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

Mechanisms of folding, assembly and remodelling chaperones

Goran Miličić

aus

Ljubljana, Slowenien

2017

Erklärung

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

Eidesstattliche Versicherung

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München, 26.06.2017

Goran Miličić

Dissertation eingereicht am: 05.12.2017

1.Gutachter: Prof. Dr. F. Ulrich Hartl

2.Gutachter: PD Dr. Dietmar E. Martin

Mündliche Prüfung am: 17.01.2018

Acknowledgments

I would like to thank Prof. Dr. F. Ulrich Hartl and Dr. Manajit Hayer-Hartl for accepting me to their group and giving me the opportunity to conduct my PhD research on a variety of exciting topics in collaboration with very smart and talented people. The freedom that was given to me during these years allowed me to grow on both scientific and personal levels and, most importantly, to attain the level of independence that I always wished for. Last but not least, their critical approach and questioning of every aspect helped me to start thinking about reality in the way it is and not as how it should be.

I am thankful to all the people who I worked with during the years: Dr. Amit Gupta and Dr. Shubhasis Haldar who I collaborated with on single molecule fluorescence projects, Dr. Thomas Hauser and Dr. Andreas Bracher who I worked with in studying the Raf1 assembly chaperone and Dr. Javaid Bhat who I collaborated with on the structural and mechanistic characterization of RsRca. I especially thank Prof. Dr. Petra Wendler who was always there to answer and clarify many of my questions regarding single particle EM data processing. I am grateful to Dr. Mike Strauss for teaching me how to prepare cryo EM samples and how to properly use the electron microscope. I would like to thank Xiao Yan, Amit Singh, Rahmi Imamoğlu for all the discussions on scientific and non-scientific topics. Last but not least I would like to thank Dr. David Balchin for many great discussions, exchange of exciting ideas and for comments on my thesis.

Nothing would have been possible without the help of Darija, Emanuel, Albert, Nadine, Romy, Ana and Silvia who make sure that the lab runs smoothly and that we can actually do our experiments. I would also like to thank all other people in the lab for the nice working atmosphere.

I am thankful to my TAC committee, Prof. Dr. Ulrich F. Hartl, Prof. Dr. Walter Neupert, Prof. Dr. Petra Wendler and Dr. Manajit Hayer-Hartl for their critical assessment of my on-going work during the years. I also have to thank members of the IMPRS office, Dr. Hans Joerg Schaeffer, Dr. Ingrid Wolf and Maximilliane Reif for their support and organization of interesting lectures, workshops and other IMPRS events.

I also want to express gratitude to members of my PhD committee for critical evaluation of my thesis.

Finally, I want to deeply thank my father Slavko, mother Lidija and sister Vesna for their love, support and encouragement in those moments when it seemed that nothing makes sense anymore. It was because of them that I stayed on the right path and not have lost the goal.

Contents

1 Summary

Proteins are the workhorses of the cell. They are involved in the collective of processes required for life. Together, the proteins of the cell constitute the proteome, and can number from 10 000 to 20 000 proteins. Increasing cellular complexity is associated with more complex proteomes. For cell survival, it is of paramount importance to preserve proteome integrity when challenged by outside stressors such as heat, oxidation, changes in pH and the presence of toxic elements. The term proteostasis (protein homeostasis) is used to describe a cell's ability to keep its proteome stable and functional in spite of such outside insults. To achieve this, cells have developed a proteostasis network of interacting proteins. Chaperones represent the majority of nodes in the proteostasis network. They help proteins to fold, assemble into oligomeric structures, maintain their structure, repair them and prevent them from aggregating.

The aim of this thesis was to investigate protein folding, assembly and repair by molecular chaperones. The chaperone system GroEL/GroES was used as a model to study in detail the mechanism of assisted protein folding. Rubisco, the enzyme that incorporates atmospheric CO_2 into the five carbon sugar ribulose-1,5-bisphosphate (RuBP) in the Calvin-Benson-Bassham cycle (CBB cycle), was used as a model substrate for studying chaperones of protein assembly and conformational repair.

The first article to which this thesis contributed, addressed in detail the mechanism by which GroEL/ES folds a model substrate called double mutant maltose binding protein (DM-MBP). GroEL and GroES form a nano-cage for a single protein molecule to fold in isolation. By using a single-molecule fluorescence approach to exclude aggregation during folding, we demonstrated that GroEL/ES plays an active role in DM-MBP folding, enhancing its folding rate up to 8-fold. Next, we showed that repetitive binding, conformational stretching and release of DM-MBP are not required for assisted folding since a single-round encapsulation event by GroES also resulted in accelerated folding. Lastly, we established that the environment of the GroEL cavity is crucial for catalysing DM-MBP folding, thus demonstrating that GroEL/ES is not a mere aggregation prevention device but an active folding catalyst.

In the second article, we investigated the conformational cycle of GroEL/ES. In the canonical description of the cycle, GroEL binds GroES on one side only, forming an asymmetrical complex (GroEL:GroES) that functions as a two-stroke machine. Under certain conditions, however, it is possible that two GroES molecules bind to GroEL, forming a symmetrical complex ($GroEL:GroES_2$). A single molecule fluorescence approach using two differently labelled populations of GroES allowed us to measure the amount of symmetrical complexes under a variety of conditions. We showed that $GroEL:GroES_2$

complexes are formed only in presence of non-foldable substrates, such as α-lactalbumin (α-LA) and α-casein, but not when challenged with foldable substrates like *Rhodospirillum rubrum* Rubisco, mitochondrial malate dehydrogenase (mMDH), mitochondrial rhodanese (mRho) or DM-MBP, or in the absence of a substrate. Additionally, we found that physiological ratios of ATP/ADP decrease the amount of symmetrical complexes. Taken together, we concluded that the asymmetrical complex is the prevalent form under physiological in vitro conditions.

The third article describes the structure of the assembly chaperone Rubisco accumulation factor 1 (Raf1), and its role in assembly of the Rubisco holoenzyme. Rubisco consists of eight large subunits (RbcL) arranged as a tetramer of antiparallel dimers capped by four small subunits (RbcS) on each side. We demonstrated using biochemical methods that Raf1 acts downstream of chaperonin assisted folding of RbcL. Additionally, X-ray crystallography, chemical crosslinking and single-particle negative stain electron microscopy (EM) allowed us to conclude that Raf1 in its dimeric functional form binds to RbcL, releases it from the chaperonin, facilitates formation of $RbcL_2$ antiparallel dimers, and drives the assembly of a $RbcL_8$:$Raf1_4$ complex. In the last step, RbcS replaces Raf1 and forms the Rubisco holoenzyme ($RbcL_8$:$RbcS_8$).

In the fourth and final article we looked at yet another type of a chaperone, which is involved in metabolic repair. We investigated the interaction and remodelling activity of red-type Rubisco activase from the bacterium *Rhodobacter sphaeroides* (RsRca). RsRca is a homohexameric AAA+ protein that uses energy derived from ATP hydrolysis to exert force on inhibited Rubisco in order to conformationally remodel it, leading to metabolic repair. A combination of various techniques, ranging from standard biochemical methods, hydrogen/deuterium exchange coupled to mass spectrometry (H/DX-MS), chemical cross-linking and single-particle cryo-EM led us to propose a model where Rca transiently binds to the corner of inhibited Rubisco, contacting both RbcS and RbcL. RsRca then interacts with the RbcL C-terminal tail to open a multi-layered active pocket and release the bound inhibitor. Moreover, we were able to show that RsRca acts locally on one active site and does not globally destabilize or unfold Rubisco.

2 Introduction

Along with lipids, nucleic acids and carbohydrates, proteins are the major macromolecules present in nature. Proteins provide a structural framework to give cells shape and are involved in a plethora of metabolic reactions, the products of which allow cells to grow and divide. Proteins also help to protect cells and organisms against outside invaders as part of highly sophisticated immune systems. Even the synthesis of proteins, which as explained by central dogma, proceeds from DNA to RNA to protein through the events of transcription and translation, is orchestrated by proteins in combination with nucleic acids. DNA replication, the hallmark of cell multiplication, is regulated and executed by proteins. Since function and structure are intimately connected, most functional proteins must attain a specific three-dimensional structure in a process called protein folding[1].

2.1 Protein structure

Proteins are polymers which consist of a genetically determined linear sequence of amino acid residues. An amino acid is an organic molecule with a chiral carbon (C_α) atom to which hydrogen, amino, carboxylic and side chain groups are attached. In nature, amino acids occur primarily in their L- enantiomeric form. Depending on the physicochemical properties of the side chain, amino acids can be divided into 7 groups: simple aliphatic, hydroxy, aminodicarboxylic and amides, basic, sulfur containing, cyclic and aromatic amino acids. Amino acids polymerize via a condensation reaction. The amino and carboxylic groups react, forming a peptide bond that connects C_α atoms through three covalent bonds in a trans configuration (in respect to NH and CO groups); C_α-N-C-C_α (Figure 1)[1].

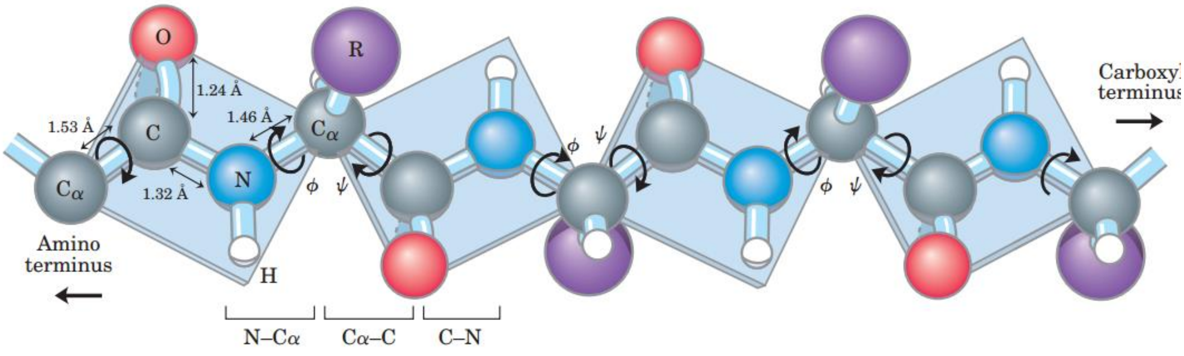


Figure 1: Schematic representation of a sequence of peptide bonds.

In the polypeptide chain depicted, bonds N-Cα and Cα-C can rotate around torsion angles phi (ϕ) and psi (ψ) to a certain degree, whereas the peptide bond C(O)-N(H) is planar (shown as plane). Distances between atoms, amino and carboxyl terminus are indicated (reproduced from Nelson *et al.*[1]).

Due to partial overlap of the carbonyl π orbital with the lone electron pair on the nitrogen atom, the peptide bond has a partial double bond character and is essentially planar. C_α atoms on both sides, however, have single bonds C_α-N and C-C_α which can rotate through torsion angles phi (ϕ) and psi (ψ), respectively. Combination of torsion angles are visualised in the so called Ramachandran plot (Figure 2)[2]. Due to steric constraints, there are two major combinations of angles that can be realized and are termed α and β. The exceptions to this rule are glycine, where the side chain is a lone hydrogen atom that allows larger span of angles, and proline, where the NH group is part of a ring which renders many torsion angles inaccessible[2].

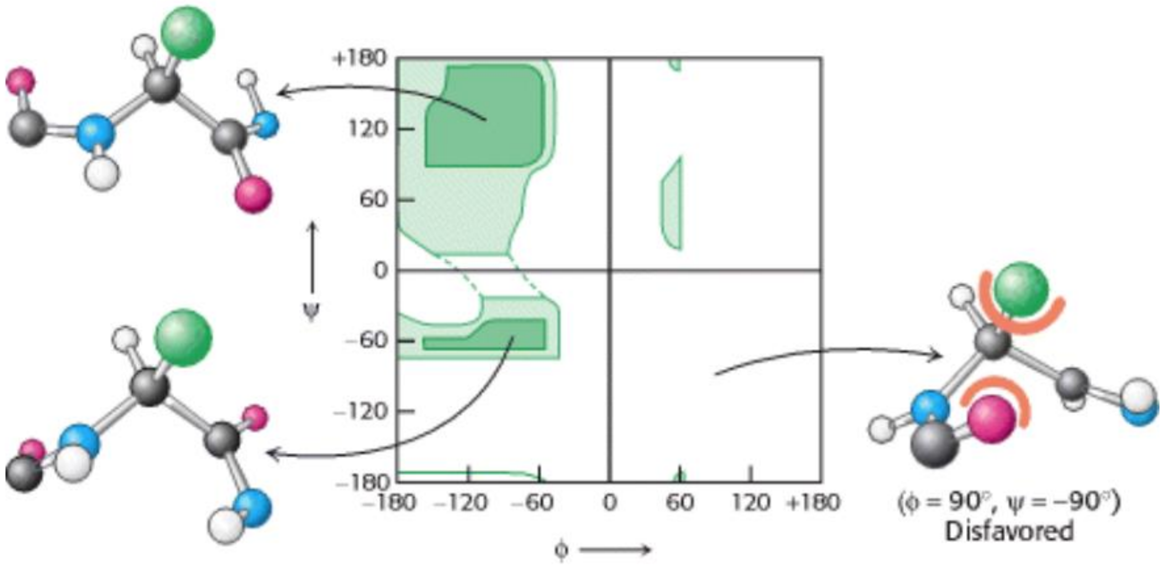


Figure 2: Ramachandran plot for L-alanine.

Combinations of phi (ϕ) and psi (ψ) angles as x and y axis, respectively, are shown on a 2D plot. Favourable combinations are shown as dark green regions. Light green regions show borderline possibilities. Structures on the left are in favourable conformation. Structure on the right is disfavoured. (reproduced from Berg *et al.*[3]).

The linear sequence of amino acid residues determines the primary structure of a protein. In relation to primary sequence, the term motif describes amino acid sequence patterns that have biological significance. Secondary structure is the local arrangement of amino acid residues into α-helices and β-sheets[4] which are formed by backbone hydrogen bonds between amine hydrogen and carbonyl oxygen. Analogous to linear motifs, structural motifs represent connectivity between secondary structure elements (super-secondary structure). Once these local elements are packed together and stabilized by π-π, π-cation, hydrogen, ionic and van der Walls interactions we use the term tertiary structure to describe them. Additionally, the expression 'fold'[5] describes how the elements of secondary structure are arranged relative to each other in space. Furthermore, several identical or different independent tertiary structures can be combined into quaternary structure. Proteins can fulfil their biological role after reaching their native structure in the intricate process of protein folding.

2.2 Protein folding

The question of how proteins fold to their final three-dimensional state, and with it the related issue of predicting the three-dimensional structure from the linear sequence, still awaits a complete answer. In a seminal experiment by Anfinsen in 1962 it was shown that denatured Ribonuclease A refolds and attains enzymatic activity upon removal of denaturant and reducing agent[6,7]. Therefore, all the information required for the protein to attain its three-dimensional form is encoded in its linear sequence of amino acids. Anfinsen's finding led him to suggest that the folding mechanism is pathway independent. Proteins would fold by random search through all the conformations in vast conformational space. However, it was not long before Levinthal opposed this view by calculating that for a protein of 100 amino acid residues, ~1000 years would be necessary to try out all possible conformations[8]. Since this is in stark contrast to the time of 5 seconds for a functional protein of 100 amino acid residues to fold at 37°C, and the fact the *E. coli* cells divide on average every 20 minutes, protein folding cannot be a random trial-and-error process. These opposing arguments are now known as Levinthal's paradox.

In order to resolve Levinthal's paradox, proteins have to achieve a stable structure (global energy minimum) on reasonable time scale (kinetic control). This could be achieved by following one specific pathway or having an option between several defined pathways. In recent years several folding models have been proposed which try to satisfy both global energy minimum and kinetic control conditions.

2.2.1 Diffusion-collision model

In the diffusion-collision model[9] the protein is considered to be composed of several parts – elementary microdomains - each short enough for all conformational alternatives to be searched rapidly, as compared with the time scale of the entire folding process (Figure 3). Overall folding would occur in a stepwise manner starting from an extended or a more collapsed state, and leading to a backbone structure close to the native state. This latter state can be described as a molten globule, which has secondary structure elements formed but lacks elements of three-dimensional structure. In the last step, the exact tertiary structure is formed by coalescence of secondary structure elements upon diffusion and collision[9].

2.2.2 Nucleation-condensation model

Nucleation-condensation model[10] postulates that neighbouring residues form native like contacts in turn forming a nucleus around which further secondary structure forms and propagates in direction out of nucleus. As a result of secondary structure formation, tertiary structure forms concomitantly (Figure 3).

2.2.3 Hydrophobic collapse model

In the hydrophobic collapse model[11] the molten globule state with a hydrophobic core forms early in the folding pathway, followed by the correct positioning of secondary structure elements in space (Figure 3).

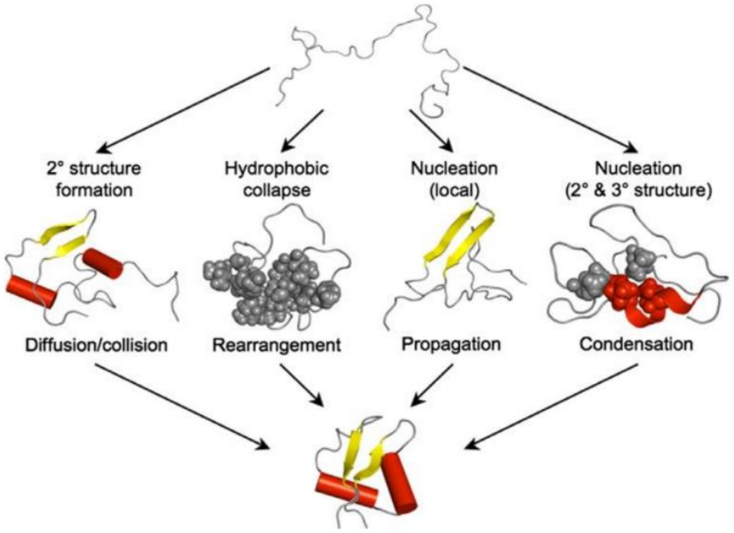


Figure 3: Classical folding mechanisms.

Diffusion-collision model states that microdomains which represent transient elements of secondary structure diffuse and collide thereby forming more stable and folded structure. In the framework model (nucleation-propagation and nucleation-condensation) initial nucleus is formed followed by outward structure formation. Hydrophobic collapse model posits the formation molten globule with hydrophobic core followed by correct positioning of secondary structure elements thus reaching a native tertiary fold (reproduced from Nickson *et al.*[12]).

2.2.4 Folding funnel model

The folding funnel model[13] is the latest in the arsenal of folding models and the most general. It describes folding as a process of native-like contacts formation causing the reduction in entropy and free energy of the polypeptide chain. Rearrangement of amino acid side chain contacts during the folding process drives the polypeptide chain along the folding landscape down the funnel towards the unique three dimensional structure with lowest free energy (Figure 4).

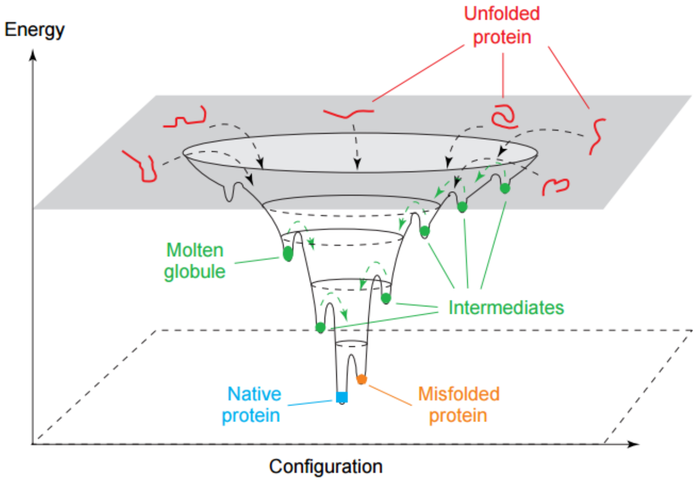


Figure 4: Folding funnel model.

Folding funnel model describes protein folding in an ensemble approach. At the beginning of folding a vast conformational space is available to protein molecules which then funnel into one unique state with a decrease in entropy and enthalpy (reproduced from Radford *et al.*[14]).

All of the folding models described above, except for nucleation-condensation, predict folding intermediates. Intermediates differ in flexibility, stability, and the number of native contacts[15]. Some intermediates can be on-pathway, and act to reduce the number of attainable states and can fold to the native state[16]. On the other hand, some intermediates can be kinetically trapped and unable to complete productive folding on their own. According to the folding funnel model, they represent local energetic minima on a rugged free energy folding landscape. In absence of external factors[17], and due to a substantial amount of exposed hydrophobic residues, such intermediates can accumulate and aggregate[18].

2.3 Methods for studying protein folding *in vitro*

To completely describe a biological process, one needs both structural information and rates of the interconversion of all the species. The complicating factor in studying protein folding is the short lifetime of folding intermediates. Nevertheless, in recent years many methods that allow the observation of protein folding have been developed. In order to approach a complete description of the folding mechanism, several of these methods have to be used in combination. Methods most commonly used in the field are summarized in Table 1.

Technique	**Timescale**	**Description**
Fluorescence		
Intrinsic fluorescence	ms	Local environment of Trp and Tyr residues

ANS binding	ms	Exposure of hydrophobic areas
FRET	ms	Distance between labelled positions
Substrate/Inhibitor binding		Formation of native contacts
Anisotropy		Dynamics of protein chains
Circular dichroism		
Far UV	ms	Secondary structure formation
Near UV	ms	Tertiary structure formation (local environment of aromatic residues)
Other spectroscopic methods		
Small-angle X-ray scattering (SAXS)	ms	Overall shape and dimensions of a protein
Absorbance (near UV)		Local environment of aromatic residues
Hydrogen exchange		
Native exchange		Global stability of a protein
Pulsed hydrogen exchange NMR		Rate of hydrogen exchange on backbone and amino acid side chain
Pulsed hydrogen exchange ESI MS		Rate of hydrogen exchange in folding population
Additional methods		
Atomic force spectroscopy (AFM)		Rates of folding and unfolding events
Solution-state NMR		Changes in environment of amino acid side chains

Table 1: List of commonly used methods in protein folding studies (Modified from Brockwell *et al.*[16] and Radford *et al.*[14]).

Abbreviations: ANS: 1-anilino naphthalene sulphonic acid; FRET: Förster energy resonance transfer; UV: ultraviolet; ESI MS: electron spray ionization mass spectrometry; NMR: nuclear magnetic resonance.

Of particular importance are the methods that allow us to study changes on short timescales, such as stopped-flow, quenched-flow and microfluidics combined with a biophysical method of choice. For example, fluorescence intensity, lifetime and anisotropy (on ensemble or on single molecule level) measurements provide information on the amount of structure, chain dynamics, inter-residue distances, changes in environment around the probe and kinetics of both local and global folding. Circular dichroism can be used to monitor the extent of secondary (far UV range) and tertiary structure (near UV range) either as steady state or kinetic measurements. In recent years, the method of hydrogen/deuterium exchange mass spectrometry (HDX MS) has provided us with an unprecedented amount of temporal and spatial information of folding processes. Nuclear magnetic resonance (NMR) and UV absorption (albeit with low resolution) are used to describe the environment around the amino acid side chain or aromatic residues, respectively. Small-angle X-ray scattering (SAXS) can be used to gain low resolution structural information on polypeptide shape and dimensions. Finally, the method of atomic force microscopy (AFM), especially when used on a single molecule level, reveals valuable information on folding and unfolding rates.

Protein engineering as realized in the method of ϕ value analysis[19] is a method of choice if one would like to interrogate the role of specific residues in the folding process. In this method, several different amino acid residues are most commonly replaced by alanine. From the change in folding and unfolding kinetics caused by such a perturbation, one can infer the degree of native-like structure around this position in the transition state or folding intermediate. Usually, ϕ values range from zero to one. When the value is close to one, it implies that this position plays significant role in stability and kinetics and that it is close to the native form. The opposite is true for values close to zero.

2.3.1 Single molecule methods in protein folding

Most of the methods listed above are used to observe folding events at an ensemble level. In an effort to gain higher resolution and describe the folding of single protein molecules, two distinct single-molecule approaches were developed: fluorescence spectroscopy and force spectroscopy. Single proteins are isolated by immobilization on a surface, trapping between optical tweezers or working at high dilution (< 100 pM).

Single molecule fluorescence spectroscopy experiments are carried out on a confocal laser microscope. Such a setup allows one to observe single protein molecules that pass through a small confocal volume (~1 fL). Proteins have to be labelled with fluorescent dyes in order to be detected. Labelling is achieved via coupling reactions of cysteine or lysine with maleimide or N-hydroxysuccinimide-ester (NHS-ester), respectively. Other methods of labelling include coupling of a fluorescence probe to N- or C-termini. Another possibility is the introduction of non-natural fluorescent amino acids. Depending on the information one would like to obtain, different fluorescence spectroscopy variants can be employed[20,21,22].

Single molecule Förster resonance energy transfer (smFRET) is radiation-less energy transfer between a suitable pair of donor and acceptor dyes (overlap integral) (Figure 5)[20]. Energy transfer or FRET efficiency is defined as the ratio of either fluorescence intensity or lifetime of the excited donor fluorophore in the presence and absence of the acceptor, and it occurs on distances of 2-10 nm. Energy transfer is highly sensitive to changes in distances between labelled positions ($E \sim r^{-6}$) and is thus suitable for detection of local properties that define different conformational states or folding intermediates. FRET measurements on a single molecule level have an advantage of obtaining quantitative information on different species present in the process which would otherwise have been averaged out in ensemble measurements.

Fluorescence correlation spectroscopy (FCS) is a method that depends on measurements of fluorescence intensity fluctuations (Figure 5)[20]. Fluctuations arise due to changes in the number of molecules that pass on a timescale of microseconds to milliseconds through the observation volume over a certain measuring time period. From the analysis of measured fluctuations the autocorrelation function, which is defined as the self-similarity of a signal as a function of time, can be constructed. The appropriate fit for the autocorrelation curve is then used and several parameters such as rotational diffusion time, blinking events, translational diffusion time and number of molecules, can be extracted[23]. FCS is commonly used to monitor events on nanosecond to hour time scale and can detect changes in size and conformation and thus has been widely used in protein folding studies[24]. Since FCS is a fluctuation method it is limited to concentrations of 1 pM to 200 nM. Below 1pM there are too few events that occur which leads to poor statistics. Going above concentrations of 200 nM however, results in reduced sensitivity to fluctuations in the confocal spot.

Photoinduced electron transfer (PET) is a photo-physical event where an electron from the indole ring of tryptophan is transferred to the fluorophore, thereby quenching the fluorophore (Figure 5)[21]. For this event to occur, tryptophan and the fluorophore have to be in van der Walls contact (<2 nm). In combination with FCS (PET-FCS) the transition between dark and bright states (blinking) can be observed as an additional exponential decay in the autocorrelation curve at short correlation times (µs). PET-FCS was used to observe fast protein chain dynamics[22], early folding events[25] and transitions between conformational states[26].

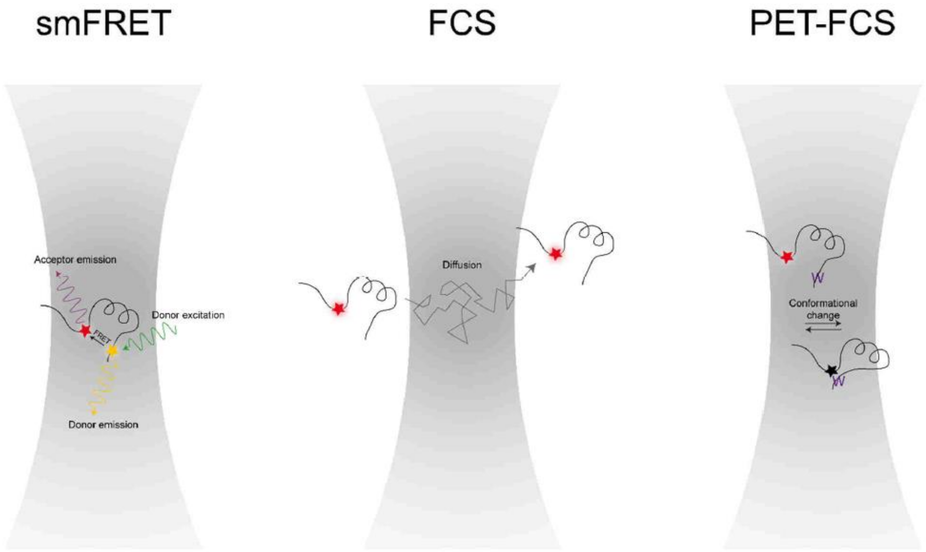


Figure 5: Comparison of single-molecule fluorescence methods.

Spectroscopic methods implemented in confocal microscopy for observation of single molecules. In smFRET, distance between labelled positions is determined from ratio of either donor/acceptor fluorescence intensity or lifetime. FCS measures fluorescence intensity fluctuations and associated processes. Most commonly, diffusion coefficient is obtained from diffusion time and the size of a confocal spot. PET-FCS is based on quenching of oxazine dye by tryptophan residue(s) when in distance of van-der-Waals contacts. Information on conformational changes or chain dynamics is obtained from such experiment (reproduced from Gupta[27]).

2.3.2 Single particle transmission electron microscopy

Single particle transmission electron microscopy (EM) is a structural method that has recently gained substantial attention[28]. Instrumental[29], and data processing progress[30] resulted in a 'resolution revolution' that now enables determination of atomic[31] or near atomic resolution of protein structures. In EM, electrons are released from an electron gun and focused on to the specimen by a set of condenser lenses. Emitted electrons are scattered by electrons and atom nuclei in macromolecules of the sample. Scattered electrons are refocused by the objective lens system resulting in the formation of an image of an object. The image is then further magnified by an intermediate lens system and finally projected by the projector lens system. A micrograph is recorded either on a photographic film, a scintillator-based digital camera or the recently developed direct electron detection camera[32] (Figure 6). Specimens for EM are prepared by applying a solution of macromolecules to an EM grid. The solution is then removed and the specimen can be either negatively stained (negative stain-EM) or plunge-frozen into liquid ethane (cryo-EM) (Figure 7).

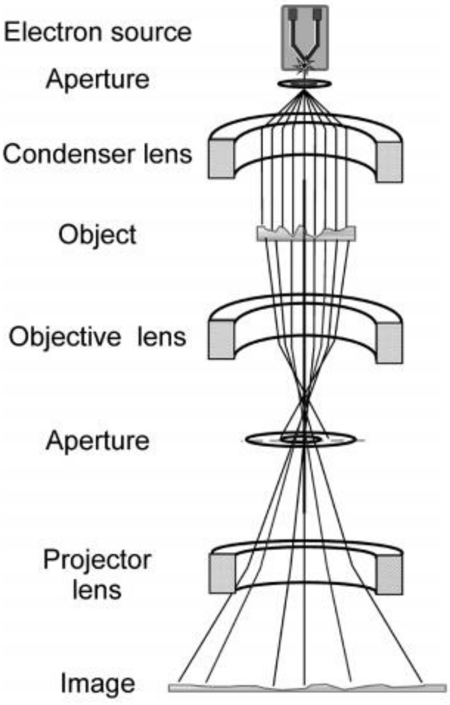


Figure 6: Simplified representation of an electron microscope.

Electrons are released from the electron gun and are focused by the condenser lens onto the specimen. The objective lens is responsible for image formation and magnification of the object. Aperture stops the electrons that are off-axis or off-energy. Further magnification is achieved with system of projector lens. Finally, the image is projected onto a plane (reproduced from Orlova *et al.*[32]).

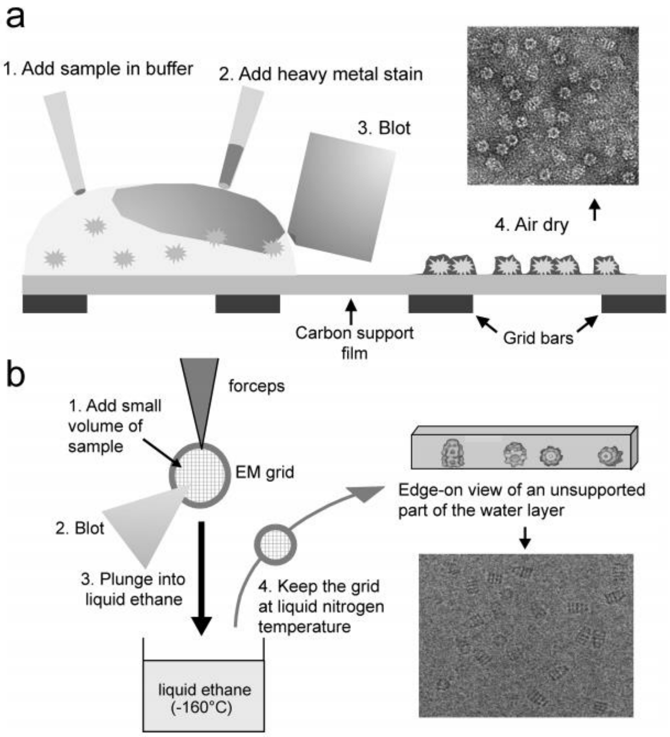


Figure 7: Sample preparation for electron microscopy.

Two approaches of sample preparation: a.) Negative staining where the sample solution is applied to the grid, stained with heavy metal stain and dried. b.) Freezing with liquid nitrogen where the sample is applied to the grid and plunge frozen into liquid ethane. Example of cryo-EM image is shown. (reproduced form Orlova *et al.*[32]).

The structure of a protein is determined computationally by combining images of many macromolecules[33]. On a micrograph, particles that are 2D projections of a molecule can adopt a wide range of orientations. Six geometric parameters define the position and spatial orientation of the particle: two in-plane coordinates (x,y), three Eulerian angles (φ, θ, ψ), and a defocus (z), which is assumed to be the same for all the particles on the micrograph. After image alignment and 2D classification which resolves compositional and configurational heterogeneity, an initial 3D model can be built. Subsequent 3D classification allows for an additional in-silico purification step. Finally, in the 3D refinement, Eulerian angles are more accurately determined and a 3D model can be built[34] (Figure 8).

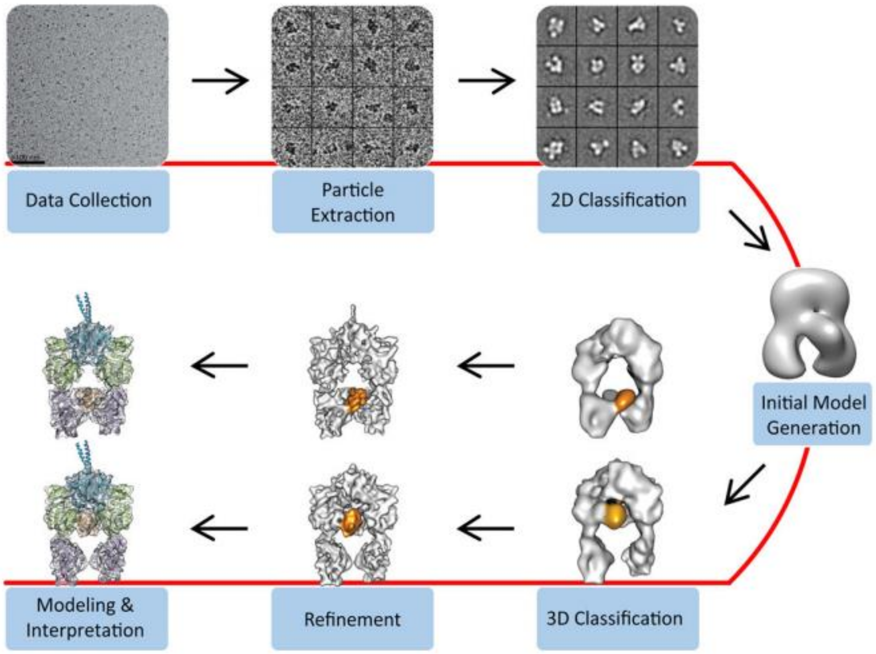


Figure 8: EM workflow.

Overview of cryo-EM single particle reconstruction. After data collection, particles are extracted and 2D classification is performed. After initial model generation, additional 3D classification can be performed to attain more conformationally homogeneous classes. Refinement results in the 3D model which can be used for de-novo model building or fitting of know structures(reproduced from Skiniotis *et al.*[34]).

Cryo-EM was previously used principally for analysing large macromolecular complexes. However, with the recent development it is possible to obtain atomic structures of smaller proteins, such as y-secretase (170 kDa)[35]. Furthermore, it is now possible to dissect various conformational states of macromolecular machines along their mechanistic cycles[36]. Cryo-EM has also become the method of choice for studying complex assemblies such as the spliceosome, which are highly dynamic in nature and resist crystalization[37]. Another advantage of EM is the small amount of material that is required for the experiment (micrograms of protein). Since it is possible to capture short lived, dynamic complexes, EM can also be used to study chaperones that are involved in the assembly of protein complexes. Intermediates that form along the assembly process can be structurally characterized. Due to the dynamic nature of intermediates, one has to take into account the compromised resolution in the final reconstruction. Nonetheless, the resulting envelope of electron density still allows for docking of X-ray determined structures (hybrid approach) and determination of the general architecture of a complex.

2.4 Protein folding in the cell

In contrast to dilute in vitro solutions, the environment in the cell is highly concentrated, with an estimated macromolecular concentration of ~300-400 mg/ml[38] (Figure 9). The term of macromolecular crowding was coined to describe this property. Such high concentrations cause macromolecular crowding[39], which results in a substantial excluded volume effect and has a profound effect on protein folding, stability, and catalysis as well as binding affinities inside a living cell[40,41].

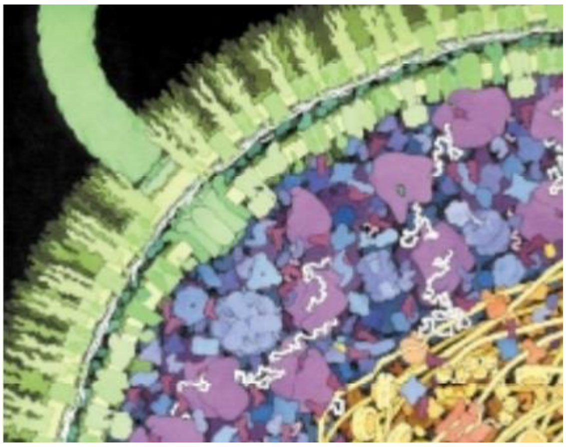


Figure 9: Macromolecular crowding in *E. coli* cell.

Environment in the cell is highly concentrated, with concentrations of macromolecules reaching up to 300-400mg/ml, resulting in substantial excluded volume effects. Approximated amount, shape and density of macromolecules inside *E. coli* cell are shown. (reproduced from Ellis *et al.*[42]).

In such an environment, non-specific interactions are more likely to occur. In particular, aggregation is favoured when partially folded proteins expose hydrophobic regions, such as the case of ribosome bound nascent chains or folding intermediates.[43]. When aggregation is faster than folding, oligomers and sometimes aggregates can form, which can have detrimental consequences for the cell[44]. Indeed, the presence of aggregates is a hallmark of many neurodegenerative diseases, such as Alzheimer's, Parkinson's, Huntingtin's and Creutzfeldt-Jakob's disease, as well as amyloid lateral sclerosis (ALS). However, it is still a matter of active debate whether oligomeric species or aggregates are responsible for toxicity[45]. For all the above listed diseases, so called amyloidoses, a common form of amyloid-like aggregates are formed that are characterized by cross-beta-sheet structure. It is therefore clear that for in-cell folding to be productive, additional factors have to be at play[43,44,46].

2.4.1 Molecular chaperones

A molecular chaperone is a protein that helps in folding or unfolding of another protein without being a part of its final structure[47]. The crucial property which gave name to chaperones is that in the act of chaperoning any unproductive/unwanted interactions are prevented. Chaperones are important in all branches of protein homeostasis (proteostasis), from biosynthesis (folding and assembly), to maintenance and degradation[44].

Many of the molecular chaperones are upregulated upon heat shock, and have since been recognized as heat shock proteins (Hsps)[48]. They are classified according to their molecular mass or function. They can act as holdases, which stabilize non-native conformations; foldases, which assist folding; or unfoldases, which unfold misfolded species or extract proteins from aggregates[49]. In general, chaperones recognize and bind exposed hydrophobic patches on a substrate protein. Binding affinities of the substrate are regulated in two markedly different manners: ATP dependent, or ATP independent. Most chaperones hydrolyse ATP (Hsp60, Hsp70, Hsp90, Hsp104 and Hsp110). ATP hydrolysis drives the chaperone through different conformational states, which remodels the substrate and modulates substrate binding affinity[50]. Examples for regulation of substrate binding affinity in an ATP independent manner include post-translational modifications such as phosphorylation (small Hsps)[51] and a reduction in the amount of stress-unfolded proteins, which can shift the equilibrium between the bound and unbound states of the substrate (Spy chaperone)[52].

Molecular chaperones act in combination with co-chaperones, which enable chaperones to interact with a broad array of substrates (Hsp40), modulate chaperone ATPase activity (Hsp40, Hsp10) and expand the number of processes in which chaperones can be involved[49].

2.4.2 Molecular chaperone network

Different molecular chaperones within the cell are organized into a complex network (Figure 10)[53]. In all three domains of life, bacteria, archaea and eukarya, proteins are synthesized on ribosomes. An average protein, while being synthesized on the ribosome, stays in a non-native state for around 15 seconds[54]. As a first adaptation to prevent early unwanted interactions, ribosomes adopt a pseudo-helical organization in polysomes[55], allowing ribosome exit tunnels to point in opposite directions. During and immediately after translation the protein enters the flux of a highly sophisticated chaperone network.

It first encounters ribosome associated chaperones[56]. Then, it is transferred downstream to Hsp70/Hsp40 system[57]. While some proteins can fold after passing through this node of network, others, require additional chaperones such as Hsp90 and barrel shaped chaperones with a central cavity (chaperonins)[58].

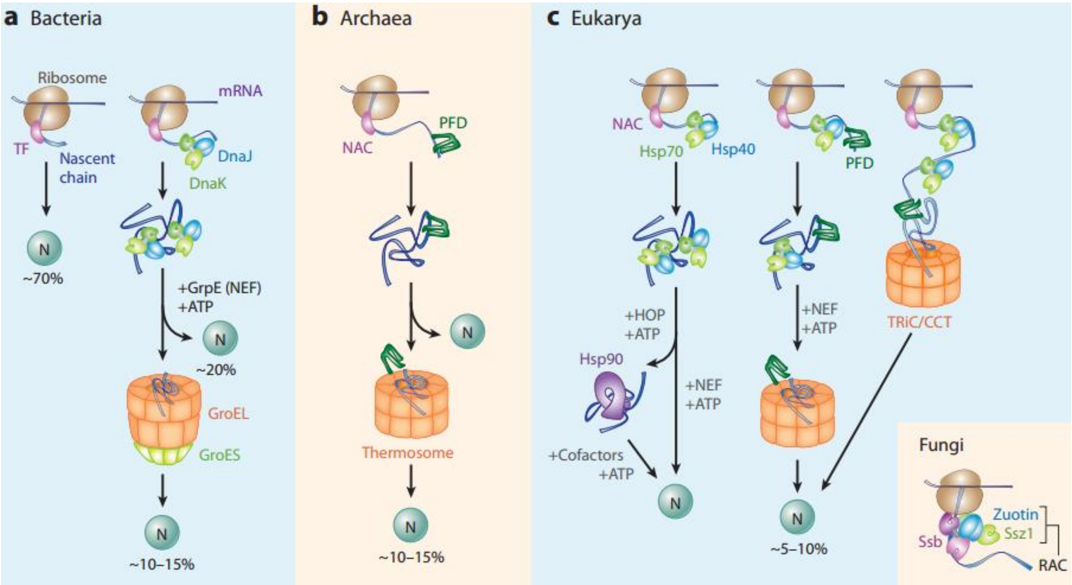


Figure 10: Chaperone network in all domains of life.

Chaperones are organized in sophisticated network in a.) bacteria b.) archaea and c.) eukarya. During translation protein chain encounters ribosome associated chaperones. Upon completion of biosynthesis some protein can spontaneously fold whereas others need assistance of downstream chaperones such as Hsp70, Hsp90 and Hsp 60. Percentages indicate the amount of proteins either independent or dependent on different chaperone systems (reproduced from Kim *et al.*[46]).

Some proteins, despite the help of a folding network, might not be able to fold due to mutations or changes in cellular environment[46] (Figure 11). To prevent unproductive cycling through folding chaperones, they have to be transferred to and degraded by either the ubiquitin/proteasome system (UPS) or autophagy/lysosome pathways. On the other hand, larger aggregates that have already formed, can be cleared by autophagy or can be remodelled and disaggregated by chaperones in cooperation with the UPS or lysosome pathways. Overload of UPS or lysosomes pathways can cause 'choking' of these systems and leads to build up of misfolded proteins and further aggregation[44].

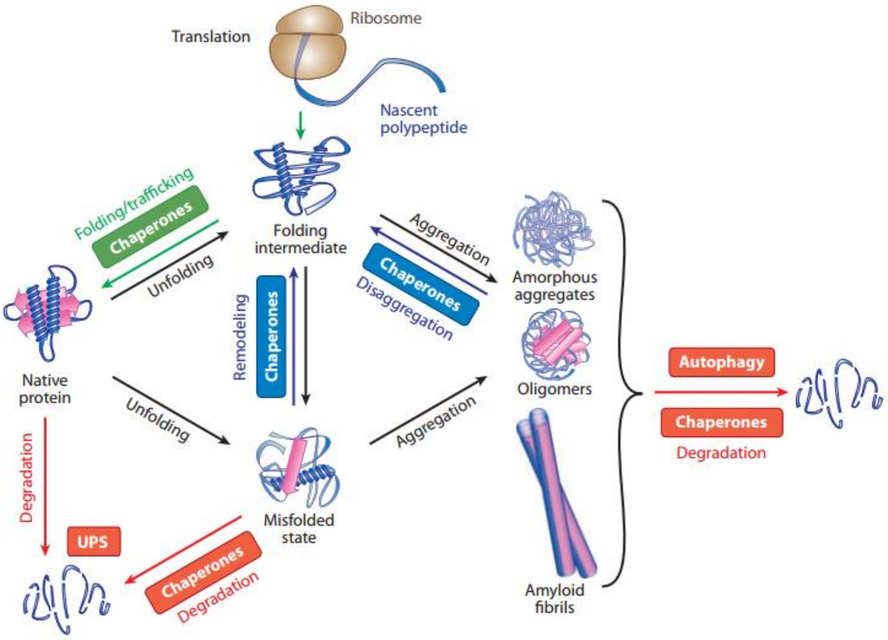


Figure 11: Proteostasis network.

Proteostasis network components are involved in synthesis, maintenance and degradation of a proteins. After the protein is synthesised on a ribosome it forms a folding intermediate which can have different fates. It can fold to native state, misfold or aggregate. These states can interconvert by assistance of proteostasis network components (reproduced from Kim *et al.*[46]).

The life and fate of a protein inside the cell does not depend only on foldases, but also chaperones which maintain its structure and at the later stage successfully transfer it for degradation(Figure 12)[43,44,46].

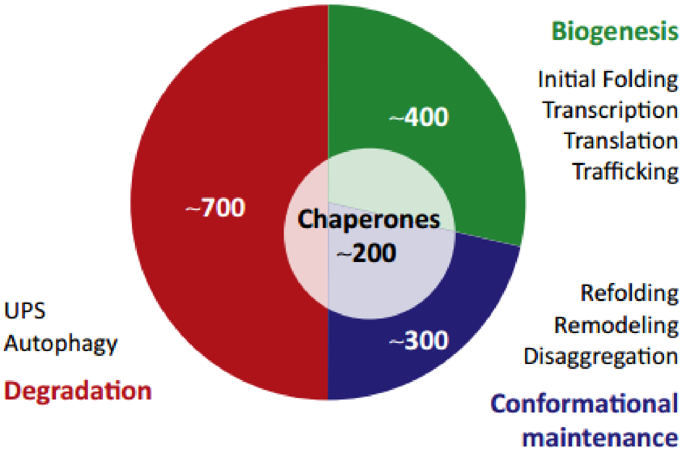


Figure 12: Components of the human proteostasis network.

Chaperones represent ~200 components of ~1400 factors comprising the proteostasis network. Proteostasis network includes proteins involved in biogenesis, conformational maintenance and degradation (reproduced from Hipp *et al.*[44]).

2.4.3 Ribosome-associated chaperones

Protein synthesis on the ribosome proceeds in a vectorial fashion from the N- to C-terminus. Only after the complete chain emerges from the ribosome exit tunnel can long-range interactions form. It is characteristic of nascent chains to expose hydrophobic patches prone to aggregation.

Bacteria have evolved the ribosome-associated chaperone Trigger factor (TF) to prevent early misfolding and aggregation events (Figure 13)[58,59]. TF is a conformationally dynamic ~50 kDa protein, consisting of three domains[60]. The dispensable peptidyl-prolyl-isomerase (PPI) domain might serve as additional substrate binding site and adds to the chaperone activity. The C-terminal domain, which bears the main chaperone activity, and the N-terminal domain are crucial for ribosome binding and chaperone function. TF binds to the ribosomal protein L23, thus positioning TF close to the ribosomal exit tunnel and allowing for interaction with the nascent chain. TF does not have a single substrate binding site but interacts with the nascent chain with its entire inner, cavity-forming surface. According to the structural model, the cavity formed between TF and ribosome could accommodate a protein domain of 14 kDa in size possibly allowing co-translational folding[61]. It is interesting to note that the action of TF delays the folding process[62]. The concentration of TF in the cell is ~50 µM and is in excess over ribosomes at ~20 µM. TF binds transiently to vacant ribosomes with an apparent K_D of 1-2 µM and k_{off} of 10-15 seconds. The K_D is decreased to 40-700 nM (2-30 fold) in the presence of the nascent chain. This allows TF to differentiate between vacant and translating ribosomes[59]. TF dissociates in an ATP independent manner. After dissociation it hands the nascent chain to downstream chaperones[57]. Due to the excess of TF over ribosomes, it was speculated that TF might be involved in other functions. Indeed, TF may function as ribosome assembly factor[59].

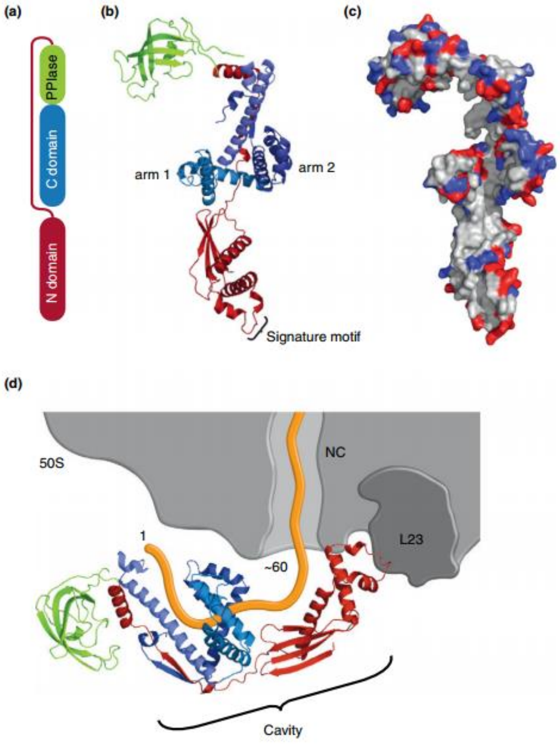


Figure 13: Trigger factor from *E. coli* – ribosome associated chaperone.

a.) Domain organization of TF. TF has N-, C- and PPIase domains. b.) Crystal structure of TF (PDB: 1W26) shown in ribbon representation. N-terminal domain (red) has a signature motif (GFRxGxxP) important for ribosome binding. Peptidyl-prolyl cis-trans isomerase (PPIase) domain (green) is connected to N-terminal domain via flexible linker. C-terminal domain (blue) is in the middle and forms two arm-like protrusions. N-domain and arm-like protrusions form a cavity for nascent polypeptide chain. c.) Distribution of positive (blue) and negative (red) charges shown in space filling model. d.) Structural model of TF bound to ribosome exit tunnel. Contact sites include the signature motif of N-terminal domain of TF (red) and ribosomal L23 protein (dark grey) (reproduced from Preissler *et al.*[59]).

TF is not present in archaea and eukarya, but is substituted by the ribosome-associated complex (RAC) which is composed of a specialized Hsp70/40 based system (Zuo/Ssz/Ssb) or nascent-chain associated complex (NAC), respectively[58,63,64].

2.4.4 The Hsp70 system

The Hsp70 chaperone family is the hub in the centre of the cytosolic chaperone network[65]. It has been recently shown in *E. coli* that Hsp70 interacts with ~700 proteins. A subset of ~180 aggregation prone proteins are especially dependent on this system[57].

Hsp70 is a ~70 kDa protein. It has an N-terminal nucleotide-binding domain (NBD) and a C-terminal substrate-binding domain (SBD), connected by a flexible linker important for allosteric regulation (Figure 14)[49]. The NBD has two large globular subdomains divided into two small subdomains. Between the subdomains there is a cleft where nucleotide binds. For efficient nucleotide binding, one Mg^{2+} and two K^+ ions are required. The SBD is subdivided into an N-terminal β-sandwich subdomain and a C-terminal α-helical subdomain. The latter functions as a lid to close over bound substrate after ATP hydrolysis.

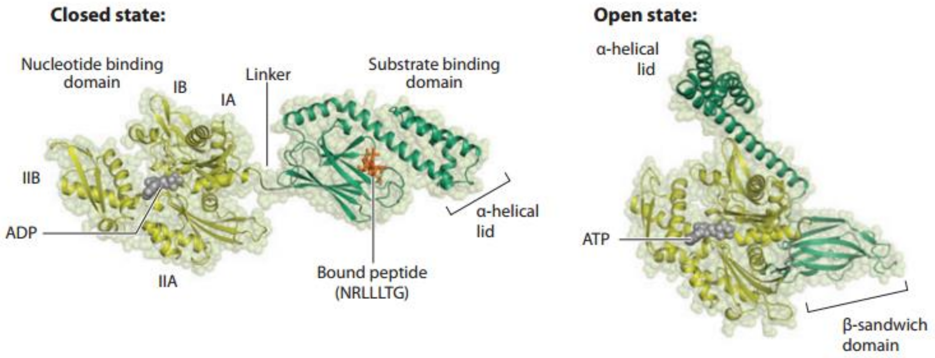


Figure 14: Structure of Hsp70 in open and closed states.

Hsp70 consist of nucleotide-binding domain (NBD) and a substrate-binding domain (SBD) connected via conserved, flexible linker. The structure of a closed state (left) (PDB: 2KHO) shown in ribbon and space filling representation was solved by hybrid approach using solution NMR and crystal structures of individual domains. In this structure ADP (violet) is bound to NBD and the α-helical lid is closed over the peptide (red) bound to SBD. A-helical domain of SBD, linker, and subdomains (IA; IB; IIA; IIB) of NBD are indicated. The open state (right) is represented by crystal structure of ATP-bound Sse1 (PDB: 2QXL). In this state β-sandwich domain contacts IA subdomain od NBD. α-helical lid contact IA and IB of NBD (reproduced from Kim *et al.*[46]).

Non-native proteins bind to Hsp70 via a hydrophobic stretch of 5-7 residues flanked by positively charged amino acid residues (Figure 15)[49]. Hsp70 can then either keep the protein in an unfolded state acting as a holdase or functions as a foldase, thus helping substrate proteins to fold[66]. In the latter case, after binding of the substrate and subsequent conformational changes of Hsp70, the substrate is released into free solution. As the protein collapses upon release, it can kinetically partition to the correct pathway and fold[67]. In an alternative explanation of its foldase activity, repetitive binding and release induce local unfolding and help overcome kinetic barriers for folding to the native state. When several cycles of repeated binding and release cannot drive the substrate to its native state, Hsp70 fulfils its role as a holdase and transfers the substrate in a non-native state to downstream chaperones[46]. Moreover, as part of a proteostasis network in yeast, Hsp70 in combination with Hsp104 can function as a disaggregase. Higher eukaryotes however, lack Hsp104; instead a combination of Hsp70 and a mixed J-protein complex is involved in disaggregation processes[68].

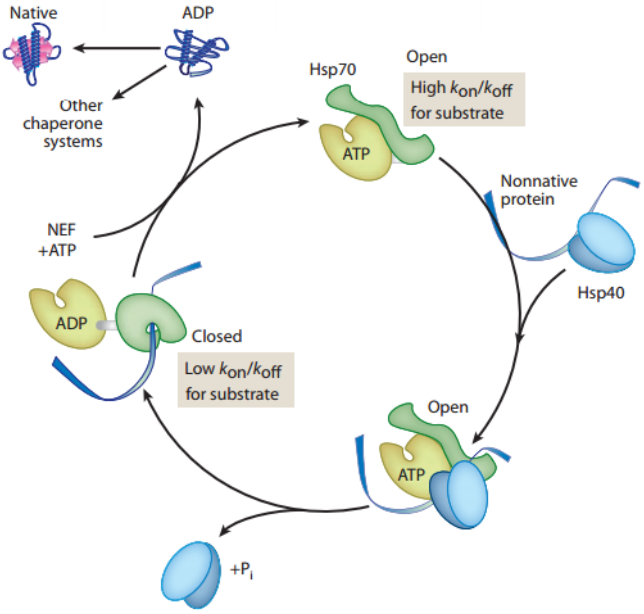


Figure 15: Hsp70/40 reaction cycle.

Hsp70 and Hsp40 are shown in cartoon representation. Hsp70 nucleotide binding domain is in yellow, Hsp70 substrate binding domain is in green. Hsp40 is in blue. Substrate protein is in ribbon representation coloured dark blue. Hsp70 with bound ATP is in open state allowing for substrate binding which is recruited by Hsp40 additionally Hsp40 stimulates Hsp70. Hsp70 with ADP bound is in closed state with α-helical lid over the bound substrate which effectively decreases on and off rates. In the last step NEFs stimulate ADP release and subsequent ATP binding causes substrate release (reproduced from Kim *et al.*[46]).

Additional roles attributed to Hsp70 are mediating the transfer of substrates across membranes and in regulating the σ^{32} heat shock transcription factor. Interestingly, Hsp70 also plays a role in the disassembly of clathrin coats, viral capsids and the nucleoprotein complex. Hsp70 can take part in such diverse tasks due to a plethora of associated co-factors[69].

In the context of proteostasis, Hsp70 has two main co-factors. The co-chaperone Hsp40 and a nucleotide exchange factor (NEF) which drive the Hsp70 conformational cycle[69]. During cycling, Hsp70 switches between ATP and ADP states which differ in substrate binding affinities. Hsp70 with bound ATP is in the 'open' state with low affinity for substrate and high on- and off-rates. By delivering the substrate protein to Hsp70, Hsp40 stimulates the ATPase activity of Hsp70 by more than ~1000 fold[70]. Substrate alone can stimulate the ATPase activity up to 2-10 fold as well, but this is not sufficient for effective cycling. ATP hydrolysis causes the closing of α-helical lid, thereby 'locking'-in the substrate. The 'closed' state is characterized by high binding affinity and low on/off rates. The NEF accelerates the release of bound ADP, causing opening of the lid and release of the substrate.

2.4.5 The chaperonins

Chaperones with cavities formed within each of two rings stacked back to back are called chaperonins. These are ~1 MDa large complexes which can bind and encapsulate non-native proteins[43]. Chaperonins act upon ~10 % of cytosolic proteins and are indispensable for cell viability[71]. The cavity of each ring is an active site where substrate protein is bound, encapsulated, folded in isolation and released. These steps are part of a complex cycling mechanism driven by ATP hydrolysis.

Chaperonins are classified into groups I and II. Group I is present in bacteria and organelles of endosymbiotic origin (mitochondria, chloroplasts), whereas group II is specific to archaea and eukarya[43]. Both types of chaperonins have cavity-forming rings stacked back to back. Chaperonins of group I have seven identical subunits which form the ring. Group II chaperonins, on the other hand, have eight to nine subunits (either identical or up to three different types) as it is the case for archaeal chaperonins, or eight different subunits within the ring, which is characteristic of eukaryotic TRiC/CCT. Chaperonins of the first group work together with lid-shaped co-chaperone GroES. Group II chaperonins do not need any external co-chaperone since they have a built-in lid in a form of an α-helical protrusion extending from the apical domain.

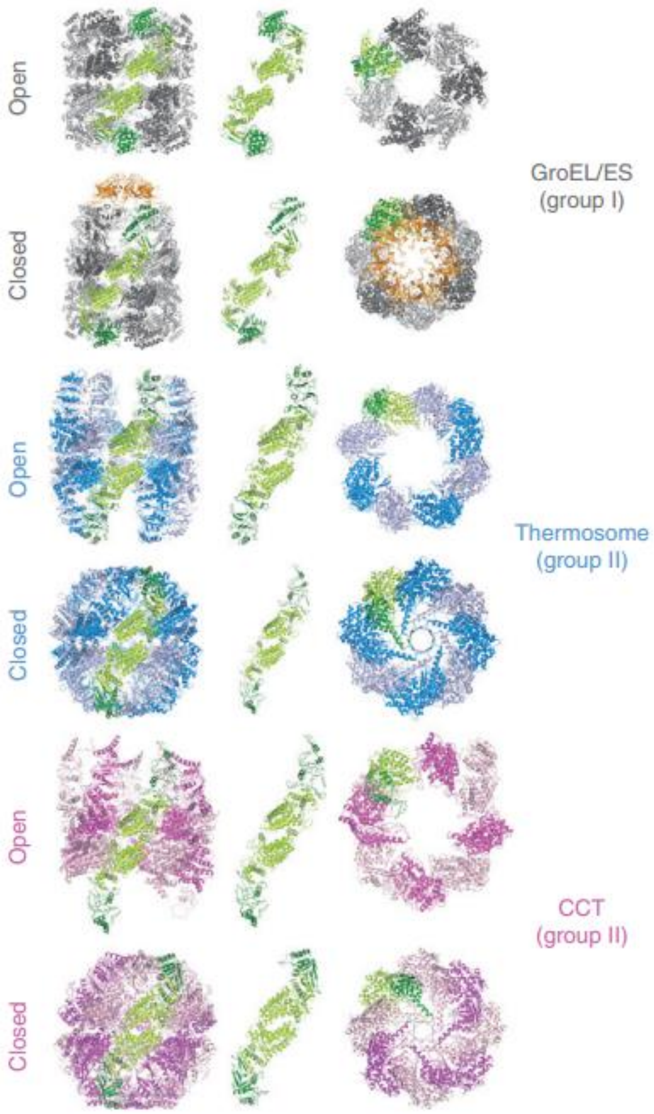


Figure 16: Structure of group I and II chaperonins.

Chaperonins consist of two rings stacked back to back. First column shows crystal structures of chaperonins in open and closed states in ribbon representation: GroEL(PDB: 3E76), GroEL/ES (PDB: 1AON); open (PDB: 3KFK) and closed (PDB: 1A6D) states of thermosome, open (PDB: 2XSM) and closed (PDB: 3IYG) states of CCT/TRiC in side views. Equatorial and intermediate domains forming inter- and intra- ring contacts, respectively, are coloured light green. Apical domain is coloured dark green. Rest of the chaperonin structures is in grey (GroEL), blue (thermosome),violet (CCT/TRiC). GroES is in orange. Middle column shows and compares isolated pair of inter-ring interacting chaperonin subunits in open and closed states. Last columns shows chaperonins in top view for open and closed state (reproduced from Yebenes *et al.*[72]).

Substrates bind the chaperonins via the apical domains of the open ring. Upon encapsulation, the substrate is given a chance to fold within the cavity. In the case of Group I chaperonins, folding starts after ATP binding and displacement of the substrate protein into chaperonin cavity by the co-chaperone. For Group II chaperonins, ATP hydrolysis triggers the closure of a built-in lid and the release of the substrate into the cavity[73].

2.4.6 The chaperonin system GroEL/ES of *E. coli*

GroEL/ES is the most studied chaperonin system thus far. EM[66], and X-ray studies[74,75] showed that GroEL consists of two heptameric rings stacked back to back (Figure 17). Each ring is composed of seven identical subunits and forms a cavity where assisted folding occurs. Each subunit, ~56 kDa in size, has three domains. The apical domain (residues 191-376) is required for the binding of a heptameric co-chaperone GroES and non-native substrate. More specifically, substrates are bound by amino acid residues on helices H (233-243) and I (255-267)[66]. Moreover, it has been shown that at least three apical domains are required for successful substrate binding[76]. The intermediate domain (residues 134-190) transmits allosteric signals and contains two pivot points (Gly192 and Gly375) around which large en-bloc movements occur[66]. The equatorial domain (residues 6-133 and 409-523) has a nucleotide binding pocket where ATP binding and hydrolysis occurs[66]. The majority of intra- and inter-ring contacts are formed between equatorial domains.

The height of the double ring complex in the apo state is ~146 Å and the width ~136 Å (Figure 17)[50]. The volume of the cavity is estimated to be 75.000 Å^3. Upon ATP and GroES binding, the dimensions of the complex increases to ~183 Å in height and ~150 Å in width (Figure 17). The volume doubles correspondingly to approximately ~150.000 Å^3 and can accommodate a substrate up to ~60 kDa in size.

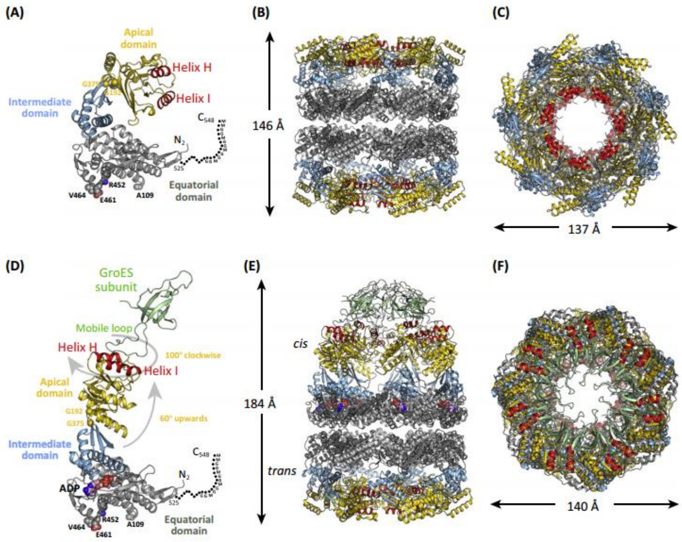


Figure 17: Structure of GroEL and GroEL/ES complex.

a.) Isolated GroEL subunit from apo GroEL (PDB: 1SS8). GroEL subunit is composed of apical (yellow), intermediate (blue) and equatorial (grey). Helices H and I important for substrate binding are in red. Hinge residues (Gly192 and Gly375) and residues forming inter-ring contacts (Ala109, Arg452, Glu461 and Val464) are shown in space-filling representation. N- and C- termini are indicated. 23 residues long C-terminal tail is modelled and shown as dots. Gly-Gly-Met repeats are indicated. b.) Side view of apo GroEL tetradecamer (PDB: 1SS8) c.) Top view of apo GroEL tetradecamer. d.-f.) GroEL/ES complex (PDB: 1PF9) in equivalent views and colouring as in a.-c.). GroES is shown in ribbon representation and coloured green. ADP is shown in space-filling representation. Grey arrows indicate conformational changes upon nucleotide binding. (reproduced from Hayer-Hartl *et al.*[50]).

The GroEL rings exert allosteric regulation[77]. Within the rings there is positive allostery which results in concerted binding of ATP by all seven subunits. Between the rings, negative allostery decreases the binding affinity for ATP of the second ring, therefore allowing only one ring to function in the low to mid-range ATP concentrations[78]. The basis for allosteric regulation are formation and breakage of salt bridges between the subunits within the rings and the subunits between the rings[79].

GroEL and GroES work together in a complex multi-step cycle driven by ATP binding and hydrolysis (Figure 18).

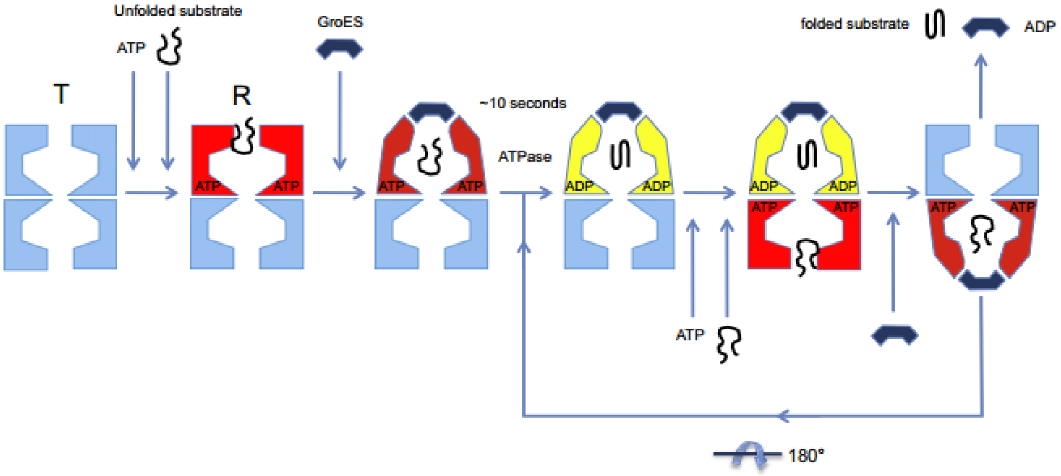


Figure 18: GroEL/ES reaction cycle.

In this figure, the starting state of the cycle is the apo GroEL tetradecamer where both rings are tense (T) state (blue). Upon substrate and ATP binding to GroEL cis ring it proceeds to relaxed (R) state (red). GroES (dark blue) binds and causes the release of the substrate protein into the cavity (dark red). Time required for ATP hydrolysis is ~10 s which allows the encapsulated protein to fold in isolation. Upon completion of ATP hydrolysis in cis ring (yellow) ATP binding to trans ring causes GroES and substrate to be released from the cis ring. Former trans ring becomes cis and the cycle repeats (reproduced from Clare *et al.*[79]).

Upon binding of ATP to the binding pocket of the equatorial domain, large rigid body movements occur. Conformational states along the pathway have been observed by fluorescence[80] and cryo-EM[79]. In the apo-state where both of the GroEL rings are in the tense state (T), ATP binding to the cis-ring causes the transition to the first relaxed state (Rs1) (Figure 19). In this state an en-bloc, 35° sideways tilt of the apical and intermediate domains around the lower hinge occurs. As a consequence, the ATP binding pocket closes. Additionally, the salt bridges R197-E386 and E255-K207 between the intermediate and apical domains, and between the equatorial domains are broken, respectively. New salt bridges are formed: K80-E386 and E255-K245 in adjacent equatorial and apical domains, respectively. Next, in the second relaxed state (Rs2), the upper hinge bends and the apical domain undergoes an additional lift without breaking the newly formed salt bridges. Finally, after breakage of the newly-formed salt bridges, the apical domain moves radially outwards and elevates for 20° to form the third relaxed state (Rs-open). This radial movement and elevation can stretch a bound substrate. As an effect, certain kinetically trapped intermediates may be resolved. Moreover, it is in this state that GroES, via its mobile loops, can bind to the GroEL apical domains. In the final step after GroES binding, the apical domains rotate 100° clockwise and complete their elevation and thus reach the R-ES conformation. This large rotation is thought to be responsible for the release of the captured substrate into the cavity, which has changed in character from hydrophobic to hydrophilic. The equatorial domains do not undergo such profound conformational changes, but only minor tilting and twisting motions. Since substrate binding and rearrangement is slower ($\sim 5s^{-1}$) than ATP binding ($\sim 100s^{-1}$), ATP would bind the GroEL cis ring in either its Rs1 or Rs2 state followed by GroES binding ($\sim 1\text{-}2s^{-1}$)[79]. The substrate is given a chance to fold in isolation for about 7-10 seconds, the time required for hydrolysis of seven ATP molecules to ADP within cis cavity (at 25°C). Afterwards, seven new ATP molecules bind to the trans ring and cause the release of ADP, GroES and encapsulated substrate. If the protein reaches the native state after one round of assisted refolding, it is released to bulk solution. Otherwise it re-binds GroEL. Now the former trans ring has become cis ring and another round of assisted refolding can commence. Essentially, the cycles as described make GroEL a two-stroke engine, allowing only one of the cavities at the time in assisting protein folding[50]. Apart from the prevalent mode of encapsulation as a means of assisting refolding, there are GroEL substrates which have a mass larger than 60 kDa[71]. For such proteins refolding may occur via trans folding, which means that the substrate is bound to apical domains of a trans ring where it undergoes conformational stretching and is released into solution after GroES binding to the opposite cis ring[81].

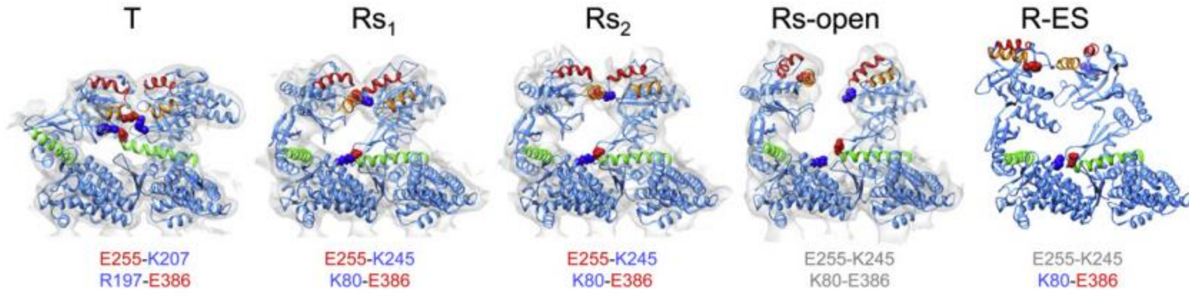


Figure 19: Series of formation and breaking of salt bridges in GroEL cis ring during reaction cycle.

GroEL crystal structure (PDB: 1OEL) shown in ribbon representation was flexibly fit into cryo-EM density (EMDB. 1997-2000) shown in white. For comparison, GroEL/ES crystal structure in ribbon representations is shown (PDB: 1SVT). Shown are two GroEL subunits interacting within a ring. Helices H, I and M are in red, orange and green, respectively. Charged residues involved in intra-ring interactions are shown in space-filling representation. Negatively charged residues are in red and positively charged in blue. GroEL ring starts in tense (T) state where E255-K207 and R197-E386 salt bridges are formed. Upon ATP binding the GroEL ring proceeds to Rs1 state characterized by formation of E255-K245 and K80-E386 salt bridges. In next state, Rs2, the salt bridges are preserved, however, when the cis ring reaches Rs-open state both former salt bridges are broken. Upon ES binding, the GroEL cis ring is in its final state R-ES where K80-E386 salt bridge is re-established (reproduced from Clare *et al.*[79]).

It has been reported that the folding of substrate in presence of GroE system experiences an increase in folding rate. There are several models proposed in the field to explain this phenomenon[82].

2.4.7 GroE mechanism of protein folding

The passive (Anfinsen) cage model postulates that the increase in a protein's refolding rate by the GroEL/ES system is due to the prevention of reversible aggregation[83]. The system plays no direct role in modulating the folding landscape of the substrate but merely allows folding to occur at infinite dilution, thus preventing any unwanted aggregation. Folding in free solution, on the other hand, generates aggregates, which can gradually disaggregate and allow monomers to fold. This reversible aggregation/disaggregation step would result in a slower apparent spontaneous refolding rate. Experimental work supporting this view was performed by dynamic light scattering measurements on the refolding of DM-MBP in absence or presence of GroEL/ES[84]. It is clear from this experiment that aggregates, which are otherwise absent in presence of chaperonin, form during spontaneous refolding. However, the reversibility of aggregation was not demonstrated.

The iterative annealing model states that the GroE system has an active role in refolding even in the absence of reversible aggregation[85,86]. The active contribution of GroEL would consist of successive rounds of substrate binding, stretching and release into either bulk solution or the chaperonin cavity. Therefore, the GroEL inner cavity plays no critical role in assisted protein folding. What is crucial in this model is stretching of the bound substrate, which exerts forces large enough to break non-native contacts and unfold kinetically trapped intermediates. Such a 'forced unfolding' event is followed by substrate release either into the cavity or into solution. The protein is then given another chance to partition between either productive or unproductive folding trajectories. The accumulated body of evidence shows that an average GroEL substrate folds in ~30-60 seconds, which at ~7-10 seconds per cycle at 25°C comes to ~3-10 cycles per folding event[71,87]. This observation would support the need for iteration. Additionally, theoretical studies[88] on polymers showed that iterative annealing can be a way to increase folding rate, however, studies with single ring variant of GroEL (SR1-EL), where absence of an allosteric signal from the second ring makes SR1-EL/ES complex long-lived[89], have shown that a single round of encapsulation is sufficient for a substrate to refold[90,91] at the same rate as during cycling. In that regard, additional, modified view of iterative annealing has been proposed based on MD simulations of a GroEL-encapsulated β-hairpin[92]. After the first and only cycle, the substrate, after being encapsulated, would iteratively bind to and release from the inner cavity wall. The wall functions as a binding surface and participates in the formation of successive folding intermediates that lead to the native state. Such an interpretation, however, already departs from the classical forced unfolding model.

The active cage model attributes an active role in refolding to the environment within the cavity[90,91,93,94]. Therefore it is of utmost importance that the substrate after the binding and stretching event is displaced into the cavity after GroES binding[95,96]. Encapsulation has a dual role, it prevents aggregation and it provides a folding environment markedly different from the one free in solution. The character of the cavity wall changes upon ATP and GroES binding from hydrophobic to hydrophilic and acquires a net negative charge of -42 clustered in two layers[75]. Charge clusters are conserved among GroEL homologs, suggesting a function in assisted protein folding[97].

Several experimental studies thus far have shown the importance of the cavity in accelerating the refolding of substrates[90,91,93]. Two obvious questions come to mind, how is this environment different from the one in bulk solution, and by what means is this realized?

A recent GroEL mutation study has shown[91], that in the GroEL mutant termed KKK2 (Figure 20), where the second negative charge cluster is mutated to positive, resulting in net neutral charge, the ability of the chaperonin to accelerate folding was abolished. This directly points to the importance of negative charges lining the inner cavity wall. Additionally, in the same study and later confirmed[90], the first experimental observation on the role of the confining effect of encapsulated, folding substrate was shown. According to polymer theory, confinement would reduce the entropy of the unfolded chain by sterically preventing extended conformations. This view was later challenged[98] by taking into account the effect of confinement not only on polypeptide chain but on solvent molecules as well. Nevertheless, the confining effect is necessary but not sufficient for accelerated folding, as demonstrated by the GroEL KKK2 mutant[91]. Despite confinement, GroEL KKK2 cannot accelerate substrate folding. One might speculate that the cavity upon encapsulation has a dual role. It isolates the protein and provides a structural framework to position negative charges, which then point into the cavity and influence the chemical environment of the cavity[90] or participate in interactions with engulfed substrate[99]. In the latter case, an intricate interplay between both substrate and inner-wall physicochemical properties would dictate the folding propensity. It was shown by MD simulation[99], that only if the cavity has a repulsive character relative to the substrate would the refolding be accelerated. Later on, another role for negative charges was suggested by MD simulation[100]. It was proposed that the negative charges in the cavity influence water structure, and in turn the folding process. According to that study[100] in wild type GroEL, solvent in the cavity displays a higher degree of structure. The degree of water structure was defined as number of water molecules 1 nm from the cavity wall. Simulation showed a correlation between experimentally observed DM-MBP refolding rates with different GroEL charge mutants and number of water molecules as determined from the simulation. Interestingly, the authors showed a higher water density at the level of the second charge cluster. Higher water density would cause the unfolded substrate to pay a larger thermodynamic cost for not being able to bury its hydrophobic amino acid residues. Essentially, as compared to bulk solution, the hydrophobic effect would be stronger inside the cavity. This led to the conclusion that the negative charges on the inner cavity wall accumulate water, and due to a stronger hydrophobic effect the encapsulated protein folds faster.

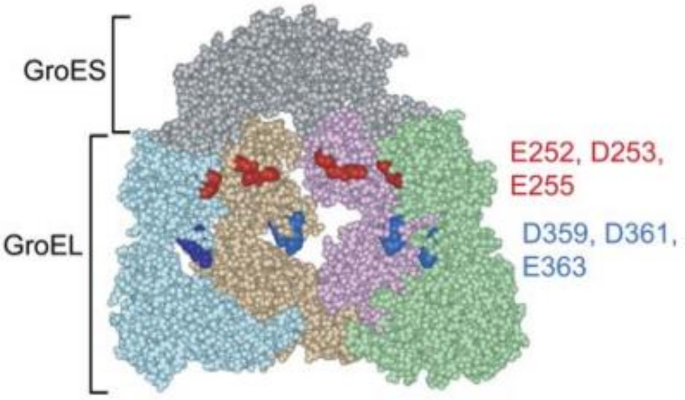


Figure 20: Inside of SR1-EL/ES complex.

Cross-section of one ring of GroEL/ES complex shown in space-filling representation (PDB: 1AON). Inside view of four GroEL/ES subunits. Two conserved charged clusters are coloured red and blue, respectively (reproduced from Tang *et al.*[91]).

So far only one experimental study[101] has attempted to detect any increase in water density or decrease in water diffusion inside the GroEL cavity upon ATP binding and GroES complex formation. For this, the authors used a single ring variant of GroEL D398A (SR1-EL D398). GroEL D398A mutant can bind ATP but it hydrolyses it at 2% rate of wild type GroEL. The D398A mutation in the context of SR1-EL makes the complex with GroES even more stable. In order to measure water properties, the GroES Y71C mutant was used. It has a cysteine residue that points into the cavity upon complex formation and is suitable for spin labelling in an NMR experiment. The result of the study was negative: no changes in water properties upon complex formation were shown. A reason could be that only local changes in proximity of charged clusters play the role, or that the water structuring effect is significant only in the presence of a substrate.

This clearly opposing results from simulation and experiment demand that more studies be performed on the question of how charges accelerate GroEL/ES assisted protein folding.

2.4.8 GroEL model substrate DM-MBP

Maltose binding protein (MBP) is an *E. coli* periplasmic protein of ~41 kDa. Removal of the signal peptide results in a cytosolic protein that can fold robustly and rapidly. MBP has two domains, discontinuous in sequence (Figure 21). The maltose binding site is located in the cleft between the domains[135]. MBP is convenient for protein folding studies due to eight intrinsic tryptophans that are spaced throughout the sequence. Upon denaturation, the tryptophan signal is reduced five-fold. As confirmed by other experiments (CD, maltose binding), the change in tryptophan fluorescence reports on global folding or unfolding.

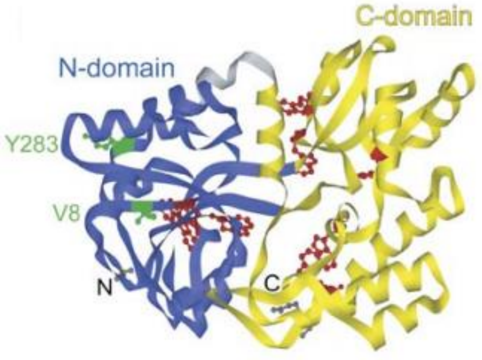


Figure 21: Structure of MBP.

Crystal structure of MBP (PDB: 1OMP) shown in ribbon representation. DM-MPB has two domains discontinuous in sequence with maltose binding pocket in the cleft between the N-domain (blue) and C-domain (yellow). Residues V8 and Y283 are shown in ball and stick representation and are coloured in green. Tryptophan residues are shown in ball and stick representation and are coloured in red. N- and C-termini are indicated (reproduced from Tang *et al.*[91]).

