Computational methods for the structure determination of highly dynamic molecular machines by cryo-EM

DISSERTATION

for the award of the degree
"Doctor rerum naturalium" (Dr. rer. nat.)
of the GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN

within the
INTERNATIONAL MAX-PLANCK-RESEARCH-SCHOOL PHYSICS OF BIOLOGICAL AND COMPLEX SYSTEMS
of the Göttingen Graduate School for Neurosciences,
Biophysics, and Molecular Biosciences (GGNB)

submitted by

Felix Lambrecht

from Kassel, Germany

Göttingen 2018

Thesis Committee

Prof. Dr. Holger Stark
Structural Dynamics, Max-Planck-Institute for Biophysical Chemistry

Prof. Dr. Helmut Grubmüller
Theoretical and Computational Biophysics, Max-Planck-Institute for Biophysical Chemistry

Dr. Michael Habeck
Inverse Statistical Problems in Biophysics, Max-Planck-Institute for Biophysical Chemistry

Members of the Examination Board

1st Referee: Prof. Dr. Holger Stark
Structural Dynamics, Max-Planck-Institute for Biophysical Chemistry

2nd Referee: Prof. Dr. Helmut Grubmüller
Theoretical and Computational Biophysics, Max-Planck-Institute for Biophysical Chemistry

Further Members of the Examination Board

Dr. Michael Habeck
Inverse Statistical Problems in Biophysics, Max-Planck-Institute for Biophysical Chemistry

Prof. Dr. Kai Tittmann
Molecular Enzymology, Albrecht-von-Haller-Institute

Prof. Dr. Blanche Schwappach
Institute of Molecular Biology, University Medical Center Göttingen

Dr. Sarah Adio
Single molecule biochemistry of macromolecular machines, Institute for Microbiology and Genetics

Date of oral examination: 14 Februar 2019

AFFIDAVIT

I hereby declare that this dissertation with the title “Computational methods for the structure determination of highly dynamic molecular machines by cryo-EM” has been written independently and 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.

Felix Lambrecht

Abstract

In the last couple of years, electron cryomicroscopy (cryo-EM) has gained of rising importance in the field of structural biology and biophysics. Not only that the routinely achievable resolution of the method has dramatically increased to routinely near-atomic resolution. First and foremost the opportunity to resolve structures which where far beyond the size limit for classical methods such as X-ray crystallography and NMR increased the popularity. These features make cryo-EM also more and more interesting for the pharmaceutical industry. However, certain challenges are being unresolved until today. One of them is the usage of the full dynamical information cryo-EM experiments bring with them. The other is the fact that classical transmission electron microscopy (TEM) on biological samples still is limited by the radiation sensitivity and the low signal in the images. This thesis will first present a new algorithm based on statistical methods to decipher the conformational landscape of a macromolecular complex from cryo-EM experiments. This information can be used to quantify biochemical phenomena like allostery on the intramolecular level. Such analyses will be discussed to be useful in the drug development pipeline and help to understand the physical foundations of the regulation and function of macromolecular complexes. Secondly, a new imaging method based on scanning TEM will be introduced. In the corresponding publication we evaluate this method in its capability to resolve non-biological samples as well as biological macromolecues. We show that the surrounding medium influences the imaging process in a way that no compromise between the applied dose and the achievable signal can be found and it will be discussed what implications can be drawn for new biological imaging modes in general.

Acknowledgment

The last years gave me the chance to experience the many different flavours of research a place like the MPI for Biophysical Chemistry radiates. This built my mind, my personality and my scientific curiousity - and for this I am deeply grateful.

First and foremost I would like to thank Holger Stark for being the most forward thinking and critical supervisor I could envision. He shaped much of the experience and always pushed work into a fruitful direction. I learned a lot from him, scientifically but even more about the world around the contents: the visions, the thinking and the enthusiasm I will try to take with me from the time with him.

The person I talked most with in the last years and who always guided me in the daily work was Mario. He was a real supervisor for me, having an open ear as well as a gentle sense of humor I will never forget. With respect to the algorithmic side of my work, he taught me the most and formed my way of thinking about the mathematics and backgrounds of the things we were doing together. Not to forget the endless coffee rounds in which topics changed rapidly. I am also grateful for working together with some real experts in our field: Niels gave me many insights in the way image processing works, especially when you want to squeeze out the little information in the worst images he had ever seen. Uwe did taught me electron microscopy 101 and Dietmar suffered a lot for me on the microscope just to get out not much.
My personal mastermind left the department still when I was there. David and me developed the CowScape algorithm and the whole idea is based on his brain. I owe him a wonderful start, a wonderful time on the proteasome structure and a few kilograms due to cake. Thank you for the great time! Another great person who left the lab at that time was Jil, who I have to thank for fun hours on the bike and an open ear to any time. I had a great time with wonderful students in the lab and I would like to thank them all but especially Georg, Sabrina and Karl with which I spent most of that time. I also supervised many students in this period of my work, but only one was a person who did a perfectly fruitful job: thank you Ka Man for your enthusiasm.
Last but not least: thank you Juliane! For everything! I would have sometimes gone crazy without you.

My Thesis committee with Helmut Grubmüller and Michael Habeck was a real help, not only in the last phase of this work. With Michael I discussed a lot and significant ideas to the CowScape algorithm originated from him and I enjoyed his way of thinking about my problems as being always amazingly productive.

Last but not least I would like to thank the people who made the Göttingen experience, a real experience. Anna, Kati, Miri, Jörn and Maxi, people I hope I will never ever miss again in my life. This also describes the person, who turned my life around and supported me incredibly throughout the last weeks. My future wife. Julia. Last but not least I have to thank the persons who made all of this possible: my parents and my grandma who all never ever questioned me at all.

Contents

Abstract II

Acknowledgment III

List of Tables VIII

List of Figures XI

1 Introduction 1
- 1.0.1 Single Particle cryo-EM . . . 3
  - 1.0.1.1 The Transmission Electron Microscope . . . 3
  - 1.0.1.2 Phase Contrast and the Contrast Transfer Function . . . 4
  - 1.0.1.3 Sample Vitrification for cryo-EM . . . 5
- 1.0.2 Single Particle Analysis (SPA) . . . 6
  - 1.0.2.1 Computational Standard Workflow . . . 8
- 1.1 Conformational Landscapes from Cryo-EM Experiments . . . 10
  - 1.1.1 Conformational Dynamics of Proteincomplexes . . . 12
    - 1.1.1.1 The Glass-Like Model of Conformational Changes . . . 12
    - 1.1.1.2 Recognition by Ensemble . . . 13
    - 1.1.1.3 Allosteric Regulation by Ensemble . . . 14
  - 1.1.2 State-Of-The-Art Visualization of Conformational Landscapes . . . 16
    - 1.1.2.1 Molecular Dynamics Simulations . . . 17
    - 1.1.2.2 X-ray crystallography . . . 17
    - 1.1.2.3 Spectroscopical Approaches to Dynamics . . . 18
  - 1.1.3 Computational Approaches for Distangling Conformations in Cryo-EM . . . 19
    - 1.1.3.1 Contributions in this work: the CowScape algorithm . . . 21
- 1.2 Single Particle Cryo-EM with iDPC-STEM . . . 22
  - 1.2.1 The Optics of a STEM . . . 22
  - 1.2.2 STEM and Imaging of Biological Molecules . . . 24
  - 1.2.3 (integrated) Differential Phase Contrast . . . 26

2 Long-range allosteric regulation of the human 26S proteasome by 20S proteasome-targeting cancer drugs 31

3 CowScape: a novel approach to quantitatively describe the conformational landscape of biological macromolecules from cryo-EM data 41

4 Biological single particle imaging with integrated differential phase contrast STEM is limited by contrast in vitreous specimen 57

5 Discussion **69**
5.1 Conformational Landscapes from Cryo-EM Experiments 69
5.1.1 PCA on Density Maps 70
5.1.1.1 The Potential of Clustering Algorithms 70
5.1.1.2 Hierarchy of Conformational Landscapes 71
5.1.2 Estimation of the Landscape 72
5.1.2.1 Choice of Principal Components 72
5.1.2.2 Interpolation of the Landscape 73
5.1.3 Interpreting Allostery with cryo-EM 74
5.1.4 Quality Assesment and Biochemical Optimization 75
5.1.5 The Interplay with MD-Simulation 77
5.1.6 Drug Development 78
5.1.6.1 Mechanistic studies 79
5.1.7 Conclusion 80
5.2 Single Particle Cryo-EM with iDPC-STEM 81
5.2.1 Contribution of the Background 81
5.2.2 The contribution of the contrast transfer function (CTF) and phase plates 83
5.2.3 Orthogonal Information from STEM 85
5.2.4 STEM-tomography 86
5.2.5 Conclusion 87

6 Outlook **89**

A Abbreviations **91**

B Supplementary information **93**
B.1 Supplementary Information to *Lambrecht (submitted)* 93

C References **99**

Curriculum Vitae **115**

List of Tables

1.1 Comparism of different algorithmical approaches. Modified from [37]. R = Ratcheted, R+Efg = Ratcheted with Efg bound, UE = Unratcheted with an occupied E-Site, UL1 = Unratcheted with free E-site and L1 domain out, LSU = Large Subunit only, ... 20

5.1 Overview on usage cases of the CowScape idea 69

List of Figures

1.1 **Simplified Ray Diagram of a TEM**[99] 3

1.2 **Experimental workflow of vitrification** The sample solution is pipetted on to a support grid and access liquid is removed by removing it with a filterpaper ("Blotting") to reduce the ice thickness. Next, and in a very quick manner, this grid is shot into liquid ethane. This sample is then imaged in the electron microscope. Modified from [109] 5

1.3 **Common refinement workflow for cryo-EM experiments** After a initial preprocessing in which from the acquired images a initial model is generated, 3D classification is used to sort out dirt or uninterpretable images. Only a subset of the clases (namely usually only the ones with a high enough resolution) are retained. The final few classes having survived this "sieving" are then refined to high resolution (here, HR-Volumes A and B in two different conformations). . . . 9

1.4 **Schematic drawing of a slice through the energy funnel of protein folding:** shown in green is the overall folding funnel, in which the upper part has a high number of degrees of freedom and is folding is governed by the hydrophobic effect. In the lower region - governed by specific interactions - the native state is located. In the circle this folded, native state is zoomed in. It involves misfolded parts as well as the functional landscape. adapted from [97, 145] 11

1.5 **The dynamics of one particular state is hierarchically ordered:** the dynamical contributions to the overall broadening of the state is described by different tiers, which describe the imposed ruggedness on different scales of energy and time. The decreasing energy differences between minima and maxima in the landscape lead to shorter kinetics on the corresponding tiers. The differences between minima (Δ) can be described by 1.6. Adapted from [47, 60] 12

1.6 **Overview on different binding mechanisms:** a) shows the classical key-and-lock model by Fischer, where the binding site is already predefined to just fit the substrate in. b) sketches the induced fit model, which is split in a two-step-mechanism, first binding the substrate weakly, which secondly alters the binding site and allows full recognition. c) shows the conformational selection model, where the three conformers are prototypically for the full conformational ensemble. 14

1.7 **Partition function of a tetrameric protein with a binding partner:** Each subunit of the complex can exist in in T and R state (square and circle, respectively). The substrate binding (triangle) happens consecutively. In a) the sequential KNF model is shown. Upon binding the teramer sequentially changes from one to the other conformation by a induced fit model. In b) the WMC model the substrate binding can occur in all yellow states with different affinities. Both models do not take the never-yellow partitions into account, which is in accordance with the EAM model. Adapted from [63]. 15

1.8 **Optics of a STEM** (a) The general optical path consists of a condenser lens system, the scan coils, which are allowing to scan the converged beam over the sample and an objective lens, forming the final probe. Dependent on the detection angle either a brightfield or a darfkfield image is acquired. (b) The optics in a STEM and a CTEM are basically equivalent, just turned around. That implies that most of the optics in the STEM case are before the sample plane and in TEM it is the other way. Both images taken from [101] 23

1.9 **The focal depth depence on the convergence angle** The point spread function is simulated for different convergence angles. At 2 mrad convergence angle more than 1 μm of stable focus can be achieved, whereas the focal depth decays quadratically according to equation 1.8, when increasing the convergence angle. [64] 24

1.10 **Detection of single metal ions in a protein** Elad et al. [33] showed that they are able to resolve single binding sites of Zn-ions in vitrified ferretin usin DF-STEM. However, the protein structure is only on minor resolution. 25

1.11 **Detector Setup for iDPC STEM** The left side of the image shows the incident beam on the specimen, which demagnifies under a convergence angle α. Then the beam magnifies again on a split detector, which is shown again with the unperturbed beam in the right part of the image. Modified from [88] 26

1.12 **Working Principle of Differential Phase Contrast** (a) In a 1D case, when no charge is present, the beam is aligned as such it is illuminating both detectors equally. At the next scan position, the beam approaches a charge and becomes deflected to the right detector more than to the left, such that the subtraction $I(X) - I(Y)$ is negative. At the next scan position the sign changes. [132] (b) The image seen on the detector while traversing from the first to the second scan position equals a shift of the intensity of the electron rochigram [93]. 27

1.13 **Theoretical iDPC CTF** The theoretical CTF for the first term in equation 1.12, where $W(k)$ is chosen to account for a four-fold symmetrical detector, is shown as a function of defocus from the disc with best contrast. The ratio k_p/k_{bf} is just one possible choice for k. [80] 29

5.1 Based on the analysis of full ensembles rather then single structures one can imagine a hierarchical refinement scheme, which resembles the tier-structure shown in figure 1.5. By firstly relaxing the compositional misture and focussing on the stable protein complex the overall energy landscape of the stable complex can be refined. Here comparing studies can be done with other biochemical modifications such as inhibitor or substrate binding. "Zooming" further in one of the minima, single gaussian minima should be accessible, which should refine to highest resolution. The variability within one of these minima could be analysed by MD simulations. 76

5.2 SBDD in the early phases can be divided into three major phases. Where Target ID and Hit ID are primarily high-resolution guided process, Lead optimization can win from medium-resolution dynamical data. (ID=Identification, OPT=optimization), [117] 78

5.3 Transmission electron microscope (TEM, a) phase contrast and scanning transmission electron microscope (STEM, b) transfer as functions of the spatial frequency, g for different defoci. (a) Bright-field TEM (BF-TEM) contrast transfer functions (CTFs) (blue and green) for a thin-phase object and (b) incoherent STEM optical transfer function (OTF, red and orange) for a thin-amplitude object for the same spherical aberration, $C_s = 2.0$ mm, at 200 kV. [36] 83

5.4 a) Overview on the different options to introduce phase contrast: defocusing, VPP, ZPP. Note the central hole of ZPPs that introduce a cutoff in the low-resolution regime of the image, taken from [19] b) Experimental CTF of a standard TEM and a TEM equipped with a VPP, taken from [22]. c) Theoretical CTF of a ZPP for different phase shifts, taken from [20] 84

1 Introduction

Latest since the Direct Electron Detector (DED) displaced the charged coupled devices (CCD) literally out of the microscope, electron cryo-microscopy (cryo-EM) catched up to the established methods in structural biology[2, 43, 90]. Next, Kuehlbrandt [78] exclaimed the "Resolution Revolution" and the whole development peaked with the Nobel Prize for Chemistry in 2017 to Joachim Frank, Jacques Debouchet and Richard Henderson[131]. Together with this boost in hardware, also recently developed algorithmical improvements, such as Maximum-Likelihood based reconstruction methods [123, 136] increased the capability of the method to resolve "near-atomic" resolution. There are numerous reviews in the literature (a pubmed search of just two-years old reviews containing "cryo-EM" yields more then 160 results)[13, 15, 30, 31, 38, 44, 90, 137, 140]. But apart from the numerous technological advances in the last years, what differentiates cryo-EM from the established methods of X-ray crystallography and nuclear magnetic resonance spectroscopy (NMR)? NMR reconstructs the position of nuclei from spin interactions. However, as the size of the observed molecules increases, the number of interactions do as well and the reconstruction is limited by a lack of degeneracy of the signals[49]. Therefore, NMR is limited to small proteins. On the other hand, NMR is an in-solution method, which allows to directly visualize the conformational ensemble of a (small) biomolecule in solution. As will be outlined later, NMR until now played the major role in understanding thermodynamical or kinetic macroscopic observations (e.g. allostery) on the ensemble scale.

In contrast to this, X-ray crystallography is based on a crystallized sample of the investigated molecule, which is then investigated by X-ray diffraction. X-ray crystallography is the benchmark of understanding the chemistry of any biomolecule. Once a suitable crystallization condition is found (and this is still the major bottleneck in some cases), modern beamline technology allows the imaging of rather large unit cells with "chemical" resolution [127]. However, deriving dynamics from a crystal structure is hardly possible and never resembles the native in solution situation.

cryo-EM has its strength in the determination of rather large complexes - up to full viruses - even though the lower limit constantly decreases. However, at the moment the routinably achievable mean resolution is somewhere between 3 and 4 Å[109]. Even though, higher resolved structures have been reported. Nevertheless, cryo-EM aims to conquer the role of crystallography in resolving the chemistry of for example drug binding [140], not only since a provocative *Nature* editorial stating that "The Revolution will not be crystallized" [11].

NMR explained the physical basis of biochemical phenomena like binding, allostery and intra-molecular signalling for small molecules already on the ensemble level. Cryo-EM could in contrast explain all of this for the large central hubs in cellular metabolism and signalling: as cryo-EM captures a frozen snapshot of the in-solution sample, it is perfectly suited to decipher the conformational motions underlying the large macromolecular (brownian) machines. The question to answer can be nicely depicted by imagine the following scenario: a few thousand blindfolded dancers are put on a stage. They all know the same choreography and when the music starts to play, one after the other starts to dance. Now you take a snapshot of the dancers to a random time point. This will look completely random, because any of the dancers started to a different - random - point in time. The question is now: given the observation of all the dancers in different positions, what is the underlying choreography? The first part of this thesis will tackle that question and will even go one step further asking what the underlying energetical and functional relation between thosre movements are. Finally, it should be clear that even this feature might pave cryo-EM's role into the pharmaceutical research much faster than waiting for resolution to be routinely on the level that truly atomic resolutions can be achieved. Understanding the impact of drug binding to the dynamics of a macromolecular complex on the structural level, might offer a new way of validating and understanding drug targets.

The two major limiting factors for cryo-EM are resolution and the signal-to-noise ratio (SNR). Talking about dynamics, the SNR is what limits the distangling of individual conformers. This is in statistics known as the identifiability problem [137], which in a nutshell says that the ability of identifying subpopulations in a sample depends on the noise in the observation. The SNR can be approximated as the signal divided by the squareroot of the background whereas the contrast is the ratio of signal divided by the background [110]. The typical biological sample is a weak scatterer, hence, contrast is very low. Amplitude contrast is estimated to be only 4% [99] in a typical cryo-EM image at 300kV. Hence high resolution transmission electron micrsocopy (HR-TEM) relies on the introduction of phase contrast. In the second part of the thesis a new method for the generation of phase contrast images from TEM is introduced and tested on biological samples: integrated differential phase contrast scanning TEM (iDPC STEM).

1.0.1 Single Particle cryo-EM

1.0.1.1 The Transmission Electron Microscope

It is important for the general understanding of the image formation process that the optical setup of a TEM is briefly explained. Even in modern electron microscopes, the ray path still resembles the one of the light microscope (figure 1.1). At the very top, the electron source is - ordered by increasing coherence - either a tungsten filament, a LaB_6 crystal or a field emission gun (FEG). In all of them the physical principle of electron production is always the same: the kinetic energy of the electrons is increased by heating the source such that the kinetic energy is larger than the work function of the filament or crystal. Through anode/cathode pairs, the electrons are then first extracted and next accelarated (usually by 300kV for single particle cryo-EM). The following aperture increases the coherence of the beam and the condenser lens forms a parallel illumination. This ray bundle then interacts with the object. After being focussed by the objective lens again, electrons scattered to the outside of the beam are filtered out by the objective aperture and the projector systems adds a final magnification to the image. This image can then be recorded.[99]

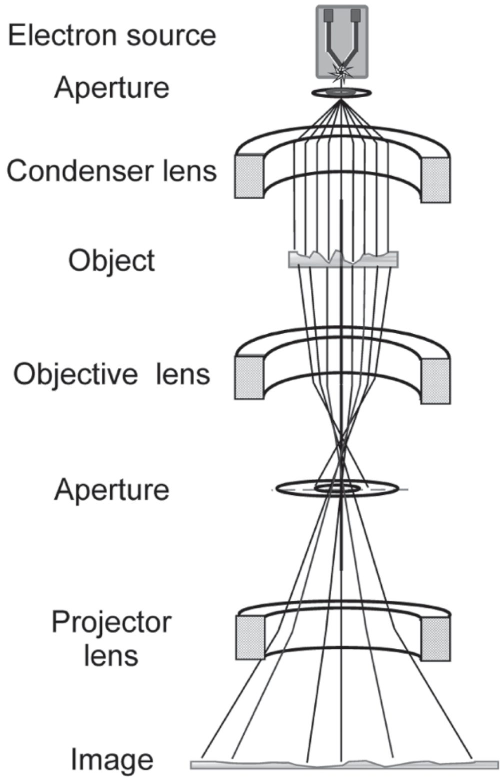


Fig. 1.1: Simplified Ray Diagram of a TEM[99]

There are two things that should be of superior interest when discussing contrast formation later on: firstly, electron microscopy is prone to abberations due to the imperfections of the electron lenses. The primary occuring are spherical abberations, chromatic abberations, astigmatism and coma[99][74]. Spherical abberations (C_s) originate from the distance an electron traverses through a lense from the optical center axis. Even though spherical abberation are used in the theoretical framework to describe image contrast - as will be discussed in the next section - it has been found useful to correct those to increase in absolute resolution [43, 54].
Secondly, the interaction cross section for light (bio)elements is larger for inelastic than for elastic scattering[59]. Consequently, these inelastically scattered electrons have a large energy spread, which is due to plasmon excitation in the sample [35]. This implies that chromatic abberations in the image forming optics are induced, which leads to the fact that those lower-energy electrons have another focal plane after they left the specimen, leading to hazy features in the detector plane (which is not in focus for those electrons anymore)[35][74]. A part of the inelastically scattered electrons is already filtered out by the objective aperture[99].

1.0.1.2 Phase Contrast and the Contrast Transfer Function

One aim of testing STEM for biological imaging in the third publication shown here was that the contrast formation mechanism is different and the hope arose that this limitation could be overcome. Contrast in TEM is achieved by the interaction of elastically scattered and unscattered electrons and the phase difference results from the path difference the elastically scattered electrons have with respect to the unscattered. The image formation process in a TEM can be described by fourier optics and here the notation of [150] will be used:
The transmission function describing the exiting wave from the sample is

$$\tau(r, z) = \tau_0 exp(i\phi) \tag{1.1}$$

with r being the two dimensional position in the image plane, z being the third dimension along the beam, $\tau_0 = \tau(r, 0)$ being the incoming wave front and ϕ being the phase shift, which depends on the specimen potential U:

$$\phi(r) = \pi\lambda \int dz'(U(r, z'))$$

as biological specimen are usually thin and do not contain strong scatterers, the weak phase object approximation (WPOA) can be applied: inspecting the last term in equation

1.1 by a first order taylor expansion and using $\phi << 1$ equation 1.1 [57]can be rewritten as

$$\tau(r,z) = \tau_0[1 + i\phi(r)] \tag{1.2}$$

As the intensity in the image plane is $\tau\tau*$, where the asterisk denotes the complex conjugate, such a weak phase object will have basically constant contrast. Namely, the path difference between scattered and unscattered electrons is just not different enough.

When the phase of the scattered beam experience an additional phase shift of $\pi/2$ the situation changes [99] and

$$I(r,z) = \tau\tau* \approx 1 - 2\phi(r,z) \tag{1.3}$$

This phase shift is usually introduced by manipulating the fourier pattern in the back focal plane (BFP) either by introducing a physical object (a phase plate [94]) or by using image distortions. The latter can be achieved by changing the contrast transfer function (CTF, i.e. the point spread function in the backfocal plane) through either defocussing or increasing abberations. The CTF of a TEM is

$$K(R) = sin(\frac{2\pi}{\lambda}(\frac{-\Delta\lambda R^2}{2} + \frac{C_s\lambda^3R^4}{4})) \tag{1.4}$$

with R being the reciprocal space coordinate, Δ being the defocus and C_s being the spherical abberation coefficient [150].
From equation 1.4 it becomes clear that the contrast in TEM relies on the defocus stronger than on the spherical abberation. However, the high spatial resolution signal fades with increasing defocus. TEM imaging is therefore always a compromise between high-resolution information and contrast. This is the major motivation for developing phase plates, which optically manipulate the BFP and introduce the phase contrast while being close to focus [19].

1.0.1.3 Sample Vitrification for cryo-EM

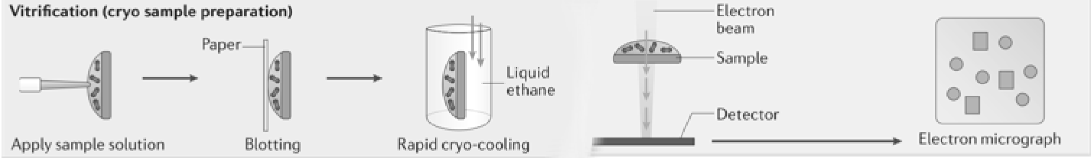


Fig. 1.2: Experimental workflow of vitrification The sample solution is pipetted on to a support grid and access liquid is removed by removing it with a filterpaper ("Blotting") to reduce the ice thickness. Next, and in a very quick manner, this grid is shot into liquid ethane. This sample is then imaged in the electron microscope. Modified from [109]

The sample for single particle cryo-EM needs to withstand on the one hand an extreme local energy density and on the other hand the applied vaccum in the microscope coloumn.

Therefore, vitrification was introduced by [29]. The sample is generated by rapid freezing of a solution of biomolecules on to a support grid. The molecules are sometimes stabilised by the addition of fixating agents, which modifies certain side chains [139]. Figure 1.2 shows the general idea of vitrification. A purified protein solution is applied to a support grid usually by pipetting. Then, the excess liquid is removed by blotting with a filter paper. This step basically determines the ice thickness on the final grid to a certain extend, which will be of importance later in the discussion of the results of the iDPC STEM. Right after the blotting step, the grid is plunged into liquid ethane. The cooling rate of liquid ethane is that fast that the ice does not crystalize but rather form a vitreous ("glass-like") structure [29]. As the pressure, the duration of the blotting procedures, the time between removing the blotting paper and the actual plunging as well as the environmental humidity are key paramters for the reproducibility of the process, this process is automated in most labs. On the other hand, automation of the full process, including the positioning of the droplets is still experimentally and one burden was just recently overcome by "self-blotting" grids [66, 107].

One thing should be emphasized here: the process of cooling is assumed to be extremely fast, as the vitrification of water happens at $10^7 K^s{-}1$. That would mean that to cool a sample by 100 K, one would in the slowest case need 10 μs. One major factor for this is apart from using liquid ethane the preferrable surface/volume ratio of a thin layer [48]. At this cooling rates, usually no relaxation of large and intermediate molecular motions can happen, as just very local motions happen on faster time scales as will be shown later. Therefore it is fair to assume that the conformational distribution found in solution is represented one to one in the vitrified sample. Based on this thinking, the first part of this thesis will focus on how this conformational information can be analysed from a cryo-EM dataset.

1.0.2 Single Particle Analysis (SPA)

Having acquired an image of the sample in the microscope, reconstructions of the single molecules can be calculated. The randomness of the grid preparation process implies not only that all the conformational states are represented in the data but also that - at least for a perfectly spherical object - all viewing directions are represented within the data. By applying reconstruction methods, the 3D volume can be calculated from their 2D projections, which was shown already in 1986 by *Klug and de Rosier*[26]. However, biomolecules are prone to being damaged by radiation. Hence, the images are acquired at very low doses and consequently have a very poor signal-to-noise ratio (SNR), which is the main obstacle of SPA. This is solved by averaging images in the same orientation together.

The first publications on computationally extracting and aligning such images go back to the late 70s and early 80s [146–148]. From this times, much has be done in the field, but one of the largest impacts on the computational side, was the introduction of Maximum-Likelihood Methods first by *Sigworth* [135] and later by *Scheres* [123] for the 2D-level.
The overall scheme is always the same: firstly, from the micrographs, single particles have to be extracted. This is at its heart a recongnition problem, which as of today is mostly solved by pattern matching algorithms, either with low-resolved models of the molecule as template or with gaussian blobs [114, 122]. These images then need to be aligned and classified. Alignment is still solved by an exhaustive search of the translational and rotational degrees of freedom in plane. One and actually the historically oldest way to classify cryo-EM images was based on fundamentally the same multivariate statistical methods as will be used in this thesis [147]. Multivariate statistical analysis (MSA) is based on that each $n \times n$-image can be seen as being a single experiment composed of $n \times n = N$ observations - i.e. random variables. Based on this statistical methods can be used to describe the differences between those "experiments". Whereas early work was based on correspondance analysis (CA), later principle component analysis (PCA) was used. CA is based on a ξ^2 metric, PCA uses the covariance matrix Σ of the experiments. Both matrices are symmetrical such that the eigenvector matrix Γ and the corresponding eigenvalue matrix Λ can be calculated [68]:

$$\Lambda = \Gamma^T \Sigma \Gamma \tag{1.5}$$

In the case of CA Σ would be the distance matrix. As will be described in the second publication (*Lambrecht et al., submitted*) in more detail, the eigenvectors can be used as the orthonmormal basis (ONB) of a new coordinate system. Classification of 2D images can hence be done by reducing the dimensionality of the images into a few basis functions and by clustering in this space [149].

The later introduced Maximum-Likelihood techniques [135] iteratively calculate a model of parameters describing the alignment parameters as well as the class affiliation. It optimizes the overall probability of the data describing this model in a Expectation-Maximization algorithm [28][123].These 2D classes can be used to evaluate the underlying data and to clean out non-particle or broken particle images. Also, these classes do not only represent different views on the molecule - i.e. different projection directions. Rather already on this level, structural heterogeneity can be differentiated [32].
From this cleaned dataset a 3D volume can be reconstructed, either by reprojecting class sums or single particles back into 3D. Again today, primarily Maximum-Lkelihood refinements are used [119]. Here, two more paramters are added to the model, which are the

left two eulerian angles. Even though also other 3D-Maximum-Likelihood implementation exist, the primary algorithm used is RELION [119][122], which increased significantly in speed since the implementation of CUDA kernels [73]. RELION developed the idea of the Maximum-Likelihood based refinement and classification further by using a bayesian maximum *a-posteriori* function. The *a-priori* information added to the model was that in fourier space, the scattering potential has to be smoothly varying, as chemical bonds do not impose discrete jumps in the density [120].

As the possibility to marginalize the class affiliation was already implemented in the 2D case, it was straight-forward to also allow the classification of 3D volumes, which basically just differs from the 2D case by the computational complexity [124]. Even though there are in the meanwhile also different flavors of 3D classification algorithms [85, 106], RELION is still the bench mark and will be primarily used throughout this thesis. As will be discussed now in greater details, the combination of 3D classification ($k > 1$, where k is the number of reference volumes) and 3D refinement ($k = 1$) algorithms has evolved a "standard" workflow in the field. In the work here, it will be shown that the methodology of applying MSA to 2D images is also applicable to 3D volumes. Furthermore in *Haselbach and Schrader (2017)* and *Lambrecht (submitted)*, the interpretation of the eigenvectors with respect to conformational motions will be described.

1.0.2.1 Computational Standard Workflow

The question is, after a first preprocessing phase (figure 1.3, upper region), how are 3D classification and 3D refinement combined to yield a high-resolution structure? The center part in figure 1.3 shows the ususal case: a small number of classes are used and in a sieve-like approach classes are thrown away after visual inspection. Usually this does not just imply two iterations but rather up to ten different "sieving"-steps. And in all of those steps particles, even if they contain valuable information, are discarded by the fact that the underlying conformational state is not populated enough to refine to high resolution. This strategy has some implications:

1. By choosing a low number of classes it is likely that particles belonging to neither of the main conformational states are collected in a dirt-class, even though those contain valuable information.

2. The uncertainty of the class affiliation at the beginning of the classification procedure is very high, hence the probability of doing a mistake right at the beginning is very high.

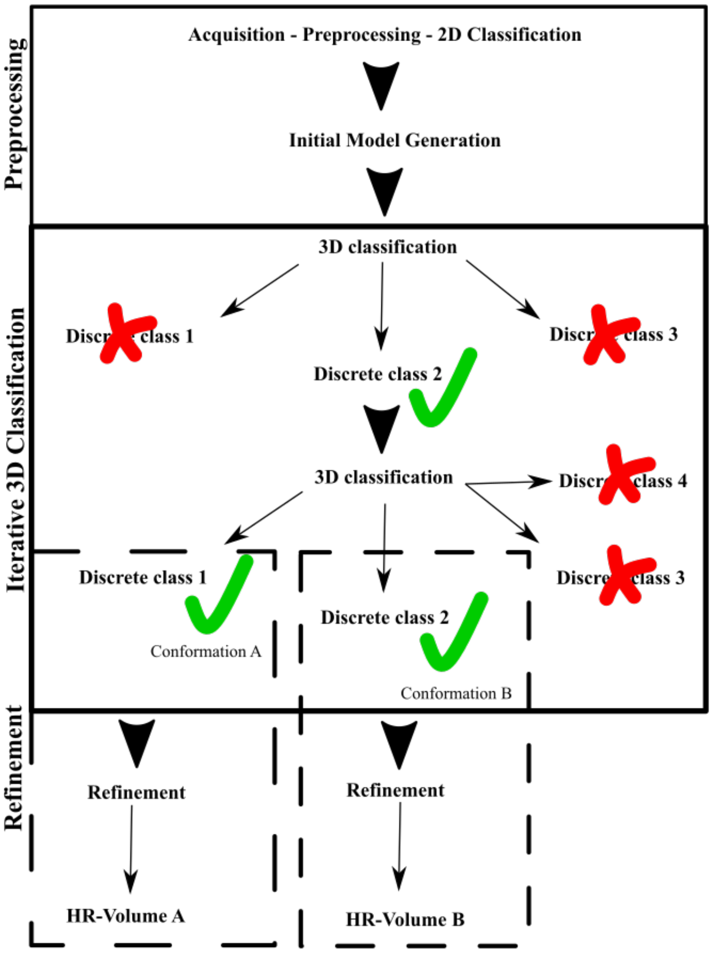

Preprocessing
Acquisition - Preprocessing - 2D Classification
Initial Model Generation
Iterative 3D Classification
3D classification
Discrete class 1
Discrete class 2
Discrete class 3
3D classification
Discrete class 4
Discrete class 1
Conformation A
Discrete class 2
Conformation B
Discrete class 3
Refinement
Refinement
Refinement
HR-Volume A
HR-Volume B

Fig. 1.3: Common refinement workflow for cryo-EM experiments After a initial preprocessing in which from the acquired images a initial model is generated, 3D classification is used to sort out dirt or uninterpretable images. Only a subset of the clases (namely usually only the ones with a high enough resolution) are retained. The final few classes having survived this "sieving" are then refined to high resolution (here, HR-Volumes A and B in two different conformations).

3. As will be described in the next chapter in more detail, the low number of classes discretize a continuos space to a very large extend. Consequently, the physical information of the pathways between the major minima in the conformational space is lost.

In the first two publications in this thesis, we advised a way to circumvent these limitations by applying a less discrete approach to the classification problem. We basically merge the two historically most prominent solutions to the classifcation problem: Maximum-Likelihood 3D classification, as it came up in the late 90s and Multivariate Statistical Analysis as it was first described in the early 80s. Here, we will use Maximum-Likelihood based classification to generate an ensemble of states, which will then be interpreted by multivariate statistics. The foundation of our description are the underlying thermodynamical principles of protein' motion, namely the glass-like state model. It will be shown in the next chapter, how the glass-like state model and the derived conformational landscapes influence the interpretation on important biophysical concepts such as binding and regulation. The overall aim of this part of the thesis shall be to describe how, using the landscape-mindset, such phenomena can be described by using an electron microscope.

1.1 Conformational Landscapes from Cryo-EM Experiments

> "The living organism seems to be a macroscopic system which in part of its behaviour approaches to that purely mechanical (as contrasted with thermodynamical) conduct to which all systems tend, as the temperature approaches the absolute zero and the molecular disorder is removed."
>
> *Erwin Schrödinger: What is Life?*

As described above, the major group of investigated molecules by cryo-EM are macromolecular complexes consisting of mostly polypeptides (e.g. the proteasome as studied here) or a mixture from polypeptides and nucleic acids (e.g. the ribosome). All of them are usually built up from many individual polypeptide chains which assemble together to form the mature complex. The three dimensional structure of peptides and nucleic acids is determined by the composition ot their bulding blocks, which successively fold and assemble to form the native form of the complex.

The process of folding is extremely interesting from a energetical point of view: the folding increases the ordering, which means it decreases the entropy. As folding is able to happen spontaneously, there has to be another driving force [97, 98]. Figure 1.4 shows the typical schematic picture of a folding funnel. The x-Axis shows the conformational coordinate, while the y-axis shows the free energy of the system. While the configurational entropy decreases during the folding process, the enthalpic contribution has to increase to gain free energy from this process. In the lower region of the funnel, where the native structure is located, this contribution comes mainly from the formation of specific interactions. In the upper part of the funnel, the chain is still too elongated to form these interactions and a random walk through this space would probably last forever. Here, another effect come into play, which helps forming a compacted form of the chain and thereby lowering the search space. The hydrophobic effect drives the hydrophobic regions of the protein to the center, whereas the hydrophilic ones come closer togehter and stronger (polarised and charged) interactions can be formed.

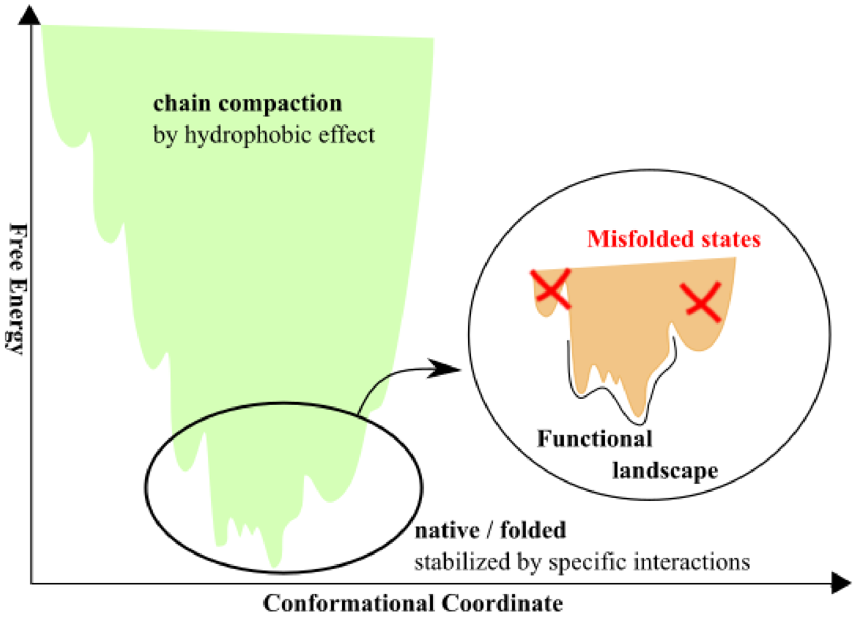


Fig. 1.4: Schematic drawing of a slice through the energy funnel of protein folding: shown in green is the overall folding funnel, in which the upper part has a high number of degrees of freedom and is folding is governed by the hydrophobic effect. In the lower region - governed by specific interactions - the native state is located. In the circle this folded, native state is zoomed in. It involves misfolded parts as well as the functional landscape. adapted from [97, 145]

Each configuration in this funnel is in equilibrium with all the others and can in theory be reached by a purely statistical process [79, 158]. This is still true when one is looking at the very bottom of the funnel, which covers the "native" configurations of the complex. This part of the landscape is in the literature referred to as the "functional landscape" [158], the part of the landscape which is biologically (i.e. functionally) relevant.

1.1.1 Conformational Dynamics of Proteincomplexes

Structural biologists look at the dynamics of a protein complex usually by describing its different conformations. These conformations are distinct stabilised configurations of the complex in the funnel shown above. The different conformations of a protein usually reflect different functional states. A very illustrative example is myglobin, which is the oxygen carrier in the muscle. Dioxygen is bound via a porphyrin-coordinated Fe(II) atom. The sixth free binding position of the iron is taken by a histidin. The "bound" conformation of myoglobin positions the iron such that oxygen is allowed to bind. When the conformation changes to the unbound state, the histidin pulls the iron out of the porphyrin ring and the dioxygen dissociates. Each of these states are well stabilised minima in the conformational landscape of the myoglobin, as will be explained now in more theoretical depth.

1.1.1.1 The Glass-Like Model of Conformational Changes

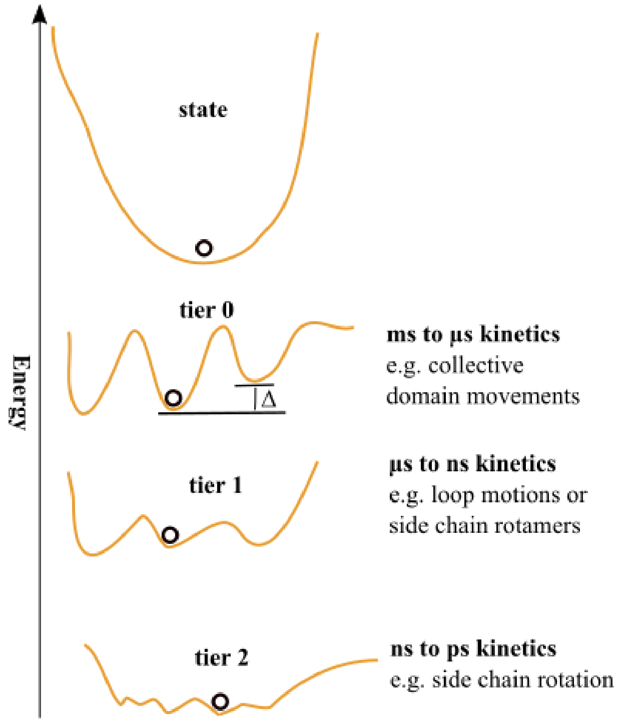


Fig. 1.5: The dynamics of one particular state is hierarchically ordered: the dynamical contributions to the overall broadening of the state is described by different tiers, which describe the imposed ruggedness on different scales of energy and time. The decreasing energy differences between minima and maxima in the landscape lead to shorter kinetics on the corresponding tiers. The differences between minima (Δ) can be described by 1.6. Adapted from [47, 60]

Myglobin is a historically important example. Frauenfelder, Parak and Young showed by elegant laser specroscopy in 1988 that proteins in solution are inhomogeneous and they proved that this is due to several conformational substates [46]. Frauenfelder elaborated further on this idea and found analogies in the physical chemistry of glasses [47][9]. From these he introduced the idea of a rugged energy landscape, which is a hypersurface in the configurational space of all atoms. In that publication also the idea of structuredness on different lengthscales as well as energyscales (figure 1.5) was introduced. This structuredness implies a certain hierarchy ("tiers"[47]) which is similar to the ones discussed above. To each configuration of a domain, several configurations of secondary structures are possible and within each of those, different rotamers of a certain side chain might be accesible.

