TUDelft

Delft University of Technology

Imaging Life at the Molecular Scale using Electrons

Huber, S.

DOI
10.4233/uuid:9673dae3-5a3d-4544-a48e-dcd088c72625

Publication date
2024

Document Version
Final published version

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Imaging Life

at the Molecular Scale

using Electrons

Stefan Huber

Imaging Life at the Molecular Scale using Electrons

Imaging Life at the Molecular Scale using Electrons

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology
by the authority of the Rector Magnificus prof. dr. ir. T.H.J.J. van der Hagen
chair of the Board for Doctorates
to be defended publicly on
Wednesday, 4 September 2024 at 17:30

by

Stefan HUBER

Master of Science in Molecular Biotechnology,
Universität Heidelberg, Germany

born in Gengenbach, Germany.

This dissertation has been approved by the promotors.

Composition of the doctoral committee:

Rector Magnificus	chairperson
Prof. dr. A.M. Dogterom	Technische Universiteit Delft, *promotor*
Dr. A. Jakobi	Technische Universiteit Delft, *copromotor*

Independent members:

Dr. S.M. Depken	Technische Universiteit Delft
Prof. dr. R. Efremov	Vrije Universiteit Brussel
Prof. dr. C. Joo	Technische Universiteit Delft
Prof. dr. A. Perrakis	Utrecht University, Netherlands Cancer Institute
Prof. dr. F. Pfeifer	Technische Universität Darmstadt
Prof. dr. B.D. Rowland	Netherlands Cancer Institute,
	Technische Universiteit Delft, *reservelid*

[...] just look at the thing!

Richard Feynman

Contents

1

Introduction

Cryogenic electron microscopy (cryo-EM) has become a powerful technique to understand the structure and function of biological macromolecules. This thesis encompasses parts ranging from sample preparation techniques in cryo-EM to structural studies of protein complexes using cryo-EM. The thesis title "Imaging Life at the Molecular Scale using Electrons" attempts to put these parts under a single thematic framework.

The initial segment of this introductory chapter presents a brief overview of the core concepts in structural biology, with a particular emphasis on the historical progression and recent advancements in cryo-EM, as well as the ongoing challenges faced in sample preparation techniques. The subsequent section centers on a microbial motility structure known as a gas vesicle, which is a primary subject of structural studies in this thesis, and offers an overview of pioneering structural studies on this protein assembly in the last decades.

Parts of this chapter are based on published articles:

Huber, S. T., Sarajlic, E., Huijink, R., Weis, F., Evers, W. H. & Jakobi, A. J. Nanofluidic chips for cryo-EM structure determination from picoliter sample volumes. *Elife* **11,** e72629 (2022)

Huber, S. T., Jakobi, A. J. Structural biology of microbial gas vesicles: Historical milestones and current knowledge. *Biochemical Society Transactions* (2024)

Huber, S. T., Terwiel, D., Evers, W. H., Maresca, D. & Jakobi, A. J. Cryo-EM structure of gas vesicles for buoyancy-controlled motility. *Cell* **186,** 975–986 (2023)

1

1.1 Studying the Molecular Machinery of Life

The human genome comprises approximately 20,000 genes, each encoding a protein that can be viewed as a tiny molecular machine (Figure 1.1). Structural biology aims to resolve the atomic arrangements in such biological macromolecules. By determining the precise configurations of atoms within these complex structures, the field provides fundamental insights into the function of the cellular machinery. This knowledge may subsequently facilitate targeted interventions in health and disease, for instance by enabling rational drug design.

In the context of atomic dimensions, proteins are giant entities. The median protein size in eucaryotes is 353 amino acids[3], bound together as a linear polypeptide chain. This corresponds to a molecular mass of ~40 kDa or ~40,000 hydrogen atoms. The chain folds into a (usually) defined three-dimensional shape of only a few nanometers in diameter - tiny with respect to common length scales humans typically experience. Light microscopy can extend our observational capabilities, yet its resolution is inherently limited to the wavelength range of visible light, approximately 400-650 nanometers - two orders of magnitude larger than a protein.

The scientific pursuit to understand biological macromolecules has catalysed continuous advancements in techniques, resulting in tools that can effectively probe the nanoscale dimensions of biological macromolecular structures. One such technique is X-ray crystallography. A single protein is too small to gather enough information from it to elucidate its structure. That is why the technique involves crystallization of millions of copies of the proteins into regular three-dimensional arrays. Subsequent illumination of the crystal with X-rays leads to diffraction patterns that can be analysed to deduce the arrangements of atoms in the repeating unit of the crystal[4].

The intricacies of biological macromolecules have also driven the adoption and refinement of nuclear magnetic resonance (NMR) spectroscopy for biomolecular studies. In NMR, the magnetic properties of atomic nuclei are exploited to obtain information about their local environments, which reveals conformational dynamics, and in the case of small proteins also the three-dimensional structure of the macromolecule. Unlike X-ray crystallography, which requires crystalline samples, NMR can be applied to molecules in solution, thereby circumventing the bottleneck of crystallisation and allowing for the study of structures in a more native-like state[5].

A third technique bypasses the need for crystallization as well: cryogenic electron microscopy (cryo-EM). Conceptually more straightforward, cryo-EM allows for the direct imaging of macromolecules through an electron microscope, exploiting the extremely short picometer-scale wavelength of highly accelerated electrons. You literally "just look at the thing". The next chapter will focus on the developments and capabilities of this technique in structural biology.

ATGAGCATTCAGAAAAGCACCAACAGCAGCAGCCTGGCGGAAGTGATTGATCGCATTCTGGATAAAG
GCATTGTGATTGATGCGTTTGCGCGCGTGAGCGTGGTGGGCATTGAAATTCTGACCATTGAAGCGCG
CGTGGTGATTGCGAGCGTGGATACCTGGCTGCGCTATGCGGAAGCGGTGGGCCTGCTGCGCGATGAT
GTGGAAGAAAACGGCCTGCCGGAACGCAGCAACAGCAGCGAAGGCCAGCCGCGCTTTAGCATTTAG
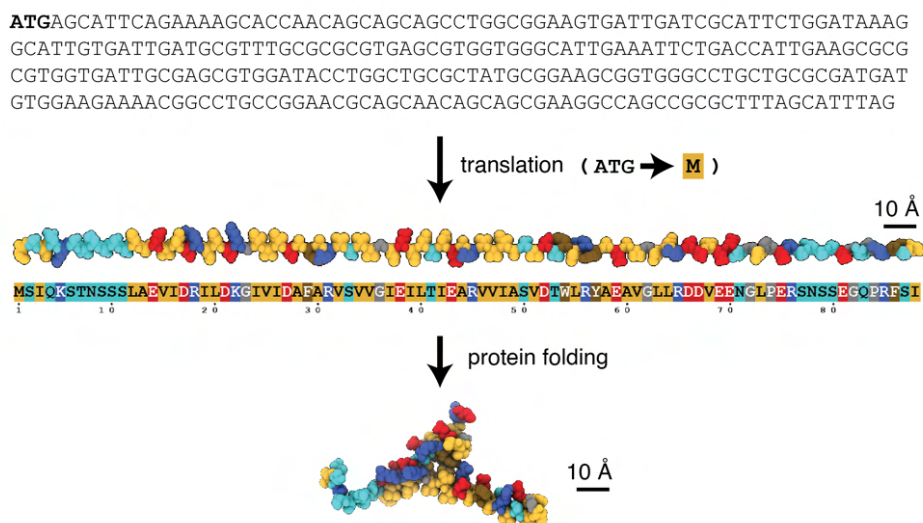
Figure 1.1: | **A mechanistic view of molecular biology.** In the genome, information is stored as a linear chain of the four nucleotides A, C, G and T. Genomic information gets 'translated' by biological systems into proteins, linear molecules made from 20 different amino acids with different properties (depicted as colors). The chains fold into three-dimensional shapes - and act as tiny molecular machines that perform a wide variety of tasks. Understanding the relationship between protein sequence, structure, and function is a main goal of structural biology.

1

1.2 The Emergence of Cryogenic Electron Microscopy

In his seminal 1959 lecture on nanotechnology, Richard Feynman anticipated the potential of electron microscopy in advancing biological research.

> "It is very easy to answer [...] fundamental biological questions; you just look at the thing! You will see the order of bases in the chain; you will see the structure of the microsome. Unfortunately, the present microscope sees at a scale which is just a bit too crude. Make the microscope one hundred times more powerful, and many problems of biology would be made very much easier.
>
> *Richard Feynman, "There is plenty of room on the bottom" 1959*[6]

Unfortunately, applying this suggestion turned out more difficult. Life "operates" in aqueous environments at temperatures where water remains liquid. Electron microscopes, on the other hand, require a vacuum to function effectively, as electrons have a very short mean-free path in air at atmospheric pressure. Introducing a liquid protein sample into such a vacuum would lead to boiling. Sample freezing techniques for use in cryogenic electron microscopy (cryo-EM) were developed as a solution to this problem. Originating from the seminal work of Jacques Dubochet and Alasdair McDowall in the 1980s[7], the technique involves rapidly immersing biological samples in a cryogen to create a thin layer of vitreous (glass-like) ice. This method preserves the near-native, hydrated state of the protein sample, while rendering it solid - compatible with the vacuum conditions in the electron microscope.

Cryo-EM addresses an additional problem not highlighted in Feynman's presentation: sample damage from the high-energy electrons typically used in transmission electron microscopy (TEM). Keeping the sample at cryogenic temperatures during imaging partially mitigates this effect, as damage spreads less at low temperatures. Nonetheless, even at low temperatures, acquiring a detailed 3D structure of a single biomolecule remains unattainable[8]. The molecule deteriorates before a sufficient signal is collected. To overcome this limitation, datasets ranging from thousands to hundreds of thousands of individual particle images are typically aligned and averaged, thereby achieving the signal-to-noise ratio necessary for detailed 3D structure determination.

Advances in cryo-EM technology during the 2010s, such as the introduction of direct electron detectors and new image processing algorithms, have significantly expanded the technique's capabilities[9]. These innovations have improved resolutions from the nanometer scale to a typical range of 2-4 Å, which is sufficiently detailed to see the fold of protein chains and positions of side-chains, allowing detailed chemical interpretation. For samples with especially favorable characteristics such as high symmetry, stability and large size, even atomic resolutions as high as 1.2 Å have been achieved[10,11], allowing atoms to be resolved as individual spheres.

1.3 Current approaches for cryo-EM sample preparation

Despite these significant advancements, cryo-EM still faces considerable challenges, the most prominent being reliable sample preparation. The technique's rising popularity in recent years could create unrealistic expectations for researchers outside the field. While some may resolve a high-resolution structure with cryo-EM on their first attempt, sample optimization is more typically a months-long process with unclear outcome. Proteins have evolved to function in a specific bulk liquid environment, surrounded by other biological molecules in a defined chemical environment. Standard cryo-EM sample preparation places proteins in a 10-100 nanometer thin liquid layer surrounded by an air-water interface, a far from native environment for a protein. Within this layer, ideally not much thicker than the particle itself, the molecules must adopt random orientations to allow 3D reconstruction of their structure from 2D images. Often the orientation is not random and viewing directions are missing, leading to artefacts in the 3D reconstruction.

The prevailing method for making these thin specimens suitable for cryo-EM has not changed substantially since its introduction several decades ago: a filter paper is pressed against one or both sides of a fenestrated support film spanning a metallic mesh grid across which a droplet of the biological sample is suspended, and then rapidly plunged into a liquid cryogen[12]. A recognised shortcoming of this method is that a reproducibly thin and consistently uniform ice layer can often not be achieved[13,14]. Frequently, time-consuming parameter optimisation is required to find conditions suitable for high-resolution imaging, and these conditions are typically different for each sample. Even if such conditions are found, additional problems encountered in practice include preferential orientation of particles[15,16], preferential interaction of particles with the support film[17] and partial denaturation of particles due to exposure to the large air-water interface[16,18–21].

The current limitations in sample preparation reproducibility have prompted extensive efforts to innovate production methods for thin ice layers. These include the development of self-wicking grids[22], spraying small-volume droplets onto holey support films[23–26] and direct thin film dispensing onto cryo-EM grids[27–29]. These developments have increased the speed of sample preparation and reduced the required sample volume by orders of magnitude. Perhaps surprisingly though, the core design of fenestrated EM grid supports has remained largely unchanged. Despite the accelerated sequence from sample deposition to vitrification, the exposure of particles to the air-water interface inherent to grid-based sample supports continues to pose a major obstacle for robust sample preparation workflows[30,31].

In Chapter 2 we introduce an alternative sample type to cryo-EM sample preparation that mitigates several existing challenges. Rather than generating a thin liquid layer with an undesirable air-water interface, our method utilizes nanofluidic channels enveloped by hydrophilic silicon nitride membranes. These membranes serve as a substitute for the problematic air-water interface. Our proof-of-principle study demonstrates that this new sample support enables protein imaging and the reconstruction of their three-dimensional structures with resolutions as high as 3 Å.

1

1.4 Time-resolved cryo-EM can capture short-lived states

Single-particle cryo-EM enables the high-resolution analysis of protein structures immobilized in thin layers of ice. During sample preparation, molecular conformations are rapidly arrested by freezing, which occurs within less than a millisecond[7]. This rapid process provides an opportunity to capture and analyze transient conformational states, just after a reaction has been initiated. Traditional sample preparation involves pipetting liquids onto an electron microscopy grid followed by excess liquid removal through blotting, a process that takes several seconds and thus limits the ability to trap short-lived states. Time-resolved cryo-EM (trEM) encompasses technologies for rapid sample preparation, permitting the arrest of molecular reactions with a specified time delay.

One method in trEM employs light flashes to initiate reactions. Several studies have developed systems that combine the illumination of a cryo-electron microscopy (cryo-EM) grid with rapid plunge-freezing into liquid ethane, typically within approximately 20 milliseconds[32–34]. A straightforward target of this method are proteins that are naturally sensitive to light, such as bacteriorhodopsin, a photoreactive proton pump found in Archaea. The protein undergoes conformational changes upon light absorption, which was investigated at a resolution of 3-4 Å using electron crystallography[35].

In addition to naturally light-sensitive proteins, flash photolysis can be used for trEM studies of chemically induced reactions. This is achieved by employing 'caged' compounds, where 'cage' refers to a light-sensitive chemical group, such as 2-nitrophenyl-eythyl (NPE)[36]. Upon photon absorption, the light-sensitive group is released, thereby activating the effector molecule, such as ATP, into its functional form. An application of this method was demonstrated by Menetret et al.[33] in their study of the temporal dynamics in the interaction between actin filaments and myosin following the release of ATP from a caged ATP compound.

Flash photolysis can also be combined with a novel technique designed to melt ice within cryo-EM samples. This method utilizes a laser beam to locally melt the sample, which then rapidly revitrifies within a few microseconds. Studies have confirmed that this rapid melting and revitrification process does not compromise the integrity of proteins[37]. Integrating this technique with a caged acidic compound, researchers were able to investigate the dynamics of a viral capsid under pH change, achieving microsecond time resolution in their observations[38].

In addition to flash photolysis, reactions can be started by rapidly mixing two liquid compounds together. This approach has been implemented by applying the first liquid by pipetting, and the second by spraying it onto the grid just before plunge-freezing[39]. This method was used to structurally investigate the transiently open-channel form of acetylcholine receptors upon binding of acetylcholine[40]. Direct mixing of two independent sprays on the grid is also possible and has been shown for the Spotiton setup[41], which was applied to several biologcal systems, achieving a temporal resolution of 100 milliseconds.

Direct mixing of liquids on the grid in cryo-electron microscopy represents a complex and relatively uncontrolled procedure. While multiple studies showed that the liquids do mix, the exact kinetics are poorly understood, leaving uncertainties about the exact time the reaction is ongoing. To alleviate this problem, several studies designed setups where

mixing occurs *before* applying the samples to the grid[25,26,42–45], which allows for better control in mixing, and also enables mixing of large molecules with low diffusion constants.

All these methods, flash-photolysis, spray-and-mix and mix-and-spray have advantages and disadvanges. Flash-photolysis necessitates caged compounds, which precludes it from studying protein-protein interactions. Spray-and-mix approaches have uncertainty about how long and how well the liquids are mixed, and have difficulties with large molecules. Mix-and-spray methods need technically difficult microfluidics setups and often require a large sample volume. Inherent to spraying methods is the difficulty to achieve controlled, homogenous and thin ice layers. Even though research on trEM has been ongoing since the 1990's, none of the approaches have so far become a widely adopted standard in the cryo-EM field.

In Chapter 3, we design and build a time-resolved setup for our research group, tailored to specific requirements in the group. Our central biological focus is the cellular defense mechanism mediated by guanylate-binding proteins (GBPs)[46]. These are GTPases featuring a GTP-binding domain and an effector domain, the conformation of which varies depending on the bound nucleotide. Employing time-resolved cryo-EM (trEM) for the controlled release of GTP and subsequent arrest of the reaction at known delay times can yield critical insights into the dynamics and transient states of this system. Given the availability of caged GTP compounds, and the relative simplicity of blotting and flash-photolysis compared to more technically involved methods, we have chosen to build such a setup for time-resolved sample preparation. Chapter 3 describes the design, manufacturing and preliminary testing of such a setup.

1.5 Structural biology of microbial gas vesicles: Historical milestones

In this thesis, cryo-electron microscopy (cryo-EM) was employed for the structural study of gas vesicles - a protein-based organelle enabling buoyancy control in a diverse range of bacteria and archaea. Gas vesicles exhibit an unusual structure - they form cylindrical gas-filled shells with a mass on the order of hundreds of megadaltons, predominantly composed of subunits approximately 7 kDa in size. In recent decades, scientists have used numerous research approaches to investigate gas vesicle structure. The following section presents a brief review over those studies and also serves as a historic overview of techniques that were available at the time.

1.5.1 Pioneering Structural Studies on Gas Vesicles

Gas vesicles are intracellular, gas-filled nanostructures that allow microbes to float vertically to the surface of their aqueous habitat. Gas vesicles are unique among prokaryotic motility systems in that their structure alone imparts the function of vertical motility. Aquatic cyanobacteria utilise this motility to optimise light harvesting conditions for photosynthesis by rising to the surface. Massive accumulation of such cyanobacterial colonies can be observed in freshwater lakes, colloquially referred to as algal blooms. The presence of gas vesicles reduces the density of cells and prevents them from sinking; about three to ten per cent of the cell volume must be taken up by gas vesicles for a cell to float to the surface[47]. While the function of gas vesicles is straightforward — enabling buoyancy — the molecular

1
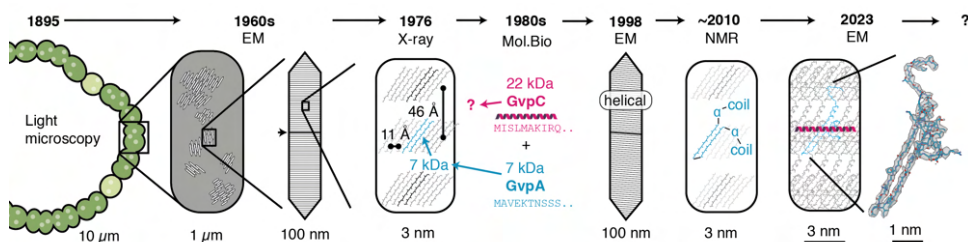
Figure 1.2: | **Historical milestones of investigating gas vesicle structure.** This timeline charts the development from 19th century light microscopy to atomic structure determination. It highlights key findings and adopts an open-ended perspective, drawing attention to persisting unresolved structural questions discussed in the Outlook (Chapter 7) of this thesis.

architecture required to fulfill this role is complex and intriguing. The structural complexity of these organelles poses a unique scientific challenge. This inherent complexity has fueled multiple waves of research employing a diverse set of methods. Each research phase has elucidated structural and functional aspects of gas vesicles, yet each has also raised new questions warranting further study [Figure 1.2].

Bright, refractile bodies in cyanobacteria were first described in 1895 using light microscopy, and identified as crucial components imparting buoyancy in these cells[48]. The gas-filled nature of these "Gasvacuolen" (engl. gas vacuoles) was established[49] and their sensitivity to pressure was shown experimentally[50]. These early studies established the function of these intracellular gas spaces. Yet, their sub-micrometer dimensions hindered detailed structural analysis with the existing methodologies.

How can a cell produce and maintain a gas-filled structure? The increased availability of commercial electron microscopes in the 1960s[51] enabled the first detailed ultrastructural examination of gas vacuoles. The gas turned out to be contained in cylindrical objects, initally called "Hohlspindeln" (engl. hollow spindles)[52]. Other descriptions used gas cylinders, gas vacuoles or gas vesicles, with the latter becoming the most widely adopted term today. Utilizing various electron microscopy techniques, studies such as those by Jost[52], Bowen and Jensen[53], Smith and Peat[54], and Walsby and Eichelberger[55], provided comprehensive insights into the overall morphology of gas vesicles. These studies established important ultrastructural features of gas vesicles: their cylinder structure with diameters of 65-75 nm, the packing of gas vesicles into hexagonal clusters inside cells, gas-filled central lumens, surface ribs spaced 4-5 nm apart, lengths of around 200 to 1000 nm with conical ends and wall thicknesses of around 2 nm.

What molecular building blocks are gas vesicles composed of? Purification techniques are crucial for chemical analysis to determine the composition of gas vesicles. Walsby and Buckland[56] developed methods to isolate gas vesicles from *Anabaena flos-aquae*, employing a 0.7 M sucrose solution for cell lysis and controlled centrifugation. Purified gas vesicles also allowed to measure pressure-collapse curves for assessing the vesicles' stability and determining maximum safe centrifugation speeds. Their research, alongside a similar study on *Microcystis aeruginosa*[57], demonstrated that purified gas vesicles are exclusively made from protein, devoid of other material such as lipids or carotenoids. However, the concept

of proteins forming extended gas-filled structures akin to balloons is exceptionally unusual and intriguing.

How can these structures self-assemble inside the cell? Waaland and Branton[58] provided key observations on the biogenesis of gas vesicles in 1969, noting a distinct central rib in the vesicle wall and the growth pattern of these vesicles over time. Gas vesicles appeared to start growing as small biconical structures, and grow from the middle outwards by elongating the cylindrical center. Walsby[59] suggested that this central rib results from the union of two identical halves of the gas vesicle, with unique molecular interactions occurring at this juncture.

What is the structural unit added during gas vesicle growth? Employing fiber diffraction, a technique famously associated with revealing the structure of DNA in the 1950s[60], Blaurock and Walsby[61] in 1976 analyzed gas vesicles from *Anabaena flos-aquae*. Their findings indicated a crystalline wall structure comprising 11 Å wide subunits, each 46 Å high and 20 Å thick, arranged along the ribs. The molecular weight was estimated at approximately 8 kDa. Diffractograms showed strong 4.7 Å peaks, consistent with the canoncical distance of neighbouring beta strands in a bet sheet assembly. Along with angular orientation and the overall subunit dimensions deduced from the fiber diffraction patterns, this led to the suggestion that gas vesicles are formed by tilted beta sheets consisting of staggered, horizontally aligned beta-hairpins. The principal structural characteristics of the essential subunit being established called for the identification of the protein which can form this unique quasi-crystalline assembly.

1.5.2 The molecular constituents GvpA and GvpC

Partial sequences of the gas vesicle wall protein were identified in the 1980s through analysis of tryptic digests of purified gas vesicles[62,63] from *Anabaena flos-aquae*. However, full sequence determination was hampered by its poor solubility in SDS, complicating SDS-PAGE separation. An observed recurring SDS-PAGE band indicated a molecular weight of 20.6 kDa[63], inconsistent with the 8 kDa estimated from X-ray diffraction. In parallel to Walsby's efforts and based on the partial polypeptide sequences derived from this work, de Marsac[64] cloned a gas vesicle gene from *Calothrix PCC7601*, a cyanobacterium highly similar to *Anabaena flos-aquae*, which revealed the complete genetic sequence of GvpA in 1985. Protein sequencing from *Anabaena flos-aquae* gas vesicles later corroborated this, identifying a 70-residue protein with a mass of 7.3 kDa[65], which matched the earlier crystallographic data[61].

The ~20 kDa band on SDS-PAGE gels from gas vesicles, originally hypothesized to be a GvpA trimer[65] was later attributed to a second structural gas vesicle protein, GvpC, with a molecular weight of 22 kDa[66,67]. Detergent treatments can remove GvpC and result in reduced gas vesicle stability to externally applied pressure, which suggested a vital role for GvpC in reinforcing the wall[68]. Based on the ratio of GvpA to GvpC which was experimentally determined to be 25:1[69], various binding geometries have been proposed in which GvpC either aligns along or spans across the gas vesicle's ribs.

While genetic and biochemical studies on GvpA and GvpC clarified their sequence, identity, and relative abundance in purified gas vesicles, they raised new questions about their fold and 3D structure. In view of the geometrical restraints deduced from the X-ray diffraction studies on the asymmetric unit, GvpA's primary sequence pattern is inconsistent

1

with GvpA consisting entirely of beta strands[65] and hinted at a more complex three-dimensional fold. Using Atomic Force Microscopy (AFM), McMaster et al.[70] achieved detailed imaging of the wall ultrastructure in collapsed gas vesicles in 1996. Their estimates of 46.4 nm ribs and 11.2 Å periodicities along the rib formed by GvpA subunits, and beta strands spaced 4.7 Å apart were consistent those of the earlier fiber diffraction patterns. Intriguingly, AFM revealed interruptions between the gas vesicle ribs, suggesting a more complex fold for GvpA beyond simple beta strands leaving important questions how succesive ribs are held together.

In the 1990s, a more macroscopical structural question about gas vesicles still persisted: are they helical or not? Since the 1960s, gas vesicles had been described as ribbed structures with a pitch ranging from 4.5 to 5 nm. Two competing models, stacked hoops and a continuous helix, were proposed to explain these ribs[47,59]. Offner et al.[71] provided conclusive evidence for the helical model in 1998. Deviating from previous sample preparation techniques, they used frozen hydrated samples of gas vesicles in electron cryomicroscopy (cryo-EM). This approach keeps gas vesicle in a much more native state, unlike for example the structural flattening in dried, metal-stained samples. The analysis of moiré patterns, created by the interference between the gas vesicle's front and back ribs, led to the definitive conclusion that the vesicle wall is formed by a continuous helical rib, not stacked hoops.

1.5.3 New techniques in gas vesicle studies

The quest to understand the structure of GvpA, the main protein in the gas vesicle wall, advanced significantly in the 2010s with the aid of solid-state NMR. This technique, distinct in its underlying physical principles from X-ray diffraction and electron microscopy, played a crucial role in revealing GvpA's definite secondary structure. Resonance assignment of 81% of the GvpA sequence[72] uncovered a coil-α-β-β-α-coil pattern. This pattern, with 27 out of 71 residues forming the central beta-hairpin flanked by alpha helices and coils, corroborated prior hypotheses based on sequence analysis[65] and atomic force microscopy (AFM) data[70]. Additionally, the NMR studies proposed the existence of inequivalent GvpA monomers in the gas vesicle's structure[73,74], suggesting these monomers were arranged in antiparallel orientations with different folds within the ribs. This hypothesis also influenced computational models of GvpA around 2012[75,76]. However, these antiparallel models were eventually invalidated by more precise structural insights gained from subsequent research.

What was known at this point is the unit cell size, the molecular constituents GvpA and GvpC, and the secondary structure pattern of GvpA. The challenge remained in deducing the precise three-dimensional conformation of the gas vesicle wall which could explain its unique properties. Gas vesicles can neither be crystallized into 3D crystals for X-ray crystallography, nor could their 3D structure be solved by NMR. The 2010s saw dramatic improvements in cryogenic electron microscopy (cryo-EM)[9] with the introduction of better electron detectors, microscopes and algorithms, which led to the 2017 Nobel price in Chemistry for Jacques Dubochet, Joachim Frank, and Richard Henderson. These improvements in cryo-EM have enabled the direct determination of the three-dimensional structures of many proteins from two-dimensional particle images, achieving resolutions sufficient to interpret structural details at the atomic level. Cryo-EM offered a route to finally determine the 3D structure of the gas vesicle wall by direct imaging.

Pioneering efforts to elucidate the gas vesicle structure using cryo-EM techniques were

1

undertaken by Bollschweiler[77] in 2015. This study focused on a I34M GvpA mutant from *Halobacterium salinarum*, known for producing thinner and more homogenous gas vesicles compared to the wild-type[76]. A key observation was the high sensitivity of gas vesicles to electron exposure, with structural degradation occurring even under very low fluence. To mitigate this, tilt series were acquired using only a fraction of the typical electron exposure, allowing the reconstruction of the wall structure at an approximate resolution of ~9 Å. The analysis revealed hook-like features in the wall, with a reversal in orientation at the gas vesicle's central rib, while maintaining consistent helical handedness. However, this resolution was not sufficient to discern the secondary structural elements of GvpA, specifically the coil-α-β-β-α-coil pattern, thus leaving the precise fold of GvpA unresolved.

In a more recent 2023 study, Dutka et al.[78] focused on the gas vesicles of *Anabaena flos-aquae*, noted for their smaller size and greater stability compared to those in *H. salinarum*. Advances in cryo-EM technology over the preceding eight years likely contributed to the acquisition of higher quality data. To counteract these challenges, cryo tomograms were captured at lower than usual doses, facilitating wall reconstructions at an approximate resolution of 8 Å. This resolution allowed for the identification of a band of density along the rib, interpreted as the polymerizing beta strands of GvpA monomers. These beta strands contributed to a corrugated pattern in the wall's cross-section, aiding gas vesicle stability. Moreover, the study provided insights into the ultrastructure of GvpC binding to the gas vesicle's exterior, establishing its alignment along the ribs.

Both cryo-EM studies highlighted a common challenge in resolving molecular details of gas vesicles: resolution along the gas vesicle ribs was much poorer compared to other spatial directions. This difficulty stems from the lack of large scale features along the gas vesicle rib, which are crucial for aiding particle alignment and averaging in cryo-EM. The rib structure is primarily characterized by a 4.7 Å spacing between beta strands of GvpA and, to a lesser extent, by a 12 Å spacing between the GvpA monomers. At resolutions below 12 Å, no distinctive features along the ribs are discernible. Alignment of these tiny 4.7 and 12 Å structural features is particularly challenging under conditions of low signal-to-noise ratio (SNR) caused by the limitations on tolerable electron exposure to mitigate radiation damage[77] and gas vesicle shrinking[78], imperfect tilt series alignment and increased beam pathlength through the ice at high tilts, and the presence of thick ice layers in the sample. These conditions impede resolving atomic-level structural features, leading to anisotropic resolution with much less detail along the gas vesicle ribs.

To mitigate these difficulties, we eliminated tilt-series acquisition (a tomography approach) in favor of 2D imaging, which allows for reduced electron exposure to prevent radiation damage. We prepared samples of gas vesicles in ultra-thin monolayers to reduce background noise and to maximise SNR. Not employing tomography also avoids quality loss from imperfect tomogram reconstruction and imaging through long path lengths of ice at high tilts. The 2D cryo-EM images obtained, coupled with helical reconstruction techniques[79], enabled atomic structure determination of the gas vesicle wall for the first time, which is described in detail in Chapter 4.

A COLOR SCHEME FOR GAS VESICLE PROTEINS

The gas vesicle protein GvpA is characterized by a distinct amino acid distribution, featuring a predominantly hydrophobic side and a hydrophilic N-terminus. While protein sequences
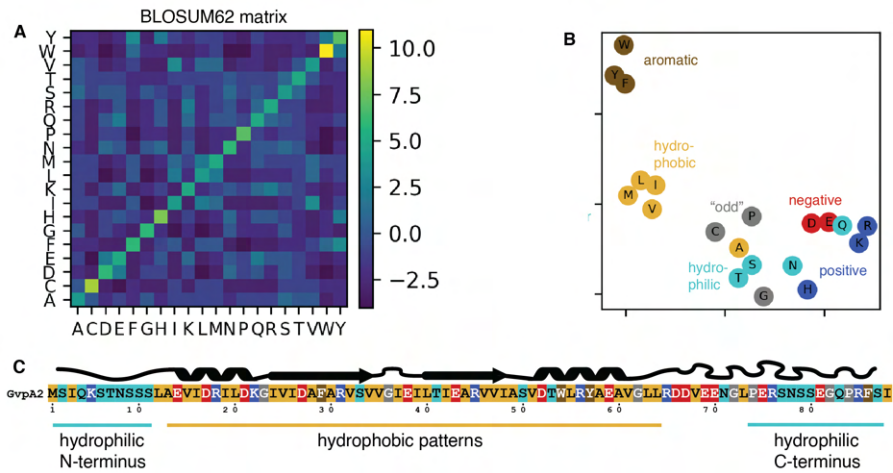
Figure 1.3: | **Proposed colorscheme for protein residue coloring.** (A) BLOSUM62 matrix[80] quantifies evolutionary conservation for each amino acid substitution. Using a column in this matrix as a position of each amino acid in high-dimensional space, (B) shows a 2D t-SNE[81] embedding of this space to give an impression of amino acid similarity. Assigned categories and colors roughly fall into close-by clusterings.

are inherently information-dense, they often pose challenges for intuitive interpretation. In this thesis, a tailored color scheme is employed to emphasize amino acid characteristics relevant for gas vesicle proteins, which is helpful in enriching multiple sequence alignments with additional intuitively understandable information. The specifics of this color scheme are outlined in Table 1.1.

Table 1.1: Proposed color scheme for amino acids in gas vesicle proteins

Amino acids	Category	One letter codes	Color	HEX code
Methionine, isoleucine, leucine, alanine, valine	hydrophobic	M,I,L,A,V	yellow	#e1b13e
Serine, threonine, asparagine, glutamine	hydrophilic	S,T,N,Q	turquoise	#49c2c6
Aspartic acid, glutamic acid	negative	D,E	red	#cb2026
Lysine, arginine, histidine	positive	K,R,H	blue	#3e58a8
Phenylalanine, tyrosine, tryptophane	aromatic	F,Y,W	brown	#715321
Glycine, proline, cysteine	"odd"	G,P,C	gray	#7b7b7b

A BLOSUM62 similarity matrix[82] [Figure 1.3A] was utilized to quantify the similarity among amino acids and to check whether the devised color scheme effectively groups similar amino acids. By representing each amino acid as a position vector in the BLOSUM62 matrix and applying dimensionality reduction through t-SNE[81], amino acids are broadly

positioned according to their assigned properties [Figure 1.3B]. This categorization is especially evident for aromatic and hydrophobic amino acids. However, other amino acids could be alternatively classified based on different criteria, such as side chain volume, and are less well clustered. The application of this color scheme to the GvpA sequence from gas vesicles highlights specific features, like the hydrophilic N-terminus and the hydrophobic core [Figure 1.3C].

1

1

1.6 Outline of the thesis

The following thesis chapters contain work on both technical aspects of cryo-EM as well as applications to biological problems. The thesis starts with two chapters on cryo-EM sample preparation techniques. **Chapter 2** introduces a new sample support called a CryoChip that contains the protein in nanofluidic channels, which addresses several problems with current techniques. **Chapter 3** outlines the design of a setup for time-resolved EM studies using flash photolysis, which can be applied to proteins studied in our research group. The subsequent chapters turn towards biological questions, presenting the atomic structure of gas vesicles (**Chapter 4**) solved by cryo-EM. Based on this structure, a study of the gas vesicle protein GvpC is presented, revealing its binding geometry and surface (**Chapter 5**). The next chapter presents a creative application of data from the other chapters: the creation of 3D-printed educational models, derived from structural data accumulated throughout the thesis (**Chapter 6**). The thesis finishes with conclusions and an outlook (**Chapter 7**) to put the work into perspective and discuss open challenges.

1.7 References

1. Huber, S. T., Sarajlic, E., Huijink, R., Weis, F., Evers, W. H. & Jakobi, A. J. Nanofluidic chips for cryo-EM structure determination from picoliter sample volumes. *Elife* **11,** e72629 (2022).

2. Huber, S. T., Terwiel, D., Evers, W. H., Maresca, D. & Jakobi, A. J. Cryo-EM structure of gas vesicles for buoyancy-controlled motility. *Cell* **186,** 975–986 (2023).

3. Nevers, Y., Glover, N. M., Dessimoz, C. & Lecompte, O. Protein length distribution is remarkably uniform across the tree of life. *Genome Biology* **24,** 135 (2023).

4. Rupp, B. *Biomolecular crystallography: principles, practice, and application to structural biology* (Garland Science, 2009).

5. Markley, J. L. View from nuclear magnetic resonance spectroscopy. *Integrative Structural Biology with Hybrid Methods,* 19–22 (2018).

6. Feynman, R. P. *Plenty of Room at the Bottom* in *APS annual meeting* (1959), 1–7.

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

8. R., H. The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules. *Q. Rev. Biophys* **28,** 171–193 (1995).

9. Callaway, E. The Revolution Will Not Be Crystallized. *Nature* **525,** 172–174. ISSN: 0163-6545 (2015).

10. Yip, K. M., Fischer, N., Paknia, E., Chari, A. & Stark, H. Atomic-resolution protein structure determination by cryo-EM. *Nature* **587,** 157–161 (2020).

11. Nakane, T., Kotecha, A., Sente, A., McMullan, G., Masiulis, S., Brown, P. M., Grigoras, I. T., Malinauskaite, L., Malinauskas, T., Miehling, J., *et al.* Single-particle cryo-EM at atomic resolution. *Nature* **587,** 152–156 (2020).

12. Adrian, M., Dubochet, J., Lepault, J. & McDowall, A. W. Cryo-electron microscopy of viruses. en. *Nature* **308,** 32–36 (1984).

13. Carragher, B., Cheng, Y., Frost, A., Glaeser, R., Lander, G., Nogales, E. & Wang, H. Current outcomes when optimizing 'standard' sample preparation for single-particle cryo-EM. *J. Microsc.* **276,** 39–45 (2019).

14. Armstrong, M., Han, B.-G., Gomez, S., Turner, J., Fletcher, D. A. & Glaeser, R. M. Microscale Fluid Behavior during Cryo-EM Sample Blotting. en. *Biophys. J.* **118,** 708–719 (2020).

15. Tan, Y. Z., Baldwin, P. R., Davis, J. H., Williamson, J. R., Potter, C. S., Carragher, B. & Lyumkis, D. Addressing preferred specimen orientation in single-particle cryo-EM through tilting. en. *Nat. Methods* **14,** 793–796 (2017).

16. Noble, A. J., Wei, H., Dandey, V. P., Zhang, Z., Tan, Y. Z., Potter, C. S. & Carragher, B. Reducing effects of particle adsorption to the air–water interface in cryo-EM. *Nat. Methods* **15,** 793–795 (2018).

1

17. Meyerson, J. R., Rao, P., Kumar, J., Chittori, S., Banerjee, S., Pierson, J., Mayer, M. L. & Subramaniam, S. Self-assembled monolayers improve protein distribution on holey carbon cryo-EM supports. en. *Sci. Rep.* **4,** 7084 (2014).

18. Donaldson, T. L., Boonstra, E. F. & Hammond, J. M. Kinetics of protein denaturation at gas—liquid interfaces. *J. Colloid Interface Sci.* **74,** 441–450 (1980).

19. Glaeser, R. M. & Han, B.-G. Opinion: hazards faced by macromolecules when confined to thin aqueous films. *Biophys. Rep.* **3,** 1–7 (2017).

20. Glaeser, R. M. Proteins, Interfaces, and cryo-EM grids. en. *Curr. Opin. Colloid Interface Sci.* **34,** 1–8 (2018).

21. D'Imprima, E., Floris, D., Joppe, M., Sánchez, R., Grininger, M. & Kühlbrandt, W. Protein denaturation at the air-water interface and how to prevent it. en. *eLife* **8** (2019).

22. Wei, H., Dandey, V. P., Zhang, Z., Raczkowski, A., Rice, W. J., Carragher, B. & Potter, C. S. Optimizing "self-wicking" nanowire grids. en. *J. Struct. Biol.* **202,** 170–174 (2018).

23. Feng, X., Fu, Z., Kaledhonkar, S., Jia, Y., Shah, B., Jin, A., Liu, Z., Sun, M., Chen, B., Grassucci, R. A., Ren, Y., Jiang, H., Frank, J. & Lin, Q. A Fast and Effective Microfluidic Spraying-Plunging Method for High-Resolution Single-Particle Cryo-EM. *Structure* **25,** 663–670.e3 (2017).

24. Rubinstein, J. L., Guo, H., Ripstein, Z. A., Haydaroglu, A., Au, A., Yip, C. M., Di Trani, J. M., Benlekbir, S. & Kwok, T. Shake-it-off: a simple ultrasonic cryo-EM specimen-preparation device. en. *Acta Crystallogr. D* **75,** 1063–1070 (2019).

25. Kontziampasis, D., Klebl, D. P., Iadanza, M. G., Scarff, C. A., Kopf, F., Sobott, F., Monteiro, D. C. F., Trebbin, M., Muench, S. P. & White, H. D. A cryo-EM grid preparation device for time-resolved structural studies. *IUCrJ* **6,** 1024–1031 (2019).

26. Torino, S., Dhurandhar, M., Stroobants, A., Claessens, R. & Efremov, R. G. Time-resolved cryo-EM using a combination of droplet microfluidics with on-demand jetting. en. *Nature Methods* **20,** 1400–1408. ISSN: 1548-7091, 1548-7105 (Sept. 2023).

27. Jain, T., Sheehan, P., Crum, J., Carragher, B. & Potter, C. S. Spotiton: A prototype for an integrated inkjet dispense and vitrification system for cryo-TEM. *J. Struct. Biol.* **179,** 68–75 (2012).

28. Arnold, S. A., Albiez, S., Bieri, A., Syntychaki, A., Adaixo, R., McLeod, R. A., Goldie, K. N., Stahlberg, H. & Braun, T. Blotting-free and lossless cryo-electron microscopy grid preparation from nanoliter-sized protein samples and single-cell extracts. *J. Struct. Bio.* **197,** 220–226 (2017).

29. Ravelli, R. B. G., Nijpels, F. J. T., Henderikx, R. J. M., Weissenberger, G., Thewessem, S., Gijsbers, A., Beulen, B. W. A. M. M., López-Iglesias, C. & Peters, P. J. Cryo-EM structures from sub-nl volumes using pin-printing and jet vitrification. en. *Nat. Commun.* **11,** 2563 (2020).

30. Klebl, D. P., Gravett, M. S., Kontziampasis, D., Wright, D. J., Bon, R. S., Monteiro, D. C., Trebbin, M., Sobott, F., White, H. D., Darrow, M. C., Thompson, R. F. & Muench, S. P. Need for Speed: Examining Protein Behavior during CryoEM Grid Preparation at Different Timescales. *Structure* **28,** 1238–1248.e4 (2020).

31. Glaeser, R. M. Preparing Better Samples for Cryo–Electron Microscopy: Biochemical Challenges Do not End with Isolation and Purification. *Annu. Rev. Biochem.* **90** (2021).

32. Subramaniam, S., Gerstein, M., Oesterhelt, D. & Henderson, R. Electron diffraction analysis of structural changes in the photocycle of bacteriorhodopsin. en. *The EMBO Journal* **12,** 1–8. ISSN: 02614189 (Jan. 1993).

33. Ménétret, J.-F., Hofmann, W., Schröder, R., Rapp, G. & Goody, R. Time-resolved cryo-electron microscopic study of the dissociation of actomyosin induced by photolysis of photolabile nucleotides. en. *Journal of Molecular Biology* **219,** 139–144. ISSN: 00222836 (May 1991).

34. Shaikh, T. R., Barnard, D., Meng, X. & Wagenknecht, T. Implementation of a flash-photolysis system for time-resolved cryo-electron microscopy. *Journal of Structural Biology* **165,** 184–189. ISSN: 10478477 (Mar. 2009).

35. Subramaniam, S. & Henderson, R. Molecular mechanism of vectorial proton transloca-tion by bacteriorhodopsin. en. *Nature* **406,** 653–657. ISSN: 0028-0836, 1476-4687 (Aug. 2000).

36. McCray, J. A. Properties and Uses of Photoreactive Caged Compounds. en. *Annu. Rev. Biophys. Biophys. Chem.* (1989).

37. Bongiovanni, G., Harder, O. F., Voss, J. M., Drabbels, M. & Lorenz, U. J. Near-atomic resolution reconstructions from *in situ* revitrified cryo samples. *Acta Crystallographica Section D Structural Biology* **79,** 473–478. ISSN: 2059-7983 (June 2023).

38. Harder, O. F., Barrass, S. V., Drabbels, M. & Lorenz, U. J. Fast viral dynamics revealed by microsecond time-resolved cryo-EM. en. *Nature Communications* **14,** 5649. ISSN: 2041-1723 (Sept. 2023).

39. Berriman, J. & Unwin, N. Analysis of transient structures by cryo-microscopy com-bined with rapid mixing of spray droplets. en. *Ultramicroscopy* **56,** 241–252. ISSN: 03043991 (Dec. 1994).

40. Unwin, N. Acetylcholine receptor channel imaged in the open state. en. *Nature* **373,** 37–43. ISSN: 0028-0836, 1476-4687 (Jan. 1995).

41. Dandey, V., Budell, W., Wei, H., Bobe, D., Maruthi, K., Kopylov, M., Eng, E., Kahn, P., Hinshaw, J., Kundu, N., Nimigean, C., Fan, C., Sukomon, N., Darst, S., Saecker, R., Chen, J., Malone, B., Potter, C. & Carragher, B. Time-resolved cryoEM using Spotiton. *Nature Methods.* Publisher: Springer US. ISSN: 1548-7105 (2020).

42. Lu, Z., Shaikh, T. R., Barnard, D., Meng, X., Mohamed, H., Yassin, A., Mannella, C. A., Agrawal, R. K., Lu, T.-M. & Wagenknecht, T. Monolithic microfluidic mixing–spraying devices for time-resolved cryo-electron microscopy. en. *Journal of Structural Biology* **168,** 388–395. ISSN: 10478477 (Dec. 2009).

43. Shaikh, T. R., Yassin, A. S., Lu, Z., Barnard, D., Meng, X., Lu, T.-M., Wagenknecht, T. & Agrawal, R. K. Initial bridges between two ribosomal subunits are formed within 9.4 milliseconds, as studied by time-resolved cryo-EM. en. *Proceedings of the National Academy of Sciences* **111,** 9822–9827. ISSN: 0027-8424, 1091-6490 (July 2014).

1

1

44. Chen, B., Kaledhonkar, S., Sun, M., Shen, B., Lu, Z., Barnard, D., Lu, T.-M., Gonzalez, R. L. & Frank, J. Structural Dynamics of Ribosome Subunit Association Studied by Mixing-Spraying Time-Resolved Cryogenic Electron Microscopy. en. *Structure* **23,** 1097–1105. ISSN: 09692126 (June 2015).

45. Mäeots, M.-E., Lee, B., Nans, A., Jeong, S.-G., Esfahani, M. M. N., Ding, S., Smith, D. J., Lee, C.-S., Lee, S. S., Peter, M. & Enchev, R. I. Modular microfluidics enables kinetic insight from time-resolved cryo-EM. en. *Nature Communications* **11,** 3465. ISSN: 2041-1723 (July 2020).

46. Kuhm, T. I., Gross, L., de Agrela Pinto, C., Huber, S. T., Taisne, C., Giannopoulou, E. A., Pardon, E., Steyaert, J., Tans, S. J. & Jakobi, A. J. Structural basis of membrane targeting and coatomer assembly by human GBP1. *bioRxiv,* 2023–03 (2023).

47. Walsby, A. E. Gas vesicles. *Microbiological Reviews* **58,** 94–144. ISSN: 01460749 (1994).

48. Strodtmann, S. Die Ursache des Schwebvermögens bei den Cyanophyceen. *Biologisches Centralblatt* **15,** 113–115 (1895).

49. Klebahn, H. Gasvakuolen, ein Bestandteil der Zellen der Wasserblütebildenden Phycochromaceen. *Flora (Jena)* **80,** 241 (1895).

50. Ahlborn, F. Über die Wasserblüte Byssusflos-aquae und ihr Verhalten gegen Druck. *Verhandlungen des Naturwissenschaftlichen Vereins in Hamburg* **III,** 25 (1895).

51. Ruska, E. The development of the electron microscope and of electron microscopy. *Reviews of modern physics* **59,** 627 (1987).

52. Jost, M. *Die Ultrastruktur von Oscillatoria rubescens DC* PhD thesis (ETH Zurich, 1965).

53. Bowen, C. & Jensen, T. Blue-green algae: fine structure of the gas vacuoles. *Science* **147,** 1460–1462 (1965).

54. Smith, R. V. & Peat, A. Comparative structure of the gas-vacuoles of blue-green algae. *Archiv für Mikrobiologie* **57,** 111–122 (1967).

55. Walsby, A. & Eichelberger, H. The fine structure of gas-vacuoles released from cells of the blue-green alga Anabaena flos-aquae. *Archiv für Mikrobiologie* **60,** 76–83 (1968).

56. Walsby, A. & Buckland, B. Isolation and purification of intact gas vesicles from a Blue–Green alga. *Nature* **224,** 716–717 (1969).

57. Jones, D. D. & Jost, M. Isolation and Chemical Characterization of Gas-Vacuole Membranes from Microcystis Aeruginosa Kuetz. Emend. Elenkin. en. *Archiv für Mikrobiologie* **70,** 43–64. ISSN: 0302-8933, 1432-072X (1970).

58. Waaland, J. R. & Branton, D. Gas vacuole development in a blue-green alga. *Science* **163,** 1339–1341 (1969).

59. Walsby, A. E. Structure and function of gas vacuoles. *Bacteriological reviews* **36,** 1–32. ISSN: 00053678 (1972).

60. Franklin, R. E. & Gosling, R. G. Molecular configuration in sodium thymonucleate. *Nature* **171,** 740–741 (1953).

61. Blaurock, A. E. & Walsby, A. E. Crystalline structure of the gas vesicle wall from Anabaena flos-aquae. *Journal of Molecular Biology* **105,** 183–199. ISSN: 00222836 (1976).

62. Weathers, P., Jost, M. & Lamport, D. The gas vacuole membrane of Microcystis aeruginosa. en. *Archives of Biochemistry and Biophysics* **178,** 226–244. ISSN: 00039861 (Jan. 1977).

63. Walker, J. E. & Walsby, A. E. Molecular weight of gas-vesicle protein from the planktonic cyanobacterium *Anabaena flos-aquae* and implications for structure of the vesicle. en. *Biochemical Journal* **209,** 809–815. ISSN: 0264-6021 (Mar. 1983).

64. De Marsac, N. T., Mazel, D., Bryant, D. A. & Houmard, J. Molecular cloning and nucleotide sequence of a developmentally regulated gene from the cyanobacterium *Calothrix* PCC 7601: a gas vesicle protein gene. en. *Nucleic Acids Research* **13,** 7223–7236. ISSN: 0305-1048, 1362-4962 (1985).

