

# Conformational Dynamics of Large Proteins and Protein Complexes

Dissertation  
for the award of the degree  
"Doctor rerum naturalium" (Dr. rer. nat.)  
of the Georg-August-Universität Göttingen

within the doctoral program IMPRS Molecular Biology  
of the Göttingen Graduate School for Neuroscience, Biophysics and Molecular  
Biosciences (GGNB)

submitted by

David Haselbach

from Cottbus, Germany

Göttingen, 2014



---

Thesis Committee

Prof. Dr. Holger Stark

3D Electron Cryo-Microscopy, Max Planck Institute for biophysical Chemistry

Prof. Dr. Kai Tittmann

Bioanalytics, Albrecht-Haller-Institute

Prof. Dr. Jörg Enderlein

Single Molecule Spectroscopy and Imaging for Biophysics and Complex Systems, 3rd Institute for Physics

Members of the Examination Board

Reviewer: Prof. Dr. Holger Stark

3D Electron Cryo-Microscopy, Max Planck Institute for biophysical Chemistry

Second Reviewer: Prof. Dr. Kai Tittmann

Bioanalytics, Albrecht-Haller-Institute

Further members of the Examination Board:

Prof. Dr. Jörg Enderlein

Single Molecule Spectroscopy and Imaging for Biophysics and Complex Systems, 3rd Institute for Physics

Prof. Dr. Ralf Ficner

Molecular Structural Biology, Institute for Microbiology and Genetics

Prof. Dr. Henning Urlaub

Bioanalytical Massspectrometry, Max Planck Institute for biophysical Chemistry

Prof. Dr. Detlef Doenecke

Department of Molecular Biology, University Medical Center

Date of the oral examination: 13.10.2014

# Affidavit

I hereby declare that this PhD thesis 'Conformational Dynamics of Large Protein Complexes' has been written independently with no other aids or sources than quoted. This thesis (wholly or in part) has not been submitted elsewhere for any academic award or qualification.

David Haselbach

# Contents

<b>Contents</b>	<b>i</b>
List of Figures . . . . .	v
List of Tables . . . . .	vii
<b>Acknowledgements</b>	<b>ix</b>
<b>Abstract</b>	<b>xiii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Molecular Machines . . . . .	1
1.2 Conformational Dynamics . . . . .	4
1.2.1 What are Conformational Dynamics? . . . . .	4
1.2.2 Physics of Conformational Dynamics . . . . .	7
1.2.3 Methods to analyse Conformational Dynamics . . . . .	11
1.2.3.1 MD Simulation . . . . .	11
1.2.3.2 Spectroscopy . . . . .	12
1.2.3.3 Single Molecule Approaches . . . . .	13
1.2.3.4 Structural Methods . . . . .	13
1.2.3.5 Nuclear Magnetic Resonance . . . . .	14
1.2.3.6 X-ray Crystallography . . . . .	14
1.2.3.7 Single Particle Electron Cryo-Microscopy . . . . .	15
1.2.3.8 Method Combinations . . . . .	15
1.3 Single particle cryo EM as tool to analyze conformational dynamics . . . . .	16
1.3.1 General workflow . . . . .	16
1.3.2 Conformational Analysis . . . . .	19
1.3.2.1 Biochemical and biophysical strategies . . . . .	20
1.3.2.2 Computational Strategies . . . . .	21
1.4 Biological Model Systems . . . . .	22
1.4.1 Chromosome Maintenance Factor 1 . . . . .	23
1.4.1.1 Nucleocytoplasmatic transport . . . . .	23
1.4.1.2 Structural Characteristics of Exportins . . . . .	24
1.4.2 Pyruvate Dehydrogenase Complex . . . . .	25
1.4.2.1 Biochemistry of the PDHc . . . . .	25
1.4.2.2 Structural Insights into the PDHc . . . . .	26

1.4.3	<i>Biomphalaria glabrata</i> Achetyl Choline Binding protein . . . . .	29
1.5	Aim of the work . . . . .	30
<b>2</b>	<b>Materials and Methods</b>	<b>31</b>
2.1	Materials . . . . .	31
2.1.1	Software . . . . .	31
2.1.2	Special Equipment . . . . .	31
2.1.3	Chemicals . . . . .	32
2.2	Microbiological Methods . . . . .	34
2.2.1	<i>Escherichia coli</i> Cultivation . . . . .	34
2.2.2	<i>Chaetomium thermophilum</i> Cultivation . . . . .	34
2.2.3	Spore Cultivation . . . . .	34
2.2.4	Mycelium Growth and Disruption . . . . .	34
2.3	Biochemical Methods . . . . .	35
2.3.1	General Biochemical Methods . . . . .	35
2.3.2	ProteoPlex . . . . .	35
2.3.3	Purification of CRM1 . . . . .	38
2.3.4	Purification of PDH . . . . .	39
2.3.5	Functional Assays of the PDH . . . . .	39
2.3.6	Purification of AChBP . . . . .	40
2.4	Fundamental Concepts of Single Particle Electron Microscopy . . . . .	41
2.4.1	The Electron Microscope . . . . .	41
2.4.1.1	The Electron Source . . . . .	42
2.4.2	Magnetic Lenses and their Aberrations . . . . .	43
2.4.3	Electron Detection . . . . .	44
2.4.4	Short Excursion into Fourier Space . . . . .	45
2.4.5	Image Formation . . . . .	45
2.4.6	Image Restoration . . . . .	48
2.4.7	Image Processing Basics . . . . .	49
2.4.8	Image Preprocessing . . . . .	51
2.4.9	Alignment and Classification . . . . .	51
2.4.9.1	Alignment . . . . .	52
2.4.9.2	Classification . . . . .	54
2.4.10	Model Building . . . . .	57
2.4.10.1	Projection Matching . . . . .	57
2.4.10.2	Angular Reconstitution . . . . .	57
2.4.10.3	Random Conical Tilt . . . . .	58
2.4.11	Reconstruction . . . . .	59
2.4.12	Refinement . . . . .	60

---

2.4.13	Resolution Assessment . . . . .	61
2.4.14	Validation . . . . .	64
2.4.15	Conformational Analysis . . . . .	64
2.4.15.1	Detection of conformational heterogeneity . . . . .	64
2.4.15.2	Analysis of large amplitude movements . . . . .	65
2.4.15.3	Analysis of large domain movements . . . . .	66
2.4.15.4	Analysis of small amplitude movements . . . . .	67
2.5	Sample Preparation for Electron Microscopy . . . . .	68
2.5.1	GraFix . . . . .	68
2.5.2	Negative Staining Preparation . . . . .	68
2.5.3	Vitrification . . . . .	69
2.6	Imaging . . . . .	69
2.7	Image Processing . . . . .	69
2.7.1	Micrograph Analysis and Particle Picking . . . . .	69
2.7.2	CTF-Correction . . . . .	70
2.7.3	2D-Processing . . . . .	70
2.7.4	Initial Model Building . . . . .	70
2.7.5	Conformational Sorting . . . . .	71
2.7.6	Refinement and Resolution Determination . . . . .	71
2.7.7	Structural assessment . . . . .	71
<b>3</b>	<b>Results</b>	<b>73</b>
3.1	Method Development . . . . .	73
3.1.1	ProteoPlex . . . . .	73
3.1.1.1	Theoretical Description of Thermofluor Curves Obtained from Single Domain Proteins . . . . .	75
3.1.1.2	Theoretical Unfolding Behavior of a Protein Complex . . . . .	78
3.1.1.3	An Heuristic Approach used for Curve Interpretation . . . . .	81
3.1.1.4	Proof of Principle . . . . .	82
3.1.1.5	Binding of Stabilizing Ligands . . . . .	84
3.1.1.6	Reconstitution and Dissociation of Complexes aided by ProteoPlex . . . . .	86
3.1.1.7	Automation . . . . .	87
3.1.1.8	ProteoPlex and Crystallization . . . . .	89
3.1.1.9	ProteoPlex: Data Overview . . . . .	90
3.1.2	Modifications of the GraFix methodology . . . . .	92
3.1.3	<i>Chaetomium thermophilum</i> . . . . .	93
3.2	Structural and Dynamical Insights into CRM1 . . . . .	96
3.2.1	<i>Chaetomium thermophilum</i> CRM1 . . . . .	96

3.2.2	Human CRM1 . . . . .	101
3.3	Structure and Dynamics of the <i>E.coli</i> Pyruvate Dehydrogenase Complex	104
3.3.1	Purification of <i>E.coli</i> PDH . . . . .	104
3.3.2	Reconstitution Trials . . . . .	106
3.3.3	Structural Analysis of <i>E.coli</i> PDH . . . . .	107
3.4	Structure of the <i>Biomphalaria glabrata</i> Acetylcholine Binding Protein . .	111
<b>4</b>	<b>Discussion</b>	<b>115</b>
4.1	Methodological Progress . . . . .	115
4.1.1	ProteoPlex . . . . .	115
4.1.2	Modifications of the GraFix Methodology . . . . .	118
4.1.3	<i>Chaetomium thermophilum</i> . . . . .	119
4.2	Conformational Landscape of CRM1 . . . . .	121
4.3	Structural Investigation on the <i>E.coli</i> Pyruvate Dehydrogenase Complex	124
4.4	Conformational Flexibility of the <i>Biomphalaria glabrata</i> Acetylcholine binding protein. . . . .	126
<b>5</b>	<b>Summary and Outlook</b>	<b>127</b>
<b>Appendices</b>		
<b>A</b>	<b>Abbreviations</b>	<b>I</b>
<b>Bibliography</b>		<b>V</b>
<b>Related Publications</b>		<b>XXXIII</b>
<b>Curriculum Vitae</b>		<b>XXXV</b>



# List of Figures

1.1	The Cell as a Factory . . . . .	2
1.2	Parts of a human and nature built machine. . . . .	3
1.3	Time Scales of Protein Dynamics and their Processes . . . . .	5
1.4	Three Examples for Protein Dynamics . . . . .	6
1.5	Schematic of an Energy Landscape . . . . .	10
1.6	Time Scales of Protein Dynamics and their Processes . . . . .	11
1.7	Single Particle cryo EM Workflow . . . . .	18
1.8	Structural Features of CRM1 . . . . .	25
1.9	Catalytic Cycle of the Pyruvate Dehydrogenase Complex . . . . .	26
1.10	Structural Architecture of the E.coli PDHc. . . . .	28
2.1	Scheme of a TEM . . . . .	42
2.2	Most Important Aberrations for a TEM . . . . .	44
2.3	Use of Fourier Filters . . . . .	46
2.4	Effect of Defocus . . . . .	49
2.5	Alignment and Classification . . . . .	52
2.6	Demonstration of Model Bias . . . . .	54
2.7	Exemplified Principal Component Analysis . . . . .	56
2.8	What Detail can be seen at which Resolution? . . . . .	61
2.9	Principal Fourier Shell Correlation Curves . . . . .	62
3.1	Schematic representation of possible unfolding scenarios . . . . .	74
3.2	Temperature Fluorescence unfolding curves of a three-subunit complex . . . . .	79
3.3	Evaluation of ProteoPlex fits . . . . .	80
3.4	ProteoPlex Proof of Principle SelA . . . . .	83
3.5	ProteoPlex proof of Principle BgHb . . . . .	84
3.6	ProteoPlex for identifying stabilizing Ligands . . . . .	85
3.7	ProteoPlex assisted reconstitution of PDHc . . . . .	87
3.8	Destabilisation of PDH . . . . .	87
3.9	Automation Set-up and Reproduceability . . . . .	88
3.10	Crystallisation Success after ProteoPlex . . . . .	90
3.11	pH Distribution of optimal Buffers . . . . .	91
3.12	Behavior of different crosslinkers for GraFix . . . . .	92
3.13	Growth Conditions of <i>Chaetomium thermophilum</i> . . . . .	94

---

3.14	Test Purification of <i>Chaetomium thermophilum</i> complexes . . . . .	95
3.15	<i>Chaetomium</i> CRM1 Raw Data . . . . .	97
3.16	<i>Chaetomium thermophilum</i> CRM1 conformations . . . . .	98
3.17	Mode 7 of NMA of CRM1 . . . . .	99
3.18	Energy Landscape <i>C.thermophilum</i> CRM1 . . . . .	100
3.19	Significance of the C-terminal Helix . . . . .	101
3.20	Human CRM1 Raw Data . . . . .	102
3.21	Human CRM1 Conformations . . . . .	102
3.22	3D-MSA of human CRM1 . . . . .	103
3.23	PDHc Purification . . . . .	105
3.24	Micrographs from PDHc Reconstitution Trials . . . . .	106
3.25	PDHc Raw Data and Initial Class Averages . . . . .	107
3.26	Initial 3D Models of PDHc . . . . .	108
3.27	Exemplified Conformational Change of PDHc . . . . .	109
3.28	Effect of CryoFix on the PDHc . . . . .	109
3.29	Initial Analysis <i>Biomphalaria glabrata</i> Acetylcholine Binding Protein . . . . .	111
3.30	Refined Model of AChBP at 4 Å . . . . .	112
3.31	Local Resolution Map of AChBP . . . . .	112
3.32	Conformational fluctuation of AChBP . . . . .	113
4.1	Functional Cycle of CRM1 . . . . .	122

# List of Tables

1.1	Physical Parameters in the Microscopic World . . . . .	7
1.2	Overview about the three main Methods of Structural Biology . . . . .	13
2.1	Software used in this thesis . . . . .	31
2.2	Special Equipment used in this thesis . . . . .	31
2.3	Chemicals used in this thesis . . . . .	32
2.4	Chemicals used in the ProteoPlex Buffer Screen . . . . .	36
2.5	Chemicals used in the ProteoPlex additive screen . . . . .	37
2.6	GraFix conditions for the analyzed samples. . . . .	68
3.1	Correlation between Crystallization Buffer and ProteoPlex Buffer . . . . .	89
3.2	Origin of optimized Samples . . . . .	90
3.3	Sizes of optimized Samples . . . . .	90
3.4	Subcellular Localization of optimized Samples . . . . .	90
3.5	Function of the optimized Complexes . . . . .	91



# Acknowledgements

The past four years have been an incredible experience in my life. They formed me scientifically and also as a person. That and this thesis itself have been made possible through many nice people in and around the lab to whom I am deeply grateful.

First of all and mostly, I would like to thank Holger Stark for introducing me to the world of structural biology. His never ending support and help were one of the motors of this thesis. He not only took his time for supervise me but also to many non-scientific chats. His everlasting enthusiasm was a real source of motivation as well as the sweets he used to bring for the lab members including myself. I cannot imagine that someone could better fulfill the metaphorical role as "Doctor father" and I will always be grateful to him.

Secondly, I am profoundly grateful to Ashwin Chari, who always emphasized that he is not my supervisor, but definitely supported me in terms of biochemical topics. We definitely had our ups and downs, but without his support and patience regarding my questions and problems this thesis would not have been possible. Additionally, I am thankful towards all of the current and past members of the Stark lab, which I consider not only my colleagues but also my friends. All of them helped me a lot during the time as a PhD student and made my time in the lab enjoyable. Florian Hauer and Niels Fischer were my first supervisors in the group who taught me most of the beginner's knowledge with much patience and humour. Prakash Dube introduced me into many different EM techniques and was supportive in any situation. Additionally, he got me to eat Indian cuisine for the first time in my life, and I quite enjoyed everything he prepared. One of the biggest thanks goes to the "IT guys": Mario Lüttich, Boris Busche, Jan Martin Kirves, Georg Bunzel and Martin Schmeißer. All of them were very supportive and helped with any computational problem as stupid as it might be. I am enormously thankful for all the things they did "especially" for me. No matter how complicated the request was, all of them took their time to help and support me. Frank Würrihausen and Karl-Heinz Knauber always helped with any technical problems on machines and computers. Further, I am very happy that I got the chance to teach and support others. I am deeply proud on my two very talented Masters students Jan Erik Schliep and Jil Schrader, who took most of my requests and advice seriously and hopefully, learned a lot, which hopefully will help them in the future. Additionally, I am thankful to Michael Hons and Stephanie Schell, who were helpful office mates. Further, I would like to thank the new members of the group Fabian Henneberg, Jan Kaesler and Lukas Schulte for keeping up the good spirit throughout the group. I

further would like to acknowledge Dietmar Riedel, for his help in EM and bike issues and the delicious syrups he used to bring. Especially his quick and dirty solutions always saved lots of time and never failed.

My special thanks also go out to Juliane Moses, who was the best secretary one could ever wish for. Last but not least, I would like to thank Wen-ti Liu, who first was my fellow student and friend and additionally became one of the most supportive and most talented colleagues. I cannot be grateful enough for all the nice discussions in long lab hours and the awesome pictures she drew for me and the lab.

Also, outside of the Stark lab, I am very grateful for many people who supported and helped with my work. First to mention is Kai Tittmann and his group, with whom I worked on the pyruvate dehydrogenase complex, who was always enthusiastic and helpful and took his time to be in my thesis advisory committee, where his good hints were a great help. Further, I would like to thank all the members of his lab, but especially Florian Brodhun, who is almost a member of the Stark lab with all his kindness and hard work.

I am also grateful to Jörg Enderlein, who was the third member of my thesis advisory committee and always contributed from the physical point of view.

Further, there is a large number of collaborators who also were always helpful partners: without any order I am grateful to: Thomas Monecke, Achim Dickmanns, Ralf Ficner, Ralph Kehlenbach, Sarah Port, Danilo Meyer, Irene Oechsner, Gabi Heyne, Hossein Kohansal, Thomas Conrad, Ulrich Steuerwald Winfried Lendeckel, Monika Raabe, Henning Urlaub, Frank Peske, Ingo Wohlgemuth, Franziska Hummel, Michael Zimmermann and many more.

I also would like to thank Steffen Burkhardt, Kerstin Grüniger and Ivana Bacakova for all the support they provided through the International Max- Planck Research School.

Additionally, I am deeply honored to have received a fellowship funded by the Boehringer Ingelheim Fonds which provided both, generous financial and personal support.

In the end, I want to thank all my friends. Of the people who supported me throughout the years, I like to mention especially: Claudi, Kristina, Caro, Carsten, Mirko, Felix, Felizitas, Caro and Ingo whom I knew before I was coming to Göttingen. Within Göttingen I would like to thank Lena, Sinem, Koray, Chepe, Melanie, Christian, Natalia, Aaron, Helena, Iris, Debora and Sebastian. All of them contributed enormously to an exciting life in Göttingen. Especially I would like to thank Cadu, who was already a good friend even before we met in Göttingen. He is a great adviser who taught me much of the simplicity of being and a very good friend. I also would like to thank the other two P and C members, Simone and Jenny who also became good friends.

Moreover, I would like to thank Lisa Upmann, a lovely and wonderful young woman, whom I had the honor just to meet recently. From the first day we met she supported me in my nerdiness and made me smile in so many ways.

Last but not least, I thank my family being my mum and my grandparents, who supported me in everything I ever did without any questioning since I can remember.





# Abstract

Life on earth is only possible through the enormous capabilities of proteins and their assemblies. Which function a certain protein fulfills is encoded in its amino acid sequence, which gives rise to a defined structure. However, this structure is not static. The thermal energy of the surrounding medium forces the molecule into different conformations. It can be assumed that in most cases these movements play an important role for a protein's function. Therefore, it is crucial to gain high-resolution 3D information of those movements. For small proteins, Nuclear Magnetic Resonance (NMR) Experiments can provide this information in good detail. For larger proteins and protein complexes single particle Electron cryo Microscopy (cryo EM) is the method of choice, which is used in this thesis. However, analysing the complete conformational landscape is not yet routine for cryo EM. Therefore, several methodological developments were made with this thesis.

Since the most crucial prerequisite for any structural analysis is an intact and homogeneous sample, the development of a method, finding stabilizing conditions was the first aim. The new method called *ProteoPlex* uses a newly developed extended theoretical framework on the existing fluorescence-based stability screen called *Thermofluor*. Therewith, Thermofluor data obtained from large multidomain proteins and protein complexes can be analyzed. In total, stabilizing conditions could be found for more than 80 complexes from all branches of life. Additionally, the usefulness of ProteoPlex towards assembly and disassembly experiments was demonstrated.

Furthermore, two strategies were employed to decrease the number of adopted conformations by decreased temperature. Firstly, a crosslinking strategy at -10 °C was successfully used. Secondly, the fermentation of the thermophile fungus *Chaetomium thermophilum* was established, and the native purification of thermophile protein complexes was successfully demonstrated.

To analyze conformational dynamics in practice three model systems were chosen: the 120 kDa single-chain nuclear export factor chromosome region maintenance factor 1 (CRM1), the 3-5 MDa bacterial pyruvate dehydrogenase complex (PDHc) and the icosahedral, 1.5 MDa Acetylcholine binding protein (AchBP) from the snail *Biomphalaria glabrata*.

CRM1 is with 120 kDa rather small for electron microscopic analysis. Nevertheless, it was demonstrated for *Chaetomium thermophilum* CRM1, that the apo protein cycles between an open superhelical and a closed ring-shaped conformation. Hereby, the full energy

landscape of this movement could be described. The energy landscape is in general flat, and CRM1 can change freely between the open and closed conformation. Furthermore, a C-terminal helix is responsible for a slight enthalpic stabilisation of the closed state.

The *E.coli* PDHc could be purified in large quantities from native source. Literature describes the overall structure to be of octahedral symmetry with a cubic core made of the E2 component surrounded by a cubic shell made of the E1 and E3 component. This symmetry could only be partially confirmed. While the core indeed seems to be cubic, no overall cubic model could be obtained. However, also no high-resolution structure could be calculated. It can be hypothesized that the overall structure is very flexible and thus structural investigation is largely hindered. Nonetheless, crystals could be obtained which could lead to structural insights into the complex in the future.

For the AChBP, a 3.6 Å resolution structure could be calculated and two conformational substates could be identified. The different states reveal significant differences in loop regions, subunit interfaces and even in  $\beta$ -strands. Here, it is the first time that such small fluctuations could be visualized by cryo EM.

In summary, this thesis provides new techniques and approaches towards the elucidation of the conformational landscape of large proteins and protein complexes.

**Keywords:** cryo-electron microscopy, single particle image processing, structural dynamics of protein complexes

# Chapter 1

## Introduction

### 1.1 Molecular Machines

*Machines take me by surprise with great frequency.*

— Alan Turing

Modern human life would not be imaginable without a plethora of machines and tools. As soon as one gets up in the morning the first machine is used - be it a coffee machine or an electric toothbrush.

Machines are ubiquitous for us since a couple of centuries. Nature, however, masters them already for millions of years. Especially in nanotechnological machines, Nature excels us in an almost inimitable way. Every single cell is packed with biological nanomachines (termed *molecular machines* or *protein complexes*). As in a factory these machines work together to establish a goal (see figure 1.1) such as building a certain molecule or reacting to a certain stimulus [1]. Astonishingly, they are built only from a few different materials, mostly of proteins, or proteins in combination with ribonucleic acid chains.

Already the variety of tasks those machines can fulfil is fascinating. Cellular machines can precisely break and form chemical bonds, detect and collect single photons, transport molecules or even vesicles for micrometers and even build and destruct such machines. Some of the most studied machines are involved in the life cycle of a protein, which is illustrated in figure 1.1. First, a machine called the *RNA-polymerase* copies a gene from the DNA into a transportable form, the mRNA. This mRNA is then translated into a polypeptide chain on the *ribosome*. Directly after protein synthesis another set of machines -the *chaperones* (called e.g. GroEL, HSP60) assist the polypeptide chain to find its three-dimensional structure. At the very end of the life cycle, the protein is finally degraded by a machine called the *proteasome*.

How these machines accomplish this seems to be an incomprehensible and daunting task on first sight. However, the effort of understanding them is necessary to ever fully un-



**Figure 1.1: The Cell as a Factory.** Like in a medieval household (left) several machines are used to keep a hypothetical simplified cell going (right). Depicted are the RNA-polymerase (blue/green) which translates the DNA (orange) into mRNA (pink). The mRNA is translated into a protein (white) by the ribosome (yellow/blue). Finally, the protein is folded by a chaperone (purple/pink). The energy and metabolites used for that are produced by many metabolic enzyme complexes. Here, the Pyruvate Dehydrogenase complex (yellow/green), the Fatty Acid Synthetase (green) and the F-ATPase (purple/yellow/red) are depicted.

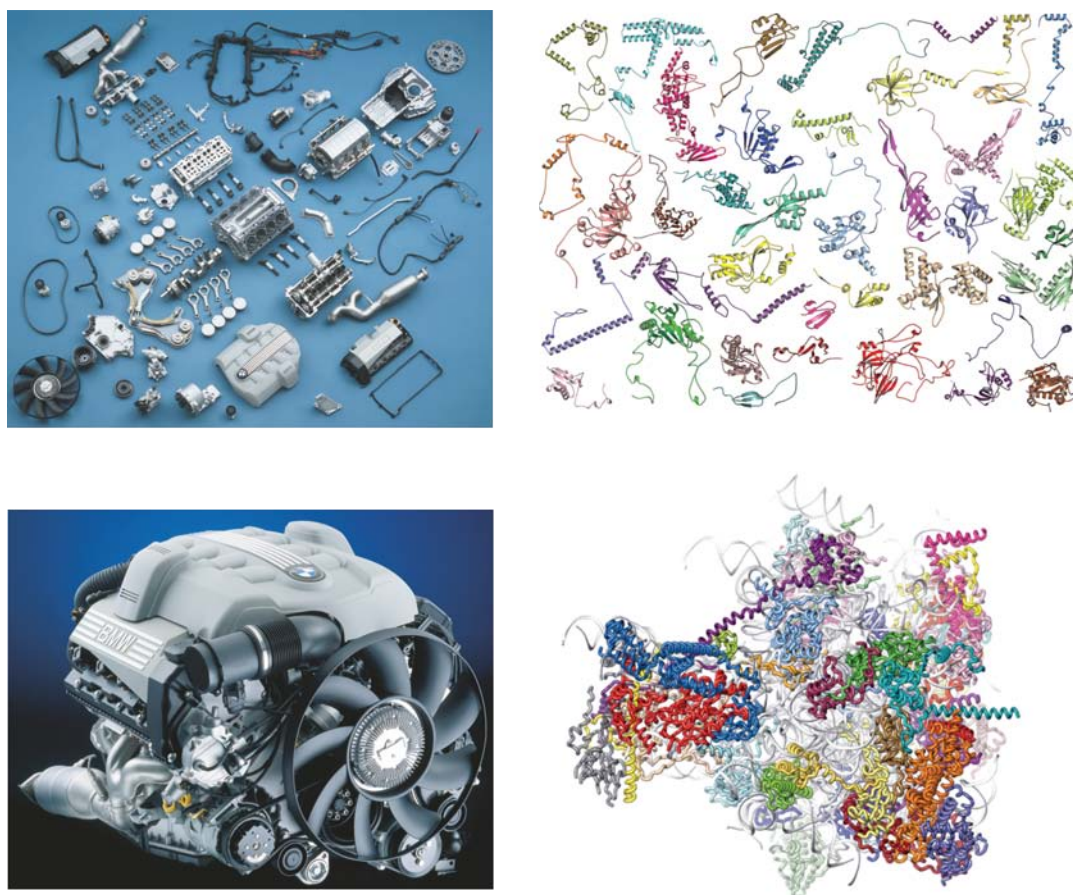
derstand cells or organisms. Actually, in many ways they can be understood as their macroscopic human-made counterparts.

It seems natural to approach an understanding of a cellular machine as an engineer would analyze a human-made machine. He would first look for modules of distinct function like a power supply, mechanical parts, electronic parts or thermal parts of which any machine is made of. Those components are commonly built in a way enabling their independent use in different machines with completely different purposes. Thus, it should be an easy task to understand the full machine after the components or modules are identified. In a car, for instance, there is the engine which produces mechanical movement, the gear unit transferring this movement to the wheels and electronic parts controlling the whole process.

Similar features can be seen in protein machines [62]. The 26 S proteasome, for instance, can be analyzed in the same manner. This protein complex degrades unwanted proteins. Therefore, it is built out of three parts: a destruction chamber - the 20 S core particle, catalyzing the degradation itself, ubiquitin binding proteins recognizing the target and a AAA+-ATPase unfolding the target protein [51]. Following the engineer's approach, one should be able to understand how those parts work together to fulfil the complexes function. Moreover, those or very similar parts are used in other machines. For instance,

a homologue of the mentioned ATPase is used to rescue ribosomes from stalled protein chains [118] while another one disassembles proteins needed for vesicle fusion [79].

However, this is not the entire story. The 26S proteasome, for instance, has many more components and modules, but for only few the function is known [51]. On top of that the modularity is not common to all such machines. In any way, the next logical step in the understanding of the machine needs to be the understanding of the components of the module and how they act together to fulfil the module's function. At some point, it will be very difficult to understand the significance of a certain part of the modules or the machines in an isolated form. Even if one sees all the parts of a car motor, it is hard to imagine how the full machine looks like and even harder how it functions (Fig. 1.2).



**Figure 1.2: Parts of a human and nature built machine.** Left: The parts of a human-built automotive engine in a disassembled and in an assembled state are shown. Right: In comparison, the parts of a cellular machine – the ribosome are shown in a disassembled representation and an assembled representation. Taking a random part from either of them and trying to understand it on its own is a daunting task. Also, guessing its function for the whole machine will not be possible without detailed knowledge about the architecture of the full machine. (Images with permission taken from BMW Group (left) and [194] (right))

The same holds true for most of Nature's machines. Key to a modest understanding can, therefore, only be the investigation of the full machine. This approach however

is technically quite challenging and cannot be done in a straightforward manner using routine methods. As human-built machines, molecular machines are built from several different parts. In most cases, they are complexes of 10 or more different biomolecular chains (mostly proteins) [62], which hold together by rather weak forces and have to be handled with great care. Since this is a prerequisite for any analysis, this thesis is regarded, first of all, to the development of new methods to handle such machines with best possible care.

It is also very necessary to investigate a machine while it is in action. Only looking on a picture or a static model of a motor like in figure 1.2 will only give vague clues about its function. However, seeing it moving will quickly lead to an understanding of its purpose as well as the significance of its individual parts. However, do all biological machines need to move? For many complexes, it is rather clear that they have to move. Motor proteins like *kinesin* for instance have to move cargoes from point A to B, ribosomes move along an mRNA chain shifting tRNAs through their interior, proteasomes have to unfold proteins and translocate them into their interior, GroEL has to do that in the opposite direction. For complexes such as storage proteins like *myoglobin* [56], which binds oxygen in muscle cells or enzymes like *dihydrofolate reductase* [18] which catalyzes the exchange of single atoms, this does not seem to be obviously necessary but many studies still propose significant and necessary movements in them. As described in detail in the next section (1.2), the chemical nature and the physical environment will induce movements in any given protein or protein complex. Thus, being a general feature of molecular machines, the present thesis is an attempt to analyze the movements within them experimentally and, therefore, contribute to their understanding.

## 1.2 Conformational Dynamics

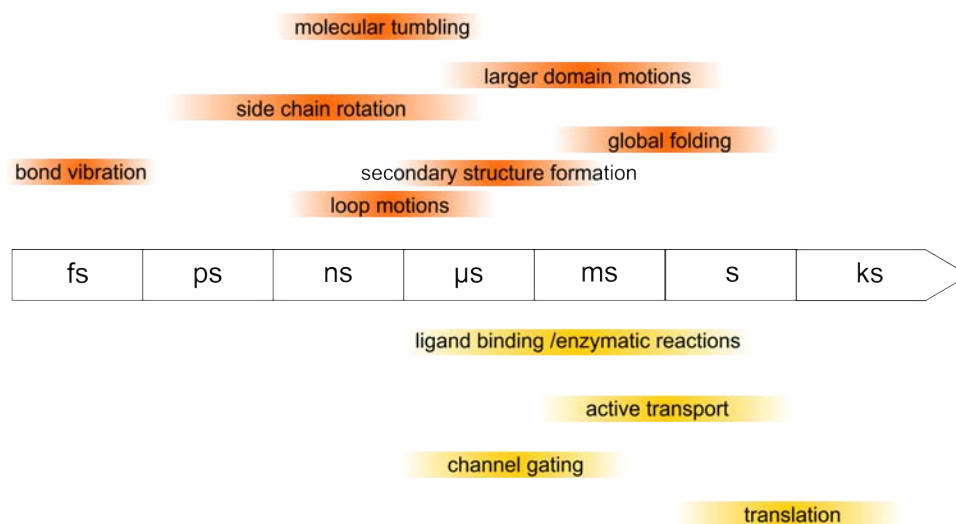
*Everything that living things do can be understood in terms of the jiggings and wiggings of atoms.*

— Richard Feynmann, *Lectures on Physics*

### 1.2.1 What are Conformational Dynamics?

Biological textbooks and many schemes in the modern scientific papers illustrate most proteins and protein complexes as static and solid, almost stone-like objects, only moving as complete units via diffusion through space. As pointed out in the previous sections, this does not reflect the reality. Intrinsic movements determine a proteins structure and shape right from the start of its life time. As soon as a linear polypeptide chain is produced by

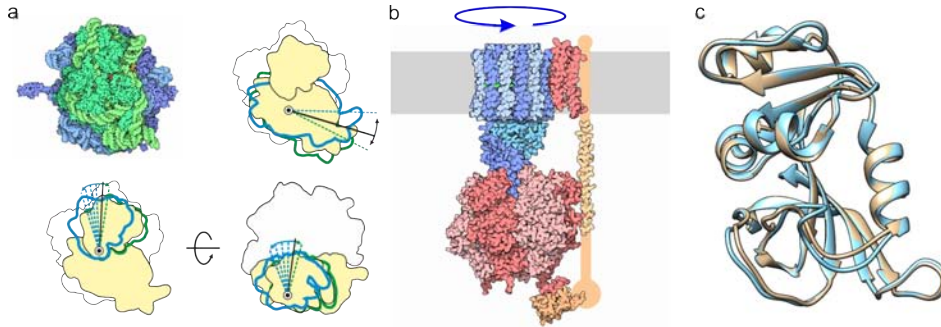
the ribosome it starts folding into a defined three-dimensional structure. This, however, is not a stable structure but rather a continuum of different states. In principle, every part of a protein can move. Thus to circumvent confusion I will use the term "conformational dynamics" for *any change in atomic coordinates of a given molecule over time*. However, it has to be noted that the characteristic time and length scales for different conformational movements span over many orders of magnitudes and are consequently quite different in their physical nature [88, 142]. The tiniest relevant movements are bond vibrations that displace atomic nuclei only for a few picometers within the time frame of picoseconds. On the other end of the scale, there are the displacements of large protein domains or subunits for several nanometers that takes mostly milliseconds or even seconds of time. A comprehensive overview over the different time scales is given in figure 1.3.



**Figure 1.3: Time Scales of Protein Dynamics and their Processes.** A comprehensive overview of the time scales of individual intrinsic movements is shown in orange. The time scales of important biological processes are depicted in yellow. The figure is based on [88, 142].

However, how relevant are those movements? First experimental evidence that even small thermal movements are relevant to a proteins function are given by Rasmussen et al. in 1992 [164]. They showed that at a certain temperature, where non-harmonic movements are diminished, the pancreatic enzyme RNase A does not work anymore, while it is still structurally intact. Similar studies were made for protein complexes like the ribosome that slows down translation below a certain temperature while keeping an intact structure [66]. What can be learned from this is, that dynamics of higher energy are needed to enable a protein to perform its function. Figure 1.3 shows that the time scales of biological relevant processes like enzymatic catalysis or translation mainly correlate with large domain movements. This makes them the most interesting movements for functional analysis of protein complexes. However, one has to keep in mind that those larger movements are, of course, based on many tiny steps on shorter length and time scales.

Again biological literature minimizes the multitude of different conformational dynamics often to "conformational changes" meaning the transition from one compact folded state to only one or a few others. Moreover, transitions are regularly described as fast, and direct transitions termed "power strokes". Apart from the tremendous simplifications in many textbooks, scientific literature describes examples for the continuous movements of proteins relevant to their function.



**Figure 1.4: Three Examples for Protein Dynamics.** (a) The relevant dynamics of the 70 S ribosome are shown. Its largest movement is the rotation of the small subunit relative to the large one. This movement is called ratcheting. The second largest movement is the so-called head swivelling which is the rotation of the small subunits head domain perpendicular to the ratcheting rotation. These movements are supposed to cause translocation of the tRNAs and thus elongation of the emerging protein chain. Schemes with permission taken from [52] (b) The main conformational change seen for the F-type ATPase is illustrated. Blue membrane anchored part can rotate relatively to the red part and thus cause conformational changes in the red head domain. This leads to ATP production. However, it is known that this process can work backwards as well (c) Structural overlay of two conformational states of DHFR (PDB structures 1rx2 and 1rx6 from [18]) shows two key intermediates of DHFR dynamics only slight movements of loops are necessary to catalyse the reaction. The overall structure does not change significantly.

Figure 1.4 illustrates a few of the most prominent examples. Besides motor proteins such as myosin, kinesin or dynamin, the ribosome and the F-Type ATPase are conformationally well characterized examples: Both are central to all life on earth. During translation the ribosome moves over a mRNA template, decoding its information into a newly synthesized polypeptide chain. Three obvious and directed movements have to be undergone: The displacement of the mRNA into the ribosome, the movement of tRNAs as carriers of the amino acids through the ribosome and the pushing of the newly synthesized chain out of the ribosome. These movements can be very well correlated with two more abstract movements, which look unrelated at first sight: Firstly, the small and the large subunit of the ribosome can rotate against each other in a process called ratcheting. Secondly, the head domain of the small subunit can swivel. Many studies directly correlate these conformational movements to the previously mentioned functional movements.[132]

The second well studied example - the F-type ATPase produces ATP powered by a proton gradient. A constant flow of protons through membrane part of the molecule rotates its soluble part. A rigid stator domain thereby forces the moving parts into new conformations and enables therewith the formation of ATP from ADP.[138]



The last example that needs to be mentioned here is the Dihydrofolate reductase (DHFR). This enzyme catalyses the reaction from folic acid to dihydrofolic acid or from dihydrofolic acid to tetrahydrofolic acid. In contrast to the previous two examples, a movement of any kind does not seem to be obviously required for this redox reaction. Further, the structural studies suggest that there are no large domain movements occurring. However, as seen in many recent studies, the tiny fluctuations occurring in the nanosecond to microsecond time scale seem to be concerted and act on the catalytic cycle of the enzyme[212]. These fluctuations seem to be even evolutionarily conserved.

Those three examples show that there is a broad relevance in the conformational dynamics of proteins and protein complexes for their function. Therefore, the fourth dimension necessarily needs to be added to the static molecular description often found in biology textbooks. An understanding of the nature of conformational dynamics requires a physical framework as outlined in the next section.

## 1.2.2 Physics of Conformational Dynamics

Almost everyone has experienced a tempest. Wind blows from seemingly every direction and pushes the unfortunate person from one direction to another. Especially if the person carries an umbrella, the storm might even be able to push arms independently from the body in some direction.

Protein complexes experience a similar environment. In the microscopic world gravity and inertia do not play any role but the Brownian motions induced by the surrounding medium and thus Stokes' law dictate the behaviour of the molecular complexes (see table 1.1). Coulomb forces that are negligible at the meter scale are the significant force given the nature of the medium [204]. At this point and onwards the analogies with the macroscopic world are not sufficient anymore to understand the functioning of molecular machines.

**Table 1.1: Physical Parameters in the Microscopic World.** The physical parameters defining the microscopy environment are exemplified for a 70 S ribosome at a temperature of 310 K. Numbers are mostly analogues to [132].  $g$  is the gravitation acceleration of  $9.81 \frac{m}{s^2}$  and  $\eta$  the viscosity of the medium.

radius $r$	$9 \cdot 10^{-9} m = 9 nm$
mass $m$	$2.7 MDa = 4.4810^{-21} kg = 4.5 ag$
density $\sigma$	$1.67 \frac{g}{cm^3}$
gravitational force ( $F = mg$ )	$4.4 \cdot 10^{-20} N = 44 zN$
frictional coefficient ( $\gamma = 6\pi\eta r$ )	$1.63 \cdot 10^{-10} \frac{Ns}{m} = 163 pJs$
diffusion coefficient ( $D = \frac{kT}{\gamma}$ )	$2.63 \cdot 10^{-11} \frac{m^2}{s}$
average velocity ( $v = \sqrt{\frac{3kT}{m}}$ )	$1.66 \frac{m}{s}$
frictional force ( $F = \gamma v$ )	$270 pN$

Solutions of complex polymers, as which cells can be considered to some extent, can be described as neither liquid nor solid. This behavior is caused by the fact that biopolymers are mesoscopic, meaning they are far larger than the atomic scale but still too small for a macroscopic description. To describe these solutions accurately, classical mechanics and physics are not sufficient. Therefore, a new field of Physics, the Soft Matter Physics, founded by Pierre Gilles de Gennes<sup>1</sup>, is necessary to describe them.

While soft matter physics is a complex field, only a few relevant features of soft matter objects have to be stated here: The forces between individual structural parts are weak and thus enthalpic contributions are comparably low. Entropic effects, however, are strong due to a large number of degrees of freedom. Therefore, an intricate balance between these contributions strongly influences the free energy [118]. Thus, the structures underlie big thermal fluctuations and exhibit stochastic trajectories. Moreover, due to the constant temperature within most cells (within reasonable time scales) the machines have to have isothermal engines that operate far from their thermal equilibrium. [112]

With these properties in mind, one can start to build up a general understanding of the movements of such machines. After the scales of the interesting motions were defined (see figure 1.3), one has to raise the question that effects drive such a motion. The average answer of most biology or even biochemistry textbooks is ATP. The release of the energy stored within the phosphodiester bond between the  $\beta$ - and  $\gamma$ -phosphates should be a major driving force for most biological processes and thus for many conformational changes.

At any temperature, the surrounding medium has an average kinetic energy of  $k_B T$  which is at room temperature about  $4 \cdot 10^{-21}$  J. This can very well be compared with a strong tempest, considering that the most used energy source by proteins is ATP-hydrolysis, which releases only  $5 \cdot 10^{-20}$  J. This is only an order of magnitude higher than the thermal noise. This effect becomes more drastic considering the power ATP cleavage can produce. A typical ATPase cleaves 100-1000 ATP molecules per second yielding  $10^{-16}$  to  $10^{-17}$  W. With relaxation times of  $10^{-13}$  s the thermal power exceeds this with about  $10^{-8}$  W by more than eight orders of magnitude [9]. Therefore, it is clear that ATP hydrolysis alone can hardly be the driving force for most relevant structural changes. Moreover, there are many examples of dynamic molecules that do not even directly interact with any nucleotide that could explain their motions. In more general terms, it is hardly believable that any chemical source can provide enough energy to overcome the overwhelming thermal noise. However, knowing that, how can directed motions, as observed in experiments, be possible? A simple but still accurate description

---

<sup>1</sup>The Nobel Prize in Physics 1991 was awarded to Pierre-Gilles de Gennes "for discovering that methods developed for studying order phenomena in simple systems can be generalized to more complex forms of matter, in particular to liquid crystals and polymers".

is necessary. A physical model, termed the energy landscape is the most accurate approach for understanding thermal fluctuations. It is a multidimensional construct resembling in its 3-dimensional form to a mountainous landscape. Height describes energy or occupation probability. Valleys are states of low energy and, therefore, describe conformational states. The mountains in between describe the energy needed to change from one conformation to another.

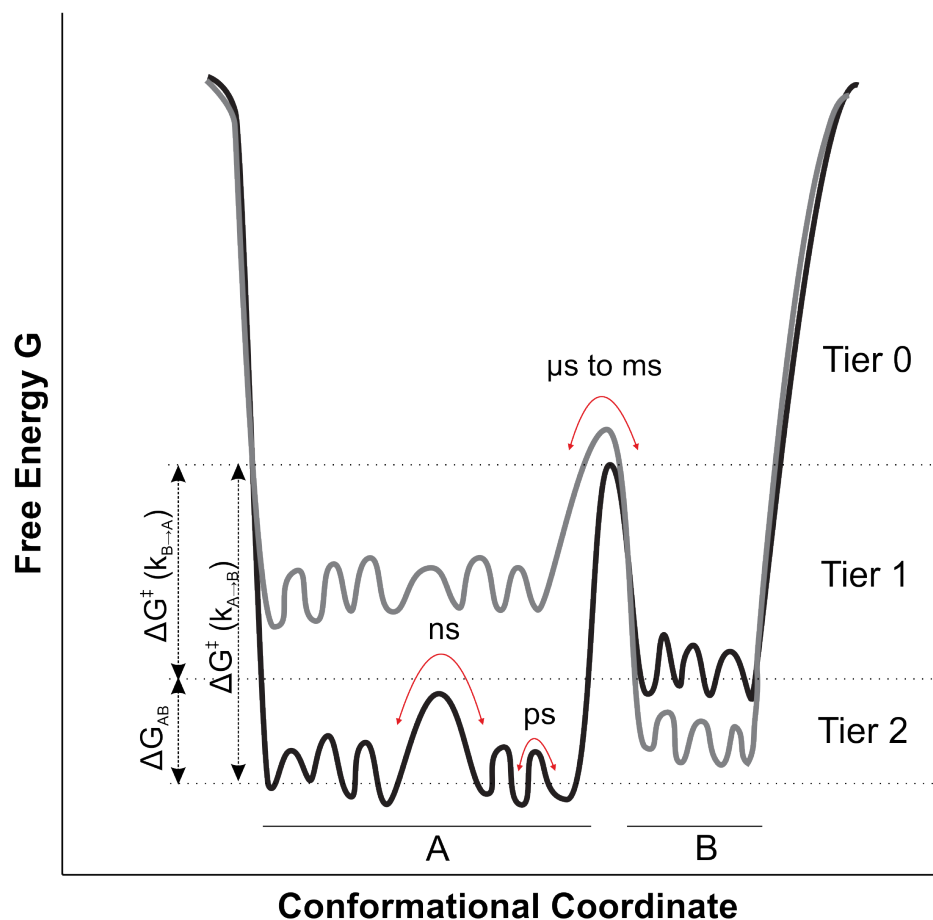
This concept is well known by biologists in terms of protein folding, where the energy landscape resembles more a canyon or simply a funnel [143]. Less known is that this concept was already used for almost 40 years to describe the conformational dynamics of proteins [10]. Hans Frauenfelder, the pioneer of conformational dynamics and his co-workers, could explain myoglobin's function and dynamics with the aid of an energy landscape, with almost 20 years of research. They were able to explore the full landscape including the heights of the energy barriers between the conformational states [56].

Figure 1.5 shows a principle 2-dimensional representation of such an energy landscape. The diagram itself is a powerful tool since it holds information on thermodynamics and kinetics of a sample. It can be constructed from all different kinds of data: either from kinetic data describing the lifetime of a certain conformational state, respectively the rates of the transition from one state to another or from thermodynamic data such as the number of molecules in a certain state at a given time point: The free energy change  $\Delta\Delta G^\circ$  in units of thermal noise  $k_B T$  can be calculated as:

$$\Delta\Delta G^\circ = k_B \cdot T \ln \frac{p_i}{p_0} \quad (1.1)$$

where  $p_i$  is the probability of finding the protein in conformational state  $i$  and  $p_0$  the probability for finding it in its ground state corresponding to the state with the highest occupancy [15]. The probabilities can then either be expressed through the absolute numbers of molecules in a given state or their mean occupancy times. However, one always has to keep in mind that a certain energy landscape is only true for a certain set of conditions (being pressure, solvent and temperature) while those are still the primary determinants of the landscape, and the molecules behaviour in it [88]. They determine how well and in which time scale a molecule can overcome a certain energy barrier. Therefore, those factors are, of course, also major experimental tools to shape and change the landscape to simplify the system for analysis.

Having constructed such a complex abstract model, what can be learned from it? First of all, in contrast to the energy landscape in protein folding, the conformational landscapes of molecular machines have a rather flat surface with a somewhat rough structure [212, 88, 52]. This means only a few energy equivalents of the thermal noise (described



**Figure 1.5: Schematic of an Energy Landscape.** The cross-section through a typical energy landscape is shown. A hierarchical representation of the energy levels (tiers) was chosen as suggested by Fraunfelder [6]. In Tier 0 high energy is needed to overcome the energy barrier from state A to B. Therefore, these transitions are seldom and only occur on a  $\mu\text{s}$  to  $\text{ms}$  time scale. In higher tiers energy barriers become smaller as well as the necessary time scales.

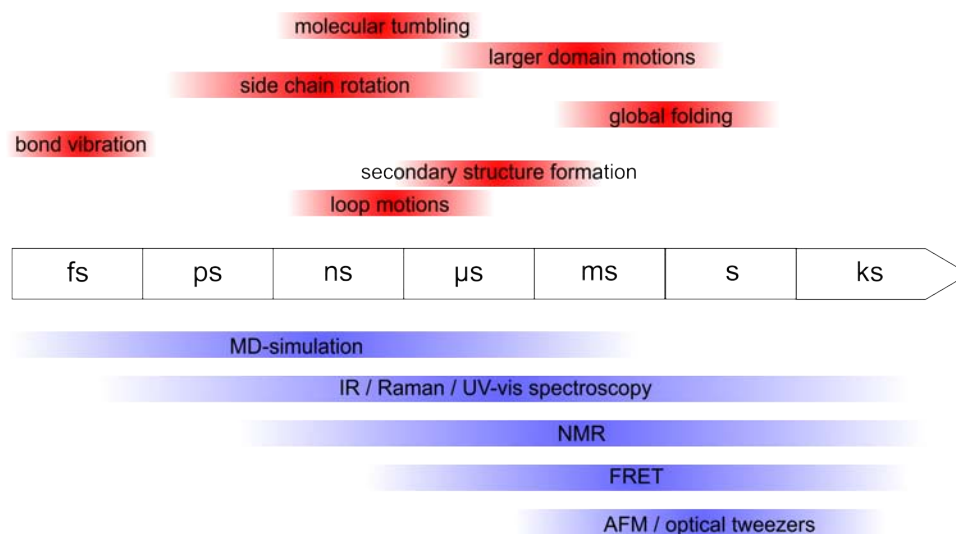
by  $k_{\text{B}}T$ ) are needed to drive the machine from one point on the surface to another. Thus, thermal energy itself is sufficient to make the molecules move. Following this idea, the molecular machine will perform a random walk over its energy surface becoming ultimately a Brownian machine.

As pointed out, a rich physical framework is available to understand the dynamics and the functions of molecular machines. Still there are a number of open questions such as: What is the role of chemical energy in the storm of thermal noise? Are there design principles of molecular machines forcing them towards a certain dynamic behaviour encoded in the structure? Can dynamics be predicted? Are some dynamics just meaningless random movements?

Surely this thesis seeks answers to these questions but so did many studies before. Of course, no finite answers can be given but at least more details contributing to the bigger picture can be revealed. New methods will be developed and established for an easier approach to better understand the dynamics of large macromolecular complexes.

### 1.2.3 Methods to analyse Conformational Dynamics

To get a full insight into the mode of action of a given molecular machine one would like to have the entire energy landscape of the machine at certain conditions. Ideally that means to gain structural as well as kinetic data at high resolution. One would like to follow experimentally (1) a single molecule (2) in full atomic detail (3) on at least microsecond time scale. This, however, is and most likely will be impossible. Nevertheless, the combination of several different methods can come close to that ideal case. In the following, I will shortly introduce the principal methods used to analyse conformational dynamics until now. Figure 1.6 points out the most relevant time scales for most of the methods presented in the following paragraphs.



**Figure 1.6: Time Scales of Protein Dynamics and their Processes.** A comprehensive overview of the time scales of individual intrinsic movements is shown in red. The time scales accessible to important methods are depicted in blue. The figure is based on [88, 142].

#### 1.2.3.1 MD Simulation

Actually, there is a method matching the above mentioned three criteria very well. With Molecular Dynamics (MD)-simulations one can really look at any movement of a given particle in full atomic detail on a femtosecond time scale [99]. The method's development

started in the seventies and earned the Nobel prize for chemistry in 2013<sup>2</sup>. A structural model of the molecule, mostly gained by experimental means is used. Movements of individual atoms are principally calculated by applying Newtonian mechanics to every atom of the molecule and the surrounding medium. This can give detailed insights into the dynamics of a molecule in an experimentally inaccessible timescale. However, reaching the experimentally accessible timescale of micro- to milliseconds is still computationally demanding to an extent, that it is hardly achieved by any study. An experimental proof for the correctness of the used models and insights is still missing to a large extent. Nevertheless, MD-simulations give valuable insights into protein dynamics that are otherwise hardly accessible experimentally. The flow of individual water molecules through an aquaporin molecule [73] or the path of tRNAs moving through the ribosome [172] in atomic detail was elucidated.

### 1.2.3.2 Spectroscopy

Classical dynamics of proteins were studied by spectroscopy. The interaction of light with molecules can provide much information about them. The number of possibilities is vast:

1. IR-spectroscopy can provide information on the vibrational movements of atoms.
2. CD-spectroscopy can give insights into the secondary structure of a protein.
3. UV-Vis spectroscopy and fluorescence will report about the changes and distances in light absorbing regions of a protein.
4. With light scattering the diffusion time and derived from that the size and movement speed of a particle can be analysed.

All of the mentioned techniques, however, have one problem in common. They need to combine signals from billions of molecules and thus average all the individual molecules no matter in which state they are. Even within a molecule it is hardly possible to localize a certain effect. Nevertheless, these methods provide valuable information and good first approximations on the dynamics of molecular machines. Many studies employed synchronisation of the molecules as solution for the mentioned problems. For example, molecules could either be triggered by temperature jumps or the addition of their binding partners or substrates. Still dynamics on very small timescales will average out, but the overall dynamics can be readily investigated in this way. Moreover, specific regions of a protein can be labelled to confer resolution in space. [88]

---

<sup>2</sup>The 2013 Nobel in chemistry was awarded to Martin Karplus, Michael Levitt and Arieh Warshel "for the development of multi scale models for complex chemical systems"

### 1.2.3.3 Single Molecule Approaches

Real insights into the dynamics of a molecule on small time scales can hardly be gained by the previously described batch methods. Molecule dynamics needed to be rather synchronized to see detailed dynamic properties. In the 1990s, this obstacle was overcome with two conceptionally very similar but technically very different approaches named *single molecule detection* and *single molecule manipulation*. With better light sources and detectors it is now possible to detect a single fluorophore bound to a single molecule in an ultra small volume. With the aid of techniques like Foerster Resonance Energy Transfer (FRET) the temporal change of a certain distance within a molecule can be measured[97]. Another branch of development even led to the possibility of moving and manipulating single molecules. This is possible in Atomic Force Microscope (AFM) as well as in optical or magnetic trap experiments. Optical or magnetic traps are experimental set-ups that can precisely move and track micrometer-sized spheres. By binding a few molecules onto the surface of those spheres their movement can be tracked, and force can be applied to them. This led to beautiful insights into the dynamics of molecular machines like polymerases or ribosomes [29, 53]. The AFM, on the other hand, is an instrument utilizing a tiny lever arm (termed cantilever) on which a fine tip is mounted. This tip, which is only a few atoms in diameter can be used to scan a surface to visualize the molecules on it (imaging mode) or can bind a molecule and displace it or apply a force on it (force mode). In force mode, the force applied between tip and surface through the molecule can be precisely measured in a microsecond time scale. In imaging mode, it became recently possible through high speed AFM to continuously scan a molecule on a millisecond time scale and follow its overall dynamics [101].