Each conformation is in equilibrium with all the others and can consequently be reached by overcoming a certain activation energy. Again lower-order tiers (0,1) will have higher activation barriers as higher-order tiers, as moving a whole domain is more expensive then just e.g. rotating a phenole-moiety. However, the probability with which one molecule can move from one conformation to the other is directly related to all activation energies on all paths from one conformation to the other. Phenomenologically, or experimentally it is hard to asses these energies for a single molecule. It is easier to monitor how many complexes exist in each state. The difference in Gibb's free energy which is necessary to come from one conformation to the other is then dependent on the ratio of single molecules in the two states and is given by the Boltzman-factor

$$\Delta G = -k_B T ln(\frac{f_1}{f_0}) \tag{1.6}$$

where, k_B is Boltzmann's constant, T is the absolute temperature and f_1 and f_0 are the observation frequencies of molecules in state 1 and 0, respectively. This concept is a pure physical one, based in statistical thermodynamics, generally describing the distribution of a group of particles in different energetic states. Such a group of particles is referred to as an esemble and the ensemble view on to protein biochemistry is a general feature of this part of thesis[5, 79].

1.1.1.2 Recognition by Ensemble

From the concept of rugged energy landscapes of proteins developed by Frauenfelder, important considerations can be drawn for systemically interesting features of proteins: recognition and regulation. Recognition and regulation build up the "intelligence" of the cellular interior up to the organismic level. Recognition of substrates and binding partners was first described in 1894 by Emil Fischer, who coined the "key-and-lock" model [41]. In 1958 Koshland *et al.* [76] then defined the induced fit model, in which the substrate basically enforces the tight binding conformation in the protein. This concept overlived another fortyone years until the discussions about energy landscapes led to the development of "conformational selection" [86].

All three concepts are sketched in figure 1.6. In contrast to key-and-lock and induced fit, conformational selection takes into account that the ensemble of one protein exists in different conformational states and binding does only happen in preferred ones. The difference to the induced fit model is that the bound conformation is already existing in the ensemble. In the limit of an extreme unflexible protein, which would only exist in one native conformation able to bind its partner, this model simplifies again to Fischer's lock-and-key model (figure 1.6, middle reaction of panel c). As the binding imposes new enthalpic binding contributions and restricts the degrees of freedom in which the polypeptide chain can move, binding usually stabilises the conformational state at least locally [17, 52, 53].

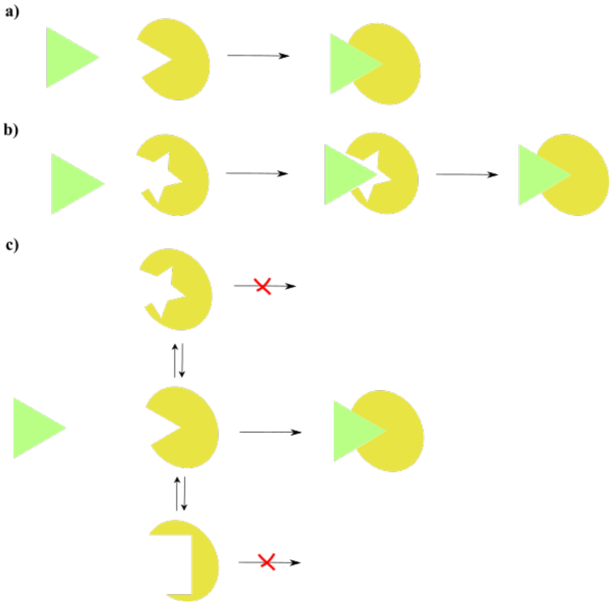


Fig. 1.6: Overview on different binding mechanisms: a) shows the classical key-and-lock model by Fischer, where the binding site is already predefined to just fit the substrate in. b) sketches the induced fit model, which is split in a two-step-mechanism, first binding the substrate weakly, which secondly alters the binding site and allows full recognition. c) shows the conformational selection model, where the three conformers are prototypically for the full conformational ensemble.

1.1.1.3 Allosteric Regulation by Ensemble

Allostery originates from the greek "*a*llos", which simply means "other" and refers to a two-side regulation, one side being the catalytic active site and the other being the effector side, where the regulator binds to the complex far away from the active site. The examples for this feature especially of metabolic proteins are numerous and the impact on cellular regulation are consistently large. One famous example is the aspartate transcarbamoylase (ATCase), which is the first enzyme in the pyrimidine biosynthesis leading to the formation of N-Carbamoylaspartate from Carbamoyl phosphate and aspartate under the release of a phosphate ion [87]. The ATCase is allosterically down-regulated by CTP - cytidine triphosphate -, which itself is an endproduct of the pyrimidine biosynthesis. Clasically two models described allosteric regulation: the symmetry model of Monod, Wyman and Changeaux (MWC [91]) and the sequential model of Koshland, Nemethy and Filmer (KNF [77]). Not only from the appearence of Daniel Koshland in this an the preceding chapter the overlap with recognition shall be apparent. Both models assume distinct states R and T of the molecule, which stand for relaxed (able to bind) and tense (unable to bind). In the KNF model, which focusses more on the effector binding, the effector binds preferentially to one conformation, which undergoes the transition from tense to relaxed, consequently. By a cooperative mechanism, this transition happens then also in the adjacent subunit. The MWC model in contrast assumes that the two conformations T and R exist in equilibrium and there are different probabilities for an effector to bind to one of them. From there on, cooperativity leads to the propagation of the binding event. Figure 1.7 shows the difference of the two models by taking into account all possible permutation of states (i.e. a partition function) for this tetramer [63].

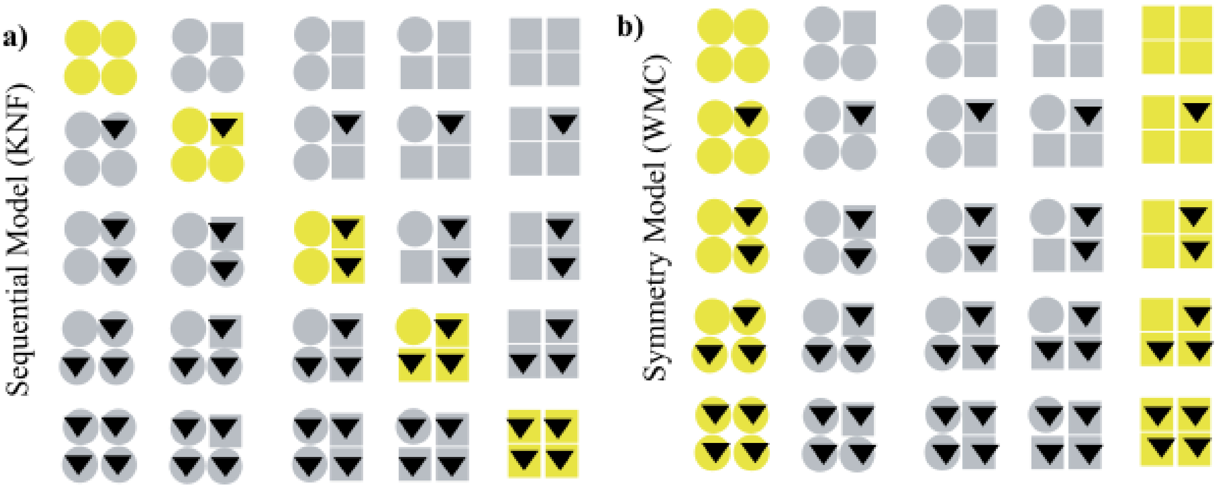


Fig. 1.7: Partition function of a tetrameric protein with a binding partner: Each subunit of the complex can exist in in T and R state (square and circle, respectively). The substrate binding (triangle) happens consecutively. In a) the sequential KNF model is shown. Upon binding the teramer sequentially changes from one to the other conformation by a induced fit model. In b) the WMC model the substrate binding can occur in all yellow states with different affinities. Both models do not take the never-yellow partitions into account, which is in accordance with the EAM model. Adapted from [63].

Both models were well aware of the existence of all the intermediate states but tried to phenomenologically describe kinetic experiments. Surely important findings were not covered by this models, as the one of Cooper and Dryer [83], who showed that allostery might appear without changing the overall structure of the complex but by just adjusting entropy, which broadens the conformational distribution rather than changing the overall look of the conformation. The MWC model already has the discussion about energy landscapes in mind, when it assumes both conformations T and R to be present all the time. With the previous chapter about energy landscapes in mind, one easily comes to the discussion that binding on the one hand happens in different regions of the energy landscape - i.e. to different states of the ensemble - with other probabilities. Cooper and Dryer furthermore showed that the binding event leads to changes of the energy landscape which might imply cooperative energies up to a few kcal per mol [92]. That implies that an effector binding to a complex changes the enrgy landscape and shifts preferred conformations, leading to either states which bind substrates or binding partner more effectively (positive cooperativity) or to states which do not (negative cooperativity). Based on these observations Hilser *et al.* [63] formulated an ensemble based model of allostery: the ensemble allosteric model (EAM). While the classical models describe only a part of the full configurational space, the EAM describes all the states in figure 1.7. to learn something about the transition from one major state to the other. Pan *et al.* showed the shift of the ensemble as the micro-mechanism of allostery for the Dihydrofolatereductase from *E. coli*. On the extreme one could then even ask where allostery starts and ends, as any interaction or modification will affect the energy landscape to a certain extend. This question "Is allostery an intrinsic property of all dynamic proteins?" was discussed in Gunasekaran *et al.* [53]. These findings underpin the importance of studying the energy landscape of biological complexes, as the principal structure of such a molecule would always be the same. Studying the functional energy landscape of a protein is probably the only key to understanding the full picture of influences of binding partners and modifications.

As will be outlined in the next section, the microscopic (in contrast to the averaged macroscopic constant) observation of phenomena such as allostery is restricted to a few methods, where the primary one used in the past was NMR. It was already briefly introduced that NMR is limited by the size of the molecule. Consequently, detailed observations on the mechanisms underlying regulatory effects are lacking in thre literature. Here, cryo-EM might help the molecular biphysics field to gain deeper insights in how such phenomena work at larger scales. A first glimpse will be given in the first publication in this thesis (*Haselbach and Schrader, 2017*).

1.1.2 State-Of-The-Art Visualization of Conformational Landscapes

Richard Feynman once said: "Everything that living things do can be understood in terms of the jigglings and wigglings of atoms" [40]. The dynamics of a complex harbor the information about its function and its regulation. The question is now: how can one visualize or measure the "jiggling and wiggling"? And how to put this back in the context of the full conformational space? Here, only methods applicable to the lower-order tiers will be discussed, as the scope of this thesis is cryo-EM.

1.1.2.1 Molecular Dynamics Simulations

In the discussion about energy landscapes above, it was implicitley assumed that the main driving force of all movements is thermal energy. The multiplier $k_B T$ in equation 1.6 has the dimension of an energy [1] and is nothing else than the energy which is available by brownian motion at a certain temperature T. This brownian motion can be simulated and applied to an experimental model of a molecule, which is either atomistic or more frequently grained [116]to a certain extend by treating the building blocks of this molecule by classical (Newton's) mechanics. By applying potential energy functions (the force field) one can calculate the forces acting on a atom or grained particle in a short period of time Δt. One then updates the position of the atoms and calculates the next Δt iteratively until the system converges [69]. Thereby, the dynamics of certain regions, conformational changes and even molecular recognition can be visualized and analyzed. The limitations are clearly at the computation intensive calculations, which makes compromises between calculation time, molecular size and the degree of coarse-graining applied necessary [103], especially for macromolecular complexes.

There are several ways to calculate free energies from the simulations which would be necessary to build up an energy landscape and those are extensivley reviewed elsewhere [102]. One particularly interesting method is umbrella sampling [70]. In a nutshell umbrella sampling applies an additional umbrella potential to the simulation, which enforces a particular reaction coordinate to be sampled. The difference in free energy is calculated along the reaction coordinate in snapshots. By summing over all energy differences from one snapshot to the other, the pathintegral is calculated and the free energy difference of the sampled transition is the result. Of course, this is not the full energy landscape as discussed above, but the information is most valuable to judge e.g. conformational transitions from a open to a closed state or the binding of a ligand. Umbrella sampling has some parellels to the algorithm presented in this part of the thesis and will be discussed again later.

1.1.2.2 X-ray crystallography

The first works discussing dynamics of protein complexes and especially allsotery used X-ray crystallography, where the protein is translated into a crystal lattice by chance and its diffraction is recorded upon an incident X-ray beam. E.g. the fundamental works by Max

---

[1] The Boltzmann constant has the dimension of $J \times K^{-1}$. Hence $k_B T$ has the dimension of Joule, which is the SI unit of an energy

Ferdinand Perutz on hemoglobin function and notably allostery [104] where done using X-ray crystallography. Deciphering dynamics is challenging using crystals: firstly, one crystal hardly contains more than one conformation. That restricts one experiment usually to one conformer. Secondly, there are only a few very well stabilised (tier 2-3) conformers, which will crystallise at all, and thirdly, even if a protein inside a crystal allows enough degrees of freedom to see (as in Perutz' case for oxygen-binding) dynamics, they are just locally, as the global dynamics are restricted by the crystal contacts. Consequently, crystallography is for sure the method capable of the highest resolutions, but also not best suited to decipher the dynamics of especially large macromolecular complexes [125]. On top of this, building up a full landscape from crystallographic experiments is not possible.

1.1.2.3 Spectroscopical Approaches to Dynamics

Measuring dynamics with specroscopy rather than 3D stuctures is another way to look at conformational landscapes. The clear advantage of such methods is that the native [2] in-solution dynamics are accesible rather than the dynamics in a non-native state, e.g. a crystal. Here, two often used methods are described: fluorescence spectroscopy with resonance energy transfer (FRET), and nuclear magnetic resonance specroscopy (NMR) [61].

FRET is only one variety of classical photonic spectroscopic methods including absorbance and quenching which can today be used as single-molecule techniques. FRET uses two fluorescent labels on the protein to measure the distance dependent energy transfer between the acceptor and the donor [128]. By this studies on e.g. the folding pathway and the conformational dynamics of protein complexes were done on the single-molecule level leading to a description of trajectories through the energy landscape[105]. The downside is clear: the gained information is a one-dimensional path trough the multidimensional conformational space, which is barely interpretable without any structural information.

Nuclear magnetic resonance spectroscopy in contrast analyses the pertubation of spin couplings of the nuclear spins in a magnetic field. These resonance properties are used in medical imaging (magnetic resonance imaging, MRI) as well as in spectroscopy. In spectroscopy the chemical environment of a nucleus is probed and the informations are treated as a 3D distance-geometry problem, which can be solved by additional restraints [155].

Different types of experiments where designed, including so called dispersion experiments, which pushed the field of deciphering structure with dynamics further [7]. Those methods are extensivley reviewed in the literature and out of scope for this text. Even if in the meanwhile dynamics information of of large macromolecular complexes such as the GroEL chaperonin and the ribosome where feasible [62], those are rare and the natural limitation of NMR to resolve the signals coming from different nuclei are the major limitation when it comes to larger complexes. Hence, NMR is the method of choice when it comes to small to medium-size polypeptides and nucleic acids also for intrinsically disordered proteins []. As the full structural and dynamical spectrum of the protein in solution is accessible, free energy landscapes are derived quite frequently in the literature [126] and especially the discussion in the previous chapters abot different binding and allostery models is primarily fueled by the NMR field. However, in contrast to the single-molecule techniques such as FRET, the information is not accessible for each individual molecule, as also in NMR an averaging occurs.

---

[2] where "native" does not imply that the extremely crowded cellular environment is in any method taken into account

1.1.3 Computational Approaches for Distangling Conformations in Cryo-EM

Deciphering of macromolecular motions is important for the in depth understanding of their function. In chapter one it was explained that cryo-EM basically captures a snapshot of the in-solution state. Hence, the full conformational space as it exists in solution is captured and the observation frequencies directly reflect the energetical proportions via equation 1.6. It is currently under research primarily by the MD community if the speed of vitrification is really sufficient to not imply constraints on the resolution of the landscape (oral communication Prof. Dr. Erik Lindahl, Stockholm University).

One "lab rat"-molecule taken as a benchmark for algorithms deciphering the different conformational states in the past was the ribosome, and a specific dataset by Joachim Frank which shows distinct tRNA movements and the accompanying motions of subunit ratchetting, was used throughout all publications until about 2013. Elmlund *et al.* [37] showed the performance of different approaches on this dataset in a comparative review. Table 1.1 shows an overview over the main algorithms for conformational sorting.

The first three approaches include simple and also more advanced clustering algorithms which cluster conformational heterogeneity based on different metrics. More statistically advances methods are shown in rows four and five. Elad et al. [32] also in the end applies a clustering but does this on a pre-adjusted space from Principal Component Analysis (PCA), as will be also used here. This analysis is basically similar to feature extraction approaches in face recognition. Still, this approach works in 2D and is therefore sensitive to poor SNR. With Bootstrapping a completely different approach was introduced, which finally works in 3D but has its limitations in the non-optimizing statistics. The introduction of Maximum-Likelihood-Methods as already described above for the refinement of a single structure also impacted the analysis of dynamics. In this approach k randomized volumes are used as starting models and refined competetively, where the class affiliation is used as another hidden variable and is consequently marginalized in an expectation maximization algorithm[118]. Based on these findings, focussed classification, where the rigid part of the molecule is masked out and signal subtraction, where the rigid part is even cut away from the reference images, where developed [3].

Table 1.1: Comparism of different algorithmical approaches. Modified from [37]. R = Ratcheted, R+Efg = Ratcheted with Efg bound, UE = Unratcheted with an occupied E-Site, UL1 = Unratcheted with free E-site and L1 domain out, LSU = Large Subunit only,

Approach	R	R+Efg	UE	UL1	LSU
Supervised classification [51], 2004	x		x		
Cluster tracking [50], 2006	x		x		
Clustering in 3D [130], 2010	x		x		
Multivariate statistical analysis [32], 2008	x			x	
Bootstrapping in 3D [50], 2006	x		x		
Maximum Likelihood [121], 2007	x	x	x		x

All of these approaches imply a discretisation (*"classes"*) of the continuous space described in the chapters before. In the work presented here, a continuos approach is achieved by using PCA in 3D, i.e. in voxel-space. The idea to this was first introduced by [138]. They used PCA on a set of bootstrapped volumes, by which they where able to extract significant heterogeneity-describing eigenvectors. These eigenvectors are linear approximation to the heterogeneity in the dataset, which as discussed in the scond publication in this work, might correspond to conformational changes.

In the last years, two approaches significantly gained importance. Firstly, the usage of elastic network models like normal modes analysis (NMA). This approach goes back to Tama *et al.* [142], but never found usage until just recently [142]. Additionally, Nakane et al. [95] published multi-body-refinement as an additional feature of RELION. Here different mobile domains are approximated as rigid bodies and refined independently, calculating relative shifts by a PCA-based algorithm. This approach is clearly influenced by the first publication in this work. The main difference is the coarsing of the analytics of dynamics to rigid bodies and the consequent reduction in calculation time, which enables an iteration-wise update of the calculation.

However, all of these approaches do only contain a one-dimensional view on a movement, i.e. they visualize only one principal movement on a very coarse-grained level. A more "landscape-like" approach was described already a few years earlier: Fischer et al. [42] showed in a fundamental study that the energy landscape of retro-translocation of the ribosome can be described from hand-sorted single-particle data. They could even show temperature-dependent shifts in the distribution, which underpinned the interpretation of the ribosome as a brownian machine. Another more automated alogorithm where single-particle images were analyzed by manifolds-embedding [23, 45] also yields a landscape representation of the conformational movements, but until today no working implementation was made publicly available. Anyways, in these cases the images are directly projected on to a manifold describing the conformational landscape. One path through this landscape corresponds to a non-linear trajectory through real-space. It can due to the lack of a accessible software only be assumed that this approach is limited to the low singal-to-noise ratio in the individual images. A thorough discussion of the differences is given in the supplemental materials to *Lambrecht (submitted).*

1.1.3.1 Contributions in this work: the CowScape algorithm

In the last chapter three important developments or mindsets were introduced for the analysis of dynamics by cryo-EM:

- Maximum-Likelihood based methods allow to classify a vast number of states throughout a dataset.
- Principal Component Analysis on the 3D level allows to find eigenvectors which correspond directly to the heterogeneity in the dataset and allow to describe this continuously.
- The distribution of particles with respect to these movements can be used to describe a higher-dimensional landscape, where the number of particle observations directly correlate to the energy differences between them.

The first two pubications presented here combine these three concepts and apply the methodology to the problems of the analysis of dynamics in general, allostery in particular, and drug binding as an important application. They built up on the scripts developed in my Master's thesis which already tried around with a similar dataset.

The publication *Haselbach and Schrader et al. (2017)* shows how this analysis can be done by classical methods of cryo-EM: classification and interpretation of B-factors. On the other hand it was the first benchmark of our methodology. We calcultated an ensemble of volumes by exhaustive classification and applied PCA to them. This yields interpretable eigenvectors in terms of dynamics and that we can use these to span a new orthonormal basis in which we can then describe the conformational energy landscape of the human 26S proteasome with and without a bound inhibitor. In this first version, these energy landscapes where still interpolated by triangulation.

In the second publication, we applied a more physical description of the landscape by exchanging the triangulation by kernel-density-estimation with a gaussian kernel function.

We benchmarked our approach on more examples, namely the interpretation of compositional heterogeneity in the Anaphase Promoting Complex (APC), the conformational motions of the 26S proteasome's lid complex and effector binding to the TRPM1 ion channel. Both publications will show the impact of the method on to the interpetation on binding events and allostery in an ensemble view and especially the second publication yields the full description of the algorithm and its implementation.

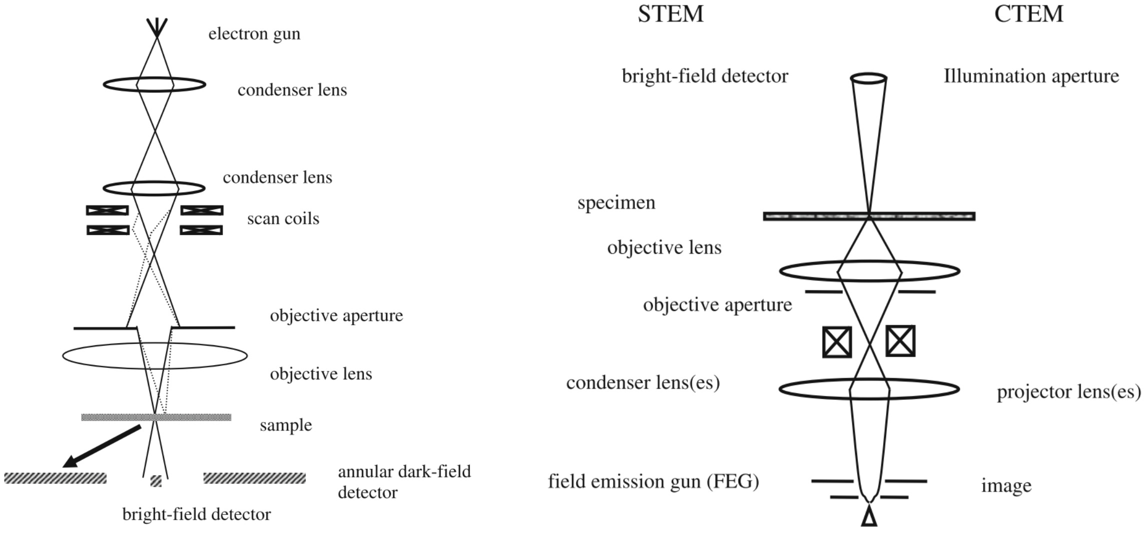
1.2 Single Particle Cryo-EM with iDPC-STEM

Contrast and resolution are the limiting factors for structure determination. It turned out that for biomolecules both are related to radiation damage and therefore with dose. Just 6 years after Ernst Ruska published "Das Elektronenmikroskop" (the TEM) in 1932 [75], Manfred von Ardenne published his first theoretical ideas for a scanning transmission electron microscope ("Das Elektronen-Rastermikroskop", STEM). However, STEM and biology was not really a success-story until now. This is basically due to the extreme dose one needs for sufficient contrast on organic molecules - to an extend that would never allow SPA on biomolecules. In this part of the thesis, I will try to introduce a new imaging mode for STEM - namely integrated differential phase contrast (iDPC) and why it might allow to use STEM in single particle analysis.

1.2.1 The Optics of a STEM

Figure 1.8(a) shows the general ray path inside a STEM. The aim of the optics is to first produce a small spot - the electron probe - which can be described as being "converged" or "demagnified". Therefore first a few condenser lenses focus the spot on to a final aperture. Before this objective aperture, the scan coils are located. By changing the circular electric current elliptically, these coils enable to scan the probe over the sample plane. The final lens, the objective lens does most of the magnification, therefore its abberrations are practically dominating the shape of the incoming wavefront. After the sample, the beam magnifies again on the detector. In STEM the detection modes are basically divided into bright- and darkfield techniques. Resembling their photonic counterparts, darkfield-techniques collect electrons far away from the optical axis, whereas brightfield STEM uses the electrons in the center.

Comparing the ray path to a conventional TEM (CTEM), one does find that there is a reciprocity: if one considers only elastic scattering, namely all the electrons have equal energies, the processes are time reversal and the path the electrons take are exactly the same in both cases. This implies that CTEM and STEM should have the same imaging properties but with a fundamental difference. That is that the optics in the STEM case


electron gun
condenser lens
condenser lens
scan coils
objective aperture
objective lens
sample
annular dark-field detector
bright-field detector
STEM
CTEM
bright-field detector
Illumination aperture
specimen
objective lens
objective aperture
condenser lens(es)
projector lens(es)
field emission gun (FEG)
image

(a) (b)

Fig. 1.8: Optics of a STEM (a) The general optical path consists of a condenser lens system, the scan coils, which are allowing to scan the converged beam over the sample and an objective lens, forming the final probe. Dependent on the detection angle either a brightfield or a darfkfield image is acquired. (b) The optics in a STEM and a CTEM are basically equivalent, just turned around. That implies that most of the optics in the STEM case are before the sample plane and in TEM it is the other way. Both images taken from [101]

are all in front of the sample and in the CTEM they are mostly behind (see also figure 1.1). Considering now inelastic scattering in the sample, STEM turns out to be less sensitive to it, because in TEM this inelastic scattering contributes significantly to chromatic abberations [35]. Hence, as the inelastic mean free path is significantly shorter than the specimen thickness, STEM allows to image much thicker specimen, without being limited by noise through inelastic events [112].

The resolution determining factor in any STEM setup is the diameter of the probe d_{Probe}[67]. An important parameter for STEM imaging is therefore the angle formed by the converged spot and the optical axis, the convergence angle α (see figure 1.11 left part):

$$d_{Probe} = \frac{\lambda}{\alpha} \quad ; \quad \lambda = \text{De Broglie wavelength} \tag{1.7}$$

There is a second reason, why STEM is perfectly suited for thick specimen, apart from the "blindness" towards inelastic scattering. Tuning the convergence angle steeper and steeper, the focal depth T increases significantly:

$$T = \frac{d_{Probe}}{\alpha} = \frac{\lambda}{\alpha^2} \tag{1.8}$$

The effect is nicely depicted in figure 1.9, which shows a simulation of defocus series for darfkfield STEM. Importantly, STEM has the best contrast transfer in focus, which is a significant difference to CTEM. Hence it is very appreciable to gain a very large region of stable focus, as this implies that also the maximum contrast transfer happens at this point and no additional modifications to the backfocal plane are necessary to introduce contrast like it is done in conventional cryo-EM. It should become clear in the next section that this is especially useful for some biological use cases.

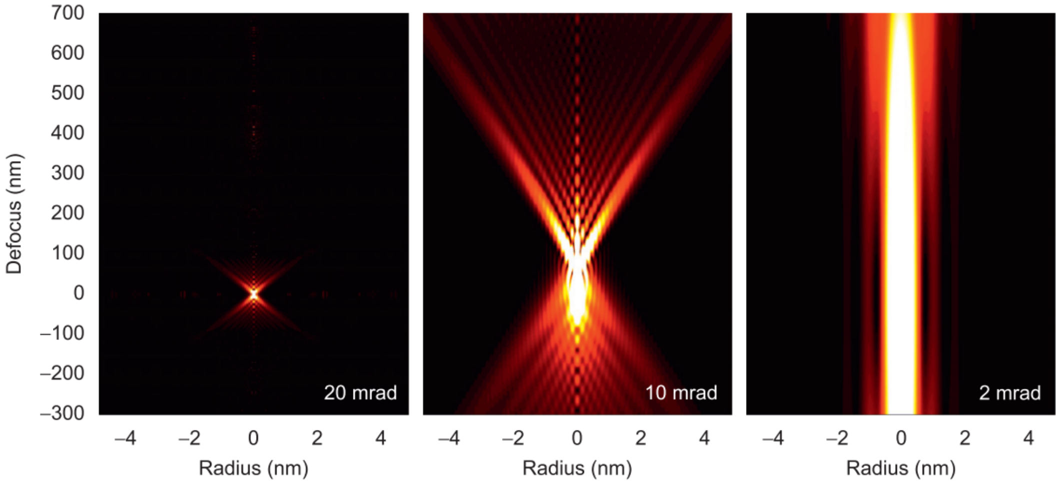


Fig. 1.9: The focal depth depence on the convergence angle The point spread function is simulated for different convergence angles. At 2 mrad convergence angle more than 1 μm of stable focus can be achieved, whereas the focal depth decays quadratically according to equation 1.8, when increasing the convergence angle. [64]

1.2.2 STEM and Imaging of Biological Molecules

The history of biomolecular STEM started in the 80s when negative-stained filaments where visualized by darkfield STEM [39]. A third flavour of STEM was then used on unstained biomolecules: electronic energy loss spectroscopy (EELS). EELS in contrast to the imaging methods analyses the inelastically scattered electrons, which fly close to the optical axis and can therefore be detected together with a darkfield-STEM image [39]. Additionally, Wall and Hainfeld [151] showed that the contrast of darkfield detectors is sensitive enough to perform mass measurements on single DNA strands. This application became later superfluous as the mass spectrometry field advanced.

Buban et al. [10] tried to find suitable imaging conditions for biological imaging on a high angle darkfield setup. They reported that at low-dose conditions, image artifacts where overwhelming the image and at least 100 $\frac{e^-}{\text{Å}^2}$ where necessary to resolve an image of an inorganic specimen. Interestingly, they discuss an important fact: STEM images are usually taken with large oversampling, i.e. the projected pixel size is much larger than the probe diameter. For single particle recordings that would imply that only a few particles per micrograph would be acquirable, which leads to a much larger acquisition time for a data set.

Another example for the imaging of single particles with a STEM was just recently published by Elad et al. [33]. They acquired vitrified ferritin with a darkfield setup at a dose which was only slightly larger than the one Buban et al. [10] used. They were able to resolve single zinc- and iorn binding sites within the protein surface of the molecule (figure 1.10). However, the protein structure itself was not even refined to the secondary-structure level (third isosurface level in figure 1.10), which was firstly due to the fact that only a small number of particles was processed and secondly, that darkfield STEM does not produce sufficient contrast to resolve light, biological molecules well.

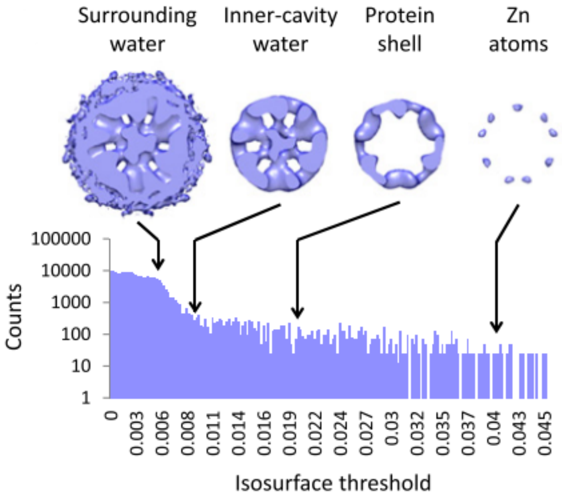


Fig. 1.10: Detection of single metal ions in a protein Elad et al. [33] showed that they are able to resolve single binding sites of Zn-ions in vitrified ferretin usin DF-STEM. However, the protein structure is only on minor resolution.

Much more successfull was in contrast the usage of STEM for the tomography field. Here, ususally brightfield and darkfield signals were combined. In a groundbreaking publication, Wolf et al. [152] showed that they can make use of the extended focal depth of STEM and image the full volume of a *A. tumefaciens*. Afterwards they could also show that by the quantitative STEM contrast the amounts of phopshporus [153] and calcium [154] in intracellular storage compartments could be quantified. Last but not least, they made a very intersting observation when comparing the image acquisition with CTEM: H_2 formation through radiation damage became visible in STEM after applying a much larger cumulative dose than in CTEM. This raises the question whether STEM would probably allow to image single particles at much lower doses as was described in the previous publications, since they all used darkfield techniques. The downside of these techniques is, that only a minor fraction of the electrons passing through the specimen is scattered to such high angles. Hence, the ration between effectively collected electrons and total dose is very small. Using brightfield STEM on biological molecules, would allow to detect much more electrons. However, conventional brightfield imaging is not sensitive enough to gain sufficient signal [33, 35]. Hence, another method using a larger fraction of the total dose with a sufficently high contrast would be necessary. Due to this fact we investigated the recording of differential phase contrast from the brightfield disc as one option. With this we aimed to on the one hand, gain the an in-focus imaging and a improved damaging mechanism, while having enough contrast to resolve biomolecules at resolution comparable to CTEM.

1.2.3 (integrated) Differential Phase Contrast

Phase contrast in CTEM is introduced by varying the defocus, which leads to an additional phase shift for electrons that have not passed through the sample. As for STEM all the electrons go through the beam probe, the mechanism of contrast generation is different. Differential phase contrast uses the difference in space, which originates from the beam being deflected while traversing through the object. DPC STEM was already formulated by Dekkers and de Lang [27] and Rose [113]. The fundamental difference to other STEM methods is that a split detector is used, as it is shown in figure 1.11. However, at that time this detector design was challenging, and it needed a few years until DPC was first brought to the same atomic resolution, as one is used to from STEM. *Shibata et al.* reported in 2012 that they have recorded the local electric field of a magnetic specimen [96, 132].

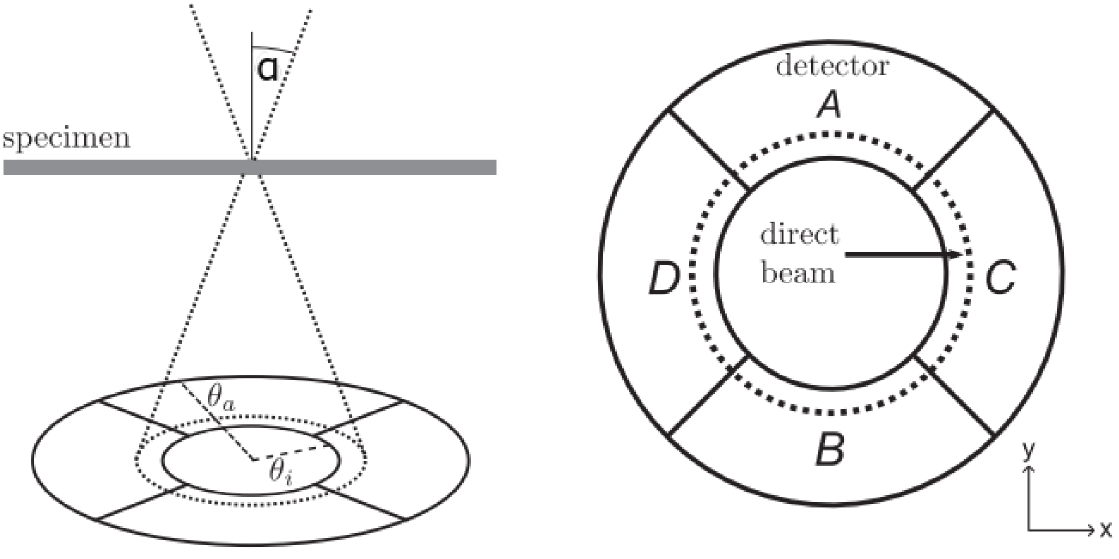
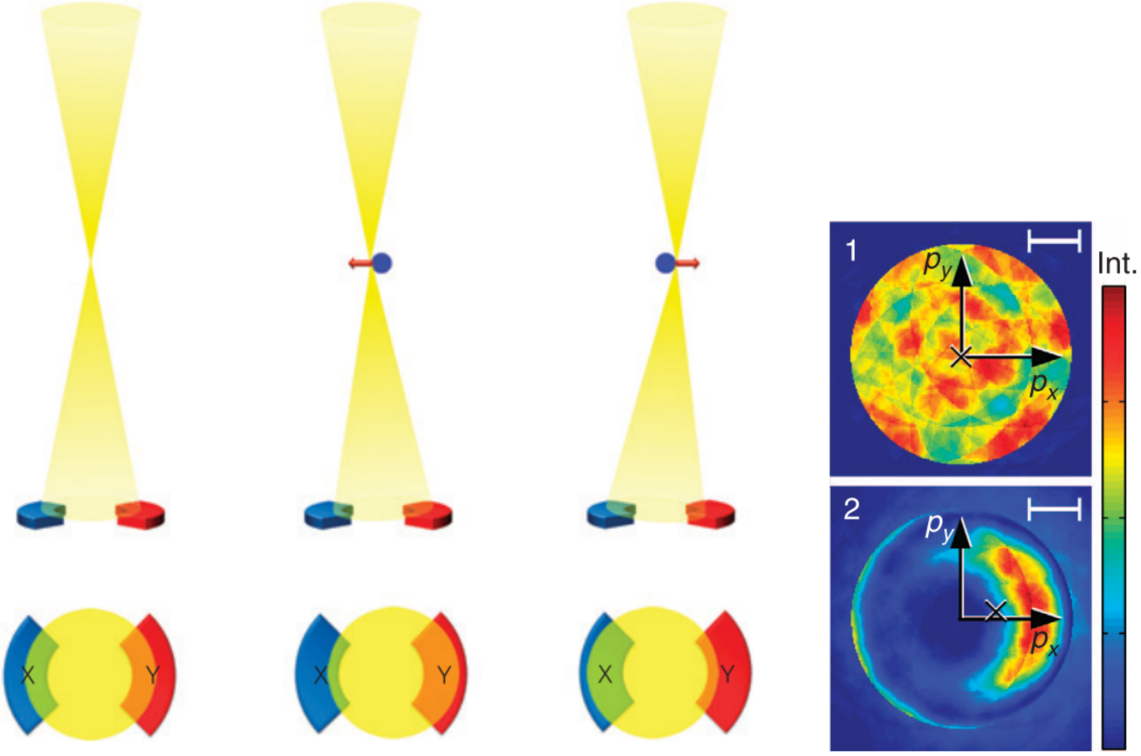


Fig. 1.11: Detector Setup for iDPC STEM The left side of the image shows the incident beam on the specimen, which demagnifies under a convergence angle α. Then the beam magnifies again on a split detector, which is shown again with the unperturbed beam in the right part of the image. Modified from [88]

The principle of DPC is depicted in figure 1.12(a) for the 1D-case. As we are looking at the 2D projection, the real detector is composed of four quadrants, as it is shown in figure 1.11. The observable is the shift of the intensity distribution in the brightfield disc (the so called ronchigram, figure 1.12(b)).

The electron ronchigram undergoes a momentum transfer due to the underlying non-zero electric field. The momentum transfer is proportional to the underlying Lorentz force. According to Ehrenfest's theorem, this still holds for the expectation value of the momentum transfer [93]. The expectation value of the momentum transfer relates to the shift of the center of mass in the ronchigram upon applying the electric field of the specimen, and the center of mass in fact can be approximated via the differential signal of the split



(a) (b)

Fig. 1.12: Working Principle of Differential Phase Contrast (a) In a 1D case, when no charge is present, the beam is aligned as such it is illuminating both detectors equally. At the next scan position, the beam approaches a charge and becomes deflected to the right detector more than to the left, such that the subtraction $I(X) - I(Y)$ is negative. At the next scan position the sign changes. [132] (b) The image seen on the detector while traversing from the first to the second scan position equals a shift of the intensity of the electron rochigram [93].

detector [80]. Using the notation of [81] and the detctor enumeration of figure 1.11 it follows that

$$\begin{aligned} I^{DPC_x}(r_p) &= \frac{\pi k_{Bf}}{2\sqrt{2}}(I^{Q_D}(r_p) - I^{Q_C}(r_p)) \\ I^{DPC_y}(r_p) &= \frac{\pi k_{Bf}}{2\sqrt{2}}(I^{Q_A}(r_p) - I^{Q_B}(r_p)) \end{aligned} \tag{1.9}$$

r_p is the scan position and k_{Bf} is the diameter of the brightfield disc. Both images together can be linear combined into one vector image

$$\vec{I}^{DPC}(r_p) = I^{DPC_x} \cdot \vec{e_x} + I^{DPC_y} \cdot \vec{e_y} \tag{1.10}$$

According to the nature of the underlying shift of the ronchigram, this is the projected representation of the electric field E of the specimen. As

$$E = -\nabla V \tag{1.11}$$

it follows that by integrating the vector image $\vec{I}^{DPC}(r_p) = \nabla I^{iDPC}$ the electric potential is derived, which directly scales with the atomic number, allowing even the "light" biological elements to be imaged[80, 81]. Lazić et al. [81] derived the same expression from imaging theory, showing that iDPC imaging yields a linear representation of the electric potential:

$$F\{I^{iDPC}(r_p)\}(k) = CTF(k, W(k)) \cdot F\{\sigma V(r_p)\}(k) \tag{1.12}$$

the operator $F\{\}$ denotes the Fourier transform, σ collects all the constants, k is the reciprocal space ccordinate and $W(k)$ is a function describing the detector geometry. W(k) is the only deviation the CTF has from ideal center-of-mass imaging [80]. The CTF of iDPC is shown in figure 1.13 and resembles the common features for contrast transfer in any STEM:

- The maximum contrast transfer happens in the very low resolution regime and drops fast to high resolutions
- Apart from the very out of focus case, there are barely zero crossings, which is in contrast to CTEM where the CTF is sinusoidal
- At higher convergence angles, the focus might even not be enough to cover full molecular machines (about 80 Å for the 30 mrad convergence angle simulated here).