There are several destabilizing mutants of MBP which have reduced folding kinetics[136,137]. A specific mutant of MBP (DM-MBP) with two point mutations (V8G, Y283D) in the N-domain, was shown to strongly interact with GroEL. Furthermore, it was shown that GroEL accelerates the folding of DM-MBP up to 10-fold[91]. Recent studies have shown that DM-MBP, upon dilution from denaturant, forms a collapsed kinetically trapped intermediate on a millisecond time scale. This intermediate has almost no secondary structure and is structurally flexible[90]. GroEL binds this intermediate on a millisecond timescale[138] and causes partial expansion. Upon ATP binding, the subsequent conformational changes cause further DM-MBP stretching. Following GroES binding DM-MBP is displaced into the cavity, where it adopts compact conformation and has the opportunity to fold to the native state.

2.5 GroEL/ES substrates

Under normal cellular conditions no more than 10% of cytosolic proteins utilize the GroEL/ES system for folding (Figure 22)[71]. The identification of GroEL/ES substrates from *E. coli* spheroplasts was achieved by affinity chromatography and subsequent mass spectrometry. This study led to the identification of ~250 proteins, among which were components of the transcription/translation machinery, metabolic enzymes, chaperones and structural proteins. The identified proteins were divided into three classes based on their GroEL dependency[71]. Class I substrates (38 proteins), were found inside GroEL/ES complex, but do not stringently need the chaperonin for folding. Class II substrates (126 proteins) are defined as those that require GroEL/ES for productive folding, but can be folded by Hsp70 system as well. Class III substrates (84 proteins) are the most aggregation prone and can fold at 37 °C only in the presence of GroEL/ES. Interestingly, among class III substrates, almost 50% percent have a $(\alpha\beta)_8$ TIM-barrel fold with a complex topology and many long-range interactions. This is significantly more than the 6.8% occurrence of the TIM-barrel fold among all *E. coli* lysate proteins. A later study[102] recapitulated earlier findings and found a reduction in the number of these substrates upon GroEL depletion in *E. coli* cells. Based on depletion experiments a new class IV within the former class III was defined. Class IV substrates (49 proteins) are defined as those that aggregate or are degraded upon depletion of GroEL.

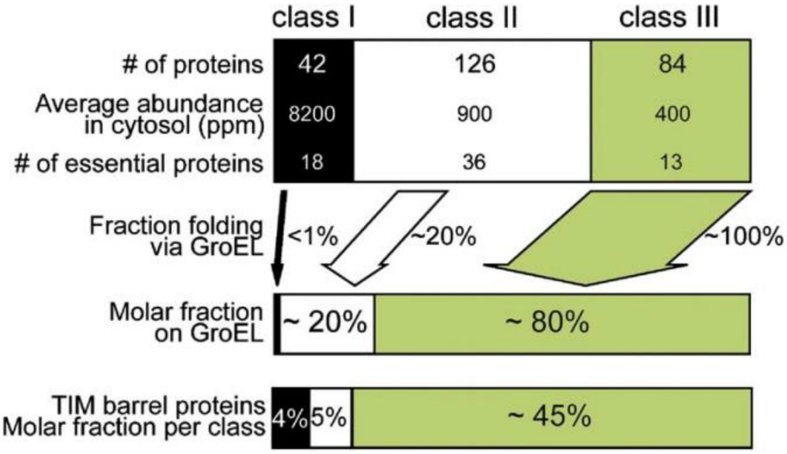


Figure 22: GroEL interactors.

There are ~250 proteins that interact with GroEL and can be divided into three classes. Class I substrates do not require GroEL for successful folding. Class II substrates can be folded by DnaK/J system and do not necessarily require GroEL. Class III substrates are stringent GroEL substrates and can fold at 37°C only in presence of GroEL. Interestingly, almost 50% of class III substrates have TIM barrel fold (reproduced from Kerner *et al.*[71]).

The analysis of GroEL substrates showed that most of them have a mass lower than 60 kDa[71], which allows them to fit into the cavity (Figure 23). Most of the substrates have a net-negative charge which coincides with net-negative charge of the cavity wall. A comparison with GroEL independent substrates showed that GroEL substrates are less hydrophobic. Moreover, *E. coli* homologs from an organism *Ureaplasma urealyticum*, which has no GroEL, have more hydrophobic amino acid residues. It might be that an increase in hydrophobicity causes a more efficient collapse due to increased hydrophobic effect. This would be analogous with the effect due to increased water structure in GroEL cavity.

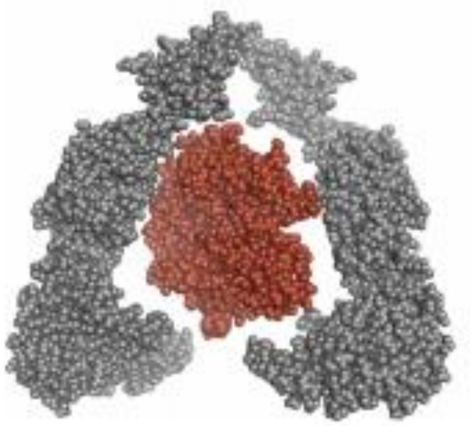


Figure 23: Substrate protein S-adenosylmethionine synthase (MetK) inside GroEL cavity.

Cross-section of one ring GroEL/ES complex (PDB: 1PF9) shown in space-filling representation. Inside the GroEL/ES cavity crystal structure of MetK (42kDa) (PDB: 5H9U) shown in space-filling representation coloured red is modelled. In its molten globule state MetK would have ~30% expanded volume (reproduced from Hayer-Hartl *et al.*[50]).

Besides using natural *E. coli* GroEL substrates, it is possible to study GroEL function using heterologous proteins which normally interact with mitochondrial or chloroplastic chaperonins, such as mitochondrial rhodanese (mRho) or *Rhodospirillum rubrum* Rubisco, respectively. Another type of substrate is a double mutant of maltose binding protein (DM-MBP) which in its natural setting never interacts with the chaperonin. Recently[95,96], it has been proposed, for two heterologous substrates, rhodanese and DM-MBP, that the substrate can 'escape' from the cavity during the refolding. Escape was described as an event where the aforementioned substrates approach and protrude through ‘windows’ (~10 Å in size) in the GroEL/ES complex cavity wall. Moreover, even more substrate escapes in the presence of a specific antibody against that substrate. Such an escape was termed 'forced escape'. Another study[103] addressed the question of escape by comparing the rates of ATP, substrate and GroES binding, and conformational changes of apical domains of SR1-EL. According to that study, 'substrate escape' is displacement of the substrate upon GroES binding due to overlap in binding sites of substrate and GroES. Additionally, one has to consider differences between non-native states of homologous and heterologous substrates when binding GroEL. Most probably, in case of heterologous non-native substrates, fewer hydrophobic residues are exposed which results in weaker binding. Moreover, addition of ATP and associated conformational changes of apical domains further increase the K_D for substrate binding, which allows for easier displacement by GroES. For homologous or highly aggregation prone heterologous substrates, escape was never observed for either GroEL or its single ring variant.

2.6 RuBisCo – a highly chaperone dependent protein

Ribulose-1,5-bisphosphate-carboxylase/-oxygenase (Rubisco) is the key enzyme in the Calvin-Benson-Bassham cycle (CBB). It is responsible for carbon fixation in all photosynthetic organisms[104]. It catalyses the reaction of carboxylation of a sugar substrate ribulose-1,5-bisphosphate (RuBP), which from an unstable carboxyketone (Figure 24) splits into two molecules of 3-phosphoglycerate (3PG). The latter is then converted to glyceraldehyde-3-phosphate (G3P). G3P is used for the synthesis of sugar molecules, fatty acids and amino acids, as well as regeneration of RuBP. Due to the electrostatic similarity of CO_2 and O_2, and the atmospheric abundance of O_2 (21%) over CO_2 (0.04%), Rubisco cannot completely discriminate between them, which leads to oxygenation of RuBP in a process called photorespiration[105]. Oxygenation of RuBP results in one molecule of 3PG and one molecule of 2-phosphoglycolate (2P-glycolate). 2P-glycolate is toxic and has to be recycled in an ATP-dependent manner via the mitochondrial-peroxisome pathway, resulting in a loss of CO_2. Photorespiration has long been perceived as an energetically wasteful process. However, recent studies show that it plays an important role in other plant metabolic pathways[106], such as nitrate assimilation[107].

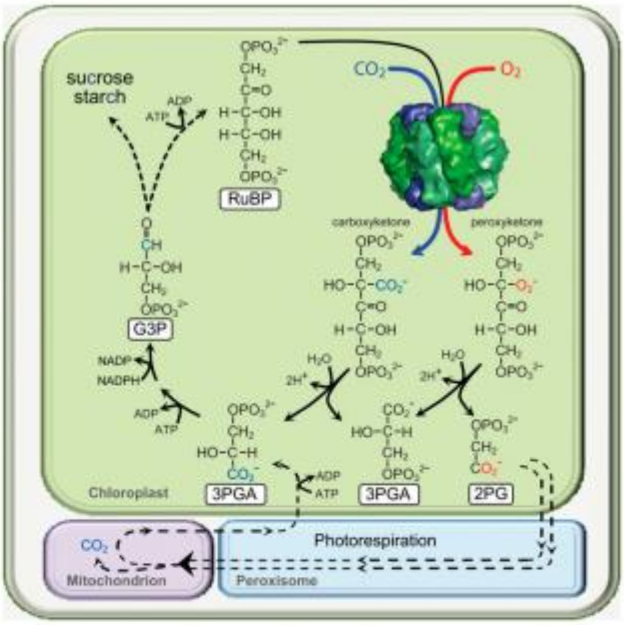


Figure 24: Rubisco catalytic cycle.

Rubisco substrate RuBP is carboxylated and transformed to unstable carboxyketone which then splits to two molecules 3-phosphoglycerate (3PG). 3PG is then converted via series of steps to glyceraldehyde-3-

phosphate (G3P) which is then used for the biosynthesis of other macromolecules. In a competing reaction O_2 reacts with RuBP which forms peroxyketone that splits to one molecule of 3PG and one molecule of 2-phosphoglycolate (2-PG). 2-PG is toxic to cell and is recycled to 3PG through photorespiration involving peroxisome and mitochondrion. In the process ATP is used and fixed CO_2 is lost (reproduced from Whitney *et al.*[108]).

2.6.1 Structure of Rubisco large and small subunits

There are several structural forms of Rubisco found in nature[109]. Plants, algae, cyanobacteria and proteobacteria have form I Rubisco. Form I is a hexadecameric structure composed of eight large (RbcL, ~50 kDa) and eight small (RbcS, ~15 kDa) subunits. Eight RbcL subunits are arranged as a tetramer of anti-parallel dimers ($RbcL_8$ core complex). The core complex is capped on the top and bottom by four RbcS subunits. The holoenzyme ($RbcL_8RbcS_8$) with a total mass of ~520 kDa has a cylindrical shape with a diameter of ~110 Å and a height of ~100 Å (Figure 25). Form II Rubisco, which is found in certain bacteria and dinoflagellates, is simpler. It consists only of two RbcL subunits, which form a dimer without RbcS subunits.

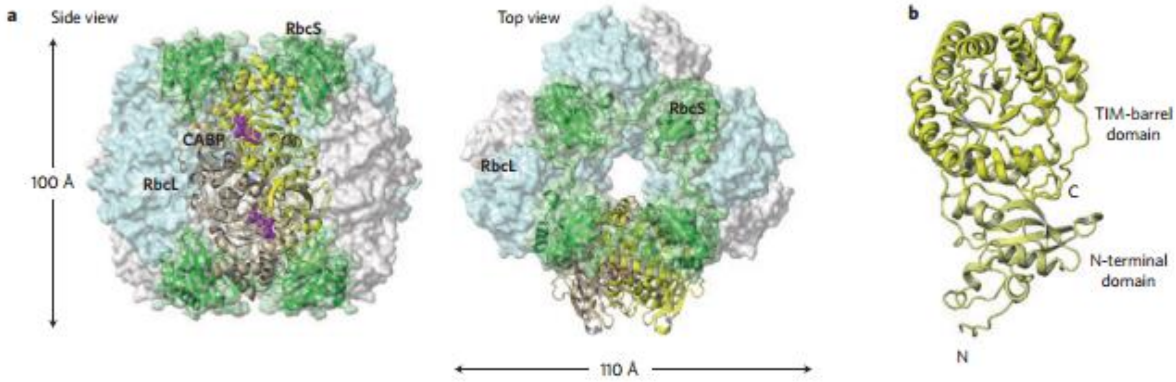


Figure 25: Rubisco form I structure.

a.) Structure of Form I Rubisco (PDB: 1RBL) from *Synechococcus elongatus* PCC6301 shown in surface representation. Alternating large subunits are coloured cyan and grey. Small subunits are coloured green. One $RbcL_2$ antiparallel dimer is shown in ribbon representation coloured grey and yellow. Inhibitor caraboxyarabinitol-1,5-bisphosphate (CABP) bound in the active site is shown in space-filling representation and is coloured violet. b.) Crystal structure of single Rubisco large subunit. It has two domains, N-terminal α+β domain and C-terminal TIM barrel domain. N- and C- termini are indicated (reproduced from Hauser *et al.*[104]).

Form III Rubiscos found in archaea consist of three to five RbcL dimers arranged in ring-like assemblies. Unlike Form I and Form II Rubiscos, Form III Rubisco is not involved in photosynthesis, but is involved in regeneration of RuBP in nucleotide biosynthesis[109]. Despite high sequence conservation[110] (~80% amino acid identity) form I RbcL sequences can be phylogenetically classified into green-type enzymes (Figure 26) (forms IA and IB of plants, green algae, cyanobacteria) and red-type enzymes (form IC and ID of non-green algae and phototrophic bacteria).

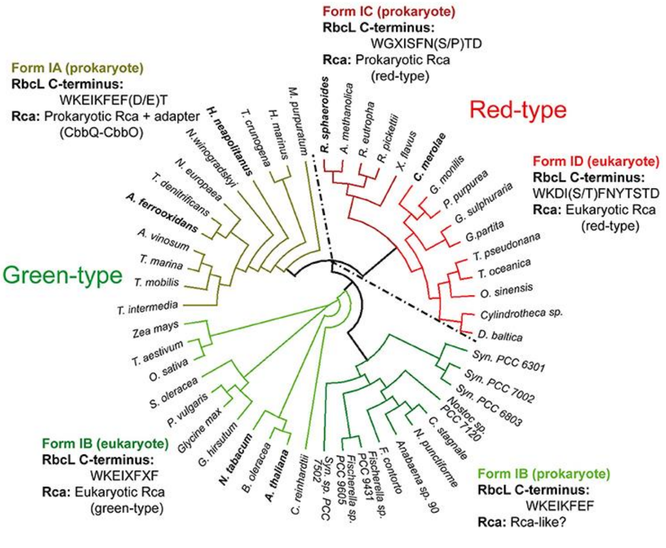


Figure 26: Rubisco phylogenetic tree.

Phylogenetic tree of selected Rubisco sequences. Form IA and IB belong to green-type Rubiscos whereas forms IC and ID to red-type Rubiscos. The RbcL C-terminal sequences and their cognate activases are indicated. X represents variable residues. For species in bold activases have been biochemically and/or structurally characterized (reproduced from Bhat *et al.*[105]).

In contrast to RbcL, RbcS subunits have lower sequence conservation[110] (~30-40% amino acid identity). The diversity of RbcS subunit sequences is responsible for functional differences between Rubisco enzymes. In eukaryotes, RbcL is always plastid encoded. RbcS however, is plastid encoded in non-green algae and nuclear encoded in green algae and plants[111]. Nuclear encoded RbcS have a cleavable N-terminal targeting sequence for chloroplast import.

The RbcL subunit consists of an N-terminal $\alpha+\beta$ domain of ~150 amino acids and a ~325 amino acids long C-terminal domain which consists of $\beta_8\alpha_8$ triosephosphate isomerase (TIM) barrel domain (~320 amino acids) and an unstructured C-terminal tail (~15 amino acids in green type Rubisco and~25 amino acids in red type rubisco)[109] (Figure 25). The antiparallel RbcL dimer has two active sites formed by residues of the opposing N-terminal domain of one subunit and the C-terminal TIM-barrel domain of the adjacent subunit[109].

2.6.2 Rubisco catalytic cycle

In order for Rubisco to successfully catalyse CO_2 fixation, it first has to be activated. Activation of Rubisco enzyme (E) is achieved by a spontaneous carbamylation reaction where one non-substrate molecule of CO_2 reacts with active site Lysine 201, followed by binding of an Mg^{2+} ion[112]. Rubisco is now in its activated form (ECM). Activation of Rubisco is necessary because chemical modification of the active site allows for correct positioning of the RuBP which is essential for a successful electrophilic attack by the second molecule of CO_2 that will be fixed into RuBP[110]. Binding of RuBP to the active site of activated

Rubisco is followed by movement of the retracted mobile loop 6 of the TIM barrel over the active site. Loop 6 is pinned down in its new position by the unstructured C-terminal tail through a so called C-terminal latch[113]. These conformational changes result in closing of the active site and its dehydration, which is mandatory for a successful electrophilic reaction.

2.6.3 Folding and assembly of Rubisco

Due to its complex topology[109] the RbcL subunit needs help from chaperonins to reach its unique three-dimensional structure[114]. Plastid encoded RbcL relies on Cpn60/Cpn10, which are chloroplast homologs of bacterial GroEL/GroES. In-vitro experiments have shown that Form II RbcL can assemble into active enzyme after GroEL/ES assisted refolding[114]. Form I RbcL subunit, however, fails to assemble and stays associated with GroEL, despite having reached its native state[115]. This observation suggested that there have to be additional factors which facilitate assembly of the $RbcL_8$ core followed by formation of $RbcL_8S_8$ holoenzyme. The protein encoded by the *rbcX* gene located in the intergenic space of the *rbcL* and *rbcS* genes seemed as a possible candidate for the assembly chaperone. Indeed, co-expression of RbcX with RbcL and RbcS in *E. coli* increased the yield of enzymatically active Rubisco[115]. Further biochemical and structural studies showed that RbcX is a homodimer of ~15 kDa subunits which functions as a specific assembly chaperone. Cyanobacterial Form I Rubisco could recently be reconstituted in vitro in the presence of GroEL/ES and RbcX. After assisted refolding, RbcX binds the unstructured C-terminal of RbcL and mediates proper antiparallel alignment of two RbcL subunits, thus functioning as a molecular clamp. Each RbcL anti-parallel dimer has RbcX bound on top and on bottom. Such stabilization of the dimer then shifts the equilibrium towards the $RbcL_8{:}RbcX_8$ assembly intermediate. In the final step, RbcS displaces RbcX and enzymatically active $RbcL_8{:}RbcS_8$ is formed (Figure 27).

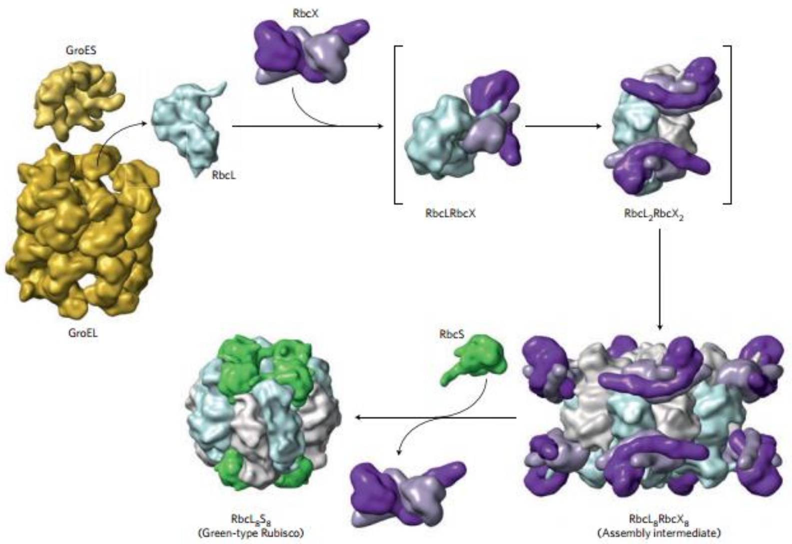


Figure 27: Green-type Rubisco folding and assembly.

Upon GroEL/GroES assisted refolding and release of cyanobacterial RbcL, RbcX binds the unstructured C-terminal tail and mediates proper alignment of the RbcL dimer. $RbcL_2$:$RbcX_2$ complex is followed by assembly into $RbcL_8$:$RbcX_8$ complex. RbcS binding to pre-formed binding interface in $RbcL_8$:$RbcX_8$ leads to structuring of RbcL N-terminus and 60s loop which results in displacement of RbcX and formation of $RbcL_8$:$RbcS_8$ holoenzyme(reproduced from Hauser *et al.*[104]).

2.6.4 Metabolic maintenance of Rubisco

Despite its pivotal role, Rubisco is an inefficient enzyme. It fixes only ~3-10 molecules of CO_2 per second in plants. The slow rate of Rubisco catalysed incorporation of CO_2 is to a degree compensated by the significant amount of Rubisco, which can reach up to 50% of soluble leaf protein[116]. Another drawback of Rubisco is that RuBP can bind to the active site before the enzyme activation event takes place, leading to an inactive enzyme. Additionally, there are other inhibitory sugar-phosphates, some of which are mis-fire products of Rubisco's multistep catalytic reaction (XuBP, PDBP), while others (CA1P) function as a night-time inhibitors which are produced by plant under low-light conditions[117]. In all cases mentioned above, Rubisco is trapped in its closed, inhibited form (EI or ECMI) (Figure 28).

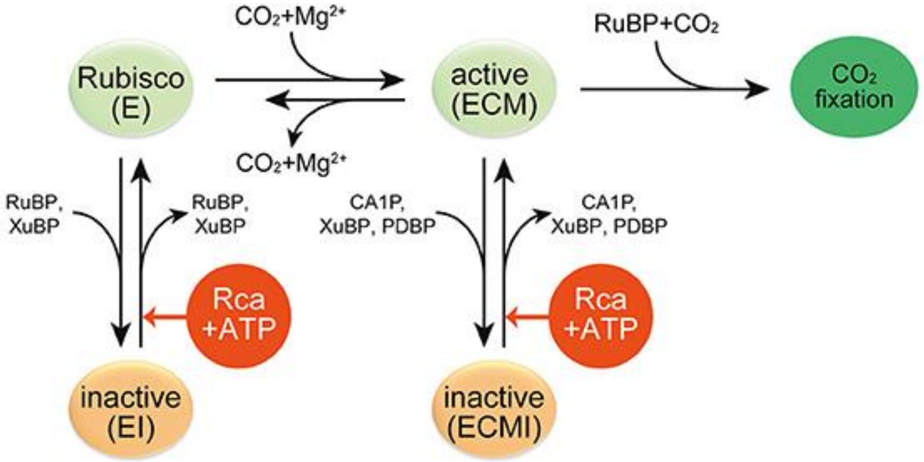


Figure 28: Sugar phosphates involved in Rubisco inhibition.

Rubisco activity is regulated or inhibited by various sugar phosphates. E, the non-carbamylated enzyme; ECM, carbamylated and Mg^{2+} ion bound enzyme; EI, the sugar phosphate inhibited E form; ECMI, the inhibited ECM form; Rca, Rubisco activase (reproduced from Bhat *et al.*[105]).

Rubisco can be rescued from its inactive form by spontaneous opening of the active site, however to reactivate Rubisco on a biologically relevant time scale, evolution has equipped many photosynthetic organisms with an enzyme called Rubisco activase (Rca)[118].

2.6.5 Rubisco activases

Rca belongs to a superfamily of proteins termed AAA+ (ATPases associated with various cellular activities)[118]. These are macromolecular machines that are powered by ATP hydrolysis for unfolding, disassembly or remodelling of their substrates[119]. Rca metabolically repairs Rubisco by opening the active site and releasing the inhibitory sugar[120]. Rubisco can then be reactivated and acquire its catalytically competent state. Rca proteins have been found to reactivate both red- and green-type Rubisco in organisms ranging from chemoautotrophic bacteria to higher plants[121,122,123]. Despite considerable sequence variability among different activases[124], they all share a common domain architecture characteristic of AAA+ proteins[125], consisting of N-terminal α/β nucleotide binding subdomain and a C-terminal α-helical subdomain[105]. The functional form of the activase is a ring-shaped hexamer with a central pore, where pore loop residues indispensable for Rubisco remodelling are located[126,127] .

2.6.5.1 Red-type activases

Red-type activases, which remodel form IC and form ID Rubisco (red-type) have an additional α-helical extension at the N-terminus[126]. The two subdomains of the AAA+ core module are separated by a flexible linker. Walker A and Walker B nucleotide binding motifs are located in the N-terminal α/β subdomain (Figure 29).

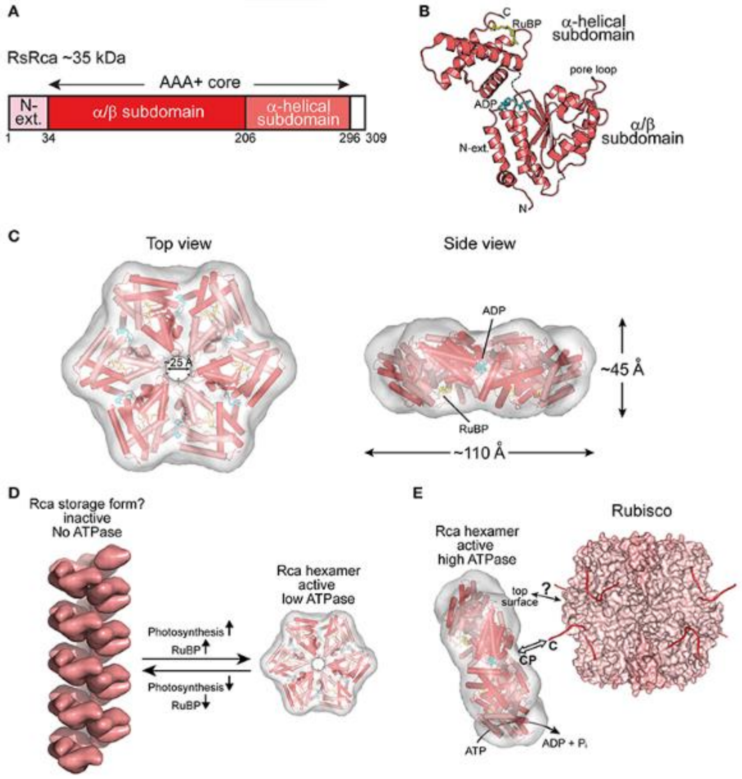


Figure 29: Red type Rubisco activase.

a.) Domain organization of red type RsRca. b.) Crystal structure of RsRca protomer (PDB: 3SYL) shown in ribbon representation. α-helical and α/β subdomains are indicated. Bound ADP; RuBP, pore loops and

N-terminal extension are indicated. c.) rsRca ring-like hexameric model fitted into EM reconstruction (EMDB: 1932) shown in top and side views: Bound ADP and RuBP are indicated in side view. d.) rsRca filament formed in absence of RuBP. Filamentous form would serve as a storage for rsRca in absence of photosynthesis. Start of photosynthesis causes the accumulation of RuBP to millimolar concentrations. Free RuBP binds to Rca leading to its hexamerization. e.) Model of rsRca interaction with Rubisco where rsRca docks onto Rubisco and transiently pulls by Rubisco C-terminal tail into the central pore. This opens the Rubisco active site and releases inhibitory sugar. Rca is shown as in c.) Rubisco structure (PDB: 4F0K) is shown in surface representation. RbcL and RbcS are shown in different shades of pink. RbcL C-terminal tails are depicted in red lines (reproduced from Bhat *et al.*[105]).

Red-type activase from the proteobacterium *Rhodobacter sphaeroides* (RsRca) attains its functional homohexameric structure in the presence of RuBP and ATP[126] (Figure 29). RuBP and Rubisco stimulate RsRca ATPase activity. In the absence of RuBP, RsRca forms high-molecular weight filaments which are ATPase inactive and are believed to represent a storage form for the activase in absence of photosynthesis (Figure 29). Reactivation of the red-type form IC Rubisco by the remodelling activity of RsRca depends on two elements: the extended C-terminal tail of RsRbcL[126] and conserved pore loop residues (Tyr/Ile/Gly) of RsRca lining the central channel (diameter of ~25Å) of the active hexamer[126]. Pore loops were implicated in exerting the pulling force on the extended C-terminal tail of the RbcL subunit. This puts forward a mechanism where RsRca docks onto Rubisco with its conserved top surface and pulls the C-terminal tail into the pore thereby opening the active site which results in the release of the inhibitory sugar[105] (Figure 29).

Eukaryotic Rca from red algae *Cyanidioschyzon merolae* (CmRca) is a hexamer consisting of two different protomers in a 1:1 ratio[122], and exerts its remodelling activity on Form ID Rubisco. One of the protomers is nuclear encoded and the other is plastid encoded. Both of them have 60-70% identity with RsRca[105]. In the case of CmRca, RuBP is not required for hexamerization. However, RuBP has role as an allosteric regulator of CmRca ATPase activity and thus couples the state of CBB cycle and Rubisco activity.

2.6.5.2 Green-type activases

Recently[123], prokaryotic Rubisco activases of green-type Rubisco form IA from the chemoautotrophic bacteria *Acidithiobacillus ferrooxidans* and *Halothiobacillus neapolitanus* have been described. These proteins are homohexamers (AfRcaI; HnRca) that need an additional adaptor protein CbbO to interact with Rubisco. AfRcaI and HnRca α/β subdomains can be classified as MoxR group of prokaryotic AAA+ proteins[128], which interact with proteins that have a von Willebrand factor A (VWA) domain (Figure 30).

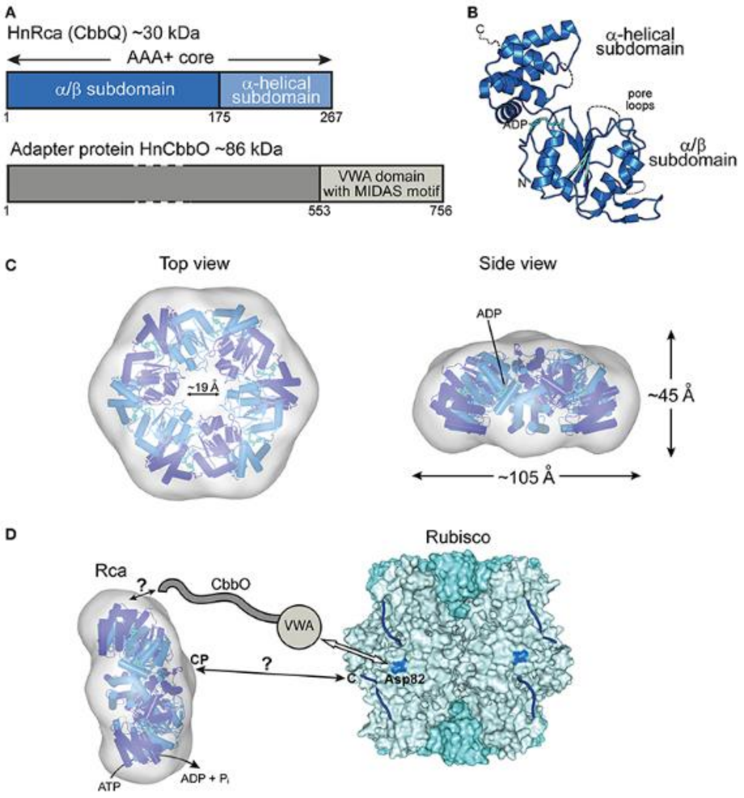


Figure 30: Prokaryotic green type Rubisco activase.

a.) Domain organization of HnRca and its adaptor protein HnCbbO. b.) Crystal structure of HnRca protomer (PDB. 5C3C) in ribbon representation. α-helical and α/β subdomains are indicated. Bound ADP and pore loops are indicated. c.) Ring-like hexameric model of HnRca fitted into EM reconstruction (EMDB: 6477 shown in top and side views. Bound ADP is indicated in side view. d.) Model of HnRca interaction with Rubisco where HnRca docks onto Rubisco via VWA domain of CbbO recognizing exposed Asp82 (marine blue). It is still not clear whether HnRca central pore engages with Rubisco. Rca is shown as in c.) Rubisco (PDB:1SVD) is shown in surface representation. RbcL and RbcS are shown in different shades of blue. RbcL C-terminal tails are depicted with blue lines (reproduced from Bhat *et al.*[105]).

Interestingly, CbbO has a VWA domain, and as is characteristic for this domain, a metal-ion-dependent adhesion site (MIDAS) mediates protein-protein interaction via a cation (usually Mg^{2+})[129] (Figure 30). The ATPase activity of AfRcaI and HnRca is stimulated by RuBP but inhibited by Rubisco and CbbO. The mode of remodelling action of AfRcaI and HnRca remains to be elucidated since recent experiments[123] showed that its remodelling activity, in contrast to RsRca, does not depend on pore loop residues. However, interaction with the C-terminal tail of RbcL is still required for reactivation[123].

One of the first Rubisco activases to be discovered[130], Rca in *Arabidopsis thaliana*, is a eukaryotic Rubisco activase of green-type Rubisco form IB. Eukaryotic activases of green-type Rubisco form IB have, in addition to the AAA+ core module, an N-terminal domain important for binding to Rubisco (localized at the top surface of the hexamer[127] for *Nicotiana tabacum*) and a C-terminal extension crucial for constitutive ATPase activity[105] (Figure 31).

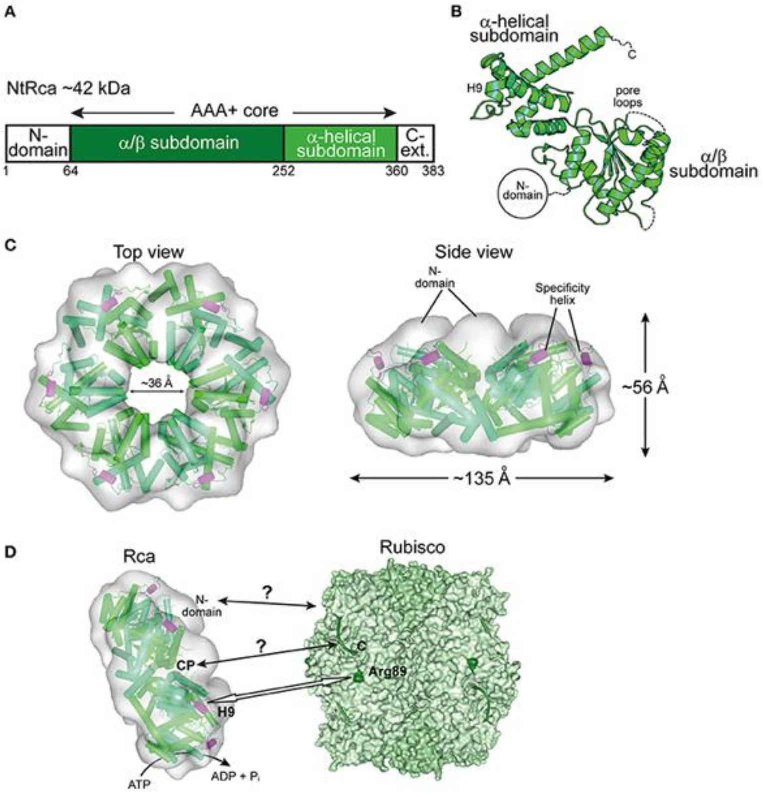


Figure 31: Eukaryotic green-type Rubisco activase.

a.) Domain organization of NtRca. b.) Crystal structure of NtRca protomer (PDB: 3T15) shown in ribbon representation. α-helical and α/β subdomains are indicated. Pore loops, specificity helix 9, N-terminal domain are indicated. c.) Ring-like hexameric model of NtRca fitted into EM reconstruction (EMDB: 1940) shown in top and side views. Additional density at the top is attributed to N-terminal domain. Specificity helix 9 is coloured in purple d.) Model of NtRca interaction with Rubisco where NtRca docks onto Rubisco via N-terminal domain and specificity helix 9 recognizes Arg89. How the ntRca interacts with Rubisco C-terminal tail still remains to be to be investigated. Rca is shown as in c.) Rubisco is shown in surface representation (PDB: 1EJ7). RbcL and RbcS are shown in different shades of green. RbcL C-terminal tails are depicted as green lines(reproduced from Bhat *et al.*[105]).

Regulation of activase hexamer formation and activity in higher plants (*A. thaliana*, rice, barley, maize and cotton) is far more intricate than for the other activases described above. One such regulatory mechanism is the formation of disulphide bonds in the C-terminal extension under oxidizing conditions (night time, in the absence of photosynthesis) which inhibits ATP binding to the activase and thereby Rubisco activation[131]. Moreover, hexamer formation of plant activases is sensitive to the ATP:ADP ratio[132], with a higher ratio favouring hexamer formation and thus the active form. Additionally, catalytic activity also depends on concentration of Mg^{2+} ions (higher Mg^{2+} concentration increases the activity in NtRca) as a response to changes in available light[133]. The mechanism of remodelling of plant Rubisco is still not established. From experiments thus far we know that NtRca engages with inhibited Rubisco via the N-terminal domain and needs conserved loop segments that face the pore to reactivate Rubisco. What is not clear, is how the activase exerts force on plant Rubisco lacking an extended C-terminal tail[105]. Interestingly, no inhibited forms of cyanobacterial Rubiscos have been reported thus far, although cyanobacteria contain genes encoding Rca-like proteins[134].

3 Publications

3.1 Article 1 - Active Cage Mechanism of Chaperonin-Assisted Protein Folding Demonstrated at Single-Molecule Level

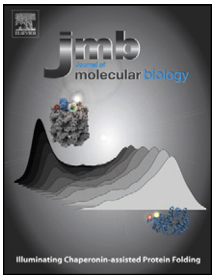
This study resolved the long-standing question whether the GroEL/ES system functions as a folding catalyst[84,93]. We also reassessed the role of the GroEL cavity in the process of assisted folding[85]. To address these questions, we used a combination of biochemical, ensemble and single molecule fluorescence, and other spectroscopic methods. Additionally, we developed a novel single molecule fluorescence approach for measuring the spontaneous and assisted folding rate of the model substrate DM-MBP. Single molecule FRET, fluorescence correlation spectroscopy (FCS) and fluorescence cross-correlation spectroscopy (FCCS) measurements were conducted at very low protein concentration (100 pM), where any unwanted protein aggregation is prevented. With this study, it has become clear that the observed increase in DM-MBP folding rate is due to an active role of the GroEL/ES system and not solely due to aggregation prevention by the GroEL/ES system. Furthermore, we showed that a single round of folding is enough to achieve an increase in folding rate, thus excluding the need for several binding, unfolding and release events as stated by the iterative annealing model. Lastly, we showed unequivocally that the substrate, once encapsulated, cannot escape from the GroEL cavity. Successful encapsulation is of utmost importance for the substrate to experience the environment within the GroEL cavity which is markedly different from one in bulk solution. Thus, the GroEL cavity does not only allow the substrate to fold in isolation, but it actively modulates the folding landscape of an encapsulated protein, resulting in an increased folding rate.

Journal of Molecular Biology (2014)[139]

Gutpa AJ, Haldar S, **Miličić G**, Hartl FU, Hayer-Hartl M.

Contribution: This project was performed in collaboration with Amit Gupta and Shubhasis Haldar. I performed encapsulation experiments using analytical gel filtration chromatography, SDS gel electrophoresis, fluorogel imaging and western blotting in order to find appropriate conditions (buffer, salt concentration, pH) for successful encapsulation of the DM-MBP by SR1-EL. These experiments showed that there is no substrate escape after encapsulation. Moreover, the established conditions for successful encapsulation by SR1-EL allowed for subsequent ensemble and single molecule fluorescence experiments to be conducted.

Featured Article



Active Cage Mechanism of Chaperonin-Assisted Protein Folding Demonstrated at Single-Molecule Level

Amit J. Gupta[†], **Shubhasis Haldar**[†], **Goran Miličić**,
F. Ulrich Hartl and Manajit Hayer-Hartl

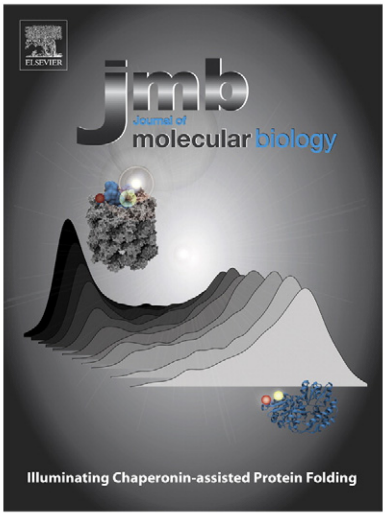
Department of Cellular Biochemistry, *Max Planck Institute of Biochemistry, Am Klopferspitz 18, 82152 Martinsried, Germany*

Correspondence to
F. Ulrich Hartl and Manajit Hayer-Hartl:
uhartl@biochem.mpg.de; mhartl@biochem.mpg.de
http://dx.doi.org/10.1016/j.jmb.2014.04.018
Edited by R. L. Gonzalez

Amit J. Gupta, Shubhasis Haldar and Goran Miličić

Abstract

The cylindrical chaperonin GroEL and its lid-shaped cofactor GroES of *Escherichia coli* have an essential role in assisting protein folding by transiently encapsulating non-native substrate in an ATP-regulated mechanism. It remains controversial whether the chaperonin system functions solely as an infinite dilution chamber, preventing off-pathway aggregation, or actively enhances folding kinetics by modulating the folding energy landscape. Here we developed single-molecule approaches to distinguish between passive and active chaperonin mechanisms. Using low protein concentrations (100 pM) to exclude aggregation, we measured the spontaneous and GroEL/ES-assisted folding of double-mutant maltose binding protein (DM-MBP) by single-pair fluorescence resonance energy transfer and fluorescence correlation spectroscopy. We find that GroEL/ES accelerates folding of DM-MBP up to 8-fold over the spontaneous folding rate. Accelerated folding is achieved by encapsulation of folding intermediate in the GroEL/ES cage, independent of repetitive cycles of protein binding and release from GroEL. Moreover, photoinduced electron transfer experiments provided direct physical evidence that the confining environment of the chaperonin restricts polypeptide chain dynamics. This effect is mediated by the net-negatively charged wall of the GroEL/ES cavity, as shown using the GroEL mutant EL(KKK2) in which the net-negative charge is removed. EL(KKK2)/ES functions as a passive cage in which folding occurs at the slow spontaneous rate. Taken together our findings suggest that protein encapsulation can accelerate folding by entropically destabilizing folding intermediates, in strong support of an active chaperonin



Legend. *The GroEL/ES-mediated folding of the model substrate protein DM-MBP is monitored by single-molecule FRET. Folding intermediate bound to GroEL shows a low FRET efficiency distribution. Upon refolding, DM-MBP converts to its compact native structure that exhibits a high FRET efficiency distribution.*

J. Mol. Biol. (2014) **426**, 2739–2754

mechanism in the folding of some proteins. Accelerated folding is biologically significant as it adjusts folding rates relative to the speed of protein synthesis.

Introduction

Chaperonins are large ATP-driven macromolecular machines composed of two rings of ~60 kDa subunits stacked back to back. They have an essential role in assisting protein folding in bacteria, archaea and eukarya [1–6]. Group I chaperonins occur in bacteria (GroEL), mitochondria (Hsp60) and chloroplasts (Cpn60). They consist of heptameric rings and functionally cooperate with lid-like cofactors (GroES in bacteria, Hsp10 in mitochondria and Cpn10/Cpn20 in chloroplasts) that function to transiently encapsulate non-native substrate protein in a cage-like compartment. Group II chaperonins in archaea and the cytosol of eukaryotic cells have rings of 8–9 subunits and employ a mechanism of opening and closing their central cavity that is in-built into the structure of the chaperonin ring.

The group I chaperonins GroEL and GroES of *Escherichia coli* have been investigated most widely and numerous biochemical and structural studies indicate that they function as a nano-compartment for single protein molecules to fold in isolation [7–23]. However, whether protein encapsulation actively promotes folding remains controversial [12,17,19,22,24–27].

Each subunit of GroEL is divided into apical, intermediate and equatorial domains [28] (Fig. 1a). The apical domains form the ring opening and expose hydrophobic amino acid residues for the binding of molten globule-like folding intermediates [29–31]. ATP binding and hydrolysis in the equatorial domains results in conformational changes that are transduced to the apical domains via the hinge-like intermediate domains, regulating substrate affinity and GroES binding through an allosteric reaction cycle [4,32]. In the current model, non-native protein substrate binds to the open ring (trans-ring) of an asymmetrical GroEL/ES complex (Fig. 1b) (The corresponding video file for the animated Abstract can be found online at doi:10.1016/j.jmb.2014.04.018). Subsequent ATP binding causes apical domain movements that may result in stretching the bound polypeptide [18,33–36], and at the same time, ADP and GroES dissociate from the opposite ring (Fig. 1b). ATP binding is closely followed by GroES binding, resulting in the displacement of the bound substrate and its encapsulation in the newly formed GroEL/ES cage (cis-ring). Encapsulated protein (up to ~60 kDa in size) is now free to fold unimpaired by aggregation in a cage with a hydrophilic, net-negatively charged wall (in cage-folding). The time allowed for folding is dependent on the rate of hydrolysis of the seven ATP in the cis-ring (~5–10 s at 25 °C). After completion of ATP hydrolysis, ATP binding to the GroEL trans-ring causes dissociation of ADP and GroES (Fig. 1b). Folded protein is released, while not-yet folded protein will be rapidly recaptured for possible stretching, encapsulation and folding. Symmetrical GroEL/ES complexes with GroES bound to both GroEL rings have also been observed, but their function in the reaction cycle is still under investigation [27,37,38].

Three models have been proposed to explain how this basic chaperonin cycle promotes protein folding. The "passive cage" (also known as "Anfinsen cage") model suggests that GroEL/ES essentially provides an infinite dilution chamber [25,39,40]. The rate of spontaneous folding, when measured in the absence of reversible aggregation, is identical with the rate of folding inside the cage. The model implies that GroEL/ES-dependent proteins fold at a biologically relevant timescale as long as aggregation is prevented. In contrast, the "active cage" model states that, besides preventing aggregation, the physical environment of the cage also modulates the folding energy landscape, resulting in accelerated folding of certain proteins. This is attributed to an effect of steric confinement that limits the conformational space to be explored during folding [12,17,19,22,41–43]. The active cage model implies that cells contain a set of proteins with kinetically frustrated folding pathways that require folding catalysis to reach their native states at a biologically relevant speed. Finally, the "iterative annealing" model posits that the function of GroEL/ES is to unfold misfolded proteins through cycles of binding and release, with folding occurring either inside or outside the cage [27,44] (out of cage-folding; Fig. 1b). In this model, accelerated folding may result from the active unfolding of kinetically trapped species that can then partition between productive and unproductive folding trajectories. The transient encapsulation of substrate is thought to be a mere by-product of the unfolding reaction [27].

In an effort to distinguish between these models, we developed novel approaches to investigate protein folding by GroEL/ES at single-molecule level. Using single-pair fluorescence resonance energy transfer (spFRET), dual-color fluorescence cross-correlation spectroscopy (dcFCCS) and photoinduced electron transfer (PET), we could exclude reversible aggregation as the cause of slow spontaneous folding and unequivocally distinguish between active and passive chaperonin mechanisms. We find that GroEL/ES accelerates the folding of a double-mutant maltose binding protein (DM-MBP) up to ~8-fold relative to its

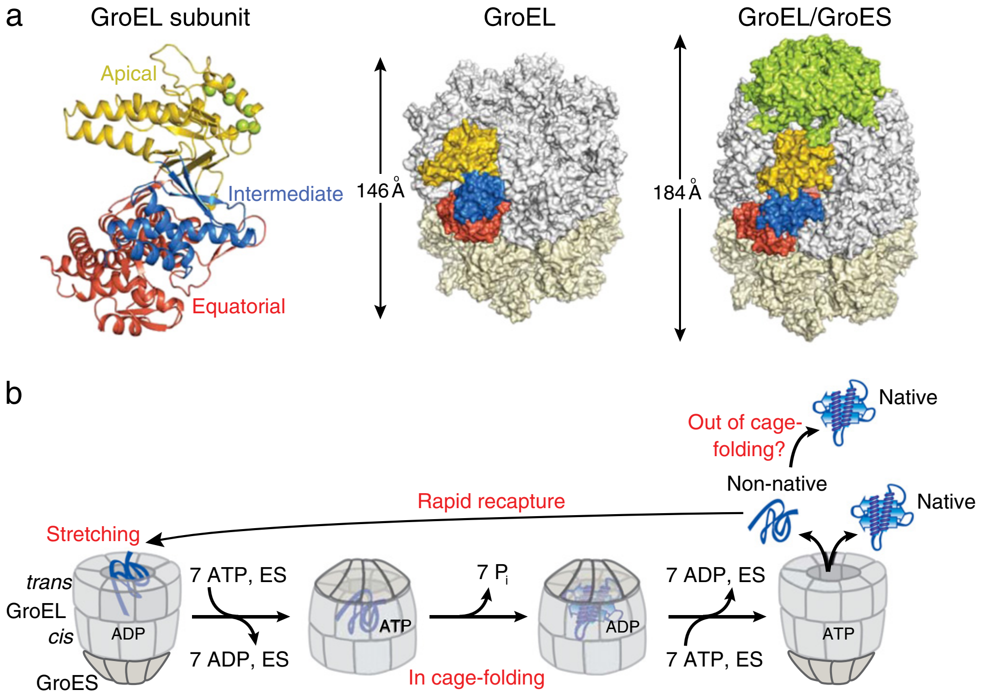


Fig. 1. Structure and function of the GroEL/ES chaperonin. (a) Left: Structure of the GroEL subunit in the apo-state in ribbon representation (PDB ID 1SS8). Apical domain, yellow; intermediate domain, blue; equatorial domain, red. Middle: Structure of the apo-GroEL 14-mer (PDB ID 1SS8). One subunit of GroEL is shown in color. Right: Structure of the GroEL/GroES complex in the ADP state (PDB ID 1PF9). The two rings of GroEL are shown in gray and beige, and GroES is shown in green. (b) Model of the GroEL/GroES reaction cycle. See Introduction and Discussion for details.

spontaneous rate. Accelerated folding occurs upon a single round of encapsulation, as demonstrated using SREL, a single-ring variant of GroEL that results in stable protein encapsulation without GroES dissociation. Thus, multiple rounds of substrate binding and release as proposed by the iterative annealing model are not required for folding catalysis. Instead, accelerated folding is due to the physical confinement of non-native protein in the net-negatively charged GroEL/ES cage. Consistent with this mechanism, we demonstrate that, during the GroEL/ES reaction cycle, the substrate protein spends ~82% of its time inside the chaperonin cage and only ~18% in the GroEL-bound state, with negligible amounts of non-native protein being free in solution.

Results

Transient aggregation is not the cause of slow spontaneous folding

We used DM-MBP (~41 kDa), a double mutant of MBP, which has previously been used as a model substrate to compare the rates of spontaneous and GroEL/ES-assisted folding [17,22,45]. DM-MBP carries mutations V8G and Y283D that delay the rate-limiting folding step of the N-domain [46] (Fig. 2a). As a result, the spontaneous refolding of DM-MBP is slow ($t_{1/2}$ ~ 35 min at 25 °C and physiological salt concentration) but nevertheless fully efficient [17,24], with only largely unstructured intermediate and the native state being populated during folding [22]. The GroEL/ES-assisted folding of DM-MBP has a 6- to 10-fold faster rate. However, there is disagreement whether the observed rate acceleration is due to GroEL/ES actively modulating the folding energy landscape [17] or to GroEL/ES passively preventing reversible aggregation that would slow spontaneous folding [24,26].

To establish conditions of spontaneous refolding in which transient aggregation is excluded, we resorted to single-molecule fluorescence measurements. A single cysteine mutant of DM-MBP, DM-MBP(312C), was labeled with the fluorophore Atto532 or Atto647N and used in dual-color fluorescence cross-correlation (dcFCCS) experiments. The native, labeled proteins were mixed 1:1 at a final concentration of 50 pM each. The probability, at 100 pM, for two monomeric DM-MBP molecules to be simultaneously present in the observation volume (1 fL) is ≤1%. As expected, no cross-correlation

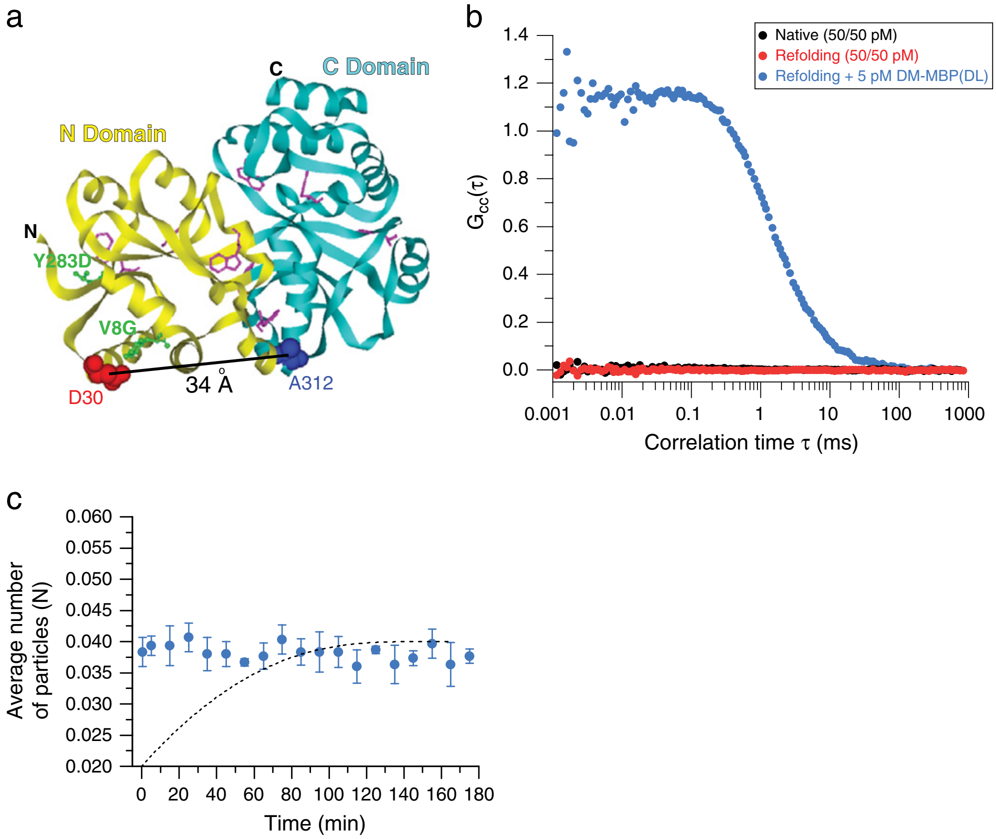


Fig. 2. Absence of DM-MBP aggregation during spontaneous refolding. (a) Ribbon structure of maltose binding protein (MBP) (PDB ID 1OMP). The discontinuous N- and C-domains are shown in yellow and blue, respectively. The positions of mutations Y283D and V8G (green) in DM-MBP, the 8 Trp residues (pink) and the cysteine substitutions at D30 and A312 used for fluorescence labeling are indicated. (b) Absence of FCCS signal $G_{cc}(\tau)$ during spontaneous refolding of DM-MBP. A 1:1 mixture of DM-MBP(312C) labeled with either Atto532 or Atto647N was denatured in 6 M GuHCl and 10 mM DTT for 1 h at 20 °C. Refolding was initiated by 200-fold dilution into buffer A to a final concentration of 50 pM each. dcFCCS was recorded with pulsed interleaved excitation [60] within the first 10 min of refolding (red). As a positive control, 5 pM native, double-labeled protein, DM-MBP(DL), was added to the mixture of single-labeled, denatured proteins during refolding to simulate the presence of an oligomeric (dimeric) species (blue). A 1:1 mixture of the native, single-labeled proteins (50 pM each) was analyzed as a negative control (black). Shown are representative results of least three independent experiments. (c) Average number of fluorescent-labeled particles inside the confocal observation volume during the course of spontaneous refolding of 100 pM DM-MBP(312C) labeled with Atto647N in buffer A. FCS was recorded over the course of 3 h. FCS analysis was performed for continuous time windows of 10 min and the average number of particles was extracted from the amplitude of the fit. Averages ± SD from three independent experiments are shown. Dotted line is the simulated increase in number of particles assuming that reversible aggregation limits spontaneous refolding.

signal was observed (Fig. 2b). To investigate DM-MBP under refolding conditions, we denatured the differently labeled DM-MBP molecules in GuHCl as a 1:1 mixture and allowed them to refold upon dilution from denaturant at 100 pM final concentration. No cross-correlation signal was observed during refolding (Fig. 2b). The sensitivity of the method was demonstrated using the double cysteine mutant DM-MBP(30C/312C) labeled with Atto532 and Atto647N [DM-MBP(DL)]. A strong cross-correlation signal was observed when 5 pM of the double-labeled protein, mimicking the presence of dimeric aggregates, was added to the 100 pM mixture of the single-labeled refolding proteins (Fig. 1b). These measurements show clearly that, at 100 pM, the labeled DM-MBP(312C) proteins are monomeric during refolding and do not form oligomers.

Analysis by fluorescence correlation spectroscopy (FCS) further confirmed the absence of reversible aggregates by demonstrating that the number of Atto647N-labeled DM-MBP(312C) particles in the observation volume remained constant over the refolding time (Fig. 2c). In contrast, if reversible aggregation was to limit the rate of spontaneous folding, the number of diffusing particles would be expected to increase as native monomeric protein is produced (Fig. 2c, simulated broken curve).

Spontaneous and GroEL/ES-assisted folding measured at single-molecule level

Having established conditions of spontaneous refolding in the absence of aggregation, we next developed a spFRET approach to measure the rates of spontaneous and GroEL/ES-assisted refolding at single-molecule level. Specifically, we tested the prediction of the passive cage model that, under conditions equivalent to infinite dilution, no rate acceleration by chaperonin would be observed [24]. As shown previously, DM-MBP(DL) in its native state and the unfolded protein when bound to GroEL have different FRET efficiency (f_E) distributions in single-molecule spFRET measurements [18]. The native protein shows a distribution of compact conformations with a peak at a high f_E of 0.72 (Fig. 3a). In contrast, GroEL-bound DM-MBP(DL) has ~40% molecules at a low f_E of 0.06, consistent with a highly expanded conformation, with the remainder of molecules populating a broad distribution of less expanded states around an intermediate f_E of 0.38 (Fig. 3b). To obtain the kinetics of spontaneous folding, we took advantage of the ability of GroEL to bind folding intermediates, but not the native protein, thereby stopping refolding and reverting not-yet folded DM-MBP(DL) molecules to the low f_E distribution (stretching; Fig. 1b). Assisted folding in the presence of GroEL/ES and ATP was stopped by the addition of apyrase, resulting in rapid hydrolysis of ATP and ADP to AMP. Quantification of the low and high f_E peak areas (f_E of 0.06 and 0.72, respectively) at different times enabled us to extract protein folding rates at a concentration of 100 pM DM-MBP(DL) (Fig. 3c and d). The rate of spontaneous refolding was ~5.6-fold slower than the assisted rate (Fig. 3e). To validate our findings, we measured the refolding rate of DM-MBP(DL) at a protein concentration of 100 nM. In these ensemble experiments we took advantage of the finding that, following initial collapse (with decrease in donor fluorescence due to FRET), the donor fluorescence of DM-MBP(DL) increases during folding, apparently due to changes in the chemical environment of the fluorophore. The observed rates were in good agreement with the single-molecule data and showed a ~7.7-fold acceleration of folding by chaperonin (Fig. 3f). Moreover, similar folding rates were previously measured for the unlabeled DM-MBP(30C/312C) by tryptophan fluorescence [18].

As an alternative single-molecule approach to measure folding rates, we utilized the difference in diffusion coefficients (D) of GroEL bound (~49 $\mu m^2\ s^{-1}$) and free DM-MBP (~160 $\mu m^2\ s^{-1}$) by FCS (Fig. 4a). We recorded the time-dependent change in the average diffusion rate during refolding of 100 pM DM-MBP(DL) using the Atto647N fluorescence signal (Fig. S1). Again, spontaneous folding was stopped by addition of excess GroEL and the assisted folding with apyrase. The folding rates obtained in these measurements (Fig. 4b) were in agreement with those obtained by spFRET (Fig. 3e).

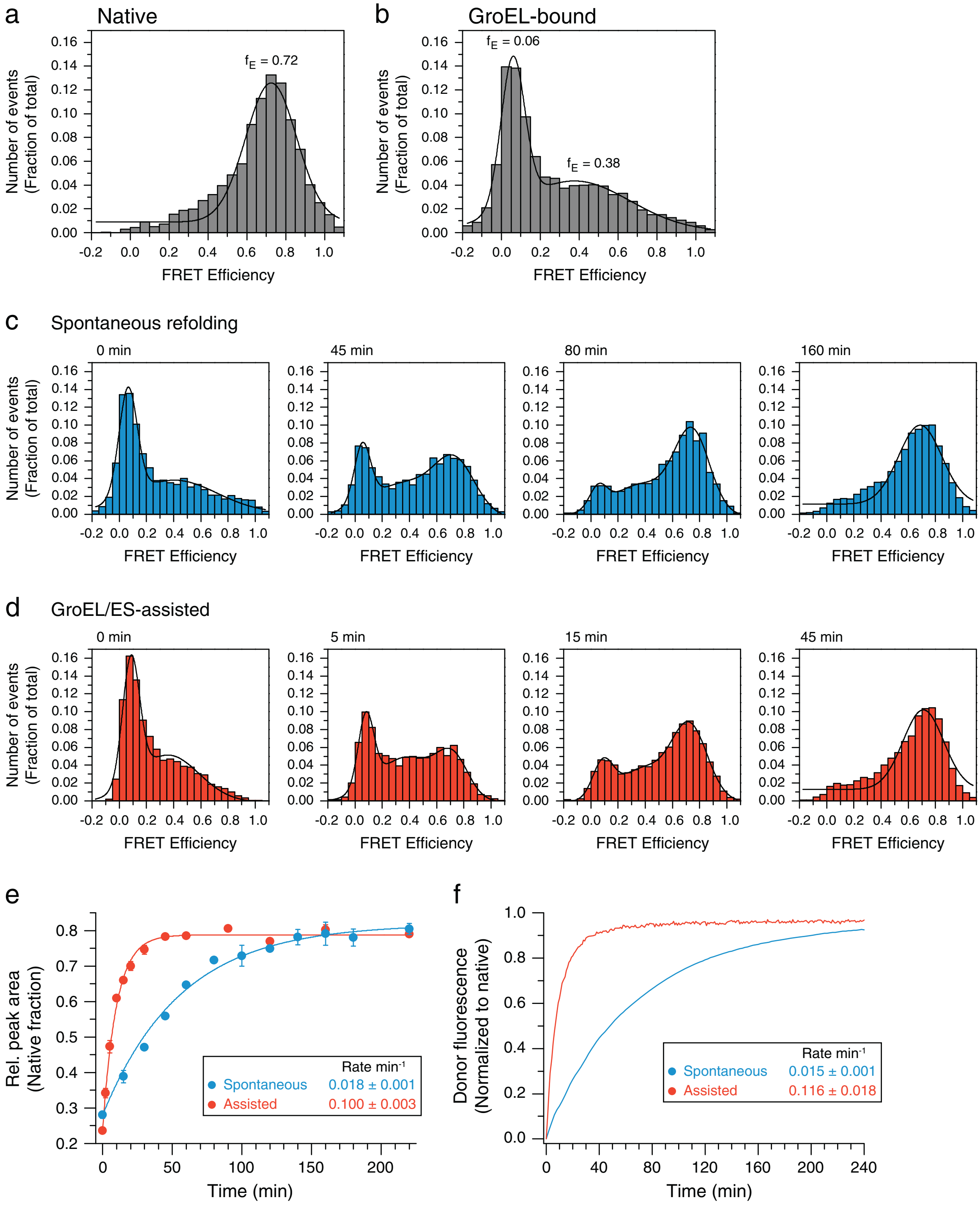
The previously reported effect of chloride salt to slow the spontaneous refolding of DM-MBP [22,26] was preserved under single-molecule conditions where aggregation is excluded (Fig. S2). Consequently, chloride salt does not decelerate spontaneous refolding by increasing aggregation [26] but by modulating the intrinsic folding properties of DM-MBP [22]. The electrostatic environment of the GroEL/ES cage apparently renders DM-MBP refolding salt insensitive [22].

PET-FCS as a measure of chain motion during folding

The active cage model of chaperonin action posits that encapsulation of non-native protein can reduce chain entropy, thereby accelerating folding kinetics [12,17,22]. Here we used fluorescence quenching via PET to test this hypothesis. In PET, the fluorescence of an oxazine dye (Atto655) is quenched via van der Waals contact with a Trp residue by direct transfer of an electron. Atto655 has the advantage of showing virtually no triplet blinking or other photophysical fluctuations [47,48] and thus is well suited to assess conformationally induced fluctuations at fast timescales. MBP contains 8 Trp residues spaced throughout the sequence (Fig. 2a) (note that GroEL and GroES do not contain Trp). The combination of PET with FCS serves as a powerful method to measure structural fluctuations in proteins at the single-molecule level on timescales from nanoseconds to milliseconds [48]. PET-FCS has been used to study denatured state dynamics and early events in protein folding [47].

The auto-correlation signal of Atto655-labeled DM-MBP(312C) [DM-MBP(Atto655)] was measured 30 s after dilution from denaturant, when essentially all DM-MBP populates a dynamic folding intermediate that converts only slowly to the native state [22]. The auto-correlation curve could not be fitted with a single component diffusion model due to the presence of a fast fluctuating component (Fig. 5a). It was fitted by a one-exponential one-diffusion equation (Fig. 5b), with the exponential term describing the PET amplitude F, which is proportional to the abundance of conformationally dynamic molecules, and the relaxation time (τ_R) providing a measure of chain motion (Fig. S3a). Based on these measurements, the slow folding intermediate of DM-MBP(Atto655) shows fast fluctuations of the fluorescence signal at a τ_R of 40 ± 3 μs, indicating high chain entropy. As a control, Atto655-labeled wild-type MBP(312C) showed no fast fluorescence fluctuation during folding (Fig. 5c), as this rapidly folding protein ($t_{1/2}$ ~ 23 s) [17] does not significantly populate the dynamic intermediate in which the dye can approach Trp residues. Similarly, when DM-MBP(Atto655) was allowed to refold to completion, the fluorescence fluctuations at the short correlation times were no longer detected (Fig. 5d). This is consistent with the fact that no Trp residue is in contact distance to residue 312 in the native state. Thus, fast fluorescence fluctuation is characteristic of the dynamic folding intermediate of DM-MBP.

The amplitude *F* of the PET-FCS signal (Fig. S3a), reflecting the fraction of conformationally dynamic molecules, decreased during spontaneous refolding (Fig. 5e) at a rate equivalent to that measured for folding of the unlabeled protein by Trp fluorescence



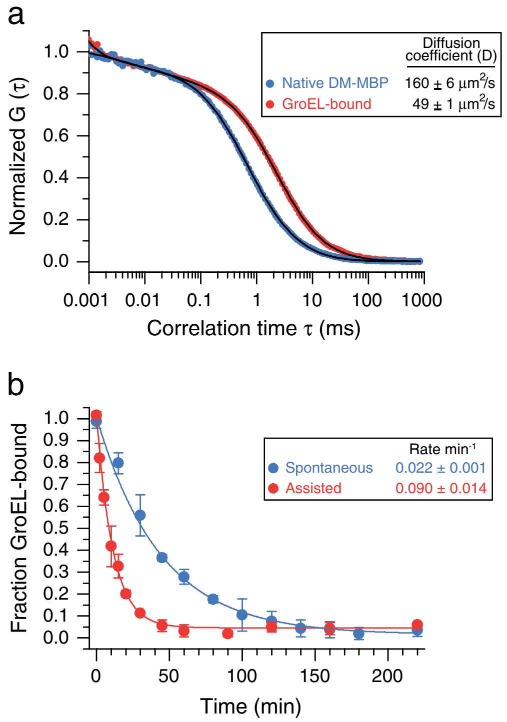


Fig. 4. Spontaneous and GroEL/ES-assisted refolding of DM-MBP measured by FCS. (a) Normalized fluorescence auto-correlation amplitudes *G* (τ) of Atto647N fluorescence for 100 pM GroEL bound and spontaneously refolded DM-MBP(DL) in buffer A. Diffusion coefficients (*D*) are indicated. ±SD from at least 3 independent measurements. (b) Refolding kinetics based on the difference in mean diffusion time through the confocal volume between GroEL-bound and free, native protein. Refolding was performed as in Fig. 3 and stopped either by adding GroEL (spontaneous refolding) or apyrase (assisted refolding), shifting not-yet folded DM-MBP to the slow diffusion time of the GroEL complex. Refolding is plotted as fraction of GroEL-bound protein and fitted to single exponential rates. ±SD from at least 3 independent measurements.

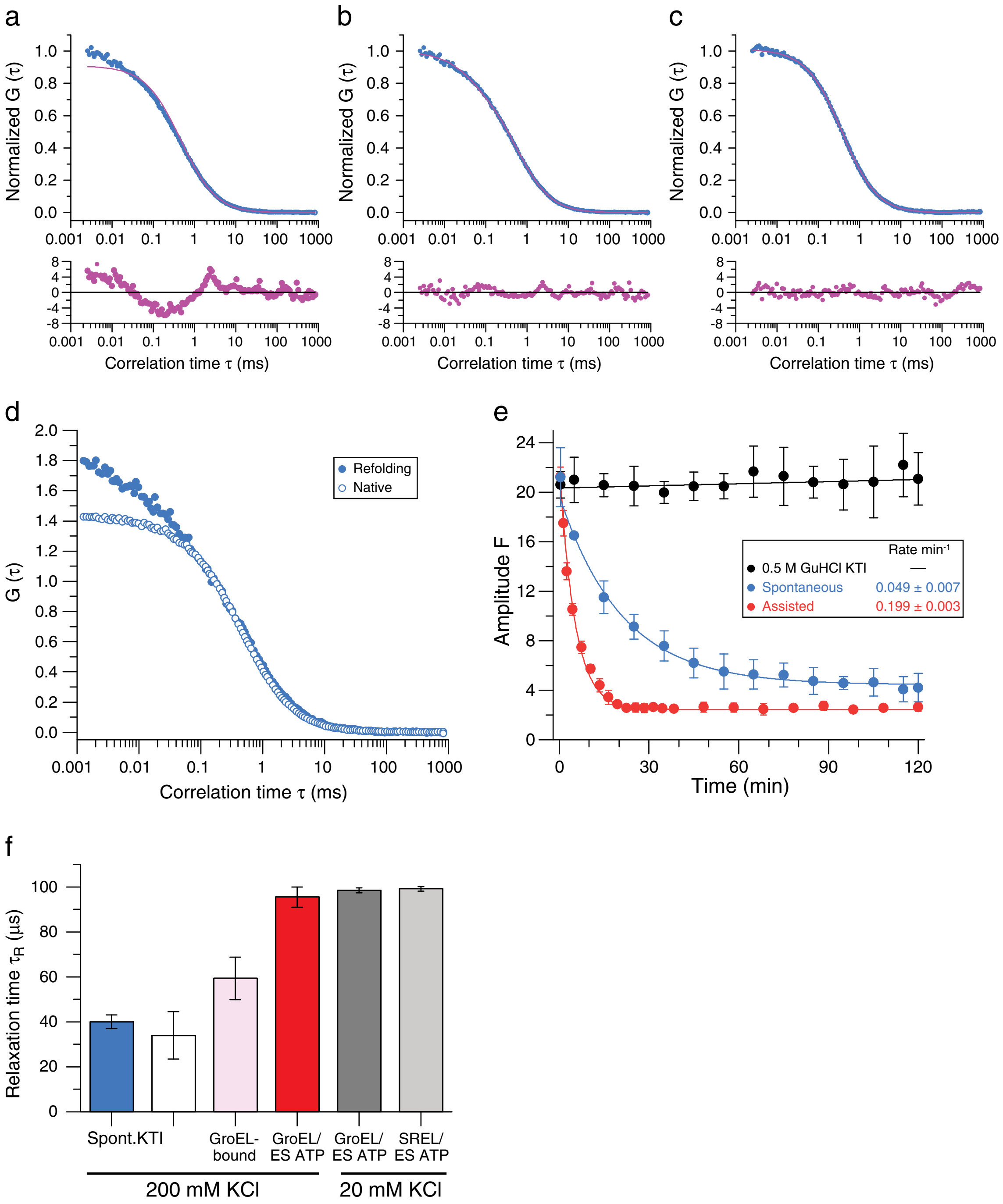
(Fig. S3b). This would be expected if refolding is limited by a kinetic energy barrier with a large entropic component. The rate of decrease in amplitude *F* was accelerated ~4-fold during GroEL/ES-assisted refolding (Fig. 5e), consistent with the chaperonin system reducing the entropic component of the energy barrier. In contrast, a constant high PET-FCS signal was observed when the labeled protein was diluted into buffer containing 0.5 M GuHCl, which stabilizes DM-MBP in its dynamic intermediate state [22] (Fig. 5e). The rate of spontaneous refolding of DM-MBP(Atto655) measured by PET-FCS was concentration independent over a concentration range of 4 orders of magnitude (100 pM to 1 μM) (Fig. S3c), further excluding aggregation as the cause of slow spontaneous refolding.

Evidence for conformational confinement in the GroEL/ES cage

While the amplitude of the PET-FCS signal is proportional to the concentration of dynamic particles, the τ_R of the PET signal is indicative of the kinetics of chain motion. During the first minute of spontaneous refolding, the τ_R of DM-MBP(Atto655) was 40 ± 3 μs, similar to the τ_R measured for the kinetically trapped intermediate in 0.5 M GuHCl (34 ± 10 μs) (Fig. 5f). Note that no τ_R could be measured for the native protein (Fig. 5d). The τ_R of the GroEL-bound protein was 59 ± 10 μs, indicating that the interaction of the unfolded DM-MBP with the apical GroEL domains reduces chain dynamics (Fig. 5f). Interestingly, the τ_R during the first minute of GroEL/ES-assisted refolding (≤20% molecules folded) was increased to 96 ± 5 μs (Fig. 5f), suggesting that the encapsulated protein is significantly restricted in chain motion, even when compared to the GroEL-bound state. To test this further, we measured chain mobility upon stable encapsulation of unfolded DM-MBP in the non-cycling SREL/ES complex (Fig. 5f). SREL is a single-ring variant of

Fig. 3. Spontaneous and GroEL/ES-assisted refolding of DM-MBP measured by spFRET. (a and b) FRET efficiency (f_E) histograms from spFRET measurements of native (a) and GroEL-bound (b) DM-MBP(DL) at 100 pM concentration (GroEL, 2 μM). (c) f_E histograms upon spontaneous refolding of 100 pM DM-MBP(DL). Refolding was performed in buffer A and stopped at the times indicated by addition of 2 μM GroEL, followed by spFRET analysis for 1 h. Peak values of a Gaussian fit to the f_E distributions are indicated. Representative histograms for the refolding times of 0, 45, 80 and 160 min are shown. (d) f_E histograms upon GroEL/ES-assisted refolding of 100 pM DM-MBP(DL). Unfolded protein was 200-fold diluted from 6 M GuHCl into buffer A containing 2 μM GroEL. Refolding was initiated by addition of 4 μM GroES/5 mM ATP and stopped by addition of apyrase at the times indicated. Representative histograms for the assisted refolding times of 0, 5, 15 and 45 min are shown. Histograms shown in (a–d) are representative of least three independent experiments. (e) Kinetics for spontaneous and assisted refolding were obtained from spFRET measurements as in (c) and (d) by plotting the time-dependent increase of the area of the high f_E peak, corresponding to native DM-MBP(DL). Data were fitted with a single exponential rate. Data represent averages ± SD from at least 3 independent measurements. (f) Refolding kinetics of 100 nM DM-MBP(DL) measured by conventional fluorescence spectroscopy in an ensemble approach. Refolding curves in buffer A were monitored over time at donor excitation and emission wavelengths of 532 nm and 550 nm, respectively, and are plotted as donor fluorescence relative to native. Averages ± SD from at least 3 independent measurements are shown.

GroEL that binds unfolded protein, ATP and GroES but ceases to hydrolyze ATP after a single round due to the absence of the allosteric signal from the GroEL trans-ring (see Fig. 1b) [8]. The SREL/ES complex is salt sensitive [9,49], and thus, the experiments with SREL were performed in low salt buffer using urea-denatured DM-MBP(Atto655). Stable substrate encapsulation was confirmed by size-exclusion chromatography (Fig. S4). Approximately 90–95% of DM-MBP co-fractionated with SREL/ES during folding, with the remainder eluting at a low molecular weight corresponding to free DM-MBP (Fig. S4a and b). Refolded DM-MBP was retained in the SREL/ES complex for more than 30 min (Fig. S4b, top panel) but was rapidly released when the SREL/ES complex was dissociated by addition of Mg chelator and high salt (Fig. S4b, bottom panel). The τ_R during the first minute of encapsulation in SREL/ES-assisted refolding was 99 ± 1 μs, identical with the value measured with the cycling GroEL/ES system under the same low



salt buffer condition (Fig. 5f). Thus, the environment of the chaperonin cage causes a considerable reduction of chain entropy, presumably promoting the conversion of dynamic folding intermediate to the native state.

Substrate protein spends most of its time in the encapsulated state during folding

The finding that, during GroEL/ES cycling, the substrate protein is conformationally restricted to the same extent as upon stable encapsulation in SREL/ES suggested that the vast majority of folding protein is in the encapsulated state. This would be consistent with recapture of non-native DM-MBP by GroEL occurring in less than 0.3 s at 25 °C [22]. Indeed the diffusion time of DM-MBP(DL) measured during the first minute of folding with GroEL/ES/ATP was virtually identical with that of the GroEL-bound protein and well discriminated from the fast diffusion time of spontaneously refolding DM-MBP(DL) (Fig. S5a). To quantify the relative amounts of GroEL--bound and encapsulated DM-MBP during the GroEL/ES reaction, we used single-molecule spFRET. DM-MBP(DL) showed similar FRET efficiency distributions when bound to GroEL or SREL with ~34–40% of molecules being in an expanded state (f_E of 0.06 and 0.1, respectively) and the remainder populating a broad distribution around an intermediate f_E of ~0.4 (Fig. 6a and b, left panels). During the first minute of encapsulation in SREL/ES, essentially all DM-MBP(DL) molecules populated a broad range of collapsed states around an f_E of 0.66 (Fig. 6b, right panel). In contrast, during the first minute of the GroEL/ES reaction, a bimodal f_E distribution was observed, with the low f_E peak representing GroEL-bound molecules and the high f_E population representing encapsulated and folded molecules (Fig. 6a, middle panel). The fraction of folded molecules was ~12%, as determined by addition of apyrase after 1 min to stop folding and revert the not-yet folded molecules to the bound state with low f_E value (Fig. 6a, right panel). Taking into consideration that the ~12% of folded DM-MBP(DL) were no longer GroEL associated, we calculated that ~18% of GroEL-associated molecules were bound and ~82% were encapsulated. Thus, during the cycling GroEL/ES reaction, the vast majority of substrate protein is in the encapsulated state during folding, in agreement with the PET-FCS measurements.

Next, we measured the duration of the GroEL/ES ATPase cycle as a function of substrate concentration. GroEL hydrolyzed ATP at a rate of ~53 ATP min^{-1} at 20 °C and GroES reduced the rate to ~21 ATP min^{-1} [50] (Fig. S5b). The presence of non-native substrate protein has been reported to stimulate the ATPase [29] by triggering ADP and GroES release from the GroEL trans-ring [38,51,52]. A single round of substrate binding and encapsulation occurs in the time it takes GroEL to hydrolyze 7 ATP (the hemi-cycle) in the presence of GroES and non-native substrate. Because spontaneous folding of DM-MBP is slow, we could measure the GroEL ATPase under conditions of substrate excess. At a concentration of 0.2 μM GroEL and 0.4 μM GroES, the initial ATPase rate reached a maximum of ~59 ATP per GroEL 14-mer min^{-1} at ~0.8 μM denatured DM-MBP and remained constant at higher DM-MBP concentrations (Fig. 6c and Fig. S4b). Thus, the duration of the GroEL/ES hemi-cycle at substrate

Fig. 5. Analysis of DM-MBP refolding and conformational dynamics by PET-FCS. (a–c) Normalized fluorescence auto-correlation amplitudes $G(\tau)$ for DM-MBP(Atto655) (a and b) and for Atto655-labeled wild-type MBP(312C) (c) during spontaneous refolding at final concentration of 1 nM in buffer A at 20 °C. Auto-correlation data collected during the first minute after 200-fold dilution of denatured protein into refolding buffer A (final MBP, 1 nM) were fitted with a one-diffusion model (a and c) or with a one-diffusion one-exponential model (b), which contains an exponential time constant (τ_R) with an amplitude F in addition to a diffusion component (τ_D and ρ) (Fig. S3a). Residuals of the fits are shown. Auto-correlation curves shown are representative of least three independent experiments. (d) PET-FCS curves for DM-MBP(Atto655) after completion of spontaneous refolding (open circles) and immediately upon dilution from denaturant into refolding buffer A (folding intermediate) (filled circles). Auto-correlation curves shown are representative of least three independent experiments. (e) Rates of spontaneous and GroEL/ES-assisted refolding of DM-MBP(Atto655) measured as the time-dependent decrease in PET-FCS amplitude (F) (Fig. S3a). FCS recording was started immediately upon initiation of refolding and was continued for 2 h. F was also analyzed upon dilution of denatured DM-MBP(Atto655) into buffer A containing 0.5 M GuHCl to stabilize the kinetically trapped intermediate (KTI). Time windows of 2 min for early time points and 10 min for late time points (GroEL/ES-assisted) or time windows of 10 min for early and late time points (spontaneous refolding and kinetically trapped intermediate) were correlated and fitted as in (b). Refolding rates were extracted by single exponential fits to plots of the amplitude of the exponential component F *versus* refolding time. Averages ± SD from at least 3 independent measurements are shown. (f) Conformational dynamics of DM-MBP is decreased inside the chaperonin cage. Relaxation time (τ_R) of the PET signal of DM-MBP(Atto655) upon 200-fold dilution from 6 M GuHCl into refolding buffer A (200 mM KCl) (spontaneous refolding; spont.) or into buffer A containing 0.5 M GuHCl to form KTI or into buffer A containing GroEL (2 μM) (GroEL-bound) or during the first minute of GroEL/ES-assisted refolding in buffer A. τ_R was also analyzed during the first minute of GroEL/ES- and SREL/ES-assisted refolding of 10 M urea-denatured protein in buffer B (20 mM KCl). FCS auto-correlation curves were fitted to a one-diffusion one-exponential model and τ_R values were extracted, which are inversely proportional to protein chain motion. Data represent averages ± SD from at least 3 independent measurements.

