid composition in the primary structure, not only determines the solubility and ionic properties but also the native structure of the respective protein. The secondary structure arises from formation of α-helical, β-strand/sheets, and loop/turn elements stabilized by hydrogen bonds. The secondary structures combine due to residue interactions, driven in part by the burying of hydrophobic surfaces, to assume the tertiary structure. The tertiary structure is the final overall conformation of a single protein including the formation of α-helical bundles, β-sheets, and/or (α/β) TIM-barrels, which are stabilized by hydrogen bonds, salt bridges, posttranslational modifications, and/or disulfide bond formation. In the tertiary structure, regions that were previously spatially separated in the primary structure come into close contact due to the compaction of the polypeptide chain. The quaternary structure arises from the stable interaction of two or more protein molecules through hydrophobic or polar contacts of side chain residues to form a functional complex. The quaternary structure is stabilized by hydrogen bonds, van-der-Waals contacts, and/or ionic forces.

Many small proteins *in vitro* are able to spontaneously fold to their native conformation leading to the hypothesis that the amino acid sequence of a polypeptide chain contains all the required information necessary for assuming the correct three-dimensional structure (Anfinsen 1973). The main driving force of spontaneous folding is the reduction in free-energy arising from polypeptide compaction. However, this ability to spontaneously fold is typically not reflected in large, complex, multi-domain proteins where off-pathway

reactions lead prominently to aggregation products; indicating that the primary sequence alone is not sufficient to drive correct protein folding.

Inherently, the native state of a protein is the most thermodynamically stable, and accordingly protein folding follows the pathway from the high energy extended conformation to the low energy native structure (Frydman 2001). It is well established that protein folding occurs rapidly (<1 s). Therefore demonstrating that protein folding is not a random process, and proteins do not sample all possible conformations before reaching their native structure. The formation of initial structural features leads to the further positioning of successive structural elements; otherwise folding would not proceed in a biologically relevant timeframe (Dobson et al. 1998; Levinthal 1968). As illustrated by the funnel shaped energy landscape representing the free-energy potential during folding of a protein chain (Figure 2.1), there are local energy minima and maxima which must be overcome to obtain the stable, low-energy native state (Hartl and Hayer-Hartl 2009). The ruggedness of the energy landscape indicates that during folding, while the protein may become trapped at local energy minima, proteins adopt conformations/intermediate states with constantly decreasing energy potentials (Jahn and Radford 2005). The final native state is the most stable and energetically efficient structure which can be obtained. Collapse of the elongated non-native protein is initiated by the burial of hydrophobic residues whereupon hydrogen bonds, charge-charge interactions, and van-der-Waals forces contribute to stabilization.

In contrast to *in vitro* conditions, protein folding in the cell can result in aberrant, off-pathway intermediates due to the extremely crowded environment in the cell. Upon release from the ribosome, newly translated proteins are subjected to the highly crowded cellular environment, which is confounded with very high concentrations of macromolecules (300-400 mg/ml) (Zimmerman and Trach 1991) creating an environment that is non-conducive for spontaneous folding. Furthermore, during translation of a polypeptide chain, the entire sequence is not free for folding as the N-terminus is exposed to the cytosol, while the C-terminus is enclosed by the ribosome exit tunnel during synthesis. Hence, the hypothesis that protein folding can also occur cotranslationally. However, if in the native state essential interactions are present between the exposed N-terminus and ribosome enclosed C-terminus, then the translating polypeptide must be in an extended form and as a consequence residues which are hidden in the native structure would be exposed leading to misfolding before the nascent chain is released from the ribosome exit tunnel. Thus, the extended polypeptide must be somehow protected from aberrant interactions during translation (Hartl 1996).
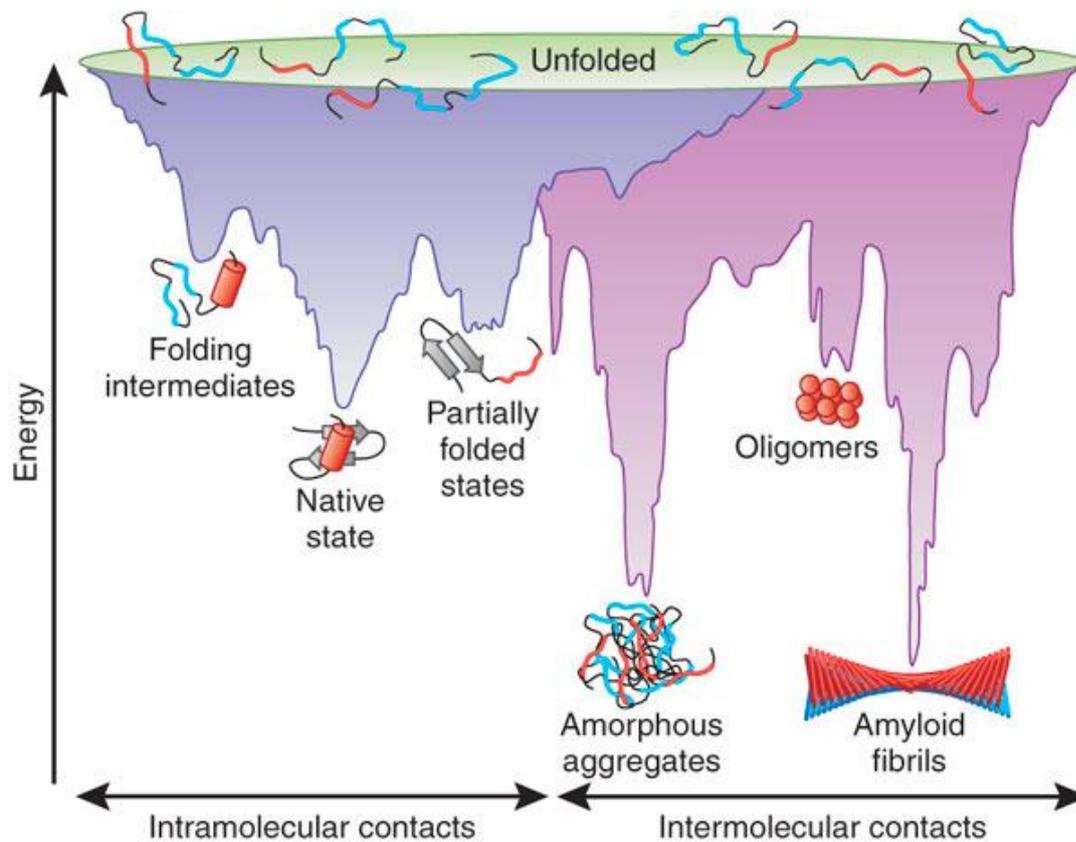
Figure 2.1 Energy landscape of protein folding.
Left side: purple funnel indicates the energy barriers present and intermediates formed during protein folding to the native state. Right side: pink funnel indicates the energy barriers present and intermediates formed when aggregation species are prevalent. Note that overlap occurs between the two surfaces. Intermediates formed during protein folding can lead to the formation of aggregates. Adapted from: (Hartl and Hayer-Hartl 2009)

The major consequence of aberrant interactions and misfolding is aggregation, especially in the presence of kinetically-trapped folding intermediates. Many small, topologically simple proteins maintain the ability to rapidly fold spontaneously without chaperone assistance, however in the case of large, topologically complex proteins, spontaneous folding can lead to the formation of aggregates where chaperones are required for productive folding (Figure 2.2 A and B). Aggregation, similar to the initial stages of protein folding, is driven by hydrophobic forces (Vabulas et al. 2010). During aggregation, non-native proteins interact resulting in the formation of unstructured amorphous elements, which in the earliest forms can still be degraded/reversed and processed by the cell (Lin and Rye 2006). Under some conditions, these amorphous elements can attain highly ordered, stable structures leading to the formation of fibrillar amyloid aggregates (Figure 2.2 C) (Hartl and Hayer-Hartl 2009), which are highly resistant to the cellular degradation machinery. During the formation of degradation-resistant amyloid fibrils, less ordered, soluble oligomeric species are present which adversely affect cell function and may be toxic (Olzscha et al. 2011). These oligomeric species are consequently linked to neurodegenerative diseases such

as Alzheimer's, Parkinson's, Huntington's, and prion disease (Haass and Selkoe 2007). To ensure that aggregation is limited during protein synthesis, folding, and stress conditions, many proteins rely on other macromolecules, chaperones, for assistance to obtain correct native structure.
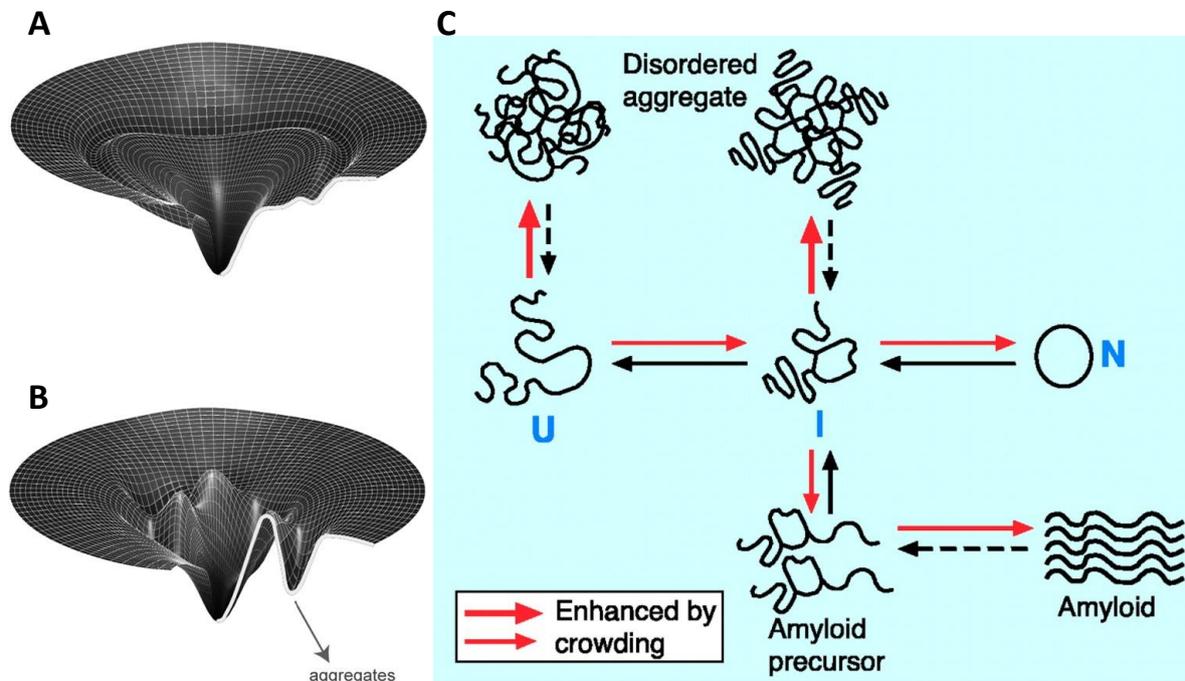
Figure 2.2 Protein aggregation pathways in the cell.
(A) The folding-energy landscape of a small, rapidly folding protein that requires no additional chaperone assistance. (B) The folding-energy landscape of a large protein that requires chaperone assistance to achieve native state due to aggregate formation propensity. A and B adapted from: (Lin and Rye 2006) (C) Red arrows indicate pathway enhancement by cellular crowding. U: newly synthesized unfolded protein chain, I: intermediate, partially folded, N: native protein. Amyloid formation is predicted to be enhanced by molecular crowding, but this has yet to be experimentally proven. Adapted from: (Hartl and Hayer-Hartl 2002)

2.2   Molecular chaperones

The term molecular chaperone was coined in 1978 to describe a particular nuclear protein assisting nucleosome assembly (Laskey et al. 1978). However, the use of this term has broadened to describe a class of macromolecules that supports folding and assembly of non-native proteins without becoming a part of or influencing their final functional conformation (Ellis 1987). Molecular chaperones are important for *de novo* protein folding, oligomeric assembly, intracellular transport, refolding of stress-denatured proteins, and assist in proteolytic degradation. Here, the involvement of molecular chaperones in protein folding and assembly will be discussed.

2.2.1 *De novo* protein folding

As mentioned above, during nascent polypeptide chain synthesis and emergence from the ribosome, extended regions of the chain are exposed to the crowded cellular environment.

Importantly, many of these exposed regions i.e., hydrophobic side chains, become buried upon formation of the native protein and their transient exposure could lead to unfavorable non-native interactions. To avoid off-pathway, aberrant side reactions during translation, a special class of molecular chaperones specifically interact with nascent chains emerging from the ribosome. This first class of chaperones can be classified as the nascent-chain interacting chaperones, consisting of Ribosome-associated chaperones and the Hsp70 system. Proteins that fail to fold by these specific chaperones are transferred to the downstream chaperones called chaperonins, which provide a sequestered environment for folding. *De novo* protein folding is ubiquitously present in all three kingdoms of life and is delineated by the initial interaction of specific chaperones with the translating polypeptide chain followed by transfer to downstream chaperones specialized in finalizing the folding process (Figure 2.3) (Hartl and Hayer-Hartl 2009).
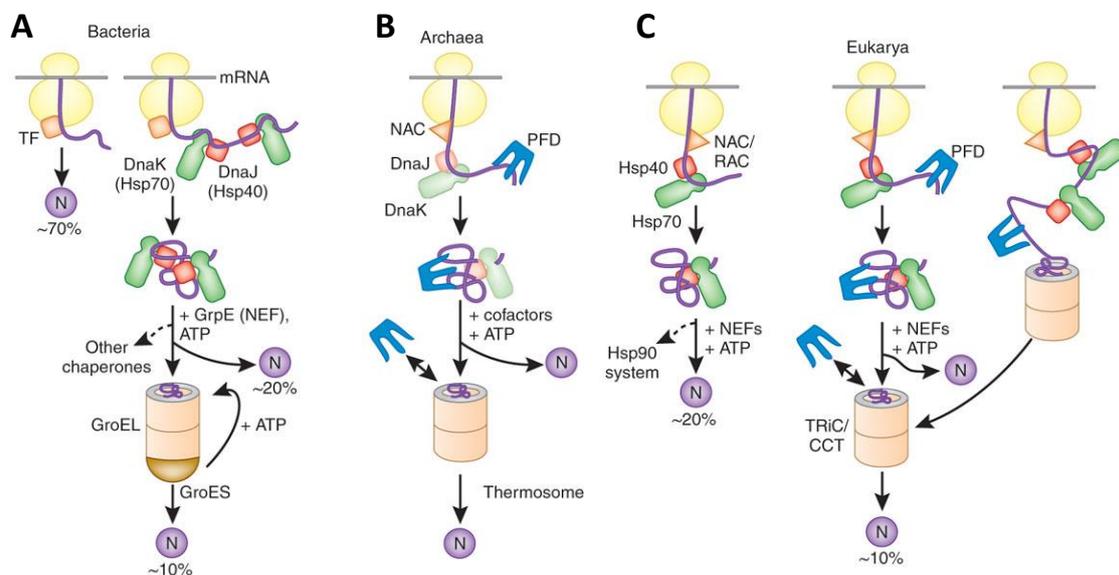
Figure 2.3 Model of chaperone-assisted folding of nascent chains in the cytosol.
(A) Bacteria: newly synthesized polypeptide chains interact with Trigger factor (TF) and small proteins, 70% of total, fold to the native state (N) after this interaction. Larger polypeptides, 20% of total, further require the assistance of the ATP-dependent capture and release of the Hsp70 system (DnaK/DnaJ/GrpE). If the native state is still not achieved, 10% of total, these non-native proteins are subsequently transferred to the chaperonin system (GroEL/GroES) for cycles of ATP-dependent encapsulation and release. (B) Archaea: nascent chain-associated complex (NAC), prefoldin (PFD), in some archaeal species DnaK and DnaJ are present, and interact with nascent chains and assist with the transfer to chaperonin, thermosome. (C) Eukarya: NAC most likely interacts with nascent chains but its actual role in folding is not fully understood. 20% of nascent chains interact with ribosome-associated complex (RAC) and the Hsp70 system to obtain their native state, however some are further transferred to Hsp90. 10% of nascent chains are co- or post-translationally transferred to the chaperonin system (TRiC/CCT) from Hsp70 and PFD. Adapted from: (Hartl and Hayer-Hartl 2009)

2.2.1.1 Ribosome-associated chaperones

Ribosome-associated chaperones not only directly interact with the ribosome but also act on nascent chains emerging from the ribosomal exit tunnel. These chaperones are found in all domains of life and include: bacterial Trigger factor, eukaryotic and archaeal nascent chain-associated complex (NAC), and the eukaryotic heat shock protein 70 (Hsp70) system

called the ribosome-associated complex (RAC) (Figure 2.3 B and C) (Hartl and Hayer-Hartl 2009). Bacterial Trigger factor is the best characterized ribosome-associated chaperone; therefore its role in *de novo* protein folding will be described in further detail.
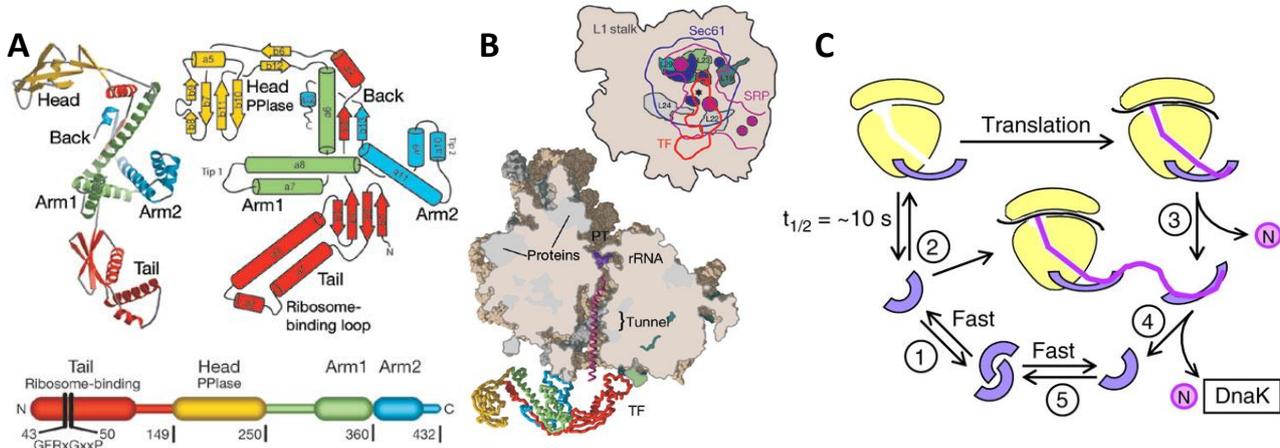
Figure 2.4 Structure of Trigger factor and Trigger factor interaction with the ribosome.
(A) Ribbon and schematic domain representations of Trigger factor structure left and right panel, respectively (PDB 1W26 (Ferbitz et al. 2004)). Bottom: linear sequence of domain organization. N-terminal tail is red, C-terminal Arm 1 is green, C-terminal Arm 2 is blue, and Head PPIase domain is yellow. (B) Trigger factor in complex with the ribosomal 50S complex, coloring as in, A. Nascent chain in magenta extending from the exit tunnel is modeled. Location of ribosomal proteins is indicated. A and B Adapted from: (Ferbitz et al. 2004) (C) Model of TF ribosome and substrate binding. 1. Unbound TF is in equilibrium between the monomeric and dimeric states. 2. Conformational change of TF upon ribosomal binding, which may cause activation for substrate interaction. 3. Translating peptide interacts with TF at the ribosomal exit tunnel. 4. After translation polypeptide and TF are released from the ribosome, the polypeptide is either still bound with TF or folded to its native state. Structurally complex proteins are transferred to DnaK for further folding assistance. C Adapted from: (Vabulas et al. 2010)

Trigger factor (TF) is a 48 kDa protein that interacts with the large ribosomal subunit near the exit tunnel in close proximity to ribosomal proteins L23 and L29 with 1:1 stoichiometry (Ferbitz et al. 2004). TF consists of three domains: the N-terminal ribosome-binding domain, the peptidyl-prolyl isomerase (PPIase) domain, and the C-terminal domain. As observed in the crystal structure, the C-terminal domain is located in between the N-terminal and PPIase domains exhibiting two arm-like extensions, Arm 1 and Arm 2 (Figure 2.4 A and B). While all three domains are implicated in binding hydrophobic regions of extended polypeptide chains, the C-terminal arm-like protrusions and the N-terminal domain primarily interact with the extended polypeptide chain (Lakshmipathy et al. 2007). In free solution TF forms a homodimer, however, TF is monomeric when bound to the ribosome (Figure 2.4 C) (Maier et al. 2003). TF interaction with the ribosome is mediated via residues of the N-terminal domain (Lakshmipathy et al. 2007). Ribosomal binding mediates a conformational change in TF, most likely preparing TF for substrate binding (Kaiser et al. 2006). During translation of especially hydrophobic nascent chains, TF binds the ribosome and attaches to the polypeptide followed by subsequent detachment from the ribosome

(Agashe et al. 2004). Substrate binding and release is ATP-independent, additionally binding is mediated by stretches of exposed hydrophobic patches on the extended nascent chain while polypeptide release apparently results from the collapse of the chain and burial of hydrophobic residues (Figure 2.4 C) (Vabulas et al. 2010). Due to TF binding close to the ribosomal exit tunnel, the majority of translating-nascent chains interact with TF and approximately 70% can reach their native state upon release from TF (Figure 2.3 A). However, if the protein does not achieve native state upon dissociation from TF, the non-native proteins are transferred to downstream chaperone systems for further attempts at chaperone-assisted protein folding.

2.2.1.2 Hsp70 system

Hsp70s act downstream of TF and are not only found as constituently expressed forms in the cell but also as highly inducible forms during stress conditions (Hartl 1996). Hsp70 proteins are important for *de novo* folding of nascent chains and are furthermore involved in protein trafficking and assisting in proteolytic degradation (Bukau and Horwich 1998; Young et al. 2003). The yield of folded protein is enhanced when both TF and the Hsp70 system act on the nascent chain (Agashe et al. 2004). However, Hsp70 and TF are partially functionally redundant. Deletion of TF or Hsp70 alone is not deleterious to cell growth, conversely cells are not viable if both of these proteins are deleted (Teter et al. 1999), thus demonstrating substantial overlap in the subset of substrates which bind to and are assisted in folding by TF and Hsp70. TF and Hsp70 substrate overlap is interesting considering these chaperones display intrinsic mechanistic differences. TF binds and releases substrate independent of nucleotide, whereas Hsp70 substrate interaction is nucleotide regulated. In bacterial systems Hsp70 (DnaK) acts downstream of TF and typically does not bind the ribosome. In the following section, the bacterial Hsp70 system (DnaK/DnaJ/GrpE system) will be discussed in detail as the bacterial system is the best characterized. Homologues of the entire or specific components of the DnaK/DnaJ/GrpE system are found in the cytosol, mitochondria, chloroplast, and endoplasmic reticulum of eukaryotic cells.

DnaK is composed of an N-terminal nucleotide binding domain and a C-terminal substrate binding domain (Figure 2.5). The C-terminal domain binds polypeptide substrates in an extended conformation and is composed of a β-sandwich domain with an α-helical lid-like segment, while the N-terminal domain has a pocket for ATP binding and is the site of ATP hydrolysis (Zhu et al. 1996). The C-terminal substrate binding domain specifically binds to polypeptide regions exposing approximately five hydrophobic residues in a row with leucine

and isoleucine residues being preferred (Ruediger et al. 1997). Regions of this character occur approximately every 50-100 amino acids in the majority of protein sequences. In this way, DnaK binds exposed hydrophobic regions preventing non-native contacts, which could occur before the native structure is reached and hydrophobic surfaces are buried. DnaK thereby functions in smoothing the energy landscape to achieve efficient protein folding. DnaK has two conformational forms displaying differential affinities to substrate: the open or closed state (Mayer et al. 2000). In the open conformation ATP is bound and DnaK binds polypeptide substrate with relatively low affinity. The cochaperone Hsp40 (DnaJ), a J-domain protein, attaches to the ATP bound DnaK N-terminus with its N-terminal J-domain. DnaJ functions by increasing the rate of ATP hydrolysis of DnaK, thereby enhancing substrate capture (Szabo et al. 1994). Note that DnaJ can also bind the substrate and transfer it to DnaK (Genevaux et al. 2007). ATP hydrolysis leads to the formation of the ADP-bound closed conformation of DnaK, which binds polypeptide substrate with relatively high affinity (Figure 2.5).
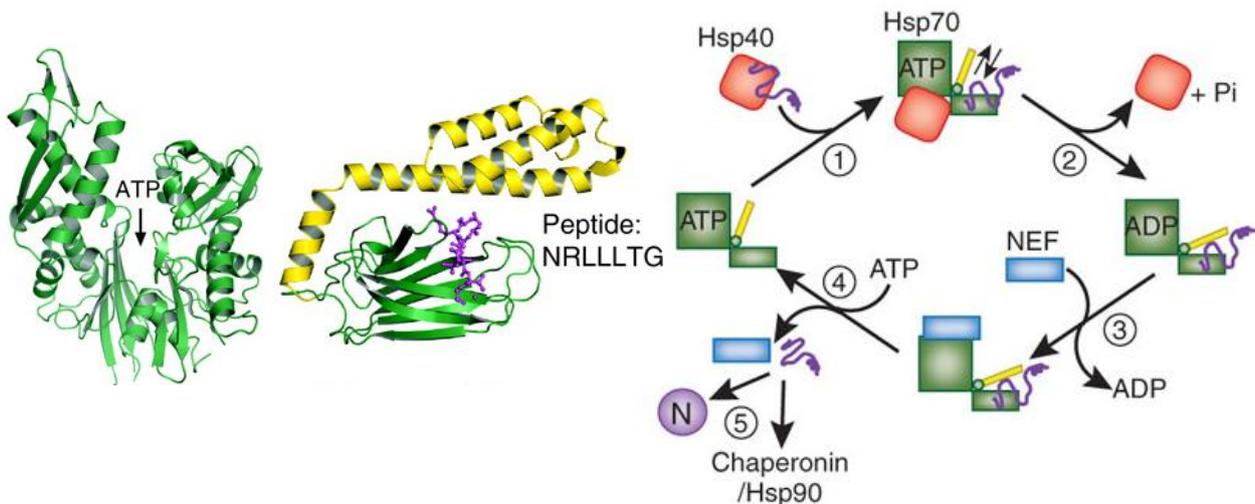
Figure 2.5 Structure of Hsp70 and model of substrate interaction.
Left panel: structure of the DnaK N-terminal ATPase domain with ATP binding site indicated (PDB 1DKG (Harrison et al. 1997)) and the C-terminal peptide binding domain, α-helical lid is indicated in yellow and substrate peptide is indicated in purple and its sequence shown in single-letter code (PDB 1DKZ (Zhu et al. 1996)). Right panel: model of the Hsp70 reaction cycle. 1. Hsp40 delivers substrate to Hsp70 with bound ATP. 2. Hsp40 interaction with Hsp70 accelerates ATP to ADP hydrolysis, Hsp70 in ADP-state results in lid closure and high affinity for substrate. 3. Nucleotide exchange factor (NEF) binds to Hsp70 and catalyzes the release of ADP. 4. ATP binding causes Hsp70 lid to open and low affinity for substrate followed by substrate and NEF release. 5. Released substrate will either fold to native state (N), rebind Hsp70, or be transferred to downstream chaperones. Adapted from: (Hartl and Hayer-Hartl 2009)

Collaboration with an additional protein, the nucleotide exchange factor (NEF) GrpE, is necessary for efficient substrate release from DnaK (Brehmer et al. 2004; Langer et al. 1992). GrpE binds to the N-terminal domain of DnaK and catalyzes the exchange of ADP to ATP, thereby opening the lid of the C-terminus resulting in substrate release (Figure 2.5). If substrate protein has not reached its native state, further rounds of binding and release are

repeated with DnaK. Nearly 20% of all *E. coli* proteins are capable of reaching their native state upon DnaK interaction, however 10% require further folding assistance (Figure 2.3 A) (Teter et al. 1999). Non-native proteins that are unable to reach the native state with the Hsp70 system are transferred to a specialized set of chaperones termed chaperonins, which provide a sequestered environment for folding.

2.2.1.3 Chaperonins

The term chaperonins refers to a class of ubiquitous, essential chaperone complexes. Chaperonins are large, oligomeric proteins (~0.8-1 MDa) consisting of two ring structures stacked back-to-back (Hemmingsen et al. 1988). They exhibit ATPase activity and bind non-native proteins or stress-denatured proteins through hydrophobic interactions. Two families of chaperonins exist in nature as determined by mechanistic and sequence variations: group I and group II (Horwich et al. 2007). Group I chaperonins occur in bacteria (GroEL), chloroplasts (ch-Cpn60), and mitochondria (mt-Hsp60). These chaperonins characteristically contain two seven-membered rings and operate with the lid-like cochaperone Hsp10 (GroES in bacteria, ch-Cpn10 and ch-Cpn20 in chloroplasts, and mt-Hsp10 in mitochondria). In contrast, group II chaperonins occur in the cytosol of archaea (thermosome) and eukaryotes (TRiC/CCT) and are organized as either two eight or nine-membered rings consisting of homooligomeric or heterooligomeric homologous subunits (Frydman 2001). The group II eukaryotic TRiC chaperonin is not stress inducible, in contrast to the group I and thermosome chaperonins (Horwich et al. 2007; Klunker et al. 2003). Furthermore, group II chaperonins do not require a separate cochaperone for their function. Instead, they contain α-helical protrusions that extend from the apical domain of the rings, which behave as a built-in-lid. Both groups of chaperonins exhibit ATP hydrolysis and fold substrate through rounds of encapsulation in an isolated environment followed by release of folded substrate (Gatenby 1992).

2.2.1.3.1 *E. coli* chaperonin system

GroEL from *E. coli* is the most extensively studied and characterized chaperonin. At least 13 essential proteins absolutely require this chaperonin for folding, therefore proper GroEL function is essential for cell viability under all growth conditions (Kerner et al. 2005). GroEL operates downstream of TF and the Hsp70 (DnaK) chaperone system, and interacts with approximately 10% (250 proteins out of 2400 proteins) of the total proteome in the *E. coli* cytosol (Figure 2.3 A). Generally, substrates of GroEL are in the range of 20-50 kDa in

size with complex α/β domain topologies; nevertheless proteins of up to 150 kDa have also been identified as GroEL substrates. Furthermore, proteins that interact with GroEL during folding can be divided into three specific substrate classes (Kerner et al. 2005). Class I substrates maintain the ability to spontaneously fold when diluted from denaturant with low aggregation propensity and exhibit only a slight requirement for GroEL for folding. Class II substrates have higher propensity for aggregation upon dilution from denaturant and chaperones are necessary for folding, but this can be accomplished by either GroEL/ES or the DnaK chaperone system. Class I and class II substrates are generally more abundant proteins with only partial dependence on GroEL for folding. Approximately 85 proteins have been characterized as class III GroEL substrate proteins. Proteins of this class are generally of low abundance in the cytosol and strictly dependent on GroEL/ES for folding. Although these substrates bind DnaK, the DnaK chaperone system is not able to mediate folding. DnaK binds class III substrates and keeps them in a folding competent state (Kerner et al. 2005). The mechanism of GroEL/ES-mediated substrate folding is distinct from the Hsp70 chaperone system in that substrate folding occurs inside an enclosed, globally confined environment.
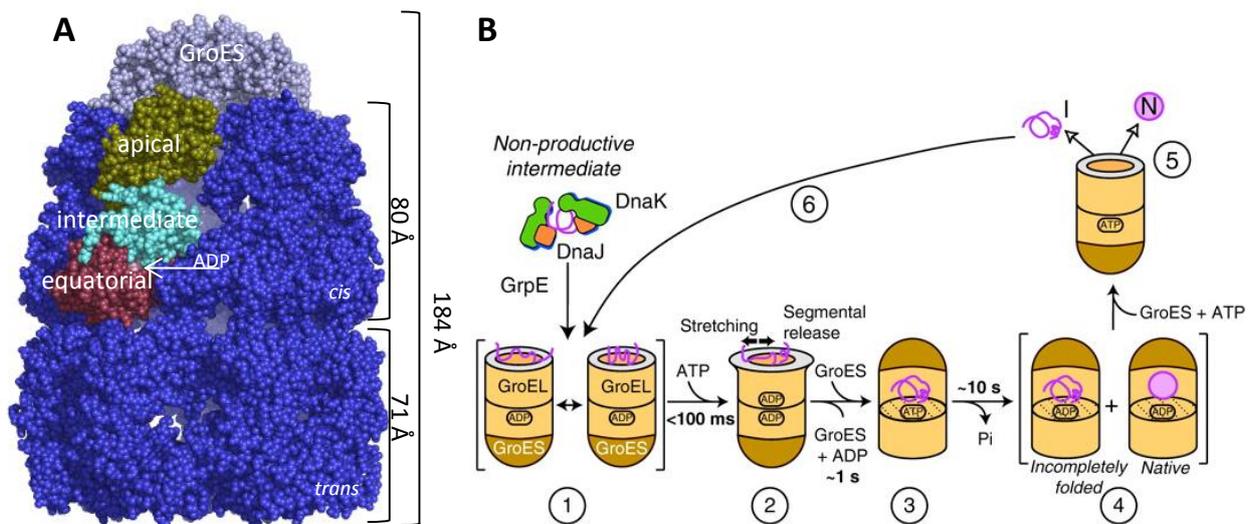
Figure 2.6 Structure of GroEL/ES and model of GroEL/ES-mediated substrate folding.
(A) Crystal structure of GroEL in complex with GroES and seven ADP bound. Dimensions of the *cis* and *trans*-ring and of the entire complex are indicated. Equatorial, intermediate, and apical domains are labeled for one subunit in the *cis*-ring as well as the position of one bound ADP. (PDB 1AON (Xu et al. 1997)). (B) Model of GroEL/ES-mediated substrate folding. 1. Substrate protein is delivered to *cis*-ring of GroEL by DnaK system. 2. Apical domain movement in an ATP-dependent manner causes bound substrate stretching. 3. Substrate is encapsulated by GroES in the cavity. 4. Substrate folding in the chaperonin cavity and ATP hydrolysis. 5. Substrate is released into solution either in the native state, or an incompletely folded intermediate state. 6. Incompletely refolded substrate is rebound to GroEL for further cycles of refolding. N: native, I: folding intermediate. Adapted from: (Vabulas et al. 2010)

GroEL is an ~800 kDa tetradecameric complex consisting of ~58 kDa identical subunits. Each subunit has an equatorial domain, an intermediate domain, and an apical domain (Braig et al. 1994). The equatorial domain forms the interface between the two rings and is also the site of nucleotide binding. The intermediate domain is located between the

equatorial and apical domains and facilitates allosteric communication between these two domains. The flexible apical domain exposes hydrophobic residues and is the binding site for both substrate proteins and GroES. The lid-like cochaperone of GroEL, GroES, is a heptameric ring composed of ~10 kDa identical subunits (Figure 2.6 A) (Xu et al. 1997). The mechanism of GroEL/ES action has been investigated in depth. GroES binding occurs in an ATP-regulated manner and induces conformational shifts in GroEL. Furthermore, binding of GroES to GroEL causes an expansion of the folding cavity and conversion of the cavity from hydrophobic to hydrophilic with an overall net negative charge (Chen et al. 1994). Substrate protein is encapsulated in the GroEL/ES cavity for ~10 s at 25°C, which is the amount of time necessary for ATP hydrolysis in the *cis*-ring (GroES bound ring). Binding of seven ATPs to the opposite *trans*-ring causes release of ADP, dissociation of GroES, and release of substrate. Protein that has not reached its final native state can rebind GroEL for further a round of encapsulation and folding (Figure 2.6 B). Intriguingly, GroEL can also support the folding of substrates too large to be enclosed in the cavity by GroES, thereby indicating that large substrates may be folded by the *trans*-ring of GroEL without encapsulation (Chaudhuri et al. 2001).

The GroEL/ES cavity creates an environment where substrate protein is sequestered from the highly crowded cellular surroundings (Martin 2004). Enclosure of substrate in the cavity reduces the possibility of off-pathway aberrant interactions and causes a smoothing of the energy landscape (Figure 2.2 B) (Brinker et al. 2001). Two divergent theories exist, depicting the mechanism of the GroEL/ES cavity as either passive or active. It has been suggested that the cavity is passive and behaves as an Anfinsen cage signifying that the cavity acts strictly by providing a space for the protein to fold, but folding is determined by the amino acid sequence and intrinsic thermodynamics of the peptide chain (Ellis 1994). In this model the GroEL/ES cavity functions solely by protecting against the crowded cellular environment and thereby impeding aggregation (Apetri and Horwich 2008; Horwich et al. 2009). The passive model assumes that protein folding of the intermediates in the GroEL cage proceeds along a productive folding pathway. However, this model is limited in that it does not address the possibility that during folding, kinetically trapped intermediate states might form that could not further access the native state.

Alternatively, increasingly more data support the hypothesis that the GroEL/ES cage can actively accelerate folding of the sequestered protein (Brinker et al. 2001; Chakraborty et al. 2010; Tang et al. 2006). Folding rate acceleration is thought to arise from confinement resulting in the modification of the energy landscape of folding or by reversal of the

formation of kinetically trapped intermediates through iterative annealing (Lin et al. 2008). It has been demonstrated by a mutant of the model protein maltose-binding protein (MBP) that encapsulation in the GroEL/ES cavity resulted in accelerated refolding rate compared to the spontaneous rate of refolding; furthermore, the folding rate acceleration was not due to prevention of irreversible of aggregates (Chakraborty et al. 2010). The polar nature as well as the size of the folding cavity is extremely important for this process of potentially catalyzing structural annealing as revealed by mutational analyses (Tang et al. 2008). Initial binding and continual rebinding of substrate to GroEL has been shown to induce forced unfolding (iterative annealing) through conformational rearrangement of GroEL upon ATP or GroES binding, thus remodeling kinetically trapped intermediates (Lin et al. 2008; Sharma et al. 2008). It seems likely that the GroEL/ES cage is not only an aggregation prevention cage but may also modulate the rate of folding by providing a sterically confined, polar environment where kinetically trapped intermediates are destabilized.

2.2.1.3.2 Cyanobacterial chaperonin

Detailed studies on the function/mechanism of chaperonin from cyanobacteria have been hindered due to its high degree of instability and tendency to disassociate upon isolation outside of the cell (Kovacs et al. 2001). Interestingly, sequences of complete cyanobacteria genomes have revealed the presence of at least two *groEL* gene homologues (designated here as Cyan-GroEL1 and Cyan-GroEL2). Cyanobacterial genomes containing two *cyan-groEL* homolog genes have only one *cyan-groES* gene in the genome. However, when three *cyan-groEL* genes are present there are correspondingly two *cyan-groES* genes present (Lund 2009). Note that the occurrence of three *groEL* homologues in the cyanobacterial genome is rare, therefore the former instance will be primarily discussed. The *cyan-groEL1* gene is located on a bicistronic operon with *cyan-groES* (Webb et al. 1990), whereas the operon for *cyan-groEL2* is monocistronic with differential genomic localization (Lehel et al. 1993). Cyan-GroEL1 and Cyan-GroEL2 share up to approximately 65% sequence identity with each other and *E. coli* GroEL. Examination of the amino acid sequence reveals intriguing differences, especially upon closer inspection of the C-terminal tails (Figure 2.7). *E. coli* GroEL displays distinct GGM repeats in the C-terminal tail portion of its sequence, which have been indicated to be critical for substrate folding (Tang et al. 2006). These GGM C-terminal tail repeats are also present in Cyan-GroEL2, but not in Cyan-GroEL1. Cyan-GroEL1 may have Gly or Met in the C-terminal tail, but it is not present as a repeated segment (Tanaka et al. 1997).

The cyanobacterial GroEL homologues are both induced upon heat shock; however Cyan-GroEL1 displays higher fold induction levels than Cyan-GroEL2. The ability of Cyan-GroEL homologues to functionally replace *E. coli* GroEL has shown that Cyan-GroEL1 but not Cyan-GroEL2 can overcome the deficiencies in *E. coli* GroEL defective cells (Furuki et al. 1996). Furthermore, light also plays an important role in the efficiency of heat shock induction; indicating an important role of light for activation of the cyanobacterial heat shock response (Asadulghani et al. 2003). Importantly, *Anabaena* sp. PCC7120 photosynthetically active cells upon heat shock localized Cyan-GroEL to the thylakoid membrane and carboxysomes (Jager and Bergman 1990; Kovacs et al. 1994). This indicates that GroEL from cyanobacteria is most likely important for the folding of a variety of different proteins and in particular Rubisco since it is was found to localize to the carboxysomes upon heat shock.
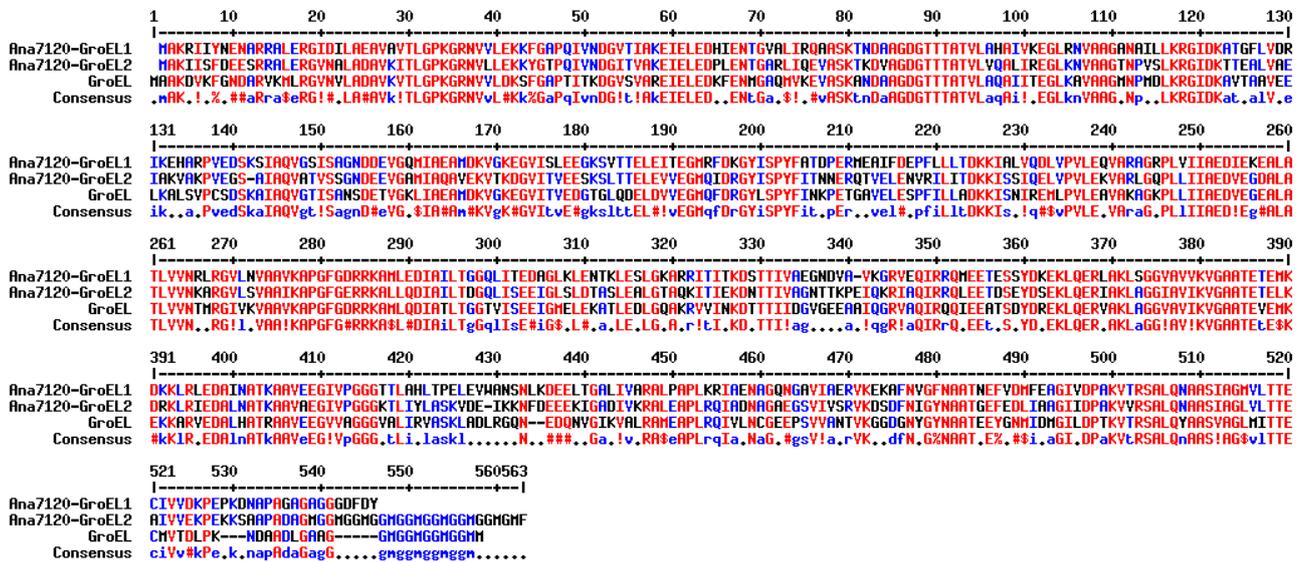
Figure 2.7 Sequence alignment of Cyan-GroEL1, Cyan-GroEL2, and *E. coli* GroEL.
Amino acid alignment of GroEL1 and GroEL2 from *Anabaena* sp. PCC7120 (Ana7120-GroEL1 and Ana7120-GroEL2) and GroEL from *E. coli* (GroEL). Completely conserved residues are shown in red and residues found in two of the sequences are shown in blue. Black indicates residues which are not conserved. The GGM repeat motif at the ultimate C-terminus is found in GroEL as well as Ana7120-GroEL2, but not in Ana7120-GroEL1. Alignment created using MultAlin.

The regulation of the cyanobacterial heat shock response is still unclear. There has, however, been the identification of a CIRCE (Controlling Inverted Repeat of Chaperone Expression) element either at the transcriptional start site of one or both *cyan-groEL* genes (Asadulghani et al. 2003; Kovacs et al. 2001; Rajaram et al. 2001). The CIRCE-HrcA system has previously been shown to be involved in the regulation of genes involved in the heat shock response in other bacterial species (Zuber and Schumann 1994). In this system under non-heat shock conditions, the dimeric HrcA protein (80 kDa) binds to the 9 bp inverted repeat sequence, impeding transcription of the genes downstream of the CIRCE element. Upon heat shock, the HrcA protein is denatured and the downstream heat shock genes are

translated. Accumulation of heat shock proteins enables the refolding of HrcA, which can once again bind to the CIRCE sequence (Liu et al. 2005). Although the CIRCE element is present upstream of many *cyan-groEL* genes, it may not be the only regulator of the heat shock response or of the transcription of *groEL* genes in cyanobacteria. Mutational studies recently demonstrated that additional heat-regulatory elements, H-box and K-box, are present in the *Anabaena* L-31 genome (Rajaram and Apte 2010). The H-box element has been shown to be important for negative heat regulation. The K-box has been identified in *groESL* promoter regions in all sequenced *Anabaena* strains and is involved in positive light-regulation of the heat shock response in these cyanobacterial species. It is very likely that there are other transcriptional regulators of chaperonin expression in cyanobacteria due to the differential induction of the *cyan-groEL* genes upon heat shock.

Detailed studies of Cyan-GroEL are primarily limited to studies *in situ* due to the inherent instability of this protein upon purification (Lehel et al. 1992). Cyan-GroEL has a high tendency to dissociate into monomeric or dimeric forms. Recombinant expression of Cyan-GroEL from *E. coli* has demonstrated the difficulty of purification of cyanobacterial GroEL and caution must be practiced upon interpretation of the results (Huq et al. 2010). Special buffer conditions must be implemented along with high protein concentrations to ensure tetradecameric stability, which can be different for different species. Importantly, once the Cyan-GroEL tetradecamer dissociates, the re-oligomerization of the monomers has been so far not possible (Lehel et al. 1992).

There are various theories for the presence of multiple chaperonin genes in an organism and how this phenomenon arose (Lund 2009). Multiple copies of genes in an organism can occur from either gene duplication or horizontal gene transfer. In either case, the additional copy enables flexibility in gene function, which could lead to the divergent evolution of the genes; one gene can further carry out its previous function while the other is free to evolve new roles or functions. In cyanobacteria it seems likely that the two GroEL proteins have slightly different roles in the cell due to the above descriptions of differential expression upon heat shock and the ability of Cyan-GroEL1 but not Cyan-GroEL2 to complement *E. coli* GroEL. It is enticing to speculate that the two GroEL proteins evolved to accommodate photosynthesis in cyanobacteria (Lund 2009); one may be important for general housekeeping in the cell, while the other may be highly involved in and necessary for Rubisco folding.

2.2.1.3.3 Chloroplast chaperonin

The occurrence of multiple group I chaperonin genes in an organism is not limited to bacteria, as chloroplast chaperonin possesses two homologous nuclear encoded α and β components (Hemmingsen and Ellis 1986). Moreover, it has been established that the chloroplast chaperonin (ch-Cpn60) is a heterooligomeric complex consisting of seven α and seven β subunits as determined by SDS-PAGE quantification (Hemmingsen et al. 1988; Nishio et al. 1999). This is confounded by the existence of multiple homologues of each α and β subunit as well as several cochaperone homologues within some species e.g., *Arabidopsis thaliana*. The α and β subunits within a species are homologous but divergent, displaying approximately 50% sequence identity to each other and *E. coli* GroEL (Martel et al. 1990). Ch-Cpn60 was first discovered in the early 1980's during isolation of the enzyme ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) and was termed the Rubisco subunit binding protein (Barraclough and Ellis 1980; Lorimer 2001). However, ch-Cpn60 is not only implicated in Rubisco folding and assembly, but it has also been indicated in association with a variety of proteins in the chloroplast, including imported proteins as well as chloroplast encoded proteins (Lubben et al. 1989). Comparison of the GroEL protein sequence with ch-Cpn60 protein sequence showed that these two proteins were evolutionary homologues with about 46% sequence identity, leading to the assumption that they would also exhibit similar functional behavior in the cell (Hemmingsen et al. 1988).

Chaperonins located in the chloroplast are nuclear encoded and display high sequence homology to Cyan-GroEL2. From phylogenetic studies, it is predicted that during green algae and higher plant evolution, cyanobacteria were taken up by eukaryotic cells through endosymbiosis (Gould et al. 2008). The chloroplast could then be formed through morphogenesis of the cyanobacteria in these early endosymbiotic organisms. Additionally, gene flow from the symbiont to the host resulted in the transfer of the *groEL2* and *groES* genes to the host nuclear genome correlating with development of protein transfer mechanisms in the host. Gene duplication events of the *groEL2* gene most likely produced the homologous α and β subunits now present in photosynthetic eukaryotic organisms (Wastl et al. 1999).

Ch-Cpn60 exhibits interesting features which distinguish it from GroEL, and could indicate possible deviations in function (Levy-Rimler et al. 2002). As mentioned above, ch-Cpn60 is a tetradecamer group I chaperonin similar to GroEL, is heat-stress induced (Thompson et al. 1995), and also correspondingly interacts with cochaperone in an ATP-dependent manner (Figure 2.8) (Viitanen et al. 1995); however the tetradecamer of ch-Cpn60

contains two different subunits whose arrangement in the assembled oligomer is still unknown. It has been demonstrated, by recombinant expression, that the ch-Cpn60α subunits are not capable of assembly to the tetradecameric complexes and accordingly maintain a monomeric form (Dickson et al. 2000). Ch-Cpn60β subunits have the propensity to form homo-tetradecameric complexes, however these homooligomers function differently in regard to cochaperone binding compared to ch-Cpn60αβ heterooligomers aiding in the assumption that the heterooligomer is the biologically relevant configuration (Nishio et al. 1999). Additionally, in contrast to GroEL, ch-Cpn60 is less stable and less readily purified especially in the presence of ATP (Hemmingsen and Ellis 1986; Lissin 1995; Musgrove et al. 1987), impeding mechanistic studies.
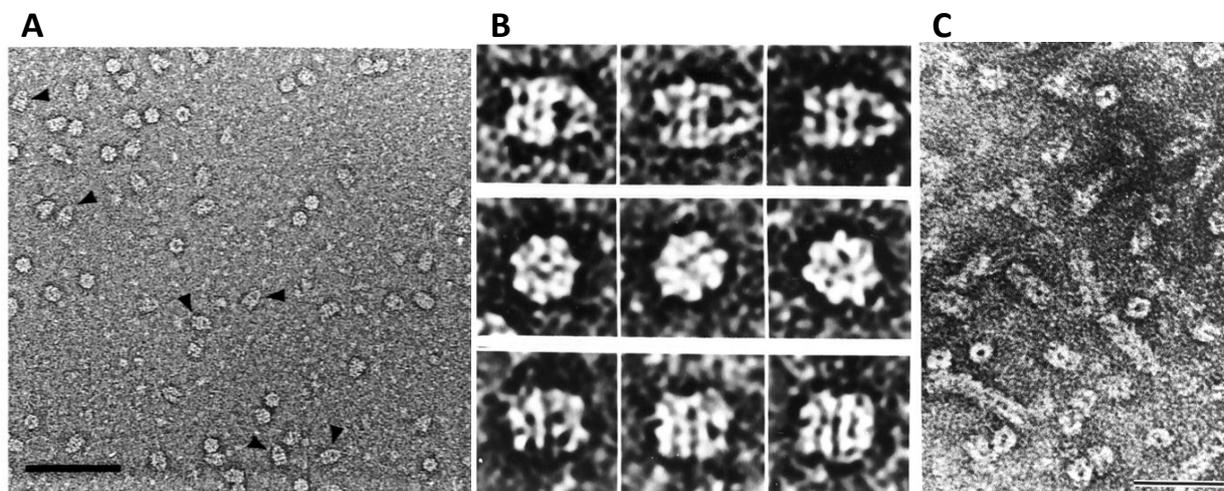
Figure 2.8 EM structure of ch-Cpn60/ch-Cpn21 complex and ch-Cpn21.
(A) Electron micrograph of pea ch-Cpn60 from chloroplasts incubated with recombinant spinach ch-Cph21 in the presence of ADP. Bullet-shaped (ch-Cpn60 with ch-Cpn21) complexes are indicated with arrows. Scale bar measures 100 nm. (B) Enlargement of bullet-shaped complexes (ch-Cpn60 with ch-Cpn21) top and middle and ch-Cpn60 alone bottom. A and B adapted from: (Viitanen et al. 1995) (C) Electron micrograph of recombinant spinach ch-Cpn21 alone, scale bar is 20 nm. Adapted from: (Baneyx et al. 1995)

There are also important differences in the cochaperones of GroEL and ch-Cpn60; not only is there at least one 10 kDa cochaperone (ch-Cpn10) present in some species, but there are also up to two 20 kDa ch-Cpn60 cochaperones (ch-Cpn21) which have been identified (Bertsch et al. 1992; Koumoto et al. 2001). The ch-Cpn21 cochaperones seemingly evolved from an endoduplication event resulting in the doubling in size from 10 kDa to 20 kDa (Bertsch et al. 1992). Ch-Cpn10, as revealed from *in vitro* studies, presumably forms a heptamer in solution (Sharkia et al. 2003). The ch-Cpn21 consists of two tandem GroES-like domains, separated by a short linker. Furthermore, the biologically relevant oligomeric configuration of the chloroplast cochaperones is still unknown, but EM pictures of ch-Cpn21 shows a ring-like complex similar to GroES when bound to ch-Cpn60 (Figure 2.8) (Baneyx et al. 1995; Viitanen et al. 1995). *In vitro* studies have indicated that ch-Cpn21 most likely forms

tetramers (Koumoto et al. 1999), which questions the manner of association of a tetrameric complex to a heptameric ring of ch-Cpn60; various theoretical models of ch-Cpn21 assembly are shown in Figure 2.9 (Weiss et al. 2009). Interestingly, as analyzed by mutational experiments, both domains of the ch-Cpn21 are necessary for its function, however, the N-terminal GroES-like domain appears to have a more prominent role in activity than the C-terminal GroES-like domain (Bertsch and Soll 1995; Bonshtien et al. 2007). Presumably, ch-Cpn21 could have evolved in plants as an adaptation to the various paralogs of chloroplast chaperonin (Boston et al. 1996).
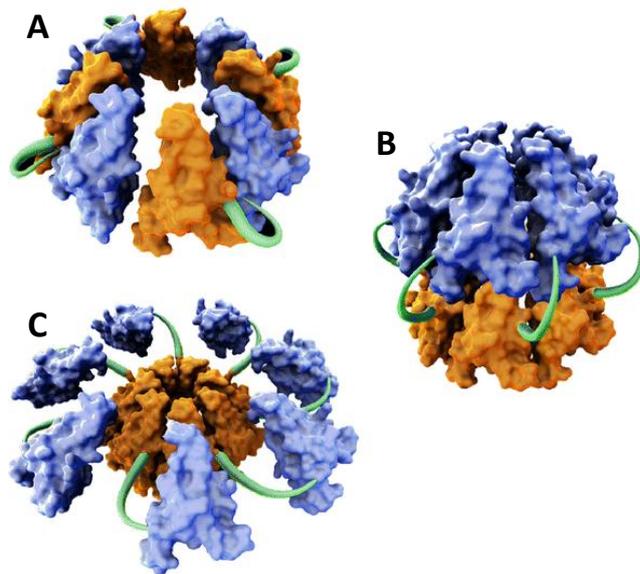
Figure 2.9 Model of co chaperone ch-Cpn21 complex assembly.
Models based on the subunit structure of *Mycobacterium tuberculosis* Cpn10 (PDB 1HX5_2). (A) Tetrameric model of ch-Cpn21, N-terminal and C-terminal would bind to the chaperonin. (B) Heptameric model, one domain would bind directly to the chaperonin while the adjacent domain would stack on top of the bound domain. (C) Heptameric model, one domain would bind directly to chaperonin while the adjacent domain extents outwards possibly for interactions with other proteins. Adapted from: (Weiss et al. 2009)

The complexity of the ch-Cpn60 system is best demonstrated in *A. thaliana* (Hill and Hemmingsen 2001). The *A. thaliana* genome encodes two ch-Cpn60α and four ch-Cpn60β isotypes. Ch-Cpn60α1 and ch-Cpn60α2 share approximately 60% sequence identity, while the three of the four ch-Cpn60β subunits share up to 90% sequence identity. The fourth ch-Cpn60β retains 60% sequence identity to the other three. In addition, there are also several homologues of cochaperone: one 21 kDa (ch-Cpn21) cochaperone and two 10 kDa (ch-Cpn10) cochaperones. These cochaperones have also been identified in pea (Schlicher and Soll 1996) and spinach (Bertsch and Soll 1995). Additionally, two homologues of ch-Cpn21 have been identified in *Chlamydomonas reinhardtii*, ch-Cpn21 and ch-Cpn23 (Weiss et al. 2009). The cochaperone homologues exhibit 50% sequence identity to each other; furthermore ch-Cpn21 displays 50% identity within its own GroES-like domains. The identification of several genes for ch-Cpn60 and cochaperone raises many questions that remain to be elucidated. For instance, it is unknown whether there are functional/mechanistic differences between these chaperonin complexes, or if the homologues are advantageous for substrate specificity or differential substrate affinity. Recent studies focused on substrate

specificity of the ch-Cpn60α and β paralogs in *A. thaliana* have demonstrated that the low abundant ch-Cpn60β4 is crucially necessary for the folding of the NdhH protein, which is a protein associated with the NADH dehydrogenase-like (NDH) complex in chloroplasts (Peng et al. 2011). This result suggests that the different ch-Cpn60 subunits display different specificities for certain substrates, which could explain the occurrence of more than one homolog of chaperonin subunits in the chloroplast.