65. Hayes, P. K., Walsby, A. E. & Walker, J. E. Complete amino acid sequence of cyanobacterial gas-vesicle protein indicates a 70-residue molecule that corresponds in size to the crystallographic unit cell. *Biochemical journal* **236,** 31–36 (1986).

66. Damerval, T., Houmard, J., Guglielmi, G., Csiszàr, K. & De Marsac, N. T. A developmentally regulated gvpABC operon is involved in the formation of gas vesicles in the cyanobacterium Calothrix 7601. en. *Gene* **54,** 83–92. ISSN: 03781119 (Jan. 1987).

67. Hayes, P., Lazarus, C., Bees, A., Walker, J. & Walsby, A. The protein encoded by gvpC is a minor component of gas vesicles isolated from the cyanobacteria Anabaena flos-aquae and Microcyctis sp. *Molecular microbiology* **2,** 545–552 (1988).

68. Walsby, A. E. & Hayes, P. K. The Minor Cyanobacterial Gas Vesicle Protein, GVPc, Is Attached to the Outer Surface of the Gas Vesicle. *Microbiology* **134,** 2647–2657. ISSN: 1350-0872 (1988).

69. Buchholz, B. E., Hayes, P. K. & Walsby, A. E. The distribution of the outer gas vesicle protein, GvpC, on the Anabaena gas vesicle, and its ratio to GvpA. *Journal of General Microbiology* **139,** 2353–2363. ISSN: 00221287 (1993).

70. McMaster, T. J., Miles, M. J. & Walsby, A. E. Direct observation of protein secondary structure in gas vesicles by atomic force microscopy. *Biophysical Journal* **70,** 2432–2436. ISSN: 00063495 (1996).

71. Offner, S., Ziese, U., Wanner, G., Typke, D. & Pfeifer, F. Structural characteristics of halobacterial gas vesicles. *Microbiology* **144,** 1331–1342. ISSN: 13500872 (1998).

72. Sivertsen, A. C., Bayro, M. J., Belenky, M., Griffin, R. G. & Herzfeld, J. Solid-state NMR characterization of gas vesicle structure. *Biophysical Journal* **99,** 1932–1939. ISSN: 15420086 (2010).

73. Sivertsen, A. C., Bayro, M. J., Belenky, M., Griffin, R. G. & Herzfeld, J. Solid-State NMR Evidence for Inequivalent GvpA Subunits in Gas Vesicles. *Journal of Molecular Biology* **387,** 1032–1039. ISSN: 00222836 (2009).

74. Bayro, M. J., Daviso, E., Belenky, M., Griffin, R. G. & Herzfeld, J. An amyloid organelle, solid-state NMR evidence for cross-β assembly of gas vesicles. *Journal of Biological Chemistry* **287,** 3479–3484. ISSN: 00219258 (2012).

75. Ezzeldin, H. M., Klauda, J. B. & Solares, S. D. Modeling of the major gas vesicle protein, GvpA: From protein sequence to vesicle wall structure. *Journal of Structural Biology* **179,** 18–28. ISSN: 10478477 (2012).

76. Strunk, T., Hamacher, K., Hoffgaard, F., Engelhardt, H., Zillig, M. D., Faist, K., Wenzel, W. & Pfeifer, F. Structural model of the gas vesicle protein GvpA and analysis of GvpA mutants in vivo. *Molecular Microbiology* **81,** 56–68. ISSN: 0950382x (2011).

77. Bollschweiler, D. Study of the archaeal motility system of Halobacterium salinarum by cryo-electron tomography, 1–118 (2015).

78. Dutka, P., Metskas, L. A., Hurt, R. C., Salahshoor, H., Wang, T.-Y., Malounda, D., Lu, G., Chou, T.-F., Shapiro, M. G. & Jensen, G. J. Structure of Anabaena flos-aquae gas vesicles revealed by cryo-ET. *Structure* **122,** 40a (2023).

79. Sachse, C. Single-particle based helical reconstruction—how to make the most of real and Fourier space. *Aims Biophysics* **2,** 219–244 (2015).

80. Henikoff, S. & Henikoff, J. G. Amino acid substitution matrices from protein blocks. en. *Proceedings of the National Academy of Sciences* **89,** 10915–10919. ISSN: 0027-8424, 1091-6490 (Nov. 1992).

81. Van der Maaten, L. & Hinton, G. Visualizing data using t-SNE. *Journal of machine learning research* **9** (2008).

82. Henikoff, S. & Henikoff, J. G. Amino acid substitution matrices from protein blocks. *Proceedings of the National Academy of Sciences* **89,** 10915–10919 (1992).

2

Nanofluidic chips for cryo-EM structure determination from picoliter sample volumes

Cryogenic electron microscopy has become an essential tool for structure determination of biological macromolecules. In practice, the difficulty to reliably prepare samples with uniform ice thickness still represents a barrier for routine high-resolution imaging and limits the current throughput of the technique. We show that a nanofluidic sample support with well-defined geometry can be used to prepare cryo-EM specimens with reproducible ice thickness from picoliter sample volumes. The sample solution is contained in electron-transparent nanochannels that provide uniform thickness gradients without further optimisation and eliminate the potentially destructive air-water interface. We demonstrate the possibility to perform high-resolution structure determination with three standard protein specimens. Nanofabricated sample supports bear potential to automate the cryo-EM workflow, and to explore new frontiers for cryo-EM applications such as time-resolved imaging and high-throughput screening.

2.1 Graphical Abstract
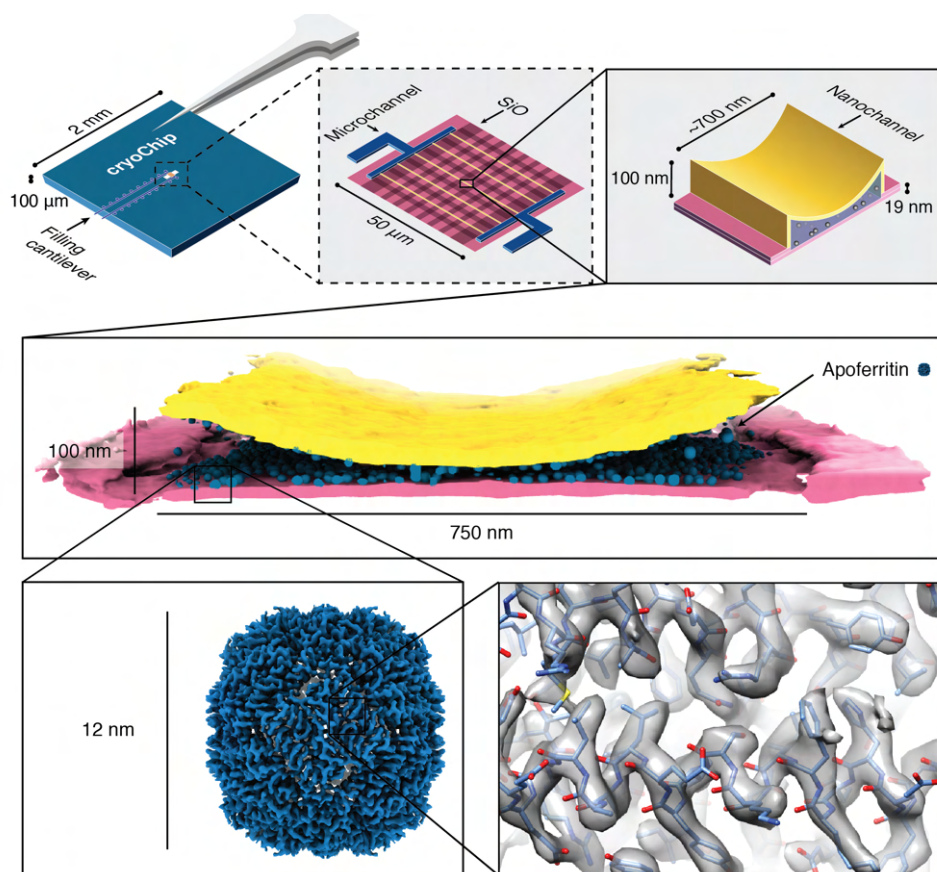
2.2 INTRODUCTION

This chapter focuses on investigating a novel approach to cryo-EM sample preparation, distinct from the conventional use of fenestrated supports with holey films, which inevitably expose the liquid sample to an air-water interface. The detrimental effects of the air-water interface can be avoided entirely if the sample were encapsulated within an enclosure made of a substrate sufficiently electron-transparent for high-contrast imaging of weak phase objects. Liquid-phase transmission electron microscopy (LP-EM) has been widely applied to image liquids inside the vacuum of an electron microscope by enclosing them in liquid cells[2,3]. Microfabrication of LP-EM sample cells has drawn on advances in semiconductor-manufacturing processes, in particular the use of free-standing silicon nitride (SiN) membranes[4–6]. This approach has also been successfully employed for room temperature LP-EM of biological macromolecular specimens[7–9]. To render liquid cells mechanically robust against elastic deformation caused by the pressure differential between the fluid-filled cell and the vacuum in the microscope column, LP-EM cells contain relatively thick (~30-50 nm) SiN membranes, whose additional scattering contributes structural noise that limits the attainable resolution.

In vitrified samples the solidified water behaves as an ultra-viscous fluid[10]. In this case elastic deflection of fluidic cells under a pressure difference is negligible, which offers opportunities for employing much thinner cells and membranes. Inspired by this prospect, we set out to design a nanofluidic sample cell for cryo-EM imaging. By miniaturising a sample-containing volume into a microelectromechanical system (MEMS) with ultrathin SiN membranes, we fabricated a cryo-EM sample support that allows cryo-EM imaging from picoliter sample volumes. We demonstrate the performance of this device by high-resolution 3D structure determination of three protein model systems, apoferritin, T20S proteasome and tobacco mosaic virus.

2.3 RESULTS

2.3.1 A NANOFLUIDIC SAMPLE SUPPORT FOR CRYO-EM IMAGING

To address common issues in cryo-EM grid preparation such as poor reproducibility of ice thickness, the destructive effect of the air-water interface and to broaden the scope of cryo-EM applications with the unique design flexibility of MEMS technology, we developed a novel cryo-EM sample support using nanofabrication techniques. We used a sacrificial polysilicon process for fabrication of a device featuring nanofluidic channels for controlled cryo-EM sample preparation and imaging [Figure 2.1, Supplementary Figure 2.7]. The nanochannels are enclosed by ultrathin membranes made from silicon-rich nitride (SiN_x) that allow for transmission of the electron beam when a thin film containing the sample is encapsulated in the space between them. The core design of this device, which we refer to as cryoChip, consists of a 2 x 2 mm silicon base with 100 μm thickness [Figure 2.1a]. A hollow microcantilever extends over the side of the chip base for sample application via capillary action. An aperture in the center of the silicon base exposes a 50 x 50 μm free-standing SiN_x membrane containing parallel rows of nanochannels that are connected to the central supply and exit channels via multifurcations. Transversal silicon oxide layers provide mechanical stability and generate a regular grid of observation windows that are 6x4 μm

in size [Figure 2.1b]. Each observation window exposes a segment of the nanochannel for TEM imaging [Figure 2.1c]. CryoChips differ in several fundamental aspects from grid-based sample supports: Instead of wicking a suspended droplet to a liquid film thin enough for transmission imaging, sample thickness in cryoChips is controlled by the geometry of electron-transparent nanochannels itself. The closed channel architecture avoids the air-water interface, which is replaced by a solid-liquid interface between the SiN_x membrane and the sample solution. With minimal adaptions, cryoChips are fully compatible with the most frequently used cryo-EM sample carrier cartridges [Supplementary Figure 2.8].

2.3.2 Geometric control of ice thickness

Control over ice thickness is a critical step for cryo-EM sample preparation. In standard blotting-based sample preparation, parameters such as blot force and time, absorbent properties of the filter paper, as well as relative humidity and temperature of the sample environment need to be tightly controlled. Despite the availability of robotic vitrification systems, achieving good reproducibility of these parameters has remained difficult and substantial across-grid and grid-to-grid variation of ice quality are common[11]. MEMS-based nanofabrication provides a means to control sample thickness directly through channel geometry, obviating the need for a particular sample environment and optimisation of blotting conditions. To quantify the reproducibility of ice thickness in cryoChip nanochannels, we first determined the average thickness of the silicon nitride layer by replicating the SiN_x deposition and etching protocol employed for device fabrication on a blank wafer area. Using ellipsometry we measured a film thickness of 9.5 ± 0.2 nm for the final SiN_x layer [Figure 2.1d]. We then filled the chip and used zero-loss imaging to determine the thickness of ice-filled nanochannels from the ratio of the integrated intensity of the zero-loss peak I_0 relative to the integral of the whole spectrum I [Figure 2.1e]. From the zero-loss log ratio of a SiN_x area adjacent to the nanochannel (0.89) and the geometric film thickness determined by ellipsometry (19 nm, corresponding to the combined thickness of top and bottom membrane), we calculated the mean free path for inelastic scattering (IMFP) at 300 keV in SiN_x as 163 nm, from which we estimate a combined water-equivalent thickness of 39 nm for the two SiN_x membranes forming the nanochannel assuming an IMFP of 320 nm for vitreous ice[12]. We next mapped the apparent and absolute ice thickness at each image position. The channels display a concave profile in cross-section, leading to a gradient with a maximum thickness of ~100 nm at the channel wall and 0 nm at the channel center. These estimates are in agreement with those obtained from a tomographic reconstruction from the same image area [Figure 2.1e]. We evaluated the consistency of nanochannel geometry both across individual cryoChips and between different chips. The precision of the sacrificial layer deposition in the fabrication process permits tight control over the maximum channel height. The width of the nanochannels is consistent across individual chips, but can range from 650 nm to 1000 nm between different chips which we attribute to process variability and/or the chip position on the wafer during fabrication [Figure 2.1f].

2.3.3 Rapid, loss-less preparation of cryo-EM samples via microcantilevers

To facilitate filling of the device from nanoliter sample droplets, we designed a hollow microchannel cantilever that allows sample application via spontaneous capillary flow
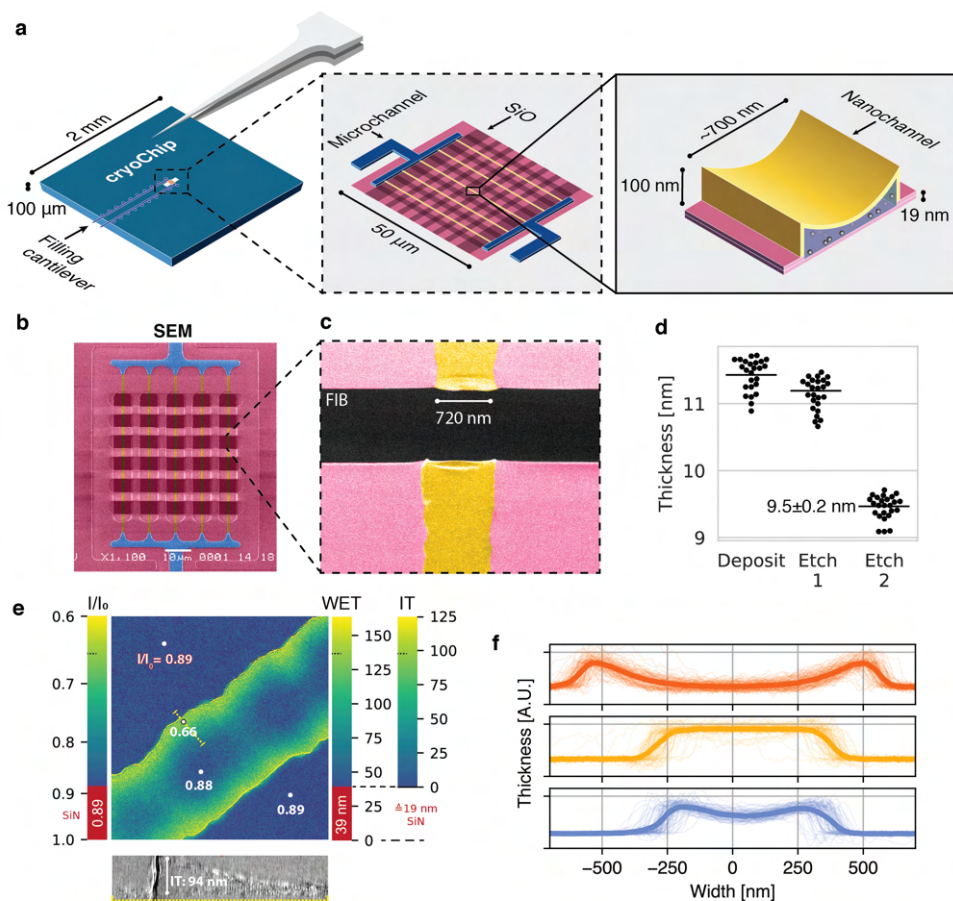
Figure 2.1: **A nanofluidic sample support for cryo-EM.** (a) Schematic representation of the MEMS-based cryoChip. The 2 x 2 mm chip base features two microcantilevers for sample filling and air escape. The cantilevers are connected to microchannels leading to a free-standing observation membrane located within an aperture at the chip center (middle inset). A multifurcation connects the microchannels (blue) to an array of five parallel nanochannels (yellow) of approximately 700 nm width and 100 nm height (rightmost inset). The top membrane is smoothly sagging, while the bottom membrane is flat. (b) False-colour scanning electron microscopy (SEM) image of the observation membrane. (c) False-colour SEM image of a single nanochannel with FIB-milled rectangular opening across the observation membrane (d) Ellipsometric determination of SiN$_x$ membrane thickness after deposition and etching steps. (e) Relative thickness map of an ice-filled nanochannel from zero-loss imaging. WET=water-equivalent thickness, IT=ice thickness. The thickness profile correlates with a 2D slice of a tomogram acquired of the same area (bottom). (f) Thickness cross-sections of nanochannels from three chips of the same wafer. Thin lines are individual nanochannel cross-sections measured across a whole chip, thick lines represent their mean.
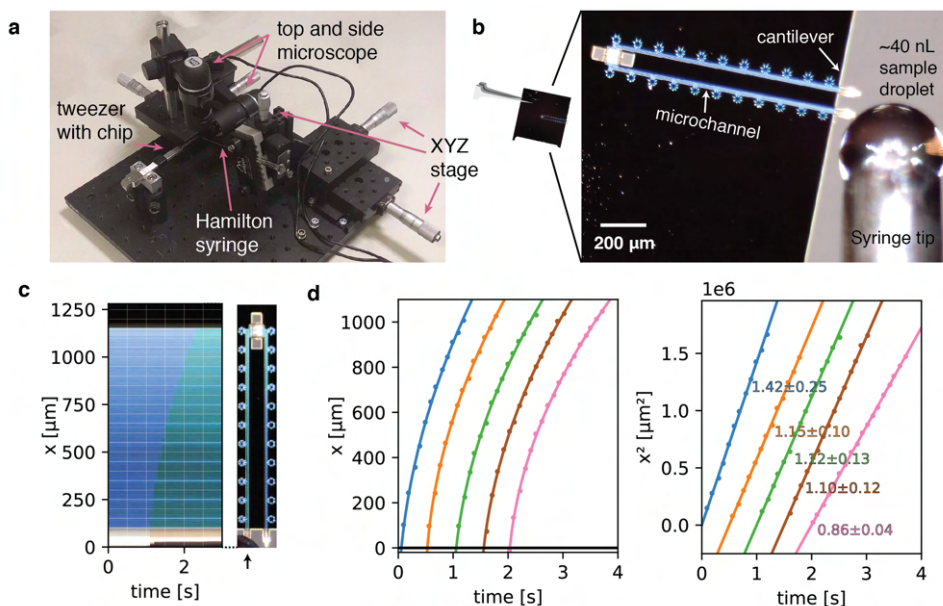
Figure 2.2: **Sample preparation with cryoChips.** (a) Manual filling station with side view and top view microscope, microvolume glass syringe, xyz precision stages for translation of the syringe and a mount for tweezers holding a cryoChip. (b) Sample loading of a cryoChip seen with the top view microscope. One of the cantilevers is approached with a ~40 nL sample droplet. (c) Kymograph of the filling process showing the progression of the sample meniscus over time. The arrow indicates the cantilever and the supply channel. (d) Filling kinetics of cryoChips. The meniscus position and associated time value were recorded at 10 fps with the top-view microscope, revealing a decreasing filling rate with progression of the fluid front (left panel). Each graph represents data from one filling experiment. Graphs from individual experiments are spaced apart by 0.5 s for better visualisation. The squared meniscus position (right panel) displays a linear relationship with elapsed filling time in agreement with expectation from Washburn kinetics. Solid lines are non-linear fits to the data. Slopes (10^6 $\mu m^2 s^{-1}$) and their standard deviations from bootstrapping statistics are shown.

without requirement for external appliances. The hollow microchannel cantilever extends only 100 μm over the chip base. To allow efficient targeting of the cantilever, we built a loading station consisting of a microvolume positive displacement syringe mounted on a linear XYZ precision stage, as well as top and side-view cameras to monitor the approach of the nozzle and filling process [Figure 2.2a]. The syringe nozzle is translated into close vicinity of the cantilever tip and a nanoliter droplet is formed. Manual handling of the syringe and the geometry of the nozzle currently limit the minimum droplet volume to about 10-40 nL [Figure 2.2b]. The cantilever is then dipped into the nanodroplet. Upon contact with the cantilever tip, a small sample volume is drawn into the cantilever by capillary forces and subsequently distributed across the central nanochannels via the multifurcation connected to the entry microchannel [Figure 2.2]. From the geometry of the overall channel design, we estimate the total filling volume to be ~4 pL (see Methods), which is about six orders of magnitude less compared to typical sample volumes of conventional blotting-based methods. Residual sample of the filling droplet can be readily recovered.

To analyse fluid kinetics in the microchannels, we filled a series of cryoChips via the sample application cantilever to generate kymographs and determined the capillary flow rate by monitoring the position of the filling meniscus over time. We observed the movement of the liquid front to be proportional to the square root of time, which is in accordance with classical Washburn kinetics for capillary flow[13]. Filling is rapid and complete filling of the entry channel proceeds at an average of ~1 mm/s [Figure 2.2c]. The filling speed is fast at the beginning and slows down with progression of the liquid front, as predicted by the Washburn equation. We found similar behaviour across the chips that were analysed, suggesting that we can achieve high channel-to-channel reproducibility during fabrication [Figure 2.2d, Supplementary Figure 2.9]. The entire filling process can be executed in a lab environment without humidity control and there is no further requirement for parameter optimisation other than sample concentration. After sample filling, the chip can be transferred to any manual or robotic vitrification device for flash cooling in liquid cryogen. In our experiments, the total time from filling to vitrification was approximately 10 seconds. This process could be readily accelerated by integrating the filling mechanism directly on a vitrification system.

2.3.4 Automated cryo-EM data collection with cryoChips

The checkerboard pattern from the silicon oxide layers on the observation membrane structures the nanochannel array into a regular grid of 30 observation squares, each comprising a segment of an individual nanochannel and flanking areas of the SiN_x membrane [Figure 2.3a,b]. Due to the different layout of the cryoChip, standardised data acquisition protocols established for holey support film substrates cannot be applied without modification. To develop an automated data acquisition strategy for cryoChips we first acquired a grid map of the entire observation membrane using SerialEM. For each observation square we then selected six equally spaced acquisition points positioned linearly along the midline of the nanochannel [Figure 2.3c]. We used virtual maps[14] generated from medium magnification images to accurately realign to the acquisition points during high-magnification acquisition. For each of the acquisition points, three images were acquired radially around the midpoint using beam-image shift, resulting in up to 540 images per cryoChip.
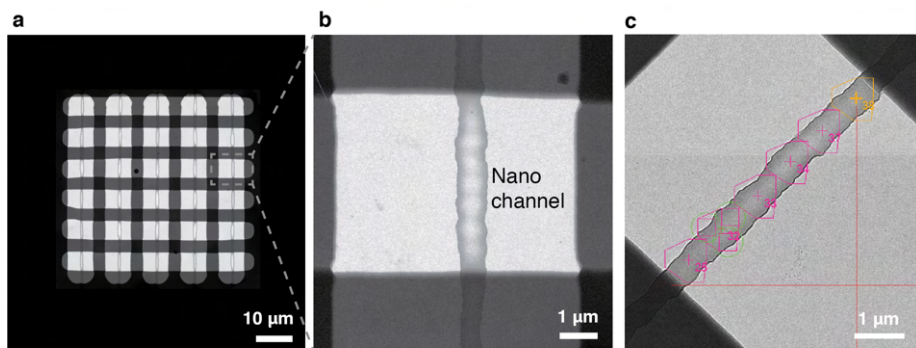
2

Figure 2.3: **Automated data acquisition with cryoChips.** (a) TEM image of the observation membrane with five nanochannels. The checkerboard structure is a stiffening layer made from silicon oxide that divides the nanochannels in observation windows. (b) Close-up of inset in (a). (c) Medium magnification image of an observation square with an overlay of the SerialEM multishot acquisition scheme containing six acquisition points along the nanochannel. Each acquisition point is centered on a polygon outlining the limits of three imaging positions achieved by beam-image shift, leading to a total of 18 images per observation square.

2.3.5 3D STRUCTURE DETERMINATION FROM CRYOCHIP SAMPLES

To validate the suitability of cryoChips for 3D structure determination, we collected datasets for three test specimen (*P. furiosus* ferritin (ApoFtn, 480 kDa, O), *T. acidophilum* 20S proteasome (T20S, 700 kDa, D7) and tobacco mosaic virus (TMV, 37 MDa, helical)) that we selected for differences in size, shape and symmetry. The low magnification maps revealed that most observation squares were suitable for data collection. We occasionally observed partially broken or cracked membranes. Observation squares in such areas were omitted for data collection. Micrographs for all three specimens showed good image contrast [Figure 2.4a]. To examine the compatibility of cryoChips with the condition to form vitreous ice, we evaluated the fraction of images containing crystalline ice by manual inspection for the presence of ice rings or crystalline reflections in the computed power spectra from all images. This led to a fraction of 86% (815/945), 75% (91/121) and 52% (43/82) of images that we classified as containing vitreous ice for ApoFtn, T20S and TMV respectively [Figure 2.4e, Supplementary Figure 2.10]. This is consistent with convective heat-transfer simulations of a simplified cryoChip model, suggesting that cooling below the glass-transition temperature occurs at timescales ($<10^{-5}$) seconds compatible with vitrification [Supplementary Figure 2.11]. It is possible that the fraction of vitreous images is underestimated since we did not discriminate between ice features originating from surface contamination and those originating from crystalline ice in the nanochannels. To investigate whether non-vitreous images are a result of inefficient vitrification or phase transition to cubic ice during exposure, we analysed a subset of images classified as icy. From a maximum intensity projection of all power spectra of all images, we find evidence for hexagonal ice indicating incomplete vitrification [Supplementary Figure 2.12]. We did not observe evidence for a correlation of the probability for forming crystalline ice and the location on the observation membrane.

The SiN_x membrane forming the walls of the nanochannels leads to additional back-
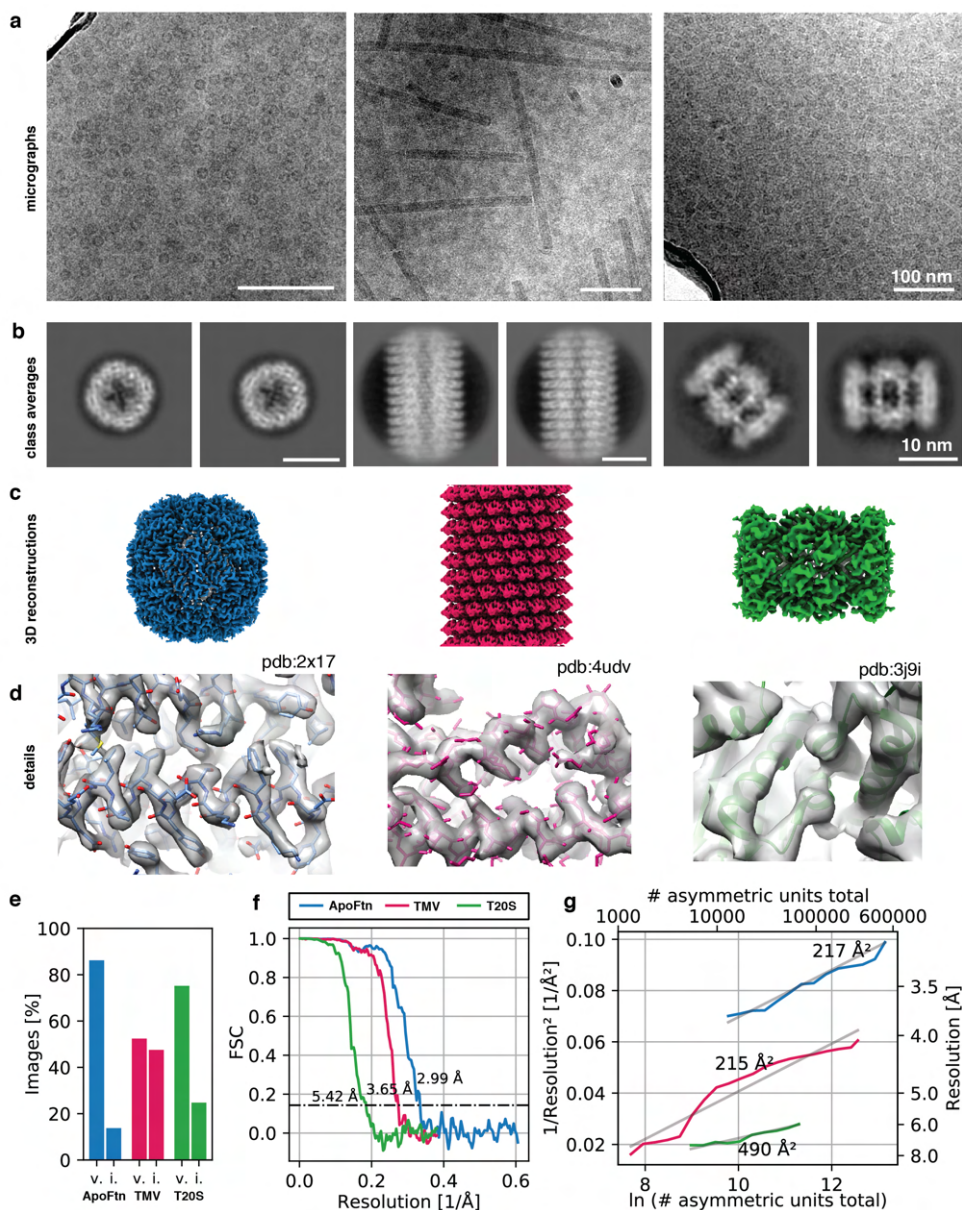
Figure 2.4: **3D reconstruction of ApoFtn, T20S proteasome and TMV.**
(a) Raw micrographs, (b) 2D class averages, (c) 3D cryo-EM maps and (d) close-up densities displaying discernible details for the three test specimens. Atomic models were overlaid as visual aid: ApoFtn pdb:2x17[15], TMV pdb:4udv[16], T20S pdb:3j9i[17] (e) Percentage of images in the dataset classified as vitreous (v) and icy (i). (f) Fourier Shell Correlation (FSC) curves with 0.143 cutoff. (g) ResLog plots for reconstructions. B-factors were estimated as 2 over the fitted slope. Colour coding is equivalent for panels (c-g). Micrographs in (a) were taken at a defocus of -1.2 μm for ApoFtn, -1.8 μm for TMV and -2.8 μm for T20S. Scale bars are equivalent for images in (a) and class averages in (b).

2

ground noise from an increase in the fraction of electrons that are scattered inelastically. This effect is equivalent to an apparent increase in ice thickness, which even for particles in the thinnest parts of the nanochannel may lead to an (apparent) ice thickness significantly greater than the size of the particle itself. This is undesirable because it leads to a corresponding loss of useful signal in the image, thereby reducing image contrast. To test whether this additional scattering limits our ability to extract high-resolution signal for particle alignment, we used automatic particle picking in cryoSPARC 3.1[18] to extract particles (segments for TMV) and performed 2D classification. For all three specimens we observe high-contrast 2D class averages with visible features indicative of secondary structure [Figure 2.4b, Supplementary Figure 2.13, Supplementary Figure 2.14, Supplementary Figure 2.15], suggesting that images contain sufficient contrast for high-resolution structure determination. To probe the resolution limit of data collected using the cryoChip, we merged data from three chips filled with ApoFtn acquired on a Titan Krios equipped with a post-GIF K2 camera and processed it in a standard single-particle analysis workflow. We used 2D and 3D classification to select a final set of 21,238 particles (O symmetry) from 815 micrographs and after 3D refinement obtained a reconstruction at an overall resolution of 3.0 Å displaying visible side chain densities [Figure 2.4c/d]. The reference model for 3D alignment was generated from the data itself using the stochastic gradient descent method [Supplementary Figure 2.13], suggesting that the quality of data collected from cryoChips is sufficient to allow initial model generation even for samples considered difficult for reference model generation. To demonstrate the versatility of cryoChips for structure determination from a variety of biological specimens, we also processed the datasets for TMV and T20S proteasome [Supplementary Figure 2.14, Supplementary Figure 2.15]. These datasets were collected on a JEOL JEM3200-FSC with a K2 detector and yielded 3D reconstructions with overall resolutions of 3.7 Å and 5.4 Å obtained from 14,238 segments (20 asu/segment) and 5,750 particles (D7), respectively [Figure 2.4f, Table 2.1]. The quality of a dataset is often quantified from a plot of the inverse-squared resolution as a function of the number of particles used in the 3D reconstruction. The slope of this function is related to a B-factor that describes the width of a Gaussian dampening function in reciprocal space used to model the loss of high-resolution information in particle images owing to a combined effect of performance-reducing factors related to the sample, data collection and instrumentation, and data processing[19]. We quantified the B-factors for each of the three datasets [Figure 2.4g]. For comparison, we also acquired datasets for ApoFtn and T20S on conventional holey carbon supports with the same acquisition parameters, yielding reconstructions at an overall resolution of 2.4 and 3.2 Å [Supplementary Figure 2.16, Table 2.2]. This leads to B-factors of 217 Å^2 vs. 178 Å^2 for ApoFtn and 490 Å^2 vs. 328 Å^2 for T20S determined for cryoChip and holey carbon datasets, respectively. We also note a consistent upward shift of the offset for holey carbon B-factor plots compared to those from cryoChips, consistent with the differences in overall signal-to-noise ratio caused by additional scattering from the SiN_x membranes.

2.3.6 Beam-induced specimen motion is comparable to standard cryo-EM supports

Beam-induced motion is a major factor limiting the achievable resolution of 3D reconstructions. The highest rate of change in particle position is typically largest at cumulative

exposures below ~4e$^-$/Å^2[20], coinciding with the frames where the sample is yet least affected by radiation damage and high-resolution signal is best preserved. The fast read-out rates of direct electron detection technology have made it possible to acquire cryo-EM data as dose-fractionated movies where the total exposure is subdivided into a series of individual frames. To reduce information loss, alignment of these frames can be used to determine the overall inter-frame movement of the specimen in the micrograph during the exposure and then correct for this motion before summation of the frame stack. To investigate beam-induced sample movement in cryoChips, we computed and analysed the alignment trajectories for the ApoFtn dataset [Figure 2.5a]. The calculated traces show that movement is largely unidirectional, but clusters into several distinct directions of motion. To further investigate the trajectories, we separated them either by chip or by the beam-image shift condition corresponding to the shot position in the multishot acquisition scheme for each acquisition point [Figure 2.5b, Supplementary Figure 2.17]. For each of the chips analysed, we observe three clusters of directional motion, each associated with one of the three shot positions. While the precise origin for this clustering needs further investigation, we suggest that different thickness and geometry of the confined ice layer in the subregions for each of the multishot positions may lead to preferential directionality of specimen movement, possibly combined with microlensing effects resulting from adjacent exposures that induce the build-up of semi-static charge on the non-conductive SiN$_x$ membrane.

To compare average specimen movement in cryoChips with other commonly used specimen supports we also computed the alignment trajectories for four datasets in the EMPIAR database[21] and ApoFtn on holey carbon [Figure 2.5c, Supplementary Figure 2.16b,c]. These datasets have been acquired using different specimen supports but resulted in high-resolution 3D reconstructions [Table 2.4]. We then computed the accumulated movement as the root mean squared deviation (RMSD) of all trajectories from the ApoFtn cryoChip dataset and the EMPIAR datasets and evaluated them over an equivalent exposure range [Figure 2.5d]. The datasets show small differences in accumulated movement over the whole exposure, ranging from 3.8 (QF-R2/2) to 5.1 Å (cryoChip) RMSD over 0-40 e$^-$/Å^2. These movements are in the order of 0.1 pixel per frame and are therefore amenable to motion correction. These overall motions represent a combination of mechanical (stage) drift and beam-induced motion. Limiting motion in early frames is essential to maximise recovery of high-resolution detail. EMPIAR-10306 has a remarkably low initial movement of 0.7 Å, while both our dataset and EMPIAR-10389 display a accumulated motion of 1.8 Å. EMPIAR-10389 led to a 2.0 Å reconstruction of *Y. enterocolitica* urease, indicating that initial specimen movement of that order is not limiting for achieving resolution beyond 2.0 Å.

2.3.7 Preferential partitioning to the solid-liquid interface effectively concentrates protein particles

Interaction with the air-water interface may cause perturbation of molecular structure in thin film-based cryo-EM sample preparations. The closed sample environment of the cryoChip avoids the problems of an air-water interface, but the large surface-to-volume ratio of the nanochannels forms an extensive solid-liquid interface between the SiN$_x$ surface and the particle solution. We performed tomographic reconstructions of protein-loaded cryoChips to investigate particle distributions inside the nanochannels. The tomographic
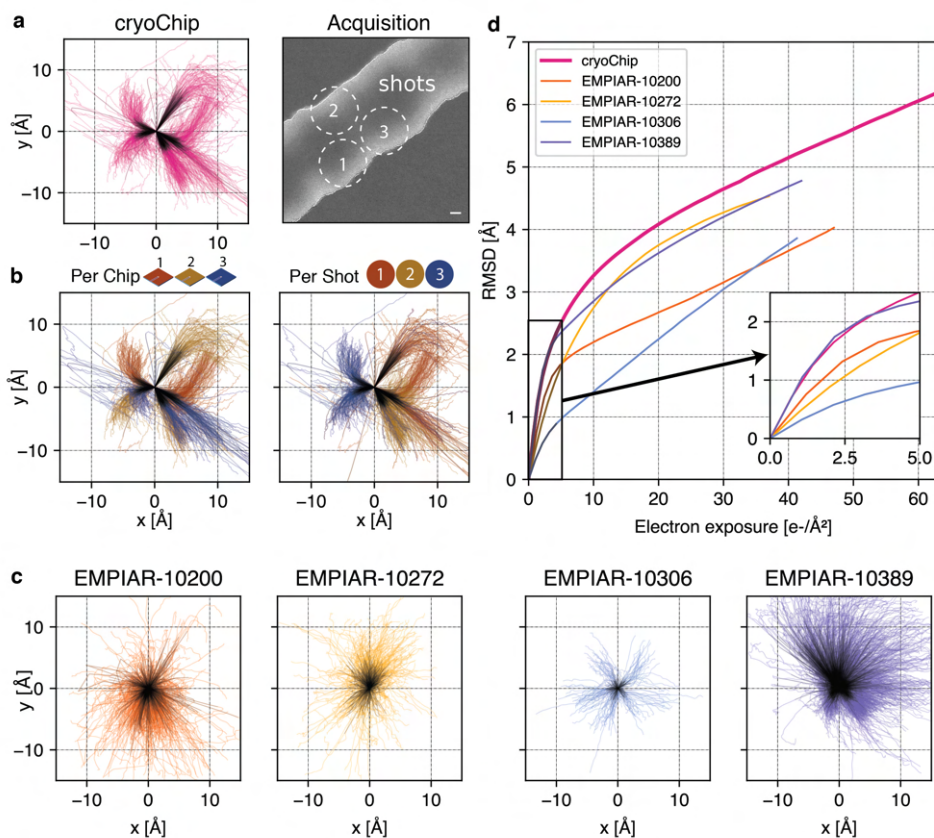
Figure 2.5: **Comparison of beam-induced motion for cryoChip and reference datasets from conventional sample preparation.** (a) Whole-frame motion trajectories of dose-fractionated movies during exposure with 63 e$^-$/Å^2 from three different cryoChips. (b) Trajectories coloured by chip or by imaging position in a repeating three-shot pattern. (c) Whole-frame motion trajectories of reference datasets. (d) Root-mean-square deviation (RMSD) of trajectories over electron exposure between the cryoChip dataset and reference datasets. The inset shows the first part of the exposure most critical for reconstructing high-resolution details. Scale bar in (a) is 100 nm. First five e$^-$/Å^2 in trajectories are coloured in black.

reconstructions confirm the overall geometry of the channels with an estimated height in cross-section ranging between 0-100 nm and a channel width of ~750 nm [Figure 2.6a/b]. In all cases we find the majority of particles close to or associated with the SiN_x membrane [Figure 2.6c, Supplementary Figure 2.18, Figure 2.6]. Analysis of the orientation distribution showed increased abundance of views along the fourfold rotation axis for ApoFtn, indicating preferential alignment of particles with the SiN_x membrane [Supplementary Figure 2.13]. Similarly, we find preferential abundance of side views for T20S particles [Supplementary Figure 2.15]. We did not observe convincing correlation of preferred orientations with the molecular electrostatic potential [Supplementary Figure 2.20]. We note that the current analysis is limited to particles with internal symmetry for which preferred orientations are less problematic. Further investigation will be required to conclude whether systematic effects for preferred protein adsorption to the solid-liquid interface exist and need to be addressed. Micrographs from datasets of all three specimen display particle densities exceeding those expected from conventional cryo-EM grids prepared with the same specimens and sample concentrations. To quantify this concentration effect, we estimated the particle concentration in the nanochannels by segmenting and counting ApoFtn particles, and measuring the channel volume from the reconstructed tomogram [Supplementary Figure 2.18]. We find 1483 ApoFtn particles within a 9.4×10^6 nm^3 volume, equivalent to a concentration of 262 µM or 126 mg/mL assuming 480 kDa molecular mass. This is a 37-fold increase in concentration compared to the initial sample concentration of 3.4 mg/mL. Interaction with the solid-liquid interface at the SiN_x channel walls thus strongly affects the effective sample concentration in the nanochannels. Similar concentration effects have also been reported for traditional sample preparation methods, such as repeated sample application[22] and in the presence of very thin ice for which densely packed protein monolayers have been observed[23].
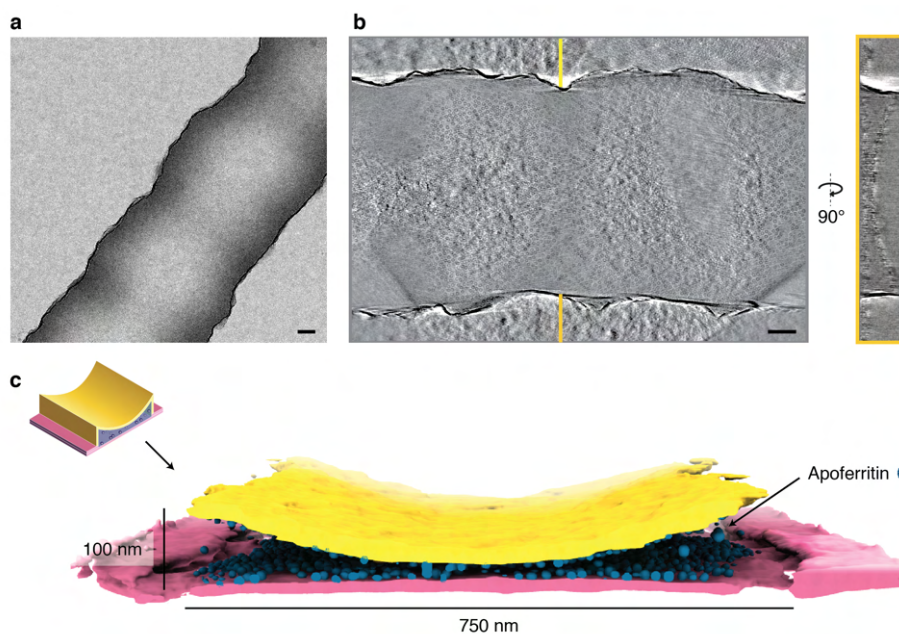
Figure 2.6: **Tomographic reconstruction of a cryoChip nanochannel filled with ApoFtn.** (a) Overview image of tomogram area of a nanochannel segment filled with ApoFtn. (b) Slice through the reconstructed tomogram and yz cross-section across the nanochannel indicated by lines. (c) Segmented reconstruction showing the nanochannel membrane (bottom: pink, top: yellow) and ApoFtn in the channel lumen (blue). The thickness of the membrane is affected by the missing wedge and not to scale. All scale bars: 100 nm.

2.4 Discussion

The process of sample preparation is widely perceived as a major bottleneck for realising the full potential of cryo-EM. All existing approaches to cryo-EM sample preparation rely on the ill-controlled formation of thin ice layers across fenestrated support films, which results in poor reproducibility and may detrimentally affect particle distribution, orientation and integrity[24–27]. Recent improvements to the support grids and to how the sample is applied have begun to address these issues[24,28–33]. Despite these advances important issues remain, most prominently the hydrophobic environment of the air-water interface[34,35]. Here, we have shown that MEMS-based nanofluidic chips in which the sample is enclosed and protected from the air-water interface within electron-transparent nanochannels provide a viable alternative for cryo-EM sample preparation that mitigates some of the prevalent issues observed for grid-based sample supports. We demonstrated that biological specimens in cryoChips can be successfully vitrified and that the thickness uniformity of the ice layer can be reproducibly controlled by channel geometry. We found the width and height of the nanochannels to be highly consistent across individual cryoChips, providing uniform imaging conditions for all parts of a chip.

The most widely applied methods for cryo-EM sample preparation currently require ~3 μL of sample volume per grid. Assuming a final liquid film of 20-100 nm thickness spanning a 3.05 mm EM support grid, the required sample volume for imaging is only 1.5-7.0 fL, which is nine orders of magnitude less than the initial sample volume. Several recent developments are trying to reduce this gap: The Chameleon/Spotiton instrument eliminates sample waste from the blotting step by applying nanoliter sample volumes onto self-wicking grids via a piezo-electric droplet dispenser[31,36]. Similarly, the VitroJet and Cryo-writer (pin-)print sub-nanoliter volumes directly onto a small area of the cryo-EM grid[30,32]. The cryoChip presented here further reduces the required sample volume by 1000-fold to only a few picoliters, approaching the practical minimum for liquid handling in cryo-EM structure determination. The cantilever-based sample application efficiently drives chip filling via capillary action, thus allowing essentially loss-less cryo-EM sample preparation. Together with the highly reproducible sample thickness we expect these properties to accelerate developments towards full automation of the cryo-EM sample preparation process. Notably, sample preparation using cryoChips does not require humidity control and is compatible with any standard plunger available in cryo-EM laboratories.

Using cryoChip supports we demonstrated *ab initio* structure determination for three different test specimen resolved at resolutions suitable for mechanistic interpretation even from datasets with a comparatively small set of particles. The diversity in shape, size and physico-chemical properties of the macromolecular complexes suggests that cryoChip sample preparation should be applicable for routine cryo-EM imaging of a wide variety of different macromolecular targets. Several aspects bear potential for future improvement. Despite our predictions from heat transfer simulations we observed that between 10-40% of images contain non-vitreous ice. Several factors may contribute to this discrepancy. Firstly, it is possible that cryogen is displaced by the chip base upon entering the cryogen, preventing complete wetting of the chip; or that air is trapped in the depression on the chip base below the free-standing observation membrane, preventing contact with the coolant

and impeding heat transfer from the nanochannels. Secondly, the observation membrane is located at the chip center which could affect probability of vitrification, consistent with the spatial distribution of isotherms reported for standard holey film grids during immersion cooling[37]. Parameters for existing plunging instruments have been optimised to maximise vitrification efficiency for holey film supports and it is conceivable that different plunging speed or depth, or jet-based cooling, may be more appropriate for cryoChips.

In standard cryo-EM sample preparation using holey support grids, the blotting process leaves a very thin layer of sample that can be imaged with high contrast and minimal secondary scattering from the particle-embedding ice layer. In cryoChips, the concave cross-sectional shape of the nanochannel produces a defined thickness gradient within which particles naturally partition into monolayers according to their molecular dimensions in the thinner parts of the channel, thus minimising background scattering from surrounding ice. The sample confinement in cryoChip nanochannels, however, comes with the penalty of additional incoherent scattering from the enclosing SiN_x membranes, which results in reduced image contrast compared to ideal samples. By quantifying the additional scattering from the nanochannel membranes as an effective increase in ice thickness, we estimate the minimal scattering mass of particles that can still be successfully detected and aligned under these conditions to ~200 kDa[38,39]. This should render cryoChips suitable for imaging a wide range of relevant protein targets for cryo-EM structure determination.

Apart from incoherent scattering contributions by the SiN_x membrane, several other factors may currently limit resolution: Firstly, spectral properties of structural noise contributed by the SiN_x membranes are different from those of vitreous ice, and secondly, there may be interference of signal from SiN_x membranes and particles located at different focal heights within the nanochannel that may affect accuracy of CTF estimation[40] [Supplementary Figure 2.19]. It is possible that assumptions on image statistics in current 3D reconstruction algorithms are sub-optimal for particles from cryoChips and improvements could be gained with modified software.

Tomographic reconstructions of sample-filled nanochannels showed that proteins preferentially partition to the solid-liquid interface of the nanochannel surface. This is consistent with observations using modified EM grids that contain surface substrates spanning the holey support film[24,28,41]. While this may appear to raise issues similar to adsorption at the air-water interface, the nature of the interaction with these surfaces is fundamentally different. The air-water interface is hydrophobic and adsorption of proteins to the interface can drive partial or complete unfolding to minimise the free energy of interaction[24,25,27]. In contrast, similar to oxidised graphene, the polar SiN_x surface is hydrophilic and adsorption of proteins is primarily mediated by the zeta potential of the interfacial electrical double layer[42]. Consistently, for the three cases we investigated we do not observe that this interaction promotes denaturation or aberrant structural changes. It is entirely possible that the charge distribution between the molecular electrostatic potential and at the solid-liquid interface could promote preferential alignment and hence carry the risk of systematic bias in particle orientation. For the cases studied we did not find evidence for pronounced preferred orientation and could not identify strong correlation of

enriched views with the molecular electrostatic potential.