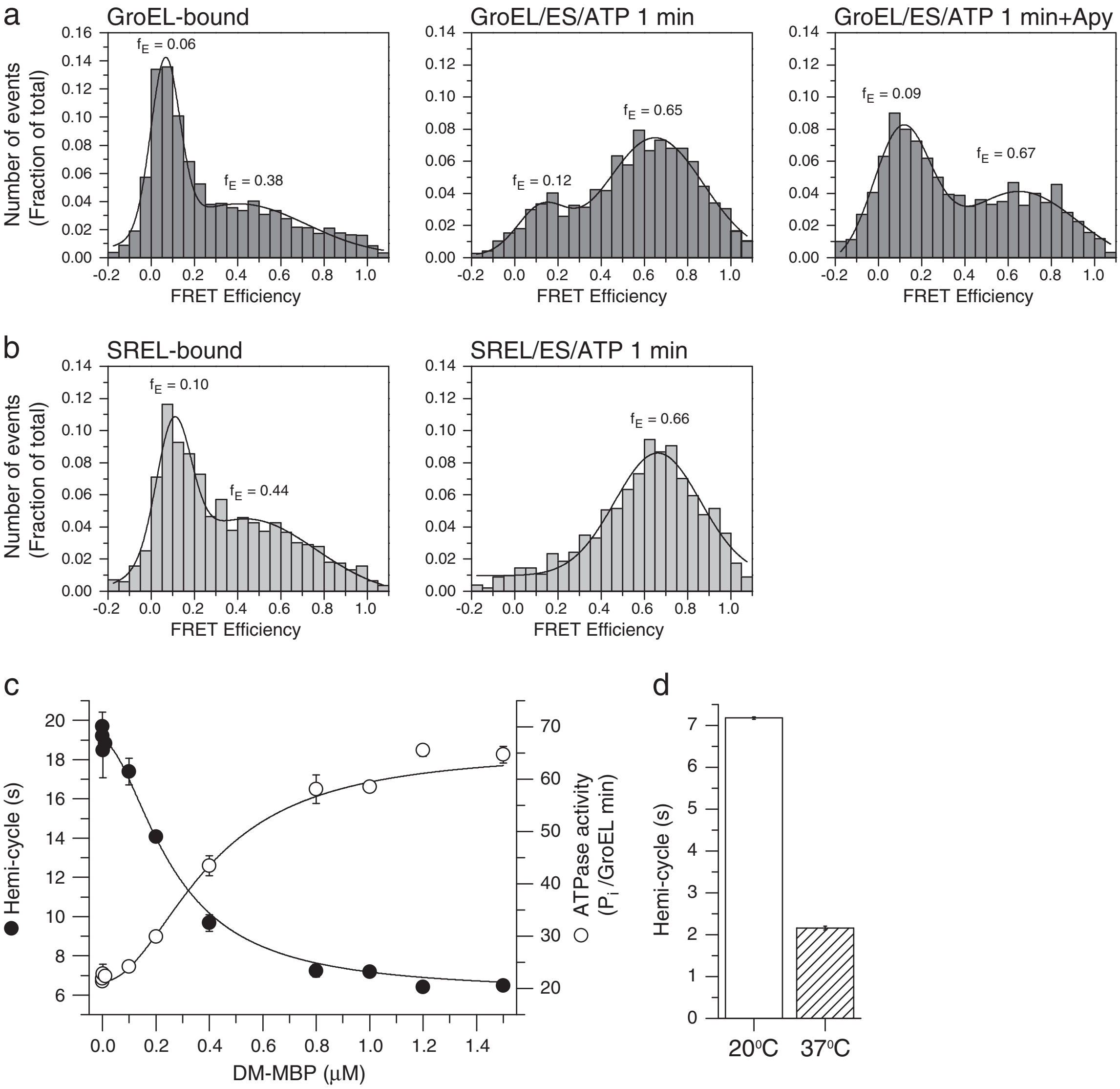


Fig. 6. Analysis of GroEL-bound and encapsulated DM-MBP during assisted refolding. (a) FRET efficiency (f_E) histograms from spFRET measurements of GroEL-bound DM-MBP(DL), as well as during the first minute of DM-MBP(DL) refolding and after stopping refolding at 1 min with apyrase (Apy). Measurements were performed as in Fig. 3a and d. Histograms representative of at least three independent experiments are shown. (b) f_E histograms as in (a) performed for SREL-bound DM-MBP(DL) and during the first minute of DM-MBP(DL) refolding while encapsulated in SREL/ES (buffer B at 20 °C). Peak values of a Gaussian fit to the f_E distributions are indicated. Histograms representative of at least three independent experiments are shown. (c) Measuring the GroEL ATPase (open circles) and the duration of the GroEL hemi-cycle (filled circles) in the presence of GroES and increasing concentrations of non-native DM-MBP. Denatured DM-MBP was diluted 200-fold to final concentrations from 0 to 1.5 μM into buffer A containing 0.2 μM GroEL and 0.4 μM GroES. ATPase activities were measured photometrically using a NADH coupled enzymatic assay at 20 °C (see Materials and Methods). ± SD from at least 3 independent measurements. (d) Duration of GroEL ATPase hemi-cycle as measured in (c) in the presence of 1 μM non-native DM-MBP at 20 °C and 37 °C. Data represent averages ± SD from at least 3 independent measurements.

saturation is ~7 s at 20 °C (Fig. 6c and d). As measured by stopped-flow mixing, binding of denatured DM-MBP to GroEL is complete after ~0.3 s and encapsulation upon addition of ATP and GroES is complete after ~0.5 s [18]. This would mean that the substrate spends ~1 s in the GroEL-bound state and ~6 s in the GroEL/ES cage, corresponding to ~14% and ~86% of hemi-cycle duration, respectively, in good agreement with the fraction of bound (~18%) and encapsulated substrate (~82%)

determined by spFRET mentioned above. The hemi-cycle duration in the presence of excess substrate at 37 °C was ~2.2 s (Fig. 6d). Thus, the reaction has a Q_{10} temperature coefficient of ~2, consistent with classical Arrhenius behavior and suggesting that all steps of the GroEL/ES mechanism undergo similar temperature-dependent rate acceleration.

Role of the net-negative charged GroEL cis-cavity wall

The wall of the GroEL cis-cavity has a net charge of minus 42 resulting from a cluster of exposed negatively charged residues (Glu252, Asp253, Glu255, Asp359, Asp361 and Glu363) [17]. Most of these are highly conserved among GroEL homologs but have no apparent function in binding of substrate or GroES. We have previously reported that mutating three of these residues (Asp359, Asp361 and Glu363) in SREL to lysines [SR(KKK2)], converting the cavity net charge to 0, impairs the ability of SR(KKK2)/ES to fold DM-MBP and bacterial ribulose bisphosphate carboxylase/oxygenase (RuBisCO), but not mitochondrial rhodanese [17]. Both DM-MBP and RuBisCO experience a significant rate acceleration of folding with wild-type SREL/ES [12,17], while rhodanese does not, suggesting that the negative charges have a role in enhancing folding kinetics. Here we analyzed this possibility using the SR(KKK2) and EL(KKK2) mutants.

Under standard conditions of physiological salt concentration, GroEL/ES accelerated the spontaneous folding of DM-MBP(312C) by ~4.5-fold. In contrast, no rate acceleration was observed with EL(KKK2)/ES (Fig. 7a). Remarkably, EL(KKK2)/ES did not restrict the chain dynamics of DM-MBP(Atto655) during folding as measured by PET-FCS, in striking contrast to GroEL/ES (Fig. 7a). Notably, DM-MBP(DL) when bound to EL(KKK2) displayed the same conformational properties as when bound to GroEL, as demonstrated by spFRET measurements (Fig. S5c). Moreover, during the first minute of folding with EL(KKK2)/ES/ATP, the diffusion time of DM-MBP(DL) was identical with that of the EL(KKK2)-bound protein (Fig. S5a), indicating that essentially all substrate was chaperonin associated. The fraction of bound and encapsulated substrate analyzed from spFRET histograms was ~16% and ~84%, respectively, close to the values obtained with GroEL/ES (Fig. S4c and Fig. 6a). The ATPase activity of EL(KKK2) was similar to that of GroEL and was inhibited by GroES. However, unlike GroEL, excess non-native DM-MBP had only a minor effect in stimulating the ATPase of EL(KKK2)/ES (Fig. S5b). These results suggested that the charge properties of the cis-cavity wall may have a dual role in entropically confining encapsulated substrate protein and in coupling the presence of substrate with the ATPase activity of GroEL.

To uncouple these two effects, we next used SR(KKK2) to analyze the chain dynamics of DM-MBP during folding when stably encapsulated. GroES-mediated substrate encapsulation by SR(KKK2) at low salt (Fig. S5d and e) was as efficient as with SREL (Fig. S4). When bound to SR(KKK2), DM-MBP(DL) displayed similar conformational properties as when bound to GroEL or SREL (Fig. S5c and f). During the first minute of encapsulation, DM-MBP(DL) populated compact conformations, as observed with SREL/ES (Fig. S5f). GroEL/ES and SREL/ES mediated the refolding of DM-MBP(312C) at essentially the same rate (measured at low salt) (Fig. 7b), and folding by

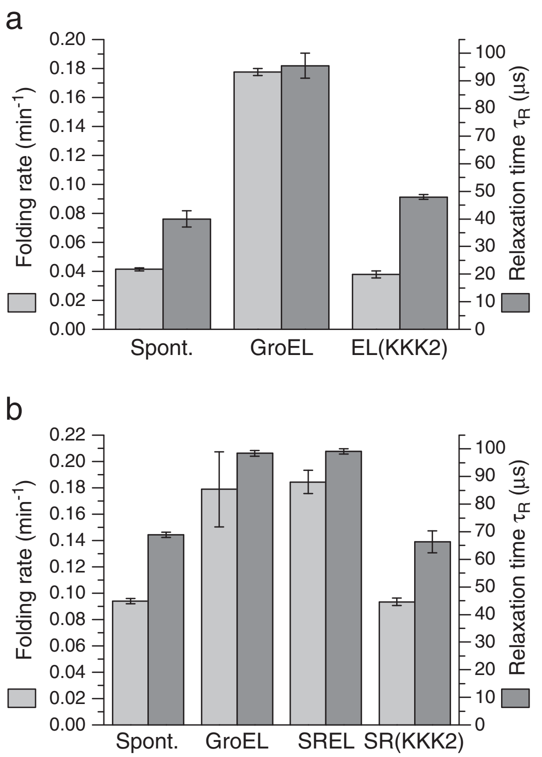


Fig. 7. Effect of net-negative charges in the chaperonin cis-cavity on DM-MBP folding and conformational dynamics. (a) Spontaneous (Spont.), GroEL/ES-assisted and EL(KKK2)/ES-assisted refolding of unlabeled DM-MBP(312C) in buffer A (200 mM KCl) were measured by monitoring the increase in Trp fluorescence at 20 °C as described in Fig. S3. Relaxation times (τ_R) of the PET signal of DM-MBP(Atto655) were measured during the first minute of spontaneous and assisted refolding. ±SD from at least 3 independent measurements. (b) Spontaneous and assisted refolding with GroEL/ES, SREL/ES and SR(KKK2)/ES were measured in buffer B (20 mM KCl) and τ_R values were also analyzed during the first minute of refolding as in (a). Data represent averages ± SD from at least 3 independent measurements.

SR(KKK2)/ES was not accelerated beyond the spontaneous rate (Fig. 7b). Interestingly, while the folding rates with GroEL/ES and SREL/ES are salt independent, the KKK2 mutant displays a similar salt dependence of folding rate as spontaneous renaturation (Fig. 7a and b and Fig. S2) [22]. Compared to high salt, the low salt condition resulted in reduced chain dynamics (slower τ_R) and ~2-fold faster folding for the spontaneous renaturation (Fig. 7a and b). Accordingly, in low salt, folding in SR(KKK2)/ES is ~2-fold faster than in EL(KKK2)/ES at high salt (Fig. 7a and b). Importantly, DM-MBP(Atto655) when inside SR(KKK2)/ES nevertheless displayed significantly higher chain dynamics (τ_R of 66 ± 4 μs) compared to SREL/ES (τ_R of 99 ± 1 μs) (Fig. 7b). Together these findings indicate that the net-negative charge of the cis-cavity plays a critical role in entropically confining dynamic folding intermediates of encapsulated substrate, thereby accelerating their conversion to the native state.

Discussion

Active *versus* passive GroEL/ES function

It has been argued that accelerated folding of DM-MBP by GroEL/ES [17,19,22] is due to the ability of chaperonin to prevent reversible aggregation that would otherwise reduce the rate, but not the yield, of spontaneous folding [24,26]. The basic assumption of this passive cage model is that GroEL/ES functions solely as an anti-aggregation device and that folding inside the GroEL/ES cage occurs at the same rate as it would during spontaneous folding at infinite dilution [25]. Here we have tested this idea by monitoring folding by single-molecule experiments at 100 pM DM-MBP. Since the probability of more than one molecule of DM-MBP residing in the confocal observation volume is ≤1% at this concentration, we were able to compare the rates of GroEL/ES-assisted folding and spontaneous folding at de facto infinite dilution. The absence of aggregation during spontaneous folding was confirmed by FCS and dcFCCS measurements. Using GroEL to sort non-native from native molecules, we showed by spFRET and FCS that GroEL/ES accelerates fluorescent-labeled DM-MBP refolding 4- to 8-fold, demonstrating the active chaperonin mechanism.

Active cage model *versus* iterative annealing

According to the iterative annealing model, substrate binding and release during ATP-dependent GroEL/ES cycling serve to unfold kinetically trapped folding intermediates, affording them a chance at productive folding either inside or outside the GroEL/ES cage [27,44]. Using SREL/ES, we demonstrated that a single round of encapsulation inside the chaperonin cage is sufficient to achieve accelerated DM-MBP folding at full yield. This argues for encapsulation being the active principle. However, an active cage may function synergistically with iterative annealing for certain substrates, for example, when a fraction of molecules were to form long-lived, misfolded states during assisted folding. We also note that the SREL experiments do not rule out the possible contribution of a single round of unfolding to subsequent accelerated folding inside the cage. A conformational expansion of substrate protein upon binding to GroEL and additional stretching upon ATP-mediated apical domain movement has been observed [18,34,35] (as well as this study) (Fig. 1b). Furthermore, during GroES-mediated encapsulation, the protein is released in a step-wise fashion, with the less hydrophobic and thus less tightly bound, sequence elements dissociating before the more hydrophobic ones [18]. This mode of release would modulate the mechanism of hydrophobic collapse and may contribute to avoiding the formation of kinetically trapped folding intermediates.

The iterative annealing model assigns no specific function to the chaperonin cage in aggregation prevention or accelerating folding and accordingly suggests that folding may equally occur inside the GroEL/ES cage or outside [27] (Fig. 1b). Out of cage-folding would require that not-yet folded protein spends significant time in free solution during GroEL/ES cycling. This is inconsistent with our PET-FCS and spFRET measurements, which demonstrate that, in the cycling reaction, non-native DM-MBP spends ~80% of the duration of the GroEL/ES hemi-cycle (7 ATP hydrolyzed; ~7 s at 20 °C) in the encapsulated state and the remainder bound to the apical GroEL domains. Considering that GroEL recaptures non-native DM-MBP from solution within 0.3 s or less (at 25 °C) [18], the fraction of out of cage folding would be insignificant (<5%). Moreover, even vastly increasing the concentration of GroEL/ES relative to substrate, as in our single-molecule experiments (100 pM DM-MBP, 2 μM GroEL/4 μM GroES), did not slow folding kinetics or yield, which can only be explained by in cage folding. It has also been argued that, at 37 °C, the dwell time of substrate inside the GroEL/ES cage becomes too short for efficient folding [27]. We measured the duration of the GroEL/ES hemi-cycle under substrate saturation at 37 °C to be 2.2 s. Making the reasonable assumption that all the steps of the GroEL/ES reaction undergo temperature-dependent acceleration, the vast majority of folding would nevertheless occur in cage. Indeed, authentic GroEL-dependent proteins are typically highly aggregation prone in a temperature-dependent manner [16,53,54]. Under so-called non-permissive conditions, where irreversible aggregation prohibits spontaneous folding, the GroEL/ES-assisted folding of these proteins stops immediately when rapid recapture of folding intermediate by GroEL is prevented [12], indicating that folding must occur in cage.

Function of the GroEL/ES cage

Protein encapsulation by chaperonin has been proposed to serve a dual purpose: it prevents aggregation and accelerates folding for certain proteins, adjusting folding speed relative to the rate of protein synthesis and thereby preventing the accumulation of unfolded or misfolded protein molecules [12]. Rate enhancement of folding was attributed to an effect of steric confinement, entropically destabilizing dynamic folding intermediates and thereby facilitating their conversion to the native state [17,19,22]. Two physical properties of the GroEL/ES cage may be critical in this regard: the volume of the cage relative to the size of the folding protein (i.e., steric confinement proper [17,19,41,42,55]) and the negative net charge of the cavity wall [17,19] that has been suggested to increase the hydrophobic effect by ordering water structure [56]. In the present study, we used PET-FCS to directly measure the effect of encapsulation on polypeptide chain dynamics. While binding of unfolded protein to GroEL restricts chain dynamics only moderately, we find that encapsulation results in a marked restriction of flexibility, as reflected in an increase of the τ_R of the DM-MBP folding intermediate from 40 ± 3 μs during spontaneous refolding to 99 ± 1 μs when encapsulated. Notably, τ_R of the folding intermediate was very similar when measured during active GroEL/ES cycling or upon stable encapsulation in SREL/ES. This provides direct evidence that, during folding, the protein spends the vast majority of the time in the encapsulated state. Contrary to a recent report [57], our experiments with SREL/ES did not reveal a functionally significant "escape" of DM-MBP from the non-cycling SREL/ES cage.

Using EL(KKK2) or SR(KKK2), a mutant in which the negative net charge of the cavity wall of 42 is reduced to 0, we analyzed the effect of cavity charge on substrate chain dynamics by PET-FCS. Strikingly, these measurements showed that the KKK2 cage is unable to restrict the conformational motion of encapsulated protein relative to the folding intermediate in free solution. This correlates with a loss of function of the KKK2 mutant in accelerating DM-MBP folding. Thus, in essence, removal of the negative net charge converts the active GroEL/ES cage to a passive cage. Moreover, we found that the negatively charged residues exposed in the cis-cavity are required for the stimulation of the GroEL ATPase by substrate protein. It thus appears that the cage senses the encapsulated protein. In summary, we suggest that steric confinement and the negatively charged cavity wall function cooperatively in promoting folding.

Materials and Methods

Strains, plasmids and proteins

The *E. coli* strains DH5α and BL21 (DE3) Gold (Stratagene) were used for cloning and protein expression, respectively. DM-MBP (MBP V8G, Y283D) and its cysteine variants, DM-MBP(312C) and DM-MBP(30C/312C), were constructed in pCH vectors (T7 promoter) by site-directed mutagenesis using QuikChange (Stratagene) and were expressed, purified and labeled with fluorophore [17,18] (see Supplementary Methods for details).

GroEL, GroES, SREL, EL(KKK2) and SR(KKK2) were expressed and purified as previously described [17,19]. GroEL preparations were subjected to rigorous quality control by the following assays: ATPase activity in presence and absence of GroES [19], rhodanese aggregation prevention [58] and DM-MBP refolding [17]. GroES preparations were controlled for efficient inhibition of GroEL ATPase activity and ability to accelerate DM-MBP refolding.

DM-MBP refolding

DM-MBP was unfolded in 6 M GuHCl/10 mM DTT or in 10 M urea/10 mM DTT for 1 h at 50 °C, as indicated in the figure legends. Spontaneous refolding was initiated upon 200-fold dilution into refolding buffer A [20 mM Tris–HCl (pH 7.5), 200 mM KCl and 5 mM $Mg(C_2H_3O_2)_2$] or refolding buffer B [50 mM Hepes–NaOH (pH 7.5), 20 mM KCl and 10 mM $MgCl_2$] as indicated in the figure legends. For assisted refolding, denatured DM-MBP was diluted 200-fold into refolding buffer containing either 2 μM GroEL or 1 μM SREL and refolding was initiated upon addition of 4 μM GroES and 5 mM ATP.

In ensemble experiments (100 nM DM-MBP), folding was recorded by the time-dependent increase in the intrinsic Trp fluorescence (excitation: 295 nm; emission: 345 nm) of DM-MBP and its variants on a Fluorolog F3-22 spectrofluorometer (Horiba), equipped with Peletier thermostat set to 20 °C (GroEL and GroES do not contain Trp) [17]. Photobleaching was carefully avoided by limiting the excitation slit width to 2 nm, with the emission slit width being set to 8 nm. Fluorescence signal was excited and emission was collected every 30 s for a 100-ms window using an automated shutter. Note that, at low concentrations of DM-MBP, bleaching of Trp fluorescence may result in measuring seemingly faster folding rates than at higher concentrations.

For DM-MBP(DL), refolding could not be measured by Trp fluorescence due to quenching. However, fluorescence of the donor dye in the folding intermediate was significantly lower than in the native state of DM-MBP(DL). The time-dependent increase in donor fluorescence was used as a measure of DM-MBP(DL) folding in ensemble measurements at 100 nM (excitation wavelength, 532 nm; emission wavelength, 550 nm).

Analysis of protein encapsulation

Encapsulation experiments were performed in buffer B. DM-MBP(Atto655) was unfolded in 10 M urea/10 mM DTT for 1 h at 50 °C. The denatured protein was diluted 200-fold (30 nM final) into refolding buffer containing 1 μM SREL or SR(KKK2). The reaction was incubated for 5 min at room temperature. Refolding was started by addition of 3 μM GroES and 1 mM ATP at 20 °C. The reaction was separated on a Superdex 200 PC3.2/30 gel-filtration column (Amersham Biosciences), equilibrated in buffer B/50 mM urea/1 mM ATP, either immediately or after 30–60 min of incubation at 20 °C or after 30–60 min of incubation at 20 °C and dissociation of the SREL/ES complex by the addition of 50 mM CDTA/70 mM GuHCl/200 mM KCl. Fractions were collected and analyzed by 15% SDS-PAGE, Coomassie staining and fluorescence imaging (FujiFilm FLA3000) and were quantified by densitometry.

ATPase assay

The ATPase activity of 0.2 μM GroEL or EL(KKK2) was measured in buffer A at 20 °C in the absence or in the presence of 0.4 μM GroES or 0.4 μM GroES with increasing concentration of denatured DM-MBP (diluted 200-fold from 6 M GuHCl). Control reactions received equivalent amounts of GuHCl (30 mM final). The hydrolysis of ATP to ADP was followed photometrically using an NADH coupled enzymatic assay (2 mM phosphoenolpyruvate, 30/20 U ml^{-1} pyruvate kinase/lactate dehydrogenase, 0.5 mM NADH and 1 mM ATP) at 20 °C in a spectrophotometer (Jasco) essentially as previously described [59].

Single-molecule spectroscopy

Single-molecule spectroscopy was performed on a Microtime 200 inverse time-resolved fluorescence microscope (PicoQuant) using pulsed interleaved excitation [60] (see Supplementary Methods for details). The instrument was maintained at a constant temperature of 20 °C. dcFCCS and FCS were used to investigate the oligomeric state of DM-MBP during refolding [18,22], while spFRET and FCS were used to assess folding rates of DM-MBP at 100 pM. The significant size difference of folded DM-MBP monomer (~41 kDa) and non-native DM-MBP in complex with GroEL (~830 kDa) results in different diffusion rates (106 ± 6 μm^2 s^{-1} and 49 ± 1 μm^2 s^{-1}, respectively) that can be monitored using FCS. The auto-correlation data were fitted with the following one triplet one-diffusion equation using the Symphotime software (PicoQuant):

$$G(\tau) = \left[1 - T + T \times e^{\left(-\frac{\tau}{\tau_T}\right)}\right] \times \left[\rho \times \left(1 + \frac{\tau}{\tau_D}\right)^{-1} \times \left(1 + \frac{\tau}{\tau_D \times \kappa^2}\right)^{-1/2}\right].$$

The mean diffusion time τ_D of particles through the focal spot is described by the structural parameter $\kappa = z_0/\omega_0$ where z_0 and ω_0 denote the axial and radial dimensions of the confocal volume, respectively. The amplitude of the correlation function is denoted by ρ. The first term is used to compensate for fast dynamics arising from dye photophysics such as triplet blinking with the amplitude T on the timescale τ_T [61]. Diffusion coefficients were calculated using the following equation

$$D = \frac{\left(V_{\text{eff}} \times \pi^{-3/2} \times \kappa^{-1}\right)^{2/3}}{4 \times \tau_D}$$

by calibrating the confocal volume V_{eff} with Atto655 dye, for which accurate diffusion parameters have been published [62]. To analyze refolding kinetics, we plotted the mean diffusion time, reflecting a shift of DM-MBP molecules from GroEL bound to free against the refolding time and fitted it with a single exponential rate.

PET-FCS was used as an approach to assess conformational dynamics of refolding DM-MBP molecules [47,48]. DM-MBP(Atto655) (or Atto655-labeled wild-type MBP) was unfolded in 6 M GuHCl/10 mM DTT for 1 h at 20 °C or 10 M urea/10 mM DTT for 1 h at 50 °C. Refolding was started by 200-fold dilution of the protein (final 100 pM to 1 μM) into refolding buffer A or buffer B as indicated in the figure legends. FCS measurements were performed immediately. In order to resolve fast dynamics in the microsecond timescale, we recorded fluorescence on two detectors simultaneously. Cross-correlation of the signals allowed removal of detector after pulsing. The correlated data were fitted with either a single component diffusion model

$$G(\tau) = \rho \times \left(1 + \frac{\tau}{\tau_D}\right)^{-1} \times \left(1 + \frac{\tau}{\tau_D \times \kappa^2}\right)^{-1/2}$$

or the following one-exponential one-diffusion equation, with the exponential term describing amplitude F and relaxation time τ_R of PET

$$G(\tau) = \left[1 - F + F \times e^{\left(-\frac{\tau}{\tau_R}\right)}\right] \times \left[\rho \times \left(1 + \frac{\tau}{\tau_D}\right)^{-1} \times \left(1 + \frac{\tau}{\tau_D \times \kappa^2}\right)^{-1/2}\right]$$

in Symphotime (PicoQuant). For relaxation time extraction, only the first or the last 30 s of a 2-h refolding experiment was considered. For folding rate extraction, the measurement was subdivided into 2 min time windows and extracted values for F were plotted against refolding time. These data were fitted with a single exponential function in Origin (OriginLabs) to give refolding rates.

Acknowledgements

We gratefully acknowledge the expert technical assistance by N. Wischnewski and A. R. Lange.

Author Contributions: A.J.G. and S.H. designed and performed the experiments and wrote the first draft of the manuscript; G.M. performed the analyses of protein encapsulation. F.U.H. and M.H.H. advised and supervised the project and wrote the paper.

Conflict of Interest Statement: The authors declare that they have no conflict of interest.

Appendix A. Supplementary data

Supplementary data include Supplementary Methods and five figures. Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jmb.2014.04.018.

Received 22 February 2014;
Received in revised form 16 April 2014;
Accepted 21 April 2014
Available online 6 May 2014

Keywords:
chaperonins;
GroEL;
GroES;
DM-MBP;
single-molecule spectroscopy

†A.J.G. and S.H. contributed equally to this work, and they share first authorship.

Abbreviations used:
DM-MBP, double-mutant maltose binding protein; spFRET, single-pair fluorescence resonance energy transfer; dcFCCS, dual-color fluorescence cross-correlation spectroscopy; FCS, fluorescence correlation spectroscopy; PET, photoinduced electron transfer.

References

[1] Frydman J. Folding of newly translated proteins *in vivo*: the role of molecular chaperones. Annu Rev Biochem 2001;70:603–47.
[2] Fenton WA, Horwich AL. Chaperonin-mediated protein folding: fate of substrate polypeptide. Q Rev Biophys 2003;36:229–56.
[3] Gershenson A, Gierasch LM. Protein folding in the cell: challenges and progress. Curr Opin Struct Biol 2011;21:32–41.
[4] Saibil HR, Fenton WA, Clare DK, Horwich AL. Structure and allostery of the chaperonin GroEL. J Mol Biol 2013;425:1476–87.
[5] Kim YE, Hipp MS, Bracher A, Hayer-Hartl M, Hartl FU. Molecular chaperone functions in protein folding and proteostasis. Annu Rev Biochem 2013;82:323–55.
[6] Vitlin Gruber A, Nisemblat S, Azem A, Weiss C. The complexity of chloroplast chaperonins. Trends Plant Sci 2013;18:688–94.
[7] Mayhew M, Da Silva ACR, Martin J, Erdjument-Bromage H, Tempst P, Hartl FU. Protein folding in the central cavity of the GroEL–GroES chaperonin complex. Nature 1996;379:420–6.
[8] Weissman JS, Rye HS, Fenton WA, Beechem JM, Horwich AL. Characterization of the active intermediate of a GroEL-GroES-mediated protein folding reaction. Cell 1996;84:481–90.
[9] Hayer-Hartl MK, Weber F, Hartl FU. Mechanism of chaperonin action: GroES binding and release can drive GroEL-mediated protein folding in the absence of ATP hydrolysis. EMBO J 1996;15:6111–21.
[10] Xu ZH, Horwich AL, Sigler PB. The crystal structure of the asymmetric GroEL-GroES-(ADP)7 chaperonin complex. Nature 1997;388:741–9.
[11] Rye HS, Burston SG, Fenton WA, Beechem JM, Xu Z, Sigler PB, et al. Distinct actions of cis and trans ATP within the double ring of the chaperonin GroEL. Nature 1997;388:792–8.
[12] Brinker A, Pfeifer G, Kerner MJ, Naylor DJ, Hartl FU, Hayer-Hartl M. Dual function of protein confinement in chaperonin-assisted protein folding. Cell 2001;107:223–33.
[13] Wang JD, Herman C, Tipton KA, Gross CA, Weissman JS. Directed evolution of substrate-optimized GroEL/S chaperonins. Cell 2002;111:1027–39.
[14] Meyer AS, Gillespie JR, Walther D, Millet IS, Doniach S, Frydman J. Closing the folding chamber of the eukaryotic chaperonin requires the transition state of ATP hydrolysis. Cell 2003;113:369–81.
[15] Figueiredo L, Klunker D, Ang D, Naylor DJ, Kerner MJ, Georgopoulos C, et al. Functional characterization of an archaeal GroEL/GroES chaperonin system—significance of substrate encapsulation. J Biol Chem 2004;279:1090–9.
[16] Kerner MJ, Naylor DJ, Ishihama Y, Maier T, Chang HC, Stines AP, et al. Proteome-wide analysis of chaperonin-dependent protein folding in *Escherichia coli*. Cell 2005;122:209–20.
[17] Tang YC, Chang HC, Roeben A, Wischnewski D, Wischnewski N, Kerner MJ, et al. Structural features of the GroEL-GroES nano-cage required for rapid folding of encapsulated protein. Cell 2006;125:903–14.
[18] Sharma S, Chakraborty K, Mueller BK, Astola N, Tang YC, Lamb DC, et al. Monitoring protein conformation along the pathway of chaperonin-assisted folding. Cell 2008;133:142–53.
[19] Tang YC, Chang HC, Chakraborty K, Hartl FU, Hayer-Hartl M. Essential role of the chaperonin folding compartment *in vivo*. EMBO J 2008;27:1458–68.
[20] Clare DK, Bakkes PJ, van Heerikhuizen H, van der Vies SM, Saibil HR. Chaperonin complex with a newly folded protein encapsulated in the folding chamber. Nature 2009;457:107–13.
[21] Zhang J, Baker ML, Schroder GF, Douglas NR, Reissmann S, Jakana J, et al. Mechanism of folding chamber closure in a group II chaperonin. Nature 2010;463:379–83.
[22] Chakraborty K, Chatila M, Sinha J, Shi Q, Poschner BC, Sikor M, et al. Chaperonin-catalyzed rescue of kinetically trapped states in protein folding. Cell 2010;142:112–22.
[23] Chen D-H, Madan D, Weaver J, Lin Z, Schroder GF, Chiu W, et al. Visualizing GroEL/ES in the act of encapsulating a folding protein. Cell 2013;153:1354–65.
[24] Apetri AC, Horwich AL. Chaperonin chamber accelerates protein folding through passive action of preventing aggregation. Proc Natl Acad Sci USA 2008;105:17351–5.
[25] Horwich AL, Apetri AC, Fenton WA. The GroEL/GroES cis cavity as a passive anti-aggregation device. FEBS Lett 2009;583:2654–62.
[26] Tyagi NK, Fenton WA, Deniz AA, Horwich AL. Double mutant MBP refolds at same rate in free solution as inside the GroEL/GroES chaperonin chamber when aggregation in free solution is prevented. FEBS Lett 2011;585:1969–72.
[27] Yang D, Ye X, Lorimer GH. Symmetric GroEL:GroES2 complexes are the protein-folding functional form of the chaperonin nanomachine. Proc Natl Acad Sci USA 2013;110:E4298–305.
[28] Braig K, Otwinowski Z, Hegde R, Boisvert DC, Joachimiak A, Horwich AL, et al. The crystal structure of the bacterial chaperonin GroEL at 2.8 Å. Nature 1994;371:578–86.

[29] Martin J, Langer T, Boteva R, Schramel A, Horwich AL, Hartl FU. Chaperonin-mediated protein folding at the surface of GroEL through a “molten globule”-like intermediate. Nature 1991;352:36–42.

[30] Fenton WA, Kashi Y, Furtak K, Horwich AL. Residues in chaperonin GroEL required for polypeptide binding and release. Nature 1994;371:614–9.

[31] Hayer-Hartl MK, Ewbank JJ, Creighton TE, Hartl FU. Conformational specificity of the chaperonin GroEL for the compact folding intermediates of alpha-lactalbumin. EMBO J 1994;13:3192–202.

[32] Horovitz A, Willison KR. Allosteric regulation of chaperonins. Curr Opin Struct Biol 2005;15:646–51.

[33] Lin Z, Rye HS. Expansion and compression of a protein folding intermediate by GroEL. Mol Cell 2004;16:23–34.

[34] Lin Z, Madan D, Rye HS. GroEL stimulates protein folding through forced unfolding. Nat Struct Mol Biol 2008;15:303–11.

[35] Kim SY, Miller EJ, Frydman J, Moerner WE. Action of the chaperonin GroEL/ES on a non-native substrate observed with single-molecule FRET. J Mol Biol 2010;401:553–63.

[36] Lin Z, Puchalla J, Shoup D, Rye HS. Repetitive protein unfolding by the trans ring of the GroEL-GroES chaperonin complex stimulates folding. J Biol Chem 2013;288:30944–55.

[37] Takei Y, Iizuka R, Ueno T, Funatsu T. Single-molecule observation of protein folding in symmetric GroEL-(GroES)2 complexes. J Biol Chem 2012;287:41118–25.

[38] Ye X, Lorimer GH. Substrate protein switches groE chaperonins from asymmetric to symmetric cycling by catalyzing nucleotide exchange. Proc Natl Acad Sci USA 2013;110:E4289–97.

[39] Ellis RJ. Molecular chaperones. Opening and closing the Anfinsen cage. Curr Biol 1994;4:633–5.

[40] Ellis RJ, Hartl FU. Protein folding in the cell: competing models of chaperonin function. FASEB J 1996;10:20–6.

[41] Baumketner A, Jewett A, Shea JE. Effects of confinement in chaperonin assisted protein folding: rate enhancement by decreasing the roughness of the folding energy landscape. J Mol Biol 2003;332:701–13.

[42] Hayer-Hartl M, Minton AP. A simple semiempirical model for the effect of molecular confinement upon the rate of protein folding. Biochemistry 2006;45:13356–60.

[43] Lucent D, Vishal V, Pande VS. Protein folding under confinement: a role for solvent. Proc Natl Acad Sci USA 2007;104:10430–4.

[44] Thirumalai D, Lorimer GH. Chaperonin-mediated protein folding. Annu Rev Biophys Biomol Struct 2001;30:245–69.

[45] Sparrer H, Lilie H, Buchner J. Dynamics of the GroEL protein complex: effects of nucleotides and folding mutants. J Mol Biol 1996;258:74–87.

[46] Chun SY, Strobel S, Bassford P, Randall LL. Folding of maltose-binding protein. Evidence for the identity of the rate-determining step *in vivo* and *in vitro*. J Biol Chem 1993;268:20855–62.

[47] Neuweiler H, Johnson CM, Fersht AR. Direct observation of ultrafast folding and denatured state dynamics in single protein molecules. Proc Natl Acad Sci USA 2009;106:18575–80.

[48] Sauer M, Neuweiler H. PET-FCS: probing rapid structural fluctuations of proteins and nucleic acids by single-molecule fluorescence quenching. Methods Mol Biol 2014;1076:597–615.

[49] Motojima F, Motojima-Miyazaki Y, Yoshida M. Revisiting the contribution of negative charges on the chaperonin cage wall to the acceleration of protein folding. Proc Natl Acad Sci USA 2012;109:15740–5.

[50] Chandrasekhar GN, Tilly K, Woolford C, Hendrix R, Georgopoulos C. Purification and properties of the GroES morphogenetic protein of *Escherichia coli*. J Biol Chem 1986;261:12414–9.

[51] Martin J, Mayhew M, Langer T, Hartl FU. The reaction cycle of GroEL and GroES in chaperonin-assisted protein folding. Nature 1993;366:228–33.

[52] Hayer-Hartl MK, Martin J, Hartl FU. Asymmetrical interaction of GroEL and GroES in the ATPase cycle of assisted protein folding. Science 1995;269:836–41.

[53] Fujiwara K, Ishihama Y, Nakahigashi K, Soga T, Taguchi H. A systematic survey of *in vivo* obligate chaperonin-dependent substrates. EMBO J 2010;29:1552–64.

[54] Calloni G, Chen T, Schermann SM, Chang H-C, Genevaux P, Agostini F, et al. DnaK functions as a central hub in the *E. coli* chaperone network. Cell Rep 2012;1:251–64.

[55] Sirur A, Best RB. Effects of interactions with the GroEL cavity on protein folding rates. Biophys J 2013;104:1098–106.

[56] England JL, Lucent D, Pande VS. A role for confined water in chaperonin function. J Am Chem Soc 2008;130:11838–9.

[57] Motojima F, Yoshida M. Polypeptide in the chaperonin cage partly protrudes out and then folds inside or escapes outside. EMBO J 2010;29:4008–19.

[58] Weber F, Hayer-Hartl M. Prevention of rhodanese aggregation by the chaperonin GroEL. In: Schneider C, editor. Chaperonin Protocols: Methods in Molecular Biology, 140. Totowa, NJ: Humana Press; 2000. p. 111–5.

[59] Poso D, Clarke AR, Burston SG. A kinetic analysis of the nucleotide-induced allosteric transitions in a single-ring mutant of GroEL. J Mol Biol 2004;338:969–77.

[60] Muller BK, Zaychikov E, Brauchle C, Lamb DC. Pulsed interleaved excitation. Biophys J 2005;89:3508–22.

[61] Widengren J, Mets U, Rigler R. Fluorescence correlation spectroscopy of triplet states in solution: a theoretical and experimental study. J Phys Chem 1995;99:13368–79.

[62] Müller CB, Loman A, Pacheco V, Koberling F, Willbold D, Richtering W, et al. Precise measurement of diffusion by multi-color dual-focus fluorescence correlation spectroscopy. Europhys Lett 2008;83:46001.

Supplementary data

Active cage mechanism of chaperonin-assisted protein folding demonstrated at single molecule level

Amit J. Gupta[1,†], Shubhasis Haldar[1,†], Goran Miličić[1], F. Ulrich Hartl[1] and Manajit Hayer-Hartl[1]

1 - Department of Cellular Biochemistry, *Max Planck Institute of Biochemistry, Am Klopferspitz 18, 82152 Martinsried, Germany*

[†] A.J.G. and S.H. contributed equally to this work, and they share first authorship.

Correspondence to F.U. Hartl, Tel.: +49 (0)8985782233; Fax: +49 (0)85782211; E-mail: uhartl@biochem.mpg.de **and M. Hayer-Hartl**, Tel.: +49 (0)8985782204; Fax: +49 (0)85782211; E-mail: mhartl@biochem.mpg.de

Supplementary Methods

Fluorescent labeling

For single molecule experiments, genetically engineered cysteine mutants of MBP and DM-MBP were constructed, taking advantage of the lack of Cys residues in the MBP sequence. The introduced, surface-exposed cystein residues were modified with fluorescent probes (AttoTec) using maleimide chemistry. The purified protein in buffer A/10 mM DTT was first buffer exchanged on a NAP5 column (Amersham Biosciences) equilibrated in buffer A and immediately mixed with a 1.2-molar excess of dye molecules (for single labeling) and incubated for 30 min at 20°C. In case of double-labeling of DM-MBP(30C/312C), the protein was incubated with a 3-fold molar excess of a 1:1 mixture of donor and acceptor dye (Atto532 and Atto647N, respectively). The labeling reaction was quenched by addition of 10 mM DTT. Free dye was removed and native conformation ensured by binding the labeled protein to an Amylose column (NEB) followed by extensive washing with buffer A/10 mM DTT. The labeled protein was then eluted using buffer A/10 mM DTT/5 mM maltose. Subsequently the buffer was exchanged on a NAP5 column (Amersham Biosciences) to buffer A/10 mM DTT. The protein was concentrated using Vivaspin concentrators (MWCO 3 kDa, GE Healthcare). The degree of labeling (DOL) was controlled by absorption spectroscopy using the following extinction coefficients (MBP, ε_{280} = 64860 M^{-1} cm^{-1}; Atto532, ε_{max} = 115000 M^{-1} cm^{-1} cf_{280} = 0.11; Atto647N, ε_{max} = 150000 M^{-1} cm^{-1} cf_{280} = 0.05; Atto655, ε_{max} = 125000 M^{-1} cm^{-1} cf_{280} = 0.08) and the equation,

$$\mathrm{DOL} = \frac{\mathrm{A}_{\mathrm{dye}} \times \varepsilon_{\mathrm{dye}}}{\left(-\left(\mathrm{A}_{\mathrm{dye}} \times \mathrm{cf}_{280}\right) + \mathrm{A}_{280}\right) \times \varepsilon_{\mathrm{protein}}}$$

DOL was >90 %. The absence of free dye in the sample was confirmed by FCS.

Fluorescence correlation spectroscopy (FCS)

FCS measurements using pulsed interleaved excitation (PIE) [60] were performed on a Microtime 200 inverse time-resolved fluorescence microscope (PicoQuant), which was maintained at a constant temperature of 20°C. For excitation of Atto647N, Atto655 and Atto532, ps pulsed diode lasers at 640 nm (LDH-PC-640B) and at 530 nm (LDH-P-FA-530) were used. Each laser had a laser power of 60 µW measured before the major dichroic. The lasers were pulsed with a rate of 27 mHz. The excitation light was guided through a water immersion objective (60 × 1.2 NA, Olympus) into the sample cuvette (Ibidi). The emitted fluorescence was separated from excitation light by a dichroic mirror (Z532/635RPC), guided through a pinhole (75 µm) and in case of cross-correlation split according to wavelength by a beamsplitter (600 DCXR) onto photon avalanche diodes (SPADs) (PDM series, MPD). The emission light was cleaned up by emission bandpass filters (HQ 690/70 and HQ 580/70, Chromas) in front of the respective detector. Detection was performed using time correlated single photon counting, making it possible to correlate any given photon with the excitation source. In case of auto-correlation measurements, after-pulsing artifacts were removed using fluorescence lifetime filters (Symphotime, PicoQuant) [63].

Dual color fluorescence cross-correlation spectroscopy (dcFCCS)

dcFCCS was employed to demonstrate the absence of inter-molecular association during spontaneous refolding of DM-MBP at 100 pM. DM-MBP(312C) was either labeled with Atto532 or Atto647N as described above. The labeled proteins were denatured at 10 nM in 6 M GuHCl/10 mM DTT for 1 h at 20°C. Refolding was induced by 200-fold dilution into buffer A to a final concentration of 50 pM each. FCCS was recorded with PIE [60] during refolding at 20^{o}C. As a positive control, 5 pM of DM-MBP(30C/312C), double labeled with Atto532 and Atto647N was mixed with 50 pM of each of the single labeled DM-MBP(312C) proteins, to mimic the presence of a dimeric species (the smallest aggregate) and to demonstrate the high sensitivity of this approach. A mixture of 50 pM native Atto532-labeled DM-MBP(312C) and 50 pM native Atto647N-labeled DM-MBP(312C) was used as a negative control.

Analysis of refolding by spFRET

DM-MBP(30C/312C) was double-labeled with Atto532 and Atto647N (DM-MBP(DL)), a commonly used FRET pair with a Förster radius of 52 Å [18,22]. DM-MBP(DL) was unfolded at a concentration of 20 nM in 6 M GuHCl/10 mM DTT for 1 h at 20°C.

Spontaneous refolding was induced upon 200-fold dilution into buffer A. Refolding at 100 pM was allowed to proceed at 20°C and was stopped at different time points by addition of 2 µM GroEL. Non-native conformers of DM-MBP are recognized by GroEL and by binding to the GroEL apical domains, are converted to a low f_E population. In contrast, natively folded protein molecules have high f_E and are not recognized by GroEL. For assisted refolding reactions, unfolded DM-MBP(DL) was diluted into buffer A containing 2 µM GroEL. Refolding was started by addition of 4 µM GroES and 5 mM ATP. Refolding was stopped by addition of 5 U apyrase (Sigma). Stopping was instant as no refolding was measured when apyrase was added immediately after initiating refolding with ATP (0 time point of refolding). Upon depletion of ATP, GroEL reverts to the apo-state that binds non-native DM-MBP(DL). After stopping refolding, reactions were transferred to the confocal microscope for spFRET analysis.

Measurements were performed on a Microtime 200 instrument in two color mode with PIE [18,60] as described under fluorescence correlation spectroscopy. Data was analyzed using a burst intensity approach [64,65] in Symphotime (PicoQuant). A single molecule diffusing through the confocal observation volume results in a burst in fluorescence intensity. A burst was considered as a significant event when it contained more than 25 photons in a 1 ms time window. In addition a threshold of 15 photons following red excitation was used to confirm the presence of a functional acceptor fluorophore. FRET efficiencies were calculated from fluorescence intensities of Donor I_D and Acceptor I_A fluorophore by the equation:

$$E = \frac{I_A}{I_A + \gamma I_D}$$

Where $\gamma = (\Phi_A \eta_A / \Phi_D \eta_D)$ denotes a correction factor for differences in quantum yields (Φ) and detection efficiencies (η) [18,66] and has been found to be 0.9 for the FRET pair used. Average intensity values of spectral crosstalk and direct excitation of acceptor fluorophores by the green laser were subtracted.

The resulting FRET efficiency histograms were analyzed using Origin (OriginLabs). To quantify the fraction of native molecules, the area of the histogram corresponding to native molecules was divided by the total area of the histogram. This fraction was plotted against refolding time and then fitted with a single exponential function, yielding the rate of refolding. Importantly, very similar results were obtained when the disappearance of the peak area corresponding to GroEL-bound molecules was analyzed. For each time point during refolding a minimum of 1000 particles was analyzed. All experiments were done at least in triplicate to ensure reproducibility.

In addition to steady state measurements, spFRET measurements were performed during the first minute of assisted refolding. DM-MBP(DL) was denatured in 6 M GuHCl/10 mM DTT at 20°C for GroEL/ES- and EL(KKK2)/ES-assisted refolding and in 10 M urea/10 mM DTT at 50°C for SREL/ES and SR(KKK2)/ES experiments. Unfolded protein was diluted 200-fold to a final concentration of 100 pM into buffer A or buffer B containing either 2 µM GroEL or 1 µM SREL, respectively**.** Refolding was initiated by addition of 4 µM GroES and 5 mM ATP and data collected immediately for 1 min. The experiment was repeated to obtain statistically significant FRET efficiency values from at least 1000 particles. From the resulting FRET efficiency histograms the relative peak areas for low FRET (chaperonin-bound molecules, f_E ~0.1) and high FRET (folded and encapsulated molecules, f_E ~0.65) were calculated. The amount of folded molecules within the first minute of the reaction was estimated by stopping the reaction with apyrase after 1 min. From the resulting FRET efficiency histogram, the bound state histogram was subtracted to give the amount of folded molecules. This amount of folded molecules was subtracted from the high FRET efficiency distribution to obtain the relative fraction of chaperonin-bound and encapsulated molecules.

Supplementary References

[63] Enderlein, J, Gregor, I, Patra, D, Fitter, J. Statistical analysis of diffusion coefficient determination by fluorescence correlation spectroscopy. J Fluoresc 2005;15:415-22.

[64] Deniz, AA, Dahan, M, Grunwell, JR, Ha, T, Faulhaber, AE, Chemla, DS, Weiss, S, Schultz, PG. Single-pair fluorescence resonance energy transfer on freely diffusing molecules: Observation of Förster distance dependence and subpopulations. Proc Natl Acad Sci USA 1999;96:3670-5.

[65] Zander, C, Sauer, M, Drexhage, KH, Ko, DS, Schulz, A, Wolfrum, J, Brand, L, Eggeling, C, Seidel, CAM. Detection and characterization of single molecules in aqueous solution. Appl Phys B 1996;63:517-23.

[66] Lee, NK, Kapanidis, AN, Wang, Y, Michalet, X, Mukhopadhyay, J, Ebright, RH, Weiss, S. Accurate FRET measurements within single diffusing biomolecules using alternating-laser excitation. Biophys J 2005;88:2939-53.

Supplementary Figures

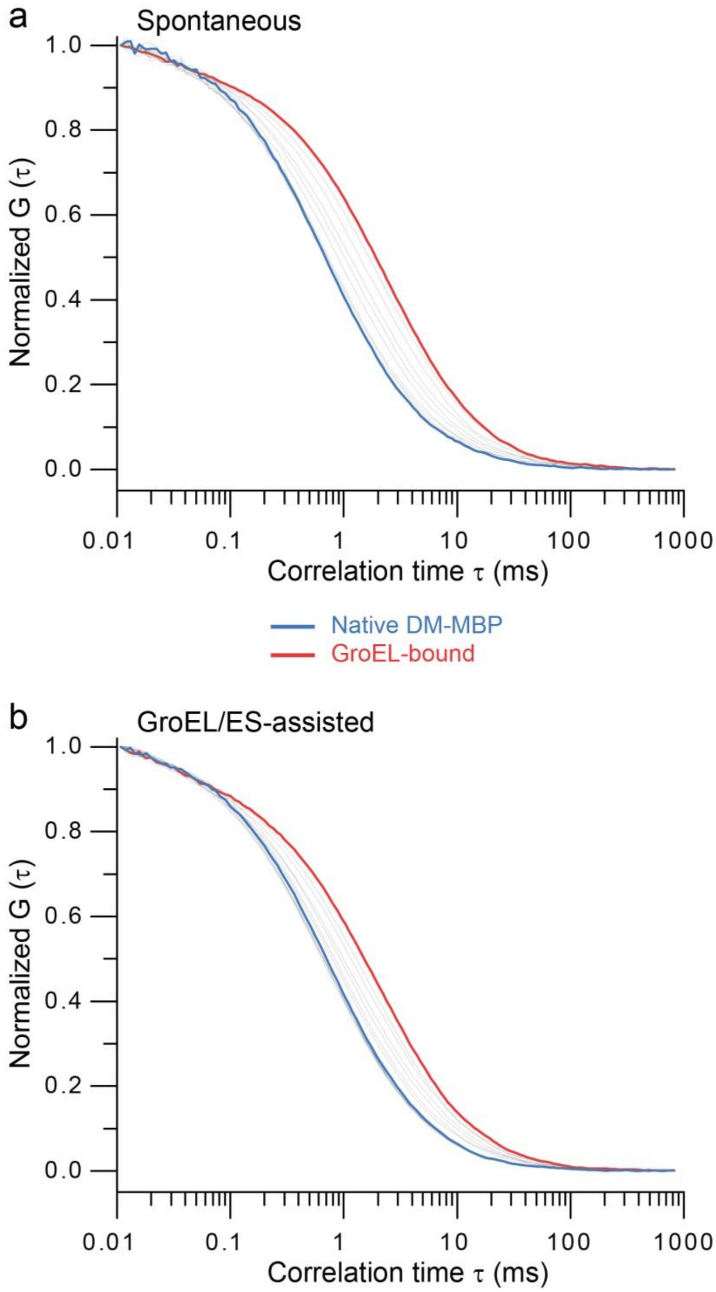


Fig. S1. Spontaneous and GroEL/ES-assisted refolding of DM-MBP measured by FCS.
(a) Auto-correlation curves of Atto647N fluorescence of DM-MBP(DL) at different times of spontaneous refolding as in Fig. 4b. The sample was subsequently analyzed by FCS. Correlation curves of GroEL-bound (red) and free native DM-MBP(DL) (blue), as well as all the time points taken during refolding (black) are shown.
(b) Auto-correlation curves as in (a) for assisted refolding.

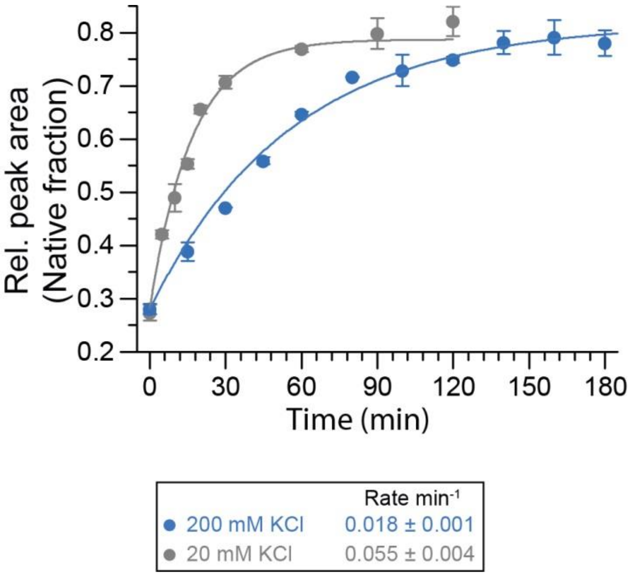


Fig. S2. Kinetics for spontaneous refolding of DM-MBP(DL) (denatured in 10 M urea/10 mM DTT) in buffer A containing 20 mM KCl were obtained from spFRET measurements as in Fig. 3e and compared to spontaneous refolding in bufferA/200 mM KCl (data of Fig 3e). Data were fitted with a single exponential rate. Data represent averages ± s.d. from at least 3 independent measurements.

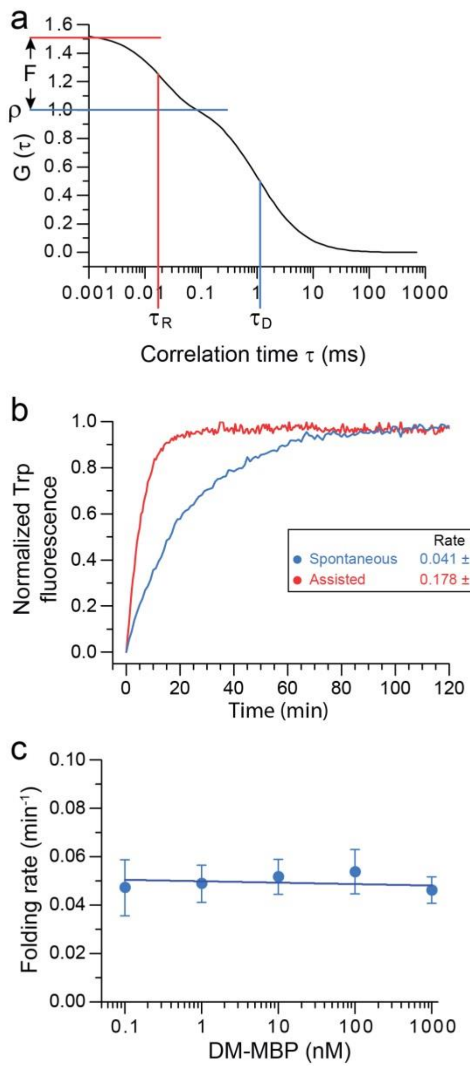


Fig. S3. Analysis of DM-MBP refolding by PET-FCS.

(a) Simulation of a characteristic FCS measurement with one diffusion and one exponential component. Fit parameters are indicated. τ_D and ρ indicate diffusion time and amplitude of the diffusion component, respectively, and F and τ_R indicate amplitude of the exponential term arising from PET and relaxation time, respectively.

(b) Spontaneous and assisted refolding of DM-MBP(312C) analyzed by Trp fluorescence at a final concentration of 100 nM in buffer A at 20°C. Averages ± s.d. from at least 3 independent measurements are shown.

(c) Rates of spontaneous refolding of DM-MBP(Atto655) either alone at 100 pM or 1 nM, or at 1 nM in the presence of 10, 100 or 1000 nM of unlabeled, refolding DM-MBP(312C). Refolding was monitored by recording the time-dependent decrease in the amplitude of the exponential component F in PET-FCS as in Fig. 5e. Rates were extracted by single exponential fit. Data represent averages ± s.d. from at least 3 independent measurements.

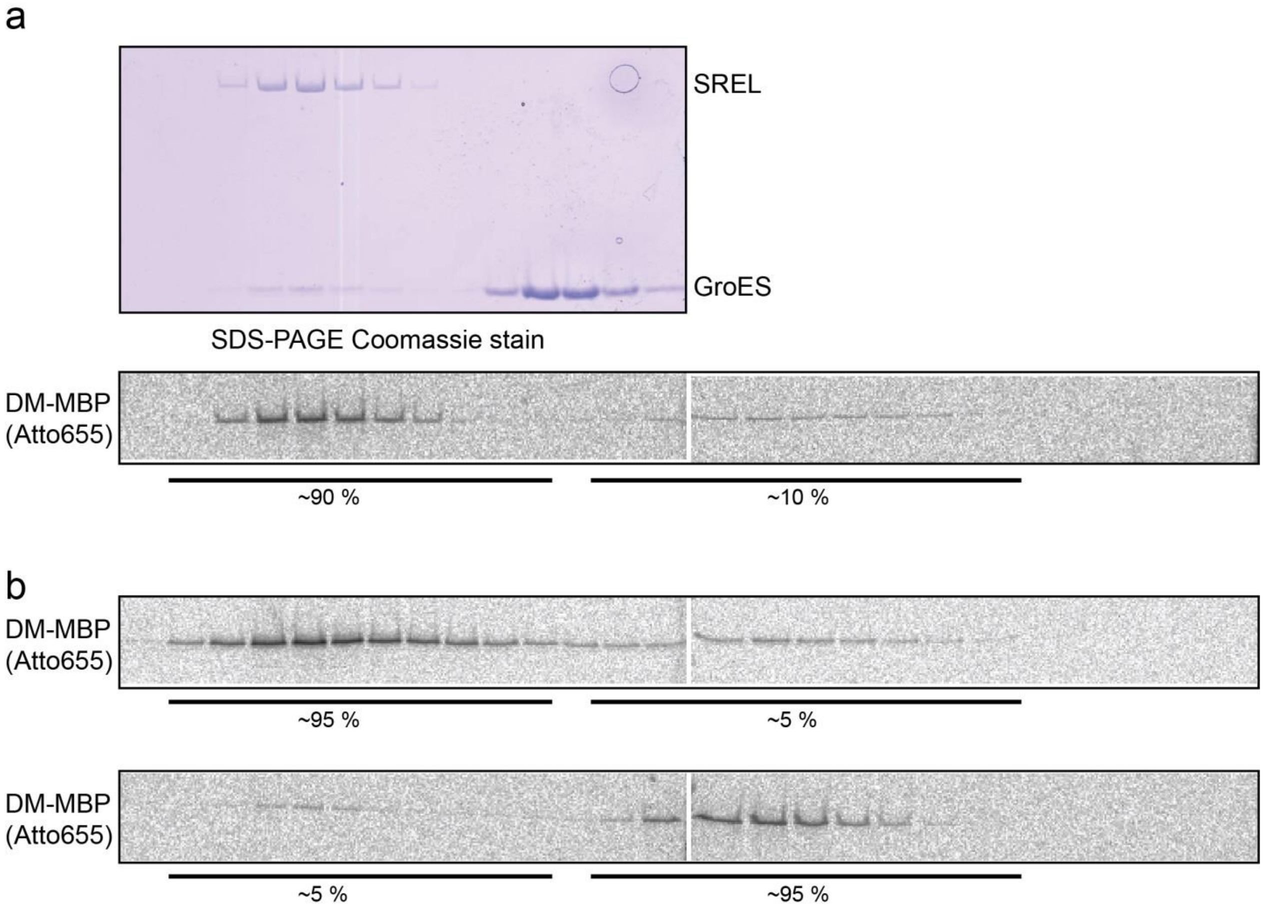


Fig. S4. Encapsulation of DM-MBP(Atto655) by SREL/ES analyzed by size-exclusion chromatography.

(a) DM-MBP(Atto655) was unfolded in 10 M urea/10 mM DTT for 1 h at 50°C and diluted 200-fold into refolding buffer B at 20^{o}C containing 1 µM SREL. After 5 min encapsulation was initiated by addition of 4 µM GroES and 1 mM ATP, and the reaction was applied to a Superdex 200 gel filtration column equilibrated in buffer B/50 mM urea/1 mM ATP. Fractions (50 µl) were collected over a period of 30 min. Top-panel: Coomassie stained 15 % SDS-PAGE shows coelution of SREL and GroES. Visualization of the SDS-PAGE with a fluorescence imager showed that DM-MBP(Atto655) coelutes with SREL. Quantification of the DM-MBP(Atto655) bands by densitometry is indicated.

(b) In a control experiment the SREL/ES-assisted refolding reaction as in (a) was incubated for 30 min at 20^{o}C, followed by size-exclusion chromatography without further additions or after addition of 50 mM CDTA/200 mM KCl /70 mM GuHCl to induce the dissociation of the SREL/ES complex and allowing the release of folded, encapsulated protein. The fractions were analyzed as in (a).

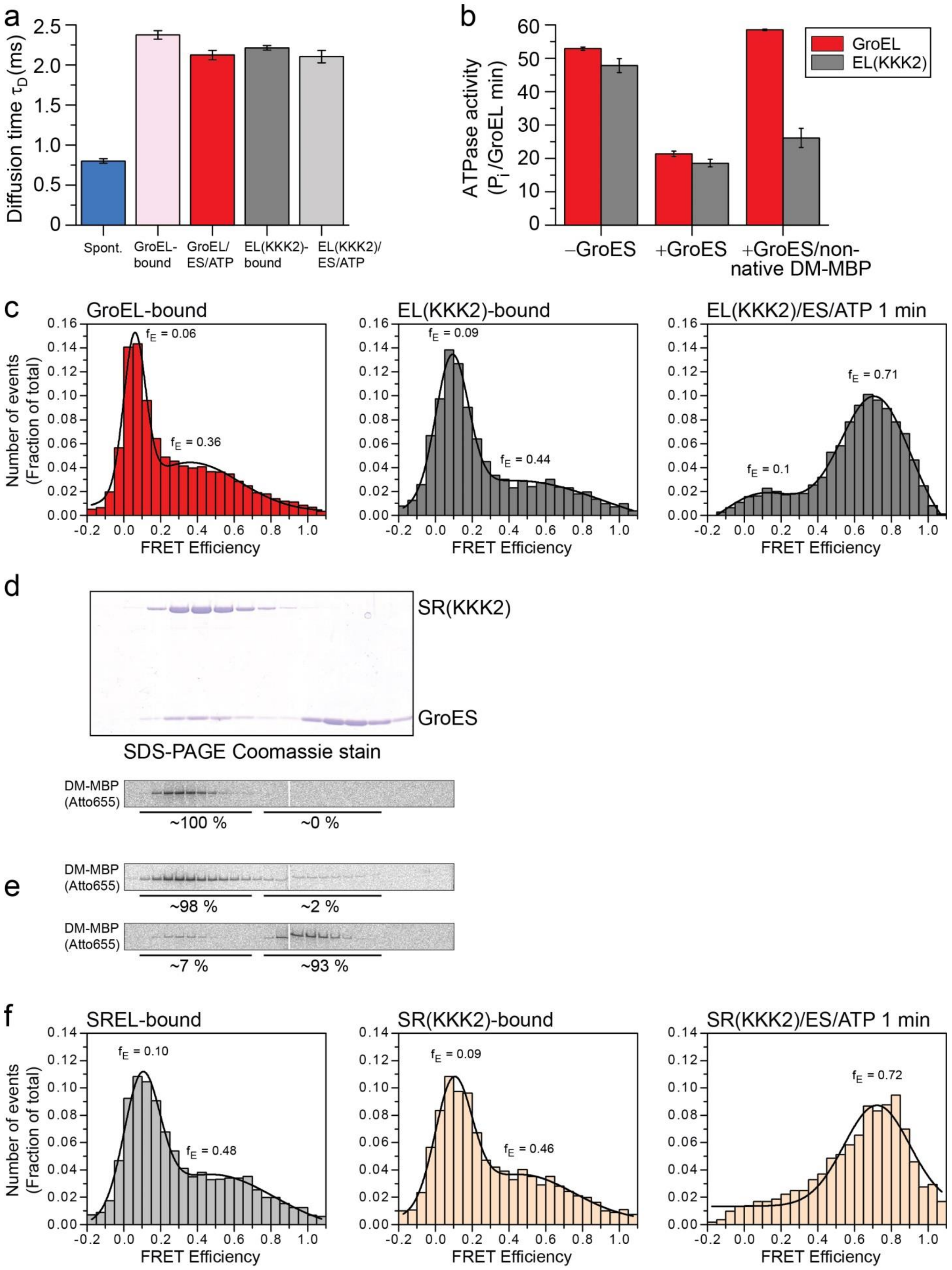

a
Diffusion time τD (ms)
2.5
2.0
1.5
1.0
0.5
0
Spont.
GroEL-bound
GroEL/ES/ATP
EL(KKK2)-bound
EL(KKK2)/ES/ATP
b
ATPase activity (Pi/GroEL min)
60
50
40
30
20
10
0
GroEL
EL(KKK2)
−GroES
+GroES
+GroES/non-native DM-MBP
c
GroEL-bound
EL(KKK2)-bound
EL(KKK2)/ES/ATP 1 min
Number of events (Fraction of total)
fE = 0.06
fE = 0.36
fE = 0.09
fE = 0.44
fE = 0.71
fE = 0.1
FRET Efficiency
d
SR(KKK2)
GroES
SDS-PAGE Coomassie stain
DM-MBP (Atto655)
~100 %
~0 %
e
DM-MBP (Atto655)
~98 %
~2 %
DM-MBP (Atto655)
~7 %
~93 %
f
SREL-bound
SR(KKK2)-bound
SR(KKK2)/ES/ATP 1 min
Number of events (Fraction of total)
fE = 0.10
fE = 0.48
fE = 0.09
fE = 0.46
fE = 0.72
FRET Efficiency

Fig. S5. Effect of net-negative charges in the chaperonin cis-cavity on DM-MBP folding and conformational dynamics.

(a) Diffusion times (τ_D) of DM-MBP(DL) measured by FCS (see Fig. S3a) during the first minute of spontaneous refolding or bound to GroEL or EL(KKK2) or during the first minute of refolding with GroEL/ES/ATP or EL(KKK2)/ES/ATP in buffer A at 20^{o}C. Data represent averages ± s.d. from at least 3 independent measurements.

(b) ATPase activity of GroEL (0.2 µM) was measured in buffer A at 20°C in absence or presence of GroES (0.4 µM) or 0.4µM GroES (0.4 µM)/non-native DM-MPB (1 µM). Measurements were performed as in Fig. 6c-d. Data represent averages ± s.d. from at least 3 independent measurements.

(c) FRET efficiency (f_E) histograms from spFRET measurements of GroEL- and EL(KKK2)-bound DM-MBP(DL), as well as during the first minute of DM-MBP(DL) refolding with EL(KKK2)/ES . Measurements were performed and analyzed as in Fig. 6a. Histograms shown are representative of at least three independent experiments.

(d) Encapsulation of DM-MBP(Atto655) by SR(KKK2)/ES analyzed by size-exclusion chromatography as in Fig. S4a.

(e) In a control experiment the SR(KKK2)/ES-assisted refolding reaction as in (c) was incubated for 60 min at 20^{o}C, followed by size-exclusion chromatography without further additions or after dissociation of the SR(KKK2)/ES complex as in Fig. S4b. The fractions were analyzed as in Fig. S4a.

(f) f_E histograms from spFRET measurements of SREL- and SR(KKK2)-bound DM-MBP(DL), as well as during the first minute of DM-MBP(DL) refolding with SR(KKK2)/ES. Measurements were performed and analyzed as in Fig. 6b. Histograms shown are representative of least three independent experiments.

3.2 Article 2 - Chaperonin-assisted protein folding: Relative population of asymmetric and symmetric GroEL:GroES complexes

In the GroEL field there is an ongoing debate whether the functional form of the GroEL/ES system in the folding cycle is an asymmetrical, bullet shaped GroEL/ES complex with GroES bound to one side of GroEL or a symmetrical, American football shaped GroEL/ES complex with GroES bound to both sides of GroEL[86,140,141,142]. The following study addressed this question by measuring cross-correlation between two populations of differently labelled GroES molecules on a single molecule level, under variety of conditions. We also employed a range of other biochemical and biophysical methods. We showed that symmetrical complexes (GroEL:GroES$_2$) are populated when the GroEL/ES system interacts with model non-foldable substrates, such as α-lactalbumin (α-LA) and α-casein, which functionally uncouple the two GroEL rings. Uncoupling of the rings leads to a diminished negative allostery between the two rings and allows GroES to bind on both sides of GroEL. In the absence of any substrate, or in the presence of foldable substrates, for example mitochondrial rhodanese (mRho), mitochondrial malate dehydrogenase (mMDH), DM-MBP or *Rhodospirillum rubrum* Rubisco, the canonical asymmetric form of the GroEL/ES complex is the prevailing species. Additionally, we showed that physiological ratios of ATP:ADP favour the formation of bullet shaped GroEL/ES complexes. These observations led us to conclude that the physiologically relevant form of the GroEL/ES complex is asymmetrically shaped and that GroEL/ES system works as a two-stroke machine.

Journal of Molecular Biology (2015)[143]

Haldar S, Gutpa AJ, Yan X, **Miličić G**, Hartl FU, Hayer-Hartl M.

Contribution: This project was performed in collaboration with Shubhasis Haldar, Amit Gupta and Xiao Yan. I conducted negative stain electron microscopy experiments of GroEL:GroES complexes under variety of conditions. In addition, I helped in performing spectroscopic measurements of the GroEL ATPase activity, which led to understanding the substrate effect on decoupling of both rings.

Chaperonin-Assisted Protein Folding: Relative Population of Asymmetric and Symmetric GroEL:GroES Complexes

Shubhasis Haldar[1], Amit J. Gupta[1], Xiao Yan, Goran Miličić, F. Ulrich Hartl and Manajit Hayer-Hartl

Department of Cellular Biochemistry, *Max Planck Institute of Biochemistry, Am Klopferspitz 18, 82152 Martinsried, Germany*

Correspondence to F. Ulrich Hartl and Manajit Hayer-Hartl: uhartl@biochem.mpg.de; mhartl@biochem.mpg.de

http://dx.doi.org/10.1016/j.jmb.2015.04.009
Edited by S. Radford

Abstract

The chaperonin GroEL, a cylindrical complex consisting of two stacked heptameric rings, and its lid-like cofactor GroES form a nano-cage in which a single polypeptide chain is transiently enclosed and allowed to fold unimpaired by aggregation. GroEL and GroES undergo an ATP-regulated interaction cycle that serves to close and open the folding cage. Recent reports suggest that the presence of non-native substrate protein alters the GroEL/ES reaction by shifting it from asymmetric to symmetric complexes. In the asymmetric reaction mode, only one ring of GroEL is GroES bound and the two rings function sequentially, coupled by negative allostery. In the symmetric mode, both GroEL rings are GroES bound and are folding active simultaneously. Here, we find that the results of assays based on fluorescence resonance energy transfer recently used to quantify symmetric complexes depend strongly on the fluorophore pair used. We therefore developed a novel assay based on fluorescence cross-correlation spectroscopy to accurately measure GroEL:GroES stoichiometry. This assay avoids fluorophore labeling of GroEL and the use of GroEL cysteine mutants. Our results show that symmetric GroEL:GroES$_2$ complexes are substantially populated only in the presence of non-foldable model proteins, such as α-lactalbumin and α-casein, which "over-stimulate" the GroEL ATPase and uncouple the negative GroEL inter-ring allostery. In contrast, asymmetric complexes are dominant both in the absence of substrate and in the presence of foldable substrate proteins. Moreover, uncoupling of the GroEL rings and formation of symmetric GroEL:GroES$_2$ complexes is suppressed at physiological ATP:ADP concentration. We conclude that the asymmetric GroEL:GroES complex represents the main folding active form of the chaperonin.

Introduction

The chaperonin GroEL and its cofactor GroES have an essential function in folding a subset of proteins in the bacterial cytosol [1]. GroEL is an ATP-driven macromolecular machine of ~800 kDa consisting of two rings with seven ~57-kDa subunits stacked back to back. GroES is a single heptameric ring of ~10-kDa subunits that attaches coaxially to the ends of the GroEL cylinder [2,3] (Fig. 1). Protein substrate binds as a molten globule-like folding intermediate in the ring center of GroEL to hydrophobic amino acid residues exposed by the apical GroEL domains. Upon binding of ATP to GroEL, GroES caps the GroEL ring that holds the substrate (*cis*-ring), resulting in its displacement into an enclosed chamber large enough for proteins up to ~60 kDa. This step is accompanied by conformational changes that render the GroEL cavity hydrophilic. The encapsulated protein is free to fold during the time required for the hydrolysis of the seven ATP molecules in the *cis*-ring. According to current understanding, binding of ATP to the opposite GroEL ring (*trans*-ring), followed closely by GroES binding, then causes the dissociation of GroES from the *cis*-ring and substrate release into solution. Prior to ATP binding, the *trans*-ring may capture a new substrate molecule.

The GroEL/ES system is allosterically regulated with positive cooperativity of ATP binding and hydrolysis within rings and negative cooperativity between rings [3,4]. In the chaperonin reaction cycle

0022-2836/© 2015 Elsevier Ltd. All rights reserved.

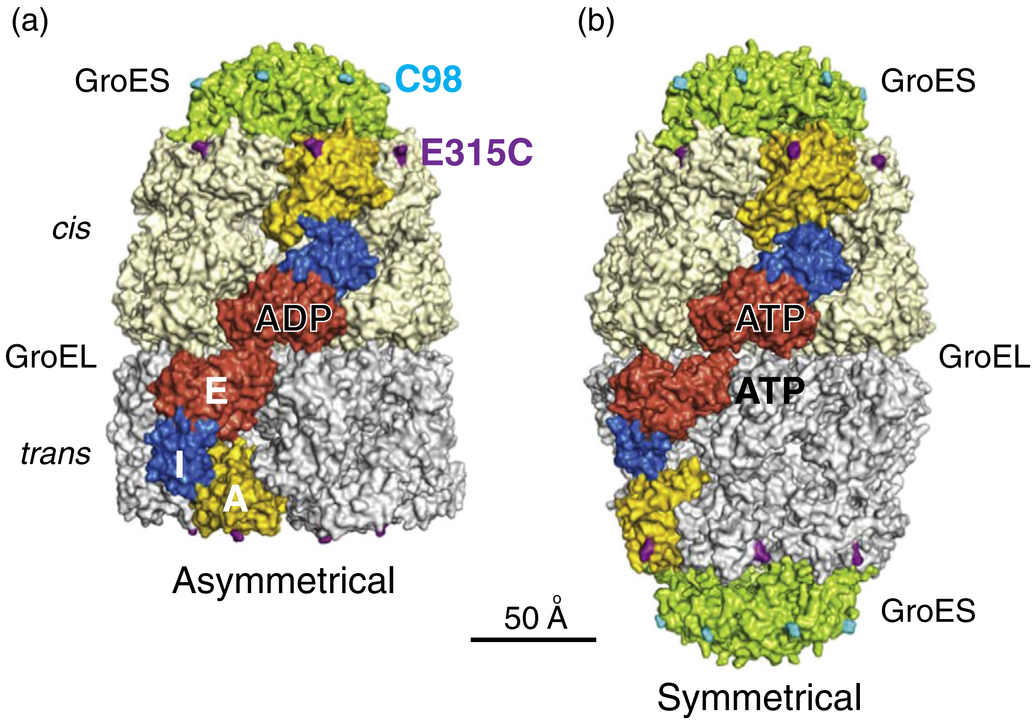


Fig. 1. Structures of asymmetric and symmetric GroEL/ES complexes. (a) Space-filling representation of the crystal structure of the asymmetric GroEL/ES:ADP complex (PDB ID 1AON) [40] with the *cis*- and *trans*-rings of GroEL colored beige and gray, respectively. For one subunit of GroEL in each of the rings, the apical (A), intermediate (I) and equatorial (E) domains are colored gold, blue and red, respectively. The cofactor GroES is colored green. The fluorophore labeling positions on GroEL(E315C) and GroES(98C) are indicated. (b) The crystal structure of the symmetric EL(D52A/D398A)/ES-ATP complex (PDB ID 3WVL) [41], colored as in (a).

as described above, the two GroEL rings function in an alternating fashion with the GroES-bound *cis*-ring forming an enclosed folding environment while the *trans*-ring is open to receive a new substrate protein. Such asymmetric GroEL:GroES complexes [5,6] are often referred to as "bullet" shaped [7] (Fig. 1a). In addition, symmetric GroEL:GroES$_2$ complexes, with both GroEL rings being capped by GroES, have also been observed (referred to as "football" shaped) [7–9] (Fig. 1b). The functional significance of these symmetric complexes has been subject of debate [10–15], but there was agreement that they may occur as a transient intermediate in the reaction cycle, when a GroES binds to the *trans*-ring before release of GroES from the *cis*-ring is complete. However, recent reports marked a notable deviation from this basic concept [16–19]. These studies, employing calibrated fluorescence resonance energy transfer (FRET) to monitor the GroEL:GroES interaction, suggest that symmetric complexes are highly populated during the functional cycle and thus are proposed to be the main, if not the only, folding active form of the chaperonin. Specifically, it was reported that nearly 100% symmetric complexes are induced by the binding of substrate protein to GroEL, with the two rings functioning independently and GroES binding and release occurring in a stochastic fashion [17–19]. Thus, fundamentally different mechanisms would underlie the GroEL/ES reaction cycle in the absence and presence of substrate protein.

Here, we use dual-color fluorescence cross-correlation spectroscopy (dcFCCS) to accurately measure the occurrence of the different GroEL:GroES complexes under a variety of conditions. By using two populations of differently labeled GroES, the dcFCCS method allowed us to analyze the interaction of GroES with wild-type GroEL in solution, thereby avoiding possible artifacts associated with the use of fluorophore-labeled mutant GroEL. Our results show that symmetric complexes are substantially populated only in the presence of proteins that cannot be folded by the chaperonin, a condition under which the GroEL rings become uncoupled. In contrast, symmetric complexes are a minor species either in the absence of substrate or in the presence of foldable substrate proteins or at a physiological concentration ratio of ATP and ADP.

We conclude that the asymmetric GroEL:GroES complex represents the main folding active form of the chaperonin and that fundamentally similar mechanisms govern the GroEL/ES interaction cycle in the absence and presence of substrate protein.

Results

Analysis of GroEL:GroES complexes by FRET

It is well established that GroEL and GroES form asymmetric complexes in the presence of ADP [5,15,20], whereas symmetric complexes form in the presence of ATP/BeF_X (a beryllium fluoride complex) [18,21]. Under the former condition, only the *cis*-ring has ADP bound. In contrast, in the presence of ATP/BeF_X, all 14 nucleotide sites are trapped in the hydrolysis transition state with ADP/BeF_X bound and the functional asymmetry between the two rings is lost [21] (Fig. 1). However, there is disagreement as to the relative occurrence of asymmetric and symmetric complexes during the functional chaperonin cycle in the presence of ATP and substrate protein.

Much of the uncertainty regarding the functional significance of the different GroEL:GroES complexes is associated with the methods used for their detection and quantification. Recently, two groups have used calibrated FRET between fluorophore-labeled GroEL and GroES to measure complex formation and determine complex stoichiometry [17–19]. While the labeling positions are the same in the two studies (E315C in the apical substrate binding domain of GroEL; 98C in GroES), the fluorophore pairs are different {Cy3 and Cy5 [17]; IAEDANS (5-[2-[(2-iodo-1-oxoethyl)amino]ethylamino]-1-naphthalenesulfonic acid) and F5M (fluorescein-5-maleimide) [18,19]}. Notably, the results obtained with the different fluorophore pairs are not consistent, varying dramatically in the proportions of symmetric complexes during the GroEL/ES functional cycle. We therefore reassessed the FRET-based quantification. As in the previous studies, we introduced the E315C mutation for site-specific labeling of GroEL and added a C-terminal cysteine residue (position 98) in GroES. Labeling efficiencies and stoichiometries with the respective fluorophore pairs were essentially as previously reported [16–18]. To calibrate the amount of FRET between IAEDANS/F5M and Cy3/Cy5 upon formation of asymmetric or symmetric complexes, we measured fluorescence spectra upon donor excitation (IAEDANS, 336 nm; Cy3, 550 nm) either in the absence of nucleotide or in the presence of ADP/BeF_X or ATP/BeF_X at 25 °C (Fig. 2a and b). GroES does not bind GroEL in the absence of nucleotide, ADP/BeF_X results in the formation of only asymmetric complexes and symmetric complexes are populated in the presence of ATP/BeF_X [18,21]. As expected, for both fluorophore pairs, we observed FRET upon GroES binding to GroEL, reflected in a decrease in donor fluorescence and a concomitant increase in acceptor fluorescence (Fig. 2a and b). In the presence of ADP/BeF_X (asymmetric complexes), donor fluorescence decreased by ~27% for Cy3/Cy5 and ~31% for IAEDANS/F5M relative to the control in the absence of nucleotide (Fig. 2a and b). In the presence of ATP/BeF_X (symmetric complexes), donor fluorescence decreased by ~55% for Cy3/Cy5 and ~56% for IAEDANS/F5M (Fig. 2a and b). Note that the increase in acceptor fluorescence resulting from FRET is relatively small due to the 5-fold excess of GroES over GroEL used in the assay [18], but the relative change in donor fluorescence associated with the formation of symmetric complexes is similar for the two fluorophore pairs. However, during functional cycling of GroEL/ES in the presence of ATP and ATP regenerating system, donor fluorescence for Cy3/Cy5 decreased substantially (by ~21%) below the value obtained with ADP/BeF_X, whereas only a marginal decrease (by ~8%) was measured with IAEDANS/F5M (Fig. 2a and b). Based on the relative increase in FRET efficiency (f_E) compared to the 100% symmetric complex control (ATP/BeF_X), these values correspond to ~59% and ~16% symmetric GroEL:$GroES_2$ complexes, respectively, being populated during cycling with ATP (Fig. 2c). Similar results were obtained when the measurements were performed at 37 °C (Fig. S1). This striking difference in the proportions of symmetric complexes observed here and reported previously [17,18] indicates that the calibrated FRET assay is highly dependent on the fluorophore pair used.