### 1.2.3.4 Structural Methods

While the previously named methods give only sparse information on the atomic coordinates of a molecule, structural methods rarely have a good time resolution. An overview of the technical limits of the three major structural methods is given in table 1.2.

**Table 1.2: Overview about the three main Methods of Structural Biology.**

	NMR	X-ray crystallography	cryo EM
accessible size range	mostly < 100 kDa	mostly < 200 kDa	mostly > 500 kDa
sample requirement	> 5 mg	at least several mg	a few $\mu\text{g}$
achievable resolution	not applicable	for molecular machines $\approx 3 \text{ \AA}$	2
time scales of dynamics	$\mu\text{s}$ to days	rarely possible	possible at most seconds to minutes

### 1.2.3.5 Nuclear Magnetic Resonance

Nuclear Magnetic Resonance (NMR) gains structural information through the magnetic properties of a molecule's nuclei (nuclei with an odd number of nuclear particles). [102] The sample is brought into a strong magnetic field (normally several Tesla), which aligns the spin of certain nuclei with the field direction. Through a radio wave pulse, the magnetic moment of the nuclei is rotated. After the pulse, the nuclei relax back to the aligned state and will thus emit the previously absorbed energy as radio wave. How much energy is absorbed, strongly depends on the chemical environment of the nuclei. With sophisticated series of radio wave pulses, distance measurements between individual atoms are possible and also dynamics can thus be measured. This makes NMR a valuable tool since it can show a protein's dynamics through time scales of nanoseconds up to days in Å-resolution [93]. Even though NMR brought great insights into the dynamics of many small proteins like ubiquitin [113] or DHFR [18], the investigation of larger assemblies can hardly be performed. Larger molecules have more nuclei, and one needs to resolve more spectral peaks from each other to gain good insights into the structure. Moreover, larger molecules have larger rotational correlation times, broadening the peaks even more. However, smart labelling made NMR studies of larger assemblies such as ribosomes [34, 92] at least partly possible. Still, the analysis of full assemblies will remain technically challenging if not impossible for many years.

### 1.2.3.6 X-ray Crystallography

X-ray Crystallography is by far the oldest of the structural investigation methods<sup>3</sup>. For this methods, proteins have to form crystals which is an intricate, barely understood process requiring a high number of empirical trials to find the right conditions. Those crystals diffract X-rays significantly, which leads to characteristic diffraction patterns. Thus, structural information has to be gained from the diffraction pattern. Only the amplitudes of interfering X-ray waves and not the corresponding phases can be recorded from such a diffraction experiment. The phases have to be determined separately via different approaches. Until today, crystallography yields the highest resolution for proteins but has its limits. For large macromolecular machines as they are analysed in this thesis only a few examples have yielded reliable structures - mostly ribosomes, polymerases and viruses were solved reaching resolutions up to 3 Å [133]. Moreover, dynamics can hardly be acquired since experimenters rely mostly on luck to find the same protein crystallizing in different conformations. In very rare cases the molecules are still functional in their crystalline state, which gives the possibility for time resolved crystallography [77]

---

<sup>3</sup>Of note: This year(2014)is named the year of crystallography by the UNESCO.



### 1.2.3.7 Single Particle Electron Cryo-Microscopy

Structural investigation of biomolecules using electron microscopes is considered the youngest among all structural techniques. [145] In the most commonly used single particle approach projection images of individual molecules or complexes are recorded in a transmission electron microscope (TEM). These can be combined computationally to yield a 3D structural model. The technique will be explained in great detail in the next section (1.3). In contrast to the other structural investigation methods, cryo EM is a true single molecule method. The signals from individual particles can be easily distinguished and thus different conformational states can in principal be separated into different structural models. However, molecules need to be frozen or embedded to withstand the vacuum of an operating TEM. This almost fully diminishes any possibility for time resolution that has to be mostly gained from other methods. Still the possibility to investigate the complete structural landscape of a large protein complex by directly counting the molecules in a certain state is so attractive that this is the main method used in this thesis.

### 1.2.3.8 Method Combinations

As must be now evident, none of the described methods alone is fully capable of determining the dynamics of a protein complex in the desired detail (see 1.2.3). Nowadays many studies combine the mentioned methods to get greater insights into the full dynamics of molecular machines. Among recent successes are the description of the tRNA movement through the ribosome with the aid of kinetic data, structural information from cryo EM and X-ray crystallography combined with MD-simulations [17] and a structural model of the HIV capsid through a combination of cryo EM, NMR and MD-simulation [229].

## 1.3 Single particle cryo EM as tool to analyze conformational dynamics

*It is very easy to answer many of these fundamental biological questions; you just look at the thing!*

— Richard Feynmann

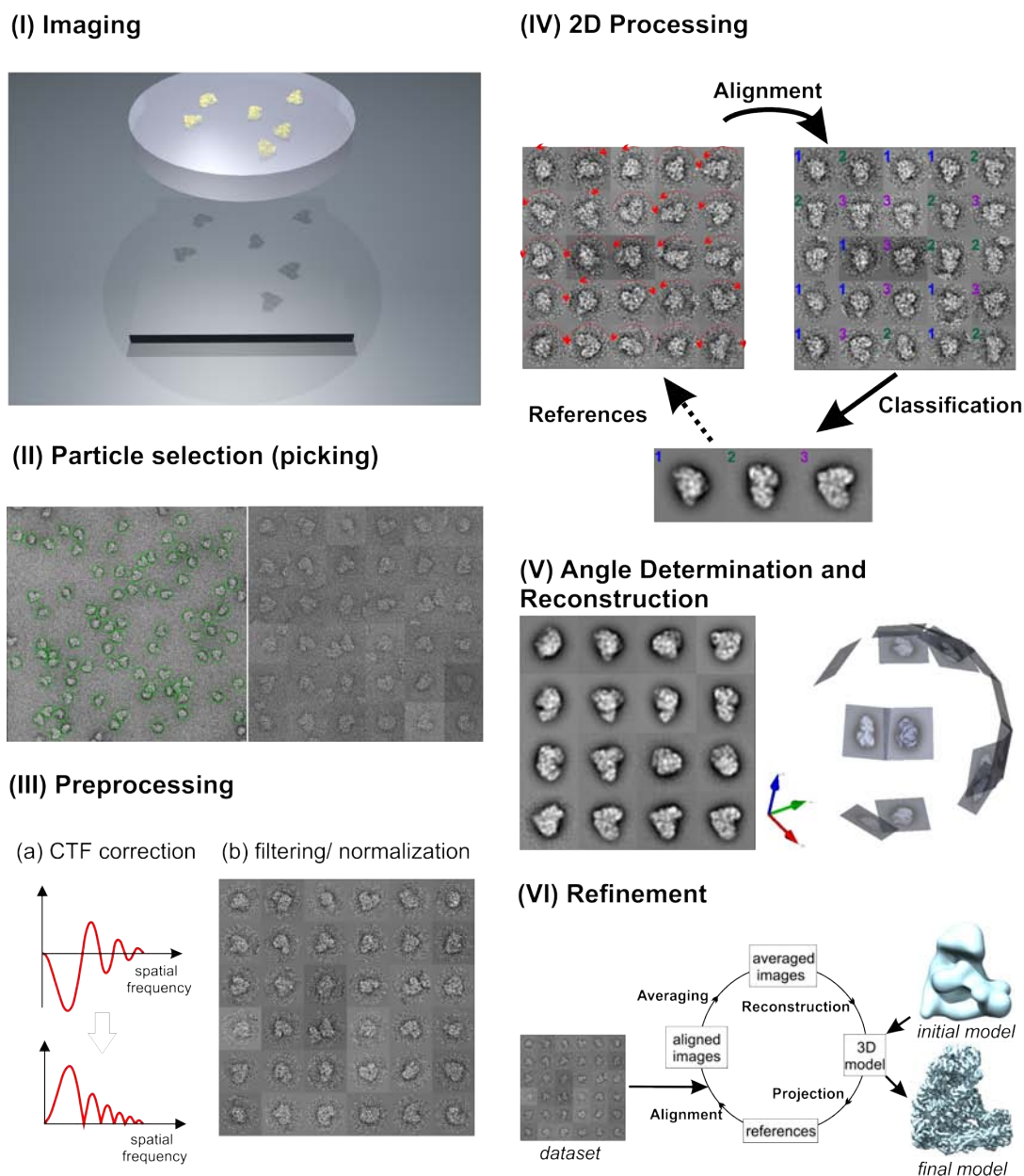
### 1.3.1 General workflow

Single Particle Electron Cryo Microscopy (cryo EM) is the main method used in this thesis to analyze the structure and finally the dynamics of molecular machines. The main advantage of this method is the high degree of directness. This means that raw data can readily be interpreted by eye and the room for the interpretation is rather narrow. In principle, it can even be seen as single molecule method since signals from individual molecules can be clearly distinguished and thus be sorted very accurately.

The general workflow and the possibility to use it for conformational dynamics is outlined in this section and will be explained in detail in the materials and methods section (see 2.4). From a very abstract point of view many similarities to biochemical purifications can be found (see figure 1.7). Very exhaustive and excellent reviews can be found in the textbooks of Joachim Frank and Michael F. Moody [55, 131]

1. As in any biophysical method the workflow starts with the preparation of the purified sample for the method. Electron microscopes operate under high vacuum for which the sample has to be stabilized. In biomolecular EM the two frequently used methods are freezing the molecules in a thin film of buffer (cryo conditions) or embed and dry the sample with surrounding heavy metal stain (negative staining), both on the carbon surface of small metal grids as support.
2. After embedding of the sample, it is introduced into a TEM and images of several regions of the grid are taken. Each image depicts many copies of the molecule in random orientation and distribution. The images are close to perfect parallel projections of the molecule, meaning the 3D-information is integrated into a 2D image. These 2D images, however, have a very low Signal-to-Noise-Ratio (SNR) and due to the high energy of the electron beam, the exposure time of the sample to electrons is kept to a bare minimum. The orientation of a molecule on the carrier grid can be described by six degrees of freedom: the three translations along the coordinate axes and the three rotations around them. To determine these parameters, the images will be subjected to exhaustive computational image processing.

3. In a first step, the particles have to be identified (picked) within the image and an image stack containing the individual particles is created.
4. Due to the low interactions of the electron beam with biological specimen, the images have to be taken with underfocus, leading to a broad and complicated point spread function (PSF) smearing the image information over a large area. This can be restored *in-silico* in a process called CTF-correction.
5. The corrected particle images need to be further prepared for analysis. To increase the SNR for initial analysis, they are down sampled (coarsed) and Fourier filtered. Further, the images have to be normalized, since they arise from different micrographs and different parts of the grid.
6. To further improve the SNR, individual particle images are averaged. Since the particles were randomly orientated on the grid different particle images depict different views of the particles and have to be sorted for their orientation and properly superimposed. The processes necessary are called *alignment* and *classification*. Alignment orients the particles images in a way that they superimpose a set of given references. Classification identifies similar images and sorts them into a given number of classes. Thus, the particle images showing the same orientation, can be averaged to so-called class averages.
7. Once average images of similar orientation are calculated, a 3D-reconstruction can be attempted. Beforehand, the orientation of the averages in space with respect to each other (as Euler angles) has to be determined. This can be done mathematically *in-silico* in a process called angular reconstitution or experimentally with a technique called random conical tilt (RCT). After the angles are determined, a 3D model can be reconstructed.
8. Initial structures are low in resolution. To improve the structure, several refinement cycles are performed. In every cycle, the accuracy of alignment and orientation angle assignment is improved.
9. Once convergence of the refinement is reached, the model needs to be validated, if this was not done before. Furthermore, the resolution of the structure should be determined. The resolution describes which degree of detail can be seen in the structure meaning how far structural features need to be apart to be distinguishable.
10. Finally, interpretation of the model can be attempted. To interpret the computed 3D structures, they can be segmented and structural models gained from other methods can be fitted into the density model. Alternatively, conformational sorting can be performed as outlined in the following section.



**Figure 1.7: Single Particle cryo EM Workflow.** The general steps in single particle cryo EM are illustrated using the Anaphase Promoting Complex (APC) as an example. (I) In the microscope projection images are taken of the ice embedded molecules. Particles appear as dark densities in light background which is inverted for image processing. (II) In a first image processing step, molecules have to be selected in a process called particle picking. Here, the molecules are encircled in green. The identified particle images are cropped from the micrograph (right). (IIIa) Every particle is distorted with a point spread function. The Fourier transformation of this is called Contrast Transfer Function (CTF) which is depicted here. It can be seen that in certain areas of this function the contrast is negative, meaning inverted. This is fixed during CTF correction by phase flipping (IIIb) The corrected images are filtered to optimally prepare them for the alignment. (IV) In an initial 2D processing, images that show the same orientation of the molecule in space are superimposed in a process called alignment and grouped and averaged in a process called classification. (V) In the next step, for all good class averages (left) their relative orientation in space is determined (right). After orientation they are projected back into 3D space to calculate a first 3D model. (VI) The first 3D model is gradually improved in iterative refinement steps. Thereby, the 3D model is projected into 2D images which are used as references for an alignment against the full dataset. After alignment, the images are averaged with respect to the best matching reference and a 3D model is reconstructed from the averages. The whole procedure is reiterated with decreasing projection distance until it converges. Shown is a 7.4 Å model of the APC/C (EMDB-2651).

### 1.3.2 Conformational Analysis

One of the main advantages of cryo EM is that the signal obtained from individual particles can be undoubtedly distinguished even by the human eye. Although new developments in X-ray crystallography like the Free-Electron Laser (FEL) point into the same direction [186], cryo EM will keep being the only technique capable of that.

Unlike most single molecule methods, the complete atomic structure is encoded into the micrograph. Due to the bad SNR and the integration of information along the z-axis through projection the particles still have to be averaged to get a 3D-model. However, the number of necessary particles to be averaged to get atomic detail is low (only several 10000) and "smart" averages, meaning only particles showing the same conformation can be averaged. In principle, all conformations of a molecule within a data set could be potentially distinguished and yield high resolution structures. Thus, sophisticated computational sorting of individual particle images in their respective state is necessary. This kind of analysis is relatively new to the field and as such not many robust methods are available. Nevertheless, several published strategies already gained outstanding results and are thus outlined in this section.

First of all, it has to be mentioned that conformational dynamics introduce new degrees of freedom into the analysis. While for a particle existing in a single conformation five degrees of freedom have to be determined, the conformation is an extra parameter. On top of that, not every heterogeneity present in the data set represents a relevant conformational state. Also, broken particles or contaminations are responsible for an unknown degree of structural heterogeneity. They should be excluded from the analysis as much as possible before any conformational sorting is performed.

At which stage conformational analysis itself has to come into play depends very much on the kind of conformational dynamics present in the sample. As outlined in section 1.2.1, conformational changes of proteins take place at different time scales correlated with different amplitudes of movements. We can, in principle, distinguish between four different conformation types which have to be handled by different means.

1. Conformational changes strongly altering the overall shape, have to be handled right at the start of the 3D analysis. Combining all particles in a single structure will result in a featureless blob that does not even resemble the molecule at all.
2. The movement of large domains of a structure not altering the shape of the molecule can be handled later during 3D analysis. The average of all particles will result in high resolution for the static part but will smear out the flexible parts.

3. Movements of small domains or binding of ligands, representing only a small fraction of the mass of the structure, can be handled at a later stage of the analysis. The average of all such particles will again result in the vanishing of the flexible part or a smearing effect.
4. Conformational changes occurring on short time scales, like bond vibrations or low amplitude movements of structural features are currently hard to handle. However, averaging particles that have only small conformational differences will only slightly reduce the gained resolution. This can be accepted for most biological questions.

The different techniques are outlined with examples in the following sections.

### 1.3.2.1 Biochemical and biophysical strategies

Prerequisite for any structural analysis of a biomolecule is its integrity and its compositional homogeneity. This has to be assured during the whole purification procedure. To facilitate that especially for electron microscopic analysis, the *GraFix* methodology was introduced recently [100]. Hereby, the molecule of interest is loaded on top of a density gradient, which contains additionally a fixative gradient. Through centrifugational sedimentation, the protein complex is separated from lighter broken complexes and heavier aggregates, and covalent crosslinks assure its structural integrity for the harsh electron microscopic preparations. It has to be stated that the crosslink will not alter the molecules structure, however, it might be possible that the occupation of certain conformations changes.

The most used techniques to tackle conformational dynamics are not of computational but of biochemical nature. With the aid of a chemical compound or a protein, the molecule of interest can be trapped in a certain conformation. The advantage of this technique is that the significance of the resulting state can be clearly linked with its function by knowing the function of the compound or its effect. However, this is not a generic strategy. There are certainly no compounds available to block any molecule in any given state. A direct screening for compounds with libraries of thousands of chemicals using a complete structural analysis is practically impossible. Thus, reasonable choices have to be taken from the available biochemical data. The number of accessible silenced conformations is rather small. But, even if such a state is found, hardly any compound is imaginable silencing the dynamics of a protein complex completely. However, this approach is the most used one in publications to date and gave great insights into snapshots of certain conformations. Thus, a few examples are addressed in the following paragraph.

As for any single particle EM technique, most work was certainly done on the ribosome. A vast number of different antibiotics was used to stall the ribosome in distinct states

[221]. In this way, almost the whole set of states in the translational cycle could be locked and investigated structurally [132]. Other successful examples come from ATPases: the importance of ATP in our cells as energy source stimulated research and led to a large number of different ATP-analogues that resemble different hydrolysis states. Commonly used compounds are non-hydrolyzable analogues like ATP $\gamma$ S, AMPPCP or AMPPNP, intermediate state analogues like ADP – AlF<sub>4</sub> or ADP – BeF<sub>3</sub> or simply the hydrolyzed product ADP. In this way, a molecule can be trapped in any stage of the ATP hydrolysis. Similar results could be gained by simple point mutations like Walker-A and -B mutations. This whole strategy was very successful in studies of motor proteins [190], chaperones [163] or AAA+-proteins [170, 32, 192] and with GTP analogues translation factors [132].

As stated before, these lucky results are exceptions. Mostly, one has to become more inventive to trap a molecule in a functional, relevant state. For example, the 30S proteasome was trapped in a translocation state by using a slowly translocating GFP, resulting in a conformation that was not seen before [124].

### 1.3.2.2 Computational Strategies

Ideally, one would like to calculate the full ensemble of possible structures from one dataset without any biophysical trick to restrict the number of conformations. Only a few studies exist so far which attempted this. For the ribosome, this was demonstrated in 2010 [52]. The authors used a hierarchical computational sorting scheme to analyze a dataset containing two million particles. Thereby, it was first sorted for the rotation of the subunits against each other - the so called ratcheting. In a second step, it was classified for different rotations of the 30S head against the body of the 30S subunit. Lastly, the different obtained conformations were analyzed for the positions of their tRNAs. In total more than 50 structures were obtained and sorted computationally. Even though this study was definitely a breakthrough, no similar study was conducted since for other molecules. The reasons for that are missing generic procedure. The hierarchical scheme used for the ribosome could be conducted since the knowledge of the ribosome is vast and the major movements were described before and thus could be accurately modelled. However, many recent studies at least identify a few conformations in parallel.

Nevertheless, many methods are published, which can aid the conformational sorting. If the conformational differences are large, meaning the overall shape of the molecule is mostly different, the conformational sorting should occur right at the start of the 3D structural analysis. A method called *Random Conical Tilt (RCT)* is used mostly for that [162]. Hereby, the missing angular parameters (Euler angles) for the recorded particles are determined experimentally by tilting the specimen holder in the TEM. With this, initial structures can be calculated from a few images. However, these models are slightly

distorted and noisy. This can be improved by 3D alignment and classification of the models and subsequent averaging [175]. Recently, images from several conformationally as well as compositionally heterogeneous samples were successfully sorted into distinct structures that could be refined. For example, the dynamics of the very flexible E3 ligase Itn1 was elucidated in several conformations [122] and all intermediates of the assembly pathway of the 30S ribosomal subunit were resolved [134].

Apart from that, a few other approaches exist to obtain different initial models for a conformationally heterogeneous dataset. As used in the previously stated, ribosome analysis modelling of initial structures based on biochemical knowledge was successfully employed. If at least the structure of one conformation is known simple simulations like Normal Mode Analysis (NMA) can be used [96]. This predicts the thermal modes of motion based on mechanical considerations of the molecule.

If the overall shape is not largely altered within the conformational landscape, meaning if only a small portion of the total mass is moving or the movement amplitude is small, conformational analysis can be postponed to a point where the refinement fails to improve the resolution of the structure. Many methods were described and successfully employed [116]. Most of them are based on a stochastic approach called *bootstrapping*. Thereby many subsets of the dataset are chosen randomly with replacement and for every subset a 3D model is calculated. The set of different models is analyzed statistically [156]. Thereby, new models from different conformations can arise, which need to be validated and refined. A more detailed overview over the available methods is given in the methods section (see section 2.4.15).

## 1.4 Biological Model Systems

In this thesis, the conformational dynamics of three protein complexes are analyzed.

Firstly, the export adapter protein Chromosome Region Maintenance factor 1 (CRM1) from the fungus *Chaetomium thermophilum* and its human homologue will be analyzed. For a cryo EM analysis, this is a major challenge since it is with 120 kDa fairly below the standard size (>500 kDa) of proteins subjected to EM analysis. Here, I suspect the dynamics to have a dramatic impact on the achievable resolution.

Secondly, the very dynamic *E.coli* Pyruvate Dehydrogenase Complex (PDHc) will be investigated. The complex is known to have very loosely attached subunits and thus is one of the most dynamic model systems while it is with more than 5 MDa an ideal object for single particle cryo EM, size wise.



Lastly, the Acetyl Choline Binding Protein (AChBP) from *Biomphalaria glabrata* will be analyzed. It is an icosahedral protein and refines thus very quickly to subnanometer resolution. Relevant dynamics for this complex will be found in small amplitudes and time scales, which were not accessed by cryo EM so far.

### 1.4.1 Chromosome Maintenance Factor 1

#### 1.4.1.1 Nucleocytoplasmatic transport

A medium sized industrial is most commonly split into different departments. One department is responsible for the production, one does the logistics, and one is managing. It is an universal principle that those departments are spatially divided, for example, to protect some workers from noise or danger from other workplaces.

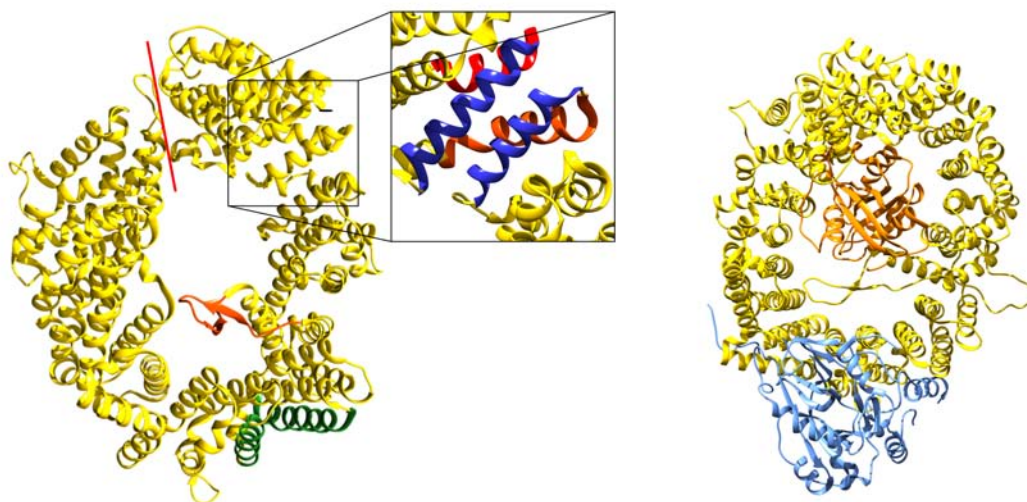
Eukaryotic cells separate their workspaces in a similar manner. The so-called compartmentalization is one of the main traits allowing higher life forms to exist. The reasoning for this is twofold. Firstly, some cellular processes need a special chemical environment to work efficiently. For example, degrading of biomolecules as occurring in lysosomes works best in an acidic environment, while this would be harmful to most other processes. Secondly, cells build complex machines made of many different molecules, which have to be assembled. It is crucial that an assembly of such machines is complete before the machine is performing any task. Therefore, cells evolved in a way that the places of assembly and function are well separated from each other. Ribosomes, for example, synthesize proteins in the cytoplasm but are assembled in the nucleus, while spliceosomal snRNPs act in the nucleus but are assembled in the cytoplasm. Moreover, most factors are required only at a given moment and should be mostly hidden from their site of action, like transcription factors.

However, these big advantages have to be met with sophisticated logistic systems transporting proteins from one compartment to another. Those systems need to be directed and tightly controlled. All membrane-enclosed compartments have their own set of transport mechanism which can be rather distinct. The most remarkable transport is found between the nucleus and the cytoplasm. Unlike other cases, proteins and even protein complexes get transported in a folded and assembled state. While many pathways, purposes and transporter molecules for nucleo-cytoplasmic-transport form a vast complexity, the main principles are simple and shared among them. First of all, the single accessible physical pathway leads through only one kind of channel: the nuclear pore complex (NPC). Moreover, there is only a single universal energy source: a gradient of the GTP loaded small GTPase called Ran. [67]

### 1.4.1.2 Structural Characteristics of Exportins

As stated before, all pathways in and out of the nucleus lead through the interior of the NPC and require Ran-GTP. Since the major principle seems to be shared between all transport complexes, it is not surprising that the different transport receptors are structurally very similar. All the known transport factors, transporting proteins belong to the  $\beta$ -karyopherin family (also importin  $\beta$ -like) and share a common architecture comprising mostly of the same structural motif – the Huntingtin, elongation factor 3, protein phosphatase 2A and TOR1 (HEAT)-repeat. The canonical HEAT-repeat is  $\approx 40$  amino acids long and consists of two antiparallel helices (termed A and B) connected by a short linker sequence [5]. However, there is no consensus sequence for this motif and also overall homology between  $\beta$ -karyopherin is only 15 % [39]. Two consecutive HEAT-repeats bind each other in a parallel fashion, where the A-helix forms the outer convex and the B-helix the inner concave surface of a superhelical arrangement (see figure 1.8). This architecture leads to an intrinsic flexibility which is thought to provide enough versatility for the receptors to bind a multitude of different cargoes [38]. These conformational dynamics make  $\beta$ -karyopherin family an ideal model system for this thesis. Exemplified the conformational cycle of the exportin Chromosome Region Maintenance factor 1 (CRM1) should be analyzed. Unlike many of the other nuclear transport receptors, the set of cargoes it can transport is large and versatile [203]. Among them are for example the large and the small ribosomal subunit, translation factors, the signal recognition particles, Cyclin-D1, and several mRNA as well as snRNA substrates. Cargo recognition requires a leucin-rich nuclear export signal (NES) and mostly several adaptor proteins.

Structural insights into CRM1's detailed architecture were made only recently [128]. Most contributions towards the structural understanding of CRM1 were made during the time course of this work and are discussed thoroughly in later sections. However, the first complete X-ray structure of CRM1 with cargo and Ran-GTP already revealed the overall architecture [128]. CRM1 comprises of 21 HEAT-repeats oriented in a superhelical fashion. As stated before, the overall homology within the  $\beta$ -karyopherin-family is rather low. The first three HEAT-repeats - the so-called CRM1 - Importin  $\beta$  etc. (CRIME) domain, interacts with Ran. Another important element responsible for Ran binding is the *acidic* loop: a  $\beta$ -hairpin after HEAT 9 (see figure 1.8). The cargo is bound around HEAT 11-16, where a hydrophobic cleft is formed between A-helices 11 and 12 recognizing the cargoes nuclear export signal (NES)(see figure 1.8). In contrast to other nuclear transport receptors CRM1 binds its substrates on the convex outer surface, not encapsulating it. On the one hand, this seems inevitable given the large sizes of CRM1's cargoes. On the other hand, the cargo is in this way fully exposed to the interior of the NPC.



**Figure 1.8: Structural Features of CRM1.** Left: A structural overview over CRM1's architecture is shown. It consists of 21 HEAT-repeats. Two of these elements are magnified. They consist of an A (red) and a B helix (blue) which stand antiparallel to each other while consecutive HEAT-repeats stack in a parallel fashion. Further, the acidic loop (orange) which is important for binding of Ran and the hydrophobic cleft (green), where cargo can bind, are marked. N- and C- terminus close the ring structure at the red line. Right: A fully assembled nuclear export complex containing CRM1 (yellow), Ran (orange) and cargo (blue, here: snurportin).

While the question of how CRM1 recognizes its substrates is generally agreed [130], the full conformational cycle of CRM1 was unknown at the beginning of this thesis. In the nucleoplasm, apo-CRM1 has to bind cargo and Ran-GTP. The ternary complex diffuses through the NPC after which it needs to disassemble upon GTP hydrolysis in the cytoplasm. Subsequently, apo-CRM1 diffuses back to the nucleus.

The structural changes within CRM1 and its complexes render it an attractive model system. Additionally, it is a challenging object since its size of 120 kDa would rather disqualify it for EM-analysis.

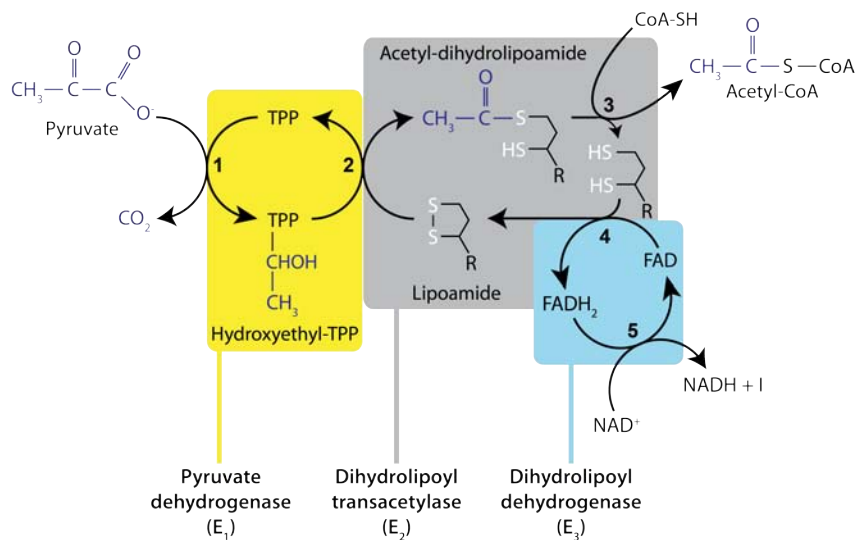
## 1.4.2 Pyruvate Dehydrogenase Complex

The Pyruvate Dehydrogenase Complex (PDHc) is one of the largest complexes in the cell with a diameter of up to 50 nm and a mass of up to 10 MDa.

### 1.4.2.1 Biochemistry of the PDHc

PDHc is found in respiratory eubacteria and the mitochondrial matrix of eukaryotes. It catalyses the irreversible reaction from pyruvate to acetyl-CoA and is, therefore, one of the central junctions within metabolism. Acetyl-CoA serves as a rich source of metabolic energy as well as it is a widely used precursor in lipid metabolism, where pyruvate is used

to replenish intermediates of the tricarboxylic acid cycle as well as a precursor for amino acids and carbohydrates via gluconeogenesis [14, 149, 148].



**Figure 1.9: Catalytic cycle of the Pyruvate Dehydrogenase Complex.** Scheme was rebuilt after Voet, Voet, Pratt Biochemistry, 2004, [214]

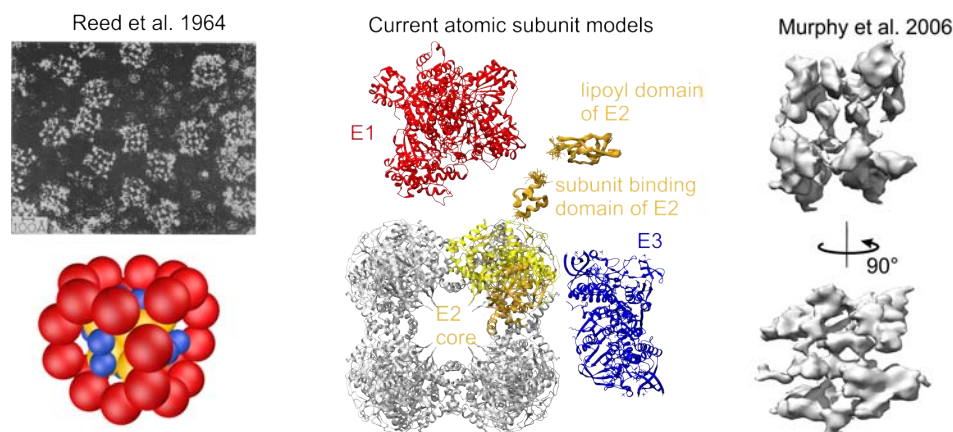
The complex is a multienzyme complex consisting of three different catalytic components, named E1, E2 and E3. In eukaryotes, the complex contains additionally the E3-binding protein (E3BP) and a number of regulating kinases. E1 is the thiamin diphosphate (TPP) dependent *pyruvate dehydrogenase* catalyzing the irreversible decarboxylation of pyruvate and the formation of hydroxyethyl-TPP. The resulting acetyl group is transferred to the N6-(lipoyl)lysine of E2 - the *dihydrolipoyl transacetylase*, which transfers it further to Coenzyme A. The E3 component, the *dihydrolipoyl dehydrogenase*, regenerates the oxidized lipoamide with a non-covalently attached FAD which is in return reduced itself by NADH. An overview of the reaction cycle is shown in figure 1.9.

#### 1.4.2.2 Structural Insights into the PDHc

Despite its apparent biochemical importance and the more than 60 years since its discovery [107], detailed structural knowledge about the whole PDHc could not be gained. That might be surprising taking into account that almost every biochemistry textbook depicts a structural model of the *E. coli* complex (e.g. [214]). This model was created in the 1960s, derived from electron micrographs. Lester Reed and colleagues recorded in several studies micrographs from the fully assembled complex as well as from the core structure [167, 220, 166, 165]. From visual inspection of the particles they concluded that the E2 component forms a cubic core structure, which is encapsulated by a cubic shell composed of E1 and E3 components. Shortly afterwards, the stoichiometry of the individual components was determined to be E1:E2:E3 = 2:2:1, where the E1 chains

form homodimers and the E2 chains homotrimers which are placed on the corners of the cubic core structure [168, 157]. This stoichiometry led to the "edges and faces" model, predicting the 12 E1 dimers to form the edges and the 6 E3 dimers the faces of a cubic shell around the cubic core. This architecture was supported by two further EM studies [16, 218]. The model for the overall architecture, however, got its first structural support only almost 30 years after its proposal by the crystal structure of a truncated E2 core from *Azotobacter vilandii* also showing an octahedral symmetry [123]. The quaternary structure of the shell is still unknown. Even though the full complex yielded crystals already in the 1970s [215], no structure was solved yet. The first published attempt of a calculated structural model of the complete complex was made in 2005 using Electron Cryotomography [136]. The authors propose that the cubic core is surrounded by flexibly tethered E1 and E3 components in 11 nm distance. They conclude the whole complex is inherently inhomogeneous and asymmetric.

The reason for the very flexible nature of the whole complex is supposed to be the mobile linker connecting the C- and the N-terminal domains of the E2. While the C-terminal acyltransferase domain catalyzes the reaction and forms the stably folded core the N-terminal contains a varying number of lipoyl domains only connected by flexible alanine and proline-rich linkers to each other and the C-terminus [197]. The shell components are shown to be only flexibly attached to the core in a distance of several nanometers to the core [217]. Overall the architecture of the *E.coli* PDHc still seems to be unclear and many questions remain unanswered. Dynamics seem to play a significant role, since at least the lipoyl domains have to move from the E1 to the E3 enzymes to fulfil a full catalytic cycle. A disordered arrangement, as suggested by the most recent publications [217, 136], would be unique so far but remains unproven.



**Figure 1.10: Structural Architecture of the *E. coli* PDHc.** Left panel: the first structural insights into the PDH came from micrographs like the one depicted [166]. A structural model was derived (below) from that. The E2 component forms a cubic core (yellow) while the E1 subunits sit on the edges (red) and the E3 (blue) on the faces of this cube. Middle panel: The available high resolution structures are depicted. The C-terminal catalytic domain of the E2 (yellow) forms trimers which are assembled to a cubic core structure (grey, PDB code: 1eaa). The N-terminal domains E3/E1 binding domain (orange, PDB code: 2pdd) and of the three lipoyl domains (orange, PDB code: 1qjo) which are only flexibly attached are depicted. E1 dimers (red, PDB code: 1l8a) and E3 dimers (blue, PDB code: 4jq9) arrange in an unknown manner on the surface of the E2 core. Right panel: The overall architecture of the complex as determined by cryo electron tomography, is shown [136]

Also, the architecture of the eukaryotic PDHc was studied since the 1960s. In contrast to the *E. coli* model, the overall symmetry is icosahedral [94], which is also true for the PDHc from gram negative bacteria. First hints towards this symmetry came from micrographs depicting the human E2 core structure in the 1960s [166]. Again, a core structure is built of E2 proteins while E1 and E3 form a shell structure. Additionally the E2 core is decorated with multiple copies of the E2-binding protein. Structural models of the full complex were solely gained from cryo EM experiments. The first structures of an icosahedral complex came from *Bacillus stearothermophilus*. Again, the core structure could be crystallized [94]. In following EM studies, reconstituted complexes containing E3 and E2 [127] or E1 and E2 [126] were used exclusively. In both cases, an icosahedral E2 core complex is surrounded by a shell of E1 respectively E3 subunits. Remarkably, a clear gap of 11 nm is visible between the core structure and the shell. The same phenomenon was found for the complexes purified from *Bos taurus* [231] and *Saccharomyces cerevisiae* [74]. It even was suggested that the core structure is mobile in the interior chamber formed by the shell [74]. While the mammalian core was solved by electron microscopy [223], the inner architecture of the shell was not revealed yet. Only one study suggests binding sites for the E3 on the intact core [213].

While in eukaryotes the architecture of the PDHc seems to be proven, its stoichiometry is controversial. Originally, it was suggested that 60 E2-subunits form the core in a do-

decahedral fashion, where 12 E3BP bind to the gaps of the structure [176]. Biophysical analysis, however, suggests a rather shared dodecahedral core out of 48 E3 with 12 E3BP subunits [91].

While the exact quaternary structure of both PDHc types lacks some important details, the atomic details of the individual components are almost fully available through crystal, EM and NMR structures (*E. coli*: [7, 69, 31], human: [35, 36, 223]). In both cases, however, structural detail is still missing about the complete architecture of the E2 component. Until now, no high-resolution structure of the complete E2 molecule is available. The reason for that is surely the flexible N-terminus, comprising of a varying number of independently folded lipoyl domains only flexibly linked to each other and the C-terminal acyltransferase domain. The E1 and the E3 component are assumed to bind to a flexibly attached binding domain as well.

### 1.4.3 *Biomphalaria glabrata* Achetyl Choline Binding protein

The first specimen that delivered atomic resolution through single particle cryo EM were icosahedral viruses [71]. Until now the highest resolution structure found in the EMDB is the Cytoplasmatic Polyhedrosis Virus (CPV) particle at 3.1 Å [225]. While recently reconstructions of the asymmetric ribosome reached the same resolution level [4], it is still considerably easier to achieve high resolution with icosahedral particles. The advantages of icosahedral structures are twofold. First of all, this symmetry leads to an enormous rigidity and stability of the structure, so that only harsh treatments lead to disruption of the particles. Moreover, the asymmetric subunit is rather small and thus only a 60th of the angular space has to be considered. Thus, considerably fewer particles are needed to gain high resolution. This speeds up all calculations. Both facts contribute to the high resolution that can be readily achieved from icosahedral particles. Therefore, these particles make it possible to analyze small amplitudes (respectively time scale) with little effort. The only challenge icosahedral viruses provide, is their large size. The earlier mentioned CPV is with 70 nm diameter almost three times as big as the prokaryotic ribosome. Therefore, large images are necessary to be able to calculate high-resolution structures from those particles. Recently, a small icosahedral non-viral particle has been found and analyzed by single particle cryo EM [178]. With 22 nm diameter the Acetyl Choline Binding Protein (AChBP) has only the size of a ribosome and is, thus, ideally suited for cryo EM. The protein complex is located in the hemolymph of the snail *Biomphalaria glabrata* and binds acetylcholine. While its function for the mollusk is debated [193, 13], the protein is highly homologous to the soluble ligand binding domain of the nicotinic acetylcholine receptor of higher animals. However, a remarkable number of structures [2] of this receptor were already solved in many different states, addressing the

most relevant biological questions. Thus, the AChBP serves here only as a model system for high-resolution dynamics, and no biological questions will be addressed.

## 1.5 Aim of the work

The aim of this work is to investigate the conformational dynamics of large proteins and their complexes. To achieve that, several methodological developments have to be performed. Initially, it is always important to make sure that the protein complex of interest is correctly folded and assembled. To assure this, soft purification methods will be applied. Moreover, a high throughput screen based on the Thermofluor method [140] will be developed. This is a fluorescence screen employed for a long time to measure the stability of a given single domain protein. Here, the methodology will be extended to protein complexes. Additionally, it is helpful to restrict the complex's conformational flexibility to the necessary minimum. Therefore, modifications of the established GraFix methodology [100] will be developed.

Once strategies for the purification and stabilization of complexes were found, three model systems will be analysed towards their conformational dynamics, utilizing cryo EM: Firstly, the nuclear export receptor CRM1 will be investigated. While the cargo bound structure was recently established, the conformation of the apo-CRM1 remained unclear. It is one aim to identify the states the protein samples and construct its conformational energy landscape. This will give answers to the remaining question how CRM1 achieves cooperative binding to its cargo and Ran-GTP [39]. Secondly, the pyruvate dehydrogenase complex from *Escherichia coli* serves as an example for a highly dynamic complex. As explained earlier, the overall architecture of the PDHc is heavily debated and recently thought to be rather unordered. This will be a challenging investigation since no *bona fide* methods exist for the analysis of complexes with extreme shape shifts. Lastly, small amplitude respectively small time scale dynamics shall be investigated. As a test object, the *Biomphalaria glabrata* Acetyl Choline Binding Protein will be used, which already previously refined to 6 Å. It is very likely that this restriction in resolution derives from local dynamics of the protein. Thus, sorting for these minor differences in the protein structure will be necessary to reach atomic resolution.

In summary, the three examples and the described biochemical methods should contribute to a generally deployable workflow to investigate the dynamics of any given large (>100 kDa) protein complex.



# Chapter 2

## Materials and Methods

### 2.1 Materials

#### 2.1.1 Software

Table 2.1: Software used in this thesis is listed.

Software	Source	Citation
Amira 4.1	TGS Europe, Merignac Cedx, France	[196]
coweyes	custom made at Stark lab, Göttingen	[28, 82]
Eman 2	<a href="http://blake.bcm.edu/emanwiki/EMAN2">http://blake.bcm.edu/emanwiki/EMAN2</a>	[202]
Imagic	Image Science, Berlin, Germany	[211]
imageJ	<a href="http://imagej.nih.gov/ij/">http://imagej.nih.gov/ij/</a>	[187]
johnHenry	custom made at Stark lab, Göttingen	[28]
Python 2.7	<a href="https://www.python.org/downloads/">https://www.python.org/downloads/</a>	
Relion 1.2	<a href="http://www2.mrc-lmb.cam.ac.uk/relion">http://www2.mrc-lmb.cam.ac.uk/relion</a>	[181]
Simple PRIME	simple.stanford.edu	[49]
UCSF Chimera	<a href="http://www.cgl.ucsf.edu/chimera/">http://www.cgl.ucsf.edu/chimera/</a>	[159]
Xmipp 3.1	<a href="http://xmipp.cnb.csic.es">http://xmipp.cnb.csic.es</a>	[40]

#### 2.1.2 Special Equipment

Table 2.2: Machines and special equipment used in this thesis are listed

Chemical	Manufacturer
Äkta Prime	GE Healthcare
Balances	Sartorius
Blender	Waring
Centrifugal Mill ZM 200	Retsch
Centrifuge Avanti	Beckman Coulter
Centrifuge WX Ultra 90	Sorvall
Copper EM grids	Plano
Gradient Master ip	Biocomp

Incubator Multitron Pro Shaker	Infors HT
Liquid Handler StarLET	Hammilton
Mortar RM200	Retsch
Orbital Shaker RS0S20	Pheonix Instrument
Photometer Lambda Bio X	Perkin Elmer
Plate Sealer ALPS 3000	Thermo Scientific
Pipetting Robot Microlab Star LET	Hamilton
Quantifoil grid R3.5/1 Cu 200 mesh	Quantifoil Micotools GmbH
rt-PCR machine CFX connect	bio-rad
Rotor TH660	Sorvall
Rotor SW40	Beckmann Coulter
Rotor SW32	Beckmann Coulter
TEM CM200 FEG	Philips
TEM Titan Krios	FEI Company
Vitrobot	FEI Company

### 2.1.3 Chemicals

**Table 2.3: Chemicals used in this thesis are listed**

Chemical	Supplier
Agar	Roth
Ampicillin	Roth
Aluminium oxide	Sigma Aldrich
Aspartate	Sigma Aldrich
Bis[sulfosuccinimidyl] suberate (BS3)	Thermo Scientific
Bis(2-hydroxyethyl)amino-tris(hydroxymethyl)methan (BisTris)	Sigma Aldrich
Bovine serum albumin (BSA)	Sigma Aldrich
Bradford Assay reagent	BioRad
Coenzyme A (CoA)	Roth
Dextrin	Sigma Aldrich
Dimethylsulphoxide (DMSO)	AppliChem
Dinatriumhydrogen phosphate ( $\text{Na}_2\text{HPO}_4$ )	Merck
Dipotassium phosphate ( $\text{K}_2\text{HPO}_4$ )	Merck
1,4-Dithiothreitol (DTT)	Roth
DNase I	Sigma Aldrich
Ethanol	Merck
Ethylenediaminetetraacetic acid (EDTA)	Merck
Glucose	Merck

---

Glutaraldehyde	Science Services
Glutathion (reduced)	Sigma Aldrich
Glycerol	Merck
GSH-Sepharose	Pierce
4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES)	Sigma Aldrich
Imidazole	Sigma Aldrich
Iron(III) sulfate hydrate	Merck
Isopropanol	Merck
Isopropyl $\beta$ -D-1-thiogalactopyranoside (IPTG)	Roth
Magnesium chloride	Merck
Magnesium sulfate	Merck
Nicotinamide adenine dinucleotide (NAD)	Roth
Peptone	BD
Polyethylene glycol (PEG) 6000	Sigma Aldrich
Potassium chloride	Merck
Potassiumdihydrogen phosphate ( $\text{KH}_2\text{PO}_4$ )	Merck
Protamin sulphate	Sigma Aldrich
Protease inhibitor tablet , EDTA free	Roche
RNase A	Sigma Aldrich
Sucrose	Merck
Sodium chloride	Merck
sodium pyruvate	Merck
SDS-PAGE running buffer	Roth
Sulfo-MBS	Thermo Scientific
Thiamine pyro phosphate (TPP)	Applichem
Tryptone	BD
Urea	Merck
Yeast extract	BD

---

## 2.2 Microbiological Methods

### 2.2.1 *Escherichia coli* Cultivation

For PDHc preparation BL21\*-*E.coli* (life technologies) cells were cultivated in PA-5052 medium [199] omitting lactose. No antibiotic was added to the medium. As pre-culture 200 mL medium were inoculated directly from a glycerol stock of BL21\* cells and grown over night at 37 °C. For the main culture, 500 mL medium in a 2 L baffled flasks were inoculated with this starter culture to reach an OD 600 of 0.1. The culture was shaken with 180 rpm at 37 °C until it reached an OD 600 of 4–5 (after 5–6 hours) and harvested 20 min centrifugation at 5000 g. The cell pellets were washed with PBS(137 mM NaCl 2.7 mM KCl, 10 mM Na<sub>2</sub>HPO<sub>4</sub>, 10 mM KH<sub>2</sub>PO<sub>4</sub> pH 7.4) to remove left over medium and subsequently frozen with liquid nitrogen. Cell pellets were stored at -80 °C.

### 2.2.2 *Chaetomium thermophilum* Cultivation

### 2.2.3 Spore Cultivation

Initially, *Chaetomium thermophilum* spores were provided by the group of Eduard Hurt (ZMBH, University of Heidelberg, Germany). New spores were cultivated on rice meal agar. 75 g wild rice (*Oryza*) was cooked for 2 h in 1 L water. This was filtered and the flowthrough was autoclaved after addition of 1.5 % (w/v) agar. 250 mL of this agar were filled into a 1 L flask and inoculated with 1 mL spore stock solution. The flask was sealed with parafilm and incubated at 55 °C for 4 weeks. Afterwards, the surface was scratched with a sterile inoculating loop, 50 mL of 1 M sorbitol were added and the flask was shaken for 20 min at 150 rpm. The resulting spore stock solution was aliquoted.

### 2.2.4 Mycelium Growth and Disruption

The mycelium growth was performed on CCM medium (3 g/L sucrose, 0.5 g/L NaCl, 0.65 g/L K<sub>2</sub>HPO<sub>4</sub>·3H<sub>2</sub>O, 0.5 g/L MgSO<sub>4</sub>·7H<sub>2</sub>O, 0.01 g/L Fe<sub>2</sub>(SO<sub>4</sub>)<sub>3</sub>·xH<sub>2</sub>O, 5 g/L trypton, 1 g/L peptone, 1 g/L yeast extract (BD), 15 g/L dextrin (dissolved in hot water)). Every culture contained 0.1 mg/mL ampicillin. First, a CCM agar plate was inoculated with spores. The plate was wrapped in parafilm to prevent it from drying out and incubated for 2 days at 50 °C. The mycel carpet from four plates was carefully removed from the agar and cut into small pieces. 500 mL CCM medium were inoculated with these pieces and shaken vigorously at 50 °C for one day. The culture was blended in a Waring blender

for 3 x 45 s. Each culture of 1 L was inoculated with 50 mL of the blended material. The cultures were shaken vigorously for 14 h. The mycel was harvested through a sieve, dried and washed via vacuum applied to a Buchner funnel. The dried mycel was crushed in a mortar cooled by liquid nitrogen. To open the cells, the crushed material was finely ground in a centrifugational mill under liquid nitrogen temperatures. The resulting powder was stored at -80 °C.

## 2.3 Biochemical Methods

### 2.3.1 General Biochemical Methods

Purification results were mostly analyzed by SDS-PAGE. Precast polyacrylamide gradient gels (Bio-Rad) with a percentage of 4–15% were mostly used. Gels were run in SDS Page Buffer (Roth) at a constant voltage of 200 V until the dye front reached the lower edge of the gel. Most commonly gels were stained with Comassie solution.

Protein concentrations were determined via the Bradford Protein assay [24]. Always, a dilution series of bovine serum albumin was prepared and measured as references for the unknown concentration.

### 2.3.2 ProteoPlex

The ProteoPlex screen was performed in 20  $\mu$ l reactions in a 96-well plate (white, low-profile, multiplate 96, Bio-Rad). Each individual 20  $\mu$ l reaction was composed of 16  $\mu$ l of the macromolecular complex under study at a concentration of 0.1 – 1  $\mu$ L, 2  $\mu$ L of SYPRO orange (Life Technologies) at final concentration of 10x and 2  $\mu$ L of the buffer/chemical under investigation. Buffer substance screens at various pH values were performed with a custom made buffer screen, where individual stock solutions were 1 M (see Table 2.4). Small molecule ligands were also screened by a custom made additive screen (see Table 2.5). Each plate contained control wells which contained the purification buffer of the respective protein and wells without protein. Unfolding transitions were recorded in a *CFX connect* real-time PCR machine (Bio-Rad). A reader program was utilized, where the entire plate was equilibrated to 30 °C for 2 min followed by fluorescence measurement. Subsequently, the samples were incrementally heated in 1 °C steps, equilibration for 30 s and fluorescence measured. The temperature ramping, equilibration and fluorescence measurement procedures were repeated 65 times and finally the sample was held at 95 °C for 2 minutes followed by a final fluorescence measurement.