Until today, there is little experimental evidence on the performance of iDPC STEM. However, as the ratio of detected electrons per full electron dose is very large (as one is

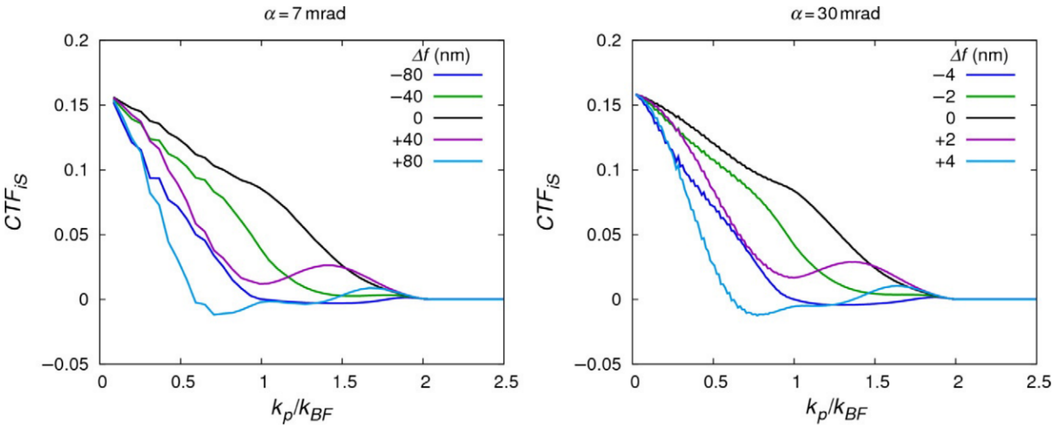


Fig. 1.13: Theoretical iDPC CTF The theoretical CTF for the first term in equation 1.12, where $W(k)$ is chosen to account for a four-fold symmetrical detector, is shown as a function of defocus from the disc with best contrast. The ratio k_p/k_{bf} is just one possible choice for k. [80]

observing changes in the brightfield disc) and the contrast at least theoretically let even the biological relevant molecules become visible, it was assumed that the imaging of radiation sensitive specimen might be finally feasible by a high-resolution STEM technique. Zeolites are crystalline materials which are characterized by their porous physical properties. One unit cell is generally composed out of a cation which is coordinated by four oxygen atoms donored from silicate or aluminate ions. Those are prone to radiation damage at doses higher than 5000 $e^-/\text{Å}^2$ [12]. Carlsson et al. [12] and Yücelen et al. [156] have shown that the imaging of zeolites up to the sub-Å level is possible with iDPC-STEM at doses comparable to 1000 $e^-/\text{Å}^2$ resolving even the oxygen atoms, which were until now hidden by means of conventional STEM techniques. This opens up the field of imaging radiation sensitive metal-organic-frameworks (MOFs).

However, this dose might be potentially even too high for radiation sensitive biomolecules. Therefore, the aim of the third publication *Lambrecht and Riedel, in preparation* was, to find suitable imaging condition for biomolecules optimizing the parameters dose, convergence angle (which scales with resolution) and focus. This would allow to overcome the thickness limitations in cryo-EM and additionally, to overcome the need to introduce phase contrast through image distortions as described above. Thereby the quality of the resulting EM maps would hopefully increase and make interpretation of the atomic structure more reliable. Furthermore, chances are high that due to the increased axial contrast intracellular imaging would be feasible which could pave the way for intracellular visiual proteomics.

2 Long-range allosteric regulation of the human 26S proteasome by 20S proteasome-targeting cancer drugs

Cited as: Haselbach and Schrader (2017)

Own contribution:

- Analysis of the conformational landscape
- Manual model building and refinement
- Analysis of B-factor distribution
- Analysis of symmetry breaking
- Analysis of the dynamical path through the molecule
- Figure 1 b-f
- Figure 2 a and b
- Figure 3 in parts

nature COMMUNICATIONS

ARTICLE

Received 30 Nov 2016 | Accepted 7 Apr 2017 | Published 25 May 2017 DOI: 10.1038/ncomms15578 OPEN

Long-range allosteric regulation of the human 26S proteasome by 20S proteasome-targeting cancer drugs

David Haselbach[1,*], Jil Schrader[1,*], Felix Lambrecht[1], Fabian Henneberg[1], Ashwin Chari[1] & Holger Stark[1]

The proteasome holoenzyme is the major non-lysosomal protease; its proteolytic activity is essential for cellular homeostasis. Thus, it is an attractive target for the development of chemotherapeutics. While the structural basis of core particle (CP) inhibitors is largely understood, their structural impact on the proteasome holoenzyme remains entirely elusive. Here, we determined the structure of the 26S proteasome with and without the inhibitor Oprozomib. Drug binding modifies the energy landscape of conformational motion in the proteasome regulatory particle (RP). Structurally, the energy barrier created by Oprozomib triggers a long-range allosteric regulation, resulting in the stabilization of a non-productive state. Thereby, the chemical drug-binding signal is converted, propagated and amplified into structural changes over a distance of more than 150 Å from the proteolytic site to the ubiquitin receptor Rpn10. The direct visualization of changes in conformational dynamics upon drug binding allows new ways to screen and develop future allosteric proteasome inhibitors.

---

[1] Department for Structural Dynamics, Max-Planck Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany. * These authors contributed equally to this work. Correspondence and requests for materials should be addressed to A.C. (email: ashwin.chari@mpibpc.mpg.de) or to H.S. (email: holger.stark@mpibpc.mpg.de).

The proteasome holoenzyme is composed of the catalytic core particle (CP, 750 kDa) and in addition either one or two molecules of the regulatory particle (RP, 900 kDa), to form the 26S (1.6 MDa) and 30S (2.5 MDa) proteasome holoenzyme[1], respectively. The CP consists of four co-axially stacked rings of seven distinct α and β subunits, whereas the RP consists of an AAA+ ATPase assembly (Rpt1–6) and 12 non-ATPase subunits (Rpn1–3 and Rpn5–13)[2] (Fig. 1a,b). Its main task is the degradation of polyubiquitinated substrates. Consequently, cellular homeostasis including diverse functions such as the control of the cell division cycle, transcription regulation, protein quality control, apoptosis and many more pathways, depends on its proteolytic activity[3]. For the mechanistic understanding of its cellular activities and its therapeutic targeting in disease, the elucidation of high-resolution structures of the proteasome holoenzyme in complex with drugs are therefore of paramount importance. In particular, the structural impact of 20S inhibitors on the proteasome holoenzyme remains entirely elusive.

Here we determined the high-resolution structures of the human 26S proteasome holoenzyme bound to the chemotherapeutic Oprozomib and its apo form using single particle cryo-electron microscopy (cryo-EM) and additionally determined the dynamic properties of the proteasome. We identify a clear restriction of the conformational landscape of the holoenzyme upon drug binding using a newly developed method to map the conformational and energy landscape of the 26S proteasome. From the energy landscape, we can infer that drug binding introduces an energy barrier minimizing the possibility of the RP to rotate on the CP. This rotation however is necessary for deubiquitination of the substrate[4] and its translocation into the proteolytic core of the proteasome. We thus find a long-range allosteric regulation that spans more than 150 Å from the location of drug binding towards the ubiquitin recognizing and regulatory regions in the 19S subunit.

Results

Structure determination of the inhibited 26S proteasome. The treatment of the proteasome holoenzyme with 20S inhibitors leads to stabilization and suppresses disassembly[5]. To address the question how 20S inhibitors affect the proteasome holoenzyme structure, we initially considered the conformational motions of the RP, which have been described previously (Fig. 1g)[4,6,7]. In essence, two conformational states have been described: a non-rotated- and a rotated-state in which the non-ATPase segments of the RP are rotated by up to 25° around the long axis of the 20S CP. This motion is coupled to the ATPase part of the RP with its Rpt4/Rpt5 coiled coil contacting the ubiquitin receptor Rpn10 (Supplementary Movie 1). Treating the 26S proteasome with either the drug Oprozomib or the natural product Epoxomicin, which both belong to the epoxyketone class of 20S proteasome inhibitors, we found the proteasome holoenzyme to be stabilized in the non-rotated state (Fig. 1h). The treatment of the 26S proteasome with 20S proteasome inhibitors leads to a stabilization of polyubiquitinated substrates[8], which may remain bound to the proteasome holoenzyme. We therefore investigated if polyubiquitinated substrates are accumulated in our proteasome preparations upon inhibition by drugs using anti-ubiquitin western Blot analysis. As shown in Supplementary Fig. 1, we have found no profound accumulation of polyubiquitinated substrates in our inhibited 26S proteasome preparation over non-treated controls. This indicates no correlation between the accumulation of polyubiquitinated substrates and proteasome inhibition. As a consequence, the allosteric regulation of RP rotation described in this manuscript is exclusively dependent on inhibitor binding (Supplementary Fig. 1). We then proceeded to reconstruct the three-dimensional (3D) structure of both the non-inhibited and Oprozomib-inhibited 26S proteasome at 4.8 Å/3.8 Å resolution, respectively. To achieve this, we utilized the identical image processing and classification protocol (Supplementary Fig. 2) for the reconstruction with and without Oprozomib. Notably, in both non-inhibited and inhibited structures the structure of the proteasome holoenzyme strongly resembles (real-space correlation >0.9) the two other high-resolution EM structures reported recently[8,9]. While additional density is absent in the β5 active sites of the non-inhibited structure (Fig. 1f), a clear density for Oprozomib is visible in the β5 active sites of the inhibited structure (Fig. 1e, Supplementary Fig. 3).

Model of the Oprozomib-bound human 26S proteasome. The Oprozomib–26S proteasome structure (Supplementary Tables 1 and 2) exhibits well-defined densities for the entire proteasome holoenzyme (with the exception of Rpn1, Supplementary Fig. 4), showing numerous amino-acid side chains in most parts of the molecule and relatively small variations in local resolution (Fig. 1c). Specifically, regions encompassing both β subunit rings and the α subunit ring of the CP bound to the RP, as well as the ATPase of the RP are resolved at a resolution range of 3.5–4.5 Å. The local resolution of the structure decreases with increasing distance from the CP (Fig. 1c) to the upper regions of the proteasome lid structure, which appear to be the most mobile regions within the proteasome. Owing to the visible side-chain densities, we built an accurate atomic model of the holoenzyme with the unequivocal assignment of amino acid registry for regions with B factors smaller than 110 Å^2 (Fig. 1d, Supplementary Fig. 5). The B-factors of the model correlate well with the local resolution differences visible in the EM density map (Fig. 1c,d).

Long-range allosteric effects of drug binding to the proteasome. We utilized only a relatively small subset of particles from both non-inhibited and Oprozomib-inhibited datasets (4% and 12%, respectively) to obtain reconstructions at high resolution described above, which are nearly identical in conformation. Thus, the differences between proteasomes with and without drug are expected to be manifested in the particle images that did not contribute to the high-resolution structures and more likely affect the conformational space adopted by the proteasome. To harness this information, we studied the conformational variations in the RP in a quantitative manner by extensive 3D classification (Supplementary Fig. 6), calculated the corresponding energy landscape (Fig. 2) and analysed how drug binding modifies this energy landscape of the RP. Briefly, we focused on the conformational variability in the RP by aligning all 26S 3D structures with respect to their 20S part only and applied principal component analysis (PCA) to reveal the eigenvectors as major modes of motion of the RP (see Fig. 2, Methods, Supplementary Note 1 and Supplementary Movie 1). The known particle number for each conformation allows the transformation of a conformational landscape into an energy landscape that describes the complexity of RP motion in a comprehensive, quantifiable manner. It also enables the direct visualization of the changes in proteasome dynamics upon Oprozomib binding (Fig. 2b, Supplementary Movie 1). According to this analysis, the energy landscape of the non-inhibited proteasome is rather flat and allows the RP to sample a large number of conformations making use of thermal energy only. In contrast, drug binding considerably decreases the available conformational space and creates an energy barrier for the molecules making it less likely to reach a fully rotated state. This is

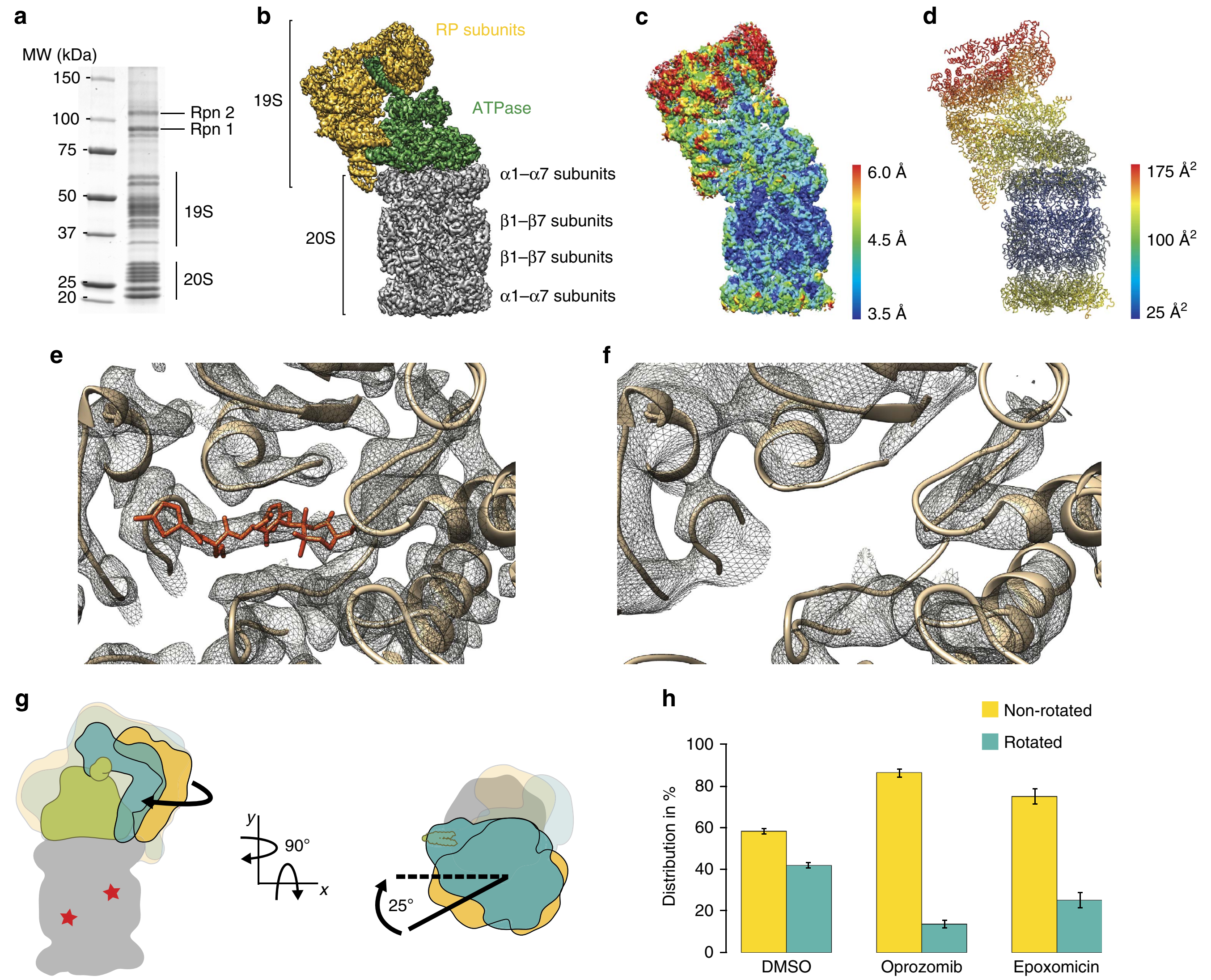


Figure 1 | Effect of Oprozomib. (**a**) SDS-PAGE of purified human proteasomes. (**b**) Surface view of the human Oprozomib-bound 26S proteasome cryo-EM density map at 3.8 Å resolution. The CP (20S) subcomplex is depicted in grey, the AAA+ ATPase subcomplex in green and the remaining RP (19S) components in yellow. (**c**) Local resolution map of the structure shown in **b** Each part of the density is coloured according to the local resolution as specified in the colour bar. The resolution ranges from 3.5 Å (blue) to 6 Å (red). (**d**) Atomic model of the complete 26S proteasome. The model is coloured according to the B-factor distribution. B factors range from 25 Å^2 (blue) to 175 Å^2 (red). (**e**) Close-up view of the Oprozomib binding site in the β5 subunit of the CP. The Oprozomib model is coloured in red the CP subunits are shown in brown. (**f**) Close-up view of the empty Oprozomib binding site in the β5 subunit of the CP. The Oprozomib model is coloured in red the CP subunits are shown in brown. (**g**) Schematic representation of the two major rotational modes of the RP reveals a rotation of the RP along the long axis of the 26S proteasome as indicated in a cartoon representation as a visual aid. (**h**) Histogram of the relative distribution of 26S proteasome particles found in either the rotated or the non-rotated state which can be modified by epoxyketone inhibitor binding. The control dataset (DMSO) reveals an almost balanced distribution with ~41% of the particles in the rotated state. The number of particles in the rotated state is significantly reduced upon Oprozomib (~13% rotated) or Epoxomicin (~25% rotated) binding. Error bars displaying s.d. indicate a high reproducibility based on data from three independent proteasome preparations (*n*=3).

manifested along several eigenvectors (data not shown) and indicates that although a subset (14%) of 26S proteasomes is observed in a rotated state upon Oprozomib inhibition, the maximal amplitude of rotation attained is only 20°. At this state of RP rotation, Rpn10 approaches the coiled coil of Rpt4/5, whereas non-inhibited 26S proteasomes rotate the RP up to 25° with high probability, where the Rpt4/5 coiled coil reaches the Rpn10/Rpn9 interface (Fig. 1g, Supplementary Movie 1).

Our data clearly suggests that in addition to a competitive, irreversible inhibition of the proteolytic activity[10], CP inhibitors also affect the conformational landscape of the proteasome holoenzyme by restriction of its capability to adopt the rotated state. The most surprising feature about the allosteric regulation elicited by Oprozomib on the 26S proteasome is that the inhibition reaction is converted to a conformational signal that is relayed over a distance of over 150 Å.

Discussion

In this paper, we have determined high-resolution structures of inhibited and non-inhibited endogenously purified 26S proteasomes. Inhibition of the proteasome allows a structure of higher resolution to be determined. Importantly, by employing a novel image analysis procedure, we can show that inhibitor binding causes a long-range allosteric regulation of the proteasome holoenzyme. Surprisingly, both species bind similar amounts of

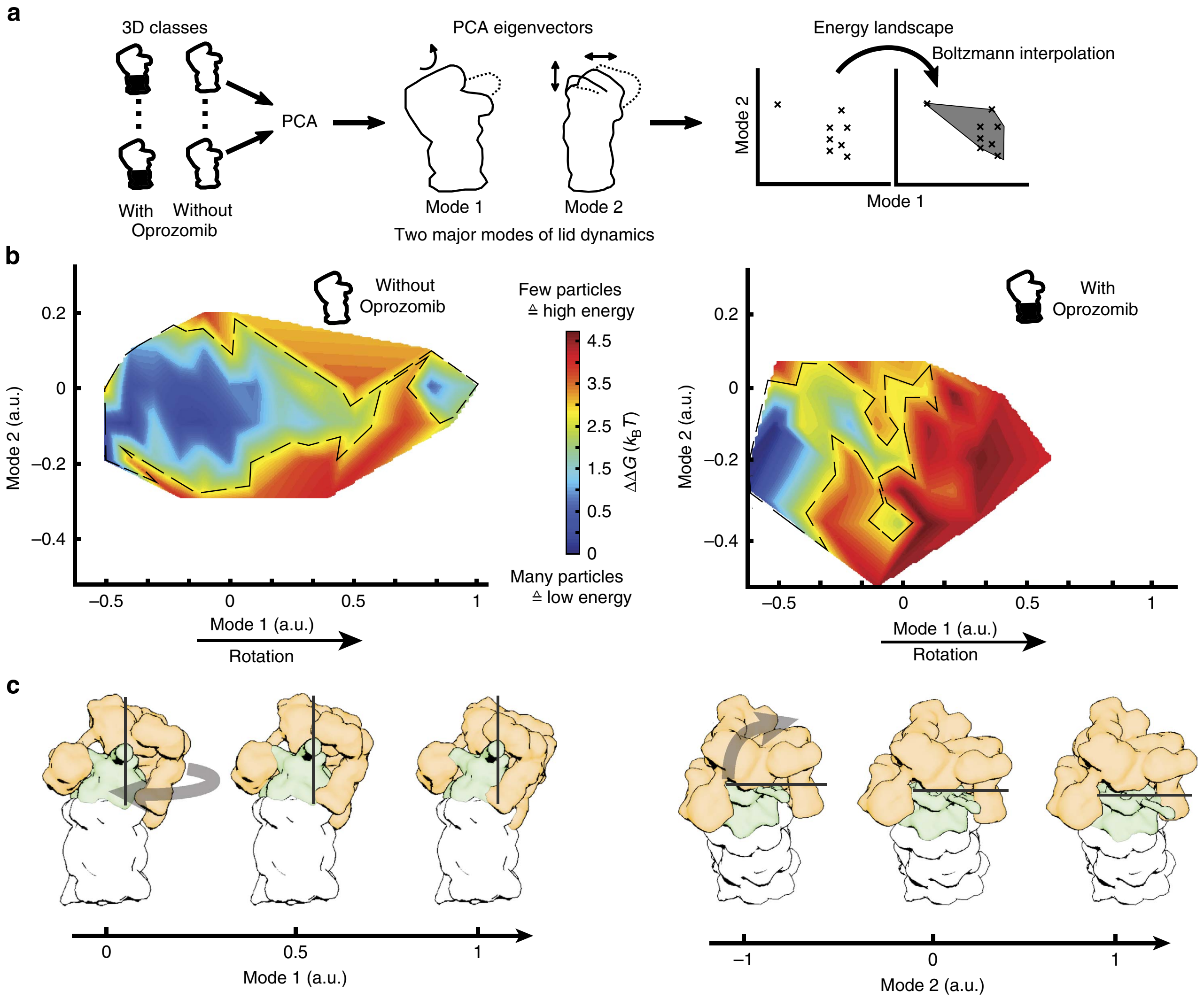


Figure 2 | Modification of the energy landscape upon Oprozomib binding. (**a**) Schematic representation of the strategy to obtain an energy landscape, which corresponds to the conformational states adopted by the proteasome holoenzyme: Principal component analysis of a large number ($n = 346$) of 3D structures yields eigenvectors corresponding to the major modes of conformational variability. The first mode corresponds to a lid rotation with respect to the 20S subunit and the second mode to a more complex rotational rearrangement of the lid (Supplementary Movies 1 and 2). The relative particle numbers, which can be obtained for the various conformational states, are used to calculate the energy landscape according to Boltzmann's law. (**b**) Energy landscapes with and without Oprozomib are depicted. Without Oprozomib, the energy landscape is rather broad and flat which allows proteasomes to sample a wide range of conformations without facing a significant energy barrier. In contrast, upon drug binding, well-separated minima can be observed next to a significant energy barrier (red) which restricts the conformational space that can be sampled by the proteasome. The 3.8 Å resolution structure was determined from particles belonging to this proteasome conformation in a local energy minimum (dark blue). (**c**) Graphical representation of the movement modes. The average structure has been segmented in its three subcomplexes and fitted as rigid bodies in the eigenmodes.

polyubiquitinated substrates as shown by western Blot analysis. We can therefore conclude that the observed long-range conformational changes of the proteasome upon drug binding are exclusively due to the binding of the inhibitor Oprozomib. This raises the interesting question about possible determinants that enable such long-range regulation. While an accurate description of this will require many more high-resolution structures, mechanistic biochemistry and molecular dynamics simulations, a detailed analysis of the Oprozomib-inhibited structure allows a first glimpse on the components that might be involved in signal relay and amplification (Supplementary Fig. 4). The chemical inhibition signal is located in the β-ring of the proteasome and needs to be relayed over several tiers (CP alpha ring, two ATPase rings) to the top of the lid structure of the RP. Two correlated criteria, such as decreased local resolutions in the EM reconstruction (Fig. 1c,d) and correspondingly regions of higher B-factors in the atomic model (Supplementary Fig. 5) indicate conformational mobility, which we used to monitor the signal pathway from the inhibition site to the upper parts of the RP. Moving vertically in tiers from the site of inhibition (the 20S β5 active site), an asymmetric B-factor elevation on α-subunits 1, 2, 3 and 4 becomes evident (Fig. 3a). This asymmetric conformational mobility of the α-subunits is further relayed in two directions onto the RP: 1) A lateral transmission onto the adjacent RP subunits Rpn5 and 6 and 2) a vertical transmission to the next higher tier (Fig. 3b), the ATPase. At present, which pathway of transmission of conformational mobility (the lateral or vertical) is dominant remains unclear and speculative but both are most likely synergistic and mutually potentiate each other. Irrespective of the precise pathway of the propagation of conformational dynamics, the ATPase subunits (Rpt2, 3 and 6) adjacent to Rpn5 and 6 and directly above the

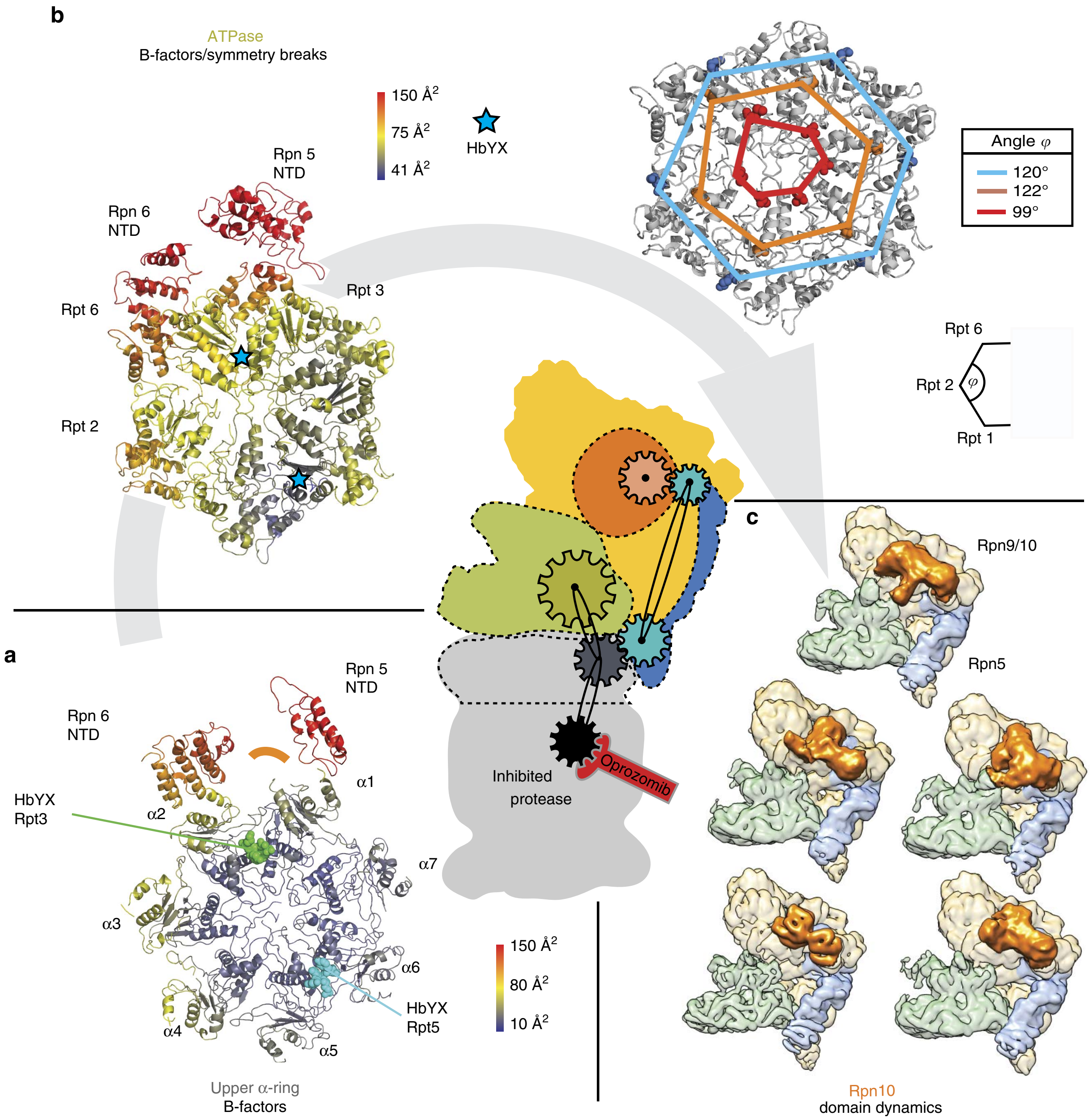

b
ATPase
B-factors/symmetry breaks
150 Å^2
75 Å^2
41 Å^2
HbYX
Rpn 5
NTD
Rpn 6
NTD
Rpt 6
Rpt 3
Rpt 2
Angle φ
120°
122°
99°
Rpt 6
Rpt 2
φ
Rpt 1
c
Rpn9/10
Rpn5
a
Rpn 5
NTD
Rpn 6
NTD
HbYX
Rpt3
α1
α2
α3
α4
α5
α6
α7
HbYX
Rpt5
Inhibited
protease
Oprozomib
150 Å^2
80 Å^2
10 Å^2
Upper α-ring
B-factors
Rpn10
domain dynamics

Figure 3 | Representation of potential bi-directional signalling pathway. (**a**) Model of the upper α-ring and the adjacent regions of Rpn5 and Rpn6 are shown coloured according to their B-factors. For better orientation, the HbYX motifes of Rpt3 and Rpt5 are depicted in their binding pockets (green and cyan) and the orange arch corresponds to the position of Rpt3's region with highest B-factors. A significant increase in B-factors can be seen on the outer parts of α2, α3 and α 4 which all are adjacent to the Oprozomib binding subunit β5. The regulatory subunits Rpn5 and Rpn6, which directly bind the α-ring, show the highest B-factors, presenting a potential communication path indicated by flexible parts. (**b**) The model of the ATPase ring and the adjacent parts of Rpn5 and 6 are shown coloured by B-factors. The perspective onto the segment is the same as in **a**. A similar distribution of B-Factors as in **a** can be seen. Subunits Rpt2, 6 and 3 show increased B-factors as they are adjacent to the more mobile α-subunits in **a**. Similarly, the regulatory subunits Rpn5 and 6 show very high B-factors. Right: to analyse the symmetry of the ATPase, three conserved amino acids have been chosen in all six. The Cα-atoms have been connected and the inner angles of the resulting hexagons enclosed by Rpt6, Rpt2 and Rpt1 have been calculated. Whereas the C-terminal region of the ATPase forms a perfectly regular hexagon, the symmetry is clearly broken in the N-terminal region near the centre of the ATPase. This deviation from perfect symmetry indicates the required motion for ATPase activity and is consistent with an increase in model B factors. (**c**) Focused classification on the Rpn9/10 (orange) region and subsequent refinement in RELION revealed different conformational states for the receptor regions. Only exemplary conformations are shown. Similar conformations can be found by focusing on Rpn5 (blue) only.

α-subunits 1, 2, 3 and 4, exhibit an asymmetric B-factor elevation. In addition, a gradual vertical deviation from perfect six-fold symmetry is evident in the ATPase (Fig. 3b), which is even visible in a different sugar conformation of the bound nucleotide in Rpt2 (Supplementary Fig. 7). Furthermore, the motion of Rpn5 is correlated in the next vertical tier with the local conformation variations of the ubiquitin receptor Rpn10 (Fig. 3c, Supplementary Movie 2). Focused 3D classification reveals considerable mobility of Rpn10, which in some conformers is in direct contact to the coiled-coil region of Rpt4/5. In other conformers, Rpn10 completely detaches from the deubiquitinase Rpn8 (Supplementary Movie 3).

In conclusion, we have provided evidence that 20S proteasome inhibitors allosterically regulate the RP to adopt a non-rotated (presumably inactive) conformation. To our best knowledge, allosteric effectors that have such an extended reach are unprecedented and have not been described yet. We have presented evidence that the RP subunit Rpn5 most likely is the lever, which conveys the information about inhibition of the proteolytic active site over a long-range distance. This is supported by the notion that Rpn5 is an essential lid protein (so far described in yeast and plants)[11]. Previous studies at lower resolution have shown that the rotated (presumably active) state is preferentially adopted by RP binding to a slowly degraded substrate[4], or by the addition of slowly hydrolyzable nucleotide[6] to the yeast holoenzyme. Ligand binding to either the RP or the CP has distinctly opposing effects on the conformational motion of the RP, which suggests a feedback regulation between RP and CP through Rpn5. This knowledge enforces the notion that development of 20S inhibitors with novel binding sites and inhibition chemistries will have a profound impact on the allosteric regulation of the proteasome holoenzyme. Vice versa, inhibitors that target the conformational variability of the RP will profoundly influence the catalytic activity of CP active sites. We postulate that allosteric regulation of large macromolecular complexes, by catalytic active site small molecule inhibitors is a general feature.

Methods

Materials. Standard chemicals were obtained from Sigma Aldrich (Taufkirchen, Germany). Oprozomib and Epoxomicin were purchased from ApexBio (Houston, USA). The crosslinking agent BS3 was obtained from Thermo Scientific (Waltham, USA).

Protein purification. S30 HeLa cytoplasmic extract[12] was prepared by hypotonic lysis, centrifugated at 30,000*g* for 30 min at 4 °C, flash frozen in liquid nitrogen and stored at − 80 °C. The S30 extract was thawed in a water bath at 37 °C, supplemented with purification buffer to 1 × from a 10 × stock, sucrose powder to 20% (w/v), octyl glucose neopentyl glycol (from a 10% (w/v) stock solution in water) to 0.1% (w/v), iodacetamide to 10 mM, *N*-ethylmaleimide to 10 mM, benzamidine chloride to 10 mM and ATP to 7.5 mM. The extract was incubated at room temperature on a magnetic stirrer for 30 min, followed by an addition of Dithiothreitol (DTT) powder to 50 mM and a second incubation at room temperature for 30 min. The S100 extract was prepared by centrifugation at 100,000*g* for 2 h at 4 °C and the supernatant was filtered through three layers each of cheese cloth and miracloth.

The S100 extract was processed by two subsequent rounds of precipitation with PolyEthyleneGlycol400 (PEG400; number signifies the mean molecular weight of the PEG polymer). First, PEG400 was added to a concentration of 23% (v/v) to the S100 extract at 18 °C on a magnetic stirrer and incubated for 30 min. Second, the supernatant was precipitated by raising the concentration of PEG400 to 30% (v/v) as described before. The precipitate contains the human 26S/30S proteasomes and was resuspended with purification buffer supplemented with 7.5 mM ATP, 5 mM DTT and 0.01% (w/v) lauryl maltose neopentyl glycol (LMNG) in an orbital shaker at 18 °C. The resuspended material was incubated with an ATP regeneration system (10 mM sodium creatine phosphate, 5 µg ml^{-1} creatine kinase) at 30 °C for 30 min.

The sample was loaded on 20%/50% two-step sucrose cushions in purification buffer containing 7.5 mM ATP and 5 mM DTT. The cushions were centrifuged at 260,000*g* for 14 h at 4 °C, harvested in 500 µl fractions with Äkta Prime (GE Healthcare, Munich, Germany) and analysed by SDS–polyacrylamide gel electrophoresis (PAGE) to identify fractions containing 26S and 30S proteasomes. Fractions were pooled and precipitated by the addition of 40% (v/v) PEG400 for 30 min and after centrifugation (30,000*g*, 30 min), the precipitate was resuspended in purification buffer containing 5% sucrose, 7.5 mM ATP, 5 mM DTT and 0.01% (w/v) LMNG. The proteasomes were treated with Oprozomib at a concentration of 0.5 mM at 25 °C for 30 min. Proteasomes were loaded on linear 10–40% (w/v) sucrose gradients in purification buffer containing 7.5 mM ATP, 5 mM DTT, which were centrifuged at 220,000*g* for 16 h at 4 °C. In total, 400 µl fractions were analysed by SDS–PAGE, selected proteasome fractions were precipitated by the addition of 40% (v/v) PEG400 and resuspended in purification buffer containing 7.5% (w/v) sucrose, 7.5 mM ATP, 5 mM DTT and 0.01% (w/v) LMNG. As a final step, proteasomes were fractionated on linear 10–45% (w/v) sucrose gradients in purification buffer containing 7.5 mM ATP, 5 mM DTT, which were centrifuged at 260,000*g* for 16 h at 4 °C. Fractions containing 26/30S proteasomes were yet again identified by SDS–PAGE, precipitated by the addition of 40% (v/v) PEG400 and resuspended in 2 × purification buffer containing 15% (w/v) sucrose, 15 mM ATP, 10 mM DTT and 0.02% (w/v) LMNG yielding the final purified protein preparation at 30 mg ml^{-1}. Protein concentrations were determined by the Bradford assay (BioRad, Munich, Germany) using BSA as a standard. This purification procedure reproducibly yields 45 mg purified human 26/30S proteasomes, starting from 800 ml S100 HeLa cytoplasmic extract at a concentration of 10 mg ml^{-1}.

Purification buffer: 0.05 M Bis-Tris pH 6.5, 0.05 M KCl, 0.01 M $MgCl_2$,
0.01 M β-Glycerophosphate

Negative staining EM sample preparation and image analysis. Proteasomes were either supplemented with 2 mM Epoxomicin, 2 mM Oprozomib or DMSO as a control. After 30 min of incubation on ice, the respective samples were loaded on sucrose gradients (10–30% w/v sucrose). Gradient centrifugations were carried out in a TH660 Rotor (Thermo Scientific, Osterode, Germany) at a centrifugational force of 114,000*g* for 16 h at 4 °C. In total, 200 µl gradient fractions were collected.

EM grids of 26S proteasome fractions were prepared by floating a continuous carbon film in the solution for 1 min at 4 °C and staining with a saturated uranyl formate solution. Samples were imaged on a Philips CM200 microscope at a magnification of × 88,000 corresponding to a pixel size of 2.5 Å per pixel. In total, 500 micrographs per sample were collected in spot scanning mode using a TVIPS CCD camera.

Particles were selected and CTF correction was performed on the individual particle level. Resulting particles were subjected to several rounds of two-dimensional classification to remove images without particles and images containing contaminations such as ice crystals. The remaining particles were aligned against a 3D model of the rotated and independently to a model of the non-rotated state and assigned to the better fitting model according to the cross correlation. This was repeated in three iterations. After each iteration, new volumes were reconstructed from the assigned particles, low-pass filtered to the same resolution and normalized. Particles contributing to the individual classes were counted.

Cryo-EM sample preparation. BS3 (2.5 mM) was added to the purified proteasome holoenzyme (12 mg ml^{-1}) and incubated at 4 °C for 2 h. The crosslink reaction was terminated by the addition of 10 mM sodium aspartate (pH 6.5) and loaded on a GraFix gradient[13] (10–30% w/v sucrose, 0–0.1% glutaraldehyde). The gradient centrifugation was carried out in a TH660 Rotor (Thermo Scientific, Osterode, Germany) at a speed of centrifugational force of 114,000*g* for 16 h at 4 °C. A total of 200 µl gradient fractions were collected and immediately quenched by adding 20 mM of sodium aspartate (pH 6.5). The protein peak in the gradient fractions was assessed by a dotblot with Amidoblack staining. Peak fractions were analysed by negative staining EM. The fractions containing the single capped 26S proteasomes were buffer exchanged to the purification buffer without any sucrose using a Zeba spin column (Thermo Scientific, Osterode, Germany). The particles were adsorbed to a continuous carbon film for 1 min at 4 °C, attached to a Quantifoil (3,5/1) (Quantifoil, Jena, Germany) grid and freeze plunged in a Leica EM GP (Leica, Wetzlar, Germany) employing the blotting sensor at 75% humidity and 4 °C.

Cryo-EM data acquisition. The grids were imaged in a Titan Krios (FEI, Eindhoven, The Netherlands) (Supplementary Data 7) equipped with a Cs-Corrector (CEOS, Heidelberg, Germany) on a Falcon II detector. Images were taken at a nominal magnification of × 110,000, corresponding to a pixel size of 1.27 Å per pixel. The total dose (50 electrons per Å^2) was fractionated on 17 frames. The first frame revealing inhomogeneous illumination due to the camera shutter was discarded. In total, 18,707 micrographs were collected in total (Supplementary Fig. 8).

Image processing. Individual image frames were aligned and weighted according to electron dose using the software unblur[14] to reduce the effects of drift and charging. The CTF of the remaining micrographs was determined using Gctf[15]. Particles were selected in a template-free manner, using image statistical properties in combination with mass centring. Individual particle coordinates were additionally refined by alignment against twelve low resolution reference images representing different views of the proteasome (Supplementary Fig. 9).

Subsequently, we performed several image sorting steps to remove contaminations, blurred images and bad particles. In a first step, power spectra were calculated, for each particle and classified using a hierarchical clustering scheme. The resulting class averages were visually inspected for Thon ring appearance and particles belonging to strongly charged or blurred classes were discarded. Second, several rounds of multi-reference alignment and two-dimensional classification were performed. Particles belonging to classes that did not show clear molecule views were discarded.

After having applied these image sorting procedures, the best class averages were used to generate an initial 3D model using simple PRIME[16]. This 3D model was used as an initial reference in a 3D classification in RELION[17], which we used to classify the particles according to their two main conformational states. To ensure correct class assignment, all particles were aligned competitively against averaged maps obtained for the two main states. The flexible protein Rpn1 interferes with the alignments and therefore its density was masked out.

Particle images belonging to the non-rotated state of the proteasome were refined by RELION auto-refine. Subsequent hierarchical sorting discriminated further sub classes of various RP conformations. Specifically, a series of 3D classification steps without alignment using increasingly smaller masks was performed in RELION. In the first classification step, we used a mask for the whole proteasome holoenzyme excluding Rpn1, in the second iteration we used a mask for the whole RP (19S) subcomplex, in the third iteration a mask for the whole lid and finally a mask for Rpn2 only. The remaining particles (233,000) were refined to a final resolution of 3.8 Å and B-factor corrected in RELION. To further improve the map, particle polishing[18] was performed on the final particle stack in RELION.

