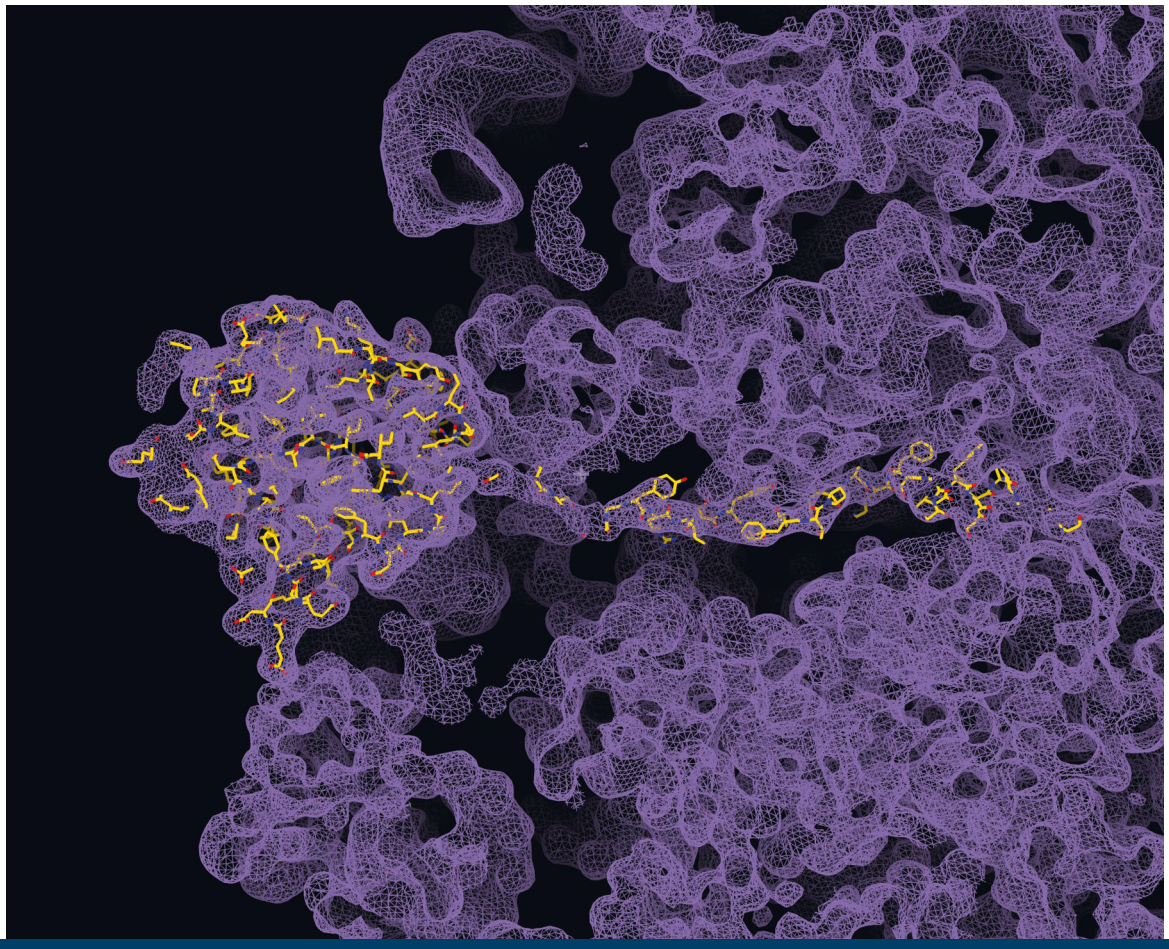


# Folding and structural studies of *saccharomyces cerevisiae* Phosphoglycerate Kinase

Nuno Bustorff

Schlüsseltechnologien / Key Technologies
Band / Volume 282
ISBN 978-3-95806-754-7

Mitglied der Helmholtz-Gemeinschaft

JÜLICH
Forschungszentrum

Schriften des Forschungszentrums Jülich
Reihe Schlüsseltechnologien / Key Technologies Band / Volume 282

Forschungszentrum Jülich GmbH
Ernst Ruska-Centrum für Mikroskopie und Spektroskopie mit Elektronen (ER-C)
Strukturbiologie (ER-C-3)

# Folding and structural studies of *saccharomyces cerevisiae* Phosphoglycerate Kinase

Nuno Bustorff

Schriften des Forschungszentrums Jülich
Reihe Schlüsseltechnologien / Key Technologies Band / Volume 282

ISSN 1866-1807 ISBN 978-3-95806-754-7

Bibliografische Information der Deutschen Nationalbibliothek.
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der
Deutschen Nationalbibliografie; detaillierte Bibliografische Daten
sind im Internet über http://dnb.d-nb.de abrufbar.

| | |
|---|---|
| Herausgeber und Vertrieb: | Forschungszentrum Jülich GmbH<br>Zentralbibliothek, Verlag<br>52425 Jülich<br>Tel.: +49 2461 61-5368<br>Fax: +49 2461 61-6103<br>zb-publikation@fz-juelich.de<br>www.fz-juelich.de/zb |
| Umschlaggestaltung: | Grafische Medien, Forschungszentrum Jülich GmbH |
| Druck: | Grafische Medien, Forschungszentrum Jülich GmbH |
| Copyright: | Forschungszentrum Jülich 2024 |

Schriften des Forschungszentrums Jülich
Reihe Schlüsseltechnologien / Key Technologies, Band / Volume 282

D 82 (Diss. RWTH Aachen University, 2024)

ISSN 1866-1807
ISBN 978-3-95806-754-7

Vollständig frei verfügbar über das Publikationsportal des Forschungszentrums Jülich (JuSER)
unter www.fz-juelich.de/zb/openaccess.

I hereby declare that I have created this work completely on my own and used no other sources or tools than the ones listed and that I have marked any citations accordingly.

Hiermit versichere ich, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie Zitate kenntlich gemacht habe.

*Aachen, March 2024*

*Nuno Rebello de Andrade Bustorff Silva*

# Acknowledgments

I would like to thank Prof. Dr. Jörg Fitter for welcoming me into his group and giving me a chance to do my Ph.D. on a topic I cherish so much. I am thankful for having access to state-of-the-art instrumentation and the opportunity to familiarize myself with all the critical biophysical techniques. Similarly, I would like to thank Prof. Dr. Carsten Sachse for accommodating our group in his institute and facilitating the institute's transition from ICS-5 to ER-C-3. The mentoring of both professors was crucial to my success during my project, and I am very thankful for this journey. Likewise, I would like to thank Prof. Dr. Schwaneberg for accepting to be my co-supervisor.

A special thanks to Dr. Alexandros Katranidis for his direct supervision and long discussions about the project. His care was essential in figuring out the right way to proceed.

Incorporating ICS-5 in ER-C-3 was a good surprise, allowing me to work with cryo-EM. I would like to thank all facility staff, Dr. Julio Ortiz, Dr. Thomas Heidler, Dr. Saba Shahzad, and Pia Sundermeyer, for keeping the microscopes in such good condition and for their vital help in setting the microscopes for data acquisition. I am very grateful to have had the opportunity to learn with Dr. Daniel Mann about processing and for his vital suggestions. Without his contribution to the computational infrastructure, this work would not be feasible.

I would like to thank every lab member who worked closely with me and helped me with experiments, suggestions, or scientific discussions. In particular, I would like to thank Dr. Olivera Korculanin, Dr. Iman Abdollahzadeh, Ramona Choy, Ilona Ritter, Olessya Yukhnovets, Farzaneh Vaghefikia, and Alida Meyer.

For all the people I met in the lab, a special thanks. The whole environment was lifted every time we were all together. I would like to give a special thanks to Dr. Sabrina Berkamp, Dr. Siavash Mostafavi, Claire Ortmann, and Lisa Jungbluth for their personal and professional support during these last four years.

Finally, I thank my wife, Inês Cabrita, for her lifelong support. To make me believe in myself and help me through all the good and bad moments during this journey.

In parallel, I would like to thank my parents, siblings, close family members, and friends for their full support and all the opportunities they facilitated during my life and for keeping me sane. Only all these things together made all this possible.

# Abstract

Proteins are synthesized within cells by ribosomes, and their functionality depend on the correct three-dimensional structure obtained through the process of folding. While the classical understanding of protein folding primarily focused on post-synthesis folding, recent research has shifted its emphasis toward unraveling the intricacies of folding during synthesis. Of particular interest in this pursuit are multi-domain proteins, which constitute over 70% of proteins in cells. Using classical single-molecule fluorescence resonance transfer efficiency (sm-FRET) studies, the unfolding/refolding transitions of a two-domain yeast phosphoglycerate kinase (yPGK) was explored as a model for multiple domain proteins. To enhance our understanding of a transition within a single-domain of full-length yPGK, I assessed two FRET pair variants within the N-terminal domain. Together with previous data we compared in total six different variants, for which we observed three distinct transitions in both domains: the first segment (positions 1–88 and 202–256, respectively) showed no transition (i.e. no change in the distance), while the second segment (positions 34–135 and 256–290, respectively) underwent a conventional two-state transition. Intriguingly, labeling the complete Rossmann motif (positions 1–135 and 202–290, respectively) revealed a compact intermediate state during GuHCl-induced unfolding in transitory conditions. The detailed understanding of N-terminal domain transitions was essential for comparative analyses with shorter-length proteins synthesized for co-translational folding studies. In order to investigate the ribosomal folding process of yPGK, I employed cryo-electron microscopy (cryo-EM). Four distinct ribosome nascent chain complex (RNC) structures were solved, each representing a nascent polypeptide of a varying length. Most structures depicted nascent chain density outside the ribosomal tunnel, and we observed for the first time the structure of a full-length protein nascent chain attached to the ribosome. In summary, the presented research advances our knowledge of yPGK folding transitions and offers novel routes for studying co-translational folding processes within RNC complexes using sm-FRET and cryo-EM.

# Überblick

Proteine werden in Zellen von Ribosomen synthetisiert, und ihre Funktionsfähigkeit hängt von der korrekten dreidimensionalen Struktur ab, die durch den Faltungsprozess erreicht wird. Während das klassische Verständnis der Proteinfaltung sich hauptsächlich auf die Faltung nach der Synthese konzentrierte, hat ein Teil der jüngeren Forschung ihren Schwerpunkt auf die Entschlüsselung der Feinheiten der Faltung während der Synthese verschoben. Besonderes Interesse gilt dabei den Multi-Domänen-Proteinen, die mehr als 70% aller Proteine in der Zelle ausmachen. Mithilfe klassischer Einzelmolekül-Fluoreszenz-Resonanz-Transfer-Effizienz (sm-FRET)-Studien wurden Entfaltungs-/Rückfaltungsübergänge eines zwei-domänen Proteins, der Phoglyceratkinase aus Hefe (yPGK) als Modell für Proteine mit mehreren Domänen untersucht. Um unser Verständnis für Faltungsübergänge einzelner Domänen in der yPGK zu vertiefen, habe ich zwei FRET-Paar-Varianten innerhalb der N-terminalen Domäne der yPGK vermessen und charakterisiert. Zusammen mit vorangegangenen Studien wurden insgesamt sechs Varianten verglichen, wobei beide Domänen drei charakteristische Übergänge zeigten: Das erste Segment (Positionen 1–88 und 202–256) zeigte keinen Übergang (d.h. ohne Abstandsänderung), während das zweite Segment (Positionen 34–135 und 256–290) einen herkömmlichen Zweizustandsübergang durchlief. Interessanterweise zeigte sich bei der Markierung des vollständigen Rossmann-Motivs (Positionen 1–135 und 202–290) ein kompakter Zwischenzustand während des GuHCl-induzierten Entfaltens. Das detaillierte Verständnis de Übergänge in der N-terminalen Domäne war entscheidend für vergleichende Analysen mit Proteinen kürzerer Länge, die für ko-translationale Faltungsstudien synthetisiert wurden. Um den ribosomalen Faltungsprozess von yPGK zu untersuchen, habe ich Kryo-Elektronenmikroskopie (Kryo-EM) verwendet. Es wurden vier verschiedene Strukturen von Ribosom-Nascent-Chain-Komplexen (RNC) bestimmt, die jeweils ein synthetisiertes Polypeptid unterschiedlicher Länge repräsentieren. Die meisten Strukturen zeigten eine Dichte des synthetisierten Kettenmaterials außerhalb des ribosomalen Tunnels, und wir haben zum ersten Mal die Struktur einer Voll-Länge-Protein-Nascent-Kette,

die an das Ribosom gebunden ist, bei mittlerer Auflösung beobachtet. Zusammenfassend trägt diese Forschung dazu bei, unser Wissen über die Faltungsübergänge von yPGK zu erweitern und bietet neue Wege zur Untersuchung ko-translationaler Faltungsprozesse innerhalb von RNC-Komplexen mithilfe von sm-FRET und Kryo-EM.

# Contents

# List of Figures

# List of Tables

# List of Publications

Part of the work produced in the thesis resulted in the following publications:

[I] Yukhnovets O, Höfig H, Bustorff N, Katranidis A, Fitter J. Impact of Molecule Concentration, Diffusion Rates and Surface Passivation on Single-Molecule Fluorescence Studies in Solution. Biomolecules. 2022 Mar 18;12(3):468. doi: 10.3390/biom12030468. PMID: 35327660; PMCID: PMC8946791.

[II] Bustorff N, Fitter J. Features of Protein Unfolding Transitions and Their Relation to Domain Topology Probed by Single-Molecule FRET. Biomolecules. 2023; 13(9):1280. https://doi.org/10.3390/biom13091280.

# List of Acronyms

**2GS** Long unstructured linker

**ADP** Adenosine diphosphate

**ALEX** Alternating-laser excitation

**APTES** (3-Aminopropyl)triethoxysilane

**ATP** Adenosine triphosphate

**AV** Accessible volume

**AzF** Azido-phenylalanine

**BSA** Bovine serum albumin

**CpK** Cyclopropene-lysine

**cryo-TEM** Transmission electron microscopy at cryogenic temperatures

**CV** Column volumes

**DED** Direct-electron detectors

**DTT** Dithiothreitol

**FCS** fluorescence correlation spectroscopy

**FRET** Föster resosnance energy transfer

**GuHCl** Guanidine hydrochloride

**HPG** L-Homopropargylglycine

**IPTG** Isopropyl ß-D-1-thiogalactopyranoside

**LB** Luria broth

**NMR** Nuclear magnetic resonance

**Ni-NTA** Nickel-Nitriloacetic acid

| | |
|---|---|
| **O.N** | Overnight |
| **PEG-NHS** | Phosphoethanolamine-N-hydroxysuccinimide |
| **yPGK** | Yeast phosphoglycerate kinase |
| **PIE** | Pulse interleaved excitation |
| **POI** | Protein of interest |
| **PTC** | Peptidyl-transferase center |
| **PURE** | Protein synthesis using recombinant elements |
| **RAP** | Ribosome arrest peptides |
| **RNC** | Ribosome nascent chain complex |
| **RF1** | Release factor 1 |
| **RNAP** | T7 RNA polymerase |
| **RL** | Structured linker |
| **SD** | Shine-Dalgarno |
| **smFRET** | Single-molecule Förster resonance energy transfer |
| **SOC** | Super optimal broth with catabolite repression medium |
| **SPAD** | Single photon avalanche photodiode |
| **TIRF** | Total internal reflection |
| **TB** | Terrific Broth medium |
| **UAA** | Unnatural amino acids |

# Chapter 1

# Introduction

## 1.1 Protein folding and structure formation

Proteins are essential for the proper functioning of the cell. This class of macromolecules is responsible for structural[1], metabolic[2], signalling[3], regulatory[4], and defensive[5] functions. Therefore, the scientific community constantly seeks more information and aims to develop tools to understand proteins in more detail. Proteins are composed of amino acids connected by amide bonds formed between the carboxyl group of the first amino acid and the amine group of the following one[6]. The functions of proteins are intricately linked to their final three-dimensional structure. The formation of the 3D geometry relies on the initial string of amino acids maturing and rearranging itself in three-dimensional space to form a more compact native structure. Understanding how proteins fold is crucial since the misfolding of proteins lies at the center of many diseases and cellular malfunctions[7].

Christian Anfinsen was the first to propose that the amino acid sequence of a protein contains all the information required for its correct folding[8]. In general, the native structure of proteins is associated with the minimization of the thermodynamic free energy. During the folding process, the proteins starts from a broad unfavorable high-energy state and follows a stochastic search towards more favorable intermediate states with lower energy, ultimately reaching the global minimum energy of the native state. In general, the number of possible configurations is reduced from a large number of conformational arrangements ($10^{16}$–$10^{19}$, depending on the size of the protein) to a single final native configuration[9]. The folding pathway can be modulated and resembles an energetic funnel with local minimum barriers that proteins need to overcome to reach the global minimum[10]. Ultimately, the crucial point is understanding how the energy landscape is shaped based on the specific sequence of amino acids in question[11] (Figure 1.1).

After the conceptual frame had been laid out, several advances were achieved on the comprehension of the folding problem. The timescale at which protein structure formation happens was investigated by conducting computational simulations and biophysical experiments[12]. The gradual acquisition of a native-like structure promotes the formation of ordered protein conformations, including secondary structural elements like $\alpha$-helices and $\beta$-sheets, as well as tertiary structures like Rossmann fold or $\beta$-barrels[13,14]. Secondary structures are characterized by the formation of hydrogen bonds between the carbonyl and amide groups of the peptide main chain. As observed in Figure 1.1, the increase in amino acid contacts, number of hydrogen bonds, and other molecular interactions improve protein stability and decrease free energy. Evidence suggests that the increasing complexity of folding is closely associated with the contact order[15]. The farther apart the core residues are from each other in the peptide sequence, the more complex and time-consuming the folding pathway is[11].

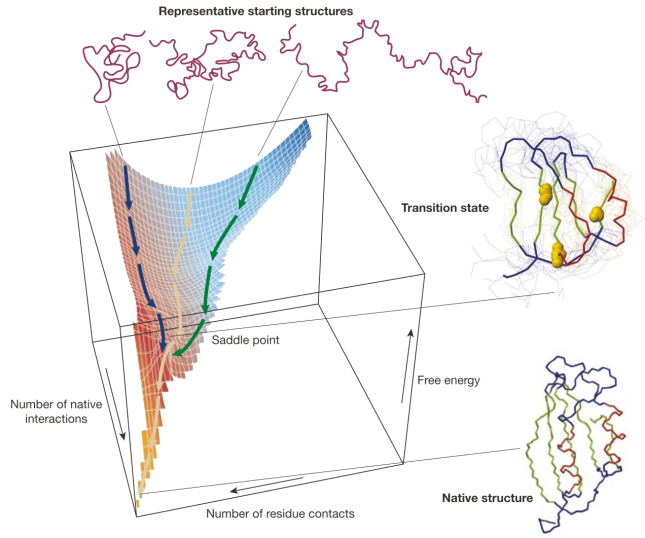



**Figure 1.1:** Protein folding energy landscape diagram. Simulation of a simplified model of small protein folding. The energetic surface funnel of possible pathways of the small unfolded protein is shown at the top, and its unique native structure is shown at the bottom. Saddle points represent the energetic barrier that must be crossed to reach the native fold. Energy is minimized by increasing native interactions of amino acids as well as by increasing the number of residues in contact. The exemplified protein efficiently overcomes the energetic barrier of the saddle point upon achievement of native interactions between the three residues represented in yellow in the transition state. Retrieved from [11].

Folding experiments of proteins larger than 100 residues revealed the existence of modular folding, meaning that persistent intermediates were found during the folding process[11]. The presence of long-lived intermediates supported the hypothesis that folding could take place independently within large segments or domains of the protein[9]. Conventional experiments to study folding focused on biophysical spectroscopic techniques using *in vitro* unfolding-refolding measurements. Chemical denaturants and temperature were used to unfold the proteins, and thermodynamic parameters were calculated[16–20]. Most of these experiments were performed as ensemble measurements, examining pools of $10^{14}$–$10^{17}$ molecules at one time[21]. However, in cells, proteins start folding as they are being synthesized by the ribosome[22]. Cells require a large machinery to efficiently synthesize the necessary proteins. The mechanism of translation has been elucidated in detail (Figure 1.2). Briefly, in bacteria, most mRNAs have an initiation sequence designated as Shine-Dalgarno (SD), which interacts with 16S rRNA of the 30S subunit[23]. To start translation, three initiation factors (IF1, IF2, and IF3) are involved[24] in coordinating translation initiation and ensuring the correct placement of all molecules, thus avoiding mistakes. After the initiation complex is formed with the binding of the 50S subunit and the presence of the fMet-tRNA$^{\mathrm{fMet}}$ in the P-site, decoding and elongation can start[23]. The next aminoacyl-tRNA-EF-Tu (aa-tRNA-EF-Tu) ternary complex is delivered to the ribosome, and the peptidyl-tRNA is formed at the peptidyl-transferase center (PTC). Meanwhile, the P-site tRNA is deacylated, and the nascent chain bound to the second tRNA translocates from the A-site to the P-site. This translocation allows the elongation process to continue, ensuring the correct placement of amino acids in the growing polypeptide chain[23]. The elongation is completed by removing the deacyl-tRNA from the E-site with the help of the EF-G[25]. This step clears the way for the next cycle of aminoacyl-tRNA binding and elongation.

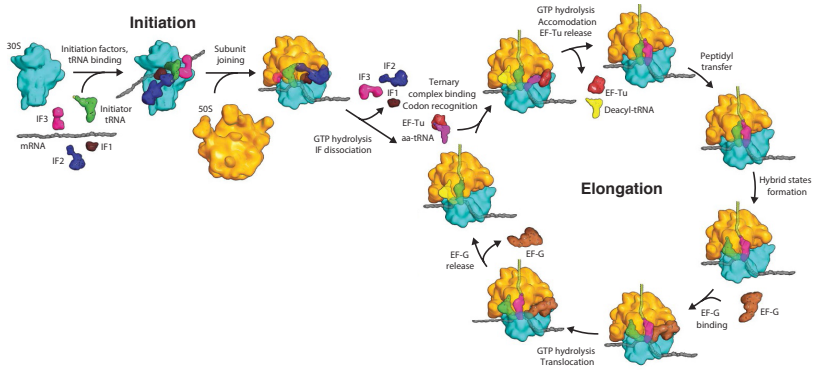



**Figure 1.2:** Key factors of bacterial translation. Simplified diagram of bacterial initiation and elongation. aa-tRNA – aminoacyl-tRNA; EF – elongation factor; IF – initiation factor. Adapted from [25].

As depicted in Figure 1.2, the protein biosynthesis process exhibits considerable complexity even in its simplified form. Cells generally have limited tolerance for unfolded and non-functional proteins, creating selective pressure for co-translational folding[26]. By folding proteins co-translationally, cells minimize the costs associated with protein quality control, because the prompt identification and elimination of misfolded and unassembled proteins is enabled[26].

## 1.2 Co-translational protein folding

The ribosome has been shown to play a significant role in guiding and facilitating the folding process during translation by providing spatial constraints and chaperone-like functions. The polypeptide is synthesized from the N-terminus to the C-terminus in a vectorial fashion, and evidence suggests that folding follows a sequential trend[27]. This results in a gradual structure formation while the protein emerges from the tunnel that traverses the 50S subunit[6,28]. The ribosomal tunnel has a length of approximately 100 Å and can accommodate 30–60 amino acids, depending on the structure of the nascent chain[29–32]. The tunnel can be divided into four distinct parts. It begins with the PTC, which serves as the entry point to the tunnel. Mov-

ing along, there are two regions; the upper tunnel and the lower tunnel, separated from each other by a constriction site. The final part of the tunnel is the vestibule, a cavity that acts as a transition zone, facilitating further processing and folding of the newly synthesized protein[33,34].

The major structural components of the tunnel are the 23S rRNA and a few ribosomal proteins. The constriction part is formed 30 Å away from the PTC by the uL4 and uL22 proteins[33]. Afterwards, the tunnel becomes gradually wider, reaching its widest point at the vestibule. The shape of the vestibule is determined by proteins uL23, uL24, and uL29[34,35]. The tunnel surface changes its electrostatic potential along its length[36], playing a role in the translation rate[37]. Depending on the peptide sequence, it promotes translational pauses due to the interaction of the amino acid's side chains with the tunnel surface[38]. Patches of positively charged amino acids can slow down translation[39,40]. Furthermore, the properties of the emerging nascent chain influence its propensity to form secondary structures such as $\alpha$-helices[32,41,42]. Co-translational folding already starts inside the ribosomal tunnel[43]. Therefore, there has been interest in better understanding the role of the tunnel in protein folding and how the geometry and its physicochemical properties impact the energetic landscape of proteins in the formation of different intermediates[37]. Many structural studies show the formation of $\alpha$-helices in different regions of the tunnel[44–47]. Additionally, force measurements revealed that even single domains are able to form inside the tunnel[48]. The interaction of the nascent chain with the ribosome surface is of importance for understanding co-translational folding, as it promotes a holdase activity that prevents premature or incorrect folding[34,49,50]. The effects are thought to be caused by an increase in the local concentration of the chemical groups of side chains exposed to the ribosomal surface[51].

The translation is relatively slow compared to most folding events[52,53]. In bacteria, the protein translation rate is about 20 amino acids per second, whereas the folding timescale spans from milliseconds to seconds[22,53]. The synthesis rate thus seems to be the limiting factor in the folding of the nascent chain emerging from the ribosome exit tunnel[35] (Figure 1.3). During *in vivo* synthesis, different cell mechanisms help the protein to form the correct structure. In crowded cellular environments, the

local concentration of the different components involved in translation is important and controls the translation rate[54,55]. Moreover, the regulation of unwinding of the mRNA secondary structure or post-transcriptional modifications also plays a role in the quality control of protein co-translational folding[55,56]. Ribosome interactions with the nascent chain and folding assistance by ribosome-associated chaperones also play an important role in the modulation of ribosome-associated nascent chain folding[57].

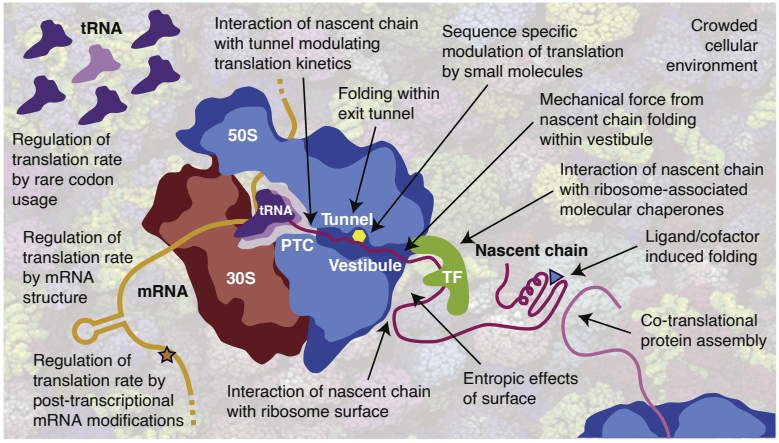



**Figure 1.3:** Possible cellular pathways that modulate protein folding co-translationally. The background shows crowding simulated from[58]. NC – Nascent chain; PTC – Peptidyl-transferase center; TF – Trigger factor. Retrieved from [59].

## 1.3 Study of multi-domain proteins

Many studies examining co-translational folding were performed using small, single-domain proteins[32,48]. However, more than 70% of proteins in cells have more than one domain. Generally, multi-domain proteins possess higher folding complexity and larger folding funnel landscape[60]. Most multi-domain proteins result from a combination of domains with distinct functions[61]. Domains can be organized by superfamilies in the genome, and their distribution forms clusters of highly abundant superfamilies[62,63]. The formation of domains with different functions is thought to

be a result of gene duplication and further sequence divergence or gene recombination[61]. The combination of more than one domain results in a series of domains, which define the multi-domain protein architecture. The structure of multi-domain proteins can be organized in continuous and discontinuous domains. In the first case, the whole domain is composed of neighboring residues, whereas discontinuous protein domains have stretches of the domain or neighboring domains separated far away in the primary sequence[64].

The three-dimensional geometry of multi-domain proteins seems to be conserved from an assembly perspective, meaning that the arrangement order of the different domains is important for the global multi-domain protein structure[63]. As described in section 1.1, amino acid contacts are important for energy minimization, and even more so across domain interfaces in multi-domain proteins. Two prominent theories propose that domain organization of multi-domain proteins resulted either from an evolutionary pathway of a previous recombination event or from a functional requirement for the correct protein activity[65,66]. Establishing the validity of a specific theory necessitates a thorough investigation of whether adjacent domains demonstrate mutually reinforcing influences. Moreover, from the folding viewpoint, it is known whether a particular domain has the same folding pathway in a single or multi-domain context[67], specifically if significant changes are observed in the transition state and appearance of distinct intermediates of the domain in each context[68].

The folding process of multi-domain proteins is complex due to the presence of multiple structural domains. Strong evidence shows that co-translational folding plays a vital role in ensuring the correct folding of these proteins[69]. During translation, as the protein chain is being synthesized, co-translational folding helps prevent the formation of incorrect or non-native structures. This is important because such non-native structures can result in misfolded proteins and subsequent aggregation[68].

Two-domain proteins can be used as a simplified model to study the complex folding pathways of multi-domain proteins. One of the most studied proteins, yeast phosphoglycerate kinase (yPGK), is a monomeric protein with a molecular weight of around 45 kDa, classified as a metabolic enzyme involved in the production of adenosine triphosphate (ATP) during glycolysis[2]. In fact, yPGK is responsible for

the catalysis of the conversion of 1,3-biphosphoglycerate and adenosine diphosphate (ADP) to 3-phosphoglycerate and ATP, respectively[70]. The structure of yPGK was determined by X-ray crystallography in 1974[70]. The structure revealed the catalytic reaction mechanism of the enzyme by unveiling the substrate binding sites[71]. yPGK is composed of two similar domains separated by a conserved $\alpha$-helix hinge[70,72] (Figure 1.4). The N-terminal domain is composed of the first 185 amino acids, and the C-terminal domain contains the last 230 amino acids. Both domains can fold independently and form a Rossmann fold topology[70,72–76]. Interestingly, in the full-length protein, the structure of the N-terminal domain is completed by the C-terminal end of the polypeptide chain[76,77]. Thus, it would be important to study other multi-domain proteins to understand how the presence of additional domains impacts the folding of the full-length protein. As an example, studies with the elongation factor G (EF-G), a four-domain enzyme involved in translation during protein synthesis, demonstrated that for the domain II to stably fold, it was necessary to synthesize the full-length protein[78].

Folding studies using yPGK were done using conventional biophysical methods and techniques. yPGK folding states were observed using circular dichroism[75], tryptophan fluorescence[74], NMR[79], and Förster resonance energy transfer (FRET)[80].

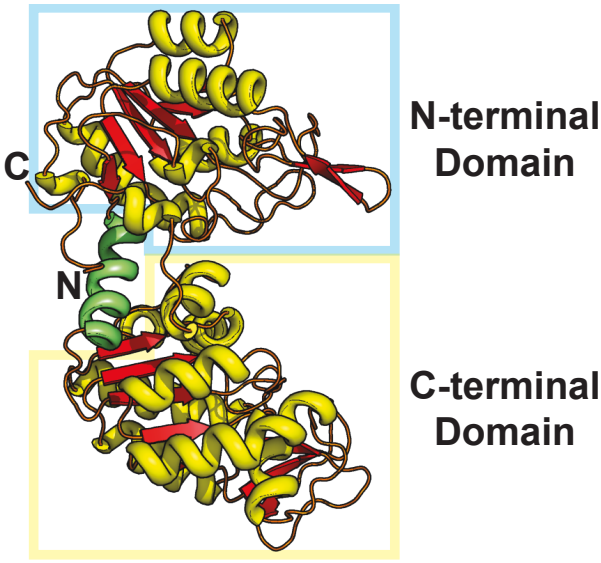



**Figure 1.4:** 3D structure of yeast PGK. PDB model 1qpg displayed in cartoon representation and colored with PyMOL[81]. Individual domains were highlighted with light blue (N-terminal) and yellow (C-terminal). Unstructured regions are colored in orange, Beta-strands are colored in red, and $\alpha$-helices are colored in yellow. Both domains show a Rossmann fold and are separated by a hinged helix colored in green.