**Table 2.4: Chemicals used in the ProteoPlex buffer screen.** 1 M SPG Buffer contains: 0.125 M Succinic Acid, 0.5 M  $\text{NaH}_2\text{PO}_4$ , 0.375 M Glycine adjusted to the appropriate pH with 10 M NaOH. 1 M MMT Buffer contains: 0.2 M DL-Malic Acid, 0.4 M MES Monohydrate, 0.4 M Tris adjusted to the appropriate pH with either 10 M HCl or 10 M NaOH. 3) 1 M PCB Buffer contains: 0.4 M Sodium Propionate, 0.2 M Sodium Cacodylate trihydrate, 0.4 M Bis-Tris Propane adjusted to the appropriate pH with 10 M HCl.) MES = (2-(N-morpholino)ethanesulfonic acid) monohydrate, Bis-Tris: Bis(2-hydroxyethyl)amino-tris(hydroxymethyl)methane. ADA = N-(2-Acetamido)iminodiacetic acid, HEPES = 4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid. Tris = 2-Amino-2-hydroxymethyl-propane-1,3-diol

Substance	pH	Well ID
Succinic acid, $\text{NaH}_2\text{PO}_4$ , Glycine	5.8, 6.2, 6.6, 7.0, 7.4, 7.8, 8.2	A1, B1, C1, D1, E1, F1, G1, H1
Malic acid, MES, Tris	5.8, 6.2, 6.6, 7.0, 7.4, 7.8, 8.2	A2, B2, C2, D2, E2, F2, G2, H2
Propionate, Cacodylate, Bis- Tris-Propane	5.8, 6.2, 6.6, 7.0, 7.4, 7.8, 8.2	A3, B3, C3, D3, E3, F3, G3, H3
Imidazole	6.2, 6.45, 6.7, 6.95, 7.2, 7.45, 7.7	A4, B4, C4, D4, E4, F4, G4, H4
HEPES	6.9, 7.1, 7.3, 7.5, 7.7, 7.9, 8.1	A5, B5, C5, D5, E5, F5, G5, H5
Tris	7.5, 7.75, 8.0, 8.25, 8.5, 8.75, 9.0	A5, B5, C5, D5, E5, F5, G5, H5

**Table 2.5: Chemicals used in the ProteoPlex additive screen.** Trace element mix = 50  $\mu$  M FeCl<sub>3</sub>, 20  $\mu$ M CaCl<sub>2</sub>, 10  $\mu$ M MnCl<sub>2</sub>, 10  $\mu$ M ZnSO<sub>4</sub>, 2  $\mu$ M CoCl<sub>2</sub>, 2  $\mu$ M CuCl<sub>2</sub>, 2  $\mu$ M NiCl<sub>2</sub>, 2  $\mu$ M Na<sub>2</sub>MoO<sub>4</sub>, 2  $\mu$ M Na<sub>2</sub>SeO<sub>3</sub>, 2  $\mu$ M H<sub>3</sub>BO<sub>3</sub>. peptide mix = NH<sub>2</sub> – RANDOM6AA – CONH<sub>2</sub> (hydrophobic aa: F, A, L, M, I, W, P and V are excluded)

Id	A	B	C	D	E	F	G	H
1	1.5 M NaCl, pH 7.0	1.5 M KCl, pH 7.0	0.25 M MgCl <sub>2</sub> , pH 7.0	0.25M CaCl <sub>2</sub> , pH 7.0	1.5 M LiCl, pH 7.0	1.5 M NH <sub>4</sub> Cl, pH 7.0	0.25 M ZnCl <sub>2</sub>	0.25 M MnCl <sub>2</sub> , pH 7.0
2	0.5 M K <sub>2</sub> SO <sub>4</sub> , pH 7.0	1 M KF, pH 7.0	1 M KBr, pH 7.0	1 M KI, pH 7.0	1 M KOAc, pH 7.0	1 M K <sub>2</sub> S	1 M KNO <sub>3</sub> , pH 7.0	1 M KHPO <sub>4</sub> , pH 7.0
3	50% v/v Glycerol, pH 7.0	50% w/v Sucrose, pH 7.0	50% w/v Glucose, pH 7.0	50% w/v Maltose, pH 7.0	50% w/v Trehalose, pH 7.0	50% w/v Galactose, pH 7.0	10% v/v PEG 400, pH 7.0	10% v/v Ethylene Glycol, pH 7.0
4	50 mM DTT, pH 7.0	50 mM TCEP, pH 7.0	50 mM reduced Glutathion, pH 7.0	50 mM Oxidized Glutathion, pH 7.0	50 mM EDTA, pH 7.0	50 mM Ascorbic Acid, pH 7.0	50 mM $\beta$ -mercaptoethanol, pH 7.0	50 mM MESNA, pH 7.0
5	50 mM Spermine, pH 7.0	50 mM Spermidine, pH 7.0	50 mM Putrescine, pH 7.0	50 mM Taurine, pH 7.0	50 mM Betaine, pH 7.0	0.5 M L-Arginine, pH 7.0	0.5 M L-Lysine, pH 7.0	0.15 M L-Asparagine, pH 7.0
6	0.5 M L-Glutamic Acid, pH 7.0	0.5 M L-Aspartic Acid, pH 7.0	0.5 M L-Cysteine, pH 7.0	0.0015 M L-Tyrosine, pH 7.0	0.5 M L-Proline, pH 7.0	0.5 M Glycine, pH 7.0	0.25M L-Histidine, pH 7.0	0.5 M L-Serine, pH 7.0
7	10 mM ATP, pH 7.0	10 mM ADP, pH 7.0	10 mM AMP-PCP, pH 7.0	10 mM CMP, GMP, TMP, UMP pH 7.0	10 mM NAD <sup>+</sup> , pH 7.0	10 mM NADH, pH 7.0	10 mM Biotin, pH 7.0	10 mM Thiamin pyrophosphate, pH 7.0
8	50 mM Na Malonate, pH 7.0	50 mM Na pyruvate, pH 7.0	50 mM L-Citrulline, pH 7.0	50 mM Na Citrate, pH 7.0	50 mM Na Glutarate, pH 7.0	50 mM Na Succinate, pH 7.0	1 mM Coenzyme A, pH 7.0	Trace-element mix, pH 7.0
9	20% v/v Methanol, pH 7.0	20% v/v Ethanol, pH 7.0	20% v/v Propanol, pH 7.0	20% v/v DMSO, pH 7.0	2% v/v Phenol, pH 7.0	2% v/v Acetone, pH 7.0	20% v/v DMF, pH 7.0	20% v/v Acetonitril, pH 7.0
10	100 mM Guanidinium-chloride, pH 7.0	100 mM Urea, pH 7.0	100 mM NH <sub>4</sub> SO <sub>4</sub> , pH 7.0	1% w/v Protamine Sulfate, pH 7.0	100 mM Trimethylamine, pH 7.0	10% w/v PEG 6000, pH 7.0	10% w/v PEG 1000, pH 7.0	0.05% w/v OGP, pH 7.0
11	100 mM ITPG, pH 7.0	50 mM AICA, pH 7.0	50 mM Polyvinyl Pyrrolidone, pH 7.0	1% w/v myo-Inositol, pH 7.0	1% w/v tRNA, pH 7.0	1% w/v Glycogene, pH 7.0	1% w/v DNA pieces, pH 7.0	1% w/v Peptide mix, pH 7.0

### 2.3.3 Purification of CRM1

Cloning, expression and purification of human and *C.thermophilum* CRM1 was performed by Thomas Monecke (WG Ficner, University of Göttingen), who provided the following protocols. The cDNA of *Chaetomium thermophilum* CRM1 encoding the full-length protein (amino acids 1 - 1077) was cloned by using *EcoRI* and *NotI* restriction sites into a modified pET24d vector. The constructed GST-CRM1 fusion protein contains an inter-jacent tobacco etch virus (TEV) protease cleavable site and was expressed in *Escherichia coli* BL21(DE3) (Merck). Transformed cells were grown in kanamycin containing 2YT medium supplemented with 2 % (w/v)  $\alpha$ -D-glucose at 37 °C to an OD600 of 0.6. Temperature was set to 20 °C and protein expression was induced at an OD600 of 0.9 by adding IPTG to a final concentration of 0.6 mM. Cells were harvested after 20 h of induction (5,000 g, 20 min, 4 °C) and resuspended in lysis buffer (500 mM NaCl, 50 mM Hepes/NaOH, pH 7.5, 2 mM MgCl<sub>2</sub>, and 4 mM DTT). Cells were disrupted using a Microfluidizer 110S (Microfluidics), and the clarified lysate (30,000 g, 30 min, 4 °C) was subsequently loaded onto a GStrap column (GE Healthcare) equilibrated with lysis buffer. Unbound proteins were removed by washing with two column volumes of lysis buffer, and the bound fusion protein was eluted with a buffer containing additionally 30 mM reduced glutathione. For cleavage, the GST - [TEV] - CRM1 fusion protein was incubated with TEV protease (Invitrogen) at 4 °C overnight in a 1:100 molar ratio of protease:fusion protein. To remove remaining contamination with DNA and/or RNA, 5  $\mu$ L Benzonase (Merck) was added to the solution. The TEV-cleaved protein was further purified by using a Superdex S200 (26/60) gel filtration column (GE Healthcare) in a buffer containing 150 mM NaCl, 20 mM Hepes/NaOH, pH 7.5, 2 mM MgCl<sub>2</sub>, and 4 mM DTT. Fractions containing CRM1 were pooled and finally passed over another GStrap to remove remaining traces of GST and uncleaved fusion protein.

Human His<sub>6</sub>-CRM1 was expressed in *Escherichia coli* TG1 as described previously [75]. Cells were resuspended in lysis buffer (500 mM NaCl, 50 mM HEPES/NaOH pH 7.5, 2 mM MgCl<sub>2</sub>, 30 mM imidazole and 2 mM DTT) and disrupted using a microfluidizer 110S (Microfluidics). The clarified lysate (30,000 xg, 30 min, 4 °C) was loaded onto a HisTrap column (GE Healthcare) equilibrated with lysis buffer. Unbound proteins were removed by washing with 2 column volumes (CV) of lysis buffer and bound His<sub>6</sub>-CRM1 was eluted with a linear gradient of elution buffer (0-100% in 6 CV) containing additionally 400 mM imidazole. In order to remove C-terminal degradation products of His<sub>6</sub>-CRM1 the pooled protein fractions were desalted (50 mM NaCl, 50 mM HEPES/-NaOH pH 7.5, 2 mM MgOAc and 2 mM DTT) and loaded onto a Source 30Q anion exchange column equilibrated with the same buffer. Full length His<sub>6</sub>-CRM1 was eluted



with a linear gradient (0–100% in 30 CV) with a high salt buffer containing 500 mM NaCl, 50 mM HEPES/NaOH pH 7.5, 2 mM MgOAc and 2 mM DTT.

### 2.3.4 Purification of PDH

Preparations of the endogenous *E.coli* PDHc were performed as follows: 100 g of frozen cells were slightly thawed to the consistency of pastry dough, then mixed with 50 g of aluminium oxide and 20 mL of PDHc-buffer (50 mM imidazole, pH 6.5, 150 mM KCl, 5 mM MgCl<sub>2</sub>). The cells were broken in an electrical mortar for 45 min at 4 °C. During the last 10 min a pinch of DNase and RNase was added as well as an EDTA free protease inhibitor tablet. The ground mixture was diluted with 200 mL PDHc buffer and then centrifuged for 10 min at 10000 g to remove the aluminium oxide. The supernatant was further centrifuged for 30 min at 30000 g to pellet the cell debris. To remove nucleic acid contaminations and ribosomes, 0.1% of protamin sulphate was added while stirring. The mixture was centrifuged for 20 min at 30000 g and the pellet was discarded. Carefully, 6% (w/v) PEG6000 (from a 50 % stock solution) was added while stirring and incubated for 30 min at 4 °C. After pelleting for 30 min at 15000 g, PEG6000 to a final concentration of 9% was added and incubated for 30 min at 4 °C and then centrifuged for 30 min at 15000 g. The supernatant was discarded and the pellet was dissolved on a rotary shaker in 2 mL of PDHc buffer. The dissolved pellet was then loaded onto six SW40 10–30% (w/v) sucrose gradients and centrifuged for 16 h at 18000 rpm at 4 °C. The gradients were fractionated and analyzed by SDS-PAGE. Fractions containing the PDHc were pooled and supplemented with 10% (w/v) PEG 6000 incubated for 30 min at 4 °C and then centrifuged for 30 min at 15000 g. The pellet was dissolved on an orbital shaker in 200 µl PDHc buffer.

### 2.3.5 Functional Assays of the PDH

The activity of the PDHc was monitored in a coupled assay via the absorbance signal produced by NADH. Assays were performed in 100 mM TrisBase buffer pH 8.0. For one assay reaction, the sample was mixed with 2 mM MgCl<sub>2</sub>, 0.2 mM TPP, 2 mM NAD and 0.2 mM CoA. The reaction was started via addition of 2 mM pyruvate and the absorbance signal was measured at wavelength of 380 nm for 1 minute at 20 °C. The initial slope was used to estimate the overall activity of PDH in the solution.

### 2.3.6 Purification of AChBP

Purification of the *Biomphalaria glabrata* AChBP was performed by the group of Jürgen Markl (University of Mainz). The protein was over expressed in *E.coli* and purified via Nickel-NTA agarose and refolded according to Saur et al.[178]. The final gel filtration was replaced by a GraFix-gradient (cp section 2.5.1).

## 2.4 Fundamental Concepts of Single Particle Electron Microscopy

Single particle cryo EM employs a number of techniques specially designed for it. The concepts and ideas are unique in the field of biological research. Thus, the major principles used are explained in this section.

### 2.4.1 The Electron Microscope

The invention of the microscope was one of the significant advances for modern biology. By seeing tissues and single cells, biological research flourished tremendously. However, using visible light as information carrier imposes a crucial limit - the diffraction limit. Ernst Abbe was the first to describe this phenomenon. The maximal resolution, which can be achieved by any kind of microscope is limited by the wavelength through:

$$d = \frac{\lambda}{2n \sin(\alpha)}, \quad (2.1)$$

where  $d$  is the maximal resolution, meaning the distance between two structural features, which can be distinguished, and  $n \sin(\alpha)$  the numerical aperture of the objective lens. Thus, in the very best case, the resolution can be half of the wavelength. Typically, this is a couple of hundred nanometer. With this resolution, the visibility of cellular structural features is limited to organelles. The only way to overcome this challenge is to use waves of shorter wavelengths. Louis-Victor Pierre Raymond de Broglie showed in 1924 that every particle in principle has a wavelength  $\lambda$ :

$$\lambda = \frac{h}{m \cdot v}, \quad (2.2)$$

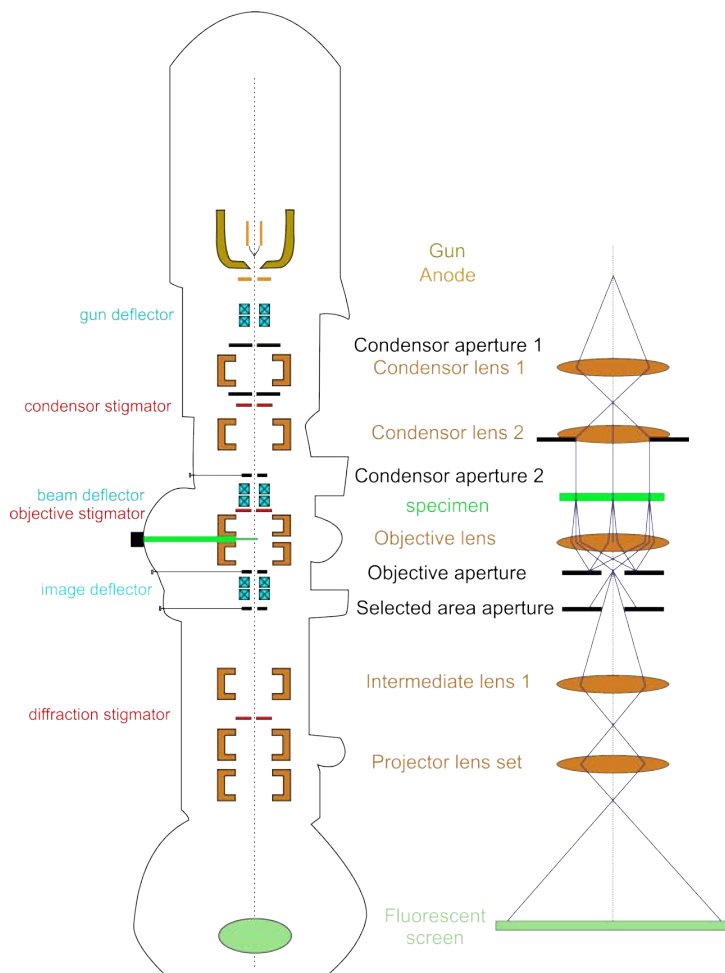
where  $h$  is the Planck constant,  $m$  the mass and  $v$  the velocity of the particle. Thus, one needs rather big particles with high speed. Already two years later Max Knoll invented an electromagnetic lens which could be used for an *electron microscope*, which he also patented. However, the first electron microscope was built by Ernst Ruska<sup>1</sup> together with Max Knoll in 1931 [105]. Even though its imaging capability improved tremendously in the last 80 years the overall architecture of a modern transmission electron microscope (TEM) resembles still the first one (see figure 2.1). In very general terms, a TEM can be compared to a light microscope. It consists of a light (here electron) source, where an electron beam is generated, a set of condensor lenses that insure perfect illumination of a specimen, a specimen holder, an objective lens, which produces the first image and a set of projector

---

<sup>1</sup>Trivia: Ernst Ruska was unaware of the de Broglie law when building the first electron microscope.

or intermediate lenses, which magnify the image on a detection device which digitizes the image information. While this seems indeed to be in principle very analogous to a light microscope, there are a few important differences, which are outlined in the following section.

**Figure 2.1: Scheme of a TEM.** A scheme of a typically used TEM is shown with all essential components. On the right, a principal ray diagram is shown illustrating the optical influence of the most important elements on the beam. Left: A typical electron microscope is shown with the most important components. Components relevant for the alignment of the beam within the column are labelled on the left side while the components important for the image formation are labelled on the right side of the scheme. For every component, the color of the label matches the color of the component in the scheme. Right: A typical ray diagram for the displayed microscope is shown. The beam is shown as a black outline. The figure was reproduced and altered with permission from C J Russo, MRC Laboratory of Molecular Biology, Introduction to Electron Microscopy Lectures, 2014



#### 2.4.1.1 The Electron Source

The most common electron source is a tungsten filament. It is heated to approximately 3000 °C, where the energy of the electron is large enough to leave the metal at a high vacuum, applied to the whole microscope imaging system. The 'evaporated' electrons are accelerated by an electric field with several hundred kilovolts. A hole in the negatively charged anode surrounded by a negatively charged cup (the Wehnelt cylinder) focuses the electrons (see figure 2.1).

Modern electron microscopes, however, use a Field Emission Gun (FEG). Hereby, the tungsten filament is reduced to a single pointed tungsten crystal (a few nm in diameter), coated with zirconium dioxide. The anode is flat and produces a strong field near the tip.

Electron beams generated in this way are both spatially and temporally more coherent and brighter.

In any way, electrons are accelerated until their kinetic energy matches the electrical energy of the applied electrical field ( $\Phi$ ) following:

$$e \cdot \Phi = \frac{1}{2}m_e v^2, \quad (2.3)$$

where  $e$  is the electron's charge and  $m_e$  its mass. Combining this with de Broglie's equation 2.2 leads to:

$$\lambda = \frac{h}{\sqrt{2m_e e \Phi}} \approx \frac{12[\text{\AA}\sqrt{V}]}{\sqrt{\Phi}}. \quad (2.4)$$

Modern microscopes are operated at acceleration voltages of 100-400 kV and can thus reach subangstrom resolution. Since the electrons reach more than half the speed of light, relativistic correction has to be applied, leading to the approximation formula [12]

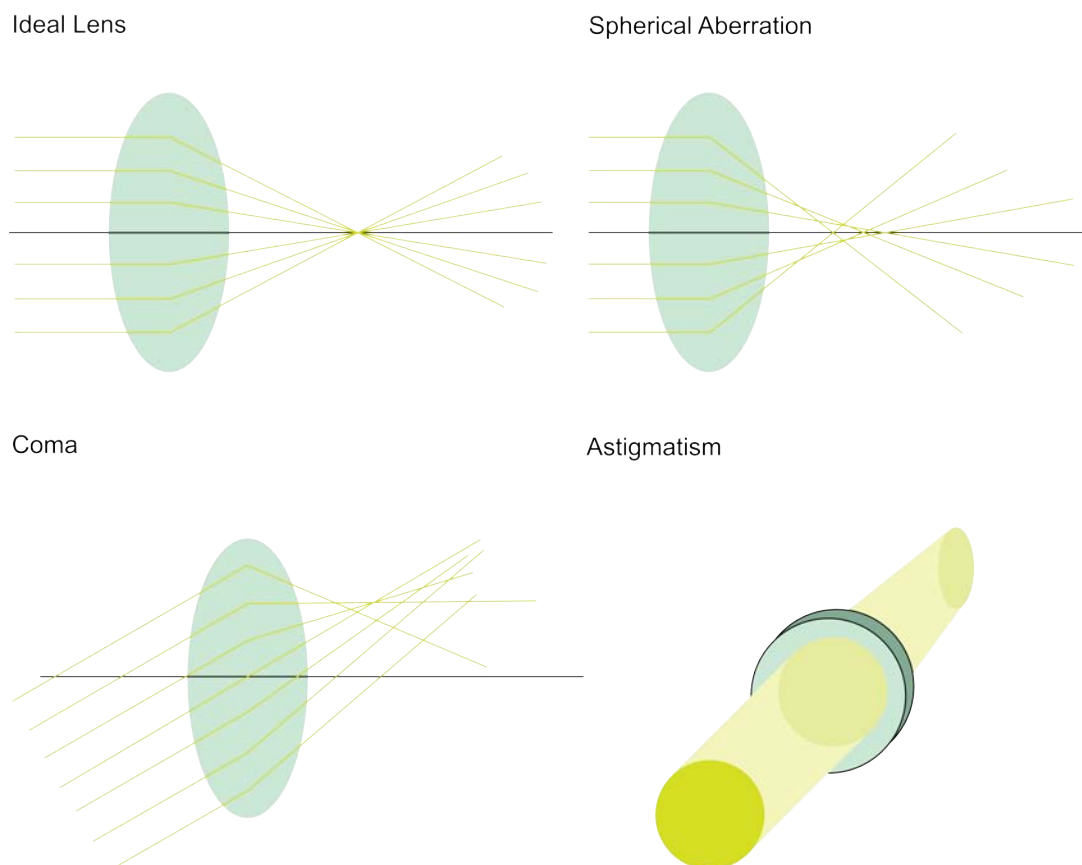
$$\lambda \approx \frac{12.3}{\sqrt{\Phi + 0.000000978\Phi^2}}. \quad (2.5)$$

The produced beam can then be readily focused by magnetic lenses.

## 2.4.2 Magnetic Lenses and their Aberrations

In great contrast to light microscopy, lenses of an electron microscope are only shaped magnetic fields and contain no matter. Due to the nature of magnetic fields, not all imaginable lens shapes can be created by them. This leads to imperfections of the lenses resulting in severe aberrations.

The most crucial of those are astigmatism, coma and spherical aberration (see figure 2.2). If two waves, propagating perpendicular to each other, have different foci, one speaks of astigmatism. If the wave front of the electron beam is not parallel to the optical axis when hitting a lens, comatic aberration will occur. Spherical lenses as they are found in an electron microscope are not able to focus all waves to single focal points. Waves closer to the optical axis will have a different focal point than those away from the optical axis. Here one speaks of spherical aberration. All of these lead to distortions of the images and need to be avoided as much as possible. Using a technique called *coma free alignment* [226], astigmatism and coma can be reduced to a minimum. This can even be improved by using a spherical aberration corrector [76].



**Figure 2.2: Most Important Aberrations for a TEM.** Schematically the optical definition of the three most common aberrations are shown. The electron beam is depicted in yellow, the lens in blue and the optical axis in black. For a perfect lens, all parts of the beam are focused in a single point. Lenses with spherical aberration focus beam parts closer to the optical axis in a different point than those farer away from the optical axis. In case of misalignment, the beam will hit the lens, not in a perpendicular fashion and different parts of the beam will have different focal points. This results in coma. If waves that are perpendicular to each other have different focal points one speaks of astigmatism. A circular beam will thus be transformed into an elliptical one.

### 2.4.3 Electron Detection

Originally, images of an electron microscope were detected using photographic films. Those directly detect the electrons through the reduction of silver ions to silver grains on the surface of the micrograph. Moreover, they cover a large area of the specimen and thus many particles can be seen per film. However, for later computer-assisted image processing, they have to be developed and scanned to digitize them, which is a work intensive and time-consuming task. Thus, charge coupled device (CCD)-cameras are nowadays more frequently used. Moreover, even a binned CCD can perform better than film in terms of electrons required to gain a particular contrast in the obtained image [174]. Since high energy electrons readily damage the device, certainly a conversion of the electron beam into a light beam via a scintillator is needed. This conversion, however, introduces extra noise. This challenge was recently overcome by the advance of the direct detection device (DDD) [50]. For example CMOS chips are hardened against electron radiation.

They produce images with an improved SNR due to a better DQE [125]. Moreover, they are fast devices and can in principle count single electrons and record multiple frames of an image. These frames were successfully used to correct for drift of the specimen [11, 180].

#### 2.4.4 Short Excursion into Fourier Space

At this point, an important mathematical principle has to be explained. Fourier transformations are a valuable tool to describe the diffraction patterns. Further, they are very helpful for image processing. [23]

The Fourier Transformation is a linear coordinate transformation, translating real space information like time  $t$  and distance  $d$  into reciprocal space information like temporal or spatial frequency  $k$ . In principle any given continuous and periodic function  $f(x)$  can be transformed back and forth with the equations:

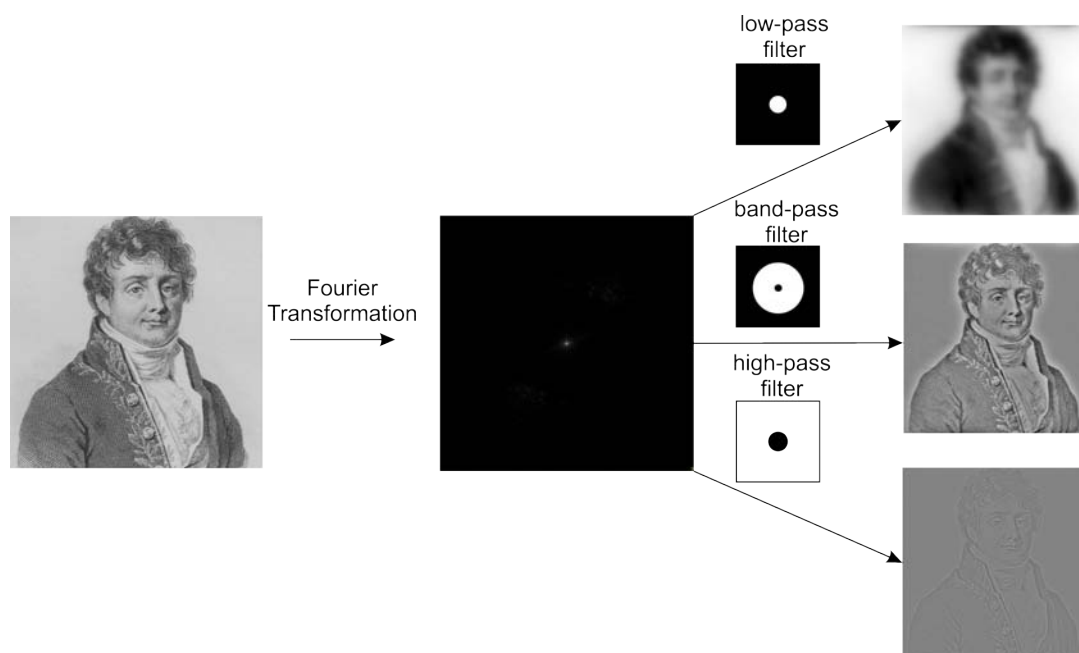
$$F(k) = \int_{-\infty}^{\infty} f(x)e^{2\pi ikx} dx \Leftrightarrow f(x) = \int_{-\infty}^{\infty} F(k)e^{2\pi ikx} dk \quad (2.6)$$

As any sequence of values can be understood as a function, any such sequence can be transformed. Images can thus be transformed as well. The most meaningful way to do so is the 2D Fourier transform, which first transforms every line of pixels into frequency space and afterwards every column of pixels of the onetime transformed image. Of note is, that the Fourier space representation consists of a real image describing amplitudes of the transformation and an imaginary part, describing phases that can not sufficiently be depicted in a single image. The value of the transformation lies in its representation. Small distances anywhere in the real space image correspond to high spatial frequencies, which are depicted near the edge of the image and large distances correspond to low spatial frequencies located in the center of the transformed image. Thus, the amount of detail in an image can be easily changed by simple filters (see figure 2.3).

Moreover, a few properties of the Fourier transform are valuable for image processing they will be outlined through the further description of the workflow.

#### 2.4.5 Image Formation

The image is formed through the direct interaction of the electrons with matter. An electron penetrating a specimen can either go through it unchanged, be deflected or absorbed. When they collide or nearly collide with a nucleus, they transfer energy to the specimen (inelastic scattering). This energy transfer can result in knock-on of nuclei, specimen ionization, X-ray emission, formation of free radicals or secondary electron



**Figure 2.3: Use of Fourier Filters.** Here a picture of Jean Baptiste Joseph Fourier was Fourier transformed and different masks (black parts of the mask will be masked out, white parts remain) were applied on the transformed image. Subsequently, the image was transformed back. Masking the outer part of the Fourier transformed image leads to removal of high spatial frequencies, meaning image details. Masking out the center part will remove low spatial frequencies that define the overall shape.

scattering. All of them will damage the specimen and contribute to the noise in the image. Due to those effects, the specimen cannot withstand high dosage of electrons and thus the resulting SNR is low. Inelastically scattered electrons will be removed from the beam or are scattered to high angles and can be removed by an objective aperture. They contribute to the so-called *amplitude contrast*. Further devices, called *energy filters*, can be added to the microscope column to remove low energy electrons to increase that contrast. Electrons can also be deflected by the specimen without losing energy (elastic scattering). Those electrons will contribute to the contrast in the image and do not harm the molecules and are thus favored interactions. However, inelastic scattering can by no means be avoided and is even three times more frequent than elastic scattering [83]. Moreover, the deflection angles of elastically scattered electrons are relatively small. So they do not contribute to the amplitude contrast. Still, the path length of the electron wave through the specimen is altered, resulting in a phase shift. Since the density of the biomolecular specimen ( $1.33 \frac{g}{cm^3}$  for protein respectively  $1.84 \frac{g}{cm^3}$  for nucleic acid) is slightly larger than the one of the surrounding ice ( $0.92 \frac{g}{cm^3}$ ), the phase shift is stronger in the specimen than in the ice and thus location dependent [131]. The incident wave  $\Psi_0$  is modified as it travels through the specimen (The derivation follows in principle [55]). The



density and thus the Coulomb potential  $\Phi(x, y, z)$  distribution can be described. Thus, the exiting wave contains this information in the form:

$$\Phi(x, y) = \int_{-t/2}^{t/2} \Phi(x, y, z) dz, \quad (2.7)$$

where  $t$  is the specimen thickness. The value of the integrated Coulomb potential is at least proportional to the phase shift applied on the exiting wave  $\Psi_{ex}$  :

$$\Psi_{ex}(x, y) = \Psi_0 e^{i\sigma\phi(x,y)}, \quad \text{with} \quad \sigma = \frac{m_e \lambda}{2\pi \hbar^2} \quad (2.8)$$

For biological specimen, which contain only small nuclei,  $\Phi(x, y)$  is small. They are called *weak phase objects*. Thus, using Euler's formula<sup>2</sup>, this leads to:

$$\Psi_{ex}(x, y) \approx \Psi_0(1 + i\sigma\Phi(x, y)) \quad (2.9)$$

One can approximate the exiting wave as the sum of the incident wave altered by the perpendicular scattering wave. The intensity distribution  $I(x, y)$  can be estimated as:

$$I(x, y) = \Psi_{ex}(x, y) \cdot \overline{\Psi_{ex}(x, y)} \approx 1 + (\sigma\Phi(x, y))^2, \quad (2.10)$$

where  $\overline{\Psi_{ex}(x, y)}$  is the complex conjugate of  $\Psi_{ex}(x, y)$ . Since  $\Phi(x, y) \ll 1$  the contrast is practically not detectable. By introducing an additional phase shift of  $90^\circ$ , the exit wave is described by:

$$\Psi_{ex}(x, y) \approx \Psi_0(1 - \sigma\Phi(x, y)) \quad (2.11)$$

leading to an intensity distribution of:

$$I(x, y) = \Psi_{ex}(x, y) \cdot \overline{\Psi_{ex}(x, y)} \approx 1 - 2\sigma\Phi(x, y), \quad (2.12)$$

which is now linearly dependent on the phase shift. In phase contrast light microscopy, such a phase shift is introduced through a phase plate. Development of a similar device for electron microscopy is ongoing since more than 60 years [137] and has still not led to a satisfying solution. The main problem is that most ideas for phase plates need to introduce devices in the column, which will quickly get contaminated or charged.

Alternatively, one could use the imperfections of the electron lenses to introduce a phase shift. In reality, the recorded image is not the direct mapping of the Coulomb densities. The imaging properties of the microscope and the noisy nature of the signal lead to signal

---

<sup>2</sup>Euler's formula:  $e^{ix} = \cos(x) + i \sin(x)$

degradation via the convolution of the exiting wave with a point spread function (PSF) to the observed wave  $\Psi_{\text{obs}}(x, y)$ :

$$\Psi_{\text{obs}}(x, y) = \Psi_{\text{ex}}(x, y) \star \text{PSF}(x, y) \quad (2.13)$$

Thus, every density point in the object is smeared out over a greater area in the image. A more convenient way to describe this is the use of Fourier transformations, where convolutions become multiplications.

$$F(\Psi_{\text{obs}}(x, y)) = F(\Psi_{\text{ex}}(x, y)) \cdot F(\text{PSF}(x, y)) = F(\Psi_{\text{ex}}(x, y)) \cdot \text{PhCTF}(\vec{R}) \cdot E(\vec{R}), \quad (2.14)$$

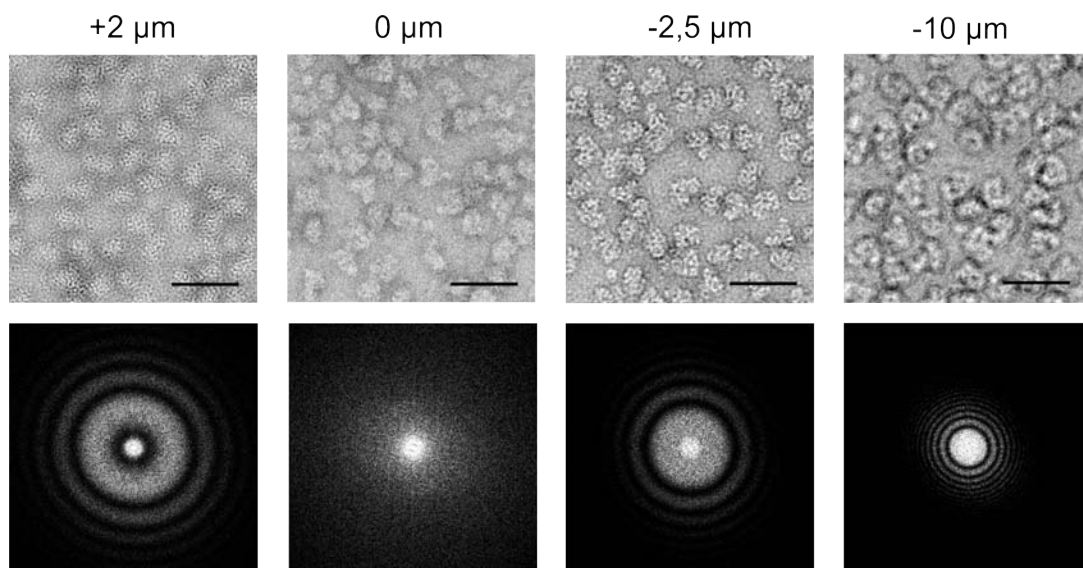
where the Fourier transform of the PSF is called the phase contrast transfer function (PhCTF) and E is an envelope function describing the decay of the signal in higher spatial frequency ranges and  $\vec{R}$  the Fourier space coordinates (the spatial frequencies). The PhCTF itself can be approximated as:

$$\text{PhCTF}(\vec{R}) = e^{-i\gamma(\vec{R})}, \quad \text{with} \quad \gamma(\vec{R}) = -2\pi \left( \frac{1}{2} \Delta \lambda \vec{R}^2 - \frac{1}{4} C_s \lambda^3 \vec{R}^4 \right), \quad (2.15)$$

where  $\gamma$  is the introduced additional phase shift,  $\lambda$  the wavelength of the electrons,  $\Delta$  the defocus and  $C_s$  the spherical aberration constant [185]. Thus, defocusing the beam will lead to an extra phase shift and a phase contrast image with visible amplitudes can be recorded.

## 2.4.6 Image Restoration

While defocusing leads to better contrast, the resulting images are heavily degraded as figure 2.4 suggests. The contrast oscillates for different spatial frequencies between positive and negative contrast. Of course, in an ideal image, the contrast should be transferred equally for every spatial frequency. To provide valuable information, the image has to be restored in a process called *CTF-correction*. Easiest would be to divide the Fourier transform of the image by a PhCTF (cp. equation 2.14). However, at certain frequencies the PhCTF becomes 0, and thus division is not defined at these spots. Therefore, mostly an heuristic multi-step procedure is used. First, the PhCTF has to be determined. The squared Fourier transformation of the image called *power spectrum* depicts a ring pattern - the Thon rings. These can be approximated numerically to estimate the parameters of the PhCTF ( $C_s, \lambda, \Delta$  and the astigmatism) and the envelope function. With these parameters, a PhCTF can be modelled, and the zero crossings of it can be determined.



**Figure 2.4: Effect of Defocus.** The Anaphase promoting complex is shown in negative stain at different focus values. Upper row shows a micrograph at the given focus value. Positive values represent overfocus, negative values underfocus. In the lower row, the power spectrum of the respective micrograph is shown. The higher the focus, the closer are the Thon rings.

In a step called *phase flipping*, the Fourier transformed image is multiplied with the sign of the modelled PhCTF:

$$F(\Psi_{restored}(x, y)) = F(\Psi_{obs}(x, y)) \cdot \text{sgn}(\text{PhCTF}(\vec{R})) \quad (2.16)$$

In this way, all image parts with negative contrast are removed. However, there are still the roots of the PhCTF, meaning frequencies, where all information is lost. Those can only be restored by averaging images taken at different defoci and thus different PhCTFs with different roots. Lastly, the decaying envelope function has to be removed from the data. In an ideal image, the amplitude as every spatial frequency of the signal should be equally high. Easiest, this amplitudes can be restored by multiplying with a B-factor-decay curve of the form  $e^{-\frac{B}{4d^2}}$ , with  $d$  being the real space coordinate. In an alternative approach, amplitudes could also be gained through solution X-ray scattering experiment [59] or theoretical scattering curves [169].

The image restoration procedures are diverse and different software packages use different approaches [151].

## 2.4.7 Image Processing Basics

In order to provide a consistent overview over the image processing workflow, a few mathematical remarks regarding images need to be made. Images can be described mathematically in different ways: first and most easily an image  $I$  containing  $n$  pixels can be

described as set of all its grey values  $p_i$ . On this very simple basis, set statistics can be used. The mean  $\bar{p}$  and standard deviation  $\sigma_p$  of  $p$  can be determined:

$$\bar{p} = \frac{1}{n} \sum_{p_i \in p} p_i \quad \sigma_p = \sqrt{\frac{1}{n-1} \sum_{p_i \in p} (p_i - \bar{p})^2} \quad (2.17)$$

and the grey values can be normalized using to a new mean  $\bar{p}_{new}$  and standard-deviation  $\sigma_{p,new}$  :

$$p_{norm} = \frac{(p - \bar{p}_{current} + \bar{p}_{new}) \cdot \sigma_{p,new}}{\sigma_{p,current}} \quad (2.18)$$

Even though this kind of notation is easy to handle, the spatial relationship between the individual pixels is lost in this way. In the simplest way this can be achieved by treating the set of grey values as a sequence  $s$  of grey values that is formally a relation:

$$s : \{1, \dots, n\} \rightarrow p \quad (2.19)$$

$$i \mapsto s_i$$

An equivalent notation would be a vectorial notation:

$$\vec{s} = \begin{pmatrix} s_1 \\ s_2 \\ \vdots \\ s_n \end{pmatrix} \quad (2.20)$$

Thus, the position of each pixel is unique within the sequence, and the spatial information is decoded and can be treated with mathematical procedures. This notation is especially useful for pixel-wise statistical and algebraic treatments like the image averaging. Since the spatial relationship between two pixels is important for rotation or translation of the image, exact coordinates of each pixel are required. Therefore, every pixel  $p_i$  can be notated as three-dimensional vector  $\vec{p}_i$ :

$$\vec{p}_i = \begin{pmatrix} x_i \\ y_i \\ 0 \\ p_i \end{pmatrix} \quad (2.21)$$

and the whole image can be summarized as Matrix I:

$$I = \begin{pmatrix} x_1 & y_1 & 0 & p_1 \\ x_2 & y_2 & 0 & p_2 \\ \vdots & \vdots & \vdots & \\ x_n & y_n & 0 & p_n \end{pmatrix} \quad (2.22)$$

With these notations in principle, all kinds of mathematical treatments of images used in the further explanation can be described.

### 2.4.8 Image Preprocessing

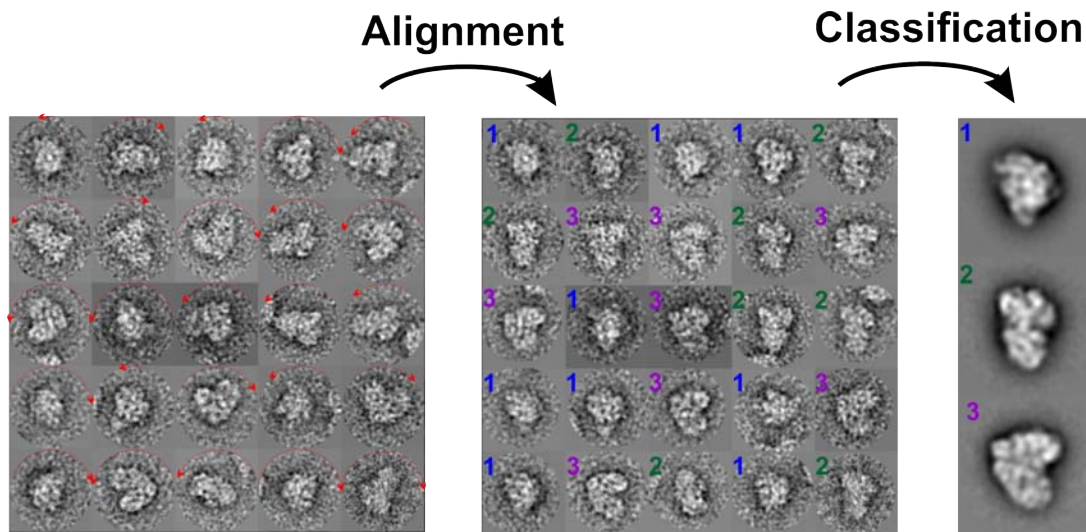
Once microscopic images, called *micrographs*, are generated, individual particles need to be identified and extracted from the images. This step, called *particle picking*, is often tedious, and time-consuming, but still needs special care and accuracy since it prepares the raw data for every following step. Due to the low SNR, particles are barely visible and often artefacts on the grid like ethane or ice crystals can be easily confused as particles. Only a few years ago, picking was done manually. Nowadays, a large number of software packages are available. Those use template matching algorithms [33], image statistics [216], machine learning [114] or combinations of those [230] to identify the particles. None of those performs perfectly on every data set, but all have their advantages [232]. Thus, additional steps to remove artefacts or false positives from the data set are required.

Once the particle images are extracted, they are preprocessed to enhance the weak SNR. For the first few image processing steps, the images are down-sampled (coarsed), Fourier filtered to remove gradients as well as noise and normalized.

### 2.4.9 Alignment and Classification

Once the data is collected and preprocessed, the goal is to reconstruct the 3D volume of the molecule. During sample preparation, the molecules are distributed in different orientations on the surface of the grid. The molecules thereby have 6 degrees of freedom to orient themselves on the grid: The three translations (x,y,z) along and the three rotations, called *Euler angles* ( $\alpha, \beta, \gamma$ ), around the coordinate axes are sufficient to describe any orientation of a molecule on the grid. Since the TEM produces projections along the z-axis, the information in z direction is integrated in the dataset. To be able to reconstruct a volume of the molecule, the theoretical molecule has to be set to the origin of coordinates. For this, all remaining five parameters (x- and y- shift,  $\alpha$ -,  $\beta$ - and  $\gamma$ -rotation) have to be determined for each particle image. The previously described picking procedure should

already have roughly determined translational parameters. Thus, only minor shifts of a few pixels are still necessary. The determination of the missing parameters needs various computational image processing steps to be iteratively repeated.



**Figure 2.5: Alignment and Classification.** The principle of alignment and classification is demonstrated on the example of real data from the Anaphase Promoting Complex. Alignment orients the particle images according to references, so that they superimpose best with the reference. Therefore, translation and rotation of the images is necessary. Here, only the rotation is shown, visualized as red arrows. After alignment, the images are grouped into similar appearing images via classification. Here, the images are sorted into three groups. Subsequently images of the same class are averaged. The averages can then be used as references for a new iteration of alignment.

Before beginning the angular parameter determination, the molecules have to be sorted according to their orientation by a procedure called *classification*. To facilitate this, particle images representing the same orientation have to be superimposed using an *alignment* procedure. Both of them are demonstrated in figure 2.5 and will be outlined further in the following sections.

#### 2.4.9.1 Alignment

To superimpose (align) two particle images, three of the five degrees of freedom have to be determined: the x- and y-shift and the  $\alpha$ -rotation called in-plane rotation. The easiest way to apply such a transformation to an image is the use of an affine transformation with the matrix  $T$ :

$$T = \begin{pmatrix} \cos\alpha & -\sin\alpha & \Delta x & 0 \\ \sin\alpha & \cos\alpha & \Delta y & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (2.23)$$

The aim is to optimize the parameters  $\alpha$ ,  $\Delta x$  and  $\Delta y$  in such a way that the distance between a reference image  $I_{ref}$  and the transformed image  $I_t = I \cdot T$  is minimal. Commonly, the distance is calculated as the Cross Correlation Coefficient (CCC) between the reference image  $s_{ref}$  and the transformed image  $s_t$  using the sequence representation (see formula 2.20) as:

$$CCC = \frac{n \sum_{i=1}^n s_{ref,i} \cdot s_{t,i} - \sum_{i=1}^n s_{ref,i} \sum_{i=1}^n s_{t,i}}{\sqrt{\left(n \sum_{i=1}^n s_{ref,i}^2 - \left(\sum_{i=1}^n s_{ref,i}\right)^2\right) \sqrt{\left(n \sum_{i=1}^n s_{t,i}^2 - \left(\sum_{i=1}^n s_{t,i}\right)^2\right)}} \quad (2.24)$$

This measure is especially advantageous since it simplifies many different calculation steps to only a few. The CCC is simply just a value of the Cross Correlation Function (CCF), which is the convolution of the particle image with the mirrored reference image:

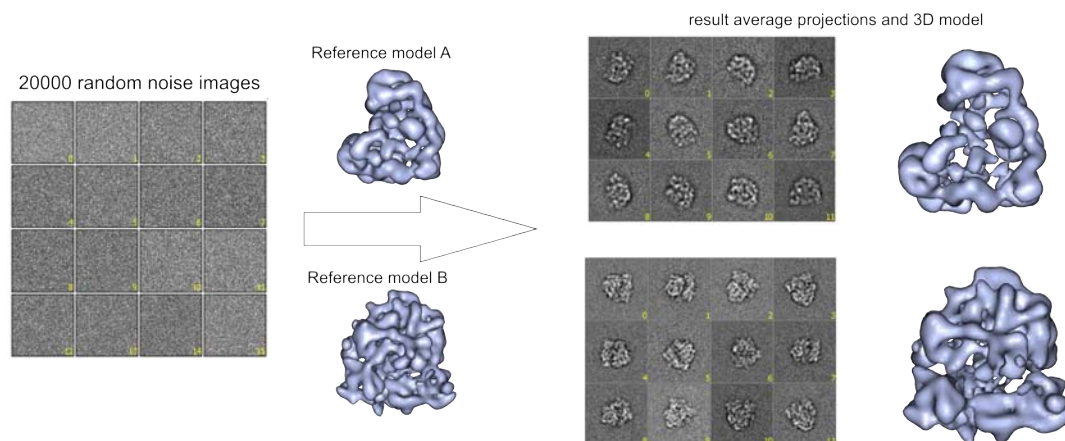
$$CCF(\vec{s}) = f(\vec{s}) \star g(-\vec{s}) = F^{-1}(F(f(\vec{s})) \cdot F(g(-\vec{s}))), \quad (2.25)$$

which is easily calculated as the multiplication of the Fourier transformed images. The resulting function will peak at coordinates corresponding to the shift to be applied. In a similar manner, the rotational parameters can be determined.

However, there are numerous different approaches described in the literature of how the actual alignment is calculated. Some are exhaustive, meaning all possible alignment parameter combinations are tried, some use an invariant transformation to simplify the search [98]. Today, the best results can be gained from an exhaustive alignment with resampling to polar coordinates [173]. After successful alignment, the three transformation parameter of the best fitting orientation is applied to each image. Also, a Maximum Likelihood (ML) approach could be used, and every possible orientation will be weighted with its probability. The final resulting aligned image is then the weighted average of all possible transformations of the original image. Even though this procedure is very computational demanding, it provides excellent results and is thus heavily used [189].

Since it is not a good idea to align a noisy image to another one [131], the question of the origin of the reference image remains. The easiest way is to gain it from an already previously solved model. However, this is often not available and even dangerous due to model bias. At this point, it is important to mention that alignments are very sensible to model bias. Many studies have shown, that alignments can reproduce any given reference just by aligning pure random noise [209, 85, 188, 198, 19]. Figure 2.6 illustrates this behaviour.

Thus, one has to choose references very carefully. The best approach is to derive a reference from the data set itself. At least, the reference should have roughly the size of the molecule and be rather noise free. In a procedure termed originally *alignment via*



**Figure 2.6: Demonstration of Model Bias.** To demonstrate the effect of model bias 20000 images containing only random noise were created *in-silico*. Those were aligned against an APC/C model and a 70S ribosome mode. The resulting 2D sums and the calculated 3D models are shown on the right. It can be seen that aligned noise is sufficient to reproduce any arbitrary model perfectly.

*classification*<sup>3</sup>, all particle images are averaged, resulting in a circle with a diameter of the particles size [47]. The data is then translationally aligned to this average. This is sufficient to center the particle images in the image frame and therefore reduces the translational degree of freedom to a minimum. The aligned images are classified (see section 2.4.9.2) and with a sufficiently large data set, a few resulting class averages will be good enough to serve as better references for a subsequent round of alignment. Since now, several reference are available, a procedure called *multi reference alignment* (MRA) is used. Here, the best alignment parameters are searched for every combination of particle image and reference. In the easiest case, the images are then assigned to the reference, they superimposed best. Alternatively, a maximum likelihood approach could be used to assign the images to all possible references with a weight proportional to the probability of this assignment [183].

### 2.4.9.2 Classification

While alignment orients the particles in the 2-dimensional image frame, it also has a certain capacity for sorting the images in different classes, if multiple references are given. However, depending on the SNR of the images this procedure can be very much biased towards the set of given references and no class other than already given by the set of references will occur.

To find the images that represent the same projection angle of the molecule, one needs to classify the images in a rather unbiased statistical procedure, coming from the Multivariate Statistical Analysis methods. This approach was suggested by van Heel and Co-workers [21] and is outlined here again. When describing the images (consisting of

<sup>3</sup>In many reviews this procedure is referred to as *reference-free alignment*.



n pixels) as vectors with n dimensions, similarity can be directly expressed as the distance between these vectors. However, the direct calculation and classification of those distances in datasets with usually more than 20000 images with each at least 10000 pixels is a computationally almost unbearable task. Moreover, the data consists to a large extent of random noise, which holds no information and should not be the basis of any similarity measure. Thus, the images should be reduced to only a few data points, which describe only the signal in the data best. This is done best by a principal component analysis (PCA). This analysis searches for new coordinate axis of the n-dimensional space, which describes the biggest variances in it best. Thus, the images can be described as a linear combination of those variances. The standard approach to calculate such variances is via an eigenvector-eigenvalue problem.

In this case, the problem will be formulated as

$$D\vec{v}_i = \lambda\vec{v}_i, \quad (2.26)$$

where  $\vec{v}_i$  are the eigenvectors that fulfil the equation to a certain eigenvalue  $\lambda$ . D is a matrix, where each element  $d_{kl}$  describes the covariance of the grey value  $x_k$  of the pixel on position k with the grey value  $x_l$  on position l throughout a dataset containing m images with n pixels. Every element  $d_{kl}$  of this covariance matrix can be thus calculated as:

$$d_{kl} = \sum_{j=1}^m (x_{kj} - \bar{x}_k)(x_{lj} - \bar{x}_l) \quad (2.27)$$

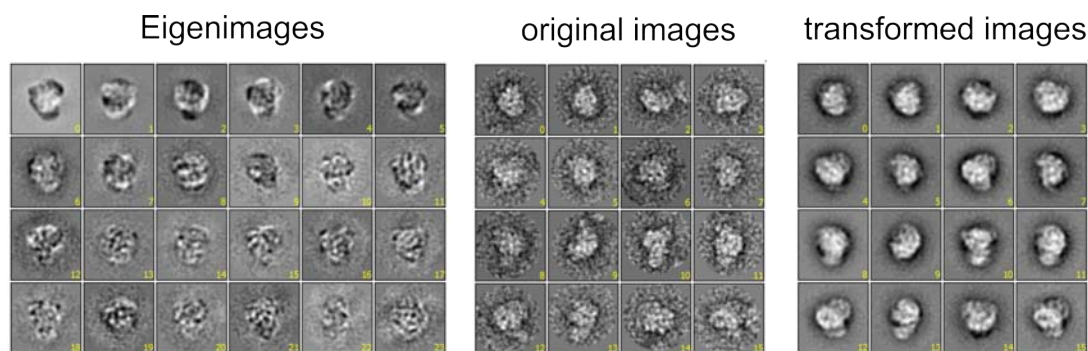
The solution to equation 2.26 can be gained through diagonalization of the matrix D. However, stochastic algorithms will lead quickly to a good result and are thus favoured.