A local resolution map was calculated in ResMap[19] by calculating local FSC values in a sphere with a diameter of 13 voxels moving over the entire 3D volume. In addition, the signal of CP was subtracted from the raw particles[20]. These subtracted particles were centred and again refined in RELION. Masks for the Rpt2/6, Rpn9/10 and Rpn5/6 regions were created in Chimera and focussed 3D classification without alignment was performed on the computationally generated 19S particles. Resulting 3D classes were refined.

Conformational landscape analysis of the RP subunit. A comprehensive explanation of the method is given in the Supplementary Note 1 of this manuscript. In brief, more than 346 3D class averages were obtained in RELION revealing conformational differences in the RP. The conformational variability was analysed quantitatively using PCA Eigenvolumes. By determining the linear factors of each volume towards those Eigenvolumes an energy landscape was calculated in MATLAB. The MATLAB scripts can be provided upon request.

Model building. The initial atomic coordinate model for the 20S particle was taken from a crystal structure of the Oprozomib-inhibited human 20S complex (PDB 5LEY)[10]. Models of each RP (19S) subunit were generated with Robetta[21,22] and docked as rigid bodies into the EM density map with UCSF Chimera[23]. The six nucleotides of the ATPase subunits were placed by fitting the crystal structure of PAN (PDB 3H4M)[24] into our density. Additional aid for regions, which had to be modelled at least partly *de novo*, was obtained using the secondary structure prediction server psipred[25].

An initial rigid body refinement was performed using real space refinement in Phenix[26] and subsequent manual modelling in coot[27]. Next, secondary structure restraints were generated using *phenix.ksdssp*. All secondary structure restraints were visually inspected and additional restraints were added manually. Several iterative rounds of real space refinement in Phenix and manual modelling in coot followed, where the last Phenix refinements included ADP refinement to calculate B-factors.

The present map quality does not allow to distinguish clearly between ATP and ADP in the ATPase and hence we modelled all nucleotides as ADP. In addition, to account for local resolution differences in the EM density map, we used calculated model B-factor distributions as a guideline to define the level of structural details interpreted in the final model (Supplementary Fig. 5). Accordingly, we analysed B factors in segments of five amino acids. Side chains were only modelled if the mean atomic B-factor per segment was smaller than 110 $Å^2$, segments with mean B-factors between 110 and 150 $Å^2$ were truncated to poly-alanine. Residues with mean B-factors higher than 150 $Å^2$ were not included in the final deposited PDB model.

Data availability. EM density maps have been deposited in the EMDB with accession number 4,146. Raw micrographs have been uploaded to EMPIAR database. Modelled atomic coordinates have been deposited in the Protein Data Bank with the accession number 5M32. The data that support the findings of this study are available from the corresponding authors upon request.

References

1. Saeki, Y. & Tanaka, K. Assembly and function of the proteasome. *Methods Mol. Biol.* **832,** 315–337 (2012).
2. Kish-Trier, E. & Hill, C. P. Structural biology of the proteasome. *Annu. Rev. Biophys.* **42,** 29–49 (2013).
3. Finley, D. Recognition and processing of ubiquitin-protein conjugates by the proteasome. *Annu. Rev. Biochem.* **78,** 477–513 (2009).
4. Matyskiela, M. E., Lander, G. C. & Martin, A. Conformational switching of the 26S proteasome enables substrate degradation. *Nat. Struct. Mol. Biol.* **20,** 781–788 (2013).
5. Kleijnen, M. F. *et al.* Stability of the proteasome can be regulated allosterically through engagement of its proteolytic active sites. *Nat. Struct. Mol. Biol.* **14,** 1180–1188 (2007).
6. Sledź, P. *et al.* Structure of the 26S proteasome with ATP-γS bound provides insights into the mechanism of nucleotide-dependent substrate translocation. *Proc. Natl. Acad. Sci. USA* **110,** 7264–7269 (2013).
7. Chen, S. *et al.* Structural basis for dynamic regulation of the human 26S proteasome. *Proc. Natl. Acad. Sci. USA* **113,** 12991–12996 (2016).
8. Schweitzer, A. *et al.* Structure of the human 26S proteasome at a resolution of 3.9 Å. *Proc. Natl. Acad. Sci. USA* **113,** 7816–7821 (2016).
9. Huang, X., Luan, B., Wu, J. & Shi, Y. An atomic structure of the human 26S proteasome. *Nat. Struct. Mol. Biol.* **18,** 1–10 (2016).
10. Schrader, J. *et al.* The inhibition mechanism of human 20S proteasomes enables next-generation inhibitor design. *Science* **353,** 594–598 (2016).
11. Yen, H. C. S., Espiritu, C. & Chang, E. C. Rpn5 is a conserved proteasome subunit and required for proper proteasome localization and assembly. *J. Biol. Chem.* **278,** 30669–30676 (2003).
12. Dignamr, J. D., Lebovitz, R. M. & Roeder, R. G. Accurate transcription initiation by RNA polymerasae II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res.* **1,** 1475–1489 (1983).
13. Kastner, B. *et al.* GraFix: sample preparation for single- particle electron cryomicroscopy. *Nat. Methods* **5,** 53–55 (2008).
14. Grant, T. & Grigorieff, N. Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6. *Elife* **4,** 1–19 (2015).
15. Zhang, K. Gctf: real-time CTF determination and correction. *J. Struct. Biol.* **193,** 1–12 (2016).
16. Elmlund, H., Elmlund, D. & Bengio, S. PRIME: probabilistic initial 3D model generation for single-particle cryo-electron microscopy. *Structure* **21,** 1299–1306 (2013).
17. Scheres, S. H. W. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180,** 519–530 (2012).
18. Scheres, S. H. Beam-induced motion correction for sub-megadalton cryo-EM particles. *Elife* **3,** 1–8 (2014).
19. Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat. Methods* **11,** 63–65 (2014).
20. Bai, X., Rajendra, E., Yang, G., Shi, Y. & Scheres, S. Sampling the conformational space of the catalytic subunit of human γ-secretase. *Elife* **4,** e11182 (2015).
21. Song, Y. *et al.* High-resolution comparative modeling with RosettaCM. *Structure* **21,** 1735–1742 (2013).
22. Kim, D. E., Chivian, D. & Baker, D. Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res.* **32,** 526–531 (2004).
23. Goddard, T. D., Huang, C. C. & Ferrin, T. E. Visualizing density maps with UCSF chimera. *J. Struct. Biol.* **157,** 281–287 (2007).
24. Zhang, F. *et al.* Structural insights into the regulatory particle of the proteasome from *Methanocaldococcus jannaschii. Mol. Cell* **34,** 473–484 (2009).
25. Buchan, D. W. A., Minneci, F., Nugent, T. C. O., Bryson, K. & Jones, D. T. Scalable web services for the PSIPRED protein analysis workbench. *Nucleic Acids Res.* **41,** 349–357 (2013).
26. Adams, P. D. *et al.* PHENIX: a comprehensive python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66,** 213–221 (2010).
27. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60,** 2126–2132 (2004).

Acknowledgements

We thank Thomas Conrad for HeLa cell growth and Hossein Kohansal for HeLa cell extract preparation. Monika Raabe and Henning Urlaub for mass spectrometry service. Mario Lüttich, Boris Busche, Jan-Martin Kirves, Georg Bunzel and Lukas Schulte for development of image processing software. Wen-ti Liu for graphical design support and Sabrina Fiedler for help with MATLAB and Brenda A. Schulman and Lars Bock for fruitful discussions. This work was funded by grants of the Deutsche Forschungsgemeinschaft (DFG; CH1098-1/1 to A.C., SFB860-TP A5 to H.S.). H.S. and A.C. received support by an R&D Instruct grant as part of the European Strategy Forum on Research Infrastructures (ESFRI), which is supported by national member subscriptions.

Author contributions

J.S. and A.C. developed and performed proteasome purification. D.H. performed electron microscopy, image processing. A.C. and F.H. screened several proteasome inhibitors. D.H. and F.L. performed the energy landscape analysis and model building. A.C. and H.S. designed and supervised research. All authors contributed to the preparation of the manuscript.

Additional information

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Haselbach, D. *et al.* Long-range allosteric regulation of the human 26S proteasome by 20S proteasome-targeting cancer drugs. *Nat. Commun.* **8**, 15578 doi: 10.1038/ncomms15578 (2017).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/

© The Author(s) 2017

3 CowScape: a novel approach to quantitatively describe the conformational landscape of biological macromolecules from cryo-EM data

Cited as: Lambrecht (submitted)

Own contribution:

- Concept and implementation of the algorithm
- Analysis of all ensembles shown in the paper
- All figures
- Manuscript in parts
- Supplemental information (see supplement)

CowScape: a novel approach to quantitatively describe the conformational landscape of biological macromolecules from cryo-EM data

Felix Lambrecht[1], Michael Habeck[2], Mario Lüttich[1], David Haselbach[1,3] and Holger Stark[1]**
[1] *Max-Planck-Institute for Biophysical Chemistry, Department of Structural Dynamics, Goettingen, Germany*
[2] *Max Planck Institute for Biophysical Chemistry, Research Group Inverse Statistical Problems in Biophysics, Goettingen, Germany*
[3] *current affiliation: Institute for Molecular Pathology, Vienna, Austria*
* e-mail: hstark1@gwdg.de; david.haselbach@imp.ac.at

Abstract

Typical cryo-EM data processing often discards images that belong to structures other than the main state under investigation; however, this likewise discards enormous amounts of information that could be valuable for determining three-dimensional ensembles of structures. Here we describe an approach ("CowScape") that analyses all particle data in a cryo-EM dataset to obtain a quantitative description of structural and conformational variability, which can represent the entire biochemically relevant conformational space of macromolecular complexes. By combining extensive image classification and principal component analysis (PCA) of the classified 3D volumes, CowScape can be used as a quantitative tool to analyze this variability. PCA orders all calculated 3D structures along the major modes of structural variability, and the known number of particle images in a given state can be used to calculate an energy landscape based on Boltzmann conversion. Thus, by revealing allosteric interactions in macromolecular complexes, CowScape allows dynamic changes in macromolecular complexes during function and regulation to be distinguished and interpreted.

Most biological processes in the cell are not driven by individual proteins but rather by large macromolecular assemblies comprising numerous protein and/or nucleic acid components[1]. These complexes can be considered as molecular machines, as they are usually very dynamic and undergo significant conformational rearrangements during their reaction cycles. These molecular machines are regulated in their natural biological context, for instance by binding external factors that influence the conformational freedom of a macromolecular complex in a functionally relevant manner. Understanding the function of such macromolecular machines requires taking the entire conformational landscape that the macromolecular complex adopts during regulation and its reaction cycle into account[2-4]. This implies a drastic step beyond the current state-of-art in cry-EM, in which the focus is on high-resolution structure determination of a single static state. The optimal situation would be to determine a high-resolution molecular "movie" of the entire conformational space that can be covered. This would allow for instance a better understanding of allosteric regulation effects as a consequence of a ligand (such as a drug) binding event that affects the activity of the complex.

Historically, X-ray crystallography has dominated as the method for determining high-resolution structure of macromolecular complexes, yet it is mainly limited to static structure determination, which can be roughly understood as molecules with the same conformation packed into a crystalline lattice. In cryo-

EM, in contrast, molecules are imaged individually in a frozen hydrated state with no bias to adopt any preferred conformation as a result of crystal packing. The range of structural heterogeneity (conformational dynamics) in cryo-EM is therefore significantly higher than in X-ray crystallography, which can be very problematic for high-resolution structure determination. The current Electron Microscopy Data Bank (EMDB) statistics exemplify how detrimental structural heterogeneity can be, as they reveal a mean value of deposited cryo-EM structures only in a range of ~6 Ångstrom resolution, despite the ground-breaking technical developments in cryo-EM that now allow up to sub-3Å resolution[5]. Indeed, any mixture of conformations (due to either low sample quality or the presence of multiple conformations) will always be resolution-limiting in cryo-EM.

Some aspects of this heterogeneity problem can be tackled using advanced computational image processing tools, which can sort cryo-EM images according to differences in composition and conformation. A large dataset of cryo-EM images can thus be purified *a posteriori* in silico to compensate for a higher level of conformational and structural heterogeneity. Extremely helpful computational tools have been recently developed for this computational sorting, including two-dimensional (2D) PCA [6], maximum likelihood-based classification in 2D/3D [7, 8], and manifold embedding of raw images [9]. However, while these computational tools effectively distinguish between distinct conformational states of the molecules, their ability to separate continuous conformational differences is usually limited. Currently, the most widely used classifications in the cryo-EM field are maximum-likelihood–based ones[10] that do not require previous knowledge, as these are easy to use within powerful software packages, such as Relion[7]. The major focus in most cryo-EM projects to date is to computationally purify particle images that belong to the main conformational state present in the cryo-EM dataset in an automated manner, and to get rid of the particle images that do not match based on several rounds of classifications. Subsequently, this “purified” particle subpopulation can be refined to high resolution. In some cases, this approach must discard more than 90% of all images in a given dataset to reach a level of structural homogeneity that allows a high-resolution structure to be calculated. Thus, in most applications to date, the majority of images does not belong to the best subpopulation. By large, the discarded data is not analysed, so that it is often unclear why the majority of molecular images do not contribute to the high-resolution structure. Nonetheless, it is probable that this discarded cryo-EM data contain interesting and relevant information. In an ideal situation, all data would be used to computationally purify all well-defined states that exist in a given dataset, rather than only the most evident one, and each image in the recorded dataset would be classified to its corresponding three-dimensional (3D) structure. However, such a procedure would require massive image statistics, as the conformational freedom of the entire molecule should be computationally sorted to understand the overall dynamics of the macromolecular complex. An alternative approach, which is implemented in Relion, uses local masks on the structure and tries to solve the locally different alignment parameters on multiple flexible regions of a given complex (multi-body refinement); this has the advantage of a more moderate requirement for particle image statistics [11].

The first attempt to determine the entire conformational landscape from a large image dataset was the exhaustive computational analysis of an *E. coli* 70S ribosome trapped in intermediate states of tRNA retro-translocation [2]. In this study, 2 million particle images were used to separate 50 conformational states by a hierarchical computational purification procedure as well as by visual inspection of the calculated 3D structures based on the tRNA positions on the ribosome. Following the tRNA motions by visual comparison of all structures allowed the order of the structures to be determined along a conformational trajectory. After taking the particle statistics into account, the first energy landscape based on cryo-EM data was determined[2]. The entire procedure required extensive manual image processing, as well as prior knowledge about tRNA motions, to find the order of all states which was later confirmed by molecular dynamic simulations[12]. It is therefore impossible to apply this approach to samples for which little or no prior knowledge is available.

We have now developed an approach, "CowScape" (Fig. 1a), to obtain a quantitative description of conformational variability of macromolecular complexes. CowScape combines extensive maximum-likelihood 3D classification and PCA (3D-PCA) to determine all possible conformational states without the need for any prior knowledge about the complexes. CowScape analyses the similarities and differences of the states in a meaningful and automated manner and then orders the states by 3D-PCA of all calculated and aligned 3D volumes. The number of states can range from in the tens to several hundred structures. The main modes of conformational variations are directly obtained by the major eigenvectors of the 3D-PCA analysis (Fig. 1b). Taking two major modes allows the conformational landscape to be plotted, which provides a likely trajectory of conformational changes (Fig. 1c). Furthermore, as the number of particle images used to calculate the structures of each conformational state is known, CowScape can convert the conformational landscape into an energy landscape by applying Boltzmann's law. We have successfully used CowScape analysis for two large macromolecular complexes, the human B^{act} spliceosome[3] and the human 26S proteasome[4]. For both complexes, CowScape allowed us to determine structures of unknown conformations and to obtain new structural insight into functionally relevant regulatory mechanisms. We provide here the details of how CowScape works and a further validation of its general applicability to cryo-EM data by using CowScape for other macromolecular complexes.

RESULTS

The CowScape algorithm, available in the COW image processing suite (www.cow-em.de), provides the 3D classification and refinement tools as well as the tools required for generating a conformational landscape and an energy landscape, for plotting data into a 2D heat map, and subsequently for visualizing these data in a 3D landscape. The first step in the analysis is an extensive 3D classification of all data identified as true particle images within a cryo-EM dataset (see overview, Fig. 1). This 3D classification can be performed in various image processing packages[7, 8, 13, 14], including the COW package. After classification, each subpopulation is refined to the highest possible resolution. Notably, the 3D classification step provides not only a qualitative estimation about the conformational ensemble but also quantitative information, by taking the number of particles into account that belong to each class. The idea behind this is straightforward: any thermodynamically stable conformation of a macromolecular complex will be observed at a higher particle frequency than a thermodynamically less-favoured state. The particle number per conformational subpopulation thus directly translates into free energy differences and, as such, can be used to determine an energy landscape via the Boltzmann factor (for details, see Supplementary Methods).

Classification provides a large number of 3D structures in different conformations and the corresponding particle numbers that were used to calculate the structures. However, at this first stage, the set of 3D structures lacks any order with respect to each other; instead, it simply reflects the conformation ensemble present in the dataset. To obtain a quantitative description of the major structural similarities, all 3D structures are then subjected to 3D-PCA. The most important output of the 3D-PCA are the eigenvectors (which themselves are 3D volumes) that describe the 3D structural variance in a hierarchical manner; that is, the first few eigenvectors can be used to describe most of the heterogeneity in the dataset. CowScape can use this information to then visualize the continuous motions along those major eigenvectors in a movie-like manner, which can illustrate the major modes of motions present in the dataset as well as reveal any potential coupling of movable parts in the macromolecular complex. The major eigenvectors also serve as coordinates that span the energy landscape—that is, a heatmap with all classified 3D structures positioned within a coordinate system that describes the major conformational variability. The x-y position in the energy landscape plot thus depicts the respective conformation of the complex, and the colour shows the relative Gibbs free energy difference (Fig. 2). The energy profile provides the user with valuable information about the number of conformational states and an estimate of the energy barriers between them.

To show the general applicability of CowScape, we have previously used it in our studies of the 26S proteasome and the human spliceosome [3, 4] and have now used it for publically available cryo-EM datasets in the Electron Microscopy Public Image Archive (EMPIAR) database (Supplementary Methods). These datasets also serve as different examples of what can be learned from the CowScape analysis. All datasets were downloaded and processed in a similar manner (see Methods). The 3D volumes were then subjected to 3D-PCA, and the corresponding energy landscapes were calculated (Fig. 2) by Kernel Density Estimation with a Gaussian kernel function (Supplementary Methods). Some general features of macromolecular complexes can be seen directly in the energy landscape plots (Fig. 3). For some biochemically well-behaved complexes, such as the non-selective cation channel TRPM4 (Fig. 4), the energy landscape reveals well-defined thermodynamic minima (blue), suggesting that a large part of the data contributes to one major structure, and that there is only a relatively small number of defined states in the dataset. In contrast, other complexes, such as the spliceosome, reveal a continuously populated energy landscape, making it more difficult to find a sufficient number of molecular images to determine structures of all conformational states at high resolution. For these complexes, the energy barriers between the various states are much smaller, and the entire energy landscape can usually be more easily modulated by ligands that bind to the complex.

It is also noteworthy that, depending on the kind of macromolecular complex, the energy landscape not only reflects the potential conformational changes of the complex but can also contain subpopulations of complexes that differ in their composition. CowScape always analyses the entire variability in the dataset, which comprises both conformational and compositional changes. If the biochemical sample is known to be very pure, the population landscapes can be interpreted as differences in conformational sampling of the same complex, and thus assumed to correspond to an energy landscape. In either case, the landscape provides the user with the unique possibility of discovering previously unknown conformations, for which one can subsequently try to select the raw data and refine them to high-resolution. For instance, the landscape determined for the mitotic checkpoint complex (MCC) bound to the anaphase promoting complex (APC) [15] revealed a high level of biochemical heterogeneity within the complex; this therefore cannot be used as a depiction of an energy landscape (Fig. 2). The two major eigenvectors of this dataset describe the stable integration of MCC into the APC complex, and the presence or absence of the protein APC2. Such a landscape can thus be used as a tool to quantitatively monitor any improvement in complex preparation and purification, with the goal of maximizing the number of molecules that have all desired components stably integrated into the complex. Energy landscapes can only be determined at a later stage after successful biochemical optimization.

To focus on conformational variability, we analysed examples that can be considered to be biochemically optimized, namely, the 26S proteasome and the TRPM4 channel[16] (Figs. 3 and 4). For the 26S proteasome (Fig. 3), our analysis verified the macromolecular complex conformations that had been previously described[17, 18]. However, the energy landscapes provided a more quantitative view about the data in general that goes beyond previous findings. Specifically, we applied CowScape to two highly dynamic proteasome samples, focusing on the subunit RPN1 within the 26S proteasome holocomplex (Fig. 2a) and the regulatory 19S subcomplex bound to the chaperone p28 (Fig. 2b). A focused classification and subsequent CowScape analysis revealed an almost continuous pendulum-like motion for RPN1, which would explain why this protein has so far been elusive in high-resolution structure studies of the 26S proteasome. The pendulum-like two-state distribution has its pivot point within the N-terminal coiled coil of RPT1 and RPT2 and might represent a regulatory mechanism by which the Rpt1/2 interface can be blocked (Supplemental Fig. 1). In contrast to other observed motions, RPN1 pendulum mobility does not seem to be directly coupled to the larger modes of motions that can be observed for the entire 26S proteasome. For the 19S proteasome, we observed opening of the Rpt2/Rpt6 interface in the ATPase ring structure with a simultaneous closing of the Rpt3/4 interface. These changes correlate with a second interface closure at the RPN3 / RPN7 interface in the non-ATPase part of the complex (Supplementary Fig. 2). Thus, using CowScape, we can not only automatically recover the order of minimum changes through the conformational snapshots but also directly distinguish between non-coupled and coupled motions within a macromolecular complex, which is likely to provide valuable information about the function of these large assemblies.

A further striking case is shown for the TRPM4 channel with and without bound calcium ions (EMPIAR-10127 and EMPIAR-10126, respectively; Fig. 4)[16]. We detected an interesting conformational flexibility at the cytoplasmic side of the channel, which to our knowledge has not yet been described in the literature (Fig. 4b, Supplementary Fig. 3). By comparing the landscape with and without bound calcium, we can speculate that calcium binding increases the number of contact sites between the soluble channel parts with the central coiled coil. These soluble parts might be involved in second messenger signalling, and increased flexibility might hence precede channel opening. The location of calcium binding is significantly distant from the detected conformational changes, which implies allosteric signalling. The functional relevance of these changes in conformation sampling after binding small ligands has yet to be determined.

We previously showed for the 26S proteasome that the major conformation of the entire complex was almost identical irrespective of the presence or absence of the cancer drug oprozomib[4]. However, after drug binding to the 26S proteasome, the overall conformational space differed considerably[4]. A similar situation was observed for the TRPM4 channel. Calcium binding to the TRPM4 channel is analogous to oprozomib binding to the 26S proteasome, with direct consequences on the ability of the large macromolecular complex to adopt the set of conformations that are relevant for its function (Supplemental figure 4). While the exact details of how this happens are still not known, it is possible to study such effects by calculating energy landscapes based on the analyses of all particle images present in a given dataset. This also implies that CowScape is a powerful tool for the structural interpretation of ligand binding to a macromolecular complex and for allowing functionally relevant conformational changes to be observed that would otherwise remain invisible (e.g., if they do not affect the stability of the major conformation).

The above examples illustrate that the quantitative analysis of conformational sampling is even more informative when two or more energy landscapes of the same complex but under different biochemical conditions can be compared. Such comparative studies are ideally suited to elucidating how regulatory factors influence macromolecular machines in a more quantitative manner. However, as analysis requires a minimal particle image frequency for a conformational state to be discovered, short-lived intermediates of a macromolecular complexes (such as those in transition states in catalysis) will not show up in this analysis, as they are not sufficiently populated. How sensitive CowScape can become at visualizing the conformational variability in macromolecular complexes in a quantitative manner in the future will depend on the image statistics used for the CowScape analysis and on the power of the applied 3D classification algorithms and the computational processing strategy. While this may be a limitation at the moment, one can expect that detectors and computers become significantly faster in the near future, and that novel 3D classification algorithms can be developed that will be able to determine ever smaller conformational differences. The CowScape algorithm itself is very fast and will not be a limiting factor even if analysis were to use several tens of thousands of conformations.

DISCUSSION

CowScape currently offers a comprehensive way to analyse and display "motions" in a cryo-EM dataset and to estimate quantitative free energy differences, which can be used to deduce mechanisms underlying motions or allosteric signal propagation[19]. This is the most appealing difference to other approaches for modelling molecular motions from cryo-EM data, which are highly selective and/or stop at the purely descriptive level. Furthermore, in contrast to other methods, CowScape does not need any *a priori* information, which makes it possible to use the approach at a very early stage of a project, thereby allowing the macromolecular complex conformations to be identified for which high-resolution structure refinement are most possible. The kind of insight generated by CowScape goes beyond the purely structural point of view and enables information to be gained that is comparable with for instance spectroscopic data, for which the distribution of states is usually accessed by only a one-dimensional output. This powerful feature of the CowScape analysis is strongly coupled to the idea of not discarding any true particle images from a given dataset but rather using all the available data to obtain a quantitative understanding of the ensemble. Doing so may make more detailed quantitative experiments possible in the future that describe how variables (e.g., temperature, pH values, salt conditions, or specific drugs) interfere with a macromolecular complex. This information will be difficult to obtain by any other method yet is critical for our understanding of how large macromolecular complexes act as "molecular machines".

Figures

Figure 1

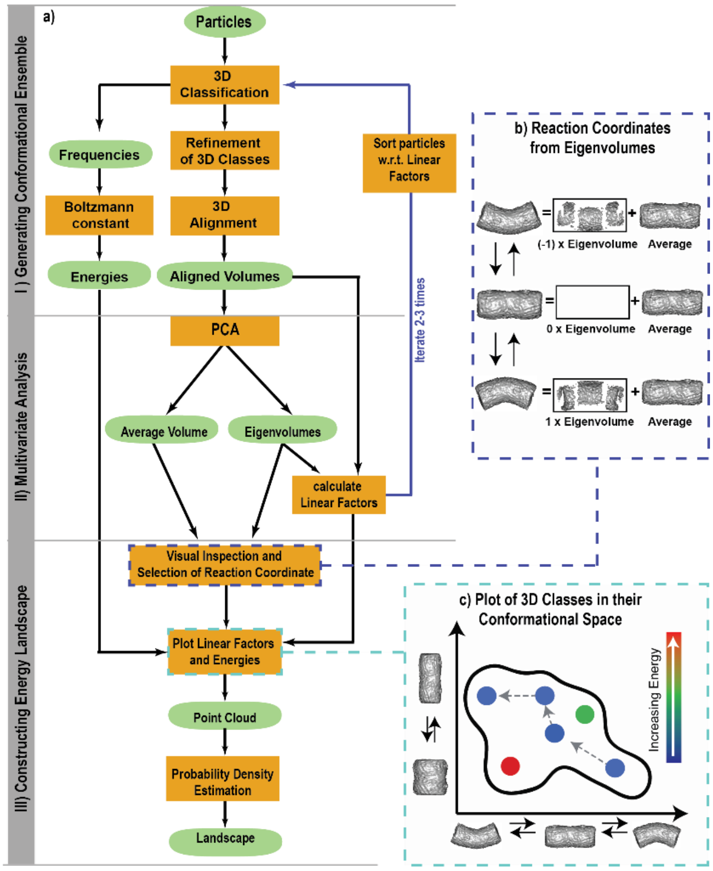

a)
I) Generating Conformational Ensemble
II) Multivariate Analysis
III) Constructing Energy Landscape
Particles
3D Classification
Frequencies
Refinement of 3D Classes
Sort particles w.r.t. Linear Factors
Boltzmann constant
3D Alignment
Energies
Aligned Volumes
Iterate 2-3 times
PCA
Average Volume
Eigenvolumes
calculate Linear Factors
Visual Inspection and Selection of Reaction Coordinate
Plot Linear Factors and Energies
Point Cloud
Probability Density Estimation
Landscape
b) Reaction Coordinates from Eigenvolumes
=
+
(-1) x Eigenvolume
Average
0 x Eigenvolume
Average
1 x Eigenvolume
Average
c) Plot of 3D Classes in their Conformational Space
Increasing Energy

Fig. 1: **The CowScape algorithm landscape** (for more details, see Supplementary Methods)

- **a)** Overview of the algorithm. CowScape is roughly divided into three steps: I) generate the conformational ensemble by applying classification algorithms; II) carry out PCA; and III) make a final selection of reaction coordinates and construct the energy landscape.
- **b)** Interpretation of eigenvolumes. Each volume of a trajectory (left column) can be described as a linear combination of the eigenvolume multiplied with a linear factor and the average volume. Hence, each eigenvolume describes a conformational motion.
- **c)** Using the linear factors, volumes can be plotted into a 2D conformational space with the respective energies (indicated by a colour gradient). Energetic minima (in blue) guide the most probable overall path through the landscape (grey arrows).

Figure 2

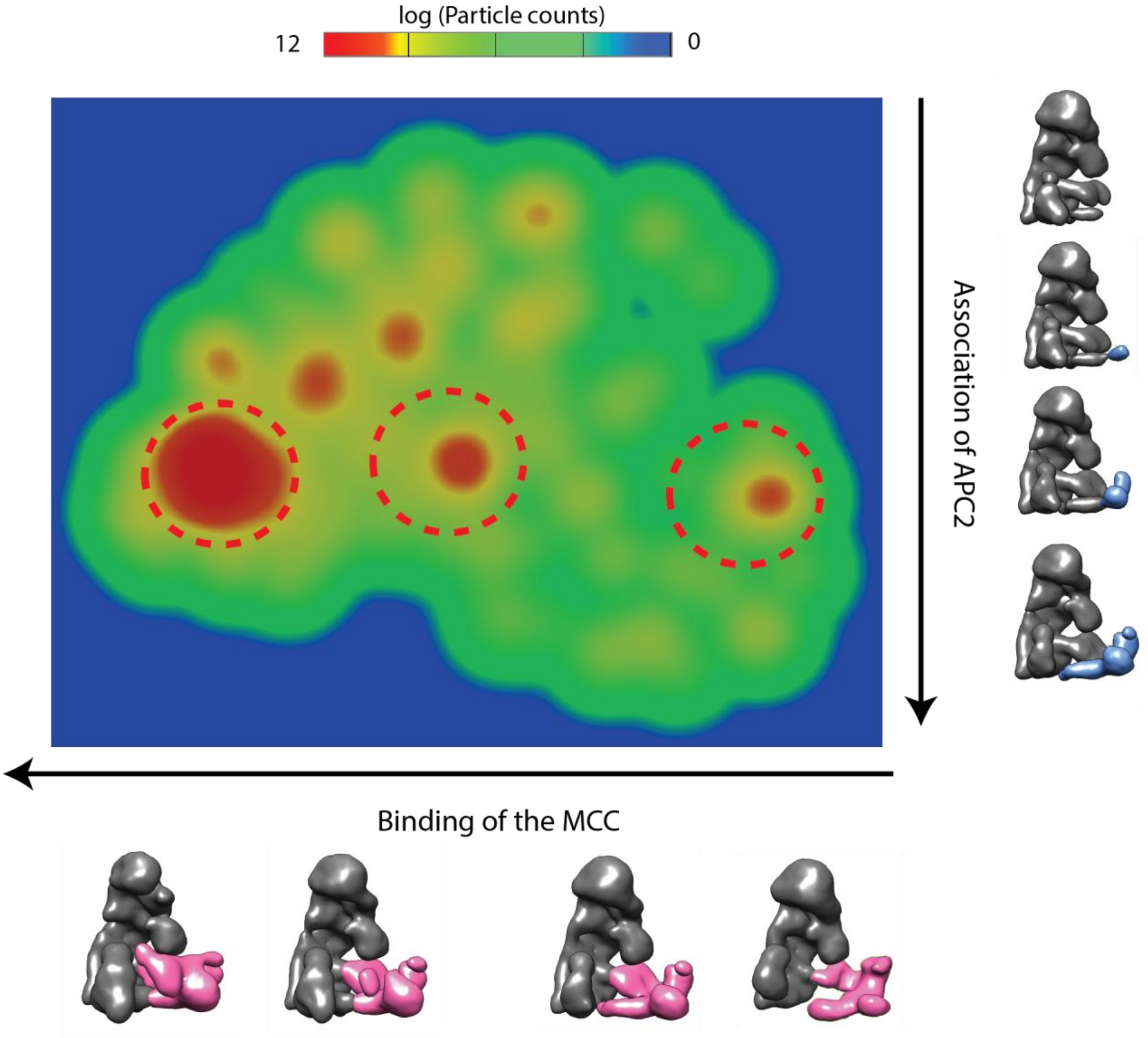


Fig. 2: Compositional landscape of APC

This mainly compositional landscape visualizes the different assembly states of APC in a sample with a large degree of biochemical heterogeneity (i.e., it is not yet optimized in an early state of the project). The three major maxima correspond (red circles) to the different binding states of MCC (along the x-axis; shown in magenta). Any structural changes along this x-axis correspond to compositional heterogeneity of MCC binding as well as to some motion of the MCC towards the active center of the APC. The y-axis reveals the level of substoichiometric binding of the protein APC2 (shown in blue). The plot can be used as a quantitative tool to monitor and improve the biochemistry of complex purification and preparation. As such, it cannot be interpreted as an energy landscape, yet it still provides valuable information about the quality of the complex to be studied.

Figure 3

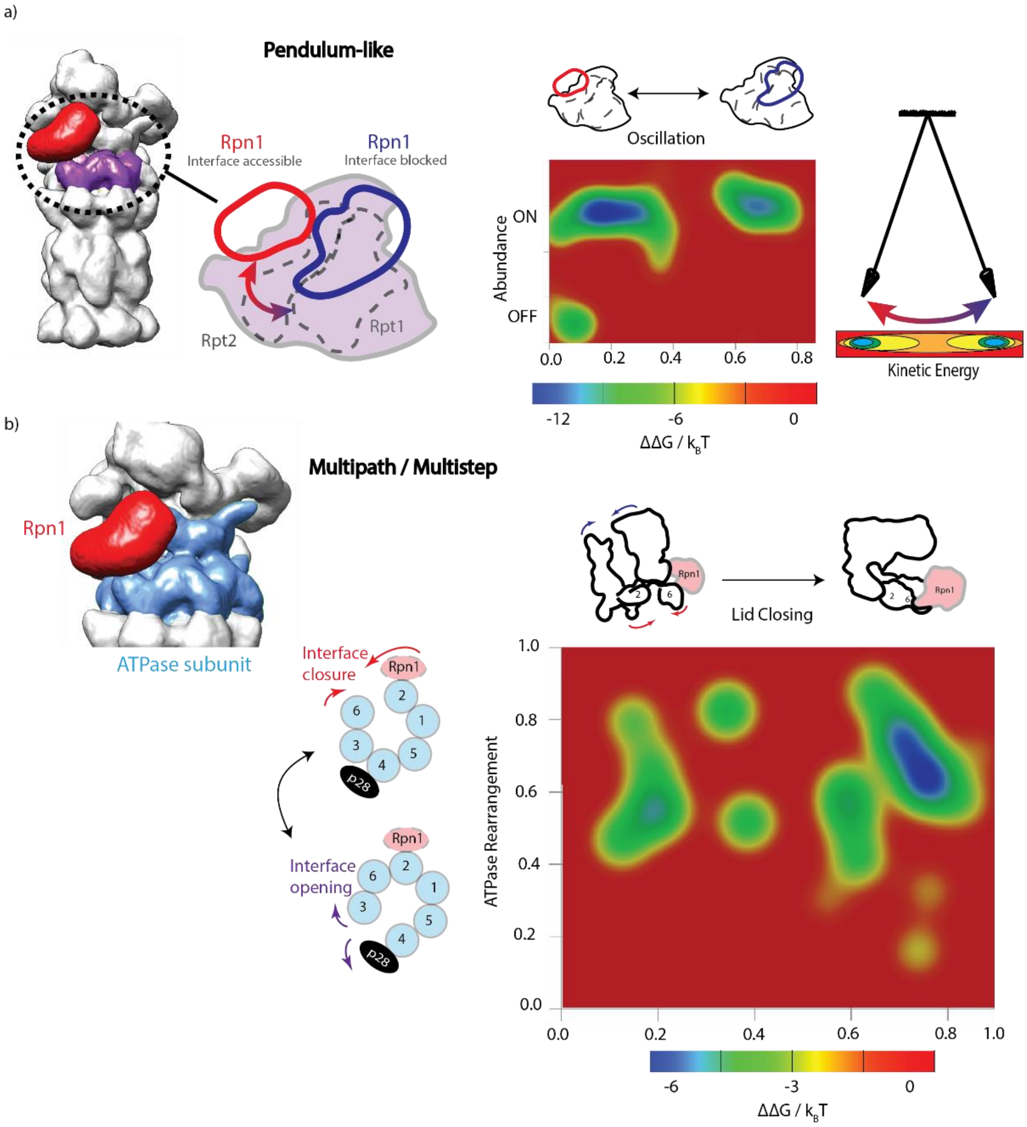


Fig. 3: Dynamic aspects within the 26S Proteasome

- **a)** RPN1 is a very dynamic protein that can adopt various conformations on the 26S proteasome. CowScape analysis reveals the two major conformations of RPN1 with respect to the Rpt2/Rpt1 interface (purple) and a requirement for a pendulum-like motion in order to proceed from one major state to the other. The major variability along the x-axis of the energy landscape can be described by this pendulum-like movement (see also Supplementary Fig. 1) whereas the y-axis in the plot describes RPN1 abundance. The RPN1-ON state reveals an energy profile similar to a two-state energy profile of a pendulum, as schematically depicted.
- **b)** Detailed analysis of the dynamic behaviour of the ATPase subunit (blue) as a major component of the 19S regulatory subunit of the 26S proteasome. Whereas the major mode describes an opening and closing of the 19S subunit along the x-axis of the plot, the y-axis reveals a novel conformational change in the ATPase subunit that can be summarized as an opening and closing of the ATPase ring system via different subunit interfaces.

Figure 4

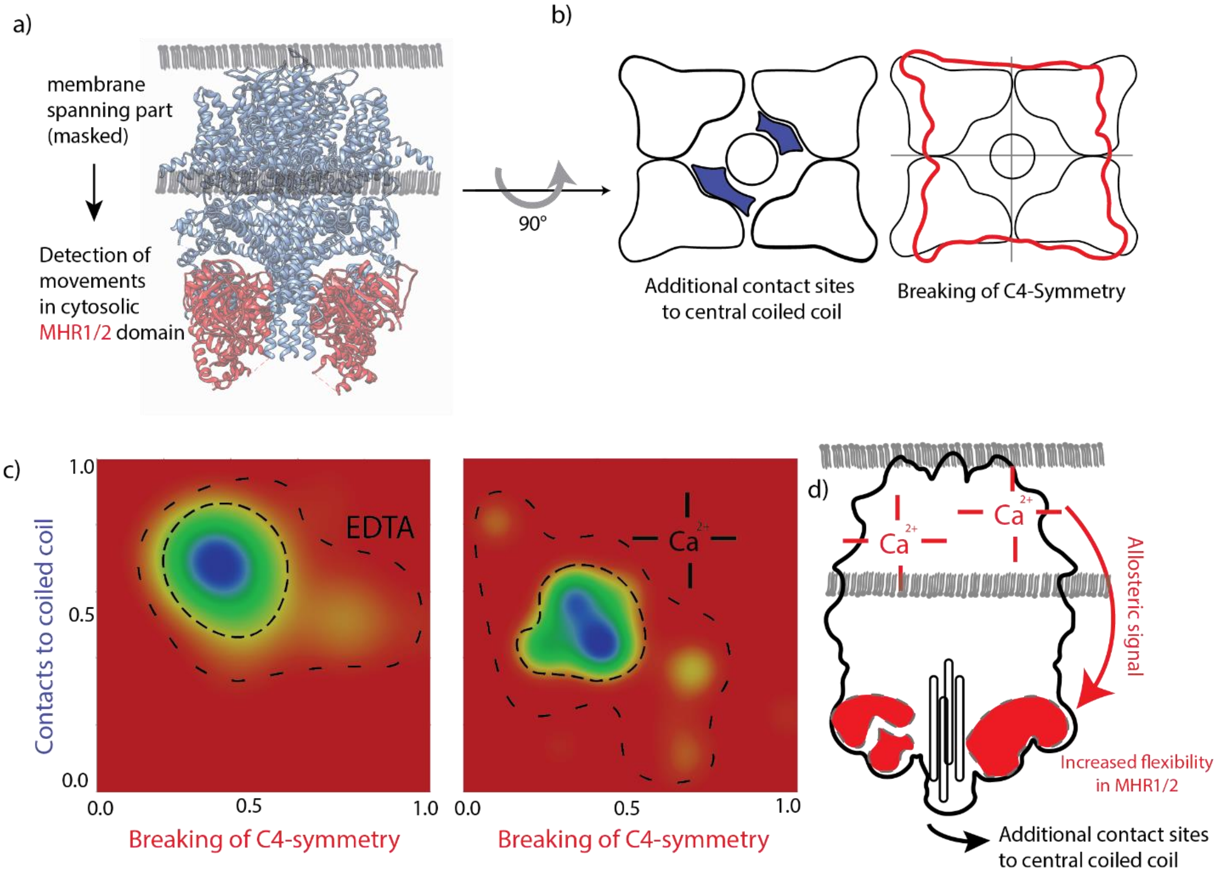
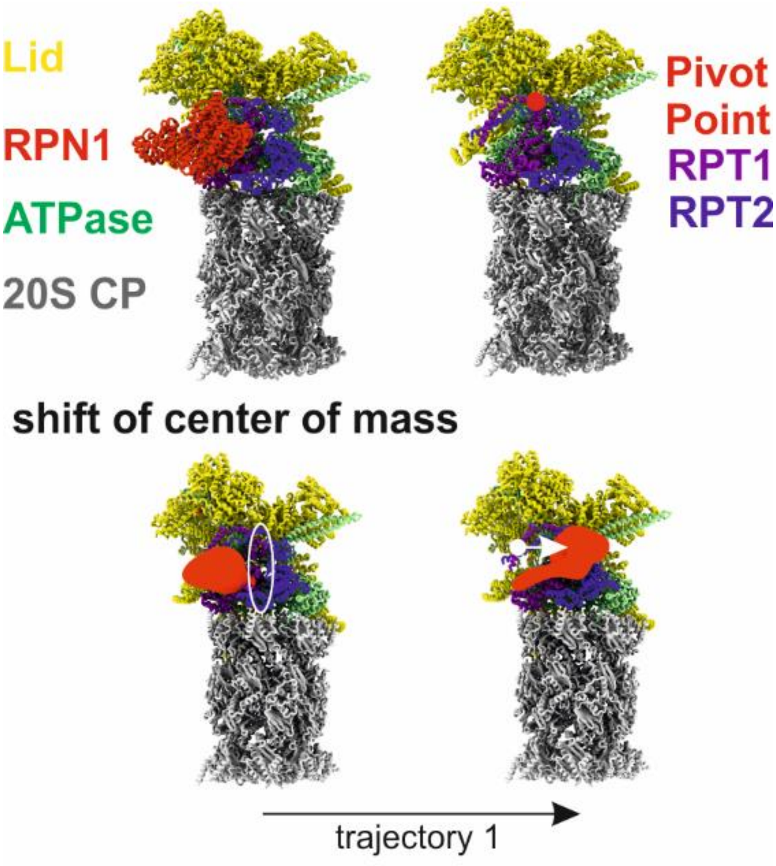

a)
membrane
spanning part
(masked)
Detection of
movements
in cytosolic
MHR1/2 domain
90°
b)
Additional contact sites
to central coiled coil
Breaking of C4-Symmetry
c)
1.0
0.5
0.0
Contacts to coiled coil
EDTA
0.0
0.5
1.0
Breaking of C4-symmetry
Ca2+
0.0
0.5
1.0
Breaking of C4-symmetry
d)
Ca2+
Ca2+
Allosteric signal
Increased flexibility
in MHR1/2
Additional contact sites
to central coiled coil

Figure 4: Allosteric Ca^{2+} signaling in the TRPM4 channel

- **a)** CowScape analysis of the TRPM4 (transient receptor potential melastatin member 4) channel not only revealed the overall conformations [16] but also conformational variations that can be attributed to an overall change from a C4 to a (more-or-less) C2 symmetry behaviour (x-axis). Changes along the y-axis can be described by additional contact sites of the MHR1/2 domains being established to the central coiled-coil domain.
- b) Binding of calcium leads to an overall change in symmetry and to a conformational change that creates new internal contact sites.
- c) Even though the binding sites for calcium is on the membrane site, an increased flexibility of the MHR1/2 domains located at the opposite site can be observed, strongly suggesting that an allosteric signaling (d) pathway affects the entire TRPM4 channel and finally leads to channel opening.