A proteomics study analyzing *A. thaliana* stroma was recently conducted, which included Native-PAGE electrophoresis coupled with mass spectrometry (Peltier et al. 2006). In this analysis the α and β paralogs ch-Cpn60α2, ch-Cpn60β1, ch-Cpn60β2, and ch-Cpn60β3 were identified in the isolated tetradecameric complex; the presence of ch-Cpn60α1 and ch-Cpn60β4 in the complex could not be detected. Interestingly, cochaperones were also not detected in the complex; however ch-Cpn21 and ch-Cpn10 were instead isolated alone in complexes of 150-170 kDa. The cochaperone ch-Cpn21 was found to be the most highly expressed chaperonin component in the chloroplast (Peltier et al. 2006). The dissociation of cochaperone from ch-Cpn60 could possibly be explained by the absence of additional nucleotide during the isolation process. This does not, however, account for the exceedingly high protein levels of cochaperone in the chloroplast in relation to chaperonin levels, which could signify possible additional, divergent roles of this protein (Weiss et al. 2009). Indeed, ch-Cpn21 has been indicated in seed development for *Vitis vinifera*, *Nicotiana benthamiana*, and *Lycopersicon esculentum* (Hanania et al. 2007). It is highly probable that the chloroplast chaperonin system mechanistically functions similar to GroEL/ES, and is necessary for the proper folding of a variety of chloroplast proteins, most importantly Rubisco.

2.3   Photosynthesis

Life on earth is dependent on photosynthesis. Photosynthesis is the transfer of solar energy into physiologically usable chemical energy converting an electron donor, usually water, and atmospheric CO_2 into carbohydrates (Horton et al. 2002). Organisms capable of photosynthesis, phototrophs, include prokaryotes: e.g., cyanobacteria and purple bacteria as well as eukaryotes: diatoms, algae, and plants. During oxygenic photosynthesis, which is found in eukaryotes and some prokaryotes, water is used as the electron donor resulting in the release of O_2.

In eukaryotic phototrophs, photosynthesis occurs in specialized organelles called chloroplasts (Figure 2.10). This organelle has a special double outer membrane that allows the diffusion of CO_2 (Horton et al. 2002). The chloroplast arose from the engulfment of ancient

cyanobacteria into eukaryotic cells (Xiong et al. 2009). The thylakoid membrane, located on the inside of chloroplasts, is highly folded and usually arranged as flattened-stacked vesicles called grana. The machinery necessary for the capture of solar energy and conversion to chemical energy (light-dependent reactions) is embedded in the thylakoid membrane, including: photosystems I and II, electron carrier chains, chloroplast ATP-synthetase, light harvesting complexes, as well as other complexes. The aqueous phase inside the thylakoid membrane is the lumen, while the aqueous phase outside the thylakoid membrane is the stroma. The stroma contains the soluble proteins and molecules necessary for the reduction of atmospheric carbon into organic carbon (light-independent reactions).
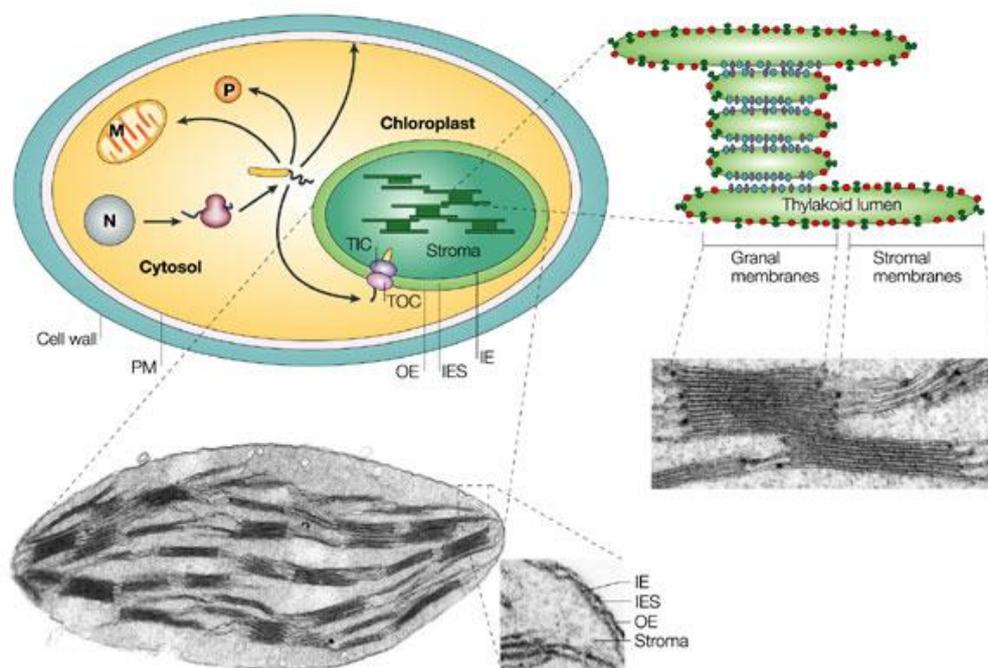
Figure 2.10 Structure of the chloroplast.
Schematic diagram of a plant cell with zoom-in of an electron micrograph of a chloroplast from *Pisum sativum.* The chloroplast contains three membrane systems: the outer envelope (OE), the inner envelope (IE), and the thylakoid membrane. The chloroplast also contains three aqueous phases: the thylakoid lumen, the stroma, and the inter envelope space (IES). Some nuclear encoded proteins must be imported to their correct cellular compartment, which is accomplished by signal transit peptides shown in yellow. In the case of nuclear encoded chloroplast proteins this occurs through two translocons: TOC and TIC, located in the OE and IE, respectively. The right panel illustrates the thylakoid organization. Granal membranes (segments of the thylakoid membrane that do not contact the stroma) are enriched with photosystem II (blue) and the cytochrome b_6f complex (purple) while stromal membranes (segments of the thylakoid membrane that do contact the stroma) are enriched with photosystem I (red) and ATP synthetase (green). N: nucleus, M: mitochondria, P: peroxisomes, PM: plasma membrane. Adapted from: (Soll and Schleiff 2004)

Prokaryotic phototrophs, for instance cyanobacteria, also carry out oxygenic photosynthesis, thus releasing O_2 as the byproduct. Cyanobacteria, the predecessors of contemporary eukaryotic chloroplasts, do not contain a separate, specialized organelle for photosynthesis. Instead, the light-dependent reaction of photosynthesis is localized on invaginations of the cell membrane called the thylakoid membrane (Figure 2.11). Light is harvested through the antennae of photosystem I and II, thylakoid bound phycobilisomes,

where pigment molecules such as phycocyanin are attached. Conversely, anoxygenic photosynthetic species e.g., purple sulfur bacteria, usually only contain one photosystem and use hydrogen sulfide and not water as an electron donor resulting in the release of sulfur from the organism. Photosystems occurring alone do not have the oxidative strength to remove electrons from water resulting in the necessity of these organisms to use alternative electron donors for photosynthesis (Leslie 2009; Tavano and Donohue 2006). Furthermore, anoxygenic photosynthetic species contain different pigment molecules which are termed bacteriochlorophyll.
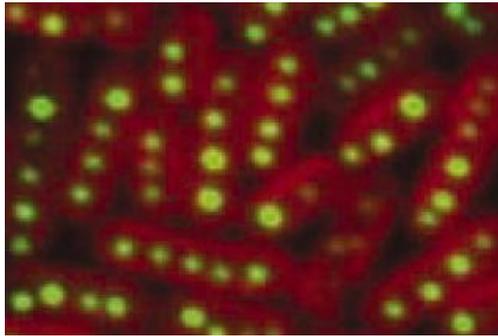
Figure 2.11 Cyanobacteria cell structure. Immunofluorescence microscopy of *Synechococcus* sp. PCC7942 cells. Thylakoid membrane is shown in red and location of Rubisco in the carboxysomes is shown in green.
Adapted from: (Savage et al. 2010)

2.3.1 Light-dependent reactions

As mentioned above, the complexes involved in the light-dependent reactions are located in the thylakoid membrane. During the light-dependent reactions solar energy is converted to ATP and NADPH, which are necessary as energy sources and reductive equivalents during the light-independent reactions. Photons of light are captured by chlorophyll or other accessory pigment molecules. Chlorophyll absorbs light through a network of conjugated double bonds; upon light absorption a ground-state electron is energized to a higher molecular state. Some pigments are located within two specific complexes in conjunction with other proteins and molecules, which are collectively called photosystems.

Oxygenic phototrophs possess two photosystems, I and II. The photosystems contain a reaction center consisting of a chlorophyll pair. The reaction center of photosystem I absorbs light maximally at 700 nm, correspondingly this reactive center is named P700. Whereas, the reaction center of photosystem II absorbs light maximally at 680 nm and is accordingly named P680 (Horton et al. 2002). Water is only able to be used as an electron source when both photosystems are present. Photosystem II becomes a strong oxidant upon excitation, which can then oxidize the oxygen atom of water releasing elementary oxygen and H^+ as a consequence (Leslie 2009). (The recently solved high resolution crystal structure of photosystem II gives detailed insight to the manganese cluster where water is oxidized

(Umena et al. 2011)). The Z-scheme schematically illustrates the movement of electrons and reduction potentials of the electron carrier components, which create the high energy molecules ATP and NADPH (Figure 2.12 A). Light energy, which is channeled from antennae pigments and light harvesting complexes and further passed to the reaction center, is used for the electron flow from the photosystems through protein transfer chains. In this process NADP is reduced to NADPH. The transfer of electrons is not only necessary for NADP reduction, but also results in the translocation of protons through the thylakoid membrane into the lumen (Figure 2.12 B). The proton gradient created between the lumen and stroma is then utilized by the chloroplast ATP-synthetase for the photophosphorylation of ADP to ATP (Graciet et al. 2004).

Figure 2.12 Light-dependent reactions in photosynthesis.
(A) Z-scheme of electron transport from H_2O to NADP in photosynthesis. Antenna complexes (colored spheres) absorb photons of light and the absorbed energy is transferred to PSII (P680) and PSI (P700). The manganese center of P680 removes electrons from H_2O which are transferred to the P680 reaction center. The excited P680 transfers the electrons to pheophytin (Pheo) then plastoquinone (Q_A and Q_B), which are then passed over a chain of intermediate electron carriers (cytochrome b_6f complex (Cyt) and plastocyanin (PC)), and finally to the reaction center of P700. The electron flow from PSII to PSI results in H^+ being transferred to the lumen resulting in a proton-motive force which is used by ATP-synthetase for ATP formation. P700 is excited by light and electrons are transferred to ferredoxin (FD) where NADP is reduced to NADPH. Adapted from: (Orr and Govindjee 2010) (B) Diagram of electron and H^+ transfer in the chloroplast. Plastoquinones (PQ), blue arrows: electron transfer that results in reduction, orange arrows: electron transfer that results in oxidation. Electron transfer from H_2O to PSII and PSI results in the movement of H+ into the lumen as well as the reduction of NADP resulting in the formation of ATP and NADPH, respectively. ATP and NADPH are utilized by the light-independent reaction of photosynthesis for carbohydrate synthesis. Adapted from: (Peltier et al. 2010)
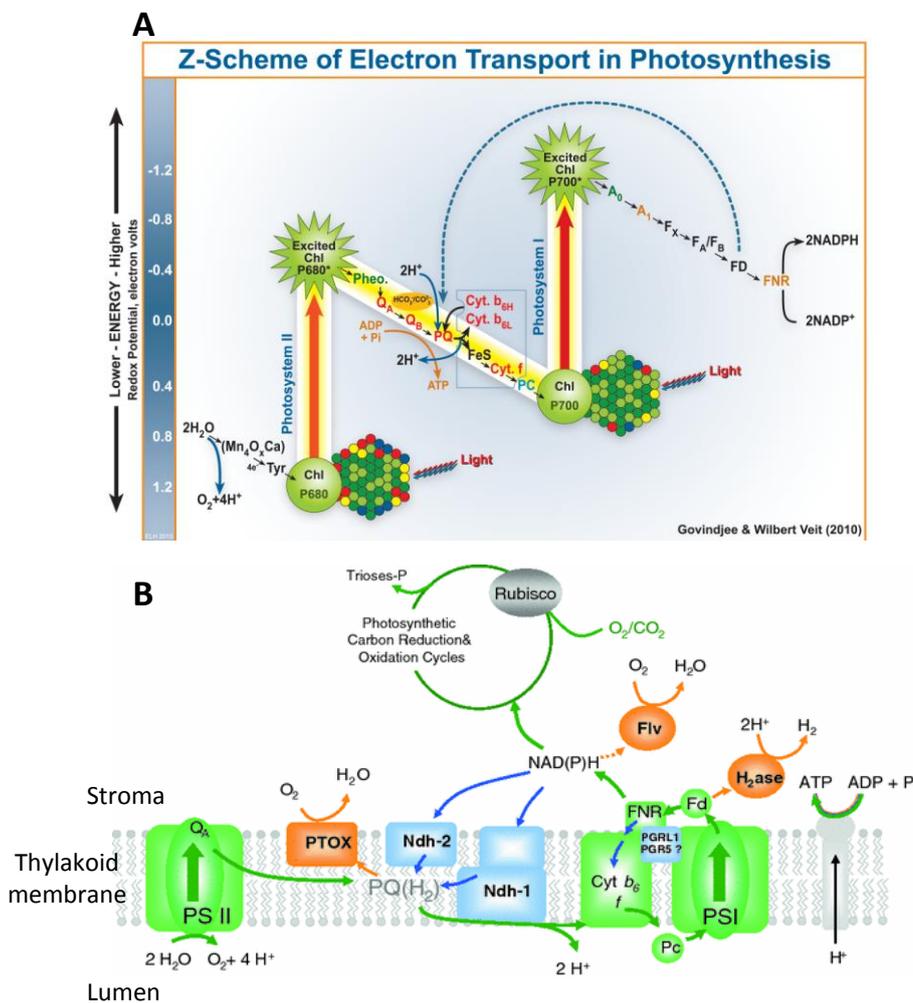
2.3.2 Light-independent reactions

The NADPH and ATP derived from the light-dependent reactions in the thylakoid membrane are necessary for carbon assimilation in the light-independent reactions through the Calvin-Benson-Bassham (CBB) cycle (Horton et al. 2002). The CBB cycle occurs in the stroma of the chloroplast of eukaryotes or cytosol of bacteria and is divided into three phases (Figure 2.13): carbon fixation, reduction (carbohydrate synthesis), and regeneration of the CO_2 acceptor (ribulose-1,5-bisphosphate (RuBP)). The carbon fixation reaction is catalyzed by the enzyme ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco). The five carbon sugar substrate of Rubisco, RuBP, is carboxylated, and cleavage of which results in the formation of two molecules of 3-phosphoglycerate (3-PGA). Subsequently, in the reduction phase, 3-PGA is ATP-dependently phosphorylated to 1,3-bisphosphoglycerate followed by NADPH-dependent reduction to form glyceraldehyde-3-phosphate (G3P). G3P is either used for the synthesis of carbohydrates or is kept in the CBB cycle and used for the regeneration of the carbon acceptor, RuBP. The regeneration phase also relies on ATP as an energy source and involves the formation of three- to seven-carbon sugars through transaldolase and transketolase reactions. Overall during the carboxylation of six five carbon sugar substrates, the CBB cycle consumes 18 ATP molecules and oxidizes 12 NADPH molecules for the formation of two G3P molecules and regeneration of 6 molecules of RuBP.
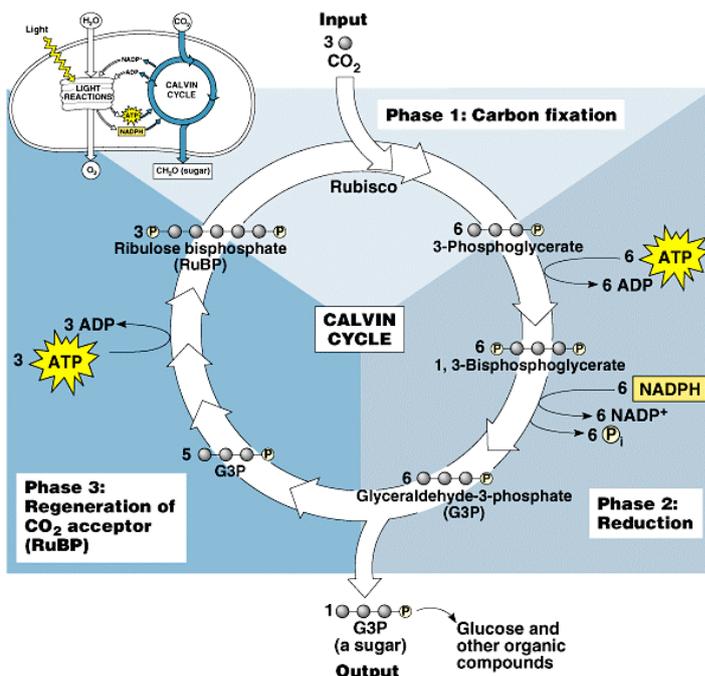
Figure 2.13 Calvin-Benson-Bassham cycle. Phase 1 carbon fixation: CO_2 is fixed by the action of Rubisco into the five carbon substrate molecule RuBP and the six carbon intermediate is rapidly cleaved into two molecules of 3-phosphoglycerate (3-PGA). Phase 2 reduction: 3-PGA is phosphorylated, mediated by ATP hydrolysis, and further reduced by oxidation of NADPH resulting in the production of glyceraldehyde-3-phosphate (G3P). G3P either enters Phase 3 regeneration of RuBP, or it is used for the generation of carbohydrates. Copyright © Pearson Education, Inc., publishing as Benjamin Cummings

2.3.3 Carbon concentrating mechanisms

During the initial phase of the light-independent reactions, Rubisco not only carboxylates RuBP, but Rubisco also oxygenates RuBP when O_2 is in the active site initiating a side reaction to photosynthesis called photorespiration (Bauwe 2011; Foyer et al. 2009; Somerville and Ogren 1982). The affinity of Rubisco for CO_2 (K_m: 9 µM) is greater than for O_2 (K_m: 350 µM), however the atmospheric concentration of O_2 is 500 times higher than CO_2 resulting in a significant amount of O_2 also diffusing to the Rubisco active site leading to oxygenation of RuBP with approximately every fourth Rubisco catalyzed reaction cycle (Peterhansel and Maurino 2011). Oxygenation of RuBP results in the formation of a five carbon intermediate, which is rapidly cleaved to one molecule of 3-PGA (three carbon molecule) and one molecule of 2-phosphoglycolate (PPG) (two carbon molecule). 3-PGA is utilized by the CBB cycle to regenerate the five carbon Rubisco substrate RuBP, whereas PPG must undergo modifications in the glycolate pathway involving reactions through the peroxisome and mitochondria before returning to the CBB cycle in a usable form i.e., 3-PGA (Figure 2.14). The recycling of PPG to 3PGA during photorespiration through two other cellular compartments leads to a net loss of a carbon atom and consumption of ATP.

The function of photorespiration is highly disputed. It is hypothesized to be either a wasteful side reaction of photosynthesis or a reaction that is necessary for optimal C3 plant survival. Photorespiration can be viewed as wasteful because it lowers photosynthetic efficiency and results in consumption of energy which would otherwise be used for the formation of carbohydrates (Peterhansel and Maurino 2011). However, photorespiration in C3 plants is potentially necessary for the proper utilization of soil nitrates, thereby allowing the CBB cycle to persist in an oxygen rich environment (Bloom et al. 2010). As indicated by mutations in C3 plants, inhibition of components of the photorespiratory pathway results in slowed plant growth (Somerville and Ogren 1982); furthermore oxygenation of RuBP may play a role in energy dissipation in plants (Osmond et al. 1997). Nevertheless, organisms e.g., plants, algae, and cyanobacteria have evolved carbon concentrating mechanisms (CCM) to reduce the oxygenation of RuBP by concentrating carbon at the active site of Rubisco and effectively increasing photosynthetic efficiency.

Figure 2.14 Photorespiration cycle.
Photorespiration is initiated by the oxygenation of RuBP by Rubisco resulting in a five carbon intermediate that is cleaved to the three carbon 3-phosphoglycerate (3-PGA) and the two carbon 2-phosphoglycolate (PPG). 3-PGA is utilized in the CBB cycle to regenerate substrate RuBP while PPG transfers to the peroxisome where it is further processed and transferred to the mitochondria before the atoms are in a usable form and return to the CBB cycle. Phosphoglycolate phosphatase (PGP), glycolate oxidase (GOX), serine-glyoxylate aminotransferase (SGT), glutamate-glyoxylate aminotransferase (GGT), glycine decarboxylase (GDC), serine hydroxymethyltransferase (SHMT), peroxisomal hydroxypyruvate reductase (HPR1), glycerate kinase (GLYK), catalase (CAT), hydroxypyruvate reductase (HPR2), peroxisomal malate dehydrogenase (pMDH), glutamine synthetase (GS2), ferredoxin-dependent glutamate synthase (GOGAT). Adapted from: (Bauwe 2011)

Terrestrial plants possess three photosynthetic pathways: the C3, the C4, and the Crassulacean acid metabolism (CAM) pathway. The C3 pathway (e.g., in rice, wheat, and barely) is the most common, and as described above atmospheric CO_2 is fixed in the CBB cycle by Rubisco, which is localized in the chloroplast of mesophyll cells (West-Eberhard et al. 2011). In contrast, the C4 pathway (e.g., maize, sugarcane, and sorghum) and the CAM pathway (e.g., pineapple, agaves, and orchids) display mechanisms of carbon concentrating within the plant either by physical or temporal separation, respectively. Carbon concentrating mechanisms are not only a feature of some plants, they are also observed in other species such as algae (pyrenoid) and cyanobacteria (carboxysomes).

2.3.3.1 C4 Plants

C4 photosynthesis arose from environmental conditions facilitating increased tendencies for Rubisco to behave as an oxygenase i.e., higher temperatures and reduced CO_2 levels in the atmosphere (Tipple and Pagani 2007). C4 photosynthetic plants exhibit

differential leaf anatomy in comparison to C3 plants (Figure 2.15). C4 plants are distinguished by the characteristic arrangement of the bundle sheath cells around the vascular tissue followed by the positioning of the mesophyll cells directly on the outer surface of the bundle sheath cells. This specialized cellular arrangement is known as Kranz anatomy. As described below, the Kranz anatomy functions to create an environment of CO_2 concentration for efficient Rubisco performance (Moroney and Ynalvez 2007). In this way, C4 plants spatially separate initial CO_2 fixation from the light-independent reaction of photosynthesis in order to vastly reduce the oxygenase activity of Rubisco.
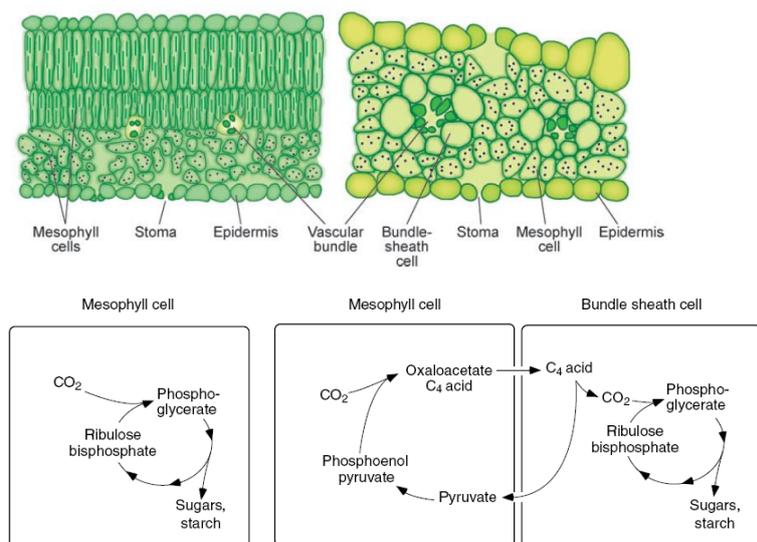
Figure 2.15 C3 and C4 leaf anatomy and photosynthetic mechanism.
Left upper panel indicates the C3 leaf anatomy and the left lower panel indicates the localization of the CBB cycle. Right upper panel indicates the C4 leaf anatomy depicting the characteristic Kranz anatomy and the right lower panel indicates the localization in the mesophyll cell of initial CO_2 fixation by PEP carboxylase followed by the four carbon intermediate transfer to the bundle sheath and further decarboxylation releasing CO_2 for utilization in the CBB cycle. Adapted from: (Tipple and Pagani 2007) and (Ehleringer and Cerling 2002)

The terminology of C3 and C4 photosynthesis is derived from the form of the carbon product after initial CO_2 fixation. During C3 photosynthesis two three carbon molecules are formed, while C4 photosynthesis initially forms a four carbon molecule. The mechanism of C3 photosynthesis, in which Rubisco in the mesophyll cells acts directly on atmospheric CO_2 resulting in three carbon intermediates, has been described above in section 2.3.2. In contrast, C4 photosynthetic plants use phosphoenolpyruvate (PEP) carboxylase to initially fix atmospheric CO_2 (hydrolyzed to HCO_3^- by carbonic anhydrase) to phosphoenolpyruvate in the mesophyll cytosol leading to the formation of a four carbon intermediate oxaloacetate (West-Eberhard et al. 2011). Oxaloacetate is either converted to malate or aspartate, which can be transported to the bundle sheath cells and undergo decarboxylation (Figure 2.15); thereby releasing CO_2 directly in the compartment where it can be fixed by Rubisco in the CBB cycle in the bundle sheath cells (Ehleringer and Cerling 2002; Kennedy 2000). C4 photosynthesis is capable of spatially isolating the CBB cycle in an environment with high concentrations of CO_2 leading to a vast reduction of oxygenase activity of Rubisco due to cellular confinement and isolation of the CBB cycle. As described previously, C3 plants may require photorespiration for the proper uptake of and reduction of nitrates. Conversely, C4

plants virtually eliminate photorespiration, this is however not detrimental to C4 plants since malic acid and NADH can be produced in the mesophyll cytoplasm, which is essential for nitrate reduction, in contrast to C3 plants that may rely on photorespiration for malic acid formation (Bloom et al. 2010).

2.3.3.2 Crassulacean acid metabolism (CAM) photosynthesis

CAM photosynthesis is an additional carbon concentrating adaptation of plants in hot, arid conditions (Keeley and Rundel 2003; Ting 1985). These plants have developed solutions to control water loss during CO_2 uptake as well as to avoid increased Rubisco oxygenase activity at the higher temperatures. Firstly, stomata of CAM plants are open and uptake CO_2 during the night to reduce water loss. Secondly, CAM photosynthesis is comparable to C4 photosynthesis, in that PEP carboxylase initially fixes CO_2 resulting in a four carbon intermediate and the initial step of carbon fixation is separated from the CBB cycle. However, formation of the four carbon intermediate occurs temporally and not spatially in relation to the CBB cycle (West-Eberhard et al. 2011). As illustrated in Figure 2.16, PEP carboxylase is active during the night in concert with the opening of the stomata, creating a pool of four carbon molecules, malic acid, in the vacuoles (Keeley and Rundel 2003). Furthermore, stoma closure during the day also reduces CO_2 loss upon decarboxylation of the four carbon intermediates leading to high concentrations of CO_2 available for Rubisco carboxylase activity in the CBB cycle during the day.
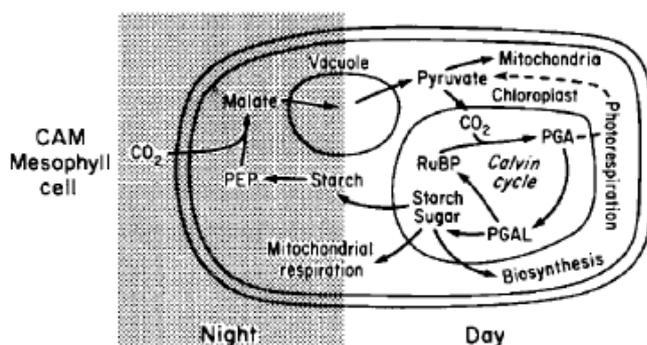
Figure 2.16 Mechanistic model of CAM photosynthesis.
CAM plants temporally separate atmospheric CO_2 fixation from the CBB cycle, allowing these two reactions to occur in the same cellular compartment, the mesophyll cell. Adapted from: (Keeley and Rundel 2003)

2.3.3.3 Pyrenoid

The pyrenoid is a non-membrane bound, proteinaceous compartment positioned in the stroma of chloroplasts in some algal species. This compartment is believed to have developed to facilitate the concentration of CO_2 to Rubisco in environments where CO_2 diffusion and concentration is low i.e., aquatic habitats. High concentrations of Rubisco are localized to the pyrenoid, and it therefore functions as the site of carbon fixation (Badger et al. 1980;

Moroney and Ynalvez 2007) (Figure 2.17A). Rubisco concentration within the pyrenoid is so high that it can be crystalline in appearance (Figure 2.17 B). The exact composition of the pyrenoid is still to be elucidated since this compartment is not easily isolated; nonetheless the pyrenoid is usually surrounded by starch deposits especially when algae are grown under extremely low CO_2 conditions (Ramazanov et al. 1994).
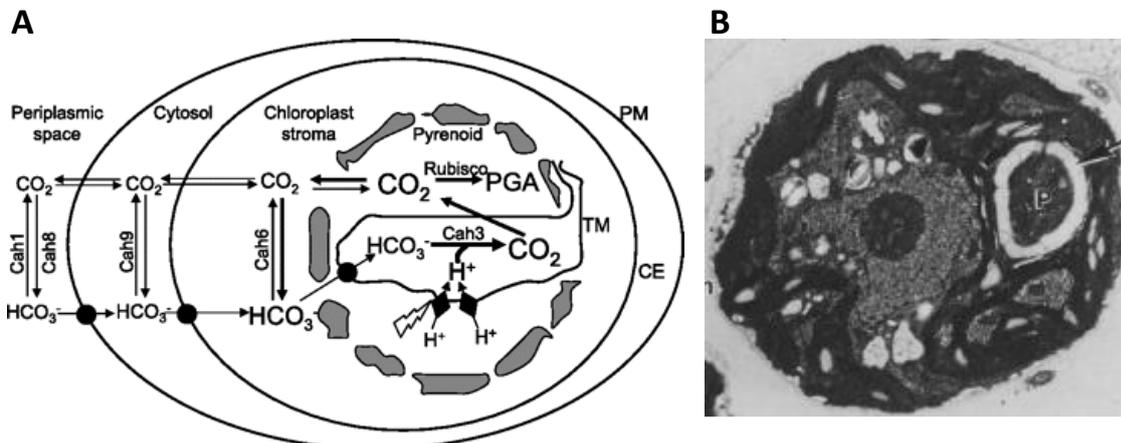
Figure 2.17 Pyrenoid localization in algal cell and mechanistic model of CO_2 uptake.
(A) Schematic model of varying transport mechanisms of CO_2 and HCO_3^- uptake in an algal cell. Size of lettering is proportional to relative concentrations; CO_2 and HCO_3^- concentrations within the pyrenoid are significantly higher than external concentrations. Cah: isoforms of carbonic anhydrase, PGA: 3-phosphoglycerate, PM: plasma membrane, CE: chloroplast envelope, TM: thylakoid membrane. Possible CO_2 or HCO_3^- transporters are depicted as solid circles while the photosynthetic electron transport chain on the thylakoid membrane is depicted as solid diamonds. Adapted from: (Moroney and Ynalvez 2007) (B) Electron micrograph of *Chlamydomonas reinhardtii*, pyrenoid is labeled as P and the surrounding starch is highlighted with an arrow. Adapted from: (Ramazanov et al. 1994)

2.3.3.4 Carboxysome

The carboxysome is a bacterial microcompartment encased by an icosahedral proteinaceous outer shell consisting of hexameric and pentameric subunits as illustrated in Figure 2.18 A-C (Tanaka et al. 2008). This microcompartment is located within cyanobacteria and houses the machinery for CO_2 fixation (Figure 2.11), specifically Rubisco and carbonic anhydrase. Cyanobacteria are categorized as either α or β, depending on the form of Rubisco present in the carboxysome. Typically, α-cyanobacteria are located in marine environments and contain form IA Rubisco and α-carboxysomes. Whereas, β-cyanobacteria are located in freshwater environments and contain form IB Rubisco and β-carboxysomes (Badger et al. 2002). (Further discussion of the various forms of Rubisco can be found in section 2.3.4.2.)

The carboxysome is essential for autotrophic growth and allows the entry of Rubisco and HCO_3^-; thereby effectively concentrating CO_2 at the active site of Rubisco and consequently increasing Rubisco carboxylase efficiency and reducing photorespiration (Figure 2.18 D). The spatial arrangement of the carboxysomes along the axis of cyanobacteria cells is non-random. During cell division the cytoskeleton navigates the segregation of carboxysomes to the daughter cells, ensuring that each daughter cell receives equal amounts

of carboxysomes. Non-random segregation of carboxysomes thus optimizes carbon fixation efficiency and cellular fitness of the progeny (Savage et al. 2010). Carbon concentrating mechanisms for organisms inhabiting aquatic environments and ensuring progeny fitness is essential especially since the diffusion of CO_2 in water is exceedingly slower than in air (Kennedy 2000).
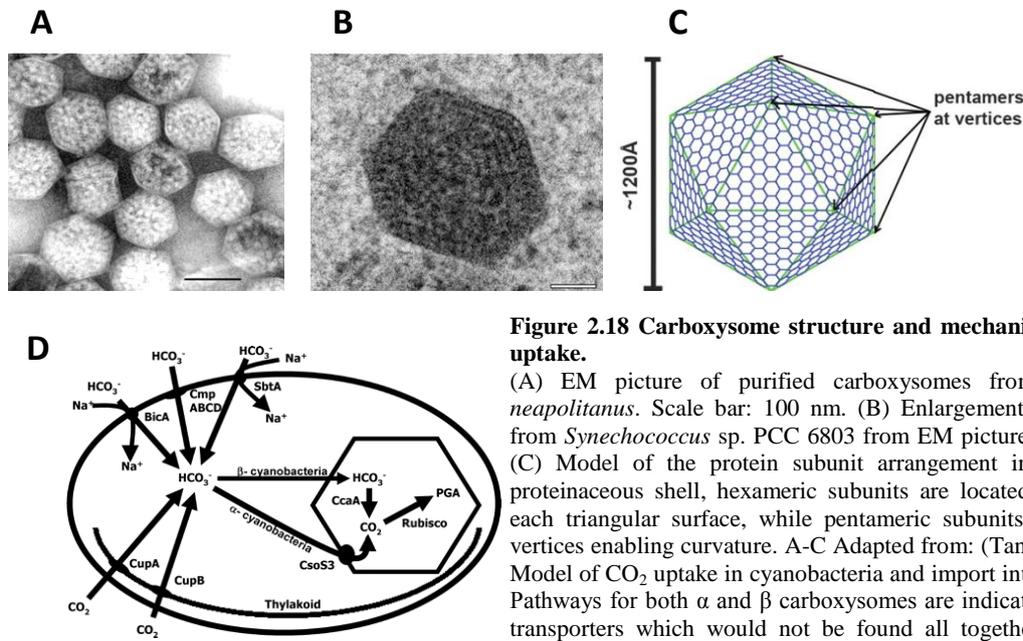
Figure 2.18 Carboxysome structure and mechanistic model of CO_2 uptake.
(A) EM picture of purified carboxysomes from *Halothiobacillus neapolitanus*. Scale bar: 100 nm. (B) Enlargement of a carboxysome from *Synechococcus* sp. PCC 6803 from EM picture. Scale bar: 50 nm. (C) Model of the protein subunit arrangement in the carboxysome proteinaceous shell, hexameric subunits are located along the face of each triangular surface, while pentameric subunits are located at the vertices enabling curvature. A-C Adapted from: (Tanaka et al. 2008) (D) Model of CO_2 uptake in cyanobacteria and import into the carboxysome. Pathways for both α and β carboxysomes are indicated as well as many transporters which would not be found all together within the same cyanobacteria. HCO_3^- or CO_2 are imported into the cell and HCO_3^- is transported to the carboxysome where it is converted to CO_2 by carbonic anhydrase and can be fixed by Rubisco. D Adapted from: (Moroney and Ynalvez 2007)

2.3.4 Ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco)

Ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) is an enzyme that has been under scrutinous investigation for the past four decades. Rubisco functions in the catalysis of the initial step of CO_2 assimilation in the CBB cycle. Unfortunately, this enzyme has several shortcomings including a slow catalytic rate (three CO_2 molecules fixed per second per Rubisco) and it does not discriminate well between the assimilation of CO_2 or O_2 to RuBP (Badger and Bek 2007; Gatenby and Ellis 1990). Photosynthetic organisms produce tremendous amounts of Rubisco to alleviate these shortcomings making Rubisco the most abundant protein on earth (Ellis 1979). Rubisco accounts for 30-50% (240 mg/ml) of total leaf protein (Parry et al. 2003). Therefore, significant quantities of nitrogen stores are invested in the production of Rubisco (Parry et al. 2008). These drawbacks of Rubisco have important consequences in the agricultural industry. Furthermore, Rubisco also has implications in relation to increasing CO_2 concentrations and temperatures in the context of global warming. Increasing temperatures can have detrimental impacts on Rubisco regulation especially in

combination with increasing CO_2 concentrations in the atmosphere. Increased concentrations of atmospheric CO_2 should potentially decrease the oxygenase side-reaction of Rubisco and increase carboxylase efficiency, however increases in CO_2 concentrations decrease plant stromal conductance leading to difficulties in transpiration and an even greater increase in leaf temperature (Sage et al. 2008; Salvucci 2008). Increases in temperature also decrease photosynthetic capacity due to impacts on electron transport and thermal denaturation of Rubisco activase, a protein necessary for Rubisco regulation through removal of inhibitors (described in section 2.3.4.1). Rubisco oxygenase activity also increases with rising temperatures due to reduced specificity of Rubisco to CO_2 and decreased concentrations of CO_2 compared to O_2 in the chloroplast at higher temperatures (Ehleringer et al. 1991). Oxygenation of RuBP not only decreases the amount of RuBP available for carboxylation in the CBB cycle but also decreases CO_2 assimilation efficiency by 50% (Andersson and Backlund 2008). Carbon fixation by Rubisco is the rate limiting step of the CBB cycle and rate limiting in photosynthesis.

Increasing photosynthetic capacity through improvement of Rubisco CO_2 specificity and/or acceleration of substrate turnover rate has been the focus of various avenues of research. Attempts have been made to improve Rubisco function by specific mutation of residues involved in catalysis and CO_2 specificity. Mutation of K128 (spinach numbering), an active site residue, in *Synechococcus* sp. PCC6301 (Syn6301) Rubisco resulted in a decrease in carboxylation (Bainbridge et al. 1998); in contrast mutation of residue C172 (spinach numbering) in *Chlamydomonas reinhardtii* to a serine resulted in an 11% increase in CO_2 specificity (Garcia-Murria et al. 2008). Although, improvements in the catalytic function of Rubisco have yielded limited enhancement, these experiments give important insights into amino acids involved in Rubisco catalytic function.

Numerous advancements have been made in our understanding of the behavior of this enzyme. It is now possible to transform tobacco plants with Rubisco from other species and determine if the foreign Rubisco remains soluble and has the capacity to substitute for inherit Rubisco e.g., Rubisco from Syn6301 and sunflower (Kanevski et al. 1999), *Rhodospirillum rubrum* (Whitney and Andrews 2001), and non-green algae (Whitney et al. 2001). The mRNA of Syn6301 Rubisco was transcribed but was not translated into protein, whereas sunflower Rubisco was able to assemble and exhibited 20% carboxylase activity of wildtype plants. Furthermore, *Rhodospirillum rubrum* Rubisco was assembled properly in tobacco leaves and could support photosynthetic growth at increased CO_2 concentrations while non-green algal Rubisco protein failed to properly fold and assemble. These results indicate that Rubisco from

different species most likely either have different mechanisms of assembly and/or accessory factors are needed which are not present in the tobacco plastid.

2.3.4.1 Rubisco regulation

The ATP and NADPH produced from the light-dependent reactions are essential for the CBB cycle as a means for chemical energy and reducing agents. Furthermore, the regulation of enzymes in the CBB cycle including Rubisco are coupled to the flux of the light-dependent reactions. Light-dependent reactions change conditions in the stroma such as pH increase through H^+ transport into the lumen counteracted by increases in Mg^{2+}, NADPH, and reduced ferredoxin amounts, which are important for regulation of CBB cycle enzymes (Berg et al. 2002). The increase in pH and Mg^{2+} concentrations in the stroma are essential for proper Rubisco function in that these conditions are necessary for carbamate formation in the activation of Rubisco.
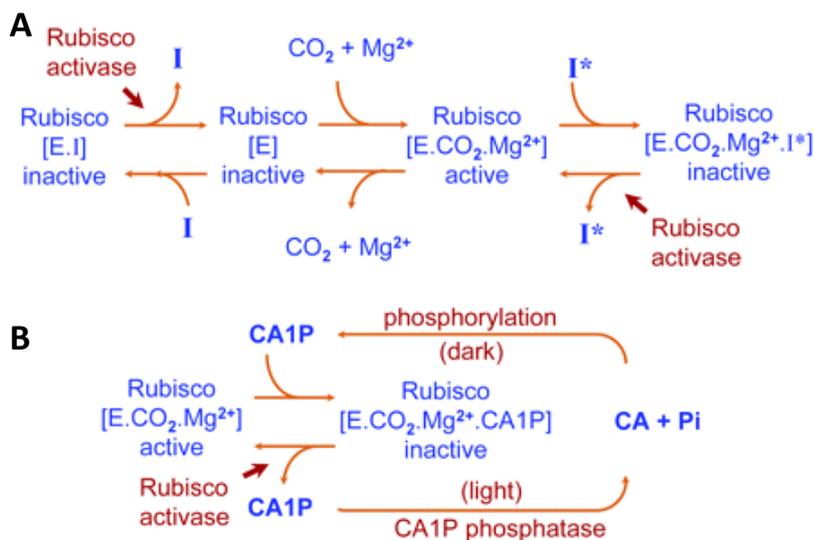
Figure 2.19 Rubisco regulation. (A) Decarbamylated Rubisco, E, can bind substrate, EI, and become inhibited. Carbamylated Rubisco, $E.CO_2.Mg^{2+}$, can also become inhibited when bound with substrate analogues, $E.CO_2.Mg^{2+}.I^*$. Rubisco activase is needed in both of these conditions to remove the inhibitory sugar-phosphate compound. (B) Carbamylated Rubisco can be inhibited by 2-carboxy-arabinitol-1-phosphate (CA1P) or D-glycero-2,3-pentodiulose-1,5-bisphosphate (PDBP) in the dark and reversibly activated by Rubisco activase in the light. Adapted from: (Parry et al. 2008)

Rubisco must be activated for proper function, which begins by the carbamylation of active site residue K201 (spinach numbering) by a non-substrate CO_2, changing the charge of K201 from positive to negative. The carbamylated K201 is stabilized by coordinating a Mg^{2+} ion at the active site in concert with residues D203 and E204 (Cleland et al. 1998; Kellogg and Juliano 1997). The addition of Mg^{2+} at the Rubisco active site is important for positioning of the RuBP enediolate intermediate state by O-2 and O-3 coordination; formation and stabilization of an enol-substrate intermediate is conserved among members of the Rubisco super family (Imker et al. 2007). K201 also acts as the general base that abstracts a proton from C-3 of substrate RuBP to create the enolization state of RuBP, whose formation aids in cleavage of the resulting six-carbon intermediate upon addition of substrate CO_2 to RuBP.

Formation of the carbamate at K201 with CO_2 is spontaneous *in vitro* (Houtz and Portis 2003).

However, RuBP can bind to Rubisco before carbamylation creating a dead-end enzymatically-inactive complex (Figure 2.19). Additionally, other naturally occurring sugar-phosphate analogs, such as 2-carboxy-arabinitol-1-phosphate (CA1P), D-xylulose-1,5-bisphosphate (XuBP), and D-glycero-2,3-pentodiulose-1,5-bisphopsphate (PDBP), can also tightly bind to the active site thereby inhibiting Rubisco catalytic activity (Lorimer et al. 1977; Parry et al. 2003; Parry et al. 2008). This enzymatically-inactive complex can be rescued by the protein Rubisco activase. Rubisco activase is a AAA^+ protein that releases inhibitor substrates from the active site of Rubisco in an ATP-dependent manner followed by activation of Rubisco with CO_2 and Mg^{2+} at the active site (Kellogg and Juliano 1997). In this effect, Rubisco activase regulates the activation state of Rubisco. Rubisco activase is regulated by the ratio of ATP to ADP (being inhibited by high concentrations of ADP), therefore Rubisco is activated by Rubisco activase in response to light intensity due to the increase in ATP concentrations from the light-dependent reactions during the day (Graciet et al. 2004; Parry et al. 2003). Rubisco activase has been found in all genomes of C3 and C4 plant species as well as in green algae (Spreitzer and Salvucci 2002), indicating the general importance of this protein for Rubisco regulation.

2.3.4.2 Rubisco structure

There are four different forms of Rubisco in the Rubisco super family found in nature (Tabita et al. 2008). The most abundant, form I Rubisco, includes four subclasses (A-D) distinguished as green-type (form IA and form IB) and red-type (form IC and form ID), which are found in higher plants, algae, cyanobacteria, and some proteobacteria. Form I Rubisco is hexadecameric consisting of a core of eight 50-55 kDa large subunits (RbcL), arranged as four anti-parallel dimers, which is capped on the top and bottom with eight 12-18 kDa small subunits (RbcS) to give a $RbcL_8S_8$ complex of approximately 550 kDa (Figure 2.20 A).

Form II Rubisco and form III Rubisco are composed exclusively of large subunits, lacking small subunits (Figure 2.20 B and C, respectively). Form II Rubisco is active as an RbcL anti-parallel dimer and is found in photosynthetic proteobacteria and dinoflagellates. Form II Rubisco is necessary for phototrophic growth of these organisms; whereas, form III Rubisco is found in some thermophilic archaea which do not have a CBB cycle (Altekar and Rajagopalan 1990; Watson et al. 1999), however form III Rubisco exhibits carboxylase activity. Form III Rubisco is typically arranged as a ring consisting of five anti-parallel dimers

$(RbcL_2)_5$ (Kitano et al. 2001). In archaeal organisms, form III Rubisco is implicated in AMP metabolism; converting AMP, phosphate, CO_2, and H_2O to adenine and two molecules of 3-PGA in concert with various enzymes (Sato et al. 2007).
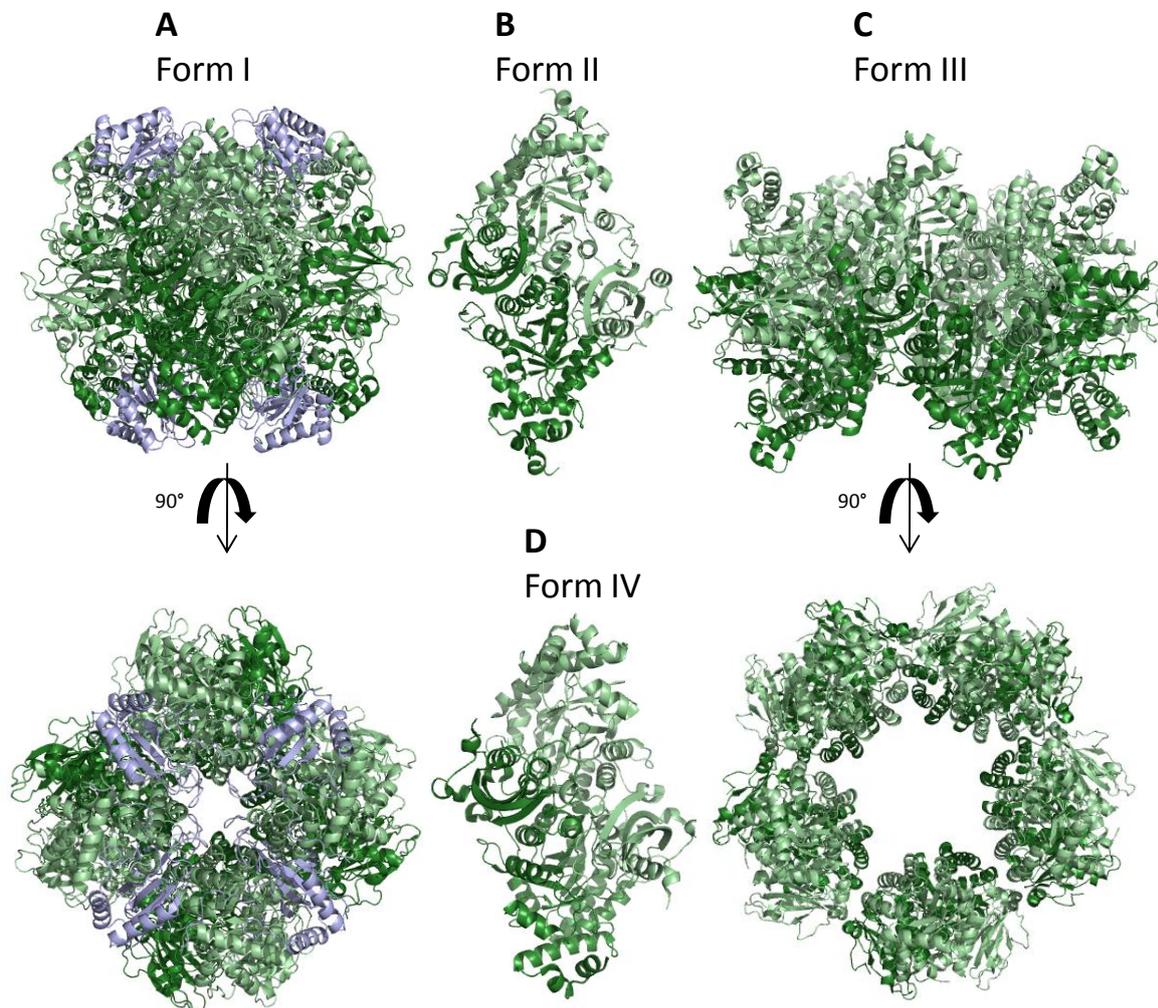
Figure 2.20 Rubisco super family structures.
(A) Side-view and top-view of form I Rubisco from spinach. Large subunits are shown in green tones and small subunits are shown in light blue (PDB 1RCX (Taylor and Andersson 1997)). (B) Form II Rubisco from *Rhodospirillum rubrum*, large subunit monomers are indicated with green tones (PDB 5RUB (Schneider et al. 1990)). (C) Side-view and top-view of form III Rubisco from *Thermococcus kodakarensis*, large subunit monomers are indicated with green tones (PDB 1GEH (Kitano et al. 2001)). (D) Form IV Rubisco from *Chlorobium tepidum*, large subunit monomers are indicated with green tones (PDB 1YKW (Li et al. 2005)).

The newest member of the Rubisco super family, form IV Rubisco-like-protein (RLP), is also solely composed of large subunits and is active as a dimer (Figure 2.20 D). RLP occur in organisms that do not have a CBB cycle and furthermore this class is divided in to six clades based on primary sequences (Tabita et al. 2008). Form IV Rubisco is termed RLP since it does not harbor any carboxylase activity due to critical substitutions in the active site (Li et al. 2005). The function of RLP is not known for all organisms in which it is found, but it is indicated in sulfur metabolism (Tabita et al. 2007). In *Geobacillus kaustophilus* and *Bacillus*

subtilis the RLP is implicated in the methionine salvage pathway and functions as an enolase, stabilizing the enol form of active site substrate (Imker et al. 2007).
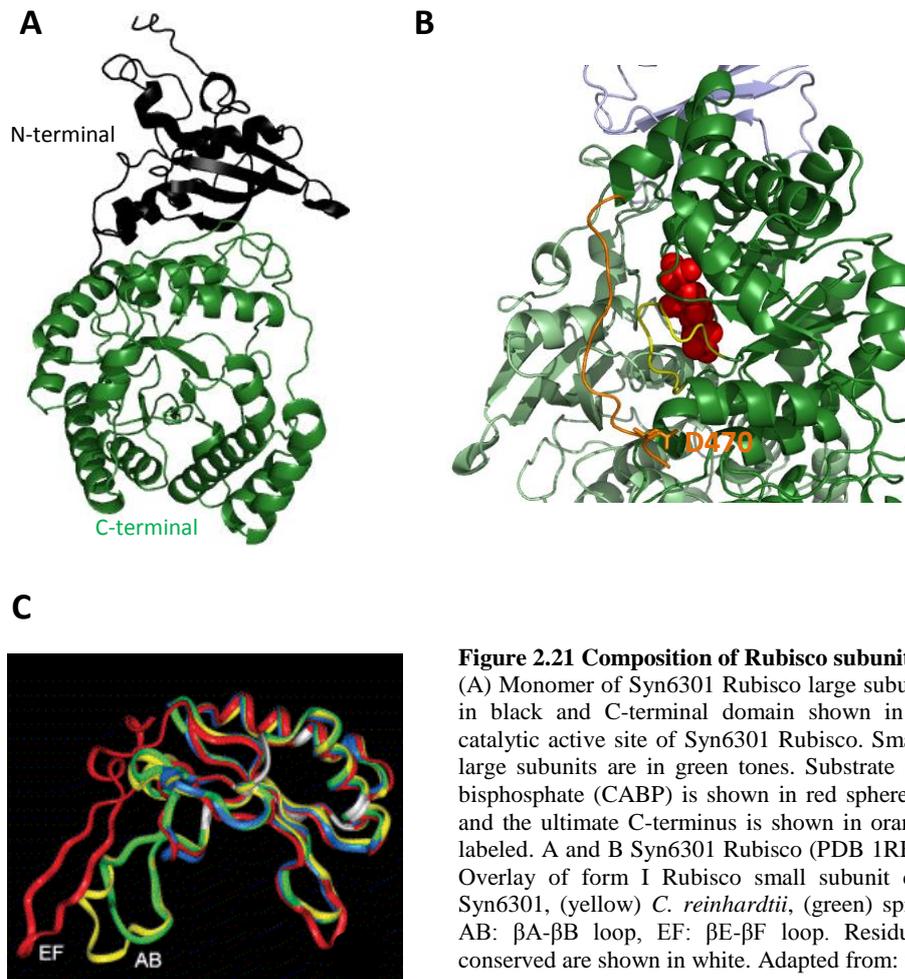
Figure 2.21 Composition of Rubisco subunits and active site.
(A) Monomer of Syn6301 Rubisco large subunit, N-terminal domain shown in black and C-terminal domain shown in green. (B) Close-up of one catalytic active site of Syn6301 Rubisco. Small subunit is in light blue and large subunits are in green tones. Substrate analog carboxy-arabinitol-1,5-bisphosphate (CABP) is shown in red spheres. Loop 6 is shown in yellow and the ultimate C-terminus is shown in orange with anchor residue D470 labeled. A and B Syn6301 Rubisco (PDB 1RBL (Newman et al. 1993)). (C) Overlay of form I Rubisco small subunit crystal structures from (blue) Syn6301, (yellow) *C. reinhardtii*, (green) spinach, (red) *Galdieria partita*. AB: βA-βB loop, EF: βE-βF loop. Residues that are more than 95% conserved are shown in white. Adapted from: (Spreitzer 2003)

The overall tertiary structures of the Rubisco large subunit dimer is conserved in all forms of Rubisco (Figure 2.20), including high conservation of active site residues. RbcL is comprised of two domains, the N-terminal domain and the C-terminal domain. The N-terminal domain consists of 150 amino acids structured as a five stranded β-sheet with two α-helices on one side (Figure 2.21 A). The C-terminal domain is structured as an eight stranded α/β-barrel (TIM-barrel) and the active site is located on one side of the barrel in conjunction with N-terminal residues of the adjacent subunit (Houtz and Portis 2003; Leegood et al. 2000). The RbcL anti-parallel dimer is the minimal functional unit of Rubisco since two subunits are necessary to complete formation of the two active sites (Andersson and Backlund 2008). During catalytic function following Rubisco activation, loop 6 of the α/β barrel closes over the active site followed by the ultimate C-terminus which is anchored by the latch residue D470 by ionic interactions with R131 and H307 within the same RbcL subunit (Figure 2.21 B, Syn6301 numbering) (Andersson and Backlund 2008; Duff et al. 2000; Kellogg and Juliano 1997).

Form I Rubisco is divergent from the other Rubisco in the super family in that RbcS is necessary for complete catalytic competence and structural integrity of the holoenzyme (Andrews 1988; Spreitzer and Salvucci 2002). Even though the $RbcL_8$ core can perform limited catalysis in the absence of RbcS, RbcS seems to be necessary for CO_2/O_2 specificity (Read and Tabita 1992). Interestingly, form I large subunits are highly homologous with up to 90% sequence identity from divergent species. RbcS is however poorly conserved between species. Examination of the tertiary structure of RbcS from different species shows intriguing differences (Spreitzer 2003). The dominant structural variation of RbcS is the addition of a prominent extension of the βA-βB loop e.g., in plants and green algae, into the $RbcL_8$ pore creating interactions that are not present in species without this loop extension e.g., cyanobacteria (Figure 2.21 C) (Knight et al. 1990). The role of this loop extension is still uncertain, but mutation of this loop showed that catalysis and specificity of Rubisco were affected (Esquivel et al. 2002). The RbcS from species of non-green algae contain two extended β-strands that extend into the large subunit central-solvent channel (Figure 2.21 C). The exact role of the small subunit is still unknown. However, experiments involving formation of chimeric Rubisco from different species have indicated that there can be incompatibility between Rubisco subunits (Andrews and Whitney 2003) and importantly that the Rubisco small subunit can strikingly influence CO_2/O_2 affinity (Andrews and Lorimer 1985) and catalysis (Karkehabadi et al. 2005).