We also observed that the strong interaction at the solid-liquid interface effectively concentrates the solute particles in the nanochannels, consistent with observations for other substrates providing solid-liquid interfaces such as hydrogenated graphene[41]. While this is principally beneficial for concentrating dilute protein solutions to particle densities suitable for imaging, it also affects design considerations for the filling channels. In early design iterations of the cryoChips that featured narrow filling and exit channels, we observed apparent depletion of the protein sample before it reached the observation nanochannels. This effect was further aggrevated with samples containing high surface charge densities. We attribute this to the very large surface-to-volume ratio of nanochannels, which we suspect may bear the risk to deplete solute proteins through interactions with the SiN walls in the filling channel, thus gradually reducing the free protein concentration in solution as they travel towards the observation membrane. The increased diameter of the filling channel in the current cryoChip design substantially improved the efficacy for reaching suitable particle concentrations in the nanochannels of the observation membrane. With solid-liquid interfaces gaining importance also for existing cryo-EM supports[24,28,43], further studies are needed to reveal the detailed dynamics and underlying mechanisms of particle adsorption to such interfaces, and to systematically investigate their effects on particle integrity and orientation.

2.5 Methods and Materials

Chip fabrication

The cryoChip is fabricated using a wafer-scale surface micromachining process. The fluidic components, such as hollow cantilevers, thick-walled supply- and exit channels, and thin-walled nanochannels, are all formed by sacrificial polysilicon etching techniques. The fabrication process starts on a 4 inches, <100>-oriented silicon wafer of 380 μm thickness. The wafer is locally thinned from the backside to form silicon membranes with a thickness of 180 μm, in which later the cryoChip membrane will be formed. The backside processing is performed by anisotropic wet etching of silicon in a 25% potassium hydroxide (KOH) solution at 75°C. A 50 nm thick stoichiometric silicon nitride (Si_3N_4) layer is used as the masking material for the KOH etch. After the KOH etching the masking layer is removed in concentrated hydrofluoric acid (HF 49%). On the locally thinned wafer, a 300 nm thick silicon oxide layer is deposited in a low pressure chemical vapour deposition (LPCVD) oven by pyrolysis of tetraethylorthosilicate (TEOS). The silicon oxide layer is patterned by buffered hydrofluoric acid (BHF). Silicon oxide is removed from an area where ultra-thin walled observation nanochannels will be formed. On the patterned substrate, an 11 nm silicon-rich nitride (SiN_x) layer is deposited by LPCVD. This low-stress SiN_x layer serves as a bottom wall of the TEM observation windows, which will be created later. Next, a sacrificial layer of polysilicon with a thickness of 300 nm is deposited by LPCVD. This relatively thick polysilicon layer will be used to create the supply- and exit channels and a large fluidic channel inside the hollow cantilevers. The deposited polysilicon layer is patterned by tetramethylammonium hydroxide (TMAH). The layer is removed from

a region where later the observation nanochannels will be formed. After the TMAH patterning, another sacrificial polysilicon layer with a thickness of 100 nm is deposited by LPCVD The thickness of the second sacrificial layer determines the height of the small fluidic channels and the height of the TEM observation windows. On the other hand, the aggregate thickness of the first and the second polysilicon layer, which is 400 nm in our design, determines the height of the larger fluidic channels. After the deposition, sacrificial polysilicon is patterned by TMAH etching to define an outline of the fluidic channels. A rather thin thermally grown silicon oxide layer (40 nm) is used as an etching mask. It should be noted that both large (400 nm) and small (100 nm) fluidic channels are etched at the same time. The underetching of the mask, due to an isotropic nature of the TMAH etch process, is used to reduce the width of the fluidic channels in which the TEM observation windows will be formed. After patterning of the fluidic channels, the second silicon-rich nitride layer is deposited by LPCVD. The 11 nm thick low-stress SiN_x layer will define a top wall of the TEM observation windows. Thereafter, a 300 nm thick silicon oxide layer is deposited by TEOS pyrolysis. A multilayer of SiN_x and silicon oxide will form the walls of the hollow cantilevers and some fluidic channels. The relatively large thickness of these walls provides mechanical stability of these components. After patterning of the second TEOS layer in BHF the TEM observation windows are completely defined. Subsequently, the $TEOS/SiN_x/SiN_x/TEOS$ multilayer is etched by reactive ion etching (RIE). In this processing step the contour of the chip and the hollow cantilevers are patterned. After the RIE etching, an access to the bulk silicon is created. By using Deep Reactive Ion Etching (DRIE) the bulk silicon is etched from the front side to a depth of 150 µm. Finally, the wafer is placed in a 25% TMAH solution heated at 90°C. During the TMAH etching, sacrificial polysilicon is removed through the access holes and the fluidic channels are created. These temporary access holes are automatically sealed after final drying of the samples. At the same time the bulk silicon is exposed to TMAH. A part of the bulk silicon underneath the TEM observation window is removed creating a suspended membrane with integrated thin-walled fluidic nanochannels. The bulk underneath the cantilevers is also removed and the cantilevers are released. During the TMAH etching the backside of the wafer remains unprotected. Due to the removal of silicon the cryoChip is further thinned down to its final thickness of 100 µm.

Filling volume

The filling volume entering the chip through the cantilever was estimated from the chip geometry. The supply microchannel is about 1 mm long, 0.4 µm high and 4 µm wide, which results in an estimated volume of ~2 pL. The five nanochannels for imaging are 50 µm long, ~750 nm wide and 100 nm high, which (assuming a parabolic profile) yields an aggregate volume of 6.25 fL, negligible compared to the microchannel volume. Assuming the liquid fills the entire entry microchannel and exit microchannel, the total filling volume is ~4 pL.

SEM Imaging and FIB

SEM imaging of cryoChips was performed on a Helios G4 CX DualBeam FIB/SEM system (ThermoFisher Scientific) at 35000x magnification using the TLD detector. A gallium beam was used to mill a rectangular opening in the observation membrane to image the cross-section of the nanochannel. Grayscale SEM images were coloured in Adobe Photoshop

according to the schematic representations in Figure 1.

Thickness measurements of silicon nitride

For thickness determination of the ultrathin silicon nitride layer surrounding the nanochannels, we deposited a silicon-rich nitride layer on a dummy wafer using the same deposition parameters as used for cryoChip fabrication. The dummy wafer was annealed at 1100°C. Next, the wafer was etched for 1 min in HF 1% solution ('etch 1'). In our fabrication process, the HF 1% dip is used to remove a native oxide layer from silicon prior to the TMAH etchings. Then the dummy wafer was etched in TMAH 25% solution at 90°C ('etch 2'). The same etching time is used in our fabrication process to release the samples. The thickness of the SiN_x layer after deposition and both etching steps was measured on 25 positions of the wafer by ellipsometry.

For thickness determination by energy-filtered TEM (EFTEM), we acquired images of a filled nanochannel at 8000x magnification (4.706 Å/pixel) on a JEOL JEM3200-FSC operated at 300 kV (K2 detector) with and without a 20 eV slit (centered at 0 eV) inserted below the omega filter. Images were recorded on a Gatan K2 direct electron detector. The resulting images were aligned and I/I_0 was computed by dividing the pixel values of the zero-loss image by those of the unfiltered image. The known total SiN thickness of 19 nm from ellipsometry measurements was used to calibrate the mean free path for inelastic scattering of silicon-rich nitride to 163 nm at 300 kV. We assumed a literature value of 320 nm[12] for the mean free path of vitreous ice. The derived quantity 'water-equivalent thickness' (*WET*) is the thickness of water/vitreous ice that would lead to the same I/I_0, and is computed as follows:

$$WET = 320 \text{ nm} \cdot \ln \frac{I_0}{I} \tag{2.1}$$

The quantity 'ice thickness' (*IT*) is computed by removing the contribution of the silicon nitride from *WET*:

$$IT = WET - \left[163 \text{ nm} \cdot \ln \frac{1}{0.89} \right] \tag{2.2}$$

WET cross-sections of nanochannels of whole chips were computed from grid-maps of three chips at 2500x (72.64 Å/pixel) on a TVIPS XF416 detector with and without a 20 eV slit inserted. Observation windows in unfiltered and zero-loss images were aligned by cross-correlation, and profiles across each nanochannel computed as averages over five stripes per observation window. Additionally, the average over all profiles was computed for each chip.

Protein expression and purification

ApoFtn: Pyrococcus furiosus apoferritin (Uniprot Q8U2T8, in pET-24d(+)) was expressed in E.coli BL21-DE3 cells using autoinduction[44]. Cells were initially grown at 37°C for 4 hours after which temperature was reduced to 20°C. After 20h, cells were harvested, resuspended

in lysis buffer (50 mM Tris, 300 mM NaCl, 0.1% Triton X-100) and disrupted by sonication. The lysate was cleared by centrifugation at 17,000 g at 4°C for 45 minutes. The cleared supernatant was boiled for 20 minutes at 80°C and centrifuged for 45 minutes at 17,000 g at 4°C. The supernatant was dialysed into ion exchange buffer (20 mM NaCl, 20 mM HEPES, pH=6.8) overnight and loaded onto a 1 mL HiTrap Q XL column (Cytiva). Following a wash step with 25 column volumes (cv) ion exchange buffer, bound protein was eluted with a 20-500 mM NaCl gradient over 50 cv. Fractions containing target protein in high purity were selected by SDS-PAGE and pooled. Pooled fractions were concentrated and finally purified by size exclusion chromatography on a Superdex200 10/300 column (Cytiva) in 20 mM Tris, 50 mM NaCl, pH 7.5. Fractions of the elution peak were pooled and concentrated to a final concentration of 3.4 mg/mL.

TMV: TMV was purified from infected tobacco leaves as described[45] and concentrated to 35 mg/mL.

T20S: Thermoplasma acidophilium 20S proteasome (pRSF-T20S, Addgene plasmid 110805) was expressed and purified essentially as described previously[46]. In brief, protein was expressed in E.coli BL21-DE3 cells using autoinduction medium following the same expression protocol as for ApoFtn. After 20 hours, cells were harvested and lysed by sonication in lysis buffer (50 mM NaH2PO4, 150 mM NaCl, pH 8.0), followed by addition of three volumes of boiling lysis buffer to precipitate non-thermostable proteins. The supernatant was further purified by immobilised metal affinity chromatography (IMAC) followed by cleavage of the His-tag by TEV protease at room temperature for 48 hours. Final polishing was done by size exclusion chromatography on a Superose 6 column (Cytiva) with SEC buffer (20 mM Tris, 50 mM NaCl, 0.1 mM EDTA, pH 8.0). Fractions of the elution peak were pooled and concentrated to 1.4 mg/mL.

cryoChip filling and vitrification

A microvolume filling station for cryoChips was built on an optical breadboard (Thorlabs, USA). A 10 μL positive displacement microvolume glass syringe (Hamilton) was mounted on a xyz translation stage for precise position control of a sample droplet relative to the sample application cantilever of a cryoChip. Two Toolcraft USB-microscopes (Conrad Electronics GmbH, Germany) were used for a top and side view of the filling process with video acquisition control on a computer. The transparent front cap was removed from the microscopes to increase the working distance. Both microscopes were mounted on an xy translation stage to orient the field of view. All experiments were done under ambient conditions without temperature or humidity control. A cryoChip was gripped on its edge with tweezers and positioned on a metal bearing for rapid and reproducible positioning in the field of view. Approximately 0.2 μL of sample was aspirated with the microvolume syringe. The syringe tip was dried with a tissue to avoid wetting of the tip during droplet extrusion. The syringe was mounted on the loading station and a small droplet of ~10-40 nL was extruded from the syringe. Using the xyz stage, the droplet was brought into contact with the cantilever tip. Within seconds, the liquid fills the entry microchannel, the nanochannels and parts of the exit microchannel. The liquid in the syringe was typically sufficient to fill many chips by drying the tip again with a tissue and extruding a new droplet. We did not observe a change in hydrophilic character of the cryoChip or reduction in chip filling capacity over a period of several month. Approximately 2-3 sec after immersing

the cantilever tip into the sample droplet, the tweezer with the chip was removed from the loading station and transferred on a Leica EM GP2 vitrification robot set to 20°C. The plunger was set to blot for zero seconds and the chip was plunged immediately into liquid ethane kept at -180°C. The time between removing the tweezer and chip from the filling station and plunge freezing was ~10 seconds.

Electron cryo-tomography

A cryoChip with ApoFtn at 3.4 mg/mL concentration was prepared as described above and stored in liquid nitrogen until imaging. Tilt series of cryoChips were acquired on a JEOL JEM3200-FSC with a bidirectional tilt scheme starting from 0° to -60° and subsequently from 2° to 60°, in increments of 2° with a total exposure of 94.6 e$^-$/Å^2 , a pixel size of 4.706 Å and a defocus target of 6 μm. Micrographs acquired as dose-fractionated movies of 30 frames were motion-corrected in MotionCor2 v1.0[47] and dose-weighted according to their accumulated dose[48]. The tilt series was aligned using patch tracking and reconstructed using back-projection as implemented in Etomo from the IMOD package[49]. Segmentation of silicon nitride membrane and particles was done in the tomoseg program of EMAN2 using a machine learning approach[50]. Segmented tomograms were visualised in ChimeraX[51]. A subvolume of the segmented ApoFtn tomogram with 376 x 750 x 100 nm^3 channel volume was segmented in ChimeraX to count individual ApoFtn particles for concentration determination using the 'segment map' command. Tomogram acquisition and reconstruction for cryoChips with TMV and T20S were done using the same protocol, using a pixel size of 4.706 Å for TMV and 3.668 Å for T20S.

Single-particle imaging of ApoFtn

Three cryoChips with ApoFtn at 3.4 mg/mL were sent to EMBL Heidelberg for data acquisition. A grid map of the whole observation membrane was acquired for each chip. Acquisition points were set up to automatically acquire three images in a circle by beam-image shift in SerialEM[52] using the 'multiple Record Setup' setting. Six acquisition points were placed in the nanochannel section of each observation window, which leads to 18 images per nanochannel section. Using py-EM[14], reference maps were generated from the grid map to find back and align to the acquisition points with high precision. Coma-free alignment was performed on the collapsed silicon nitride membrane adjacent to the nanochannels. Movies of 90 frames were acquired during a 9 second exposure on a Krios (ThermoFisher Scientific) microscope with a Quantum-K2 detector. The total exposure of the image was 63 e$^-$/Å^2, the pixel size 0.8127 Å/pix and the dose rate ~2 e$^-$/pixel/sec. The defocus range was set to -1 to -2 μm. In total 945 movies were acquired from three cryoChips. A reference dataset of 276 movies were acquired for ApoFtn on QF1.2/1.3 holey carbon grids using the same data acquisition settings.

Single-particle imaging of T20S and TMV

Datasets comprising 121 movies of the T20S proteasome at 1.4 mg/mL and 82 movies of TMV at 1.1 mg/mL were acquired on a JEOL JEM 3200-FSC microscope operated at 300 kV. For T20S, we used cryoChips with a slightly different design featuring orthogonal nanochannels arranged in a clover-like pattern instead of straight nanochannels [Supplementary Figure 2.18c]. Both TMV and T20S single-particle datasets were acquired on a Quantum-K2

detector at a magnification of 30,000x (1.288 Å/pixel) and an exposure time of 10.95 seconds with a total exposure of 59.68 e⁻/Å² over 73 frames. A reference dataset of 879 movies were acquired for T20 on QF1.2/1.3 holey carbon grids using the same data acquisition settings.

Single-particle image processing

945 movies from three cryoChips filled with ApoFtn were processed in cryoSPARC 3.1[18] [Supplementary Figure 2.13].Patch-motion correction and patched CTF estimation was followed by manual curation of micrographs to remove images with visible ice reflections in the power spectra, retaining 815 micrographs. Particles were picked with the blob-picker and subjected to 2D classification to obtain a picking reference. The reference was low-pass filtered to 20 Å and was used to pick 318,058 ApoFtn particles. After 2D classification, selection and local-motion correction in cryoSPARC, 97,431 particles were retained. Three more rounds of 2D classification were used to select for classes with visible secondary structure features, resulting in 34,102 particles. Retained particles were used for initial model generation by cryoSPARC's stochastic gradient descent method while imposing O symmetry. The resulting reference volume was low-pass filtered to 15 Å and used in homogeneous refinement, leading to a 3D reconstruction at 3.7 Å resolution. Per-particle defocus refinement and beam-tilt refinement improved the resolution to 3.0 Å. Heterogenous refinement with 3 classes was used to further prune particles and lead to a 3.0 Å resolution reconstruction from 21,238 particles. In all cases, O symmetry was imposed during reconstruction. 276 movies of a reference dataset of ApoFtn on QF1.2/1.3 holey carbon grids were processed in the same way and led to a reconstruction at 2.4 Å resolution.

82 movies from two cryoChips filled with TMV were imported in cryoSPARC 3.1[18], followed by patch-motion correction and patched CTF estimation [Supplementary Figure 2.14]. 43 of 82 movies with vitreous ice were retained for further processing. Manual picking and 2D classification was used to obtain picking references. Filament tracing was used to trace and segment virus particles with a step size of 20 asymmetric units (1.41 Å each), resulting in 31,623 segments. Two rounds of 2D classification were used to prune segments. Local motion correction and re-extraction yielded 14,238 TMV segments. Helical refinement was started from a cylindrical initial model and an initial symmetry of 22.04° twist and 1.41 Å rise, while refining the symmetry. This yielded a 3D reconstruction at 4.3 Å reconstruction. Two rounds of per-particle defocus and global CTF (tilt & trefoil) refinement improved the reconstruction to 3.7 Å. The final refined symmetry converged to 22.036° twist and 1.415 Å helical rise.

121 movies of T20S proteasome were processed in cryoSPARC 3.1[18] [Supplementary Figure 2.14]. Patch-motion correction and patched CTF estimation were performed, followed by manual curation of micrographs to remove movies with visible reflections from crystalline ice in their power spectra. 91 movies were retained for further processing. Blob-picking and 2D classification was used to generate references for automated picking. Five reference classes were low-pass filtered to 20 Å for template picking, resulting in 47,484 particles. After two rounds of 2D classification and selection, 12,351 particles were retained and used for ab-initio 3D reconstruction imposing D7 symmetry. The initial model thus obtained was low-pass filtered to 30 Å and used for homogenous refinement, yielding a reconstruction at 7.9 Å resolution. After heterogenous refinement with three classes, particles in the class

with highest resolution (5,750 particles) were selected and subjected to another round of homogenous refinement yielding the final reconstruction at 6.0 Å resolution. Per-particle defocus and global CTF refinement (tilt & trefoil) further improved the resolution to 5.4 Å. All reconstructions had D7 symmetry imposed. 879 movies of a reference dataset of T20S on QF1.2/1.3 holey carbon grids were processed in the same way and led to a final reconstruction at 3.2 Å resolution. For all final reconstructions, five independent runs of ResLog[53] analysis were performed and the results averaged for ResLog curves. Local resolution was estimated in cryoSPARC using a windowed FSC method[54] and visualised in ChimeraX[51]. Map sharpening was done in cryoSPARC by applying the overall B-factor estimated from Guinier plots.

Atomic model building

Atomic models for TMV (PDB ID 4udv) and T20S (PDB ID 3j9i) were rigid body-fitted into the density maps and not further refined. The atomic model from an available crystal structure of ApoFtn (PDB ID 2x17)[15] was rigid body-fitted into the cryo-EM density, and real-space refinement in Phenix 1.13[55] was iterated with manual model building in Coot 0.8.9.2[56]. Model validation statistics were retrieved from MOLPROBITY[57] as implemented in Phenix and overfitting was probed by half-map refinement.

Beam-induced motion trajectory analysis

Four datasets (10200[58], 10272[28], 10306[59], 10389[60]) from the Electron Microscopy Public Image Archive (EMPIAR[21]) were subjected to full-frame motion correction in cryoSPARC[18] to calculate motion trajectories and were compared with equivalently derived trajectories from the cryoChip ApoFtn dataset.

Kymograph generation

Movies of the filling process for each cryoChip were acquired with a Toolcraft USB-microscope (Conrad Electronics GmbH, Germany) at ten frames per second and a pixel size of 1.6 µm. A ten pixel-wide section along the supply or exit microchannels was averaged and plotted over time. For each time point coordinates of the liquid front were manually picked and individually plotted for each chip. A square root function of the form

$$x(t) = \sqrt{a(t-b)} + c \tag{2.3}$$

was fitted to each filling trajectory using the curve_fit function from the scipy package[61]. For error estimation of the fitting parameters, a bootstrapping approach was used, performing the fit 100 times with a random set of n out of n coordinates with replacement. The parameter and its uncertainty was estimated as mean and standard deviation over the 100 fits. For the linearised plots of x^2 versus t, the shift parameter c was subtracted from the curve.

Convective heat transfer simulations

Simulations were performed in COMSOL Multiphysics 5.4 using a simplified 2D geometry of cryoChips represented by a 2 mm x 100 µm silicon base and 50 µm x 20 nm SiN observation membrane. The initial temperature of the chip was set to room temperature (293.15 K) and

the ethane bath was at 90 K. Heat flux was modelled by a external forced convection model with 1.5 m/s fluid velocity and a plate length of 2 mm. This does not take into account geometry of the plunging process or fluid dynamics and uses an average heat flux for every surface of the chip. Temperature development over time was evaluated over a profile along the observation membrane and plotted in different time windows. Material properties were obtained from the COMSOL material library unless stated otherwise. The specific heat capacity c_p for silicon nitride was assumed as 700 $J kg^{-1} K^{-1}$ and its density as 3100 $kg m^{-3}$. For thermal conductivity of SiN, we averaged values from[62] for thin membranes over the relevant temperature range to obtain 2.3 $W m^{-1} K^{-1}$. For silicon, heat capacity was assumed as 700 $J kg^{-1} K^{-1}$, thermal conductivity as 130 $W m^{-1} K^{-1}$ and density as 2329 $kg m^{-3}$.

2.6 Supplementary Figures and Tables

2

Table 2.1: Data collection and refinement statistics for collected datasets

	ApoFtn **EMD-12901**	**TMV** **EMD-12903**	**T20S** **EMD-12915**
Specimen	3 cryoChips filled with 3.4 mg/mL ApoFtn	2 cryoChips filled with 1.1 mg/mL TMV	1 cryoChip filled with 1.4 mg/mL T20S proteasome
Microscope	TFS Titan Krios	JEOL 3200-FSC	JEOL 3200-FSC
Voltage (kV)	300	300	300
Detector	K2-XP	K2-XP	K2-XP
Energy filter	Gatan BioQuantum	In-column omega filter	In-column omega filter
Number of movies (vitreous/crystalline)	945 (815 / 130)	82 (43 / 39)	121 (91/30)
Pixel size (Å)	0.8127	1.288	1.288
Electron exposure (e$^-$/Å^2)	63	59.7	59.7
# of frames	90	73	73
Exposure time (s)	9	10.95	10.95
Defocus range (μm)	-1 to -2	-1 to -3.5	-2 to -4
Symmetry imposed	O	helical rotation: 22.036°, helical rise: 1.415 Å	D7
Map sharpening B-factor (Å^2)	-88.3	-65.8	-353.5
Final number of particles / asymmetric units	21,238 / 509,712	14,238 / 284,760	5750 / 80,500
Final map resolution (Å)	3.0	3.7	5.4
Map resolution range (Å)	2.9-3.5	3.5-4.3	5.2-6.5
EMPIAR	10708	10708	10708

Table 2.2: Data collection and refinement statistics for reference datasets on holey carbon grids

	ApoFtn-QF	**T20S-QF**
Specimen	Quantifoil R2/1 300mesh 3.4 mg/mL ApoFtn	Quantifoil R2/1 300mesh 1.4 mg/mL T20S
Microscope	TFS Titan Krios	JEOL 3200-FSC
Voltage (kV)	300	300
Detector	K2-XP	K2-XP
Energy filter	Gatan BioQuantum	In-column omega filter
Number of movies	276	879
Pixel size (Å)	0.8127	1.288
Electron exposure (e$^-$/Å^2)	63	59.7
# of frames	90	93
Exposure time (s)	9	10.95
Defocus range (μm)	-1 to -2	-2 to -4
Symmetry imposed	O	D7
Map sharpening B-factor (Å^2)	-73.6	-175.2
Final number of particles / asymmetric units	21,683 / 520,392	279,332 / 3,910,648
Final map resolution (Å)	2.4	3.2

Table 2.3: Model refinement statistics for ApoFtn atomic model.

	ApoFtn 7ohf
Initial model used (PDB code)	2x17
Model resolution (Å, FSC = 0.5)	3.3
CC mask	0.776
Model composition	
Nonhydrogen atoms	24 x 1399
Protein residues	24 × 169
RMSD	
Bond lengths (Å)	0.008
Bond angles (°)	0.864
Validation	
MolProbity score	0.88
Clashscore	1.41
Rotamer outliers (%)	1
Ramachandran plot	
Favored (%)	99.4
Allowed (%)	0.6
Disallowed (%)	0

Table 2.4: | **Data and 3D reconstruction parameters of reference datasets.**

EMPIAR	Protein	Support film	Resolution	B-factor	# particles	Pixel size (Å)	Acq. scheme	Ref.
10200	ApoF	QF-R2/2 holey carbon	1.65 Å	66 Å^2 *	426,450	0.814	multi-shot	[58]
10272	ApoF	Ultrafoil-R0.6/1 + graphene	2.14 Å	54 Å^2 **	41,202	0.649	single-shot	[28]
10306	TMV	CF-2/2-2C holey carbon	1.90 Å	41 Å^2 **	20,000	0.638	multi-shot	[59]
10389	Urease	QF-R1.2/1.3 holey carbon	1.98 Å	38 Å^2 **	119,020	0.639	both	[60]
10708	ApoFtn	cryoChip SiN$_x$	2.99 Å	88.3 Å^2 **	21,238	0.813	multi-shot	

*B-factor from ResLog plot. **B-factors as determined from Guinier plot.
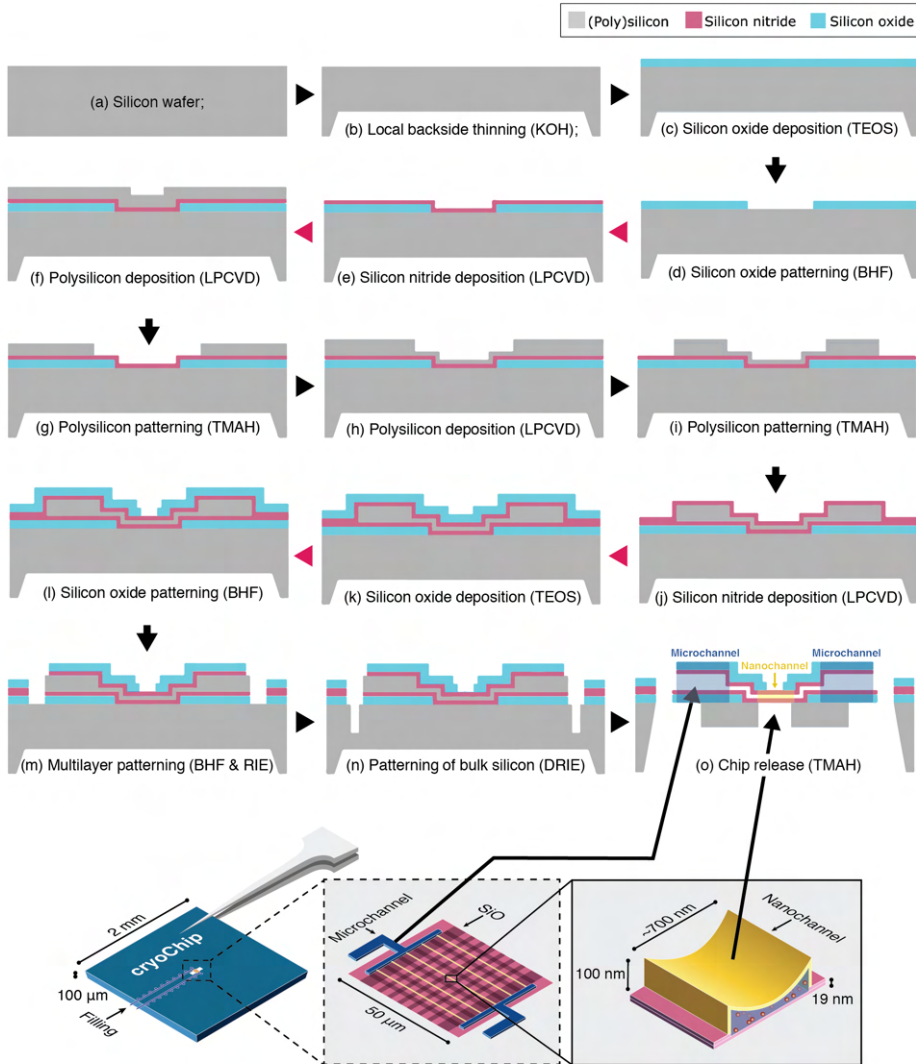ApoF: horse spleen apoferritin; ApoFtn: *P. furiosus* apoferritin

Figure 2.7: **Fabrication schematic of cryoChips.** All steps for nanofabrication of the cryoChip are shown in cross-section. The graphic on the bottom shows the location of microchannels (supply channels for filling) and nanochannels (for imaging) in the schematic.
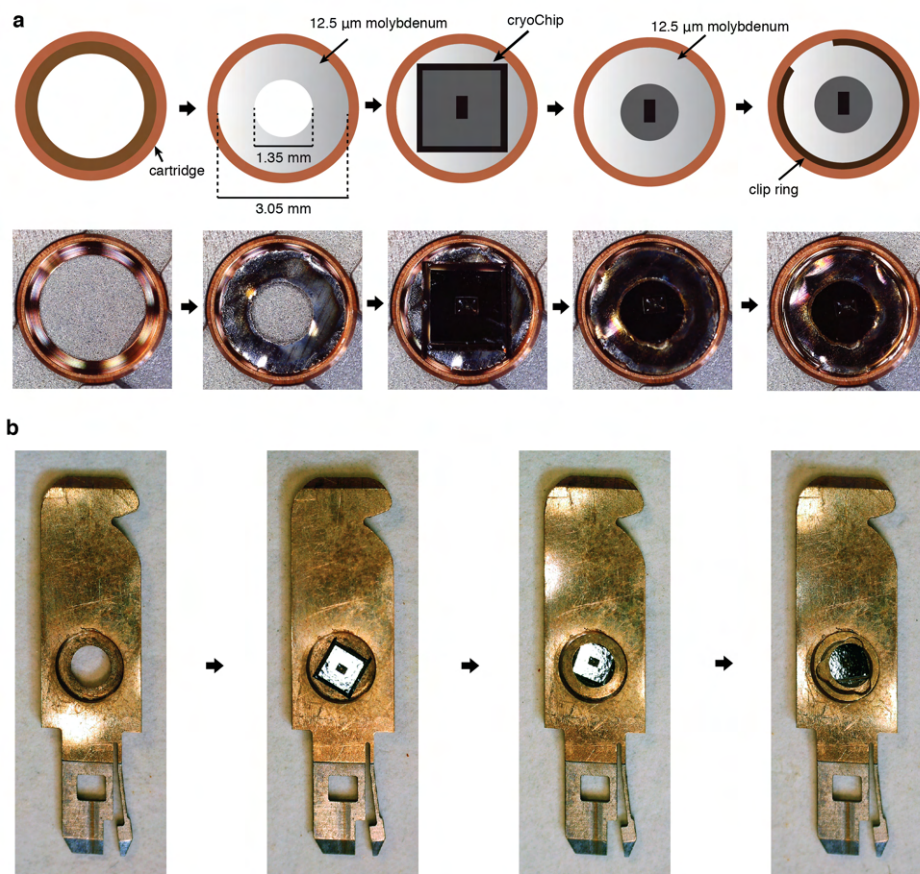
2

Figure 2.8: **Preparation of cryoChips for TEM imaging.** (a) Mounting in Autogrid cartridges (Thermo Fisher Scientific). A cryoChip is sandwiched between two molybdenum adapter rings with a circular aperture (laser cut from 12.5 μm molybdenum foil) and fixated with the Autogrid clip ring. Chips are mounted with the observation membrane facing the cartridge base and the slanted KOH-etch side walls facing the clip ring (b) Mounting in cartridges for the JEOL JEM3200-FSC microscope. The chip is placed with the observation membrane facing the cartridge base. The default metal spacer ring is placed on top and the assembly is secured by a screw ring.
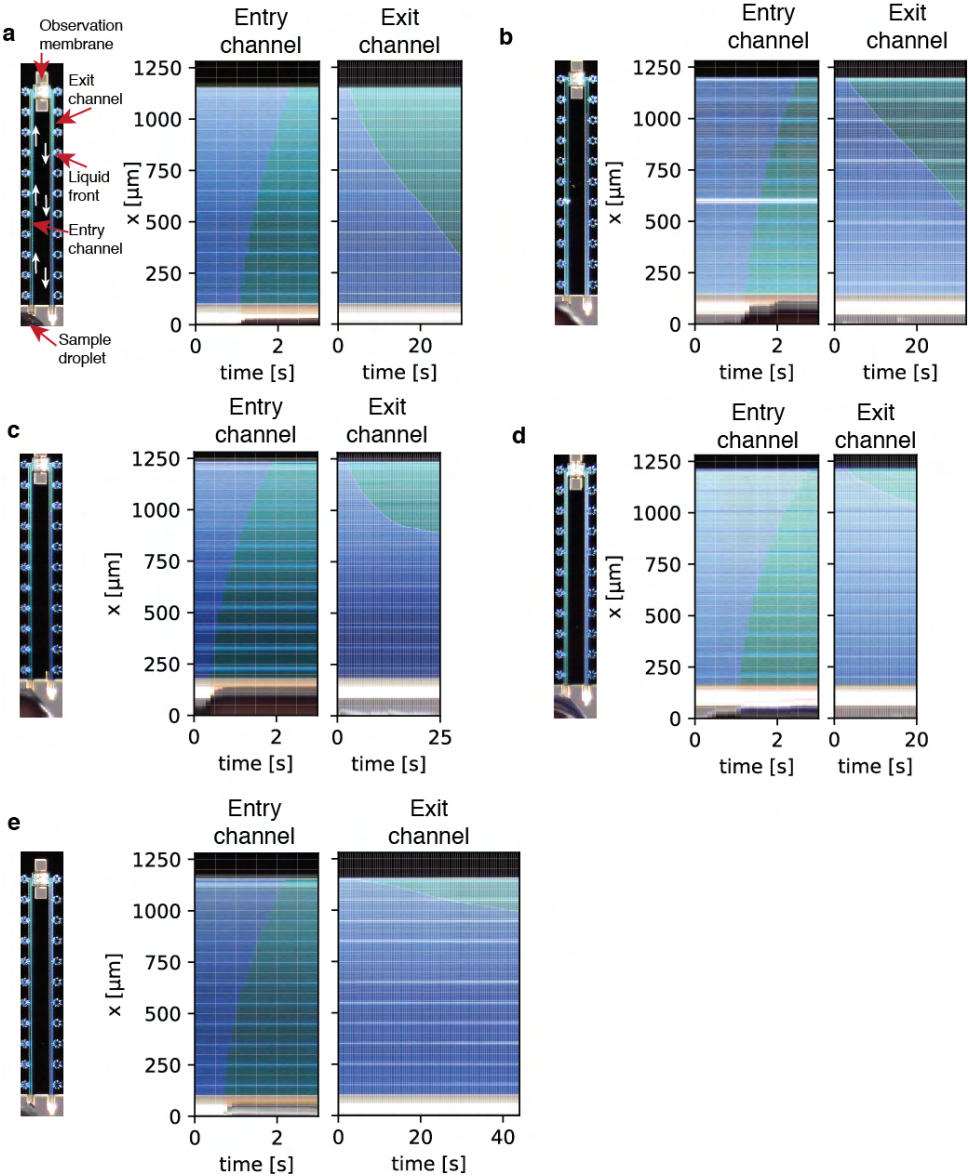
Figure 2.9: Kymographs showing liquid entry and exit through the supply- and exit microchannels. The liquid takes around one second to flow from the cantilever inside the whole entry channel. Due to the small cross-section of the nanochannels in the observation membrane, liquid flows out of the exit channel very slowly.
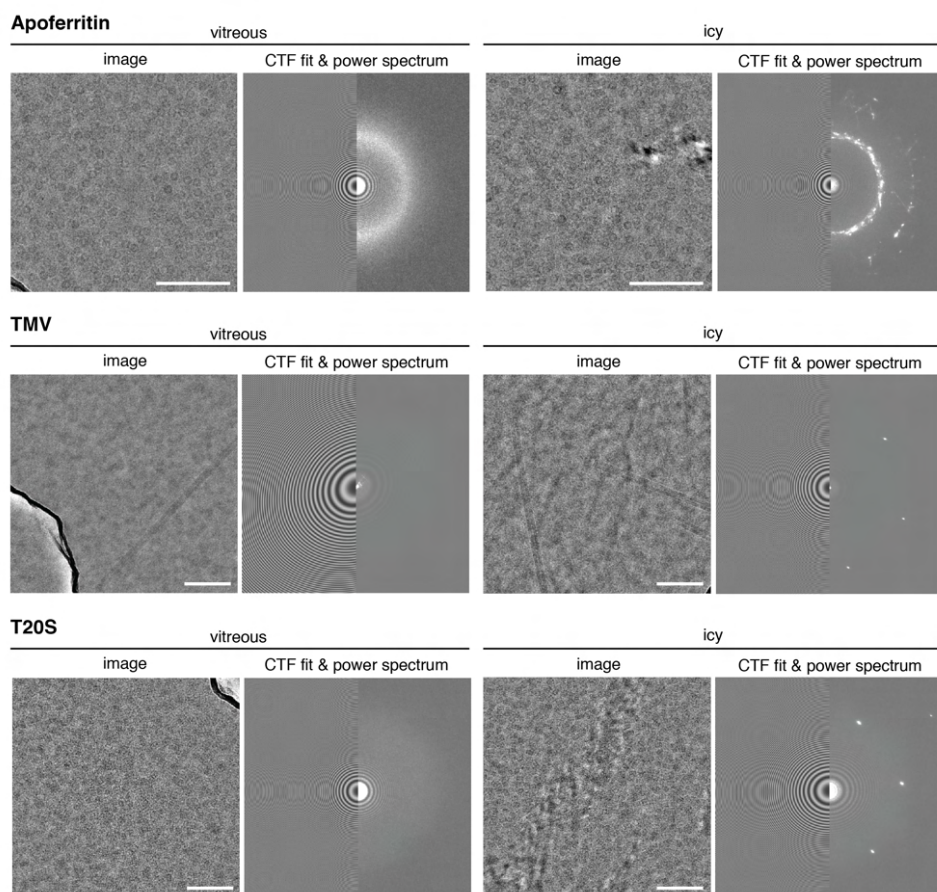
Figure 2.10: Vitrification efficiency analysis. Representative images with vitreous ice (left panels) and crystalline ice (right panels) from the three test specimens. CTF fits to the power spectra are also shown. Thon rings from SiN_x can be used for coma-free alignment. Scale bar is 100 nm.

Figure 2.11: Simulation of plunge freezing of a simplified model of cryoChips in COMSOL 5.4. (a) Temperature profiles over 50 μm observation membrane from 1-10 μs (pink), 1-10 ms (orange) and 10-100 ms (blue). Dashed line shows glass-transition temperature of water at 136 K. (b) 2D model of cryoChips with 2 mm x 100 μm silicon base and 50 μm x 20 nm SiN_x observation membrane used in simulations.

2

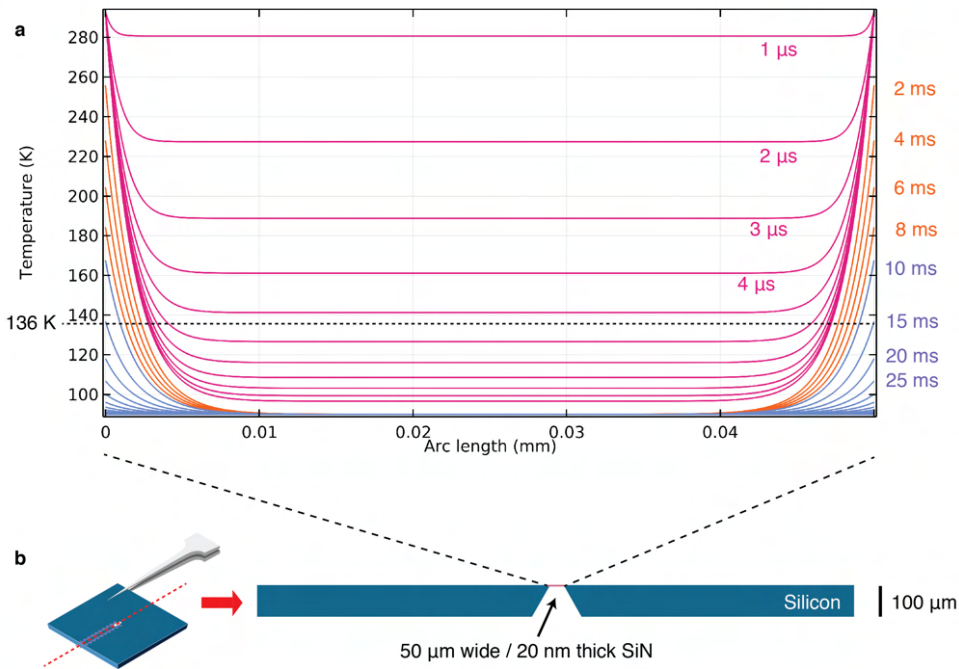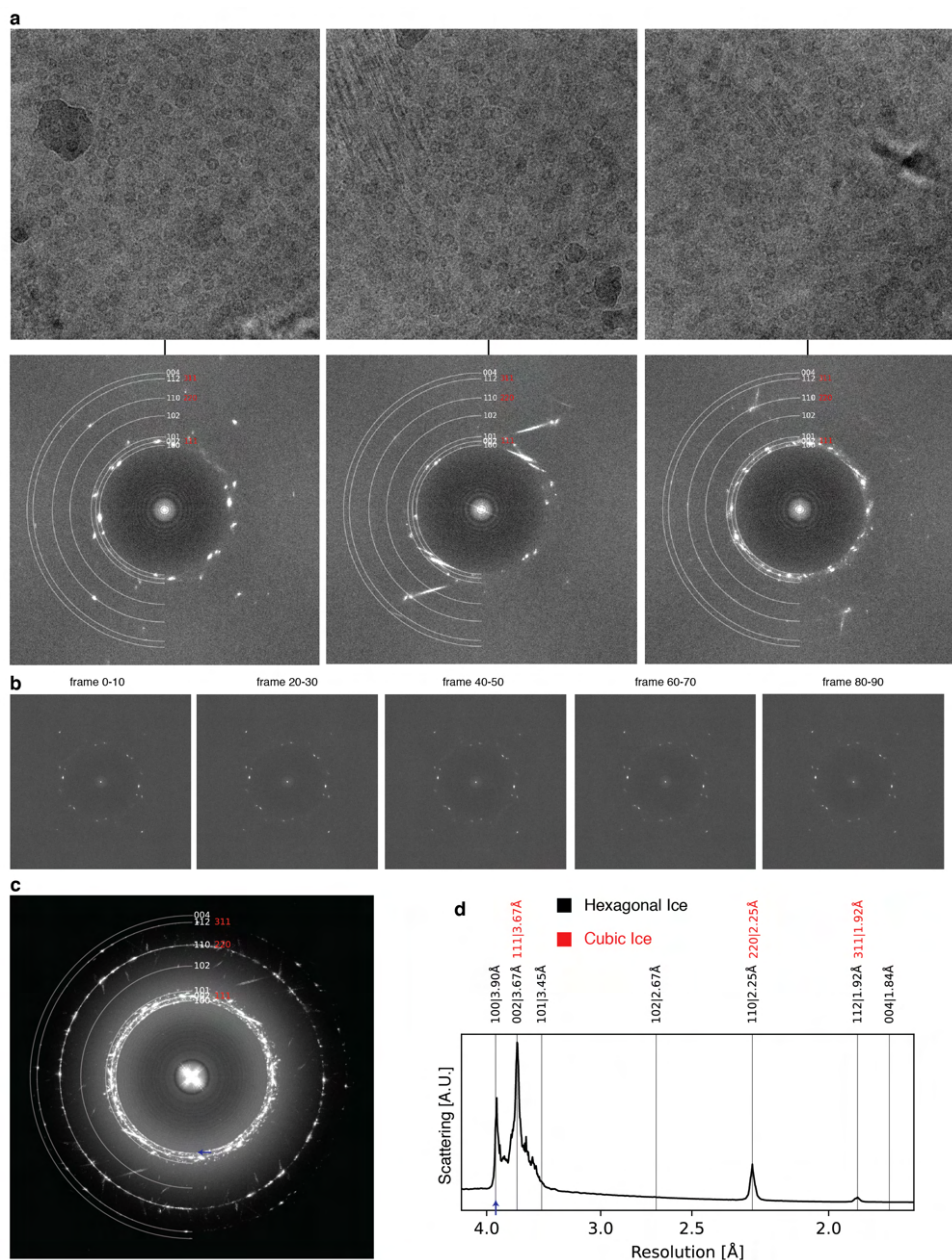Figure 2.12: Analysis of icy micrographs from ApoFtn dataset. (a) Micrograph (top) and power spectrum (bottom) for three examples with crystalline ice. Solid rings indicate expected Bragg spacings diffraction spots for hexagonal ice (white) and cubic ice (red) (derived from[63]). (b) Power spectra for frame subsets from the first example in (a) indicate no visible intensity changes of ice reflections throughout the exposure, suggestive of non-vitreous sections in the nanochannels and not surface ice contamination. (c) Maximum intensity projection of all power spectra from the entire ApoFtn dataset. (d) Rotational average of (c) shows four discernible diffraction peaks, including the (100) reflection (3.90 Å) unique for hexagonal ice (blue arrow).
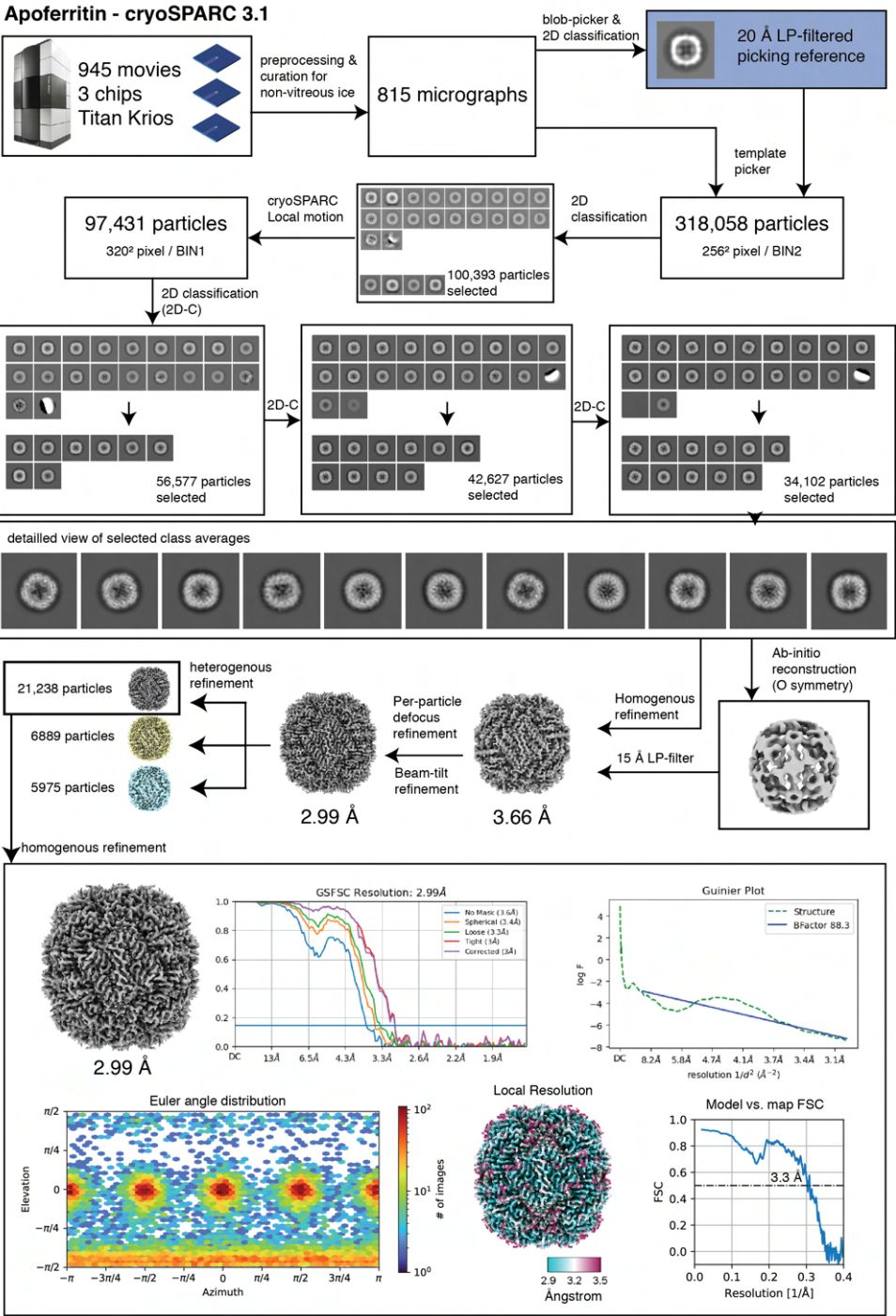
Figure 2.13: Single-particle analysis workflow for *P. furiosus apoferritin* (ApoFtn)
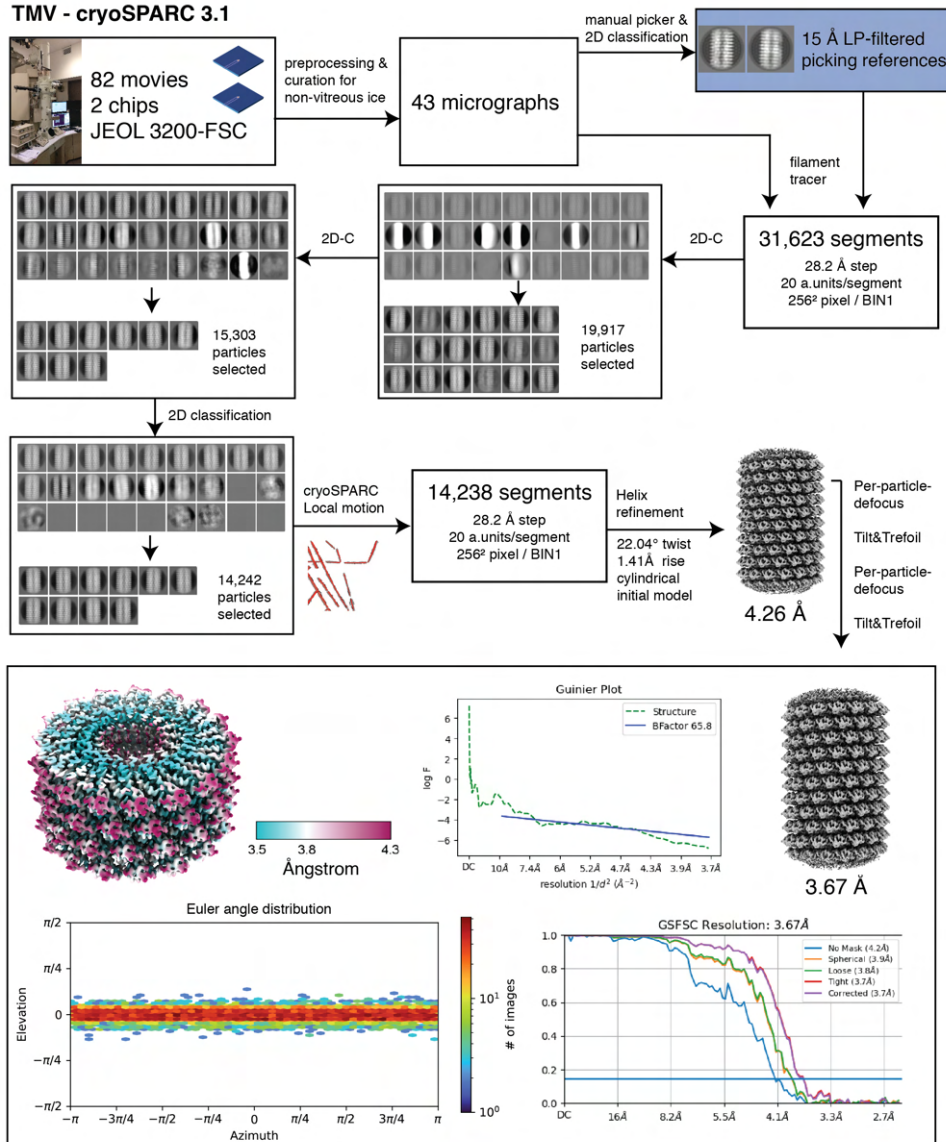
Figure 2.14: Single-particle analysis workflow for tobacco mosaic virus (TMV)
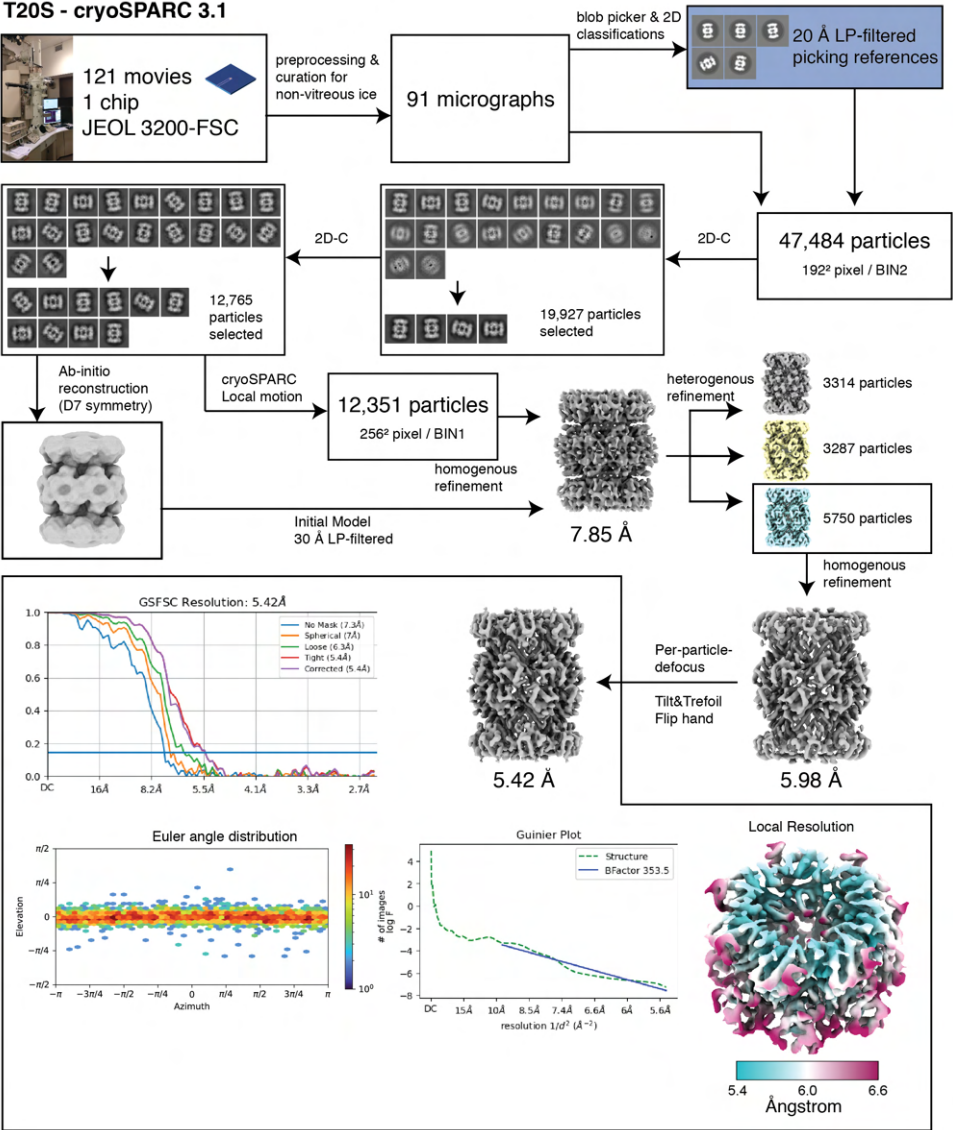
Figure 2.15: Single-particle analysis workflow for *T. acidophilum* T20S proteasome (T20S)
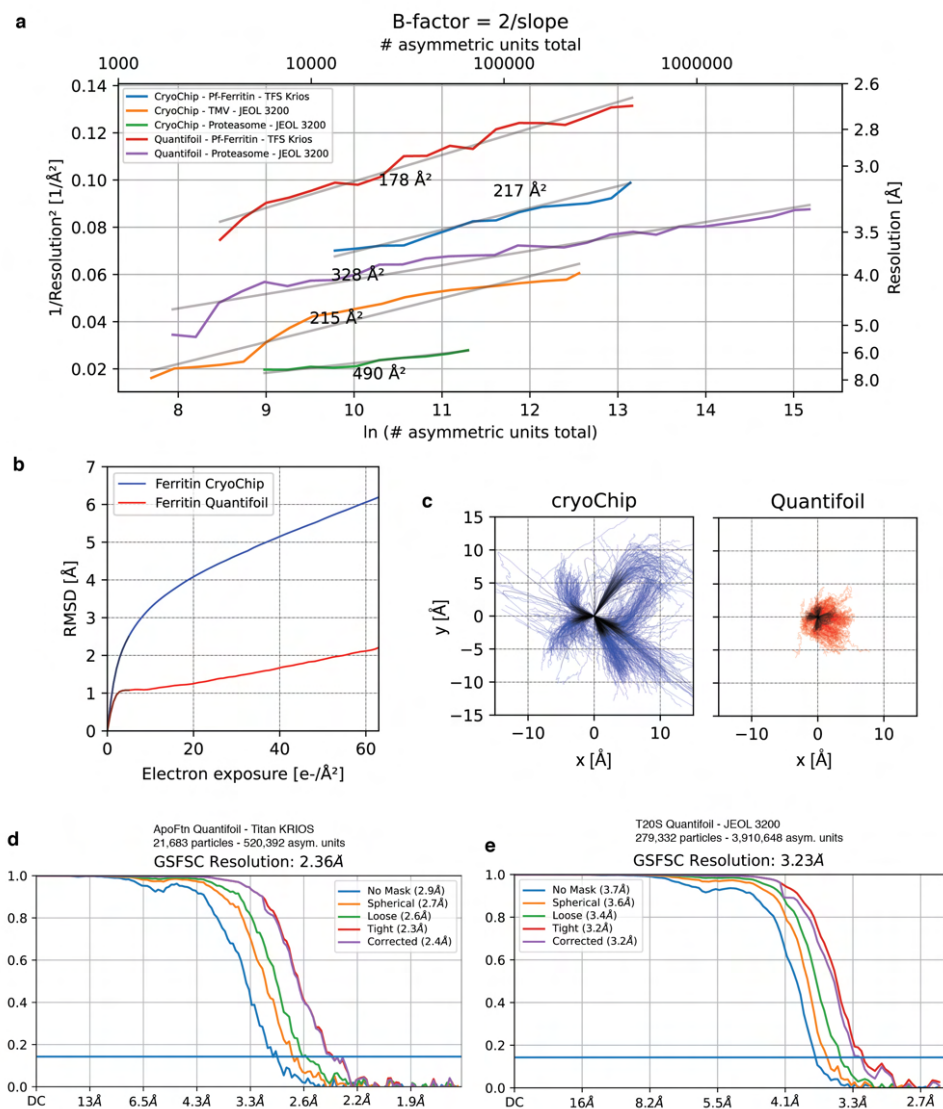
Figure 2.16: Comparison of cryoChip datasets with reference datasets on conventional holey carbon (QF). (a) B-factor plots for ApoFtn, T20S and TMV from cryoChip and holey carbon (T20S and ApoFtn) datasets obtained using the same acquisition settings and processing workflows. (b) Motion trajectories for ApoFtn datasets from cryoChip and holey carbon grids. The overall specimen movement for cryoChips is higher, but initial movement at 0-5 e⁻Å/^{2} is on the order of 0.5 pixel/frame. (c) Trajectories from (b). Fourier Shell Correlation (FSC) curves for 3D reconstructions of (d) ApoFtn and (e) T20S from holey carbon reference datasets.