## 1.4 Strategy to study co-translational protein folding

Studying *in vivo* co-translational folding poses significant challenges[82]. The cellular environment, characterized by its crowded and primarily asynchronous nature, presents obstacles in directly observing the targeted molecules[82]. First, the presence of cellular membranes acts as a physical barrier, preventing the access of external labeling agents such as dyes. Second, the presence of many different molecules that attenuate the signal of interest close to the background noise makes signal detection hard. Furthermore, the measurement and assessment of proper folding are complicated by the presence of polysomes and chaperones. Moreover, it is also necessary to exclude non-specifically bound molecules to the ribosome to ensure accurate analysis[82]. Therefore, achieving selectivity and specificity in detecting co-translational folding *in vivo* is not trivial. Alternatively, the development of cell-free synthesis

systems allows for true *in situ* measurements[83].

Cell-free expression systems emerged in the 1970s as a groundbreaking development, enabling the synthesis of proteins in test tubes[84]. These systems contained all the components necessary for protein synthesis. This was achieved by isolating the cellular machinery responsible for protein synthesis. Synthesizing a protein in a test tube requires the addition of the template mRNA to the cell-free system to initiate the reaction, which gives precise control of protein production. Additionally, cell-free systems offers the flexibility to fine-tune key components of the synthesis reaction, such as translation factors and chaperones. Notably, the use of cell-free systems was instrumental in introducing non-natural amino acids, which played a critical role in overcoming limitations related to labeling selectivity and specificity of ribosome nascent chain complexes.[82].

Cell-free expression systems can be classified into two main groups: lysate-based and minimal-based systems. Lysate-based systems were developed from cell lysates and are also referred to as extract systems. In contrast, protein synthesis using recombinant elements (PURE) was created as a minimal system[85]. In PURE systems, each protein component of the *E. coli* translation apparatus was purified individually through a histidine tag, allowing a higher control of the reaction mixture.

Additionally, coupled transcription/translation systems were developed with the purpose of protecting the liable mRNA molecules. These systems utilize DNA as a template and employ T7 RNA polymerase (RNAP) to generate coupled systems, allowing for the simultaneous incorporation of transcription and translation[84,86]. The DNA template can be supplied in three different forms: circular or linear plasmid and as a linearized PCR product. The biggest advantage of the PURExpress systems over lysates is the stability of the system[85]. Since every component can be added individually, it is possible to avoid the presence of nucleases and proteases that digest both template and product. Moreover, the system also gives total control over the components used in the system. It is possible to supply engineered components and remove endogenous components. With such versatility, a cell-free system represents an ideal platform for the expression of ribosome nascent chain complexes[87].

The first approach to obtain ribosome nascent chain complexes was to use mRNA

lacking a stop codon. This change in the RNA molecule only affects translation termination. During protein synthesis, the ribosome is able to translate the nascent chain normally. However, when it reaches the end of the mRNA, it stays bound to the nascent chain, which remains attached covalently to the tRNA[26]. Additionally, with the discovery of gene expression regulators called ribosome arrest peptides (RAPs), more tools were developed to generate RNCs. RAPs arrest translation activity by interacting with internal components of the ribosome and pause translation. The arrest is mediated by amino acid sequence independent of codon usage[88]. The activity of the arrest sequence is centered in the interior of the ribosome, with most arresting sequences being approximately 20 residues in length. Most of the key amino acids are spaced accordingly to perform their stalling activity. Multiple arrest points are observed in nature. One of the most targeted *loci* is the PTC, with several arrest peptides blocking it. The peptides targeting PTC prevent the peptide bond formation between peptidyl-tRNA and aminoacyl-tRNA, keeping the nascent chain on the P-site[89,90].

SecM is part of the PTC arrest peptides sequence group and belongs to the intrinsic peptide class that naturally holds translation. In nature, the sequence of this class of peptides needs to be long enough to be able to interact with effector molecules that enable the re-initiation of translation[88]. SecM is part of the secretory system of bacteria. It regulates the expression of the SecA ATPase enzyme that activates the SecYEG translocon[91]. Only when SecM is arrested can the ribosome express SecA due to the disruption of mRNA secondary structure and exposure of its SD sequence[92]. The normal stalling efficiency of SecM sequences is around 20%[93]. This stalling efficiency is relatively low for assays requiring the isolation of RNCs. Fortunately, an enhanced stalling sequence of SecM (SecMstr) was developed, increasing the stalling efficiency to 90%[93].

Another important requisite for studying RNCs is the ability to label the nascent chain selectively. Therefore, it is necessary to incorporate non-natural amino acids with unique chemically reactive side chains to avoid cross-reactivity with non-desired molecules. Most unnatural amino acids (UAAs) can react orthogonally with fluorescent dyes using click-chemistry[94]. Incorporating UUAs into the nascent chains can

be achieved by supplementing cell-free synthesis systems with engineered suppressor tRNAs designed to recognize a specific stop codon[95]. The presence of the suppressor tRNA enables the specific incorporation of UAAs into the nascent chain, allowing for selective labeling.

## 1.5 Methods used to study nascent chain folding events on the ribosome

The first approaches used to study co-translational folding relied on biochemical assays, such as limited proteolysis assays to probe the folding state, crosslinking, following enzymatic activity acquisition, or structural epitope interaction of the nascent chain with antibodies[96]. The combination of these techniques is a powerful tool to elucidate information about the nascent chain folding events. Ribosome profiling demonstrated that in mammalian cells, the domain structure is formed briefly after the whole sequence emerges outside the tunnel[97]. The first structural studies of RNCs were obtained using solution-state nuclear magnetic resonance (NMR). Again, these studies are challenging since the sample needs to be selectively labeled with $^{13}C/^{15}N$[98]. The structures obtained support the appearance of structural elements as the protein is emerging from the exit tunnel[27,98]. Other techniques, such as transmission electron microscopy at cryogenic temperatures (cryo-EM), were limited to low-resolutions above 7 Å[99]. However, the introduction of direct-electron detectors (DED) revolutionized cryo-EM for high-resolution structure determination. Seminal studies conducted by Beckmann's and Wilson's groups solved structures of the nascent chain sequences inside the exit tunnel. These structures had Coulomb potential density of $\alpha$-helical structures along the ribosomal tunnel[100–102]. Both NMR and cryo-EM techniques provide important information on the atomic positions of ribosome nascent chain complexes. One major challenge remains in dealing with the heterogeneity of RNCs samples due to the intrinsic dynamic nature of translation. Recent advances in 3D-reconstruction processing algorithms have the potential to help deal with some of the challenges arising from the intrinsic heterogeneity of

RNCs[103].

Finally, fluorescence methods can add further information on the structure and dynamics of RNCs by elucidating transition rates of unfolding/folding states and sensing local mobility and motion of the protein complexes[104–106]. It is possible to perform time-resolved anisotropy experiments solving nanosecond-scale motion events, elucidating spatial restrictions of the exit-tunnel[104]. Steady-state measurements rely mostly on FRET to elucidate the distance between two points labeled with a donor and acceptor[83]. Similar to structural studies, the analysis of fluorescence measurements is not trivial due to sample heterogeneity, especially in co-translational protein folding experiments. One way to overcome the challenges posed by co-translational protein folding studies can be addressed by single-molecule measurements. This type of experiment offers a solution to deal with processes that are difficult to observe due to their asynchronous or stochastic origin and, therefore, remain hidden in ensemble measurements.

### 1.5.1 Single-molecule fluorescence

Fluorescence is particularly appropriate for single-molecule measurements since it offers high sensitivity and selectivity. Achieving the desired conditions for single-molecule level measurements necessitates the control of sample concentration, achieved either through substantial dilutions or by reducing the observation window to femtoliter (fL) volumes. Fluorescence microscopy setups allow the efficient collection and detection of signals. The intensity of excitation light is normally many orders of magnitude higher than the emission light signal. In this respect, fluorescence microscopes offer the possibility to separate excitation light from emission light signals due to the optical properties of their components. According to Jablonski's diagram[107], after the excitation of a fluorophore, the excited electrons lose energy in vibrational transitions, and the emitted photons exhibit lower energy levels with longer wavelengths. Therefore, using proper optical elements such as dichroic mirrors, one can block the excitation light wavelength and let the emission photons through to the detectors. Moreover, the use of high numerical aperture

objectives increases the sensitivity of the fluorescence detection by gathering more light. The fluorescence events can then be detected with high detection efficiency by single photon avalanche photodiode (SPAD) detectors with low background signal. Typically, there are two conventional ways of measuring at a single-molecule level. One can use a confocal or widefield setup. The major difference resides in the way the light is focused on the sample. The confocal setup focuses the light in a diffraction-limited spot. Confocal measurements can capture freely-diffusing molecules in solution. On the other hand, a widefield setup can illuminate a larger region of interest and can form an image through the optics of the microscope.

Confocal setups have increased contrast and resolution as a consequence of reducing the out-of-focus light collected by the detectors. This reduction is achieved by the introduction of a pinhole in the system at the diffraction focus plane. The smaller the size of the pinhole is, the larger the amount of out-of-focus light is removed. In single-molecule measurements, confocal setups use pinhole sizes ranging from 25 µm up to 100 µm, creating detection volumes on the order of a few femtoliters. Confocal experiments with free-diffusing molecules detect particles on the order of a few milliseconds (ms), which corresponds to the dwell time of the fluorescent molecules traveling through the confocal volume. Measurements with proteins are routinely performed with surface-treated high-precision glass to reduce the adsorption of labeled molecules to the surface. The treatment of the surface confers additional reliability to the measurements and avoids surface interactions induced artifacts.

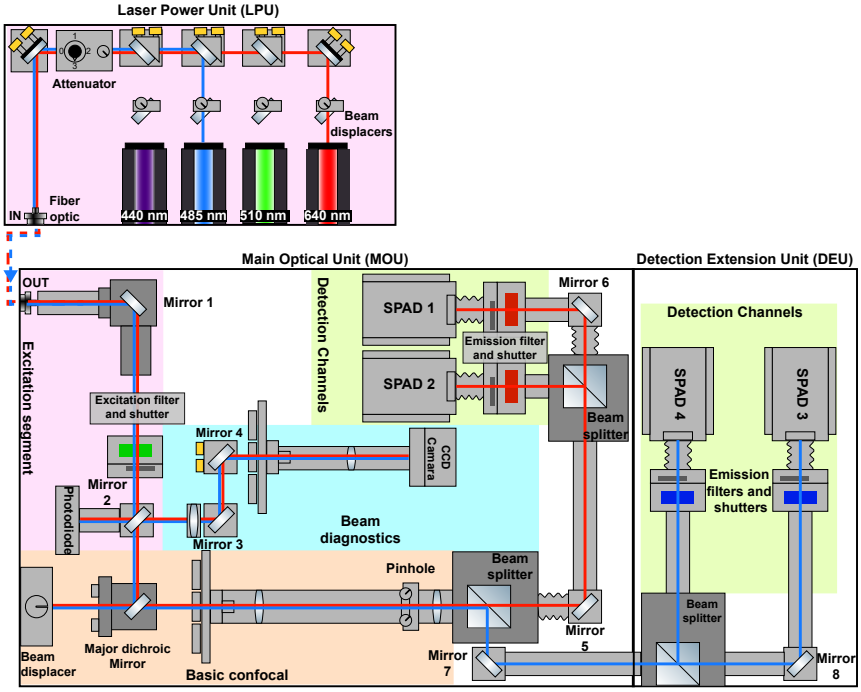



**Figure 1.5:** Schematic diagram of microtime200 from PicoQuant. Designed with the help of ComponentLibrary, created by Alexander Franzen.

The confocal microscope setup in the lab is characterized by two central units (Figure 1.5). Excitation pulses are emitted by the laser power unit (LPU) and transmitted to the main unit (MOU and DEU) through an optical fiber. Afterwards, the laser light is reflected by the major dichroic mirror to reach the objective. Then, emitted photons are collected by the SPAD detectors after passing through the pinhole. In two-color fluorescence measurements, the signal is divided into donor and acceptor channels using the beam splitter next to mirror 5 in Figure 1.5[108].

In the current work, two fluorescence techniques were employed. First, Fluorescence correlation spectroscopy (FCS) supported single-molecule fluorescence measurements, helping with the accurate determination of analytical properties of the sample, such as concentration ($C$) and hydrodynamic radius ($R_H$), and characterizing the confocal volume ($V_{eff}$). FCS makes use of the intensity fluctuation of the

signal generated from biomolecules going in and out of the detection volume[109]. Briefly, in a confocal setup, the auto-correlation function ($G(t)$) can be described as the similarity of the fluorescence intensity trajectories at a certain time ($I(t)$) with the intensity at various lag times ($I(t+\tau)$)[109] (Figure 1.6 A and B, see Equation 1.1).

$$G(\tau) = \frac{\langle I(t)I(t+\tau)\rangle}{\langle I(t)\rangle^2} - 1 \tag{1.1}$$

The shape of the observation volume can be represented by a 3D Gaussian profile, which is determined by fitting the auto-correlation curve with the following model:

$$G(\tau) = \frac{1}{N}\left(1+\frac{\tau}{\tau_D}\right)^{-1}\left(1+\frac{\tau}{\tau_D}\cdot\frac{r_0^2}{z_0^2}\right)^{-\frac{1}{2}} \tag{1.2}$$

where the average molecule number present at the same time in the confocal volume is given by ($N$), and ($\tau_D$) describes the average time that the molecules take to diffuse through the detection volume. The lateral and radial-axial dimensions of the detection volume are obtained from ($r_0$) and ($z_0$), respectively.

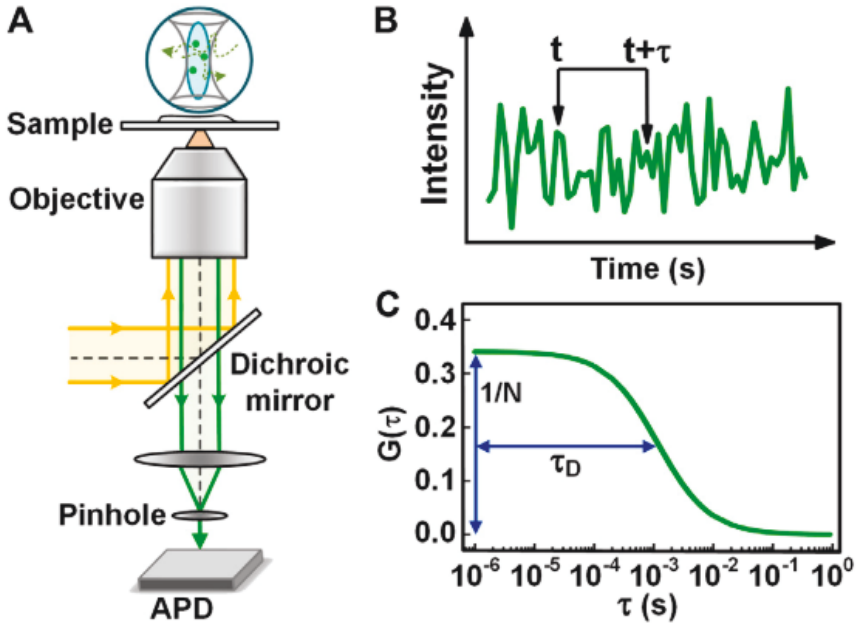



**Figure 1.6:** FCS data analysis. A – Schematics of the optical path of confocal-based setup. B – Intensity time trace observed in confocal measurement. C – Auto-correlation curve calculated from intensity time trace. The number of molecules in the confocal volume is inversely proportional to the amplitude of the auto-correlation curve at $\tau = 0$. Abbreviations: APD – Avalanche photodiode detector Retrieved from [109].

Figure 1.6 C represents the fit obtained from Equation 1.2 to the auto-correlation curve. The amplitude of the curve at $(G(0))$ is inversely proportional to the number of fluorescent molecules $(N)$. Moreover, as the size of the particle increases, the slower the particle diffuses in solution, and the larger its diffusion time $(\tau_D)$. Afterwards, the diffusion coefficient $(D)$ can be determined from the width of the correlation curve after the calibration of the system to determine $(r_0)$ (Equation 1.3).

$$D = \frac{r_0^2}{4\tau_D} \tag{1.3}$$

Furthermore, it is possible to obtain the hydrodynamic radius of the particles from

the Stokes-Einstein relationship[109] (Equation 1.4).

$$R_H = \frac{k_B T}{6\pi\eta D} \tag{1.4}$$

where $(D)$ is the diffusion coefficient, $(k_B)$ is the Boltzmann constant, $(T)$ is the absolute temperature in Kelvin, $(\eta)$ is the dynamic viscosity of the solvent in which the particles are dissolved, and $(R_H)$ is the hydrodynamic radius of the particles detected in the confocal volume.

Benefiting from FCS entails maintaining the concentration of fluorescent molecules within the range of 100 pM to 10 nM. When concentrations surpass the upper limit, the significance of correlated photons originating from individual molecules diminishes in comparison to the significance of uncorrelated photons emanating from distinct molecules. Conversely, falling below the lower limit results in an inadequate signal-to-noise ratio[110].

Second, FRET measurements were performed at the single-molecule level. Analyzing populations with different interfluorophore distances within the same sample can give information about the presence of different conformational states and how often they are visited[111]. Theodor Förster described FRET in detail as the energy transfer between two fluorophores (donor and acceptor)[112]. The energy transfer efficiency depends on the overlap between the emission spectrum of the donor and the absorption spectrum of the acceptor[112]. More importantly, the energy transfer $(E)$ is dependent on the distance $(R)$ between the two fluorophores and Förster radius $(R_0)$ according to equation 1.5.

$$E = \frac{1}{1 + \left(\frac{R}{R_0}\right)^6} \tag{1.5}$$

Briefly, the Förster radius is a characteristic distance at a FRET efficiency of 50%, which is unique for individual FRET pairs (Figure 1.7). $R_0$ can be determined by the following equation:

$$R_0 = \left[8.79 \times 10^{-5} (\kappa^2 n^{-4} \phi_D J(\lambda))\right]^{1/6} \text{ Å} \tag{1.6}$$

Here $(\kappa^2)$ refers to the orientation factor between the two fluorophores dipole moments, which was taken as 2/3 assuming freely rotating fluorophores. $(n)$ represents the refractive index of the medium, $(\phi_D)$ is the quantum yield of the donor, and $(J(\lambda))$ is the spectral overlap integral between donor fluorescence and the absorbance of the acceptor.

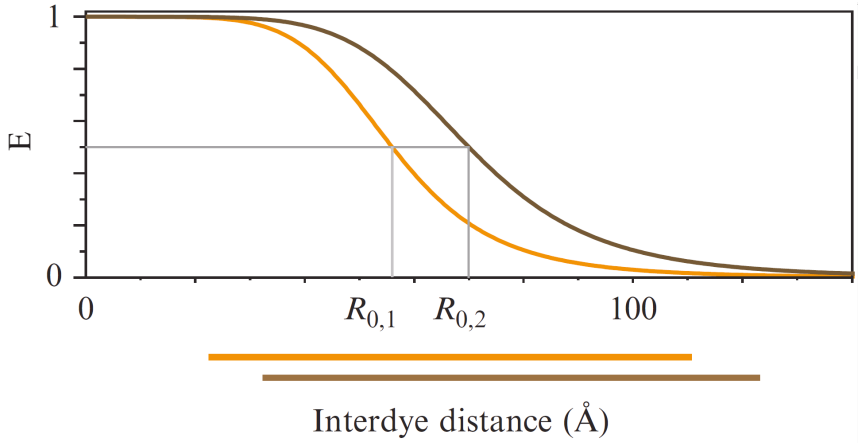



**Figure 1.7:** Plot of FRET efficiency $(E)$ against the interdye distance for two different $R_0$ values. $R_{0,1} = 56$ Å (orange), and $R_{0,2} = 70$ Å (brown). Adapted from [113].

When calculating FRET efficiencies in single-molecule measurements, each photon emitted by excited molecules is captured by the corresponding Single-Photon Avalanche Diode (SPAD) and categorized as a fluorescence burst[108,114]. Therefore, FRET efficiency can be expressed as the ratio between photons detected in acceptor channels and all photons detected[114] (Equation 1.7).

$$E = \frac{F_D^A}{F_D^D + F_D^A} \tag{1.7}$$

where the FRET efficiency is given by $(E)$, the number of photons of acceptor emission after donor excitation is represented by $(F_D^A)$, and the number of photons of

donor emission after donor excitation is given by ($F_D^D$).

Further details about burst acceptance and analysis can be found in the references[108,111,114]. An example of burst selection can be found in Figure 1.8. Briefly, the burst selection is based on the interphoton time distance between two consecutive photons. Background photons have larger interphoton time distances than photons generated from fluorescent molecules. Signal dips with enough number of photons shown in Figure 1.8 can be defined as bursts[114].

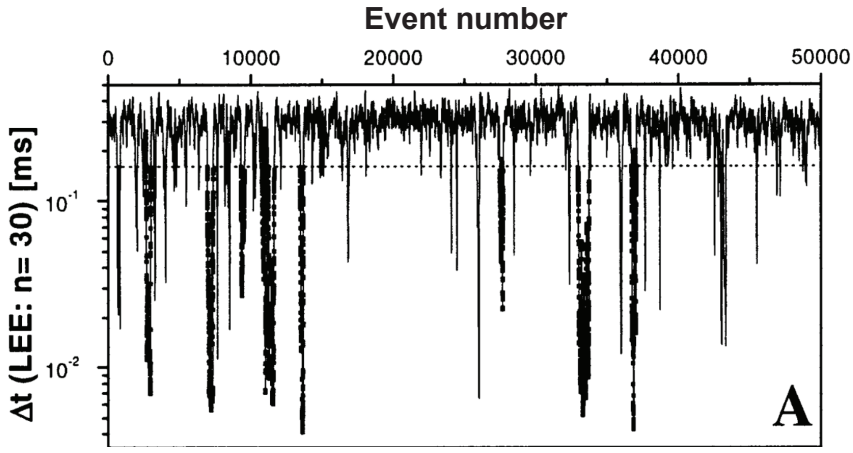



**Figure 1.8:** sm-FRET time-dependent burst analysis. Interphoton time ($\Delta t$) between two consecutive photons smoothed with a Lee filter. The selection threshold is shown as a dotted line for $\Delta t$ = 160 µs. Selected photons are represented with big black squares. Adapted from [114].

Individual bursts contain different numbers of photons according to the traveling trajectory of labeled particles and labeling characteristics, e.g., molecules with two donors or two acceptors. Alternating-laser excitation (ALEX) schemes such as ns-ALEX and Pulse interleaved excitation (PIE) facilitate the sorting of the different bursts according to their properties[108,111]. Moreover, emissions can additionally be presented as a distance-independent ratio, called stoichiometry ($S$)[115]. $S$ is closely related to the label stoichiometry and sensitive to the relative brightness of fluorophores in single molecules, enabling the monitoring of oligomerization and local environment changes, respectively[115]. The definition of ($S$) can be found in Equation

1.8.

$$S = \frac{F_D^A + F_D^D}{F_D^D + F_D^A + F_A^A} \tag{1.8}$$

where the stoichiometry ratio is given by ($S$), the number of photons after donor excitation is represented by ($F_D^A$ and $F_D^D$), and the number of photons after acceptor excitation is given by ($F_A^A$).

Considering Equation 1.7 and Equation 1.8, fluorescent samples measured in sm-FRET can be sorted in 2D plots as shown in Figure 1.9.

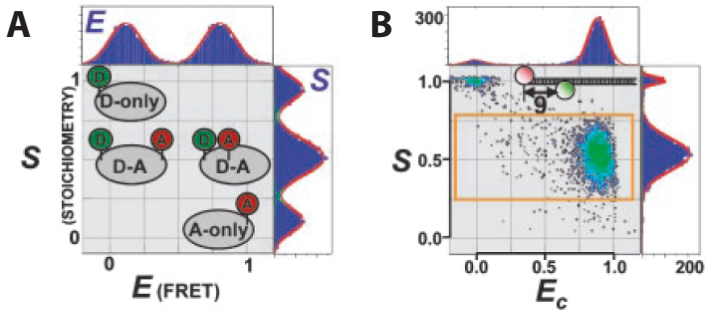



**Figure 1.9:** Sorting of single-molecule burst with 2D plots. A – $E$-$S$ histogram for donor-only, acceptor-only, and donor-acceptor species with different dye distances. $E$ or $E_c$ sorts species according to FRET and fluorophores distance; $S$ sorts species according to dye stoichiometry (Donor-Acceptor). B – High FRET DNA. (1D histograms of $E$ and $s$ in blue; Sum of Gaussian fits in red) Adapted from [115].

Table 1.1 summarizes the distribution of fluorescent species observed in sm-FRET measurements. Finally, to quantitatively determine the distance between the two fluorophores, it is necessary to take into consideration the background, spectral crosstalk, direct excitation, and gamma correction factors[108].

**Table 1.1:** Fluorescent species in single-molecule FRET measurements.

| Species | Stoichiometry | Efficiency |
|---|---|---|
| Donor-only | 1 | 0 |
| Acceptor-only | 0 | 1 |
| FRET | 0.5 | 0–1 |

### 1.5.2 Cryo-EM

Structural studies give important insights at the atomic level of biological processes. One of the techniques with increasing utility in the last two decades was transmission electron microscopy (TEM) of biological samples. Ernst Ruska invented the first transmission electron microscope at the beginning of the 20$^{\text{th}}$ century. Like light microscopes, TEM aims to focus the illumination source on the sample to create a magnified image with a high level of detail. In the electron microscope, the electrons generated from the high current passing through the electron source (e.g., tungsten filament, field emission gun (FEG)) are accelerated through the microscope column, pass through the sample, and collected at the detection point[116]. The image of the sample is then generated from elastic scattering events of the electrons that interact with the sample. Biological samples are composed mainly of light atoms such as carbon, nitrogen, oxygen, and hydrogen. Therefore, scattering events of electrons have low contrast and are hard to image. Nevertheless, contrary to light microscopes where the light wavelength limits the resolution, TEM accelerates electrons to such a speed that their wavelength is between 2 and 4 pm, which is much lower than the resolution of the molecules obtained from TEM in practice[117].

Initially, viral particle specimens were imaged at room temperature using staining solutions with heavy atoms to generate higher contrast[118]. The sample was adsorbed to the grid and embedded with a heavy metal stain (e.g., uranyl acetate, uranyl formate, ammonium molybdate) for a few seconds. Afterwards, the excess liquid was removed to form a thin layer on top of the sample. This technique is limited to

resolutions of ≈ 20 Å since the sample is dried during its preparation, and in the microscope, only the outline is imaged[118]. Other disadvantages of negative stains are the potential changes in the molecule shape induced by the stain and dehydration. Therefore, the presence of a heavy atom-shell leads to the collapse of large molecules such as viruses or liposomes.

Alternatively, developing TEM under cryogenic conditions (cryo-EM) addressed some of the previous problems. Imaging at liquid nitrogen temperatures helped to protect the sample against radiation damage, allowed the sample to survive the high vacuum of the microscope column, prevented sample evaporation, and enabled the use of stain-free samples. Debouchet[119] developed to that end a sample vitrification technique to plunge-freeze a small amount of sample very quickly into liquid ethane at temperatures below 110 Kelvin. With the rapid transition from room temperature to frozen states, the aqueous solution vitrifies into amorphous ice without the formation of crystalline ice, thus remaining electron-transparent.

During imaging, 2D projections of the biological specimen are generated in the electron detector. Since the sample is vitrified in solution, the 2D projections should correspond to random 3D orientations of the specimen, which are fundamental to obtaining a reliable 3D reconstruction. As mentioned above, biological samples have low contrast in TEM. Therefore, the sample must be imaged with large amounts of defocus, improving its contrast. Imaging the sample with a defocus introduces aberrations and distortions, mainly through contrast inversion. The aberrations and distortions in images can be described by the contrast transfer function (CTF), which allows the recovery of all pertinent information[120].

$$CTF(r) = -\sin\left(\frac{\pi}{2}C_S\lambda^3 r^4 + \pi f \lambda r^2\right) \tag{1.9}$$

where the information transfer at each spatial frequency ($r$) is affected by the spherical aberrations of the microscope generated by lens imperfections ($C_S$), depends on the electron wavelength ($\lambda$), and the defocus value applied ($f$). The CTF curve is characterized by an oscillating function where every time the curve crosses the horizontal axis, there is no information at that spatial frequency. In order to acquire

sample information across a wide range of frequencies, data is collected by varying defocus values, thereby shifting the oscillations and CTF cross-over points in each image[121]. Moreover, at the end of the 3D reconstruction, the attenuated high spatial frequencies need to be compensated by a B-factor to interpret the structure based on all available frequencies[122].

Another milestone to get high-resolution structures from cryo-EM was the development of fast direct electron detectors. These detectors greatly improved the detective quantum efficiency (DQE) and allowed the fast detection of single electron events[123]. The introduction of fast detectors also enabled the acquisition of movies instead of images, which allowed the correction of beam-induced motion. Overall, direct electron detectors helped push the field of cryo-EM further. Moreover, to reconstruct the sample to high resolution, it is necessary to collect many movies. Since the images are taken with a low dose of electrons, i.e. 20–40 $e^-/Å^2$, they are extremely noisy. Therefore, we need to average many particles to boost the protein signal. After image collection, applying the strategies described above to sample all possible spatial frequencies, the 2D projections need to be processed to reconstruct a 3D object. The reconstruction is based on the Fourier slice theorem, which postulates that a 2D projection of an object represents a single slice through the 3D Fourier transform of that object passing through its center[124]. The process of finding the correct Fourier slice that matches the right 2D projection is done iteratively by determining each projection's orientation and translation angles and refining them after each iteration. The Fourier shell correlation (FSC) determines the final reconstruction resolution. In order to estimate the resolution, the dataset is divided into two halves, with separate reconstructions performed on each half[125]. Fourier shells of each volume at each spatial frequency are compared by cross-correlation, and the final resolution is determined from the spatial frequency with cross-correlation below 0.143[125].

Finally, after determining the consensus map with the correct B-factor applied, it is possible to start building an atomic model of the protein investigated. In general terms, coordinates of reference structures or amino acids are fitted in 3D EM density[126]. Afterwards, refinement and validation of the atomic model follow. Ulti-

mately, based on the solved atomic model and together with prior knowledge available about the protein investigated, it is possible to interpret its biological functions and mechanisms.

## 1.6 Aims

The overarching goal of the thesis was to elucidate the process of folding closer to physiological conditions, meaning that I set to look at co-translational folding events by investigating nascent chains synthesized to different lengths of the multi-domain protein from *saccharomyces cerevisiae* phosphoglycerate kinase. In order to have a complete understanding of the folding events, I used classical unfolding/folding studies of the full-length protein to compare with the results of co-translational folding. To this end, I used single-molecule FRET to monitor distances and conformational changes in order to determine the dynamic interplay between translation, domain folding, and final protein conformation.

First, I expressed two new variants of yPGK targeting two extra distances in the N-terminal domain of the protein in addition to the labeling positions previously measured, three in the C-terminal, one in the N-terminal domain, and one inter-domain. The two new position pairs in the N-terminal domain were important for having more distance values when compared with the protein's shortest truncations.

Second, to go one step further and study co-translational folding, we used double-labeled truncated versions of the protein at strategic lengths that would allow us to draw conclusions on the folding intermediates compared to the full-length protein.

Third, I aimed to observe the same nascent chains attached to the ribosome by applying a unique labeling technique to label the nascent chains specifically.

Finally, I aimed to address co-translational folding globally by resolving structural details of the ribosome together with the nascent chain in the ribosomal exit tunnel using cryo-EM.