2.3.4.3 Rubisco folding and assembly

Correct folding and assembly of the Rubisco holoenzyme is essential for biological function. For the simpler form II Rubisco this entails the proper folding of the large subunit monomers and assembly of the RbcL dimer, which occurs in the cytosol. Prokaryotic form I Rubisco folding and assembly is also located in the cytosol and *rbcL* and *rbcS* are located on the same operon. Red-type form I *rbcL* and r*bcS* genes from non-green algae are encoded on the same operon in the chloroplast genome where the proteins are also synthesized (Andrews and Whitney 2003). In contrast, higher plant and green algae green-type form I Rubisco encode the large subunit in the chloroplast genome (encoded as a single copy) with synthesis on the chloroplast ribosome (Blair and Ellis 1973), whereas, the small subunit occurs as a homologous multi-gene family and is encoded in the nuclear genome with synthesis on cytosolic ribosomes (Dobberstein et al. 1977; Hartman and Harpel 1994). The specific role of multiple RbcS proteins in Rubisco is still unknown. RbcS is first synthesized as a precursor molecule in the cytosol with an N-terminal transit peptide for targeting to the chloroplasts

where it is subsequently processed into full-length RbcS. RbcS then assembles on the $RbcL_8$ core, forming the holoenzyme (Bradley et al. 1986; Chua and Schmidt 1978; Goloubinoff et al. 1989b; Hartman and Harpel 1994).

Form II Rubisco and prokaryotic form I Rubisco can be recombinantly expressed in *E. coli* (Gatenby et al. 1985; Somerville and Somerville 1984). However, no soluble eukaryotic form I Rubisco has been expressed outside of the chloroplast (Gatenby 1984) even if chloroplast chaperonin is also coexpressed in *E. coli* (Cloney et al. 1993). Early experiments of Rubisco revealed that a binding protein (Cpn60) was associated with newly synthesized Rubisco in chloroplasts (Barraclough and Ellis 1980; Bloom et al. 1983; Roy et al. 1988), yet at that time the exact role of this protein was not known. An important breakthrough in understanding a critical step in Rubisco biogenesis arose when it was demonstrated that a certain subset of chaperones, the chaperonins, are required to obtain assembled Rubisco folding in a bacterial host system (Goloubinoff et al. 1989b) and for *in vitro* refolding of form II Rubisco (Goloubinoff et al. 1989a). The bacterial DnaK/DnaJ/GrpE chaperone system also aids in folding of recombinantly expressed Rubisco (Checa and Viale 1997), but the chaperonin system was found to be essential for the folding and possibly the assembly of Rubisco large subunits. This advance was due to the realization that the bacterial chaperonin system, GroEL and GroES, are evolutionary homologues of the chloroplast chaperonin system, Cpn60 and Cpn21 (Hemmingsen et al. 1988). The action of chaperonin in assisting RbcL folding is essential for Rubisco assembly. Once the $RbcL_8$ core of form I Rubisco is formed, RbcS binds and assembles to give the active holoenzyme in an ATP-independent manner (Gatenby and Ellis 1990; Hartman and Harpel 1994).

To date higher plant Rubisco has failed to reconstitute upon expression in *E. coli*. However, the reason for this remains ambiguous especially since prokaryotic and eukaryotic form I Rubisco from different species exhibit high sequence identity and structural similarity. One possible explanation is that eukaryotic Rubisco undergoes co- and post-translational modifications which are not fulfilled upon expression in *E. coli* e.g., deformylation, acetylation, methylation, and N-terminal proteolytic processing of the holoenzyme (Houtz et al. 2008). This possibility seems less likely because many of these modifications occur after assembly and influence overall stability or activity and not necessarily assembly. It is more likely that specific folding or assembly factors are missing in *E. coli* that occur in the chloroplast which are necessary for holoenzyme formation. Large protein complexes consisting of multiple subunits have been shown to require assistance from chaperones not only for folding but for correct assembly. For example accessory assembly factors are utilized

for the assembly of the nucleosome (Laskey et al. 1978) and the proteasome (Witt et al.
2000). Indeed, for Rubisco, one such assembly factor, RbcX, has been characterized and
shown to be involved in cyanobacterial form I Rubisco assembly (Li and Tabita 1997; Liu et
al. 2010; Saschenbrecker et al. 2007; Tarnawski et al. 2008).

2.3.4.4 RbcX
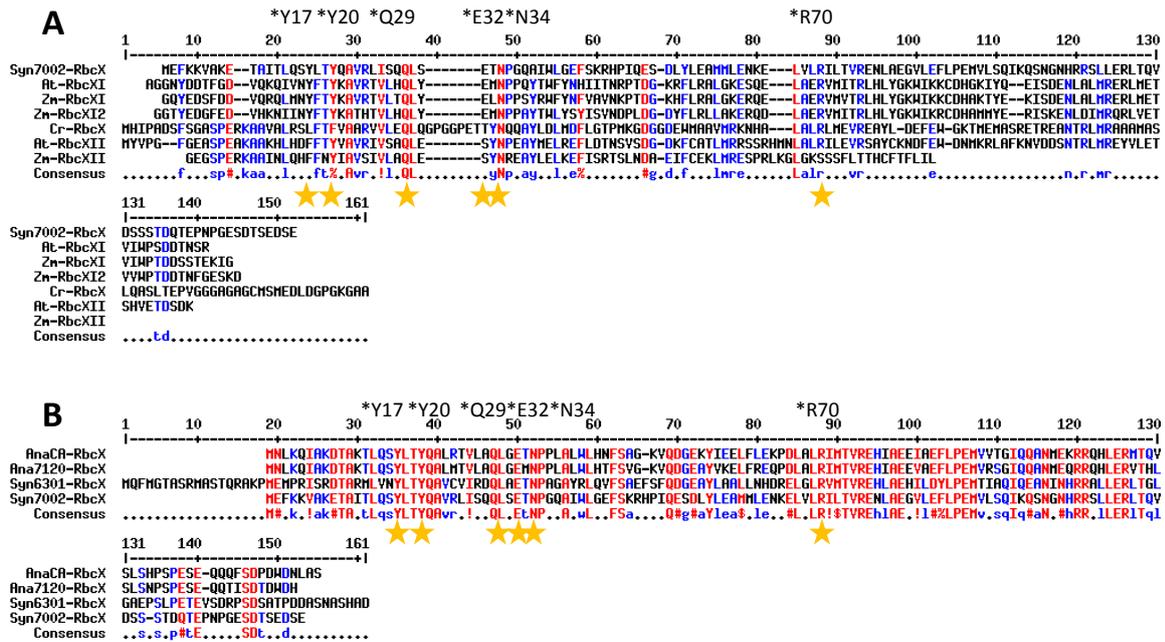
Figure 2.22 Amino acid sequence alignment of RbcX₂ from prokaryotes and eukaryotes.
(A) Amino acid sequence alignment of RbcX₂ from Syn7002 compared to eukaryotic species without putative-transit
peptide: *C. reinhardtii* (Cr), *A. thaliana* (At), *Zea mays* (Zm) without putative transit peptides. Isoforms of eukaryotic RbcX₂
are indicated as RbcXI and RbcXII. Conserved and important regions are indicated with a star and the numbering above the
sequences is according to Syn7002. (B) Amino acid alignment of RbcX₂ from cyanobacteria, labeling as in A. Conserved
residues are shown in red and residues found in two of the sequences are shown in blue. Black indicates no homology.
Alignment created using MultAlin.

The discovery of the small assembly chaperone, RbcX, revealed that besides the
chaperonin system there is another protein involved in Rubisco biogenesis (Li and Tabita
1997). Homologues of *rbcX* have been found in genomes of species encoding form IB
Rubisco including cyanobacteria, green algae, and higher plants. In some cyanobacteria
species (*Anabaena* sp. PCC7120, Ana7120; *Anabaena* sp. CA, AnaCA; *Synechococcus* sp.
PCC7002, Syn7002) the *rbcX* gene is encoded on the *rbc* operon between the *rbcL* and *rbcS*
genes, and deletion of the *rbcX* gene region hinders the production of soluble Rubisco
(Larimer and Soper 1993; Onizuka et al. 2004). In *Synechococcus* sp. PCC6301 (Syn6301)
the *rbcX* gene is located downstream of the *rbcL* and *rbcS* operon. Syn6301 Rubisco can
assemble in *E. coli* without expression of the *rbcX* gene; however coexpression of r*bcX*
enhances the production of assembled Rubisco (Saschenbrecker et al. 2007). The *rbcX* gene is

encoded in the nuclear genome of green algae and higher plants with at least two isoforms of the *rbcX* gene present in the genome of higher plants (Figure 2.22 A).

RbcX is a homodimer of ~15 kDa subunits consisting of four α-helices aligned in an anti-parallel fashion along the α4 helix held together by uncharged/hydrophobic interactions (Figure 2.23). $RbcX_2$ functions as a stabilizer of folded RbcL. The $RbcX_2$ protein recognizes a highly conserved C-terminal sequence of the large subunit of form IA and form IB Rubisco, termed the C-terminal recognition motif: EIKFEFD (Figure 2.24). After RbcL interacts with and is folded by the chaperonin system, $RbcX_2$ stabilizes the folded RbcL and then assists in the formation of the $RbcL_8$ core. $RbcX_2$ does not remain a part of the final holoenzyme complex, instead $RbcX_2$ has a dynamic relationship with RbcL. $RbcX_2$ is displaced by RbcS binding, probably due to a conformational change in the large subunits (Liu et al. 2010; Saschenbrecker et al. 2007). However, the exact mechanism by which RbcS binding displaces $RbcX_2$ from the $RbcL_8$ core is still unknown.
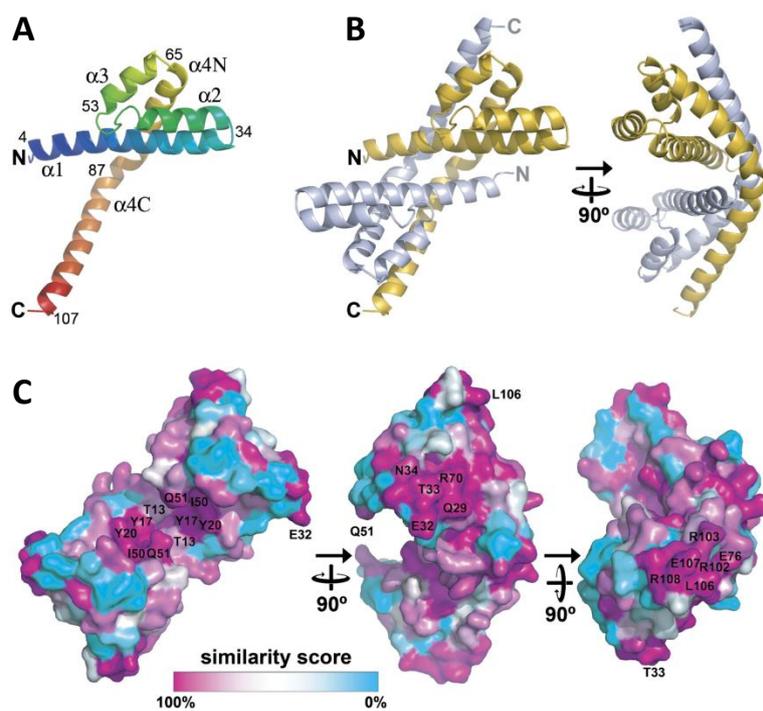
Figure 2.23 Crystal structure of Syn7002-$RbcX_2$ and $RbcX_2$ conserved surface regions.
(A) Structure of the RbcX monomer. N-to C-terminal is indicated by a gradient of cold to warm colors, α-helices 1-4 are labeled along with numbering of specific amino acids along the chain. (B) RbcX dimer structure. Subunits are depicted in yellow and light blue; N- and C-termini are labeled. (C) Amino acid surface conservation in $RbcX_2$. 151 cyanobacterial $RbcX_2$ sequences were aligned and the similarity score plotted onto the exposed surface on $RbcX_2$ using the PFAM database. Highly conserved residues are shown in magenta and variable residues are shown in cyan. Conserved surface residues are labeled. Adapted from: (Saschenbrecker et al. 2007)

Similarity scores from an alignment of 151 $RbcX_2$ sequences from cyanobacteria revealed that there are two highly conserved regions on $RbcX_2$: the 5.4 Å wide central hydrophobic crevice and the peripheral polar surfaces at the corners of the dimer (Figure 2.23). From mutational studies, it was found that both of these regions are critical for $RbcX_2$ function. The central crevice is important for the production of soluble RbcL by binding to the extended C-terminal recognition motif. Two phenylalanines (462 and 464, Figure 2.24 C: numbering according to Syn7002 Rubisco) extend into the hydrophobic groove of $RbcX_2$. The peripheral surface is important for efficient $RbcL_8$ core complex formation. Importantly,

residues that are necessary for cyanobacterial $RbcX_2$ function are also highly conserved in the sequences of higher plant $RbcX_2$ i.e., Y17, Y20, Q29, E32, N34, R70 (Figure 2.22). The dynamic relationship of the RbcL/RbcX complex is critical for the formation of the holoenzyme. Comparison of $RbcX_2$ from cyanobacteria has shown that they share only about 50% sequence identity (Figure 2.22 B), however crystal structures of various prokaryotic $RbcX_2$ proteins have indicated high similarity of tertiary and quaternary structures (Saschenbrecker et al. 2007; Tanaka et al. 2007).
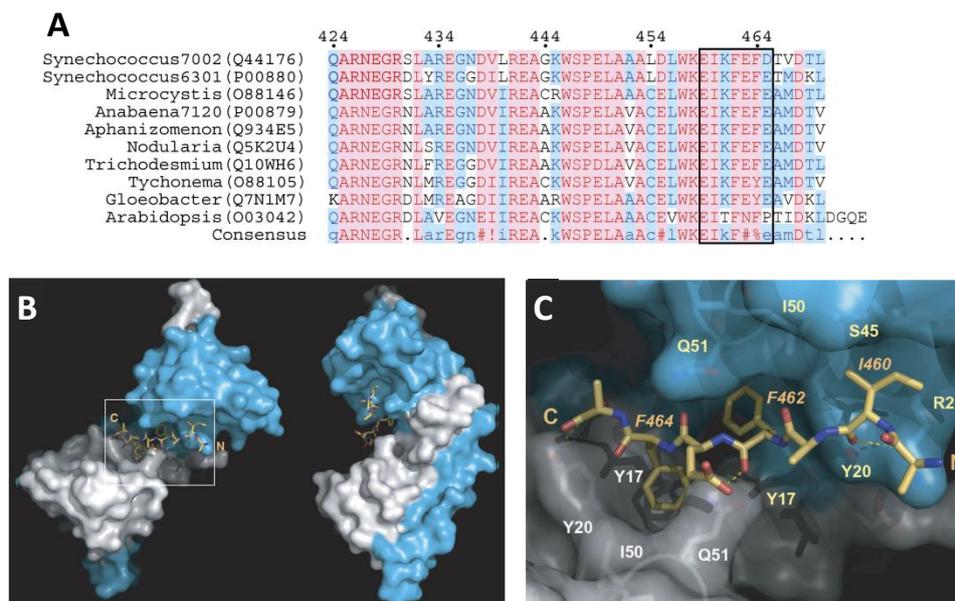
Figure 2.24 Binding of $RbcX_2$ to RbcL C-terminal recognition motif.
(A) Alignment of the C-terminal residues of RbcL from cyanobacteria and higher plant species as indicated using MultAlin. Swiss-Prot accession numbers are shown in brackets. Red coloring indicates high (greater than 90%) consensus level and blue coloring indicates low (less than 50% consensus level). Conserved RbcL C-terminal recognition motif is boxed (B) Surface representation of $RbcX_2$ crystal structure (individual subunits shown in white and blue) in complex with the peptide EIKFEFD in stick representation. N- and C- termini of the peptide are shown. (C) Close-up of the boxed-in area in B, the interactions of the EIKFEFD peptide with the groove region of $RbcX_2$ are highlighted. Hydrogen bonds are indicated with dashed lines. Residues of $RbcX_2$ that participate in peptide binding are shown in stick representation below the transparent surface and colored white or yellow for each subunit. The phenylalanines of the C-terminal peptide that bind the hydrophobic pockets in the $RbcX_2$ groove are also labeled. Adapted from: (Saschenbrecker et al. 2007)

It was observed that when RbcL and $RbcX_2$ from different species are used together, a stable, but nonetheless dead-end complex was formed (e.g., Syn7002-RbcL with AnaCA-$RbcX_2$). The affinity of AnaCA-$RbcX_2$ to the C-terminal recognition motif of Syn7002-RbcL is extremely high (K_D: 5 μM) and Syn7002-RbcS is not able to replace AnaCA-$RbcX_2$, hindering holoenzyme formation (Saschenbrecker et al. 2007). $RbcX_2$ is a substrate specific assembly chaperone; thus, these previous results indicate that the RbcL and $RbcX_2$ dynamic interaction evolved in a way where displacement of $RbcX_2$ by RbcS is optimal for proteins from the same species. The dynamic nature of the RbcX/RbcL complex has hindered attempts to characterize the regions of RbcL that interact with the peripheral regions of $RbcX_2$. In the current model of cyanobacterial form I Rubisco folding and assembly, the large subunit

undergoes folding in the GroEL/ES cage. Upon release of the large subunit from the chaperonin, $RbcX_2$ binds the C-terminal recognition motif and promotes the formation of $RbcL_8$ complexes. Binding of RbcS displaces $RbcX_2$ thereby creating the functional holoenzyme (Figure 2.25) (Liu et al. 2010; Saschenbrecker et al. 2007).
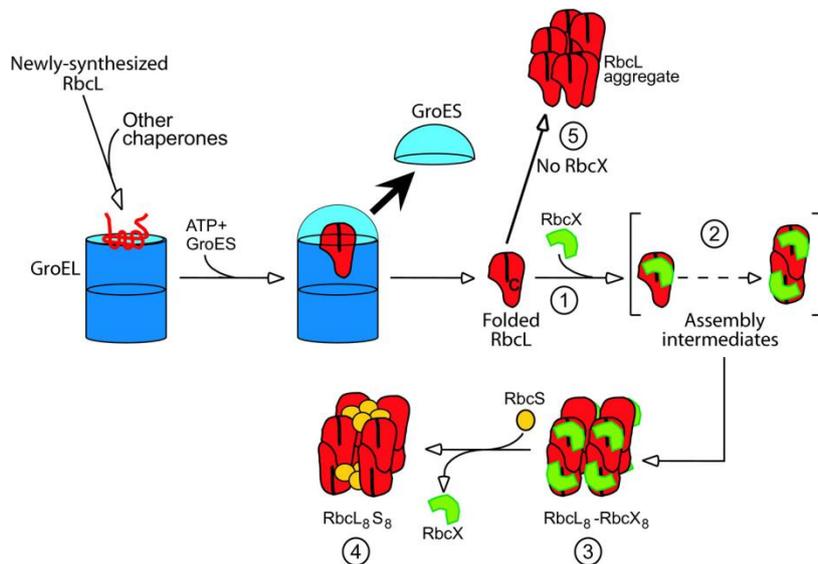
Figure 2.25 Model of form I Rubisco folding and assembly.
1. $RbcX_2$ interaction occurs downstream of GroEL/ES action. 2. Binding of $RbcX_2$ stabilizes folded RbcL at dimer or monomer level. 3. Subsequently followed by formation of the $RbcL_8$ core. 4. The RbcL/RbcX interaction is dynamic, and RbcS binding to the $RbcL_8$ core results in the displacement of bound $RbcX_2$ resulting in holoenzyme formation.
Adapted from: (Saschenbrecker et al. 2007)

Form II Rubisco was the only Rubisco to be successfully reconstituted *in vitro* from the denatured state using the GroEL/ES system (Goloubinoff et al. 1989a). However, more recently, form I Rubisco from Syn6301 cyanobacteria was successfully reconstituted *in vitro* (Liu et al. 2010). The GroEL/ES system and the assembly chaperone $RbcX_2$ are the two key chaperones used for the *in vitro* reconstitution. It was demonstrated that when $RbcX_2$ was not present in the assay, Syn6301-RbcL was not released in an assembly competent manner, and hence rebound to GroEL. Furthermore, addition of cognate $RbcX_2$ to the assay also failed to produce assembled $RbcL_8$ core complexes. Folded RbcL could only be removed from GroEL if the high affinity heterologous $RbcX_2$ was added to the system. The high affinity $RbcX_2$ could compete with GroEL for binding of RbcL and push the equilibrium to RbcL/RbcX assembly competent intermediates. As noted above, high affinity heterologous $RbcX_2$ inhibits proper formation of holoenzyme in that it is not displaced by RbcS creating a dead-end complex; in the *in vitro* assay it was observed that increasing concentrations of high affinity $RbcX_2$ was detrimental to catalytic activity. However, addition of the RbcL C-terminal recognition motif peptide, after assembly of the RbcL/RbcX complex aided in removal of $RbcX_2$ from the $RbcL_8$ core, and as observed on Native-PAGE the holoenzyme, $RbcL_8S_8$, could form (Liu et al. 2010). The development of an *in vitro* reconstitution assay for form I Rubisco is a vital advancement for future efforts to engineer a 'better' Rubisco enzyme by creating a fast and efficient method for screening many Rubisco mutants.

2.4 Aim of study

The ability to engineer a more efficient Rubisco would be advantageous in several ways. For instance, a more efficient Rubisco would reduce photosynthetic water usage, increase plant growth yield thereby creating more food and fuel possibilities for the growing human population, reduce the amount of nitrogen needed for plant growth, and a more efficient Rubisco could potentially process more atmospheric CO_2, thus reducing atmospheric CO_2 concentrations responsible for the green house effect. However, attempts to create a 'better' Rubisco have only resulted in minimal improvements and efforts have been hindered by the inherent complexity of this enzyme. Though much advancement has been made in understanding the folding and assembly of form I Rubisco, the exact and detailed structural mechanism of form I Rubisco assembly remains to be elucidated. Thus, the aim of this study was to gain detailed insight into the $RbcX_2$-mediated assembly mechanism of cyanobacterial form I Rubisco.

To obtain molecular detail of the $RbcX_2$-mediated assembly of form I Rubisco, cryo-EM and X-ray crystallography in concert with mutational analysis were employed. Mutational analysis based on the crystal structure of the $RbcL_8/(RbcX_2)_8$ assembly intermediate were utilized to determine the precise interaction site between the body of RbcL and the peripheral region of $RbcX_2$. From this methodology a critical salt bridge could be identified that is crucial as a guide-point for correct RbcL dimer positioning and it was established that $RbcX_2$ exclusively mediates Rubisco dimer assembly. Furthermore, the mechanism of $RbcX_2$ displacement from the $RbcL_8$ core by RbcS binding was revealed as well as an indication of how RbcS docking on the $RbcL_8$ core would be imperative for full Rubisco catalytic function. Finally, initial attempts in *in vitro* reconstitution of eukaryotic Rubisco are described along with characterization of *Arabidopsis thaliana* binding to the C-terminal recognition motif of the Rubisco large subunit from various species.

3 Materials and Methods

3.1 Materials

3.1.1 Chemicals

All chemicals, unless otherwise stated, were of *pro analysis* quality and were purchased from **Sigma-Aldrich** (Steinheim, Germany) or **Merck** (Darmstadt, Germany).

Amersham Biosciences (Freiburg, Germany): ECL Solutions, NaH^{14}CO$_3$, ^{35}S-methionine

Bio-Rad (Munich, Germany): Bradford protein assay dye reagent, AG 50W-X8 Resin hydrogen form

Biozym (Hessisch Oldendorf, Germany): Biozym LE Agarose

Bachem (Bubendorf, Germany): *para*-benzoyl-L-phenylalanine

Difco (Heidelberg, Germany): Bacto peptone, Bacto tryptone, Bacto yeast extract

Fermentas (St. Leon-Rot, Germany): GeneRuler 1kb DNA Ladder

Invitrogen (Karlsruhe, Germany): dNTP set, protein marker for SDS-PAGE, SYBR Safe DNA gel stain

J.M Gabler Saliter GmbH & Co. KG (Obergünzburg, Germany): skim milk powder

PEQLAB (Erlangen, Germany): IPTG

Qiagen (Hilden, Germany): Ni-NTA Agarose

Roche (Basel, Switzerland): Chloramphenicol, Complete protease inhibitor cocktail, DTT

Roth (Karlsruhe, Germany): Ampicillin, glycine, Rotiszint

Serva (Heidelberg, Germany): Acrylamide-Bis, BSA, PMSF, SDS, Serva Blue G/R

3.1.2 Reagents and Purification kits

Hampton Research (Aliso Viejo / CA, USA): Crystallization screens

Post Genome Institute Co., Ltd (Tokyo, Japan): PURE system

Promega (Mannheim, Germany): PureYield Plasmid Midiprep System, Wizard *Plus* SV Minipreps DNA Purification System, Wizard SV Gel and PCR Clean-Up System

Roche (Basel, Switzerland): RTS 100 *E. coli* HY Kit

3.1.3 Strains

Stratagene (Heidelberg, Germany): *E. coli* BL21 (DE3)

Novagen (Darmstadt, Germany): *E. coli* DH5α

3.1.4 Plasmids, DNA, and oligonucleotides

Novagen (Darmstadt, Germany): pET11a, pET28b, pET30b, pCOLADuet-1

Metabion (Martinsried, Germany): oligonucleotides (primers)

The pHUE plasmid was a kind gift from Dr. S. Whitney (ANU, Canberra, Australia). pUC-*rbcLS* (Smith and Tabita 2003) and pLS32 (Li and Tabita 1997) were kind gifts from Dr. F.R. Tabita and encode the *rbc*-operons from *Synechococcus* sp. PCC6301 and *Anabaena* sp. CA, respectively. The plasmid *p*Sup-BpaRS-6tRN, which allows the incorporation of the photo-reactive amino acid analogue, *para*-benzoyl-L-phenylalanine (*p*Bpa) at *amber* stop codon sites, was a kind gift from Prof P.G. Schultz (Ryu and Schultz 2006). This plasmid encodes the mutant amino-acyl tRNA synthetase (aaRS) BpaRS tyrosyl-tRNA synthetase, *Mj*TyrRS, and two monocistronic operons containing three copies of the *amber* stop codon suppressor tRNA $Mjt\text{RNA}^{Tyr}_{CUA}$ gene. The genomic DNA of *Anabaena* sp. PCC7120 was a kind gift from Dr. S. Sato (Kazusa DNA Research Institute, Japan). The plasmids Synpcc7942_2313(GroEL1)-pET21a and Synpcc7942_2314(GroES)-pET21a were kind gifts from Saaima (Hokkaido System Science Co., Ltd., Japan), and encode the *groEL1* and *groES* genes from *Synechococcus* PCC7942 (6301). The plasmid pBAD33*ES/EL* (Ewalt et al. 1997) allows the arabinose inducible expression of the *E. coli* chaperonins, GroEL and GroES. The plasmid pG-*KJE8* (Nishihara et al. 2000) was utilized to express the *E. coli* proteins GroEL/ES and DnaK/DnaJ/GrpE through tetracycline and arabinose induction, respectively.

The plasmids listed below were used in this study and were constructed by Dr. S. Saschenbrecker and N. Krahmer (TAG mutants):

Plasmid name	Description
Ana7120-*rbcX*-pET11a	Plasmid encoding Rubisco assembly chaperone RbcX from *Anabaena* sp. PCC7120 in pET11a expression vector
AnaCA-*rbcX*$_{N\text{-}FLAG}$-pET30b	Rubisco assembly chaperone RbcX$_{N\text{-}FLAG}$ from *Anabaena* sp. CA in pET30b expression vector (Saschenbrecker et al. 2007)
Syn6301-*rbcL(ΔC)*-pET11a	Plasmid encoding Rubisco large subunit RbcL(ΔC-last 12 amino acids missing) from *Synechococcus* sp. PCC6301 in pET11a expression vector (Liu et al. 2010)
Syn6301-*rbcL(F464A)*-pET11a	Plasmid encoding Rubisco large subunit RbcL(F464A) from *Synechococcus* sp. PCC6301 in pET11a expression vector (Liu et al. 2010)
Syn6301-*rbcL(F466A)*-pET11a	Plasmid encoding Rubisco large subunit RbcL(F466A) from *Synechococcus* sp. PCC6301 in pET11a expression vector (Liu et al. 2010)
Syn6301-*rbcL(F464A,F466A)*-	Plasmid encoding Rubisco large subunit

Plasmid name	Description
pET11a	RbcL(F464A,F466A) from *Synechococcus* sp. PCC6301 in pET11a expression vector (Liu et al. 2010)
Syn7002-*rbcX*-pET11a	Plasmid encoding Rubisco assembly chaperone from *Synechococcus* sp. PCC7002 in pET11a expression vector (Saschenbrecker et al. 2007)
Syn6301-*rbcL*-pET11a	Plasmid encoding Rubisco large subunit from *Synechococcus* sp. PCC6301 in pET11a expression vector (Saschenbrecker et al. 2007)
Syn6301-*rbcS*-pET11a	Plasmid encoding Rubisco small subunit from *Synechococcus* sp. PCC6301 in pET11a expression vector (Saschenbrecker et al. 2007)
Syn7002-*rbcL*-pET11a	Plasmid encoding Rubisco large subunit from *Synechococcus* sp. PCC7002 in pET11a expression vector (Saschenbrecker et al. 2007)
Syn7002-*rbcS*-pET11a	Plasmid encoding Rubisco small subunit from *Synechococcus* sp. PCC7002 in pET11a expression vector (Saschenbrecker et al. 2007)
Syn7002-*rbcL*/Syn7002-*rbcX$_{N-6His}$*-pET28b	Coexpression plasmid encoding *Synechococcus* sp. PCC7002 Rubisco large subunit and RbcX$_{N-6His}$ in pET28b expression vector (Liu et al. 2010)
Syn7002-*rbcL*/Syn7002-*rbcX(K5TAG)$_{N-6His}$*-pET28b	Coexpression plasmid encoding *Synechococcus* sp. PCC7002 Rubisco large subunit and RbcX$_{N-6His}$ with TAG mutation at position K5 in pET28b expression vector (Liu et al. 2010)
Syn7002-*rbcL*/Syn7002-*rbcX(L25TAG)$_{N-6His}$*-pET28b	Coexpression plasmid encoding of *Synechococcus* sp. PCC7002 Rubisco large subunit and RbcX$_{N-6His}$ with TAG mutation at position L25 in pET28b expression vector (Liu et al. 2010)
Syn7002-*rbcL*/Syn7002-*rbcX(Q28TAG)$_{N-6His}$*-pET28b	Coexpression plasmid encoding *Synechococcus* sp. PCC7002 Rubisco large subunit and RbcX$_{N-6His}$ with TAG mutation at position Q28 in pET28b expression vector (Liu et al. 2010)
Syn7002-*rbcL*/Syn7002-*rbcX(T33TAG)$_{N-6His}$*-pET28b	Coexpression plasmid encoding *Synechococcus* sp. PCC7002 Rubisco large subunit and RbcX$_{N-6His}$ with TAG mutation at position T33 in pET28b expression vector (Liu et al. 2010)
Syn7002-*rbcL*/Syn7002-*rbcX(G42TAG)$_{N-6His}$*-pET28b	Coexpression plasmid encoding *Synechococcus* sp. PCC7002 Rubisco large subunit and RbcX with TAG mutation at position G42 and N-terminal 6His tag in pET28b expression vector (Liu et al. 2010)
Syn7002-*rbcL*/Syn7002-*rbcX(L55TAG)$_{N-6His}$*-pET28b	Coexpression plasmid encoding *Synechococcus* sp. PCC7002 Rubisco large subunit and RbcX$_{N-6His}$ with TAG mutation at position L55 in pET28b expression vector (Liu et al. 2010)
Syn7002-*rbcL*/Syn7002-*rbcX(E58TAG)$_{N-6His}$*-pET28b	Coexpression plasmid encoding *Synechococcus* sp. PCC7002 Rubisco large subunit and RbcX$_{N-6His}$ with TAG mutation at position E58 in pET28b expression vector (Liu et al. 2010)
Syn7002-*rbcL*/Syn7002-*rbcX(K65TAG)$_{N-6His}$*-pET28b	Coexpression plasmid encoding *Synechococcus* sp. PCC7002 Rubisco large subunit and RbcX$_{N-6His}$ with TAG

Plasmid name	Description
	mutation at position K65 in pET28b expression vector (Liu et al. 2010)
Syn7002-*rbcL*/Syn7002-*rbcX(E66TAG)*_{N-6His}-pET28b	Coexpression plasmid encoding *Synechococcus* sp. PCC7002 Rubisco large subunit and RbcX$_{N-6His}$ with TAG mutation at position E66 in pET28b expression vector (Liu et al. 2010)
Syn7002-*rbcL*/Syn7002-*rbcX(V74TAG)*_{N-6His}-pET28b	Coexpression plasmid encoding *Synechococcus* sp. PCC7002 Rubisco large subunit and RbcX$_{N-6His}$ with TAG mutation at position V74 in pET28b expression vector (Liu et al. 2010)
Syn7002-*rbcL*/Syn7002-*rbcX(E80TAG)*_{N-6His}-pET28b	Coexpression plasmid encoding *Synechococcus* sp. PCC7002 Rubisco large subunit and RbcX$_{N-6His}$ with TAG mutation at position E80 in pET28b expression vector (Liu et al. 2010)
Syn7002-*rbcL*/Syn7002-*rbcX(K95TAG)*_{N-6His}-pET28b	Coexpression plasmid encoding *Synechococcus* sp. PCC7002 Rubisco large subunit and RbcX$_{N-6His}$ with TAG mutation at position K95 in pET28b expression vector (Liu et al. 2010)
Syn7002-*rbcL*/Syn7002-*rbcX(L106TAG)*_{N-6His}-pET28b	Coexpression plasmid encoding *Synechococcus* sp. PCC7002 Rubisco large subunit and RbcX$_{N-6His}$ with TAG mutation at position L106 in pET28b expression vector (Liu et al. 2010)

The plasmids listed below were used in this study and were constructed by Dr. C. Liu:

Plasmid name	Description
Cr-*rbcL*-pET11a	Plasmid encoding Rubisco large subunit from *Chlamydomonas reinhardtii* in pET11a expression vector
Cr-*rbcS2*-pET11a	Plasmid encoding Rubisco small subunit 2 from *Chlamydomonas reinhardtii* lacking 45 residues of the transit peptide in pET11a expression vector
Cr-*rbcX*	Plasmid encoding Rubisco assembly chaperone RbcX from *Chlamydomonas reinhardtii* lacking 34 residues of the transit peptide
Syn6301-*rbcL*-pHUE	Plasmid encoding Rubisco large subunit from *Synechococcus* sp. PCC6301 in pHUE expression vector
Syn6301-*rbcX*_{N-FLAG}-pET11a	Plasmid encoding Rubisco assembly chaperone RbcX$_{N-FLAG}$ from *Synechococcus* sp. PCC6301 in pET11a expression vector (Liu et al. 2010)
Syn6301-*rbcX(Y17A,Y20L)*_{N-FLAG}-pET30b	Plasmid encoding Rubisco assembly chaperone RbcX(Y17A,Y21L)$_{N-FLAG}$ from *Synechococcus* sp. PCC6301 in pET30b expression vector (Liu et al. 2010)
Syn6301-*rbcX(Q29A)*_{N-FLAG}-pET30b	Plasmid encoding Rubisco assembly chaperone RbcX(Q29A)$_{N-FLAG}$ from *Synechococcus* sp. PCC6301 in pET30b expression vector (Liu et al. 2010)

The plasmids listed below were used in this study and were constructed by Dr. K. Vasudeva Rao and Dr. B. Vasudeva Rao:

Plasmid name	Description
At-*rbcL*-pET11a	Plasmid encoding Rubisco large subunit from *Arabidopsis thaliana* in pET11a expression vector (Rao 2009a; Rao 2009b)
At-*rbcS1A*-pET11a	Plasmid encoding Rubisco small subunit S1A from *Arabidopsis thaliana* lacking 55 residues of the transit peptide in pET11a expression vector (Rao 2009a; Rao 2009b)
At-*rbcXA1*$_{N\text{-}FLAG}$-pHUE	Plasmid encoding Rubisco assembly chaperone RbcX$_{N\text{-}FLAG}$ A1 from *Arabidopsis thaliana* lacking 82 residues of the transit peptide in pHUE expression vector (Rao 2009a; Rao 2009b)

Plasmids constructed for this study are listed in appendix 7.1 and their construction is further described below in section 3.3.7.

3.1.5 Enzymes, proteins, peptides, and antibodies

Amersham Bioscience (Freiburg, Germany): RNAguard Ribonuclease Inhibitor (Porcine)

Invitrogen (Karlsruhe, Germany): Protein marker for SDS-PAGE

JPT Peptide Technologies GmbH (Berlin, Germany): PepSpot Peptides Epitope mapping

Merck (Darmstadt, Germany): Benzonase

MPI of Biochemistry (Martinsried, Germany): antisera (produced in rabbits) against purified AnaCA-RbcX$_{N\text{-}FLAG}$, Chlamy-RbcL, Syn6301-RbcL, and Syn7002-RbcX

MPI of Biochemistry, Department of Cellular Biochemistry (Martinsried, Germany): purified protein stocks of GroEL, GroES, and Rr-RbcL

MPI of Biochemistry, Peptide Synthesis Service (Martinsried, Germany): oligopeptides

New England Biolabs (Frankfurt a. Main, Germany): Prestained protein marker for SDS-PAGE, restriction endonucleases, T4 DNA ligase

Promega (Mannheim, Germany): *Pfu* DNA polymerase

Roche (Basel, Switzerland): Hexokinase, shrimp alkaline phosphatase

Sigma-Aldrich (Steinheim, Germany): BSA, HRP-conjugated secondary antibodies, EZview Red ANTI-FLAG M2 Affinity Gel, lysozyme, mouse monoclonal anti-FLAG M2 antibody, subtilisin

3.1.6 Media

Media were prepared with deionized double-distilled water and autoclaved following preparation.

LB medium: 10 g/l tryptone, 5 g/l yeast extract, 10 g/l NaCl, (15 g/l agar for solid medium)

3.1.7 Buffers and Standard Solutions

Buffers were prepared with deionized water with 18.2 MΩ·cm electrical resistance.

Antibiotic solutions (1000x): 100 mg/ml ampicillin, 20 mg/ml chloramphenicol, 30 mg/ml kanamycin

Buffer A: 50 mM Tris-HCl, pH 8.0, 300 mM NaCl, 10 mM KCl, 2 mM Mg-acetate, 10 mM β-mercaptoethanol, 10 mM imidazole

Buffer B: 50 mM Tris-HCl, pH 8.0, 10 mM KCl, 2 mM Mg-acetate, 5% glycerol

Buffer C: 40 mM Tris-HCl, pH 8.0, 0.25 M sucrose, 1% (v/v) Triton X-100

Buffer D: 40 mM Tris-HCl, pH 8.0, 0.25 M sucrose, 10 mM EDTA, 1% Triton X-100, 2 M Urea

Buffer E: 40 mM Tris-HCl, pH 8.0, 0.25 M sucrose, 10 mM EDTA

Buffer F: 40 mM Tris-HCl, pH 8.0, 6 M GdnHCl, 1 mM EDTA, 5 mM DTT

Buffer G: 50 mM Tris-HCl, pH 8.0, 50 mM NaCl, 10 mM $MgCl_2$

Buffer H: 20 mM Tris-HCl, pH 8.0, 20 mM NaCl, 1 mM EDTA

Buffer I: 20 mM imidazole-HCl, pH 6.4, 20 mM NaCl

Buffer J: 20 mM Tris-HCl, pH 8.0, 50 mM NaCl, 10 mM $MgCl_2$

Buffer K: 20 mM Tris-HCl, pH 7.5, 500 mM NaCl

Buffer L: 50 mM $NaHPO_4$-HCl, pH 8.0, 500 mM NaCl, 0.25 M sucrose

Buffer M: 20 mM bicine-NaOH, pH 9.0, 50 mM NaCl

Buffer N: 20 mM Tris-HCl, pH 9.0, 250 mM NaCl, 50 mM $NaHCO_3$, 4 mM $MgCl_2$

Buffer O: 20 mM Tris-HCl, pH 9.0, 100 mM NaCl, 1 mM EDTA

Buffer P: 50 mM NaH_2PO_4-HCl, pH 8.0, 20 mM NaCl, 1 mM EDTA

Buffer Q: 750 mM K_3PO_4-HCl, pH 7.0, 1 mM PMSF, 1 mM benzamide, 1 mM aminocaproic acid, 1 mg/ml lysozyme

^{14}**C mix**: Rubisco assay buffer, 1 mM DTT, 300 mM $NaHCO_3$, 50 mM $MgCl_2$, 10 μCi $NaH^{14}CO_3$

CN Bis-tris gel buffer (3x): 150 mM bis-tris-HCl, pH 7.0, 1.5 M 6-aminocaproic acid

CN Bis-tris cathode buffer (5x): 60 mM bis-tris-HCl, pH 7.0, 250 mM tricine

CN Bis-tris anode buffer (5x): 250 mM bis-tris-HCl, pH 7.0

CN Bis-tris-sample buffer (2x): 100 mM bis-tris-HCl, pH 7.0, 1 M 6-aminocaproic acid, 30% (v/v) glycerol, 2 mM $MgCl_2$, 20 mM KCl, 2x Complete protease inhibitor EDTA-free (Roche), 0.008% Ponceau-S

Coomassie destaining solution: 10% (v/v) ethanol, 7% (v/v) acetic acid

Coomassie staining solution: 0.1% (w/v) Serva Coomassie Blue R-250, 40% (v/v) ethanol, 8% (v/v) acetic acid

Denaturation buffer: 6 M GdnHCl, 10 mM DTT in Refolding buffer

DNA-loading buffer (6x): 10 mM Tris-HCl pH 7.6, 0.03% (w/v) bromophenol blue, 0.03% (w/v) xylene cyanol FF, 60% (v/v) glycerol, 60 mM EDTA

ECL PBS (10x): 48.3 mM Na_2HPO_4 x 2 H_2O, 27 mM KCl, 14 mM KH_2PO_4, 1.37 M NaCl

ECL PBS-T: 1x PBS, 0.1% (v/v) Tween-20

ECL PBS-TM: 1x PBS-T, 5% (w/v) milk powder

Electrophoresis migration buffer (10x): 250 mM tris, 1.92 M glycine, 1% SDS was additionally added for SDS-PAGE

Flag buffer: 50 mM Tris-HCl pH 8.0, 150 mM NaCl, 5 mM $MgCl_2$, 1mM EDTA, 0.1% (v/v) Triton X-100

Malachite green reagent: 350 mg malachite green in 75 ml water, 10.5 g ammonium molybdate in 250 ml 4 N HCl, 0.1% (v/v) triton X-100

Native-sample buffer (2x): 50% (v/v) glycerol, 0.25% (w/v) bromophenol blue, in electrophoresis migration buffer without SDS

PBS (10x): 92 mM Na_2HPO_4 x 2 H_2O, 16 mM NaH_2PO_4 x H_2O, 1.5 M NaCl, pH 7.2 with NaOH

Ponceau S stain: 0.1% (w/v) Ponceau S, 5% (v/v) acetic acid

Refolding buffer: 20 mM MOPS-KOH, pH 7.5, 100 mM KCl, 5 mM Mg-acetate

Regeneration buffer I: 62.5 mM Tris-HCl pH 6.7, 2% (w/v) SDS, 100 mM β-mercaptoethanol

Regeneration buffer IIA: 8 M Urea, 1% (w/v) SDS, 0.1% β-mercaptoethanol

Regeneration buffer IIB: 400 ml H_2O, 500 ml ethanol, 100 ml acetic acid

Rubisco assay buffer: 100 mM Tris-HCl, pH 7.5, 10 mM KCl, 2 mM Mg-acetate

SDS-sample buffer (2x): 125 mM Tris-HCl pH 6.8, 20% (v/v) glycerol, 4% (w/v) SDS, 10% (v/v) β-mercaptoethanol, 0.005% bromophenol blue

Strip solution: 0.2 M NaOH

Subtilisin protease buffer: 20 mM HEPES-NaOH, pH 7.5, 50 mM NaCl

TAE buffer (50x): 2 M Tris, 5.71% (v/v) glacial acetic acid, 50 mM EDTA

TBS (10x): 500 mM Tris-HCl pH 8.0, 1.37 M NaCl, 27 mM KCl

T-TBS (10x): 200 mM Tris-HCl pH 7.5, 1.37 M NaCl, 1 % (v/v) Tween-20

1xTBSM: 1x TBS, 1% (w/v) milk powder

1xT-TBSM: 1x T-TBS, 1% (w/v) milk powder

TFB I (Transformation buffer I): 30 mM K-acetate pH 5.8 (acetic acid), 50 mM $MnCl_2$, 100 mM RbCl, 10 mM $CaCl_2$, 15% (v/v) glycerol

TFB II (Transformation buffer II): 10 mM MOPS-NaOH pH 7.0, 75 mM $CaCl_2$, 10 mM $RbCl_2$, 15% (v/v) glycerol

TNME buffer: 50 mM Tris-HCl, pH 8.0, 20 mM NaCl, 5 mM $MgCl_2$, 1 mM EDTA

TSS solution: 10% (v/v) PEG600, 5% (v/v) DMSO, 50 mM $MgCl_2$, in LB medium

Western blot buffer: 50 mM Tris, 20% (v/v) methanol, 192 mM glycine

3.2 Instruments

Abimed (Langenfeld, Germany): Gilson Pipetman (2, 10, 20, 100, 200, 1000 µl)

Bachofer (Reutlingen, Germany): Hybridization Oven

Beckmann Coulter GmbH (Krefeld, Germany): centrifuges (J6-MI, GS-6R, Avanit 30, Avanit J-25I, Optima LE80K ultra centrifuge), spectrophotometers (DU640, DU800), LS 6500 mulit-purpose scintillation counter

Biometra (Göttingen, Germany): PCR thermocycler

Bio-Rad (Munich, Germany): ChemiDoc, Gene Pulser Xcell electroporation system, electroporation cuvettes, horizontal agarose gel electrophoresis system (Wide) Mini-SUB CELL GT, Mini-PROTEAN electrophoresis chambers, Model 583 gel dryer, Power Pac 300

Branson (Connecticut, USA): Sonifer cell disrupter B15

DeLano Scientific LLC (San Francisco, USA): PyMOL Executable Build

Drummond Scientific (Broomall / PA, USA): Pipet-aid

Eppendorf (Hamburg,Germany): centrifuges (5415D, 5417R), Thermomixer comfort

Fisher Scientific (Schwerte, Germany): Accumet Basic pH meter

Forma Scientific (Marietta / OH, USA): Orbital Shaker 4581

Fuji/Raytest (Straubenahrdt, Germany): Fuji-LAS3000 luminescence and densitometry system, Image Reader Fuji-FLA2000, Phosphoimager FLA2000, software AIDA v.3.52.046

GE Healthcare (Munich, Germany): Äkta Explorer, Äkta Purifier, prepacked chromatography columns

Hampton Research (Aliso Viejo / CA, USA): VDX Plates

Hoefer Scientific Instruments (San Francisco, USA): SemiPhor blotting systems

Invitrogen (Karlsruhe, Germany): XCell SureLock Mini-Cell, NuPAGE 4-12% Bis-Tris Gel 1.0 mm, 10 well, NuPAGE buffers

LOT-Oriel (Darmstadt, Germany): 500W Mercury Arc lamp

Mettler Toledo (Geiβen, Germany): Balances (AG285, PB602)

MicroCal (Northampton / MA, USA): MicroCal VP-ITC MicroCalorimeter

Millipore (Bedford / MA, USA): Amicon ultra centrifugal filter, Millex-HA filters, Steritop vacuum filters

Misonix (Farmingdale / NY, USA): Sonicator 3000

MPI of Biochemistry (Martinsried, Germany): Western blot system (semi-dry)

MWG Biotech AG (Ebersberg, Germany): Gel documentation system

New Brunswick Scientific (Nürtigen, Germany): Innova 4430 incubator

Pall (Portsmouth, England): PVDF membrane

Scientific Industries, Inc. (Bohemia / NY, USA): Vortex-Genie 2

Savant/Thermoquest (Engelsback, Germany): Stacked Gel Dryer SGD300

Whatman GmbH (Dassel, Germany): Whatman Klari-flex, Whatman Protran nitrocellulose transfer membrane

WTW (Weilheim, Germany): pH meter

Wyatt Technology (Santa Barbara / CA, USA): FFF-MALS system, software ASTRA

3.3   Molecular biological methods

3.3.1 DNA analytical methods

UV spectrometry was used to measure DNA concentration at a wavelength of 260 nm, where one absorption unit corresponds to 50 ng/µl double stranded DNA in water. DNA-free water was used as a reference. Impurities present in the DNA can be determined by the 260/280 nm absorbance ratio, which should be approximately 1.8; deviations from this absorbance ratio indicate contaminations such as RNA or protein (Sambrook et al. 1989).

DNA samples were run on 1% (w/v) agarose in 1x TAE buffer supplemented with SYBR safe for DNA visualization. Before loading on the agarose gel, the DNA samples were mixed with 6x DNA-loading buffer. Agarose gel electrophoresis was carried out in 1x TAE at a constant voltage of 80 V. Cloned plasmid constructs were verified by restriction enzyme digest analysis and DNA-sequencing of the protein-coding regions by the Core Facility at the MPI of Biochemistry in Martinsried, Germany. The sequence chromatograms were visualized by the program Chromas (C. McCarthy, Griffith University, Australia) and the program

MultAlin (http://multalin.toulouse.inra.fr/multalin/multalin.html) was used for comparative sequence alignment analyses (Corpet 1988).

3.3.2 Competent *E. coli* cell preparation and transformation

3.3.2.1 Chemocompetent *E. coli* cells and chemical transformation

The RbCl method was used to prepare chemically competent *E. coli* cells (Hanahan 1983). 500 ml LB medium were inoculated with 10 ml of overnight culture of the respective *E. coli* strain, and incubated at 37°C to an OD_{600} of 0.5. Cells were then kept for 15 min on ice followed by centrifugation for 10 min at 1000 g and 4°C in sterilized centrifugation tubes. Supernatant was removed and pellet kept on ice. Cells were resuspended in 125 ml pre-cooled TFB I and incubated on ice for an additional 20 min. Subsequently, the cells were centrifuged again in sterile centrifugation tubes for 10 min at 1000 g and 4°C. Following supernatant decantation, the harvested cells were placed on ice and resuspended in 20 ml of pre-cooled TFB II. While keeping the cells on ice, 50 µl aliquots were added to pre-chilled tubes, frozen on dry ice, and stored at -80°C.

The prepared 50 µl chemocompetent cell aliquots were then used for heat shock transformation purposes. The 50 µl aliquots were thawed on ice and 50-100 ng of plasmid DNA or 10 µl of ligation reaction was added. The cells were incubated on ice for 30 min, followed by a 90 s heat shock at 42°C. After heat shock, the cells were once again incubated on ice for 5 min. 800 µl of LB was added to the cells and incubated for 1 hr at 37°C with gentle shaking. The cells were then plated on selective LB agar plates and incubated overnight at 37°C.

3.3.2.2 Electrocompetent *E. coli* cells and electroporation

Electrocompetent *E. coli* cells were made by inoculating 500 ml of LB medium with 2.5 ml overnight culture of the respective strain. Cells were grown to OD_{600} of 0.5 followed by incubation on ice for 15 min. Next, cells were centrifuged in sterilized centrifuge tubes for 15 min at 4000 rpm (Beckmann centrifuge Avanti JLA 10.500 rotor) and 4°C. Supernatant was removed and harvested cells were resuspended in 500 ml cold sterile water followed by centrifugation as above. This washing step was repeated. After the second washing step, the cells were then resuspended in 40 ml pre-cooled, sterile 10% (v/v) glycerol. Centrifugation was carried out for 10 min at 3500 rpm (Beckmann centrifuge Avanti JLA 10.500) and 4°C. Harvested cells were then resuspended in 1 ml of pre-cooled, sterile 10% (v/v) glycerol and 50 µl was aliquoted to pre-chilled tubes, frozen on dry ice, and stored at -80°C.

Electroporation of the 50 µl electrocompetent *E. coli* cells was initiated by thawing the cells on ice and adding 50 ng of plasmid DNA to the cells. This was then transferred to a cooled 1 mm electroporation cuvette and pulsed (1250 V, 25 µF, 200 Ω). Directly after the pulse, the cells were taken up in 800 µl LB medium and incubated for 1 hr at 37°C with gentle shaking. The cells could then be plated on selective LB agar plates and incubated overnight at 37°C.

3.3.2.3 TSS transformation

This method was used for fast transformation purposes (Chung et al. 1989) and for transformations of more than one plasmid DNA in *E. coli* cells. The respective *E. coli* strain, potentially already transformed with one DNA plasmid, was plated on selective LB agar plates and grown overnight at 37°C; a single colony from the plate was used to inoculate 5 ml of LB medium containing respective antibiotic and accordingly incubated at 37°C to an OD_{600} of 0.4. An aliquot of 2 ml was centrifuged at 16,000 g for 1 min at RT; the supernatant was decanted and the resulting pellet was resuspended in 100 µl TSS solution containing 0.5-1 µg of plasmid DNA. After incubation on ice for 30 min, 0.5 ml of LB medium was added and the culture was incubated at 37°C with agitation. The cells could then be plated on selective LB agar plates and incubated overnight at 37°C.

3.3.3 Plasmid DNA and DNA fragment purification

E. coli DH5α was used for the amplification of plasmid DNA. The plasmid DNA was isolated and purified from overnight cultures grown in LB medium with appropriate antibiotics (Sambrook et al. 1989) by anion exchange chromatography using the Wizard *Plus* SV Miniprep DNA Purification System or the PureYield Plasmid Midiprep System according to the manufacturer's instructions.

Plasmid DNA or DNA fragments were purified after enzymatic reactions (e.g. after restriction digest, see section 3.3.6) or agarose gel electrophoresis (e.g. after PCR, see section 3.3.4) by anion exchange chromatography using the Wizard SV Gel and PCR Clean-Up System. Purification and isolation were performed according to the manufacturer's instructions.

3.3.4 PCR (polymerase chain reaction)

Polymerase chain reaction was used to amplify DNA target sequences according to the standard protocol and cycling conditions indicated below in Table 3-1 and Table 3-2. When

necessary, appropriate modifications or adjustments were made to the standard protocol and/or PCR thermocycling conditions. Primer pairs used for PCR are listed in appendix 7.2.

Total volume 50 µl	
Buffer	1x *Pfu* Polymerase buffer
dNTPs	0.25 mM
Primers	40 pmol each
Polymerase	3.75 U *Pfu* DNA Polymerase
Template DNA	100-200 ng plasmid or genomic DNA

Table 3-1 Standard PCR amplification protocol

Lid temperature of PCR thermocycler: 105°C				
Step	**Purpose**	**Temperature**	**Duration**	**Number of Cycles**
1	Initial denaturation	95°C	300 s	1
2	Cycle denaturation	95°C	60 s	
3	Primer annealing	58°C	45 s	35 (steps 2-4)
4	Primer extension	72°C	2 min/kb	
5	Final extension	72°C	600 s	1
6	Cooling	4°C	∞	1

Table 3-2 Thermocycling conditions for PCR amplification

3.3.5 Site-directed mutagenesis

Whole plasmid PCR was used for mutation introduction in template DNA. A complete list of mutant plasmids constructed in this study can be found in appendix 7.1. Generally, the standard protocol and thermocycling conditions listed in Table 3-3 and Table 3-4, respectively, were used for site-directed mutagenesis; modifications to the procedure were made when necessary. Primer pairs used for site-directed mutagenesis are listed in appendix 7.2. Following whole plasmid PCR and for processing purposes of the resulting constructs, 10 U of *DpnI* was added to the sample and incubated for 1 hr at 37°C. *DpnI* specifically digests methylated DNA; therefore, exclusively parental template DNA strands were removed from the sample (Weiner et al. 1994). The entire sample was accordingly transformed into chemocompetent *E. coli* DH5α cells.

Total volume 20 µl	
Buffer	1x *Pfu* Polymerase buffer
dNTPs	0.2 mM
Primers	40 pmol each
Polymerase	1.2 U *Pfu* DNA Polymerase
Template DNA	50 ng of plasmid DNA

Table 3-3 Standard protocol for site-directed mutagenesis

Lid temperature of PCR thermocycler: 105°C				
Step	**Purpose**	**Temperature**	**Duration**	**Number of Cycles**
1	Initial denaturation	95°C	30 s	1
2	Cycle denaturation	95°C	30 s	
3	Primer annealing	55°C	60 s	18 (steps 2-4)
4	Primer extension	68°C	2 min/kb	
5	Final extension	68°C	600 s	1
6	Cooling	4°C	∞	1

Table 3-4 Thermocycling conditions for site-directed mutagenesis

3.3.6 Restriction digest and DNA ligation

Amplified PCR product or vector containing gene of interest and expression vector of interest were digested with appropriate restriction enzymes in the respective buffer conditions as follows: 30 µl total reaction volume, 3 µl of appropriate buffer, 1 µl of restriction enzymes, 15 µl water, and 10 µl plasmid. Samples were incubated for 1 hr at 37°C. The vector was further dephosphorylated with 1 µl shrimp alkaline phosphorylase for an additional 1 hr at 37°C. The digests were purified by running the samples on 1% agarose followed by band extraction as described above (section 3.3.3).

Following purification and extraction of the vector and insert bands, ligation was conducted to ligate the gene of interest with the expression vector. Conditions for ligation were as follows: 20 µl total reaction volume, 2 µl of vector, 4 µl insert, 2 µl T4 ligase buffer, 1 µl T4 ligase, and 11 µl water with incubation at room temperature for 1 hr. To control for complete dephosphorylation of the expression vector, ligation was also carried out with no addition of insert DNA in the ligation reaction, water was added instead. 10 µl of ligation was heat shock transformed into DH5α *E. coli* competent cells. All of the transformed cells were plated on selective agar plates and incubated overnight at 37°C. Correct constructs were

determined by isolating the plasmid DNA from colonies on the selective plates followed by restriction analysis and sequencing of the purified plasmid DNA at the sequencings services of the MPI of Biochemistry core facility in Martinsried, Germany. The sequence chromatograms were visualized by the program Chromas (C. McCarthy, Griffith University, Australia) and the program MultAlin (http://multalin.toulouse.inra.fr/multalin/multalin.html) was used for comparative sequence alignment analyses (Corpet 1988).