The analysis results in at most n eigenvectors, which are all orthogonal to each other. Thus, they are defining a new coordinate system. This in turn describes the information in the data much better since the eigenvectors point hierarchically into the direction of the highest variances. The higher a corresponding eigenvalue, the more significant the described variance is for the data. Now the data can be described as a linear combination of those eigenvectors:

$$\vec{s}_i = a_1\vec{v}_1 + a_2\vec{v}_2 + \dots + a_n\vec{v}_n \quad (2.28)$$

with the linear factors  $a_r$ . In praxis only a few eigenvectors are used and thus only a few parameters  $a_r$  are necessary to describe an image sufficiently. Mostly, the similarities and differences between the images can be sufficiently described with less than 100 eigenvectors. This enormous data reduction will also remove most of the noise from the

transformed images. Of note: the eigenvectors are themselves vectors of the dimension of the image vectors and thus can be understood as images (eigenimages) themselves (see figure 2.7). Since they describe the biggest variances in the data set, they are important tools in an EM analysis and have to be evaluated carefully.



**Figure 2.7: Exemplified Principal Component Analysis.** A Principal Component Analysis was performed for 1000 particles images of the Anaphase Promoting Complex. The first few eigenimages as shown. It can be seen, that while the first ones represent the molecule’s shape very well, the later ones appear to be much noisier. With the shown eigenimages 16 exemplified original particle images were transformed in the eigenspace. The resulting images much better represent the molecule and are noise free. However, their information content is reduced tremendously.

Once the data reduction is complete, the compressed data can be classified. There are several different classification algorithms available. The most common are hierarchical clustering and k-means clustering [54].

In hierarchical ascendant clustering, each image represents a single class, and all interclass distances are calculated. The two classes with the smallest distance are merged, and its intra-class variance is calculated. The algorithm progresses by merging classes with small interclass distances while keeping the intra-class variance minimal. Once a given number of classes are formed, the algorithm stops. The whole algorithm can also be fulfilled in the opposite direction by starting from one class containing all images and successively splitting it (hierarchical descendant clustering). Starting from many classes, will have the advantage that outliers stay alone in their own class, whereas the starting from one class will result in a more uniform class size and thus uniform SNRs among the classes. Thus, both approaches find useful applications during image processing.

While the described approaches are deterministic, meaning with the same set of images and parameters, the classes will always be the same, the k-means approach is stochastic. Here k random points in the hyperspace are chosen. All the images get assigned to the point nearest to them. Once this is done, the center of the classes is calculated. Now again, the images are assigned to the nearest class center, and the whole process is iterated. After a few iterations, the class center position will converge. However, the procedure is mostly dependent on the initial seeds and can result in rather heterogeneous classes. To overcome

this problem, the method can be either repeated a couple of times and only the classes which reoccur are used or it can be combined with a maximum likelihood approach.

The distance metric between the particles is one of the most important parameters. Most commonly the direct Euclidean distance between the points in hyperspace is chosen. However, there are also  $\chi^2$  or modulation distances in use [55].

The resulting class averages, meaning the average image of all class members, derived by either method are evaluated, and good ones are used as references for a new round of alignment and classification. The number of particles per class will be successively decreased, and the procedure is reiterated until stable class averages with good SNRs are obtained.

## 2.4.10 Model Building

The resulting stable class averages should now represent all the different projection directions of the molecule present in the data set. They should be mass centered, and thus, the translational degrees of freedom are determined. For the model, only the rotational degrees of freedom need to be defined. In principle, three different main approaches and a few newly described ones exist. They are called projection matching, angular reconstitution and random conical tilt (RCT). In the following sections they are outlined.

### 2.4.10.1 Projection Matching

Projection matching is the most simple, but also the most biased method to determine the angles of particles images [154]. Similar to crystallographic molecular replacement methods, a model of the investigated molecule or a homologue is already known. This model is projected *in-silico* to create reference images for an alignment. During alignment, each experimental image gets assigned to the best matching reference and its corresponding projection angles. While this is a fast and easy method to determine the angles, it is heavily biased as the alignment itself is (see 2.4.9.1) and cannot be used, unless the model was correctly validated. Nevertheless, projection matching is used for the later refinement of the structure since the angular accuracy for a correct model is the highest among all methods.

### 2.4.10.2 Angular Reconstitution

Mostly, no initial model is available and thus the projection angles have to be derived directly from the data itself. A purely mathematical approach, needing no extra informa-

tion besides the data set itself is called angular reconstitution. It is based on the Fourier slice theorem, which says [22, 23]:

**Theorem 1.** *The Fourier transform of a parallel 2D projection of a 3D object equals a central section through the Fourier transformation of the 3D object. It follows: If two projections are derived from the same 3D object, their 2D Fourier Transformations will share at least one central section through the center of the transformed image.*

Apart from highly symmetrical objects, with only two projections the angle between those projections cannot be determined. But already a set of three projections gives rise to three common lines, which is sufficient to determine the angular orientations.

While equal line sections are hard to determine in Fourier space, a real space analogous theorem, the common lines theorem, was developed [208]. If one transforms the common line from Fourier space back into Real space, they represent 1D projections of the original 2D projections. Since they are equal in Fourier space, they also have to be equal in real space. One can thus derive that two 2D projections from the same 3D object will share at least one 1D projection.

With this, one only needs to find this common line. To do so, Radon transformations, also called sinograms, are calculated. These are the set of all possible 1D projections of a given 2D image. In a second step, all combinations of 1D projections are compared for similarity and thus the Euler angles are determined.

In a common line, the image information is condensed to only a few pixels. Thus, only a few parameters can be used to compare them pairwise. A recently suggested *voting algorithm* solves this problem by performing a consistency check with all the projection images [191].

### 2.4.10.3 Random Conical Tilt

Ideally one would like to gain angular information through an experiment. The only possible approach to do so is to take several images of the same particle from different directions. This can be achieved by tilting the specimen holder in the TEM by a defined angle. The angle between the two projections then corresponds directly to the experimental rotation angle applied to the EM specimen holder. From the locations of corresponding particles in the tilted and non-tilted image this holder rotation can be even calculated and thus be verified. The distances between two particles will change following geometrical rotation operations and can thus be calculated [161]. The caveat of this method is, that the sample has to be exposed to the electron beam more than once. To minimize the radiation damage only a pair of images from the same particles are recorded: in a method called random conical tilt (RCT), one image at no tilt and one at a

high tilt angle (typically 45 -60 °). The particle images are selected while keeping the pair relation. Consequently, the untilted images are preprocessed and several iterations of 2D processing involving alignments and classifications are performed until convergence. A class average contains about 40-100 images which are averaged to only one image. Due to a random in-plane rotation in the untilted image, the corresponding tilted images orient all around the circumference of a cone. With this, from each class average, a 3D model can be calculated. However, these 3D models lack information in certain directions, describable as a *missing cone* in Fourier space and are thus distorted in an direction in real space. To minimize this effect, several RCT-volumes could be averaged after 3D alignment and classification [175]. Nevertheless, they can be used as initial models for a refinement of the whole data set. Alternatively, recently a modification of RCT called *Orthogonal Tilt* was developed [117]. Hereby, both images are tilted one +45° and one -45°. The total rotation is thus 90°, which does not lead to a missing cone. However this method requires an even distribution of different orientations of the particles on the grid. Moreover, tilting the image is technically challenging. Tilting often leads to charging of the specimen's surface resulting in blurring of the image under cryo conditions. In negative staining, dark stain rims around the particles can arise in the tilt images if they are not sufficiently embedded in stain. Thus, one should always be careful with tilt methods and their results.

### 2.4.11 Reconstruction

Once all the missing orientational parameters are determined, a 3D model can be calculated. Again many different reconstruction methods exist, which all come to very similar results [152].

The task of reconstruction is to calculate a discrete density distribution of a molecule in 3D from 2D projections. This volume is commonly represented by a 3D-array of the  $N \times N \times N$  grey values called voxels, and the 2D projections are 2D-arrays of  $N \times N$  grey values.

In principal the reconstruction can be calculated in real space or in Fourier space. Real-space methods can be subdivided in *backprojection methods* and *algebraic methods*. Algebraic methods are based on the fact that the numerical representation of a projection can be described as linear equation system. The searched volume is a discrete array of  $N \times N \times N$  3D-pixels (voxels). One pixel of a projection of this volume in a given direction is the weighted sum of the grey values of the voxels, intersecting with a normal vector on that pixel. Each pixel of every projection can thus be represented as one linear equation. A dataset giving rise to  $M$  projections can be translated in  $M \times N^2$  equations. One searches for  $N^3$  unknowns. Thus at least  $N$  projections are needed to find one algebraic solution.

However, in reality, projections are not consistent with each other, because of the noise. Exact algebraic solutions are at least impracticable. The available algebraic methods try thus only to optimize iteratively the difference between measured and calculated projections, without actually attempting to solve the equations. [152]

A simpler and thus faster approach is the back projection. Each projection is hereby simply smeared out into 3D space along its normal vector. The grey value of a given pixel in a given projection is added to the grey value of each voxel that intersects with the normal vector of the pixel. This will result in smeared out details and a surrounding halo. This, however, can be simply corrected by a filter function, which is convoluted with the original projections. Which filter should be used, depends on the symmetry, the noise level, the number of projections and on the size of the volume. [80]

The Fourier space reconstruction is one of the first applied techniques [41]. It is based on the Central Section Theorem (cp. section 2.4.10.2). Projections of a 3D object in real space are central sections of the Fourier space representation of the object. Thus, a Fourier volume can be simply assembled from the projections knowing their angular relation, which can be readily translated into real space with all the problems related to the required interpolations in Fourier space.

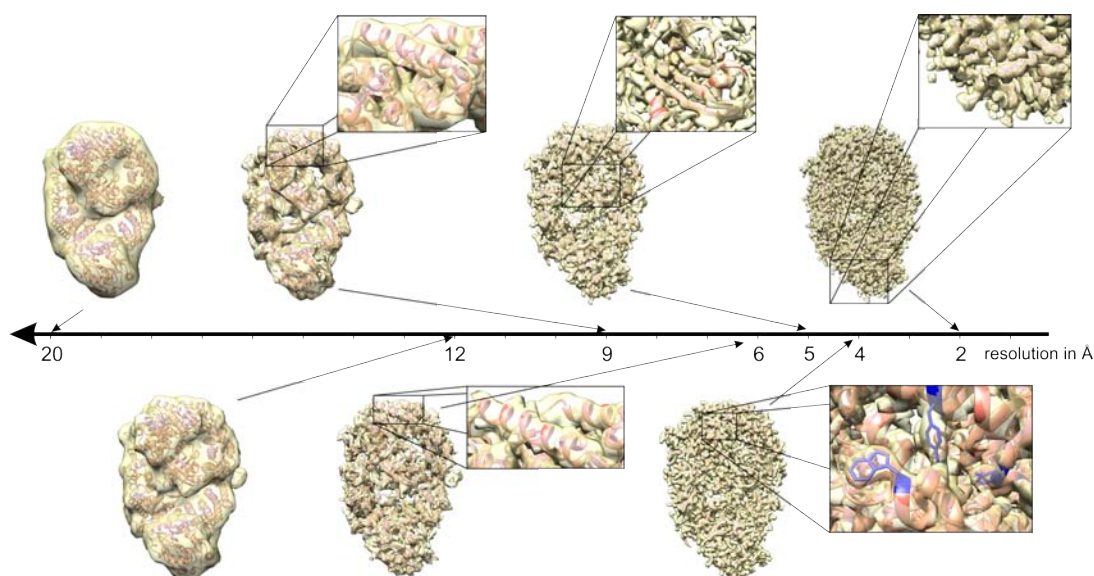
### 2.4.12 Refinement

All of the methods described here will yield only a first estimate of the unknown orientation parameters and an initial low resolution 3D model. There are several reasons for that: First of all, reference-free alignment and classification will hardly yield class averages depicting all possible orientations of the molecule. Furthermore, the images are downsampled and filtered initially to increase the SNR to facilitate a first reconstruction. Thus, structural features of a first model are rough. To get a better resolved structure, the initial model is refined. This is mostly done through a projection matching approach (cp. section 2.4.10.1). Hereby, equally spaced projections are calculated from the initial model. These are used as references for an alignment. Images aligning to the same projection are averaged and the projection angles are assigned to the average. A new structure is calculated, which will show more structural details. This process is iterated while the angular grid of projection angles is decreased and the initial filters are relaxed.

This procedure can be well automated and several software packages are available. The most prominent ones are *Frealign* [70] and *Relion* [181]. Relion, which will be used during this thesis, follows a maximum likelihood strategy combined with employing an empirical Bayesian approach.

### 2.4.13 Resolution Assessment

After successful refinement, a well-resolved structure should be achieved. To be able to interpret a structure accurately, it is important to know, what can be deduced from it and to what level of resolution. The resolution in general describes the minimal distance of two object details in the model, which can be discerned significantly from each other by the method of choice. The intuitive way to determine the resolution is to investigate directly the structure towards which chemical and structural features can be identified (cp. Figure 2.8).



**Figure 2.8: What Detail can be seen at which Resolution?** Here, the atomic model of a CRM1-Ran(GTP)-snurportin complex (pdb: 3gix) was filtered to different resolutions. At 20 Å only the overall shape is visible. Individual domains can be recognized starting from 12 Å. At around 9 Å long  $\alpha$ -helices become visible as sausage like structures while all of those should become clearly recognizable at 7 Å. At 5 Å  $\beta$ - sheets get separated. Starting from 4 Å bulky side chains are visible, while at 2 Å all side chains should be visible.

However, this is subjective and not very accurate and resolutions better than 10 Å need to be reached to do that. Alternatively, if an atomic model of at least parts of the structure is available, one could filter this to the estimated resolution and compare the result with the reconstructed volume. But this is barely possible since usually no atomic models are available. A better way is to derive the resolution mathematically without knowledge of the structure at all to be as unbiased as possible.

Whether two features can be distinguished in practice mainly depends on the SNR. The resolution could best be described by the SNR as a function of the spatial frequency. Mostly, this is realized by the Fourier ring correlation (FRC) [179] in the 2D case or the Fourier shell correlation (FSC) [80] in the 3D case. For this, the data set is split

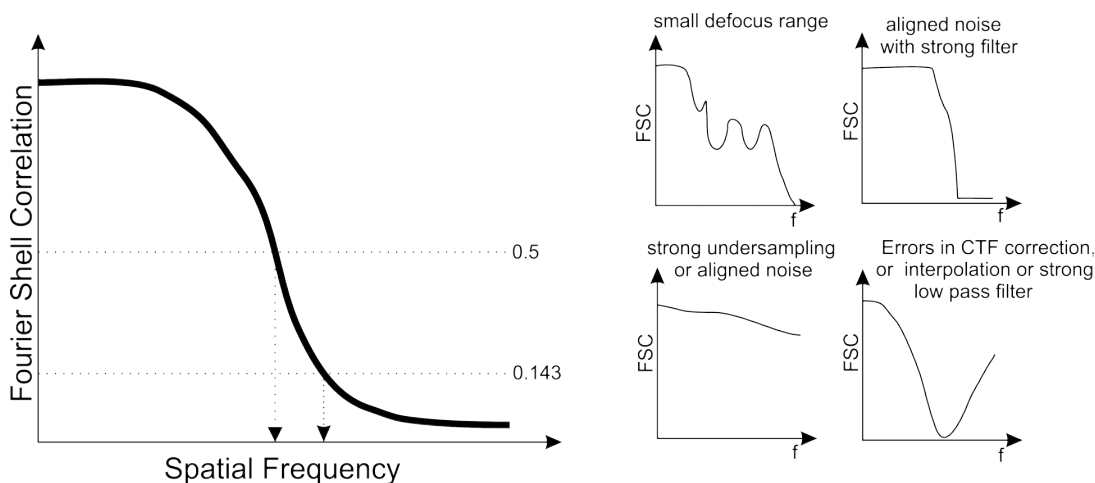
randomly in two equally large datasets which are refined independently. The two final reconstructions are compared in shells in Fourier space:

$$FSC(r_i) = \frac{\sum_{r_i \in r} F_1(r_i) \cdot \overline{F_2(r_i)}}{\sqrt{\sum_{r_i \in r} |F_1(r_i)|^2 \cdot \sum_{r_i \in r} |F_2(r_i)|^2}}, \quad (2.29)$$

where  $F_1(r_i)$  is the structure of the first volume, the radius  $r_i$  (corresponding to a certain spatial frequency) and  $\overline{F_2(r_i)}$  the complex conjugate at the same radius. At low spatial frequencies (inner shell radius) the correlation will be maximal, but at some point the signal will drop to almost no correlation. An ideal FSC and exemplified erroneous curves are shown in figure 2.9. The FSC correlates with the spectral Signal-to-Noise-Ratio (SSNR) following [150]:

$$SSNR(r_i) = \frac{FSC(r_i)}{1 - FSC(r_i)}. \quad (2.30)$$

The actual resolution can only be estimated by applying a threshold to those functions. A large number of different threshold criteria exists and all have a rational behind them [153]. Most commonly spatial frequencies at which the FSC drops below 0.5 or 0.143 are used as resolution measures [210]. A value of 0.5 would correspond to a SSNR of 1 and at a value of 0.143 the correlation between a experimental map and a perfect reference map correlates with the crystallographic figure-of-merit [169].



**Figure 2.9: Principal Fourier Shell Correlation Curves.** Left: An ideal Fourier shell correlation curve is shown. Commonly the resolution will be determined by the spatial frequency, at which the FSC drops below a value of 0.5 or a value of 0.143. Right: Typical erroneous FSC curves are shown. [153]

However, none of these measures was proven to perfectly correlate with the physics and chemistry of the molecule due to several problems. The most common problem arises from *overfitting* of the structure. For each image, five degrees of freedom have to be determined. Thus, the number of parameters to be determined exceeds the number of measured variables by far, which is called an *ill-posed problem* in mathematics. That is



why iterative methods need to be used, which in turn can lead to the accumulation of noise in the structure. This noise tends to mimic fine structural features, which can easily be mistaken as helices or even amino acid side chains. To prevent that, the *gold-standard refinement* was introduced, which recollects the original idea of splitting the data set into random halves right from the start and independently refine these [182]. Another way of evaluating the resolution estimate is to compare it to the actually achievable resolution. Physically, the resolution is limited by the wavelength of the electrons, the diffraction at the microscopic lenses and consequently their aberrations. The limit arising from the diffraction of the wave can be ignored as it is in the range of sub angstroms. However, the imperfections of magnetic lenses can give rise to a significant lowering of resolution. The microscope should always be well aligned to minimize any aberrations. Nevertheless, this becomes significant only in the very high resolution regime. The absolute mathematical resolution limit of the calculated 3D structure is derived from the sampling of the image. Following the Nyquist-Shannon sampling theorem, the maximum resolution of an image sampled with a pixel size of  $d$  Å will be  $2d$  Å. This can be intuitively understood through the discrete nature of the sampling. If two structural features need to be separated, they need to be apart by at least the distance of two pixels. The frequency  $f = \frac{1}{2d}$  is called the *Nyquist-frequency*. In a similar way, the angular sampling determines the final resolution. If one samples the angular space with an angular difference of  $\Delta\phi$ , then the maximal attainable resolution  $d$  is:

$$d = D \sin^2\left(\frac{\Delta\phi}{2}\right) = \frac{2D}{M}, \quad (2.31)$$

where  $D$  is the diameter of the projection image, and  $M$  the number of projections when evenly sampled with  $\Delta\phi$  as angular sampling step [80]. Interestingly, from this the number of necessary projections needed to achieve a certain resolution can be derived [84]. For instance to achieve a resolution of 3 Å at least 12600 perfect projection images are required (assuming a SNR of 0.05). Even though this is the lower estimate for the number of needed particles, the order of magnitude should be correct. While there are studies reaching that [11], many publications used millions of particles just to achieve sub-nanometer resolution (e.g. [207, 8]). If more images are used than necessary, obviously, there is more room for errors being introduced by a higher risk of overfitting.

Apart from the mentioned criteria, structural heterogeneity has a major impact on the achievable resolution. If a part of a structure moves relative to the rest, it will smear out and result in a lower resolution in that part and the resolution gets anisotropic. The FSC will always estimate the resolution over the full molecule. To estimate the local resolution better, a local FSC map can be calculated. Therefore, the FSC is calculated in several small boxes (windows) over the full volume and the resulting resolution estimate is assigned to the central voxel of each window [90]. Unfortunately the result of this is very much dependent on the chosen window size. An alternative was proposed recently:

a software called *ResMap* estimates the local resolution using local sinusoidal features [110].

## 2.4.14 Validation

Due to the large noise content, the whole workflow starting from the image acquisition to the final 3D model is extremely error prone. Validation at every step is thus more than needed. However, there is so far not much done towards that. To improve the situation, the EMDB formed a *Validation Task Force* [86]. Until now, the methods to validate EM structures are sparse. The only widely applied method was introduced 2011 by Richard Henderson. He proposed a tilt-validation strategy, where, like for RCT, pairs of tilted and untilted images are recorded. The untilted particles are treated as usual to produce a structure. The tilted images are aligned to the produced structure and a tilt angle based on the model is predicted. Finally, the predicted tilt angle is plotted and compared to the applied tilt angle. Validity would be reached if they are similar. However, as outlined before (cp. section 2.4.10.3), tilt methods are problematic and itself produce artefacts like stain rims or charging, which can lead to false negative results. Apart from that an omit-map strategy is sometimes used [52]. To validate a certain structure, a significant part of it is removed from the model and the data is refined against it. If the removed part emerges from the data again, it can be considered to be valid.

With only two methods at hand, the future has to bring more ideas to gain trustable results from single particle cryo EM.

## 2.4.15 Conformational Analysis

### 2.4.15.1 Detection of conformational heterogeneity

Naturally, before particle images can be sorted according to their conformations, the conformational heterogeneity has to be detected. First indication is given by a premature convergence of the refinement procedure at lower resolution than expected. Similarly, a local resolution map could be calculated to identify regions of low resolution which most probably correlate with regions of high flexibility [90, 109]. However, this is not a very quantitative measure. In statistics, the measure to quantify heterogeneity in the data is the variance. Similarly a 3D variance map will give insights into the heterogeneity. Unfortunately, the dataset obtained by cryo EM is a 2D dataset. To calculate a 3D variance one needs to split this data set into random parts, determine and refine 3D structures from the sub data sets and determine the variance from this. This approach requires a very large dataset and time intensive calculation. An alternative method uses

*bootstrapping*: many random subsets from the dataset are chosen. Hereby the same particle can be in different subsets. Normally, a couple of thousand subsets are chosen, a volume is reconstructed from every subset and the overall variance is calculated. By overlaying the variance map with the original map, regions of high variance become visible and strategies to cope with them can be developed. [155] This, however, only works when conformational heterogeneity is localized. If the overall shape of the structure changes, assignment of angles to different projections is not trivial and the variance would not depict a useful density.

#### 2.4.15.2 Analysis of large amplitude movements

Ideally, one would like to analyse the whole conformational space of a protein with only one data set. This, in turn, will require a computational separation of particles with different conformations. If the protein's shape changes dramatically, the standard EM methods will fail. Thus, particles with different conformations have to be sorted right from the start of the analysis. To date there are only a few methods available to overcome this challenge.

The most reliable among them is random conical tilt (RCT) [161] combined with 3D alignment and classification [175]. In this method, the noisy RCT volumes will be aligned to each other, classified and averaged. The advantage lies in the fact that no matter how different the shapes of the different conformations are, the method will be able to identify them. Only very featureless and globular structures will cause problems. This strategy was already successfully employed on very dynamic proteins like the spliceosome [65], the fatty acid synthetase [25], chromatin remodelling complexes [116] or the E3-ligase Itn1 [122] or heterogeneous sample like an assembling ribosome preparation [134].

The most common alternative to generate an initial model structure is called angular reconstitution based on common lines. In principle, it should be possible to sort the common lines into different subsets [89], which was already demonstrated with real data [58, 78]. However, common line approaches are complicated for conformationally homogeneous samples, due to the small amount of data the image is reduced to. Thus, no applications for heterogeneous samples were published yet.

Apart from that, a few other algorithms delivering promising results for heterogeneous datasets like *Simple Prime* [49] were published, but their validity was not proven so far.

Alternatively, if at least a stable ground state structure is available, one can use Normal Mode Analysis (NMA). This method is able to simulate possible conformational changes from a single starting structure. It is assumed that this starting structure represents the

energy minimum, and all movements are the combinations of simple harmonic oscillations of the atoms, which are coupled through their bonds to the neighbouring atoms [205]. Due to the given geometry certain collective movements arise from the coupling. This kind of calculations were used for crystallographic refinements [103] and fitting of atomic coordinates [201] for a long time. However, EM-maps lack atomic coordinates. Thus, a coarse-grained strategy employing pseudo atoms is used: the volume is approximated as a set of small spheres that are coupled with their neighbouring spheres [141]. Thus, from one energy minimized structure models with possible mechanic deformations can be predicted. It has to be kept in mind here that this is a purely topological or mechanical approach not taking any chemical or even physical forces like electrostatics, into account. However, refining the resulting structures against the whole data set yielded biological relevant results [26, 96]. Nevertheless, the final structures have to be validated with great care since alignment procedures are heavily biased towards any given reference [209].

Apart from that, if prior biochemical knowledge about the conformational dynamics is available one could use it and specifically model structural states as initial references. Again, model bias is a large concern, and validation can not be omitted. One could simply split the data set according to the modelled structures and refine the different subsets either against a model omitting the changed region or depicting another state. In both cases, the data set must reproduce the originally modelled volume to be considered valid. This strategy was successful for the description of the ratcheting movement of the ribosome [52].

### 2.4.15.3 Analysis of large domain movements

In many cases, large domains of a protein complex can move for several nanometers or even bind and unbind while the overall structure of the complex remains almost unchanged. A well studied example of this kind is the ratcheting of the small ribosomal subunit. Since the ribosome is one of the favourite objects of the EM community, many methods are developed and available to handle such problems. In most cases, references need to be produced in some way. One could use NMA or modelling. Additionally, bootstrapping could be used again. Resulting volumes could be directly used as references or could be classified to give more fitting references in 3D.

In a procedure called supervised classification a set of promising references is projected into 2D space and all images are aligned against all references. Depending on the best fitting reference image the particle images are assigned to a model. In this way, a competitive refinement can be performed, where in every refinement step all images are aligned against all available references. These methods gave greater insights into the ribosomes dynamics. To reduce reference bias during this procedure and find better convergence Maximum

Likelihood (ML) algorithms can be used [184]. Hereby, the assignment of a particle image to a certain reference is weighted with its probability. Even if the initial references are identical, the resulting 3D structures converge into different conformations.

An alternative approach uses principal component analysis (PCA) which is already often used in the regular single particle analysis, as previously discussed (see section 2.4.9.2). Ligand occupation and size differences can be detected in such a way [48, 144]. Even more complicated conformational changes for the symmetric chaperonine GroEL could be found [37]. However, this method depends on a lot of experience in interpreting the eigenimages, and it is not yet clear whether more complicated conformational changes are comprehensible in this way.

#### 2.4.15.4 Analysis of small amplitude movements

Once the overall conformational changes were sorted sufficiently, one has to take care of the remaining small local changes. This can be done by bootstrapping. However, this time the resulting volumes are classified with focus on the dynamic region. This means, the stable part of the structure is masked out during analysis, resulting in models showing different states of the flexible region while leaving the overall structure unchanged. These models again can be refined against the whole data set. Once more, the best example is the ribosome. The tRNA movement during the elongation cycle was analyzed in this way [52].

Finally, the best strategy towards the analysis of conformational heterogeneous complexes will need to employ a combination of the mentioned methods. This could be done in a hierarchical fashion, orienting on the hierarchical nature of the energy landscape (see figure 1.5). Tier-0 movements should be addressed firstly followed by the other tiers [52].

## 2.5 Sample Preparation for Electron Microscopy

### 2.5.1 GraFix

Samples were prepared for EM analysis using the GraFix methodology [100]. In short: Sucrose density gradients in the optimal protein buffer were prepared. Additionally, the sucrose gradient contained a glutaraldehyde gradient. Mostly, gradients contained 0–0.1 % (v/v) glutaraldehyde. The sample was loaded on the gradient, and the centrifugation conditions were chosen to have the molecule of interest sediment to 2/3 of the gradient. Conditions for the used samples are outlined in table 2.6.

**Table 2.6:** GraFix conditions for the analyzed samples.

GraFix conditions mentioned	are for a	TH660 rotor	Speed [rpm]	Time [h]	Temperature [°C]	Speeds (Thermo Scientific).
Sample	gradient [% (w/v) sucrose]					
CtCRM1	5 – 25		45000	16	4	
HsCRM1	5 – 25		45000	16	-10	
EcPDHc	10 – 30		21000	16	4	
BgAChBP	10 – 30		30000	16	4	

The gradients were fractionated in 200  $\mu$ L steps. The fractions were either analysed by SDS-PAGE or dotblots. For dotblotting, 1  $\mu$ L samples of each fraction were added on a nitrocellulose membrane (Whatman) and subsequently stained with amido black [81]. Peak fractions were analyzed by EM.

### 2.5.2 Negative Staining Preparation

An approximately 2% (w/v) solution of uranyl formate was freshly prepared in double distilled water. The solution was vigorously mixed and the insoluble precipitate was pelleted via centrifugation in a table top centrifuge at 13000 rpm for 10–15 min. A piece of an in-house prepared carbon film was floated on top of the protein solution for mostly 1 min. The foil was absorbed to a holey carrier carbon film on the surface of a copper grid. Excess liquid was removed from the grid via a filter paper and the particle-adsorbed film was brought onto the stain solution for approximately 1 min. Again, the excess liquid was removed via a filter paper and the grid was allowed to air dry at room temperature.

### 2.5.3 Vitrification

For cryo grid preparation, sucrose is removed from the GraFix fractions via a Zeba spin desalting column. As for negative staining, a continuous thin carbon film was freshly floated onto the sample solution for 1 minute. The foil was adsorbed to a quantifoil grid containing a thick carbon film with a regular whole pattern. 4  $\mu\text{L}$  of water are applied to the surface and the excess liquid was blotted in a Vitrobot apparatus at 100 % humidity and 4 °C. Subsequently, the grid was plunge shot into a liquid ethan container and thus the remaining thin water film with particles in it was vitrified. Grids were stored in liquid nitrogen before imaging.

## 2.6 Imaging

Negative stain grids were routinely analyzed in a CM200 FEG microscope. Images were taken at defocus values of approximately 1  $\mu\text{m}$  on a twice binned CCD camera. Regularly, magnifications of 88000x or 155000x were used, which correspond to a pixel size of 2.45  $\text{\AA}/\text{px}$  respectively 1.85  $\text{\AA}/\text{px}$ .

Cryo grids were analyzed on a Titan Krios TEM equipped with a  $C_s$ -corrector and a Falcon direct electron detector. Before usage, the alignment of the microscope was adjusted and image distortions due to astigmatism, coma or spherical aberration were minimized by alignment of the  $C_s$ -corrector. Images were recorded at a dosage of 25 electrons per  $\text{\AA}^2$  with pixel sizes of 2  $\text{\AA}$  respectively 1.25  $\text{\AA}$ . For RCT, images were taken with 0° and 45° tilt with the same electron dose.

## 2.7 Image Processing

### 2.7.1 Micrograph Analysis and Particle Picking

Prior to particle identification, all micrographs were visually inspected. Micrographs, showing a vast contamination, devitrification, charging or drift were discarded right away. Particles were picked with the custom written software *John Henry* [28]. An approach based on image statistics was used to initially identify particles. The resulting coordinates were refined using local alignments of the picked images on a coarsely projected model. Individual particles were mostly cropped with a window roughly two times the diameter of the particle. Images for high resolution datasets were cropped in windows three to

four times the diameter of the particle to be able to restore the delocated high resolution information in the CTF-correction.

### 2.7.2 CTF-Correction

To achieve best possible image restoration, the phase flipping method described by Sander et al. [173] was used. In brief: The particles were cropped in windows 2-4 times the diameter of the particle. Power spectra were calculated and filtered to optimally visualise the Thon-rings. Multivariate statistical analysis, consisting of PCA and classification, was performed on the power spectra. The PhCTF parameters for each class average were estimated through a curve fitting algorithm and theoretical power spectra were calculated. These were compared to class averages by eye and class averages not corresponding to the theoretical power spectra were excluded from further analysis. The remaining images were individually phase flipped and the window size was removed to 1.5 times the particle size.

### 2.7.3 2D-Processing

For 2D analysis, particles were down sampled to a pixel size of 4-8 Å/px depending on the initial size as well as filtered and normalized. Reference free class averages were calculated [47] and iteratively refined. Additionally, class averages using a 2D maximum likelihood approach [183] were calculated using Relion [181]. Images giving rise to class averages depicting artefacts like carbon edges, stain artefacts or ethane/ice crystals were removed from further analysis.

### 2.7.4 Initial Model Building

Several techniques were used to produce initial models. First of all, angular reconstitution was performed for all three model systems. To gain starting models for different conformations, this was combined with a bootstrapping strategy: for a few thousand times, a set of 10-20 class averages was drawn from the complete stack and a model was calculated via angular reconstitution. The models were analyzed visually and consistent models were used for the further analysis. In an independent approach, class averages were subjected to Simple PRIME [49]. This produces 3D models based on an iterative projection matching approach, without an initial model. Lastly, for the PDHc, RCT was performed [175].



### 2.7.5 Conformational Sorting

Additional, conformational sorting was performed. Mostly, bootstrapping in combination with 3D-MSA was performed [52]: several thousand times, a subset of 10–20 images was randomly drawn from an aligned stack of images. The angular assignment was taken from the alignment and 3D-models were calculated. These volumes were classified using 3D-MSA, and 3D-class averages were inspected visually and sorted for their different shapes. Those models were used for competitive refinements termed *supervised classification* [60]. Alternatively 3D-maximum likelihood classification [184] was performed in Relion [181]. This *unsupervised refinement* produced a set of seeds and by distributing the particles images randomly to a set of equal references and performing several rounds of competitive refinements. Also Normal Mode Analysis (NMA) was performed to identify new conformations. For EM-maps the 3DEM Loupe algorithm [141] in XMIPP [40] and for atomic coordinate models the ElNemo server [200] was used. Resulting models were analyzed with supervised classification.

### 2.7.6 Refinement and Resolution Determination

Final structures were mostly refined using Relion [181]. CRM1, however, was refined in Coweyes using a corrim-based refinement strategy [173]. The resolution was estimated using FSC [80, 210] by the gold-standard approach [182].

### 2.7.7 Structural assessment

Structures are visualized in UCSF Chimera [159]. If available, atomic models were fitted in this software.

To validate a structural model, a cross reference test was used. If a dataset yielded models in two different conformations, the particles were split into corresponding stacks. The stacks were consequently refined on the wrong model. If the final reconstruction relaxed back to the previously gained model, the sorting was considered valid.



# Chapter 3

## Results

### 3.1 Method Development

In this thesis, the dynamics of large proteins and protein complexes are analyzed using three examples. To facilitate the analysis a number of technical advances had to be made. Prerequisite for the analysis of dynamics is to assure that the protein of interest is intact, folded and assembled. In the first part of this thesis, a high throughput screen for stabilizing conditions was developed. Further, different cross linking strategies based on the GraFix methodology were compared towards their stabilizing effects on the proteins structure. Lastly, the influence of temperature is analyzed by either combining GraFix with low temperatures or using proteins from a thermophile organism.

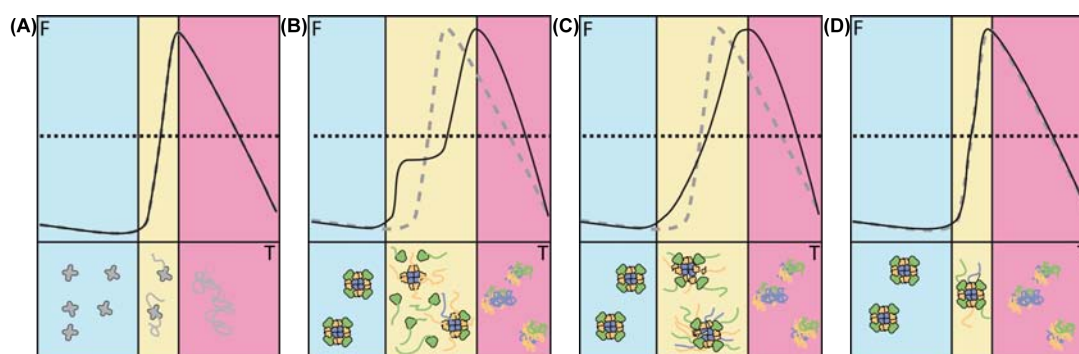
#### 3.1.1 ProteoPlex

One of the major challenges to successfully work with protein complexes is their stability. Most of them are composed of more than ten individual biomolecules [62] which only bind with rather low energy to each other. Therefore, in the isolated state they readily lose their structural integrity, decompose or even aggregate. This will lead to samples containing many different compositions of the complex, which becomes a new degree of freedom. This complicates any structural analysis. To prevent that, a new method termed *ProteoPlex* was developed in this thesis.

The basic concept is that complexes are stable for many hours or even days, weeks and years in their natural environment - the context of a cell. Mimicking this environment as good as possible will improve the structural integrity of any given biomolecule. However, cells are not just "teabags" randomly filled with biomolecules. Cells provide a number of different environments, which are far from homogeneity. Therefore, it is absolutely impossible to find a generic *in-vitro* condition that is suitable for any given complex. Thus, one needs to find a suitable condition for a given complex by screening many different conditions. The classical way would be to purify the complex in the to-be-tested buffer and analyze its stability by density-gradient centrifugation followed by SDS-PAGE

analysis of the individual fractions. The high amount of sample needed and hands-on-time limits the number of testable conditions to only a few.

Therefore, a method is needed which is very sensitive and quick. Nevertheless, it should not need special equipment or expensive chemicals. Luckily, such a method termed *Thermofluor* was described recently [146]. In this method protein samples in a screening buffer are continuously heated in presence of a environment sensitive (*solvatochromic*) dye. Upon unfolding of the protein, hydrophobic side chains are exposed and the dye gets hyperfluorescent. With increasing concentration of unfolded protein in the solution, it starts to aggregate and the hydrophobic patches are hidden again leading to a loss of fluorescence signal. From this experiment one can determine the melting temperature of the protein within this buffer composition, which is determined as the inflection point of transition phase. This is the temperature at which 50 % of the proteins are unfolded. The higher the needed temperature, the more stable the protein. One can now easily screen hundreds of conditions in parallel in a second generation RT-PCR machine. While this approach has yielded good results for single chain mono-domain proteins, it was disregarded for protein complexes in the literature [140]. However, I challenged that idea and tried to find reasons, why results from the *Thermofluor* do not give sufficient information to screen for protein complex stability.



**Figure 3.1:** Schematic representation of possible unfolding scenarios. Different scenarios depicting typical fluorescence-temperature curves are shown. Each curve can be subdivided into the native phase (blue), the unfolding phase (yellow) and the aggregation phase (pink). (A) Unfolding of a single domain protein. During the unfolding phase, the protein rapidly unfolds leading to a fast transition from folded to unfolded state and thus a rapid increase in fluorescence signal. This resembles the perfect curve and is depicted as dashed grey line in every of the other scenarios. (B) If a complex is very fragile, it disassembles before it unfolds and its different components give rise to step pattern in the transition. (C) If the component's melting point is rather similar, the step pattern fuses to a less steep transition. (D) At full cooperativity, the complex's unfolding curve resembles the single domain unfolding curve.

To the current understanding: single proteins yield unfolding transitions reflecting a two-state behavior (cp figure 3.1A), it is assumed, that complexes would exhibit multiphasic unfolding and thus uninterpretable curves (cp figure 3.1B and C). These unfolding transitions would reflect the sum of the unfolding of all individual subunits of a given macromolecular complex and therefore not contain any information about the stability of the

entire complex. This is indeed true for complexes under destabilizing buffer conditions. However, as outlined in section 3.1.1.4, it could be found for over 80 complexes that even large macromolecular assemblies can behave as apparent two-state folders under optimal buffer conditions. Thus, the unfolding curves of large complexes under stabilizing buffer conditions are not simply the sum of unfolding transitions of each individual subunit. In contrast, the dissociation of macromolecular complexes into subunits and their unfolding appears to occur in a near concerted manner in a stable buffer environment (cp figure 3.1). The ProteoPlex transitions of such stabilized complexes show near two-state unfolding characteristics very similar to those of single-domain proteins, which make this method suitable for the determination of stabilizing conditions for any given protein complex.

In a first step, a theoretical framework for the analysis of curves was established and experimentally challenged as described in the following sections.

### 3.1.1.1 Theoretical Description of Thermofluor Curves Obtained from Single Domain Proteins

The theory behind unfolding experiments was well-established by the protein folding community. The most simple model system for the thermal unfolding is a single domain protein. Assuming a thermal unfolding shifts the following equilibrium:



where N represents the native folded state and U the unfolded state of the protein. This can be further described by:

$$K = \frac{[U]}{[N]} = \frac{f_U}{f_N}, \quad (3.2)$$

where K is the equilibrium constant, while  $f_N$  resp.  $f_U$  are the fractions of the two states with  $f_N + f_U = 1$ .

The measured signal in a Thermofluor assay is the hyperfluorescence of dye - unfolded-protein complexes. With an excess of dye, one can assume that all unfolded protein is bound, and the signal is directly proportional to the concentration of the unfolded state. The measured fluorescence signal F is put together as follows:

$$F = f_U \cdot F_U + f_N \cdot F_N + F_0 \quad (3.3)$$

The autofluorescence  $F_0$  of the unbound dye can be neglected in the further course of the derivation since it will be corrected for that with control experiments. The fractions of the two protein states can than described with:

$$f_U = \frac{F_N - F}{F_N - F_U} \quad \text{resp.} \quad f_N = \frac{F - F_U}{F_N - F_U} \quad (3.4)$$

With equation 3.2 it follows:

$$F = \frac{F_N + F_U \cdot K}{1 + K} \quad (3.5)$$

The experimental set-up dictates a temperature dependence which can be included via van't Hoff's equation:

$$F = \frac{F_N + F_U \cdot e^{-\frac{\Delta G}{RT}}}{1 + e^{-\frac{\Delta G}{RT}}}, \quad (3.6)$$

with  $\Delta G$  as free enthalpy of the reaction, R the universal gas constant and T the applied temperature in K.

Since the unfolding is temperature induced, one has to consider the temperature dependence of  $\Delta G$  which is described with the Gibbs-Helmholtz equation:

$$\Delta G = \Delta H - T\Delta S \quad (3.7)$$

The melting point  $T_m$  of the domain is defined as the temperature, where the fractions of U and N are equal. Therefore,  $K = 1$  is concluding in  $\Delta G = 0 \frac{\text{kJ}}{\text{mol}}$ . It follows directly:

$$\Delta S = \frac{\Delta H}{T_m} \quad (3.8)$$

In good approximation, one can neglect the temperature dependence of reaction entropy and enthalpy in the considered range concluding in:

$$\Delta G = \Delta H - \frac{T\Delta H}{T_m} = \Delta H \frac{T_m - T}{T_m} \quad (3.9)$$

With equation 3.6 it follows:

$$F = \frac{F_N + F_U \cdot e^{-\frac{\Delta H}{RT} \frac{T_m - T}{T_m}}}{1 + e^{-\frac{\Delta H}{RT} \frac{T_m - T}{T_m}}} \quad (3.10)$$

Furthermore, one has to consider that  $F_N$  and  $F_U$  are both temperature dependent. The temperature dependence of  $F_N$  can be described mostly in a linear fashion, with the slope of  $m_N$ . Only small interactions with the natively folded complex and the dye induce this behavior

$$F_N(T) = F_{N,0} + m_N \cdot T \quad (3.11)$$

But in case detergent is added, vast effects can be seen since the dye - itself being an amphiphilic substance - can heavily interact with the detergent. This leads to large background signals, that cannot be easily removed from the data. Therefore, detergents cannot be screened and neither can membrane protein in this approach.

$F_U$  will decrease with growing temperature due to entropic effects. The unfolded protein will appear at higher concentrations at higher temperatures. This results in aggregation proteins exposed hydrophobic patches and thus release of dye, since this will increase the entropy of the dye dramatically. In the *Thermofluor* assay this gets complicated, because in this case time and temperature dependent effects of aggregation overlap. At a given temperature, the protein will aggregate with a certain rate. This can be simply modelled by an exponential equation since this is a first order reaction. One can assume that dye release and aggregation depend linearly on each other so that the decrease of signal can be described at a given temperature as a single exponential. But at the same time temperature changes, so the rate changes also exponentially. The resulting model has to be a hyper exponential model. But this is only applicable when the temperature would change linearly in time but the Thermofluor experiment changes temperature in a stepwise fashion in order to allow adjustment of the unfolding equilibrium. Since the real model would introduce too many parameters to fit, I model it as a single exponential with an average aggregation rate  $m_U$  and the parameters  $a_{U,0}$  and  $b_U$ :

$$F_U(T) = a_{U,0} \cdot e^{m_U \cdot T} + b_U \quad (3.12)$$

This finally concludes in:

$$F = \frac{F_{N,0} + m_N \cdot T + (a_{U,0} \cdot e^{m_U \cdot T} + b_U) \cdot e^{\frac{-\Delta H}{RT} \frac{T_m - T}{T_m}}}{1 + e^{\frac{-\Delta H}{RT} \frac{T_m - T}{T_m}}} \quad (3.13)$$

The obtained Thermofluor-Data can be fully fitted and described with this model. The aim of the procedure is still to find a valid measure for protein stability at room temperature which is  $\Delta G$  extrapolated to room temperature. This can be done as follows:

$$\frac{d\Delta H}{dT} = T \cdot \frac{d\Delta S}{dT} = C_P(U) - C_P(N) = \Delta C_P, \quad (3.14)$$

where  $\Delta C_P$  is the difference of the heat capacity of folded and unfolded domain usually assumed to be  $5.150 \frac{kJ}{mol \cdot K}$ . It follows:

$$\Delta H(T) = \Delta H(T_m) + \Delta C_P \cdot (T - T_m) \quad (3.15)$$

and:

$$\Delta S(T) = \Delta S(T_m) + \Delta C_P \cdot (\ln(T) - \ln(T_m)) \quad (3.16)$$

Finally, this results combined with the Gibbs-Helmholtz-equation in:

$$\Delta G(T) = \Delta H \frac{T_m - T}{T_m} - \Delta C_P [T_m - T + T \cdot \ln\left(\frac{T}{T_m}\right)] \quad (3.17)$$

This all neglects a major problem in the Thermofluor assay which is the temperature dependence of buffers. This simply depends on the hydration enthalpy change  $\Delta H_{hyd}$  with:

$$\frac{d \ln K}{dT} = \frac{\Delta H_{hyd}}{R \cdot T^2} \quad (3.18)$$

Therefore, the obtained  $\Delta H$  has to be corrected for its value at the melting point as well as one has to keep in mind that the measured stability is measured for the melting point pH and not for the initial pH.

### 3.1.1.2 Theoretical Unfolding Behavior of a Protein Complex

Protein complexes are built of several chains of individually folded proteins. These chains bind to each other to form the complex. Thus, the overall signal measured in a Thermofluor experiment of such a complex can be described as:

$$F = \sum_{i=1}^n (f_{U,i} \cdot F_{U,i} + f_{N,i} \cdot F_{N,i}) + F_0, \quad (3.19)$$

where  $n$  is the number of individual components, meaning the single proteins as well as subcomplexes.

The number of possible dependencies of the unfolding behaviour of individual components on each other is high and difficult to predict. However, simple expressions for the extreme behaviours can be found. An unstable complex would disassemble very rapidly or is already disassembled. Thus, its individual components unfold independently. In this case the measured unfolding curve is simply the superimposition of the individual unfolding curves following equation 3.13:

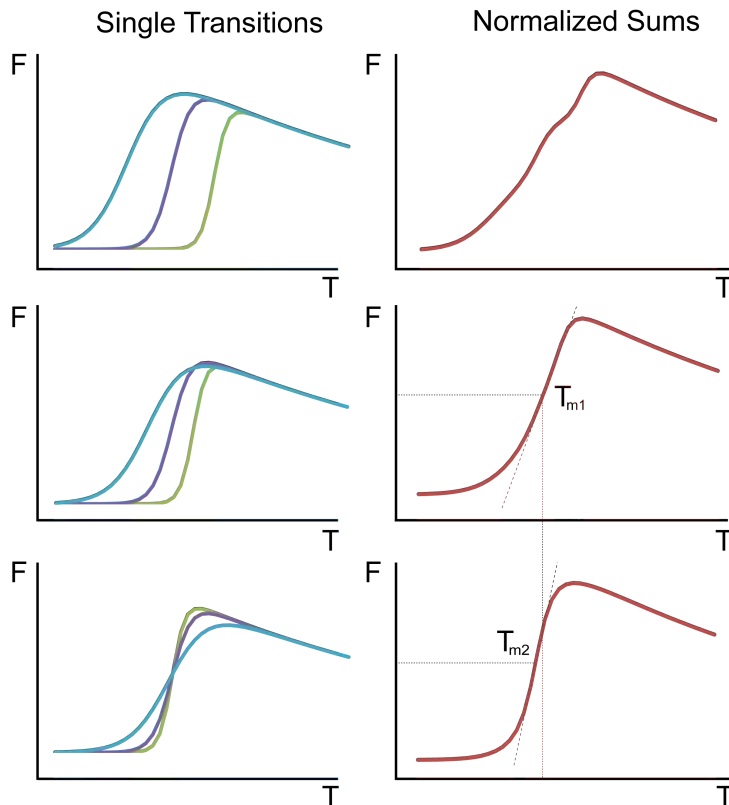


$$F = \sum_{i=1}^n \frac{F_{N,i} + m_{N,i} \cdot T + (a_{U,i} \cdot e^{m_{U,i} \cdot T}) + b_{U,i} \cdot e^{-\frac{\Delta H}{RT} \frac{T_{m,i}-T}{T_{m,i}}}}{1 + e^{-\frac{\Delta H}{RT} \frac{T_{m,i}-T}{T_{m,i}}}} \quad (3.20)$$

This behavior is illustrated in figure 3.2 (upper panel). The opposite extreme case would be the assumption of full cooperativity, meaning the disassembly and the unfolding of the complex occur concertedly. This seems unlikely, however, the physical forces driving the formation of a complex and keeping it together are the same as the ones that drive protein folding. Thus, one can describe the disassembly of the complex via the portion of disassembled complex ( $f_{dis}$ ) in analogy to equation 3.5 as:

$$f_{dis} = \frac{K_{dis}}{1 + K_{dis}} \quad \text{with} \quad K_{dis} = e^{-\frac{\Delta H_{dis}(T-T_{m,dis})}{RT_{m,dis}}}, \quad (3.21)$$

where  $\Delta H_{dis}$  is the dissociation enthalpy and  $T_{m,dis}$  the melting temperature of the complex, describing the disassembly entropy. In a Thermofluor experiment  $f_{dis}$  cannot be measured directly. However, assuming full cooperativity, the unfolding happens concertedly, meaning  $f_{dis}$  is directly proportional to the measured fluorescence signal. Thus, the signal can be described with equation 3.13, which is actually only a simplification of equation 3.20 with  $n = 1$ . This scenario is depicted in figure 3.2 (bottom panel). This is a simple two-state folding model as described in the previous section.

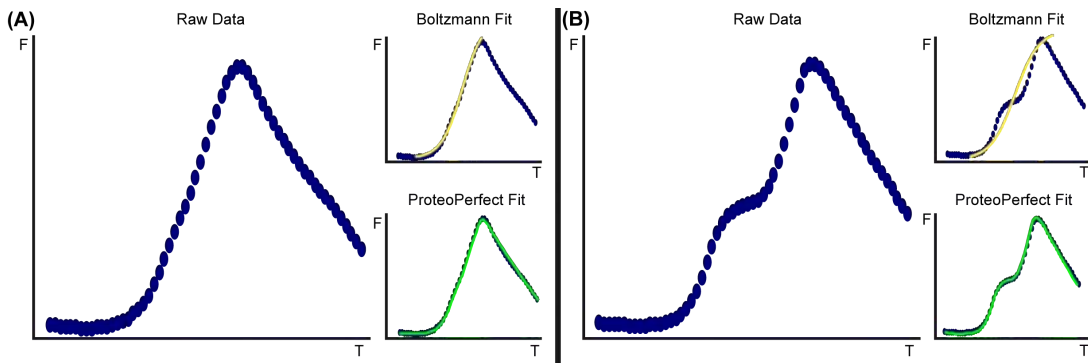


**Figure 3.2: Temperature-Fluorescence unfolding curves of a three-subunit complex.** Left: The single-transitions of the individual components are depicted. Right: ProteoPlex measurements depict a normalized sum of three individual curves. Upper panel: The three components unfold independently from each other at different temperatures. The sum is a curve showing several independent transitions. Middle Panel: The components unfold at similar temperatures but still independent from each other. The resulting sum resembles a two-state unfolding curve. Lower Panel: The components unfold cooperatively at the same temperature. This also results in a two-state curve, which is steeper than in the middle case. However, the melting temperature of the shallower uncooperative unfolding transitions results in an apparently higher melting point  $T_{m1}$ , than the cooperative transition ( $T_{m2}$ ) as visualized by the dotted lines.

Everything between these two extreme cases can be arbitrarily complicated since several different equilibria can be involved. This intermediate scenario describes a slowly disassembling complex, which exposes some individual unfolding components. This is illustrated in 3.2 (middle panel). This general case can be described as follows: the overall signal appears as the sum of the signals from all  $n$  individual components (single proteins and subcomplexes). Each component arises from an intricate equilibrium described with the constant  $K_{dis,i}$ :

$$F(T) = F = \sum_{i=1}^n \frac{F_{N,i} + m_{N,i} \cdot T + (a_{U,i} \cdot e^{m_{U,i} \cdot T}) + b_{U,i} \cdot e^{-\frac{\Delta H}{RT} \frac{T_{m,i}-T}{T_{m,i}}}}{1 + e^{-\frac{\Delta H}{RT} \frac{T_{m,i}-T}{T_{m,i}}}} \cdot \frac{K_{dis,i}(T)}{1 + K_{dis,i}(T)}. \quad (3.22)$$

The other two scenarios arise from this model. When all  $K_{dis,i}$  are large, meaning the dissociation is favored, the equation can be simplified to equation 3.20. If a high level of cooperativity exists all  $K_{dis,i}$  are equal simplifying the equation to an equivalent of 3.13. One could even model the dependence of the different equilibria on each other. However, already with equation 3.22 the number of unknown parameters exceeds the number of measured data points. Thus, overfitting of such a model becomes a serious problem and there is no straight forward way to distinguish those models. This in turn is absolutely necessary, since a complex composed of components with similar melting temperatures following equation 3.22 can hardly be distinguished from a two-state folding model and thus obtained parameters can be estimated wrongly.



