Structure and function of cyanobacterial Rubisco activase from *Nostoc* sp. PCC 7120

Mirkko Flecken

Aachen Dezember 2021

Dissertation zur Erlangung des Doktorgrades
der Fakultät für Chemie und Pharmazie
der Ludwig-Maximilians-Universität München

Structure and function of cyanobacterial Rubisco activase from *Nostoc* sp. PCC 7120

Mirkko Flecken

aus

Bonn, Deutschland

2021

Eidesstattliche Versicherung

Erklärung

Diese Dissertation wurde im Sinne von §7 der Promotionsordnung vom 28. November 2011 von Herrn *Prof. Dr. F. Ulrich Hartl* betreut.

Eidesstattliche Versicherung

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

Aachen, 16. Dezember 2021

Mirkko Flecken

Dissertation eingereicht am:	20.12.2021
1. Gutachter:	Prof. Dr. F. Ulrich Hartl
2. Gutachter:	Prof. Dr. Dario Leister
Mündliche Prüfung am:	04.02.2022

Acknowledgments

Foremost, I want to express my gratitude to Prof. Dr. F. Ulrich Hartl and Dr. Manajit Hayer-Hartl for the opportunity to work in the Department of Cellular Biochemistry at the MPIB under their supervision. I sincerely thank my collaborators Dr. Andreas Bracher and Dr. Huping Wang for their valuable contribution to this project and helpful discussions. I am especially grateful to Dr. Andreas Bracher, Dr. Gabriel Thieulin-Pardo and Dr. Robert Wilson for good advice and guidance during my PhD research.

Moreover, I want to thank all members of the Rubisco office including Dr. Gabriel Thieulin-Pardo, Dr. Robert Wilson, Dr. Harald Aigner, Dr. Kun Zhang, Dr. Huping Wang, Lorenzo Calisse and Alonso Carvajal Alvarez for a fun and supportive working atmosphere and valuable discussions. Working with all of you made my first steps in the Rubisco / Rubisco activase field and my start in Martinsried easy. Lorenzo, thanks for much to rare climbing and ski touring adventures and your italian verve and enthusiasm, and Alonso, thanks for many inspiring conversations about Open Science, R programming and for introducing me to some of the few pieces of Latin-American culture that the Munich cultural landscape has to offer.

I cordially thank the staff members of the lab: Albert Ries, Nadine Wischnewski, Syliva Gärtner, Romy Lange, Massimo Bossi, Emmanuel Burghardt and Darija Pompino for their support and perpetual efforts to keep the lab running. I am grateful for the supportive and collaborative environment in the Hartl department and beyond. I am especially grateful for the outstanding research infrastructure, the know-how and collaborative atmosphere provided at the MPIB. I want to thank Dr. Daniel Bollschweiler and Dr. Tillman Schäfer from the CryoEM Facility, and Dr. Stephan Uebel from the Biochemistry Core Facility for competent support and advice. I am grateful for technical support by the Crystallization Facility and the Immunization Service. Moreover, I want to sincerely thank Fabien Bonneau and Dr. Claire Basquin (Department of Structural Cell Biology, MPIB) for their generous help in finding a suitable buffer for β-cyanobacterial Rubisco activase and Dr. Stefan Pettera for peptide synthesis. I want to thank the members of my PhD committee Prof. Dr. Dario Leister, PD Dr. Bettina Bölter, Prof. Dr. Klaus Förstemann, Prof. Dr. Martin Biel and Prof. Dr. Karl-Peter Hopfner for critically reading and evaluating this thesis.

Last but certainly not least I want to thank my friends and family for their sympathy, confidence and continuous support. I will always be deeply indebted to my own little family for their trust, understanding, patience and love.

Abstract

The great majority of all organic carbon within the Earth's biogeosphere - including fossil resources - has once been sequestered by the action of the key enzyme of the Calvin-Benson-Bassham cycle, Rubisco. Rubisco is scattered across the tree of life, enabling diverse organisms to grow autotrophically. Form IB Rubisco of plants, algae and cyanobacteria is a tetramer of antiparallel Rubisco large subunit (RbcL) dimers that associates with eight Rubisco small subunits (RbcS) to form a hetero-hexadecameric complex of $\sim$ 520 kDa. Each dimer features two active sites formed at the interface of the C-terminal domain of one and the N-terminal domain of the other subunit. RbcL requires chaperonin assistance for folding and their association in the holoenzyme is aided by additional assembly chaperones. Moreover, Rubisco is prone to inactivation by inhibitory sugar phosphates and depends on an additional chaperone, the hexameric AAA+ ATPase Rubisco activase (Rca), for conformational repair in a process which is largely enigmatic.

Here, we focus on Rca and Rubisco from the cyanobacterium *Nostoc* sp. PCC 7120 and seek to investigate Rubisco reactivation in cyanobacteria as an example of form IB Rubisco reactivation. Cyanobacterial Rca lacks the N-domain typically found in the well studied plant Rca, but features a C-terminal extension which shares sequence similarity with RbcS and the small subunit-like (SSUL) domains of CcmM. CcmM has been shown to bind Rubisco via its SSUL domains and induces the formation of Rubisco condensates with liquid-like properties during the formation of carboxysomes, proteinaceous compartments that accommodate catalytically active Rubisco in cyanobacteria.

The recombinant cyanobacterial Rubisco, generated by co-expression of endogenous chaperonins, was not susceptible to inhibition by RuBP nor XuBP, but was efficiently inhibited by the artificial inhibitor CABP. In biochemical *in vitro* assays, we found that recombinant cyanobacterial Rca was able to efficiently reactivate CABP-inhibited Rubisco without requiring the SSUL domain. By cryo-EM, we obtained a structural snapshot of a post-remodeling state of Rubisco, with Rca still bound to the active site after inhibitor release. The RbcL N-terminal peptide was engaged in interactions with pore loops located in the central solvent channel of the Rca hexamer, which formed small cavities to accommodate the alternating small side chains of the RbcL N-terminal peptide. Our structural analysis indicated, that Rca remodels Rubisco by a combination of pulling the N-terminus of one and displacing

the C-terminus of the other RbcL within the antiparallel dimer where the active site pocket is formed. In addition to engaging the RbcL N-terminal peptide, we identified three additional interaction sites on the surface of Rubisco. Mutational analysis of the corresponding residues in Rca, rendered Rca reactivation incompetent, demonstrating their importance for Rubisco reactivation. Likewise, Rubisco mutants of *Nostoc* sp. PCC 7120, Arabidopsis and Tobacco lacking the RbcL N-terminal peptide, could no longer be reactivated by their cognate Rca, indicating that the engagement of the RbcL N-terminal peptide is an essential, conserved feature of form IB Rubisco reactivation. Moreover, we found that the SSUL domain of cyanobacterial Rca establishes the formation of liquid-like condensates in presence of Rubisco. These condensates are able to undergo fusion events, and as in CcmM, Rubisco condensate formation is modulated by the redox state of two cysteines located in the SSUL domain. As analyzed by biochemical competition experiments, mutational analysis and cryo-EM, the SSUL modules of CcmM and Rca bind Rubisco at the dimer-dimer cleft close to the equator of Rubisco. Our results are in accordance with a model wherein the reduced SSUL domain mediates the co-encapsulation of Rubisco and Rca during carboxysome formation to ensure their co-localization in the mature, oxidizing carboxysome where Rca is released and enabled to fulfill its reactivation function.

Contents

List of Figures

List of Tables

Abbreviations

Table 0.1.: List of Abbreviations

Abbreviation	Explanation
2PG	2-phosphoglycolate
3PGA	3-phosphoglycerate
AAA+	ATPase associated with various cellular activities
ASCE subgroup	Additional strand, conserved glutamate subgroup (of AAA+ proteins)
At- (prefix)	Gene or protein from *Arabidopsis thaliana*
ATP	Adenosine triphosphate
ATPase	ATP hydrolyzing enzyme
AU	Arbitrary units
BSD2	Bundle sheath defective 2
CA	Carbonic anhydrase
CA1P	2-carboxy-D-arabinitol-1-phosphate
CABP	2-carboxyarabinitol-1,5-bisphosphate
CAM	Crassulacean acid metabolism
CBB	Calvin-Benson-Bassham
CCM	Carbon Concentrating Mechanism
CES	Control by epistasy of synthesis
CHAPS	3-[(3-Cholamidopropyl)dimethylammonio]-1-Propanesulfonate
CTBP	2-carboxytetritol-1,4-bisphosphate
DHAP	Dihydroxyacetone
DTT	Dithiothreitol
E	Decarbamylated, inactive Rubisco
E.RuBP	Decarbamylated, inactive Rubisco, inhibited by RuBP
E.XuBP	Decarbamylated, inactive Rubisco, inhibited by XuBP
ECM	Carbamylated, active Rubisco
ECM.CABP	Carbamylated Rubisco, inhibited by CABP
EDTA	Ethylenediaminetetraacetic acid
ER	Endoplasmic reticulum
FRAP	Fluorescence recovery after photobleaching

...continues on next page...

Abbreviation	Explanation
G3P	Glyceraldehyde-3-phosphate
GOE	Great oxidation event
H_6ubi	His_6-ubiquitin
h2i	Helix-2-insert
HGT	Horizontal gene transfer
Hsp	Heat shock factor
IPTG	β-d-1-thiogalactopyranoside
ISP	Inhibitory sugar phosphate
ISS	Inter-subunit signaling
KABP	3-ketoarabinitol-1,5-bisphosphate
k_{cat}	Catalytic rate
L_2	RbcL dimer
L_8	Octameric RbcL core
L_8S_8	Hexadecameric Rubisco holoenzyme
M35	Short isoform of CcmM, $\sim$ 35 kDa
MIDAS	Metal-ion-dependent adhesion site
MPIB	Max Planck Institute of Biochemistry
NAC	Nascent-chain-associated complex
NanA	N-acetylneuraminic acid aldolase
NCL	Nucleotide communication loop
Nos- (prefix)	Gene or protein from *Nostoc* sp. PCC 7120
Nt- (prefix)	Gene or protein from *Nicotiana tabacum*
NTP	Nucleoside triphosphate
OEC	Oxygen evolving complex
P-loop NTPase	Pore loop type, nucleoside triphosphate binding and hydrolyzing ATPase
P680	Pigment with absorption maximum at 680 nm
P700	Pigment with absorption maximum at 700 nm
PAGE	Polyacrylamide gel electrophoresis
PDB	Accession / identifier according to the protein data base
PDBP	D-glycero-2,3-pentodiulose-1,5-bisphosphate
PEP	Phosphoenolpyruvate
PL	Pore loop
ps1	Pre-sensor 1
ps1βh	Pre-sensor 1 β hairpin

...continues on next page...

Abbreviation	Explanation
PS-I / II	Photosystem I / II
RAC	Ribosome-associated complex
Raf1	Rubisco accumlation factor 1
RbcL	Rubisco large subunit
RbcS	Rubisco small subunit
RC-I / II	Reactioncenter I / II
Rca	Rubisco activase
Rca-α	Long isoform of Rca from plants
Rca-β	Short isoform of Rca from plants
RLP	Rubisco-like protein
RLSMT	Rubisco large subunit methyltransferase
Rubisco	Ribulose-1,5-diphosphate carboxylase/oxygenase
RuBP	Ribulose-1,5-bisphosphate
$S_{C/O}$	Specificity of Rubisco; a measure for the prevalance of the carboxylation reaction over the oxygenation reaction
Se- (prefix)	Gene or protein from *Synechococcus elongatus* PCC 7942
SEC	Size exclusion chromatography
Sm- (prefix)	Gene or protein from *Serratia marcescens*
SRH	Second region of homology
TF	Trigger factor
TIM	Triose-phosphate isomerase
VCP	Valosin-containing protein
VWA	Von Willebrand factor Type A
XuBP	D-xylulose-1,5-bisphosphate
α-*At*RbcL	Antibody directed against Rubisco from *Arabidopsis thaliana*
α-*Nos*RbcS	Antibody directed against RbcS from *Nostoc* sp. PCC 7120
α-*Se*RbcL	Antibody directed against RbcL from *S. Elongatus* PCC 6301
α-*Nt*RbcL	Antibody directed against Rubisco from *Nicotiana tabacum*

1. Introduction

1.1. Protein folding

Almost all cellular processes depend on proteins, abundant cellular macromolecules essential for life, which perform a wide range of functions as enzymes, structural elements, regulators or messengers. The cellular concentrations of specific proteins vary over several orders of magnitude and are tightly regulated by their rate of synthesis and proteolytic degradation. Proteins are composed of linear chains of amino acids that are linked by peptide bonds. The sequence of amino acids, the primary structure of a protein, is encoded by the nucleotide sequence of the corresponding gene, a piece of deoxyribonucleic acid (DNA) essentially comprising the same information in inheritable form. Segments in the polypeptide chain of most proteins form secondary structures, such as α-helices and β-sheets, by local interactions of backbone atoms. Secondary structure elements frequently form simple reoccurring motifs that associate to compact domains. The distinct three dimensional conformation of a (multidomain) protein is referred to as the proteins tertiary structure. If multiple polypeptide chains associate in a protein complex, their arrangement is referred to as quarternary structure. How much of the information required to adopt a proteins three dimensional conformation is contained solely in its primary sequence, and to what extent the cellular environment shapes its conformation, is a central biological question and still under active investigation [Berg et al., 2018].

1.1.1. Protein folding *in vitro*

Upon synthesis on ribosomes as linear chains of amino acids, the majority of proteins have to fold into a unique three-dimensional structure in order to acquire functional activity. Folding is governed by the formation of mainly non-covalent interactions, like hydrophobic interactions and electrostatic interactions (including ionic interactions, hydrogen bonds and van der Waals forces) and sometimes by covalent bonds (disulphides) or the coordination of metals. Due to the high entropic costs of an extended, linear polypeptide chain for the formation of a bulk water hydrogen-bond network, the formation of intramolecular interactions occurs spontaneously. Thus folding proceeds from a high free-energy state to a thermodynamically more favorable state of lower free-energy. This “collapse” of the polypeptide chain is thought

to be primarily driven by the formation of hydrophobic interactions between nonpolar amino acid sidechains and the concomitant entropy gain in aqueous, hydrophilic solvents of biological systems. The need to adopt a thermodynamically stable state, while maintaining the flexibility most proteins require to carry out enzymatic reactions, manifests in an evolutionary tradeoff between stability and functionality [Foit et al., 2009; Kim et al., 1998; Mayer, 2010]. The result of this tradeoff is an intrinsic susceptibility of proteins to misfold and aggregate, as they are only marginally stable at physiological conditions, which is reflected by the narrow operational temperature range of enzymes.

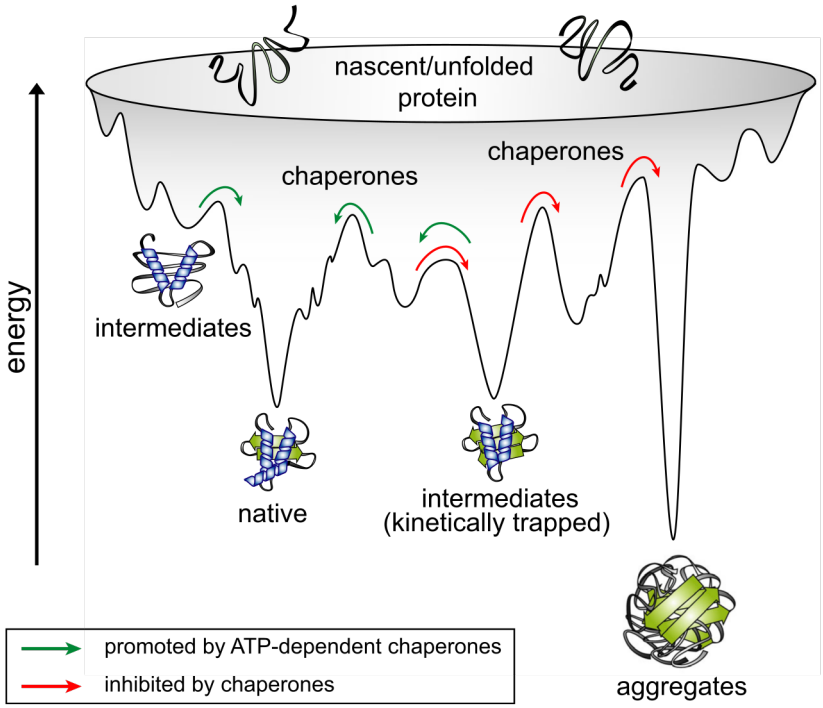

Figure 1.1.: The energy landscape of protein folding shaped by chaperones. Folding proteins are thought to navigate a multidimensional, funnel-shaped, free-energy surface towards a state of lower energy. Meta-stable intermediates or aggregates form when proteins encounter minima, that cause folding to slow down. Molecular chaperones prevent aggregation, resolve kinetically trapped conformations and assist the folding by lowering free-energy barriers. The native conformation is the conformation that primarily persists under the physiological conditions for which it was evolutionary selected. Figure reproduced from [Balchin et al., 2020].

How does a protein find its native, functional conformation? In their famous experiments Anfinsen et al. [1961] demonstrated that folding of bovine pancreatic ribonuclease in the test tube is reversible, suggesting that i) the native state of a protein is thermodynamically the most stable conformation and ii) that the primary sequence contains all necessary information to restore the three-dimensional state [Anfinsen, 1973]. Given the large number of possible conformations that a protein chain may adopt ($> 10^{30}$ for a 100-aa protein [Balchin et al., 2016]), folding is an intrinsically

error-prone process. Considering the enormous time consumption associated with the iterative sampling of all possible conformations, an accelerated, kinetically favored folding pathway must exist to explain the observed fast *in vitro* folding rates of proteins [Levinthal, 1969]. This folding pathway - a preferred sequence of events carrying the protein from the extended polypeptide state to the native functional conformation - implies, that the conformational space of the folding protein is restricted [Levinthal, 1968]. This concept of a folding pathway has been extended by Sali et al. [1994] to a folding landscape, which allows to account for the multitude of alternative conformations that a non-native polypeptide chain may adopt as it folds [Baldwin, 1994]. This folding landscape is often depicted as a multidimensional free-energy diagram as a function of selected parameters (bonding angles) that define the possible conformations of a protein. As the protein travels from a non-native to the native fold, an increasing number of favorable intramolecular interactions are established (concomitant with e.g. the formation of a hydrophobic core and increasing compactness of the molecule), that progressively narrow the available conformational space and thus accelerate folding.

En route to the native state, some folding proteins may enter a kinetically trapped, non-functional, intermediate folding state, where additional energy is required to achieve a state of higher free-energy by breaking previously formed, thermodynamically favorable intermolecular interactions, which allows the protein to return to a productive folding path. According to the concept of a free-energy landscape of protein folding, a "rugged" landscape reflects the high propensity of a protein to enter trapped intermediate conformations and may result in a low yield of the folding reaction. However, in the cellular environment folding is aided by other proteins known as molecular chaperones that either prevent the formation of favorable but unwanted interactions or actively break favorable interactions to resolve a trapped state (Figure **1.1**).

1.1.2. Protein folding *in vivo*

1.1.2.1. Co-translational folding

Inside the cell folding occurs co-translationally, as the nascent chain appears in a stepwise fashion on the ribosome - a fundamental difference to renaturation experiments, where folding is initiated with the full sequence available and long-range interactions are possible from the start. Importantly, ribosomes have been shown to modulate protein folding by altering the speed of translation (e.g. via specific codon

usage) [Zhou et al., 2013], restricting conformational space within the exit tunnel (for the first emerging 30-40 amino acids) [Kosolapov and Deutsch, 2009; O'Brien et al., 2011], providing the ribosomal surface to prevent misfolding of incomplete chains [Kaiser et al., 2011] and acting as hub for ribosome associated factors that aid the subsequent folding process (see below). In addition, folding is affected by the high concentration of macromolecules (300 - 400 $g\,L^{-1}$) *in vivo*, establishing excluded volume effects that favor intermolecular interactions [Ellis, 2001]. Due to the unstructured polypeptide backbone and exposed hydrophobic residues, which are usually buried in the core of a cytoplasmic protein, nascent chains are particularly prone to entangle with one another and form aggregates. Moreover, environmental stresses may alter the biochemical properties of the proteome on a large scale which eventually leads to partial unfolding and aggregation of less rigid proteins. Since some aggregates and soluble, non-native oligomeric species commonly expose hydrophobic surfaces and sometimes aggregation-prone unpaired β-strand structures, they are prone to drive other meta-stable proteins into aggregation and thus are often toxic to cells. For example, many neurodegenerative diseases, like Alzheimer's and Parkinson's disease, are associated with aggregate deposition [Ross and Poirier, 2004].

1.1.2.2. Molecular chaperones

In order to protect themselves from aberrant protein aggregation, cells in all domains of life prevent misfolding and dissolve aggregates to maintain protein homeostasis. To achieve the former, cells express molecular chaperones - proteins that aid the folding and/or assembly of another protein, without being part of its final structure [Hartl, 1996; Hartl et al., 2011]. The term "molecular chaperone" comprises both chaperones that are broadly involved in *de novo* protein folding and refolding as well as client specific chaperones which function downstream of protein folding in the assembly of oligomeric complexes [Hartl et al., 2011]. The former group comprises multiple conserved protein families which cooperate in a chaperone network of conserved organizational structure and guide a nascent chain from its first appearance at the ribosome to the native conformation, allowing it to fold to the native state at a biologically relevant (time)scale [Balchin et al., 2016]. Many members of this network are upregulated upon stress conditions (e.g. elevated temperature) and hence are referred to as stress proteins or heat shock proteins (Hsp) which are classified by molecular weight (e.g. Hsp60, Hsp70, Hsp90). Generally, members of these classes of molecular chaperones act as molecular machines that promote folding via ATP- and cofactor-regulated cycles of substrate binding and release [Hartl et al., 2011].

They transiently interact with sequences enriched in hydrophobic residues to prevent non-native interactions and concomitantly affect the conformational space of the folding protein chain, which may impact on its folding pathway. Smaller proteins are more likely to obtain their native conformation spontaneously in association with the ribosome whereas larger proteins have a greater tendency to require chaperone activity, because part of their sequence is not yet available for potentially required, long-range intermolecular interactions preventing productive folding.

The majority of nascent protein chains interact cotranslationally with ribosome-associated chaperones which delay chain compaction and prevent premature misfolding [Balchin et al., 2016]. Chaperones belonging to this class are trigger factor (TF) in bacteria and the structurally unrelated ribosome-associated complex (RAC) and nascent-chain-associated complex (NAC) in eukaryotes. TF interacts with the majority of proteins exceeding a length of $\sim$ 100 amino acids [Preissler and Deuerling, 2012]. The binding of TF to the nascent chain competes with the tendency of the bound polypeptide to bury hydrophobic segments in the course of translation. TF is likely released when the bound region collapses [Hoffmann et al., 2010]. The functionally related RAC and NAC of eukaryotes have additional functions, e.g. NAC has been shown to be important for subcellular protein sorting by reducing the affinity of ribosomes to the translocon within the endoplasmic reticulum (ER) membrane [Gamerdinger et al., 2015].

The Hsp70 system operates downstream of ribosome-associated chaperones and assists the folding of proteins that require further assistance to reach their native conformation. Hsp70 is an ATPase, and the bound nucleotide determines its affinity for substrate and its binding kinetics [Hartl, 1996; Luengo et al., 2019; Rosenzweig et al., 2019]. Hsp70 functionally cooperates with Hsp40 proteins and nucleotide exchange factors (DnaK/DnaJ/GrpE in bacteria, respectively), co-chaperones that control the ATP-hydrolysis activity of Hsp70, and thereby govern an iterative cycle of substrate binding and release. The Hsp70 system is thought to function by unfolding of trapped folding intermediates and by biasing the folding pathway to accelerate folding in comparison to spontaneous folding in absence of aggregation (reviewed in Balchin et al. [2020]). Next to their function in *de novo* protein folding, Hsp70s are involved in membrane translocation, disassembly of protein complexes and solubilization and refolding of aggregated proteins [Rosenzweig et al., 2019]. They also function as a hub and guide non-native proteins to the more specialized Hsp60 and Hsp90 systems [Balchin et al., 2016].

The mechanistic details of Hsp90 ATPases are less clear (reviewed in Jackson [2013] and Luengo et al. [2019]). The ATPase activity of eukaryotic Hsp90 is controlled by numerous co-chaperones. Eukaryotic Hsp90s receive their clients directly from the Hsp70 system or via adapter proteins. These molecular chaperones are thought to act on conformationally more mature protein folding intermediates, and their substrate pool is enriched for proteins participating in cell signaling, including kinases, steroid-hormone receptors and transcription factors [Morgner et al., 2015]. In contrast to the eukaryotic Hsp90, bacterial Hsp90 (HtpG) does not rely on co-chaperones for client recruiting or regulation [Balchin et al., 2016].

1.1.2.3. The chaperonins

Chaperonins are $\sim$ 60 kDa, ATP hydrolyzing proteins that form oligomeric, double-ring toroids of 0.8-1 MDa. Generally, the two chaperonin rings function as separate folding chambers for single, non-native substrate proteins to fold in isolation and unimpaired by aggregation [Hayer-Hartl et al., 2016]. In each kingdom, chaperonins are essential and can be divided into three groups. The group I chaperonins, GroEL, Hsp60 and Cpn60 are found in the cytosol of bacteria, in mitochondria and in chloroplasts, respectively, and function in cooperation with a co-chaperone (GroES, Hsp10, Cpn10, Cpn20). The group II chaperonins, including TRIC (also CCT) and the thermosome, which occur in the cytosol of eukaryotes and archaea, respectively, are composed of eight subunits per ring and do not co-operate with a co-chaperone like GroES. Instead, their apical domain is enlarged and contains a built-in lid which shields their central cavity from bulk solution [Lopez et al., 2015]. More recently, group III chaperonins have been discovered in deep branching bacterial species [Techtmann and Robb, 2010]. These are phylogenetically divergent from group I and II chaperonins, but resemble the hexadecameric structure of group II chaperonins consisting of a doublet of octameric rings and feature a built-in lid [An et al., 2016, 2017]. Group III chaperonins differ from group II chaperonins in the structure of the lid which is thought to affect the mechanism of inter-ring cooperation [An et al., 2017].

The archetypical bacterial chaperonin GroEL consists of $\sim$ 57 kDa subunits, each comprising an equatorial ATPase domain, an intermediate hinge domain and an apical domain that exposes hydrophobic residues for binding of non-native substrate [Boisvert et al., 1996] (Figure **1.2**). These subunits associate to form a cylindrical complex of two heptameric rings, which are stacked back-to-back in a staggered arrangement. The C-terminal, flexible extension of GroEL, a hydrophobic motif

of four Gly-Gly-Met repeats, protrudes into the central cavity. GroES, the 10 kDa cofactor of GroEL, forms a heptameric, lid-shaped ring and creates the closed folding chamber upon binding to the apical domain of GroEL [Xu et al., 1997]. In the apo state, ATP hydrolysis proceeds by positive cooperativity within the same GroEL ring and negative allostery toward the other ring, favoring an equilibrium between a low affinity state in one, and a high affinity state for ATP in the other ring. The high affinity of GroES to nucleotide-bound GroEL is thought to favor the formation of an asymmetric complex, with the heptameric GroES cofactor bound to one of the two GroEL rings only.

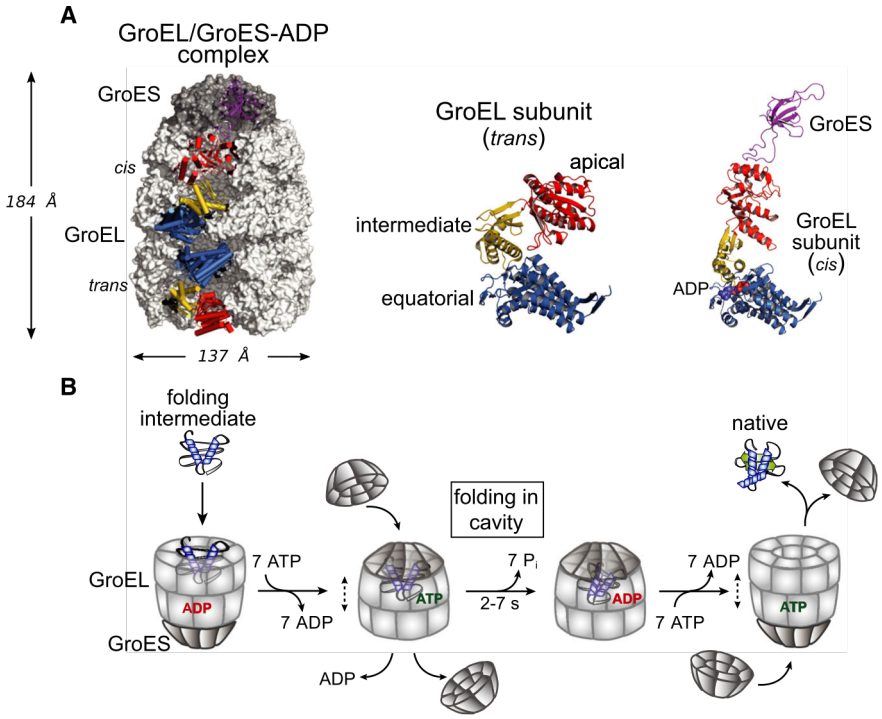

Figure 1.2.: Structure and reaction cycle of the GroEL/ES chaperonin.
A) Structure and organization of the GroEL/ES complex. Left, the asymmetrical, ADP-bound GroEL:GroES complex in space-filling representation. One subunit of each GroEL ring is shown in ribbon representation with the equatorial domain in blue, the intermediate domain in yellow and the apical domain in red. The interacting GroES subunit is shown in magenta. The open conformation of the trans subunit (middle) and the GroES-bound, closed conformation of the cis subunit (right) are highlighted. PDB: 1aon [Xu et al., 1997]. **B) The reaction cycle of GroEL/ES according to the asymmetric model.** The open ring of the GroEL-GroES-ADP complex binds the folding intermediate, which causes a conformational change in the GroEL apical domains that allows GroES to bind. Upon binding of GroES, the substrate is displaced in the folding cage ("*cis*-cavity") and GroES, ADP and the previously bound substrate dissociate from the opposite GroEL ring. The encapsulated substrate is allowed to fold until all ATP is hydrolyzed. Release of the properly folded substrate is induced by binding of ATP, new substrate, and GroES to the opposite GroEL ring. If the substrate has not reached its native fold, the cycle proceeds iteratively, until the substrate adopts the native conformation. From [Balchin et al., 2020] and [Balchin et al., 2016]. Reprinted with permission from AAAS.

A folding cycle is initiated by non-native substrate binding to the hydrophobic residues exposed by the apical domains of GroEL. Upon binding of ATP and GroES,

the substrate is displaced into the folding chamber, which is shielded from the cytosol by GroES. Concomitantly, conformational changes increase the volume of the substrate-encapsulating folding chamber (*cis* ring) and switch the inner surface properties from hydrophobic to hydrophilic [Xu et al., 1997]. Folding of the encapsulated substrate is allowed to proceed for the duration of ATP hydrolysis (7 ATPs, 2 - 7 s depending on temperature). Cytosolic proteins in their native conformation are less likely to interact with the chaperonin, because hydrophobic segments are buried in their core and shielded from solvent by inter- or intramolecular interactions. Proteins which failed to reach their native conformation within a single encapsulation event are prone to rebind to GroEL to continue with another folding cycle.

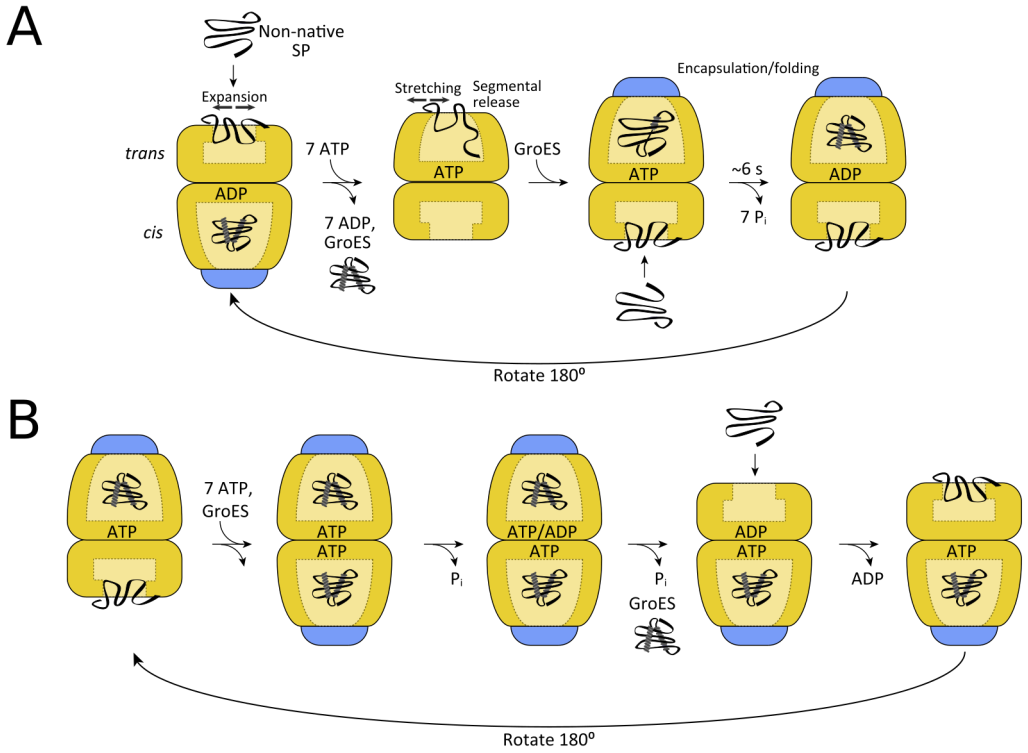

Figure 1.3.: Models of the GroEL/GroES interaction cycle.
The asymmetric (**A**) and symmetric (**B**) GroEL/ES interaction cycles in presence of substrate are shown. Conformational transitions of the substrate protein (SP) are indicated in **A**. Figure reproduced from [Hayer-Hartl et al., 2016] with permission.

Two models are used to describe the interdependence of the individual GroEL rings in the folding cycle of the GroEL/ES complex (Figure **1.3**) [Hayer-Hartl et al., 2016]. In the asymmetric GroEL/ES interaction cycle, folding reactions occur sequentially in each of the rings. ATP hydrolysis in the *cis*-ring induces conformational rearrangements of the same ring, which reduce the negative allostery toward the *trans*-ring. Upon binding of ATP and GroES to the *trans* ring a new *cis*-ring forms and ADP and GroES are released from the previous *cis* ring via an allosteric mechanism. This negative allosteric coupling of the rings, which has been shown to result in the tran-

sient separation of two GroEL rings in vitro [Yan et al., 2018], facilitates substrate release from the *trans* ring. According to the model of the symmetric GroEL/ES interaction cycle, folding reactions occur simultaneously in each of the GroEL rings, favoring the symmetric GroEL/ES complex with two heptameric GroES cofactors bound to either side of the GroEL double ring. Binding of the non-native substrate protein to the *trans*-ring is thought to accelerate the release of ADP from the same ring and renders it compatible for a new folding reaction by binding of ATP and GroES before ATP hydrolysis in the *cis*-ring is complete. According to the symmetric GroEL/ES interaction cycle, ADP and GroES release are uncoupled. Evidence for both models has been reported and is reviewed in [Hayer-Hartl et al., 2016].

In comparison to spontaneous folding in the absence of aggregation, GroEL-mediated folding has been shown to be accelerated 20 - 100 fold with several substrates [Georgescauld et al., 2014; Singh et al., 2020]. This acceleration has been attributed to re-modeling of mis-folded states, induced by movements of the apical domain of GroEL [Lin et al., 2008], the net negative charge of the inner surface of the *cis* cavity [Gupta et al., 2014], the relation between the *cis* cavity volume and the substrate's size and the flexible, hydrophobic C-terminus of GroEL [Tang et al., 2006]. The mechanism of folding by chaperonins is discussed in the context of three models - the Passive Cage Model, the Active Cage Model and the Iterative Annealing Model (reviewed in Hayer-Hartl et al. [2016]) -, that primarily distinguish between sole aggregation prevention by the chaperonin cage (Passive Cage), steric confinement by the cage (Active Cage) and the property of GroEL to provide recurrent option to fold productively by iterative unfolding of non-native protein, without the strict need for encapsulation (Iterative Annealing). Open questions of GroEL/ES mediated folding are the relative importance of symmetric (both rings have GroES bound) and asymmetric GroEL complexes (only the cis ring has GroES bound) or the role of the flexible C-terminal GroEL peptide [Balchin et al., 2016].

Approximately 5 % of the bacterial proteome and 10 % of the proteome of eukaryotes associate with the Hsp60 chaperonins and their eukaryotic homolog TRIC, respectively, to fold into their native conformation [Houry et al., 1999; Kerner et al., 2005; Yam et al., 2008]. In *Escherichia coli*, proteins containing a $\beta_8\alpha_8$ triose-phosphate isomerase (TIM)-barrel domain within their three-dimensional structure are enriched among the stringent chaperonin clients [Kerner et al., 2005]. Interestingly, the TIM barrel fold itself does not impose the requirement for assisted folding. Many TIM barrel domain proteins are not stringent chaperonin substrates, for example N-acetylneuraminic acid aldolase (NanA) from Mycoplasma synoviae, a bacterium lacking GroEL [Georgescauld et al., 2014]. Notably, the TIM barrel domain protein

Rubisco large subunit (RbcL) is an obligatory GroEL substrate [Brinker et al., 2001; Ellis and Van Der Vies, 1988; Goloubinoff et al., 1989a].

1.2. Photosynthesis

Photosynthesis refers to the light dependent sequestration of CO_2 into organic compounds. It encompasses two functionally cooperating modules - a light dependent electron transport chain converting sunlight into chemical energy and reducing potential, and a pathway for the sequestration and reduction of gaseous CO_2, allocated to the organism's metabolism. Oxygenic photosynthesis involves the oxidation of water to molecular O_2 and is always associated with the predominant pathway for carbon sequestration in the extant biosphere, the Calvin-Benson-Bassham (CBB) cycle [Fuchs, 2011]. Oxygenic photosynthesis developed in an ancestral cyanobacterium. The engulfment of such an organism by an ancestral heterotrophic eukaryotic cell via "endosymbiosis" lead to the development of chloroplasts and initiated the rise of the archaeplastida, including glaukophytes, rhodophyceae (red algae) and chloroplastida (includes green algae and land plants) [Adl et al., 2005; Raven et al., 2009]. Hence, oxygenic photosynthesis in extant algae and plants is of cyanobacterial origin. In contrast to oxygenic photosynthesis, photosynthesis in non-cyanobacterial phototrophic bacteria depends on organic or inorganic reductants other than water and thus is referred to as anoxygenic, as no O_2 is produced.

In the words of Fuchs [2011], "the fixation of inorganic carbon into organic material (autotrophy) is a prerequisite for life and sets the starting point of biological evolution". Global net primary production is estimated to $\sim 105 \times 10^{15}$ g carbon per year, with nearly equal contributions from marine (annual $\sim 50 \times 10^{15}$ g C) and terrestrial (annual $\sim 55 \times 10^{15}$ g C) systems [Field et al., 1998] and oxygenic photosynthesis contributes the bulk of it [Raven, 2009]. Hence, with regard to mass flow, oxygenic photosynthesis constitutes the most-important process in the biogeosphere [Bauwe et al., 2010]. Chlorophyll-based phototrophy (chlorophototrophy) most likely evolved at deep sea hydrothermal vents using the small fraction of visible light among the low-intensity, long-wavelength geothermal light and allowed chemoautotrophs to use other reductants than H_2. The subsequent evolution of anoxygenic photosynthesis predates the oxygenic photosynthesis and is thought to have evolved via the intermediate step of H_2S oxidation, freeing primary production from H_2 exhalation at vents. The advent of oxygenic photosynthesis made the ubiquitous H_2O usable

as reductant and stimulated the greatest ecological niche expansion in Earth's history [Martin et al., 2018]. According to Raven [2009], the transition from H_2-based chemosynthesis to anoxygenic photosynthesis stimulated global annual primary production by a factor of 2000 which was further increased by an additional factor of 30 upon transition to oxygenic photosynthesis. The concomitant rise of biogenic oxygen has been coined "the great oxidation event" (GOE) and cumulated in what is thought to be the greatest mass extinction of the former biosphere.

In addition to the CBB cycle, anoxygenic photosynthesis may employ three of the five additional alternative pathways for carbon fixation [Martin et al., 2018] that have been discovered in bacteria so far. For example, the photomixotrophic green non-sulfur bacterium *Chloroflexus aurantiacus* has the ability to sequester CO_2 via the 3-hydroxypropionate CO_2 fixation cycle [Fuchs, 2011; Zarzycki et al., 2009]. Although their contribution to global net productivity is small, on a local scale these alternative carbon fixation pathways may play an important role in some ecosystems (reviewed in Hügler and Sievert [2010]).

Photoautotrophy is widely but sparsely distributed across separate bacterial lineages and inevitably involved the distribution of genetic information by horizontal gene transfer (HGT) [Martin et al., 2018]. Since photochemistry and carbon sequestration are not always coupled in extant organisms and have different evolutionary histories [Fischer et al., 2016], these processes are examined separately in the following subsections.

1.2.1. The photosynthetic electron transport chain

The core of the photosynthetic electron transport chain encompasses the photosynthetic reaction center(s), which mediates the conversion of light energy to chemical energy and reducing equivalents. The quest for the lineage that first acquired (chloro)photosynthesis is related to the phylogeny of these reaction centers, and stimulated opposing hypotheses regarding their evolution and distribution across separate bacterial lineages [Martin et al., 2018]. While phototrophic members of the lineages proteobacteria, chloroflexi and gemmatimonadetes express a type II reaction center (RCII) which later evolved to the photosystem II (PSII) of cyanobacteria, some phototrophic members of chlorobi, firmicutes and acidobacteria express a type I reaction center (RCI), which later became the ancestral photosystem I (PSI) of cyanobacteria [Fischer et al., 2016]. While cyanobacteria and all archaeplastida (including algae and plants) employ a linear electron flow across two heterodimeric

photosystems, other bacterial lineages maintain a cyclic electron transport and rely on a single homodimeric reaction center.

According to the "export" hypothesis suggested by [Allen, 2005; Olson and Pierson, 1987] and extended and coined by Martin et al. [2018], both reaction centers evolved in the ancestor of recent cyanobacteria (protocyanobacterium). While RCI likely appeared first and provided an H_2 independent pathway for the reduction of ferredoxin (for subsequent production of NADH for reductive carbon fixation), RCII appeared in consequence of an ancient gene duplication event in the same organism. Subsequently, all other phototrophic organisms received either copy of the reaction centers by HGT from this protocyanobacterium, before PSII acquired the ability to oxidize water. This theory is mainly supported by a single shared pathway for chlorophyllide a synthesis, the precursor of all bacteriochlorophylls and chlorophyll a and b, which is indicative of a gene duplication event followed by functional diversification in the same lineage. Moreover, the presence of plasmids carrying phototrophic genes which are mobile in *Roseobacter* [Petersen et al., 2012] adds substantial evidence to this hypothesis. Importantly, acquisition of phototrophy by a protocyanobacterium via HGT may also explain the occurrence of the CBB cycle in non-cyanobacterial bacteria, due to transfer of the photochemistry module with or without the CBB cycle genes. The opposing "merger" hypothesis [Blankenship, 2017; Hohmann-Marriott and Blankenship, 2011] proposes, that the two reaction centers evolved in a lineage-specific manner by vertical divergence and before a protocyanobacterium obtained a second copy of the reaction center by HGT. This hypothesis raises the question, why a beneficial trait like photoautotrophy was lost by so many representatives of the lineage directly branching before the cyanobacteria, given that there are only two known cases among cyanobacteria, where photoautotrophy was lost due to the adaptation to a symbiotic lifestyle [Bombar et al., 2014; Thompson et al., 2012]. Thus, the evolution of RCI and RCII is still an active debate.

The linear photosynthetic electron transport chain in oxygenic photosynthesis is briefly described below and depicted in Figure **1.4**: The linear electron transfer of extant cyanobacteria starts with the excitation of a central chlorophyll pair (P680, pigment with absorption maximum at 680 nm) in PSII. Rapid charge separation results in the transient formation of the radical anion pheophytin$^-$ pigment which transfers the electron to the bound plastoquinone molecule. The radical cation P680$^+$ (the most oxidizing chemical species in biology [Rappaport and Diner, 2008]) is reverted to its ground state by a redox-active Tyr residue (within $\sim 100\,\mu s$ [Cox et al., 2013]). The tyrosyl radical in turn is neutralized by the electrons derived from the oxygen evolving complex (OEC, also referred to as water oxidizing complex), which

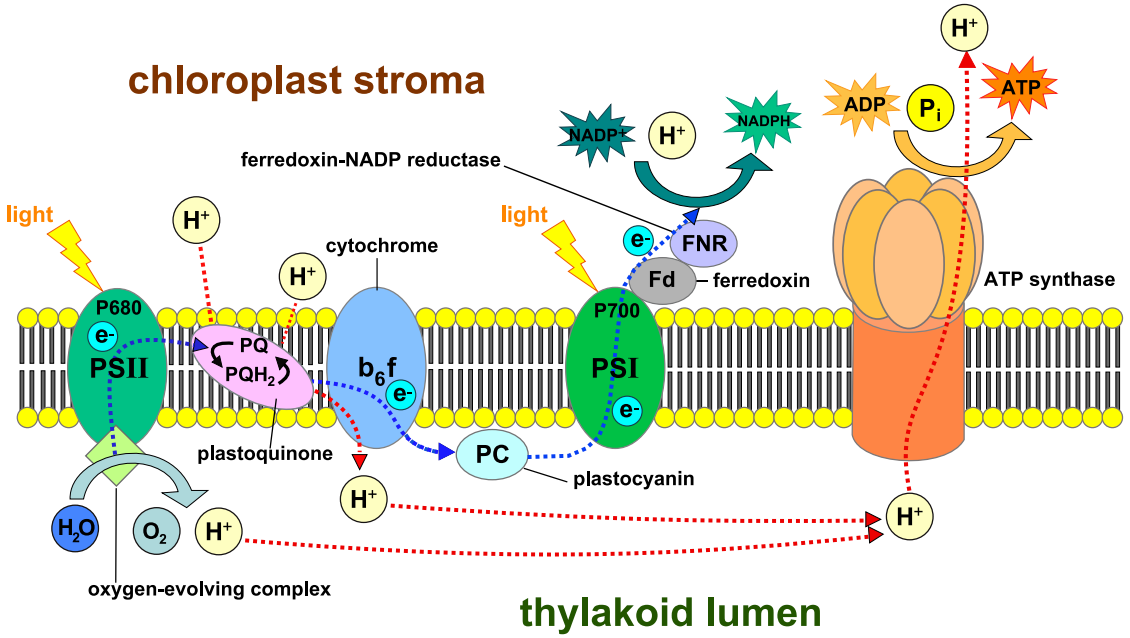

Figure 1.4.: The linear photosynthetic electron transport chain in oxygenic photosynthesis.
The light induced excitation of PSII leads to the successive reduction of the plastoquinone pool. While PSII becomes neutralized by electrons derived from the oxidation of water, plastoquinone transfers electrons via the cytochrome b_6f complex to plastocyanin. Excitation of PSI leads to the reduction of ferredoxin, allowing the ferredoxin-NADP reductase to generate NADPH. PSI in turn is neutralized by electrons from plastocyanin. The generated proton motive force serves to generate ATP by ATP-synthase. See text for details. Figure modified from [Somepics, 2015] (CC BY-SA 4.0).

is associated with PSII. After a second excitation, the plastosemiquinone is fully reduced to plastoquinol and released into the membrane [Fischer et al., 2016]. The OEC comprises a tetranuclear manganese complex and functions in a five-step cycle as first proposed by Kok et al. [1970] and provides the electrons required for charge neutralization at P680. Every four photons two H_2O are oxidized and one molecule of O_2 is produced. Per excitation event, one electron is transferred to P680 via the tyrosyl radical and one H^+ is produced. After release from PSII, reduced plastoquinol is oxidized at the membrane-spanning cytochrome b_6f complex, which involves the shuttling of additional protons across the membrane, and results in the transfer of electrons to the soluble plastocyanin. The excitation of the central chlorophyll pair of PSI (P700, absorption maximum at 700 nm) in a process similar to excitation at PSII, uses phytomenadione as intermediate electron acceptor and channels electrons across three iron-sulfur centers to ferredoxin. Finally, a ferredoxin-$NADP^+$ reductase catalyses the reduction of $NADP^+$ to NADPH, which is available for reductive biosynthetic processes, e.g. in the CBB cycle. The electrochemical potential across the membrane is used by ATP-synthase for the generation of chemical energy in form of ATP.

The central chlorophyll or bacteriochlorophyll pair in the center of cyanobacterial photosystems or proteobacterial reaction centers, respectively, has a narrow absorption range, and thus absorbs only a fraction of the spectral irradiance. With regard to their evolutionary origin and by adaptation to their ecological niche, phototrophs acquired additional accessory pigments that allowed them to broaden their spectrum of photosynthetically available radiation. These pigments are organized in macromolecular structures and are associated with reaction centers and photosystems, e.g. in the chlorosomes of anoxygenic photosynthetic bacteria, the phycobillisomes of cyanobacteria, rhodophytes and glaukophytes or the light harvesting complexes of green algae and plants. Inside these structures the covalently (phycobillisome) or non-covalently linked (chlorosome) light harvesting pigments are arranged in a manner that maximizes energy transfer to the central chlorophyll pair of reaction centers and photosystems.

In cyanobacteria, both photosynthetic electron transport chain and respiratory electron transport chain are located in the plasma membrane and the ubiquinon pools are likely interconnected resulting in cross-talk of these opposed processes. In algae and plants, these elements are located in the thylakoid membrane of chloroplasts, organelles that evolved from the ancestral cyanobacterial endosymbiont. While protons are released from the acidic thylakoid lumen, ATP is generated in the stroma of the chloroplast, which is separated from the cytosol by a highly selective inner and permeable outer membrane.

1.2.2. The Calvin-Benson-Bassham cycle and photorespiration

The CBB cycle likely evolved in the ancestors of cyanobacteria [Hügler and Sievert, 2010], although some of its core enzymes are of more ancient origin and likely were present prior to the separation of archaea and bacteria. Interestingly, in recent organisms, it is not always associated with phototrophy. For example, numerous aerobic chemoautotrophic proteobacteria that oxidize dihydrogen, ferrous iron, sulfur, or ammonia use the CBB cycle [Claassens et al., 2020], suggesting that their CBB cycle genes were acquired by HGT.

The Calvin-Benson-Bassham (CBB) cycle comprises three main functions - carbon sequestration, reduction, and substrate regeneration. Central to the CBB cycle is the enzyme Rubisco, which catalyzes the addition of a gaseous CO_2 to a five-carbon

sugar, which results in a net gain of organic carbon for the organism. The net reaction catalyzed by the CBB cycle is the following:

$$3\,CO_2 + 6\,NADPH + 6\,H^+ + 9\,ATP \longrightarrow G3P + 6\,NADP^+ + 3\,H_2O + 9\,ADP + 8\,P_i \quad (1.1)$$

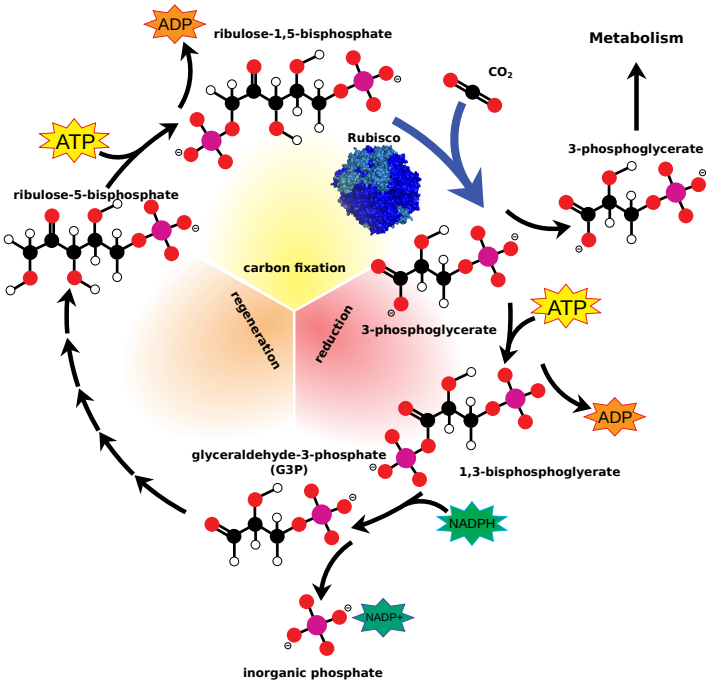

Figure 1.5.: Schematic representation of the CBB cycle.
Rubisco, the key enzyme of the CBB cycle, catalyzes the addition of gaseous CO_2 to the five-carbon sugar RuBP, thereby generating two molecules of 3PGA. After phosphorylation and subsequent reduction, the three-carbon sugar is regenerated to RuBP. One three-carbon sugar is gained every three turns of the cycle and serves the organisms metabolism. Figure modified from [Jones, 2010] (CC BY-SA 3.0).

In a first step, Rubisco catalyses the carboxylation of ribulose 1,5-bisphosphate (RuBP) to two molecules of 3-phosphoglycerate (3PGA). The subsequent reduction of 3 PGA to glyceraldehyde-3-phosphate (G3P) proceeds via the intermediate 1,3-bisphosphoglycerate and is catalyzed by the enzymes phosphoglycerate kinase and glyceraldehyde-3-phosphate dehydrogenase. The former reaction involves an ATP dependent phosphorylation, while the latter requires the cofactor NADPH as hydrogen donor and releases inorganic phosphate. Fixation and reduction of three CO_2 molecules results in the net sequestration of one molecule of G3P, which leaves the cycle and eventually is converted into dihydroxyacetone phosphate (DHAP) which is available for the synthesize of reduced carbon compounds (sugar, starch, lipids, amino acids). The remaining reactions in the CBB cycle serve the regeneration of the five-carbon sugar from the three-carbon compound G3P.

Most enzymes catalyzing reactions for RuBP regeneration overlap with those working in the non-oxidative phase of the pentose phosphate pathway for the generation of reducing equivalents and pentose sugars, e.g for the synthesis of nucleotides. In brief, DHAP and G3P are converted to the six-carbon sugar fructose 6-phosphate. Fructose 6-phosphate and G3P are subject to catalysis by transketolase, yielding the four-carbon compound erythrose 4-phosphate and the five-carbon compound xylulose 5-phosphate. Aldolase catalyses the conversion of erythrose 4-phosphate and DHAP into sedoheptulose 1,7-bisphosphate, which is dephosphorylated to sedoheptulose 7-phosphate. Further processing with G3P by transketolase yields the five-carbon compounds ribose 5-phosphate and xylulose 5-phosphate. Both the former and the latter are converted to ribulose 5-phosphate by ribose 5-phosphate isomerase and phosphopentose epimerase, respectively. Catalysis by the enzyme phosphoribulokinase, a final phosphorylation regenerates RuBP from ribulose 5-phosphate. A schematic representation of the CBB cycle is given in Figure **1.5**.

Next to the carboxylation reaction, Rubisco may accept O_2 as an alternative gaseous substrate. RuBP oxygenation produces only one molecule of 3PGA and another molecule of 2-phosphoglycolate (2PG). Accumulation of 2PG or its metabolites glyoxylate and glycolate is toxic [Anderson, 1971; Kleczkowski, 1994]. In moderate climates, CO_2 loss due to RuBP oxygenation in plants amounts to 20 % of net-photosynthesis, but this is thought to be much higher in arid and warm climates [Cegelski and Schaefer, 2006]. Despite a much higher affinity of Rubisco for CO_2, 2PG is produced in vast amounts, primarily caused by the composition of today's atmosphere, where there are $\sim$ 500 times more O_2 than CO_2 molecules.

In order to counterbalance 2PG production, organisms employing the CBB cycle developed "2PG-salvage pathways" for the recycling of 2PG to 3PGA. Photoautotrophs employ a C2-cycle referred to as "photorespiration", which, in terms of mass flow, is considered the second-most important process in the biogeosphere, next to photosynthesis itself [Bauwe et al., 2010]. The regeneration of 3PGA by photorespiration is metabolically costly, requiring ATP, reducing equivalents and involves the loss of a previously fixed carbon. Hence, in terms of carbon sequestration, it is considered a wasteful process. In plants, photorespiration is spread across three compartments (four including alternative pathways in the cytosol), including chloroplasts, peroxisomes and mitochondria and involves minimally nine different enzymes, not including enzymes required for co-factor regeneration [Bauwe et al., 2010; Peterhansel et al., 2010]. The process is outlined in Figure **1.6**.

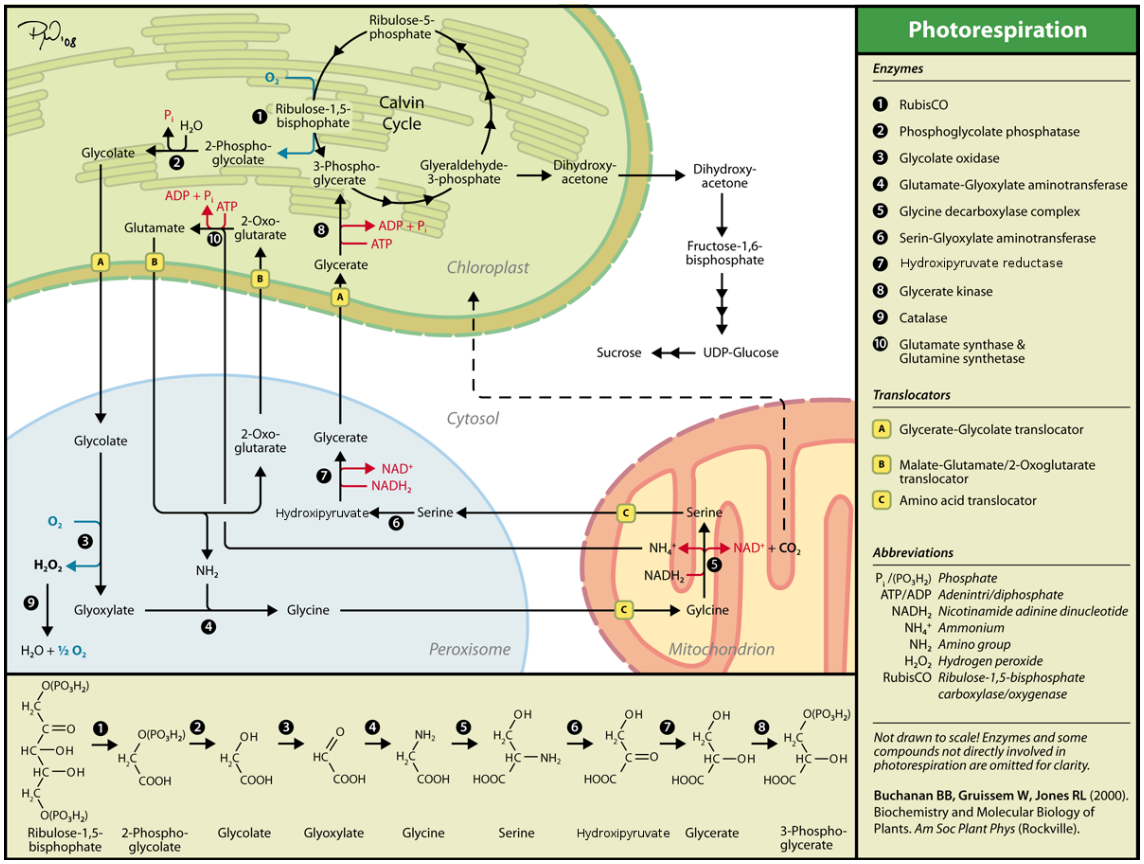

Figure 1.6.: Photorespiration or the 2PG-salvage pathway in plants.
The product of the oxygenation reaction of Rubisco, 2PG, and its metabolites are toxic and have to be converted to 3PGA. The pathway comprises eight enzymes, distributed across the chloroplast, the cytosol, the peroxisome and the mitochondrion. Figure reproduced from [Crenim, 2008] (CC BY-SA 3.0). See bottom right corner for reference to the original version.

While in plants photorespiration is the only 2PG salvage pathway currently known, cyanobacteria employ several partially redundant 2PG-salvage pathways [Eisenhut et al., 2008]. Photorespiration likely co-evolved with oxygenic photosynthesis and the CBB cycle in cyanobacteria. Importantly, the oxygenation reaction of Rubisco is also of major importance for non-oxygenic, aerobic chemolithoautotrophic bacteria, which operate the CBB cycle. As was recently shown for *Cupriavidus necator* H16, these proteobacteria do not rely on the C2 cycle but operate a bacterial-like glycerate pathway which is supported by a malate cycle for 2PG-salvage [Claassens et al., 2020].

In addition to the removal of toxic 2PG and the salvage of 75 % of carbon, photorespiration has numerous other beneficial effects. For example, it has been shown to protect from photoinhibition [Kozaki and Takeba, 1996], implicated in pathogen resistance [Taler et al., 2004] and is interconnected with other essential metabolic path-

ways like nitrogen assimilation [Rachmilevitch et al., 2004] and sulfur metabolism [Abadie and Tcherkez, 2019] to name a few.

1.2.3. Carbon concentrating mechanisms

To decrease the probability of the oxygenation reaction and concomitant carbon loss, autotrophic organisms employing the CBB cycle developed several strategies to increase the local CO_2 concentration in the vicinity of Rubisco, so-called carbon concentrating mechanisms (CCM).

1.2.3.1. C4 photosynthesis and the crassulacean acid metabolism of plants

Next to the C3 photosynthesis described above, plants developed two additional forms of carbon sequestration: the C4 and crassulacean acid metabolism (CAM), which rely on either spatial and temporal separation of CO_2 uptake and carbon fixation, respectively. C4 photosynthesis relies on the addition of CO_2 derived from HCO_3^- to phosphoenolpyruvate (PEP) yielding the four-carbon compound oxaloacetate. Oxaloacetate is produced and converted to aspartate or malate in mesophyll cells, which are transported across plasmodesmata to a specialized cell type, the bundle sheath cells that form ring-like structures immediately surrounding the vascular tissue. In the bundle sheath cells, the prefixed CO_2 is released by decarboxylation and serves as a substrate in the CBB cycle. Most plants employing the CAM pathway populate arid and hot regions. Unlike C3 plants, that would experience severe water loss due to the day time gas exchange via leave stomata, CAM plants restrict gas exchange to the colder night time. During the night, CO_2 is prefixed to PEP yielding oxaloacetate which is converted to malate and stored in the vacuole. During the day, CO_2 is released from malate by decarboxylation and serves as substrate in the CBB cycle. The closed stomata prevent the influx of additional O_2 and CO_2.

1.2.3.2. The algal CCM

The CCM of algae and hornworts relies on the active uptake of CO_2 and HCO_3^- and likely on the properties of the pyrenoid, a specialized compartment in the chloroplast which is encapsulated by starch-like oligosaccharides (in green algae) and consists of a spheroidal protein matrix of Rubisco. The role of the starch sheath is still unclear. Although starch accumulation was reported to coincide with the induction of the CCM and thus proposed to act as a diffusion barrier for gaseous CO_2 required for

CCM function, [Ramazanov et al., 1994], a *Chlamydomonas reinhardtii* mutant unable to synthesize starch displayed an increased affinity to inorganic carbon under carbon limiting conditions, typical for a functional CCM [Villarejo et al., 1996]. The pyrenoid is traversed by membranous tubules which connect to the photosynthetic thylakoid membranes in the stroma [Engel et al., 2015]. Embedded minitubules connecting the chloroplast stroma with the pyrenoid lumen may serve for metabolite exchange with other proteins of the CBB cycle [Hennacy and Jonikas, 2020]. While recent progress in the EM technique cryo-electron tomography in combination with biophysical measurements and *in vitro* studies focusing on the pyrenoid of *Chlamydomonas reinhardtii* lead to novel insights into the pyrenoid organization [Engel et al., 2015; Freeman Rosenzweig et al., 2017; Itakura et al., 2019; Meyer et al., 2020, 2012], a detailed understanding of the pyrenoid is still elusive. Importantly, many but not all algae contain a pyrenoid, thus pyrenoids are not obligatory for a functional CCM [Meyer et al., 2017].

1.2.3.3. The cyanobacterial CCM

Cyanobacterial CCMs rely on the active uptake of inorganic carbon by membrane localized CO_2 and HCO_3^- transporters as well as on carboxysomes, specialized microcompartments that function in carbon sequestration. Carboxysomes are proteinaceous, icosahedral, bacterial microcompartments (BMC) equipped with large quantities of Rubisco and carbonic anhydrase (CA) (Figure **1.7**). The selective carboxysomal shell acts as a diffusion barrier for gaseous CO_2 but allows entry of small polar molecules, like HCO_3^- [Cai et al., 2009; Dou et al., 2008]. This organization has been reported to lead to up to 1000-fold increased levels of inorganic carbon in the carboxysomal interior (in comparison to the bulk medium) in vicinity of Rubisco [Rae et al., 2013] (Figure **1.7 D**).

1.2.3.3.1. Two divergent carboxysomes with analogous function

Two types of carboxysomes, the α- and β-carboxysomes are distinguished by their phylogenetic origin, carboxysomal gene organization, their assembly pathway and the form of Rubisco they accommodate [Rae et al., 2013]. α-Carboxysomes accommodate form IA Rubisco together with a β-CA (CsoSCA) and are found in α-cyanobacteria and chemoautotrophs, whereas β-carboxysomes accommodate form IB Rubisco together with a γ-CA and are exclusively found in β-cyanobacteria. The β-carboxysome associated γ-CA always fulfills a structural role in carboxysome organization and may or may not display CA function [de Araujo et al., 2014;

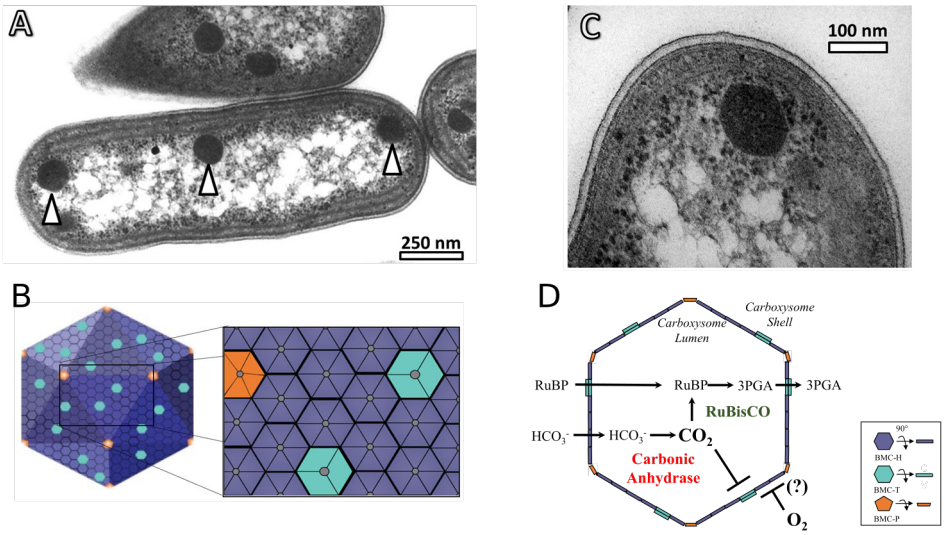

Figure 1.7.: β-carboxysomes of *S. elongatus* PCC 7942 and a model of their function.
A) Subcellular organization of β-carboxysome in *S. elongatus* PCC 7942. The white arrows point to carboxysomes. **B) Architecture of a carboxysome shell.** BMC-H protein in purple, BMC-T protein in light blue and the BMC-P protein at the vertices in orange. Shell protein's pores are indicated as grey circles. **C) Close-up of a β-carboxysome from *S. elongatus* PCC 7942. D) Model of carboxysome function.** Carbonic anhydrase releases CO_2 from HCO_3^- in the carboxysomal lumen. Because the carboxysomal shell represents a diffusion barrier for gases, the partial pressure of CO_2 is dramatically increased, repressing the oxygenation reaction of the co-encapsulated Rubisco. Small molecules (RuBP, 3PGA, HCO_3^-) are thought to be able to pass through pores within the carboxysomal shell. Figures **A/C** and **B/D** reproduced with permission from Rae et al. [2013] and Turmo et al. [2017], respectively.

Pena et al., 2010]. β-Carboxysomes containing the latter accommodate an ancillary β-CA (CcaA) [Cot et al., 2008; McGurn et al., 2016; Price et al., 1992]. While α-carboxysomal genes are largely organized within a single *cso* operon, β-carboxysomal genes are more scattered across the genome and comprise several satellite loci of the core *ccm* operon. The carboxysomal shell is formed by hexameric BMC-H, pentameric BMC-P and pseudohexameric BMC-T proteins. In α-carboxysomes, these are represented by CsoS1/S1E, CsoS4A/S4B and CsoS1D, whereas in β-cyanobacteria these are represented by CcmK1/K2/K3/K4, CcmL and CcmO/CcmP, respectively [Kerfeld and Melnicki, 2016] (Figure **1.7 B**). While the BMC-P proteins are encoded in the carboxysomal core gene clusters and constitute the vertices of the icosahedron, BMC-H and BMC-T are the main building blocks of the shell. BMC-H and BMC-T each comprise a family of several proteins which may be encoded within the carboxysomal core gene cluster or in satellite loci. They differ in regions proximal to the channel formed at the center of each hexamer, which is thought to selectively facilitate access of metabolites [Sommer et al., 2017]. Interestingly, some BMC-T proteins have been shown to form stacked and gated shell proteins (e.g. Cso1D [Klein et al., 2009] and CcmP [Cai et al., 2013]), suggest-

ing the exchange of metabolites or small peptides across the carboxysomal shell. A phylogeny of BMC vertex proteins provided evidence for a deep divergence of α- and β-carboxysomes and the β-carboxysome was suggested to be closer related to metabolosomes of heterotrophs than to its functional α-carboxysomal counterpart [Kerfeld and Melnicki, 2016]. While in β-carboxysomes the cargo organizes carboxysome formation, the cargo and shell of the α-carboxysomes are thought to assemble simultaneously [Cameron et al., 2013; Iancu et al., 2010]. A more detailed description of the biogenesis of β-carboxysomes is given in Chapter **1.5.1.1**.

1.3. Rubisco

With its crucial function in the CBB cycle, Rubisco plays an essential role in the recent biogeochemical carbon cycle and began to shape our extant bio-, geo- and atmosphere since its first appearance during the Archean Eon [Nisbet et al., 2007]. Different Rubisco forms occur in all three domains of life, which are likely of monophyletic origin, and thus the first appearance of the enzyme activity is thought to predate the branching of bacteria and archaea [Tabita et al., 2008]. Rubisco catalyzes the exergonic carboxylation and oxygenation of ribulose-1,5-bisphosphate (RuBP) to either two molecules of 3-phosphoglycerate (3PGA) or one molecule 3PGA and another molecule 2-phosphoglycolate (2PG) (see Equation (1.2) and (1.3)), respectively, but only the former contributes to the net gain of organic carbon.

$$\mathrm{CO_2 + RuBP \longrightarrow 2\,PGA + 2\,H^+} \tag{1.2}$$

$$\mathrm{O_2 + RuBP \longrightarrow PGA + PG + 2\,H^+} \tag{1.3}$$

Rubisco is a comparatively slow enzyme with a catalytic rate (k_{cat}) between 2-14 s^{-1}; for comparison carbonic anhydrase, one of the fastest enzymes known, has a k_{cat} of $10^6\,s^{-1}$ [Khalifah, 1971]. Hence, many organisms express Rubisco to high abundance, presumably to compensate for the slow reaction kinetics, resulting in an estimated 0.7×10^{15} g for the accumulated mass of Rubisco on Earth [Bar-On and Milo, 2019].

According to phylogenetic analyses of Rubisco large subunit sequences, four forms of Rubisco can be distinguished. While Rubisco of form I, II and III all catalyze the carboxylation of RuBP, form IV, so-called Rubisco-like protein (RLP), lacks several

catalytic residues and thus is unable to catalyze CO_2 fixation [Hanson and Tabita, 2001; Tabita et al., 2008]. Form I Rubisco occurs in most photoautotrophic eukaryotes, in cyanobacteria and proteobacteria. These enzymes can be further subdivided in form IA, IB, IC and ID. Sometimes form IA and IB are referred to as "green-type" and form IC and ID as "red-type" Rubisco, according to their primary occurrence in the "green" (cyanobacteria, algae, and plants) or "red" plastid lineage (non-green algae), respectively [Tabita et al., 2007]. Form IA is found in α-cyanobacteria and various proteobacteria and can be distinguished according to their association with carboxysomal gene clusters as form IAc (carboxysomal gene cluster associated) or IAq (CbbQ/O associated - see Chapter **1.4.6.3**). Form IB Rubisco is found in β-cyanobacteria, green algae and plants, whereas form IC and form ID Rubisco occur primarily in α- and β-proteobacteria and rhodophytes, respectively. A brief overview of the distribution of Rubisco forms is presented by Badger and Bek [2008]. From this overview it is apparent that the presence of a single Rubisco gene per genome correlates with obligate autotrophy, whereas facultative autotrophs and mixotrophs may encode several Rubiscos of different forms. Interestingly, differential expression patterns have been observed for form I and II Rubisco in proteobacterial species encoding multiple Rubiscos [Badger and Bek, 2008; Kusian and Bowien, 1997]. E.g. while form II Rubisco was expressed independently of the CO_2 level, both form I Rubiscos were only expressed at low CO_2 levels, consistent with the on average higher selectivity of form I Rubisco for the carboxylation reaction. Only under low CO_2 conditions the formation of carboxysomes was noted [Yoshizawa et al., 2004].

There are remarkable differences between form I Rubiscos with regard to their preference for the carboxylation reaction (specificity), which is frequently expressed as $S_{C/O} = \frac{k_{cat}^{CO2}}{K_m^{CO_2}} / \frac{k_{cat}^{O2}}{K_m^{O_2}}$, with the specificity $S_{C/O}$ and the maximal catalytic rates $k_{cat}^{CO2}, k_{cat}^{O2}$ and the Michelis-Menten constants K_m^{CO2}, K_m^{O2} for the carboxylation and oxygenation reaction, respectively. On average form ID Rubiscos have a comparably high $S_{C/O}$ ($\sim$ 140) but low K_m^{CO2} ($\sim 30\,\mu mol\,L^{-1}$), whereas cyanobacterial form IB Rubiscos have a lower $S_{C/O}$ ($\sim$ 45) but a higher K_m^{CO2} ($\sim 220\,\mu mol\,L^{-1}$) [Whitney et al., 2011]. A higher $S_{C/O}$ has been suggested to be one adaptation to rising O_2 concentrations [Jordan and Ogren, 1983]. Of note, very recently another Rubisco type, coined form I', has been identified in metagenome-assembled genomes (MAG) of Chloroflexi that forms a separate cluster in proximity to form I Rubisco in phylogenetic analyses. Form I' Rubisco from *Candidatus* 'Promineofilum breve' displays k_{cat}^{CO2} and K_m^{CO2} similar to some form ID Rubiscos, but their $S_{C/O}$ is in the range of cyanobacterial Rubiscos [Banda et al., 2020].

Form II Rubisco mainly occurs in proteobacteria and dinoflagellates, but also in archaea. They display $\sim 30\,\%$ sequence identity with form I Rubisco [Tabita et al., 2007] and display variable kinetic properties, although their $S_{C/O}$ is comparably low (~ 20) [Whitney et al., 2011]. Similar to form I Rubisco, form II Rubisco functions in the CBB cycle, although, as of now, no evidence suggests form II Rubisco to function in conjunction with a CCM [Badger and Bek, 2008].

Form III Rubiscos are only found in archaea. Although they catalyze the same reaction as Rubisco of from I and II, they do not function in the CBB cycle or any other photosynthesis related pathway. Work from Aono et al. [2012] suggested, that Rubisco of the archaeon *Thermococcus kodakarensis* functions at the final step within a short RNA degradation pathway. Genome analyses of several archaea demonstrated that the pathway is widely distributed among archaea, but not present in eukaryotes or bacteria.

Form IV Rubisco or Rubisco-like proteins (RLPs) display $\sim 23\,\%$ sequence similarity with form I Rubisco [Ashida et al., 2005], but do not catalyze the carboxylation of RuBP. Instead Ashida et al. [2003] demonstrated that the RLP from *Bacillus subtilis* catalyzes the 2,3-diketo-5-methylthiopentyl-1-phosphate enolase reaction in the methionine salvage pathway. Moreover, they demonstrated that the RLP deficient phenotype of *B. subtilis* could be rescued by expression of form II Rubisco from *Rhodospirillum rubrum*. Because RLP expressing bacteria and Archaea are more ancient than the invention of the CBB cycle, Ashida et al. [2005] suggested that RLPs may be the ancestral enzymes of photosynthetic Rubisco.

1.3.1. Structure of Rubisco

Rubisco was first isolated from *Spinacia oleracea* leaves in 1947 by Wildman and Bonner [1947] as fraction I protein. The first crystals of Rubisco were obtained from a form I enzyme, Rubisco of *Nicotiana tabacum*, as early as 1977 by Baker et al. [1977], and crystals of form I Rubiscos from a chemoautotrophic proteobacterium and spinach followed [Bowien et al., 1980; Andersson and Brändén, 1984]. However, the first high resolution structure was obtained from form II Rubisco of *Rhodospirillum rubrum* in 1986, which revealed the detailed three-dimensional structure of a dimer of large subunits of Rubisco [Schneider et al., 1986]. Because Rubisco from *Rhodospirillum rubrum* (*Rr*Rubisco) could be heterologously expressed in *E. coli*, it became a model protein for the investigation of Rubisco structure-function relationships. In the 1990ies, this structure was refined to 1.7 Å [Schneider et al., 1990]

and several form I Rubisco structures of plants were solved. For an overview of early available Rubisco structures and their functional state, see Duff et al. [2000].

1.3.1.1. The RbcL dimer as functional unit

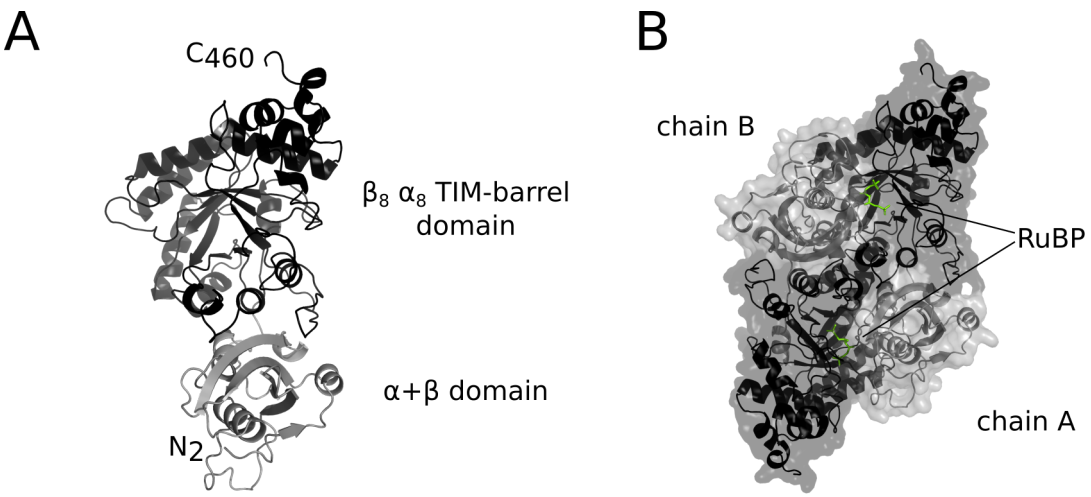

Figure 1.8.: Structure of RbcL.
A) Crystal structure of the RbcL subunit (chain A) of Rubisco from *Rhodospirillum rubrum* in ribbon representation. The N-terminal $\alpha+\beta$ domain, the C-terminal β_8 α_8 TIM-barrel domain and chain termini are indicated. **B)** Anti-parallel arrangement of RbcL subunits in the functional Rubisco dimer. The position of the active site pockets is indicated by the bound RuBP substrate (PDB: 5RUB, [Lundqvist and Schneider, 1991]).

The functional unit of Rubisco is an antiparallel dimer of large subunits of about $\sim$ 52 kDa (Figure **1.8**). The RbcL N-terminal domain of $\sim$ 150 residues forms a $\alpha + \beta$ domain, whereas the remaining $\sim$ 325 residues consist of a β_8 α_8 triose phosphate isomerase (TIM)-barrel domain ($\sim$ 310 residues) and a flexible C-terminal peptide ($\sim$ 15 residues) [Bracher et al., 2017; Knight et al., 1990; Schneider et al., 1990]. Two active sites are located at the interface between the N-terminal domain of one subunit and the C-terminal TIM-barrel domain of the antiparallel RbcL subunit. The majority of active site residues are contributed by the C-terminal domain (Figure **1.17 C**) [Duff et al., 2000; Taylor and Andersson, 1997].

1.3.1.2. The quarternary arrangement of RbcL dimers

An overview of the arrangement of RbcL dimers in Rubisco holoenzymes of different forms is given in Figure **1.9**.