The calibrated FRET studies also presented data that substrate protein induces the formation of symmetric complexes [17–19]. We therefore tested the effect of substrate using the different FRET pairs. The model substrate α-lactalbumin (αLA), a disulfide-bonded secretory protein, is unable to fold in the presence of DTT and in the absence of calcium ions. Under these conditions, αLA populates an ensemble of molten globule-like folding intermediates with affinity for GroEL [22]. Addition of this non-foldable protein to the IAEDANS/F5M chaperonin reaction (at a 30-fold excess over GroEL [19]) resulted in an increase in symmetric complexes from ~16% during cycling in the absence of substrate to ~64% in the presence of αLA at 25 °C (Fig. 2c) and to ~76% at 37 °C (Fig. S1c), which is less than the reported value of 100% [19]. In contrast, with the Cy3/Cy5 fluorophore pair, similar proportions of symmetric complexes (~50–60%) were measured in the absence and presence of αLA at both temperatures (Fig. 2c and Fig. S1c). To analyze the effect of a foldable substrate, we used a double mutant of

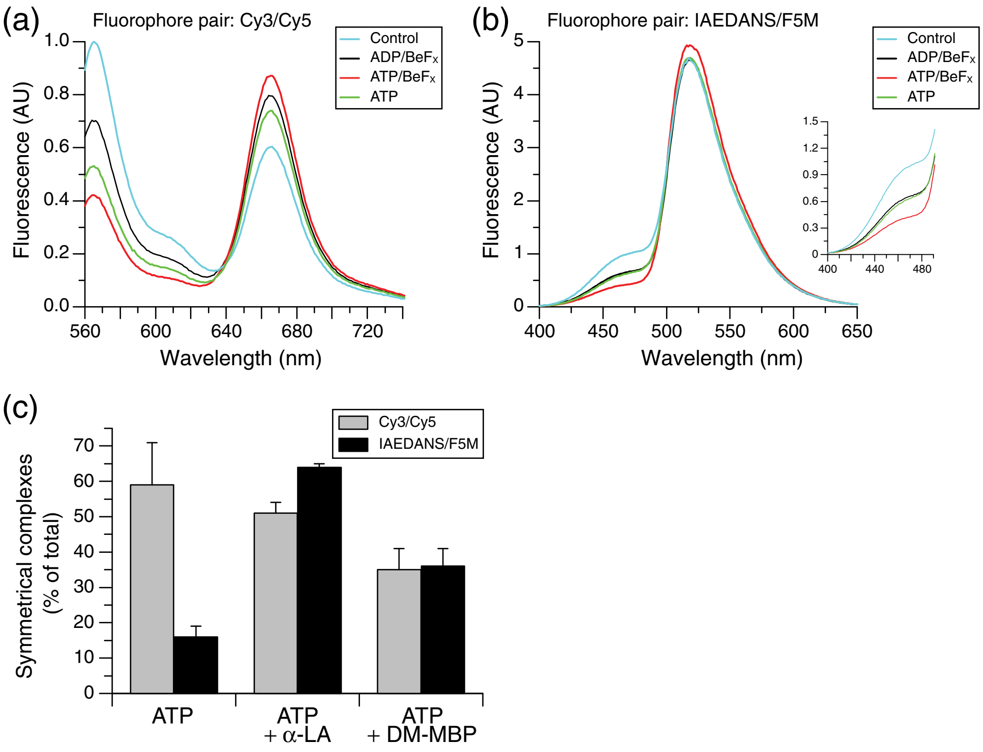


Fig. 2. Analysis of GroEL/ES symmetric complexes by calibrated FRET. The fluorophore pairs Cy3/Cy5 (a) and IAEDANS/F5M (b) were used as FRET reporters [17,18]. We analyzed 140 nM GroEL(E315C) labeled with IAEDANS or Cy3 and 700 nM GroES(98C) labeled with F5M or Cy5 in buffer C at 25 °C in the absence of nucleotide (Control) or in the presence of either ADP/BeF$_X$ or ATP/BeF$_X$ or ATP with ATP regenerating system. Fluorescence spectra were recorded immediately after nucleotide addition with donor excitation at 550 nm (a) or 336 nm (b). The inset in (b) shows a zoom-in on the donor fluorescence emission. (c) The proportion of symmetric GroEL:GroES$_2$ complexes was estimated based on the donor fluorescence relative to the donor fluorescence in the presence of ADP/BeF$_X$ (0% symmetric complexes) and in the presence of ATP/BeF$_X$ (100% symmetric complexes) (see Materials and Methods). Donor fluorescence was measured at 565 nm (Cy3) and 465 nm (IAEDANS). The proportion of symmetric complexes was also measured in the presence of a 30-fold excess of non-foldable substrate protein αLA or a 6-fold excess of foldable DM-MBP. Averages ± SD from three independent experiments are shown.

maltose binding protein (DM-MBP). This protein has a high affinity for GroEL in the non-native state and its spontaneous folding is accelerated 5- to 10-fold by GroEL/ES [23,24]. Since the spontaneous folding is slow ($t_{1/2}$ ~ 30 min at 25 °C) but highly efficient, DM-MBP can be added to the chaperonin reaction in excess over GroEL [25]. In the case of GroEL and GroES labeled with IAEDANS and F5M, respectively, DM-MBP also induced the formation of symmetric complexes, although to a lesser extent than αLA (Fig. 2c). In contrast, with the Cy3/Cy5-labeled proteins, DM-MBP substantially reduced the proportion of symmetric complexes compared to the reaction with ATP in the absence of substrate (Fig. 2c). Similar results were obtained at 37 °C (Fig. S1c). Thus, dependent on the FRET pair used, protein substrates seem to either induce or reduce the occurrence of symmetric complexes, indicating that the calibrated FRET assay is not a reliable method to measure GroEL:GroES stoichiometries in the functional chaperonin reaction.

Quantitative analysis of GroEL:GroES complexes by dcFCCS

Given the uncertainties associated with the FRET measurements, we developed a novel approach to quantify the relative proportion of asymmetric and symmetric GroEL:GroES complexes using dcFCCS. In this assay, the co-diffusion of two differently labeled GroES molecules through the confocal observation volume results in a quantifiable cross-correlation signal, the amplitude of which is proportional to the relative amount of complexes that contain both labeled species. Here, we labeled GroES(98C) with either a single molecule of Atto532 (ES532) or Atto655 (ES655) per heptamer and GroEL remained unlabeled. The labeled GroES bound efficiently to GroEL in a nucleotide-dependent manner, as demonstrated by measuring cross-correlation of labeled GroEL and GroES (Fig. S2). An important advantage of the dcFCCS assay is that cross-correlation is only observed for GroEL:GroES$_2$ complexes. Moreover, any form of mutant GroEL can be readily analyzed, as labeling of GroEL is not required. Note that the diffusion time of GroEL complexes through the confocal volume was measured to be ~1 ms at 20 °C [25,26], at least 3 orders of magnitude faster than the half-life of GroEL:GroES complexes that is determined by ATPase activity (see Ref. [25] and Fig. 5 below).

As expected, only background cross-correlation was measured with 100 nM GroEL and 200 nM each of ES532 and ES655 in the presence of ADP or ADP/BeF$_X$ (Fig. 3a and b), where 100% asymmetric complexes are populated. A clear cross-correlation signal was observed with the ATP-hydrolysis-deficient GroEL mutant, EL-D398A (Fig. 3a), which binds ATP but hydrolyzes it very slowly [15] and generates 100% symmetric complexes with GroES [16,27]. Cross-correlation in the presence of ATP/BeF$_X$ with wild-type GroEL was similar to that measured with EL-D398A and ATP (Fig. 3b). Interestingly, only ~10% symmetric complexes were detected with GroEL/ES in the presence of ATP (with ATP regenerating system), indicating that mainly asymmetric complexes are populated during functional cycling in the absence of substrate protein (Fig. 3a and b). The cross-correlation signal with EL-D398A/ES ATP was linearly dependent on the concentration of EL-D398A, demonstrating the quantitative nature and high sensitivity of the assay (Fig. 3c).

Differential effect of foldable and non-foldable substrate protein

Next we analyzed the effect of substrate protein on the relative proportion of asymmetric and symmetric GroEL:GroES complexes. To exclude photophysical artifacts, we measured the auto-correlation for the two labeled GroES species individually in the presence of GroEL/ATP and increasing amounts of non-native αLA or DM-MBP. We found the auto-correlation amplitudes to be independent of the presence of

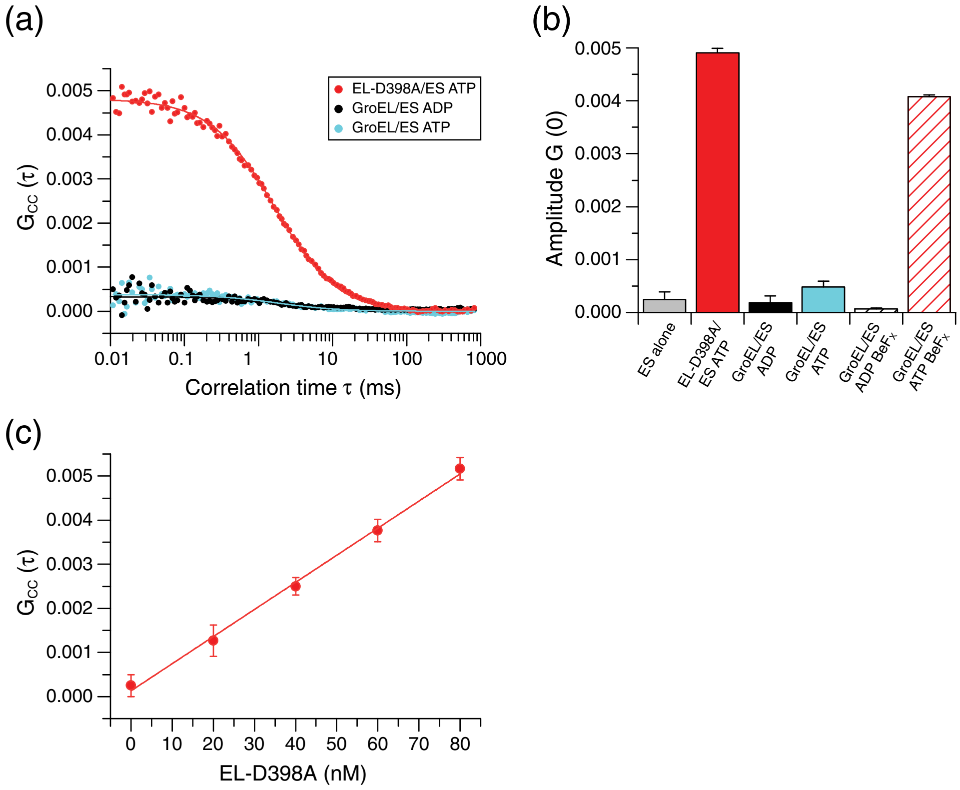


Fig. 3. Quantification of GroEL/ES complexes by dcFCCS. (a) Cross-correlation of ES655 and ES532 indicates formation of symmetric GroEL/ES complexes. A 1:1 mixture of Atto655- and Atto532-labeled GroES(98C) (400 nM total) was added to buffer A containing 100 nM of either EL-D398A or GroEL. Complex formation was initiated by the addition of 1 mM ATP (with ATP regenerating system) or ADP and the cross-correlation was measured for a time window of 10 min at 20 °C. (b) Cross-correlation amplitudes of GroES obtained as above either with EL-D398A or with GroEL in the nucleotide conditions indicated. Background cross-correlation was measured in the absence of GroEL (ES alone). Averages ± SD from three independent experiments are shown. (c) Cross-correlation was measured as in (a) with increasing concentrations of EL-D398A. The amplitude of the cross-correlation signal is plotted against the EL-D398A concentration. Averages ± SD from three independent experiments are shown.

substrate (Fig. S3), excluding fluorescence quenching in the presence of substrate protein. Cross-correlation experiments with various substrate proteins were performed in the presence of an ATP regenerating system. The non-foldable αLA resulted in a cross-correlation signal that saturated at an ~25-fold excess of αLA over GroEL and corresponded to ~54% symmetric complexes (Fig. 4a and g, as well as data not shown), indicating that symmetric complexes are readily observed by the dcFCCS method under conditions of active GroEL/ES cycling. α-Casein, another non-foldable protein with an affinity for GroEL

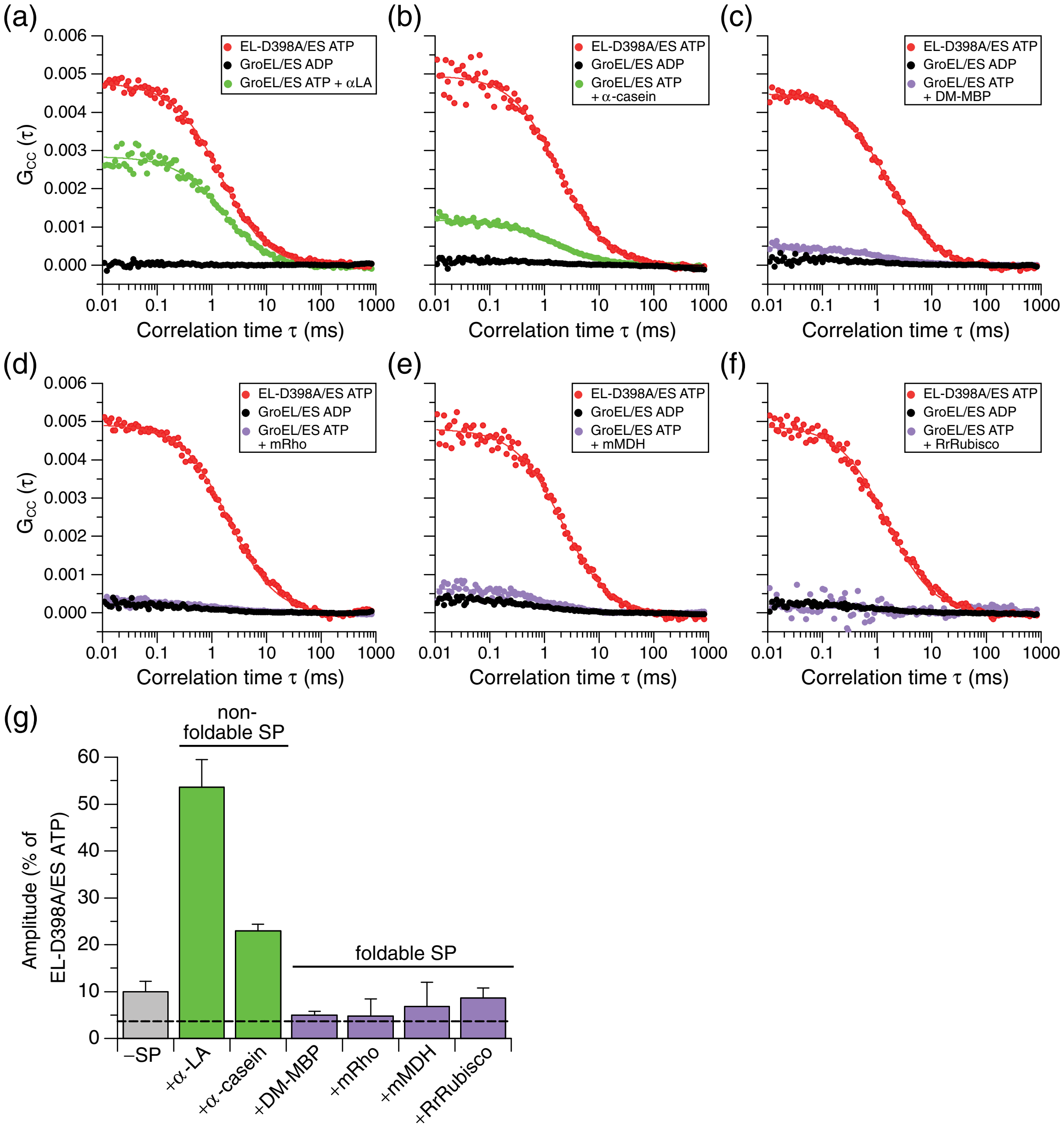


Fig. 4. Effect of non-foldable and foldable substrate proteins on formation of symmetric GroEL:GroES$_2$ complexes. Symmetric complexes were quantified by dcFCCS as in Fig. 3 in the presence of ATP and ATP regenerating system. Effect of non-foldable substrate proteins αLA and α-casein (a and b) and foldable substrate proteins DM-MBP, mRho, mMDH and RrRubisco (c–f) on the cross-correlation signal. Cross-correlation was recorded for a time window of 10 min for non-foldable substrates and 1 min for foldable substrate proteins. The signals for EL-D398A/ES ATP and for GroEL/ES ADP are shown as controls. (g) Quantification of symmetric complexes. The cross-correlation signal obtained with EL-D398A/ES ATP is set to 100%. The dotted line indicates the average cross-correlation signal with GroEL/ES ADP. SP, substrate protein. Averages ± SD from three independent experiments are shown.

[28], also induced symmetric complexes, albeit to a lesser extent of ~23% (Fig. 4b and g). In contrast, four different foldable substrate proteins, DM-MBP, mRho (mitochondrial rhodanese), mMDH (mitochondrial malate dehydrogenase) and RrRubisco (*Rhodospirillum rubrum* ribulose bisphosphate carboxylase oxygenase), when added from denaturant, did not induce measurable GroEL:GroES$_2$ complexes (Fig. 4c–f and g). Thus, we were unable to reproduce recent reports using calibrated FRET that foldable substrate proteins induce symmetrical complexes to 80–90% [18,19]. Note that our measurements were performed with saturating amounts of substrate protein and during the first minute of refolding to ensure substrate occupation of GroEL (see Materials and Methods). Moreover, in the case of DM-MBP, which does not aggregate significantly during spontaneous refolding [23,24], substrate saturation of GroEL was ensured by using a 6-fold molar excess over GroEL [25]. The residual concentration of denaturant (~6 mM GuHCl) in the measurements did not destabilize symmetric complexes, as demonstrated by addition of GuHCl to reactions in the presence of αLA (Fig. S4). These results suggest that only non-foldable but not foldable substrate proteins induce the formation of symmetric GroEL:GroES$_2$ complexes.

Link between GroEL ATPase activity and formation of symmetric complexes

It has been suggested that substrate protein shifts GroEL to a state in which both rings hydrolyze ATP independently and bind and release GroES stochastically [16,18,19]. To test this possibility, we measured the steady-state ATPase activity of GroEL in the presence of GroES and the various substrate proteins. While the ability of non-native substrate to stimulate the GroEL ATPase is well established [6,25,28–30], we were surprised to find that both non-foldable substrates, αLA and α-casein, caused an ~2-fold higher ATPase activity than the foldable substrates (Fig. 5). This "over-stimulation" of the ATPase correlated with the formation of symmetric complexes (Fig. 4g), presumably as a result of weakening of the negative allosteric coupling of the two GroEL rings [29]. This would allow both rings to bind and hydrolyze ATP simultaneously. These findings also imply that if foldable substrates were to induce symmetric complexes, they would have a longer half-life than those induced by non-foldable substrates and should have been detected by dcFCCS.

Physiological ADP concentration suppresses formation of symmetric complexes

The negative cooperativity between GroEL rings with respect to ATP binding is known to increase in the presence of ADP, ensuring that the two rings of GroEL are not in phase [31,32]. We hypothesized that physiological concentrations of ADP should therefore reduce the formation of GroEL:GroES$_2$ complexes. The molar ratio of ATP:ADP in *Escherichia coli* in mid-log-phase has been measured at ~10:1, with the ATP concentration being ~3 mM [33]. Using this ratio of ATP:ADP at 1 mM, ATP reduced the proportion of symmetric complexes in the presence of αLA from ~54% to ~27% (Fig. 6a). At an ATP:ADP ratio of 5:1, only a background level of cross-correlation was observed (Fig. 6a). Refolding experiments with a 4-fold excess of DM-MBP over GroEL demonstrated that the chaperonin system in the presence of ATP:ADP was functional in accelerating the spontaneous folding reaction (Fig. 6b). Note that when DM-MBP is in excess over GroEL, the measured folding rate is a combination of accelerated folding in the GroEL/ES cage and slow spontaneous folding in bulk solution. The presence of ADP slightly reduced the flux of DM-MBP through GroEL, reflected in a decrease in overall folding rate, presumably by slowing GroES cycling. These results demonstrate that, at a physiological ATP:ADP ratio, the propensity of GroEL/ES to populate symmetric complexes is strongly diminished and the asymmetric GroEL:GroES complex functions as the folding active form of the chaperonin.

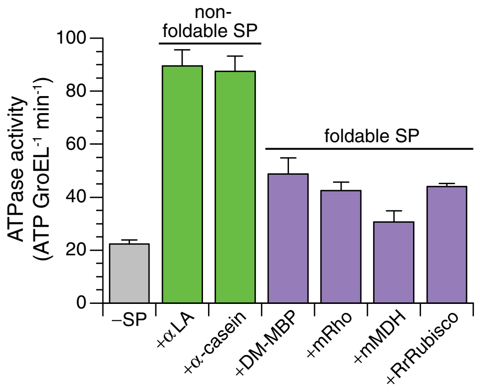


Fig. 5. ATPase activity of GroEL in the presence of non-foldable and foldable substrate proteins. Steady-state rates of ATP hydrolysis of 100 nM GroEL/400 nM GroES in the absence or presence of substrate proteins as in Fig. 4 were measured in buffer A using a coupled enzymatic assay [39]. SP, substrate protein. Averages ± SD from three independent experiments are shown.

Discussion

Here, we have reinvestigated the functional significance of asymmetric and symmetric GroEL/ES complexes during chaperonin-assisted protein folding. We show that recent FRET-based assays

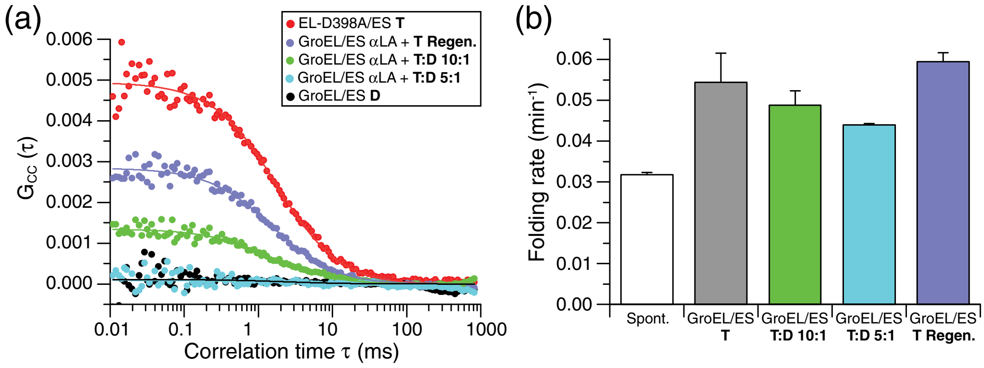


Fig. 6. Effect of physiological ATP:ADP concentration ratio on formation of symmetric GroEL:GroES$_2$ complexes. (a) Cross-correlation signals as in Fig. 4a in the presence of αLA and 1 mM ATP with ATP regenerating system, or 1 mM ATP/0.1 mM ADP (10:1 ratio) or 1 mM ATP/0.2 mM ADP (5:1 ratio). The signals for EL-D398A/ES ATP and for GroEL/ES ADP are shown as controls. (b) GroEL/ES-assisted refolding of DM-MBP in the presence of nucleotide concentrations as in (a). DM-MBP was diluted 200-fold from 6 M GuHCl into buffer A containing 100 nM GroEL/400 nM GroES to a final concentration of 400 nM. Refolding was initiated by the addition of nucleotide and tryptophan fluorescence monitored at 345 nm. The rate of spontaneous refolding (Spont.) of DM-MBP is shown for reference. Folding rates are indicated as averages ± SD from three independent experiments. T, ATP; D, ADP; T Regen., ATP with regenerating system.

used to detect these complexes provide inconsistent results for the proportion of GroEL:GroES$_2$ complexes in the absence of substrate and overestimate their population in the presence of substrate protein. We attribute these problems to the limited sensitivity of the calibrated FRET assay that measures the signal indicative of symmetric complexes over a large background FRET signal (minimally 50%) from asymmetric complexes. Possible adverse effects from fluorophore labeling the apical domains of GroEL, which house the substrate and GroES binding regions, may also be an issue. We therefore developed a novel and robust in-solution dcFCCS assay for GroEL:GroES complex quantification in which labeling of GroEL is avoided and only the symmetric GroEL:GroES$_2$ complex generates the signal, providing superior sensitivity.

Our data demonstrate that symmetric GroEL:GroES$_2$ complexes are not significantly populated in the absence of substrate or presence of foldable substrate proteins, in contrast to recent reports [17–19]. However, non-foldable substrate proteins such as αLA and α-casein can induce the formation of symmetric complexes. This effect of non-foldable substrates correlated with an ~2-fold over-stimulation of the GroEL ATPase rate (in presence of GroES) above that measured with foldable substrates. Thus, the interaction with non-foldable substrates appears to weaken the negative allosteric coupling between the two GroEL rings to an extent that allows them to bind and hydrolyze ATP simultaneously, and accordingly, each ring can bind GroES. The reason for the different behavior of the chaperonin system in the presence of non-foldable and foldable substrate proteins remains to be investigated. The inability of non-foldable proteins to rapidly compact and bury hydrophobic residues upon encapsulation in the GroEL/ES folding cage may play a role in this context [34,35]. Furthermore, experiments performed at physiological ATP:ADP concentration ratios demonstrated that the presence of ADP partially or completely suppressed the ability of non-foldable substrate proteins to induce symmetric GroEL:GroES$_2$ complexes. This is consistent with findings that ADP is a potent allosteric effector of the chaperonin ATPase cycle [31,32]. Specifically, ADP enhances the negative allosteric coupling between GroEL rings, such that they bind and hydrolyze ATP sequentially rather than simultaneously. As a consequence, asymmetric complexes dominate.

Taken together, our results do not support the view that the symmetric GroEL:GroES$_2$ complex is the main or only folding active unit of the chaperonin reaction. Instead, we find that asymmetric complexes dominate both in the absence and in the presence of substrate protein, supporting a model in which the GroEL rings function sequentially in chaperonin-assisted folding.

Materials and Methods

Strains, plasmids and proteins

The *E. coli* strains DH5α and BL21 (DE3) Gold (Stratagene) were used for cloning and protein expression, respectively. GroEL, GroES and chaperonin cysteine mutants were expressed and purified as previously described [23,36]. Purification of substrate proteins DM-MBP and RrRubisco followed published procedures [23,37]. mMDH, mRho, αLA and α-casein were purchased from Sigma.

Preparation of GroEL:substrate protein complexes

mRho, mMDH and RrRubisco were denatured in 6 M GuHCl/5 mM DTT for 1 h at 25 °C. GroEL substrate complexes were prepared by diluting the denatured substrate protein 300-fold into buffer A [20 mM Mops/NaOH (pH 7.4), 200 mM KCl and 10 mM $MgCl_2$]/5 mM DTT containing GroEL (1 μM), with gentle shaking. Aggregated protein was removed by centrifugation at 20,000*g* for 10 min at 15 °C. Denatured substrate protein addition and centrifugation was repeated three times to ensure GroEL saturation with substrate. For ATPase assays and dcFCCS experiments, the GroEL substrate complexes were diluted 10-fold to a final GroEL concentration of 100 nM. The final GuHCl concentration was ∼6 mM. In experiments with the non-aggregating substrate protein DM-MBP, the protein in 6 M GuHCl was diluted 200-fold into buffer A/5 mM DTT to give a 4- to 6-fold excess over GroEL [25]. The non-foldable substrate αLA was unfolded in presence of 5 mM DTT and added in up to 30-fold excess over GroEL, while the non-foldable substrate α-casein, which exposes hydrophobic residues in its native state, was added directly to buffer A/5 mM DTT containing GroEL [28]. All concentrations of GroEL and GroES indicated refer to the oligomeric complexes.

Protein labeling

Purified GroEL(E315C), GroEL(D490C) or GroES(98C) in storage buffer [20 mM Mops/NaOH (pH 7.5), 100 mM NaCl and 10 mM DTT] was buffer exchanged on a NAP5 column (Amersham Biosciences) equilibrated in buffer B [20 mM Tris/HCl (pH 7.5) and 100 mM KCl] and immediately mixed with an equimolar concentration of dye molecules per oligomer to label a single subunit or with the respective excess of dye for labeling multiple subunits and incubated for 30 min at 20 °C in case of GroES(98C) and 15 min in case of GroEL cysteine proteins (which contain three endogenous cysteine residues). Note the endogenous GroEL cysteine residues are not accessible to dye molecules during the labeling time used. The labeling reaction was quenched by addition of 10 mM DTT. Free dye was removed by gel filtration on commercial G25 columns (GE) equilibrated in buffer B/10 mM DTT. The protein was concentrated using Vivaspin concentrators (molecular mass cutoff, 10 kDa; GE Healthcare). The degree of labeling (DOL) was controlled by absorption spectroscopy using the following extinction coefficients (GroEL 14-mer, ε_{280} = 146,020 M^{-1} cm^{-1}; Cy3, ε_{max} = 136,000 M^{-1} cm^{-1} and cf_{280} = 0.08; Cy5, ε_{max} = 250,000 M^{-1} cm^{-1} and cf_{280} = 0.05; F5M, ε_{max} = 83,000 M^{-1} cm^{-1} and cf_{280} = 0.3; IAEDANS, ε_{max} = 5700 M^{-1} cm^{-1} and cf_{280} = 0.27; Atto532, ε_{max} = 115,000 M^{-1} cm^{-1} and cf_{280} = 0.11; Atto655 ε_{max} = 125,000 M^{-1} cm^{-1} and cf_{280} = 0.08) and the equation,

$$\mathrm{DOL} = \frac{A_{\mathrm{dye}} \times \varepsilon_{\mathrm{dye}}}{\left(-\left(A_{\mathrm{dye}} \times cf_{280}\right) + A_{280}\right) \times \varepsilon_{\mathrm{protein}}}$$

In case of GroES labeling due to the low GroES concentration after labeling (low micromolar range) and the low extinction coefficient (GroES 7-mer, ε_{280} = 10,430 M^{-1} cm^{-1}), the protein concentration (c_{protein}) was determined by Bradford assay (Bio-Rad). An unlabeled GroES sample of known concentration (determined photometrically at a high concentration of 100 μM) was used as a standard. The dye concentration was determined by absorption spectroscopy and the DOL determined using the equation:

$$\mathrm{DOL} = \frac{A_{\mathrm{dye}} \times \varepsilon_{\mathrm{dye}}}{c_{\mathrm{protein}}}$$

DOL for a given dye molecule was in all cases >90%. The absence of free dye in the sample was confirmed by fluorescence correlation spectroscopy (measured diffusion curves did not show additional components corresponding to the diffusion time expected for free dye). For calibrated FRET measurements (Fig. 2), GroEL(E315C) contained ∼6 dye molecules (IAEDANS) or ∼1 dye molecule (Cy3) per oligomer and GroES(98C) contained ∼3 dye molecules (F5M or Cy5) per oligomer [16,18]. In the case of cross-correlation measurements, GroES(98C) contained ∼1 dye molecule per oligomer. GroEL(D490C) also contained ∼1 dye molecule per oligomer (Fig. S2).

Ensemble FRET

Fluorescence measurements were performed on a Fluorolog F3-22 spectrofluorometer (Horiba), equipped with a Peltier element set to 25 °C or 37 °C. Scans were taken in increments of 1 nm with an integration time of 0.1 s. GroEL(E315C) (140 nM) labeled with IAEDANS or Cy3 was mixed with GroES(98C) (700 nM) labeled with F5M or Cy5, respectively, in buffer C [50 mM Tris/HCl (pH 7.5), 200 mM KCl, 10 mM $MgCl_2$ and 1 mM DTT]. The mixture was supplemented either with ADP/BeF_X (10 mM KF/1 mM $BeSO_4$/1 mM ADP) to populate asymmetric complexes or with ATP/BeF_X (10 mM KF/1 mM $BeSO_4$/1 mM ATP) to populate symmetric complexes or with 1 mM ATP/ATP regenerating system [4.5 mM phosphoenolpyruvate (PEP)/0.2 U pyruvate kinase (Pyk)] to analyze functional chaperonin cycling. Fluorescence spectra were recorded immediately after nucleotide addition with donor excitation at 336 nm (IAEDANS) or 550 nm (Cy3). To analyze the effect of substrate proteins, we prepared GroEL:substrate complexes as described above. FRET efficiency (f_E) was calculated with the following equation:

$$f_{\mathrm{E}} = 1 - (I_{\mathrm{W}}/I_{\mathrm{WO}})$$

where I_W is the donor fluorescence in the presence of nucleotide and I_{WO} is the donor fluorescence in the absence of nucleotide. To calculate the fraction of symmetric complexes, we used the following equation:

$$\mathrm{Fraction}_{\mathrm{SYM}} = (f_{\mathrm{E}} - f_{\mathrm{E\ ASYM}})/(f_{\mathrm{E\ SYM}} - f_{\mathrm{E\ ASYM}})$$

where $f_{E\ ASYM}$ is the f_E in the presence of ADP/BeF_X (100% asymmetric complexes) and $f_{E\ SYM}$ is the f_E in the presence of ATP/BeF_X (100% symmetric complexes).

Dual-color fluorescence cross-correlation spectroscopy

Correlation spectroscopy was performed on a MicroTime 200 inverse time-resolved fluorescence microscope (PicoQuant) using pulsed interleaved excitation [25,38].

For excitation, picosecond pulsed diode lasers at 640 nm (LDH-PC-640B) and at 530 nm (LDH-P-FA-530) were used at a pulse rate of 27 mHz. As cross-correlation experiments were performed at high concentrations of the fluorophore (400 nM), the laser power was decreased to 5 μW measured before the major dichroic. Reactions in buffer A with additions as indicated in figure legends were measured in a microscopy cuvette (ibidi) that was blocked with 6 mg/ml bovine serum albumin for 1 min and rigorously washed with water to remove unbound bovine serum albumin. The microscope was maintained at a temperature of 20 °C. Measurements were performed for 10 min in case of non-foldable substrate proteins and 1 min in case of foldable substrate proteins. Recorded fluorescence traces were either auto-correlated or cross-correlated. The general function to express auto-correlation or dual-color cross-correlation of fluorescence fluctuation is

$$G_{\mathrm{GR}}(\tau) = \frac{\langle \delta F_{\mathrm{G}}(t) \cdot \delta F_{\mathrm{R}}(t+\tau)\rangle}{\langle F_{\mathrm{G}}(t)\rangle \cdot \langle F_{\mathrm{R}}(t)\rangle}$$

where δF_G and δF_R denote the fluctuation of the signal of green and red fluorescence at time points t and $t + \tau$. In case of auto-correlation of a single color, G = R. The resulting correlated data are approximated by a single diffusion component fit in case of cross-correlation or a diffusion and triplet model in case of auto-correlation (note that, in dcFCCS, blinking and triplet formation is not correlated and does not need to be accounted for):

$$G(\tau) = \rho \cdot \left(1 + \frac{\tau}{\tau_{\mathrm{D}}}\right)^{-1} \cdot \left(1 + \frac{\tau}{\tau_{\mathrm{D}} \cdot \kappa^2}\right)^{-\frac{1}{2}}$$

$$G(\tau) = \left[1 - T + T \cdot e^{\left(-\frac{\tau}{\tau_{\mathrm{T}}}\right)}\right]\left[\rho \cdot \left(1 + \frac{\tau}{\tau_{\mathrm{D}}}\right)^{-1} \cdot \left(1 + \frac{\tau}{\tau_{\mathrm{D}} \cdot \kappa^2}\right)^{-\frac{1}{2}}\right]$$

In both functions, τ denotes the correlation time, τ_D denotes the diffusion time, ρ is the amplitude of the correlation function and κ is the structural parameter that describes the quotient of the axial and radial dimensions of the confocal volume. In case of the latter triplet model, used for fitting of auto-correlation data, T denotes the triplet state fraction on the timescale τ_T. In most cases, only the parameter ρ is of interest as the amplitude of the correlated data in auto-correlation is inversely proportional to the concentration of particles and in case of cross-correlation is directly proportional to the concentration of double-labeled particles:

$$\langle C_{\mathrm{RG}} \rangle = \frac{G_{\mathrm{cc}}(0)}{G_{\mathrm{R}}(0) \cdot G_{\mathrm{G}}(0) \cdot V_{\mathrm{eff}}}$$

The average concentration of double-labeled particles is directly proportional to the amplitude of the cross-correlation function $G_{cc}(0)$ and inversely proportional to the amplitudes of the auto-correlation functions for particles labeled red $G_R(0)$ and green $G_G(0)$. V_{eff} denotes the volume of the focal spot. For constant values of $G_R(0)$ and $G_G(0)$, $G_{cc}(0)$ can be used as a direct measure for changes in complex concentrations.

ATPase assay

The ATPase activity of GroEL (100 nM) was measured in buffer A at 25 °C in presence of GroES (400 nM) with or without substrate proteins as described above. Control reactions were performed in presence of equivalent amounts of denaturant (GuHCl or DTT). ATP hydrolysis was followed using a NADH-coupled enzymatic assay (2 mM PEP, 20 U/ml Pyk, 20 U/ml lactate dehydrogenase, 0.5 mM NADH and 1 mM ATP) at 25 °C in a temperature-controlled spectrophotometer (Jasco) [39].

DM-MBP refolding

Refolding was measured by monitoring the time-dependent increase in the intrinsic tryptophan fluorescence of DM-MBP (excitation: 295 nm; emission: 345 nm) [25]. DM-MBP was unfolded in 6 M GuHCl for 1 h at 25 °C. Denatured DM-MBP was diluted 200-fold into buffer A at 25 °C for spontaneous refolding or into buffer A containing GroEL (100 nM)/GroES (400 nM) for assisted refolding with refolding initiated by addition of either 1 mM ATP, 1 mM ATP/0.1 mM ADP, 1 mM ATP/0.2 mM ADP or 1 mM ATP/4.5 mM PEP/0.2 U Pyk.

Acknowledgements

We acknowledge the expert technical assistance by R. Lange, N. Wischnewski and A. Jungclaus. We thank D. Balchin for discussion and A. Bracher for Fig. 1.

Author Contributions: S.H. and A.J.G. designed and performed FRET and dcFCCS experiments, X.Y. and G.M. performed the ATPase and refolding assays and F.U.H. and M.H.H. advised and supervised the project and wrote the paper with contributions from all authors.

Conflict of Interest Statement: The authors declare that they have no conflict of interest.

Appendix A. Supplementary data

Supplementary data includes four figures. Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jmb.2015.04.009.

Received 18 March 2015;
Received in revised form 15 April 2015;
Accepted 15 April 2015
Available online 23 April 2015

Keywords:
chaperonin;
GroEL;
GroES;
protein folding;
fluorescence correlation spectroscopy

S.H. and A.J.G. contributed equally to this work and share first authorship.

Abbreviations used:

αLA, α-lactalbumin; dcFCCS, dual-color fluorescence cross-correlation spectroscopy; DM-MBP, double mutant of maltose binding protein; DOL, degree of labeling; FRET, fluorescence resonance energy transfer; PEP, phosphoenolpyruvate; Pyk, pyruvate kinase.

References

[1] Hartl FU, Bracher A, Hayer-Hartl M. Molecular chaperones in protein folding and proteostasis. Nature 2011;475:324–32.

[2] Kim YE, Hipp MS, Bracher A, Hayer-Hartl M, Hartl FU. Molecular chaperone functions in protein folding and proteostasis. Annu Rev Biochem 2013;82:323–55.

[3] Saibil HR, Fenton WA, Clare DK, Horwich AL. Structure and allostery of the chaperonin GroEL. J Mol Biol 2013;425: 1476–87.

[4] Horovitz A, Willison KR. Allosteric regulation of chaperonins. Curr Opin Struct Biol 2005;15:646–51.

[5] Langer T, Pfeifer G, Martin J, Baumeister W, Hartl FU. Chaperonin-mediated protein folding: GroES binds to one end of the GroEL cylinder, which accommodates the protein substrate within its central cavity. EMBO J 1992;11:4757–65.

[6] Martin J, Mayhew M, Langer T, Hartl FU. The reaction cycle of GroEL and GroES in chaperonin-assisted protein folding. Nature 1993;366:228–33.

[7] Schmidt M, Rutkat K, Rachel R, Pfeifer G, Jaenicke R, Viitanen P, et al. Symmetric complexes of GroE chaperonins as part of the functional cycle. Science 1994;265:656–9.

[8] Azem A, Kessel M, Goloubinoff P. Characterization of a functional $GroEL_{14}(GroES_7)_2$ chaperonin hetero-oligomer. Science 1994;265:653–6.

[9] Llorca O, Marco S, Carrascosa JL, Valpuesta JM. The formation of symmetrical GroEL-GroES complexes in the presence of ATP. FEBS Lett 1994;345:181–6.

[10] Engel A, Hayer-Hartl MK, Goldie KN, Pfeifer G, Hegerl R, Muller S, et al. Functional significance of symmetrical versus asymmetrical GroEL-GroES chaperonin complexes. Science 1995;269:832–6.

[11] Hayer-Hartl MK, Martin J, Hartl FU. Asymmetrical interaction of GroEL and GroES in the ATPase cycle of assisted protein folding. Science 1995;269:836–41.

[12] Hayer-Hartl MK, Ewalt KL, Hartl FU. On the role of symmetrical and asymmetrical chaperonin complexes in assisted protein folding. Biol Chem 1999;380:531–40.

[13] Rye HS, Burston SG, Fenton WA, Beechem JM, Xu Z, Sigler PB, et al. Distinct actions of *cis* and *trans* ATP within the double ring of the chaperonin GroEL. Nature 1997;388:792–8.

[14] Sparrer H, Rutkat K, Buchner J. Catalysis of protein folding by symmetric chaperone complexes. Proc Natl Acad Sci U S A 1997;94:1096–100.

[15] Rye HS, Roseman AM, Chen S, Furtak K, Fenton WA, Saibil HR, et al. GroEL-GroES cycling: ATP and nonnative polypeptide direct alternation of folding-active rings. Cell 1999;97:325–38.

[16] Sameshima T, Ueno T, Iizuka R, Ishii N, Terada N, Okabe K, et al. Football- and bullet-shaped GroEL-GroES complexes coexist during the reaction cycle. J Biol Chem 2008;283: 23765–73.

[17] Sameshima T, Iizuka R, Ueno T, Funatsu T. Denatured proteins facilitate the formation of the football-shaped GroEL-$(GroES)_2$ complex. Biochem J 2010;427:247–54.

[18] Yang D, Ye X, Lorimer GH. Symmetric $GroEL:GroES_2$ complexes are the protein-folding functional form of the chaperonin nanomachine. Proc Natl Acad Sci U S A 2013; 110:E4298–305.

[19] Ye X, Lorimer GH. Substrate protein switches GroE chaperonins from asymmetric to symmetric cycling by catalyzing nucleotide exchange. Proc Natl Acad Sci U S A 2013;110:E4289–97.

[20] Roseman AM, Chen SX, White H, Braig K, Saibil HR. The chaperonin ATPase cycle—mechanism of allosteric switching and movements of substrate-binding domains in GroEL. Cell 1996;87:241–51.

[21] Taguchi H, Tsukuda K, Motojima F, Koike-Takeshita A, Yoshida M. BeF(x) stops the chaperonin cycle of GroEL-GroES and generates a complex with double folding chambers. J Biol Chem 2004;279:45737–43.

[22] Hayer-Hartl MK, Ewbank JJ, Creighton TE, Hartl FU. Conformational specificity of the chaperonin GroEL for the compact folding intermediates of alpha-lactalbumin. EMBO J 1994;13:3192–202.

[23] Tang YC, Chang HC, Roeben A, Wischnewski D, Wischnewski N, Kerner MJ, et al. Structural features of the GroEL-GroES nano-cage required for rapid folding of encapsulated protein. Cell 2006;125:903–14.

[24] Chakraborty K, Chatila M, Sinha J, Shi Q, Poschner BC, Sikor M, et al. Chaperonin-catalyzed rescue of kinetically trapped states in protein folding. Cell 2010;142:112–22.

[25] Gupta AJ, Haldar S, Milicic G, Hartl FU, Hayer-Hartl M. Active cage mechanism of chaperonin-assisted protein folding demonstrated at single molecule level. J Mol Biol 2014;426: 2739–54.

[26] Georgescauld F, Popova K, Gupta AJ, Bracher A, Engen JR, Hayer-Hartl M, et al. GroEL/ES chaperonin modulates the mechanism and accelerates the rate of TIM-barrel domain folding. Cell 2014;157:922–34.

[27] Koike-Takeshita A, Yoshida M, Taguchi H. Revisiting the GroEL-GroES reaction cycle via the symmetric intermediate implied by novel aspects of the GroEL(D398A) mutant. J Biol Chem 2008;283:23774–81.

[28] Martin J, Langer T, Boteva R, Schramel A, Horwich AL, Hartl FU. Chaperonin-mediated protein folding at the surface of GroEL through a "molten globule"-like intermediate. Nature 1991;352:36–42.

[29] Yifrach O, Horovitz A. Allosteric control by ATP of non-folded protein binding to GroEL. J Mol Biol 1996;255:356–61.

[30] Grason JP, Gresham JS, Widjaja L, Wehri SC, Lorimer GH. Setting the chaperonin timer: the effects of K^+ and substrate protein on ATP hydrolysis. Proc Natl Acad Sci U S A 2008; 105:17334–8.

[31] Kad NM, Ranson NA, Cliff MJ, Clarke AR. Asymmetry, commitment and inhibition in the GroE ATPase cycle impose alternating functions on the two GroEL rings. J Mol Biol 1998; 278:267–78.

[32] Horovitz A, Fridmann Y, Kafri G, Yifrach O. Allostery in chaperonins. J Struct Biol 2001;135:104–14.

[33] Buckstein MH, He J, Rubin H. Characterization of nucleotide pools as a function of physiological state in *Escherichia coli*. J Bacteriol 2008;190:718–26.

[34] Lin Z, Rye HS. Expansion and compression of a protein folding intermediate by GroEL. Mol Cell 2004;16:23–34.

[35] Sharma S, Chakraborty K, Mueller BK, Astola N, Tang YC, Lamb DC, et al. Monitoring protein conformation along the pathway of chaperonin-assisted folding. Cell 2008;133: 142–53.

[36] Tang YC, Chang HC, Chakraborty K, Hartl FU, Hayer-Hartl M. Essential role of the chaperonin folding compartment *in vivo*. EMBO J 2008;27:1458–68.

[37] Brinker A, Pfeifer G, Kerner MJ, Naylor DJ, Hartl FU, Hayer-Hartl M. Dual function of protein confinement in chaperonin-assisted protein folding. Cell 2001;107:223–33.

[38] Muller BK, Zaychikov E, Brauchle C, Lamb DC. Pulsed interleaved excitation. Biophys J 2005;89:3508–22.

[39] Poso D, Clarke AR, Burston SG. A kinetic analysis of the nucleotide-induced allosteric transitions in a single-ring mutant of GroEL. J Mol Biol 2004;338:969–77.

[40] Xu ZH, Horwich AL, Sigler PB. The crystal structure of the asymmetric GroEL-GroES-$(ADP)_7$ chaperonin complex. Nature 1997;388:741–9.

[41] Koike-Takeshita A, Arakawa T, Taguchi H, Shimamura T. Crystal structure of a symmetric football-shaped GroEL:$GroES_2$-ATP_{14} complex determined at 3.8 Å reveals rearrangement between two GroEL rings. J Mol Biol 2014; 426:3634–41.

Supplementary Data

Chaperonin-assisted protein folding: Relative population of asymmetric and symmetric GroEL-GroES complexes

Shubhasis Haldar[†], Amit J. Gupta[†], Xiao Yan, Goran Miličić, F. Ulrich Hartl
and Manajit Hayer-Hartl

Department of Cellular Biochemistry, Max Planck Institute of Biochemistry, Am Klopferspitz 18, 82152 Martinsried, Germany

[†] S.H and A.J.G. contributed equally to this work and share first authorship.

Correspondence to F. Ulrich Hartl and Manajit Hayer-Hartl: *uhartl@biochem.mpg.de; mhartl@biochem.mpg.de*

Supplementary Figures

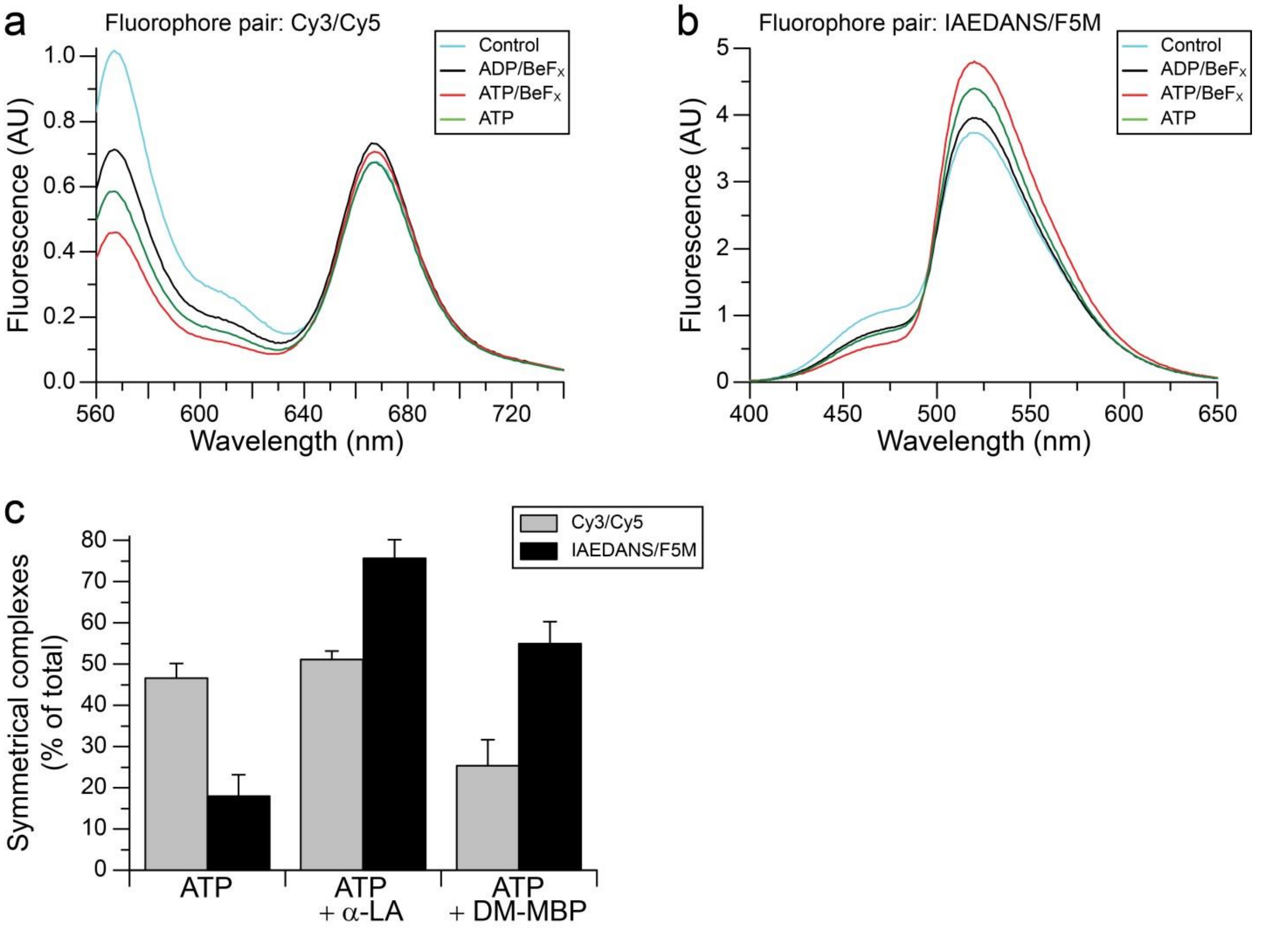


Fig. S1. Analysis of GroEL-GroES complexes by calibrated FRET. The fluorophore pairs Cy3/Cy5 (a) and IAEDANS/F5M (b) were used as FRET reporters [16,17]. 140 nM GroEL(E315C) labeled with IAEDANS or Cy3 and 700 nM GroES(98C) labeled with F5M or Cy5 in buffer C at 37^{o}C was analyzed in the absence of nucleotide (Control) or in the presence of either ADP/BeF$_X$ or ATP/BeF$_X$ or ATP with ATP regenerating system. Fluorescence spectra were recorded immediately after nucleotide addition with donor excitation at 550 nm (a) or 336 nm (b). (c) The proportion of symmetric GroEL:GroES$_2$ complexes was estimated based on the donor fluorescence relative to the donor fluorescence in the presence of ADP/BeF$_X$ (0% symmetric complexes) and in the presence of ATP/BeF$_X$ (100% symmetric complexes) (see Materials and Methods). Donor fluorescence was measured at 565 nm (Cy3) and 465 nm (IAEDANS). The proportion of symmetric complexes was also measured in the presence of a 30-fold excess of non-foldable substrate protein α-LA or a 6-fold excess of foldable DM-MBP. Averages ± SD from three independent experiments are shown.

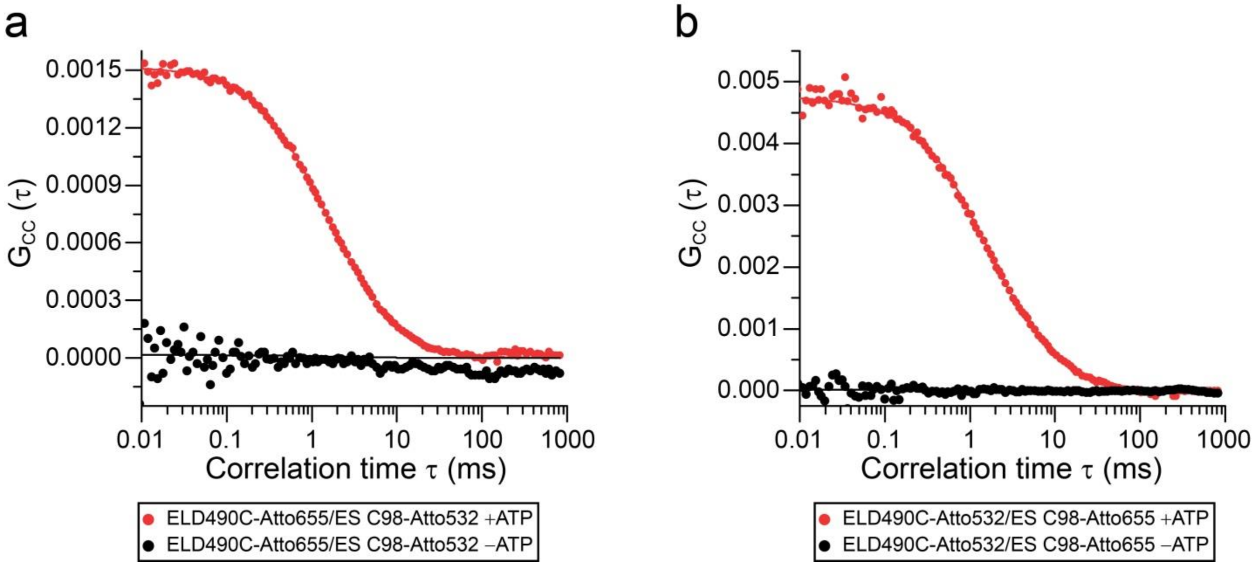


Fig. S2. Fluorophore labeled GroES is competent for GroEL interaction. (a) Cross correlation signal of 100 nM GroEL(D490C) labeled with Atto655 and 400 nM GroES(98C) labeled with Atto532 in absence or presence of 1 mM ATP. (b) Cross correlation signal of 100 nM GroEL(D490C) labeled with Atto532 and 400 nM GroES(98C) labeled with Atto655 in absence or presence of 1 mM ATP. Both experiments were carried out in buffer A and in the presence of ATP regenerating system.

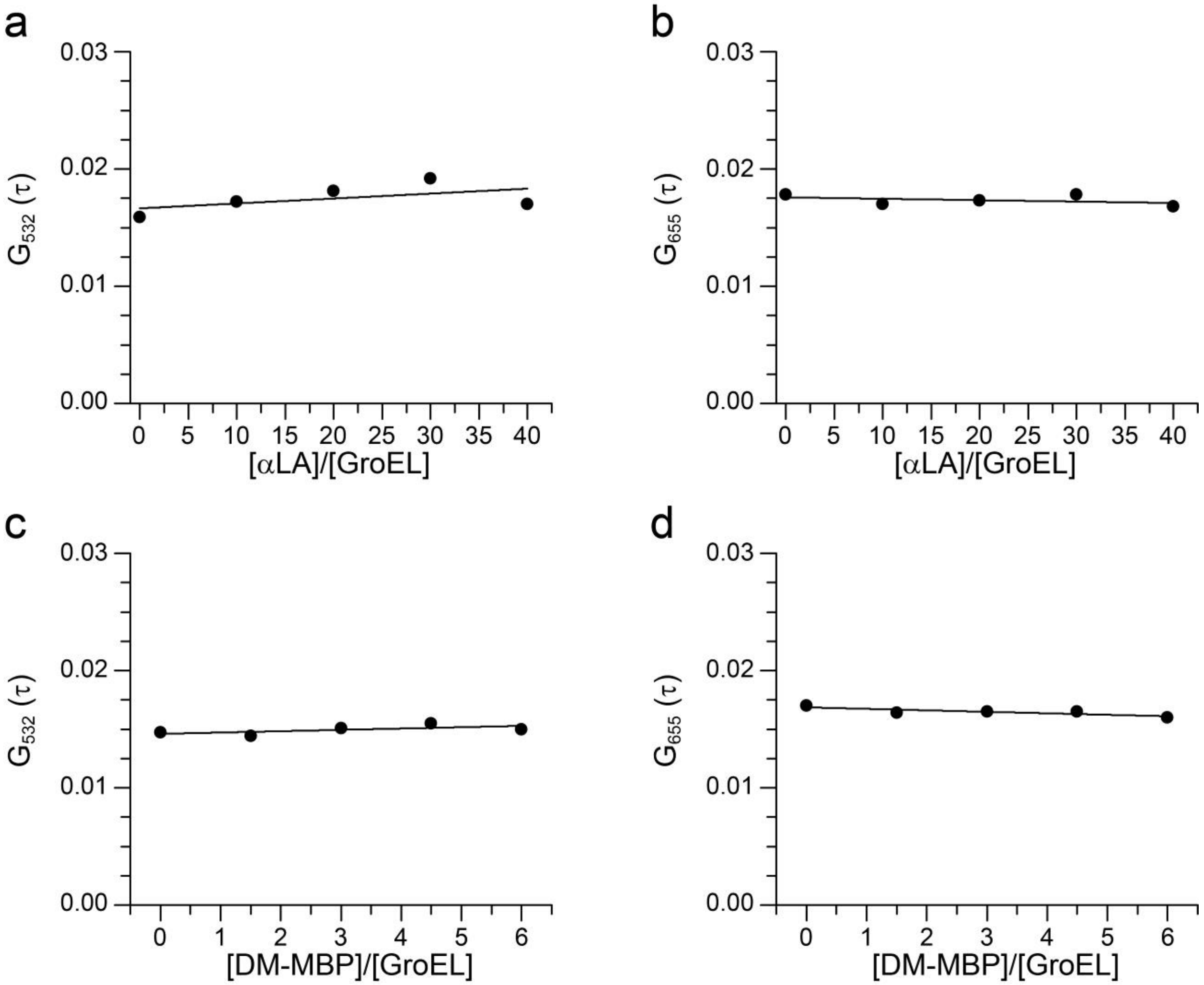


Fig. S3. Presence of substrate protein does not affect dye fluorescence and correlation results. Auto correlation signals of separately labeled GroES (ES532 or ES655, 200 nM each) in the presence 100 nM GroEL and 1 mM ATP and increasing concentrations of substrate proteins αLA and DM-MBP relative to GroEL. αLA and DM-MBP were unfolded in presence of 5 mM DTT and 6 M GuHCl, respectively. Unfolded substrate protein was added in varying amounts to 100 nM GroEL and 400 nM labeled GroES in buffer A including ATP regenerating system. GroEL/ES cycling was initiated by addition of 1 mM ATP. Cross correlation was measured immediately after nucleotide addition for a time window of 1 min. The auto correlation amplitudes of ES532 (a and c) and ES655 (b and d) were approximated by fit and plotted against either αLA (a and b) or DM-MBP (c and d) concentration.

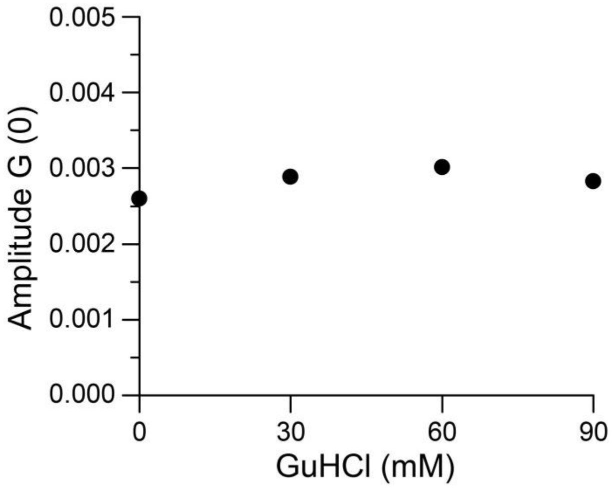


Fig. S4. Residual GuHCl does not destabilize symmetric GroEL:GroES$_2$ complexes. Cross correlation amplitudes of GroES in reactions containing GroEL, ATP and ATP regenerating system were measured as in Fig. 4a in the presence of a 30-fold excess of αLA over GroEL. GuHCl was added at the concentrations indicated.

3.3 Article 3 - Structure and mechanism of the Rubisco-assembly chaperone Raf1

Rubisco is an enzyme that catalyses the incorporation of atmospheric CO_2 into organic carbon during photosynthesis. Rubisco consist of eight large subunits (RbcL) arranged in a tetramer of antiparallel dimers capped on the top and bottom by four small subunits (RbcS)[104]. In this project, we carried out a biochemical and structural characterization of Rubisco accumulation factor 1 (Raf1). We also investigated the role of Raf1 as an assembly chaperone acting downstream of chaperonin-assisted RbcL folding. Our X-ray structure of Raf1 showed that it is a ~40 kDa protein consisting of a β-sheet dimerization domain and a flexibly linked α-helical domain. It functions as a dimer in the $RbcL_8$ assembly process. Chemical crosslinking coupled to mass spectrometry and a negative stain electron microscopy reconstruction showed that the dimerization domain of the Raf1 dimer interacts along the equator of each antiparallel RbcL dimer and that the α-helical domains embrace the top and bottom edges of $RbcL_2$. In such an intermediate, the rebinding of a structurally labile RbcL to GroEL is prevented. We also concluded that Raf1 would bring RbcL subunits into proximity promoting $RbcL_2$ formation. This would lead to an assembly of the larger $RbcL_8$:$Raf1_4$ complex. In the final step, RbcS binding displaces Raf1 and completes the assembly of the holoenzyme.

Nature Structural and Molecular Biology (2015)[144]

Hauser T, Bhat J, **Miličić G**, Wendler P, Hartl FU, Bracher A, Hayer-Hartl M.

Contribution: This project was performed in collaboration with Thomas Hauser, Javaid Bhat, Petra Wendler and Andreas Bracher. I prepared the samples for negative stain electron microscopy, collected micrographs, analysed them and built the 3D reconstruction of the $RbcL_8$-$Raf1_4$ complex.

nature
structural &
molecular biology

Structure and mechanism of the Rubisco-assembly chaperone Raf1

Thomas Hauser[1], Javaid Y Bhat[1], Goran Miličić[1], Petra Wendler[2], F Ulrich Hartl[1], Andreas Bracher[1] & Manajit Hayer-Hartl[1]

Biogenesis of the photosynthetic enzyme Rubisco, a complex of eight large (RbcL) and eight small (RbcS) subunits, requires assembly chaperones. Here we analyzed the role of Rubisco accumulation factor1 (Raf1), a dimer of ~40-kDa subunits. We find that Raf1 from *Synechococcus elongatus* acts downstream of chaperonin-assisted RbcL folding by stabilizing RbcL antiparallel dimers for assembly into $RbcL_8$ complexes with four Raf1 dimers bound. Raf1 displacement by RbcS results in holoenzyme formation. Crystal structures show that Raf1 from *Arabidopsis thaliana* consists of a β-sheet dimerization domain and a flexibly linked α-helical domain. Chemical cross-linking and EM reconstruction indicate that the β-domains bind along the equator of each $RbcL_2$ unit, and the α-helical domains embrace the top and bottom edges of $RbcL_2$. Raf1 fulfills a role similar to that of the assembly chaperone RbcX, thus suggesting that functionally redundant factors ensure efficient Rubisco biogenesis.

Life on earth depends on the ability of photosynthetic organisms to convert atmospheric CO_2 into organic carbon. The key enzyme responsible for this process is ribulose-1,5-bisphosphate carboxylase-oxygenase (Rubisco). Rubisco catalyzes the rate-limiting step of carbon fixation by carboxylating the five-carbon sugar substrate ribulose-1,5-bisphosphate (RuBP). In plants, algae, cyanobacteria and some proteobacteria, Rubisco (form I) is an oligomeric complex of eight large (RbcL, ~50 kDa) and eight small (RbcS, ~15 kDa) subunits[1]. The RbcL subunits are arranged as a tetramer of antiparallel dimers ($RbcL_8$ core) capped by four RbcS subunits at the top and four at the bottom ($RbcL_8S_8$ holoenzyme). Remarkably, Rubisco is a rather inefficient enzyme, fixing only around three to ten carbon atoms per second[2,3]. Moreover, certain Rubiscos confuse CO_2 with molecular oxygen as often as once in four catalytic cycles[4], thus resulting in wasteful photorespiration[5–7]. Hence, engineering a more efficient Rubisco enzyme is central to efforts to increase global food production[8–11].

Although the structure and mechanism of Rubisco are well understood[12,13], knowledge of the chaperone machinery required for Rubisco biogenesis remains limited. As recently demonstrated by *in vitro* reconstitution[14], folding of cyanobacterial RbcL requires the chaperonin GroEL and its cofactor GroES (Cpn60, Cpn10 and Cpn20 in chloroplasts[15]). The spontaneous assembly of folded RbcL is inefficient and has been shown to be facilitated by a specific assembly chaperone, RbcX[14,16]. In several cyanobacteria, RbcX is encoded in an operon between the *rbcL* and *rbcS* genes, and coexpression with RbcX has been shown to enhance the production of active form I Rubisco in *Escherichia coli*[16–21]. RbcX, a homodimer of ~15-kDa subunits, functions as a molecular clamp in stabilizing the antiparallel RbcL dimer and promotes formation of an $RbcL_8$–$RbcX_8$ assembly intermediate[14,22] from which RbcX is displaced by RbcS[14]. However, RbcX is not essential in *S. elongatus* PCC7942 (Syn7942)[23], thus suggesting functional redundancy with other factors. Indeed, a recent screen of photosynthetic mutants in maize identified a nuclear-encoded chloroplast protein, Raf1 (Rubisco accumulation factor 1), that is required for efficient Rubisco biogenesis[24]. Raf1 is conserved in all photosynthetic organisms containing RbcX and functions in Rubisco assembly *in vitro* and *in vivo*[25,26].

Here we set out to functionally and structurally characterize the plant and cyanobacterial Raf1 proteins. We solved the crystal structures of the *A. thaliana* Raf1 domains and analyzed the interaction of Raf1 with RbcL by multiple biochemical and biophysical approaches. Our results show that Raf1 brackets the antiparallel RbcL dimer and stabilizes it in a state competent for assembly of higher oligomers up to $RbcL_8$. Displacement of Raf1 by RbcS leads to formation of the functional holoenzyme. Thus, Raf1 fulfills a role similar to that of RbcX but uses a different structural scaffold and mechanism.

RESULTS

Reconstitution of Raf1 function in Rubisco assembly

A. thaliana (At) contains two Raf1 homologs encoded by the *AT3G04550* (*Atraf1.2*) and *AT5G28500* (*Atraf1.1*) genes, whereas cyanobacteria have only one *raf1* gene. The respective Raf1 proteins share ~38% sequence similarity, and their cyanobacterial homologs are ~55% similar (**Supplementary Fig. 1**). We recombinantly expressed AtRaf1.1 (42.6 kDa) and AtRaf1.2 (43.5 kDa) without their predicted transit peptides, as well as Raf1 from *S. elongatus* PCC7942 (Syn7942-Raf1; 40.2 kDa) and *Synechococcus* sp. PCC7002 (Syn7002-Raf1; 39.9 kDa). The purified Raf1 proteins (**Supplementary Fig. 2a**) formed homodimers, as determined by size-exclusion chromatography

---

[1]Department of Cellular Biochemistry, Max Planck Institute of Biochemistry, Martinsried, Germany. [2]Gene Center Munich, Ludwig-Maximilians-Universität München, Munich, Germany. Correspondence should be addressed to A.B. (bracher@biochem.mpg.de) or M.H.-H. (mhartl@biochem.mpg.de).

Received 22 October 2014; accepted 1 July 2015; published online 3 August 2015; doi:10.1038/nsmb.3062

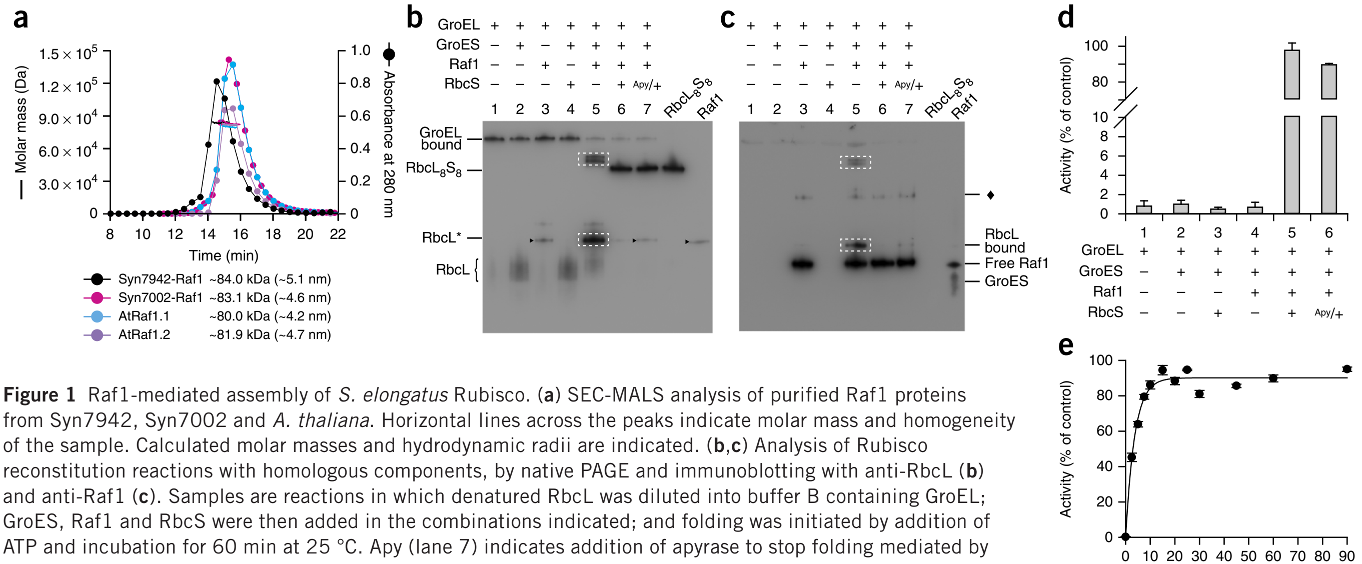


Figure 1 Raf1-mediated assembly of *S. elongatus* Rubisco. (**a**) SEC-MALS analysis of purified Raf1 proteins from Syn7942, Syn7002 and *A. thaliana*. Horizontal lines across the peaks indicate molar mass and homogeneity of the sample. Calculated molar masses and hydrodynamic radii are indicated. (**b**,**c**) Analysis of Rubisco reconstitution reactions with homologous components, by native PAGE and immunoblotting with anti-RbcL (**b**) and anti-Raf1 (**c**). Samples are reactions in which denatured RbcL was diluted into buffer B containing GroEL; GroES, Raf1 and RbcS were then added in the combinations indicated; and folding was initiated by addition of ATP and incubation for 60 min at 25 °C. Apy (lane 7) indicates addition of apyrase to stop folding mediated by GroEL–ES before addition of RbcS. White dashed rectangles mark the positions of the RbcL–Raf1 complex RbcL* and of HMW RbcL–Raf1 complexes. Arrowheads in **b** point to a minor contaminant present in some Raf1 preparations that cross-reacts with anti-RbcL and migrates similarly to RbcL* in native PAGE. In **c**, the position of GroES (70 kDa) is shown, and the filled diamond indicates higher oligomers of Raf1, which are apparently not bound to RbcL. (**d**) Rubisco activity in reactions from **b** after addition of RbcS, which had been absent during reconstitution. Activities are expressed as a percentage of purified $RbcL_8$ standard supplemented with RbcS (control). (**e**) Time course of RbcL-subunit folding. Samples are folding reactions performed in the presence of GroEL–ES and Raf1 as in **b** and stopped with apyrase. Rubisco activities were measured after addition of RbcS. Error bars in **d** and **e** are s.d. (*n* = 3 independent experiments).

combined with static light scattering (SEC-MALS) (**Fig. 1a**). Henceforth, concentrations of Raf1 refer to the dimer.

To analyze the function of Raf1, we performed reconstitution experiments with RbcL, RbcS and Raf1 from *S. elongatus* PCC7942 (Syn7942). Upon folding of denatured RbcL with GroEL–GroES (GroEL–ES) in the presence of ATP, most RbcL remained associated with GroEL, as analyzed by native PAGE (**Fig. 1b**, lanes 1 and 2). Only a small amount of RbcL was released from GroEL, and it migrated as a diffuse band (**Fig. 1b**, lane 2). This species failed to assemble into holoenzyme with RbcS (**Fig. 1b**, lane 4). Addition of Raf1 at equimolar concentration to RbcL resulted in efficient release of RbcL from GroEL–ES and in formation of a well-defined RbcL species (denoted RbcL*) (**Fig. 1b**, lane 5). In addition, we detected small amounts of high-molecular-weight (HMW) complexes migrating above the holoenzyme standard ($RbcL_8S_8$) (**Fig. 1b**, lane 5). Both RbcL* and the HMW forms represented complexes of RbcL with Raf1, as detected by immunoblotting with anti-Raf1 (in which antibody reactivity with Raf1 was reduced in the HMW complexes) (**Fig. 1c**, lane 5). The RbcL–Raf1 complexes formed only in the presence of GroEL and GroES (**Fig. 1b**, lanes 3 and 5), thus indicating that Raf1 acts on folded RbcL. Efficient formation of enzymatically active $RbcL_8S_8$ occurred in the presence of Raf1 and RbcS (**Fig. 1b**, lane 6, and **Fig. 1d**, lane 5). We also observed holoenzyme assembly when RbcS was added after treatment with apyrase, which hydrolyzes ATP to AMP and stops RbcL folding by GroEL–ES (**Fig. 1b**, lane 7, and **Fig. 1d**, lane 6). Thus, Raf1 maintained folded RbcL in an assembly-competent state. RbcL folding occurred with an apparent half-time of ~3 min (**Fig. 1e**), a result consistent with previous measurements[14].

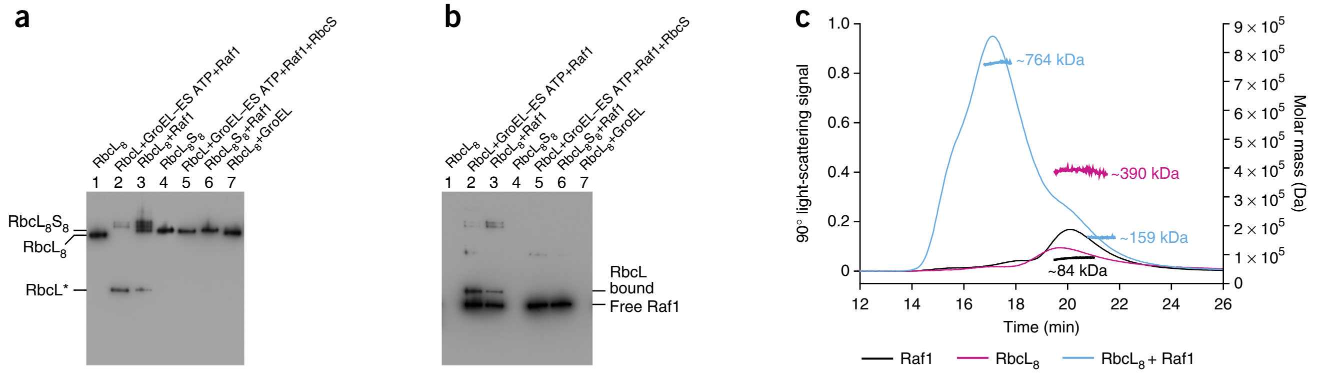
Assembly-competent RbcL* also formed during reconstitution with the heterologous Syn7002-Raf1 or AtRaf1.2 but not with AtRaf1.1 (**Supplementary Fig. 2b**). Reconstitution was ~70% efficient with Syn7002-Raf1 and only ~10% efficient with AtRaf1.2. We obtained no active enzyme with AtRaf1.1 or a heterodimer of AtRaf1.1 and AtRaf1.2 (**Supplementary Fig. 2c**). Thus, Raf1 proteins appear to be adapted to their cognate RbcL proteins, results consistent with those in a recent report[26].

RbcL–Raf1 assembly intermediates

The results described above showed that Raf1 assists in the formation of RbcL assembly intermediates that are competent to form holoenzyme with RbcS. Next we tested whether Raf1 could also bind to preformed $RbcL_8$. Recombinantly produced $RbcL_8$ migrated as a stable complex in native PAGE[16] (**Fig. 2a**, lane 1). Addition of Raf1 resulted in an upshift of $RbcL_8$ indicative of the formation of $RbcL_8$–Raf1 complexes (**Fig. 2a**, lane 3), which are similar to the HMW complexes obtained during reconstitution (**Fig. 2a**, lane 2). RbcS displaced Raf1, thus resulting in holoenzyme formation (**Supplementary Fig. 2d**). Binding of Raf1 to $RbcL_8$ also produced a small amount of RbcL–Raf1 complex that migrated indistinguishably from RbcL* (**Fig. 2a,b**, lanes 3). Raf1 did not interact with $RbcL_8S_8$ (**Fig. 2a,b**, lanes 6), and $RbcL_8$ remained stable in the presence of GroEL (**Fig. 2a,b**, lanes 7).

To determine the relative stoichiometry of RbcL and Raf1 in the RbcL* complex, we excised RbcL* from native PAGE gels and reanalyzed it by SDS-PAGE, which was followed by Coomassie staining and densitometry. Taking into account that RbcL stains approximately two-fold more intensely than Raf1, this analysis suggests that RbcL* is a complex of one Raf1 dimer and two RbcL subunits (**Supplementary Fig. 2e**), consistently with the antiparallel RbcL dimer being the building block of $RbcL_8$.

Next we characterized the HMW complex consisting of RbcL and Raf1 by SEC-MALS and native mass spectrometry (native MS). The molar mass of $RbcL_8$ was ~390 kDa by SEC-MALS and ~420 kDa by native MS, values close to the theoretical mass of 420.4 kDa (**Fig. 2c,d**). Addition of Raf1 to $RbcL_8$ resulted in a complex of ~764 kDa by SEC-MALS and ~741 kDa by native MS (**Fig. 2c,d**), consistently with four Raf1 dimers being bound to $RbcL_8$ (theoretical mass 740.9 kDa). SEC-MALS also detected a species of ~159 kDa,



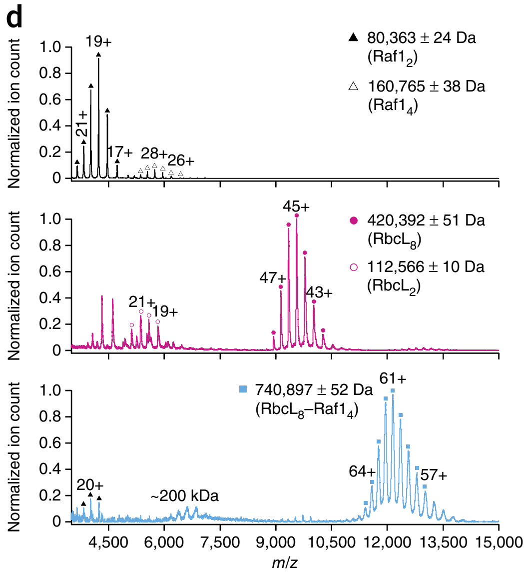


Figure 2 Interaction of Raf1 with preformed $\mathrm{RbcL_8}$ complexes. (**a**,**b**) Complex formation of $\mathrm{RbcL_8}$ with Raf1, analyzed by native PAGE and immunoblotting with anti-RbcL (**a**) and anti-Raf1 (**b**). Samples are purified $\mathrm{RbcL_8}$ or $\mathrm{RbcL_8S_8}$, incubated with Raf1 or GroEL as indicated for 15 min at 25 °C. Reconstitution reactions (as in **Fig. 1b**) in the presence of Raf1 or Raf1 and RbcS were analyzed in parallel. (**c**) SEC-MALS analysis of $\mathrm{RbcL_8}$–Raf1 complex in solution. Samples are purified $\mathrm{RbcL_8}$ incubated with Raf1 for 15 min at 25 °C; $\mathrm{RbcL_8}$ alone; and Raf1 alone. Horizontal lines across the peaks indicate molar mass and homogeneity. Calculated molar masses are indicated. (**d**) Nano-ESI native MS spectra of Raf1, $\mathrm{RbcL_8}$ and $\mathrm{RbcL_8}$–$\mathrm{Raf1_4}$ complex generated by incubation of $\mathrm{RbcL_8}$ with Raf1 for 15 min at 25 °C. Symbols indicate charge-state distributions; charge states are shown for some peaks. The calculated mass around the *m/z* values of the respective protein complexes and the accuracy of mass values calculated from the different *m/z* peaks are indicated.

which is close to the theoretical mass of one Raf1 dimer bound to $\mathrm{RbcL_2}$ (RbcL*; theoretical mass 185.6 kDa) (**Fig. 2c**). We detected an ~200-kDa complex by native MS, but it could not be assigned with confidence (**Fig. 2d**).

These results support a model in which the $\mathrm{RbcL_2}$–Raf1 complex is the first assembly intermediate, and the $\mathrm{RbcL_8}$–$\mathrm{Raf1_4}$ complex is the end state of Raf1-mediated assembly. However, the latter was only a minor species during reconstitution in the absence of RbcS, and the equilibrium was shifted to the $\mathrm{RbcL_2}$–Raf1 complex (**Fig. 2a**,**b**, lane 2). Notably, the yield of active Rubisco decreased when Raf1 was present in excess over RbcL (**Supplementary Fig. 2f**).

Crystal structures of Raf1 domains

Secondary-structure prediction suggested that Raf1 proteins consist of an α-helical N-terminal domain of ~185 residues and a C-terminal β-sheet domain of ~150 residues, connected by a linker of ~10–27 residues (**Fig. 3a**). Although Syn7942-Raf1 failed to crystallize, we obtained crystals of AtRaf1.2 containing either residues 62–275 (Raf1α) or 281–449 (Raf1β), as determined by MS analysis, thus indicating protease sensitivity of the interdomain linker. We cloned and recombinantly expressed the respective α- and β-domains of the different Raf1 homologs for further analysis (**Supplementary Fig. 2a**). The Raf1α domains were monomeric, and the Raf1β domains behaved as dimers (**Supplementary Fig. 3a**). The domains on their own were essentially inactive in the Rubisco reconstitution assay (**Supplementary Fig. 4**).

We solved the crystal structures of AtRaf1.2α and AtRaf1.2β by single isomorphous replacement with anomalous scattering (SIRAS) at resolutions of 1.95 Å and 2.6–2.8 Å, respectively (**Table 1** and **Supplementary Fig. 5a**,**b**). Raf1α is a slightly curved rod with dimensions of 30 × 30 × 70 Å, composed of 11 stacked α-helices, the last three of which form a regular three-helix bundle (**Fig. 3b**). Only one face of the domain exhibits substantial surface-residue conservation among Raf1 homologs; it forms a rather flat surface with hydrophobic and positively charged regions (**Fig. 3c** and **Supplementary Fig. 5c**). Particularly conspicuous is the hydrophobic face of helix 4 (Asn130, Ile133, Val134 and Ser141) and the conserved charged residues (Arg167, Arg174, Arg196, Lys199, Arg228 and Glu232) of the adjacent helices 8 and 9.