Supplementary Figures

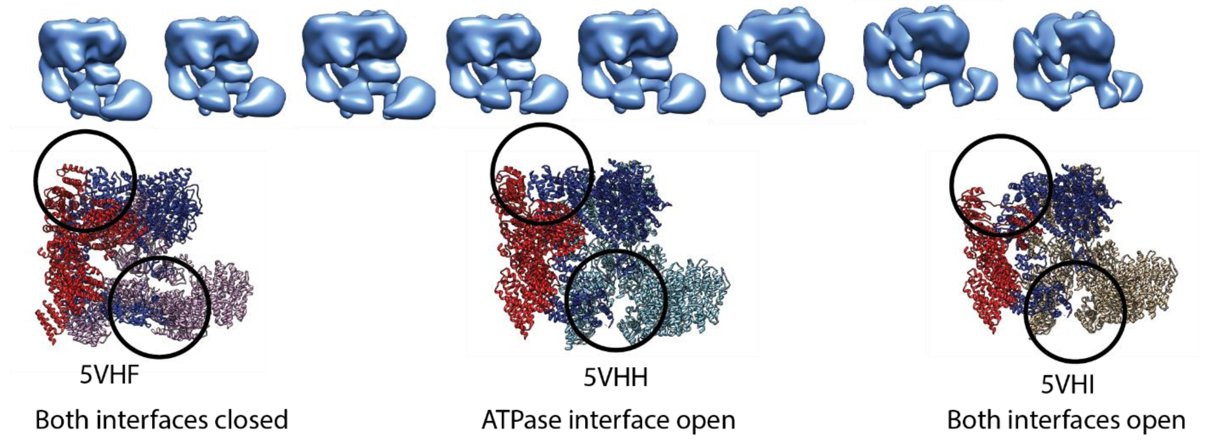
Supplementary Fig. 1


Lid
RPN1
ATPase
20S CP
Pivot
Point
RPT1
RPT2
shift of center of mass
trajectory 1

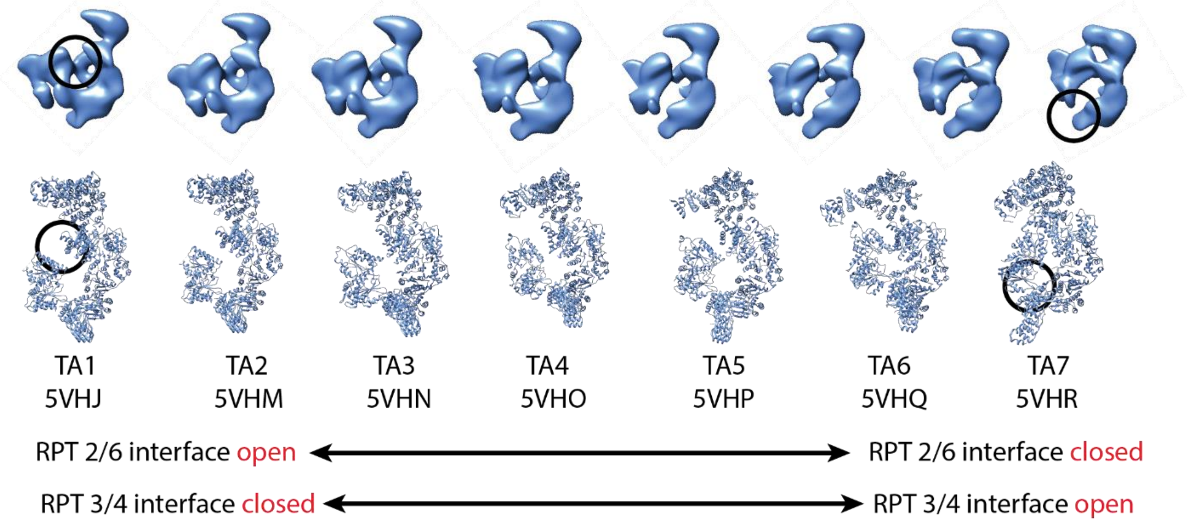
Supplementary Fig. 1: RPN1 movement.
The upper panel shows the same view of the human 26S Proteasome (PDB 5M32) with and without the protein RPN1. RPN1 is labelled in red and the proteins RPT1 and RPT2 are coloured in purple and blue, respectively. CowScape revealed a combined shift and pendulum-like movement of RPN1 described in the main text which reflects the major variability along the first eigenvector. RPN1's center of mass completely covers the RPT1/2 interface (white circle). RPN1 shifts from its position in front of RPT1 (white dot) towards RPT2 following the indicated movement vector.

Supplementary Fig. 2

a) Lid Closing


5VHF
Both interfaces closed
5VHH
ATPase interface open
5VHI
Both interfaces open

b) ATPase Rearrangement


TA1
5VHJ
TA2
5VHM
TA3
5VHN
TA4
5VHO
TA5
5VHP
TA6
5VHQ
TA7
5VHR
RPT 2/6 interface open
RPT 2/6 interface closed
RPT 3/4 interface closed
RPT 3/4 interface open

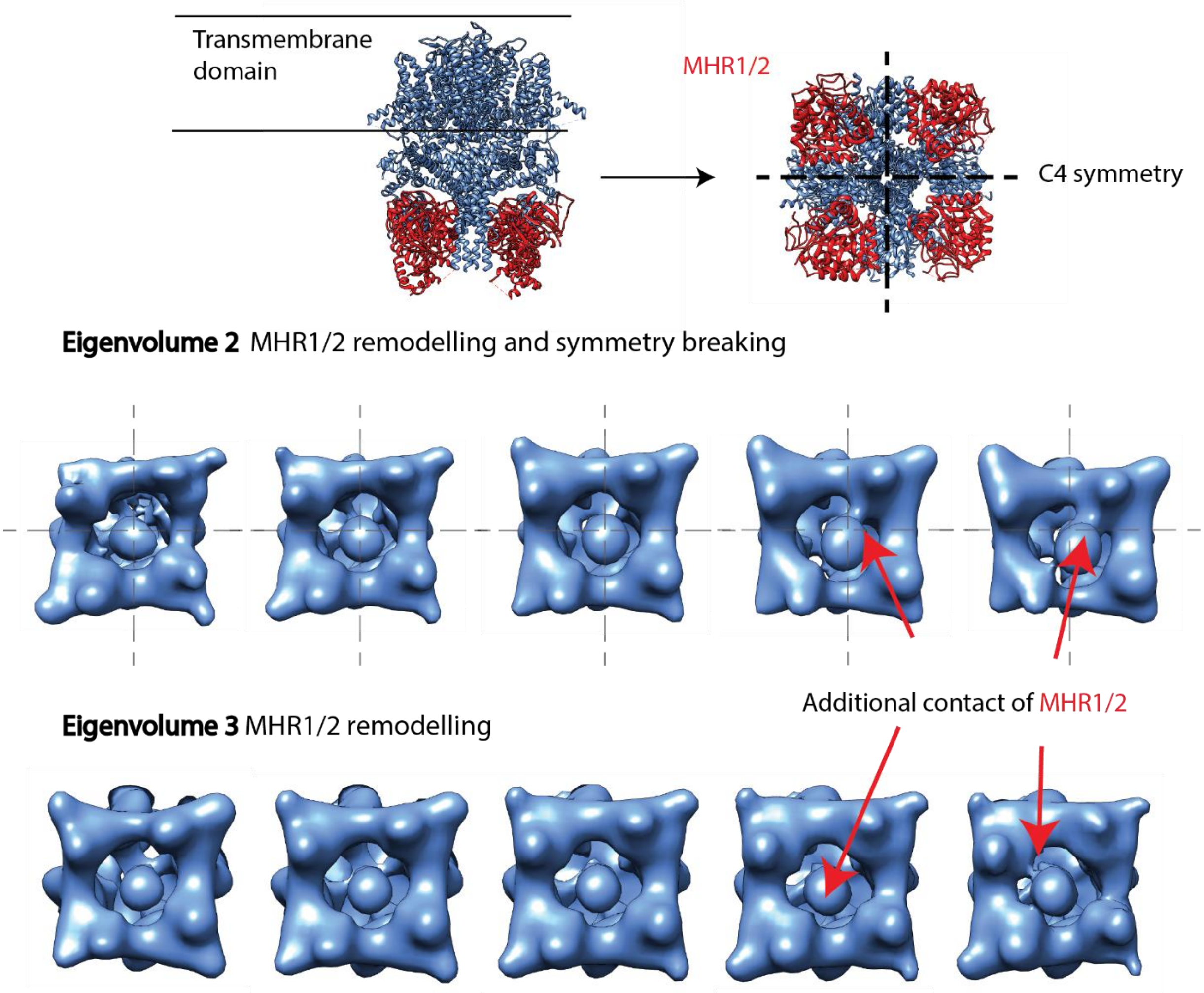
Supplementary Fig. 2: p28-bound 19S proteasome.

Binding of p28 is a necessary component of the 26S assembly pathway, as it positions the 19S correctly on the 20S core particle, to form the full 26 proteasome. CowScape detected—automatically and without *a priori* information—the findings described by Lu *et al.*[17].

a) The results of the PCA analysis confirm the lid closing observed by Lu et al.

b) Using CowScape we determined another motion located in the ATPase ring system that correlates with the lid closing. Here we observe an alternating opening of the RPT2/6 and RPT3/4 interfaces. Based on the energy landscape shown in Fig. 2, we can conclude that the lid closes upon p28 binding. Simultaneously, the ATPase interface changes in going from an open RPT2/6 interface to an open RPT3/4 interface.

Supplementary Fig. 3


Transmembrane domain
MHR1/2
C4 symmetry
Eigenvolume 2 MHR1/2 remodelling and symmetry breaking
Additional contact of MHR1/2
Eigenvolume 3 MHR1/2 remodelling

Supplementary Fig. 3: Analysis of motions in the TRPM4 channel.

TRPM4 has been implicated in various severe diseases, such as diffuse large B-cell lymphomas and complete heart block. Our analysis focused on the cytoplasmic domain by applying a mask during PCA. Additionally, we did not restrict our analysis to C4 symmetry. The original structure (6BQV) is shown in the upper panel, with the MHR1/2 domains marked in red. We detected large symmetry breaks in the second eigenvolume as well as additional contact sites of the MHR1/2 domains with the central coiled coil in the eigenvolumes 2 and 3. These contact sites were not observed in the original analysis by Autzen et al.[16]. Interestingly, the MHR1/2 domains harbour putative binding sites for secondary messengers, such as PIP2 and PIP3. Flexibility seems to increase upon binding of calcium ions, which precedes channel opening. The observed additional binding sites to the coiled-coil might be the reason why Autzen et al. observed more side chains to be visible in the calcium bound state.

References

1. Gavin, A.C. et al. Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415**, 141-147 (2002).
2. Fischer, N., Konevega, A.L., Wintermeyer, W., Rodnina, M.V. & Stark, H. Ribosome dynamics and tRNA movement by time-resolved electron cryomicroscopy. *Nature* **466**, 329-333 (2010).
3. Haselbach, D. et al. Structure and Conformational Dynamics of the Human Spliceosomal B(act) Complex. *Cell* **172**, 454-464 e411 (2018).
4. Haselbach, D. et al. Long-range allosteric regulation of the human 26S proteasome by 20S proteasome-targeting cancer drugs. *Nature communications* **8**, 15578 (2017).
5. Merk, A. et al. Breaking Cryo-EM Resolution Barriers to Facilitate Drug Discovery. *Cell* **165**, 1698-1707 (2016).
6. Elad, N., Clare, D.K., Saibil, H.R. & Orlova, E.V. Detection and separation of heterogeneity in molecular complexes by statistical analysis of their two-dimensional projections. *Journal of structural biology* **162**, 108-120 (2008).
7. Scheres, S.H. Processing of Structurally Heterogeneous Cryo-EM Data in RELION. *Methods in enzymology* **579**, 125-157 (2016).
8. Moriya, T. et al. High-resolution Single Particle Analysis from Electron Cryo-microscopy Images Using SPHIRE. *Journal of visualized experiments : JoVE* (2017).
9. Dashti, A. et al. Trajectories of the ribosome as a Brownian nanomachine. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 17492-17497 (2014).
10. Sigworth, F.J. A maximum-likelihood approach to single-particle image refinement. *Journal of structural biology* **122**, 328-339 (1998).
11. Nakane, T., Kimanius, D., Lindahl, E. & Scheres, S.H. Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION. *eLife* **7** (2018).
12. Bock, L.V. et al. Energy barriers and driving forces in tRNA translocation through the ribosome. *Nature structural & molecular biology* **20**, 1390-1396 (2013).
13. Grant, T., Rohou, A. & Grigorieff, N. cisTEM, user-friendly software for single-particle image processing. *eLife* **7** (2018).
14. Punjani, A., Rubinstein, J.L., Fleet, D.J. & Brubaker, M.A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* **14**, 290-296 (2017).
15. Brown, N.G. et al. Dual RING E3 Architectures Regulate Multiubiquitination and Ubiquitin Chain Elongation by APC/C. *Cell* **165**, 1440-1453 (2016).
16. Autzen, H.E. et al. Structure of the human TRPM4 ion channel in a lipid nanodisc. *Science* **359**, 228-232 (2018).
17. Lu, Y. et al. Conformational Landscape of the p28-Bound Human Proteasome Regulatory Particle. *Molecular cell* **67**, 322-333 e326 (2017).
18. Wang, X. et al. Molecular Details Underlying Dynamic Structures and Regulation of the Human 26S Proteasome. *Molecular & cellular proteomics : MCP* **16**, 840-854 (2017).
19. Bahar, I., Lezon, T.R., Yang, L.W. & Eyal, E. Global dynamics of proteins: bridging between structure and function. *Annual review of biophysics* **39**, 23-42 (2010).

Acknowledgment

We thank Lars Bock for initial discussions on how to best calculate and plot energy landscapes. This work was funded by a grant of the Deutsche Forschungsgemeinschaft (SFB860-TP) to H.S.

4 Biological single particle imaging with integrated differential phase contrast STEM is limited by contrast in vitreous specimen

Cited as: Lambrecht and Riedel (in preparation)

Own contribution:

- Analysis of all data sets
- Argumentation with theoretical work
- Manuscript apart from Materials and Methods
- All figures

Biological Single Particle Imaging with integrated differential phase contrast STEM is limited by contrast in vitreous specimen

Felix Lambrecht[1], Dietmar Riedel[1], Uwe Luecken[1], KaMan Yip[1], Niels Fischer[1] and Holger Stark[1] *

[1] Max-Planck-Institute for biophysical Chemistry, Am Fassberg 11, 37077 Göttingen

Introduction

Imaging of biological molecules with electron microscopy experienced a rising importance in the last decade[1,2]. This ascendency is mostly governed by the advances in the field of conventional high-resolution transmission electron microscopy (CTEM). While in CTEM a parallel illumination is used to acquire images over the full field of view, scanning transmission electron microscopy (STEM) uses a converged beam and scans each position of the acquired image individually[3]. STEM imaging modes can be divided by the collection angle range into brightfield and darkfield techniques. The former makes use of electrons which where barely scattered and are consequently detected close to the central optical axis of the microscope. Darkfield techniques in contrast use the electrons which are collected at higher angles[4].

As for the biologically relevant light elements brightfield STEM does not produce enough contrast, darkfield methods were usually used. Those techniques have the downside of collecting only a fraction of the full electron dose, which for single biomolecule observation implies that damaging doses would be necessary to collect enough signal[5]. Generating differential phase contrast (DPC) is one way to reduce the necessary number of electrons by providing a more efficient contrast formation mechanism[6]. DPC is a special case of center of mass imaging, where the center of mass of the converged beam electron diffraction (CBED) is determined at each probe position r. The phase difference from one scanning position to the other is determined by the displacement the center of mass upon experiencing the local electric forces in the sample (Fig 1a)[7,8]. A split detector is used to determine the relative shift of the center of mass by observing the difference between the two detectors. Here, this detector is divided in four quadrants, to determine the shift on a two-dimensional plane (Fig 1b)[8]. The intensity of the x- and y-components is the subtraction of the two corresponding detector intensities at each scan position $\vec{r}$:

$$I^{DPC_x}(r) \approx I^A(r) - I^C(r)$$

$$I^{DPC_y}(r) \approx \; I^B(r) - I^D(r)$$

By applying a linear combination of both images, the vector image $I^{DPC} = \nabla I^{iDPC}$ can be formed. Lazic *et al.* showed that the integration of I^{DPC} yields the projection of the electric potential of the sample[8,9].

There are several benefits over darkfield STEM: firstly, a larger number of electrons is collected which is superior for light biological elements[5]. Secondly, the intensity scales approximately linear with the atomic number[7]. Both should in theory allow to image at low dose conditions, which is necessary to overcome limitations by radiation damage and charging for biomolecules. Indeed, it was shown that zeolites could be imaged under such conditions with iDPC[10,11].

Apart from such specialized usage cases iDPC STEM should also have beneficial features over CTEM for single-particle analysis. Contrary to CTEM, where defocusing is necessary to introduce phase contrast via an oscillating transfer function, iDPC STEM has maximum contrast transfer in focus and the transfer function is positively defined over a large interval in frequency space as well as a large range of defoci[5]. The latter, namely the large focal depth, makes STEM in general interesting for thick specimen, as it should practically allow to image a few nanometers away from the absolute focus and still achieve maximum contrast transfer. Taking advantage of this property, Wolf et al. were able to tomographically inspect the full volume of a bacterial cell by darkfield STEM[3,12]. The second interesting feature is the specificity of the contrast to the atomic mass, which can be used to quantify single metal binding sites in biological macromolecules[13].

The large focal depth comes by a cost. Both the focal depth (squared) and the resolution (linearly) scale directly with the opening angle. That means the drawback of aiming for high resolution is a compromise in the focal depth, which is necessary to have optimal information transfer throughout the specimen[3,5]. On top if this, increasing the resolution consequently reduces the beam current as the spot size is decreased. This in turn decreases the SNR in the image significantly [14].

On the other hand, dose is always an important parameter when working with biomolecules, as those are naturally prone to radiation damage and usually in particular high resolution features decay rapidly already after a few electrons[15]. When using tomographical darkfield STEM, Wolf *et al*. experienced that the accumulated dose before radiation damage occured was much larger for STEM then for CTEM[16] and additionally Elad *et al.* showed that they could resolve the overall shape of ferritin at more than 120 $\frac{e^-}{\text{Å}^2}$ [13]. This might be an indication for a less harmful damaging behavior of STEM with respect to CTEM.

Implementing a useful single particle workflow for iDPC STEM hence on the one hand includes finding an applicable dose to not generate radiation damage and on the other hand having enough contrast to analyze the acquired images. Here we show that it is possible to achieve a sufficient compromise between dose and contrast on non-biological samples. However, single particle analysis of the acquired images turns out to be hindered by the enhanced low-resolution signal. This intense low-resolution signal scales significantly with the background introduced by the embedding medium in single-particle imaging, which does not allow to gain sufficient contrast at acceptable doses.

Materials and methods

For cryo-EM preparation of ribosomes, five microliters of 0.05 µM vacant *Escherichia coli* 70S ribosomes (kindly provided by M.V. Rodnina) were applied onto cryo-EM grids (Quantifoil 3.5/1 µm, Jena) covered with custom-made pre-floated continuous carbon film; grids were blotted manually under controlled environmental conditions at 4°C and 95% humidity.

Results

As discussed above, firstly a convergence angle had to be found, which offers a compromise between focal depth and resolution, where optimal (small) resolution and optimal (large) focal depth behave reciprocally to each other. From a theoretical point of view [3] we should be able to achieve a resolution of close to 3 Å with a µm-long focal depth with a convergence angle of 4 mrad. Due to stability and handling reasons, the imaging conditions were first assessed on non-vitrified inorganic samples. We tested the achievable resolution at 4 mrad on graphitized carbon black (GCB) under first high-dose conditions and found a clear 3.4 Å signal (Supplemental figure 1). Subsequently the contrast transfer on bio-compatible doses was tested at 68 $\frac{e^-}{\text{Å}^2}$ for different semi-convergence angles (first column in figure 2a). While at 4 and 7 mrad, the signal was nicely visible in the power spectra, it was not detectable at 10 mrad. Additionally, imaging at 10 mrad turned out to be challenging as the focal depth on our non-aberration-corrected setup was tough to adjust. As the spot size decreases with increasing opening angle, we tried to recover the signal by increasing the beam current and indeed found that we construct a constant contrast by increasing the dose with two times the convergence angle squared (Fig 2a). However, for 7 and 10 mrad the dose was far too high for biological imaging.

While passing along this constant contrast diagonal in Fig 2a we recognized that the background signal increased significantly and made the interpretation of the images in real space challenging (Fig 2b, first column, fig. 2c). This overlaying background signal could be completely eliminated by extensive high-pass filtering with a cut-off frequency of $10 \times f_{Nyq}$ on the whole micrograph. As a proof-of-principle, we acquired six micrographs of GCB with $\alpha = 4$ mrad at the same location with each having about 10 $\frac{e^-}{\text{Å}^2}$ per acquisition and aligned and averaged them with cross-correlation based alignment algorithms usually used for single-particle processing. For non-filtered micrographs we could not retain the 3.5 Å signal in the average image after alignment ("alisum"). The alisum of the high-pass filtered micrographs in contrast showed an increased signal with respect to the non-averaged micrographs (Fig 2d). This gave us confidence that the acquisition at $\alpha = 4$ mrad should yield sufficient contrast and resolution to perform single particle alignments while having a theoretically large focal depth. This focal depth we cross-validated by performing focus series on a very thin layer of a silicon specimen, verifying more than 500 nm with about constant contrast transfer (Supplemental figure 2). Most important, the doses at higher convergence angles where far too high to perform imaging on biological specimen.

With this parameters, we aimed to image a vitrified sample of the *A. thermophilus* 20S proteasome. This molecule is especially suited for prototyping, as it is D7 symmetrical and therefore a rather small number of images should be necessary to refine to high resolution. Either colloidal gold added to the sample or the edge of the grid hole where used to come close to absolute focus. We acquired a dataset with the same imaging settings as used above for GCB and filtered the raw micrographs accordingly (Fig 3a, upper right panel). However, single particles were barely visible and automated extraction algorithms such as gautomatch even failed with template matching. Consequently, the dose was increased by a factor of two and single particles were nicely visible (Fig 3a, upper left panel). However, there were still large differences in contrast within one and the same micrograph. As a control, the same specimen was acquired with $\alpha = 7$ mrad but no particle was visible at all, with comparable and even doubled dose.

In contrast to CTEM there are two resolution determining features in STEM: the brightfield disc diameter k_{Bf}, which depends on the semi-convergence angle and as in CTEM the pixel size, which depends on the magnification and the resolution's dependence originates from Nyquist's theorem. Consequently, we determined the dependence of the resolution and contrast on the pixel size at $\alpha = 4$ mrad (Fig 3b). Interestingly, the contrast increased with larger pixel sizes and slight undersampling might be advantageous.

As the imaging of the 20S turned out to be non-straightforward, we also tested the 70S ribosome, as a second model system, with a larger diameter as well as a higher contrast due to the containing RNA. Firstly, the same imaging settings were used as for the 20S proteasome, resulting in extremely bright and nearly shapeless spots. Anyways, the contrast increased extremely. We decreased the dose to 34 $\frac{e^-}{\text{Å}^2}$ and could acquire images with shaped particles and a contrast which allowed us to extract the single particles with automated, template-free tools. Notably, for all biomolecules imaged, we could never see a clear boundary between molecule and solvent, as one is used to from

CTEM (Supplemental Figure 3), but the outermost solvent layer around the molecule was rather fuzzy. Additionally, we recognized different qualities of contrast in one and the same micrograph.

Next, single particles were extracted manually from the proteasome data (Fig 3a) and the ribosome data (template-free, Fig 4a). The coordinates were determined on the highpass-filtered micrographs and particles were then extracted from the non-filtered ones, to have more flexibility with respect to filtering. The two 4 mrad proteasome datasets shown in Fig. 3a showed a clear difference with respect to the background. Where in the 64 $\frac{e^-}{\text{Å}^2}$ dataset boundaries between background and particle could be achieved in the 2D class sum, this was not possible, when doubling the dose. Also, another dataset at 3 mrad was acquired at intermediate dose, which showed the most isotropic and contrast-rich class sums. For the ribosomes, we could not detect any shape in the class sums acquired with more than 34 $\frac{e^-}{\text{Å}^2}$, but rather damaged particles where visible (fig. 4b and c), even though in some rare particles higher resolution features could be found (fig. 4b). Still, even the best class averages did only show really low resolved features like domain separation (e.g. small and large subunit in the ribosome or central pore in the proteasome). The outer particle boundary was never well defined in any of the class sums, regardless of the imaging condition or the molecule. Additionally for the ribosome it became apparent that for higher doses radiation damage led to blurring of the images (fig. 4b). All in all, no class sum could be calculated which was even close to the theoretical resolution, even though the resolution and the contrast on the inorganic GCB looked fairly promising.

Discussion

We have shown that iDPC STEM is capable of resolving near-atomic information at low semi-convergence angles and low doses with a really comfortable focal depth on GCB, which was shown before for the radiation sensitive zeolites [7,10,17]. These images could be aligned and averaged by single-particle algorithms, after filtering out low-resolution background components, which increase with the semi-convergence angle. Also, the contrast itself depends on the semi-convergence angle but can be compensated for by increasing the dose. This might be a statistical problem of the number of electrons hitting the quadrants while decreasing the probe size. However, as we could show that a semi-convergence angle of 4 mrad yielded enough resolution, we focused on this value for the analysis of biological samples as the necessary dose was in a regime that no radiation damage could be expected and the focal depth did not limit the unautomated acquisition.

When going to vitrified samples, the background increased dramatically. Particle extraction was only possible in high-pass filtered images as otherwise the cloudy features overwhelmed the image as was previously described for single-particle images in darkfield STEM [13]. We found that in a respectable dose regime below 200 $\frac{e^-}{\text{Å}^2}$ the contrast was only sufficient with a semi-convergence angle of 4 mrad and below, underpinning the trend seen for GCB. Rez [18] discussed from a theoretical point of view that ice might limit the contrast of STEM for single particle cryo-EM compared to CTEM. As CTEM's contrast transfer function suppresses low-resolution signals, the background signal is not visible in CTEM[13]. Additionally, for STEM the sample does not behave as a weak phase object as in CTEM. Therefore the scattering of the embedding medium adds sufficiently to the image formation. This in turn leads to a thickness-dependent suppression of contrast throughout the frequency spectrum, which is about a factor of two compared to CTEM when the surrounding is as thick as the molecule but might become a factor of 5 already for a 20 nm thick ice film[19]. Recent studies on the ice thickness of 20S proteasome preparations even assume a distribution of ice thicknesses ranging from 50 to 200 nm [20]. Unfavorable is also in accordance with our finding that slight undersampling increases the visibility of the particles: contrast scales with the pixel dimension[19,21] and consequently, we could recover contrast loss by the ice thickness by increasing the pixel size.

If we assume that the large low-resolution background signal originates significantly from the surrounding medium and is proportional to the thickness, then this also explains the observation that the contrast changes within a few ten nanometers in one field of view. This on the one hand emphasizes the sensitivity of iDPC STEM to light elements in the water but on the other hand this also implies that individual filter values for certain regions or even particle would be needed. During the processing of the data shown above, we found micrographs, which needed to be filtered to a frequency which affected the low-resolution frequencies of the individual particles. As alignment algorithms are usually sensitive to these frequencies, affecting those might lead to a decrease in alignment accuracy. On the other hand, the wavenumber to which the background extends, might be a reasonable estimate of the evenness of the ice layer.

Interestingly, proteasomes and ribosomes behaved differently in our experiments. Ribosomes showed some low-resolution features at significantly lower doses, which can be explained by ribosomes being larger in diameter and contain RNA. Increasing the doses led to immediate burning of the ribosomes, which became apparent in the shapeless class sums shown in figure 3. Based on our results we were not able to find a sufficient compromise between contrast – which should increase with dose – and radiation damage. This contradicts the hypothesis that STEM might be less harmful for single particle analysis, which was shown before for tomography.

The promising results of STEM tomography published in the last years [12,18] might lead the way to a more suited application for iDPC-STEM than in single particle analysis. Another opportunity which might lead to interesting biological findings is the evaluation of the iDPC-Signal together with the HAADF signal as both detectors can be inserted in parallel and the analysis of metal species with STEM was already shown before [13]. For the use in single particle analysis, contrast enhancement would be necessary which might possibly come from more sensitive detector technology in the future.

1. Renaud, J.-P. *et al.* Cryo-EM in drug discovery: achievements, limitations and prospects. *Nat. Rev. Drug Discov.* **17,** 471–492 (2018).
2. Cheng, Y. Single-particle cryo-EM—How did it get here and where will it go. *Science (80-.).* **361,** 876–880 (2018).
3. Elbaum, M., Wolf, S. G. & Houben, L. Cryo-scanning transmission electron tomography of biological cells. *MRS Bull.* **41,** 542–548 (2016).
4. Pennycook, S. J. & Nellist, P. D. *Scanning Transmission Electron Microscopy*. *Imaging and Analysis* (Springer New York, 2011). doi:10.1007/978-1-4419-7200-2
5. Elbaum, M. Quantitative Cryo-Scanning Transmission Electron Microscopy of Biological Materials. *Adv. Mater.* **30,** 1706681 (2018).
6. Dekkers, N. H. & de Lang, H. DIFFERENTIAL PHASE CONTRAST IN A STEM. *Optik (Stuttg).* **41,** 452–456 (1974).
7. Yücelen, E., Lazić, I. & Bosch, E. G. T. Phase contrast scanning transmission electron microscopy imaging of light and heavy atoms at the limit of contrast and resolution. *Sci. Rep.* **8,** 2676 (2018).

8. Lazić, I. & Bosch, E. G. T. *Analytical Review of Direct Stem Imaging Techniques for Thin Samples*. *Advances in Imaging and Electron Physics* **199,** (2017).

9. Lazić, I., Bosch, E. G. T. & Lazar, S. Phase contrast STEM for thin samples: Integrated differential phase contrast. *Ultramicroscopy* **160,** 265–280 (2016).

10. Carlsson, A., Alexandrou, I., Yücelen, E., Bosch, E. G. T. & Lazić, I. Low Dose Imaging Using Simultaneous iDPC- and ADF-STEM for Beam Sensitive Crystalline Structures. *Microsc. Microanal* **24,** (2018).

11. Bosch, E. G. T., Lazic, I. & Lazar, S. Integrated Differential Phase Contrast (iDPC) STEM: A New Atomic Resolution STEM Technique To Image All Elements Across the Periodic Table. *Microsc. Microanal.* **22,** 306–307 (2016).

12. Wolf, S. G., Houben, L. & Elbaum, M. Cryo-scanning transmission electron tomography of vitrified cells. *Nat. Methods* **11,** 423–428 (2014).

13. Elad, N., Clare, D. K., Saibil, H. R. & Orlova, E. V. Detection and separation of heterogeneity in molecular complexes by statistical analysis of their two-dimensional projections. *J. Struct. Biol.* **162,** 108–20 (2008).

14. Buban, J. P., Ramasse, Q., Gipson, B., Browning, N. D. & Stahlberg, H. High-resolution low-dose scanning transmission electron microscopy. *J. Electron Microsc. (Tokyo).* **59,** 103–12 (2010).

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

16. Townsend, P. D. *et al.* Global low-frequency motions in protein allostery: CAP as a model system. doi:10.1007/s12551-015-0163-9

17. Lazic, I., Bosch, E. G. T., Lazar, S., Wirix, M. & Yücelen, E. Integrated Differential Phase Contrast (iDPC)–Direct Phase Imaging in STEM for Thin Samples. *Microsc. Microanal.* **22,** 36–37 (2016).

18. Rez, P., Larsen, T. & Elbaum, M. Exploring the theoretical basis and limitations of cryo-STEM tomography for thick biological specimens. *J. Struct. Biol.* **196,** 466–478 (2016).

19. Rez, P. Comparison of phase contrast transmission electron microscopy with optimized scanning transmission annular dark field imaging for protein imaging. *Ultramicroscopy* **96,** 117–124 (2003).

20. Rice, W. J. *et al.* Routine determination of ice thickness for cryo-EM grids. *J. Struct. Biol.* **204,** 38–44 (2018).

21. Henderson, R. The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules. *Q. Rev. Biophys.* **28,** 171 (1995).

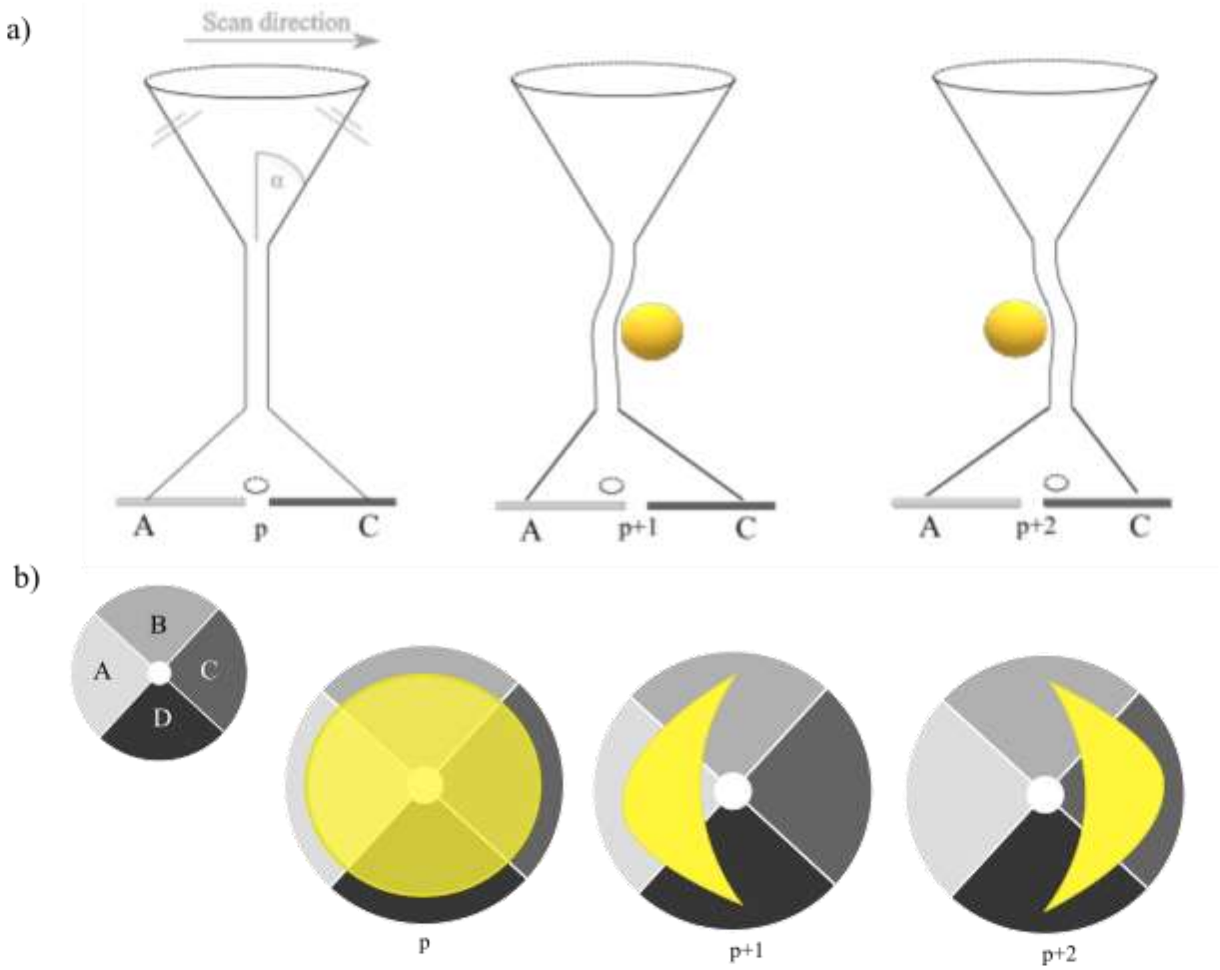


Figure 1: General principle of differential phase contrast STEM with a four-quadrant detector

a) Shows a slice through a generalized ray-path. The beam converges to a defined probe under the semi-convergence angle α. The focus stays constant over a certain range (the focal depths) before the ray diverges again and hits the detector, forming the CBED pattern. Without a sample, the center of mass (COM, circle in a) of the CBED pattern is aligned in the center of the detector. Upon scanning, the probe experiences the sample and is deflected, which leads to a shift of the COM on the detector. The full four-quadrant detector is shown in

b) , where the deformation of the probe away from the charged particle is shown. The deflection in x direction is the difference of the quadrants A and C, whereas the deflection in y is the difference of the quadrants B and D

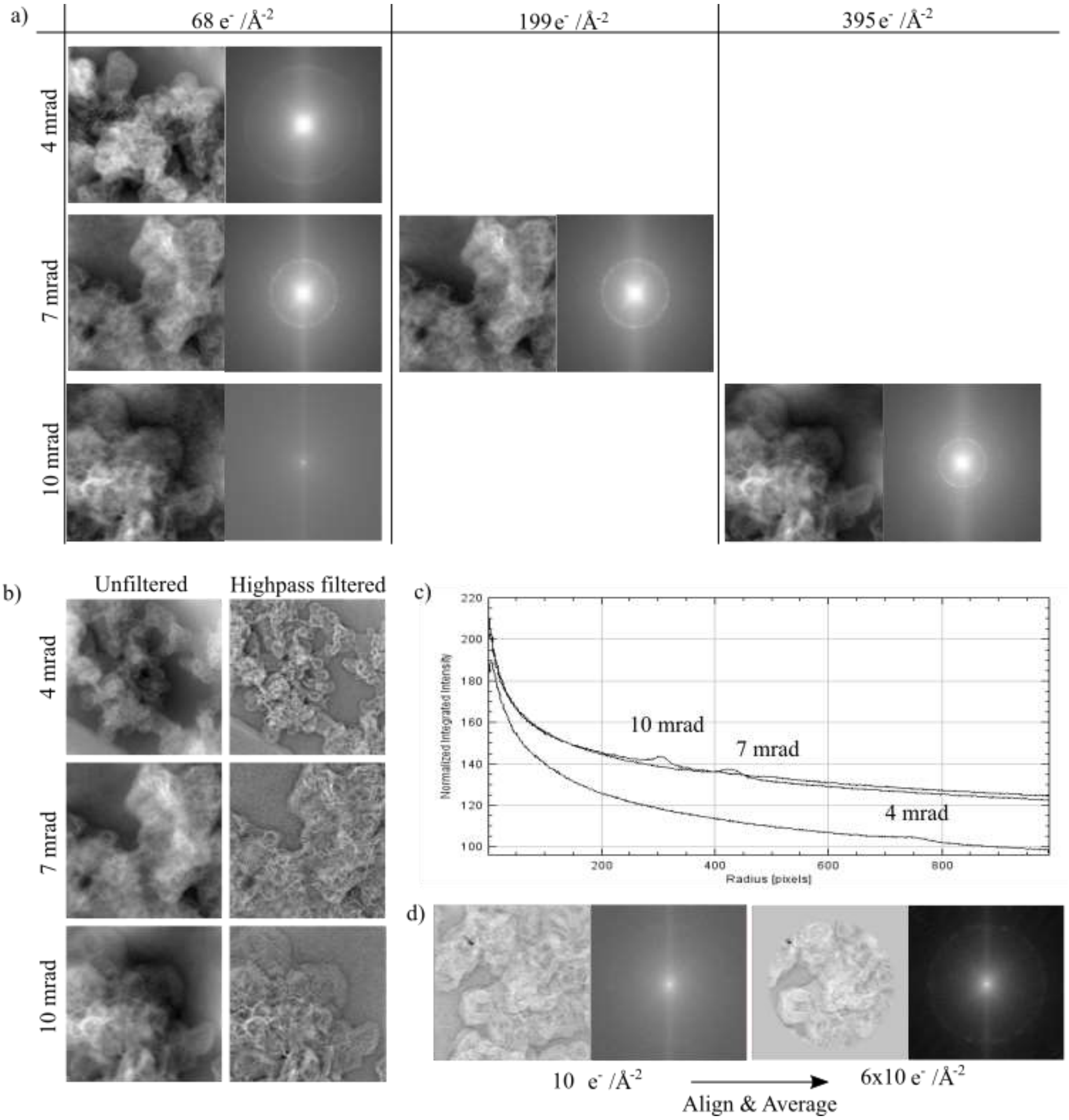

a)
68 e⁻ /Å⁻²
199 e⁻ /Å⁻²
395 e⁻ /Å⁻²
4 mrad
7 mrad
10 mrad
b)
Unfiltered
Highpass filtered
4 mrad
7 mrad
10 mrad
c)
Normalized Integrated Intensity
Radius [pixels]
220
200
180
160
140
120
100
200
400
600
800
10 mrad
7 mrad
4 mrad
d)
10 e⁻ /Å⁻²
Align & Average
6x10 e⁻ /Å⁻²

Figure 2: Graphitized Carbon Black (GCB) imaged under low-dose conditions for different α

Imaging biological molecules requires to go to low-dose conditions, while the theoretical achievable resolution should be better than 3Å. All signals visible in the Fourier transforms of the individual images correspond to the 3.5 Å signal of GCB and the relative positions originate from the different pixel sizes.

a) Shows that keeping the dose constant at low $\frac{e^{-}}{\text{Å}^{2}}$ results in a loss of contrast which even results in the signal not being detectable at α = 10 mrad. We found that to keep the constant about equal while increasing the opening angle, the dose has to scale with two times the semi-convergence angle squared. The pronunciation of the low-resolution region also increases with increasing opening angle.
b) Shows that extensive high-pass filtering at a cut-off frequency corresponding to 0.1 times the Nyquist frequency is necessary to recover the full information in the images.
c) Rotational power spectra of the unfiltered images from b) showing the increase in background for larger convergence angles

d) After applying high-pass filtering as shown in figure b) the signal to noise ratio could be increased by averaging.