**Figure 3.3: Evaluation of ProteoPlex fits.** Two exemplified data curves are shown as blue dots. Commonly only the transition part of the data is fitted by a Boltzmann model as shown in yellow [140]. The best data approximation from ProteoPlex is shown in green. While Boltzmann data approximation still gives acceptable results for a near two-state unfolding behavior (left), multiple transitions cannot be approximated by a simple Boltzmann model (right). In contrast, ProteoPlex still describes the obtained curve well. Of note: The ProteoPlex model is able to fit the whole curve and thus will obtain more accurate parameters.

Fitting the full model (equation 3.22) is not helpful due to overfitting issues. So the data can only be approximated with the two extreme cases and at most 6 components ( $n = 6$ ) can

be considered. In practice, every dataset was fitted with equation 3.20 with  $n = 1$  to 6. Testing a number of cases reveals that this is sufficient and even superior to a simple Boltzmann approximation as it was used before.

### 3.1.1.3 An Heuristic Approach used for Curve Interpretation

The main reason why Thermofluor is disregarded for protein complexes is the experience that the obtained parameters do not match with results obtained by other methods. This is mostly due to the analysis strategy. Mostly, the transition only is approximated to estimate the melting point  $T_m$  of the whole curve, which is the inflection point of the transition. The higher this melting point is, the higher is the unfolding entropy and thus the protein stability. However, applying this strategy to a complex can go terribly wrong. In a total uncooperative case as depicted in 3.2 (upper panel), several of such melting points exist and the analysis fails. However, for the other two cases (middle and lower panel), the melting point analysis can be formally performed and can result in misleading results. As shown, the apparent melting point of the less stable complex can be higher than the cooperatively unfolding complex. The reason for this is, that the superimposed curve of the less stable protein (middle panel) is less steep and the inflection point is reached later. However, this value is meaningless since only for the fully cooperative case the estimated parameters have a real physical meaning. The task for a valid analysis is therefore to distinguish these two scenarios. This is not directly possible only through curve fitting. The more complex model (equation 3.22) will always approximate the data better than the less complicated model (equation 3.13). A Thermofluor screen, however, consists of many measurements of the same complex in different chemical environments. From this whole data set, one can make deductions. As seen before, the slope is a good parameter to distinguish the two models. The higher the slope, the more probable the cooperative model gets. Physically in the cooperative model the slope is proportional to the dissociation enthalpy. Theoretically, following Hess's law, the unfolding enthalpy should be equal for each condition, assuming no direct interaction with the surrounding medium. From this, one can build an heuristic approach for the analysis of the curves:

1. The curves have to be tested if they are valid for analysis. Curves have to at least show one valid transition and no more than six local maxima should be detectable. This excludes curves with low SNR or curves obtained from wrongly pipetted experiments. Further, the controls containing no protein have to be valid, meaning here no detectable transition should be found.
2. The reference curves obtained from a buffer dye mixture are averaged and subtracted from all curves. The curves are normalized, where the lowest local minimum

before the transition is set to 0 and the absolute maximum is set to 1000 arbitrary units.

3. A number of artifacts are described in the literature [20]. Among them are air bubbles and aggregates leading to a high negative slope at the beginning of the curve. Also detergent can have such an effect. If the initial slope exceeds a certain threshold, the curve is discarded from the analysis.
4. Fits are found using the approximation equation 3.20 with  $n = 1$  to 6 on every curve.
5. For all fits, the coefficient of determination  $R^2$  is calculated. Only if this value is below, a certain threshold the next best fitting model is chosen for further analysis.
6. The melting temperature  $T_m$  and the enthalpy  $\Delta H$  are extracted from the chosen fit. Fits showing  $\Delta H$  values in the order of magnitude of the highest measured in the whole dataset are used for further analysis.
7. Finally, remaining curves are sorted as before by  $T_m$ .

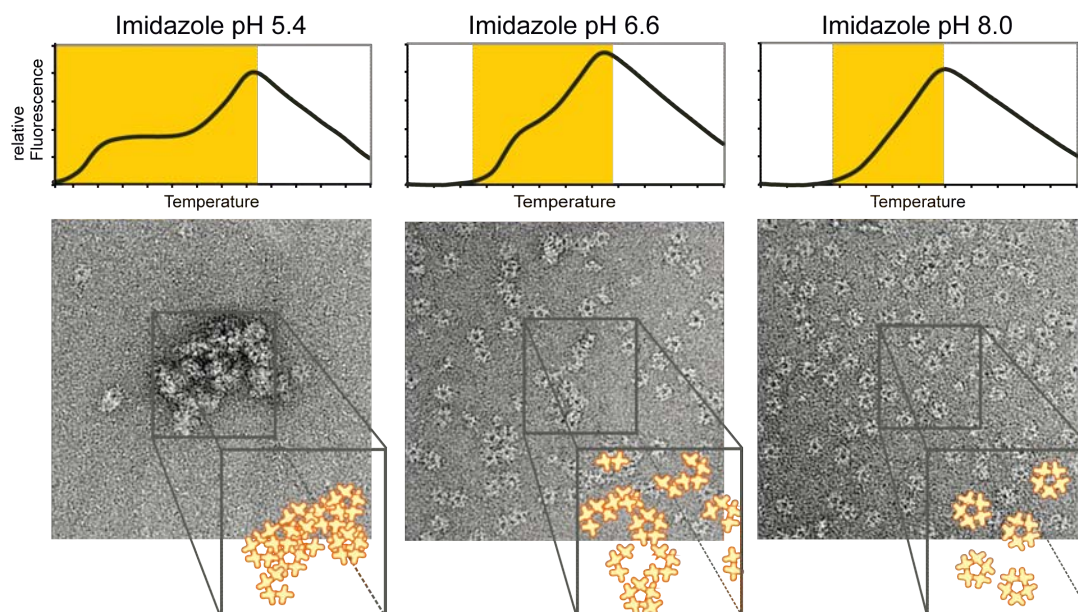
Further parameters like the area below the curve can be added to this heuristic approach. And even though this seems to be a quantitative analysis, it should be noted that it provides only qualitative results. The values obtained have high errors and should not be used for any biophysical interpretation apart from the stated stability.

The complete algorithm was implemented into a stand-alone software package by Jan-Martin Kirves [104].

#### 3.1.1.4 Proof of Principle

To test the developed analysis method, proof of principle experiments were performed together with Dr. Ashwin Chari, who also initiated the ProteoPlex approach. A 96-well buffer screen of several test complexes using the usual Thermofluor protocol was performed. The data was analyzed with the described ProteoPlex approach. In parallel, the complexes were brought to the tested buffer systems and electron micrographs were recorded to estimate the homogeneity and dispersity of the sample. As a first test system, the selenocysteine-tRNA-synthetase *SelA*, a homo-decameric protein with a weight of 500 kDa, provided by Dr. Oleg Ganychkin and Professor Markus Wahl (FU Berlin) was used.

Figure 3.4 illustrates these experiments. In conditions, leading to polyphasic transitions in the ProteoPlex experiment, also large aggregates could be seen in the micrographs (left panel). Further, if the transition got less polyphasic also the visible aggregation decreased



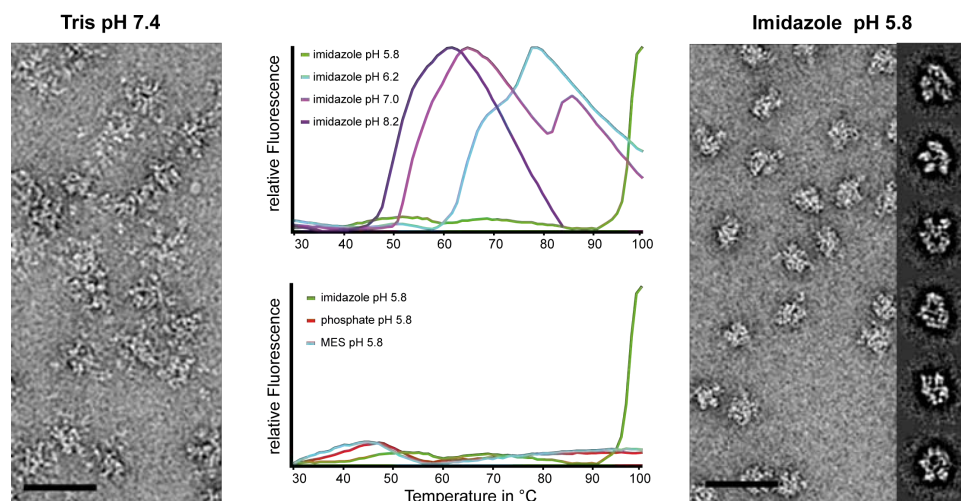
**Figure 3.4: ProteoPlex Proof of Principle SelA.** The nature of unfolding transitions is indicative of the mono-dispersity of macromolecular complexes. Shown are unfolding transitions of tRNA synthetase SelA at different buffer conditions (top panel) and corresponding EM micrographs at the same conditions (bottom panel). Only single unfolding transitions reveal a mono-disperse field of particles by EM (e.g. Imidazole pH 8.0).

(middle panel). And finally a curve fully compatible with the two-state unfolding model (right panel, cp. equation 3.13) resulted in a monodisperse particle distribution.

Those experiments were repeated with several complexes of known structure and only in conditions, where a real two-state behavior could be assumed the complexes showed a monodisperse distribution on the grid.

As final proof of principle, the *Biomphalaria glabrata* hemoglobin (BgHb) was used. This is a 1.5 MDa extracellular homooligomeric protein of unknown structure. The protein was kindly provided by Vanessa Moeller and Professor Jürgen Markl (University of Mainz). BgHb was purified by anion exchange chromatography as previously described [119]. Electron micrographs from the original revealed a poly-disperse sample with various degrees of dissociated particles and aggregates. In a ProteoPlex assay, a gradual stabilization of BgHb in Imidazole buffer from alkaline to acidic pH values was observed (Figure 3.5). In fact, the thermal stability of BgHb was enhanced by a total of 45 K. Therefore, I rebuffed BgHb into Imidazole at pH 5.8 and re-evaluated its aggregation state by EM. In these conditions, EM micrographs with a mono-disperse field of particles (Figure 3.5, middle panel) were obtained, which are clearly amenable to structural biology. Notably, this entire procedure from the sample to EM images of intact molecules was accomplished within 24 hours.

Taken together, these experiments indicate that the optimization of macromolecular complex stability is achieved without any prior biochemical information about the sample



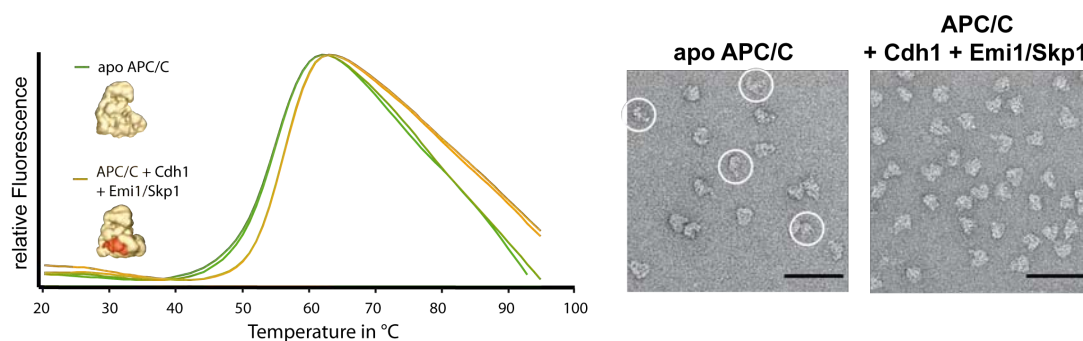
**Figure 3.5: ProteoPlex Proof of Principle BgHb.** Analysis of *Biomphalaria glabrata* hemoglobin complex (BgHb, 1.5 MDa native molecular weight) - a protein of unknown structure. Under standard purification conditions BgHb is mostly present as aggregated particles in negatively stained EM images (left panel, scale bar = 50 nm). The upper pane of the middle panel depicts a subset of unfolding transitions from a ProteoPlex pH screen. The stability of BgHb is gradually increased from alkaline to acidic imidazole buffer conditions, with a final enhanced stabilization of 45 K at pH 5.8 compared to pH 8.2. The lower pane of the middle panel reveals that interpretable unfolding transitions of the complex only occurs in imidazole buffer, which suggests the role of imidazole additionally as a stabilizing ligand. In imidazole pH 5.8, a monodisperse field of view with compact particles is seen (right panel, scale bar 50 nm).

under study in a completely automated workflow. Through many collaborations 86 different complexes could be subjected to the ProteoPlex methodology. Strikingly, all of those complexes showed at least one condition in which an apparent two-state folding model was evident.

### 3.1.1.5 Binding of Stabilizing Ligands

When one compares unfolding transitions of BgHb in other buffer systems and the optimal pH 5.8 derived from ProteoPlex measurements, it can be noticed that only Imidazole displayed this beneficial stabilizing effect (Figure 3.5, right panel). In an attempt to explain the selective stabilization effect of Imidazole, I recognized that all globins contain proximal heme-stabilizing histidines that are absent in BgHb. Thus, in addition to its function as a pH buffer molecule, the dramatic stabilization of BgHb structure emerging from imidazole is likely to result from its effect as a heme stabilizing ligand. Similar drastic effects could be shown for other complexes. For instance the addition of the co-factor thiamin diphosphate (TPP) to an *E.coli* Pyruvate Dehydrogenase complex preparation resulted in an increase in melting temperature of more than 20 K and the addition of the unhydrolyzable ATP-analogue ATP $\gamma$ S to the ATPase cohesin leads to an increase in slope and melting temperature of the transition. Thus, the stabilizing interaction of small molecules could be verified. This raised the question whether the addition of proteinaceous

interactors could also be found. Therefore, the Anaphase promoting complex /cyclosome (APC/C) was used as model system. Recently, the protein EMI1 was found to exert an inhibitory activity on the complex upon binding at the center of it [57].



**Figure 3.6: ProteoPlex for identifying stabilizing ligands. Identification of proteinaceous stabilizing ligands.** Unfolding transitions are shown (in duplicates) of the human Anaphase Promoting Complex (APC/C) with (orange) and without (green) Cdh/EMI1/Skp1 (left). Upon binding of the protein ligands a significant temperature shift and a higher slope of the transition are seen. EM structures with (EMDB-2354, EMI1 marked in red) and without protein ligands (EMDB-2353) are shown. EMI1 binds in a central position in the complex, contacting many subunits, which explains the observed stabilizing. Note also that dissociated particles are only visible in the apo complex (white circles).

To verify a structural stabilization of the APC/C by EMI1, the complex was subjected to ProteoPlex both in the presence and absence of Cdh1/EMI1/Skp1 (Figure 3.6). A stabilization of the APC/C in the presence of Cdh1/EMI1/Skp1, such that the melting point was shifted by 1.6 K, was observed. This stabilization effect was also readily visible in EM micrographs, where broken particles were observed only in the absence of Cdh1/EMI1/Skp1 (Figure 3.6, compare middle and right panels). The stabilization effect is also explained by recently determined 3D structures in which EMI1 is observed to reside in a position where it bridges the platform structure with the catalytic sites (Figure 3.6) stabilizing the relatively fragile connection between the APC/C platform and arc-lamp domains. This illustrates, that the method is in principle suitable to find stabilizing small molecule and proteinaceous ligands of macromolecular complexes.

However, not all interactions can be readily verified. In collaboration with the Kornberg Lab (Stanford, USA), a transcriptional pre-initiation complex consisting of RNA-Polymerase II and several general transcription factors was measured with and without a bound special transcription factor. The resulting curves superimposed perfectly and no contribution of the ligand was seen, even though it clearly bound to the complex. However, from earlier studies it is known this factor exclusively binds peripheral to a single subunit. For this no cooperative binding is anyways expected.

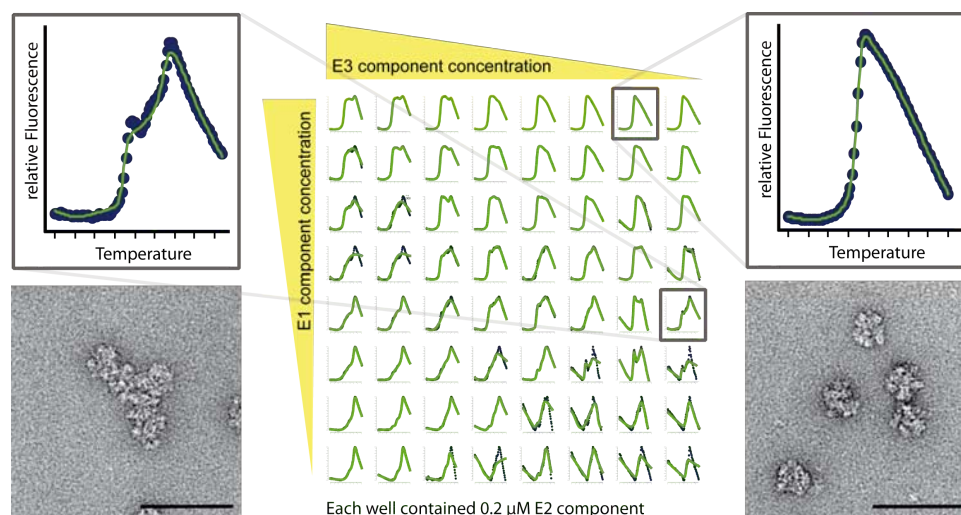
### 3.1.1.6 Reconstitution and Dissociation of Complexes aided by ProteoPlex

Complexes in cells are often found in various states and compositions to adapt to certain circumstances. For structural biology, this can be especially challenging since heterogeneous complex might not form crystals or are hard to interpret with other methods. Thus, biomolecular complexes are often reconstituted *in-vitro* from recombinant components. However, finding efficient reconstitution protocols remain a trial and error process requiring substantial amounts of protein. On the other hand, as described above, ProteoPlex is capable of efficiently finding stabilizing conditions for complexes with little amounts of sample. I thus asked the question if one could utilize the present method to find suitable conditions for the reconstitution of macromolecular complexes. A good model system for this is the *E.coli* PDHc, where all three subunits (E1, E2 and E3) are available in a purified state. To determine conditions, where PDHc could be efficiently reconstituted from individual subunits, I designed an experiment in which the concentration of E2 was kept constant throughout the entire experiment (Figure 3.7): along the x-axis I titrated molar equivalents of E3, while along the y-axis molar values of E1 were varied. In conditions where polyphasic unfolding transitions were obtained, electron micrographs revealed aggregated particles (Figure 3.7, left panel). In contrast, conditions which yielded single unfolding transitions, intact particles were visualized by EM that closely resembled particles assembled *in vivo* (Figure 3.7, right panel). Thus, ProteoPlex can be utilized to find efficient strategies for the reconstitution of macromolecular complexes from purified individual subunits with minimum sample requirements.

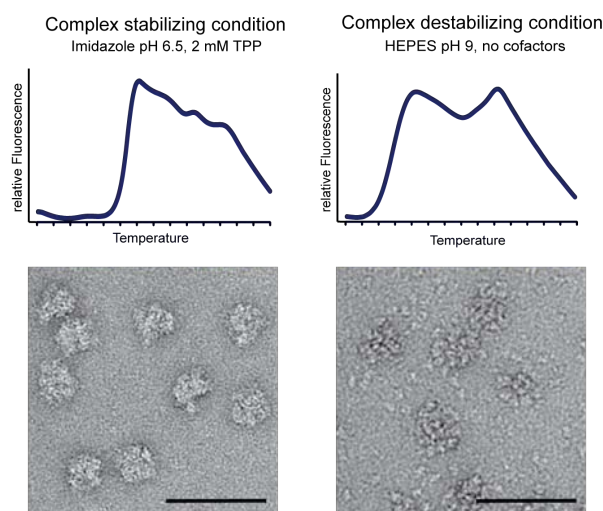
Sometimes it is necessary to destabilize a complex. For example, if a certain integral component should be purified. If a curve fits the model expressed in equation 3.20 with  $n > 4$  very well and clearly distinct transitions can be seen, the measured complex will have fallen apart. This is demonstrated again with the pyruvate dehydrogenase complex. It is known, that this complex disassembles in core complex and individual shell components at alkaline pH values.

Still, I ran a ProteoPlex screen over the full pH range. Strikingly, transition curves measured in alkaline condition showed two clearly separated transitions (cp figure 3.8 right panel), while more acidic conditions resulted in two-state-like curves. One can conclude that at least two species can be found in a PDHc sample under alkaline conditions, which matches with the obtained micrographs showing intact core particles and disassembles shells (cp figure 3.8 right panel).





**Figure 3.7: ProteoPlex assisted reconstitution of PDHc.** Reconstitution of PDHc from its individually purified subunits (E1, E2, E3). A constant concentration of the core E2 component was mixed with increasing amounts of E1 along the y-axis and increasing E3 amounts along the x-axis (5, 4 and 2.5, 1.5, 1, 0.5, 0.25 and 0.125 fold molar excess of E2 subunit) and assayed with ProteoPlex. Experimental curves (blue dots) and fits (green) from the screen (middle) show two-state unfolding behavior in the case of a high excess of E1 over E2 and low amounts of E3 and mono-disperse, compact particles in EM images (right panel). Whereas, low amounts of E1 and E3 in comparison to E2 yields polyphasic transitions and aggregated particles (left panel). The scale bar corresponds to 50 nm in the EM images.

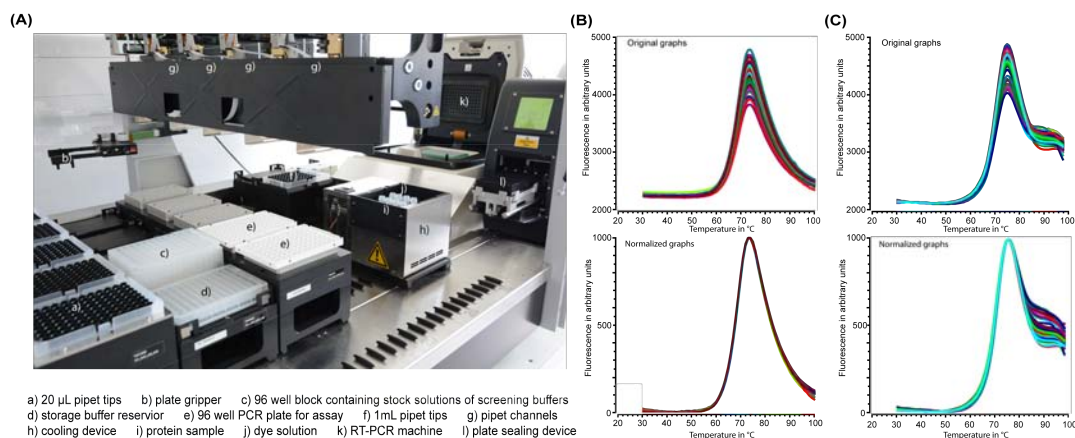


**Figure 3.8: Destabilisation of PDH.** ProteoPlex analysis can lead to two-state unfolding behavior, revealing monodisperse intact complexes in micrograph (scalebar = 50 nm) (left). However, in the same measurement destabilizing conditions (multistate behavior of the curve) are also found (right). EM micrographs of this particles in these conditions reveal a disassembled shell but intact core components (scale bar = 50 nm).

### 3.1.1.7 Automation

ProteoPlex experiments are not easy to pipette. To perform high throughput screens with little amounts of sample, small volumes have to be pipetted. As consequence, the reagents to be tested and the dye are highly concentrated and need to be dispensed in low volumes. Further, when mixing a protein containing solution, one can easily introduce air bubbles which scatter light and result in false values.

These parameters additionally contribute to the inaccuracy in measurement plate preparation. Therefore, considerable effort was spent to develop a fully automated platform (see figure 3.9 A), which handles all steps from sample preparation, plate sealing and the



**Figure 3.9: Automation Set-up and Reproducibility.** (A) Shown is an image of the fully automated integrated ProteoPlex solution. It consists of three main parts: 1) A liquid handling robot (Hamilton), which supplies the main platform on which all other peripheral devices are placed. 2) The peripheral devices, which are an automated plate sealer (Thermo Scientific) and a RT-PCR machine (BioRad). 3) A desktop computer (not shown), controls the liquid handling robot, acquisition of unfolding transitions in the RT-PCR device and data analysis. (B) The robot was set up to pipette the identical condition of a test protein sample in all 96 wells of the measurement plate. The resulting curves are shown in original and in normalized form. Although enormous variations in the absolute amplitude of unfolding transitions are seen for different wells (upper panel), upon normalization one recognizes that all graphs superimpose well. Typical standard deviations are around 0.5 K in  $T_m$ , which is the temperature inaccuracy of the RT-PCR device. This indicates that the pipetting errors of the liquid handling robot are lower than detectable by this measurement method. (C) Skilled humans were asked as well to pipette the identical condition of a test protein sample in 96 wells of the measurement plate. While the normalized graphs seem to reasonably overlap, a clear variance is visible (note the thickness of the curves in the beginning of the transition). Typical standard deviations are around 2 K in  $T_m$ .

measurement of unfolding transitions in an automated manner. The developed system is composed of a Hamilton Starlet liquid handler with 4 pipetting channels, a plate handling arm, a ThermoScientific sealing- and a BioRad Connect RT-PCR device (see figure 3.9 A). The experimenter has to provide sample, storage buffer (i.e. the buffer in which the protein of interest was purified) and a Sypro Orange dye solution diluted in storage buffer, which are kept at 4  $^{\circ}$ C in the course of plate preparation. In a typical screen, the machine will first pipette a 3-log dilution series of the protein and mix it with dye on a 96-well plate to determine the optimal protein concentration for screening. The plate is then automatically sealed and placed in the RT-PCR machine, which executes the measurement. Resulting unfolding transitions are automatically analyzed according to the thermodynamic framework described above and the concentration, which corresponds to an appropriate signal over background fluorescence, will be chosen as the protein stock solution to be utilized for screening. Afterwards, a screen of reagents of interest, provided in a deep well block will be pipetted and analyzed in the same way. It was increasingly experienced that automated liquid handling is essential for reproducibility (see figure 3.9 B). Even skilled experimenters tend to introduce a variance in pipetting, which would then falsify results (see figure 3.9 C). The resulting unfolding transitions are normalized and analyzed as described earlier. The system will provide a suggestion for the addition

of the respective screening solution to be included in the next round of purification. The whole procedure takes about 3 h and is ideally re-iterated with freshly prepared sample using yet another buffer and/or a small molecule ligand screen until convergence.

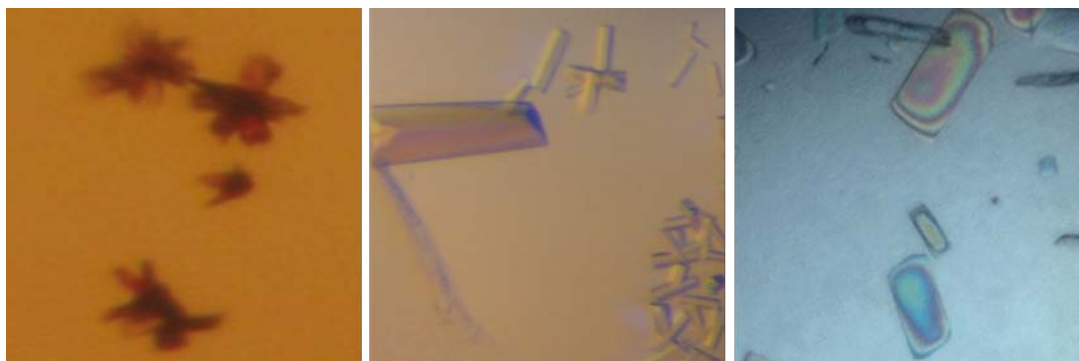
### 3.1.1.8 ProteoPlex and Crystallization

Throughout the explained test experiments, electron microscopy was used for validation. A definite relationship between two-state folding curve shapes and a monodisperse distribution of particles could be seen. Thus, ProteoPlex can be used as a tool to analyze samples towards the applicability of single particle cryo EM. However, most structural biologist still use crystallography as the main method. Thus, correlations between ProteoPlex curve shapes and crystallizability were searched. The general ThermoFluor approach is known to give a good correlation between melting temperatures and crystal growth. To validate that for ProteoPlex I searched the literature for crystallization conditions for the tested 86 complexes. In only five cases successful crystallization was reported. However, in those cases strikingly, the obtained ProteoPlex optimized conditions matched almost perfectly with the published crystallization conditions (see table 3.1).

**Table 3.1: Correlation between Crystallization Buffer and ProteoPlex-buffer.** For six of the 86 tested complexes published crystallization conditions were found. All of those match with the optimized result buffer obtained by ProteoPlex.

Protein Complex	ProteoPlex optimized buffer	Crystallization buffer	Citation
EFG-70S ribosome complex	MES, pH 6.5	MES, pH 6.5	[61]
p97	HEPES, pH 6.8	HEPES, pH 7	[42]
7S	Tris, pH 8.0	Tris, pH 7-6-8.2	[228]
6S	HEPES, pH 6.8	HEPES pH 7.0	[72]
8S	HEPES, pH 8.0	HEPES pH 8.0	[72]
crm1-Ran(GTP)-snurportin	HEPES, pH 8.0	Tris pH 8.0	[128]

To further check this correlation, crystallization trials for three of the 86 complexes without known crystallization conditions were performed in the optimized buffer conditions. Encouragingly, for all of them crystals were found after a short time (see figure 3.10).



**Figure 3.10: Crystallisation Success after ProteoPlex.** Three different previously uncrystallized protein complexes crystallized right away in the ProteoPlex result buffer.

### 3.1.1.9 ProteoPlex: Data Overview

To test and validate ProteoPlex 86 different complexes were analysed. An overview of all of them can be found in the tables 3.2, 3.5, 3.3 and 3.4.

**Table 3.2: Origin of optimized Samples**

Organism group	Number of analysed complexes
Bacteria	10
Plantae	1
Fungi	23
Animalia	47
Archea	1
viral	2

**Table 3.3: Sizes of optimized Samples**

Number of Chains	Number of analysed complexes
1 (multi-domain)	13
2 to 4	18
5 to 8	19
9 to 15	4
more than 15	30

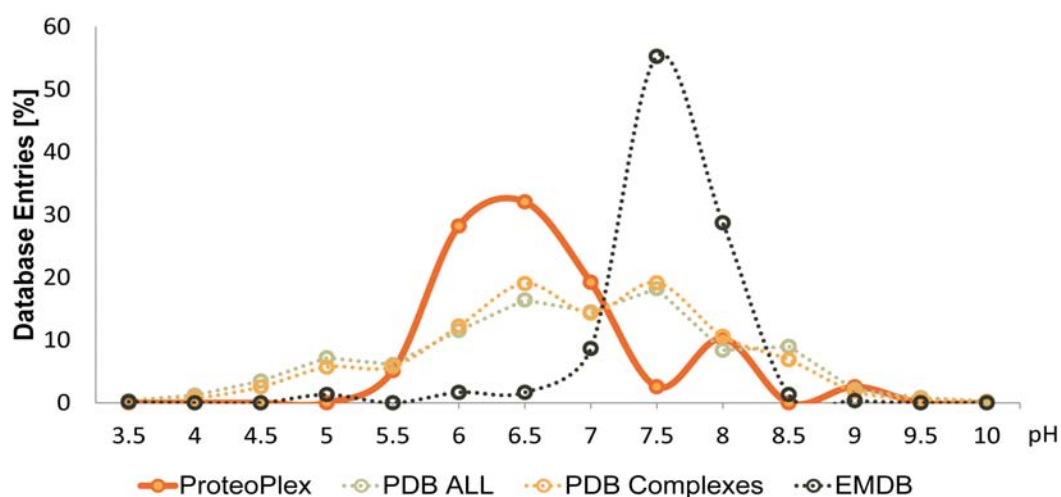
**Table 3.4: Localization of optimized Samples**

Subcellular Localisation	Number of analysed Complexes
Extracellular	7
Cytoplasmic	40
Mitochondrial	6
Nucleoplasmic	31

**Table 3.5: Function of the optimized Complexes**

Function	Number of analysed complexes
Metabolic enzyme	4
Scaffolding protein	3
Chaperone	7
mRNA-metabolism	25
Transcription	5
Translation	6
Proteostasis	6
Subcellular transport	20
molecular motor	4
Miscellaneous	4

Samples were obtained from all domains of life, and sub cellular localization as well as sizes. While a small bias towards certain complexes like those involved in RNA-metabolism can be seen, this is still a sufficient sample to make a few general remarks.

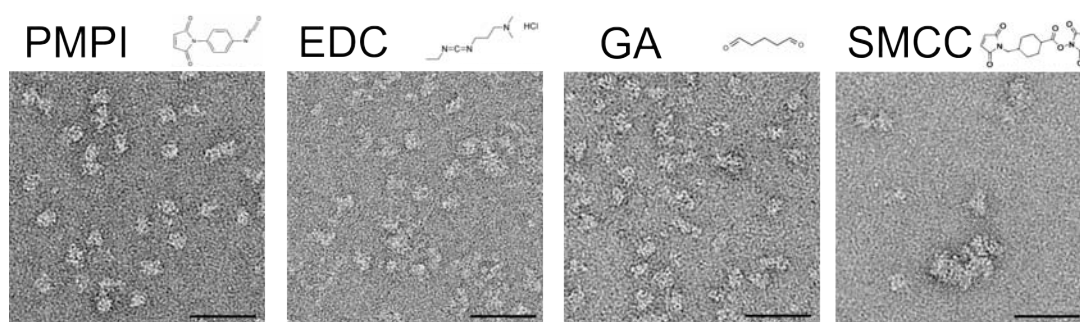


**Figure 3.11: pH Distribution of optimal Buffers.** The pH distribution of optimized buffers is plotted in orange (solid line). Further the distribution of the enlisted buffers from all relevant EMDB and PDB entries is shown as dotted lines. While at least one peak of the pdb entries matches our result. There is no match with the EMDB entries.

First of all looking on the distribution of result pH-values (see figure 3.11) one notices it to be rather broad and several Gaussian shaped peaks can be identified. Surprisingly, these peaks are not at the neutral pH but rather slightly acidic or alkaline. To set this in a context for comparison, surveys on the PDB and EMDB were made. While the crystallization buffer pH-value distribution obtained from the PDB matches the ProteoPlex distribution quite well, a clear pH-value peak at the neutral range can be seen for the EMDB-pH values. Similar broad distributions are found for the buffer chemistry and the additives. As expected there is not one generic buffer system which makes ProteoPlex a prerequisite for a successful structural analysis of a given complex.

### 3.1.2 Modifications of the GraFix methodology

Even if the most stabilizing conditions for a complex can be found using the ProteoPlex approach, most complexes are still fragile objects and need to be further stabilized to withstand the adsorption to the carbon film during grid preparation or the buffer exchange into a sugar free buffer for vitrification. To assure this the GraFix methodology was introduced in 2008. As it was originally introduced, glycerol was used as density gradient and glutaraldehyde as cross-linking agent. For six different complexes, different crosslinking agents as well as different sugars as densitizing agents were used. Figure 3.12 shows an example of such a study. Here the same preparation of the RNA induced silencing complex (RISC) loading complex was used as test sample. The sample was provided by Stephanie Schell (AG Ficner, University of Göttingen, Germany). Here, the crosslinking agents, glutaraldehyde, N-(p-Maleimidophenyl)isocyanate (PMPI), 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide) (EDC) and Succinimidyl-4-(N-maleimidomethyl)cyclohexane-1-carboxylate (SMCC) were tested. As seen in the micrograph a homogeneous monodisperse distribution of complexes in the micrograph is best obtained using the crosslinkers glutaraldehyde and PMPI. As shown previously [120] a combination of PMPI and glutaraldehyde is beneficial for several RNPs. However, this cannot be generalized for all complexes and testing available crosslinking agents is essential to find the best preparation conditions. The same result was found for different density gradient forming agents. While sucrose performs good in most cases, other sugars may be better suited in some cases.



**Figure 3.12:** Behavior of different crosslinkers for GraFix. The RISC loading complex was subjected to conventional GraFix using different crosslinkers: N-(p-Maleimidophenyl)isocyanate (PMPI), 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide(EDC), glutaraldehyde (GA) and Succinimidyl-4-(N-maleimidomethyl)cyclohexane-1-carboxylate (SMCC). In all cases the peak fraction was analysed as judged by Dot-blot. For PMPI and GA a monodisperse field of view can be seen. For EDC a few aggregates are visible while SMCC produces almost only aggregates. (scalebar = 50 nm)

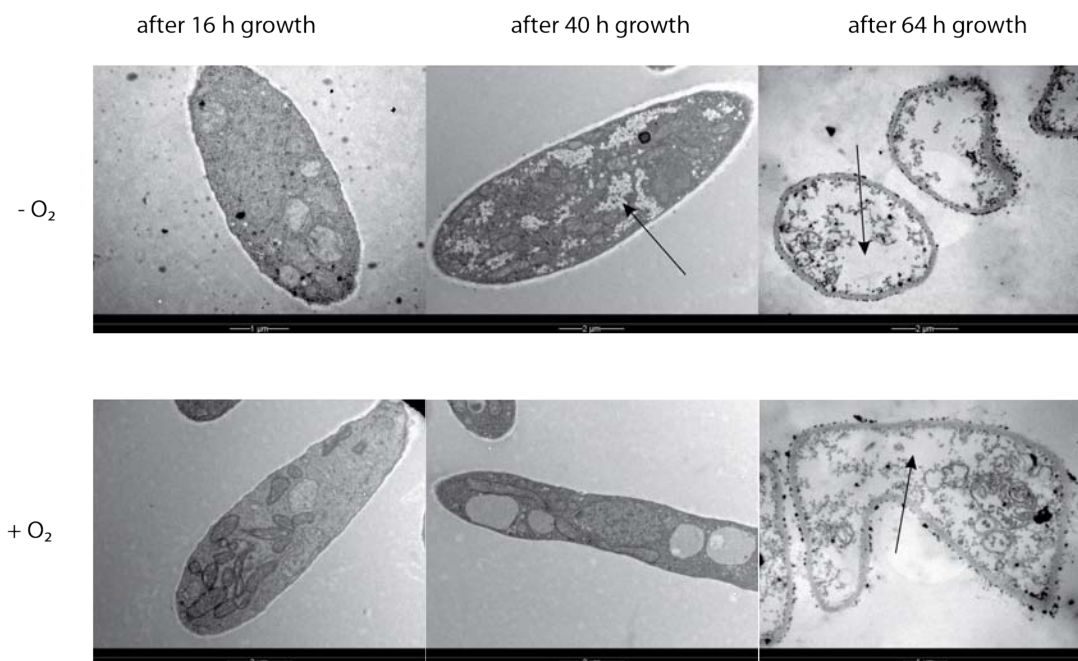
As shown later (see section 3.3.3), structural analysis of the stabilized PDHc was severely impaired by classification problems. This hints towards a large heterogeneity in the dataset. This is either explainable by many differently broken complexes or a continuum of conformational structural states. Based on the Stark labs work on ribosome dynam-

ics [52], it is known that lowering the temperature leads to a significant stabilization of complexes in their thermodynamically favored conformation. However, complexes are already analyzed routinely at 4 °C in the Stark lab. For further cooling of particles and enrichment of ground state molecules, a new method termed cryoFix was developed (Initial experiments towards this methods were made by Florian Platzmann [160]): Solutions with high sucrose content (> 10% (w/v)) freeze at a temperature below -10 °C, implying complexes can be cooled down to this temperature without freezing. Unfortunately, high sucrose content reduces the contrast of particles in the electron microscope. To overcome this problem, GraFix was again modified. Gradients are cooled here down to -10 °C during the run. In this way, the complex is slowly cooled to the negative temperature, populating a ground state and is finally trapped in this state by the cross linking agent. After centrifugation, the complex can be transferred to 4 °C, the sucrose can be removed and electron microscopic analysis can be performed by standard procedures. This thesis contributes to the exploration phase of this by testing several proteins as described.

### 3.1.3 *Chaetomium thermophilum*

Even though the number of different conformations in the sample can decrease significantly using CryoFix, there can be still too many states to solve a high resolution structure. Since further cooling below -10 °C without freezing the solution is not possible, I started working with the thermophilic organism *Chaetomium thermophilum*. Similar growth rates of this fungus compared to closely related fungi let assume that dynamics of the thermophilic proteins at 50 °C should be comparable to the ones of mesophilic organisms at 30 °C. Extrapolating this to 4 °C or even -10 °C, one can assume a low number of different conformational states for the thermophilic proteins at those temperatures. Despite this thermal adaptation of proteins, the sequence similarity of most *Chaetomium* complexes to human homologs is high.

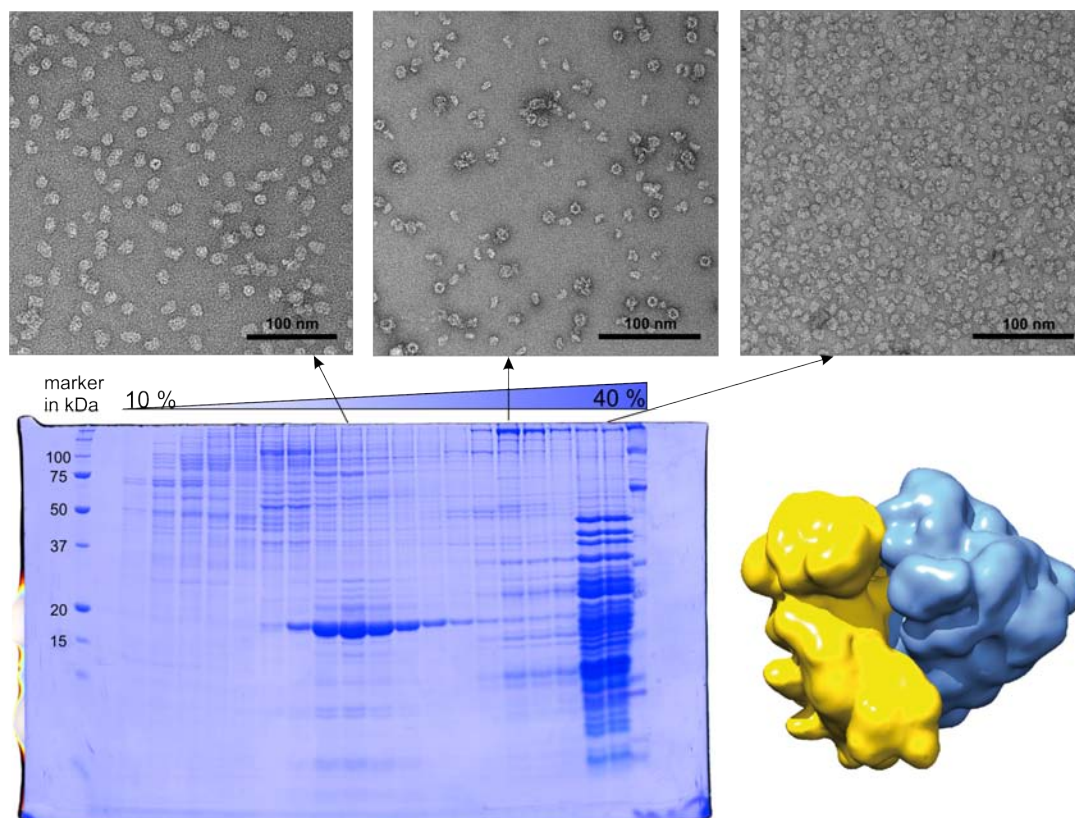
With this I approached cultivation of the fungus (with kind help of the Hurt group, University of Heidelberg). Several trials to scale up a shaking flask cultivation up to a fermenter scale failed initially. The first extracts revealed almost no protein, making growth conditions optimization necessary. The most important cultivation parameters were found to be the aeration of the culture and the fermentation time. Figure 3.13 shows exemplified cells sections at different time points of a fermentation. With increasing time the cells seem to degrade their cytoplasm, especially if the aeration is not sufficient. A fermentation time of 16 h was chosen to assure intact cells. The cause for degradation of the whole cell interior remains still unclear and needs to be further addressed in future work.



**Figure 3.13: Growth Conditions of *Chaetomium thermophilum*.** The same starter culture was grown with no and with good aeration. Samples were taken, embedded, dissected and analyzed by electron microscopy. The EM preparation was performed by Dr. Dietmar Riedel. The empty space visible in the cells after long fermentation (black arrows) indicates that the cells dissolve with higher fermentation times especially if no oxygen is given.

With this up-scaling, a total cell mass of 1.6 kg of cells was obtained from a 100 L fermenter. Standard purification protocols for complexes from human and yeast source could readily adapted to this new organism. With this, an attempt to purify endogenous complexes from the fungus was possible. In a proof of principle experiment, a full cytoplasm extract was prepared, fractionated by PEG precipitation and density gradient sedimentation. The results are depicted in figure 3.14. It can be seen that right away in this initial experiment monodisperse almost pure complexes were obtained and imaged on the micrographs. Thus, using complexes originating from this fungus could aid structural investigation.





**Figure 3.14: Test purification of *Chaetomium thermophilum* complexes.** A full cell extract was prepared from 100 g of *Chaetomium thermophilum* cells. The extract was precipitated with 10 % of PEG6000. Subsequently, the dissolved pellet was loaded onto a 10-40 % sucrose gradient and sedimented. The individual fractions were analyzed by SDS-PAGE (lower panel left) and by negative stain EM. Distinct, mono disperse complexes are visible. In the last fractions (right) clearly ribosomes are visible. A structural analysis of those resulted right away in the depicted 3D structure, which was calculated by Wen-ti Liu (lower panel right).

## 3.2 Structural and Dynamical Insights into CRM1

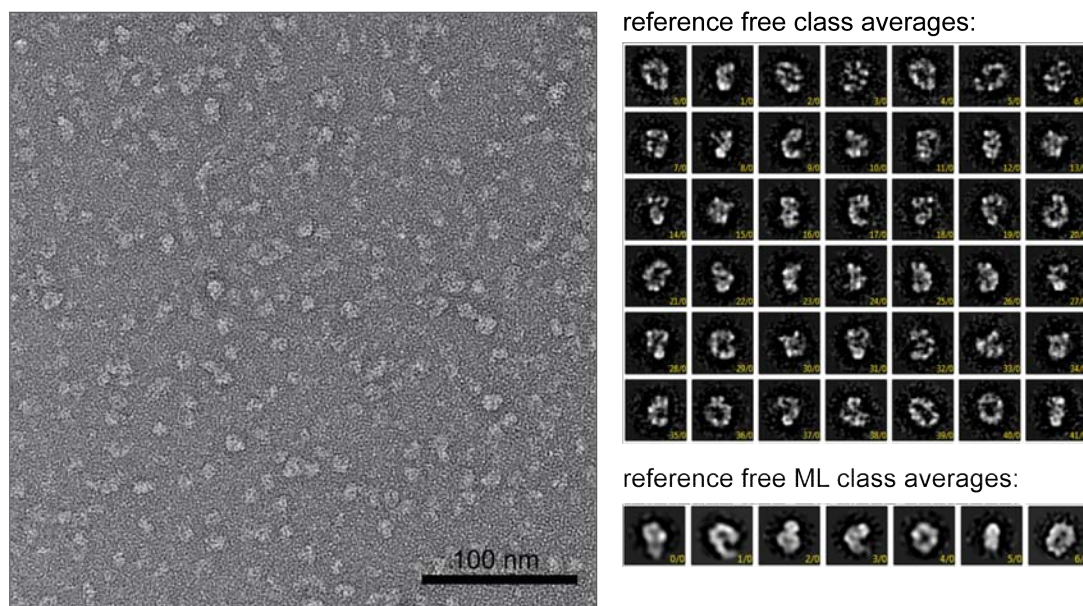
The first biological model system I used in this thesis for the analysis of conformational dynamics was the nuclear export factor CRM1. With a molecular weight of 120 kDa, CRM1 is rather small for EM analysis and therefore especially challenging.

To fulfil its task, it binds a cargo and Ran GTP in the nucleus, diffuses as complex through the NPC, releases both upon hydrolysis of GTP in the cytoplasm and diffuses back into the nucleus. To be able to understand this functional and conformational cycle of CRM1, it is crucial to gain structural insights into the apo-state as well as the cargo bound state. At the beginning of this thesis, several CRM1 crystal structures with bound cargo or Ran were already solved [130]. However, there was no insight into the structure of apo-CRM1. During the course of the work on this thesis, the lab of Ralf Ficner (University of Göttingen) succeeded in crystallizing free CRM1 from the thermophile *Chaetomium thermophilum* [129]. Surprisingly, the structure was very different to the already obtained structures of yeast and human CRM1-cargo/Ran-complexes. While the complexed CRM1 appears to adopt a closed ring structure, the newly solved apo structure shows an open superhelical shape. However, crystallization can force molecules in unflavored conformations. Thus, Electron microscopy analysis was performed to analyze the conformational variability of CRM1 in solution. This was also repeated with the human protein. The results of both investigations, which were already published in [129] and [46], are summarized in the following sections.

### 3.2.1 *Chaetomium thermophilum* CRM1

*C. thermophilum* CRM1 was expressed recombinantly in *E. coli* and purified by Dr. Thomas Monecke (AG Ficner, University of Göttingen). A full ProteoPlex screen was performed yielding conditions identical to the successful crystallization conditions (Hepes pH 8.0), which were used for all further analysis. The protein was subjected to the standard GraFix procedure and subsequently analyzed by negative stain EM. An exemplified micrograph is shown in figure 3.15.

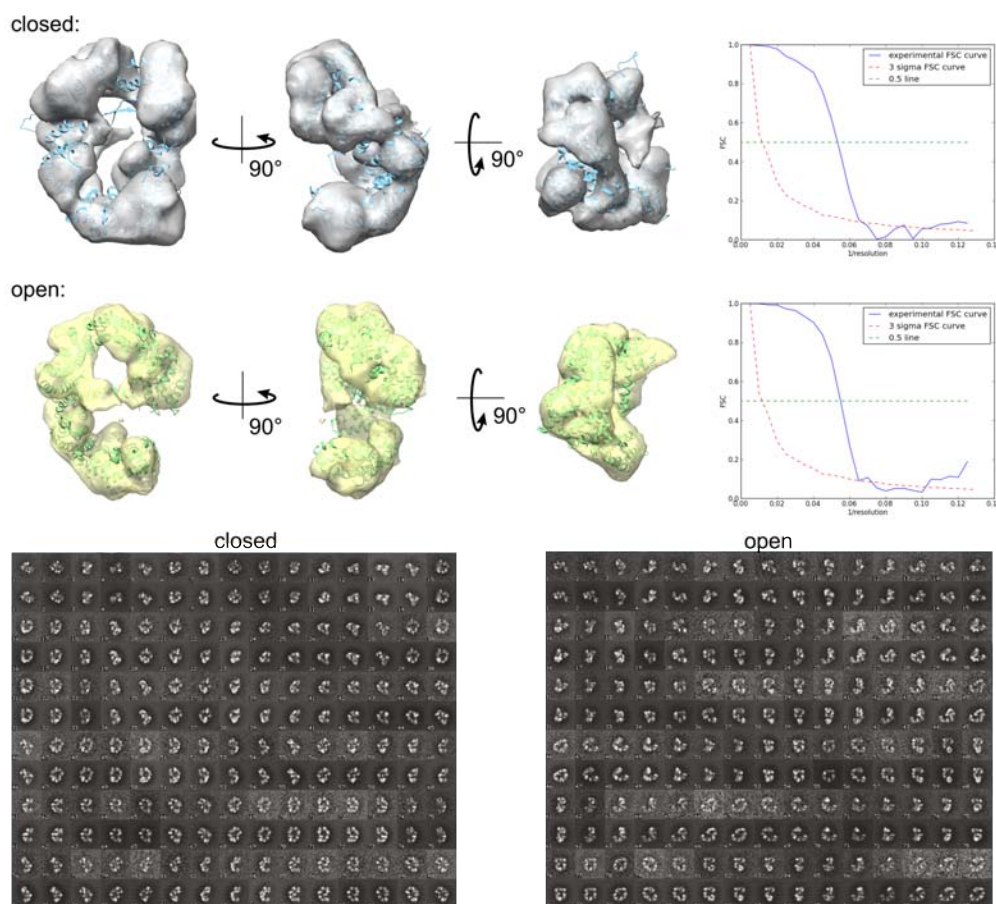
A monodisperse field of particle views was obtained allowing standard image analysis techniques to be applied. 300 micrographs with a pixel size of 1.85 Å/px were recorded and 31970 particles were selected. To avoid model bias reference-free class averages were calculated via reference-free alignment using *imagic* and 2D maximum-likelihood classification using *relion*. In figure 3.15 representative class averages are shown. Already at this stage class averages can be found that hardly fit into a closed ring structure as the previous crystal structures suggest. To obtain a valid initial 3D model, several independent



**Figure 3.15: *Chaetomium* CRM1 Raw Data.** Raw data and initial class averages of *Chaetomium* CRM1. Left: A representative negative stain raw micrograph is shown. Right: typical class averages obtained from the data are depicted by standard classification and maximum likelihood classification.

strategies were used in parallel. Firstly, angular reconstitution was used in combination with resampling techniques. Therefore, the obtained 1000 class averages were resampled with replacement into 250 sets of 12 class averages and an angular reconstitution 3D model was calculated. The resulting 3D models were manually inspected and sorted for similarity. The very same class averages were also subjected to *simple PRIME*, a software using a reference-free projection matching approach. Both strategies yielded very similar 3D models. In both cases the same two distinct conformations could be identified.

One model resembles the ring shaped, closed form as known from the cargo complexes and the other appears to be an open superhelical, pitched structure, which matches the newly obtained crystal structure model (cp. figure 3.16). Both models were refined until convergence competitively against the full dataset. To assure the validity of both models, a cross validation strategy was used. Therefore, the dataset was split into two groups according to the best fit to one of the two models. Subsequently, images assigned to the closed ring conformation were refined against the open superhelical model and vice versa. In both cases the resulting model relaxed back from the wrongly assigned model to the original model after several iterations. This way, the final models were obtained directly from the split dataset almost without model bias and can be considered valid. Further verification of the obtained models could be gained by fitting the previously mentioned crystal structures into the models as shown in figure 3.16. To remove model bias, the crystal structures were filtered to a resolution of 20 Å. Both models fitted very well to the crystal structures.