# Chapter 2

# Materials

## 2.1 Instrumentation

Absorption spectrometer

- NANODROP 2000c spectrophotometer (ThermoFisher Sientific Inc., Waltham, MA, USA)

Autoclave (Fedegari GmbH & Co KG, München, Germany)

Electron microscopes

- Krios G4 Cryo-TEM for Life Sciences (Thermo Scientific™ Waltham, MA, USA)
- Talos™ Arctica™ TEM for Life Sciences (Thermo Scientific™ Waltham, MA, USA)
- Talos™ L120C TEM for Life Sciences (Thermo Scientific™ Waltham, MA, USA)

Light microscope

- Microtime 200 (PicoQuant GmbH, Berlin, Germany)
  - Microscope: Olympus IX-81 (Olympus Germany, Hamburg, Germany)
  - Objective: UplanSApo, 60×, NA 1.2 (Olympus Germany, Hamburg, Germany)
  - Lasers: 485 nm and 640 nm
    LDH-D-C 485 and LDH-D-C 640 (PicoQuant GmbH, Berlin, Germany)

  - TCSPC-Module: HydraHarp 400 (PicoQuant GmbH, Berlin, Germany)
  - Two Avalanche Photodiode for each channel.
    485 nm: $\tau$-SPAD (PicoQuant GmbH, Berlin, Germany)
    640 nm: SPCM-CD3077-H and SPCM-AQR-14 (Perkin-Elmer Inc., Waltham, MA, USA)
  - Major dichroic: XF2401 (Omega Optical Inc., Brattleboro, VT, USA)
  - Minor dichroic: T600lpxr (Chroma Technology Corp., Bellows Falls, VT, USA)
  - 50/50 beam-splitter (Olympus Germany, Hamburg, Germany)
  - Emissions filter:
    485 nm: 530/55 Bandpass filter (Semrock Inc., Rochester, NY, USA)
    640 nm: 635 Longpass filter (Semrock Inc., Rochester, NY, USA)

Fluorometer

- Spectrofluorophotometer QM-7 (Photon Technology International, Birmingham, NJ, USA)

FPLC-Systems

- ÄKTAexplorer 10 (GE Healthcare, Chalfont St. Giles, UK)
- ÄKTApure (GE Healthcare, Chalfont St. Giles, UK)
- ÄKTAmicro (GE Healthcare, Chalfont St. Giles, UK)

Gel imaging

- ChemiDoc™ MP Imaging System (Biorad Laboratories Inc., Hercules, CA, USA)

Incubator

- Multitron Standard (Infors AG, Bottmingen, Schweiz)

Cell-disruptor

– Alpha 1-2 LD plus (Martin Christ Gefriertrocknungsanlagen GmbH, Osterode, Germany)

pH-meter

– S20-SevenEasy™ with InLab® Microelectrode (Mettler Toledo International Inc., Columbus, OH, USA)

Pipettes

– Eppendorf Research® plus: P1000, P200, P100, P20 and P10 (Eppendorf, Hamburg, Germany)

Plasma cleaner

– Zepto plasma cleaner (Plasma Surface Technology, Diener Electronic GmbH, Ebhausen, Germany)

Power supply units

– Power Pack P25 (Biometra, Göttingen, Germany)
– Power Pac HC (Biorad Laboratories Inc., Hercules, CA, USA)
– PowerPac™ Basic Power Supply (Biorad Laboratories Inc., Hercules, CA, USA)

Sterile cabinet

– UNIFLOW UVUB (Uniequip Laborgerätebau- und Vertriebs GmbH, Planegg, Germany)

PCR thermocycler

– FlexCycler$^2$ (Analytik Jena AG, Jena, Germany)

Centrifuges

– Optima XPN-80 ultracentrifuge / Rotor: Ti 70 with Polycarbonate tubes (26.3 mL); SW32Ti with Polypropylene tubes (38.5 mL) (Beckmann Coulter Inc., Brea, CA, USA)

- AvantiTM J-20 XP centrifuge / Rotor: JLA-8.1 with Polypropylene tubes (1000 mL) (Beckmann Coulter Inc., Brea, CA, USA)
- Optima MAX-XP / Rotor: TLA-110 with Polycarbonate tubes (4 mL) (Beckmann Coulter Inc., Brea, CA, USA)

Gradient station

- Gradient Station™ Base Unit (BioComp Instruments, Fredericton, Canada)
- Model 2110 Fraction Collector (Biorad Laboratories Inc., Hercules, CA, USA)
- Triax Flow Cell model FC-1 for 260 nm scans (FC Software included) (BioComp Instruments, Fredericton, Canada)

## 2.2 Consumables

Chromatography

- Sephadex G25 (Sigma Aldrich Corp., St. Louis, MO, USA)
- NiNTA-Agarose (Qiagen GmbH, Hilden, Germany)
- column glass (Different sizes, Biorad Laboratories Inc., Hercules, CA, USA)
- HiTrapTM Benzamidine FF, 2x 1ml (GE Healthcare, Chalfont St. Giles, UK)
- HighLoadTM Superdex 200 (16/60), präparativ (prep grade) (GE Healthcare, Chalfont St. Giles, UK)
- MonoQTM (5/50 GL) (GE Healthcare, Chalfont St. Giles, UK)

Cover slides

- High-precision cover slide No. 1.5H (Paul Marienfeld GmbH&Co.KG, Lauda-Königshofen, Germany)

Clear tubes

- EPPENDORF TUBES® Protein LowBind/DNA LowBind Tube, 1,5 ml (Safe-Lock Tubes), PCR clean (Eppendorf AG, Hamburg, Germany)

Concentrators

- Amicon® Ultra 4 / 15 Ultracel®, MWCO 10kDa/50kDa/100kDa (Merck Millipore Ldt., Tullagreen, Irland)

Kits

- QIAprep® Spin Miniprep Kit (Qiagen GmbH, Hilden, Germany)
- NucleoSpin Gel and PCR Clean-up, Mini kit (MACHEREY-NAGEL GmbH & Co. KG, Düren, Germany)
- In-Fusion® Snap Assembly Starter Bundle (TAKARA BIO INC., Shiga, Japan)
- PURExpress® (New England Biolabs, MA, USA)
- RTS 100 E. coli HY (Biotechrabbit GmbH, Berlin, Germany)

## 2.3 Chemicals

Sigma Aldrich Corp., St. Louis, MO, USA

- Ammonium sulfate
- Bis-(2-hydroxyethyl)-Amino-Tris-(hydroxymethyl)-Methane (Bis-Tris)
- Bromphenol blue
- Calcium chloride
- Dimethylsulfoxide (DMSO), anhydrous
- Dithiothreitol (DTT)
- Acetic acid (99 %)
- ß-Mercaptoethanol

- Sulfuric acid (95-98 %)
- Tris (2-carboxyethyl) phosphinhydrochloride (TCEP)
- Hydrogen peroxide (30 %)

AppliChem GmbH, Darmstadt, Germany

- Agar
- Coomassie Brilliant Blue G 250
- Guanidinium hydrochloride (GuHCl)
- Imidazole
- Potassium dihydrogen phosphate
- Potassium hydroxide
- 3-(N-morpholino)propanesulfonic acid (MOPS)
- Sodium chloride
- Sodium hydroxide
- Sodium dodecyl sulfate (SDS)
- Tris(hydroxymethyl)aminomethane (TRIS)
- Urea

Carl Roth GmbH & Co.KG, Karlsruhe, Germany

- Acrylamide/Bisacrylamide: Rotiphorese® Gel 30 (37.5:1)
- Roti®-Block (10x concentrate)
- Ethanol (96%)
- Ethanol (for HPLC, 99.9%)
- Glycerin
- Hydrochloric acid (32%)
- Tween 20

Merck KGaA, Darmstadt, Germany

- Uvasol Acetone (for spectroscopy, 99.9%)
- Uvasol Ethanol (for spectroscopy, 99.9%)
- Magnesium sulfate heptahydrate

Other chemicals

- Agarose: SeaKem® LE Agarose (Lonza Group AG, Basel, Switzerland)
- Coomassie Brilliant Blue R (SERVA Electrophoresis GmbH, Heidelberg, Germany)
- Ethylenediaminetetraacetic acid (EDTA) (MP Biomedicals LLC, Santa Ana, CA, USA)
- GelRed (Biotium Inc., Hayward, CA, USA)
- Glycine (MP Biomedicals LLC, Santa Ana, CA, USA)
- Isopropyl $\beta$-D-1-thiogalactopyranoside (IPTG) (VWR International LLC, Radnor, PA, USA)
- Methanol (for HPLC) (VWR International LLC, Radnor, PA, USA)
- Nitrogen (Linde AG, Munich, Germany)

## 2.4 Fluorophores

Alexa Fluor® 488 C5 Maleimid (Thermo Fisher Scientific Inc., Waltham, MA, USA)

Alexa Fluor® 647 C2 Maleimid (Thermo Fisher Scientific Inc., Waltham, MA, USA)

Alexa Fluor® 488 NHS (Thermo Fisher Scientific Inc., Waltham, MA, USA)

Alexa Fluor® 488 sDIBO Alkyne (Thermo Fisher Scientific Inc., Waltham, MA, USA)

Alexa Fluor® 647 sDIBO Alkyne (Thermo Fisher Scientific Inc., Waltham, MA, USA)

Alexa Fluor® 647 Azide, Triethylammonium Salt (Thermo Fisher Scientific Inc., Waltham, MA, USA)

ATTO 655 NHS (ATTO-TEC GmbH, Siegen, Germany)

ATTO 488 tetrazine (MeTet) (ATTO-TEC GmbH, Siegen, Germany)

ATTO 647N tetrazine (MeTet) (ATTO-TEC GmbH, Siegen, Germany)

## 2.5 Buffers

Buffers and other aqueous solutions were prepared using MilliQ water.

Agarose gel electrophoresis

- TAE:
  40 mM Tris/Acetic Acid, pH 8 1 mM EDTA

yPGK isolation and purification

- Lysis buffer: 30 mM MOPS, pH 7.5; 500 mM NaCl, 0.5 mM TCEP
- Wash buffer I: Lysis buffer + 2.5 mM imidazole
- Wash buffer II: Lysis buffer + 5 mM imidazole
- Elution buffer: Lysis buffer + 250 mM imidazole

Cell-free system

- Buffer W: 100 mM Tris-HCl pH 8.0, 150 mM NaCl
- Blocking buffer: Buffer W + 0.5% BSA
- Cell-free buffer: 10 mM Tris-Acetate pH 8.2, 14 mM $Mg(OAc)_2$, 0.6 mM KOAc
- Buffer BXT: 100 mM Tris-HCl pH 8.0, 150 mM NaCl, 50 mM biotin, 1 mM EDTA

Western-blot

- Blocking buffer: 20 mM Tris/HCl, pH 7.5; 150 mM NaCl; 0.1% Tween 20; 3% BSA
- Amersham™ ECL™ Prime Western Blotting Detection Reagent (enhanced chemiluminescence) (GE Healthcare, Chalfont St. Giles, UK)

SDS-PAGE

- 5x sample buffer:
  60 mM Tris/HCl, pH 6.8; 8 M Urea; 14.4 mM $\beta$-Mercaptoethanol; 2% (w/v) SDS; 25% (w/v) Glycerin; 0.1% (w/v) Bromophenol blue

RNC purification

- Buffer A: 50 mM HEPES pH 7.5, 25 mM $Mg(OAc)_2$, 10 mM KOAc

## 2.6 Antibiotics

Ampicillin Sodium Salt (AppliChem GmbH, Darmstadt, Germany)

Puromycin Dihydrochloride (Merck KGaA, Darmstadt, Germany)

Chloramphenicol (Merck KGaA, Darmstadt, Germany)

## 2.7 Inhibitors

Protease Inhibitor Cocktail (Tablets), cOmplete, EDTA-free (Roche Diagnostics GmbH, Mannheim, Germany)

## 2.8 Antibodies

Anti-PGK1 antibody [22C5D8] (ab113687) (Mouse) (abcam, Cambridge, UK)

Goat Anti-Mouse IgG H&L (HRP) (ab205719) (abcam, Cambridge, UK)

## 2.9 Markers

DNA Standard

- 1 Kb DNA Ladder (Thermo Fisher Scientific Inc., Waltham, MA, USA)

Protein Standards

- Page Ruler™ Prestained Protein Ladder (Thermo Fisher Scientific Inc., Waltham, MA, USA)
- Precision Plus Protein™ All Blue Prestained (Biorad Laboratories Inc., Hercules, CA, USA)

## 2.10 Software

Adobe illustrator CS6 (Adobe Inc., USA)

CCP-EM v1[127]

CCP4[128]

ChimeraX 1.5[129]

Chimera 1.14[130]

Coot 0.9[127]

CryoSPARC 4[131]

EPU (Thermo Fisher Scientific Inc., USA)

LaTeX (LaTeX Project Public License (LPPL)

LigPlot+[132]

LocScale[133]

Microsoft Office 365 (Microsoft Corp., USA)

Origin 2019 (OriginLab Corporation, USA)

PAM[134]

Phenix 1.20[135]

PyMOL (Version 2.3 Schrödinger, LLC., USA)

RELION 3 and 4[136]

SBgrid[137]

Segger 1.9.5[138,139]

SymphoTime 64 2.7 (PicoQuant, Germany)

# Chapter 3

# Methods

## 3.1 Molecular biology

### 3.1.1 In-fusion cloning

The different yPGK variants used in this project were obtained through in-fusion® cloning (Takara Bio, Japan). This is a versatile seamless cloning method That uses 3'-exonuclease digestion and in vivo repair to re-circularize the recombinant product. In-fusion® was mostly used during the project for single-insert cloning and site-directed mutagenesis. The standard workflow of the in-fusion® protocol is to linearize the vector by inverse PCR, as well as PCR-amplify the insert, followed by the in-fusion reaction, bacterial transformation, and clone screening. It is imperative that the insert contains a 15 bp homology to the corresponding vector ends to facilitate cloning and give the desired directionality.

PCR reactions were conducted for the vector and the insert separately using a 1 × Hifi PCR mix (Takara Bio, cat. no. 639298) in a three-step PCR cycle. Double DNA strands were denatured for 10 seconds at 98 °C, followed by 15 seconds annealing at 55 °C and DNA extension for 15 seconds/kbp at 72 °C. This cycle was repeated 35 times with a final 10 min DNA extension at 72 °C. The PCR products were purified by agarose gel extraction and PCR clean-up, respectively (Macherey-Nagel, cat. no. 740609.50). For the in-fusion reaction, 100 ng of insert and vector were mixed with 1 × in-fusion® mix (Takara Bio, cat. no. 638947) and the reaction was incubated for 10 min at 50 °C before the bacterial transformation.

### 3.1.2 Transformation

Commercial chemically competent *E. coli* cells (Stellar®, Takara bio, cat. no. 636763) with increased transformation efficiency were mixed with 2.5 µL of in-fusion reaction mix and incubated on ice for 15 min. Afterwards, the cells-DNA mixture was incubated at 42 °C for 30 s to promote the internalization of DNA and was further cooled on ice for 2 min before incubation for 1 h at 37 °C with 400 µL of super optimal broth with catabolite repression medium (SOC, Takara Bio, Japan). Lastly, 100 µL were plated and incubated at 37 °C overnight (O.N) in Luria broth (LB) agar plates supplemented with 100 µg/mL carbenicillin to grow the transformed cells selectively. On the next day, colonies were picked and sent for sequencing to confirm whether the cloning was successful.

## 3.2 Biochemistry

### 3.2.1 Bacterial expression of double cysteine variants for sm-FRET characterization

All yPGK variants were expressed in *E. coli* using the BL21(DE3) competent cells (New England Biolabs, cat. no. C2527H). Cells were grown at 37 °C in Terrific Broth medium (TB) (MP Biomedicals, cat. no. 113046022-CF) with shaking starting from $O.D_{600nm} = 0.2$ up to $O.D_{600nm} = 1$. Afterwards, protein expression was induced by the addition of 1 mM Isopropyl ß-D-1-thiogalactopyranoside (IPTG), and incubation was continued for 3 h at the same temperature. Finally, protein expression was stopped by harvesting cells at 6000 × g for 15 min at 4 °C. The cell pellet was resuspended in lysis buffer (30 mM MOPS pH 7.5, 500 mM NaCl, 0.001 mM TCEP) (5 mL/g) supplemented with cOmplete™ Protease Inhibitor Cocktail (1 tablet/50 mL) (Roche, cat. no. 11697498001). Cells were opened using a cell disruptor by passing them 2 × at a pressure of 1.7 mbar to increase protein recovery. The Cell lysate was centrifuged at 30.000 × g for 45 min to remove cell debris. A salting-

out step using Ammonium sulfate precipitation followed. First, the clear lysate was mixed with 40 % (w/v) $(NH_4)_2SO_4$ (AppliChem GmbH, cat. no. 141140) for 1 h at 4 °C. Non-precipitated proteins were recovered after centrifuging the solution at 5.000 × g for > 45 min. The process was repeated with 60 % (w/v) of $(NH_4)_2SO_4$. Soluble proteins were further purified with Nickel-Nitriloacetic acid (Ni-NTA) affinity chromatography. Batch purification was performed using 5 mL of resin from a 50 % (w/v) suspension Protino® Ni-NTA agarose (Macherey-Nagel, cat. no. 745400.500). Beads were equilibrated in lysis buffer before incubating them for 1 h with the protein solution. The beads were washed twice with 5 column volumes (CV) of lysis buffer containing 10 mM and 20 mM imidazole. Lastly, the protein of interest (POI) was eluted with 5 CVs of lysis buffer supplemented with 250 mM imidazole. The elution was dialyzed overnight in 400 times volume excess of lysis buffer plus 1:10 (w/w) of TEV-His×6 protease (produced in-house) to remove the purification tag. TEV protease and uncut protein were removed through reverse Ni-NTA affinity chromatography as described above, where the protein of interest was recovered in the flow-through. Finally, the protein preparation was polished in lysis buffer with size exclusion chromatography using preparative HiLoad 16/600 Superdex 200 pg (Cytiva, cat. no. 28989335). The presence of protein in the fractions of interest was confirmed by SDS-PAGE and Western blot.

### 3.2.2 Cell-free protein synthesis and purification

Sample preparation of coupled transcription/ translation systems was done with either RTS 100 *E. coli* HY kit (Biotechrabbit, cat. no. BR1400101) for sm-FRET measurements of truncated yPGK variants, or with PURExpress® *In Vitro* Protein Synthesis kit (New England Biolabs, cat. no. E6800S) for cryo-EM sample preparation.

#### 3.2.2.1 Expression of yeast PGK truncation variants for sm-FRET

Cell-free protein synthesis reaction was performed together with Strep-Tactin®XT Superflow® high-capacity resin (IBA, cat. no. 2-4030-002). First, 20 µL of resin slurry was washed three times with 50 µL buffer W (100 mM Tris-HCl pH 8.0, 150 mM NaCl) and centrifuged 1 min at 5.000 × g. Furthermore, the resin was blocked for 1 h at room temperature with 50 µL of buffer W supplemented with 0.5 % bovine serum albumin (BSA). Afterwards, the resin was washed three times with buffer W. Finally, the resin was equilibrated in 50 µL cell-free buffer (10 mM Tris-Acetate pH 8.2, 14 mM $Mg(OAc)_2$, 0.6 mM KOAc) three times and resuspended in 10 µL of the cell-free buffer. For *In vitro* translation, 50 µL reactions were incubated for 5 hours at 29 °C with 14.000 rpm shaking by mixing RTS 100 *E. coli* HY kit (Biotechrabbit, cat. no. BR1400101) components as described by the manufacturer with 5.5 nM DNA template and 10 µL of previously prepared Strep-Tactin resin. After incubation, the reaction mix was washed four times with 40 µL buffer W. All proteins were eluted by two incubations with 20 µL buffer BXT (100 mM Tris-HCl pH 8.0, 150 mM NaCl, 50 mM biotin, 1 mM EDTA). Protein concentration was determined from the SDS-PAGE gel using a standard curve of BSA. Elution fractions with POI were pooled together and stored in lysis buffer after buffer exchange using a Zeba spin desalting column with a cutoff of 7 kDa (Thermo Scientific™, cat. no. 89882).

#### 3.2.2.2 Expression of ribosome nascent chain complexes (RNCs) for sm-FRET

50 µL reaction of PURExpress® *In Vitro* Protein Synthesis kit (New England Biolabs, cat. no. E6800S) was used to express modified RNCs for sm-FRET characterization. Solution A and B were premixed with 0.4 µL of RNA oligo anti-tRNAser and incubated for 5 min at 37 °C to deactivate the endogenous tRNAser(AGU). Subsequently, 5 µM DNA oligo antisense tmRNA and 5.5 nM DNA template were added,

as well as two sets of components for incorporation of azido-phenylalanine (AzF) and cyclopropene-lysine (CpK). Each set contained the corresponding tRNA (tRNAamber or tRNAser(AGU)), tRNA synthetase (AzF-RS or CpK-RS), and unnatural amino acid (AzF or CpK), respectively. After mixing all components, the translation reaction was incubated at 37 °C for 1 h with 650 × rpm shaking. Following the incubation, the reaction mix was desalted using a Zeba spin desalting column with a cutoff of 40 kDa (Thermo Scientific™, cat. no. 87766) to remove small components. The sample was stored at 4 °C before labeling. A variation of the cell-free protein synthesis reaction was used to incorporate 2 × L-Homopropargylglycine (HPG).

#### 3.2.2.3 Sample preparation of ribosome nascent chain complexes (RNCs) for cryo-EM

The cryo-EM samples were prepared by starting a 750 µL translation reaction. The reaction mix was incubated at 37 °C for 1 h with 650 rpm shaking after mixing PURExpress® *In Vitro* Protein Synthesis kit (New England Biolabs, cat. no. E6800S) components according to the manufacturer with 5 µM DNA oligo antisense tmRNA and 5.5 nM DNA template. Afterwards, the reaction mix was incubated for 3 minutes at room temperature with 100 µM puromycin and immediately centrifuged in a 10 %–40 % sucrose gradient at 4 °C at 175.000 × g for 5 h using a SW32Ti (Beckman Coulter, cat. no. 369650). Polysome fractions were collected after fractionation (Biocomp Instruments, cat. no. 153) and sedimented with ultracentrifugation using TLA-110 (Beckman Coutler, cat. no. 366735) rotor at 164.000 × g for 3 h at 4 °C followed. The pellet was resuspended in 50 µL buffer A (50 mM HEPES pH 7.5, 25 mM $Mg(OAc)_2$, 10 mM KOAc) and mixed with 50 µL 100 µM DNA oligo anti-mRNA, 4 µL RNase H (New England BioLabs, cat. no. M0297S), 2µL 1M Dithiothreitol (DTT). RNase H reaction mix was incubated 1 h at 25 °C and stored overnight (O.N) at 4 °C. Monosomes were separated from polysomes by repeating the density gradient ultra-centrifugation. The 70S fraction was collected after fractionation, sedimented, and stored at 4 °C until further use. Fractions of all steps were collected for sample characterization by Western blot. A variation of

the protocol was followed to try to improve the resolution of the nascent chain in the tunnel vestibule and outside it. Briefly, after the RNAseH digestion reaction, the density gradient was performed together with a fixation agent (glutaraldehyde, sigma cat. no. G5882) gradient according to reference[140]. Everything else was kept as described above.

### 3.2.3 Labeling of proteins for single-molecule FRET experiments

The protein characterization with fluorescence techniques follows by attaching fluorescent labels to the protein amino acids. The most straightforward way to label proteins is to target primary amines and label them with fluorophores coupled with an N-hydroxyl-succinimidyl-ester group. The reactive group reacts irreversibly at room temperature. When the reaction is performed at pH >8.0, the selectivity of the reaction is increased to target primary amines. However, proteins normally have many lysines and arginines in their sequences (the primary source of primary amines in a protein), making this labeling strategy useless to determine pairwise distances with FRET. Therefore, these positions need to be modified to a more sporadically occurring amino acid so that they can be specifically labeled. The most common amino acid used is cysteine. This amino acid has a thiol group at its side chain, which can react with fluorophores coupled with a maleimide group. In some cases, proteins have many cysteines in their primary sequence, which decreases the practicality of this approach. Likewise, in the current work, the cysteines of the RNCs ribosomal proteins contribute to the available labeling positions, making it impossible to label the nascent chain specifically. Therefore, we introduced unnatural amino acids into the nascent chain to label it specifically.

#### 3.2.3.1 Cysteine labeling

The labeling reaction was performed with cysteine double variants and simultaneously for both positions by mixing equal parts of 10 × the final molar excess of

maleimide dyes (Thermo Scientific™ Alexa 488 cat. no. A10254, and Alexa 647 cat. no. A20347) with 10 µM yPGK. Changes in the labeling scheme were used for some yPGK variants by mixing different amounts of donor and acceptor into the protein solution to maximize labeling yield. The reaction was incubated for 2 h in the dark at room temperature and continued overnight at 4 °C. Excess-free dye was removed by passing the labeling mixture through a long Zeba™ dye and biotin removal 2 mL spin column (Thermo Scientific™ cat. no. A44298). The sample was kept at 4 °C until used in the confocal microscope.

#### 3.2.3.2 Unnatural amino acids labeling

The labeling reaction was performed sequentially by mixing 10 times molar excess of tetrazine dye (ATTO 488, or ATTO 647N) with 10 µM yPGK. The reaction was incubated for 4 h in the dark at room temperature. Excess-free dye was removed by passing the labeling mixture through a long Zeba™ dye and biotin removal 2 mL spin column (Thermo Scientific™ cat. no. A44298). Afterwards, the same steps were followed for a second DIBO dye (Thermo Scientific™ Alexa 488 cat. no. C20020, or Alexa 647 cat. no. C20022). The sample was kept at 4 °C until used in the confocal microscope.

## 3.3 Fluorescence measurements

### 3.3.1 Preparation of cover slides

All fluorescence measurements were performed with treated cover slides prepared as follows. High-precision glass 20 × 20 mm was submerged in piranha solution (1 part 35 % Hydrogen peroxidase ($H_2O_2$) in 2 parts Sulfuric acid ($H_2SO_4$) for 1 hour to chemically clean the surface. The slides were thoroughly washed with miliQ water and dried with nitrogen gas ($N_2$) before being subjected to plasma cleaning treat-

ment for 10 min to functionalize the silica molecules of the glass. Afterwards, the slides were submerged for 20 min in 2 % (3-Aminopropyl)triethoxysilane (APTES) diluted in acetone for silanization. The slides were washed 3 times with acetone and methanol, dried with $N_2$ gas, and incubated overnight with a 50 µL solution of 50 mg of phosphoethanolamine-N-hydroxysuccinimide (PEG-NHS) dissolved in sodium carbonate. After the whole procedure, the surface was ready to be loaded with fluorescent molecules.

### 3.3.2 Fluorescence correlation spectroscopy

Fluorescently labeled samples were measured in the confocal microscope (PicoQuant, Microtime 200) with fluorescence correlation spectroscopy (FCS) to determine the presence of double-labeled species, diffusion coefficient, and correct dilution for single-molecule measurements. First, the confocal volume of each laser (485nm and 640nm) was determined using standard fluorophores (Atto 655-NHS ester or Alexa488-NHS ester) with known diffusion coefficients[141,142]. The calibration mix was prepared with approximately 2 nM of Atto655-NHS ester and 4 nM of Alexa488-NHS ester. The calibration measurement was followed by pipetting 30 µL of calibration mix on a small cover slide 22 × 22 mm with 0.17 mm thickness (Marienfeld, cat. no. 0107052) on top of a water immersion high numerical aperture objective (Olympus UPLanSApo 60×/1.2 W cat. no. N1480800). Laser power was adjusted to 15 µW for the 640 nm laser and to 4 µW for the 485 nm laser to be in the linear regime of the dye's molecular brightness. Measurements were performed 3 × for 3 min each 10 µm above the second focal plane of the cover slide. Auto-correlation curves and determine confocal volumes from each laser. Afterwards, the sample was diluted to 4 nM concentration, and the measurement was repeated as described above. Samples that met the criteria necessary (presence of double-labeled species and correct diffusion coefficient) were diluted appropriately for single-molecule measurements.

### 3.3.3 Single-molecule FRET measurements

Single-molecule measurements were performed by diluting the fluorescently labeled sample to 13 pM of the highest concentrated fluorophore on a small cover slide sealed with parafilm to reduce sample evaporation. Measurements were performed in PIE mode with 20 MHz pulses. Laser power was adjusted to 9 µW for the 485nm and 640 nm lasers to be in the linear regime of the dye's molecular brightness. Samples were measured 10 µm above the second focal plane of the cover slide. Measurements were performed for at least 2 h. Longer measurements were performed if the number of events obtained was not enough. Data from the measurements was analyzed with the PAM toolkit from LMU University[134].

## 3.4 Cryo electron microscopy (cryo-EM)

### 3.4.1 Sample preparation for negative stain

Grids for negative stain were prepared with 5 µL of ribosomal sample at 0.05 – 0.1 mg/mL applied on glow discharged (90 s 15 mA) 300 mesh copper support grids (Electron Microscopy Sciences cat. no. CF300-Cu-50). After waiting for 30 s, the excess sample was blotted on filter paper following two washes in buffer A. Finally, the grid was dipped in 3 µL 2 % Uranyl acetate solution and immediately blotted. The process was repeated once more with 2 min waiting time before blotting. The sample was allowed to dry for 2 min before it was stored in a grid box for later imaging on the Talos 120.

### 3.4.2 Image recording of negatively stained EM grids

Negatively stained grids were imaged using ThermoFisher Talos L120C G2 transmission electron microscope operated at 120 kV acceleration voltage. Images were

taken with a Thermo scientific 4k × 4k CETA 16M CEMOS camera at 115× magnification for micrograph overview and 92.000x for high magnification. All images were collected using TEM imaging & analysis software (TIA, ThermoFisher Scientific).

### 3.4.3 Sample preparation for cryo-EM

Sample preparation for imaging with a transmission electron microscope in cryogenic conditions followed the general principle of plunge freezing. This principle was applied in the same way for all types of grids used, and it was independent of the device. Briefly, the grids were prepared using a vitrobot Mk IV (Thermofisher Scientific) or EM GP2 automatic plunge (Leica) freezer after 90 s glow discharging at 15 mA. 5 µL were applied on the grid inside a climate-controlled chamber at 4 °C and 95 % humidity. The sample was allowed to adsorb to the grid for 30 s before blotting it for 2.5 s and quickly plunging it into liquid ethane at -180 °C. Small variations of the protocol were done to maximize the number of grid squares with thin (∼50 nm) vitreous ice. Finally, grids were clipped and stored in liquid nitrogen storage before imaging.

### 3.4.4 Data acquisition in cryo-EM

Cryo-EM grids were imaged with automated data collection software EPU (ThermoFisher) either on a Titan Krios G4 electron microscope operated at 300 kV voltage acceleration or in a Talos Arctica operated at 200 kV voltage acceleration. The Krios images were acquired with a Falcon IV detector at a 96.000× magnification, yielding a pixel size of 0.808 Å, whereas Arctica images were acquired with a K3 camera coupled with Biocontinuum imaging filter (Gatan) in super-resolution counting mode at 100.000x magnification with objective aperture of 100 µm, yielding a pixel size of 0.408 Å. Multi-frame movies (total dose: 50 $e^-/Å^2$, 50 frames, dose per frame: 1 $e^-/Å^2$) were recorded with nominal defocus range of -0.5 µm to -2.5 µm in 0.5 µm steps.

### 3.4.5 Data processing

All movies were processed with the software Cryosparc[131] and RELION[136]. The final maps presented in the thesis were obtained from the following pipeline. Pre-processing was performed with the Cryosparc live tool, which does the processing on-the-fly, while the images are being recorded on the microscope. Briefly, the movies were merged into micrographs after running the Patch motion correction job with default parameters. Afterwards, patch CTF estimation job was done to determine the accurate defocus value of each image and later correct their CTF. Particles were picked with the blob picker tool, using a particle diameter range of 200–300 Å. Particles were extracted with a box size of 480 px and binned to 160 px. The 2D classification was performed with 50 classes to generate templates from different particle views to pick particles with the template-picking job. Particles were picked with 300 Å diameter and a 150 Å distance between particles. Full-size particles were extracted with a box size of 480 px. At least two rounds of 2D classification with 150 classes each were implemented to remove junk particles. Particles from good classes were used to generate a template-free initial model using the *Ab-initio* job. Homogeneous refinement with per-particle defocus optimization, and per-micrograph CTF parameters optimization (tilt, trefoil, and anisotropic magnification) was executed. Particle stacks were curated by removing damaged particles after force-hard classification heterogeneous refinement with 4 classes using four copies of the homogeneous refinement volume as starting reconstructions. Consensus volume was generated by repeating the homogeneous refinement job described before from the high-resolution classes. Particles were down-sampled to 160 px for all downstream processing. The final consensus reconstruction was followed with full particles. The signal of the full ribosome excluding tRNA sites was subtracted from particle images with the particle subtraction tool after masking the remaining volume of the ribosome density map in chimera[130] with Segger[138,139]. Focused 3D classification of subtracted particles was performed using a local mask covering the density of all tRNA sites. Particles showing the presence of all tRNAs were selected for further processing. Afterwards, 3D variability analysis was completed using a local mask

outside the ribosomal tunnel to select particles showing the presence of the peptide nascent chain. A Final consensus refinement of ribosome nascent chain complexes was obtained from full particles as described above.