In form I Rubisco, RbcL antiparallel dimers associate to form a tetramer of dimers (Figure **1.9 A**). In form IA, IB, IC and ID Rubisco this octameric cube is capped

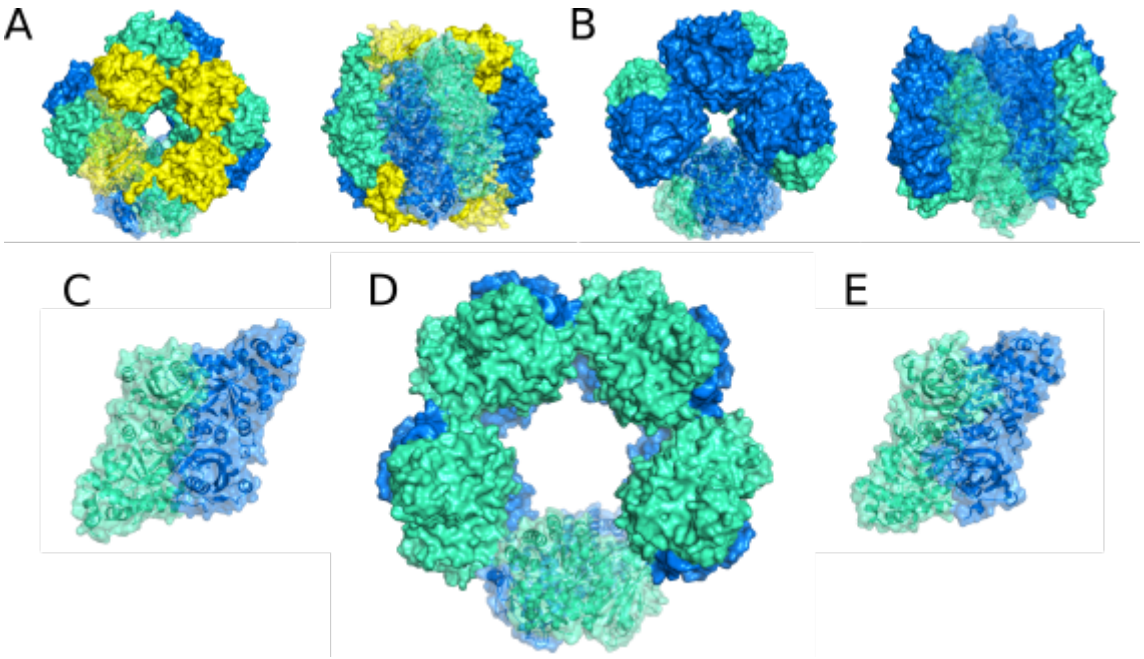

Figure 1.9.: Structure of Rubisco.
A) Top- (left) and side view (right) form IB Rubisco from *Synechococcus elongatus* PCC 6301 (PDB: 1RBL [Newman et al., 1993]). **B)** Top- (left) and side view (right) of form I' Rubisco from *Candidatus* Promineofilum breve (PDB: 6URA [Banda et al., 2020]). **C)** Side view of form II Rubisco of *Rhodospirillum rubrum* (PDB:5RUB [Schneider et al., 1990]). **D** Top view of form III Rubisco from *Thermococcus kodakarensis* KOD1 (PDB:1GEH [Kitano et al., 2001]). **E** Side view of form IV Rubisco-like protein from *Chlorobaculum tepidum* (PDB:1YKW [Li et al., 2005b]). All Rubiscos in surface representation and antiparallel RbcL chains are displayed in alternating colors (marine and greencyan, respectively). RbcS is shown in yellow. The surface of one antiparallel RbcL-dimer (including RbcS in **B**) per Rubisco is transparent and individual protein chains are shown as ribbons.

by four small subunits at the top and four at the bottom to form the functional hexadecameric holoenzyme of $\sim$ 520 kDa (Figure **1.10 B**). The recruitment of RbcS is thought to be an evolutionary adaptation to the rise of atmospheric oxygen due to global photosynthetic activity. Since plants and green algae encode multiple differentially expressed RbcS isoforms, they were suggested to fulfill a regulatory role. RbcS comprises a core fold of a four-stranded antiparallel β-sheet which is covered by two helices on one side [Knight et al., 1990] (Figure **1.10 A**). In the functional hexadecameric holoenzyme, RbcS does not contribute directly to the active site, but is required for catalytic activity via long-range effects and thus thought to stabilize the active site.

Interestingly, a major structural difference between RbcS of red- and green-type organisms manifests in the presence of an additional antiparallel β-hairpin motif at the RbcS C-terminus (Figure **1.10 A**). The β-hairpins of RbcS combine to form an eight-stranded β-barrel structure, which lines the central solvent channel at both ends of the form IC and ID Rubisco cubes [Hansen et al., 1999; Joshi et al., 2015;

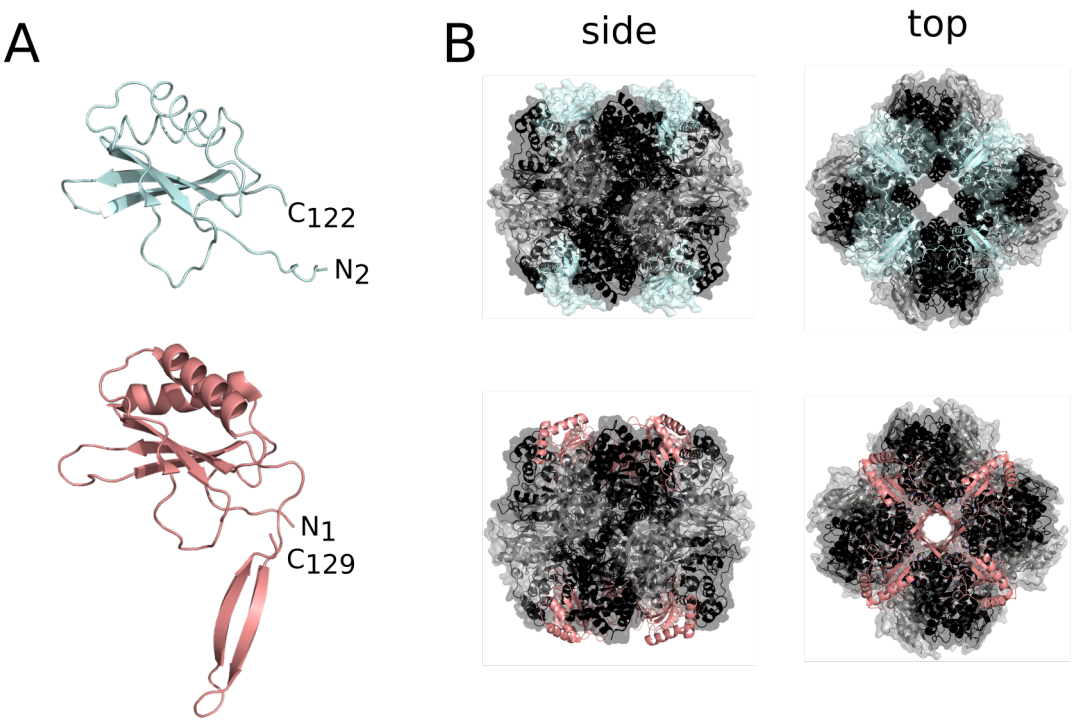

Figure 1.10.: Structure of RbcS and its arrangement in the Rubisco from I holoenzyme.
The structures of green- (**top**) and red-type (**bottom**) RbcS (**A**) and their respective arrangement in the holoenzymes of Rubisco form IB from *Synechococcus elongatus* PCC 6301 (PDB:1RBL) and IC from *Rhodobacter sphaeroides* (PDB:5NV3) (**B**) are shown in cyan and red, respectively. Protein chains are represented as ribbons and chain termini are indicated. In **B**, the α+β domain and the β_8 α_8 TIM-barrel domain are shown in grey and black, respectively, and the surface of the holoenzyme is indicated.

Sugawara et al., 1999]. These β-hairpins have been implicated to be involved in the assembly of the Rubisco form IC holoenzyme [Joshi et al., 2015].

Importantly, form I' Rubisco does not associate with RbcS and functions as tetramer of dimers (Figure **1.9 B**). The discovery of form I' Rubisco came as a surprise to the Rubisco community, since previously RbcS was thought to be required for the stabilization of the octameric core of RbcL in existent Rubiscos. The only form I Rubisco RbcL known to assemble to an octameric core is more aggregation-prone than the hexadecameric holoenzyme and almost carboxylation inactive [Goloubinoff et al., 1989b].

Form I Rubisco is the only Rubisco associated with RbcS. Form II Rubisco generally occupy a range of oligomeric species. They either function as dimers (e.g. Rubisco of *Rhodospirillum rubrum*) or associate to higher oligomers (Figure **1.9 C**). Association to trimers of RbcL dimers have been reported on several accounts, e.g. in the case of *Rhodopseudomonas palustris* [Satagopan et al., 2014]. Archaeal form III

Rubisco occurs as dimer or forms pentamers of antiparallel dimers, like Rubisco from *Thermococcus kodakarensis* (Figure **1.9 D**). The Rubisco-like proteins of form IV Rubisco are closely related to form II Rubisco and have not been reported to form assemblies large than the antiparallel dimer [Ashida et al., 2003; Imker et al., 2007] (Figure **1.9 E**).

1.3.2. Rubisco biogenesis

1.3.2.1. RbcL is an obligate chaperonin client and requires folding assistance to reach its native state

Like many TIM-barrel domain proteins [Kerner et al., 2005], folding of RbcL of both Form I and II Rubisco relies on the action of the chaperonins. In fact, a plant chloroplast homologue of the bacterial GroEL and mitochondrial Hsp60 was first coined "Rubisco subunit binding protein", based on the observation that newly translated RbcL subunits associated with a large protein complex, that was mainly composed of $\sim$ 60 kDa subunits and later termed Cpn60 [Barraclough and Ellis, 1980; Ellis and Van Der Vies, 1988]. Later, Goloubinoff et al. [1989a] demonstrated that the denatured form II Rubisco from *Rhodospirillum rubrum* (*Rr*Rubisco) can be reconstituted *in vitro*, which critically depends on the bacterial chaperonin GroEL and its cofactor GroES and requires Mg–ATP. Folding of the RbcL subunit of *Rr*Rubisco has been demonstrated to occur in the confined environment of the chaperonin *cis* cavity, often referred to as "cage", formed between one GroEL ring and its cofactor GroES (see Chapter **1.1.2.3** and Figure **1.2**). Folding inside the cage has been shown to be accelerated several times in comparison to spontaneous folding in free solution under conditions that prevent aggregation [Brinker et al., 2001].

While chaperonin-folded RbcL of form II Rubisco from *Rhodospirillum rubrum* have been shown to spontaneously assemble into soluble dimers *in vitro* [Goloubinoff et al., 1989a], attempts to reconstitute the hexadecameric form IB Rubisco of plants, green algae and cyanobacteria have failed in presence of the chaperonins alone [Liu et al., 2010; Tabita, 1999].

1.3.2.2. Assembly of form IB Rubisco of cyanobacteria

To date, two assembly factors have been characterized that are involved in the assembly of form IB Rubisco of cyanobacteria - RbcX and Raf1. Albeit at low levels, some functional hexadecameric Form IB Rubisco can be recombinantly expressed

with co-overexpression of *E. coli* chaperonins GroES and GroEL (*Ec*SEL), like e.g. Rubisco from *Synechococcus elongatus* PCC 7942 [Goloubinoff et al., 1989b], the yield of assembled Rubisco is increased by co-expression of RbcX, Raf1 or both [Emlyn-Jones et al., 2006; Hauser et al., 2015; Kolesinski et al., 2014; Li and Tabita, 1997; Onizuka et al., 2004; Saschenbrecker et al., 2007; Whitney et al., 2015].

1.3.2.2.1. The assembly chaperone RbcX

Many cyanobacteria contain an additional gene, *rbcX*, located between *rbcL* and *rbcS* within the Rubisco operon (e.g. *Synechococcus* sp. PCC 7002, *Synechocystis* sp. PCC 6803 or *Thermosynechococcus elongatus* BP-1) - in others it is encoded elsewhere (e.g. *Synechococcus* PCC 7942). Observations, that form IB Rubisco cannot be reconstituted *in vitro* by the chaperonins alone, stimulated the idea that perhaps additional assembly factors are required to mediate its assembly. RbcX being co-transcribed with the Rubisco genes in some β-cyanobacterial species, was a likely candidate for this function.

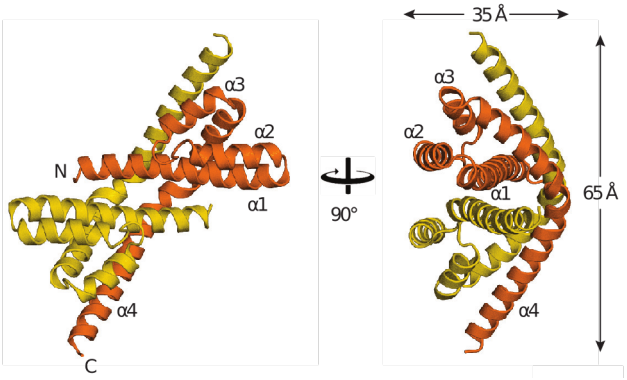

Figure 1.11.: Crystal structure of RbcX from *Synechococcus* sp. PCC 7002.
Two perpendicular views of the RbcX dimer (PDB: 2PEN, [Saschenbrecker et al., 2007]) are shown and protein chains are displayed in ribbon representation. Individual protomers are indicated in orange and gold. Figure adapted from Bracher et al. [2017] with permission.

Functional and structural analysis of *Synechococcus* sp. PCC 7002 RbcX revealed that RbcX is a $\sim$ 15 kDa protein of α-helical structure arranged in a 4-helix bundle (Figure **1.11**) [Bracher et al., 2011; Saschenbrecker et al., 2007]. One helix is elongated and contains a $\sim 60°$ kink. RbcX forms anti-parallel, arc-shaped homodimers with the elongated helix serving as a dimerization interface. The conserved central cleft of the homodimer is hydrophobic and binds the RbcL C-terminal sequence EIKFE(F/Y)X [Saschenbrecker et al., 2007]. The crystal structure and an EM density of an end-state assembly intermediate of form I Rubisco revealed that up to eight RbcX dimers can bind to the octameric RbcL core (L_8) [Bracher et al.,

2011; Liu et al., 2010]. The RbcX dimer binds the antiparallel RbcL dimer (L_2) in a clamp-like manner, engaging the RbcL C-terminus and peripheral regions on RbcL, thereby stabilizing L_2 and likely guiding the assembly of the L_8 core. Subsequent binding of RbcS involves the ordering of the previously unstructured 60s loop required for Rubisco function, which results in the displacement of the RbcX dimer [Bracher et al., 2011]. Biochemical data indicated, that RbcX acts after chaperonin mediated folding of RbcL and prevents it from rebinding to GroEL [Liu et al., 2010]. Notably, RbcX from *Anabaena* sp. CA, that displayed a $\sim$ 30 fold higher affinity to *Synechococcus* sp. PCC 7002 Rubisco in comparison to the endogenous RbcX, could no longer be replaced by RbcS, thus demonstrating that the relative stoichiometry and affinities of RbcX and RbcS to Rubisco seem to be important for function, and determine the fate of the L_8 core. Figure **1.12** depicts the current model of RbcX-assisted cyanobacterial form IB Rubisco assembly.

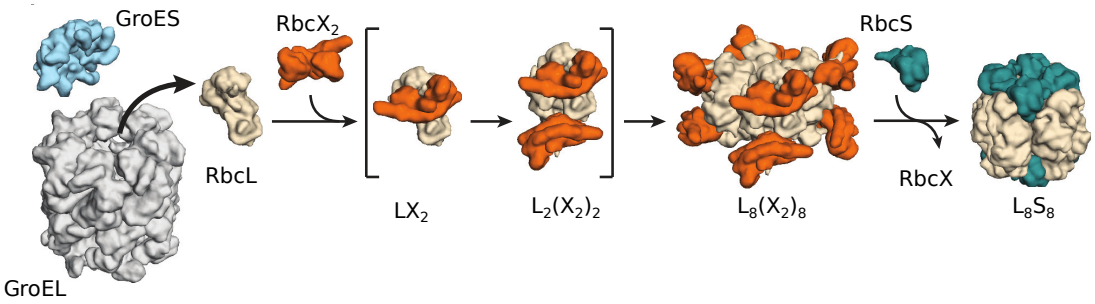

Figure 1.12.: Model of RbcX-assisted assembly of cyanobacterial form IB Rubisco.
Upon its release from GroEL-GroES, RbcL exposes its flexible C-terminus enabling the RbcX dimer ($RbcX_2$) to bind either the RbcL monomer (LX_2) or dimer ($L_2(X_2)_2$) which prevents RbcL aggregation. Likely guided by RbcX, the stabilized $L_2(X_2)_2$ tetramerize to form $L_8(X_2)_8$. Subsequent binding of RbcS structures both the RbcL N-terminus and the 60s loop and ultimatively displaces RbcX to form the L_8S_8 holoenzyme. Figure adapted from Bracher et al. [2017] with permission.

However, experiments that sought to analyze the contribution of RbcX on heterologous Rubisco production in *E. coli*, were controversial. While co-expression of RbcX enhanced the yield of soluble Rubisco from *Anabaena* sp. CA, *Synechococcus* sp. PCC 7002 and *Synechococcus elongatus* PCC 7942 [Emlyn-Jones et al., 2006; Li and Tabita, 1997; Onizuka et al., 2004], coexpression of RbcX had no effect on the production of Rubisco from *Nostoc* sp. PCC 7120 [Larimer and Soper, 1993]. Moreover, *in vivo* data on cyanobacterial Rubisco assembly by RbcX is currently sparse and sometimes conflicting. To what extent cyanobacterial hexadecameric Form 1B Rubisco formation depends on assembly factors is not clear and apparently varies between species. For example, the attempt to inactivate *rbcX* in *Synechococcus* sp. PCC 7002 resulted in incomplete segregation of transformants and was either interpreted as obligate requirement for *rbcX* or questioned as a pleiotropic disruption

of rbcL and rbcS, given their proximity within the Rubisco operon [Emlyn-Jones et al., 2006; Onizuka et al., 2004]. In *S. elongatus* PCC 7942, *rbcX* is not essential, and - according to Emlyn-Jones et al. [2006] - its deletion did affect neither growth nor Rubisco content or partitioning into carboxysomes, while in a more recent study deletion of *rbcX* was reported to increase Rubisco abundance without impairment of holoenzye formation [Huang et al., 2019]. Due to additional observed effects on carboxysome formation and distribution, Huang et al. [2019] suggested a role for RbcX in initiation of carboxysome formation. Interestingly, even the simultaneous deletion of *rbcX* and *raf1* allowed *S. elongatus* PCC 7942 cells to survive in air, indicating that neither of the genes are as important for Rubisco assembly in *S. elongatus* PCC 7942 as previously speculated [Huang et al., 2020]. However, it should be noted that Rubisco assembly in *S. elongatus* PCC 7942 may not be representative for the bulk of β-cyanobacteria, given the distinct position of *rbcX* outside the Rubisco operon as well as its relaxation from the requirement for assembly factors, as demonstrated by Goloubinoff et al. [1989b].

1.3.2.2.2. The assembly chaperone Raf1

Rubisco accumulation factor 1, Raf1, was first identified in a Photosynthetic Mutant Library in maize and its destruction was shown to result in lethality at the seedling stage [Stern et al., 2004]. Comparative biochemical analysis of the *raf1* and wild-type maize, implicated Raf1 in Rubisco holoenzyme formation and a role similar to RbcX was suggested for Raf1 in stabilizing L_2, thus aiding its release from the plant chaperonins Cpn60α/β [Feiz et al., 2012]. Unlike RbcX, Raf1 expression was verified on the protein level in enriched fractions of bundle sheath cells in maize, as expected for a chaperone of Rubisco [Feiz et al., 2012]. Raf1 is encoded in all cyanobacterial and eukaryotic organisms encoding RbcX. Moreover, heterologous expression of Rubisco from *Thermosynechococcus elongatus* BP-1 with its cognate Raf1 in *E. coli* and *in vitro* reconstitution experiments emphasized its involvement in cyanobacterial Rubisco biogenesis [Kolesinski et al., 2014].

Raf1 functions as a dimer with a subunit molecular weight of $\sim$ 40 kDa (Figure **1.13**). Its N-terminal domain adopts an α-helical arrangement, whereas the C-terminal domain forms β-sheets and is important for homodimerization [Hauser et al., 2015]. Reconstitution experiments of unfolded RbcL from *Synechococcus elongatus* PCC 7942 in presence of various combinations of chaperonins, Raf1 and RbcS indicated that Raf1 acts after chaperonin-mediated folding, like RbcX. A combination of native and crosslinking mass spectrometry as well as low resolution EM reconstructions indicated, that four Raf1 dimers bind to the L_8 core, with one Raf1

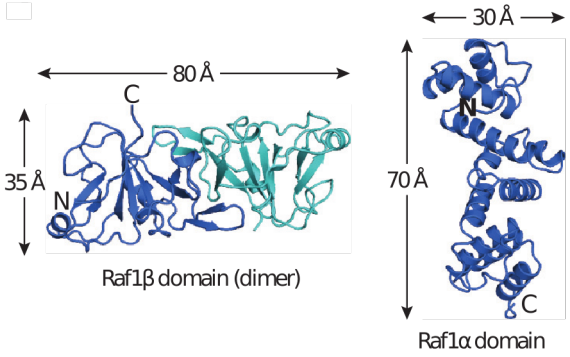

Figure 1.13.: Crystal structures of the C-terminal dimerization domain (β domain) and the N-terminal α domain of Raf1 from *Arabidopsis thaliana*.
Protein chains in the β domain (PDB: 4WT3) and the α domain (PDB: 4WT4) are displayed in ribbon representation and the individual protomers in the β domain are indicated in blue and cyan. Figure adapted from Bracher et al. [2017] and **?** with permission.

dimer bracketing one antiparallel RbcL dimer with the α-helical domains, while the β-sheet dimerization domain is positioned at the equator of L_2 [Hauser et al., 2015]. Upon addition of RbcS to the $L_8(Raf1_2)_4$ assembly intermediate, Raf1 was displaced and the holoenzyme formed. Hence, the role of Raf1 in Rubisco biogenesis, resembles the role of RbcX in stabilizing the RbcL dimer to keep it in an assembly competent state. Figure **1.14** depicts the current model of Raf1-assisted cyanobacterial form IB Rubisco assembly.

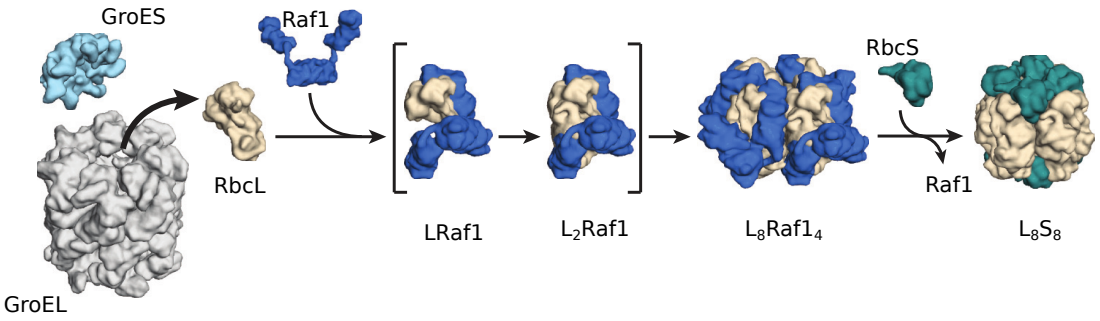

Figure 1.14.: Model of Raf1-mediated assembly of cyanobacterial form IB Rubisco.
Upon its release from GroEL-GroES, RbcL either dimerizes spontaneously or dimerization is supported by Raf1 likely involving the flexible RbcL C-terminus. Raf1 prevents aggregation of the otherwise labile RbcL dimer (L_2Raf1). Insertion of the C-terminus of Raf1 into the active site pocket of Rubisco contributes to the tetramerization of L_2Raf1 to L_8Raf1_4. Due to overlapping binding sites of the α-helical domain of Raf1 and RbcS, Raf1 is displaced by RbcS, eventually forming the functional holoenzyme L_8S_8. Figure adapted from Bracher et al. [2017] with permission.

In vivo, Raf1 was found to be important but not essential for Rubisco assembly, because deletion of *raf1* in *S. elongatus* PCC 7942 caused only reduced cell growth and Rubisco protein content ($\sim 30\,\%$ decrease in Rubisco holoenzyme formation) [Huang et al., 2020]. Interestingly, severe effects on carboxysome formation were observed

in Δ*raf1* which were partially rescued in the Δ*rbcx*Δ*raf1* mutant, suggesting that RbcX and Raf1 are not functionally redundant but may act in an antagonistic way in mediating Rubisco assembly in *S. elongatus* PCC 7942 [Huang et al., 2020].

1.3.2.3. Assembly of form IB Rubisco of plants

An overview of molecular chaperones participating in plant Rubisco biogenesis is depicted in Figure **1.15**.

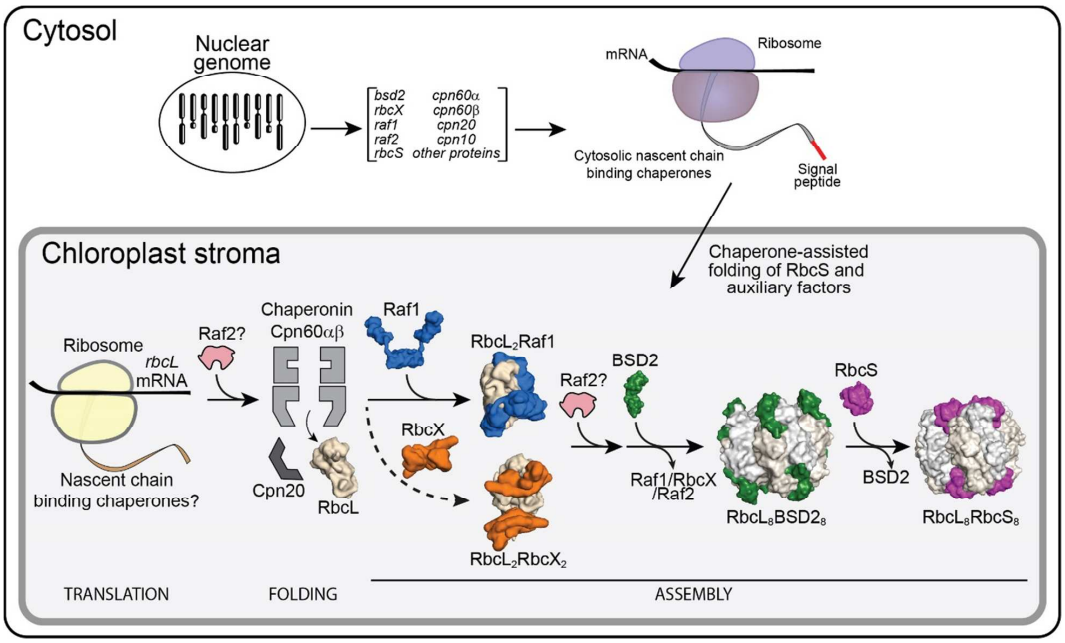

Figure 1.15.: Molecular chaperones involved in plant Rubisco biogenesis. While RbcL is chloroplast-encoded, the remaining factors participating in Rubisco holoenzyme formation (e.g. the chaperonins, assembly factors and RbcS) are nuclear-encoded and have to be imported into the chloroplast. The import into the chloroplast proceeds in an unfolded state [Jarvis and Kessler, 2014]. The transit peptide, ensuring targeting to the chloroplast, is cleaved and folding to the native state may be assisted by stromal chaperones. Upon folding of RbcL by the stromal chaperonins Cpn60α/β / Cpn20, RbcL may assemble into dimers and higher oligomers assisted by Raf1 and/or RbcX. BSD2 displaces Raf1 and/or RbcX and stabilizes the octameric RbcL core in an assembly competent state. Following the import of RbcS, RbcS displaces BSD2 and forms the functional holoenzyme. Although being required for plant Rubisco biogenesis in *E. coli*, the exact role of Raf2 is not yet clear. Reprinted (adapted) with permission from Wilson and Hayer-Hartl [2018]. Copyright 2018 American Chemical Society.

In plants and algae, photosynthesis and Rubisco assembly is restricted to the chloroplast, while many components are encoded in the nucleus, rendering Rubisco biogenesis more complex. While RbcL is encoded in the chloroplast genome (plastome), the family of RbcS homologues and the assembly factors identified so far are all nuclear-encoded and have to be imported into the chloroplast. The spatial separation of gene expression of subunits (and assembly factors), which participate in

the formation of the Rubisco holoenzyme, requires a tight coordination to ensure their timely expression and to maintain the balance in the amounts of the subunits and assembly factors. A prerequisite therefore is sophisticated signaling between the chloroplast and the nucleus of the plant cell. Additional processes are in place to achieve the concerted accumulation of factors inside the chloroplast. For example, free RbcS has been shown to be subject to rapid degradation inside the chloroplast [Schmidt and Mishkind, 1983]. Moreover, some chloroplast-encoded proteins - among them RbcL - of green algae and plants have been shown to be subject of assembly-dependent translational regulation [Khrebtukova and Spreitzer, 1996; Wostrikoff and Stern, 2007]. Assembly-dependent translational regulation refers to a mechanism, whereby accumulation of one assembly partner affects the translation of the other, which is also termed control by epistasy of synthesis (CES).

Folding of newly-synthesized plant and algal RbcL is mediated by the chloroplast homolog of bacterial chaperonin, Cpn60. Cpn60 is a heterooligomeric complex composed of Cpn60α and Cpn60β subunits that cooperate with the lid-like cofactors Cpn20 and Cpn10 [Gruber et al., 2013]. The assembly chaperones RbcX and Raf1 are conserved in plants, however the genome of plants may encode multiple homologs of both proteins. E.g. *Arabidopsis thaliana* encodes two RbcX proteins (RbcX1 and RbcX2) that were both shown to enhance the expression of cyanobacterial Rubisco but are differentially expressed and localized *in vivo* [Kolesiński et al., 2011] and both *A. thaliana* and *N. tabacum* encode two Raf1 homologous (Raf1.1 and Raf1.2) [Whitney et al., 2015]. Apart from RbcX and Raf1, plants and algae encode the additional assembly chaperones Raf2 and Bundle Sheath Defective 2 (BSD2).

1.3.2.3.1. The assembly chaperone Raf2

Like Raf1, Raf2 was identified in a Photosynthetic Mutant Library in maize and its deletion is lethal at the seedling stage [Stern et al., 2004]. *raf2* is unrelated to *raf1* and has distant homology to a pterin-4α-carbinolamine dehydratase (PCD). However, Raf2 of several plant and algal species lack PCD activity [Naponelli et al., 2008] and forms multimers. Because *raf2* mutant maize displayed a reduced Rubisco content despite of unaltered levels of transcription and translation of Rubisco subunits, it was proposed to be implicated in Rubisco assembly [Feiz et al., 2014]. Accordingly, the majority of RbcL was found to be associated with the chloroplast chaperonins in *raf2* mutant maize. Interestingly, co-immunoprecipitation experiments after crosslinking, suggested that Raf2, but also Raf1 and BSD2, preferentially associate with RbcS rather than with RbcL. This refreshed the discussion whether RbcS, which refolds spontaneously *in vitro*, requires chaperone assistance

upon import in the chloroplast. However, a mechanism for Raf2 is still elusive. Interestingly, homologs of Raf2 (also acRaf, α-carboxysome Rubisco accumulation factor) have also been identified in all prokaryotes expressing form IAc Rubisco, including α-cyanobacteria. There, the *raf2* gene, is encoded in the α-carboxysomal operon together with *rbcL* and *rbcS* [Wheatley et al., 2014]. AcRaf, like its eukaryotic homolog, lacked PCD activity but increased the yield of soluble Rubisco heterologously expressed in *E. coli*.

1.3.2.3.2. The assembly chaperone BSD2

BSD2 was first identified in maize by Roth et al. [1996] as the cause for an aberrant bundle sheath chloroplast structure phenotype in a mutant maize plant. *bds2* mutant maize is seedling-lethal and has been reported to contain neither RbcL nor RbcS protein despite of normal transcription of the corresponding genes. Hence a role for BSD2 in post-transcriptional control of rbcL transcript accumulation or translation was suggested. However, *in vitro* data demonstrated that BSD2, a monomer of $\sim$ 10 kDa, acts as an assembly chaperone of Rubisco with eight BSD2 binding to one $RbcL_8$ octamer core [Aigner et al., 2017]. BSD2-chaperoned $RbcL_8$ was proposed to represent the end-state assembly intermediate from which BSD2 is displaced by RbcS. BSD2 is thought to stabilize the octameric core of L_8 and prevent its aggregation in a manner similar to cyanobacterial RbcX. Notably, while Raf1, Raf2 and RbcX all have homologs in cyanobacteria, BSD2 is the only assembly chaperone specific to chloroplasts. Hence, in agreement with the absence or unexpectedly low protein levels of RbcX in chloroplasts [Feiz et al., 2012], BSD2 may diminish the role of RbcX in plants. Consistently, intermediary Rubisco-BSD2 assembly complexes have been reported *in vivo* [Conlan et al., 2019]. However, the mechanistic details regarding the role of BSD2 in transcriptional regulation of RbcL as proposed by [Brutnell et al., 1999; Doron et al., 2014; Roth et al., 1996] are still elusive.

Concomitant with the mechanistic details of BSD2 function, Aigner et al. [2017] presented the first heterologous expression of plant (*Arabidopsis thaliana*) Rubisco in *E. coli*, which constitutes a major achievement towards the manipulation of plant Rubisco. However, as of now, four years after publication of the first heterologous plant Rubisco synthesis, *in vitro* reconstitution of plant Rubisco has not been reported.

1.3.3. Reaction mechanism

1.3.3.1. Rubisco activation

In order to become catalytically active, Rubisco has to undergo an activation process, commonly referred to as "carbamylation". Activation requires the covalent attachment of a non-substrate CO_2 to the conserved Lys201 (plant numbering) and subsequent coordination of a Mg^{2+} ion [Taylor and Andersson, 1996].

A possible mechanism of Rubisco activation was proposed by Stec [2012] (Figure **1.16**): Access to the active site of Rubisco is gated by the conserved residue His335 (His327 in plants), which can adopt two alternative conformations. The outward rotation of His335 allows entry of Mg^{2+} surrounded by two hydroxyl groups and one water molecule, that are stabilized by His300 and His302 and Gln409 (His292, His294 and Gln401 in plants). Upon binding of the activating, non-substrate CO_2 to the Mg^{2+} ion, the ϵ-amine group of Lys210 transfers a proton to one of the Mg^{2+}-bound hydroxyl groups and its N-atom locates in proximity to the carbon atom of the CO_2. Deprotonation of the ϵ-amine group of Lys210 (Lys201 in plants) by the remaining Mg^{2+}-bound hydroxylgroup, induces the nucleophilic attack of the ϵN-atom on the CO_2 molecule and results in the formation of the carbamate. Destabilization of the Mg^{2+}-ion and a conformational change in the carbamyl group of Lys210 induce the relocation of the Mg^{2+}-ion, which is coordinated by the carboxyl groups of Asp212 and Glu213 (Asp203 and Glu204 in plants) in the activated enzyme.

1.3.3.2. The carboxylation reaction

Rubisco functions as a two-state enzyme, with an active site that can adopt an open or closed conformation (Figure **1.17 A** and **B**) [Taylor and Andersson, 1996]. Binding of RuBP to the active site involves interactions with both phosphates and between the substrate backbone and the activating Mg^{2+}-ion (Figure **1.17 C**) [Taylor and Andersson, 1997]. A model for Rubisco active site closure in response to substrate binding has been suggested by Duff et al. [2000]. Structural constraints enforced by the productive interaction of RuBP with the Mg^{2+}-ion and one of the phosphate binding sites require the RuBP backbone to bend whereby the degree of the curvature is determined by the bound Mg^{2+}-ion [Duff et al., 2000]. The bent conformation of RuBP brings both phosphate binding sites in close distance, which is thought to induce the quasi-rigid body rotation of the N-terminal domain towards the α/β barrel domain required for active site closure. Rotation of the N-terminal

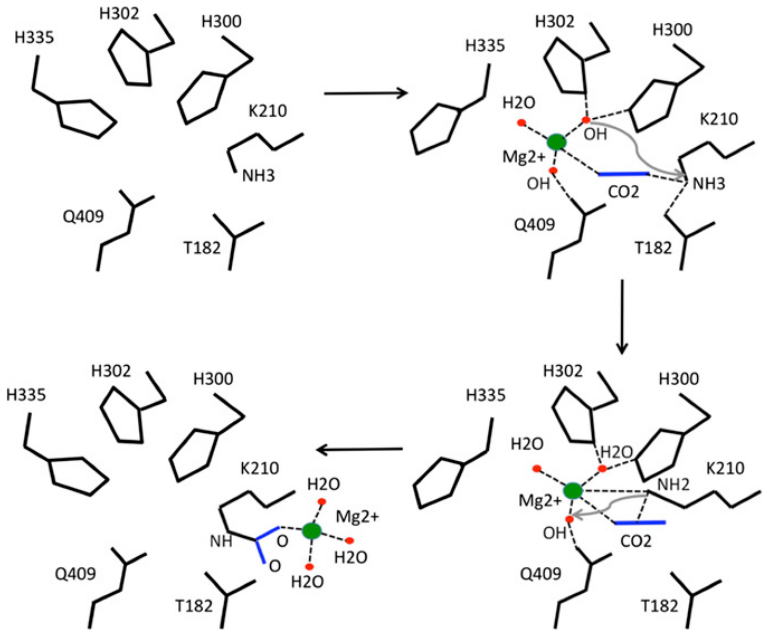

Figure 1.16.: Model of Rubisco activation.
The model of Rubisco activation by carboxylation of Lys210 (Lys201 in plants) and subsequent Mg^{2+} coordination is based on crystal structures of Rubisco from *Galdieria partita*. Residue numbering corresponds to this Rubisco enzyme. See text for explanation. Figure adapted from [Stec, 2012].

domain allows loop 6 of the adjacent subunit to fold over the active site. The N-terminal peptide folds on top of the 60s loop, and an Asp latch site (Asp473 in plant RbcL or Asp474 in *Nos*RbcL), which is conserved at the C-terminus of form I Rubiscos, locks the C-terminal peptide of the adjacent subunit in an extended position above the active site. Thereby loop 6 is pinned down in a stably closed conformation, contributing to the active site environment that facilitates catalysis (Figure **1.17 B**).

The conversion of RuBP into two molecules of 3PGA occurs mostly by acid-base chemistry in five steps. The participation of catalytic residues in the carboxylation reaction has been described by Taylor and Andersson [1997] on the basis of the structure of spinach Rubisco (refer to Figure 7 in Taylor and Andersson [1997], who display the full model including participating active site residues). The basic reaction mechanism of RuBP carboxylation and oxygenation are depicted in Figure **1.18**: Briefly, in a first step, C3 of RuBP is deprotonated, forming an unstable enediolate intermediate. The carbamate group of Lys201 is the most likely candidate serving as a catalytic base in this process. Next, the physico-chemical environment established in the active site with participation from the lid-like loop 6 contributes to the polarization of the incoming CO_2 and facilitates its electrophilic attack on C2 of RuBP. Activation of a nearby water molecule, likely involving the carbamate group

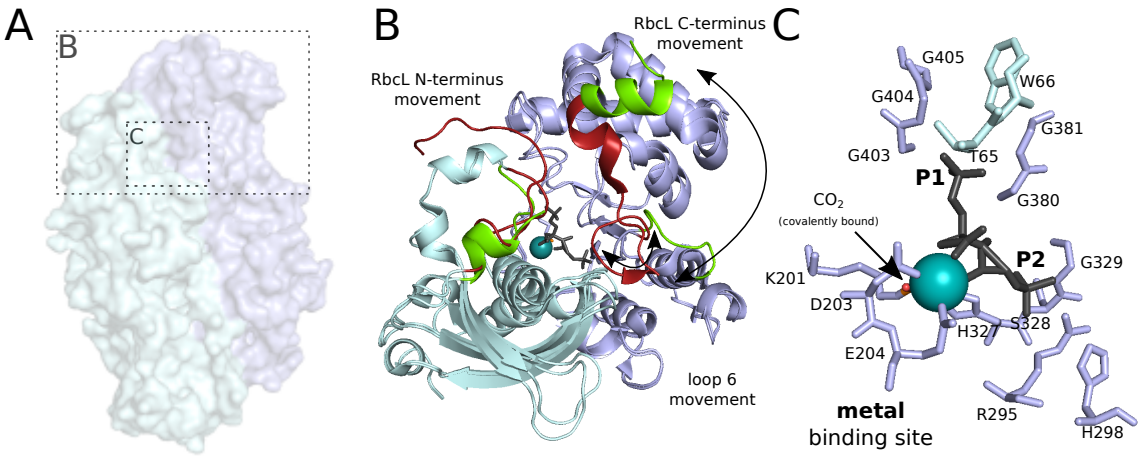

Figure 1.17.: Rubisco active site architecture.
A) The antiparallel RbcL dimer of plant Rubisco in surface representation. Individual chains are colored in pale cyan and light blue. Rectangles mark the position of close up views on B and C. **B)** Alignment of RbcL dimers of Rubisco from *N. tabacum* in open and closed state. Major differences between the closed and open state are highlighted in red and green, respectively. The backbones of protein chains are shown as ribbons. **C)** Close up of the bound CABP in the closed state. Residues constituting the major contact sites (P1, metal site and P2) are labeled. The bound CABP and the Mg^{2+}-ion are shown in grey stick representation and as sphere in deep teal, respectively. The covalently attached, activating CO_2 is highlighted in orange. PDB: 4RUB (closed state) and 1EJ7 (open state).

and His325, allows for its nucleophilic attack on C3, which completes the formation of the short-lived, branched six carbon sugar 2-carboxy-3-keto-D-arabinitol. The keto group in this compound undergoes a hydration to a geminal diol at C3. The following rapid cleavage of the C2-C3 bond is initiated by deprotonation of the hydroxy groups at C3 and the resultant protons are likely abstracted by Lys201 and His294. While this process yields one 3PGA molecule directly (from C3-C5), the other has to be generated by stereospecific protonation. The position of the charged Lys175 at the opposite site of the substrate renders it a likely proton donor in this process.

1.3.3.3. The oxygenation reaction

As outlined before, the highly reactive enediolate required for carboxylation is also prone to oxidation, which results in the generation of a hydroperoxy derivative that breaks down to 2PG and 3PGA. A potential mechanism of Rubisco-catalyzed RuBP oxygenation has been suggested by [Lorimer, 1981] and reviewed by Tcherkez [2016]. Only recently, the proposed mechanism of O_2 activation by single electron transfer, whereby the RuBP enolate reduces molecular O_2 which yields a radical intermediate that recombines to form RuBP peroxide, was experimentally supported by Bathellier et al. [2020].

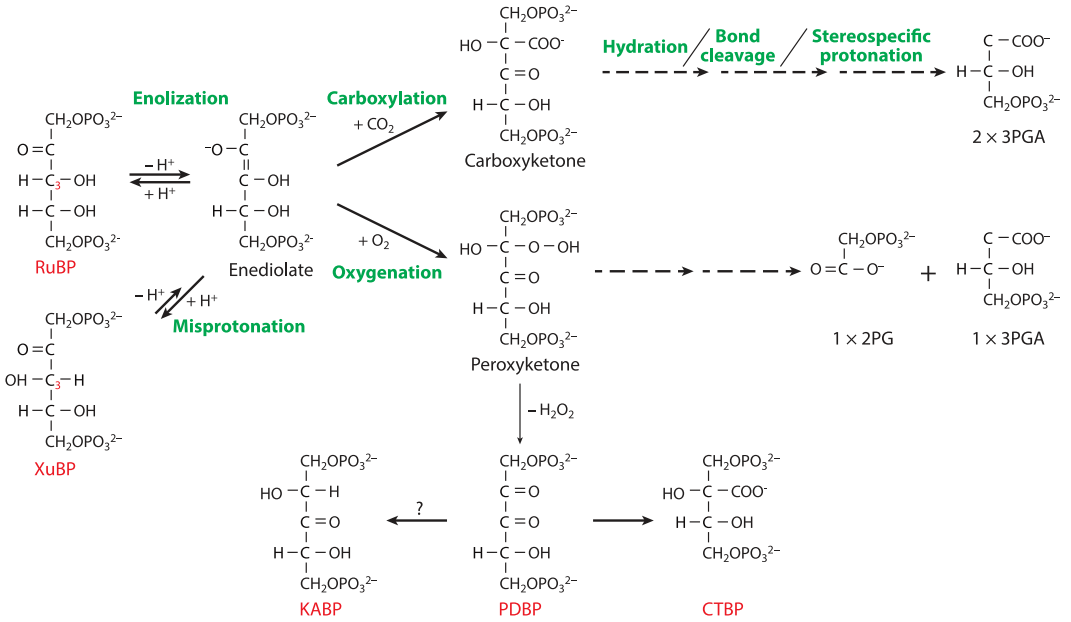

Figure 1.18.: Rubisco catalyzed carboxylation and oxygenation of RuBP. Critical reaction intermediates of the carboxylation and oxygenation reaction catalyzed by Rubisco are shown. Carboxylation proceeds via RuBP (ribulose 1,5-bisphosphate) enolization, carboxylation, hydration, bond cleavage and stereospecific protonation and produces two molecules of 3PGA, whereas substrate oxygenation yields one molecule of 3PGA and one molecule of 2PG. Their common reaction intermediate, the enediolate, is highly reactive and prone to misprotonation, which results in formation of the inhibitory XuBP (D-xylulose 1,5-bisphosphate). Peroxyketone, a reaction intermediate of the oxygenation reaction may result in the generation of additional inhibitory sugar phosphates - PDBP (D-glycero 2,3-pentodiulose 1,5-bisphosphate), CTBP (2-carboxytetritol 1,4-bisphosphate) and KABP (3-ketoarabinitol 1,5-bisphosphate). Inhibitory compounds are indicated in red. Notably, RuBP binding to the inactive catalytic pocket prevents the carbamate formation and thus, may have an inhibitory effect on Rubisco. Figure adapted from Bracher et al. [2017] with permission.

1.3.3.4. Rubisco inhibition by misfire products

Rubisco catalysis is prone to the generation of "misfire" byproducts, inhibitory sugar phosphates, that impair Rubisco catalysis (Figure **1.18**). Importantly, the deprotonation leading to the formation of the reactive enediolate in the first step of Rubisco catalysis is reversible and mis-protonation can result in the formation of D-xylulose 1,5-bisphosphate (XuBP) [Edmondson et al., 1990]. Although XuBP can be carboxylated [Yokota, 1991], deprotonation of XuBP is comparatively slow and thus impairs catalysis by active site blockage. XuBP impairs Rubisco activity both by binding to the carbamylated and unmodified catalytic site. In the former case, XuBP acts as a competitive inhibitor to the RuBP substrate, whereas in the latter case, it impedes Rubisco activation by prevention of CO_2 and Mg^{2+} binding. Moreover, RuBP itself can impede Rubisco activation by binding to the unmodified Rubisco catalytic site [Jordan and Chollet, 1983]. H_2O_2 elimination from the peroxyketone intermediate of the oxygenation reaction gives rise to another misfire product, D-glycero 2,3-

pentodiulose 1,5-bisphosphate (PDBP). PDBP accumulates at the active site and causes a decline of the carboxylation rate under substrate saturating conditions, termed "fallover". Notably, Kane et al. [1998] demonstrated that PDBP may also form non-enzymatically by oxidation of RuBP in presence of metals. Early studies did not account for this non-enzymatic PDBP formation which very likely lead to an overestimation of the importance of PDBP caused fallover. More recent studies estimated PDBP formation to occur once in 230 to 260 turnovers [Kim and Portis, 2004; Pearce, 2006]. For comparison, the rate of XuBP formation was $\sim$ 4-fold higher in similar experiments [Pearce, 2006], likely rendering XuBP inhibition more important. An additional inhibitor of Rubisco activity, 2-carboxy D-arabinitol 1-phosphate (CA1P) is no erroneous byproduct of Rubisco catalysis, but instead is generated from 2-carboxy D-arabinitol (CA) by phosphorylation in the dark [Moore and Seemann, 1992]. CA1P accumulates in the chloroplasts [Moore et al., 1995] and tightly binds to the carbamylated active site of Rubisco *in vivo* [Moore and Seemann, 1994]. Given the variable production of CA1P in different species, its importance is likely species specific [Moore et al., 1991]. Tight binding inhibitors of Rubisco have been proposed to confer resistance against proteolytic breakdown, because inhibition by CA1P protected Rubisco from proteolytic cleavage by chloroplast proteases *in vitro* [Khan et al., 1999] and protection against oxidative stress has been suggested [Parry et al., 2008]. Other less extensively studied Rubisco inhibitors comprise 2-carboxytetritol 1,4-bisphosphate (CTBP) and 3-ketoarabinitol 1,5-bisphosphate (KABP) (reviewed by Parry et al. [2008]). Notably, the artificial analogue of the carboxyketone intermediate of the carboxylation reaction 2-carboxyarabinitol 1,5-diphosphate (CABP) is the most tightly binding inhibitor with a dissociation constant $K_d \sim 10^{-11}\,\mathrm{mol\,L^{-1}}$ [Pierce et al., 1980] and thus frequently used in structural studies for its ability to rigidify the active site of Rubisco leading to an overall more compact enzyme.

Efficient reversal of Rubisco inhibition by removal of the inhibitory sugar phosphates from the catalytic pocket requires the action of Rubisco activase (Rca) (see Chapter **1.4.6**). Notably, inhibitors with high affinity to the carbamylated active site would eventually rebind upon removal by Rca, rendering the reactivation process inefficient. Interestingly, selective phosphatases for both CA1P and XuBP exist. The XuBPase, product of the *cbbY* gene of the chemoautotrophic bacterium *R. sphaeroides*, was found to selectively degrade XuBP in presence of millimolar concentration of RuBP [Bracher et al., 2015b]. Conservation of the XuBPase in some cyanobacteria, green algae, red algae and plants indicates that XuBP production is likely a problem for all autotrophs that rely on the CBB cycle, although the rate of XuBP formation varies between Rubiscos of different forms [Pearce, 2006]. CA1P is dephosphorylated by the

selective plant phosphatase CA1Pase in response to the redox state of the cell, which is governed by the light dependent activity of the photosynthetic electron transport chain [Heo and Holbrook, 1999; Holbrook et al., 1989]. The same phosphatase has also been implicated in the degradation of PDBP [Andralojc et al., 2012].

1.4. The AAA+ protein Rubisco activase

1.4.1. AAA+ proteins

Members of the protein family of "ATP hydrolyzing enzymes (ATPase) associated with various cellular activities" (AAA+ proteins) are involved in a plethora of cellular processes such as DNA replication, vesicle formation, membrane fusion, protein homeostasis, cytoskeletal rearrangements, assembly and disassembly of macromolecular complexes and more [Hanson and Whiteheart, 2005; Puchades et al., 2020]. AAA+ proteins assemble to homo- or heterooligomeric complexes and their functional state is often a ring-shaped hexamer with a central pore. Despite of their functional diversity, all AAA+ proteins comprise one or multiple conserved ATPase domains for the conversion of chemical energy, derived from ATP hydrolysis, into mechanical force. Eventually, the concomitant domain rearrangements may be transferred onto a macromolecular substrate molecule and result in its remodeling. Since the nucleotide binding site of AAA+ proteins is located at the interface of the AAA+ domains between adjacent subunits in the oligomeric complex, individual subunits always function in cooperation.

AAA+ proteins belong to the "additional strand, conserved glutamate" (ASCE) subgroup of the P-loop type nucleoside triphosphate (NTP)-binding and hydrolyzing proteins (P-loop NTPase). P-loop NTPases constitute between 10 and 18 % of the predicted proteins of prokaryotic and eukaryotic genomes [Koonin et al., 2000] and their nucleotide binding fold has been described as the most prevalent fold in the protein domain universe [Leipe et al., 2003]. P-loop NTPases are of monophyletic origin and are found in all three domains of life. Due to their ancient origin, the sequence homology between P-loop NTPases (and also AAA+ proteins) is vestigial, but they share a structurally similar core fold. As first pointed out by Walker et al. [1982], the only common hallmark manifested in primary sequence of all P-loop NTPases are the P-loop or Walker A and the Walker B sequence motifs - both of which are conserved among P-loop NTPases and implicated in nucleotide binding and hydrolysis. AAA+ proteins, like all members of the ASCE group, are distinguished

from other members of the P-loop NTPases (nucleotide kinases and GTPases) according to the presence of an additional strand between the β-sheet leading to the Walker A loop and the Walker B strand as well as a conserved glutamate at the end the Walker B motif [Leipe et al., 2002].

A distinguishing feature of AAA+ proteins is the presence of a small α-helical domain, also termed α-helical bundle or lid, C-terminal to the larger nucleotide binding α/β-subdomain [Neuwald et al., 1999]. Despite their high sequence diversity, the small α-helical domains preserved a core fold consisting of two helical hairpins, each connected by short strap loops that form a parallel two-stranded β-sheet with one another, arranged in a left-handed superhelix [Ammelburg et al., 2006]. The α/β subdomain has a Rossman fold architecture typical for AAA+ proteins, which resembles a three-layered sandwich, where the outer β-sheet connecting α-helices, clasp around the central, parallel, five-stranded β-sheet in an β5-β1-β4-β3-β2 arrangement (Figure **1.19**). Unlike other members of the ASCE group, AAA+ proteins do not deviate from this central β-sheet arrangement and do not include additional β-sheet strands [Erzberger and Berger, 2006].

Fusion of the AAA+ domain with a multitude of domains of distinct function throughout evolution lead to an increase of functional diversity among AAA+ domain containing proteins [Puchades et al., 2020]. According to the number of AAA+ domains contained, AAA+ proteins with one or two domains are classified as type I or type II ATPases, respectively. On basis of their sequence and structure, the AAA+ proteins have been divided into several clades: clamp loader, initiator, classic and the pre-sensor 1 β-hairpin superclade (ps1βh) which comprises the superfamily helicase III, HCLR, helix-2-insert (h2i) and pre-sensor 2 (ps2) clades [Ammelburg et al., 2006; Iyer et al., 2004]. Notably, in contrast to Iyer et al. [2004], the SF3 helicases are excluded by Ammelburg et al. [2006] due to the presence of a C-terminal domain unrelated to the core topology of the small α-helical subdomain. While the minimal topology of AAA+ proteins outlined above, represents the most basic clade of AAA+ proteins, the clamp loader clade, each of the other subclades contains clade-specific insertions of specific secondary structural elements within the α/β subdomain. Irrespective of the clade, the small α-helical domains are more variable in both sequence and structure and often contain additional insertions, deletions and replacements in their core fold.

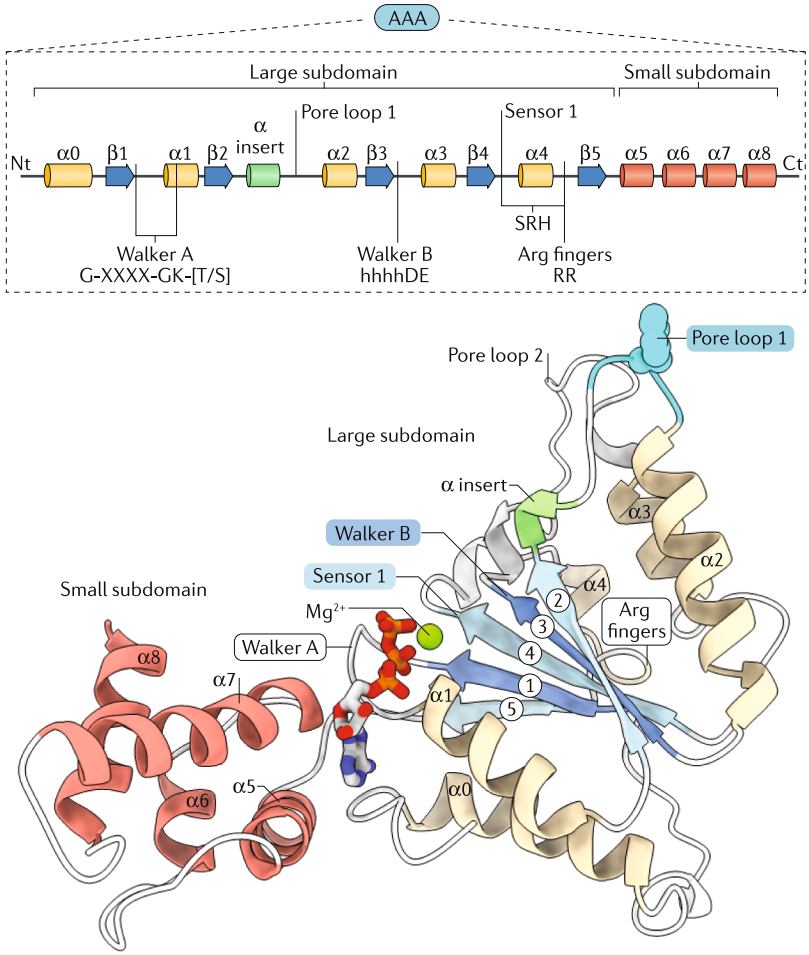

Figure 1.19.: The conserved organization of the AAA+ domain of classic clade AAA+ ATPases.
The secondary (**top**) and tertiary structure (**bottom**) of the classical AAA+ domain are depicted in corresponding color schemes. Large (α/β-) subdomain and small (α-helical) subdomain as well as Walker A, Walker B and the second region of homology (SRH) including sensor-1 and the arginine finger motif are indicated. The atomic model is based on YME1 (pdb:6AZ0). Figure reproduced from Puchades et al. [2020].

1.4.2. Defining features of AAA+ proteins

1.4.2.1. Walker A and B

Walker A and B are the shared hallmarks of all P-loop NTPases. Most often, the Walker A motif consists of a Gly-X-X-X-X-Gly-Lys-[Thr/Ser] sequence, where X denotes any amino acid and the last residue is either a threonine or serine [Walker et al., 1982]. The Walker A motif is located in a loop (P-loop) between β1 and α1 within the AAA+ minimal topology (Figure **1.19**). Although strongly conserved in AAA+ proteins, some AAA+ members display clade specific deviations from this motif. The Lys residue of the Walker A motif has been shown to directly interact

with the phosphate groups of the bound nucleotide and the Thr residue helps in the coordination of the Mg^{2+}-ion, together with Walker B residues [Liu et al., 2000]. Walker B comprises a hydrophobic strand, β3, of the central β-sheet and follows the sequence pattern h-h-h-h-Asp-[Asp/Glu], where h denotes any hydrophobic amino acid and the last residue is either Asp or Glu. The Asp or Glu located at the C-terminal end of β3 acts as catalytic base, activating a water molecule for nucleophilic attack on the γ-phosphate of the bound nucleotide (Figure **1.20 C**). Additionally, the catalytic glutamate of some AAA+ proteins can occupy either a hydrolysis competent or incompetent state, which is linked to the substrate binding pore loops via sidechain interactions. This coupling of substrate binding to ATPase activity allows many AAA+ proteins to enhance their ATPase activity upon substrate engagement. Due to its ability to switch between the catalytically competent and incompetent states, this mechanism is often referred to as "glutamate switch" [Zhang and Wigley, 2008]. Because of its role in catalysis, the catalytic Walker B residue is frequently mutated in experiments to distinguish between the effects of ATP hydrolysis and ATP binding on AAA+ activity, e.g. [Kress et al., 2009; Weibezahn et al., 2003]. Likewise, the mutation of the Lys residue in the Walker A motif abolishes nucleotide binding and is often introduced to analyze the role of nucleotide binding on the state of the AAA+ domain, e.g. [Singh and Maurizi, 1994].

1.4.2.2. Second region of homology

In addition to Walker A and B motifs, AAA+ proteins feature an additional homologous region of about $\sim$ 20 residues, the second region of homology (SRH), located between β4 and β5, spanning the entire helix α4 (Figure **1.19**) [Erzberger and Berger, 2006]. This contains both the sensor 1 and the arginine finger motifs, that coordinate nucleotide hydrolysis and inter-subunit communication (Figure **1.20 C**). Sensor 1 is a polar residue that assists in the coordination of the water molecule required for nucleophilic attack on the γ-phosphate. Sensor 1 is located at the C-terminal end of β4 and is often Asn [Miller and Enemark, 2016]. The Arg finger is located in the loop following helix α4. In contrast to the *cis*-acting Walker A/B and sensor 1, the Arg-finger(s) acts in *trans* (i.e. from the adjacent subunit) by reaching into the nucleotide binding site of the adjacent subunit, interacting with the γ-phosphate of the bound nucleotide and is thought to stabilize a transition state during hydrolysis [Ogura et al., 2004] (Figure **1.20 C**).

1.4.2.3. Sensor 2, 3 and the N-linker

Sensor 2 is conserved in all AAA+ proteins, apart from members of the classic clade [Miller and Enemark, 2016]. It is a single positively charged residue, Arg or Lys, located at the base of the helix $\alpha 7$ (Figure **1.19**). It directly interacts with the α-phosphate of the bound nucleotide and its mutation affects nucleotide binding as well as the propagation of the nucleotide state between adjacent subunits [Enemark and Joshua-Tor, 2006]. Sensor 2 is *cis*-acting, unless the C-terminal lid deviates from the canonical fold of the α-helical subdomain, in which case sensor 2 might be positioned as a *trans*-acting residue, e.g. in case of the MCM helicase [Miller et al., 2014]. Notably, according to the classification by Ammelburg et al. [2006], MCM helicase would not constitute a member of the AAA+ protein family, due to this deviation from the small subdomain core fold. Some helicases, among them MCM, have been shown to possess a third *trans*-acting sensor residue that is either His or Arg, which was shown to interact with the Walker B catalytic base. Sensor 3 is thought to stabilize the tight inter-subunit interface in the ATP-bound state. [Enemark and Joshua-Tor, 2006; Li et al., 2015]. The N-linker, a conserved sequence motif just preceding the N-terminal helix $\alpha 0$, has been reported to strongly correlate with the SRH. AAA+ members that contain two Arg residues in the SRH, now members of the classic clade of AAA+ proteins, often have a Gly-Gly motif as N-linker, whereas AAA+ members with mostly one Arg in this region have a hydrophobic-Gly motif preceding $\alpha 0$. Functionally, the N-linker contributes to the nucleotide binding site and is thought to be involved in the propagation of nucleotide-dependent conformational changes to distant elements [Hanson and Whiteheart, 2005; Smith et al., 2004].

1.4.3. Classic clade of AAA+ proteins

Members of the so-called classic clade belong to the structurally and mechanistically best described AAA+ proteins. Prominent members of the classic clade are e.g. the microtubule severing enzymes katanin and spastin [Roll-Mecak and Vale, 2008; Zehr et al., 2017], the membrane protease YME1 [Puchades et al., 2017], Cdc48 / p97 involved in the extraction of polyubiquitinated membrane proteins [Cooney et al., 2019; Stach et al., 2020; Twomey et al., 2019], the SNARE-disassembly protein NSF, which is involved in synaptic vesicle fusion [White et al., 2018; Zhao et al., 2015], or the proteasome ATPase subunits [de la Peña et al., 2018]. Green-type Rca from plants also belongs to the classic clade of AAA+ proteins.

AAA+ domains of proteins within the classic clade are characterized by a short α-helical insertion between $\alpha 2$ and $\beta 2$ [Miller and Enemark, 2016] (Figure **1.19**). In the hexameric ring, this helix forms a loop in proximity to the central solvent channel and precedes the pore loop 1 (PL1), an element involved in substrate binding and translocation, which is conserved in most members of the classic clade [Deville et al., 2019; Germany et al., 2018; Han et al., 2017; Puchades et al., 2017; Sandate et al., 2019; Zehr et al., 2017]. Moreover, the Arg residue of the sensor 2 motif at the base of helix $\alpha 7$ is missing in members of the classic clade and frequently replaced by an Ala residue [Ogura et al., 2004]. Unlike non-classic clade AAA+ proteins, members of the classic clade contain two Arg residues, which function as arginine fingers within the SRH and can be subdivided in the FtsH-, katanin-, TIP49-, AFG1-, Proteasomal-, NSF/Cdc48/Pex-, and ClpABC- protein families [Miller and Enemark, 2016]. Notably, the katanin protein family as a subclade of the classic AAA+ ATPases is synonymously used with the meiotic clade, which was named in reference to the common functional role of its founding members - katanin, spastin and fidgetin - in meiosis.

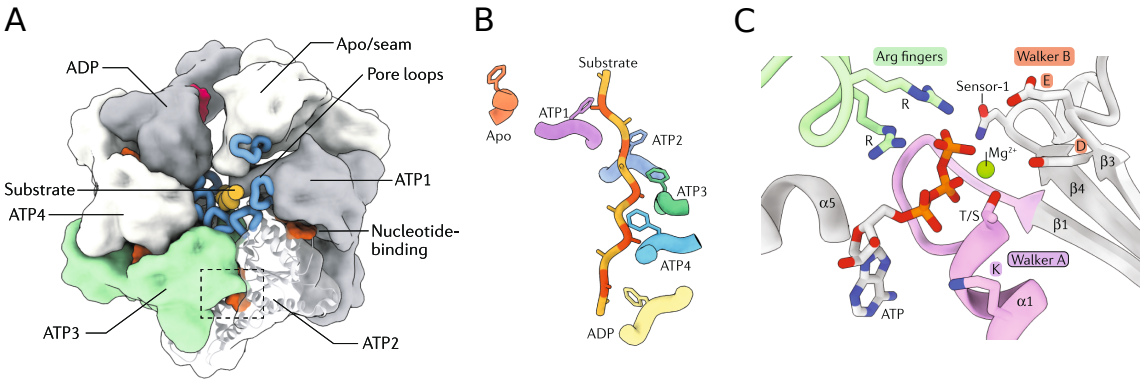

Figure 1.20.: Organization of the classical substrate-bound AAA+ ATPase. **A) Overview of the hexameric ATPase.** Most AAA+ hexamer adopts a ring-like, right-handed spiral, with the pore loops (blue) encircling the bound substrate (orange). The nucleotide-binding sites are formed at the inter-subunit interfaces and the nucleotide (brown) state of each subunit is highlighted (bound to ATP (ATP1-4), bound to ADP, nucleotide-free (apo)). The dashed rectangle indicates the nucleotide-binding site enlarged in **C**. **B) The pore-loops 1 reflect the spiraling organization of the hexamer.** Every two residues PL1 aromatic residues of the ATP-bound subunits intercalate with the substrate backbone. The conformation of PL1 is highly correlated with the nucleotide state. The ADP-bound and apo subunits have limited or no contact to the substrate, respectively. **C) Organization of the nucleotide-binding site.** Walker A, Walker B, arginine-fingers, Sensor-1 and conserved residues are highlighted. The arginine-fingers of subunit ATP3 (green) protrude into the nucleotide-binding site in *cis*. Figure modified from Puchades et al. [2020]

1.4.4. Substrate translocation mechanisms

Structural and functional commonalities of mostly substrate translocating AAA+ proteins have stimulated a quest for a common mechanism of force generation and substrate translocation in this class of proteins.

1.4.4.1. AAA+ spiral formation and pore loop function

As pointed out by Puchades et al. [2020], all structures of substrate bound AAA+ proteins of the classic clade have been found to adopt a spiraling conformation. In these structures, the pore loop 1 (PL1), an aromatic-hydrophobic sequence motif conserved in most classic clade members, that forms a loop lining the central solvent channel of the hexameric AAA+ ring, was found to intercalate with the translocating peptide backbone (Figure **1.20 A** and **B**). For example, in type II AAA+ proteins containing both an ATPase active and inactive AAA+ domain, only the ATPase active AAA+ domain adopts a spiraling conformation around the pore-loop engaged substrate peptide, whereas the ATPase inactive AAA+ domain forms a symmetric planar ring above or below the substrate engaged spiral [Blok et al., 2015; Gardner et al., 2018; White et al., 2018]. In fact, the ATPase active type I ATPase Cdc48 was shown to form planar rings in the absence of substrate, whereas it adopted a spiraling conformation in the substrate bound state, indicating that the binding of substrate itself may determine the quaternary arrangement of the AAA+ domain [Cooney et al., 2019; Twomey et al., 2019]. The PL1-peptide interactions are mostly mediated by hydrogen bonds with the peptide backbone and by steric interactions with the intercalating aromatic residue, allowing for peptide translocation largely irrespective of sequence or directionality. Mutation of the PL1 aromatic residue renders AAA+ translocases translocation inactive, while basal ATPase activity is mostly unaffected.

1.4.4.2. Nucleotide state and models for substrate translocation

High resolution structures of substrate bound AAA+ proteins have revealed a correlation between the nucleotide state of a given subunit and its position within the hexameric spiral and the PL1 conformation. While ATP-bound subunits intercalate with the substrate peptide, ADP-bound subunits establish marginal or no contacts at all with the peptide and are mostly found in the lowermost subunit (positioned furthest from the plane which is perpendicular to the central axis and

located between the uppermost AAA+ domain and the substrate molecule - Figure **1.20 B**)). These observations lead to the proposal of a "hand-over-hand" model of substrate translocation, whereby four subunits ("hands") at a time are peptide engaged, whereas the lowermost is hydrolyzing ATP, thereby weakening the interface to the adjacent subunit, which ultimately leads to its detachment and transitioning to the uppermost position. This model is also referred to as "sequential clockwise / two residue steps" (SC/2R), because subunit movement is suggested to proceed in clockwise direction (with ATP-hydrolysis proceeding in anti-clockwise direction) and with two substrate residues translocated per ATP hydrolysis event [Fei et al., 2020]. While ATP hydrolysis and phosphate release are thought to be restricted to the lowermost position, nucleotide exchange is bound to occur during the transition step in this model. Following ADP release and transition to the upper position, the pore loop would intercalate with the peptide at the uppermost position and ATP binding would stabilize the interface to the adjacent subunit captured in the context of the spiral [Puchades et al., 2020]. According to this model, every subunit would cycle through each of the 6 positions, solely by vertical movement parallel to the central axis of the spiral. This concomitant movement of subunits is thought to be caused by interdomain rotations between the large and small subdomains with respect to each other.

However, the hand-over-hand or SC/2R model is still controversial, given that it is primarily based on structural snapshots of substrate engaged AAA+ motors. Notably, for the classic clade AAA+ ATPase Vps4, both a clockwise [Han et al., 2017; Monroe et al., 2017] and anticlockwise [Su et al., 2017] sequential ATP hydrolysis model has been proposed. Moreover, cryo-EM snapshots of the biochemically extensively studied HCLR-type AAA+ ATPase ClpX resemble structures of other classic clade AAA+ proteins, for which a hand-over-hand model has been proposed [Dong et al., 2019; Fei et al., 2020; de la Peña et al., 2018; Puchades et al., 2017]. However, a large body of biochemical and biophysical evidence collected across the ClpX community is incompatible with a strict hand-over-hand mechanism. The step size of $\sim$ 6 - 24 residues per ATP hydrolyzed in ClpX quantified by optical trapping experiments is larger than expected [Aubin-Tam et al., 2011; Maillard et al., 2011; Olivares et al., 2017] and ATP hydrolysis can occur in more than one position in the functioning ClpX hexamer [Martin et al., 2005]. In addition, ClpX has been demonstrated to translocate poly-proline stretches, which are unable to adopt the extended conformation required for translocation according to the SC/2R model [Barkow et al., 2009; Fei et al., 2020]. The SC/2R model and several alternatively proposed models for ClpX-mediated substrate translocation are discussed in a re-

cent publication by Fei et al. [2020]. However, if one of the proposed models is able to fully integrate the observed AAA+ motor characteristics and whether a common mechanism exists at all, is an ongoing debate.

1.4.5. Nucleotide sensing and inter-subunit communication in classic clade AAA+ proteins

Several recent publications unraveled additional elements involved in the communication of the nucleotide state to adjacent subunits for the classic clade AAA+ ATPases YME1, Msp1 and spastin (reviewed in [Puchades et al., 2020]).

The mitochondrial inner membrane protease YME1 is involved in mitochondrial protein quality control, important for regulation of e.g. the respiratory electron transport chain, protein import and lipid biosynthesis. YME1 employs an inter-subunit signaling (ISS) Asp-Gly-Phe motif, which is located at the C-terminal end of $\alpha 5$ at the interface between two α/β-subdomains of adjacent subunits. This motif is conserved across homologs and paralogs of YME1, as well as in the ATPases of the 26S proteasome [Puchades et al., 2017]. Dependent on the nucleotide state, the ISS bridges the nucleotide binding pocket as a loop and engages in inter-subunit hydrophobic packing and π-stacking interactions with the anticlockwise neighboring subunit. According to Puchades et al. [2017], loss of the γ-phosphate upon ATP hydrolysis abolishes the inter-subunit interactions as the ISS folds into a helix.

Another example constitutes the "nucleotide communication loop" (NCL) from the meiotic clade AAA+ ATPase Msp1, which extracts mis-localized tail-anchored proteins from membranes and targets them for degradation [Wang et al., 2020]. The NCL is unique to Msp1 and immediately follows a Trp-Asp residue pair (WD motif), which engages in cation-π and ionic interactions with the arginine fingers, thereby sensing the nucleotide state of the clockwise adjacent subunit. Upon ATP hydrolysis and γ-phosphate release, both the arginine fingers as well as the WD motif become less ordered and result in the "melting" of the NCL, weakening the inter-subunit interface [Wang et al., 2020].

Yet another example of inter-subunit communication of AAA+ proteins was recently described for the meiotic clade AAA+ ATPase and microtubule severing enzyme spastin. Sandate et al. [2019] elucidated a network of salt bridges connecting the substrate-engaging PL1 of one spastin subunit to the nucleotide binding pocket of the neighboring subunit. Since this allosteric mechanism does not involve the reorganization of secondary structure elements, it was speculated to allow the rapid

propagation of conformational changes, potentially allowing near-simultaneous ATP hydrolysis events in all subunits [Puchades et al., 2020].

1.4.6. Rubisco activase

The discovery of Rubisco activase (Rca) dates back to a screen designed to identify photorespiratory pathway mutants in *Arabidopsis thaliana* and the characterization of a mutant with a high CO_2 requiring phenotype and a low *in vivo* Rubisco activity. As shown by ^{14}C batch labeling, this mutant accumulated RuBP, as expected for a mutant impaired in Rubisco activity [Somerville et al., 1982]. However, the purified Rubisco displayed normal activity (after EDTA inactivation and subsequent activation in high CO_2 / high Mg^{2+} conditions) and was later shown to be impaired in its activation status, thus named *rca* for "ribulosebisphosphate carboxylase activation". Somerville et al. [1982] was the first to raise the two competing hypotheses that either the absence of an activator or the presence of an inhibitor could be responsible for the observed phenotype. While both processes play a role in the light dependent regulation of Rubisco activity, Salvucci et al. [1985] later demonstrated that the high CO_2 requiring phenotype of the *rca* mutant was linked to the absence of a chloroplastic protein (of two subunits in *A. thaliana*), indicating that the process of Rubisco activation, which was previously thought to happen spontaneously [Lorimer et al., 1976], is enzymatically catalyzed *in vivo*. Only a serendipitous contamination of the RuBP preparation with ATP facilitated the development of an *in vitro* assay for Rca activity [Portis and Salvucci, 2005] and pointed Streusand and Portis [1987] to the ATP dependence of Rubisco reactivation. Initial experiments used uncarbamylated and RuBP-inhibited Rubisco as substrate for Rca [Portis et al., 1986], causing ambiguity between Rubisco activation by carbamylation and activation by inhibitor removal in response to Rca action [Portis and Salvucci, 2005]. Only when Robinson and Portis [1988] demonstrated, that Rca facilitated the release of the CA1P inhibitor from activated (carbamylated) active sites, inhibitor removal was unambiguously established as key to Rca function (Figure **1.21**).

Interestingly, in addition to green-type Rca of plants, convergent evolution in different kingdoms has led to at least three other types of Rubisco activating enzymes, which evolved from distinct AAA+ proteins. These three types of Rubisco activating enzymes are introduced below and a comparison of their AAA+ modules is given in Figure **4.2**.

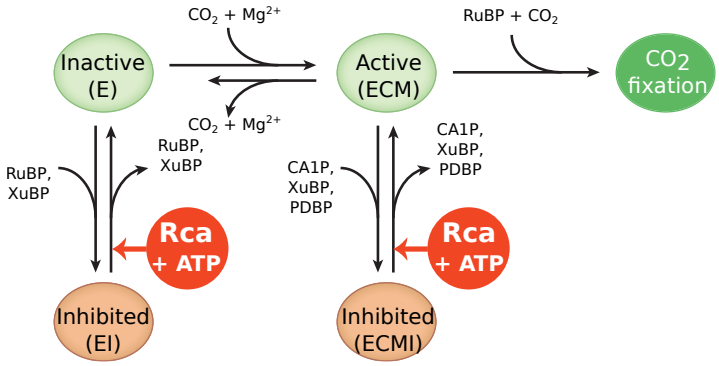

Figure 1.21.: Rubisco maintenance by Rca.
Inhibitory sugar phosphates are removed form Rubisco active sites by Rca. The free or inhibitor bound uncarbamylated (E or EI, respectively) and carbamylated (ECM or ECMI, respectively) Rubiscos are depicted. Abbreviation for inhibitory sugar phosphates are stated in Figure **1.18**. Figure adapted from Bracher et al. [2017] with permission.

1.4.6.1. Green-type Rca

Green-type Rca (in the following abbreviated as Rca, unless specified otherwise) from plants is nuclear encoded and upon expression targeted and imported into chloroplasts due to a N-terminal chloroplast transit peptide. The AAA domains in Rca are preceded by an N-terminal $\sim$ 70-residue domain of unknown structure, and followed by a C-terminal tail (Figure **1.22** (**top**)).

The first structure of Rca from plants was Rca from *Nicotiana tabacum* (Figure **1.22** (**bottom**)) solved by Stotz et al. [2011], which confirmed its classification as a AAA+ protein [Neuwald et al., 1999]. Later, the structure of Rca from *Arabidopsis thaliana* [Hasse et al., 2015] was determined. In addition, the structure of the small domain from *Larrea tridentata* was solved [Henderson et al., 2011]. A cluster analysis of AAA+ sequences, identified based on phylogenetic relations of their C-terminal α-helical subdomains, revealed homology to the ATPase inactive D2 domain of NSF [Ammelburg et al., 2006], whereas structural similarities to the ATPase active D1 domain of p97/Valosin-containing protein (VCP)/Cdc48 were found by DALI search [Hasse et al., 2015]. In the available crystal structures of the chaperone from *Nicotiana tabacum* and *Arabidopsis thaliana*, the AAA domains of Rca assumed a spiral staircase arrangement; N- and C-terminal domains were disordered [Hasse et al., 2015; Stotz et al., 2011]. Unlike other AAA+ proteins, Rca lacks a canonical pore loop motif and does not contain an aromatic residue in PL1. However, two other loops lining the pore, PL2 and PL3, which were disordered in the structures, contain aromatic residues, and mutational analysis demonstrated their importance for activase function, indicating an active role in Rubisco reactivation [Stotz et al., 2011]. Rca may therefore function in analogy to other hexameric AAA+ proteins by

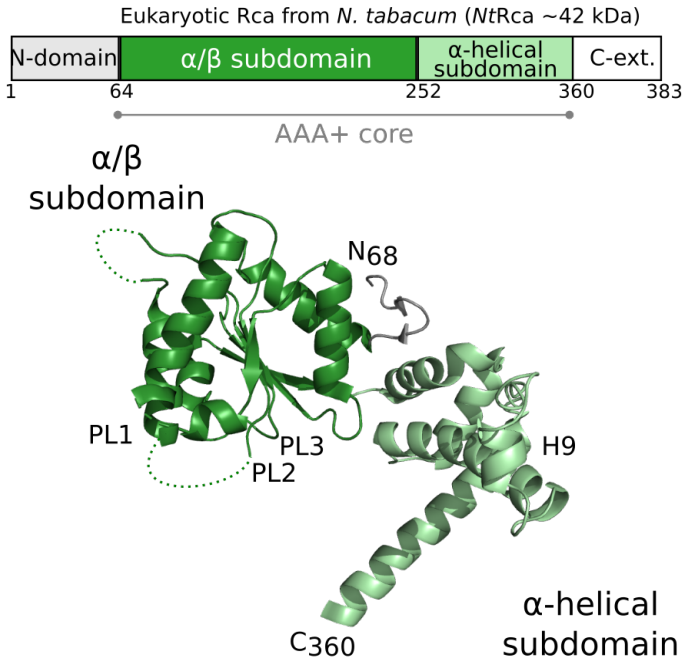

Figure 1.22.: Domain composition and structure of *Nt*Rca.
Top: Schematic representation of the domain composition of *Nt*Rca. **Bottom:** Ribbon representation of *Nt*Rca (PDB: 3T15). α/β and α-helical subdomains are indicated in dark and light green, respectively. Pore loops (PL1, PL2 and PL3), helix H9, and N- and C-termini are indicated.

partial threading of an extended peptide or loop segment through its pore, albeit mechanistic details of the reactivation mechanism of Rca are elusive.