2
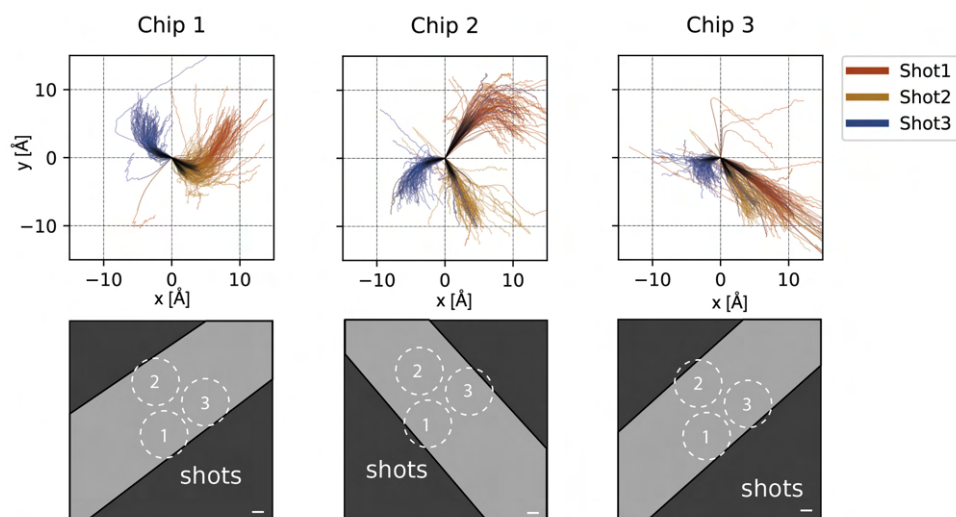
Figure 2.17: Beam-induced motion of ApoFtn dataset in cryoChips. Motion trajectories separated by chip and colour-coded by acquisition location reveal preferred directions of movement specific for each shot position in the nanochannels. Pictograms show the direction of nanochannels in the acquisition movies. Scale bar: 100 nm.

2
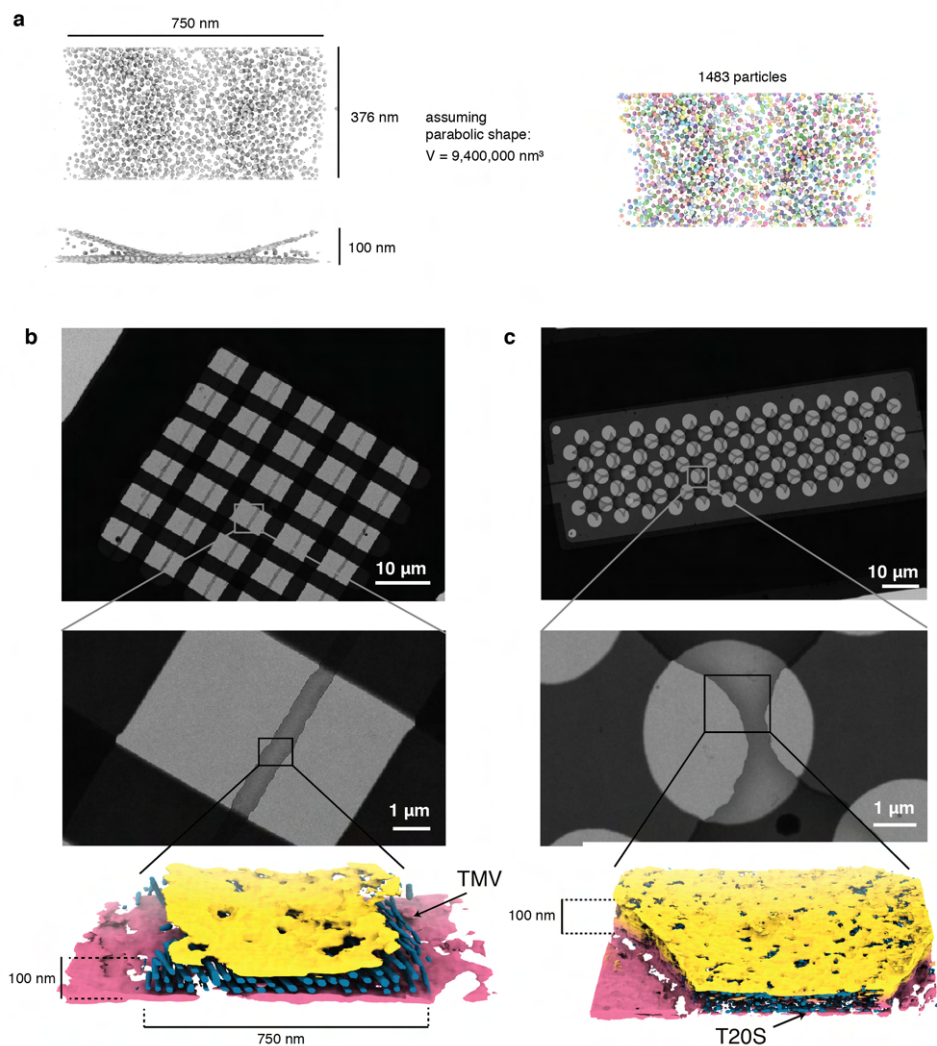
Figure 2.18: Tomograms of cryoChips with TMV and T20S. (a) Schematic illustration of volume determination and particle counting for a part of the ApoFtn tomogram. Segmented ApoFtn particles are shown as spheres. Low (top) and medium (center) magnification micrographs and segmented tomograms of nanochannels filled with (b) TMV (EMD-12914) and (c) T20S proteasome (EMD-12917). The cryoChip for T20S has a clover-patterned design of nanochannel and circular observation windows.
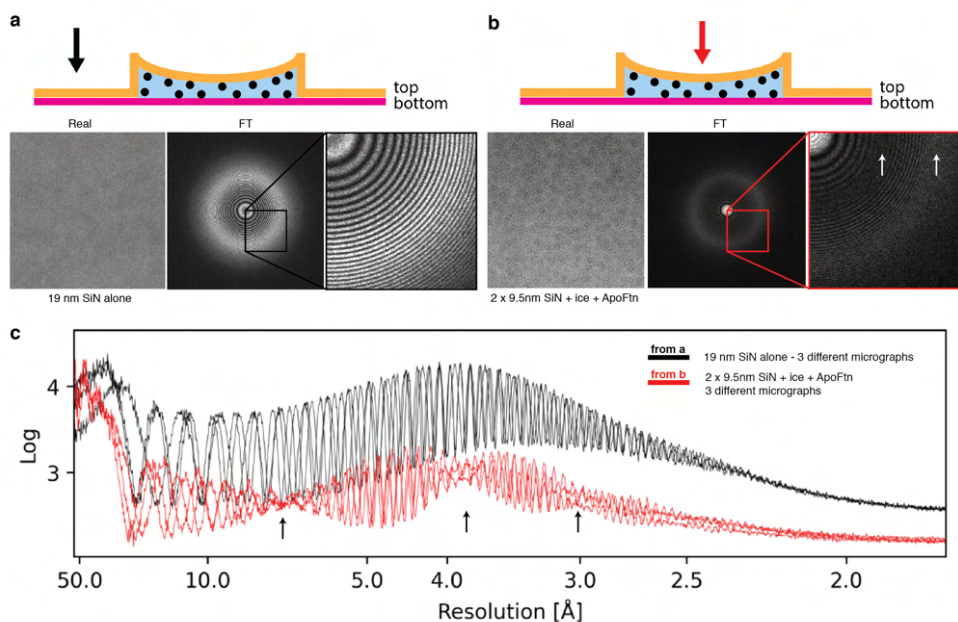
Figure 2.19: Spectral analysis of SiN. Real space image, Fourier amplitude spectrum and detail of Fourier amplitude spectrum for (a) 19 nm silicon nitride membrane adjacent to a nanochannel (b) ApoFtn in vitreous ice in a silicon nitride nanochannel. Arrows in the schematics indicate imaging position. The top and bottom membrane of the nanochannel are between 0 and 100 nm apart, leading to interference of their CTFs resulting in nodes (arrows) in their overall Thon ring pattern. (c) Radially averaged power spectra of three micrographs at different defoci show a broad peak of scattering intensity for amorphous SiN at about 3.9 Å. CTF nodes are visible in the nanochannel Fourier spectra (arrows).
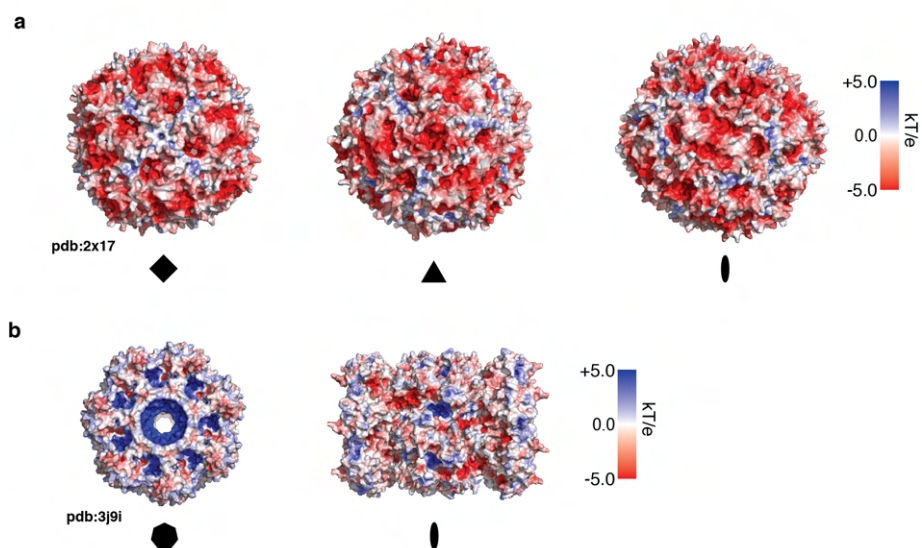
Figure 2.20: Electrostatic potential maps for (a) ApoFtn, displayed along four-fold, three-fold and two-fold axes and (b) T20S proteasome displayed along the seven-fold and two-fold axes. Electrostatic potentials have been computed from PDB ID 2x17[15] and PDB ID 3j9i[17] using Adaptive Poisson Boltzmann server (ABPS)[64].

2.7 Elife Evaluation

This article was peer-reviewed by the Elife journal[1]. The Editor's evaluation is printed along with the abstract for further context and is reproduced below:

Public Evaluation Summary

Sample preparation for single-particle electron cryo-microscopy (cryo-EM) remains a bottleneck of this technique. The sample ice thickness cannot be accurately controlled, molecules may display strongly preferred orientations that make more elaborate data collection schemes necessary, and the sample may degrade at the air-water interface before it is finally frozen. In their pioneering work, the authors describe a prototype of a new microfluidic device that addresses some of these problems, including a refreshingly objective and critical discussion about the pros and cons of this novel approach. While some development will be required for this method to become mainstream, it has the potential to become a powerful alternative to the conventional workflow of single-particle cryo-EM, enabling full automation and making sample preparation highly reproducible.

Reviewer comments

Detailed reviewer comments and author responses can be found online:
https://www.biorxiv.org/content/10.1101/2021.05.25.444805v1review

2.8 References

1. Huber, S. T., Sarajlic, E., Huijink, R., Weis, F., Evers, W. H. & Jakobi, A. J. Nanofluidic chips for cryo-EM structure determination from picoliter sample volumes. *Elife* **11,** e72629 (2022).

2. Peckys, D. B., Veith, G. M., Joy, D. C. & de Jonge, N. Nanoscale Imaging of Whole Cells Using a Liquid Enclosure and a Scanning Transmission Electron Microscope. *PLOS One* **4,** 1–7 (2009).

3. Nishiyama, H., Suga, M., Ogura, T., Maruyama, Y., Koizumi, M., Mio, K., Kitamura, S. & Sato, C. Atmospheric scanning electron microscope observes cells and tissues in open medium through silicon nitride film. *J. Struct. Biol.* **169,** 438–449 (2010).

4. Ciarlo, D. R. Silicon Nitride Thin Windows for Biomedical Microdevices. *Biomed. Microdevices* **4,** 63–68 (2002).

5. Ring, E. A., Peckys, D. B., Dukes, M. J., Baudoin, J. P. & de Jonge, N. Silicon nitride windows for electron microscopy of whole cells. *J. Microsc.* **243,** 273–283 (2011).

6. Gorelick, S., Alan, T., Sadek, A. Z., Tjeung, R. T. & Marco, A. d. Nanofluidic and monolithic environmental cells for cryogenic microscopy. *Nanotechnology* **30,** 085301 (2019).

7. Mirsaidov, U. M., Zheng, H., Casana, Y. & Matsudaira, P. Imaging protein structure in water at 2.7 nm resolution by transmission electron microscopy. *Biophys. J.* **102,** L15–l17 (2012).

8. Cameron Varano, A., Rahimi, A., Dukes, M. J., Poelzing, S., McDonald, S. M. & Kelly, D. F. Visualizing virus particle mobility in liquid at the nanoscale. *Chem. Comm.* **51,** 16176–16179 (2015).

9. Koo, K., Park, J., Ji, S., Toleukhanova, S. & Yuk, J. M. Liquid-Flowing Graphene Chip-Based High-Resolution Electron Microscopy. en. *Advanced Materials* **33,** 2005468. ISSN: 0935-9648, 1521-4095 (Jan. 2021).

10. Angell, C. A. Liquid Fragility and the Glass Transition in Water and Aqueous Solutions. *Chem. Rev.* **102,** 2627–2650 (2002).

11. Carragher, B., Cheng, Y., Frost, A., Glaeser, R., Lander, G., Nogales, E. & Wang, H. Current outcomes when optimizing 'standard' sample preparation for single-particle cryo-EM. *J. Microsc.* **276,** 39–45 (2019).

12. Yesibolati, M. N., Laganá, S., Kadkhodazadeh, S., Mikkelsen, E. K., Sun, H., Kasama, T., Hansen, O., Zaluzec, N. J. & Mølhave, K. Electron inelastic mean free path in water. en. *Nanoscale* **12,** 20649–20657 (2020).

13. Washburn, E. W. The Dynamics of Capillary Flow. *Physical Review* **17,** 73–283 (1921).

14. Schorb, M., Haberbosch, I., Hagen, W. J. H., Schwab, Y. & Mastronarde, D. N. Software tools for automated transmission electron microscopy. en. *Nat. Methods* **16,** 471–477 (2019).

15. Kasyutich, O., Ilari, A., Fiorillo, A., Tatchev, D., Hoell, A. & Ceci, P. Silver ion incorporation and nanoparticle formation inside the cavity of Pyrococcus furiosus ferritin: structural and size-distribution analyses. en. *J. Am. Chem. Soc.* **132,** 3621–3627 (2010).

16. Fromm, S. A., Bharat, T. A. M., Jakobi, A. J., Hagen, W. J. H. & Sachse, C. Seeing tobacco mosaic virus through direct electron detectors. en. *J. Struct. Biol.* **189,** 87–97 (2015).

17. Li, X., Mooney, P., Zheng, S., Booth, C. R., Braunfeld, M. B., Gubbens, S., Agard, D. A. & Cheng, Y. Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. en. *Nat. Methods* **10,** 584–590 (2013).

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

19. Rosenthal, P. B. & Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* **333,** 721–745 (2003).

20. Vinothkumar, K. R. & Henderson, R. Single particle electron cryomicroscopy: trends, issues and future perspective. en. *Q. Rev. Biophys.* **49,** e13 (2016).

21. Iudin, A., Korir, P. K., Salavert-Torres, J., Kleywegt, G. J. & Patwardhan, A. EMPIAR: a public archive for raw electron microscopy image data. en. *Nat. Methods* **13,** 387–388 (2016).

22. Snijder, J., Borst, A., Dosey, A., Walls, A., Burrell, A., Reddy, V., Kollman, J. & Veesler, D. Vitrification after multiple rounds of sample application and blotting improves particle density on cryo-electron microscopy grids. *J. Struct. Biol.* **198,** 38–42 (2017).

23. Yip, K. M., Fischer, N., Paknia, E., Chari, A. & Stark, H. Atomic-resolution protein structure determination by cryo-EM. *Nature* **587,** 157–161 (2020).

24. D'Imprima, E., Floris, D., Joppe, M., Sánchez, R., Grininger, M. & Kühlbrandt, W. Protein denaturation at the air-water interface and how to prevent it. en. *eLife* **8** (2019).

25. Noble, A. J., Wei, H., Dandey, V. P., Zhang, Z., Tan, Y. Z., Potter, C. S. & Carragher, B. Reducing effects of particle adsorption to the air–water interface in cryo-EM. *Nat. Methods* **15,** 793–795 (2018).

26. Glaeser, R. M. & Han, B.-G. Opinion: hazards faced by macromolecules when confined to thin aqueous films. *Biophys. Rep.* **3,** 1–7 (2017).

27. Glaeser, R. M. Proteins, Interfaces, and cryo-EM grids. en. *Curr. Opin. Colloid Interface Sci.* **34,** 1–8 (2018).

28. Naydenova, K., Peet, M. J. & Russo, C. J. Multifunctional graphene supports for electron cryomicroscopy. en. *Proc. Natl. Acad. Sci. U. S. A.* **116,** 11718–11724 (2019).

29. Tan, Y. Z., Baldwin, P. R., Davis, J. H., Williamson, J. R., Potter, C. S., Carragher, B. & Lyumkis, D. Addressing preferred specimen orientation in single-particle cryo-EM through tilting. en. *Nat. Methods* **14,** 793–796 (2017).

30. Ravelli, R. B. G., Nijpels, F. J. T., Henderikx, R. J. M., Weissenberger, G., Thewessem, S., Gijsbers, A., Beulen, B. W. A. M. M., López-Iglesias, C. & Peters, P. J. Cryo-EM structures from sub-nl volumes using pin-printing and jet vitrification. en. *Nat. Commun.* **11,** 2563 (2020).

31. Jain, T., Sheehan, P., Crum, J., Carragher, B. & Potter, C. S. Spotiton: A prototype for an integrated inkjet dispense and vitrification system for cryo-TEM. *J. Struct. Biol.* **179,** 68–75 (2012).

32. Arnold, S. A., Albiez, S., Bieri, A., Syntychaki, A., Adaixo, R., McLeod, R. A., Goldie, K. N., Stahlberg, H. & Braun, T. Blotting-free and lossless cryo-electron microscopy grid preparation from nanoliter-sized protein samples and single-cell extracts. *J. Struct. Bio.* **197,** 220–226 (2017).

33. Drulyte, I., Johnson, R. M., Hesketh, E. L., Hurdiss, D. L., Scarff, C. A., Porav, S. A., Ranson, N. A., Muench, S. P. & Thompson, R. F. Approaches to altering particle distributions in cryo-electron microscopy sample preparation. *Acta Cryst. D Struct. Biol.* **74,** 560–571 (2018).

34. Klebl, D. P., Gravett, M. S., Kontziampasis, D., Wright, D. J., Bon, R. S., Monteiro, D. C., Trebbin, M., Sobott, F., White, H. D., Darrow, M. C., Thompson, R. F. & Muench, S. P. Need for Speed: Examining Protein Behavior during CryoEM Grid Preparation at Different Timescales. *Structure* **28,** 1238–1248.e4 (2020).

35. Glaeser, R. M. Preparing Better Samples for Cryo–Electron Microscopy: Biochemical Challenges Do not End with Isolation and Purification. *Annu. Rev. Biochem.* **90** (2021).

36. Wei, H., Dandey, V. P., Zhang, Z., Raczkowski, A., Rice, W. J., Carragher, B. & Potter, C. S. Optimizing "self-wicking" nanowire grids. en. *J. Struct. Biol.* **202,** 170–174 (2018).

37. Kasas, S., Dumas, G., Dietler, G., Catsicas, S. & Adrian, M. Vitrification of cryoelectron microscopy specimens revealed by high-speed photographic imaging. *J. Microsc.* **211,** 48–53 (2003).

38. R., H. The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules. *Q. Rev. Biophys* **28,** 171–193 (1995).

39. Zhang, Y., Tammaro, R., Peters, P. J. & Ravelli, R. B. G. Could Egg White Lysozyme be Solved by Single Particle Cryo-EM? *J. Chem. Inf. Model.* **60,** 2605–2613 (2020).

40. Tichelaar, W., Hagen, W. J., Gorelik, T. E., Xue, L. & Mahamid, J. TEM bright field imaging of thick specimens: nodes in Thon ring patterns. *Ultramicroscopy* **216,** 113023 (2020).

41. Russo, C. J. & Passmore, L. A. Ultrastable gold substrates for electron cryomicroscopy. *Science* **346,** 1377–1380 (2014).

42. Norde, W. Adsorption of proteins from solution at the solid-liquid interface. *Adv. Colloid Interface Sci.* **25,** 267–340 (1986).

43. Naydenova, K., Jia, P. & Russo, C. J. Cryo-EM with sub–1 Å specimen movement. *Science* **370,** 223–226 (2020).

44. Studier, F. W. & William Studier, F. Protein production by auto-induction in high-density shaking cultures. *Protein Expr. Purif.* **41,** 207–234 (2005).

45. Gooding Jr, G. V. & Hebert, T. T. A simple technique for purification of tobacco mosaic virus in large quantities. en. *Phytopathology* **57,** 1285 (1967).

46. Yu, Y., Smith, D. M., Kim, H. M., Rodriguez, V., Goldberg, A. L. & Cheng, Y. Interactions of PAN's C-termini with archaeal 20S proteasome and implications for the eukaryotic proteasome–ATPase interactions. *Embo J.* **29,** 692–702 (2010).

47. Zheng, S. Q., Palovcak, E., Armache, J.-P., Verba, K. A., Cheng, Y. & Agard, D. A. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. en. *Nat. Methods* **14,** 331–332 (2017).

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

49. Mastronarde, D. N. & Held, S. R. Automated tilt series alignment and tomographic reconstruction in IMOD. en. *J. Struct. Biol.* **197,** 102–113 (2017).

50. Chen, M., Dai, W., Sun, S. Y., Jonasch, D., He, C. Y., Schmid, M. F., Chiu, W. & Ludtke, S. J. Convolutional neural networks for automated annotation of cellular cryo-electron tomograms. *Nat. Methods* **14,** 983–985 (2017).

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

52. Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. en. *J. Struct. Biol.* **152,** 36–51 (2005).

53. Stagg, S. M., Noble, A. J., Spilman, M. & Chapman, M. S. ResLog plots as an empirical metric of the quality of cryo-EM reconstructions. *J. Struct. Biol.* **185,** 418–426 (2014).

54. Cardone, G., Heymann, J. B. & Steven, A. C. One number does not fit all: Mapping local variations in resolution in cryo-EM reconstructions. *J. Struct. Biol.* **184,** 226–236 (2013).

55. Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Cryst. D Struct. Biol.* **66,** 213–221 (2010).

56. Emsley, P., Lohkamp, B., Scott, W. & Cowtan, K. Features and Development of\textit {Coot}. *Acta Crystallogr D Struct. Biol.* (2010).

57. Chen, V. B., Arendall III, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. *MolProbity*: all-atom structure validation for macromolecular crystallography. *Acta Cryst Section D Struct. Biol.* **66** (2010).

58. Zivanov, J., Nakane, T., Forsberg, B. O., Kimanius, D., Hagen, W. J., Lindahl, E. & Scheres, S. H. New tools for automated high-resolution cryo-EM structure determination in RELION-3. en. *Elife* **7** (2018).

59. Weis, F., Beckers, M., von der Hocht, I. & Sachse, C. Elucidation of the viral disassembly switch of tobacco mosaic virus. en. *EMBO Rep.* **20,** e48451 (2019).

60. Righetto, R. D., Anton, L., Adaixo, R., Jakob, R. P., Zivanov, J., Mahi, M.-A., Ringler, P., Schwede, T., Maier, T. & Stahlberg, H. High-resolution cryo-EM structure of urease from the pathogen Yersinia enterocolitica. en. *Nat. Commun.* **11,** 5101 (2020).

2

61. Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. en. *Nat. Methods* **17,** 261–272 (2020).

62. Ftouni, H., Blanc, C., Tainoff, D., Fefferman, A. D., Defoort, M., Lulla, K. J., Richard, J., Collin, E. & Bourgeois, O. Thermal conductivity of silicon nitride membranes is not sensitive to stress. *Physical Review B* **92,** 125439 (2015).

63. Kumai, M. Hexagonal and Cubic Ice at Low Temperatures. *Journal of Glaciology* **7,** 95–108 (1968).

64. Jurrus, E. *et al.* Improvements to the APBS biomolecular solvation software suite. *Protein Sci. : a publication of the Protein Society* **27,** 112–128 (2018).

2

3

Towards a time-resolved cryo-EM flash photolysis setup

Single-particle cryogenic electron microscopy (cryo-EM) allows for high-resolution structural analysis of proteins immobilized in thin ice layers. With sufficiently fast sample preparation, transient states of a biochemical reaction can be captured with a technique called time-resolved EM (trEM). In this chapter, we design and build a setup that can initiate reactions with light flashes and arrest them after milliseconds by rapid plunge-freezing.

3.1 Graphical Abstract
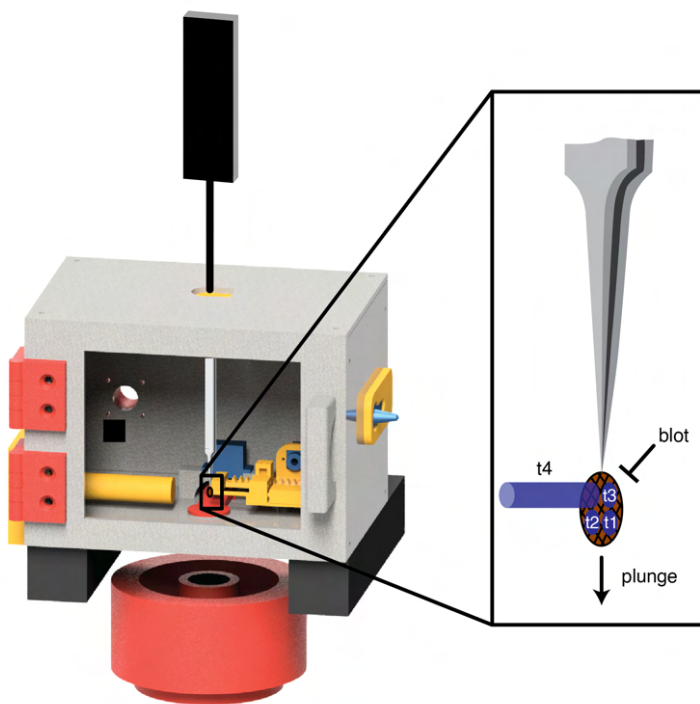
3.2 Introduction

Single-particle cryo-EM involves preparing protein samples in thin ice layers. The freezing process in sample preparation occurs in less than a millisecond[1], which in principle allows the capture of transient conformational states in a method called time-resolved cryo-EM (trEM). However, the conventional blotting and plunging steps typically involved need seconds, negating this advantage. A short review of previous sample preparation setups to achieve time-resolution is provided in Chapter 1. Here, we set out to design and build a sample preparation setup for trEM based on the long-established blotting method to achieve thin ice layers - in combination with flash-photolysis of caged chemical compounds to initiate reactions.

3.3 Design of a flash-photolysis setup

In trEM, a reaction is initiated in a controlled way and arrested after a specified time delay to catch intermediate states present at that time point. One important requirement is to initiate the reaction simultaneously throughout the system. Currently, two main methods are used for this purpose: The first involves mixing of two liquids together, either directly on an electron micoscopy grid[2,3] or in a mixing device just before application to the grid[4–7]. The second method is flash-photolysis, which achieves synchronous activation by illumination of photolabile compounds such as caged ATP prior to plunge freezing[8–11].

Several reasons prompted us to build a trEM setup based on flash-photolysis. A main topic in our research group is to study guanylate-binding proteins (GBPs), which are GTPases that perform membrane remodelling by undergoing conformational changes upon GTP binding and hydrolysis. The availability of caged GTP compounds makes this group of proteins amenable to trEM studies by flash photolysis. Another reason to choose flash-photolysis is the relative technical simplicity. Other trEM methods rely on spraying liquid onto a grid, which introduces a major obstacle in producing thin, continuous and reproducible ice layers for imaging large datasets, as well as technically complex mixing devices and often large sample consumption. In a flash-photolysis setup, long-established blotting-based thinning of ice layers can be employed, just prior to illuminating the sample with light. While blotting methods have some problems with reproducibility (see Chapter 2), they can lead to large and thin ice layers, with the possibility to achieve thin ice on the entire 3 mm diameter grid.

The possibility of large sample areas with good ice conditions opens up the chance to access several time-delay points on the same sample by illuminating different regions at different time points. To allow sampling of different time points, we decided to use four LED light sources that can be independently triggered to illuminate four different regions on the same grid, with the possibility of future extension to more spots. The flash-photolysis approach is also fully compatible with the cryoChip samples described in Chapter 2. In this case, the protein would be loaded into the nanofluidic channels of the chip by capillary action. The blotting step can be skipped, and the ice thickness accurately controlled by the channel geometry, ranging in height from 0 to 100 nanometers. The current chip design features a single 50 μm sample area, so only a single area can be illuminated with this setup. The chip design could be adapted to contain multiple, widely spaced areas for separate illumination with the independent LED sources.

Activation with light can be fast for compounds with good quantum yield and/or a strong enough light source, which makes the physical motion of flash-freezing the sample in liquid ethane the time-limiting step. An inherent limit with photocleavable compounds however is the release half-time, which is around 7 milliseconds for the typically used photocleavable ATP compound NPE-ATP[12]. This sets the order of magnitude of time-resolution achievable with this setup to around 10 milliseconds. We decided to integrate an industry-grade linear motor (PS01-23x160H-HP-R, LinMot, Switzerland), capable of 5.2 m/s velocity and 138 Newtons of force. For an approximate weight of the moving slider with a tweezer of 200 grams, the theoretically achievable acceleration is 690 m/s**2 or ~70 g (1 g = earth's gravitational acceleration).

3.3.1 Theoretical Considerations

The initiation of the release reaction of a photocaged compound is facilitated using a fiber-coupled LED light source. The photons absorbed generate excited intermediates that decay to release the ligand during 'dark' reaction steps[12]. These have characteristic rates and quantum yield (the fraction of ligand released per absorbed photons). The efficiency of a photocage is measured by the ability to absorb light (i.e. the molar absorption coefficient or the cross-section) and by the quantum yield. The following elaboration details the estimation of the necessary LED power to achieve a specific concentration of reactants such as ATP or GTP - in this example for 1-(2-nitrophenyl)ethyl (NPE)-ATP.
 Given parameters:

- Extinction coefficient (ε): $430\,\mathrm{M^{-1}cm^{-1}}$ [Josts et al. 2018[13]]
- Quantum yield (η): 0.65 [Josts et al. 2018[13]]
- LED wavelength (λ): $365\,\mathrm{nm}$
- Beam diameter (d): $0.5\,\mathrm{mm}$
- Path length within the sample (l): $20\,\mathrm{nm}$ or $2\times10^{-6}\,\mathrm{cm}$
- Concentration of the sample (c): $10\,\mathrm{mM}$ or $0.01\,\mathrm{M}$
- Target reaction rate (R): $0.1\,\mathrm{mM/ms}$
- Avogadro's number (N_A): $6.022\times10^{23}\,\mathrm{mol^{-1}}$

The absorbance (A) of the caged compound within a thin layer is determined using the Beer-Lambert Law:

$$A = \varepsilon \cdot c \cdot l \tag{3.1}$$
$$= (430\,\mathrm{M^{-1}cm^{-1}}) \cdot (0.01\,\mathrm{M}) \cdot (2\times10^{-6}\,\mathrm{cm})$$
$$= 8.6\times10^{-6} \tag{3.2}$$

The energy associated with each photon (E_{photon}) is given by:

$$E_{\mathrm{photon}} = \frac{hc}{\lambda} \tag{3.3}$$

Calculation of the molar photon flux (Φ), indicative of the photon count per second emitted by the light source:

$$\Phi = \frac{P}{E_{\text{photon}} \cdot N_A} \tag{3.4}$$

The molar photon flux absorbed by the sample given by the Beer-Lambert law (Φ_{abs}):

$$\Phi_{\text{abs}} = \Phi \cdot [1 - 10^{-A}] \tag{3.5}$$

The reaction rate (R), expressed in moles per liter per second, is calculated based on Φ_{abs}, the irradiated volume, and the quantum yield of the compound:

$$R = \frac{\Phi_{\text{abs}} \cdot \eta}{V} = \frac{P \cdot [1 - 10^{-A}] \cdot \eta}{E_{\text{photon}} \cdot V \cdot N_A} \tag{3.6}$$

Rearranging the above to solve for the necessary laser power (P):

$$P = \frac{R \cdot E_{\text{photon}} \cdot V \cdot N_A}{[1 - 10^{-A}] \cdot \eta} \tag{3.7}$$

Substituting the known values to calculate the required power for a desired uncaged compound creation rate of 0.1 mM per millisecond:

$$P = 10.56 \, \text{mW} \tag{3.8}$$

To be useful in experiments to study kinetics, the dark reactions should be fast compared to the timescale of downstream events studied. In the present example, applying a 10 ms illumination with a laser power of approximately 10 mW enables the generation of a 1 mM concentration of uncaged ATP in the sample. This duration is comparable to the half-release time of 7 ms for the commonly utilized caged compound NPE-ATP[12]. Consequently, our experimental setup incorporates four fiber-coupled LEDs (M365FP1, Thorlabs), each delivering 15 mW of power. These LEDs are independently controllable, allowing for the creation of distinct time-delay regions on the same grid.

3.3.2 Setup description

The main components of the setup are arranged around the position where a tweezer holds an EM grid prior to plunge-freezing [Figure 3.1A]. A 3D printed humidity chamber keeps the grid at 100% humidity. The endpoint of the LED fibers is placed on the left of the grid. An entry point for a pipet is on right side of the humidity chamber to allow pipetting a sample onto the grid. The blotter arm controlled by a servo motor is on the right side of the grid to blot away excess sample liquid with a strip of filter paper.

Below the grid is a shutter controlled by a servo motor [Figure 3.1A], which only opens for plunge-freezing. Below the humidity chamber is a liquid nitrogen dewar with a ethane cup. The height of the ethane cup can be controlled by a mechanical lift on a belt-driven linear unit. The position of the tweezer holding the grid is controlled by a fast linear motor.

An external humidifier chamber with an integrated aquarium "fogger" produces humidified air. It is attached to a water reservoir with controlled flow using a T-junction. A

3
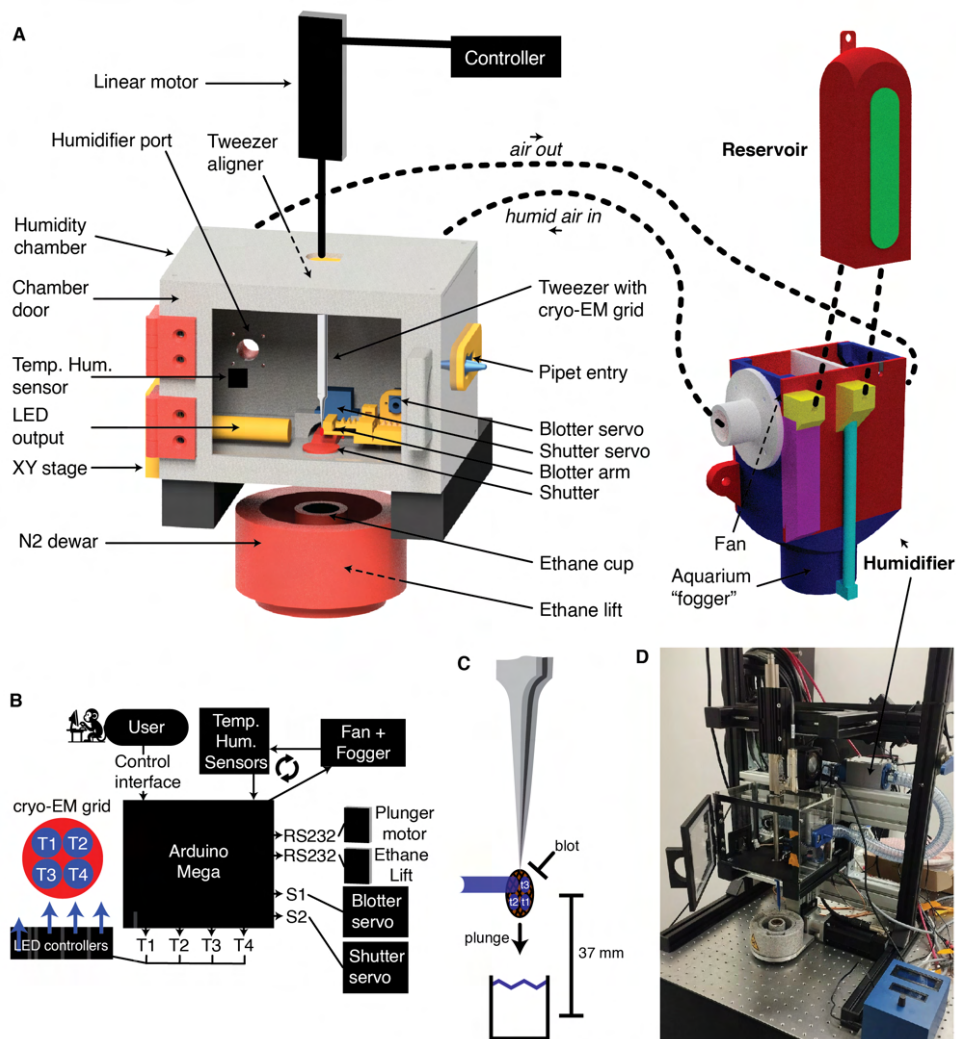
Figure 3.1: | **A time-resolved cryo-EM setup employing flash photolysis.** (A) Schematic overview of the setup. Humidity chamber with hardware for sample blotting, illumination and trajectory execution. Humidity in the chamber is generated by a humidifier attached to a water reservoir. (B) The setup is controlled by a Arduino Mega, which triggers LEDs, controls the servo motors and sends commands to the linear motors of plunging arm and ethane lift, and controls humidity (C) The distance of illumination position and end position of plunge freezing is 37 mm. (D) Image of the fully assembled setup.

switchable fan pumps humidified air through tubing into the humidity chamber and forms a closed loop with a second tube coming out of the chamber.

An Arduino Mega 2560 is used to control the setup. Digital on/off signals are sent to control the four LEDs. PWM signals control the servo motor for the blotter to set distances for touching the grid and retracting. The humidifier and fan are switched with digital on/off signals through a relay. The ethane lift and linear motor for the plunging arm require more complicated signaling, as explained in a later section.

The blot mechanism uses a spring mechanism to hold a ~1 cm wide strip of filter paper in place. The user can push the spring mechanism to release the paper and move it forward for every blot to use a different area on the filter paper [Figure 3.2A]. The mechanism is designed to hold the paper with a few degrees of tilt accounting for the angle of the tweezer tip that is holding the grid.

The user interface guides the user through a four-step program: (1) 'Load tweezer' moves the tweezer attachment point of the plunging arm in a position below the humidity chamber to allow attaching a tweezer with a cryo-EM grid. (2) 'Raise tweezer' moves the grid and tweezer up into the humidity chamber to a position where the sample can be applied by pipetting. (3) 'Raise ethane' moves the ethane cup up to a position just below the humidity chamber to minimize the distance for the grid to travel during plunge-freezing. (4) 'Plunge' activates the blot mechanism, blots away excess liquid, illuminates four spots on the grid with LEDs, and then plunges the grid into liquid ethane. Settings for all these processes can be adjusted in sub-menus [Figure 3.2B].

3.3.3 Communication with ethane lift and plunging arm

Communication to both the lift and linear motor works through the LinRS protocol of the manufacturer (LinMot) [see Supplementary Information]. Messages are compiled for a command and sent through a serial connection using the RS-232 standard to the controller, which will execute the command.

Upon startup of the lift and motor, commands for initialisation and motor homing are sent to the controllers. They establish a consistent homing position and put the motors in operative mode ready to receive commands. All other commands needed for this setup are GoTo commands to move the arm/lift in a certain position with a specified maximum speed, acceleration and deceleration. The Arduino function for these commands can be found in the supplementary information of this chapter.

One peculiarity of communication with the arm/lift is the global variable 'Nibble-counter'. It is a boolean variable functioning as an internal safety mechanism to prevent executing the same command twice. It needs to be flipped after every executed command to make sure a different value is sent each time.

3.3.4 Trajectory programming and analysis

Using this GoTo function, all necessary trajectories can be programmed by setting speed, acceleration and deceleration. We performed preliminary testing on the arm to verify that trajectories are executed correctly. The LinMot controller is capable of saving time curves that can be analysed.