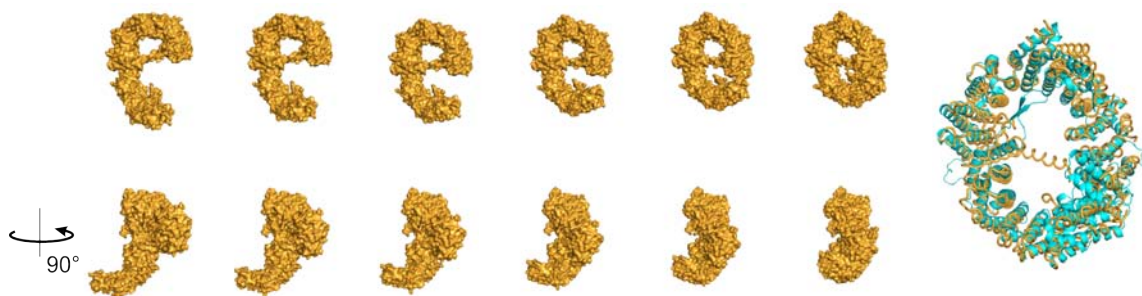


**Figure 3.16: *Chaetomium thermophilum* CRM1 conformations.** Density models of *Chaetomium* CRM1 conformations. The resulting two conformations- the compact, closed (grey) and the superhelical, open (yellow) are shown. In both cases the atomic models taken from the PDB were fitted: for the closed model PDB: 3gix and for the open model: PDB: 4fgv

After the validation of the two models, a first approximation of the energy landscape can be attempted. The distribution of the particle numbers assigned to the model allow an estimation of the energy difference between the two conformations. 18953 particles were assigned to the closed ring shaped model and 13017 to the superhelical model. Using equation 1.1 one can approximate an energy difference of  $\Delta G = -1.5 \frac{\text{kJ}}{\text{mol}}$  corresponding to  $0.7 \text{ k}_B\text{T}$ . Thus, the energy of the surrounding medium is sufficient to shift the protein from the one conformation to the other.

To gain a better resolution of the energy landscape, it was attempted to resolve more conformational states. As the two independent approaches used to find the initial 3D structures yielded the same two conformations, one can assume that no other major conformational change takes place. This means that further conformations should only be intermediates between the known two. To identify those, new reference models were calculated using a normal mode analysis. This was done using both the obtained EM structure and the crystal structure respectively as ground state model. Interestingly, in

both cases the first non-trivial mode<sup>1</sup> describes exactly the opening movement as it can be seen in figure 3.17. Interestingly, one of the simulated 3D models from the NMA of the open superhelical model matches the experimentally gained closed structure almost perfectly (cp figure 3.17 right)



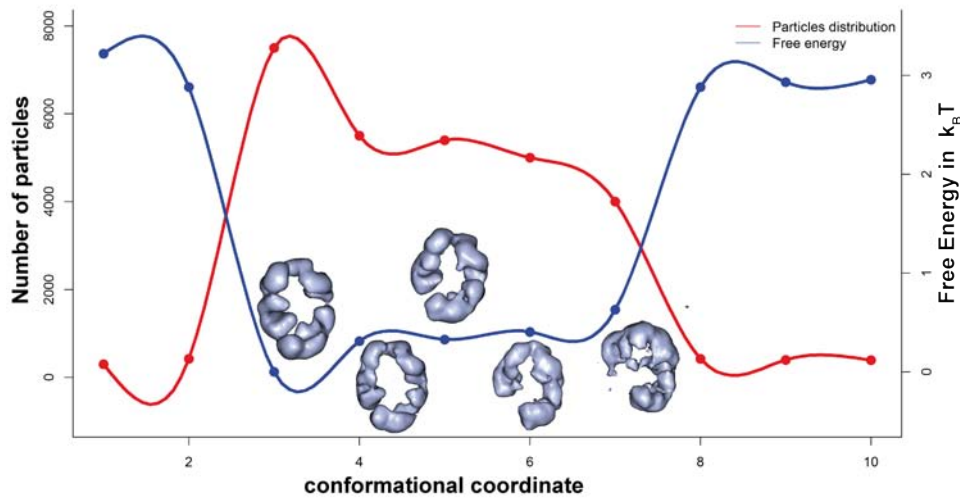
**Figure 3.17: Mode 7 of NMA of CRM1.** Mode 7 of the normal mode analysis of CRM1 is shown. Starting from the open CRM1 crystal structure an NMA was calculated. Intermediates of the first non-trivial mode (mode 7) are shown. On the right the superimposition simulated closed state (orange) and the experimentally determined closed state (green) is shown.

This further indicates, that the opening and closing of the ring is an energetically favoured movement which occurs rapidly, because the NMA modes are by default sorted with respect to increasing energy. To identify intermediate states in the dataset, 10 equally spaced structures from the NMA trajectory were used as references for a competitive alignment against the full dataset. To minimize the model bias, structures were calculated by angular reconstitution from class averages not by projection matching in the first few iterations. The resulting structures were subjected to three further rounds of projection matching. Five of the ten tested conformations yielded reasonable structures. Thus, the 13000 closed state particles are split in 7500 closed and 5500 almost closed particles, while the 18000 which aligned to the open conformation were in three almost equally populated states (see figure 3.18).

The newly found almost closed conformation of CRM1 is a compact ring structure, which shows a small but significant gap between the N- and C-terminal regions. MD simulations performed in the group of Helmut Grubmüller (Max Planck Institute for biophysical Chemistry, Göttingen, Germany) support the existence this intermediate.

In general, this more detailed analysis provides a more accurate description of the energy landscape of conformational transitions. While the various open conformations are mostly equally populated, meaning there is almost no energy barrier between them, the closed conformation is slightly higher populated and therefore slightly stabilized. The free energy difference between the open and the closed state ensemble can now be better estimated as  $0.4k_B T = 0.9 \frac{\text{kJ}}{\text{mol}}$ .

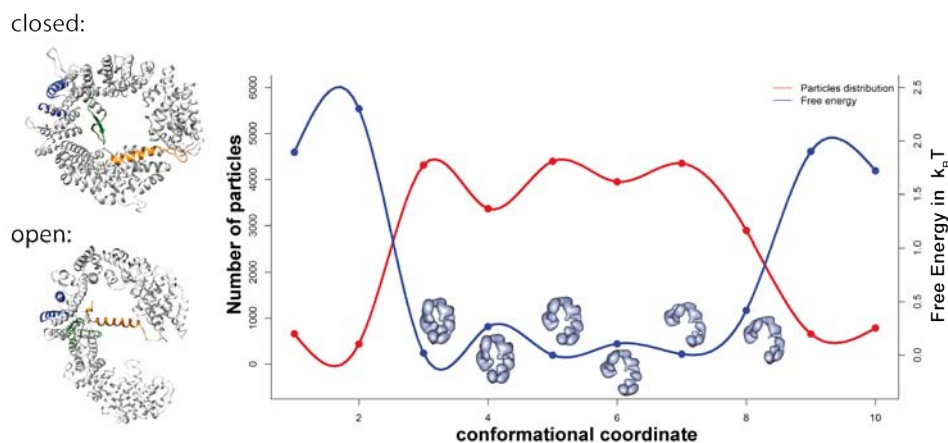
<sup>1</sup>The first six modes gained from an NMA describe the three possible translations and rotations of the full molecule. They are thus called trivial since they contain no information on the molecules flexibility.



**Figure 3.18: Energy Landscape *C. thermophilum* CRM1.** An approximation on the energy landscape of free CRM1 is shown. Ten equally spaced intermediate structures from NMA mode 7 were produced and all particles images were refined against these models. Particles were assigned according to the best fitting model. The number of particles in the individual substates is plotted against the state as conformational coordinate. Further, the free energy was calculated following equation 1.1 to gain a first estimate on the conformational landscape. For the significantly populated states calculated structures are depicted.

Even though this energy difference is rather small, it is significant and most probably important for CRM1's conformational cycle. Thus, it is necessary to find the molecular reason for this difference. In principal, it can either arise from an entropic or enthalpic contribution or for both. Entropically the open state is much more favored since more different conformational states can be populated. Thus, the energy difference has to be explained by the enthalpic contribution. Most likely it is the binding energy of the two termini to each other. The crystal structures determined by Thomas Monecke predict that the major determinant for the two conformations is a C-terminal helix. This expands to the NES-cleft in the open conformation, but is folded back to the terminus in the closed conformation, most likely to facilitate binding of the termini (see figure 3.19 left).

To investigate this assumption, a variant of the protein deleting this helix was produced. While the helix itself is invisible at the current resolution range, the energy landscape can still be analyzed. It has to be noted that the overall quality of the sample got much worse due to the deletion and significant aggregation of CRM1 particles was observed in the micrographs. Nevertheless, through manual selection of the isolated particles the full computational analysis could be performed. Here, the resulting energy landscape was flat, meaning the particles were rather evenly distributed among all of the conformations (see figure 3.19 right). The deletion of the helix does result in open and closed structures that do no longer differ significantly in energy, confirming the hypothesis.



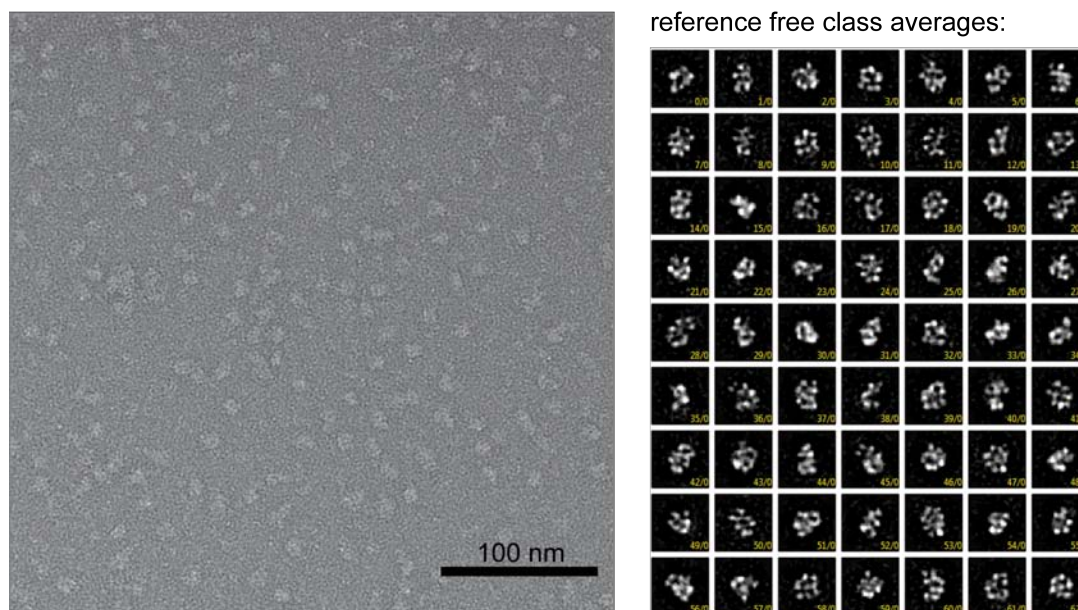
**Figure 3.19: Significance of the C-terminal helix.** Left: The conformational change of the C-terminal helix (orange) is shown in the crystal structures. In the open conformation, the helix expands towards the acidic loop (green) and NES-cleft (blue) while it folds back to the interaction surface between the termini in the closed conformation. Right: An representation of the energy landscape of free CRM1 with deleted C-terminus is shown. Ten equally spaced intermediate structures from the described NMA mode 7 were produced and all particles images were refined against these models. Particles were assigned to the model they fitted best to. The number of particles in the individual substates is plotted against the state as reaction coordinate. Further, the free energy was calculated following equation 1.1 to gain a first estimate on the conformational landscape. For the significantly populated states, the refined structures are shown.

To finalize the analysis, higher order NMA modes were analyzed in the same manner. This, however, resulted in no significantly different structures. Thus, the energy landscape of the *C. thermophilum* CRM1 seems to be almost fully described by the opening and closing movement, which seems to be purely driven by thermal energy.

### 3.2.2 Human CRM1

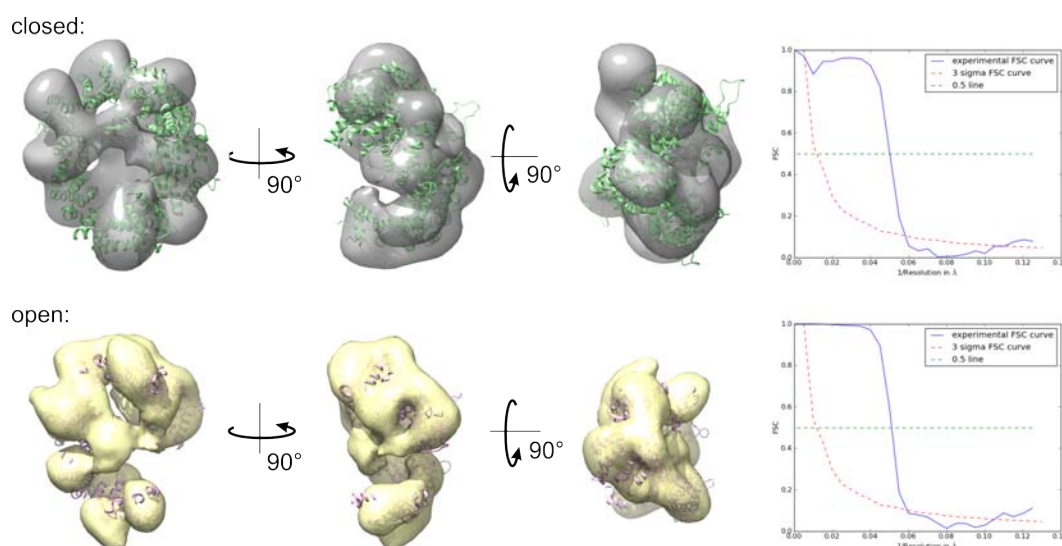
For the human CRM1 the same approach as for the *C. thermophilum* CRM1 was used. However, no good micrographs and class averages could be obtained from the standard GraFix approach. It was assumed that the conformational flexibility could be the reason for that. Thus, the previously described (see section 3.1.2) cryoFix method was applied. With this approach good micrographs and class averages could be obtained (see figure 3.20 ).

For initial 3D models the same methodology as described for the *C. thermophilum* CRM1 was used. Again, the analysis resulted in two distinct states describing the open and closed conformation of the protein which look very similar to the obtained *C.thermophilum* structure. The only significant difference between the models from the two species is that the center of the human CRM1 shows significant density while the fungal structures are empty. After refinement both structures reached a resolution of about 20 Å, which is considerably lower as for the *C.thermophilum*. Due to the lack of atomic models for the



**Figure 3.20: Human CRM1 Raw Data** Raw data and initial class averages of human CRM1. Left: a representative negative stain raw micrograph is shown. Right: typical class averages obtained from the data are depicted.

human protein, the X-ray models from the *C.thermophilum* CRM1 were fitted. Since the HEAT-repeats are only moderately homologous between the two CRM1 the obtained poor quality fit was not surprising. To corroborate the correctness of the results, a cross validation experiment was performed as described for the fungal models.



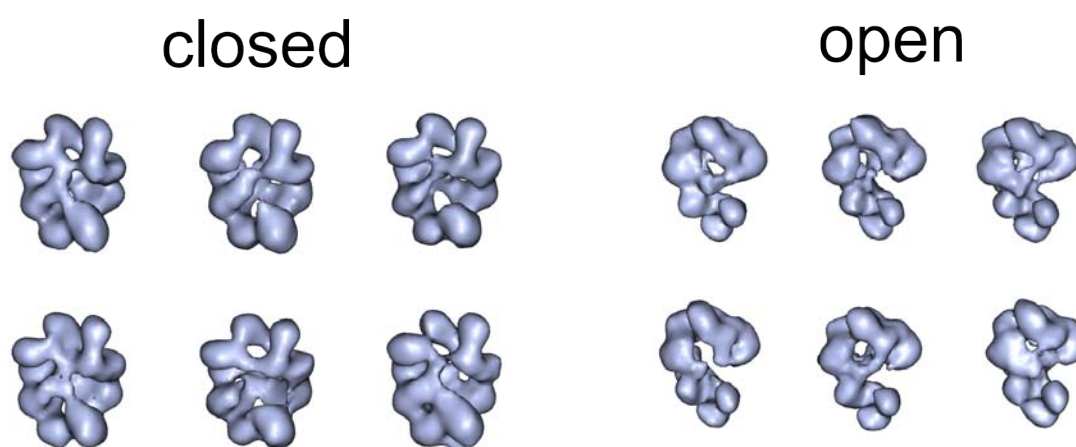
**Figure 3.21: Human CRM1 Conformations** Density models of human CRM1 conformations. The resulting two conformations - the compact, closed (grey) and the superhelical, open (yellow) are shown. In both cases the atomic models taken from the PDB were fitted: for the closed model PDB: 3gjx and for the open model: PDB: 4fgv. On the right the FSC-curves are shown for the respective model. At a threshold of 0.5 both models have a resolution of about 20 Å.

After successful validation of the models, the energy between the two states can be estimated. Curiously, both models are almost equally populated with 19254 in open superhe-



lical conformation and 22854 in the closed, compact conformation. This ratio concludes that the energy difference between the two states is minor, suggesting that human CRM1 can freely sample its full conformational space because a significant energy barrier is missing.

As for the *C. thermophilum*, the obtained models should be subdivided into sub states. However, while NMA worked very well for the thermophile CRM1, all attempts towards this failed for the human CRM1. This failure is most likely due to the extra internal mass which stiffens the model at least for the NMA analysis. Thus, the conformational flexibility was analyzed by applying a resampling approach to the aligned particles. All of those were refined and are shown in figure 3.22.



**Figure 3.22: 3D-MSA of human CRM1.** Refined structures from the most abundant 3D-MSA results are shown. It can be seen that for both states the central density is the major source of heterogeneity.

The results obtained revealed a large number of distinct structures all differing in the structural appearance of the central density. This explains the finding of the strong central density that was obtained. It can be speculated that different interactions of acidic loop and C-terminal helix give rise to this density. This, however, cannot be shown at this level of resolution. Thus, it can be concluded that both of the two main states sample a large number of substates which would result yet again in a very flat energy landscape.

### 3.3 Structure and Dynamics of the *E.coli* Pyruvate Dehydrogenase Complex

The *E.coli* Pyruvate Dehydrogenase Complex (PDHc) was used as a model system to analyse the conformational heterogeneity of a large macromolecular complex. With more than 60 years [107] of research on that complex, its biochemistry is very well understood. However, almost no structural insights into the complete complex could be gained this far. Here, I describe the results obtained from structural investigation attempts. This was done in collaboration with the Tittmann group (University of Göttingen). At the beginning of this thesis, many PDHc purification protocols were already established and documented in the literature [68, 106, 195]. Since structural investigation often needs substantial amounts of material, an overexpression strategy was initially developed by the Tittmann group. However, the overexpression strategy of the full *E.coli* operon leads to a nonstoichiometric over abundance of the E3 component which - as shown via reconstitution attempts - counteracts a regular assembly. To avoid assembly problems, I started working on purification a strategy of the endogenous complex.

#### 3.3.1 Purification of *E.coli* PDH

Research on the PDHc was initiated more than 60 years ago, at a time where no recombinant techniques were available. Thus, protocols for a native purification are established [68, 106, 195]. However, many of those protocols could not be properly reproduced. This is most likely due to various chemicals such as polyethylene glycol or protamine sulphate, required for these protocols, are different in composition and purity. Because of the difficulties in the exact reproduction of the old protocols, new purification protocols were developed, taking the old published protocols as guideline. I assumed that the natural expression of the PDHc will in some way be related to carbon source provided in the bacterial medium. Thus, growth curves were analyzed for growing the bacteria in glucose, pyruvate and acetate. Best growth and PDHc activity of extracts were found for media containing glucose.

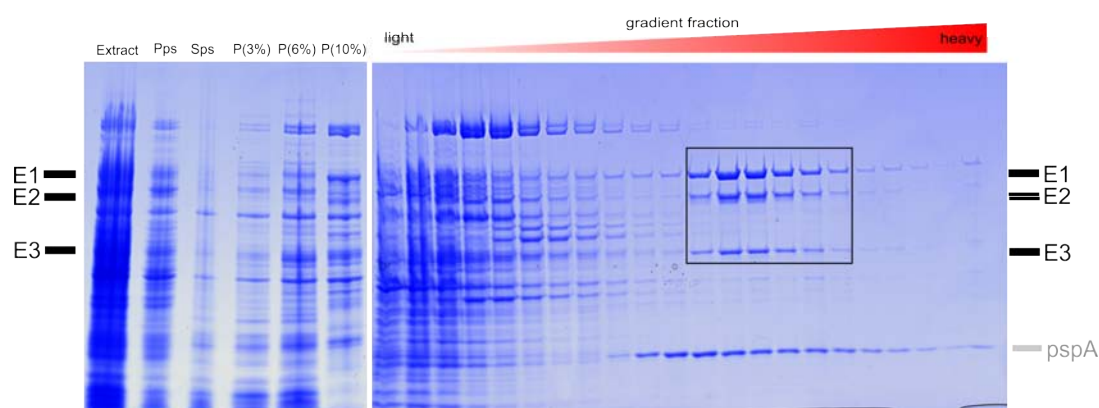
Having established good growth conditions, several cell breakage methods were tested. While opening cells by sonication or french press results in reduced yields and activity, grinding of the cells with aluminium oxide provides the highest activities in the extracts.

Since the PDHc is one of the largest complexes in the cell, a purification strategy based on precipitation and sedimentation was chosen, because it can be expected that there are only few possible other complexes in the cell that are of similar size. In this way, binding

to a matrix while isolating the complex could be avoided, which has proven to damage numerous other macromolecular complexes, studied in the Stark lab. The final optimized purification strategy involved four steps:

1. Cell breakage with aluminium oxide using an electrical mortar.
2. Protamine sulphate precipitation to remove most of the ribosomes.
3. Fractionated PEG precipitation.
4. Sedimentation in a sucrose gradient.

The individual steps were performed at 4 °C.

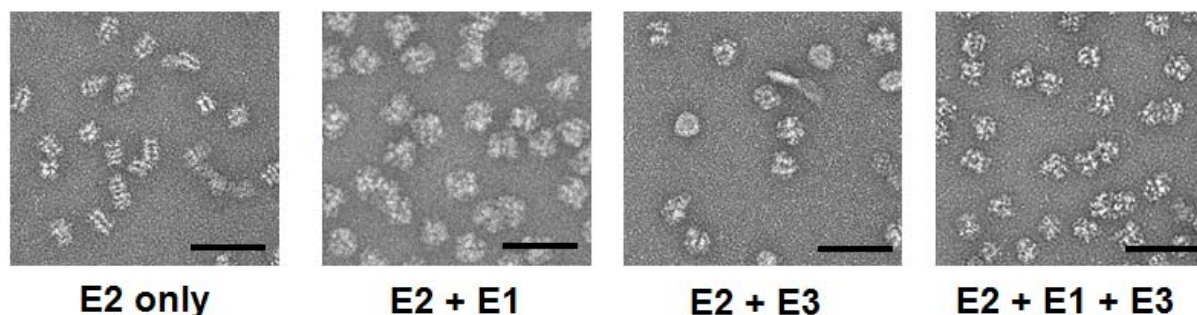


**Figure 3.23: PDHc Purification.** Every step of the purification was shown on an SDS-PAGE gradient gel (4-15% acrylamid). Thereby, Pps is the dissolved pellet fraction from the initial protamine sulphate precipitation, Sps its supernatant. The dissolved pellets from the 0-3 % (P(3%)), 3-6 % (P(6%)) and 6-10 % (P(10%)) PEG6000 precipitation step are adjacent. The dissolved 6-10 % pellet was separated on a 10-30 % (w/v) sucrose gradient. The gradient fractions are shown. The PDHc-fractions are framed. The respective lanes were analyzed by mass spectrometry, and identified proteins are marked on the side of the gel.

In each step, the active fractions were pooled and used for the further steps. As shown later, this protocol yielded nice monodisperse PDHc as judged by EM. The identity of the complex was verified by mass spectrometry in the group of Henning Urlaub (Max Planck Institute for biophysical Chemistry, Goettingen, Germany). The E1, E2 and E3 could be unambiguously identified. Some preparations, however, contained *phage shock protein A* (*psp A*) in significant amounts that co-purified very well. The micrographs showed no extra protein apart from the complex. Thus, it is unclear whether *psp A* is a significant part of the PDHc under certain condition or an artefact of the purification procedure. The purified complexes were subjected to ProteoPlex to find optimum buffer conditions that lead a maximum stabilization of PDHc for structural studies. As a result of the optimization screen a slightly acidic imidazole buffer pH 6.5, containing magnesium ions and necessarily thiamin diphosphate (TPP), was determined. Strikingly, the activity of this optimized complex even increased by 10%.

### 3.3.2 Reconstitution Trials

In parallel to the native purification of endogenous PDHc, reconstitution trials were performed. The individual components were expressed and purified by Michael Tietzel respectively Danilo Meyer from the Tittmann group (University of Göttingen, Germany). Several combinations of the three components in equimolar ratios were incubated for 30 min on 4 °C and then loaded onto a GraFix gradient. The peak fractions were analyzed with negative staining EM.



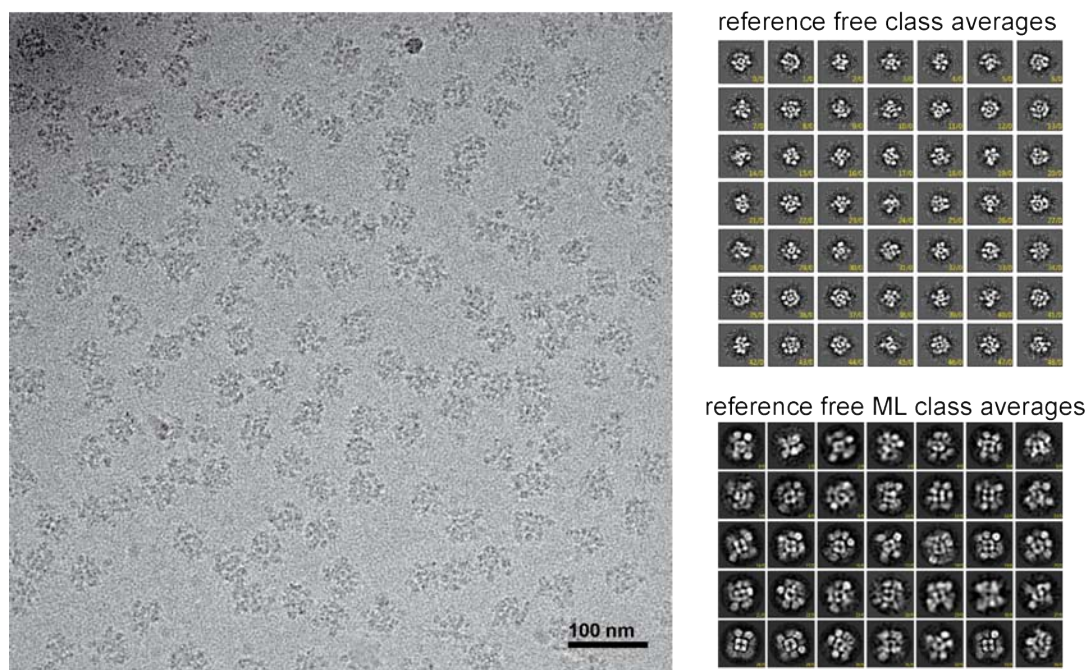
**Figure 3.24: Micrographs from PDHc Reconstitution Trials.** The reconstitution of the PDHc was attempted stepwise. Thereby, the individual components were mixed equimolar, incubated for several hours on 4 °C and loaded on a GraFix gradients. Micrographs of the peak fractions are shown. (scale bar = 100 nm).

For the E2, linear oligomers composed of the cubic cores were formed even under optimized assembly conditions. Only a small fraction of the cores seem to be monomeric respectively at most dimeric. Thus, most of the material pelleted in the gradient. Surprisingly, the addition of an excess of E1 results in the solution of those aggregates. In the GraFix gradient, a sharp peak of protein concentration is detectable, and almost no protein pelleted. This is in line with dynamic light scattering experiments performed by Michael Tietzel showing the same result. In contrast, the addition of the E3 component does not have the same effect, even though it binds at the same positions of the E2. While a mixture of E2 and E3 only forms a small amount of complexes leaving the E2 aggregates intact, the mixture of all components in equal stoichiometry led to fewer aggregates, but still not the homogeneity of an E1, E2 mixture. Thus, a more thorough search for the right stoichiometry was performed as explained earlier in section 3.1.1.6, using the newly developed ProteoPlex methodology. The idea was to measure the thermal stability of reconstituted complexes under varying concentrations of E1 and E3 with constant concentration of E2. The analysis suggests using a stoichiometry of E1:E2:E3 = 16:4:1 or 8:2:1. Those values are way beyond the published stoichiometry for complex assembly, which are 2:2:1 [168] or 1:1:1 [167] respectively. When analyzing these ProteoPlex optimized reconstitution mixtures on a gradient, a large excess of E1 remains on top of the gradient, indicating that it was not incorporated into the complex. Furthermore, the complex itself appears in a monodisperse peak in the gradient, indicating a

homogeneous assembly. If less excess of E1 is used, E2 aggregates appear as extra peak in the gradient. Thus, one can hypothesize that the E1 acts as a chaperone for the assembly additional to its catalytic function. However, analyzing the resulting particles in negative stain EM results in a monodisperse field of view. Nevertheless, class averages looked fuzzy and smeared out, suggesting heterogeneity. Also, the average diameter of the reconstituted particles appears to be about 10 % smaller than the natively purified ones. Since the reconstituted PDHc still revealed these structural differences with respect to the endogenous material, further structural investigations were performed on the native preparations.

### 3.3.3 Structural Analysis of *E. coli* PDH

The natively prepared samples yielded a monodisperse field of view in negative stain as well as with cryogenic imaging conditions. An example micrograph is shown in figure 3.25. Class averages were obtained via the classical reference-free alignment approach and the reference free Maximum-likelihood approach (cp. figure 3.25 right).

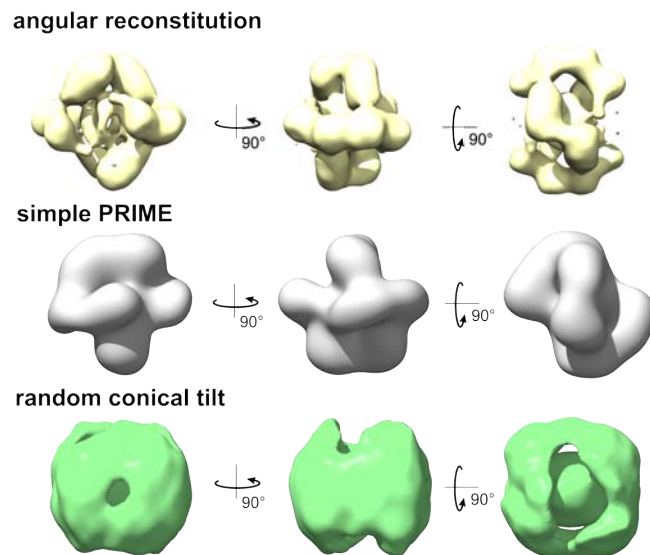


**Figure 3.25: PDHc Raw Data and Initial Class Averages.** Raw data and initial class averages of *E. coli* PDHc are shown. Left: A representative cryo raw micrograph at a magnification of 37000x is shown. Right: typical class averages obtained from the data are depicted. Class averages were obtained using the classical reference free alignment (hereby the particles were down sampled by a factor of 2) and classification approach and via a maximum likelihood (ML) classification approach.

The class averages from both methods do not show distinct, homogeneous densities, except for a solid core density. The shell densities are smeared out and not very clear in most of the averages. This indicates a heterogeneity of the outer shell, which can either

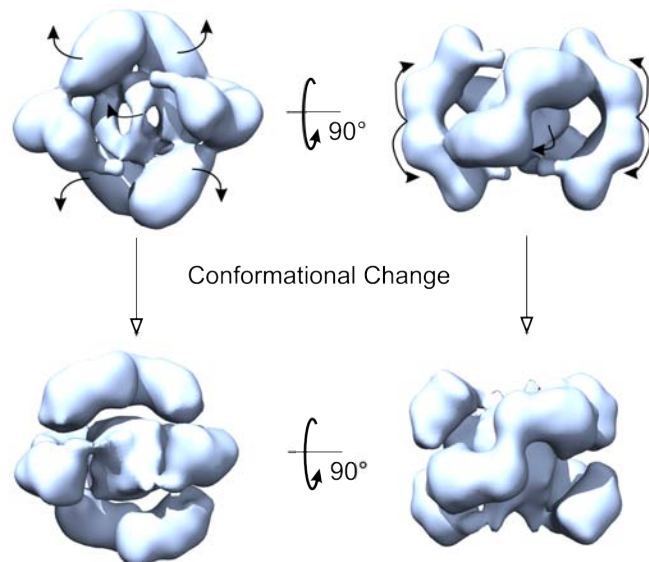
be compositional or conformational. Additionally, it becomes clear that these cannot be explained by an octahedral model. Euler angle determination and reconstruction imposing octahedral symmetry clearly reveals an octahedral core surrounded by a shell of fuzzy density in the 3D structure. This result is incompatible with a defined octahedral symmetry for the entire PDH complex which is, however, the current textbook model for the 3D architecture of the *E.coli* PDHc. Because of these discrepancies more detailed computational analysis was performed to identify the correct symmetry. Three different methods were used to calculate independent initial models from class averages. Firstly, angular reconstitution combined with resampling was used, imposing different symmetry constraints. The D2 symmetry produced the most reliable results. The resampling resulted in several different models. The most reliable one is shown in figure 3.26. Here a cubic core structure is surrounded by six densities with the appearance of dimers. Additionally, *simple PRIME* was used to generate another independent initial 3D model. Here, no symmetry was applied and surprisingly the overall symmetry seemed to be D2 again, and the model resembled the angular reconstitution model very well, with the difference, that no clear core density is recognizable. Lastly, random conical tilt was performed, followed by 3D- maximum likelihood alignment of the resulting volumes. The overall 3D structure rather resembled a cubic structure with an almost octahedral core. However, this structure could neither explain most of the obtained class averages nor the raw particles. This is most likely to the poor 2D classification results obtained from the data.

**Figure 3.26: Initial 3D Models of PDHc.** Three different methods were used to build the initial model. For angular reconstitution class averages were resampled and several hundred individual structures were calculated. Here a D2 symmetry was imposed. The simple PRIME reconstruction was obtained from 250 chosen class averages with good contrast. No symmetry was imposed. 400 Random conical tilt volumes were calculated from 18000 particle images. An average of those volumes is shown after 3D-ML-alignment.



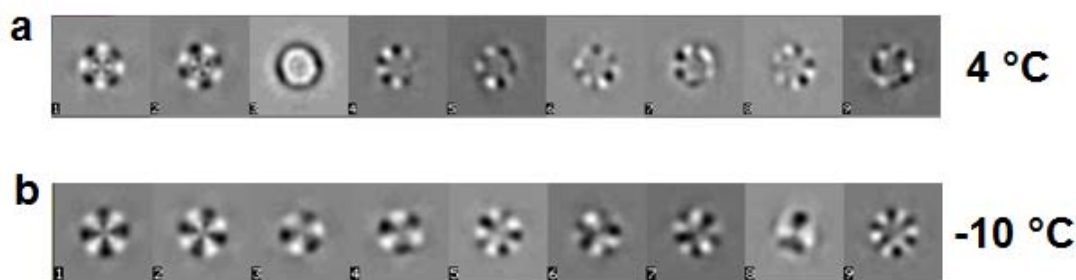
All of the structures were used for refinement against the full dataset. However, none of them proved to be sufficient to give better resolution than 30-40 Å. The reason for that can be found by analyzing the other re-sampling structures. Exemplified two different structures are shown in figure 3.27, which differ significantly while computed reprojections seem to be consistent with the obtained class averages. These observations indicate that

only the core of the PDHc appears to be symmetric and stable while the outer shell reveals a low level of symmetry (if at all) and a high degree of structural dynamics. Moreover, the number of different structures is vast and there seems to be a continuum of different movements on the outer shell. Since core and outer shell seem to move relative



**Figure 3.27: Exemplified Conformational Change of PDHc.** One exemplified detectable conformational change of the PDHc is shown. The two structures were obtained by resampling of initial class averages combined with angular reconstitution.

to each other and the outer shell is also not stable, it is very hard to calculate reliable 3D structures by standard EM methods. Thus, an attempt was made to potentially reduce the dynamics of the outer shell by the newly developed cryoFix method. In both cases 20000 particles were selected, mass centred and eigenimages were calculated (see figure 3.28).



**Figure 3.28: Effect of CryoFix on the PDHc.** The natively purified PDHc treated with standard GraFix (a) and once with cryoFix (b). In both cases 20000 particles were picked, mass centred and eigenimages were calculated. While the central portion of the images is symmetric in both cases, the outer ring only gets symmetric in the -10 °C case.

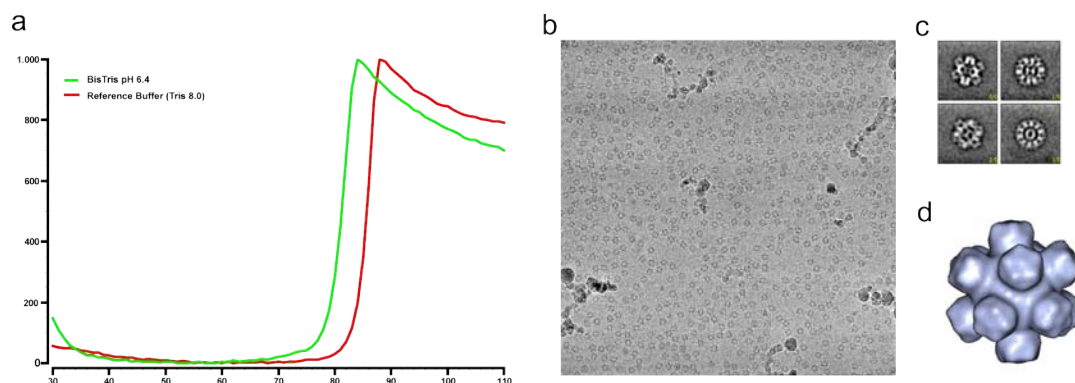
The eigenimages reveal a higher degree of symmetry, indicating a more stable and symmetric arrangement of the outer PDHc shell upon cryo-fixation. While in the 4 °C sample core seems to be more symmetric component, the symmetry expands over the full the structure in the -10 °C sample. This itself is already a success for the cryoFix method, however, the improvement at the level of 2D class averages was still not sufficient to significantly improve the 3D structure.

The major limitation in 3D structure determination can clearly be attributed to the outer shell region of the complex. This can be due to fluctuations in the structure as well as to inherent compositional differences. Structurally and functionally conformational dynamics are more likely, but also compositional heterogeneity cannot be entirely excluded. The currently available techniques for *in silico* purification of heterogeneous datasets are not sufficient to achieve high-resolution structures due to a very high level of variations in the particles. Assuming that the purified PDHc is structurally homogeneous and varies in conformation only, crystallisation may still be possible. Setting up initial crystallisation experiments gave first crystals indeed. Different crystal forms could be reproducibly grown. However, no diffraction pattern was obtained yet, but since the crystal formation requires a homogeneous conformation by default, this could be the way to gain a high-resolution structure of the PDHc.



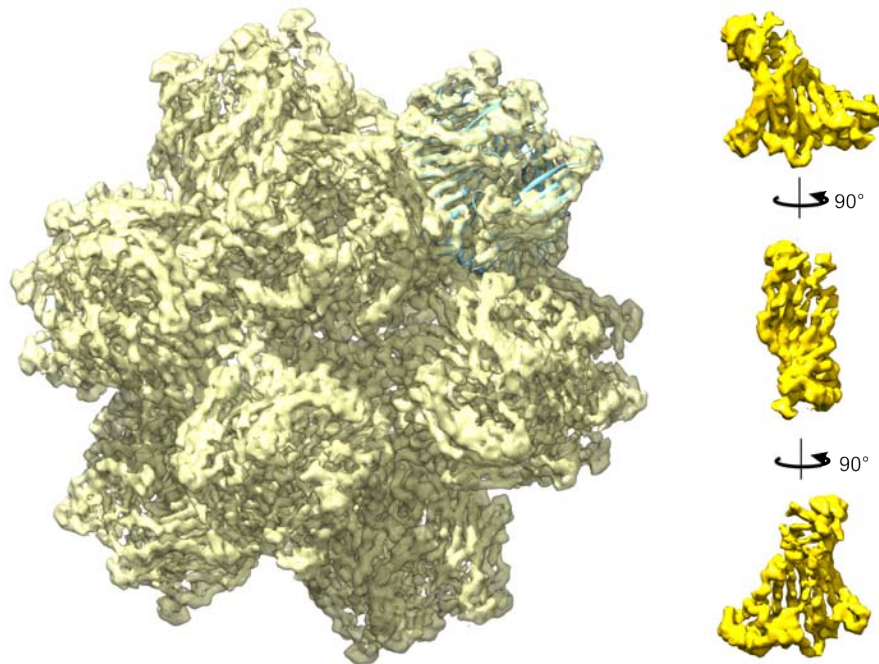
### 3.4 Structure of the *Biomphalaria glabrata* Acetylcholine Binding Protein

In this section, structural changes with low amplitudes are analyzed. Since these are more subtle than the high amplitude movements, higher resolution is required already at the starting point of the analysis. Therefore, I chose an icosahedral object which is known to refine quickly to sub-nanometer resolution. The Acetylcholine Binding Protein from the snail *Biomphalaria glabrata* seemed to be the best model system since the building blocks of the icosahedral molecule can be expressed recombinantly in *E.coli*. The fully assembled complex is smaller than most icosahedral particles but still in an ideal size range for electron microscopy. The sample was obtained from Jürgen Markl's group (University of Mainz).



**Figure 3.29: Initial Analysis of the *Biomphalaria glabrata* Acetyl Cholin Binding Protein.** (a) ProteoPlex unfolding curves of the original buffer and the resulting buffer are shown. A shift of the melting point by 4 °C can be seen. (b) Exemplified cryo-micrograph at a pixel size of 2 Å per pixel is depicted. (c) Representative class averages from the initial analysis are shown. (d) The initial 3D model obtained from the class averages in (c) is shown. Icosahedral symmetry was imposed on the reconstruction.

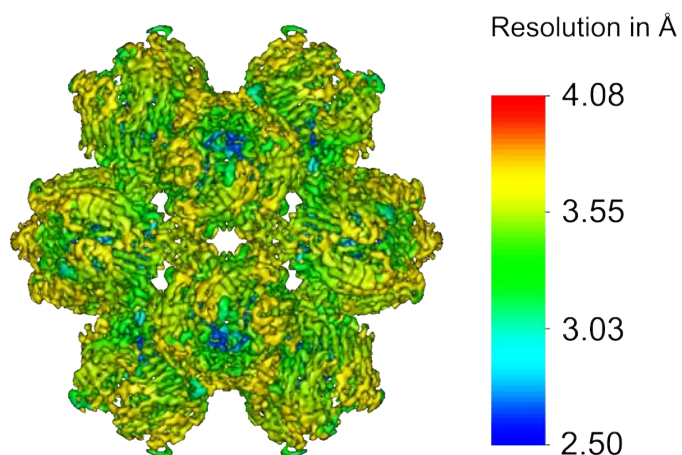
In a first step the buffer was optimized using ProteoPlex (see figure 3.29 a). The original purification buffer was Tris pH 8.0 and resulted in a melting point of 78 °C. The ProteoPlex optimized buffer was BisTris pH 6.8 with a melting point of 82 °C. All melting curves revealed an ideal steepness indicating high enthalpic stability which is most probably due to the icosahedral architecture of the complex. After buffer optimization, the GraFix methodology was applied to stabilize the sample optimally for cryo grid preparation. Figure 3.29 shows exemplified an resulting electron micrograph (b) and reference free class averages (c). Facilitated by the symmetry, a first 3D model was readily obtained via angular reconstitution (cp. figure 3.29d). The structure could be refined to a resolution of 4 Å with only 6000 particles of the 146000 initially recorded particle images. The structure is shown in figure 3.30.



**Figure 3.30: Refined Model of AChBP at 4 Å.** The refined structure of the Acetylcholine Binding protein is shown. The structure refined to 4 Å resolution with icosahedral symmetry imposed. The previously published atomic model (pdb: 4aod) of a penton is fitted. Further, a single asymmetric subunit was segmented out of the volume and is shown from all sides.

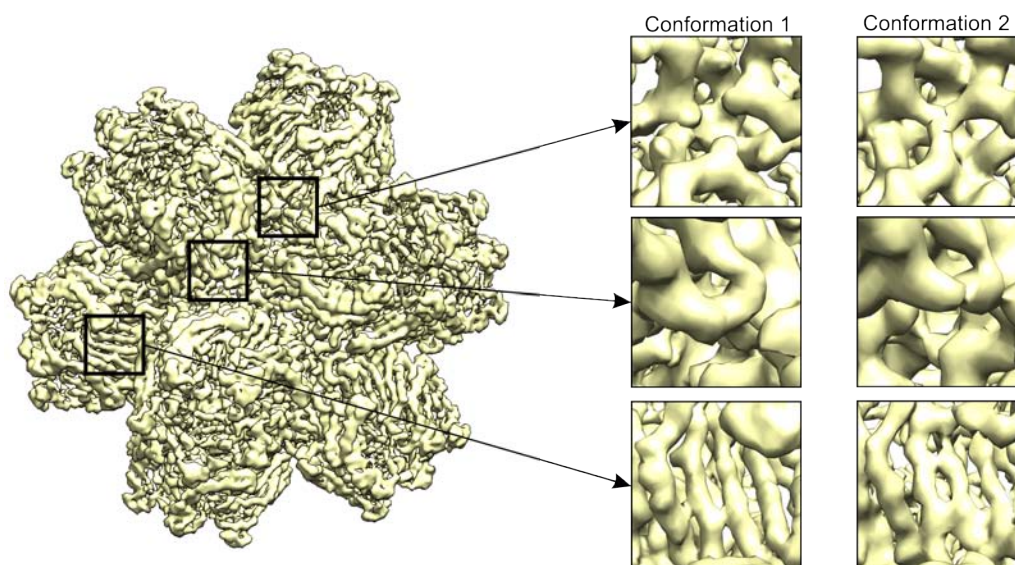
The resolution was further investigated by a local resolution map which is shown in figure 3.31. The inner parts of the pentons are well resolved, reaching a resolution of almost Nyquist frequency. In contrast, the outer parts are significantly less well resolved. At the same position, the structure is anisotropic. One can, thus, hypothesize that the outer parts are rather flexible.

**Figure 3.31: Local Resolution Map of AChBP.** A local resolution map was determined by calculating a local FSC in cowEyes [104]. Since the map was calculated with gold standard refinement, the 0.143 threshold was used. It can be seen that the central part in blue reaches the Nyquist frequency while the distal parts of the structure are less well resolved.



To resolve this structure better, resampling followed by 3D-MSA was performed to calculate models of different conformations. Three different classes could be identified. Those were competitively refined against the dataset and two of them refined well to 4.2 and

3.6 Å resolution. Subsequently, the dataset was split according to which model the particle images fit best, and both sub datasets were refined against the initial low resolution model shown in figure 3.29d. The refined structures are shown in figure 3.32.



**Figure 3.32: Conformational Fluctuations of AChBP.** The originally obtained data set was successfully split in two distinct structures. Conformation 1 was refined to a resolution of 3.6 Å, while conformation 2 reached 4.2 Å resolution. Three positions showing significant differences are magnified.

Besides loop movements, which are fast and frequently occurring fluctuations, one can see contact changes in the beta-sheets that actually usually require substantial amounts of energy. All those differences are independent of ligand binding since the binding pocket is unoccupied in both structures. In conclusion, two structures with very high resolution, which differ minimally but significantly, were obtained from noisy raw images. If indeed these differences are fluctuation in the structure, it is the first time that such small amplitude dynamics were visualized.



# Chapter 4

## Discussion

For a long time, single particle cryo EM was not able to reach the resolution of X-ray crystallography. However, within the last ten years enormous progress was made in improving the resolution by cryo EM. First atomic resolution models were obtained for symmetric objects like viruses [224] and chaperons [227]. This showed that indeed single particle cryo-EM can resolve protein side chain density. Very recently, also the first asymmetric particle reached atomic resolution. Not surprisingly, it was the ribosome [4]. In the cryo-EM field the recent progress in obtaining high resolution is considered to be facilitated by the development of *direct electron detection devices* and a correction for the so-called *beam-induced motion* [111, 64]. The latter is the hypothetical effect by which the specimen moves through melting of its surrounding ice and thus causing drift in the image. Since the newly developed CMOS-based direct detectors record several frames per image, it is possible now to account for this phenomenon. Moreover, those cameras significantly improve through a better detector quantum efficiency and a better SNR. Though, these two effects are arguably significant, it is hardly believable, that they are the primary limiting factors. The initial atomic resolution structures did not need any of these [224, 227]. Thus, other factors must be clearly limiting. This thesis approaches two other possibilities. As for any workflow, errors that arise at the beginning of it are hardly to be revised later. Thus, I am convinced that sample preparation and sample preservation is crucial before any analysis. This aspect is discussed in the light of the gained methodological development in the following section. Once this is assured, the major limiting factor is from my point of view the conformational heterogeneity which will be discussed the subsequent sections.

### 4.1 Methodological Progress

#### 4.1.1 ProteoPlex

Cells are complex in composition and organization. Crowding and micro environments are a major determinants of their chemical structure. This vast complexity has to be

reduced for any controlled laboratory experiment. Not surprisingly, almost any structural study reduces the medium in which the molecule of interest is purified to the absolute minimum — mostly a buffering substance, a mixture of salts, redox and crowding agents. Importantly, due to the complexity of the cell, one single generic buffer solution, that works best for all complexes, does not exist. Each and every molecule needs its own optimized buffer. In crystallography this is well established. As figure 3.11 shows, the range of pH values used for successful crystallization covers almost the entire pH range. In contrast to this broad pH distribution for the PDB entries, the corresponding cryo EM database (EMDB) reveals a very narrow distribution with a clean peak at pH 7.5. This discrepancy is already surprising since both methods work best with optimally stabilized samples.

However, how are the physiologic values? As it can be found in almost any biology textbook [214], pH 7.4-7.5 is considered to be physiologic. This, however, is only true for extracellular fluids like blood. The intracellular pH is considered to be lower with values ranging from 6.5 to 7.2 [30, 43, 147, 222]. Moreover, in contrast to the extracellular medium, the pH within cells is much more heterogeneous in correlation with biochemical activities [30]. The distribution of pH values in the PDB entries even suggests two clearly distinct peaks - one with an optimum around pH 7.5 and one with an optimum at pH 6.5 within the rather broad distribution.

From these considerations, one would expect to have a strong bias towards extra cellular protein structures in the EMDb. This, however, is not the case, but many entries come from the same few molecules - ribosomes, viruses, chaperones, proteasomes and polymerases. These are known to be rather stable and thus could easily withstand non-optimal pH. Almost all other entries do not reach sub-nanometer resolution. Thus, one could hypothesize, that most structural investigation attempts with EM were done without buffer optimization. The reasons for that are twofold. First of all, while structural integrity is a prerequisite for crystallization and screening for good conditions is mandatory to gain any structure, single particle EM can produce incorrect models from any data. The software tools used do not distinguish between intact particles and broken particles. Done in a wrong way, EM methods can even generate a well-defined model from pure noise [85]. This is surprising since the enormous power of EM is that single molecules are distinguishable and can be judged by their appearance. However, comparing the same molecule in different buffer conditions by EM can be difficult, because it is a tedious effort and hardly quantifiable. Until now, no alternative method existed.

Thus, in the first part of my thesis I searched for a new suitable method to screen for an ideal buffer condition for EM samples. This method should be fast, amenable to high throughput, and use only cheap equipment and chemicals. Luckily, in the last decade such a method was already developed. This method termed *ThermoFluor* or *Differential*

*Scanning Fluorometry*, helped to identify buffer conditions suitable for the crystallization of many single domain proteins quickly by measuring and analysing thermal unfolding curves of proteins in different buffer conditions.[20]. However, the literature [140] and even the filed patent on the method [87] disregard its application on larger protein assemblies. The argumentation used is that the proteins within complexes will unfold more or less independently from each other and result thus in superimposed multi-transition curves, meaning curves with several steps visible. Such curves can be indeed obtained for very destabilizing conditions. However, these complex multi transition curves can still be evaluated and explained and do not need to be disregarded. Moreover, even complex samples with many components can result in much simpler curve shapes. I even found at least one condition that showed an apparent single transition curve in the assay for every of the tested 86 complexes. The existence of such cases was also reported recently in the literature [20, 63]. While single transition curves were not expected in the literature for complex samples [140], this curve form can theoretically very well arise from cooperative unfolding of a complex. A fully cooperative unfolding curve can even be analyzed by the originally proposed method. However, one has to be very careful with that since this originally described analysis can be misleading. One can imagine a case, where the individual components of a complex have similar melting temperatures. Thus, the measured signal, which is the superimposition of the individual unfolding curves of the components of a broken sample, can appear as single transition curve. To exclude this apparent cooperativity from real cooperativity in unfolding, a new framework under the name ProteoPlex was developed in this thesis.

Altogether 86 different complexes were tested and highly cooperative unfolding could be found for all of them, even for the largest ones like the PDHc, the ribosome or the spliceosome. Nevertheless, it seems very likely that this behavior is not general for any imaginable protein complex. The antigen-antibody complexes, for instance, used as an argument in the ThermoFluor patent [87] do certainly not unfold in a concerted manner, which led to the stated conclusion. However, one can hypothesize that naturally occurring complexes might have evolved to fold and thus also unfold cooperatively, since this would guaranty a substantial stabilization. Interestingly, even complete proteomes will unfold cooperatively [45]. This makes sense, since unfolded proteins often lead to aggregates that are toxic for cells. In the view of an organism it is favorable rather to destroy the cell, containing unstable and broken proteins, than to keep it alive. The same argumentation holds true for complexes. It is better to destroy the whole complex than to have missing parts.