We obtained two crystal forms for AtRaf1.2β of space groups *C*2 (2.8-Å resolution) and $P2_12_12_1$ (2.6-Å resolution) (**Table 1**). In both crystal forms, Raf1β exhibited mostly β-structured dimers with pseudo-two-fold symmetry (**Fig. 3d**). The three independent copies were closely similar (r.m.s. deviation 0.469–1.139 Å), except for the domain swapping of a loop in the $P2_12_12_1$ crystal form (**Supplementary Fig. 5d**,**e**). The overall appearance of the dimer was again rod shaped, with dimensions of 30 × 35 × 80 Å, and with the monomer core forming a curved, mixed β-sheet composed of strands A–J–G–F–B (**Fig. 3d**, side view). Satellite β-sheets (strands I–H–J and B and E) branch off the elongated β-strands B and J and curl back onto the central sheet on both sides. Three short α-helices are interspersed between the strands. The two-fold symmetry is broken where the connectors between the first and second β-strand meet, close to the dyad axis (**Fig. 3d** and **Supplementary Fig. 5f**). The dimer interface is formed by the β-hairpin protrusions (strands C and D) from one subunit and β-strands B and E from the other. The interface is highly hydrophobic, and it buries 1,400 $Å^2$ in each subunit (**Supplementary Fig. 5g**). Only the face of the dimer from which the linkers to the α-domains extend is conserved among Raf1 homologs (**Fig. 3e**). This surface is rather flat, and it contains

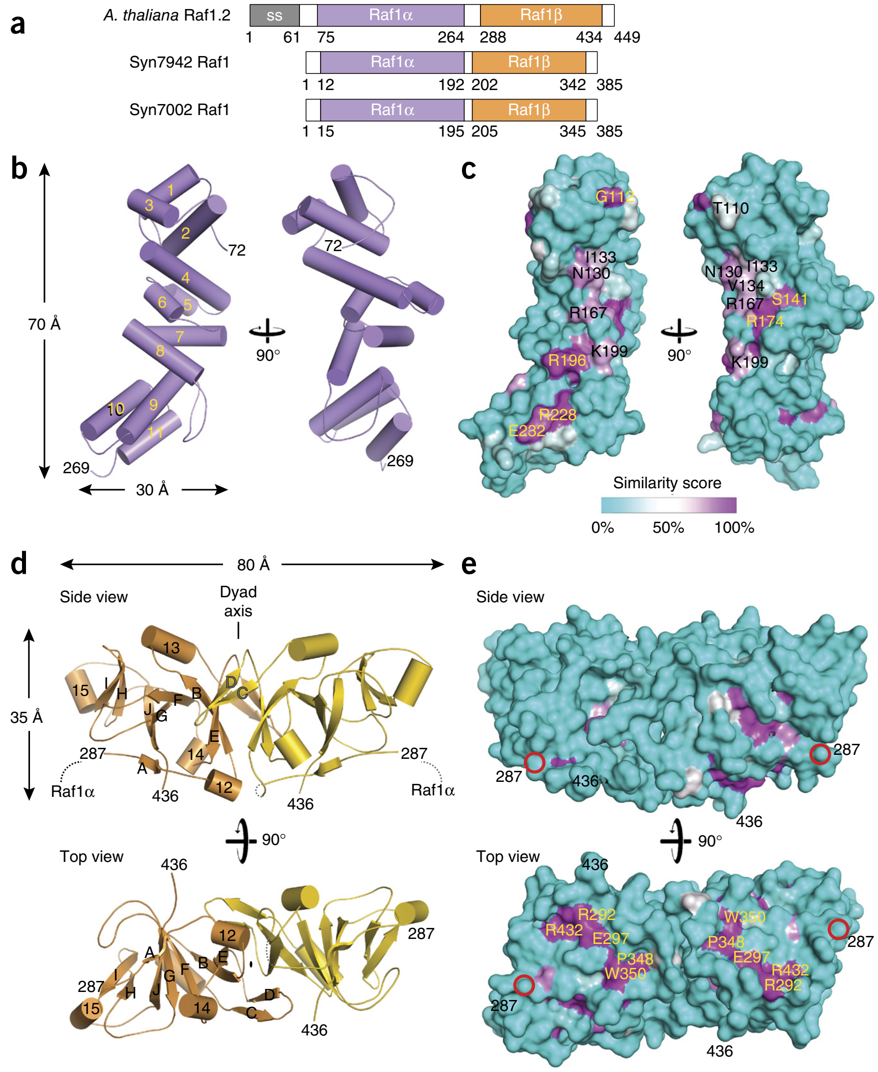


Figure 3 Crystal structures of Raf1 domains. (**a**) Schematic representation of the domain structures of Raf1 from *A. thaliana* (isoform 2; AtRaf1.2), *S. elongatus* PCC7942 (Syn7942) and *Synechococcus* sp. PCC7002 (Syn7002). The predicted chloroplast signal sequence (SS) of AtRaf1.2 is shown in gray. The α-helical and β-sheet domains are shown purple and orange, respectively. Variable linker regions are indicated in white. The domain boundaries for the cyanobacterial Raf1 proteins are based on the sequence alignment in **Supplementary Figure 1**. (**b**) Crystal structure of AtRaf1.2α. Views related by 90° rotation are shown. Helices are represented as cylinders. (**c**) Surface conservation in AtRaf1.2α. AtRaf1.2α is oriented as in **b**. (**d**) Structure of the AtRaf1.2β dimer. The two subunits are shown in orange and yellow. Secondary-structure elements are indicated by numbers for α-helices and by letters for β-strands. The position of the pseudo-two-fold axis is shown. (**e**) Surface conservation in the AtRaf1.2β dimer, analyzed as in **c**.

conserved hydrophobic (Pro348 and Trp350) and charged (Arg292, Glu297 and Arg432) residues in adjacent β-strands A and J. The C-terminal 12 residues were disordered in all crystal lattices.

Small-angle X-ray scattering (SAXS) measurements indicated maximum dimensions (D_{max}) of the isolated domains similar to those derived from the crystal structures (**Supplementary Fig. 3b–d**). The D_{max} of the full-length AtRaf1.2 dimer was ~208 Å, and the radius of gyration (R_g) was 52 Å; we obtained similar values for Syn7942-Raf1. This suggested that the highly charged, flexible linker allows dynamic movements of the Raf1α domains relative to the β-domain dimer, as supported by an ensemble model of the AtRaf1.2 structure (**Supplementary Fig. 3e**).

Mutational analysis of Raf1

Next we performed a mutational analysis to determine the relevance of specific Raf1 residues for interaction with RbcL. Relative to wild-type Raf1, all mutants of the α-domain, except E159A, displayed enhanced formation of HMW $RbcL_8$–Raf1 complexes at the expense of the $RbcL_2$–Raf1 intermediate (**Fig. 4a**). In contrast, the β-domain mutants almost exclusively populated the $RbcL_2$–Raf1 complex, similarly to wild-type (**Fig. 4a**). Although most of the Raf1 mutants supported holoenzyme assembly upon addition of RbcS, two mutants with conserved positive charges in the α-domain (R104Q and K126A K129A, equivalent to Arg174, Arg196 and Lys199 in AtRaf1.2), resulted in a 45–70% reduced yield of activity (**Fig. 4b,c**). Mutant R104Q showed lower amounts of HMW RbcL–Raf1 complex and an increase in diffusely migrating, low-molecular-weight RbcL, presumably representing unassembled protein (**Fig. 4b**). This suggested that decreased binding affinity of Raf1 R104Q for RbcL generated unstable assembly intermediates. We also noted that several of the β-domain mutants resulted in incomplete conversion of $RbcL_2$–Raf1 to $RbcL_8S_8$ (**Fig. 4b**), which correlated with a milder reduction in the yield of active enzyme by 10–20% (**Fig. 4c**).

When taken together, our results suggest that mutations in the α-domain have a more pronounced effect on the binding properties of Raf1 than mutations in the β-domain, consistently with the larger conserved surface of the α-domains. The strength of the RbcL-Raf1 interaction appears to be carefully tuned to allow efficient holoenzyme formation.

Analysis of RbcL-Raf1 interaction by chemical cross-linking

To identify the contact regions between Raf1 and RbcL, we next performed cross-linking coupled to mass spectrometry (CXMS). We added Raf1 to *S. elongatus* $RbcL_8$ to form the $RbcL_8$–$Raf1_4$ complex and then incubated the complex with a 1:1 H_{12} and D_{12} isotopic mixture of the lysine-specific cross-linker disuccinimidylsuberate (DSS) (**Supplementary Fig. 6**). We performed these experiments either with the homologous Syn7942-Raf1 or with the functionally active heterologous Syn7002-Raf1 (**Supplementary Fig. 2b**,**c**), which differ considerably in the number and distribution of lysine residues (**Supplementary Fig. 1**). We separated cross-linked products by SDS-PAGE and analyzed bands >170 kDa, which were likely to contain both Raf1 and RbcL (**Supplementary Fig. 6d**). The median Cα-Cα distance spanned by DSS is ~16.4 Å, with an upper boundary of ~36 Å for the structurally most dynamic regions[27].

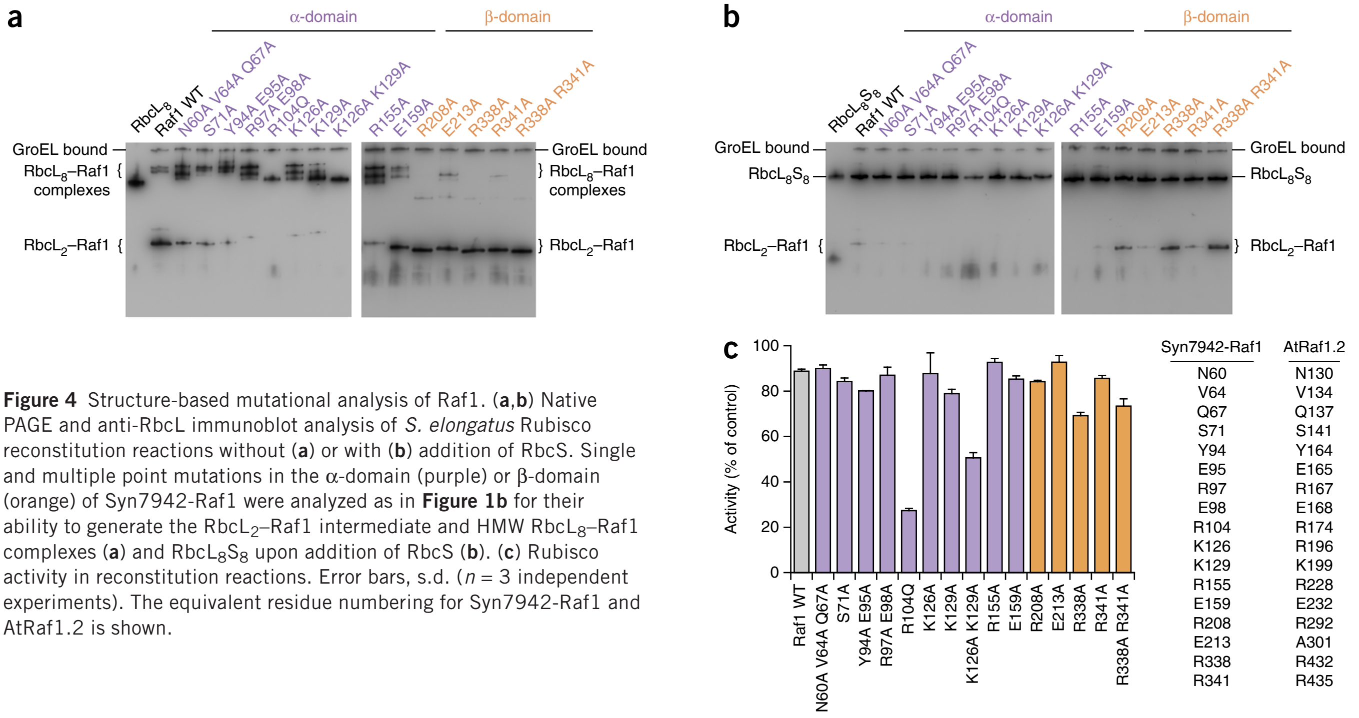
In total, we identified 39 cross-linked peptide pairs in the presence of Syn7942-Raf1 and 45 in the presence of Syn7002-Raf1

Table 1 Data collection and refinement statistics

	AtRaf1.2α native	AtRaf1.2α $K_2[PtCl_4]$	AtRaf1.2β native I	AtRaf1.2β HMBA-I	AtRaf1.2β native II
Data collection					
Space group	$P4_12_12$	$P4_12_12$	$C2$	$C2$	$P2_12_12_1$
Cell dimensions					
a, *b*, *c* (Å)	29.77, 29.77, 457.09	29.36, 29.36, 454.86	157.54, 34.36, 106.89	155.72, 34.70, 115.86	39.68, 60.79, 143.27
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 93.67, 90	90, 97.84, 90	90, 90, 90
		Peak		*Peak*	
Wavelength		1.072		1.009	
Resolution (Å)[a]	45.71–1.95 (2.06–1.95)[a]	45.49–2.9 (3.05–2.9)	45.51–2.8 (2.96–2.8)	49.38–3.4 (3.58–3.4)	47.76–2.57 (2.71–2.57)
R_{merge}	0.068 (1.067)	0.158 (0.706)	0.044 (0.697)	0.102 (0.941)	0.097 (0.909)
$I / \sigma I$	20.2 (2.1)	10.9 (2.9)	20.0 (1.9)	19.4 (2.6)	11.9 (2.1)
Completeness (%)	100 (99.8)	99.7 (98.2)	99.2 (96.5)	99.7 (98.3)	96.7 (81.8)
Redundancy	13.3 (12.6)	14.4 (12.9)	3.6 (3.5)	11.0 (11.2)	4.3 (4.2)
Refinement					
Resolution (Å)	30–1.95	–	30–2.8	–	30–2.57
No. reflections	16,514	–	13,676	–	10,713
R_{work} / R_{free}	0.211 / 0.246	–	0.240 / 0.289	–	0.210 / 0.279
No. atoms					
Protein	1,630	–	4,443	–	2,282
Phosphates[b]	–	–	15	–	–
Water	60	–	–	–	9
B factors					
Protein	64.52	–	94.61	–	60.44
Phosphates	–	–	138.06	–	–
Water	51.42	–	–	–	45.77
r.m.s. deviations					
Bond lengths (Å)	0.011	–	0.004	–	0.008
Bond angles (°)	1.210	–	1.063	–	1.272

[a]Values in parentheses are for highest-resolution shell. [b]From precipitant.

(**Supplementary Data Set 1a–d** and **Supplementary Note**). We were able to assign plausible intra- or intermolecular distances of 7 to 20 Å for 12 out of the 29 RbcL-RbcL cross-links, on the basis of the $RbcL_8S_8$ crystal structure[28]. The remaining 17 RbcL-RbcL cross-links could not be assigned distances because one or both of the cross-linked residues are located at the flexible N or C terminus. We identified



Syn7942-Raf1	AtRaf1.2
N60	N130
V64	V134
Q67	Q137
S71	S141
Y94	Y164
E95	E165
R97	R167
E98	E168
R104	R174
K126	R196
K129	K199
R155	R228
E159	E232
R208	R292
E213	A301
R338	R432
R341	R435

Figure 4 Structure-based mutational analysis of Raf1. (**a**,**b**) Native PAGE and anti-RbcL immunoblot analysis of *S. elongatus* Rubisco reconstitution reactions without (**a**) or with (**b**) addition of RbcS. Single and multiple point mutations in the α-domain (purple) or β-domain (orange) of Syn7942-Raf1 were analyzed as in **Figure 1b** for their ability to generate the $RbcL_2$–Raf1 intermediate and HMW $RbcL_8$–Raf1 complexes (**a**) and $RbcL_8S_8$ upon addition of RbcS (**b**). (**c**) Rubisco activity in reconstitution reactions. Error bars, s.d. ($n = 3$ independent experiments). The equivalent residue numbering for Syn7942-Raf1 and AtRaf1.2 is shown.

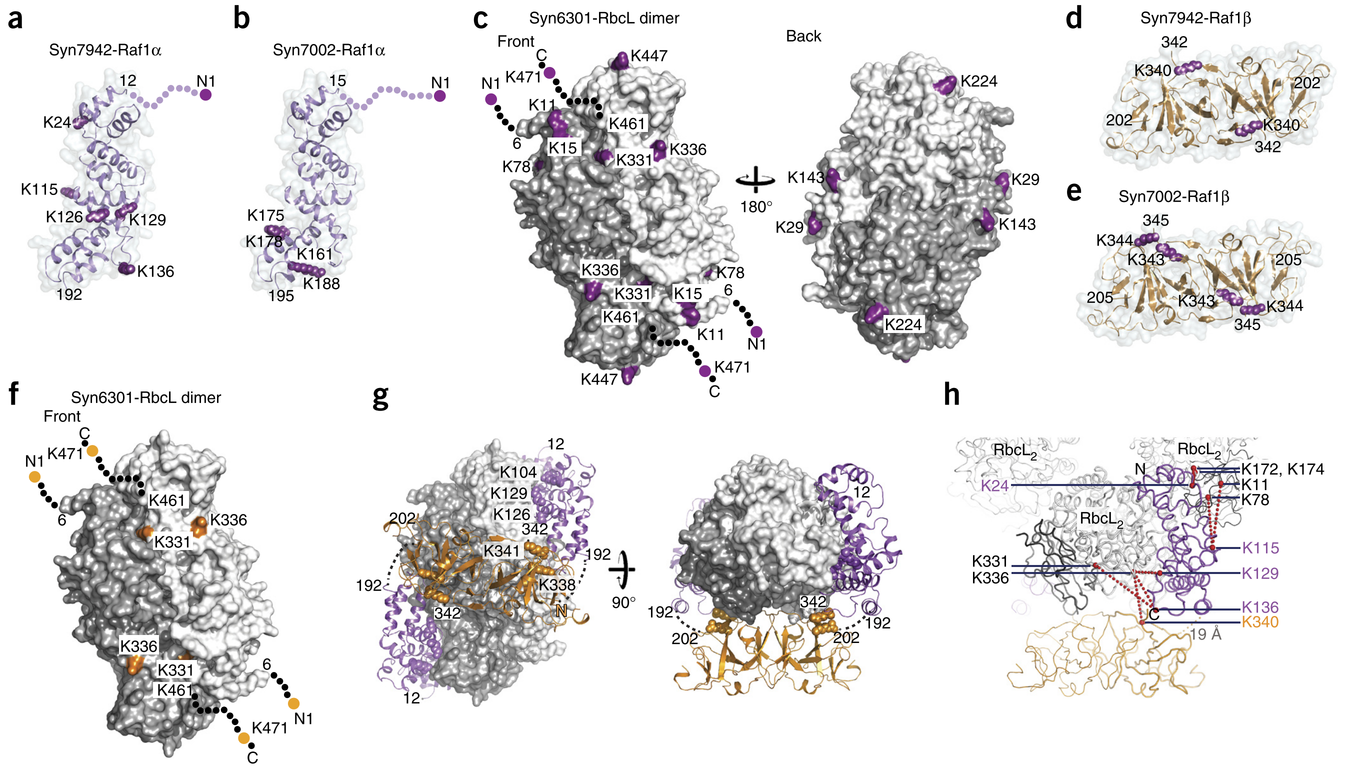


Figure 5 Probing the RbcL–Raf1 complex by chemical cross-linking. (**a**,**b**) Cross-linking sites in the Raf1 α-domains of Syn7942-Raf1 and Syn7002-Raf1. Cross-linked lysine residues and the N-terminal amino group are shown in space-filling representation. The backbone is shown in ribbon representation underneath a translucent molecular surface. Disordered residues are indicated by dots. The orientation corresponds to **Figure 3b**, left. (**c**) Cross-linking sites from Raf1α on the surface of the RbcL dimer. The model is based on the crystal structure of Rubisco[28]. The RbcL subunits are shown in white and gray. Residues Lys172 and Lys174 of RbcL cross-linked to Raf1α are not indicated because they are partially buried in the RbcL dimer interface, and Lys161 is pointing inwards. (**d**,**e**) Cross-linking sites in the Raf1β domains of Syn7942-Raf1 and Syn7002-Raf1, displayed in the same style as in **a** and **b**. Top views are shown. (**f**) Cross-linking sites from Raf1β on the surface of the RbcL dimer. (**g**) Tentative model of the $RbcL_2$–Raf_1 protomer of the $RbcL_8$–$Raf1_4$ complex, based on the cross-linking data. The RbcL dimer is depicted as above; Raf1 is in ribbon representation. Functionally critical Raf1 residues are shown in space-filling representation. (**h**) Cross-links mapped onto the $RbcL_8$–$Raf1_4$ complex. Dotted red lines indicate plausible cross-links between lysine residues of RbcL subunits and Raf1.

31 cross-links between RbcL and Raf1, 19 involving the α-domain, 6 involving the β-domain, 4 involving the flexible N terminus and 2 involving the interdomain linker of Raf1 (**Supplementary Data Set 1b**,**d** and **Supplementary Note**).

According to homology models for the Syn7942-Raf1 and Syn7002-Raf1 domains, most of the α-domain cross-links originated from the periphery of the conserved surface, and we found them multiple times (**Fig. 3c**, **Supplementary Data Set 1b**,**d** and **Supplementary Note**). They connected to the N domain of RbcL, including the flexible N terminus and residues Lys11, Lys15, Lys29, Lys78 and Lys143, as well as to the TIM-barrel domain (Lys161, Lys172, Lys174, Lys224, Lys331 and Lys336) and the C-terminal domain (Lys447 and Lys471) (**Fig. 5a**–**c**, **Supplementary Data Set 1b**,**d** and **Supplementary Note**). The cross-links from the β-domain also involved lysines near its conserved top surface (**Figs. 3e** and **5d**,**e**). The Syn7942-Raf1β domain has only one lysine (Lys340), which cross-linked repeatedly to Lys336 on RbcL (**Fig. 5d**,**f**, **Supplementary Data Set 1b** and **Supplementary Note**). The β-domain of Syn7002-Raf1 has six lysines, and of these Lys343 cross-linked exclusively to Lys336 on RbcL, whereas Lys344 cross-linked to either Lys331 or Lys336 on RbcL (**Fig. 5e**,**f** and **Supplementary Data Set 1d** and **Supplementary Note**). Both the α- and β-domains cross-linked to residues Lys331 and Lys336 of RbcL (**Supplementary Data Set 1b**,**d** and **Supplementary Note**), thus suggesting that the Raf1 domains are in proximity, in accordance with the limited linker length of ten residues.

Structural model of the RbcL-Raf1 interaction

We obtained plausible distances for the cross-links between the β-domain and RbcL (19–21 Å) by positioning the β-domain dimer coaxially at the equatorial front face of each RbcL antiparallel dimer (**Fig. 5g**,**h**, **Supplementary Data Set 1b**,**d** and **Supplementary Note**). In contrast, positioning the β-domain dimer on the two-fold axis between RbcL dimers resulted in substantially increased cross-link distances (43–67 Å; data not shown), and thus such a topology seems less plausible. Indeed, we found similar cross-links during folding and assembly (**Supplementary Data Set 1e**,**f** and **Supplementary Note**), where mostly the $RbcL_2$–Raf1 intermediate was populated (**Fig. 2a**,**b**). From the equatorial positioning of the β-domains, and taking the length of the α-β domain linker into account, we obtained optimal distances for the cross-links between the α-domain and RbcL (**Supplementary Data Set 1b**,**d** and **Supplementary Note**) by placing the α-domains so that they embraced the top and bottom edges of each $RbcL_2$ unit. The shallow groove in the conserved surface of the α-domain (**Fig. 3c**) would cradle the back of the C-terminal domain of RbcL, consistently with mutations in this region weakening the interaction with RbcL (**Fig. 5g** and **Fig. 4**). In addition, the α-domains are also within cross-linking distance to the adjacent $RbcL_2$ unit in the $RbcL_8$–$Raf1_4$ complex (**Fig. 5h**). The resulting tentative model for the $RbcL_8$–$Raf1_4$ complex (**Fig. 5g**,**h**) is consistent with a role of Raf1 in stabilizing $RbcL_2$ and allowing its assembly into the $RbcL_8$ core complex. In the structure of $RbcL_8$ determined by cryo-EM, the ~60 C-terminal residues of RbcL

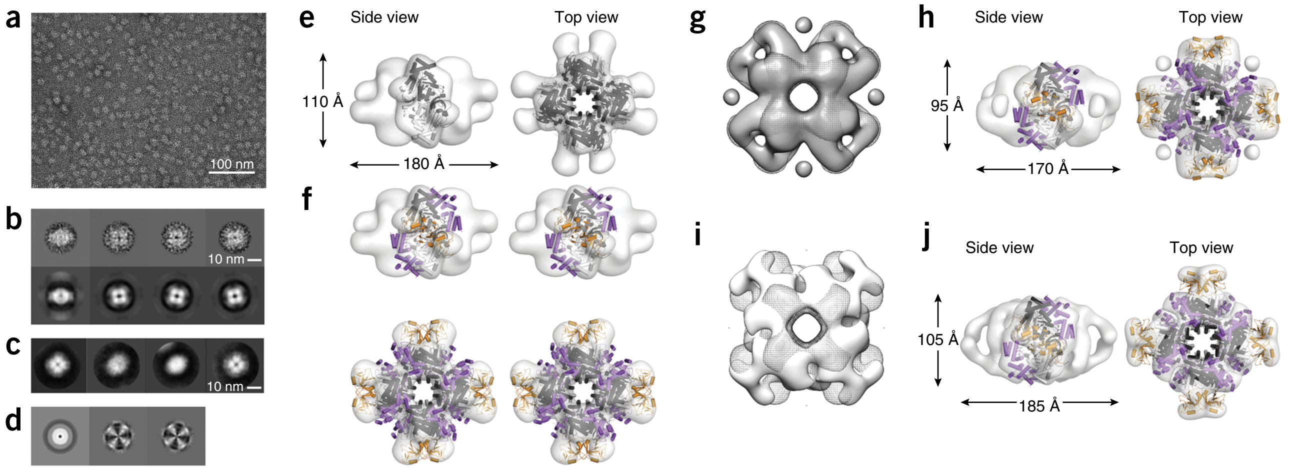


Figure 6 Negative-stain EM and 3D reconstructions of RbcL$_8$–Raf1$_4$ complex. (**a**) Micrograph of negatively stained complexes of *S. elongatus* RbcL$_8$ and Syn7942-Raf1 complex. (**b**) Class averages of the complexes derived from multivariate statistical analysis (MSA) in IMAGIC (upper row) and corresponding reprojections (bottom rows) of the initial 3D reconstruction. Each class average contains ~30 particles. (**c**) Class averages of the complexes from two-dimensional classification in RELION-1.3. (**d**) Eigenimages derived from MSA of top views in IMAGIC. (**e**) Rigid-body domain fitting of *S. elongatus* RbcL$_8$ into the final 3D reconstruction of RbcL$_8$–Raf1$_4$. Side and top views are shown. RbcL subunits are in gray and black. (**f**) Rigid-body domain fitting of *S. elongatus* RbcL$_8$ and the Syn7942-Raf1 α- and β-domains into the final 3D reconstruction of RbcL$_8$–Raf1$_4$. Side and top views are shown in stereo views. Gray and black, RbcL subunits; purple, Raf1α; orange, Raf1β. (**g**) Negative-stain EM density of *S. elongatus* RbcL$_8$–Raf1$_4$ (dark mesh) overlaid on EM density of the same cross-linked complex (dark gray surface) in top view; the contour level is set to enclose 740 kDa. (**h**) Structural model of RbcL$_8$ and the Syn7942-Raf1 α- and β-domains docked into the cross-linked *S. elongatus* RbcL$_8$–Raf1$_4$ EM map shown as in **f**. (**i**) Negative-stain EM density of *S. elongatus* RbcL$_8$–Raf1$_4$ (dark mesh) overlaid on EM density of the *T. elongatus* RbcL$_8$–Raf1$_4$ complex (white surface) in top view. (**j**) Structural model of RbcL$_8$ and the Syn7942-Raf1 α- and β-domains docked into the *T. elongatus* RbcL$_8$–Raf1$_4$ EM map shown as in **f**.

are disordered but are ordered in the complex with the assembly chaperone RbcX[14,22]. Thus, it seems plausible that the C-terminal residues of RbcL are also ordered in the complex with Raf1.

To obtain additional structural information, we next performed negative-stain EM and single-particle image analysis of the *S. elongatus* RbcL$_8$–Raf1$_4$ complex. Our reference-free analysis of two-dimensional class averages (7,602 particles) revealed fourfold symmetry in top views (**Fig. 6a–c**) and eigenimages from end views of the complex (**Fig. 6d**). Further analysis of a final data set of ~6,200 particles resulted in a 25-Å density map when dihedral fourfold symmetry was imposed (**Fig. 6e** and **Supplementary Fig. 7a**). When rendered at a threshold of 740-kDa mass, the particle dimensions were 110 × 180 Å. Compared to RbcL$_8$ in the holoenzyme crystal structure[28], the RbcL$_8$–Raf1$_4$ complex was ~10 Å taller and ~70 Å wider. However, it was ~35 Å taller than the cryo-EM structure of RbcL$_8$ alone, in which the C-terminal ~60 residues of RbcL are disordered[14].

To determine the position of Raf1 in the complex, we first fitted the RbcL$_8$ core from the holoenzyme structure into the density with Chimera[29] (**Fig. 6e**). The Raf1β dimer was docked into the protruding densities at the equatorial rim, and the α-domains were placed into additional densities at the top and bottom edges of RbcL$_2$ units (**Fig. 6f**). The cross-links mapped on this model with plausible distances (**Supplementary Data Set 1b** and **Supplementary Note**). Overall the EM- and CXMS-derived structural models are in reasonable agreement (**Supplementary Fig. 7b**) with the differences in domain orientations, thus possibly reflecting the dynamic nature of the interaction. We note that we were able to obtain a somewhat improved fit into the density when assuming that the ~60 C-terminal residues in RbcL are disordered (**Supplementary Fig. 7c**, **Supplementary Data Set 1b** and **Supplementary Note**). However, in this model the conserved surface of Raf1α faced the solvent, and the distance between the Raf1 α- and β-domains would be beyond the contour length of the linker residues.

To potentially stabilize the RbcL$_8$–Raf1$_4$ complex, we performed the EM analysis after DSS cross-linking (**Supplementary Fig. 7a**,**d**). The three-dimensional (3D) reconstruction obtained from 5,183 particles resembled the non-cross-linked complex, with a cross-correlation coefficient of 0.9834 (**Fig. 6g**). Although the cross-linked complex was slightly smaller (95 × 170 Å), and the Raf1β dimer was stabilized (**Fig. 6h**), cross-linking did not reduce heterogeneity in the EM data set (**Supplementary Fig. 7d**), thus suggesting that structural dynamics is an intrinsic functional property of the RbcL$_8$-Raf1 interaction.

To further validate the EM structural model, we analyzed the RbcL–Raf1 complex of the thermophilic cyanobacterium *Thermosynechococcus elongatus* (**Supplementary Fig. 7a**,**e**). The purified complex, obtained upon coexpression of *T. elongatus* RbcL and Raf1 in *E. coli*, again contained four Raf1 dimers bound to RbcL$_8$ (**Supplementary Fig. 8**). The EM density of the *T. elongatus* RbcL$_8$–Raf1$_4$ complex was closely similar to that of *S. elongatus* (cross-correlation coefficient of 0.9742) (**Fig. 6i**), although its central pore was slightly wider (**Fig. 6j**).

Together the cross-linking data and the EM reconstructions support a model in which Raf1 brackets the RbcL antiparallel dimer, thus stabilizing it in a state competent for assembly to higher oligomers up to RbcL$_8$–Raf1$_4$.

DISCUSSION

Assembly of oligomeric protein complexes is widely considered to be a spontaneous process, and relatively little is known about the machineries that support the formation of specific multiprotein complexes. The biogenesis of hexadecameric Rubisco has emerged as a paradigm of assisted assembly[30,31]. Here we analyzed the structure

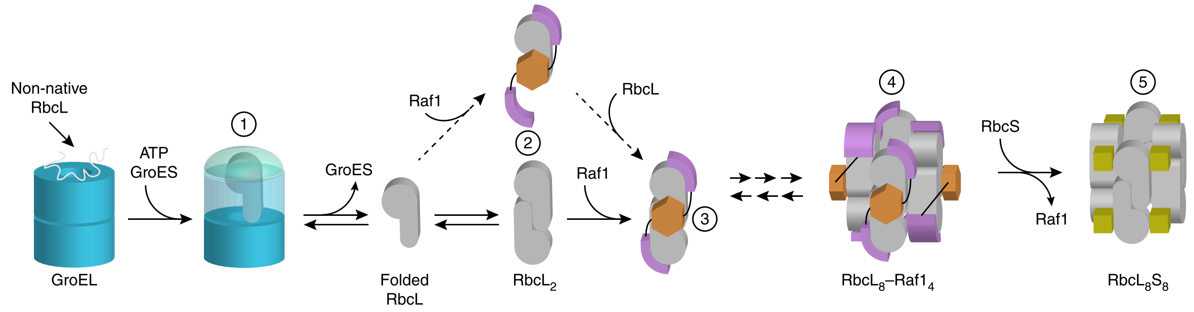


Figure 7 Model of Raf1-mediated Rubisco assembly. After folding by GroEL–ES (1), the released RbcL subunits either dimerize spontaneously or do so with assistance by Raf1 (2). The antiparallel RbcL dimer is stabilized by Raf1 (3) and is in dynamic equilibrium with higher oligomers up to RbcL$_8$–Raf1$_4$ complexes (4). The RbcL–Raf1 complex is dynamic, thus allowing RbcS binding to shift the equilibrium toward holoenzyme formation (5).

and mechanism of the Rubisco-assembly chaperone Raf1 (ref. 24). Our results demonstrate that the dimeric Raf1 functions downstream of RbcL-subunit folding by the GroEL–ES chaperonin system. The major intermediate populated during assembly is a complex in which Raf1 brackets the antiparallel RbcL dimer (**Fig. 7**), thus preventing rebinding of the structurally labile RbcL to GroEL. It is also possible that Raf1 brings RbcL subunits into proximity, thereby promoting dimer formation. The RbcL$_2$–Raf1 complex is competent for assembly into higher oligomeric states, with the RbcL$_8$–Raf1$_4$ complex as the endpoint (**Fig. 7**). Finally, binding of RbcS displaces Raf1 and completes assembly of the holoenzyme.

Like the structurally unrelated Rubisco-assembly chaperone RbcX[16,32,33], Raf1 is dimeric and engages in bivalent interactions with RbcL, a principle that probably relates to the antiparallel RbcL dimer being the building block of the RbcL$_8$ complex. As shown by X-ray crystallographic analysis, Raf1 consists of an N-terminal α-domain, a flexible linker segment and a C-terminal β-sheet domain that mediates dimerization. Both domains expose conserved interaction surfaces. From crystal structures, biochemical analysis, chemical cross-linking and negative-stain EM reconstruction, we propose a model of the RbcL$_8$–Raf1$_4$ assembly intermediate. In this model, the α-domains mediate the majority of functionally important contacts with RbcL by bracketing each RbcL dimer at the top and bottom, and the β-domain dimers are positioned coaxially in front of the RbcL$_2$ units. Because the α-domain alone is essentially inactive, dimerization is critical for Raf1 to achieve the necessary avidity for complex formation with RbcL assembly intermediates. At the same time, the interaction must remain dynamic in order to allow RbcS binding, because the Raf1 α-domain and RbcS have overlapping binding sites on RbcL.

Interestingly, the structurally distinct Raf1 and RbcX assembly proteins perform the same function, but they do so by using different interaction sites on RbcL. The boomerang-shaped RbcX dimer binds the C-terminal peptide tail of one RbcL subunit in a central cleft and contacts the N-terminal domain of the adjacent RbcL subunit via a peripheral region[16,22]. In contrast, the Raf1-RbcL interaction surface appears to be more extensive. An interesting question therefore concerns whether Raf1 and RbcX act in parallel assembly pathways or functionally cooperate *in vivo* to achieve efficient assembly at a biologically relevant timescale. Taking into consideration that the binding sites for RbcX and RbcS have little if any overlap[22], the previously described RbcL$_8$–RbcX$_8$ complex[14,22] may be an additional assembly intermediate, especially when insufficient RbcS is present. A cooperation of Raf1 and RbcX, rather than mere functional redundancy, would be consistent with the strict co-occurrence of the two factors in photosynthetic organisms. Given that the RbcL subunits of all form I Rubiscos are highly homologous, it is surprising that the Rubisco of higher plants has so far been refractory to reconstitution in *E. coli* or *in vitro*. Perhaps additional factors are required for the assembly of the plant enzyme[24,34,35].

The chaperone requirement for the folding and assembly of hexadecameric (form I) Rubisco is remarkably complex, thus raising the question of why such an important and abundant enzyme has failed to evolve a more robust biogenesis pathway. Form I Rubisco evolved from the simpler dimeric form II enzyme about 2.5 billion years ago to adapt to the increasing levels of oxygen[36]. Both the form I and form II RbcL subunits require the GroEL–ES chaperonin for folding, like many other TIM-barrel proteins[37,38]. However, only the RbcL of form I Rubisco remains structurally labile after folding and tends to rebind to chaperonin rather than to spontaneously assemble. The dependence on assembly chaperones, in turn, is likely to have slowed the evolution of form I Rubisco. Recent attempts to improve the enzymatic properties of the enzyme by directed evolution have shown that although the chaperonin tends to increase the number of permissible mutations, the assembly chaperone RbcX retards evolvability[11]. It remains to be seen whether Raf1 confers similar evolutionary constraints.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Coordinates for the structural domains of the protein Raf1.2 from *A. thaliana* have been deposited in the Protein Data Bank under accession codes 4WT3, 4WT4 and 4WT5. Negative-stain EM maps of the *S. elongatus* and *T. elongatus* RbcL$_8$–Raf1$_4$ complexes have been deposited in the Electron Microscopy Data Bank under accession codes EMD-3051 (*T. elongatus* RbcL$_8$–Raf1$_4$), EMD-3052 (cross-linked *S. elongatus* RbcL$_8$–Raf1$_4$) and EMD-3053 (*S. elongatus* RbcL$_8$–Raf1$_4$).

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

A. thaliana cDNA was a kind gift from B. Bölter (Ludwig-Maximilians-Universität München), and *S. elongatus* PCC7942 DNA was a kind gift from M. Hagemann (Universität Rostock). We thank A. Jungclaus for assistance with protein purification and R. Körner for performing the MS on the cross-linked samples. We also thank A. Sinz (Martin Luther University Halle-Wittenberg) for making available StavroX software for the cross-linking MS analysis. Assistance by K. Valer and J. Basquin at the Max Planck Institute of Biochemistry (MPIB) crystallization facility, as well as the staff at Swiss Light Source beamlines X10SA-PX-II and X06DA-PXIII, is gratefully acknowledged. SAXS data were collected at European

Synchrotron Radiation Facility beamline BM29 with the assistance of A. Round. Cryo-EM analysis was performed in the Department of W. Baumeister at MPIB. The authors thank A. Kahraman (Universität Zürich) for helpful discussions regarding the interpretation of cross-linking data and D. Balchin and L. Popilka for critically reading the manuscript. P.W. is supported by an Emmy Noether grant of the German Research council (WE 4628/1).

AUTHOR CONTRIBUTIONS

T.H. performed the biochemical and functional analysis of Raf1 and obtained the Raf1-domain crystals; A.B. solved the crystal structures; J.Y.B. performed the native MS and analyzed the cross-linking MS data; M.H.-H. did the SEC-MALS measurements; and G.M. and P.W. performed the cryo-EM analysis and reconstruction. M.H.-H., F.U.H. and A.B. supervised the experimental design and data interpretation, and wrote the manuscript with contributions from all authors.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

---

1. Andersson, I. & Backlund, A. Structure and function of Rubisco. *Plant Physiol. Biochem.* **46**, 275–291 (2008).
2. Hartman, F.C. & Harpel, M.R. Structure, function, regulation, and assembly of D-ribulose-1,5-bisphosphate carboxylase/oxygenase. *Annu. Rev. Biochem.* **63**, 197–234 (1994).
3. Ellis, R.J. Biochemistry: tackling unintelligent design. *Nature* **463**, 164–165 (2010).
4. Sharkey, T.D. Estimating the rate of photorespiration. *Physiol. Plant.* **73**, 147–152 (1988).
5. Portis, A.R. & Parry, M.A.J. Discoveries in Rubisco (ribulose 1,5-bisphosphate carboxylase/oxygenase): a historical perspective. *Photosynth. Res.* **94**, 121–143 (2007).
6. Peterhansel, C., Niessen, M. & Kebeish, R.M. Metabolic engineering towards the enhancement of photosynthesis. *Photochem. Photobiol.* **84**, 1317–1323 (2008).
7. Maurino, V.G. & Peterhansel, C. Photorespiration: current status and approaches for metabolic engineering. *Curr. Opin. Plant Biol.* **13**, 249–256 (2010).
8. Whitney, S.M., Houtz, R.L. & Alonso, H. Advancing our understanding and capacity to engineer nature's CO_2-sequestering enzyme, Rubisco. *Plant Physiol.* **155**, 27–35 (2011).
9. Parry, M.A.J. *et al.* Rubisco activity and regulation as targets for crop improvement. *J. Exp. Bot.* **64**, 717–730 (2013).
10. Lin, M.T., Occhialini, A., Andralojc, P.J., Parry, M.A. & Hanson, M.R. A faster Rubisco with potential to increase photosynthesis in crops. *Nature* **513**, 547–550 (2014).
11. Durão, P. *et al.* Opposing effects of folding and assembly chaperones on evolvability of Rubisco. *Nat. Chem. Biol.* **11**, 148–155 (2015).
12. Cleland, W.W., Andrews, T.J., Gutteridge, S., Hartman, F.C. & Lorimer, G.H. Mechanism of Rubisco: the carbamate as general base. *Chem. Rev.* **98**, 549–562 (1998).
13. Andersson, I. Catalysis and regulation in Rubisco. *J. Exp. Bot.* **59**, 1555–1568 (2008).
14. Liu, C. *et al.* Coupled chaperone action in folding and assembly of hexadecameric Rubisco. *Nature* **463**, 197–202 (2010).
15. Vitlin Gruber, A., Nisemblat, S., Azem, A. & Weiss, C. The complexity of chloroplast chaperonins. *Trends Plant Sci.* **18**, 688–694 (2013).
16. Saschenbrecker, S. *et al.* Structure and function of RbcX, an assembly chaperone for hexadecameric Rubisco. *Cell* **129**, 1189–1200 (2007).
17. van der Vies, S.M., Bradley, D. & Gatenby, A.A. Assembly of cyanobacterial and higher plant ribulose bisphosphate carboxylase subunits into functional homologous and heterologous enzyme molecules in *Escherichia coli*. *EMBO J.* **5**, 2439–2444 (1986).
18. Andrews, T.J. Catalysis by cyanobacterial ribulose-bisphosphate carboxylase large subunits in the complete absence of small subunits. *J. Biol. Chem.* **263**, 12213–12219 (1988).
19. Goloubinoff, P., Gatenby, A.A. & Lorimer, G.H. GroE heat-shock proteins promote assembly of foreign prokaryotic ribulose bisphosphate carboxylase oligomers in *Escherichia coli*. *Nature* **337**, 44–47 (1989).
20. Li, L.-A. & Tabita, F.R. Maximum activity of recombinant ribulose 1,5-bisphosphate carboxylase/oxygenase of *Anabaena sp.* strain CA requires the product of the *rbcx* gene. *J. Bacteriol.* **179**, 3793–3796 (1997).
21. Onizuka, T. *et al.* The *rbcX* gene product promotes the production and assembly of ribulose-1,5-bisphosphate carboxylase/oxygenase of *Synechococcus sp.* PCC7002 in *Escherichia coli*. *Plant Cell Physiol.* **45**, 1390–1395 (2004).
22. Bracher, A., Starling-Windhof, A., Hartl, F.U. & Hayer-Hartl, M. Crystal structure of a chaperone-bound assembly intermediate of form I Rubisco. *Nat. Struct. Mol. Biol.* **18**, 875–880 (2011).
23. Emlyn-Jones, D., Woodger, F.J., Price, G.D. & Whitney, S.M. RbcX can function as a Rubisco chaperonin, but is non-essential in *Synechococcus* PCC7942. *Plant Cell Physiol.* **47**, 1630–1640 (2006).
24. Feiz, L. *et al.* Ribulose-1,5-bis-phosphate carboxylase/oxygenase accumulation factor1 is required for holoenzyme assembly in maize. *Plant Cell* **24**, 3435–3446 (2012).
25. Kolesinski, P., Belusiak, I., Czarnocki-Cieciura, M. & Szczepaniak, A. 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*. *FEBS J.* **281**, 3920–3932 (2014).
26. Whitney, S.M., Birch, R., Kelso, C., Beck, J.L. & Kapralov, M.V. Improving recombinant Rubisco biogenesis, plant photosynthesis and growth by coexpressing its ancillary RAF1 chaperone. *Proc. Natl. Acad. Sci. USA* **112**, 3564–3569 (2015).
27. Leitner, A. *et al.* The molecular architecture of the eukaryotic chaperonin TRiC/CCT. *Structure* **20**, 814–825 (2012).
28. Newman, J., Bränden, C.I. & Jones, T.A. Structure determination and refinement of ribulose 1,5-bisphosphate carboxylase/oxygenase from *Synechococcus* PCC6301. *Acta Crystallogr. D Biol. Crystallogr.* **49**, 548–560 (1993).
29. Pettersen, E.F. *et al.* UCSF chimera: a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
30. Chari, A. & Fischer, U. Cellular strategies for the assembly of molecular machines. *Trends Biochem. Sci.* **35**, 676–683 (2010).
31. Ellis, R.J. Assembly chaperones: a perspective. *Phil. Trans. R. Soc. Lond. B* **368**, 20110398 (2013).
32. Tanaka, S., Sawaya, M.R., Kerfeld, C.A. & Yeates, T.O. Structure of the RuBisCO chaperone RbcX from *Synechocystis sp.* PCC6803. *Acta Crystallogr. D Biol. Crystallogr.* **63**, 1109–1112 (2007).
33. Kolesinski, P. *et al.* Insights into eukaryotic Rubisco assembly: crystal structures of RbcX chaperones from *Arabidopsis thaliana*. *Biochim. Biophys. Acta* **1830**, 2899–2906 (2013).
34. Wheatley, N.M., Sundberg, C.D., Gidaniyan, S.D., Cascio, D. & Yeates, T.O. Structure and identification of a pterin dehydratase-like protein as a ribulose-bisphosphate carboxylase/oxygenase (RuBisCO) assembly factor in the alpha-carboxysome. *J. Biol. Chem.* **289**, 7973–7981 (2014).
35. Feiz, L. *et al.* RAF2: a novel Rubisco biogenesis factor in maize. *Plant J.* **80**, 862–869 (2014).
36. Hohmann-Marriott, M.F. & Blankenship, R.E. Evolution of photosynthesis. *Annu. Rev. Plant Biol.* **62**, 515–548 (2011).
37. Kerner, M.J. *et al.* Proteome-wide analysis of chaperonin-dependent protein folding in *Escherichia coli*. *Cell* **122**, 209–220 (2005).
38. Georgescauld, F. *et al.* GroEL/ES chaperonin modulates the mechanism and accelerates the rate of TIM-barrel domain folding. *Cell* **157**, 922–934 (2014).

ONLINE METHODS

Cloning and plasmids. Open reading frames for Syn7942-Raf1, Syn7002-Raf1, AtRaf1.1 and AtRaf1.2 were amplified by PCR from genomic DNA of *S. elongatus* PCC7942, *Synechococcus* sp. PCC7002 (ATCC no. 27264) and *A. thaliana* cDNA, respectively, and cloned between the SacII and SacI restriction sites of the pHue plasmid[39], thus resulting in the following constructs: pHueSyn7942raf1, pHueSyn7002raf1, pHueAtraf1.1 and pHueAtraf1.2. The chloroplast transit peptides of AtRaf1.1 and AtRaf1.2 were predicted with TargetP (http://www.cbs.dtu.dk/services/TargetP/) or adopted from the plant proteome database (http://ppdb.tc.cornell.edu/). The bicistronic plasmid pHueAtraf1.1raf1.2 was created by amplifying Atraf1.2 from pHueAtraf1.2 and inserting it into pHueAtraf1.1 with SacI and NotI restriction sites. The respective primer sequences are listed in **Supplementary Table 1**. SacII and SacI sites in the protein-coding regions were removed with whole-plasmid site-directed mutagenesis. The sequences for the N-terminal (Raf1α) and C-terminal (Raf1β) Raf1 domains were analogously cloned into pHue. *T. elongatus* BP-1 genes encoding RbcL and Raf1, and were synthesized (Life Technologies) and cloned between the NcoI and NotI and NdeI and XhoI restriction sites of the bicistronic pCDF-Duet-1 plasmid, respectively (Novagen). Raf1 was synthesized to contain a TEV protease–cleavable N-terminal His-tag (MGSSHHHHHHENLYFQG). For FLAG-tagged constructs, the sequence encoding the FLAG tag (MDYKDDDDKAG) was inserted upstream of the respective start codon (as described above). Point mutants were produced by PCR-based mutagenesis. All plasmid inserts were verified by DNA sequencing.

Protein expression and purification. All purification steps were performed at 4 °C, and protein concentrations were determined spectrophotometrically at 280 nm. Raf1 proteins were expressed as N-terminal His_6-ubiquitin fusion proteins in *E. coli* BL21(DE3) cells containing the given pHue expression plasmid. Cells were grown to an OD_{600} of 0.5 at 37 °C in Luria-Bertani medium; this was followed by induction for 16 h with 0.5 mM isopropyl β-D-thiogalactoside at 23 °C. Cells were lysed in 50 mM Tris-HCl, pH 8.0, 20 mM NaCl, 1 mM EDTA, 0.5 mg mL^{-1} lysozyme and 5 mM phenylmethylsulfonyl fluoride for 30 min on ice, and this was followed by ultrasonication (Misonix Sonicator 3000). The supernatant obtained after high-speed centrifugation (48,000*g*, 40 min, 4 °C) was applied to an Ni-IMAC column (GE Biotech) to capture the His_6-ubiquitin-Raf1 fusion protein. This was followed by overnight cleavage of the His_6-ubiquitin moiety at 23 °C with the deubiquitinating enzyme Usp2 (ref. 40). The protein-containing fraction was dialyzed against buffer A (20 mM Tris-HCl, pH 8.0, and 50 mM NaCl) and applied to a MonoQ (GE Biotech) column equilibrated with buffer A. Proteins were eluted with a linear salt gradient to 1 M NaCl. Fractions containing Raf1 were combined and concentrated, 5% glycerol was added, and fractions were flash frozen in liquid nitrogen and stored at –80 °C. Raf1 for X-ray crystallographic studies was purified further by Superdex200 (GE Biotech) size-exclusion chromatography in buffer A.

T. elongatus RbcL–Raf1 complex was expressed and purified essentially as described for Raf1 proteins. After purification with an Ni-IMAC column, the complex was dialyzed against buffer A. This was followed by purification with an ion-exchange column (MonoQ) and a final size-exclusion-chromatography step in buffer A (Superdex200). Fractions containing the HMW *T. elongatus* RbcL–Raf1 complex were combined and concentrated, 5% glycerol was added, and fractions were flash frozen in liquid nitrogen and stored at –80 °C.

S. elongatus 6301 (Syn6301) $RbcL_8S_8$, $RbcL_8$ and RbcS, as well as unassembled RbcL, were purified as previously described[14,16]. The Rubisco proteins of Syn6301 and Syn7942 are 100% identical in sequence (UniProt P00880, Q31NB3, P04716 and Q31NB2). Rabbit antibodies against *S. elongatus* RbcL (ref. 14), Syn7942-Raf1 and AtRaf1.2 were produced with standard procedures (**Supplementary Data Set 2**). The *E. coli* chaperonins GroEL and GroES were purified as described previously[41].

Rubisco reconstitution. GroEL–ES–mediated RbcL folding was performed as in ref. 14, with modifications. Denatured *S. elongatus* RbcL was diluted 200-fold from 6 M GuHCl (final concentration 0.5 µM) into ice-cold buffer B (20 mM MOPS-KOH, pH 7.5, 100 mM KCl, and 5 mM $MgOAc_2$) containing 5 mM DTT, 1 mg mL^{-1} BSA and 1 µM GroEL, and was incubated on ice and then centrifuged to remove any aggregated protein. GroES (2 µM), *S. elongatus* RbcS (5 µM) and Raf1 were added to the supernatant containing GroEL-bound RbcL when indicated in figure legends. Folding and assembly was initiated by addition of 4 mM ATP and incubation at 25 °C. Reactions were stopped at the indicated times by addition of apyrase (Sigma) to a final concentration of 0.25 U μl^{-1} to inhibit GroEL–ES activity. (Apyrase hydrolyzes ATP and ADP to AMP.) Formation of assembled complexes was analyzed by continuous Tris-acetate native-PAGE gradient gels (5–15% acrylamide) and immunoblotting with anti-RbcL and anti-Raf1 antibodies.

For measurement of Rubisco enzymatic activity, aliquots of the folding assay were supplemented with 5 µM *S. elongatus* RbcS when it had not previously been present in the reaction, and $RbcL_8S_8$ assembly was allowed to proceed for 15 min at 25 °C. Rubisco carboxylation activity was determined after incubation at 25 °C for 10 min in 50 mM Tris-HCl, pH 8.0, 10 mM $MgCl_2$, and 30 mM $NaH^{14}CO_3$ (25 Bq/nmol), and the amount of fixed carbon was quantified as described previously[41]. Activities are expressed as a percentage of purified *S. elongatus* $RbcL_8$ standard supplemented with *S. elongatus* RbcS (control).

Interaction of Raf1 with preformed $RbcL_8$ complexes. Purified $RbcL_8$ and $RbcL_8S_8$ complexes were incubated with Raf1 for 15 min at 25 °C in buffer B containing 5 mM DTT. Formation of assembled complexes was analyzed by continuous Tris-acetate native PAGE (5–15% acrylamide) and immunoblotting with anti-RbcL and anti-Raf1 antibodies and by SEC-MALS and native MS.

Size-exclusion chromatography coupled to multi-angle static light scattering (SEC-MALS). Protein samples (~30 to 60 µg) were analyzed with static and dynamic light scattering by autoinjection of the sample onto a SEC column (5 µm, 7.8 × 300 mm column, Wyatt Technology WTC-030N5) at a flow rate of 0.2 mL min^{-1} in buffer C (25 mM Tris-HCl, pH 8.0, and 50 mM NaCl) at 25 °C or at a flow rate of 0.15 mL min^{-1} in buffer B for analysis of the $RbcL_8$–Raf1 complexes in solution. The column is inline with the following detectors: a variable UV-absorbance detector set at 280 nm (Agilent 1100 series), a DAWN EOS MALS detector (Wyatt Technology, 690-nm laser) and an Optilab rEX refractive-index detector (Wyatt Technology, 690-nm laser)[42]. Masses were calculated with ASTRA (Wyatt Technology) with the dn/dc value set to 0.185 mL g^{-1}. Bovine serum albumin (Thermo) was used as the calibration standard.

Native mass spectrometry. The purified *S. elongatus* $RbcL_8$ (1.25 µM oligomer), Syn7942-Raf1 (5 µM dimer) and the binding reaction (15 min at 25 °C) of Syn7942-Raf1 (5 µM dimer) to $RbcL_8$ (1.25 µM oligomer) in buffer D (20 mM MOPS-KOH, pH 7.5, 50 mM KCl, and 5 mM $MgOAc_2$) containing 5 mM DTT were buffer-exchanged into 100 mM ammonium acetate, pH 7.5–8.0 (Fluka, Sigma) with Micro Bio-Spin 6 chromatography columns (Bio-Rad). The purified *T. elongatus* RbcL–Raf1 complex (~8 µM) was buffer-exchanged as above. Native-MS analyses were performed in positive-ion mode on an electrospray ionization quadrupole time-of-flight (ESI-TOF) hybrid mass spectrometer (Synapt G2-Si, Waters) equipped with a Z-spray nano-ESI source. Gold-plated 10-µm nano-ESI pipettes (Mascom) were used as capillaries. Optimized capillary and sample cone voltages were 1–1.6 kV and 100–150 V, respectively. Spectra were calibrated with 30 mg mL^{-1} cesium iodide dissolved in 1:1 acetonitrile/water.

Electron microscopy (EM) and 3D reconstruction. $RbcL_8$–$Raf1_4$ complex (~50 µg mL^{-1}) in buffer B containing 5 mM DTT was applied to freshly plasma-cleaned, carbon-coated grids (Quantifoil) and was negative stained with 2% (w/v) uranyl acetate. Images were digitally recorded on a Philips CM200 FEG electron microscope equipped with a TemCam F415MP 4k at a nominal magnification of 50,000×, with a pixel size of 2.16 Å/pixel at specimen level and defocus ranging from 300 to 1,500 nm. The microscope was operated under low-dose conditions at 160 kV. The defocus and astigmatism of the images were determined with CTFFIND4 (ref. 43) in RELION-1.3 (ref. 44). A total of 62, 49 and 90 micrographs were selected for image processing of the *S. elongatus* $RbcL_8$–$Raf1_4$, the cross-linked *S. elongatus* $RbcL_8$–$Raf1_4$ and the *T. elongatus* $RbcL_8$–$Raf1_4$ complexes, respectively. Manual particle selection with RELION-1.3 yielded a final data set of 5,183 particles for the cross-linked *S. elongatus* $RbcL_8$–$Raf1_4$ complex and 5,471 particles for the *T. elongatus* $RbcL_8$–$Raf1_4$ complex. For the native *S. elongatus* $RbcL_8$–$Raf1_4$ complex, a total of 1,057 particles from 13 micrographs were manually selected with RELION-1.3 to generate references for automatic particle picking. After automatic particle selection in RELION-1.3 and manual cleaning of the data set on the basis of z-score characteristics, a stack of 7,602 particles was obtained. The extracted particle data set was subjected to 2D classification in RELION-1.3, and classes showing erroneously picked features and aggregates were discarded, thus resulting in 6,191 particles in the final data set. Initial 3D reconstruction of the *S. elongatus* RbcL$_8$–Raf1$_4$ complex and the *T. elongatus* RbcL$_8$–Raf1$_4$ complex were generated with IMAGIC-5 (ref. 45). Particle images were band-pass-filtered between 250 and 10 Å, normalized and centered by iterative alignment to their rotationally averaged sum. Initial class averages containing ~30 particles were obtained by two rounds of classification on the basis of multivariate statistical analysis (MSA) and subsequent multireference alignment with homogenous classes as new references. For symmetry analysis, top views of the complexes were extracted, randomly rotated and subjected to MSA. A density map with imposed dihedral four-fold symmetry was created by angular reconstitution. 3D refinement of the initial model was carried out in RELION-1.3 with application of four-fold symmetry and no mask. The resolution of the final refinement by gold-standard Fourier shell correlation (FSC) at cutoffs of 0.143 and 0.5 is shown in **Supplementary Figure 7a**. The PDB structures of RbcL$_8$ (from the Rubisco crystal structure[28]) and Raf1 domains were fitted manually as rigid bodies with UCSF Chimera[29].

Cross-linking coupled to mass spectrometry (CXMS). Purified RbcL$_8$ (1.25 µM oligomer) was incubated with Raf1 (10 µM) in buffer D for 15 min at 25 °C. This was followed by addition of 2 mM of a 1:1 isotopic mixture of H_{12} and D_{12} disuccinimidylsuberate, H_{12}/D_{12}-DSS (Creative Molecules) for 30 min (**Supplementary Fig. 6**). Reactions were quenched by addition of 150 mM NH_4HCO_3. (CXMS) with the bifunctional lysine-specific cross-linker disuccinimidylsuberate (DSS). Cross-linking during reconstitution was performed with N-terminally FLAG-tagged Raf1; this was followed by isolation of cross-linked Raf1 by anti-FLAG immunoprecipitation (anti-FLAG beads, Sigma). Cross-linked complexes were separated on 4–12% Bis-Tris SDS-PAGE (Life Technologies) and visualized by Coomassie staining. Excision of gel bands, reduction, alkylation, in-gel digestion and desalting were performed as previously described[46,47]. Desalted peptides were dissolved in 5% formic acid and analyzed on an Easy-nLC 1000 UPLC system (Thermo) connected to a Q-Exactive Orbitrap mass spectrometer (Thermo). Peptides were separated on nanospray columns (ID 75 µm, 20 cm long with 8-µm tip opening, New Objective) containing 1.9-µm C18 beads (Reprosil-Pur C18-AQ, Dr. Maisch) on a 60-min gradient with buffers (0.2% formic acid in water and 0.2% formic acid in acetonitrile). Sample loading to the column was performed by the Thermo Easy-nLC 1000 autosampler without a trap column, at a flow rate of 0.5 µl per min. The UPLC flow rate during sample analysis was set to 0.2 µl/min. MS^2 analysis was performed with standard data-dependent mode settings, with alternation between one high-resolution (resolution 70,000) MS1 scan (*m*/*z* of 400–1,750) and ten MS2 scans (resolution 17,500) of the ten most intense ions with charge states of three or more, as run on Thermo Xcalibur software.

For data analysis, Thermo Xcalibur .raw files were converted to .mgf (Mascot generic file) format, with Proteome Discoverer 1.4 (Thermo). To identify cross-linked peptides, the .mgf files were analyzed on StavroX 3.1.19 (ref. 48) with the following parameters: MS1 tolerance of 3 p.p.m.; MS2 tolerance of 0.8 Da; missed cleavages for lysine and arginine of 3 and 1, respectively; signal-to-noise ratio ≥2; and fixed and variable modifications of cysteine carbamidomethylation and methionine oxidation, respectively. The potential cross-linking sites for DSS are lysine residues and the free amino group at the N terminus. All the identified linked peptides were validated manually for b- and y-ion assignment and were included in the final list only if precursors of the respective peptides contained doublets with mass difference of 12.0753 Da (mass difference between H_{12}-DSS and D_{12}-DSS). The cross-linked peptides were identified from two independent experiments for each Raf1 protein.

Small-angle X-ray scattering (SAXS). SAXS measurements were performed at beamline BM29 of the European Synchrotron Radiation Source (ESRF), Grenoble, France. Protein samples at three different concentrations in 50 mM Tris-HCl, pH 8.0, 50 mM NaCl and 5% (v/v) glycerol were exposed for 1 s. Scattering data from ten repeats were averaged. Buffer background was subtracted. The protein scattering data were processed with Primus[49,50]. Radii of gyration were determined with the Guinier approximation. Scattering curves were fitted with GNOM[51]. An ensemble model of the AtRaf1.2 structure was generated with EOM 2.0 (refs. 52,53).

Crystallization. Initial crystals of AtRaf1.2α were obtained by the sitting-drop vapor-diffusion method at 18 °C, with mixture of equal volumes of AtRaf1.2(62–449) (18 mg mL^{-1} in buffer C) with a precipitant containing 0.1 M MES-NaOH, pH 6.0, and 20% (w/v) PEG 6000 (Qiagen pHClear Suite condition D3. Crystals of AtRaf1.2α(62–274) were obtained at 20 °C with 0.1 M MES-NaOH, pH 6.0, and 26% (w/v) PEG-6000 as a precipitant. For cryoprotection, crystals were transferred stepwise into 0.1 M MES-NaOH, pH 6.0, 26% (w/v) PEG 6000 and 15% (v/v) glycerol before being flash frozen in liquid nitrogen.

Orthorhombic crystals of AtRaf1.2β were obtained by the sitting-drop vapor-diffusion method at 4 °C, with mixture of equal volumes of AtRaf1.2(62–449) (17 mg mL^{-1} in buffer C) with a precipitant containing 20% PEG 6000, 50 mM K_2HPO_4 and 200 mM KH_2PO_4. Monoclinic crystals of AtRaf1.2β(281–449) were obtained with 10% PEG3350 as a precipitant.

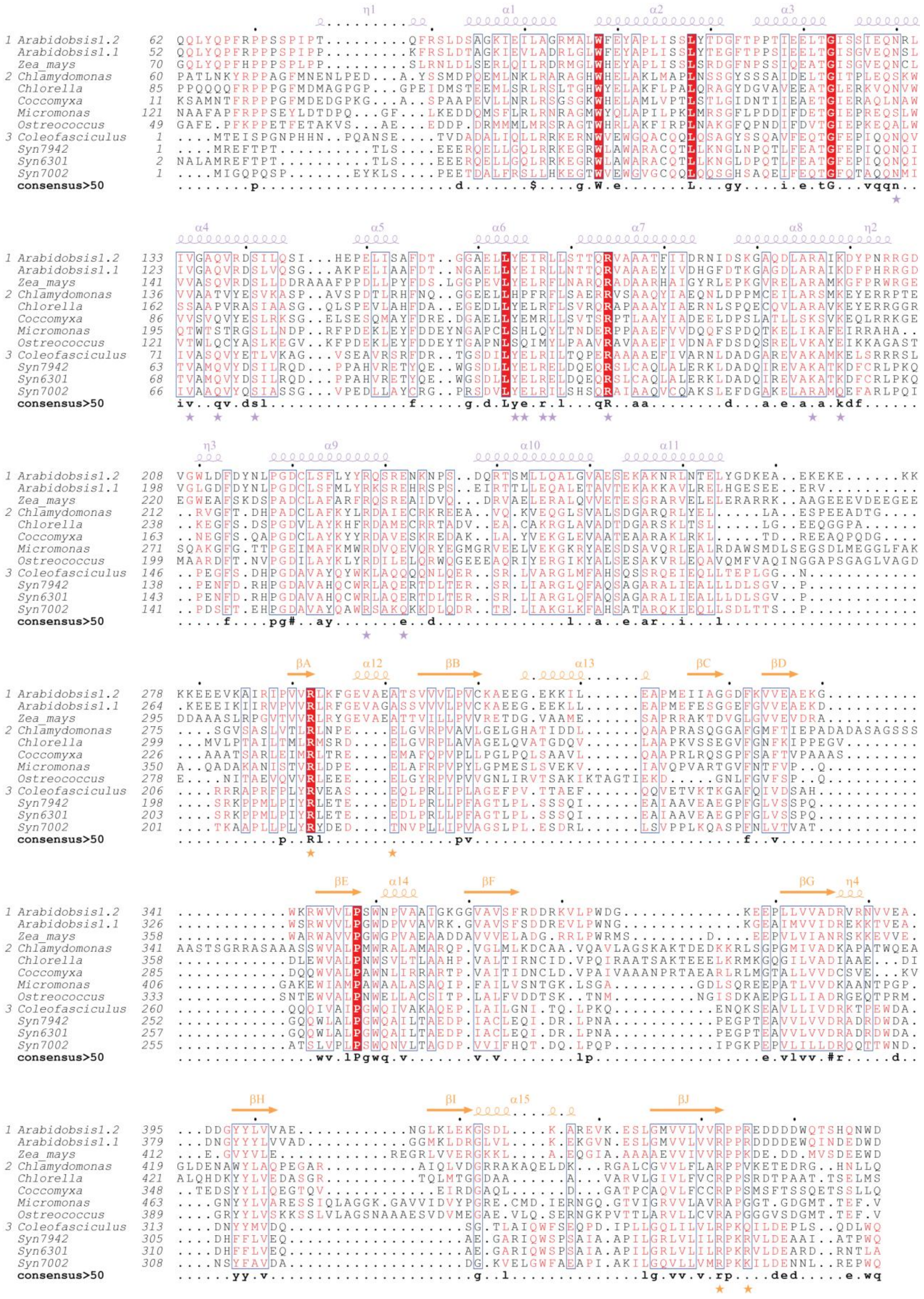
Structure determination. Diffraction data were collected at beamlines X10SA and X06DA of the Swiss Light Source (SLS) in Villigen, Switzerland. The data were processed with XDS[54] and transferred into the CCP4 format with Pointless[55], Scala[56] and Truncate[57]. The structures of AtRaf1.2α and monoclinic AtRaf1.2β were solved by SIRAS with crystals soaked with 1 mM $K_2[PtCl_4]$ and sodium *p*-(hydroxymercuri)benzoate (HMBA), respectively. Two platinum and six mercury sites were found with SHELXD[58], as implemented in HKL2MAP[59], for AtRaf1.2α and AtRaf1.2β, respectively. These solutions were further refined with Sharp[60]. Density modification was subsequently carried out with Resolve[61]. A preliminary model for AtRaf1.2α was manually built in the resulting map with Coot[62]. For final model building and refinement, a nearly isomorphous native data of AtRaf1.2α(62–274) was used. Buccaneer[63] built a partial model of AtRaf1.2β, which was completed manually with Coot. Orthorhombic AtRaf1.2β was solved by molecular replacement with Molrep[64]. The models were improved by iterative cycles of refinement with Refmac[65], as implemented in the CCP4 interface[66], and this was followed by manual model building. The final refinement of AtRaf1.2α was carried out with phenix.refine[67]. Nonglycine residues facing solvent channels without discernible side chain density were modeled as alanines.

Structure analysis. Coordinates were aligned with Lsqman[68]. The sequence alignment was prepared with ClustalW[69,70] and ESPript[71]. Secondary-structure prediction was performed with Jpred-3 (ref. 72). Figures were generated with PyMOL (http://www.pymol.org/).

39. Catanzariti, A.M., Soboleva, T.A., Jans, D.A., Board, P.G. & Baker, R.T. An efficient system for high-level expression and easy purification of authentic recombinant proteins. *Protein Sci.* **13**, 1331–1339 (2004).
40. Baker, R.T. *et al.* Using deubiquitylating enzymes as research tools. *Methods Enzymol.* **398**, 540–554 (2005).
41. Brinker, A. *et al.* Dual function of protein confinement in chaperonin-assisted protein folding. *Cell* **107**, 223–233 (2001).
42. Wyatt, P.J. Light scattering and the absolute characterization of macromolecules. *Anal. Chim. Acta* **272**, 1–40 (1993).
43. Mindell, J.A. & Grigorieff, N. Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* **142**, 334–347 (2003).
44. Scheres, S.H. Semi-automated selection of cryo-EM particles in RELION-1.3. *J. Struct. Biol.* **189**, 114–122 (2015).
45. van Heel, M., Harauz, G., Orlova, E.V., Schmidt, R. & Schatz, M. A new generation of the IMAGIC image processing system. *J. Struct. Biol.* **116**, 17–24 (1996).
46. Shevchenko, A., Wilm, M., Vorm, O. & Mann, M. Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal. Chem.* **68**, 850–858 (1996).
47. Rappsilber, J., Ishihama, Y. & Mann, M. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem.* **75**, 663–670 (2003).
48. Götze, M. *et al.* Automated assignment of MS/MS cleavable crosslinks in protein 3D-structure analysis. *J. Am. Soc. Mass Spectrom.* **26**, 83–97 (2015).
49. Vachette, P., Koch, M.H. & Svergun, D.I. Looking behind the beamstop: X-ray solution scattering studies of structure and conformational changes of biological macromolecules. *Methods Enzymol.* **374**, 584–615 (2003).
50. Konarev, P.V., Volkov, V.V., Sokolova, A.V., Koch, M.H.J. & Svergun, D.I. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* **36**, 1277–1282 (2003).
51. Svergun, D.I. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* **25**, 495–503 (1992).
52. Bernadó, P., Mylonas, E., Petoukhov, M.V., Blackledge, M. & Svergun, D.I. Structural characterization of flexible proteins using small-angle X-ray scattering. *J. Am. Chem. Soc.* **129**, 5656–5664 (2007).
53. Petoukhov, M.V. *et al.* New developments in the program package for small-angle scattering data analysis. *J. Appl. Crystallogr.* **45**, 342–350 (2012).
54. Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
55. Evans, P. Scaling and assessment of data quality. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 72–82 (2006).

doi:10.1038/nsmb.3062

56. Evans, P.R. Scala. *Joint CCP4 ESF-EACBM Newsl. Prot. Crystallogr.* **33**, 22–24 (1997).
57. French, G. & Wilson, K. On the treatment of negative intensity observations. *Acta Crystallogr. A* **34**, 517–525 (1978).
58. Sheldrick, G.M. Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 479–485 (2010).
59. Pape, T. & Schneider, T.R. HKL2MAP: a graphical user interface for phasing with SHELX programs. *J. Appl. Crystallogr.* **37**, 843–844 (2004).
60. Vonrhein, C., Blanc, E., Roversi, P. & Bricogne, G. Automated structure solution with autoSHARP. *Methods Mol. Biol.* **364**, 215–230 (2007).
61. Terwilliger, T.C. Maximum-likelihood density modification. *Acta Crystallogr. D Biol. Crystallogr.* **56**, 965–972 (2000).
62. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
63. Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 1002–1011 (2006).
64. Vagin, A.A. & Isupov, M.N. Spherically averaged phased translation function and its application to the search for molecules and fragments in electron-density maps. *Acta Crystallogr. D Biol. Crystallogr.* **57**, 1451–1456 (2001).
65. Murshudov, G.N., Vagin, A.A. & Dodson, E.J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53**, 240–255 (1997).
66. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **50**, 760–763 (1994).
67. Afonine, P.V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 352–367 (2012).
68. Kleywegt, G.T. & Jones, T.A. A super position. *CCP4 ESF-EACBM Newsl. Prot. Crystallogr.* **31**, 9–14 (1994).
69. Thompson, J.D., Higgins, D.G. & Gibson, T.J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).
70. Larkin, M.A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007).
71. Gouet, P., Courcelle, E., Stuart, D.I. & Metoz, F. ESPript: multiple sequence alignments in PostScript. *Bioinformatics* **15**, 305–308 (1999).
72. Cole, C., Barber, J.D. & Barton, G.J. The Jpred 3 secondary structure prediction server. *Nucleic Acids Res.* **36**, W197–W201 (2008).

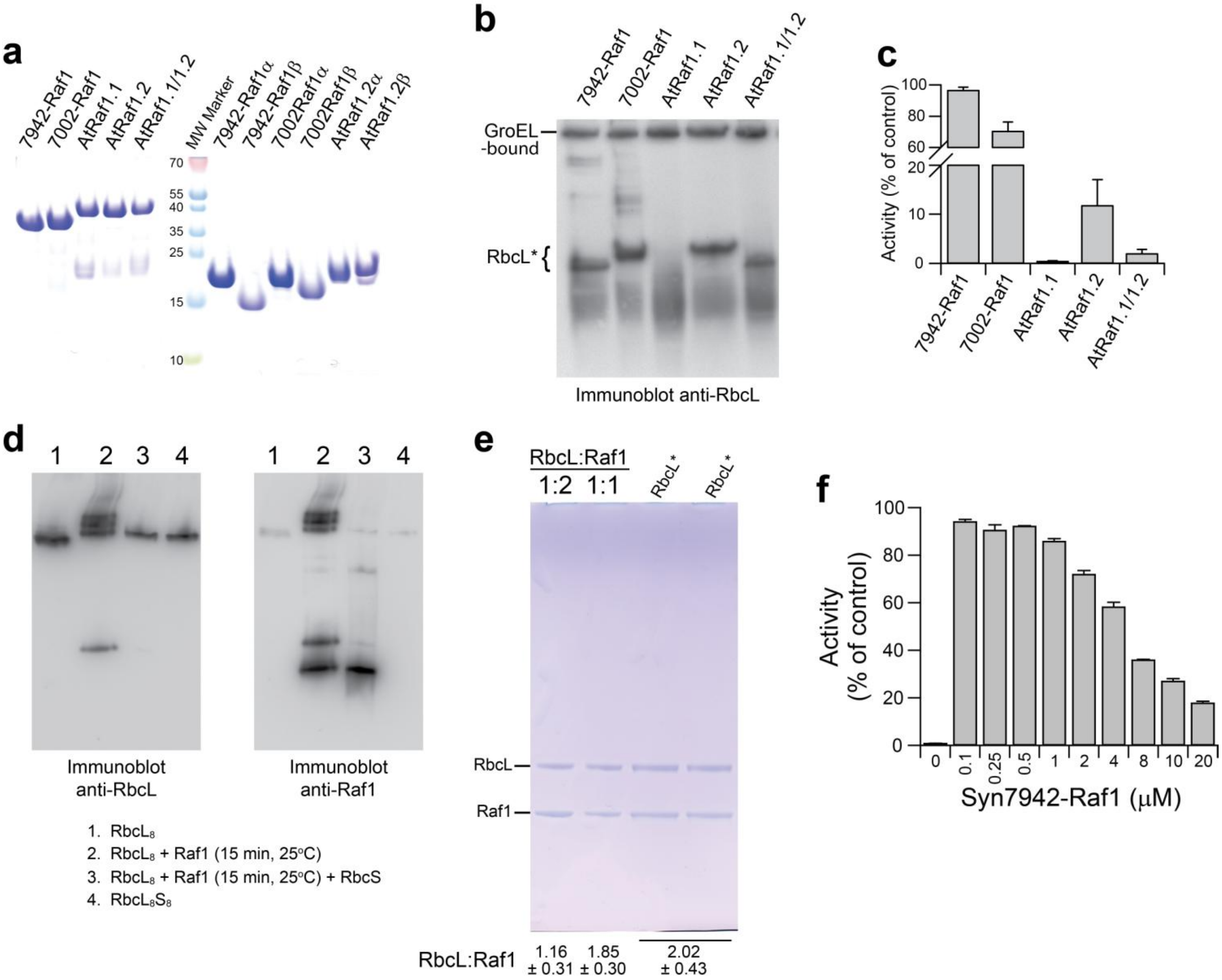


Nature Structural & Molecular Biology: doi:10.1038/nsmb.3062

Supplementary Figure 1

Alignment of Raf1 sequences.

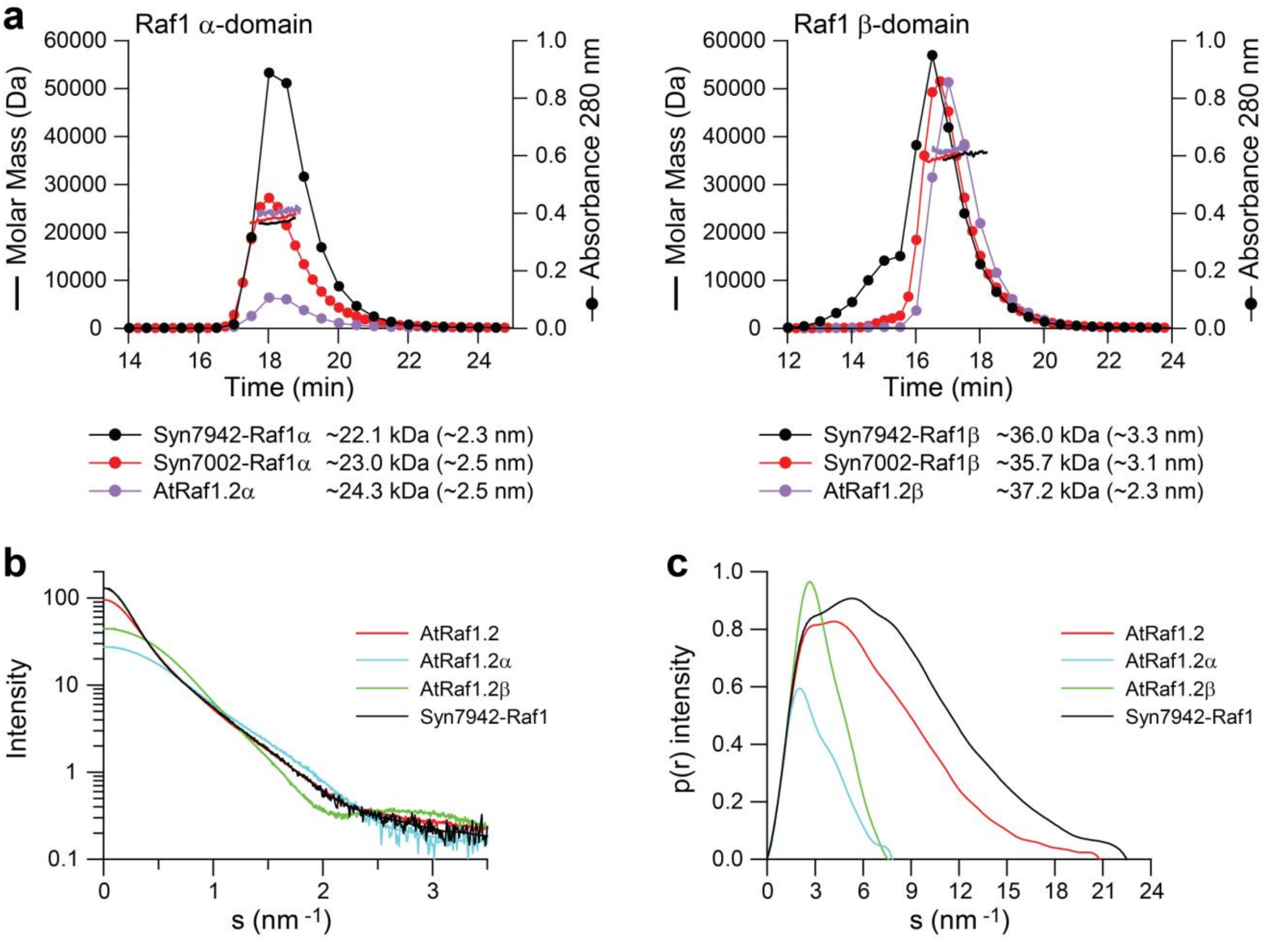
Amino acid sequences of a representative set of Raf1 homologs were aligned using the EBI Clustal-Ω server. Secondary structure elements for Raf1.2 from *Arabidopsis thaliana* are indicated above the sequences. The Raf1 domain structure is indicated by purple and orange coloring of secondary structure elements in the Raf1α and Raf1β domains, respectively. The sequences from plants (1), green algae (2) and cyanobacteria (3) are grouped separately. Similar residues are shown in red and identical residues in white on a red background. Blue frames indicate homologous regions. The consensus sequence is shown at the bottom. The chloroplast signal sequences are not shown. Asterisks below the sequence indicate mutations in Syn7942-Raf1 analyzed in this study (Fig. 4). The Uniprot accession codes for the sequences are: Q9SR19, *Arabidopsis thaliana* Raf1.2; Q9LKR8, *Arabidopsis thaliana* Raf1.1; B4FR29, *Zea mays*; I0YJW5, *Coccomyxa subellipsoidea* C-169; E1ZGR5, *Chlorella variabilis*; Cre06.g308450.t1.2, *Chlamydomonas reinhardtii*; Q00S02, *Ostreococcus tauri*; C1FI81, *Micromonas* sp. (strain RCC299 / NOUM17); B4VSU9, *Coleofasciculus chthonoplastes* PCC7420; Q31Q05, *Synechococcus elongatus* PCC7942; Q5N472, *Synechococcus elongatus* PCC6301; B1XK11, *Synechococcus* sp. PCC7002.



Supplementary Figure 2

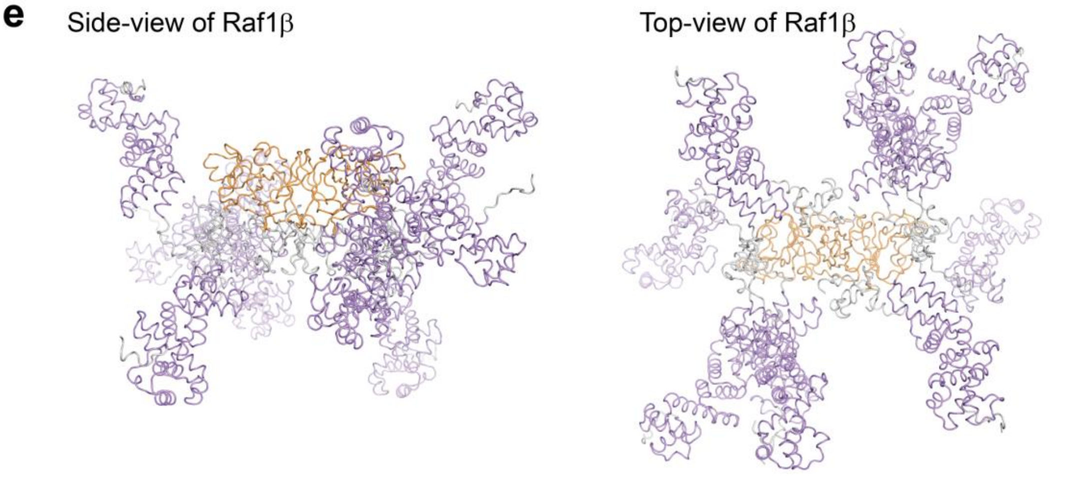
Functional analysis of Raf1 homologs.