3.3.7 Cloning strategies

The complete list of plasmids constructed in this study and primer pairs used can be found in appendices 7.1 and 7.2, respectively. The *rbcX* gene was amplified by PCR, using either the plasmid AnaCA-*rbcX*-pET11a or Syn6301-*rbcX*$_{N\text{-}FLAG}$-pET11a as a template, via primers flanked with *NdeI* and *BamHI* restriction sites. To replace the TAG stop codon in AnaCA-*rbcX* with the TAA stop codon (for site-specific crosslinking purposes), the necessary mutation was present in the reverse primer. An *NheI* restriction site was also present after the stop codon in the *BamHI* reverse flanking primer. Using the *NdeI* and *BamHI* restriction sites, the amplified PCR product was inserted into plasmid pET28b, downstream of the T7 promoter sequence, resulting in N-terminal 6-His tagged constructs: AnaCA-*rbcX(TAA)*$_{N\text{-}6His}$-pET28b, AnaCA-*rbcX*$_{N\text{-}6His}$-pET28b or Syn6301-*rbcX*$_{N\text{-}6His}$-pET28. Syn6301-*rbcL* could then be inserted downstream of *rbcX*$_{N\text{-}6His}$, but upstream of T7 terminator by sequential digestion using *NheI* (*XbaI* for Syn6301-*rbcL*) and *BamHI* restriction sites resulting in the coexpression plasmids AnaCA-*rbcX(TAA)*$_{N\text{-}6His}$/Syn6301-*rbcL*-pET28b, AnaCA-*rbcX*$_{N\text{-}6His}$/Syn6301-*rbcL*-pET28b, or Syn6301-*rbcX*$_{N\text{-}6His}$/Syn6301-*rbcL*-pET28b.

Whole plasmid site-directed mutagenesis was then used to point mutate the residue of interest in the previously listed plasmids as well as in the plasmids Syn6301-*rbcL*-pET11a, Syn7002-*rbcL*/Syn7002-*rbcX*$_{N\text{-}6His}$-pET28b, Syn7002-*rbcL*-pET11a, Syn6301-*rbcX*$_{N\text{-}FLAG}$/Syn6301-*rbcL*-pET11a, Syn6301-*rbcS*-pET11a according to the methods described in section 3.3.5. N-terminally truncated *rbcL* was cloned by utilizing modified forward primers, having an *NdeI* restriction site. The PCR product was then inserted in plasmid pET11a, followed by restriction enzyme digest and ligation downstream of *rbcX*$_{N\text{-}FLAG}$ as described above (Syn6301-*rbcX*$_{N\text{-}FLAG}$-pET11a also contains the *NheI* and *BamHI* sites).

The gDNA from *Anabaena* sp PCC7120 was used to amplify the chaperonin and cochaperone genes Ana7120-*groEL1*, Ana7120-*groEL2*, and Ana7120-*groES*. Flanking primers with *NdeI* and *BamHI* restriction sites were used. The amplified products were inserted in the pET11a or pCOLAduet-1 plasmids resulting in Ana7120-*groEL1*-pET11a,

Ana7120-*groEL2*-pET11a, Ana7120-*groEL2*-pCOLAduet-1, and Ana7120-*groES*-pET11a. The plasmids Synpcc7942_2313(GroEL1)-pET21a and Synpcc7942_2314(GroES)-pET21a were used to amplify the Syn6301 chaperonin Syn6301-*groEL1* and cochaperone Syn6301-*groES* genes also using flanking primers with *NdeI* and *BamHI* restriction sites. The amplified products could be inserted in the pET11a plasmid, using *NdeI* and *BamHI* restriction sites, producing constructs Syn6301-*groEL1*-pET11a and Syn6301-*groES*-pET11a.

Integrity of constructs generated in this study was confirmed by sequencing (core facility sequencing service at the MPI of Biochemistry in Martinsried, Germany). Comparative sequence analysis was performed using the program MultAlin; http://bioinfo.genopoletoulouse.prd.fr/multalin/multalin.html (Corpet 1988).

3.4   Protein biochemical and biophysical methods

3.4.1   Protein analytical methods

3.4.1.1 Protein quantification and sequence alignments

The concentration of purified protein was determined spectrophotometrically by measuring the absorption at 280 nm; protein-free buffer was used as a reference. Sample concentration could be determined from the theoretical molar extinction coefficient (M^{-1} cm^{-1}) of the particular protein sequence using the ExPASy Proteomics Server ProtParam function (http://expasy.org/tools/protparam.html). Additionally, concentration of lysate samples or samples containing a mixture of proteins was determined colorimetrically by Bradford assay using Bio-Rad protein assay reagent at 595 nm according to manufacturer's procedures (Bradford 1976). This method takes advantage of the absorbance maximum shift coomassie blue dye exhibits when bound to protein; the intensity of absorbance at 595 nm is directly proportional to the amount of protein present in the sample. Protein concentration of native proteins is given for their respective oligomeric state unless otherwise stated; whereas, the concentration of denatured proteins is in the monomeric state. Comparative sequence analyses were performed using the program MultAlin; http://bioinfo.genopoletoulouse.prd.fr/multalin /multalin.html (Corpet 1988).

3.4.1.2 SDS-PAGE

In order to separate proteins on a gel matrix based on molecular weight, SDS-PAGE (sodium dodecyl sulfate polyacrylamide gel electrophoresis) was implemented. Protein analysis was, accordingly, performed under denaturing and reducing conditions (Laemmli 1970). Gel preparation was carried out as described in Table 3-5; furthermore, isopropanol

was layered over the resolving gel directly after casting to create an even plane between the resolving and stacking gels. Samples were mixed with 2x SDS-sample buffer, boiled for 5 min at 95°C, and centrifuged for 1 min at 16,000 g. Gel electrophoresis was completed in Mini-PROTEAN electrophoresis chambers in the presence of Electrophoresis migration buffer (with SDS added) employing a starting constant voltage of 100 V followed by 160 V (upon soluble front migration into the resolution gel) for the rest of the run.

Solutions (for 4 gels)	Resolution gel			Stacking gel	Final buffer concentration
	10% AA	12.5% AA	16% AA	4% AA	
ddH$_2$O	8 ml	6.4 ml	4.1 ml	6.1 ml	
1.5 M Tris-HCl, pH 8.8	5 ml	5 ml	5 ml	--	375 mM
0.5 M Tris-HCl, pH 6.8	--	--	--	2.5 ml	125 mM
10% (w/v) SDS	200 µl	200 µl	200 µl	100 µl	0.1%
30% AA/ 0.8% bis-AA	6.8 ml	8.4 ml	10.7 ml	1.3 ml	
10% (w/v) APS	100 µl	100 µl	100 µl	50 µl	0.05%
TEMED	10 µl	10 µl	10 µl	10 µl	0.05% & 0.1%

Table 3-5 SDS-PAGE gel preparation

3.4.1.3 Gradient SDS-PAGE

Gradient SDS-PAGE also separates proteins based on their molecular weight; however, the gradient allows better separation of protein bands. Gradient SDS-PAGE gels were poured in either 20x20cm or 16x16 cm glass plates with 1 mm spacers. The bottom of the glass plates was sealed by addition of plug solution. After plug polymerization, a gradient maker was used to create the appropriate gradient. Gel solutions were prepared as described in Table 3-6. After addition of the gradient resolution gel by means of the gradient maker, isopropanol was layered on top directly after casting to create an even plane between the resolving and stacking gels. Samples were mixed with 2x SDS-sample buffer (with or without reducing agents). If reducing agents were present in the sample buffer, the samples were further boiled for 5 min at 95°C followed by centrifugation at 16,000 g. Gel electrophoresis was completed in a large gel electrophoresis chamber in the presence of Electrophoresis migration buffer (with SDS added). Electrophoresis was applied overnight at constant current of 5.5 mA.

Solutions (for 1 gel)	Resolution gel		Stacking gel	Plug
	10% AA	15% AA	4% AA	25% AA
ddH$_2$O	6.96 ml	--	7 ml	--
2 M Tris-HCl, pH 8.8	3 ml	3 ml	--	--
0.5 M Tris-HCl, pH 6.8	--	--	1.5 ml	--
2 M Sucrose	--	4.43 ml	2.25 ml	0.5 ml
25% SDS (w/v)	60 µl	60 µl	50 µl	--
30% AA/ 0.8% bis-AA	4.96 ml	7.5 ml	1.65 ml	2.5 ml
10% (w/v) APS	60 µl	60 µl	66.7 µl	27 µl
TEMED	9 µl	9 µl	6.3 µl	5 µl

Table 3-6 Gradient SDS-PAGE gel preparation

3.4.1.4 Native-PAGE

Native-PAGE was used for separation of protein complexes on a gel matrix under native conditions. In this method protein samples are not separated based on their molecular weight; instead, the distance of migration is dependent on the given protein's overall globular structure and sequence charge. Gels were prepared as described in Table 3-7, and isopropanol was layered over the resolving gel directly after casting to create an even plane between the resolving and stacking gels. Samples were mixed with 2x Native-sample buffer. Gel electrophoresis was completed in Mini-PROTEAN electrophoresis chambers in the presence of Electrophoresis migration buffer at 4°C employing a starting constant voltage of 100 V followed by 160 V (upon soluble front migration into the resolution gel) for the rest of the run.

Solutions (for 2 gels)	Resolution gel		Stacking gel	Final buffer concentration
	6% AA	10% AA	3% AA	
ddH$_2$O	4.27 ml	3.23 ml	3.22 ml	
1.5 M Tris-HCl, pH 8.8	2.1 ml	2.1 ml	--	394 mM
0.5 M Tris-HCl, pH 6.8	--	--	1.25 ml	125 mM
30% AA/ 0.8% bis-AA	1.6 ml	2.67 ml	0.5 ml	
10% (w/v) APS	64 µl	64 µl	68 µl	0.08%
TEMED	4.4 µl	4.4 µl	6.8 µl	0.06% & 0.1%

Table 3-7 Native-PAGE gel preparation

3.4.1.5 CN Bis-tris-PAGE

CN (clear native) Bis-tris-PAGE is similar to the above described method of Native-PAGE; however, discontinuous CN Bis-tris Native-PAGE allows enhanced resolution of higher molecular weight species. Gels were prepared with 20x20 cm glass plates and 1mm spacers. The bottom of the glass plates was sealed with thick, water-resistant tape before addition of the plug. After plug polymerization, a gradient maker was used to create the

appropriate gradient. Gel solutions were prepared as described in Table 3-8. After addition of the gradient resolution gel by means of the gradient maker, isopropanol was layered on top directly after casting to create an even plane between the resolving and stacking gels. Samples were mixed with 2x CN Bis-tris-sample buffer. Gel electrophoresis was completed in a large gel electrophoresis chamber in the presence of CN Bis-tris cathode and CN Bis-tris anode buffers. Electrophoresis was applied overnight at 4°C starting at a constant voltage of 100 V for 3 hr followed by 230 V for the rest of the run.

Solutions (for 1 gel)	Resolution gel		Stacking gel	Plug
	5% AA	13% AA	3% AA	9% AA
ddH$_2$O	5.4 ml	0.8 ml	3.4 ml	1.4 ml
Gel buffer (3x)	4 ml	4 ml	2 ml	--
Glycerol	0.6 ml	2 ml	--	--
30% AA/ 0.8% bis-AA	2 ml	5.2 ml	0.6 ml	0.6 ml
10% (w/v) APS	58 µl	42 µl	50 µl	20 µl
TEMED	5.8 µl	4.2 µl	5 µl	2 µl

Table 3-8 CN Bis-tris-PAGE gel preparation

3.4.1.6 Coomassie blue staining of polyacrylamide gels

Coomassie blue is a triphenylmethane coloring agent that binds to proteins. This dye integrates with cationic, nonpolar, and hydrophobic side chains of a polypeptide. Therefore, coomassie blue staining was used for the visualization of protein bands in polyacrylamide gels. Protein bands were fixed and stained in the gel matrix by incubating the gels for 1 hr with Coomassie staining solution. Background staining was removed by further incubation with Coomassie destaining solution.

3.4.1.7 Western blotting and immunodetection

Western blotting is a more sensitive method, as compared to coomassie blue staining, for the visualization of proteins separated on polyacrylamide gels. The method described for western blotting and immunodetection is based on a previously published report (Towbin et al. 1979) with modifications. Subsequent to electrophoresis, gels were incubated in Western blot buffer (0.1% SDS (w/v) supplementation for Native-PAGE gels to denature proteins for more efficient transfer) for 30 min. Protein bands were transferred to nitrocellulose membrane using a semi-dry western blot system (SemiPhor). Transfer was performed in Western blot buffer and accomplished by employing a constant current dependent on blot surface area (length x width x 0.8) for 1 hr. Transfer efficiency was determined by staining the membranes for 2 min with Ponceau S stain solution, followed by water washings to remove background

staining. Membranes were either stored overnight or immediately subjected to immunodetection.

Immunodetection began by blocking membrane for 1 hr with ECL PBS-TM solution. The membrane was subsequently incubated for 1 hr with 1:5,000 or 1:10,000 dilutions of primary antibody in ECL PBS-TM. Unbound antibody was removed by three washing steps (15 min, 5 min, and 5 min) with ECL PBS-T. After the washing step, the membrane was incubated with 1:5,000 or 1:10,000 dilutions of horseradish peroxidase (HRP)-conjugated secondary antibody in ECL PBS-TM for 1 hr. Unbound antibody was again removed by three washing steps (15 min, 5 min, and 5 min) with ECL PBS-T. Solutions for chemiluminescent detection (ECL I and ECL II) were prepared fresh in a 1:1 ratio, additionally, extra H_2O_2 was added to the mixture (0.02% (v/v)). Membranes were covered with the freshly prepared ECL solution and bands were detected and documented using Bio-Rad ChemiDoc system. When necessary, membranes were stripped with 0.2 M NaOH for 5 min followed by a 5 min wash with water.

3.4.1.8 Autoradiography

Autoradiography was utilized for the visualization of minuscule amounts of radioactively (^{35}S) labeled protein. Following electrophoresis of Native-PAGE or SDS-PAGE, gels were stained with coomassie and further vacuum dried on whatman paper; distance of standard protein migration was marked with small amounts of radioactive dye on the paper. The radioactive band pattern was transferred to a phosphorimaging plate during overnight incubation and the signals read by a laser scanner (Image Reader Fuji-FLA2000) and analyzed using AIDA software.

3.4.1.9 Generation of anti-serum

Production of polyclonal antibodies against AnaCA-RbcX$_{N-FLAG}$ was performed at the MPI of Biochemistry animal facilities Martinsried, Germany. Prior to injection, an emulsion of protein with adjuvant was prepared by mixing 500µl of 2 mg/ml purified AnaCA-RbcX$_{2(N-FLAG)}$ with 500µl of TiterMax Gold Adjuvant using a 3 ml syringe with an 18 gauge double hub emulsifying needle. 250 µl of the water-in-oil emulsion was used for subcutaneous injection in rabbit. This same procedure was conducted for three further injections in four week intervals with Freund's Adjuvant Incomplete. Test blood was obtained 10 days after injections.

Anti-serum was prepared from the test blood. The test blood was stirred for 1 min with a glass rod and then kept for 1 hr at room temperature. The blood was stirred again for 1 min followed by storage at 4°C overnight. The next morning the blood was mixed again and subsequently centrifuged for 30 min at 5,000 g. The supernatant (anti-serum) was collected and stored at -20°C. Specificity of the anti-serum was tested by immunodetection using purified protein or lysate samples blotted to nitrocellulose membrane.

3.4.1.10   TCA precipitation

Protein samples, having the same amount of total protein, were incubated for 15 min at 4°C with 13% (w/v) TCA (trichloroacetic acid), by adding a two-fold volume of 20% (w/v) TCA to protein solutions. Samples were centrifuged for 15 min at 16,000 g at 4°C and supernatant removed. Protein pellets were washed with 1 ml cold acetone, centrifuged for 10 min at 16,000 g at 4°C, and supernatant decanted. Cold acetone washing step was repeated and afterwards excess acetone evaporated. 1x SDS sample buffer was added; if the solutions appeared yellow, indicating incomplete removal of TCA, small amounts of 1 M Tris-HCl, pH 7.5 was added until the solutions turned blue. Samples were boiled for 5 min at 95°C followed by SDS-PAGE analysis.

3.4.1.11   SEC-MALS

SEC-MALS (size exclusion chromatography-multiangle light scattering), performed by Dr. Manajit Hayer-Hartl, was used for molecular mass determination of wildtype and mutant $RbcL_8/(RbcX_2)_8$ complexes. 60-125 µg of protein complexes were injected on a gel filtration column (TSK Super 3000 SW Size: 4.6 mm x 30 cm) using a 100 µl injection loop, 0.25 ml/min flow rate, and UV detector set to 500 mAU. The system was connected to DAWN EOS MALS detector (690 nm laser), variable wavelength detector (absorbance at 280 nm, Agilent 1100 series), and Optilab DSP refractive index detector (690 nm). The ASTRA program was used for the calculation of masses with the dn/dc value set to 0.185 ml/g. System was run in Buffer O, however without EDTA addition.

3.4.1.12   Proteolytic digestion and Edman degradation

Protein samples were subjected to subtilisin protease treatment to determine protein complex stability. 1 mg/ml of purified protein complexes were incubated with increasing concentrations (0.5, 1, 2, 5, 50, and 500 µg/ml) of subtilisin in Subtilisin protease buffer for 60 min on ice. Digestion was quenched with 10 mM PMSF. After proteolysis, samples were

subjected to SDS-PAGE and transferred to PVDF membrane via western blotting. Before transfer, the PVDF membrane was washed with methanol and equilibrated in Western blot buffer. Following transfer, the membrane was stained with Ponceau S stain or Coomassie blue for band visualization, and the band of interest was excised for Edman degradation analysis. N-terminal sequencing was carried out by R. Mentele at the MPI of Biochemistry Martinsried, Germany. This method was used to determine if the N-terminal of a given protein was cleaved during proteolytic digestion of protein samples.

3.4.1.13   Mass spectrometry

ESI-MS (electrospray ionization mass spectrometry) was implemented to determine the mass of subtilisin protease treated samples or to measure the mass of purified protein complexes. ESI-MS was performed by Dr. Yi-chin Tsai or E. Weyher (MPI of Biochemistry core facility) on quenched samples following proteolytic digestion or samples of purified protein. Mass spectrometry was also used for determination of protein composition of a specific band on NuPAGE Bis-Tris gels. The band of interest was digested by trypsin protease and the purified peptides were analyzed on MS by Dr. S. Schermann. NuPAGE Bis-Tris gels were run according to the manufacturer's recommendations.

3.4.1.14   CD (circular dichroism) spectroscopy

The secondary structure of purified $RbcX_2$ variants was determined by CD spectroscopy in 0.1 cm cuvettes using a Jasco 715 spectrometer with peltier-thermostat. A protein concentration of 0.1 mg/ml in 200 μl Buffer J was analyzed at 4°C. Wavelength scans ranging from 260-195 nm were conducted four times for each sample followed by baseline correction using a buffer sample with no protein. The data were processed with CDPro analysis and Spectra manager software and fitted using the program CONTIN.

3.4.2 Protein crystallization

Initial crystal screens were conducted with purified $At\text{-}RbcXI_2$ protein at the crystallization facility of the MPI of Biochemistry Martinsried, Germany. Conditions tested were the pHClear Suite and the pHClear II Suite at 4°C and room temperature using 150 nl drop size. However, these screens yielded no observable crystals, which is most likely due to the inability of the $At\text{-}RbcXI_2$ protein to be concentrated above 1 mg/ml (32 μM dimer).

Crystallization of the Syn6301-RbcL$_8$/(AnaCA-RbcX$_2$)$_8$ complex was performed by Dr. A. Bracher and preliminary crystal screening by Dr. S. Saschenbrecker (Bracher et al. 2011).

3.4.3 Protein expression and purification

Protein purification was performed at 4°C unless stated otherwise. All purification steps were monitored by SDS-PAGE and protein concentration measured by Bradford assay.

3.4.3.1 Syn6301-RbcL$_8$

E. coli BL21 (DE3) cells harboring the plasmids Syn6301-*rbcL*-pHUE and pBAD33*ES/EL* were grown at 37°C in LB supplemented with amp and cam until an OD$_{600}$ of 0.5; induction of GroEL/ES preceded RbcL expression in order to increase RbcL solubility. Arabinose was added to the cultures at a final concentration of 0.4% (w/v) and cells were incubated for a further 2 hr at 30°C. Cells were harvested (20 min, 4200 rpm) and transferred to fresh media (without arabinose) and RbcL was induced with 1 mM IPTG for 3 hr at 30°C. The cultures were then harvested by centrifugation (25 min, 4200 rpm) and the pellet stored at -20°C overnight.

Cell pellets were suspended in Buffer A (supplemented with 10 U/ml benzonase, Complete protease inhibitor cocktail, and 0.5 mg/ml lysozyme) and stored on ice for 30 min. Cells were disrupted by sonification, and cell debris removed by ultracentrifugation (30 min 40,000 rpm Ti45). The resulting soluble fraction was sterile filtered and applied on a Ni-NTA-agarose column that was previously equilibrated with Buffer A. This protein tended to aggregate upon addition to the Ni-NTA-agarose (even if batch binding was performed), however, aggregate could be removed by centrifugation of the beads at low rpm and removal of the white aggregation layer. After removal of aggregates, a step-wise washing with Buffer A itself and Buffer A containing 25 mM imidazole was completed and the bound protein eluted with 250 mM imidazole. To cleave the ubiquitin and 6His tag from Syn6301-RbcL, Ub2 protease was added in a 1:100 (mg protease: mg soluble protein) ratio in the presence of 20% (v/v) glycerol and incubated overnight at 16°C.

If sample turned cloudy, it was ultracentrifuged (30 min, 40,000 g Ti45) to remove aggregation before further purification steps. The sample was then run on a HiPrep 26/10 Desalting column before being applied to a MonoQ column; both were pre-equilibrated with Buffer B. The MonoQ column was developed with a linear salt gradient from 0-1 M NaCl in Buffer B, cleaved ubiquitin eluted in the void volume, while fractions containing RbcL$_8$ were

collected and concentrated (MWCO 100,000). Concentrated fractions were further applied to a Superdex200 column for size exclusion chromatography in Buffer B. $RbcL_8$ containing fractions were again collected and concentrated (MWCO 100,000), protein concentration was measured spectrophotometrically at 280 nm, aliquots flash frozen in liquid nitrogen, and stored at -80°C.

3.4.3.2 Prokaryotic and Eukaryotic RbcL

RbcL was also purified in denatured form from inclusion bodies by taking advantage of the fact that much of RbcL (or all, depending on the species) aggregates upon expression in *E. coli* cells. *E. coli* BL21 (DE3) cells transformed with plasmid pET11a encoding the *rbcL* species gene of interest, were grown in LB amp at 37°C to an OD_{600} of 0.6. 1mM IPTG was added to the cells to induce the production of RbcL and cells were further incubated for 5 hr at 30°C. Cells were harvested (30 min, 4200 rpm) and pellet stored overnight at -20°C.

Pelleted cells were suspended in Buffer C (supplemented with Complete protease inhibitor cocktail and 0.5 mg/ml lysozyme) and incubated on ice for 30 min, whereupon 0.5 M EDTA was added. Cells were disrupted by sonification and inclusion bodies harvested by centrifugation (1 hr, 22,000 g). Inclusion bodies were transferred to a homogenizer and suspended in Buffer D. Washed inclusion bodies were centrifuged (45 min, 22,000 g) and again homogenized with Buffer E followed by centrifugation. In the final step, the purified denatured protein was homogenized with Buffer F followed by sterile-filtration. Protein concentration was measured spectrophotometrically at 280 nm. Aliquots of denatured protein were snap frozen in liquid nitrogen and stored at -80°C.

3.4.3.3 Prokaryotic and Eukaryotic RbcS

Expression, disruption, and purification of RbcS from *E. coli* BL21 (DE3) cells was performed essentially as described for the purification of RbcL from inclusion bodies (Coligan et al. 2011; Somerville et al. 1986). However, after homogenization in Buffer F and sterile-filtration, denatured purified protein was diluted in Buffer F to a final protein concentration of 0.5 mg/ml. This was dialyzed against Buffer G, where the protein could spontaneously refold into its native state. Aggregated protein was removed by ultracentrifugation (30 min, 40,000 rpm Ti70), refolded purified protein was further concentrated (MWCO 10,000), glycerol added to 10% (v/v), protein concentration measured spectrophotometrically at 280 nm, aliquots snap frozen in liquid nitrogen, and stored at -80°C.

3.4.3.4 Cyanobacterial-RbcX$_2$

$E.~coli$ BL21 (DE3) cells transformed with a pET vector containing the cyanobacterial $rbcX$ gene of interest were grown at 37°C to an OD$_{600}$ of 0.6. The temperature was then reduced to 30°C and cells induced with 1mM IPTG for 4.5 hr. Cells were harvested by centrifugation (30 min, 4200 rpm) and pellets stored overnight at -20°C. The cell pellets were resuspended in Buffer H supplemented with Complete protease inhibitor cocktail and 0.5 mg/ml lysozyme and incubated on ice for 30 min. Cells were disrupted by sonification and cell debris removed by ultracentrifugation (35 min, 40,000 rpm Ti45). The soluble fraction was applied to a SourceQ column for ion exchange chromatography, pre-equilibrated in Buffer H. The SourceQ column was developed with a linear salt gradient from 20-500 mM NaCl in Buffer H. Fractions containing protein of interest were pooled, dialyzed in Buffer I, and further subjected to ion exchange chromatography with MonoQ pre-equilibrated in Buffer I. The MonoQ column was developed with a linear salt gradient from 20-500 mM NaCl in Buffer I. RbcX$_2$ containing fractions were collected and dialyzed against Buffer J. The dialyzed protein was applied again to a MonoQ column pre-equilibrated against Buffer J and the protein eluted with a linear salt gradient from 50-400 mM NaCl Buffer J. RbcX$_2$ containing fractions were pooled and concentrated (MWCO 10,000). Protein was then subjected to size exclusion chromatography using Superdex75 column pre-equilibrated against Buffer J. RbcX$_2$ fractions were pooled and concentrated (MWCO 10,000), glycerol added to 10% (v/v), protein concentration measured spectrophotometrically at 280 nm, and aliquots snap frozen in liquid nitrogen and stored at -80°C.

3.4.3.5 Cyanobacterial-RbcX$_{2(N-6His)}$

$E.~coli$ BL21 (DE3) cells transformed with $rbcX_{N-6His}$-pET28b were grown, induced, and disrupted as described for Cyanobacterial-RbcX$_2$. However, the pellet was suspended in Buffer K, supplemented with 10 mM imidazole, 10 U/ml benzonase, Complete protease inhibitor cocktail, and 0.5 mg/ml lysozyme. The soluble fraction of the disrupted cells was applied to Ni-NTA-agarose equilibrated in Buffer K with 10 mM imidazole added. Protein was washed and eluted from the Ni-NTA-agarose step-wise with Buffer K containing 10, 50, and 250 mM imidazole. Glycerol was added to 10% (v/v) and protein concentration measured spectrophotometrically at 280 nm. Thereafter, aliquots were snap frozen in liquid nitrogen and stored -80°C.

3.4.3.6 At-RbcXI$_2$

E. coli BL21 (DE3) cells transformed with plasmid At-*rbcXA1*$_{N\text{-}FLAG}$-pHUE were grown at 37°C to an OD$_{600}$ of 0.6. Whereupon, the temperature was decreased to 30°C and cells induced with 1 mM IPTG for 5 hr. Cells were harvested (30 min, 4200 rpm) and pellets stored overnight at -20°C. Pellets were suspended in Buffer L supplemented with 10 mM imidazole, 10 U/ml benzonase, Complete protease inhibitor cocktail, and 0.5 mg/ml lysozyme and incubated on ice for 30 min. Cells were disrupted by sonification and cell debris removed by ultracentrifugation (45 min, 40,000 rpm Ti45). The soluble lysate fraction was applied to Ni-NTA-agarose equilibrated with Buffer L supplemented with 10 mM imidazole. The Ni-NTA-agarose was washed and protein eluted with a step-wise gradient of Buffer L containing 10, 50, and 250 mM imidazole. The eluted protein was incubated with Ub2 protease, to cleave the ubiquitin tag and 6-His tag from RbcX$_2$, in a 1:100 (mg protease: mg protein) ratio overnight at 16°C in the presence of 1 mM DTT and 3% (v/v) glycerol.

Cleaved samples were dialyzed against Buffer M and further applied to MonoQ column equilibrated with Buffer M. The MonoQ column was developed with a linear salt gradient from 0.05-1 M NaCl in Buffer M, and RbcX$_2$ containing fractions were pooled and concentrated (MWCO 10,000). However, upon protein concentration, the *Arabidopsis* RbcXI$_2$ would aggregate, the volume of the sample would decrease, but the concentration of soluble protein would not increase. Instead, many flake-like substances would appear that would not redissolve in buffer; therefore, only a concentration of 1 mg/ml of protein could be achieved (protein concentration was measured spectrophotometrically at 280 nm). This sample was, nevertheless, sent to the MPI crystallization facility for initial crystal screens as described in section 3.4.2. 10% (v/v) glycerol was added to the rest of the sample, snap frozen in liquid nitrogen, and stored -80°C.

3.4.3.7 RbcL$_8$/(RbcX$_2$)$_8$ complex

RbcL$_8$/(RbcX$_{2(N\text{-}6His)}$)$_8$ complexes were expressed in *E. coli* BL21 (DE3) cells transformed with *rbcX*$_{N\text{-}6His}$/*rbcL*-pET28b and the plasmid pG-*KJE8*. The transformed cells were grown at 37°C to an OD$_{600}$ of 0.5, at which point the temperature was reduced to 30°C and arabinose (0.4% (v/v)) and tetracycline (20 ng/ml) were added to the cultures for 2 hr induction of DnaK/DnaJ/GrpE and GroEL/ES, respectively; DnaK/DnaJ/GrpE and GroEL/ES overexpression preceded RbcL/RbcX expression to increase RbcL solubility. Cells were harvested (20 min, 4200 rpm) and fresh LB medium was added containing 1 mM IPTG; RbcL

and RbcX were induced for 3.5 hr at 30°C. Cells were further centrifuged (25 min, 4200 rpm) and pellets stored -20°C overnight.

Cell pellets were incubated for 30 min on ice with Buffer N supplemented with 10 mM imidazole, 10 U/ml benzonase, Complete protease inhibitor cocktail, and 0.5 mg/ml lysozyme, followed by cell disruption by sonification. Cell debris was removed by ultracentrifugation (30 min, 20,000 rpm Ti45). Soluble cell lysate was applied to Ni-NTA-agarose pre-equilibrated in Buffer N with 10 mM imidazole. Ni-NTA-agarose was washed with Buffer N containing 10 and 50 mM imidazole and protein eluted with Buffer N containing 250 mM imidazole. $RbcL_8/(RbcX_2)_8$ complexes were isolated by size exclusion chromatography using Superose 6 or Sephacryl S-300 column equilibrated with Buffer O. Fractions containing the $RbcL_8/(RbcX_2)_8$ complexes were pooled and concentrated (MWCO 10,000), glycerol added to 10% (v/v), protein concentration measured by Bradford assay, snap frozen in liquid nitrogen, and stored -80°C.

3.4.3.8 Syn6301-RbcL$_8$/(AnaCA-RbcX(TAG)$_{2(N-6His)}$)$_8$ complexes

Syn6301-RbcL$_8$/(AnaCA-RbcX(TAG)$_{2(N-6His)}$)$_8$ complexes were expressed in BL21 (DE3) *E. coli* cells cotransformed with the appropriate pET28b vector containing the wildtype or mutant genes and the plasmid *p*Sup-BpaRS-6tRN (enabling the incorporation of the photo-reactive amino acid analog *p*Bpa at TAG sites). Cells were grown in LB kan/cam in the presence of 50 mM K_2HPO_4, 20 mM KH_2PO_4, and 1 mM *p*Bpa (0.1 M *p*Bpa stocks were made fresh in 1 N HCl) at 37°C to an OD_{600} of 0.6. Protein expression was induced with 1 mM IPTG overnight at 30°C. Cells were harvested by centrifugation (30 min, 4200 rpm) and incubated for 30 min in Buffer K supplemented with 10 mM imidazole, 10 U/ml benzonase, Complete protease inhibitor cocktail, and 0.5 mg/ml lysozyme. Cells were disrupted by sonification and cell debris removed by ultracentrifugation (30 min, 20,000 rpm Ti45). Soluble cell lysate was sterile-filtered and applied to Ni-NTA-agarose equilibrated with Buffer K with 10 mM imidazole. Ni-NTA-agarose was washed step-wise with Buffer K containing 10 and 50 mM imidazole and protein eluted with 250 mM imidazole. Glycerol was added to 10% (v/v), protein concentration measured by Bradford assay, snap frozen in liquid nitrogen, and stored at -80°C.

3.4.4 Functional analyses

3.4.4.1 *Amber* stop codon suppression

The efficiency of *p*Bpa incorporation at *amber* stop codon (TAG) sites in mutated AnaCA-RbcX$_{N-6His}$ was analyzed by growing cells according to 3.4.3.8, however a control batch of cells was also grown in the absence of *p*Bpa. The soluble lysate of induced cells was analyzed by SDS-PAGE followed by western blotting and immunodetection with anti-RbcX antibodies. Therefore, the efficiency of *amber* stop codon suppression could be determined by visualization of full length AnaCA-RbcX(TAG)$_{N-6His}$ mutants. Soluble lysate was also analyzed by Native-PAGE followed by western blotting and immunodetection with anti-RbcL antibodies to examine Syn6301-RbcL$_8$/(AnaCA-RbcX(TAG)$_{2(N-6His)}$)$_8$ complex formation.

3.4.4.2 Site-specific crosslinking

Syn6301-RbcL$_8$/(AnaCA-RbcX(TAG)$_{2(N-6His)}$)$_8$ complexes (where the photo-reactive amino acid analog *p*Bpa was incorporated at specific sites in RbcX) were purified as described in section 3.4.3.8. 1 ml of the purified sample was exposed to UV light (365 nm) for 30 min under a 500 W Mercury Arc lamp on ice. The crosslinked samples were analyzed by Gradient SDS-PAGE; a control sample that had not been exposed to UV light was also present for comparison. Protein bands were transferred by western blotting to nitrocellulose membrane and immunodetected with anti-RbcL and anti-RbcX antibodies. Bands containing crosslinked Syn6301-RbcL and AnaCA-RbcX$_{N-6His}$ were excised and digested with trypsin, the resulting purified peptides were subjected to mass spectrometry to determine the exact site on Syn6301-RbcL that crosslinked with AnaCA-RbcX$_{N-6His}$; however, the crosslinked peptides could not be established. AnaCA-RbcX$_{N-6His}$ is not optimally digested with trypsin resulting in few but long peptides, therefore, the peptides did not fly well in mass spectrometry and crosslinked peptide could not be detected.

3.4.4.3 Rubisco carboxylation activity assay of *E. coli* lysate

E. coli BL21 (DE3) cells containing the *rbcL* gene of interest were grown to an OD$_{600}$ of 0.6 in 5 ml LB with appropriate antibiotic. RbcL (with or without RbcS and/or RbcX coexpression) was then induced with 1 mM IPTG for 5 hr at 30°C. Cells were harvested by centrifugation (5 min, 3500 rpm) and pellets stored -20°C. Equivalent amount of cells were resuspended with TNME buffer. Cells were incubated for 30 min on ice with Complete protease inhibitor cocktail, 0.1% (v/v) Triton X-100, and 0.1 mg/ml lysozyme supplementation. The cells were disrupted by sonication and cell debris removed by

centrifugation (10 min, 16,000 g). The amount of soluble RbcL and oligomeric complex formation were monitored by SDS-PAGE and Native-PAGE, respectively followed by western blotting and immunodetection with anti-RbcL antibodies.

Total protein was measured by Bradford assay, and 20 µg of total protein was diluted into Rubisco assay buffer supplemented with 2 mg/ml BSA and 1 mM DTT. If RbcS was not already present, it was added to a final concentration of 3-10 µM and complex formation was completed by incubation at room temperature for 15 min. Subsequent to complex assembly, ^{14}C mix was added to the samples (final concentrations: 0.2 mM DTT, 60 mM NaHCO$_3$, 10 mM MgCl$_2$, and 2 µCi NaH^{14}CO$_3$ in Rubisco assay buffer) followed by further incubation for 5 min at room temperature, enabling carbamate formation. Rubisco carboxylation was initiated with 2.5 mM RuBP and stopped after 10 min at room temperature with 3 N acetic acid. Non-fixed radioactive carbon was removed by evaporation at 95°C, the non-volatile components were dissolved in 100 µl water, followed by addition of 1 ml scintillation fluid. The radioactivity of the fixed carbon was quantified using a LS 6500 multi-purpose scintillation counter (Goloubinoff et al. 1989a; Goloubinoff et al. 1989b; Viitanen et al. 1990).

3.4.4.4 RTS: *in vitro* translation of Rubisco

The RTS100 *E. coli* HY Kit was used for *in vitro* translation purposes of T7-promoter-driven plasmids. 1 µg of wildtype or mutant *rbcL* in pET11a vector were translated in the reconstituted *E. coli* lysate supplemented with 0.5 U/µl RNAguard Ribonuclease inhibitor, Complete protease inhibitor cocktail, 6% (v/v) ^{35}S-methionine, 50 µM unlabelled methionine, and purified chaperone components (concentration and composition indicated in the respective figures). Translations occurred at 30°C for 90 min, reactions were stopped with either cam (200 µg/ml) addition and put on ice, or samples were directly put on ice. Samples were then centrifuged (30 min, 16,000 g, 4°C) and soluble fractions analyzed by SDS-PAGE and Native-PAGE followed by autoradiography (Agashe et al. 2004).

Pulse-chase experiments were performed basically as described above with modifications. 1 µg of plasmid was added to *E. coli* reconstituted lysate, which was supplemented with 0.5 U/µl RNAguard Ribonuclease inhibitor and Complete protease inhibitor cocktail (also chaperone where indicated). This was incubated for 6 min at 30°C, whereupon 6 µl ^{35}S-methionine was added for a further incubation of 90 s at 30°C (zero time point taken). The samples were chased with 3 mM unlabelled methionine at 30°C. Aliquots were taken at indicated time points and stopped by putting on ice. After centrifugation (45

min, 16,000 g, 4°C), the soluble fraction was analyzed by Native-PAGE followed by autoradiography. For Rubisco carboxylation activity measurement purposes, there was no addition of ^{35}S-methionine to the samples. Instead, the equivalent amount of unlabelled methionine was supplemented to the samples, which could then be used for Rubisco carboxylation activity assays with ^{14}C as described in section 3.4.4.3.

3.4.4.5 PURE system

The PURE system was utilized to translate Syn6301-*rbcL*-pET11a in the presence of 0.5 µM GroEL and 1 µM GroES (Shimizu et al. 2005). The PURE system contains only the purified components necessary for transcription and translation, it is not an *E. coli* lysate; therefore, the folding requirement of RbcL with specific chaperones, e.g. GroEL/ES system, can be examined. Assays were prepared by mixing Solution A and Solution B from the PURE system with 0.66 µM unlabelled-methionine, 0.5 U/µl RNAguard Ribonuclease inhibitor, GroEL/ES and transcription/translation initiated by the addition of 20 ng/µl plasmid. Reactions were incubated at 30°C for 90 min, and reactions stopped by addition of cam (0.3 mg/ml) and samples put on ice. The soluble fraction and pellet fractions were separated by centrifugation (15 min, 16,000 g, 4°C). Total, soluble, and pellet fractions were analyzed by SDS-PAGE, followed by western blotting and immunodetection with anti-RbcL antibodies. The soluble fraction was also used to measure the carboxylase activity of the translated RbcL. 10 µM of RbcS was added to the samples and the carboxylation activity measured as described in section 3.4.4.3.

3.4.4.6 CABP synthesis from RuBP

The synthesis of carboxy-arabinitol-1,5-bisphosphate (CABP) was performed essentially as in (Pierce et al. 1980; Zhu and Jensen 1990) with modifications. 67.5 µmole of ribulose-1,5-bisphosphate (RuBP) was dissolved in 1.5 ml 100 mM Tris-acetate, pH 8.3 and this solution was added to 84.5 µmole of KCN powder and incubated for at least 48 hr in the fume hood in a closed tube to synthesize CABP (an equal amount of CRBP, carboxy-ribitol-1,5-bisphosphate, is also synthesized, but displays less affinity for the Rubisco active site (Smrcka et al. 1991)). 4 ml of water was added to 4 ml AG 50W-X8 resin (H+) to make a slurry, the resin was acid washed with 3 column volumes (CV) 20% formic acid and equilibrated with 3 CV water. In the fume hood, CABP was passed over the resin and washed with 4 CV water; flow through and washes were collected in glass beaker. Beaker with flow

through and washings, was placed at 25°C and a gentle stream of nitrogen was used for drying. The resulting residue was suspended in 16 ml 50 mM Bicine-NaOH, pH 9.3.

The concentration of the synthesized CABP was measured by Malachite green assay (Chifflet et al. 1988) through the establishment of a K_2Pi standard. Increasing concentrations of K_2Pi were incubated with Malachite green reagent with 3.8% (w/v) sodium citrate added, incubated for 1 min, and absorbance measured spectrophotometrically at 640 nm. After the establishment of the phosphate standard, 20 μl and 10μl of synthesized CABP was incubated with 2.5 units of shrimp alkaline phosphorylase (SAP) at 37°C for 30 min and 1 hr. SAP was inactivated by incubation for 15 min at 65°C. Malachite green reagent and sodium citrate were added and absorbance was measured as above. The concentration of synthesized CABP/CRBP could be determined from the slope of the phosphate standard created. Inhibition of the prepared CABP was tested by incubating increasing concentrations with Rubisco and measuring carboxylation activity as described in 3.4.4.3. Aliquots were stored at -20°C.

3.4.4.7 Disulfide crosslinking

Purification of $RbcL_8/(RbcX_2)_8$ wildtype and mutant complexes containing cysteine residues at the RbcL/RbcX interface, were purified as described in section 3.4.3.8; however, 5 mM oxidized glutathione (GSSG) and 0.5 mM reduced glutathione (GSH) were present in Buffer K during elution from the Ni-NTA-agarose to facilitate a redox environment and enhance disulfide bridge formation. The isolated $RbcL_8/(RbcX_2)_8$ complexes were incubated for 15 min with 100 mM iodoacetamide to covalently block reduced cysteines (Park et al. 2007). The crosslinked samples were analyzed under non-reducing conditions on 10-15% gradient SDS-PAGE, followed by immunoblotting with anti-RbcL antibodies.

3.4.4.8 *In vivo* expression of cyanobacterial GroEL and GroES in *E. coli*

The oligomeric state of cyanobacterial GroEL1 and GroEL2 was analyzed by recombinant expression and/or coexpression in BL21 (DE3) *E. coli* cells. Cells transformed with the plasmids Ana7120-*groEL1*-pET11a, Ana7120-*groEL2*-pET11a, Ana7120-*groES*-pET11a, Ana7120-*groEL2*-pCOLAduet-1 cotransformed with Ana7120-*groEL1*-pET11a, Syn6301-*groEL1*-pET11a, or Syn6301-*groES*-pET11a were grown to an OD_{600} of 0.6 at 37°C and GroEL/GroES was induced with 1 mM IPTG for 5 hr at 30°C. Cells were harvested by centrifugation (10 min, 16,000 g) and pellets suspended in different buffer preparations (as indicated in the respective figures). Equivalent amounts of cells were disrupted by

sonification and cell debris removed by centrifugation (10 min, 16,000 g). Soluble cell lysate was subsequently applied to SDS-PAGE and Native-PAGE gels followed by coomassie staining/destaining.

3.4.4.9 *In vivo* assembly of $RbcL_8/(RbcX_2)_8$ complexes in *E. coli*

Expression of wildtype or mutant Syn6301-RbcL or Syn6301-RbcL/RbcX from the respective pET-vector and preparation of soluble lysate was performed essentially as above. Equivalent amounts of cells were lysed in the presence of TNME buffer supplemented with 0.1% (v/v) Triton X-100, 0.5 mg/ml lysozyme, and Complete protease inhibitor cocktail. 1 µM of AnaCA-$RbcX_2$ was added to the soluble lysate expressing Syn6301-RbcL and respective mutants followed by incubated for 1 hr at 25°C. When indicated, 3 µM Syn6301-RbcS was also added with further incubation for 1 hr at 25°C. Formation of $RbcL_8/(RbcX_2)_8$ complexes was analyzed by Native-PAGE followed by immunoblotting with anti-RbcL antibodies. Soluble RbcL was also analyzed by SDS-PAGE and immunoblotting with anti-RbcL antibodies.

3.4.4.10 Analytical gel filtration of $RbcL_8/(RbcX_2)_8$ complexes

The purification and isolation of $RbcL_8/(RbcX_2)_8$ complexes was done essentially as described in 3.4.3.7. Fractions containing $RbcL_8/(RbcX_2)_8$ complexes were pooled. 1mg/ml of the purified wildtype or mutant complexes were applied to a Superose 6 gel filtration column using the Ettan system, with or without addition of 25 µM wildtype or mutant Syn6301-RbcS at 25°C for 1 hr. The displacement of $RbcX_2$ by RbcS was monitored by SDS-PAGE of elution fractions.

3.4.4.11 Co-immunoprecipitation

Preparation of soluble lysate was essentially as above using cells coexpressing chaperones (GroEL/ES, DnaK/DnaJ/GrpE) in plasmid pG-*KJE8* and wildtype or mutant RbcL with FLAG-tagged RbcX in the plasmid Syn6301-*rbcX*FLAG/Syn6301-*rbcL*-pET11a. All centrifugation steps with FLAG beads were performed for 30 s at 8,200 g and 4°C, beads were kept at 4°C, and all incubation steps were performed with gentle rotation. Before use, gel beads were washed twice with Flag buffer and vortexing. 500 µl soluble cell lysate was incubated with 40 µl anti-FLAG M2 Affinity Gel beads for 1 hr at 4°C. The beads were washed 3 times 5 min with Flag buffer and the bound complexes were then incubated for 1 hr at 4°C with increasing concentrations of Syn6301-RbcS. Beads were then washed 4 times 15

min and RbcL protein that remained bound after washing with Flag buffer was eluted by incubating the beads for 5 min at 95°C in non-reducing SDS sample buffer and analyzed on 12.5 % SDS-PAGE gels, followed by immunoblotting with anti-RbcL antibodies.

3.4.4.12   Rubisco *in vitro* refolding

Denatured RbcL, of the respective species being analyzed, was diluted from Denaturation buffer 200-fold (0.5 µM final concentration) into ice cold Refolding buffer supplemented with 5 mM DTT, 1 mg/ml BSA in the presence of 0.5 µM chaperonin. RbcL was carefully added with constant vortexing. After addition of RbcL, the sample was incubated at 25°C for 5 min, followed by centrifugation for 10 min at 16,000 g and 4°C. The supernatant was transferred to new tubes, kept at room temperature, and a small aliquot was taken to analyze substrate binding efficiency by SDS-PAGE. Subsequently, at concentrations indicated in the figure legends, cochaperone, RbcX$_2$, and RbcS were added. Refolding was initiated by the addition of 4 mM ATP at 25°C. Reactions were stopped at specific time points by the addition of 0.25 U/µl apyrase or 1 U/µl hexokinase with 10 mM glucose. As specified, 200 µM Syn6301-RbcL C-terminal peptide (KEIFEFETMD) and 10 µM RbcS (when not previously present in the refolding reaction) were added, and incubated at room temperature for 15 min to allow complex assembly. Samples were split in to two parts with one half subjected to Bis-tris Native-PAGE and immunoblotting with anti-RbcL antibodies, and the other half was used to measure carboxylation activity of the refolded RbcL according to the methods described in 3.4.4.3.

3.4.4.13   Peptide binding screen

Analysis of RbcX$_2$ binding to the C-terminal peptide region of RbcL was performed using a peptide array membrane prepared by JPT Peptide Technologies GmbH; the epitope map is prepared by SPOT-synthesis. The dodecapeptides, with 10 amino acid overlap covering the sequence of the last 40 amino acids in various RbcL species, are covalently bound to Whatman 50 cellulose support by the C-terminus end and have an acetylated N-terminus that is not bound. Each spot contains approximately 5 nmol of peptide.

Before RbcX$_2$ was bound to the membrane, and after each membrane regeneration, the membrane was detected without protein addition to see if there was any background binding of antibody to the spots or if residual protein was still bound after regeneration. Blank detection was begun by washing the membrane for 5 min with 100% methanol. The methanol wash is performed to avoid hydrophobic peptide precipitation in the following TBS washing

steps. The membrane was then washed 3 times 10 min with 1xTBS and incubated overnight with 1xTBSM. The next morning the membrane was washed 3 times 5 min with 1xT-TBS and incubated for 45 min with 1:2,000-5,000 dilutions of primary antibody in 1xT-TBSM. Excess antibody was removed by washing the membrane 2 times 5 min with 1xT-TBS. Then, the membrane was incubated with 1:5,000 dilutions of HRP-conjugated secondary antibody in 1xT-TBSM for 45 min. Excess antibody was removed by washing 3 times 5 min with 1xT-TBS and chemiluminescent detection performed as described in 3.4.1.7. Incubation of the membrane with $RbcX_2$ was performed as described above, however after initial washing with 1xTBS, the membrane was blanked for 3 hr with 1xTBSM followed by overnight incubation of 10 µg/ml $RbcX_2$ protein in 1xTBSM.

After protein incubation and ECL detection, the protein bound to the membrane was removed by regenerating the membrane as described by the manufacturer; two different regeneration protocols are detailed, Regeneration protocol I and Regeneration protocol II. For Regeneration protocol I the membrane was washed 3 times 10 min with water after ECL detection. The membrane then was incubated at 50°C 4 times 30 min in Regeneration buffer I. The membrane was washed 3 times for 20 min with 10xPBS, followed by washing 3 times 20 min with 1xT-TBS. The membrane could then be blank detected to determine the extent of protein removal. If protein remained bound to the membrane, Regeneration protocol II was performed. Membrane was washed 2 times 10 min with water and then incubated 3 times 10 min with Regeneration buffer IIA, followed by incubation 3 times 10 min with Regeneration buffer IIB. The membrane was subsequently washed 3 times 10 min with water and 3 times 10 min with 1xT-TBS, whereupon blank detection was carried out. Regeneration of the membrane was repeated until all protein was removed or only minimal amounts of protein remained bound.

4   Results

4.1   Characterization of cyanobacterial chaperonin systems

To determine the potential usage of cyanobacterial chaperonin systems in refolding Rubisco, components of the chaperonin systems from two cyanobacteria species (Syn6301 and Ana7120) were recombinantly expressed in *E. coli*. As seen by SDS-PAGE in Figure 4.1 A, all proteins were soluble and efficiently expressed in *E. coli*. The cyanobacterial co-chaperone GroES migrated very similar to *E. coli* GroES on Native-PAGE suggesting identical oligomeric states (Figure 4.1 B lanes 2, 5, and 7) i.e., cyanobacterial GroES most likely also formed heptameric ring structures. In contrast, the cyanobacterial chaperonins GroEL1 and GroEL2 exhibited faster electrophoretic mobility than the *E. coli* GroEL chaperonin on Native-PAGE (Figure 4.1 B lanes 1, 3, 4, and 6). This indicated that cyanobacterial GroEL did not form higher order oligomers in *E. coli*, instead the chaperonins seem to migrate as monomers or dimers.
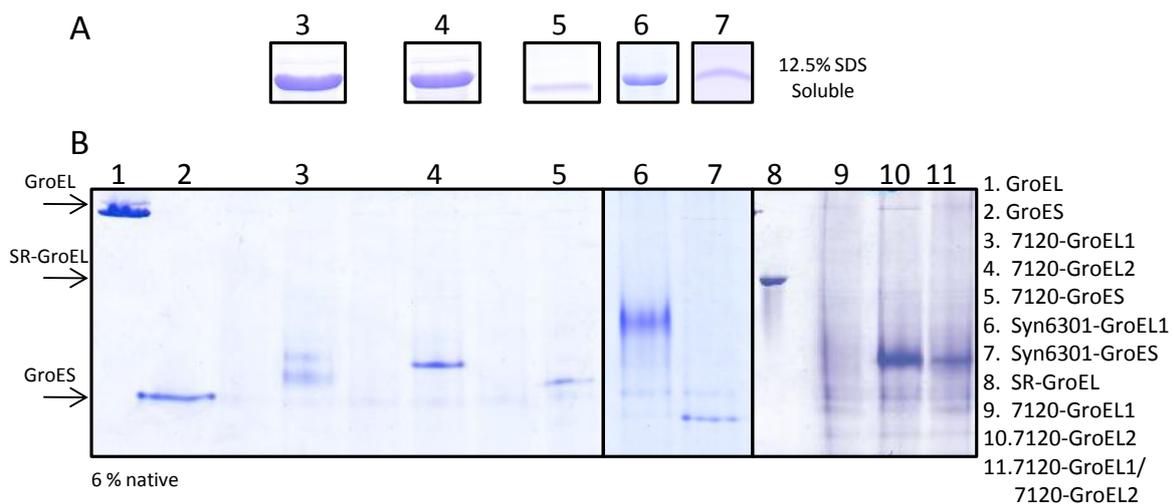
Figure 4.1 Recombinant expression of cyanobacterial GroEL and GroES in *E. coli*.
GroEL and GroES from *Anabaena* sp. PCC7120 and *Synechococcus* sp. PCC6301 were overexpressed in *E. coli* cells and soluble lysate analyzed by (A) SDS-PAGE and (B) Native-PAGE. Buffer P was used during cell lysis of lance 3-7, while Buffer Q was used for lanes 8-11 as recommended by (Lehel et al. 1992).

Previous studies have also shown that the cyanobacterial chaperonins are highly unstable, and oligomeric complexes difficult to isolate (Huq et al. 2010; Kovacs et al. 2001). Intriguingly, it has been suggested that GroEL1 and GroEL2 may form heterooligomers (Huq et al. 2010); two genes for GroEL exist in cyanobacteria, therefore, it is plausible that both GroEL's are important for oligomerization, as is the case for plant chaperonin (Nishio et al. 1999). However, no known reports exist of coexpression experiments of GroEL1 and GroEL2 from cyanobacteria. Here, the coexpression of GroEL1 and GroEL2 from *Anabaena* 7120 in *E. coli* did not result in the formation of higher molecular weight species (Figure 4.1 B lane

11). Furthermore, oligomerization was not observed even if cell lysis occurred in the presence of Buffer Q (Figure 4.1 B lanes 9-11), which was reported to support cyanobacterial GroEL oligomeric formation/stabilization (Lehel et al. 1992). Buffer Q conditions, K_3PO_4 at pH 7.0, compared to Buffer P conditions, NaH_2PO_4 at pH 8.0, actually seemed to be detrimental to the formation of a coherent species for 7120-GroEL1 because only a smear could be visualized (Figure 4.1 lane 9 compared to lane 3). This was not observed for 7120-GroEL2 or 7120-GroEL2/GroEL1 coexpression; here clear bands were present in the middle of the Native-PAGE gel (Figure 4.1 lanes 10 and 11). Nevertheless, no shift to the formation of higher oligomeric species was observed since this band clearly migrated below the heptameric single-ring GroEL (SR-GroEL) (Figure 4.1 lane 8). Conditions for cyanobacterial GroEL oligomer formation unfortunately could not be established in this study. Hence, the *E. coli* chaperonin system GroEL/ES was used instead to investigate the biogenesis of cyanobacterial Rubisco.

4.2 Role of GroEL/ES and RbcX$_2$ in Rubisco folding and assembly

Chaperonin and the assembly chaperone $RbcX_2$ are necessary for proper form I Rubisco folding and assembly (Li and Tabita 1997; Onizuka et al. 2004; Saschenbrecker et al. 2007; Tarnawski et al. 2008). In the RTS100 HY *E. coli in vitro* transcription/translation system form I Rubisco from Syn7002 was found to be absolutely dependent on $RbcX_2$ for productive $RbcL_8$ formation (Saschenbrecker et al. 2007). In contrast, Syn630l Rubisco was able to fold and assemble into holoenzyme when *E. coli* GroEL and GroES were present as the only supplemented proteins in this system (Figure 4.2 A, lane C). In this assay Rubisco activity can then be quantified by ^{14}C-CO_2 fixation assays following the addition of purified small subunit to the reaction. Enhanced solubility with proper $RbcL_8$ formation and correspondingly increased Rubisco activity was observed upon addition of 40 µM Syn7002-$RbcX_2$ or 20 µM Syn6301-$RbcX_2$ (Figure 4.2 A, lanes D and F). The carboxylase activity was greatly reduced in the presence of 40 µM AnaCA-$RbcX_2$ and the formation of an intermediate complex ($RbcL_8$/($RbcX_2$)$_8$) was observed (Figure 4.2 A, lane E). Interestingly, there was a two-fold increase in Rubisco carboxylation activity when Syn6301-RbcS was also present in the transcription/translation reaction in the absence of $RbcX_2$. However, maximum carboxylation activity was only achieved when Syn6301-RbcS and Syn6301-$RbcX_2$ were both present in the reaction mixture (Figure 4.2 A, lanes H and I). These results confirm that, consistent with data obtained *in vivo* (Emlyn-Jones et al. 2006; Goloubinoff et al. 1989b), in the case of Syn6301

Rubisco assembly of the RbcL$_8$ core is not dependent on RbcX$_2$, nevertheless enhanced assembly was observed in presence of the latter.