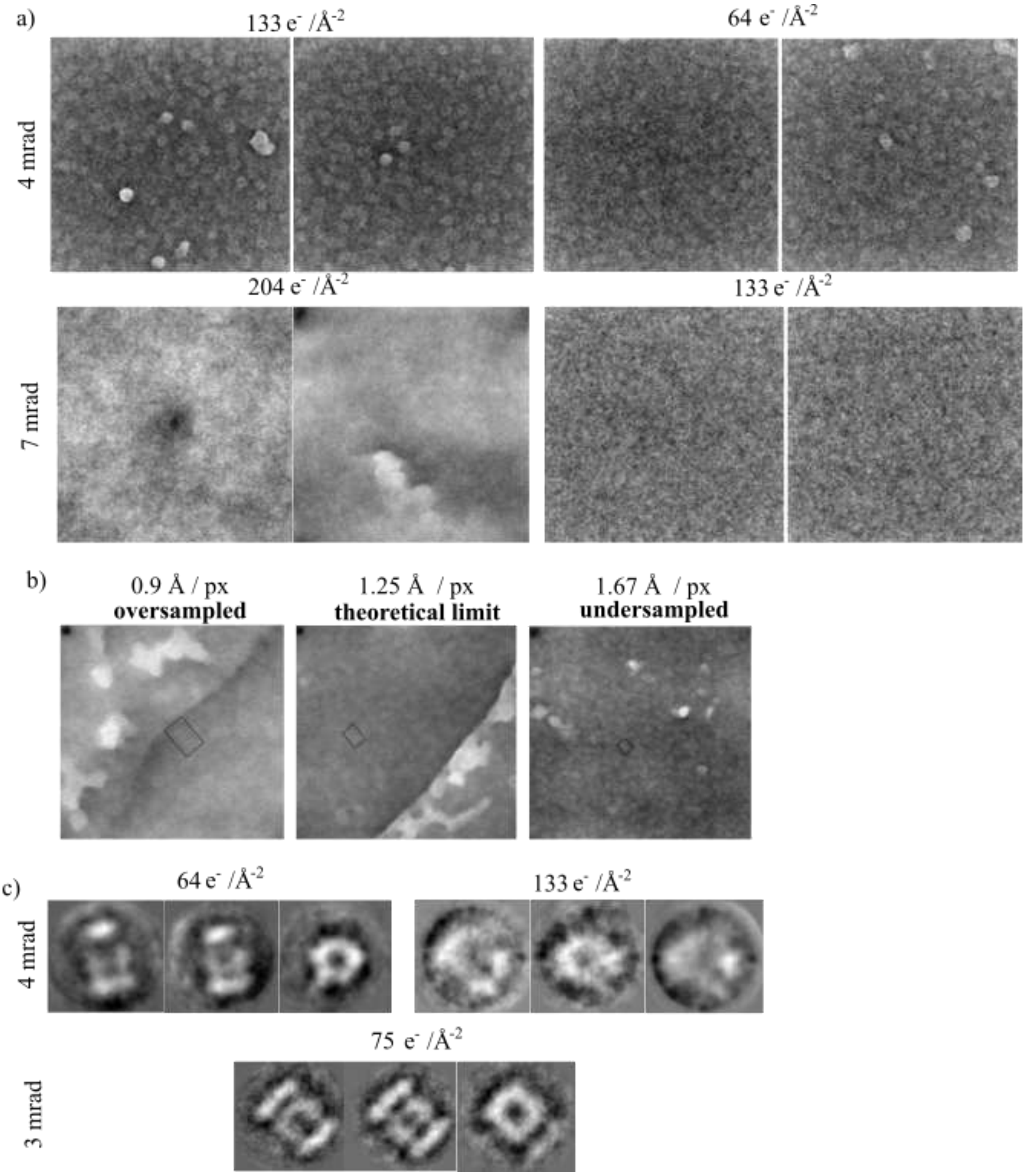

a)
133 e⁻ /Å⁻²
64 e⁻ /Å⁻²
4 mrad
204 e⁻ /Å⁻²
133 e⁻ /Å⁻²
7 mrad
b)
0.9 Å / px
oversampled
1.25 Å / px
theoretical limit
1.67 Å / px
undersampled
c)
64 e⁻ /Å⁻²
133 e⁻ /Å⁻²
4 mrad
75 e⁻ /Å⁻²
3 mrad

Figure 3: Imaging of 20S proteasomes with iDPC STEM

a) Exemplary micrographs showing 20S proteasomes imaged with different dose at 4 and 7 mrad, respectively. At 7 mrad semi-convergence angle the contrast was not sufficient do distinguish single particles. At 4 mrad semi-convergence angle, a dose of 64 $\frac{e^-}{\text{Å}^2}$ hardly allowed to recognize particles. At 133 $\frac{e^-}{\text{Å}^2}$ nearly all micrographs showed sufficient contrast to distinguish single proteasomes. All micrographs were high-pass filtered to remove the enhanced low-resolution signal.

b) Shows the dependence on the pixel size. Sufficient sampling is judged by the theoretically achievable resolution at 4 mrad. We see that a larger pixel size increases the contrast significantly. The particles imaged at 64 $\frac{e^{-}}{\text{Å}^{2}}$ (undersampled) and 133 $\frac{e^{-}}{\text{Å}^{2}}$ were extracted and class sums are shown in c)

c) Additionally a dataset was taken at 3 mrad at intermediate dose, which showed nice contrast in the raw micrographs.

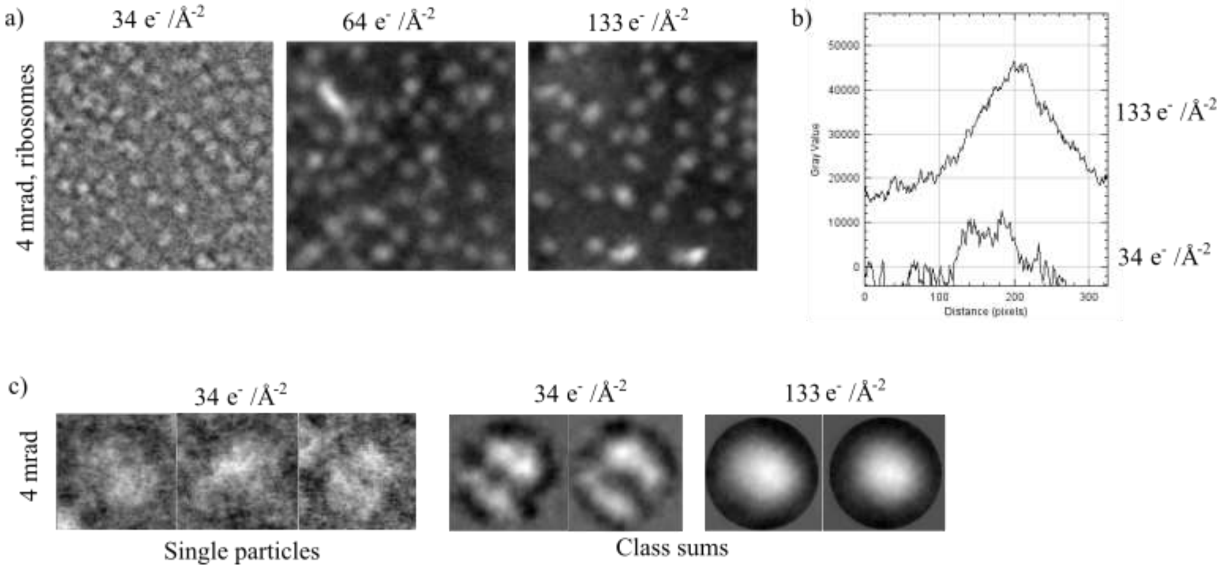

a)
34 e⁻ /Å⁻²
64 e⁻ /Å⁻²
133 e⁻ /Å⁻²
4 mrad, ribosomes
b)
50000
40000
30000
20000
10000
0
Gray Value
0
100
200
300
Distance (pixels)
133 e⁻ /Å⁻²
34 e⁻ /Å⁻²
c)
34 e⁻ /Å⁻²
34 e⁻ /Å⁻²
133 e⁻ /Å⁻²
4 mrad
Single particles
Class sums

Figure 4: Imaging of ribosomes with iDPC STEM

a) Ribosomes in contrast could be resolved at lower doses. The best defined features in the exemplary micrographs could be achieved at 34 $\frac{e^{-}}{\text{Å}^{2}}$, whereas larger doses led to a loss in features, which is in contrast with the respective class sums shown in

b) Line plots over single ribosome particles underpin the trend to lose information already hown in the micrographs

c) Even though rarely single ribosome particles could be found showing higher resolved features, no class sum showed features, even close to the theoretically feasible resolution.

5 Discussion

5.1 Conformational Landscapes from Cryo-EM Experiments

In the first two publications presented here, the concept of deriving conformational coordinates from cryo-EM experiments by extensive 3D classification and subsequent 3D multivariate statistical analysis was introduced. These conformational coordinates come as eigenvectors of the voxel-wise covariance matrix. The high-dimensional ($d = N^3$) volumes can then be projected into a low-dimensional (usually $d = 2$) space, representing their conformational state.

This approached was used to solve different questions in the past two years, since the publication of *Haselbach and Schrader (2017)* which are listed in 5.1. The plethora of usage cases in different biological contexts already demonstrates the usefulness for the interpretation of molecular dynamics from experimental data. In the following discussion, first, algorithmic extensions and different options in developing CowScape further shall be shown and then its future usage will be discussed with interfaces to different disciplines.

Table 5.1: Overview on usage cases of the CowScape idea

Interest	**Macromolecular Complex**	**Reference**
Judging upon the quality of the assembly	APC	Lambrecht (submitted)
Describing the conformational motion of single subunits	Rpn1, 26S proteasome	Lambrecht (submitted), Haselbach and Schrader (2017)
Order the states on an unknown reaction pathway	19S proteasome, spliceosomal Bact	Lambrecht (submitted), Haselbach et al. [58]
Find new and unknown conformational states	spliceosomal Bact	Haselbach et al. [58]
Find allosteric mechanisms by comparing landscapes	26S proteasome, TRPM1, FAS	Haselbach and Schrader 2017, Lambrecht (submitted), Singh (in preparation)
Detect concerted movements	APC, 19S proteasome	Lambrecht (submitted)

5.1.1 PCA on Density Maps

There have been attempts to describe the heterogeneity in the sample by means of PCA, mostly in the 2D case. Penczek et al. [100] et al. first showed that it is in priciple possible to estimate the covariance of a sample by a bootstrapping algorithm. Nameley, they resampled all 2D images into random volumes and calculated the covanriance matrix from them. This approach is generally not limited by the pure number of volumes, but the information content of the individual volumes is low. Furthermore, as bootstrapping allows single observations to be contained in more than one volume, no valid estimate of the energies can be done. However, the CowScape approach is based on the same idea to estimate covariances directly from the volumes, but the 3D classification prior to the PCA analysis enhances the interpretability of the data, which was shown in the supplement to *Haselbach and Schrader (2017).*

It is nevertheless noteworthy that in the past years, there have been attempts to directly estimate the covariance matrix or the principal components of the 3D space directly from the 2D observations [71, 141]. Even though such concepts would be a computational shortcut to the approach presented here, we where not able to implement one of the algorithms in a form that yielded a converged and interpretable result and no testable version of the algorithms is available until today.

5.1.1.1 The Potential of Clustering Algorithms

For the analyis of the conformational ensemble, PCA has the advantage that it directly orders each volume in a space, which we define as the conformational space. However, if one carefully looks through the volumes of an ensemble, then it becomes clear that volumes representing the same or equal conformations exist in the ensemble more then once. This seems to be an intrinsic property of the way the ensemble is generated by the 3D classification and might be uvercome by recent implementations in our department (data not shown). However, as one would like to reconstruct the average volume of all volumes belonging to one energetical minimum, clustering is necessary.

A first naive approach would be to cluster the volumes in the N^3-dimensional voxel-space. However this imposes two problems, the first being computational complexity and the second being redundancy. Redundancy describes the property of the volumes that neighbouring voxels will contain nearly the same information content. This leads to a less robust clustering result as most of the metrics used in clustering will calculate a voxel-wise distance [130]. Therefore it is appreciable to cluster after PCA was applied to reduce this redundancy. In statistics and unsupervised machine learning, this pre-treatment by PCA is referred to as "whitening" [65]. In cryo-EM, this feature was already used for years in 2D classification prior to 3D reconstruction [149]. When applying the same strategy to 3D the importance of whitening will increase as also redundancy increases when projecting the information from 2D to 3D. It will be a matter of testing how many and which principal compomnents are necessary to guide clustering. In the most "human" form of clustering, it will be already sufficient to let the user cluster within the 2-dimensional landscape. A more physical correct form might be grouping together all volumes falling in one minimum in the energy landscape, for which basically just an energy threshold has to be defined. This resembles to some extent a flood-fill algorithm. However, it will be interesting to apply clustering algorithms in higher-dimensional spaces and learn about, how this "conformational cleaning" can affect resolution. Thereby the usual high-resolution refinement scheme presented in chapter 1 is quasi-inverted and refinements could in future be done starting from many classes and converge to a few stable ones.

5.1.1.2 Hierarchy of Conformational Landscapes

When thinking about clustering volumes in one minimum together, the discussion about the glass-like state model comes into play again. It was discussed in the introduction that there is a hierarchy of states and "zooming" into one well of the energy landscape should allow to resolve more, energetically flatter and lower-amplitude motions. Translating this into the properties of PCA that would mean that within the minima defined by the low-order PCs the higher-order PCs should allow us to identify other minima. Materese et al. [89] showed in a very intersting study that one can use PCA on ensembles generated from NMR and MD and that those have an intrinsic hierarchy. Namely, when they focussed on one minimum in a landscape constructed by the low-order PCs they could resolve new minima and this scheme hold for five iterations. By traversing down in the hierarchy the probability distribution shaped more and more gaussian.

This iterative scheme was also tested with our approach but already after two hierarchy levels, no further minima were obtained and the shape looked completely gaussian. It is fair to assume that for our testcases the statistics was just not sufficient and hence the resolution did only allow to further inspect the first few eigenvectors.

A similar scheme was applied already in Fischer et al. [42], Haselbach et al. [58]. They ordered the volumes on one PC and clustered in this one dimension. These clusters where then again classified and hence the local motions could be refined to more detail. In *Lambrecht (submitted)*, we described this part as an iterative variant of the CowScape algorithm. However, both methods, the dissection of the landscape by the hierarchical organization of the PCs as well as the "relaxation" [42] of one conformational coordinate to improve the refinement of the substates resemble the glass-like state model. This emphasizes that enhanced cryo-EM classification schemes and the analysis by PCA can recover the true physical nature of the hierarchical organization of conformational states. The limitation seems to be the resolution of the individual maps. As described above, the nature of the higher-order tiers are secondary structure and side-chain movements. As with the current numbers of particles and the classification algorithms we can hardly resolve low-populated states to medium resolution, these tiers are nowadays out of scope.

5.1.2 Estimation of the Landscape

The primary aim of the algorithm presented in *Haselbach and Schrader* and *Lambrecht* was to find an interpretable visualization of the conformational space of the underlying sample. The overall workflow after calculating PCs firstly includes the choice of the PCs in which the landscape should be projected and secondly the interpolation of the final landscape.

5.1.2.1 Choice of Principal Components

The principal components (PCs) in general encode the biological information which will be assessable through the landscape. In general that implies two things:

- In the end, the PCA will calculate PCs on any dataset given to it. That is shown in *Lambrecht, 2018*, where in the first figure compositional rather then conformational heterogeneity is displayed. This resembles the general "garbage-in-garbage-out"-philosophy in many image processing applications that if the dataset is not of good quality, the landscape will display non-sense information.
- The implementation is done in a way that any system of PCs can be used as basis function for a given dataset. Otherwise, generalized components (e.g. "Define a common basis function set for all AAA+ ATPases") or comparism between datasets would not be possible. That also implies that by mischoosing the PCs, non-sense information can be generated, again.

However, in the current implementation, the choice of the PC is a left to the user, after optically inspecting them (*Lambrecht,2018*). As discussed above, the motions displayed by PCA should be hierarchically ordered into different levels that relate more to locality then to significance. Another quite good estimate for significance is shape of the probability distribution of the eigenvector, which *in lingo* of statistics would act as a discriminant measure. As a PC just displays one single state and the fluctuations within, it converges more and more to a single gaussian [1, 24]. Consequently, an automated way of advising "significant" PCs to the landscape might in the future involve calculating the deviation from a gaussian function [89]. Nevertheless, the interpretation of the movement is still at the core of structural biology and a visual inspection of the encoded motions will allways be necessary.

5.1.2.2 Interpolation of the Landscape

The aim of interpolating all the intermediate steps is at the moment governed by the easier visual interpretation and foremost the identification of potential pathways. However in the two publications in this thesis the general look of the landscapes differ significantly. The first versions of the algorithms did not have their own visulization and interpolation tool, hence the interpolation was done in MATLAB. The used algorithm was based on Delaunay triangulation [25]. Triangulation is a geometrical process, which interpolates the height values linearly on an area between three adjacent points.
This is for sure not the most physical correct interpretation of the data, as it yields a non-continuos probability function. Hence, the final implementation of the CowScape algorithm used Kernel Density Estimation (KDE) with a gaussian kernel function to estimate the underlying probability density [129]. In KDE, each data point is described by the gaussian and each point is the sum of all gaussians normalized by the number of data points in the plot (a thorough description can be found in the supplement to *Lambrecht (submitted)*). KDE also offers its own solution to the process of forming probability-based clustering in the landscape as discussed above, as the clustering in this case depends on the minima found by KDE and hence on the widths of the underlying kernel functions, which can be estimated from the data e.g. by cross-validation as implemented in the current version of the algorithm [55].
When generalizing the concept of KDE a bit more, i.e. unfixes the means of the gaussians, a gaussian mixture model (GMM) could be used. GMM's describe a probability distribution by a number of k gaussians each having their own parameter set $\Theta_k = (\mu_k, \Sigma_k)$, with μ_k being the mean and Σ_k being the covariance matrix of the gaussian mixture component. As we can assume that especially the higher-order tiers will have only a few gaussian shaped peaks, keeping $k << n$, (with n being the number of data points), might yield a more robust version of the algorithm presented here. However, one major shortcoming of the approach can not be overcome by any interpolation method, which is the fact that the short-lived transition states representing the barrier between the local energetical minima are not resolved at all. A complete discription of the energetical path from one conformation to the other is thus not possible.

5.1.3 Interpreting Allostery with cryo-EM

Cryo-EM is powerful when it comes to resolving different conformational states of a molecule in one sample or under different biochemical conditions and the examples in the literature are numerous. By comparing differences between the resolved structures conclusions about the significant motions can be drawn and by interpolation methods similar to the ones used here trajectories can be drawn[95]. However, when it comes to allostery this single-structure based view is retained in the field. Roseman et al. [115] discussed already on a very low-resolved model of the GroEL chaperonin how ATP-binding cooperatively affects the subunits of the complex. Still 17 years later, when cryo-EM was already near atomic the interpretation of allostery came from a single-structure point of view: Zhang et al. [157] discusses changes in intersubunit binding sites, interestingly while oberserving single histidines at 4.5 Åresolution. Lü et al. [84] uses measures of assymetry to determine allosteric communications between different states of a NMDA receptor and Banerjee et al. [4] already showed the structural differences of a human p97 map with and wthout an allosteric inhibitor. In *Haselbach and Schrader (2017)* we have basically used all these features as well: detection of symmetry breaking, changes in subunit interfaces and the influence of binding of a small inhibitor.

We added yet another level of description to allostery, which is based on the ensemble nature discussed in the introduction. By comparing the landscapes with and without the inhibitor we could conclude about the type information transfer: the main conformation is always the same but the chance to observe another conformation was far lower in the inhibitor-bound state. This implies rigidification or stabilisation of the 19S part upon inhibitor binding in the 20S (figure 2), which resembles the characteristics of a conformational selection model (binding and stabilisation of a distinct state) and not of an induced-fit like behavior. In contrast to this, the TRPM1 channel undergoes a slight shift in the preferred conformation upon calcium binding (*Lambrecht (submitted)*, figure 3).

In general this shows the capability of the method to dinstinguish different models describing binding or allostery. However, Induced fit and conformational selection actually describe the *binding* of the molecule (Oprozomib or Calcium in our case) and not the results of binding up to a few hundred Å away from the binding site. In the two examples presented here, the analysis is restricted to the fact that binding occured in the rigid parts of the molecule and the affected, dynamic parts of the molecule were analysed (whereas the rigid parts were masked out). This also limits the interpretability of the results with respect to the classic macroscopic models. However, current investigations in the group show that there are molecules, in which the amplitudes of the effector and affector regions are about comparable so that masking might not be necessary and a more thorough investigation might be possible.

On the other hand, the fact that binding especially of small molecules happens of course on small scales whereas the affected regions (the central coiled-coil in the TRPM1 or the 19S in the proteasome) remodel on larger scales seems to be common theme. This could be a signal-enhancement mechanism on the molecular scale. While signal enhancement on the systemic, cellular level is quite understood (e.g. by cascading signal transduction as in kinase cascades) this molecular level in large macromolecular machines could be a prior enhancing step. This seems logical if one assumes that the binding of a small molecule like calcium might just lead to a charge difference of two in the binding pocket (which leads to very local deformations) but for the opening of a channel gate much larger remodeling is necessary. Understanding mechanisms of signal enhancement might even lead to new concepts of biosensors for the biotechnological application.

5.1.4 Quality Assesment and Biochemical Optimization

It was already discussed in the section about the choice of conformational coordinates that these coordinates can also encode compositional rather then conformational heterogeneity. Until now this property was discussed as a downside of the method but in fact it is rather the opposite: it is a guide. Elmlund et al. [38] state the sample preparation step as being one of the major "nuts and bolts" for high-resolution structure determination and indeed if one assumes that the major limit for high-resolution is still statistics it makes sense to not waste acquisition time to image broken particles. Several methods where introduced to stabilize the macomolecular complex prior to imaging, primary crosslinking gradient fixation (Grafix [139]) and buffer screening methods based on differential scanning fluoromoetry (Proteoplex [14]). However, there is yet not one objective measure on how "broken" a sample is. We have shown that by computing a landscape of the full sample, compositional heterogeneity can be visualized in a comprehensive way. By using one set of eigenvectors (one ONB) calculated once on one or more samples, one can easily compare samples not only with respect to their conformational but also compositional heterogeneity. For optimization purposes, it would be e.g. a fair strategy to screen the few most stablest conditions from a proteoplex screen and visualize by a CowScape the intactness of the different samples. This introduces another layer of stability screening especially for more fragile samples. Using just the most stablest fraction of a sample for subsequent refinements and dynamics analyses makes these algorithms probably converge to higher resolutions. Taking into account all these considerations one can advice an "heterogeneity-based" refinement scheme for cryo-EM single particle analysis which in the future has to be compared against the classical workflow (figure 5.1).

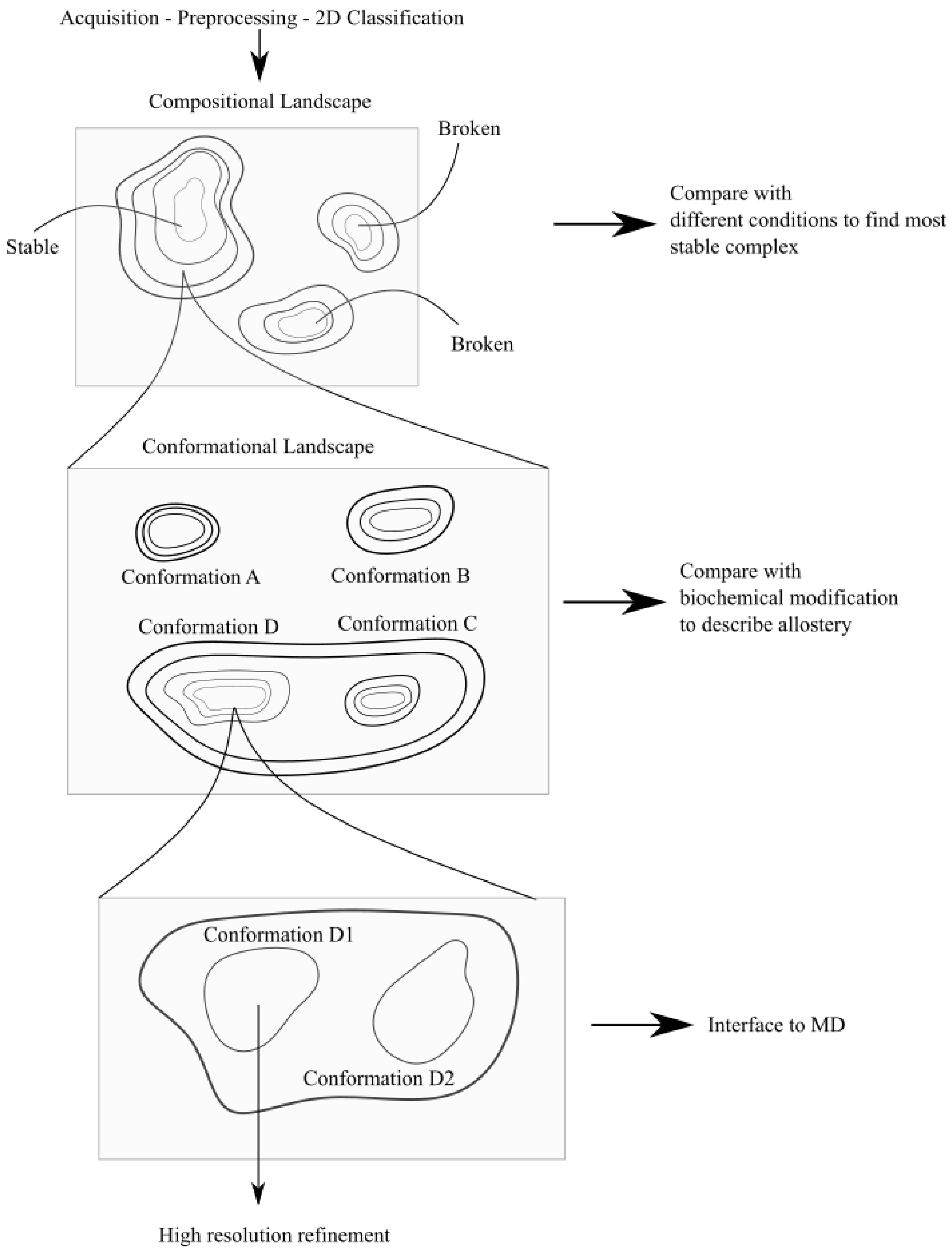

Acquisition - Preprocessing - 2D Classification
Compositional Landscape
Broken
Stable
Broken
Compare with
different conditions to find most
stable complex
Conformational Landscape
Conformation A
Conformation B
Conformation D
Conformation C
Compare with
biochemical modification
to describe allostery
Conformation D1
Conformation D2
Interface to MD
High resolution refinement

Fig. 5.1: Based on the analysis of full ensembles rather then single structures one can imagine a hierarchical refinement scheme, which resembles the tier-structure shown in figure 1.5. By firstly relaxing the compositional misture and focussing on the stable protein complex the overall energy landscape of the stable complex can be refined. Here comparing studies can be done with other biochemical modifications such as inhibitor or substrate binding. "Zooming" further in one of the minima, single gaussian minima should be accessible, which should refine to highest resolution. The variability within one of these minima could be analysed by MD simulations.

5.1.5 The Interplay with MD-Simulation

Let us summarise the major limitations worked out until now: due to the resolution (and hence due to limited statistics) the resolution of the individual volumes is too low to describe energetics in the higher-order tiers. One consequence of this is that for example the local changes upon inhibitor binding can hardly be resolved. Additionally we discussed in *Lambrecht (submitted)* that we are not able to acces the height of activation barriers as those are not sampled in sufficient numbers.
MD-simulations play an important role in retrieving information on the physical fundamentals of biomolecuar motions and are well suited to predict motions and epecially explain the energetics and kinetics underlying allosteric communication. However, as the underlying forces are calculated for each time point and each atom in the simulation box, the computational complexity makes long simulation times for macromolecular complexes such as the proteasome problematic. As the transition from one lower-order tier to another happens usually on the μs to ms timescale, they are usually not accesible for all-atom MD and compromises have to be found as for example coarse graining.

It is obvious that both methods complement each other [6]. The time- and lengthscales could be sampled by a hybrid method: for each minimum in the conformational landscape from EM, a rigid body model could be fitted into the best resolved structure. Then energy minimization is performed and from the energy minimized starting model a MD simulation can sample the higher-order tiers "on the bottom" of the low-order tier found by the CowScape algorithm. By applying an additional potential, generated from all points in the sampled minimum, one could potentially restrict the simulation to sample only the aimed locality of the landscape.

Applying an additional potential is also a way to complement another limitation of cryo-EM: the activation barriers (and intermediate but still infrequent sampled areas) in between the maxima. One potential method to solve this might be umbrella sampling. In umbrella sampling, free energy differences are calculated along a trajectory between two states by forcing the simulation to sample this trajectory by an additional potential. The trajectory could be chosen as being similar to the eigenvectors found by PCA.
These examples should make clear how potentially in the future cryo-EM and MD simulations could compensate each others weaknesses when it comes to the analysis of dynamics and energetics in macromolecular complexes.

5.1.6 Drug Development

Structure based drug discovery (SBDD) is today routinely applied in the pharmaceutical industry. It relies on the fact that the structure of a drug target is known and for example the binding pose of a inhibitor library can be predetermined *in silico*. This step narrows the search space significantly. However, with cryo-EM it is nowadays not routinely possible to achieve resolutions which can be used for predictions on the atomistic level. Renaud et al. [109] discusses another role of cryo-EM in the hit-to-lead characterization and lead optimization. Lead optimization describes the process of understanding the mechanism of action underlying a substance, which was found to be active in high-throughput screenings [117]. This involves also the analysis of the biological context and also on the underlying - conformational - mechanism. It was shown above that the CowScape algorithm is capable of describing the conformational changes induced by drug-binding in a comprehensive way and here it should be further explained how this feature could be utilized in the pharmaceutic development.

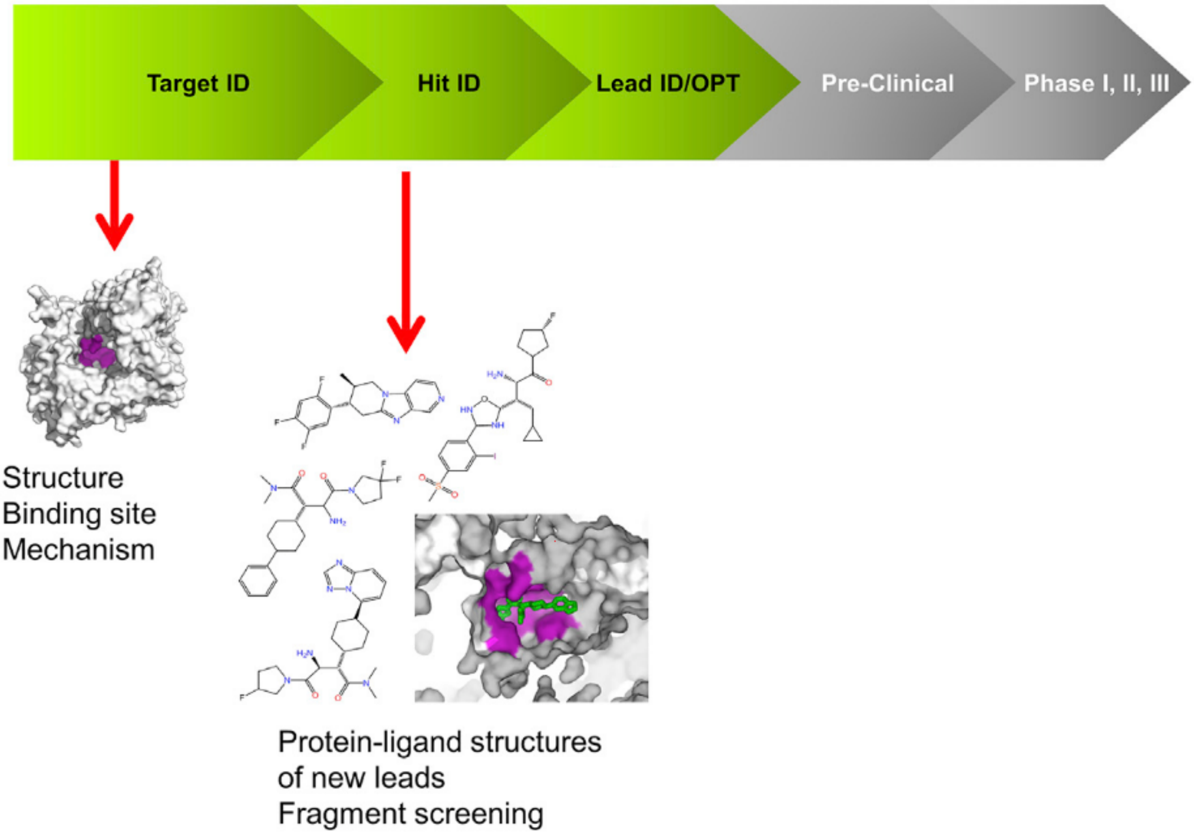


Fig. 5.2: SBDD in the early phases can be divided into three major phases. Where Target ID and Hit ID are primarily high-resolution guided process, Lead optimization can win from medium-resolution dynamical data. (ID=Identification, OPT=optimization), [117]

5.1.6.1 Mechanistic studies

In the literature several examples can be found in which drug-bound protein complexes where resolved and *Haselbach and Schrader (2017)* is just one example. Only a few of them show resolutions better then 3Å [109], and for some of them the validity of the map has to be challenged by objective means (missing water molecules, inplanarity of the aromatic systems). This sub-3-Å-resolution, which can routinely be achieved by crystallography once a sufficient crystallisation condition was found [108], allows to interprete side-chain information. In fact, this side chain information is necessary to draw conclusions about the underlying binding model. Additionally, to investigate the underlying binding model on an ensemble basis, it would be necessary to resolve the full ensemble on an atomistic scale. Such descriptions are as of today only possible by NMR on small proteins [7].

Routinely overcoming this resolution barrier by computational means (e.g. Ewald sphere correction or beam tilt correction [143, 159]) as well as by instrumentation and pipelining [117] might be possible within the next few years. Then there is an particular advantage over NMR as protein and inhibitor conformation can be visualized together [18]. However, for some applications this resolution might not be necessary at all. Especially for large macromolecular complexes the understanding of large scale remodelling to understand and detect the influence a drug has on the whole complex might be important [134, 144]. Cryo-EM has shown its power in detecting such motions. Apart from the examples given in *Haselbach and Schrader (2017)* and *Lambrecht (submitted)* interesting results where also generated for microtubule stabilizers [72], which have shown to be promising leads in oncology. Another quite early example was the stabilization of the γ-secretase by the dipeptidic inhibitor DAPT, used to suppress β-amyoloid plaque formation in primary neurons [3].

From discussions above, it is fair to assume that an active allosteric inhibitor will cause large-scale remodeling within the macromolecular target complex. This remodelling will always affect the energy landscape, even if the principal mean structure does not change [16]. It might therefore be a new paradigm to characterise a drug-protein complex not only by a mean structure (the "crystallographic" point of view) but rather by the change drug binding applies to the conformational landscape. One could for example imagine a workflow in which the apo-structure of the target complex was solved with large statistics and PCA is used to calculate the apo-energy landscape - probably also in a mixture with a set of known affectors or substrates. The result would be a "benchmarking" orthonormal-basis (ONB). Then, preparations with new lead candidates could be used to assess if there is a significant change in the landscape of the candidate (in the benchmark ONB) with respect to the apo-benchmark. As the coordinates of the spaces directly encode conformational motions, implications on the functional level could be drawn. Thereby inhibitors, which affect regions in the complex, which should not be touched (e.g. signalling hotspots in an inhibitor which should just slow-down the catalytic function), could directly be sorted out. Thereby, a more dynamics-based point of view (in contrast to the static structure picture) would be achieved. However as of today, this vision is limited by the computational complexity, which requires the analysis to run over a few iterations for more than a day.

5.1.7 Conclusion

In the publications *Haselbach and Schrader (2017)* and *Lambrecht (submitted)* the usability of a ensemble based analysis of cryo-EM data yields important information on the allosteric pathway the durg binding activated. Especially in *Lambrecht (2018)*, the concept was applied to different kinds of motions and different molecules. Also it was shown that it is possible to detect and analyse allosteric mechanism of protein regulation as well as having a tool to guide the biochemical optimization of macromolecular complexes.
In this chapter further improvements were suggested. By applying clustering algorithms and using the ensemble view, one can envision a new paradigm in the analyis of single-particle data. This is actually based on the tier-like description of protein motion introduced by *Frauenfelder* in the late 80's. It can only be assumed at this point that by applying a physically more realistic scheme, based on the actual hierarchy in the dynamics of proteins, the refinement performance will also increase. However, such a scheme enables harnessing the full information content with respect to the classical scheme introduced in chapter 1, where all "non-interpretable" results are discarded to a certain extend. Furthermore, the opportunities for drug development were shown, where in particularly the interpretation of conformational changes upon inhibitor binding is a task which can hardly be performed by any other method at the moment.

However, at the moment the computational task of all the refinement steps included is a burden, but it is a general interest in the field to speed up that process. Then, this method is perfectly suited to be integrated in an automated workflow, as it allows to directly assess and compare the quality of different samples and the influence of inhibitors and stabilising agents. All in all, a versatile and flexible method was presented which presents a cryo-EM experiment as what it is: the observation of the in-solution ensemble of a biomolecular sample.

5.2 Single Particle Cryo-EM with iDPC-STEM

In *Lambrecht and Riedel (in preparation)* we tried to optimize iDPC STEM for the imaging of biological macromolecules. We could show that on organic but non-biological specimen such as GCB we were able to image at conditions which should in theory allow to image cryo-EM samples up to a resolution of 3.4 Å. However, when we examined proteasomel and ribosomel samples in vitrified ice, the contrast immediatley dropped and the background contribution to the image increased drastically. When trying to compensate this effect by increasing the dose, radiation damage became apparent. Here, I will outline the contribution of the background from the point of view from imaging theory. In particular, the shape of the CTF can be compared to other imaging modes and allows to speculate how cryo-EM can be further contrast-optimized.

5.2.1 Contribution of the Background

Typically, biomolecules are rather weak scatterers which is due to the elemental composition. For each carbon atom there are in average 1.585 hydrogen atoms, 0.285 nitrogen atoms, 0.3 oxygen atoms, and 0.01 sulfur atoms in proteins. This fact, and the fact that the samples in cryo-EM are typically very thin, allows to treat the samples as weak phase object. This means that the phase shift $\phi(x, y) << 1$. For equation 1.1, the concept of the WPOA was introduced. The WPOA also implies that if the phase contrast is weak the surrounding medium can usually be neglected in the description of the background signal and only ineleastical collision events and non-scattered electrons contribute to background noise. Two different theoretical approaches describe background noise formation for phase contrast TEM under assumption of the WPOA: Rez [110] and Henderson [59]. Both basically come to the same conclusion, but the model of Rez [110] is directly comparable to ADF-STEM. According to this work, the intensity of the background in the case of phase contrast TEM is independent of the surrounding

$$I_B^{WPO} = Jd^2 \tag{5.1}$$

with J being the elctron flux and d being the pixel dimension. In contrast to this, for ADF-STEM, the WPOA does not hold, the embedding medium contributes to the background and

$$I_B^{PO} = N_w t \sigma_w J d^2 \tag{5.2}$$

with N_w being the number of water molecules and σ_w is the scattering cross section of water. Indeed, for all the iDPC relevant literature deriving the image formation theory numerically, the WPOA was never used[80, 82]. This theory is basically based on Bosch and Lazić [8], who in contrast to the WPOA derives another fomulation which in the end uses a comparable linear approach to describe the dependence of the exiting wave on the sample. However, none of the authors ever discussed the embedding medium as a source of contrast suppression. Indeed, new primary results from numerical simulations based on the above cited theory show that there is a rather strong dependence on the embedding medium, which can not be fully explained by the theoretical description until now (E. Bosch, oral communication). Nevertheless, this seams to imply that the background contributes in a way that is theoretically not understood and needs to be investigated in depths using a better framework which accounts for the embedding medium.

Interestingly, Lazić and Bosch [80], assumed that the integration of the differential phase contrast signal should even improve the contrast in the images by the following mechanism: as outlined in the introduction, the DPC image is a representation of the projected electric volume field having two dimensions. Integrating this image yields a representation of the electric potential. However, noise should not be integratable as it is non-conservative field. Hence, the integration should yield a better signal-to-noise ratio. However, one reason for iDPC producing that much background in contrast to CTEM might be, that it is sensitive to buffer ions and other charges in the sovent (or even the solvent dipole itself). This would again yield a conservative vector field which can be integrated and information is retained, leading to a smaller SNR in the images.

Another discussion about the theoretical description of the DPC image was published by Majert and Kohl [88]. They in contrast used the WPOA to describe DPC imaging and found that the inner cutoff angle (defined by the cetral hole in figure 1.11) of the detection is limiting the resolution and is actually damping the high- and medium-resolution area. This dependence might be also explainable with Lazić and Bosch [80], where they found a statistical limit to resolve the high-resolution regime. However, we also found that while the low-resolution signal is overwhelming, the high-resolution data is rather suppressed. The finding of Majert and Kohl [88] implies that the physical hole size of the detector has to be changed to improve the damping of the high-resolution regime. This seems to be not necessary for non-ice embedded specimen [156]. As the contrast transfer is already low for the ice-embedded specimen, this might be the only optical feature which can be improved to enhance contrast in the limit of the radiation damage. From the results presented above, it is hard to imagine that without changing the optical parameters of the setup, any improvement of the imaging conditions for biomolecules can be achieved.