Nevertheless, the most stabilizing buffer conditions for a given macromolecular complex do not necessarily reflect the conditions of highest functional activity. One may expect a complex to be even more active in a buffered environment that does not reflect conditions

of highest thermodynamic stabilization. Therefore, it was surprising to find that the *E. coli* Pyruvate Dehydrogenase Complex turned out to be more active under stabilizing conditions compared to standard buffer conditions. While a generalization of this statement is impossible because for most of the complexes studied no straightforward functional assays are available, it indicates that for this specific case the highest functional activity and the best structural stability happen to require the same buffer conditions.

If indeed all natural occurring complexes are evolved to fold and unfold cooperatively, we have found a generic method to identify optimal buffer conditions. The capacity of the method itself can even be extended. As Thermofluor itself, several additional use can be found for ProteoPlex. Within this thesis, the potential application towards screening of stabilizing protein interactors and even a complete reconstitution was demonstrated. The only limitations towards analyzable samples are membrane proteins. Since the dye heavily interacts with detergents, extensive background signals emerge. Detergents, however, are necessary to stabilize membrane proteins. Thus, membrane proteins have to be excluded for now.

However, a few words of caution with the newly developed methods have to be made. Even though a quantitative, thermodynamic framework is used within the interpretation of one screen, the obtained thermodynamic constants are only very rough estimates of the real values. Most likely, only their order of magnitude is correct. Thus, measuring of binding constants, as it is possible for Thermofluor, will be hardly accomplishable. This can be especially critical when pipetting the assay by hand. Therefore, in collaboration with FEI Company (Eindhoven, The Netherlands) we developed a fully automated platform, which performs the full screen reproducibly from the purified sample until the identification of the optimized buffer.

From the broad statistics on pH stability optima obtained by ProteoPlex in this thesis, one can conclude that the narrow pH range evident in the EMDB raises some questions about sample stability in present EM structural work by the single particle method. The strong correlation of sample pH with monodispersity and stability of complexes in EM images indicates that single particle EM structures might benefit significantly in resolution and accuracy upon by buffer optimization. In conclusion, by the employment of the sample optimization workflow presented in this thesis, I envisage a substantial boost in the functional understanding and structure determination of molecular machines.

### 4.1.2 Modifications of the GraFix Methodology

A complex in a buffer environment optimized for stability is likely when kept cold. However, the preparation for electron microscopy can introduce additional destabilizing forces.



First of all, the complexes bind to a carbon film which is known to provide sufficient forces to destroy fragile complexes [100]. Moreover, for a cryo grid preparation any sugar has to be removed from the sample to assure good contrast. However, the required removal of sugars is well known to destabilize complexes. To prevent this destabilization from destroying the complex, the GraFix methodology was introduced [100]. This alone was already a valuable tool for many studies [139, 219, 177, 27, 135]. One of the goals of this thesis was to develop the GraFix protocol further to improve the situation also for very dynamic complexes.

The dynamics of complexes are tremendously reduced at lower temperatures. At lower temperatures, the thermodynamically stable complex conformations become most populated, as it was shown for the ribosome [52]. Standard GraFix gradients are already performed at 4 °C. However, since sucrose prevents aqueous solutions from freezing, even lower temperatures could be used during the GraFix run. As in a standard GraFix a cross linking agent is likely to trap the molecules in thermodynamically stabilized states even if the temperature is raised again. For human CRM1 and the *E.coli* PDHc this methodology gave significant changes indeed. No good class averages could be obtained from human CRM1 samples graFixed at 4 °C, while at -10 °C even 3D analysis was achieved. The reason for that might be that at 4 °C CRM1 is so flexible that a vast number of different states is found in the sample thus only a few particles from the whole dataset can be combined to a homogeneous class. At -10 °C, however, the number of different conformations seems to be reduced and thus they can be averaged better to homogeneous class averages and are amenable to 3D analysis. The improvement for the PDHc, however was only detectable in the increased symmetry of initial class averages. Still, this method could be a valuable tool for many molecules.

### 4.1.3 *Chaetomium thermophilum*

Another very similar concept comes from the use of thermophile organisms. Since the doubling times of thermophile organisms are in the range of related mesophile organisms, one can assume that also the protein dynamics are comparable at the living temperature. Thus, a protein from an organisms living at 60 °C is expected to behave at 4 °C as its mesophile homologue at -30 °C. This concept has worked well for many prokaryotic complexes like the ribosome [108] or the ATP-synthetase [115] taken from thermophilic bacteria. However, thermophile eukaryotes were missing for a long time. Only recently, the genome of the thermophile fungus *Chaetomium thermophilum* was sequenced [3]. Starting from a culturing method obtained from the Hurt lab, a fermentation strategy was developed with this thesis. At culturing conditions, the fungus dies very rapidly leaving only cell "ghosts". What exact cell biological effect caused this, was not investigated here.

Nevertheless, shortly cultivated cells were perfectly viable and several endogenous *C. thermophilum* complexes were successfully purified from these cells. Interestingly, thermal stability tests of *C. thermophilum* complexes indicated that melting temperatures were not shifted to higher temperatures to the expected extent. *Chaetomium thermophilum* enriches its cytoplasm with trehalose [95], which is known to protect organisms from harm by very low or very high temperatures. Therefore, there seems no need for thermophile proteins.

Despite this rather unexpected similarity of melting temperatures obtained for mesophile and *C. thermophilum* complexes, the conformational dynamics of *C. thermophilum* complexes as analyzed by EM seem to be much reduced. Such conformational stabilizations are equally important for high resolution structure determination, indicating why *Chaetomium thermophilum* proteins are interesting targets for structural biology.

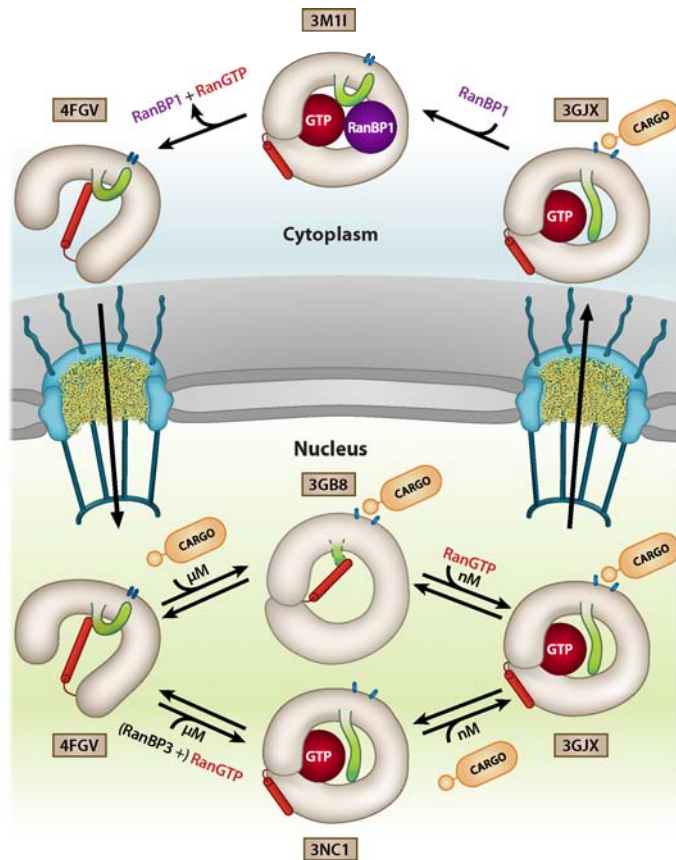
## 4.2 Conformational Landscape of CRM1

CRM1 is one of the major nuclear export receptors in eukaryotic cells. Its cargo spectrum is extremely versatile and ranges from whole ribosome subunits to factors that are only a few kDa in mass. During its duty cycle, it binds the small GTPase Ran and its cargo in the nucleus, diffuses as complex through the NPC and releases both in the cytoplasm. The binding of Ran and cargo was shown to be highly cooperative while either of them alone has only minor affinity for CRM1 [158]. Since the binding spots of cargo (the NES cleft) and Ran (the terminal region) are several nanometers apart, it was not clear how the signal of binding is transferred through the HEAT-repeat structure. To solve this problem, an integrative structural approach, employing X-ray crystallography, small angle X-ray scattering, molecular dynamics simulations and single particle cryo EM, was used. This thesis describes the electron microscopic analysis of free CRM1 mostly. The main result of the EM analysis reveals that CRM1 samples a conformational landscape between a closed ring-shaped and an open superhelical conformation in a ratchet-like motion. The energy landscape between different open conformations is very flat, implying that thermal energy is sufficient to drive this conformational flexibility of CRM1. The closed conformation is only slightly favored and shows a minor stabilization which is, however, still easily breakable by the thermal energy of the surrounding medium. Newly [129] and previously [128] obtained crystal structures of open and modelled closed conformations by the Ficner group (University of Göttingen, Germany) provide structural details of the two conformations at atomic details. As shown in figure 3.19, the C-terminal helix of CRM1 expands to the acidic loop in the open conformation transferring a contact between C-terminus (the binding side of Ran) of CRM1 and the NES-cleft (the cargo binding side). To investigate the contribution of the C-terminal helix, a deletion variant was analyzed. The removal of the C-terminal helix leads to a flat overall energy landscape, where each conformation is almost equally populated. Thus, indeed the backfolded helix leads to an interaction of the termini. This interaction with the acidic loop does not seem to be required to maintain the superhelical conformation. Nevertheless, the interaction as well as the superhelical shape itself sterically hinder the binding of Ran and alter the NES-cleft preferably into a closed conformation.

In conclusion, to bind either of the two interaction partners, CRM1 has to adopt a closed conformation. This, however, requires the protein to be more compact to minimize its flexibility and consequently its entropy. Thus, binding of either of the two interaction partners to CRM1 is energetically unfavoured. After CRM1 bound either of the ligands, the closed conformation is stabilized and the second ligand can easily bind. It can be concluded that the ligands bind rather by a conformational selection mechanism, than

by an induced fit and can be described as unordered bimolecular reaction. A theoretic model is summarized in figure 4.1.

**Figure 4.1: Functional Cycle of CRM1.** The hypothetical functional cycle of CRM1 is shown. CRM1 can associate with either Cargo or Ran-GTP in the nucleus. Both bind with low affinity. Upon binding by either of them binding of the other is facilitated. Thereby, a binding between the acidic loop (green) and the C-terminal helix plays a crucial role. For every stage of this assembly, crystal structures are available under the listed pdb codes. This figure was kindly provided by Dr. Thomas Monecke and Lin-Ta Hsu (University of Göttingen)



By thermal sampling of the conformational space, CRM1 adopts randomly the closed conformation. In this conformation, the C-terminal helix is displaced and the NES-cleft is open. This favors the binding of the two ligands. Upon binding of either of them, sterical hindering prohibits closing of the NES cleft respectively movement of the C-terminal helix. Therefore, CRM1 is stabilized in the closed conformation and the binding of the other ligand is favored. This complex diffuses then in the cytoplasm and both ligands are released from CRM1. The structural reasons for the release have yet to be found in future studies.

All of these findings could be easily obtained from *C.thermophilum* CRM1, while it was not possible to determine an energy landscape for human CRM1 by cryo EM analysis and image processing. In case of human CRM1 the two extreme states of the movement were found only. Even though intermediates of those could not be found, many substates were identified. These involve the interior of CRM1, most likely the C-terminal helix as well as the acidic loop.

After publication of our findings, two further publications confirmed these in 2013. Dian et al. [44] analyzed a C-terminal deletion mutant of human CRM1 more closely and were even able to crystalize it. Their findings, however, are perfectly in line with our interpreta-

---

tions of CRM1's energy landscape. Further, Saito et al. [171] published a crystal structure of yeast CRM1, which also agrees perfectly with the described interpretation.

### 4.3 Structural Investigation on the *E.coli* Pyruvate Dehydrogenase Complex

The pyruvate dehydrogenase complex is one of the pacemaker enzymes in most living cells. It catalyzes the reaction from pyruvate - the final product of glycolysis - to Acetyl-CoA, an intermediate for many anabolic as well as catabolic reactions. Thus, its biochemical significance is documented in all major textbooks [214]. In case of the *E.coli* PDHc, even a quaternary structure can be found in all text books following the publication of Lester Reed and colleagues in the 1960ies [167, 166, 165, 168]. According to the authors, the E2 component forms a cubic core structure which is surrounded by a cubic shell made of the E1 and E3 component. This interpretation is purely based on the visual inspection and interpretation of micrographs from the complex. This, however, has never been verified with modern computational methods and no complete 3D structure is available. One of the goals in this thesis was to determine the PDHc structure by single particle cryo EM. Reassuringly, the calculated class averages resemble well the images obtained by Reed et al.. However, while any kind of symmetry should facilitate the assignment of projection angles to PDHc class averages, imposing octahedral symmetry led to a well-defined E2 core but not to an overall symmetrical structure. The symmetrical E2 core indeed agrees with obtained octahedral crystal structures of the E2 core [123]. However, the shell consisting of E1/E3 is clearly not in agreement with the assumed octahedral symmetry.

To analyze the structure of the shell, three independent initial 3D models were constructed by different methods using the same cryo EM data. Two of them showed features of a D2 symmetry, which was not proposed before. It seems surprising that the symmetries of the outer shell and the core do not match, since this would mean that the numbers of active centers of the individual components do not match. However, symmetry mismatch can also be seen in other complexes like the 26 S proteasome, where a C6 AAA-Protein binds to a D7 core particle [206]. Moreover, in depth analysis of the EM data revealed several possible low-symmetry modules that agree well with the EM raw data at low-resolution. High resolution structure determination was not yet possible probably because of a very large number of possible conformers present in the same sample. With the current EM software these conformational fluctuations cannot yet be computationally separated. Any high resolution PDH structure by cryo-EM will therefore require new methodological developments related to the improvement of computational sorting of cryo EM images. Murphy et al. [136] even suggest an hyper flexible structure with no clearly defined shell. This seems not very likely in the light of evolution since this would slow down a substrate channeling through the three components. The calculated reconstructions in

this thesis always contain a fully connected shell structure which seems to be rather flexible. It remains to be determined how this PDH shell is organized in detail. Since the E1 and the E3 bind to the same position on the E2, there is also the possibility that the observed heterogeneity does not come as a result of a considerable conformational flexibility. It could also equally well explained by a more or less random assembly process with various different stoichiometries. Assuming that this is true and each of the 24 subunit binding domains of the E2 core is bound by a random shell component (E1 or E3) would already give rise to several thousand different possible configurations. This random assembly however, was never seen for any other macromolecular complex so far. It would be certainly entropically favored but give rise to inhomogeneous catalysis results which is unlikely for such an important pacemaker enzyme. Moreover, the fact that crystals were obtained from purified PDH speaks against this hypothesis. Complex mixtures of PDHc in different configurations would hinder crystallization quite tremendously. Thus, a very dynamic nature of the complex is currently the most likely explanation for the failed attempts towards a high resolution structure by cryo EM. Nevertheless, further structural investigations and different approaches like single molecule spectroscopy could shed light on this puzzling molecule.

## 4.4 Conformational Flexibility of the *Biomphalaria glabrata* Acetylcholine binding protein.

For CRM1 tier-0 movements could be successfully analysed (cp figure 1.5). To demonstrate that also tier-1 and 2 movements can be analysed, the icosahedral *Biomphalaria glabrata* Acetylcholine Binding protein was chosen. A 6 Å structure was already available at the beginning of this project [178]. To gain such high resolution, the molecule has to be rather static and thus significant tier-0 movements can be excluded. In a first attempt to solve the structure after sample optimization and imaging in a high-end microscope resulted already in 4 Å resolution. This improvement was possible by optimized buffer conditions using the ProteoPlex methodology in combination with stabilization by GraFix. Of note: Often concerns about the restriction of attainable resolution are raised on conferences (personal communication with Professor Dr. Holger Stark). Here it was shown for the first time that atomic resolution is achievable for a grafixed sample. However, at first 4 Å resolution seemed to be a hard barrier for any refinement attempt. Even using a lower pixel size that is required for higher resolution structures did not help, and these structures also converged at 4 Å resolution. This was a good indication for the sample itself being the resolution limiting factor possibly because of dynamic conformational fluctuations present in the molecules of the dataset. In an attempt to address possible conformational fluctuations resampling techniques and 3D-MSA were used. Thereby, three conformations could be identified. Upon refinement, two of them converged to well defined, and distinct structures and small fluctuations became visible. Mainly loop but also  $\beta$ -sheet rearrangements were found. This is the first time that such small fluctuations could be visualized by cryo EM. The more populated structure even improved to a resolution of 3.6 Å which is already a first validation for the sorting procedure.

Unfortunately, this approach is not able to identify all possible structures and define an energy landscape. Since it is based on stochastic resampling results emerge randomly and not the full spectrum of movements is likely to be identified. Therefore, it is a challenge for the future to develop better methods that are able to analyze all possible movement parameters.



# Chapter 5

## Summary and Outlook

As shown in the recent literature (e.g. [4, 121]) and in this thesis, high resolution structures are obtainable by single particle cryo EM. To reach that, a fully intact sample is the absolute prerequisite, which most likely also the biggest hurdle. However, as also seen in this thesis, even a fully optimized sample can still impair structural analysis by conformation fluctuations. To still be able to gain high resolution information the sample has to be analyzed regarding its conformational energy landscapes comprising of all possible conformations. This is still a demanding challenge for cryo EM. Even though many approaches exist, none of them seems to work generically for any given problem and clearly new developments at all steps of the single particle cryo EM workflow are necessary to aid this analysis.

Here, especially sample optimization techniques were demonstrated to improve the quality of the obtained 3D structure. The newly developed ProteoPlex method is a great advance for structural biology. By stabilizing the sample effectively, it can alter projects from hopeless to promising. This technology can be improved further. For example, membrane proteins can currently not be analyzed, since the dye interacts with the detergent required for stable membrane protein preparations, resulting in a complex background signal that cannot easily be subtracted. Further, more information can be extracted from unfolding experiments. While monodispersity is only inferred through a ProteoPlex analysis, the technology could be extended by directly monitored during the course of the experiment by adding light scattering measurements. Also, the improvements upon GraFix can be further developed to have a complete toolkit for the stabilization of macromolecular complexes. More sophisticated strategies could be used. For instance, using more specific crosslinking agents or a complex mixture would be imaginable. Additionally, the usefulness of the thermophile eukaryote *Chaetomium thermophilum* towards structural biology was demonstrated in this thesis. However, the biology of the fungus is not entirely understood, and even the trait that makes its proteins withstand high temperatures is not clear. More investigations towards a deeper understanding in the fungus itself have certainly to be made.

With this work, the structural determinants of CRM1's allosteric binding of ligands were elucidated. However, there are still structural questions remaining. The direct interaction of CRM1 with the nuclear pore components, for instance, is structurally unknown. While the assembly of CRM1-export complexes seems to be spontaneous, the disassembly is supposed to require cytoplasmic components of the Nuclear Pore Complex such as NUP214. A structural investigation on this might give insights into the disassembly of the complex. Moreover, CRM1 binds an enormous cargo spectrum and most of the structural analysis was made on only a few examples. In particular for large cargoes like ribosomal subunits the binding mode might be different and therefore represents a very attractive target for future studies. According to the current level of understanding, CRM1 moves cargoes similar to a tugboat through the NPC, which is hardly imaginable for larger cargoes.

The pyruvate dehydrogenase complex is still a puzzling object especially since many open questions could not be answered within this thesis. However, the obtained crystals might lead to first structural insights in the near future. Additionally, the soft preparation techniques used for the preparation of the bacterial PDHc can be readily adjusted for the purification of eukaryotic PDHc. While its structure and overall architecture are much better understood, also no high-resolution structure is available. In this case the proven icosahedral symmetry may facilitate this. Also, this should be elucidated.

As the era of atomic resolution structures has begun recently for single particle cryo EM, more structures will emerge, asking for improved tools to handle conformational flexibility. This thesis demonstrated that slight conformational movements in the range of a few angstroms can be resolved by cryo EM. However, to gain insights into a conformational landscape to get physical values, a more systematic approach has to be followed. Modelling aided by normal mode analysis worked well in the case of CRM1 at low resolution. It should be investigated in the future whether it will also improve high-resolution structural investigation.

In summary, single particle cryo EM has emerged as an exciting field of research. With the new developments, it became a competitive method for crystallography and with more technical developments ahead and the developments made in this thesis it might even become the leading method for the structural investigation of large proteins and protein complexes.

# Appendix A

## Abbreviations

**1D** one-dimensional

**2D** two-dimensional

**3D** three-dimensional

**AAA+** ATPases Associated with various cellular Activities

**AChBP** Acetyl Choline Binding Protein

**AFM** Atomic Force Microscope

**ART** algebraic reconstruction technique

**ADP** Adenosine di phosphate

**AMPPCP** 5'-adenylyl ( $\beta$ ,  $\gamma$ -methylene)diphosphonate

**AMPPNP** Adenosine 5' ( $\beta$ ,  $\gamma$ -imido)triphosphate

**APC/C** Anaphase promoting complex /cyclosome

**ATP** Adenosine tri phosphate

**BgHb** *Biomphalaria glabrata* hemoglobin

**CCC** Cross Correlation Coefficient

**CCD** charge coupled device

**CCF** Cross Correlation Function

**CD** Circular Dichroism

**CMOS** complementary metal oxide semiconductors

**CoA** coenzyme A

**CPU** central processing unit

**CPV** Cytoplasmatic Polyhedrosis Virus

**CRIME** CRM1 - Importin  $\beta$  etc.

**CRM1** Chromosome Region Maintenance factor 1

**cryo EM** electron cryo microscopy

**CTF** contrast transfer function

**DDD** direct detection device

**DFT** discrete Fourier transformation

**DHFR** Dihydrofolate reductase

**DNA** Desoxy-Ribonucleic Acid

**DQE** Detector Quantum Efficiency

**E3BP** E3-binding protein

**EDC** 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide)

**EM** electron microscopy

**EMDB** Electron Microscopy Data Bank

**ER** endoplasmatic reticulum

**FAD** Flavin adenine dinucleotide

**FEG** Field Emission Gun

**FEL** Free-Electron Laser

**FRC** Fourier ring correlation

**FRET** Foerster Resonance Energy Transfer

**FSC** Fourier shell correlation

**GFP** green fluorescent protein

**GPU** graphical processing unit

**GraFix** Gradient Fixation

**GST** Gluthathion-S-transferase

**GTP** Guanosine tri phosphate

**HEAT** Huntingtin, elongation factor 3, protein phosphatase 2A and TOR1

**HIV** Human Immunodeficiency Virus

**IR** Infrared

**MD** Molecular Dynamics

---

**ML** Maximum Likelihood

**MRA** multi-reference alignment

**mRNA** messenger Ribonucleic Acid

**MSA** multivariate statistical analysis

**MTF** modulation transfer function

**NAD** Nicotinamide adenine dinucleotide

**NES** nuclear export signal

**NMA** Normal Mode Analysis

**NMR** Nuclear Magnetic Resonance

**NPC** nuclear pore complex

**PCA** principal component analysis

**PDB** Protein Data Bank

**PDHc** Pyruvate Dehydrogenase Complex

**PhCTF** phase contrast transfer function

**PMPI** N-(p-Maleimidophenyl)isocyanate

**PSF** point spread function

**psp A** phage shock protein A

**RCT** random conical tilt

**RISC** RNA induced silencing complex

**RNP** Ribonucleo protein

**RT-PCR** Real time polymerase chain reaction

**SART** simultaneous algebraic reconstruction technique

**SDS-PAGE** sodium dodecyl sulfate- poly acryl amide gel electrophoresis

**SIRT** simultaneous iterative reconstruction technique

**SMCC** Succinimidyl- 4-(N-maleimidomethyl)cyclohexane-1-carboxylate

**SNR** Signal-to-Noise-Ratio

**snRNP** small nuclear Ribonucleic Protein

**SSNR** spectral Signal-to-Noise-Ratio

**TEM** transmission electron microscope

**tRNA** transfer Ribonucleic Acid

**TPP** thiamin diphosphate

**UV-Vis** Ultraviolett - visible light

# Bibliography

- [1] B. Alberts. The cell as a collection of protein machines: preparing the next generation of molecular biologists. *Cell*, 92(3):291–4, Feb. 1998. ISSN 0092-8674. URL <http://www.ncbi.nlm.nih.gov/pubmed/9476889>.
- [2] E. X. Albuquerque, E. F. Pereira, M. Alkondon, and S. W. Rogers. Mammalian Nicotinic Acetylcholine Receptors: From Structure to Function. *Physiol Rev.*, 89(1):73–120, 2009. doi: 10.1152/physrev.00015.2008.Mammalian.
- [3] S. Amlacher, P. Sarges, D. Flemming, V. van Noort, R. Kunze, D. P. Devos, M. Arumugam, P. Bork, and E. Hurt. Insight into structure and assembly of the nuclear pore complex by utilizing the genome of a eukaryotic thermophile. *Cell*, 146(2):277–89, July 2011. ISSN 1097-4172. doi: 10.1016/j.cell.2011.06.039. URL <http://www.sciencedirect.com/science/article/pii/S0092867411007148>.
- [4] A. Amunts, A. Brown, X.-c. Bai, J. L. Ll acer, T. Hussain, P. Emsley, F. Long, G. Murshudov, S. H. W. Scheres, and V. Ramakrishnan. Structure of the Yeast Mitochondrial Large Ribosomal Subunit. *Science*, 343(6178):1485–1489, Mar. 2014. doi: 10.1126/science.1249410. URL <http://www.sciencemag.org/content/343/6178/1485.abstract>.
- [5] M. A. Andrade, C. Petosa, S. I. O’Donoghue, C. W. M uller, and P. Bork. Comparison of ARM and HEAT protein repeats. *Journal of molecular biology*, 309(1):1–18, May 2001. ISSN 0022-2836. doi: 10.1006/jmbi.2001.4624. URL <http://www.sciencedirect.com/science/article/pii/S0022283601946248>.
- [6] A. Ansari, J. Berendzen, S. F. Bowne, H. Frauenfelder, I. E. Iben, T. B. Sauke, E. Shyamsunder, and R. D. Young. Protein states and proteinquakes. *Proceedings of the National Academy of Sciences*, 82(15):5000–5004, Aug. 1985. ISSN 0027-8424. doi: 10.1073/pnas.82.15.5000. URL <http://www.pnas.org/cgi/content/long/82/15/5000>.
- [7] P. Arjunan, N. Nemeria, A. Brunskill, K. Chandrasekhar, M. Sax, Y. Yan, F. Jordan, J. R. Guest, and W. Furey. Structure of the pyruvate dehydrogenase multienzyme complex E1 component from *Escherichia coli* at 1.85   resolution. *Biochemistry*, 41(16):5213–21, Apr. 2002. ISSN 0006-2960. URL <http://www.ncbi.nlm.nih.gov/pubmed/11955070>.

- [8] J.-P. Armache, A. Jarasch, A. M. Anger, E. Villa, T. Becker, S. Bhushan, F. Jossinet, M. Habeck, G. Dindar, S. Franckenberg, V. Marquez, T. Mielke, M. Thomm, O. Berninghausen, B. Beatrix, J. Söding, E. Westhof, D. N. Wilson, and R. Beckmann. Cryo-EM structure and rRNA model of a translating eukaryotic 80S ribosome at 5.5-Å resolution. *Proceedings of the National Academy of Sciences of the United States of America*, 107(46):19748–53, Nov. 2010. ISSN 1091-6490. doi: 10.1073/pnas.1009999107. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2993355&tool=pmcentrez&rendertype=abstract>.
- [9] R. D. Astumian and P. Hänggi. B ROWNIAN. *Physics today*, pages 33–39, 2002.
- [10] R. H. Austin, K. W. Beeson, L. Eisenstein, H. Frauenfelder, and I. C. Gunsalus. Dynamics of ligand binding to myoglobin. *Biochemistry*, 14(24):5355–73, Dec. 1975. ISSN 0006-2960. URL <http://www.ncbi.nlm.nih.gov/pubmed/1191643>.
- [11] X.-C. Bai, I. S. Fernandez, G. McMullan, and S. H. Scheres. Ribosome structures to near-atomic resolution from thirty thousand cryo-EM particles. *eLife*, 2:e00461, Jan. 2013. ISSN 2050-084X. doi: 10.7554/eLife.00461. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3576727&tool=pmcentrez&rendertype=abstract>.
- [12] T. S. Baker and R. Henderson. Electron cryomicroscopy of biological macromolecules. In *International Tables for Crystallography*, volume F, chapter 19.6, pages 593–614. 2012.
- [13] G. Banks, I. Kemenes, M. Schofield, M. O’Shea, and S. a. Korneev. Acetylcholine binding protein of mollusks is unlikely to act as a regulator of cholinergic neurotransmission at neurite-neurite synaptic sites in vivo. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*, 23(9):3030–6, Sept. 2009. ISSN 1530-6860. doi: 10.1096/fj.08-117135. URL <http://www.ncbi.nlm.nih.gov/pubmed/19395478>.
- [14] D. L. Bates, M. J. Danson, G. Hale, E. A. Hooper, and R. N. Perham. Self-assembly and catalytic activity of the pyruvate dehydrogenase multienzyme complex from *Escherichia coli*. *Nature*, 268:313–316, 1977.
- [15] D. L. Beveridge and F. M. DiCapua. Free energy via molecular simulation: applications to chemical and biomolecular systems. *Annual review of biophysics and biophysical chemistry*, 18:431–92, Jan. 1989. ISSN 0883-9182. doi: 10.1146/annurev.bb.18.060189.002243. URL <http://www.ncbi.nlm.nih.gov/pubmed/2660832>.
- [16] D. M. Bleile, P. Munk, R. M. Oliver, and L. J. Reed. Subunit structure of dihydrolipoyl transacetylase component of pyruvate dehydrogenase complex from *Es-*



- cherichia coli. *Proceedings of the National Academy of Sciences*, 76(9):4385–4389, 1979.
- [17] L. V. Bock, C. Blau, G. F. Schröder, I. I. Davydov, N. Fischer, H. Stark, M. V. Rodnina, A. C. Vaiana, and H. Grubmüller. Energy barriers and driving forces in tRNA translocation through the ribosome. *Nature structural & molecular biology*, 20(12):1390–6, Dec. 2013. ISSN 1545-9985. doi: 10.1038/nsmb.2690. URL <http://www.ncbi.nlm.nih.gov/pubmed/24186064>.
- [18] D. D. Boehr, D. McElheny, H. J. Dyson, and P. E. Wright. The dynamic energy landscape of dihydrofolate reductase catalysis. *Science (New York, N.Y.)*, 313(5793):1638–42, Sept. 2006. ISSN 1095-9203. doi: 10.1126/science.1130258. URL <http://www.ncbi.nlm.nih.gov/pubmed/16973882>.
- [19] E. J. Boekema, J. A. Berden, and M. G. van Heel. Structure of mitochondrial F1-ATPase studied by electron microscopy and image processing. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 851(3):353–360, Oct. 1986. ISSN 00052728. doi: 10.1016/0005-2728(86)90071-X. URL <http://www.sciencedirect.com/science/article/pii/000527288690071X>.
- [20] S. Boivin, S. Kozak, and R. Meijers. Optimization of protein purification and characterization using Thermofluor screens. *Protein expression and purification*, 91(2):192–206, Oct. 2013. ISSN 1096-0279. doi: 10.1016/j.pep.2013.08.002. URL <http://www.sciencedirect.com/science/article/pii/S1046592813001459>.
- [21] L. Borland and M. van Heel. Classification of image data in conjugate representation spaces. *Journal of the Optical Society of America A*, 7(4):601, Apr. 1990. ISSN 1084-7529. doi: 10.1364/JOSAA.7.000601. URL <http://www.opticsinfobase.org/abstract.cfm?URI=josaa-7-4-601>.
- [22] R. Bracewell. Strip integration in radio astronomy. *Australian Journal of Physics*, 1956. URL <http://www.publish.csiro.au/?paper=PH560198>.
- [23] R. N. Bracewell. Numerical transforms. *Science (New York, N.Y.)*, 248(4956):697–704, May 1990. ISSN 0036-8075. doi: 10.1126/science.248.4956.697. URL <http://www.ncbi.nlm.nih.gov/pubmed/17812072>.
- [24] M. M. Bradford. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry*, 72(1-2):248–254, May 1976. ISSN 00032697. doi: 10.1016/0003-2697(76)90527-3. URL <http://www.sciencedirect.com/science/article/pii/0003269776905273>.

- [25] E. J. Brignole, S. Smith, and F. J. Asturias. Conformational flexibility of metazoan fatty acid synthase enables catalysis. *Nature structural & molecular biology*, 16(2):190–7, Feb. 2009. ISSN 1545-9985. doi: 10.1038/nsmb.1532. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2653270&tool=pmcentrez&rendertype=abstract>.
- [26] J. Brink, S. J. Ludtke, Y. Kong, S. J. Wakil, J. Ma, and W. Chiu. Experimental verification of conformational variation of human fatty acid synthase as predicted by normal mode analysis. *Structure (London, England : 1993)*, 12(2):185–91, Feb. 2004. ISSN 0969-2126. doi: 10.1016/j.str.2004.01.015. URL <http://www.sciencedirect.com/science/article/pii/S0969212604000036>.
- [27] C. Bröcker, A. Kuhlee, C. Gatsogiannis, H. J. K. Balderhaar, C. Hönscher, S. Engelbrecht-Vandré, C. Ungermann, and S. Raunser. Molecular architecture of the multisubunit homotypic fusion and vacuole protein sorting (HOPS) tethering complex. *Proceedings of the National Academy of Sciences of the United States of America*, 109(6):1991–6, Feb. 2012. ISSN 1091-6490. doi: 10.1073/pnas.1117797109. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3277535&tool=pmcentrez&rendertype=abstract>.
- [28] B. Busche. *New Algorithms for Automated Processing of Electronmicroscopic Images*. PhD thesis, Georg-August University Goettingen, 2013.
- [29] C. Bustamante, W. Cheng, Y. X. Meija, and Y. X. Meija. Revisiting the central dogma one molecule at a time. *Cell*, 144(4):480–97, Feb. 2011. ISSN 1097-4172. doi: 10.1016/j.cell.2011.01.033. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3063003&tool=pmcentrez&rendertype=abstract>.
- [30] J. R. Casey, S. Grinstein, and J. Orłowski. Sensors and regulators of intracellular pH. *Nature reviews. Molecular cell biology*, 11(1):50–61, Jan. 2010. ISSN 1471-0080. doi: 10.1038/nrm2820. URL <http://www.ncbi.nlm.nih.gov/pubmed/19997129>.
- [31] K. Chandrasekhar, J. Wang, P. Arjunan, M. Sax, Y.-H. Park, N. S. Nemeria, S. Kumaran, J. Song, F. Jordan, and W. Furey. Insight to the interaction of the dihydrolipoamide acetyltransferase (E2) core with the peripheral components in the Escherichia coli pyruvate dehydrogenase complex via multifaceted structural approaches. *The Journal of biological chemistry*, 288(21):15402–17, May 2013. ISSN 1083-351X. doi: 10.1074/jbc.M113.466789. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3663558&tool=pmcentrez&rendertype=abstract>.
- [32] B. Chen, M. Doucleff, D. E. Wemmer, S. De Carlo, H. H. Huang, E. Nogales, T. R. Hoover, E. Kondrashkina, L. Guo, and B. T. Nixon. ATP ground- and

- transition states of bacterial enhancer binding AAA+ ATPases support complex formation with their target protein, sigma54. *Structure (London, England : 1993)*, 15(4):429–40, Apr. 2007. ISSN 0969-2126. doi: 10.1016/j.str.2007.02.007. URL <http://www.sciencedirect.com/science/article/pii/S0969212607001062>.
- [33] J. Z. Chen and N. Grigorieff. SIGNATURE: a single-particle selection system for molecular electron microscopy. *Journal of structural biology*, 157(1):168–73, Jan. 2007. ISSN 1047-8477. doi: 10.1016/j.jsb.2006.06.001. URL <http://www.ncbi.nlm.nih.gov/pubmed/16870473>.
- [34] J. Christodoulou, G. Larsson, P. Fucini, S. R. Connell, T. A. Pertinhez, C. L. Hanson, C. Redfield, K. H. Nierhaus, C. V. Robinson, J. Schleucher, and C. M. Dobson. Heteronuclear NMR investigations of dynamic regions of intact *Escherichia coli* ribosomes. *Proceedings of the National Academy of Sciences of the United States of America*, 101(30):10949–54, July 2004. ISSN 0027-8424. doi: 10.1073/pnas.0400928101. URL <http://www.pnas.org/cgi/content/long/101/30/10949>.
- [35] E. M. Ciszak, L. G. Korotchkina, P. M. Dominiak, S. Sidhu, and M. S. Patel. Structural basis for flip-flop action of thiamin pyrophosphate-dependent enzymes revealed by human pyruvate dehydrogenase. *The Journal of biological chemistry*, 278(23):21240–6, June 2003. ISSN 0021-9258. doi: 10.1074/jbc.M300339200. URL <http://www.ncbi.nlm.nih.gov/pubmed/12651851>.
- [36] E. M. Ciszak, A. Makal, Y. S. Hong, A. K. Vettaikkorumakankauv, L. G. Korotchkina, and M. S. Patel. How dihydrolipoamide dehydrogenase-binding protein binds dihydrolipoamide dehydrogenase in the human pyruvate dehydrogenase complex. *The Journal of biological chemistry*, 281(1):648–55, Jan. 2006. ISSN 0021-9258. doi: 10.1074/jbc.M507850200. URL <http://www.ncbi.nlm.nih.gov/pubmed/16263718>.
- [37] D. K. Clare, D. Vasishtan, S. Stagg, J. Quispe, G. W. Farr, M. Topf, A. L. Horwich, and H. R. Saibil. ATP-triggered conformational changes delineate substrate-binding and -folding mechanics of the GroEL chaperonin. *Cell*, 149(1):113–23, Mar. 2012. ISSN 1097-4172. doi: 10.1016/j.cell.2012.02.047. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3326522&tool=pmcentrez&rendertype=abstract>.
- [38] E. Conti, C. W. Müller, and M. Stewart. Karyopherin flexibility in nucleocytoplasmic transport. *Current opinion in structural biology*, 16(2):237–44, Apr. 2006. ISSN 0959-440X. doi: 10.1016/j.sbi.2006.03.010. URL <http://www.sciencedirect.com/science/article/pii/S0959440X06000480>.

- [39] A. Cook, F. Bono, M. Jinek, and E. Conti. Structural biology of nucleocytoplasmic transport. *Annual review of biochemistry*, 76:647–71, Jan. 2007. ISSN 0066-4154. doi: 10.1146/annurev.biochem.76.052705.161529. URL <http://www.ncbi.nlm.nih.gov/pubmed/17506639>.
- [40] J. M. de la Rosa-Trevín, J. Otón, R. Marabini, A. Zaldívar, J. Vargas, J. M. Carazo, and C. O. S. Sorzano. Xmipp 3.0: an improved software suite for image processing in electron microscopy. *Journal of structural biology*, 184(2):321–8, Nov. 2013. ISSN 1095-8657. doi: 10.1016/j.jsb.2013.09.015. URL <http://www.sciencedirect.com/science/article/pii/S1047847713002566>.
- [41] D. de Rosier and A. Klug. Reconstruction of three dimensional structures from electron micrographs. *Nature*, 1968. URL <http://europepmc.org/abstract/MED/23610788>.
- [42] B. DeLaBarre and A. T. Brunger. Nucleotide dependent motion and mechanism of action of p97/VCP. *Journal of molecular biology*, 347(2):437–52, Mar. 2005. ISSN 0022-2836. doi: 10.1016/j.jmb.2005.01.060. URL <http://www.ncbi.nlm.nih.gov/pubmed/15740751>.
- [43] C. Deutsch, J. S. Taylor, and D. F. Wilson. Regulation of intracellular pH by human peripheral blood lymphocytes as measured by  $^{19}\text{F}$  NMR. *Proceedings of the National Academy of Sciences of the United States of America*, 79(24):7944–8, Dec. 1982. ISSN 0027-8424. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=347466&tool=pmcentrez&rendertype=abstract>.
- [44] C. Dian, F. Bernaudat, K. Langer, M. F. Oliva, M. Fornerod, G. Schoehn, C. W. Müller, and C. Petosa. Structure of a truncation mutant of the nuclear export factor CRM1 provides insights into the auto-inhibitory role of its C-terminal helix. *Structure (London, England : 1993)*, 21(8):1338–49, Aug. 2013. ISSN 1878-4186. doi: 10.1016/j.str.2013.06.003. URL <http://www.sciencedirect.com/science/article/pii/S0969212613002025>.
- [45] K. a. Dill, K. Ghosh, and J. D. Schmit. Physical limits of cells and proteomes. *Proceedings of the National Academy of Sciences of the United States of America*, 108(44):17876–82, Nov. 2011. ISSN 1091-6490. doi: 10.1073/pnas.1114477108. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3207669&tool=pmcentrez&rendertype=abstract>.
- [46] N. Dölker, C. E. Blanchet, B. Voß, D. Haselbach, C. Kappel, T. Monecke, D. I. Svergun, H. Stark, R. Ficner, U. Zachariae, H. Grubmüller, and A. Dickmanns. Structural determinants and mechanism of mammalian CRM1 allostery. *Structure (London, England : 1993)*, 21(8):1350–60, Aug. 2013. ISSN 1878-4186. doi: 10.1016/

- j.str.2013.05.015. URL <http://www.sciencedirect.com/science/article/pii/S0969212613001974>.
- [47] P. Dube, P. Tavares, R. Lurz, and M. van Heel. The portal protein of bacteriophage SPP1: a DNA pump with 13-fold symmetry. *The EMBO journal*, 12(4):1303–9, Apr. 1993. ISSN 0261-4189. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=413341&tool=pmcentrez&rendertype=abstract>.
- [48] N. Elad, D. K. Clare, H. R. Saibil, and E. V. Orlova. Detection and separation of heterogeneity in molecular complexes by statistical analysis of their two-dimensional projections. *Journal of structural biology*, 162(1):108–20, Apr. 2008. ISSN 1095-8657. doi: 10.1016/j.jsb.2007.11.007. URL <http://www.ncbi.nlm.nih.gov/pubmed/18166488>.
- [49] H. Elmlund, D. Elmlund, and S. Bengio. PRIME: probabilistic initial 3D model generation for single-particle cryo-electron microscopy. *Structure (London, England : 1993)*, 21(8):1299–306, Aug. 2013. ISSN 1878-4186. doi: 10.1016/j.str.2013.07.002. URL <http://www.sciencedirect.com/science/article/pii/S0969212613002505>.
- [50] A. R. Faruqi and R. Henderson. Electronic detectors for electron microscopy. *Current opinion in structural biology*, 17(5):549–55, Oct. 2007. ISSN 0959-440X. doi: 10.1016/j.sbi.2007.08.014. URL <http://www.sciencedirect.com/science/article/pii/S0959440X07001212>.
- [51] D. Finley. Recognition and processing of ubiquitin-protein conjugates by the proteasome. *Annual review of biochemistry*, 78:477–513, Jan. 2009. ISSN 1545-4509. doi: 10.1146/annurev.biochem.78.081507.101607. URL <http://www.ncbi.nlm.nih.gov/pubmed/19489727>.
- [52] N. Fischer, A. L. Konevega, W. Wintermeyer, M. V. Rodnina, and H. Stark. Ribosome dynamics and tRNA movement by time-resolved electron cryomicroscopy. *Nature*, 466(7304):329–333, 2010. ISSN 0028-0836. doi: 10.1038/nature09206. URL <http://dx.doi.org/10.1038/nature09206>.
- [53] R. a. Forties and M. D. Wang. Discovering the Power of Single Molecules. *Cell*, 157(1):4–7, Mar. 2014. ISSN 1097-4172. doi: 10.1016/j.cell.2014.02.011. URL <http://www.ncbi.nlm.nih.gov/pubmed/24679521>.
- [54] J. Frank. Classification of macromolecular assemblies studied as 'single particles'. *Quarterly reviews of biophysics*, 23(3):281–329, Aug. 1990. ISSN 0033-5835. URL <http://www.ncbi.nlm.nih.gov/pubmed/2204955>.

- [55] J. Frank. *Three-Dimensional Electron Microscopy of Macromolecular Assemblies*. Oxford University Press, 2nd edition, 2006. ISBN 9780195182187.
- [56] H. Frauenfelder, S. Sligar, and P. Wolynes. The energy landscapes and motions of proteins. *Science*, 254:1598–1603, 1991. URL <http://www.sciencemag.org/content/254/5038/1598.short>.
- [57] J. J. Frye, N. G. Brown, G. Petzold, E. R. Watson, C. R. R. Grace, A. Nourse, M. a. Jarvis, R. W. Kriwacki, J.-M. Peters, H. Stark, and B. a. Schulman. Electron microscopy structure of human APC/C(CDH1)-EMI1 reveals multimodal mechanism of E3 ligase shutdown. *Nature structural & molecular biology*, 20(7):827–35, July 2013. ISSN 1545-9985. doi: 10.1038/nsmb.2593. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3742808&tool=pmcentrez&rendertype=abstract>.
- [58] J. Fu, H. Gao, and J. Frank. Unsupervised classification of single particles by cluster tracking in multi-dimensional space. *Journal of structural biology*, 157(1): 226–39, Jan. 2007. ISSN 1047-8477. doi: 10.1016/j.jsb.2006.06.012. URL <http://www.sciencedirect.com/science/article/pii/S1047847706002000>.
- [59] I. S. Gabashvili, R. K. Agrawal, C. M. Spahn, R. A. Grassucci, D. I. Svergun, J. Frank, and P. Penczek. Solution Structure of the E. coli 70S Ribosome at 11.5 Å Resolution. *Cell*, 100(5):537–549, Mar. 2000. ISSN 00928674. doi: 10.1016/S0092-8674(00)80690-X. URL <http://www.sciencedirect.com/science/article/pii/S009286740080690X>.
- [60] H. Gao, M. Valle, M. n. Ehrenberg, and J. Frank. Dynamics of EF-G interaction with the ribosome explored by classification of a heterogeneous cryo-EM dataset. *Journal of structural biology*, 147(3):283–90, Sept. 2004. ISSN 1047-8477. doi: 10.1016/j.jsb.2004.02.008. URL <http://www.sciencedirect.com/science/article/pii/S1047847704000437>.
- [61] Y.-G. Gao, M. Selmer, C. M. Dunham, A. Weixlbaumer, A. C. Kelley, and V. Ramakrishnan. The structure of the ribosome with elongation factor G trapped in the posttranslocational state. *Science (New York, N.Y.)*, 326(5953):694–9, Oct. 2009. ISSN 1095-9203. doi: 10.1126/science.1179709. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3763468&tool=pmcentrez&rendertype=abstract>.
- [62] A.-C. Gavin, P. Aloy, P. Grandi, R. Krause, M. Boesche, M. Marzioch, C. Rau, L. J. Jensen, S. Bastuck, B. Dümpelfeld, A. Edelmann, M.-A. Heurtier, V. Hoffman, C. Hoefert, K. Klein, M. Hudak, A.-M. Michon, M. Schelder, M. Schirle, M. Remor, T. Rudi, S. Hooper, A. Bauer, T. Bouwmeester, G. Casari, G. Drewes, G. Neubauer,

- J. M. Rick, B. Kuster, P. Bork, R. B. Russell, and G. Superti-Furga. Proteome survey reveals modularity of the yeast cell machinery. *Nature*, 440(7084):631–6, Mar. 2006. ISSN 1476-4687. doi: 10.1038/nature04532. URL <http://www.ncbi.nlm.nih.gov/pubmed/16429126>.
- [63] T. W. Geders, K. Gustafson, and B. C. Finzel. Use of differential scanning fluorimetry to optimize the purification and crystallization of PLP-dependent enzymes. *Acta crystallographica. Section F, Structural biology and crystallization communications*, 68(Pt 5):596–600, May 2012. ISSN 1744-3091. doi: 10.1107/S1744309112012912. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3374521&tool=pmcentrez&rendertype=abstract>.
- [64] R. M. Glaeser and R. J. Hall. Reaching the information limit in cryo-EM of biological macromolecules: experimental aspects. *Biophysical journal*, 100(10):2331–7, May 2011. ISSN 1542-0086. doi: 10.1016/j.bpj.2011.04.018. URL <http://www.sciencedirect.com/science/article/pii/S0006349511004619>.
- [65] M. M. Golas, B. Sander, C. L. Will, R. Lührmann, and H. Stark. Major conformational change in the complex SF3b upon integration into the spliceosomal U11/U12 di-snRNP as revealed by electron cryomicroscopy. *Molecular cell*, 17(6):869–83, Mar. 2005. ISSN 1097-2765. doi: 10.1016/j.molcel.2005.02.016. URL <http://www.sciencedirect.com/science/article/pii/S1097276505011159>.
- [66] A. Goldstein, D. B. Goldstein, and L. I. Lowney. Protein synthesis at 0°C in *Escherichia coli*. *Journal of Molecular Biology*, 9(1):213–235, July 1964. ISSN 00222836. doi: 10.1016/S0022-2836(64)80102-9. URL <http://www.sciencedirect.com/science/article/pii/S0022283664801029>.
- [67] D. Görlich and U. Kutay. Transport between the cell nucleus and the cytoplasm. *Annual review of cell and developmental . . .*, 1999. URL <http://www.annualreviews.org/doi/abs/10.1146/annurev.cellbio.15.1.607>.
- [68] A. D. Gounaris and L. P. Hager. Resolution of the *Escherichia coli* Pyruvate Dehydrogenase Complex. *jbc*, 236(4):1013–1018, 1961.
- [69] J. D. Green, E. D. Laue, R. N. Perham, S. T. Ali, and J. R. Guest. Three-dimensional structure of a lipoyl domain from the dihydrolipoyl acetyltransferase component of the pyruvate dehydrogenase multienzyme complex of *Escherichia coli*. *Journal of Molecular Biology*, 248(2):328–343, Jan. 1995. ISSN 00222836. doi: 10.1016/S0022-2836(95)80054-9. URL <http://www.sciencedirect.com/science/article/pii/S0022283695800549>.

- [70] N. Grigorieff. FREALIGN: high-resolution refinement of single particle structures. *Journal of structural biology*, 157(1):117–25, Jan. 2007. ISSN 1047-8477. doi: 10.1016/j.jsb.2006.05.004. URL <http://www.ncbi.nlm.nih.gov/pubmed/16828314>.
- [71] N. Grigorieff and S. C. Harrison. Near-atomic resolution reconstructions of icosahedral viruses from electron cryo-microscopy. *Current opinion in structural biology*, 21(2):265–73, Apr. 2011. ISSN 1879-033X. doi: 10.1016/j.sbi.2011.01.008. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3088881&tool=pmcentrez&rendertype=abstract>.
- [72] C. Grimm, A. Chari, J.-P. Pelz, J. Kuper, C. Kisker, K. Diederichs, H. Stark, H. Schindelin, and U. Fischer. Structural basis of assembly chaperone-mediated snRNP formation. *Molecular cell*, 49(4):692–703, Mar. 2013. ISSN 1097-4164. doi: 10.1016/j.molcel.2012.12.009. URL <http://www.sciencedirect.com/science/article/pii/S1097276512010180>.
- [73] B. L. D. Groot and H. Grubmüller. Water Permeation Across Biological Membranes : Mechanism and Dynamics of Aquaporin-1 and GlpF. *Science*, 294:2353–2357, 2001. doi: 10.1126/science.1066115.
- [74] Y. Gu, Z. H. Zhou, D. B. McCarthy, L. J. Reed, and J. K. Stoops. 3D electron microscopy reveals the variable deposition and protein dynamics of the peripheral pyruvate dehydrogenase component about the core. *Proceedings of the National Academy of Sciences of the United States of America*, 100(12):7015–20, June 2003. ISSN 0027-8424. doi: 10.1073/pnas.0732060100. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=165822&tool=pmcentrez&rendertype=abstract>.
- [75] T. Guan and R. Kehlenbach. Nup50, a nucleoplasmically oriented nucleoporin with a role in nuclear protein export. *...and cellular biology*, 20(15):5619–5630, 2000. doi: 10.1128/MCB.20.15.5619-5630.2000.Updated. URL <http://mcb.asm.org/content/20/15/5619.short>.
- [76] M. Haider, S. Uhlemann, E. Schwan, and H. Rose. Electron microscopy image enhanced. *Nature*, 392(April):5–6, 1998. URL <http://adsabs.harvard.edu/abs/1998Natur.392..768H>.
- [77] J. Hajdu, R. Neutze, T. Sjögren, K. Edman, a. Szöke, R. C. Wilmouth, and C. M. Wilmot. Analyzing protein functions in four dimensions. *Nature structural biology*, 7(11):1006–12, Nov. 2000. ISSN 1072-8368. doi: 10.1038/80911. URL <http://www.ncbi.nlm.nih.gov/pubmed/11062553>.
- [78] R. J. Hall, B. Siridechadilok, and E. Nogales. Cross-correlation of common lines: a novel approach for single-particle reconstruction of a structure containing a flexible



- domain. *Journal of structural biology*, 159(3):474–82, Sept. 2007. ISSN 1047-8477. doi: 10.1016/j.jsb.2007.05.007. URL <http://www.sciencedirect.com/science/article/pii/S1047847707001360>.
- [79] P. I. Hanson and S. W. Whiteheart. AAA+ proteins: have engine, will work. *Nature reviews. Molecular cell biology*, 6(7):519–29, July 2005. ISSN 1471-0072. doi: 10.1038/nrm1684. URL <http://www.ncbi.nlm.nih.gov/pubmed/16072036>.
- [80] G. Harauz and M. van Heel. Exact algorithms for general geometry three dimensional reconstruction. *Optik*, 78(4):146–156, 1986. URL [http://sbio.uct.ac.za/wp-content/uploads/2013/11/optik\\_73\\_4\\_1986.pdf](http://sbio.uct.ac.za/wp-content/uploads/2013/11/optik_73_4_1986.pdf).
- [81] S. Harper and D. W. Speicher. Detection of proteins on blot membranes. *Current protocols in protein science / editorial board, John E. Coligan ... [et al.]*, Chapter 10 (1995):Unit 10.8, May 2001. ISSN 1934-3663. doi: 10.1002/0471140864.ps1008s00. URL <http://www.ncbi.nlm.nih.gov/pubmed/19016450>.
- [82] B. C. Heisen. *New Algorithms for Macromolecular Structure Determination*. PhD thesis, Georg-August-University Goettingen, 2009.
- [83] R. Henderson. Image contrast in high-resolution electron microscopy of biological macromolecules: TMV in ice. *Ultramicroscopy*, 46(1-4):1–18, Oct. 1992. ISSN 03043991. doi: 10.1016/0304-3991(92)90003-3. URL <http://www.sciencedirect.com/science/article/pii/0304399192900033>.
- [84] R. Henderson. The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules. *Quarterly reviews of biophysics*, 28(2):171–93, May 1995. ISSN 0033-5835. URL <http://www.ncbi.nlm.nih.gov/pubmed/7568675>.
- [85] R. Henderson. Avoiding the pitfalls of single particle cryo-electron microscopy: Einstein from noise. *Proceedings of the National Academy of Sciences of the United States of America*, 110(45):18037–41, Nov. 2013. ISSN 1091-6490. doi: 10.1073/pnas.1314449110. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3831464&tool=pmcentrez&rendertype=abstract>.
- [86] R. Henderson, A. Sali, M. L. Baker, B. Carragher, B. Devkota, K. H. Downing, E. H. Egelman, Z. Feng, J. Frank, N. Grigorieff, W. Jiang, S. J. Ludtke, O. Medalia, P. A. Penczek, P. B. Rosenthal, M. G. Rossmann, M. F. Schmid, G. F. Schröder, A. C. Steven, D. L. Stokes, J. D. Westbrook, W. Wriggers, H. Yang, J. Young, H. M. Berman, W. Chiu, G. J. Kleywegt, and C. L. Lawson. Outcome of the first electron microscopy validation task force meeting. *Structure (London, England : 1993)*, 20(2):205–14, Feb. 2012. ISSN 1878-4186. doi: 10.1016/j.str.2011.12.014. URL <http://www.sciencedirect.com/science/article/pii/S0969212612000147>.