### 3.4.6 Model building

The peptide chain emerging from the ribosomal tunnel was modeled *de novo* in the 70S ribosome 7k00.pdb[143] using Coot. Additionally, the crystal structure of yPGK 1QPG.pdb[144] was rigid body fitted to the RNC map using Chimera software[130] and connected to the nascent chain in Coot[127]. The model was refined using the Phenix auto refinement tool[135], and the rotamer outliers were fixed in Coot[127]. This process was repeated iteratively until the best model was obtained. The final map was sharpened according to the model B-Factors with LocScale[133]. Finally, the model was validated with Phenix Molprobity tool[135].

# Chapter 4

# Results and Discussion

## 4.1 sm-FRET

### 4.1.1 *In silico* analysis of putative labeling positions on the N-terminal domain of yeast PGK

Sm-FRET can detect conformational/structural states and study unfolding/refolding events by mapping distances between different pair positions along the nascent chain. In practice, we compare the different pair distances for the truncated nascent chains against the same pair distances from the native protein. This comparison would allow us to gain insight into the folding pathway and folding degree of emerging nascent chains. Therefore, we initially characterized and determined classic unfolding/refolding experiments for all target pairs with full-length protein.

This project followed previous studies from the lab, which properly characterized most pairs we aimed to use during the course of this study[60].

**Table 4.1:** Yeast PGK predicted distances for known FRET pairs. ID – inter-domain. Data retrieved from [60].

| Dye attachment positions | Domain | Short Name | Theoretical distance (Å) |
|---|---|---|---|
| 135-290 | ID | ID1 | 40 |
| 1-135 | N | N1 | 48 |
| 202-256 | C | C1 | 49 |
| 202-290 | | C2 | 57 |
| 256-290 | | C3 | 25 |

We could see from Table 4.1 that the N-terminal domain had only one FRET pair

compared to the three pairs within the C-terminal domain. Therefore, we screened the N-terminal domain to find two additional appropriate pairs. Furthermore, the selected FRET pairs in the N-terminal domain were later characterized by FCS and sm-FRET (see summary on Table 4.2 and Figure 4.5).

The selection of the new FRET pairs was done according to the following criteria. First, we chose unstructured regions (loops), thus allowing the protein to maintain its native biophysical properties, even upon sequence changes. Next, we searched for pairs with a distance between each residue in the range of 0.55 to 1.5 times the Förster radius of the specific donor-acceptor fluorophore pair used for the measurements. For this study, we used the fluorophores Alexa 488 and Alexa 647, and the Förster radius was determined to be 56 Å (see [60] for detailed discussion). Therefore, accepted distances should be greater than 31 Å and smaller than 84 Å.

The selection of the right attachment points is a crucial step when finding new FRET pairs. As shown in Section 1.5.1 Figure 1.7, the highest variation in FRET energy transfer is observed for distances close to the Förster radius. In this region, it is possible to observe a sharp fall-off of the FRET efficiency as the distance between fluorophores increases. Therefore, two residues that are away from each other at a distance within this regime are the most sensitive positions to detect structural changes in the residues enclosed by the fluorescent dyes.

We selected two new FRET pairs in the N-terminal domain to match the number of FRET pairs present in the C-terminal domain. The first pair was S1–G88, and the second pair was T34–Q135. Two of the four positions, namely S1 and Q135, were previously characterized, and it was therefore known that they do not disturb the structure/folding of our protein of interest (positions marked in orange on Figure 4.1).

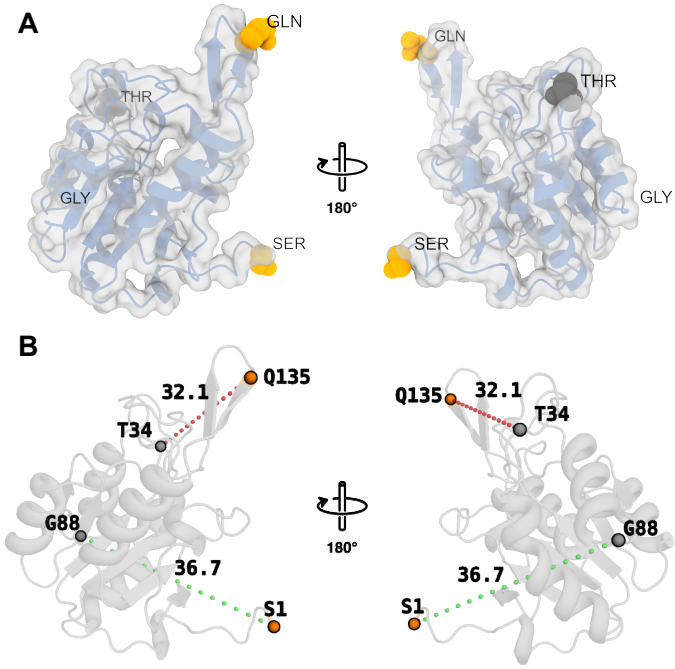



**Figure 4.1:** Selection of FRET pairs in yeast PGK N-terminal domain. A – Representation of surface accessible residues used for new FRET pairs on the N-terminal domain of yPGK (1qpg.pdb). The protein surface, excluding side chains, is depicted in gray. Colored in gray are the accessible residues which were not tested experimentally. Marked in orange are the positions that were tested previously. B – Distance between selected pairs (distances were selected for values 0.6 × above the Förster radius of the fluorophores used). In green is marked the distance in Angstroms of the new pair between position 1 and 88. In red is marked the distance in Angstroms of the new pair between position 34 and 135.

To accurately predict energy transfer efficiency based on the physical distance between the two attachment positions in the protein, we needed to determine the theoretical distance between the two attachment positions, as shown in table 4.2. The theoretical distances for the new FRET pairs were determined using the tools developed previously by the Claus Seidel group[145]. Considering Equation 1.5 (see Section 1.5.1), we could convert the predicted distance ($R$) to the expected energy transfer efficiency ($E$).

**Table 4.2:** Calculation of accessible volume distances for new N-terminal domain FRET pairs. The distances were determined with the toolkit developed by the Claus Seidel group [145].

| Dye attachment positions | Domain | Short Name | Theoretical distance (Å) | Theoretical FRET ($E$) |
|---|---|---|---|---|
| 1–88 | N | N2 | 53 | 0.58 |
| 34–135 | | N3 | 43 | 0.83 |

### 4.1.2 Biochemical characterization of N-terminal domain FRET pair yPGK variants

Having selected the two new FRET pairs, we proceeded with their expression and characterization. Purification of yeast PGK was previously optimized in the lab. The most important phases in purification were two sequential salting out steps with 40 % and 60 % ammonium sulfate, affinity chromatography using 6 × His-tag (6×His), and size exclusion chromatography. As expected, yPGK migrates with the correct size in the representative purification SDS-PAGE gel (Figure 4.3), and after Nickel-NTA (Ni-NTA) affinity chromatography, it is mostly pure. In the final polishing step, yPGK is eluted in a single peak at the expected retention volume.

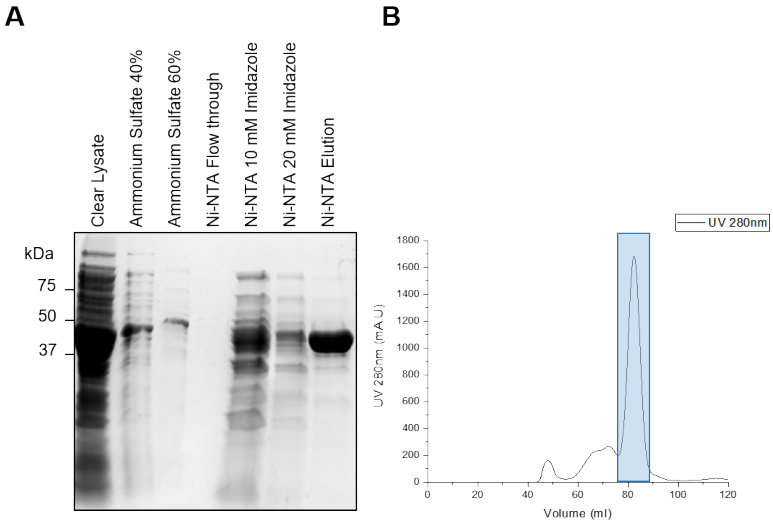



**Figure 4.2:** Purification of yeast PGK N-terminal domain FRET pair variants. A. SDS-PAGE showing all yPGK purification steps. B. Size exclusion chromatography showing the eluted yPGK peak.

### 4.1.3 sm-FRET measurements of N-terminal domain FRET pair yPGK variants

We labeled each yPGK variant with Alexa dyes (Alexa488 and Alexa647) to measure FRET efficiency using a confocal setup. Dye ratios were optimized for each FRET pair to achieve the highest possible double labeling. We used Ni-NTA magnetic beads, taking advantage of the 6 × His and TEV cleavage site to remove all free dye after the labeling reaction (Figure 4.3). Lanes FT, W1, W2, and W3 show unbound yPGK. Our fraction of interest, FT2, shows the presence of both cleaved yPGK and TEV protease. However, only yPGK is labeled (Figure 4.3 — (black triangle) UV panel). Therefore, the sample should be ready for confocal measurements. The remaining lanes, E and Beads, were used as a control to judge how much yPGK was lost during the labeling procedure. By comparing FT2 with both control lanes, we observed a minimal loss compared to the band intensity of the yPGK band present in the FT2 lane.

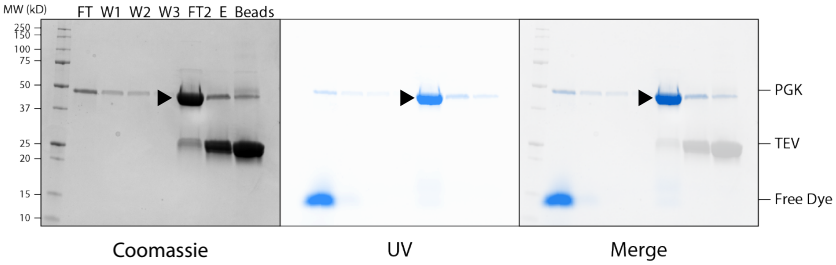



**Figure 4.3:** SDS-PAGE gel of labeling reaction. Panel Coomassie shows proteins labeled with instant blue. Panel UV shows proteins coupled with Alexa488. Panel Merge overlays panel Coomassie and UV. Lane FT has all proteins of the labeling reaction that did not bind to the magnetic beads. Lanes W1-W3 show protein that was removed from the magnetic beads with washing steps. Lane FT2 shows the unbound protein after incubation with TEV protease. Lane E corresponds to the elution of uncleaved protein after incubation of the beads with buffer supplemented with biotin. Lane Beads show protein that was not eluted from the previous step.

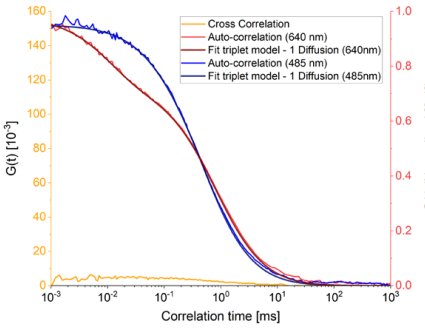
Before measuring the sample at a single-molecule level, we performed measurements in the nM concentration regime and analyzed them with FCS (Figure 4.4 A and 4.4 B). The diffusion coefficient of yPGK was calculated from the temporal autocorrelation function obtained from the time measurement (see example Figure 1.6 B), fitted with the triplet state model with one diffusion species (Equation 4.1). Later, the value obtained was compared to the theoretical value of 72 $\mu m^2/s$ (determined from the crystal structure of yeast PGK (1qpg.pdb) using the HullRad server (http://52.14.70.9/index.html)). The diffusion coefficient for 640 nm and 485 nm channels from yPGK variant 34–135 were within an acceptable range compared to the predicted theoretical value of yPGK. The yPGK variant 1–88 also showed a diffusion coefficient of $\sim$60 $\mu m^2/s$, which is similar to that of the yPGK variant 34–135. After determining the diffusion coefficient, we calculated the respective hydrodynamic radii of the diffusing particles, resulting in values of approximately 3.2 nm and 3.6 nm, respectively. Both variants showed hydrodynamic radii within an acceptable range compared to the theoretical value of 3.0 nm.

$$G(t) = \left[1 + T\left[exp\left(-\frac{t}{\tau_{Trip}}\right) - 1\right]\right] \sum_{i=0}^{n_{Diff}-1} \frac{\rho[i]}{\left[1 + \frac{t}{\tau_{Diff}[i]}\right]\left[1 + \frac{t}{\tau_{Diff}[i]\kappa^2}\right]^{0.5}} \quad (4.1)$$

where $T$ is the dark (triplet) fraction of molecules, ($\tau_{Trip}$) is the lifetime of the dark (triplet) state, ($t$) is the time, ($n_{Diff}$) is the number of diffusing species, ($\rho$) is the contribution of the diffusing species, ($\tau_{Diff}$) is the diffusion time, and ($\kappa$) is the length to diameter ratio of the focal volume.

The fitting model can evaluate the presence of several diffusing species and characterize them. From the labeling protocol, the sample could be contaminated with free dye, or it could have induced protein aggregation. Due to the large difference between the diffusion coefficients of free dye (∼300–400 $\mu m^2/s$), protein (∼50–150 $\mu m^2/s$), and aggregates (∼5–10 $\mu m^2/s$), FCS allowed the correct characterization of protein species present in the sample. Moreover, the triplet model accounts for dark states of the dyes that can be induced by the local environment of the attachment points. Overall, the obtained results gave a strong indication that the observed labeled species were indeed monomers of yPGK. Moreover, from the cross-correlation FCS curve, it is possible to judge whether any double-labeled species were present in the sample. In Figure 4.4 A, the orange curve showed a small amount of cross-correlation between channels 640 nm and 485 nm. Therefore, I assumed that there were double-labeled species in the sample.

**A**




**B**

| | 640 nm Channel | 485 nm Channel |
|---|---|---|
| $V_{eff}$(fl) | 2.8 | 1.3 |
| N | 9.1 | 6.4 |
| C (nM) | 5.4 | 8.1 |
| D (μm²/s) | 68.0 | 64.0 |

**Figure 4.4:** FCS data analysis of yPGK variant 34–135. A – FCS curves and fits obtained from triplet model. B – Fit parameters obtained from auto-correlation fit model.

After the biophysical characterization with FCS analysis, the sample was diluted to single-molecule concentrations to determine FRET efficiencies (Figure 4.5 A and

B). In Figure 4.5 A, a time window of 20 s measurement with a bin size of 1 ms is depicted. Bursts were selected in PAM[134] using the interphoton time with the Lee filter model. Conventionally, we selected bursts with more than 40 photons, with interphoton times between 150–300 µs after smoothing with Lee filter $N = 2$. After the selection of the bursts, we obtained the 2D plot from which we can determine the FRET efficiency (Figure 4.5 B). Before applying any correction factor, the non-corrected FRET efficiency is designated as proximity ratio ($PR$). First, the background correction was determined from the average count rate of the time traces for each channel individually. Then, the correction factors of cross-talk and direct excitation were determined from the donor-only and acceptor-only populations, respectively. As explained in Hohlbein et al.[108], cross-talk (ct) and direct excitation (de) are calculated from equations 4.2 and 4.3, respectively. Finally, we took the value for $\gamma$ correction factor from Cerminara et al.[60]. The same procedure was repeated for the second variant and all concentrations of guanidine hydrochloride GuHCl as well.

$$ct = \frac{PR^{D-only}}{1 - PR^{D-only}} \tag{4.2}$$

$$de = \frac{S_{raw}^{A-only}}{1 - S_{raw}^{A-only}} \tag{4.3}$$

where $PR$ and $S_{raw}$ are the proximity ratio and raw stoichiometry after background correction, respectively.

Some improvements could have been performed to obtain optimal 2D plots. In Figure 4.5 B, the FRET population had a $S$ close to 0.3. Ideally, the population should be at $S = 0.5$. One way to avoid a shift of the FRET events to lower stoichiometry values is to tune the power output of the acceptor and donor channels. Since the $S$ value depends on the number of photons present in each burst in all individual channels, it is important to optimize the photons emitted according to the fluorophores' photophysical properties, as explained by Hohlbein et al.[108]. To be sure that the laser intensities are appropriate for the dye pair used during the

measurements, one could use a sample with 100 % double-labeled species (e.g., dsDNA) and repeat single-molecule measurements with different laser intensities until the FRET population has $S = 0.5$.

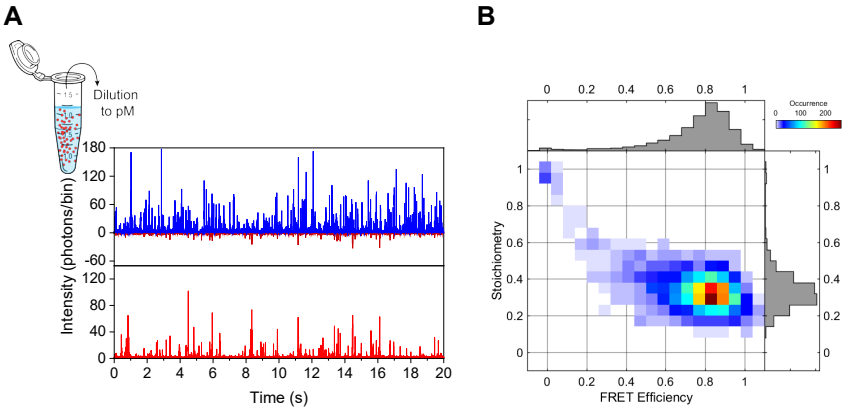



**Figure 4.5:** Single-molecule FRET data analysis of yPGK variant 34–135. A – Example of time trace of single-molecule measurement visualized with PAM[134]. Upper panel: Signal collected by donor and acceptor channels after donor excitation. Lower panel: Signal collected by acceptor channel after acceptor excitation. B – 2D plot of accepted bursts. The plot was corrected for background, cross-talk, direct excitation, and $\gamma$ factor. Applied $\gamma$ correction factor was taken from[60].

#### 4.1.3.1 Unfolding transitions in yeast PGK with label positions in the N-terminal domain

The following section was based on the manuscript published by Bustorff et al. in *Biomolecules* 2023 [II].

Single-molecule fluorescent measurements of both yPGK variants in native conditions showed a FRET efficiency similar to the ones predicted in Table 4.2 (Figure 4.6 [GuHCl] – 0 M). These results support the fact that both variants show a native fold.

After measuring the yPGK variants in native conditions, we added increasing amounts of GuHCl and performed single-molecule FRET measurements to determine FRET efficiency ($E$) in different denaturing conditions. We started from low

amounts of GuHCl up to a concentration that would completely unfold yPGK (1.5 M) shown previously in the literature[60] (full denaturing series can be found in Appendix A).

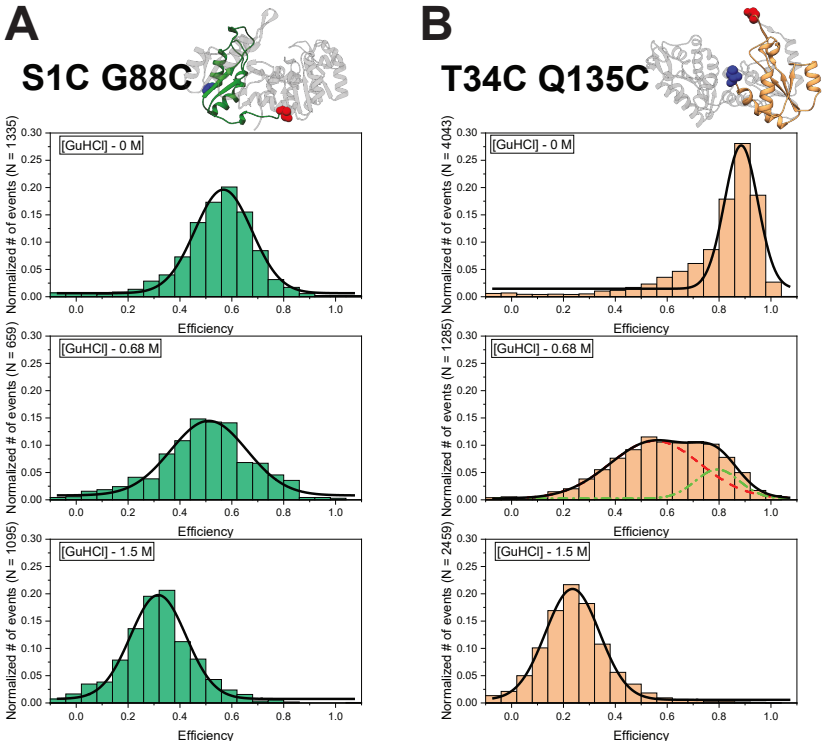



**Figure 4.6:** Denaturation series of N-terminal domain yPGK variants using GuHCl. The attachment positions of the dyes are highlighted as spheres. The structural elements present between the attachment positions are colored in green – variant N2 and copper – variant N3. The corresponding transfer efficiency histograms for each measurement with specific GuHCl concentrations are represented below. The solid and dashed lines represent a Gaussian fit with one or two peaks.

After performing the denaturation series, we observed two distinct behaviors (Figure 4.7). The FRET efficiency of the yPGK variant 1–88 showed a small decrease in E (∼0.6 to ∼0.4) with increasing amounts of the denaturant GuHCl. On the other hand, the yPGK variant 34–135 showed a shift from high FRET to low FRET with increasing concentration of GuHCl. At intermediate concentrations of GuHCl (0.5

– 0.8 M), we observed two populations of yPGK 34–135. One with low FRET efficiency and another with high FRET efficiency.

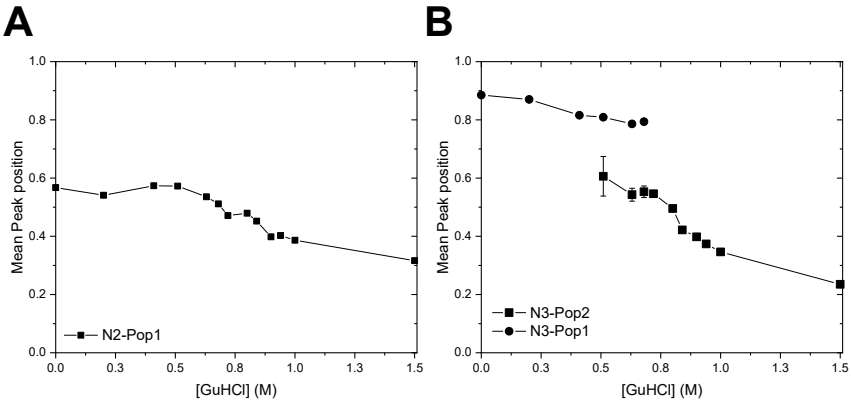



**Figure 4.7:** Effect of Guanidine hydrochloride on the FRET efficiency of yPGK variants labeled in the N-terminal domain. The mean peak position of FRET efficiency histograms. The N2 variant had only one population, whereas the N3 variant had a high FRET population and a low FRET one.

Overall, we can observe two separate unfolding transitions for variants 1–88 and 34–135. In variant 1–88, no transition was observed at any GuHCl concentration, while variant 34–135 showed a classical two-state transition from high to low FRET upon increasing amounts of the denaturant. Interestingly, the C-terminal domain folding pathways showed similar behavior, which was previously described by Cerminara et al.[60].

**Table 4.3:** Summary of distance changes observed in sm-FRET measurements of yPGK intra-domain labeling pairs. The inter-dye distances ($R_{DA}$) were calculated in Å from the mean peak position of the FRET histogram of yPGK in native, intermediate, and unfolded states. The theoretical distance based on a polymer chain model was estimated for each label pair, giving a prediction for the unfolded distance. According to the measured distances, we classified each label pair unfolding transition as one of the following: NT – no transition; CI – Compact intermediate; TS – Two states. Adapted from [II] and [60].

| | **N-domain** | | | **C-domain** | | |
|---|---|---|---|---|---|---|
| Label positions | 1–88 | 1–135 | 34–135 | 202–256 | 202–290 | 256–290 |
| | **N2-variant** | **N1-variant** | **N3-variant** | **C1-variant** | **C2-variant** | **C3-variant** |
| # Residues | 88 | 135 | 101 | 54 | 88 | 34 |
| Native $R_{DA}$ | 48 Å | 51 Å | 38 Å | 51 Å | 57 Å | 31 Å |
| Intermediate $R_{DA}$ | | 43 Å | | | 45 Å | |
| Unfolded $R_{DA}$ | 52 Å | 75 Å | 65 Å | 51 Å | 63 Å | 54 Å |
| Polymer $\sqrt{\langle R^2 \rangle}$ | 61 Å | 76 Å | 65 Å | 48 Å | 61 Å | 38 Å |
| Type of transition | **NT** | **CI** | **TS** | **NT** | **CI** | **TS** |

Together with data published earlier, Table 4.3 summarized the existence of three distinct unfolding transitions present in each domain. In terms of the observed native inter-dye distance, all label positions were in good agreement with the distances calculated from the crystal structure of yPGK. As expected, apart from the no-transition label pairs, all other pairs showed a larger inter-dye distance in the unfolded state. Therefore, we observed the expansion of the protein chain in the presence of a denaturant due to the rearrangement of the bonds formed between the amino acids.

The predicted distance of the unfolding state between the attachment positions was determined assuming a random coil arrangement that can be calculated according to the following equation:

$$\left\langle R^2 \right\rangle = n \cdot l^2 \tag{4.4}$$

where $n$ corresponds to the number of residues and $l$ is a constant named persistence length[146]. Since the formula ignores the neighboring residues from the whole protein, it was likely that we would not observe a complete match between the predicted and measured distances. Nevertheless, all distances with the exception of N2 and C3 were in good agreement.

The support of the data obtained from the N-terminal domain label pairs was fun-

damental to characterize the unfolding footprints of the Rossmann fold. This is particularly true for yPGK since both domains have almost identical folds. It was expected that the Rossmann fold displayed similar transitions independently of its primary sequence. Indeed, our data supports this hypothesis but also raises further questions. Should we expect similar behaviors of Rossmann folds present in multi-domain proteins? Moreover, would it be plausible to observe similar pathways in different motifs such as *beta*-barrels?

Earlier studies reported the presence of compact intermediates in similar $\alpha/\beta$ motif proteins, which the authors attributed to a molten globule state (MG)[147]. The folding state was previously described as a thermodynamically stable state with a less compact arrangement of secondary structure elements[148]. In our case, we interpreted the compact intermediate as a transient misfolded state of the protein during the unfolding/folding transition.

To better understand the recorded data concerning the unfolding patterns of the protein structure, we looked at the topological diagrams and 3D structure of the individual domains, highlighting the residues enclosed by the dye attachment positions and suggested a possible link between the observed transition type and the protein structures (Figure 4.8 and 4.9). These diagrams showed that a similar secondary structure organization encloses the label positions from N- and C-terminal transitions. In both domains, the region closest to the N-terminus of the protein displayed no transition, and similarly, both regions closest to the C-terminus displayed a two-state transition.

Inspecting the full Rossmann fold with the dyes attached to positions enclosing 4 of the 6 parallel $\beta$-sheets, we observed an unfolding pathway with a compact intermediate. However, the upstream portion of the Rossmann fold enclosing just 2 of the parallel $\beta$-sheets did not show any FRET mean peak transition, which we named the 'no transition pathway.' In contrast, in the second portion of the Rossmann fold, we observed a classic two-state unfolding pathway.

Finally, the domain topologies provided some evidence that the presence of a compact intermediate might be related to the position at which the dye is placed in the Rossmann fold. Either the coverage of a higher number of parallel $\beta$-sheets to-

wards the N-terminus (position 1 and 202 versus 34 and 256) revealed the compact intermediate, or this transition was observed due to the attachment point present in the anti-parallel $\beta$-sheet of position 135 and 290. In the presence of GuHCl at intermediate denaturant concentrations, the elements of the Rossmann fold come in close proximity to another anti-parallel $\beta$-sheet. At this point, the protein transiently acquires a non-native packing of its secondary structure, which is used as an off-pathway to the native state. Similar evidence can be found in the literature[147,149].

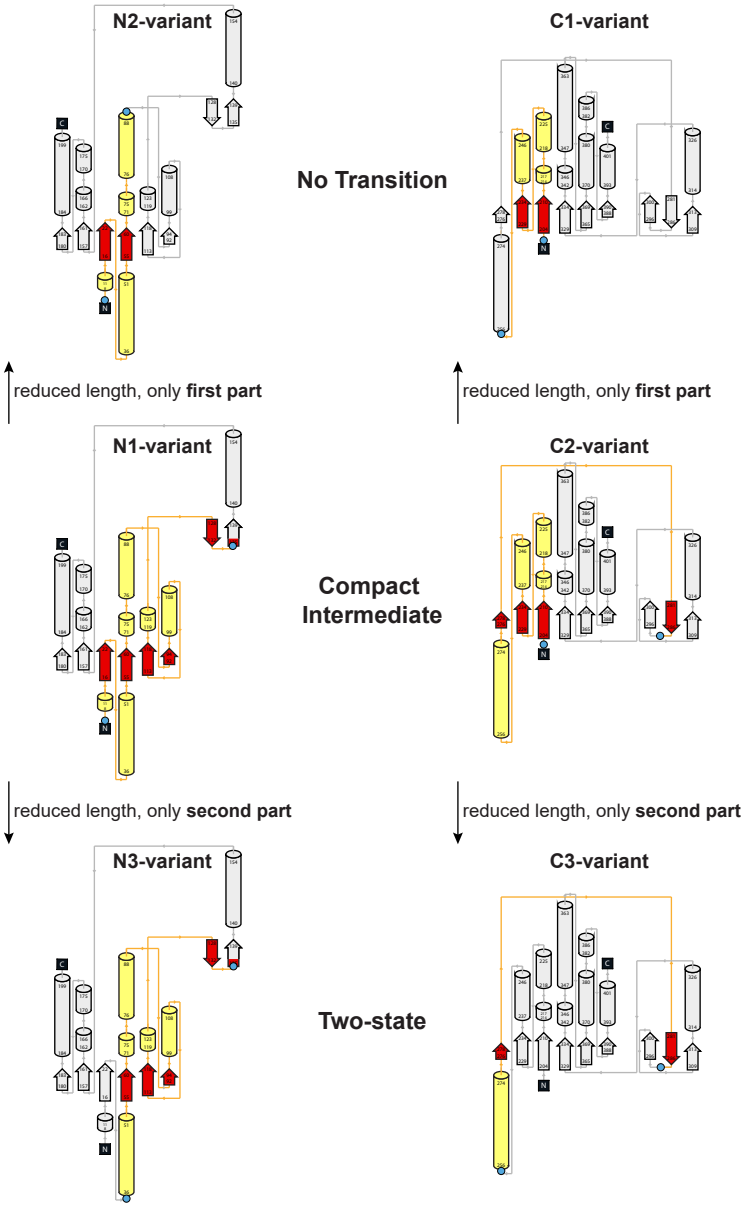



**Figure 4.8:** Topology scheme of unfolding transitions present in yPGK single domains. Diagrams of the N-terminal domain (left column) and the C-terminal domain (right column) were displayed vertically by unfolding transition. Structural elements were highlighted in yellow ($\alpha$-helices) and red ($\beta$-strands/sheets) colors. Attachment positions were represented as blue circles. Adapted from [II].