Most analyzed plant species express a long ($\sim$ 45 kDa) and a shorter ($\sim$ 41 kDa) isoform of Rca, Rca-α and Rca-β respectively, which differ in their C-termini and may be the product of alternative splicing (e.g. in *A. thaliana*, spinach, barley or rice) or of separate genes (barley) [Rundle and Zielinski, 1991; Salvucci et al., 1985; To et al., 1999; Werneke et al., 1989]. Other plants express only one isoform (e.g. tobacco). Rca-β lacks a so-called C-terminal extension, a 40 residue stretch, which comprises redox-sensitive cysteines as demonstrated by *in vitro* experiments in presence of reduced thioredoxin-f with Rca from *A. thaliana* and spinach [Zhang and Portis, 1999]. Rca activity is regulated by the redox state and subject to product inhibition by ADP. Under oxidative conditions, Rca-α was inactive in Rubisco reactivation, whereas Rca activity increased severalfold under reducing conditions. In subunit mixing experiments, Zhang et al. [2001] showed that inhibition of Rca activity by oxidation was most effective at a 1:1 molar ratio of Rca-β:Rca-α; higher ratios were less effective in reactivating inhibited Rubisco. Consistently, in *A. thaliana* leaves, Rca isoforms accumulated at a 1:1 ratio, both on transcript and protein level, but other ratios have been reported in other species as well [Eckardt et al., 1997; Salvucci

et al., 1987]. On basis of their findings, Zhang et al. [2001] proposed a model where reduction of Rca-α allows for a rapid monomer-dimer exchange upon mixing, whereby the conformation of the C-terminal extension modulates the activity of both short and long isoforms in response to the redox status. Consistently, upon individual expression of either of the Rca isoforms in *A. thaliana*, the Rubisco activation level was regulated by light intensity only in plants expressing Rca-α, whereas in plants expressing exclusively Rca-β the Rubisco activation level was largely unresponsive to light conditions [Zhang et al., 2002]. Moreover, Rca-α was found to be more susceptible to ADP inhibition under oxidizing conditions than under reducing conditions or than the short isoform and displayed reduced affinity for ATP and ATPγS [Shen et al., 1991; Wang and Portis, 2006; Zhang et al., 2001]. The increased susceptibility towards ADP inhibition was dependent on the Cys residues and negatively charged residues located in the C-terminal extension [Wang and Portis, 2006]. By site specific UV-crosslinking, Wang and Portis [2006] obtained evidence that the C-terminal extension is in proximity to the active site and likely competes with the binding and hydrolysis of ATP. Hence, they proposed a model whereby the C-terminal extension of the long isoform Rca-α has an inhibitory effect on ATP hydrolysis which depends on the redox state of the environment. Consistently, mutation of a conserved Tyr within the C-terminal extension (Tyr361 in *Nt*Rca), renders Rca ATPase and reactivation inactive [Stotz et al., 2011].

In addition to the C-terminal extension, mature Rca features a N-terminal domain (of ~ 60 residues). Deletion of this N-domain renders Rca reactivation inactive while its ATPase activity is unaltered. This effect could be attributed to a conserved Trp residue (W16 and W11 in Rca from tobacco and spinach, respectively), because a single residue mutation to Ala caused the same effect [Esau et al., 1996; Stotz et al., 2011]. The N-domain of *Nt*Rca has been shown to be susceptible to proteolytic cleavage, which indicates that the N-domain is flexibly attached to the AAA+ core and variable in conformation [Stotz et al., 2011]. Notably, substrate interaction of the flexible regions N-terminal to the AAA+ domain, is a common feature of AAA+ proteins like spastin. Moreover, in crosslinking experiments interdomain contacts between the N-domain and the C-terminal extension were identified, demonstrating their proximity in the tertiary and quarternary arrangement [Salvucci, 2004]. Since the presence of the N-domain was required for the crosslinking of residues in the C-terminal extension, Peterson-Forbrook et al. [2017] proposed that the N-domain induces a disorder-to-order transition in the C-terminal peptide, bringing them into crosslinking distance. However, the functional relevance of these findings is not yet clear.

Rca assumes a series of oligomeric states, clearly exceeding hexamers. The oligomeric state of Rca has been shown to depend on Rca concentration, the presence of divalent cations and nucleotides as well as on temperature and the presence of crowding agents in *in vitro* experiments [Barta et al., 2010; Keown et al., 2013; Keown and Pearce, 2014; Peterson-Forbrook et al., 2017; Salvucci, 1992; Wang et al., 1993]. Moreover, the oligomeric state of the individual isoforms may differ from each other and the heterooligomeric complex [Keown and Pearce, 2014]. Only the mutations R294A and R294V, which are thought to modify the subunit-subunit interface, result in stable hexamers in the presence of ATPγS [Blayney et al., 2011; Stotz et al., 2011]. Despite of the highly polydisperse nature of Rca, structural data of an ATPase and reactivation active, stable hexamer forming mutant in combination with mutant subunit mixing experiments strongly suggested that the functional state of Rca is hexameric [Stotz et al., 2011]. Consistently, combined measurements of steady state ATP turnover and dissociation constants by fluorescence-correlation spectroscopy indicated that assemblies larger than hexamers are hydrolytically inactive [Serban et al., 2018]. They proposed a model, according to which hexameric rings form transiently by association of dimers and tetramers and dissociate into these lower oligomeric species upon hydrolysis, thereby coupling the oligomeric state to the ATP hydrolysis cycle of Rca from tobacco.

Comparison of Rcas from different plant species shows considerable variation not only with respect to their oligomerization properties, but also with regard to their ATP turnover kinetics as quantified in standardized *in vitro* assays. Typical ATP hydrolysis rates may vary between 15 - 60 min^{-1} depending on species and ADP:ATP ratio [Hazra et al., 2015] and unlike many other AAA+ proteins, Rca is not stimulated in presence of its substrate, Rubisco. Rca and Rubisco proteins from different plants are remarkably cross-compatible, although substrate specificity is clearly affected by coevolution, as demonstrated by the inability of Rca from solanaceous species to activate Rubisco from non-solanaceous species and vice versa [Wang et al., 1992]. On the Rubisco side, mutational analysis attributed this specificity to two surface residues located in the equatorial region of Rubisco [Larson et al., 1997; Ott et al., 2000]. While solanaceous plants encode an arginine (Arg89) and lysine (Lys94) at these positions, non-solanaceous species encode a proline (Pro89) and aspartate (Asp94) or glutamate (Glu94), respectively [Portis et al., 2008]. On the Rca side, this specificity was subsequently attributed to residues located in the sensor-2 region within the α-helical subdomain [Li et al., 2005a]. Mutational analysis pointed to a direct interaction between charged residues on Rca and Rubisco, and indicated that a conserved charge switch is responsible for the ability to distinguish solanaceous

from non-solanaceous Rubisco [Portis et al., 2008]. Additional Rca residues have been identified and implicated in Rca-Rubisco interaction, because their mutation lead to strongly compromised reactivation ability with negligible effects on ATP hydrolysis. For example, efforts to understand and transfer the increased thermotolerance of *Agave tequilana* Rca to Rca from rice, lead to the implication of residues located on the α4-β4 loop (Gln215 and Lys216 - numbering according to Rca from rice) of Rca in Rubisco interactions [Shivhare and Mueller-Cajar, 2017]. Moreover, Shivhare et al. [2019] used a structure guided approach to map additional Rca-Rubisco interacting residues and identified a phosphomimetic substitution (S23D) in the N-terminal domain as well as several residues (Lys148 and Asp159 - rice numbering) in the highly conserved helix α3 that uncoupled ATPase from Rubisco reactivation.

1.4.6.2. CbbX-type Rca

The first red-type Rca that was characterized was CbbX from *Rhodobacter sphaeroides* (*Rs*CbbX) [Mueller-Cajar et al., 2011]. The *ccbX* gene is associated with the Rubisco operon in this species [Gibson and Tabita, 1997]. In prokaryotes, this type of Rca was found to be associated with Rubisco form IC. The AAA domains in CbbX are N-terminally preceded by a compact α-helical domain of $\sim$ 35 residues (Figure **1.23** (**top**)).

Structurally, CbbX resembles highly processive protease-associated motors like HslU and ClpX and the helicase RuvB of the HCLR (HslU/ClpX, the C-terminal AAA+ module of ClpABC, Lon and RuvB) clade of AAA+ proteins [Hasse et al., 2015], which are characterized by a single pre-sensor 1 β-hairpin insertion (ps1βh) between α3 and β4 in the minimal core of the AAA+ domain (Figure **1.19**) [Miller and Enemark, 2016]. Surprisingly, the eponymous element of the ps1βh clade is absent in CbbX (Figure **1.23** (**bottom**)).

Activity of CbbX from *R. sphaeroides* requires the presence of the Rubisco substrate RuBP. *In vitro*, RuBP-free CbbX forms ATPase inactive, fibrillar, high molecular weight assemblies, which have been suggested to represent a storage form. In presence of ATP, binding of RuBP induces the formation of ring-shaped hexamers, that hydrolyze ATP at a rate of $5\,\mathrm{min}^{-1}$ and are capable of reactivating Rubisco inhibited with RuBP, CABP, or XuBP [Bhat et al., 2017; Loganathan et al., 2016; Mueller-Cajar et al., 2011]. Notably, both inhibited and active Rubisco stimulate ATP hydrolysis by CbbX several fold. RuBP binding can be competed by sulfate

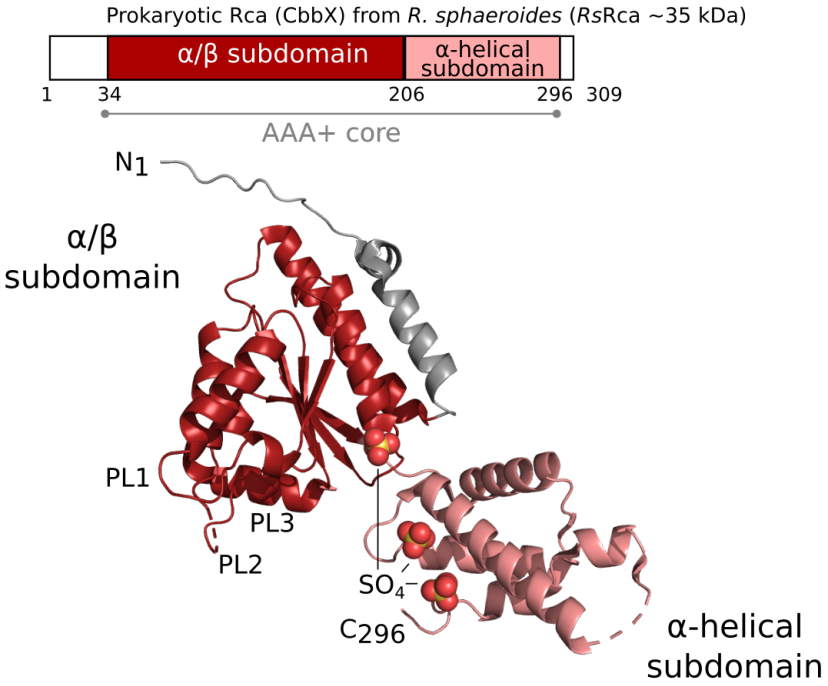

Figure 1.23.: Domain composition and structure of *Rs*CbbX.
Top: Schematic representation of the domain composition of *Rs*CbbX. **Bottom:** Ribbon representation of *Rs*CbbX (PDB: 3SYL). α/β and α-helical subdomains are indicated in dark and light red, respectively. The domain N-terminal to the AAA domains is indicated in grey. Pore loops (PL1, PL2 and PL3), and N- and C-termini are indicated.

ions and mutations of conserved, mostly positively charged residues within the proposed RuBP binding pocket abolish both RuBP-induced ATPase active hexamer formation and Rubisco reactivation [Mueller-Cajar et al., 2011].

CbbX has a canonical AAA+ protein family pore loop (PL1) sequence, Gln-Tyr-Ile-Gly. Mutating the Tyr in PL1 stimulated basal ATPase activity twofold in presence of RuBP, but rendered CbbX unreceptive to Rubisco-induced acceleration of ATP hydrolysis and abolished Rubisco reactivation [Mueller-Cajar et al., 2011]. For reactivation CbbX engages one Rubisco active site at a time [Bhat et al., 2017; Mueller-Cajar et al., 2011]. As demonstrated by mutational analysis, reactivation of Rubisco by CbbX requires an extended C-terminal peptide tail of RbcL [Mueller-Cajar et al., 2011], which is only present in Rubisco of form IC and ID. Importantly, point mutations of the conserved RbcL C-terminal Asp residues (Asp474 and Asp 477), which are thought to provide a major anchor point for the RbcL C-terminus during active site closure, have also been shown to abolish the engagement by CbbX. Hence, the fixed RbcL C-terminal peptide running across the closed active site pocket might serve as a cue for CbbX to engage Rubisco [Bhat et al., 2017].

Interestingly, reactivation of red-type Rubisco is conserved in organisms expressing a Form ID Rubisco, like red algae, diatoms, brown algae, and cryptophytes. For example, the red algae *Cyanidioschyzon merolae* expresses two copies of a protein homologous to the prokaryotic CbbX; one of them being plastid-encoded(CbbX-P), the other being nuclear- or nucleomorph-encoded (CbbX-N) [Fujita et al., 2008; Loganathan et al., 2016]. While the individual copies of both algal CbbX homologs have been shown to be ATPase and reactivation inactive, the copurified and *in vitro* assembled heterohexamer CbbX-NP displayed a low basal ATPase rate at physiological temperatures and reactivated both prokaryotic Form IC and eukaryotic Form ID Rubisco of the red lineage from inhibition by RuBP or CABP binding [Loganathan et al., 2016]. Rubisco reactivation by CbbX-NP requires a functional PL1 and experiments with C-terminally truncated *Rs*Rubisco demonstrated that the CbbX-NP functions by engagement of the RbcL C-terminal peptide, like prokaryotic CbbX. However, perhaps caused by the nature of the heterohexamer, its regulation is different from prokaryotic CbbX, as it does not form ATPase inactive high molecular weight complexes. In contrast, the CbbX-NP forms stable hexameric complexes across a wide concentration range and its ATPase rate is stimulated both by RuBP and Rubisco, although in comparison to the prokaryotic CbbX, stimulation by Rubisco is only modest.

1.4.6.3. CbbQ-type Rca

The most recently identified type of Rca is the CbbQ/O heterooligomer from *Acidithiobacillus ferrooxidans* (*Af*CbbQ/O) (Figure **1.24**). CbbQ/O is frequently associated with prokaryotic form IA and form II Rubisco in chemoautotrophic proteobacteria [Tsai et al., 2015]. Genomes of chemoautotrophic proteobacteria often encode multiple different forms of Rubisco, including Form IAc, Form IAq, one or more Form IC, and Form II [Badger and Bek, 2008]. Experiments in *Hydrogenovibrio marinus* and *Halothiobacillus neapolitanus* demonstrated differential expression of these Rubisco forms in response to the environmental CO_2 concentration [Baker et al., 1998; Yoshizawa et al., 2004], which was proposed to be a general phenomenon explaining the coexistence of several Rubisco forms in chemoautotrophic proteobacteria.

CbbQ belongs to the MoxR protein family, that is part of the helix-2-insert (h2i) clade of AAA+ proteins. The h2i clade is characterized by containing an additional β-hairpin insertion in $\alpha 2$ that is referred to as the helix-2-insert in addition to the single β-hairpin insertion between $\alpha 3$ and $\beta 4$ that distinguishes members of the pre-sensor 1 β-hairpin superclade (ps1βh) from those with the minimal core

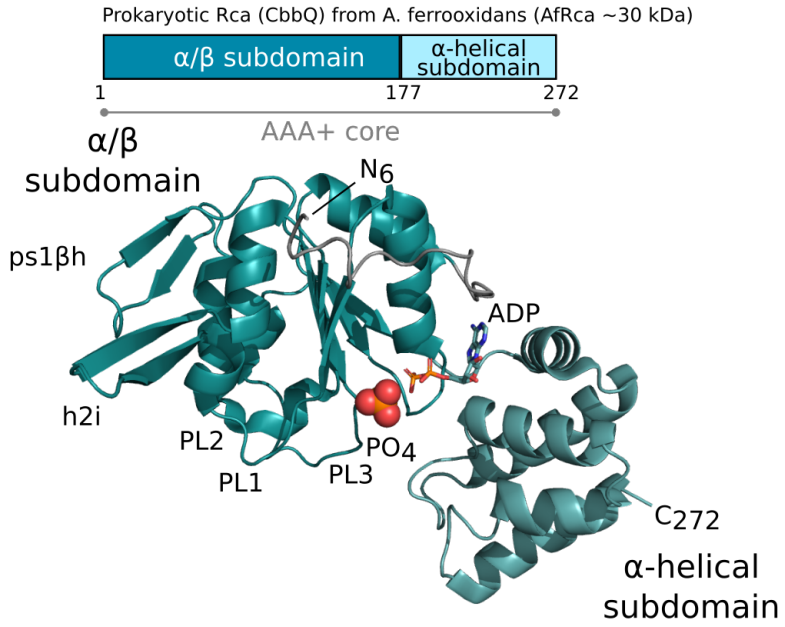

Figure 1.24.: Domain composition and structure of *Af*CbbQ.
Top: Schematic representation of the domain composition of *Af*CbbQ. **Bottom:** Ribbon representation of *Af*CbbQ (PDB: 6L1Q). α/β and α-helical subdomains are indicated in dark and light teal, respectively. The bound ADP and the PO_4 ion are represented as sticks and spheres, respectively. Pore loops (PL1, PL2 and PL3), ps1βh, h2i and N- and C-termini are indicated.

of the AAA+ domain (Figure **1.24**) [Iyer et al., 2004]. Miller and Enemark [2016] further classify MoxR family members as proteins belonging to the pre-sensor 2 insert (ps2i) clade, which differ in an additional α-helical insertion after α5 from other h2i members. However, CbbQ lacks a ps2i and its sensor 2 residue is *cis*-acting like in members of the HCLR clade and most other AAA+ proteins. In typical MoxR family proteins, the α-helical subdomain is positioned on the opposite site of the α/β-subdomain and as a consequence their sensor 2 residue contributes to the active site pocket of the adjacent subunit [Miller and Enemark, 2016].

MoxR family proteins frequently function together with a Von Willebrand factor Type A (VWA) domain containing protein encoded within the same operon [Snider and Houry, 2006]. In *Acidithiobacillus ferrooxidans*, the gene encoding the VWA domain containing protein CbbO (*cbbo*) is positioned directly adjacent to *cbbq* in the same operon. *Acidithiobacillus ferrooxidans* encodes three CbbQO homologues, Q1O1 associated with Rubisco form IAq, Q3O3 associated with Rubisco form IAc, and Q2O2 associated with Rubisco form II. In phylogenetic analyses of CbbQO homologues, the three types form separate clusters across chemoautotrophic proteobacteria [Tsai et al., 2015].

CbbQ forms ring-shaped ATPase active hexamers but activates Rubisco only in co-operation with CbbO, which binds CbbQ in a $CbbQ_6$:$CbbO_1$ stoichiometry [Tsai et al., 2015]. Interestingly, both Q1O1 and Q2O2 isoforms display bona fide Rubisco activase activity toward their associated Rubisco, but are unable to reactivate the non-associated Rubisco encoded within the same genome. Importantly, Rubisco activase function has not been demonstrated for the carboxysomal associated Q3O3 so far [Sutter et al., 2015].

Experiments with a C-terminal truncated form of CbbO2 demonstrated that CbbO binds the CbbQ hexamer via its N-terminus and subsequent structural experiments elucidated that CbbO2 engages the concave side of the CbbQ hexamer [Tsai et al., 2015, 2020]. Binding of CbbO stimulates ATP hydrolysis of CbbQ $\sim$ threefold, which is further stimulated in presence of the RuBP-or CABP- inhibited Rubisco, but not by the activated Rubisco. The central pore of CbbQ is formed by the h2i domain which shields PL1, which in turn is inaccessible from the convex side of the hexamer [Sutter et al., 2015; Tsai et al., 2020]. Mutation of the h2i located key residue Leu85, that faces the concave side of the hexamer, to a charged residue abolishes reactivation completely and renders the CbbQ ATPase unresponsive to Rubisco [Tsai et al., 2020]. Interestingly, the mutation of directly adjacent residues Asp86 and Asp88, located at the convex side of the CbbQ hexamer, did not impair CbbQ2O2 function. All CbbQ isoforms lack the aromatic residue in the pore loop (PL1). Importantly, mutations of PL1 residues of CbbQ2 maintain CbbQ2O2 reactivation-competent, unless these mutations display a severely reduced ATPase rate. Specifically, reduced reactivation at near wild type ATPase activity have not been observed, indicating that PL1 does not participate in a pore loop threading mechanism [Tsai et al., 2020].

CbbO was shown to bind Rubisco via a metal-ion-dependent adhesion site (MIDAS). While several mutations in the MIDAS motif still allowed CbbO2 to form ATPase active Q_6O_1 complexes, they abolished reactivation of inhibited form II Rubisco [Tsai et al., 2015]. Concomitantly, these mutations also abolished further stimulation of the ATPase rate by the inhibited substrate, probably, because substrate sensing was impaired. Substrate sensing and Rubisco reactivation were equally impaired upon mutation of the corresponding conserved acidic residues Asp82 and Glu75 on the surface of both form I and form II Rubisco, respectively [Tsai et al., 2015]. Moreover, Rubisco reactivation by CbbQO stringently requires the last two RbcL C-terminal residues [Tsai et al., 2020]. Both form IAq and Form II Rubisco display a conserved His-Lys/Arg motif at their most RbcL C-terminal residues. Mutation of the conserved His is sufficient to completely abolish Rubisco reactivation. Interestingly, Form IAc Rubisco lacks this conserved His, which may point to a distinct role for

CbbQ3O3, that is maybe different from the roles of the CbbQ1O1 and CbbQ2O2 homologs.

A combination of mutational analysis, biochemical and structural experiments allowed Tsai et al. [2020] to draw a mechanistic model of Rubisco engagement by the CbbO adapter and subsequent signal communication to the CbbQ motor, which is required for Rubisco stimulated enhancement of its ATPase activity. According to their model, engagement of Rubisco by CbbO via the MIDAS motif and the conserved acidic surface residue of Rubisco induces a conformational change in the last two C-terminal residues of CbbO, which are part of the MIDAS domain. The conformational change can be mimicked by C-terminal truncations of CbbO, which lead to a severe (over)stimulation of the ATPase activity irrespective of the presence of Rubisco. Binding of Rubisco is communicated to the CbbQ motor via a hydrophobic residue (Leu161) located in the β4-α4 loop of CbbQ. Further communication to the active site of CbbQ may involve a network of salt bridges formed between a spacially adjacent positive residue (Lys166) located in the β4-α4 loop and the negatively charged sidechains of Glu72 in PL1 and Glu114, which is located in proximity to the Walker B glutamate (Glu111). Both mutations of Lys166 and Glu72 stimulate ATPase activity. Leu161-communicated Rubisco sensing depends on the key residue Leu85 located in the h2i. This may position elements like Walker B glutamate and the Arg finger (R172) into conformations more conducive to ATP hydrolysis via this network of salt bridges [Tsai et al., 2020]. Thus, Leu85 was hypothesized to fulfill a key function in communication between the CbbQ motor and the CbbO adapter - both with regard to communication of Rubisco binding and force transmission from the CbbQ hexamer to CbbO that is required for Rubisco remodeling.

A more detailed view on force transmission from the CbbQ motor to the CbbO adapter as well as mechanistic insights into the actual remodeling process at the interface of CbbO and Rubisco are still elusive and await further investigation.

1.5. Phase separation and its role in carboxysome assembly

Many biochemical reactions are spatially constrained in specific compartments, e.g. for the simultaneous execution of reactions that would interfere with each other. Classic compartments in eukaryotic cells are enclosed by a lipid bilayer, which restricts diffusion and is permeable for a set of specific molecules only. For example,

the lipid bilayer of lysosomes restricts the hydrolytic degradation machinery and its acidic environment to this organelle. Other compartments, often referred to as membrane-less organelles, are devoid of a selective membrane and thus for a long time it was elusive how these concentrate molecules, maintain and regulate their structure, control their composition and modulate internal biochemical activities [Banani et al., 2017]. The latter group includes the nucleolus, involved in ribosome biogenesis [Boisvert et al., 2007], Cajal bodies, involved in RNA modifications [Staněk, 2016], stress granules, that form upon inhibition of translation initiation [Buchan and Parker, 2009], or p or germ granules, involved in establishment of the germ line of sexually reproducing cells [Seydoux and Braun, 2006]. One of the earliest descriptions of the physico-chemical properties of a membrane-less organelle came from P granules of *Caenorhabditis elegans*. Brangwynne et al. [2009] demonstrated that P granules overall behave like liquid droplets, as P granules fuse, drip, wet and are easily deformed by flows and have a viscosity similar to runny honey [Hyman et al., 2014]. Subsequently, liquid-like properties were demonstrated for many of these membrane-less organelles and thus the term, "biomolecular condensate" was coined as a common denominator of all these compartments, which are thought to form by a shared mechanism - liquid-liquid phase separation (LLPS) [Banani et al., 2017]. Notably, LLPS is not restricted to eukaryotes. Carboxysomes, membrane-less, biological microcompartments (BMC) with specialized function in carbon sequestration of photo- and chemoautotrophic prokaryotes have been demonstrated to rely on proteins that form biomolecular condensates in an LLPS induced process.

1.5.1. Phase separation in biological systems

Liquid-liquid phase separation (LLPS) refers to a material state, in which two or more liquids coexist but do not mix. Although individual molecules are able to cross the boundary of the two liquids, the net flux between the compartments equals zero because the chemical potential of the individual liquids are equal [Hyman et al., 2014]. While the gain in entropy, a measure for disorder in a given system, drives the mixing of different molecules by diffusion, demixing in separate phases is driven by the cumulative strength of interactions between those molecules. If solute:solute or solvent:solvent interactions are favored over solute:solvent interactions and the gain of free energy by associating solute molecules and solvent molecules overcomes the contribution of entropy to form a homogeneous mixture, the mixture of molecules begins to demix and forms separate phases [Feng et al., 2019]. In biological systems the solvent is typically aqueous, thus the property to demix is solute dependent.

As solute:solute interactions are required for macromolecular condensation, most proteins found in liquid-like condensates are multivalent. For proteinaceous solutes, multivalency can be achieved by repeats of folded domains, which are prone to interact with themselves or other components of the condensate or by repetitive, low complexity motifs in intrinsically disordered proteins (IDP), that do not adopt a defined three-dimensional fold [Banani et al., 2017].

An example for the former was presented by Li et al. [2012], who used an engineered two component system based on repeat domains often found in signaling proteins. Modulating multivalency by increasing the number of fused interaction motifs lead to a higher propensity of these artificial proteins to demix from solution. The formed droplets were strongly enriched in protein in comparison to the bulk solution and exhibited liquid-like properties. Droplet formation could be impeded by addition of high affinity monovalent ligands. In fluorescence recovery after photobleaching (FRAP) experiments, the authors noticed that the recovery rate was inversely correlated with monomer:monomer affinity and valency, pointing to a sol-gel transition inside the droplets. Subsequently Li et al. [2012] recapitulated these findings for the endogenous nephrin:NCK:N-WASP system of the actin-regulatory signaling pathway. Notably, in the carboxysomal scaffolding protein CcmM multivalency is also achieved by repeats of folded domains [Wang et al., 2019].

IDPs participating in LLPS are often characterized by low sequence complexity and are enriched in a small number of amino acids, such as glycine, serine, glutamine, asparagine, phenylalanine and tyrosine [Banani et al., 2017]. For example, cation-π interactions between numerous Phe-Gly motifs and Arg sidechains as well as π-stacking interactions are thought to generate the multivalent weakly adhesive intermolecular interactions required for LLPS in the N- and C-termini of the DEAD box helicase Ddx4 [Nott et al., 2015], a major component of germ granules which are thought to host components of an RNAi pathway, which guard spermatocytes and spermatids against transposable elements [Kotaja and Sassone-Corsi, 2007]. Moreover, many IDPs which participate in LLPS contain clusters of amino acids with net negative or positive charge or contain multiple Gln, Asn or Ser residues which contribute to weak intermolecular interactions via electrostatic potential and dipolar interactions, respectively. Interestingly, IDR proteins are often found in biomolecular condensates, which also enrich RNA, such as P bodies, stress granules, or germ granules.

In comparison to membrane-enclosed organelles, biomolecular condensates represent a means for a more flexible and faster regulation of associated reactions by

phase transition. Phase transitions may be invoked in response to environmental cues like temperature, pH, salt concentration or by posttranslational modification or, on larger time scales, changes in gene expression of scaffolding proteins. At the same time, biomolecular condensates still allow unhindered access by diffusion for small molecules, including substrates for biochemical reactions. Phase transitions in biological systems are a field of active research, because corresponding theories from colloidal chemistry are able to describe the organization of membranes and the cytoplasm from a more general physico-chemical perspective, than previous attempts, which were primarily focused on the association of rigid macromolecules of defined molecular architecture [Hyman and Simons, 2012]. Importantly, because a large number of protein aggregates coinciding with neurodegenerative disease have been proposed to be a product of misregulated phase transition [Hyman and Simons, 2012], the maturation, that is the alteration of the physico-chemical properties of biomolecular condensates with time (e.g. sol-gel transitions), has received great attention in this context (reviewed e.g. by Aguzzi and Altmeyer [2016]).

1.5.1.1. Biogenesis of β-carboxysomes via LLPS

The formation of β-carboxysomes is governed by LLPS. In β-cyanobacteria, carboxysome assembly begins with the aggregation of Rubisco and the γ-CA domain containing scaffolding protein CcmM into procarboxysomes (PC) [Cameron et al., 2013] (Figure **1.26**). Full-length CcmM is a $\sim$ 58 kDa multidomain protein, that forms trimers and comprises an N-terminal γ-CA domain and three to five small subunit-like (SSUL) domains, with sequence similarity to RbcS. Moreover, an internal ribosomal binding site located C-terminal to the CA domain ensures the translation of a $\sim$ 35 kDa isoform, M35, comprising only the SSUL repeat domains. Rubisco aggregation has been proposed to be mediated by interconnected SSUL domains of CcmM, which individually bind to Rubisco and thereby connect (link) multiple Rubisco molecules [Long et al., 2007, 2010, 2011].

Indeed, the interface between an individual CcmM SSUL domain and Rubisco has recently been identified. A combination of crystallography and cryo-EM allowed Wang et al. [2019] to locate the SSUL:Rubisco interface at the equator of Rubisco in the cleft formed between two antiparallel RbcL dimers. Interactions between SSUL and both RbcL and RbcS were mostly mediated by electrostatic and van der Waals interactions, with the former critically depending on three conserved Arg residues (Arg251, Arg252, Arg254 in *S. elongatus* PCC 7942) (Figure **1.25**). Mutation of two of these Arg residues, but also the absence of RbcS from the octameric RbcL core,

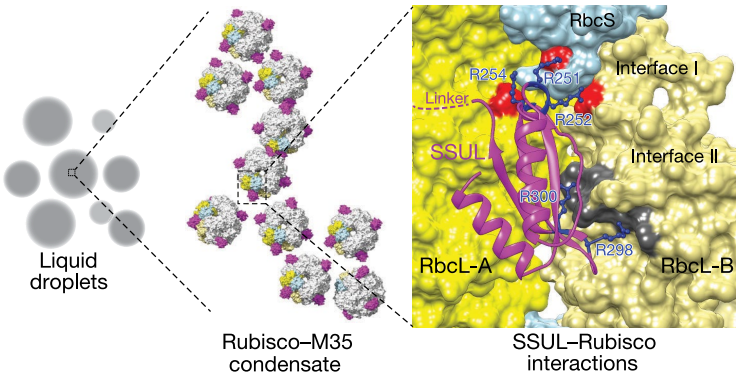

Figure 1.25.: Model of the liquid-like Rubisco-CcmM (M35) condensate.
The C-terminal SSUL domains of CcmM bind close to the equator of Rubisco at the interface between two antiparallel dimers via two interfaces. Their interaction crucially depends on conserved arginine residues (Arg251, Arg252, Arg254). See text for details. Figure modified from Wang et al. [2019].

abolished SSUL binding. Moreover, *in vitro* experiments demonstrated that Rubisco and M35 demix from solvent and colocalize in spherical droplets with liquid-like properties as judged by the observation of fusion events between individual droplets [Wang et al., 2019], consistent with condensate formation by LLPS. Condensate formation was susceptible to high salt concentrations and could be competed by the addition of individual SSUL domains. Redox-dependent disulfide formation within some SSUL domains of CcmM decreased the affinity of CcmM to Rubisco and lead to increased mobility of Rubisco inside the condensate, as estimated by fluorescence recovery after photobleaching (FRAP), in comparison to the reduced state.

Next to its role in Rubisco condensate formation, full length CcmM interacts with CcaA [Cot et al., 2008] and is required to mediate the association of CcmN, which is indispensable for β-carboxysome biogenesis, with the PC [Cameron et al., 2013; Kinney et al., 2012]. In turn, CcmN has been shown to interact with the hexameric shell protein CcmK2 via a C-terminal encapsulation peptide, which adopts an alpha-helical structure and is conserved in other BMCs. In absence of CcmN shell recruitment is abolished [Cameron et al., 2013] and carboxysome formation is stalled. After association of CcmN with the PC, the shell proteins CcmK2 and CcmO are recruited and simultaneously required to form the icosahedron. Finally, the pentameric protein CcmL is thought to close the vertices to prevent leakage of CO_2 [Cameron et al., 2013]. Accordingly, a mutant of the pentameric vertex protein of α-carboxysomes, CsoS4, has been shown to lose its limited permeability to CO_2 in *Halothiobacillus neapolitanus* [Cai et al., 2009] (Figure **1.26**).

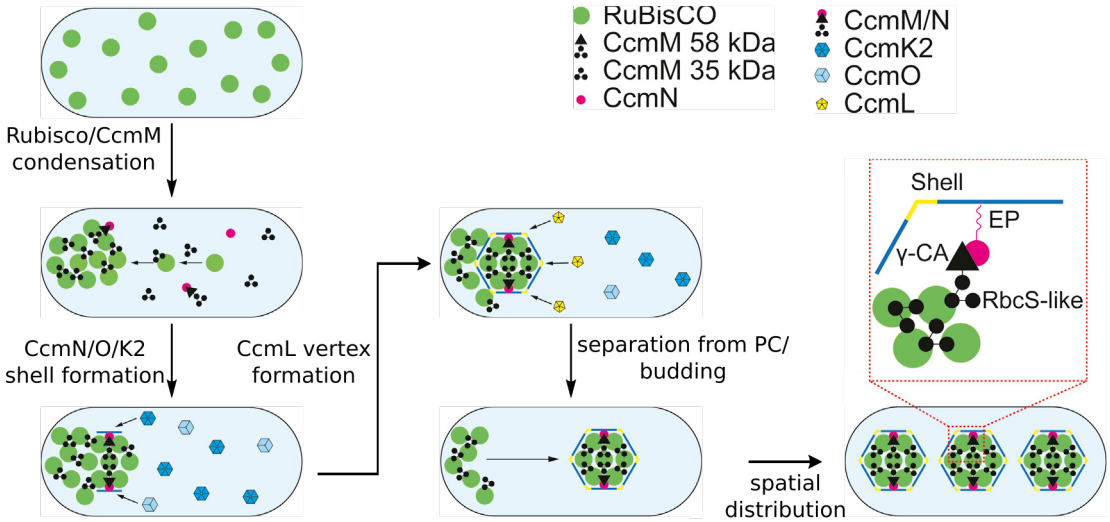

Figure 1.26.: Formation of the β-carboxysome.
Carboxysome formation is initialized by Rubisco/CcmM condensate formation. Subsequently, CcmN mediated interactions with CcmK allow the shell proteins to associate. The carboxysome is closed by the pentameric vertex protein CcmL and separates from the procarboxysomal condensate. See text for details. Figure modified from Cameron et al. [2013].

1.5.1.2. LLPS in other carbon concentrating mechanisms

Interestingly, proteins with a function analogous to the β-carboxysomal scaffolding protein CcmM also exist in α-carboxysomes and the pyrenoid of eukaryotic algae and hornworts.

The α-carboxysomal protein CsoS2 is intrinsically disordered and interacts both with Rubisco and the shell protein CsoS1 [Cai et al., 2015]. While the binding interface with CsoS1 is currently unclear, CsoS2 binds Rubisco via sequence repeats localized in the N-terminal domain of CsoS2. These repeats have been reported to transiently form α-helices containing positively charged side chains, which bind to a negatively charged patch at the dimer-dimer cleft of Rubisco via ionic- and π-cation interactions [Liu et al., 2018; Oltrogge et al., 2020]. Like the SSUL of CcmM, binding of the CsoS2 repeat motif depends on residues on RbcS, likely to ensure encapsulation of the functional hexadecameric Rubisco, but the smaller interface allows simultaneous binding of eight CsoS2 binding motifs instead of four SSUL domains in CcmM. Notably, the N-terminal domain of CsoS2 was required and sufficient to induce demixing of Rubisco via LLPS at low salt concentrations only. According to a model proposed by Oltrogge et al. [2020], the susceptibility to physiological salt concentrations is overcome by a local increase of CsoS2 concentration by prior association of multiple CsoS2 with the carboxysomal shell.

Likewise, the algal protein required for pyrenoid formation, EPYC1, is mostly disordered but comprises five evenly spaced, α-helical, low-affinity Rubisco-binding regions [He et al., 2020]. In the hexadecameric holoenzyme, each RbcS provides an independent binding site for one of the Rubisco-binding regions of EPYC1, which binds primarily via ionic- and hydrophobic interactions. Like CcmM or CsoS2, EPYC1 forms liquid-like condensates by LLPS that are able to fuse and are affected by salt, temperature and the concentration of EPYC1 [Wunder et al., 2018].

1.6. Aim of the study

Throughout the last decade marked progress has been made towards elucidating structure, function and regulation of CbbX-, CbbQ and Rca-type Rubisco activases. Although the structure of several plant Rcas has now been solved and residues critical to function have been identified, the detailed mechanism of form IB Rubisco reactivation is still enigmatic.

A homolog of plant Rca occurs in β-cyanobacteria whose ancestors are thought to be the progenitor of plant chloroplasts and contain a form IB Rubisco. This "Rca-like" protein was previously suggested to have Rca function, based on the finding that the Rca-like protein from *Anabaena* sp. strain CA is required for elevated Rubisco activity and growth under high light conditions [Li et al., 1999]. However, the corresponding Rubisco did not exhibit inhibition by RuBP nor was it susceptible to the so called "fallover", the decrease in activity due to progressive accumulation of inhibiting misfire products [Li and Tabita, 1997]. Hence, direct evidence for Rca function of Rca-like protein is still lacking. Cyanobacterial Rca-like protein lacks the N-domain typically found in plant Rca, but features a C-terminal extension which shares sequence similarity with RbcS and the SSUL domains of the carboxysomal scaffolding protein CcmM. This SSUL domain was proposed to interact with Rubisco [Zarzycki et al., 2013].

Here, we focus on Rca-like protein and Rubisco from the cyanobacterium *Nostoc* sp. PCC 7120 and seek to investigate Rubisco reactivation in cyanobacteria as a model system to study form IB Rubisco reactivation. We analyze the structure and mechanism of cyanobacterial Rca using a combination of crystallography, cryo-EM and *in vitro* biochemical assays. By examining both the AAA+ domain and SSUL domain of cyanobacterial Rca separately, we hope to elucidate the interdependence of functions associated with these domains and hope to deepen our understanding of the role of the SSUL domain in the context of carboxysomal localization of Rca.

2. Material and Methods

2.1. Molecular Biology

2.1.1. Standard DNA preparation

All constructs were generated using standard molecular biology techniques. Commercial kits were used as specified in the supplier manual. In short, Polymerase chain reaction (PCR) was carried out using *Herculase II Fusion DNA Polymerase* (Agilent), PCR products were analyzed and purified by Tris-acetate-EDTA (TAE) -buffered Agarose gel-electrophoresis according to Aaij and Borst [1972] and Armstrong and Schulz [2015] and purified using QIAquick Gel Extraction Kit (Qiagen). *NEBuilder HiFi DNA Assembly Master Mix* (New England Biolabs) was used for Gibson assembly. For the amplification of plasmid DNA, assembled constructs were transformed into chemically competent *E. coli* cells (DH5α) employing the Hanahan method [Hanahan, 1983; Renzette, 2011], positive clones were selected based on growth on corresponding selective medium and subsequently cultured in Luria-Bertani (LB) medium (containing the respective antibiotic) for 12 - 16 h before being harvested using the *QIAprep Spin Miniprep Kit* (Qiagen). DNA concentrations were determined by UV spectroscopy at 260 nm on a *Nanodrop One* (Thermo Fisher Scientific). All constructs were verified by mail-order sequencing at either Eurofins Genomics or Microsynth. For visualization of sequencing traces and comparative sequence analysis the online tool benchling (https://www.benchling.com/) was used.

2.1.2. Culturing and preparation of genomic DNA from *Nostoc* sp. PCC 7120

For the preparation of genomic DNA from *Nostoc* sp. PCC 7120, cells were obtained from the Culture Collection of Algae (UTEX - Cat. 2576) and cultured in BG-11 medium (`http://web.biosci.utexas.edu/utex/Media%20PDF/bg-11-medium.pdf`) at 30 °C and 50 rpm in continuous light. *Nostoc* sp. PCC 7120 cells (1 mL) cultured to high density were pelleted by centrifugation at 10000 x g for 10 min at 4 °C using the Eppendorf Centrifuge 5417R equipped with a FA 45-30-11 rotor (Eppendorf).

The cell pellet was resuspended in 100 µL buffer (50 mM Tris-HCl pH 8.0 / 50 mM NaCl) and cells were lysed by 5 cycles of heating (3 min at 95 °C) and snap freezing in liquid nitrogen. The lysate was centrifuged at 20000 x g for 10 min at 4 °C and 1 µL of the supernatant was used as template in the PCR reactions.

2.1.3. General cloning strategy

In order to generate pET11a-*Ec*GroSEL, the *groES-groEL* operon derived from the pBAD33-EcSEL plasmid [Kerner et al., 2005] was amplified and cloned into the pET11a backbone. pCDF1b-*Nos*RbcX and pCDF1b-*Nos*Raf1 were generated by amplification of the *rbcX* and *raf1* genes from genomic DNA of *Nostoc* sp. PCC 7120 (*Nos*RbcX and *Nos*Raf1), respectively, and subsequent cloning into the pCDF1b backbone. pCDF1b-*Nos*RbcX-*Nos*Raf1 was obtained by amplification of *raf1* from pCDF1b-*Nos*Raf1 and subsequent cloning into pCDF1b-*Nos*RbcX. pET11a-*Nos*GroSEL was obtained by amplifying the *groES-groEL* operon from genomic DNA of *Nostoc* sp. PCC 7120 (*Nos*GroSEL) by PCR and subsequent cloning into the pET11a backbone via Gibson assembly. The monocistronic GroEL was amplified from genomic DNA of *Nostoc* sp. PCC 7120 (*Nos*GroEL2) and subsequently cloned into pET11a-*Nos*GroSEL, thereby generating pET11a-*Nos*GroSEL-GroEL2. The genes in the Rubisco operon *rbcL*, *rbcX* and *rbcS* as well as *raf1* from *Nostoc* sp. PCC 7120 (encoding *Nos*RbcL, *Nos*RbcX, *Nos*RbcS and *Nos*Raf1, respectively) were obtained from genomic DNA by PCR and initially assembled in a pET28b backbone, yielding pET28b-*Nos*LXSR and pET28b-*Nos*LS by Gibson assembly. The His_6-ubiquitin (H_6ubi) tag was amplified from the pHUE plasmid [Baker et al., 2005; Catanzariti et al., 2004] and inserted as a N-terminal fusion to *Nos*RbcL, generating pET28b-H_6ubi*Nos*LS. H_6ubi*Nos*LS was then moved into pET11a-*Nos*GroSEL to obtain pET11a-*Nos*GroSEL-H_6ubi-*Nos*LS. pET11a-*Nos*GroSEL-H_6ubi-*Nos*LΔN12-S and pET11a-*Nos*GroSEL-H_6ubi-*Nos*L(*Cr*N11/ΔN12)-S were generated by deletion of the first twelve N-terminal residues of *Nos*RbcL (MSYAQTKTQTKS) and by replacing the same residues with the first eleven N-terminal residues of *Chlamydomonas reinhardtii* RbcL (MVPQTETKAGA) by PCR and Gibson assembly, respectively. pET11a-*Nos*GroSEL-H_6ubi-*Nos*LΔC2-S, pET11a-*Nos*GroSEL-H_6ubi-*Nos*LΔC4-S and pET11a-*Nos*GroSEL-H_6ubi-*Nos*LΔC9-S were obtained by changing *Nos*RbcL residues 475/476, 473/474 or 468/469 into two consecutive stop codons by site directed mutagenesis. pHUE-*Nos*RbcS was generated by amplifying *Nos*RbcS from pET11a-*Nos*GroSEL-H_6ubi-*Nos*LS and subsequent cloning into the pHUE backbone. pCDF-*Nt*XS1R1R2B2 encoding RbcS1, RbcX, Raf1, Raf2 and BSD2 from

Nicotiana tabacum and pET11a-*At*Cpn60α/β-*At*Cpn20*Nt*RbcL were a kind gift from Lin et al. [2020] and the latter was modified to obtain pET11a-*At*Cpn60α/β-*At*Cpn20-H_6ubi-*Nt*RbcLΔN9 and pET11a-*At*Cpn60α/β-*At*Cpn20-H_6ubi-*At*RbcLΔN9 by deletion of the first nine N-terminal residues (PQTETKASV) of *Nt*RbcL (*Nt*RbcLΔN9) and *At*RbcL (*At*RbcLΔN9) and subsequent introduction of the His_6-ubiquitin tag from pHUE backbone as a N-terminal fusion to *Nt*RbcLΔN9 or *At*RbcLΔN9 (H_6ubi-*Nt*RbcLΔN9 or H_6ubi-*At*RbcLΔN9, respectively). pET11a-*At*Cpn60α/β-*At*Cpn20-H_6ubi-*At*RbcLΔN9-*At*S was generated by PCR amplification of RbcS from *A. thaliana* obtained from the pET28-*At*RbcLS plasmid [Wilson et al., 2019] and subsequent cloning in pET11a-*At*Cpn60α/β-*At*Cpn20-H_6ubi-*At*RbcLΔN9 C-terminal to H_6ubi-*At*RbcLΔN9. pHUE-*Nos*Rca was obtained by amplification of the *rca* gene from genomic DNA (*Nostoc* sp. PCC 7120) and subsequent cloning of residues 2-414 into the pHUE backbone. To generate pHUE-*Nos*RcaΔC, pHUE-*Nos*Rca was truncated by mutating residue 292 and 293 into two consecutive stop codons by site-directed mutagenesis. Both pHUE-*Nos*Rca and pHUE-*Nos*RcaΔC were generated by Leonhard Popilka. All mutations in pHUE-*Nos*Rca (*Nos*Rca R352E/R353E, *Nos*Rca V91E/R92G/G93N/L244D/N248A/L250A, *Nos*Rca K19D, *Nos*Rca D227K, *Nos*Rca R352E/R352E and *Nos*Rca Y116A) and pHUE-*Nos*RcaΔC (*Nos*RcaΔC V91E/R92G/G93N, *Nos*RcaΔC P140G/Y143A/D144A, *Nos*RcaΔC L244D/N248A/L250A) were introduced by QuikChange mutagenesis (Agilent). pHUE-*Nos*SSUL was generated by cloning the C-terminal SSUL domain of *Nos*Rca (residues 325-414) into the pHUE backbone. The *ccmM* gene from *Nostoc* sp. PCC 7120 (*Nos*CcmM) was amplified from genomic DNA and cloned into the pHUE backbone to obtain pHUE-*Nos*CcmM for the expression of full-length *Nos*CcmM. pHUE-*Nos*M35 was generated by cloning the C-terminal domain of *Nos*CcmM (residues 226-555), starting from an internal ribosomal binding side [Long et al., 2007] and comprising three SSUL modules that constitute the short isoform of *Nos*CcmM (*Nos*M35), into the pHUE backbone. A comprehensive list of all plasmids used in this study can be found in Table **2.1**.

Table 2.1.: List of plasmids

Designation	**Source**
pBAD33-*Ec*SEL	[McLennan and Masters, 1998]
pET11a-*Ec*SEL	This study
pET11a-*Nos*GroSEL-GroEL2	This study
pET11a-*Nos*GroSEL	This study

...continues on next page...

Designation	**Source**
pCDF1b-*Nos*RbcX	This study
pCDF1b-*Nos*Raf1	This study
pCDF1b-*Nos*RbcXNosRaf1	This study
pET28b-*Nos*LS	This study
pET28b-*Nos*LXSRaf1	This study
pET28b-H_6ubi*Nos*LS	This study
pET11a-*Nos*GroSEL-H_6ubi-*Nos*LS	This study
pET11a-*Nos*GroSEL-H_6ubi-*Nos*LΔN12-S	This study
pET11a-*Nos*GroSEL-H_6ubi-*Nos*L(*Cr*N11/ΔN12)-S	This study
pET11a-*Nos*GroSEL-H_6ubi-*Nos*LΔC2-S	This study
pET11a-*Nos*GroSEL-H_6ubi-*Nos*LΔC4-S	This study
pET11a-*Nos*GroSEL-H_6ubi-*Nos*LΔC9-S	This study
pHUE-*Nos*RbcS	This study
pCDF-*Nt*XS1R1R2B2	Lin et al. [2020]
pET11a-*At*Cpn60α/β-*At*Cpn20*Nt*RbcL	Lin et al. [2020]
pET11a-*At*Cpn60α/β-*At*Cpn20-*Nt*RbcLΔN9	This study
pET11a-*At*Cpn60α/β-*At*Cpn20-H_6ubi-*Nt*RbcL	This study
pET11a-*At*Cpn60α/β-*At*Cpn20-H_6ubi-*Nt*RbcLΔN9	This study
pET11a-*At*Cpn60α/β-*At*Cpn20-H_6ubi-*At*RbcL	This study
pET11a-*At*Cpn60α/β-*At*Cpn20-H_6ubi-*At*RbcL-*At*S	This study
pET11a-*At*Cpn60α/β-*At*Cpn20-H_6ubi-*At*RbcLΔN9	This study
pET11a-*At*Cpn60α/β-*At*Cpn20-H_6ubi-*At*RbcLΔN9-*At*S	This study
pDuet-*At*R1/R2/RX/B2	Wilson et al. [2019]
pET28-*At*RbcLS	Wilson et al. [2019]
pHUE-*Nos*Rca	Leonhard Popilka[1]
pHUE-*Nos*Rca R352E/R353	This study
pHUE-*Nos*Rca V91E/R92G/G93N/L244D/N248A/L250A	This study
pHUE-*Nos*RcaΔC	Leonhard Popilka[1]
pHUE-*Nos*RcaΔC V91E/R92G/G93N	This study
pHUE-*Nos*RcaΔC P140G/Y143A/D144A	This study
pHUE-*Nos*RcaΔC L244D/N248A/L250A	This study
pHUE-*Nos*SSUL	Leonhard Popilka[1]
pHUE-*Nos*CcmM	This study
pHUE-*Nos*M35	This study

[1] former graduate student

2.1.4. Standard protein analytical methods

2.1.4.1. Protein quantification

Total protein concentration of cell lysates and of partially purified proteins were determined according to Bradford [1976]. To this end, the *Coomassie Plus Protein Assay Reagent* (Pierce) was used in accordance to the supplier's instructions and protein standard curves were obtained from fresh preparations of dilution series of bovine serum albumin (BSA) in H_2O. Protein concentrations of highly pure proteins were determined on a *Nanodrop One* (Thermo Fisher Scientific) by UV spectroscopy at 280 nm using the theoretical extinction coefficient of the respective protein calculated with protparam [Gasteiger et al., 2005]. Biophysical parameters of the proteins used in this study are listed in Table **2.2**.

2.1.4.2. Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and native-PAGE

At every stage of the protein purification procedure, protein expression and protein purity were analyzed by SDS-PAGE and/or native-PAGE. For standard SDS-PAGE protein aliquots were dissolved in denaturing SDS sample buffer (60 mM Tris - HCl pH 6.8 / 2 % (w/v) SDS / 0.05 % (w/v) bromophenol blue / 10 % (v/v) glycerol / 5 % (v/v) β-mercaptoethanol) and heated at 95 °C for 5 min. Approximately 12 µg of total protein were applied to 4 - 12 % Bis-Tris gels (*NuPAGE 4 - 12 % Bis-Tris Gel*, Invitrogen) per lane. Gels were run using a *Mini Gel Tank* (Invitrogen) in 2-(N-morpholino)ethanesulfonic acid (MES) running buffer (50 mM MES, 50 mM Tris base, 1 mM EDTA, 0.1 % (w/v) SDS) at 180 V for 45 - 60 min.

To ensure integrity of protein complexes and for the gel shift assay, proteins were also analyzed by native-PAGE. To this end, protein samples were dissolved in native sample buffer (60 mM Tris - HCl pH 6.8 / 0.05 % (w/v) bromophenol blue / 10 % (v/v) glycerol) and approx. 0.5 - 3 µg (purified protein) or 15 - 20 µg (clarified protein lysate) were applied to *Novex Value 4 - 12 % Tris-Glycine Gels* (Invitrogen). Gels were run using a *Mini Gel Tank* (Invitrogen) in Tris-Glycine running buffer (25 mM Tris base, 50 mM glycine) at 150 V for 120 min at 4 °C. To analyze *Nos*Rca - Rubisco interaction in the gel shift assay, reducing agent was included as described in Chapter **2.10.4**.

Table 2.2.: Biophysical parameters of purified proteins[1]

Description	**Aminoacids**	**MW**[2] **/g mol^{-1}**	**pI**[3]	**Ext. coef.**[4] **/M^{-1} cm^{-1}**
*Nos*RbcL	476	53045.29	6.25	77810
*Nos*RbcLΔN12	464	51689.76	6.05	76320
*Nos*RbcL*Cr*N11/ΔN12	475	52804.04	6.04	76320
*Nos*RbcLΔC2	474	52845.05	6.25	77810
*Nos*RbcLΔC4	472	52598.77	6.38	77810
*Nos*RbcLΔC9	467	51975.10	6.68	77810
*Nos*RbcS	108	12695.58	7.95	22920
re*Nos*Rca	413	46536.84	6.18	28880
ox*Nos*Rca	413	46536.84	6.18	29005
*Nos*Rca Y116A	413	46444.74	6.18	27390
*Nos*Rca Y289A	413	46444.74	6.18	27390
*Nos*Rca R352E/R353E	413	46482.69	5.70	28880
*Nos*Rca V91E/R92G/G93N/L244D/N248A/L250A	413	46441.56	5.81	28880
*Nos*RcaΔC	290	32734.38	5.41	18910
*Nos*RcaΔC V91E/R92G/G93N	290	32722.28	5.20	18910
*Nos*RcaΔC P140G/Y143A/D144A	290	32558.21	5.55	17420
*Nos*RcaΔC L244D/N248A/L250A	290	32651.20	5.29	18910
*Nos*SSUL	90	10250.54	6.46	8480
*Nos*M35	330	35434.51	8.63	29910
*Nt*RbcL	475	52679.96	6.44	70820
*Nt*RbcLΔN9	466	51737.93	6.43	70820
*Nt*RbcS	123	14559.65	5.19	42400
*At*RbcL	477	52736.79	5.88	69330
*At*RbcLΔN9	468	51794.75	5.87	69330
*At*RbcS	126	14797.04	5.71	33920

[1]all parameters were determined using Expasy-ProtParam [Gasteiger et al., 2005] and are given per monomer ; [2]molecular weight; [3]isoelectric point; [4]extinction coefficient.

2.1.4.3. Coomassie staining of protein gels after SDS- and native-PAGE

For the fixation and visualization of gel-bound proteins, protein gels were incubated in the Coomassie-based stain *Instant Blue* (Expedeon) for 1 h. Subsequently, gels were destained in deionized H_2O for approx. 12 h. For documentation, a flatbed scanner was used to scan the gel image.

2.1.4.4. Western blotting and immunodetection of proteins

For standard immunodetection of proteins after SDS- or native-PAGE, proteins were transferred onto nitrocellulose membranes (*Roti NC*, Carl Roth) using the semi-dry blotting procedure (*SEMI-PHOR*) in 50 mM Tris / 192 mM glycine / 0.1 % (w/v) SDS / 20 % methanol) at a constant current of 250 mA for 60 min. Due to the large size of protein complexes in the gel shift assay, an improved transfer protocol was employed. To facilitate protein complex egression after native-PAGE, proteins were transferred in a methanol-free buffer (50 mM Tris / 192 mM glycine / 0.1 % (w/v) SDS) on a highly hydrophobic polyvinylidene fluoride (PVDF) membrane, which prior to use had been pre-wetted in 50 % methanol, rinsed once in deionized H_2O and equilibrated in buffer. Proteins were transferred at a constant current of 250 mA for 2 h at 4 °C using a *Mini Trans-Blot Electrophoretic Transfer Cell* (Bio-Rad). Following protein transfer, membranes were incubated in 10 % (w/v) milk powder in Tris-buffered saline (TBS) (20 mM Tris-HCl pH 7.5 / 137 mM NaCl) to saturate unspecific binding sites. Subsequently, membranes were incubated in 1 % milk in TBS containing the primary antibody (Table **2.3**) for 1 h at 25 °C or over night at 4 °C. Incubation of membranes in 1 % milk in TBS containing the horseradish peroxidase (HRP) coupled secondary antibody was preceded and followed by three washes (10 min each) in TBS + 1 % Tween 20 (TBST). After brief incubation in Luminol containing substrate (*Luminata^{TM} Crescendo Western HRP substrate*), chemoluminescence was detected on a Fuji-LAS3000 detection system (Fujifilm-Corporation).

2.2. Size-exclusion chromatography coupled to multi-angle static light scattering (SEC-MALS)

SEC-MALS data were obtained in collaboration with Manajit Hayer-Hartl and Monica Zobawa. SEC-MALS was performed at indicated protein concentrations or

Table 2.3.: Antibodies

Antibody	Source	Reference
α-*Nt*Rubisco	rabbit	[Whitney et al., 2001]
α-*Se*RbcL	rabbit	[Bracher et al., 2015a]
α-*Nos*RbcS[1]	rabbit	This study*
α-*Nos*SSUL[2]	rabbit	This study*
α-*Nos*Rca[3]	rabbit	This study*

[1] raised against RbcS from *Nostoc* sp. PCC 7120
[2] raised against the SSUL module of Rca from *Nostoc* sp. PCC 7120 (*Nos*SSUL)
[3] raised against the AAA+ domain of *Nostoc* sp. PCC 7120 (*Nos*RcaΔC)
* available upon request to the corresponding author of [Flecken et al., 2020]

at $2\,\mathrm{g\,L^{-1}}$ of purified proteins. Static and dynamic light scattering was analyzed upon auto-injection of the sample onto a size exclusion chromatography (SEC) column (5 mm, 4.6x300 mm column, Wyatt Technology, WTC-030N5) equilibrated in 50 mM MMT (a tripartite composite buffer containing DL-malic acid, MES monohydrate and Tris base at a 1:2:2 molar ratio), pH 8.4 / 300 mM KCl / 10 mM KCl at $0.2\,\mathrm{mL\,min^{-1}}$ and 25 °C. Nucleotides (ADP, ATP or ATPγS) were added as indicated. The column was in line with the following detectors: a variable UV absorbance detector set at 280 nm (Agilent 1100 series), the DAWN EOS MALS detector (Wyatt Technology, 690 nm laser) and the Optilab rEXTM refractive index detector (Wyatt Technology, 690 nm laser) [Wyatt, 1993]. Molecular masses were calculated using the ASTRA software (Wyatt Technology) with the dn/dc value set to $0.185\,\mathrm{mL\,g^{-1}}$. The calibration standard was Bovine serum albumin (BSA) (Thermo Fisher Scientific). Alternatively, indicated samples were injected on a Superdex 200 HR 10/30 column (Cytiva Lifesciences formerly GE Healthcare) and analyzed using the Wyatt Dawn HELEOS 8-angle light scattering detector and Wyatt Optilab TrEX refractive index monitor linked to a 1260 infinity HPLC system (Agilent) comprising a 1310B isopump, a 1329B autosampler and G7114A DAD UV detector. Samples were prepared as described above and analyzed using the Astra software suite (Wyatt Technology).

2.3. Analysis of the distribution of Rca proteins across β-cyanobacterial groups and Rubisco RbcL subunit phylogeny construction

All 175 cyanobacterial genome assemblies and 507 corresponding nucleotide sequences with full genome representation and „refseq“ status available at the time of analysis (November 20th, 2020) were received from NCBI (https://www.ncbi.nlm.nih.gov/). Most information (e.g. taxonomic information, nucleotide-, assembly- and protein summary files (esummaries) and links) were accessed via the molecular biology database system Entrez of NCBI. The analysis was carried out using the statistical software R [The R Core team, 2020], specifically using the tidyverse ecosystem [Wickham et al., 2019] and communication with Entrez was mediated by rentrez [Winter, 2017]. Sequence data was handled using Biotrings [Pagès et al., 2020]. RbcL sequences were identified utilizing the stand-alone BLAST applications (BLAST+; version 2.10.1+) [Camacho et al., 2009] using RbcL from *Nostoc* sp. PCC 7120 (WP_010995693.1) as query. For exclusion of false hits an identity threshold of 74 % proved useful. Three assemblies (GCF_000025125.1, GCF_000829235.1, GCF_003574135.1) of endosymbiotic cyanobacterial species were excluded from further analysis due to the absence of RbcL sequences in these species. The remaining 172 cyanobacterial assemblies were represented by 118 unique RbcL sequences and subjected to multiple sequence alignment (msa) [Bodenhofer et al., 2015], calculation of a distance matrix (seqinr) [Charif and Lobry, 2007] and phylogenetic tree construction (ape, treeio, tidytree, ggtree) [Paradis and Schliep, 2019; Yu, 2020]. For the annotation of non-redundant RbcL sequences representing multiple assemblies, assembly information was randomly sampled from the set of corresponding assemblies that were represented by the given RbcL sequence. If this set of assemblies encompassed the assembly of a model species, assembly information of the model species was used to annotate the corresponding RbcL sequence.

Depicted RbcL sequence alignments were conducted with the full-length proteins of selected cyanobacterial and plant RbcL sequences that encompassed WP_010995693.1; *Nostoc* sp. PCC 7120 = FACHB-418, WP_012628807.1; *Cyanothece* sp. PCC 7425, WP_015126647.1; *Calothrix* sp. PCC 7507, WP_015173841.1; *Geitlerinema* sp. PCC 7407, WP_015168382.1; *Synechococcus* sp. PCC 7502, WP_015157218.1; *Chroococcidiopsis thermalis* PCC 7203, WP_015205060.1; *Crinalium epipsammum* PCC 9333, WP_068389226.1; *Leptolyngbya* sp. NIES-3755, WP_006194087.1; *Nodularia spumigena* UHCC 0039, WP_172355939.1; *Thermoleptolyngbya* sp. PKUAC-SCTA183,

F2X0C4; *Hevea brasiliensis*, Q09MH0; *Citrus sinensis*, Q0P3J3; *Ostreococcus tauri*, P00877; *Chlamydomonas reinhardtii*, P00876; *Nicotiana tabacum*, O03042; *Arabidopsis thaliana*, P00875; *Spinacia oleracea*, P0C510; *Oryza sativa*, P05698; *Hordeum vulgare*, A0A0M3TGJ0; *Larrea tridentata*, NP_904194.1; *Physcomitrium patens*, WP_045869249.1; *Fremyella diplosiphon* PCC 7601.

Putative Rca nucleotide sequences were identified by stand-alone BLAST using Rca from *Nostoc* sp. PCC 7120 (WP_010995702.1) as query. Rca sequences were selected based on a local search of a pre-formatted conserved domains database (ftp://ftp.ncbi.nih.gov/pub/mmdb/cdd/little_endian/Cdd_LE.tar.gz) by stand-alone BLAST using corresponding Rca protein sequences as query. The combined presence of one ccd motif CDD:365803 (pfam00004, AAA+ protein) and minimum one cdd motif CDD:226858 (COG4451, RbcS) were interpreted as indicative for cyanobacterial Rca sequences. Although CDD:215031 (PLN00020, Rca; provisional) was identified in three additional putative Rca sequences, these were excluded due to the absence of a predicted RbcS-like motif. Distribution of Rca across β-cyanobacterial groups was analyzed and mapped to the RbcL phylogenetic tree together with the taxonomic group of the represented assembly.

The alignments of Rca proteins were either conducted with the full-length protein or maturated protein (without the transit peptide required for chloroplast import) of cyanobacteria and plants, respectively, or with the AAA+ domain alone, as indicated. The transit peptides were predicted and processed using plant-specific default parameters in TargetP-2.0 ([Armenteros et al., 2019]) and for alignments of AAA+ domains alone, sequences of selected cyanobacterial and plant Rcas were restricted to intervals matching the query (residue 1-300 of *Nostoc* sp. PCC 7120, WP_010995702.1) in a stand-alone BLAST+ search. The selected Rca sequences comprised WP_010995702.1; *Nostoc* sp. PCC 7120 = FACHB-418, WP_012629660.1; *Cyanothece* sp. PCC 7425, WP_015126639.1; *Calothrix* sp. PCC 7507, WP_015173838.1; *Geitlerinema* sp. PCC 7407, WP_015167325.1; *Synechococcus* sp. PCC 7502, WP_015157222.1; *Chroococcidiopsis thermalis* PCC 7203, WP_015202507.1; *Crinalium epipsammum* PCC 9333, WP_068382484.1; *Leptolyngbya* sp. NIES-3755, WP_006194083.1; *Nodularia spumigena* UHCC 0039, WP_172356093.1; *Thermoleptolyngbya* sp. PKUAC-SCTA183, XP_021656011.1; *Hevea brasiliensis*, XP_006486397.1; *Citrus sinensis*, A0A454Y6I1; *Ostreococcus tauri*, P23489; *Chlamydomonas reinhardtii*, Q40460; *Nicotiana tabacum*, P10896; *Arabidopsis thaliana*, P10871; *Spinacia oleracea*, P93431; *Oryza sativa* subsp. japonica, Q40073; *Hordeum vulgare*, Q7X9A0; *Larrea tridentata*, XP_024392724.1; *Physcomitrium patens*, WP_045869747.1; *Fremyella diplosiphon*

PCC 7601. All alignments were displayed using ESPript3.0 ([Gouet et al., 1999], http://espript.ibcp.fr).

2.4. Buffer screen

The buffer screen was performed using the (capillary-based) differential scanning fluorimetry (nanoDSF) technique [Martin et al., 2015]. *Nos*Rca and *Nos*RcaΔC, obtained from initial crude purifications, were prepared in 50 mM Ethanolamine (ETA) / 100 mM NaCl / 10 mM $MgCl_2$ at $\sim 0.11\,g\,L^{-1}$ (corresponding to 2.4 and 3.4 µM, respectively) on ice using 0.65 mL low binding *SafeSeal Microcentrifuge Tubes* (Sorenson). Initial protein purification of *Nos*Rca and *Nos*RcaΔC was essentially carried out as described in chapter **2.5.2** with the modification that the buffer exchange during the final gelfitration into Tris based buffer was omitted and instead contained 50 mM ETA / 100 mM NaCl / 10 mM $MgCl_2$ / 5 % glycerol. For reproducible and comparable results, a highly standardized sample preparation was required, e.g. an identical sample volume and a consistent technique of capillary loading were found necessary. To this end, 18 µL of the prepared protein solution were added to 2 µL of each 10-fold buffer of the set of buffers to be tested (see Tables **2.4** and **2.5**) at a final concentration of $0.1\,g\,L^{-1}$ (corresponding to 2.1 and 3.0 µM, respectively). After 10 min, the individual protein solutions were moved to room temperature and loaded into *nanoDSF Grade Standard Capillaries* (Nanotemper) which were mounted on the capillary rack of the *Prometheus NT.48* (Nanotemper). All 48 individual capillaries were filled and loaded within 25 min. To assess the variability of the procedure and to detect potential sample loading bias, control samples in 50 mM ETA / 100 mM NaCl / 10 mM $MgCl_2$ were evenly distributed across the 48 positions and measured in quadruplicates. Upon excitation at 280 nm +/- 5 nm the machine records fluorescence at 350 nm and 330 nm. Thus not only tryptophane residues but also tyrosine residues are being excited. The experiment was conducted at a laser intensity of 20 % and a heating ramp from 20 °C to 80 °C with a heating rate of $1\,K\,min^{-1}$. Melting curve records were exported and processed according to an in-house R package *promelt* (kindly provided by Fabien Bonneau) within the tidyverse ecosystem [Wickham et al., 2019] in the statistical software R [The R Core team, 2020].

Table 2.4.: Design of the buffer screen[1]

Capillary	Buffer	pH	pH group
1	SPG	5.60	low
2	MMT	6.00	low
3	PCB	5.60	low
4	MES	5.70	low
5	Na-Citrate	5.50	low
6	BisTris	6.00	low
7	Na/K-P	5.80	low
8	ADA	6.20	low
9	HEPES	6.80	medium-low
10	Imidazol	6.50	medium-low
11	Tris	7.50	medium
12	ETA	9.20	high
13	SPG	6.40	medium-low
14	MMT	6.80	medium-low
15	PCB	6.40	medium-low
16	MES	6.10	low
17	Na-Citrate	6.00	low
18	BisTris	6.40	medium-low
19	Na/K-P	6.40	medium-low
20	ADA	6.60	medium-low
21	HEPES	7.20	medium
22	Imidazol	7.00	medium
23	Tris	7.90	medium-high
24	ETA	9.20	high
25	SPG	7.20	medium
26	MMT	7.60	medium
27	PCB	7.20	medium
28	MES	6.50	medium-low
29	Na-Citrate	6.50	medium-low
30	BisTris	6.80	medium-low
31	Na/K-P	7.00	medium
32	ADA	7.00	medium
33	HEPES	7.60	medium

...continues on next page...

Capillary	Buffer	pH	pH group
34	Imidazol	7.52	medium
35	Triphosphates	8.30	medium-high
36	ETA	9.20	high
37	SPG	8.00	medium-high
38	MMT	8.40	medium-high
39	PCB	8.00	medium-high
40	MES	6.90	medium-low
41	Na-Citrate	7.00	medium
42	BisTris	7.20	medium
43	Na/K-P	7.60	medium
44	ADA	7.40	medium
45	HEPES	8.00	medium-high
46	Imidazol	8.00	medium-high
47	Tris	8.80	high
48	ETA	9.20	high

[1]For buffer composition, see Table **2.5**

In brief, a slanted Boltzmann model was fitted to the intrinsic (tryptophan) fluorescence ratio $\frac{350\,\mathrm{nm}}{330\,\mathrm{nm}}$ (WFR) as a function of temperature (T):

$$WFR(T) = y_n(T) * \frac{(y_d(T) - y_n(T))}{(1 + exp(maxSlope * (T_m - T)))} \tag{2.1}$$

with $y_n(T)$ representing the baseline fluorescence ratio in the native state,

$$y_n(T) = a_n * T + b_n \tag{2.2}$$

$y_d(T)$ representing the baseline fluorescence ratio in the denatured state,

$$y_d(T) = a_d * T + b_d \tag{2.3}$$

and the melting temperature T_m as well as the slope of the model $maxSlope$. Melting curves with $S > 95\,\%$ quantile showed clear deviations from two-stage unfolding and were excluded from further analysis. Upon manual inspection, the remaining melting curves were inconspicuous and thus, parameters T_m, $maxSlope$ and S were extracted

and hierarchically ranked in a protein-wise fashion. Correlation of T_m and *maxSlope* emphasized that high T_m coincide with steep *maxSlope* and revealed buffers with potentially stabilizing effects on *Nos*Rca and *Nos*RcaΔC. The screen was repeated for each protein to verify reproducibility.

Table 2.5.: Composition of buffers in Bufferscreen[1]

Buffer	Composition[2]
SPG	0.125 M Succinic Acid / 0.5 M NaH_2PO_4 / 0.375 M Glycine, NaOH
MMT	0.2 M DL-Malic Acid / 0.4 M MES monohydrate / 0.4 M Tris, NaOH
PCB	0.4 M Sodium Propionate / 0.2 M Sodium Cacodylate trihydrate / 0.4 M Bis-Tris Propane, HCl
MES	1 M (2-(N-morpholino)ethanesulfonic acid) monohydrate
Na-Citrate	1 M Trisodium 2-hydroxypropane-1,2,3-tricarboxylate
BisTris	1 M Bis(2-hydroxyethyl)amino-tris(hydroxymethyl)methane
Na/K-P	1 M K-phosphate monobasic / Na-phosphate dibasic
ADA	1 M N-(2-Acetamido)iminodiacetic acid
HEPES	1 M 4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid
Imidazole	1 M Imidazole
Tris	1 M 2-Amino-2-hydroxymethyl-propane-1,3-diol
ETA	1 M Ethanolamine

[1] All buffers are amenable to cryo EM imaging of macromolecular protein complexes and produce low contrast images [Chari et al., 2015]
[2] Buffers represent a 10-fold concentrated stock solution.

2.5. Protein expression and purification

2.5.1. *Nos*Rca and mutants

Nostoc sp. PCC 7120 Rubisco activase (*Nos*Rca) was expressed as H_6ubi-fusion protein in *E. coli* BL21 Star cells harbouring pBAD33-*Ec*GroSEL and pHUE-*Nos*Rca plasmids for inducible expression. Cells were grown in 2x YT medium (5 g L^{-1} NaCl / 10 g L^{-1} yeast extract / 16 g L^{-1} tryptone / 0.5 % glycerol (v/v)) containing 10 mM KCl [Larimer and Soper, 1993] and appropriate antibiotics (35 mg L^{-1} chloramphenicol / 100 mg L^{-1} ampicillin) at 37 °C and 180 rpm to an OD_{600} 0.3-0.4. GroSEL was induced by addition of 0.4 % (w/v) L-arabinose. 1 h after chaperonin induction, the

temperature was reduced to 18 °C. Cells were grown for another hour before H_6ubi-*Nos*Rca expression was induced by addition of 0.2 mM β-D-1-thiogalactopyranoside (IPTG) and shaking was reduced to 120 rpm for 18 h. Cells were harvested by centrifugation at 4200 rpm for 25 min at 4 °C using an Avanti J6-MI centrifuge equipped with a JS 4.2 swinging-bucket rotor (Beckman Coulter), washed once in 50 mM Tris, pH 8 / 100 mM NaCl / 10 mM $MgCl_2$ and either flash frozen in liquid N_2 for storage or directly processed as described next. The harvested cells were resuspended and incubated in buffer A (50 mM MMT pH 8.4, 300 mM KCl, 10 mM $MgCl_2$, 5 % glycerol) containing 1 g L^{-1} lysozyme, 2.5 U mL^{-1} *Serratia marcescens* DNAse (*Sm*DNAse), Complete protease inhibitor cocktail (Roche) for 1 h prior to mechanical lysis using an EmulsiFlex C5 (Avestin, Inc) French press cell disruptor. After high-speed centrifugation at 100,000 x g for 40 min at 4 °C using an Optima XPN-100 ultracentrifuge equipped with a 45-Ti rotor (Beckman Coulter), the supernatant was loaded onto a gravity TALON metal affinity column (Takara), washed with 10 column volumes (CV) buffer A / 20 mM imidazole and eluted with buffer A / 200 mM imidazole. To avoid high total protein concentration and precipitation of the protein upon cleavage of the H_6ubi moiety in the next step, the TALON eluate was diluted threefold in buffer A / 5 mM 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS) / 10 % glycerol / 5 mM 2-mercaptoethanol to reach a final total protein concentration below ~ 0.5 g L^{-1}. The H_6ubi moiety was cleaved by the His-tagged deubiquitinating enzyme Usp2 [Baker et al., 2005; Catanzariti et al., 2004] at 10 °C overnight. Note that Usp2 cleaves the H_6ubi moiety after a short Gly-Gly linker C-terminal of ubiquitin and leaves the target protein essentially unaltered. The cleaved protein was desalted with a HiPrep 26/10 desalting column (Cytiva lifesciences) equilibrated in 50 mM MMT / 10 mM KCl / 5 % glycerol and subsequently loaded on a MonoQ column (GE). The protein was eluted with a 10 CV gradient (0.01 - 0.5 M KCl). Fractions containing the protein of interest were concentrated to ~ 5 mL using Vivaspin 20 MWCO 10000 concentrators (Cytiva lifesciences) and subsequently applied to a HiLoad 16/60 Superdex 200 pg (Cytiva lifesciences) column equilibrated in 50 mM MMT pH 8.4, 100 mM KCl, 10 mM $MgCl_2$, 5 % glycerol. Fractions containing *Nos*Rca were pooled and concentrated (Vivaspin 20 MWCO 10000) to > 50 µM (hexamer) and flash frozen in liquid N_2. Reduced *Nos*Rca was kept in reducing condition by addition of 5 mM DTT to all buffers after cleavage of the H_6ubi moiety. Oxidized *Nos*Rca was purified in buffer free of reducing agent after cleavage of the H_6ubi moiety and left on ice for 48 hours to air oxidize prior to storage at -80 °C.

*Nos*Rca mutants containing the SSUL module were expressed and purified under reducing conditions as described above. *NosRca* mutants lacking the SSUL module (residues 2 - 291; ΔC) were expressed and purified under reducing conditions as described above, but the threefold dilution step prior to cleavage of the H_6ubi moiety was omitted, since these proteins were less prone to aggregate and precipitate.

2.5.2. *Nos*RcaΔC and *Nos*SSUL for crystallography

Rca from *Nostoc* sp. PCC 7120 lacking the C-terminal SSUL module (residues 2-291; *Nos*RcaΔC) was expressed as H_6ubi-fusion protein in *E. coli* BL21 harbouring arabinose-inducible pBAD33-*Ec*GroSEL and pHUE-*Nos*Rca plasmids. Cells were grown in Luria-Bertani (LB) medium containing appropriate antibiotics (35 mg L^{-1} chloramphenicol / 100 mg L^{-1} ampicillin) at 37 °C and 180 rpm to OD_{600} 0.3-0.4. *Ec*GroSEL was induced by addition of 1 % (w/v) L-arabinose. Two hours after induction, the temperature was reduced to 18 °C and expression of H_6ubi-*Nos*RcaΔC was induced by addition of 0.5 mM IPTG for 20 h. Cells were harvested by centrifugation at 4000 rpm for 25 min at 4 °C using an Avanti J6-MI centrifuge equipped with a JS 4.2 swinging-bucket rotor (Beckman Coulter) and either flash frozen in liquid N_2 for storage or directly processed as described next. Cells were resuspended in ETA lysis buffer (50 mM ETA, pH 8 / 300 mM NaCl / 10 mM imidazole / 5 % (v/v) glycerol) containing 1 mM PMSF, 10 mM 2-mercaptoethanol, 1 g L^{-1} lysozyme and 5 U mL^{-1} benzonase and were lysed by ultrasonication using Sonicator 3000 (Misonix) over 15 cycles of 15 s bursts followed by 75 s of intermission. After centrifugation at 20,000 x g for 30 min at 4 °C using an Avanti J-25 centrifuge equipped with a JA 25.50 rotor (Beckman Coulter), the supernatant was loaded on a gravity nickel nitrilotriacetic acid (Ni-NTA) metal affinity column (Qiagen), washed with 10 CV ETA lysis buffer / 25 mM imidazole and eluted with 30 mL ETA lysis buffer pH 9.2 / 200 mM imidazole. Protein-containing fractions were pooled and the H_6ubi moiety was cleaved by Usp2 [Baker et al., 2005; Catanzariti et al., 2004] at 4 °C overnight. After dialysis (ZelloTrans) against 50 mM ETA pH 9.2 / 10 mM NaCl, *Nos*RcaΔC was loaded on a Mono Q HR 16/10 column (Cytiva lifesciences). After elution with a 10 CV gradient (0.01 - 1 M NaCl) fractions containing *Nos*RcaΔC were pooled, concentrated using Centricons (Millipore) and subsequently applied to a Superdex 200 10/300 GL (Cytiva lifesciences) column in 20 mM Tris-HCl, pH 8 / 50 mM NaCl / 5 mM $MgCl_2$ / 5 % glycerol. Fractions containing *Nos*RcaΔC were pooled, concentrated to > 12 µM (hexamer) and flash frozen in liquid N_2.