(**a**) Purified full-length Raf1 proteins and the respective α- and β-domains of Syn7942, Syn7002 and *A. thaliana*. AtRaf1.1/1.2 is a complex of AtRaf1.1 and AtRaf1.2 produced from a biscistronic plasmid. (**b**) Native-PAGE analysis of Rubisco reconstitution reactions as in Fig. 1b, containing the Raf1 proteins indicated. (**c**) Rubisco activities in reactions shown in (b) after supplementation with RbcS as described in Fig. 1d. Error bars, s.d. (*n* = 3 independent experiments). (**d**) Displacement of Raf1 from $RbcL_8$ by RbcS. Purified *S. elongatus* $RbcL_8$ was incubated with Syn7942-Raf1 as in Fig. 2a, followed by addition of RbcS (**5** μM) and analysis by native-PAGE and immunoblotting with anti-RbcL (left) and anti-Raf1 (right). *S. elongatus* $RbcL_8$ and $RbcL_8S_8$ were used as standards. (**e**) Stoichiometry of RbcL and Raf1 in RbcL* complexes. RbcL* was excised from native-PAGE and separated by SDS-PAGE followed by Coomassie staining and quantitation by densitometry. Top, molar ratios of RbcL and Raf1 standards as quantified by extinction coefficients. Bottom, ratios of RbcL to Raf1 stain intensities are indicated as averages ±S.D from four measurements. The molar ratio of RbcL to Raf1 in RbcL* is close 1:1. Shown is a representative Coomassie stained gel. (**f**) Dependence of Rubisco assembly on Raf1 concentration. Reconstitution reactions were performed as in Fig. 1e at increasing concentrations of Raf1 and the Rubisco activities obtained after 60 min are indicated as percentage of control. Error bars, s.d. (*n* = 3 independent experiments).



d

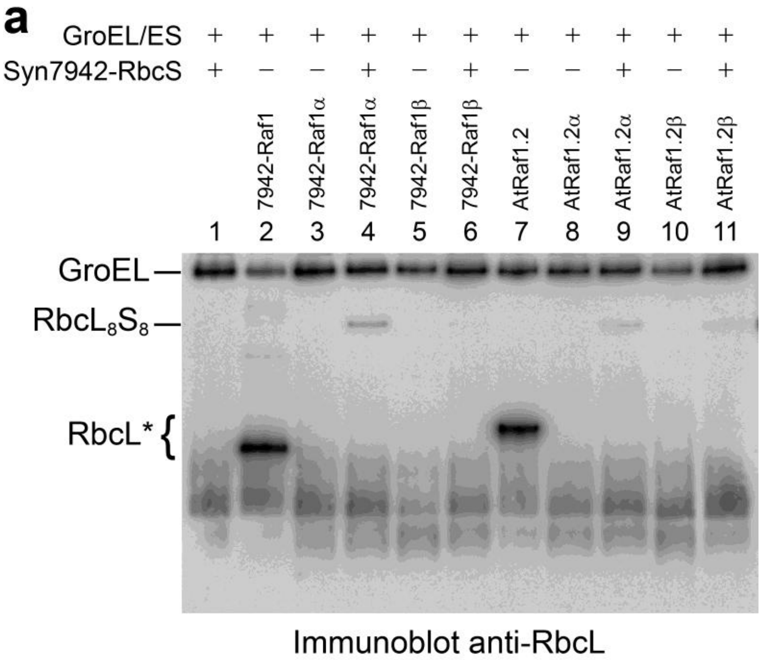
	R_g (Å)	D_{max} (Å)
AtRaf1.2	52	208
AtRaf1.2α	24	78
AtRaf1.2β	25	76
Syn7942-Raf1	59	225

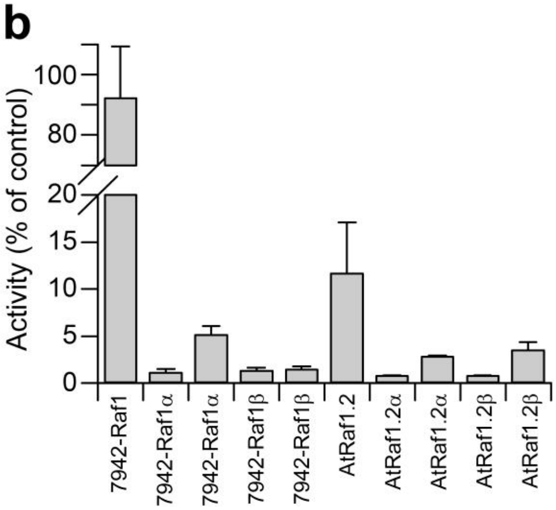


Supplementary Figure 3

SEC-MALS and SAXS analysis of Raf1 proteins.

(**a**) SEC-MALS analysis of purified Raf1 domains from Syn7942, Syn7002 and *A. thaliana*. Data showing measurements for ~30 μg of the respective proteins. Horizontal lines across the peaks indicate molar mass and homogeneity of the sample. Calculated molar masses and hydrodynamic radii are indicated. (**b**) Representative X-ray scattering curves of AtRaf1.2 (red), AtRaf1.2α (blue) and AtRaf1.2β (green) and Syn7942-Raf1 (black). The curves represent background-corrected averages of ten measurements. The GNOM-fitted[51] curves are overlaid. (**c**) Density distributions for AtRaf1.2 (red), AtRaf1.2α (blue), AtRaf1.2β (green) and Syn7942-Raf1 (black) calculated with GNOM. AtRaf1.2α and AtRaf1.2β appear rod-shaped and globular, respectively. The curves for AtRaf1.2 and Syn7942-Raf1 suggests flexibly linked domains.(**d**) Parameters from SAXS data analysis. Radii of gyration were determined using the Guinier approximation. Scattering curves were fitted with GNOM. (**e**) Ensemble model for the AtRaf1.2 dimer. Two perpendicular views are shown. The backbones are represented as coils. A subset of five models matching the experimental scattering curve (Chi value 3.978) was picked from a library of 10,000 conformations by a genetic algorithm implemented in the program EOM 2.0[52,53]. The position of the dimeric β-domain (orange) was fixed at the dyad symmetry axis. The α-domains are represented in purple; the flexible termini and inter-domain linkers are shown in gray.

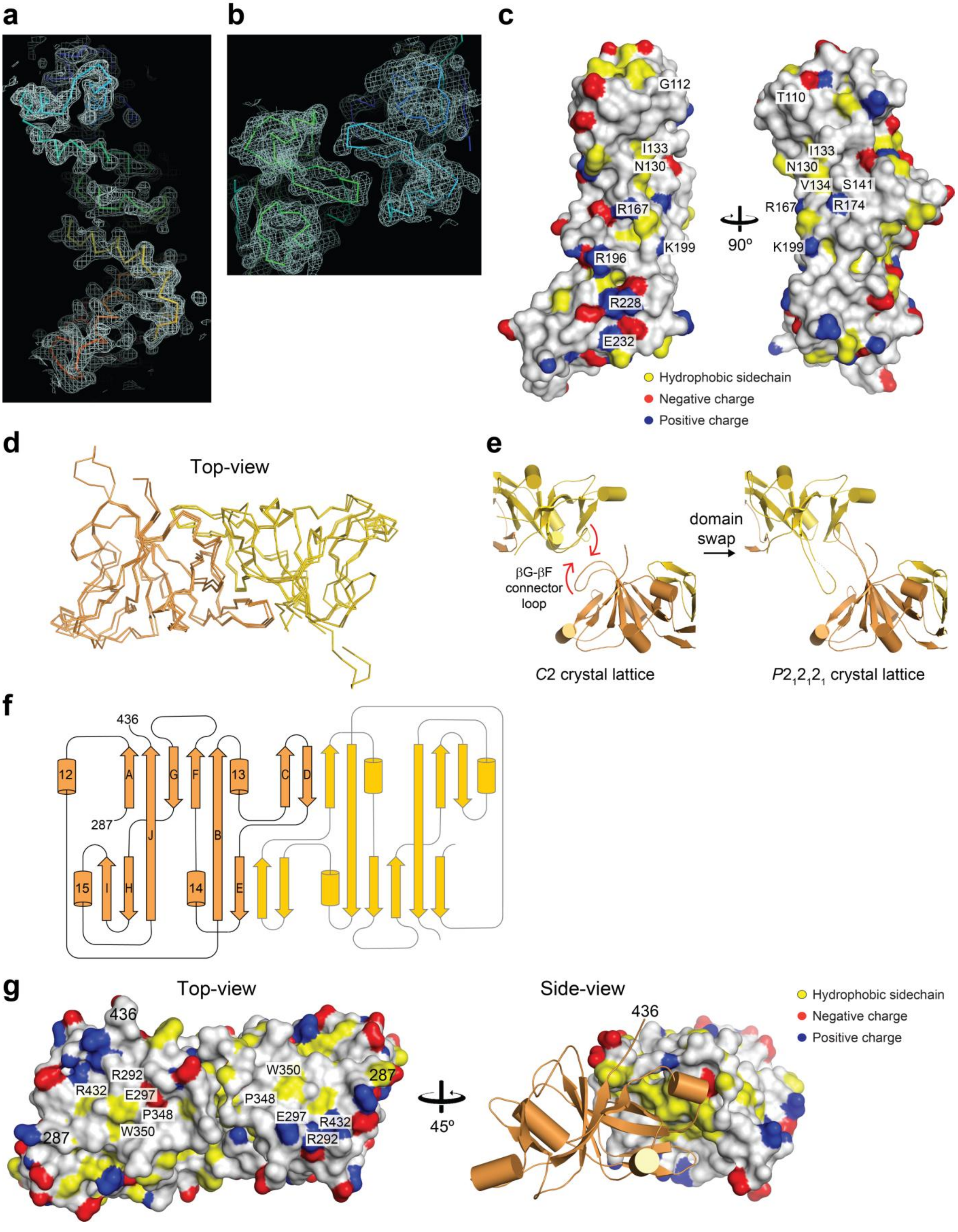




Supplementary Figure 4

Functional analysis of Raf1 homologs.

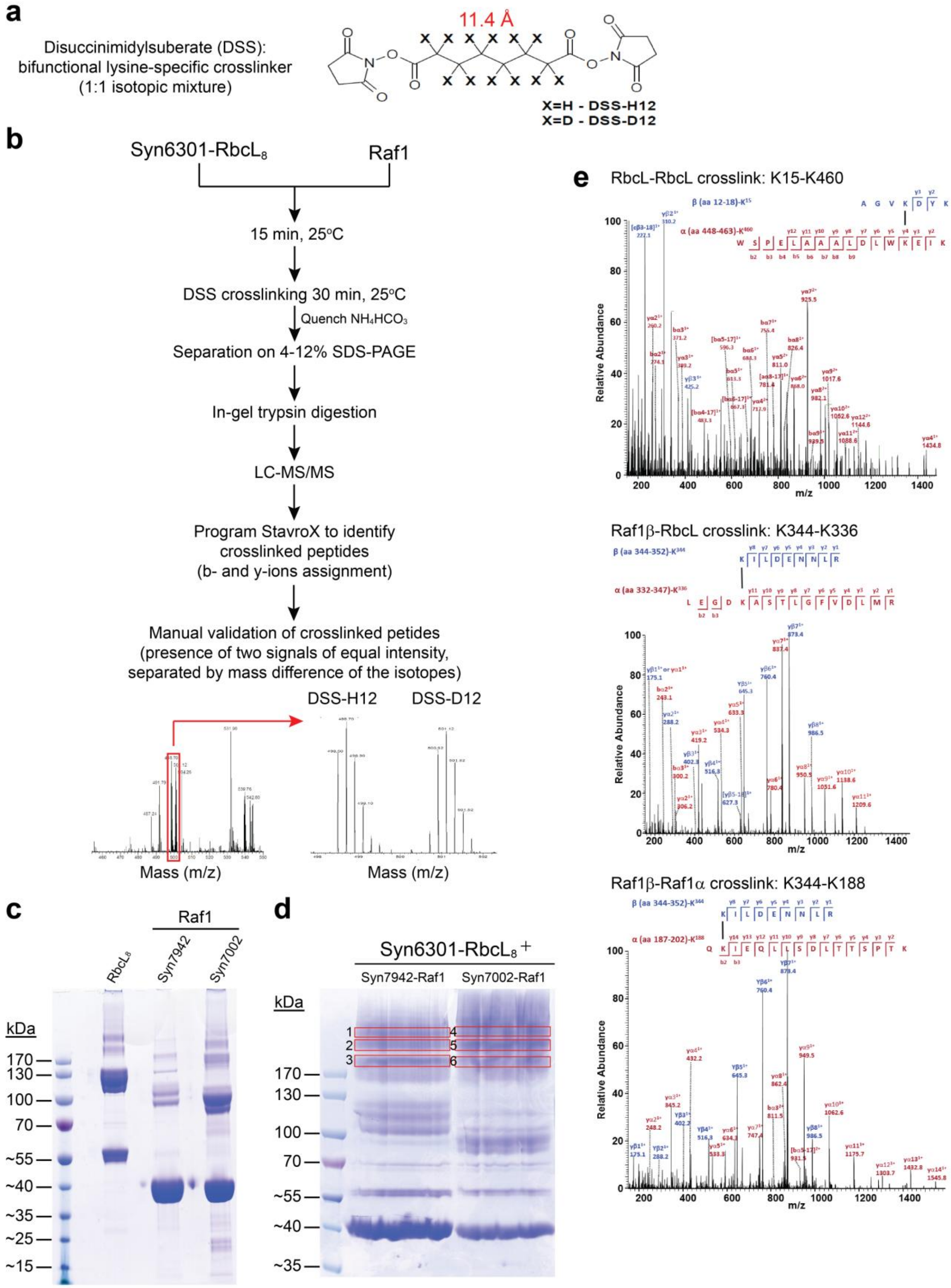
(**a**) Native-PAGE analysis of Rubisco reconstitution reactions as in Fig. 1b, containing purified full-length Raf1 and the α- and β-domains from Syn7942 and *A. thaliana*. RbcS was present when indicated. (**b**) Rubisco activities in reactions shown in (a) after supplementation with RbcS when absent, as described in Fig. 1d. Error bars, s.d. (*n* = 3 independent experiments).


a
b
c
G112
T110
I133
N130
I133
N130
V134
S141
R167
R174
R167
K199
K199
90º
R196
R228
E232
Hydrophobic sidechain
Negative charge
Positive charge
d
Top-view
e
domain swap
βG-βF connector loop
C2 crystal lattice
P2₁2₁2₁ crystal lattice
f
436
287
12
A
J
G
F
B
13
C
D
15
I
H
14
E
g
Top-view
Side-view
436
R292
R432
E297
P348
W350
287
W350
287
P348
E297
R432
R292
45º
436
Hydrophobic sidechain
Negative charge
Positive charge

Supplementary Figure 5

Crystal structures of AtRaf1.2 domains.

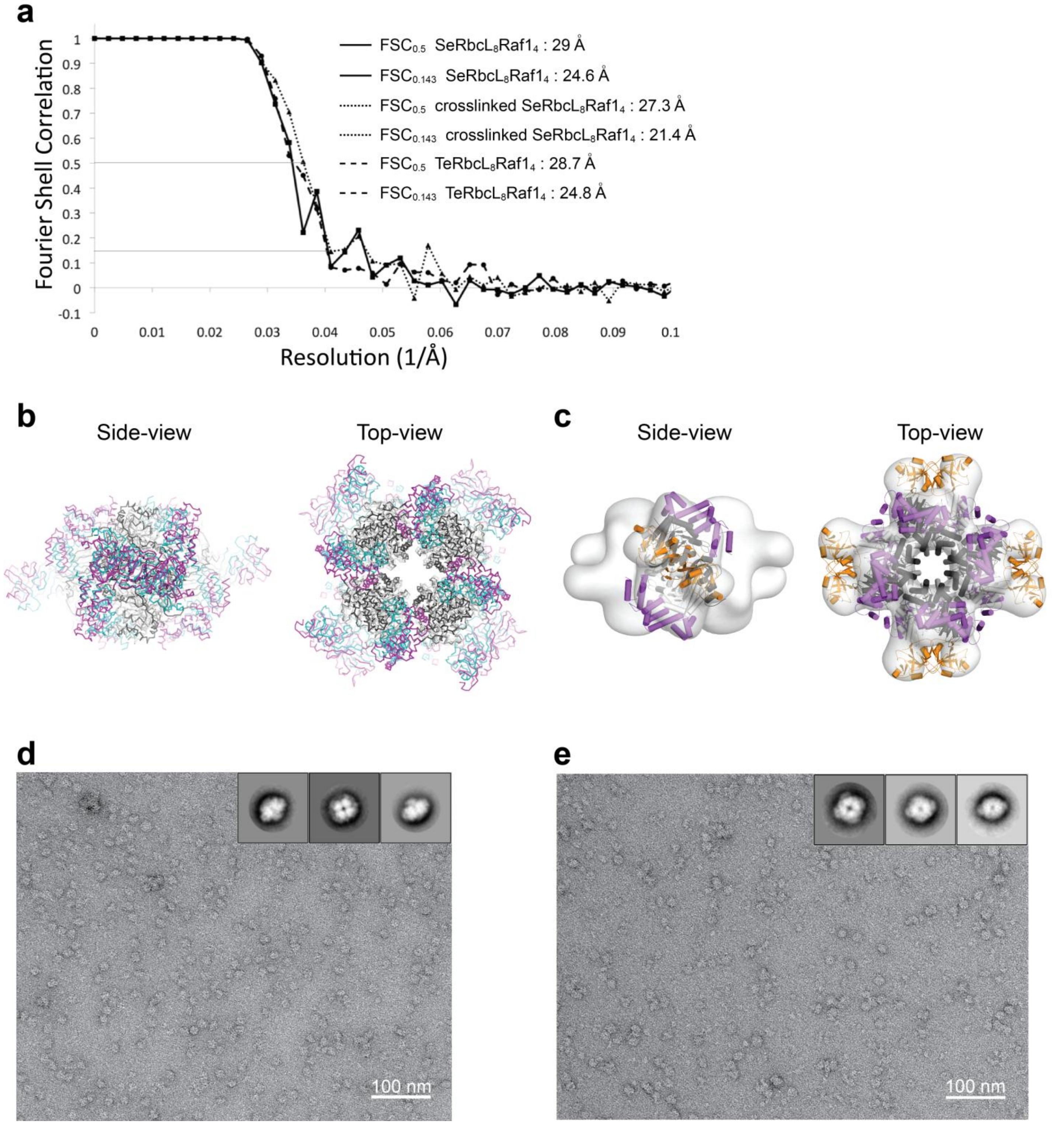
(**a,b**) Experimental electron density maps for AtRaf1.2α and AtRaf1.2β. Representative regions are shown. The meshwork represents the isocontour surface at 1.0 σ level. The nominal resolutions of the AtRaf1.2α Pt-SIRAS and AtRaf1.2β Hg-SIRAS maps are 2.75 and 3.0 Å, respectively. Panel B shows a contact between two AtRaf1.2β dimers. (**c**) Surface properties of AtRaf1.2α. The same views as in Fig. 3c are shown. Positively and negatively charged groups are shown in blue and red, respectively. Yellow color signifies hydrophobic sidechains. (**d**) Superposition of three crystallographically independent copies of the AtRaf1.2β dimer. The models are represented as Cα traces. The orientation corresponds to the top-view in Fig. 3d.(**e**) Domain swapping in the $P2_12_12_1$ crystal lattice of AtRaf1.2β. In this lattice the long βF-βG connecting loops reach across between adjacent dimers, making contacts to a hydrophobic pit. In the *C*2 crystal form, the hydrophobic residues at the loop apex fold back onto the respective hydrophobic area of the same chain, realizing an analogous intramolecular contact. Outside of the crystal lattice the conformation observed in the *C*2 crystal form should be strongly favored. (**f**) Topology of the secondary structure in the AtRaf1.2β dimer. α-Helices and β-strands are represented by cylinders and arrows, using the same color scheme as in the main text. The monomer shown in orange differs from the second by insertion of helix 12. (**g**) Features of the AtRaf1.2β dimer interface. On the left, the surface properties of the area facing the RbcL dimer are show using the same representation as in in panel c. On the right, one monomer is shown as backbone ribbon, the other in surface representation to reveal the AtRaf1.2β dimer interface. Yellow color signifies hydrophobic sidechains.


a
Disuccinimidylsuberate (DSS):
bifunctional lysine-specific crosslinker
(1:1 isotopic mixture)
11.4 Å
X=H - DSS-H12
X=D - DSS-D12
b
Syn6301-RbcL8
Raf1
15 min, 25°C
DSS crosslinking 30 min, 25°C
Quench NH4HCO3
Separation on 4-12% SDS-PAGE
In-gel trypsin digestion
LC-MS/MS
Program StavroX to identify crosslinked peptides (b- and y-ions assignment)
Manual validation of crosslinked petides (presence of two signals of equal intensity, separated by mass difference of the isotopes)
DSS-H12
DSS-D12
Mass (m/z)
Mass (m/z)
c
Raf1
RbcL8
Syn7942
Syn7002
kDa
170
130
100
70
~55
~40
~35
~25
~15
d
Syn6301-RbcL8 +
Syn7942-Raf1
Syn7002-Raf1
kDa
1
2
3
4
5
6
170
130
100
70
~55
~40
~35
e
RbcL-RbcL crosslink: K15-K460
β (aa 12-18)-K15
α (aa 448-463)-K460
A G V K D Y K
W S P E L A A A L D L W K E I K
Relative Abundance
m/z
Raf1β-RbcL crosslink: K344-K336
β (aa 344-352)-K344
α (aa 332-347)-K336
K I L D E N N L R
L E G D K A S T L G F V D L M R
Relative Abundance
m/z
Raf1β-Raf1α crosslink: K344-K188
β (aa 344-352)-K344
α (aa 187-202)-K188
K I L D E N N L R
Q K I E Q L L S D L T T S P T K
Relative Abundance
m/z

Supplementary Figure 6

Cross-linking coupled to mass spectrometry (CXMS).

(**a**) Structure of the bifunctional crosslinker disuccinimidylsuberate (DSS). The crosslinker is a 1:1 mixture of unlabeled (light; H_{12}) and deuterium labeled (heavy; D_{12}) compounds with a mass difference of 12.0753 Da. (**b**) Workflow for analysis of crosslinked protein bands marked and numbered in (c) by in-gel trypsin digestion, followed by LC–MS. (**c**) Crosslinking products of individual proteins *S. elongatus* $RbcL_8$, Syn7942-Raf1 and Syn7002-Raf1. The purified proteins (1.25 μM $RbcL_8$ and 10 μM of the respective Raf1 proteins) were incubated with H_{12}:D_{12}–DSS (2 mM) for 30 min at 25^{o}C, followed by quenching of the crosslinking reaction with NH_4HCO_3 (150 mM) and analysis by SDS-PAGE. (**d**) Crosslinking products of $RbcL_8$ (1.25 μM) with Syn7942-Raf1 or Syn7002-Raf1 (10 μM each). Boxed areas were analyzed as in (b). (**e**) Representative MS/MS spectra for the crosslinks RbcL–RbcL (Lys15–Lys460), Raf1β–RbcL (Lys344–Lys336) and Raf1β–Raf1α (Lys344–Lys188).

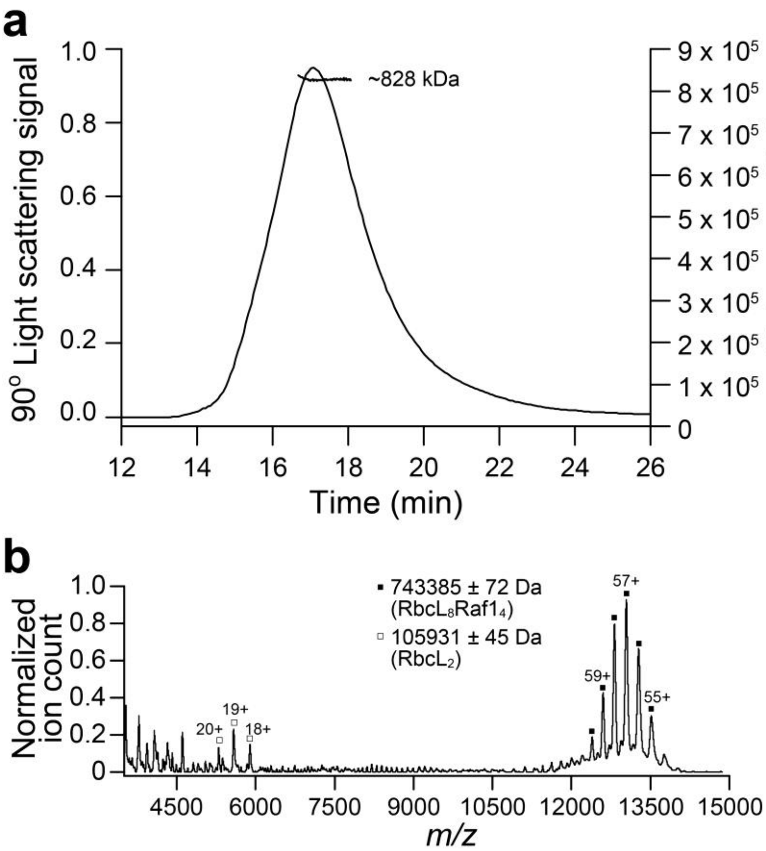


Supplementary Figure 7

Negative-stain EM analysis.

(a) Fourier Shell Correlation (FSC) curves of *S. elongatus* (Se) RbcL$_8$–Raf1$_4$, crosslinked SeRbcL$_8$–Raf1$_4$, and *T. elongatus* (Te)

$RbcL_8$–$Raf1_4$ as determined by gold standard FSC procedure in RELION-1.3. The resolution of the maps estimated by FSC with 0.5 and 0.143 correlation cut-off and no masking are given. **(b)** Comparison of the $RbcL_8$–$Raf1_4$ models derived from CXMS distance restraints (**Fig. 5g,h**) and EM reconstruction (**Fig. 6f,h,j**) (assuming the C-terminal 65 residues of RbcL are structured). The backbones are represented by Cα traces. Raf1 and RbcL in the CXMS model are shown in magenta and white, respectively. Raf1 and RbcL in the EM reconstruction are shown in cyan and gray, respectively. **(c)** Rigid body domain fitting of SeRaf1α- and β-domains and $RbcL_8$ missing the C-terminal 65 amino acids into the 3D reconstruction of $SeRbcL_8$–$Raf1_4$. RbcL subunits in gray and black; Raf1α in purple and Raf1β in orange. Side- and top-views are shown. **(d,e)** Negative stain electron micrograph of crosslinked $SeRbcL_8$–$Raf1_4$ (**d**) and of $TeRbcL_8$–$Raf1_4$ (**e**). Exemplary class averages of the respective complexes obtained from 2D classification in RELION-1.3 are shown in the insets.



Supplementary Figure 8

Characterization of the RbcL–Raf1 complex of the thermophilic cyanobacterium *T. elongatus*.

T. elongatus RbcL and Raf1 proteins were coexpressed in *E. coli* and purified as a high molecular weight complex. (**a**) SEC-MALS analysis of RbcL-Raf1 complex in solution (~40 μg). The horizontal line across the peak indicates the calculated molar mass. Note that the sample contained a small amount of aggregated protein which leads to a higher average molar mass (~828 kDa) than expected for the $RbcL_8$–$Raf1_4$ complex (~740 kDa). (**b**) nano-ESI native MS spectra of RbcL–Raf1 complex (~8 μM), Symbols indicate the charge state distributions with the charge states shown for some peaks; the calculated mass around the m/z values of the respective protein complexes is indicated. S.D. refers to the accuracy of mass values calculated from the different m/z peaks. The theoretical masses for $RbcL_8$–$Raf1_4$ and $RbcL_2$ are 741628.8 Da and 106265.4 Da, respectively.

SUPPLEMENTARY INFORMATION

Structure and mechanism of the Rubisco assembly chaperone Raf1

Thomas Hauser, Javaid Y. Bhat, Goran Miličić, Petra Wendler, F. Ulrich Hartl, Andreas Bracher

and Manajit Hayer-Hartl

SUPPLEMENTARY TABLES

Supplementary Table 1. PCR Primer Sequences for Cloning of Raf1 Constructs. The SacI (GAGCTC), SacII (CCGCGG) and NotI (GCGGCCGC) restriction sites are underlined, and the start codons are italicized.

Plasmid		**Oligos**
Syn7942raf1	Syn7942Raf1: 5’	CCGCGGTGGT*ATG*CGTGAGTTCACCCCCAC
	Syn7942Raf1: 3’	GAGCTCCTATTCTTCAAACTGCCAAGGTG
Syn7002raf1	Syn7002Raf1: 5’	CCGCGGTGGT*ATG*ATTGGACAACCGCAATCTCC
	Syn7002Raf1: 3’	GAGCTCTTAATCGTCCATTTGCCAGGGTTCC
Atraf1.1 (52-434)	AtRaf1.1: 5’	CCGCGGTGGT*CAA*CAACTCTATCAACCATTCCG
	AtRaf1.1: 3’	GAGCTCCTAATCCCAATCTTCATCATTTATC
Atraf1.2 (62-449)	AtRaf1.2: 5’	CCGCGGTGGT*CAA*CAGCTCTACCAACCGTTCCG
	AtRaf1.2: 3’	GAGCTCTCAGTCCCAGTTCTGATGACTTG
Atraf1.1/1.2	AtRaf1.1/1.2: 5’	GAGCTCAAGCTTAGATCCGGCTGCTAACAAAGCCCGTAAGAAGGAGATATACC*ATG*CAACAGCTCTACCAACCGTTC
	AtRaf1.1/1.2: 3’	GCGGCCGCTCAGTCCCAGTTCTGATGACTTGTCTGC
Syn7942raf1α (1-192)	Syn7942Raf1α: 5’	CCGCGGTGGT*ATG*CGTGAGTTCACC
	Syn7942Raf1α: 3’	GAGCTCTTAATCTAGCAGCAATGCTTCGATGAGG
Syn7942raf1β (194-356)	Syn7942Raf1β: 5’	CCGCGGTGGT*CCG*ATGCTGCCGATCTATCGACTG
	Syn7942Raf1β: 3’	GAGCTCCTATTCTTCAAACTGCCAAGGTGTTG
Syn7002raf1α (1-195)	Syn7002Raf1 α: 5’	CCGCGGTGGT*ATG*ATTGGACAACCGCAATCTCC
	Syn7002Raf1 α: 3’	GAGCTCTTAGTCACTGAGGAGTTGCTCGAT
Syn7002raf1β (197-359)	Syn7002Raf1 β: 5’	CCGCGGTGGT*CCG*CTCCTCCCCCTCTACCGTTAC
	Syn7002Raf1 β: 3’	GAGCTCTTAATCGTCCATTTGCCAGGG
Atraf1.2α (62-274)	AtRaf1.2α: 5’	CCGCGGTGGT*CAA*CAGCTCTACCAACCGTTCCG
	AtRaf1.2α: 3’	GAGCTCTCACTTCTCCTTCTCTGCTTCCTTGTC
Atraf1.2β (281-449)	AtRaf1.2β: 5’	CCGCGGTGGT*GAG*GAAGTCAAAGCTATTCGGATTCC
	AtRaf1.2β: 3’	GAGCTCTCAGTCCCAGTTCTGATGACTTG

SUPPLEMENTARY NOTE

Supplementary Data Set 1 DSS crosslinks identified in the complex of *S. elongatus* RbcL$_8$ and Raf1 (**a**) *S. elongatus* RbcL$_8$ and Syn7942-Raf1 complex. Protein band refers to the band number on SDS-PAGE in Supplementary Fig. 6d; A, experiment 1; B, experiment 2. m/z, mass over charge. The crosslinked peptides, as well as their detected and calculated mass, and the positions of the crosslinked residues are indicated. m, denotes oxidized methionine. Note that the N-terminal amino acid of a protein can be crosslinked with DSS. See Online Methods and Supplementary Fig. 6 for details. (**b**) Consolidated list of crosslink sites from Supplementary Data Set 1a. Domain categories: N, N-terminal residues 1–11 in Syn7942-Raf1 and residues 1–147 in *S. elongatus* RbcL; TIM, TIM-barrel residues 148–410 in *S. elongatus* RbcL; C, C-terminal residues 411–472 in *S. elongatus* RbcL; α, α-domain residues 12–192 in Syn7942-Raf1; β, β-domain residues 202–342 in Syn7942-Raf1; linker, residues 193–201 in Syn7942-Raf1. Crosslink distances between RbcL residues indicate the intra- or intermolecular distance in RbcL$_2$ in context of the RbcL$_8$ core complex. Crosslink distances between RbcL and Raf1 residues refer to the distances of the respective residues in the structural model derived from crosslinking data (assuming the C-terminal 65 residues of RbcL are structured) or EM reconstruction (assuming the C-terminal 65 residues of the RbcL subunit are either structured or disordered). Distances of Raf1 to the adjacent RbcL and the next closest RbcL are reported. Crosslink distances between Raf1 residues indicate the intra- or intermolecular distance in the Raf1 dimer in the structural models derived from crosslinking data (assuming the C-terminal 65 residues of RbcL are structured) or EM reconstruction (assuming the C-terminal 65 residues of the RbcL subunit are either structured or disordered). [1], RbcL residues in the flexible N- or C-terminal sequence precluding distance assignment. [2], Raf1 residues in the flexible N-terminal sequence and interdomain linker precluding distance assignment. *, denotes a crosslink between Raf1 α-domains that cannot occur in the context of the structural models of the RbcL$_8$Raf1$_4$ complex and must be due to the dynamic nature of the interaction. (**c**) *S. elongatus* RbcL$_8$ and Syn7002-Raf1 complex. Protein band refers to the band number on SDS-PAGE in Supplementary Fig. 6d; A, experiment 1; B, experiment 2. m/z, mass over charge. The crosslinked peptides, as well as their detected and calculated mass, and the positions of the crosslinked residues are indicated. m, denotes oxidized methionine. Note that the N-terminal amino acid of a protein can be crosslinked with DSS. See Online Methods and Supplementary Fig. 6 for details. (**d**) Consolidated list of crosslink sites from Supplementary Data Set 1c. Domain categories: N, N-terminal residues 1–14 in Syn7002-Raf1 and residues 1–147 in *S. elongatus* RbcL; TIM, TIM-barrel residues 148–410 in *S. elongatus* RbcL; C, C-terminal residues 411–472 in *S. elongatus* RbcL; α, α-domain residues 15–195 in Syn7002-Raf1; β, β-domain residues 205–345 in Syn7002-Raf1; linker, residues 196–204 in Syn7002-Raf1. Crosslink distances between RbcL residues indicate the intra- or intermolecular distance in RbcL$_2$ in context of the RbcL$_8$ core complex. Intermolecular distances of RbcL to the next closest RbcL of a RbcL$_2$ are also reported. Crosslink distances between RbcL and Raf1 residues refer to the distances of the respective residues in the structural model derived from crosslinking data (assuming the C-terminal 65 residues of RbcL are structured). Distances of Raf1 to the adjacent RbcL and the next closest RbcL are reported. Crosslink distances between Raf1 residues indicate the intra- or intermolecular distance in the Raf1 dimer in the structural model derived from crosslinking data (assuming the C-terminal 65 residues of RbcL are structured). [1], RbcL residues in the flexible N- or C-terminal sequence precluding distance assignment. [2], Raf1 residues in the flexible N-terminal sequence and interdomain linker precluding distance assignment. (**e**) Reconstitution of SeRbcL in the presence Syn7942-Raf1. Protein band refers to the band number on SDS-PAGE. A, experiment 1; B, experiment 2. m/z, mass over charge. The crosslinked peptides, as well as their detected and calculated mass, and the positions of the crosslinked residues are indicated. m, denotes oxidized methionine. Note that the N-terminal amino acid of a protein can be crosslinked with DSS. See Online Methods and Supplementary Fig. 6 for details. (**f**) Consolidated list of crosslink sites from Supplementary Data Set 1e. Domain categories: N, N-terminal residues 1–14 in Syn7002-Raf1 and residues 1–147 in *S. elongatus* RbcL; TIM, TIM-barrel residues 148–410 in *S. elongatus* RbcL; α, α-domain residues 15–195 in Syn7002-Raf1; β, β-domain residues 205–345 in Syn7002-Raf1; linker, residues 196–204 in Syn7002-Raf1. Crosslink distances between RbcL residues indicate the intra- or intermolecular distance in RbcL_2 in context of the RbcL_8 core complex. Crosslink distances between RbcL and Raf1 residues refer to the distances of the respective residues in the structural model derived from crosslinking data (assuming the C-terminal 65 residues of RbcL are structured). Crosslink distances between Raf1 residues indicate the intramolecular distance in the Raf1 dimer in the structural model derived from crosslinking data (assuming the C-terminal 65 residues of RbcL are structured). [2], Raf1 residues in the flexible interdomain linker precluding distance assignment.

3.4 Article 4 - Mechanism of Enzyme Repair by the AAA+ Chaperone Rubisco Activase

Rubisco is a highly inefficient enzyme fixing only ~3-10 CO_2 molecules per second. It is also prone to inhibition by premature RuBP binding and binding of misfire products (XuBP, PDBP). Rubisco activase (Rca) is a AAA+ protein that conformationally remodels Rubisco and rescues it from the inhibited state[126]. In this study, we have characterized in detail how red-type Rca from *Rhodobacter sphaeroides* repairs the corresponding Rubisco enzyme by using hydrogen/deuterium exchange coupled to mass spectrometry, chemical crosslinking coupled to mass spectrometry, cryo-electron microscopy and biochemical assays. The results of our experiments show that Rca binds transiently to the corner of inhibited Rubisco, contacting both large and small Rubisco subunits. Rca exerts its remodelling action locally on Rubisco's multi-layered active site consisting of the '60s loop of the N-terminal domain of RbcL, loop 6 of the C-terminal domain of the adjacent RbcL subunit, and the extended C-terminal tail of RbcL. Rca repairs the active site by pulling the C-terminal tail thereby breaking the C-terminal latch. This opens the active site and releases the inhibitor. Conformational remodelling by Rca is limited to the catalytic site, avoiding global destabilization of Rubisco complex.

Molecular Cell (2017)[170]

Bhat J, **Miličić G**, Thieulin-Pardo G, Bracher A, Maxwell A, Ciniawsky S, Mueller-Cajar O, Engen J, Hartl F, Wendler P, Hayer-Hartl M.

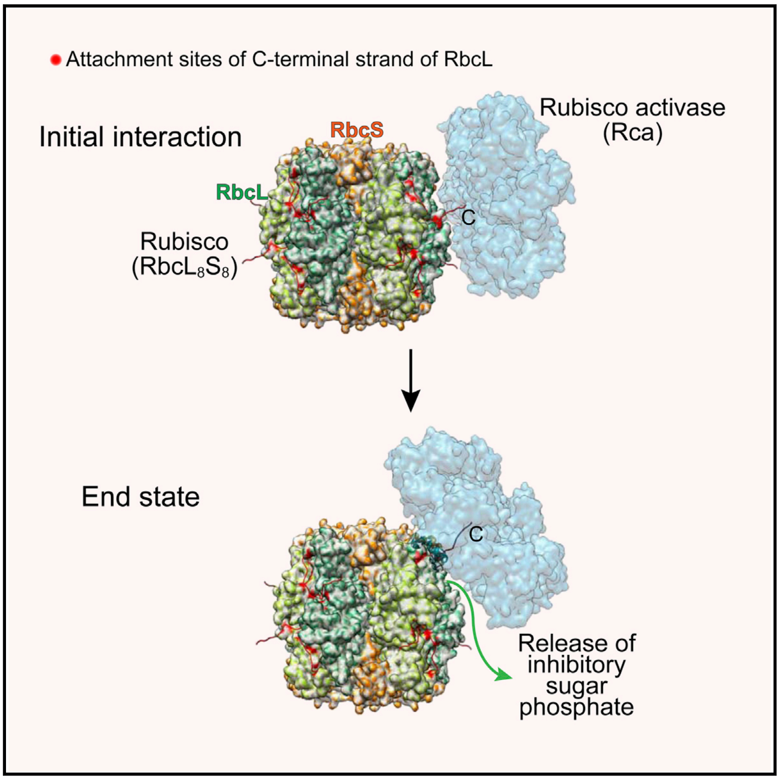
Contribution: This project was performed in collaboration with Javaid Bhat, Gabriel Thieulin-Pardo, Andreas Bracher, Andrew Maxwell, Susane Ciniawsky, Oliver Mueller Cajar, John Engen and Petra Wendler. I was involved in analysis and single particle cryo-EM 3D reconstruction of the Rubisco and Rubisco-Rca complex. I also performed the structural analysis of the Rubisco-Rca complex.

Molecular Cell

Article

Mechanism of Enzyme Repair by the AAA$^+$ Chaperone Rubisco Activase

Graphical Abstract



Highlights

- Rca contacts both large (RbcL) and small (RbcS) subunits of inactive Rubisco
- Rca docks onto Rubisco side-on, engaging the flexible C-terminal strand of RbcL
- Rca conformationally repairs one Rubisco catalytic site at a time
- Rca performs "molecular surgery" with minimal structural perturbation

Authors

Javaid Y. Bhat, Goran Miličić,
Gabriel Thieulin-Pardo, ...,
F. Ulrich Hartl, Petra Wendler,
Manajit Hayer-Hartl

Correspondence

petra.wendler@uni-potsdam.de (P.W.),
mhartl@biochem.mpg.de (M.H.-H.)

In Brief

The key photosynthetic enzyme Rubisco undergoes inhibition by substrate and non-substrate sugar phosphates. Inhibited Rubisco requires metabolic repair by the AAA+ chaperone Rubisco activase. Using an integrated approach of biochemical and structural techniques, Bhat et al. show that activase repairs the defective enzyme with remarkable precision, avoiding global structural perturbation.

Data Resources

5NV3

Bhat et al., 2017, Molecular Cell 67, 744–756
September 7, 2017
http://dx.doi.org/10.1016/j.molcel.2017.07.004

CellPress

Molecular Cell
Article

Mechanism of Enzyme Repair by the AAA$^+$ Chaperone Rubisco Activase

Javaid Y. Bhat,[1] Goran Miličić,[1] Gabriel Thieulin-Pardo,[1] Andreas Bracher,[1] Andrew Maxwell,[1,4] Susanne Ciniawsky,[2,5] Oliver Mueller-Cajar,[1,6] John R. Engen,[3] F. Ulrich Hartl,[1] Petra Wendler,[2,7,*] and Manajit Hayer-Hartl[1,8,*]
[1]Department of Cellular Biochemistry, Max Planck Institute of Biochemistry, Am Klopferspitz 18, 82152 Martinsried, Germany
[2]Gene Center Munich, Ludwig-Maximilians-Universität München, Feodor-Lynen-Strasse 25, 81377 Munich, Germany
[3]Department of Chemistry and Chemical Biology, Northeastern University, 360 Huntington Avenue, Boston, MA 02115-5000, USA
[4]Present address: University of York, Heslington, York YO10 5DD, UK
[5]Present address: Gene Synthesis and Molecular Biology, Eurofins Medigenomix, GmbH, Anzinger Strasse 7a, 85560 Ebersberg, Germany
[6]Present address: School of Biological Sciences, Nanyang Technological University, 60 Nanyang Drive, Singapore 637551, Singapore
[7]Present address: Institut für Biochemie und Biologie, Universität Potsdam, Karl-Liebknecht-Strasse 24-25, 14476 Potsdam-Golm, Germany
[8]Lead Contact
*Correspondence: petra.wendler@uni-potsdam.de (P.W.), mhartl@biochem.mpg.de (M.H.-H.)
http://dx.doi.org/10.1016/j.molcel.2017.07.004

SUMMARY

How AAA+ chaperones conformationally remodel specific target proteins in an ATP-dependent manner is not well understood. Here, we investigated the mechanism of the AAA+ protein Rubisco activase (Rca) in metabolic repair of the photosynthetic enzyme Rubisco, a complex of eight large (RbcL) and eight small (RbcS) subunits containing eight catalytic sites. Rubisco is prone to inhibition by tight-binding sugar phosphates, whose removal is catalyzed by Rca. We engineered a stable Rca hexamer ring and analyzed its functional interaction with Rubisco. Hydrogen/deuterium exchange and chemical crosslinking showed that Rca structurally destabilizes elements of the Rubisco active site with remarkable selectivity. Cryo-electron microscopy revealed that Rca docks onto Rubisco over one active site at a time, positioning the C-terminal strand of RbcL, which stabilizes the catalytic center, for access to the Rca hexamer pore. The pulling force of Rca is fine-tuned to avoid global destabilization and allow for precise enzyme repair.

INTRODUCTION

Members of the AAA+ (ATPases associated with various cellular activities) family of molecular chaperones are involved in a wide range of functions in which the energy of ATP is used to unfold, disassemble, or conformationally remodel various substrate proteins (Olivares et al., 2016; Snider et al., 2008). These hexameric ring complexes exert mechanical force by pulling extended sequences or loop segments into their central pore. How they recognize their clients and apply conformational force to effect a specific outcome is not well understood. The AAA+ protein Rubisco activase (Rca) has emerged as an important paradigm to understand these mechanisms (Bhat et al., 2017; Bracher et al., 2017; Hauser et al., 2015). Rca is required for the metabolic repair of ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco), the most abundant enzyme in nature.

Rubisco mediates the fixation of carbon dioxide from the atmosphere into organic carbon during photosynthesis by catalyzing the carboxylation of the five-carbon sugar, ribulose-1,5-bisphosphate (RuBP). This reaction is directly or indirectly responsible for the production of all biomass. The most common form of Rubisco (form 1) is highly conserved from photosynthetic bacteria to algae and plants. It consist of eight large (RbcL) and eight small (RbcS) subunits, forming a hexadecameric cylindrical complex of ~550 kDa (Figure 1A). RbcL consists of an N-terminal α+β domain of ~150 residues and a C-terminal domain of ~325 residues, which contains a $\beta_8\alpha_8$ triosephosphate isomerase (TIM)-barrel domain of ~310 residues and a flexible C-terminal tail of ~15 residues (Figure 1A). This C-terminal tail is extended to ~24 residues in the so-called red-type Rubiscos of red algae and certain photosynthetic bacteria, such as *Rhodobacter sphaeroides* (Bhat et al., 2017; Bracher et al., 2017; Hauser et al., 2015). The RbcL subunits are arranged as a tetramer of antiparallel dimers with two catalytic centers located at the interface of each dimer between the N-terminal domain of one subunit and the $\beta_8\alpha_8$ domain of the other subunit (Andersson and Backlund, 2008) (Figure 1A). Catalysis requires the active site to adopt a closed, solvent-inaccessible state (Duff et al., 2000). While the structural features of the RbcL subunits are highly conserved, the RbcS subunits are more diverse. The common core structure of the RbcS subunit is a four-stranded anti-parallel β sheet covered on one side by two helices (Andersson and Backlund, 2008). The variation among RbcS subunits is mainly in the loops between β strands A and B and at the C terminus, resulting in more or less extensive interactions with the RbcL subunits. Although the RbcS subunits do not participate directly in the formation of the active site, they are required for catalytic activity through long-range effects (Bracher et al., 2011).

Despite its pivotal role, Rubisco is remarkably inefficient as an enzyme: its catalytic rate is only two to ten CO_2 molecules fixed per second (Bracher et al., 2017). Moreover, the multistep reaction of carboxylation is error prone, producing sugar by-products, such as xylulose-1,5-bisphosphate (XuBP), that bind tightly to

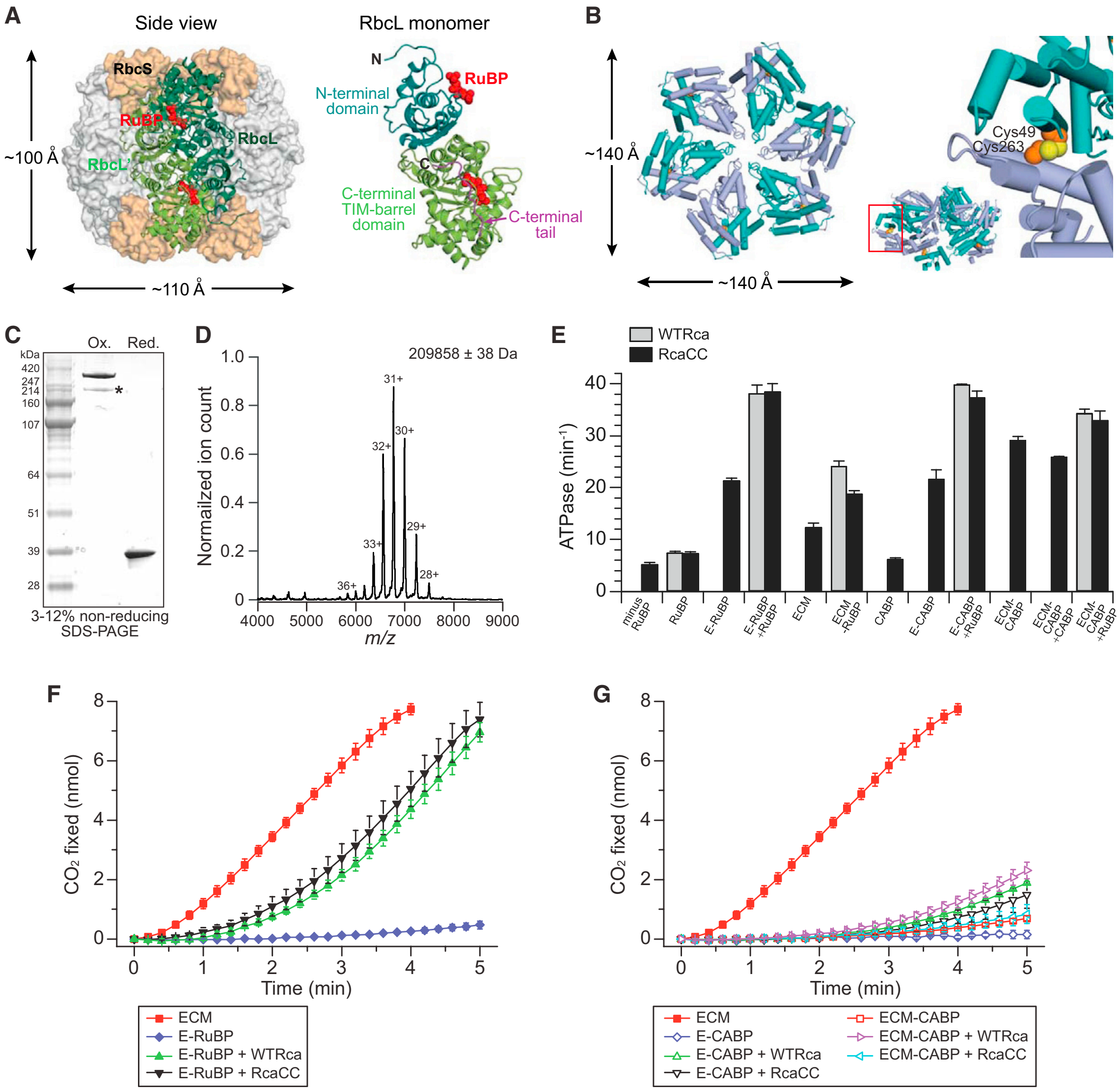


Figure 1. Structural and Functional Characterization of RcaCC

(A) Crystal structure of Rubisco (from spinach) (PDB: 1RCX) shown in side view. Subunits of one antiparallel RbcL dimer (RbcL and RbcL') are shown in dark and light green, respectively, in ribbon representation, with the substrate RuBP (red) in space-filling representation. Other subunits are depicted as molecular surfaces, with RbcL in gray and RbcS in beige. The structure of the RbcL monomer is shown on the right, with the N-terminal domain in cyan and the C-terminal TIM-barrel domain in green.

(B) Structure of the hexamer of RsRca double-cysteine mutant L49C/A263C (RcaCC) modeled based on the structure of RsRca (Mueller-Cajar et al., 2011). A top view is shown with alternating subunits in teal and light blue. The right panel zooms in on residues C49 and C263, with cysteine side chains in space-filling representation.

(C) RcaCC in oxidized (Ox.) and reduced (Red.) states analyzed by non-reducing SDS-PAGE. Asterisk indicates incompletely oxidized RcaCC.

(D) Native MS spectrum of oxidized RcaCC. The low-intensity charge states of mass/charge (m/z) values of 4,000–5,000 belong to a minor contaminant of ~70 kDa (possibly, the chaperone DnaK).

(E) ATPase rates of oxidized RcaCC and wild-type RsRca, determined in the absence or presence of sugar phosphate ligands, as indicated, and in the presence of non-carbamylated (E) or carbamylated (ECM) Rubisco from *R. sphaeroides*. 0.83 μM WTRca or RcaCC (hexamer) and 0.375 μM non-carbamylated (E) or carbamylated (ECM) Rubisco (hexadecamer) were used.

(F and G) Reactivation of inhibited Rubisco by RcaCC. CO_2 fixation assays were performed with ECM or E-RuBP (F) or E-CABP (G) Rubisco (0.02 μM hexadecamer or 0.15 μM RbcL protomer) in the absence or presence of RsRca or RcaCC (0.67 μM hexamer). See STAR Methods for details.

All SDs are from at least three independent experiments.

the active site and inhibit Rubisco function (Hauser et al., 2015; Parry et al., 2008). Inhibition of Rubisco can also occur as a consequence of the complex reaction that is necessary for catalytic activation of the enzyme. Prior to the binding of its substrates, RuBP and CO_2, Rubisco must be carbamylated by a non-substrate CO_2 molecule at the active-site lysine and bind Mg^{2+} ion as a cofactor

(Bracher et al., 2017). Upon binding of RuBP, the mobile loop 6 region of the TIM barrel and the movable C-terminal RbcL sequence form a multi-layered lid that "closes" the active site. Together with a loop sequence from residues 59 to 82 (tobacco numbering) in the N-terminal domain of the adjacent subunit (the so-called "60s loop") (Duff et al., 2000), this generates the physical environment required for electrophilic attack of RuBP by CO_2 or O_2. Premature binding of RuBP to non-carbamylated Rubisco results in a closed, inhibited enzyme (Parry et al., 2008). Reactivation of the inhibited Rubisco is mediated by Rca (Bhat et al., 2017; Bracher et al., 2017; Hauser et al., 2015). Deletion of Rca in the plant *Arabidopsis thaliana* and the proteobacterium *R. sphaeroides* results in severe photoautotrophic growth defects (Bracher et al., 2017).

Recent structural and biochemical studies show that Rca proteins from different organisms are type 1 AAA+ chaperones (Bhat et al., 2017). They all share the AAA+ domain architecture, consisting of an N-terminal α/β nucleotide-binding subdomain and a C-terminal α-helical subdomain, and function as hexamers with a central pore (Bhat et al., 2017). Residues in the central pore loops have been shown to be required for Rubisco remodeling and, in the case of red-type Rubiscos, have been implicated in exerting a pulling force on the extended C-terminal tail of the RbcL subunits (Loganathan et al., 2016; Mueller-Cajar et al., 2011; Stotz et al., 2011). The extent of conformational remodeling resulting from this action and how it causes the release of inhibitory sugar phosphate remains to be elucidated. Two geometries of Rubisco-Rca interactions have been envisioned: a "top-on" model of Rca involving substantial molecular contacts with the RbcS subunits of Rubisco and a "side-on" model with the Rca hexamer binding so that its central pore is centered over the 2-fold rotational axis of the RbcL antiparallel dimer (Wachter et al., 2013).

We engineered the Rca from the photosynthetic proteobacterium *R. sphaeroides* to form a stable, fully functional hexamer and used a hybrid approach of three techniques to analyze its transient interaction with cognate Rubisco. Hydrogen/deuterium exchange (H/DX) mass spectrometry (MS) is a biophysical method that provides information on protein structural dynamics (Harrison and Engen, 2016), while the combination of chemical crosslinking (coupled to MS) and cryo-electron microscopy with single-particle reconstruction (Leitner et al., 2016) was used to provide structural information on protein-protein interactions. All three methods showed that the conformational changes in Rubisco effected by Rca are precisely restricted to the catalytic site and its close vicinity. We find that Rca interacts with the cube-shaped Rubisco side-on, with the conserved top surface of the hexamer contacting both RbcL and RbcS subunits. This topology positions the C-terminal tail of RbcL so that it can transiently be pulled into the Rca hexamer pore, thereby catalyzing the opening of the active site. Importantly, this interaction preserves the structural integrity of the Rubisco complex, allowing Rca to repair the inhibited enzyme with minimal structural perturbation.

RESULTS

Stable Rca Hexamers

Rca from *R. sphaeroides* (RsRca) functions as a hexamer in the presence of ATP and RuBP but forms amorphous oligomers in the apo-state or fibrillar structures with ATP alone (Mueller-Cajar et al., 2011). RuBP, the substrate of the target protein Rubisco, binds to Rca as an allosteric effector of hexamer formation. To stabilize the RsRca in the hexamer state, we introduced cysteine residues at positions Leu49 and Ala263 (forming RcaCC) so as to allow inter-subunit disulfide bond formation (Figure 1B). Purified recombinant RcaCC was incubated with ATP and RuBP, followed by oxidation with copper chloride. Oxidized RcaCC migrated as a high-molecular-weight complex on non-reducing SDS-PAGE (Figure 1C). A small amount of a faster migrating species was also detected (Figure 1C), probably representing incompletely oxidized RcaCC. Size exclusion chromatography of oxidized RcaCC demonstrated a homogeneous molecule distribution consistent with that of hexamers, whereas reduced RcaCC and wild-type Rca (WTRca) populated a range of oliogomeric states (Figure S1A). Native MS determined a mass of the oxidized RcaCC of ~210 kDa, ~1.7 kDa greater than the calculated hexamer mass (208,117 Da) (Figure 1D), suggesting the presence of bound adenine nucleotide and/or RuBP. Henceforth, RcaCC refers to the oxidized hexamer complex.

RcaCC showed basal ATPase activity (~7 min^{-1}) in the absence of RuBP, consistent with the oxidized protein being independent of the allosteric effector for hexamer formation (Figure 1E). Importantly, this basal activity was stimulated by inhibited Rubisco enzyme (E-RuBP; non-carbamylated Rubisco with bound RuBP) to ~20 min^{-1}. In the presence of excess free RuBP, the ATPase activity of RcaCC increased to ~40 min^{-1}, equivalent to the ATPase activity of WTRca (Figure 1E). Indeed, RcaCC was fully functional in reactivating E-RuBP (Figure 1F) and preserved functional activity upon reduction of the disulfide bonds (Figure S1B). Active Rubisco enzyme (ECM, for enzyme with bound CO_2 at the active lysine and bound Mg^{2+}) with or without RuBP stimulated the ATPase of RcaCC to a lower level (~10–20 min^{-1}) than the inactive Rubisco (E-RuBP) (Figure 1E). The ATPase activity of RcaCC was also stimulated to high levels when the tight-binding synthetic inhibitor 2-carboxyarabinitol-1,5-bisphosphate (CABP) (Schloss, 1988) was bound to inhibited or active Rubisco enzyme (E-CABP or ECM-CABP, respectively) (Figure 1E). Note that CABP is not able to replace RuBP as allosteric effector of Rca hexamer formation, explaining why E-CABP failed to stimulate the ATPase of WTRca (Figure 1E). However, reactivation of E-CABP (measured in presence of excess RuBP) occurred only with low efficiency (Figure 1G), consistent with the much higher affinity of CABP for Rubisco (K_D, ~10^{-11} M) compared to RuBP (K_D, ~10^{-6} M) (Brooks and Portis, 1988; Pierce et al., 1980; Wang and Tabita, 1992) (see STAR Methods for details). Thus, CABP is an effective competitive inhibitor, resulting in Rca interacting continuously with Rubisco. In conclusion, RcaCC is fully active as a stable hexamer, facilitating structural and functional analysis.

Remodeling of Inhibited Rubisco by Rca

To obtain insight into the conformational changes that occur in Rubisco during remodeling by Rca, we first performed H/DX coupled to MS of intact Rubisco (Figure 2A). The RbcL of ECM incorporated 78 ± 1 deuterons (~17% of 463 possible deuteration sites) during a 60s D_2O pulse, while the RbcL from inactive E-RuBP incorporated only 57 ± 1 deuterons (Figure 2B),

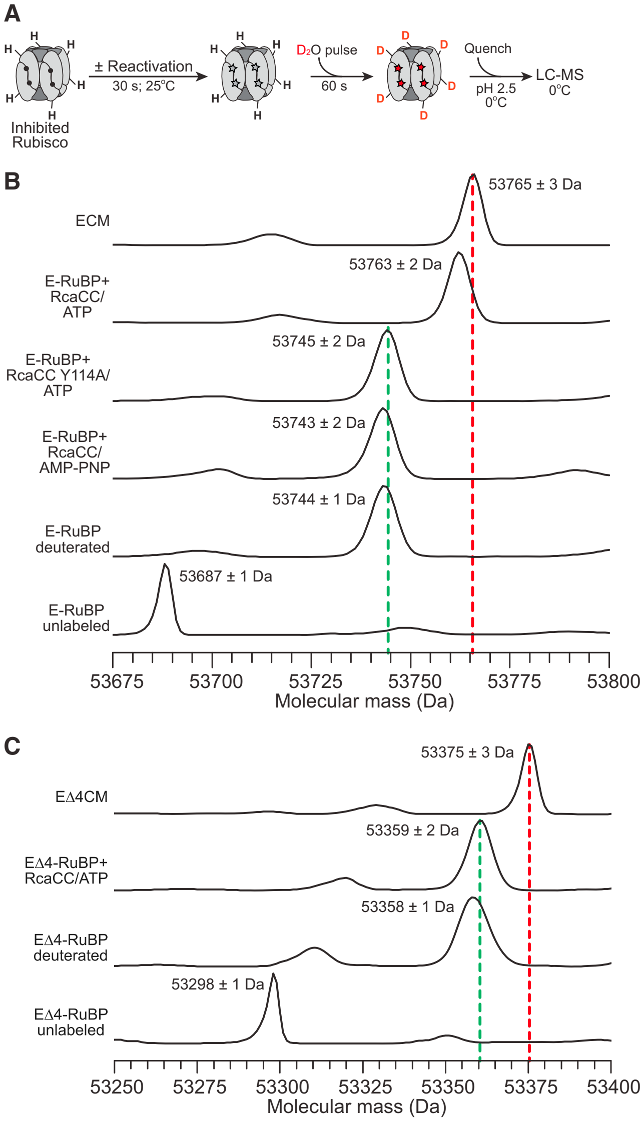


Figure 2. RcaCC-Mediated Remodeling of Inhibited Rubisco Monitored by H/DX-MS Analysis

(A) Schematic of the global H/DX-MS pulse experiment. After 30 s of reactivation of inhibited Rubisco (E-RuBP) with 8-fold excess of RcaCC (hexamer) over Rubisco catalytic sites, proteins were pulse-labeled with D_2O buffer for 60 s, followed by acid quenching and LC-MS analysis. See STAR Methods for details. H, hydrogen; D, deuterium.

(B and C) Mass spectra showing deuterium uptake in RbcL (B) and RbcLΔ4 (C) from reactions containing RcaCC or RcaCC Y114A and ATP or AMP-PNP (10 mM) as indicated. The positions of RbcL from the carbamylated Rubisco (ECM; EΔ4CM) and inhibited Rubisco (E-RuBP; EΔ4-RuBP) are indicated by red and green dotted lines, respectively. E-RuBP and EΔ4-RuBP not subjected to D_2O pulse are shown as controls. Mass values are averages of five independent experiments.

consistent with active Rubisco being more dynamic than the inhibited enzyme. After reactivation of E-RuBP with excess RcaCC in the presence of ATP, the RbcL showed essentially the same deuterium incorporation as the active ECM state (Figure 2B), indicating that Rca does not increase the flexibility of the RbcL subunit beyond that in the functional enzyme. Remodeling of E-RuBP was not observed in the presence of the non-hydrolysable ATP analog, AMP-PNP, or with the RcaCC pore loop mutant Y114A (Mueller-Cajar et al., 2011) (Figure 2B), confirming that ATP hydrolysis and the central pore of the hexamer are critical for Rca function. Similar results were obtained when Rubisco was inhibited with the misfire product XuBP (Bracher et al., 2015) (Figure S1C). In contrast, no significant increase in deuterium uptake was observed when CABP-bound Rubisco (E-CABP) was incubated with RcaCC and ATP (Figure S1D), consistent with the low efficiency of reactivation in the presence of CABP (Figure 1G).

We previously found that deletion of the four C-terminal residues from *R. sphaeroides* RbcL (RbcLΔ4) rendered the inhibited Rubisco non-activatable by Rca, while removal of the last two residues reduced reactivation by 80% (Mueller-Cajar et al., 2011). In its active state, the RbcLΔ4 (EΔ4CM) incorporated 77 ± 3 deuterons (~17% of 460 possible deuteration sites), and the inhibited EΔ4-RuBP incorporated 60 ± 2 deuterons (Figure 2C), similar to wild-type RbcL. No shift in deuterium incorporation of EΔ4-RuBP occurred in the presence of RcaCC and ATP (Figure 2C), confirming that remodeling requires Rca to engage the extreme C terminus of RbcL (Mueller-Cajar et al., 2011). As expected, RcaCC failed to reactivate EΔ4-RuBP (Figure S1E).

H/DX analysis of the RbcS subunits showed incorporation of 17 ± 1 deuterons (~14% of 120 possible deuteration sites), and no difference in deuterium incorporation between the active, inactive, and reactivated states was detected (Figure S1F). This suggests that remodeling by Rca either is restricted to the RbcL subunit or involves only flexible regions of RbcS.

Remodeling Is Restricted to the Rubisco Active Site

To identify the regions on RbcL that undergo time-dependent conformational changes during reactivation by Rca, we next monitored H/DX protection at peptide resolution. E-RuBP was incubated with RcaCC at a ratio, for Rubisco catalytic sites to RcaCC, of 1:2.5. After different times (10 s to 7 min), the reactions were labeled with a short D_2O pulse of 10 s, sufficient to label regions with low or no structure but not long enough to label structured regions (Engen and Wales, 2015). Further exchange was then inhibited by acid quench, followed by pepsin digest and analysis of peptides by liquid chromatography-tandem mass spectrometry (LC-MS/MS) (Figure 3A). Pepsin digestion produced 73 unique and overlapping peptides of RbcL and 22 peptides of RbcS, corresponding to 86% and 91.5% sequence coverage, respectively (Figures S2A and S2B). All the peptides analyzed showed unimodal exchange kinetics (i.e., a single binomial isotope distribution; Figure S2C), which is consistent with a single population of protein molecules (Engen and Smith, 2001). Observation of unimodal exchange kinetics indicates that the proteins do not undergo large folding-unfolding transitions during the deuterium pulse. Such transitions would result in a bimodal isotope pattern. Pulse labeling of E-RuBP in the absence of RcaCC resulted mainly in deuteration of N- and C-terminal peptides of RbcL (Figure S3A). Incubation with RcaCC and ATP resulted in additional deuterium incorporation in a time-dependent manner, as shown in a difference plot (Figure 3B). Saturation of exchange was reached after 2–7 min of

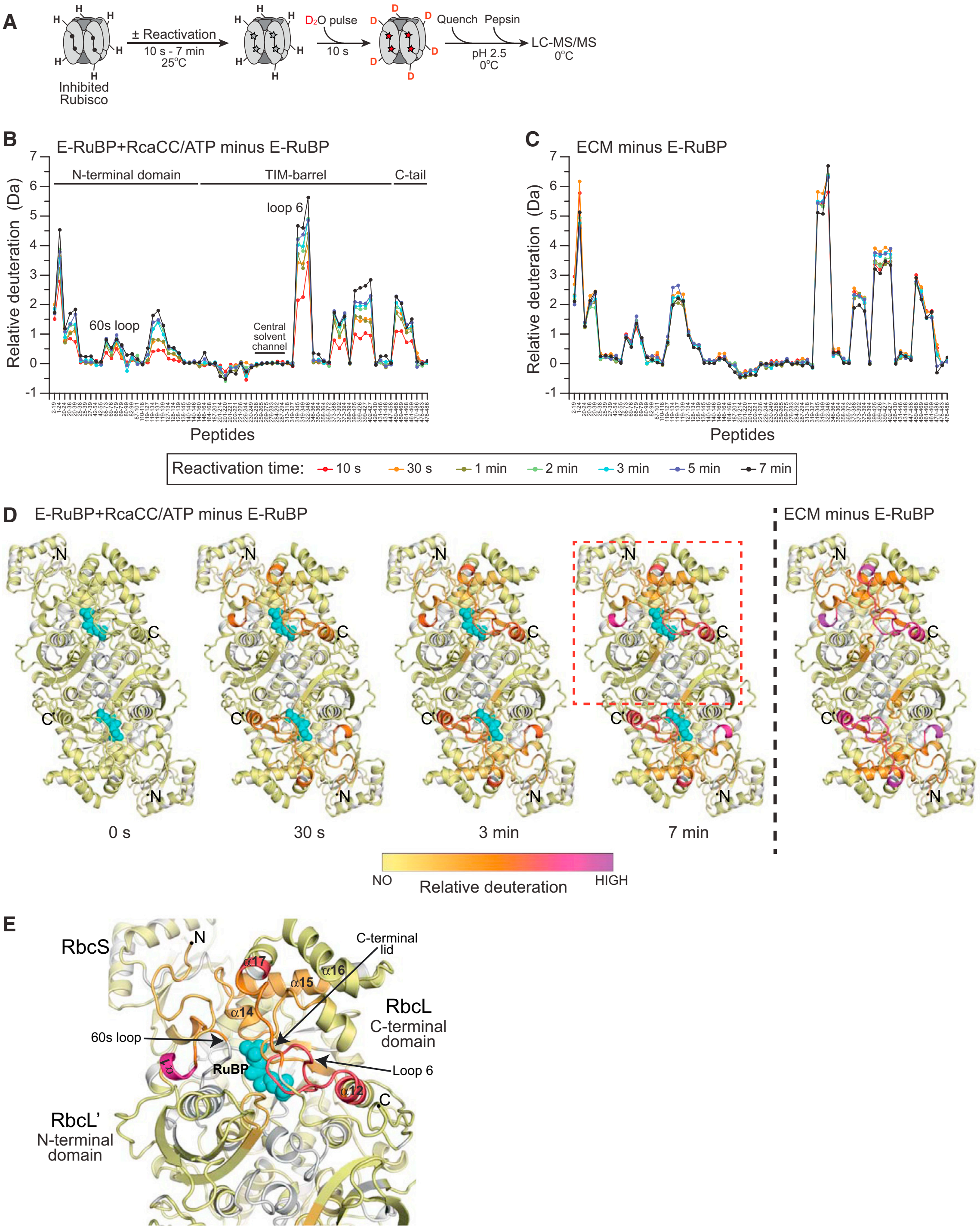


Figure 3. Deuterium Uptake into RbcL Peptides during Rubisco Activation by RcaCC

(A) Schematic of H/DX pulse experiment at peptide resolution. E-RuBP was reactivated by a 2.5-fold excess of RcaCC (hexamer) over Rubisco catalytic sites for 10 s to 7 min, followed by a 10-s D_2O pulse, acid quench, pepsin digestion, and LC-MS/MS. See STAR Methods for details. H, hydrogen; D, deuterium.

(B and C) Difference plots showing deuterium incorporation into RbcL peptides after reactivation of E-RuBP by RcaCC for the times indicated (B) and upon incubation of ECM in the absence of RcaCC (C). Deuterium incorporation measured for E-RuBP alone is subtracted. Changes in deuteration >0.5 Da are meaningful. All H/DX-MS data represent averages of four to six independent experiments (also, see related Figure S4).

(D) Time course of RbcL remodeling by RcaCC. The difference in deuterium uptake into RbcL peptides shown in (B) was mapped onto a homology model of the RbcL antiparallel dimer based on the crystal structure of CABP-bound Rubisco from *Galdieria partita* (PDB: 1BWV). A color gradient from pale yellow (no change in deuterium uptake relative to E-RuBP) to magenta (high relative deuterium uptake) indicates time-dependent remodeling, with the end state having deuteration

(legend continued on next page)

activation, dependent on the specific peptide analyzed (Figure S4), with the end state having the same H/DX properties of RbcL as in the active enzyme (Figures 3C, 3D, and S3B). Remarkably, increased deuterium incorporation during activation by RcaCC occurred exclusively in peptides that mapped to the solvent-accessible face of the Rubisco active site: in the N domain, the flexible N terminus (1–19), helix $\alpha1$ (residues 20–24), residues of the so-called 60s loop (residues 68–79), and helix $\alpha4$ to strand $\beta4$ (residues 119–139); in the C domain, loop 6 (residues 319–346), residues 373–394 and 399–427 of the TIM barrel, and the C terminus (residues 459–469) (Figure 3B). Mapping these regions on the $RbcL_2$ unit illustrates the time-dependent deprotection during activation (Figures 3D and 3E). Mobilization of loop 6, the C-terminal lid, and the 60s loop would reflect the opening of the active site (Figure 3E). Notably, peptides of the central solvent channel (RbcL residues 253–294) of the Rubisco complex remained protected to deuterium incorporation during reactivation (Figure 3B).

The deuteration properties of RbcLΔ4 Rubisco were the same as for the wild-type enzyme when the active (EΔ4CM-RuBP) and inactive (EΔ4-RuBP) forms were compared (Figures 3C and S3C). However, no additional deuterium incorporation was observed upon incubation of the inactive EΔ4-RuBP with RcaCC and ATP (Figure S3D), confirming that engagement of the RbcL C terminus is critical for remodeling. RbcS peptides did not show significant changes in deuteration between the active and inactive states of Rubisco (data not shown), consistent with the H/DX analysis of the intact protein (Figure S1F).

These results indicate that the remodeling of RbcL by Rca is restricted to active-site regions that are conformationally dynamic during the functional open-to-closed transition of the enzyme (Duff et al., 2000). Thus, Rca faithfully restores the inhibited Rubisco to a state identical in structural dynamics to that of the active enzyme, avoiding global destabilization of RbcL in the process.

Analysis of the Rubisco-Rca Interface

Having established the conformational changes in Rubisco that are associated with Rca-mediated reactivation, we next used chemical crosslinking coupled to MS (CXMS) to analyze the interacting regions of Rubisco and Rca. We added the lysine-specific crosslinker disuccinimidyl suberate (DSS) to reactions containing RcaCC and Rubisco inhibited with CABP (ECM-CABP), with or without ATP and RuBP. CABP was used as inhibitor in these experiments because it cannot be metabolized by Rubisco; thus, RcaCC continuously works on Rubisco, increasing the probability of capturing the transient RcaCC-Rubisco interaction. Crosslinking was stopped after 10–30 min, and crosslinked products were visualized by native-PAGE analysis (Figure 4A). Three discrete high-molecular-weight (HMW) bands representing crosslinked complexes were detected only in the presence of ATP (asterisks in Figure 4A, lanes 6 and 7). Note that ECM-CABP and RcaCC migrate on native PAGE at the same position in the absence of DSS (Figure 4A, lanes 1, 3, and 5), although they are very different in mass. When treated with DSS, ECM-CABP migrates faster and RcaCC runs slower, as lysine modification will change the charge property of the protein (Figure 4A, lanes 2 and 4).

Next, we analyzed the crosslinking reactions by native MS. The mass of ECM-CABP alone when incubated with DSS was ~555 kDa and that of RcaCC was ~212 kDa (Figure 4B). The crosslinking reaction of RcaCC and Rubisco in the presence of ATP contained additional complexes of ~768, ~978, and ~1,223 kDa, corresponding to Rubisco with one, two, or three RcaCC hexamers bound, respectively (Figure 4B). These complexes were not observed in the absence of ATP (Figure S5A), consistent with the observation by native PAGE (Figure 4A).

To identify the crosslink sites between Rubisco and RcaCC, we used a 1:1 H_{12} and D_{12} isotopic mixture of DSS, which produces peptide doublets of near-equal intensity separated by a mass of 12 Da in MS (Figure S5B). The crosslinked HMW protein complexes containing one, two, or three RcaCC hexamers bound to Rubisco were excised from native PAGE, subjected to in-gel digestion, and analyzed by LC-MS/MS (Figures S5C and S5D) (Sinz, 2006). We identified 57 unique crosslinked peptide pairs from two independent experiments (Tables S1A and S1B): 23 intra- and intermolecular crosslinks for RcaCC, 7 for RbcL, and 1 for RbcS; 19 intermolecular crosslinks for RbcL-RcaCC; and 7 for RbcS-RcaCC. Based on the hexameric structural model of RsRca (Mueller-Cajar et al., 2011) and a structural model of the *R. sphaeroides* $RbcL_8S_8$ holoenzyme, crosslinks within RcaCC and within Rubisco generally satisfied the plausible $C\alpha$-$C\alpha$ distance spanned by DSS (~14–36 Å) (Leitner et al., 2016) (Table S1B).

Multiple RbcL-RcaCC and RbcS-RcaCC crosslinks were identified at least twice (Figure 4C; Table S1B), suggesting that Rca contacts both RbcL and RbcS subunits of the inhibited Rubisco. Lysines 99, 123, and 127 of each RcaCC subunit, which form a charge cluster on the top surface of the hexamer, crosslinked with residues in both RbcL and RbcS (Figure 4C). The amino-group of Met1 at the flexible N terminus of RcaCC also made crosslinks to RbcL and RbcS. Lysines 123 and 127 are conserved and belong to the polar face of helix $\alpha5$ of Rca (Figure 4C). The corresponding crosslinked residues in Rubisco located to the N-terminal domain (K4, K9, K15, K20, and K133) and the C-terminal domain (K306) of RbcL, as well as to the adjoining RbcS subunits (K62 and K73) (Figure 4C). Note that the RbcS sequence contains only these two lysines. Similar crosslink pairs were identified, regardless of whether complexes containing one, two, or three RcaCC hexamers per Rubisco were analyzed (Figure S5D; Table S1B). Overall, the crosslinking footprint suggests that the RsRca hexamer docks onto Rubisco side-on over one or both active sites of an antiparallel RbcL dimer, making extensive contacts with RbcL and RbcS. This

---

properties of the active, carbamylated Rubisco (ECM) (shown as control). Regions for which no peptide coverage was obtained are shown in gray. RuBP is shown in cyan, using space-filling representation. Positions of the N and C termini are indicated.

(E) The boxed area in (D) is magnified to demonstrate that deuterium uptake by RbcL during reactivation occurs in peptide regions close to the catalytic site pocket, including the 60s loop, loop 6, and the flexible C-terminal lid.

See also Figure S4.

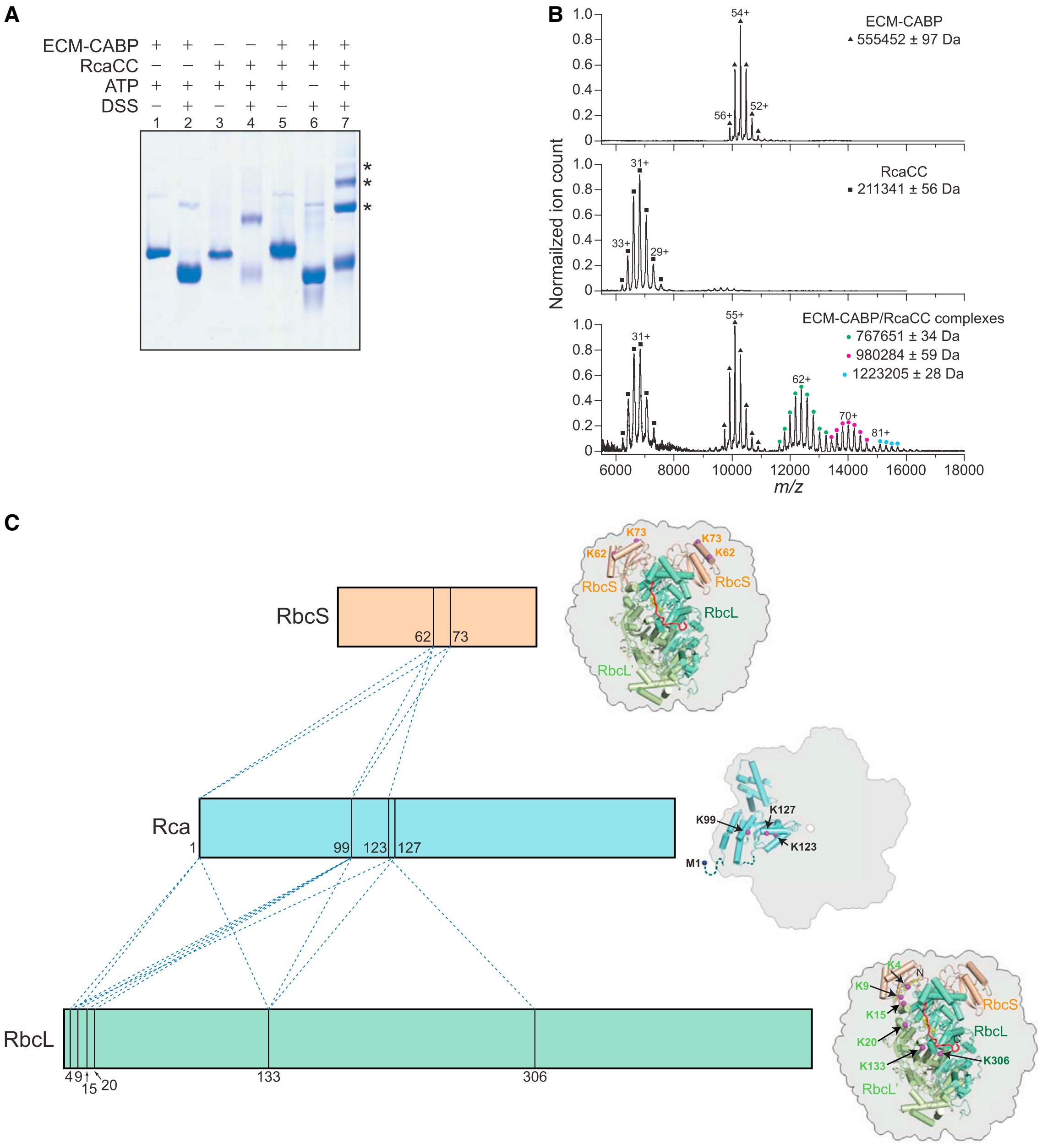


Figure 4. Analysis of the Rubisco-RcaCC Interface by Chemical Crosslinking

(A) Rubisco with bound inhibitor CABP (ECM-CABP) was preincubated for 1 min with RcaCC/RuBP and ATP when indicated. RuBP was present in all reactions. Reactions were then incubated for 30 min with or without DSS crosslinker, followed by analysis by native PAGE. High-molecular-weight complexes representing crosslinked species are marked with asterisks. See STAR Methods for details.

(B) Nano-ESI native-MS spectra of DSS-crosslinked (10 min at 25°C) ECM-CABP and RcaCC/RuBP/ATP and of ECM-CABP/RcaCC complexes obtained in the presence of RuBP and ATP. Symbols indicate charge-state distributions. Calculated mass around m/z values and accuracy of mass values are indicated. The charge state distributions marked with green, magenta, and blue symbols in the bottom panel represent ECM-CABP complexed to one, two, or three RcaCC hexamers, respectively. Nano-ESI, nano-electrospray ionization.