For this test, we decided to accelerate for 75 mm, keep the speed constant for 45 mm and decelerate for 15 mm, while planning maximum velocities ranging from 1 m/s to 4 m/s

3
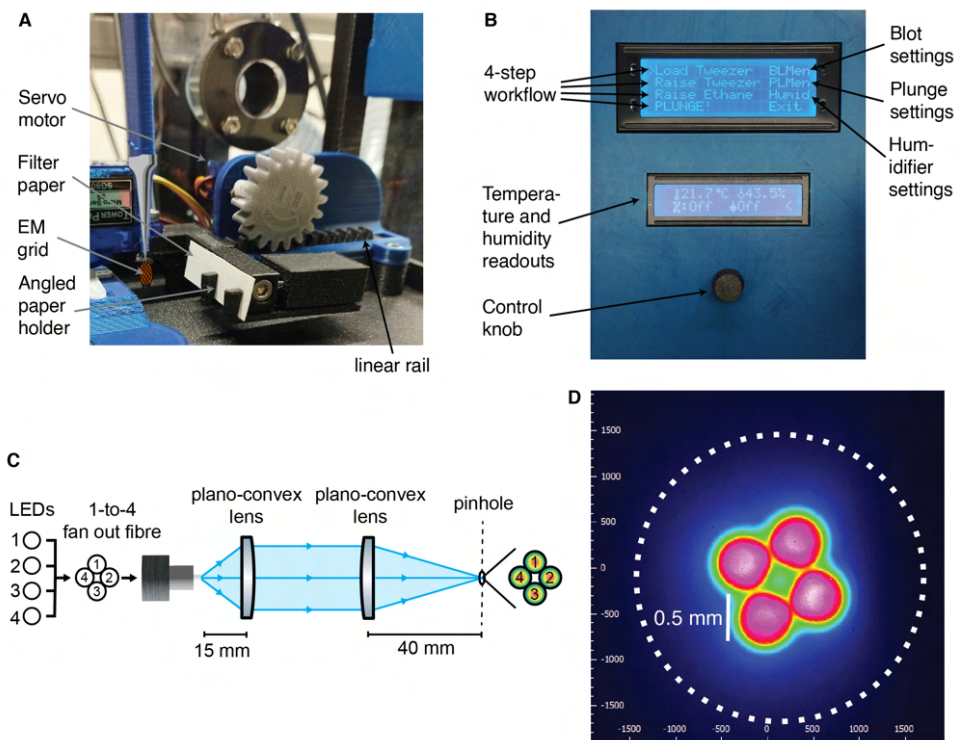
Figure 3.2: | **Details of blotting mechanism, user interface and optics system.** (A) Blotting mechanism with linear rail operated with a servo motor. The strip of filter paper is angled to help align with the angled tweezer tip that is holding the grid. (B) User interface. Workflow with four steps to control the setup: Load tweezer, Raise tweezer, Raise Ethane, PLUNGE. Settings for blotting, plunging and LED illumination can be changed in the respective submenus. (C) Optical setup for focussing four LEDs into adjacent positions. (D) Beam profile produced by the LEDs show a 0.5 mm diameter circular spot. Dotted line: 3 mm diameter EM grid.

Table 3.1: | **Plunging parameters for trajectory validation.** Acceleration, deceleration and max speed values were input into the GoTo function to execute the moves described in Figure 3.3

Max speed [m/s]	Acc. [m/s**2]	[g]	Acceleration distance [m]	Dec. [m/s**2]	[g]	Deceleration distance [m]
1	6.667	0.680	0.075	33.333	3.398	0.015
1.5	15.000	1.529	0.075	75.000	7.645	0.015
2	26.667	2.718	0.075	133.333	13.592	0.015
2.5	41.667	4.247	0.075	208.333	21.237	0.015
3	60.000	6.116	0.075	300.000	30.581	0.015
3.5	81.667	8.325	0.075	408.333	41.624	0.015
4	106.667	10.873	0.075	533.333	54.366	0.015

[Figure 3.3A]. The parameters for the GoTo function can be derived from the kinematic equations for accelerated motion and are summarised in Table 3.1.

Planned and actual trajectories matched well for the slow acceleration of up to 10 g, and the desired velocities of 1-4 m/s were reached. The deceleration portion of the trajectory showed overshoot behavior for decelerations above 10 g. The final position of the plunge was overshot by up to 5 mm. This may cause problems for sample safety, because the bottom of the ethane cup may be hit. It also indicates that the theoretically possible 70 g of acceleration is likely an overestimate and the actually achievable value is lower.
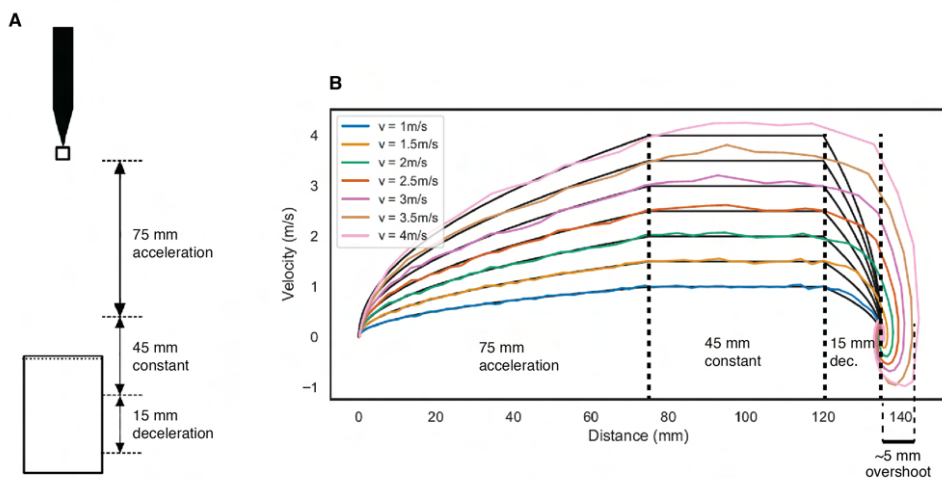
3

Figure 3.3: | **Trajectory testing.** (A) Schematic of the experiment (B) Trajectories plotted as velocity-time curves. The planned 15 mm deceleration path length was exceeded by the linear motor by up to 5 mm at a maximum tested speed of 4 m/s.

3.4 Discussion and Outlook

We described the design and construction of a time-resolved sample preparation setup. The setup can keep samples at 100% humidity, illuminate a grid with 365 nm 15 mW LED lights to initiate reactions and can execute various plunging trajectories. Measured trajectories showed limitations of the setup to accelerate the sample with more than 10 g, which would lead to plunging times of 34 ms for a path length of 30 mm with 15 mm acceleration and 15 mm deceleration distance. This is slightly slower than the time scale flash-photolysis can achieve, limited by the release half-time of photocleavable compounds such as NPE-ATP[12]. Further optimisation of the setup can help push the hardware to the theoretical limit of 70 g, which would allow a plunge time of ~13 ms over the same trajectory.

An example calculation shows that 1 mM of ATP can be released in roughly 7 milliseconds using the integrated 15 mW LED in the design. This is on the same timescale than the release half-time of this compound[12] of 7 milliseconds. Higher ATP concentrations can be reached in our setup by further increasing the concentration of caged compound. The illumination power in our setup (7.6 W/cm^2 for a 0.5 mm circle, or 1.9 W/cm^2 for a 1 mm circle) is in the same order of magnitude as in a recently described similar setup[11] using a single LED to illuminate the entire grid (1.3 W/cm^2).

The aim of the blotting-based method of generating thin ice layers in this setup is to generate large sample areas with good imaging conditions, such that four different areas can be illuminated at different time points and imaged on the same sample. Creation of good ice layers, and the involved blotting time and blotting pressure has to be benchmarked and optimized. Preliminary testing using a tobacco mosaic test sample shows that the blotting mechanism could achieve thin ice over large parts of the grid with a blot time of 8

seconds when blotting from the front or the back-side of the grid [Figure 3.4].

To assign exact timings to cryo-EM images of a reaction, the delay of sample illumination and plunging needs to be calibrated externally, for example using a high-speed camera. The movies can capture the illumination of the four different areas on the EM grid and accurately measure the timing until the grid hits the surface of the liquid ethane during plunge freezing.

After validation tests are performed, a biological system can be probed with the setup. In Chapter 2, we used apoferritin as a benchmarking protein for nanofluidic devices for sample preparation. Apoferritin undergoes disassembly at pH below 3.4[14]. Structures of intermediate states of this disassembly process could be captured using this time-resolved setup employing a NPE-caged proton compound to lower the pH. Structurally characterising intermediate assembly states of apoferritin could lead to new insights into the assembly path of the protein and to insights into biotechnological applications as nanocages.

In our research group, guanylate binding proteins (GBPs) are a main topic of study. GBPs are interferon-inducible GTPases that act as effector molecules during infection in a process called cell-autonomous immunity[15]. GBPs can form coatomers on lipid membranes in their nucleotide-bound state[16] as part of their mechanism to aid in cellular defense. Using this flash-photolysis setup, GTP can be released in the sample just prior to plunge-freezing, and the time-dependent nature of this process can be studied in the future.

3.5 Materials

Table 3.2: Equipment and Components List

Category	Details
Linear Motor	
Stator	LinMot PS01-23x160H-HP-R (Ø 23x160mm)
Slide	LinMot PL01-12x420/380-HP (Ø 12mm, L=420mm)
Drive	C1100-GP-XC-0S-000
Power Supply	Linmot SPH500-7207 (72VDC / 500W)
Cryogen Lift	
Linear Belt Drive Unit	Unimotion MTJ-40-ECO-160-S-1R (max. 160 mm stroke)
Motor Flange	VK MTJ40 72 x 72 GESM14
Coupling	GESM14 F11C F14
Servo Motor	Sangalli DSM5.22.31SZ.88
Drive	Linmot C1100-GP-XC-0S-000
Power Supply	Linmot SPH500-7207
Optical Table & Setup	
Optical Table	Thorlabs T1020D (1 m x 2 m x 310 mm Metric)
LED	365 nm, 15.5 mW Fiber-Coupled (M365FP1) x4
LED Driver	T-Cube LED (LEDD1B) x4
Fiber Bundle	1-to-4 Fan-Out, Ø400 µm Core Multimode Fiber x4
Fiber Adapter Plate	SMA with External SM1
Lens 1	N-BK7 Plano-Convex, Ø1", f = 25 mm, Uncoated
Lens 2	Aspheric Condenser, Ø1", f=16 mm, NA=0.79, Uncoated
Lens Tube	SM1, Adjustable (0.31" Travel Range)
Power Meter	PM100D with thermal sensor head (0.19 - 20 µm, 10 µW - 1 W)

3.6 Supplementary Information
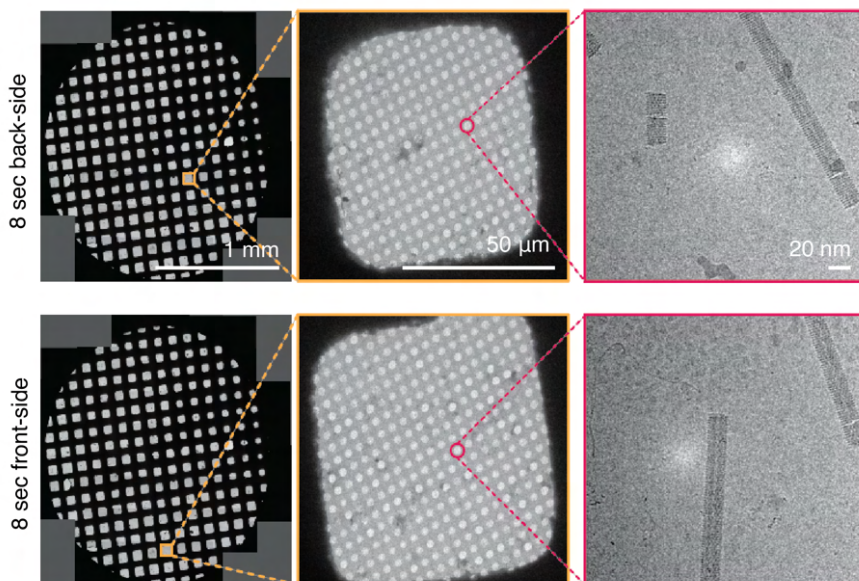
Figure 3.4: | **Cryo-EM of tobacco mosaic virus.** Testing of cryo-EM grid preparation with front-side or back-side blotting using the setup shows large areas of thin ice suitable for imaging.

Example 3.1: Motor homing codes

```
1  void sendLiftCommand(byte code, int delayTime) {
2    byte message[9] = {0x01, 0x00, 0x05, 0x02, 0x00, 0x01, code, 0x00, 0x04};
3    Serial1.write(message, 9);
4    delay(delayTime);
5  }
6
7  //Deactivate Lock
8  sendLiftCommand(0x00, 2000);
9
10 //Homing
11 sendLiftCommand(0x3F, 6000);
12
13 //Set to normal Operation
14 sendLiftCommand(0x00, 1000);
```

Example 3.2: Go to function

```
1  void ArmGoToWithParams(byte *result, int size, float pos, float maxvel, float maxacc,
    float maxdec) {
2      // Check if pos is within valid range
3      if (pos > 160 || pos < 0) return;
4
5      // Scale the float values to integers by multiplying them
6      unsigned long pos_acc = pos * 10000;
7      unsigned long maxvel_acc = maxvel * 1000000;
8      unsigned long maxacc_acc = maxacc * 100000;
```

```
 9        unsigned long maxdec_acc = maxdec * 100000;
10
11        // Use a single byte array for the data
12        byte tmp[size] = {
13          0x01, 0x00, 0x15, 0x02, 0x00, 0x02, ArmNibblecounter, 0x01,
14          (byte)(pos_acc & 0xFF), (byte)((pos_acc >> 8) & 0xFF),
15          (byte)(maxvel_acc & 0xFF), (byte)((maxvel_acc >> 8) & 0xFF),
16          (byte)(maxacc_acc & 0xFF), (byte)((maxacc_acc >> 8) & 0xFF),
17          (byte)(maxdec_acc & 0xFF), (byte)((maxdec_acc >> 8) & 0xFF),
18          0x04
19        };
20
21        // Toggle ArmNibblecounter
22        ArmNibblecounter = !ArmNibblecounter;
23
24        // Copy the data to the result array
25        for (int i = 0; i < size; i++) {
26          result[i] = tmp[i];
27        }
28  }
29
30  byte message25[25];
31  ArmGoToWithParams(message25, sizeof(message25), gotoheight, 0.2, 0.2, 0.2);
32  Serial1.write(message25, sizeof(message25));
```

3

3.7 References

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

2. Berriman, J. & Unwin, N. Analysis of transient structures by cryo-microscopy combined with rapid mixing of spray droplets. en. *Ultramicroscopy* **56,** 241–252. ISSN: 03043991 (Dec. 1994).

3. Dandey, V., Budell, W., Wei, H., Bobe, D., Maruthi, K., Kopylov, M., Eng, E., Kahn, P., Hinshaw, J., Kundu, N., Nimigean, C., Fan, C., Sukomon, N., Darst, S., Saecker, R., Chen, J., Malone, B., Potter, C. & Carragher, B. Time-resolved cryoEM using Spotiton. *Nature Methods.* Publisher: Springer US. ISSN: 1548-7105 (2020).

4. Frank, J. Time-resolved cryo-electron microscopy: Recent progress. en. *Journal of Structural Biology* **200,** 303–306. ISSN: 10478477 (Dec. 2017).

5. Mäeots, M.-E., Lee, B., Nans, A., Jeong, S.-G., Esfahani, M. M. N., Ding, S., Smith, D. J., Lee, C.-S., Lee, S. S., Peter, M. & Enchev, R. I. Modular microfluidics enables kinetic insight from time-resolved cryo-EM. en. *Nature Communications* **11,** 3465. ISSN: 2041-1723 (July 2020).

6. Kontziampasis, D., Klebl, D. P., Iadanza, M. G., Scarff, C. A., Kopf, F., Sobott, F., Monteiro, D. C. F., Trebbin, M., Muench, S. P. & White, H. D. A cryo-EM grid preparation device for time-resolved structural studies. *IUCrJ* **6,** 1024–1031 (2019).

7. Torino, S., Dhurandhar, M., Stroobants, A., Claessens, R. & Efremov, R. G. Time-resolved cryo-EM using a combination of droplet microfluidics with on-demand jetting. en. *Nature Methods* **20,** 1400–1408. ISSN: 1548-7091, 1548-7105 (Sept. 2023).

8. Subramaniam, S., Gerstein, M., Oesterhelt, D. & Henderson, R. Electron diffraction analysis of structural changes in the photocycle of bacteriorhodopsin. en. *The EMBO Journal* **12,** 1–8. ISSN: 02614189 (Jan. 1993).

9. Ménétret, J.-F., Hofmann, W., Schröder, R., Rapp, G. & Goody, R. Time-resolved cryo-electron microscopic study of the dissociation of actomyosin induced by photolysis of photolabile nucleotides. en. *Journal of Molecular Biology* **219,** 139–144. ISSN: 00222836 (May 1991).

10. Shaikh, T. R., Barnard, D., Meng, X. & Wagenknecht, T. Implementation of a flash-photolysis system for time-resolved cryo-electron microscopy. *Journal of Structural Biology* **165,** 184–189. ISSN: 10478477 (Mar. 2009).

11. Montaño Romero, A., Bonin, C. & Twomey, E. C. C-SPAM: an open-source time-resolved specimen vitrification device with light-activated molecules. *IUCrJ* **11** (Jan. 2024).

12. Gurney, A. M. *Flash photolysis of caged compounds* in (Citeseer, 1994), 389–406.

13. Josts, I., Niebling, S., Gao, Y., Levantino, M., Tidow, H. & Monteiro, D. Photocage-initiated time-resolved solution X-ray scattering investigation of protein dimerization. en. *IUCrJ* **5,** 667–672. ISSN: 2052-2525 (Nov. 2018).

14. Kim, M., Rho, Y., Jin, K. S., Ahn, B., Jung, S., Kim, H. & Ree, M. pH-dependent structures of ferritin and apoferritin in solution: disassembly and reassembly. *Biomacromolecules* **12,** 1629–1640 (2011).

15. Tretina, K., Park, E.-S., Maminska, A. & MacMicking, J. D. Interferon-induced guanylate-binding proteins: Guardians of host defense in health and disease. *Journal of Experimental Medicine* **216,** 482–500 (2019).

16. Kuhm, T. I., Gross, L., de Agrela Pinto, C., Huber, S. T., Taisne, C., Giannopoulou, E. A., Pardon, E., Steyaert, J., Tans, S. J. & Jakobi, A. J. Structural basis of membrane targeting and coatomer assembly by human GBP1. *bioRxiv,* 2023–03 (2023).

3

4

Cryo-EM structure of gas vesicles for buoyancy-controlled motility

Gas vesicles are gas-filled nanocompartments that allow a diverse group of bacteria and archaea to control their buoyancy. The molecular basis of their properties and assembly remains unclear. Here, we report the 3.2 Å cryo-EM structure of the gas vesicle shell made from the structural protein GvpA that self-assembles into hollow helical cylinders closed off by cone-shaped tips. Two helical half shells connect through a characteristic arrangement of GvpA monomers, suggesting a mechanism of gas vesicle biogenesis. The fold of GvpA features a corrugated wall structure typical for force-bearing thin-walled cylinders. Small pores enable gas molecules to diffuse across the shell, while the exceptionally hydrophobic interior surface effectively repels water. Comparative structural analysis confirms the evolutionary conservation of gas vesicle assemblies and demonstrates molecular features of shell reinforcement by GvpC. Our findings will further research into gas vesicle biology and facilitate molecular engineering of gas vesicles for ultrasound imaging.

4.1 Graphical Abstract
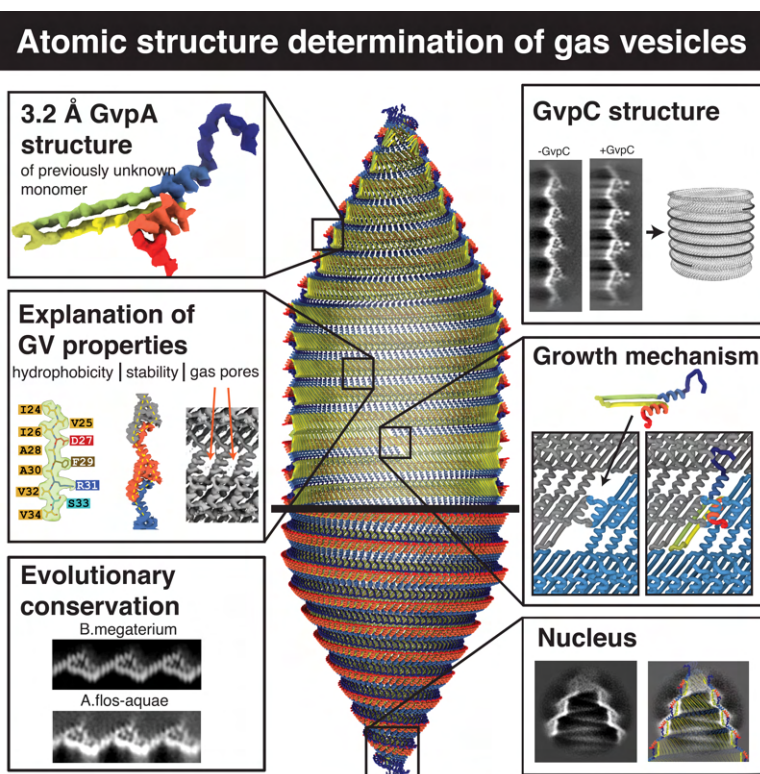
4.2 Introduction

Microorganisms utilize active motility systems to move toward or away from a variety of environmental stimuli such as chemicals and light[2]. These include swimming by rotation of rigid flagella and movement over solid surfaces with filamentous appendages[3]. Other forms of motility do not rely on active propulsion. Aquatic bacteria and archaea have evolved mechanisms to regulate buoyancy and can—similar to ballast tanks in submarines—create and eliminate gas-filled compartments to allow vertical migration in the water column. The cellular compartments providing positive buoyancy are formed by gas-filled protein shells called gas vesicles (GVs)[4].

There are very specific requirements for such structures: to achieve net buoyancy, GVs must occupy a substantial proportion of the cell, which involves forming compartments that extend over hundreds of nanometers in size. To maximize buoyancy the shell must be constructed from minimal material. At the same time, the shell needs to provide resistance to hydrostatic pressure to maintain buoyancy with changes in water depth[5].

GVs have therefore evolved as rigid, thin-walled structures composed of a single protein unit that typically polymerizes into large cylindrical shells closed off by conical tips[6,7]. The shell allows gas to diffuse passively between the GV lumen and the surrounding liquid, while effectively repelling water[8]. All GVs identified to date appear to be constructed from the same components[9]. The 7 kDa primary GV protein GvpA forms the core of the GV shell and the cone-shaped tips.

With molar masses exceeding hundreds of MDa, GVs range among the largest protein-based macromolecular assemblies reported to date. Despite intensive efforts[6,10–16], the molecular structure of GVs and therefore a molecular-level understanding of their distinctive properties have remained elusive. A comprehensive summary of the structural studies spanning several decades can be found in the introduction of this thesis (Chapter 1). Here, we solved the atomic structure of the canonical GV shell, providing detailed insight into the biogenesis of GVs and the unique evolutionary adaptions that enable buoyancy-controlled motility.

4.3 Results

4.3.1 Cryo-EM structure of the gas vesicle wall

We expressed and purified *Bacillus megaterium* GVs that form narrow tubes most suitable for cryo-EM analysis (Figure 4.1). The small diameter enables preparation of samples with thin ice to maximize contrast and reduces the propensity of GVs to deform from ideal cylindrical shape. The native *B. megaterium* GV gene cluster contains two almost identical GvpA homologs, named GvpA and GvpB; for consistency in naming convention, we will refer to them as GvpA1 and GvpA2. A minimal gene cluster including GvpA2 but lacking GvpA1 is sufficient for GV assembly in E. coli[17,18] (Figure 4.1A). Cryo-EM images showed GVs forming 0.1–1 μm long cylinders with varying diameters (55 ± 7 nm), consistent with previous data[19]. A subset of images (~16%) contained GVs with diameters as small as 34–42 nm (Figure 4.1B), which had their cylinder shape best preserved in the thin liquid film of cryo-EM samples (Figures 4.1B–1F).

We used a combination of two- and three-dimensional (2D and 3D) classification techniques to select 4% of the small GV subset corresponding to 35.6 nm diameter (Figure
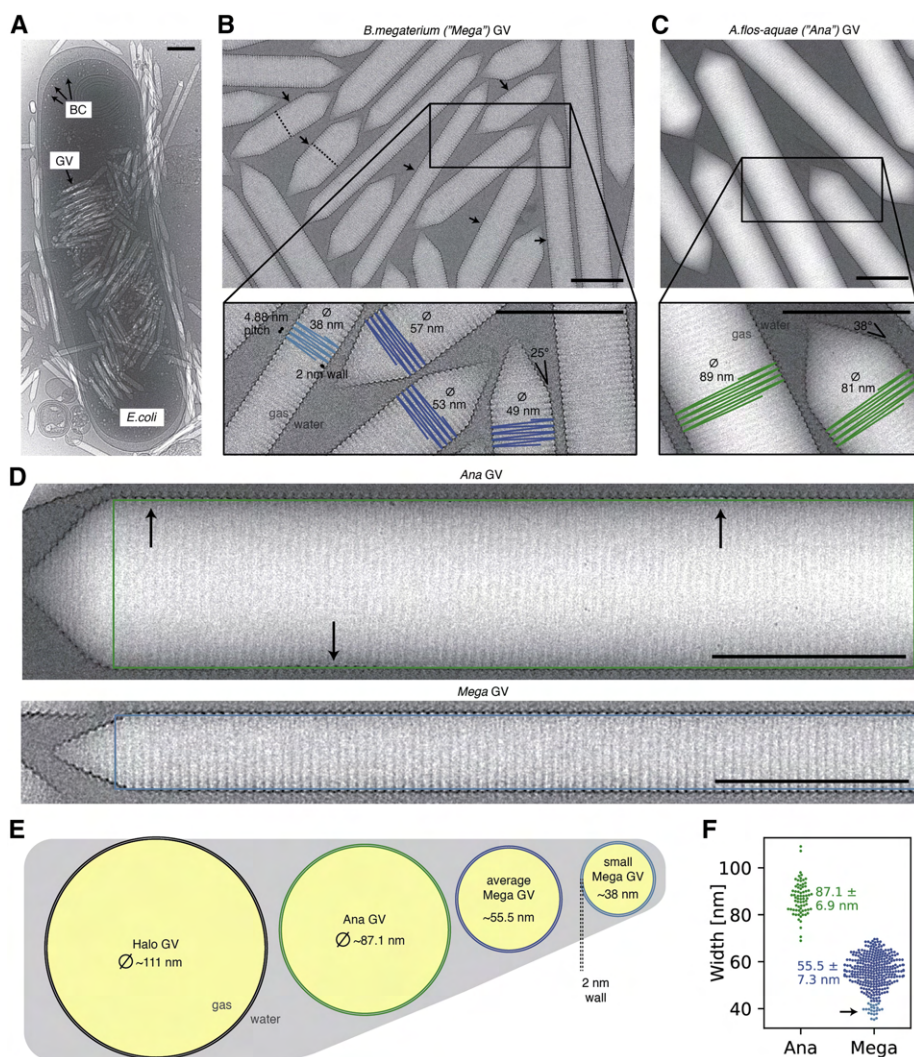
4

Figure 4.1: | **Gas vesicles.** (A) Cryo-EM micrograph of an E. coli cell heterologously producing *B. megaterium* gas vesicles (GVs). Mature GVs and small bicone (BC) nuclei are visible inside bacteria and in the surrounding medium. (B) Cryo-EM micrograph of purified *B. megaterium* GVs. GVs appear brighter than the surrounding solvent due to the lower density of the gas inside the GV lumen. Dashed lines on a subset of GVs represent midpoints of the helical GV segment; arrows indicate the location of the seam. The inset shows close-up examples of average-sized GVs (blue) and a small diameter GV (light blue) used for structure determination. The 4.88 Å helical pitch, the 2 nm-thick wall, and the 25° cone angle are annotated. (C) Representative cryo-EM micrograph of Ana GVs with inset showing a close-up indicating diameters and the ~38° cone angle. (D) A. flos-aquae (Ana) GVs with mean diameter of 87 nm deform in cryo-EM sample preparation and do not stay perfectly cylindrical. The green rectangle depicts the ideal, non-deformed shape of an ideal cylinder in projection. Arrows point at deviations of the GV from the ideal shape. Helical reconstruction in cryo-EM assumes perfect helical crystals of the repeating unit. This assumption is violated when GVs "squish" during sample preparation. Very thin *B. megaterium* (Mega) GVs maintain a cylindrical shape in the thin ice layer of the cryo-EM sample. (E) Schematic cross-sections of H. salinarum, A. flos-aquae, and *B. megaterium* GVs drawn to scale. Average diameter of H. salinarum GV from Dutka et al.[19] (F) Width measurements in 20 micrographs from both *A. flos-aquae* and *B. megaterium* GV datasets (this study). A long tail of small-diameter outliers with 34–42 nm diameter (black arrow) was observed in *B. megaterium* GVs. All scale bars 100 nm.

4.2) and used helical reconstruction to obtain a cryo-EM density of 3.2 Å resolution (Figure 4.3). The final reconstruction yielded a cylindrical GV shell assembly with ~93 GvpA2 monomers per helical turn, which represents one member of a range of helical polymorphs with diameters ranging from 34 to 70 nm and with 90–183 monomers per helical turn.

The reconstructed density allowed de novo atomic model building of GvpA2 in the structural context of its native assembly (Figure 4.3; Table 4.1). The cylindrical shell is constructed from thousands of GvpA2 monomers polymerized side by side into ribs spiraling into a left-handed helix with a helical pitch of 48.8 Å and -3.87° helical twist, resulting in 92.93 GvpA2 units per helical turn (Figures 4.3A and 4.3B). Contrary to postulated models[13,15], GvpA monomers and not antiparallel dimers form the repeating units of the helical assembly. GvpA2 adopts a coil-α-β-β-α-coil fold (Figures 4.3A and 4.3B). The carboxyl-terminal residues Asp67-Ile88 are flexible and not resolved in our structure. The helical lattice forms an array of ribs consisting of densely packed GvpA2 subunits with a lateral center-to-center distance of ~12 Å. The central β-hairpin, tilted at -36° relative to the long axis of the cylinder, forms the core of the GV ribs. Helix α2 folds back onto the hairpin, and helix α1 forms a bridge across the ~16 Å gap separating adjacent ribs. The GV wall is therefore only one or two peptide layers thick (Figure 4.3B). The inner wall of the GV shell forms a continuous hydrophobic surface consisting of a dense pattern of hydrophobic residues located on the luminal side of the β-hairpin and helix α1 (Figures 4.3C and 4.3D). Connections between the ribs of the GV shell are formed by the predominantly polar N terminus, which extends perpendicular to helix α1 and folds across the β-hairpin of the adjacent rib, stabilized by interactions with several residues in the β-hairpin and helix α2 (Figure 4.3E).

The extreme hydrophobicity of the luminal GV surface constitutes an energetic barrier for diffusion of liquid water or condensation of gaseous H2O. Consistently, GVs have been shown to be impermeable to water but to be highly permeable to gas molecules[8]. How gas molecules pass through the GV wall has so far been unclear. We located pores in the GV shell formed by slit-like openings between α1 helices of adjacent GpvA2 monomers (Figure 4.3F). We quantified the resulting pore size in the GV assembly computationally using Voronoi diagrams[21] and retrieved three different access routes with minimal constrictions ranging from 2.4 to 3.8 Å, compatible with the collisional cross-sections of gases dissolved in the cytosol (Figure 4.3F)[22].

Despite its limited thickness, the GV shell can resist several atmospheres of pressure without collapse[5]. GvpA2 monomers are held together tightly by lateral connections along the GV ribs formed by an extensive hydrogen-bonding network between the β strand backbones (Figure 4.3G). The hydrogen bonds are oriented at an angle of 54° relative to the cylinder axis, which is close to the "magic angle" 54.7° at which transverse and longitudinal stresses are equal in the wall of a cylinder[4,23]. Additional reinforcements are made by a continuous network of salt bridges formed by Glu43-Arg31 between two monomers and Arg31-Glu38 within a monomer. The GvpA2 shell consists of alternating line segments and triangular cross-sections providing force-bearing elements (Figure 4.3E). The corrugations increase stiffness along the rib direction, while the linear segments provide compliant hinge elements that increases elasticity of GVs, increasing their capacity to accommodate deformations orthogonal to the rib without collapse[24].
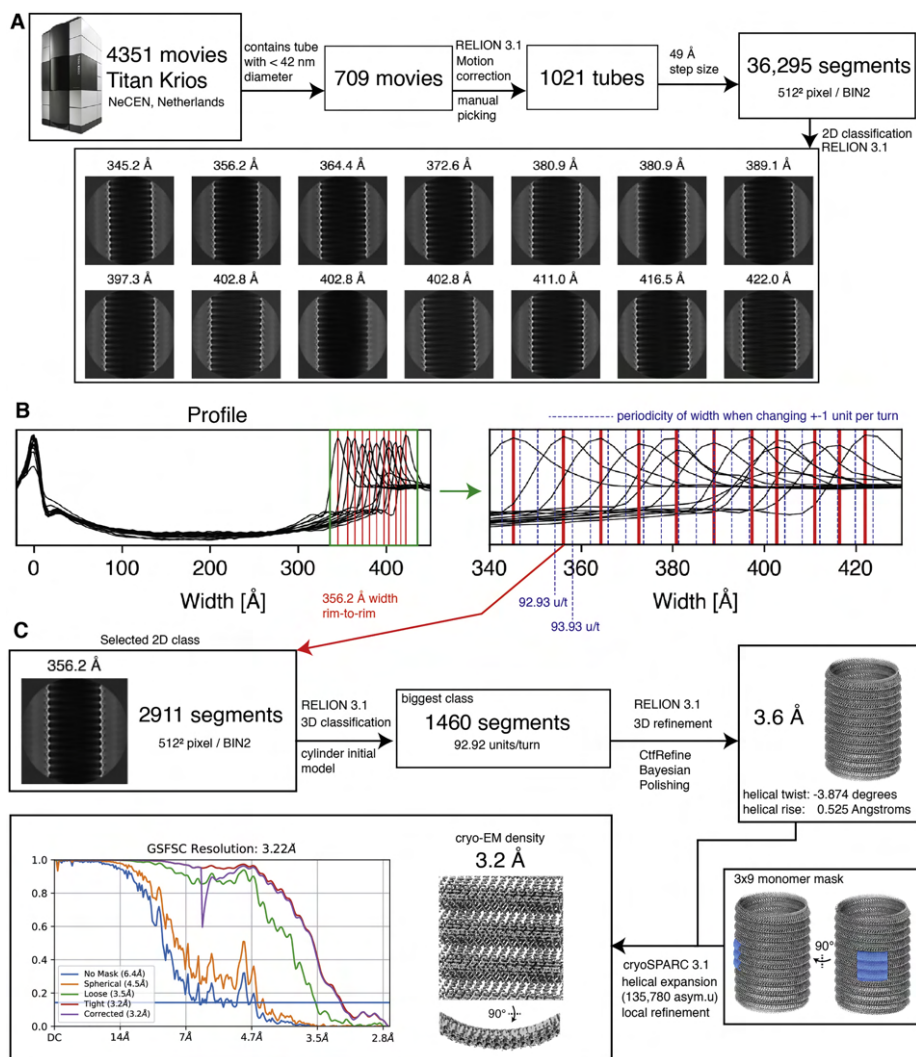
4

Figure 4.2: │ **Data processing of *B. megaterium* gas vesicle dataset.** (A) Preprocessing, manual picking, segment extraction and 2D classification leads to 2D class averages of GVs with different diameters. (B) The 2D classes were projected along the helical axis to generate profiles. The profiles were aligned with respect to the left peak. Zooming into the right peaks shows the distribution of GV widths in the 2D classes. Peaks are marked with a vertical red line. Blue lines indicate the periodicity of widths when an increment of one monomer per helical turn is assumed, based on a side-to-side distance of monomers of 12 Å, leading to a diameter increment of $12/\pi$ = 3.8 Å. Two to three different helical polymorphs are part of the particle subset belonging to a single 2D class average. (C) Processing steps starting from 2911 selected segments of a particular 2D class. The particle subset from the 2D class was further selected by 3D classification, imposing possible symmetry candidates between 90 and 95 units per wrung to select 1460 segments. Focused refinement on a 3x9 monomer segment of the wall in cryoSPARC 3.1[20] after symmetry expansion to 135,780 asymmetric units leads to the final result of a 3.2 Å resolution cryo-EM density of the GV wall.
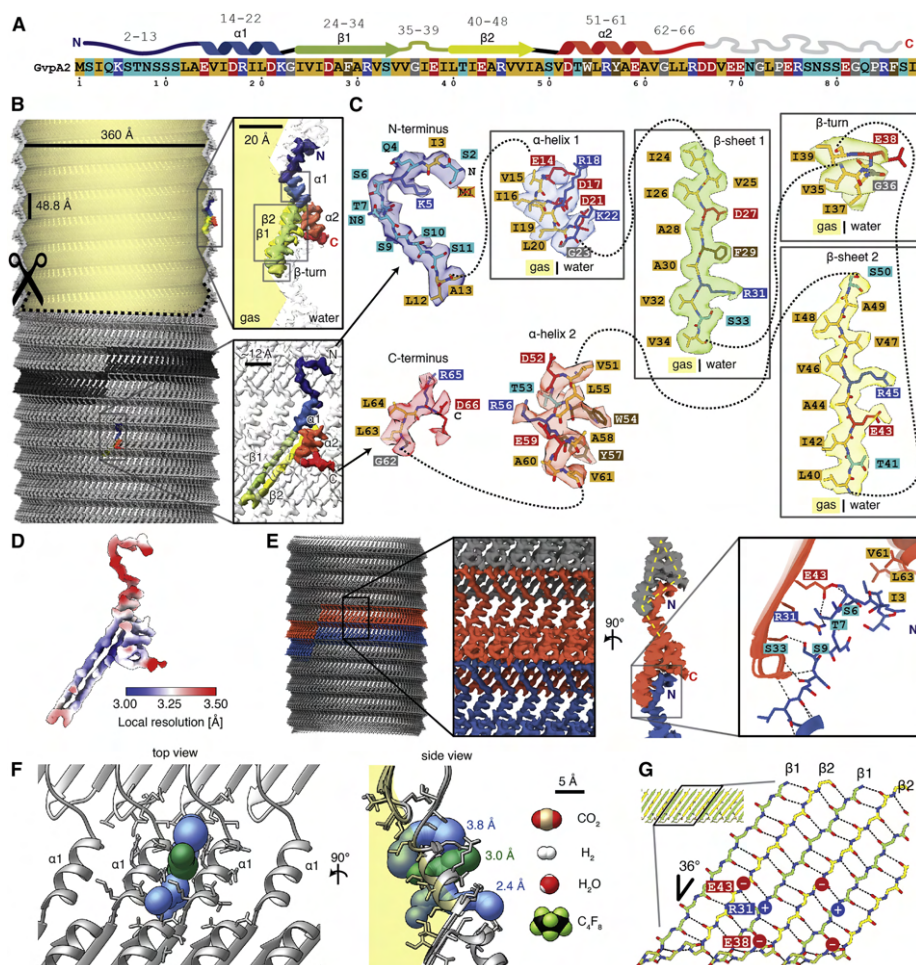
Figure 4.3: | **Cryo-EM structure of the gas vesicle wall.** (A) Primary and secondary structure of *B. megaterium* GvpA2. Residues in the primary structure are colored based on physicochemical properties. (B) GvpA2 monomers form thin-walled gas-filled protein shells assembled into a left-handed helix. One individual rib formed by 93 monomers is highlighted in dark gray. The top part of the GV is cut open to visualize the gas space. The 3.2 Å resolution cryo-EM density is shown with a single monomer colored according to sequence (N terminus, blue; C terminus, red) from a side view (top inset) and front view (bottom inset). (C) Entire cryo-EM density of a GvpA2 monomer annotated with respective amino acid one-letter-codes and the atomic model. Aliphatic residues (Ala, Val, Leu, Ile) line the gas-facing side of the GV wall. (D) Estimate of local resolution mapped onto the monomer density. (E) Cryo-EM density of the GV wall with two ribs highlighted in orange and blue. Side view illustrating corrugated zigzag structure and triangular cross-sections of the GV wall (yellow lines). Close up of inter-rib interactions mediated by the N terminus (blue), which binds across the β-hairpin and the C terminus (orange) of adjacent ribs, stabilized by hydrogen bonds from backbone, polar side chains (Ser6, Thr7, Ser9), and hydrophobic contacts (Ile3). (F) A slit between α1 helices allows diffusion of gas through the wall. Three computed tunnels approximate the slit and have bottleneck sizes ranging from 2.4 to 3.8 Å. The van der Waals surfaces of several gas molecules known to diffuse through the wall are shown to scale, with perfluorocyclobutane being the largest with 6.3 Å collision diameter.[8] (G) Schematic of the β strand rib providing most lateral connections for the assembly through backbone hydrogen bonding (dotted lines) and electrostatic interactions (Glu43-Arg31-Glu38).

Table 4.1: Cryo-EM data collection and model refinement statistics

Data collection	
Microscope	Titan Krios
Magnification	64,000
Voltage (kv)	300
Electron exposure ($e^-/Å^2$)	30
Exposure time (s)	2.4
Number of fractions	60
Number of movies	4351
Defocus range (μm)	0.25 - 1.25
Pixel size (Å)	1.37
Detector	K3
Dose rate detector (e^-/pix/s)	24
Data processing	
Helical rise (Å)	0.525
Helical twist (°)	-3.874
Final no. of asym. units	135,780
Global map resolution (Å) (FSC = 0.143)	3.2
Map sharpening B-factor ($Å^2$)	60.8
Model refinement	
Model starting point	*de-novo*
Sequence	GvpB - Uniprot O68677
Model composition	
Non-hydrogen atoms	497
Protein residues	65
Validation	
MolProbity score	0.69
Clashscore	0
Rotamer outliers (%)	0
Ramachandran plot	
Favored (residues)	61
Allowed (residues)	1
Disallowed (residues)	1 (V35)

4.3.2 Gas vesicles consist of two half shells in reverse orientation

The thin film of the cryo-EM sample orients the large GVs into a sideways orientation, providing a consistent viewing direction onto the GV edges in projection. Detailed inspection of GV edges revealed that GvpA2 monomers are always oriented with their β-turns pointing toward the center of the GV cylinder (Figure 4.4A), which contains a structural irregularity that has previously been referred to as a seam[12]. 2D class averages of GV edges around this seam showed two oppositely oriented GvpA2 monomers that make contact via their β-turns (Figures 4.4A and 4.5A) implying that this is the contact site of two GV half shells in reverse orientation.

While two contacting cylinders or cones have the same contact geometry along the entire circumference, GVs are assembled from two contacting helices. Continuity of the helicity implies there must be a unique polarity reversal point (PRP) at a point along the circumference, where an upwards- and a downwards-oriented GvpA2 monomer meet side by side. Apart from classes showing contacting β-turns, there is one less frequent set of 2D classes of the seam that displays two overlapping GvpA2 monomers in reverse orientation (Figures 4.4A and 4.5B). We posit that this 2D class is a projection view of the PRP located at the GV edge. As GVs can freely rotate around the cylinder axis in the thin ice layer, the PRP will be located exactly at the edge only at special rotation angles, explaining why this class is observed less frequently. The mirror symmetry in the 2D class implies a viewing direction perpendicular to a 180° (C2/D1) rotation axis (Figure 4.5C) that points through the PRP.

Using restraints from our 3D reconstruction and 2D class averages, we constructed a pseudo-atomic model of a GV half shell. Starting at the PRP, we extended the model from the D1 symmetry axis using the known helical symmetry for the cylindrical part of the GV shell and allowed transitioning into the conical tips by linearly decreasing the radius set by the 25° semi-angle of the cone and while refining structural adaption of GvpA monomers at defined hinge-points to match the experimental data (details in Methods and Figure 4.6). We then duplicated the half shell by rotating around the D1 axis, leading to a complete GV model consisting of 1,730 monomers and a total molecular mass of 12.2 MDa. Simulated density projections from this model closely match the experimental 2D classes (Figures 4.5D and 4.5E).