*Nos*SSUL, the SSUL module of *Nos*Rca (residues 325 - 414), was expressed as a H_6ubi-fusion protein in the absence of *E. coli* chaperonins from pHUE-*Nos*SSUL. Cells were grown in Luria-Bertani (LB) medium containing 100 mg L^{-1} ampicillin at 37 °C and 180 rpm to OD_{600} 0.3-0.4. *Nos*SSUL expression was induced by addition of 0.2 mM IPTG and cells were shifted to 22 °C and 120 rpm for 18 h. Cells were harvested by centrifugation at 4000 rpm for 25 min at 4 °C using an Avanti J6-MI centrifuge equipped with a JS 4.2 swinging-bucket rotor (Beckman Coulter) and either flash frozen in liquid N_2 for storage or directly processed as described next. Cells were resuspended in ETA lysis buffer (50 mM ETA, pH 8 / 300 mM NaCl / 10 mM imidazole / 5 % (v/v) glycerol) containing 1 mM PMSF, 10 mM 2-mercaptoethanol, 1 g L^{-1} lysozyme and 5 U mL^{-1} benzonase and were lysed by ultrasonication on ice using a Sonicator 3000 (Misonix) over 15 cycles of 15 s bursts followed by 75 s of cooling. After centrifugation at 20,000 x g for 30 min at 4 °C using an Avanti J-25 centrifuge equipped with a JA 25.50 rotor (Beckman Coulter) the supernatant was loaded on a nickel nitrilotriacetic acid (Ni-NTA) metal affinity column (Qiagen) operated with gravity flow, washed with 10 CV ETA lysis buffer / 25 mM imidazole and eluted with 30 mL ETA lysis buffer pH 9.2 / 200 mM imidazole. Protein-containing fractions were pooled and the H_6ubi moiety was cleaved with Usp2 [Baker et al., 2005; Catanzariti et al., 2004] at 4 °C over night. Subsequent to dialysis against ETA lysis buffer after the first Ni-NTA column, *Nos*SSUL was applied to a second Ni-NTA column for the removal of Usp2, the H_6ubi moiety and any uncleaved protein. The flow through was concentrated and applied to a HiLoad 16/60 Superdex 75 pg (Cytiva lifesciences) column in 20 mM Tris-HCl pH 8 / 50 mM NaCl / 5 mM $MgCl_2$ / 5 % glycerol. *Nos*SSUL containing fractions were pooled, concentrated and flash frozen in liquid N_2.

The crystallized proteins *Nos*RcaΔC and *Nos*SSUL were generated by Leonhard Popilka, a previous graduate student. Protein expression and purification was modified (see Chapter **2.4** and **2.5.1**) to achieve higher yields and to enable the production of full-length *Nos*Rca, which showed a high propensity to precipitate under the described conditions.

2.5.3. *Nos*Rubisco and mutants

Rubisco from *Nostoc* sp. PCC 7120 (*Nos*Rubisco) was expressed as a RbcL N-terminal H_6ubi-fusion protein in *E. coli* BL21 Star cells harbouring the IPTG-inducible pET11a-*Nos*GroSEL-H_6ubi-*Nos*LS plasmid. Cells were grown in 2x YT medium (5 g L^{-1} NaCl / 10 g L^{-1} yeast extract / 16 g L^{-1} tryptone / 0.5 % glycerol

(v/v)) containing 10 mM KCl [Larimer and Soper, 1993] and 100 mg L^{-1} ampicillin at 37 °C and 180 rpm to OD_{600} 0.6-0.8. *Nostoc* sp. PCC 7120 chaperonins (*Nos*GroSEL) and *Nos*Rubisco were induced by addition of 0.5 mM IPTG and shifted to 22 °C and 120 rpm for 22 h. Cells were harvested by centrifugation at 4200 rpm for 25 min at 4 °C using Avanti J6-MI equipped with a JS 4.2 swinging-bucket rotor (Beckman Coulter), washed once in 50 mM Tris, pH 8 / 100 mM NaCl / 10 mM $MgCl_2$ and either flash frozen in liquid N_2 for storage or directly processed as described next. The harvested cells were resuspended and incubated in buffer A (50 mM MMT pH 8.4, 300 mM KCl, 10 mM $MgCl_2$, 5 % glycerol) containing 1 g L^{-1} lysozyme, 2.5 U mL^{-1} *Sm*DNAse, complete protease inhibitor cocktail (Roche) for 1 h prior to mechanical lysis using EmulsiFlex C5 (Avestin, Inc). After high speed centrifugation at 100 000 x g for 40 min at 4 °C using an Optima XPN-100 ultracentrifuge equipped with a 45-Ti rotor (Beckman Coulter) the supernatant was loaded onto a gravity TALON metal affinity column (Takara), washed with 10 CV buffer A / 20 mM imidazole and eluted with buffer A / 200 mM imidazole. Cleavage of the H_6ubi moiety by Usp2 was carried out overnight at 10 °C in presence of 5 mM 2-mercaptoethanol. After removal of imidazole on a HiPrep 26/10 desalting column (Cytiva lifesciences) equilibrated in buffer A, the eluate was applied to a second TALON column for removal of Usp2, the H_6ubi moiety and any uncleaved protein. The flow through was concentrated to $\sim$ 5 mL using Vivaspin 20 MWCO 10000 concentrators (Cytiva lifesciences) and subsequently applied to a HiLoad 16/60 Superdex 200 pg (Cytiva lifesciences) equilibrated in 50 mM MMT pH 8.4, 100 mM KCl, 30 mM $NaHCO_3$, 10 mM $MgCl_2$, 5 % glycerol. Fractions containing *Nos*Rubisco were pooled, concentrated to $\sim$ 24 µM (hexadecamer), aliquoted and flash frozen in liquid N_2.

2.5.3.1. Optimization of *Nos*Rubisco production

To optimize recombinant Rubisco production, test expressions were carried out at small scale (0.1 L) and expression conditions were adapted to the expression system used. For Rubisco expression using pBAD33-*Ec*GroSEL and pET28b-*Nos*LXSRaf1 or pET28b-*Nos*LS, BL21 (DE) cells harboring the corresponding plasmids were grown in 2x YT medium containing 10 mM KCl, 35 mg L^{-1} chloramphenicol and 50 mg L^{-1} kanamycin at 37 °C and 180 rpm to OD_{600} 0.3-0.4. Expression of *Ec*GroSEL was induced by addition of 0.4 % (w/v) L-arabinose. 1 h after chaperonin induction, the temperature was reduced to 22 °C. Cells were grown for another hour before the expression of *Nos*Rubisco genes and assembly factors was induced by addition of 0.4 mM IPTG and shaking was reduced to 120 rpm for 24 h. For Rubisco expression using pET28b-*Nos*LXSRaf1 and either pET11-*Ec*GroSEL or pET11a-*Nos*GroSEL

or pET11a-*Nos*GroSEL1-GroEL2, BL21 Star cells harboring corresponding plasmids were grown in 2x YT medium containing 10 mM KCl, 100 mg L^{-1} ampicillin and 50 mg L^{-1} kanamycin at 37 °C and 180 rpm. When OD_{600} approximated 0.3-0.4., temperature was reduced to 22 °C and cells were grown until OD_{600} 0.6-0.8 was reached ($\sim$ 1 h). Simultaneous expression of the chaperonins, *Nos*Rubisco genes and assembly factors was induced by the addition of 0.5 mM IPTG. Shaking was reduced to 120 rpm and cells were cultured for additional 24 h. Rubisco expression tests using pET28b-*Nos*LS and either pET11a-*Nos*GroSEL or pET11a-*Ec*GroSEL were carried out alike. Similarly, for Rubisco expression using the pET11a-*Nos*GroSEL1-GroEL2, pET28b-*Nos*LS, and either the pCDF1b-*Nos*RbcX or pCDF1b-*Nos*Raf1 or pCDF1b-*Nos*RbcX-Raf1, expression conditions were essentially the same. Small modifications comprised the addition of 100 mg L^{-1} spectinomycin for the selection for the pCDF plasmid and an increase of the IPTG concentration to 0.6 mM.

For analysis, 2 mL aliquots of the cultures were harvested by centrifugation at 6800 x g for 5 min at 4 °C using the Eppendorf centrifuge 5417R equipped with a FA 45-30-11 rotor (Eppendorf). Cells were resuspended and incubated in buffer A containing 1 g L^{-1} lysozyme, 2.5 U mL^{-1} *Sm*DNAse, complete protease inhibitor cocktail (Roche) for 30 min prior to lysis by sonication using a Branson sonifier 250 (Emerson) with output control set to 2 using a 80 % duty cycle and a total processing time of 3 min. The soluble fraction of the lysate was obtained by removing the cell debris at 25000 x g for 20 min at 4 °C using the Eppendorf Centrifuge 5417R equipped with a FA 45-30-11 rotor (Eppendorf). Both the non-fractionated lysate (total protein, T) and the soluble protein (S) were analyzed by SDS- and native PAGE and immunoblotting. The carboxylation activity of Rubisco in the cleared supernatant was analyzed as described in Chapter **2.10.1**.

2.5.4. *Nt*RubiscoΔN9 and *At*RubiscoΔN9

*Nt*RubiscoΔN9 was recombinantly expressed in *E. coli* BL21 Star cells harbouring the IPTG-inducible plasmids pET11a-*At*Cpn60α/βCpn20-H_6ubi-*Nt*LΔN9 and pCDF-*Nt*X$SR_1R_2B_2$. *At*RubiscoΔN9 was expressed in *E. coli* BL21 Star cells harbouring the IPTG-inducible plasmids pET11a-*At*Cpn60α/βCpn20-H_6ubi-*At*LΔN9-*At*S and p*At*R1/R2/RX/B2 [Aigner et al., 2017]. *Nt*RubiscoΔN9 and *At*RubiscoΔN9 were expressed as RbcL N-terminal H_6ubi-fusion proteins. Cells were grown in 0.1 L ZYP-5052 auto-induction medium [Lin et al., 2020; Studier, 2005] at 37 °C with shaking at 160 rpm for 6 h. Protein expression was continued for a further 22 h at the reduced temperature of 23 °C with shaking at 120 rpm. Cells were harvested by

centrifugation at 4200 rpm for 25 min at 4 °C using Avanti J6-MI equipped with a JS 4.2 swinging-bucket rotor (Beckman Coulter), washed once in 50 mM Tris, pH 8 / 100 mM NaCl / 10 mM $MgCl_2$ and either flash-frozen in liquid N_2 for storage or directly processed as described next. The harvested cells were re-suspended and incubated in buffer A (50 mM MMT, pH 8.4 / 300 mM KCl / 10 mM $MgCl_2$ / 5 % glycerol) containing 1 g L^{-1} lysozyme, 2.5 U mL^{-1} *Sm*DNAse, Complete protease inhibitor cocktail (Roche) for 1 h prior to lysis by ultrasonication (Sonicator 3000, Misonix). After high-speed centrifugation at 100,000 x g for 40 min at 4 °C (Optima XPN-100; 45-Ti rotor - Beckman Coulter), the supernatant was loaded onto a Hi-Trap (GE) TALON Crude (Takara) metal affinity column equilibrated in buffer A. After a 10 CV wash with buffer A, the bound protein was eluted with a gradient from 0 - 200 mM imidazole for the separation of H_6ubi-Rubisco from *At*Cpn60α/β. The absence of *At*Cpn60α/β subunits ($\sim$ 58 kDa) from fractions containing H_6ubi-*Nt*RbcLΔN9 or H_6ubi-*At*RbcLΔN9 ($\sim$ 62 kDa) was verified by immunodetection using α-*Nt*RbcL and α-*At*Cpn60α/β. Fractions showing the presence of Rubisco but not of Cpn60α/β were pooled and subjected to Usp2 digest for 8 h at 10 °C in presence of 5 mM 2-mercaptoethanol to cleave the H_6ubi moiety. Note, that N-terminal processing of RbcL in the *E. coli* lysate is avoided, because Usp2 mediated cleavage occurs at a partially purified stage and generates the intact *Nt*RbcLΔN or *At*RbcLΔN, respectively [Ng et al., 2020]. After removal of imidazole using a HiPrep 26/10 desalting column (Cytiva lifesciences) equilibrated in 50 mM MMT pH 8.4 / 100 mM KCl / 5 % glycerol, the desalted sample was applied to a second TALON (Takara) metal affinity column for removal of Usp2, the H_6ubi moiety and any uncleaved protein. The flow through was concentrated using Vivaspin 20 MWCO 10000 concentrators to $\sim$ 8 µM (hexadecamer), aliquoted and flash-frozen in liquid N_2.

2.5.5. *Nos*SSUL, *Nos*Rbcs, *Nos*M35

RbcS, M35 and the SSUL module of Rca (residues 325 - 414) from *Nostoc* sp. PCC 7120 (*Nos*Rbcs, *Nos*M35, *Nos*SSUL, respectively) were expressed as H_6ubi-fusion proteins from pHUE-*Nos*RbcS, pHUE-*Nos*M35 and pHUE-*Nos*SSUL, respectively, in *E. coli* BL21 cells. Cells were grown in 2x YT medium (5 g L^{-1} NaCl / 10 g L^{-1} yeast extract / 16 g L^{-1} tryptone / 0.5 % glycerol (v/v)) containing 100 mg L^{-1} ampicillin at 37 °C and shaking at 180 rpm to OD_{600} 0.6-0.8. Protein expression was induced by addition of 0.2 mM IPTG and cells were shifted to 22 °C and 120 rpm for 18 h. Cells were harvested by centrifugation at 4200 rpm for 25 min at 4 °C using an Avanti J6-MI centrifuge equipped with a JS 4.2 swinging-bucket rotor (Beckman

Coulter), washed once in 50 mM Tris, pH 8 / 100 mM NaCl / 10 mM $MgCl_2$ and either flash-frozen in liquid N_2 for storage at −80 °C or directly processed as described next. *Nos*Rbcs, *Nos*M35, *Nos*SSUL were purified employing the two-stage IMAC purification as described for *Nos*Rubisco. Briefly, the harvested cells were resuspended and incubated in buffer A containing 1 g L^{-1} lysozyme, 2.5 U mL^{-1} *Sm*DNAse, Complete protease inhibitor cocktail (Roche) for 1 h prior to mechanical lysis. After high-speed centrifugation at 100,000 x g for 40 min at 4 °C the supernatant was loaded onto a gravity TALON metal affinity column (Takara), washed with 10 CV buffer A / 20 mM imidazole and eluted with buffer A / 200 mM imidazole. Cleavage of the H_6ubi moiety with Usp2 was carried out over night at 10 °C in presence of 5 mM 2-mercaptoethanol. After cleavage, the protein solution was passed through a 0.22 µm syringe filter with PVDF membrane (Starlab) and applied onto a HiPrep 26/10 desalting column (Cytiva lifesciences) for removal of imidazole. The eluate was applied to a second TALON column for removal of Usp2, the H_6ubi moiety and any uncleaved protein. The flow through was concentrated to ∼ 5 mL using Vivaspin 20 MWCO 3000 concentrators (Cytiva lifesciences) and subsequently applied to a HiLoad 16/60 Superdex 75 pg (Cytiva lifesciences) equilibrated in 50 mM MMT pH 8.4, 100 mM KCl, 10 mM $MgCl_2$, 5 % glycerol. Fractions containing the respective proteins were pooled, concentrated, aliquoted and flash-frozen in liquid N_2. Reduced *Nos*M35 and *Nos*SSUL were kept in reducing conditions by addition 5 mM DTT to all buffers after cleavage of the H_6ubi moiety. Oxidized *Nos*M35 and *Nos*SSUL was purified in buffer free of reducing agent after cleavage of the H_6ubi moiety and left on ice for 48 hours to air-oxidize prior to storage at −80 °C. *Nos*RbcS was purified in buffer devoid of reducing agent after cleavage of the H_6ubi moiety by Usp2 and flash-frozen in liquid N_2 for storage at −80 °C immediately after fractionation in 50 µL aliquots.

2.6. CABP-Synthesis

2-Carboxyarabinitol-1,5-diphosphate (CABP) was synthesized according to Pierce et al. [1980]. In brief, 100 µmol RuBP were incubated with 2-fold molar excess KCN in 5 mL 0.1 M Tris-acetate pH 8.3 for 48 h at 25 °C in 5 mL Microcentrifuge Tubes (Eppendorf). The racemic mixture of 2-carboxyribitol-1,5-diphosphate (CRBP) and CABP was treated with AG50W-X8 H^+ resin (Bio-Rad) for the removal of potassium and other cationic impurities, and the resin was removed by filtration through empty Poly-Prep chromatography columns (Bio-Rad). The resin was washed with 0.5 CV deionized water and the flow-through was lyophilizated. For the separation of the

enantiomers, the lactonized products were dissolved in 3 mM HCl and applied to a 120 mL AG1X8 Cl^- (Bio-Rad) column equilibrated in 3 mM HCl. The products were eluted with a 4 L gradient of 0 - 0.4 M LiCl and each of the 50 mL fractions were assayed for total phosphate [Chifflet et al., 1988]. CABP containing fractions were pooled and reduced to a volume of 50 mL in a rotary evaporator at 30 °C. Addition of a 3-fold molar excess of barium acetate to the expected amount of CABP (maximum 50 µmol) precipitated the CABP as barium salt (1 h at −20 °C). The precipitate was collected by centrifugation at 5,000 x g for 20 min at 4 °C using an Allegra X-15R centrifuge equipped with the SX4750 swinging-bucket rotor (BeckmanCoulter) and re-dissolved by addition of acid-washed AG50W-X8 H^+ resin. Subsequently, the resin was removed by passage through an empty Poly-Prep chromatography column (Bio-Rad) and the purified CABP was freeze-dried and re-dissolved in 50 mM Bicine-NaOH pH 9.3. Complete saponification was ensured by incubation on ice for 24 h. The final CABP concentration was determined according to [Chifflet et al., 1988] and $\sim$ 30 mM aliquots were flash frozen in liquid N_2 and stored at −80 °C until use.

2.7. Generation of Antisera

Antisera for the detection of *Nos*RcaΔC, *Nos*SSUL and *Nos*RbcS was generated in the Animal Facility of the MPIB. Recombinantly produced *Nos*RcaΔC, *Nos*SSUL or *Nos*RbcS were prepared in 250 µL of 50 mM MMT, pH 8.4 / 50 mM KCl / 10 mM $MgCl_2$ at 0.8 g L^{-1}. Stable water-in-oil emulsions were prepared by 1:1 dilution of the protein solutions in adjuvants and vigorous mixing. As adjuvant TiterMax Gold-Adjuvans (Merck) was used for the first immunization while both subsequent boosts (day 40 and day 70, respectively) were prepared using Freund's Adjuvant incomplete. Injections were applied subcutaneously into the rabbits. Ten days after injection of the third boost the blood from the sacrificed animals was received from the Animal Facility. The blood was stirred for 1 min with a glass rod, incubated at 20 °C for 1 h, stirred again and stored at 4 °C for $\sim$ 12 h. Coagulated components were removed by centrifugation at 5000 g for 30 min at 4 °C using an Allegra X-15R centrifuge equipped with the SX4750 swinging-bucket rotor (BeckmanCoulter). The supernatant was collected, divided in $\sim$ 1 mL aliquots and stored at −20 °C. The serum was qualitatively analyzed by immunoblotting. All antisera are available upon request to the corresponding author of [Flecken et al., 2020].

2.8. Crystallography

The crystal structures of *Nos*RcaΔC and *Nos*SSUL were obtained in collaboration with Leonhard Popilka, Andreas Bracher and David Cobessi. Details regarding crystallization, data collection, structure solution, refinement and analysis are published in [Flecken et al., 2020]. The crystallographic models and structure factors for *Nos*RcaΔC-Gd complex, *Nos*RcaΔC and *Nos*SSUL have been deposited to the Worldwide Protein Data Bank (wwPDB) under accession codes 6Z1D, 6Z1E and 6HAS, respectively.

2.9. Cryo-EM analysis and single particle reconstruction

The cryo-EM analysis and reconstruction were conducted in collaboration with Huping Wang and Andreas Bracher. Details regarding sample preparation, grid freezing, data collection, image processing, model building, refinement and analysis are published in [Flecken et al., 2020]. The models and the electron density maps for *Nos*RcaΔC:Rubisco and *Nos*Rca:Rubisco have been deposited to the wwPDB under PDB/EMDB accession codes 6Z1F/EMD-11028 and 6Z1G/EMD-11029, respectively. The local electron density map for *Nos*RcaΔC in the *Nos*RcaΔC:Rubisco complex has been deposited under EMDB accession code EMD-11575.

2.10. Biochemical assays

2.10.1. Rubisco activity and reactivation assays

Rubisco reactivation was measured by incorporation of radiolabelled CO_2 into acid-stable products as described previously [Carpentier, 2011; Mueller-Cajar et al., 2011; Robinson and Portis, 1989]. Rubisco was activated by incubation at 15 µM (active site concentration) in 100 mM MMT pH 8.4 / 60 mM $NaHCO_3$ / 20 mM $MgCl_2$ at 25 °C for 10 min. Subsequently, ECM and ECM.CABP were formed by 1:1 dilution in H_2O or 15 µM CABP dissolved in H_2O, respectively. The decarbamylated Rubisco (E) was obtained by buffer exchanging Rubisco into 100 mM MMT pH 8.4 / 50 mM KCl / 4 mM EDTA using a *Micro Bio-Spin 6* column (Bio-Rad). RuBP-inhibited Rubisco (E.RuBP) was formed by incubation of E in 0.5 mM RuBP for 30 min at

25 °C. XuBP-inhibited Rubisco (E.XuBP) was formed by incubation in XuBP at 50-, 200-, 400- or 600- fold molar excess over active sites for 1 h at 25 °C. Rubisco activity and reactivation assays were performed in 50 mM MMT pH 8.4, 50 mM KCl, 30 mM [14 C]$NaHCO_3$ (14 Bq / nmol), 10 mM $MgCl_2$, 3 mM phosphocreatine, 50 U / ml phosphocreatine kinase at 25 °C. Time course experiments were initiated by addition of ECM, ECM.CABP, E, E.RuBP or E.XuBP to the reaction mix containing RuBP (0.4 mM), Rca and ATP as indicated. Every 2 min a 60 µL aliquot was quenched in 20 µL 50 % formic acid. Excess [14 C]$NaHCO_3$ was evaporated and incorporation of the non-volatile product was determined by liquid scintillation counting on a Hitachi LSC 8000. To robustly estimate the specific activity and background counts, corresponding samples were measured in 5 or 6 replicates per assay. For the demonstration of ATP dependence, *Nos*Rca or *Nos*RcaΔC were included at 0.5 µM (hexamer) in presence or absence of ATP (3 mM). Dose dependent reactivation was always performed in presence of ATP (3 mM) and at indicated *Nos*RcaΔC concentrations (0.125, 0.25, 0.5, 1.25 µM hexamer). 8 min (single point) reactivation assays were conducted at 1 mM RuBP to ensure steady state kinetics of the ECM control. The Rubisco concentration was constant at 0.25 µM active sites for all Rubisco activity and Rubisco reactivation experiments.

To determine Rubisco activity in cell lysates, the total soluble protein concentration was determined according to Bradford [1976], adjusted to $3\,g\,L^{-1}$ and aliquots of 20 µL were assayed in 50 mM MMT, pH 8.4 / 50 mM KCl, 30 mM $NaH\,^{14/12}CO_3$ (17.5 Bq / nmol), 10 mM $MgCl_2$. The reactions were started by addition of RuBP at a final concentration of 0.4 mM. Reactions were stopped after 4 min and processed as described above. For comparison, the amount of CO_2 fixed within 1 min was normalized to 1 mg total protein of the lysate analyzed.

2.10.2. ATP hydrolysis assay

To measure the ATPase activity of *Nos*Rca, *Nos*RcaΔC and mutants, ATP hydrolysis was enzymatically coupled to NADH oxidation and measured spectrophotometrically at 340 nm and 25 °C [Carpentier, 2011; Mueller-Cajar et al., 2011; Robinson and Portis, 1989]. The assay was carried out in ATPase buffer (100 mM MMT pH 8.4 / 20 mM KCl / 10 mM $MgCl_2$ / 5 mM DTT / 2 mM phosphoenolpyruvate / 0.3 mM NADH / 2 mM ATP / 30 U/ml pyruvate kinase / 45 U/ml lactic dehydrogenase (Sigma P0294)) and the measurement was started by addition of *Nos*Rca, *Nos*RcaΔC, or mutants (*Nos*Rca R352E/R352E; *Nos*Rca V91E/R92G/G93N/L244D/N248A/L250A; *Nos*Rca Y116A; *Nos*Rca Y289A; *Nos*RcaΔC P140G/Y143A/D144A;

*Nos*RcaΔC V91E/R92G/G93N; *Nos*RcaΔC L244D/N248A/L250A) to a final concentration of 0.5 µM (hexamer). To measure ATPase activity of *Nos*Rca and *Nos*RcaΔC in the presence of activated (*Nos*ECM) or inhibited (*Nos*ECM.CABP) Rubisco (0.25 µM holoenzyme), the salt concentration in the assay buffer was increased to a final of 200 mM KCl, to avoid scattering due to LLPS of Rca and Rubisco at 20 mM KCl. All ATP turnover rates are stated with respect to the concentration of ATP binding sites (one per subunit).

2.10.3. Ellman assay

Free protein thiols were quantified by the 5,5'-dithiobis-(2-nitrobenzoic acid) (DTNB) / cystamine method as described by Riener et al. [2002] with modifications. *Nos*M35 or *Nos*Rca ($\sim$ 100 µM) purified in presence of reducing agent was buffer exchanged on a *Micro Bio-Spin 6* column (Biorad) to 50 mM MMT, pH 8.4 / 100 mM KCl / 10 mM $MgCl_2$ / 5 % glycerol to remove the reducing agent DTT and then heat-denatured by incubation for 5 min at 95 °C. Following centrifugation at 18 000 x g for 5 min, the protein pellets were dissolved in 5 M guanidinium hydrochloride (GdnHCl) / 100 mM PBS, pH 8.2 / 1 mM EDTA. Proteins purified in absence of reducing agent were treated equally, without prior buffer exchange. As control, a fully reduced protein sample was generated by incubation of denatured *Nos*M35 or *Nos*Rca in 30 mM DTT for 15 min at 25 °C and the reducing agent was removed by a second buffer exchange in 5 M guanidinium hydrochloride (GdnHCl) / 100 mM PBS, pH 8.2 / 1 mM EDTA prior to the analysis. Since GdnHCl affects the absorbance coefficient of DTNB, a cysteine standard for the relation of absorbance readings to the number of thiols was prepared as previously described [Riener et al., 2002]. The assay reagent was prepared as combined Ellman/cystamine stock solution (10 mM DTNB / 10 mM cystamine, pH 7.0), immediately frozen at −20 °C and thawed aliquots were used within 30 min. The assay was conducted at a final concentration of 178 µM Ellman/cystamine in 5 M GdnHCl) / 100 mM PBS, pH 8.2 / 1 mM EDTA and the protein concentration was adjusted to match the linear range within the detection interval ($\sim$ 0.5 - 35 µM free thiols). The assay was conducted in a total volume of 56 µL in a 384 well microplate format (Greiner, clear/polystyrene) and free thiols were quantified by measuring absorbance at 412 nm (Abs_{412}) on a CLARIOstar micro plate reader (BMG labtech). The Abs_{412} signal of protein samples and the cysteine standard was recorded every 2 min and shown to reach saturation within 2 min. Abs_{412} values obtained from the protein solution alone were subtracted from Abs_{412} of samples and the cysteine standard was used to quantify the number of free

thiols. The oxidation state per protein is given as $\frac{[thiols_{quant}]}{[thiols_{max}]}$ in %, with the concentration of quantified thiols $[thiols_{quant}]$ and $[thiols_{max}]$ being the product of protein concentration and total number of cysteines per protein. The protein concentration per sample was determined by UV spectroscopy in 5 M GdnHCl / 100 mM PBS, pH 8.2 / 1 mM EDTA based on the theoretical extinction coefficient of the protein (Table **2.2**).

2.10.4. Gel shift assay

Complex formation between *Nos*Rubisco and *Nos*Rca was monitored by electrophoretic mobility in native PAGE. *Nos*Rubisco (ECM and ECM.CABP) was co-incubated with *Nos*Rca at indicated oligomer ratios (1:0.5, 1:1, 1:2) in sample buffer (60 mM Tris - HCl pH 6.8 / 0.05 % (w/v) bromophenol blue / 10 % (v/v) glycerol) containing 2 mM DTT for 30 min on ice. As control, *Nos*Rubisco and 1.5 µM *Nos*Rca (hexamer) were prepared individually. The concentration of *Nos*Rubisco was constant at 0.75 µM (holoenzyme). Slightly different protein concentrations were used to demonstrate the species-specific interaction of the SSUL domain (indicated in the figure legend). Following incubation on ice, 4 µl of the reaction mix were analyzed by native PAGE, Coomassie staining and immunodetection using the α-*Se*RbcL or α-*Nt*RbcL antibody raised against RbcL from *Synechococcus elongatus* PCC 6301 [Bracher et al., 2015a] or Rubisco from *N. tabacum* (a kind gift from S. Whitney [Whitney et al., 2001]), respectively. *Nos*ECM and *Nos*ECM.CABP were essentially prepared as described in chapter **2.10.1**.

2.10.5. Turbidity assay

Formation of condensates of *Nos*Rubisco, *Nos*Rca and *Nos*Rubisco and *Nos*M35 (short isoform of CcmM) via LLPS was monitored by light scattering. The experiments were conducted at 25 °C in turbidity assay buffer (50 mM MMT pH 8.4 / 10 mM $MgCl_2$) in presence or absence of 5 mM DTT and at various concentrations of KCl as indicated. If not specified in the Figure legends, the standard concentration of KCl was 20 mM. All proteins were prepared as 10-fold concentrated stock solution and diluted in assay buffer to the following final concentrations: 0.25 µM *Nos*Rubisco (holoenzyme), 0.125/0.25/0.5/1.25 µM *Nos*Rca (hexamer), 1 µM *Nos*M35. Due to carry-over from protein stock solutions in storage buffer all assays included $\sim$ 1 % glycerol. Carry-over of salt was accounted for in the assay buffer. Upon addition of Rubisco to the assay buffer, the reaction was equilibrated at 25 °C for 5 min.

Rca or M35 were then added at indicated concentrations, mixed by vortexing, and rapidly transferred to a quartz cuvette (1.4 mL; pathlength 10 mm) and absorbance at 340 nm was recorded as a function of time on a Jasco V-560 spectrophotometer.

For direct comparison of kinetic parameters between M35 from *S. elongatus* PCC 7942 and *Nostoc* sp. PCC 7120, the determination of the protein concentration resulting in half maximal turbidity signal for oxidized and reduced *Nos*M35 was performed as previously described [Wang et al., 2019] in 50 mM Tris-HCl, pH 8.0 / 50 mM KCl / 10 mM $Mg(OAc)_2$. The protein concentration required to reach the half-maximal turbidity signal was estimated by fitting a logistic model ($f(x) = c + \frac{0-c}{1+exp(b*(x-a))}$).

3. Results

The main findings of this study are presented in Chapters **3.1** and **3.2**. Further findings regarding the production of recombinant proteins and key reagents are presented in the Appendix **A**.

3.1. Biochemical and structural characterization of Rca from *Nostoc* sp. PCC 7120

Based on the currently available set of sequenced and fully assembled cyanobacterial genomes, approximately 50 % of β-cyanobacteria encode a protein with sequences highly similar to the structurally and functionally well described Rca from plants. The respective gene (*rca-like*) is not found in α-cyanobacteria and its occurrence within cyanobacterial taxonomic subgroups varies, although it is especially prominent in the taxon *Nostocales* ($> 90\,\%$) (Figure **3.1**). As the subject of this study we therefore chose the model organism *Nostoc* sp. PCC 7120. In *Synechococcales* and *Oscillatoriophycideae* $\sim 25\,\%$ and $35\,\%$ of the corresponding genomes comprise *rca-like*. Based on all currently available nucleotide sequences in NCBI assigned to *Gleobacteria* and *Pleurocapsales*, its absent from bacteria of these taxons (Figure **3.1**).

Based on sequence homology, the product of *rca-like* (Rca-like) comprises an N-terminal AAA+ domain (pfam00004) and a C-terminal domain similar to the Rubisco small subunit (cd03527) and the small subunit like (SSUL) modules of CcmM, which is essential for carboxysome formation and interacts with Rubisco via the SSUL domain, connected by a variable linker (Figure **3.2**). On average, the AAA+ domains of cyanobacterial Rca-like share $\sim 79\,\%$ sequence similarity and $61\,\%$ sequence identity with plant Rca. Specifically, the AAA+ domains of the functionally well-characterized Rca proteins from *Nicotiana tabacum* (*Nt*Rca) and *Arabidopsis thaliana* (*At*Rca) share $\sim 79.4\,\%$ and $79.7\,\%$ sequence similarity and $60.9\,\%$ and $62.1\,\%$ sequence identity with the AAA+ domain of Rca-like from *Nostoc* sp. PCC 7120.

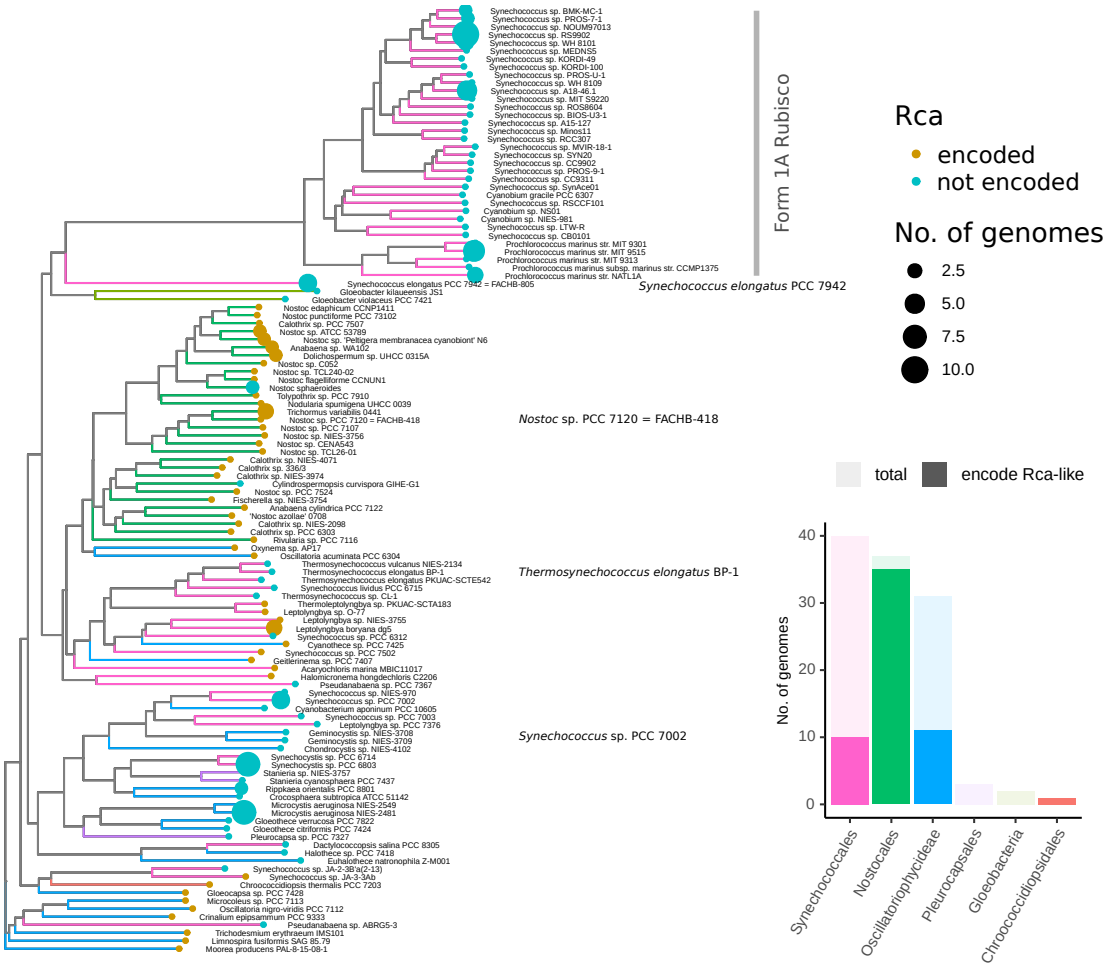

Figure 3.1.: Approximately 50 % of β-cyanobacterial genomes encode *rca-like*. Unrooted phylogenetic tree of 118 unique RbcL sequences derived from 172 sequenced and fully assembled cyanobacterial genomes. α-cyanobacterial Form 1A Rubisco forms a cluster separate from β-cyanobacterial Form 1B Rubisco. The number of genomes represented by individual RbcL sequences is mapped to the size of the tip points of the tree. The colors of the tree branches represent taxonomic subgroups (see inset) and the colors of the tip points refer to the presence of Rca in the genome represented by individual RbcL sequence. Important model species are highlighted. **Bar-chart inset:** Distribution of *rca-like* in β-cyanobacterial taxonomic subgroups. A separate query against all nucleotide sequences of *Gleobacteria* or *Pleurocapsales*, respectively, did not reveal any occurrence of *rca-like* in sequences assigned to these bacterial groups.

3.1.1. *Nos*Rca is a Rubisco activase

To analyze the function of the AAA+ domain of Rca-like from *Nostoc* sp. PCC 7120, two constructs were generated and recombinantly produced in *E. coli*: the full-length protein, *Nos*Rca, and a construct lacking the linker and the SSUL domain (residues 2-291), *Nos*RcaΔC. Further details on the purification of *Nos*Rca and *Nos*RcaΔC including representative purifications of both proteins are provided in Chapter **A.1.1.1**. Consistent with the prediction as AAA+ protein, both purified *Nos*Rca and *Nos*RcaΔC hydrolyzed ATP, but the presence of the SSUL domain

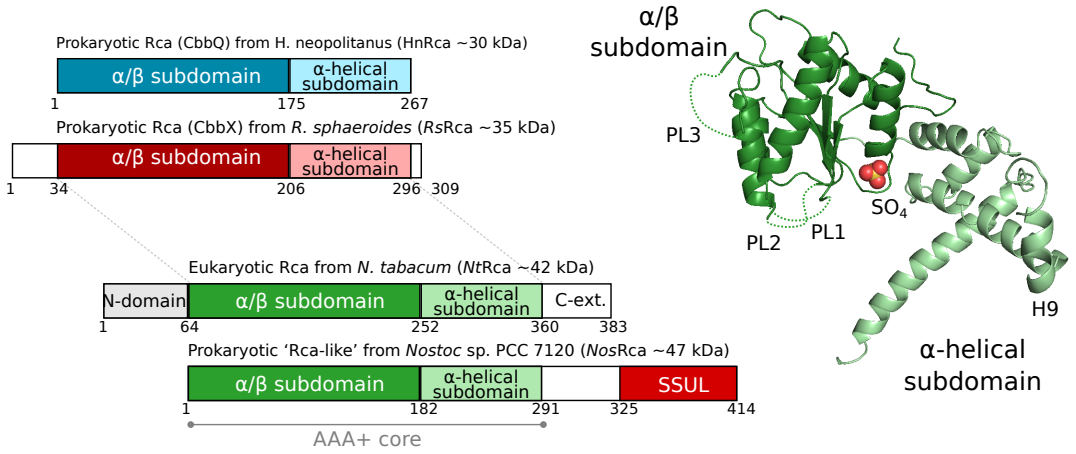

Figure 3.2.: The AAA+ domain of *Nos*Rca-like is highly similar to AAA+ domain of plant Rca.
Left: Schematic representation of the domain structures of Rcas across phylogenetic groups - CbbQ from the chemoautotroph *Halothiobacillus neapolitanus*, CbbX from *Rhodobacter sphaeroides* (prokaryotic, red-type), Rca from *Nicotiana tabacum* (eukaryotic, green-type) and cyanobacterial Rca-like (prokaryotic, green-type) from *Nostoc* sp. PCC 7120. In contrast to red-type Rcas, the AAA+ domains of cyanobacterial Rca-like and plant Rca are highly homologous. Notably, cyanobacterial Rca-like proteins feature a C-terminal extension (SSUL domain), that is not found in Rcas of other phylogenetic groups. **Right:** Ribbon representation of *At*Rca (PDB: 4W5W). α/β and α-helical subdomains are indicated in dark and light green, respectively. Pore loops (PL1, PL2 and PL3), helix H9, and N- and C-termini are indicated. A sulfate ion bound to the nucleotide binding pocket is represented as space-filling model.

exerted an inhibitory effect on ATP hydrolysis (Figure **3.3 A**). With ATP hydrolysis rates of $6.5\,\mathrm{min}^{-1}$ and $9.5\,\mathrm{min}^{-1}$ for *Nos*Rca and *Nos*RcaΔC, respectively, ATP turnover was considerably lower than reported for plant Rca, which varies between approx. 15 - $60\,\mathrm{min}^{-1}$ depending on the plant species and isoform [Hazra et al., 2015]. Although plant Rca functions as a hexamer, it is frequently reported to populate a range of oligomeric states in solution. Notably, when we analyzed the oligomeric state of *Nos*Rca by SEC-MALS, we found that the protein formed stable hexamers over a wide concentration range from 0.1 - $2\,\mathrm{g\,L}^{-1}$ (2 - 40 µM *Nos*Rca protomer) without addition of nucleotides (Figure **3.3 C** and Table **3.1**).

To test whether *Nos*Rca functions as a Rubisco activase, we recombinantly expressed Rubisco from *Nostoc* sp. PCC 7120 (*Nos*Rubisco) by coexpression of the endogenous cyanobacterial chaperonins. Further details on the expression and purification of *Nos*Rubisco including a representative purification are provided in Chapter **A.1.1.2.1** and **A.1.1.2.2**. The purified and activated Rubisco (ECM) catalyzed the carboxylation of RuBP at ∼ 4 molecules CO_2 per active site min^{-1} under non-substrate-saturating conditions. As reported for other cyanobacterial Rubiscos [Li et al., 1999], the purified and decarbamylated *Nos*Rubisco (E) was not inhibited by

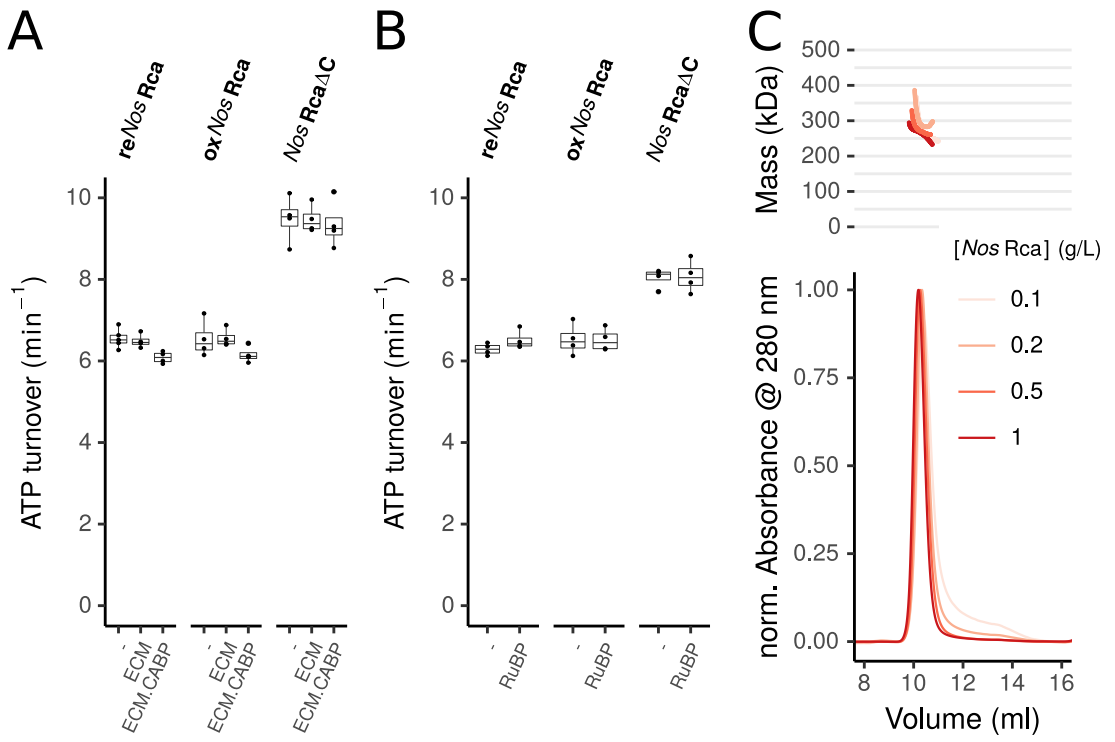

Figure 3.3.: ATPase rate and oligomeric state of *Nos*Rca.
ATPase activities of *Nos*Rca and *Nos*RcaΔC are not stimulated in presence of Rubisco (A) or RuBP (B). ATP hydrolysis was measured at 0.5 µM *Nos*Rca and 0.25 µM (holoenzyme) active Rubisco (ECM) or inactive Rubisco (ECM.CABP) (**A**) or in presence of 0.2 mM RuBP (**B**). To prevent scattering due to protein aggregation in samples containing *Nos*Rubisco, ATPase rates were measured at 200 mM KCl, whereas all other ATPase rates were determined at 20 mM KCl (see chapter **2.10.2** for details). Data from four independent experiments are presented in box and whisker plots (Tukey; lower and upper hinges correspond to 25th and 75th percentiles, respectively, and the whiskers extend from the hinges to the smallest and largest value, respectively, but maximally 1.5 x the inter-quartile range). Note, that the SSUL of *Nos*Rca contains redox-sensitive cysteins which can be oxidized (see chapter **3.2** and **2.5.1** for details). **C) *Nos*Rca forms stable hexamers across a wide concentration range.** The oligomeric state of *Nos*Rca was analyzed at indicated concentrations by SEC-MALS. 0.1 g L^{-1} / ~ 0.3 µM (hexamer): ~ 268 kDa; 0.2 g L^{-1} / ~ 0.7 µM: ~ 294 kDa; 0.5 g L^{-1} / ~ 1.8 µM: ~ 273 kDa; 1 g L^{-1} / ~ 3.6 µM: ~ 265 kDa. Data was obtained in collaboration with Monica Zobawa from the MPI Core facility.

RuBP (Figure **3.4 A**), in contrast to plant Rca [Wang and Portis, 1992]. XuBP competitively inhibits Rubisco from plants and *Rhodobacter sphaeroides* and forms spontaneously due to erroneous misprotonation of the enediolate intermediate of Rubisco catalysis. Thus, we tested the effect of XuBP on *Nos*Rubisco. When we incubated decarbamylated *Nos*Rubisco with increasing concentrations of XuBP (50 - 150 µM) in presence of low concentrations of RuBP (400 µM) we found that XuBP competed with the carboxylation of RuBP in a concentration dependent manner (Figure **3.4 B**). Notably, the addition of *Nos*Rca partially alleviated the XuBP competition in an ATP dependent manner (Figure **3.4 C**), suggesting that *Nos*Rca facilitated the removal of XuBP from Rubisco active sites. No complete relieve from

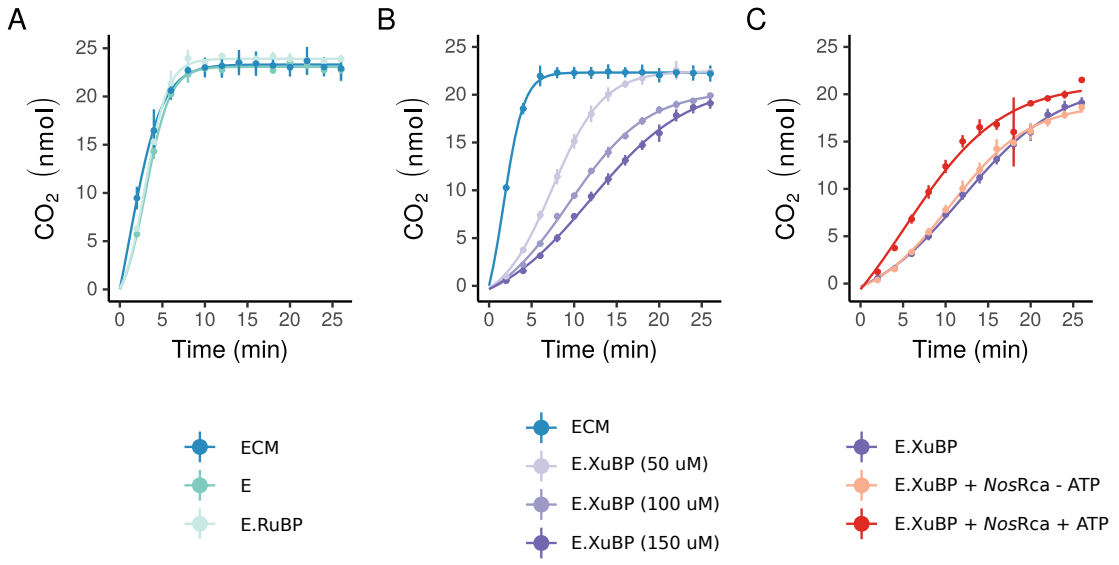

Figure 3.4.: Inhibition of *Nos*Rubisco activity.
Rubisco carboxylation and reactivation assays were conducted with 0.25 µM Rubisco (active sites) at 0.4 mM RuBP and CO_2 fixation was monitored over a time course of 26 min. Reactivation assays included 3 mM ATP or not, as indicated. Data represents mean and standard deviation (*sd*) of three independent experiments. For visualization, a four-parameter logistic model of the form $f(x) = c + \frac{d-c}{1+exp(b*(x-a))}$ was fit to the data. **A) RuBP does not inhibit uncarbamylated *Nos*Rubisco.** Catalytically competent, carbamylated Rubisco (ECM) was decarbamylated and subsequently inactivated by incubation in 0.5 mM RuBP (E.RuBP) or not (E). **B) XuBP acts as a competitive inhibitor at high concentrations.** XuBP inhibited Rubisco (E.XuBP) was formed by incubation of decarbamylated Rubisco in XuBP at 200, 400 or 600 fold molar excess over active sites for 1 h. The legend states the final XuBP concentration in the assay. **C) *Nos*Rca reactivates XuBP-inhibited *Nos*Rubisco in an ATP-dependent manner.** 0.5 µM *Nos*Rca was added to XuBP inhibited Rubisco (150 µM) in presence or absence of ATP. Note, that data **B)** and **C)** derives from the same set of experiments. For clarity, data for E.XuBP (150 µM) shown in **C)** has been reproduced from **B)**.

XuBP competition was observed, because XuBP was not degraded and rebound to activated active sites. The comparatively low affinity of XuBP toward cyanobacterial Rubisco may be the result of the suggested high flexibility of active sites in cyanobacterial Rubisco, presumably required for fast carboxylation rates [Savir et al., 2010; Tcherkez et al., 2006]. Consistently, we did not detect any gene homologous to the XuBP phosphatase Cbby [Bracher et al., 2015b] in *Nostocales* genomes, indicating that XuBP is likely no important inhibitor in these cyanobacteria. For complete inhibition of *Nos*Rubisco, we tested the effect of the tight binding inhibitor 2-carboxyarabinitol-1,5-diphosphate (CABP), the analogue of the short-lived carboxyketone intermediate of the carboxylation reaction, on *Nos*Rubisco. Already at concentrations equimolar to Rubisco active sites (0.25 µM), *Nos*Rubisco was completely and efficiently inhibited. Notably, both *Nos*Rca and *Nos*RcaΔC reactivated CABP-inhibited *Nos*Rubisco in presence of ATP (Figure **3.5 A**), demonstrating

that the SSUL domain is not required for Rubisco reactivation. In fact, reactivation by *Nos*RcaΔC was slightly more effective than by *Nos*Rca, indicating that the SSUL domain had a slight inhibitory effect on Rubisco reactivation, too. Moreover, Rubisco reactivation was dependent on Rca concentration, as demonstrated by addition of increasing concentrations of *Nos*RcaΔC (0.125 - 1.25 µM) to inhibited *Nos*Rubisco (Figure **3.5 B**). Thus, *Nos*Rca is a bona fide Rubisco activase, although the physiologically relevant inhibitor is currently not known.

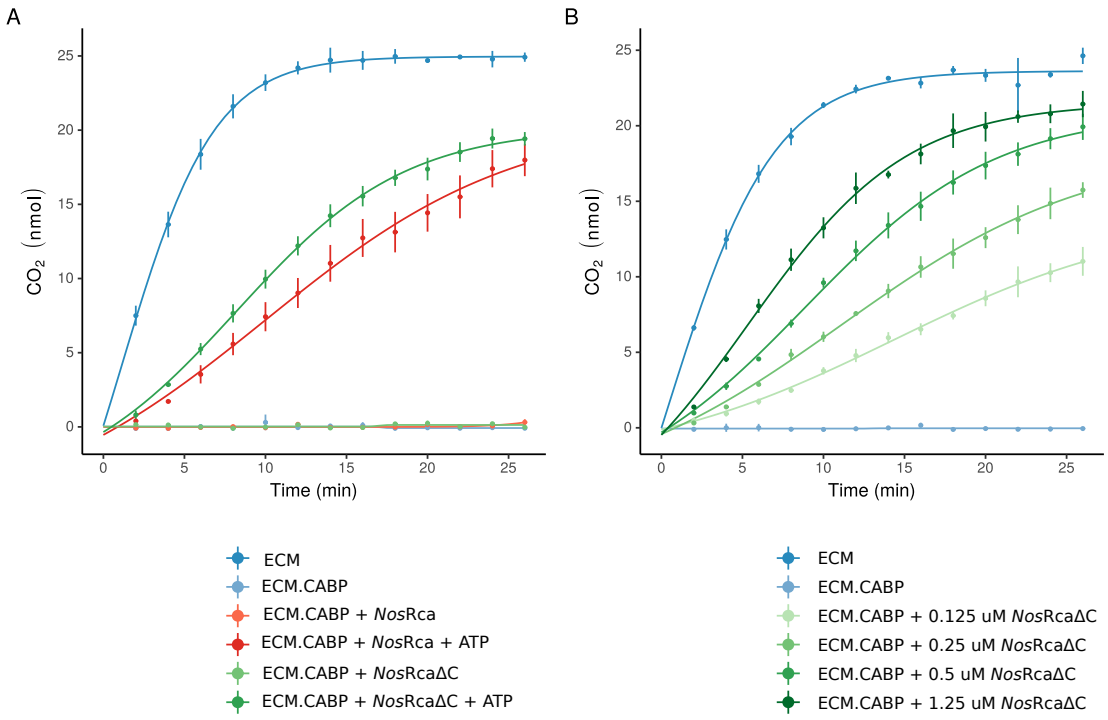

Figure 3.5.: The C-terminal SSUL domain of *Nos*Rca is not required for reactivation.
A) The C-terminal SSUL domain of *Nos*Rca is not required for reactivation. *Nos*Rubisco was inactivated by incubation of ECM in CABP at equimolar ratio (CABP:active site) for 30 min. Reactivation reactions were carried out at 0.25 µM *Nos*Rubisco (ECM or ECM.CABP) in presence of 0.4 mM RuBP and 0.5 µM *Nos*Rca or *Nos*RcaΔC (hexamer) in presence or absence of 3 mM ATP. ECM and ECM.CABP are shown as controls. **B) *Nos*RcaΔC reactivates CABP-inhibited *Nos*Rubisco in a concentration dependent manner.** Reactivation reactions using *Nos*RcaΔC were carried out as described in **A)** in presence of 3 mM ATP. Data represent the mean +/- *sd* of three independent experiments. For visualization a logistic model was fit to the data as in Figure **3.4**.

In contrast to the prokaryotic red-type Rubisco activase from *R. sphaeroides*, which forms ATPase active hexameric arrangements only in presence of ATP and RuBP and is strongly stimulated in presence of Rubisco [Bhat et al., 2017; Mueller-Cajar et al., 2011], *Nos*Rca is neither stimulated in presence of its cognate active or inactive substrate Rubisco nor in presence of RuBP (Figure **3.3 B**), similar to plant Rca [Robinson and Portis, 1989].

3.1.2. Crystal structure of *Nos*Rca∆C

For structural comparison of plant and cyanobacterial Rca we next sought to solve the crystal structure of *Nos*Rca. Protein constructs for crystallography purposes were designed and purified by Leonhard Popilka who also obtained the crystals in collaboration with the Crystallography Core Facility of the MPIB. Diffraction data were analyzed and a model was built by Andreas Bracher who further helped in the interpretation of crystallography data. Further information about crystallization, data analysis and model building is published in [Flecken et al., 2020].

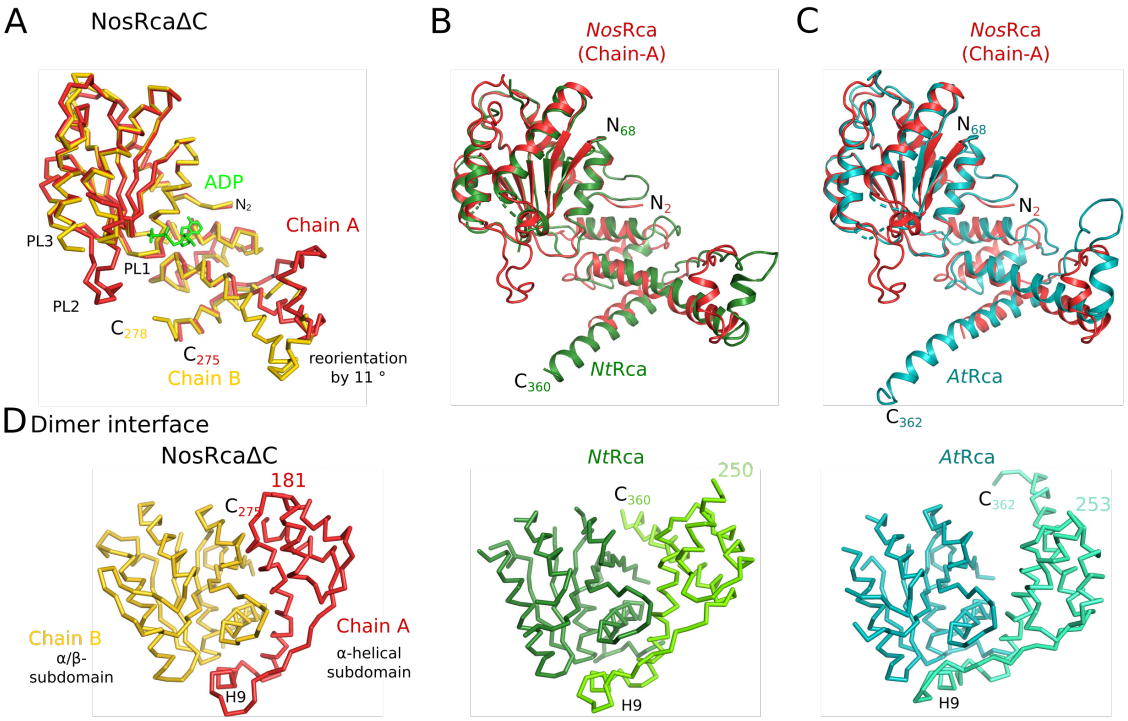

Figure 3.6.: Crystal structure of *Nos*Rca∆C.
A) Superposition of the two crystallographically independent chains of *Nos*Rca∆C within the asymmetric unit of the *Nos*Rca∆C crystal. Chain A and B are represented as Cα traces and colored in red and yellow, respectively. The ADP bound to chain B is displayed in a green wire-frame representation. Chain termini, pore loops and helix H9 are indicated. In chain B, pore-loop (PL) 2 (residues 105 - 115) and the connecting linker between helix α9 (H9) and α10 were disordered. Notably, the interdomain angle between the α/βsubdomain and the α-helical lid domain differed by approx. 11°. **Structural comparison of *Nos*Rca∆C with *Nt*Rca and *At*Rca.** Superposition of *Nos*Rca∆C chain A with *Nt*Rca (PDB: 3T15) [Stotz et al., 2011] **B)** or *At*Rca (PDB: 4W5W) [Hasse et al., 2015] **C)** are shown. Protein chains of *Nos*Rca, *Nt*Rca and *At*Rca are represented as ribbons in red, green and teal, respectively. **D) The subunit-subunit interaction in *Nos*Rca (left), *Nt*Rca (middle) and *At*Rca (right) is conserved.** The interface is formed between the α/β subdomain of one domain and the helices α8, H9 and the connecting linker to α10 in the α-helical subdomain of the adjacent subunit. Cα traces of the dimer interfaces are shown in yellow and red (*Nos*Rca), in dark and light green (*Nt*Rca) and teal and cyan (*At*Rca), respectively. H9 and chain termini are indicated. Data obtained in collaboration with Andreas Bracher and Leonhard Popilka. Figure modified from [Flecken et al., 2020].

While *Nos*Rca resisted any crystallization attempts, *Nos*RcaΔC crystallized readily using 2.2 or 2.3 mM Na-acetate pH 7 as precipitant by hanging-drop vapor diffusion. Diffraction data of the *Nos*RcaΔC crystals were recorded at beamline X06DA at the Swiss Synchrotron Light Source (SLS) in Villigen, Switzerland. The structure was solved by Gd-multi-wavelength anomalous diffraction of a Gadolinium ($GdCl_3$) derivative and the final native structure was solved at 2.45 Å by molecular replacement.

The asymmetric unit comprised two independent copies of *Nos*RcaΔC of distinct conformation. While the nucleotide binding pocket of one chain (Chain A) was accessible, the other was occupied by ADP (Chain B) (Figure **3.6 A** and **3.7**). Notably, no nucleotide was added during crystallization, thus the nucleotide bound during recombinant expression in *E. coli*. The entire backbone of chain A was ordered (residues 2 - 275), apart from the C-terminal 16 residues. In contrast, chain B comprised three unstructured regions - pore loop 2 (PL2) (residues 105 - 115), the connecting linker between helix α9 and α10 (residues 247 - 255) and the C-terminal 13 residues.

Like many typical AAA+ proteins, *Nos*RcaΔC consists of an N-terminal α/β subdomain (residues 7 - 177) and a small α-helical domain (residues 182 - 274) (Figure **3.2 A** and **3.6 A**). The α/β subdomain follows a Rossman fold typical for AAA+ proteins, which resembles a three-layered sandwich, where the outer β-sheet connecting α-helices, clasp around the central, parallel, five-stranded β-sheet in an β5-β1-β4-β3-β2 arrangement. The presence of a helical insertion (α2 in *Nos*RcaΔC) between the α2 and β2 of the typical AAA+ topology and a lacking Sensor-2 Arg at the beginning of α7 (α8 in *Nos*Rca) mark *Nos*Rca as a member of the Classic Clade of AAA+ proteins, like Rca from plants (e.g. *Nt*Rca: 3T15). The core of the α-helical subdomain is a 4-helix bundle which is formed by α6, α7, α8 and α10. Similar to Rca from plants (e.g. *Nt*Rca, 3T15; *At*Rca, 4W5W; and Rca from *Larrea tridentata*, 3THG), α8 is elongated and together with the connecting linker to α10 forms a paddle-like structure that detaches a short α-helical insertion α9 (H9) from the main body of the protein (Figure **3.6 A-C**). Note that H9 is absent in *L. tridentata*. In plants, H9 has been denoted as specificity helix, because it harbors residues that distinguish solanaceous from non-solanaceous Rubisco.

Superposition of the independent *Nos*RcaΔC chains revealed their nearly identical subdomain conformations (Figure **3.6 A**). Both subunits differed primarily in the interdomain angle between the α/β and the α-helical subdomain, whereby the ADP-bound state (Chain B) correlated with an inward rotation of the α-helical

domain by 11°. Interdomain reorientations in response to nucleotide binding are frequently observed within the AAA+ protein family (e.g. Fei et al. [2020]; Wang et al. [2001]).

Upon superposition of *Nos*RcaΔC chain A with the individual chains of *Nt*Rca (rmsd 1.46 Å for 206 matching Cα positions) and *At*Rca (rmsd 1.96 Å for 219 matching Cα positions), we found a remarkable structural similarity with the AAA+ cores of plant Rca (Figure **3.6 B and C**). Structural conservation became even more apparent, when we superposed the individual subdomains separately (0.53 Å and 0.52 Å for *Nt*Rca and *At*Rca, respectively). The dimer interface is formed between the α/β subdomain and helix H9 and the C-terminal part of α8 of two adjacent subunits. While the orientation of H9 to the α-helical bundle of the same subunit is flexible, the C-terminal part of α8 and H9 with respect to the α/β subdomain of the adjacent subunit appears to form a rigid module independent of the nucleotide state. Importantly, this rigid module is also found in crystal lattices of plant Rca and their orientation towards each other is conserved (Figure **3.6 D**).

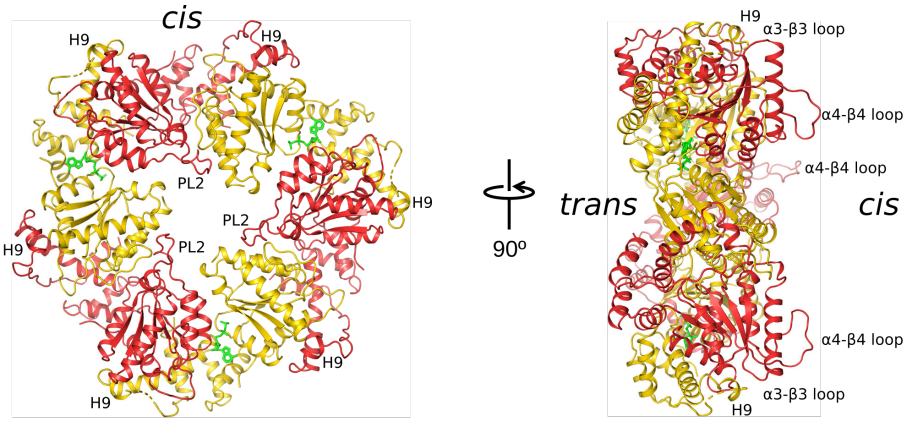

Figure 3.7.: *Nos*RcaΔC forms hexamers with alternating subunits in the rhombohedral crystal lattice.
Chain A and B of the asymetric unit are represented as red and yellow ribbons, respectively. ADP is shown in a wire-frame representation in green. The specificity helix α9 (H9), pore-loop PL2 and the α4-β4 loop are indicated. The *cis* surface is defined as the surface from which H9 and the α4-β4 loop protrude. In contrast, the *trans* surface comprises the proteins C-termini which fade into the linker to the SSUL domain in the full-length protein. Data obtained in collaboration with Andreas Bracher and Leonhard Popilka. Figure modified from Flecken et al. [2020].

In the rhombohedral crystal lattice the NosRcaΔC dimers form a hexamer with alternating subunit conformation (Figure **3.7**). The hexamer pore appears wide enough to accommodate a sphere of up to 12 Å. The *cis* surface is defined as the surface from which H9 and the α4-β4 loop protrude. In contrast, the *trans* surface comprises the proteins C-termini which fade into the linker to the SSUL domain in the full-length protein.

3.1.3. *Nos*Rca binds the N-terminus of RbcL in its central pore

To elucidate the reactivation mechanism, a cryo-EM single particle analysis of the *Nos*Rca-Rubisco complex was conducted in collaboration with Huping Wang and Andreas Bracher. Since reactivation does not require the SSUL domain, *Nos*RcaΔC was used, which facilitated sample preparation. To this end, a nucleotide substitution strategy was employed, where the CABP inhibited Rubisco was shortly incubated in ATP (10 s), followed by addition of ATPγS and incubation for 10 min prior to freeze plunging. This method is particularly useful to capture different nucleotide states and the specific conformations during conformational cycling as demonstrated (see Dong et al. [2019]).

However, using this technique we obtained only one predominant conformation of the *Nos*RcaΔC:Rubisco complex, the structure of which was solved at 2.86 Å (Figure **3.8 A**). The *Nos*RcaΔC density was however at much lower local resolution, requiring a local refinement strategy to resolve more details. The Rubisco subtracted density of *Nos*RcaΔC was finally resolved at 3.29 Å (Figure **3.13 A**). In our model, *Nos*RcaΔC faced Rubisco with the N-terminal α/β subdomain and the protruding α4-β4 loop (*cis* surface) pointing towards Rubisco (Figure **3.8 A**). The rim of the Rca hexamer wheel was approximately located above one Rubisco active site formed by an antiparallel RbcL dimer (N-terminal domain of RbcL-A and C-terminal domain of RbcL-B) at the edge of the cube-shaped Rubisco holoenzyme, while the central axis of the *Nos*RcaΔC hexamer pointed approximately to the cleft between two RbcL-dimers. The engaged catalytic site of Rubisco was lacking discernible density for the CABP inhibitor, whereas all other catalytic sites showed well-resolved density for the inhibitor. Hence, the captured interaction between *Nos*RcaΔC and *Nos*Rubisco represents the end-stage of the reactivation reaction, prior to the dissociation of Rca from Rubisco.

Importantly, we found a well-resolved density within the *Nos*RcaΔC hexamer pore, fitting to a 13 amino acid peptide in extended conformation (Figure **3.8 B**). From C-terminal truncation studies of Rubisco from *Rhodobacter sphaeroides*, we know that the extended C-terminal peptide is required for reactivation in red-type proteobacteria [Mueller-Cajar et al., 2011], suggesting a similar mechanism for green-type reactivation, despite of an 11 residues shorter C-terminus of *Nos*Rubisco. However, the C-terminus of *Nos*Rubisco RbcL-B was resolved until Glu471 and the remaining six unresolved residues (residues 472-477) would not have been able to cover the distance to the entrance of the Rca pore. In contrast, the 13 N-terminal residues Ser2 - Tyr14 of subunit RbcL-A fit the density well. Residues 15-23 of RbcL-A were unresolved,

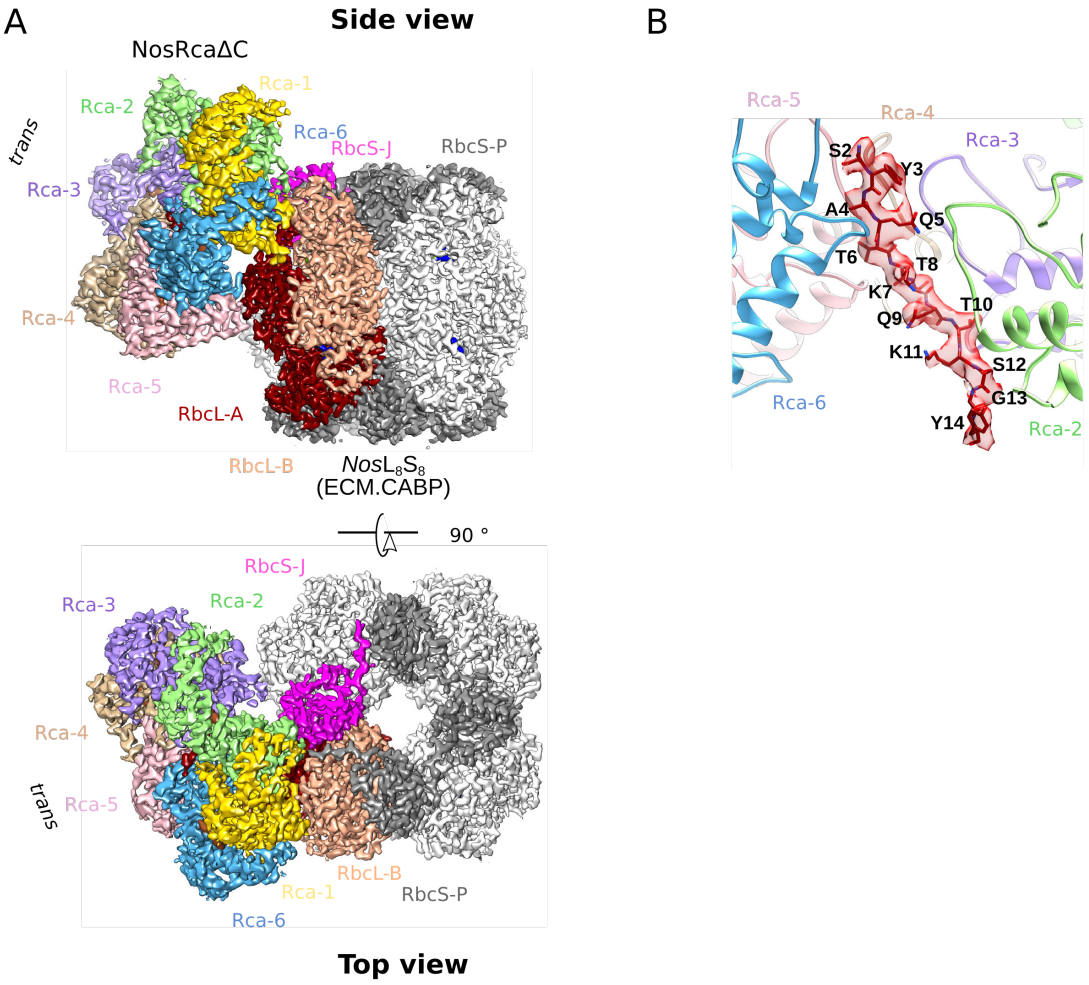

Figure 3.8.: Cryo-EM reconstruction of the *Nos*RcaΔC:Rubisco complex. **A) The hexameric *Nos*Rca faces Rubisco with the *cis* surface during reactivation.** The *Nos*RcaΔC:Rubisco complex was solved by a nucleotide substitution strategy at 2.86 Å. Top and side view of the complex are shown related to the two-fold and four-fold symmetry axis of Rubisco, respectively. NosRca subunits (Rca1-6) and the contacting Rubisco subunits (antiparallel RbcL-A / RbcL-B dimer as well as RbcS-J and RbcS-P) are indicated. **B) The N-terminal peptide of RbcL-A fits the additional density in the Rca central pore.** Engaged Rca subunits are indicated and displayed as ribbons. Amino acid residues of 2 to 14 of the bound peptide are indicated in single letter code. Data obtained in collaboration with Huping Wang and Andreas Bracher. Figure modified from Flecken et al. [2020].

indicative of a structurally flexible connection to the next resolved residue Tyr24. In all other RbcL subunits, the last resolved residue was Gly13. Interestingly, the comparison of *Nos*Rubisco N-terminal sequences with other cyanobacterial, green algae and plant N-terminal Rubisco sequences revealed a conserved pattern of alternating small and bulky side chains (Figure **3.14 A**).

To biochemically test the involvement of the RbcL N-terminus in reactivation of *Nos*Rubisco, we generated N- and C-terminally truncated *Nos*Rubisco mutants, with RbcL lacking either the first twelve (*Nos*L_8S_8ΔN) or 23 N-terminal residues or the last two (*Nos*L_8S_8ΔC2), four (*Nos*L_8S_8ΔC4) or nine (*Nos*L_8S_8ΔC9) C-terminal

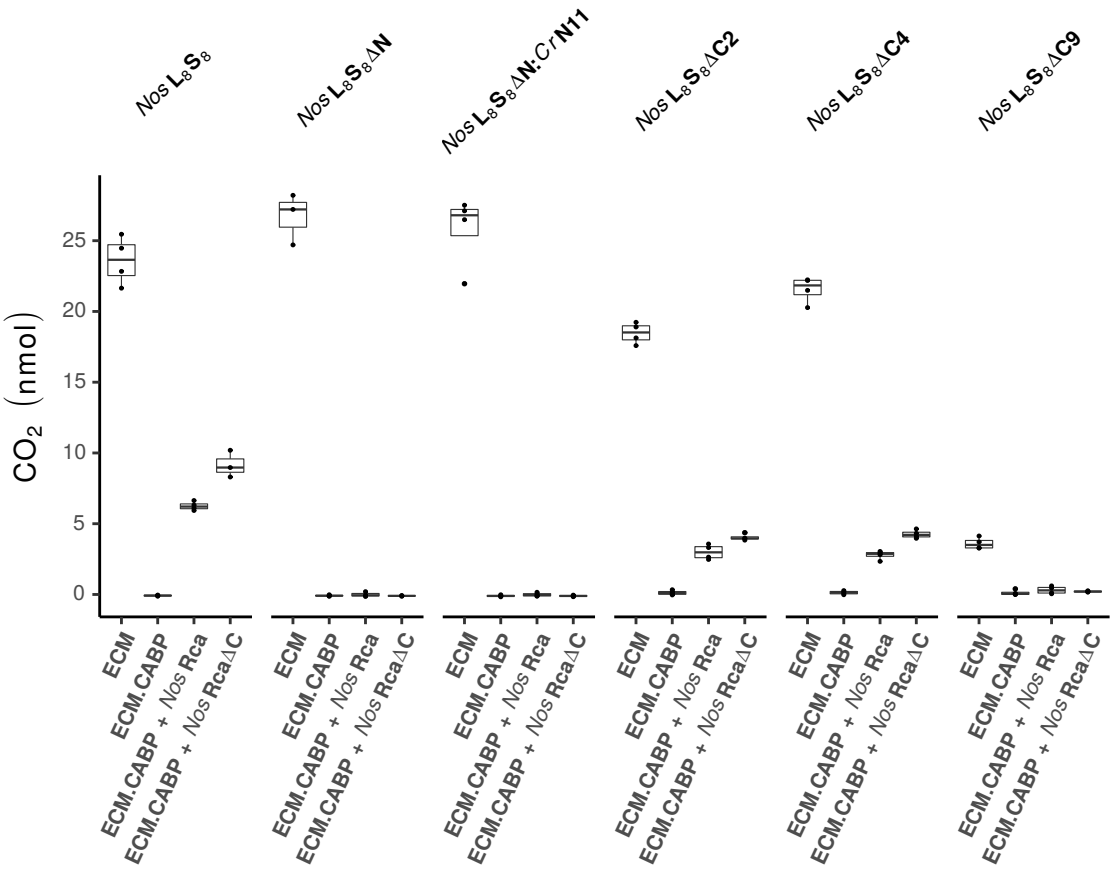

Figure 3.9.: *Nos*Rca requires the RbcL N-terminus for *Nos*Rubisco reactivation, but not its C-terminus.
*Nos*Rubisco or *Nos*Rubisco C- or N-terminal truncation mutants inhibited with 0.25 µM CABP ($NosL_8S_8\Delta N$, $NosL_8S_8\Delta N$:*Cr*N11, $NosL_8S_8\Delta C2$, $NosL_8S_8\Delta C4$ or $NosL_8S_8\Delta C9$) were co-incubated with *Nos*Rca. Net CO_2 fixation was stopped and recorded after 8 min. Corresponding ECM and ECM.CABP reactions for each Rubisco mutant are shown to distinguish reactivation failure from carboxylation impaired Rubisco mutants. To ensure steady state kinetics of the ECM controls, the RuBP concentration was increased to 1 mM. Box and whisker plots (Tukey) represent data from four independent replicates.

residues. Additionally, a *Nos*Rubisco chimeric protein was generated where we exchanged the first 12 N-terminal *Nos*RbcL residues with the first 11 N-terminal residues from *Chlamydomonas reinhardtii* ($Nos L_8S_8\Delta N$:*Cr*N11) to assess the importance of N-terminal sequence identity. While the N-terminal truncation of 23 residues rendered Rubisco carboxylation-inactive (data not shown), $NosL_8S_8\Delta N$ displayed wild type carboxylation activity, but could no longer be reactivated by *Nos*Rca (Figure **3.9**). Although the carboxylation activity of both $NosL_8S_8\Delta C2$ and $NosL_8S_8\Delta C4$ were impaired, they could still be reactivated upon CABP-inhibition. Note, that the reduced CO_2 fixation in these reactivation reactions were approximately proportional to the reduction in Rubisco carboxylation activity. CO_2 fixation of $NosL_8S_8\Delta C9$ was reduced by 80 % which hampered the analysis of reactivation

by *Nos*Rca (Figure **3.9**). Hence, reactivation solely depends on the RbcL N-terminal peptide with no hint for contributions of the C-terminus to the reactivation reaction. Like *Nos*$L_8S_8\Delta$N, the chimeric mutant *Nos*$L_8S_8\Delta$N:*Cr*N11 displayed approximately wild type carboxylation activity but could no longer be reactivated, suggesting some degree of sequence specificity beyond a common pattern of alternating small and bulky N-terminal RbcL sidechains. Of note, addition of the N-terminal peptide in *trans* to *Nos*Rubisco reactivation assays with *Nos*RcaΔC (at 1-, 10- and 100- fold excess over Rca pores) did not compete reactivation (data not shown), suggesting that additional cues, perhaps from the Rubisco surface, might be required for peptide binding in the *Nos*Rca pore.

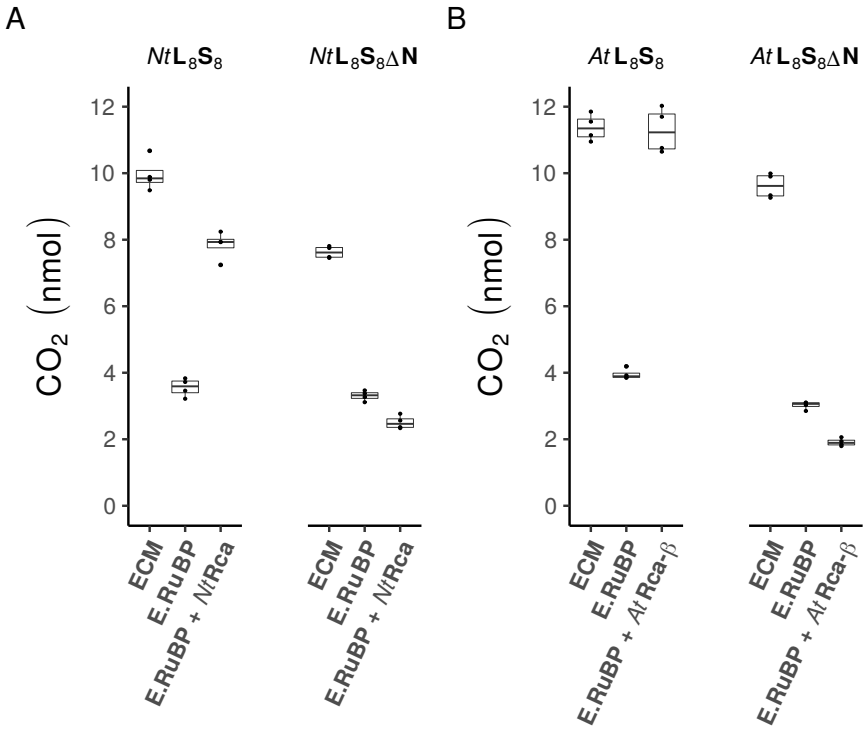

Figure 3.10.: Form IB Rubisco reactivation is conserved in green type organisms.
Wild type or N-terminal truncated plant Rubisco from *Nicotiana tabacum* (**A**) or *Arabidopsis thaliana* (**B**) was inhibited by incubation of the decarbamylated enzyme in 0.5 mM RuBP. Reactivation reactions were carried out in presence of *Nt*Rca or *At*Rca-β, respectively. The assays were carried out as described in Figure **3.9** with the concentration of Rubisco fixed at 0.25 µM active site. Note that wild type Rubisco was of endogenous source, whereas the N-terminal truncated enzymes were recombinantly produced in *E. coli*. Box and whisker plots (Tukey) represent data from four independent replicates.