- [87] A. J. Henry and C. Meier. *The Thermofluor Method*, 2010.
- [88] K. Henzler-Wildman and D. Kern. Dynamic personalities of proteins. *Nature*, 450(7172):964–72, Dec. 2007. ISSN 1476-4687. doi: 10.1038/nature06522. URL <http://www.ncbi.nlm.nih.gov/pubmed/18075575>.
- [89] G. T. Herman and M. Kalinowski. Classification of heterogeneous electron microscopic projections into homogeneous subsets. *Ultramicroscopy*, 108(4):327–38, Mar. 2008. ISSN 0304-3991. doi: 10.1016/j.ultramic.2007.05.005. URL <http://www.sciencedirect.com/science/article/pii/S0304399107001374>.
- [90] J. B. Heymann and D. M. Belnap. Bsoft: image processing and molecular modeling for electron microscopy. *Journal of structural biology*, 157(1):3–18, Jan. 2007. ISSN 1047-8477. doi: 10.1016/j.jsb.2006.06.006. URL <http://www.sciencedirect.com/science/article/pii/S1047847706001997>.
- [91] Y. Hiromasa, T. Fujisawa, Y. Aso, and T. E. Roche. Organization of the cores of the mammalian pyruvate dehydrogenase complex formed by E2 and E2 plus the E3-binding protein and their capacities to bind the E1 and E3 components. *The Journal of biological chemistry*, 279(8):6921–33, Feb. 2004. ISSN 0021-9258. doi: 10.1074/jbc.M308172200. URL <http://www.ncbi.nlm.nih.gov/pubmed/14638692>.
- [92] S.-T. D. Hsu, P. Fucini, L. D. Cabrita, H. Launay, C. M. Dobson, and J. Christodoulou. Structure and dynamics of a ribosome-bound nascent chain by NMR spectroscopy. *Proceedings of the National Academy of Sciences of the United States of America*, 104(42):16516–21, Oct. 2007. ISSN 0027-8424. doi: 10.1073/pnas.0704664104. URL <http://www.pnas.org/cgi/content/long/104/42/16516>.
- [93] R. Ishima and D. a. Torchia. Protein dynamics from NMR. *Nature structural biology*, 7(9):740–3, Sept. 2000. ISSN 1072-8368. doi: 10.1038/78963. URL <http://www.ncbi.nlm.nih.gov/pubmed/10966641>.
- [94] T. Izard, a. Aevarsson, M. D. Allen, a. H. Westphal, R. N. Perham, a. de Kok, and W. G. Hol. Principles of quasi-equivalence and Euclidean geometry govern the assembly of cubic and dodecahedral cores of pyruvate dehydrogenase complexes. *Proceedings of the National Academy of Sciences of the United States of America*, 96(4):1240–5, Feb. 1999. ISSN 0027-8424. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=15447&tool=pmcentrez&rendertype=abstract>.
- [95] H. F. Jepsen and B. Jensen. Accumulation of trehalose in the thermophilic fungus *Chaetomium thermophilum* var. *coprophilum* in response to heat or salt stress. *Soil Biology and Biochemistry*, 36(10):1669–1674, Oct. 2004. ISSN 00380717. doi: 10.1016/j.soilbio.2004.07.010. URL <http://www.sciencedirect.com/science/article/pii/S0038071704002007>.

- [96] Q. Jin, C. O. S. Sorzano, J. M. de la Rosa-Trevín, J. R. Bilbao-Castro, R. Núñez Ramírez, O. Llorca, F. Tama, and S. Jonić. Iterative Elastic 3D-to-2D Alignment Method Using Normal Modes for Studying Structural Dynamics of Large Macromolecular Complexes. *Structure (London, England : 1993)*, pages 1–11, Feb. 2014. ISSN 1878-4186. doi: 10.1016/j.str.2014.01.004. URL <http://www.ncbi.nlm.nih.gov/pubmed/24508340>.
- [97] C. Joo, H. Balci, Y. Ishitsuka, C. Buranachai, and T. Ha. Advances in single-molecule fluorescence methods for molecular biology. *Annual review of biochemistry*, 77:51–76, Jan. 2008. ISSN 0066-4154. doi: 10.1146/annurev.biochem.77.070606.101543. URL <http://www.ncbi.nlm.nih.gov/pubmed/18412538>.
- [98] L. Joyeux and P. a. Penczek. Efficiency of 2D alignment methods. *Ultramicroscopy*, 92(2):33–46, July 2002. ISSN 0304-3991. URL <http://www.ncbi.nlm.nih.gov/pubmed/12138941>.
- [99] M. Karplus and J. A. McCammon. Molecular dynamics simulations of biomolecules. *Nature structural biology*, 9(9):646–52, Sept. 2002. ISSN 1072-8368. doi: 10.1038/nsb0902-646. URL <http://www.ncbi.nlm.nih.gov/pubmed/12198485>.
- [100] B. Kastner, N. Fischer, M. M. Golas, B. Sander, P. Dube, D. Boehringer, K. Hartmuth, J. Deckert, F. Hauer, E. Wolf, H. Uchtenhagen, H. Urlaub, F. Herzog, J. M. Peters, D. Poerschke, and R. Lu. GraFix : sample preparation for single- particle electron cryomicroscopy. *Nature Methods*, 5(1):53–55, 2008. doi: 10.1038/NMETH1139.
- [101] A. J. Katan and C. Dekker. High-speed AFM reveals the dynamics of single biomolecules at the nanometer scale. *Cell*, 147(5):979–82, Nov. 2011. ISSN 1097-4172. doi: 10.1016/j.cell.2011.11.017. URL <http://www.ncbi.nlm.nih.gov/pubmed/22118456>.
- [102] L. E. Kay. NMR studies of protein structure and dynamics. *Journal of magnetic resonance (San Diego, Calif. : 1997)*, 173(2):193–207, Apr. 2005. ISSN 1090-7807. doi: 10.1016/j.jmr.2004.11.021. URL <http://www.sciencedirect.com/science/article/pii/S1090780704003854>.
- [103] A. Kidera and N. Go. Refinement of protein dynamic structure: normal mode refinement. *Proceedings of the National Academy of Sciences*, 87(10):3718–3722, May 1990. ISSN 0027-8424. doi: 10.1073/pnas.87.10.3718. URL <http://www.pnas.org/cgi/content/long/87/10/3718>.
- [104] J.-M. Kirves. *New Algorithms for Single Particle Cryo Electron Microscopic Image Processing*. PhD thesis, University of Goettingen, 2014.

- [105] M. Knoll and E. Ruska. Das Elektronenmikroskop. *Zeitschrift fuer Physik*, 78: 318–339, 1932.
- [106] M. Koike, L. J. Reed, and W. R. Carrol. a-Keto Acid Dehydrogenation Complexes. I: Purification and Properties of Pyruvate and a-ketoglutarate dehydrogenation Complexes of *Escherichia coli*. *Journal of Biological Chemistry*, 235(7):1924–1930, 1960.
- [107] S. Korkes, A. del Campillo, I. C. Gunsalus, and S. Ochoa. ENZYMATIC SYNTHESIS OF CITRIC ACID : IV . PYRUVATE AS ACETYL DONOR Seymour Korkes , Alice del Campillo , I . C . *Journal of Biological Chemistry*, 193:721–735, 1951.
- [108] A. Korostelev, S. Trakhanov, M. Laurberg, and H. F. Noller. Crystal structure of a 70S ribosome-tRNA complex reveals functional interactions and rearrangements. *Cell*, 126(6):1065–77, Oct. 2006. ISSN 0092-8674. doi: 10.1016/j.cell.2006.08.032. URL <http://www.sciencedirect.com/science/article/pii/S0092867406011469>.
- [109] A. Kucukelbir, F. J. Sigworth, and H. D. Tagare. Quantifying the local resolution of cryo-EM density maps. *Nature methods*, pages 1–15, Nov. 2013. ISSN 1548-7105. doi: 10.1038/nmeth.2727. URL <http://www.ncbi.nlm.nih.gov/pubmed/24213166>.
- [110] A. Kucukelbir, F. J. Sigworth, and H. D. Tagare. Quantifying the local resolution of cryo-EM density maps. *Nature Methods*, (November):1–5, 2013. ISSN 1548-7091. doi: 10.1038/nmeth.2727. URL <http://dx.doi.org/10.1038/nmeth.2727>.
- [111] W. Kühlbrandt. The Resolution Revolution. *Science*, 343:1443–1444, Oct. 2014. ISSN 1470-8752. doi: 10.1042/BST0371042. URL <http://www.ncbi.nlm.nih.gov/pubmed/24675944>.
- [112] Kurzynski. *The Thermodynamic Machinery of Life*. Springer, 2006. ISBN 9783540238881.
- [113] O. F. Lange, N.-A. Lakomek, C. Farès, G. F. Schröder, K. F. a. Walter, S. Becker, J. Meiler, H. Grubmüller, C. Griesinger, and B. L. de Groot. Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. *Science (New York, N.Y.)*, 320(5882):1471–5, June 2008. ISSN 1095-9203. doi: 10.1126/science.1157092. URL <http://www.ncbi.nlm.nih.gov/pubmed/18556554>.
- [114] R. Langlois, J. Pallesen, and J. Frank. Reference-free particle selection enhanced with semi-supervised machine learning for cryo-electron microscopy. *Journal of structural biology*, 175(3):353–61, Sept. 2011. ISSN 1095-8657. doi: 10.1016/j.

- jsb.2011.06.004. URL <http://www.sciencedirect.com/science/article/pii/S1047847711001705>.
- [115] W. C. Y. Lau and J. L. Rubinstein. Subnanometre-resolution structure of the intact *Thermus thermophilus* H<sup>+</sup>-driven ATP synthase. *Nature*, 481(7380):214–8, Jan. 2012. ISSN 1476-4687. doi: 10.1038/nature10699. URL <http://www.ncbi.nlm.nih.gov/pubmed/22178924>.
- [116] A. E. Leschziner and E. Nogales. Visualizing Flexibility at Molecular Resolution : Analysis of Heterogeneity in Single-Particle Electron Microscopy. *Annual reviews in Biophysics*, pages 43–62, 2007. doi: 10.1146/annurev.biophys.36.040306.132742.
- [117] A. E. Leschziner, A. Saha, J. Wittmeyer, Y. Zhang, C. Bustamante, B. R. Cairns, and E. Nogales. Conformational flexibility in the chromatin remodeler RSC observed by electron microscopy and the orthogonal tilt reconstruction method. *Proceedings of the National Academy of Sciences of the United States of America*, 104(12):4913–8, Mar. 2007. ISSN 0027-8424. doi: 10.1073/pnas.0700706104. URL <http://www.pnas.org/cgi/content/long/104/12/4913>.
- [118] S. Li and B. Sun. *Advances in Soft Matter Physics*. Springer, 1st edition, 2012. ISBN 9787040317299.
- [119] B. Lieb, K. Dimitrova, H.-S. Kang, S. Braun, W. Gebauer, A. Martin, B. Hanelt, S. a. Saenz, C. M. Adema, and J. Markl. Red blood with blue-blood ancestry: intriguing structure of a snail hemoglobin. *Proceedings of the National Academy of Sciences of the United States of America*, 103(32):12011–6, Aug. 2006. ISSN 0027-8424. doi: 10.1073/pnas.0601861103. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1567689&tool=pmcentrez&rendertype=abstract>.
- [120] W.-t. Liu. *Strategies to stabilize RNP complexes for structural determination by 3D cryo-electron microscopy*. PhD thesis, Georg-August University Goettingen, 2013.
- [121] P. Lu, X.-c. Bai, D. Ma, T. Xie, C. Yan, L. Sun, G. Yang, Y. Zhao, R. Zhou, S. H. W. Scheres, and Y. Shi. Three-dimensional structure of human  $\gamma$ -secretase. *Nature*, 512(7513):166–170, June 2014. ISSN 0028-0836. doi: 10.1038/nature13567. URL <http://www.nature.com/doifinder/10.1038/nature13567>.
- [122] D. Lyumkis, S. K. Doamekpor, M. H. Bengtson, J.-W. Lee, T. B. Toro, M. D. Petroski, C. D. Lima, C. S. Potter, B. Carragher, and C. a. P. Joazeiro. Single-particle EM reveals extensive conformational variability of the Ltn1 E3 ligase. *Proceedings of the National Academy of Sciences of the United States of America*, 110(5):1702–7, Jan. 2013. ISSN 1091-6490. doi: 10.1073/pnas.1210041110. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3562785&tool=pmcentrez&rendertype=abstract>.

- [123] A. Mattevi, G. Obmolova, E. Schulze, K. H. Kalk, a. H. Westphal, A. de Kok, and W. G. Hol. Atomic structure of the cubic core of the pyruvate dehydrogenase multienzyme complex. *Science (New York, N.Y.)*, 255(5051):1544–50, Mar. 1992. ISSN 0036-8075. URL <http://www.ncbi.nlm.nih.gov/pubmed/1549782>.
- [124] M. E. Matyskiela, G. C. Lander, and A. Martin. Conformational switching of the 26S proteasome enables substrate degradation. *Nature Structural & Molecular Biology*, 20(7):781–788, June 2013. ISSN 1545-9993. doi: 10.1038/nsmb.2616. URL <http://www.nature.com/doifinder/10.1038/nsmb.2616>.
- [125] A.-C. Milazzo, A. Cheng, A. Moeller, D. Lyumkis, E. Jacovetty, J. Polukas, M. H. Ellisman, N.-H. Xuong, B. Carragher, and C. S. Potter. Initial evaluation of a direct detection device detector for single particle cryo-electron microscopy. *Journal of structural biology*, 176(3):404–8, Dec. 2011. ISSN 1095-8657. doi: 10.1016/j.jsb.2011.09.002. URL <http://www.sciencedirect.com/science/article/pii/S1047847711002577>.
- [126] J. L. S. Milne, D. Shi, P. B. Rosenthal, J. S. Sunshine, G. J. Domingo, X. Wu, B. R. Brooks, R. N. Perham, R. Henderson, and S. Subramaniam. Molecular architecture and mechanism of an icosahedral pyruvate dehydrogenase complex: a multifunctional catalytic machine. *The EMBO journal*, 21(21):5587–98, Nov. 2002. ISSN 0261-4189. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=131071&tool=pmcentrez&rendertype=abstract>.
- [127] J. L. S. Milne, X. Wu, M. J. Borgnia, J. S. Lengyel, B. R. Brooks, D. Shi, R. N. Perham, and S. Subramaniam. Molecular Structure of a 9-MDa Icosahedral Pyruvate Dehydrogenase Subcomplex Containing the E2 and E3 Enzymes Using Cryoelectron Microscopy \*. *Journal of Biological Chemistry*, 281(7):4364–4370, 2006. doi: 10.1074/jbc.M504363200.
- [128] T. Monecke, T. Güttler, P. Neumann, A. Dickmanns, D. Görlich, and R. Ficner. Crystal Structure of the Nuclear Export. *Science*, 69(May):1087–1091, 2009.
- [129] T. Monecke, D. Haselbach, B. Voß, A. Russek, P. Neumann, E. Thomson, and E. Hurt. Structural basis for cooperativity of CRM1 export complex formation. *Proceedings of the National Academy of Sciences*, 2012. doi: 10.1073/pnas.1215214110/-/DCSupplemental.[www.pnas.org/cgi/doi/10.1073/pnas.1215214110](http://www.pnas.org/cgi/doi/10.1073/pnas.1215214110).
- [130] T. Monecke, A. Dickmanns, and R. Ficner. Allosteric control of the exportin CRM1 unraveled by crystal structure analysis. *The FEBS journal*, 49(0), May 2014. ISSN 1742-4658. doi: 10.1111/febs.12842. URL <http://www.ncbi.nlm.nih.gov/pubmed/24823279>.

- [131] M. F. Moody. *Structural Biology using electrons and x-rays*. Elsevier, 2011. ISBN 9780123705815.
- [132] P. B. Moore. How should we think about the ribosome? *Annual review of biophysics*, 41:1–19, Jan. 2012. ISSN 1936-1238. doi: 10.1146/annurev-biophys-050511-102314. URL <http://www.ncbi.nlm.nih.gov/pubmed/22577819>.
- [133] M. Mueller, S. Jenni, and N. Ban. Strategies for crystallization and structure determination of very large macromolecular assemblies. *Current opinion in structural biology*, 17(5):572–9, Oct. 2007. ISSN 0959-440X. doi: 10.1016/j.sbi.2007.09.004. URL <http://www.sciencedirect.com/science/article/pii/S0959440X07001443>.
- [134] A. M. Mulder, C. Yoshioka, A. H. Beck, A. E. Bunner, R. a. Milligan, C. S. Potter, B. Carragher, and J. R. Williamson. Visualizing ribosome biogenesis: parallel assembly pathways for the 30S subunit. *Science (New York, N.Y.)*, 330(6004):673–7, Oct. 2010. ISSN 1095-9203. doi: 10.1126/science.1193220. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2990404&tool=pmcentrez&rendertype=abstract>.
- [135] K. Murakami, H. Elmlund, N. Kalisman, D. a. Bushnell, C. M. Adams, M. Azubel, D. Elmlund, Y. Levi-Kalisman, X. Liu, B. J. Gibbons, M. Levitt, and R. D. Kornberg. Architecture of an RNA polymerase II transcription pre-initiation complex. *Science (New York, N.Y.)*, 342(6159):1238724, Nov. 2013. ISSN 1095-9203. doi: 10.1126/science.1238724. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4039082&tool=pmcentrez&rendertype=abstract>.
- [136] G. E. Murphy and G. J. Jensen. Electron cryotomography of the E. coli pyruvate and 2-oxoglutarate dehydrogenase complexes. *Structure (London, England : 1993)*, 13(12):1765–73, Dec. 2005. ISSN 0969-2126. doi: 10.1016/j.str.2005.08.016. URL <http://www.ncbi.nlm.nih.gov/pubmed/16338405>.
- [137] K. Nagayama. Another 60 years in electron microscopy: development of phase-plate electron microscopy and biological applications. *Journal of electron microscopy*, 60 Suppl 1(Supplement 1):S43–62, Jan. 2011. ISSN 1477-9986. doi: 10.1093/jmicro/dfr037. URL <http://www.ncbi.nlm.nih.gov/pubmed/21844600>.
- [138] M. Nakanishi-Matsui, M. Sekiya, R. K. Nakamoto, and M. Futai. The mechanism of rotating proton pumping ATPases. *Biochimica et biophysica acta*, 1797(8):1343–52, Aug. 2010. ISSN 0006-3002. doi: 10.1016/j.bbabi.2010.02.014. URL <http://www.sciencedirect.com/science/article/pii/S000527281000068X>.
- [139] V. Q. Nguyen, A. Ranjan, F. Stengel, D. Wei, R. Aebersold, C. Wu, and A. E. Leschziner. Molecular architecture of the ATP-dependent chromatin-remodeling complex SWR1. *Cell*, 154(6):1220–31, Sept. 2013. ISSN 1097-4172. doi: 10.1016/

- j.cell.2013.08.018. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3776929&tool=pmcentrez&rendertype=abstract>.
- [140] F. H. Niesen, H. Berglund, and M. Vedadi. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nature protocols*, 2(9):2212–2221, 2007. doi: 10.1038/nprot.2007.321.
- [141] R. Nogales-Cadenas, S. Jonic, F. Tama, a. a. Arteni, D. Tabas-Madrid, M. Vázquez, a. Pascual-Montano, and C. O. S. Sorzano. 3DEM Loupe: Analysis of macromolecular dynamics using structures from electron microscopy. *Nucleic acids research*, 41(Web Server issue):W363–7, July 2013. ISSN 1362-4962. doi: 10.1093/nar/gkt385. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3692114&tool=pmcentrez&rendertype=abstract>.
- [142] H. Ode, M. Nakashima, S. Kitamura, W. Sugiura, and H. Sato. Molecular dynamics simulation in virus research. *Frontiers in microbiology*, 3(July):258, Jan. 2012. ISSN 1664-302X. doi: 10.3389/fmicb.2012.00258. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3400276&tool=pmcentrez&rendertype=abstract>.
- [143] J. N. Onuchic, Z. Luthey-Schulten, and P. G. Wolynes. Theory of protein folding: the energy landscape perspective. *Annual review of physical chemistry*, 48(1):545–600, Jan. 1997. ISSN 0066-426X. doi: 10.1146/annurev.physchem.48.1.545. URL <http://www.ncbi.nlm.nih.gov/pubmed/9348663>.
- [144] E. V. Orlova and H. R. Saibil. *Methods for three-dimensional reconstruction of heterogeneous assemblies.*, volume 482. Elsevier Inc., 1 edition, Jan. 2010. doi: 10.1016/S0076-6879(10)82013-0. URL <http://www.ncbi.nlm.nih.gov/pubmed/20888967>.
- [145] E. V. Orlova and H. R. Saibil. Structural analysis of macromolecular assemblies by electron microscopy. *Chemical reviews*, 111(12):7710–48, Dec. 2011. ISSN 1520-6890. doi: 10.1021/cr100353t. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3239172&tool=pmcentrez&rendertype=abstract>.
- [146] M. W. Pantaliano, F. R. Salemme, and T. E. Carver Jr. High throughput method for functional classifying proteins identified using a genomics approach, 2002.
- [147] E. Pastoriza-Munoz, R. M. Harrington, and M. L. Graber. Axial Heterogeneity of Intracellular pH in Rat Proximal Convolute Tubule Nephron segment identification. *J. Clin. Invest.*, 80:207–215, 1987. doi: 10.1172/JCI113049.imal.



- [148] M. S. Patel and L. G. Korotchkina. Mini-Series : Modern Metabolic Concepts The Biochemistry of the Pyruvate Dehydrogenase Complex \*. *Biochemistry and Molecular Biology education*, 3(iii):5–15, 2003.
- [149] M. S. Patel and T. E. Roche. Molecular biology and biochemistry of pyruvate dehydrogenase complexes. *FASEB J*, 4:3224–3233, 1990.
- [150] P. A. Penczek. Three-dimensional spectral signal-to-noise ratio for a class of reconstruction algorithms. *Journal of Structural Biology*, 138(1-2):34–46, Apr. 2002. ISSN 10478477. doi: 10.1016/S1047-8477(02)00033-3. URL <http://www.sciencedirect.com/science/article/pii/S1047847702000333>.
- [151] P. A. Penczek. *Image restoration in cryo-electron microscopy.*, volume 482. Elsevier Inc., 1 edition, Jan. 2010. doi: 10.1016/S0076-6879(10)82002-6. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3166661&tool=pmcentrez&rendertype=abstract>.
- [152] P. A. Penczek. *Fundamentals of three-dimensional reconstruction from projections.*, volume 482. Elsevier Inc., 1 edition, Jan. 2010. doi: 10.1016/S0076-6879(10)82001-4. URL <http://www.ncbi.nlm.nih.gov/pubmed/20888956>.
- [153] P. A. Penczek. *Resolution measures in molecular electron microscopy.*, volume 482. Elsevier Inc., 1 edition, Jan. 2010. doi: 10.1016/S0076-6879(10)82003-8. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3165049&tool=pmcentrez&rendertype=abstract>.
- [154] P. A. Penczek, R. A. Grassucci, and J. Frank. The ribosome at improved resolution: New techniques for merging and orientation refinement in 3D cryo-electron microscopy of biological particles. *Ultramicroscopy*, 53(3):251–270, Mar. 1994. ISSN 03043991. doi: 10.1016/0304-3991(94)90038-8. URL <http://www.sciencedirect.com/science/article/pii/0304399194900388>.
- [155] P. a. Penczek, J. Frank, and C. M. T. Spahn. A method of focused classification, based on the bootstrap 3D variance analysis, and its application to EF-G-dependent translocation. *Journal of structural biology*, 154(2):184–94, May 2006. ISSN 1047-8477. doi: 10.1016/j.jsb.2005.12.013. URL <http://www.ncbi.nlm.nih.gov/pubmed/16520062>.
- [156] P. A. Penczek, C. Yang, J. Frank, and C. M. T. Spahn. Estimation of variance in single-particle reconstruction using the bootstrap technique. *Journal of structural biology*, 154(2):168–83, May 2006. ISSN 1047-8477. doi: 10.1016/j.jsb.2006.01.003. URL <http://www.sciencedirect.com/science/article/pii/S1047847706000219>.

- [157] R. N. Perham and E. A. Hooper. Polypeptide chain stoichiometry in the self-assembly of the pyruvate dehydrogenase multienzyme complex of *Escherichia coli*. *FEBS Letters*, 73(2):137–140, Feb. 1977. ISSN 00145793. doi: 10.1016/0014-5793(77)80965-4. URL <http://www.sciencedirect.com/science/article/pii/0014579377809654>.
- [158] C. Petosa, G. Schoehn, P. Askjaer, U. Bauer, M. Moulin, U. Steuerwald, M. Soler-López, F. Baudin, I. W. Mattaj, and C. W. Müller. Architecture of CRM1/Exportin1 suggests how cooperativity is achieved during formation of a nuclear export complex. *Molecular cell*, 16(5):761–75, Dec. 2004. ISSN 1097-2765. doi: 10.1016/j.molcel.2004.11.018. URL <http://www.sciencedirect.com/science/article/pii/S1097276504006896>.
- [159] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin. UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of computational chemistry*, 25(13):1605–12, Oct. 2004. ISSN 0192-8651. doi: 10.1002/jcc.20084. URL <http://www.ncbi.nlm.nih.gov/pubmed/15264254>.
- [160] F. Platzmann. *Methods to improve the sample quality of macromolecular complexes for structure determination by 3D Electron Cryo-Microscopy*. PhD thesis, Georg-August University Gottingen, 2012.
- [161] M. Radermacher. Three-dimensional reconstruction of single particles from random and nonrandom tilt series. *Journal of electron microscopy technique*, 9(4):359–94, Aug. 1988. ISSN 0741-0581. doi: 10.1002/jemt.1060090405. URL <http://www.ncbi.nlm.nih.gov/pubmed/3058896>.
- [162] M. Radermacher, T. Wagenknecht, A. Verschoor, and J. Frank. Three-dimensional reconstruction from a single-exposure random conical tilt series applied to the 50S ribosomal subunit of *Escherichia coli*. *Journal of Microscopy*, 146:113–136, 1987.
- [163] N. A. Ranson, G. W. Farr, A. M. Roseman, B. Gowen, W. A. Fenton, A. L. Horwich, and H. R. Saibil. ATP-Bound States of GroEL Captured by Cryo-Electron Microscopy. *Cell*, 107(7):869–879, Dec. 2001. ISSN 00928674. doi: 10.1016/S0092-8674(01)00617-1. URL <http://www.sciencedirect.com/science/article/pii/S0092867401006171>.
- [164] B. Rasmussen, A. Stock, D. Ringe, and G. Petsko. Crystalline ribonuclease A loses function below the dynamical transition at 220 K. *Nature*, 357:423–424, 1992. URL <http://www.nature.com/nature/journal/v357/n6377/abs/357423a0.html>.
- [165] L. J. Reed. Pyruvate Dehydrogenase Complex. *Current Topics in Cellular Regulation*, pages 233–251, 1969.

- [166] L. J. Reed and R. M. Oliver. The multienzyme  $\alpha$ -keto Acid Dehydrogenase Complexes. *Brookhaven Symp. Biol.*, 21:397–412, 1968.
- [167] L. J. Reed, M. Koike, and C. R. Willms. Electron Microscopic and Biochemical Studies of Pyruvate Dehydrogenase Complex of *Escherichia coli*. *Science*, 145:930–932, 1964.
- [168] L. J. Reed, F. H. Pettit, M. H. Eley, L. Hamilton, J. I. M. H. Collins, and M. Oliver. Reconstitution of the *Escherichia coli* pyruvate dehydrogenase complex *Biochemistry* :. *PNAS*, 72(8):3068–3072, 1975.
- [169] P. B. Rosenthal and R. Henderson. Optimal Determination of Particle Orientation, Absolute Hand, and Contrast Loss in Single-particle Electron Cryomicroscopy. *Journal of Molecular Biology*, 333(4):721–745, Oct. 2003. ISSN 00222836. doi: 10.1016/j.jmb.2003.07.013. URL <http://linkinghub.elsevier.com/retrieve/pii/S0022283603010222>.
- [170] I. Rouiller, V. M. Butel, M. Latterich, R. A. Milligan, and E. M. Wilson-Kubalek. A Major Conformational Change in p97 AAA ATPase upon ATP Binding. *Molecular Cell*, 6(6):1485–1490, Dec. 2000. ISSN 10972765. doi: 10.1016/S1097-2765(00)00144-1. URL <http://www.sciencedirect.com/science/article/pii/S1097276500001441>.
- [171] N. Saito and Y. Matsuura. A 2.1-Å-resolution crystal structure of unliganded CRM1 reveals the mechanism of autoinhibition. *Journal of molecular biology*, 425(2):350–64, Jan. 2013. ISSN 1089-8638. doi: 10.1016/j.jmb.2012.11.014. URL <http://www.sciencedirect.com/science/article/pii/S0022283612008868>.
- [172] K. Y. Sanbonmatsu, S. Joseph, and C.-S. Tung. Simulating movement of tRNA into the ribosome during decoding. *Proceedings of the National Academy of Sciences of the United States of America*, 102(44):15854–9, Nov. 2005. ISSN 0027-8424. doi: 10.1073/pnas.0503456102. URL <http://www.pnas.org/cgi/content/long/102/44/15854>.
- [173] B. Sander, M. M. Golas, and H. Stark. Corrim-based alignment for improved speed in single-particle image processing. *Journal of Structural Biology*, 143:219–228, 2003. doi: 10.1016/j.jsb.2003.08.001.
- [174] B. Sander, M. M. Golas, and H. Stark. Advantages of CCD detectors for de novo three-dimensional structure determination in single-particle electron microscopy. *Journal of structural biology*, 151(1):92–105, July 2005. ISSN 1047-8477. doi: 10.1016/j.jsb.2005.04.004. URL <http://www.sciencedirect.com/science/article/pii/S1047847705001061>.

- [175] B. Sander, M. M. Golas, and R. Lu. Ways & Means An Approach for De Novo Structure Determination of Dynamic Molecular Assemblies by Electron Cryomicroscopy. *Structure*, 18:667–676, 2010. doi: 10.1016/j.str.2010.05.001.
- [176] S. J. Sanderson, C. Miller, and J. G. Lindsay. Stoichiometry, organisation and catalytic function of protein X of the pyruvate dehydrogenase complex from bovine heart. *European journal of biochemistry / FEBS*, 236(1):68–77, Feb. 1996. ISSN 0014-2956. URL <http://www.ncbi.nlm.nih.gov/pubmed/8617288>.
- [177] A. Sauerwald, S. Sandin, G. Cristofari, S. H. W. Scheres, J. Lingner, and D. Rhodes. Structure of active dimeric human telomerase. *Nature structural & molecular biology*, 20(4):454–60, Apr. 2013. ISSN 1545-9985. doi: 10.1038/nsmb.2530. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3785136&tool=pmcentrez&rendertype=abstract>.
- [178] M. Saur, V. Moeller, K. Kapetanopoulos, S. Braukmann, W. Gebauer, S. Tenzer, and J. Markl. Acetylcholine-binding protein in the hemolymph of the planorbid snail *Biomphalaria glabrata* is a pentagonal dodecahedron (60 subunits). *PloS one*, 7(8):e43685, Jan. 2012. ISSN 1932-6203. doi: 10.1371/journal.pone.0043685. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3423370&tool=pmcentrez&rendertype=abstract>.
- [179] W. O. Saxton and W. Baumeister. The correlation averaging of a regularly arranged bacterial cell envelope protein. *Journal of microscopy*, 127(Pt 2):127–138, 1982. ISSN 0022-2720. doi: 10.1111/j.1365-2818.1982.tb00405.x.
- [180] S. H. Scheres. Beam-induced motion correction for sub-megadalton cryo-EM particles. *eLife*, 3:1–8, Aug. 2014. ISSN 2050-084X. doi: 10.7554/eLife.03665. URL <http://elifesciences.org/lookup/doi/10.7554/eLife.03665>.
- [181] S. H. W. Scheres. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *Journal of structural biology*, 180(3):519–30, Dec. 2012. ISSN 1095-8657. doi: 10.1016/j.jsb.2012.09.006. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3690530&tool=pmcentrez&rendertype=abstract>.
- [182] S. H. W. Scheres and S. Chen. Prevention of overfitting in cryo-EM structure determination. *Nature methods*, 9(9):853–4, Sept. 2012. ISSN 1548-7105. doi: 10.1038/nmeth.2115. URL <http://www.ncbi.nlm.nih.gov/pubmed/22842542>.
- [183] S. H. W. Scheres, M. Valle, R. Nuñez, C. O. S. Sorzano, R. Marabini, G. T. Herman, and J.-M. Carazo. Maximum-likelihood multi-reference refinement for electron microscopy images. *Journal of molecular biology*, 348(1):139–49, Apr. 2005. ISSN

- 0022-2836. doi: 10.1016/j.jmb.2005.02.031. URL <http://www.sciencedirect.com/science/article/pii/S0022283605001932>.
- [184] S. H. W. Scheres, H. Gao, M. Valle, G. T. Herman, P. P. B. Eggermont, J. Frank, and J.-m. Carazo. Disentangling conformational states of macromolecules in 3D-EM through likelihood optimization. *4*(1):27–29, 2007. doi: 10.1038/NMETH992.
- [185] O. Scherzer. The Theoretical Resolution Limit of the Electron Microscope. *Journal of Applied Physics*, 20(1):20, 1949. ISSN 00218979. doi: 10.1063/1.1698233. URL <http://scitation.aip.org/content/aip/journal/jap/20/1/10.1063/1.1698233>.
- [186] I. Schlichting and J. Miao. Emerging opportunities in structural biology with X-ray free-electron lasers. *Current opinion in structural biology*, 22(5):613–26, Oct. 2012. ISSN 1879-033X. doi: 10.1016/j.sbi.2012.07.015. URL <http://www.sciencedirect.com/science/article/pii/S0959440X12001261>.
- [187] C. a. Schneider, W. S. Rasband, and K. W. Eliceiri. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, 9(7):671–675, June 2012. ISSN 1548-7091. doi: 10.1038/nmeth.2089. URL <http://www.nature.com/doifinder/10.1038/nmeth.2089>.
- [188] M. Shatsky, R. J. Hall, S. E. Brenner, and R. M. Glaeser. A method for the alignment of heterogeneous macromolecules from electron microscopy. *Journal of structural biology*, 166(1):67–78, Apr. 2009. ISSN 1095-8657. doi: 10.1016/j.jsb.2008.12.008. URL <http://www.sciencedirect.com/science/article/pii/S1047847708003018>.
- [189] F. J. Sigworth, P. C. Doerschuk, J.-M. Carazo, and S. H. W. Scheres. An introduction to maximum-likelihood methods in cryo-EM. *Methods in enzymology*, 482(10):263–94, Jan. 2010. ISSN 1557-7988. doi: 10.1016/S0076-6879(10)82011-7. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3051286&tool=pmcentrez&rendertype=abstract>.
- [190] C. V. Sindelar. A seesaw model for intermolecular gating in the kinesin motor protein. *Biophysical reviews*, 3(2):85–100, June 2011. ISSN 1867-2450. doi: 10.1007/s12551-011-0049-4. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3117274&tool=pmcentrez&rendertype=abstract>.
- [191] A. Singer, R. R. Coifman, F. J. Sigworth, D. W. Chester, and Y. Shkolnisky. Detecting consistent common lines in cryo-EM by voting. *Journal of structural biology*, 169(3):312–22, Mar. 2010. ISSN 1095-8657. doi: 10.1016/j.jsb.2009.11.003. URL <http://www.sciencedirect.com/science/article/pii/S1047847709003062>.

- [192] P. Sledz, F. Förster, and W. Baumeister. Allosteric Effects in the Regulation of 26S Proteasome Activities. *Journal of molecular biology*, pages 1–9, Feb. 2013. ISSN 1089-8638. doi: 10.1016/j.jmb.2013.01.036. URL <http://www.ncbi.nlm.nih.gov/pubmed/23416139>.
- [193] A. B. Smit, N. I. Syed, D. Schaap, J. V. Minnen, J. Klumperman, K. S. Kits, H. Lodder, A. V. Elk, B. Sorgedrager, A. Brejc, T. K. Sixma, W. P. M. Geraerts, and R. C. V. D. Schors. A glia-derived acetylcholine-binding protein that modulates synaptic transmission. *Nature*, 411(May):261–268, 2001.
- [194] M. T. J. Smith and J. L. Rubinstein. Beyond blob-ology. *Science*, 345(6197):617–619, Aug. 2014. ISSN 0036-8075. doi: 10.1126/science.1256358. URL <http://www.sciencemag.org/cgi/doi/10.1126/science.1256358>.
- [195] D. C. Speckhard and P. A. Frey. Escherichia coli pyruvate dehydrogenase complex: Improved purification and the flavin content. *Biochemical and Biophysical Research Communications*, 62(3):614–620, Feb. 1975. ISSN 0006291X. doi: 10.1016/0006-291X(75)90443-X. URL <http://www.sciencedirect.com/science/article/pii/0006291X7590443X>.
- [196] D. Stalling, M. Westerhoff, and H.-c. Hege. Amira - a Highly Interactive System for Visual Data Analysis. *The Visualisation Handbook*, 1:1–18, 2005.
- [197] P. E. Stephens, M. G. Darlison, H. M. Lewis, and J. R. Guest. The pyruvate dehydrogenase complex of Escherichia coli K12. Nucleotide sequence encoding the dihydrolipoamide acetyltransferase component. *European journal of biochemistry / FEBS*, 133(3):481–9, July 1983. ISSN 0014-2956. URL <http://www.ncbi.nlm.nih.gov/pubmed/6345153>.
- [198] A. Stewart and N. Grigorieff. Noise bias in the refinement of structures derived from single particles. *Ultramicroscopy*, 102(1):67–84, Dec. 2004. ISSN 0304-3991. doi: 10.1016/j.ultramic.2004.08.008. URL <http://www.sciencedirect.com/science/article/pii/S0304399104001706>.
- [199] F. W. Studier. Protein production by auto-induction in high-density shaking cultures. *Protein Expression and Purification*, 41(1):207–234, May 2005. ISSN 10465928. doi: 10.1016/j.pep.2005.01.016. URL <http://linkinghub.elsevier.com/retrieve/pii/S1046592805000264>.
- [200] K. Suhre and Y.-H. Sanejouand. ElNemo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement. *Nucleic acids research*, 32(Web Server issue):W610–4, July 2004. ISSN 1362-4962. doi: 10.1093/nar/gkh368. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=441506&tool=pmcentrez&rendertype=abstract>.

- [201] F. Tama, O. Miyashita, and C. L. Brooks. Normal mode based flexible fitting of high-resolution structure into low-resolution experimental data from cryo-EM. *Journal of structural biology*, 147(3):315–26, Sept. 2004. ISSN 1047-8477. doi: 10.1016/j.jsb.2004.03.002. URL <http://www.sciencedirect.com/science/article/pii/S1047847704000590>.
- [202] G. Tang, L. Peng, P. R. Baldwin, D. S. Mann, W. Jiang, I. Rees, and S. J. Ludtke. EMAN2: an extensible image processing suite for electron microscopy. *Journal of structural biology*, 157(1):38–46, Jan. 2007. ISSN 1047-8477. doi: 10.1016/j.jsb.2006.05.009. URL <http://www.ncbi.nlm.nih.gov/pubmed/16859925>.
- [203] K. Thakar, S. Karaca, S. a. Port, H. Urlaub, and R. H. Kehlenbach. Identification of CRM1-dependent Nuclear Export Cargos Using Quantitative Mass Spectrometry. *Molecular & cellular proteomics : MCP*, 12(3):664–78, Mar. 2013. ISSN 1535-9484. doi: 10.1074/mcp.M112.024877. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3591659&tool=pmcentrez&rendertype=abstract>.
- [204] D. W. Thompson. *On Growth and Form*. Cambridge University Press, 1917.
- [205] M. Tirion. Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis. *Physical review letters*, 77(9):1905–1908, Aug. 1996. ISSN 1079-7114. URL <http://www.ncbi.nlm.nih.gov/pubmed/10063201>.
- [206] R. J. Tomko Jr and M. Hochstrasser. Molecular Architecture and Assembly of the Eukaryotic Proteasome. *Annual review of biochemistry*, (March):1–31, Mar. 2013. ISSN 1545-4509. doi: 10.1146/annurev-biochem-060410-150257. URL <http://www.ncbi.nlm.nih.gov/pubmed/23495936>.
- [207] P. Unverdorben, F. Beck, P. ÅñLedÅñ, a. Schweitzer, G. Pfeifer, J. M. Plitzko, W. Baumeister, and F. Forster. Deep classification of a large cryo-EM dataset defines the conformational landscape of the 26S proteasome. *Proceedings of the National Academy of Sciences*, Mar. 2014. ISSN 0027-8424. doi: 10.1073/pnas.1403409111. URL <http://www.pnas.org/cgi/doi/10.1073/pnas.1403409111>.
- [208] M. Van Heel. Angular reconstitution: A posteriori assignment of projection directions for 3D reconstruction. *Ultramicroscopy*, 21(2):111–123, Jan. 1987. ISSN 03043991. doi: 10.1016/0304-3991(87)90078-7. URL <http://www.sciencedirect.com/science/article/pii/0304399187900787>.
- [209] M. van Heel. Finding trimeric HIV-1 envelope glycoproteins in random noise. *Proceedings of the National Academy of Sciences of the United States of America*, 110(45):E4175–7, Nov. 2013. ISSN 1091-6490. doi: 10.1073/pnas.1314353110. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3831429&tool=pmcentrez&rendertype=abstract>.

- [210] M. van Heel and M. Schatz. Fourier shell correlation threshold criteria. *Journal of structural biology*, 151(3):250–62, Sept. 2005. ISSN 1047-8477. doi: 10.1016/j.jsb.2005.05.009. URL <http://www.sciencedirect.com/science/article/pii/S1047847705001292>.
- [211] M. van Heel, G. Harauz, E. V. Orlova, R. Schmidt, and M. Schatz. A new generation of the IMAGIC image processing system. *Journal of structural biology*, 116(1):17–24, 1996. ISSN 1047-8477. doi: 10.1006/jsbi.1996.0004. URL <http://www.ncbi.nlm.nih.gov/pubmed/8742718>.
- [212] M. Vendruscolo and C. M. Dobson. Structural biology. Dynamic visions of enzymatic reactions. *Science (New York, N.Y.)*, 313(5793):1586–7, Sept. 2006. ISSN 1095-9203. doi: 10.1126/science.1132851. URL <http://www.ncbi.nlm.nih.gov/pubmed/16973868>.
- [213] S. Vijayakrishnan, S. M. Kelly, R. J. C. Gilbert, P. Callow, D. Bhella, T. Forsyth, J. G. Lindsay, and O. Byron. Solution Structure and Characterisation of the Human Pyruvate Dehydrogenase Complex Core Assembly. *Journal of Molecular Biology*, 399(1):71–93, 2010. ISSN 0022-2836. doi: 10.1016/j.jmb.2010.03.043. URL <http://dx.doi.org/10.1016/j.jmb.2010.03.043>.
- [214] D. Voet and J. G. Voet. *Biochemistry*. Wiley, 3rd edition, 2004.
- [215] O. Vogel, B. Hoehn, and U. Henning. Molecular structure of the pyruvate dehydrogenase complex from *Escherichia coli* K-12. *Proceedings of the National Academy of Sciences of the United States of America*, 69(6):1615–9, June 1972. ISSN 0027-8424. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=426760&tool=pmcentrez&rendertype=abstract>.
- [216] N. Voss, C. Yoshioka, M. Radermacher, C. Potter, and B. Carragher. DoG Picker and TiltPicker: Software tools to facilitate particle selection in single particle electron microscopy. *Journal of Structural Biology*, 166(2):205–213, May 2009. ISSN 10478477. doi: 10.1016/j.jsb.2009.01.004. URL <http://www.sciencedirect.com/science/article/pii/S1047847709000197>.
- [217] T. Wagenknecht, R. Grassucci, J. Berkowitz, and C. Forneris. Configuration of interdomain linkers in pyruvate dehydrogenase complex of *Escherichia coli* as determined by cryoelectron microscopy. *Journal of Structural Biology*, 109(1):70–77, July 1992. ISSN 10478477. doi: 10.1016/1047-8477(92)90069-M. URL <http://www.sciencedirect.com/science/article/pii/104784779290069M>.
- [218] J. S. Walls and P. A. Frey. Quaternary Structure of Pyruvate Dehydrogenase Complex from. *Proceedings of the National Academy of Sciences*, 260(30):16049–16051, 1985.



- [219] H.-w. Wang, C. Noland, B. Siridechadilok, D. W. Taylor, E. Ma, K. Felderer, J. A. Doudna, and E. Nogales. Structural insights into RNA processing by the human RISC-loading complex. *Nature Structural & Molecular Biology*, 16(11):1148–1153, 2009. ISSN 1545-9993. doi: 10.1038/nsmb.1673. URL <http://dx.doi.org/10.1038/nsmb.1673>.
- [220] C. R. Willms, R. M. Oliver, R. Henney, B. B. Mukherjee, and J. Reed. a Keto Acid Dehydrogenase Complexes: Dissociation and Reconstitution of the Dihydrolipoyl Transacetylase of Escherichia Coli. *jbc*, 242:889–897, 1967.
- [221] D. N. Wilson. The A-Z of bacterial translation inhibitors. *Critical reviews in biochemistry and molecular biology*, 44(6):393–433, 2009. ISSN 1549-7798. doi: 10.3109/10409230903307311. URL <http://www.ncbi.nlm.nih.gov/pubmed/19929179>.
- [222] S. Wray. Smooth muscle intracellular pH: measurement, regulation, and function. *The American journal of physiology*, 254(2 Pt 1):C213–25, Feb. 1988. ISSN 0002-9513. URL <http://www.ncbi.nlm.nih.gov/pubmed/3279796>.
- [223] X. Yu, Y. Hiromasa, H. Tsen, J. K. Stoops, T. E. Roche, and Z. H. Zhou. Structures of the Human Pyruvate Dehydrogenase Complex Cores : A Highly Conserved Catalytic Center with Flexible N-Terminal Domains. *Structure*, (January):104–114, 2008. doi: 10.1016/j.str.2007.10.024.
- [224] X. Yu, L. Jin, and Z. H. Zhou. 3.88 Å structure of cytoplasmic polyhedrosis virus by cryo-electron microscopy. *Nature*, 453:415–419, 2008. ISSN 0028-0836. doi: 10.1038/nature06893.
- [225] X. Yu, P. Ge, J. Jiang, I. Atanasov, and Z. H. Zhou. Atomic model of CPV reveals the mechanism used by this single-shelled virus to economically carry out functions conserved in multishelled reoviruses. *Structure (London, England : 1993)*, 19(5):652–61, May 2011. ISSN 1878-4186. doi: 10.1016/j.str.2011.03.003. URL <http://www.sciencedirect.com/science/article/pii/S0969212611001006>.
- [226] F. Zemlin, K. Weiss, P. Schiske, W. Kunath, and K.-H. Herrmann. Coma-free alignment of high resolution electron microscopes with the aid of optical diffractograms. *Ultramicroscopy*, 3:49–60, Jan. 1978. ISSN 03043991. doi: 10.1016/S0304-3991(78)80006-0. URL <http://www.sciencedirect.com/science/article/pii/S0304399178800060>.
- [227] J. Zhang, M. L. Baker, G. F. Schröder, N. R. Douglas, S. Reissmann, J. Jakana, M. Dougherty, C. J. Fu, M. Levitt, S. J. Ludtke, J. Frydman, and W. Chiu. Mechanism of folding chamber closure in a group II chaperonin. *Nature*, 463:379–383, 2010. ISSN 0028-0836. doi: 10.1038/nature08701.

- [228] R. Zhang, B. R. So, P. Li, J. Yong, T. Glisovic, L. Wan, and G. Dreyfuss. Structure of a key intermediate of the SMN complex reveals Gemin2's crucial function in snRNP assembly. *Cell*, 146(3):384–95, Aug. 2011. ISSN 1097-4172. doi: 10.1016/j.cell.2011.06.043. URL <http://www.sciencedirect.com/science/article/pii/S0092867411007185>.
- [229] G. Zhao, J. R. Perilla, E. L. Yufenyuy, X. Meng, B. Chen, J. Ning, J. Ahn, A. M. Gronenborn, K. Schulten, C. Aiken, and P. Zhang. Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. *Nature*, 497(7451):643–6, May 2013. ISSN 1476-4687. doi: 10.1038/nature12162. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3729984&tool=pmcentrez&rendertype=abstract>.
- [230] J. Zhao, M. A. Brubaker, and J. L. Rubinstein. TMaCS: a hybrid template matching and classification system for partially-automated particle selection. *Journal of structural biology*, 181(3):234–42, Mar. 2013. ISSN 1095-8657. doi: 10.1016/j.jsb.2012.12.010. URL <http://www.sciencedirect.com/science/article/pii/S1047847713000026>.
- [231] Z. H. Zhou, D. B. McCarthy, C. M. O'Connor, L. J. Reed, and J. K. Stoops. The remarkable structural and functional organization of the eukaryotic pyruvate dehydrogenase complexes. *Proceedings of the National Academy of Sciences of the United States of America*, 98(26):14802–7, Dec. 2001. ISSN 0027-8424. doi: 10.1073/pnas.011597698. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=64939&tool=pmcentrez&rendertype=abstract>.
- [232] Y. Zhu. Automatic particle selection: results of a comparative study. *Journal of Structural Biology*, 145(1-2):3–14, Jan. 2004. ISSN 10478477. doi: 10.1016/j.jsb.2003.09.033. URL <http://www.sciencedirect.com/science/article/pii/S1047847703002004>.

# Related Publications

Parts of this thesis have been published:

1. N. Dölker\*, CE. Blanchet\*, B. VoSS\*, **D. Haselbach**\*, C. Kappel, T. Monecke, DI. Svergun, H. Stark, R. Ficner, U. Zachariae, H. Grubmüller, A. Dickmanns (2013). Structural determinants and mechanism of mammalian CRM1 allostery. *Structure* *21(8):1350-60*
2. T. Monecke, **D. Haselbach**, B. VoSS, A. Russek, P. Neumann, E. Thompson, E. Hurt, U Zachariae, H. Stark, H. Grubmüller, A. Dickmanns, R. Ficner (2013). Structural basis for cooperativity of CRM1 export complex formation. *Proc Natl Acad Sci USA*. *110(3):960-5*
3. **D. Haselbach**\* , JM. Kirves\*, A. Chari, and H. Stark. ProteoPlex ? A thermodynamic and computational framework to optimize the stability of macromolecular complexes by thermal unfolding *manuscript submitted*
4. A. Chari, **D. Haselbach**, JM. Kirves, J. Ohmer, E. Paknia, N. Fischer, O. Ganichkin, V. Möller, G. Petzold, JM. Peters, BA. Schulman, K. Tittmann, J. Markl, U. Fischer and H. Stark. ProteoPlex - maximizing the stability of macromolecular complexes by systematic screening of chemical space *manuscript submitted*
5. H. Stark, A. Chari, **D. Haselbach**, JM Kirves. Methods for analyzing biological macromolecular complexes and use thereof. Patent, WO2013034160 A1, PCT/EP2011/004475

\* Authors contributed equally.



# Curriculum Vitae

David Haselbach

born 24<sup>th</sup> June 1985 in Cottbus, Germany

Am Gewende 3, 37083 Göttingen

Phone +49 (0) 163 7753201

E-Mail dhaselb@gwdg.de

---

## University education

- |                   |                                                                                                                                                                                                                                                                                                       |
|-------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 09/2010 - present | Max Planck Institute for biophysical Chemistry, Göttingen: Doctoral studies within the doctoral program IMPRS Molecular Biology of the Göttingen Graduate School for Neuroscience, Biophysics and Molecular Biosciences (GGNB). PhD thesis: <i>Conformational Dynamics of Large Protein Complexes</i> |
| 09/2008 - 03/2010 | Georg-August University Göttingen: Studies in Molecular Biology (MSc.). Master Thesis: <i>Unfolding Mechanics of the knotted proteins HCAIII and UCH-L3</i> at the Physics Department at TU Munich, Chair for Biophysics, E22                                                                         |
| 10/2005 - 08/2008 | University of Potsdam: Studies in Life Science (Specialisation on Biochemistry) (BSc.). Bachelor Thesis: <i>Detektion einzelsträngiger DNA mit Hilfe von DNAszymen</i> at the Max Planck Institute for Molecular Genetics                                                                             |

---

## School education

- |                   |                                                   |
|-------------------|---------------------------------------------------|
| 08/1998 - 07/2005 | Secondary school Max-Steenbeck-Gymnasium, Cottbus |
| 08/1989 - 07/1993 | Primary school 20. Grundschule Cottbus            |
-