A

	A	B	C	D	E	F	G	H	I
Syn6301-RbcL	+	+	+	+	+	+	+	+	+
GroEL 0.5 μM		+	+	+	+	+	+	+	+
GroES 1 μM			+	+	+	+	+	+	+
Syn7002-RbcX (μM)				40					
AnaCA-RbcX (μM)					40				
Syn6301-RbcX (μM)						20	5		20
Syn6301-RbcS 10 μM								+	+

GroEL/RbcL →
RbcL$_8$/(RbcX$_2$)$_8$ →
RbcL$_8$ →

Native 5-13%   10   2   39   73   80   61   97   100%

B

	A	B	C	D	E
Syn6301-RbcL	+	+	+	+	+
GroEL (μM)		0.5	0.5	0.25	0.1
GroES (μM)			1	0.5	0.2

Figure 4.2 *In vitro* translation of Syn6301-RbcL.
(A) RTS100 HY *E. coli in vitro* transcription/translation system: Syn6301-*rbcL*-pET11a was translated using the RTS system in the presence of ^{35}S-met or unlabelled methionine for 1.5 hr at 30°C. GroEL, GroES, RbcX$_2$, and RbcS were added to the reactions at the indicated concentrations. Samples were either subjected to 5-13% gradient Native gel analysis followed by autoradiography or carboxylation activity assays. Note, 10 μM Syn6301-RbcS was supplemented when not previously present for activity assays.
(B) PURE system: Syn6301-*rbcL*-pET11a was translated in the PURE system in the presence of unlabelled methionine for 1.5 hr at 30°C. Where indicated GroEL and GroES were added to the reactions at the specified concentrations. After stopping reactions with cam, the samples were incubated for 15 min with 10 μM Syn6301-RbcS before carboxylation activity of samples was measured.

The RTS transcription/translation system does not represent a truly minimal system, since the entire *E. coli* chaperone machinery is present, albeit at low concentrations. To establish whether the chaperonin system alone is capable of mediating folding and assembly of Syn6301-RbcL, the PURE system was utilized. The PURE system is constituted from recombinant purified components that are known to be essential for transcription and translation in *E. coli* (Shimizu et al. 2005); therefore, the refolding and assembly of Syn6301-RbcL in a translating system can be analyzed without influence from any other chaperoning components. As shown by in Figure 4.2 B, lanes C-E, significant amounts of holoenzyme were produced in the PURE system when GroEL and GroES were the only chaperones present. Increased carboxylation activity of Rubisco was observed with increasing GroEL/ES concentrations, indicating that formation of holoenzyme from translated Syn6301-RbcL polypeptide was not inhibited by high GroEL/ES concentrations. Instead, it seems that newly translated Syn6301-RbcL polypeptide could compete with GroEL-bound and folded RbcL subunits, thereby forcing folded RbcL to be released from the GroEL/ES cavity, which then

further oligomerized into $RbcL_8$ cores. Furthermore, proper folding and assembly of Syn6301 Rubisco was dependent on GroEL/ES encapsulation and release in a translating system because no carboxylase activity could be determined without chaperonin or cochaperone supplementation (Figure 4.2 B, lanes A and B).

4.2.1 RbcX$_2$ binding motif

Recently, it has been reported that the central crevice of $RbcX_2$ from cyanobacteria binds to a specific region on the C-terminal tail of cyanobacterial RbcL (Liu et al. 2010; Saschenbrecker et al. 2007), referred to as the C-terminal recognition motif. This motif is present in all green-type form I Rubisco sequences which leads to the hypothesis that this region could be essential for $RbcX_2$ binding in other species as well, for example eukaryotic species. Form I Rubisco occurs in four classes A-D (Badger and Bek 2007). Form IA and form IB are referred to as green-type and form IC and form ID are referred to as red-type Rubisco (Tabita 1999). Form IA Rubisco is found in marine cyanobacteria and while the RbcL of form IA does have the conserved C-terminal recognition motif, no *rbcX* has been located in their genome. Form IB Rubisco is found in plants, green algae, and fresh water cyanobacteria. The form IB Rubisco large subunit also has the conserved C-terminal recognition motif and correspondingly a discernable *rbcX* homolog in the genome. Form IC (in proteobacteria) and form ID Rubisco (in proteobacteria and red algae) do not have a traditional C-terminal recognition motif and to date no *rbcX* homolog has been located in their genome (Tabita 2007).

In order to test the binding of three cyanobacterial $RbcX_2$ proteins to the C-terminal tail of different classes of Rubisco, a peptide membrane binding assay was implemented (Figure 4.3 A-C). Figure 4.3 D shows the binding consensus in red and the general binding pattern to the RbcL C-terminal tail was similar for the three species. However, Syn6301-$RbcX_2$ and Syn7002-$RbcX_2$ displayed more diversity in their binding, in that they also bound well to eukaryotic RbcL C-terminal peptides, while AnaCA-$RbcX_2$ binding was mainly restricted to prokaryotic species. Note that AnaCA-$RbcX_2$ did bind to the consensus motif of two eukaryotic species, i.e. *Chlamydomonas reinhardtii* and *Arabidopsis thaliana* (Figure 4.3 C, lanes 9* and 13*). Importantly, there was specific binding to the C-terminal recognition motif (EIKFEF) and no binding to any non-related sequences or to species that do not have the C-terminal recognition motif i.e., form II and form III Rubisco. Syn7002-$RbcX_2$ had slightly more promiscuous binding, and there was some weak interaction with the C-terminal peptide of red-type form I Rubisco, which could be due to similarity in this sequence pattern

to the C-terminal recognition motif i.e., the hydrophobic residues in the highlighted area would potentially be able to interact with the hydrophobic pocket in the RbcX$_2$ groove.

D

Form IC	1. *Rhodobacter sphaeroides*	LRAAAKWSKPLEAALDTWGNITFNYTSTDTSDFVPTASVA
Form IA	2. *Thiobacillus denitrificans*	GKTVLTEAAKNSPELKIAMETWKEIKFEFDTVDKLDVAHK
	3. *Allochromatium vinosum*	KDVLTKAAASSPELKIAMETWKEIKFEFDTVDKLDIAHHK
	4. *Nitrobacter winogradskyi*	KDILTAAAAQSPELKIAMETWKEIKFEFETMDKLAIANK
	5. *Synechococcus sp* WH 8102	KESRDILMEAGKHSPELAIALETWKEIKFEFDTVDKLDVQN
Form IB	6. *Synechococcus sp* PCC 7002	SLAREGNDVLREAGKWSPELAAALDLWKEIKFEFDTVDTL
	7. *Synechococcus elongates* PCC 6301	DLYREGGDILREAGKWSPELAAALDLWKEIKFEFETMDKL
	8. *Anabaena* PCC 7120	NLAREGNDVIREAAKWSPELAVASELWKEIKFEFEAMDTV
	9. *C. reinhardtii*	DLAREGGDVIRSASKWSPELAAASEVWKEIKFEFDTIDKL
Form ID	10. *Galdieria sulfuraria*	REAAKNSGALRTALDLWKDITFNYTSTDTSDFVETPTANI
	11. *Griffithsia pacifica*	RDAAKTSGPLQTALDLWKDITFNYTSTDTADFVETPTANV
	12. *Suhria vittata*	RDAAKTSGPLQTALDLWKDISFNYTSTDTADFVETPTANV
Form IB	13. *A. thaliana*	EGNEIIREASKWSPELAAASEVWKEITFNFPTIDKLDGQE
	14. *Nicotiana tabacum*	AQEGNEIIREASKWSPELAAASEVWKEIVFNFAAVDVLDK
	15. *O. sativa*	AREGNEIIRSASKWSPELAAASEIWKAIKFEFEPVDKLDS
	16. *Spinacia oleracea*	DLAREGNTIIREATKWSPELAAASEVWKEIKFEFPAMDTV
	17. *Rhapis excelsa*	DLAREGNEIIREASKWSPELAAASEVWKEIKFEFEPVDKLD
	18. *Zea mays*	DLAREGNEIIKAASKWSAELAAASEIWKEIKFDGFKAMDTI
Form II	19. *Rhodospirillum rubrum*	AREHKELARAFESFPGDADQIYPGWRKALGVEDTRSALPA
Form III	20. *Thermococcus kodakaraensis* KOD1	ARAVRQAIDAIMQGIPLDEYAKTHKELARALEKWGHVTPV

E

Syn6301-RbcL peptide: KEIKFEFETMD

	K_A (M^{-1})	K_D (µM)
Syn6301-RbcX$_{2(N-FLAG)}$	9545 ± 781	105 ± 8.5
AnaCA-RbcX$_{2(N-FLAG)}$	507100 ± 104200	1.97 ± 0.45
Syn7002 -RbcX$_2$	3337 ± 1042	300 ± 104

Figure 4.3 Binding of Syn6301-RbcX$_{2(N-FLAG)}$, Syn7002-RbcX$_2$, and AnaCA-RbcX$_{2(N-FLAG)}$ to RbcL C-terminal.
(A-C) Purified RbcX$_2$ was incubated with a membrane array of dodecapeptides, with a 10 residue overlap, covering the sequence of the last 40 amino-acids of various Rubisco large subunits as listed. Peptide-bound RbcX$_2$ was visualized by chemiluminescent immunodetection with anti-FLAG or anti-Syn7002- RbcX antibodies. Rows 16-18 (in red) on membrane C indicate background binding. (D) Consensus motif of RbcX$_2$ binding indicated in red. (E) K_A and K_D values for the interaction of Syn6301-RbcL C-terminal peptide with RbcX$_2$ proteins determined by isothermal titration calorimetry (ITC) (Liu et al. 2010).

In accordance with a previous report (Saschenbrecker et al. 2007) it is evident that prokaryotic $RbcX_2$ bound specifically to the extended C-terminal tail of RbcL, which is conserved in green-type form I Rubisco. As discussed in the introduction, the interaction between $RbcX_2$ and RbcL must be dynamic for holoenzyme formation to occur. Interestingly, it was found that $RbcX_2$ from different species have very divergent binding affinities for this motif (Figure 4.3 E). Using isothermal titration calorimetry (ITC) it was demonstrated that AnaCA-$RbcX_2$ binds the C-terminal recognition motif peptide with strikingly high affinity, followed by Syn6301-$RbcX_2$, and then Syn7002-$RbcX_2$. The particularly high binding affinity of AnaCA-$RbcX_2$ to the Syn6301-RbcL C-terminal recognition motif led to the formation of a stable $RbcL_8/(RbcX_2)_8$ complex as observed in Figure 4.2 A, lane E. The binding between these two proteins was so tight that holoenzyme formation was impeded, resulting in no measurable carboxylase activity.
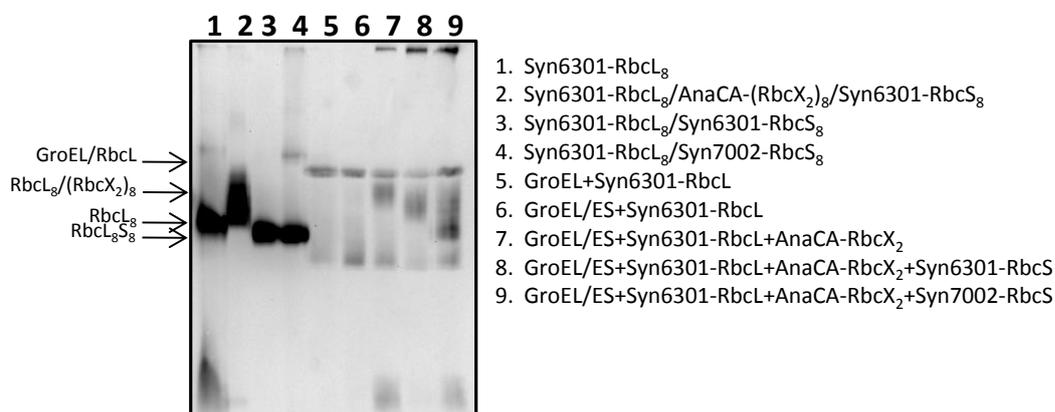
Figure 4.4 *In vitro* refolding: displacement of AnaCA-$RbcX_2$ with Syn7002-RbcS.
Lanes 1-4 Protein standards: (1) 0.25 µM $RbcL_8$ (monomer); (2) 0.25 µM $RbcL_8$ (monomer) incubated for 1 hr 25°C with 1.5 µM AnaCA-$RbcX_2$ (dimer) and 2.5 µM Syn6301-RbcS; (3) 0.25 µM $RbcL_8$ (monomer) incubated for 1 hr 25°C with 2.5 µM Syn6301-RbcS; (4) 0.25 µM $RbcL_8$ (monomer) incubated for 1 hr 25°C with 2.5 µM Syn7002-RbcS. Lanes 5-9 *in vitro* refolding: 0.5 µM denatured Syn6301-RbcL was diluted 100x into Refolding buffer containing 0.5 µM GroEL. As indicated 1 µM GroES, 1.5 µM AnaCA-$RbcX_2$, and/or 10 µM Syn6301-RbcS or Syn7002-RbcS was added to the samples. Refolding assay was carried out for 1 hr at 25°C and samples subjected to 5-13% Native-PAGE analysis followed by immunoblotting with anti-RbcL antibody.

The observation of high binding affinity between RbcL and $RbcX_2$ from different species proved important for the development of the *in vitro* reconstitution assay of Syn6301 Rubisco (Liu et al. 2010) (see section 2.3.4.4 for a detailed description of this assay). When AnaCA-$RbcX_2$ was not present, RbcL appeared not to be released from GroEL during *in vitro* refolding of Syn6301-RbcL (Figure 4.4 lane 5 and 6). *In vitro*, the use of a high affinity $RbcX_2$ was necessary to push the equilibrium towards the assembly intermediates and reduce RbcL rebinding to GroEL (Figure 4.4 lane 7). Addition of the cognate RbcS was unable to completely remove AnaCA-$RbcX_2$ from the Syn6301-$RbcL_8$ cores (Figure 4.4 lane 8). Conversely, addition of Syn7002-RbcS to the *in vitro* refolding reaction allowed the formation of holoenzyme and the removal of AnaCA-$RbcX_2$ (Figure 4.4 lane 9). Interestingly,

a smaller, additional assembly intermediate (presumably $RbcL_2/(RbcX_2)_2$) was discernible in the presence of AnaCA-RbcX$_2$, with or without Syn7002-RbcS addition (Figure 4.4 lanes 7 and 9), but not in the presence of Syn6301-RbcS (Figure 4.4 lane 8), suggesting that the cognate small subunit was able to fully recruit assembly intermediates for oligomerization whereas the heterologous protein may be impaired in this respect.
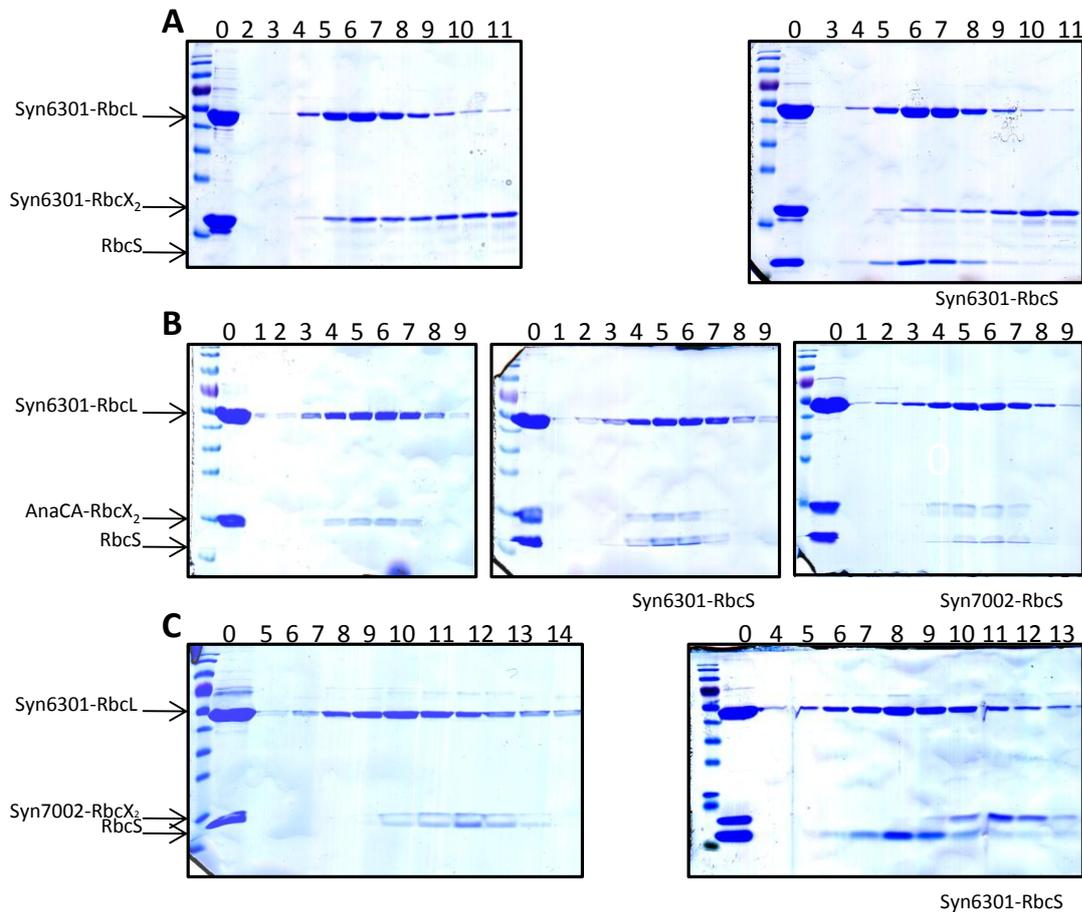
Figure 4.5 RbcS displacement of RbcX$_2$.
(A) Left panel, 10 μM (monomer) Syn6301-RbcL$_8$ was incubated for 1 hr 25°C with 5 μM (dimer) Syn6301-RbcX$_{2(N-FLAG)}$ followed by size exclusion chromatography with Superose6 on the Ettan system and elution fractions analyzed on 12.5% SDS-PAGE followed by coomassie staining. Right panel, same as previous except RbcL/RbcX$_2$ incubation was followed by 1 hr incubation with 20 μM Syn6301-RbcS at 25°C. (B) Incubation procedure and analysis same as in A, Left panel Syn6301-RbcL$_8$ incubated with AnaCA-RbcX$_2$; Middle panel, Syn6301-RbcL$_8$/AnaCA-RbcX$_{2(N-FLAG)}$ with Syn6301-RbcS addition; Right panel, Syn6301-RbcL$_8$/AnaCA-RbcX$_{2(N-FLAG)}$ with Syn7002-RbcS addition. (C) Incubation procedure and analysis same as in A, Left panel, Syn6301-RbcL$_8$ incubated with Syn7002-RbcX$_2$; Right panel, Syn6301-RbcL$_8$/Syn7002-RbcX$_2$ with Syn6301-RbcS addition. Lane 0 is the input fraction. RbcL concentration is calculated for the monomer while RbcX concentration is calculated for the dimer.

To further investigate the possibility that Syn7002-RbcS displaces AnaCA-RbcX$_2$ from Syn6301-RbcL$_8$ cores and to better visualize the RbcL/RbcX interaction, size exclusion chromatography experiments were performed (Figure 4.5). Syn6301-RbcL$_8$ was incubated with Syn6301-RbcX$_2$ (Figure 4.5 A), AnaCA-RbcX$_2$ (Figure 4.5 B), or Syn7002-RbcX$_2$ (Figure 4.5 C) followed by Syn6301-RbcS or Syn7002-RbcS addition. In the cognate condition before RbcS addition, Syn6301-RbcL$_8$ and Syn6301-RbcX$_2$ co-eluted indicating

that complex formation occurred; $RbcX_2$ that did not bind stably eluted later (Figure 4.5 A, left panel). When Syn6301-RbcS was added to the $Syn6301-RbcL_8$/$Syn6301-RbcX_2$ complex, most $Syn6301-RbcX_2$ was removed and the $Syn6301-RbcL_8$/$Syn6301-RbcS_8$ complex was formed, as seen by the slight shift in the elution volume of RbcL, and the decrease in $RbcX_2$ co-eluting with RbcL (Figure 4.5 A, right panel).

When $AnaCA-RbcX_2$ was incubated with $Syn6301-RbcL_8$ cores, the formation of the stable $RbcL_8$/$(RbcX_2)_8$ complex was apparent; all of the $AnaCA-RbcX_2$ bound to and eluted with $Syn6301-RbcL_8$ (Figure 4.5 B, left panel). Furthermore, addition of Syn6301-RbcS or Syn7002-RbcS could not displace this high affinity $RbcX_2$ (Figure 4.5 B, middle and right panels, respectively). A ternary RbcL/RbcX/RbcS complex was formed. This revealed that even though in Figure 4.4 lane 9, it seemed that $Syn7002-RbcS$ could displace $AnaCA-RbcX_2$ from $Syn6301-RbcL_8$, size exclusion chromatography clearly demonstrated that Syn7002-RbcS was not able cause displacement. The composition/configuration of this RbcL/RbcX/RbcS complex was however not further analyzed.

In ITC experiments, $Syn7002-RbcX_2$ exhibited the lowest affinity for the Syn6301-RbcL C-terminal recognition motif (Figure 4.3 E). This low affinity was also apparent in size exclusion chromatography (Figure 4.5 C). Here, the complex formed between $Syn6301-RbcL_8$ and $Syn7002-RbcX_2$ was not stable enough to withstand gel filtration analysis and $RbcX_2$ did not co-elute with $RbcL_8$ (Figure 4.5 C, left panel). The elution fractions containing RbcL were shifted towards lower molecular weight compared to the experiments containing $Syn6301-RbcX_2$ or $AnaCA-RbcX_2$ (Figure 4.5 A and B, left panel) indicating $RbcL_8$ without $RbcX_2$ bound (Figure 4.5 C, left panel). Therefore, the size exclusion experiments corroborate with the ITC measurements of the affinity of $RbcX_2$ to $RbcL_8$.
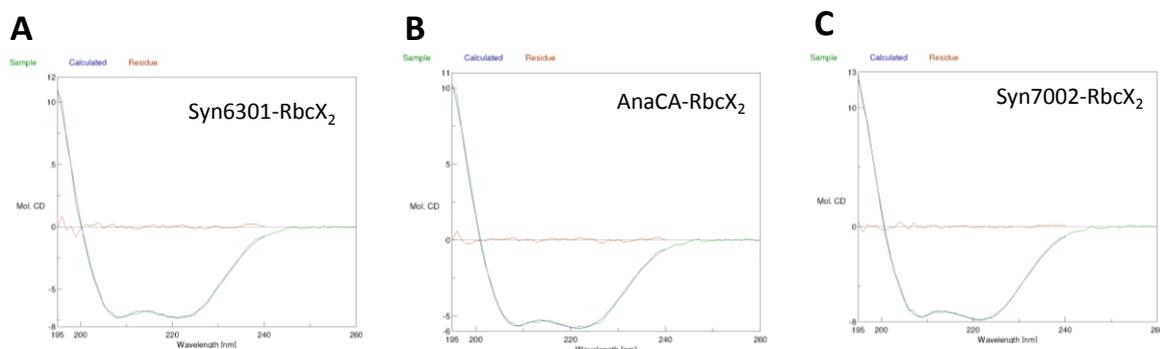
Figure 4.6 CD spectra of cyanobacterial $RbcX_2$.
The CD spectra of $RbcX_2$ (0.1 mg/ml at 4°C in Buffer J) from three different cyanobacteria species as indicated. (A) $Syn6301-RbcX_{2(N-FLAG)}$ 61.2% alpha-helical. (B) $AnaCA-RbcX_{2(N-FLAG)}$ 57.8% alpha-helical. (C) $Syn7002-RbcX_2$ 66.8% alpha-helical.

The precise cause for the difference in $RbcX_2$ affinity to the C-terminal recognition motif still remains to be elucidated. The overall secondary structures of the three RbcX's were

quite similar (Figure 4.6). As illustrated by the CD spectra, all were predominantly alpha-helical, and from previous crystal structure analyses of various $RbcX_2$ species it is known that the tertiary/quaternary structures of $RbcX_2$ are also highly similar (Saschenbrecker et al. 2007; Tanaka et al. 2007). However, there were several differences discerned from comparison of the $Syn7002-RbcX_2$ and $AnaCA-RbcX_2$ crystal structures which may play a role in $AnaCA-RbcX_2$'s high binding affinity to $RbcL_8$. The loop connecting helices α2 and α3 is shortened by one amino acid in the crystal structure of $AnaCA-RbcX_2$ and a Lys is present at position 48, introducing a positive charge at the binding cleft possibly increasing the affinity for the negatively charged C-terminal tail of RbcL (Saschenbrecker et al. 2007), but exact determination for the increase in affinity is difficult since this loop is disordered in the $AnaCA-RbcX_2$ crystal structure. Therefore, to potentially increase the charge/charge interaction of $Syn7002-RbcX_2$ to the C-terminal tail of RbcL, G42 and E43 were mutated to His and Asn, respectively to try to create a higher affinity assembly chaperone. These residues were chosen due to their close location to the binding cleft of $RbcX_2$ and as a way to introduce a positively charged and a polar amino acid by replacing a non polar and a negatively charged amino acid. In this way the sequence of the $Syn7002-RbcX_2$ protein would be more similar to the $AnaCA-RbcX_2$ sequence which may help to increase $Syn7002-RbcX_2$ affinity for the C-terminal tail of $RbcL_8$. Unexpectedly, mutation of these residues to the corresponding amino acids present in the $AnaCA-RbcX_2$ sequence did not result in the formation of the $RbcL_8/(RbcX_2)_8$ complex on Native-PAGE (Figure 4.7 lane 2). AnaCA-$RbcX_2$'s high affinity for $Syn6301-RbcL_8$ cores is most likely not only due to its interaction with the C-terminal tail of RbcL, but the peripheral interaction between RbcX and RbcL must also play an important role in determining $RbcX_2$ affinity to $RbcL_8$. In addition there could be other structural features of $AnaCA-RbcX_2$ that contribute to its high affinity to the C-terminal tail of RbcL including the shortened loop between the α2 and α3 helices.

1. Syn7002-RbcL/Syn7002-RbcX
2. Syn7002-RbcL/Syn7002-RbcX(G42H,E43N)
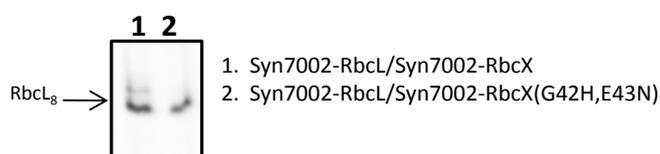
Figure 4.7 High affinity Syn7002-RbcX$_2$.
Syn7002-RbcL was coexpressed in *E. coli* with wildtype $Syn7002-RbcX_2$ (lane 1) or mutant $Syn7002-RbcX_2$ (lane 2). Soluble lysate was subjected to Native-PAGE followed by immunodetection with Anti-RbcL antibodies.

4.3   Analysis of the $RbcL_8/(RbcX_2)_8$ interaction

4.3.1   Cryo-EM structure of the $RbcL_8/(RbcX_2)_8$ complex

4.3.1.1 Examination of the cryo-EM structure

Detailed analysis of the functional $RbcL_8/(RbcX_2)_8$ complex has been hindered by the dynamic character of the interaction between RbcL and RbcX. It has therefore not been

possible to obtain information on the exact binding position of RbcX$_2$ in relation to RbcL. However, by taking advantage of the above mentioned high affinity interaction between Syn6301-RbcL and AnaCA-RbcX$_2$, stable Syn6301-RbcL$_8$/(AnaCA-RbcX$_2$)$_8$ complexes could be purified from *E. coli* and visualized by cryo-EM, along with visualization of Syn6301-RbcL$_8$ core complexes for comparison (Figure 4.8). (Cryo-EM pictures were creating in collaboration with Dr. A. Young and Prof. Dr. R. Beckmann.)
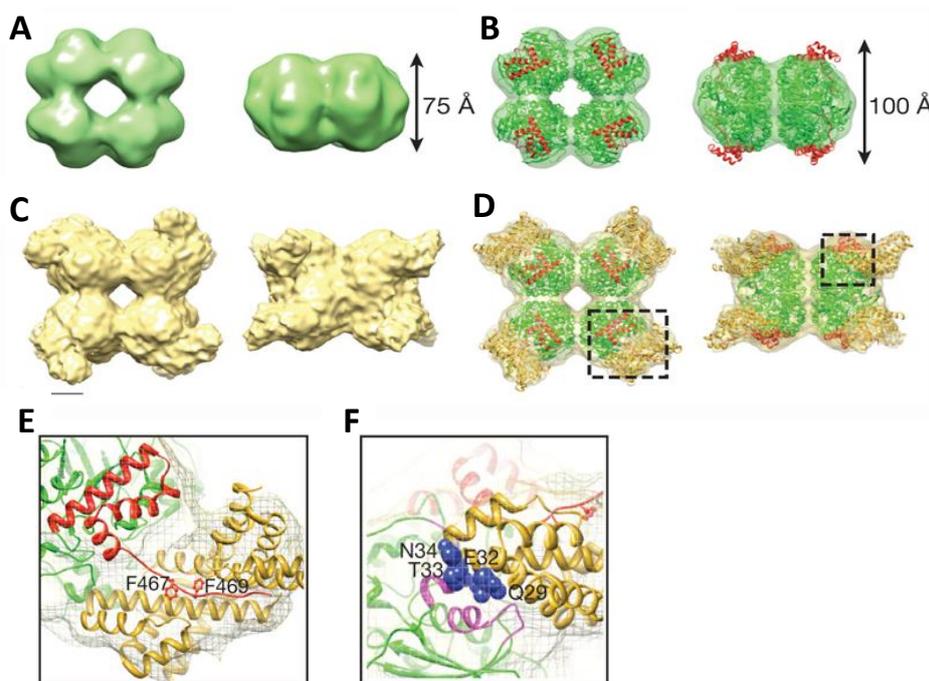
Figure 4.8 Cryo-EM maps of RbcL$_8$ and RbcL$_8$/(RbcX$_2$)$_8$ complexes.
(A) Surface representation of Syn6301-RbcL$_8$ density map, top view and side view; green. (B) Top view and side view of Syn6301-RbcL$_8$ crystal structure, ribbon representation, fitted as a rigid body in the surface representation density map from, A; transparent green. (C) Surface representation of Syn6301-RbcL$_8$/(AnaCA-RbcX$_2$)$_8$ density map, top view and side view; yellow. Scale bar is 50 Å. (D) Top view and side view of Syn6301-RbcL$_8$ crystal structure, green ribbon representation, and AnaCA-RbcX$_2$, gold ribbon representation, fitted as a rigid body in the surface representation density map from, C; transparent yellow. (E) Zoom-in of dashed region in, D left panel. C-terminus of Syn6301-RbcL is indicated in red (residues 413-475: spinach numbering). RbcL C-terminal tail is represented in its extended form and conserved phenylalanines (467 and 469: spinach numbering) are shown inserted into the hydrophobic pockets of the groove region of AnaCA-RbcX$_2$ which is shown in gold. (F) Zoom-in of dashed region in, D right panel. Conserved RbcX residues (Q29, E32, T33, N34) interacting with the adjacent RbcL are depicted by blue spheres, areas of RbcL interaction with RbcX are shown in pink. Adjacent RbcL subunit C-terminus binding to the RbcX$_2$ groove region is shown in red. (Liu et al. 2010)

The arrangement of subunits in the 16 Å cryo-EM density map of Syn6301-RbcL$_8$ (Figure 4.8 A) was similar to previously solved Rubisco holoenzyme structures (Newman and Gutteridge 1994). The RbcL$_8$ core was clearly present, displayed as four RbcL dimers, with a diameter of 110 Å. However, the height of the cryo-EM Syn6301-RbcL$_8$ structure was 25 Å less than the height of RbcL$_8$S$_8$ crystal structures. Upon fitting of Syn6301-RbcL$_8$ core from Syn6301-RbcL$_8$S$_8$ crystal structure into the Syn6301-RbcL$_8$ cryo- EM density map, specific regions of the crystal structure protruded outside of the density map (Figure 4.8 B). The protrusions correlated to the last 60 amino acids (residues 410-472, Syn6301 Rubisco numbering) of the RbcL C-terminus. These residues contain the final helix of the RbcL TIM-

barrel domain and also contain regions important for contacting RbcS (residues 1-37). The inability to resolve this C-terminal portion of RbcL indicated its inherent structural flexibility and could possibly explain why folded RbcL monomers exhibit high affinity and rebind to GroEL instead of being fully released during *in vitro* reconstitution. Ordering of RbcL C-terminal helices by $RbcX_2$ binding may be critical to prevent rebinding to GroEL.

Examination of the cryo-EM density for the $RbcL_8/(RbcX_2)_8$ complex revealed the location of $RbcX_2$ binding in relation to $RbcL_8$ cores (Figure 4.8 C). The $RbcL_8$ core region was readily identified and furthermore eight additional densities extended from the body of the $RbcL_8$ core. These densities corresponded to eight RbcX dimers located at the top and bottom of $RbcL_8$. Interestingly, upon fitting of $RbcL_8$ from the holoenzyme crystal structure into the $RbcL_8/(RbcX_2)_8$ density map, the RbcL C-terminal helices are visible and fit in the density map. This indicates that the RbcL C-terminal helices become structured with $RbcX_2$ bound to the $RbcL_8$ core (Figure 4.8 D). Increased structuring of this region by $RbcX_2$ was paralleled by the observed increase in protease protection of the $RbcL_8/(RbcX_2)_8$ complex compared to $RbcL_8$, and as expected, maximum protease protection was obtained upon holoenzyme formation (Figure 4.9 A). AnaCA-$RbcX_2$ also exhibited reduced protease sensitivity in the presence of RbcL C-terminal peptide; even though flexibility in the C-terminal portion was still evident (Figure 4.9 B). The limited-proteolysis results in Figure 4.9 B demonstrate that the α4 helix of $RbcX_2$ is not stably structured. Therefore, a new crystal form of AnaCA-$RbcX_2$ where residues 106-115 were disordered was used for rigid body fitting in the $RbcL_8/(RbcX_2)_8$ cryo-EM envelope. This shorter $RbcX_2$ version could be positioned significantly better than full length in the obtained density map further demonstrating the flexibility in the α4 helix as this portion was not visualized by cryo-EM.

It is important to note that $RbcX_2$ was optimally positioned, in relation to $RbcL_8$, for binding the extended RbcL C-terminal tail (Figure 4.8 E). Moreover, additional contacts of $RbcX_2$ with RbcL were evident. $RbcX_2$ not only made known contacts to the extended C-terminus of one RbcL subunit, but it also made contacts to the N-terminal domain of the adjacent RbcL subunit. $RbcX_2$ contacts two RbcL monomers within the RbcL anti-parallel dimer and seemingly acts like a 'molecular staple' within an RbcL dimer. $RbcX_2$ would facilitate dimer stabilization during $RbcL_8$ core formation. As observed from fitting of $RbcX_2$ in the cryo-EM structure (Figure 4.8 F), it was apparent that the residues of $RbcX_2$ contacting the RbcL N-terminus were the conserved peripheral residues Q29, E32, T33, and N34. Importantly, residues Q29 and E32 were reported to be functionally critical (Saschenbrecker et al. 2007).
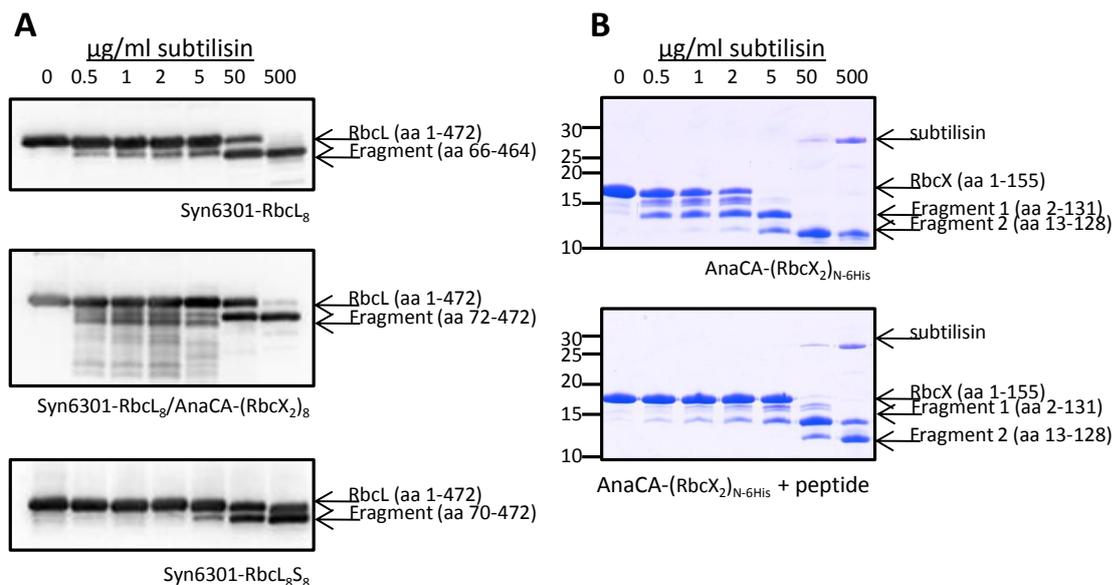
Figure 4.9 Limited proteolysis of RbcL$_8$, RbcL$_8$/(RbcX$_2$)$_8$, RbcL$_8$S$_8$, RbcX$_2$, and RbcX$_2$ with peptide.
(A) 1 mg/ml of Syn6301-RbcL$_8$, Syn6301-RbcL$_8$/(AnaCA-RbcX$_{2(N-6His)}$)$_8$, or Syn6301-RbcL$_8$S$_8$ was digested on ice with the indicated concentrations of subtilisin for 1 hr, whereupon digestion was quenched with 10 mM PMSF. Samples were analyzed on 16% SDS-PAGE followed by chemiluminescent detection with anti-RbcL antibodies. Fragment composition was determined by Edman degradation and mass spectrometry. (B) 1 mg/ml of AnaCA-RbcX$_{2(N-6His)}$ alone or with addition of Syn6301-RbcL C-terminal peptide was digested, analyzed, and fragment composition determined as above. Digested bands were visualized by coomassie staining of the gels.

4.3.1.2 Crosslinking of RbcL and RbcX

The peripheral interaction of RbcX$_2$ with RbcL, as observed from cryo-EM, was confirmed by site-specific crosslinker introduction into RbcX. The photo-reactive phenylalanine analog *para*-benzoyl-L-phenylalanine (*p*Bpa) was incorporated at specific peripheral sites in the RbcX dimer. This was accomplished by taking advantage of a recently developed orthogonal tRNA/aminoacyl-tRNA synthetase pairing method (Ryu and Schultz 2006). Using this method, unnatural amino acids can be specifically incorporated into the protein of interest. RbcX-expressing cells were cotransformed with the plasmid pSup-BpaRS-6TRN, which encodes six copies of the orthogonal *Amber* stop codon suppressor tRNA gene and a copy of an orthogonal tyrosyl-tRNA synthetase. The tRNA synthetase specifically charges the suppressor tRNA with *p*Bpa, which then subsequently incorporates *p*Bpa at TAG triplets in mutated RbcX (referred to as RbcX(TAG)$_2$). Efficient suppression of the *Amber* stop codon in AnaCA-RbcX$_2$ and Syn7002-RbcX$_2$ was determined, and only when cells were induced in the presence of *p*Bpa (I+: cells induced along with *p*Bpa addition) was full-length RbcX obtained (Figure 4.10 A). Suppression of the *Amber* stop codon was not 100% efficient; therefore in some instances, lower molecular weight bands were present, which were not degraded, indicating stable truncated versions of RbcX$_2$. This truncated RbcX$_2$ would not,

however, influence crosslinking in a detrimental way except that crosslinking efficiency could potentially be lowered due to binding competition of RbcX(TAG)$_2$ with truncated RbcX$_2$.
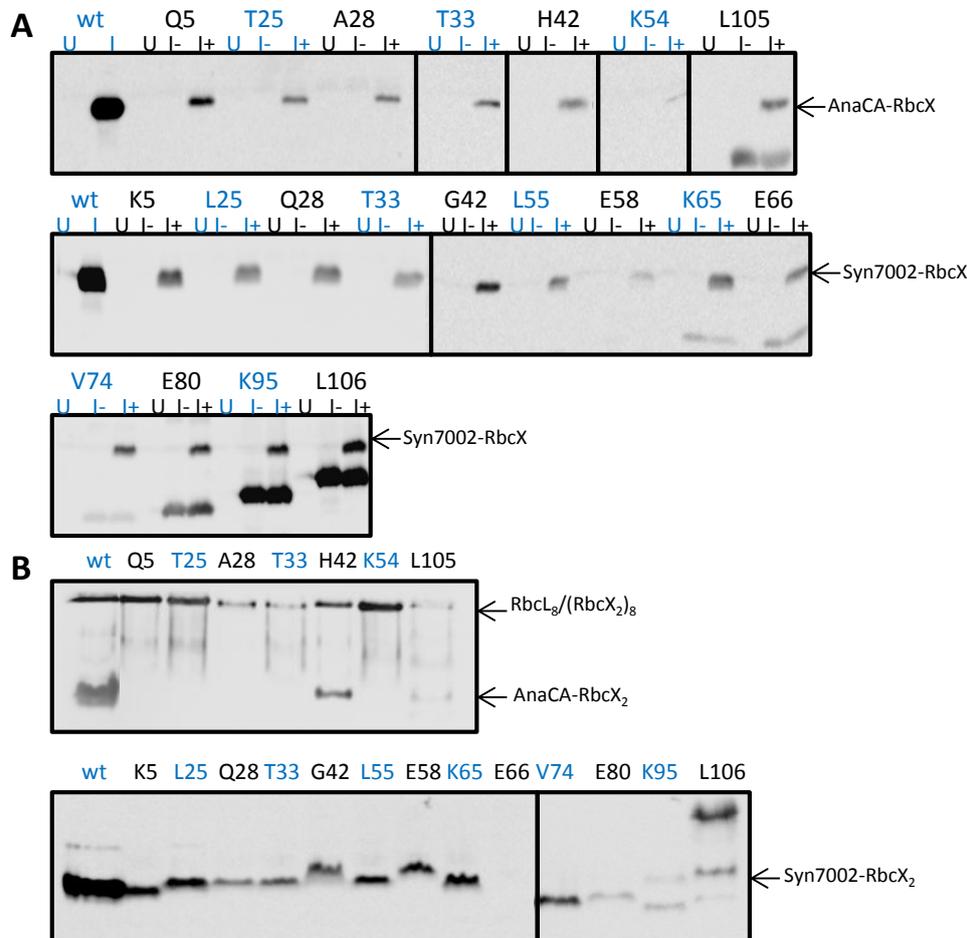
Figure 4.10 *Amber* stop codon suppression and RbcX$_2$(TAG) dimerization.
(A) Efficiency of *Amber* stop codon suppression in RbcX(TAG) mutants was determined by inducing (I) Syn6301-RbcL/AnaCA-RbcX (top) or Syn7002-RbcL/Syn7002-RbcX (bottom) in *E. coli* BL21 cells cotransformed with the *p*Sup-BpaRS-6tRN plasmid, either in the presence (I+) or absence (I-) of 1 mM *p*Bpa after induction. (U) uninduced. (wt) wildtype RbcX. Formation of full length RbcX after induction was analyzed by 16% SDS-PAGE followed by immunodetection with Anti-RbcX antibodies. (B) Proper folding of RbcX with site-specific *p*Bpa incorporation was visualized by 10% Native-PAGE of soluble cell lysate followed by immunodetection with Anti-RbcX antibodies. Syn6301-RbcL/AnaCA-RbcX (top) and Syn7002-RbcL/Syn7002-RbcX (bottom).

It was also necessary to ensure that the RbcX(TAG)$_2$ proteins folded properly and maintained native interactions with RbcL. It was demonstrated by Native-PAGE that incorporation of *p*Bpa in RbcX(TAG) did not influence RbcL$_8$/(RbcX$_2$)$_8$ complex assembly during coexpression of Syn6301-RbcL with AnaCA-RbcX, or RbcX dimer assembly during coexpression of Syn7002-RbcL with Syn7002-RbcX (Figure 4.10 B). Note that essentially all of the AnaCA-expressed RbcX(TAG)$_2$ was incorporated into the RbcL$_8$/(RbcX$_2$)$_8$ complex with little free RbcX$_2$ detected due to the low expression levels of the TAG mutant proteins compared to wildtype.
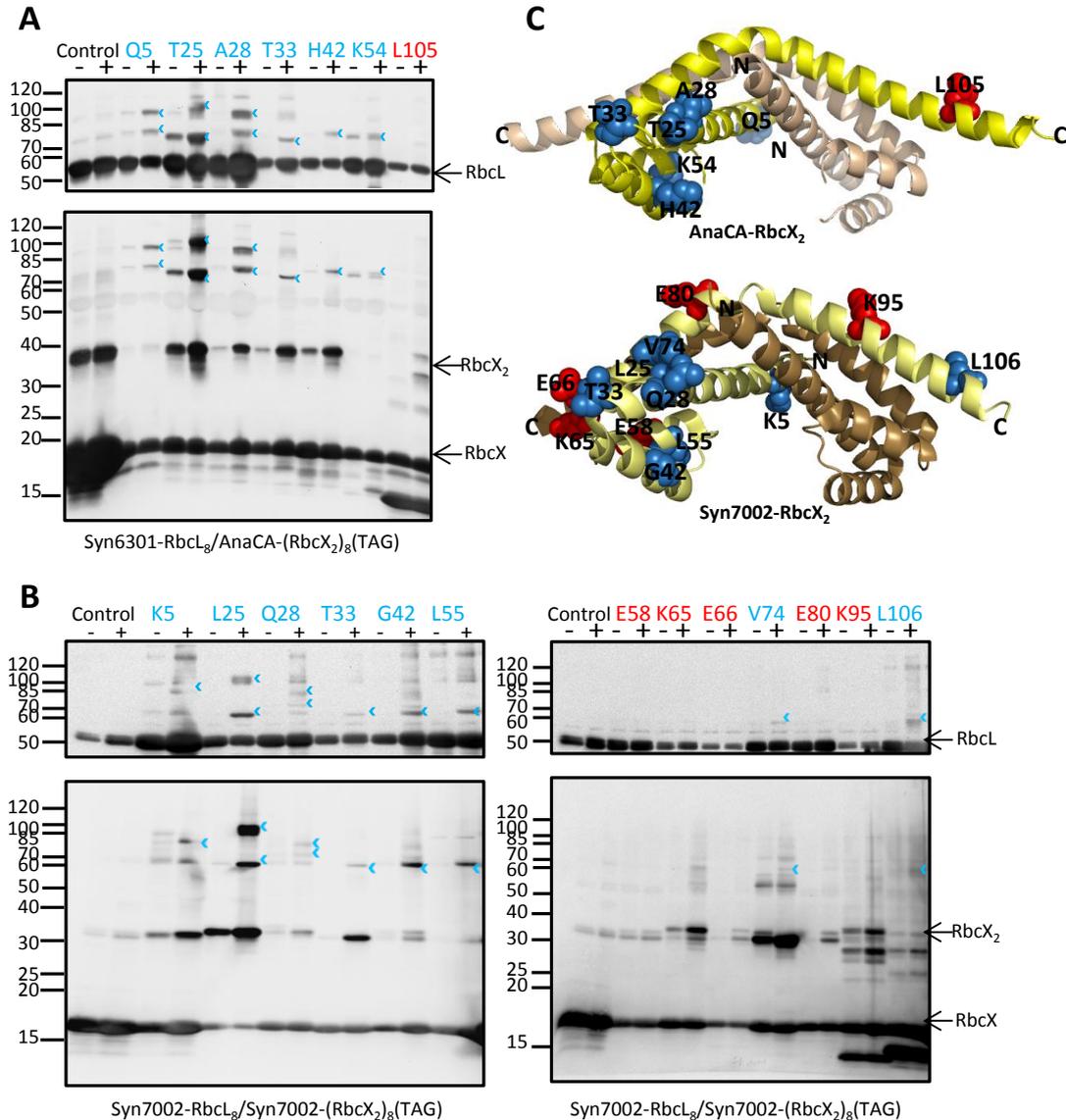
Figure 4.11 Crosslinking of heterologous and cognate RbcL$_8$/(RbcX$_2$)$_8$ complex with *p*Bpa.
(A) Purified Syn6301-RbcL$_8$/(AnaCA-RbcX$_2$)$_8$ complexes, with AnaCA-RbcX$_2$ containing the photo-reactive amino acid analog *p*Bpa, were isolated and crosslinked by exposure to UV light. UV exposed (+) and UV non-exposed samples (-) were subjected to 10-15% SDS-PAGE followed by immunodetection with anti-RbcL (top) and anti-RbcX (bottom) antibodies. Control lane is the wildtype Syn6301-RbcL$_8$/(AnaCA-RbcX$_2$)$_8$ complex without *p*Bpa incorporation in RbcX$_2$. Photoadducts are indicated with blue open arrowheads and positions with crosslinks are indicated in blue, positions that resulted in no crosslinking are indicated in red. (B) Purified Syn7002-RbcL$_8$/(Syn7002-RbcX$_2$)$_8$ complexes were isolated with *p*Bpa incorporation in Syn7002-RbcX$_2$. Samples were analyzed as in A. (C) Ribbon diagram of AnaCA-RbcX$_2$ and Syn7002-RbcX$_2$ indicating amino acid positions (one chain labeled) with *p*Bpa incorporation, positions resulting in crosslinks are shown in blue and non-crosslinked positions are in red. RbcX chains are indicated by two shades of yellow.

Crosslinking was accordingly performed in the heterologous Syn6301-RbcL/AnaCA-RbcX system (Figure 4.11 A) as well as in the dynamic cognate Syn7002-RbcLX system (Figure 4.11 B). Crosslinks were obtained after purification of RbcL$_8$/(RbcX$_2$)$_8$ complexes followed by irradiation under UV light. Positions which were UV-dependent and present during immunodetection with both anti-RbcL and anti-RbcX antibodies corresponded to a cluster of residues at the peripheral region of RbcX$_2$ (Figure 4.11 C). Similar crosslinks were obtained for both systems confirming that the contacts observed in the heterologous system

could be extrapolated to the biologically relevant/dynamic system. In some instances, more than one crosslink was indicated, which corresponded to additional crosslinking within the RbcX dimer itself (Figure 4.11 A and B). For RbcX(TAG)$_2$ mutants, the emergence of the RbcX$_2$ band on SDS-PAGE was UV-dependent. RbcX$_2$ was also observed in wildtype AnaCA-RbcX$_2$ on SDS-PAGE, but these crosslinks were UV-independent, indicating that AnaCA-RbcX$_2$ was partially SDS-resistant.

4.3.1.3 Mechanism of RbcX$_2$ during *in vitro* refolding of RbcL

The ability of AnaCA-RbcX$_2$ to assist Syn6301-RbcL assembly *in vitro* is possibly due to the high affinity of AnaCA-RbcX$_2$ to the C-terminal tail of Syn6301-RbcL. Complex formation appears to be helped by AnaCA-RbcX$_2$ by retrieving folded Syn6301-RbcL subunits from the GroEL/ES cage. However, the mechanism by which this is accomplished is still unknown. Therefore, various Syn6301-RbcL and AnaCA-RbcX mutants were analyzed for the formation of assembly intermediates during *in vitro* reconstitution.
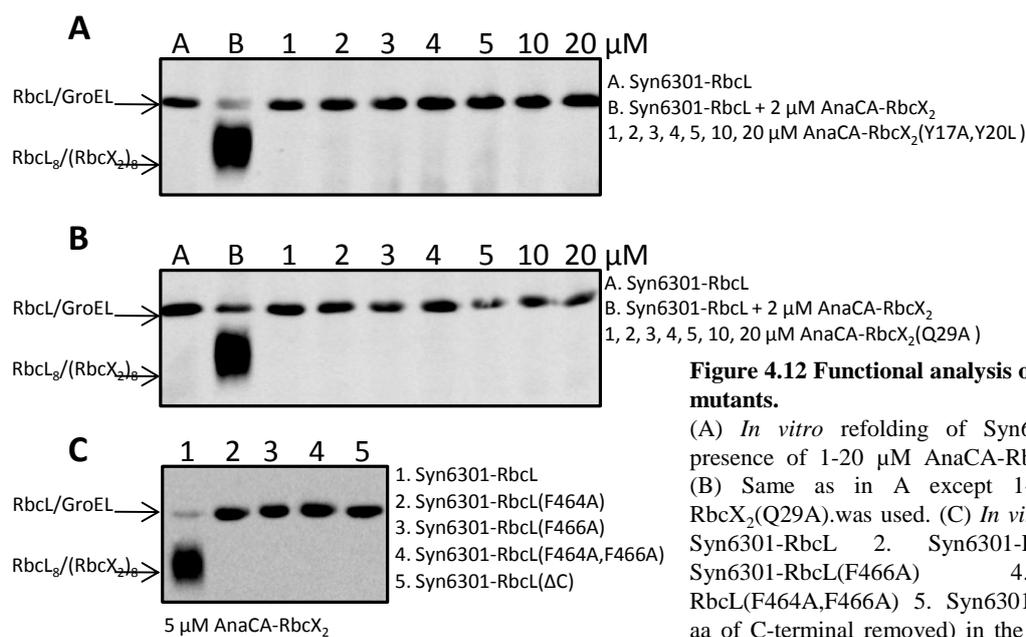
Figure 4.12 Functional analysis of RbcL and RbcX mutants.
(A) *In vitro* refolding of Syn6301-RbcL in the presence of 1-20 μM AnaCA-RbcX$_2$(Y17A,Y20L). (B) Same as in A except 1-20 μM AnaCA-RbcX$_2$(Q29A).was used. (C) *In vitro* refolding of 1. Syn6301-RbcL 2. Syn6301-RbcL(F464A) 3. Syn6301-RbcL(F466A) 4. Syn6301-RbcL(F464A,F466A) 5. Syn6301-RbcL(ΔC-last 12 aa of C-terminal removed) in the presence of 5 μM AnaCA-RbcX$_2$. Refolding assays were analyzed on 5-13% Native-PAGE followed by chemiluminescent detection with anti-RbcL antibodies. RbcL bound to GroEL is indicated as well as the migration of the RbcL$_8$/(RbcX$_2$)$_8$ complex.

As seen in Figure 4.12 A and B, when AnaCA-RbcX was mutated at the conserved binding groove (Y17A,Y20L) or peripheral (Q29A) regions, respectively, *in vitro* refolding and assembly of Syn6301-RbcL$_8$/(AnaCA-RbcX$_2$)$_8$ complex was not realized. This indicates that both functional regions of the RbcX homodimer are essential for preventing the re-association of folded RbcL to GroEL. Furthermore, mutation of the C-terminal residues

(F464A or F466A) or deletion of the entire C-terminal portion of Syn6301-RbcL also did not result in formation of assembled product and RbcL remained bound to GroEL even in the presence of wildtype AnaCA-RbcX$_2$ (Figure 4.12 C); denoting that RbcX$_2$ does not function by simply shielding this portion of RbcL to prevent rebinding to GroEL. It has previously been shown that deletion of the C-terminal recognition motif in Syn6301-RbcL abolishes RbcX$_2$-independent assembly to RbcL$_8$ cores (Saschenbrecker et al. 2007), leading to the conclusion that the extended C-terminal on RbcL has a direct role in RbcL$_8$ assembly or stability. Here, the importance of the C-terminal recognition motif for its role in efficient complex assembly is further demonstrated.

4.3.1.4 Role of RbcX$_2$ in RbcL$_8$ C-terminal stabilization

Superimposition of the cryo-EM density maps from Syn6301-RbcL$_8$ and Syn6301-RbcL$_8$/(AnaCA-RbcX$_2$)$_8$ complexes highlighted extra densities protruding from the top and bottom of the RbcL$_8$ cores (Figure 4.13 A and B). These not only corresponded to the densities of eight bound RbcX$_2$, but also to the last C-terminal helices, which were not resolved in the RbcL$_8$ cryo-EM density map. Upon binding of RbcX$_2$ to the C-terminal recognition motif of RbcL there seemed to be stabilization and concurrent ordering of these segments; the unstabilized RbcL segments appeared to be pulled back against the body of the RbcL$_8$ core complex, accounting for their visualization during cyro-EM imaging when RbcX$_2$ was bound. Stabilization of the C-terminal helices by binding of RbcX$_2$ to RbcL could be important for creating a docking platform for RbcS. In accordance with this suggestion, as seen above during *in vitro* translation system experiments (Figure 4.2 A), maximum solubility and activity of RbcL was only established when RbcX$_2$ is present with RbcS.
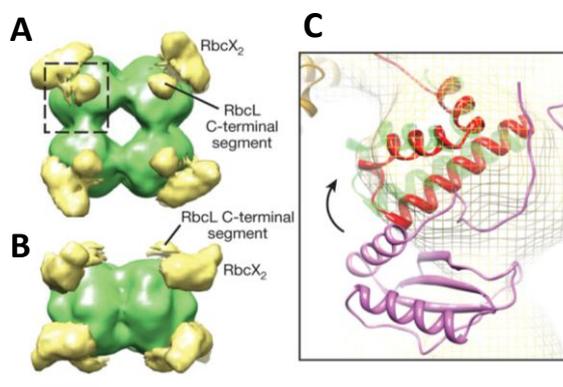
Figure 4.13 RbcL C-terminal shift with RbcX$_2$ bound.
Superimposition of Syn6301-RbcL$_8$ and Syn6301-RbcL$_8$/(AnaCA-RbcX$_2$)$_8$ density maps shown at a contour level of one: (A) Top view. (B) Side view. Areas of extra density from the Syn6301-RbcL$_8$/(AnaCA-RbcX$_2$)$_8$ density map are shown in yellow, corresponding to AnaCA-RbcX$_2$ and the ordered C-terminus of RbcL. Scale bar is 50 Å. (C) Zoom-in of dashed area indicated in A. Syn6301-RbcL C-terminus in red indicates position when AnaCA-RbcX$_2$ bound compared to position of holoenzyme crystal structure in green. Position of Syn6301-RbcS in the holoenzyme crystal structure is shown in purple, RbcL$_8$/(RbcX$_2$)$_8$ density map is shown in yellow mesh. Movement of the RbcL C-terminus upon RbcS binding is indicated by the arrow. When RbcX$_2$ is bound, a steric clash is observed between the C-terminus of RbcL (red) and RbcS (purple). (Liu et al. 2010)

Intriguingly, comparison of the C-terminal helices arrangement in the EM model with the crystal structure of the holoenzyme revealed a significant shift (8 Å) in the positioning of

these helices (Figure 4.13 C). The binding of RbcS to RbcL$_8$ thus seemed to induce a conformational shift in this region. When RbcS binds between RbcL dimers, the shift in the C-terminal helices could possibly be important for destabilizing the interaction of RbcX$_2$ with the body of RbcL$_8$, thereby inducing its proper displacement and release. Furthermore, RbcX$_2$ prevents rebinding of RbcL to chaperonin and enhances dimer formation by stapling two RbcL subunits together, since RbcL was not able to efficiently fold and assemble into an assembly competent form with chaperonin alone.