To assess whether the requirement of the RbcL N-terminal peptide for reactivation is conserved in plants, we introduced the analogous N-terminal truncations to Rubisco from Tobacco (*Nt*$L_8S_8\Delta$N) and Arabidopsis (*At*$L_8S_8\Delta$N). Both *Nt*$L_8S_8\Delta$N and *At*$L_8S_8\Delta$N displayed carboxylation activity similar to their respective wild

type counterparts purified from endogenous sources. However, upon partial inhibition of the decarbamylated Rubiscos by RuBP (E.RuBPΔN), the corresponding Rcas, *Nt*Rca$_{R294V}$ and *At*Rca-β, respectively, were no longer able to alleviate the inhibition, in contrast to wild type Rubisco (E.RuBP) (Figure **3.10**). Note, that *Nt*Rca$_{R294V}$ shows wild type ATPase- and reactivation activity, while it forms stable hexamers in solution in presence of ATP [Stotz et al., 2011].

3.1.4. Architecture of the substrate-bound *Nos*RcaΔC

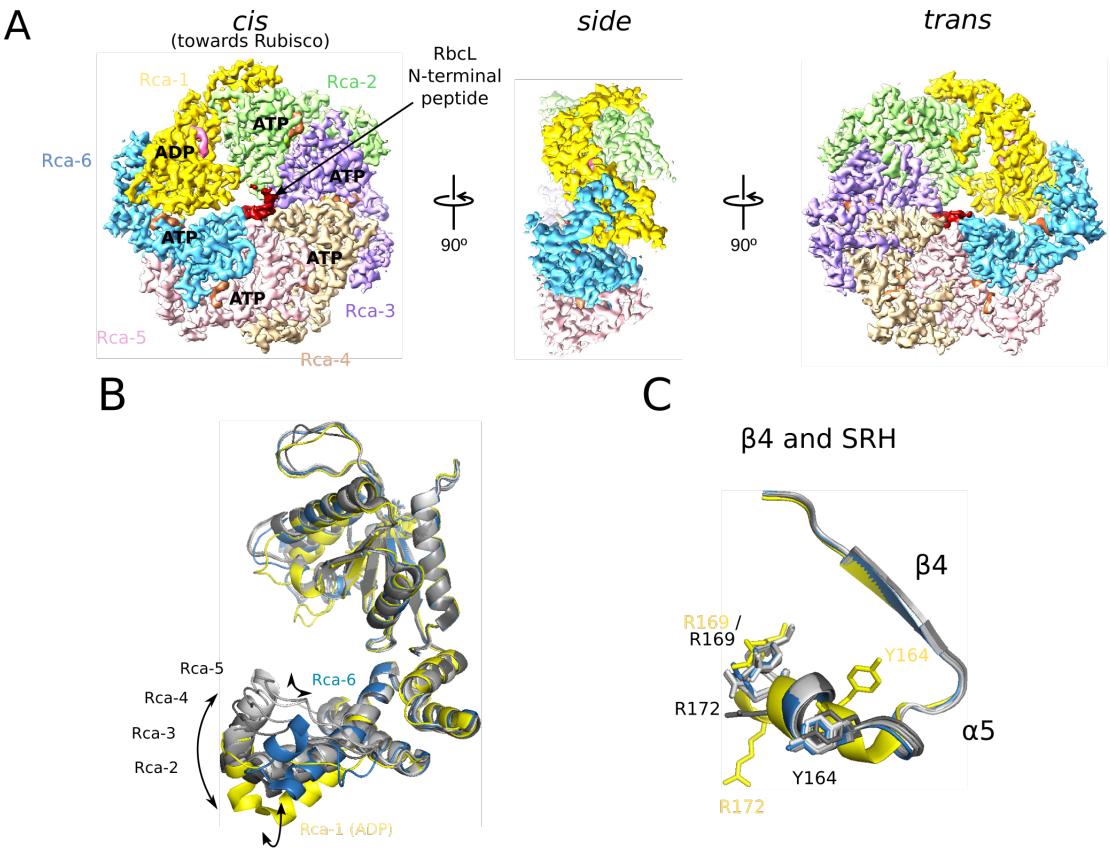

Figure 3.11.: Architecture of the substrate-bound *Nos*RcaΔC.
A) Cryo-EM density of *Nos*RcaΔC. Rca subunits and bound nucleotides are indicated. Figure adapted from [Flecken et al., 2020]. **B and C) Alignment of individual subunits of the *Nos*RcaΔC complex reveals movement of H9 with respect to the α-helical subdomain and highlights distinct positions of the arginine fingers in the ADP-bound state.** The full-length protein chain (**B**) and β4 followed by the SRH (**C**) are represented in ribbon representation. The arginine fingers Arg169 and Arg172 as well as Tyr164 are displayed in stick representation. To underline the gradual translation of H9, subunits Rca-2 (dark) to Rca-5 (light) are colored in grey shades. The color scheme of Rca-6 and Rca-5 corresponds to **A**.

The substrate bound *Nos*RcaΔC was in a right-handed helical „split-washer“ conformation (Figure **3.11**), as previously shown for multiple other AAA+ proteins [Han et al., 2017; Puchades et al., 2020; Sandate et al., 2019; Zehr et al., 2017]. The Rca

subunits 1 to 6 were specified with Rca-6 elongating the hexameric helix in *trans* direction and Rca-1 elongating the helix in *cis* direction (towards Rubisco). All nucleotide binding sites showed clear density of bound nucleotide (Figure **3.12**). The binding site formed by the adjacent subunits Rca-1 and Rca-2 accommodated ADP, whereas all other binding sites accommodated either ATP or ATPγS (Figure **3.11** and **3.12**). The conserved, *trans*-acting arginine residues Arg169 and Arg172, important for ATP hydrolysis and sensing of resultant conformational change (also called „arginine fingers“), were in close distance to the bound nucleotides at the interface toward the adjacent subunit. The EM density of these arginine fingers was generally well defined in all subunits, except for Rca-1, which correlates to its position at the split end of the helical arrangement of *Nos*RcaΔC (Figure **3.12**). The lower resolution of the EM density of these two arginine residues in Rca-1 might hint towards an ongoing ATP-hydrolysis event and resultant heterogeneous conformations.

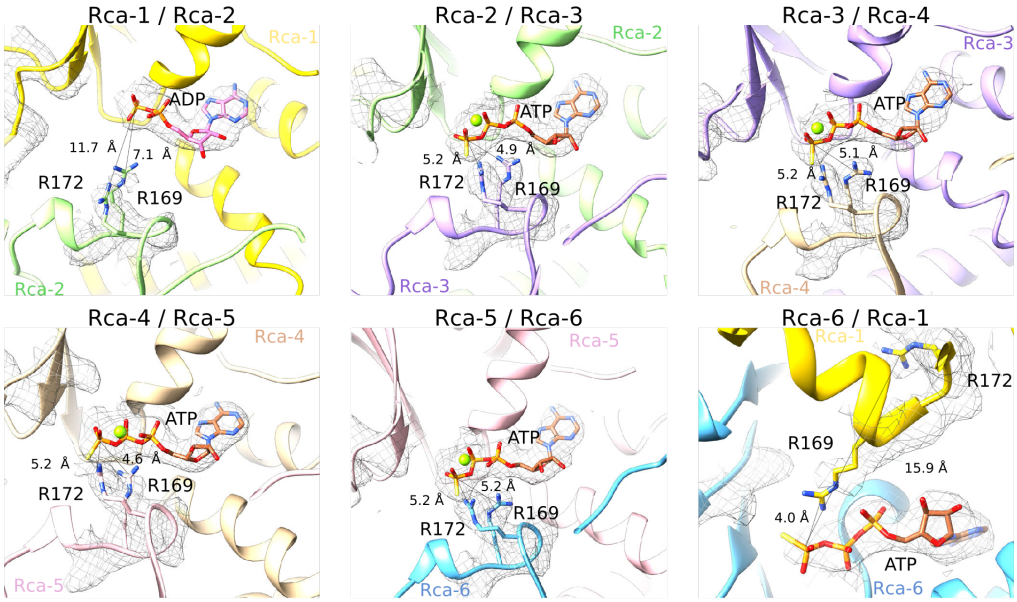

Figure 3.12.: Conformation of the „arginine fingers“ (Arg169/Arg172) with respect to the bound nucleotides in the nucleotide binding sites of the *Nos*RcaΔC spiral.
Nucleotide binding sites of the Rca spiral at subunit/subunit interfaces are shown. Indicated, color-coded Rca subunits are represented as ribbons. The EM densities for the bound nucleotide and Arg169/Arg172 are represented as meshwork and their distance (measured from the carbon atom in the guanidino group) to the last phosphorus atom of the bound nucleotide is indicated. Nucleotide in stick representation and the Mg^{2+}-ion shown as green sphere. Figures adapted from [Flecken et al., 2020].

The positioning of the arginine fingers is likely guided by the proposed interaction between Tyr164 and Asp104 of the adjacent subunit. Notably, the carboxyl groups of Asp104 are not resolved in any subunit, whereas Tyr164 is overall well resolved

and is retracted from the Rca-6:Rca-1 interface (Figure **3.11 C**). Alignment of the individual Rca subunits of the *Nos*RcaΔC:Rubisco complex indicated, that major movements within *Nos*RcaΔC are restricted to changes of the orientation of the C-terminal part of helix α8, H9, and the linker to helix α10 with respect to the small α-helical subdomain (Figure **3.11 B**). This angle is widest in the ADP-bound subunit (Rca-1) and progressively decreases in the ATP-bound subunits (position of Rca-2 to Rca-5). Notably, in the subunit anti-clockwise adjacent to the ADP-bound subunit (Rca-6), H9 and the C-terminal end of helix α8 are translated outward, following the movement of the α/β-subdomain of the ADP-bound subunit to which it is attached. Apart from this flexible joint to the adjacent α/β-subdomain, the individual domains are nearly invariant, including major loops like the α4-β4 loop and the pore loops. Only in the ADP-bound subunit (Rca-1), both PL1 and PL2 adopt a conformation different from the substrate engaged pore loops (Rca-2 to Rca-6).

3.1.5. Pore-loop interactions with the N-terminus of RbcL

All pore loops PL1 (67-72), PL2 (105-119) and PL3 (160-165) within all six *Nos*RcaΔC subunits were well-resolved and adopted a staircase-like arrangement (Figure **3.13 A, B** and **C**). Like in all substrate-bound AAA+ proteins that have been studied so far, the nucleotide state defines the conformation and organization of the pore loops which can engage with substrate [Puchades et al., 2020]. Accordingly, the ADP bound Rca-1 was the only subunit disengaged from the substrate peptide, while all five ATP/ATPγS bound subunits contacted the RbcL N-terminal peptide via both PL1 and PL2 (Figure **3.13 A** and **B**).

AAA+ protein translocases but also *Rhodobacter sphaeroides* Rca (*Rs*Rca) feature a canonical PL1 aromatic residue, which intercalates with the substrate peptide backbone and allows the translocating peptide to be pulled in a largely sequence-independent manner. Besides, some AAA+ proteins feature a substrate intercalating residue in PL2, like the yeast quality control protease YME1 [Puchades et al., 2017]. Cyanobacterial and plant Rca lack the canonical PL1 aromatic residue, but contain a highly conserved aromatic residue in PL2 which is required for ATP hydrolysis and reactivation activity in *Nt*Rca (Tyr188, [Stotz et al., 2011]) and *Nos*Rca (Tyr116, see Figure **3.13 E** and **F**). Sequence and functional conservation of this residue suggested a common mechanism of substrate engagement similar to the function of aromatic residues in pore loops of other AAA+ proteins. To our surprise, we found that the PL2 Tyr116 in *Nos*RcaΔC neither intercalated with the substrate backbone

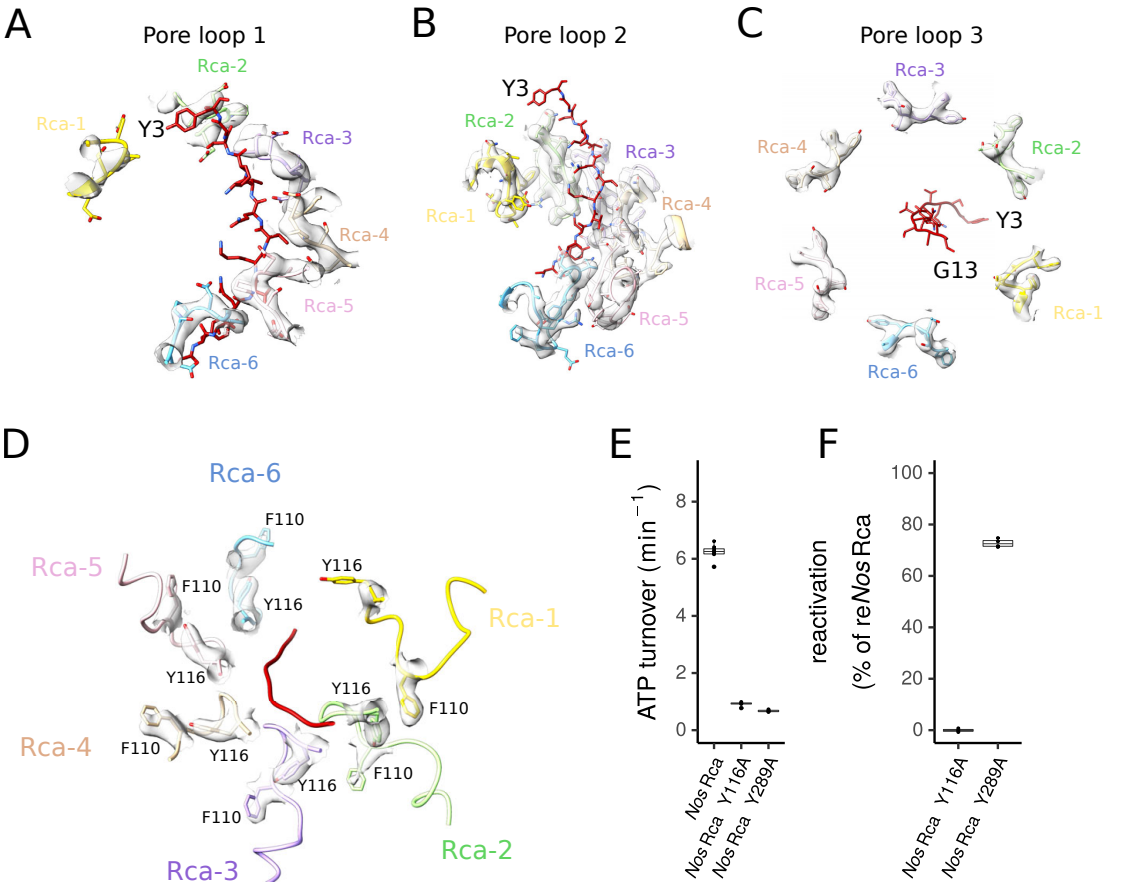

Figure 3.13.: Details of the engagement of the RbcL N-terminal peptide. Non-canonical PL1 (A) and PL2 (B) adopt a staircase-like arrangement and interact with the RbcL N-terminal peptide via van der Waals and hydrogen bonds, while PL3 (C) is disengaged from the substrate peptide. EM density of the pore loops is shown with a superposed atomic model in stick representation. The RbcL N-terminal peptide is shown in stick representation in red. *Nos*RcaΔC subunits are indicated. **D) The functional conserved aromatic residue in PL2 of cyanobacterial, algae and plant Rca does not intercalate with the substrate peptide.** Hydrogen bonded residues Tyr116 and Phe110 within the same subunit are indicated. The RbcL N-terminal peptide is represented as Cα-chain in red. **The conserved PL2 Tyr116 severely impacts ATP hydrolysis (E) and is crucial for reactivation (F).** ATP hydrolysis was measured in absence of Rubisco at 20 mM KCl and at 0.5 µM (hexamer) *Nos*Rca, *Nos*Rca Y116A or *Nos*Rca Y289A. CO_2 fixation assays were carried out for 8 min at 1 mM RuBP and reactivation of mutant Rca was normalized to reactivation by *Nos*Rca (100 %). Box and whisker plots (Tukey) represent data from four independent experiments. Figures modified from [Flecken et al., 2020].

nor did it directly interact with the peptide at all. Instead, Tyr116 formed hydrogen bonds with Phe110 of the same subunit and Gln121 in the subsequent subunit, likely resulting in a rigidification of the central pore (Figure **3.13 D**).

How is the N-terminal RbcL peptide stabilized in the central Rca pore? The RbcL N-terminal peptide follows a pattern of alternating small and bulkier side chains (Figure **3.14 A**). Interestingly, only the small side chains of the RbcL N-terminal peptide residues Ala4, Thr6, Thr8 and Thr10 point into successive binding pockets

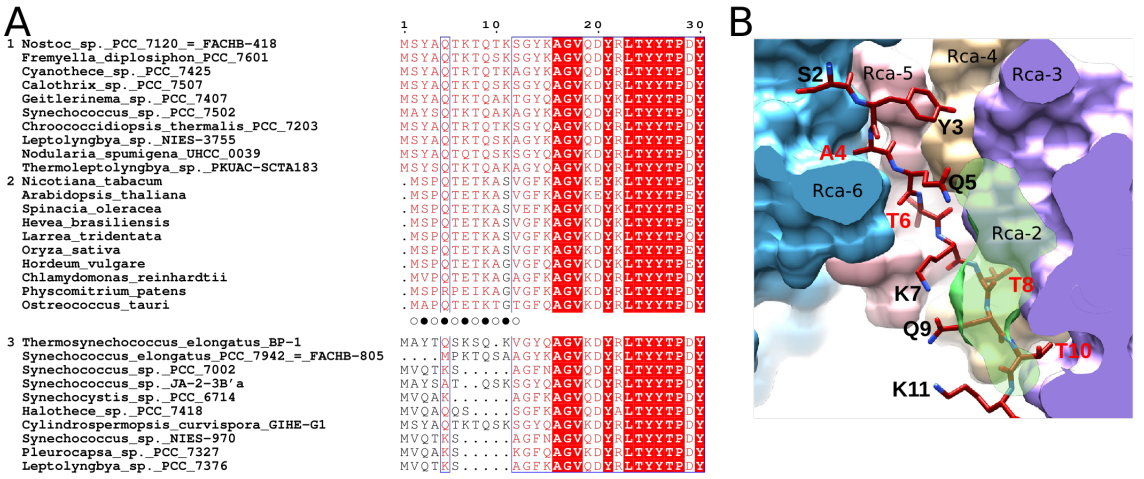

Figure 3.14.: A RbcL N-terminal pattern of alternating small and bulky residues is conserved in plants and *rca*-encoding β-cyanobacteria and likely important for binding to the Rca pore.
A) The RbcL N-terminal pattern of alternating small and bulky residues is conserved in *rca*-encoding β-cyanobacteria and plants, but not in β-cyanobacteria which do not encode *rca*. Selected RbcL sequences were aligned as described in Chapter **2.3** and the first 30 residues are displayed. Residues in white on red background highlight identical residues, while residues in red indicate similarity. Homologous regions are framed in blue. 1: *rca*-encoding β-cyanobacteria; 2: plants and algae; 3: not *rca*-encoding β-cyanobacteria; Open circles: small residue; solid circles: bulky residue. **B) The sidechains of alternating small residues of the RbcL N-terminal peptide point into cavities formed in the Rca pore.** Cavities are formed by pore loops PL1 and PL2 of adjacent subunits of Rca-2 to Rca-6. Bulkier residues face the solvent channel. The RbcL N-terminal peptide is shown in stick representation and the *Nos*Rca central pore in surface representation. Figure adapted from [Flecken et al., 2020].

formed between PL1 and PL2 within the central pore. In contrast, the bulkier side chains of Gln5, Lys7, Gln9, and Lys11 are facing the central solvent channel (Figure **3.14 B**). Notably, the pattern of alternating small and bulkier side chains does not extend beyond residue 12 and is generally conserved in RbcL from Form 1B Rubisco. Interactions between pore loops PL1 and PL2 and the bound substrate are predominantly mediated by van der Waals interactions. Additionally, the peptide backbone is in hydrogen-bond distance to the carbonyl group of Pro69 located in PL1 of Rca-3 to Rca-6. In contrast, PL3 makes no contacts to the substrate peptide (Figure **3.13 C**). However, its location between PL2 and the nucleotide binding site suggests a role in allosteric regulation, presumably mediated by the spatially adjacent Arg169 and Arg172. In total, the RbcL N-terminal peptide buries a substantial surface of 1018 $Å^2$ in the *Nos*RcaΔC pore.

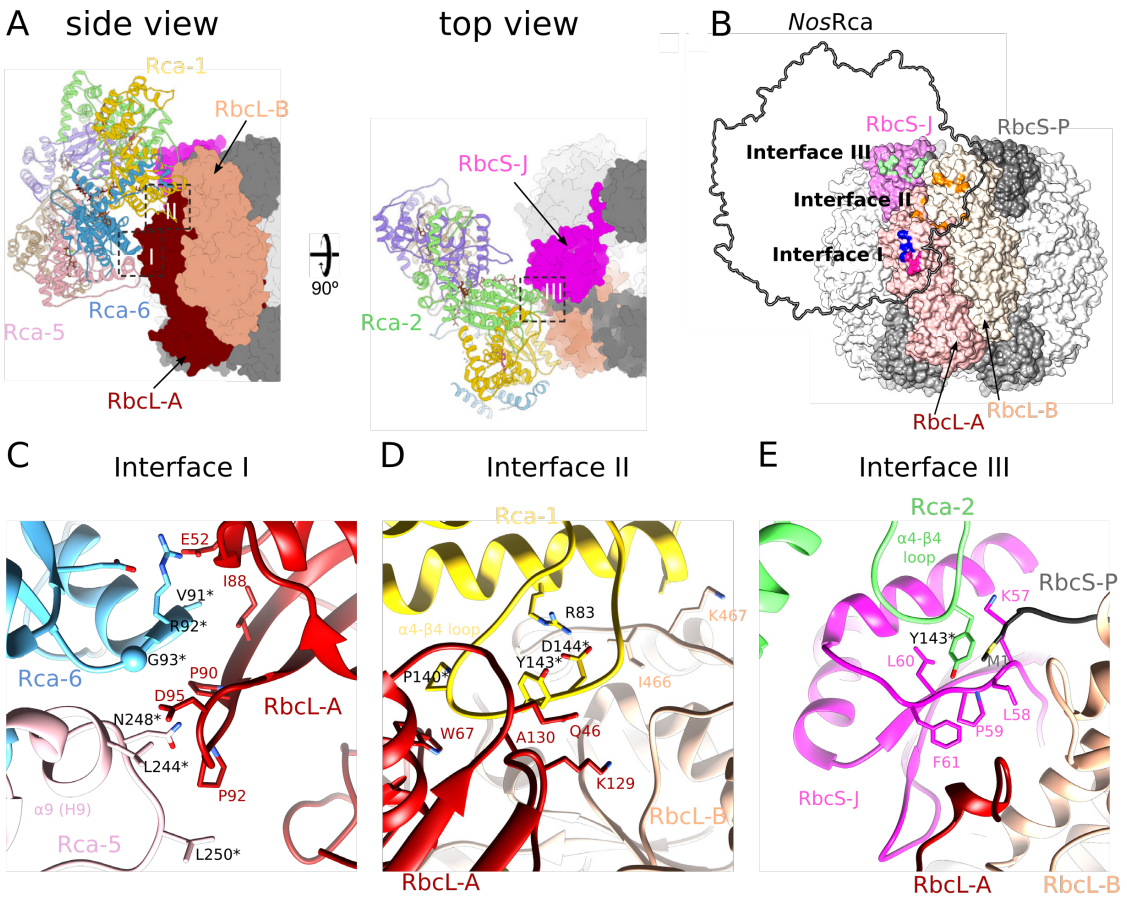

Figure 3.15.: The *Nos*RcaΔC spiral contacts the Rubisco surface at three interfaces.
A and B) Location of the three *Nos*Rubisco interfaces in contact with *Nos*RcaΔC. A) Side- and top view of the *Nos*RcaΔC:Rubisco complex with regard to Rubisco. *Nos*RcaΔC protein chains and Rubisco are represented as ribbons or in surface representation, respectively. **B)** *Nos*Rubisco is shown is surface representation. The position of the *Nos*RcaΔC spiral is outlined. Rubisco subunits in contact with Rca are indicated and interfaces are color-coded according to the engaged Rca subunit as in Figure **3.8-A**. **C) - D) Molecular details of the *Nos*RcaΔC:Rubisco interfaces.** Co lour-coded Rca subunits are shown in ribbon representation. At interface I, Rca-5 residues located in and close to H9 and helix α3 of Rca-6 interact with RbcL-A. At interface II, the $\alpha 4\beta 4$-loop of Rca-1 contacts Trp67 in loop 6 of RbcL-A. At interface III, the $\alpha 4\beta 4$-loop of Rca-2 interacts with residues on RbcS-J/P. Side chains of residues contacting Rubisco are shown in stick representation. * highlights mutated residues. Figures adapted from [Flecken et al., 2020].

3.1.6. Details of the *Nos*RcaΔC:Rubisco interface

Aside from the contacts established with the RbcL N-terminal peptide, *Nos*RcaΔC engaged with the substrate at three separate interfaces (Figure **3.15 A** and **B**), covering a total surface of 1445 Å^2. Interface I comprised interactions between Rca-5/Rca-6 with RbcL-A and covered a surface of 419 Å^2 (Figure **3.15 C**). This interface was mainly characterized by a high surface shape complementarity. Van der Waals contacts are established between H9 residues Leu244 and Asn248 on Rca-5 and Asp95 and Pro92 on RbcL-A, respectively. Pro92 is further engaged in van

der Waals contacts with Leu250 on Rca-5, and Val91 as well as Gly93 on Rca-6. Additional van der Waals contacts are formed between Rca-6 residue Val91 and Ile88 and Glu52. Moreover, a salt bridge is formed between helix α3 located, Rca-6 residue Arg92 and Glu52 on RbcL-A. The same Arg92 forms a intramolecular van der Waals bond with Glu57. Both interfaces II and III involve interactions with α4β4-loops of Rca-1/Rca-2. At interface II, which covers a surface of 600 Å^2, the Rca-1 α4β4-loop is inserted between the RbcL-A/RbcL-B dimer (Figure **3.15 D**). The intramolecular salt bridge between Arg83 and Asp144 seems to stabilize the conformation of the protruding loop. In addition, the presence of a proline (Pro92) adds conformational rigidity. The α4β4-loop contacts Trp67 of the substrate binding pocket within the 60s loop (residues 60-69) of RbcL-A. Van der Waals contacts are formed between Rca-1 residue Tyr143 and RbcL-A residues Gln46, Lys129 and Ala130. Of note, Rca-1 residue Val91 weakly interacts with the C-terminal residues Ile466 and Lys467 of RbcL-B, perhaps aiding inhibitor release by keeping the active site open. At interface III, covering 426 Å^2 the α4β4-loop engages in interactions with RbcS-J/P (Figure **3.15 E**). The Rca-2 residue Tyr143 forms van der Waals interactions with Lys57, Leu58, Pro59, Leu60 and Phe61 located at the β1α2-loop of RbcS-J and Met1 of RbcS-P. Interestingly, interface I residues Gly93 and Asn248 as well as interface II/III residue Tyr143 are conserved in algae and plant Rca (Figure **3.16**), strengthening the arguments for a conserved mechanism of Form 1B Rubisco remodeling.

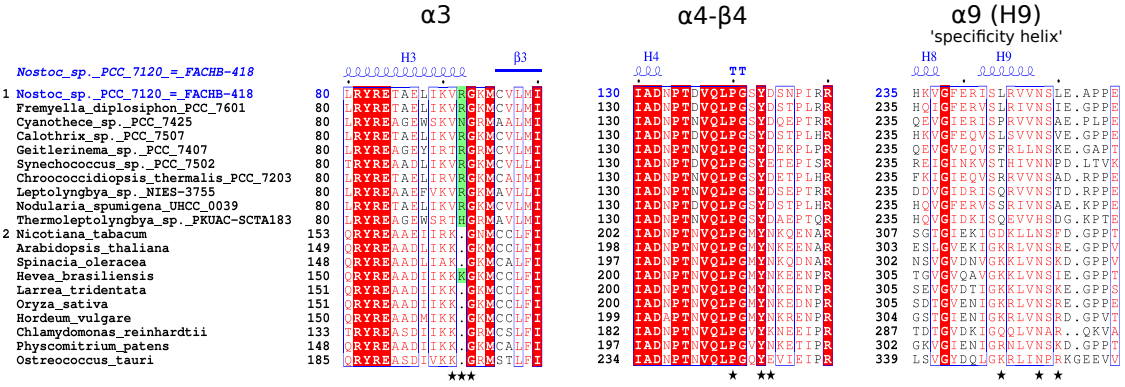

Figure 3.16.: Alignment of selected cyanobacterial and plant Rca sequences reveals conserved residues in helix α3, the α4-β4 loop and H9.
Selected Rca sequences were aligned as described in Chapter **2.3** and helix α3, the α4-β4 loop and H9 are displayed. Residues in white on red background highlight identical residues, while Residues in red indicate similarity. Homologous regions are framed in blue. The green background indicates systematic differences between cyanobacterial and plant Rca. Mutated residues important for the interaction of *Nos*Rca with Rubisco are indicated by black stars.

With the aim to assess contributions of the α4β4-loop, the helix a3 and H9 in *Nos*Rubisco reactivation individually, we next disturbed the *Nos*RcaΔC:*Nos*Rubisco

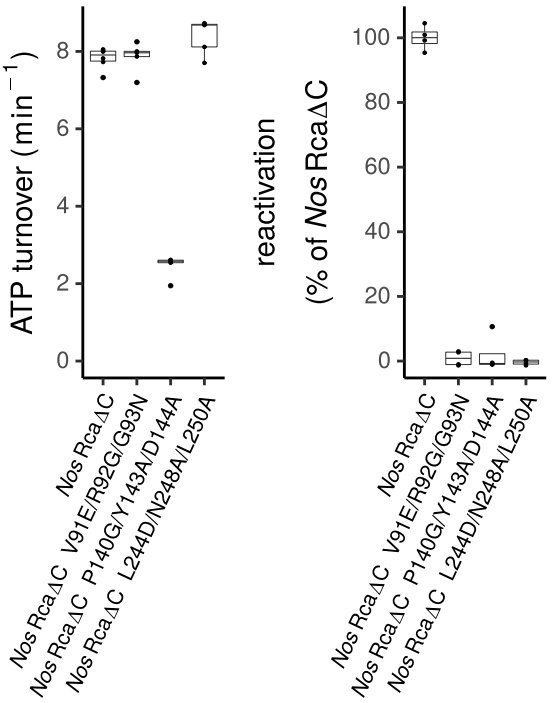

Figure 3.17.: Mutations in the Rca:Rubisco interface disrupt the reactivation interaction.
ATPase rates (**Left**) and reactivation capabilities (**Right**) of *Nos*RcaΔC and Rubisco:Rca interface mutants. ATP hydrolysis was measured in the absence of Rubisco at 20 mM KCl and at 0.5 µM (hexamer) *Nos*RcaΔC or any Rca:Rubisco interface mutant (*Nos*RcaΔC V91E/R92G/G93N, *Nos*RcaΔC P140G/Y143A/D144A, *Nos*RcaΔC L244D/N248A/L250A). CO_2 fixation assays were carried out for 8 min at 1 mM RuBP and reactivation capabilities are stated in relation to *Nos*RcaΔC (100 %). Box and whisker plots (Tukey) represent data from four independent experiments.

interfaces by mutagenesis. *Nos*RcaΔC V91E/R92G/G93N was generated to disrupt the intermolecular salt bridge to Glu52 on *Nos*RbcL as well as van der Waals bonds to Ile88 and P90. *Nos*RcaΔC L244D/N248A/L250A was generated to disrupt van der Waals interactions between H9 and the Asp95 and P92. In reactivation assays, both interface I mutants failed to reactivate inhibited *Nos*Rubisco, despite of wild type ATP hydrolysis rates (Figure **3.17**). The interface II/III mutant *Nos*RcaΔC P140G/Y143A/D144A, designed to make the $\alpha 4\beta 4$-loop more flexible and break its interactions with Rubisco, likewise showed no reactivation activity but was impaired in ATP hydrolysis (approx. 30 % of wild type ATP hydrolysis). Of note, another mutant Tyr289 (corresponds to Tyr361 in *Nt*Rca), although only lightly impaired in reactivation (> 70 % of wild type Rca reactivation) displayed a strongly reduced ATP hydrolysis rate (Figure **3.13 E** and **F**), indicating that ATP hydrolysis is likely not rate limiting for reactivation under specified conditions. Taken together, the $\alpha 4\beta 4$-loop, the helix $a3$ and H9 are required for *Nos*Rubisco reactivation. Importantly,

residues within the $\alpha 4\beta 4$-loop, the helices $a3$ and $a9$ (H9) of plant Rca have been implicated in Rubisco remodeling before [Li et al., 2005a; Ott et al., 2000; Portis et al., 2008; Shivhare et al., 2019; Shivhare and Mueller-Cajar, 2017].

3.1.7. Active site conformations and remodeling mechanism

As previously described, the Rca engaged Rubisco active site in the *Nos*RcaΔC:*Nos*Rubisco complex was lacking density for the inhibitory sugar phosphate, thus representing a post-remodeling state. In contrast, all other active sites displayed clear density for the bound inhibitor. Superposition of the Rubisco active site pocket in open and closed conformation allowed the analysis of structural rearrangements and revealed insights into the remodeling mechanism. Several structural changes between the open and closed state were recognized (see Figure **3.18**):

In the closed state, RbcL N-terminal residues were well resolved until Gly13, suggesting that the peptide running across the active site pocket was stably attached to the Rubisco surface. Furthermore, the RbcL C-terminus of the adjacent subunit of the antiparallel dimer locked loop 6 in a position where it closed above the active site pocket, together forming the „multi-layered lid“ required for catalysis [Bracher et al., 2017]. In the open state, binding of the RbcL N-terminal peptide to the Rca central pore destabilized the subsequent RbcL N-terminal residues (15-23) which were disordered and likely affected the conformation of the Rubisco 60s loop that stabilized the active site in the closed state and was displaced in the open state. The RbcL C-terminal peptide running across the Rubisco catalytic site in the closed state was displaced and its interactions with loop 6 were disrupted. Both the 60s loop of RbcL-A and loop 6 of RbcL-B pointed to solvent in the open state. In summary, Rca mediated retraction of both the 60s loop and loop 6 opened the Rubisco active site pocket and facilitated inhibitor release.

3.2. Function of the SSUL domains and the AAA+ core in Rca:Rubisco condensates

In all cyanobacteria, Rca's substrate Rubisco is localized in carboxysomes. In order to fulfill its function, close proximity of Rca to Rubisco is required. Indeed, a recent study by Lechno-Yossef et al. [2020] demonstrated by *in vivo* confocal microscopy that Rca co-localizes with Rubisco in carboxysomes of *Fremyella diplosiphon* - a

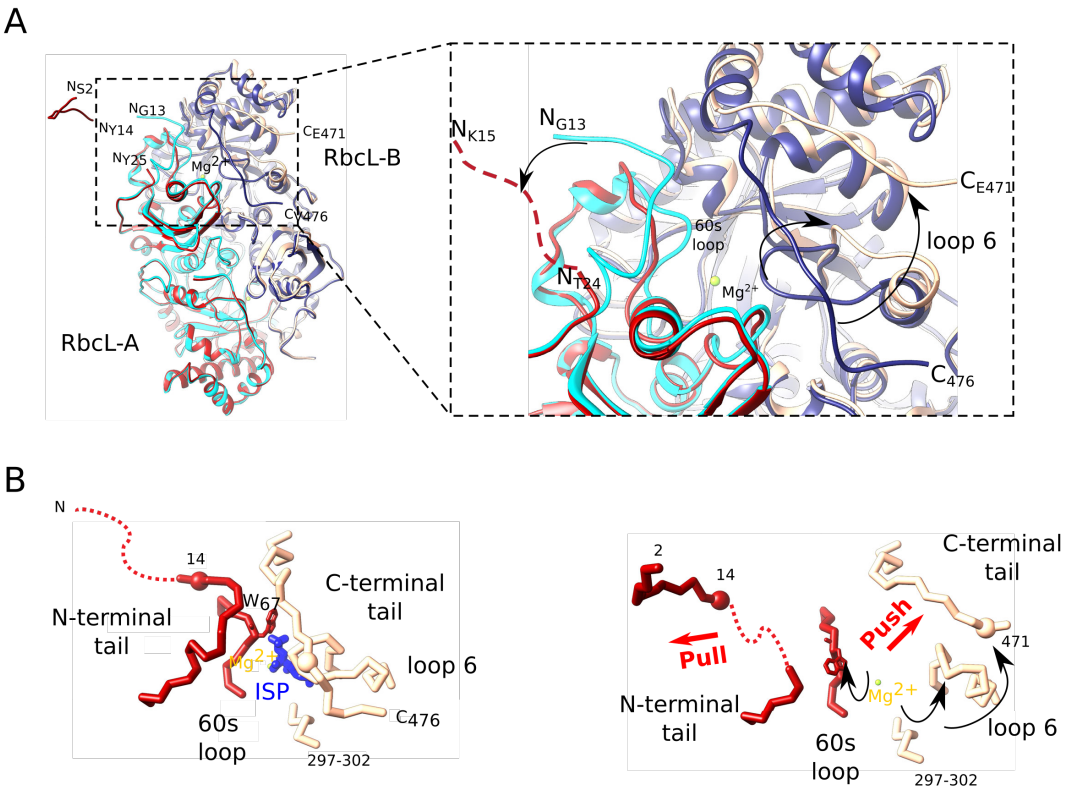

Figure 3.18.: Retraction of both the 60s loop and loop 6 opens the Rubisco active site pocket.
A) Superposition of Rubisco active sites in the closed and open state reveal mechanistic details of remodeling by Rca. The superposed atomic models are represented as ribbons. The Rca engaged antiparallel RbcL dimer formed by RbcL-A and RbcL-B is shown in red and peach, respectively. The closed state is represented in cyan (corresponding RbcL-A position) and blue (corresponding RbcL-B position). The arrows highlight movements necessary to open the active site pocket and remodeled elements (RbcL N-terminal peptide, 60s loop, RbcL C-terminus and loop 6) are indicated. **B) Schematic representation of the opening mechanism.** Focus is on remodeled elements around the Rubisco active site pocket, only. All elements are represented as Cα-chains, elements corresponding to RbcL-A and RbcL-B positions are shown in the closed state (left) and open state (right) in red and peach, respectively. The inhibitory sugar phosphate (ISP) is shown in blue. The bound Mg^{2+} ion, indicative of the carboxylation active state, is shown in yellow. Black arrows highlight movements necessary to open the active site, red arrows indicate anticipated „push and pull“ action mediated by Rca that are required for the opening of the Rubisco active site.

species closely related to *Nostoc* PCC 7120. But how is the co-encapsulation of Rca achieved during carboxysome biogenesis?

3.2.1. The SSUL of *Nos*Rca shares sequence and structural similarity with the SSUL domains of CcmM

A unique feature of cyanobacterial Rca is the presence of a C-terminal small subunit like domain (SSUL) whose name derives from distant sequence similarity to the small

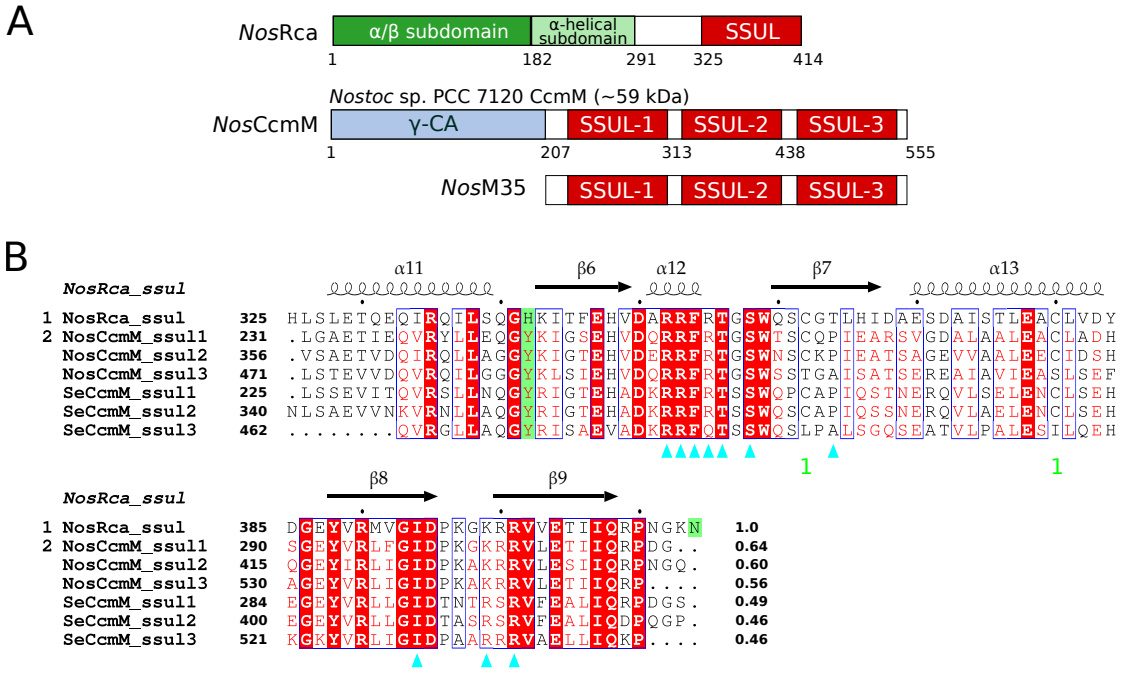

Figure 3.19.: The SSUL domain of *Nos*Rca shares approximately 60 % sequence identity with the SSUL domains of CcmM
A) Schematic representation of *Nos*CcmM in comparison to *Nos*Rca. Due to the presence of an internal ribosomal binding site in the *Nos*CcmM coding sequence, *Nos*CcmM is expressed as a long (*Nos*CcmM) and short isoform (*Nos*M35) - the latter lacking the N-terminal γ-carbonic anhydrase domain [Long et al., 2007]. **B) Sequence alignment of SSUL domains from *Nos*Rca (*Nos*SSUL) and *Nos*CcmM in comparison with the three SSUL domains from CcmM of *S. elongatus* PCC 7942 (*Se*CcmM).** Residues in white on red background highlight identical residues, while Residues in red indicate similarity. Homologous regions are framed in blue. The green background indicates systematic differences between *Nos*SSUL and SSULs of CcmM. Contact residues with Rubisco are indicated by triangles in cyan and green characters „1“ indicate the position of the disulfide bridge in the crystal structure of SSUL1 of *Se*CcmM. Sequence identities of each SSUL domain to *Nos*SSUL is given in the last column of the alignment. According to Fitch matrix comparison [Fitch, 1966], the SSUL of *Nos*Rca and the SSUL domains of *Nos*CcmM share ∼ 80 % sequence similarity. NCBI RefSeq accession codes: WP_010995702.1, *Nos*Rca; WP_010995039.1, *Nos*CcmM; WP_011378029.1, *Se*CcmM.

subunit of Rubisco (RbcS). This module is also found at the C-terminus of the γ-carbonic anhydrase (γ-CA) CcmM (Figure **3.19 A**), which has carbonic anhydrase function or not depending on the species [Pena et al., 2010], but in either case acts as scaffold during carboxysome formation [Long et al., 2007, 2010]. CcmM occurs in two isoforms. The full length CcmM comprises the γ-CA domain and three to five C-terminal SSUL domains and forms trimers is solution, while the short isoform, M35, is a monomer and consists of the SSUL domains only. These SSUL domains have been shown to aggregate Rubisco to liquid-like condensates and redox sensitive cysteine residues mediate the redox dependence of this process [Long et al., 2007; Wang et al., 2019]. The SSUL domain (residues 325-414) of *Nos*Rca (*Nos*SSUL) shares 56 - 64 % sequence identity with the SSUL domains of CcmM from *Nostoc* sp. PCC 7120 (*Nos*CcmM; short form *Nos*M35) (Figure **3.19 B**) and is attached to

the AAA+ domain by a presumably flexible linker of ∼ 35 amino acids. Interestingly, all residues but Pro263 in the SSUL from *Se*CcmM which were shown to contact *Se*Rubisco within the CcmM:Rubisco condensate are conserved in *Nos*Rca (Figure **3.19 B**).

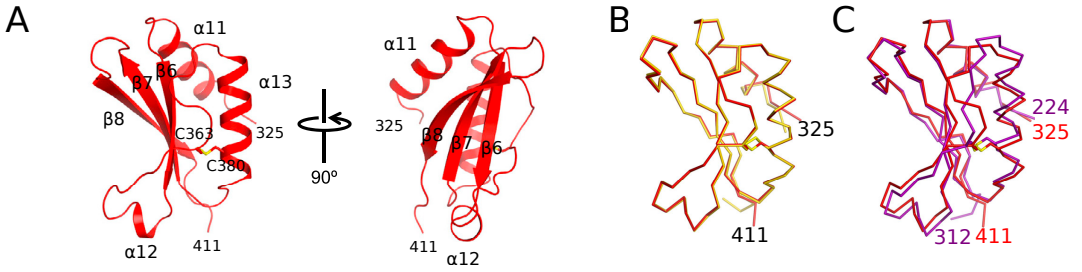

Figure 3.20.: The crystal structure of ox*Nos*SSUL closely resembles the α/β-fold of the SSUL1 from *Se*M35.
A) Crystal structure of oxidized *Nos*SSUL. *Nos*SSUL chain A is represented in ribbon representation in red and chain termini, disulfide bridge and secondary structure elements are indicated. **B) Superposition of chain A and B of *Nos*SSUL in the asymmetric unit of the crystal lattice.** Cα-traces of chain A and B are represented in red and yellow, respectively. The rmsd for matching Cα positions is 0.27 Å. **C) Superposition of *Nos*SSUL with the SSUL1 domain of CcmM from *S. elongatus* (*Se*SSUL; PDB: 6HBA).** Cα-traces of *Nos*SSUL chain A and *Se*SSUL1 are represented in red and magenta, respectively. The rmsd for 82 matching Cα positions is 1.0 Å. Crystallographic data of *Nos*SSUL was obtained and interpreted in collaboration with Leo Popilka and Andreas Bracher. Figures adapted from [Flecken et al., 2020].

To investigate whether the sequence similarity extends to the structural level, we produced recombinant *Nos*SSUL and solved its crystal structure by MAD (Figure **3.20 A** and **B**). A representative purification of *Nos*SSUL is provided in Chapter **A.1.1.3**. The asymmetric unit contained two nearly identical copies of the domain (rmsd 0.43 Å for all Cα positions), revealing a α/β-fold similar to the SSUL domain structures of CcmM from *S. elongatus* (rmsd 1 Å for 82 matching Cα positions) (Figure **3.20 B, C**). Like the first SSUL domain of *Se*CcmM in oxidizing conditions, the two cysteines (Cys363 and Cys380) of *Nos*SSUL formed disulfides.

The presence of a SSUL domain at the C-terminus of *Nos*Rca, which shares considerable sequence and structural similarity with the Rubisco interacting SSUL domains of CcmM, suggested a similar mechanism of interaction with Rubisco to ensure the co-localization with its substrate in the carboxysome.

3.2.2. *Nos*SSUL interacts with Rubisco

To analyze whether the SSUL of *Nos*Rca interacts with Rubisco, we performed a gel-shift assay. Upon co-incubation of *Nos*Rca and *Nos*Rubisco in presence of reducing agent and subsequent native PAGE analysis, we found that *Nos*Rca but not

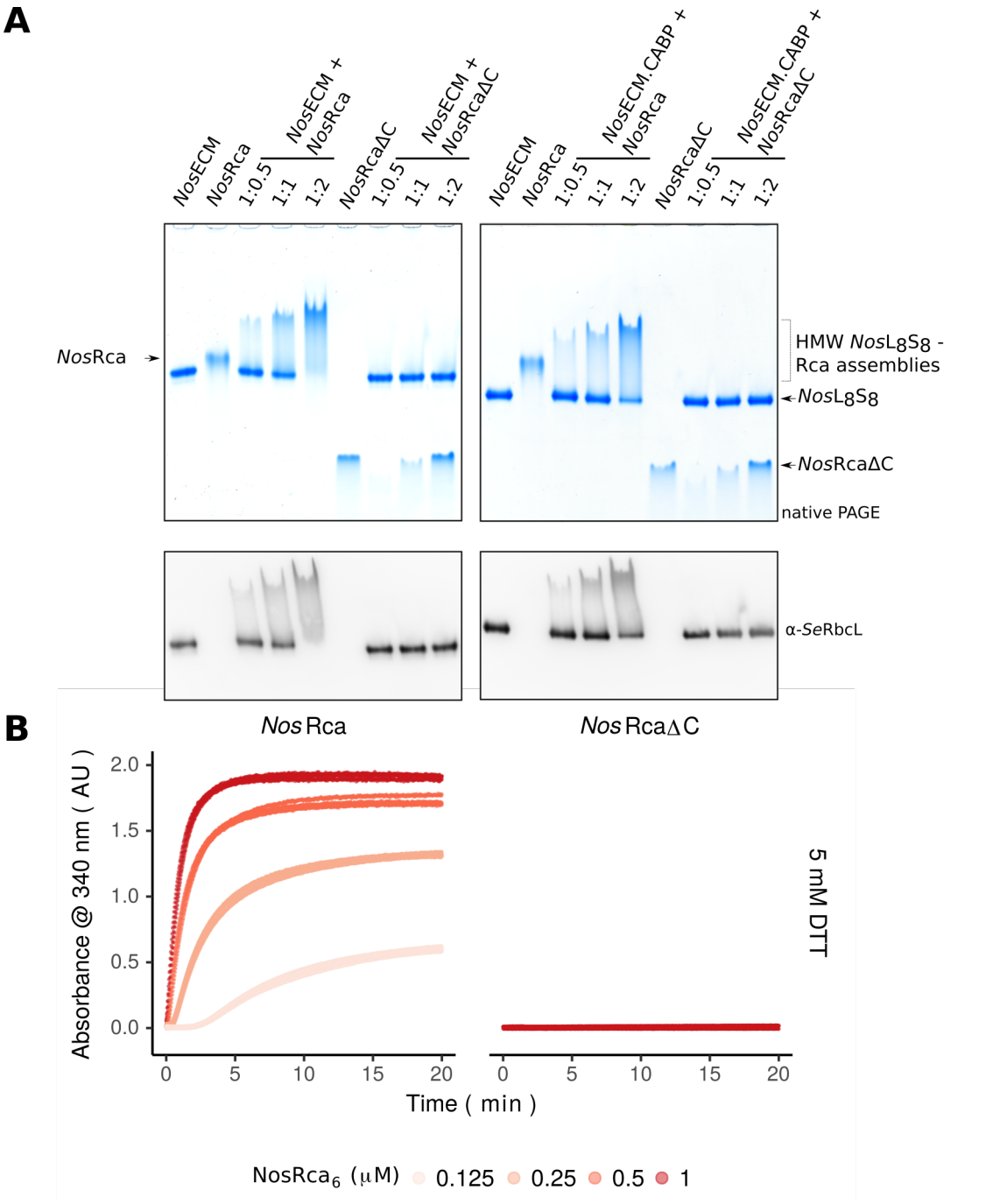

Figure 3.21.: *Nos*Rca interacts with Rubisco via the SSUL domain.
A) *Nos*Rca but not *Nos*RcaΔC shifts *Nos*Rubisco to high molecular weight assemblies. *Nos*ECM or *Nos*ECM.CABP (0.75 µM holoenzyme) were co-incubated with *Nos*Rca or *Nos*RcaΔC (0.325 µM, 0.75 µM, 1.5 µM) at indicated ratios for 30 min in native sample buffer including 2 mM DTT and subsequently analyzed by native PAGE and immunostaining with the α-*Se*RbcL antibody. **B) Upon co-incubation of *Nos*Rca but not *Nos*RcaΔC with *Nos*Rubisco the reaction turned turbid.** In presence of 5 mM DTT, *Nos*Rca or *Nos*RcaΔC were added to 0.25 µM *Nos*Rubisco at indicated concentrations and scattering was monitored at 340 nm. *Nos*RcaΔC was added at the highest concentration (1 µM) only. Data of three individual replicates are shown. AU: arbitrary units.

*Nos*RcaΔC shifted Rubisco to a higher molecular weight species in a concentration-dependent manner (Figure **3.21 A**). This effect was largely independent of the inhibition state of Rubisco, since Rubisco in both ECM and ECM.CABP form were shifted. Only after prolonged incubation in CABP (> 3 h) was the extent of the shift reduced, presumably by global structural rearrangements of Rubisco induced by tight binding of CABP (data not shown).

The SSUL-mediated interaction between *Nos*Rca and Rubisco was recapitulated by a turbidity assay, a method that monitors light scattering at 340 nm wavelength and which has been used previously to study the M35:Rubisco interaction in *S. elongatus* [Wang et al., 2019]. Upon addition of *Nos*Rca to *Nos*Rubisco in reducing conditions, the solution rapidly turned turbid and scattering developed in a concentration-dependent manner (Figure **3.21 B**). As observed in gel-shift assays, only addition of *Nos*Rca, but not *Nos*RcaΔC, resulted in intense scattering, while solutions of *Nos*Rca or *Nos*Rubisco alone remained clear (not shown).

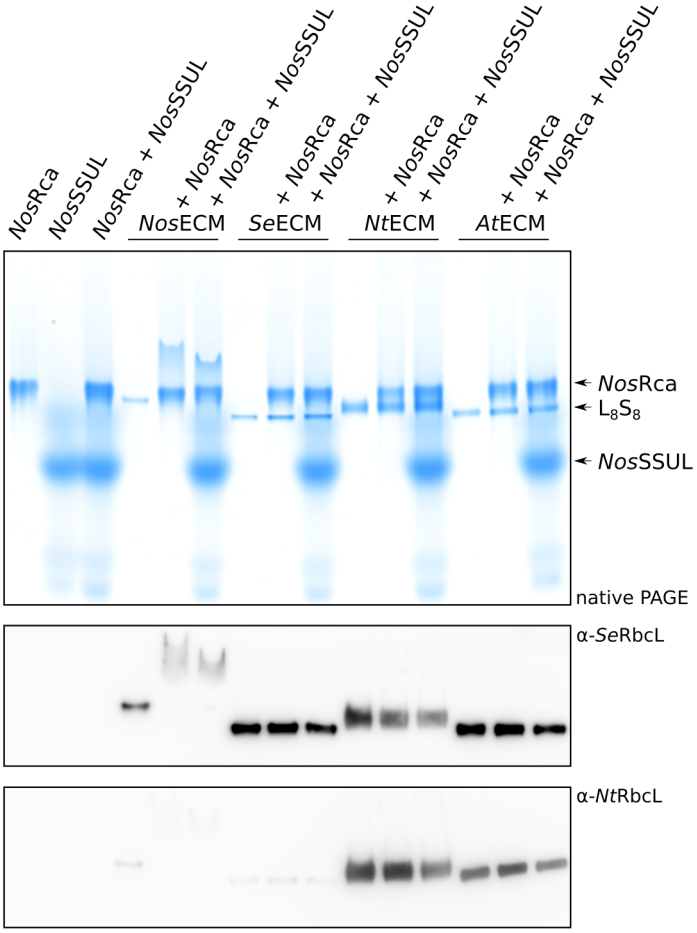

Figure 3.22.: The *Nos*SSUL:Rubisco interaction is specific.
0.156 µM *Nos*ECM, *Se*ECM, *Nt*ECM and *At*ECM were co-incubated with 2.5 µM *Nos*Rca, or *Nos*Rca and 150 µM *Nos*SSUL as described in Figure **3.21**.

Interestingly, the Rubisco of the cyanobacterium *Synechococcus elongatus* PCC 7942 (*S. elongatus*, *Se*Rubisco), which does not encode a cyanobacterial *rca* gene, failed to interact with *Nos*Rca, like Rubisco from Arabidopsis (*At*ECM) and Tobacco (*Nt*ECM) (Figure **3.21 B**), suggesting specificity of this interaction. Moreover, addition of the purified *Nos*SSUL alone to *Nos*Rca-*Nos*Rubisco reactions appeared to compete the *Nos*Rca-*Nos*Rubisco interaction as it led to a decrease of the extend of the shift.

3.2.3. Properties of the *Nos*SSUL-mediated interaction

Like the first two SSUL domains of M35 from *S. elongatus* (*Se*M35) and *Nos*M35, *Nos*SSUL contains two cysteines that formed a disulfide bond in the crystal structure (Figure **3.19 B** and **3.20**). Two additional cysteines are located in the α/β-subdomain of *Nos*Rca which cannot form disulfide bonds in their native conformation. To test the redox dependence of turbidity formation we generated an oxidized *Nos*Rca by purification in absence of reducing agent followed by air oxidation (ox*Nos*Rca; opposed to re*Nos*Rca, which had been purified in presence of reducing agent). $\sim 100\,\%$ of cysteines within the SSUL domain of ox*Nos*Rca formed disulfide bonds, as estimated by a modified Ellman assay (Figure **3.25 C**). Besides, only $\sim 50\,\%$ of cysteines within the SSUL domain of re*Nos*Rca were reduced. Upon addition of ox*Nos*Rca to *Nos*Rubisco in absence of reducing agent the solution remained clear (Figure **3.23 A**). However, DTT addition after 5 min to the ox*Nos*Rca:*Nos*Rubisco reaction resulted in rapid formation of scattering that reached levels similar to the re*Nos*Rca:*Nos*Rubisco reaction after 20 min. Moreover, turbidity formation of re*Nos*Rca and *Nos*Rubisco was dependent on salt (Figure **3.23 B**) and pH (not shown), reminiscent of properties of the phase-separated liquid-like M35:Rubisco condensate required for carboxysome biogenesis [Wang et al., 2019]. To test whether *Nos*Rca and Rubisco indeed formed phase-separated droplets, we N-terminally labeled *Nos*Rubisco with Alexa Fluor 532 (Rubisco-A532) and *Nos*Rca (re*Nos*Rca, ox*Nos*Rca or *Nos*RcaΔC) with Alexa Fluor 405 and monitored droplet formation upon co-incubation of differentially labeled proteins in presence of ATPγS (Figure **3.23 C**). We found that re*Nos*Rca and *Nos*Rubisco co-localized in liquid-like phase-separated droplets, while upon co-incubation of ox*Nos*Rca with Rubisco or *Nos*RcaΔC only very small or no droplets formed, respectively, consistent with turbidity experiments. Of note, the redox state of the SSUL located thiols neither affected ATP hydrolysis nor reactivation activity (Figure **3.24 B, C**) or the oligomeric state of the hexameric protein complex, as analyzed by size exclusion chromatog-

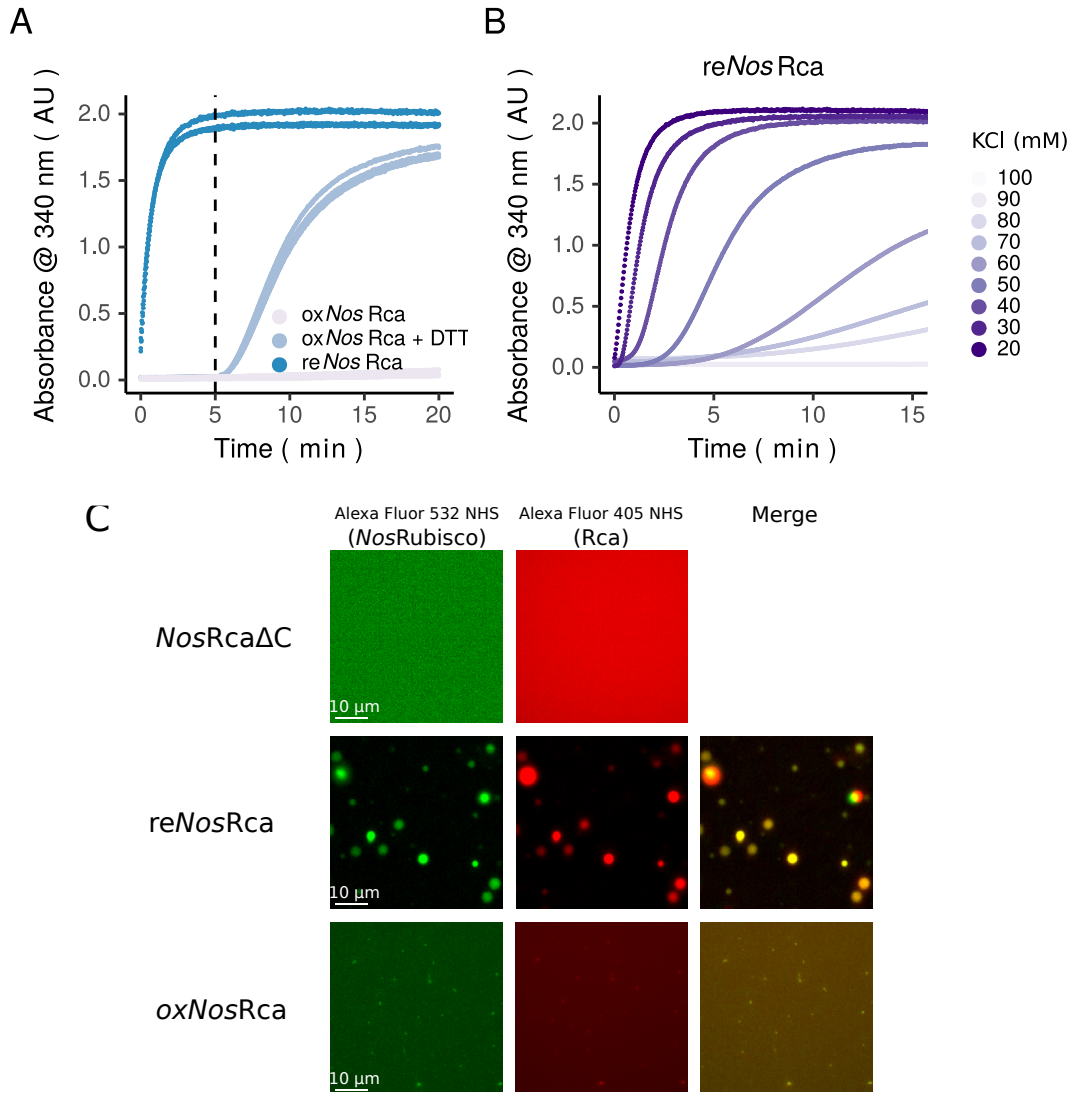

Figure 3.23.: Properties of the *Nos*Rca:Rubisco condensate.
A) Condensate formation is redox dependent. 1 µM (hexamer) ox*Nos*Rca was added to 0.25 µM *Nos*Rubisco and after 5 min 30 mM DTT was added to the reaction or not. Condensate formation is shown in comparison to re*Nos*Rca. Data of three individual experiments are shown. **B) Condensate formation is salt dependent.** 0.5 µM re*Nos*Rca was added to 0.25 µM *Nos*Rubisco at indicated KCl concentrations and turbidity was monitored. **C) re*Nos*Rca and Rubisco co-localized in phase separated liquid-like droplets.** Fluorescently labeled proteins (A405-re*Nos*Rca, A405-ox*Nos*Rca, A405-*Nos*RcaΔC and A532-*Nos*Rubisco) were diluted 1:10 in respective unlabeled proteins and coincubated at 25 °C (0.5 µM Rca and 0.25 µM Rubisco) as indicated and droplet formation was monitored in presence of 5 mM DTT and 2 mM ATPγS. Confocal microscopy data was obtained in collaboration with Huping Wang. Figure adapted from [Flecken et al., 2020].

raphy (not shown). Importantly, this SSUL-mediated, redox-regulated affinity to *Nos*Rubisco may explain how Rca co-localizes with Rubisco during carboxysome biogenesis in the reducing cytosol and is subsequently released in the oxidative environment of the mature carboxysome to fulfill its remodeling function. Interestingly, while in M35 the LLPS facilitating avidity is generated by multiple SSUL domains

per monomer, in cyanobacterial Rca each subunit of the hexamer has only one SSUL domain.

3.2.4. The binding site of *Nos*SSUL and the SSULs of CcmM on Rubisco are conserved

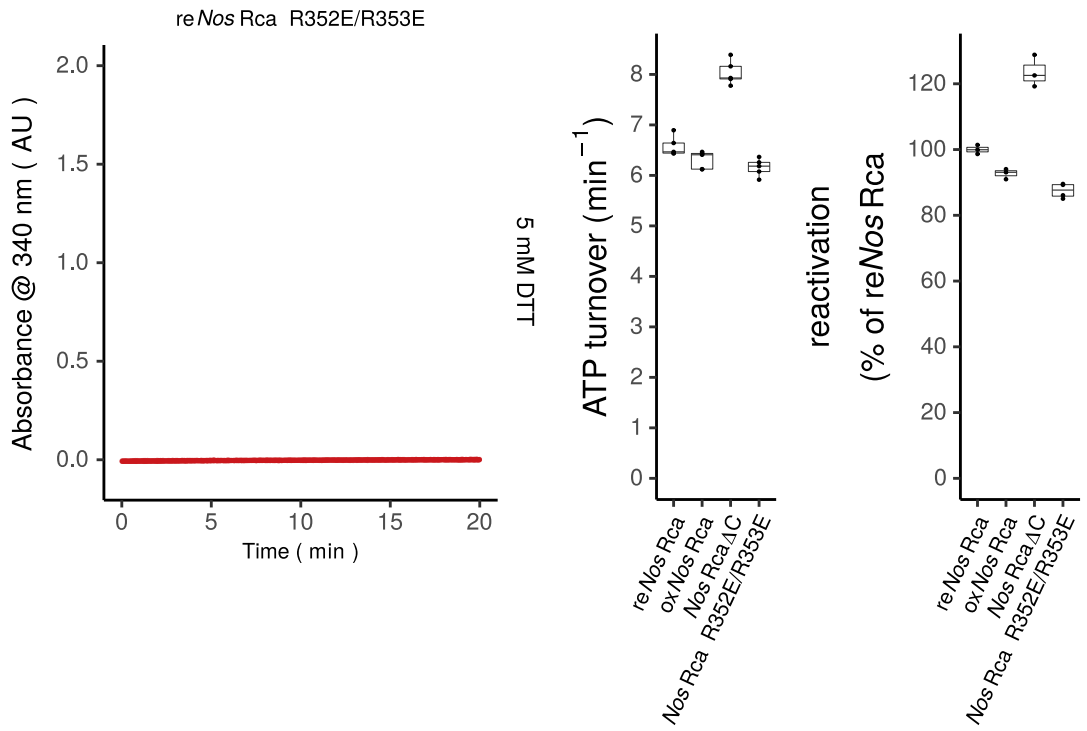

Figure 3.24.: Conserved arginine residues are required for the SSUL-mediated condensate formation by *Nos*Rca, while their mutation to glutamate does neither affect ATP hydrolysis nor Rubisco reactivation.
A) *Nos*Rca-Rubisco condensate formation depends on *Nos*SSUL-located Arg 352/353 which are conserved between CcmM and Rca. *Nos*Rca R352E/R353E (1 µM hexamer) was added to 0.25 µM *Nos*Rubisco and condensate formation was monitored by scattering at 340 nm. Data has been recorded simultaneously with data displayed in Figure **3.21** and three individual replicates are shown. **Neither the redox state of *Nos*Rca nor the Arg mutations (R352E/R353E) which abolish condensate formation affect ATP hydrolysis (B) or Rubisco reactivation (C).** ATP hydrolysis was measured as described in chapter **2.10.2** in the absence of *Nos*Rubisco at 20 mM KCl. Reactivation is normalized to the reactivation capabilities of re*Nos*Rca (100 %). Box and whisker plots (Tukey) represent data from four independent replicates.

The high conservation of Rubisco-contacting residues of *Se*M35 in *Nos*SSUL (Figure **3.19 B**), suggested a conserved binding mode of the SSUL domains to Rubisco. Specifically, two critical arginine residues had been identified previously in the first two SSUL domains of *Se*M35 that formed salt bridges with interacting residues

at the dimer-dimer cleft of *Se*Rubisco and upon mutagenesis abolished condensate formation. To test our hypothesis, we introduced a charge switch and mutated the corresponding arginine residues in the SSUL of *Nos*Rca to glutamate (*Nos*Rca R352E/R353E). As expected, *Nos*Rca R352E/R353E was only slightly impaired in ATP hydrolysis and Rubisco reactivation, but condensate formation with Rubisco was abolished (Figure **3.24**).

To test whether the SSUL domains of Rca and CcmM indeed bind Rubisco at the same binding site, we analyzed the displacement of *Nos*M35 from *Nos*M35:Rubisco condensates by *Nos*SSUL. To this end, we generated reduced and oxidized *Nos*M35 (re*Nos*M35 and ox*Nos*M35, respectively) by purification in presence or absence of reducing agent as described for *Nos*Rca. A representative purification of *Nos*M35 is provided in Chapter **A.1.1.3**. According to the modified Ellman assay, re*Nos*M35 and ox*Nos*M35 were $\sim$ 25 % and $\sim$ 75 % reduced, respectively (Figure **3.25 C**). As proxy for affinity, we determined *Nos*M35 concentrations required to reach half-maximal turbidity values in presence of a constant concentration of *Nos*Rubisco, as previously reported for *Se*M35 [Wang et al., 2019]. With 0.26 +/- 0.01 μM and 1.33 +/- 0.03 μM for re*Nos*M35 and ox*Nos*M35, respectively, these concentrations were similar to those obtained for *Se*M35 (reduced: 0.206 +/- 0.003 μM; oxidized: 1.08 +/- 0.05 μM) suggesting similar importance of the redox regulation of *Nos*M35 for Rubisco condensate formation despite the presence of cyanobacterial Rca in *Nostoc* sp. PCC 7120 (Figure **3.25 B**). For the competition experiments, re*Nos*M35 or re*Nos*Rca were incubated with *Nos*Rubisco and after 10 min either *Nos*SSUL or *Nos*RbcS (as control) were added to the reaction. As expected, addition of *Nos*SSUL but not *Nos*RbcS to both re*Nos*M35:Rubisco and re*Nos*Rca:Rubisco established condensates resulted in decreasing turbidity and demonstrated competition for the same binding site (Figure **3.25 A**).

Ultimately, the biochemical findings were strongly supported by a low resolution density map of the *Nos*Rca:*Nos*Rubisco complex obtained under reducing conditions by cryo-EM. Docking of the atomic model of *Nos*SSUL into the 8.3 Å Rca:Rubisco density map resolved the binding site of *Nos*Rca to Rubisco. The SSUL of *Nos*Rca bound close to the equator of the *Nos*Rubisco holoenzyme within a groove in the RbcL-dimer cleft of Rubisco, as previously shown for the SSUL(s) in the M35:Rubisco complex from *S. elongatus* PCC 7942.

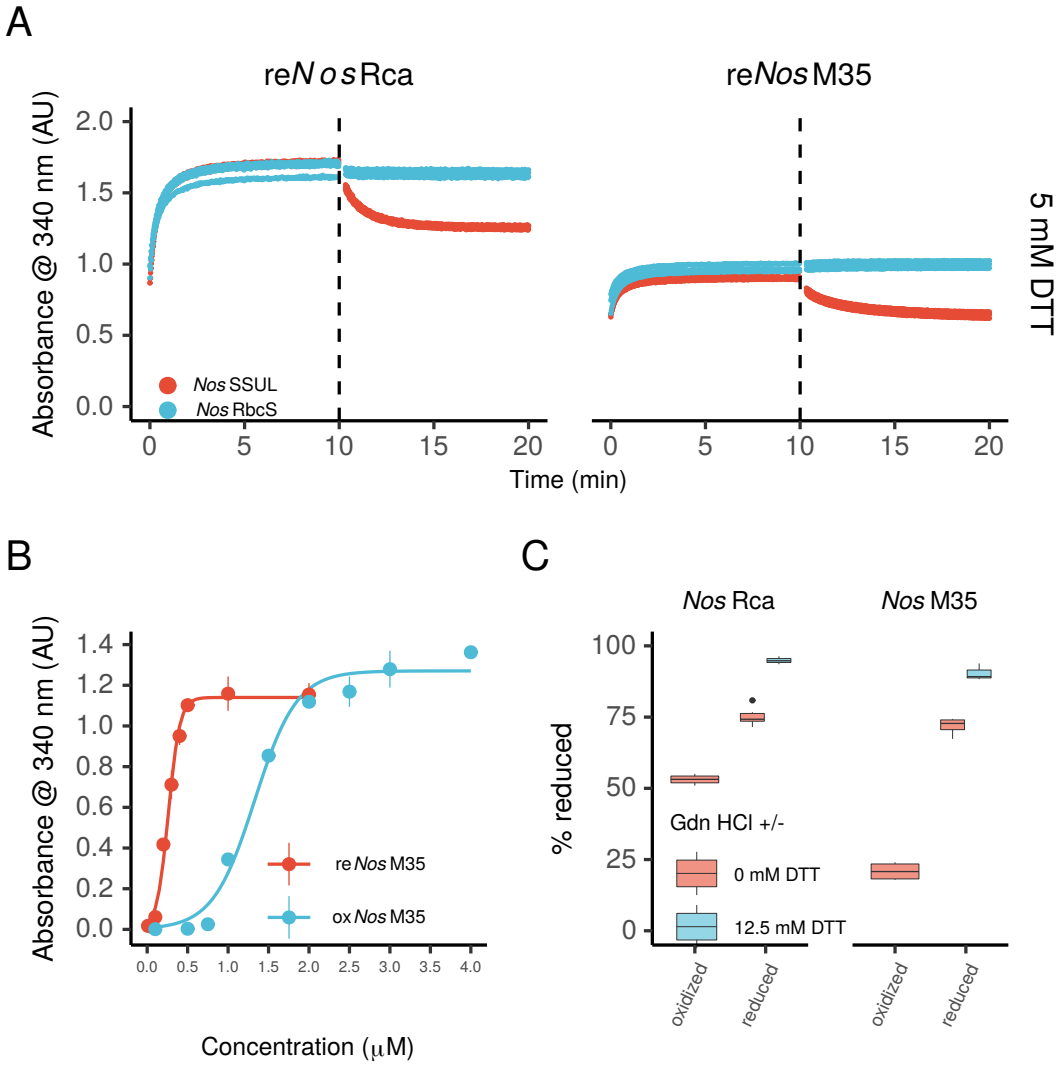

Figure 3.25.: The SSUL modules of *Nos*Rca and *Nos*M35 compete for the same binding site. A) *Nos*SSUL competes the formation of condensates of *Nos*Rubisco and *Nos*M35 or *Nos*Rca, respectively. Condensate formation was induced by addition of 1 µM re*Nos*M35 or 0.25 µM re*Nos*Rca to 0.25 µM *Nos*Rubisco. After 10 min *Nos*SSUL or *Nos*RbcS were added to a final concentration of 25 µM. To conduct the experiment at a shorter time scale, 2 mM ADP was included in the reaction mix, which increases kinetics of *Nos*Rca condensate formation. Data of three individual experiments are shown. **B) *Nos*M35 shows redox dependent Rubisco condensate formation similar to *Se*M35.** 0.25 µM *Nos*Rubisco was incubated with increasing concentrations of *Nos*M35 and turbidity values reached after 10 min were recorded as a function of *Nos*M35 concentration. The concentration required to reach the half-maximal turbidity (0.26 +/- 0.01 and 1.33 +/- 0.03 for re*Nos*M35 and ox*Nos*M35, respectively) was estimated by fitting a logistic model ($f(x) = c + \frac{0-c}{1+exp(b*(x-a))}$). Data represent the mean +/- *sd* of three independent experiments. **C) Redox state of *Nos*M35 and *Nos*Rca purified in presence or absence of reducing agent.** The fraction of free thiols was measured according to Riener et al. [2002] with modifications (see chapter **2.10.3**) in 5 mol L^{-1} GdnHCl. As control, the heat denatured re*Nos*M35 or re*Nos*Rca were solubilized in 5 mol L^{-1} GdnHCl / 10 mM DTT and any reducing agent was removed by gel filtration before analysis. Of note, *Nos*Rca contains a total of four cysteines while only the two in the SSUL domain can form disulfide bonds.

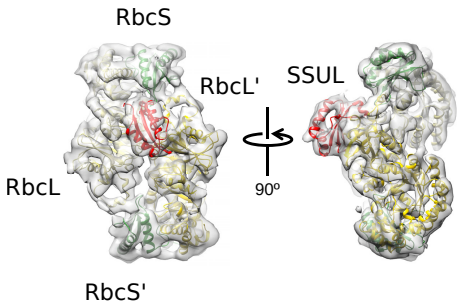

Figure 3.26.: Like SSUL(s) from *Se*M35, *Nos*SSUL binds Rubisco close to the equator in the cleft between RbcL dimers.
The density map of 2RbcL-2RbcS-SSUL units was solved at ∼ 8.2 Å resolution following a previously published single-particle analysis and 3D reconstitution procedure [Wang et al., 2019]. The RbcL subunits in gold, RbcS in green and the SSUL domain in red ribbon presentation are docked in the cryo-EM density map. Data was obtained in collaboration with Huping Wang. Figure adapted from [Flecken et al., 2020]

3.2.5. The AAA+ core of *Nos*Rca contributes to Rubisco condensate formation via avidity

As noted previously by Lechno-Yossef et al. [2020], we observed that the turbidity formed upon co-incubation of *Nos*Rca and *Nos*Rubisco was modulated by nucleotides (Figure **3.27 A**). This suggests that the AAA+ domain participated in condensate formation. To test the importance of the AAA+ domain for the SSUL-mediated binding to Rubisco, we generated a *Nos*Rca mutant with a disrupted interface I (*Nos*Rca V91E/R92G/G93N/L244D/N248A/L250A = *Nos*Rca-IF$_{mut}$). Interestingly, co-incubation of *Nos*Rca-IF$_{mut}$ with *Nos*Rubisco resulted in considerably decreased turbidity (Figure **3.27 B**). Consistently, the size of the *Nos*Rca-IF$_{mut}$:Rubisco droplets was greatly reduced in comparison to the re*Nos*Rca:Rubisco condensates in LLPS experiments (not shown), indicating that the binding to *Nos*Rubisco was strongly impaired. Note that the oligomerization of *Nos*Rca-IF$_{mut}$ was not affected (Table **3.1**), excluding that the reduced affinity to Rubisco is the result of a lower oligomerization state and reduced SSUL mediated avidity.