(C) Crosslinks from RcaCC to Rubisco RbcL and RbcS subunits. Crosslinks are shown schematically along the protein sequences on the left. Crosslinked amino acids are indicated (see Table S1B). The locations of crosslinking sites in the 3D structures are shown on the right. Protein complexes are shown in outline; one subunit of Rca and a dimer of RbcL (RbcL is indicated in olive green; RbcL' is indicated in light green) with two RbcS subunits (orange) bound are shown in cartoon representation (homology model based on the *G. partita* Rubisco structure; PDB: 1BWV); helices are represented as cylinders. Cα atoms of crosslinked Lys residues are indicated by magenta spheres, and the N-terminal Met is indicated in dark blue. Disordered residues are indicated by dotted lines. For clarity, crosslinked residues are shown only in one half of the RbcL dimer. The C-terminal peptide of RbcL is shown in red, and bound CABP is shown in yellow.
See also Table S1B.

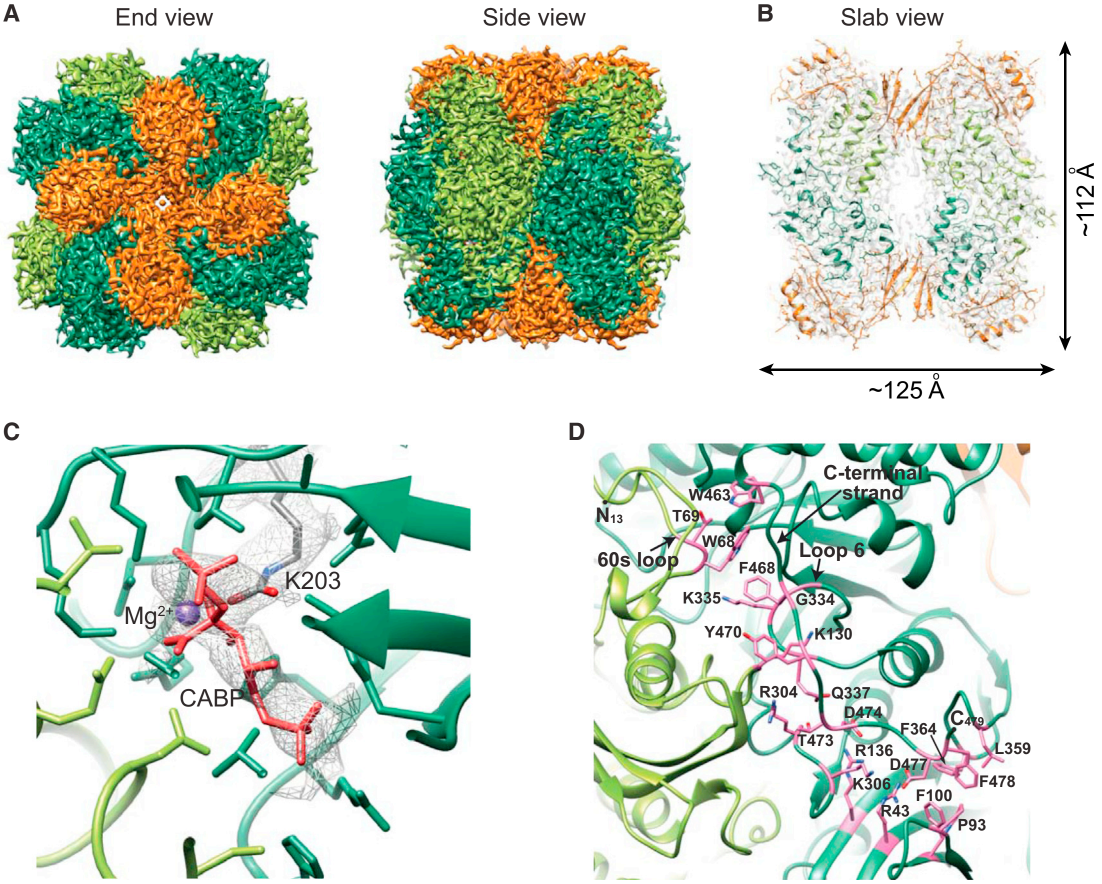


Figure 5. Cryo-EM and 3D Reconstructions of *R. sphaeroides* Rubisco

(A) 3D reconstruction of ECM-CABP complexes at 3.4-Å resolution shown as end and side views. D4 symmetry was applied during 3D reconstruction. Densities for RbcL and RbcS subunits are indicated in green and orange, respectively.

(B) 15-Å slice through the ECM-CABP complex with D4 symmetry with the sequences of *R. sphaeroides* RbcL and RbcS fitted into the electron density. The EM reconstruction is shown as transparent surface, and the model is colored as in (A).

(C and D) Cryo-EM density of the catalytic site pocket of Rubisco, showing density for CABP and carbamylated Lys 203 (C) and the interactions of the C-terminal strand with the body of the closed enzyme (D). Interacting residues are shown in stick representation in pink (Trp463 interacts with Trp68 and Thr69 of the 60s loop; Phe468 interacts with Trp68 of the 60s loop and Gly334 and Lys335 of loop 6; Tyr470 interacts with Lys130; Thr473 interacts with Arg304; Asp474 interacts with Arg136 and Glu337; Asp477 interacts with Arg43 and Lys306; Phe477 interacts with Pro93, Phe100, Phe364, and Leu359).

topology is in agreement with the conformational changes observed in the active-site region by H/DX analysis. A major contact region on RsRca is helix α5, consistent with mutation K123A resulting in an ∼80% loss of activase function (Mueller-Cajar et al., 2011).

EM Analysis of Rubisco and Rubisco-Rca Complex

To obtain additional structural information on the interaction between Rubisco and Rca, we performed negative-stain electron microscopy (EM) and cryo-EM with single-particle image analysis. In preliminary experiments, we failed to detect significant numbers of Rubisco-RcaCC complexes in the presence of ATP and/or ATP analog, consistent with the transient nature of the interaction (apparent K_d = 3.1 μM ± 0.8 μM) (Mueller-Cajar et al., 2011). To capture the interaction, reactions containing ECM-CABP, RcaCC, and ATP were incubated with the crosslinker glutaraldehyde (GA), which has a shorter crosslinking distance than DSS. Crosslinked HMW complexes, similar to those observed with DSS (Figures 4A and S6A), were enriched by size exclusion chromatography (Figure S6B). 2D class averages of the negatively stained complexes (Figure S6B, fraction 12) showed either one or two RcaCC hexamers bound to Rubisco (∼60% and ∼40% of particles, respectively) (Figures S6C and S6D). The two Rcas were bound to either adjacent or opposing surfaces of Rubisco (Figure S6D). In rare cases, up to three RcaCC hexamers were bound, as indicated by native-MS analysis (Figure 4B).

Structure of Inhibited Rubisco

Next, we performed cryo-EM analysis of fraction 13 from the size exclusion chromatography (Figure S6B), which contained free Rubisco and complexes with mainly one RcaCC bound (Figure S6E). The dataset was subdivided accordingly based on visual identification of 2D class averages (Figures S6F and S6G). To

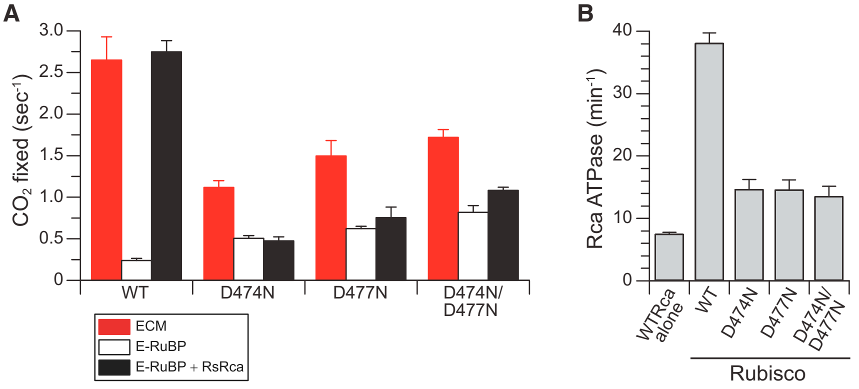


Figure 6. Attachment of RbcL C-Terminal Strand to the Body of Rubisco

(A) Carboxylation activity of activated RsRubisco (ECM) of wild-type (WT) and mutants D474N, D477N, and D474N/D477N, as well as reactivation of inhibited WT and mutant Rubisco (E-RuBP) with or without RsRca. Carboxylation and reactivation assays were performed as in Figures 1F and 1G.

(B) ATPase activities of Rca upon incubation with mutant or WT Rubisco are indicated. ATPase assays were performed as in Figure 1E.

All SDs are from at least three independent experiments.

obtain insight into the structure of inhibitor-bound *R. sphaeroides* Rubisco, we first solved the structure of Rubisco with bound CABP at a resolution of ∼3.4 Å, applying D4 symmetry, and at ∼3.5 Å, without imposing symmetry (Figures 5A and S7A–S7C). The sequences of *R. sphaeroides* RbcL and RbcS were manually built into the B-factor sharpened electron density, starting from the crystal structure of the highly homologous red-type Rubisco of *Galdieria partita* with bound CABP (PDB: 1BWV; see STAR Methods for details). Overall, the structure is closely similar to known crystal structures of other red-type Rubiscos, with root-mean-square deviations (RMSDs) of 0.683–0.800 Å for RbcL and 0.624–0.873 Å for RbcS, indicating that GA crosslinking did not result in structural distortions.

The model encompasses residues 13–479 of the 486-residue RbcL and the complete sequence (residues 1–129) of RbcS. Amino-acid side chains are generally well resolved, with the exception of residues 13–17 and 462–479 at the N and C termini of RbcL, respectively (Figures 5B and S7C). A close-up of the active site shows that Lys203 is carbamylated and that density for bound CABP is clearly discernible (Figure 5C). The C-terminal strand (residues 462–479) that pins down loop 6 over the catalytic site appears to be dynamic, exhibiting higher B factors (Figure S7D). B-factor and refinement statistics suggest an atomic occupancy of ∼40% for the C-terminal residues in the closed conformation (see STAR Methods), mediated by a series of van der Waals, electrostatic, and hydrogen-bond contacts from residues Trp463, Phe468, Tyr470, Thr473, Asp474, Asp477, and Phe478 (Figure 5D). Asp474 is conserved in all form I Rubiscos and is positioned to form a salt bridge with Arg136 and van der Waals contacts with Glu337. These interactions are thought to provide a major anchor point for the C-terminal strand (Figure 5D) (Duff et al., 2000). Interestingly, the extended C-terminal sequence of red-type Rubisco has an additional conserved Asp-Phe motif (residues 477 and 478), which contributes to stabilizing the C-terminal strand in the closed state. Asp477 makes electrostatic contacts with Arg43 and Lys306, while Phe478 engages in extensive van der Waals or π-π stacking interactions with Pro93, Phe100, Leu359, and Phe364 (Figure 5D). The ultimate C-terminal residues 480–486 are disordered and are not included in the structural model. These residues are critical for Rca-mediated remodeling (Figures 2C and S3D).

To investigate whether anchoring the C-terminal strand of RbcL to the body of the enzyme is critical for Rca to engage the RbcL C terminus, we mutated residues Asp474 and Asp477 to asparagine individually and in combination. The three Rubisco mutants had an ∼30%–60% reduced CO_2 fixation activity compared to wild-type Rubisco (Figure 6A), consistent with destabilization of the closed state of the enzyme. This is supported by an increase in spontaneous reactivation of the inhibited mutant enzymes reaching ∼50% of full activity compared to only ∼10% for wild-type Rubisco (Figure 6A). Strikingly, no significant increase in reactivation by WTRca was observed for any of the mutants (Figure 6A). Moreover, the Rubisco mutants only partially stimulated the ATPase activity of Rca (Figure 6B), to a value similar to that observed upon deletion of the two C-terminal residues of RbcL (Mueller-Cajar et al., 2011). These results suggest that an anchored C-terminal strand is required for engagement of the RbcL C terminus by Rca.

Structure of Rubisco in Complex with Rca

The structure of the Rubisco-RcaCC complex was resolved to ∼7.6 Å, with the Rca hexamer ring bound at one corner of the cube-shaped Rubisco (Figures 7A and S7A). The RcaCC density has a height of ∼43 Å, similar to the unbound hexamer (EMDB: 1932; PDB: 3ZUH), but its diameter exceeds that of the free hexamer (∼147 Å versus ∼110 Å) (Figure 7A). Accordingly, rigid-body docking of the RsRca hexamer model from the crystal structure resulted in unoccupied density in the periphery and also showed RsRca protruding from the density in several places (Figure 7A). This suggests that the Rubisco-crosslinked RcaCC represents an average of somewhat different topologies, which could not be separated by classifications. Structural heterogeneity can occur within the substrate-engaged ATPase ring and by slightly varying attachment points of RcaCC on the Rubisco complex. The vast majority of complexes adopts a binding mode where the central pore of the hexamer is positioned over the last C-terminal helix of RbcL (helix α17), from which the flexible C-terminal tail that covers the active site emanates (Figure 7B). This position may represent the end state of Rca action after the central pore has engaged the C-terminal tail. Determination of the local resolution of the map reveals lower resolution of structural elements adjacent to the contact area between both subcomplexes (Figure S7E), suggesting structural re-arrangements.

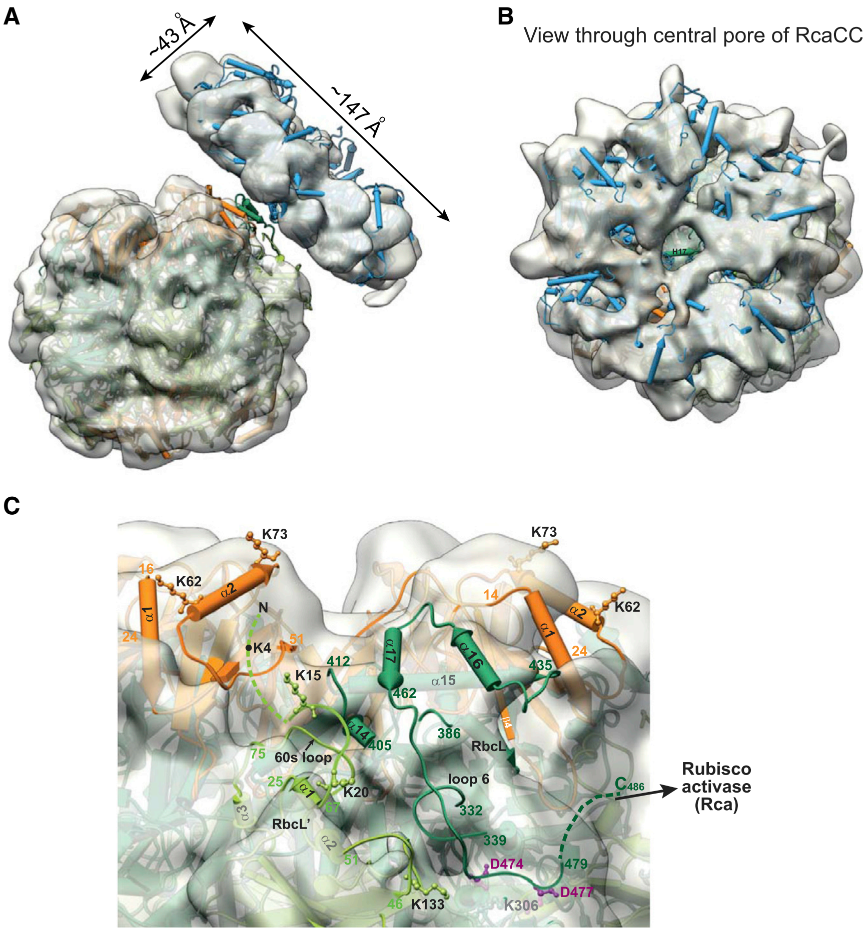


Figure 7. Cryo-EM and 3D Reconstructions of Rubisco-RcaCC Complex

(A) 3D reconstruction of the complex of inactive Rubisco (ECM-CABP) with RcaCC hexamer bound at ~7.6-Å resolution shown as side view, with RsRubisco and RsRca (PDB: 3ZUH) fitted as rigid bodies.

(B) End view of RcaCC bound to Rubisco as shown in (A). The C-terminal helix α17 of RbcL shown in green is visible through the central pore of RcaCC.

(C) Close-up view of the RbcL and RbcS subunits proximal to RcaCC. For clarity, the density map of RcaCC is removed. The flexible residues at the N and C termini of the RbcL and RbcL' subunits, not resolved in the 3.4-Å structure of Rubisco (Figure 5), are indicated by dashed lines. Structural elements outside the density map, representing regions that are conformationally destabilized by Rca, are labeled. Residues that crosslinked to RcaCC are indicated either in ball-and-stick representation or as a black circle in the unstructured N terminus of RbcL'. The positions of Asp474 and Asp477 that anchor the C-terminal strand to the body of the enzyme are shown in magenta. Arrow denotes transient pulling by Rca on the RbcL C terminus.

The Rubisco structure determined as described earlier (Figure 5A) was well accommodated by the electron density of the Rubisco-RcaCC complex, except for the regions of the RbcL dimer interacting with RcaCC (Figures 7A and 7C). Interestingly, at the interaction site, density for the ~50 C-terminal residues (encompassing helices α16, α17, and residues 462–486 of the C-terminal tail) and residues 332–339 of loop 6 that covers the active site in the closed state is missing in one of the RbcL subunits (Figure 7C). In the adjacent subunit, density for the N-terminal residues 1–23, 67–76 (60s loop), 44–51, and 130–133 is missing (Figure 7C). The 60s loop forms the major interface between the N-terminal domain of one RbcL subunit and the α/β TIM-barrel domain of the adjacent RbcL subunit and is involved in stabilizing the closed state of the catalytic site (Duff et al., 2000). The same regions were observed to undergo increased deuterium uptake in the H/DX analysis of Rubisco during reactivation by Rca (Figure 3). Moreover, the residues of RbcL that crosslinked with RcaCC are located within this destabilized interface (Figures 4C and 7C; Table S1B). The observed destabilization suggests that, in the RbcL dimer, the catalytic site proximal to RcaCC is in the open state. Such destabilization is not observed in the other catalytic sites of the Rubisco-Rca complex, indicating that Rca remodels one site at a time. The single-particle reconstruction also showed missing density in helices α1 and α2 of the two neighboring RbcS subunits (Figure 7C), consistent with the observed RbcS-RcaCC crosslinks (Figure 4C; Table S1B). Weakening the RbcL-RbcS interface would contribute to mobilizing the 60s loop and, hence, destabilize the catalytic center (Bracher et al., 2011).

To obtain a higher resolution structure of the Rubisco-RcaCC complex, we masked the outer parts of the RcaCC density during refinement. This resulted in a reconstruction at ~6.0-Å resolution (Figures S7A and S7F). We again observed destabilization of the same regions in RbcL and RbcS (Figure S7G) as for the lower resolution map (Figure 7C), confirming that the catalytic site is in the open conformation.

DISCUSSION

To understand how force generation by AAA+ chaperones can be utilized to perform specific tasks, we have analyzed the function of the AAA+ protein Rca in metabolic repair of the photosynthetic key enzyme Rubisco. H/DX and chemical crosslinking were used to define the regions of Rubisco that are conformationally remodeled by Rca and determine the sites that Rca contacts on the Rubisco complex. Cryo-EM of Rubisco in complex with Rca provided a snapshot of Rca in action. This integrated approach offered detailed insight into the mechanism by which Rca performs enzyme repair avoiding global destabilization of the Rubisco complex.

Rubisco Remodeling by Rca

Perhaps the most striking result of this study is how precisely the conformational remodeling of Rubisco by Rca is limited to the catalytic site, as demonstrated by the H/DX analysis. This contrasts with the ability of structurally similar AAA+ proteins, such as bacterial ClpX, to mediate global protein unfolding (Sauer and Baker, 2011). In the cryo-EM structure of the Rubisco-Rca complex, the Rca hexamer ring is associated at one corner of Rubisco in a tilted manner, contacting the N domain of one RbcL subunit and the C domain of the adjacent RbcL of the dimer, with the catalytic site located at the inter-subunit interface. The central pore of the hexamer is positioned over the last C-terminal helix of RbcL (helix α17), from which the C-terminal tail that covers the catalytic site emanates (Figure 7B). As shown by crosslinking, a major contact region on the flat surface of the Rca ring is helix α5, containing the functionally critical lysine residue 123 (Mueller-Cajar et al., 2011) (Figure 4C). Rca makes additional crosslink contacts to the two proximal RbcS subunits—specifically, their helices α1 and α2 (Figure 4C), which are destabilized as indicated by missing EM density in the complex (Figure 7C).

Remodeling of the closed catalytic site likely involves transient pulling on the C-terminal RbcL tail by Rca (Figure 7C). Mobilization of the tail region is supported by the H/DX data (Figure 3) and is consistent with the finding that truncation of the C-terminal two to four amino acids renders inhibited Rubisco non-activatable. The pulling force necessary to mobilize the C-terminal tail appears to be limited, considering that, in the inhibitor-bound enzyme, the C-terminal strand is already relatively dynamic (Figures 5 and S7D). This would be in line with the proposal that, when closed, the C-terminal strand is maintained under tension (Duff et al., 2000). Our structural analysis suggests that Asp474, Asp477, and Phe478, which are conserved in red-type Rubiscos, form a triad that is important in attaching the C-terminal strand in a latch-like manner over the catalytic site. Breaking this interaction might, therefore, lead to retraction of the C-terminal strand. Mutation of Asp474 and Asp477 reduced Rubisco activity and prohibited reactivation of inhibited Rubisco by Rca, suggesting that anchoring the C-terminal strand to the body of the enzyme is a prerequisite for the central pore of Rca to productively engage the RbcL C terminus. Interaction of Rca with the C-terminal strand would trigger a sequence of coupled conformational changes that ultimately lead to release of the inhibitory sugar phosphate. During initial recognition of the C terminus, the central pore of Rca is likely positioned closer to the equator of Rubisco than observed in the cryo-EM reconstruction of the complex. We, therefore, suggest that, as the C-terminal strand is being detached from the body of the enzyme and its anchor points are released, Rca moves from a side-on to a tilted position. This movement may be coupled to destabilizing the RbcS-RbcL interface that contributes to structuring residues 72–78 of the 60s loop (Bracher et al., 2011). Detachment of the C-terminal strand and destabilization of the 60s loop results in the retraction of loop 6 and opening of the catalytic site.

Limited Remodeling Versus Global Destabilization

What are the functional properties underlying Rca's ability to perform precise conformational remodeling and avoid global unfolding? It would appear that Rca interacts only transiently with its target substrate, in contrast to AAA+ unfoldases like ClpX, which act in a highly processive manner (Olivares et al., 2016). Whereas the unfoldases interact with a broad range of substrates, binding of Rca to Rubisco is specific and requires both the central pore and contacts of the hexamer ring surface with RbcL and RbcS. The grip strength of AAA+ proteins depends on their pore loop structure and on how pore loop movement is coupled to the ATPase cycle of the AAA+ module (Olivares et al., 2016; Rodriguez-Aliaga et al., 2016). The pulling force of Rca is likely lower than that of AAA+ unfoldases, consistent with the ATPase rate of RsRca in the presence of substrate being about ~7-fold lower than that of ClpX (Burton et al., 2003; Mueller-Cajar et al., 2011).

AAA+ proteins with Rca function exist in prokaryotic and eukaryotic photosynthetic organisms (Bhat et al., 2017; Bracher et al., 2017). The divergence in primary sequence of these proteins suggests that a process of convergent evolution underlies the use of the common AAA+ module in the Rubisco repair mechanism. Moreover, Rubiscos have coevolved with their cognate activases, as exemplified by the C-terminal extension in red-type RbcL. Accordingly, Rca from *R. sphaeroides* cannot activate plant Rubisco, and, conversely, plant Rca is unable to activate RsRubisco. It will be interesting to understand the mechanisms of client-specific recognition and repair by this important group of AAA+ proteins.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- METHOD DETAILS
  - Proteins
  - Preparation of Oxidized RcaCC
  - Size-Exclusion Chromatography
  - ATPase Assay
  - Rubisco Activity Assay
  - Chemical Crosslinking Coupled to Mass Spectrometry
  - Identification of Crosslinks
  - Native Mass Spectrometry
  - Hydrogen-deuterium Exchange (H/DX) and Mass Spectrometry
  - Electron Microscopy (EM) and Data Processing
- DATA AVAILABILITY
  - Data Resources

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures and one table and can be found with this article online at http://dx.doi.org/10.1016/j.molcel.2017.07.004.

AUTHOR CONTRIBUTIONS

J.Y.B. planned, performed, and analyzed the biochemical and mass spectrometric experiments. O.M.-C. and A.B. designed the disulfide-bonded Rca hexamer. G.T.-P. and A.M. performed related experiments not included in the revised manuscript. H/DX measurements were carried out in collaboration with J.R.E. EM specimen preparation and in-house data collection were carried out by S.C.; image processing was carried out by S.C., G.M., and

P.W.; and data analysis, interpretation, and presentation were carried out by A.B., G.M., and P.W. EM project supervision was provided by P.W. Project conception and overall supervision was by M.H.-H. M.H.-H. and F.U.H. wrote the manuscript, with contributions from P.W. and all other authors.

ACKNOWLEDGMENTS

We thank N. Wischnewski, R. Lange, A. Jungclaus, and A. Ries for technical assistance. We also thank R. Körner for help in MS data acquisition, D. Balchin for support with H/DX experiments, and A. Sinz (Halle-Wittenberg) for providing the StavroX software and training. P.W. was supported by the German Research Council (Emmy Noether grant WE4628/1), and M.H.-H. was supported by the Minerva Foundation. Cryo-EM data were collected at the Netherlands Centre for Electron Nanoscopy (NeCen), Leiden, the Netherlands. The use of resources of Instruct, a Landmark ESFRI project, and the Leibniz Supercomputing Centre (LRZ; https://www.lrz.de) is acknowledged. Materials used for this study are available upon request.

Received: March 9, 2017
Revised: June 7, 2017
Accepted: July 1, 2017
Published: August 10, 2017

SUPPORTING CITATIONS

The following references appear in the Supplemental Information: Schrodinger (2010).

REFERENCES

Andersson, I., and Backlund, A. (2008). Structure and function of Rubisco. Plant Physiol. Biochem. *46*, 275–291.

Bhat, J.Y., Thieulin-Pardo, G., Hartl, F.U., and Hayer-Hartl, M. (2017). Rubisco activases: AAA+ chaperones adapted to enzyme repair. Front. Mol. Biosci. *4*, 20.

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. Nat. Struct. Mol. Biol. *18*, 875–880.

Bracher, A., Sharma, A., Starling-Windhof, A., Hartl, F.U., and Hayer-Hartl, M. (2015). Degradation of potent Rubisco inhibitor by selective sugar phosphatase. Nat Plants *1*, 14002.

Bracher, A., Whitney, S.M., Hartl, F.U., and Hayer-Hartl, M. (2017). Biogenesis and metabolic maintenance of Rubisco. Annu. Rev. Plant Biol. *68*, 29–60.

Brooks, A., and Portis, A.R. (1988). Protein-bound ribulose bisphosphate correlates with deactivation of ribulose bisphosphate carboxylase in leaves. Plant Physiol. *87*, 244–249.

Brown, A., Long, F., Nicholls, R.A., Toots, J., Emsley, P., and Murshudov, G. (2015). Tools for macromolecular model building and refinement into electron cryo-microscopy reconstructions. Acta Crystallogr. D Biol. Crystallogr. *71*, 136–153.

Burton, R.E., Baker, T.A., and Sauer, R.T. (2003). Energy-dependent degradation: linkage between ClpX-catalyzed nucleotide hydrolysis and protein-substrate processing. Protein Sci. *12*, 893–902.

Chen, V.B., Arendall, W.B., 3rd, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr. D Biol. Crystallogr. *66*, 12–21.

Crowther, R.A., Henderson, R., and Smith, J.M. (1996). MRC image processing programs. J. Struct. Biol. *116*, 9–16.

Duff, A.P., Andrews, T.J., 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. J. Mol. Biol. *298*, 903–916.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr. D Biol. Crystallogr. *66*, 486–501.

Engen, J.R., and Smith, D.L. (2001). Investigating protein structure and dynamics by hydrogen exchange MS. Anal. Chem. *73*, 256A–265A.

Engen, J.R., and Wales, T.E. (2015). Analytical aspects of hydrogen exchange mass spectrometry. Annu. Rev. Anal. Chem. (Palo Alto, Calif.) *8*, 127–148.

Frank, J., Radermacher, M., Penczek, P., Zhu, J., Li, Y., Ladjadj, M., and Leith, A. (1996). SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. J. Struct. Biol. *116*, 190–199.

Götze, M., Pettelkau, J., Schaks, S., Bosse, K., Ihling, C.H., Krauth, F., Fritzsche, R., Kühn, U., and Sinz, A. (2012). StavroX—a software for analyzing crosslinked products in protein interaction studies. J. Am. Soc. Mass Spectrom. *23*, 76–87.

Grant, T., and Grigorieff, N. (2015). Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6. eLife *4*, e06980.

Guttman, M., Wales, T.E., Whittington, D., Engen, J.R., Brown, J.M., and Lee, K.K. (2016). Tuning a high transmission ion guide to prevent gas-phase proton exchange during H/D exchange MS analysis. J. Am. Soc. Mass Spectrom. *27*, 662–668.

Harrison, R.A., and Engen, J.R. (2016). Conformational insight into multi-protein signaling assemblies by hydrogen-deuterium exchange mass spectrometry. Curr. Opin. Struct. Biol. *41*, 187–193.

Hauser, T., Popilka, L., Hartl, F.U., and Hayer-Hartl, M. (2015). Role of auxiliary proteins in Rubisco biogenesis and function. Nat Plants *1*, 15065.

Kiefer, F., Arnold, K., Künzli, M., Bordoli, L., and Schwede, T. (2009). The SWISS-MODEL Repository and associated resources. Nucleic Acids Res. *37*, D387–D392.

Kubien, D.S., Brown, C.M., and Kane, H.J. (2011). Quantifying the amount and activity of Rubisco in leaves. Methods Mol. Biol. *684*, 349–362.

Leitner, A., Faini, M., Stengel, F., and Aebersold, R. (2016). Crosslinking and mass spectrometry: an integrated technology to understand the structure and function of molecular machines. Trends Biochem. Sci. *41*, 20–32.

Loganathan, N., Tsai, Y.C., and Mueller-Cajar, O. (2016). Characterization of the heterooligomeric red-type rubisco activase from red algae. Proc. Natl. Acad. Sci. U S A *113*, 14019–14024.

Mindell, J.A., and Grigorieff, N. (2003). Accurate determination of local defocus and specimen tilt in electron microscopy. J. Struct. Biol. *142*, 334–347.

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*, 194–199.

Olivares, A.O., Baker, T.A., and Sauer, R.T. (2016). Mechanistic insights into bacterial AAA+ proteases and protein-remodelling machines. Nat. Rev. Microbiol. *14*, 33–44.

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. J. Exp. Bot. *59*, 1569–1580.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. J. Comput. Chem. *25*, 1605–1612.

Pieper, U., Webb, B.M., Barkan, D.T., Schneidman-Duhovny, D., Schlessinger, A., Braberg, H., Yang, Z., Meng, E.C., Pettersen, E.F., Huang, C.C., et al. (2011). ModBase, a database of annotated comparative protein structure models, and associated resources. Nucleic Acids Res. *39*, D465–D474.

Pierce, J., Tolbert, N.E., and Barker, R. (1980). Interaction of ribulosebisphosphate carboxylase/oxygenase with transition-state analogues. Biochemistry *19*, 934–942.

Piovesan, D., Minervini, G., and Tosatto, S.C. (2016). The RING 2.0 web server for high quality residue interaction networks. Nucleic Acids Res. *44* (W1), W367–W374.

Plumb, R.S., Johnson, K.A., Rainville, P., Smith, B.W., Wilson, I.D., Castro-Perez, J.M., and Nicholson, J.K. (2006). UPLC/MS(E); a new approach for generating molecular fragment information for biomarker structure elucidation. Rapid Commun. Mass Spectrom. *20*, 1989–1994.

Rodriguez-Aliaga, P., Ramirez, L., Kim, F., Bustamante, C., and Martin, A. (2016). Substrate-translocating loops regulate mechanochemical coupling and power production in AAA+ protease ClpXP. Nat. Struct. Mol. Biol. *23*, 974–981.

Rohou, A., and Grigorieff, N. (2015). CTFFIND4: fast and accurate defocus estimation from electron micrographs. J. Struct. Biol. *192*, 216–221.

Sauer, R.T., and Baker, T.A. (2011). AAA+ proteases: ATP-fueled machines of protein destruction. Annu. Rev. Biochem. *80*, 587–612.

Scheres, S.H. (2015). Semi-automated selection of cryo-EM particles in RELION-1.3. J. Struct. Biol. *189*, 114–122.

Schloss, J.V. (1988). Comparative affinities of the epimeric reaction-intermediate analogs 2- and 4-carboxy-D-arabinitol 1,5-bisphosphate for spinach ribulose 1,5-bisphosphate carboxylase. J. Biol. Chem. *263*, 4145–4150.

Schrodinger, LLC (2010). The PyMOL molecular graphics system, version 1.3. https://www.pymol.org.

Shaikh, T.R., Gao, H., Baxter, W.T., Asturias, F.J., Boisset, N., Leith, A., and Frank, J. (2008). SPIDER image processing for single-particle reconstruction of biological macromolecules from electron micrographs. Nat. Protoc. *3*, 1941–1974.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. *13*, 2498–2504.

Shevchenko, A., Tomas, H., Havlis, J., Olsen, J.V., and Mann, M. (2006). In-gel digestion for mass spectrometric characterization of proteins and proteomes. Nat. Protoc. *1*, 2856–2860.

Sinz, A. (2006). Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein-protein interactions. Mass Spectrom. Rev. *25*, 663–682.

Snider, J., Thibault, G., and Houry, W.A. (2008). The AAA+ superfamily of functionally diverse proteins. Genome Biol. *9*, 216.

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. Nat. Struct. Mol. Biol. *18*, 1366–1370.

Trabuco, L.G., Villa, E., Mitra, K., Frank, J., and Schulten, K. (2008). Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. Structure *16*, 673–683.

Tsai, Y.C., Lapina, M.C., Bhushan, S., and Mueller-Cajar, O. (2015). Identification and characterization of multiple rubisco activases in chemoautotrophic bacteria. Nat. Commun. *6*, 8883.

van Heel, M., Harauz, G., Orlova, E.V., Schmidt, R., and Schatz, M. (1996). A new generation of the IMAGIC image processing system. J. Struct. Biol. *116*, 17–24.

Wachter, R.M., Salvucci, M.E., Carmo-Silva, A.E., Barta, C., Genkov, T., and Spreitzer, R.J. (2013). Activation of interspecies-hybrid Rubisco enzymes to assess different models for the Rubisco-Rubisco activase interaction. Photosynth. Res. *117*, 557–566.

Wales, T.E., and Engen, J.R. (2006). Hydrogen exchange mass spectrometry for the analysis of protein dynamics. Mass Spectrom. Rev. *25*, 158–170.

Wang, X., and Tabita, F.R. (1992). Reversible inactivation and characterization of purified inactivated form I ribulose 1,5-bisphosphate carboxylase/oxygenase of *Rhodobacter sphaeroides*. J. Bacteriol. *174*, 3593–3600.

Wang, L., Pan, H., and Smith, D.L. (2002). Hydrogen exchange-mass spectrometry: optimization of digestion conditions. Mol. Cell. Proteomics *1*, 132–138.

Zhang, Z., and Smith, D.L. (1993). Determination of amide hydrogen exchange by mass spectrometry: a new tool for protein structure elucidation. Protein Sci. *2*, 522–531.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, Peptides, and Recombinant Proteins		
2-carboxyarabinitol-1,5-bisphosphate (CABP)	Custom synthesized	Pierce et al., 1980
Cesium iodide	Sigma	Cat#21655
Creatine P-kinase	Sigma	Cat#C3755
Deuterium oxide	Euriso-top	Cat#D216L
Disuccinimidyl suberate (DSS)	Thermo Scientific	Cat#21655
Disuccinimidly suberate (DSS) d0/d12	Creative Molecules	Cat # 001S
Glu-fibrinopeptide	Waters, USA	Cat#700008842-2
Glutaraldehyde (GA)	Sigma	Cat#G5882
Glyceraldehyde-3-phosphate dehydrogenase	Sigma	Cat#G0763
Glycerophosphate dehydrogenase-triosephosphate isomerase	Sigma	Cat#G1881
Lactate dehydrogenase/pyruvate kinase	Sigma	Cat#P0294
NADH	Sigma	Cat#N6785
NuPAGE 3%–8% Tris-acetate Native gel	Life Technologies	Cat#EA03755BOX
Phosphocreatine	Roche	Cat#14662432
Phosphoenol pyruvate	Roche	Cat#13516325
Phosphoglycerate kinase	Sigma	Cat#P7634
ReproSil C18-AQ	Dr. Maisch, Germany	Cat#r13.aq
Ribulose-1,5-bisphosphate (RuBP)	Sigma	Cat#R0878
Sodium bicarbonate (Na14HCO3)	PerkinElmer	Cat#NEC086H005MC
TCEP	Thermo Scientific	Cat#20491
Xylulose-1,5-bisphospahte (XuBP)	Custom synthesized	Bracher et al., 2015
Deposited Data		
RsRubisco structure with CABP, derived from EMD-3699	This study	PDB: 5NV3
RsRubisco map with CABP, D4	This study	EMD-3699
RsRubisco map with CABP, C1	This study	EMD-3700
RsRubisco map with CABP and RcaCC	This study	EMD-3701
RsRubisco with CABP and RcaCC masked	This study	EMD-3702
Mendeley Data	This study	http://dx.doi.org/10.17632/vnt5ck2kbz.1
Recombinant DNA		
pET30b-RsRbcLS	Previous study	Mueller-Cajar et al., 2011
pET30b-RsRbcLΔ4S	Previous study	Mueller-Cajar et al., 2011
pHue-RsRca	Previous study	Mueller-Cajar et al., 2011
pHue-RsRcaCC	This study	N/A
pHue-RsRcaCC Y114A	This study	N/A
pET30b-RsRbcL D474N	This study	N/A
pET30b -RsRbcL D477N	This study	N/A
pET30b-RsRbcL D474N/D477N	This study	N/A
Software and Algorithms		
Chimera	Pettersen et al., 2004	http://www.cgl.ucsf.edu/chimera
Coot	Emsley et al., 2010	https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/
CTFFIND3	Mindell and Grigorieff, 2003	http://grigoriefflab.janelia.org/ctf

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
CTFFIND4	Rohou and Grigorieff, 2015	http://grigoriefflab.janelia.org/ctf
Cytoscape 3.4.0	Shannon et al., 2003	http://www.cytoscape.org/
DynamX 3.0	Waters, USA	
IMAGIC-5	van Heel et al., 1996	https://www.imagescience.de/imagic.html
MassLynx 4.1	Waters, USA	
Mass Matrix Data File Conversion Tools v.3.9		http://www.massmatrix.net
ModBase	Pieper et al., 2011	https://salilab.org/modeller
ProteinLynx Global Server 2.4	Waters, USA	N/A
PyMOL 0.99 or 1.3	N/A	http://www.pymol.org
REFMAC	Brown et al., 2015	https://www2.mrc-lmb.cam.ac.uk/groups/murshudov/content/refmac/refmac.html
RELION 1.3 or 1.4	Scheres, 2015	http://www2.mrc-lmb.cam.ac.uk/relion
Ring v.2.0.1	Piovesan et al., 2016	http://protein.bio.unipd.it/ring/
SPIDER	Frank et al., 1996	https://spider.wadsworth.org/spider_doc/spider/docs/spider.html
StavroX 3.6.0.1	Götze et al., 2012	http://www.stavrox.com
SWISS-MODEL	Kiefer et al., 2009	https://swissmodel.expasy.org
Thermo Xcalibur	Thermo Scientific	N/A
Unblur	Grant and Grigorieff, 2015	http://grigoriefflab.janelia.org/unblur
Ximdisp	Crowther et al., 1996	http://www2.mrc-lmb.cam.ac.uk/research/locally-developed-software/image-processing-software
Other		
ACQUITY UPLC C18 BEH analytical column	Waters, USA	Cat#186002346
ACQUITY UPLC C18 BEH VanGuard pre-column	Waters, USA	Cat#186003975
Micro Bio-spin 6 column	Bio-Rad	Cat#732-6221
Nano-ESI pipettes, gold plated	Mascom, Bremen	Cat#MC-10-10u
POROS-20 AL beads	Applied Biosystems	Cat#1-6028-10
Quantifoil R3/3 holey carbon grids with 2nm carbon	Quantifoil Micro Tools, Germany	N/A
Reprosil Gold 300 C4, 3 μm (10 × 1 mm) column	Dr. Maisch, Germany	Cat#r33.4g.t0101
Reprosil Gold 300 C4, 3 μm (125 × 1 mm) column	Dr. Maisch, Germany	Cat#r33.4g.s1201
Restek HPLC column	Restek, USA	Cat#25119
Superdex 200 PC 3.2/30 column	GE Healthcare Life Sciences	Cat#29-0362-31

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to the Lead Contact M. Hayer-Hartl (mhartl@biochem.mpg.de).

METHOD DETAILS

Proteins

R. sphaeroides wild-type Rubisco, mutant Rubisco with deletion of the four C-terminal residues from RbcL (EΔ4) and the activase (RsRca) were purified as previously described (Mueller-Cajar et al., 2011). The mutants Rca L49C/A263C (RcaCC), RcaCC Y114A, RsRbcL D474N, RsRbcL D477N, and RsRbcL D474N/D477N were generated by QuikChange mutagenesis (Strategene) and recombinantly expressed and purified (Mueller-Cajar et al., 2011), except that DTT (2-5 mM) was present during purification and storage of RcaCC proteins.

Preparation of Oxidized RcaCC

Purified RcaCC was buffer exchanged into 50 mM Tris-HCl pH 8.0/50 mM NaCl and incubated with 4 mM ATP, 10 mM $MgCl_2$ and 0.5 mM RuBP. Oxidation was initiated by addition of 50 μM $CuCl_2$ for 15 min at 25°C and stopped with 50 mM EDTA. The reaction was applied to a size-exclusion chromatography column (Superdex 200 10/30; GE Healthcare) equilibrated with buffer A (20 mM Tris-HCl pH 8.0/50 mM NaCl) and the fractions containing hexameric RcaCC were pooled, concentrated, supplemented with 10% glycerol, flash frozen in liquid N_2 and stored at $-80^{\circ}C$.

Size-Exclusion Chromatography

Size-exclusion chromatography was performed on a Superdex 200 3.2/300 column attached to an AKTA Basic HPLC system with a flow rate of 80 $\mu L\ min^{-1}$ (GE Healthcare). RcaCC (50 μL of $\sim$4.0 μM hexamer) with ATP/RuBP (1 mM each) was injected into the column equilibrated with buffer A or buffer A containing 3 mM DTT for non-reducing and reducing conditions, respectively. The elution of the protein was monitored at 280 nm.

ATPase Assay

ECM, ECM-CABP, EΔ4CM, E-RuBP, E-CABP, E-XuBP and EΔ4-RuBP were prepared as previously described (Bracher et al., 2015; Mueller-Cajar et al., 2011). Briefly, active carbamylated Rubisco (ECM or EΔ4CM; 25-50 μM active sites) was prepared by incubation in buffer A/40 mM $NaHCO_3$/10 mM $MgCl_2$ for 10 min at 25°C. The inhibited non-carbamylated Rubisco (E-RuBP or EΔ4-RuBP; 25-50 μM active sites) was prepared by incubation in buffer A/4 mM EDTA for 10 min followed by addition of 0.8 mM RuBP and incubation for 10 min. The inhibited enzymes ECM-CABP, E-CABP or E-XuBP, were first incubated for10 min in the respective buffers described above, followed by incubation for 10 min in the presence of 0.8 mM carboxypentitol bisphosphate (a 1:1 mixture of the inhibitor CABP and its lower affinity stereoisomer carboxyribitol-1,5-bisphosphate) (Pierce et al., 1980) or 0.5 mM XuBP. ATPase activity was measured spectrophotometrically by monitoring NADH oxidation in a coupled enzymatic assay (Mueller-Cajar, 2011 #15). ATPase assays were performed at 25°C with active or inhibited Rubisco (0.375 μM hexadecamer) in the presence of wild-type Rca or RcaCC (0.83 μM hexamer) in buffer (50 mM Bis-Tris pH 7.5/50 mM NaCl/100 mM KCl/10 mM $MgCl_2$/ 2 mM phosphoenol pyruvate/1 mM ATP/0.5 mM NADH and ATP regenerating enzymes (2.5 U pyruvate kinase and 3.5 U lactate dehydrogenase). RuBP (1 mM) was included when indicated.

Rubisco Activity Assay

Rubisco CO_2 fixation assays were measured spectrophotometrically in a coupled enzymatic assay monitoring NADH oxidation as a function of CO_2 uptake and 3-PGA release (Kubien et al., 2011; Tsai et al., 2015). Reactions with wild-type Rubisco (0.15 μM active sites or 0.02 μM hexadecamer) in the presence or absence of Rca or RcaCC (0.67 μM hexamer) in buffer (0.1 M Tricine-NaOH pH 8.0/ 5 mM $MgCl_2$/coupling enzymes (2.5 U creatine P-kinase, 2.5 U glyceraldehyde-3-P-dehydrogenase, 2.5 U 3-phosphoglycerate kinase, 20 U triose-P-isomerase and 2 U glycerol-3-P-dehydrogenase)/0.5 mM NADH/10 mM phosphocreatine/20 mM $NaHCO_3$/ 1 mM RuBP/1 mM ATP) were performed at 25°C. Activity assays of ECM-CABP or E-CABP also contained $\sim$5 μM carboxypentitol bisphosphate, as a result of diluting the stock solutions of inhibited Rubisco (see above) into the enzyme reaction. Buffers used for measurements under reducing conditions contained 3 mM DTT.

Chemical Crosslinking Coupled to Mass Spectrometry

Prior to chemical crosslinking, proteins were buffer exchanged into 20 mM HEPES pH 7.5/50 mM NaCl using Micro BioSpin 6 columns (BioRad). Reactions containing Rubisco (ECM-CABP; 1.25 μM hexadecamer) and RcaCC (3.33 μM hexamer) either alone or together were incubated with 10 mM ATP/20 mM $MgCl_2$/1 mM RuBP for 1 min at 25°C. Crosslinking was initiated by addition of 1 mM of an isotopic mixture of H_{12}/D_{12} (1:1) disuccinimidyl suberate (Creative Molecules) and the reaction allowed to proceed for 30 min, before quenching with NH_4HCO_3 (150 mM). Crosslinked proteins were resolved on a NuPAGE 3%–8% Tris-acetate native gel (Life Technologies). The high molecular weight (HMW) crosslinked bands were excised and subjected to in-gel trypsin-digestion and desalted as described previously (Shevchenko et al., 2006). Desalted peptides in 5% formic acid were analyzed on nanoACQUITY UPLC system (Waters Corp., USA) connected to LTQ-Orbitrap (Thermo). Peptide separation was performed on home-made nanoLC columns (ID 100 μm, length 20 cm with 15 μm tip opening, New Objective), packed with 3 μm C18 beads (Reprosil-Pur C18-AQ, Dr. Maisch GmbH) using a 120 min gradient from 0.2% formic acid in water to 0.2% formic acid in acetonitrile. Sample loading to the column (0.5 μl/min) was performed with a nanoACQUITY UPLC autosampler (Waters Corp., USA) without a trap column, with the UPLC flow rate during sample analysis maintained at 0.4 μL per min. MS analysis, using Thermo Xcalibur software, was performed in a standard data dependent mode of 1 high resolution 60000 FWHM MS1 scan (m/z 330- 1700) followed by MS^2 scans of 8 most intense ions, with charge states 1 and 2 excluded.

Identification of Crosslinks

To identify the crosslinked peptides, Thermo Xcalibur raw files were converted to Mascot generic file (mgf) format, using Mass Matrix Data File Conversion tools v 3.9 (http://www.massmatrix.net). Analysis of the mgf files was performed on StavroX 3.6.0.1 (http://www.stavrox.com/Download_StavroX_Win.htm) (Götze et al., 2012) with the following filters: FDR cut off $\leq$ 5%, MS^1 tolerance = 3 ppm, MS^2 tolerance = 0.3 Da, missed cleavages for K and R = 3 and 1, respectively, signal to noise ratio $\geq$ 2, fixed and variable modifications = Cys-carbamidomethylation and Met-oxidation, respectively, minimum length of each of the two crosslinked peptides = 5 amino acids, crosslinks with terminal lysines excluded. Lysine and N-termini were considered as potential crosslinking sites for DSS. Manual validation of all identified crosslinked peptides included the assignment of b- and y-ion and presence of

MS^1 doublets separated by 12.0753 Da mass for DSS-H_{12}/D_{12}. Crosslink sites (Table S1B) were only considered when identified in at least two experiments.

Native Mass Spectrometry

Crosslinking was performed as described above except that the DSS (Pierce) did not contain the D_{12} heavy isotope and the crosslinking was quenched after 10 min. Crosslinked proteins (1.25 μM ECM-CABP hexadecamer, 3.33 μM RcaCC hexamer and ECM-CABP/RcaCC complex) were buffer exchanged into 100 mM ammonium acetate (Sigma) pH 8.0, using Micro Bio-Spin 6 columns (BioRad). For native-MS analysis of non-crosslinked RcaCC (Figure 1D), the protein (2.5 μM hexamer) was buffer exchanged as described above. Native-MS analysis was performed on quadrupole IM time-of-flight hybrid mass spectrometer with a Z-spray nano-ESI source (Synapt G2-Si, Waters Corp., USA) in positive ion mode, using 10 μm gold-platted nano-ESI pipettes (Mascom, Bremen) as capillaries. Optimized capillary and sample cone voltages were 1-1.5 kV and 150 V respectively, at a source temperature of 30°C. The instrument was calibrated with acetonitrile: water solution of 30 mg mL^{-1} cesium iodide.

Hydrogen-deuterium Exchange (H/DX) and Mass Spectrometry

Intact Protein H/DX MS Analysis

In remodeling experiments of inhibited Rubisco, RcaCC (20 μM hexamer) was incubated with E-RuBP, E-XuBP, ECM-CABP or EΔ4-RuBP (2.5 μM catalytic sites each) in buffer C (25 mM Tris-HCl pH 7.5/20 mM $MgCl_2$/10 mM ATP) containing 1 mM RuBP for 30 s at 25°C (Figures 2 and S1). Hydrogen-deuterium exchange was initiated by addition of 14-fold excess of 99.9% D_2O buffer (25 mM Tris-HCl pD 7.5) for 60 s, followed by shifting the pH to ~2.5 with ice cold quench buffer (200 mM phosphate buffer pH 2.2/3 M GuHCl/150 mM TCEP). Hydrogen-deuterium exchange was initiated by addition of 14-fold excess of 99.9% D_2O buffer (25 mM MOPS pD 7.8/5 mM KCl) for 60 s, followed by shifting the pH to ~2.5 with ice cold quench buffer. The samples ECM, E-RuBP, E-XuBP, EΔ4CM, EΔ4-RuBP and ECM-CABP in the absence of RcaCC were incubated, labeled, and quenched as described above in either buffer C or D. Quenched samples (50 μl) were injected into a nanoACQUITY UPLC HD/X manager (Waters Corp., USA) housing reverse phase C4 desalting (10 × 1 mm) and analytical (125 × 1 mm) columns (ReproSil Gold 300 C4, 3 μm, Dr. Maisch GmbH, Germany). Samples were desalted for 3 min at a flow rate of 100 μl min^{-1} and separated on an analytical column at a flow rate of 40 μl min^{-1} using a 8 min gradient from 0.1% formic acid/15% acetonitrile to 0.1% formic acid/85% acetonitrile. To minimize back exchange, all chromatographic steps were performed at ~0°C and pH 2.5 (Zhang and Smith, 1993). The proteins were directly eluted into a Synapt G1 or G2-Si mass spectrometer (Waters Corp., USA) equipped with a standard electrospray interface. Mass accuracy of < 20 ppm was confirmed using horse heart myoglobin as a standard. Experiments were repeated 3-5 times, allowing comparison of relative deuterium uptake levels. The averaged masses as relative deuterium uptake levels were not corrected for back exchange (Wales and Engen, 2006). Protein mass spectra were deconvoluted using MassLynx 4.1 (Waters Corp., USA).

Peptide Level HD/X MS Analysis

H/DX reactions were performed essentially as above, except that the concentration of Rubisco catalytic sites was 6 μM and RcaCC 2.5 μM (hexamer). Deuterium labeling was performed for 10 s after incubating inactive Rubisco (E-RuBP) from 10 s to 7 min in the presence of RcaCC and ATP (10 mM), followed by acid quenching at 0°C. Quenched samples (~15 pmoles) were injected into a nanoACQUITY UPLC HD/X manager and digested on an in-line stainless steel column of 2.1 mm inner diameter and 50 mm length (Restek Corporation, USA), self-packed with immobilized pepsin (Wang et al., 2002) on POROS-20AL beads (Applied Biosystems) at a flow rate of 100 μl min^{-1} at 20°C. The peptic peptides were trapped and desalted on a ACQUITY UPLC C18 BEH VanGuard precolumn (2.1 × 5 mm, Waters Corp., USA) containing 1.7 μm C18 beads for 3 min at a flow rate of 100 μl min^{-1} and separated in 6 min using a 8%–40% gradient of acetonitrile in 0.1% formic acid pH 2.5 on a ACQUITY UPLC BEH C18 analytical column (1.0 × 100 mm, 1.7 μm, Waters Corp., USA). Both pre- and analytical columns were maintained in the nanoACQUITY UPLC HD/X manager maintained at ~0°C. The use of protonated solvents during UPLC allowed exchange of deuterium from side chains and amino/carboxyl termini with hydrogen that exchange faster than backbone amides. The eluted peptic peptides were directly analyzed on a Synapt G2-Si mass spectrometer equipped with a standard ESI source (Waters Corp., USA) over a mass range of 50-1700 m/z. Glu-fibrinopeptide (Sigma) was continuously sprayed through the lock mass channel to maintain calibration at < 3 ppm. The mass spectrometer was run in MS^E and ion mobility mode with the following settings: ESI positive ion mode; capillary voltage, 3000 V; cone voltage, 40 V; desolvation temperature, 175°C; source temperature, 80°C; nitrogen desolvation gas flow, 600 l hr^{-1}; scan rate, 0.23 scans s^{-1}. Step wave settings used were as described (Guttman et al., 2016). Peptic fragment identification was performed with $Identity^E$ software implemented in the ProteinLynx Global Server 2.4 (Waters Corp., USA) from 4-5 MS^E analyzed replicates (Plumb et al., 2006). MS^E was performed by a series of low-high collision energies ramping from 5–30 V, ensuring proper fragmentation of the peptic peptides. Deuterium exchange levels were determined by DynamX 3.0 software (Waters Corp., USA) by identifying the isotopic distribution (from +1 to +6 charge state, depending on the peptide). Isotope distribution and peak selection was verified manually for all peptides. The relative deuterium incorporation was calculated by subtracting the centroid of the isotopic distribution for peptide ions of the unlabeled reference from the centroid of the isotopic distribution for peptide ions from each deuterated sample. All the experiments were performed under identical experimental conditions and hence, deuterium levels are uncorrected for back exchange and are reported as relative uptake levels (Wales and Engen, 2006). Experiments were repeated 3-5 times.

Electron Microscopy (EM) and Data Processing

Negative-Stain EM

Crosslinked ECM-CABP-RcaCC complexes were prepared by mixing RcaCC (20 μM monomer) with E.C.M-CABP (10 μM monomer) in a reaction containing 20 mM HEPES pH 7.5, 50 mM NaCl, 10 mM $MgCl_2$, 10 mM ATP and 1mM RuBP, for 1 min at 25°C prior to addition of 0.125% of glutaraldehyde (GA). After 10 min the reaction was quenched by addition of 0.1M Tris-HCl pH 8 followed by gel filtration on a Superdex 200 PC 3.2/30 column (GE Healthcare). The fractions were eluted in buffer A and analyzed on a 6% native gel. Fractions 12 and 13 (Figure S6B) containing HMW complexes with the least amount of free Rubisco were chosen for negative-stain and cryo-EM, respectively. The crosslinked ECM-CABP-RcaCC complexes were diluted to 30-35μg mL^{-1} in 20 mM Tris-HCl pH 8.0, 50 mM NaCl, 1 mM ATP, 1 mM ATPγS and 1 mM RuBP, and negative staining performed using 2% uranyl acetate. Images for 2D class averages were taken on a Tecnai G2 Spirit transmission electron microscope, operating at 120 kV, using a FEI 2,048 × 2,048-pixel CCD camera. The pictures were taken at a nominal magnification of 90,600 x resulting in a pixel size of 3.1 Å at the specimen level with the defocus varying between 350 nm and 1000 nm. The defocus and astigmatism of micrographs were determined using CTFFIND3 (Mindell and Grigorieff, 2003) and CTF correction was done in SPIDER (Frank et al., 1996; Shaikh et al., 2008). A total of 1191 particles were manually selected using the MRC program Ximdisp (Crowther et al., 1996). Boxed particles were analyzed using IMAGIC-5 (van Heel et al., 1996). Single particles were normalized and band-pass filtered between 200 and 10 Å and centered by iterative alignments to their rotationally averaged sum. Resulting class averages (~8 particles/class) served as input for one round of competitive alignment to separate the dataset into classes having either one or two RcaCC bound per Rubisco based on multivariate statistical analysis followed by multi-reference alignment.

Cryo-EM

Using a FEI Vitrobot Mark VI, 3.5 μl GA crosslinked ECM-CABP-RcaCC complexes (~120 μg mL^{-1}) prepared as above were applied on glow discharged Quantifoil R3/3 holey carbon grids with 2 nm carbon support film for 45 s at 4°C and 100% humidity before blotting for 2 s and vitrification. Cryo-EM images were acquired under low-dose conditions at 300 keV on a Titan Krios transmission electron microscope equipped with a Falcon II detector at the Netherlands Center for Electron Nanoscopy (NeCEN) financed by iNext (Instruct). A total of 4824 micrographs were collected at a nominal magnification of 134615 with defocus ranging from -200 to -3600 nm and sampled at 1.04 Å per pixel at the specimen level. 22 frames were recorded over 1.25 s giving an accumulated dose of 50 e-/Å^2. All image processing and 3D reconstruction was performed using RELION-1.3 or -1.4 (Scheres, 2015). Contrast transfer functions were determined by CTFFIND4 (Rohou and Grigorieff, 2015) and movie frame alignment was performed using Unblur (Grant and Grigorieff, 2015). Micrographs recorded with strong astigmatism, at overfocus, very low defocus or with broken ice or strong ice contamination were discarded, yielding 4239 micrographs for further processing steps. 528 manually selected particles were 2D-classified and a 2D class average of ECM-CABP with RcaCC bound was used as template for semi-automated particle selection (Scheres, 2015) resulting in an initial dataset of 862590 particles. The dataset was cleaned and sorted using the z-score function and iterative 2D classifications in RELION-1.3, resulting in a sub-dataset of ECM-CABP of 333122 particles and a sub-dataset of ECM-CABP-RcaCC of 333711 particles. One third of the ECM-CABP dataset was selected for a first 3D refinement using a homology model of *R. sphaeroides* Rubisco based on the crystal structure of *Alcaligenes eutrophus* Rubisco (pdb: 1BXN) filtered to 50 Å resolution as reference. The resulting (initial) 3D reconstruction was filtered to 20 Å resolution and used for 3D refinement of the aligned movie frame averages of the ECM-CABP dataset in RELION-1.3. Auto-refinement without mask resulted in a 3.4 Å and 3.8 Å reconstruction for ECM-CABP with applied D4 symmetry and without symmetry, respectively.

3D refinement of a smaller dataset of the Rubisco-RcaCC complex of ~37500 particles, using the down filtered ECM-CABP D4 reconstruction as input model, generated an initial reconstruction for the complex. The initial reconstruction was filtered to 20 Å resolution and used for 3D refinement of the entire ECM-CABP-RcaCC dataset. Different masks were applied during 3D refinement, yielding ECM-CABP reconstructions with a) the full RcaCC density at 7.56 Å resolution, and b) only the central density of the hexameric RcaCC ring at 6.05 Å resolution. Movie frame weighting was performed in RELION-1.4 using all movie frames. Subsequent 3D refinement of the weighted movie dataset resulted in a 3.4 Å symmetrized ECM-CABP reconstruction and a 3.5 Å reconstruction without applied symmetry. Auto-refinement of the weighted movie ECM-CABP-RcaCC dataset did not improve the reconstruction. The dataset of ECM-CABP-RcaCC was also 3D-classified into 2 classes using the initial 3D reconstruction filtered to 20 Å as a reference and the resulting classes showed no difference in the binding mode of RcaCC. All refinements used gold standard Fourier shell correlation (FSC) calculations and reported resolutions are based on the FSC = 0.143 criterion of mask-corrected FSC curves. All final maps were masked and sharpened using automatically determined negative B-factors.

Modeling

The RbcL and RbcS subunit structures of *R. sphaeroides* Rubisco were initially modeled with SWISS-MODEL (Kiefer et al., 2009) and ModBase (Pieper et al., 2011). The sequences of RsRbcL and RsRbcS were manually built into the B-factor sharpened electron density using the program Coot (Emsley et al., 2010), based on the crystal structure of the highly homologous red-type Rubisco of *Galdieria partita* with bound CABP (pdb: 1BWV). A B-factor of -200 Å^2 was applied to sharpen the electron density map for visualization. For reciprocal-space refinement of the model with REFMAC5 (Brown et al., 2015) in the EM mode, the RELION map was converted first into a pseudo-diffraction dataset mtz file using REFMAC5 (Brown et al., 2015). During refinement, the weight for the model geometry versus the X-ray target was adjusted for plausible stereochemistry. The final model has one outlier (RbcS Val116) and 94.5% in the favored regions in the Ramachandran plot, according to the criteria of MolProbity (Chen et al., 2010). Occupancies 0.2, 0.4, 0.5, 0.6 and 0.8 were tested for residues 462-479 in REFMAC5 refinement, because of high B-factors in this segment at full occupancy. At occupancy 0.4, the C-terminal strand had B-factors similar to residues contacting the C-terminal tail (290-396 $Å^2$) and the corresponding model the best refinement statistics (R-factor 0.3511, average Fourier shell correlation 0.8138). The Rubisco model and the hexameric model of RsRca were docked into the cryo-EM maps as rigid bodies using the Chimera fit (Pettersen et al., 2004) in map function by simulating maps from atomic coordinates at the resolution of the EM maps and optimizing map in map correlation. The models were then refined by molecular dynamics flexible fitting (Trabuco et al., 2008). Visualization of cryo-EM maps and models was performed using Chimera (Pettersen et al., 2004) and PyMOL 0.99.

Analysis of Amino Acid Residue Interaction Network

The analysis of interacting amino acid residues in structural model of RsRubisco was performed with the Residue Interaction Network Generator (RING) v2.0.1 web server (Piovesan et al., 2016). All possible interactions were taken into account (covalent, salt bridges, hydrogen bonds, van der Waals interactions, $\pi-\pi$ stacking, π cation) with distance threshold set to either strict or relaxed (for comparison) and network policy to closest. Per residue pair multiple interaction types were determined. The interaction network was visualized with PyMOL 0.99 and Cytoscape 3.4.0 (Shannon et al., 2003).

DATA AVAILABILITY

Data Resources

The EM density maps for the complexes RsRubisco D4; RsRubisco C1; ECM-CABP-RcaCC and ECM-CABP-RcaCC masked have been deposited to the Electron Microscopy Data (EMD) Bank under accession codes EMD: 3699, 3700, 3701, and 3702, respectively. The accession number for the Rubisco structural model reported in this paper is PDB: 5NV3. The full gels of Figures 1C, 4A, S5, and S6 can be viewed at Mendeley Data (URL: http://dx.doi.org/10.17632/vnt5ck2kbz.1).

Molecular Cell, Volume *67*

Supplemental Information

Mechanism of Enzyme Repair by the AAA^{+} Chaperone Rubisco Activase

Javaid Y. Bhat, Goran Miličić, Gabriel Thieulin-Pardo, Andreas Bracher, Andrew Maxwell, Susanne Ciniawsky, Oliver Mueller-Cajar, John R. Engen, F. Ulrich Hartl, Petra Wendler, and Manajit Hayer-Hartl

SUPPLEMENTAL INFORMATION

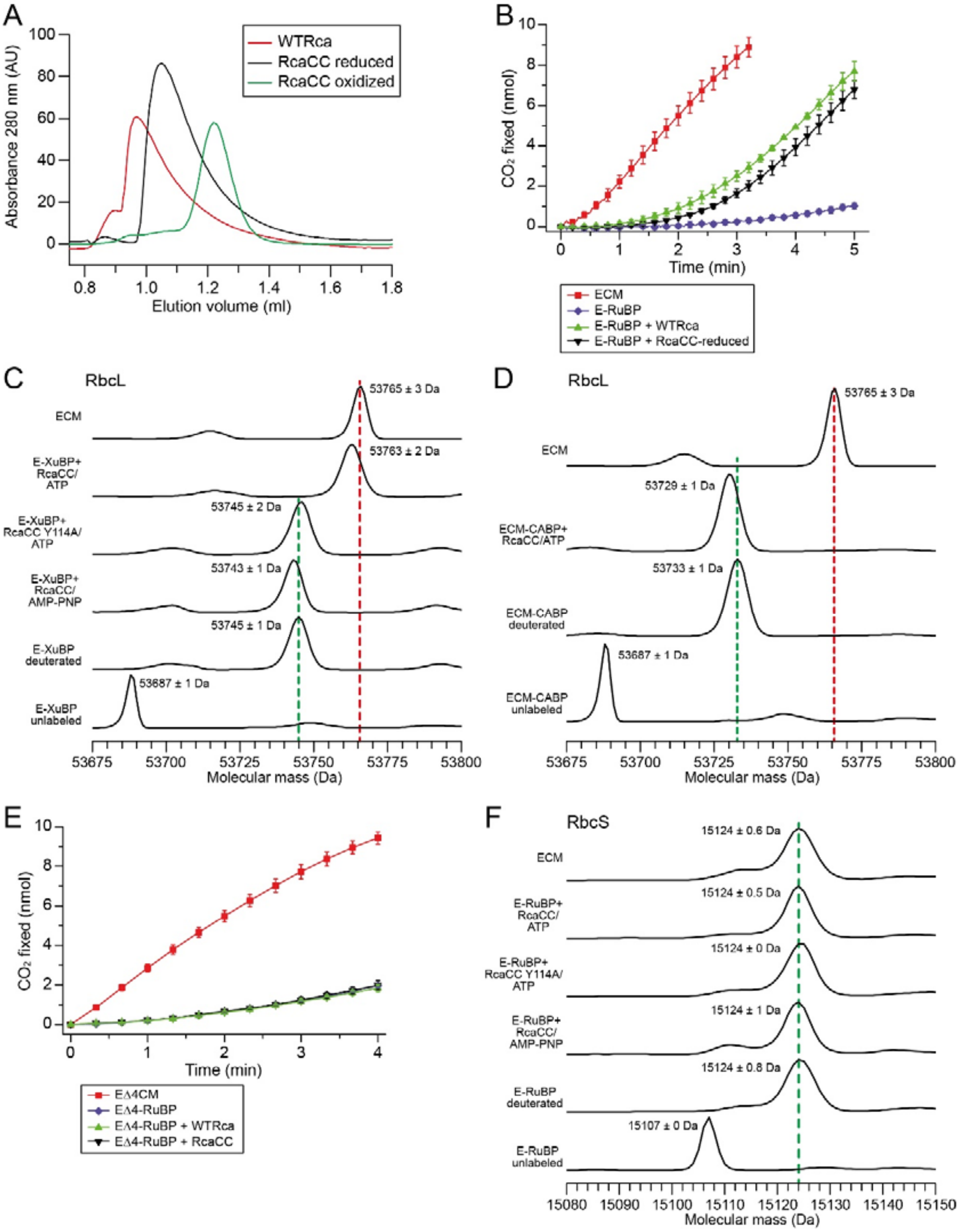


Figure S1. RcaCC-mediated Remodeling of Inhibited Rubisco, Related to Figures 1 and 2
(**A**) Size exclusion chromatography of WTRca, oxidized RcaCC and reduced RcaCC on a Superdex 200 3.2/300 column in presence of ATP/RuBP. See Star Methods for details. (**B**) Reactivation of inhibited Rubisco by reduced RcaCC. CO2 fixation assays were performed with active (ECM) or inhibited (E-RuBP) Rubisco as in Figure 1F in the absence or presence of WTRca or RcaCC under reducing conditions (3 mM DTT). SDs are from at least three independent experiments. (**C**-**D**) Remodeling of inhibited Rubisco monitored by global H/DX-MS analysis. Mass spectra showing deuterium uptake in RbcL from Rubisco inhibited with XuBP (A) or CABP (B). Measurements were performed as in Figure 2. (**E**) Inability of WTRca and RcaCC to reactivate Rubisco containing C-terminally truncated RbcL (EΔ4-RuBP). Activation assays were performed as in Figure 1F. SDs are from at least three independent experiments. (**F**) Remodeling of inhibited Rubisco monitored by global H/DX-MS analysis. Mass spectra showing deuterium uptake in RbcS from Rubisco inhibited with RuBP in comparison to the carbamylated active form. Measurements were performed as in Figure 2. See Star Methods for details.

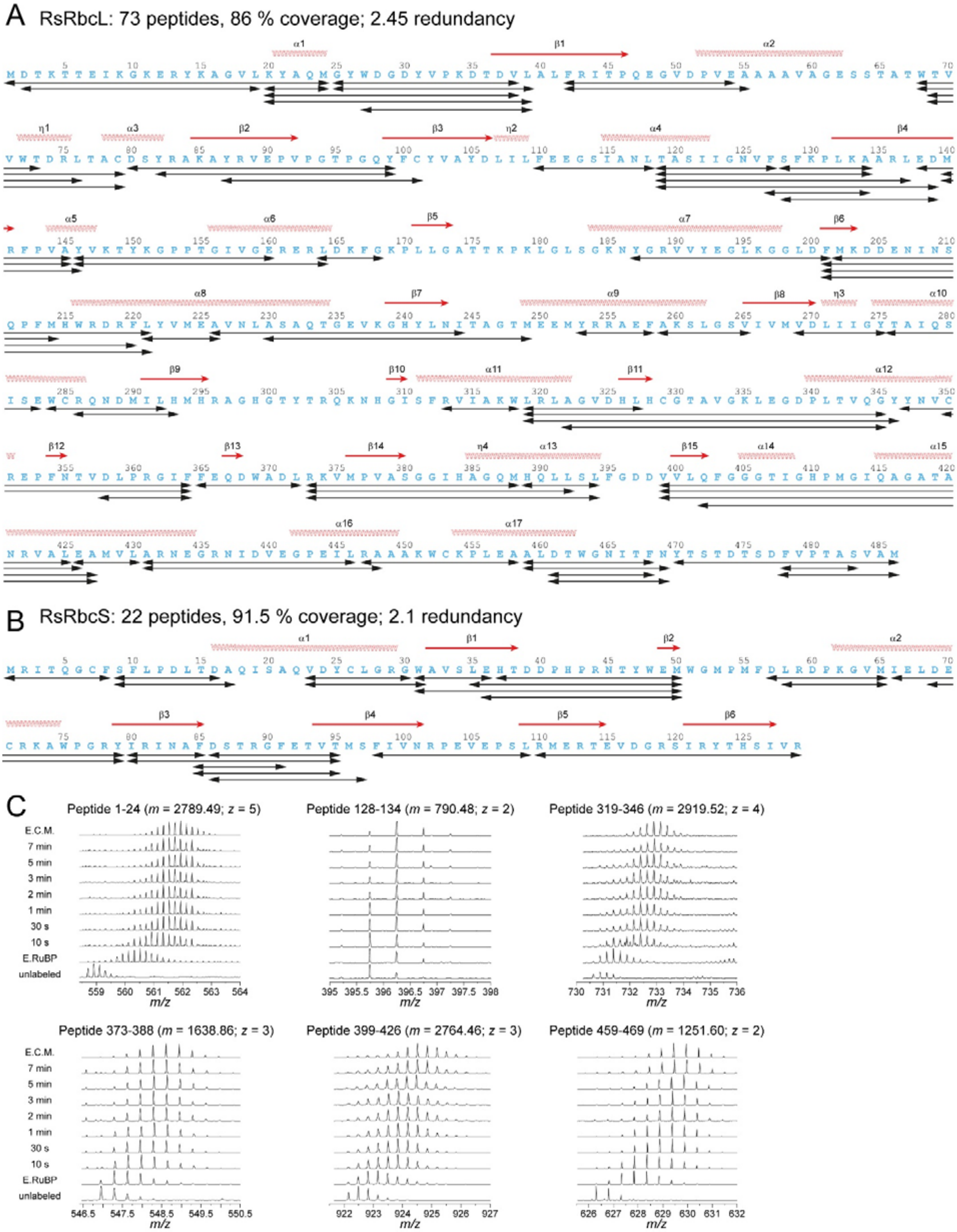

A RsRbcL: 73 peptides, 86 % coverage; 2.45 redundancy
B RsRbcS: 22 peptides, 91.5 % coverage; 2.1 redundancy
C
Peptide 1-24 (m = 2789.49; z = 5)
Peptide 128-134 (m = 790.48; z = 2)
Peptide 319-346 (m = 2919.52; z = 4)
Peptide 373-388 (m = 1638.86; z = 3)
Peptide 399-426 (m = 2764.46; z = 3)
Peptide 459-469 (m = 1251.60; z = 2)
E.C.M.
7 min
5 min
3 min
2 min
1 min
30 s
10 s
E.RuBP
unlabeled
m/z

Figure S2. H/DX-MS Analysis of Rubisco Reactivation at Peptide Resolution, Related to Figure 3

(**A-B**) Sequence coverage in H/DX measurements. Peptic peptides (black double-headed arrows) of RbcL (A) and RbcS (B) analyzed in H/DX-MS measurements of Figure 3 are mapped on the amino acid sequences. α-Helices and β-strands are shown in red on top of the sequences as helices and arrows, respectively. (**C**) Mass spectra showing deuterium incorporation into representative peptides of RbcL during the time course of reactivation of E-RuBP by RcaCC (see Figures 3B and S4). Mass and charge state are shown for each peptide. Deuterium uptake by ECM and spectra of unlabeled E-RuBP are included as reference.

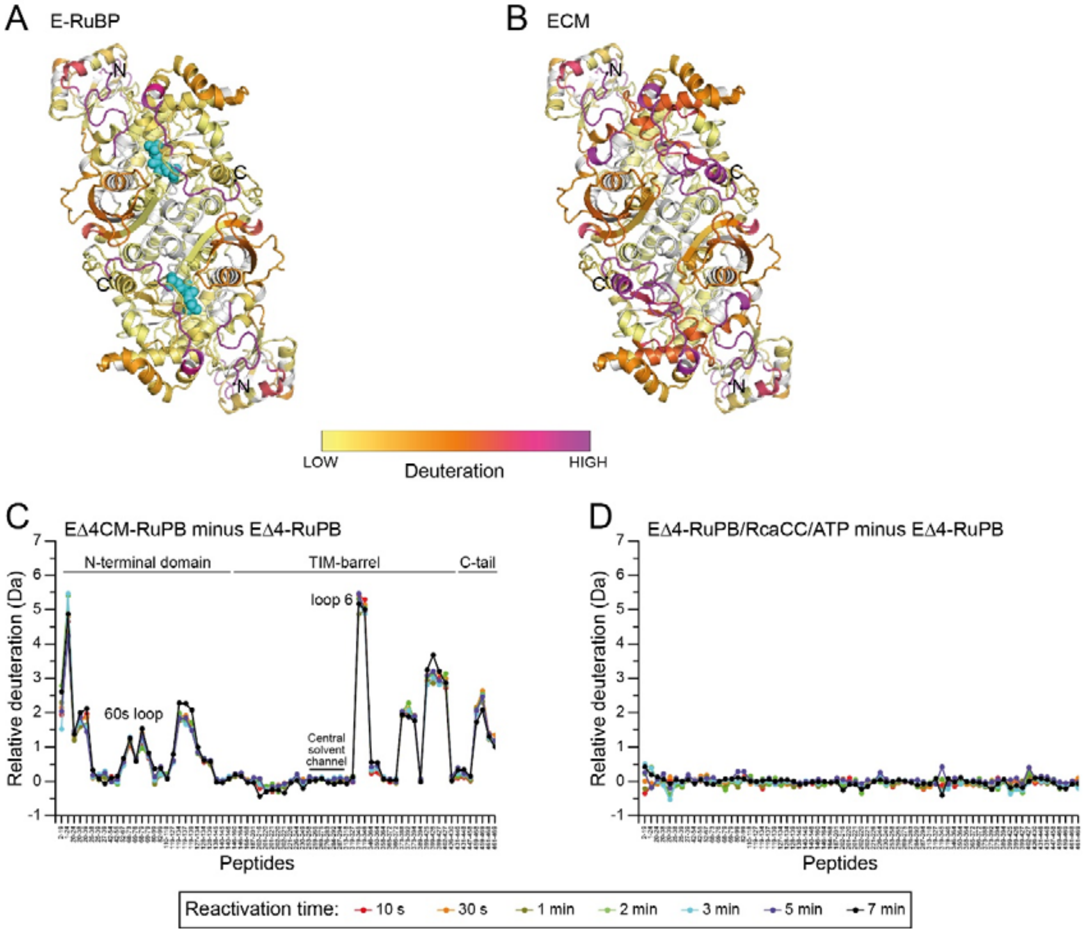


Figure S3. Deuterium Uptake into RbcL Peptides, Related to Figure 3
(**A-B**) The relative deuterium incorporation into RbcL peptides of inhibited Rubisco (E-RuBP) (A) and active Rubisco (ECM) (B) is mapped onto a homology model of the RbcL antiparallel dimer and shown as a color gradient from pale yellow (low relative deuterium uptake) to magenta (high relative deuterium uptake). Regions for which no peptide coverage was obtained are in grey. RuBP is shown in cyan using space-filling representation. Positions of the N- and C-termini are indicated. See Figure 3 and Star Methods for details. (**C-D**) Difference plots showing deuterium incorporation into RbcL peptides upon incubation of EΔ4CM in the absence of RcaCC (C) and upon incubation of EΔ4-RuBP with RcaCC (D) over a time course of 10 s to 7 min of reactivation. Deuterium incorporation measured with EΔ4-RuBP alone is subtracted. Changes in deuteration >0.5 Da are meaningful. H/DX-MS data represents averages of 4-6 independent experiments per time point.

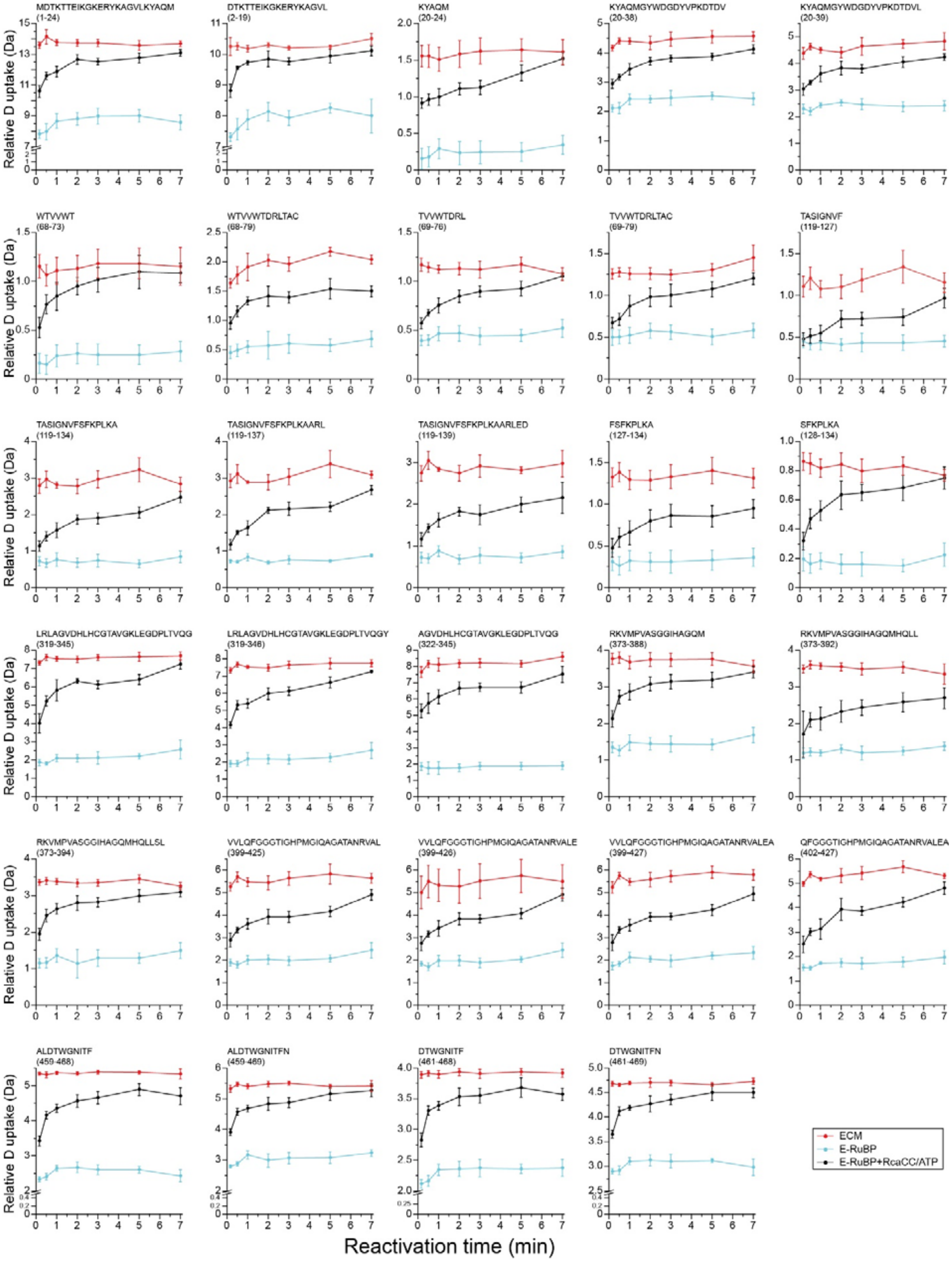

MDTKTTEIKGKERYKAGVLKYAQM (1-24)
DTKTTEIKGKERYKAGVL (2-19)
KYAQM (20-24)
KYAQMGYWDGDYVPKDTDV (20-38)
KYAQMGYWDGDYVPKDTDVL (20-39)
WTVVWT (68-73)
WTVVWTDRLTAC (68-79)
TVVWTDRL (69-76)
TVVWTDRLTAC (69-79)
TASIGNVF (119-127)
TASIGNVFSFKPLKA (119-134)
TASIGNVFSFKPLKAARL (119-137)
TASIGNVFSFKPLKAARLED (119-139)
FSFKPLKA (127-134)
SFKPLKA (128-134)
LRLAGVDHLHCGTAVGKLEGDPLTVQG (319-345)
LRLAGVDHLHCGTAVGKLEGDPLTVQGY (319-346)
AGVDHLHCGTAVGKLEGDPLTVQG (322-345)
RKVMPVASGGIHAGQM (373-388)
RKVMPVASGGIHAGQMHQLL (373-392)
RKVMPVASGGIHAGQMHQLLSL (373-394)
VVLQFGGGTIGHPMGIQAGATANRVAL (399-425)
VVLQFGGGTIGHPMGIQAGATANRVALE (399-426)
VVLQFGGGTIGHPMGIQAGATANRVALEA (399-427)
QFGGGTIGHPMGIQAGATANRVALEA (402-427)
ALDTWGNITF (459-468)
ALDTWGNITFN (459-469)
DTWGNITF (461-468)
DTWGNITFN (461-469)
Relative D uptake (Da)
Reactivation time (min)
0 1 2 3 4 5 6 7
ECM
E-RuBP
E-RuBP+RcaCC/ATP

Figure S4. Relative Deuterium Uptake into RbcL Peptides during Rubisco Reactivation, Related to Figure 3

Time-dependent deuterium incorporation into representative RbcL peptic peptides spanning the RbcL sequence during incubation of E-RuBP with RcaCC (black) and of ECM (red) and E-RuBP (cyan) in the absence of RcaCC. Shown are the raw data (4-6 replicates per time-point) after DynamX analysis of each peptide in all charge states followed by manual inspection and validation. D, deuterium. See Star Methods for details.