4.3.2 Crystal structure of RbcL$_8$/(RbcX$_2$)$_8$ complex

By further taking advantage of the high affinity interaction of AnaCA-RbcX$_2$ with the C-terminal tail of Syn6301-RbcL, crystals of the Syn6301-RbcL$_8$/(AnaCA-RbcX$_2$)$_8$ hetero-complex could be obtained from purified samples and the structure solved by molecular replacement to 3.2 Å by Dr. Andreas Bracher (Figure 4.14). The solved crystal structure fitted well to the above described cryo-EM envelope as evident by the high similarity in real space correlation coefficient (0.4863 and 0.4654). Accordingly, RbcL formed a core of four anti-parallel dimers bound by eight RbcX dimers, importantly RbcX$_2$ was located in between adjacent RbcL monomers with contacts to the C-terminal of one subunit and the N-terminal domain of the neighboring subunit.
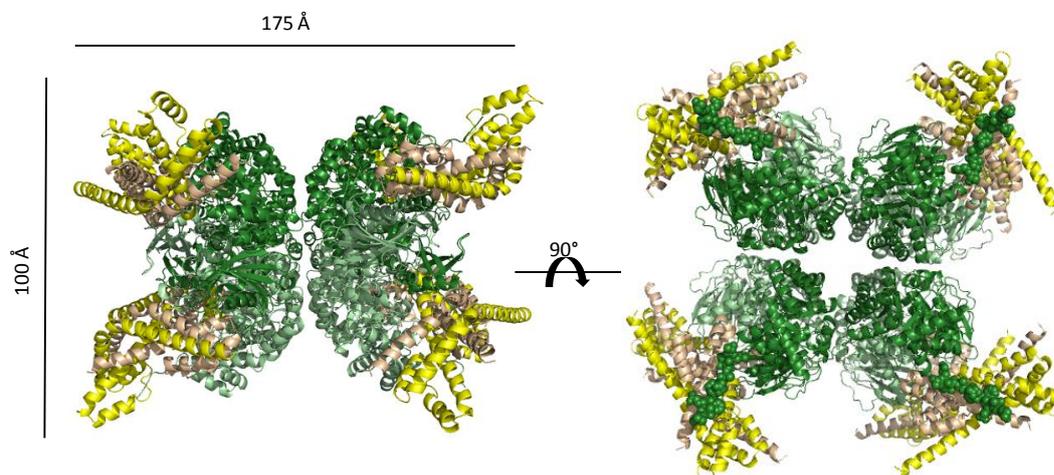
Figure 4.14 Crystal structure of the RbcL$_8$/(RbcX$_2$)$_8$ complex.
Side and top view of the RbcL$_8$/(RbcX$_2$)$_8$ crystal structure, respectively. RbcL chains are shown in green tones and RbcX chains are shown in yellow tones. Top view: RbcL C-terminal tail region bound to the RbcX$_2$ groove is indicated by sphere representation. Dimensions of the complex are indicated.

There were however, underlying differences between the cryo-EM envelope and the crystal structure of the RbcL$_8$/(RbcX$_2$)$_8$ complex. The molecules of RbcX$_2$ were more closely attached to the RbcL$_8$ body in the crystal structure, and the α4 helix of one chain of the RbcX dimer was 8 residues longer than its RbcX counterpart chain. The most striking difference

was observed in the RbcL C-terminal helices. As described above for the $RbcL_8/(RbcX_2)_8$ complex, there was an 8 Å displacement in the positioning of these helices upon $RbcX_2$ binding relative to the holoenzyme. However, this shift was not as prevalent in the crystal structure, where there was only a maximum displacement of 1.5 Å compared to the holoenzyme (possibly due to crystal packing); suggesting that this movement in the C-terminal domain would not produce a dramatic enough conformational shift to displace $RbcX_2$ from $RbcL_8$ upon RbcS binding. It is nonetheless demonstrated that $RbcX_2$ structures this region, since density was missing for the C-terminal helices in the cryo-EM reconstitution of the $RbcL_8$ core alone (Figure 4.8), which would promote RbcS binding for holoenzyme formation.

4.3.2.1 RbcL/RbcX interface

There were three observed areas of contact at the interface of anti-parallel $RbcL_2$ and $RbcX_2$. The largest (~800 Å^2) interface, area I, consisted of the RbcL extended C-terminus and the $RbcX_2$ binding groove (Figure 4.15 A and B). This was contributed by residues 458-468 of the RbcL C-terminal tail (large subunit residues are numbered according to the *Synechococcus* sp PCC6301 sequence, add 3 to obtain corresponding spinach Rubisco numbering). The highly conserved phenylalanines, F464^L and F466^L, were bound within the hydrophobic pockets of the $RbcX_2$ binding groove, which is highly similar to the previously described crystal structure of Syn7002-$RbcX_2$ in complex with RbcL C-terminal peptide EIKFEFD (Saschenbrecker et al. 2007). The orientation of $RbcX_2$ with the body of RbcL was most likely stabilized and rigidified by the indole side chain of W459^L and further by W382^L. A smaller RbcL/RbcX interaction area, area II encompassing residues L332^L and E333^L, was also observed on the C-terminal domain of this RbcL subunit, which are active site residues in loop 6. This loop was stabilized in the 'open' structure by the α1 helix of $RbcX_2$, but may be less important for $RbcX_2$ binding affinity, as judged by poor definition in the crystal structure.

Not only was there an interface formed between $RbcX_2$ with the C-terminal domain of one RbcL subunit, but there was also an interface formed with the N-terminal domain of the adjacent RbcL subunit, area III (Figure 4.15 A). The N-terminal RbcL interface involved in area III contacts consisted of residues 42-46, 49 and 53 in the αB helix including the preceding loop, and loop residues 123-126 which connect αC helix to βE beta-strand. The residues at this interface are highly conserved in form I RbcL sequences. This small interface harbored only ~450 Å^2 of buried RbcL surface; however there was a high degree of surface shape complementarity. Importantly, there were two polar, high-energy interactions

contributing to this interface, $E49^L$ with $R69^X$ and $G123^L$ with $E32^X$ (Figure 4.15 C), indicating that this area may be involved in the binding affinity of $RbcX_2$ to RbcL, and further that this interface has an important function in stabilizing the RbcL dimer. Close examination of the corresponding interface of $RbcX_2$ revealed that most of the contacts to RbcL, except for the binding groove interaction, were mediated by only one RbcX subunit of the dimer (Figure 4.15 B). The four-helix bundle of one protomer was clamped between the RbcL dimer while the other was positioned out towards the solvent.
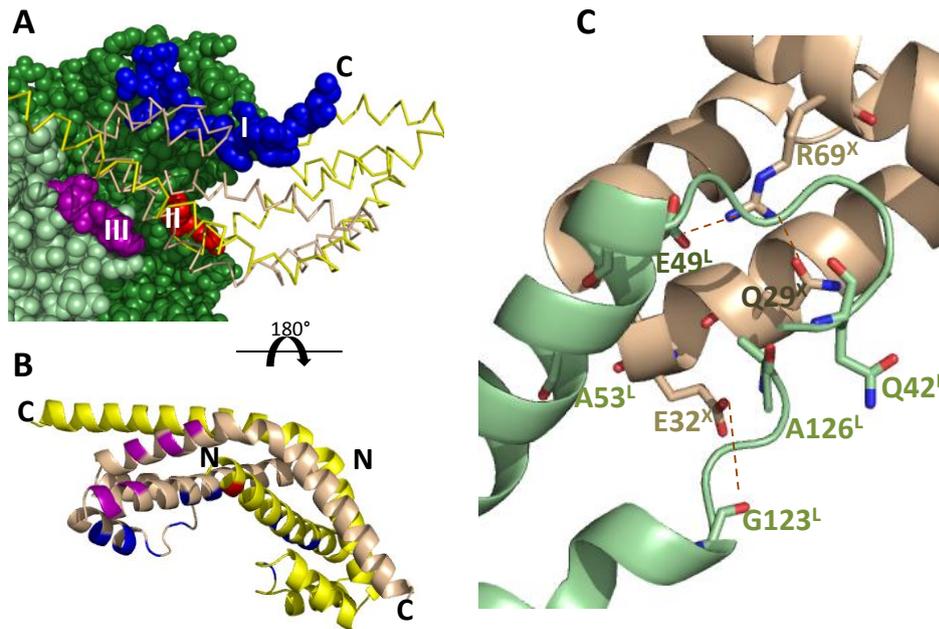
Figure 4.15 RbcL and RbcX interaction.
(A) Areas on RbcL (green) that interact with RbcX (yellow line representation). RbcL area I (blue) and area II (red) are located on one RbcL monomer while area III (purple) is located on the adjacent RbcL monomer. (B) RbcX dimer with RbcL interaction areas indicated, coloring as in, A with 180° rotation. (C) Close-up of area III RbcL/RbcX interaction, critical residues are shown in stick representation (oxygen and nitrogen labeled with red and blue, respectively).

The interactions of RbcX with RbcL visualized by the crystal structure coincide with observations from previous mutational studies of Syn7002-$RbcX_2$ (Saschenbrecker et al. 2007). The residues $Y17^X$ and $Y20^X$ in the $RbcX_2$ binding groove are essential for RbcL C-terminal peptide interaction (area I). It was reported that $E32^X$ and $R69^X$ (R70 for Syn7002-$RbcX_2$) are necessary for proper Rubisco assembly (Saschenbrecker et al. 2007). Here it is clearly demonstrated that they contributed crucial polar contacts to the adjacent RbcL subunit (area III). Furthermore, $Q29^X$ helped align $R69^X$ and stabilized the interaction with $E49^L$ while additionally providing van-der-Waals contacts to residues $Q42^L$, $V45^L$, and $A126^L$ (Figure 4.15 C). The conserved residues $T33^X$ and $V73^X$ (V74 for Syn7002-$RbcX_2$), which were found to be less mutation sensitive (Saschenbrecker et al. 2007), contributed hydrophobic contacts to RbcL. Residues whose mutation did not result in a defect ($N34^X$,

W40^X, E75^X (E76 for Syn7002-RbcX$_2$) (Saschenbrecker et al. 2007)) did not directly contact RbcL.
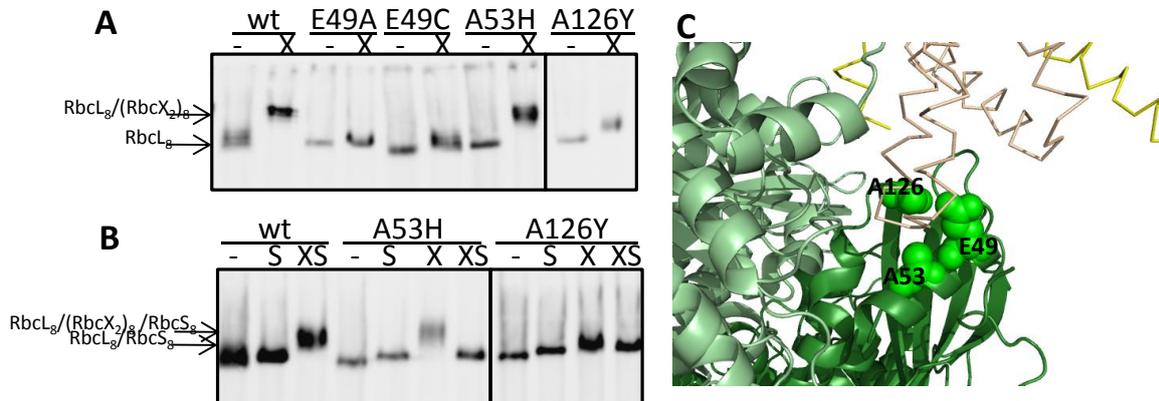
Figure 4.16 Analysis of the RbcL/RbcX area III interface.
(A) Syn6301-RbcL (wt or mutant: E49A, E49C, A53H, A126Y) was expressed in *E. coli* BL21 cells, soluble lysate was either directly (-) subjected to 6% Native-PAGE analysis followed by immunodetection with anti-RbcL antibodies, or after incubation for 1 hr at 25°C with 1 μM purified AnaCA-RbcX$_{2(N-FLAG)}$ (X). (B) Same as in A, except soluble lysate was incubated with 3 μM Syn6301-RbcS for 1 hr at 25°C alone (S) or following incubation with 1 μM AnaCA-RbcX$_{2(N-FLAG)}$ (XS). (C) Area III close-up of RbcL residues mutated in A and B shown by sphere representation. RbcL: green ribbon representation. RbcX$_2$: yellow Cα trace representation.

In order to validate the RbcX$_2$ contacts to the N-terminal domain of RbcL (area III, Figure 4.15 A), a series of mutations were introduced at conserved sites on RbcL. The ability of these mutants to form the RbcL$_8$/(RbcX$_2$)$_8$ complex was analyzed by Native-PAGE (Figure 4.16 A and B). Position E49 in RbcL was mutated to either an alanine or cysteine to identify the importance of salt bridge formation with R69^X while histidine and tyrosine were used to replace A53 and A126, respectively to disrupt the close surface shape complementarity contacts of RbcX with RbcL area III (Figure 4.16 C). Complex formation of wildtype Syn6301-RbcL$_8$ with wildtype AnaCA-RbcX$_2$ was clearly visualized as a migration up-shift in relation to RbcL$_8$ positioning on Native-PAGE, whereas this shift was not observed when position E49 on RbcL was mutated to alanine or cysteine and it was reduced when A126 was mutated to a tyrosine (Figure 4.16 A). This confirmed that salt bridge formation of E49^L with R69^X and the close contacts enabled by A126^L are critical for high-affinity complex formation. Furthermore, when position A53 on RbcL was mutated to a histidine, complex formation with AnaCA-RbcX$_2$ did not seem to be hindered; however, AnaCA-RbcX$_2$ could be completely displaced from the Syn6301-RbcL$_8$ core upon Syn6301-RbcS addition (also observed for A126Y) as seen by similar migration with the Syn6301-L$_8$S$_8$ holoenzyme, which was in contrast to the wildtype complex (Figure 4.16 B). This strongly suggests that affinity of AnaCA-RbcX$_2$ to Syn6301-RbcL$_8$ is reduced when residue 53 in RbcL is mutated to histidine, further exemplifying the importance of close contact of the RbcX$_2$ to the RbcL N-terminus for proper interaction.
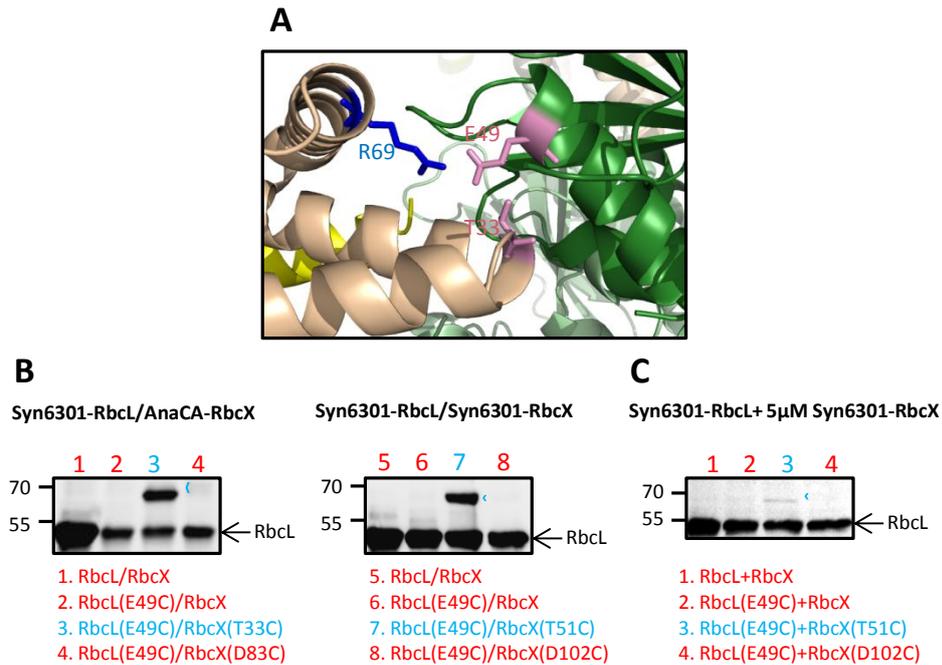
Figure 4.17 Site-specific crosslinking of heterologous and cognate RbcL$_8$/(RbcX$_2$)$_8$ complex.
(A) Visualization of the residues involved in the disulfide crosslink from the Syn6301-RbcL$_8$/(AnaCA-RbcX$_2$)$_8$ crystal structure shown in stick representation. RbcL is in green and RbcX is in yellow tones. Positions mutated to cysteines are colored pink. R69 on RbcX is also indicated in blue stick form. (B) RbcL and RbcX$_{N-6His}$ were coexpressed and the resulting complex purified under oxidizing conditions. Analysis of disulfide formation was carried out on non-reducing 10-15% SDS-PAGE followed by immunodetection with anti-RbcL antibodies. Crosslinks are indicated with blue open arrowheads. Coexpressions resulting in a crosslink are colored blue while coexpressions not resulting in a crosslink are colored red. 1. Wildtype RbcL and RbcX coexpressed 2. Syn6301-RbcL with cysteine at position 49 and wildtype RbcX$_2$ 3. Syn6301-RbcL with cysteine at position 49 and AnaCA-RbcX$_2$ with cysteine at position 33 or Syn6301-RbcX$_2$ with cysteine at corresponding position, 51 4. Syn6301-RbcL with cysteine at position 49 and AnaCA-RbcX$_2$ with cysteine at position 83 or Syn6301-RbcX$_2$ with cysteine at corresponding position, 102. Lanes 5-8 have additional mutation, F342I, in RbcL. (C) 5 μM purified Syn6301-RbcX$_{2(N-6His)}$ (wildtype, T51C, or D102C) were incubated in oxidizing conditions with soluble lysate of cells expressing Syn6301-RbcL (wildtype or E49C). Samples analyzed by 12.5% non-reducing SDS-PAGE followed by immunodetection with anti-RbcL antibodies. Labels as in B. Note, F342I mutation in RbcL is not present in C.

Confirmation that salt bridge formation also occurs between RbcX$_2$ and the N-terminus of RbcL in the cognate complex was verified by site-specific disulfide bridge crosslinking between RbcX and RbcL. Cysteines were introduced at residues E49 in Syn6301-RbcL and T51 in Syn6301-RbcX$_2$ (corresponding to T33 in AnaCA-RbcX$_2$), which are located directly opposed to each other in the heterologous complex (Figure 4.17 A). RbcL and RbcX were coexpressed in *E. coli* and complexes purified under oxidizing conditions. It is necessary to note that the additional mutation, F342I, which stabilizes Syn6301-RbcL$_8$ (Mueller-Cajar and Whitney 2008), was necessary for toleration of the E49C mutation in RbcL with coexpression of cognate RbcX$_2$. As predicted from the structural model, site-specific disulfide bond formation occurred between C49^L and AnaCA-RbcX(C33)$_2$ (Figure 4.17 lane 3) and C49^L and Syn6301-RbcX(C51)$_2$ (Figure 4.17 lane 7). Conversely, no disulfide bond formation occurred between C49^L and AnaCA-RbcX(C83)$_2$ (Figure 4.17 lane 4) or C49^L and Syn6301-RbcL(C102)$_2$ (Figure 4.17 lane 8). Disulfide formation in the

cognate complex could also be observed in the absence of the additional F342I mutation in RbcL. Syn6301-RbcL(49C) expressed alone in *E. coli* was able to form RbcL$_8$ cores (Figure 4.16 A). Upon incubation of soluble lysate of *E. coli* cells expressing Syn6301-RbcL(49C) with purified wildtype Syn6301-RbcX$_2$ or Syn6301-RbcX$_2$ cysteine mutants, disulfide formation was only observed when Syn6301-RbcX$_2$ had a cysteine residue at position 51 (Figure 4.17 C lane 3). This demonstrates that the contacts observed in the heterologous complex can be extrapolated to the cognate complex and are biologically relevant.

4.3.2.2 Mechanism of RbcX$_2$ function in RbcL dimer assembly

Recent studies have reported that RbcX$_2$ interacts with RbcL immediately after release from chaperonin (Liu et al. 2010; Saschenbrecker et al. 2007), but so far it is unknown how this interaction supports RbcL assembly. It is highly probable that RbcX$_2$ initially interacts with the C-terminal tail of RbcL (area I) with subsequent interaction of RbcX$_2$ with the N-terminal domain of the adjacent RbcL (area III) supporting dimer formation. As demonstrated, both areas are crucial for proper chaperone function of RbcX$_2$.
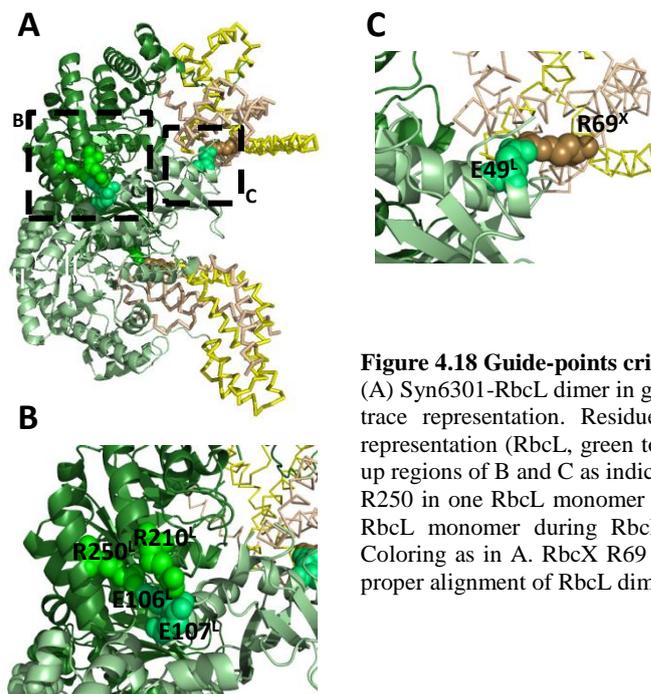
Figure 4.18 Guide-points critical for RbcL dimer formation.
(A) Syn6301-RbcL dimer in green ribbon representation, AnaCA-RbcX$_2$ in yellow Cα trace representation. Residues involved in dimerization are shown with sphere representation (RbcL, green tones and RbcX, brown). Dashed boxes designate close-up regions of B and C as indicated. (B) Close-up of dashed box B. Residues R210 and R250 in one RbcL monomer pair with E107 and E106, respectively of the opposing RbcL monomer during RbcL dimer formation. (C) Close-up of dashed box C. Coloring as in A. RbcX R69 forms a salt bridge with E49 of RbcL further ensuring proper alignment of RbcL dimers during assembly.

RbcX$_2$ most likely has a fundamental role in RbcL dimerization assembly and/or stability before final oligomerization, but the mechanism remained to be determined. Contacts at the RbcL dimer interface are quite large and hydrophobic, and improper interactions of this hydrophobic surface could lead to the misassembly/aggregation of RbcL. Moreover, there are two pairs of buried salt bridges (E106 with R250 and E107 with R210) at the RbcL dimer interface, that are directional and might function as guide-points for dimer assembly and add

to $RbcL_2$ stability (Figure 4.18 A and B). These contacts have also been reported in a recent analysis of *Chlamydomonas reinhardtii* Rubisco subunit interface dynamics, and the salt bridge contacts were discovered to be highly conserved in form I Rubisco (van Lun et al. 2011). From analysis of the $RbcL_8/(RbcX_2)_8$ crystal structure, an additional guide-point is contributed by $RbcX_2$, arising from salt bridge formation between residues $R69^X$ and $E49^L$ (Figure 4.18 C); indicating that $RbcX_2$ provides further positional information necessary for RbcL dimerization.
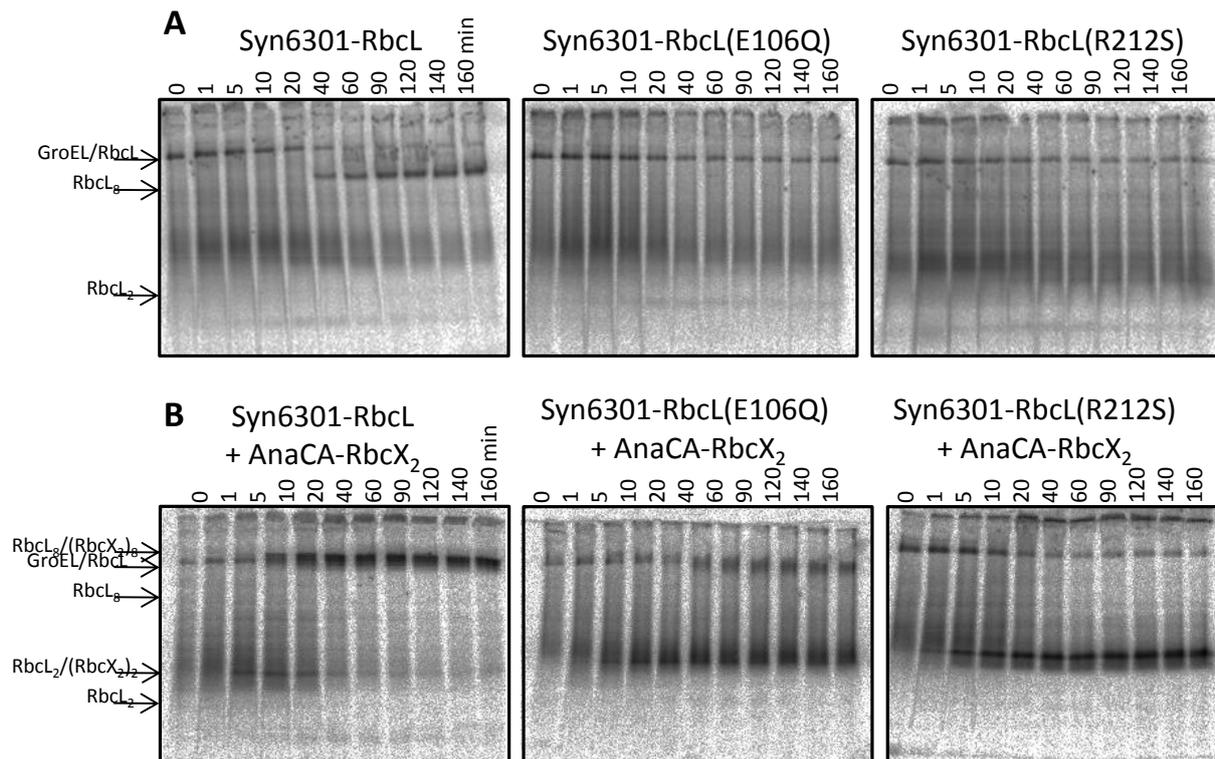
Figure 4.19 RTS pulse/chase: mechanism of $RbcX_2$ in RbcL dimer stabilization
(A) RTS100 HY *E. coli in vitro* transcription/translation system pulse/chase: Syn6301-*rbcL*-pET11a, Syn6301-*rbcL(E106Q)*-pET11a, or Syn6301-*rbcL(R212S)*-pET11a were translated with the RTS system in the presence of ^{35}S-met followed by a chase with unlabelled methionine at 30°C. Samples were taken at the indicated time points and subjected to 6% Native gel analysis followed by autoradiography. (B) Same as in A, except 40 µM AnaCA-$RbcX_2$ was present during transcription/translation.

The importance of the negative charge of E106 for RbcL dimer assembly was clearly revealed through the use of the isosteric mutation E106Q. Wildtype Syn6301-RbcL was inherently capable of $RbcL_8$ formation on Native-PAGE when expressed in an *in vitro* transcription/translation system; contrary to Syn6301-RbcL(E106Q) where no $RbcL_8$ formation was observed (Figure 4.19 A). If AnaCA-$RbcX_2$ was present during *in vitro* translation of Syn6301-RbcL(E106Q), formation of an assembly intermediate was visible and complex assembly was partially rescued (Figure 4.19 B). The role of $RbcX_2$ in RbcL dimer stabilization was further demonstrated when RbcL dimer-dimer assembly was blocked. The RbcL dimer-dimer interface is polar and exhibits three areas of surface shape

complementarity. Substitution to serine at position 212^L abolished the interaction of the arginine guanidinium group with the peptide backbone of the adjacent dimer and the Syn6301-RbcL(R212S) mutant could not assemble to $RbcL_8$ cores in an *in vitro* transcription/translation system (Figure 4.19 A). A stable assembly intermediate was only observed when $AnaCA-RbcX_2$ was present during *in vitro* translation of Syn6301-RbcL(R212S) (Figure 4.19 B), which is most likely the $RbcL_2/(RbcX_2)_2$ intermediate. The $RbcL_2/(RbcX_2)_2$ intermediate was observed initially when wildtype Syn6301-RbcL was translated *in vitro* in the presence of $AnaCA-RbcX_2$ followed by final assembly and formation of the $RbcL_8/(RbcX_2)_8$ complex (Figure 4.19 B), but in the case of Syn6301-RbcL(R212S) there was a stalling of complex formation at the $RbcL_2/(RbcX_2)_2$ intermediate stage.

A

RbcL_8/(RbcX_2)_8
RbcL_8/(X_2)_8/S →
RbcL_8 →

RbcL_2/(RbcX_2)_2 →

6% native

12.5% SDS
Soluble

1. RbcL
2. RbcL/AnaCA-RbcX
3. RbcL/AnaCA-RbcX/RbcS
4. RbcL(E106Q)
5. RbcL(E106Q)/AnaCA-RbcX
6. RbcL(E106Q)/AnaCA-RbcX/RbcS
7. RbcL(R212S)
8. RbcL(R212S)/AnaCA-RbcX
9. RbcL(R212S)/AnaCA-RbcX/RbcS

B

Figure 4.20 Role of RbcX and RbcS in RbcL dimerization.
(A) Soluble lysate of cells expressing Syn6301-RbcL (wt, E106Q, or R212S) alone or coexpressed with AnaCA-RbcX or Syn6301-RbcS was analyzed by 6% Native-PAGE and 12.5% SDS-PAGE followed by immunodetection with anti-RbcL antibodies. (B) Location of mutated residues, E106Q and R212S, on RbcL crystal structure involved in dimerization and oligomerization, respectively.

Further support of the role of $RbcX_2$ in RbcL dimer stabilization was also evident during *in vivo* coexpression of Syn6301-RbcL(E106Q) and Syn6301-RbcL(R212S) with $AnaCA-RbcX_2$ in *E. coli*. There was a high tendency of Syn6301-RbcL(E106Q) to aggregate when expressed alone (Figure 4.20 A lane 4). However, increased solubility and partial rescue of the $RbcL_2/(RbcX_2)_2$ intermediate could been seen as well as the slight tendency towards higher oligomer formation when $AnaCA-RbcX_2$ was coexpressed (Figure 4.20 A lane 5). Notably, complete rescue of complex formation of Syn6301-RbcL(E106Q) was observed when both $AnaCA-RbcX_2$ and Syn6301-RbcS were coexpressed (Figure 4.20 A lane 6) In contrast, when Syn6301-RbcL(R212S) was expressed alone, a significant amount remained soluble, although there was no assembly to $RbcL_8$ (Figure 4.20 A lane 7) instead there was a faint smear migrating around the level of RbcL dimers. Strong stabilization of the

$RbcL_2/(RbcX_2)_2$ intermediate was obtained when $AnaCA-RbcX_2$ was coexpressed (Figure 4.20 A lane 8) as well as complete rescue of complex formation when both $AnaCA-RbcX_2$ and Syn6301-RbcS were coexpressed (Figure 4.20 A lane 9). These results indicate how RbcS shifts the equilibrium towards complex formation by reinforcing stability of the mutant complexes.

4.3.2.3 The role of RbcS in $RbcX_2$ displacement and Rubisco catalysis

The role of $RbcX_2$ in mediating RbcL dimer formation has been described above; however the mechanism of $RbcX_2$ displacement by RbcS from $RbcL_8$ cores remained to be elucidated. Comparison of the $RbcL_8/(RbcX_2)_8$ structure with the holoenzyme structure revealed possible insight into the displacement mechanism. In the $RbcL_8/(RbcX_2)_8$ structure, the density of the N-terminal 18 amino acids of RbcL was missing as well as the densities for residues 64-70 comprising the '60ies loop' (residues 58-76) (Figure 4.21 A). These regions were structured in the holoenzyme due to RbcL dimer interactions within itself and with RbcS (Figure 4.21 B). Interestingly, in the $RbcL_8/(RbcX_2)_8$ complex, the '60ies loop' and the N-terminus must be displaced from holoenzyme positions, due to the steric clash that would ensue between the RbcL '60ies loop' and the N-terminus upon $RbcX_2$ binding to area III (Figure 4.21 C). The binding of RbcS in between two RbcL dimers should promote structuring of these regions, thereby effecting $RbcX_2$ release. As previously noted in section 4.2.1 during *in vitro* reconstitution of Syn6301-RbcL, a smeared band appeared when $AnaCA-RbcX_2$ was present, coinciding with the $RbcL_8/(RbcX_2)_8$ complex. Upon addition of RbcS, this band was slightly shifted, however it did not comigrate with $RbcL_8S_8$ (Figure 4.4 lane 2 and 8) and it was unclear what the configuration of this complex was. From the examination of the $RbcL_8/(RbcX_2)_8$ crystal structure it is highly unlikely that $RbcX_2$ and RbcS bind simultaneously on $RbcL_8$. However, it may be possible for $AnaCA-RbcX_2$ to remain bound to the C-terminus of RbcL upon cognate RbcS binding in a different manner than in the crystal structure and the smeared band represents this $RbcL_8/(RbcX_2)_8/RbcS_8$ complex.

To test the proposed mechanism of $RbcX_2$ displacement from $RbcL_8$ by RbcS, the removal of Syn6301-RbcL from immobilized Syn6301-RbcX$_2$ by Syn6301-RbcS was measured (Figure 4.22 A). As observed for the wildtype system, addition of increasing concentrations of Syn6301-RbcS to Syn6301-RbcL$_8$ bound to Syn6301-RbcX$_{2(N-FLAG)}$ on FLAG beads, caused the removal of $RbcL_8$ from $RbcX_2$; this was also observed for Syn6301-RbcL with the first 11 amino acids deleted. Evidently enough of the N-terminus is present that upon RbcS binding, $RbcX_2$ could still be displaced. In contrast, the RbcL mutant

W67A,L70G,L71A and the RbcS mutant L60A,F63A, which contain mutations affecting the RbcL-RbcS interface, could not be displaced from $RbcX_2$ by RbcS (Figure 4.22 A and D). All of the mutants still maintained the ability to bind RbcS in the absence of $RbcX_2$ (Figure 4.22 B and C); however, in the case of the RbcS mutant L60A,F63A the binding to RbcL was not as efficient as the RbcL mutants binding to wildtype RbcS. It is important to note, that there was no apparent detectable catalytic activity for any of the above described mutants (Figure 4.22 B).
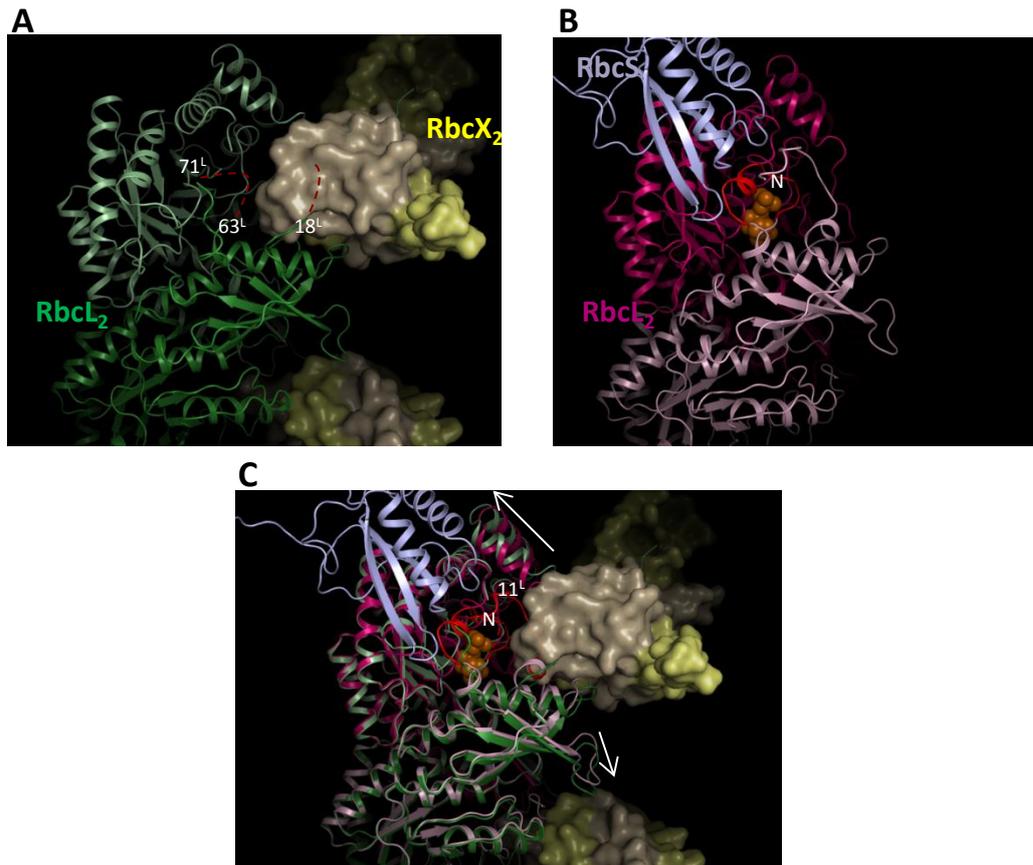
Figure 4.21 Superimposition of $RbcL_8/(RbcX_2)_8$ complex and CABP bound $RbcL_8S_8$.
(A) Close-up of the RbcLX interaction. $RbcL_2$ is shown in green ribbon representation and $RbcX_2$ in yellow surface representation. The '60ies loop' and the ultimate N-terminus which are disordered are represented by dashed maroon lines. (B) Close-up of the $RbcL_8S_8$ crystal structure. $RbcL_2$ is shown by pink ribbon representation and RbcS is indicated by light blue ribbon representation. The bound inhibitor CABP is shown by orange sphere representation showing the holoenzyme active site. The '60ies loop' of RbcL is colored red and the RbcL N-terminus is labeled. (C) Superimposition of A and B. The steric clash between $RbcX_2$ and structured RbcL '60ies loop' and N-terminal can be seen, red. Arrows indicate RbcL movement when $RbcX_2$ bound. Figure was prepared by Dr. Andreas Bracher.

RbcS is necessary for full catalytic function of form I Rubisco, but the reason for this requirement is still unclear since RbcS does not directly contact the reactive centers of the RbcL dimer interfaces (Andersson 2008; Spreitzer 2003). Nonetheless, inspection of the catalytically inactive $RbcL_8/(RbcX_2)_8$ crystal structure provided clarification of this issue. The densities of the first 18 amino acids of the $RbcL_8/(RbcX_2)_8$ crystal structure were missing; however only the first five amino acids are not resolved in the Rubisco holoenzyme crystal

structure (Newman et al. 1993). Consistent with the results shown here (Figure 4.22 B), previous studies have shown that removal of the first five amino acids of Rubisco does not deleteriously impact catalytic activity, while the first 11 amino acids are vital (Gutteridge et al. 1986). Accordingly, creation of a chimeric RbcL by replacement of the first 12 amino acids with homologous sequences, displayed characteristics similar to wildtype; however deletion of the first six amino acids and replacement of the next six amino acids with an unrelated amino acid sequence resulted in 90% reduction in catalytic activity compared to wildtype (Kettleborough et al. 1987) leading to the conclusion that the N-terminal is in some way involved in forming the binding site for RuBP. Here, deletion of the first five amino acids of the large subunit did not detrimentally affect catalytic activity when RbcS was bound, while activity was basically nonexistent when the first 11 amino acids were deleted (Figure 4.22 B). This further reveals the importance of the RbcL ultimate N-terminus in Rubisco catalytic activity. Upon RbcS binding to $RbcL_8$, the N-terminus is structured leading to proper active site formation. Accordingly, in the $RbcL_8/(RbcX_2)_8$ crystal structure these residues were disordered, hence this complex was catalytically inactive.

Detailed examination of the crystal structure of the Rubisco holoenzyme active site in the closed conformation with the transition state analog CABP bound, reveals important active site contacts with the '60ies loop'. The side chains of $T62^L$ and $W63^L$ are in direct contact with CABP (Newman et al. 1993) and in all form I Rubisco holoenzyme structures the '60ies loop' is ordered (except for one example where crystal packing most likely causes disorder (Duff et al. 2000)). In the $RbcL_8/(RbcX_2)_8$ crystal structure, however, these residues of the '60ies loop' were displaced and residues 64-70 disordered (Figure 4.21 A); furthermore, binding of RbcS to the RbcL mutant W67A,L70G,L71A did not produce an enzymatically activity Rubisco (Figure 4.22 B). Moreover, RbcS mutant L60A,F63A addition to wildtype RbcL resulted in no detectable catalytic activity compared to wildtype (Figure 4.22 B). Residues of RbcS which contact the RbcL '60ies loop' are highly conserved and mutations of these RbcS regions in *Chlamydomonas reinhardtii* compromise overall Rubisco function (Flachmann and Bohnert 1992; Genkov and Spreitzer 2009; Read and Tabita 1992). These results indicate that the '60ies loop' in the above described mutations remains unstructured and RbcS binding is necessary to stabilize the '60ies loop' in a catalytically competent conformation, providing compelling structural evidence of the role of RbcS in Rubisco catalysis.
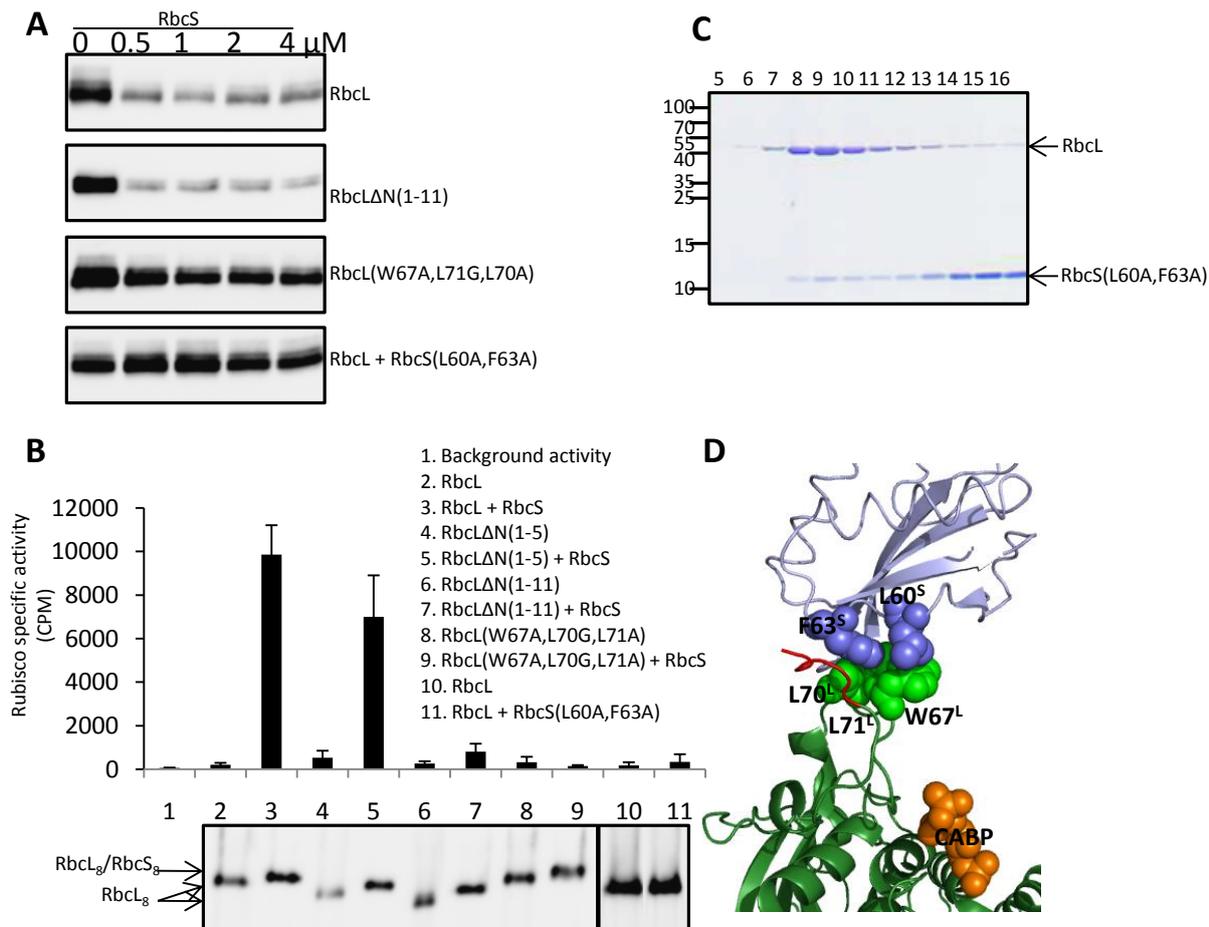
Figure 4.22 RbcS displacement of RbcX₂ from RbcL₈.
(A) Syn6301-RbcL/(Syn6301-RbcX$_{2(N-FLAG)}$), Syn6301-RbcLΔN(1-11)/(Syn6301-RbcX$_{2(N-FLAG)}$), or Syn6301-RbcL(W67A,L70G,L71A)/(Syn6301-RbcX$_{2(N-FLAG)}$) complexes were coimmunoprecipitated with EZview Red ANTI-FLAG M2 Affinity Gel. Displacement of RbcX₂ from RbcL₈ was analyzed by adding increasing concentrations of Syn6301-RbcS or Syn6301-RbcS(L60A,F63A) to the bound complexes as indicated. Protein that remained bound after RbcS addition was removed under denaturing conditions and subjected to 12.5% SDS-PAGE analysis followed by immunodetection with anti-RbcL antibodies. (B) Upper panel: specific activity of mutants described in A. Lower panel: RbcS and RbcS(L60A,F63A) addition to RbcL wt or RbcL mutants as indicated. Soluble lysate was incubated for 1 hr at 25°C with 3 μM RbcS or RbcS(L60A,F63A) followed by analysis on 6% Native-PAGE and immunodetection with anti-RbcL antibodies. (C) Size exclusion chromatography of 10 μM Syn6301-RbcL₈ (monomer concentration) incubated for 1 hr 25°C with 20 μM Syn6301-RbcS(L60A,F63A) analyzed by 12.5% SDS-PAGE followed by coomassie staining. (D) Location of mutants described in A-C in the Syn6301-RbcL₈S₈ crystal structure. RbcS is shown in light blue and RbcL is in green. Mutations are indicated by sphere representation. Bound CABP is shown in orange.

4.4   Analysis of eukaryotic Rubisco and refolding components

The recently achieved ability to reconstitute cyanobacterial form I Rubisco *in vitro* from purified components (Liu et al. 2010) gives insight into the constituents which are essential for efficient holoenzyme folding and assembly. Specifically, in the described method using Syn6301-RbcL as the substrate, GroEL/ES must be present for refolding of RbcL subunits and RbcX₂ for assembly. The RbcX₂ used for *in vitro* refolding of Syn6301 Rubisco must harbor a high enough binding affinity to prevent recapture by chaperonin. The concentration of high affinity heterologous RbcX₂ must be precisely controlled since high concentrations are inhibitory for the holoenzyme formation. Importantly, to effectively remove high affinity

$RbcX_2$ from the *in vitro* assembled $RbcL_8$ cores, the C-terminal peptide has to be added to the assay as a competitor after assembly (Liu et al. 2010). The *in vitro* reconstitution of cyanobacterial Rubisco is well established, and it was hypothesized that the same components or their homologues would also be sufficient for *in vitro* reconstitution of eukaryotic Rubisco. Therefore, the various components deemed to be essential for Rubisco *in vitro* reconstitution were analyzed using eukaryotic homologues i.e., *Arabidopsis thaliana* $RbcX_2$ and chaperonin.

A. thaliana contains two $RbcX_2$ homologues in its genome, $RbcXI_2$ and $RbcXII_2$ (Figure 2.22). The binding behavior of both analogues to the RbcL C-terminal recognition motif was analyzed here. Purification of $RbcX_2$ from *A. thaliana* (At-$RbcX_2$), without putative transit peptide as indicated in Figure legends, revealed significant differences in its behavior compared to its cyanobacterial homolog. Efficient expression and solubility of At-$RbcX_2$ was only obtained using the plasmid pHUE, which adds a cleavable ubiquitin tag to expression proteins. At-$RbcX_2$ could be cleaved of its ubiquitin tag with minimal loss of solubility as long as the pH was ≥ 8. However, At-$RbcX_2$ precipitated and aggregated at concentrations above 1 mg/ml. For this reason crystals of this eukaryotic assembly chaperone could not be obtained for structural analysis, nevertheless the concentration was sufficient for use in biochemical analyses to determine if At-$RbcX_2$ has a similar function in assembly as its bacterial homolog.

A fundamental characteristic of cyanobacterial $RbcX_2$ is its ability to bind the conserved C-terminal recognition motif of RbcL (Figure 4.3). This binding characteristic was also analyzed for the two *A. thaliana* $RbcX_2$ homologues (Figure 4.23). Incubation of purified At-$RbcXI_2$ and At-$RbcXII_2$ with a peptide array consisting of the last 40 amino acids of RbcL from 20 different species revealed that these eukaryotic $RbcX_2$ homologues also bound the RbcL C-terminal recognition motif (Figure 4.23), exhibiting the same general binding pattern as cyanobacterial $RbcX_2$ (Figure 4.3). This finding implies that binding to the RbcL C-terminal recognition motif may be a general property of all $RbcX_2$ homologues. Intriguingly, while the overall binding pattern between At-$RbcXI_2$ and At-$RbcXII_2$ was similar, At-$RbcXII_2$ displayed seemingly improved recognition of this motif compared to At-$RbcXI_2$ (Figure 4.23 A and B). Both bound the C-terminal recognition motif of cyanobacterial species, however At-$RbcXI_2$ did not seem to bind the C-terminal recognition motif of the majority of eukaryotic species tested while At-$RbcXII_2$ did. Furthermore, At-$RbcXII_2$ bound the C-terminal recognition motif of *A. thaliana* while there was no detectable binding to this portion by At-$RbcXI_2$ (Figure 4.23 A and B row 13). The observed differential binding could suggest functional divergence of the two *A. thaliana* $RbcX_2$'s.

A

At-RbcXI$_{2(N\text{-}FLAG)}$

B
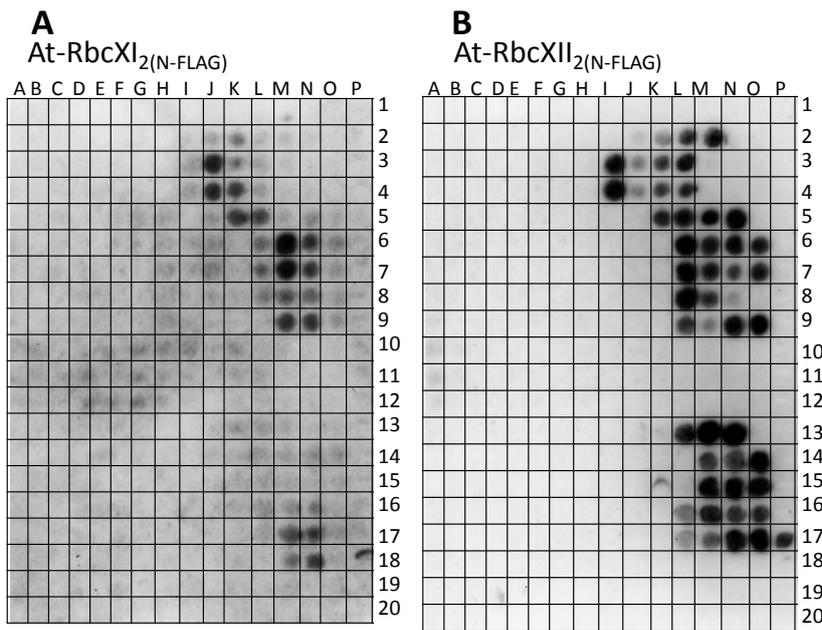
At-RbcXII$_{2(N\text{-}FLAG)}$

Figure 4.23 Binding of At-RbcXI$_{2(N\text{-}FLAG)}$ and At-RbcXII$_{2(N\text{-}FLAG)}$ to RbcL C-terminal tail.
(A-B) Purified RbcXI$_{2(N\text{-}FLAG)}$ (N-terminal Δ62) or RbcXII$_{2(N\text{-}FLAG)}$ (N-terminal Δ44) from *Arabidopsis thaliana* (At) was incubated with a membrane array of dodecapeptides, with a 10 residue overlap, covering the sequence of the last 40 amino-acids of various Rubisco large subunits as listed. Peptide-bound RbcX$_2$ was visualized by chemiluminescent immunodetection with anti-FLAG antibodies. At-RbcXI$_{2(N\text{-}FLAG)}$ (N-terminal Δ62) and At-RbcXII$_{2(N\text{-}FLAG)}$ (N-terminal Δ44) were constructed and purified by T. Hauser. (C) Consensus motif of At-RbcX$_2$ binding indicated in red.

C

Form	Species	Sequence
Form IC	1. *Rhodobacter sphaeroides*	LRAAAKWSKPLEAALDTWGNITFNYTSTDTSDFVPTASVA
Form IA	2. *Thiobacillus denitrificans*	GKTVLTEAAKNSPELKIAMETWKEIKFEFDTVDKLDVAHK
	3. *Allochromatium vinosum*	KDVLTKAAASSPELKIAMETWKEIKFEFDTVDKLDIAHHK
	4. *Nitrobacter winogradskyi*	KDILTAAAAQSPELKIAMETWKEIKFEFETMDKLAIANK
	5. *Synechococcus sp* WH 8102	KESRDILMEAGKHSPELAIALETWKEIKFEFDTVDKLDVQN
Form IB	6. *Synechococcus sp* PCC 7002	SLAREGNDVLREAGKWSPELAAALDLWKEIKFEFDTVDTL
	7. *Synechococcus elongates* PCC 6301	DLYREGGDILREAGKWSPELAAALDLWKEIKFEFETMDKL
	8. *Anabaena* PCC 7120	NLAREGNDVIREAAKWSPELAVASELWKEIKFEFEAMDTV
	9. *C. reinhardtii*	DLAREGGDVIRSASKWSPELAAASEVWKEIKFEFDTIDKL
Form ID	10. *Galdieria sulfuraria*	REAAKNSGALRTALDLWKDITFNYTSTDTSDFVETPTANI
	11. *Griffithsia pacifica*	RDAAKTSGPLQTALDLWKDITFNYTSTDTADFVETPTANV
	12. *Suhria vittata*	RDAAKTSGPLQTALDLWKDISFNYTSTDTADFVETPTANV
Form IB	13. *A. thaliana*	EGNEIIREASKWSPELAAASEVWKEITFNFPTIDKLDGQE
	14. *Nicotiana tabacum*	AQEGNEIIREASKWSPELAAASEVWKEIVFNFAAVDVLDK
	15. *O. sativa*	AREGNEIIRSASKWSPELAAASEIWKAIKFEFEPVDKLDS
	16. *Spinacia oleracea*	DLAREGNTIIREATKWSPELAAASEVWKEIKFEFPAMDTV
	17. *Rhapis excelsa*	DLAREGNEIIREASKWSPELAAASEVWKEIKFEFEPVDKLD
	18. *Zea mays*	DLAREGNEIIKAASKWSAELAAASEIWKEIKFDGFKAMDTI
Form II	19. *Rhodospirillum rubrum*	AREHKELARAFESFPGDADQIYPGWRKALGVEDTRSALPA
Form III	20. *Thermococcus kodakaraensis* KOD1	ARAVRQAIDAIMQGIPLDEYAKTHKELARALEKWGHVTPV

Since RbcX$_2$ from *A. thaliana* bound the C-terminal recognition motif of cyanobacterial RbcL, its ability to assemble Syn6301-RbcL during *in vitro* reconstitution was analyzed (AnaCA-RbcX$_2$ used as a positive control). As visualized by immunodetection of Native-PAGE, a band corresponding to RbcL$_8$/(RbcX$_2$)$_8$ could be seen when At-RbcXI$_2$ was present during refolding. Furthermore, upon subsequential addition of Syn6301-RbcS, carboxylation activity was observed, indicating holoenzyme formation (Figure 4.24 A lanes 3 and 4). Thus demonstrating that eukaryotic RbcX$_2$ can act as an assembly chaperone for Rubisco assembly similar to cyanobacterial RbcX$_2$.