To better understand the nucleotide dependence of condensate formation, we analyzed the oligomeric state of *Nos*Rca in presence of nucleotide by SEC-MALS (Figure **3.27 C**). The presence of 2 mM ATP or ATPγS did not affect the oligomeric state as the molecular weight was consistent with a hexamer (no nucleotide: ∼ 265 - 274 kDa; ATP: 255.2 kDa; ATPγS: 255.8 kDa; Table **3.1**). However, the presence of 2 mM ADP resulted in a continuous, polydisperse distribution of oligomeric states, suggesting that a majority populated oligomeric states with more than nine, presumably transiently exchanging subunits. Notably, oligomerization to high molecular

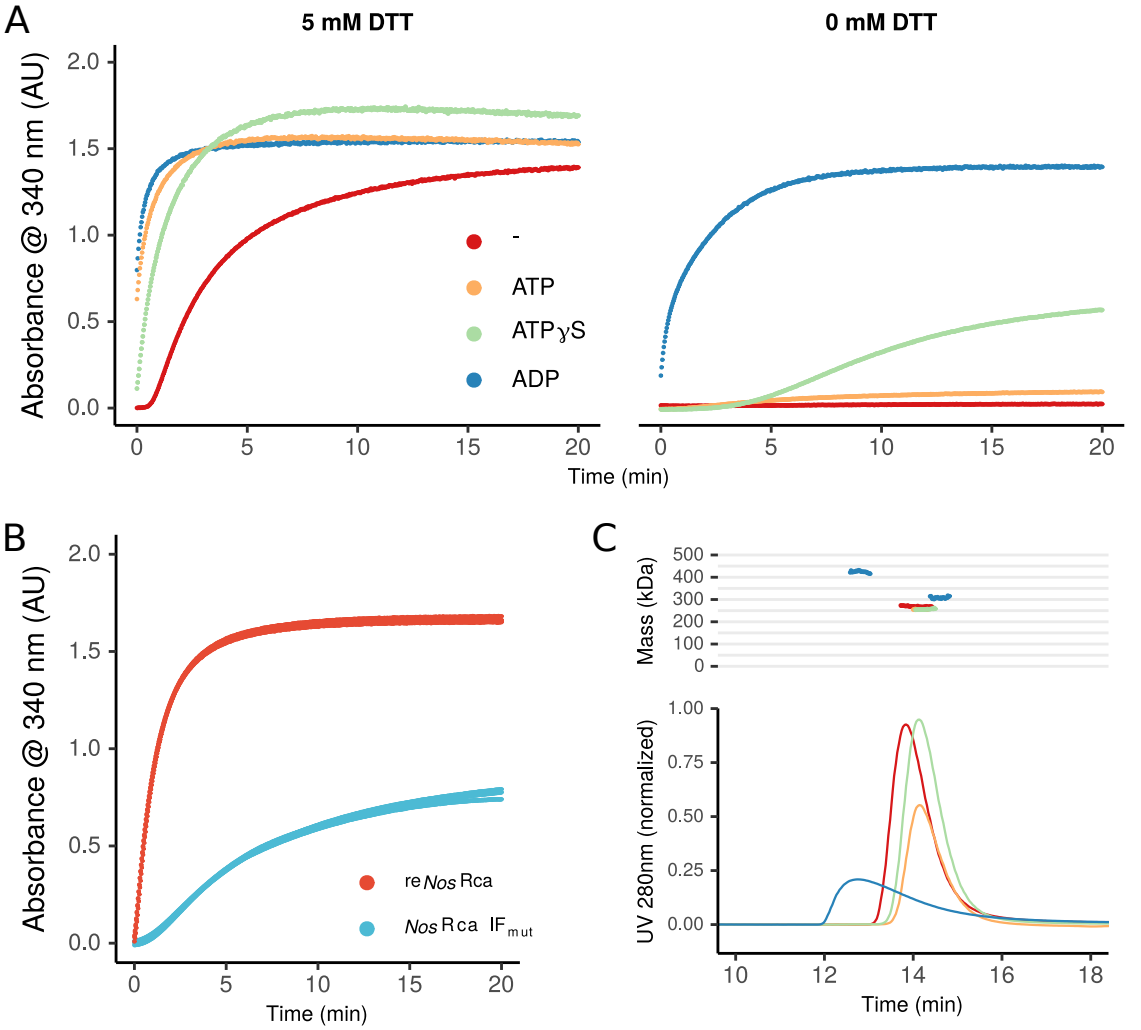

Figure 3.27.: Importance of the AAA+ domain in Rubisco condensate formation.
A) Presence of nucleotides enhances condensate formation under reducing and oxidizing conditions. Turbidity assays were performed as described in Figure (**3.21 B**) with 0.25 µM NosRubisco plus 0.25 µM reNosRca and 5 mM DTT or oxNosRca without DTT and in presence or absence of 2 mM nucleotide as indicated. **B) The AAA+ domain of *Nos*Rca contributes additional valency upon Rubisco condensate formation.** Turbidity assays were conducted at 0.5 µM re*Nos*Rca or *Nos*Rca IF_{mut} and 0.25 µM *Nos*Rubisco at 5 mM DTT. Data from three independent replicates are shown. **C) Neither ATP nor ATPγS, but ADP affects oligomerization behavior of *Nos*Rca.** 2 g L^{-1} *Nos*Rca or *Nos*RcaΔC were analyzed by SEC-MALS in presence or absence of 1 mM nucleotides as indicated. Data was obtained in collaboration with Manajit Hayer-Hartl.

weight species in presence of ADP has been reported previously for Rca from spinach [Keown and Pearce, 2014]. Hence the enhanced condensate formation in presence of ADP, as observed in turbidity experiments Figure (**3.27 A**), may be explained by the formation of high-avidity species due to transient oligomerization to complexes of higher oligomeric state than a hexamer - conditions known to favor LLPS [Shin and Brangwynne, 2017]. How ATP and ATPγS affect the affinity of Rca to Rubisco is currently unclear. Evidence that the oligomeric state indeed affects the affinity

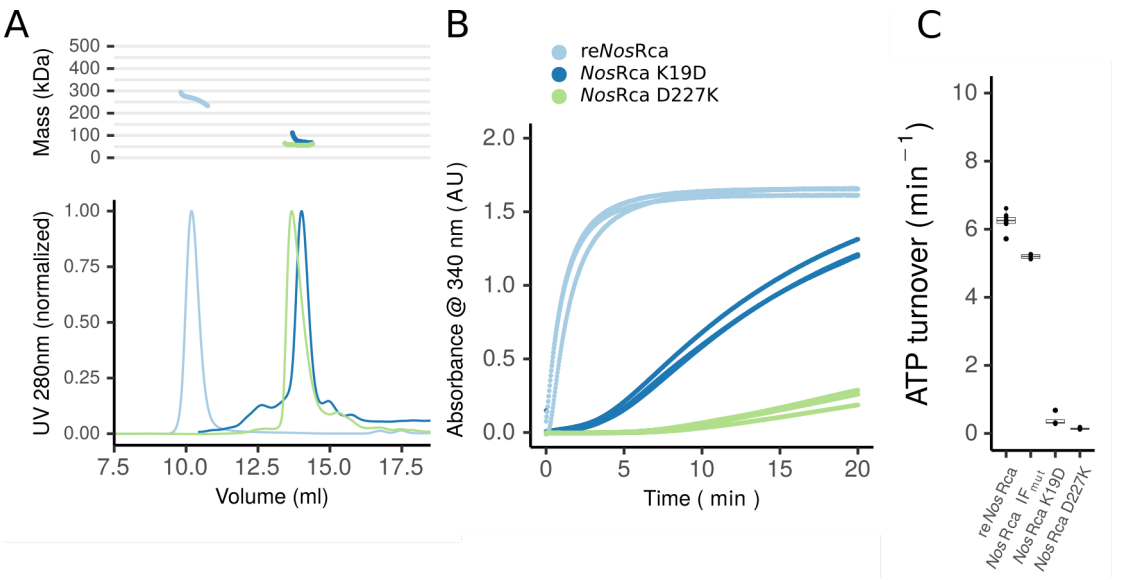

Figure 3.28.: *Nos*Rca mutants impaired in hexamer formation show reduced Rubisco condensate formation.
A,B) SEC-MALS analysis of *Nos*Rca oligomerization mutants. The oligomeric state of *Nos*Rca was analyzed by SEC-MALS in 50 mM MMT, pH 8.4 / 100 mM KCl / 10 mM $MgCl_2$. *Nos*Rca, *Nos*Rca K19D and *Nos*Rca D227K were applied at 1 g L^{-1}. : *Nos*Rca: ∼ 265 kDa; *Nos*Rca K19D: ∼ 73.8 kDa; *Nos*Rca D227K: ∼ 58.4 kDa. For clarity, data from re*Nos*Rca is reproduced from Figure **3.3**, since experiments were conducted on the same day. Data was obtained in collaboration with Monica Zobawa from the MPIB Core facility. **C) Rca oligomerization mutants are impaired in Rubisco condensate formation.** Assay was conducted as in Figure (**3.21 B**) with dimer-forming *Nos*Rca K19D and monomeric *Nos*Rca D227K. Data from three independent replicates are shown. **D) The oligomerization mutants, but not *Nos*Rca IF$_{mut}$ are impaired in ATP hydrolysis.** Assays were conducted as described in Figure **3.24 B**. Box and whisker plots (Tukey) represent data from four independent replicates.

of Rca towards Rubisco comes from the Rca-Rca interface mutants *Nos*Rca K19D and *Nos*Rca D227K, which cannot form stable oligomers due to disruption of a conserved salt bridge between Lys19 of the α/β-subdomain and Asp277 in the α-helical subdomain of a adjacent subunit. Their determined molecular weight (73.8 kDa and 58.4 kDa, respectively) corresponds to a dimer (93 kDa) or monomer (46.5 kDa), as analyzed by SEC-MALS (Table **3.1** and Figure **3.28 A**). Both mutants cannot hydrolyze ATP and thus are reactivation inactive (Figure **3.28 C**). Consistently, they are impaired in condensate formation as observed by turbidity experiments, emphasizing the importance of the hexameric state for the binding to Rubisco (Figure **3.28 B**).

Taken together, our results suggest a mechanism where the cyanobacterial Rca hexamer uses the SSUL to piggyback on Rubisco to ensure their co-localization in the carboxysome being formed. The redox dependence of this process favors the association of Rca and Rubisco in the reducing conditions of the cytosol, whereas the

Table 3.1.: Molar mass and hydrodynamic radius of proteins analyzed by SEC-MALS.
Proteins were analyzed in 50 mM MMT pH 8.4 / 100 mM KCl / 10 mM $MgCl_2$ at 25 °C. Multiple values given per protein represent replicates. Otherwise data refer to single experiments. Data was obtained in collaboration with Manajit Hayer-Hartl and Monica Zobawa (MPIB Biochemical Core facility).

Protein	**th. MW_1**[1]	**th. MW_6**[2]	**act. MW**[3]	**$r_{hydrodyn.}$**[4]
***Nos*Rca**	46668.03	280008.18	274000 265000	6.2
***Nos*Rca / ATP**			255200	5.7
***Nos*Rca / ATPγS**			255800	6.0
***Nos*Rca** (V91E/R92G/G93N/L244D/N248A/L250A)	46441.56	278649.36	267400	5.9
***Nos*Rca** (K19D)	46523.75	279142.5	73800	
***Nos*Rca** (D227K)	46549.92	279299.52	58400	
***Nos*RcaΔC**	32865.57	197193.42	199700	5.4
***Nos*RcaΔC / ATP**			212600	5.5
***Nos*RcaΔC / ATPγS**			196200	5.0
***Nos*RcaΔC** (V91E/R92G/G93N)	32722.28	196333.68	195000	5.3
***Nos*RcaΔC** (P140G/Y143A/D144A)	32558.21	195349.26	197000	5.0
***Nos*RcaΔC** (L244D/N248A/L250A)	32651.2	195907.2	197400	5.1
***Nos*SSUL**	10938		10430	1.7

[1] theoretical molecular weight of the protomer in Da; [2] theoretical molecular weight of the hexamer in Da; [3] determined molecular weight in Da; [4] determined hydrodynamic radius in nm.

comparatively oxidizing environment of mature and fully enclosed carboxysomes allows Rca to dissociate and fulfill its function.

4. Discussion

The discovery of Rubisco activase as an important regulator of Rubisco and thus photosynthetic carbon sequestration dates back ∼ 35 years. The following extensive biochemical and structural studies have elucidated key aspects of Rca regulation and function but did not resolve the mechanistic details of the mechanism underlying Rubisco remodeling. Here we applied a combination of biochemical and structural analysis to the cyanobacterial Rubisco:Rca system of *Nostoc* sp. PCC 7120, which allowed us to propose a model for the reactivation of cyanobacterial Rubisco. Importantly, we find that some aspects of the proposed mechanism also apply to the reactivation of Rubisco from plants, thus suggesting a conserved process of Rubisco form IB reactivation. Moreover, we found that the additional SSUL domain in cyanobacterial Rca, which is not required for Rubisco reactivation as such, also binds Rubisco and induces liquid-like phase separation of Rca and Rubisco, much like the carboxysomal protein CcmM. Based on biochemical and microscopy data, we propose a model wherein Rca requires this domain to piggyback on Rubisco to ensure co-localization during carboxysome formation. As an additional aspect of this study, we noticed that cyanobacterial chaperonins strongly attenuate the dependence of form IB Rubisco synthesis on additional assembly factors, which were previously considered essential for assembly of most form IB Rubiscos.

4.1. *Nos*Rubisco biogenesis

To optimize the production of cyanobacterial *Nos*Rubisco in *E. coli*, we investigated the requirement of the cognate assembly chaperones *Nos*RbcX and *Nos*Raf1 for Rubisco production upon co-overexpression of either cyanobacterial chaperonins or the endogenous *E. coli* chaperonins. Importantly, co-overexpression of *Nos*SEL was sufficient to produce high levels of recombinant functional *Nos*Rubisco, whereas co-overexpression of *Ec*SEL strictly required co-expression of either or both assembly chaperones to produce functional *Nos*Rubisco.

The GroEL-mediated folding is thought to be an iterative process comprising a cycle of substrate binding, encapsulation, and substrate release, which eventually results in the release of a natively folded substrate in its native state. Potentially, each of these steps can be affected in *Nos*RbcL folding by non-endogenous *Ec*SEL and

impede the production of folded *Nos*Rubisco. Since co-expression of the assembly factors Raf1 and RbcX, acting downstream of chaperonin, seem to partially overcome the inability of *Ec*SEL to fold *Nos*Rubisco, it seems intuitive to assume that the release of the RbcL from chaperonin is affected in *Ec*SEL, as previously shown for RbcL from *S. elongatus* PCC 6301 [Liu et al., 2010] (the Rubisco operons from *S. elongatus* PCC strains 6301 and 7942 are identical). Both RbcX and Raf1 have been shown to stabilize the RbcL dimer and to prevent rebinding to *Ec*SEL [Hauser et al., 2015; Liu et al., 2010]. Given the high abundance of soluble *Ec*SEL, prolonged engagement of *Ec*SEL and RbcL due to impaired substrate release should reveal some chaperonin-bound RbcL in the soluble fraction by SDS-PAGE analysis. However, the absence of any soluble RbcL indicates that the release of RbcL from *Ec*SEL is likely not impaired. In fact, this might indicate that *Ec*SEL does not interact with non-native *Nos*RbcL at all or, more likely, that an improperly folded, aggregation prone intermediate is produced, neither able to re-engage with *Ec*SEL for continuous productive folding nor to properly interact with RbcS, presumably aiding in the stabilization of the $RbcL_8$ core and ultimately the formation of the Rubisco holoenzyme. Since *Ec*SEL stimulates recombinant Rubisco synthesis upon co-overexpression in *E. coli* and has been shown to bind and encapsulate RbcL [Brinker et al., 2001; Goloubinoff et al., 1989b,a; Liu et al., 2010; Saschenbrecker et al., 2007], the former possibility can be excluded.

Notably, in an extensive biochemical analysis Huq et al. [2010] found that the two GroEL homologues of *S. elongatus* PCC 7942 displayed different properties which markedly differed from the *E. coli* chaperonins, with respect to aggregation prevention, ATPase rate and its stimulation by GroES, folding activities on GroEL model substrates and oligomerization properties. In light of their findings, it is reasonable to assume that these biochemical differences may also contribute to differential folding of cyanobacterial RbcL. Notably, during preparation of this work, Xia et al. [2020] undertook a structural and functional characterization of *Nos*Raf1 and demonstrated its role in *Nos*Rubisco assembly. Although the authors expressed *Nos*Rubisco by co-expression of *Nos*SEL, they reported the near absence of soluble *Nos*Rubisco after nickel-affinity purification of the His_6-tagged Rubisco (RbcL was expressed as a C-terminal His_6 fusion protein) when *Nos*Rubisco was co-expressed with *Nos*SEL alone. In contrast, upon co-expression of *Nos*Rubisco, *Nos*SEL and *Nos*Raf1, they affinity purified large amounts of soluble and active *Nos*Rubisco. How these conflicting findings may be explained is currently unclear.

Importantly, potential specific contributions of cyanobacterial chaperonins to Rubisco synthesis have not been accounted for in numerous *in vitro* Rubisco recon-

stitution experiments and recombinant expressions in *E. coli*, which all used *E. coli* chaperonins [Emlyn-Jones et al., 2006; Hauser et al., 2015; Huang et al., 2019; Kolesinski et al., 2014; Li and Tabita, 1997; Liu et al., 2010; Onizuka et al., 2004; Saschenbrecker et al., 2007]. The importance of some of the earlier findings may be challenged, if the observed differential action of *Nos*SEL on Rubisco synthesis proves to be a general property of cyanobacterial chaperonins. Certainly, it would be interesting to further investigate why *Nos*SEL and *Ec*SEL affect Rubisco production differentially and to test whether cyanobacterial chaperonins allow to reconstitute cyanobacterial Form 1B Rubisco in the absence of assembly factors *in vitro*. Moreover, applying the set of methodologies used by Huang et al. [2019, 2020] to study the *in vivo* function of assembly factors in *S. elongatus* PCC 7942 to cyanobacteria that encode RbcX in the Rubisco operon will likely help to elucidate the *in vivo* function of RbcX and Raf1 further. Recognizing a specialized function of cyanobacterial chaperonin in Rubisco folding would add another layer of complexity to the multilayered folding and assembly pathway of this most abundant enzyme and will inspire exciting research.

4.2. Mechanism of Rubisco form IB reactivation by Rca and comparison to CbbX- and CbbQ/O-type Rcas

The functional unit of Rubisco is the antiparallel dimer of large subunits that comprises two active sites at the interface formed by the C-terminal domain of one subunit and the N-terminal domain of the antiparallel subunit. Despite differences in the quaternary arrangement of large subunit dimers of form I and II Rubisco, both position and architecture of the active sites are highly conserved (Figure **4.1 A**) [Andersson and Backlund, 2008; Satagopan et al., 2014; Sugawara et al., 1999; Taylor and Andersson, 1996, 1997]. In fact, conservation of the active site extends to the primary sequence level [Satagopan et al., 2014; Sugawara et al., 1999]. Yet, the wide distribution of Rubisco enzymes across the tree of life, concomitant with the adaptation of their hosts to highly diverse environments and lifestyles [Badger and Bek, 2008; Tabita et al., 2008], gave rise to the independent appearance of proteins with Rubisco activase function, which occurred at least three times in evolution. Rca, CbbX and CbbQ/O originated from highly distinct domains of sequence space, but all employ a versatile AAA+ motor protein for the ATP-dependent remodeling of

the active site of their substrate [Ammelburg et al., 2006; Mueller-Cajar, 2017] (Figure **4.2 A**). Because of their shared function in Rubisco reactivation, they constitute an excellent paradigm of convergent evolution.

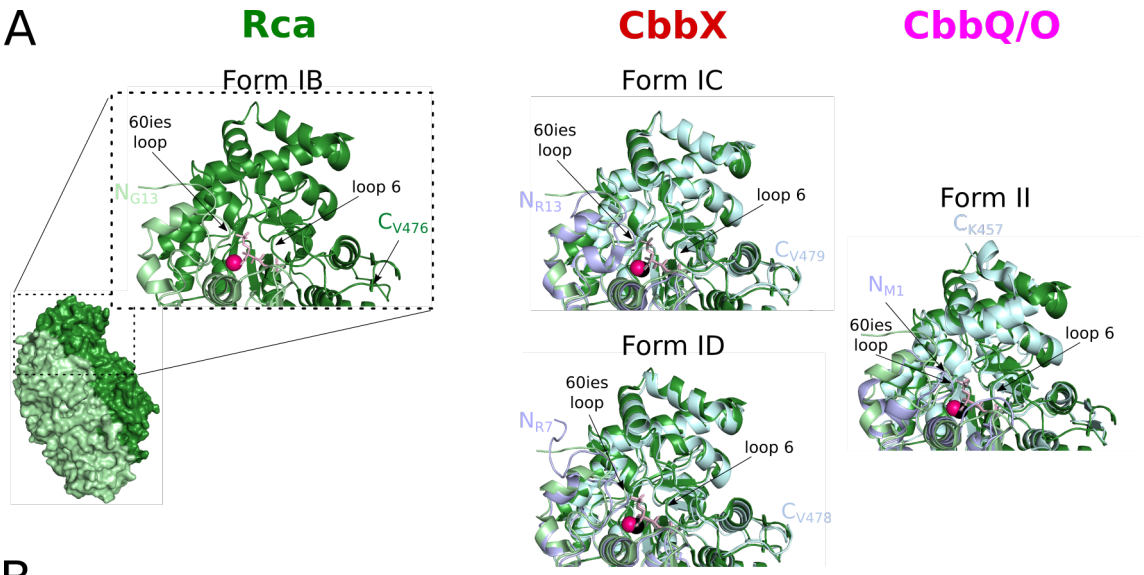

Figure 4.1.: Active site geometry is mostly conserved between Rubisco orthologs targeted by Rca, CbbX or CbbQ but length and conformation of the large subunit N- and C-termini vary.
A) View on the upper third of a *Nos*RbcL dimer comprising the CABP-bound active site in closed conformation (*left*) and superposition on corresponding regions of CABP-bound form IC, ID and II Rubisco (*right*). Protein backbones are displayed as ribbons. Chain C and D of the *Nos*RbcL dimer are shown in light and dark green, chain A and B of the other Rubisco large subunit dimers are shown in light blue and pale cyan, respectively. The Mg^{2+} ion and CABP bound to form IB Rubisco are shown as sphere and in stick representation in pink colors, respectively, those bound to other Rubiscos are shown in grey colors. 60ies loop, loop 6 and N- and C-termini are indicated. Structures were obtained from pdb accessions 5NV3, 1BWV, 4LF1 for form IC Rubisco from *Rhodobacter sphaeroides*, form ID Rubisco from *Galdieria partita* and form II Rubisco from *Rhodopseudomonas palustris* CGA009, respectively. Currently, no structure of a form IAq Rubisco is available. **B) Alignment of N- and C-termini of large subunits representative of form IAq, IAc, IB, IC, ID and form II Rubisco.** White letters on red background highlight identical residues, while residues in red indicate similarity. Homologous regions are framed in blue. NCBI accession codes: WP_012823801.1, *Halothiobacillus neapolitanus*; WP_004871237.1, *Acidithiobacillus caldus*; WP_009566926.1, *Acidithiobacillus ferrooxidans*; WP_011112458.1, *Nitrosomonas europaea*; WP_002721829.1, *Rhodobacter*; WP_119019987.1, *Rhodopseudomonas palustris*; NP_904194.1, *Physcomitrium patens*; BAB77890.1, *Nostoc* sp. PCC 7120 = FACHB-418; QCF29635.1, *Galdieria partita*; ADB27539.1, *Thalassiosira oceanica* CCMP1005; WP_029912625.1, *Hydrogenovibrio marinus*; WP_012537012.1, *Acidithiobacillus ferrooxidans*. Position of secondary structure elements were obtained from pdb accession codes mentioned above.

As pointed out previously, the high degree of conservation of Rubisco active sites across phylogenetic kingdoms suggested that each of the independently evolved Rubisco activases might employ a similar mechanism for active site opening and inhibitor release [Mueller-Cajar, 2017; Tsai et al., 2020]. However, our study reveals insights into the mechanism of a cyanobacterial Rca that seem to be representative for Rubisco form IB reactivation in general. This mechanism differs from the previously described mechanisms of form IC and ID reactivation by CbbX- and from reactivation of Rubisco from form IAq and II reactivation by CbbQ/O-type Rcas.

As previously shown for the red-type proteobacterial *Rs*Rca [Bhat et al., 2017], the hexameric *Nos*Rca docks onto Rubisco and remodels one inhibitor-bound active site at a time. C-terminal truncations of RbcL in *Rs*Rubisco provided evidence for engagement of the C-terminus by *Rs*Rca [Mueller-Cajar et al., 2011]. In contrast, *Nos*Rca is rotated clockwise by $\sim 50°$ and engages the RbcL N-terminal peptide (Figure **4.2 B**). Our single-conformation, post-remodeling snapshot of the Rca-Rubisco reactivation may indicate that Rca reaches a stable state at the end of the reactivation process. While the first RbcL N-terminal residues (residues 2-14) are stably bound in the Rca pore, the residues connecting Tyr14 with Thr24 on the surface of Rubisco are disordered, suggesting that they are conformationally flexible, perhaps because ATP dependent peptide threading comes to a halt. The split washer conformation adopted by *Nos*Rca, with the ADP bound subunit positioned at its slightly opened seam, is commonly observed in structures of substrate threading AAA+ proteins that hydrolyze ATP around the ring of the hexamer. This is in line with a partial substrate threading mechanism according to the hand-over-hand model. As we demonstrated by N-terminal truncations of plant RbcL that render plant Rubisco non-reactivatable, insertion of the RbcL N-terminal peptide is most likely a conserved feature of form IB Rubisco reactivation. Likely, plant Rubisco reactivation is modulated by the presence of an additional flexible N-terminal domain, which is missing in cyanobacterial Rca [Esau et al., 1996; Stotz et al., 2011].

According to our current model, form IB Rubisco is remodeled in three steps: (1) The RbcL N-terminal peptide (residues 2-14) binds in the Rca pore, which results in destabilization of the subsequent RbcL N-terminal region that is disordered in the remodeled state but not in the closed state. (2) Destabilization of the RbcL N-terminal region allows the 60s loop to reorient, which in turn destabilizes the active site. (3) Concomitantly, the RbcL C-terminal peptide of the antiparallel subunit is displaced from its position above the active site pocket by the bulk of the Rca-1 α4-β4 loop and helix α3. Consequently, interactions of the RbcL C-terminal peptide

with loop 6 are disrupted and result in the retraction of loop 6, thereby transforming the active site into its completely open state and allowing inhibitor release.

Close examination of N- and C-terminal sequences of Rubisco large subunits reveals differences, which may have stimulated the evolution of several distinct modes of reactivation. The C-termini of form I Rubiscos feature a conserved Asp residue (D474 in *Nos*RbcL), which serves as an important anchor that fixes the RbcL C-terminal peptide on the surface of Rubisco [Duff et al., 2000; Taylor and Andersson, 1996, 1997]. This Asp latch residue is required for proper active site closure and critical for catalysis [Bhat et al., 2017; Satagopan and Spreitzer, 2004]. RbcL C-termini of red-type form IC and ID Rubisco are extended by approximately 12 residues beyond this Asp latch residue, whereas in form IB Rubisco this Asp is rarely followed by more than two residues. Exceptions are *At*RbcL and RbcL of *Hordeum vulgare* with six residues after the conserved Asp (Figure **4.1 B**).

Presumably, the flexible tails C-terminal to the Asp latch residue in form IB Rubisco are not long enough to reach into the Rca pore of green-type Rca, whereas the extended C-terminus of red-type Rubiscos provides sufficient grip for red-type Rca to achieve active site opening by partial threading of the RbcL C-terminal peptide [Bhat et al., 2017; Loganathan et al., 2016; Mueller-Cajar et al., 2011]. This action is thought to disrupt interactions between the C-terminus and the Rubisco surface at the latch side and results in the release of loop 6 from its locked position, thereby opening the active site and enabling inhibitor release [Bhat et al., 2017]. Consequently, the RbcL C-terminal peptide cannot be engaged in form IB Rubisco reactivation and interaction with the RbcL N-terminal peptide, together with displacement of the C-terminal peptide by the bulk of the Rca core, evolved as an alternative mechanism for active site opening.

Similarly, the CbbQO-type Rca is constrained to remodel its target substrate form II Rubisco without partial threading of the large subunit C-terminus. Notably, the C-terminus of form II Rubisco does not layer on top of the substrate binding pocket and, hence, does not participate in active site closure [Satagopan et al., 2014]. Instead, the flexible C-terminus points in the opposite direction, likely due to a kink in the penultimate C-terminal α-helix that prevents the C-terminal peptide from adopting an extended conformation on top of the active site [Satagopan et al., 2014]. Consistently, form II Rubisco lacks the conserved Asp latch residue and features both shorter N- and C-termini (Figure **4.1**). In the closed state, the first six residues N-terminal to the conserved Tyr7 (Tyr21 in *Nos*RbcL) are stably attached to the surface of Rubisco. Hence both the distinct function of the form II Rubisco

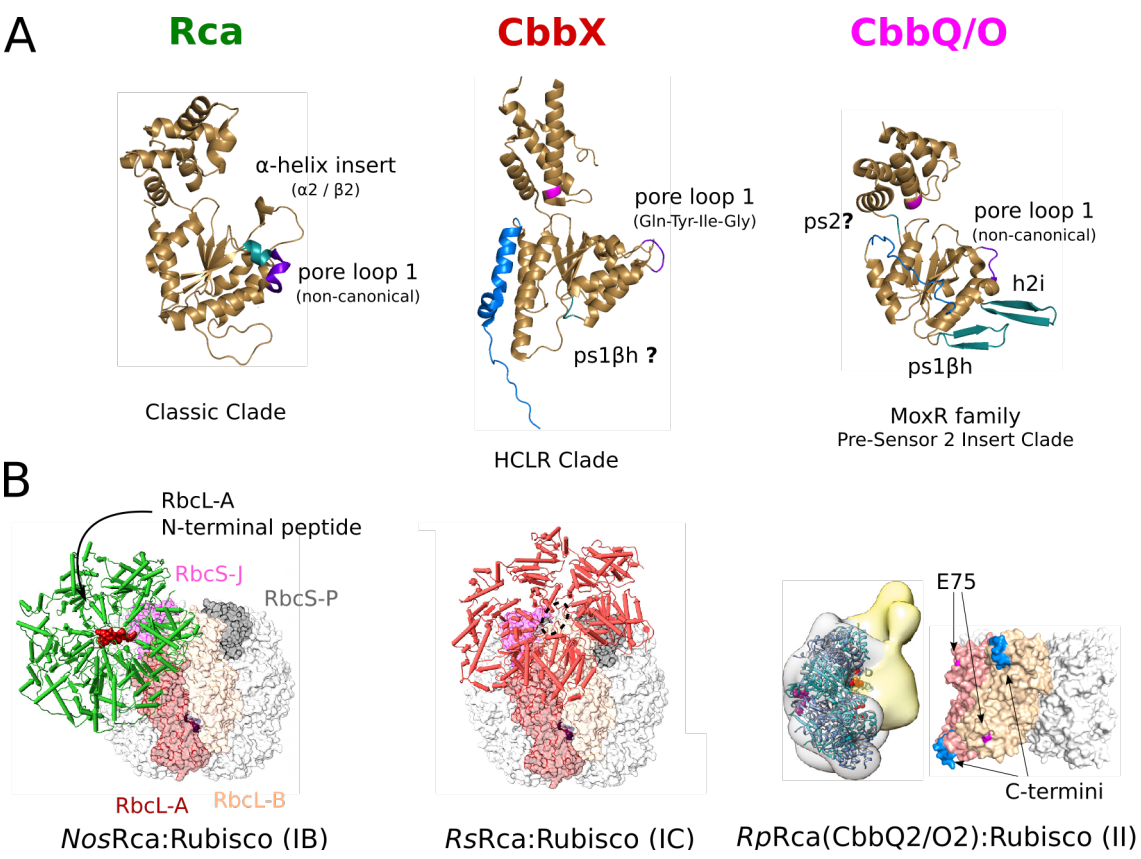

Figure 4.2.: Rca, CbbX and CbbQ/O - Rubisco activases from diverse phylogenetic backgrounds reactivate an essentially conserved substrate by different means.
A) Structural comparison of Rca, CbbX and CbbQ. Protein chains are shown as ribbons. Clade specific elements and pore loop 1 are labeled and highlighted in dark teal and purple, respectively. Sensor 2 and N-terminal extensions are shown in pink and blue, respectively. A question mark denotes the position of the labeled element in other members of the same clade, which are absent in the shown structures. h2i, helix-2 insert; ps1βh, pre-sensor 1 β-hairpin; ps2, pre-sensor 2 insert. Structure coordinates from PDB accessions 3T15, 3SYL and 6L1Q were used to display Rca, CbbX and CbbQ, respectively. **B) Comparison of the complexes of *Nos*RcaΔC (*left*) and *Rs*Rca (*middle*) with their cognate Rubisco and a hypothetical model of the CbbQ2/O2:Rubisco interaction in *Rhodopseudomonas palustris* (*right*).** Rubisco is shown in surface representation and Rca, CbbX and CbbQ are shown as ribbons. *Nos*Rubisco and *Rs*Rubisco are shown in the same orientation and the antiparallel RbcL dimer and the RbcS subunits involved in the interaction with Rca are indicated. The black-dashed circle indicates the central pore of *Rs*Rca. The *Rp*Rca structure is superposed on a low resolution EM density of the CbbQ2/O2 complex, and the position of the VWA domain is indicated. The position of the *Rp*Rubisco C-termini is indicated in blue and important residues are highlighted. *Rs*Rca:Rubisco model, EMDB: EMD-3701/PDB: 3ZUH and PDB: 5NV3; *Rp*Rca:Rubisco model, EMDB: EMD-0789/PDB: 6L1Q and PDB: 4LF1. Figures from B were adapted from [Flecken et al., 2020] and [Tsai et al., 2020].

C-terminus together with the unavailability of a flexible N-terminus, likely stimulated the evolution of a reactivation mechanism that does not rely on the threading of either of the large subunit termini. Instead, the VWA domain in the adapter protein CbbO recognizes Rubisco by a mechanism that involves the interaction of a functional MIDAS motif in the VWA domain with a conserved acidic residue on the

surface of Rubisco as well as the intact Rubisco C-terminus. Substrate binding is communicated to the CbbQ motor by a series of conserved residues and stimulates ATP hydrolysis. The mechanical force generated by CbbQ is thought to be utilized for active site remodeling by force transmission via CbbO [Tsai et al., 2015, 2020].

It is remarkable that although the general reaction mechanism of Rubisco has been highly conserved in a wide range of organisms (with only fine differences in otherwise conserved active sites), entirely different Rca types have evolved for the shared purpose of reactivating Rubisco.

4.3. Binding of the RbcL N-terminus

A key finding of our structure of the *Nos*RcaΔC:Rubisco post-remodeling complex was the presence of the extended N-terminal RbcL peptide bound in the central pore of Rca. Close inspection of the N-terminal peptide sequence revealed a pattern of alternating small and bulky sidechains. Surprisingly, we found that the bulky side chains were solvent-exposed and always pointed towards the central channel, whereas the smaller side chains pointed into successive binding pockets formed by PL1 and PL2 of Rca. While form 1B RbcL sequences of Rca-encoding cyanobacteria and eukaryotes adhere to this alternating pattern of amino acids with small and bulky side chains, with some variation in sequence, RbcL N-termini of related β-cyanobacteria not encoding the *rca* gene appear more variable in length and sequence. This is indicative of an alleviated adaptive pressure on the RbcL N-termini of form 1B Rubiscos that do not rely on remodeling by green-type Rca. Binding of the RbcL N-terminal peptide in the central Rca channel is mostly mediated by van der Waals interactions and hydrogen bonds between pore loop residues and the peptide backbone, indicating a relaxed preference for specific side chains. Hence we were surprised to see that the Rubisco mutant *Nos*L_8S_8ΔN:*Cr*N11, in which the 12 N-terminal RbcL residues of *Nos*Rubisco are exchanged for 9 N-terminal RbcL residues of *C. reinhardtii*, could no longer be reactivated by *Nos*Rca, despite of the presence of an alternating pattern of small and bulky side chains similar to the one observed in cyanobacteria.

Notably, in plants and algae it has been shown that the N-terminus of RbcL is post-translationally processed, resulting in cleavage of the first two residues before the conserved Pro1 and its acetylation by a yet unknown factor [Houtz et al., 1992, 1989]. Moreover, the RbcL N-terminus of some plants (e.g. tobacco, pea, potato,

pepper, but not *C. reinhardtii*, the liverwort *Marchantia*, wheat or spinach) is subject to trimethylation of Lys14 by the chloroplast-localized Rubisco large subunit methyltransferase (RLSMT) [Houtz et al., 1992; Raunser et al., 2009]. Perhaps the post-translationally modified N-terminal peptide of eukaryotic form 1B RbcL is more bulky and conformationally restrained than its cyanobacterial counterpart and thus may demand for more flexible pore loop actions upon Rca engagement. Intriguingly, we noticed that PL 1 residue Pro69, which is conserved in cyanobacteria, rigidifies PL 1 and forms hydrogen bonds to the substrate peptide, is replaced by a potentially flexible Gly residue in plants and algae. Although our recombinantly produced Rubisco is not subject to post-translational modifications, the Pro1 in the *Cr*RbcL N-terminal peptide might pose a structural constraint sufficient to interfere with peptide engagement, thus possibly explaining the failure of *Nos*Rca to reactivate the *Nos*$L_8S_8\Delta$N:*Cr*N11 mutant.

In the same context, the conserved charge reversal at residue 7 of the RbcL N-terminal peptides (*Nos*RbcL Lys7 corresponds to *Cr*RbcL Glu4), which mostly is a positive residue in cyanobacteria and always a negative residue in plants and algae, is likely less relevant for reactivation, because interactions between pore loops and the substrate peptide are not mediated by electrostatic interactions. Notably, Ng et al. [2020] demonstrated that the mutation of the corresponding residue in *At*RbcL to Ala did not affect reactivation by *At*Rca. Their work was published during the preparation of this thesis and independently demonstrated that reactivation of *At*Rubisco by *At*Rca requires the RbcL N-terminal peptide. Furthermore, Ng et al. [2020] demonstrated that Rubisco reactivation critically depends on length and sequence of the RbcL N-terminal peptide, as both deletion of the Thr-Glu-Thr motif and the replacement by an Ala_3-peptide as well as the insertion of an Ala_3-peptide after the Thr-Glu-Thr motif abolished reactivation. Moreover, individual Thr/Ala point mutations within this motif are sufficient to severely reduce reactivation by *At*Rca, demonstrating its importance for peptide insertion or binding inside the *At*Rca pore.

4.4. Conservation of the Rca:Rubisco interface in plants

At the end of the reactivation process, after release of the inhibitory sugar phosphate, the Rca:Rubisco complex presumably reaches a stable state, as suggested by our captured *Nos*RcaΔC:Rubisco complex. In this complex, Rca engages Rubisco

at three surface regions, interfaces I-III. In interface I, subunits Rca-5 and Rca-6 engage the βC-βD loop of RbcL-A. While Rca-5 contributes residues located in H9 to interface I, Rca-6 contributes residues located in the C-terminal region of helix α3 and the following α3-β3-loop. In plants, this α3-β3-loop is one residue shorter and the „specificity helix“ H9 harbors residues that distinguish Rubisco from solanaceous and non-solanaceous plants. A distinguishing feature of solanaceous Rubiscos are the conserved RbcL residues Arg89 and Lys94. In non-solanaceous Rubiscos and cyanobacterial Rubiscos these residues correspond to Pro89 and Asp94 or Pro90 and Asp95, respectively. In our *Nos*RcaΔC:Rubisco complex, *Nos*RbcL Pro90 engages Val91 and Gly93 located in the α3-β3-loop of Rca-6, while Asp95 makes van der Waals contacts to Leu244 in helix H9 of Rca-5. To elucidate the structural constraints involved in H9 conferring specificity for solanaceous Rubiscos, we modeled the *At*Rca- and *Nt*Rca:Rubisco complexes as representatives for the non-solanaceous and solanaceous Rca-Rubisco system, respectively, on the basis of their available crystal structures. Modeling the *At*Rca:Rubisco complex suggested that Asp94 in *At*RbcL engages in electrostatic interaction with Lys313 in *At*Rca, which corresponds to Leu244 in *Nos*Rca. These residues correspond to Lys94 in *Nt*RbcL and Asp316 in *Nt*Rca, indicating that a charge reversal causes the specificity switch, which enables Rca to discriminate solanaceous from non-solanaceous Rubiscos. Indeed, modeling of the *Nt*Rca:Rubisco complex suggests a sufficiently close proximity for these residues to engage electrostatically and thus supports this model which is consistent with mutational analyses [Li et al., 2005a; Ott et al., 2000; Portis et al., 2008]. Interface I comprises an additional salt bridge between *Nos*RbcL Glu52 and *Nos*Rca-6 Arg92 that is conserved in *A. thaliana* (RbcL Glu51 - Rca Lys161) and *N. tabacum* (RbcL Glu51 - Rca Lys164).

Interactions in interface II and III are mediated by the α4-β4 loops of *Nos*Rca-1 and *Nos*Rca-2, respectively. In interface II, this loop is inserted just above the open substrate binding pocket at the interface formed by the N-terminal domain of one subunit and the C-terminal TIM barrel domain of the antiparallel RbcL subunit. In interface III, the α4-β4 loop engages both adjacent RbcS-J and RbcS-P, serving as an additional anchoring point for Rca. Notably, in addition to the residues in H9, both residues in helix α3 and the α4-β4 loop of plant Rca have been implicated in direct contacts with Rubisco and several mutations in these regions have been shown to impair or abolish reactivation with variable effects on ATP hydrolysis [Shivhare et al., 2019; Shivhare and Mueller-Cajar, 2017].

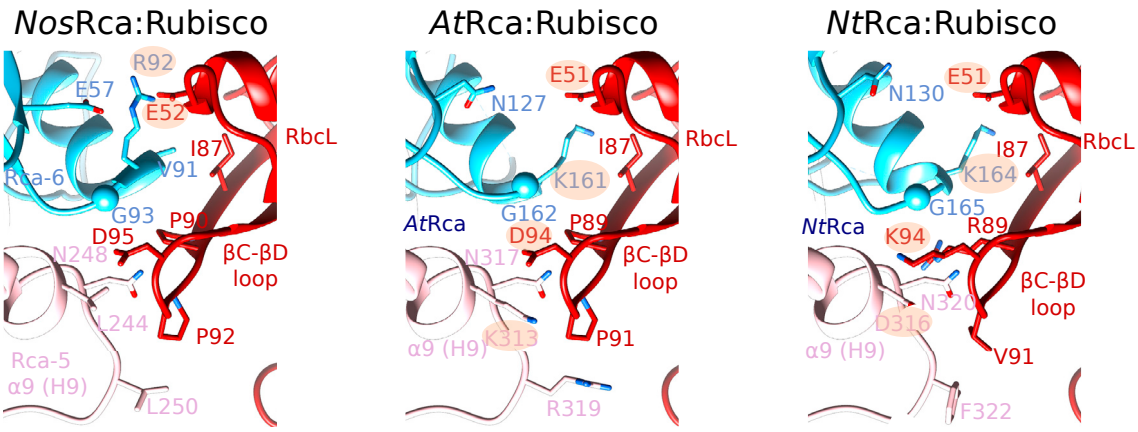

Figure 4.3.: Structural models of *At*- and *Nt*Rca:Rubisco complexes support a „charge switch“ as basis for discrimination between solanaceous and non-solanaceous Rubisco.
Left Interface I of the *Nos*RcaΔC:Rubisco complex. Structural models for interface I in the *At*Rca:*At*Rubisco (**middle**) and *Nt*Rca:*Nt*Rubisco (**right**) complexes. Protein backbones in ribbon representation and interacting amino acids in stick representation. Residues in intermolecular electrostatic interaction are highlighted. Sequence differences to *Nostoc* and side-chain conformers were manually adjusted. Figure adapted from [Flecken et al., 2020].

Consequently, reactivation of cyanobacterial Rubisco by its cognate Rca presents many features of plant Rubisco reactivation, which reinforced our hypothesis that the basic mechanism of Rubisco form IB reactivation is likely conserved.

4.5. Functional implications for Rca's mode of action and comparison with other classical AAA+ proteins

According to Ammelburg et al. [2006], Rca is phylogenetically related to the AT-Pase inactive D2 domain of NSF and structurally resembles p97/VCP/Cdc48 [Hasse et al., 2015] and spastin [Stotz et al., 2011]. For both NSF and spastin as well as for several other AAA+ ATPases the hand-over-hand or SC/2R mechanism of substrate translocation has been proposed (see Chapter **1.4.4.2**). AAA+ ATPase oligomers, for which the SC/2R model has been proposed, adopt a spiral staircase conformation and bind their substrate peptides by staggered interactions of pore loops, with ATP hydrolysis occurring in a sequential manner in an anti-clockwise direction. According to this model, the substrate is translocated in two-residue steps by ATP hydrolysis at the far side of the spiral (with respect to the substrate), which induces

the inward rotation of individual subunits and concomitantly leads to their translational movement by one position. Upon nucleotide exchange, the subunit at the far side of the spiral transitions to the uppermost (substrate facing) end, locking the previous uppermost subunit in the spiral, and binds two additional substrate residues adjacent to the engaged peptide.

Likewise, the substrate-engaged *Nos*Rca hexamer adopts a spiral staircase organization with one subunit closing the spiral by bridging both ends of the right-handed helix. The staircase-like organization of subunits is reproduced by the substrate-engaged PL1 and PL2, which form small cavities to accommodate small sidechains of substrate peptide residues. To our knowledge, Rca is the only active classical AAA+ ATPase investigated so far that does not depend on an intercalating PL1 aromatic residue for substrate translocation. The ATPase inactive D2 module of NSF also deviates from the canonical Tyr-Val-Lys motif, but nevertheless contains a PL1 aromatic residue (Phe) which does not bind the substrate peptide [Nagiec et al., 1995; White et al., 2018]. Although PL1 alone intercalates with the substrate peptide in the majority of available AAA+ ATPase structures, e.g. in Hsp104 [Gates et al., 2017], Vps4 [Han et al., 2017], NSF [White et al., 2018], ClpB [Deville et al., 2019] or the proteasomal subunits [de la Peña et al., 2018], engagement of the substrate peptide by both PL1 and PL2 has been observed in katanin [Zehr et al., 2020], spastin [Sandate et al., 2019] and YME1 [Puchades et al., 2017]. Although Rca contains an aromatic residue in PL2 (Tyr116), it does not intercalate with the bound substrate, but seemingly rigidifies the central pore via interactions with Phe110 instead.

Moreover, only one subunit at the seam, which is ADP-bound, does not contact the RbcL N-terminal peptide in Rca. This nucleotide occupation pattern is indicative of a sequential, "around the hexamer ring" ATP hydrolysis model. However, deducing directionality of ATP hydrolysis from the structure alone is challenging as the nucleotide free state is not resolved in our structure, unlike in the model of Msp1 [Wang et al., 2020] or YME1 [Puchades et al., 2017]. As demonstrated for *Nt*Rca, which requires both arginine fingers for ATP hydrolysis [Li et al., 2006], ATP hydrolysis of *Nos*Rca likely depends on the contribution of the *trans*-acting arginine fingers to the nucleotide binding site of the adjacent subunit. In our structure both arginine fingers of Rca-3 (R169:4.9 Å, R172: 5.2 Å), but also one arginine finger of Rca-1 (R169: 4.0 Å), seem to be located sufficiently close to the γ-phosphates bound to their corresponding nucleotide binding site, to contribute to hydrolysis at interfaces Rca-6:Rca-1 and Rca-2:Rca-3.

Notably in katanin, which also depends on both arginine fingers for ATP hydrolysis, a clockwise model for ATP hydrolysis has been proposed, where the subunit clockwise adjacent to the seam subunit would hydrolyze the bound ATP, thereby weakening the same interface resulting in its detachment and translocation to the opposite end of the helix [Zehr et al., 2020, 2017]. Thus, according to this model, the subunit which had just hydrolyzed ATP detaches itself from the spiral. Anticlockwise ATP hydrolysis was considered unlikely, since the arginine fingers of the ADP bound subunit were not in sufficiently close position to reach the ATP bound to the adjacent subunit in anticlockwise direction. On the contrary, in YME1, Msp1 or Vps4 ATP hydrolysis proceeds in the opposite direction, with ATP hydrolysis of the anti-clockwise adjacent subunit being required for full detachment from the spiral to allow translocation of the ADP-bound or nucleotide free subunit that hydrolyzed ATP in the previous step (Figure 4.4). Hence, nucleotide hydrolysis has to be communicated between subunits.

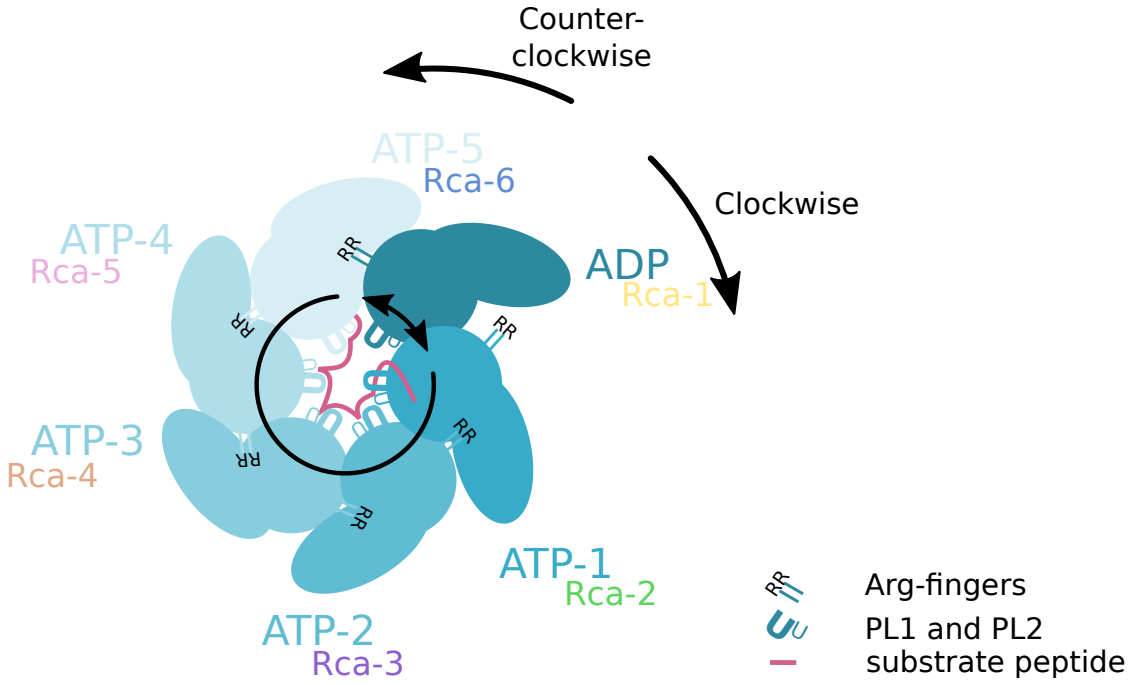

Figure 4.4.: Schematic of clockwise and anti-clockwise sequential ATP hydrolysis.
Clockwise sequential ATP hydrolysis implicates that the ATP hydrolyzing subunit itself translocates to the far side of the spiral. Counterclockwise sequential ATP hydrolysis requires intersubunit communication, because only ATP hydrolysis of the anti-clockwise adjacent subunit allows full detachment of the ADP-bound or nucleotide free subunit from the context of the spiral and its concomitant translocation. The translocating subunit does not contact the substrate peptide. In *Nos*Rca the ADP bound position is occupied by Rca-1 at the "top" of the spiral (facing Rubisco), whereas the ATP-5 position is occupied by Rca-6. Arg-fingers, PL1 and PL2 and substrate peptide are indicated.

Although the position of the arginine fingers in *Nos*Rca would likely allow ATP hydrolysis in both directions, clockwise ATP hydrolysis is unlikely, as it would not lead to productive translocation of the RbcL N-terminal peptide because the uppermost subunit in the spiral context (closest to Rubisco) would translocate to the far side of the spiral. Moreover, both arginine fingers at the Rca-6:Rca-1 interface were less resolved than those of remaining substrate engaged subunits, which may point to ongoing ATP hydrolysis in anti-clockwise direction. This mechanism could perhaps result in productive translocation of the RbcL N-terminal peptide and generate the force required for active site opening. However, as demonstrated for the HCLR AAA+ ATPase ClpX (discussed in[Fei et al., 2020]) or the AAA+ protease Lon [Shin et al., 2020], structural properties do not always reflect the full spectrum of AAA+ motor functions. Thus, further biochemical and single molecule studies are required to understand the finer details of force generation by Rca. Because ATP binding alone is not sufficient for Rubisco reactivation, which requires ATP hydrolysis, pulling the RbcL N-terminal peptide seems a plausible mechanism for force transmission to the closed active site. However, it is currently unclear whether this mechanism results in continuous translocation of a stretch of RbcL N-terminal residues or whether it generates a counterforce just large enough to facilitate active site opening primarily by displacing the RbcL C-terminus from its closed position above the active site, using the α4-β4-loop as a lever. Interestingly, the unusual properties of the Rca pore loops seem to prevent the generation of high forces, and thus likely limit the impact of Rubisco remodeling to the active site.

4.6. Rubisco:Rca condensate formation is mediated by the SSUL domain

The SSUL domain, flexibly attached to the C-terminus of the AAA+ core of cyanobacterial Rca, has previously been implicated in Rubisco aggregation in *Fremyella diplosiphon*. Lechno-Yossef et al. [2020] presented evidence for *in vivo* co-localization of Rubisco and Rca in carboxysomes and linked the function of Rubisco condensation to the SSUL domain by *in vitro* experiments.

As previously shown for the SSUL domain of CcmM [Wang et al., 2019], the SSUL domain of *Nos*Rca binds to a groove within the cleft formed by two neighboring, antiparallel RbcL dimers close to the equator of Rubisco. Indeed, the isolated SSUL domain of Rca competes with condensate formation of Rubisco with Rca or CcmM.

The observed fusion events between condensate droplets point to liquid-like properties of the Rca:Rubisco condensate, reminiscent of condensates formed with CcmM. Binding to Rubisco relies on critical arginine residues first identified in CcmM, which are conserved in the SSUL domain of Rca. Importantly, the multivalency required for LLPS is achieved by the presence of six SSUL domains in the Rca hexamer. Thus, factors affecting the oligomeric state of Rca, like the presence of ADP, will change its valency, which in turn impacts LLPS formation, as was previously reported for several proteins undergoing LLPS [Banani et al., 2016; Banjade and Rosen, 2014; Li et al., 2012; Nott et al., 2015; Su et al., 2016]. In contrast, the multivalency in CcmM is established by presence of several SSUL domains in the same protomer. While the short isoform M35 is a monomer, full length CcmM forms trimers. Therefore, differential translation of either of the isoforms will drastically affect the properties and composition of the Rubisco condensate [Huang et al., 2020; Long et al., 2010]. Interestingly, the AAA+ core of Rca contributes additional valency to Rubisco condensate formation, which may explain the residual propensity of monomeric Rca mutants to form condensates in presence of Rubisco. Although the increased propensity for condensate formation in presence of ADP is likely caused by changes of the oligomeric state and thus valency, the enhanced condensate formation in presence of ATP and ATPγS, which do not affect the oligomeric state of Rca must be of distinct origin. Perhaps binding of ATP or ATPγS increases the affinity of the AAA+ domain of Rca for substrate, or it affects the availability of the SSUL domain for Rubisco binding. Concordantly, experiments with Rca from plants indicated that the Rca C-terminus likely interacts with the nucleotide binding pocket in a nucleotide dependent manner [Wang and Portis, 2006]. Hence, similar interactions between regions C-terminal to the AAA+ core and the nucleotide binding site of cyanobacterial Rca are conceivable, especially given that a conserved Tyr residue at the end of the AAA+ core is involved in the regulation of the ATPase activity in both plant [Stotz et al., 2011] and cyanobacterial Rca (this study).

Previous work with M35 from *S. elongatus* PCC 7942 illustrated that Rubisco condensation is regulated by redox sensitive cysteines contained in some SSUL domains. Likewise, *Nos*M35 recapitulates the reduced affinity to Rubisco in the oxidized state and binds Rubisco with overall similar affinities. The redox sensitive cysteines are conserved in the SSUL domains of cyanobacterial Rca orthologs and the reduced affinity of oxidized *Nos*Rca demonstrates that Rubisco condensate formation mediated by both Rca and CcmM are governed by the redox state of their local environment. Hence our findings are consistent with a model, whereby the SSUL domain enables Rca to piggyback on Rubisco during carboxysome formation in the

cytosol, to ensure their joint encapsulation during carboxysome formation. Since the carboxysomal shell is thought to prevent diffusion of reducing agents into the carboxysome, the maturating carboxysome becomes oxidizing [Chen et al., 2013]. Hence, in the mature carboxysome the affinity of Rca and Rubisco is reduced, which likely enables Rca to more freely dissociate in the carboxysome interior to fulfill its reactivation function (see figure **4.5**).

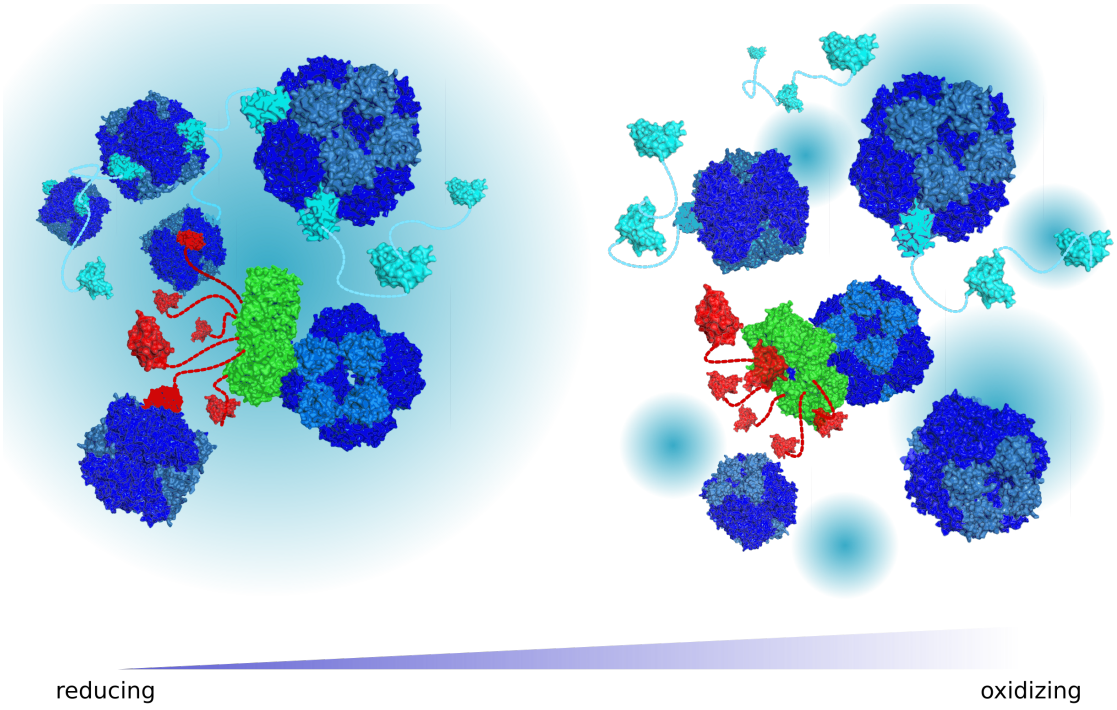

Figure 4.5.: Cyanobacterial Rca mimics SSUL driven Rubisco condensate formation by M35 to colocalize with Rubisco in carboxysomes.
Reducing conditions in the cyanobacterial cytosol allow Rubisco condensation, important for carboxysome assembly. Condensate formation is largely driven by SSUL modules of M35 and Rca uses its SSUL to piggyback on Rubisco to ensure its co-encapsulation in carboxysomes, which are known to form from the inside to the outside. As demonstrated, the SSUL-Rubisco interaction is sensitive to the oxidation state of the SSUL. Hence, oxidizing conditions in the mature carboxysome weaken the SSUL-mediated interaction and allow Rca to fulfill its reactivation function in the carboxysome. Rubisco holoenzyme in blue/lightblue, M35 in cyan, the AAA+ module of Rca in green and the linker together with the SSUL in red. Figure modified from [Flecken et al., 2020].

Since Lechno-Yossef et al. [2020] did not detect a function in Rubisco reactivation for Rca from *Fremyella diplosiphon*, they proposed a regulatory role for Rca in carboxysome formation, based on findings, that Δrca mutants did not respond to the increase of CO_2 with a reduction of the carboxysome number per cell, which is typically observed in wild type cells. Although we cannot rule out an additional regulatory role of Rca in carboxysome formation, we strongly favor a model which is centered on the reactivation function of cyanobacterial Rca. Accordingly, the find-

ings, that both *Nos*M35 and *Se*M35 show essentially the same affinity to Rubisco and similar regulation in response to alterations of the redox environment, is not indicative of a specialized function of M35 in species that do contain Rca. Importantly, unlike CcmM, cyanobacterial Rca is not essential for carboxysome formation, as demonstrated by transmission electron microscopy (TEM) images showing unaltered carboxysomes in Δ*rca* cells in comparison to wild type cells of *F. diplosiphon*. Given the close phylogenetic relation between *F. diplosiphon* and *Nostoc* sp., a species specific function of Rca seems highly unlikely.

From an evolutionary perspective, the acquisition of the SSUL domain by cyanobacterial Rca likely occurred with the appearance of carboxysomes $\sim$ 350 million years ago [Rae et al., 2013], long after the primary endosymbiotic event leading to the evolution of chloroplasts, which explains the absence of remnants of the C-terminal SSUL domain in plant Rca.

5. Implications

A functional requirement for cyanobacterial Rca has long been questioned, because its cognate Rubisco is only weakly inhibited by known Rubisco inhibitors. Moreover, not all cyanobacterial genomes do encode *rca* homologs and putative candidates of both α- and β-cyanobacterial Rca proteins have been reported to fail in reactivating inhibited Rubisco. In this study we present the first biochemical evidence for a cyanobacterial Rca. Although the physiologically relevant inhibitor of cyanobacterial Rubisco is currently not known, the presence of a functional Rca in many β-cyanobacteria indicates that cyanobacterial Rubisco is susceptible to inhibition by sugar-phosphates as well and consequently requires Rca to maintain its functional state. Our mechanistic model of cyanobacterial form IB Rubisco reactivation differs from models previously suggested for reactivation of form IC and ID Rubisco of chemoautotrophic prokaryotes and red algae by CbbX-type Rca as well as from form II Rubisco reactivation by CbbQ-type Rca. Importantly, we identify the engagement of the RbcL N-terminal peptide as a conserved feature of Rubisco form IB reactivation in plants and cyanobacteria. Although the requirement of the RbcL N-terminal peptide represents a major step towards understanding the mechanism of form IB Rubisco reactivation, some specific aspects of Rubisco reactivation in plants are still enigmatic, such as the mechanism of the N-domain of plant Rca, a feature not found in Rca from cyanobacteria, as well as the modulation of Rca activity by the C-terminal extension of the large isoform.

The importance of Rca for the functional maintenance of cyanobacterial Rubisco has implications for attempts to enhance crop productivity by reconstituting the cyanobacterial CCM in chloroplasts of plants. Our results suggest that co-expression of cyanobacterial Rca may benefit Rubisco fidelity and perhaps enhances plant productivity. Moreover, knowledge of the redox dependent targeting of cyanobacterial Rca to the carboxysome may present a handle to affect the efficacy of the CCM.

6. Bibliography

Aaij, C. and Borst, P. (1972). The gel electrophoresis of DNA. *Biochimica et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis*, 269(2):192–200. 2.1.1

Abadie, C. and Tcherkez, G. (2019). Plant sulphur metabolism is stimulated by photorespiration. *Communications Biology*, 2(1):379. 1.2.2

Adl, S. M., Simpson, A. G. B., Farmer, M. A., Andersen, R. A., Anderson, O. R., Barta, J. R., Bowser, S. S., Brugerolle, G., Fensome, R. A., Fredericq, S., James, T. Y., Karpov, S., Kugrens, P., Krug, J., Lane, C. E., Lewis, L. A., Lodge, J., Lynn, D. H., Mann, D. G., Mccourt, R. M., Mendoza, L., Moestrup, Ø., Mozley-Standridge, S. E., Nerad, T. A., Shearer, C. A., Smirnov, A. V., Spiegel, F. W., and Taylor, M. F. J. R. (2005). The New Higher Level Classification of Eukaryotes with Emphasis on the Taxonomy of Protists. *Journal of Eukaryotic Microbiology*, 52(5):399–451. 1.2

Aguzzi, A. and Altmeyer, M. (2016). Phase Separation: Linking Cellular Compartmentalization to Disease. *Trends in Cell Biology*, 26(7):547–558. 1.5.1

Aigner, H., Wilson, R. H., Bracher, A., Calisse, L., Bhat, J. Y., Hartl, F. U., and Hayer-Hartl, M. (2017). Plant RuBisCo assembly in E. coli with five chloroplast chaperones including BSD2. *Science*, 358(6368):1272–1278. 1.3.2.3.2, 2.5.4

Allen, J. F. (2005). A redox switch hypothesis for the origin of two light reactions in photosynthesis. *FEBS Letters*, 579(5):963–968. 1.2.1

Ammelburg, M., Frickey, T., and Lupas, A. N. (2006). Classification of AAA+ proteins. *Journal of Structural Biology*, 156(1):2–11. 1.4.1, 1.4.1, 1.4.2.3, 1.4.6.1, 4.2, 4.5

An, Y. J., Rowland, S. E., Na, J.-H., Spigolon, D., Hong, S. K., Yoon, Y. J., Lee, J.-H., Robb, F. T., and Cha, S.-S. (2017). Structural and mechanistic characterization of an archaeal-like chaperonin from a thermophilic bacterium. *Nature Communications*, 8(1):827. 1.1.2.3

An, Y. J., Rowland, S. E., Robb, F. T., and Cha, S.-S. (2016). Purification, crystallization, and preliminary X-ray crystallographic analysis of the Group III chaperonin from Carboxydothermus hydrogenoformans. *Journal of Microbiology*, 54(6):440–444. 1.1.2.3

Anderson, L. E. (1971). CHLOROPLAST AND CYTOPLASMIC ENZYMES n. PEA LEAF TRIOSE PHOSPHATE ISOMERASES. *BIOCHIMICA ET BIOPHYSICA ACTA*, 235:237–244. 1.2.2

Andersson, I. and Backlund, A. (2008). Structure and function of Rubisco. *Plant Physiology and Biochemistry*, 46(3):275–291. 4.2

Andersson, I. and Brändén, C.-I. (1984). Large single crystals of spinach 1,5-bisphosphate carboxylase/oxygenase suitable for X-ray studies. *Journal of Molecular Biology*, 172(3):363–366. 1.3.1

Andralojc, P. J., Madgwick, P. J., Tao, Y., Keys, A., Ward, J. L., Beale, M. H., Loveland, J. E., Jackson, P. J., Willis, A. C., Gutteridge, S., and Parry, M. A. J. (2012). 2-Carboxy-D-arabinitol 1-phosphate (CA1P) phosphatase: Evidence for a wider role in plant Rubisco regulation. *The Biochemical Journal*, 442(3):733–742. 1.3.3.4

Anfinsen, C. B. (1973). Principles that Govern the Folding of Protein Chains. *Science*, 181(4096):223–230. 1.1.1

Anfinsen, C. B., Haber, E., Sela, M., and White, F. H. (1961). The Kinetics of Formation of Native Ribonuclease During Oxidation of the Reduced Polypeptide Chain. *Proceedings of the National Academy of Sciences*, 47(9):1309–1314. 1.1.1

Aono, R., Sato, T., Yano, A., Yoshida, S., Nishitani, Y., Miki, K., Imanaka, T., and Atomi, H. (2012). Enzymatic Characterization of AMP Phosphorylase and Ribose-1,5-Bisphosphate Isomerase Functioning in an Archaeal AMP Metabolic Pathway. *Journal of Bacteriology*, 194(24):6847–6855. 1.3

Armenteros, J. J. A., Salvatore, M., Emanuelsson, O., Winther, O., von Heijne, G., Elofsson, A., and Nielsen, H. (2019). Detecting sequence signals in targeting peptides using deep learning. *Life Science Alliance*, 2(5). 2.3

Armstrong, J. A. and Schulz, J. R. (2015). Agarose Gel Electrophoresis. *Current Protocols Essential Laboratory Techniques*, 10(1):7.2.1–7.2.22. 2.1.1

Ashida, H., Danchin, A., and Yokota, A. (2005). Was photosynthetic RuBisCO recruited by acquisitive evolution from RuBisCO-like proteins involved in sulfur metabolism? *Research in Microbiology*, 156(5-6):611–618. 1.3

Ashida, H., Saito, Y., Kojima, C., Kobayashi, K., Ogasawara, N., and Yokota, A. (2003). A Functional Link Between RuBisCO-like Protein of Bacillus and Photosynthetic RuBisCO. *Science*, 302(5643):286–290. 1.3, 1.3.1.2

Aubin-Tam, M.-E., Olivares, A. O., Sauer, R. T., Baker, T. A., and Lang, M. J. (2011). Single-Molecule Protein Unfolding and Translocation by an ATP-Fueled Proteolytic Machine. *Cell*, 145(2):257–267. 1.4.4.2

Baaske, P., Wienken, C. J., Reineck, P., Duhr, S., and Braun, D. (2010). Optical Thermophoresis for Quantifying the Buffer Dependence of Aptamer Binding. *Angewandte Chemie International Edition*, 49(12):2238–2241. A.1.1.1.1

Badger, M. R. and Bek, E. J. (2008). Multiple Rubisco forms in proteobacteria: Their functional significance in relation to CO2 acquisition by the CBB cycle. *Journal of Experimental Botany*, 59(7):1525–1541. 1.3, 1.4.6.3, 4.2

Baker, R. T., Catanzariti, A.-M., Karunasekara, Y., Soboleva, T. A., Sharwood, R., Whitney, S., and Board, P. G. (2005). Using Deubiquitylating Enzymes as Research Tools. In *Methods in Enzymology*, volume 398 of *Ubiquitin and Protein Degradation, Part A*, pages 540–554. Academic Press. 2.1.3, 2.5.1, 2.5.2, A.1.1

Baker, S. H., Jin, S., Aldrich, H. C., Howard, G. T., and Shively, J. M. (1998). Insertion mutation of the form I cbbL gene encoding ribulose bisphosphate carboxylase/oxygenase (RuBisCO) in Thiobacillus neapolitanus results in expression of form II RuBisCO, loss of carboxysomes, and an increased CO2 requirement for growth. *Journal of Bacteriology*, 180(16):4133–4139. 1.4.6.3

Baker, T. S., Suh, S. W., and EISENBERGt, D. (1977). Structure of ribulose-1,5-bisphosphate carboxylase-oxygenase: Form III crystals. *Proc. Natl. Acad. Sci. USA*, page 5. 1.3.1

Balchin, D., Hayer-Hartl, M., and Hartl, F. U. (2016). In vivo aspects of protein folding and quality control. *Science*, 353(6294). 1.1.1, 1.1.2.2, 1.2, 1.1.2.3

Balchin, D., Hayer-Hartl, M., and Hartl, F. U. (2020). Recent advances in understanding catalysis of protein folding by molecular chaperones. *FEBS Letters*, 594(17):2770–2781. 1.1, 1.1.2.2, 1.2

Baldwin, R. L. (1994). Matching speed and stability. *Nature*, 369(6477):183–184. 1.1.1

Banani, S. F., Lee, H. O., Hyman, A. A., and Rosen, M. K. (2017). Biomolecular condensates: Organizers of cellular biochemistry. *Nature Reviews Molecular Cell Biology*, 18(5):285–298. 1.5, 1.5.1

Banani, S. F., Rice, A. M., Peeples, W. B., Lin, Y., Jain, S., Parker, R., and Rosen, M. K. (2016). Compositional Control of Phase-Separated Cellular Bodies. *Cell*, 166(3):651–663. 4.6

Banda, D. M., Pereira, J. H., Liu, A. K., Orr, D. J., Hammel, M., He, C., Parry, M. A. J., Carmo-Silva, E., Adams, P. D., Banfield, J. F., and Shih, P. M. (2020). Novel bacterial clade reveals origin of form I Rubisco. *Nature Plants*, 6(9):1158–1166. 1.3, 1.9

Banjade, S. and Rosen, M. K. (2014). Phase transitions of multivalent proteins can promote clustering of membrane receptors. *eLife*, 3:e04123. 4.6

Bar-On, Y. M. and Milo, R. (2019). The global mass and average rate of rubisco. *Proceedings of the National Academy of Sciences of the United States of America*, 116(10):4738–4743. 1.3

Barkow, S. R., Levchenko, I., Baker, T. A., and Sauer, R. T. (2009). Polypeptide Translocation by the AAA+ ClpXP Protease Machine. *Chemistry & Biology*, 16(6):605–612. 1.4.4.2

Barraclough, R. and Ellis, R. J. (1980). Protein synthesis in chloroplasts. IX. Assembly of newly-synthesized large subunits into ribulose bisphosphate carboxylase in isolated intact pea chloroplasts. *Biochimica Et Biophysica Acta*, 608(1):19–31. 1.3.2.1

Barta, C., Dunkle, A. M., Wachter, R. M., and Salvucci, M. E. (2010). Structural changes associated with the acute thermal instability of Rubisco activase. *Archives of Biochemistry and Biophysics*, 499(1):17–25. 1.4.6.1

Bathellier, C., Yu, L.-J., Farquhar, G. D., Coote, M. L., Lorimer, G. H., and Tcherkez, G. (2020). Ribulose 1,5-bisphosphate carboxylase/oxygenase activates O2 by electron transfer. *Proceedings of the National Academy of Sciences*, 117(39):24234–24242. 1.3.3.3

Bauwe, H., Hagemann, M., and Fernie, A. R. (2010). Photorespiration: Players, partners and origin. *Trends in Plant Science*, 15(6):330–336. 1.2, 1.2.2

Berg, J. M., Tymoczko, J. L., Jr, G. J. G., and Stryer, L. (2018). *Stryer Biochemie*. Springer Spektrum, eighth edition. 1.1

Bhat, J. Y., Miličić, G., Thieulin-Pardo, G., Bracher, A., Maxwell, A., Ciniawsky, S., Mueller-Cajar, O., Engen, J. R., Hartl, F. U., Wendler, P., and Hayer-Hartl, M. (2017). Mechanism of Enzyme Repair by the AAA+ Chaperone Rubisco Activase. *Molecular Cell*, 67(5):744–756.e6. 1.4.6.2, 3.1.1, 4.2, 4.2

Blankenship, R. E. (2017). How Cyanobacteria went green. *Science*, 355(6332):1372–1373. 1.2.1

Blayney, M. J., Whitney, S. M., and Beck, J. L. (2011). NanoESI Mass Spectrometry of Rubisco and Rubisco Activase Structures and Their Interactions with Nucleotides and Sugar Phosphates. *Journal of The American Society for Mass Spectrometry*. 1.4.6.1

Blok, N. B., Tan, D., Wang, R. Y.-R., Penczek, P. A., Baker, D., DiMaio, F., Rapoport, T. A., and Walz, T. (2015). Unique double-ring structure of the peroxisomal Pex1/Pex6 ATPase complex revealed by cryo-electron microscopy. *Proceedings of the National Academy of Sciences of the United States of America*, 112(30):E4017–E4025. 1.4.4.1

Bodenhofer, U., Bonatesta, E., Horejš-Kainrath, C., and Hochreiter, S. (2015). Msa: An R package for multiple sequence alignment. *Bioinformatics*, 31(24):3997–3999. 2.3

Boisvert, D. C., Wang, J., Otwinowski, Z., Norwich, A. L., and Sigler, P. B. (1996). The 2.4 Å crystal structure of the bacterial chaperonin GroEL complexed with ATPγS. *Nature Structural Biology*, 3(2):170–177. 1.1.2.3

Boisvert, F.-M., van Koningsbruggen, S., Navascués, J., and Lamond, A. I. (2007). The multifunctional nucleolus. *Nature Reviews Molecular Cell Biology*, 8(7):574–585. 1.5

Bombar, D., Heller, P., Sanchez-Baracaldo, P., Carter, B. J., and Zehr, J. P. (2014). Comparative genomics reveals surprising divergence of two closely related strains of uncultivated UCYN-A cyanobacteria. *The ISME journal*, 8(12):2530–2542. 1.2.1

Bowien, B., Mayer, F., Spiess, E., Pähler, A., Englisch, U., and Saenger, W. (1980). On the Structure of Crystalline Ribulosebisphosphate Carboxylase from Alcaligenes eutrophus. *European Journal of Biochemistry*, 106(2):405–410. 1.3.1

Bracher, A., Hauser, T., Liu, C., Hartl, F. U., and Hayer-Hartl, M. (2015a). Structural Analysis of the Rubisco-Assembly Chaperone RbcX-II from Chlamydomonas reinhardtii. *PLOS ONE*, 10(8):e0135448. 2.3, 2.10.4

Bracher, A., Sharma, A., Starling-Windhof, A., Hartl, F. U., and Hayer-Hartl, M. (2015b). Degradation of potent Rubisco inhibitor by selective sugar phosphatase. *Nature Plants*, 1(1):1–7. 1.3.3.4, 3.1.1

Bracher, A., Starling-Windhof, A., Hartl, F. U., and Hayer-Hartl, M. (2011). Crystal structure of a chaperone-bound assembly intermediate of form I Rubisco. *Nature Structural & Molecular Biology*, 18(8):875–880. 1.3.2.2.1

Bracher, A., Whitney, S. M., Hartl, F. U., and Hayer-Hartl, M. (2017). Biogenesis and Metabolic Maintenance of Rubisco. *Annual Review of Plant Biology*, 68(1):29–60. 1.3.1.1, 1.11, 1.12, 1.13, 1.14, 1.18, 1.21, 3.1.7

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry*, 72(1):248–254. 2.1.4.1, 2.10.1, A.10, A.11

Brangwynne, C. P., Eckmann, C. R., Courson, D. S., Rybarska, A., Hoege, C., Gharakhani, J., Jülicher, F., and Hyman, A. A. (2009). Germline P Granules Are Liquid Droplets That Localize by Controlled Dissolution/Condensation. *Science*, 324(5935):1729–1732. 1.5

Brinker, A., Pfeifer, G., Kerner, M. J., Naylor, D. J., Hartl, F. U., and Hayer-Hartl, M. (2001). Dual Function of Protein Confinement in Chaperonin-Assisted Protein Folding. *Cell*, 107(2):223–233. 1.1.2.3, 1.3.2.1, 4.1

Brutnell, T. P., Sawers, R. J., Mant, A., and Langdale, J. A. (1999). BUNDLE SHEATH DEFECTIVE2, a novel protein required for post-translational regulation of the rbcL gene of maize. *The Plant Cell*, 11(5):849–864. 1.3.2.3.2

Buchan, J. R. and Parker, R. (2009). Eukaryotic Stress Granules: The Ins and Outs of Translation. *Molecular Cell*, 36(6):932–941. 1.5

Cai, F., Dou, Z., Bernstein, S., Leverenz, R., Williams, E., Heinhorst, S., Shively, J., Cannon, G., and Kerfeld, C. (2015). Advances in Understanding Carboxysome Assembly in Prochlorococcus and Synechococcus Implicate CsoS2 as a Critical Component. *Life*, 5(2):1141–1171. 1.5.1.2

Cai, F., Menon, B. B., Cannon, G. C., Curry, K. J., Shively, J. M., and Heinhorst, S. (2009). The Pentameric Vertex Proteins Are Necessary for the Icosahedral Carboxysome Shell to Function as a CO2 Leakage Barrier. *PLOS ONE*, 4(10):e7521. 1.2.3.3, 1.5.1.1

Cai, F., Sutter, M., Cameron, J. C., Stanley, D. N., Kinney, J. N., and Kerfeld, C. A. (2013). The Structure of CcmP, a Tandem Bacterial Microcompartment Domain Protein from the β-Carboxysome, Forms a Subcompartment Within a Microcompartment. *Journal of Biological Chemistry*, 288(22):16055–16063. 1.2.3.3.1

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, 10(1):421. 2.3

Cameron, J. C., Wilson, S. C., Bernstein, S. L., and Kerfeld, C. A. (2013). Biogenesis of a Bacterial Organelle: The Carboxysome Assembly Pathway. *Cell*, 155(5):1131–1140. 1.2.3.3.1, 1.5.1.1, 1.5.1.1, 1.26

Carpentier, R., editor (2011). *Photosynthesis Research Protocols*. Number v. 684 in Methods in Molecular Biology, 1064-3745. Humana Press, New York, 2nd ed edition. 2.10.1, 2.10.2

Catanzariti, A.-M., Soboleva, T. A., Jans, D. A., Board, P. G., and Baker, R. T. (2004). An efficient system for high-level expression and easy purification of authentic recombinant proteins. *Protein Science*, 13(5):1331–1339. 2.1.3, 2.5.1, 2.5.2, A.1.1

Cegelski, L. and Schaefer, J. (2006). NMR determination of photorespiration in intact leaves using in vivo 13CO2 labeling. *Journal of Magnetic Resonance*, 178(1):1–10. 1.2.2

Chari, A., Haselbach, D., Kirves, J.-M., Ohmer, J., Paknia, E., Fischer, N., Ganichkin, O., Möller, V., Frye, J. J., Petzold, G., Jarvis, M., Tietzel, M., Grimm, C., Peters, J.-M., Schulman, B. A., Tittmann, K., Markl, J., Fischer, U., and Stark, H. (2015). ProteoPlex: Stability optimization of macromolecular complexes by sparse-matrix screening of chemical space. *Nature Methods*, 12(9):859–865. 2.5, A.1.1.1.1

Charif, D. and Lobry, J. R. (2007). SeqinR 1.0-2: A Contributed Package to the R Project for Statistical Computing Devoted to Biological Sequences Retrieval and Analysis. In Bastolla, U., Porto, M., Roman, H. E., and Vendruscolo, M., editors, *Structural Approaches to Sequence Evolution: Molecules, Networks, Populations*, Biological and Medical Physics, Biomedical Engineering, pages 207–232. Springer, Berlin, Heidelberg. 2.3

Chen, A. H., Robinson-Mosher, A., Savage, D. F., Silver, P. A., and Polka, J. K. (2013). The Bacterial Carbon-Fixing Organelle Is Formed by Shell Envelopment of Preassembled Cargo. *PLoS ONE*, 8(9):e76127. 4.6

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):1–4. 2.6, A.17

Claassens, N. J., Scarinci, G., Fischer, A., Flamholz, A. I., Newell, W., Frielingsdorf, S., Lenz, O., and Bar-Even, A. (2020). Phosphoglycolate salvage in a chemolithoautotroph using the Calvin cycle. *Proceedings of the National Academy of Sciences of the United States of America*, 117(36):22452–22461. 1.2.2, 1.2.2