5.2.2 The contribution of the contrast transfer function (CTF) and phase plates

The CTF adds another layer of complexity on to the problem of imaging biomolecules. Whereas the previous paragraph focussed on how the contrast is generated by scattering cross-section differences, here the focus is on how this contrast is transferred to the image. It was already discussed in the beginning of this thesis that the fourier transform of the PSF is the CTF, which basically determines how the produced contrast in the image is visible in the image.

Elbaum et al. [36] discusses the influences of the different CTF's of TEM and STEM.

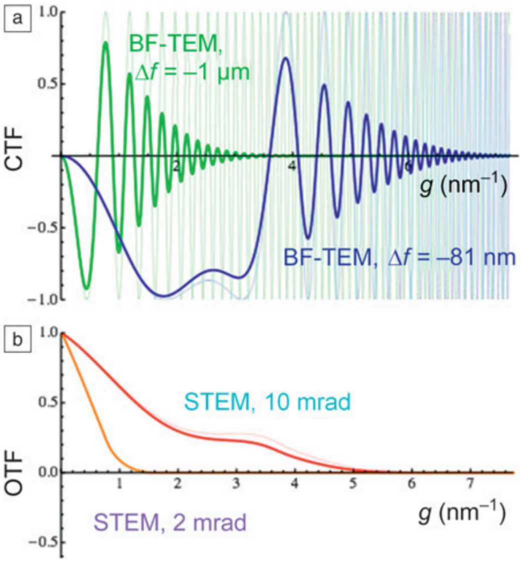
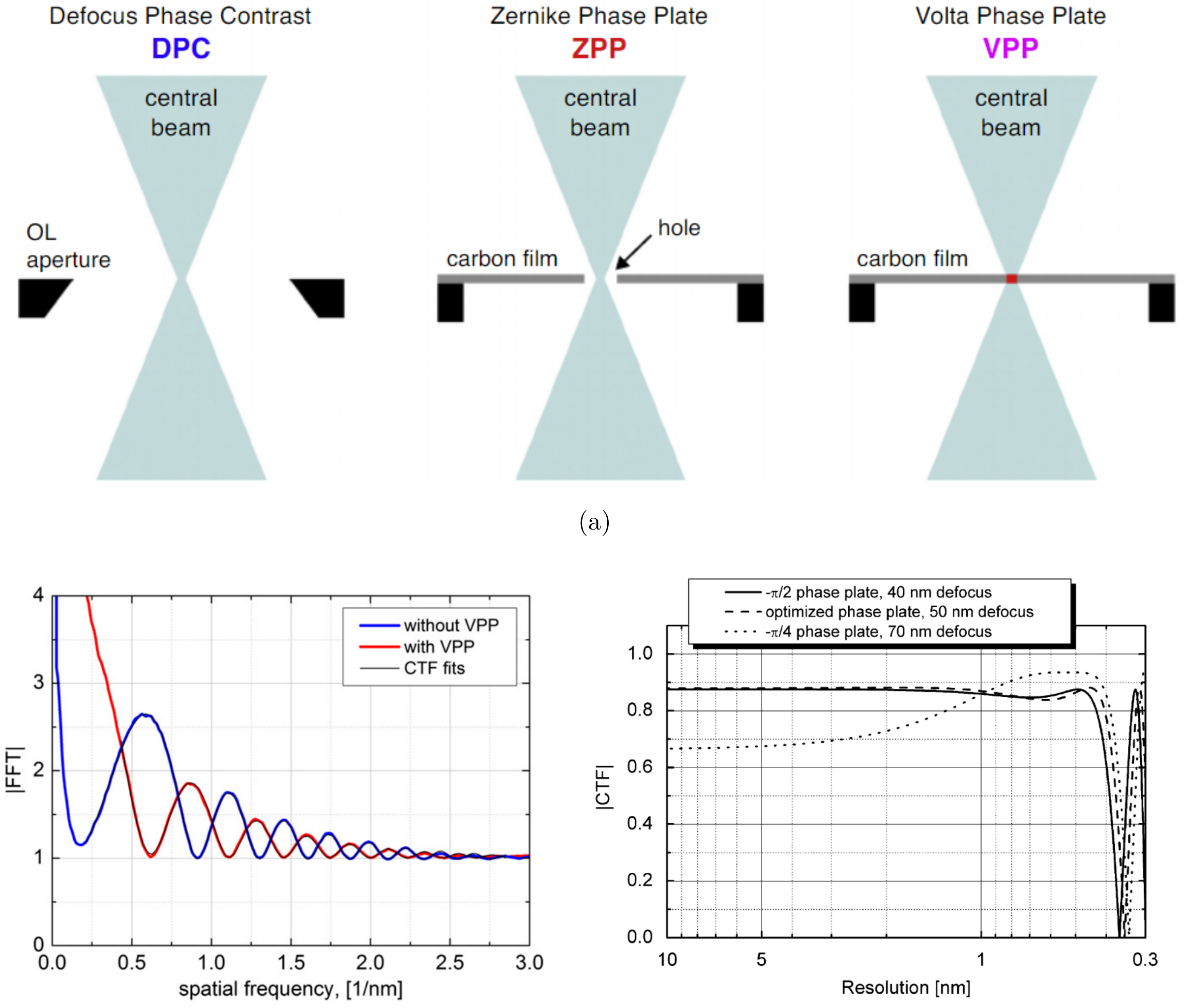


Fig. 5.3: Transmission electron microscope (TEM, a) phase contrast and scanning transmission electron microscope (STEM, b) transfer as functions of the spatial frequency, g for different defoci. (a) Bright-field TEM (BF-TEM) contrast transfer functions (CTFs) (blue and green) for a thin-phase object and (b) incoherent STEM optical transfer function (OTF, red and orange) for a thin-amplitude object for the same spherical aberration, $C_s = 2.0$ mm, at 200 kV. [36]

As can be seen in figure 5.3, the sinuosidal nature of the CTF in TEM supresses the low-resolution features of the image. Thereby it acts as a "optical high-pass filter" [33]. This is in contrast to STEM, having an evenly decaying contrast transfer, which peaks at the very center of the detection beam, namely at the lowest resolution. This implies that the above discussed damping of high-resolution features and the overwhelming low-resolution are optically instrinsic features of each STEM setup. This is in accordance with the observation of Elad et al. [33]. When they imaged single particles by HAADF-STEM they also reported about extremely high low-resolution noise in the images. This underpins the statement already done in the last paragraph about the necessary changes to the optics, when STEM on single particle cryo-samples shall be successful.

In the field of TEM there are for years discussions on how to improve the contrast, as this would enable imaging of smaller molecules. The most discussed opportunity is the usage of phase plates similar to photonic microscopy. Two concepts emerged to be succesfull, even though the practical usage is not fully proven yet, due to long-term stability[22]: the Zernike phase plate (ZPP) and the Volta phase plate (VPP). Both are placed in the back focal plane of the microscope (quasi in fourier space) and modulate the spectrum accordingly. The ZPP was the first sucessfully applied phase plate in single-particle analysis and tomography [19–21]. It is based on a thin layer of amorphous carbon and its thickness is adjusted according to acceleration voltage such that a $\frac{\pi}{4}$ phaseshift is achieved. A central hole is needed to align the microscope properly[21]. In contrast to this, the VPP is a continuos and heated carbon layer, and the phase shift is potentially generated by inducing a volta potential on the surface of the carbon layer [22]. These two have a difference in



(a)

(b)

(c)

Fig. 5.4: a) Overview on the different options to introduce phase contrast: defocusing, VPP, ZPP. Note the central hole of ZPPs that introduce a cutoff in the low-resolution regime of the image, taken from [19] b) Experimental CTF of a standard TEM and a TEM equipped with a VPP, taken from [22]. c) Theoretical CTF of a ZPP for different phase shifts, taken from [20]

the first zero crossing (figure 5.4). Both plates shift the first zero-crossing much further to the positiv direction as compared to the CTF without any additional phase contrast. The CTF of the ZPP is theoretical. It lacks one important feature, namely the cutoff in the low-resolution regime. The CTF of the VPP instead is a measured quantity, therefore not showing that it actually starts at (0,0). This is in contrast to the ZPP and the STEM CTF, which both have the maximum contrast transfer in the lowest frequency. From the findings above about the problems of low-resolution noise in the STEM case, one might draw the conclusion that the extreme low-resolution signal enhancement of the ZPP might lead to unwanted background contributions, if this is not compensated for by the central hole.

Another complication was not discussed before which might be a general concern also for phase plates. For the STEM measurements, the focus had to be adjusted especially for high convergence angles. In the literature, the focus sensitivity of phase plates is rarely discussed. In the STEM images presented here, some pictures where clearly out of focus or were pre-exposed due to the manual focussing procedures. For both, STEM and PP-TEM a robust and pre-exposure free focussing algorithm has to be implemented in the acquisition software as manual focussing does not allow to come up with the necessary throughput to routinely acquire images with both techniques.

5.2.3 Orthogonal Information from STEM

Until now only the pure iDPC signal was discussed. However as it turns out that this will not yield promising insights into biomolecules, maybe orthogonal techniques can be used. As already explained above the actually measured quantity is not the iDPC signal but rather the DPC - the differential phase contrast in x- and y-direction. Lazić and Bosch [80] showed that the DPC is directly proportional to the electric field of a charge and indeed in material sciences DPC is mostly used to assess the fields in magnetic samples [96, 132]. Just recently Shibata et al. [133] imaged the local electric field of a single atom using the DPC signal of a four-quadrant detector in a crystalline lattice. In the same manner, one could try to gain in-depth insight of the coordination sphere of protein-bound metals, with a significant electric field vector to the stabilising charged side chains. A comparable study was done before with ferritin imaged by HAADF [33]. They could resolve the overall location of the metal centers but again due to the large background signal and the insufficient resolution, no further mechanistic conclusions could be drawn. However, Shibata et al. [133] showed that the resolution and localization of the charges is far better in DPC as in ADF. Additionally, we could show that at least low-resolution protein features are detectable in iDPC. Therefore, a study on a metal-bound protein might reveal another usage case for quadrant detectors in biology.

Additionally, the four-quadrant detector can be used in parallel with the HAADF-detector. Therefore, an interesting experiment might be the parallel visualization of a protein (or again protein-metal complex) by both detectors. This might at least increase the number of electrons gained from one experiment and an algebraic combination of different STEM signals has be done before [152].

5.2.4 STEM-tomography

In the last couple of years, a variety of methods arose which made electron tomography a powerful tool to dissect cellular structures. Those are extensively reviewed e.g. in Hanssen [56]. Primarily two departments at the Weizmann Institute focussed on STEM tomography [36, 152]. This work focussed on ADF-STEM, even thouh also combinations of DF and BF signals were used [152]. The main advantage is that STEM has as discussed above a higher focal depths then TEM and hence it was shown that a bacterial cell could be measured with high z-detail in one tomogram [34, 36]. Additionally, Wolf et al. [152] showed that the charging behavior in STEM tomography is less harmful then in TEM, as bubbling occured at a much higher cumulative dose. Another advantage is that as STEM contrast is quantitative, the inclusion of for example phosphorus or calcium can be visualized and approximately quantified [153, 154].
Hoewever, even though STEM tomography fits perfectly within the "thickness-gap" between soft X-ray tomography and focussed ion beam surface electron microscopy (FIB-SEM), it has still some limitations to reach through even higher thicknesses. Nearly all electrons that have reached the ADF detector had to be scattered more than once, and the probability scales with the thickness of the specimen. Therefore a volume of more than two times the mean free path length for the aquaeous medium seems feasible without a resolution penalty [112]. This effect is not as severe for electrons scattered on the brightfield detector, as those will have experienced less scattering events. iDPC here offers to get the best from two worlds: the bright-field phase contrast used to from TEM and the large depth of focus from a STEM[35]. In the section before, several limitations were worked out, which should not matter for cellular tomography anymore. Firstly, we discussed the large background noise from the embedding medium. For cellular tomography the density of protein and nucleic acids is higher than for a single particle experiment [112] and furthermore the results with ADF STEM tomography show that highpass-filtering out the low-resolution background does still yield robust tomographic slices [152]. Secondly, the low contrast for proteins in the vitrified medium was discussed in *Lambrecht and Riedel (in preparation)*. Again in tomography, this should not play a major role, as the contrast directly scales with the size of the molecule or the feature under investigation [110, 111]. As these features are much larger in cellular tomography, the contrast should also be much better then it was reported for the single-particle case here.

5.2.5 Conclusion

We have shown in *Lambrecht and Riedel (submitted)* that a direct visulization of high-resolved single particle images will not be feasible with iDPC STEM due to the high background and the low contrast - at least not for radiation sensitive biomolecules. It was made clear here that the large contribution of background is not in agreement with a WPO and that the current theoretical framework for iDPC STEM does not cover the embedding medium in a sufficient extend. It can only be speculated that by physically tuning the detector, the contrast for biomolecules can be improved. However, further usage cases especially for bioinorganic complexes and in STEM tomography were worked out and in particular the usage in tomography is very promising, due to the detection of the dignal in the brightfield disc.

6 Outlook

In the first part of this thesis and the corresponding publications we have shown that cryo-EM is capable to assess the conformational landscape of a macromolecular complex. Analysing cryo-EM data from the ensemble point of view rather than from a pure-structural point of view, opens up a plethora of possible usage cases, where the quality assessment and pharmaceutical charactization of a drug-binding event in terms of allostery are just two in particularly interesting ones.

In the second part we have shown that improving the contrast with the help of iDPC STEM is not straight-forward. We have learned there that even if a method works sucessfully on a "simple" specimen such as we have shown it with graphitized carbon black, radiation-sensitive, vitrified cryo-EM samples are introducing even new and challenging condition for any setup.

Even if both topics seem to be very unrelated, the motivation behind them interdepend. As was introduced right in the beginning, distangling the heterogeneity in cryo-EM samples is a problem of indentifiability. Or in other words: in the case of infinite contrast, one could see everything. Advances in the field of conformational dynamics might be triggered through the development of a more sophisticated optic, increasing the contrast of the images. On the other hand, such optics might need better behaved specimen, for example with respect to ice thickness or sample impurities. At this point, the quality assessment use case described above might be helpful.

In the end it is obvious that software can not improve endlessly without optics and limitation in optics might be compensated by software. The current focus of the cryo-EM community seems to be primarily on the software. But in the end, an investment in optics might potentiate the power of the algorithmic improvements. This thesis tackles both sites more or less successful and it might be appreciable to represent both - optics and image processing - in the focus of the future research.

A Abbreviations

AAA+ ATPases ATPases associated with diverse cellular activities
ATP Adenosinetriphosphate
BFP Back focal plane
CA Correspondence analysis
CBED Converged beam electron diffraction
CCD Charged coupled device
C_s Spherical abberation
CTF Contrast transfer function
DED Direct electron detector
DPC Differential phase contrast
EELS Electron energy loss spectroscopy
EM Electron microscope
FEG Field emission gun
FRET Fluorescence resonance energy transfer
GMM Gaussian mixture model
HAADF High-angle annular darkfield
iDPC Integrated differential phase contrast
KDE Kernel density estimation
KNF Koshland, Némethy and Filmer model of allostery
MDS Molecular dynamics simulation
MSA Multivariate statistical analysis
MWC Monod-Wyman-Changeux model
NMA Normal modes analysis
NMDA N-methyl-D-aspartate
NMR Nuclear magnetic resonance
ONB Orthonormal basis
PCA Principal component analysis
PSF Point spread function
SBDD Structure based drug design
SNR Signal to noise ratio
SPA Single particle analysis
STEM Scanning transmission electron microscope
TEM Transmission electron microscope
TRPM Transient receptor potential ion channels
VPP Volta phase plate
ZPP Zernike phase plate

B Supplementary information

B.1 Supplementary Information to *Lambrecht (submitted)*

Supplementary Methods

Data analysis.

All raw data were downloaded from the EMPIAR Database (spliceosome, 10160; proteasome, 10166; 19S particle with p28, 10091; and the TRPM4 channel, 10126 and 10127) and subjected to several rounds of 2D classification in the COW image processing suite to remove non-particle views.

As described in the main text, the general workflow covers three steps of image *processing:* I) generate the conformational ensemble by applying classification algorithms; II) carry out PCA; and III) construct the energy landscape.

Ensemble generation and alignment.

An ensemble was defined as a set of 3D volumes calculated from one given dataset in which the particle images may be representatives of more than just one structure. Our approach can be applied to any set of 3D volumes that share the same image frame size. Here, we describe an application to 3D volumes generated by extensive Maximum-Likelihood 3D Classification, which was tested in RELION, SPHIRE, and the COW suite. For all examples described in the manuscript, classification runs were calculated using 200 000 particles separated into 40 classes (note that, if the initial dataset was smaller, fewer classes were used to maintain the same ratio between number of particles and number of classes). For this separation into classes, the entire population of particles must first be split randomly into multiple subsets, which are then classified separately, to reduce the computational effort (which increases dramatically with the number of particles). After classification, individual volumes were autorefined, and the set of all 3D volumes was subjected to PCA to: *i*) identify the major sources of variability among all 3D volumes; and *ii*) order them with respect to those major variabilities. Because PCA is sensitive to any differences between the volumes, aligning them against the same reference is crucial. This alignment was carried out either in the COW suite using the Logic "3DAlignment" with standard parameters, or in UCSF Chimera[1] with the FitinMap procedure. 3DAlignment maximizes the real-space cross-correlation to a reference by applying shifts and rotations to the volume. Special measures applied to some of the datasets are listed below:

Dataset	Remarks to classification and processing
26S/RPN1	No alignment was necessary, as the classification was only local, and the 20S part remained stable throughout the classification runs
TRPM4	The classification and subsequent refinement was—in contrast to the original publication—performed without applying symmetry restraints

All subsequent steps were performed within the COW image processing suite. The volumes were low-pass filtered to the lowest resolution among all ensemble members. Again, this is necessary to avoid the detection of resolution differences via PCA. The Fourier space cutoff of the filter is given by

$$\gamma = \frac{2 \times \text{pixel Size}}{\text{lowest Resolution}}$$

Subsequently, all volumes were normalized to a mean of zero and a σ of 2. The σ is an important variable to fine-tune the information content of the eigenvolumes, as it directly influences the detection of information by PCA in the next step.

Principal Component Analysis (PCA).

PCA finds a representation of the input data in a lower dimensional space that preserves the correlation structure of data as much as possible (for a thorough review of PCA for single-particle analysis, see van Heel et al[1]. The coordinate axes of the lower dimensional space are obtained by calculating the eigenvectors of the data covariance matrix. The eigenvectors are orthogonal to each other and sorted by their eigenvalue (which correlates with the fraction of the total variance described by this eigenvector). CowEyes' implementation of PCA is memory efficient especially for large volumes, as it uses the $\mathrm{N} \times \mathrm{N}$ dimensional covariance matrix, where N is the number of volumes. Selecting a subset of the eigenvectors means that these span a new lower dimensional space, which represents the input. The position of the volume in this lower-dimensional space is determined by a vector of linear factors a, where a_k (belonging to the k-th volume) has as many entries as the new space has dimensions. COW calculates the linear factors by a matrix multiplication performed in the logic "PCATransform".

Visual inspection of eigenvectors.

To decide which conformational transitions are captured by the eigenvectors, one needs to visually inspect them. Therefore, trajectories are simulated from one eigenvector. The algorithm for generation of trajectories is implemented as "TrajectoryVisualizer" in CowEyes. It shows the information represented by a single user–specified eigenvolume by simulating a trajectory of N volumes. From the linear factors calculated in "PCATransform", one could approximate each volume V_k in the dataset through the relation

$$V_k = \sum_{i=0}^{n} a_{k,i} \times E_i + A \qquad (1)$$

where A is the average volume calculated in the PCA, V_k is the k-th volume in the dataset, E_i is the i-th eigenvolume describing a conformational motion, and $a_{k,i}$ is the linear factor that acts as a coordinate for V_k in a space spanned by E_i.

The TrajectoryVisualizer module in the COW software suite simulates a volume that only displays the information encoded by one eigenvolume E_i . Hence (1) simplifies to:

$$V_{sim} = a_i \times E_i + A \qquad (2)$$

To show the N snapshots along the trajectory, the linear factors of all volumes with respect to the specified eigenvector E_i are collected and then sorted into one vector $\mathbf{a}_\mathrm{i} = \{\mathrm{a}_{\mathrm{i},0}, \dots , \mathrm{a}_{\mathrm{i},\mathrm{n}}\}$, where $\mathrm{a}_{\mathrm{i},0}$ and $\mathrm{a}_{\mathrm{i},\mathrm{n}}$ are the lowest and highest linear factors, respectively. The average volume A is constant throughout the dataset. TrajectoryVisualizer now offers two modes: either the real linear factors are used for the calculation (i.e. all elements of a_i), or an equally sampled interval between $\mathrm{a}_{\mathrm{i},0}$ and $\mathrm{a}_{\mathrm{i},\mathrm{n}}$ is used. In the latter case, a sampling with a distance of $\frac{a_{I,n} - a_{I,0}}{N}$ is visualized. By applying (2) for each element in the vector $\mathbf{a}_i$, or for each sample in the interval $[a_{i,0}, a_{i,n}]$, a volume series is calculated.

The final snapshot volumes of the trajectory can directly be visualized by the 3D viewer implemented in CowEyes. Supplementary Videos were generated by loading the full trajectory into UCSF Chimera and recording a movie of a morph between those snapshots.

Iterative refinement of classification.

As the dataset was initially split in random subsets, the first 3D classification results in a very rough estimation of the total conformational landscape, and lower populated states will be clearly missed. Thus, particles are sorted by the linear factor towards the first eigenvector and split in subsets. These sorted particle subsets are again subjected (individually) to 3D classification. The resulting classes are joined, and 3D PCA analysis is repeated as above. The entire procedure is repeated several times until populations are stable. In most cases, three iterations are sufficient.

Energy landscape estimation.

The previous analysis steps mainly serve to find an efficient representation of the input 3D volumes in a low-dimensional space. What is still missing is a quantitative continuous description of this space that would also allow the estimation of the thermodynamic properties of the system. The algorithm for the estimation of a continuous surface from the linear factor representation is also implemented in CowEyes as EnergyLandscape. From two chosen eigenvolumes E_1 and E_2, the linear factors are saved in two vectors $\mathbf{a}_{k=1,I} = \{a_{0,1}, \dots, a_1\}$ and $\mathbf{a}_{k=2,I} = \{a_{0,2}, \dots, a_{n,2}\}$, respectively. Additionally, a third vector $f = \{f_0, \dots, f_n\}$ containing the observation frequencies or particle counts of the corresponding volume is saved. Each data point of the form $(a_{1,k} \mid a_{2,k} \mid f_k)$ is then placed on a grid forming a weighted point cloud. To estimate a continuous surface, kernel density estimation (KDE) with a Gaussian kernel function

$$f(x,y) = A\, e^{-\left(\frac{(x-x_0)^2}{2\sigma^2} + \frac{(y-y_0)^2}{2\sigma^2}\right)}$$

is used, whereby x and y are the coordinates on the sampling grid, and x_0 and y_0 are the coordinates of the points specified by $a_{1,k}$ and $a_{2,k}$, respectively. Hence, the kernel function is evaluated for each combination of grid point (x,y) and data point (x_0, y_0). The sum of all Gaussians fitted to the data points is then the continuous probability surface. σ determines the width of the Gaussian and has to be user-adjusted. With the standard grid spacing, all landscapes shown are plotted with a σ of 300 px. A determines the height of the Gaussian and can either be the observation frequency f_k or the corresponding temperature-normalized Gibb's free energy

$$\left(\frac{\Delta\Delta G}{k_B T}\right)_k = \ln \frac{f_k}{\max(f)}$$

A final 3D-rendered visualization of the landscape can be obtained by CowEyes' built in surface viewer, which is based on the Qt Data Visualization library.

Supplemental discussion.

An outline of the algorithmic workflow to analyze an ensemble of cryo-EM volumes with respect to their intrinsic modes of motion is provided herein. These motions are then used to describe the functional region of the energy landscape that the macromolecular machine samples in solution.

A competitive maximum likelihood refinement ('3D classification'), whereby one particle is assigned to a randomly initialized reference, generates all the volumetric ensembles used in this work. Nevertheless, generation of this ensemble is extremely flexible and not restricted to 3D. Rather, an ensemble of 2D images, all of which represent the same view, could also be used; this would not require either a large datasets or the 3D classification step (which is computationally quite expensive). A prerequisite for an analysis in 3D space is that a certain subset of 2D images have been averaged during 3D reconstruction. This strategy has a major disadvantage that can however be offset through its advantage. The disadvantage is that it discretizes the continuous movement of macromolecular machines into only those volumes that form the ensemble. However, averaging 2D images suppresses noise and hence increases the robustness and information content of the statistical analysis by PCA. The last point is a clear advantage over statistical methods based purely on 2D images.

PCA has been previously used to detect conformational motions in cryo-EM data on both 2D [2] and 3D data [3-5]. However, PCA only finds a linear approximation to the motions in the dataset. This also implies that the real motions of the molecule are always a linear combination of all eigenvolumes found for the dataset. Non-linear approaches are in contrast prone to overfitting, and their validity has to be verified thoroughly in the future. Further, CowScape uses the eigenvolumes primarily for dimensionality reduction of the high-dimensional volumes. For this purpose, PCA offers a robust way of finding dimensions representing conformational motions), and being able to apply it to volumes rather than to coarse models further increases the information content.

After embedding the volumes into a low-dimensional space, CowScape quantitatively interprets the experiment by deriving Gibb's free energies. These free energies are calculated by the Boltzmann coefficient from the number of particles per 3D class. Counting information resulting from 3D classification algorithms has been presented in the literature only rarely. The estimated energies are prone to errors for the following reasons: i) generation of 3D volumes needs a minimum of a few thousand particles that should represent different projection directions. On the one hand, this makes it hard to detect minimally-populated states; on the other hand, in certain projection directions, the conformational change might hardly be detectable. In both cases, the final landscape will be smoother, and the energies will be underestimated; ii) even with the most careful sorting, broken particles will always be contained in the dataset and thus will be randomly assigned to the volumes. This will lead to an underestimation of the energies and to a smoothening of the landscape; iii) higher resolved structures tend to attract more particles even if they do not fit in with respect to their conformation. This must be addressed by constant filtering of obtained 3D volumes; iv) refinement after 3D classification discretizes the particle assignment to a single conformation, which might due to the noise in the individual image not be fully justified. This argument is addressed by estimating energy surfaces using Gaussian kernel density estimation (KDE). KDE will *a posteriori* soften the assignments in the 3D classification and increase the robustness of the energies. A Gaussian kernel has two parameters: its mean μ, and its standard deviation or bandwidth σ. Since the means are fixed by the position in 2D space, the bandwidth is critical for the absolute height of the energies and still has to be adjusted. In CowScape, this is left to the expertise of the user. Future improvements of the model have to include a stable estimation procedure of the bandwidth parameter.

The computationally most complex step is the generation of the ensemble prior to the actual analysis with CowScape, which in terms of computation time is negligible compared to the first step. Hence, the analysis and testing of different parameters can be done directly after the ensemble has been generated once.

References

1. Heel, M.v., Portugal, R.V. & Schatz, M. Multivariate Statistical Analysis of Large Datasets: Single Particle Electron Microscopy. *Open Journal of Statistics* **06**, 701-739 (2016).
2. White, H.E., Saibil, H.R., Ignatiou, A. & Orlova, E.V. Recognition and separation of single particles with size variation by statistical analysis of their images. *J Mol Biol* **336**, 453-460 (2004).
3. Fischer, N., Konevega, A.L., Wintermeyer, W., Rodnina, M.V. & Stark, H. Ribosome dynamics and tRNA movement by time-resolved electron cryomicroscopy. *Nature* **466**, 329-333 (2010).
4. Klaholz, B.P., Myasnikov, A.G. & Van Heel, M. Visualization of release factor 3 on the ribosome during termination of protein synthesis. *Nature* **427**, 862-865 (2004).
5. Penczek, P.A., Frank, J. & Spahn, C.M. 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**, 184-194 (2006).

C References

For PDF version only: by clicking on the DOI link, you will automatically get forwarded to an online version of the corresponding reference.

[1] A Amadei. Essential dynamics of protien. *... : Structure, Function, and ...*, 1993. URL `http://onlinelibrary.wiley.com/doi/10.1002/prot.340170408/full`.

[2] A. Amunts, A. Brown, X.-c. Bai, J. L. Llacer, 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. ISSN 0036-8075. doi: 10.1126/science.1249410. URL `http://www.ncbi.nlm.nih.gov/pubmed/24675956http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4046073http://www.sciencemag.org/cgi/doi/10.1126/science.1249410`.

[3] Xiao Chen Bai, Eeson Rajendra, Guanghui Yang, Yigong Shi, and Sjors HW Scheres. Sampling the conformational space of the catalytic subunit of human g-secretase. *eLife*, 4(December2015), 2015. ISSN 2050084X. doi: 10.7554/eLife.11182. URL `http://www.alzforum.org/mutations`.

[4] Soojay Banerjee, Alberto Bartesaghi, Alan Merk, Prashant Rao, Stacie L. Bulfer, Yongzhao Yan, Neal Green, Barbara Mroczkowski, R. Jeffrey Neitz, Peter Wipf, Veronica Falconieri, Raymond J. Deshaies, Jacqueline L.S. Milne, Donna Huryn, Michelle Arkin, and Sriram Subramaniam. 2.3 Å resolution cryo-EM structure of human p97 and mechanism of allosteric inhibition. *Science*, 351(6275):871–875, 2016. ISSN 10959203. doi: 10.1126/science.aad7974.

[5] 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.annualreviews.org/doi/abs/10.1146/annurev.bb.18.060189.002243`.

[6] Lars V Bock, Christian Blau, Gunnar F Schröder, Iakov I Davydov, Niels Fischer, Holger Stark, Marina V Rodnina, Andrea C Vaiana, and Helmut Grubmüller. Energy barriers and driving forces in tRNA translocation through the ribosome. *Nature Structural and Molecular Biology*, 20(12):1390–1396, dec 2013. ISSN 15459993. doi: 10.1038/nsmb.2690. URL `http://www.nature.com/articles/nsmb.2690`.

[7] David D Boehr, Dan McElheny, H Jane Dyson, and Peter E Wright. The dynamic energy landscape of dihydrofolate reductase catalysis. *Science (New York, N.Y.)*, 313(5793):1638–42, sep 2006. ISSN 1095-9203. doi: 10.1126/science.1130258. URL `http://www.ncbi.nlm.nih.gov/pubmed/16973882`.

[8] Eric G T Bosch and Ivan Lazić. Analysis of HR-STEM theory for thin specimen. *Ultramicroscopy*, 156:59–72, sep 2015. ISSN 1879-2723. doi: 10.1016/j.ultramic.2015.02.004. URL `http://www.sciencedirect.com/science/article/pii/S0304399115000194`.

[9] J D Bryngelson and P G Wolynes. Spin glasses and the statistical mechanics of protein folding. *Proceedings of the National Academy of Sciences of the United States of America*, 84(21):7524–8, nov 1987. ISSN 0027-8424. doi: 10.1073/

PNAS.84.21.7524. URL `http://www.ncbi.nlm.nih.gov/pubmed/3478708http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC299331`.

[10] James P Buban, Quentin Ramasse, Bryant Gipson, Nigel D Browning, and Henning Stahlberg. High-resolution low-dose scanning transmission electron microscopy. *Journal of electron microscopy*, 59(2):103–12, 2010. ISSN 1477-9986. doi: 10.1093/jmicro/dfp052. URL `http://www.ncbi.nlm.nih.gov/pubmed/19915208http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2857930`.

[11] Ewen Callaway. The revolution will not be crystallized: a new method sweeps through structural biology. *Nature*, 525(7568):172–174, sep 2015. ISSN 0028-0836. doi: 10.1038/525172a. URL `http://www.nature.com/doifinder/10.1038/525172a`.

[12] Anna Carlsson, Ioannis Alexandrou, Emrah Yücelen, Eric G T Bosch, and Ivan Lazić. Low Dose Imaging Using Simultaneous iDPC- and ADF-STEM for Beam Sensitive Crystalline Structures. *Microsc. Microanal*, 24(1), 2018. doi: 10.1017/S1431927618001101. URL `https://doi.org/10.1017/S1431927618001101`.

[13] Marta Carroni and Helen R. Saibil. Cryo electron microscopy to determine the structure of macromolecular complexes. *Methods*, 95:78–85, feb 2016. ISSN 10462023. doi: 10.1016/j.ymeth.2015.11.023. URL `http://www.ncbi.nlm.nih.gov/pubmed/26638773http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5405050https://linkinghub.elsevier.com/retrieve/pii/S1046202315301638`.

[14] Ashwin Chari, David Haselbach, Jan-Martin Kirves, Juergen Ohmer, Elham Paknia, Niels Fischer, Oleg Ganichkin, Vanessa Möller, Jeremiah J Frye, Georg Petzold, Marc Jarvis, Michael Tietzel, Clemens Grimm, Jan-Michael Peters, Brenda A Schulman, Kai Tittmann, Jürgen Markl, Utz Fischer, and Holger Stark. ProteoPlex: stability optimization of macromolecular complexes by sparse-matrix screening of chemical space. *Nature Methods*, 12(9):859–865, sep 2015. ISSN 1548-7091. doi: 10.1038/nmeth.3493. URL `http://www.nature.com/articles/nmeth.3493`.

[15] Yifan Cheng. Single-particle cryo-EM—How did it get here and where will it go. *Science*, 361(6405):876–880, aug 2018. ISSN 0036-8075. doi: 10.1126/science.aat4346. URL `http://www.ncbi.nlm.nih.gov/pubmed/30166484http://www.sciencemag.org/lookup/doi/10.1126/science.aat4346`.

[16] A. Cooper and D. T.F. Dryden. Allostery without conformational change - A plausible model. *European Biophysics Journal*, 11(2):103–109, oct 1984. ISSN 01757571. doi: 10.1007/BF00276625. URL `http://link.springer.com/10.1007/BF00276625`.

[17] Peter Csermely, Robin Palotai, and Ruth Nussinov. Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. *Trends in Biochemical Sciences*, 35(10):539–546, oct 2010. ISSN 0968-0004. doi: 10.1016/J.TIBS.2010.04.009. URL `https://www.sciencedirect.com/science/article/pii/S0968000410000885`.

[18] Paula C.A. da Fonseca and Edward P. Morris. Cryo-EM reveals the conformation of a substrate analogue in the human 20S proteasome core. *Nature Communications*, 6(1):7573, dec 2015. ISSN 2041-1723. doi: 10.1038/ncomms8573. URL `http://www.nature.com/articles/ncomms8573`.

[19] Radostin Danev and Wolfgang Baumeister. Expanding the boundaries of cryo-EM with phase plates. *Current Opinion in Structural Biology*, 46:87–94, oct 2017. ISSN 0959440X. doi: 10.1016/j.sbi.2017.06.006. URL `https://linkinghub.elsevier.com/retrieve/pii/S0959440X1730043X`.

[20] Radostin Danev, Robert M Glaeser, and Kuniaki Nagayama. Practical factors affecting the performance of a thin-film phase plate for transmission electron microscopy. *Ultramicroscopy*, 109(4):312–325, 2009. ISSN 03043991. doi: 10.1016/j.ultramic.2008.12.006. URL `www.elsevier.com/locate/ultramic`.

[21] Radostin Danev, Shuji Kanamaru, Michael Marko, and Kuniaki Nagayama. Zernike phase contrast cryo-electron tomography. *Journal of Structural Biology*, 171(2):174–181, 2010. ISSN 10478477. doi: 10.1016/j.jsb.2010.03.013. URL `https://ac.els-cdn.com/S1047847710000900/1-s2.0-S1047847710000900-main.pdf?{_}tid=7c7d7815-6dae-4faa-9c65-686424ceaa5e{&}acdnat=1543160604{_}0487cd819209ebabaf6549c8dabf189e`.

[22] Radostin Danev, Bart Buijsse, Maryam Khoshouei, Jürgen M Plitzko, and Wolfgang Baumeister. Volta potential phase plate for in-focus phase contrast transmission electron microscopy. *Proceedings of the National Academy of Sciences*, 111(44):15635–15640, 2014. ISSN 0027-8424. doi: 10.1073/pnas.1418377111. URL `www.pnas.org/cgi/doi/10.1073/pnas.1418377111http://www.pnas.org/cgi/doi/10.1073/pnas.1418377111`.

[23] Ali Dashti, Peter Schwander, Robert Langlois, Russell Fung, Wen Li, Ahmad Hosseinizadeh, Hstau Y Liao, Jesper Pallesen, Gyanesh Sharma, Vera A Stupina, Anne E Simon, Jonathan D Dinman, Joachim Frank, and Abbas Ourmazd. Trajectories of the ribosome as a Brownian nanomachine. *Proceedings of the National Academy of Sciences of the United States of America*, 111(49):17492–7, dec 2014. ISSN 1091-6490. doi: 10.1073/pnas.1419276111. URL `http://www.ncbi.nlm.nih.gov/pubmed/25422471http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4267381`.

[24] Charles C. David and Donald J. Jacobs. Principal Component Analysis: A Method for Determining the Essential Dynamics of Proteins. pages 193–226. Humana Press, Totowa, NJ, 2014. doi: 10.1007/978-1-62703-658-0_11. URL `http://link.springer.com/10.1007/978-1-62703-658-0{_}11`.

[25] Mark de Berg, Otfried Cheong, Marc van Kreveld, and Mark Overmars. *Computational Geometry: Algorithms and Applications*. Springer-Verlag TELOS, Santa Clara, CA, USA, 3rd ed. edition, 2008. ISBN 3540779736, 9783540779735.

[26] D. J. DE ROSIER and A. KLUG. Reconstruction of Three Dimensional Structures from Electron Micrographs. *Nature*, 217(5124):130–134, jan 1968. ISSN 0028-0836. doi: 10.1038/217130a0. URL `http://www.nature.com/doifinder/10.1038/217130a0`.

[27] N. H. Dekkers and H. de Lang. DIFFERENTIAL PHASE CONTRAST IN A STEM. *Optik*, 41(4):452–456, 1974. ISSN 00304026. URL `https://www.scopus.com/record/display.uri?eid=2-s2.0-0016329294{&}origin=inward{&}txGid=b8c2fb92f4327db70aeab91dab5e8616`.

[28] A. P. Dempster, N.M. Laird, and D.B. Rubin. *Maximum Likelihood from Incomplete Data via the EM Algorithm*, volume 39. 1976. ISBN 0000000779. doi: 10.2307/2984875.

[29] J. Dubochet, M. Adrian, J. Lepault, and A. W. McDowall. Emerging techniques: Cryo-electron microscopy of vitrified biological specimens. *Trends in Biochemical Sciences*, 10 (4):143–146, 1985. ISSN 09680004. doi: 10.1016/0968-0004(85)90150-1. URL `https://www.cell.com/trends/biochemical-sciences/pdf/0968-0004(85)90150-1.pdf`.

[30] Lesley A Earl, Veronica Falconieri, Jacqueline LS Milne, and Sriram Subramaniam. Cryo-EM: beyond the microscope. *Current Opinion in Structural Biology*, 46:71–78, oct 2017. ISSN 0959440X. doi: 10.1016/j.sbi.2017.06.002. URL `http://www.ncbi.nlm.nih.gov/`

pubmed/28646653http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5683925https://linkinghub.elsevier.com/retrieve/pii/S0959440X17300180.

[31] Edward H. Egelman. The Current Revolution in Cryo-EM. *Biophysical Journal*, 110(5):1008–1012, mar 2016. ISSN 00063495. doi: 10.1016/j.bpj.2016.02.001. URL http://www.ncbi.nlm.nih.gov/pubmed/26958874http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4788751https://linkinghub.elsevier.com/retrieve/pii/S0006349516001429.

[32] Nadav Elad, Daniel K Clare, Helen R Saibil, and Elena 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, 2008. ISSN 1095-8657. doi: 10.1016/j.jsb.2007.11.007. URL http://www.ncbi.nlm.nih.gov/pubmed/18166488.

[33] Nadav Elad, Giuliano Bellapadrona, Lothar Houben, Irit Sagi, and Michael Elbaum. Detection of isolated protein-bound metal ions by single-particle cryo-STEM. *Proceedings of the National Academy of Sciences*, 114(42):11139–11144, 2017. ISSN 0027-8424. doi: 10.1073/pnas.1708609114. URL http://www.pnas.org/lookup/doi/10.1073/pnas.1708609114.

[34] Michael Elbaum. Expanding horizons of cryo-tomography to larger volumes. *Current Opinion in Microbiology*, 43:155–161, jun 2018. ISSN 13695274. doi: 10.1016/j.mib.2018.01.001. URL https://linkinghub.elsevier.com/retrieve/pii/S1369527417300498.

[35] Michael Elbaum. Quantitative Cryo-Scanning Transmission Electron Microscopy of Biological Materials. *Advanced Materials*, 30(41):1706681, oct 2018. ISSN 09359648. doi: 10.1002/adma.201706681. URL http://doi.wiley.com/10.1002/adma.201706681.

[36] Michael Elbaum, Sharon G. Wolf, and Lothar Houben. Cryo-scanning transmission electron tomography of biological cells. *MRS Bulletin*, 41(07):542–548, jul 2016. ISSN 0883-7694. doi: 10.1557/mrs.2016.136. URL http://www.journals.cambridge.org/abstract{_}S0883769416001366.

[37] Dominika Elmlund and Hans Elmlund. Cryogenic Electron Microscopy and Single-Particle Analysis. *Annual Review of Biochemistry*, 84(1):499–517, jun 2015. doi: 10.1146/annurev-biochem-060614-034226. URL www.annualreviews.org.

[38] Dominika Elmlund, Sarah N Le, and Hans Elmlund. High-resolution cryo-EM: the nuts and bolts. *Current Opinion in Structural Biology*, 46:1–6, oct 2017. ISSN 0959440X. doi: 10.1016/j.sbi.2017.03.003. URL https://linkinghub.elsevier.com/retrieve/pii/S0959440X16301956.

[39] Andreas Engel and Christian Colliex. Application of scanning transmission electron microscopy to the study of biological structure. Technical report, 1993. URL https://ac.els-cdn.com/095816699390005H/1-s2.0-095816699390005H-main.pdf?{_}tid=7f0a3437-1639-4746-9f7f-0e4e68b2dc08{&}acdnat=1544349546{_}f0155ec4b5b4ad41a574c2a37bd518ca.