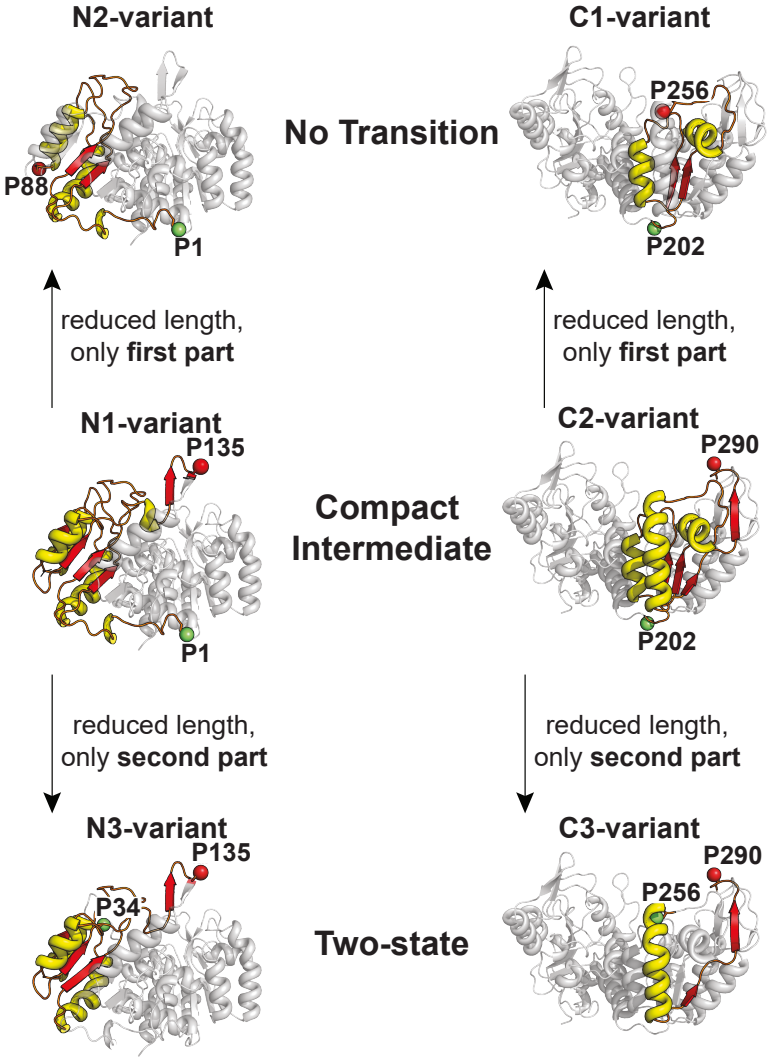



**Figure 4.9:** Comparison of folding transitions of yPGK domains. The 3D structures of yPGK are shown with individual attachment positions (red and green spheres) for all intra-domain distances within the N-terminal domain (left column) and within the C-terminal domain (right column). Structural elements were highlighted in yellow ($\alpha$-helices) and red ($\beta$-strands/sheets) colors. The 3D structures were organized according to their unfolding transitions. Adapted from [II].

### 4.1.4 Incorporation of unnatural amino acids

#### 4.1.4.1 Limitations of cysteine labeling

After completing the characterization of the new yPGK variants and having a better understanding of the folding pathways of free full-length yPGK in solution (discussed in section 4.1.3.1), we proceeded with experiments to study yPGK folding coupled with the ribosome as a ribosome-nascent chain complex (RNC). Our target was to specifically label the nascent chain to compare the results of the RNCs with the ones of the free full-length protein. Upon examining the surface proteins of the ribosomes, we can easily identify at least ten accessible cysteines (Figure 4.10). Therefore, using cysteine labeling for this project would result in labeling multiple proteins instead of specifically labeling the nascent chain.

We circumvented this challenge by incorporating unnatural amino acids. Multiple strategies could have been used to incorporate the desired residues. We followed the work developed by Sadoine et al.[150,151], demonstrating site-specific incorporation of azido-phenylalanine (AzF) using the nonsense amber stop codon UAG. Initially, we attempted to incorporate 2 × AzF using a release factor 1 (RF1)-depleted commercial kit from Biotechrabbit. Unfortunately, we could not obtain data showing bursts for molecules double-labeled with donor and acceptor dyes.

Subsequently, we tried a selective labeling approach by incorporating a pair of sense (AGU serine) and nonsense (UAG amber stop) codons[152]. With this strategy, we could target the nascent chain selectively by introducing two different labeling chemistries.

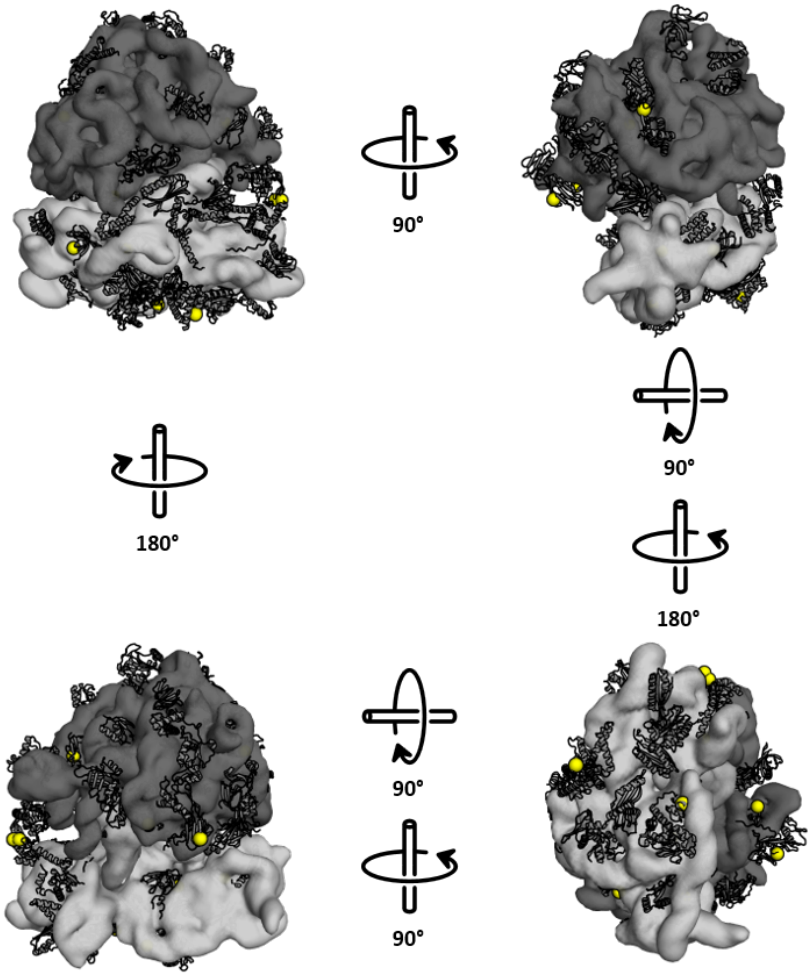
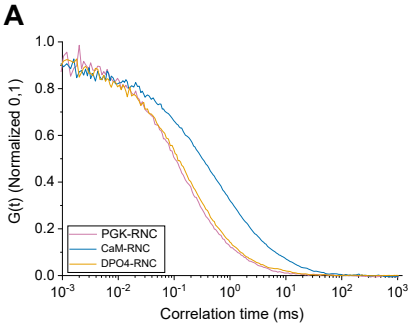



**Figure 4.10:** 70S ribosome surface accessible cysteines. 7k00.pdb structure with rRNA and tRNAs shown as surface. Surface-accessible cysteines are highlighted as yellow spheres. Spheres were scaled 3x to be more visible.

#### 4.1.4.2 Limitations of protein size in the incorporation of unnatural amino acids

Unfortunately, we faced unexpected limitations that did not allow this approach to be successful. To check whether the total number of amino acids of the full-length

nascent chain plays a role in the successful formation of RNCs, we prepared RNCs of 3 different proteins with varying nascent chain lengths (yPGK – 415 amino acids; DNA polymerase 4 (DPO4) – 352 amino acids; and Calmodulin 1 (CaM) – 149 amino acids). As shown in Figure 4.11 A and B, longer nascent chains did not give the expected results in FCS. Auto-correlation curves from yPGK-RNC and DPO4-RNC samples dropped earlier than the CaM-RNC (left shift of auto-correlation curve (pink and orange)), meaning that both complexes diffused faster than CaM-RNC. Analyzing the fitted parameters of diffusion coefficient (obtained by fitting the Equation 4.1 from Section 4.1.3), we deduced that CaM-RNC diffuses with values similar to those of ribosome diffusing particles, whereas yPGK-RNC and DPO4-RNC have diffusion coefficients that indicate the existence of free protein. Presumably, longer constructs, i.e., PGK and DPO4, were not stalled properly by the presence of the SecMstr arrest sequence, in contrast to shorter constructs, i.e., CaM. Several factors may influence the outcome of the reaction. First, it is crucial to suppress the naturally occurring $tRNA^{Ser}(AGU)$. Otherwise, serine will be preferentially incorporated over the unnatural CpK. To this end, an antisense-modified RNA oligo was added to bind and block the endogenous tRNA.




B

| | CaM-RNC | PGK-RNC | DPO4-RNC |
|---|---|---|---|
| $V_{eff}$ (fl) | 1.23 | | |
| N | 14.0 | 18.0 | 4.5 |
| C (nM) | 19.0 | 24.0 | 6.1 |
| D ($\mu m^2/s$) | 22.0 | 104.0 | 79.0 |

**Figure 4.11:** Comparison of fluorescent measurements of RNCs with varying nascent chain lengths. A – FCS curve normalized of different protein-RNC labeled complexes. B – FCS fit parameters according to fixed values of detection volume.

Second, it is essential to prevent the premature release of the nascent chain. The amount of RF1 was reduced to increase the chance of the unnatural AzF to incorporate and suppress the release of the polypeptide chain. Since all these steps are

not entirely efficient, this leads to incomplete reactions. In addition, the further away from the N-terminus the two unnatural amino acids are being incorporated, the greater the chances of an error are. The combination of the above facts could be why we could not effectively express and label RNCs of proteins larger than 17 kDa.

Following the incorporation of CaM-RNC, we could show that the protein is active while still bound to the ribosome (Figure 4.12 A and B). When changing the buffer of calmodulin from calcium-free buffer (apo) to calcium-rich buffer (holo), we observed a shift in FRET efficiency from low-FRET to high-FRET. This observation is in line with results from the literature, confirming that CaM adopts a compact conformation when bound to $Ca^{2+}$.[150]

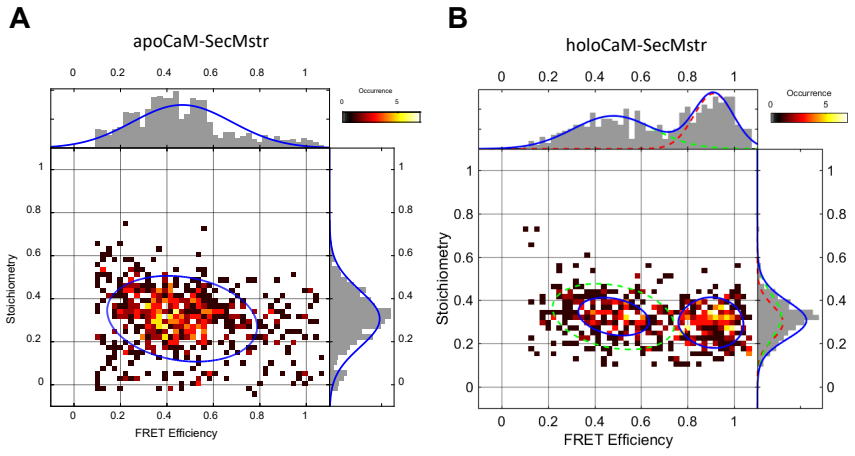



**Figure 4.12:** Characterization of CaM-RNC. A – 2D plot of CaM-RNC in calcium-free buffer. B – 2D plot of CaM-RNC in buffer supplemented with $Ca^{2+}$. The blue line represents the Gaussian fit of the histograms. Dashed lines represent individual Gaussian peaks for the case of two subpopulations.

#### 4.1.4.3 Alternatives to using CpK and AzF

We tried to increase the incorporation yield by changing the reaction to have only natural tRNA incorporation. In this approach, we used the $tRNA^{Met}$(AUG) with a Met-depleted PURE system to incorporate our unnatural amino acid site-

specifically. Instead of using Methionine, we supplemented the system with the methionine analog homopropargylglycine (HPG), which was used by the methionyl-tRNA synthetase as a substrate for aminoacylation of the $tRNA^{Met}(AUG)$[87].

We compared the incorporation of the unnatural amino acid HPG to the incorporation of CpK/AzF indirectly with in-gel fluorescence. Since the labeling chemistries were different, they could have an impact on the in-gel results (this was not taken into consideration for the analysis of the data). The incorporation of the unnatural amino acid AzF was not detectable by in-gel fluorescence, presumably due to low reaction yield (Figure 4.13 A). Surprisingly, the reaction without messenger RNA (mRNA) (negative control) showed three fluorescent bands (Figure 4.13 A, lane 3). The lowest of them came from the labeled tRNA, but the two upper ones seemed to be unspecific binding of the dye to other components of the reaction mix, even though click chemistry should be specific[153]. These three bands were also present in the reactions with the two constructs (Figure 4.13 A, lane 1 and 2). The expected size of the yPGK-SecMstr construct is  70 kDa, and it is clearly absent in the gel (Figure 4.13 A, lane 1). However, the yPGK(1-160)-SecMstr construct seems to give a band at the expected size of  35 kDa. This band overlaps in part with the upper unspecific band, but it is clearly lower (Figure 4.13 A, lane 2). This supports the results discussed in section 4.1.4.2 that shorter constructs have better chances of being synthesized successfully.

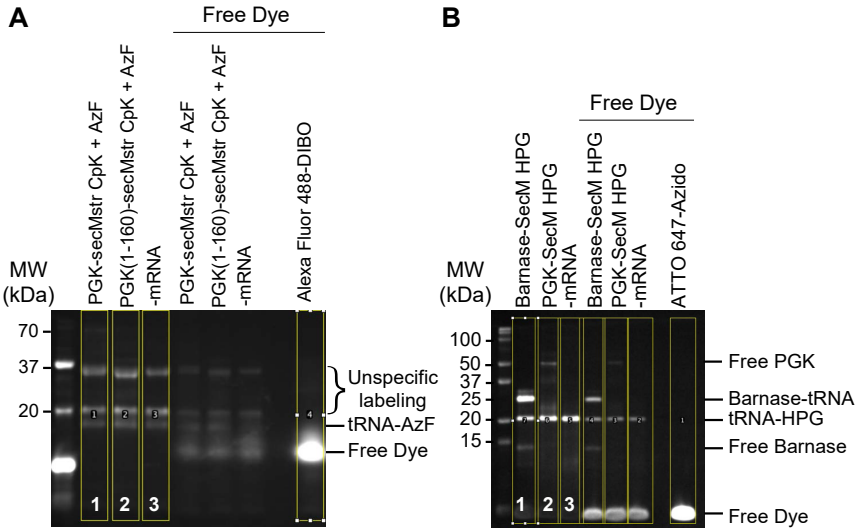



**Figure 4.13:** Comparison of incorporation of HPG with AzF and CpK. A – SDS gel of yPGK-RNC complexes labeled with copper-free click chemistry. B – SDS gel of protein-RNC complexes labeled with copper click chemistry.

Nevertheless, the incorporation of unnatural amino acids via synthetic tRNAs showed suboptimal results compared to the natural $tRNA^{Met}$(AUG). On the other hand, the incorporation of natural HPG-$tRNA^{Met}$ worked as expected. The negative control reaction showed fluorescence coming only from the free tRNA (Figure 4.13 B, lane 3). As expected, in the reactions with mRNA, an additional fluorescent product was observed, namely Barnase and yPGK, respectively (Figure 4.13 B, lane 1 and 2). The secM did not efficiently stall the proteins in the ribosomes since part of barnase, and the complete yPGK seemed to be released (free barnase and free yPGK) (Figure 4.13 B, lane 1 and 2).

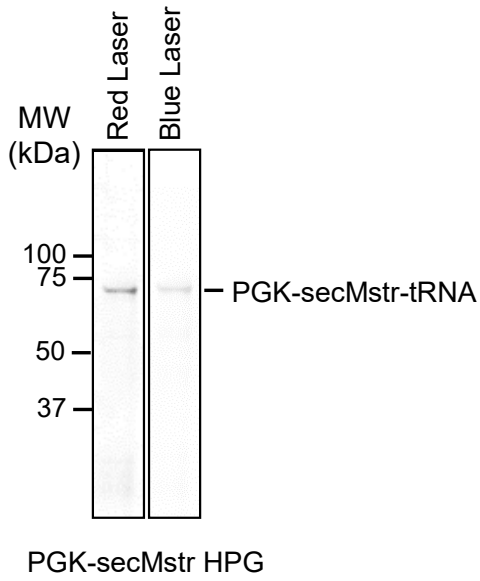



**Figure 4.14:** Labeling of yPGK-SecMstr with Alexa flour dyes conjugated with azido. SDS gel of yPGK-RNC labeled with copper click chemistry.

Reproducing the incorporation of HPG in yPGK using the stronger arresting sequence variant SecMstr resulted in the expected stalled yPGK migrating as a complex with the tRNA (Figure 4.14). Unfortunately, the FCS analysis of the sample showed a high amount of aggregates and no double incorporation. Due to difficulties in proceeding with the synthesis of yPGK-RNC with HPG incorporation, alternative strategies were considered. The main complication was the copper click-chemistry required to label the yPGK-RNC complexes that demanded special conditions and equipment. The reaction was performed in anaerobic conditions that could not be easily established in the lab. In the end, additional purification steps would be required to remove the N-terminus HPG. Otherwise, complexes would have three dyes. Considering the previous points, the synthesis with incorporation of HPG was abandoned.

### 4.1.5 Production of yPGK released truncations for cysteine labeling

As an alternative to the single-molecule FRET measurements with truncated nascent chains bound to the ribosome, we opted to synthesize released protein of the same yPGK truncations and labeling pairs (Figure 4.15). By having released nascent chains, it was possible to use the incorporation of cysteines at the labeling positions, which improved the reaction yield and simplified the labeling scheme.

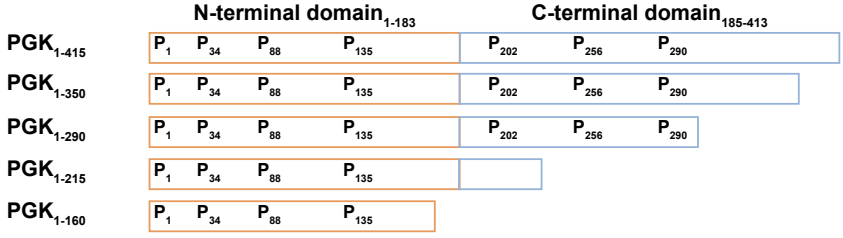



**Figure 4.15:** Labeling scheme of released yPGK truncations synthesized with a cell-free synthesis system. Possible labeling positions were represented in each truncation ($P_n$). Each yPGK variation was synthesized with a single labeling pair (P1–P88, P1–P135, P34–P135, P135–P290, P202–P256, P202–P290, P256–P290), covering the pairs present within each truncation sequence.

Proteins were synthesized with a cell-free system to be able to express a large number of constructs in parallel. Starting with the full-length, we confirmed that similar results to cell-based yPGK measurements[60] were obtained (Figure 4.16).

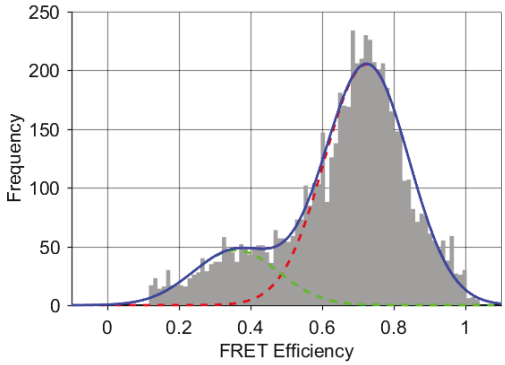



**Figure 4.16:** Single-molecule FRET (E) histogram of cell-free synthesis yPGK interdomain cysteine pair. The solid represents the sum of the Gaussian fits, and dashed lines represent the individual Gaussian fits.

When expressing the truncations, small variances were observed during the purification of the different FRET pairs, with the appearance of second lower band in the purified product (Appendix B). As an example of the purification products, a picture of the gel lane of the elution step of one label pair of each truncation was selected (Figure 4.17). The pictures of the elution lanes of each truncation had a single band with the correct size. A few labeling attempts were performed. However, due to time constraints, it was not possible to obtain sm-FRET histograms of the truncated variants in time for the submission of the thesis.

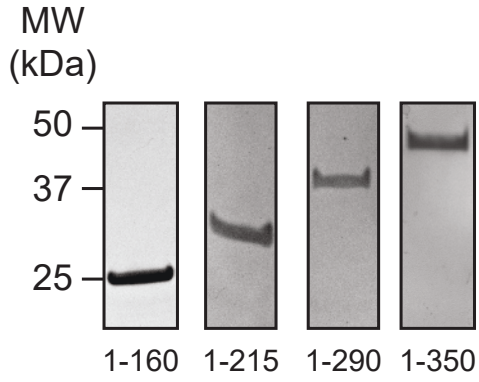



**Figure 4.17:** Synthesis of yPGK truncations with cell-free synthesis system. SDS gel of purified yPGK truncations synthesized to different lengths.

## 4.2 Cryo-EM

### 4.2.1 cryo-EM sample preparation and characterization

The structure of yPGK-RNC with different lengths was later analyzed by cryo-EM. In the first attempt, the construct prepared for single-molecule FRET measurements served as a template for the cryo-EM sample preparation. Initial results were unfavorable since we could not see any density of the nascent chain outside the ribosome. We hypothesized that the lack of density in our EM density map was due to the flexible emerging nascent chain blurring the density close to the exit tunnel outside the ribosome. Based on findings from Bushan et al.[45], where the authors showed that a sequence of five times EAAAK was able to form a secondary structure in the tunnel of the ribosome, especially in the tunnel's vestibule, we decided to change the long unstructured linker that we used (2GS) with the more structured rigid linker (RL). The goal was to reduce the mobility of the newly synthesized protein and restrain its position close to the exit tunnel.

Additionally, we optimized the expression time to maximize the RNC yield. After synthesizing yPGK-RNCs at five different time points, we developed a western blot using a mouse anti-yPGK antibody to detect the reaction products (Figure 4.18 A and B). The western blot showed three bands that we interpreted as different reaction products. The top band of each lane should correspond to a fully synthesized nascent chain bound to the tRNA (identified as RNC). The middle band should correspond to smaller incomplete nascent chains bound to the tRNA (polysomes) since ribosomes attached to the same mRNA initiate translation and do not finish it. The first ribosome is paused at the stalling position and blocks the remaining ribosomes in the mRNA. Finally, the unbound, fully synthesized nascent chain (released) should migrate faster than the previous species. Due to incomplete stalling, some ribosomes should be able to finish translation and release the nascent chain to later re-initiate translation in a different mRNA. The previous results evidence that a total of 60 min expression yielded the maximum amount of RNCs. Afterwards,

the stalling efficiency decreases; after a 120 min reaction, the release reaches the maximum. Therefore, from this point onwards, all reactions were incubated for 60 min to minimize nascent chain release.

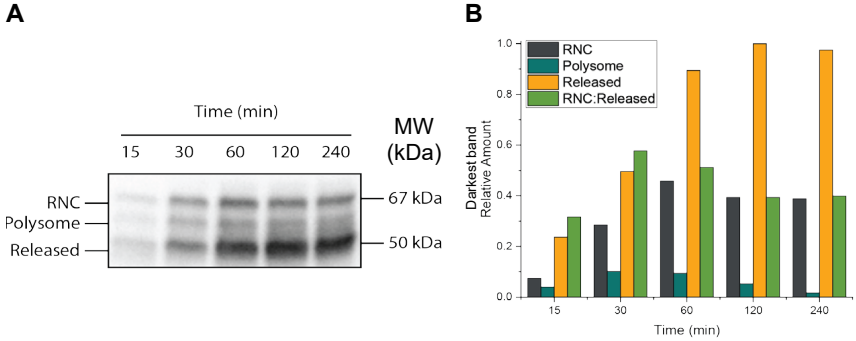



**Figure 4.18:** Time points expression of RNCs. A – Western blot of yPGK-RNC synthesized at different time points detected with anti yPGK-mAb. B – Densitometry analysis of western blot bands relative to the darkest band. RNC – full-length nascent chain bound to tRNA; Polysome – incomplete nascent chain bound to tRNA; Released – free full-length nascent chain; RNC:Released – ration between RNC and Released band.

We identified the different species by treating the reaction product with puromycin and RNase A. The products were detected by western blotting, similarly to the previous results. The upper and middle bands disappeared when the sample was treated with RNase A. Present RNA molecules were degraded while the bond between the nascent chain and tRNA was broken. Therefore, yPGK-positive species in the sample migrate as a single band according to the size of the released protein. Additionally, adding puromycin to the reaction evaluated the presence of polysomes. The fully synthesized nascent chains are puromycin-resistant, whereas incomplete nascent chains are sensitive to the antibiotic. Therefore, the sample treated with puromycin showed only two bands in the western blot, confirming the presence of polysomes (Figure 4.19).

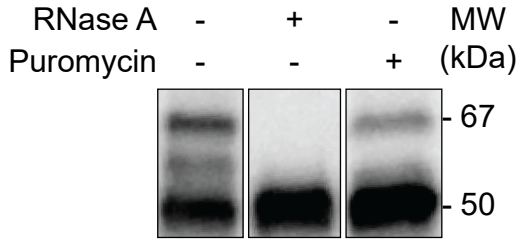



**Figure 4.19:** RNase A and puromycin treatment of RNC sample confirmed the presence of a nascent chain bound to tRNA. Western blot of yPGK-RNC treated with RNase A and puromycin. Bands were detected with anti yPGK-mAb.

#### 4.2.1.1 Biochemical optimization of sample for cryo-EM

We proceeded with further purification steps to increase sample homogeneity. One of the employed protocols was incubating the reaction mix with affinity tag resin. For the present work, strepII-tag was incorporated at the N-terminus of yPGK(1-184)-RL-secMstr for affinity purification. After the synthesis reaction, the sample was washed with buffer W to remove unbound molecules. Finally, the sample was eluted by supplementing buffer W with 50 mM biotin. To evaluate the success of the affinity chromatography method, we detected the sample by western blotting. Additional negative stain grids were prepared to inspect the presence of ribosomes and evaluate particle concentration (Figure 4.20). The data showed that most RNCs did not bind to the resin and were removed in the flow-through. Nevertheless, a band of yPGK(1-184)-RNC was visible in the western blot in both elution 1 and 2. After inspecting the images of the negative stain, it was possible to identify ribosomal particles with the correct size (20 nm) in the imaged fractions. Most ribosomes were present in elution 1 with good coverage of the field of view. Ultimately, this protocol was not followed since the concentration and volume of RNCs solution was insufficient for sample vitrification.

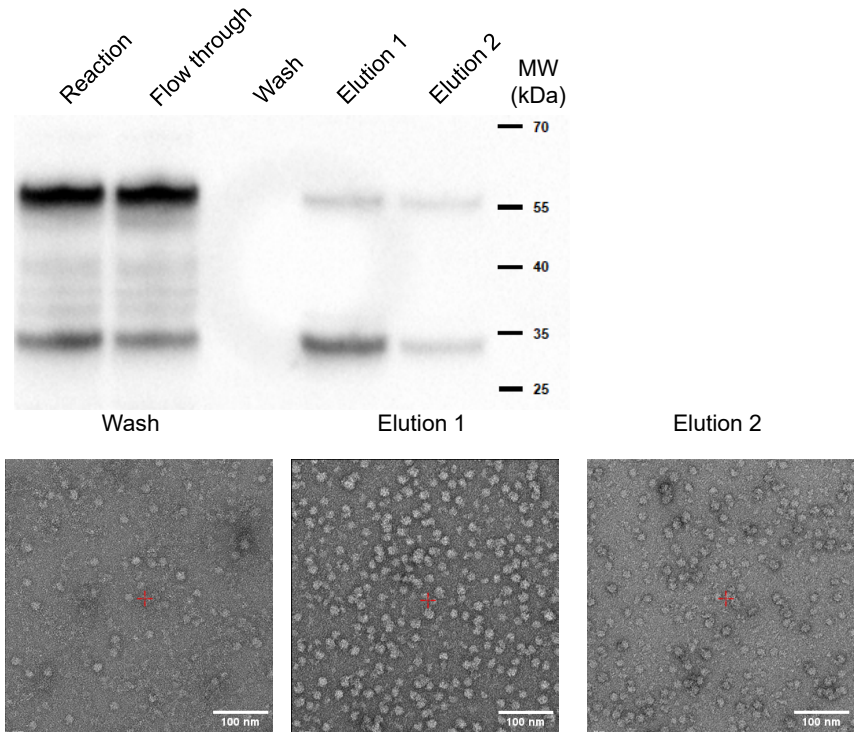



**Figure 4.20:** StrepII-tag affinity chromatography purification of yPGK-RNCs. Western blot of yPGK-RNC purification fractions detected with anti yPGK-mAb. Negative stain of purification fractions imaged in Talos 120 kV.

Considering the necessity of large amounts of RNCs, the reaction size was increased fifteen times, and a step of sucrose gradient purification was added to the protocol. The sucrose gradient step was introduced previously in a similar work by van der Stel and colleagues to solve the TnaC structure[154]. The sample was fractionated on a sucrose density gradient to separate the different molecular complexes in the reaction. We were interested in isolating yPGK-RNCs, which should be the largest complexes in the solution. Therefore, they should migrate the longest distance through the gradient (Polysomes area on Figure 4.21 A). The presence of yPGK in each fraction was tracked with a western blot, similar to the western blots described in section 4.2.1. Reaction and puromycin fractions were collected before density

gradient ultracentrifugation and fractionation. Both fractions detected the presence of RNCs and released protein. After puromycin treatment, the polysome bands in the reaction lane were not detected on the puromycin lane, confirming the action of puromycin described in section 4.2.1. After the first density gradient ultracentrifugation, the released protein remained in the top fraction (Figure 4.21 B). The 70S fraction did not detect the presence of any yPGK species. On the other hand, the polysome fraction detected the presence of yPGK species migrating with the size of RNCs, polysomes, and released. These results support that most ribosomes in the reaction were inactive. Only a small fraction of all ribosomes formed polysomes. Moreover, incomplete nascent chains still bound to ribosomes and released protein bands in the polysome lane were unexpected. Nevertheless, the sample was further processed. Afterwards, the polysomes were mixed with RNase H and a DNA oligo to cleave the mRNA and form monomers. Most polysomes were cleaved to 70S and contained yPGK-RNC.

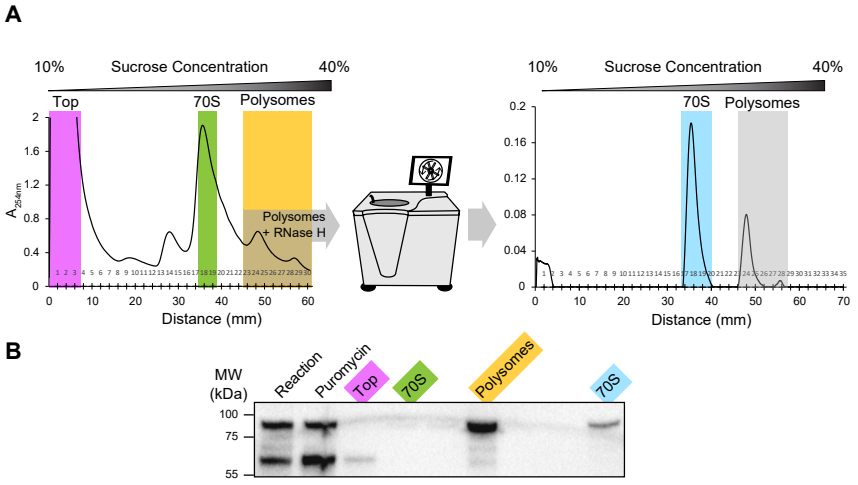



**Figure 4.21:** Sucrose gradient purification. A – Sucrose gradient profile of cell-free protein synthesis and the following polysome RNase H treated fraction. B – Western blot of the critical fraction of RNC purification detected with anti yPGK-mAb.