4.3.3 Molecular mechanism of gas vesicle biogenesis

Our pseudo-atomic model of the GV assembly shows that half shells interact through contacts at the GvpA β-turns around the circumference of the seam, as well as at the PRP where the GvpA2 rib reverses its polarity (Figure 4.4C). The pattern of side-by-side contacts between β sheets of GvpA2 monomers (d=d=d=d) along a rib is identical everywhere except at the PRP, where β-hairpins align inversely (d=d-p=p). The PRP is therefore a unique point in the GV assembly that may be recognized by the molecular machinery that facilitates GV growth and is likely the point at which new GvpA molecules are inserted during GV growth. Expansion of the GV assembly could occur either by inserting two oppositely oriented monomers, leading to symmetric expansion of both GV half shells, or by stochastic insertion of monomers into either half shell. If GV growth occurs by insertion of dimers, the PRP should always be located exactly at the midpoint of the cylindrical segment. Instead,
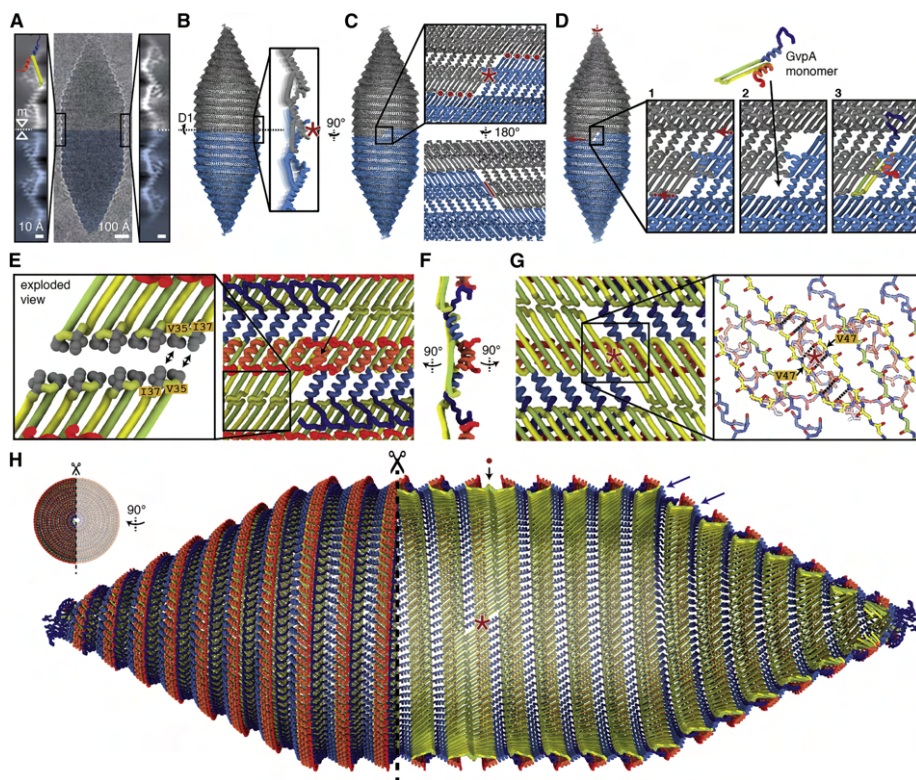
Figure 4.4: | **Pseudo-atomic model of a gas vesicle with two half shells in reverse orientation.** (A) Raw cryo-EM image of a single GV (with inverted contrast). β-hairpins of GvpA2 (cartoon) always point toward the seam at the center of the GV cylinder. Two different types of 2D class averages of the seam (left and right) are observed. The mirror symmetry (mirror axis: m) suggests a 180° symmetry axis (D1) at the point where two inversely oriented GV half shells meet. (B) Pseudo-atomic model of a GV constructed from two identical halves (gray, blue) with close up showing a side view of the polarity reversal point (PRP, red asterisk). (C) Close-up view onto the GvpA2 lattice around the PRP (red asterisk). Red dots indicate molecular contacts along the GV circumference where β-turns contact. The red line indicates contact between parts of hairpin strands β2 at the PRP. (D) Model of monomer insertion at the PRP. The two GV halves rotate against each other, with the β-hairpin contacts sliding over each other (red arrow) in a ratcheting fashion to allow monomer insertion in the resulting gap. Insertion of the monomer in the orientation opposite to that depicted is geometrically equivalent and would enlarge the other GV half shell. (E) View onto the seam between the two GV half shells with color scheme as in Figure 2B. The seam is sealed by contacting β-turns and the hydrophobic side chains Val35 and Ile37 (exploded view in inset). The lateral orientation of the hairpins is not uniquely determined by the data. An unresolved sterical clash between α2 helices from GvpA monomers around the PRP is marked by an arrow. (F) Side view of the PRP with color scheme as in Figure 2B. (G) View of the PRP from inside the GV. Oppositely oriented monomers next to the PRP are connected by six hydrogen bonds formed between the backbones of the β2 strands, with amino acid Val47 being located around the D1 symmetry axis (red asterisk). (H) Enlarged, sideways-rotated, and cut-open GV model with PRP (asterisk) and seam (dot) annotated. The proposed hinging motion of the N terminus required for adaption to the cylinder-to-cone transition and the reducing diameter of the tip is shown (blue arrows).
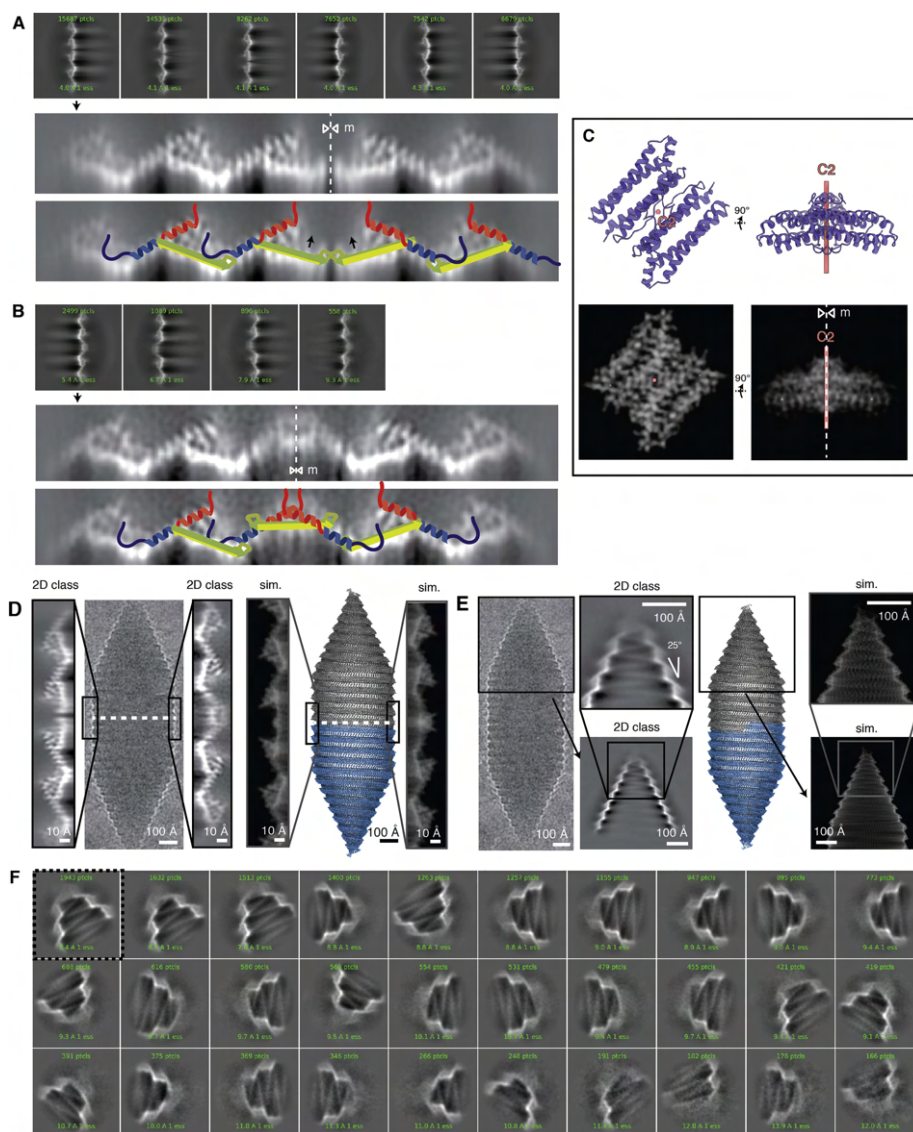
Figure 4.5: | **2D classification of *B. megaterium* seams and tips.** (A) 2D classes from the seam show perfect or near-perfect mirror symmetry. β-hairpins seem to hinge upwards at the seam (black arrows). (B) 2D classes from the putative PRP. Selected classes were magnified and sharpened for easier depiction. The mirror line is shown (m, dotted white line). A cartoon is drawn on the 2D classes to visualize GvpA molecules with the N-terminus in blue and the C-terminus in red. (C) Demonstration of the fact that views orthogonal to a 180° rotation axis show mirror symmetry in projection. Apoferritin dimer (PDB: 7ohf) is shown along the C2 axis and orthogonal to the C2 axis. 2D projection images (right) orthogonal to the C2 axis show mirror symmetry. (D) 2D class from the seam and PRP and simulated EM density from the pseudo-atomic model are in close agreement. (E) Pseudo-atomic model of a GV with simulated 2D projections of the tip, closely matching the experimental data. The 2D class average of GV tips with large box size reveals a linear decrease in radius at the tips with a cone angle of 25°. (F) 2D classes of GV tips with smaller box size reveal more detail, but all end in a blurry density at the tip. Alignment of secondary structure features is not possible and indicated strong structural heterogeneity at the tips.
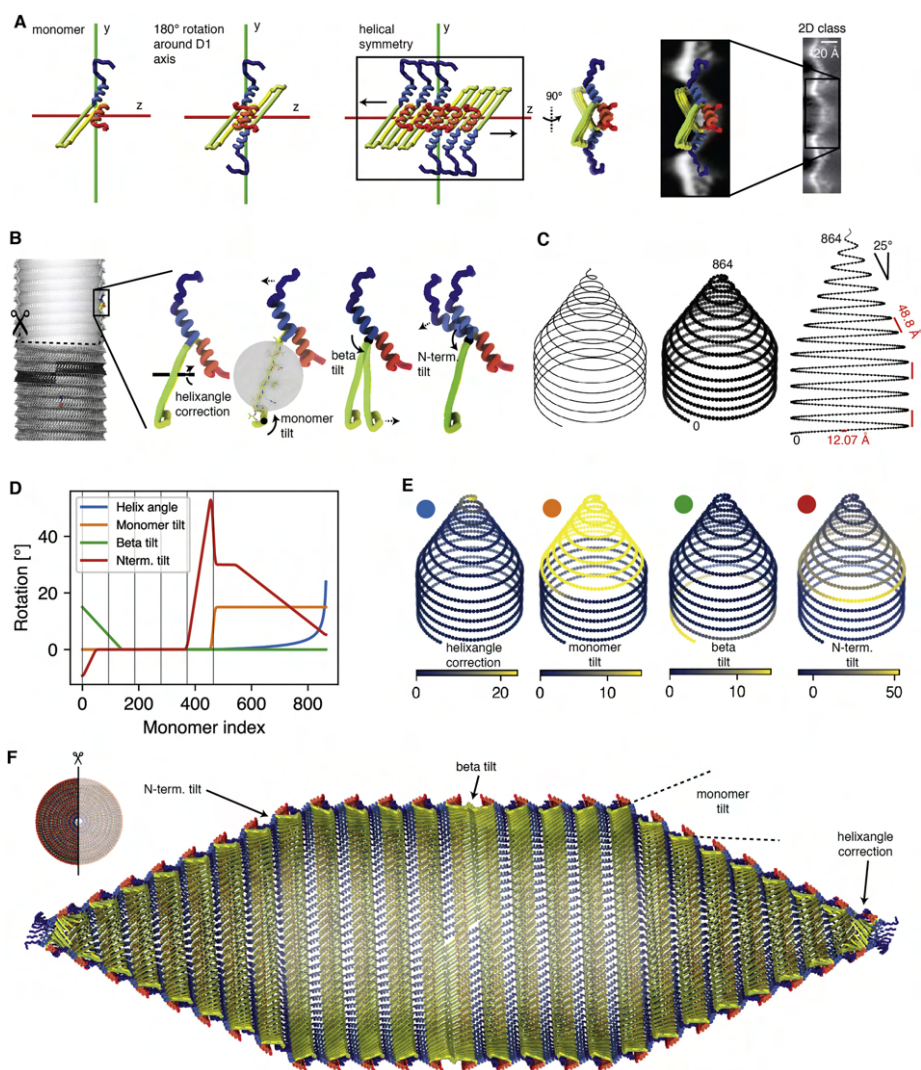
Figure 4.6: | **Construction of pseudo-atomic model of a whole *B. megaterium* gas vesicle.** (A) A GvpA2 monomer was placed next to the symmetry axis (along x) such that a 180° rotation would reproduce a view corresponding to the experimental 2D class average. The β-sheets meet in an angle at this stage, which is later corrected by tilting the sheets. (B) Four rotation parameters (helixangle correction, monomer tilt, β-hairpin tilt and N-term. tilt) used in the model are visualized. (C) The model is based on a helical curve in space with a linearly decreasing radius in the cones. The pitch in both the cylinder and cone is 48.8 Å. 835 monomers are placed equidistantly on the curve with an inter-monomer distance of 12.07 Å. (D) The helixangle parameter was extracted from the curve, while the other three parameters were manually tuned to fit the 2D class average data, make the β-sheets line up in the PRP and avoid gaps by adjusting the N-termini. Vertical lines every 93 units depict the monomers on each wrung of the cylindrical part. (E) Parameters mapped onto the monomers (F) Cross-section of the final model highlighting the impact of the four model parameters. The N-terminal tilt is required to maintain the binding site of the N-terminus to the adjacent rib despite the constricting cone radius. The beta tilt is required to make the orientation of the beta-hairpin compatible with the 2D classes at the seam. The monomer tilt is required to make the monomer orientation compatible with the 2D classes of GV tips. The helixangle correction is required to accommodate the increasing helical angle in the constricting cone toward the tips.

we observe that the PRP can be located away from the midpoint of the helical GV segment (Figure 4.1B), consistent with previous observations.26 We therefore propose that insertion of new monomers on either side of the PRP occurs stochastically through ratcheting of the two GV half shells by rotation relative to each other, generating a single monomer gap at the PRP (Figure 4.4D). This would involve breaking the lateral hydrogen bonds between the two monomers around the PRP, along with breaking and re-establishing hydrophobic contacts (Val35, Ile37) of the β-turns at the seam with no net loss in energy (Figure 4.4E). In our model, two factors suggest that the PRP represents the weakest point in the assembly. First, the monomer orientation at the PRP leads to steric hindrance by the α2 helices of the two symmetry-related monomers (Figure 4.4E). Second, the two oppositely oriented monomers at the PRP are connected by only 6 hydrogen bonds between segments of strand β2 around Val47 as compared to 11 hydrogen bonds formed between monomers along the rib (Figure 4.4G). We propose that the energetic disadvantage of this conformation facilitates opening of the seam for addition of new monomers during growth.

If the GV grows by adding monomers at the PRP, backward extrapolation starting from a mature GV leads to a state at which the seam forms between two conical half shells (Figure 4.7A). This is the point where conical growth transitions into cylindrical growth. This transition requires adaption of the relative orientation of adjacent ribs, which we propose is mediated through a hinging motion of the N terminus relative to helix α1 to accommodate for the reduction in diameter at the transition point and along the cone toward the tip (Figure 4.4H).

Continuing this extrapolation leads to a biconical nucleus that must initially form to start GV biogenesis. According to our model, the original nucleus would remain at the cone tips on both half shells after maturation. The 2D classes and the fitted pseudo-atomic model suggest that N-terminal residues of GvpA2 monomers crowd together at the tip (Figures 4.7B and 4.7C). At diameters lower than 50 Å, the 2D class averages of these cone tips display weak density (Figure 4.5F) sealing off the opening. The fuzzy appearance of the density at the tip is indicative of structural heterogeneity in the nucleating monomers.

4.3.4 CONSERVATION OF GAS VESICLE SHELL ARCHITECTURE

Sequence alignments between GvpAs of three evolutionary diverse bacterial and archaeal species producing GVs reveal a high degree of sequence conservation in the structural parts of GvpA, suggesting that the overall mode of GvpA shell assembly must be similar. This is supported by computational predictions using AlphaFold2, revealing highly similar assemblies (ribs) of GvpA oligomers for the evolutionarily distant GvpAs in firmicutes (*B. megaterium*), cyanobacteria (*A. flos-aquae*) and the archaeon *H. salinarum*, that all resemble our experimental structure (Figures 4.8A and 4.8B).

To verify these predictions, we employed cryo-EM to image GVs of *A. flos-aquae*. Differences in the GvpA sequences between *A. flos-aquae* and *B. megaterium* GvpA2 locate mainly to the N-terminal and C-terminal regions of the folded core (Figure 4.8C). *A. flos-aquae* GVs formed cone-ended cylinders with mean diameter of 87 ± 7 nm, consistent with previous observations[19]. Despite significant effort, 3D refinement using a similar helical reconstruction strategy as for *B. megaterium* did not converge. Closer inspection of full-length GVs revealed them to flatten in the thin ice of the cryo-EM sample, breaking symmetry assumptions of helical reconstruction. We therefore resorted to obtain high-
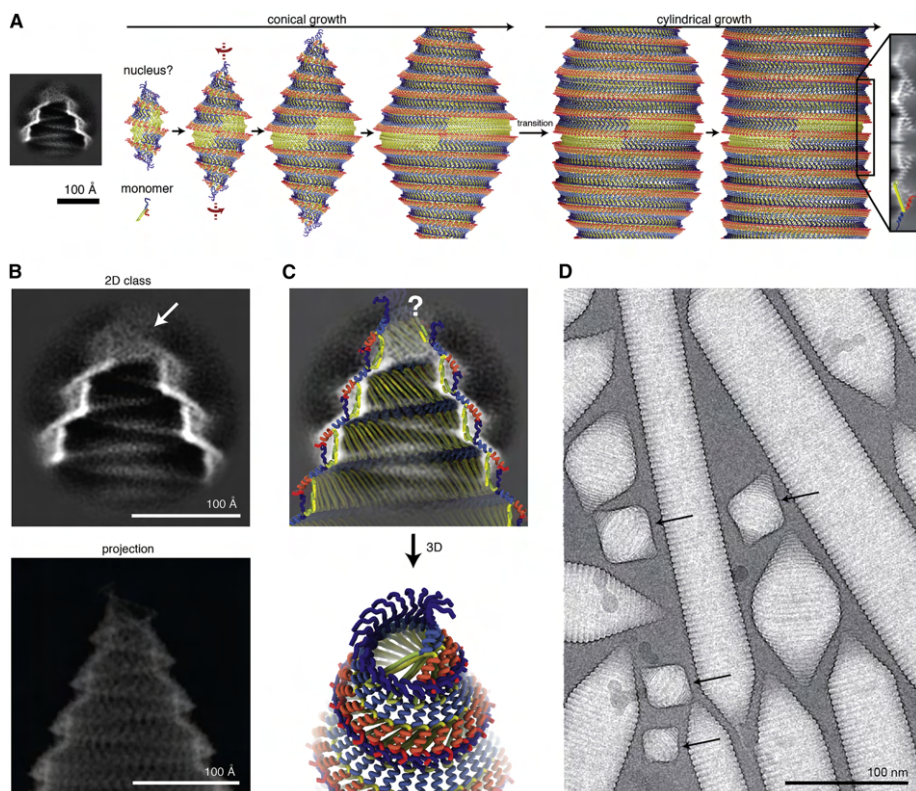
Figure 4.7: | **Model of gas vesicle growth from an initial nucleus.** (A) GV growth from a hypothetical nucleus to mature GVs. The transition from conical to cylindrical growth is annotated. A 2D class of GV tips with proposed nucleus remnants is shown to scale. (B) Class average of the GV tip shows no molecular order at diameters lower than 50 Å toward the cone end (top). 2D projection of the pseudo-atomic model is shown to scale (bottom). (C) 2D class overlaid with cut-through of the pseudo-atomic model (top) and 3D view onto the tip (bottom). The simplified model of a helix with linearly decreasing radius breaks down at the very tip, leading to steric clashes between the main chains. N termini from GvpA monomers at the tip come into proximity and might close off the gas space. (D) Cryo-EM micrograph with immature GV bicones (arrows).

Figure 4.8: | **Highly conserved GvpA from *B. megaterium* and *A. flos-aquae* adopt the same fold and assembly.** (A) Sequence alignment of *B. megaterium* (Mega) GvpA1 (GvpA) and GvpA2 (GvpB), *A. flos-aquae* (Ana) GvpA, and *H.salinarum* (Halo) GvpA1 and GvpA2 show high degree of conservation despite forming GVs of different diameters. (B) AF2 prediction of GvpA 5-mers for selected genes in (A) compared to the cryo-EM struct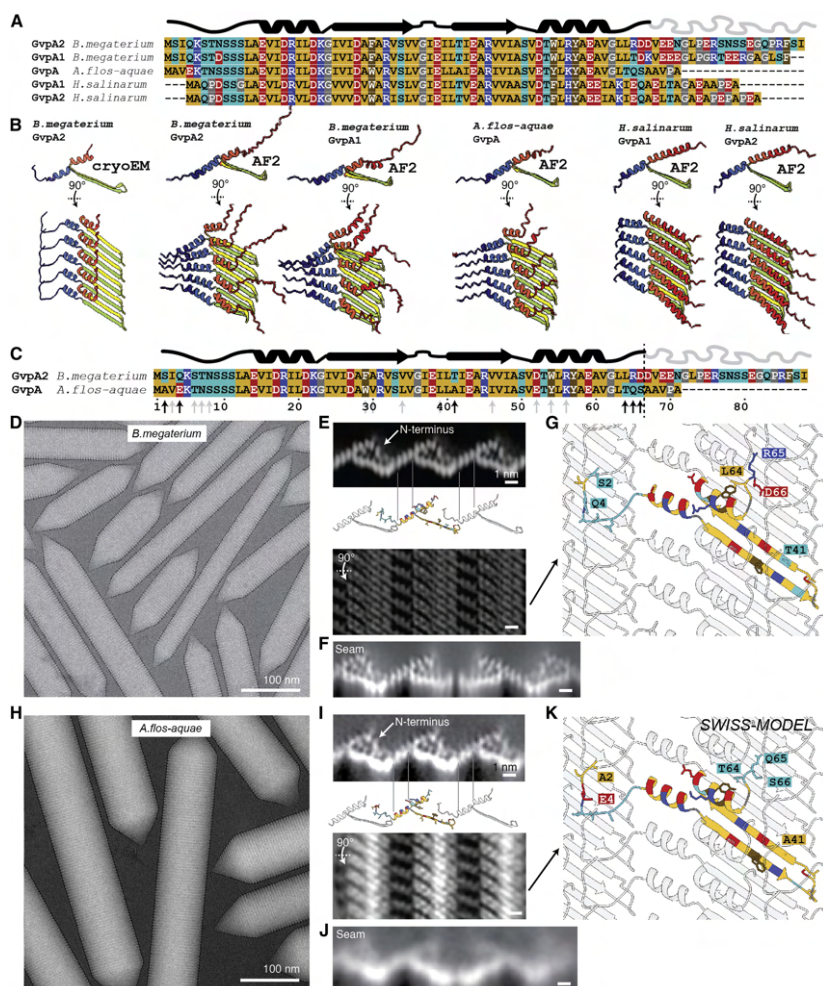ure. Only the middle monomer is shown in side view. AF2 predicts the general arrangement of a GvpA rib and the angle between the β-hairpin and α helix 1 accurately. The conformation of the N-terminal coil appears different, and the distance between α helix 1 (where gas pores are located) is not accurately modeled by AF2. (C) Protein primary sequences of wall-forming protein GvpA from both *B. megaterium* and A. flos-aquae are very similar. Black arrows show 6 property-changing mutations; gray arrows show 12 property-conserving mutations in the ordered part of the structure. (D) Representative cryo-EM micrograph of *B. megaterium* GVs. (E) 2D-projected side view and top view of the 3.2 Å cryo-EM density of *B. megaterium* GVs. (F) 2D class average of seam between two GV half shells. (G) Atomic model of *B. megaterium* GV wall with different side chains between both species displayed and property-changing mutations highlighted by one-letter-code. Residues are colored according to side chain chemistry as indicated in (A). (H–K) Equivalent data for *A. flos-aquae* GVs shows the high degree of conservation of GvpA fold and assembly. The 2D views in (I) are computed by 2D classification of GV edges and collapsed GVs and have a resolution better than 5.4 Å or 4.8 Å as the α-helical pitch or the β strands are resolved. The homology model of the *A. flos-aquae* GV wall was computed using SWISS-MODEL[25].

resolution structural information from 2D classification. Class averages of GV edges showed a corrugated zigzag pattern of the ~2 nm-thick wall formed by GvpA (Figure 4.9A). Another set of 2D class averages, obtained from collapsed GVs also present in the data, revealed the *A. flos-aquae* GV wall to consist of a periodic array of ribs consisting of dense 5.0 × 1.25 nm GvpA subunits tilted at -36° relative to the long axis of the cylinder (Figure 4.9B). The class averages with a resolution of better than 4.8 Å allow discerning the α-helical repeat of helix α1 bridging the gap of adjacent ribs and the individual β strands of the polymerizing β-hairpin. A cumulative Fourier spectrum computed from in-plane rotated GV segments showed a typical Fourier transform of a helix (Figure 4.9C), consistent with our 3D reconstruction from *B. megaterium* GVs (Figure 4.9D). We compared the *A. flos-aquae* 2D classes of GV wall side and top views with equivalent projections of the *B. megaterium* GV structure. On the level of the main-chain fold, the two GV shell assemblies are indistinguishable, hence confirming the conserved architecture of GvpA assembly of the GV shell (Figure 4.8). Our model of two contacting half shells also applies to *A. flos-aquae* GVs as we observe similar contacting β-turns in 2D class averages of the seam (Figures 4.8F and 4.8J).

More generally, our results establish that 2D classification of GV edges in cryo-EM data can give valuable structural insight into GV architecture for cases when 3D reconstruction is out of reach. This is due to the small unit cell of the GV wall, where not many structural features overlap in side views at the GV edges. 2D class averages can be computed to sufficient resolution to see secondary structure elements and even large side chains. We suggest that this approach can be used for comparative studies of GVs from different species with larger sequence divergence to reveal different assembly modes, such as resulting from differences in binding modes of the evolutionarily less conserved GvpA N terminus (Figure 4.10A).
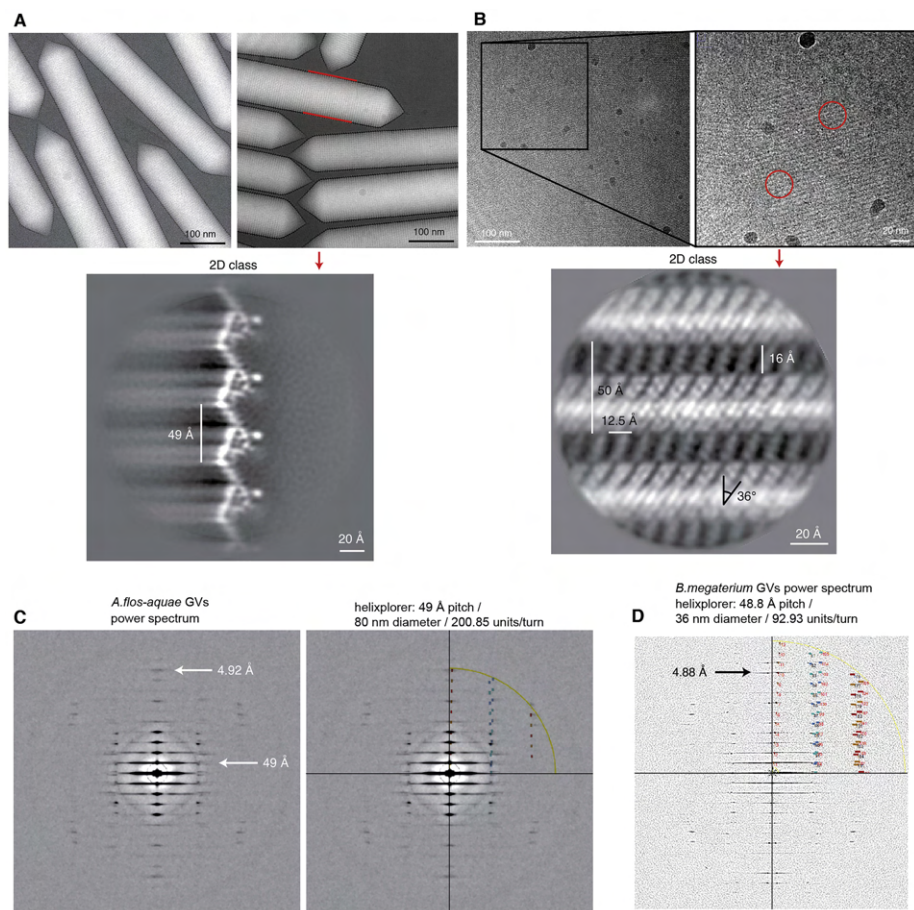
Figure 4.9: | **Cryo-EM of *A. flos-aquae* gas vesicles.** (A) Representative micrographs of *A. flos-aquae* GVs. GV edges were analyzed by 2D class averaging to give a low-noise high-resolution 2D view of the edges, to reveal a repetitive zigzag pattern. The 2D view shows details of at least 5.4 Å resolution corresponding to the α-helical pitch. (B) The same dataset contained collapsed GV wall segments. Those can be averaged as well by 2D class averaging to reveal a high-resolution top-view of the GV wall with better than 4.8 Å resolution as the β-strands are resolved. (C) Computing the sum of in-plane rotated power spectra of segments of all GVs in the dataset gives rise to a layer-line pattern typical for helical assemblies. This approach can be seen as a form of fiber diffraction where fibers are aligned computationally. Overlay with computed layer line patters (Helixplorer, http://rico.ibs.fr/helixplorer/) shows good agreement to a helix with 49 Å pitch and 200.85 units per helical turn. (D) Power spectrum of *B. megaterium* power spectrum with first layer line at 48.8 Å. The upper right quadrant is overlaid by a screenshot from Helixplorer (http://rico.ibs.fr/helixplorer/) used for interactive exploration of helical symmetry.

4

Figure 4.10: | **Evolutionary conservation of gas vesicle proteins.** (A) Multiple sequence alignment of *B. megaterium* GvpA2 with 50 sampled Uniref. 90 clusters of ~40–100% sequence identity. (B) MSA of sequences similar to *A. flos-aquae* GvpC 33 AA repeats. Alignment of 33 amino acid repeat consensus sequence of Ana GvpC with 91 similar sequences from Uniref. 90 clusters with ~40–100% sequence identity reveals a highly conserved pattern including leucine, phenylalanine and Arg19 residues. The consensus sequences on top were computed in MView and show residues conserved on at least 70%, 80%, 90% and 100% of sequences.

4.4 Discussion

GVs represent a remarkable example of biomolecular self-assembly. Our results provide a canonical structural framework for the unique molecular properties of GVs, including their selective permeability to gases[8,26], their mechanical stability[24], and their distinctive ability to grow without compromising the integrity of its shell[4]. Our work establishes an atomic resolution model of the mature GV shell formed exclusively by GvpA. A key question is how GvpA nucleates to form an elementary bicone from which the shell extends during GV growth. Besides GvpA, many GV gene clusters contain genes encoding the proteins GvpJ, GvpM, or GvpS[27] that exhibit high sequence homology and predicted folds similar to that of GvpA. A dominant structural role of these homologs in mature GVs is unlikely, as none of them were found in intact GVs[28], suggesting a putative role as nucleation factors. Additional support for this role comes from the observation that, in some species, GvpJ and GvpM are expressed as part of a separate transcript exclusively during early exponential growth[9], whereas a transcript of gvpACNO is expressed at later stages, possibly to enlarge the formed nuclei. How the initial nucleus is structured and by which mechanism other Gvps assist in growth remain open questions[29].

In most organisms, biconical GVs transition their growth mode when reaching a certain diameter and continue extending cylindrically. This transition occurs over a range of diameters, different for each individual GV. We observed mature cylindrical *B. megaterium* GVs with a diameter of 55.5 ± 7.3 nm corresponding to 145 ± 19 monomers per helical turn and mature *A. flos-aquae* GVs with diameter of 87.1 ± 6.9 nm corresponding to 227 ± 18 monomers per helical turn. The mechanism for the bicone-to-cylinder transition is unknown. The GvpA ribs in the cones are highly curved and may be energetically disfavored. Insertion of new monomers in a bicone results in enlargement of the cone and reduces rib curvature. However, the expanding cone does not provide defined interactions between monomers of adjacent ribs. In contrast, cylindrical GV segments have crystalline order where the N terminus always has identical contacts to the adjacent rib. The interplay of curvature preference and the energetic advantage of crystallinity could favor a cone-to-cyclinder transition at a certain critical diameter. This suggests that the GvpA N terminus could play a role in determining the size distribution of GVs. Consistently, the sequence of the N terminus is most divergent. Further evidence for the decisive role of GvpA sequence in defining the final diameter comes from a hybrid GV gene cluster where *A. flos-aquae* GvpA integrated into the native gene cluster of *B. megaterium*, resulting in GVs with diameters consistent with native *A. flos-aquae* GVs[30].

Our structure reveals that the gas permeability of the GV wall can be ascribed to a large number of molecular pores formed between α1 helices of the GvpA shell, the size of which is compatible with the collision diameters (2.65–3.64 Å) of typical atmospheric gases dissolved in the cytosol[22]. Surprisingly, the GV wall has been shown to be permeable to perfluorocyclobutane (C4F8) with a collision diameter (6.3 Å) exceeding the estimated pore size in our structure. While atmospheric gases appear to diffuse freely through the GV wall[26], the diffusion coefficient of C4F8 is consistent with a very small number (~11) of such pores[8]. Based on our pseudo-atomic model, such pores would need to be located at the seam, PRP or at the conical tips. Alternatively, flexibility in the GV shell may modulate size of the regular gas pores to allow passage of C4F8.

Recently, GVs have been repurposed as genetically encoded acoustic reporters[31,32].

The high contrast in density between gas-filled GVs and surrounding cellular structures makes them amenable to ultrasound imaging[33]. While native GVs display little shell deformation under ultrasound exposure, GVs that are stripped of GvpC become less stiff and scatter non-linearly above a certain pressure threshold[34]. This behavior enables amplitude modulation imaging and multiplexing of stripped and unstripped GVs in an in vivo context[34,35]. Engineering conditional binding strength of GvpC can transform GVs into biosensors with switchable acoustic properties[36,37]. The high-resolution structure of the GvpA shell may enable development of designer GVs with custom mechanical shell properties through direct engineering of the GvpA sequence.

Together, our results establish the molecular basis of a widely conserved buoyancy-controlled motility apparatus in aquatic bacteria and archaea. Our study will form the foundation for addressing a multitude of open questions on GV biogenesis such as nucleation, growth, width regulation, function of other GV gene products in GV assembly, and species-to-species variability of GV gene clusters.

Limitations of the study

While our structure provides an atomic model of the main structural protein GvpA as it forms the GV wall assembly, the function and/or structural role of other accessory proteins in GV gene clusters in GV biogenesis is yet to be determined. Some of those proteins presumably work together to create a nucleus whose precise structure remains unknown.

We also report a pseudo-atomic model of an entire GV. The precise structure of the GvpA monomer in this model is supported by cryo-EM data at 3.2 Å resolution; however, the special structural features of the assembly such as the seam, polarity reversal point, and tip are inferred from 2D projection views alone. The exact molecular details at these special positions can slightly deviate from the pseudo-atomic model we present.

4.5 Methods and Materials

B. megaterium gas vesicle expression and purification

The purification protocol for Mega GVs was derived from Lakshmanan et al.[18]. In brief, BL21(DE3)pLysS *E.coli* cells were transformed with the pST39-pNL29 plasmid (a gift from Mikhail Shapiro; Addgene number 91696)[18] and 1 L of LB containing 0.2% (w/v) glucose was inoculated with 10 mL of overnight culture. The culture was grown at 37°C until OD = 0.5 and GV expression was induced with 20 μM IPTG. Following induction, cells were grown at 30°C for 20 h.

The culture was centrifuged at 500 rcf for 2 h in 50 mL Falcon tubes. The floating fraction was collected with a peristaltic pump. This process was repeated once more. The resulting 25 mL of liquid were lysed chemically by adding 5 mL of SoluLyse reagent per 50 mL of liquid, 0.25 mg/mL lysozyme and 10 μg/mL DNaseI, and slowly rotated for 1 h at room temperature.

GVs were purified in three overnight rounds of floatation separation by centrifugation at 800 rcf in 50 mL Falcon tubes. After each centrifugation step, the GV-containing top layer was gently removed with a pipette after which the GVs were resuspended in PBS containing 6M urea (first round), and subsequently in PBS alone. Final concentration was determined as OD500 = 3.1 by optical density measurement at 500 nm against a sonicated

blank. The sample was dialyzed into imaging buffer (20 mM Tris, 50 mM NaCl, pH = 8) prior to cryo-EM sample preparation.

A. FLOS-AQUAE GAS VESICLE PURIFICATION

The purification protocol for GVs from *A. flos-aquae* was derived from Lakshmanan et al.[18]. Briefly, *A. flos-aquae* (CCAP 1403/13F), also known as *Dolichospermum flos-aquae*, were grown in 250 mL G625 medium complemented by BG11 medium (Sigma C3061) for approximately 2 weeks until confluence. The culture was centrifuged at 350 rcf for 4 h or until a floating layer of cells was observed. Subnatant was removed using a syringe before lysing the cells in 500 mM sorbitol and 10% v/v lysis buffer (SoluLyse) at 4°C for 6–8 h while gently rotating. GVs were purified by three rounds of flotation separation with 4–6 h centrifugation at 350 rcf. After each centrifugation, subnatant was removed by syringe after which GVs were resuspended in PBS at pH 7.4.

CRYO-EM OF B. MEGATERIUM GAS VESICLES

B. megaterium GVs at OD500 = 3.1 were applied to a freshly glow-discharged Quantifoil R2/1 grid and frozen using a Leica plunger set to 95% humidity, front-side blotting and 20°C with blot times ranging from 5 to 11 s. Micrographs were collected on a Titan Krios (Thermo Fisher Scientific) microscope at the Netherlands Center for Electron Nanoscopy (NeCEN) operated at 300 kV. Dose-fractionated movies were acquired on a Gatan K3 direct electron detector at a pixel size of 1.37 Å with 60 frames over an exposure of 30 e-/Å2 and a defocus range from -0.25 to -1.25 μm.

CRYO-EM OF A. FLOS-AQUAE GAS VESICLES

Native *A. flos-aquae* GVs (containing GvpC) at OD500 = 13 in imaging buffer (20 mM Tris, 50 mM NaCl, 0.5 mM EDTA, pH = 8) were applied to a freshly glow-discharged Quantifoil R2/1 grid and frozen using a Leica plunger set to backside-blotting, 95% humidity and 20°C with 10 s blot time. 1273 cryo-EM micrographs at 1.288 Å pixel size were acquired on a JEOL 3200 microscope with a K2 detector using 62 e- total exposure over 60 frames. *A. flos-aquae* GVs stripped from GvpC (OD500 = 1, in PBS buffer) were prepared and imaged in a similar same way, using 17.6 e- over 50 frames, 1.288 Å pixel size, acquiring 58 micrographs.

DATA PROCESSING AND STRUCTURE DETERMINATION OF B. MEGATERIUM GAS VESICLES

4351 collected super-resolution movies were 2x binned and motion-corrected in RELION 3.1[38]. CTF determination was performed using Gctf 1.06[39]. 709 micrographs containing thin GVs with a diameter of 42 nm or less were identified manually. 1021 tubes were manually picked in RELION by selecting start and end coordinates. 36,295 overlapping segments with 512 pixels were extracted along the cylindrical sections with a step size of 49 Å (2x binned). 2D classification was done in RELION3.1 with the 'ignore CTFs until first peak' option turned on. The resulting 2D class averages were grouped by projecting them along the helical axis and calculating the rim-to-rim distance (Figure S2B) between the two density maxima. A class with 35.6 nm edge-to-edge distance was selected containing 2911 segments.

Analysis of in-plane rotated power spectra of the segments using Helixplorer (http://rico.ibs.fr/helixplorer/)
revealed a likely helical symmetry between 90.92 and 95.92 units per helical rung, with ~49 Å helical pitch. 2D classification was not sufficient to separate all symmetries and the final set of segments originated from several assemblies with different symmetries.

3D classification starting from a featureless cylinder and imposing candidate symmetry parameters was used to further select for segments adhering to a particular symmetry, leading to 1460 segments with symmetry of 92.93 units per turn. 3D refinement of those particles with CTF refinement and Bayesian polishing led to a resolution of 3.6 Å at FSC = 0.143. Convergence of 3D refinements was only achieved when using a $'tau_f udge'$ parameter of 5.

A final round of particle polishing was used to create a new particle stack extracted only from frames 1–20 (0–10 e-/Å2) of the movies. Particles were exported to cryoSPARC 3.1.0[20] and high-pass filtered to 100 Å. 3D refinement with the helical reconstruction algorithm implemented in cryoSPARC also led to a reconstruction at 3.6 Å. To account for small deviations from helical symmetry, e.g. by flattening of the tube in ice, a mask encompassing ~3x9 GvpA2 monomers was created in ChimeraX[40]. The particle stack expanded by helical symmetry was subjected to focused refinement in cryoSPARC using the mask, which increased the final resolution to 3.2 Å at FSC = 0.143. The final maps were cropped from a box size of 512^3 voxels to a box size of 128^3 voxels centered on the refined region. Local resolution was calculated in cryoSPARC 3.1.0[20] and mapped onto the monomer structure in ChimeraX[40].

Atomic model building and refinement

To build an atomic model of a GvpA2 monomer the final map density was traced de novo using COOT[41]. The monomer was expanded using helical symmetry (rise: 0.525 Å, twist: -3.87°) into a 15 subunit segment (three ribs with 5 monomers each) to account for connections between monomers and manually adjusted in ISOLDE[42] before automatic real-space refinement in PHENIX 1.13[43]. We used 'phenix_real_space_refine' with automatic restraint weighting to refine coordinates and atomic displacement factors against the globally sharpened experimental density map using secondary structure and non-crystallographic symmetry restraints. The central monomer of the 15 subunit segment was used as the asymmetric unit for PDB deposition. Renderings of the cryo-EM density and atomic models were made in ChimeraX 1.4[40].

2D classification of edges, seams and tips

From the *B. megaterium* GVs cryo-EM dataset, several hundred particles of either seams or tips were picked manually and used to generate a template for automated picking in cryoSPARC 3.3[20]. Particles were high-pass filtered to 100 Å to eliminate the large negative contrast of the gas space in the GV interior. The picked particles were cleaned up by several rounds of 2D classification to give a clean set of particles of either the seam, the PRP or the tips. These particle sets were used to train a neural network for particle picking (TOPAZ v0.23[44]), which was then applied to the micrographs to pick seams, PRPs, and tips. Those particles were cleaned by several rounds of 2D classification and led to the final presented 2D classes. For display, the 2D classes were treated in ImageJ using an 'unsharp mask'

filter.

The two cryo-EM datasets of *A. flos-aquae* GVs with and without GvpC were used to obtain 2D class images of the edges leading to side-views of the wall. Movies were imported into cryoSPARC v3.3][20], motion-corrected and CTF-estimated. For both dataset, frames were used only until an exposure of ~15 e-/Å2 because shrinking of GVs was observed for high exposures leading to GV edges blurring out. A few hundred edges were manually selected for 2D classification to generate picking references for the cryoSPARC filament tracer. Particles were extracted with 192 pixel box size and high-pass filtered to 100 Å. Several rounds of 2D classification and selection of sharp classes led to the final 2D classes of the GV edges with and without GvpC. For display, the 2D classes were treated in ImageJ using an 'unsharp mask' filter.

Similarly, 2D classes of the GvpA lattice from collapsed GVs were calculated from the *A. flos-aquae* GV dataset above including GvpC. For this, the dataset was preprocessed in RELION 3.1[38], points on the lattice manually picked to create a 2D class, which was then used for automated particle picking. Particles were extracted with a box size of 128 pixels and aligned with 2D classification to generate a view onto the collapsed GV wall. The biggest class containing ~34,000 particles was selected and was treated in ImageJ using an 'unsharp mask' filter.

Pore analysis

Gas pores in the GV wall were analyzed using MOLE 2.5[21]. The gap between α1 helices has a slit-type shape, enabling multiple possible routes for gas diffusion. Several start and endpoints of tunnels around both sides of the slit were selected and tunnels were computed with MOLE. The slit was modeled by three tunnels and displayed in Chimera[45]. The minimal constriction of the tunnel was calculated as the diameter of the smallest sphere in the respective tunnel model.

Pseudo-atomic modeling of a complete gas vesicle

The model was generated from the solved atomic model of *B. megaterium* GvpA2. The GvpA2 monomer was placed next to the x axis (D1 axis) manually such that a 180°, symmetry copy operation around the x axis would reproduce a side view of the GV edge compatible with the determined 2D class average.

A left-handed parametric helix in 3D space was defined with the parameter t corresponding to turns of the helix: where tcaps is the number of turns in the cap, tcyl is a user-parameter of how many cylindrical turns the model should have (5), P is the pitch of the helix of 48.8 Å, rmax the radius of the assembly of 178.4 Å, and αis the cone angle of the tip (25°).

$$t_{cap} = \frac{r_{max}}{P \cdot sin(\alpha)}$$

$$r(t) = \begin{cases} r_{max}, & \text{if } t < t_{cyl} \\ r_{max} \cdot (1 - \frac{t-t_{cyl}}{t_{cap}}) & \text{if } t \geq t_{cyl} \end{cases}$$

$$x(t) = \quad r(t) \cdot cos(2\pi t)$$

$$y(t) = -r(t) \cdot sin(2\pi t)$$

$$z(t) = \begin{cases} P \cdot t, & \text{if } t < t_{cyl} \\ P \cdot t_{cyl} + P \cdot cos(\alpha) \cdot (t - t_{cyl}), & \text{if } t \geq t_{cyl} \end{cases}$$

The starting point of the curve was adjusted to go through the pivot point (in the center of the two β-sheets between amino acid (AA) 28 and AA 42 carbonyl oxygen atom) of the placed monomer by applying a shift along the z axis and rotation around the z axis. Points were placed along the curve at a distance of 12.07 Å, calculated as sqrt[(2πr)2 + P2]/ut, where ut is the number of monomers per turn defined by the solved helical symmetry (92.93). 835 points were placed with the last 4 points toward the tip being omitted.

A model only with placement of monomers does not reflect the experimental 2D classes well. Four additional parameters were introduced rotating and modifying the monomer. (1) a correction for the change of helix angle toward the ends of the tip as the helix becomes steeper with narrower radius. The rotation center is between the AA28 and AA42 carbonyl oxygen atoms, and the rotation axis normal to a fitted plane through C-α atoms of all amino acids of the β-hairpin (AA23-49).

(2) a rotation of the entire monomer to account for the tilt of monomers following the cone angle. The rotation center is carbonyl oxygen 36 and the axis normal to the plane of AA24-33 C-β atoms.

(3) a hinging motion of the two β-strands to account for the deformation visible in 2D classes of the PRP where GvpA ribs from both GVs halves meet. The rotation axis is defined by a line between the AA23 C-α atom and the AA49 C-α atom, and moves the atoms of AA23-49.

(4) a hinging motion of the α-helix 1 and the N-terminus, to account for gaps in the assembly formed when monomers tilt toward the cone angle. The axis is defined as in (3), and the rotation moves the atoms of AA2-23.

For the purpose of illustrating GV growth from nuclei to mature cylinders, the model was replicated in Blender 3.0 in a simplified form (without modifications of the monomer) and animated using varying parameters for the diameter and the tcyl parameter.

Alphafold2 structure predictions

Complexes of five monomers of GvpA corresponding to a rib section of the GV wall were predicted using AlphaFold2[46] run in a Google Colab environment[47] with each five generated models per run and six recycles. The highest-scoring model was selected and displayed in ChimeraX[40].

4.6 References

1. Huber, S. T., Terwiel, D., Evers, W. H., Maresca, D. & Jakobi, A. J. Cryo-EM structure of gas vesicles for buoyancy-controlled motility. *Cell* **186,** 975–986 (2023).

2. Szurmant, H. & Ordal, G. W. Diversity in Chemotaxis Mechanisms among the Bacteria and Archaea. *Microbiology and Molecular Biology Reviews* **68,** 301–319. ISSN: 1092-2172 (2004).

3. Jarrell, K. F. & McBride, M. J. The surprisingly diverse ways that prokaryotes move. *Nature Reviews Microbiology* **6,** 466–476. ISSN: 17401526 (2008).

4. Walsby, A. E. Gas vesicles. *Microbiological Reviews* **58,** 94–144. ISSN: 01460749 (1994).

5. Hayes, P. & Walsby, A. The inverse correlation between width and strength of gas vesicles in cyanobacteria. *British phycological journal* **21,** 191–197 (1986).

6. Blaurock, A. E. & Walsby, A. E. Crystalline structure of the gas vesicle wall from Anabaena flos-aquae. *Journal of Molecular Biology* **105,** 183–199. ISSN: 00222836 (1976).

7. Offner, S., Ziese, U., Wanner, G., Typke, D. & Pfeifer, F. Structural characteristics of halobacterial gas vesicles. *Microbiology* **144,** 1331–1342. ISSN: 13500872 (1998).

8. Walsby, A. E. Permeability of gas vesicles to perfluorocyclobutane. *Journal of General Microbiology* **128,** 1679–1684. ISSN: 00221287 (1982).

9. Pfeifer, F. Distribution, formation and regulation of gas vesicles. *Nature Reviews Microbiology* **10,** 705–715. ISSN: 17401526 (2012).

10. McMaster, T. J., Miles, M. J. & Walsby, A. E. Direct observation of protein secondary structure in gas vesicles by atomic force microscopy. *Biophysical Journal* **70,** 2432–2436. ISSN: 00063495 (1996).

11. Sivertsen, A. C., Bayro, M. J., Belenky, M., Griffin, R. G. & Herzfeld, J. Solid-State NMR Evidence for Inequivalent GvpA Subunits in Gas Vesicles. *Journal of Molecular Biology* **387,** 1032–1039. ISSN: 00222836 (2009).

12. Sivertsen, A. C., Bayro, M. J., Belenky, M., Griffin, R. G. & Herzfeld, J. Solid-state NMR characterization of gas vesicle structure. *Biophysical Journal* **99,** 1932–1939. ISSN: 15420086 (2010).

13. Strunk, T., Hamacher, K., Hoffgaard, F., Engelhardt, H., Zillig, M. D., Faist, K., Wenzel, W. & Pfeifer, F. Structural model of the gas vesicle protein GvpA and analysis of GvpA mutants in vivo. *Molecular Microbiology* **81,** 56–68. ISSN: 0950382x (2011).

14. Bayro, M. J., Daviso, E., Belenky, M., Griffin, R. G. & Herzfeld, J. An amyloid organelle, solid-state NMR evidence for cross-β assembly of gas vesicles. *Journal of Biological Chemistry* **287,** 3479–3484. ISSN: 00219258 (2012).

15. Ezzeldin, H. M., Klauda, J. B. & Solares, S. D. Modeling of the major gas vesicle protein, GvpA: From protein sequence to vesicle wall structure. *Journal of Structural Biology* **179,** 18–28. ISSN: 10478477 (2012).

16. Knitsch, R., Schneefeld, M., Weitzel, K. & Pfeifer, F. Mutations in the major gas vesicle protein GvpA and impacts on gas vesicle formation in Haloferax volcanii. *Molecular Microbiology* **106,** 530–542. ISSN: 13652958 (2017).

17. Li, N. & Cannon, M. C. Gas vesicle genes identified in Bacillus megaterium and functional expression in Escherichia coli. *Journal of Bacteriology* **180**, 2450–2458. issn: 00219193 (1998).

18. Lakshmanan, A., Lu, G. J., Farhadi, A., Nety, S. P., Kunth, M., Lee-Gosselin, A., Maresca, D., Bourdeau, R. W., Yin, M., Yan, J., Witte, C., Malounda, D., Foster, F. S., Schröder, L. & Shapiro, M. G. Preparation of biogenic gas vesicle nanostructures for use as contrast agents for ultrasound and MRI. *Nature Protocols* **12**, 2050–2080. issn: 17502799 (2017).

19. Dutka, P., Malounda, D., Metskas, L. A., Chen, S., Hurt, R. C., Lu, G. J., Jensen, G. J. & Shapiro, M. G. Measuring gas vesicle dimensions by electron microscopy. *Protein Science* **30**, 1081–1086. issn: 1469896x (2021).

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

21. Sehnal, D., Vařeková, R. S., Berka, K., Pravda, L., Navrátilová, V., Banáš, P., Ionescu, C.-M., Otyepka, M. & Koča, J. MOLE 2.0: advanced approach for analysis of biomacro-molecular channels. *Journal of cheminformatics* **5**, 1–13 (2013).

22. Ismail, A. F., Khulbe, K. C. & Matsuura, T. Gas separation membranes. *Switz. Springer* **10**, 973–978 (2015).

23. Horgan, C. & Murphy, J. Magic angles for fibrous incompressible elastic materials. *Proceedings of the Royal Society A* **474**, 20170728 (2018).

24. Walsby, A. E. The elastic compressibility of gas vesicles. *Proceedings of the Royal Society of London - Biological Sciences* **216**, 355–368. issn: 09628452 (1982).

25. Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., de Beer, T. A. P., Rempfer, C., Bordoli, L., *et al.* SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic acids research* **46**, W296–w303 (2018).

26. Walsby, A. E. The permeability of blue-green algal gas-vacuole membranes to gas. *Proceedings of the Royal Society of London. Series B. Biological Sciences* **173**, 235–255. issn: 0080-4649 (1969).

27. Tashiro, Y., Monson, R. E., Ramsay, J. P. & Salmond, G. P. Molecular genetic and physical analysis of gas vesicles in buoyant enterobacteria. *Environmental Microbiology* **18**, 1264–1276. issn: 14622920 (2016).

28. Dunton, P. G., Mawby, W. J., Shaw, V. A. & Walsby, A. E. Analysis of tryptic digests indicates regions of GvpC that bind to gas vesicles of Anabaena flos-aquae. *Microbiology* **152**, 1661–1669. issn: 13500872 (2006).

29. Völkner, K., Jost, A. & Pfeifer, F. Accessory Gvp Proteins Form a Complex During Gas Vesicle Formation of Haloarchaea. *Frontiers in Microbiology* **11.** issn: 1664302x (2020).

30. Bourdeau, R. W., Lee-Gosselin, A., Lakshmanan, A., Farhadi, A., Kumar, S. R., Nety, S. P. & Shapiro, M. G. Acoustic reporter genes for noninvasive imaging of microorganisms in mammalian hosts. *Nature* **553**, 86–90. issn: 14764687 (2018).

31. Shapiro, M. G., Goodwill, P. W., Neogy, A., Yin, M., Foster, F. S., Schaffer, D. V. & Conolly, S. M. Biogenic gas nanostructures as ultrasonic molecular reporters. *Nature nanotechnology* **9,** 311–316 (2014).

32. Maresca, D., Lakshmanan, A., Abedi, M., Bar-Zion, A., Farhadi, A., Lu, G. J., Szablowski, J. O., Wu, D., Yoo, S. & Shapiro, M. G. Biomolecular ultrasound and sonogenetics. *Annual Review of Chemical and Biomolecular Engineering* **9,** 229–252. ISSN: 19475438 (2018).

33. Heiles, B., Terwiel, D. & Maresca, D. The advent of biomolecular ultrasound imaging. *Neuroscience* **474,** 122–133 (2021).

34. Maresca, D., Lakshmanan, A., Lee-Gosselin, A., Melis, J. M., Ni, Y.-L., Bourdeau, R. W., Kochmann, D. M. & Shapiro, M. G. Nonlinear ultrasound imaging of nanoscale acoustic biomolecules. *Applied physics letters* **110,** 073704 (2017).

35. Maresca, D., Sawyer, D. P., Renaud, G., Lee-Gosselin, A. & Shapiro, M. G. Nonlinear X-wave ultrasound imaging of acoustic biomolecules. *Physical Review X* **8,** 041002 (2018).

36. Lakshmanan, A., Farhadi, A., Nety, S. P., Lee-Gosselin, A., Bourdeau, R. W., Maresca, D. & Shapiro, M. G. Molecular engineering of acoustic protein nanostructures. *ACS nano* **10,** 7314–7322 (2016).

37. Lakshmanan, A., Jin, Z., Nety, S. P., Sawyer, D. P., Lee-Gosselin, A., Malounda, D., Swift, M. B., Maresca, D. & Shapiro, M. G. Acoustic biosensors for ultrasound imaging of enzyme activity. *Nature chemical biology* **16,** 988–996 (2020).

38. Zivanov, J., Nakane, T., Forsberg, B. O., Kimanius, D., Hagen, W. J., Lindahl, E. & Scheres, S. H. New tools for automated high-resolution cryo-EM structure determination in RELION-3. en. *Elife* **7** (2018).

39. Zhang, K. Gctf: Real-time CTF determination and correction. en. *J. Struct. Biol.* **193,** 1–12 (2016).

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

41. Emsley, P., Lohkamp, B., Scott, W. & Cowtan, K. Features and Development of\textit {Coot}. *Acta Crystallogr D Struct. Biol.* (2010).

42. Croll, T. I. ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallographica Section D: Structural Biology* **74,** 519–530 (2018).

43. Liebschner, D., Afonine, P. V., Baker, M. L., Bunkóczi, G., Chen, V. B., Croll, T. I., Hintze, B., Hung, L.-W., Jain, S., McCoy, A. J., *et al.* Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallographica Section D: Structural Biology* **75,** 861–877 (2019).