The capacity of *A. thaliana* chloroplast chaperonin to facilitate reconstitution of Syn6301 Rubisco *in vitro* was also examined. As mentioned in the introduction (section

2.2.1.3.3), plant chloroplast chaperonin (Cpn60) consists of two subunits α and β. In the case of *A. thaliana* there are two homologues of α and four homologues of β present in the genome (Peltier et al. 2006). Additionally, there are two different cochaperones genes as well (At-Cpn10 and At-Cpn20). For the purposes of this study the ratio of cochaperone and chaperonin concentrations was 1:2:1 (At-Cpn10:At-Cpn20:At-Cpn60α2β3). The α and β subunits of At-Cpn60 purified by recombinant expression in *E. coli* by Dr. Yi-chin Tsai, were At-Cpn60α2 and At-Cpn60β3, which *in planta* exhibit the highest expression levels of mRNA compared to the other homologues (Peltier et al. 2006).

A

	1	2	3	4	5	6	7	8
Syn6301-RbcL 0.5 μM	+	+	+	+	+	+	+	+
GroEL/ES (0.5/1 μM)	+	+	+	+				
At-Cpn60/20/10					+	+	+	+
AnaCA-RbcX$_2$ (μM)		1.5				1.5	5	20
At-RbcXI$_2$ (μM)			2.5	5				
ATP 4 mM	+	+	+	+	+	+	+	+

B

1. Rr-RbcL/GroEL
2. Rr-RbcL/At-Cpn60
3. Syn6301-RbcL/GroEL
4. Syn6301-RbcL/At-Cpn60
5. Cr-RbcL/GroEL

Figure 4.24 *In vitro* refolding of Syn6301-RbcL with At-RbcXI$_2$ and At-Cpn60/20/10.
(A) 0.5 μM Syn6301-RbcL was refolded in the presence of GroEL/ES and At-RbcXI$_{2(N-FLAG)}$ (N-terminal Δ82) or At-Cpn60/20/10 (0.5/1/0.5 μM) and AnaCA-RbcX$_2$. Samples were analyzed by 5-13% Native-PAGE followed by immunodetection with Syn6301-RbcL antibodies. 10 μM Syn6301-RbcS was added to all samples before activity assay; furthermore 200 μM Syn6301-RbcL C-terminal peptide was added to At-Cpn60/20/10 samples prior to activity assay.
(B) 12.5% SDS-PAGE analysis of RbcL bound to chaperonin with quantification indicated below each lane.

During *in vitro* reconstitution of Syn6301-RbcL with At-Cpn60/20/10, the formation of assembly intermediates (RbcL$_2$/(RbcX$_2$)$_2$) and the RbcL$_8$/(RbcX$_2$)$_8$ complex was clearly detected on Native-PAGE when AnaCA-RbcX$_2$ was present (Figure 4.24 A lanes 6-8). The ability of At-Cpn60/20/10 to refold the form II Rubisco from *Rhodospirillum rubrum* was used as a positive control for chaperonin functionality (data not shown) (Viitanen et al. 1995). Reconstituted Syn6301-RbcL did not exhibit observable carboxylation activity upon subsequential addition of Syn6301-RbcS, even after addition of RbcL C-terminal peptide

(Figure 4.24 A lanes 6-8). It is essential to note that up to 20 µM of AnaCA-RbcX$_2$ was required to obtain detectable quantities of the RbcL$_8$/(RbcX$_2$)$_8$ complex (Figure 4.24 A lane 8), which is significantly higher than the 0.5 µM concentration of Syn6301-RbcL in the assay. It is therefore possible that at this high concentration AnaCA-RbcX$_2$ was inhibitory to proper holoenzyme assembly even though RbcL C-terminal peptide was added to compete with AnaCA-RbcX$_2$ binding to RbcL.

Further support for this argumentation arises from examination of the initial binding of RbcL to chaperonin, which was visualized by SDS-PAGE from samples taken after aggregate removal (Figure 4.24 B). Here, the inefficiency of Syn6301-RbcL binding to At-Cpn60 compared to GroEL was apparent (Figure 4.24 B lanes 3 and 4). This demonstrates that the final amount of RbcL present during Syn6301-RbcL *in vitro* reconstitution using At-Cpn60 would be substantially less than 0.5 µM, which would correspondingly increase the impact of AnaCA-RbcX$_2$ inhibition on proper holoenzyme formation due to the limited amounts of Syn6301-RbcL initially bound to At-Cpn60. Note, this inefficiency of binding to At-Cpn60 was not shared by the form II RbcL from *R. rubrum* (Figure 4.24 B lane 1 and 2) resulting in the capacity to achieve close to 100% yield of *in vitro* reconstituted form II Rubisco using either the GroEL or Cpn60 system (data not shown). Nevertheless, here it has been demonstrated by the identification of RbcL$_8$/(RbcX$_2$)$_8$ assembly intermediates by Native-PAGE that the At-Cpn60/20/10 system assists folding of Syn6301-RbcL in the *in vitro* reconstitution system upon AnaCA-RbcX$_2$ addition, albeit with significantly less efficiency than GroEL/ES.

As established above, the eukaryotic components RbcX$_2$ and Cpn60/20/10 behaved comparable to their prokaryotic counterparts during *in vitro* reconstitution of Syn6301-RbcL. Taking advantage of this, *in vitro* reconstitution of eukaryotic RbcL was attempted i.e., from *A. thaliana* (At-RbcL) and *Chlamydomonas reinhardtii* (Cr-RbcL). It has been demonstrated that expression of the chloroplast chaperonin system in *E. coli* facilitated the solubility and stability of coexpressed plant Rubisco (Cloney et al. 1993). Unfortunately, no holoenzyme was detected, which could be due to the absence of the assembly factor RbcX$_2$. Therefore, in accordance with previous attempts to express eukaryotic Rubisco *in vivo*, the eukaryotic chaperonin system (At-Cpn60/20/10) was initially utilized for eukaryotic RbcL *in vitro* refolding along with addition of various homologues of RbcX$_2$ since it remained to be determined which species of RbcX$_2$ facilitated eukaryotic Rubisco *in vitro* assembly. The RbcX$_2$ homolog would need to exhibit high enough affinity to RbcL to facilitate removal from chaperonin. Guided by the binding of cyanobacterial RbcX$_2$ to the peptide array

membrane of RbcL C-terminal peptides (Figure 4.3), different RbcX$_2$ species displaying significant levels of binding to At-RbcL or Cr-RbcL were used for *in vitro* reconstitution assays along with cognate RbcX$_2$ in the At-Cpn60/20/10 system (Figure 4.25 A). Unfortunately, there was no measurable carboxylation activity, and visualization of samples by Native-PAGE did not show the formation of assembly intermediates (Figure 4.25 A left panel). As described for Syn6301-RbcL, At-RbcL and Cr-RbcL did not bind efficiently to At-Cpn60 upon dilution into the refolding buffer from denaturant as observed from the high degree of background on Native-PAGE after immunodetection and as analyzed by SDS-PAGE (Figure 4.25 A right panel) limiting the effectiveness of this assay. This could signify an essential difference between bacterial and chloroplast chaperonin; chloroplast chaperonin may inefficiently bind certain substrates which are directly diluted from denaturant in contrast to bacterial chaperonin which efficiently binds most substrates diluted from denaturant. Future directions would include examination of the effectiveness of the chloroplast chaperonin system in an *in vitro* transcription/translation system where other chaperones are present, which could possibly facilitate efficient transfer of eukaryotic RbcL polypeptide to chaperonin.

Due to the ability of GroEL/ES to aid in the reconstitution of Syn6301-RbcL and its indiscriminate binding to a variety of substrates, this system was also utilized in the *in vitro* reconstitution assay of eukaryotic RbcL (Figure 4.25 B). As indicated by the sample without ATP addition, At-RbcL and Cr-RbcL were clearly bound to GroEL (Figure 4.25 B lanes 0). (Note, *A. thaliana* contains four homologues of RbcS, RbcS1A (N-terminal Δ55) was used for these studies based on its high level of expression *in planta* (Yoon et al. 2001), while RbcS2 (N-terminal Δ45) of *C. reinhardtii* was used for these studies since RbcS2 is the homolog that is present in the crystal structure (1IR2); furthermore the two homologues are highly identical with only four amino acid substitutions.) In the case of At-RbcL, after ATP supplementation, there was no release from chaperonin, regardless of the presence of RbcX$_2$ or At-RbcS (Figure 4.25 B lanes 2-6) including the presence of cognate RbcXI$_2$ (N-terminal Δ82, data not shown). Thus, At-RbcL clearly bound GroEL, but rebound immediately upon release. Future experiments need to be conducted to determine if At-RbcL is cycling properly on GroEL as well as verification of its encapsulation in the GroEL/ES cavity. If these criteria are met by At-RbcL with the GroEL/ES system it would point to the possibility that another factor is required either for folding or assembly to achieve holoenzyme formation of *A. thaliana* Rubisco during *in vitro* reconstitution.

A

	1	2	3	4	5	6	7	8	9	10	11	12
At-RbcL 0.5 µM	+	+	+	+	+	+						
Cr-RbcL 0.5 µM							+	+	+	+	+	
At-Cpn60/20/10	+	+	+	+	+	+	+	+	+	+	+	
At-RbcXI$_2$ (µM)	2.5	5										
Ana7120-RbcX$_2$ (µM)				2.5	5	10						
Cr-RbcX$_2$ (µM)							2.5	5				
Syn6301-RbcX$_2$ (µM)										2.5	5	10
ATP 4 mM	+	+	+	+	+	+	+	+	+	+	+	+

Cpn60/RbcL →

1. 0.5µM At-Cpn60
2. 0.5 µM At-RbcL
3. At-RbcL bound At-Cpn60
4. Cr-RbcL bound At-Cpn60

Cpn60 → RbcL → 100 51 58%

B

	0	1	2	3	4	5	6	0	7	8	9	10	11	12	13	14
At-RbcL 0.5 µM	+	+	+	+	+	+	+									
Cr-RbcL 0.5 µM								+	+	+	+	+	+	+	+	+
GroEL 0.5 µM	+		+	+	+	+	+	+		+	+	+	+	+	+	+
GroES 1 µM	+			+	+	+	+	+			+	+	+	+	+	
At-RbcS 10 µM	+				+											
Cr-RbcS 5 µM								+				+	+	+		+
Syn6301-RbcX$_2$ 5 µM	+					+										
Syn6301-RbcX$_2$ 10 µM							+									
Ana7120-RbcX$_2$ 5 µM								+					+		+	
Ana7120-RbcX$_2$ 10 µM														+		
ATP 4 mM		+	+	+	+	+	+		+	+	+	+	+	+	+	+

GroEL/RbcL →
Cr-RbcL$_8$S$_8$ →

Figure 4.25 *In vitro* refolding of *Arabidopsis thaliana* and *Chlamydomonas reinhardtii* Rubisco.
(A) Left panel: *in vitro* refolding of At-RbcL and Cr-RbcL using the At-Cpn60/20/10 system (0.5/1/0.5 µM, respectively). Concentration of RbcX$_2$ or RbcS is as indicated. Samples analyzed on 5-13% Native-PAGE followed by immunodetection with anti-Cr-RbcL antibodies. Right panel: 12.5% SDS-PAGE analysis of RbcL bound to At-Cpn60. At-Cpn60 was purified by Dr. Yi-chin Tsai. (B) *In vitro* refolding of At-RbcL or Cr-RbcL using the GroEL/ES system. RbcX$_2$ and RbcS additions are as indicated. Samples analyzed on 5-13% Native-PAGE followed by immunodetection with anti-Cr-RbcL antibodies.

The behavior of Cr-RbcL in the GroEL/ES *in vitro* reconstitution system was similar to At-RbcL in that Cr-RbcL efficiently bound to GroEL (Figure 4.25 B lanes 0 and 8) and it

was not released even if RbcX$_2$ was present (Figure 4.25 B lanes 8, 9 and 13) including during the presence of cognate RbcX$_2$ (N-terminal Δ34, data not shown). This behavior was altered when Cr-RbcS was added to the system together with ATP supplementation; Cr-RbcS led to the displacement of Cr-RbcL from GroEL irrespective of GroES addition (Figure 4.25 B lanes 10 and 14). This did not occur upon addition of At-RbcS (data not shown). It remains to be determined whether Cr-RbcS is acting as a substrate for GroEL competing with Cr-RbcL. This hypothesis is consistent with the finding that Cr-RbcL was removed from GroEL without GroES upon ATP addition, which means no encapsulation of the substrate in the cavity and therefore no proper refolding could occur.

It is nonetheless interesting that there was no detection of Cr-RbcL following immunodetection of Native-PAGE gels after *in vitro* reconstitution when Cr-RbcS was present (Figure 4.25 B lane 10 and 14); importantly, Cr-RbcL was not degraded as determined by immunodetection of SDS-PAGE gels of the samples (data not shown). Therefore, Cr-RbcS was not contaminated with or behaving as a protease since full length Cr-RbcL was still detected in the assay by SDS-PAGE; furthermore the removal from GroEL was ATP-dependent, which could indicate substrate competition. When both RbcX$_2$ and Cr-RbcS were added to the reconstitution assay, there was release from GroEL as well as an increase in the appearance of higher aggregates and a smear down the length of the gel (Figure 4.25 B lanes 11 and 12). However, any interpretation of what is taking place is purely speculative.

In order to resolve how Cr-RbcS is behaving in the *in vitro* reconstitution assay, it needs to be established if Cr-RbcS was actually behaving as a GroEL substrate; if not it would not only indicate the necessity of Cr-RbcS in reconstitution assays of Cr-RbcL for efficient removal from chaperonin but also that an essential yet unknown component is still missing. Additionally, the encapsulation of Cr-RbcL in the GroEL/ES cavity must also be ascertained as well as the verification of proper cycling of Cr-RbcL on GroEL. Further vital experiments which need to be established would be the expression of Cr-RbcL or At-RbcL in the *in vitro* transcription/translation system with either chloroplast chaperonin or bacterial chaperonin supplementation.

5   Discussion

The mechanism by which Rubisco folds and assembles has been under examination for the past several decades. This protein has been in the spotlight of photosynthesis research not only due to its exceedingly high abundance in nature, but also due to its key role in CO_2 fixation, which has implications in attempts to improve crop yield and photosynthetic efficiency. Rubisco catalyzes incorporation of CO_2 in the sugar-phosphate substrate ribulose-1,5-bisphosphate (RuBP) during the Calvin-Benson-Bassham cycle; however this enzyme is inefficient in that it has an extremely low substrate turnover rate and furthermore does not discriminate well between CO_2 and O_2. Oxygenation of RuBP leads to a decrease in photosynthetic efficiency, thus many efforts towards increasing crop plant yield are focused on enhancing Rubisco carboxylase activity through protein engineering. However a critical obstacle has hampered efforts to engineer a 'better' Rubisco: the inability of higher plant Rubisco to fold and assemble outside of the chloroplast. While there has been much advancement in the understanding of the folding and assembly of form I Rubisco, a detailed description of the mechanism is lacking. Therefore, the objective of the work presented here was to elucidate the detailed structural mechanism of cyanobacterial form I Rubisco folding and assembly.

5.1   Role of chaperonins in Rubisco folding

It is well established that the formation of bacterial Rubisco holoenzyme requires the action of chaperonins. Form II Rubisco can be successfully folded and assembled recombinantly in *E. coli* (Somerville and Somerville 1984), in tobacco chloroplasts (Whitney and Andrews 2001), or it can be properly refolded and assembled during *in vitro* reconstitution in the presence of the chaperonin system (Goloubinoff et al. 1989a). This is the simplest form of Rubisco, being a dimer of large subunits, and the chaperonin system alone can satisfy its folding and assembly requirements. Prokaryotic form I Rubisco can also be successfully folded and assembled recombinantly in *E. coli* (Goloubinoff et al. 1989b). Furthermore, as revealed in section 4.2, the GroEL/ES system alone was capable of facilitating Syn6301 Rubisco holoenzyme formation in an *in vitro* translation system (Figure 4.2 B). However, recently it has been demonstrated that prokaryotic form I Rubisco from Syn6301 could be reconstituted from denatured components *in vitro* (Liu et al. 2010). In this system not only was the GroEL/ES system essential, but also the assembly chaperone $RbcX_2$ was required. In the presence of the GroEL/ES system alone or with cognate $RbcX_2$, assembly competent intermediates were not formed during *in vitro* reconstitution; form I RbcL

remained bound to GroEL and the holoenzyme would not form. During *in vitro* reconstitution of Syn6301 form I Rubisco, addition of high affinity $RbcX_2$ was necessary to push the equilibrium towards assembly competent intermediates, whereas in the *in vitro* translation system newly synthesized RbcL was capable of competing with folded RbcL for the GroEL/ES, enforcing its release. Nevertheless, even though the GroEL/ES system alone was able to facilitate formation of Syn6301 Rubisco holoenzyme, holoenzyme formation was enhanced when the assembly chaperone $RbcX_2$ was also present (Figure 4.2 A).

Intriguingly, in contrast to form II Rubisco, prokaryotic form I RbcL was not expressed upon tobacco plastid transformation (Kanevski et al. 1999). The reason for this lack of expression is undetermined and it is unknown if the foreign RbcL was immediately degraded upon expression, as no insoluble RbcL was detected. It is possible that the failure to express was due to a mismatched chaperone system. However, in section 4.4 (Figure 4.24 A) it was shown that the eukaryotic chaperonin system from *Arabidopsis thaliana* could refold and assemble Syn6301 Rubisco in cooperation with $RbcX_2$ *in vitro*, albeit less efficiently than the GroEL/ES system. Furthermore, Syn6301 Rubisco could be assembled *in vitro* with *A. thaliana* $RbcX_2$ (Figure 4.24 A). Thus, either the tobacco chloroplast Cpn60 or $RbcX_2$ is not compatible with cyanobacterial Rubisco or the environment of the tobacco chloroplast is not conducive to cyanobacterial Rubisco folding and/or assembly resulting in protein degradation. Conversely, although attempts to express eukaryotic form I Rubisco in *E. coli* have so far been unsuccessful (Cloney et al. 1993; Gatenby 1984; Gatenby et al. 1987), the large subunit of Rubisco from sunflower can fold and assemble in tobacco chloroplasts (Kanevski et al. 1999). This indicates that its chaperone needs are met in this foreign environment and it is highly likely that higher plant Rubisco from various species follow similar folding and assembly pathways.

Higher plant Rubisco has been found in complex with chloroplast Cpn60 (Barraclough and Ellis 1980), indicating that the chaperonin is also essential for folding of eukaryotic Rubisco. It is also highly probable that $RbcX_2$ is an additional factor involved in eukaryotic Rubisco assembly since both isoforms of $RbcX_2$ from *A. thaliana* ($At-RbcX_2$) interact with the C-terminal of various species of Rubisco, including eukaryotic species (Figure 4.23) and $At-RbcX_2$ was able to effectively assemble Syn6301 Rubisco (Figure 4.24 A). This behavior is comparable to that of cyanobacterial $RbcX_2$ (Figure 4.3), which is a known assembly factor of cyanobacterial Rubisco (Liu et al. 2010; Saschenbrecker et al. 2007). These results strongly suggest that eukaryotic $RbcX_2$ is also involved in higher plant Rubisco assembly. However, as shown in section 4.4, while eukaryotic Rubisco large subunits could bind to chaperonin

(Figure 4.25), the addition of various species of $RbcX_2$ or cognate RbcS, did not result in the formation of assembly competent intermediates or Rubisco holoenzyme. Considering these results, it seems possible that one or several other factors besides $RbcX_2$ are also involved in eukaryotic Rubisco folding and/or assembly.

5.2   Highly homologous Rubisco proteins have diverse assembly pathways

A perplexing aspect of Rubisco is that this protein from various species displays up to 90% identity in the primary sequence of the large subunit but exhibits remarkable differences in enzyme kinetics. For example prokaryotic Rubisco has low CO_2/O_2 specificity values but high substrate turnover rates, whereas higher plant Rubisco has high CO_2/O_2 specificity values but low substrate turnover rates (Andersson and Backlund 2008). The small subunits of the prokaryotic form I Rubisco can be removed from the $RbcL_8$ core without causing large subunit aggregation, but this characteristic is not shared with eukaryotic Rubisco (Andrews and Lorimer 1985). Removal of RbcS from spinach $RbcL_8$ cores causes aggregation of the large subunit, indicating that the RbcS is required for stability of the $RbcL_8$ core. It is also quite possible that the pathways of assembly of non-related species of Rubisco are distinct. One line of evidence for this is the presence or absence of *rbcX* in the genome. Form IA Rubisco (in proteobacteria and marine cyanobacteria) and form IB Rubisco (in fresh water cyanobacteria, green algae, and plants) are similar in that they share the $RbcX_2$ C-terminal recognition motif; however, to date no *rbcX* homolog has been identified in the genome of species harboring form IA Rubisco. Furthermore, form IC Rubisco (in proteobacteria) and form ID Rubisco (in proteobacteria and red algae) do not possess the $RbcX_2$ C-terminal recognition motif and the genomes of these species lack *rbcX* (Tabita 2007). The absence of an *rbcX* homolog could either indicate that Rubisco folding and assembly is $RbcX_2$-independent, or that another factor(s) is necessary for folding and/or assembly of species that do not contain an *rbcX* homolog in their genome.

5.3   Assembly chaperones

The assembly of some multi-domain/oligomeric protein complexes requires the assistance of particular assembly chaperones. The necessity of assembly chaperones for support in the formation of the nucleosome (Laskey et al. 1978) and the proteasome (Witt et al. 2000) has been well established. The term molecular chaperone was initially coined to define the set of proteins involved in assisting the assembly of oligomeric complexes, and the term was later broadened to include proteins involved in the folding of non-native proteins

(Ellis 2006). As described in section 2.2, there are many examples of chaperones essential for the folding of non-native proteins but there are few examples of assembly chaperones apart from their role in nucleosome and proteasome formation and in the assembly of form I Rubisco.

Recently, the discovery of a new ATP-independent chaperone, Spy, located in the periplasm of *E. coli* was reported (Quan et al. 2011). The authors have demonstrated that this cradle shaped protein assists with folding of unstable proteins and suppresses protein aggregation. There is a striking structural resemblance between the Spy protein and $RbcX_2$ (Figure 5.1). The Spy protein is a homodimer of subunits with four α-helices, aligned in an anti-parallel fashion along the α3 helix, and the dimer is about 30 kDa in size. It has a concave 'groove-like' region that has hydrophobic patches. The Spy protein was shown to be induced by protein conformational state, and could also act as a chaperone for a variety of substrates *in vitro*. However, the exact mechanism of its function remains to be determined especially since this protein acts ATP-independently and it is unknown how it is released from substrate after binding. $RbcX_2$ also acts in an ATP-independent manner; however it is known to be released from its substrate by the binding of the small subunit to the $RbcL_8/(RbcX_2)_8$ complex in a mechanism involving the structuring of the RbcL '60ies loop' (section 4.3.2.3 and 5.2.1). $RbcX_2$ also operates as an aggregation suppressor, and the folding intermediates of Rubisco are highly aggregation prone (Gutteridge and Gatenby 1995). However, there has been no known report of $RbcX_2$ induction by unfolded protein or stress conditions; on the contrary, $RbcX_2$ is of low abundance and difficult to detect in cellular lysate (Emlyn-Jones et al. 2006). Nevertheless, a recent proteomics study was able to detect trace amounts peptides of both $RbcX_2$ isoforms in the stroma of *A. thaliana* (Olinares et al. 2010), indicating that $RbcX_2$ protein is expressed in plants and transported to the chloroplast, albeit at low levels.

Since $RbcX_2$ is not found at very high concentrations, its function may be more important during plant development when high amounts of Rubisco are being synthesized. Furthermore, Rubisco is a highly stable, long-lived protein in its holoenzyme form, making it unnecessary to constantly have $RbcX_2$ protein present once Rubisco is made. Nevertheless, the newly described periplasmic *E. coli* chaperone, Spy, could provide clues as to other possible functions of eukaryotic $RbcX_2$. An important step to understanding the function of eukaryotic $RbcX_2$ would be the analysis of the crystal structure. Unfortunately, this protein is highly aggregation prone above concentrations of 1 mg/ml (section 3.4.3.6), making crystallization difficult.
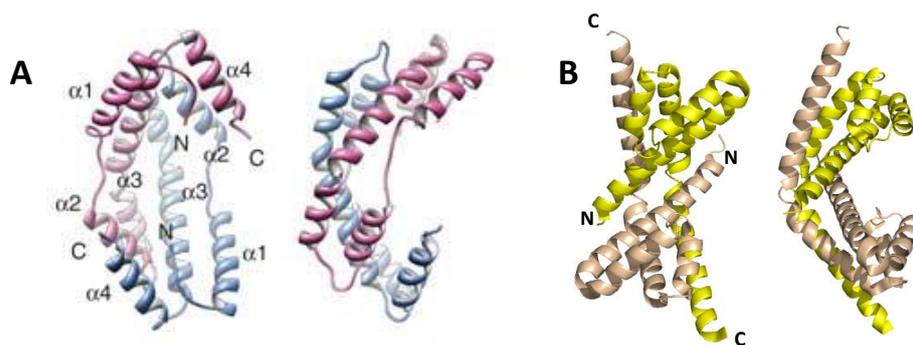
Figure 5.1 Comparison of the Spy protein crystal structure with RbcX$_2$ crystal structure.
(A) Front and side ribbon representation of the Spy protein crystal structure with subunits indicated in light blue and magenta. The α-helices 1-4 are labeled along with the N- and C-termini. Adapted from: (Quan et al. 2011). (B) Front and side ribbon representation of Syn7002-RbcX$_2$ crystal structure with subunits indicated in light brown and yellow. The N- and C-termini are labeled.

5.3.1 Mechanism of RbcX$_2$-mediated Rubisco assembly

In this study the detailed structural mechanism of RbcX$_2$-mediated cyanobacterial form I Rubisco assembly has been elucidated (section 4.3.2.3). The chaperone-assisted folding and assembly of form I Rubisco are tightly coupled; RbcX$_2$ binds RbcL immediately after encapsulation and release from GroEL/ES to ensure the formation of assembly-competent intermediates. RbcX$_2$ recognizes and binds the C-terminal of RbcL (area I) along with area II (Figure 4.15). Binding of RbcX$_2$ to RbcL prevents rebinding to chaperonin and supports RbcL dimerization, possibly by ordering of the C-terminal helices of RbcL (Figure 4.8). Furthermore, RbcX$_2$ promotes RbcL dimerization by salt bridge formation to area III of the adjacent RbcL subunit.

RbcL dimerizes in an anti-parallel fashion by alignment of three charge pairs. Two charge pairs are located at the RbcL/RbcL interface, while the additional charge pair is contributed by the bound RbcX$_2$ (Figure 4.16 and Figure 4.17). These guide-points are necessary to prevent misalignment and off-pathway aggregation by formation of the intermediate RbcL$_2$/(RbcX$_2$)$_2$. This intermediate is less prone to aggregation due to the burial of the hydrophobic surfaces located at the RbcL/RbcL interface (Figure 4.19 and Figure 4.20). The intermediate can then assemble to RbcL$_8$/(RbcX$_2$)$_8$ by surface shape complementarity and polar contacts at the RbcL$_2$/RbcL$_2$ interface. RbcX$_2$ bound to the RbcL$_8$ core promotes formation of the RbcS binding interface by ordering the C-terminal helices of RbcL (Figure 4.8). RbcS then binds in between RbcL dimers of the RbcL$_8$ core, which leads to structuring of the RbcL '60ies loop'. The structuring of the '60ies loop' in turn blocks RbcX$_2$ binding to area III of RbcL and mediates RbcX$_2$ displacement. When the '60ies loop' is structured, RbcL is rendered sterically incompatible with RbcX$_2$ binding. Figure 5.2 provides a structural description of the mechanism of RbcX$_2$-medaited Rubisco assembly.
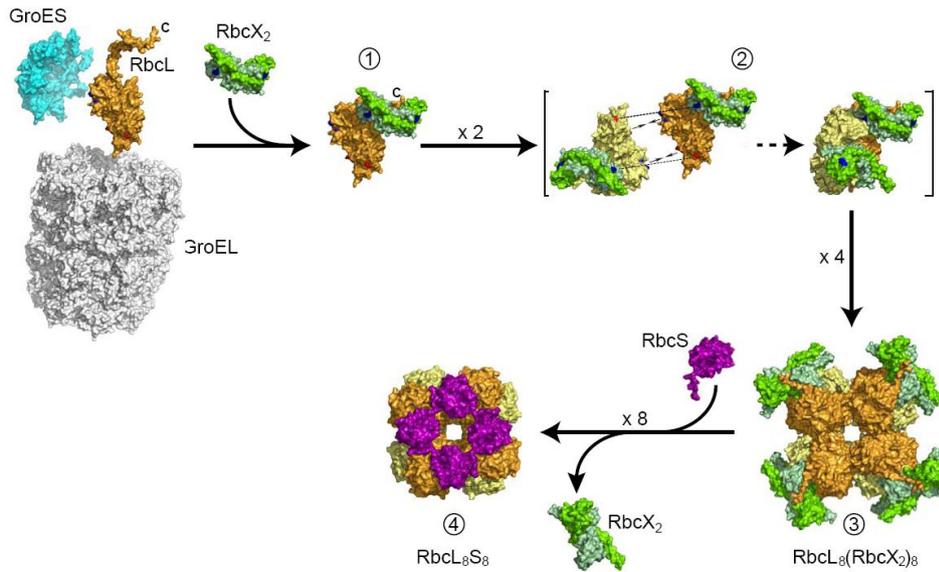
Figure 5.2 Model of RbcX$_2$-assisted Rubisco assembly.
RbcL is encapsulated and folded by GroEL/ES. (1) RbcX$_2$ binds the C-terminal recognition motif of RbcL (area I) and area II upon release from chaperonin. (2) RbcL dimerization. The indicated complementary surface charge guide-points from RbcL and RbcX$_2$ most likely play a role in ensuring proper alignment of the anti-parallel dimer. RbcX$_2$ forms a salt bridge with area III of the adjacent RbcL subunit and stabilizes the dimers by acting as a 'molecular staple'. (3) The intermediates, RbcL$_2$/(RbcX$_2$)$_2$, assemble to the RbcL$_8$/(RbcX$_2$)$_8$ complex, where RbcX$_2$ also likely supports formation of the RbcS binding interface on RbcL. (4) Binding of RbcS causes the structuring of the RbcL '60ies loop' causing steric blockage of RbcX$_2$ binding to area III on RbcL. This leads to the displacement of RbcX$_2$ and formation of the functional Rubisco holoenzyme. Figure prepared by Dr. Andreas Bracher.

5.4   Role of RbcL N-terminal region in Rubisco catalysis

Cleavage of the N-terminus of RbcL results in Rubisco enzymatic inactivity (Houtz et al. 1989). It has previously been noted that the N-terminus of the Rubisco large subunit is most likely involved in formation of the active site and plays a role in maintaining active site conformation for RuBP binding (Kettleborough et al. 1987). Binding of RbcS to the RbcL$_8$ core facilitates structuring of the N-terminus, aiding in proper active site conformation (Figure 4.21). The importance of the N-terminal region in Rubisco activity was also demonstrated in Figure 4.22. Here it was shown that the first 11 amino acids were necessary for full catalytic activity while deletion of the first five amino acids was not detrimental to catalysis. Addition of substrate or inhibitor to the Rubisco active site has been shown to prevent protease sensitivity of the N-terminal portion of RbcL due to N-terminal interaction with substrate, resulting in increased structural integrity (Khan et al. 1999). Binding of inhibitors to the Rubisco active site at night would aid in reducing N-terminal cleavage by proteases and enhance the amount of active enzyme available during the day. The N-terminal segment of the Rubisco large subunit, including the '60ies loop', remained vulnerable to protease digestion when no substrate or inhibitor was bound as demonstrated by limited-proteolysis of RbcL$_8$ cores (Figure 4.9). Interestingly, the RbcL '60ies loop' even remained partially protease sensitive with RbcS bound. This result underlines the necessity of bound

substrate in the Rubisco active site to impede proteolytic digestion of the N-terminal region of Rubisco to maintain catalytic activity. In the present study, the importance of the N-terminus for full catalytic activity has been further demonstrated (Figure 4.22).

5.5 Role of RbcS in Rubisco catalysis

RbcS must be bound for optimal catalytic function of form I Rubisco. However, the structural basis for this had been unclear since RbcS is distant from the active site of Rubisco located at the RbcL/RbcL interface. In this study, a possible role of RbcS in Rubisco catalytic activity is revealed. Structuring of the RbcL '60ies loop' by RbcS binding not only mediates displacement of $RbcX_2$ but also generates the catalytically active conformation of Rubisco (Figure 4.21 and Figure 4.22). The catalytic inactivity of the $RbcL_8/(RbcX_2)_8$ complex is partially due to the unstructured RbcL '60ies loop'. As revealed in the $RbcL_8S_8$ holoenzyme crystal structure with CABP bound in the closed state, the amino acids T62 and W63 of the RbcL '60ies loop' actually contact the bound substrate directly (Figure Figure 4.21). This loop is generally structured in form I holoenzyme crystal structures, thereby suggesting that RbcS plays a role in its stability. In support of this, Figure 4.22 demonstrated that the mutant RbcL(W67A,L70G,L71A), which interrupts the interface interaction of RbcL with RbcS, is catalytically inactive when RbcS is bound, most likely due to the inability of the '60ies loop' to become structured in this mutant. Further support came from the analysis of the mutant RbcS(L60A,F63A), which affects the RbcL/RbcS interface. Addition of this mutant RbcS to $RbcL_8$ did not result in a catalytically competent enzyme, presumable due to the inability of the mutant RbcS to structure the RbcL '60ies loop' (Figure 4.22). Mutations of RbcS in this region in *Chlamydomonas reinhardtii* also resulted in compromised Rubisco function (Flachmann and Bohnert 1992; Flachmann et al. 1997; Genkov and Spreitzer 2009). These results indicate that RbcS binding is necessary to stabilize the '60ies loop' in a catalytically competent conformation, which provides highly convincing structural evidence for the role of RbcS in Rubisco catalysis.

5.6 Outlook

Having resolved the mechanism of cyanobacterial Rubisco assembly, the next important goal is to understand the folding and assembly pathway of plant Rubisco. Given the importance of $RbcX_2$ during cyanobacterial form I Rubisco assembly, it is very likely that this assembly chaperone is also relevant for the assembly of eukaryotic form I Rubisco. This hypothesis is supported by the identification of *rbcX* homologues in the genome of eukaryotic

species encoding form IB Rubisco, and from the present study showing that *Arabidopsis thaliana* At-RbcXI$_2$ and At-RbcXII$_2$ both bind to the C-terminal recognition motif of RbcL from a variety of species, similar to the behavior of cyanobacterial RbcX$_2$ (Figure 4.23 and Figure 4.3, respectively). Furthermore, At-RbcXI$_2$ could support assembly of Syn6301-RbcL in the *in vitro* reconstitution assay (Figure 4.24), thereby providing direct evidence that a eukaryotic RbcX$_2$ could assist assembly of form I Rubisco. The RbcL$_8$ core from plants is highly unstable without RbcS bound (Andrews and Lorimer 1985); therefore a component like RbcX$_2$ could be important for the stabilization of chloroplast synthesized RbcL before nuclear synthesized RbcS is added. Future efforts to engineer a better Rubisco will need to take into account the highly complex mechanism of folding and assembly. The detailed structural mechanism explained in this study should aid further efforts in establishing a system for eukaryotic Rubisco folding and assembly.

6   References

Agashe VR, Guha S, Chang HC, Genevaux P, Hayer-Hartl M, Stemp M, Georgopoulos C, Hartl FU and Barral JM (2004) Function of trigger factor and DnaK in multidomain protein folding: Increase in yield at the expense of folding speed. Cell 117: 199-209

Altekar W and Rajagopalan R (1990) Ribulose bisphosphate carboxylase activity in halphilic archaebacterai. Archives of Microbiology 153: 169-174

Andersson I (2008) Catalysis and regulation in Rubisco. Journal of Experimental Botany 59: 1555-1568

Andersson I and Backlund A (2008) Structure and function of Rubisco. Plant Physiology and Biochemistry 46: 275-291

Andrews TJ (1988) Catalysis by cyanobacterial ribulose-bisphosphate carboxylase large subunits in the complete absence of small subunits Journal of Biological Chemistry 263: 12213-12219

Andrews TJ and Lorimer GH (1985) Catalytic properties of a hybrid between cyanobacterial large subunits and higher-plant small subunits of ribulose bisphosphate carboxylase-oxygenase. Journal of Biological Chemistry 260: 4632-4636

Andrews TJ and Whitney SM (2003) Manipulating ribulose bisphosphate carboxylase/oxygenase in the chloroplasts of higher plants. Archives of Biochemistry and Biophysics 414: 159-169

Anfinsen CB (1973) Principles that govern the folding of protein chains. Science 181: 223-230

Apetri AC and Horwich AL (2008) Chaperonin chamber accelerates protein folding through passive action of preventing aggregation. Proceedings of the National Academy of Sciences of the United States of America 105: 17351-17355

Asadulghani, Suzuki Y and Nakamoto H (2003) Light plays a key role in the modulation of heat shock response in the cyanobacterium Synechocystis sp. PCC 6803. Biochemical and Biophysical Research Communications 306: 872-879

Badger MR and Bek EJ (2007) Multiple Rubisco forms in proteobacteria: their functional significance in relation to CO_2 acquisition by the CBB cycle. Journal of Experimental Botany 59: 1525-1541

Badger MR, Hanson D and Price GD (2002) Evolution and diversity of CO_2 concentrating mechanisms in cyanobacteria. Functional Plant Biology 29: 161-173

Badger MR, Kaplan A and Berry JA (1980) Internal inorganic carbon pool of Chlamydomonas reinhardtii: evidence for a carbon dioxide concentrating mechanism. Plant Physiology 66: 407-413

Bainbridge G, Anralojc PJ, Madgwick PJ, Pitts JE and Parry MAJ (1998) Effect of mutation of lysine-128 of the large subunit of ribulose bisphosphate carboxylase/oxygenase from Anacystis nidulans. Biochemical Journal 336: 387-393

Baneyx F, Bertsch U, Kalbach CE, Vandervies SM, Soll J and Gatenby AA (1995) Spinach chloroplast Cpn21 co-chaperonin possesses two functional domains fused together in a toroidal structure and exhibits nucleotide-dependent binding to plastid chaperonin 60. Journal of Biological Chemistry 270: 10695-10702

Barraclough R and Ellis RJ (1980) Protein-synthesis is chloroplasts .9. assembly of newly-synthesized large subunits into ribulose bisphosphate carboxylase in isolated intact pea-chloroplasts. Biochimica Et Biophysica Acta 608: 19-31

Bauwe H (2011) Photorespiration: The Bridge to C-4 Photosynthesis. In: Raghavendra AS and Sage RF (eds) C4 Photosynthesis and Related CO2 Concentrating Mechanisms, pp 81-108, Springer, Dordrecht

Berg JM, Tymoczko JL and Stryer L (2002) Biochemistry, 5th edition. W.H. Freemand and Co., New York, 974 pp

Bertsch U and Soll J (1995) Functional analysis of isolated cpn10 domains and conserved amino acid residues in spinach chloroplast co-chaperonin by site-directed mutagenesis. Plant Molecular Biology 29: 1039-1055

Bertsch U, Soll J, Seetharam R and Viitanen PV (1992) Identification, characterization, and DNA sequence of a functional "double" GroES-like chaperonin from chloroplasts of higher plants. Proceedings of the National Academy of Sciences of the United States of America 89: 8696-8700

Blair GE and Ellis RJ (1973) Protein-synthesis in chloroplasts .1. light-driven synthesis of large subunit of fraction I protein by isolated pea chloroplasts. Biochimica Et Biophysica Acta 319: 223-234

Bloom AJ, Burger M, Rubio-Asensio JS and Cousins AB (2010) Carbon dioxide enrichment inhibits nitrate assimilation in wheat and Arabidopsis. Science 328: 899-903

Bloom MV, Milos P and Roy H (1983) Light-dependent assembly of ribulose-1,5-bisphosphate carboxylase. Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences 80: 1013-1017

Bonshtien AL, Weiss C, Vitlin A, Niv A, Lorimer GH and Azem A (2007) Significance of the N-terminal domain for the function of chloroplast cpn20 chaperonin. Journal of Biological Chemistry 282: 4463-4469

Boston RS, Viitanen PV and Vierling E (1996) Molecular chaperones and protein folding in plants. Plant Molecular Biology 32: 191-222

Bracher A, Starling-Windhof A, Hartl FU and Hayer-Hartl M (2011) Crystal structure of a chaperone-bound assembly intermediate of form I Rubisco. Nature Structural & Molecular Biology 18: 875-880

Bradford MM (1976) A rapid and sensitive method for the quantification of microgram quantities of protein utilizing the principle of protein-dye binding. Analytical Biochemistry 72: 248-254

Bradley D, Vandervies SM and Gatenby AA (1986) Expression of cyanobacterial and higher-plant ribulose-1,5-bisphosphate carboxylase genes in Escherichia coli. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences 313: 447-458

Braig K, Otwinowski Z, Hegde R, Boisvert DC, Joachimiak A, Horwich AL and Sigler PB (1994) The crystal-structure of the bacterial chaperonin GroEL at 2.8 Angstrom. Nature 371: 578-586

Brehmer D, Gassler C, Rist W, Mayer MP and Bukau B (2004) Influence of GrpE on DnaK-substrate interactions. Journal of Biological Chemistry 279: 27957-27964

Brinker A, Pfeifer G, Kerner MJ, Naylor DJ, Hartl FU and Hayer-Hartl M (2001) Dual function of protein confinement in chaperonin-assisted protein folding. Cell 107: 223-233

Bukau B and Horwich AL (1998) The Hsp70 and Hsp60 chaperone machines. Cell 92: 351-366

Chakraborty K, Chatila M, Sinha J, Shi QY, Poschner BC, Sikor M, Jiang GX, Lamb DC, Hartl FU and Hayer-Hartl M (2010) Chaperonin-catalyzed rescue of kinetically trapped states in protein folding. Cell 142: 112-122

Chaudhuri TK, Farr GW, Fenton WA, Rospert S and Horwich AL (2001) GroEL/GroES-mediated folding of a protein too large to be encapsulated. Cell 107: 235-246

Checa SK and Viale AM (1997) The 70-kDa heat-shock protein DnaK chaperone system is required for the productive folding of ribulose-bisphosphate carboxylase subunits in Escherichia coli. European Journal of Biochemistry 248: 848-855

Chen S, Roseman AM, Hunter AS, Wood SP, Burston SG, Ranson NA, Clarke AR and Saibil HR (1994) Location of a folding protein and shape changes in GroEL-GroES complexes imaged by cryoelectron microscopy. Nature 371: 261-264

Chifflet S, Torriglia A, Chiesa R and Tolosa S (1988) A method for the determination of inorganic phosphate in the presence of labile organic phosphate and high concentrations of protein: application to lens ATPases. Analytical Biochemistry 168: 1-4

Chua NH and Schmidt GW (1978) Post-translational transport into intact chloroplasts of a precursor to the small subunit of ribulose-1,5-bisphosphate carboxylase. Proceedings of the National Academy of Sciences of the United States of America 75: 6110-6114

Chung CT, Niemela SL and Miller RH (1989) One-step preparation of competent Escherichia coli: transformation and storage of bacterial cells in the same solution. Proceedings of the National Academy of Sciences of the United States of America 86: 2172-2175

Cleland WW, Andrews TJ, Gutteridge S, Hartman FC and Lorimer GH (1998) Mechanism of Rubisco: The carbamate as general base. Chemical Reviews 98: 549-561

Cloney LP, Bekkaoui DR and Hemmingsen SM (1993) Coexpression of plastid chaperonin genes and a synthetic plant rubisco operon in Escherichia coli. Plant Molecular Biology 23: 1285-1290

Coligan JE, Dunn BM, Speicher DW and Wingfield PT (2011) Current Protocols in Protein Science. John Wiley & Sons, Inc., New York

Corpet F (1988) Multiple sequence alignment with hierarchical-clustering. Nucleic Acids Research 16: 10881-10890

Dickson R, Weiss C, Howard RJ, Alldrick SP, Ellis RJ, Lorimer G, Azem A and Viitanen PV (2000) Reconstitution of higher plant chloroplast chaperonin 60 tetradecamers active in protein folding. Journal of Biological Chemistry 275: 11829-11835

Dobberstein B, Blobel G and Chua NH (1977) In vitro synthesis and processing of a putative precursor for small subunit of ribulose-1,5-bisphosphate carboxylase of Chlamydomonas reinhardtii. Proceedings of the National Academy of Sciences of the United States of America 74: 1082-1085

Dobson CM, Sali A and Karplus M (1998) Protein folding: A perspective from theory and experiment. Angewandte Chemie-International Edition 37: 868-893

Duff AP, Andrews TJ and Curmi PMG (2000) The transition between the open and closed states of rubisco is triggered by the inter-phosphate distance of the bound bisphosphate. Journal of Molecular Biology 298: 903-916

Ehleringer JR and Cerling TE (2002) C3 and C4 photosynthesis. In: Munn T (ed) Encyclopedia of Global Environmental Change, pp 186-190, John Wiley & Sons, Ltd., Chichester

Ehleringer JR, Sage RF, Flanagan LB and Pearcy RW (1991) Climate change and the evolution of C4 photosynthesis. Trends in Ecology and Evolution 6: 95-99

Ellis J (1987) Proteins as molecular chaperones. Nature 328: 378-379

Ellis RJ (1979) Most abundant protein in the world. Trends in Biochemical Sciences 4: 241-244

Ellis RJ (1994) Molecular chaperones: opening and closing the Anfinsen cage. Current Biology 4: 633-635

Ellis RJ (2006) Molecular chaperones: assisting assembly in addition to folding. Trends in Biochemical Sciences 31: 395-401

Emlyn-Jones D, Woodger FJ, Price GD and Whitney SM (2006) RbcX can function as a Rubisco chaperonin, but is non-essential in Synechococcus PCC7942. Plant and Cell Physiology 47: 1630-1640

Esquivel MG, Anwaruzzaman M and Spreitzer RJ (2002) Deletion of nine carboxy-terminal residues of the Rubisco small subunit decreases thermal stability but does not eliminate function. Febs Letters 520: 73-76

Ewalt KL, Hendrick JP, Houry WA and Hartl FU (1997) In vivo observation of polypeptide flux through the bacterial chaperonin system. Cell 90: 491-500

Ferbitz L, Maier T, Patzelt H, Bukau B, Deuerling E and Ban N (2004) Trigger factor in complex with the ribosome forms a molecular cradle for nascent proteins. Nature 431: 590-596

Flachmann R and Bohnert HJ (1992) Replacement of a conserved arginine in the assembly domain of ribulose-1,5-bisphospate carboxylae/oxygenase small subunit interferes with holoenzyme formation. Journal of Biological Chemistry 267: 10576-10582

Flachmann R, Zhu GH, Jensen RG and Bohnert HJ (1997) Mutations in the small subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase increase the formation of the misfire product xylulose-1,5-bisphosphate. Plant Physiology 114: 131-136

Foyer CH, Bloom AJ, Queval G and Noctor G (2009) Photorespiratory Metabolism: Genes, Mutants, Energetics, and Redox Signaling. Annual Review of Plant Biology 60: 455-484

Frydman J (2001) Folding of newly translated proteins in vivo: The role of molecular chaperones. Annual Review of Biochemistry 70: 603-647

Furuki M, Tanaka N, Hiyama T and Nakamoto H (1996) Cloning, characterization and functional analysis of groEL-like gene from thermophilic cyanobacterium Synechococcus vulcanus, which does not form an operon with groES. Biochimica Et Biophysica Acta-Protein Structure and Molecular Enzymology 1294: 106-110

Garcia-Murria MJ, Karkehabadi S, Marin-Navarro J, Satagopan S, Andersson I, Spreitzer RJ and Moreno J (2008) Structural and functional consequences of the replacement of proximal residues Cys(172) and Cys(192) in the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase from Chlamydomonas reinhardtii. Biochemical Journal 411: 241-247

Gatenby AA (1984) The properties of the large subunit of maize Ribulose bisphosphate carboxlyase oxygenase synthesized in Escherichia coli. European Journal of Biochemistry 144: 361-366

Gatenby AA (1992) Protein folding and chaperonins. Plant Molecular Biology 19: 677-687

Gatenby AA and Ellis RJ (1990) Chaperone function - the assembly of Ribulose bisphosphate carboxylase/oxygenase. Annual Review of Cell Biology 6: 125-149

Gatenby AA, Vandervies SM and Bradley D (1985) Assembly in Escherichia coli of a functional multi-subunit Ribulose bisphosphate carboxylase from a blue-green-alga. Nature 314: 617-620

Gatenby AA, Vandervies SM and Rothstein SJ (1987) Coexpression of both the maize large and wheat small subunit genes of Ribulose-bisphosphate carboxylase in Escherichia coli. European Journal of Biochemistry 168: 227-231

Genevaux P, Georgopoulos C and Kelley w (2007) The Hsp70 chaperone machines of Escherichia coli: a paradigm for the repartition of chaperone functions. Molecular Microbiology 66: 840-857

Genkov T and Spreitzer RJ (2009) Highly Conserved Small Subunit Residues Influence Rubisco Large Subunit Catalysis. Journal of Biological Chemistry 284: 30105-30112

Goloubinoff P, Christeller JT, Gatenby AA and Lorimer GH (1989a) Reconstitution of active dimeric ribulose bisphosphate carboxylase from and unfolded state depends on 2 chaperonin proteins and Mg-ATP. Nature 342: 884-889

Goloubinoff P, Gatenby AA and Lorimer GH (1989b) GroE heat-shock proteins promote assembly of foreign prokaryotic ribulose bisphosphate carboxylase oligomers in Escherichia coli. Nature 337: 44-47

Gould SB, Waller RR and McFadden GI (2008) Plastid evolution. Annual Review of Plant Biology 59: 491-517

Graciet E, Lebreton S and Gontero B (2004) Emergence of new regulatory mechanisms in the Benson-Calvin pathway via protein-protein interactions: a glyceraldehyde-3-phosphate dehydrogenase/CP12/phosphoribulokinase complex. Journal of Experimental Botany 55: 1245-1254

Gutteridge S and Gatenby AA (1995) Rubisoc synthesis, assembly, mechanism, and regulation. Plant Cell 7: 809-819

Gutteridge S, Millard BN and Parry MAJ (1986) Inactivation of ribulose-bisphosphate carboxylase by limited proteolysis. Febs Letters 196: 263-268

Haass C and Selkoe DJ (2007) Soluble protein oligomers in neurodegeneration: lessons from the Alzheimer's amyloid beta-peptide. Nature Reviews Molecular Cell Biology 8: 101-112

Hanahan D (1983) Studies on transformation of Escherichia coli with plasmids. Journal of Molecular Biology 166: 557-580

Hanania U, Velcheva M, Or E, Flaishman M, Sahar N and Perl A (2007) Silencing of chaperonin 21, that was differentially expressed in inflorescence of seedless and seeded grapes, promoted seed abortion in tobacco and tomato fruits. Transgenic Research 16: 515-525

Harrison CJ, HayerHartl M, DiLiberto M, Hartl FU and Kuriyan J (1997) Crystal structure of the nucleotide exchange factor GrpE bound to the ATPase domain of the molecular chaperone DnaK. Science 276: 431-435

Hartl FU (1996) Molecular chaperones in cellular protein folding. Nature 381: 571-580

Hartl FU and Hayer-Hartl M (2002) Protein folding - Molecular chaperones in the cytosol: from nascent chain to folded protein. Science 295: 1852-1858

Hartl FU and Hayer-Hartl M (2009) Converging concepts of protein folding in vitro and in vivo. Nature Structural & Molecular Biology 16: 574-581

Hartman FC and Harpel MR (1994) Structure, function, regulation, and assembly of D-ribulose-1,5-bisphosphate carboxylase/oxygenase. Annual Review of Biochemistry 63: 197-234

Hemmingsen SM and Ellis RJ (1986) Purification and properties of ribulose bisphosphate carboxylase large subunit bindig-protein. Plant Physiology 80: 269-276

Hemmingsen SM, Woolford C, Vandervies SM, Tilly K, Dennis DT, Georgopoulos CP, Hendrix RW and Ellis RJ (1988) Homologous plant and bacterial proteins chaperone oligomeric protein assembly. Nature 333: 330-334

Hill JE and Hemmingsen SM (2001) Arabidopsis thaliana type I and II chaperonins. Cell Stress & Chaperones 6: 190-200

Horton HR, Moran LA, Ochs RS, Rawn DJ and Scrimgeour KG (2002) Principles of Biochemistry. Prentice-Hall, Inc., New Jersey

Horwich AL, Apetri AC and Fenton WA (2009) The GroEL/GroES cis cavity as a passive anti-aggregation device. Febs Letters 583: 2654-2662

Horwich AL, Fenton WA, Chapman E and Farr GW (2007) Two families of chaperonin: Physiology and mechanism. Annual Review of Cell and Developmental Biology 23: 115-145

Houtz RL, Magnani R, Nayak NR and Dirk LMA (2008) Co- and post-translational modifications in Rubisco: unanswered questions. Journal of Experimental Botany 59: 1635-1645

Houtz RL and Portis AR (2003) The life of ribulose 1,5-bisphosphate carboxylase/oxygenase-posttranslational facts and mysteries. Archives of Biochemistry and Biophysics 414: 150-158

Houtz RL, Stults JT, Mulligan RM and Tolbert NE (1989) Post-translational modifications in the large subunit of ribulose bisphosphate carboxylase/oxygenase. Proceedings of the National Academy of Sciences of the United States of America 86: 1855-1859

Huq S, Sueoka K, Narumi S, Arisaka F and Nakamoto H (2010) Comparative Biochemical Characterization of Two GroEL Homologs from the Cyanobacterium Synechococcus elongatus PCC 7942. Bioscience, Biotechnology, and Biochemistry 74: 2273-2280

Imker HJ, Fedorov AA, Fedorov EV, Almo SC and Gerlt JA (2007) Mechanistic diversity in the Rubisco superfamily: The "Enolase" in the methionine salvage pathway in Geobacillus kaustophilus. Biochemistry 46: 4077-4089

Jager KM and Bergman B (1990) Localization of a multifunctional chaperonin (GroEL protein) in nitrogen-fixing Anabaena PCC 7120 - presence in vegetative cells and hererocysts. Planta 183: 120-125

Jahn TR and Radford SE (2005) The Yin and Yang of protein folding. Febs Journal 272: 5962-5970

Kaiser CM, Chang HC, Agashe VR, Lakshmipathy SK, Etchells SA, Hayer-Hartl M, Hartl FU and Barral JM (2006) Real-time observation of trigger factor function on translating ribosomes. Nature 444: 455-460

Kanevski I, Maliga P, Rhoades DF and Gutteridge S (1999) Plastome engineering of ribulose-1,5-bisphosphate carboxylase/oxygenase in tobacco to form a sunflower large subunit and tobacco small subunit hybrid. Plant Physiology 119: 133-141

Karkehabadi S, Peddi SR, Anwaruzzaman M, Taylor TC, Cederlund A, Genkov T, Andersson I and Spreitzer RJ (2005) Chimeric small subunits influence catalysis without causing global conformational changes in the crystal structure of ribulose-1,5-bisphosphate carboxylase/oxygenase. Biochemistry 44: 9851-9861

Keeley JE and Rundel PW (2003) Evolution of CAM and C-4 carbon-concentrating mechanisms. International Journal of Plant Sciences 164: S55-S77

Kellogg EA and Juliano ND (1997) The structure and function of Rubisco and their implications for systematic studies. American Journal of Botany 84: 413-428

Kennedy R (2000) Photosynthesis: Physiology and metabolism. In: Leegood RC, Sharkey TD and von Caemmerer S (eds) Photosynthesis: Physiology and metabolism, pp i-624, Kluwer Academic Publishers

Kerner MJ, Naylor DJ, Ishihama Y, Maier T, Chang HC, Stines AP, Georgopoulos C, Frishman D, Hayer-Hartl M, Mann M and Hartl FU (2005) Proteome-wide analysis of chaperonin-dependent protein folding in Escherichia coli. Cell 122: 209-220

Kettleborough CA, Parry MAJ, Burton S, Gutteridge S, Keys AJ and Phillips AL (1987) The role of the N-terminus of the large subunit of ribulose-bisphosphate carboxylase investigated by construction and expression of chimaeric genes. European Journal of Biochemistry 170: 335-342

Khan S, Andralojc PJ, Lea PJ and Parry MAJ (1999) 2 '-Carboxy-D-arabitinol 1-phosphate protects ribulose 1,5-bisphosphate carboxylase/oxygenase against proteolytic breakdown. European Journal of Biochemistry 266: 840-847

Kitano K, Maeda N, Fukui T, Atomi H, Imanaka T and Miki K (2001) Crystal structure of a novel-type archaeal Rubisco with pentagonal symmetry. Structure 9: 473-481

Klunker D, Haas B, Hirtreiter A, Figueiredo L, Naylor DJ, Pfeifer G, Muller V, Deppenmeier U, Gottschalk G, Hartl FU and Hayer-Hartl M (2003) Coexistence of group I and group II chaperonins in the archaeon Methanosarcina mazei. Journal of Biological Chemistry 278: 33256-33267

Knight S, Andersson I and Branden CI (1990) Crystallographic analysis of ribulose-1,5-bisphosphate carboxylase from spinach at 2.4 A resolution - subunit interactions and active-site. Journal of Molecular Biology 215: 113-160

Koumoto Y, Shimada T, Kondo M, Hara-Nishimura I and Nishimura M (2001) Chloroplasts have a novel Cpn10 in addition to Cpn20 as co-chaperonins in Arabidopsis thaliana. Journal of Biological Chemistry 276: 29688-29694

Koumoto Y, Shimada T, Kondo M, Takao T, Shimonishi Y, Hara-Nishimura I and Nishimura M (1999) Chloroplast Cpn20 forms a tetrameric structure in Arabidopsis thaliana. Plant Journal 17: 467-477

Kovacs E, Torok Z, Horvath I and Vigh L (1994) Heat-stress induces association of the GroEL-analog chaperonin with thylakoid membranes in cyanobacterium, Synechocystis PCC 6803. Plant Physiology and Biochemistry 32: 285-293