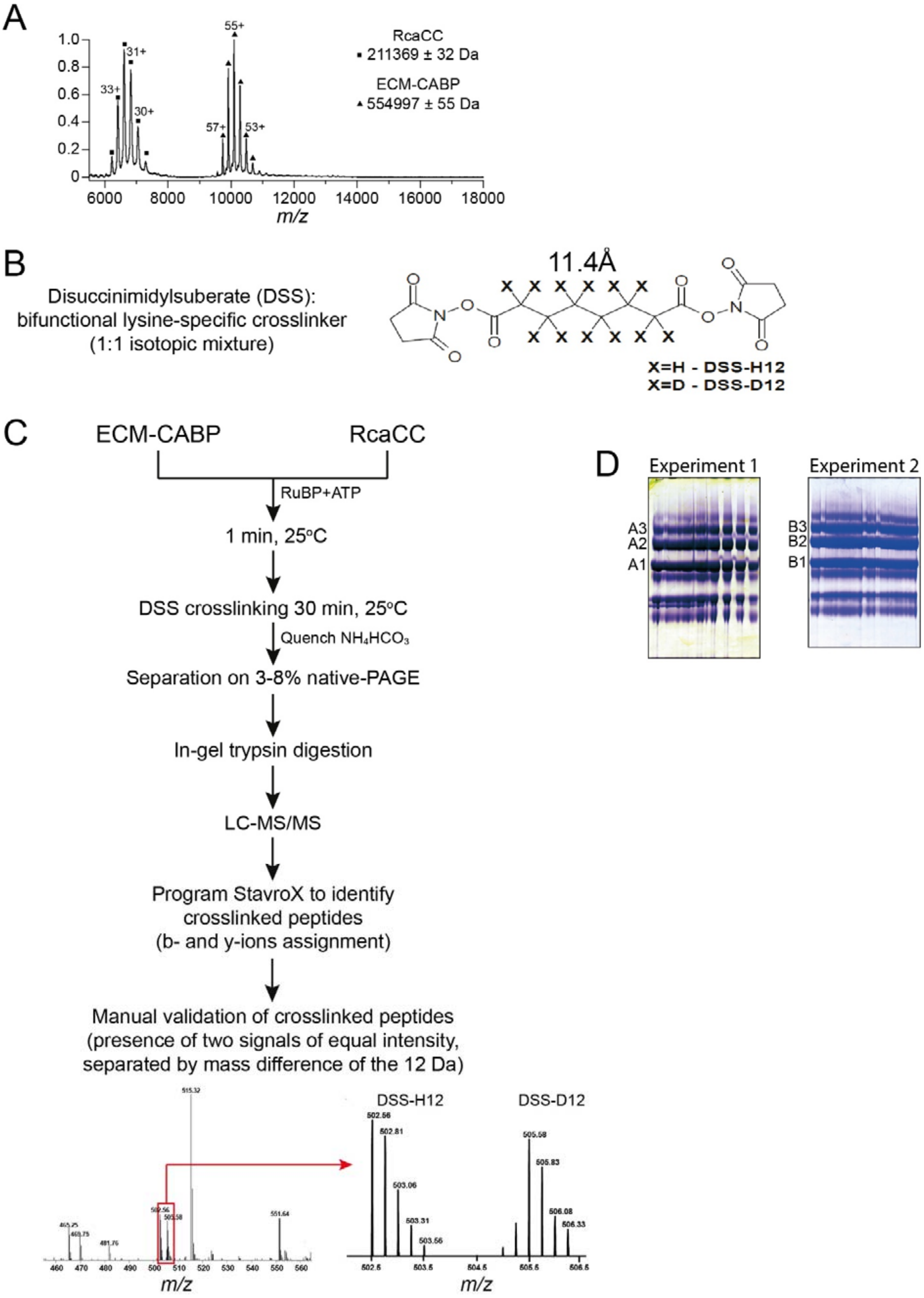

A
1.0
0.8
0.6
0.4
0.2
0
6000
8000
10000
12000
14000
16000
18000
m/z
33+
31+
30+
55+
57+
53+
RcaCC
▪ 211369 ± 32 Da
ECM-CABP
▴ 554997 ± 55 Da
B
Disuccinimidylsuberate (DSS):
bifunctional lysine-specific crosslinker
(1:1 isotopic mixture)
11.4Å
X=H - DSS-H12
X=D - DSS-D12
C
ECM-CABP
RcaCC
RuBP+ATP
1 min, 25°C
DSS crosslinking 30 min, 25°C
Quench NH4HCO3
Separation on 3-8% native-PAGE
In-gel trypsin digestion
LC-MS/MS
Program StavroX to identify crosslinked peptides (b- and y-ions assignment)
Manual validation of crosslinked peptides (presence of two signals of equal intensity, separated by mass difference of the 12 Da)
515.32
465.25
469.75
481.76
502.56
505.58
551.64
460
470
480
490
500
510
520
530
540
550
560
m/z
DSS-H12
DSS-D12
502.56
502.81
503.06
503.31
503.56
505.58
505.83
506.08
506.33
502.5
503.5
504.5
505.5
506.5
m/z
D
Experiment 1
Experiment 2
A3
A2
A1
B3
B2
B1

Figure S5. Chemical Crosslinking Coupled to MS (CXMS) Analysis of E.C.M.-CABP/RcaCC Complexes, Related to Figure 4

(**A**) Nano-ESI native-MS spectra of DSS crosslinking reaction (10 min at 25°C) of ECM-CABP and RcaCC in the presence of RuBP but absence of ATP. Symbols indicate charge-state distributions. The calculated mass around the m/z values of the respective protein complexes and the accuracy of mass values calculated from the different m/z peaks are indicated. (**B**) Structure of the bifunctional crosslinker disuccinimidyl suberate (DSS). The crosslinker is a 1:1 mixture of protonated (light; H_{12}) and deuterated (heavy; D_{12}) isotopes with a mass difference of 12.0753 Da. (**C**) Workflow for crosslinking and analysis of crosslinked peptides by in-gel trypsin digestion and LC-MS/MS. (**D**) Crosslink products of the ECM-CABP and RcaCC reaction as in (C). The proteins (10 µM ECM-CABP and 20 µM RcaCC) were incubated with 10 mM ATP/20 mM $MgCl_2$/1 mM RuBP for 1 min at 25°C, followed by crosslinking by addition of H_{12}:D_{12}–DSS (1 mM) for 30 min at 25°C. The crosslinking reaction was quenched with NH_4HCO_3 (150 mM) and analyzed by native-PAGE. High molecular weight crosslinked bands A1-A3 (Experiment 1) and B1-B3 (Experiment 2) were subjected to in-gel trypsin digestion and LC-MS/MS analysis (Table S1).

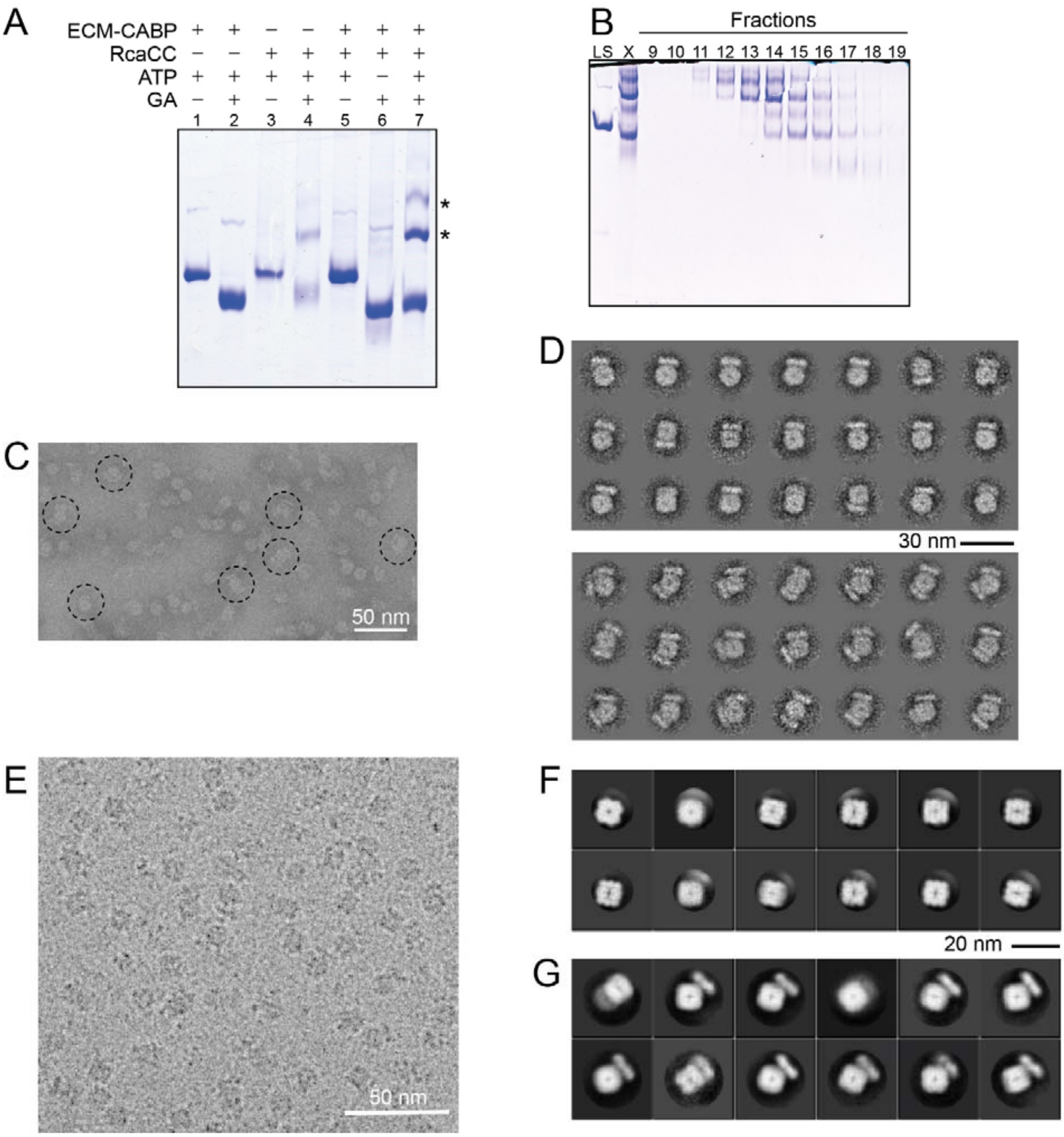


Figure S6. Analysis of Glutaraldehyde Crosslinked E.C.M.-CABP/RcaCC complexes by Negative-stain and Cryo-Electron Microscopy, Related to Figures 6 and 7

(**A**) Rubisco with the tightly bound inhibitor CABP (ECM-CABP) was preincubated for 1 min with RcaCC and ATP as indicated. RuBP was present in all reactions. Reactions were then incubated for 10 min with or without glutaraldehyde (GA), followed by analysis by native-PAGE. High molecular weight complexes representing crosslinked species are marked with asterisks. See Star Methods for details. (**B**) Fractions of the ECM-CABP/RcaCC GA crosslinking reaction obtained by size-exclusion chromatography (Superdex 200 PC 3.2/30; GE Healthcare) were analyzed by native-PAGE. Fraction 13 containing mainly the high molecular weight complexes were used for EM analysis. See Star Methods for details. LS, RbcL$_8$S$_8$; X, GA crosslinking reaction prior to size-exclusion chromatography. (**C**) Negative stain electron micrograph of ECM-CABP/RcaCC complexes (30-35μg ml^{-1}) in the presence of 1 mM ATP and ATPγS. Exemplary ECM-CABP/RcaCC complexes with one, two or three RcaCC rings bound are encircled. (**D**) Class averages of ECM-CABP/RcaCC complexes derived from reference free multivariate statistical analysis (MSA) of negatively stained images in IMAGIC. Each class average contains five to ten images. ECM-CABP bound with either one (upper panel) or two (lower panel) RcaCC complexes are shown. (**E**) Micrographs of ECM-CABP/RcaCC complexes (~120 μg ml^{-1}) in the presence of 1 mM ATP and ATPγS taken under low dose cryo conditions. (**F**) Representative reference free 2D class averages of ECM-CABP derived from cryo-EM images in RELION-1.3. (**G**) Representative reference free 2D class averages of ECM-CABP/RcaCC complexes in the presence of 1 mM ATP and ATPγS, derived from cryo-EM images in RELION-1.3.

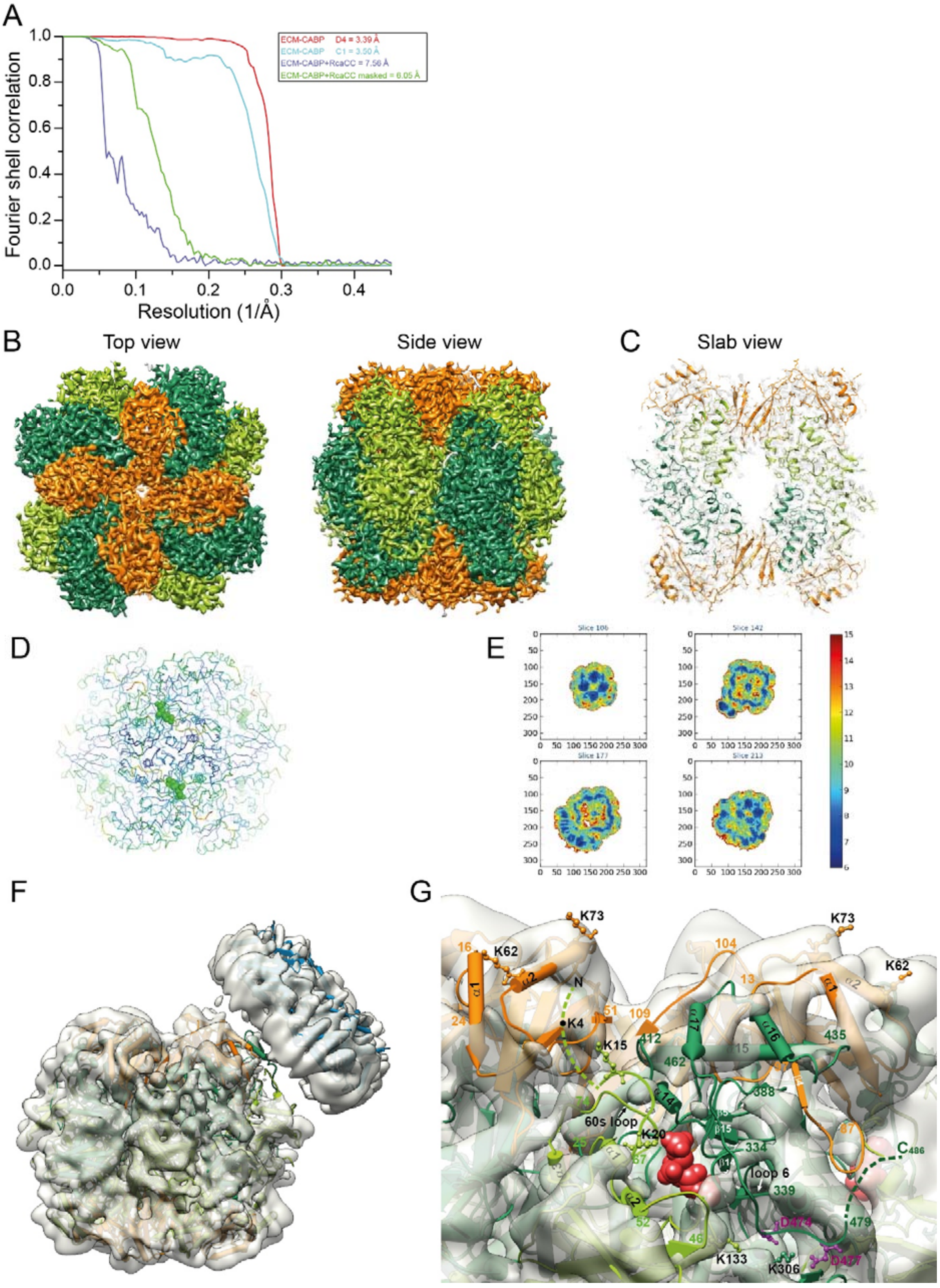

A
Fourier shell correlation
Resolution (1/Å)
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.1
0.2
0.3
0.4
ECM-CABP D4 = 3.39 Å
ECM-CABP C1 = 3.50 Å
ECM-CABP+RcaCC = 7.56 Å
ECM-CABP+RcaCC masked = 6.05 Å
B
Top view
Side view
C
Slab view
D
E
Slice 106
Slice 142
Slice 177
Slice 213
0
50
100
150
200
250
300
15
14
13
12
11
10
9
8
7
6
F
G
K73
16
K62
α1
α2
N
24
K4
51
109
104
13
K73
K62
α1
α2
K15
412
α17
α16
435
α15
462
97
388
74
60s loop
α14
β5
β15
K20
25
α1
67
334
β1
loop 6
87
C486
339
α2
52
46
D474
479
K133
K306
D477

Figure S7. Cryo-EM Structural Analysis of Rubisco and Rubisco-RcaCC complexes, Related to Figures 6 and 7

(**A**) Mask- corrected Fourier shell correlation (FSC) curves of ECM-CABP and ECM-CABP/RcaCC complexes as determined by gold standard FSC procedure in RELION-1.4. All maps were masked and sharpened using automatically determined negative B-factors. (**B**) 3D reconstruction of ECM-CABP complexes at 3.5 Å resolution shown as top and side views. No symmetry was applied during 3D reconstruction. RbcL and RbcS subunits are colored in green and orange, respectively. (**C**) 15 Å slice through the ECM-CABP complex with no symmetry applied and with the sequences of R. sphaeroides RbcL and RbcS fitted into the electron density. The EM reconstruction is shown as transparent surface and the model is colored as in (B). (**D**) B-factors of the refined *R. sphaeroides* Rubisco D4-symmetric model. A backbone trace is shown; CABP is depicted in space-filling mode. B-factors from 214 to 496 Å^2 are shown as a rainbow gradient from blue to red. The average B-factor is 277 Å^2. Residues 13-17 and 465-474 of RbcL are the most mobile parts (B-factor > 370 Å^2). (**E**) Representative slices through the Rubisco-RcaCC complex colored according to ResMap results. The color bar indicates local resolution between 6 and 15 Å. (**F**) 3D reconstruction of the ECM-CABP/RcaCC complex at ~6.0 Å resolution shown as side view with RsRubisco and RsRca (pdb: 3ZUH) fitted as rigid bodies. (**G**) Close up view of the RbcL and RbcS subunits proximal to RcaCC. For clarity, the density map of RcaCC is removed. Flexible residues at the N- and C-termini of the RbcL subunits not resolved in the 3.4 Å structure of Rubisco (Figure 5) are indicated by dashed lines. Structural elements outside the density map are labeled. Residues that were identified to crosslink to RcaCC are shown either in ball and stick representation or as black circles in the flexible N-terminus of RbcL. CABP is shown as space-filling model in red.

4 Discussion and Outlook

4.1 Article 1 - Active Cage Mechanism of Chaperonin-Assisted Protein Folding Demonstrated at Single-Molecule Level

It has been suggested that GroEL/ES accelerates protein folding by preventing reversible aggregation of its substrates[84]. In such an interpretation the GroEL/ES system would perform a passive role, functioning merely as an aggregation prevention device with no active contribution to the folding process. Our experiments on a single molecule level at DM-MBP concentrations of 100 pM, where the probability of two DM-MBP molecules to interact is <1%, allowed us to effectively exclude any influence of reversible aggregation on measured folding rates. We confirmed the absence of aggregates by FCS and dcFCCS measurements. Under such conditions and using smFRET, we compared the spontaneous folding rate of DM-MBP with the rate of GroEL/ES assisted folding and observed a 4-8 fold acceleration of the DM-MBP folding rate in presence of GroEL/ES. This finding demonstrated the active role of GroEL/ES system in the folding process. We went further and asked which part of the system is the active principle: the ATP dependent cycling of the chaperonin (repeated binding, stretching and release) as stated by the iterative annealing model[88,145] or the properties of the GroEL cavity that influence the folding of encapsulated substrate[91,100]? By using the single-ring version of GroEL (SR1-EL) which does not cycle and undergoes only one round of encapsulation[89] we measured a similar increase in DM-MBP folding rate and full yield of folding. It is possible, however, that one round of binding and stretching, followed by release into the cavity partitions the substrate on a productive folding pathway. Our SR1-EL gel filtration experiments also excluded any kind of substrate escape[96,146] from the cavity. PET-FCS measurements on a cycling GroEL/ES system showed further that the non-native DM-MBP spends ~80% of the cycle in the GroEL cavity, demonstrating that it is the GroEL cavity itself that is important for folding of the DM-MBP. This is in contrast to the statement posited by the iterative annealing model[85,88], which does not ascribe any specific role to the inner environment of the GroEL cavity and presumes that the substrate can be released either into the cavity or into free solution for folding. Further into detail on how the GroEL cavity influences the encapsulated substrate, we employed PET-FCS and measured the chain dynamics of the DM-MBP intermediate during spontaneous and GroEL/ES assisted folding. Comparison of the DM-MBP intermediate chain dynamics, which is reflected in the relaxation time (τ_R) of free (τ_R = 40±3 μs) and encapsulated (τ_R = 99±1 μs) protein, revealed a restriction in the flexibility of the DM-MBP chain upon encapsulation. A similar τ_R was measured for DM-MBP when stably encapsulated in SR1-EL. PET-FCS experiments allowed for the first time a direct observation of chain entropy for substrate within the GroEL cavity. Experimental results corroborated earlier predictions that the GroEL cavity restricts conformational entropy of the encapsulated chain by steric confinement which accelerates folding[90,91,147]. However, confinement is not the sole factor contributing to acceleration of substrate folding. Another property of the GroEL cavity that has a crucial role is the net negative charge (-42) of the inner wall. Negative charges are clustered into two circular layers at the level of the apical domains. Experiments with the GroEL KKK2 mutant, where the second layer of negatively charged amino acid residues was mutated to positive charge (D359K, D361K, E363K), changing a net negative charge (-42) of the cavity to net neutral charge (0), showed no acceleration of DM-MBP folding, despite steric confinement. Similarly, stable encapsulation by SR1-EL KKK2 showed no increase in DM-MBP folding rates. Additionally, our PET-FCS measurements showed that DM-MBP intermediate chain dynamics within GroEL KKK2 is not restricted anymore and is similar in behaviour to a highly dynamic intermediate found free in solution. The GroEL cavity clearly plays a dual role. It allows proteins to fold in isolation and thereby prevents any unwanted interactions with other molecules that might result in formation of aggregates. On the other hand, the cavity also serves as a structural framework for correct positioning of the charged residues on the cavity wall, that upon substrate encapsulation point into the cavity and influence protein folding. It is interesting to note that GroEL mutants with a net neutral charge, albeit with different mutations, do not show the same defect in assisted protein folding, pointing to a specific role of the second charge cluster[91,100]. How clusters of charged residues influence folding of an encapsulated substrate remains to be determined.

One way in which the GroEL cavity may influence protein folding is that the charge clusters would accumulate water molecules in their vicinity, thus inducing a water structuring effect[100]. A layer of structured water molecules between exposed hydrophobic residues on the encapsulated substrate intermediate and charged residues on the GroEL cavity wall would experience a high degree of entropy restriction, thus resulting in an enhanced hydrophobic effect. This would result in more efficient burial of hydrophobic residues and accelerated folding of the encapsulated substrate. To date only one experimental study[101] has been conducted to measure the structuring of water inside the GroEL cavity. The authors were not able to find any water structuring effect upon GroEL/ES complex formation. This could be due to sensitivity of the method, unfortunate choice of the residue (top of the inner side of GroES) around which water structuring effect was measured, or the absence of encapsulated substrate. A more promising approach of how to measure water structure (or water mobility), would be to specifically introduce cysteine residues for labelling on the inner wall at positions along the vertical axis of GroEL. The positions for introducing cysteines should not differ substantially between apo and complex states of GroEL, thus excluding the influence of a conformational change on the measurement. The introduced cysteines could then be labelled with a fluorescent probe and by measuring the fluorescence lifetime at different wavelengths one could obtain information on solvation dynamics on a nanosecond timescale[148]. Another option would be to introduce tryptophans instead of cysteines and perform femtosecond fluorescence spectroscopy, which gives information on solvation dynamics at picosecond timescale[149]. It would be necessary to perform these experiments in the absence and presence of the encapsulated protein. The biggest challenge would be to measure how water structure within GroEL changes as the substrates folds. One could achieve this by combining either of the two above mentioned fluorescent approaches with stopped-flow measurements.

A second potential role for the GroEL charge cluster might be direct interaction with encapsulated substrate protein. In this case, the net charge of the substrate protein would also be important since the GroEL cavity might then attract or repel substrate[99]. This could be the reason for observed differences in folding rate enhancements for different substrates[91]. To identify residues that interact with the substrate during folding, one could use an approach whereby the transient interactions would be crosslinked and subsequently analysed by hydrogen/deuterium exchange coupled to mass spectrometry or NMR. To simplify matters one would have to use a slowly folding substrate such as DM-MBP or Rubisco. In the case of fast folding substrates, for example DapA or MetF, one could then use quenched flow approach to follow interactions on shorter time scales.

It is also of potential significance that the two negative charge clusters are located at the level of the GroEL apical domains, which undergo the largest conformational changes during the GroEL cycle. Apical domain movements are important for the remodelling of the bound substrate and subsequent controlled release into the cavity. In the case of GroEL KKK2, the conformational cycle may be perturbed and impair stretching of the substrate. Thus, it may upon release into the cavity form a kinetically trapped state. One could reanalyse the GroEL KKK2 conformational states during the ATP cycle by cryo-EM single particle analysis and compare them with GroEL. It would be of utmost interest to compare the conformational states of GroEL and GroEL KKK2 in the presence of a substrate protein. Comparison of the structures during the conformational cycle with substrate present would then give a definitive answer as to whether the inability of GroEL KKK2 to fold substrates lies in structural perturbations of the conformational cycle.

Another GroEL feature that despite considerable efforts still remains unexplained are the unstructured C-terminal tails[50]. The C-terminal tails have four glycine-glycine-methionine repeats ending with an additional methionine. Interestingly, removal of the C-terminal tails did not result in any growth defects in *E. coli*[150]. It is possible that this modification would prove deleterious under a specific set of stress conditions. So far, the role for the C-terminal tails has been proposed to be in: more efficient substrate encapsulation[151], control of the GroEL ATPase activity[152] and intercalation into the encapsulated substrate during folding[91,153,154]. One could envision a combined effect of these functions, such that

C-terminal tails would help bring the substrate into the cavity, sense that is indeed inside and relay this information to the ATPase active site to increase ATPase activity and speed up the cycle. Additionally, C-terminal tails may help fold the substrate and lastly, may play a role in substrate protein release. The latter might be controlled entropically: once the encapsulated protein reaches the folded state, the entropy of the complete system, GroEL/ES and the substrate, would be too low and not compensated anymore by favourable interactions of the hydrophobic C-terminal tail with now buried hydrophobic residues of the substrate. To increase entropy, folded substrate (low in entropy) would have to be expelled from the system and the new system, GroEL alone, would relax to a high entropy state due to the conformational freedom of the C-terminal tails[155]. A potential approach to observe GroEL C-terminal tails dynamics, and thereby infer their entropy, is by using PET-FCS. One could then measure the C-terminal chain dynamics during the GroEL cycle in the presence and absence of the substrate.

All of the GroEL cavity features have to work in unison to realize the full potential of the GroEL/ES folding machinery. The GroEL cavity is necessary but not sufficient, charged clusters seem to have the biggest effect on folding, since their removal turns GroEL into a passive cage, and the C-terminal tails are implicated in a myriad of functions that remain to be understood. Taken together, the general mechanism by which GroEL folds proteins is still far from being fully explained.

4.2 Article 2 – Chaperonin-assisted protein folding: Relative population of asymmetric and symmetric GroEL:GroES complexes

Recent experiments have suggested that the functional form in the GroEL/ES cycle is a symmetrically shaped complex with two GroES molecules bound on either side of GroEL. In these studies, however, GroEL was modified by introducing cysteines at the level of the apical domains[86,141,142,156]. The apical domain cysteines were used either for coupling to biotin in order to allow tethering to a streptavidin coated surface, or were labelled with a fluorophore to monitor the FRET signal between GroEL and GroES. Such GroEL modifications might have resulted in decreased allosteric coupling between the rings, or changes in binding affinity between GroEL and GroES. This might have caused the authors to overestimate the amount of symmetrical complexes. Our experimental approach used dual colour FCCS with differently labelled populations of GroES, and thereby circumvented any potential perturbation of the allosteric communication in the GroEL rings, or changes in binding affinities. This allowed for a more accurate determination of the relative ratio of asymmetric versus symmetric complexes. In contrast to previous studies, our data showed that symmetrical complexes do not form in the absence of a substrate or in presence of a foldable substrate. We have observed GroES binding to both sides of GroEL only in presence of non-foldable substrates, such as α-lactalbumin (α-LA) and α-casein. These substrates also stimulated the GroEL ATPase activity in the presence of GroES up to 2-fold. This suggested an explanation whereby non-foldable substrates allosterically uncouple the two GroEL rings and allow ATP to bind to both rings. This in turn preconditions GroEL to bind two GroES molecules. The mechanism by which non-foldable substrates uncouple the two rings still has to be elucidated. It might be that, since non-foldable substrates are unable to effectively bury their hydrophobic residues, they 'over-interact' with the C-terminal tails emanating from the GroEL equatorial domains. This may lead to an increased interaction surface area between the GroEL rings, effectively disabling allosteric pathways. In that sense, it would be interesting to see, by using our dcFCCS approach, whether the GroEL mutant without C-terminal tails (GroEL ΔC) can form symmetrical complexes in presence of non-foldable substrates. Moreover, a comparison of single particle cryoEM 3D reconstructions of the GroEL and GroEL ΔC in presence and absence of non-foldable substrates might provide a hint on any increase in distance or change in conformation between the two rings. It might also be that there is a change from a staggered to stacked conformation of GroEL rings, which predisposes the GroEL:GroES$_2$ formation, again by a decrease in allosteric communication between the rings[157]. Along these lines of interplay between GroEL and the substrate, it is tempting to presume that a symmetrical complex might be more efficient in folding since it can fold the double amount of substrate in the same time. However, a recent native mass spectrometry study[158] showed that not all GroEL substrates can bind simultaneously to the apo state of GroEL. It is exactly for such substrates that it would be of benefit to have a free ring in an asymmetrical GroEL:GroES complex which allows for substrate binding and its solubilization. In the case of the symmetrical complex, the substrate binding site is occupied by GroES which would lead to aggregation of the substrate. Thus, the damage to the cell would be much greater than the benefit of simultaneous folding for some substrates. Of course, one has to take into consideration the actual number of substrates that can or cannot bind simultaneously to GroEL. Therefore, it would be of interest to find out the stoichiometry of binding for all GroEL substrates. This would provide a good indication of which form is favoured. Moreover, if GroEL really could function as a symmetrical machine, it is not clear what the trigger would be for the release of GroES and the substrate[50]. An additional interesting observation from our study was that at physiological ATP:ADP ratios (10:1) the population of symmetrical complexes was substantially decreased, even in the presence of non-foldable substrates. It has been previously observed that ADP is an allosteric regulator which increases the negative allostery between the rings such that they bind and hydrolyse ATP sequentially rather than simultaneously[77]. This hints that in physiological setting most probably the functional form is a GroEL:GroES asymmetric complex. To experimentally prove this hypothesis the best approach to use would be to employ recently developed method[159] of focused ion beam (FIB) milling to open an *E. coli* cell and image the interior by cryo electron-tomography. One could then determine the relative populations of apo GroEL, GroEL:GroES and GroEL:GroES$_2$ complexes inside the cell. Furthermore, such a study could then be extended further and investigate the influence of stress (heat, oxidation), starvation or different growth states on relative ratios of the aforementioned populations. The results of such a study would give a definite answer as to which form is functional and put an end to this long-standing debate. Going further, it is interesting to look at cycles of other group I chaperonins such a mitochondrial Hsp60 (mt-Hsp60) and group II chaperonins, for example thermosome or TRiC. It was shown for mt-Hsp60 that in presence of ATP and absence of substrate, complexes without Hsp10, single and doubly bound Hsp10 complexes and even single ring with and without Hsp10 were populated[160]. It seems that in this case the regulation of complex composition is not as strict as for the bacterial GroEL/ES system. This could be due to differences in interfaces between the rings, resulting in decreased allosteric communication, or it could be that dissociation/association kinetics are different and allow for intermediates to be detected. On the other hand, for group II chaperonins, it was proposed that both cavities function simultaneously, requiring a different mechanism for opening of the rings[161]. However, group II chaperonins have not been investigated in the presence of ATP using cryo electron microscopy. Such a study would show the distribution of population of open, single closed or double closed particles. The most appropriate group II chaperonin would be TRiC, because the resulting reconstructions could be compared to already published structures of TRiC along the conformational cycle trapped by different nucleotide analogs[161]. It seems that during evolution the mechanistic cycle of chaperonin conformational states has simplified but on the other side the complexity of proteins that chaperonins have to handle has increased[162]. Of course, it is a pure speculation that there is any functional correlation between these two phenomena. Nevertheless, the study of the cycle, especially of group II chaperonins, still offers a very exciting field of research.

4.3 Article 3 – Structure and mechanism of the Rubisco assembly chaperone Raf1

Rubisco is the most abundant enzyme in nature. It catalyses the incorporation of atmospheric CO_2 into five carbon sugar ribulose-1,5-bisphosphate and is therefore directly or indirectly responsible for the production of all biomass. Form I Rubisco is the most common and it consists of eight large subunits (RbcL) and eight small subunits (RbcS). RbcL subunits are arranged as a tetramer of antiparallel dimers, forming barrel-like structure which is capped on both ends by four RbcS subunits. Form I RbcL evolved from a simpler Form II RbcL[163]. RbcL of both forms needs assistance of chaperonins in order to fold. Form II RbcL is upon completion of assisted folding released from the chaperonin and can assemble into a functional dimer. On the other hand, form I RbcL, after completion of assisted refolding, remains in a structurally labile state which perpetually rebinds to chaperonin. As was shown for a cyanobacterial Form I RbcL from *Synechococcus sp 6301* (Syn6301), it is the loop region in the TIM-barrel domain between β-sheet and α-helix with a rather hydrophobic motif (LGSTIKPKLGL) that is solvent exposed and thus facilitates rebinding to GroEL[115]. To be released and assembled RbcL needs assistance from additional factors. Assembly of the final $RbcL_8$:$RbcS_8$ is preceded by formation of the $RbcL_8$ core, followed by docking of RbcS. When various cyanobacterial form I Rubiscos were expressed in heterologous host systems the level of functional enzyme reached only 0.1-10%[164]. Biogenesis of hexadecameric Rubisco has become a prime example of assisted assembly. The first described Rubisco assembly chaperone RbcX from cyanobacteria *Anabaena sp.* CA allowed for Syn6301 Rubisco reconstitution in the presence of GroEL/ES up to 40% yield[115]. The increased but still relatively low yield was due to high affinity of *Anabaena sp.* CA RbcX which allowed the release from GroEL but prevented the binding of RbcS. This finding suggested the existence of additional factors that drive Rubisco assembly to completion. Indeed, a recent photosynthetic mutant library screen in maize identified a gene, *Raf1*, responsible for the reduced amount of Rubisco holoenzyme despite unchanged levels in transcription and translation of RbcL and RbcS[165]. The RbcL subunit was found bound to GroEL suggesting that Raf1 acts downstream of the chaperonin.

We have structurally and mechanistically described plant and cyanobacterial Rubisco accumulation factor 1 (Raf1), a Rubisco assembly chaperone. Our X-ray structure of *Arabidopsis thaliana* Raf1 showed that Raf1 dimerizes through its β-sheet dimerization domain and is flanked by α-domains. Our negative stain EM reconstruction of Syn6301-$RbcL_8$:Syn7942-$Raf1_4$ showed that the Raf1 dimer interacts with the Rubisco antiparallel dimer along the equator via its β-domains and that Raf1 α-domains bracket the $RbcL_2$ antiparallel dimer by embracing it at the top and bottom sides of $RbcL_2$, overlapping with RbcS binding sites. In our low resolution EM reconstruction, parts of the missing density at the interface of $RbcL_2$ and Raf1 indicate that the complex is highly dynamic. It is exactly this dynamic nature of interaction that allows RbcS to displace Raf1 α-domains and form the final $RbcL_8$:$RbcS_8$ holoenzyme. Therefore, the $RbcL_8$:$Raf1_4$ structure represents the last structural intermediate in the assembly process. Earlier in the assembly pathway, Raf1 is most probably involved in dimerization of RbcL which prevents rebinding to the chaperonin. The $RbcL_2$:Raf1 complex is a major structural intermediate and precedes assembly of the $RbcL_8$:$Raf1_4$ complex. As already discussed, Raf1 is not the sole assembly chaperone involved in Rubisco biogenesis. RbcX, described above, is a structurally unrelated, ~30 kDa 'boomerang' shaped dimer that binds to RbcL via a set of different interaction sites, forming a less extensive interaction surface than $RbcL_2$:Raf1[115]. The different sites of interaction define a different mode of action for RbcX, which binds the C-terminal tail of one RbcL subunit and N-terminal domain of adjacent RbcL subunit, thus functioning as a molecular clamp that brings together two RbcL subunits. This leads to assembly of an $RbcL_8$:$RbcX_8$ complex. Raf1 performs essentially the same function as RbcX, which immediately raises the question whether the two assembly chaperones functionally cooperate or are redundant and thus function in parallel pathways. In contrast to Raf1, RbcX binding sites do not overlap with RbcS and thus the $RbcL_8$:$RbcX_8$ complex may be an additional assembly intermediate which keeps the RbcL core in an assembly competent state when protein levels of RbcS are high enough to complete the last step. Moreover, genetic studies showed strict co-occurrence of RbcX and Raf1 in photosynthetic organisms with green-type Form I Rubsico, providing another clue for a functional cooperation of the two chaperones.

However, not all Form I Rubisco enzymes depend on assembly chaperones. It was recently shown for a red-type Form I Rubisco from a proteobacterium *Rhodobacter sphaeroides* that RbcS alone is sufficient for assembly to the $RbcL_8$:$RbcS_8$ complex[166]. The red-type RbcS subunit has an additional C-terminal loop that has a role in Rubisco assembly. Interestingly, red type large subunits were not found to be bound to GroEL after refolding in the absence of RbcS. This suggests that red type RbcL are less prone to rebinding to GroEL and easily dimerize. Moreover, a homology search did not reveal any homologous green type assembly chaperones in prokaryotic or eukaryotic organisms with red type Rubsico. This led to the conclusion that the assembly processes evolved in different ways. There is an additional difference between green-type and red-type organisms; in non-green algae RbcS are expressed in chloroplasts whereas in green algae and plants RbcS is encoded in nuclear genome and has to be imported into cholorplasts[167]. This might explain the need for assembly chaperones in green-type organisms so that Raf1 would be involved in the assembly of the $RbcL_8$ core and RbcX would stabilize it until RbcS is imported into chloroplast. It is interesting to note that such an important enzyme as Rubisco depends on many factors for its biogenesis rather than having developed a more robust folding process.

The dependence on the assembly chaperone RbcX imposes a limit on the mutational space that Rubisco can explore, thus keeping it trapped in a form that is not enzymatically optimal[168]. On the other hand, the upstream chaperonin was shown to buffer mutations and thus allow further Rubisco evolution. The exact role of Raf1 is still to be determined. The ability to mutate and evolve plant Rubisco to a better enzyme, either in terms of its catalytic turnover or increased specificity for CO_2 over O_2 without disrupting its structure is of great interest for agriculture since such genetic manipulations would result in increased crop production. Unfortunately, to date no one has been able to reconstitute plant Rubisco in a heterologous system or in vitro. This suggest that additional factors for plant Rubisco assembly still await to be discovered.

4.4 Article 4 – Mechanism of Enzyme Repair by the AAA+ Chaperone Rubisco Activase

In addition to being a very inefficient enzyme, Rubisco is also error-prone in its multi-step carboxylation reaction[105]. Sugars that form as mis-fire products, such as xylulose-1,5-bisphosphate (XuBP), can bind tightly to the active site of Rubisco and inhibit its catalytic action. Rubisco inhibition can also occur due to premature binding of RuBP[117]. To release the inhibitor from the active site and allow for re-activation of Rubisco on a biologically relevant time scale, many photosynthetic organisms need help of a Rubisco activase (Rca)[105]. Rca belongs to a diverse family of AAA+ proteins which are involved in remodelling, disassembly or unfolding of proteins in the cell[118]. Rca remodelling action on inhibited Rubisco, which leads to metabolic repair, has emerged as an important paradigm to understand these mechanisms[126,127]. We have studied the remodelling mechanism of Rca from a proteobacterium *Rhodobacter sphaeroides* (RsRca), which harbours a red-type Rubisco. To become functional, RsRca needs to hexamerize. Hexamerization occurs in the presence of ATP and the substrate RuBP. The RsRca ring-like hexamer has a central pore with conserved pore loop residues (YIG) that are implicated in interaction with the RsRbcL C-terminal tail[126]. From our hydrogen/deuterium exchange (H/DX) experiments coupled to the mass-spectrometry (MS), we were able to identify RsRbcL regions that were remodelled by RsRca. It is astonishing that the action of RsRca is limited to a precisely defined region consisting of the C-terminal tail, loop 6 of C-terminal domain and '60s loop of the adjacent N-terminal domain of RbcL. These are the exact elements that build the multi-layered active site of Rubisco[113]. This finding argues in favour of a restricted region being remodelled as opposed to a global unfolding event. Such a mode of action is in contrast to ClpX, another AAA+ protein responsible for unfolding of substrates with an ssrA tag[169]. ClpX works in a highly processive manner, threading the substrate through its pore which in many cases leads to complete unfolding.

Another feature of the RsRca interaction with Rubisco is its transient nature (cross linking with glutaraldehyde was required to trap the complex). It might be that a difference in interaction strength determines the specific mechanism of remodelling. Additionally, the ATPase activity of ClpX is stimulated to higher levels (7 fold) by the interacting substrate as compared to RsRca. Remarkably, the pore loop residues of ClpX (YVG) are similar to RsRca, which excludes them as a reason for the difference in remodelling action. Thus, one could envision RsRca working by a 'kiss and run' mechanism where the weak interaction and slow processivity of RsRca would allow only for a short remodelling action. This would disrupt the active site just enough to open it and release the bound inhibitor. Our single particle cryo-EM reconstruction showed in more detail that RsRca binds to a corner of the Rubisco complex and is in a tilted orientation, interacting with both RbcL and RbcS.

The RsRca pore is positioned above the C-terminal helix of RbcL (helix 17), which is the starting point of the C-terminal tail that covers the active site. Large regions of the active site are missing in density, confirming that the RbcL structure is locally disrupted. Additionally, our cryoEM structure shows destabilization of helices 1 and 2 of RbcS, which was not observed by the HDX/MS experiments. RsRca itself is missing density as well, which could be due to distortions by glutaraldehyde cross-linking or more probably due to movement of RsRca on Rubisco. It is tempting to speculate that the structure we reconstructed is a morph of Rca, bound side-on to Rubisco where it initially interacts with the C-terminal tail, and the end state is closer to the top of the complex. Such a movement would peel off the C-terminal tail from the body of the enzyme. Our 3D classification, however, was not able to identify such classes. It might be that the structural differences are too small to be effectively differentiated by the algorithm. Our crosslinking experiment showed that the major interaction site on Rca is helix 5, where the functionally critical lysine residue 123 is located. As mentioned above, RsRca engages with the RbcL C-terminal tail. Crucial for the interaction, as shown by mutational analysis, are the last four residues of the tail[126]. In addition, it was speculated that immobilisation of the otherwise highly dynamic C-terminal strand is another important factor for successful interaction with RsRca. We performed an in-depth structural analysis of the closed inhibited state and observed that Asp474, Asp477 and Phe478, which are conserved in red type Rubisco, form a triad that might be crucial in locking the C-terminal strand over the catalytic site in a latch-like manner. Indeed, mutation of these aspartates into asparagines resulted in decreased Rubisco activity and inability of RsRca to reactivate inhibited Rubisco. This confirmed that the C-terminal tail has to be properly positioned to the body of enzyme such that RsRca with its central pore is able to engage with the last four residues of the C-terminal tail. This suggests a model where RsRca-induced conformational remodelling is initiated by breaking the contacts formed by Asp474 and Asp477, thus triggering a sequence of conformational changes. This results in destabilization of the 60s loop, leading to retraction of loop 6 and finally opening of the active site release of the inhibitory sugar. Only one RsRca per Rubisco active site is involved in reactivation. However, we observed from our cross-linking experiments and from our cryoEM 2D class averages that up to three RsRca can be bound, albeit in different positions, indicating simultaneous remodelling of three active sites.

It is still not clear, how RsRca differentiates between two closed states of Rubisco, namely the inhibited state, and the state where the productive carboxylation reaction is ongoing. It might be that the RsRca binding is kinetically controlled, meaning that when the enzyme is in active mode the transition between the open and the closed states is faster than when the enzyme is inhibited, thereby decreasing the time window in which RsRca can interact with Rubisco. It would therefore be of great interest to study the kinetics of opening of the active site. One would have to find appropriate positions for fluorophore labelling and try to measure kinetics by FRET. Such an experimental setup would allow the mechanism of Rubisco catalysis to be probed in even greater detail under a variety of conditions. It would also be interesting to study the activation process from the perspective of RsRca. For example, how many ATP molecules does a single RsRca need to hydrolyse in order to open the active site? Further, are ATP molecules hydrolysed simultaneously or sequentially? Dissection of activase conformational changes along the ATP hydrolysis reaction coordinate while bound to Rubisco would also be of interest. This could be achieved by 3D reconstruction of activase:Rubisco complexes in the presence of ADP-BeF_x (ATP bound state), ADP-AlF_x (ATP hydrolysis state) and ADP-VO_4 (post-hydrolysis state) given that the complexes are stable enough. Another unresolved question is the force that RsRca exerts on Rubisco, which could be experimentally approached using single-molecule force spectroscopy. Also, how exactly do pore loop residues and the ATPase activity determine grip strength and processivity of RsRca? Would it be possible to mutate RsRca in such a way that it would interact with Rubisco and completely unfold it?

It is bewildering how the different AAA+ proteins evolved to be specifically adapted to the needs of their cognate substrates. Along these lines, it has been shown that red-type activases cannot activate a green-type plant Rubisco and vice versa. How exactly the plant activase remodels plant Rubisco still remains to be answered. Plant activases have an additional N-terminal domain which was shown to be important for interaction with Rubisco[105]. Additionally it is not clear where the plant activase would pull the Rubisco in order to open the active site, since the C-terminal tail is shorter by ~10 residues as compared to red-type Rubisco[113]. Another interesting case are cyanobacterial Rubiscos, for which no inhibited forms have been described[105]. These organisms do, however, have Rca like proteins with an RbcS like domain. This points to yet another possible mode of interaction with Rubisco. The field of Rubisco activation still has many research challenges, and offers an alternative approach to engineering a better Rubisco by modulating the interaction and activity of its cognate activase.

5 References

1. Nelson, D. L. & Cox, M. M. *Lehninger Principles of Biochemistry*. (2008).
2. Ramachandran, G. N. & Sasisekharan, V. Conformation of Polypeptides and Proteins. *Adv. Protein Chem.* **23,** 283–437 (1968).
3. Berg, J. M., Tymoczko, J. L. & Stryer, L. Biochemistry. (2006).
4. Pauling, L., Corey, R. B. The pleated sheet, a new layer configuration of polypeptide chains. *Proc. Natl. Acad. Sci. U. S. A.* **37,** 521–526 (1951).
5. Murzin, A. G., Brenner, S. E., Hubbard, T. & Chothia, C. SCOP: A structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* **247,** 536–540 (1995).
6. Haber, E., Anfinsen, C. B. Side-chain interactions governing the pairing of half-cystine residues in ribonuclease. *J. Biol. Chem.* (1962).
7. Anfinsen, C. B. Principles that Govern the Folding of Protein Chains. *Science (80-.).* **181,** 223–230 (1973).
8. Levinthal, C. Are there pathways for protein folding? *J. Chim. Phys. Physico-Chimie Biol.* **65,** 44–45 (1968).
9. Karplus, M. & Weaver, D. L. Protein folding dynamics: the diffusion-collision model and experimental data. *Protein Sci.* **3,** 650–68 (1994).
10. Wetlaufer, D. B. Nucleation, Rapid Folding, and Globular Intrachain Regions in Proteins. *Proc. Natl. Acad. Sci.* **70,** 697–701 (1973).
11. Kauzmann, W. Some Factors in the Interpretation of Protein Denaturation. *Adv. Protein Chem.* **14,** 1–63 (1959).
12. Nickson, A. A. & Clarke, J. What lessons can be learned from studying the folding of homologous proteins? *Methods* **52,** 38–50 (2010).
13. Plotkin, S. S. & Onuchic, J. N. Understanding protein folding with energy landscape theory Part I: Basic concepts. *Q. Rev. Biophys.* **35,** (2002).
14. Radford, S. E. Protein folding: Progress made and promises ahead. *Trends in Biochemical Sciences* **25,** 611–618 (2000).
15. Lorch, M., Mason, J. M., Clarke, A. R. & Parker, M. J. Effects of core mutations on the folding of a β-sheet protein: Implications for backbone organization in the I-state. *Biochemistry* **38,** 1377–1385 (1999).
16. Brockwell, D. J., Smith, D. A. & Radford, S. E. Protein folding mechanisms: New methods and emerging ideas. *Current Opinion in Structural Biology* **10,** 16–25 (2000).
17. Shtilerman, M. Chaperonin Function: Folding by Forced Unfolding. *Science (80-.).* **284,** 822–825 (1999).
18. Fink, A. L. Chaperone-mediated protein folding. *Physiol. Rev.* **79,** 425–49 (1999).

19. Fersht, A. R., Matouschek, A. & Serrano, L. The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding. *J. Mol. Biol.* **224,** 771–782 (1992).

20. Haustein, E. & Schwille, P. Single-molecule spectroscopic methods. *Current Opinion in Structural Biology* **14,** 531–540 (2004).

21. Sauer, M. & Neuweiler, H. PET-FCS: Probing rapid structural fluctuations of proteins and nucleic acids by single-molecule fluorescence quenching. *Methods Mol. Biol.* **1076,** 597–615 (2014).

22. Neuweiler, H., Löllmann, M., Doose, S. & Sauer, M. Dynamics of Unfolded Polypeptide Chains in Crowded Environment Studied by Fluorescence Correlation Spectroscopy. *J. Mol. Biol.* **365,** 856–869 (2007).

23. Kim, S. A., Heinze, K. G. & Schwille, P. Fluorescence correlation spectroscopy in living cells. *Nat. Methods* **4,** 963–973 (2007).

24. Haldar, S., Mitra, S. & Chattopadhyay, K. Role of protein stabilizers on the conformation of the unfolded state of cytochrome c and its early folding kinetics: Investigation at single molecular resolution. *J. Biol. Chem.* **285,** 25314–25323 (2010).

25. Teufel, D. P., Johnson, C. M., Lum, J. K. & Neuweiler, H. Backbone-driven collapse in unfolded protein chains. *J. Mol. Biol.* **409,** 250–262 (2011).

26. Frank, G. a *et al.* Out-of-equilibrium conformational cycling of GroEL under saturating ATP concentrations. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 6270–6274 (2010).

27. Gupta, A. J. A single molecule study of the GroEL active cage mechanism. (2015).

28. Kuhlbrandt, W. The Resolution Revolution. *Science (80-.).* **343,** 1443–1444 (2014).

29. Milazzo, A.-C. *et al.* Initial evaluation of a direct detection device detector for single particle cryo-electron microscopy. *J. Struct. Biol.* **176,** 404–408 (2011).

30. Scheres, S. H. W. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180,** 519–530 (2012).

31. Bartesaghi, A. *et al.* 2.2 A resolution cryo-EM structure of -galactosidase in complex with a cell-permeant inhibitor. *Science (80-.).* **348,** 1147–1151 (2015).

32. Orlova, E. V. & Saibil, H. R. Structural analysis of macromolecular assemblies by electron microscopy. *Chemical Reviews* **111,** 7710–7748 (2011).

33. Cheng, Y. Single-particle Cryo-EM at crystallographic resolution. *Cell* **161,** 450–457 (2015).

34. Skiniotis, G. & Southworth, D. R. Single-particle cryo-electron microscopy of macromolecular complexes. *Microscopy* **65,** 9–22 (2016).

35. Lu, P. *et al.* Three-dimensional structure of human gamma-secretase. *Nature* **512,** 166–170 (2014).

36. Sun, M. *et al.* Dynamical features of the Plasmodium falciparum ribosome during

translation. *Nucleic Acids Res.* **43,** 10515–10524 (2015).

37. Yan, C., Wan, R., Bai, R., Huang, G. & Shi, Y. Structure of a yeast activated spliceosome at 3.5 A resolution. *Science (80-.).* **353,** 904–911 (2016).

38. Zimmerman, S. B. & Minton, A. P. Macromolecular crowding: biochemical, biophysical, and physiological consequences. *Annu. Rev. Biophys. Biomol. Struct.* **22,** 27–65 (1993).

39. Minton, A P. & Wilf, J. Effect of macromolecular crowding upon the structure and function of an enzyme: glyceraldehyde-3-phosphate dehydrogenase. *Biochemistry* **20,** 4821–4826 (1981).

40. Christiansen, A., Wang, Q., Cheung, M. S. & Wittung-Stafshede, P. Effects of macromolecular crowding agents on protein folding in vitro and in silico. *Biophysical Reviews* **5,** 137–145 (2013).

41. Wilcox, A. E., LoConte, M. A. & Slade, K. M. Effects of macromolecular crowding on alcohol dehydrogenase activity are substrate-dependent. *Biochemistry* **55,** 3550–3558 (2016).

42. Ellis, R. J. Macromolecular crowding: Obvious but underappreciated. *Trends in Biochemical Sciences* **26,** 597–604 (2001).

43. Balchin, D., Hayer-Hartl, M. & Hartl, F. U. In vivo aspects of protein folding and quality control. *Science (80-.).* **353,** aac4354-aac4354 (2016).

44. Hipp, M. S., Park, S. H. & Hartl, U. U. Proteostasis impairment in protein-misfolding and -aggregation diseases. *Trends in Cell Biology* **24,** 506–514 (2014).

45. Labbadia, J. & Morimoto, R. I. The Biology of Proteostasis in Aging and Disease. *Annu. Rev. Biochem.* **84,** 435–464 (2015).

46. Kim, Y. E., Hipp, M. S., Bracher, A., Hayer-Hartl, M. & Ulrich Hartl, F. Molecular Chaperone Functions in Protein Folding and Proteostasis. *Annu. Rev. Biochem.* **82,** 323–355 (2013).

47. Hartl, F. U. Molecular chaperones in cellular protein folding. *Nature* **381,** 571–579 (1996).

48. Lindquist, S. Varying patterns of protein synthesis in Drosophila during heat shock: Implications for regulation. *Dev. Biol.* **77,** 463–479 (1980).

49. Mayer, M. P. Gymnastics of molecular chaperones. *Molecular Cell* **39,** 321–331 (2010).

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

51. Haslbeck, M., Franzmann, T., Weinfurtner, D. & Buchner, J. Some like it hot: the structure and function of small heat-shock proteins. *Nat. Struct. Mol. Biol.* **12,** 842–846 (2005).

52. Stull, F., Koldewey, P., Humes, J. R., Radford, S. E. & Bardwell, J. C. A. Substrate protein folds while it is bound to the ATP-independent chaperone Spy. *Nat. Struct. Mol. Biol.* **23,** 53–58 (2015).

53. Hartl, F. U. Molecular Chaperones in the Cytosol: from Nascent Chain to Folded Protein. *Science (80-.).* **295,** 1852–1858 (2002).

54. Jaenicke, R. Protein stability and protein folding. *Ciba Found. Symp.* **161,** 206-216-221 (1991).

55. Brandt, F. *et al.* The Native 3D Organization of Bacterial Polysomes. *Cell* **136,** 261–271 (2009).

56. del Alamo, M. *et al.* Defining the specificity of cotranslationally acting chaperones by systematic analysis of mRNAs associated with ribosome-nascent chain complexes. *PLoS Biol.* **9,** (2011).

57. Calloni, G. *et al.* DnaK Functions as a Central Hub in the E. coli Chaperone Network. *Cell Rep.* **1,** 251–264 (2012).

58. Bukau, B., Deuerling, E., Pfund, C. & Craig, E. A. Getting newly synthesized proteins into shape. *Cell* **101,** 119–122 (2000).

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

60. Hoffmann, A., Bukau, B. & Kramer, G. Structure and function of the molecular chaperone Trigger Factor. *Biochim. Biophys. Acta - Mol. Cell Res.* **1803,** 650–661 (2010).

61. Oh, E. *et al.* Selective ribosome profiling reveals the cotranslational chaperone action of trigger factor in vivo. *Cell* **147,** 1295–1308 (2011).

62. O'Brien, E. P., Christodoulou, J., Vendruscolo, M. & Dobson, C. M. Trigger factor slows Co-translational folding through kinetic trapping while sterically protecting the nascent chain from aberrant cytosolic interactions. *J. Am. Chem. Soc.* **134,** 10920–10932 (2012).

63. Rospert, S., Dubaquié, Y. & Gautschi, M. Nascent-polypeptide-associated complex. *Cell. Mol. Life Sci.* **59,** 1632–1639 (2002).

64. Koplin, A. *et al.* A dual function for chaperones SSB-RAC and the NAC nascent polypeptide-associated complex on ribosomes. *J. Cell Biol.* **189,** 57–68 (2010).

65. Frydman, J. Folding of newly translated proteins in vivo: the role of molecular chaperones. *Annu. Rev. Biochem.* **70,** 603–47 (2001).

66. Saibil, H. Chaperone machines for protein folding, unfolding and disaggregation. *Nat. Rev. Mol. Cell Biol.* **14,** 630–642 (2013).

67. Sharma, S. *et al.* Monitoring protein conformation along the pathway of chaperonin-assisted folding. *Cell* **133,** 142–53 (2008).

68. Shorter, J. The mammalian disaggregase machinery: Hsp110 synergizes with Hsp70 and Hsp40 to catalyze protein disaggregation and reactivation in a cell-free system. *PLoS One* **6,** (2011).

69. Kampinga, H. H. & Craig, E. A. The HSP70 chaperone machinery: J proteins as drivers of functional specificity. *Nat. Rev. Mol. Cell Biol.* **11,** 579–592 (2010).

70. Mayer, M. P. Hsp70 chaperone dynamics and molecular mechanism. *Trends in*

Biochemical Sciences **38,** 507–514 (2013).

71. Kerner, M. J. *et al.* Proteome-wide analysis of chaperonin-dependent protein folding in Escherichia coli. *Cell* **122,** 209–220 (2005).

72. Yebenes, H., Mesa, P., Muñoz, I. G., Montoya, G. & Valpuesta, J. M. Chaperonins: Two rings for folding. *Trends in Biochemical Sciences* **36,** 424–432 (2011).

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

74. Braig, K. *et al.* The crystal structure of the bacterial chaperonin GroEL at 2.8 A. *Nature* **371,** 578–86 (1994).

75. Xu, Z., Horwich, a L. & Sigler, P. B. The crystal structure of the asymmetric GroEL-GroES-(ADP)7 chaperonin complex. *Nature* **388,** 741–750 (1997).

76. Elad, N. *et al.* Topologies of a Substrate Protein Bound to the Chaperonin GroEL. *Mol. Cell* **26,** 415–426 (2007).

77. Gruber, R. & Horovitz, A. Allosteric Mechanisms in Chaperonin Machines. *Chemical Reviews* **116,** 6588–6606 (2016).

78. Horwich, A. Chaperonin-mediated protein folding. *J. Biol. Chem.* 23622–23632 (2013). doi:10.1074/jbc.X113.497321

79. Clare, D. K. *et al.* ATP-triggered conformational changes delineate substrate-binding and -folding mechanics of the GroEL chaperonin. *Cell* **149,** 113–123 (2012).

80. Cliff, M. J., Limpkin, C., Cameron, A., Burston, S. G. & Clarke, A. R. Elucidation of steps in the capture of a protein substrate for efficient encapsulation by GroE. *J. Biol. Chem.* **281,** 21266–21275 (2006).

81. Chaudhuri, T. K., Farr, G. W., Fenton, W. a, Rospert, S. & Horwich, a L. GroEL/GroES-mediated folding of a protein too large to be encapsulated. *Cell* **107,** 235–46 (2001).

82. England, J., Lucent, D. & Pande, V. Rattling the cage: computational models of chaperonin-mediated protein folding. *Current Opinion in Structural Biology* **18,** 163–169 (2008).

83. Horwich, A. L., Apetri, A. C. & Fenton, W. A. The GroEL/GroES cis cavity as a passive anti-aggregation device. *FEBS Letters* **583,** 2654–2662 (2009).

84. Tyagi, N. K., Fenton, W. a, Deniz, A. a & Horwich, A. L. Double mutant MBP refolds at same rate in free solution as inside the GroEL/GroES chaperonin chamber when aggregation in free solution is prevented. *FEBS Lett.* **585,** 1969–72 (2011).

85. Corsepius, N. C. & Lorimer, G. H. Measuring how much work the chaperone GroEL can do. *Proc. Natl. Acad. Sci. U. S. A.* **110,** E2451-9 (2013).

86. Yang, D., Ye, X. & Lorimer, G. H. Symmetric GroEL:GroES2 complexes are the protein-folding functional form of the chaperonin nanomachine. *Proc. Natl. Acad. Sci.* **110,** E4298–E4305 (2013).

87. Georgescauld, F. *et al.* GroEL/ES chaperonin modulates the mechanism and

accelerates the rate of TIM-barrel domain folding. *Cell* **157,** 922–934 (2014).

88. Stan, G., Thirumalai, D., Lorimer, G. H. & Brooks, B. R. Annealing function of GroEL: Structural and bioinformatic analysis. *Biophysical Chemistry* **100,** 453–467 (2003).

89. Weissman, J. S. *et al.* Mechanism of GroEL action: Productive release of polypeptide from a sequestered position under groes. *Cell* **83,** 577–587 (1995).

90. Chakraborty, K. *et al.* Chaperonin-catalyzed rescue of kinetically trapped states in protein folding. *Cell* **142,** 112–22 (2010).

91. Tang, Y.-C. *et al.* Structural features of the GroEL-GroES nano-cage required for rapid folding of encapsulated protein. *Cell* **125,** 903–14 (2006).

92. Jewett, A. I., Baumketner, A. & Shea, J.-E. Accelerated folding in the weak hydrophobic environment of a chaperonin cavity: Creation of an alternate fast folding pathway. *Proc. Natl. Acad. Sci.* **101,** 13192–13197 (2004).

93. Brinker, A *et al.* Dual function of protein confinement in chaperonin-assisted protein folding. *Cell* **107,** 223–33 (2001).

94. Baumketner, A., Jewett, A. & Shea, J. E. Effects of confinement in chaperonin assisted protein folding: Rate enhancement by decreasing the roughness of the folding energy landscape. *J. Mol. Biol.* **332,** 701–713 (2003).

95. Motojima, F. & Yoshida, M. Polypeptide in the chaperonin cage partly protrudes out and then folds inside or escapes outside. *EMBO J.* **29,** 4008–19 (2010).

96. Motojima, F., Motojima-Miyazaki, Y. & Yoshida, M. Revisiting the contribution of negative charges on the chaperonin cage wall to the acceleration of protein folding. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 15740–5 (2012).

97. Brocchieri, L. & Karlin, S. Conservation among HSP60 sequences in relation to structure, function, and evolution. *Protein Sci.* **9,** 476–86 (2000).

98. Sorin, E. J. & Pande, V. S. Nanotube confinement denatures protein helices. *J. Am. Chem. Soc.* **128,** 6316–6317 (2006).

99. Sirur, A. & Best, R. B. Effects of interactions with the GroEL cavity on protein folding rates. *Biophys. J.* **104,** 1098–106 (2013).

100. England, J. L., Lucent, D. & Pande, V. S. A role for confined water in chaperonin function. *J. Am. Chem. Soc.* **130,** 11838–11839 (2008).

101. Franck, J. M. *et al.* Probing water density and dynamics in the chaperonin GroEL cavity. *J. Am. Chem. Soc.* **136,** 9396–9403 (2014).

102. Fujiwara, K., Ishihama, Y., Nakahigashi, K., Soga, T. & Taguchi, H. A systematic survey of in vivo obligate chaperonin-dependent substrates. *EMBO J.* **29,** 1552–1564 (2010).

103. Hofmann, H. *et al.* Role of Denatured-State Properties in Chaperonin Action Probed by Single-Molecule Spectroscopy. *Biophys. J.* **107,** 2891–2902 (2014).

104. Hauser, T., Popilka, L., Hartl, F. U. & Hayer-Hartl, M. Role of auxiliary proteins in Rubisco biogenesis and function. *Nat. plants* **1,** 1–11 (2015).

105. Bhat, J. Y., Thieulin-Pardo, G., Hartl, F. U. & Hayer-Hartl, M. Rubisco Activases: AAA+ Chaperones Adapted to Enzyme Repair. *Front. Mol. Biosci.* **4,** (2017).

106. Walker, B. J., VanLoocke, A., Bernacchi, C. J. & Ort, D. R. The Costs of Photorespiration to Food Production Now and in the Future. *Annu. Rev. Plant Biol.* **67,** 107–129 (2016).

107. Bloom, A. J. Photorespiration and nitrate assimilation: A major intersection between plant carbon and nitrogen. *Photosynth. Res.* **123,** 117–128 (2015).

108. Whitney, S. M., Houtz, R. L. & Alonso, H. Advancing Our Understanding and Capacity to Engineer Nature's CO2-Sequestering Enzyme, Rubisco. *PLANT Physiol.* **155,** 27–35 (2011).

109. Bracher, A., Whitney, S. M., Hartl, F. U. & Hayer-Hartl, M. Biogenesis and Metabolic Maintenance of Rubisco. *Annu. Rev. Plant Biol.* **68,** annurev-arplant-043015-111633 (2017).

110. Andersson, I. & Backlund, A. Structure and function of Rubisco. *Plant Physiol. Biochem.* **46,** 275–291 (2008).

111. Tabita, F. R. Microbial ribulose 1 , 5-bisphosphate carboxylase / oxygenase : A different perspective. *Plant Biotechnol.* **60,** 1–28 (1999).

112. Cleland, W. W., Andrews, T. J., Gutteridge, S., Hartman, F. C. & Lorimer, G. H. Mechanism of Rubisco: The Carbamate as General Base ^{X}. *Chem. Rev.* **98,** 549–562 (1998).

113. Duff, A. P., Andrews, T. J. & Curmi, P. M. . The transition between the open and closed states of rubisco is triggered by the inter-phosphate distance of the bound bisphosphate. *J. Mol. Biol.* **298,** 903–916 (2000).

114. Goloubinoff, P., Gatenby, A. A. & Lorimer, G. H. GroE heat-shock proteins promote assembly of foreign prokaryotic ribulose bisphosphate carboxylase oligomers in Escherichia coli. *Nature* **337,** 44–47 (1989).

115. Liu, C. *et al.* Coupled chaperone action in folding and assembly of hexadecameric Rubisco. *Nature* **463,** 197–202 (2010).

116. Feller, U., Anders, I. & Mae, T. Rubiscolytics: Fate of Rubisco after its enzymatic function in a cell is terminated. in *Journal of Experimental Botany* **59,** 1615–1624 (2008).

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

118. Neuwald, A. F., Aravind, L., Spouge, J. L. & Koonin, E. V. AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. *Genome Res.* **9,** 27–43 (1999).

119. Miller, J. M. & Enemark, E. J. Fundamental Characteristics of AAA+ Protein Family Structure and Function. *Archaea* **2016,** (2016).

120. Portis, A. R., Li, C., Wang, D. & Salvucci, M. E. Regulation of Rubisco activase and its interaction with Rubisco. in *Journal of Experimental Botany* **59,** 1597–1604

(2008).

121. Portis, A. R. & Salvucci, M. E. The discovery of Rubisco activase - Yet another story of serendipity. *Photosynthesis Research* **73,** 257–264 (2002).

122. Loganathan, N., Tsai, Y.-C. C. & Mueller-Cajar, O. Characterization of the heterooligomeric red-type rubisco activase from red algae. *Proc. Natl. Acad. Sci.* **113,** 14019–14024 (2016).

123. Tsai, Y.-C. C., Lapina, M. C., Bhushan, S. & Mueller-Cajar, O. Identification and characterization of multiple rubisco activases in chemoautotrophic bacteria. *Nat. Commun.* **6,** 8883 (2015).

124. Snider, J., Thibault, G. & Houry, W. A. The AAA+ superfamily of functionally diverse proteins. *Genome Biol.* **9,** 216 (2008).

125. Wendler, P., Ciniawsky, S., Kock, M. & Kube, S. Structure and function of the AAA+ nucleotide binding pocket. *Biochim. Biophys. Acta - Mol. Cell Res.* **1823,** 2–14 (2012).

126. Mueller-Cajar, O. *et al.* Structure and function of the AAA+ protein CbbX, a red-type Rubisco activase. *Nature* **479,** 194–199 (2011).

127. Stotz, M. *et al.* Structure of green-type Rubisco activase from tobacco. *Nat. Struct. Mol. Biol.* **18,** 1366–1370 (2011).

128. Wong, K. S. & Houry, W. A. Novel structural and functional insights into the MoxR family of AAA+ ATPases. *Journal of Structural Biology* **179,** 211–221 (2012).

129. Whittaker, C. A. Distribution and Evolution of von Willebrand/Integrin A Domains: Widely Dispersed Domains with Roles in Cell Adhesion and Elsewhere. *Mol. Biol. Cell* **13,** 3369–3387 (2002).

130. Somerville, C. R., Portis, A. R. & Ogren, W. L. A Mutant of Arabidopsis thaliana Which Lacks Activation of RuBP Carboxylase In Vivo. *Plant Physiol.* **70,** 381–7 (1982).

131. Zhang, N. & Portis, A. R. Mechanism of light regulation of Rubisco: A specific role for the larger Rubisco activase isoform involving reductive activation by thioredoxin-f. *Proc. Natl. Acad. Sci.* **96,** 9438–9443 (1999).

132. Carmo-Silva, A. E. & Salvucci, M. E. The Regulatory Properties of Rubisco Activase Differ among Species and Affect Photosynthetic Induction during Light Transitions. *PLANT Physiol.* **161,** 1645–1655 (2013).

133. Hazra, S., Henderson, J. N., Liles, K., Hilton, M. T. & Wachter, R. M. Regulation of ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) activase: Product inhibition, cooperativity, and magnesium activation. *J. Biol. Chem.* **290,** 24222–24236 (2015).

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

135. Spurlino, J. C., Lu, G. Y. & Quiocho, F. A. The 2.3-A resolution structure of the

maltose- or maltodextrin-binding protein, a primary receptor of bacterial active transport and chemotaxis. *J. Biol. Chem.* **266,** 5202–5219 (1991).

136. Chun, S. Y., Strobel, S., Bassford, P. & Randall, L. L. Folding of maltose-binding protein. Evidence for the identity of the rate- determining step in vivo and in vitro. *J. Biol. Chem.* **268,** 20855–20862 (1993).

137. Wang, J. D., Michelitsch, M. D. & Weissman, J. S. GroEL-GroES-mediated protein folding requires an intact central cavity. *Proc. Natl. Acad. Sci.* **95,** 12163–12168 (1998).

138. Sharma, S. *et al.* Monitoring Protein Conformation along the Pathway of Chaperonin-Assisted Folding. *Cell* **133,** 142–153 (2008).

139. Gupta, A. J., Haldar, S., Miličić, G., Hartl, F. U. & Hayer-Hartl, M. Active cage mechanism of chaperonin-assisted protein folding demonstrated at single-molecule level. *J. Mol. Biol.* **426,** 2739–2754 (2014).

140. Ye, X. & Lorimer, G. H. Substrate protein switches GroE chaperonins from asymmetric to symmetric cycling by catalyzing nucleotide exchange. *Proc. Natl. Acad. Sci.* **110,** E4289–E4297 (2013).

141. Sameshima, T. *et al.* Football- and bullet-shaped GroEL-GroES complexes coexist during the reaction cycle. *J. Biol. Chem.* **283,** 23765–23773 (2008).

142. Sameshima, T., Iizuka, R., Ueno, T. & Funatsu, T. Denatured proteins facilitate the formation of the football shaped GroEL-(GroES)2 complex. *Biochem. J.* **427,** 247–254 (2010).

143. Haldar, S. *et al.* Chaperonin-assisted protein folding: Relative population of asymmetric and symmetric GroEL:GroES complexes. *J. Mol. Biol.* **427,** 2244–2255 (2015).

144. Hauser, T. *et al.* Structure and mechanism of the Rubisco-assembly chaperone Raf1. *Nat. Struct. Mol. Biol.* **22,** 720–728 (2015).

145. Corsepius, N. C. & Lorimer, G. H. Measuring how much work the chaperone GroEL can do. *Proc. Natl. Acad. Sci.* **110,** E2451–E2459 (2013).

146. Motojima, F. & Yoshida, M. Polypeptide in the chaperonin cage partly protrudes out and then folds inside or escapes outside. *EMBO J.* **29,** 4008–19 (2010).

147. Zhou, H. X. & Dill, K. A. Stabilization of proteins in confined spaces. *Biochemistry* **40,** 11289–11293 (2001).

148. Dalal, V., Arya, S. & Mukhopadhyay, S. Confined Water in Amyloid-Competent Oligomers of the Prion Protein. *ChemPhysChem* **140306,** 2804–2807 (2016).

149. Qin, Y., Wang, L. & Zhong, D. Dynamics and mechanism of ultrafast water–protein interactions. *Proc. Natl. Acad. Sci.* **113,** 8424–8429 (2016).

150. Tang, Y.-C., Chang, H.-C., Chakraborty, K., Hartl, F. U. & Hayer-Hartl, M. Essential role of the chaperonin folding compartment in vivo. *EMBO J.* **27,** 1458–68 (2008).

151. Chen, D. H. *et al.* XVisualizing GroEL/ES in the act of encapsulating a folding protein. *Cell* **153,** (2013).

152. Farr, G. W., Fenton, W. a & Horwich, A. L. Perturbed ATPase activity and not ‘close confinement’ of substrate in the cis cavity affects rates of folding by tail-multiplied GroEL. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 5342–7 (2007).

153. Weaver, J. & Rye, H. S. The C-terminal tails of the bacterial chaperonin GroEL stimulate protein folding by directly altering the conformation of a substrate protein. *J. Biol. Chem.* **289,** 23219–23232 (2014).

154. Machida, K., Kono-Okada, A., Hongo, K., Mizobata, T. & Kawata, Y. Hydrophilic residues 526 KNDAAD 531 in the flexible C-terminal region of the chaperonin GroEL are critical for substrate protein folding within the central cavity. *J. Biol. Chem.* **283,** 6886–96 (2008).

155. Tompa, P. & Tompa, P. Intrinsically unstructured proteins. *Trends Biochem. Sci.* **27,** 527–533 (2002).

156. Sameshima, T. *et al.* Single-molecule study on the decay process of the football-shaped GroEL-GroES complex using zero-mode waveguides. *J. Biol. Chem.* **285,** 23159–23164 (2010).

157. Ranson, N. A. *et al.* Allosteric signaling of ATP hydrolysis in GroEL–GroES complexes. *Nat. Struct. Mol. Biol.* **13,** 147–152 (2006).

158. Van Duijn, E. *et al.* Tandem mass spectrometry of intact GroEL - Substrate complexes reveals substrate-specific conformational changes in the trans ring. *J. Am. Chem. Soc.* **128,** 4694–4702 (2006).

159. Villa, E., Schaffer, M., Plitzko, J. M. & Baumeister, W. Opening windows into the cell: Focused-ion-beam milling for cryo-electron tomography. *Current Opinion in Structural Biology* **23,** 771–777 (2013).

160. Levy-Rimler, G. *et al.* The effect of nucleotides and mitochondrial chaperonin 10 on the structure and chaperone activity of mitochondrial chaperonin 60. *Eur. J. Biochem.* **268,** 3465–3472 (2001).

161. Cong, Y. *et al.* Symmetry-free cryo-EM structures of the chaperonin TRiC along its ATPase-driven conformational cycle. *EMBO J.* **31,** 720–730 (2012).

162. Powers, E. T. & Balch, W. E. Diversity in the origins of proteostasis networks--a driver for protein function in evolution. *Nat. Rev. Mol. Cell Biol.* **14,** 237–48 (2013).

163. Hohmann-Marriott, M. F. & Blankenship, R. E. Evolution of Photosynthesis. *Annu. Rev. Plant Biol.* **62,** 515–548 (2011).

164. Larimer, F. W. & Soper, T. S. Overproduction of Anabaena 7120 ribulose-bisphosphate carboxylase/oxygenase in Escherichia coli. *Gene* **126,** 85–92 (1993).

165. Feiz, L. *et al.* Ribulose-1,5-Bis-Phosphate Carboxylase/Oxygenase Accumulation Factor1 Is Required for Holoenzyme Assembly in Maize. *Plant Cell* **24,** 3435–3446 (2012).

166. Joshi, J., Mueller-Cajar, O., Tsai, Y. C. C., Hartl, F. U. & Hayer-Hartl, M. Role of small subunit in mediating assembly of red-type Form I Rubisco. *J. Biol. Chem.* **290,** 1066–1074 (2015).

167. Soll, J. Protein import into chloroplasts. *Current Opinion in Plant Biology* **5,** 529–

535 (2002).

168. Durão, P. *et al.* Opposing effects of folding and assembly chaperones on evolvability of Rubisco. *Nat. Chem. Biol.* **11,** 148–155 (2015).

169. Olivares, A. O., Baker, T. A. & Sauer, R. T. Mechanistic insights into bacterial AAA+ proteases and protein-remodelling machines. *Nat. Rev. Microbiol.* **14,** 33–44 (2015).

170. Bhat, J *et al.* Mechanism of Enzyme Repair by the AAA+ Chaperone Rubisco Activase. *Mol. Cell*. 1-13 (2017)

6 Abbreviations

2PG	2-phosphoglycolate
3PG	3-phosphoglycerate
AAA+	ATPases associated with various cellular activites
Å	Ångstrom
aa	amino acid
ADP	adenosine 5’-diphosphate
AFM	atomic force microscopy
afRca1	Rubisco activase from *Acidithiobacillus ferrooxidans*
ALS	amyotrophic lateral sclerosis
At	*Arabidopsis thaliana*
ATP	adenosine-5’-triphosphate
Bis-ANS	4,4’-Dianilino-1,1’-binaphthyl-5,5’-disulfonic acid dipotassium salt
C	carbon
°C	degree Celsius
C_α	chiral carbon atom
CABP	2-carboxyarabinitol bisphosphate
CA1P	carboxyarabibitol 1-phosphate
CBB	Calvin-Benson-Bassham
CD	circular dichroism
cmRca	Rubisco activase from *Cyanidioschyzon merolae*
CO_2	carbon dioxide
Cpn	chaperonin
dcFCCS	dual colour fluorescence cross-correlation spectroscopy
DapA	dihydrodipicolinate synthase
DM-MBP	double mutant maltose binding protein
DNA	deoxyribonucleic acid
DnaJ	bacterial Hsp40 chaperone
DnaK	bacterial Hsp70 chaperone
E	non carbamylated Rubisco
ECM	carbamylated and Mg^{2+} bound Rubisco
ECMI	inhibited ECM form
EI	sugar phosphate inhibited E form
E.coli	*Escherichia coli*
EM	electron microscopy
ESI	electron spray ionization
FCS	fluorescence correlation microscopy
FIB	focused ion beam
fL	femtoliter
FRET	Förster resonance energy transfer
GroE	GroEL/ES system
GroEL	bacterial Hsp60 chaperonin
GroES	bacterial Hsp10 co-chaperonin
G3P	glyceraldehyde phosphate

hnRca	Rubisco activase from *Halothiobacillus neapolitanus*
HDX	hydrogen deuterium exchange
Hsp	heat shock protein
K^+	potassium ions
KABP	3-ketoarabinitol-1,5-bisphosphate
K_D	dissociation constant
kDa	kilodalton
k_{on}	association rate constant
k_{off}	dissociation rate constant
ms	millisecond
MS	mass spectrometry
MD	molecular dynamics
MDa	megadalton
MDH	malate dehydrogenase
MetK	S-adenosylmethionine synthase
mg	milligrams
Mg^{2+}	magnesium ions
ml	millilitres
mRNA	messenger RNA
N	nitrogen
NAC	nascent chain associated complex
NADPH	β-nicotinamide adenine dinucleotide 2'-phosphate
NBD	nucleotide binding domain
NEF	nucleotide exchange factor
NHS	N-hydroxysuccinimide
nm	nanometer
nM	nanomolar
NMR	nuclear magnetic resonance
ntRca	Rubisco activase from *Nicotiana tabacum*
O_2	oxygen
PDB	protein data bank
PDBP	2,3-pentodiulose-1,5-bisphosphate
PEP	phosphoenolpyruvate
PET	photoinduced electron transfer
PFD	prefoldin
Pi	phosphate group
pM	picomolar
PPI	peptidyl prolyl isomerase
R	relaxed
RAC	ribosome-associated complex
Raf1	Rubisco accumulation factor 1
Raf2	Rubisco accumulation factor 2
RbcL	Rubisco large subunit
RbcS	Rubisco small subunit
Rca	Rubisco activase
Rs	*Rhobobacter sphaeroides*
RsRca	Rubisco activase from *Rhobobacter sphaeroides*

Rubisco	Ribulose-1,5-bisphosphate carboxylase/oxygenase
RuBP	Ribulose-1,5-bisphosphate
RNA	ribonucleic acid
s	seconds
SAXS	small angle X-ray scattering
SBD	substrate-binding domain
SDS	sodium dodecyl sulphate
SmFRET	single-molecule Förster resonance energy transfer
SR1-EL	single ring version of GroEL
Syn	*Synechococcus*
T	tense
TF	trigger factor
TRiC/CCT	TCP-1 Ring Complex/Chaperonin containing TCP-1
TCP-1	T-complex 1
TIM	triose phosphate isomerase
τ_R	relaxation time
Trp	tryptophane
Tyr	tyrosine
µs	microsecond
UPS	ubiquitin proteasome system
UV	ultraviolet
VWA	von Willebrand factor A
XuBP	xylulose-1,5-bisphosphate