In line with the discussion in section 4.2.1, one of the limiting factors in resolving the structures of ribosome nascent chain complexes is their flexibility, i.e., the relative

mobility of the nascent chain to the 70S ribosomes. Therefore, chemical fixation can make the molecular complex more rigid and stable. Instead of using a normal sucrose gradient, the sample was passed through a sucrose/glutaraldehyde gradient (GraFix), first described by Stark and colleagues[140]. The presence of glutaraldehyde creates covalent bonds between molecules nearby, stabilizing the molecular complex[140].

We tested GraFix on the yPGK truncation 1-184, expecting improvement in the density of the nascent chain outside the tunnel of the ribosome. The sample was passed through GraFix after collecting the polysome fraction. GraFix stabilized larger molecular complexes from incomplete RNase H digestion (Figure 4.22). The curve of the standard density gradient had a large peak of the 70S, a small peak of disomes, and a tiny peak of trisomes. On the other hand, the curve from the GraFix density gradient showed a broader disome peak, a larger trisome peak, and an additional tetrasome peak.

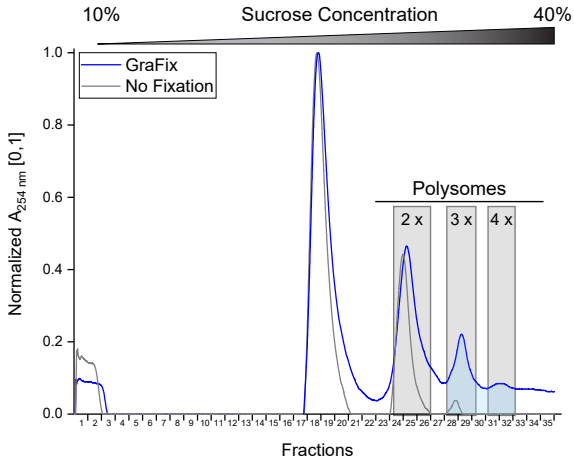



**Figure 4.22:** Sample preparation of ribosome nascent chain complexes cross-linked by GraFix. Normalized sucrose gradient profiles of the polysome RNase H treated fraction in the presence of GraFix (blue line) and without fixation agent (gray line).

### 4.2.2 Analysis of cryo-EM data

After optimizing the biochemistry of cryo-EM samples, I determined the cryo-EM structure of yPGK-RNC complexes. Cryo-EM samples were prepared according to the optimized sucrose gradient protocol without GraFix described in section 4.2.1.1. Results from samples treated with GraFix were similar to those of samples without. Several methods of vitrification were followed to increase the presence of RNC molecules in the holes of the grids. The most relevant data sets for the present work were presented in Table 4.4. Briefly, we tested data sets of the full-length yPGK with different buffer conditions, such as adding RNasin® inhibitor and chloramphenicol[155]. Moreover, efforts were made to collect data sets of additional nascent chain truncation, such as yPGK 1-215 and yPGK 1-290.

**Table 4.4:** Cryo-EM data acquisition of ribosome nascent chain complexes sample's final collection. APE – density present in tRNA cavities of A-site, P-site, and E-site simultaneously; NC – density present in the nascent chain outside the exit tunnel.

| | **yPGK (1-415)-RNC** | **yPGK (1-363)-RNC** | **yPGK (1-350)-RNC** | **yPGK (1-184)-RNC** |
|---|---|---|---|---|
| Data Collection | | | | |
| Microscope | Talos Arctica | Talos Arctica | Talos Arctica | Titan Krios |
| Camera | Gatan K3 | Gatan K3 | Gatan K3 | Falcon 4 |
| Magnification | 100000 | 100000 | 100000 | 96000 |
| Voltage (kV) | 200 | 200 | 200 | 300 |
| Electron dose ($e^-$/Å$^2$) | 50 | 50 | 50 | 50 |
| Defocus range (μm) | -2.5 to -0.5 | -2.5 to -0.5 | -2.5 to -0.5 | -2.5 to -0.5 |
| Pixel size (Å) | 0.816 | 0.816 | 0.816 | 0.808 |
| Initial particles (no.) | 531123 | 1138208 | 849076 | 992349 |
| Particles APE (no.) | 71097 | 151354 | 88104 | 117537 |
| Final particles NC (no.) | 37746 | | 28359 | 44112 |

All data sets present in Table 4.4 were collected overnight, with some requiring longer collection times due to variable grid quality. Generally, data sets with half a million to a million particles can reach final 3D reconstructions with global resolution below 4.0 Å. However, in the present work, only 3–7 % were actual particles targeted to study co-translational folding. Therefore, from the initial pool of particles, the data sets ended up with 3D reconstructions with 30,000–44,000 particles. Additional data collection could contribute to 3D reconstructions with a larger number of particles.

The sample vitrification method that showed the best results was using the vitrobot with 5 µL of sample at a concentration of 0.6 mg/mL, including a waiting time of 30 s before plunge-freezing. This method produced micrographs with good particle distribution, as shown in Figure 4.23. Because during the vitrification method, the grid is quickly submerged in liquid ethane, it is possible to see some contamination particles of ethane droplets (dark homogeneous black circles). Most contaminants were removed from the final reconstruction through the classification algorithms at the 2D and 3D levels. The initial ribosome reconstructions of the data sets collected had global resolutions of 2.5 to 2.8 Å.

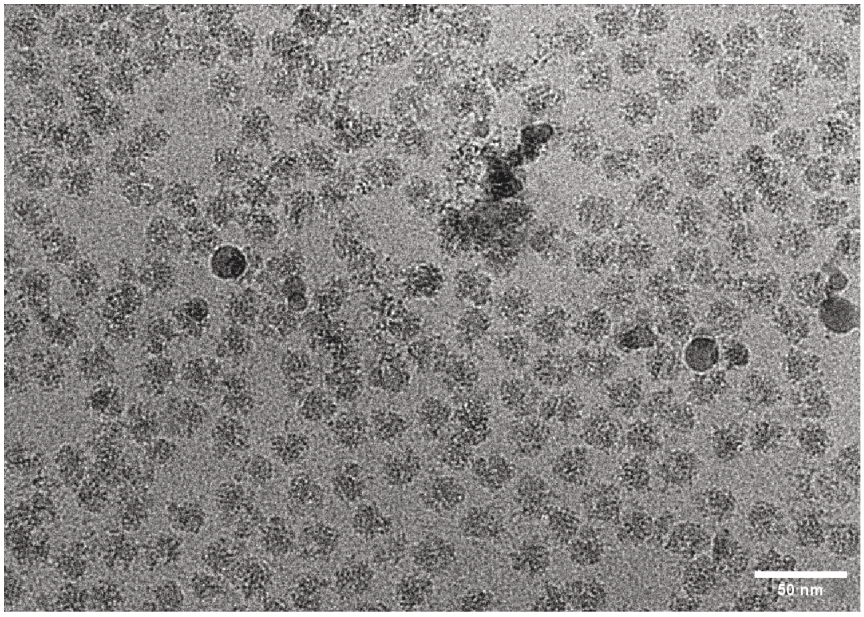



**Figure 4.23:** Particle density in motion corrected micrograph of RNCs sample.

Ribosomal particles had an approximate diameter of 20 nm, in agreement with previous structures of the bacterial ribosome[143]. Due to the high content in phosphate groups from all RNA molecules, ribosomes displayed higher contrast than simple protein complexes and facilitated the particle-picking job for downstream processing. Most micrographs across all data sets collected had a CTF fit ranging from 2–5

Å resolution with some outliers due to patches of crystalline ice or broken squares (Figure 4.24 A, B, and C).

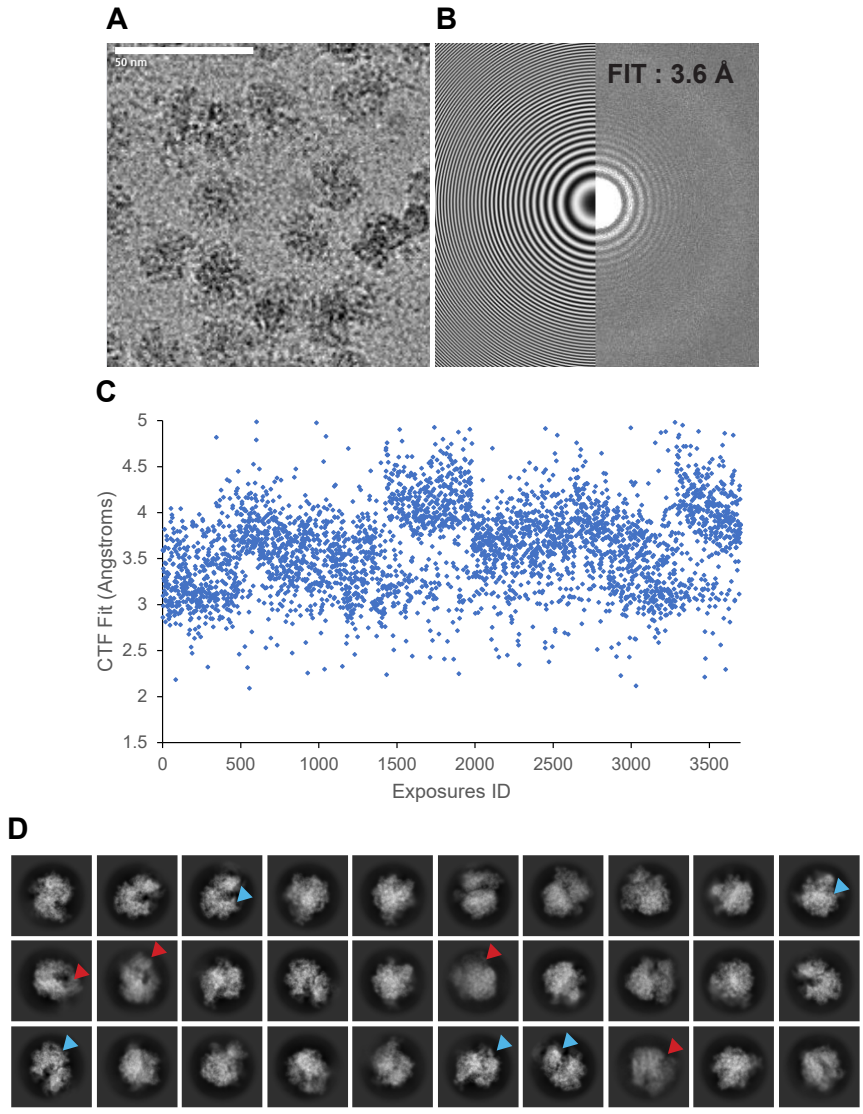



**Figure 4.24:** Data processing workflow at 2D. A – Cropped area of motion corrected micrograph of RNCs sample. B – CryoSPARC patch CTF estimation 2D fit on the left. Power-spectrum of dose-weighted motion corrected full micrograph (A) on the right. C – Variation of CTF estimation along all micrographs from a full data set. Micrographs with CTF fit above 5 Å are not displayed. D – Accepted class averages from the 2D classification job of CryoSPARC. Blue arrows mark high-resolution classes; red arrows mark low-resolution classes.

After particle extraction, unidentified particles were removed by 2D classification. A typical output of accepted classes from the 2D classification job is present in Figure 4.24 D. In the example, a few high-resolution classes are marked with blue arrows, and a few low-resolution classes are marked with red arrows. The classes with lower resolution result from a higher degree of uncertainty in the assignment of particles to those classes by the algorithm.

The initial 3D reconstruction of selected 2D classes was further classified at the 3D level. Initial heterogeneous refinement using four times the refined structured as starting volumes pooled 20 % of the particles in a damaged category, 52 % of the particles with a nascent chain, and 28 % of particles without density in the tRNA sites, classified as an empty category. $\approx$ 270,000 particles with a nascent chain density were refined to obtain a consensus 3D reconstruction. From the newly aligned particles, the signal of the large and small subunits was removed using the job particle subtraction from cryoSPARC. The particles should have signal for the tRNAs cavities, tunnel inside 50S, and exit tunnel outside the ribosome. The particles were classified with a 3D classification job without alignment in 7 different classes. The 3D classification pooled 8 % of the particles in a class with E- and P-tRNA site density, 14 % in a P- and A-tRNA site density class, 16 % in a class of P-tRNA site density, 8 % with E-tRNA site density, 28 % in an Empty tRNAs density class, 26 % with 3 tRNAs density class (Table 4.5).

**Table 4.5:** Cryo-EM data of density maps from 3D classification results. The percentage of particles of each class in the whole data set is presented in square brackets.

| Class | Damaged | Empty | P/A | P | E/P/A | E/P | E |
|---|---|---|---|---|---|---|---|
| Particle number [%] | 64,165 [20 %] | 21,106 [43 %] | 38,943 [7 %] | 44,304 [8 %] | 71,097 [14 %] | 22,402 [4 %] | 22,647 [4 %] |
| Resolution (Å) | | 3.2 | 3.0 | 3.0 | 2.9 | 3.1 | 3.1 |

Analyzing the particles classified in both 3D classification jobs (heterogeneous refinement and 3D classification without alignment), I observed that most ribosomes were empty (43 %). Additionally, particles classified as damaged were discarded since their contribution would be limited to low-resolution information. The particles of interest were those with density present in the 3 tRNAs cavities, which also had the best density of the nascent chain. In the initial reconstruction from the 3-

tRNAs density class, it was not possible to observe the density of the nascent chain in the sharpened map. To analyze the shape of the nascent chain, the map needed to be inspected unsharpened or lowpass filtered to 12 Å. The global resolution of the map was 2.9 Å, and it was possible to look at the density close to the PTC, which was important for later confirming that the particles were indeed yPGK-RNCs.

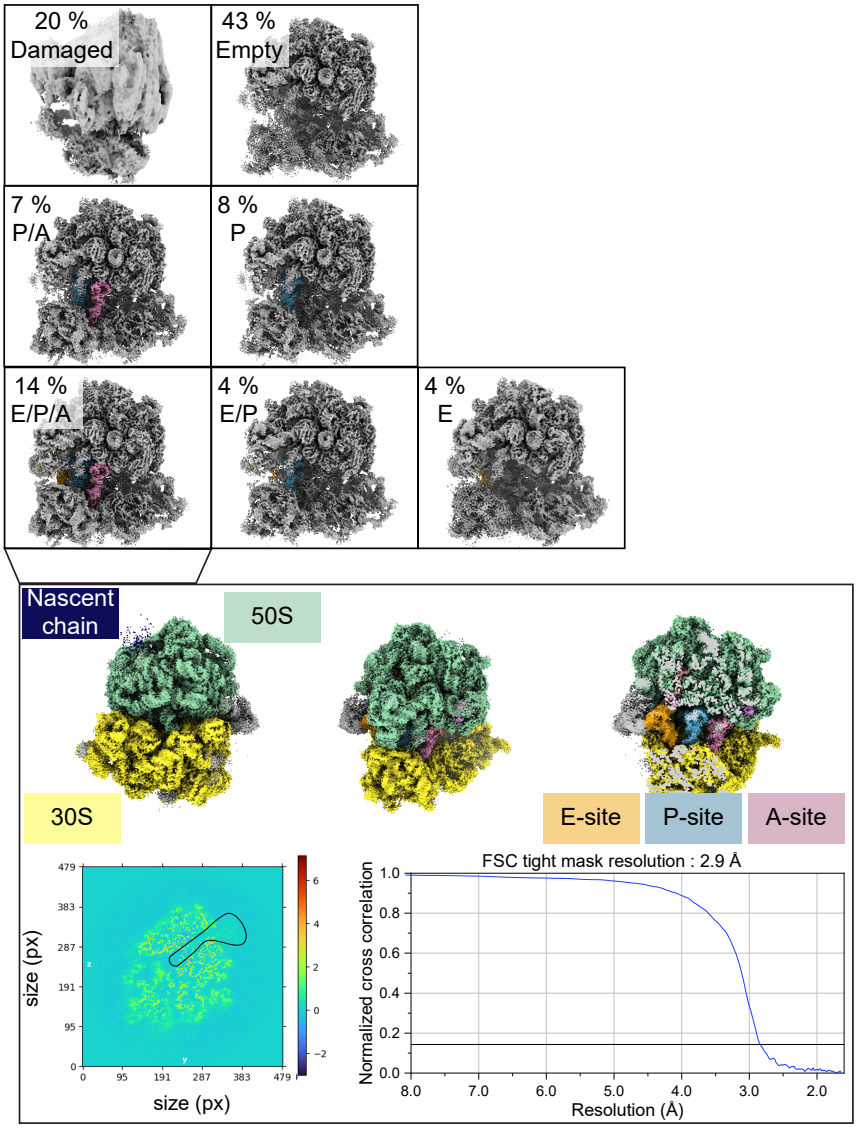



**Figure 4.25:** 3D volumes of different classes of ribosomes present in the yPGK(full-length)-RNC data set. The percentage of each class refers to the total number of particles of each class present in all 3D jobs compared to the initial particle number of the data set (it means that it is not strictly necessary that the number of particles present in the reconstruction of each volume matches the overall percentage of the class). The most important features of the RNCs maps were labeled properly. The occupancy of the tRNA sites was colored orange (E-site), blue (P-site), and pink (A-site). The large and small subunits were colored green and yellow, respectively. The FSC curve of the map E/P/A was displayed to depict the resolution obtained.

Finally, we used a local refinement job from CryoSPARC with a mask outside the exit tunnel with the shape of the yPGK to maximize the local resolution. The local refinement job improved the resolution across the yPGK map density region. Initially, the local resolution was $> 11$ Å in the C-terminal of yPGK map density. The resolution increased, especially in the C-terminal, to values $\approx 8$ Å (Figure 4.26). We could visualize the $\alpha$-helices as cylinders, and the density covered the whole crystal structure. The N-terminal domain was better resolved than the C-terminal domain since it was closer to the ribosome surface.

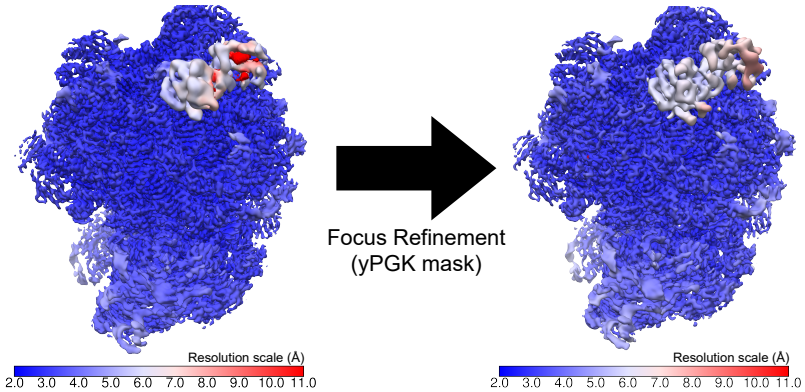



**Figure 4.26:** Local resolution maps highlighting improvement in the density of the nascent chain. The maps were colored according to the estimated local resolution from high resolution (blue) to low resolution (red).

When the map after 3D classification was inspected, it was possible to rigid body fit the yPGK crystal structure to the region outside the exit tunnel. The region covered most of the 3D model and had a shape resembling the yPGK structure. After obtaining the map with improved local resolution, the shape of the density map was tightly connected to the crystal structure. At the same time, it was possible to observe an extra density expanding the C-terminal towards the ribosomal tunnel. Due to limited resolution, it was impossible to resolve side chains of the nascent chain to investigate atomic interactions with the ribosomal surface (Figure 4.27).

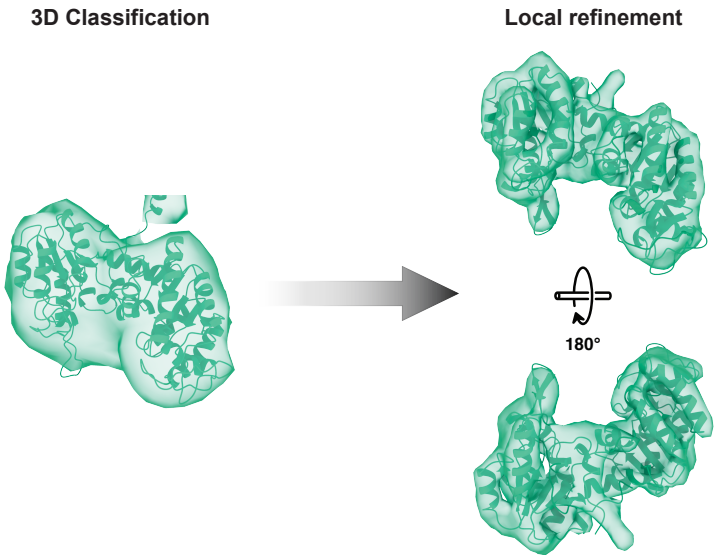



**Figure 4.27:** Local refinement shows native fold of full-length yPGK-RNC. PDB 1qpg rigid body fitted in chimera to the density of RNC particles outside the tunnel.

#### 4.2.2.1 Proof of stalling with secMstr

It was important to confirm that the structures resulted from true RNCs molecules. Since we could not solve the RNCs at high resolution outside the tunnel, I decided to look at the PTC site of the yPGK(full-length)-RNC data set. The PTC sits in the core of the ribosome, having, therefore, a strong signal. The density was resolved at $\approx 2$ Å. We could build the atomic model of the stalling sequence with a high degree of confidence up to the first 11 residues (Figure 4.28). Since we were using different arresting peptides, for which the stalling mechanism was not known, we tried to build the atomic model starting from different protein positions in the P-site. The last amino acid of the secMstr was designated position 0 at the P-site. The amino acids closer to the N-terminus were designated from -1 up to -5. In all maps, there was a clear density present at the A-site. Therefore, we built the atomic models starting from position -1. The model generated from position -1 was clearly the best-fitting model. Inspecting the map-model cross-correlation per position (Figure

4.28 A), up to the $11^{th}$ residue, the cross-correlation value was close to 0.8. All other sequences had lower cross-correlation values, with position -3 being the closest to position -1. Figure 4.28 B showed that correlation position -1 had the distribution with the smallest standard deviation and position -5 had the lowest mean value of cross-correlation. The cross-correlation values for each sequence were corroborated by the analysis of the sequences in the density. Figure 4.28 C black arrows showed the places where the model had low cross-correlation with the map. Position -1 map/model figure showed high agreement for most side chains, giving us strong support for the presence of RNCs.

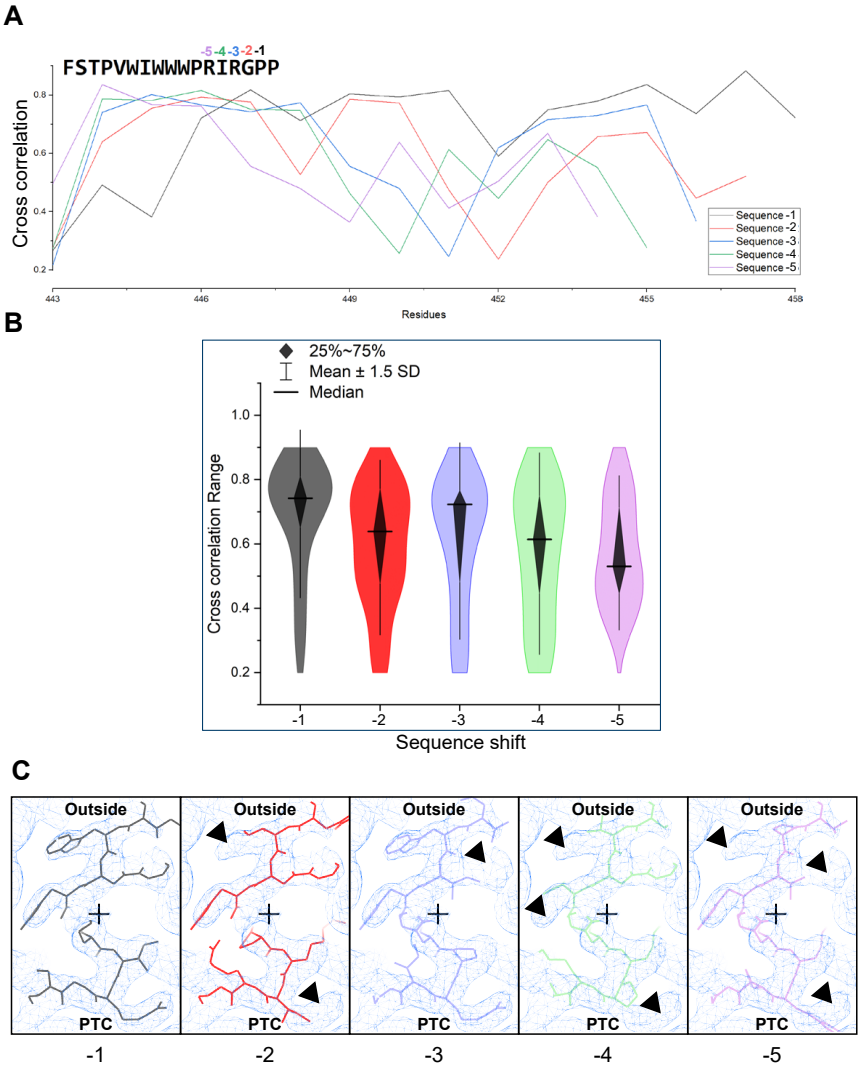



**Figure 4.28:** Cross-correlation analysis of sequence assignment in the tunnel starting in different starting points. A – Comparison of cross-correlation and atomic model in the 50S tunnel close to the PTC. B – Distribution of cross-correlation values in all positions for each sequence assigned in the 50S tunnel close to the PTC. C – Image of density and atomic model of all sequences in a part of the 50S tunnel close to the PTC.

In addition to the PTC site, we validated the presence of RNCs by inspecting the mRNA density (Figure 4.29). We could observe clear density at all tRNA sites, and

we could verify the presence of either pyrimidine or purine bases. Even if the density was not clear for the third position of the tRNA E-site, we had the RNA sequence of the construct, which helped assign the correct base to the respective density. The sequence assignment matched with the sequence present in the PTC. At the E-site, we assigned the codon GGC that codes for glycine, and the remaining P and A sites were occupied by proline codons (CCC and CCU, respectively).

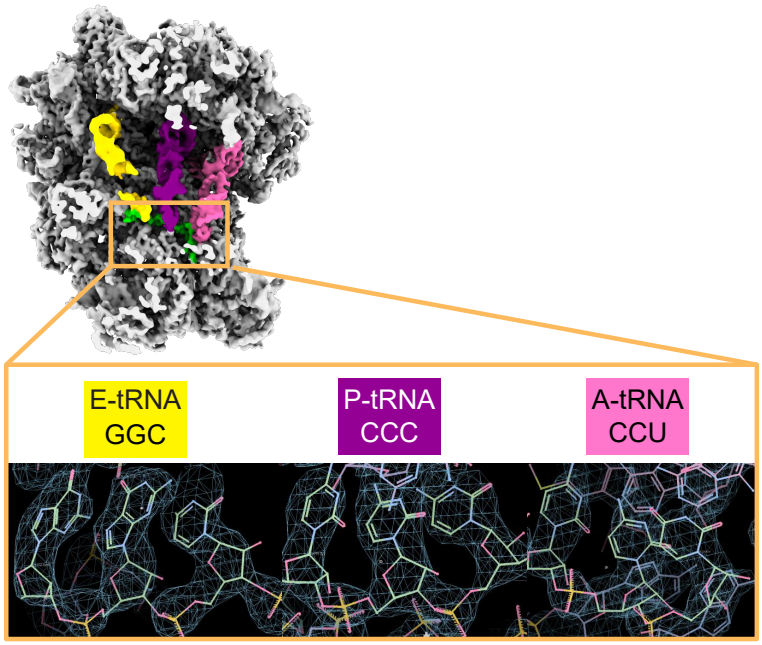



**Figure 4.29:** Sequence assignment in the mRNA density. 70S ribosome in grey, E-site in yellow, P-site in purple, A-site in pink, mRNA in green. RNA sequence assigned to density of mRNA.

#### 4.2.2.2 Heterogeneity of protein translation with stalling

The translation process is intrinsically dynamic. Due to limitations of the purification steps during sample preparation, we froze a large percentage of ribosomes that were not in the desired state. We performed several rounds of 3D classification jobs to sort all these particles out, such as heterogeneous refinement and 3D

classification without alignment in CryoSPARC. We tested jobs with different sets of classes and tried to optimize the key parameters of each job as much as possible. Overall, data sets from yPGK-RNC truncations could be separated into seven distinct classes (Figure 4.30). The most populated class was the empty 70S, reaching up to 65 % of all ribosomes present in the data set. In the literature, most RNCs were sorted from P-only ribosomes[43,154,155]. Interestingly, the 3D reconstructions of this p-class in our data sets did not show any density in the tunnel or outside the ribosome. In our data sets, only particles with P/A and all tRNAs (E/P/A) had density either in the tunnel or outside the ribosome instead. This property might be due to the different arresting peptide used in the present work. Most papers used secM[155] or TnaC[154] arresting peptides. The enhanced stalling sequence used in our project (secMstr) probably has a different stalling mechanism, resulting in a distinct arresting ribosome state. Since only the yPGK(full-length)-RNC showed density outside the tunnel in the P/A state, we did not process this class further. All data sets had density in the tunnel for the E/P/A state. This class represented only 10 % of the ribosomal particles, limiting the achievable resolution.

These results demonstrated how dynamic the translation process is and how rarely we could find our target particles. Evidently, the rare number of RNCs with visible nascent chains was one of the major limitations of the project, and it was the stage where we spent a considerable amount of time trying to optimize and improve the sample.

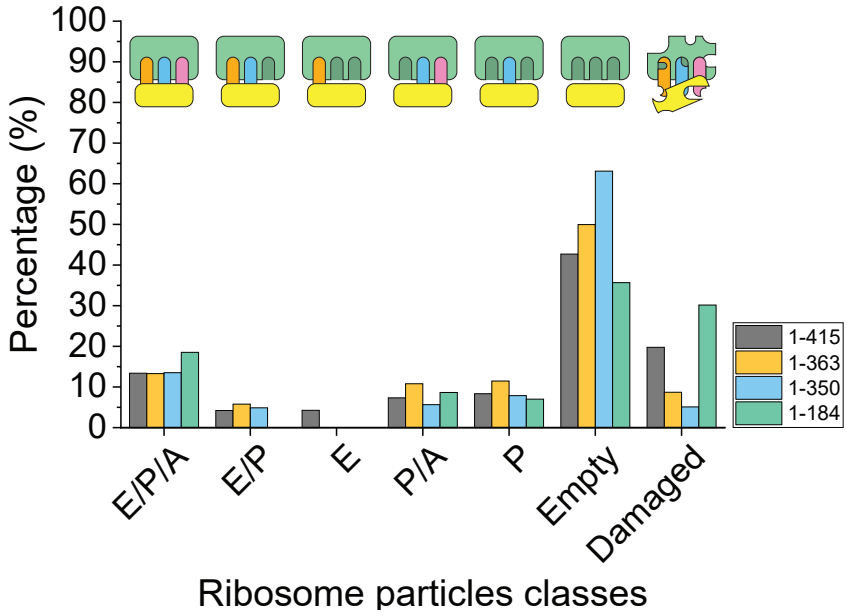



**Figure 4.30:** Particle distribution of translating ribosomes. Analysis of classes of particles present in 3D classification in all RNCs samples. Illustration of different classes shows the occupancy of the tRNA sites and particle quality.

#### 4.2.2.3 Differences between yPGK truncations

Inspecting the class of E/P/A from multiple data sets from 3D classification, we could see the density of all tRNAs (E – orange, P – blue, A – pink) and nascent chain (green). The density was stronger for the tRNA sites than for the nascent chain. From all the data sets presented, only the truncation of yPGK 1–363 did not show any density outside the tunnel (Figure 4.31). The density outside the tunnel that we observed for the remaining data sets was not completely clear. Nevertheless, after fitting the different structure models of yPGK (full-length – 1qpg.pdb; 1–350 – Alphafold prediction; 1–184 – Alphafold prediction) in the corresponding densities, it was clear that the density outside the tunnel of both truncations has approximately the size of the predicted yPGK(1–184) structure. Intriguingly, the remaining sequence of the yPGK(1–350) construct had no extra density covering

the amino acids of the C-terminal domain.