[40] Richard P Feynman, Robert B Leighton, Matthew Sands, and E M Hafner. The Feynman Lectures on Physics; Vol. I. *Quantum Computation and Quantum Information American Journal of Physics*, 33:558, 1965. doi: 10.1119/1.1972241. URL http://aapt.scitation.org/toc/ajp/33/9.

[41] Emil Fischer. Ueber den Einfluss der Konfiguration auf die Wirkung der Enzyme III. *Berichte der deutschen chemischen Gesellschaft*, 28(2):1429–1438, 1895. ISSN 10990682. doi: 10.1002/cber.18950280243.

[42] Niels Fischer, Andrey L Konevega, Wolfgang Wintermeyer, Marina V Rodnina, and Holger Stark. Ribosome dynamics and tRNA movement by time-resolved electron cryomicroscopy. *Nature*, 466(7304):329–33, jul 2010. ISSN 1476-4687. doi: 10.1038/nature09206. URL `http://dx.doi.org/10.1038/nature09206`.

[43] Niels Fischer, Piotr Neumann, Andrey L. Konevega, Lars V. Bock, Ralf Ficner, Marina V. Rodnina, and Holger Stark. Structure of the E. coli ribosome–EF-Tu complex at <3 Å resolution by Cs-corrected cryo-EM. *Nature*, 520(7548):567–570, apr 2015. ISSN 0028-0836. doi: 10.1038/nature14275. URL `http://www.ncbi.nlm.nih.gov/pubmed/25707802http://www.nature.com/articles/nature14275`.

[44] Joachim Frank. Advances in the field of single-particle cryo-electron microscopy over the last decade. *Nature Protocols*, 12(2):209–212, jan 2017. ISSN 1754-2189. doi: 10.1038/nprot.2017.004. URL `http://www.nature.com/doifinder/10.1038/nprot.2017.004`.

[45] Joachim Frank and Abbas Ourmazd. Continuous changes in structure mapped by manifold embedding of single-particle data in cryo-EM. *Methods*, 100:61–67, 2016. doi: 10.1016/j.ymeth.2016.02.007. URL `http://dx.doi.org/10.1016/j.ymeth.2016.02.007`.

[46] H Frauenfelder. Conformational Substates In Proteins. *Annual Review of Biophysics and Biomolecular Structure*, 17(1):451–479, 1988. ISSN 10568700. doi: 10.1146/annurev.biophys.17.1.451. URL `www.annualreviews.orghttp://biophys.annualreviews.org/cgi/doi/10.1146/annurev.biophys.17.1.451`.

[47] H Frauenfelder, S. Sligar, and P. Wolynes. The energy landscapes and motions of proteins. *Science*, 254(5038):1598–1603, dec 1991. ISSN 0036-8075. doi: 10.1126/science.1749933. URL `http://science.sciencemag.org/content/254/5038/1598.abstract`.

[48] Peter M. Frederik and D.H.W. Hubert. Cryoelectron Microscopy of Liposomes. *Methods in Enzymology*, 391:431–448, jan 2005. ISSN 0076-6879. doi: 10.1016/S0076-6879(05)91024-0. URL `https://www.sciencedirect.com/science/article/pii/S0076687905910240`.

[49] Dominique P Frueh, Andrew C Goodrich, Subrata H Mishra, and Scott R Nichols. NMR methods for structural studies of large monomeric and multimeric proteins, oct 2013. ISSN 0959440X. URL `http://www.ncbi.nlm.nih.gov/pubmed/23850141http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3805735`.

[50] Jie Fu, Haixiao Gao, and Joachim Frank. Unsupervised classification of single particles by cluster tracking in multi-dimensional space. *Journal of structural biology*, 157:226–239, 2007. doi: 10.1016/j.jsb.2006.06.012.

[51] Haixiao Gao, Mikel Valle, Måns Ehrenberg, and Joachim Frank. Dynamics of EF-G interaction with the ribosome explored by classification of a heterogeneous cryo-EM dataset. *Journal of Structural Biology*, 147(3):283–290, 2004. ISSN 10478477. doi: 10.1016/j.jsb.2004.02.008.

[52] Barry J Grant, Alemayehu A Gorfe, and J Andrew Mccammon. Large conformational changes in proteins: signaling and other functions. *Current Opinion in Structural Biology*, 20:142–147, 2010. doi: 10.1016/j.sbi.2009.12.004. URL `www.sciencedirect.com`.

[53] K. Gunasekaran, Buyong Ma, and Ruth Nussinov. Is allostery an intrinsic property of all dynamic proteins? *Proteins: Structure, Function, and Bioinformatics*, 57(3):433–443, jul 2004. ISSN 08873585. doi: 10.1002/prot.20232. URL `http://doi.wiley.com/10.1002/prot.20232`.

[54] M. Haider, H. Rose, S. Uhlemann, B. Kabius, and K. Urban. Towards 0.1 nm resolution with the first spherically corrected transmission electron microscope. *Journal of Electron Microscopy*, 47(5):395–405, jan 1998. ISSN 0022-0744. doi: 10.1093/oxfordjournals.jmicro.a023610. URL `https://academic.oup.com/jmicro/article-lookup/doi/10.1093/oxfordjournals.jmicro.a023610`.

[55] Peter Hall, J. S. Marron, and Byeong U. Park. Smoothed cross-validation. *Probability Theory and Related Fields*, 92(1):1–20, mar 1992. ISSN 0178-8051. doi: 10.1007/BF01205233. URL `http://link.springer.com/10.1007/BF01205233`.

[56] Eric Hanssen. Biological and Medical Physics, Biomedical Engineering Cellular Imaging Electron Tomography and Related Techniques. URL `http://www.springer.com/series/3740`.

[57] Karl Joseph Hanszen. The Optical Transfer Theory of the Electron Microscope: Fundamental Principles and Applications. *Advances in Imaging and Electron Physics*, 207: 251–341, 2018. ISSN 10765670. doi: 10.1016/bs.aiep.2018.04.001.

[58] David Haselbach, Ilya Komarov, Dmitry E. Agafonov, Klaus Hartmuth, Benjamin Graf, Olexandr Dybkov, Henning Urlaub, Berthold Kastner, Reinhard Lührmann, and Holger Stark. Structure and Conformational Dynamics of the Human Spliceosomal B act Complex. *Cell*, 172(3):454–464.e11, jan 2018. ISSN 00928674. doi: 10.1016/j.cell.2018.01.010. URL `https://linkinghub.elsevier.com/retrieve/pii/S0092867418300473`.

[59] Richard Henderson. The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules. *Quarterly Reviews of Biophysics*, 28(02):171, 1995. ISSN 14698994. doi: 10.1017/S003358350000305X. URL `http://www.journals.cambridge.org/abstract{_}S003358350000305X`.

[60] Katherine Henzler-Wildman and Dorothee Kern. Dynamic personalities of proteins. *Nature*, 450(7172):964–972, dec 2007. ISSN 0028-0836. doi: 10.1038/nature06522. URL `http://www.nature.com/doifinder/10.1038/nature06522`.

[61] Katherine A. Henzler-Wildman, Ming Lei, Vu Thai, S. Jordan Kerns, Martin Karplus, and Dorothee Kern. A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature*, 450(7171):913–916, dec 2007. ISSN 0028-0836. doi: 10.1038/nature06407. URL `http://www.nature.com/articles/nature06407`.

[62] Katherine A. Henzler-Wildman, Vu Thai, Ming Lei, Maria Ott, Magnus Wolf-Watz, Tim Fenn, Ed Pozharski, Mark A. Wilson, Gregory A. Petsko, Martin Karplus, Christian G. Hübner, and Dorothee Kern. Intrinsic motions along an enzymatic reaction trajectory. *Nature*, 450(7171):838–844, dec 2007. ISSN 0028-0836. doi: 10.1038/nature06410. URL `http://www.nature.com/articles/nature06410`.

[63] Vincent J Hilser, James O Wrabl, and Hesam N Motlagh. Structural and Energetic Basis of Allostery. *Annual Review of Biophysics*, 41(1):585–609, 2012. ISSN 1936-122X. doi: 10.1146/annurev-biophys-050511-102319. URL `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3935618/pdf/nihms455166.pdfhttp://www.annualreviews.org/doi/10.1146/annurev-biophys-050511-102319`.

[64] Jerome K. Hyun, Peter Ercius, and David A. Muller. Beam spreading and spatial resolution in thick organic specimens. *Ultramicroscopy*, 109(1):1–7, dec 2008. ISSN 03043991. doi: 10.1016/j.ultramic.2008.07.003. URL `https://www.sciencedirect.com/science/article/pii/S0304399108001915`.

[65] Aapo Hyvärinen, Jarmo Hurri, and Patrik O. Hoyer. Principal Components and Whitening. pages 93–130. Springer, London, 2009. doi: 10.1007/978-1-84882-491-1_5. URL `http://link.springer.com/10.1007/978-1-84882-491-1{_}5`.

[66] Tilak Jain, Patrick Sheehan, John Crum, Bridget Carragher, and Clinton S. Potter. Spotiton: A prototype for an integrated inkjet dispense and vitrification system for cryo-TEM. *Journal of Structural Biology*, 179(1):68–75, jul 2012. ISSN 10478477. doi: 10.1016/j.jsb.2012.04.020. URL `http://www.ncbi.nlm.nih.gov/pubmed/22569522http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3378829http://linkinghub.elsevier.com/retrieve/pii/S1047847712001232`.

[67] E M James and N D Browning. Practical aspects of atomic resolution imaging and analysis in STEM. *Ultramicroscopy*, pages 1–15, 1999. URL `papers2://publication/uuid/46909C64-FCD5-43FB-BF4D-E58F659FDE02`.

[68] I. T. Jolliffe. Choosing a Subset of Principal Components or Variables. pages 92–114. Springer, New York, NY, 1986. doi: 10.1007/978-1-4757-1904-8_6. URL `http://link.springer.com/10.1007/978-1-4757-1904-8{_}6`.

[69] Martin Karplus and J. Andrew McCammon. Molecular dynamics simulations of biomolecules. *Nature Structural Biology*, 9(9):646–652, sep 2002. ISSN 10728368. doi: 10.1038/nsb0902-646. URL `http://www.nature.com/doifinder/10.1038/nsb0902-646`.

[70] Johannes Kästner. Umbrella sampling. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 1(6):932–942, nov 2011. ISSN 17590876. doi: 10.1002/wcms.66. URL `http://doi.wiley.com/10.1002/wcms.66`.

[71] E Katsevich, A Katsevich, and A Singer. Covariance Matrix Estimation for the Cryo-EM Heterogeneity Problem. *SIAM journal on imaging sciences*, 8 (1):126–185, jan 2015. ISSN 1936-4954. doi: 10.1137/130935434. URL `http://www.ncbi.nlm.nih.gov/pubmed/25699132http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4331039`.

[72] Elizabeth H. Kellogg, Nisreen M.A. Hejab, Stuart Howes, Peter Northcote, John H. Miller, J. Fernando Díaz, Kenneth H. Downing, and Eva Nogales. Insights into the Distinct Mechanisms of Action of Taxane and Non-Taxane Microtubule Stabilizers from Cryo-EM Structures. *Journal of Molecular Biology*, 429(5):633–646, mar 2017. ISSN 00222836. doi: 10.1016/j.jmb.2017.01.001. URL `https://linkinghub.elsevier.com/retrieve/pii/S0022283617300153`.

[73] Dari Kimanius, Bjorn O Forsberg, Sjors Scheres, and Erik Lindahl. Accelerated cryo-EM structure determination with parallelisation using GPUs in RELION-2. Technical report, jun 2016. URL `http://biorxiv.org/lookup/doi/10.1101/059717`.

[74] Angus Kirkland, Lan-Yun Chang, Sarah Haigh, and Crispin Hetherington. Transmission electron microscopy without aberrations: Applications to materials science. *Current Applied Physics*, 8(3-4):425–428, may 2008. ISSN 1567-1739. doi: 10.1016/J.CAP.2007.10.065. URL `https://www.sciencedirect.com/science/article/pii/S1567173907002246`.

[75] M. Knoll and E. Ruska. Das Elektronenmikroskop. *Zeitschrift für Physik*, 78(5-6):318–339, may 1932. ISSN 1434-6001. doi: 10.1007/BF01342199. URL `http://link.springer.com/10.1007/BF01342199`.

[76] D E Koshland. Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proceedings of the National Academy of Sciences*, 44(2):98–104, 1958. ISSN 0027-8424. doi: 10.1073/pnas.44.2.98. URL `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC335371/pdf/pnas00681-0050.pdfhttp://www.pnas.org/cgi/doi/10.1073/pnas.44.2.98`.

[77] D. E. Koshland, J. G. Nemethy, and D. Filmer. Comparison of Experimental Binding Data and Theoretical Models in Proteins Containing Subunits. *Biochemistry*, 5(1):365–385, 1966. ISSN 15204995. doi: 10.1021/bi00865a047. URL `https://pubs.acs.org/sharingguidelines`.

[78] Werner Kuehlbrandt. The resolution revolution, mar 2014. ISSN 10959203. URL `http://www.ncbi.nlm.nih.gov/pubmed/24675944http://www.sciencemag.org/cgi/doi/10.1126/science.1251652`.

[79] Michal Kurzynski. *The Thermodynamic Machinery of Live*, volume 4. Springer, Heidelberg, 2006. ISBN 978-3-540-23888-1.

[80] Ivan Lazić and Eric G.T. Bosch. *Analytical Review of Direct Stem Imaging Techniques for Thin Samples*, volume 199. 2017. ISBN 9780128120910. doi: 10.1016/bs.aiep.2017.01.006.

[81] Ivan Lazić, Eric G.T. Bosch, and Sorin Lazar. Phase contrast STEM for thin samples: Integrated differential phase contrast. *Ultramicroscopy*, 160:265–280, 2016. ISSN 18792723. doi: 10.1016/j.ultramic.2015.10.011.

[82] Ivan Lazic, Eric G.T. Bosch, Sorin Lazar, Maarten Wirix, and Emrah Yücelen. Integrated Differential Phase Contrast (iDPC)–Direct Phase Imaging in STEM for Thin Samples. *Microscopy and Microanalysis*, 22(S3):36–37, 2016. ISSN 1431-9276. doi: 10.1017/S1431927616001033. URL `http://www.journals.cambridge.org/abstract{_}S1431927616001033`.

[83] Cheryl L. Limebeer, Jon P. Krohn, Shelley Cross-Mellor, Devin E. Litt, Klaus Peter Ossenkopp, and Linda A. Parker. Exposure to a context previously associated with nausea elicits conditioned gaping in rats: A model of anticipatory nausea. *Behavioural Brain Research*, 187(1):33–40, 2008. ISSN 01664328. doi: 10.1016/j.bbr.2007.08.024. URL `https://link.springer.com/content/pdf/10.1007{%}2FBF00276625.pdfhttp://www.ncbi.nlm.nih.gov/pubmed/6544679`.

[84] Wei Lü, Juan Du, April Goehring, and Eric Gouaux. Cryo-EM structures of the triheteromeric NMDA receptor and its allosteric modulation. *Science*, 355(6331), 2017. ISSN 10959203. doi: 10.1126/science.aal3729.

[85] Dmitry Lyumkis, Axel F. Brilot, Douglas L. Theobald, and Nikolaus Grigorieff. Likelihood-based classification of cryo-EM images using FREALIGN. *Journal of Structural Biology*, 183(3):377–388, sep 2013. ISSN 1047-8477. doi: 10.1016/J.JSB.2013.07.005. URL `https://www.sciencedirect.com/science/article/pii/S1047847713001858`.

[86] Buyong Ma, Sandeep Kumar, Chung-Jung Tsai, and Ruth Nussinov. Folding funnels and binding mechanisms. *Protein Engineering, Design and Selection*, 12(9):713–720, sep 1999. ISSN 1741-0134. doi: 10.1093/protein/12.9.713. URL `https://academic.oup.com/peds/article-lookup/doi/10.1093/protein/12.9.713`.

[87] Christine P Macol, Hiro Tsuruta, Boguslaw Stec, and Evan R Kantrowitz. Direct structural evidence for a concerted allosteric transition in Escherichia coli aspartate transcarbamoylase. *Nature Structural Biology*, 8(5):423–426, 2001. ISSN 10728368. doi: 10.1038/87582. URL `http://structbio.nature.com`.

[88] S Majert and H Kohl. High-resolution STEM imaging with a quadrant detector-Conditions for differential phase contrast microscopy in the weak phase object approximation. *Ultramicroscopy*, 148:81–86, 2014. doi: 10.1016/j.ultramic.2014.09.009. URL `http://dx.doi.org/10.1016/j.ultramic.2014.09.009`.

[89] Christopher Kroboth Materese, Christa Charisse Goldmon, and Garegin A Papoian. Hierarchical organization of eglin c native state dynamics is shaped by competing direct and water-mediated interactions. *Proceedings of the National Academy of Sciences of the United States of America*, 105(31):10659–64, aug 2008. ISSN 1091-6490. doi: 10.1073/pnas.0801850105. URL `http://www.pnas.org/content/105/31/10659.full`.

[90] G. McMullan, A R Faruqi, and R Henderson. Direct Electron Detectors. In *Methods in Enzymology*, volume 579, pages 1–17. 2016. ISBN 9780128053829. doi: 10.1016/bs.mie.2016.05.056. URL `http://dx.doi.org/10.1016/bs.mie.2016.05.056`.

[91] JACQUES MONOD, JEFFRIES WYMAN, and JEAN-PIERRE CHANGEUX. On the Nature of Allosteric Transitions: A Plausible Model. *Selected Papers in Molecular Biology by Jacques Monod*, pages 593–623, jan 1978. doi: 10.1016/B978-0-12-460482-7.50052-X. URL `https://www.sciencedirect.com/science/article/pii/B978012460482750052X`.

[92] Hesam N. Motlagh, James O. Wrabl, Jing Li, and Vincent J. Hilser. The ensemble nature of allostery. *Nature*, 508(7496):331–339, apr 2014. ISSN 0028-0836. doi: 10.1038/nature13001. URL `http://www.ncbi.nlm.nih.gov/pubmed/24740064http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4224315http://www.nature.com/articles/nature13001`.

[93] Knut Müller, Florian F. Krause, Armand Béché, Marco Schowalter, Vincent Galioit, Stefan Löffler, Johan Verbeeck, Josef Zweck, Peter Schattschneider, and Andreas Rosenauer. Atomic electric fields revealed by a quantum mechanical approach to electron picodiffraction. *Nature Communications*, 5(1):5653, dec 2014. ISSN 20411723. doi: 10.1038/ncomms6653. URL `http://www.nature.com/articles/ncomms6653`.

[94] Kuniaki Nagayama and Radostin Danev. Phase contrast electron microscopy: Development of thin-film phase plates and biological applications, 2008. ISSN 09628436. URL `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2453530/pdf/rstb20082268.pdf`.

[95] Takanori Nakane, Dari Kimanius, Erik Lindahl, and Sjors HW Scheres. Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION. *eLife*, 7, jun 2018. ISSN 2050084X. doi: 10.7554/eLife.36861. URL `https://elifesciences.org/articles/36861`.

[96] Peter D. Nellist. Electron microscopy: Atomic resolution comes into phase. *Nature Physics*, 8(8):586–587, 2012. ISSN 1745-2473. doi: 10.1038/nphys2357. URL `http://dx.doi.org/10.1038/nphys2357`.

[97] Ruth Nussinov and Chung Jung Tsai. Free energy diagrams for protein function, mar 2014. ISSN 10745521. URL `https://www.sciencedirect.com/science/article/pii/S107455211400009X?via{%}3Dihub`.

[98] José Nelson Onuchic and Peter G Wolynes. Theory of protein folding, feb 2004. ISSN 0959440X. URL `https://www.sciencedirect.com/science/article/pii/S0959440X04000107?via{%}3Dihub`.

[99] E V Orlova and H R Saibil. Structural Analysis of Macromolecular Assemblies by Electron Microscopy. *Chem. Rev*, 111:7710–7748, 2011. doi: 10.1021/cr100353t.

[100] Pawel A Penczek, Marek Kimmel, and Christian M T Spahn. Identifying conformational states of macromolecules by eigen-analysis of resampled cryo-EM images. *Structure*, 19 (11):1582–1590, 2011. ISSN 09692126. doi: 10.1016/j.str.2011.10.003.

[101] Stephen J. Pennycook and Peter D. Nellist. *Scanning Transmission Electron Microscopy*. Springer New York, New York, NY, 2011. ISBN 978-1-4419-7199-9. doi: 10.1007/978-1-4419-7200-2. URL `http://link.springer.com/10.1007/978-1-4419-7200-2`.

[102] Alberto Perez, Joseph A Morrone, Carlos Simmerling, and Ken A Dill. Advances in free-energy-based simulations of protein folding and ligand binding. 2016. doi: 10.1016/j.sbi.2015.12.002. URL `http://dx.doi.org/10.1016/j.sbi.2015.12.002`.

[103] Juan R Perilla, Boon Chong Goh, C Keith Cassidy, Bo Liu, Rafael C Bernardi, Till Rudack, Hang Yu, Zhe Wu, and Klaus Schulten. Molecular dynamics simulations of large macromolecular complexes. *Current Opinion in Structural Biology*, 31:64–74, apr 2015. ISSN 0959-440X. doi: 10.1016/J.SBI.2015.03.007. URL `https://www.sciencedirect.com/science/article/pii/S0959440X15000342`.

[104] M. F. PERUTZ. Stereochemistry of Cooperative Effects in Haemoglobin: Haem–Haem Interaction and the Problem of Allostery. *Nature*, 228(5273):726–734, nov 1970. ISSN 0028-0836. doi: 10.1038/228726a0. URL `http://www.nature.com/doifinder/10.1038/228726a0`.

[105] Menahem Pirchi, Guy Ziv, Inbal Riven, Sharona Sedghani Cohen, Nir Zohar, Yoav Barak, and Gilad Haran. Single-molecule fluorescence spectroscopy maps the folding landscape of a large protein. *Nature Communications*, 2(1):493, sep 2011. ISSN 2041-1723. doi: 10.1038/ncomms1504. URL `http://www.nature.com/articles/ncomms1504`.

[106] Ali Punjani, John L Rubinstein, David J Fleet, and Marcus A Brubaker. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nature Methods*, 14 (3):290–296, mar 2017. ISSN 1548-7091. doi: 10.1038/nmeth.4169. URL `http://www.nature.com/articles/nmeth.4169`.

[107] Ivan Razinkov, Venkat Dandey, Hui Wei, Zhening Zhang, David Melnekoff, William J Rice, Christoph Wigge, Clinton S Potter, and Bridget Carragher. A new method for vitrifying samples for cryoEM. *Journal of structural biology*, 195(2):190–198, 2016. ISSN 1095-8657. doi: 10.1016/j.jsb.2016.06.001. URL `http://www.ncbi.nlm.nih.gov/pubmed/27288865http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5464370`.

[108] Jean-Paul Renaud, Chun-wa Chung, U. Helena Danielson, Ursula Egner, Michael Hennig, Roderick E. Hubbard, and Herbert Nar. Biophysics in drug discovery: impact, challenges and opportunities. *Nature Reviews Drug Discovery*, 15(10):679–698, oct 2016. ISSN 1474-1776. doi: 10.1038/nrd.2016.123. URL `http://www.nature.com/articles/nrd.2016.123`.

[109] Jean-Paul Renaud, Ashwin Chari, Claudio Ciferri, Wen-ti Liu, Hervé-William Rémigy, Holger Stark, and Christian Wiesmann. Cryo-EM in drug discovery: achievements, limitations and prospects. *Nature Reviews Drug Discovery*, 17(7):471–492, jun 2018. ISSN 1474-1776. doi: 10.1038/nrd.2018.77. URL `http://www.nature.com/doifinder/10.1038/nrd.2018.77`.

[110] Peter Rez. Comparison of phase contrast transmission electron microscopy with optimized scanning transmission annular dark field imaging for protein imaging. *Ultramicroscopy*, 96(1):117–124, jul 2003. ISSN 0304-3991. doi: 10.1016/S0304-3991(02)00436-9. URL `https://www.sciencedirect.com/science/article/pii/S0304399102004369`.

[111] Peter Rez, Thomas Larsen, and Michael Elbaum. Exploring the theoretical basis and limitations of cryo-STEM tomography for thick biological specimens. *Journal of Structural Biology*, 196(3):466–478, dec 2016. ISSN 10958657. doi: 10.1016/j.jsb.2016.09.014. URL `http://www.sciencedirect.com/science/article/pii/S1047847716302039`.

[112] Peter Rez, Thomas Larsen, and Michael Elbaum. Exploring the theoretical basis and limitations of cryo-STEM tomography for thick biological specimens. *Journal of Structural Biology*, 196(3):466–478, dec 2016. ISSN 10958657. doi: 10.1016/j.jsb.2016.09.014. URL `http://www.sciencedirect.com/science/article/pii/S1047847716302039`.

[113] H Rose. NONSTANDARD IMAGING METHODS IN ELECTRON MICROSCOPY *. Technical report, 1977. URL `http://xrm.phys.northwestern.edu/research/pdf{_}papers/1977/rose{_}ultramicroscopy{_}1977.pdf`.

[114] Alan M Roseman. Particle finding in electron micrographs using a fast local correlation algorithm. *Ultramicroscopy*, 94(3-4):225–236, apr 2003. ISSN 0304-3991. doi: 10.1016/S0304-3991(02)00333-9. URL `https://www.sciencedirect.com/science/article/pii/S0304399102003339`.

[115] Alan M. Roseman, Shaoxia Chen, Helen White, Kerstin Braig, and Helen R. Saibil. The chaperonin ATPase cycle: Mechanism of allosteric switching and movements of substrate-binding domains in GroEL. *Cell*, 87(2):241–251, 1996. ISSN 00928674. doi: 10.1016/S0092-8674(00)81342-2.

[116] Marissa G Saunders and Gregory A Voth. Coarse-graining of multiprotein assemblies. *Current Opinion in Structural Biology*, 22(2):144–150, apr 2012. ISSN 0959-440X. doi: 10.1016/J.SBI.2012.01.003. URL `https://www.sciencedirect.com/science/article/pii/S0959440X1200005X`.

[117] Giovanna Scapin, Clinton S. Potter, and Bridget Carragher. Cryo-EM for Small Molecules Discovery, Design, Understanding, and Application. *Cell Chemical Biology*, aug 2018. ISSN 2451-9456. doi: 10.1016/J.CHEMBIOL.2018.07.006. URL `https://www.sciencedirect.com/science/article/pii/S2451945618302332?via{%}3Dihub`.

[118] Sjors H W Scheres. Classification of structural heterogeneity by maximum-likelihood methods. *Methods in enzymology*, 482:295–320, jan 2010. ISSN 1557-7988. doi: 10.1016/S0076-6879(10)82012-9. URL `http://www.sciencedirect.com/science/article/pii/S0076687910820129`.

[119] Sjors 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`.

[120] Sjors H W Scheres. A Bayesian View on Cryo-EM Structure Determination. *Journal of Molecular Biology*, 415:406–418, 2012. doi: 10.1016/j.jmb.2011.11.010.

[121] Sjors H W Scheres, Haixiao Gao, Mikel Valle, Gabor T Herman, Paul P B Eggermont, Joachim Frank, and Jose-maria Carazo. Disentangling conformational states of macromolecules in 3D-EM through likelihood optimization. *Nature Methods*, 4(1):27–29, dec 2007. doi: 10.1038/NMETH992.

[122] Sjors H.W. Scheres. Semi-automated selection of cryo-EM particles in RELION-1.3. *Journal of Structural Biology*, 189(2):114–122, feb 2015. ISSN 1047-8477. doi: 10.1016/J.JSB.2014.11.010. URL `https://www.sciencedirect.com/science/article/pii/S1047847714002615`.

[123] Sjors H.W. Scheres, Mikel Valle, Rafael Nuñez, Carlos O.S. Sorzano, Roberto Marabini, Gabor T Herman, and Jose Maria Carazo. Maximum-likelihood multi-reference refinement for electron microscopy images. *Journal of Molecular Biology*, 348(1):139–149, 2005. ISSN 00222836. doi: 10.1016/j.jmb.2005.02.031. URL `https://ac.els-cdn.com/S0022283605001932/1-s2.0-S0022283605001932-main.pdf?{_}tid=049531b1-5823-48d7-87d3-acaeb8d626fd{&}acdnat=1543744371{_}37dae55d5f0c7692a3e342b47315cf5e`.

[124] Sjors H.W. Scheres, Haixiao Gao, Mikel Valle, Gabor T Herman, Paul P.B. Eggermont, Joachim Frank, and Jose Maria Carazo. Disentangling conformational states of macromolecules in 3D-EM through likelihood optimization. *Nature Methods*, 4(1):27–29, 2007. ISSN 15487091. doi: 10.1038/nmeth992. URL `http://npg.nature.com/reprintsandpermissions/`.

[125] I. Schlichting. The Catalytic Pathway of Cytochrome P450cam at Atomic Resolution. *Science*, 287(5458):1615–1622, mar 2000. ISSN 00368075. doi: 10.1126/science.287.5458.1615. URL `http://www.sciencemag.org/cgi/doi/10.1126/science.287.5458.1615`.

[126] Marieke Schor, Antonia S J S Mey, and Cait E MacPhee. Analytical methods for structural ensembles and dynamics of intrinsically disordered proteins. *Biophysical reviews*, 8(4):429–439, 2016. ISSN 1867-2450. doi: 10.1007/s12551-016-0234-6. URL `http://www.ncbi.nlm.nih.gov/pubmed/28003858http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5135723`.

[127] Jil Schrader, Fabian Henneberg, Ricardo A Mata, Kai Tittmann, Thomas R Schneider, Holger Stark, Gleb Bourenkov, and Ashwin Chari. The inhibition mechanism of human 20S proteasomes enables next-generation inhibitor design. *Science*, 353(6299):594–598, aug 2016. ISSN 10959203. doi: 10.1126/science.aaf8993. URL `http://www.ncbi.nlm.nih.gov/pubmed/27493187`.

[128] Benjamin Schuler, Andrea Soranno, Hagen Hofmann, and Daniel Nettels. Single-Molecule FRET Spectroscopy and the Polymer Physics of Unfolded and Intrinsically Disordered Proteins. *Annual Review of Biophysics*, 45(1):207–231, 2016. ISSN 1936-122X. doi: 10.1146/annurev-biophys-062215-010915. URL `www.annualreviews.orghttp://www.annualreviews.org/doi/10.1146/annurev-biophys-062215-010915`.

[129] David W. Scott and Stephan R. Sain. Multidimensional Density Estimation. *Handbook of Statistics*, 24:229–261, jan 2005. ISSN 0169-7161. doi: 10.1016/S0169-7161(04)24009-3. URL `https://www.sciencedirect.com/science/article/pii/S0169716104240093`.

[130] Maxim Shatsky, Richard J Hall, Eva Nogales, Jitendra Malik, and Steven E Brenner. Automated multi-model reconstruction from single-particle electron microscopy data. *Journal of Structural Biology*, 170:98–108, 2010. doi: 10.1016/j.jsb.2010.01.007. URL `https://ac.els-cdn.com/S1047847710000195/1-s2.0-S1047847710000195-main.pdf?{_}tid=c4cd081f-bd89-4037-bf8f-f0186f9e02eb{&}acdnat=1541226349{_}265429ec7ab8c0f95c7121f7c17c38f3`.

[131] Peter S. Shen. The 2017 Nobel Prize in Chemistry: cryo-EM comes of age. *Analytical and Bioanalytical Chemistry*, 410(8):2053–2057, mar 2018. ISSN 1618-2642. doi: 10.1007/s00216-018-0899-8. URL `http://www.ncbi.nlm.nih.gov/pubmed/29423601http://link.springer.com/10.1007/s00216-018-0899-8`.

[132] Naoya Shibata, Scott D. Findlay, Yuji Kohno, Hidetaka Sawada, Yukihito Kondo, and Yuichi Ikuhara. Differential phase-contrast microscopy at atomic resolution. *Nature*

Physics, 8(8):611–615, jun 2012. ISSN 1745-2473. doi: 10.1038/nphys2337. URL `http://www.nature.com/doifinder/10.1038/nphys2337`.

[133] Naoya Shibata, Takehito Seki, Gabriel Sánchez-Santolino, Scott D. Findlay, Yuji Kohno, Takao Matsumoto, Ryo Ishikawa, and Yuichi Ikuhara. Electric field imaging of single atoms. *Nature Communications*, 8:15631, may 2017. ISSN 2041-1723. doi: 10.1038/ncomms15631. URL `http://www.nature.com/doifinder/10.1038/ncomms15631`.

[134] Ichio Shimada, Takumi Ueda, Yutaka Kofuku, Matthew T. Eddy, and Kurt Wüthrich. GPCR drug discovery: integrating solution NMR data with crystal and cryo-EM structures. *Nature Reviews Drug Discovery*, nov 2018. ISSN 1474-1776. doi: 10.1038/nrd.2018.180. URL `http://www.nature.com/doifinder/10.1038/nrd.2018.180`.

[135] F J Sigworth. A maximum-likelihood approach to single-particle image refinement. *Journal of structural biology*, 122(3):328–39, jan 1998. ISSN 1047-8477. doi: 10.1006/jsbi.1998.4014. URL `http://www.sciencedirect.com/science/article/pii/S104784779894014X`.

[136] F J Sigworth. {A} {Maximum}-{Likelihood} {Approach} to {Single}-{Particle} {Image} {Refinement}. *Journal of Structural Biology*, 122(3):328–339, 1998. ISSN 10478477. URL `https://ac.els-cdn.com/S104784779894014X/1-s2.0-S104784779894014X-main.pdf?{_}tid=1b2e5062-ba46-4630-8dda-4fe9c437448d{&}acdnat=1543744390{_}577f7880a18d80927e841b623c794d46`.

[137] Fred J. Sigworth. Principles of cryo-EM single-particle image processing. *Microscopy*, 65(1):57–67, feb 2016. ISSN 2050-5698. doi: 10.1093/jmicro/dfv370. URL `http://www.ncbi.nlm.nih.gov/pubmed/26705325http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4749045https://academic.oup.com/jmicro/article-lookup/doi/10.1093/jmicro/dfv370`.

[138] Christian Mt Spahn and Pawel A Penczek. Exploring conformational modes of macromolecular assemblies by multiparticle cryo-EM, 2009. ISSN 0959440X. URL `www.sciencedirect.com`.

[139] Holger Stark and Ashwin Chari. Sample preparation of biological macromolecular assemblies for the determination of high-resolution structures by cryo-electron microscopy. *Microscopy (Oxford, England)*, 65(1):23–34, 2016. ISSN 20505701. doi: 10.1093/jmicro/dfv367. URL `https://academic.oup.com/jmicro/article-abstract/65/1/23/2579661`.

[140] Sriram Subramaniam, Lesley A Earl, Veronica Falconieri, Jacqueline LS Milne, and Edward H Egelman. Resolution advances in cryo-EM enable application to drug discovery. *Current Opinion in Structural Biology*, 41:194–202, dec 2016. ISSN 0959440X. doi: 10.1016/j.sbi.2016.07.009. URL `http://www.ncbi.nlm.nih.gov/pubmed/27552081http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5154827https://linkinghub.elsevier.com/retrieve/pii/S0959440X16300847`.

[141] Hemant D Tagare, Alp Kucukelbir, Fred J Sigworth, Hongwei Wang, and Murali Rao. Directly reconstructing principal components of heterogeneous particles from cryo-EM images. 2015. doi: 10.1016/j.jsb.2015.05.007. URL `http://dx.doi.org/10.1016/j.jsb.2015.05.007`.

[142] Florence Tama, Willy Wriggers, and Charles L Brooks III. Exploring Global Distortions of Biological Macromolecules and Assemblies from Low-resolution Structural Information and Elastic Network Theory. doi: 10.1016/S0022-2836(02)00627-7. URL `http://www.idealibrary.com`.

[143] Yong Zi Tan, Sriram Aiyer, Mario Mietzsch, Joshua A. Hull, Robert McKenna, Joshua Grieger, R. Jude Samulski, Timothy S. Baker, Mavis Agbandje-McKenna, and Dmitry Lyumkis. Sub-2 Å Ewald curvature corrected structure of an AAV2 capsid variant. *Nature Communications*, 9(1):1–11, 2018. ISSN 20411723. doi: 10.1038/s41467-018-06076-6. URL `http://dx.doi.org/10.1038/s41467-018-06076-6`.

[144] Simon J. Teague. Implications of protein flexibility for drug discovery. *Nature Reviews Drug Discovery*, 2(7):527–541, 2003. ISSN 14741776. doi: 10.1038/nrd1129.

[145] C J Tsai, S Kumar, B Ma, and R Nussinov. Folding funnels, binding funnels, and protein function. *Protein science : a publication of the Protein Society*, 8(6):1181–90, jun 1999. ISSN 0961-8368. doi: 10.1110/ps.8.6.1181. URL `http://www.ncbi.nlm.nih.gov/pubmed/10386868http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2144348`.

[146] Marin Van Heel. Detection of objects in quantum-noise-limited images. *Ultramicroscopy*, 7 (4):331–341, jan 1982. ISSN 0304-3991. doi: 10.1016/0304-3991(82)90258-3. URL `https://www.sciencedirect.com/science/article/pii/0304399182902583`.

[147] Marin van Heel and Joachim Frank. Use of multivariates statistics in analysing the images of biological macromolecules. *Ultramicroscopy*, 6(1):187–194, jan 1981. ISSN 0304-3991. doi: 10.1016/S0304-3991(81)80197-0. URL `https://www.sciencedirect.com/science/article/pii/S0304399181801970`.

[148] Marin van Heel and Wilko Keegstra. IMAGIC: A fast, flexible and friendly image analysis software system. *Ultramicroscopy*, 7(2):113–129, jan 1981. ISSN 0304-3991. doi: 10.1016/0304-3991(81)90001-2. URL `https://www.sciencedirect.com/science/article/pii/0304399181900012`.

[149] Marin van Heel, Rodrigo V. Portugal, and Michael Schatz. Multivariate Statistical Analysis of Large Datasets: Single Particle Electron Microscopy. *Open Journal of Statistics*, 06(04):701–739, jul 2016. ISSN 2161-718X. doi: 10.4236/ojs.2016.64059. URL `http://www.scirp.org/journal/doi.aspx?DOI=10.4236/ojs.2016.64059`.

[150] R.H. Wade. A brief look at imaging and contrast transfer. *Ultramicroscopy*, 46(1-4): 145–156, oct 1992. ISSN 0304-3991. doi: 10.1016/0304-3991(92)90011-8. URL `https://www.sciencedirect.com/science/article/pii/0304399192900118?via{%}3Dihub`.

[151] J S Wall and J.F. Hainfeld. Mass Mapping With The Scanning-Transmission Electron-Microscope. *Annual Review Of Biophysics And Biophysical Chemistry*, 15:355–376, 1986.

[152] Sharon Grayer Wolf, Lothar Houben, and Michael Elbaum. Cryo-scanning transmission electron tomography of vitrified cells. *Nature Methods*, 11(4):423–428, feb 2014. ISSN 1548-7091. doi: 10.1038/nmeth.2842. URL `http://www.nature.com/doifinder/10.1038/nmeth.2842`.

[153] Sharon Grayer Wolf, Peter Rez, and Michael Elbaum. Phosphorus detection in vitrified bacteria by cryo-STEM annular dark-field analysis. *Journal of Microscopy*, 260(2):227–233, nov 2015. ISSN 13652818. doi: 10.1111/jmi.12289. URL `http://doi.wiley.com/10.1111/jmi.12289`.

[154] Sharon Grayer Wolf, Yael Mutsafi, Tali Dadosh, Tal Ilani, Zipora Lansky, Ben Horowitz, Sarah Rubin, Michael Elbaum, and Deborah Fass. 3D visualization of mitochondrial solid-phase calcium stores in whole cells. *eLife*, 6, 2017. ISSN 2050084X. doi: 10.7554/eLife.29929. URL `https://doi.org/10.7554/eLife.29929.001`.

[155] Kurt Wüthrich. Protein structure determination in solution by NMR spectroscopy. *Journal of Biological Chemistry*, 265(36):22059–22062, 1990. ISSN 00219258. doi: 10.1126/science.2911719.

[156] Emrah Yücelen, Ivan Lazić, and Eric G. T. Bosch. Phase contrast scanning transmission electron microscopy imaging of light and heavy atoms at the limit of contrast and resolution. *Scientific Reports*, 8(1):2676, dec 2018. ISSN 2045-2322. doi: 10.1038/s41598-018-20377-2. URL `http://www.nature.com/articles/s41598-018-20377-2`.

[157] Qinfen Zhang, Xinghong Dai, Yao Cong, Junjie Zhang, Dong Hua Chen, Matthew T. Dougherty, Jiangyong Wang, Steven J. Ludtke, Michael F. Schmid, and Wah Chiu. Cryo-EM structure of a molluscan hemocyanin suggests its allosteric mechanism. *Structure*, 21 (4):604–613, 2013. ISSN 09692126. doi: 10.1016/j.str.2013.02.018. URL `http://dx.doi.org/10.1016/j.str.2013.02.018`.

[158] Pavel I Zhuravlev and Garegin A Papoian. Protein functional landscapes, dynamics, allostery: a tortuous path towards a universal theoretical framework. *Quarterly reviews of biophysics*, 43(3):295–332, aug 2010. ISSN 1469-8994. doi: 10.1017/S0033583510000119. URL `http://www.ncbi.nlm.nih.gov/pubmed/20819242`.

[159] Jasenko Zivanov, Takanori Nakane, Bjorn Forsberg, Dari Kimanius, Wim J.H. Hagen, Erik Lindahl, and Sjors H.W. Scheres. RELION-3: new tools for automated high-resolution cryo-EM structure determination. *bioRxiv*, page 421123, 2018. ISSN 2050-084X. doi: 10.1101/421123. URL `https://www.biorxiv.org/content/early/2018/09/19/421123`.

Curriculum Vitae

Felix Lambrecht

Alte Dorfstraße 18a
07751 Jena, Germany

Born: March, 5th 1991, Kassel

University Education

since 2016 Max-Planck-Institute for Biophysical Chemistry, Department of Structural Dynamics. PhD Thesis on "Computational Methods for the Determination of highly flexible molecular machines by cryo-EM".

2014-2016 University of Goettingen: Master of Science Microbiology and Biochemistry. Master Thesis at the Max-Planck-Institute for Biophysical Chemistry on the "Computational Analysis of Heterogeneity in cryo-EM Datasets"

2011-2014 University of Goettingen: Bachelor of Science Biology with main subject biochemistry. Bachelor Thesis in the Department of Plant Biochemistry (Albrecht-von-Haller Institute) on "Purification and Crystallization of a bifunctional Waxsynthase"

School Education

2011 Abitur at the Albert-Schweitzer-Schule Hofgeismar