44. Bepler, T., Morin, A., Rapp, M., Brasch, J., Shapiro, L., Noble, A. J. & Berger, B. Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nature methods* **16,** 1153–1160 (2019).

4

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

46.   Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596,** 583–589 (2021).

47.   Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S. & Steinegger, M. ColabFold: making protein folding accessible to all. *Nature methods* **19,** 679–682 (2022).

4

5

# BINDING SITE AND CONSERVATION OF THE GAS VESICLE BINDING PROTEIN GvpC

Gas vesicles in many species are reinforced by a secondary structural protein GvpC that is bound to the outside of the gas vesicle wall. While the role of this protein is firmly established, its binding geometry and interaction surface remains elusive. By employing cryo-electron microscopy in conjunction with sequence analysis and the experimental examination of point mutants, we elucidate the binding geometry, interaction surface, and the specific residues involved in this interaction. This insight paves the way for future rational engineering of GvpC, opening avenues for new application in biotechnology.

5

5.1 Graphical Abstract
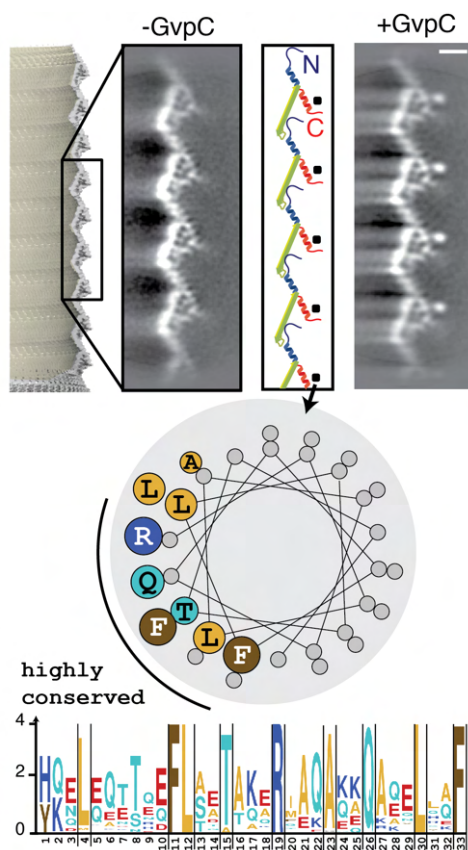
5.2 Introduction

In the previous chapter, we determined the atomic structure and the assembly of the gas vesicle wall composed of the major structural protein GvpA. Many gas vesicle gene clusters contain a second structural GV protein GvpC, which is absent from our model GV from *B. megaterium*[2]. This protein binds the exterior of the GV and provides additional structural reinforcement[3,4]. GvpC increases the critical collapse pressure of GVs[3,5]. While the essential role of GvpC has been firmly established, how it binds to and reinforces GVs has remained elusive.

Cyanobacterial GvpC proteins encompass multiple 33 amino acid repeats. Possible ways in which these repeats may form periodic interactions with the underlying crystalline GvpA assembly have been proposed[4]. These include binding along the gas vesicle ribs - or binding across gas vesicle ribs, suggested to stabilise gas vesicles by holding the ribs together. Since it has now become possible to image gas vesicles with high details using cryo-EM, this question can finally be answered.

5.3 Results

5.3.1 Reinforcement of the gas vesicle shell by GvpC

In order to find out how cyanobacterial GvpC binds to gas vesicles, we acquired cryo-EM images of *A. flos-aquae* GVs, a gas vesicle with a GvpC gene in its gene cluster. Gas vesicles from this organism have a large 87 nm cylinder diameter and are not suitable for helical reconstruction techniques, compared to the small 36 nm *B. megaterium* gas vesicles of which the wall structure was solved in Chapter 4. Despite extensive efforts to use a similar technique on those large gas vesicles, the helical reconstruction procedure did not converge to a 3D reconstruction that would resolve 3D features along the gas vesicle ribs. Structural flattening in the vitreous ice of cryo-EM samples leads to distortion of the cylinder shape and violates assumptions of helicity necessary for using helical reconstruction techniques.

We showed in Chapter 4 that detailed 2D class averages of the gas vesicle wall can be computed with resolutions exceeding 5.4 Å, resolving the α-helical pitch of the structure. To achieve this, we averaged many segments of the rims of gas vesicles picked from cryo-EM images together to increase the signal and average out the noise. To investigate the structural role of GvpC, we imaged *A. flos-aquae* GVs in the presence and absence of GvpC. Detailed 2D class averages with better than 5.4 Å resolution were computed from the datasets and showed structural features of the gas vesicle wall in a 2D side-view (Figure 5.1A).

The class average containing GvpC shows an additional dot-like density (~10 Å diameter) located in vicinity to helix α2 (Figure 5.1A). This feature and its dimensions are consistent with the projection of an α helix viewed along its helical axis. Therefore, many GvpC molecules wrap around the gas vesicle following the helical pitch of the gas vesicle ribs (Figure 5.1B). This alignment is supported by an 8-12 Ångstrom resolution molecular envelope of the *A. flos-aquae* GV wall determined by electron tomography[6]. GvpC is predicted to be all α-helical[4] and consists of five 33-residue repeats (Figures 5.1C) that are highly similar in sequence.

We analyzed the evolutionary conservation of this 33 amino acid repeat sequence. In a set of 91 GvpC repeat sequences (Figure 4.10B) from different organisms, we find
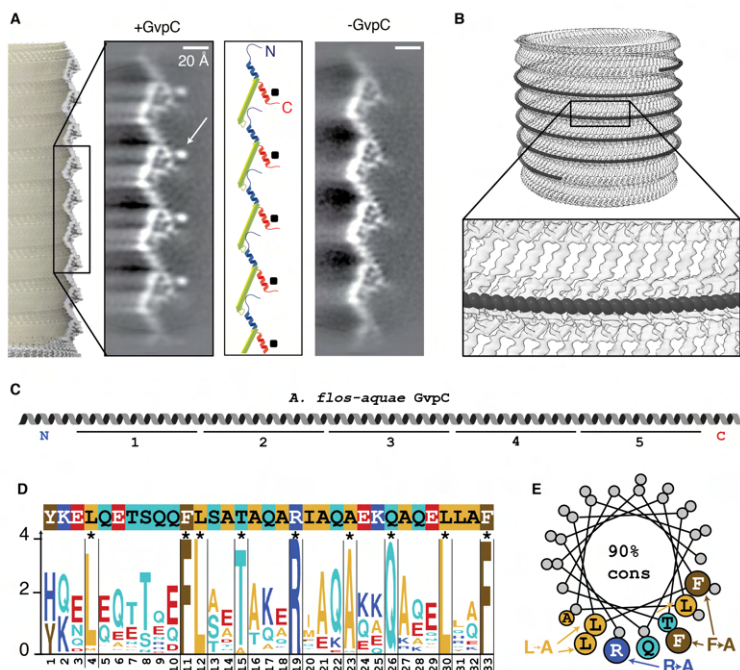
5

Figure 5.1: | **The secondary wall protein GvpC binds along the GvpA ribs of _A. flos-aquae_ GVs.** (A) Comparison of 2D class averages of GV edges with (left) and without (right) GvpC reveals an additional circular density (arrow). A cartoon model of GvpA helps locating the GvpC density to helix α2 (red). (B) Artist impression of GvpC molecules wrapping around GVs. (C) Predicted secondary structure and 33-amino-acid repeats 1–5 of _A. flos-aquae_ GvpC. (D) Consensus sequence of the GvpC repeats with logo representation of evolutionary conservation reveals nine highly conserved residues. The height of the characters depicts the degree of conservation (information content in bits).(E) Helical wheel plot of highly conserved (>90% conserved in 91 GvpC sequences from other species) amino acids reveal that there is one highly conserved face of the α-helical repeat. Experimentally tested GvpC mutants are indicated.
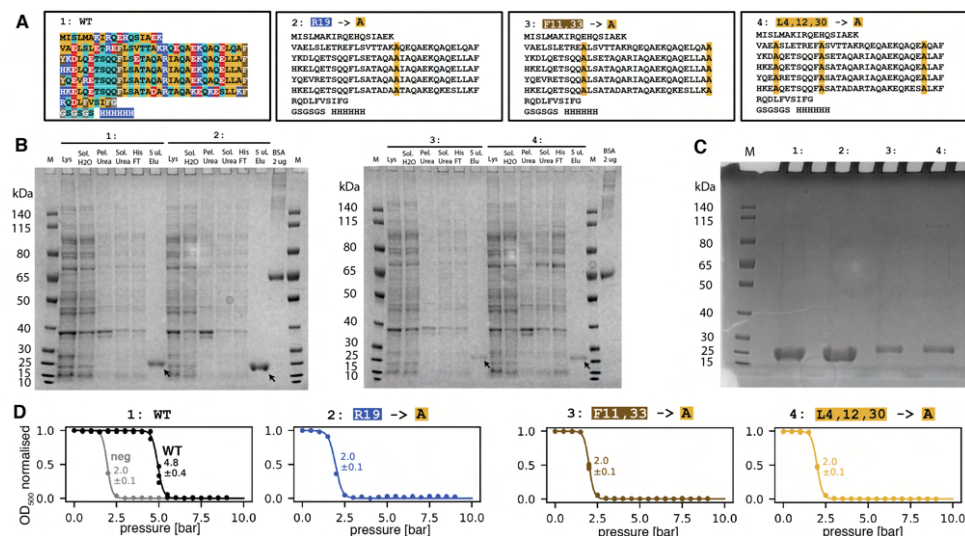
Figure 5.2: | **Collapse pressure measurements of GvpC mutants.** (A) A GvpC wild-type construct and three mutants were designed with F, L and R residues in all five repeats mutated to alanine. All constructs have a C-terminal GSGSGS linker and 6x His-tag. (B) SDS-PAGE following purification steps of wild-type GvpC and point mutants. Arrow indicates final product. Legend as follows: P:Pellet/SolH2O:Soluble fraction in aqueous solution/Pel.Urea:Pelleting fraction in urea-containing buffer/Sol.Urea:Soluble fraction in urea-containing buffer/HisFT:Flowthrough from Ni-NTA column/5uLElu:Eluted fraction from Ni-NTA column (C) Overloading an SDS-PAGE gel with GvpC mutant sample indicates high degree of purity. (D) Collapse pressure measurements of the four constructs confirm binding to *A. flos-aquae* GVs of the wild-type construct and loss of binding for the mutants. Stripped GVs before re-addition of GvpC were measured as a negative control. Circles are measurement points from three independent re-addition experiments. Solid curves are fits of a sigmoid function, with the stated number being the pressure when the normalized OD500 drops to 0.5.

nine residues to be conserved in more than 90% of those sequences, including a strongly conserved set of leucine (Leu4,12,30), phenylalanine (Phe11,33), and arginine (Arg19) residues (Figure 5.1D). A helical wheel plot of the repeat shows that all conserved residues cluster on the same face (Figure 5.1E) that likely forms the binding interface.

We set out to confirm the importance of these three residue types for the binding interface of GvpC and validate predictions based on evolutionary conservation. Four different constructs were synthesized, one wild-type GvpC, and one construct each with arginine, leucine or phenylalanine replaced by alanine for all five 33 amino acid repeats (Figure 5.2A). The protein was expressed and purified following established protocols[7] (Figure 5.2B,C). The mutant GvpC proteins were added to *A. flos-aquae* GVs that were previously stripped from GvpC[7] to create gas vesicles bound to the mutant GvpC. Collapse pressure measurements of those modified gas vesicles showed that all mutants lost their ability to stabilise gas vesicles (Figure 5.2D).

Our 2D class averages provide strong spatial restraints on the positioning of GvpC relative to the GvpA ribs in two dimensions, while the conservation pattern and mutants pinpoint residues essential for binding. We used these data as restraints to predict a model for GvpC binding by computational docking. Starting models of a GvpC consensus

repeat with perfect α-helical geometry, as well as a 7-mer of a GvpA homology model for *A. flos-aquae* based on our atomic model of *B. megaterium* were created. Additional information was input into HADDOCK by making only the nine amino acids which are highly conserved 'active' in the docking process. For GvpA, only α-helix 2 was defined as 'active' (Figure 5.3A-C).

While the resulting docking solutions do not allow us to decisively distinguish whether GvpC follows the left-handed spiral of the rib upwards toward the tips, or downwards, we obtained the highest-scoring docking solution with the downward orientation (Figure 5.3D). A single GvpC repeat spans four monomers of GvpA. In this GvpA tetrad, glutamate residues Glu52 and Glu59 in GvpA monomers 1, 2, and 4 form hydrogen bonds with GvpC. In our model, Glu52 in monomer 3 binds to the conserved Arg19 of GvpC, while Glu59 in monomer 4 binds to the conserved Gln26. The (Leu4, Leu12, Leu30) triplet inserts between the α2 helices of the GvpA tetrad, while Phe11 and Phe33 are sandwiched in between α2 in GvpA monomers 2 and 3 and face-on on helix α2 of monomer 4.

5.3.2 Binding of multiple repeats of GvpC

We inferred a binding mode of a single 33-amino-acid repeat of GvpC to four consecutive GvpA monomers assembled in the cylindrical *A. flos-aquae* GV shell. The requirement for a binding site comprised of consecutive GvpA monomers in a helical assembly is consistent with previous observations in split-GFP assays showing that monomeric GvpA and GvpC do not interact[8]. What remains open is how the five consecutive GvpC repeats are structured. It seems likely that all repeats would form identical molecular contacts with the GvpA ribs. The side-by-side distance of GvpA monomers measured between identical points of the inner β-hairpins is 12.1 Å. Due to the curvature of GVs, this distance increases radially outwards across the GV wall toward the binding site for GvpC. For an average 87 nm *A. flos-aquae* GV, this distance is ~12.7 Å. A single perfectly α-helical repeat of GvpC would span 49.5 Å (33 × 1.5 Å/residue), only slightly shorter than the 50.8 Å spanned by a stretch of four GvpA monomers. The five tandem repeats of native GvpC would bind to 20 GvpA monomers. Taking into account additional space for the N and C termini of GvpC, this corresponds well to a previously determined 1:25 M ratio of GvpA to GvpC[4]. The modest distance mismatch between one GvpC repeat and a GvpA tetrad could be accommodated by slight deviation of GvpC from perfect helicity. It has been previously suggested that this would also be required to maintain the relative orientation of consecutive repeats, as with perfect helicity they would be (100° × 33) % 360 = 60° rotated toward each other[4]. A small unstructured stretch would allow GvpC to adapt to different curvatures of the GvpA ribs in cylindrical and conical parts and in GVs of different diameters.

5.3.3 Repeating patterns in GvpC proteins

For a more rational analysis of repeat patterns in GvpC proteins, 442 unique GvpC sequences were downloaded from the UniRef100 database[9]. Each letter in the sequence was embedded into a vector space spanned by the columns in the BLOSUM62 similarity matrix[10] (Figure 5.4A). Similar amino acids are fall in similar positions in this high-dimensional space (Figure 5.4B). The embedded sequences were used for an autocorrelation analysis to reveal repetetive patterns (Figure 5.4C). The autocorrelation analysis revealed strong self-similarity of the sequence every 33 amino acids, corresponding to approximately four repeats of
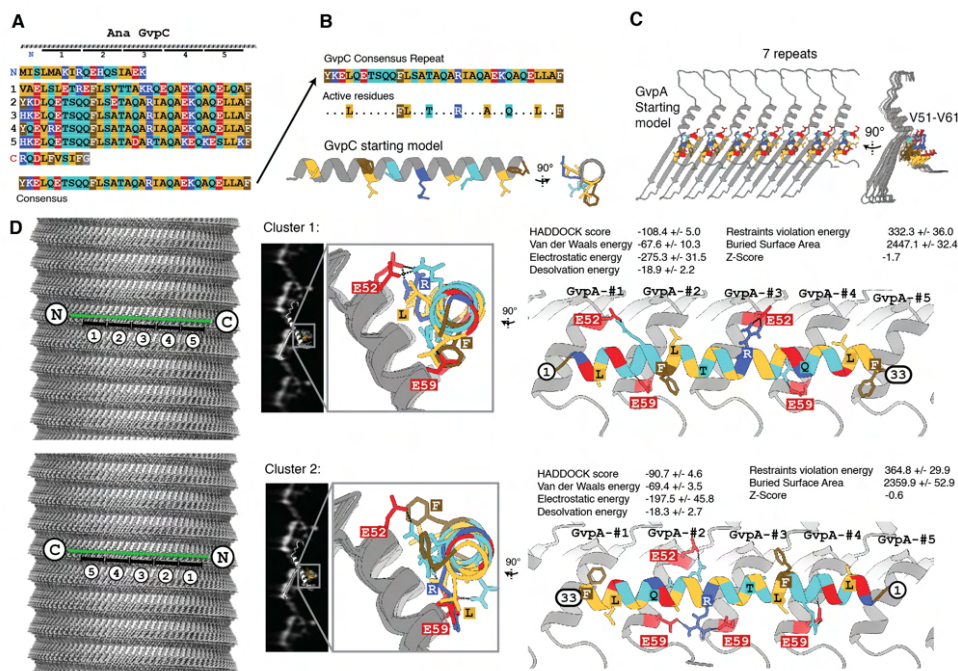
Figure 5.3: | **Binding analysis of GvpC to *A. flos-aquae* gas vesicles.** (A) Primary sequence of Ana GvpC with 5 repeats and the consensus sequence of the repeats. (B) Identified highly conserved residues in the repeat are highlighted and chosen as 'active residues' in HADDOCK. A perfect α-helical peptide is used as the starting model. (C) Amino acids V51 to V61 on α-helix 2 are chosen as the active residues (on a homology model of GvpA for *A. flos-aquae*) based on observed close proximity in 2D class averages. GvpA was repeated 7 times according to the helical symmetry of the solved *B. megaterium* assembly. (D) Two possible binding geometries of GvpC are shown on a *B. megaterium* GV density oriented with the seam on the bottom and the tip on the top: direction from C to N-terminus is following the left-handed loop of the helix, or reverse. HADDOCK protein docking results between GvpC and GvpA fall into two main clusters, one highest-scoring solution with a score of -108.4 and a solution with reverse binding polarity of GvpC with a score of -90.7. The best solution of each cluster is shown. The solution was fitted into 2D class averages of *A. flos-aquae* edges.

GvpA in the gas vesicle rib (Figure 5.4D,E). Correlation was also high every 3-4 amino acids, showing a repeating pattern in the autocorrelation function. This can be attributed to the different amino acid compositions between the binding side of the α-helix compared to the cytosolic side. Several smaller peaks at 18 or 51 amino acids appear in the autocorrelation functions, and roughly correspond to two and six GvpA repeats along the rib. These peaks may belong to GvpC sequences in the UniProt database that follow different repeat patterns.
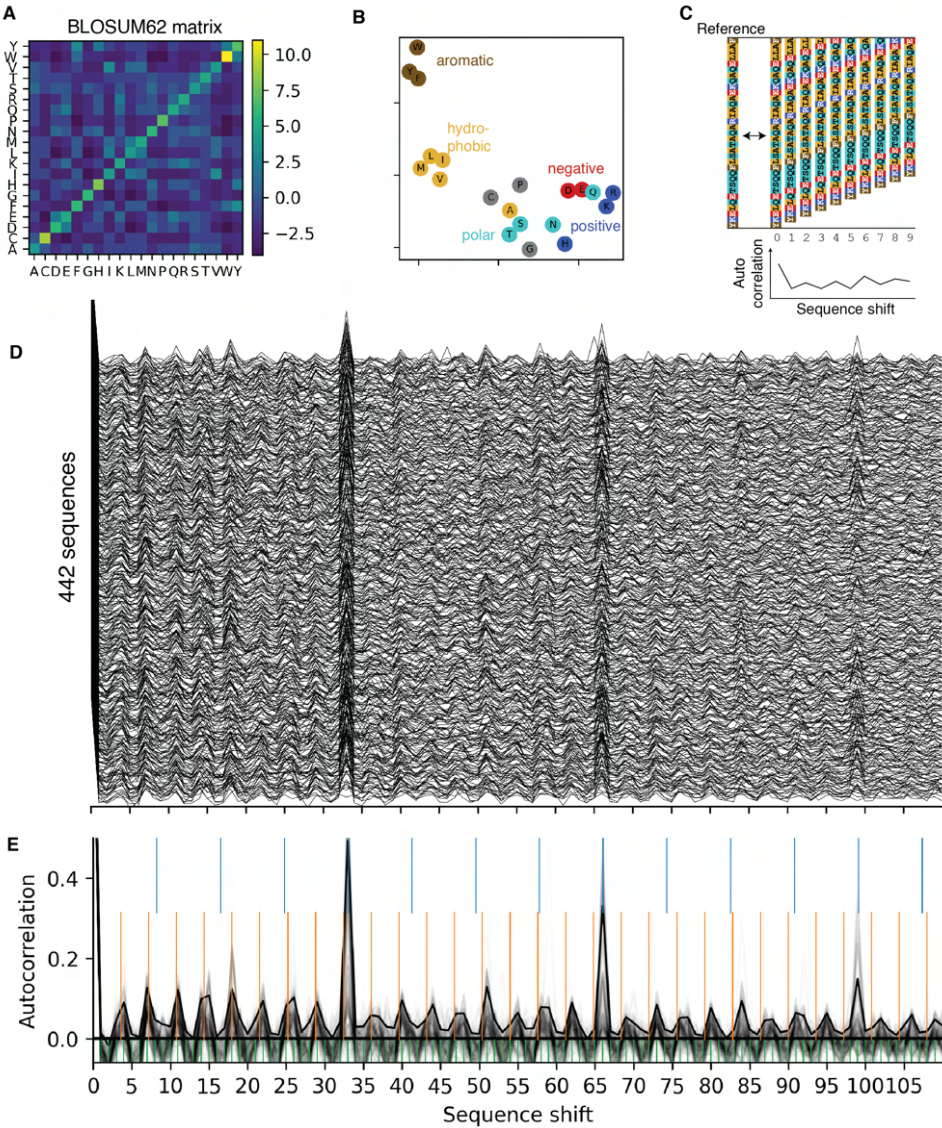
5

Figure 5.4: | **Autocorrelation analysis of GvpC sequences across the UniProt database.** (A) BLOSUM62 matrix[11] (B) Each column in the BLOSUM62 matrix can be interpreted as a vector for the position of the amino acid in a high-dimensional space. This space is visualised using a two-dimensional t-SNE embedding[12]. The amino acids are clustered in this space by their physico-chemical properties (colored). (C) Illustration of sequence autocorrelation analysis. The sequence is shifted with respect to the reference, and a similarity score based on the BLOSUM62 matrix is calculated for each shift. (D) Autocorrelation profiles of 442 sequences GvpC sequences from the UniRef100 database. (E) Aggregated profiles. Thick line shows 90th percentile of the autocorrelation traces of all sequences. GvpA spacing (blue), alpha-helical pitch spacing (orange), amino acid monomer spacing (green).

5.4 Discussion

In this Chapter, we reveal the binding geometry of GvpC to the gas vesicle wall - it binds along the ribs of gas vesicles and interacts with α-helix 2 of GvpA. The enhanced stability of gas vesicles when GvpC is attached can be explained from this finding: the wall thickness measured between the extremes of the corrugation pattern increases from ~3 nm to ~4 nm when GvpC is present. Because the buckling pressure of thin-walled cylinders is proportional to the third power of the wall thickness[13], the most straightforward explanation for enhanced stability is that GvpC bound on the outside of GVs increases the wall thickness. The real picture however is likely more complicated. Our structure from Chapter 4 reveals that GVs are assembled from two helical half shells that are connected by a seam made from interacting hydrophobic amino acids. Because of this structure, it is conceivable that the seam is a weak point of the GV where failure will occur first. In this case, GvpC would be most critical to prevent failure at the seam, which questions its requirement elsewhere on the GV shell.

Autocorrelation analysis of GvpC sequences from the UniProt database[9] showed that most sequences have a 33 amino acid repeat sequence as well as a strong bias of amino acid composition equivalent to every half turn of an α-helix. Other prominent peaks correspond to 2 and 6 repeats of the underlying GvpA rib and warrant further investigation. Are there organisms where GvpC has a different repeating pattern that deviates from the 33 amino acid window? GvpC in haloarchaea is a known example for that, with less strict repeating patterns of 32-38 amino acid sequences. Only a few sequences in the Uniprot database are from archaea, and did not significantly influence the autocorrelation analysis. It is an open question how GvpC in haloarchaea binds to haloarchaeal gas vesicles and if the geometry is equivalent to cyanobacteria such as *A. flos-aquae*.

5.5 Methods and Materials

GvpC mutant purification

Codon-optimized genes for *A. flos-aquae* wild-type and mutant (R19A; F-11,33-A; L-4,12,30-A) GvpC (Uniprot: P09413) including a C-terminal 'GSGSGS' linker and a C-terminal 6xHis-tag in a pET-28a(+) vector were obtained from Genscript (New Jersey, United States). The mutations were engineered into all five repeats.

Proteins were expressed in *E.coli* BL21-DE3 cells grown in 750 mL autoinduction medium57 for 3 h at 37°C before the temperature was lowered to 20°C for additional 20 h. The bacteria were harvested by centrifugation and lysed by freeze-thaw cycles in lysis buffer (20 mM Tris, 500 mM NaCl, 0.1% Triton X-, 20 mM imidazole; 5 mL per gram of pellet). Lysozyme (0.15 mg/mL) and DNAseI (10 μg/mL) were added and the lysate rotated for 2 h at room temperature. Isolation of inclusion body and IMAC purification was performed as described previously[7]. Protein purity was assessed by SDS-PAGE and concentration was determined according to Bradford.

Preparation of gas vesicles with recombinant GvpC and collapse pressure measurements

A. flos-aquae GVs were stripped of GvpC by resuspending in 6 M urea and 60 mM Tris buffer, using 3 rounds of flotation separation as previously described[7]. Recombinant GvpC

was added to stripped GVs according to the formulation: 2 x OD500 × 198 nM x GV volume (L) = nmol GvpC. GvpC will be present in a 2-fold excess under the assumption of a 1:25 M ratio of GvpC/GvpA[7]. The urea solution was then slowly replaced with PBS at pH 7.4 by 2 rounds of ~12 h of dialysis over a 7–10 kDA MWCO dialysis membrane. Finally, 3 rounds of floatation separation at 350 rcf removed traces of urea. GVs were diluted to OD500 = 0.1–0.4 for collapse pressure measurements. Samples were loaded in a pressure vessel and hydrostatic pressure was increased in increments of 0.5 bar using pressurized nitrogen gas. Samples were allowed to equilibrate for 5 s after pressure changes before absorption was measured using a spectrophotometer (Ocean optics) at 500 nm OD500 values were normalized between the minimum and maximum for each measurement. Three independent re-additions per GvpC mutant were performed and measured.

A sigmoid function with p0 as the inflection point and k as the width was fitted to the curves using the means and standard deviations of measured triplicates (n = 3) as input for the scipy *'curve_fit'* function.

$$OD_{500,norm} = \frac{1}{1 + e^{k(p-p0)}}$$

The error of p0 was determined using a bootstrapping approach, performing the fit 50 times with a random set of n out of n measurement points with replacement. The parameters and their uncertainty were estimated as their mean and SD over the 50 fits.

Conservation analysis

The sequences of the five 33 AA repeats of *A. flos-aquae* GvpC (Uniprot P09413) was converted into a consensus sequence 'YKELQETSQQFLSATAQARIAQAEKQAQELLAF' using the software 'cons' from the EMBOSS package[14]. The sequence was used as input for the ConSurf Server[15] to search the Uniref90 database with the HMMER web server[16] using one iteration, resulting in 91 sequences. The resulting sequence alignment was displayed and consensus sequences calculated in MView[17]. A sequence logo was calculated from the sequence alignment using WebLogo[18]. Helical wheel plots were generated in Heliquest[19].

HADDOCK modeling of GvpC binding to *A. flos-aquae* gas vesicles

A α-helical model of the 33 residue consensus repeat of *A. flos-aquae* GvpC ('YKELQET-SQQFLSATAQARIAQAEKQAQELLAF') was generated in ChimeraX[20] using ideal α-helical backbone dihedral angles φ = -57° and ψ = -47°. A homology model of the *A. flos-aquae* GvpA monomer was generated using SWISS-MODEL[21] based on the structure of *B. megaterium* structure GvpA2 and the *A. flos-aquae* sequence for GvpA (UniProt: P10397). The homology model was extended with the symmetry parameters from the *B. megaterium* assembly in ChimeraX43 into a rib of 7 adjacent monomers.

Both models were used as input for HADDOCK 2.4[22]. For GvpC, residues

4,11,12,15,19,23,26,30,33

of the consensus repeat, which all are >90% conserved in the bioinformatics analysis

were chosen as active residues. For GvpA, residues 51–61 of GvpA, part of helix α2 and adjacent to the GvpC density in the 2D classes, were chosen as active residues. All remaining settings were used at default values.

The highest scoring cluster of the docking solutions (HADDOCK score: -108.4 ± 5.0) showed GvpC binding across several GvpA monomers along the helical spiral of the rib, with the GvpC sequence oriented inversely to the direction of helical propagation (with the direction from seam to tip). A very similar cluster was found shifted by one GvpA monomer laterally with the same molecular contacts and was discarded. A second, lower scoring type of cluster (HADDOCK score: -90.7 ± 4.6) showed GvpC binding mode with the GvpC sequence oriented to the direction of helical propagation.

Docking clusters were displayed in ChimeraX[20] and hydrogen bonds between GvpA and GvpC highlighted with the 'hbonds' command. The highest-scoring solution was manually fitted into a 2D class of the *A. flos-aquae* GV wall with GvpC.

Autocorrelation of GvpC sequences

Non-redundant GvpC sequences (n=474) were retrieved from the UniRef100 database using the search term 'gvpc NOT gvpa NOT gvpn NOT gvpf'. Sequences with at least 100 residues were retained (n=442). Each position in the sequence was encoded by a vector from the BLOSUM62[10] matrix, and autocorrelaction of the sequences was computed in python, normalised by the length of the overlap window that is compared.

5.6 REFERENCES

1. Huber, S. T., Terwiel, D., Evers, W. H., Maresca, D. & Jakobi, A. J. Cryo-EM structure of gas vesicles for buoyancy-controlled motility. *Cell* **186,** 975–986 (2023).

2. Tashiro, Y., Monson, R. E., Ramsay, J. P. & Salmond, G. P. Molecular genetic and physical analysis of gas vesicles in buoyant enterobacteria. *Environmental Microbiology* **18,** 1264–1276. ISSN: 14622920 (2016).

3. Walsby, A. E. & Hayes, P. K. The Minor Cyanobacterial Gas Vesicle Protein, GVPc, Is Attached to the Outer Surface of the Gas Vesicle. *Microbiology* **134,** 2647–2657. ISSN: 1350-0872 (1988).

4. Buchholz, B. E., Hayes, P. K. & Walsby, A. E. The distribution of the outer gas vesicle protein, GvpC, on the Anabaena gas vesicle, and its ratio to GvpA. *Journal of General Microbiology* **139,** 2353–2363. ISSN: 00221287 (1993).

5. Dunton, P. G., Mawby, W. J., Shaw, V. A. & Walsby, A. E. Analysis of tryptic digests indicates regions of GvpC that bind to gas vesicles of Anabaena flos-aquae. *Microbiology* **152,** 1661–1669. ISSN: 13500872 (2006).

6. Dutka, P., Metskas, L. A., Hurt, R. C., Salahshoor, H., Wang, T.-Y., Malounda, D., Lu, G., Chou, T.-F., Shapiro, M. G. & Jensen, G. J. Structure of Anabaena flos-aquae gas vesicles revealed by cryo-ET. *Structure* **122,** 40a (2023).

7. Lakshmanan, A., Lu, G. J., Farhadi, A., Nety, S. P., Kunth, M., Lee-Gosselin, A., Maresca, D., Bourdeau, R. W., Yin, M., Yan, J., Witte, C., Malounda, D., Foster, F. S., Schröder, L. & Shapiro, M. G. Preparation of biogenic gas vesicle nanostructures for use as contrast agents for ultrasound and MRI. *Nature Protocols* **12,** 2050–2080. ISSN: 17502799 (2017).

8. Völkner, K., Jost, A. & Pfeifer, F. Accessory Gvp Proteins Form a Complex During Gas Vesicle Formation of Haloarchaea. *Frontiers in Microbiology* **11.** ISSN: 1664302x (2020).

9. Bateman, A., Martin, M.-J., Orchard, S., Magrane, M., Ahmad, S., Alpi, E., Bowler-Barnett, E. H., Britto, R., Bye-A-Jee, H., Cukura, A., *et al.* UniProt: the universal protein knowledgebase in 2023. *Nucleic Acids Research* **51** (2022).

10. Henikoff, S. & Henikoff, J. G. Amino acid substitution matrices from protein blocks. en. *Proceedings of the National Academy of Sciences* **89,** 10915–10919. ISSN: 0027-8424, 1091-6490 (Nov. 1992).

11. Henikoff, S. & Henikoff, J. G. Amino acid substitution matrices from protein blocks. *Proceedings of the National Academy of Sciences* **89,** 10915–10919 (1992).

12. Van der Maaten, L. & Hinton, G. Visualizing data using t-SNE. *Journal of machine learning research* **9** (2008).

13. Walsby, A. E. The elastic compressibility of gas vesicles. *Proceedings of the Royal Society of London - Biological Sciences* **216,** 355–368. ISSN: 09628452 (1982).

14. Rice, P., Longden, I. & Bleasby, A. EMBOSS: the European molecular biology open software suite. *Trends in genetics* **16,** 276–277 (2000).

15. Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T. & Ben-Tal, N. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic acids research* **44,** W344–w350 (2016).

5

16. Finn, R. D., Clements, J. & Eddy, S. R. HMMER web server: interactive sequence similarity searching. *Nucleic acids research* **39,** W29–w37 (2011).

17. Brown, N. P., Leroy, C. & Sander, C. MView: a web-compatible database search or multiple alignment viewer. *Bioinformatics (Oxford, England)* **14,** 380–381 (1998).

18. Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome research* **14,** 1188–1190 (2004).

19. Gautier, R., Douguet, D., Antonny, B. & Drin, G. HELIQUEST: a web server to screen sequences with specific α-helical properties. *Bioinformatics* **24,** 2101–2102 (2008).

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

21. Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., de Beer, T. A. P., Rempfer, C., Bordoli, L., *et al.* SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic acids research* **46,** W296–w303 (2018).

22. Van Zundert, G., Rodrigues, J., Trellet, M., Schmitz, C., Kastritis, P., Karaca, E., Melquiond, A., van Dijk, M., De Vries, S. & Bonvin, A. The HADDOCK2. 2 web server: user-friendly integrative modeling of biomolecular complexes. *Journal of molecular biology* **428,** 720–725 (2016).

5

6

FROM PROTEIN STRUCTURES TO EDUCATIONAL MODELS

The chapter discusses the use of 3D printing, specifically filament deposition modeling (FDM), for the fabrication of educational models of protein structures. Basic design principles relevant for this process are outlined, including transformation of structural biology data, including cryo-electron microscopy (cryo-EM) density maps and Protein Data Bank (PDB) files, into models suitable for 3D printing. The design of three example models is explained in detail, based on structural data from this thesis. These models serve to illustrate concepts such as point group symmetry in proteins, the growth process of gas vesicles, and the folding geometry of protein chains.

6

6.1 Graphical Abstract
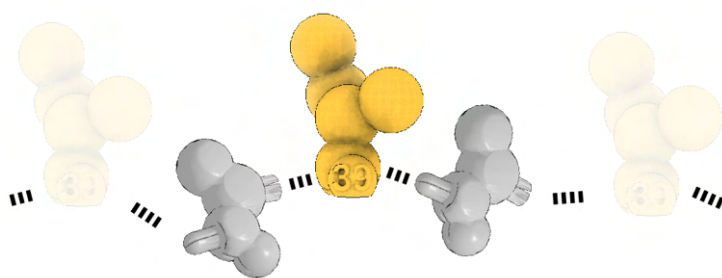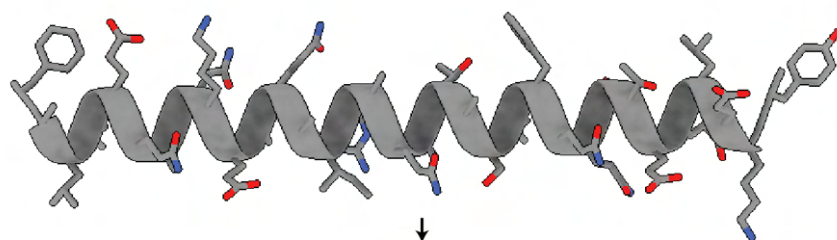
6.2 Introduction

Filament deposition modeling (FDM) 3D printers have become increasingly accessible, with costs reducing to under €500. Despite this, their adoption in scientific research laboratories has been slow, often limited by perceptions of their applicability[1]. This chapter demonstrates the utility of FDM 3D printers in a scientific setting. Throughout the duration of this thesis, a variety of lab tools have been designed and fabricated via 3D printing, including sample holders, transport containers and microscope adapters, as shown in Figure 6.1. Structural parts of the time-resolved plunge freezing setup in Chapter 3, such as the humidity chamber, water tanks and electronic casings were also quickly and affordably fabricated using a 3D printer.

This chapter extends the scope of 3D printing from its obvious role in fabricating laboratory hardware to its use in converting structural data from cryo-EM into educational models. These models can be instrumental in transmitting scientific concepts, underscoring the superiority of visual aids like images[2] and animations[3] over purely textual descriptions in communicating scientific ideas[4]. To this end, we discuss three exemplary models that are designed to enhance the understanding of (1) protein point group symmetry, (2) the growth of gas vesicles, and (3) the principles of protein folding and structure.

Humidity chamber

CryoEM transport container

Glass slider holder

Eppi pressuriser

Gravity flow column holder

Resin racks

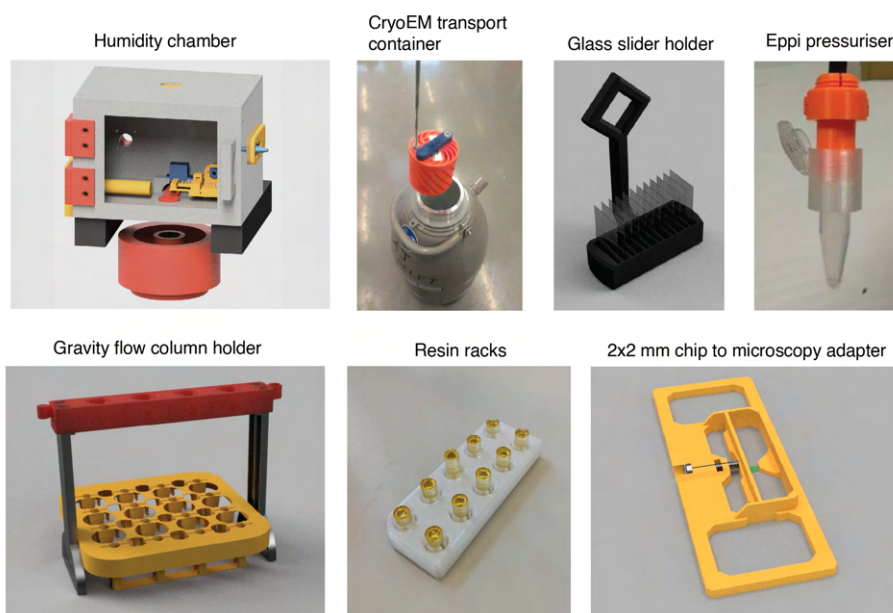2x2 mm chip to microscopy adapter

Figure 6.1: | **3D printed laboratory equipment designed and fabricated in the Jakobi group.**

6.2.1 Filament deposition modeling (FDM) 3D printing

Three-dimensional (3D) printing is an additive manufacturing technique where material is added layer by layer onto a print bed to create a final 3D object. A relatively affordable

variant of this method that gained widespead use is filament deposition modeling (FDM). In this technique, a continuous filament of thermoplastic material is melted by heat and extruded through a nozzle. The nozzle moves along a predetermined path to deposit plastic in defined positions within a 2D layer. Once a layer is completed, the nozzle moves up and deposits the following layer above, until the 3D object is finished.

A standard FDM 3D printing workflow begins with the generation of a 3D model, which can be accomplished using computer-aided design (CAD) applications or in the case of molecular structures, specialized molecular modeling programs such as ChimeraX[5]. Once the 3D structure is finalized, it is exported to a slicer software, a tool that converts the model into a printer-readable format by segmenting it into horizontal layers and generating the corresponding tool paths. These paths guide the printer's nozzle movements, ensuring the precise deposition of the material. The slicer software compiles these instructions into a file, commonly in G-code format, which is transferred to the 3D printer. The printer interprets this file and methodically executes the print job, creating the object layer by layer.

The capabilities of FDM 3D printing are subject to certain constraints, particularly concerning the geometries that can be printed. Since the process is based on layer-by-layer deposition, it is not possible to extrude plastic into empty space, which restricts the angle of overhangs to approximately 45 degrees[6]. Support structures can be introduced during slicing to create larger overhangs. However, these supports are not part of the final object and must be manually removed after printing, which can be labour-intensive. The resolution of features and the minimum wall thickness that can be printed are constrained by the nozzle diameter, which typically ranges from 0.25 to 0.8 millimeters. Adequate bed adhesion is essential for print stability, thus the object's base must be sufficiently large to prevent detachment during printing. An awareness of these constraints is critical during the design phase to ensure successful outcomes.

6.2.2 Design of printable protein models

Two types of cryo-EM data of proteins can be used as a starting point: (1) atomic models, available in pdb or mmCIF formats, and (2) cryo-EM density maps, typically in mrc format. An atomic model comprises a list of atomic coordinates which can be represented using molecular visualization software in different styles, for exampled simplified as a ribbon diagram[7]. Conversely, a cryo-EM density map represents the Coulomb potential of the molecule as a 3D array, which can be converted into an isodensity surface, representing the protein outline.

ChimeraX[5] is a popular molecular visualization software that was used throughout this thesis to generate figures of protein structures. It supports a number of display styles for atomic models [Figure 6.2], such as displaying all atoms, simplified cartoon representations or molecular surfaces. Anything that can be displayed in ChimeraX can be directly exported as a 3D model for use in 3D printing. This route was used for the example models of apoferritin and gas vesicles, as detailed later.

OpenSCAD is a free open-source software using a script-based approach to creating 3D models, allowing for the construction of shapes through parametric design. It enables users to modify dimensions and characteristics of the model by altering the input parameters within the script, providing a high level of control and repeatability. This feature is
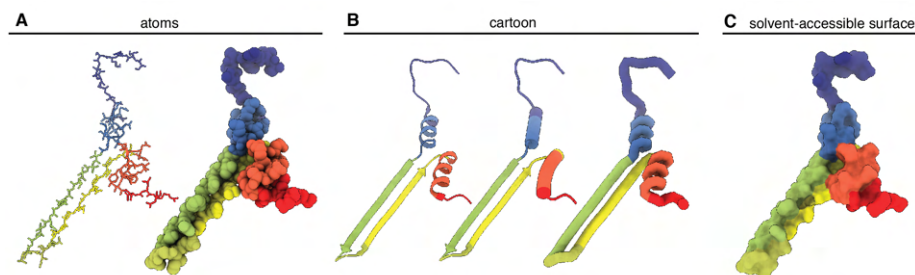
Figure 6.2: | **Display styles of atomic models in ChimeraX.** Atomic models can be represented with simplified styles e.g. in the software ChimeraX[5]. The chosen example is GvpA from this thesis [pdb:7r1c[8]]. (A) Atoms can be displayed as the endpoints of thin sticks, or as spheres with a radius equivalent to their van der Waals radius. (B) Simplified cartoons using a ribbon diagram[7], cylinders, or a chain tracing the protein backbone. (C) Solvent-accessible surface of the protein.

particularly useful in scenarios where models need to be systematically adjusted or iterated upon. This software was used for the generalized atomic model from arbitrary pdb files, showcased later in this chapter.

Blender is a free open-sorce software for general purpose 3D modeling with a wide scope including animated movies, 3D rendering, video editing and 3D printing. Blender includes parametric design tools called geometry nodes that can generate arbitraty shapes and patterns, which can then be exported as STL files for various applications. In this chapter, Blender was employed to design the GvpC helix attached to the gas vesicle model.

Fusion360 is a commercial, easy to use CAD software to create 3D models from 2D sketches. There are free alternatives available, such as FreeCAD. Fusion360 was utilized for the design of most of the laboratory equipment in Figure 6.1.

To be imported into a slicer software, data from all these softwares must be transformed into a suitable 3D model file, generally in stl or obj formats. These formats describe the surface geometry of the model using a mesh of triangles, providing an approximation of the object's shape suitable for slicing software and subsequent 3D printing. All models shown in this chapter were sliced using the free and open-source software PrusaSlicer 2.4 to 2.6 and printed on either a Prusa Mk3S or a Prusa Mk4 3D printer.

6.3 EXAMPLE MODELS

6.3.1 MODEL 1: ILLUSTRATING POINT GROUP SYMMETRY WITH FERRITIN

Proteins often assemble into symmetric oligomeric complexes consisting of two or more subunits[9]. These oligomers can exhibit cyclic symmetry (C_n), wherein n subunits are organized around a central symmetry axis. Dihedral symmetry (D_n) introduces an additional perpendicular 2-fold symmetry axis to cyclic symmetry. The remaining symmetry categories correspond to arrangements as Platonic solids. In tetrahedral symmetry (T), the assembly mimics a tetrahedron and features four 2-fold and six 3-fold symmetry axes intersecting at a central origin (Figure 6.3A). Icosahedral symmetry, frequently observed
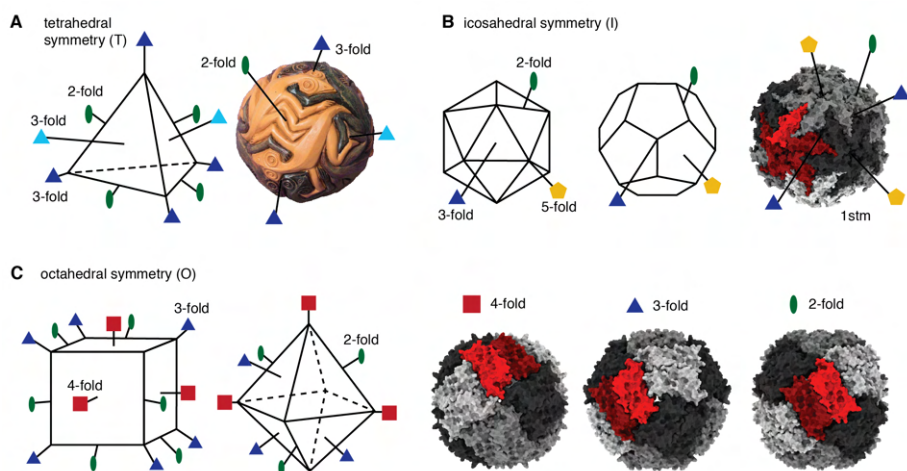
Figure 6.3: | **Point group symmetries represented by the five Platonic solids.** (A) Depiction of tetrahedral symmetry using M.C. Escher's Sphere with Lizards (maple wood, 1949). Two different shades of blue indicate the two inequivalent 3-fold axes (photograph used with permission of the 'Escher in het Paleis' museum) (B) Icosahedral symmetry represented by an icosahedron and a dodecahedron. The protein shown is the satellite panicum mosaic virus (pdb:1stm)[10], representing one of the smallest icosahedral viruses in the Protein Data Bank. (C) Octahedral symmetry represented by a cube and an octahedron. The protein example is apoferritin (pdb:7ohf)[11], shown along its 4-fold, 3-fold, and 2-fold axes. Two monomers in a dimer are highlighted in shades of red.

6

in viral capsids, comprises twelve 5-fold, twenty 2-fold, and thirty 3-fold symmetry axes (Figure 6.3B). This assembly can be represented by either an icosahedron or a dodecahedron. Lastly, octahedral (O) symmetry presents six 4-fold, eight 3-fold, and twelve 2-fold axes, and can be represented by an octahedron or a cube (Figure 6.3C). The protein ferritin adopts this symmetry, comprising 24 identical subunits of ~20 kDa.

Ferritin is ubiquitous protein for iron storage throughout the tree of life, characterized by its hollow core structure capable of encapsulating iron[12]. Its iron-free 'apo' form is commonly employed in cryo-electron microscopy (cryo-EM) as a validation sample, owing to its large size, high stability, high symmetry, and straightforward sample preparation. In the context of cryo-EM, atomic-resolution structures of up to 1.2 Å have been obtained using apoferritin as a sample[13,14]. This protein is also utilized in our own research outlined in Chapter 2, where we employ it to test a new nanofluidic cavity-based sample preparation technique.

Here, we design an educational model of ferritin comprised of 24 identical 3D-printed subunits. The shape of apoferritin monomers allows 24 such units to self-assemble into a densely packed spherical shell. This model can serve as a pedagogical tool to elucidate the geometric principles required for a protein to form a tightly packed shell with a hollow core. Moreover, it offers an intuitive understanding of the three distinct symmetry axes inherent to octahedral symmetry.

We initiated our model construction by using the cryo-EM density described in Chapter
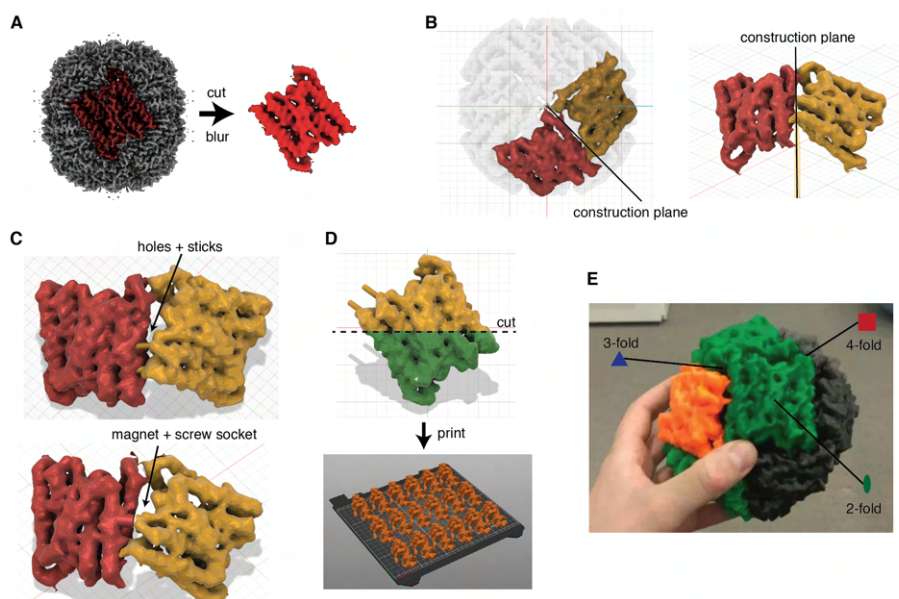
Figure 6.4: | **Stepwise design and assembly of a 3D printable ferritin model.** (A) Cryo-EM density map of *P. furiosus* apoferritin from EMDB entry 12901[11]. The density was subsequently blurred, and a dimeric subunit (shown in red) was isolated around the pdb model for further processing. (B) Imported dimer density in Fusion360 software, where a plane was constructed between two dimers to facilitate the design of attachment mechanisms. Parts of the mechanism were extruded from this plane as a base. (C) Two different attachment mechanisms, namely hole+stick and magnet+screw, were developed for dimer assembly. (D) A single dimer was bisected along its 2-fold symmetry axis and 3D-printed 24 times to create the full assembly. (E) The completed hole+stick model, with symmetry axes annotated.