Conlan, B., Birch, R., Kelso, C., Holland, S., De Souza, A. P., Long, S. P., Beck, J. L., and Whitney, S. M. (2019). BSD2 is a Rubisco-specific assembly chaperone, forms intermediary hetero-oligomeric complexes, and is nonlimiting to growth in tobacco. *Plant, Cell & Environment*, 42(4):1287–1301. 1.3.2.3.2

Cooney, I., Han, H., Stewart, M. G., Carson, R. H., Hansen, D. T., Iwasa, J. H., Price, J. C., Hill, C. P., and Shen, P. S. (2019). Structure of the Cdc48 segregase in the act of unfolding an authentic substrate. *Science*, 365(6452):502–505. 1.4.3, 1.4.4.1

Cot, S. S.-W., So, A. K.-C., and Espie, G. S. (2008). A Multiprotein Bicarbonate Dehydration Complex Essential to Carboxysome Function in Cyanobacteria. *Journal of Bacteriology*, 190(3):936–945. 1.2.3.3.1, 1.5.1.1

Cox, N., Pantazis, D. A., Neese, F., and Lubitz, W. (2013). Biological Water Oxidation. *Accounts of Chemical Research*, 46(7):1588–1596. 1.2.1

Crenim (2008). English: Photorespiration in C3-plants. How plants cope with the oxygenase reaction of RuBisCO; CC BY-SA 3.0 <http://creativecommons.org/licenses/by-sa/3.0/>, via Wikimedia Commons. 1.6

de Araujo, C., Arefeen, D., Tadesse, Y., Long, B. M., Price, G. D., Rowlett, R. S., Kimber, M. S., and Espie, G. S. (2014). Identification and characterization of a carboxysomal γ-carbonic anhydrase from the cyanobacterium Nostoc sp. PCC 7120. *Photosynthesis Research*, 121(2):135–150. 1.2.3.3.1

de la Peña, A. H., Goodall, E. A., Gates, S. N., Lander, G. C., and Martin, A. (2018). Substrate-engaged 26S proteasome structures reveal mechanisms for ATP-hydrolysis–driven translocation. *Science*, 362(6418). 1.4.3, 1.4.4.2, 4.5

Deville, C., Franke, K., Mogk, A., Bukau, B., and Saibil, H. R. (2019). Two-Step Activation Mechanism of the ClpB Disaggregase for Sequential Substrate Threading by the Main ATPase Motor. *Cell Reports*, 27(12):3433–3446.e4. 1.4.3, 4.5

Dong, Y., Zhang, S., Wu, Z., Li, X., Wang, W. L., Zhu, Y., Stoilova-McPhie, S., Lu, Y., Finley, D., and Mao, Y. (2019). Cryo-EM structures and dynamics of substrate-engaged human 26S proteasome. *Nature*, 565(7737):49–55. 1.4.4.2, 3.1.3

Doron, L., Segal, N., Gibori, H., and Shapira, M. (2014). The BSD2 ortholog in Chlamydomonas reinhardtii is a polysome-associated chaperone that co-migrates on sucrose gradients with the rbcL transcript encoding the Rubisco large subunit. *The Plant Journal*, 80(2):345–355. 1.3.2.3.2

Dou, Z., Heinhorst, S., Williams, E. B., Murin, C. D., Shively, J. M., and Cannon, G. C. (2008). CO2 Fixation Kinetics of Halothiobacillus neapolitanus Mutant Carboxysomes Lacking Carbonic Anhydrase Suggest the Shell Acts as a Diffusional Barrier for CO2*. *Journal of Biological Chemistry*, 283(16):10377–10384. 1.2.3.3

Duff, A. P., Andrews, T., and Curmi, P. M. (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(5):903–916. 1.3.1, 1.3.1.1, 1.3.3.2, 4.2

Eckardt, N. A., Snyder, G. W., Portis, A. R., and Orgen, W. L. (1997). Growth and photosynthesis under high and low irradiance of Arabidopsis thaliana antisense mutants with reduced ribulose-1,5-bisphosphate carboxylase/oxygenase activase content. *Plant Physiology*, 113(2):575–586. 1.4.6.1

Edmondson, D. L., Kane, H. J., and Andrews, T. J. (1990). Substrate isomerization inhibits ribulosebisphospate carboxylase-oxygenase during catalysis. *FEBS Letters*, 260(1):62–66. 1.3.3.4

Eisenhut, M., Ruth, W., Haimovich, M., Bauwe, H., Kaplan, A., and Hagemann, M. (2008). The photorespiratory glycolate metabolism is essential for cyanobacteria and might have been conveyed endosymbiontically to plants. *Proceedings of the National Academy of Sciences*, 105(44):17199–17204. 1.2.2

Ellis, R. J. (2001). Macromolecular crowding: An important but neglected aspect of the intracellular environment. *Current Opinion in Structural Biology*, 11(1):114–119. 1.1.2.1

Ellis, R. J. and Van Der Vies, S. M. (1988). The Rubisco subunit binding protein. *Photosynthesis Research*, 16(1-2):101–115. 1.1.2.3, 1.3.2.1

Emlyn-Jones, D., Woodger, F. J., Price, G. D., and Whitney, S. M. (2006). RbcX Can Function as a Rubisco Chaperonin, But is Non-Essential in Synechococcus PCC7942. *Plant and Cell Physiology*, 47(12):1630–1640. 1.3.2.2, 1.3.2.2.1, 4.1, A.1.1.2.1

Enemark, E. J. and Joshua-Tor, L. (2006). Mechanism of DNA translocation in a replicative hexameric helicase. *Nature*, 442(7100):270–275. 1.4.2.3

Engel, B. D., Schaffer, M., Kuhn Cuellar, L., Villa, E., Plitzko, J. M., and Baumeister, W. (2015). Native architecture of the Chlamydomonas chloroplast revealed by in situ cryo-electron tomography. *eLife*, 4. 1.2.3.2

Erzberger, J. P. and Berger, J. M. (2006). EVOLUTIONARY RELATIONSHIPS AND STRUCTURAL MECHANISMS OF AAA+ PROTEINS. *Annual Review of Biophysics and Biomolecular Structure*, 35(1):93–114. 1.4.1, 1.4.2.2

Esau, B. D., Snyder, G. W., and Portis, A. R. (1996). Differential effects of N- and C-terminal deletions on the two activities of rubisco activase. *Archives of Biochemistry and Biophysics*, 326(1):100–105. 1.4.6.1, 4.2

Fei, X., Bell, T. A., Jenni, S., Stinson, B. M., Baker, T. A., Harrison, S. C., and Sauer, R. T. (2020). Structures of the ATP-fueled ClpXP proteolytic machine bound to protein substrate. *eLife*, 9. 1.4.4.2, 3.1.2, 4.5

Feiz, L., Williams-Carrier, R., Belcher, S., Montano, M., Barkan, A., and Stern, D. B. (2014). A protein with an inactive pterin-4a-carbinolamine dehydratase domain is required for Rubisco biogenesis in plants. *The Plant Journal: For Cell and Molecular Biology*, 80(5):862–869. 1.3.2.3.1

Feiz, L., Williams-Carrier, R., Wostrikoff, K., Belcher, S., Barkan, A., and Stern, D. B. (2012). Ribulose-1,5-Bis-Phosphate Carboxylase/Oxygenase Accumulation Factor1 Is Required for Holoenzyme Assembly in Maize. *The Plant Cell*, 24(8):3435–3446. 1.3.2.2.2, 1.3.2.3.2

Feng, Z., Chen, X., Wu, X., and Zhang, M. (2019). Formation of biological condensates via phase separation: Characteristics, analytical methods, and physiological implications. *The Journal of Biological Chemistry*, 294(40):14823–14835. 1.5.1

Field, C. B., Behrenfeld, M. J., Randerson, J. T., and Falkowski, P. (1998). Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components. *Science*, 281(5374):237–240. 1.2

Fischer, W. W., Hemp, J., and Johnson, J. E. (2016). Evolution of Oxygenic Photosynthesis. *Annual Review of Earth and Planetary Sciences*, 44(1):647–683. 1.2, 1.2.1, 1.2.1

Fitch, W. M. (1966). An improved method of testing for evolutionary homology. *Journal of Molecular Biology*, 16(1):9–16. 3.19

Flecken, M., Wang, H., Popilka, L., Hartl, F. U., Bracher, A., and Hayer-Hartl, M. (2020). Dual Functions of a Rubisco Activase in Metabolic Repair and Recruitment to Carboxysomes. *Cell*, 183(2):457–473.e20. 2.3, 2.7, 2.8, 2.9, 3.1.2, 3.6, 3.7, 3.8, 3.11, 3.12, 3.13, 3.14, 3.15, 3.20, 3.23, 3.26, 4.2, 4.3, 4.5

Foit, L., Morgan, G. J., Kern, M. J., Steimer, L. R., von Hacht, A. A., Titchmarsh, J., Warriner, S. L., Radford, S. E., and Bardwell, J. C. A. (2009). Optimizing Protein Stability In Vivo. *Molecular Cell*, 36(5):861–871. 1.1.1

Freeman Rosenzweig, E. S., Xu, B., Kuhn Cuellar, L., Martinez-Sanchez, A., Schaffer, M., Strauss, M., Cartwright, H. N., Ronceray, P., Plitzko, J. M., Förster, F., Wingreen, N. S., Engel, B. D., Mackinder, L. C., and Jonikas, M. C. (2017). The Eukaryotic CO2-Concentrating Organelle Is Liquid-like and Exhibits Dynamic Reorganization. *Cell*, 171(1):148–162.e19. 1.2.3.2

Fuchs, G. (2011). Alternative Pathways of Carbon Dioxide Fixation: Insights into the Early Evolution of Life? *Annual Review of Microbiology*, 65(1):631–658. 1.2

Fujita, K., Tanaka, K., Sadaie, Y., and Ohta, N. (2008). Functional analysis of the plastid and nuclear encoded CbbX proteins of Cyanidioschyzon merolae. *Genes & Genetic Systems*, 83(2):135–142. 1.4.6.2

Gamerdinger, M., Hanebuth, M. A., Frickey, T., and Deuerling, E. (2015). The principle of antagonism ensures protein targeting specificity at the endoplasmic reticulum. *Science*, 348(6231):201–207. 1.1.2.2

Gao, K., Oerlemans, R., and Groves, M. R. (2020). Theory and applications of differential scanning fluorimetry in early-stage drug discovery. *Biophysical Reviews*, 12(1):85–104. A.1.1.1.1

Gardner, B. M., Castanzo, D. T., Chowdhury, S., Stjepanovic, G., Stefely, M. S., Hurley, J. H., Lander, G. C., and Martin, A. (2018). The peroxisomal AAA-ATPase Pex1/Pex6 unfolds substrates by processive threading. *Nature Communications*, 9. 1.4.4.1

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., and Bairoch, A. (2005). Protein Identification and Analysis Tools on the ExPASy Server. In Walker, J. M., editor, *The Proteomics Protocols Handbook*, pages 571–607. Humana Press, Totowa, NJ. 2.1.4.1, 2.2

Gates, S. N., Yokom, A. L., Lin, J., Jackrel, M. E., Rizo, A. N., Kendsersky, N. M., Buell, C. E., Sweeny, E. A., Mack, K. L., Chuang, E., Torrente, M. P., Su, M., Shorter, J., and Southworth, D. R. (2017). Ratchet-like polypeptide translocation mechanism of the AAA+ disaggregase Hsp104. *Science*, 357(6348):273–279. 4.5

Georgescauld, F., Popova, K., Gupta, A. J., Bracher, A., Engen, J. R., Hayer-Hartl, M., and Hartl, F. U. (2014). GroEL/ES Chaperonin Modulates the Mechanism and Accelerates the Rate of TIM-Barrel Domain Folding. *Cell*, 157(4):922–934. 1.1.2.3

Germany, E. M., Zahayko, N., Huebsch, M. L., Fox, J. L., Prahlad, V., and Khalimonchuk, O. (2018). The AAA ATPase Afg1 preserves mitochondrial fidelity and cellular health by maintaining mitochondrial matrix proteostasis. *Journal of Cell Science*, 131(22). 1.4.3

Ghisaidoobe, A. B. T. and Chung, S. J. (2014). Intrinsic Tryptophan Fluorescence in the Detection and Analysis of Proteins: A Focus on Förster Resonance Energy Transfer Techniques. *International Journal of Molecular Sciences*, 15(12):22518–22538. A.1.1.1.1

Gibson, J. L. and Tabita, F. R. (1997). Analysis of the cbbXYZ operon in Rhodobacter sphaeroides. *Journal of Bacteriology*, 179(3):663–669. 1.4.6.2

Goloubinoff, P., Christeller, J. T., Gatenby, A. A., and Lorimer, G. H. (1989a). Reconstitution of active dimeric ribulose bisphosphate carboxylase from an unfolded state depends on two chaperonin proteins and Mg-ATP. *Nature*, 342(6252):884–889. 1.1.2.3, 1.3.2.1, 4.1, A.1.1.2.1

Goloubinoff, P., Gatenby, A., and Lorimer, G. (1989b). GroE heat-shock proteins promote assembly of foreign prokaryotic ribulose bisphosphate carboxylase oligomers in Escherichia coli. *Nature*, 377:44–47. 1.3.1.2, 1.3.2.2, 1.3.2.2.1, 4.1, A.1.1.2.1

Gouet, P., Courcelle, E., Stuart, D., and Metoz, F. (1999). ESPript: Analysis of multiple sequence alignments in PostScript. *Bioinformatics*, 15(4):305–308. 2.3

Gruber, A. V., Nisemblat, S., Azem, A., and Weiss, C. (2013). The complexity of chloroplast chaperonins. *Trends in Plant Science*, 18(12):688–694. 1.3.2.3

Gupta, A. J., Haldar, S., Miličić, G., Hartl, F. U., and Hayer-Hartl, M. (2014). Active Cage Mechanism of Chaperonin-Assisted Protein Folding Demonstrated at Single-Molecule Level. *Journal of Molecular Biology*, 426(15):2739–2754. 1.1.2.3

Han, H., Monroe, N., Sundquist, W. I., Shen, P. S., and Hill, C. P. (2017). The AAA ATPase Vps4 binds ESCRT-III substrates through a repeating array of dipeptide-binding pockets. *eLife*, 6. 1.4.3, 1.4.4.2, 3.1.4, 4.5

Hanahan, D. (1983). Studies on transformation of Escherichia coli with plasmids. *Journal of Molecular Biology*, 166(4):557–580. 2.1.1

Hansen, S., Vollan, V. B., Hough, E., and Andersen, K. (1999). The crystal structure of rubisco from Alcaligenes eutrophus reveals a novel central eight-stranded β-barrel formed by β-strands from four subunits11Edited by R. Huber. *Journal of Molecular Biology*, 288(4):609–621. 1.3.1.2

Hanson, P. I. and Whiteheart, S. W. (2005). AAA+ proteins: Have engine, will work. *Nature Reviews Molecular Cell Biology*, 6(7):519–529. 1.4.1, 1.4.2.3

Hanson, T. E. and Tabita, F. R. (2001). A ribulose-1,5-bisphosphate carboxylase/oxygenase (RubisCO)-like protein from Chlorobium tepidum that is involved with sulfur metabolism and the response to oxidative stress. *Proceedings of the National Academy of Sciences of the United States of America*, 98(8):4397–4402. 1.3

Hartl, F. U. (1996). Molecular chaperones in cellular protein folding. *Nature*, 381(6583):571–580. 1.1.2.2

Hartl, F. U., Bracher, A., and Hayer-Hartl, M. (2011). Molecular chaperones in protein folding and proteostasis. *Nature*, 475(7356):324–332. 1.1.2.2

Hasse, D., Larsson, A. M., and Andersson, I. (2015). Structure of Arabidopsis thaliana Rubisco activase. *Acta Crystallographica Section D: Biological Crystallography*, 71(4):800–808. 1.4.6.1, 1.4.6.2, 3.6, 4.5

Hauser, T., Bhat, J. Y., Miličić, G., Wendler, P., Hartl, F. U., Bracher, A., and Hayer-Hartl, M. (2015). Structure and mechanism of the Rubisco-assembly chaperone Raf1. *Nature Structural & Molecular Biology*, 22(9):720–728. 1.3.2.2, 1.3.2.2.2, 4.1, A.1.1.2.1

Hayer-Hartl, M., Bracher, A., and Hartl, F. U. (2016). The GroEL–GroES Chaperonin Machine: A Nano-Cage for Protein Folding. *Trends in Biochemical Sciences*, 41(1):62–76. 1.1.2.3, 1.3, 1.1.2.3

Hazra, S., Henderson, J. N., Liles, K., Hilton, M. T., and Wachter, R. M. (2015). Regulation of Ribulose-1,5-bisphosphate Carboxylase/Oxygenase (Rubisco) Activase. *The Journal of Biological Chemistry*, 290(40):24222–24236. 1.4.6.1, 3.1.1

He, S., Chou, H.-T., Matthies, D., Wunder, T., Meyer, M. T., Atkinson, N., Martinez-Sanchez, A., Jeffrey, P. D., Port, S. A., Patena, W., He, G., Chen, V. K., Hughson, F. M., McCormick, A. J., Mueller-Cajar, O., Engel, B. D., Yu, Z., and Jonikas, M. C. (2020). The structural basis of Rubisco phase separation in the pyrenoid. *Nature Plants*, 6(12):1480–1490. 1.5.1.2

Henderson, J. N., Kuriata, A. M., Fromme, R., Salvucci, M. E., and Wachter, R. M. (2011). Atomic Resolution X-ray Structure of the Substrate Recognition Domain of Higher Plant Ribulose-bisphosphate Carboxylase/Oxygenase (Rubisco) Activase. *Journal of Biological Chemistry*, 286(41):35683–35688. 1.4.6.1

Hennacy, J. H. and Jonikas, M. C. (2020). Prospects for Engineering Biophysical CO2 Concentrating Mechanisms into Land Plants to Enhance Yields. *Annual Review of Plant Biology*, 71(1):461–485. 1.2.3.2

Heo, J. and Holbrook, G. P. (1999). Regulation of 2-carboxy-D-arabinitol 1-phosphate phosphatase : Activation by glutathione and interaction with thiol reagents. *Biochemistry Journal*, page 8. 1.3.3.4

Hoffmann, A., Bukau, B., and Kramer, G. (2010). Structure and function of the molecular chaperone Trigger Factor. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1803(6):650–661. 1.1.2.2

Hohmann-Marriott, M. F. and Blankenship, R. E. (2011). Evolution of photosynthesis. *Annual Review of Plant Biology*, 62:515–548. 1.2.1

Holbrook, G. P., Bowes, G., and Salvucci, M. E. (1989). Degradation of 2-Carboxyarabinitol 1-Phosphate by a Specific Chloroplast Phosphatase 1. *Plant Physiology*, 90(2):673–678. 1.3.3.4

Houry, W. A., Frishman, D., Eckerskorn, C., Lottspeich, F., and Hartl, F. U. (1999). Identifcation of in vivo substrates of the chaperonin GroEL. 402:8. 1.1.2.3

Houtz, R. L., Poneleit, L., Jones, S. B., Royer, M., and Stults, J. T. (1992). Posttranslational Modifications in the Amino- Terminal Region of the Large Subunit of Ribulose- 1,5-Bisphosphate Carboxylase/Oxygenase from Several Plant Species. *Plant Physiology*, 98(3):1170–1174. 4.3

Houtz, R. L., Stults, J. T., Mulligan, R. M., and Tolbert, N. E. (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(6):1855–1859. 4.3

Huang, F., Kong, W.-W., Sun, Y., Chen, T., Dykes, G. F., Jiang, Y.-L., and Liu, L.-N. (2020). Rubisco accumulation factor 1 (Raf1) plays essential roles in mediating Rubisco assembly and carboxysome biogenesis. *Proceedings of the National Academy of Sciences of the United States of America*, 117(29):17418–17428. 1.3.2.2.1, 1.3.2.2.2, 4.1, 4.6

Huang, F., Vasieva, O., Sun, Y., Faulkner, M., Dykes, G. F., Zhao, Z., and Liu, L.-N. (2019). Roles of RbcX in Carboxysome Biosynthesis in the Cyanobacterium Synechococcus elongatus PCC79421. *Plant Physiology*, 179(1):184–194. 1.3.2.2.1, 4.1, A.1.1.2.1

Hügler, M. and Sievert, S. M. (2010). Beyond the Calvin Cycle: Autotrophic Carbon Fixation in the Ocean. *Annual Review of Marine Science*, 3(1):261–289. 1.2, 1.2.2

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(11):2273–2280. 4.1

Hyman, A. A. and Simons, K. (2012). Beyond Oil and Water—Phase Transitions in Cells. *Science*, 337(6098):1047–1049. 1.5.1

Hyman, A. A., Weber, C. A., and Jülicher, F. (2014). Liquid-Liquid Phase Separation in Biology. *Annual Review of Cell and Developmental Biology*, 30(1):39–58. 1.5, 1.5.1

Iancu, C. V., Morris, D. M., Dou, Z., Heinhorst, S., Cannon, G. C., and Jensen, G. J. (2010). Organization, Structure, and Assembly of α-Carboxysomes Determined by Electron Cryotomography of Intact Cells. *Journal of Molecular Biology*, 396(1):105–117. 1.2.3.3.1

Imker, H. J., Fedorov, A. A., Fedorov, E. V., Almo, S. C., and Gerlt, J. A. (2007). Mechanistic diversity in the RuBisCO superfamily: The "enolase" in the methionine salvage pathway in Geobacillus kaustophilus. *Biochemistry*, 46(13):4077–4089. 1.3.1.2

Itakura, A. K., Chan, K. X., Atkinson, N., Pallesen, L., Wang, L., Reeves, G., Patena, W., Caspari, O., Roth, R., Goodenough, U., McCormick, A. J., Griffiths, H., and Jonikas, M. C. (2019). A Rubisco-binding protein is required for normal pyrenoid number and starch sheath morphology in *Chlamydomonas reinhardtii*. *Proceedings of the National Academy of Sciences*, 116(37):18445–18454. 1.2.3.2

Iyer, L. M., Leipe, D. D., Koonin, E. V., and Aravind, L. (2004). Evolutionary history and higher order classification of AAA+ ATPases. *Journal of Structural Biology*, 146(1):11–31. 1.4.1, 1.4.6.3

Jackson, S. E. (2013). Hsp90: Structure and function. *Topics in Current Chemistry*, 328:155–240. 1.1.2.2

Jarvis, P. and Kessler, F. (2014). Mechanisms of Chloroplast Protein Import in Plants. In Theg, S. M. and Wollman, F.-A., editors, *Plastid Biology*, Advances in Plant Biology, pages 241–270. Springer, New York, NY. 1.15

Jones, M. (2010). English: Overview of the Calvin cycle pathway. CC BY-SA 3.0 <https://creativecommons.org/licenses/by-sa/3.0>, via Wikimedia Commons. 1.5

Jordan, D. B. and Chollet, R. (1983). Inhibition of ribulose bisphosphate carboxylase by substrate ribulose 1,5-bisphosphate. *Journal of Biological Chemistry*, 258(22):13752–13758. 1.3.3.4

Jordan, D. B. and Ogren, W. L. (1983). Species variation in kinetic properties of ribulose 1,5-bisphosphate carboxylase/oxygenase. *Archives of Biochemistry and Biophysics*, 227(2):425–433. 1.3

Joshi, J., Mueller-Cajar, O., Tsai, Y.-C. C., Hartl, F. U., and Hayer-Hartl, M. (2015). Role of Small Subunit in Mediating Assembly of Red-type Form I Rubisco. *Journal of Biological Chemistry*, 290(2):1066–1074. 1.3.1.2

Kaiser, C. M., Goldman, D. H., Chodera, J. D., Tinoco, I., and Bustamante, C. (2011). The Ribosome Modulates Nascent Protein Folding. *Science*, 334(6063):1723–1727. 1.1.2.1

Kane, H. J., Wilkin, J.-M., Portis, A. R., and John Andrews, T. (1998). Potent Inhibition of Ribulose-Bisphosphate Carboxylase by an Oxidized Impurity in Ribulose-1,5-Bisphosphate. *Plant Physiology*, 117(3):1059–1069. 1.3.3.4

Kaneko, T., Nakamura, Y., Wolk, C. P., Kuritz, T., Sasamoto, S., Watanabe, A., Iriguchi, M., Ishikawa, A., Kawashima, K., Kimura, T., Kishida, Y., Kohara, M., Matsumoto, M., Matsuno, A., Muraki, A., Nakazaki, N., Shimpo, S., Sugimoto, M., Takazawa, M., Yamada, M., Yasuda, M., and Tabata, S. (2001). Complete genomic sequence of the filamentous nitrogen-fixing cyanobacterium Anabaena sp. strain PCC 7120. *DNA research: an international journal for rapid publication of reports on genes and genomes*, 8(5):205–213; 227–253. A.1.1.2.1

Keown, J. R., Griffin, M. D. W., Mertens, H. D. T., and Pearce, F. G. (2013). Small Oligomers of Ribulose-bisphosphate Carboxylase/Oxygenase (Rubisco) Activase Are Required for Biological Activity. *The Journal of Biological Chemistry*, 288(28):20607–20615. 1.4.6.1

Keown, J. R. and Pearce, F. G. (2014). Characterization of spinach ribulose-1,5-bisphosphate carboxylase/oxygenase activase isoforms reveals hexameric assemblies with increased thermal stability. *The Biochemical Journal*, 464(3):413–423. 1.4.6.1, 3.2.5

Kerfeld, C. A. and Melnicki, M. R. (2016). Assembly, function and evolution of cyanobacterial carboxysomes. *Current Opinion in Plant Biology*, 31:66–75. 1.2.3.3.1

Kerner, M. J., Naylor, D. J., Ishihama, Y., Maier, T., Chang, H.-C., Stines, A. P., Georgopoulos, C., Frishman, D., Hayer-Hartl, M., Mann, M., and Hartl, F. U. (2005). Proteome-wide Analysis of Chaperonin-Dependent Protein Folding in Escherichia coli. *Cell*, 122(2):209–220. 1.1.2.3, 1.3.2.1, 2.1.3, A.1.1.2.1

Khalifah, G. (1971). I. STOP-FLOW KINETIC STUDIES ON THE NATIVE HUMAN ISOENZYMES B AND C". *THE JOURNAL OF BIOLOGICAL CHEMISTRY*, 246(8):2561–2573. 1.3

Khan, S., Andralojc, P. J., Lea, P. J., and Parry, M. A. J. (1999). 2′-Carboxy-D-arabitinol 1-phosphate protects ribulose 1,5-bisphosphate carboxylase/oxygenase against proteolytic breakdown. *European Journal of Biochemistry*, 266(3):840–847. 1.3.3.4

Khrebtukova, I. and Spreitzer, R. J. (1996). Elimination of the Chlamydomonas gene family that encodes the small subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase. *Proceedings of the National Academy of Sciences of the United States of America*, 93(24):13689–13693. 1.3.2.3

Kim, D. E., Gu, H., and Baker, D. (1998). The sequences of small proteins are not extensively optimized for rapid folding by natural selection. *Proceedings of the National Academy of Sciences of the United States of America*, 95(9):4982–4986. 1.1.1

Kim, K. and Portis, A. R. (2004). Oxygen-dependent H2O2 production by Rubisco. *FEBS Letters*, 571(1-3):124–128. 1.3.3.4

Kinney, J. N., Salmeen, A., Cai, F., and Kerfeld, C. A. (2012). Elucidating Essential Role of Conserved Carboxysomal Protein CcmN Reveals Common Feature of Bacterial Microcompartment Assembly *. *Journal of Biological Chemistry*, 287(21):17729–17736. 1.5.1.1

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(6):473–481. 1.9

Kleczkowski, L. A. (1994). Inhibitors of Photosynthetic Enzymes/Carriers and Metabolism. *Annu. Rev. Plant Physiol. Plant Mol. Biol.*, 45:339–367. 1.2.2

Klein, M. G., Zwart, P., Bagby, S. C., Cai, F., Chisholm, S. W., Heinhorst, S., Cannon, G. C., and Kerfeld, C. A. (2009). Identification and Structural Analysis of a Novel Carboxysome Shell Protein with Implications for Metabolite Transport. *Journal of Molecular Biology*, 392(2):319–333. 1.2.3.3.1

Knight, S., Andersson, I., and Brändén, C.-I. (1990). Crystallographic analysis of ribulose 1,5-bisphosphate carboxylase from spinach at 2·4 Å resolution: Subunit interactions and active site. *Journal of Molecular Biology*, 215(1):113–160. 1.3.1.1, 1.3.1.2

Kok, B., Forbush, B., and McGLOIN, M. (1970). Cooperation of Charges in Photosynthetic O2 Evolution–I. a Linear Four Step Mechanism. *Photochemistry and Photobiology*, 11(6):457–475. 1.2.1

Kolesinski, P., Belusiak, I., Czarnocki-Cieciura, M., and Szczepaniak, A. (2014). Rubisco Accumulation Factor 1 from Thermosynechococcus elongatus participates in the final stages of ribulose-1,5-bisphosphate carboxylase/oxygenase assembly in Escherichia coli cells and in vitro. *The FEBS Journal*, 281(17):3920–3932. 1.3.2.2, 1.3.2.2.2, 4.1, A.1.1.2.1

Kolesiński, P., Piechota, J., and Szczepaniak, A. (2011). Initial characteristics of RbcX proteins from Arabidopsis thaliana. *Plant Molecular Biology*, 77(4-5):447–459. 1.3.2.3

Koonin, E. V., Wolf, Y. I., and Aravind, L. (2000). Protein fold recognition using sequence profiles and its application in structural genomics. In *Advances in Protein Chemistry*, volume 54 of *Analysis of Amino Acid Sequences*, pages 245–275. Academic Press. 1.4.1

Kopec, J. and Schneider, G. (2011). Comparison of fluorescence and light scattering based methods to assess formation and stability of protein–protein complexes. *Journal of Structural Biology*, 175(2):216–223. A.1.1.1.1

Kosolapov, A. and Deutsch, C. (2009). Tertiary interactions within the ribosomal exit tunnel. *Nature Structural & Molecular Biology*, 16(4):405–411. 1.1.2.1

Kotaja, N. and Sassone-Corsi, P. (2007). The chromatoid body: A germ-cell-specific RNA-processing centre. *Nature Reviews. Molecular Cell Biology*, 8(1):85–90. 1.5.1

Kovács, E., van der Vies, S. M., Glatz, A., Török, Z., Varvasovszki, V., Horváth, I., and Vígh, L. (2001). The Chaperonins of Synechocystis PCC 6803 Differ in Heat Inducibility and Chaperone Activity. *Biochemical and Biophysical Research Communications*, 289(4):908–915. A.1.1.2.1

Kozaki, A. and Takeba, G. (1996). Photorespiration protects C3 plants from photooxidation. *Nature*, 384(6609):557–560. 1.2.2

Kress, W., Mutschler, H., and Weber-Ban, E. (2009). Both ATPase Domains of ClpA Are Critical for Processing of Stable Protein Structures. *Journal of Biological Chemistry*, 284(45):31441–31452. 1.4.2.1

Kusian, B. and Bowien, B. (1997). Organization and regulation of cbb CO2 assimilation genes in autotrophic bacteria. *FEMS Microbiology Reviews*, 21(2):135–155. 1.3

Larimer, F. W. and Soper, T. S. (1993). Overproduction of Anabaena 7120 ribulose-bisphosphate carboxylase/oxygenase in Escherichia coli. *Gene*, 126(1):85–92. 1.3.2.2.1, 2.5.1, 2.5.3

Larson, E. M., O'Brien, C. M., Zhu, G., Spreitzer, R. J., and Portis, A. R. (1997). Specificity for Activase Is Changed by a Pro-89 to Arg Substitution in the Large Subunit of Ribulose-1,5-bisphosphate Carboxylase/Oxygenase. *Journal of Biological Chemistry*, 272(27):17033–17037. 1.4.6.1

Lechno-Yossef, S., Rohnke, B. A., Belza, A. C. O., Melnicki, M. R., Montgomery, B. L., and Kerfeld, C. A. (2020). Cyanobacterial carboxysomes contain an unique rubisco-activase-like protein. *New Phytologist*, 225(2):793–806. 3.2, 3.2.5, 4.6, 4.6

Leipe, D. D., Koonin, E. V., and Aravind, L. (2003). Evolution and classification of P-loop kinases and related proteins. *Journal of Molecular Biology*, 333(4):781–815. 1.4.1

Leipe, D. D., Wolf, Y. I., Koonin, E. V., and Aravind, L. (2002). Classification and evolution of P-loop GTPases and related ATPases11Edited by J. Thornton. *Journal of Molecular Biology*, 317(1):41–72. 1.4.1

Levinthal, C. (1968). Are there pathways for protein folding? *Journal de Chimie Physique*, 65:44–45. 1.1.1

Levinthal, C. (1969). How to fold graciously. In *Mossbauer Spectroscopy in Biological Systems: Proceedings of a Meeting Held at Allerton House, Monticello, Illinois*, pages 22–24, Allerton House, Monticello, Illinois. University of Illinois Press. 1.1.1

Li, C., Salvucci, M. E., and Portis, A. R. (2005a). Two Residues of Rubisco Activase Involved in Recognition of the Rubisco Substrate. *Journal of Biological Chemistry*, 280(26):24864–24869. 1.4.6.1, 3.1.6, 4.4

Li, C., Wang, D., and Portis, A. R. (2006). Identification of critical arginine residues in the functioning of Rubisco activase. *Archives of Biochemistry and Biophysics*, 450(2):176–182. 4.5

Li, H., Sawaya, M. R., Tabita, F. R., and Eisenberg, D. (2005b). Crystal Structure of a RuBisCO-like Protein from the Green Sulfur Bacterium Chlorobium tepidum. *Structure*, 13(5):779–789. 1.9

Li, L. A. and Tabita, F. R. (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(11):3793–3796. 1.3.2.2, 1.3.2.2.1, 1.6, 4.1, A.1.1.2.1

Li, L.-A., Zianni, M. R., and Tabita, F. R. (1999). Inactivation of the monocistronic rca gene in Anabaena variabilis suggests a physiological ribulose bisphosphate carboxylase/oxygenase activase-like function in heterocystous cyanobacteria. *Plant Molecular Biology*, 40:467–478. 1.6, 3.1.1

Li, N., Zhai, Y., Zhang, Y., Li, W., Yang, M., Lei, J., Tye, B.-K., and Gao, N. (2015). Structure of the eukaryotic MCM complex at 3.8 Å. *Nature*, 524(7564):186–191. 1.4.2.3

Li, P., Banjade, S., Cheng, H.-C., Kim, S., Chen, B., Guo, L., Llaguno, M., Hollingsworth, J. V., King, D. S., Banani, S. F., Russo, P. S., Jiang, Q.-X., Nixon, B. T., and Rosen, M. K. (2012). Phase Transitions in the Assembly of Multi-Valent Signaling Proteins. *Nature*, 483(7389):336–340. 1.5.1, 4.6

Lin, M. T., Stone, W. D., Chaudhari, V., and Hanson, M. R. (2020). Small subunits can determine enzyme kinetics of tobacco Rubisco expressed in Escherichia coli. *Nature Plants*, 6(10):1289–1299. 2.1.3, 2.1, 2.5.4

Lin, Z., Madan, D., and Rye, H. (2008). GroEL stimulates protein folding through forced unfolding. *Nature Structural & Molecular Biology*, 15(3):303–311. 1.1.2.3

Liu, C., Young, A. L., Starling-Windhof, A., Bracher, A., Saschenbrecker, S., Rao, B. V., Rao, K. V., Berninghausen, O., Mielke, T., Hartl, F. U., Beckmann, R., and Hayer-Hartl, M. (2010). Coupled chaperone action in folding and assembly of hexadecameric Rubisco. *Nature*, 463(7278):197–202. 1.3.2.1, 1.3.2.2.1, 4.1, A.1.1.2.1

Liu, J., Smith, C. L., DeRyckere, D., DeAngelis, K., Martin, G. S., and Berger, J. M. (2000). Structure and Function of Cdc6/Cdc18: Implications for Origin Recognition and Checkpoint Control. *Molecular Cell*, 6:637–648. 1.4.2.1

Liu, Y., He, X., Lim, W., Mueller, J., Lawrie, J., Kramer, L., Guo, J., and Niu, W. (2018). Deciphering molecular details in the assembly of alpha-type carboxysome. *Scientific Reports*, 8(1):15062. 1.5.1.2

Loganathan, N., Tsai, Y.-C. C., and Mueller-Cajar, O. (2016). Characterization of the heterooligomeric red-type rubisco activase from red algae. *Proceedings of the National Academy of Sciences*, 113(49):14019–14024. 1.4.6.2, 4.2

Long, B. M., Badger, M. R., Whitney, S. M., and Price, G. D. (2007). Analysis of carboxysomes from Synechococcus PCC7942 reveals multiple Rubisco complexes with carboxysomal proteins CcmM and CcaA. *The Journal of Biological Chemistry*, 282(40):29323–29335. 1.5.1.1, 2.1.3, 3.19, 3.2.1

Long, B. M., Rae, B. D., Badger, M. R., and Dean Price, G. (2011). Over-expression of the β-carboxysomal CcmM protein in Synechococcus PCC7942 reveals a tight co-regulation of carboxysomal carbonic anhydrase (CcaA) and M58 content. *Photosynthesis Research*, 109(1):33–45. 1.5.1.1

Long, B. M., Tucker, L., Badger, M. R., and Price, G. D. (2010). Functional Cyanobacterial β-Carboxysomes Have an Absolute Requirement for Both Long and Short Forms of the CcmM Protein. *Plant Physiology*, 153(1):285–293. 1.5.1.1, 3.2.1, 4.6

Lopez, T., Dalton, K., and Frydman, J. (2015). The Mechanism and Function of Group II Chaperonins. *Journal of Molecular Biology*, 427(18):2919–2930. 1.1.2.3

Lorimer, G. H. (1981). The Carboxylation and Oxygenation of Ribulose 1,5-Bisphosphate: The Primary Events in Photosynthesis and Photorespiration. *Annual Review of Plant Physiology*, 32(1):349–382. 1.3.3.3

Lorimer, G. H., Badger, M. R., and Andrews, T. J. (1976). The activation of ribulose-1,5-bisphosphate carboxylase by carbon dioxide and magnesium ions. Equilibria, kinetics, a suggested mechanism, and physiological implications. *Biochemistry*, 15(3):529–536. 1.4.6

Luengo, T. M., Mayer, M. P., and Rüdiger, S. G. D. (2019). The Hsp70–Hsp90 Chaperone Cascade in Protein Folding. *Trends in Cell Biology*, 29(2):164–177. 1.1.2.2

Lundqvist, T. and Schneider, G. (1991). Crystal structure of the ternary complex of ribulose-1,5-bisphosphate carboxylase, Mg(II), and activator CO2 at 2.3-A resolution. *Biochemistry*, 30(4):904–908. 1.8

Maillard, R. A., Chistol, G., Sen, M., Righini, M., Tan, J., Kaiser, C. M., Hodges, C., Martin, A., and Bustamante, C. (2011). ClpX(P) Generates Mechanical Force to Unfold and Translocate Its Protein Substrates. *Cell*, 145(3):459–469. 1.4.4.2

Martin, A., Baker, T. A., and Sauer, R. T. (2005). Rebuilt AAA + motors reveal operating principles for ATP-fuelled machines. *Nature*, 437(7062):1115–1120. 1.4.4.2

Martin, L., Schwarz, S., and Breitsprecher, D. (2015). Analyzing Thermal Unfolding of Proteins: The Prometheus NT.48. 2.4

Martin, W. F., Bryant, D. A., and Beatty, J. T. (2018). A physiological perspective on the origin and evolution of photosynthesis. *FEMS Microbiology Reviews*, 42(2):205–231. 1.2, 1.2.1

Mayer, M. P. (2010). Gymnastics of Molecular Chaperones. *Molecular Cell*, 39(3):321–331. 1.1.1

McGurn, L. D., Moazami-Goudarzi, M., White, S. A., Suwal, T., Brar, B., Tang, J. Q., Espie, G. S., and Kimber, M. S. (2016). The structure, kinetics and interactions of the β-carboxysomal β-carbonic anhydrase, CcaA. *Biochemical Journal*, 473(24):4559–4572. 1.2.3.3.1

McLennan, N. and Masters, M. (1998). GroE is vital for cell-wall synthesis. *Nature*, 392(6672):139–139. 2.1, A.8, A.1.1.2.1

Meyer, M. T., Genkov, T., Skepper, J. N., Jouhet, J., Mitchell, M. C., Spreitzer, R. J., and Griffiths, H. (2012). Rubisco small-subunit -helices control pyrenoid formation in Chlamydomonas. *Proceedings of the National Academy of Sciences*, 109(47):19474–19479. 1.2.3.2

Meyer, M. T., Itakura, A. K., Patena, W., Wang, L., He, S., Emrich-Mills, T., Lau, C. S., Yates, G., Mackinder, L. C. M., and Jonikas, M. C. (2020). Assembly of the algal CO2-fixing organelle, the pyrenoid, is guided by a Rubisco-binding motif. *Science Advances*, 6(46). 1.2.3.2

Meyer, M. T., Whittaker, C., and Griffiths, H. (2017). The algal pyrenoid: Key unanswered questions. *Journal of Experimental Botany*, 68(14):3739–3749. 1.2.3.2

Miller, J. M., Arachea, B. T., Epling, L. B., and Enemark, E. J. (2014). Analysis of the crystal structure of an active MCM hexamer. *eLife*. 1.4.2.3

Miller, J. M. and Enemark, E. J. (2016). Fundamental Characteristics of AAA+ Protein Family Structure and Function. *Archaea*, 2016:1–12. 1.4.2.2, 1.4.2.3, 1.4.3, 1.4.6.2, 1.4.6.3

Monroe, N., Han, H., Shen, P. S., Sundquist, W. I., and Hill, C. P. (2017). Structural basis of protein translocation by the Vps4-Vta1 AAA ATPase. *eLife*, 6:e24487. 1.4.4.2

Moore, B. D., Kobza, J., and Seemann, J. R. (1991). Measurement of 2-Carboxyarabinitol 1-Phosphate in Plant Leaves by Isotope Dilution. *Plant Physiology*, 96(1):208–213. 1.3.3.4

Moore, B. D. and Seemann, J. R. (1992). Metabolism of 2′-Carboxyarabinitol in Leaves. *Plant Physiology*, 99(4):1551–1555. 1.3.3.4

Moore, B. D. and Seemann, J. R. (1994). Evidence That 2-Carboxyarabinitol 1-Phosphate Binds to Ribulose-1,5-Bisphosphate Carboxylase in Vivo. *Plant Physiology*, 105(2):731–737. 1.3.3.4

Moore, B. D., Sharkey, T. D., and Seemann, J. R. (1995). Intracellular localization of CA1P and CA1P phosphatase activity in leaves of Phaseolus vulgaris L. *Photosynthesis Research*, 45(3):219–224. 1.3.3.4

Morgner, N., Schmidt, C., Beilsten-Edmands, V., Ebong, I.-o., Patel, N. A., Clerico, E. M., Kirschke, E., Daturpalli, S., Jackson, S. E., Agard, D., and Robinson, C. V. (2015). Hsp70 Forms Antiparallel Dimers Stabilized by Post-translational Modifications to Position Clients for Transfer to Hsp90. *Cell Reports*, 11(5):759–769. 1.1.2.2

Mueller-Cajar, O. (2017). The Diverse AAA+ Machines that Repair Inhibited Rubisco Active Sites. *Frontiers in Molecular Biosciences*, 4:31. 4.2, 4.2

Mueller-Cajar, O., Stotz, M., Wendler, P., Hartl, F. U., Bracher, A., and Hayer-Hartl, M. (2011). Structure and function of the AAA + protein CbbX, a red-type Rubisco activase. *Nature*, 479(7372):194–199. 1.4.6.2, 1.4.6.2, 2.10.1, 2.10.2, 3.1.1, 3.1.3, 4.2, 4.2

Nagiec, E. E., Bernstein, A., and Whiteheart, S. W. (1995). Each Domain of the N-Ethylmaleimide-sensitive Fusion Protein Contributes to Its Transport Activity (*). *Journal of Biological Chemistry*, 270(49):29182–29188. 4.5

Naponelli, V., Noiriel, A., Ziemak, M. J., Beverley, S. M., Lye, L.-F., Plume, A. M., Botella, J. R., Loizeau, K., Ravanel, S., Rébeillé, F., de Crécy-Lagard, V., and Hanson, A. D. (2008). Phylogenomic and Functional Analysis of Pterin-4a-Carbinolamine Dehydratase Family (COG2154) Proteins in Plants and Microorganisms. *Plant Physiology*, 146(4):1515–1527. 1.3.2.3.1

Neuwald, A. F., Aravind, L., Spouge, J. L., and Koonin, E. V. (1999). AAA+: A Class of Chaperone-Like ATPases Associated with the Assembly, Operation, and Disassembly of Protein Complexes. *Genome Research*, 9(1):27–43. 1.4.1, 1.4.6.1

Newman, J., Brändén, C.-I., and Jones, T. A. (1993). Structure determination and refinement of ribulose 1,5-bisphosphate carboxylase/oxygenase from Synechococcus PCC6301. *Acta Crystallographica Section D: Biological Crystallography*, 49(6):548–560. 1.9

Ng, J., Guo, Z., and Mueller-Cajar, O. (2020). Rubisco activase requires residues in the large subunit N terminus to remodel inhibited plant Rubisco. *Journal of Biological Chemistry*, 295:16427. 2.5.4, 4.3

Niesen, F. H., Berglund, H., and Vedadi, M. (2007). The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nature Protocols*, 2(9):2212–2221. A.1.1.1.1

Nisbet, E. G., Grassineau, N. V., Howe, C. J., Abell, P. I., Regelous, M., and Nisbet, R. E. R. (2007). The age of Rubisco: The evolution of oxygenic photosynthesis. *Geobiology*, 5(4):311–335. 1.3

Nott, T. J., Petsalaki, E., Farber, P., Jervis, D., Fussner, E., Plochowietz, A., Craggs, T. D., Bazett-Jones, D. P., Pawson, T., Forman-Kay, J. D., and Baldwin, A. J. (2015). Phase Transition of a Disordered Nuage Protein Generates Environmentally Responsive Membraneless Organelles. *Molecular Cell*, 57(5):936–947. 1.5.1, 4.6

O'Brien, E. P., Christodoulou, J., Vendruscolo, M., and Dobson, C. M. (2011). New scenarios of protein folding can occur on the ribosome. *Journal of the American Chemical Society*, 133(3):513–526. 1.1.2.1

Ogura, T., Whiteheart, S. W., and Wilkinson, A. J. (2004). Conserved arginine residues implicated in ATP hydrolysis, nucleotide-sensing, and inter-subunit interactions in AAA and AAA+ ATPases. *Journal of Structural Biology*, 146(1-2):106–112. 1.4.2.2, 1.4.3

Olivares, A. O., Kotamarthi, H. C., Stein, B. J., Sauer, R. T., and Baker, T. A. (2017). Effect of directional pulling on mechanical protein degradation by ATP-dependent proteolytic machines. *Proceedings of the National Academy of Sciences*, 114(31):E6306–E6313. 1.4.4.2

Olson, J. M. and Pierson, B. K. (1987). Evolution of Reaction Centers in Photosynthetic Prokaryotes. In Bourne, G. H., Jeon, K. W., and Friedlander, M., editors, *International Review of Cytology*, volume 108, pages 209–248. Academic Press. 1.2.1

Oltrogge, L. M., Chaijarasphong, T., Chen, A. W., Bolin, E. R., Marqusee, S., and Savage, D. F. (2020). Multivalent interactions between CsoS2 and Rubisco mediate α-carboxysome formation. *Nature Structural & Molecular Biology*, 27(3):281–287. 1.5.1.2

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(10):1390–1395. 1.3.2.2, 1.3.2.2.1, 4.1, A.1.1.2.1

Ott, C. M., Smith, B. D., Portis, A. R., and Spreitzer, R. J. (2000). Activase Region on Chloroplast Ribulose-1,5-bisphosphate Carboxylase/Oxygenase NONCONSERVATIVE SUBSTITUTION IN THE LARGE SUBUNIT ALTERS SPECIES SPECIFICITY OF PROTEIN INTERACTION. *Journal of Biological Chemistry*, 275(34):26241–26244. 1.4.6.1, 3.1.6, 4.4

Pagès, H., Aboyoun, P., Gentleman, R., and DebRoy, S. (2020). Biostrings: Efficient manipulation of biological strings. Bioconductor version: Release (3.12). 2.3

Paradis, E. and Schliep, K. (2019). Ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, 35(3):526–528. 2.3

Parry, M. A. J., Keys, A. J., Madgwick, P. J., Carmo-Silva, A. E., and Andralojc, P. J. (2008). Rubisco regulation: A role for inhibitors. *Journal of Experimental Botany*, 59(7):1569–1580. 1.3.3.4

Pearce, F. G. (2006). Catalytic by-product formation and ligand binding by ribulose bisphosphate carboxylases from different phylogenies. *Biochemical Journal*, 399(3):525–534. 1.3.3.4

Pena, K. L., Castel, S. E., de Araujo, C., Espie, G. S., and Kimber, M. S. (2010). Structural basis of the oxidative activation of the carboxysomal -carbonic anhydrase, CcmM. *Proceedings of the National Academy of Sciences*, 107(6):2455–2460. 1.2.3.3.1, 3.2.1

Peterhansel, C., Horst, I., Niessen, M., Blume, C., Kebeish, R., Kürkcüoglu, S., and Kreuzaler, F. (2010). Photorespiration. *The Arabidopsis Book*, 8:e0130. 1.2.2

Petersen, J., Brinkmann, H., Bunk, B., Michael, V., Päuker, O., and Pradella, S. (2012). Think pink: Photosynthesis, plasmids and the Roseobacter clade. *Environmental Microbiology*, 14(10):2661–2672. 1.2.1

Peterson-Forbrook, D. S., Hilton, M. T., Tichacek, L., Henderson, J. N., Bui, H. Q., and Wachter, R. M. (2017). Nucleotide Dependence of Subunit Rearrangements in Short-Form Rubisco Activase from Spinach. *Biochemistry*, 56(36):4906–4921. 1.4.6.1

Pierce, J., Tolbert, N. E., and Barker, R. (1980). Interaction of ribulosebisphosphate carboxylase/oxygenase with transition-state analogs. *Biochemistry*, 19(5):934–942. 1.3.3.4, 2.6, A.1.2, A.17

Portis, A. R., Li, C., Wang, D., and Salvucci, M. E. (2008). Regulation of Rubisco activase and its interaction with Rubisco. *Journal of Experimental Botany*, 59(7):1597–1604. 1.4.6.1, 3.1.6, 4.4

Portis, A. R. and Salvucci, M. E. (2005). The discovery of Rubisco activase — yet another story of serendipity. In Govindjee, Govindjee, Beatty, J. T., Gest, H., and Allen, J. F., editors, *Discoveries in Photosynthesis*, volume 20, pages 851–858. Springer Netherlands, Dordrecht. 1.4.6

Portis, A. R., Salvucci, M. E., and Ogren, W. L. (1986). Activation of Ribulosebisphosphate Carboxylase/Oxygenase at Physiological CO2 and Ribulosebisphosphate Concentrations by Rubisco Activase. *Plant Physiology*, 82(4):967–971. 1.4.6

Preissler, S. and Deuerling, E. (2012). Ribosome-associated chaperones as key players in proteostasis. *Trends in Biochemical Sciences*, 37(7):274–283. 1.1.2.2

Price, G. D., Coleman, J. R., and Badger, M. R. (1992). Association of Carbonic Anhydrase Activity with Carboxysomes Isolated from the Cyanobacterium Synechococcus PCC7942. *Plant Physiology*, 100(2):784–793. 1.2.3.3.1

Puchades, C., Rampello, A. J., Shin, M., Giuliano, C. J., Wiseman, R. L., Glynn, S. E., and Lander, G. C. (2017). Structure of the mitochondrial inner membrane AAA+ protease YME1 gives insight into substrate processing. *Science*, 358(6363). 1.4.3, 1.4.4.2, 1.4.5, 3.1.5, 4.5

Puchades, C., Sandate, C. R., and Lander, G. C. (2020). The molecular principles governing the activity and functional diversity of AAA+ proteins. *Nature Reviews Molecular Cell Biology*, 21(1):43–58. 1.4.1, 1.4.1, 1.19, 1.20, 1.4.4.1, 1.4.4.2, 1.4.5, 3.1.4, 3.1.5

Rachmilevitch, S., Cousins, A. B., and Bloom, A. J. (2004). Nitrate assimilation in plant shoots depends on photorespiration. *Proceedings of the National Academy of Sciences*, 101(31):11506–11510. 1.2.2

Rae, B. D., Long, B. M., Badger, M. R., and Price, G. D. (2013). Functions, Compositions, and Evolution of the Two Types of Carboxysomes: Polyhedral Microcompartments That Facilitate CO2 Fixation in Cyanobacteria and Some Proteobacteria. *Microbiology and Molecular Biology Reviews : MMBR*, 77(3):357–379. 1.2.3.3, 1.2.3.3.1, 1.7, 4.6

Ramazanov, Z., Rawat, M., Henk, M. C., Mason, C. B., Matthews, S. W., and Moroney, J. V. (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(2):210–216. 1.2.3.2

Rappaport, F. and Diner, B. A. (2008). Primary photochemistry and energetics leading to the oxidation of the (Mn)4Ca cluster and to the evolution of molecular oxygen in Photosystem II. *Coordination Chemistry Reviews*, 252(3):259–272. 1.2.1

Raunser, S., Magnani, R., Huang, Z., Houtz, R. L., Trievel, R. C., Penczek, P. A., and Walz, T. (2009). Rubisco in complex with Rubisco large subunit methyltransferase. *Proceedings of the National Academy of Sciences of the United States of America*, 106(9):3160–3165. 4.3

Raven, J. (2009). Contributions of anoxygenic and oxygenic phototrophy and chemolithotrophy to carbon and oxygen fluxes in aquatic environments. *Aquatic Microbial Ecology*, 56:177–192. 1.2

Raven, J. A., Beardall, J., Flynn, K. J., and Maberly, S. C. (2009). Phagotrophy in the origins of photosynthesis in eukaryotes and as a complementary mode of nutrition in phototrophs: Relation to Darwin's insectivorous plants. *Journal of Experimental Botany*, 60(14):3975–3987. 1.2

Renzette, N. (2011). Generation of Transformation Competent E. coli. *Current Protocols in Microbiology*, 22(1):A.3L.1–A.3L.5. 2.1.1

Riener, C. K., Kada, G., and Gruber, H. J. (2002). Quick measurement of protein sulfhydryls with Ellman's reagent and with 4,4′-dithiodipyridine. *Analytical and Bioanalytical Chemistry*, 373(4-5):266–276. 2.10.3, 3.25

Robinson, S. P. and Portis, A. R. (1988). Release of the nocturnal inhibitor, carboxyarabinitol- 1-phosphate, from ribulose bisphosphate carboxylase/oxygenase by rubisco activase. *FEBS LETTERS*, 233(2):4. 1.4.6

Robinson, S. P. and Portis, A. R. (1989). Adenosine triphosphate hydrolysis by purified rubisco activase. *Archives of Biochemistry and Biophysics*, 268(1):93–99. 2.10.1, 2.10.2, 3.1.1

Roll-Mecak, A. and Vale, R. D. (2008). Structural basis of microtubule severing by the hereditary spastic paraplegia protein spastin. *Nature*, 451(7176):363–367. 1.4.3

Rosenzweig, R., Nillegoda, N. B., Mayer, M. P., and Bukau, B. (2019). The Hsp70 chaperone network. *Nature Reviews Molecular Cell Biology*, 20(11):665–680. 1.1.2.2

Ross, C. A. and Poirier, M. A. (2004). Protein aggregation and neurodegenerative disease. *Nature Medicine*, 10(7):S10–S17. 1.1.2.1

Roth, R., Hall, L. N., Brutnell, T. P., and Langdale, J. A. (1996). Bundle sheath defective2, a Mutation That Disrupts the Coordinated Development of Bundle Sheath and Mesophyll Cells in the Maize Leaf. *The Plant Cell*, 8(5):915–927. 1.3.2.3.2

Rundle, S. J. and Zielinski, R. E. (1991). Organization and expression of two tandemly oriented genes encoding ribulosebisphosphate carboxylase/oxygenase activase in barley. *Journal of Biological Chemistry*, 266(8):4677–4685. 1.4.6.1

Sali, A., Shakhnovich, E., and Karplus, M. (1994). How does a protein fold? *Nature*, 369(6477):248–251. 1.1.1

Salvucci, M. E. (1992). Subunit interactions of Rubisco activase: Polyethylene glycol promotes self-association, stimulates ATPase and activation activities, and enhances interactions with Rubisco. *Archives of Biochemistry and Biophysics*, 298(2):688–696. 1.4.6.1

Salvucci, M. E. (2004). Potential for interactions between the carboxy- and amino-termini of Rubisco activase subunits. *FEBS Letters*, 560(1-3):205–209. 1.4.6.1

Salvucci, M. E., Portis, A. R., and Ogren, W. L. (1985). A soluble chloroplast protein catalyzes ribulosebisphosphate carboxylase/oxygenase activation in vivo. *Photosynthesis Research*, 7(2):193–201. 1.4.6, 1.4.6.1

Salvucci, M. E., Werneke, J. M., Ogren, W. L., and Portis, A. R. (1987). Purification and Species Distribution of Rubisco Activase. *Plant Physiology*, 84(3):930–936. 1.4.6.1

Sandate, C. R., Szyk, A., Zehr, E. A., Lander, G. C., and Roll-Mecak, A. (2019). An allosteric network in spastin couples multiple activities required for microtubule severing. *Nature Structural & Molecular Biology*, 26(8):671–678. 1.4.3, 1.4.5, 3.1.4, 4.5

Saschenbrecker, S., Bracher, A., Rao, K. V., Rao, B. V., Hartl, F. U., and Hayer-Hartl, M. (2007). Structure and Function of RbcX, an Assembly Chaperone for Hexadecameric Rubisco. *Cell*, 129(6):1189–1200. 1.3.2.2, 1.11, 1.3.2.2.1, 4.1, A.1.1.2.1, A.1.1.3

Satagopan, S., Chan, S., Perry, L. J., and Tabita, F. R. (2014). Structure-Function Studies with the Unique Hexameric Form II Ribulose-1,5-bisphosphate Carboxylase/Oxygenase (Rubisco) from Rhodopseudomonas palustris. *The Journal of Biological Chemistry*, 289(31):21433–21450. 1.3.1.2, 4.2, 4.2

Satagopan, S. and Spreitzer, R. J. (2004). Substitutions at the Asp-473 latch residue of chlamydomonas ribulosebisphosphate carboxylase/oxygenase cause decreases in carboxylation efficiency and CO(2)/O(2) specificity. *The Journal of Biological Chemistry*, 279(14):14240–14244. 4.2

Sato, M., Nimura-Matsune, K., Watanabe, S., Chibazakura, T., and Yoshikawa, H. (2007). Expression Analysis of Multiple dnaK Genes in the Cyanobacterium Synechococcus elongatus PCC 7942. *Journal of Bacteriology*, 189(10):3751–3758. A.1.1.2.1

Sato, S., Ikeuchi, M., and Nakamoto, H. (2008). Expression and function of a groEL paralog in the thermophilic cyanobacterium Thermosynechococcus elongatus under heat and cold stress. *FEBS Letters*, 582(23-24):3389–3395. A.1.1.2.1

Savir, Y., Noor, E., Milo, R., and Tlusty, T. (2010). Cross-species analysis traces adaptation of Rubisco toward optimality in a low-dimensional landscape. *Proceedings of the National Academy of Sciences*, 107(8):3475–3480. 3.1.1

Schmidt, G. W. and Mishkind, M. L. (1983). Rapid degradation of unassembled ribulose 1,5-bisphosphate carboxylase small subunits in chloroplasts. *Proceedings of the National Academy of Sciences of the United States of America*, 80(9):2632–2636. 1.3.2.3

Schneider, G., Lindqvist, Y., Brändén, C.-I., and Lorimer, G. (1986). Three-dimensional structure of ribulose-1,5-bisphosphate carboxylase/oxygenase from Rhodospirillum rubrum at 2.9 Å resolution. *The EMBO Journal*, 5(13):3409–3415. 1.3.1

Schneider, G., Lindqvist, Y., and Lundqvist, T. (1990). Crystallographic refinement and structure of ribulose-1,5-bisphosphate carboxylase from Rhodospirillum rubrum at 1.7 Å resolution. *Journal of Molecular Biology*, 211(4):989–1008. 1.3.1, 1.3.1.1, 1.9

Serban, A. J., Breen, I. L., Bui, H. Q., Levitus, M., and Wachter, R. M. (2018). Assembly–disassembly is coupled to the ATPase cycle of tobacco Rubisco activase. *Journal of Biological Chemistry*, 293(50):19451–19465. 1.4.6.1

Seydoux, G. and Braun, R. E. (2006). Pathway to Totipotency: Lessons from Germ Cells. *Cell*, 127(5):891–904. 1.5

Shen, J. B., Orozco, E. M., and Ogren, W. L. (1991). Expression of the two isoforms of spinach ribulose 1,5-bisphosphate carboxylase activase and essentiality of the conserved lysine in the consensus nucleotide-binding domain. *The Journal of Biological Chemistry*, 266(14):8963–8968. 1.4.6.1

Shin, M., Puchades, C., Asmita, A., Puri, N., Adjei, E., Wiseman, R. L., Karzai, A. W., and Lander, G. C. (2020). Structural basis for distinct operational modes and protease activation in AAA+ protease Lon. *Science Advances*, 6(21). 4.5

Shin, Y. and Brangwynne, C. P. (2017). Liquid phase condensation in cell physiology and disease. *Science*, 357(6357). 3.2.5

Shivhare, D. and Mueller-Cajar, O. (2017). In Vitro Characterization of Thermostable CAM Rubisco Activase Reveals a Rubisco Interacting Surface Loop1[OPEN]. *Plant Physiology*, 174(3):1505–1516. 1.4.6.1, 3.1.6, 4.4

Shivhare, D., Ng, J., Tsai, Y.-C. C., and Mueller-Cajar, O. (2019). Probing the rice Rubisco–Rubisco activase interaction via subunit heterooligomerization. *Proceedings of the National Academy of Sciences*, 116(48):24041–24048. 1.4.6.1, 3.1.6, 4.4

Singh, A. K., Balchin, D., Imamoglu, R., Hayer-Hartl, M., and Hartl, F. U. (2020). Efficient Catalysis of Protein Folding by GroEL/ES of the Obligate Chaperonin Substrate MetF. *Journal of Molecular Biology*, 432(7):2304–2318. 1.1.2.3

Singh, S. K. and Maurizi, M. R. (1994). Mutational analysis demonstrates different functional roles for the two ATP-binding sites in ClpAP protease from Escherichia coli. *The Journal of Biological Chemistry*, 269(47):29537–29545. 1.4.2.1

Smith, G. R., Contreras-Moreira, B., Zhang, X., and Bates, P. A. (2004). A link between sequence conservation and domain motion within the AAA+ family. *Journal of Structural Biology*, 146(1):189–204. 1.4.2.3

Snider, J. and Houry, W. A. (2006). MoxR AAA+ ATPases: A novel family of molecular chaperones? *Journal of Structural Biology*, 156(1):200–209. 1.4.6.3

Somepics (2015). Light-dependent reactions of photosynthesis in the thylakoid membrane of plant cells. <https://creativecommons.org/licenses/by-sa/4.0>, via Wikimedia Commons. 1.4

Somerville, C. R., Portis, A. R., and Ogren, W. L. (1982). A Mutant of Arabidopsis thaliana Which Lacks Activation of RuBP Carboxylase In Vivo. *Plant Physiology*, 70(2):381–387. 1.4.6

Sommer, M., Cai, F., Melnicki, M., and Kerfeld, C. A. (2017). β-Carboxysome bioinformatics: Identification and evolution of new bacterial microcompartment protein gene classes and core locus constraints. *Journal of Experimental Botany*, 68(14):3841–3855. 1.2.3.3.1

Stach, L., Morgan, R. M., Makhlouf, L., Douangamath, A., von Delft, F., Zhang, X., and Freemont, P. S. (2020). Crystal structure of the catalytic D2 domain of the AAA+ ATPase p97 reveals a putative helical split-washer-type mechanism for substrate unfolding. *Febs Letters*, 594(5):933–943. 1.4.3

Staněk, D. (2016). Cajal bodies and snRNPs - friends with benefits. *RNA Biology*, 14(6):671–679. 1.5

Stec, B. (2012). Structural mechanism of RuBisCO activation by carbamylation of the active site lysine. *Proceedings of the National Academy of Sciences*, 109(46):18785–18790. 1.3.3.1, 1.16

Stern, D. B., Hanson, M. R., and Barkan, A. (2004). Genetics and genomics of chloroplast biogenesis: Maize as a model system. *Trends in Plant Science*, 9(6):293–301. 1.3.2.2.2, 1.3.2.3.1

Stotz, M., Mueller-Cajar, O., Ciniawsky, S., Wendler, P., Hartl, F. U., Bracher, A., and Hayer-Hartl, M. (2011). Structure of green-type Rubisco activase from tobacco. *Nature Structural & Molecular Biology*, 18(12):1366–1370. 1.4.6.1, 3.6, 3.1.3, 3.1.5, 4.2, 4.5, 4.6

Streusand, V. J. and Portis, A. R. (1987). Rubisco Activase Mediates ATP-Dependent Activation of Ribulose Bisphosphate Carboxylase 1. *Plant Physiology*, 85(1):152–154. 1.4.6

Studier, F. W. (2005). Protein production by auto-induction in high-density shaking cultures. *Protein Expression and Purification*, 41(1):207–234. 2.5.4

Su, M., Guo, E. Z., Ding, X., Li, Y., Tarrasch, J. T., Brooks, C. L., Xu, Z., and Skiniotis, G. (2017). Mechanism of Vps4 hexamer function revealed by cryo-EM. *Science Advances*, 3(4):e1700325. 1.4.4.2

Su, X., Ditlev, J. A., Hui, E., Xing, W., Banjade, S., Okrut, J., King, D. S., Taunton, J., Rosen, M. K., and Vale, R. D. (2016). Phase separation of signaling molecules promotes T cell receptor signal transduction. *Science*, 352(6285):595–599. 4.6

Sugawara, H., Yamamoto, H., Shibata, N., Inoue, T., Okada, S., Miyake, C., Yokota, A., and Kai, Y. (1999). Crystal Structure of Carboxylase Reaction-oriented Ribulose 1,5-Bisphosphate Carboxylase/Oxygenase from a Thermophilic Red Alga, *Galdieria partita*. *Journal of Biological Chemistry*, 274(22):15655–15661. 1.3.1.2, 4.2

Sutter, M., Roberts, E. W., Gonzalez, R. C., Bates, C., Dawoud, S., Landry, K., Cannon, G. C., Heinhorst, S., and Kerfeld, C. A. (2015). Structural Characterization of a Newly Identified Component of α-Carboxysomes: The AAA+ Domain Protein CsoCbbQ. *Scientific Reports*, 5. 1.4.6.3

Suzuki, K., Nakanishi, H., Bower, J., Yoder, D. W., Osteryoung, K. W., and Miyagishima, S.-y. (2009). Plastid chaperonin proteins Cpn60α and Cpn60β are required for plastid division in Arabidopsis thaliana. *BMC Plant Biology*, 9:38. A.1.1.2.1

Tabita, F. R. (1999). Microbial ribulose 1,5-bisphosphate carboxylase/oxygenase: A different perspective. *Photosynthesis Research*, 26:1–28. 1.3.2.1, A.1.1.2.1

Tabita, F. R., Hanson, T. E., Li, H., 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(4):576–599. 1.3

Tabita, F. R., Satagopan, S., Hanson, T. E., Kreel, N. E., and Scott, S. S. (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(7):1515–1524. 1.3, 1.3, 4.2

Taler, D., Galperin, M., Benjamin, I., Cohen, Y., and Kenigsbuch, D. (2004). Plant eR Genes That Encode Photorespiratory Enzymes Confer Resistance against Disease. *The Plant Cell*, 16(1):172–184. 1.2.2

Tang, Y.-C., Chang, H.-C., Roeben, A., Wischnewski, D., Wischnewski, N., Kerner, M. J., Hartl, F. U., and Hayer-Hartl, M. (2006). Structural Features of the GroEL-GroES Nano-Cage Required for Rapid Folding of Encapsulated Protein. *Cell*, 125(5):903–914. 1.1.2.3

Taylor, T. C. and Andersson, I. (1996). Structural transitions during activation and ligand binding in hexadecameric Rubisco inferred from the crystal structure of the activated unliganded spinach enzyme. *Nature Structural Biology*, 3(1):95–101. 1.3.3.1, 1.3.3.2, 4.2, 4.2

Taylor, T. C. and Andersson, I. (1997). The structure of the complex between rubisco and its natural substrate ribulose 1,5-bisphosphate11Edited by R. Huber. *Journal of Molecular Biology*, 265(4):432–444. 1.3.1.1, 1.3.3.2, 1.3.3.2, 4.2, 4.2

Tcherkez, G. (2016). The mechanism of Rubisco-catalysed oxygenation. *Plant, Cell & Environment*, 39(5):983–997. 1.3.3.3

Tcherkez, G. G. B., Farquhar, G. D., and Andrews, T. J. (2006). Despite slow catalysis and confused substrate specificity, all ribulose bisphosphate carboxylases may be nearly perfectly optimized. *Proceedings of the National Academy of Sciences*, 103(19):7246–7251. 3.1.1

Techtmann, S. M. and Robb, F. T. (2010). Archaeal-like chaperonins in bacteria. *Proceedings of the National Academy of Sciences*, 107(47):20269–20274. 1.1.2.3

The R Core team (2020). R: A Language and Environment forStatistical Computing. R Foundation for Statistical Computing. 2.3, 2.4

Thompson, A. W., Foster, R. A., Krupke, A., Carter, B. J., Musat, N., Vaulot, D., Kuypers, M. M. M., and Zehr, J. P. (2012). Unicellular Cyanobacterium Symbiotic with a Single-Celled Eukaryotic Alga. *Science*, 337(6101):1546–1550. 1.2.1

To, K.-Y., Suen, D.-F., and Chen, S.-C. G. (1999). Molecular characterization of ribulose-1,5-bisphosphate carboxylase/oxygenase activase in rice leaves. *Planta*, 209(1):66–76. 1.4.6.1

Tsai, Y.-C. C., Lapina, M. C., Bhushan, S., and Mueller-Cajar, O. (2015). Identification and characterization of multiple rubisco activases in chemoautotrophic bacteria. *Nature Communications*, 6. 1.4.6.3, 1.4.6.3, 4.2

Tsai, Y.-C. C., Ye, F., Liew, L., Liu, D., Bhushan, S., Gao, Y.-G., and Mueller-Cajar, O. (2020). Insights into the mechanism and regulation of the CbbQO-type Rubisco activase, a MoxR AAA+ ATPase. *Proceedings of the National Academy of Sciences*, 117(1):381–387. 1.4.6.3, 4.2, 4.2

Turmo, A., Gonzalez-Esquer, C. R., and Kerfeld, C. A. (2017). Carboxysomes: Metabolic modules for CO2 fixation | FEMS Microbiology Letters | Oxford Academic. https://academic.oup.com/femsle/article/364/18/fnx176/4082729. 1.7

Twomey, E. C., Ji, Z., Wales, T. E., Bodnar, N. O., Ficarro, S. B., Marto, J. A., Engen, J. R., and Rapoport, T. A. (2019). Substrate processing by the Cdc48 ATPase complex is initiated by ubiquitin unfolding. *Science (New York, N.Y.)*, 365(6452). 1.4.3, 1.4.4.1

Villarejo, A., Martinez, F., Plumed, M. d. P., and Ramazanov, Z. (1996). The induction of the CO2 concentrating mechanism in a starch-less mutant of Chlamydomonas reinhardtii. *Physiologia Plantarum*, 98(4):798–802. 1.2.3.2

Walker, J., Saraste, M., Runswick, M., and Gay, N. (1982). Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *The EMBO Journal*, 1(8):945–951. 1.4.1, 1.4.2.1

Wang, D. and Portis, A. R. (2006). Increased Sensitivity of Oxidized Large Isoform of Ribulose-1,5-bisphosphate Carboxylase/Oxygenase (Rubisco) Activase to ADP Inhibition Is Due to an Interaction between Its Carboxyl Extension and Nucleotide-binding Pocket *. *Journal of Biological Chemistry*, 281(35):25241–25249. 1.4.6.1, 4.6

Wang, H., Yan, X., Aigner, H., Bracher, A., Nguyen, N. D., Hee, W. Y., Long, B. M., Price, G. D., Hartl, F. U., and Hayer-Hartl, M. (2019). Rubisco condensate formation by CcmM in β-carboxysome biogenesis. *Nature*, 566(7742):131–135. 1.5.1, 1.5.1.1, 1.25, 2.10.5, 3.2.1, 3.2.2, 3.2.3, 3.2.4, 3.26, 4.6

Wang, J., Song, J. J., Seong, I. S., Franklin, M. C., Kamtekar, S., Eom, S. H., and Chung, C. H. (2001). Nucleotide-Dependent Conformational Changes in a Protease-Associated ATPase HslU. *Structure*, 9(11):1107–1116. 3.1.2

Wang, L., Myasnikov, A., Pan, X., and Walter, P. (2020). Structure of the AAA protein Msp1 reveals mechanism of mislocalized membrane protein extraction. *eLife*, 9:e54031. 1.4.5, 4.5

Wang, Z. Y. and Portis, A. R. (1992). Dissociation of Ribulose-1,5-Bisphosphate Bound to Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase and Its Enhancement by Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase Activase-Mediated Hydrolysis of ATP. *Plant Physiology*, 99(4):1348–1353. 3.1.1

Wang, Z. Y., Ramage, R. T., and Portis, A. R. (1993). Mg2+ and ATP or adenosine 5′-[γ-thio]-triphosphate (ATPγS) enhances intrinsic fluorescence and induces aggregation which increases the activity of spinach Rubisco activase. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology*, 1202(1):47–55. 1.4.6.1

Wang, Z.-Y., Snyder, G. W., Esau, B. D., Portis, A. R., and Ogren, W. L. (1992). Species-Dependent Variation in the Interaction of Substrate-Bound Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase (Rubisco) and Rubisco Activase. *Plant Physiology*, 100(4):1858–1862. 1.4.6.1

Weibezahn, J., Schlieker, C., Bukau, B., and Mogk, A. (2003). Characterization of a Trap Mutant of the AAA+ Chaperone ClpB. *Journal of Biological Chemistry*, 278(35):32608–32617. 1.4.2.1

Weissenbach, J. (2017). Evolution of Chaperonin Gene Duplication in Stigonematalean Cyanobacteria (Subsection V). *Genome Biol Evol.*, pages 241–252. A.1.1.2.1

Werneke, J. M., Chatfield, J. M., and Ogren, W. L. (1989). Alternative mRNA splicing generates the two ribulosebisphosphate carboxylase/oxygenase activase polypeptides in spinach and Arabidopsis. *The Plant Cell*, 1(8):815–825. 1.4.6.1

Wheatley, N. M., Sundberg, C. D., Gidaniyan, S. D., Cascio, D., and Yeates, T. O. (2014). Structure and Identification of a Pterin Dehydratase-like Protein as a Ribulose-bisphosphate Carboxylase/Oxygenase (RuBisCO) Assembly Factor in the α-Carboxysome *. *Journal of Biological Chemistry*, 289(11):7973–7981. 1.3.2.3.1

White, K. I., Zhao, M., Choi, U. B., Pfuetzner, R. A., and Brunger, A. T. (2018). Structural principles of SNARE complex recognition by the AAA+ protein NSF. *eLife*, 7:e38888. 1.4.3, 1.4.4.1, 4.5

Whitney, S. M., Baldet, P., Hudson, G. S., and Andrews, T. J. (2001). Form I Rubiscos from non-green algae are expressed abundantly but not assembled in tobacco chloroplasts. *The Plant Journal*, 26(5):535–547. 2.3, 2.10.4

Whitney, S. M., Birch, R., Kelso, C., Beck, J. L., and Kapralov, M. V. (2015). Improving recombinant Rubisco biogenesis, plant photosynthesis and growth by coexpressing its ancillary RAF1 chaperone. *Proceedings of the National Academy of Sciences of the United States of America*, 112(11):3564–3569. 1.3.2.2, 1.3.2.3, A.1.1.2.1

Whitney, S. M., Houtz, R. L., and Alonso, H. (2011). Advancing Our Understanding and Capacity to Engineer Nature's CO_2-Sequestering Enzyme, Rubisco. *Plant Physiology*, 155(1):27–35. 1.3

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., and Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, 4(43):1686. 2.3, 2.4

Wildman, S. G. and Bonner, J. (1947). The proteins of green leaves; isolation, enzymatic properties and auxin content of spinach cytoplasmic proteins. *Archives of Biochemistry*, 14(3):381–413. 1.3.1

Wilson, R. H. and Hayer-Hartl, M. (2018). Complex Chaperone Dependence of Rubisco Biogenesis. *Biochemistry*, 57(23):3210–3216. 1.15

Wilson, R. H., Thieulin-Pardo, G., Hartl, F.-U., and Hayer-Hartl, M. (2019). Improved recombinant expression and purification of functional plant Rubisco. *FEBS Letters*, 593(6):611–621. 2.1.3, 2.1

Winter, D. J. (2017). Rentrez: An R package for the NCBI eUtils API. *The R Journal*, 9(2):520–526. 2.3

Wostrikoff, K. and Stern, D. (2007). Rubisco large-subunit translation is autoregulated in response to its assembly state in tobacco chloroplasts. *Proceedings of the National Academy of Sciences of the United States of America*, 104(15):6466–6471. 1.3.2.3

Wunder, T., Cheng, S. L. H., Lai, S.-K., Li, H.-Y., and Mueller-Cajar, O. (2018). The phase separation underlying the pyrenoid-based microalgal Rubisco supercharger. *Nature Communications*, 9(1):5076. 1.5.1.2

Wyatt, P. J. (1993). Light scattering and the absolute characterization of macromolecules. *Analytica Chimica Acta*, 272(1):1–40. 2.2

Xia, L.-Y., Jiang, Y.-L., Kong, W.-W., Sun, H., Li, W.-F., Chen, Y., and Zhou, C.-Z. (2020). Molecular basis for the assembly of RuBisCO assisted by the chaperone Raf1. *Nature Plants*, 6(6):708–717. 4.1

Xu, Z., Horwich, A. L., and Sigler, P. B. (1997). The crystal structure of the asymmetric GroEL–GroES–(ADP) 7 chaperonin complex. *Nature*, 388(6644):741–750. 1.1.2.3, 1.2, 1.1.2.3

Yam, A. Y., Xia, Y., Lin, H.-T. J., Burlingame, A., Gerstein, M., and Frydman, J. (2008). Defining the TRiC/CCT interactome links chaperonin function to stabilization of newly made proteins with complex topologies. *Nature Structural & Molecular Biology*, 15(12):1255–1262. 1.1.2.3

Yan, X., Shi, Q., Bracher, A., Miličić, G., Singh, A. K., Hartl, F. U., and Hayer-Hartl, M. (2018). GroEL Ring Separation and Exchange in the Chaperonin Reaction. *Cell*, 172(3):605–617.e11. 1.1.2.3

Yokota, A. (1991). Carboxylation and Detoxification of Xylulose Bisphosphate by Spinach Ribulose Bisphosphate Carboxylase/Oxygenase. *Plant and Cell Physiology*, 32(6):755–762. 1.3.3.4

Yoshizawa, Y., Toyoda, K., Arai, H., Ishii, M., and Igarashi, Y. (2004). CO2-Responsive Expression and Gene Organization of Three Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase Enzymes and Carboxysomes in Hydrogenovibrio marinus Strain MH-110. *Journal of Bacteriology*, 186(17):5685–5691. 1.3, 1.4.6.3

Yu, G. (2020). Using ggtree to Visualize Data on Tree-Like Structures. *Current Protocols in Bioinformatics*, 69(1):e96. 2.3

Zarzycki, J., Axen, S. D., Kinney, J. N., and Kerfeld, C. A. (2013). Cyanobacterial-based approaches to improving photosynthesis in plants. *Journal of Experimental Botany*, 64(3):787–798. 1.6

Zarzycki, J., Brecht, V., Müller, M., and Fuchs, G. (2009). Identifying the missing steps of the autotrophic 3-hydroxypropionate CO2 fixation cycle in Chloroflexus aurantiacus. *Proceedings of the National Academy of Sciences*, 106(50):21317–21322. 1.2

Zehr, E., Szyk, A., Piszczek, G., Szczesna, E., Zuo, X., and Roll-Mecak, A. (2017). Katanin spiral and ring structures shed light on power stroke for microtubule severing. *Nature Structural & Molecular Biology*, 24(9):717–725. 1.4.3, 3.1.4, 4.5

Zehr, E. A., Szyk, A., Szczesna, E., and Roll-Mecak, A. (2020). Katanin Grips the β-Tubulin Tail through an Electropositive Double Spiral to Sever Microtubules. *Developmental Cell*, 52(1):118–131.e6. 4.5

Zhang, N., Kallis, R. P., Ewy, R. G., and Portis, A. R. (2002). Light modulation of Rubisco in Arabidopsis requires a capacity for redox regulation of the larger Rubisco activase isoform. *Proceedings of the National Academy of Sciences*, 99(5):3330–3334. 1.4.6.1

Zhang, N. and Portis, A. R. (1999). Mechanism of light regulation of Rubisco: A specific role for the larger Rubisco activase isoform involving reductive activation by thioredoxin-f. *Proceedings of the National Academy of Sciences*, 96(16):9438–9443. 1.4.6.1

Zhang, N., Schürmann, P., and Portis, A. R. (2001). Characterization of the regulatory function of the 46-kDa isoform of Rubisco activase from Arabidopsis. *Photosynthesis Research*, 68(1):29–37. 1.4.6.1

Zhang, X. and Wigley, D. B. (2008). The 'glutamate switch' provides a link between ATPase activity and ligand binding in AAA+ proteins. *Nature Structural & Molecular Biology*, 15(11):1223–1227. 1.4.2.1

Zhao, M., Wu, S., Zhou, Q., Vivona, S., Cipriano, D. J., Cheng, Y., and Brunger, A. T. (2015). Mechanistic insights into the recycling machine of the SNARE complex. *Nature*, 518(7537):61–67. 1.4.3

Zhou, M., Guo, J., Cha, J., Chae, M., Chen, S., Barral, J. M., Sachs, M. S., and Liu, Y. (2013). Non-optimal codon usage affects expression, structure and function of clock protein FRQ. *Nature*, 495(7439):111–115. 1.1.2.1

A. Appendix

A.1. Generation of recombinant proteins and key resources

A.1.1. Generation of recombinant proteins from *Nostoc* sp. PCC 7120

All recombinant proteins generated in this study were expressed as N-terminal H_6ubi-fusion proteins [Baker et al., 2005; Catanzariti et al., 2004]. Because the H_6ubi moiety was cleaved after IMAC at a partially purified stage by addition of the deubiquitinating enzyme usp2, N-terminal processing in the lysate of the expression host was largely avoided, allowing the production of essentially unaltered proteins with intact N-termini. All recombinant proteins used in this study were > 95 % pure as judged by SDS-PAGE analysis of the final product (Figure **A.1**).