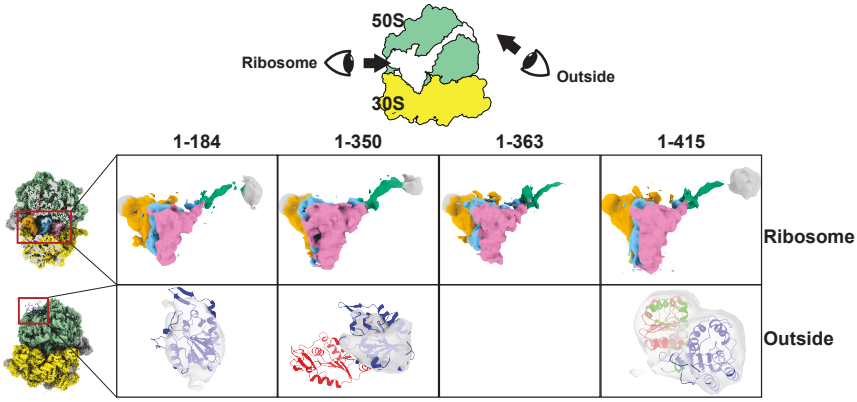



**Figure 4.31:** 3D classification shows low-resolution nascent chains outside the tunnel. Focus 3D classification with signal subtraction of different yPGK-RNC truncations. Colored in orange is the E-site tRNA, in blue is the P-site tRNA, in pink is the A-site tRNA, in green is the nascent chain inside the tunnel, and in gray is the nascent chain outside the tunnel. The Alphafold prediction of yPGK(1–184) and yPGK(1–350) were manually fitted to the density outside the tunnel of the corresponding maps. PDB 1qpg was rigid body fitted to the PGK(1–415)-RNC density outside the tunnel. PDB models were colored-coded according to the residues of each truncation: 1–184 – dark blue; 185–350 – red; 351–415 – lime.

We hypothesized that when yPGK was fully synthesized, stabilizing interactions with the ribosome were formed, allowing the formation of a clear density of yPGK outside the tunnel. Moreover, the synthesis of the full C-terminal domain sequence seems to be fundamental for its proper folding. The only time we observed any density for the C-terminal domain was in the full-length construct. The synthesis of the full C-domain sequence is likely required due to the contacts established between the C-terminus (407–415) and N-terminus (4–6) (Figure 4.32).

The yPGK truncations were not stabilized by the ribosome, but instead remained flexible exhibiting more freedom outside the tunnel, which resulted in the poorly resolved observed density. In order to validate the observed density outside the tunnel and exclude image processing artifacts, we tested the same approach for the class with empty ribosomes and, we could not observe any density after signal subtraction, corroborating our conclusions.

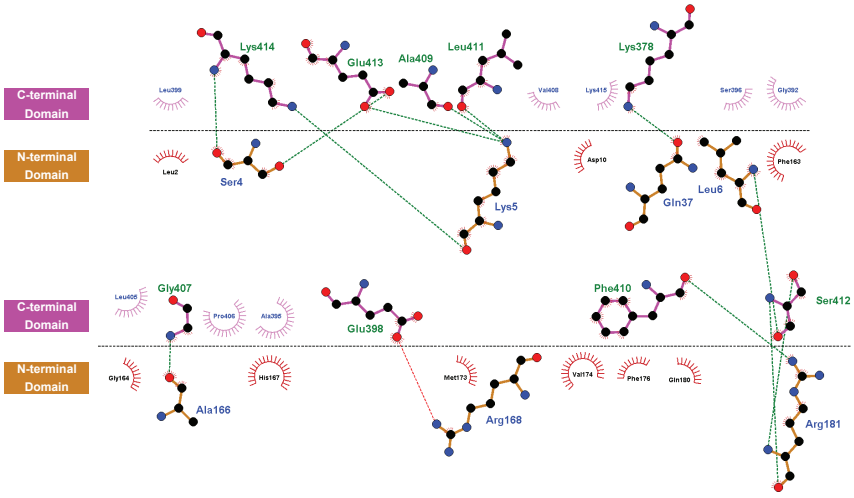



**Figure 4.32:** Depiction of protein-protein interactions between N-terminal domain and C-terminal domain of yPGK. The plot was generated with LigPlot+, with the spoked arcs representing residues making non bonded contacts, and dotted green lines representing hydrogen bonds.

After sorting the particles with density outside the tunnel, we proceeded with a homogeneous refinement with per-group CTF and per-particle defocus optimization to maximize the map resolution. Inspecting the sharpened maps (Figure 4.33), only the density of yPGK full-length was visible (density colored in dark blue). Although we could visualize density outside of the tunnel for the remaining truncations in the 3D classification job, these were not strong enough to be aligned in the homogeneous refinement job. To be able to visualize the density outside the tunnel, we lowpass filtered the maps to 20 Å resolution. The slice view of each truncation revealed the presence of a weak signal for the nascent chain along the tunnel and outside the tunnel, as well as marked inside the black area. The slice panels showed the strongest signal of the nascent chain at the full-length. Interestingly, the shortest truncation yPGK(1–184) had a stronger signal than the truncation yPGK(1–350) outside the tunnel. We hypothesized that this observation came from the fact that the 1–350 construct had a longer C-terminal unstructured region outside the tunnel, increasing its flexibility and decreasing the signal intensity. Altogether, we could visualize the full protein with native structure while attached to the ribosome. We could not

determine at which point yPGK adopts a native form. However, we could observe some degree of structure formation of smaller truncation. If yPGK truncation were completely unstructured, we could not observe any density, even at a low-resolution level.

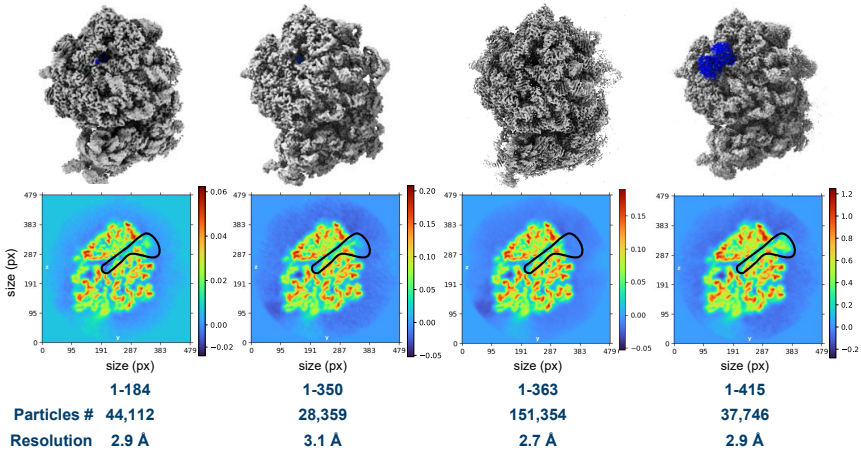



**Figure 4.33:** Final 3D volumes after data processing. Sharpened 3D volumes of homogeneous refinements with respective 2D slices with heat maps of arbitrary units of signal intensity on the bottom. Maps represented on the real space slices were lowpass filtered to 20 Å. The nascent chain region is highlighted by a black contour. 3D volumes of nascent chains are colored in dark blue, and ribosomes are colored in grey.

One possible solution to resolve better the nascent chain density of intermediate structures of yPGK during translation, could be the addition of chaperones to the purified RNC complex of yPGK truncations. Kaixian, L. et al.[78] demonstrated that the trigger factor reduced the misfolding of the inter-domain of elongation factor G during translation, enabling the correct folding of the N-terminal G-domain. The authors observed that the trigger factor directly protected the already-folded domain from destabilizing the newly emerging peptide chain. Reports of a holdase function of ribosome surface[34,49] might also be related to the destabilization of incomplete yPGK nascent chain outside the exit tunnel. Similarly to the elongation factor G, the addition of the trigger factor could stabilize the truncation samples of yPGK, revealing their structure features with higher resolution and helping in our understanding of the co-translational folding process.

Further studies of incomplete protein synthesis of EF-G sequence coupled with arrested peptide secM and a reporter fluorescent tag revealed important properties of co-translation folding[156]. In that work, the authors showed that mechanical forces from folding the first domain pulled the nascent chain and released the emerging protein. The authors presented identical data of *E. coli* PGK (ecPGK) (unpublished data). Even though it seems that chaperones were not necessary for the folding of ecPGK N-terminal domain, the yeast homologous protein is from a different organism. We cannot rule out that the two proteins have different folding pathways. Similarly, it is possible that the trigger factor could not help in stabilizing the truncation nascent chains of yPGK close to the ribosome surface. Both proteins are from different organisms. Hence, the trigger factor might have a system-specific activity. To correctly check the effect of chaperones on the co-translational folding of yPGK, one should use a yeast translation system including yeast chaperones.

# Chapter 5

# Conclusion and Future Perspectives

In the present thesis, an exhaustive study was conducted with yPGK. First, the presented results we focused on the comprehension of folding properties of released protein and addressed later a few approaches to investigate folding dynamics and structural characteristics of yPGK-RNCs complexes. The measurements of single-molecule FRET with released yPGK extended our understanding of the folding mechanisms of the individual domains of yPGK, particularly of the Rossmann fold motif. Two new FRET pairs with targeted distances within the N-terminal domain were synthesized and characterized with sm-FRET. Together with previously published data, three signature folding transitions were observed with a set of six pairs, three in each domain. We characterized and discussed the properties of the observed transitions, particularly the compact intermediate. Both domains presented similar transitions due to their close structural homology, which is in agreement with previous observations[60]. Previous sm-FRET studies mapping multiple distances in other proteins with different domain topologies that that of yPGK identified additional unfolding transition categories. For example, the cytolytic toxin ClyA (a helical monomer with 34 kDa) revealed five different unfolding transitions when exposed to GuHCl[157], two of them compact intermediates. Another study with apoflavodoxin (20 kDa) with $\alpha/\beta$ topology, similar to yPGK domain topology, revealed also a compact intermediate unfolding transition[158]. These studies present strong evidences that the particular unfolding transition type: compact intermediate, can be observed in multiple proteins with singular domain topologies. These studies provide clear evidence that the particular unfolding transition type, the compact intermediate, can be observed in various proteins with a single domain topology. In the future, studies of isolated single domains yPGK should be conducted to evaluate the effect on the neighbor domain in the unfolding transitions. Do the domains behave the same way in both, the isolated and the multi-domain context? Understanding the impact of the neighboring domain could improve our understanding of

the full-length protein folding and whether or how inter-domain interactions control the folding of individual domains.

The major goal of the thesis was to measure the folding of yPGK bound to the ribosome using sm-FRET and cryo-EM. In order to measure yPGK-RNC complexes in sm-FRET, it was necessary to incorporate unnatural amino acids. A first attempt to synthesize yPGK-RNC was successfully achieved. However, the labeling of these products was not sufficient to measure single-molecule FRET. Since the sm-FRET measurements were challenging in yPGK-RNC context, the production of released yPGK truncations using cell-free synthesis was employed as an alternative to compare sm-FRET results with those from cryo-EM. The study of yPGK-RNC complexes using cryo-EM single particle analysis produced interesting results. Developments in the field allowed us to disentangle multiple states of protein translating complexes. Most of the purified ribosomes complexes were in states where no nascent chain was present. Interestingly, the class of ribosomes with nascent chains was in a state that was never described for arrested ribosomes. In most studies with arresting peptides, ribosomes are stalled in the P-only state, whereas our ribosomes were stalled in a E/P/A-state, suggesting a different stalling mechanism of the arresting peptide used in our studies. Moreover, for the first time, we could reconstruct a ribosome with a full-length nascent chain in the native state still attached to it. When the same approach was applied for the truncation variants, the density could not be resolved with the same resolution as compared to that of the full-length yPGK. The truncated protein variants are likely more flexible, limiting the resolution obtained for the nascent chain outside the tunnel. This observation also suggests that the full polypeptide sequence of the C-terminal domain is required to form stabilizing interactions within the fold of the yPGK structure.

Information obtained from sm-FRET measurements of released yPGK truncations could support the results obtained by cryo-EM. In principle, the results of sm-FRET and cryo-EM are complementary. With single-molecule FRET, we measure not only single distances between two reference points, but instead we make use of multiple reference points to obtain as much information as possible about distance changes at specific folding states. In cryo-EM, we record endpoint pictures of translation

complexes that resolve the nascent chain with limited resolution. The lack of high resolution information might come from three different scenarios. First, the protein might fold properly but move as a whole, i.e. as a rigid body against the ribosome. Second, the protein might be fixed against the ribosome but exhibits internal flexibility, resulting from incomplete folding. Finally, it can be a combination of both scenarios. With successful sm-FRET measurements, we would be able to address this limitation of cryo-EM results. By calculating the distance between the different reference points in the truncation variants, we can understand if they have values similar to the native structure. This would indicate whether the nascent chain is folded or not. Moreover, it would be helpful to synthesize larger amounts of the different protein truncations to be able to measure circular dichroism. These measurements, would show the degree of secondary structure formation in the truncation samples.

Altogether, the project elucidated the folding transitions of yPGK in terms of domain topology and opened new avenues to study the co-translational folding of RNC complexes with sm-FRET and cryo-EM.

# Appendix A

# sm-FRET denaturant series

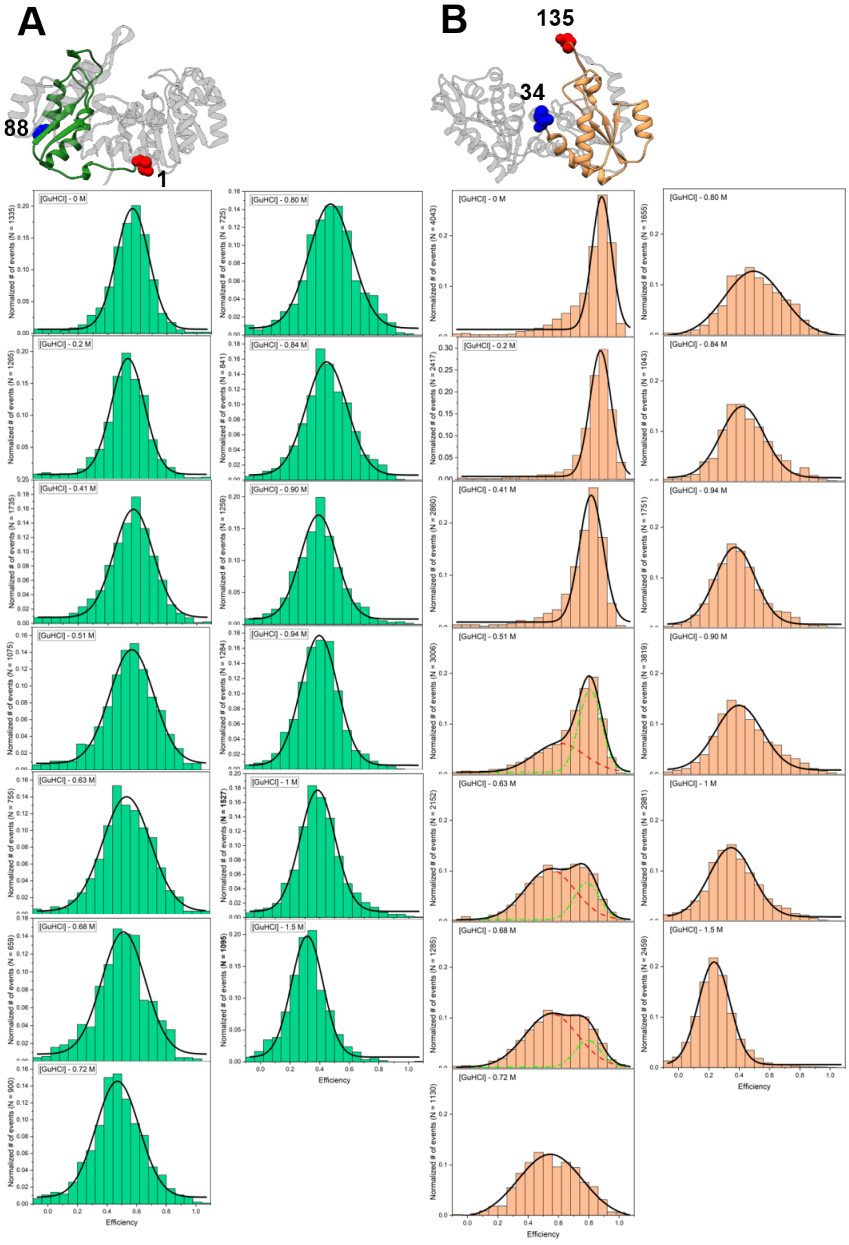



**Figure A.1:** Denaturation series of N-terminal yPGK variants using GuHCl. Attachment positions of the dyes are highlighted as spheres. The structural elements present between the attachment positions are colored in green – variant N2 and copper – variant N3. The corresponding transfer efficiency histograms for each measurement with specific GuHCl concentrations are represented below. The solid and dashed lines represent a Gaussian fit with one or two peaks.

# Appendix B

# Cell-free synthesis of yPGK truncations

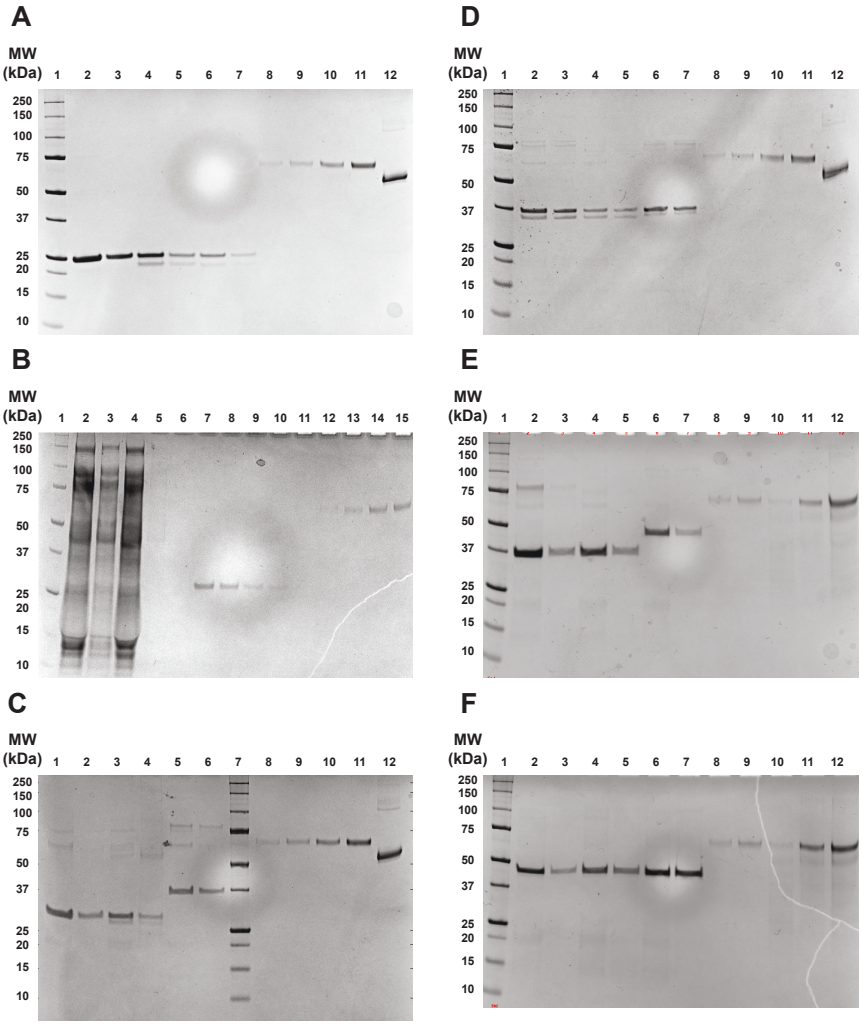



**Figure B.1:** Purification of yPGK truncation variants. Protein standards were the same for all gels. A – yPGK(1–160); lane 1: Marker, lane 2-3: Elution 1/2 pair 1–88, lane 4-5: Elution 1/2 pair 1–135, lane 6-7: Elution 1/2 pair 34–135, lane 8-12: BSA standards (30ng, 60 ng, 120 ng, 240 ng, 480 ng, respectively). B – yPGK(1–215) pair 1–135; lane 1: Marker, lane 2: negative control - no template, lane 3: reaction, lane 4: Flow-through affinity purification, lane 5-6: Wash 1/2, lane 7-10: Elution 1-4, lane 11-15: BSA standards. C – yPGK(1–215) + yPGK(1–290); lane 1-2: Elution 1/2 yPGK(1–215) pair 1–88, lane 3-4: Elution 1/2 yPGK(1–215) pair 34–135, lane 5-6: Elution 1/2 yPGK(1–290) pair 1–88, lane 7: Marker, lane 8-12: BSA standards. D – yPGK(1–290); lane 1: Marker, lane 2-3: Elution 1/2 pair 1–135, lane 4-5: Elution 1/2 pair 34–135, lane 6-7: Elution 1/2 pair 135–290, lane 8-12: BSA standards. E – yPGK(1–290) + yPGK(1–350); lane 1: Marker, lane 2-3: Elution 1/2 yPGK(1–290) pair 202–290, lane 4-5: Elution 1/2 yPGK(1–290) pair 256–290, lane 6-7: Elution 1/2 yPGK(1–350) pair 1–88, lane 8-12: BSA standards. F – yPGK(1–350); lane 1: Marker, lane 2-3: Elution 1/2 pair 1–135, lane 4-5: Elution 1/2 pair 34–135, lane 6-7: Elution 1/2 pair 135–290, lane 8-12: BSA standards.

# References


1. Dominguez R. & Holmes K. C. Actin structure and function. *Annu Rev Biophys* **40,** 169–86. doi:10.1146/annurev-biophys-042910-155359 (2011).

2. Lu Z. & Hunter T. Metabolic Kinases Moonlighting as Protein Kinases. *Trends Biochem Sci* **43,** 301–310. doi:10.1016/j.tibs.2018.01.006 (2018).

3. Posner B. I. & Laporte S. A. Cellular signalling: Peptide hormones and growth factors. *Prog Brain Res* **181,** 1–16. doi:10.1016/S0079-6123(08)81001-1 (2010).

4. Johnson P. F. & McKnight S. L. Eukaryotic transcriptional regulatory proteins. *Annu Rev Biochem* **58,** 799–839. doi:10.1146/annurev.bi.58.070189.004055 (1989).

5. Forthal D. N. Functions of Antibodies. *Microbiol Spectr* **2,** AID–0019–2014. doi:10.1128/microbiolspec.AID-0019-2014 (2014).

6. Komar A. A. A pause for thought along the co-translational folding pathway. *Trends Biochem Sci* **34,** 16–24. doi:10.1016/j.tibs.2008.10.002 (2009).

7. Luheshi L. M., Crowther D. C. & Dobson C. M. Protein misfolding and disease: from the test tube to the organism. *Curr Opin Chem Biol* **12,** 25–31. doi:10.1016/j.cbpa.2008.02.011 (2008).

8. Anfinsen C. B. Principles that govern the folding of protein chains. *Science* **181,** 223–30. doi:10.1126/science.181.4096.223 (1973).

9. Dinner A. R., Sali A., Smith L. J., Dobson C. M. & Karplus M. Understanding protein folding via free-energy surfaces from theory and experiment. *Trends Biochem Sci* **25,** 331–9. doi:10.1016/s0968-0004(00)01610-8 (2000).

10. Wolynes P. G., Onuchic J. N. & Thirumalai D. Navigating the folding routes. *Science* **267,** 1619–20. doi:10.1126/science.7886447 (1995).

11. Dobson C. M. Protein folding and misfolding. *Nature* **426,** 884–90. doi:10.1038/nature02261 (2003).

12. Eaton W. A., Munoz V., Thompson P. A., Henry E. R. & Hofrichter J. Kinetics and dynamics of loops, alpha-helices, beta-hairpins, and fast-folding proteins. *Accounts of Chemical Research* **31,** 745–753. doi:DOI10.1021/ar9700825 (1998).

13. Rao S. T. & Rossmann M. G. Comparison of super-secondary structures in proteins. *J Mol Biol* **76,** 241–56. doi:10.1016/0022-2836(73)90388-4 (1973).

14. Chou K. C. Low-frequency motions in protein molecules. Beta-sheet and beta-barrel. *Biophys J* **48,** 289–97. doi:10.1016/S0006-3495(85)83782-6 (1985).

15. Plaxco K. W., Simons K. T. & Baker D. Contact order, transition state placement and the refolding rates of single domain proteins. *J Mol Biol* **277,** 985–94. doi:10.1006/jmbi.1998.1645 (1998).

16. Djikaev Y. S. & Ruckenstein E. Temperature effects on the nucleation mechanism of protein folding and on the barrierless thermal denaturation of a native protein. *Phys Chem Chem Phys* **10,** 6281–300. doi:10.1039/b807399f (2008).

17. Naganathan A. N., Doshi U. & Munoz V. Protein folding kinetics: barrier effects in chemical and thermal denaturation experiments. *J Am Chem Soc* **129,** 5673–82. doi:10.1021/ja0689740 (2007).

18. McCrary B. S., Edmondson S. P. & Shriver J. W. Hyperthermophile protein folding thermodynamics: differential scanning calorimetry and chemical denaturation of Sac7d. *J Mol Biol* **264,** 784–805. doi:10.1006/jmbi.1996.0677 (1996).

19. Siligardi G. & Hussain R. CD spectroscopy: an essential tool for quality control of protein folding. *Methods Mol Biol* **1261,** 255–76. doi:10.1007/978-1-4939-2230-7_14 (2015).

20. Van Mierlo C. P. & Steensma E. Protein folding and stability investigated by fluorescence, circular dichroism (CD), and nuclear magnetic resonance (NMR) spectroscopy: the flavodoxin story. *J Biotechnol* **79,** 281–98. doi:10.1016/s0168-1656(00)00244-3 (2000).

21. Borgia A., Williams P. M. & Clarke J. Single-molecule studies of protein folding. *Annu Rev Biochem* **77,** 101–25. doi:10.1146/annurev.biochem.77.060706.093102 (2008).

22. Kramer G., Boehringer D., Ban N. & Bukau B. The ribosome as a platform for co-translational processing, folding and targeting of newly synthesized proteins. *Nat Struct Mol Biol* **16,** 589–97. doi:10.1038/nsmb.1614 (2009).

23. Ramakrishnan V. Ribosome structure and the mechanism of translation. *Cell* **108,** 557–72. doi:10.1016/s0092-8674(02)00619-0 (2002).

24. Gualerzi C. O. & Pon C. L. Initiation of mRNA translation in prokaryotes. *Biochemistry* **29,** 5881–9. doi:10.1021/bi00477a001 (1990).

25. Schmeing T. M. & Ramakrishnan V. What recent ribosome structures have revealed about the mechanism of translation. *Nature* **461,** 1234–42. doi:10.1038/nature08403 (2009).

26. Johnson A. E. The co-translational folding and interactions of nascent protein chains: a new approach using fluorescence resonance energy transfer. *FEBS Lett* **579,** 916–20. doi:10.1016/j.febslet.2004.11.046 (2005).

27. Hartl F. U. & Hayer-Hartl M. Converging concepts of protein folding in vitro and in vivo. *Nat Struct Mol Biol* **16,** 574–81. doi:10.1038/nsmb.1591 (2009).

28. Ban N., Nissen P., Hansen J., Capel M., Moore P. B. & Steitz T. A. Placement of protein and RNA structures into a 5 A-resolution map of the 50S ribosomal subunit. *Nature* **400,** 841–7. doi:10.1038/23641 (1999).

29. Voss N. R., Gerstein M., Steitz T. A. & Moore P. B. The geometry of the ribosomal polypeptide exit tunnel. *J Mol Biol* **360,** 893–906. doi:10.1016/j.jmb.2006.05.023 (2006).

30. Picking W. D., Picking W. L., Odom O. W. & Hardesty B. Fluorescence characterization of the environment encountered by nascent polyalanine and polyserine as they exit Escherichia coli ribosomes during translation. *Biochemistry* **31,** 2368–75. doi:10.1021/bi00123a023 (1992).

31. Malkin L. I. & Rich A. Partial resistance of nascent polypeptide chains to proteolytic digestion due to ribosomal shielding. *J Mol Biol* **26,** 329–46. doi:10.1016/0022-2836(67)90301-4 (1967).

32. Holtkamp W., Kokic G., Jager M., Mittelstaet J., Komar A. A. & Rodnina M. V. Cotranslational protein folding on the ribosome monitored in real time. *Science* **350,** 1104–7. doi:10.1126/science.aad0344 (2015).

33. Dao Duc K., Batra S. S., Bhattacharya N., Cate J. H. D. & Song Y. S. Differences in the path to exit the ribosome across the three domains of life. *Nucleic Acids Res* **47,** 4198–4210. doi:10.1093/nar/gkz106 (2019).

34. Liutkute M., Samatova E. & Rodnina M. V. Cotranslational Folding of Proteins on the Ribosome. *Biomolecules* **10.** doi:10.3390/biom10010097 (2020).

35. Gloge F., Becker A. H., Kramer G. & Bukau B. Co-translational mechanisms of protein maturation. *Curr Opin Struct Biol* **24,** 24–33. doi:10.1016/j.sbi.2013.11.004 (2014).

36. Lu J., Kobertz W. R. & Deutsch C. Mapping the electrostatic potential within the ribosomal exit tunnel. *J Mol Biol* **371,** 1378–91. doi:10.1016/j.jmb.2007.06.038 (2007).

37. Farias-Rico J. A., Ruud Selin F., Myronidi I., Fruhauf M. & von Heijne G. Effects of protein size, thermodynamic stability, and net charge on cotranslational folding on the ribosome. *Proc Natl Acad Sci U S A* **115,** E9280–E9287. doi:10.1073/pnas.1812756115 (2018).

38. Lu J. & Deutsch C. Electrostatics in the ribosomal tunnel modulate chain elongation rates. *J Mol Biol* **384,** 73–86. doi:10.1016/j.jmb.2008.08.089 (2008).

39. Charneski C. A. & Hurst L. D. Positively charged residues are the major determinants of ribosomal velocity. *PLoS Biol* **11,** e1001508. doi:10.1371/journal.pbio.1001508 (2013).

40. Requiao R. D., de Souza H. J., Rossetto S., Domitrovic T. & Palhano F. L. Increased ribosome density associated to positively charged residues is evident in ribosome profiling experiments performed in the absence of translation inhibitors. *RNA Biol* **13,** 561–8. doi:10.1080/15476286.2016.1172755 (2016).

41. Woolhead C. A., McCormick P. J. & Johnson A. E. Nascent membrane and secretory proteins differ in FRET-detected folding far inside the ribosome and in their exposure to ribosomal proteins. *Cell* **116,** 725–36. doi:10.1016/s0092-8674(04)00169-2 (2004).

42. Nissley D. A. & O'Brien E. P. Structural Origins of FRET-Observed Nascent Chain Compaction on the Ribosome. *J Phys Chem B* **122,** 9927–9937. doi:10.1021/acs.jpcb.8b07726 (2018).

43. Agirrezabala X., Samatova E., Macher M., Liutkute M., Maiti M., Gil-Carton D., Novacek J., Valle M. & Rodnina M. V. A switch from alpha-helical to beta-strand conformation during co-translational protein folding. *EMBO J* **41,** e109175. doi:10.15252/embj.2021109175 (2022).

44. Nilsson O. B., Hedman R., Marino J., Wickles S., Bischoff L., Johansson M., Muller-Lucks A., Trovato F., Puglisi J. D., O'Brien E. P., Beckmann R. & von Heijne G. Cotranslational Protein Folding inside the Ribosome Exit Tunnel. *Cell Rep* **12,** 1533–40. doi:10.1016/j.celrep.2015.07.065 (2015).

45. Bhushan S., Gartmann M., Halic M., Armache J. P., Jarasch A., Mielke T., Berninghausen O., Wilson D. N. & Beckmann R. alpha-Helical nascent polypeptide chains visualized within distinct regions of the ribosomal exit tunnel. *Nat Struct Mol Biol* **17,** 313–7. doi:10.1038/nsmb.1756 (2010).