Kovacs E, van der Vies SM, Glatz A, Torok Z, Varvasovszki V, Horvath I and Vigh L (2001) The Chaperonins of Synechocystis PCC 6803 differ in heat inducibility and chaperone activity. Biochemical and Biophysical Research Communications 289: 908-915

Laemmli UK (1970) Cleavage of structural proteins during assembly of head bacteriophage-T4. Nature 227: 680-685

Lakshmipathy SK, Tomic S, Kaiser CM, Chang HC, Genevaux P, Georgopoulos C, Barral JM, Johnson AE, Hartl FU and Etchells SA (2007) Identification of nascent chain interaction sites on trigger factor. Journal of Biological Chemistry 282: 12186-12193

Langer T, Lu C, Echols H, Flanagan J, Hayer M and Hartl FU (1992) Successive action of DnaK, DnaJ and GroEL along the pathway of chaperone-mediated protein folding. Nature 356: 683-689

Larimer FW and Soper TS (1993) Overproduction of Anabaena 7120 ribulose bisphosphate carboxylase/oxygenase in Escherichia coli. Gene 126: 85-92

Laskey RA, Honda BM, Mills AD and Finch JT (1978) Nucleosomes are assembled by an acidic protein which binds histones and transfers them to DNA Nature 275: 416-420

Leegood RC, Sharkey TD and von Caemmerer S (2000) Introduction. In: Leegood RC, Sharkey TD and von Caemmerer S (eds) Photosynthesis: Physiology and metabolism, pp 1-8, Kluwer Academic Publishers

Lehel C, Los D, Wada H, Gyorgyei J, Horvath I, Kovacs E, Murata N and Vigh L (1993) A 2nd GroEL-like gene, organized in a GroESL operon is present in the genome of Synechocystis sp. PCC 6803. Journal of Biological Chemistry 268: 1799-1804

Lehel C, Wada H, Kovacs E, Torok Z, Gombos Z, Horvath I, Murata N and Vigh L (1992) Heat-shock protein-synthesis of the cyanobacterium Synechococcus synechocystis PCC 6803 - purification of the GroEL-related chaperonin. Plant Molecular Biology 18: 327-336

Leslie M (2009) On the Origin of Photosynthesis. Science 323: 1286-1287

Levinthal C (1968) Are there pathways for protein folding? Journal de Chimie Physique et de Physico-Chimi Biologique 65: 44-45

Levy-Rimler G, Bell RE, Ben-Tal N and Azem A (2002) Type I chaperonins: not all are created equal. Febs Letters 529: 1-5

Li HY, Sawaya MR, Tabita FR and Eisenberg D (2005) Crystal structure of a Rubisco-like protein from the green sulfur bacterium Chlorobium tepidum. Structure 13: 779-789

Li LA and Tabita FR (1997) Maximum activity of recombinant ribulose 1,5-bisphosphate carboxylase/oxygenase of Anabaena sp. strain CA requires the product of the rbcX gene. Journal of Bacteriology 179: 3793-3796

Lin Z, Madan D and Rye HS (2008) GroEL stimulates protein folding through forced unfolding. Nature Structural & Molecular Biology 15: 303-311

Lin Z and Rye HS (2006) GroEL-Mediated protein folding: Making the impossible, possible. Critical Reviews in Biochemistry and Molecular Biology 41: 211-239

Lissin NM (1995) In-vitro dissociation and self-assembly of 3 chaperonin-60s - the role of ATP. Febs Letters 361: 55-60

Liu CM, Young AL, Starling-Windhof A, Bracher A, Saschenbrecker S, Rao BV, Rao KV, Berninghausen O, Mielke T, Hartl FU, Beckmann R and Hayer-Hartl M (2010) Coupled chaperone action in folding and assembly of hexadecameric Rubisco. Nature 463: 197-U81

Liu JY, Huang CD, Shin DH, Yokota H, Jancarik J, Kim JS, Adams PD, Kim R and Kim SH (2005) Crystal structure of a heat-inducible transcriptional repressor HrcA from Thermotoga maritima: Structural insight into DNA binding and dimerization. Journal of Molecular Biology 350: 987-996

Lorimer GH (2001) A personal account of chaperonin history. Plant Physiology 125: 38-41

Lorimer GH, Badger MR and Andrews TJ (1977) D-ribulose-1,5-bisphosphate carboxylase/oxygenase: improved methods for activation and assay of catalytic activities. Analytical Biochemistry 78: 66-75

Lubben TH, Donaldson GK, Viitanen PV and Gatenby AA (1989) Several proteins imported into chloroplasts form stable complexes with the GroEL-related chloroplast molecular chaperone. Plant Cell 1: 1223-1230

Lund PA (2009) Multiple chaperonins in bacteria - why so many? Fems Microbiology Reviews 33: 785-800

Maier R, Eckert B, Scholz C, Lilie H and Schmid FX (2003) Interaction of trigger factor with the ribosome. Journal of Molecular Biology 326: 585-592

Martel R, Cloney LP, Pelcher LE and Hemmingsen SM (1990) Unique composition of plastid chaperonin-60 - alpha-polypeptide-encoding and beta-polypeptide-encoding genes are highly divergent. Gene 94: 181-187

Martin J (2004) Chaperonin function - effects of crowding and confinement. Journal of Molecular Recognition 17: 465-472

Mayer MP, Rudiger S and Bukau B (2000) Molecular basis for interactions of the DnaK chaperone with substrates. Biological Chemistry 381: 877-885

Moroney JV and Ynalvez RA (2007) Proposed carbon dioxide concentrating mechanism in Chlamydomonas reinhardtii. Eukaryotic Cell 6: 1251-1259

Mueller-Cajar O and Whitney SM (2008) Evolving improved Synechococcus Rubisco functional expression in Escherichia coli. Biochemical Journal 414: 205-214

Musgrove JE, Johnson RA and Ellis RJ (1987) Dissociation of the ribulose bisphosphate-carboxylase large-subunit binding-protein into dissimilar subunits. European Journal of Biochemistry 163: 529-534

Newman J, Branden CI and Jones TA (1993) Structure determination and refinement of ribulose1,5-bisphosphate carboxylase/oxygenase from Synechococcus PCC 6301. Acta Crystallographica Section D-Biological Crystallography 49: 548-560

Newman J and Gutteridge S (1994) Structure of an effector-induced inactivated state of ribulose-1,5-bisphosphate carboxylase/oxygenase - the binary complex between enzyme and xyluose-1,5-bisphosphate. Structure 2: 495-502

Nishihara K, Kanemori M, Yanagi H and Yura T (2000) Overexpression of trigger factor prevents aggregation of recombinant proteins in Escherichia coli. Applied and Environmental Microbiology 66: 884-889

Nishio K, Hirohashi T and Nakai M (1999) Chloroplast chaperonins: Evidence for heterogeneous assembly of alpha and beta Cpn60 polypeptides into a chaperonin oligomer. Biochemical and Biophysical Research Communications 266: 584-587

Olinares PDB, Ponnala L and van Wijk KJ (2010) Megadalton Complexes in the Chloroplast Stroma of Arabidopsis thaliana Characterized by Size Exclusion Chromatography, Mass Spectrometry, and Hierarchical Clustering. Molecular & Cellular Proteomics 9: 1594-1615

Olzscha H, Schermann SM, Woerner AC, Pinkert S, Hecht MH, Tartaglia GG, Vendruscolo M, Hayer-Hartl M, Hartl FU and Vabulas RM (2011) Amyloid-like Aggregates Sequester Numerous Metastable Proteins with Essential Cellular Functions. Cell 144: 67-78

Onizuka T, Endo S, Akiyama H, Kanai S, Hirano M, Yokota A, Tanaka S and Miyasaka H (2004) The rbcX gene product promotes the production and assembly of ribulose-1,5-bisphosphate carboxylase/oxygenase of Synechococcus sp PCC7002 in Escherichia coli. Plant and Cell Physiology 45: 1390-1395

Orr L and Govindjee (2010) Photosynthesis online. Photosynthesis Research 105: 167-200

Osmond B, Badger M, Maxwell K, Bjorkman O and Leegood R (1997) Too many photos: Photorespiration, photoinhibition and photooxidation. Trends in Plant Science 2: 119-121

Park ES, Fenton WA and Horwich AL (2007) Disulfide formation as a probe of folding in GroEL-GroES reveals correct formation of long-range bonds and editing of incorrect short-range ones. Proceedings of the National Academy of Sciences of the United States of America 104: 2145-2150

Parry MAJ, Andralojc PJ, Mitchell RAC, Madgwick PJ and Keys AJ (2003) Manipulation of Rubisco: the amount, activity, function and regulation. Journal of Experimental Botany 54: 1321-1333

Parry MAJ, Keys AJ, Madgwick PJ, Carmo-Silva AE and Andralojc PJ (2008) Rubisco regulation: a role for inhibitors. Journal of Experimental Botany 59: 1569-1580

Peltier G, Tolleter D, Billon E and Cournac L (2010) Auxiliary electron transport pathways in chloroplasts of microalgae. Photosynthesis Research 106: 19-31

Peltier JB, Cai Y, Sun Q, Zabrouskov V, Giacomelli L, Rudella A, Ytterberg AJ, Rutschow H and van Wijk KJ (2006) The oligomeric stromal proteome of Arabidopsis thaliana chloroplasts. Molecular & Cellular Proteomics 5: 114-133

Peng L, Fukao Y, Myouga F, Motohashi R, Shinozaki K and Shikanai T (2011) A chaperonin subunit with unique structures is essential for folding of a specific substrate. PLoS Biology 9: e1001040. doi:10.1371/journal.pbio.1001040

Peterhansel C and Maurino VG (2011) Photorespiration Redesigned. Plant Physiology 155: 49-55

Pierce J, Tolbert NE and Barker R (1980) Interaction of ribulose bisphosphate carboxylase/oxygenase with transition-state analogs. Biochemistry 19: 934-942

Quan S, Koldewey P, Tapley T, Kirsch N, Ruane KM, Pfizenmaier J, Shi R, Hofmann S, Foit L, Ren G, Jakob U, Xu ZH, Cygler M and Bardwell JCA (2011) Genetic selection designed to stabilize proteins uncovers a chaperone called Spy. Nature Structural & Molecular Biology 18: 262-270

Rajaram H and Apte K (2010) Differential regulation of groESL operon expression in response to heat and light in Anabaena. Archives of Microbiology 192: 729-738

Rajaram H, Ballal AD, Apte SK, Wiegert T and Schumann W (2001) Cloning and characterization of the major groESL operon from a nitrogen-fixing cyanobacterium Anabaena sp strain L-31. Biochimica Et Biophysica Acta-Gene Structure and Expression 1519: 143-146

Ramazanov Z, Rawat M, Henk MC, Mason CB, Matthews SW and Moroney JV (1994) The induction of the CO2-concentrating mechanism is correlated with the formation of the starch sheath around the pyrenoid of Chlamydomonas reinhardtii. Planta 195: 210-216

Rao BV (2009a) Structural characterization of chaperone assisted folding and assemble of Rubisco. PhD Thesis, Faculty of Chemistry and Pharmacy, Ludwig-Maximilians-Universitaet, Munich

Rao KV (2009b) Chaperone assisted Rubisco folding and assembly-role of RbcX. PhD Thesis, Faculty of Chemistry and Pharmacy, Ludwig-Maximilians-Universitaet, Munich

Read BA and Tabita FR (1992) Amino acid substitutions in the small subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase that influence catalytic activity of the holoenzyme. Biochemistry 31: 519-525

Roy H, Cannon S and Gilson M (1988) Assembly of Rubisco from native subunits. Biochimica Et Biophysica Acta 957: 323-334

Ruediger S, Buchberger A and Bukau B (1997) Interaction of Hsp70 chaperones with substrates. Nature Structural Biology 4: 342-349

Ryu YH and Schultz PG (2006) Efficient incorporation of unnatural amino acids into proteins in Escherichia coli. Nature Methods 3: 263-265

Sage RF, Way DA and Kubien DS (2008) Rubisco, Rubisco activase, and global climate change. Journal of Experimental Botany 59: 1581-1595

Salvucci ME (2008) Association of Rubisco activase with chaperonin-60 beta: a possible mechanism for protecting photosynthesis during heat stress. Journal of Experimental Botany 59: 1923-1933

Sambrook J, Fritsch EF and Maniatis T (1989) Molecular cloning: a laboratory manual SECOND EDITION VOLS. 1 2 AND 3. Cold Spring Harbor Press, New York

Saschenbrecker S, Bracher A, Rao KV, Rao BV, Hartl FU and Hayer-Hartl M (2007) Structure and function of RbcX, an assembly chaperone for hexadecameric rubisco. Cell 129: 1189-1200

Sato T, Atomi H and Imanaka T (2007) Archaeal type III RuBisCOs function in a pathway for AMP metabolism. Science 315: 1003-1006

Savage DF, Afonso B, Chen AH and Silver PA (2010) Spatially Ordered Dynamics of the Bacterial Carbon Fixation Machinery. Science 327: 1258-1261

Schlicher T and Soll J (1996) Molecular chaperones are present in the thylakoid lumen of pea chloroplasts. Febs Letters 379: 302-304

Schneider G, Lindqvist Y and Lundqvist T (1990) Crystallographic refinement and structure of Ribulose-1,5-bisphosphate carboxylase from Rhodospirillum rubrum at 1.7 A resolution Journal of Molecular Biology 211: 989-1008

Sharkia R, Bonshtien AL, Mizrahi I, Weiss C, Niv A, Lustig A, Viitanen PV and Azem A (2003) On the oligomeric state of chloroplast chaperonin 10 and chaperonin 20. Biochimica Et Biophysica Acta-Proteins and Proteomics 1651: 76-84

Sharma S, Chakraborty K, Mueller BK, Astola N, Tang YC, Lamb DC, Hayer-Hartl M and Hartl FU (2008) Monitoring protein conformation along the pathway of chaperonin-assisted folding. Cell 133: 142-153

Shimizu Y, Kanamori T and Ueda T (2005) Protein synthesis by pure translation systems. Methods 36: 299-304

Smith SA and Tabita FR (2003) Positive and negative selection of mutant forms of prokaryotic (cyanobacterial) ribulose-1,5-bisphosphate carboxylase/oxygenase. Journal of Molecular Biology 331: 557-569

Smrcka AV, Bohnert HJ and Jensen RG (1991) Modulation of the tight binding carboxyarabinitol 1,5-bisphosphate to the large subunit of ribulose 1,5-bisphosphate carboxylase/oxygenase. Archives of Biochemistry and Biophysics 286: 14-19

Soll J and Schleiff E (2004) Protein import into chloroplasts. Nature Reviews Molecular Cell Biology 5: 198-208

Somerville CR, McIntosh L, Fitchen J and Gurevitz M (1986) The cloning and expression in Escherichi coli of RuBP carboxylase oxygenase large subunit genes. Methods in Enzymology 118: 419-433

Somerville CR and Ogren WL (1982) Genetic modification of photorespiration. Trends in Biochemical Sciences 7: 171-174

Somerville CR and Somerville SC (1984) Cloning and expression of the Rhodospirillum rubrum Ribulosebisphosphate carboxylase gene in E. coli. Molecular & General Genetics 193: 214-219

Spreitzer RJ (2003) Role of the small subunit in ribulose-1,5-bisphosphate carboxylase/oxygenase. Archives of Biochemistry and Biophysics 414: 141-149

Spreitzer RJ and Salvucci ME (2002) Rubisco: Structure, regulatory interactions, and possibilities for a better enzyme. Annual Review of Plant Biology 53: 449-475

Szabo A, Langer T, Schroder H, Flanagan J, Bukau B and Hartl FU (1994) The ATP hydrolysis-dependent reaction cycle of the Escherichia coli Hsp70 system-DnaK, DnaJ, and GrpE. Proceedings of the National Academy of Sciences of the United States of America 91: 10345-10349

Tabita FR (1999) Microbial ribulose 1,5-bisphosphate carboxylase/oxygenase: A different perspective. Photosynthesis Research 60: 1-28

Tabita FR (2007) Rubisco: The enzyme that keeps on giving. Cell 129: 1039-1040

Tabita FR, Hanson TE, Li HY, Satagopan S, Singh J and Chan S (2007) Function, structure, and evolution of the Rubisco-like proteins and their Rubisco homologs. Microbiology and Molecular Biology Reviews 71: 576-+

Tabita FR, Satagopan S, Hanson TE, Kreel NE and Scott SS (2008) Distinct form I, II, III, and IV Rubisco proteins from the three kingdoms of life provide clues about Rubisco evolution and structure/function relationships. Journal of Experimental Botany 59: 1515-1524

Tanaka N, Hiyama T and Nakamoto H (1997) Cloning, characterization and functional analysis of groESL operon from thermophilic cyanobacterium Synechococcus vulcanus. Biochimica Et Biophysica Acta-Protein Structure and Molecular Enzymology 1343: 335-348

Tanaka S, Kerfeld CA, Sawaya MR, Cai F, Heinhorst S, Cannon GC and Yeates TO (2008) Atomic-level models of the bacterial carboxysome shell. Science 319: 1083-1086

Tanaka S, Sawaya MR, Kerfeld CA and Yeates TO (2007) Structure of the Rubisco chaperone RbcX from Synechocystis sp PCC6803. Acta Crystallographica Section D-Biological Crystallography 63: 1109-1112

Tang YC, Chang HC, Chakraborty K, Hartl FU and Hayer-Hartl M (2008) Essential role of the chaperonin folding compartment in vivo. Embo Journal 27: 1458-1468

Tang YC, Chang HC, Roeben A, Wischnewski D, Wischnewski N, Kerner MJ, Hartl FU and Hayer-Hartl M (2006) Structural features of the GroEL-GroES nano-cage required for rapid folding of encapsulated protein. Cell 125: 903-914

Tarnawski M, Gubernator B, Kolesinski P and Szczepaniak A (2008) Heterologous expression and initial characterization of recombinant RbcX protein from Thermosynechococcus elongatus BP-1 and the role of RbcX in RuBisCO assembly. Acta Biochimica Polonica 55: 777-785

Tavano CL and Donohue TJ (2006) Development of the bacterial photosynthetic apparatus. Current Opinion in Microbiology 9: 625-631

Taylor TC and Andersson I (1997) The structure of the complex between rubisco and its natural substrate ribulose 1,5-bisphosphate. Journal of Molecular Biology 265: 432-444

Teter SA, Houry WA, Ang D, Tradler T, Rockabrand D, Fischer G, Blum P, Georgopoulos C and Hartl FU (1999) Polypeptide flux through bacterial Hsp70: DnaK cooperates with trigger factor in chaperoning nascent chains. Cell 97: 755-765

Thompson MD, Paavola CD, Lenvik TR and Gantt JS (1995) Chlamydomonas transcripts encoding three divergent plastid chaperonins are heat-inducible. Plant Molecular Biology 27: 1031-1035

Ting IP (1985) Crassulacean acid metabolism. Annual Review of Plant Physiology and Plant Molecular Biology 36: 595-622

Tipple BJ and Pagani M (2007) The early origins of terrestrial C-4 photosynthesis. Annual Review of Earth and Planetary Sciences 35: 435-461

Towbin H, Staehelin T and Gordon J (1979) Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proceedings of the National Academy of Sciences of the United States of America 76: 4350-4354

Umena Y, Kawakami K, Shen JR and Kamiya N (2011) Crystal structure of oxygen-evolving photosystem II at a resolution of 1.9 A. Nature 473: 55-65

Vabulas RM, Raychaudhuri S, Hayer-Hartl M and Hartl FU (2010) Protein folding in the cytoplasm and the heat shock response. Cold Spring Harbor Perspectives in Biology 2: a004390

van Lun M, van der Spoel D and Andersson I (2011) Subunit interface dynamics in hexadecameric Rubisco. Journal of Molecular Biology doi:10.1016/j.jmb.2011.06.052

Viitanen PV, Lubben TH, Reed J, Goloubinoff P, Okeefe DP and Lorimer GH (1990) Chaperonin-facilitated refolding of Ribulose bisphosephate carboxylase and ATP hydrolysis by chaperonin 60 (GroEL) are potassium dependent. Biochemistry 29: 5665-5671

Viitanen PV, Schmidt M, Buchner J, Suzuki T, Vierling E, Dickson R, Lorimer GH, Gatenby A and Soll J (1995) Functional characterization of the higher plant chloroplast chaperonins. Journal of Biological Chemistry 270: 18158-18164

Wastl J, Fraunholz M, Zauner S, Douglas S and Maier UG (1999) Ancient gene duplication and differential gene flow in plastid lineages: The GroEL/Cpn60 example. Journal of Molecular Evolution 48: 112-117

Watson GMF, Yu JP and Tabita FR (1999) Unusual ribulose 1,5-bisphosphate carboxylase/oxygenase of anoxic Archaea. Journal of Bacteriology 181: 1569-1575

Webb R, Reddy KJ and Sherman LA (1990) Regulation and sequence of the *Synechococcus* sp strain-PCC-7942 GroESL operon, encoding a cyanobacterial chaperonin. Journal of Bacteriology 172: 5079-5088

Weiner MP, Costa GL, Schoettlin W, Cline J, Mathur E and Bauer JC (1994) Site-directed mutagenesis of double-stranded DNA by the polymerase chain reaction. Gene 151: 119-123

Weiss C, Bonshtien A, Farchi-Pisanty O, Vitlin A and Azem A (2009) Cpn20: Siamese twins of the chaperonin world. Plant Molecular Biology 69: 227-238

West-Eberhard MJ, Smith JAC and Winter K (2011) Plant Science: Photosynthesis, reorganized. Science 332: 311-312

Whitney SM and Andrews TJ (2001) Plastome-encoded bacterial ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) supports photosynthesis and growth in tobacco. Proceedings of the National Academy of Sciences of the United States of America 98: 14738-14743

Whitney SM, Baldett P, Hudson GS and Andrews TJ (2001) Form I Rubiscos from non-green algae are expressed abundantly but not assembled in tobacco chloroplasts. Plant Journal 26: 535-547

Witt E, Zantopf D, Schmidt M, Kraft R, Kloetzel PM and Kruger E (2000) Characterisation of the newly identified human Ump1 homologue POMP and analysis of LMP7(beta 5i) incorporation into 20 S proteasomes. Journal of Molecular Biology 301: 1-9

Xiong AS, Peng RH, Zhuang J, Gao F, Zhu B, Fu XY, Xue Y, Jin XF, Tian YS, Zhao W and Yao QH (2009) Gene duplication, transfer, and evolution in the chloroplast genome. Biotechnology Advances 27: 340-347

Xu ZH, Horwich AL and Sigler PB (1997) The crystal structure of the asymmetric GroEL-GroES-(ADP)(7) chaperonin complex. Nature 388: 741-750

Yoon M, Putterill JJ, Ross GS and Laing WA (2001) Determination of the relative expression levels of rubisco small subunit genes in Arabidopsis by rapid amplification of cDNA ends. Analytical Biochemistry 291: 237-244

Young JC, Barral JM and Hartl FU (2003) More than folding: localized functions of cytosolic chaperones. Trends in Biochemical Sciences 28: 541-547

Zhu GH and Jensen RG (1990) Statues of the substrate binding-sites of ribulose bisphosphate carboxylase as determined with 2-C-carboxyarabinitol 1,5-bisphosphate. Plant Physiology 93: 244-249

Zhu XT, Zhao X, Burkholder WF, Gragerov A, Ogata CM, Gottesman ME and Hendrickson WA (1996) Structural analysis of substrate binding by the molecular chaperone DnaK. Science 272: 1606-1614

Zimmerman SB and Trach SO (1991) Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of Escherichia coli Journal of Molecular Biology 222: 599-620

Zuber U and Schumann W (1994) CIRCE, a novel heat shock element involved in regulation of heat shock operon dnaK of Bacillus subtilis. Journal of Bacteriology 176: 1359-1363

7 Appendices

7.1 Plasmids constructed in this study

Plasmid	Description
AnaCA-*rbcX(TAA)*_{N-6His}-pET28b	Plasmid encoding *Anabaena* sp. CA RbcX_{N-6His} with TAA stop codon in pET28b expression vector (Liu et al. 2010)
AnaCA-*rbcX(TAA)*_{N-6His}/Syn6301-*rbcL*-pET28b	Coexpression plasmid encoding *Synechococcus* sp. PCC6301 Rubisco large subunit and *Anabaena* sp. CA RbcX_{N-6His} with TAA stop codon in pET28b expression vector (Liu et al. 2010)
AnaCA-*rbcX(Q5TAG,TAA)*_{N-6His}/Syn6301-*rbcL*-pET28b	Coexpression plasmid encoding *Synechococcus* sp. PCC6301 Rubisco large subunit and *Anabaena* sp. CA RbcX_{N-6His} with TAG mutation at position Q5 and TAA stop codon in pET28b expression vector (Liu et al. 2010)
AnaCA-*rbcX(T25TAG,TAA)*_{N-6His}/Syn6301-*rbcL*-pET28b	Coexpression plasmid encoding *Synechococcus* sp. PCC6301 Rubisco large subunit and *Anabaena* sp. CA RbcX_{N-6His} with TAG mutation at position T25 and TAA stop codon in pET28b expression vector (Liu et al. 2010)
AnaCA-*rbcX(H42TAG,TAA)*_{N-6His}/Syn6301-*rbcL*-pET28b	Coexpression plasmid encoding *Synechococcus* sp. PCC6301 Rubisco large subunit and *Anabaena* sp. CA RbcX_{N-6His} with TAG mutation at position H42 and TAA stop codon in pET28b expression vector (Liu et al. 2010)
AnaCA-*rbcX(A28TAG,TAA)*_{N-6His}/Syn6301-*rbcL*-pET28b	Coexpression plasmid encoding *Synechococcus* sp. PCC6301 Rubisco large subunit and *Anabaena* sp. CA RbcX_{N-6His} with TAG mutation at position A28 and TAA stop codon in pET28b expression vector (Liu et al. 2010)
AnaCA-*rbcX(T33TAG,TAA)*_{N-6His}/Syn6301-*rbcL*-pET28b	Coexpression plasmid encoding *Synechococcus* sp. PCC6301 Rubisco large subunit and *Anabaena* sp. CA RbcX_{N-6His} with TAG mutation at position T33 and TAA stop codon in pET28b expression vector (Liu et al. 2010)
AnaCA-*rbcX(K54TAG,TAA)*_{N-6His}/Syn6301-*rbcL*-pET28b	Coexpression plasmid encoding *Synechococcus* sp. PCC6301 Rubisco large subunit and *Anabaena* sp. CA RbcX_{N-6His} with TAG mutation at position K54 and TAA stop codon in pET28b expression vector (Liu et al. 2010)
AnaCA-*rbcX(L105TAG,TAA)*_{N-6His}/Syn6301-*rbcL*-pET28b	Coexpression plasmid encoding *Synechococcus* sp. PCC6301 Rubisco large subunit and *Anabaena* sp. CA RbcX_{N-6His} with TAG mutation at position L105 and TAA stop codon in pET28b expression vector (Liu et al. 2010)
Syn6301-*rbcL(E49A)*-pET11a	Plasmid encoding *Synechococcus* sp. PCC6301 RbcL(E49A) in pET11a expression vector (Bracher et al. 2011)
Syn6301-*rbcL(E49C)*-pET11a	Plasmid encoding *Synechococcus* sp. PCC6301 RbcL(E49C) in pET11a expression vector (Bracher et al. 2011)
Syn6301-*rbcL(A53H)*-pET11a	Plasmid encoding *Synechococcus* sp. PCC6301 RbcL(A53H) in pET11a expression vector (Bracher et al. 2011)
Syn6301-*rbcL(A53T)*-pET11a	Plasmid encoding *Synechococcus* sp. PCC6301 RbcL(A53T) in pET11a expression vector

Plasmid	Description
Syn6301-*rbcL(A126T)*-pET11a	Plasmid encoding *Synechococcus* sp. PCC6301 RbcL(A126T) in pET11a expression vector
Syn6301-*rbcL(A126Y)*-pET11a	Plasmid encoding *Synechococcus* sp. PCC6301 RbcL(A126Y) in pET11a expression vector (Bracher et al. 2011)
AnaCA-*rbcX$_{N-6His}$*/Syn6301-*rbcL*-pET28b	Coexpression plasmid encoding *Anabaena* sp. CA RbcX$_{N-6His}$ and *Synechococcus* sp. PCC6301 RbcL in pET28b expression vector (Bracher et al. 2011)
AnaCA-*rbcX$_{N-6His}$*/Syn6301-*rbcL(E49C)*-pET28b	Coexpression plasmid encoding *Anabaena* sp. CA RbcX$_{N-6His}$ and *Synechococcus* sp. PCC6301 RbcL(E49C) in pET28b expression vector (Bracher et al. 2011)
AnaCA-*rbcX(T33C)$_{N-6His}$*/Syn6301-*rbcL(E49C)*-pET28b	Coexpression plasmid encoding *Anabaena* sp. CA RbcX(T33C)$_{N-6His}$ and *Synechococcus* sp. PCC6301 RbcL(E49C) in pET28b expression vector (Bracher et al. 2011)
AnaCA-*rbcX(E83C)$_{N-6His}$*/Syn6301-*rbcL(E49C)*-pET28b	Coexpression plasmid encoding *Anabaena* sp. CA RbcX(E83C)$_{N-6His}$ and *Synechococcus* sp. PCC6301 RbcL(E49C) in pET28b expression vector (Bracher et al. 2011)
Syn6301-*rbcX(T51C)$_{N-6His}$*-pET28b	Plasmid encoding *Synechococcus* sp. PCC6301 RbcX(T51C)$_{N-6His}$ in pET28b expression vector
Syn6301-*rbcX(D102C)$_{N-6His}$*-pET28b	Plasmid encoding *Synechococcus* sp. PCC6301 RbcX(D102C)$_{N-6His}$ in pET28b expression vector
Syn6301-*rbcX$_{N-6His}$*/Syn6301-*rbcL*-pET28b	Coexpression plasmid encoding *Synechococcus* sp. PCC6301 RbcX$_{N-6His}$ and *Synechococcus* sp. PCC6301 RbcL in pET28b expression vector
Syn6301-*rbcX$_{N-6His}$*/Syn6301-*rbcL(E49C)*-pET28b	Coexpression plasmid encoding *Synechococcus* sp. PCC6301 RbcX$_{N-6His}$ and *Synechococcus* sp. PCC6301 RbcL(E49C) in pET28b expression vector
Syn6301-*rbcX(T51C)$_{N-6His}$*/Syn6301-*rbcL(E49C)*-pET28b	Coexpression plasmid encoding *Synechococcus* sp. PCC6301 RbcX(T51C)$_{N-6His}$ and *Synechococcus* sp. PCC6301 RbcL(E49C) in pET28b expression vector
Syn6301-*rbcX(D102C)$_{N-6His}$*/Syn6301-*rbcL(E49C)*-pET28b	Coexpression plasmid encoding *Synechococcus* sp. PCC6301 RbcX(D102C)$_{N-6His}$ and *Synechococcus* sp. PCC6301 RbcL(E49C) in pET28b expression vector
Syn6301-*rbcX$_{N-6His}$*/Syn6301-*rbcL(F342I)*-pET28b	Coexpression plasmid encoding *Synechococcus* sp. PCC6301 RbcX$_{N-6His}$ and *Synechococcus* sp. PCC6301 RbcL(F342I) in pET28b expression vector (Bracher et al. 2011)
Syn6301-*rbcX$_{N-6His}$*/Syn6301-*rbcL(E49C,F342I)*-pET28b	Coexpression plasmid encoding *Synechococcus* sp. PCC6301 RbcX$_{N-6His}$ and *Synechococcus* sp. PCC6301 RbcL(E49C,F342I) in pET28b expression vector (Bracher et al. 2011)
Syn6301-*rbcX(T51C)$_{N-6His}$*/Syn6301-*rbcL(E49C,F342I)*-pET28b	Coexpression plasmid encoding *Synechococcus* sp. PCC6301 RbcX(T51C)$_{N-6His}$ and *Synechococcus* sp. PCC6301 RbcL(E49C,F342I) in pET28b expression vector (Bracher et al. 2011)
Syn6301-*rbcX(D102C)$_{N-6His}$*/Syn6301-	Coexpression plasmid encoding *Synechococcus* sp. PCC6301 RbcX(D102C)$_{N-6His}$ and *Synechococcus* sp.

Plasmid	Description
rbcL(E49C,F342I)-pET28b	PCC6301 RbcL(E49C,F342I) in pET28b expression vector (Bracher et al. 2011)
Syn7002-*rbcL (E48C)*/Syn7002-*rbcX*$_{N-6His}$-pET28b	Coexpression plasmid encoding *Synechococcus* sp. PCC7002 RbcX$_{N-6His}$ and RbcL(E48C) in pET28b expression vector
Syn7002-*rbcL (E48C)*/Syn7002-*rbcX(T33C)*$_{N-6His}$-pET28b	Coexpression plasmid encoding *Synechococcus* sp. PCC7002 RbcX(T33C)$_{N-6His}$ and RbcL(E48C) in pET28b expression vector
Syn7002-*rbcL (E48C)*/Syn7002-*rbcX(E84C)*$_{N-6His}$-pET28b	Coexpression plasmid encoding *Synechococcus* sp. PCC7002 RbcX(E84C)$_{N-6His}$ and RbcL(E48C) in pET28b expression vector
Syn7002-*rbcL (E48C,F341I)* –pET11a	Plasmid encoding *Synechococcus* sp. PCC7002 RbcL(E48C,F341I) in pET11a expression vector
Syn7002-*rbcL (F341I)*/Syn7002-*rbcX*$_{N-6His}$-pET28b	Coexpression plasmid encoding *Synechococcus* sp. PCC7002 RbcX$_{N-6His}$ and RbcL(F341I) in pET28b expression vector
Syn7002-*rbcL (E48C,F341I)*/Syn7002-*rbcX(T33C)*$_{N-6His}$-pET28b	Coexpression plasmid encoding *Synechococcus* sp. PCC7002 RbcX(T33C)$_{N-6His}$ and RbcL(E48C,F341I) in pET28b expression vector
Syn7002-*rbcL (E48C,F341I)*/Syn7002-*rbcX(E84C)*$_{N-6His}$-pET28b	Coexpression plasmid encoding *Synechococcus* sp. PCC7002 RbcX(E84C)$_{N-6His}$ and RbcL(E48C,F341I) in pET28b expression vector
Syn7002-*rbcL (F341I)*/Syn7002-*rbcX(T33C)*$_{N-6His}$-pET28b	Coexpression plasmid encoding *Synechococcus* sp. PCC7002 RbcX(T33C)$_{N-6His}$ and RbcL(F341I) in pET28b expression vector
Syn7002-*rbcL (F341I)*/Syn7002-*rbcX(E84C)*$_{N-6His}$-pET28b	Coexpression plasmid encoding *Synechococcus* sp. PCC7002 RbcX(E84C)$_{N-6His}$ and RbcL(F341I) in pET28b expression vector
Syn7002-*rbcX(G42H,E43N)*$_{N-6His}$-pET28b	Plasmid encoding *Synechococcus* sp. PCC7002 RbcX(G42H,E43N)$_{N-6His}$ in pET28b expression vector
Syn6301-*rbcL(E106Q)*-pET11a	Plasmid encoding *Synechococcus* sp. PCC6301 RbcL(E106Q) in pET11a expression vector (Bracher et al. 2011)
AnaCA-*rbcX*$_{N-6His}$/Syn6301-*rbcL(E106Q)*-pET28b	Coexpression plasmid encoding *Anabaena* sp. CA RbcX$_{N-6His}$ and *Synechococcus* sp. PCC6301 RbcL(E106Q) in pET28b expression vector (Bracher et al. 2011)
Syn6301-*rbcL(R212S)*-pET28b	Plasmid encoding *Synechococcus* sp. 6301 RbcL(R212S) in pET11a expression vector (Bracher et al. 2011)
AnaCA-*rbcX*$_{N-6His}$/Syn6301-*rbcL(R212S)*-pET28b	Coexpression plasmid encoding *Anabaena* sp. CA RbcX$_{N-6His}$ and *Synechococcus* sp. PCC6301 RbcL(R212S) in pET28b expression vector (Bracher et al. 2011)
Syn6301-*rbcX*$_{N-FLAG}$/Syn6301-*rbcL*-pET11a	Coexpression plasmid encoding *Synechococcus* sp. PCC6301 RbcX$_{N-FLAG}$ and *Synechococcus* sp. PCC6301 RbcL in pET11a expression vector (Bracher et al. 2011)
Syn6301-*rbcX*$_{N-FLAG}$/Syn6301-*rbcL(W67A,L70G,L71A)*-pET11a	Coexpression plasmid encoding *Synechococcus* sp. PCC6301 RbcX$_{N-FLAG}$ and *Synechococcus* sp. PCC6301 RbcL(W67A,L70G,L71A) in pET11a expression vector (Bracher et al. 2011)
Syn6301-	Plasmid encoding *Synechococcus* sp. PCC6301

Plasmid	Description
*rbcL(W67A,L70G,L71A)-*pET11a	RbcL(W67A,L70G,L71A) in pET11a expression vector (Bracher et al. 2011)
Syn6301-*rbcLΔN(1-11)-*pET11a	Plasmid encoding *Synechococcus* sp. PCC6301 RbcLΔN(1-11) in pET11a expression vector
Syn6301-*rbcX$_{N-FLAG}$*/Syn6301-*rbcLΔN(1-11)*-pET11a	Coexpression plasmid encoding *Synechococcus* sp. PCC6301 RbcX$_{N-FLAG}$ and *Synechococcus* sp. PCC6301 RbcLΔN(1-11) in pET11a expression vector
Syn6301-*rbcLΔN(1-5)-*pET11a	Plasmid encoding *Synechococcus* sp. PCC6301 RbcLΔN(1-5) in pET11a expression vector
Syn6301-*rbcS(L60A,F63A)-*pET11a	Plasmid encoding *Synechococcus* sp PCC6301 RbcS(L60A,F63A) in pET11a expression vector
Ana7120-*groEL1*-pET11a	Plasmid encoding *Anabaena* sp. PCC7120 GroEL 1 in pET11a expression vector
Ana7120-*groEL2*-pET11a	Plasmid encoding *Anabaena* sp. PCC7120 GroEL 2 in pET11a expression vector
Ana7120-*groES*-pET11a	Plasmid encoding *Anabaena* sp. PCC7120 GroES in pET11a expression vector
Ana7120-*groEL2-*pCOLAduet-1	Plasmid encoding *Anabaena* sp. PCC7120 GroEL 2 in pCOLAduet-1 expression vector
Syn6301-*groEL1*-pET11a	Plasmid encoding *Synechococcus* sp. PCC6301 GroEL 1 in pET11a expression vector
Syn6301-*groES*-pET11a	Plasmid encoding *Synechococcus* sp. PCC6301 GroES in pET11a expression vector

Table 7-1 Plasmids designed in this study.

7.2   Primer pairs used for PCR

Oligo name	Primer sequence 5´-3´
AnaCAX_Fwd_NdeI	ACTAGT**CAT*ATG***AACCTCAAGCAAATTGCGAAAGATACAGCC
AnaCA X(TAA)_NheI_BamHI_Rev	GAGCAG**GGATCCGCTAGC***TTA*ACTGGCTAAATTATCCCAGTCA
AnaCAX_Q5stop_Fwd	CCAAAACACTCCAAAGCTACCTGACTTATCAG
AnaCAX_Q5stop_RevPhos	CTGTATCTTTCGCAAT<u>CTA</u>CTTGAGGTTCA
AnaCAX_T25stop_Fwd	GCACTAAGG<u>TAG</u>GTGCTGGCACAGCTAGGC
AnaCAX_T25stop_RevPhos	CTGATAAGTCAGGTAGCTTTG
AnaCAX_A28stop_Fwd	GACTGTGCTG<u>TAG</u>CAGCTAGGCGAAAC
AnaCAX_A28stop_RevPhos	CTTAGTGCCTGATAAGTCAGGTAG
AnaCAX_T33stop_Fwd	CACAGCTAGGCGAA<u>TAG</u>AATCCACCGTTG
AnaCAX_T33stop_RevPhos	CCAGCACAGTCCTTAGTGCCTGATAAGTCAGG
AnaCAX_H42stop_Fwd	GCACTTTGGCTG<u>TAG</u>AACTTTTCTGC
AnaCAX_H42stop_RevPhos	CAACGGTGGATTTGTTTCGCCTAG
AnaCAX_K54stop_Fwd	GATGGCGAA<u>TAG</u>TACATCGAAGAAC
AnaCAX_K54stop_RevPhos	CTGGACTTTCCCGGCAGAAAAGTTATG
AnaCAX_L105stop_Fwd	CCGCCAGCAT<u>TAG</u>GAACGCATGACG
AnaCAX_L105stop_RevPhos	CGTTTTTCCATGTTGGCTTGCTGAATAC
7120GroEL2/Fwd_NdeI	AGAAAGGA**CAT*ATG*G**CAAAAATTATTTC
7120GroEL2/Rev_BamHI	CTCCCTGC**GGATCC***TTA*GAACATACCCATAC
7120GroEL1/Fwd_NdeI	TTTAGACA**CAT*ATG*G**CAAAGCGCATTATCTACAAC

Oligo name	Primer sequence 5´-3´
7120GroEL1/Rev_BamHI	GTGACTAC**GGATCC***TTA*GTAATCGAAGTCACCGC
7120GroES/Fwd_NdeI	GAAATTAA**CAT***ATG*GCAGCAGTATCTCTAAGC
7120GroES/Rev_BamHI	TACTTAT**GGATCC***TTA*GCCAACGACTGCTAGAATGTC
7120EL2_Fwdmid	CTAAGGTCGCCAAGCCTGTAGAAGGTAGTGC
7120EL1_Fwdmid	GCTCGTCCTGTAGAAGATTCTAAATCTATTGCTCAAGTTGG
T7_Promotor	TAATACGACTCACTATAGGG
T7_Rev_Term	GCTAGTTATTGCTCAGCGG
6301EL1_Fwd_NdeI	GAGATATA**CAT***ATG*GCTAAACGGATCATTTACAACG
6301EL1_Rev_BamHI	TGGTGGTG**GGATCC***TTA*GTAGTCGAAGTCGCCCATG
6301ES_Fwd_NdeI	GAGATATA**CAT***ATG*GCAGCTGTATCTCTGAGTGTTTCG
6301ES_Rev_BamHI	TGGTGGTG**GGATCC***CTA*GGCAACAACGGCCAAGATG
6301L_A126Y_Fwd QC	GTGTTTGGCTTCAAA<u>TAT</u>ATCCGTTCGCTG
6301L_A126Y_Rev QC	CAGCGAACGGAT<u>ATA</u>TTTGAAGCCAAACAC
6301L_A126T_Fwd QC	GTGTTTGGCTTCAAA<u>ACT</u>ATCCGTTCGCTG
6301L_A126T_Rev QC	CAGCGAACGGAT<u>AGT</u>TTTGAAGCCAAACAC
6301L_A53T_Fwd QC	GAAGCTGGTGCG<u>AC</u>GATCGCGGCTGAATCTTCGACCG
6301L_A53T_Rev QC	CGGTCGAAGATTCAGCCGCGATC<u>GT</u>CGCACCAGCTTC
6301L_A53H_Fwd QC	GAAGCTGGTGCG<u>CA</u>TATCGCGGCTGAATCTTCGACCG
6301L_A53H_Rev QC	CGGTCGAAGATTCAGCCGCGAT<u>ATG</u>CGCACCAGCTTC
6301L_E49C_Fwd QC	GTGTCCCTGCTGAC<u>TGT</u>GCTGGTGCGG
6301L_E49C_Rev QC	CCGCACCAGC<u>ACA</u>GTCAGCAGGGACAC
AnaX_T33C_Fwd QC	GCACAGCTAGGCGAA<u>TGT</u>AATCCACCGTTG
AnaX_T33C_Rev QC	CAACGGTGGATT<u>ACA</u>TTCGCCTAGCTGTGC
AnaX_E83C_Fwd QC	CTGAAGAAATCGCT<u>TGT</u>TTCCTACCAGAAATG
AnaX_E83C_Rev QC	CATTTCTGGTAGGAA<u>ACA</u>AGCGATTTCTTCAG
7002X_T33C_Fwd QC	CAGCTTAGTGAA<u>TGC</u>AATCCTGGACAGGC
7002X_T33C_Rev QC	GCCTGTCCAGGATT<u>GCA</u>TTCACTAAGCTG
7002X_E84C_Fwd QC	GAAGGAGTTCTGT<u>GT</u>TTTTTGCCAGAAATG
7002X_E84C_Rev QC	CATTTCTGGCAAAAA<u>ACA</u>CAGAACTCCTTC
6301X_T51C_Fwd QC	GATCAATTGGCTGAG<u>TGT</u>AATCCGGCCG
6301X_T51C_Rev QC	CGGCCGGATT<u>ACA</u>CTCAGCCAATTGATC
6301X_D102C_Fwd QC	GAGCATATTCTC<u>TGC</u>TACCTGCCG
6301X_D102C_Rev QC	CGGCAGGTA<u>GCA</u>GAGAATATGCTC
6301L_E106Q_Fwd QC	CTCGACCTGTTT<u>CAA</u>GAAGGGTCGGTCAC
6301L_E106Q_Rev QC	GTGACCGACCCTTC<u>TTG</u>AAACAGGTCGAG
7002X(G42H,E43N)_Fwd QC	CGATTTGGCTA<u>CATAAT</u>TTCTCTAAACGTCATC
7002X(G42H,E43N)_Rev QC	GATGACGTTTAGAGAA<u>ATTATG</u>TAGCCAAATCG
6301L_R212S_Fwd QC	GTTCCAACGCTGGT<u>CC</u>GATCGCTTCCTG
6301L_R212S_Rev QC	CAGGAAGCGATC<u>GG</u>ACCAGCGTTGGAAC
6301L_K180D_Fwd QC	GTCTGTCGGCG<u>GAT</u>AACTACGGTCGTG
6301L_K180D_Rev QC	CACGACCGTAGTT<u>ATC</u>CGCCGACAGAC
6301L_D103A_Fwd QC	GCTTACCCGCTCG<u>GCC</u>CTGTTTGAAGAAG
6301L_D103A_Rev QC	CTTCTTCAAACAG<u>GGC</u>GAGCGGGTAAGC
6301L_E49A_Fwd QC	GTGTCCCTGCTGAC<u>GCT</u>GCTGGTGCG

Oligo name	Primer sequence 5´-3´
6301L_E49A_Rev QC	CGCACCAGC<u>AGC</u>GTCAGCAGGGACAC
7002L_E48C_Fwd QC	CTGGAGTCCCCCCCGAAT<u>GTT</u>GTGCTGCGGCTGTTGC
7002L_E48C_Rev QC	GCAACAGCCGCAGCACA<u>AC</u>ATTCGGGGGGGGACTCCAG
6301LΔN(1-14)NdeI_Fwd	ATAAGGCC**CAT*ATG***AAGGACTACAAACTCACCTATTACACCC
6301LΔN(1-14)BamHI_Rev	TAGCAGCC**GGATCC***TTA*GAGCTTGTCCATCGTTTCGAATTC
6301L(W67A,L70G,L71A)_FwdQC	GACCACCGTG<u>GCT</u>ACCGAC<u>GGTGCT</u>ACCGACATGGATCG
6301L(W67A,L70G,L71A)_RevQC	CGATCCATGTCGGT<u>AGCACC</u>GTCGGT<u>AGC</u>CACGGTGGTC
6301S(L60A,F63A)_FwdQC	GATGTGGAAG<u>GCT</u>CCCCTGG<u>CT</u>GACTGCAAGAG
6301S(L60A,F63A)_RevQC	CTCTTGCAGTC<u>AGC</u>CAGGGG<u>AGC</u>CTTCCACATC
6301LΔN(1-11)NdeI_Fwd	CCGCAGGC**CAT*ATG***GCCGGGGTGAAGGACTACAAACTCAC
6301L(E49A,A53H)_FwdQC	<u>CT</u>GCTGGTGCG<u>CA</u>TATCGCGGCTGAATCTTC
6301L_(E49A,A53H)_RevQC	GAAGATTCAGCCGCGAT<u>ATG</u>CGCACCAGC<u>AG</u>
6301LΔN(1-5)NdeI_Fwd	CCGCAGGC**CAT*ATG***TCTGCCGCAGGCTATAAGGCCGG

Table 7-2 Sequences of oligonucleotides.
Introduced endonuclease restriction sites are shown in bold, start or stop codons are in italic and mutations are underlined.
QC: primers for quick change mutagenesis

7.3   Protein accession numbers

Protein	Species	Accession number
GroEL	*Escherichia coli*	YP_859815
GroES	*Escherichia coli*	NP_418566
At-Cpn60α1	*Arabidopsis thaliana*	BAD94998
At-Cpn60α2	*Arabidopsis thaliana*	AAP68223
At-Cpn60β1	*Arabidopsis thaliana*	BAB11583
At-Cpn60β2	*Arabidopsis thaliana*	BAB01754
At-Cpn60β3	*Arabidopsis thaliana*	BAH19596
At-Cpn60β4	*Arabidopsis thaliana*	BAD95277
At-Cpn20	*Arabidopsis thaliana*	O65282
At-Cpn10	*Arabidopsis thaliana*	O80504
At-RbcXI$_2$	*Arabidopsis thaliana*	AAO50682
At-RbcXII$_2$	*Arabidopsis thaliana*	AAN13142
At-RbcS	*Arabidopsis thaliana*	AEE34594
At-RbcL	*Arabidopsis thaliana*	AAB68400
Cr-RbcS2	*Chlamydomonas reinhardtii*	RBS2_CHLRE
Cr-RbcL	*Chlamydomonas reinhardtii*	NP_958405
Ana7120-GroEL1	*Anabaena* sp. PCC7120	NP_487702
Ana7120-GroEL2	*Anabaena* sp. PCC7120	NP_485936
Ana7120-GroES	*Anabaena* sp. PCC7120	NP_487701
Syn6301-GroEL1	*Synechococcus* sp. PCC6301	YP_401330
Syn6301-GroES	*Synechococcus* sp. PCC6301	YP_172498
Syn6301-RbcX$_2$	*Synechococcus* sp. PCC6301	BAD80711

Protein	Species	Accession number
Syn6301-RbcS	*Synechococcus* sp. PCC6301	BAD78319
Syn6301-RbcL	*Synechococcus* sp. PCC6301	BAD78320
Syn7002-RbcX$_2$	*Synechococcus* sp. PCC7002	YP_001735041
Syn7002-RbcS	*Synechococcus* sp. PCC6301	YP_001735040
AnaCA-RbcX$_2$	*Anabaena* sp. CA	AAA63603
Rr-RbcL	*Rhodospirillum rubrum*	Q2RRP5
Zm-RbcXI$_2$	*Zea mays*	NP_001144731.1
Zm-RbcXII$_2$	*Zea mays*	LOC100277780
Zm-RbcXI2$_2$	*Zea mays*	LOC100275464

Table 7-3. Protein accession numbers.

7.4 Abbreviations

Å	Ångström
AA	Acrylamide
aa	Amino acid
ADP	Adenosine 5'-diphosphate
AMP	Adenosine 5'-monophosphate
Amp	Ampicillin
AnaCA	*Anabaena*
APS	Ammonium peroxydisulfate
At	*Arabidopsis thaliana*
ATP	Adenosine 5'-triphosphate
BLAST	Basic Local Alignment Search Tool
bp	Base pairs
BSA	Bovine serum albumin
C-	C-terminal
°C	Degree Celsius
^{14}C	Carbon-14
CA1P	2-carboxy-D-arabinitol-1-phosphate
CABP	Carboxy-arabinitol-1,5-bisphosphate
Cam	Chloramphenicol
CBB	Calvin-Benson-Bassham
CD	Circular dichroism
cDNA	Copy DNA
ch	Chloroplast
Cpn	Chaperonin
Cr	*Chlamydomonas reinhardtii*
CRBP	Carboxy-ribitol-1,5-bisphosphate
CV	Column volume
Cyan	Cyanobacteria
Δ (delta)	Deletion
DNA	Deoxyribonucleic acid
DnaJ	Bacterial Hsp40 chaperone
DnaK	Bacterial Hsp70 chaperone
DTT	Dithiothreitol
E. coli	*Escherichia coli*

EDTA	Ethylenediaminetetraacetic acid
ESI	Electrospray ionization
FPLC	Fast protein liquid chromatography
g	Acceleration of gravity, 9.81 m/s^2
gDNA	Genomic DNA
G3P	Glyceraldehyde-3-phosphate
GdnHCL	Guanidinium hydrochloride
GroEL	Bacterial Hsp60 chaperonin
GroES	Bacterial Hsp10 cochaperone
GrpE	Bacterial nucleotide exchange factor of DnaK
GSH	Reduced glutathione
GSSG	Oxidized glutathione
hr	Hour
HEPES	N-(2-hydroxyethyl)piperayin-N'-2-ethanesulfonic acid
6His	6 histidine tag
HRP	Horseradish peroxidase
Hsp	Heat shock protein
IPTG	Isopropyl-β-D-1thiogalactopyranoside
ITC	Isothermal titration calorimetry
Kan	Kanamycin
k^C_{cat}	CO_2-saturated carboxylase velocity
K_C	Michaelis-Menten constant for carboxylation
K_D	Dissociation constant
kDa	Kilodalton
k^O_{cat}	O_2-saturated oxygenase velocity
K_O	Michaelis-Menten constant for oxygenation
LB	Luria Bertani
mA	Milliampere
MALS	Multiangle light scattering
min	Minute
MOPS	3-(N-morpholino)propanesulfonic acid
MS	Mass spectrometry
MW	Molecular weight
MWCO	Molecular weight cut off
N-	N-terminal
NAC	Nascent chain-associated complex
NADPH	β-nicotinamide adenine dinucleotide 2'-phosphate
NTA	Nitrilotriacetic acid
OD	Optical density
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PDB	Protein data bank
PFD	Prefoldin
Pfu	*Pyrococcus furiosus*
3-PGA	3-phosphoglycerate
PMSF	Phenylmethylsulfonyl fluoride
PPG	2-phosphoglycolate
PVDF	Polyvinylidene fluoride
RbcL	Rubisco large subunit
RbcS	Rubisco small subunit
RbcX	Rubisco assembly chaperone

RNA	Ribonucleic acid
rpm	Revolutions per minute
Rr	*Rhodospirillum rubrum*
RT	Room temperature
RTS	Rapid transcription/translation system
Rubisco	Ribulose-1,5-bisphosphate carboxylase/oxygenase
RuBP	Ribulose-1,5-bisphosphate
s	Second
^{35}S	Sulfur-35
SAP	Shrimp alkaline phosphatase
SDS	Sodium dodecyl sulfate
SEC	Size exclusion chromatography
So	*Spinacia oleracea*
sp.	Species
Syn	*Synechococcus*
TCA	Trichloroacetic acid
TEMED	N,N,N',N'-tetramethylethylenediamine
TF	Trigger factor
TRiC	Tailless complex peptide 1 (TCP1) ring complex
Tris	Tris(hydroxymethyl)aminomethane
tRNA	Transfer RNA
UV	Ultraviolet
V	Volts
v/v	Volume per volume
wt	Wildtype
w/v	Weight per volume
XUBP	Xylulose 1,5-bisphosphate

7.5   Publications

Bracher A*, **Starling-Windhof A***, Hartl FU and Hayer-Hartl M (2011) Crystal structure of a chaperone-bound assembly intermediate of form I Rubisco. **Nature Structural and Molecular Biology** 18: 875-880

Starling-Windhof A, Massaro M and Briskie JV (2011) Differential effects of exotic predator-control on nest success of native and introduced birds in New Zealand. **Biological Invasions** 13: 1021-1028

Liu C, Young A, **Starling-Windhof A**, Bracher A, Saschenbrecker S, Vasudeva Rao B, Vasudeva Rao K, Berninghausen O, Mielke T, Hartl FU, Beckmann R and Hayer-Hartl M (2010) Coupled chaperone action in folding and assembly of hexadecameric Rubisco. **Nature** 463: 197-202

Massero M, **Starling-Windhof A**, Briskie JV and Martin TE (2008) Introduced mammalian predators induce behavioural changes in parental care in an endemic New Zealand bird. **PloS ONE** 3: 1-7

*authors contributed equally

7.6   Curriculum vitae

Amanda Windhof
Citizenship: American

Education

2008-2011   **Ludwig Maximilian University of Munich**, Munich, Germany
                Max Planck Institute of Biochemistry: Department of Cellular Biochemistry
                Supervisors: Prof. Dr. F.U. Hartl and Dr. M. Hayer-Hartl
                PhD in Chemistry
                "Rubisco folding and oligomeric assembly: detailed analysis of an assembly intermediate"

2004-2006   **University of Canterbury**, Christchurch, New Zealand
                School of Biological Sciences
                Supervisor: Dr. James V. Briskie
                Master of Science in Zoology: First Class Honours
                "Behavioural plasticity of life history traits in the New Zealand avifauna"

1999-2003   **Wittenberg University**, Springfield, Ohio
                Bachelor of Science in Biology, Minor Management

1995-1999   **Rutherford B. Hayes High School**, Delaware, Ohio
                High School Diploma: Valedictorian

Research and Professional Experience

2008-2011   Doctoral candidate, Max Planck Institute of Biochemistry: Department of Cellular
                Biochemistry, Munich, Germany

2008   Editing, Max Planck Institute of Ornithology, Seewiesen, Germany

2007-2008   Technical Assistant, Max Planck Institute of Biochemistry: Department of Cellular
                Biochemistry, Munich, Germany

2004-2006   Master of Science field research, University of Canterbury, Christchurch, New Zealand

2003   Assistant Animal Keeper, The Columbus Zoo and Aquarium, Powell, Ohio

2001   Personal Care Aide, Sarah Moore Nursing Home, Delaware, Ohio

Munich, 01.09.2011

Amanda Windhof

Amanda Windhof