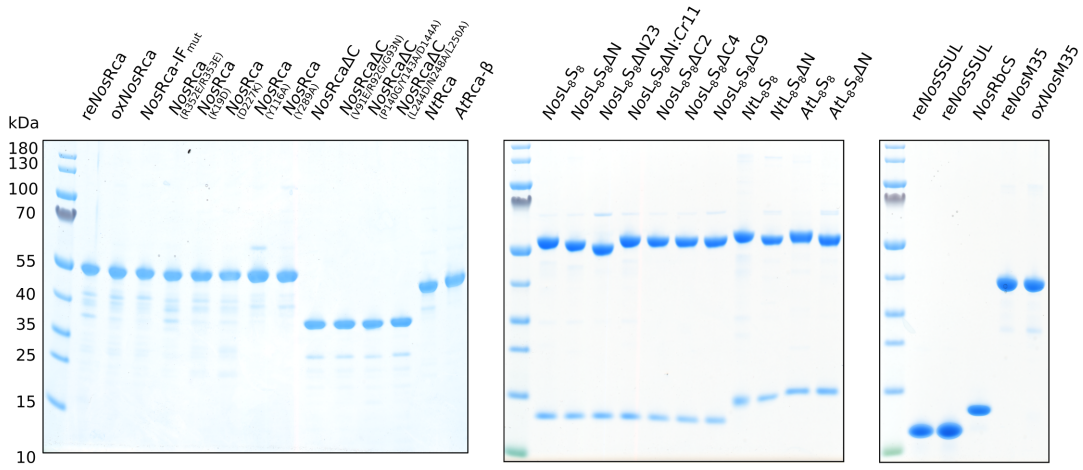

Figure A.1.: SDS-PAGE analysis of purified proteins demonstrates purity of reagents (> 95 %).
1.5 µg of each purified protein was analyzed by SDS-PAGE and Coomassie staining. Protein specific absorption coefficients as stated in Table **2.2** were used for concentration determination via UV-vis spectrometry.

A.1.1.1. *Nos*Rca, *Nos*RcaΔC and mutants

To investigate the function of the AAA$^+$ domain and the SSUL domain of *Nos*Rca separately, this study required the recombinant production of three related constructs: Full length *Nos*Rca (residues 2-414), a C-terminal truncated version lacking linker and SSUL domain, *Nos*RcaΔC (residues 2-291), and the SSUL domain, *Nos*SSUL (residues 325-414). While the purification of *Nos*RcaΔC and *Nos*SSUL at adequate yield was straightforward in Ethanolamine (ETA) buffer (50 mM ETA pH 9.2 / 100 mM NaCl, 10 mM $MgCl_2$, 5 % glycerol), purification of the full-length construct proved to be difficult. In contrast to *Nos*RcaΔC and *Nos*SSUL, initially $\sim$ 70 % of the full-length *Nos*Rca precipitated upon cleavage of the H_6ubi-moiety. Moreover, the full-length construct was intolerant to low-salt conditions, which were established prior to an ion exchange chromatography at a later stage during protein purification. To solve this problem a buffer screen was conducted to find conditions more favorable to the stability of *Nos*Rca.

A.1.1.1.1. Optimization of buffer conditions for the purification and cryo-EM analysis of *Nos*Rca

The thermal stability of *Nos*Rca in different buffer conditions was analyzed by differential scanning fluorimetry (DSF). DSF reports on conformational rearrangements or the creation of new molecular interactions upon changes in environmental conditions such as in pH, ionic strength, detergents or the presence of specific anions or cations and cofactors. These molecular rearrangements alter the proteins Gibbs free energy and thus its thermal stability [Gao et al., 2020; Niesen et al., 2007], which manifests in a shift of the melting temperature (T_m), the temperature at which 50 % of the protein is unfolded. Many conventional DSF methods rely on extrinsic binding of a hydrophobic dye to the protein of interest in the course of its unfolding event upon heat or chemical denaturation. In contrast, nanoDSF is a label-free technique and relies on changes in the intrinsic tryptophan (Trp) environment upon unfolding [Baaske et al., 2010]. Often Trp residues are buried in the core of a protein which represents a hydrophobic environment. Upon unfolding, the initially buried Trp residues become solvent-exposed and their environment becomes more polar resulting in a redshift [Ghisaidoobe and Chung, 2014] and a change in the intrinsic fluorescence ratio $\frac{350\,\mathrm{nm}}{330\,\mathrm{nm}}$ (WFR).

In conventional DSF, a Boltzmann model is frequently applied to determine the T_m as indicator for protein stability. The Boltzmann model assumes two-state unfolding, meaning that upon thermal unfolding the protein is converted from the native to the

denatured state in a single transition without occupying partially unfolded intermediate states. This assumption is often violated in DSF of macromolecular protein complexes, because the recorded melting curves represent the sum of the transitions of individual, conformationally different subunits. Dissociation and non-cooperative unfolding of individual subunits often result in polyphasic melting curves which cannot be represented by a Boltzmann fit. Hence, DSF is mostly used on monomeric proteins. Chari et al. [2015] subjected macromolecular complexes to thermal unfolding by DSF in a variety of buffer conditions and evaluated corresponding complex stability by negative stain EM. They noted, that a destabilizing environment (as judged by aggregating particles in EM micrographs) is indicated when individual subunits unfold in a non-cooperative manner, resulting in polyphasic melting curves. In contrast, they obtained monodispersed, single particles (indicative of stabilizing buffer conditions), when DSF curves approximated apparent two-state unfolding which correlated with a steepening slope of the single unfolding transition, indicative of the concerted unfolding of subunits. In fact, their findings resulted in the development of a thermodynamic framework, able to model polyphasic transitions, at the heart of an automated platform for protein stability screening by DSF. However, successful and consistent interpretation of conventional DSF data (using a simple Boltzmann model) from small protein complexes has been reported before [Kopec and Schneider, 2011].

The sequence of *Nos*Rca comprises two Trp residues - Trp178 is located between the α/β-subdomain and the α-helical subdomain of the AAA+ domain, while Trp359 is located within the SSUL module. Because the two Trp residues are located in separate domains of *Nos*Rca, we reasoned that non-cooperative unfolding would result in a bad fit (high S) when using the conventional Boltzmann model. To identify stabilizing buffer conditions, both to improve *Nos*Rca yields and to find conditions amenable for cryo-EM analysis, we combined the model parameters T_m and $maxSlope$ in our analysis of the data.

Melting curves were processed as described in chapter **2.4**. Interestingly, the majority of the melting curves showed a sigmoidal increase of the *WFR* and were represented well by the two-state unfolding Boltzmann model, as judged by visual inspection of melting curves (e.g. Figure **A.2**) and overall low values for the residual standard deviation, S (Figure **A.3 A**). In comparison with *Nos*Rca, melting curves of *Nos*RcaΔC followed two-state unfolding closer, as indicated by overall lower S values (Figure **A.3 A** - lateral density plot). The poorly fit melting curves ($S > 95\,\%$ quantile) revealed clear signs of deviations from two-state unfolding and hence were excluded from further analysis (Figure **A.3 B**). Comparison of T_m and $maxSlope$ of

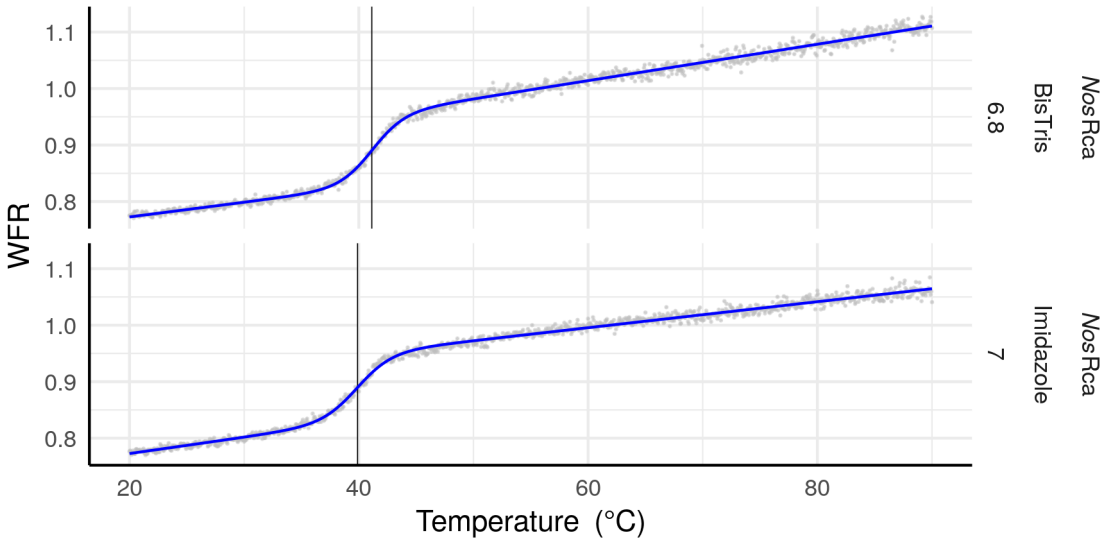

Figure A.2.: Examples of typical protein melting curves.
The intrinsic (or tryptophan) fluorescence ratio (WFR) $\frac{F_{350\,\mathrm{nm}}}{F_{330\,\mathrm{nm}}}$ of two typical protein melting curves is plotted as a function of temperature (grey points). The curve fits, a slanted Boltzmann model (see Chapter **2.4**), are shown in blue. Melting temperatures T_m are indicated by vertical lines. Protein, buffer and pH are indicated to the right of the plotting area. For buffer specifications, see Table **2.4**

the ethanolamine (ETA) pH 9.2 controls to the corresponding (protein-wise) distributions of these parameters as presented in the buffer screen (Figure **A.4 A**, **C** and **D**), indicated a potential to increase protein stability by changing the buffer system. For the remaining data the S value was considered a poor indicator of protein stability, due to its large variability in the ETA controls (Figure **A.4 B**).

On the other hand, the correlation of T_m and *maxSlope* revealed, that high T_m (> 42 °C) coincided with steep unfolding events (high *maxSlope*) (Figure **A.5**). In presence of MMT and ADA buffer these parameters were comparatively invariable to a broad pH range (ADA: 6.6 - 7.4, MMT: 6.8 - 8.4), suggesting buffer compositions that would be beneficial to protein stability. Hierarchical sorting according to T_m alone, resulted with both ADA and MMT as top hits for *Nos*Rca (rank 1-8) and *Nos*RcaΔC (rank 1-7 and 12), independent of pH. In contrast, buffers spanning the more acidic range of the pH scale, e.g. NaCitrate (pH 5.5 - 7.0) and NaKPhosphate (pH 5.8 - 7.6), showed a stronger pH dependence of T_m and *maxSlope* with overall lower T_m and *maxSlope* parameters. Compared to the ETA controls, T_m and *maxSlope* for *Nos*Rca were maximally raised from 38.6 °C to 44.0 °C in ADA pH 7.4 (concomitant with an increase of *maxSlope* from 0.74 to 0.92) and from 0.74 to 0.94 in MMT pH 8.4 (concomitant with an increase of T_m from 38.6 °C to 42.7 °C), respectively, rendering them both a suitable replacement for the ETA buffer with beneficial effects on the stability of *Nos*Rca (Figure **A.5**). However, it was unclear

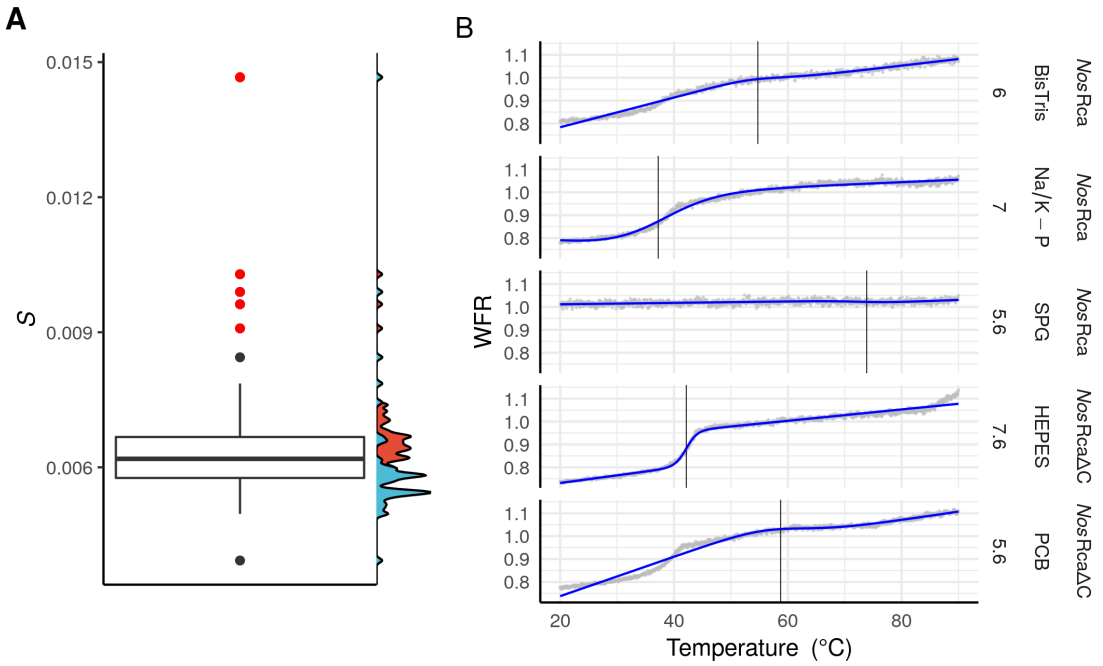

Figure A.3.: High S indicates deviation from two-stage unfolding.
A) Overall distribution of the error of the fit S. The overall distribution of S is represented by a box and whiskers plot (Tukey). Values of S outside of the 95 % quantile are considered outliers (red points). The lateral density plot distinguishes between S -values for *Nos*Rca (red) and *Nos*RcaΔC (blue). **B) S values outside of the 95 % quantile are indicative of clear deviation from two-state unfolding.** For melting curves with $S >$ 95 % quantile, the intrinsic (or tryptophan) fluorescence ratio (WFR) $\frac{F_{350\,\mathrm{nm}}}{F_{330\,\mathrm{nm}}}$ is plotted as a function of temperature (grey points). The curve fits, a slanted Boltzmann model (see Chapter **2.4**), are shown in blue. Melting temperatures T_m are indicated by vertical lines. For buffer specifications, see Table **2.4**.

whether the slightly higher T_m obtained in ADA buffer pH 7.4 or the slightly higher *maxSlope* in MMT buffer pH 8.4 was the better indicator for *Nos*Rca stability. Thus, to settle on one buffer for the next purification attempts, we reasoned that *Nos*Rca would bind tighter to the anion exchange column at elevated pH of the MMT buffer pH 8.4, facilitating its purification. Notably, MMT is also a more cost economic buffer, being ∼ 7 times cheaper than ADA buffer.

Modifications of the initial purification protocol, including a change of the buffer from ETA pH 9.2 to MMT pH 8.4, resulted in greatly increased purification yields (∼ fivefold increase in final *Nos*Rca yield). The improved purification strategy is described in detail in Chapter **2.5.1** and the outcome presented in Chapter **A.1.1.1.2**.

A.1.1.1.2. Purification of *Nos*Rca, *Nos*RcaΔC and mutants

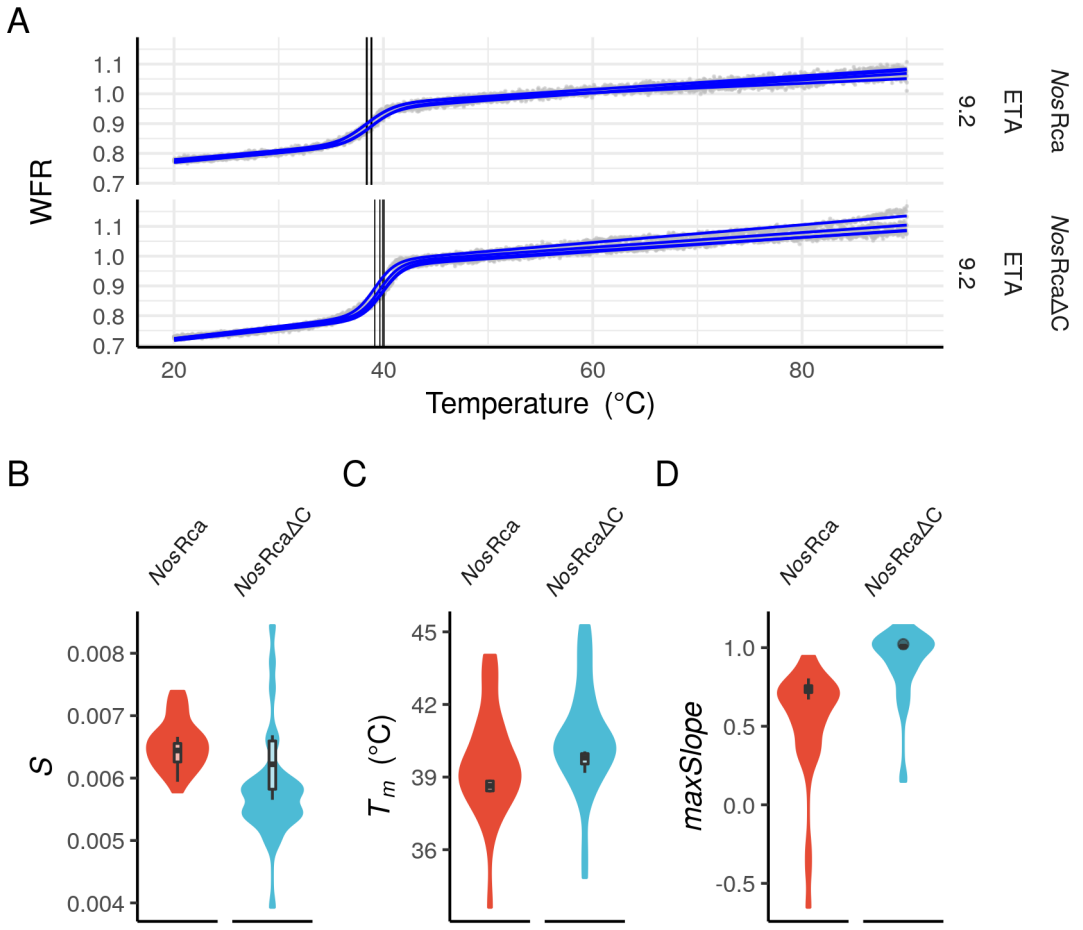

Figure A.4.: Variability and potential of the Bufferscreen.
A) The slanted Boltzmann model represents two-stage unfolding of *Nos*Rca and *Nos*RcaΔC well. WFR ratio of control reactions in ETA, pH 9.2, are plotted as function of temperature. Control reactions were distributed across the capillary rack to account for potential loading bias and were measured in quadruplicate per screen. For experimental details, see chapter **2.4** and Figure **A.3**. **Proteinwise distribution of S (A), T_m (B) and *maxSlope* (C) reveal potential for protein stabilization in the set of tested buffers.** Distributions of S, Tm and *maxSlope* per protein are shown as violin plots. Corresponding parameters for the control reactions only are represented as box and whiskers plots (Tukey) in the foreground.

*Nos*Rca, *Nos*RcaΔC and corresponding mutants were strongly enriched by IMAC and the soluble fractions obtained after cleavage of the H_6ubi-moiety were subsequently purified by IEX and SEC as described in Chapter **2.5.1**. Next to changing the purification buffer from ETA pH 9.2 to MMT pH 8.4, cleavage of the H_6ubi moiety by Usp2 digest as well as the subsequent desalting were equally critical steps for a successful purification. Only in presence of 10 % glycerol, 5 mM CHAPS, high ionic strength and at low protein concentrations ($< 0.5\,\mathrm{g\,L^{-1}}$) could the H_6ubi moiety be cleaved without causing the protein to aggregate and precipitate rapidly. As *Nos*Rca is susceptible to low salt concentrations, fast desalting by gel filtration and immediate application to IEX was necessary to prevent extensive aggregation. Rep-

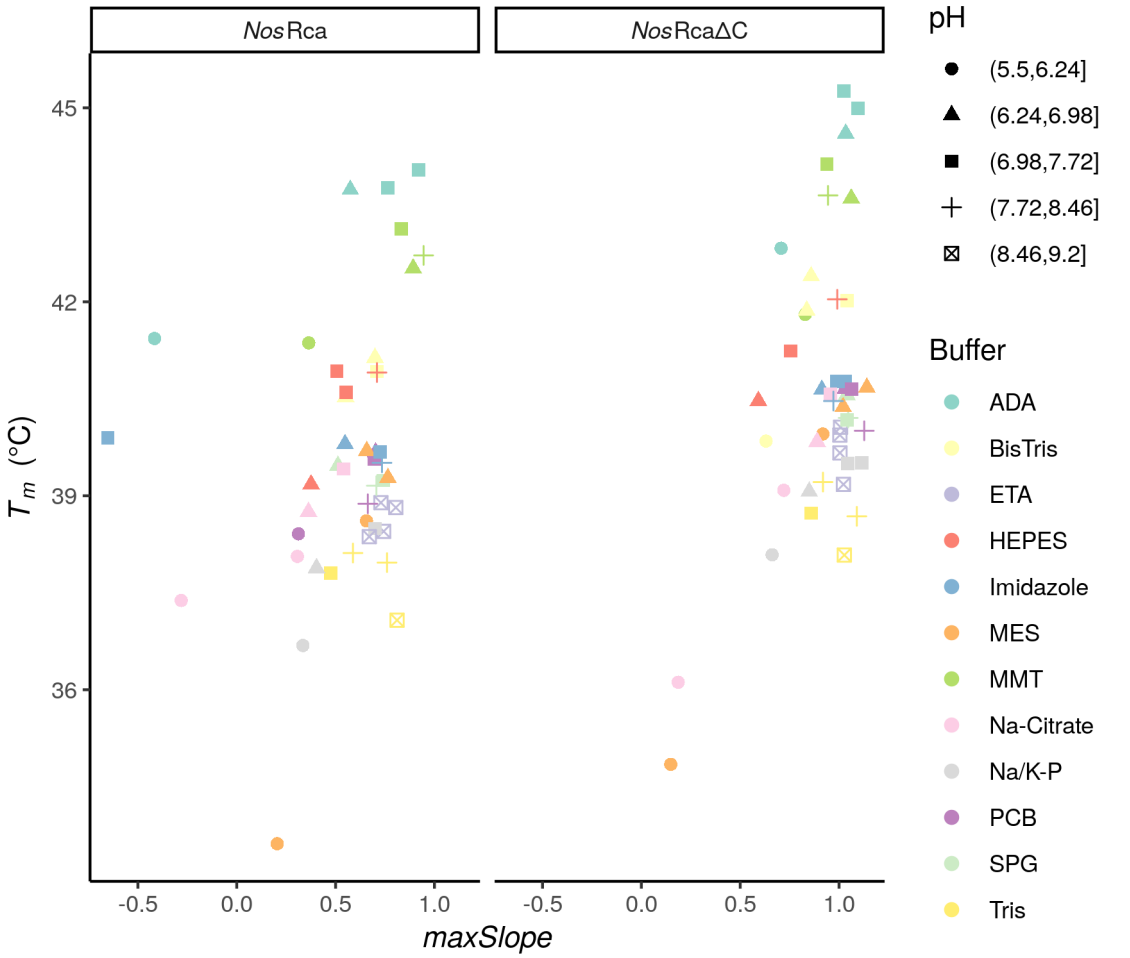

Figure A.5.: Correlation of Tm and maxSlope.
Correlating T_m and *maxSlope* reveals that high T_m coincide with steep *maxSlopes*. Buffer and pH are shown on discrete scales and the pH interval borders are given in the figure legend. The buffer substance is indicated by color, the pH range by the symbol shape. See Tables **2.5** and **2.4** for buffer details.

resentative examples for the purification of *Nos*Rca and *Nos*RcaΔC are shown in Figure **A.6** and **A.7**, respectively.

In comparison to *Nos*Rca, the aggregation propensity of *Nos*RcaΔC was greatly reduced which resulted in ∼ twofold higher purification yields for *Nos*RcaΔC. Mutations in the Rubisco-Rca interface did not affect aggregation propensity of *Nos*RcaΔC V91E/R92G/G93N, *Nos*RcaΔC P140G/Y143A/D144A and *Nos*RcaΔC L244D/N248A/L250A as judged by the obtained protein yields after purification. However, the mutant *Nos*Rca V91E/R92G/G93N/L244D/N248A/L250A showed enhanced aggregation during Usp2 cleavage which was solved by increasing the salt concentration to 0.5 M KCl during IMAC and Usp2 cleavage (in buffer A). Notably, a mutant of *Nos*Rca, with two Arg residues within the SSUL domain replaced by Ala (*Nos*Rca R352E/R353E), showed a substantially reduced aggregation propensity.

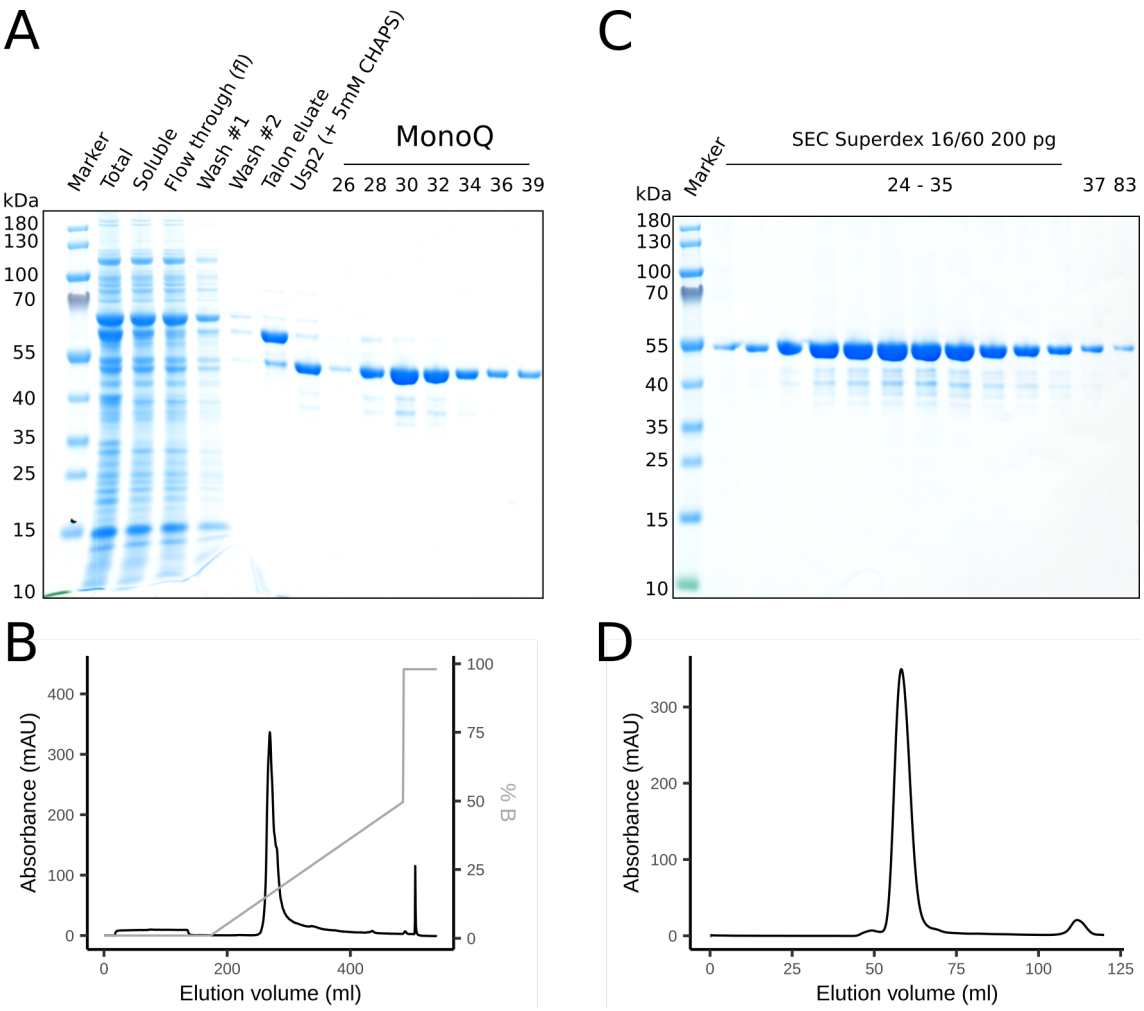

Figure A.6.: Representative purification of *Nos*Rca.
A) SDS-PAGE analysis of aliquots taken during cell lysis, IMAC (Talon) and IEX (MonoQ) purification. Lane ‚Usp2' refers to the soluble fraction from Usp2-mediated cleavage of the H_6ubi moiety. **B) Chromatogram of the IEX purification on a MonoQ HiPrep 26/10.** Protein was buffer exchanged to 10 mM KCl, loaded and washed prior to gradient elution up to 0.5 mol L^{-1} KCl. **C) SDS-PAGE analysis after SEC.** *Nos*Rca was purified to $> 95\,\%$ purity. **D) Chromatogram of the SEC purification on a Superdex 16/60 200 pg.** The single uniform peak was indicative of a monodispersed distribution of *Nos*Rca.

A.1.1.2. *Nos*Rubisco wildtype and mutants

A.1.1.2.1. Optimization of recombinant *Nos*Rubisco expression

RbcL of form I and II Rubisco is a client of the chaperonin system. While chaperonin-folded RbcL of the form II Rubisco spontaneously assembles into functional soluble dimers *in vitro* [Goloubinoff et al., 1989a], assembly of the hexadecameric form I Rubisco has failed in presence of the chaperonins alone [Liu et al., 2010; Tabita, 1999]. Notably, correctly assembled, soluble hexadecameric form I Rubiscos can be produced at low yields, when the *E. coli* chaperonins GroES and GroEL (*Ec*SEL) are co-expressed (for example, Rubisco of *Synechococcus elongatus* PCC 7942 Rubisco

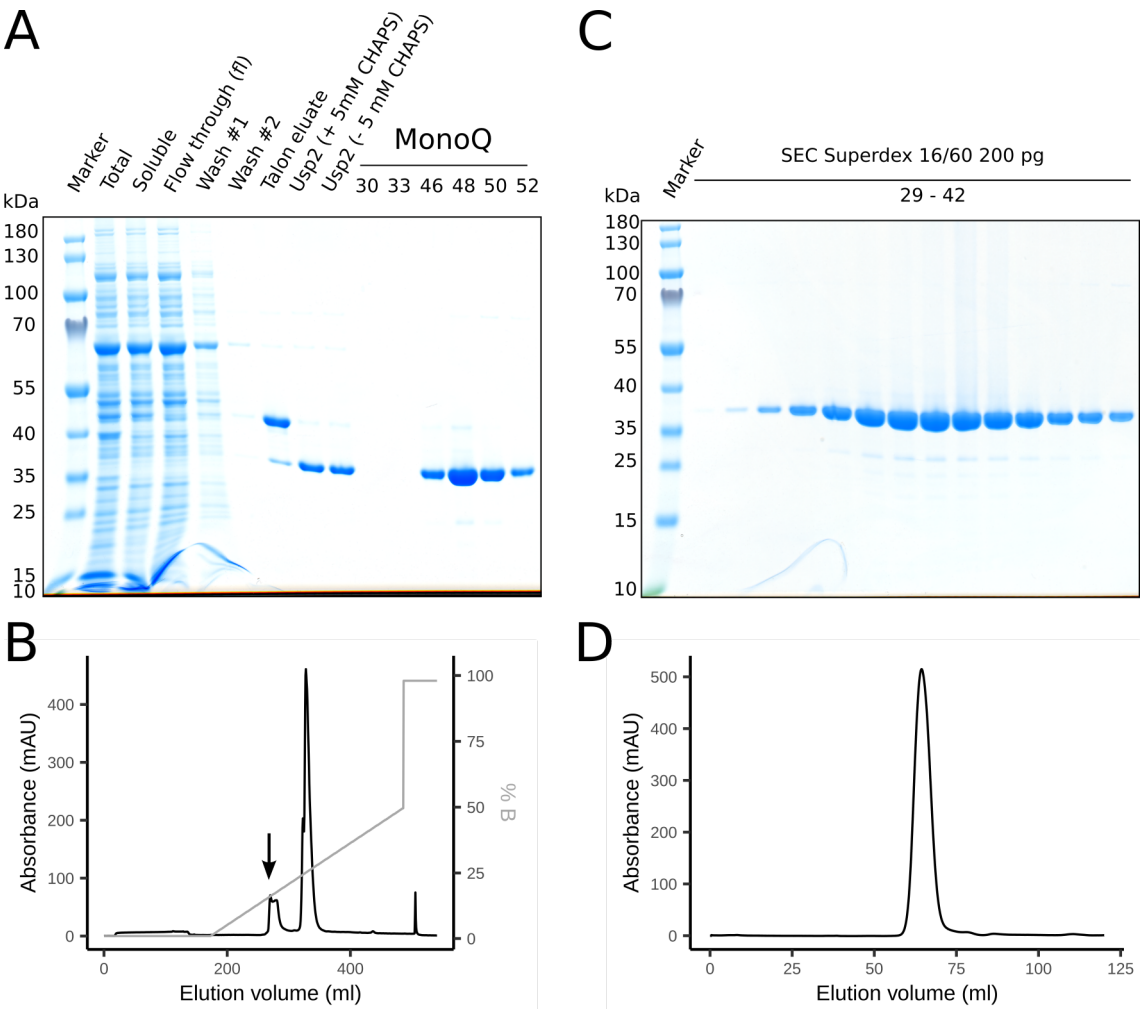

Figure A.7.: Representative purification of *Nos*RcaΔC.
A) SDS-PAGE analysis of aliquots taken during cell lysis, IMAC (Talon) and IEX (MonoQ) purification. Lane ‚Usp2' refers to the soluble fraction from Usp2-mediated cleavage of the H_6ubi moiety. **B) Chromatogram of the IEX purification on a MonoQ HiPrep 26/10.** Protein was buffer exchanged to 10 mM KCl, loaded and washed prior to gradient elution up to 0.5 mol L^{-1} KCl. Notably, the lower isoelectric point (pI) of *Nos*RcaΔC resulted in an increased affinity to the MonoQ resin. This shift in elution volume allowed the separation of *Nos*RcaΔC from a small impurity which cannot be separated upon purification of *Nos*Rca (see arrow). **C) SDS-PAGE analysis after SEC.** *Nos*RcaΔC was purified to > 95 % purity. **D) Chromatogram of the SEC purification on a Superdex 16/60 200 pg.** The single uniform peak was indicative of a monodispersed distribution of *Nos*RcaΔC.

[Goloubinoff et al., 1989b]). However, the yield of assembled form I Rubisco is greatly increased upon co-expression of the assembly chaperones RbcX or Raf1 [Emlyn-Jones et al., 2006; Hauser et al., 2015; Kolesinski et al., 2014; Li and Tabita, 1997; Onizuka et al., 2004; Saschenbrecker et al., 2007; Whitney et al., 2015]. Like in many cyanobacterial species [Huang et al., 2019], the *rbcX* gene is located in the Rubisco operon between *rbcL* and *rbcS* in *Nostoc* sp. PCC 7120, while *raf1* is located approximately 2 Mbp away from the Rubisco operon on the complement strand [Kaneko et al., 2001]. Thus, it seemed likely, that either RbcX or Raf1 or both from

Nostoc sp. PCC 7120 (*Nos*RbcX and *Nos*Raf1, respectively) would stimulate the recombinant production of the cognate Rubisco (*Nos*Rubisco).

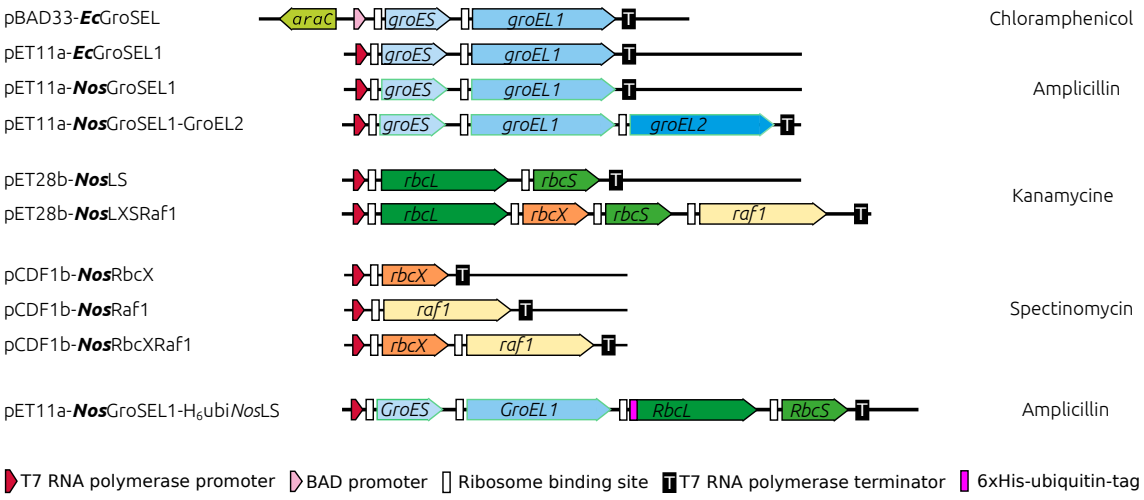

Figure A.8.: Plasmids generated for the optimization of *Nos*Rubisco expression and purification.
Apart from arabinose-inducible pBAD33-*Ec*GroSEL [McLennan and Masters, 1998], all plasmids were generated in this study.

In a first attempt to produce *Nos*Rubisco,we tested the combined requirement of *Nos*RbcX and *Nos*Raf1 for the expression of *Nos*Rubisco in presence of co-overexpressed *Ec*SEL. Initially, *Ec*SEL was expressed from the arabinose-inducible pBAD33 backbone (pBAD33-*Ec*SEL, [Kerner et al., 2005; McLennan and Masters, 1998]), while *Nos*Rubisco was expressed either with or without *Nos*RbcX and *Nos*Raf1 from the IPTG-inducible pET28b backbone (pET28b-*Nos*LXSRaf1 and pET28b-*Nos*LS, respectively - Figure **A.8**). Cells harboring the pBAD33-*Ec*SEL plasmid together with pET28b-*Nos*LXSRaf1 produced soluble Rubisco subunits upon induction of protein expression, while cells instead carrying the pET28b-*Nos*LS plasmid did not, which indicated that either Raf1 or RbcX or both were required for *Nos*Rubisco production in presence of *Ec*SEL (Figure **A.9**). According to SDS-PAGE analysis not only Rubisco but also Raf1 and RbcX expression varied substantially between individual expression batches and the partially purified Rubisco displayed variable properties. Furthermore, we noticed that cell cultures harboring the pET28b-*Nos*LXSRaf1 plasmid occasionally lysed upon IPTG induction, suggesting that either one or multiple of the encoded proteins were toxic to the expression host. In the attempt to increase the expression levels of assembly factors, we co-expressed *Nos*RbcX and *Nos*Raf1 from a third plasmid (pCD1Fb-*Nos*RbcX-*Nos*Raf1) in addition to pET28b-*Nos*LXSRaf1 and pBAD-*Ec*SEL. However, this mostly lead to cell lysis and the overall breakdown of expression cultures and Rubisco expression was

abolished in the remaining cells, suggesting that either Raf1, RbcX or both are toxic to *E. coli* cells at elevated concentrations (data not shown).

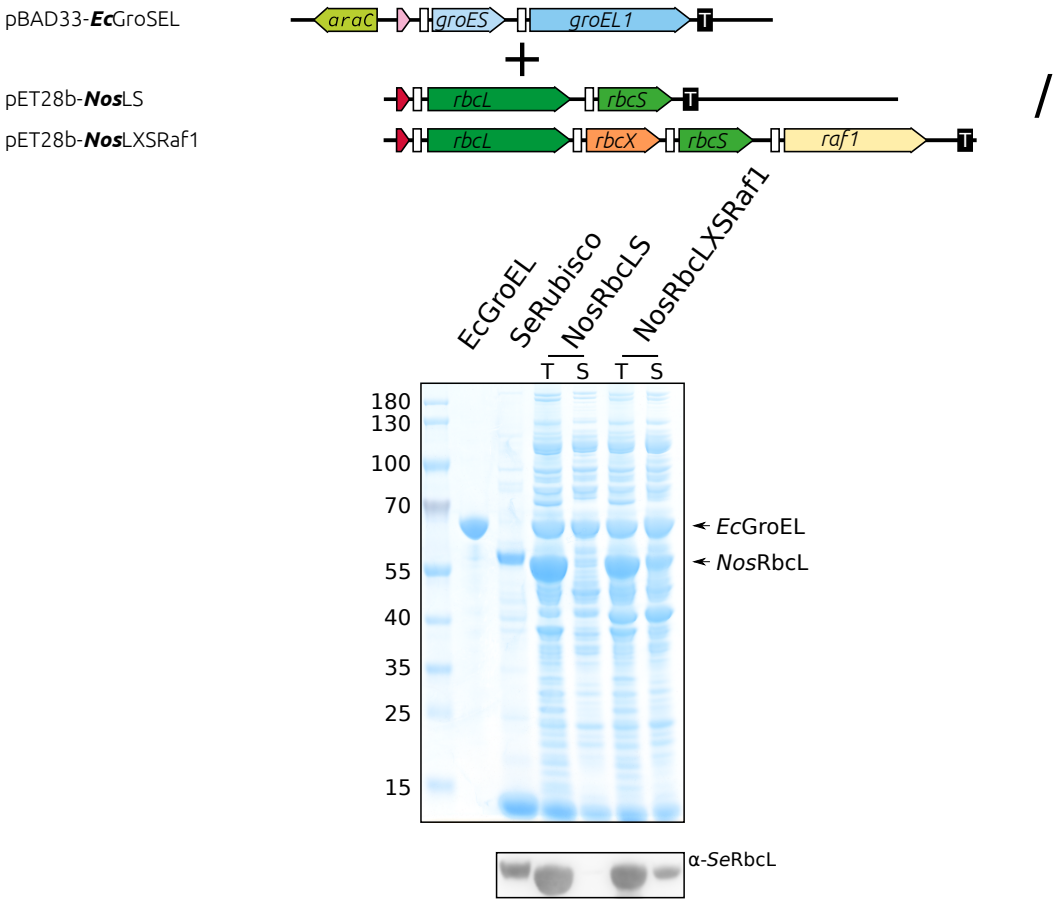

Figure A.9.: RbcL folded by *Ec*SEL requires *Nos*RbcX or *Nos*Raf1 or both for assembly to the functional holoenzyme.
BL21 cells harbored both the arabinose inducible pBAD33-*Ec*GroSEL and IPTG inducible pET28b-*Nos*LS or pET28b-*Nos*LXSRaf1, respectively. After protein expression and cell lysis, the lysate was fractionated by centrifugation at 25000 x g and the unfractionated lysate (T) and the supernatant (S) were analyzed by SDS-PAGE and immunoblotting with α-*Se*RbcL.

Because the toxic effect of one or both assembly factors seemed to limit recombinant *Nos*Rubisco production, we sought to test the contribution of the assembly chaperones and endogenous chaperonins individually. Unlike other heterocyst forming filamentous cyanobacteria, which often possess three GroEL homologs (two in operon with GroES and one additional monocistronic GroEL), *Nostoc* PCC 7120 encodes one GroEL in operon with GroES (*Nos*SEL) and one additional monocistronic GroEL (*Nos*EL2). The operon-located GroEL is thought to be the ancestor of plant Cpn60α/β [Suzuki et al., 2009; Weissenbach, 2017] and is essential in *S. elongatus* PCC 7942 [Sato et al., 2007] unlike GroEL2 [Sato et al., 2008]. Moreover, it is more folding competent for *S. elongatus* PCC 7942 Rubisco than GroEL2 [Kovács et al., 2001]. Therefore, we reasoned that *Nos*SEL fulfills a house-keeping function in *Nostoc* sp. PCC 7120 and thus would be able to fold *Nos*RbcL. However, to rule out

contributions of *Nos*EL2, which previously has been implicated in the adaptation to sub optimal growth temperatures [Sato et al., 2008], we analyzed the effect of either *Nos*SEL-EL2 or *Nos*SEL in comparison to *Ec*SEL on soluble Rubisco production and activity in presence of the assembly chaperones *Nos*Raf1 and *Nos*RbcX (Figure **A.10**). Not surprisingly, we found that co-expression of *Nos*SEL stimulated production of assembled *Nos*Rubisco $\sim$ twofold as analyzed by native PAGE and immunoblotting with α-*Se*RbcL and by measuring Rubisco activity in the cleared lysates. Additional co-expression of *Nos*EL2 did not improve Rubisco yield or activity.

Next, we aimed at analyzing the individual contribution of *Nos*RbcX and *Nos*Raf1 on Rubisco production in cells expressing *Nos*SEL-EL2 and *Nos*Rubisco from pET11a-*Nos*SEL-EL2 and pET28b-*Nos*LS, respectively. Surprisingly, already in absence of additional assembly factors we found that *Nos*Rubisco was expressed at high levels (Figure **A.11**), whereas co-overexpression of *Ec*SEL and *Nos*LS failed to produce any soluble Rubisco as observed previously (Figure **A.9**). Upon coexpression of Raf1 the amount of soluble Rubisco increased slightly as analyzed by SDS- and native PAGE and Rubisco activity in cleared lysates was approximately 10 % higher in comparison to Rubisco expressed in absence of assembly factors. Moreover, co-expression of RbcX increased the amount of soluble Rubisco to a level similar to Raf1, however Rubisco activity was reduced by nearly 50 % despite of a large pool of free *Nos*RbcS. Besides, bands corresponding to Rubisco complexes on native PAGE appeared broader when Raf1 or RbcX were co-expressed, perhaps pointing to the presence of multiple assembly intermediates. Furthermore, immunodetection of RbcS and RbcL in native PAGE analysis did not suggest a drastically altered RbcL:RbcS ratio in the Rubisco complex upon coexpression with Raf1 or RbcX. When RbcX and Raf1 were simultaneously co expressed, both the amount of soluble Rubisco as well as Rubisco activity in the cleared lysates were greatly diminished. Notably, the expression of Raf1 was also diminished in these expression tests. In summary, these results indicate that likely the joint co-expression of RbcX and Raf1 was responsible for the variable growth phenotype of cultures expressing Rubisco from the pET28b-*Nos*LXSRaf1 plasmid and limited Rubisco production in *E. coli*.

As pointed out before, *Ec*SEL-folded RbcL strictly required the assistance of assembly chaperones for soluble *Nos*Rubisco production, unlike *Nos*SEL-folded RbcL. To analyze whether co-expression of *Ec*SEL or *Nos*SEL affects *Nos*Rubisco production indeed differentially, we repeated the previous experiment, but expressed both *Ec*SEL and *Nos*SEL from the pET11a vector to rule out any contributions from the plasmid backbone (Figure **A.12**). *E. coli* Bl21 star cells harboring either

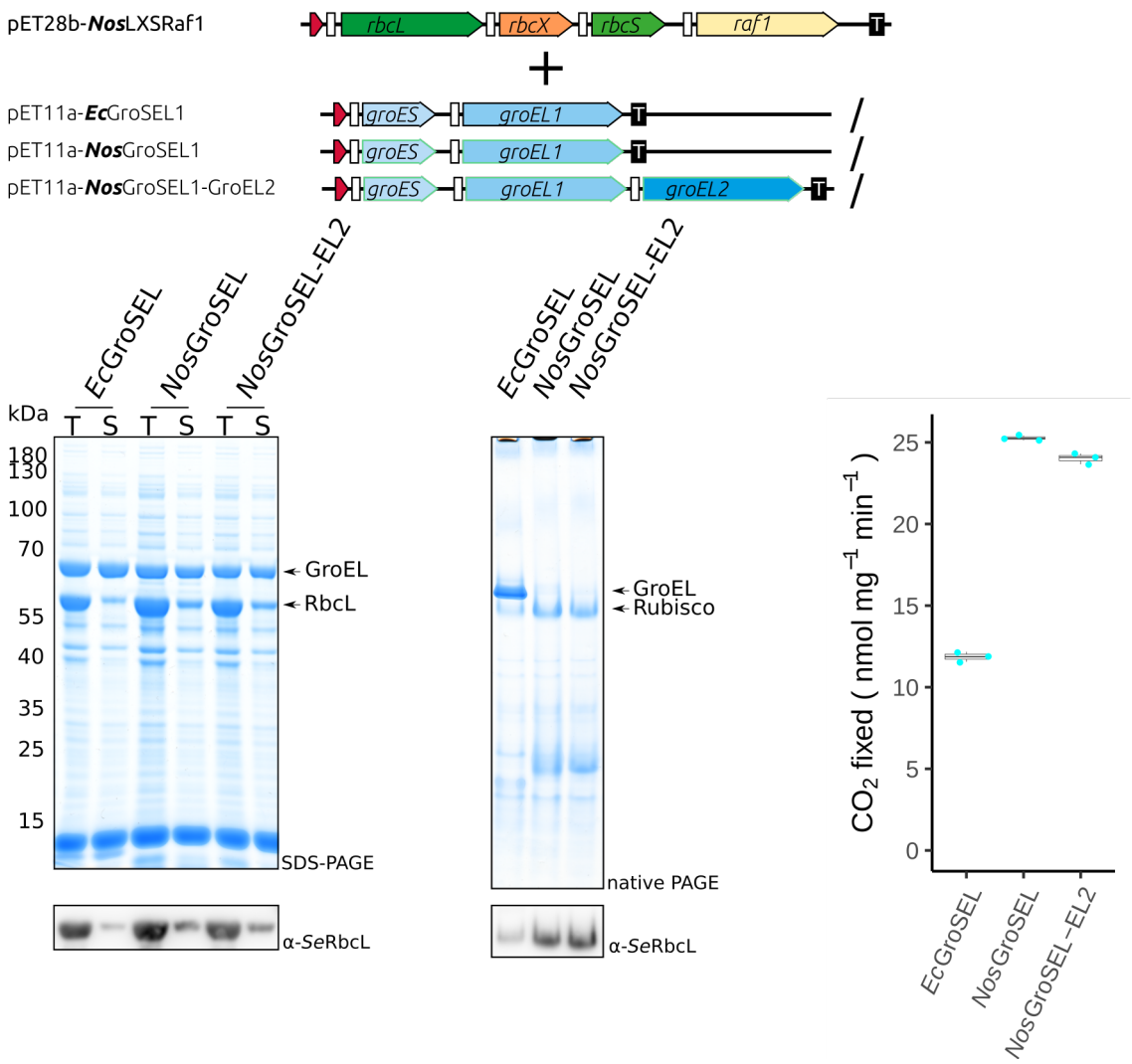

Figure A.10.: Co-expression of endogenous chaperonins increases *Nos*Rubisco expression.
BL21 star cells harbored both IPTG inducible pET28b-*Nos*LXSRaf1 and either pET11a-*Ec*GroSEL, pET11a-*Nos*GroSEL or pET11a-*Nos*GroSEL-EL2, respectively (**top**). After protein expression and cell lysis the total (T) and soluble (S) protein content was analyzed by SDS-PAGE (**left**). Additionally, the soluble protein was analyzed by native PAGE (**middle**) and assayed for Rubisco activity in ^{14}C carbon fixation experiments (**right**). Both native- and SDS-PAGE were blotted and probed with the α-*Se*RbcL antibody to detect *Nos*RbcL. The values in the ^{14}C fixation experiments are stated in relation to the total protein content of the lysate (per mg), which was analyzed according to [Bradford, 1976]. The box and whiskers plots (Tukey) represent three technical replicates.

pET11a-*Nos*SEL or pET11a-*Ec*SEL were transformed with pET28b-*Nos*LS and Rubisco expression was carried out simultaneously by IPTG induction. Analysis of Coomassie blue-stained SDS-PAGE gels indicated that both *Ec*SEL and *Nos*SEL were expressed at similar amounts. As previous experiments suggested, sole co-expression of *Nos*SEL was sufficient for strong expression of soluble *Nos*Rubisco, whereas upon co-expression of *Ec*SEL no soluble Rubisco was detected, albeit the

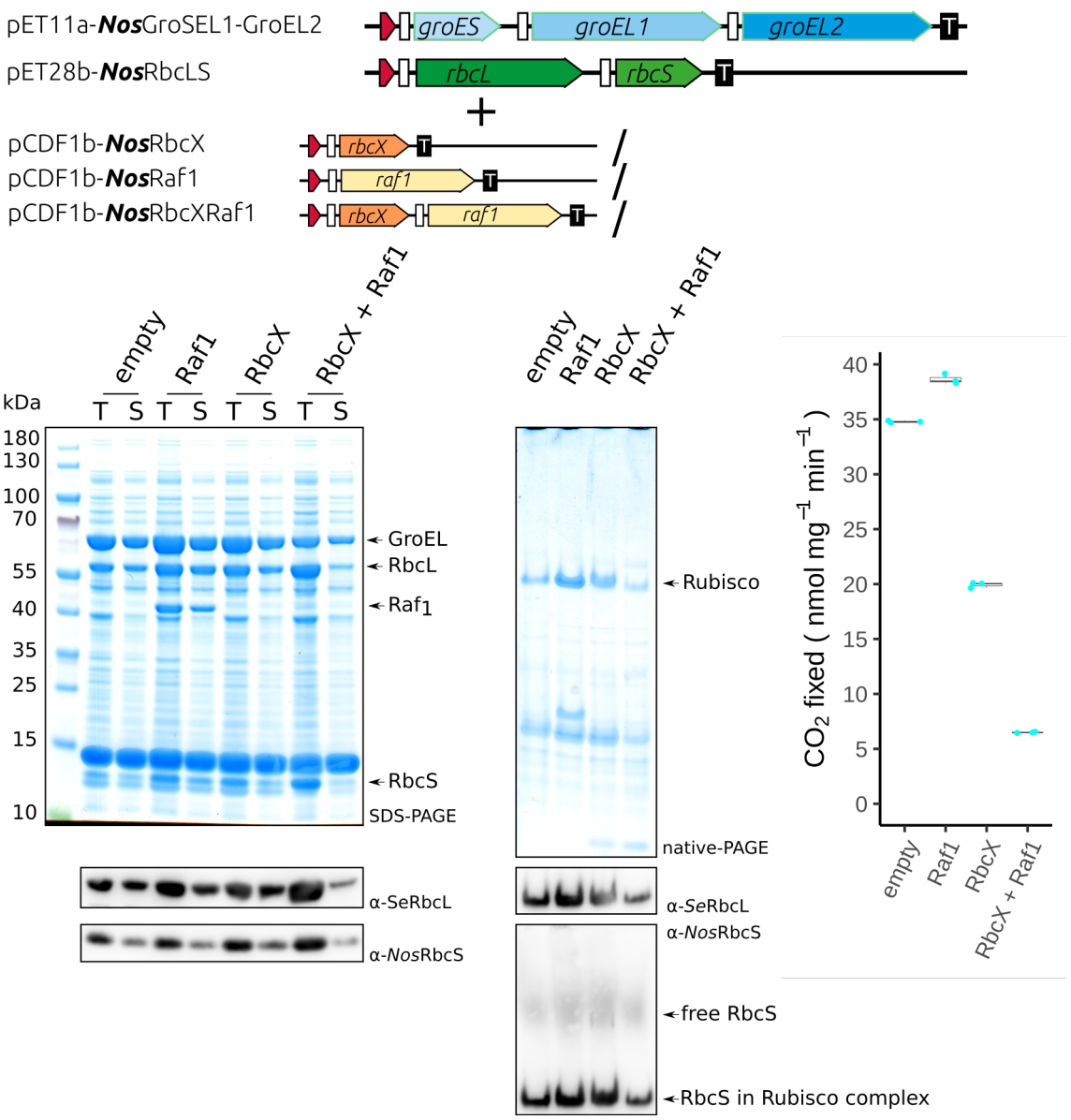

Figure A.11.: Co-expression of endogenous *Nos*Raf1 increases soluble *Nos*Rubisco content and enhances *Nos*Rubisco activity slightly.
BL21 star cells harbored both IPTG inducible pET11a-*Nos*GroSEL-EL2 and pET28b-*Nos*LS (**top**) and either pCDF1b-empty, pCDF1b-*Nos*RbcX, pCDF1b-*Nos*Raf1, or pCDF1b-*Nos*RbcX-*Nos*Raf1, respectively. After protein expression and cell lysis the total (T) and soluble (S) protein content was analyzed by SDS-PAGE (**left**). Additionally, the soluble protein was analyzed by native PAGE (**middle**) and assayed for Rubisco activity in ^{14}C carbon fixation experiments (**right**). Both native- and SDS-PAGE were blotted and probed with the *α-Se*RbcL and *α-Nos*RbcS antibodies to detect *Nos*RbcL and *Nos*RbcS, respectively. The values in the ^{14}C fixation experiments are stated in relation to the total protein content of the lysate (per mg), which was analyzed according to [Bradford, 1976]. The box and whiskers plots (Tukey) represent three technical replicates.

total expression levels of *Nos*RbcL were similar. Notably, native PAGE analysis followed by immunoblotting revealed one dominant assembly state corresponding to the functional Rubisco holoenzyme and no intermediate assembly states could be detected.

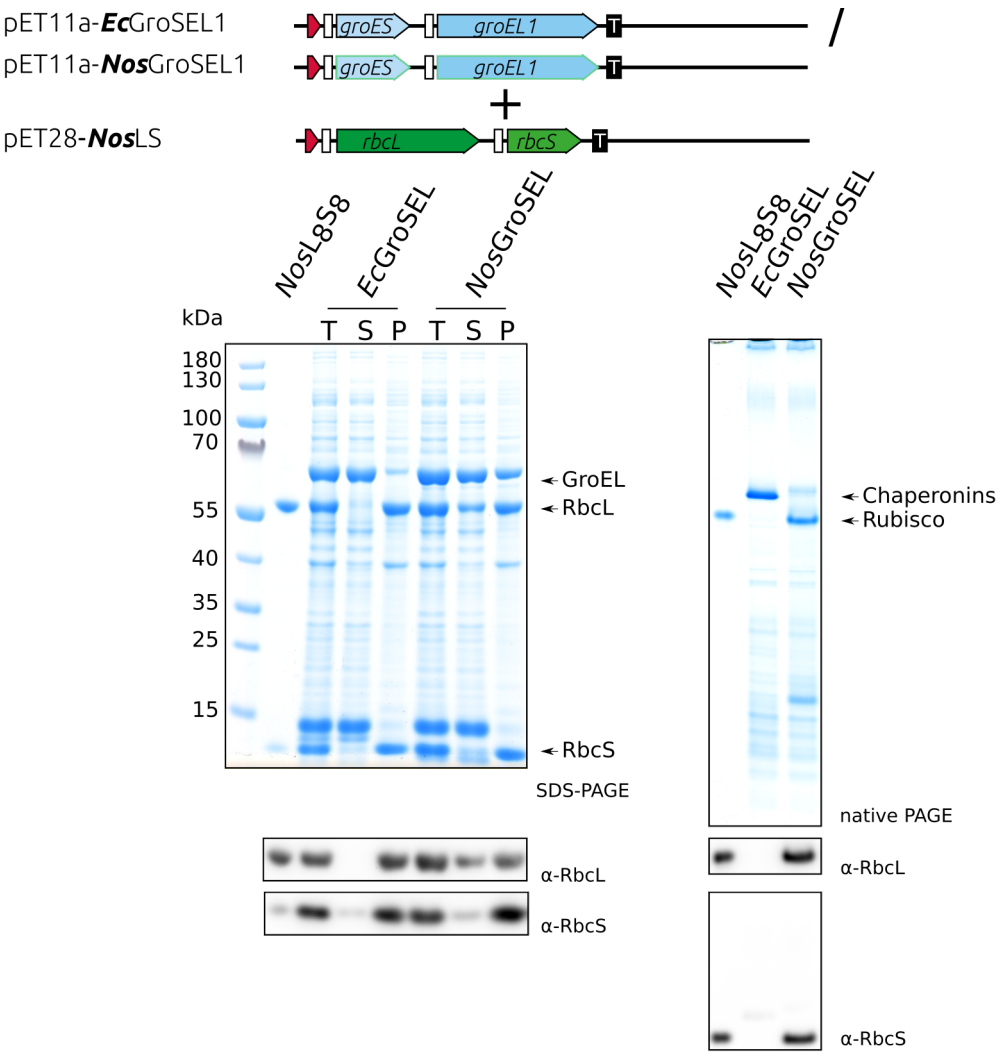

Figure A.12.: Overexpression of *Nos*SEL, but not *Ec*SEL is sufficient for the production of recombinant Rubisco in *E. coli*.
BL21 star cells harbored pET28b-*Nos*LS (**top**) and either pET11a-*Ec*GroSEL or pET11a-*Nos*GroSEL, all IPTG inducible. After protein expression and cell lysis the total (T) and soluble (S) protein content was analyzed by SDS- (**left**) and native-PAGE (**right**). Both native- and SDS-PAGE were blotted and probed with the α-*Se*RbcL and α-*Nos*RbcS antibodies to detect *Nos*RbcL and *Nos*RbcS, respectively. The experiment was conducted in two independent replicates from individual transformed colonies.

Since the contribution of *Nos*Raf1 on soluble and active Rubisco production was minor, we decided to produce *Nos*Rubisco in the absence of *Nos*Raf1 to prevent heterogeneous assembly states in the final product. Although it was previously demonstrated, that fusion of a His_6 peptide C-terminal to RbcS does neither affect functionality nor reactivation of the recombinant plant Rubisco, we still aimed to purify Rubisco using the cleavable H_6ubi affinity tag to prevent possible contributions of the His_6 moiety on the outcome of our analyses. Thus, *Nos*Rubisco was expressed as N-terminal H_6ubi-RbcL fusion protein from pET28b-H_6ubi-*Nos*LS and *Nos*SEL was co-expressed from pET11a-*Nos*SEL. The H_6ubi tag affected folding or

assembly of RbcL only marginally as judged by the soluble Rubisco yield. However upon scale up from 0.1 L test expression cultures to 1 L cultures for batch expression, soluble Rubisco yields decreased severely and repeatedly. To solve this, we moved H_6ubi-*Nos*LS into the pET11a-*Nos*SEL plasmid (pET11a-*Nos*SEL-H_6-ubi*Nos*LS) to express *Nos*SEL and H_6ubi-*Nos*Rubisco jointly from a single approximately 4.2 kb transcript. Using this approach, soluble and active *Nos*Rubisco could be routinely expressed at high levels corresponding to approximately 5 - 10 % of the soluble proteins and with consistent quality. A representative purification, here ΔN*Nos*Rubisco, is shown in Figure **A.13** in chapter **A.1.1.2.2**.

A.1.1.2.2. Purification of *Nos*Rubisco and mutants

The expression of *Nos*Rubisco and corresponding mutants was presented in detail in chapter **A.1.1.2.1**. *Nos*Rubisco and corresponding mutants were purified by two consecutive IMAC steps and subsequent SEC. The H_6ubi moiety was cleaved after the first IMAC and removed during the second IMAC step, together with Usp2 and any uncleaved protein as described in Chapter **2.5.3**. A representative purification, here ΔN*Nos*Rubisco, is shown in Figure **A.13**.

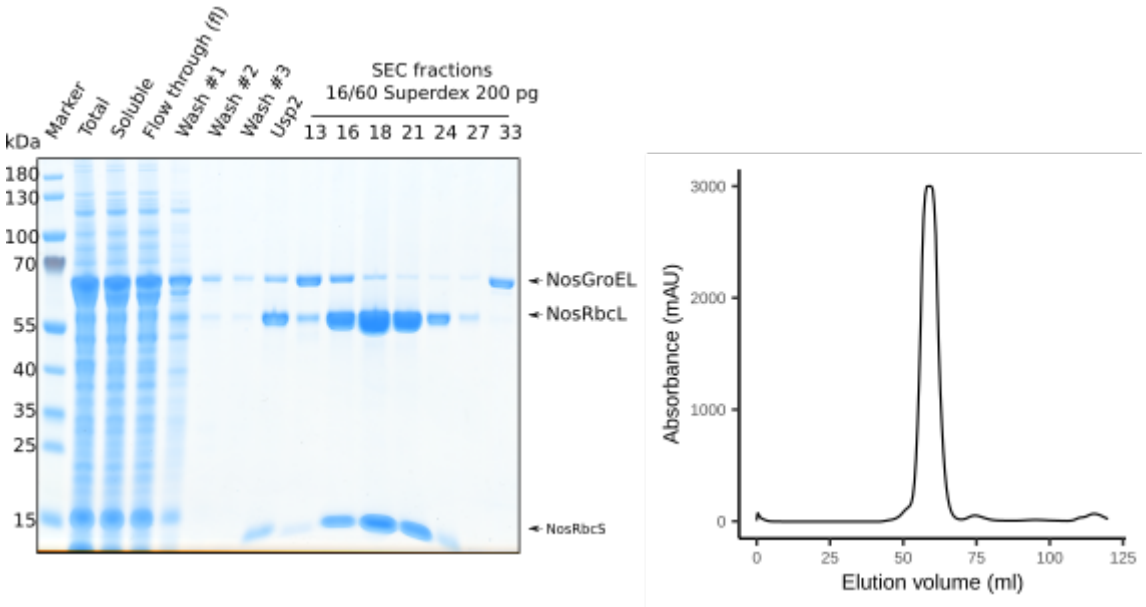

Figure A.13.: Representative purification of NosRubisco and mutants.
Left: **SDS-PAGE analysis of aliquots taken during cell lysis, IMAC I (Talon) and SEC.** Here, purification of *Nos*L_8S_8ΔN9 is shown as a representative example. Notably, H_6ubi-*Nos*RbcL cannot be discriminated from *Nos*GroSEL by SDS-PAGE analysis and Coomassie staining due to their similar size. Lane ‚Usp2' refers to the Usp2 mediated cleavage of the H_6ubi-tag. Right: **Chromatogram of the SEC purification on a Superdex 16/60 200 pg.** The single nearly uniform peak was indicative of a monodisperse distribution of *Nos*Rubisco. *Nos*Rubisco containing fractions 18-21 were pooled, concentrated and stored in liquid nitrogen. The final product was > 95 % pure as judged by SDS-PAGE analysis.

A.1.1.3. *Nos*SSUL, *Nos*M35 and *Nos*RbcS

*Nos*SSUL, *Nos*RbcS and *Nos*M35 were purified by two consecutive IMAC steps followed by SEC as described in chapter **2.5.5**. Representative purifications are shown in Figures **A.14**, **A.16** and **A.15**.

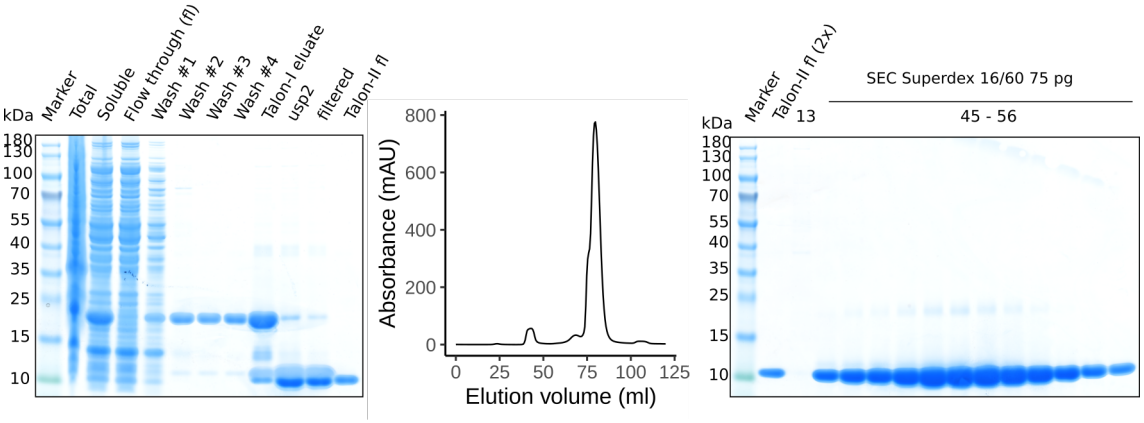

Figure A.14.: Purification of *Nos*SSUL.
From left to right: **SDS-PAGE analysis of aliquots taken during cell lysis, IMAC I and II (Talon) purification.** Lane ‚Usp2' refers to the Usp2-mediated cleavage of the His_6-ubi-tag. After cleavage of H_6ubi-moiety, the soluble fraction was passed through a 0.22 µm syringe filter (PVDF membrane) to remove any protein aggregates (lane ‚filtered'). **Chromatogram of the SEC purification on a Superdex 16/60 75 pg.** The single nearly symmetrical peak was indicative of a monodisperse distribution of *Nos*SSUL. **SDS-PAGE analysis after SEC.** *Nos*SSUL was purified to $> 95\,\%$ purity.

*Nos*M35 and *Nos*SSUL displayed a strong affinity toward cobalt ions. This became apparent upon high imidazole elution of these proteins from the TALON cobalt-affinity resin, which colored the protein containing fractions a striking blue, suggestive of the presence of cobalt (II). After SEC in presence of the reducing agent DTT and protein concentration, the reduced protein appeared brown, in contrast to the protein purified in absence of DTT, suggesting that addition of DTT modified the cobalt-bound protein. Initial attempts to remove the bound cobalt by EDTA chelation (up to 5 mM) and subsequent gel filtration failed. To thoroughly remove the cobalt ion before binding to the protein, the purification protocol was adapted as follows: Prior to cleavage by Usp2 the TALON I eluate was treated with 5 mM EDTA and buffer exchanged into 50 mM MMT pH 8.4 / 100 mM KCl / 10 mM $MgCl_2$ / 5 % glycerol. The cleaved protein was buffer exchanged in the same buffer to remove the β-mercaptoethanol prior to the application to TALON II, which was followed by a second EDTA treatment before the final gel filtration step. This procedure apparently removed all bound cobalt ions, albeit in case of *Nos*M35, the final product was found to be instable as with prolonged storage multiple cleavage

products appeared on SDS-PAGE. Therefore, the cobalt-bound (brownish) protein was used in this study.

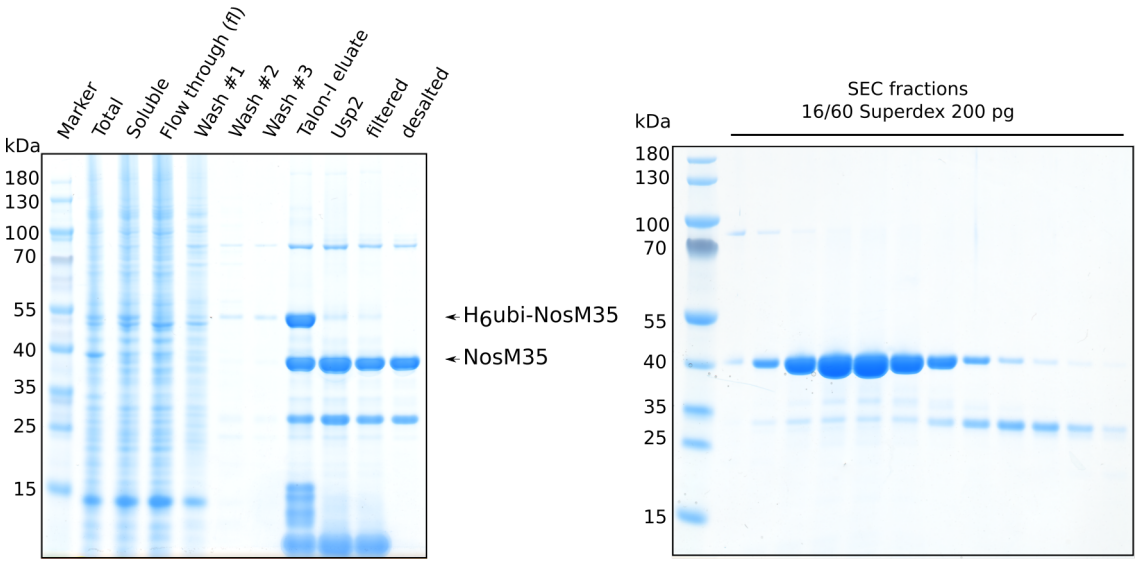

Figure A.15.: Purification of *Nos*M35.
Left: **SDS-PAGE analysis of aliquots taken during cell lysis and IMAC I (Talon) purification of ox*Nos*M35.** Lane ‚Usp2' refers to the Usp2-mediated cleavage of the H_6ubi moiety. After cleavage of H_6ubi-moiety, the soluble fraction was passed through a 0.22 µm syringe filter (PVDF membrane) to remove any protein aggregates (lane ‚filtered'). Right: **SDS-PAGE analysis after SEC.** ox*Nos*M35 was purified to $> 95\,\%$ purity.

RbcS of cyanobacteria and eukaryotes is only poorly soluble and recombinant RbcS is frequently purified from inclusion bodies [Saschenbrecker et al., 2007]. Surprisingly, the N-terminal fusion of the H_6-ubi moiety to *Nos*RbcS was found to be entirely soluble (Figure **A.16**). After Usp2-mediated cleavage of H_6ubi, *Nos*RbcS stayed soluble and showed signs of slight aggregation only at high protein concentration. The chromatogram of the final purification step by SEC showed a clear double-peak elution profile for *Nos*RbcS. The peaks corresponded to *Nos*RbcS dimers and monomers, respectively, according to SEC-MALS analysis of individual fractions. *Nos*RbcS was purified with reasonable yield (7 mg pure protein from 1 L culture) and could be concentrated to at least $8\,g\,L^{-1}$.

A.1.2. CABP Synthesis

CABP was synthesized according to a published protocol by Pierce et al. [1980], as described in chapter **2.6**. The conversion rate of a typical reaction was between 20 - 35 % (of maximum 50 % as the stereoisomers CABP and CRBP are produced at equal amounts). The elution profile was reproducible between runs and fit previously published data well [Pierce et al., 1980]. An attempt to quantify CABP purity by

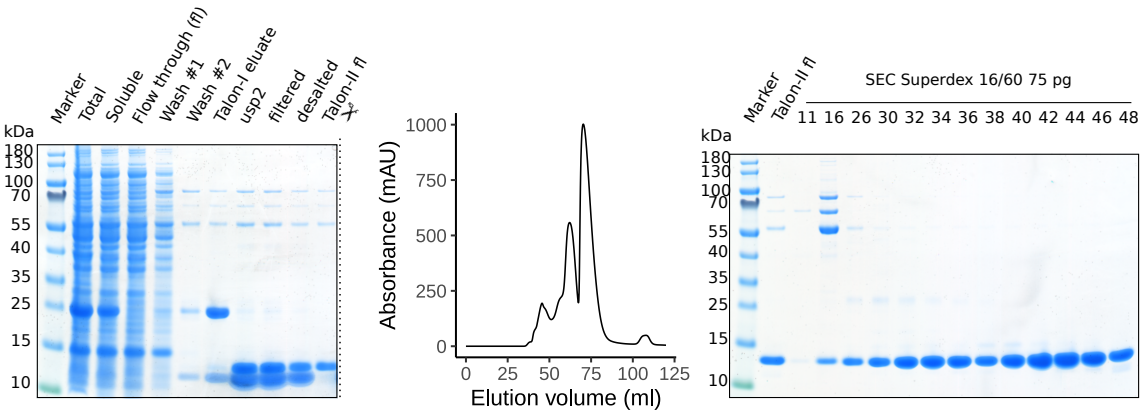

Figure A.16.: Purification of NosRbcS reveals redox dependent dimerization. From left to right: **SDS-PAGE analysis of aliquots taken during cell lysis, IMAC I and II (Talon) purification.** Lane ‚Usp2' refers to the Usp2-mediated cleavage of the His_6-ubi-tag. After cleavage of H_6ubi-moiety, the soluble fraction was passed through a 0.22 µm syringe filter (PVDF membrane) to remove any protein aggregates (lane ‚filtered'). **Double-peak elution profile of *Nos*RbcS on a Superdex 16/60 75 pg.** The peaks correspond to *Nos*RbcS dimers and monomers, according to subsequent SEC-MALS analysis of individual fractions (Collaboration with M. Hayer-Hartl). **SDS-PAGE analysis after SEC.** *Nos*SSUL was efficiently separated by larger impurities and was purified to $> 95\,\%$ purity. Notably, the double peak elution profile is also visible on SDS-PAGE.

proton NMR was not successful, due to the strong background signal of the Bicine buffer. However, every fresh CABP batch was assayed in its ability to inactivate *Nos*Rubisco. At an equimolar ratio (CABP:active site) 100 % inhibition was always met after max. 10 min.

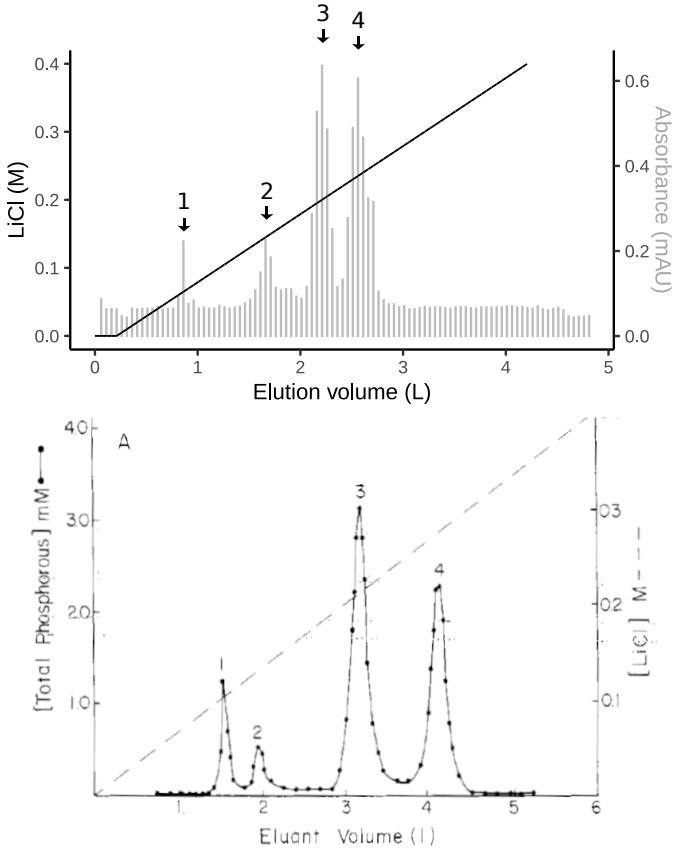

Figure A.17.: Elution profile of the racemic mixture of CRBP and CABP.
Top panel: To separate the enantiomers CRBP and CABP, the racemic mixture was loaded onto a 120 mL AG1-X8 Cl^- column equilibrated in 3 mM HCL. Products were eluted by gradient elution (0 - 0.4 mol L^{-1} LiCl / 4 L gradient) and subsequently 50 mL fractions were assayed for total phosphate according to [Chifflet et al., 1988] by UV-VIS spectroscopy of the malachite green phosphomolybdate complex at 650 nm. Peaks were assigned based on published results by Pierce et al. [1980] (**Bottom panel**) **1)** carboxyarabinitol-P_2 **2)** carboxyribitol-P_2 **3)** carboxyribitol-P_2 γ-lactone **4)** carboxyarabinitol-P_2 γ-lactone.