46. Agirrezabala X., Samatova E., Klimova M., Zamora M., Gil-Carton D., Rodnina M. V. & Valle M. Ribosome rearrangements at the onset of translational bypassing. *Sci Adv* **3,** e1700147. doi:10.1126/sciadv.1700147 (2017).

47. Su T., Cheng J., Sohmen D., Hedman R., Berninghausen O., von Heijne G., Wilson D. N. & Beckmann R. The force-sensing peptide VemP employs extreme compaction and secondary structure formation to induce ribosomal stalling. *Elife* **6.** doi:10.7554/eLife.25642 (2017).

48. Wruck F., Tian P., Kudva R., Best R. B., von Heijne G., Tans S. J. & Katranidis A. The ribosome modulates folding inside the ribosomal exit tunnel. *Commun Biol* **4,** 523. doi:10.1038/s42003-021-02055-8 (2021).

49. Alexander L. M., Goldman D. H., Wee L. M. & Bustamante C. Non-equilibrium dynamics of a nascent polypeptide during translation suppress its misfolding. *Nat Commun* **10,** 2709. doi:10.1038/s41467-019-10647-6 (2019).

50. Kaiser C. M., Goldman D. H., Chodera J. D., Tinoco I. J. & Bustamante C. The ribosome modulates nascent protein folding. *Science* **334,** 1723–7. doi:10.1126/science.1209740 (2011).

51. Cabrita L. D., Cassaignau A. M. E., Launay H. M. M., Waudby C. A., Wlodarski T., Camilloni C., Karyadi M. E., Robertson A. L., Wang X., Wentink A. S., Goodsell L., Woolhead C. A., Vendruscolo M., Dobson C. M. & Christodoulou J. A structural ensemble of a ribosome-nascent chain complex during cotranslational protein folding. *Nat Struct Mol Biol* **23,** 278–285. doi:10.1038/nsmb.3182 (2016).

52. Roder H. & Colon W. Kinetic role of early intermediates in protein folding. *Curr Opin Struct Biol* **7,** 15–28. doi:10.1016/s0959-440x(97)80004-8 (1997).

53. De Sancho D. & Munoz V. Integrated prediction of protein folding and unfolding rates from only size and structural class. *Phys Chem Chem Phys* **13,** 17030–43. doi:10.1039/c1cp20402e (2011).

54. Zhang G., Hubalewska M. & Ignatova Z. Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. *Nat Struct Mol Biol* **16,** 274–80. doi:10.1038/nsmb.1554 (2009).

55. Zhou M., Guo J., Cha J., Chae M., Chen S., Barral J. M., Sachs M. S. & Liu Y. Non-optimal codon usage affects expression, structure and function of clock protein FRQ. *Nature* **495,** 111–5. doi:10.1038/nature11833 (2013).

56. Chen C., Zhang H., Broitman S. L., Reiche M., Farrell I., Cooperman B. S. & Goldman Y. E. Dynamics of translation by single ribosomes through mRNA secondary structures. *Nat Struct Mol Biol* **20,** 582–8. doi:10.1038/nsmb.2544 (2013).

57. Hartl F. U., Bracher A. & Hayer-Hartl M. Molecular chaperones in protein folding and proteostasis. *Nature* **475,** 324–32. doi:10.1038/nature10317 (2011).

58. McGuffee S. R. & Elcock A. H. Diffusion, crowding & protein stability in a dynamic molecular model of the bacterial cytoplasm. *PLoS Comput Biol* **6,** e1000694. doi:10.1371/journal.pcbi.1000694 (2010).

59. Waudby C. A., Dobson C. M. & Christodoulou J. Nature and Regulation of Protein Folding on the Ribosome. *Trends Biochem Sci* **44,** 914–926. doi:10.1016/j.tibs.2019.06.008 (2019).

60. Cerminara M., Schöne A., Ritter I., Gabba M. & Fitter J. Mapping multiple distances in a multi-domain protein for the identification of folding intermediates. *Biophysical Journal.* doi:10.1016/j.bpj.2019.12.006 (2019).

61. Vogel C., Bashton M., Kerrison N. D., Chothia C. & Teichmann S. A. Structure, function and evolution of multidomain proteins. *Curr Opin Struct Biol* **14,** 208–16. doi:10.1016/j.sbi.2004.03.011 (2004).

62. Apic G., Huber W. & Teichmann S. A. Multi-domain protein families and domain pairs: comparison with known structures and a random model of domain recombination. *J Struct Funct Genomics* **4,** 67–78. doi:10.1023/a:1026113408773 (2003).

63. Vogel C., Berzuini C., Bashton M., Gough J. & Teichmann S. A. Supradomains: evolutionary units larger than single protein domains. *J Mol Biol* **336,** 809–23. doi:10.1016/j.jmb.2003.12.026 (2004).

64. Kaur U., Kihn K. C., Ke H., Kuo W., Gierasch L. M., Hebert D. N., Wintrode P. L., Deredge D. & Gershenson A. The conformational landscape of a serpin N-terminal subdomain facilitates folding and in-cell quality control. *bioRxiv.* doi:10.1101/2023.04.24.537978 (2023).

65. Bashton M. & Chothia C. The geometry of domain combination in proteins. *J Mol Biol* **315,** 927–39. doi:10.1006/jmbi.2001.5288 (2002).

66. Han J. H., Kerrison N., Chothia C. & Teichmann S. A. Divergence of interdomain geometry in two-domain proteins. *Structure* **14,** 935–45. doi:10.1016/j.str.2006.01.016 (2006).

67. Han J. H., Batey S., Nickson A. A., Teichmann S. A. & Clarke J. The folding and evolution of multidomain proteins. *Nat Rev Mol Cell Biol* **8,** 319–30. doi:10.1038/nrm2144 (2007).

68. Batey S., Nickson A. A. & Clarke J. Studying the folding of multidomain proteins. *HFSP J* **2,** 365–77. doi:10.2976/1.2991513 (2008).

69. Frydman J. Folding of newly translated proteins in vivo: the role of molecular chaperones. *Annu Rev Biochem* **70,** 603–47. doi:10.1146/annurev.biochem.70.1.603 (2001).

70. Watson H. C., Walker N. P., Shaw P. J., Bryant T. N., Wendell P. L., Fothergill L. A., Perkins R. E., Conroy S. C., Dobson M. J., Tuite M. F. & et al. Sequence and structure of yeast phosphoglycerate kinase. *EMBO J* **1,** 1635–40 (1982).

71. Blake C. C. & Rice D. W. Phosphoglycerate kinase. *Philos Trans R Soc Lond B Biol Sci* **293,** 93–104. doi:10.1098/rstb.1981.0063 (1981).

72. Minard P., Hall L., Betton J. M., Missiakas D. & Yon J. M. Efficient expression and characterization of isolated structural domains of yeast phosphoglycerate kinase generated by site-directed mutagenesis. *Protein Eng* **3,** 55–60. doi:10.1093/protein/3.1.55 (1989).

73. Li Q., Scholl Z. N. & Marszalek P. E. Unraveling the Mechanical Unfolding Pathways of a Multidomain Protein: Phosphoglycerate Kinase. *Biophys J* **115,** 46–58. doi:10.1016/j.bpj.2018.05.028 (2018).

74. Missiakas D., Betton J. M., Chaffotte A., Minard P. & Yon J. M. Kinetic studies of the refolding of yeast phosphoglycerate kinase: comparison with the isolated engineered domains. *Protein Sci* **1,** 1485–93. doi:10.1002/pro.5560011110 (1992).

75. Missiakas D., Betton J. M., Minard P. & Yon J. M. Unfolding-refolding of the domains in yeast phosphoglycerate kinase: comparison with the isolated engineered domains. *Biochemistry* **29,** 8683–9. doi:10.1021/bi00489a025 (1990).

76. Bryant T. N., Watson H. C. & Wendell P. L. Structure of yeast phosphoglycerate kinase. *Nature* **247,** 14–7. doi:10.1038/247014a0 (1974).

77. Lillo M. P., Szpikowska B. K., Mas M. T., Sutin J. D. & Beechem J. M. Real-Time Measurement of Multiple Intramolecular Distances during Protein Folding Reactions: A Multisite Stopped-Flow Fluorescence Energy-Transfer Study of Yeast Phosphoglycerate Kinase. *Biochemistry* **36,** 11273–81. doi:10.1021/bi970789z (1997).

78. Liu K., Maciuba K. & Kaiser C. M. The Ribosome Cooperates with a Chaperone to Guide Multi-domain Protein Folding. *Mol Cell* **74,** 310–319 e7. doi:10.1016/j.molcel.2019.01.043 (2019).

79. Kolb V. A., Makeyev E. V. & Spirin A. S. Co-translational folding of an eukaryotic multidomain protein in a prokaryotic translation system. *J Biol Chem* **275,** 16597–601. doi:10.1074/jbc.M002030200 (2000).

80. Lillo M. P., Beechem J. M., Szpikowska B. K., Sherman M. A. & Mas M. T. Design and characterization of a multisite fluorescence energy-transfer system for protein folding studies: a steady-state and time-resolved study of yeast phosphoglycerate kinase. *Biochemistry* **36,** 11261–72. doi:10.1021/bi9707887 (1997).

81. Schrodinger L. *The PyMOL Molecular Graphics System, Version 1.8* Unpublished Work. 2015.

82. Komar A. A. Unraveling co-translational protein folding: Concepts and methods. *Methods* **137,** 71–81. doi:10.1016/j.ymeth.2017.11.007 (2018).

83. Thommen M., Holtkamp W. & Rodnina M. V. Co-translational protein folding: progress and methods. *Curr Opin Struct Biol* **42,** 83–89. doi:10.1016/j.sbi.2016.11.020 (2017).

84. Zubay G. In vitro synthesis of protein in microbial systems. *Annu Rev Genet* **7,** 267–87. doi:10.1146/annurev.ge.07.120173.001411 (1973).

85. Ohashi H., Kanamori T., Shimizu Y. & Ueda T. A highly controllable reconstituted cell-free system–a breakthrough in protein synthesis research. *Curr Pharm Biotechnol* **11,** 267–71. doi:10.2174/138920110791111889 (2010).

86. DeVries J. K. & Zubay G. DNA-directed peptide synthesis. II. The synthesis of the alpha-fragment of the enzyme beta-galactosidase. *Proc Natl Acad Sci U S A* **57,** 1010–2. doi:10.1073/pnas.57.4.1010 (1967).

87. Hamadani K. M., Howe J., Jensen M. K., Wu P., Cate J. H. D. & Marqusee S. An in vitro tag-and-modify protein sample generation method for single-molecule fluorescence resonance energy transfer. *J Biol Chem* **292,** 15636–15648. doi:10.1074/jbc.M117.791723 (2017).

88. Ito K. & Chiba S. Arrest peptides: cis-acting modulators of translation. *Annu Rev Biochem* **82,** 171–202. doi:10.1146/annurev-biochem-080211-105026 (2013).

89. Muto H., Nakatogawa H. & Ito K. Genetically encoded but nonpolypeptide prolyl-tRNA functions in the A site for SecM-mediated ribosomal stall. *Mol Cell* **22,** 545–52. doi:10.1016/j.molcel.2006.03.033 (2006).

90. Vazquez-Laslop N., Thum C. & Mankin A. S. Molecular mechanism of drug-dependent ribosome stalling. *Mol Cell* **30,** 190–202. doi:10.1016/j.molcel.2008.02.026 (2008).

91. Rajapandi T., Dolan K. M. & Oliver D. B. The first gene in the Escherichia coli secA operon, gene X, encodes a nonessential secretory protein. *J Bacteriol* **173,** 7092–7. doi:10.1128/jb.173.22.7092-7097.1991 (1991).

92. Park E. & Rapoport T. A. Preserving the membrane barrier for small molecules during bacterial protein translocation. *Nature* **473,** 239–42. doi:10.1038/nature10014 (2011).

93. Kempf N., Remes C., Ledesch R., Zuchner T., Hofig H., Ritter I., Katranidis A. & Fitter J. A Novel Method to Evaluate Ribosomal Performance in Cell-Free Protein Synthesis Systems. *Sci Rep* **7,** 46753. doi:10.1038/srep46753 (2017).

94. Katranidis A. & Fitter J. Single-Molecule Techniques and Cell-Free Protein Synthesis: A Perfect Marriage. *Anal Chem* **91,** 2570–2576. doi:10.1021/acs.analchem.8b03855 (2019).

95. Noren C. J., Anthony-Cahill S. J., Griffith M. C. & Schultz P. G. A general method for site-specific incorporation of unnatural amino acids into proteins. *Science* **244,** 182–8. doi:10.1126/science.2649980 (1989).

96. Balchin D., Hayer-Hartl M. & Hartl F. U. In vivo aspects of protein folding and quality control. *Science* **353,** aac4354. doi:10.1126/science.aac4354 (2016).

97. Han Y., David A., Liu B., Magadan J. G., Bennink J. R., Yewdell J. W. & Qian S. B. Monitoring cotranslational protein folding in mammalian cells at codon resolution. *Proc Natl Acad Sci U S A* **109,** 12467–72. doi:10.1073/pnas.1208138109 (2012).

98. Hsu S. T., Fucini P., Cabrita L. D., Launay H., Dobson C. M. & Christodoulou J. Structure and dynamics of a ribosome-bound nascent chain by NMR spectroscopy. *Proc Natl Acad Sci U S A* **104,** 16516–21. doi:10.1073/pnas.0704664104 (2007).

99. Frank J. Advances in the field of single-particle cryo-electron microscopy over the last decade. *Nat Protoc* **12,** 209–212. doi:10.1038/nprot.2017.004 (2017).

100. Seidelt B., Innis C. A., Wilson D. N., Gartmann M., Armache J. P., Villa E., Trabuco L. G., Becker T., Mielke T., Schulten K., Steitz T. A. & Beckmann R. Structural insight into nascent polypeptide chain-mediated translational stalling. *Science* **326,** 1412–5. doi:10.1126/science.1177662 (2009).

101. Bhushan S., Hoffmann T., Seidelt B., Frauenfeld J., Mielke T., Berninghausen O., Wilson D. N. & Beckmann R. SecM-stalled ribosomes adopt an altered geometry at the peptidyl transferase center. *PLoS Biol* **9,** e1000581. doi:10.1371/journal.pbio.1000581 (2011).

102. Wilson D. N. & Beckmann R. The ribosomal tunnel as a functional environment for nascent polypeptide folding and translational stalling. *Curr Opin Struct Biol* **21,** 274–82. doi:10.1016/j.sbi.2011.01.007 (2011).

103. Punjani A. & Fleet D. J. 3D Variability Analysis: Directly resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM images. *J Struct Biol* **213,** 107702. doi:10.1016/j.jsb.2021.107702 (2021).

104. Ellis J. P., Bakke C. K., Kirchdoerfer R. N., Jungbauer L. M. & Cavagnero S. Chain dynamics of nascent polypeptides emerging from the ribosome. *ACS Chem Biol* **3,** 555–66. doi:10.1021/cb800059u (2008).

105. Ellis J. P., Culviner P. H. & Cavagnero S. Confined dynamics of a ribosome-bound nascent globin: Cone angle analysis of fluorescence depolarization decays in the presence of two local motions. *Protein Sci* **18,** 2003–15. doi:10.1002/pro.196 (2009).

106. Weinreis S. A., Ellis J. P. & Cavagnero S. Dynamic fluorescence depolarization: a powerful tool to explore protein folding on the ribosome. *Methods* **52,** 57–73. doi:10.1016/j.ymeth.2010.06.001 (2010).

107. Jablonski A. Efficiency of anti-stokes fluorescence in dyes. *Nature* **131,** 839–840. doi:DOI10.1038/131839b0 (1933).

108. Hohlbein J., Craggs T. D. & Cordes T. Alternating-laser excitation: single-molecule FRET and beyond. *Chem Soc Rev* **43,** 1156–71. doi:10.1039/c3cs60233h (2014).

109. Yu L., Lei Y., Ma Y., Liu M., Zheng J., Dan D. & Gao P. A Comprehensive Review of Fluorescence Correlation Spectroscopy. *Frontiers in Physics* **9.** doi:10.3389/fphy.2021.644450 (2021).

110. Enderlein J., Gregor I., Patra D. & Fitter J. Art and artefacts of fluorescence correlation spectroscopy. *Curr Pharm Biotechnol* **5,** 155–61. doi:10.2174/1389201043377020 (2004).

111. Barth A., von Voithenberg L. V. & Lamb D. C. in *Advanced Photon Counting* 129–157 (2014). doi:10.1007/4243_2014_66.

112. Förster T. Zwischenmolekulare Energiewanderung und Fluoreszenz. *Annalen der Physik* **437,** 55–75. doi:10.1002/andp.19484370105 (1948).

113. Sisamakis E., Valeri A., Kalinin S., Rothwell P. J. & Seidel C. A. Accurate single-molecule FRET studies using multiparameter fluorescence detection. *Methods Enzymol* **475,** 455–514. doi:10.1016/S0076-6879(10)75018-7 (2010).

114. Eggeling C., Berger S., Brand L., Fries J. R., Schaffer J., Volkmer A. & Seidel C. A. Data registration and selective single-molecule analysis using multiparameter fluorescence detection. *J Biotechnol* **86,** 163–80. doi:10.1016/s0168-1656(00)00412-0 (2001).

115. Kapanidis A. N., Lee N. K., Laurence T. A., Doose S., Margeat E. & Weiss S. Fluorescence-aided molecule sorting: analysis of structure and interactions by alternating-laser excitation of single molecules. *Proc Natl Acad Sci U S A* **101,** 8936–41. doi:10.1073/pnas.0401690101 (2004).

116. SUBRAMANIAN K. *Textbook on Fundamentals and Applications of Nanotechnology* (Astral International Pvt. Limited, 2018).

117. Reimer L. & Kohl H. *Transmission Electron Microscopy: Physics of Image Formation* (Springer New York, 2008).

118. Harris J. *Negative Staining and Cryoelectron Microscopy: The thin film techniques, Oxford: BIOS Scientific* 1997.

119. Dubochet J., Adrian M., Chang J. J., Homo J. C., Lepault J., McDowall A. W. & Schultz P. Cryo-electron microscopy of vitrified specimens. *Q Rev Biophys* **21,** 129–228. doi:10.1017/s0033583500004297 (1988).

120. Mindell J. A. & Grigorieff N. Accurate determination of local defocus and specimen tilt in electron microscopy. *J Struct Biol* **142,** 334–47. doi:10.1016/s1047-8477(03)00069-8 (2003).

121. Raimondi V. & Grinzato A. A basic introduction to single particles cryo-electron microscopy. *Aims Biophysics* **9,** 5–20. doi:10.3934/biophy.2022002 (2022).

122. Rosenthal P. B. & Henderson R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J Mol Biol* **333,** 721–45. doi:10.1016/j.jmb.2003.07.013 (2003).

123. McMullan G., Chen S., Henderson R. & Faruqi A. R. Detective quantum efficiency of electron area detectors in electron microscopy. *Ultramicroscopy* **109,** 1126–43. doi:10.1016/j.ultramic.2009.04.002 (2009).

124. Nogales E. & Scheres S. H. Cryo-EM: A Unique Tool for the Visualization of Macromolecular Complexity. *Mol Cell* **58,** 677–89. doi:10.1016/j.molcel.2015.02.019 (2015).

125. Scheres S. H. & Chen S. Prevention of overfitting in cryo-EM structure determination. *Nat Methods* **9,** 853–4. doi:10.1038/nmeth.2115 (2012).

126. Burley S. K., Bhikadiya C., Bi C., Bittrich S., Chao H., Chen L., Craig P. A., Crichlow G. V., Dalenberg K., Duarte J. M., Dutta S., Fayazi M., Feng Z., Flatt J. W., Ganesan S., Ghosh S., Goodsell D. S., Green R. K., Guranovic V., Henry J., Hudson B. P., Khokhriakov I., Lawson C. L., Liang Y., Lowe R., Peisach E., Persikova I., Piehl D. W., Rose Y., Sali A., Segura J., Sekharan M., Shao C., Vallat B., Voigt M., Webb B., Westbrook J. D., Whetstone S., Young J. Y., Zalevsky A. & Zardecki C. RCSB Protein Data Bank (RCSB.org): delivery of experimentally-determined PDB structures alongside one million computed structure models of proteins from artificial intelligence/machine learning. *Nucleic Acids Res* **51,** D488–D508. doi:10.1093/nar/gkac1077 (2023).

127. Emsley P., Lohkamp B., Scott W. G. & Cowtan K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66,** 486–501. doi:10.1107/S0907444910007493 (2010).

128. Burnley T., Palmer C. M. & Winn M. Recent developments in the CCP-EM software suite. *Acta Crystallogr D Struct Biol* **73,** 469–477. doi:10.1107/S2059798317007859 (2017).

129. Goddard T. D., Huang C. C., Meng E. C., Pettersen E. F., Couch G. S., Morris J. H. & Ferrin T. E. UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci* **27,** 14–25. doi:10.1002/pro.3235 (2018).

130. Pettersen E. F., Goddard T. D., Huang C. C., Couch G. S., Greenblatt D. M., Meng E. C. & Ferrin T. E. UCSF Chimera–a visualization system for exploratory research and analysis. *J Comput Chem* **25,** 1605–12. doi:10.1002/jcc.20084 (2004).

131. Punjani A., Rubinstein J. L., Fleet D. J. & Brubaker M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* **14,** 290–296. doi:10.1038/nmeth.4169 (2017).

132. Laskowski R. A. & Swindells M. B. LigPlot+: multiple ligand–protein interaction diagrams for drug discovery. doi:dx.doi.org/10.1021/ci200227u (2011).

133. Jakobi A. J., Wilmanns M. & Sachse C. Model-based local density sharpening of cryo-EM maps. *Elife* **6.** doi:10.7554/eLife.27131 (2017).

134. Schrimpf W., Barth A., Hendrix J. & Lamb D. C. PAM: A Framework for Integrated Analysis of Imaging, Single-Molecule, and Ensemble Fluorescence Data. *Biophys J* **114,** 1518–1528. doi:10.1016/j.bpj.2018.02.035 (2018).

135. Liebschner D., Afonine P. V., Baker M. L., Bunkoczi G., Chen V. B., Croll T. I., Hintze B., Hung L. W., Jain S., McCoy A. J., Moriarty N. W., Oeffner R. D., Poon B. K., Prisant M. G., Read R. J., Richardson J. S., Richardson D. C., Sammito M. D., Sobolev O. V., Stockwell D. H., Terwilliger T. C.,

Urzhumtsev A. G., Videau L. L., Williams C. J. & Adams P. D. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr D Struct Biol* **75,** 861–877. doi:10.1107/S2059798319011471 (2019).

136. Scheres S. H. Processing of Structurally Heterogeneous Cryo-EM Data in RELION. *Methods Enzymol* **579,** 125–57. doi:10.1016/bs.mie.2016.04.012 (2016).

137. McNicholas S., Potterton E., Wilson K. S. & Noble M. E. Presenting your structures: the CCP4mg molecular-graphics software. *Acta Crystallogr D Biol Crystallogr* **67,** 386–94. doi:10.1107/S0907444911007281 (2011).

138. Pintilie G. & Chiu W. Comparison of Segger and other methods for segmentation and rigid-body docking of molecular components in cryo-EM density maps. *Biopolymers* **97,** 742–60. doi:10.1002/bip.22074 (2012).

139. Pintilie G. D., Zhang J., Goddard T. D., Chiu W. & Gossard D. C. Quantitative analysis of cryo-EM density map segmentation by watershed and scale-space filtering, and fitting of structures by alignment to regions. *J Struct Biol* **170,** 427–38. doi:10.1016/j.jsb.2010.03.007 (2010).

140. Stark H. GraFix: stabilization of fragile macromolecular complexes for single particle cryo-EM. *Methods Enzymol* **481,** 109–26. doi:10.1016/S0076-6879(10)81005-5 (2010).

141. Dertinger T., Pacheco V., von der Hocht I., Hartmann R., Gregor I. & Enderlein J. Two-focus fluorescence correlation spectroscopy: a new tool for accurate and absolute diffusion measurements. *Chemphyschem* **8,** 433–43. doi:10.1002/cphc.200600638 (2007).

142. Petrasek Z. & Schwille P. Precise measurement of diffusion coefficients using scanning fluorescence correlation spectroscopy. *Biophys J* **94,** 1437–48. doi:10.1529/biophysj.107.108811 (2008).

143. Watson Z. L., Ward F. R., Meheust R., Ad O., Schepartz A., Banfield J. F. & Cate J. H. Structure of the bacterial ribosome at 2 A resolution. *Elife* **9.** doi:10.7554/eLife.60482 (2020).

144. McPhillips T. M., Hsu B. T., Sherman M. A., Mas M. T. & Rees D. C. Structure of the R65Q mutant of yeast 3-phosphoglycerate kinase complexed with Mg-AMP-PNP and 3-phospho-D-glycerate. *Biochemistry* **35,** 4118–27. doi:10.1021/bi952500o (1996).

145. Kalinin S., Peulen T., Sindbert S., Rothwell P. J., Berger S., Restle T., Goody R. S., Gohlke H. & Seidel C. A. A toolkit and benchmark study for FRET-restrained high-precision structural modeling. *Nat Methods* **9,** 1218–25. doi:10.1038/nmeth.2222 (2012).

146. Lapidus L. J., Steinbach P. J., Eaton W. A., Szabo A. & Hofrichter J. Effects of chain stiffness on the dynamics of loop formation in polypeptides. Appendix: Testing a 1-dimensional diffusion model for peptide dynamics. *Journal of Physical Chemistry B* **106,** 11628–11640. doi:10.1021/jp020829v (2002).

147. Lindhoud S., Pirchi M., Westphal A. H., Haran G. & van Mierlo C. P. Gradual Folding of an Off-Pathway Molten Globule Detected at the Single-Molecule Level. *J Mol Biol* **427,** 3148–57. doi:10.1016/j.jmb.2015.07.002 (2015).

148. Ptitsyn O. B. & Uversky V. N. The molten globule is a third thermodynamical state of protein molecules. *FEBS Lett* **341,** 15–8. doi:10.1016/0014-5793(94)80231-9 (1994).

149. Fink A. L. Compact intermediate states in protein folding. *Annu Rev Biophys Biomol Struct* **24,** 495–522. doi:10.1146/annurev.bb.24.060195.002431 (1995).

150. Sadoine M., Cerminara M., Kempf N., Gerrits M., Fitter J. & Katranidis A. Selective Double-Labeling of Cell-Free Synthesized Proteins for More Accurate smFRET Studies. *Anal Chem* **89,** 11278–11285. doi:10.1021/acs.analchem.7b01639 (2017).

151. Sadoine M., Cerminara M., Gerrits M., Fitter J. & Katranidis A. Cotranslational Incorporation into Proteins of a Fluorophore Suitable for smFRET Studies. *ACS Synth Biol* **7,** 405–411. doi:10.1021/acssynbio.7b00433 (2018).

152. Cui Z., Mureev S., Polinkovsky M. E., Tnimov Z., Guo Z., Durek T., Jones A. & Alexandrov K. Combining Sense and Nonsense Codon Reassignment for Site-Selective Protein Modification with Unnatural Amino Acids. *ACS Synth Biol* **6,** 535–544. doi:10.1021/acssynbio.6b00245 (2017).

153. Lee K. J., Kang D. & Park H. S. Site-Specific Labeling of Proteins Using Unnatural Amino Acids. *Mol Cells* **42,** 386–396. doi:10.14348/molcells.2019.0078 (2019).

154. Van der Stel A. X., Gordon E. R., Sengupta A., Martinez A. K., Klepacki D., Perry T. N., Herrero Del Valle A., Vazquez-Laslop N., Sachs M. S., Cruz-Vera L. R. & Innis C. A. Structural basis for the tryptophan sensitivity of TnaC-mediated ribosome stalling. *Nat Commun* **12,** 5340. doi:10.1038/s41467-021-25663-8 (2021).

155. Zhang J., Pan X., Yan K., Sun S., Gao N. & Sui S. F. Mechanisms of ribosome stalling by SecM at multiple elongation steps. *Elife* **4.** doi:10.7554/eLife.09684 (2015).

156. Chen X., Rajasekaran N., Liu K. & Kaiser C. M. Synthesis runs counter to directional folding of a nascent protein domain. *Nat Commun* **11,** 5096. doi:10.1038/s41467-020-18921-8 (2020).

157. Dingfelder F., Benke S., Nettels D. & Schuler B. Mapping an Equilibrium Folding Intermediate of the Cytolytic Pore Toxin ClyA with Single-Molecule FRET. *J Phys Chem B* **122,** 11251–11261. doi:10.1021/acs.jpcb.8b07026 (2018).

158. Lindhoud S., Pirchi M., Westphal A. H., Haran G. & van Mierlo C. P. Gradual Folding of an Off-Pathway Molten Globule Detected at the Single-Molecule Level. *J Mol Biol* **427,** 3148–57. doi:10.1016/j.jmb.2015.07.002 (2015).

Band / Volume 269
**The complex inositol metabolism of *Corynebacterium glutamicum* and its application for the production of rare inositols**
P. Ramp (2023), VI, 161 pp
ISBN: 978-3-95806-699-1

Band / Volume 270
**Spin- and orbital-dependent band structure of unconventional topological semimetals**
K. Hagiwara (2023), v, 115 pp
ISBN: 978-3-95806-701-1

Band / Volume 271
**Neutron scattering**
Experimental Manuals of the JCNS Laboratory Course held at Forschungszentrum Jülich and at the Heinz-Maier-Leibnitz Zentrum Garching edited by T. Brückel, S. Förster, K. Friese, M. Kruteva, M. Zobel and R. Zorn (2023), ca 150 pp
ISBN: 978-3-95806-705-9

Band / Volume 272
**Ab-initio investigation of the interplay between the hyperfine interaction and complex magnetism at the nanoscale**
S. R. S. Shehada (2023), ix, xi, 119 pp
ISBN: 978-3-95806-718-9

Band / Volume 273
**Analysis of the signal transduction cascade tuning the 2-oxoglutarate dehydrogenase activity in *Corynebacterium glutamicum***
L. Sundermeyer (2023), VI, 119 pp
ISBN: 978-3-95806-722-6

Band / Volume 274
**Multicellular defense against phage infection in *Streptomyces* – impact of secondary metabolites and mycelial development**
L. Kever (2023), iv, 246 pp
ISBN: 978-3-95806-724-0

Band / Volume 275
**Investigation of the electronic band structure of 2D transition metal dichalcogenides via angle-resolved photoemission spectroscopy**
B. Parashar (2023), xvii, 156 pp
ISBN: 978-3-95806-725-7

Band / Volume 276
**Strain- and process engineering for polyketides production with *Pseudomonas taiwanensis* VLB120 in two-phase cultivations**
T. P. Schwanemann (2023), 230 pp
ISBN: 978-3-95806-726-4

Band / Volume 277
**Quantitative atomic-level investigation of solid materials through multidimensional electron diffraction measurements**
H. L. Lalandec-Robert (2024), xxi, 152 pp
ISBN: 978-3-95806-735-6

Band / Volume 278
**Studies on the cAMP-responsive regulatory network of *Corynebacterium glutamicum***
N. Wolf (2024), iii, 122 pp
ISBN: 978-3-95806-736-3

Band / Volume 279
**Rare-earth atoms on two-dimensional materials: ab initio investigation of magnetic properties**
J. P. Carbone (2024), 235 pp
ISBN: 978-3-95806-740-0

Band / Volume 280
**Communities of Niche-optimized Strains (CoNoS) – a novel concept for improving biotechnological production**
R. Zuchowski (2024), VIII, 168 pp
ISBN: 978-3-95806-743-1

Band / Volume 281
**Enabling mixed microbial upcycling of plastic monomers**
Y. S. Ackermann (2024), XVI, 203 pp
ISBN: 978-3-95806-749-3

Band / Volume 282
**Folding and structural studies of *saccharomyces cerevisiae* Phosphoglycerate Kinase**
N. Bustorff (2024), xxvi, 126 pp
ISBN: 978-3-95806-754-7

Schlüsseltechnologien / Key Technologies
Band / Volume 282
ISBN 978-3-95806-754-7

Mitglied der Helmholtz-Gemeinschaft

JÜLICH
Forschungszentrum