2 (deposited in the EMDB as entry 12901) and imported it into ChimeraX software[5]. The resolution of 3.0 Å was too fine-grained for practical 3D printing. Therefore, we applied a Gaussian blur to the density and then carved out a dimer based on the corresponding atomic model (PDB: 7OHF) as shown in Figure 6.4A.

```
volume gaussian #1 sdev 1.8
volume zone #1 nearAtoms #2 range 2.8
```

The modified density retained key features such as secondary structural elements and alpha-helical pitch, but the elements were slightly blended together. This facilitated 3D printing of the model as a single block without the need for additional support material. We exported the model as an STL file and imported it into Fusion 360 for further processing. Alignment in space was performed using a 2D image of the density along the 4-fold axis. A construction plane was set precisely between two duplicated dimers, angled at 45 degrees to two of the 4-fold axes. This plane served as a base to sketch and extrude the interlocking mechanisms between dimers (Figure 6.4B).

Two assembly mechanisms were designed: (1) a friction-based mechanism consisting of matching holes and sticks, and (2) a magnet mechanism (Figure 6.4C). Since the resulting dimer exhibits C2 symmetry around its central axis, it was bisected to create a flat surface suitable for 3D printing (Figure 6.4D). The model was bisected in such a way that the alpha helices in the model all point in the z-direction (upwards from the print bed). In this way, there are no steep overhangs in the model and it can be printed without support material. The final model was printed 24 times, and the dimer halves were glued at their flat surfaces to form 12 dimers. These dimers were then assembled into an octahedral ferritin shell using the designed assembly mechanisms (Figure 6.4E). The resulting model illustrates how biological systems can build large structures by repetition and self-assembly, and gives an intuitive way to understand point-group symmetry with a physical educational model.

6.3.2 Model 2: Mechanism of gas vesicle growth

Gas vesices are gas-filled protein shells that enable buoyancy-based motility in microbes. In Chapter 4 we used cryo-EM to solve the atomic structure of the gas vesicle wall. Combining this information with high-resolution 2D views of the gas vesicle tips and the gas vesicle seam, we built a pseudo-atomic model of an entire gas vesicle that illustrates the overall assembly of this organelle. This second 3D print showcase outlines the design of an educational model of an entire gas vesicle, which highlights that (1) gas vesicles are made of two identical halves assembled around a 180 degree symmetry axis, (2) our proposed mechanism of how a new monomer can be inserted into the vesicle by rotation of both halves against each other, involving ratcheting of molecular contacts at the seam, and (3) the attachment of the reinforcement protein GvpC.

The starting point for this model is a pseudo-atomic model of a gas vesicle half [Figure 6.5A] (available on Zenodo: https://zenodo.org/records/6458345). A set of commands in ChimeraX[5] generate a simplified display style of the protein chains suitable for 3D printing:

```
car style protein  modeh default arrows f xsect oval width 5 thick 5 divisions 2
  barSides 4;
car style :25-34   modeh default arrows f xsect oval width 4 thick 4 divisions 2
  barSides 4;
car style :2-13    modeh default arrows f xsect oval width 8 thick 8 divisions 2
  barSides 4;
```
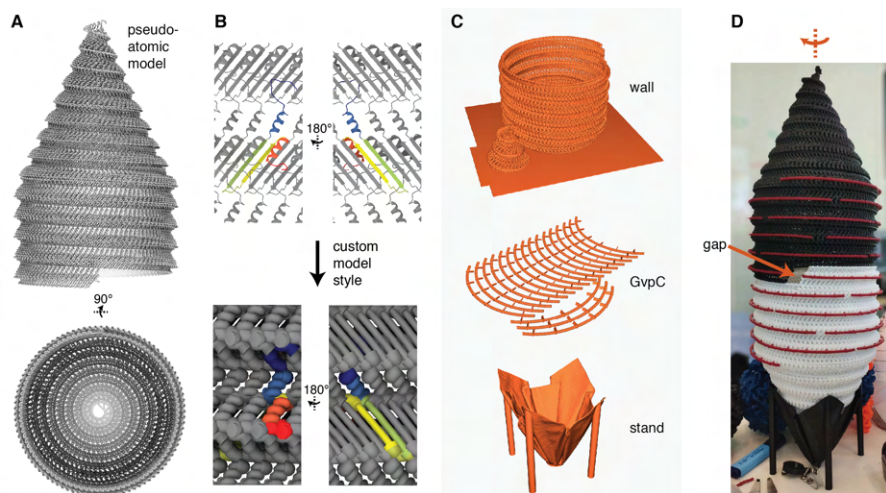
Figure 6.5: | **3D printable model illustrating the mechanism of gas vesicle growth.** (A) Pseudo-atomic model of a gas vesicle half, including conical tips and seams, as discussed in Chapter 4. (B) A custom model style is employed to maximize the visibility of alpha-helices and beta-strands while creating a connected, 3D printable model. (C) Separate prints of the gas vesicle wall, GvpC molecules, and a supporting stand. (D) The assembled model can demonstrate how rotation of the upper gas vesicle half creates a gap at the polarity reversal point (PRP), where the addition of new monomers is hypothesized.

6

The first command transforms the protein chains into 5 Ångstrom thick tubes depicting a smoothed trajectory of the main chains. The second command modifies the beta strands in the model to a thickness of only 4 Ångstrom to seperate the strands for clearer depiction. The third command increases the thickness of the N-terminus to 8 Ångstrom to ensure that the helical GvpA ribs are physically connected in the printed model [Figure 6.5B].

As discussed in Chapter 4, there is a clash of α-helix 2 of GvpA at the polarity reversal point (PRP) where GvpA ribs from both GV halves meet. To avoid that this clash interferes with the model assembly, this helix was deleted. The resulting model was exported in the stl format for 3D printing.

The printed model should not only highlight the gas vesicle wall made from GvpA, but also the attachment geometry of GvpC which we revealed in Chapter 4. GvpC forms an extended alpha helix of roughly 250 Å length that attaches to the outside of the gas vesicle along the GvpA ribs and binds to α-helix 2. The α-helix was modelled as a helical tube of 5.4 Å pitch in Blender 3.0 using the geometry nodes functionality. Regularly spaced hooks were inserted in the model at a distance corresponding to the 33 amino acid repeats in the protein. The hooks facilitate the attachment of GvpC to α-helix 2 of the gas vesicle shell by pushing them in the gaps between the helices. A small portion on the bottom of the helix was removed using PrusaSlicer to create a flat printing surface to aid print bed adhesion.

A stand was created in Blender 3.0, utilizing the 3D-printable model of the previously described gas vesicle half. Blender's cloth function, capable of simulating fabric physics, was employed to computationally 'drape' a cloth piece over the model. This process allowed the cloth model to deform, delineating the model's outline. Subsequently, the cloth was

rotated by 180 degrees to transform it into a stand, extruded to a thickness of one millimeter, and four cylindrical legs were integrated into the design.

Due to the Prusa Mk3S printer's size constraints, the model was printed in multiple sections. The gas vesicle half-shell was divided into three parts along its helical axes, and these were glued together post-printing. Two such half-shells were printed to form the full model.

GvpC was printed 38 times to populate the gas vesicle's cylindrical area. GvpC monomers are externally attached to the model. The monomers can be attached either regularly spaced or randomly. This variable positioning mirrors the lack of knowledge about GvpC distribution on the vesicle surface - whether it is random or involves coordinated assembly of C and N-termini to create a dense mesh.

The final assembled model consists of two identical helical gas vesicle half shells, which can be placed upright into the stand. The top half is only loosely placed on the bottom half and can be rotated against it. The β-turns of the GvpA monomers in the seam form small corrugations along the seam, effectively demonstating the ratcheting mechanism we proposed where the gas vesicle halves rotate by exactly one monomer into the next equivalent position to allow for new monomer insertion [Figure 6.5D].

6.3.3 Model 3: Protein folding and geometry - a parametric atomic protein model

The primary objective of cryo-EM single particle analysis (SPA) is to generate a three-dimensional density map of the target particle. Attaining a resolution below 4 Ångstroms allows for the creation of an atomic model, which is crucial for detailled chemical and biological analysis. An atomic model consists of three-dimensional coordinates that represent every atom. In this section, we introduce a generalized approach for the production of 3D printed atomic models of proteins.

A peptide comprises peptide bonds that link the amino group of one amino acid to the carboxyl group of the adjacent amino acid. The involved four atoms – O=C-N-H – typically form an almost planar structure that is very similar throughout the structure. The peptide chain also includes more variable elements, notably the C-α atoms of amino acids, to which the side chains are attached. The conformational flexibility within a peptide chain arises from the angles, known as Φ (phi) and Ψ (psi), formed between the C-α atom and its adjacent planar peptide bonds. Additionally, the side chains' conformations, referred to as rotamers, are also variable.

Owing to these structural properties, the 3D printed model is designed with planar backbones as immutable components, printed once for each amino acid in the model, as depicted in [Figure 6.6A]. The variable elements, namely the C-α atoms and their attached side chains, are calculated based on the data provided in the input atomic model file - making the model 'parametric'.

This parametric model can be generated for any input pdb file containing atomic coordinates of a peptide. The Φ (phi) and Ψ (psi) angles, derived from the atomic model, are integrated as small notches in each amino acid's C-α atom, as illustrated in [Figure 6.6B]. Corresponding notches are included into the immutable peptide backbone. During model assembly, aligning the notches of amino acid and backbone automatically leads to the correct fold of the peptide chain [Figure 6.6C]. Additionally, the rotamer conformation

of each amino acid, extracted from the pdb file, dictates the precise 3D positioning of atoms in the printed model.

The computation algorithm for this parametric model is incorporated into a Jupyter Notebook, titled 'ColabProteinPrint.' This can be executed either locally or within a Google Colab environment, with the link provided in the Supplementary material. Processing of the pdb file is facilitated by the 'gemmi' Python library[15], while the creation of 3D geometries is achieved using OpenSCAD.

For a practical demonstration of the 3D-printable atomic model, the consensus repeat 'YKELQETSQQFLSATAQARIAQAEKQAQELLAF' of GvpC, as discussed in Chapter 5, was selected as a sample print. GvpC is characterized by its extended alpha helix structure, which attaches to the exterior of gas vesicles via a conserved interface composed of leucine, phenylalanine, and arginine residues. A pdb file of this consensus repeat was generated in ChimeraX [5], employing the 'build model' function with ideal α-helical Φ and Ψ angles (φ = -57° and ψ = -47°). This pdb file was then input into the ColabProteinPrint algorithm, yielding stl files for the backbone and each individual amino acid.

The backbones and side chains of the model were printed separately on a Prusa Mk3S 3D printer, using distinct colors for differentiation. The residue number is printed onto the C-α atom of each amino acid to keep track of the correct residue order. The parts were assembled and correctly folded by aligning the notches that encode the Φ and Ψ angles, as depicted in [Figure 6.6D]. The final model of the GvpC consensus repeat distinctly displays the clustering of leucine, arginine, and phenylalanine residues on one side of the α-helix, which we hypothesised as the binding interface of GvpC's interaction with gas vesicles (Chapter 5).

Furthermore, this type of physical model can serve as an educational tool. By building and physically manipulating the model, learners can gain an appreciation of the geometry of secondary structure features of proteins and their overall assembly. This hands-on approach enhances learning, making complex molecular biology concepts more accessible and engaging. Overall, such physical models represent a bridge between theoretical knowledge and practical understanding, making them a useful tool in molecular biology education.
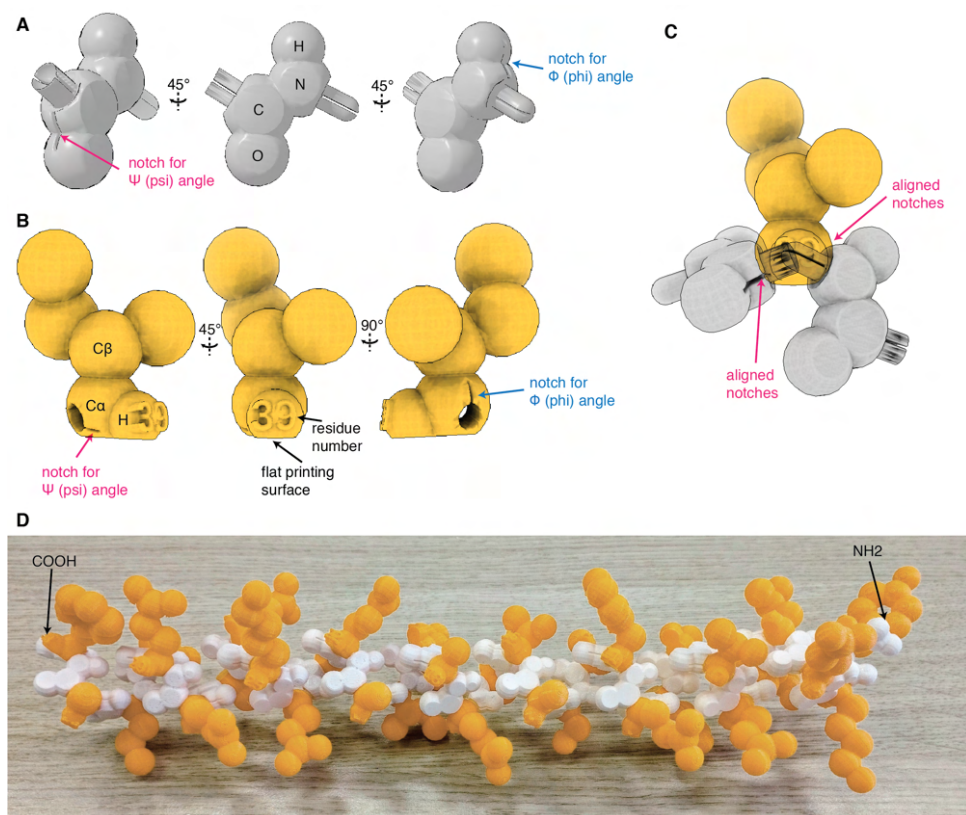
Figure 6.6: | **3D printable atomic model of a protein..** (A) A planar peptide backbone (O=C-N-H) with bonds (sticks) of differing diameters for C-C and N-H bonds to adjacent C-α atoms. Notches next to the bonds indicate phi (Φ) and psi (Ψ) angles for adjacent amino acid rotation. Flat surfaces are included to emphasize bond planarity and simplify printing. (B) The amino acid isoleucine, featuring two holes in the C-α atom for backbone attachment. Notches depict phi (Φ) and psi (Ψ) angles from the pdb file. A flat surface on the C-α atom aids in printing. (C) Proper protein folding is achieved by aligning matching notches in backbone and amino acid parts during assembly. (D) The completed model displays a completely α-helical 33 amino acid repeat sequence of GvpC (Chapter 5). Amino and carboxy groups are added as caps on both sides of the model.

6.4 Model and code availability

All 3D models described in this chapter are freely available for download and reprinting (under the CC BY-NC-SA 4.0 licence): https://www.printables.com/@StefanHuber_113966/models

Code for parametric 3D-printable protein model generation is available as a Jupyter notebook, and can be either executed locally or in the Google Colab environment online: http://github.com/stefanhuber1993/ColabProteinPrint

6

6.5 References

1. Saggiomo, V. A 3D Printer in the Lab: Not Only a Toy. en. *Advanced Science* **9**, 2202610. ISSN: 2198-3844, 2198-3844 (Sept. 2022).

2. Goodsell, D. S., Franzen, M. A. & Herman, T. From Atoms to Cells: Using Mesoscale Landscapes to Construct Visual Narratives. en. *Journal of Molecular Biology* **430**, 3954–3968. ISSN: 00222836 (Oct. 2018).

3. Iwasa, J. H. Using animation to mediate scientific discourse. en. *Nature Microbiology* **7**, 3–3. ISSN: 2058-5276 (Dec. 2021).

4. Lyall, K., Iwasa, J. H., Goodsell, D. S. & Holt, L. Communicating science through visual means. en. *Trends in Biochemical Sciences* **48**, 2–4. ISSN: 09680004 (Jan. 2023).

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

6. Redwood, B., Schffer, F. & Garret, B. *The 3D printing handbook: technologies, design and applications* (3D Hubs, 2017).

7. Richardson, J. S. Early ribbon drawings of proteins. *nature structural biology* **7**, 624–625 (2000).

8. Huber, S. T., Terwiel, D., Evers, W. H., Maresca, D. & Jakobi, A. J. Cryo-EM structure of gas vesicles for buoyancy-controlled motility. *Cell* **186**, 975–986 (2023).

9. Goodsell, D. S. & Olson, A. J. Structural symmetry and protein function. *Annual review of biophysics and biomolecular structure* **29**, 105–153 (2000).

10. Ban, N. & McPherson, A. The structure of satellite panicum mosaic virus at 1.9 Å resolution. *Nature structural biology* **2**, 882–890 (1995).

11. Huber, S. T., Sarajlic, E., Huijink, R., Weis, F., Evers, W. H. & Jakobi, A. J. Nanofluidic chips for cryo-EM structure determination from picoliter sample volumes. *Elife* **11**, e72629 (2022).

12. Arosio, P., Elia, L. & Poli, M. Ferritin, cellular iron storage and regulation. *IUBMB life* **69**, 414–422 (2017).

13. Yip, K. M., Fischer, N., Paknia, E., Chari, A. & Stark, H. Atomic-resolution protein structure determination by cryo-EM. *Nature* **587**, 157–161 (2020).

14. Nakane, T., Kotecha, A., Sente, A., McMullan, G., Masiulis, S., Brown, P. M., Grigoras, I. T., Malinauskaite, L., Malinauskas, T., Miehling, J., *et al.* Single-particle cryo-EM at atomic resolution. *Nature* **587**, 152–156 (2020).

15. Wojdyr, M. GEMMI: A library for structural biology. *Journal of Open Source Software* **7**, 4200 (2022).

7

Conclusions and Outlook

7.1 Future directions of nanofluidics-based cryo-EM sample preparation

In Chapter 2, we proposed a new cryo-EM sample preparation method called a "cryoChip" where picoliter amounts of protein sample are drawn into nanofluidic channels. Sample thickness is usually difficult to control during sample preparation. This new approach automatically determines the thickness of the sample layer by the channel geometry. We showed that three test samples, namely apoferritin, T20S proteasome and tobacco mosaic virus (TMV) can be drawn successfully into the channels, imaged and their structure reconstructed to resolutions of up to 3 Å. This is sufficient resolution to build *de novo* atomic models, but is less than what could be expected from perfect conventional sample preparation techniques involving blotting paper. Just as continuous developments have pushed the limit of achievable contrast and residual specimen movement for holey support films[2], we expect that similar progress can be attained for nanofluidic supports by further improvements in electron transparency, as well as electrical and thermal conductance of the enclosing materials to further reduce beam-induced motion. We anticipate that ongoing developments will eventually allow approaching resolutions realised for conventional samples.

We predict that the cryoChip is only the beginning of a range of future cryo-EM supports building on MEMS technology. Several immediate improvements can be achieved with minor modifications to our current design. Currently the observation membrane is 50 µm x 50 µm, representing less than 0.1% of the entire chip area. The imaging area can be readily extended by additional observation membranes, thus allowing fully automated collection of thousands of images with uniform ice thickness from a single chip. Parallel arrangements of observation membranes with separate filling channels will allow screening different samples on the same chip. Beyond this, the versatility of nanofabrication technology opens up possibilities to facilitate a range of new cryo-EM applications by bridging lab-on-a-chip approaches with structural biology. These include possibilities for on-chip mixing, parallelised high-throughput screening for structure-based drug discovery, time-resolved experiments and on-chip purification of complex protein mixtures.

We tested our cryoChip samples with three different protein samples: apoferritin, T20S proteasome and TMV. To get further insights how different proteins behave with our method, broader testing is neccessary. This should include proteins known to disintegrate at the air-water interface, as well as proteins with strong preferred orientation, and will help determine the limitations of nanofluidic channels as a cryo-EM sample preparation technique.

7.2 Future directions of structural studies on gas vesicles

Gas vesicles are intracellular, gas-filled nanostructures that allow microbes to float vertically to the surface of their aqueous habitat. In Chapter 4, we determined for the first time their atomic structure at 3.2 Å resolution[3]. This facilitated precise *de novo* atomic model building, unveiling the fold of the coil-α-β-β-α-coil secondary structure of GvpA, amino

acid side chain positions and the packing of the GvpA monomer within the gas vesicle rib. Our findings provide a comprehensive view of the gas vesicle's molecular architecture, explaining its physical properties such as the extremely hydrophobic interior, corrugations leading to structural stability, and the selective permeability through its 3.8 Å gas pores[3].

7.2.1 Structural unknowns: Growth mechanism and nucleation

Combining information of our atomic structure of GvpA and 2D projection images of the gas vesicle seam and tips, we constructed a complete model of the entire gas vesicle in Chapter 4. The model consists of two identical halves in reverse orientation, and contains unique molecular interactions of GvpA at the seam [Figure 7.1A], finally confirming theoretical considerations initially postulated by Walsby more than four decades earlier[4]. The seam is formed by hydrophobic interactions between residues V35 and I37 of the contacting GvpA β-turns from both gas vesicle halves [Figure 7.1A]. A unique and critical feature along the circumference of the seam is the 'polarity reversal point' (PRP), where the helical ribs of both gas vesicle halves contact side-by-side and the orientation of GvpA reverses.

Early gas vesicles grow as small biconical structures, presumably transitioning to cylindrical growth upon reaching a critical diameter. A hypothetical minimal biconical nucleus might consist of a single GvpA rib in each cone segment [Figure 7.1B]. While cryo-EM has captured images of small bicones[3,5], the actual existence, formation process, and structural details of such a minimal biconical nucleus remain speculative. Furthermore, it is unclear whether the bicone represents the initial nucleating structure of the gas vesicle, or if it results from the combination of two smaller nuclei in opposing orientations. Isolating and purifying these nascent stages would be crucial for a deeper understanding of gas vesicle nucleation, allowing for detailed structural and compositional analyses of these early stages.

At a critical diameter, gas vesicle bicones continue to grow with constant diameter thus marking the transition to cylindrical growth. This critical diameter is not tightly controlled and gas vesicles exhibit a range of cylinder diameters[6,7]. The mechanism underlying the cone-to-cylinder transition is not well understood. Based on structural considerations, an energetic balance between preferred curvature of GvpA ribs and binding energy gained by rib-to-rib contacts was proposed[3]. The GvpA N-terminus forms well-defined contacts with the β-turn of the adjacent rib in the crystalline, cylindrical wall segments, as compared to more disorganised contacts in the cone. Alterations in amino acids in the β-turn or N-terminus could thus influence gas vesicle morphology by modifying this energetic balance. Gas vesicles in *H.salinarum* with mutations around the beta turn (e.g. I34M, corresponding to I37 in *B.megaterium* GvpA) provide some evidence for this[8], as they exhibit different cone-to-cylinder transition behavior.

We have formulated a hypothesis for the cylindrical growth of gas vesicles[3], based on the assumption that there is a 1:1 binding of GvpA monomers from both sides of the seam. Central to this model is the creation of a gap at the polarity reversal point (PRP) for the integration of an additional GvpA monomer. The creation of this gap is facilitated by the rotational movement of one half of the gas vesicle relative to the other. Such movement necessitates a ratcheting action involving the disruption and reformation of hydrophobic contacts between V35 and I37 at the seam [Figure 7.1C]. A new GvpA monomer can then

be incorporated into the gap, reforming the fully closed seam. Critically, the total length of the seam remains unchanged in this model.

This seam ratcheting mechanism can explain how two contacting helical cylinders can grow, but is not compatible with the growth of a bicone. In a growing bicone, the size of their contacting bases increases, necessitating the seam to enlarge. We therefore propose a different mechanism of growth for the bicone stage. It involves sliding of the ribs against each other within each cone, while the the seam remains connected through its hydrophobic contacts. By this mechanism, the seam can enlarge by one monomer, while creating a similar opening at the PRP for monomer insertion [Figure 7.1C]. Such sliding between the gas vesicle ribs may be possible in the cone due to the lack of crystalline relationship between the ribs as they would occur in the crystalline cylindrical segment.

The problem with this model of bicone growth is that the seam grows by an entire GvpA monomer at each insertion step. In the limit, if 100 monomers were added in an original cone of only 20 monomers, the resulting cone would have more than 100 monomers at the seam of the 120 monomers in total, which is geometrically impossible. The real growth mechanism of the bicone is thus likely more complex and possibly involves a mix of rib sliding and seam ratcheting.

For both sliding and ratcheting mechanisms, a gap at the PRP allows for vesicle growth by insertion of new monomers. The PRP is formed by contact of two gas vesicle halves in reverse orientation and exhibits 180 degree rotational symmetry relative to the long axis of the gas vesicle. This makes it geometrically equivalent for a GvpA monomer to be inserted into either top or bottom gas vesicle halves [Figure 7.1D], which would randomly increase the size of either bicone. Indeed the seam is only approximately observed in the gas vesicle center[3,9].

Unlike our idealised model, it has been observed that both gas vesicle halves can have unequal diameter[9,10]. This may be explained by the cone-to-cylinder transition not being tightly coupled and occuring independently in both halves [Figure 7.1E]. The exact mechanism of gas vesicle growth therefore remains an highly intriguing problem in structural biology. Resolving the seam's structure with cryo-EM could yield deeper insights into the growth mechanics. Further understanding might also come from biophysical techniques that observe gas vesicle growth kinetics in reconstituted systems.

7.2.2 Structure and role of other gas vesicle gene products

Gas vesicles almost exclusively consist of the structural proteins GvpA and GvpC, yet their synthesis requires an extensive cluster of around ten genes [Figure 7.2A]. The functional roles of the genes within the clusters are not well understood. Adopting a structural biology view, we present AlphaFold2 predictions for gene clusters from *B. megaterium*, *A. flos-aquae*, and *H. salinarum*. These are annotated with recent biochemical research insights, offering potential functional interpretations [Figure 7.2B]. This presentation aims to foster a mechanistic understanding of gas vesicle biogenesis and highlights aspects that warrant further investigation.

All gene clusters feature homologs of the wall protein GvpA, such as GvpJ, GvpM, or GvpS, which are not present in the cylindrical part of the gas vesicle wall[3]. Recent research suggests that these homologs contribute to the formation of the gas vesicle's conical segments since they were found in purified bicones[5]. Considering our earlier discussion
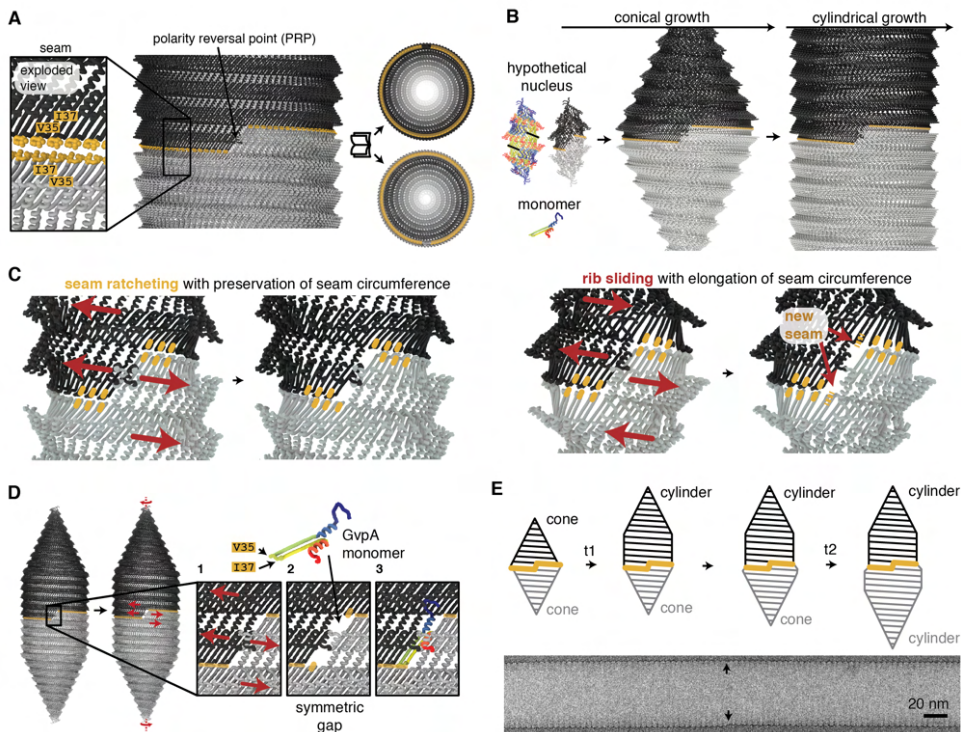
Figure 7.1: | **Model of gas vesicle biogenesis.** (A) Growth stages of gas vesicles. (B) Molecular details of the gas vesicle seam formed by hydrophobic residues V35 and I37 of GvpA hairpins of both gas vesicle halves pointing towards the seam. (C) Different growth mechanisms for bicones (where the seam needs to increase in circumference) and cylindrical gas vesicle segments (where seam needs to preserve its circumference). (D) Monomer insertionat the PRP in the cylindrical section of the gas vesicle by ratcheting motion of gas vesicle halves. (E) More complicated growth model with separate cone-to-cylinder transitions for both halves (t1 and t2) which may lead to unequal diameters of cylindrical segments in both gas vesicles halves . Cryo-EM image of *B.megaterium* gas vesicle with unequal cylinder diameters[3].

on distinct growth mechanisms for cylindrical and bicone structures, it seems possible that these GvpA homologs are integral to the unique growth processes hypothesized in bicones. Notably, the *A.flos-aquae* gene cluster differs from the others, containing only a single GvpA homolog, GvpJ, which is further distinguished by an atypical C-terminus comprising six repeats of 21 amino acids, the function of which remains to be elucidated.

GvpC, a long alpha-helical protein characterized by repetitive sequences, binds to gas vesicle ribs for stabilization[3,9,11]. In cyanobacteria, GvpC contains uniform 33 amino acid repeats that span four GvpA monomers along a rib. In contrast, haloarchaeal GvpC exhibits 32-38 amino acid repeats and includes a distinct C-terminal domain, linked by an extended linker [Figure 7.2B]. The binding affinity of GvpC is associated with a specific motif comprising arginine, phenylalanine, and leucine[3]; however, the precise atomic structure of this binding interface remains undetermined. Additionally, the pattern of GvpC's attachment to gas vesicles — whether random or regularly spaced — is not clear. Indications of GvpC self-interaction from split-GFP assays[12] hint at a more regular mode of binding.

GvpK, consistently found across gene clusters, is implicated in the initiation of gas vesicle formation, as evidenced by several studies[13–15]. Its pronounced hydrophobicity, suggested by AF2 predictions, supports a role in nucleation. Notably, within the *A. flos-aquae* gene cluster, GvpK exhibits an unusual feature: it includes a GvpA domain, suggesting it may be a fusion protein combining aspects of GvpJMS with GvpK. This attribute renders it as an interesting target for studies of GV nucleation. Understanding the structure of a minimal nucleus could greatly advance our knowledge of the nucleation process and the composition of the nucleus.

GvpN, an AAA+ ATPase, is an essential component in all gene clusters. Research indicates that the absence of GvpN arrests gas vesicle formation at the bicone stage[5,16]. Bicones can initially form without GvpN, they fail to transition from cone to cylinder or to elongate as cylinders. Given our earlier discussion on distinct growth mechanisms for cylindrical and bicone structures, GvpN may be needed to provide the energy for seam ratcheting in the cylindrical growth phase. Intriguingly, GvpN features an atypically long 30 amino acid β-hairpin within its H2 insert[17] [Figure 7.2B]. This feature bears resemblance to the extended GvpA β-hairpin, which may inform speculation at a potential functional interplay. Unraveling the role of GvpN and the underlying mechanism of its action remains a compelling challenge for structural studies on gas vesicles.

Not surprising given its highly hydrophobic nature, GvpA has been shown to be associated with GvpF or GvpL, acting as a chaperone[13–15]. The interaction site between GvpF and GvpA has been identified, involving GvpA's α-1 helix[14]. This finding suggests a chaperoning role during the protein synthesis process. Resolving the structure of GvpA in its chaperoned state would be crucial for a comprehensive understanding of this interaction, as well as to discern the structural distinctions of GvpA when bound to a chaperone versus its integrated form in the gas vesicle wall. Facilitating the solubilization of this hydrophobic protein within the cell and safely directing its integration into the gas vesicle represents a compelling subject for investigation.

A recent study has put forth suggested roles for the proteins GvpU and GvpT: they are implicated in inducing the dense packing of gas vesicles inside cells, facilitated by their interaction with the elongated C-terminus of GvpA in *B. megaterium*[18]. Tight packing has been observed in cyanobacteria through electron microscopy (EM)[19]. Intriguingly, both

the GvpU and GvpT genes are exclusive to the phylum Bacillota (previously known as Firmicutes) and lack homologous counterparts in other phyla, suggesting that gas vesicle clustering may have evolved independently in phylogenetically distant organisms.

The recent high-resolution structures of gas vesicles only mark the very beginning of the quest to understand the assembly mechanism of these unique protein-based compartments, leaving an abundance of unanswered questions. Structural studies targeting gas vesicle nuclei, tips and seams, as well as the complex molecular machinery facilitating gas vesicle biogenesis will be crucial to advance our understanding of this motility machinery.
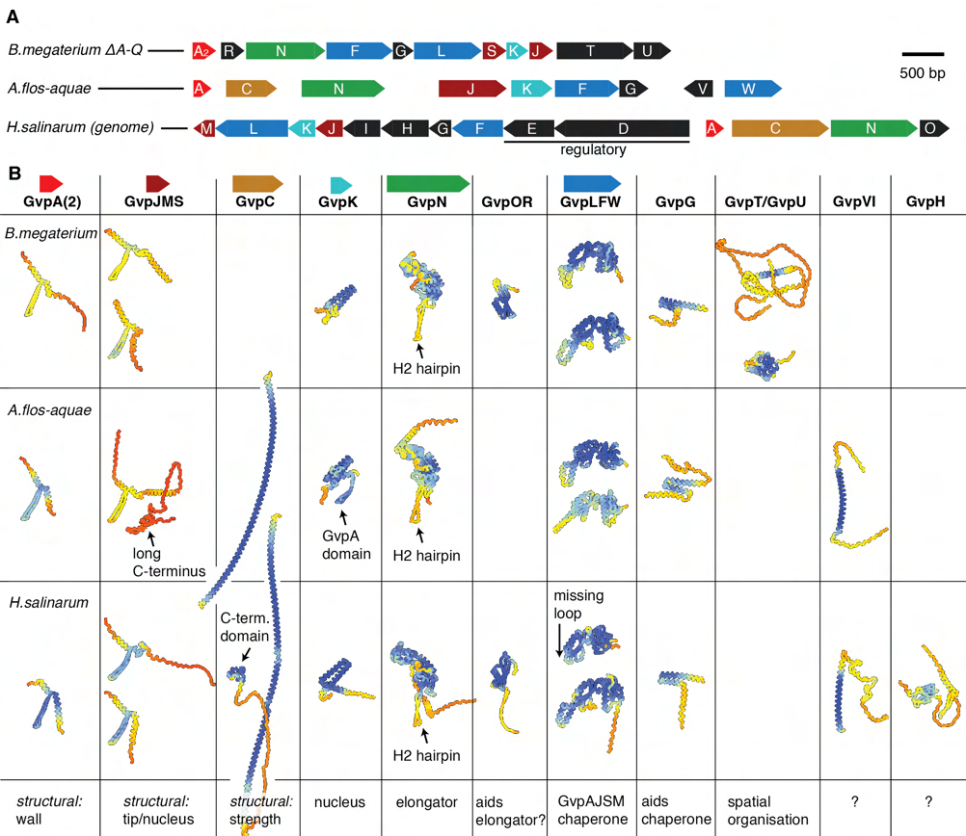
7

Figure 7.2: | **Structure predictions of gas vesicle gene clusters from three organism.** (A) Gas vesicle gene clusters from three different organisms (B) AlphaFold2[20] structure predictions for gene clusters from *B.megaterium*, *A.flos-aquae* and *H.salinarum*. All proteins are depicted to scale. GvpD and GvpE from *H.salinarum* were omitted due to their role in gene regulation[21]. The color scheme highlights prediction confidence (blue: good, yellow: bad). Predictions for GvpAJMS were performed on a homo-5-mer and the middle monomer is depicted. Arrows depict noteworthy features. Annotation sources: GvpA[3,9,10], GvpG[15], GvpLFW[13,14], GvpOR[15], GvpK[13−15], GvpN[5,16], GvpJMS[5], GvpC[11].

7.3 Structural biology and artificial intelligence

In the period of this thesis spanning from 2019 to 2023, significant advancements were made in the field of protein structure prediction with the release of the AlphaFold2 (AF2) algorithm[20]. While in the decades prior, improvements in prediction accuracies were only incremental, AF2 significantly outperformed other contributions at the "14th Critical Assessment of Protein Structure Prediction" (CASP14) competition. The protein structure predictions were close to the experimentally determined structures - an achievement that marked a breakthrough in the field.

The availability of AF2 has already influenced several parts of this thesis. Notably, in Chapter 7 (section "Structure and role of other gas vesicle gene products"), we utilized AF2 to predict the structures of various genes within the gas vesicle gene cluster. This analysis revealed intriguing characteristics in some proteins that were not previously evident. While it is important to note that these predictions are not substitutes for experimental structure determination and do not currently account for complex protein environments, they serve as invaluable tools for hypothesis generation, guiding future research directions. Such findings underscore the transformative impact of AF2 in the field of structural biology, opening up new possibilities for exploration and discovery.

In Chapter 4, AF2 was employed to predict the structures of GvpA homologs from various organisms. Prior to AF2's release, attempts at structure prediction of GvpA using methods like EVcouplings[22] failed to produce plausible results. Earlier computational modeling efforts for GvpA, as reported a decade ago[23,24], were also challenging and did not yield predictions with strong explanatory power. However, AF2 marked a turning point, successfully predicting a plausible structure for short ribs of oligomerised GvpA (such as 5-mers). The results showed remarkable consistency in the structures for *B. megaterium* and *A. flos-aquae*, while *H. salinarum* was predicted to have a slightly longer C-terminal α-helix. Notably, the predictions accurately identified a hydrophobic face in the assembly, although they fell short in detailing features like the N-terminal coil, rib-to-rib contacts or gas pores. This highlights that AF2 is not a one-to-one replacement for experimental structural biology. A careful view is that AF2 generates "exceptionally useful hypotheses"[25] that can be combined or validated with real data.

The success of AF2 to predict features even of an unusual protein such as GvpA demonstrates the profound impact it can have on structural biology, particularly in hypothesizing and guiding research where high-resolution experimental data are lacking. A recent successful combination of AF2 with experimental data is the analysis of a 7.9 Å molecular envelope of an archaeal cell sheath, determined through cryo-electron tomography (cryo-ET)[26]. The AF2 prediction of the SH protein (sheath protein) could be successfully docked into the density and used to interpret the low-resolution data by constructing an integrative model. New insights from this modelling procedure were for example the positioning of cysteins at distances that allow disulfide bonds, explaining previous data on the stability of the assembly.

The success of machine learning in resolving the long-standing challenge of protein structure prediction, a task that eluded solutions through conventional human-engineered methods, suggests prospects for tackling other previously intractable problems in the near future. One powerful potential application lies in the prediction of small molecule binders to proteins for pharmaceutical applications. Accurately forecasting how small

molecules interact with proteins could vastly accelerate drug discovery. Another promising avenue is the design of novel molecular machines, such as tailored enzymes. The ability to design complex enzymatic structures and functions could lead to the creation of enzymes with enhanced or entirely new capabilities, for instance offering solutions to ecological challenges such as plastic degradation.

The emerging ability of black box algorithms to tackle the complex sequence-to-structure relationship of proteins thus opens the door to exciting new capabilities in biotechnology, the outcomes of which remain largely unpredictable.

7

7.4 References

1. Huber, S. T., Sarajlic, E., Huijink, R., Weis, F., Evers, W. H. & Jakobi, A. J. Nanofluidic chips for cryo-EM structure determination from picoliter sample volumes. *Elife* **11,** e72629 (2022).

2. Naydenova, K., Jia, P. & Russo, C. J. Cryo-EM with sub–1 Å specimen movement. *Science* **370,** 223–226 (2020).

3. Huber, S. T., Terwiel, D., Evers, W. H., Maresca, D. & Jakobi, A. J. Cryo-EM structure of gas vesicles for buoyancy-controlled motility. *Cell* **186,** 975–986 (2023).

4. Walsby, A. E. Structure and function of gas vacuoles. *Bacteriological reviews* **36,** 1–32. ISSN: 00053678 (1972).

5. Ling, B., Gungoren, B., Yao, Y., Dutka, P., Smith, C. A. B., Lee, J., Swift, M. B. & Shapiro, M. G. *Truly tiny acoustic biomolecules for ultrasound imaging and therapy* en. preprint (Synthetic Biology, June 2023).

6. Hayes, P. K. & Walsby, A. E. The inverse correlation between width and strength of gas vesicles in cyanobacteria. **1617** (2007).

7. Dutka, P., Malounda, D., Metskas, L. A., Chen, S., Hurt, R. C., Lu, G. J., Jensen, G. J. & Shapiro, M. G. Measuring gas vesicle dimensions by electron microscopy. *Protein Science* **30,** 1081–1086. ISSN: 1469896x (2021).

8. Knitsch, R., Schneefeld, M., Weitzel, K. & Pfeifer, F. Mutations in the major gas vesicle protein GvpA and impacts on gas vesicle formation in Haloferax volcanii. *Molecular Microbiology* **106,** 530–542. ISSN: 13652958 (2017).

9. Dutka, P., Metskas, L. A., Hurt, R. C., Salahshoor, H., Wang, T.-Y., Malounda, D., Lu, G., Chou, T.-F., Shapiro, M. G. & Jensen, G. J. Structure of Anabaena flos-aquae gas vesicles revealed by cryo-ET. *Structure* **122,** 40a (2023).

10. Bollschweiler, D. Study of the archaeal motility system of Halobacterium salinarum by cryo-electron tomography, 1–118 (2015).

11. Walsby, A. E. & Hayes, P. K. The Minor Cyanobacterial Gas Vesicle Protein, GVPc, Is Attached to the Outer Surface of the Gas Vesicle. *Microbiology* **134,** 2647–2657. ISSN: 1350-0872 (1988).

12. Jost, A. & Pfeifer, F. Interaction of the gas vesicle proteins GvpA, GvpC, GvpN, and GvpO of Halobacterium salinarum. en. *Frontiers in Microbiology* **13,** 971917. ISSN: 1664-302X (July 2022).

13. Winter, K., Born, J. & Pfeifer, F. Interaction of haloarchaeal gas vesicle proteins determined by split-GFP. *Frontiers in Microbiology* **9,** 1–11. ISSN: 1664302X (2018).

14. Völkner, K., Jost, A. & Pfeifer, F. Accessory Gvp Proteins Form a Complex During Gas Vesicle Formation of Haloarchaea. *Frontiers in Microbiology* **11.** ISSN: 1664302X (2020).

15. Iburg, M., Anderson, A. P., Wong, V. T., Anton, E. D., He, A. & Lu, G. J. *Elucidating the Assembly of Gas Vesicles by Systematic Protein-Protein Interaction Analysis* en. preprint (Synthetic Biology, July 2023).

7

16. Tashiro, Y., Monson, R. E., Ramsay, J. P. & Salmond, G. P. Molecular genetic and physical analysis of gas vesicles in buoyant enterobacteria. *Environmental Microbiology* **18,** 1264–1276. ISSN: 14622920 (2016).

17. Jessop, M., Felix, J. & Gutsche, I. AAA+ ATPases: structural insertions under the magnifying glass. en. *Current Opinion in Structural Biology* **66,** 119–128. ISSN: 0959440X (Feb. 2021).

18. Li, Z., Shen, Q., Dai, Y., Anderson, A. P., Iburg, M., Lin, R., Zimmer, B., Meyer, M. D., You, L., Chilkoti, A. & Lu, G. J. *Spatial Organization of Gas Vesicles is Governed by Phase-separable GvpU* en. preprint (Synthetic Biology, June 2023).

19. Jost, M. *Die Ultrastruktur von Oscillatoria rubescens DC* PhD thesis (ETH Zurich, 1965).

20. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596,** 583–589 (2021).

21. Hofacker, A., Schmitz, K.-M., Cichonczyk, A., Sartorius-Neef, S. & Pfeifer, F. GvpE- and GvpD-mediated transcription regulation of the p-gvp genes encoding gas vesicles in Halobacterium salinarum. en. *Microbiology* **150,** 1829–1838. ISSN: 1350-0872, 1465-2080 (June 2004).

22. Hopf, T. A., Green, A. G., Schubert, B., Mersmann, S., Schärfe, C. P. I., Ingraham, J. B., Toth-Petroczy, A., Brock, K., Riesselman, A. J., Palmedo, P., Kang, C., Sheridan, R., Draizen, E. J., Dallago, C., Sander, C. & Marks, D. S. The EVcouplings Python framework for coevolutionary sequence analysis. en. *Bioinformatics* **35** (ed Valencia, A.) 1582–1584. ISSN: 1367-4803, 1367-4811 (May 2019).

23. Strunk, T., Hamacher, K., Hoffgaard, F., Engelhardt, H., Zillig, M. D., Faist, K., Wenzel, W. & Pfeifer, F. Structural model of the gas vesicle protein GvpA and analysis of GvpA mutants in vivo. *Molecular Microbiology* **81,** 56–68. ISSN: 0950382x (2011).

24. Ezzeldin, H. M., Klauda, J. B. & Solares, S. D. Modeling of the major gas vesicle protein, GvpA: From protein sequence to vesicle wall structure. *Journal of Structural Biology* **179,** 18–28. ISSN: 10478477 (2012).

25. Terwilliger, T. C., Liebschner, D., Croll, T. I., Williams, C. J., McCoy, A. J., Poon, B. K., Afonine, P. V., Oeffner, R. D., Richardson, J. S., Read, R. J. & Adams, P. D. AlphaFold predictions are valuable hypotheses and accelerate but do not replace experimental structure determination. en. *Nature Methods.* ISSN: 1548-7091, 1548-7105 (Nov. 2023).

26. Wang, H., Zhang, J., Toso, D., Liao, S., Sedighian, F., Gunsalus, R. & Zhou, Z. H. Hierarchical organization and assembly of the archaeal cell sheath from an amyloid-like protein. en. *Nature Communications* **14,** 6720. ISSN: 2041-1723 (Oct. 2023).

SUMMARY

Cryogenic electron microscopy (cryo-EM) has become a powerful technique to understand the structure and function of proteins. Thanks to substantial technical advancements in the 2010s, the frequent determination of atomic structures is now possible. This thesis encompasses chapters that contribute to both the remaining technical challenges of cryo-EM, and to the advancement of biological insight through the application of cryo-EM to a motility machinery called a gas vesicle. As an overarching theme, the thesis is titled 'Imaging Life at the Molecular Scale using Electrons'. **Chapter 1** provides an introduction into methods to study structure and function in biology at the molecular scale, and into the development of cryo-EM since its inception. A comprehensive review of previous structural work on gas vesicles follows.

Technological improvements in electron detectors have significantly enhanced the quality of protein images obtainable. However, predicting the success of cryo-EM structure determination still poses a challenge, as it is heavily dependent on the unique properties of the sample and its behavior during the sample preparation process. Proteins are typically forced into an thin liquid layer, which is frozen by quick immersion into a cryogen. The large air-water interface can cause proteins to aggregate, denature or adopt preferred orientation, hindering structure determination. The thickness of the layer is also hard to control - too thick layers lead to poor image contrast, too thin layers damage the protein. In **Chapter 2** we outline a new method of sample preparation using nanofluidic channels, which replaces the air-water interface with an air-silicon nitride interface and controls the ice thickness by the inherent channel geometry. We test this new method with three different test specimens and show in a proof-of-concept study that structure determination with up to 3 Å is possible with these devices.

During sample preparation, the protein is rapidly immobilized in its current conformation in less than a millisecond. This rapid arrest can be exploited for conducting time-resolved EM (trEM) studies. **Chapter 3** details the design and fabrication of a trEM setup, which utilizes a fast linear motor for rapid sample freezing and LEDs to initiate reactions through flash photolysis.

In **Chapter 4**, we utilize cryo-EM to investigate a protein assembly known as a gas vesicle. These organelles, essentially gas-filled protein shells, are produced by microbes to produce buoyancy, functioning as intricate motility mechanisms. The atomic structure of gas vesicles had remained elusive, presenting challenges to other structural biology methods like NMR or X-ray crystallography. By leveraging recent technical advancements in cryo-EM and adopting a novel strategy that focuses on more stable and thinner gas vesicles, we have successfully elucidated the atomic structure of these vesicles for the first time. The structure explains intriguing properties of gas vesicles such as their stability, gas permeability and ability to grow while excluding water.

We follow up on this work in **Chapter 5**, which focusses on a second structural protein in gas vesicles that binds to the outside of the shell. Cryo-EM observation, bioinformatics

analysis and experimental mutant analysis are combined to reveal the geometry and binding site of the GvpC-gas vesicle interaction.

Chapter 6 utilises structural data from previous thesis chapters to create educational models by 3D printing. Such models were useful during this thesis to reason and argue about proposed structural features. Example models in this chapter showcase (1) the concept of point group symmetry in proteins, (2) our proposed growth mechanism of gas vesicles and (3) the general atomic structure and fold of proteins.

Samenvatting

Cryogene elektronenmicroscopie (cryo-EM) is een krachtige techniek geworden om de structuur en functie van eiwitten te begrijpen. Dankzij aanzienlijke technische vooruitgang in de jaren 2010 is het nu mogelijk om regelmatig atomaire structuren te bepalen. Deze thesis omvat hoofdstukken die bijdragen aan zowel de resterende technische uitdagingen van cryo-EM, als aan de vooruitgang van biologisch inzicht door de toepassing van cryo-EM op een motility-apparaat genaamd een gasblaasje of gas vesicle in het Engels. Als overkoepelend thema heeft de thesis de titel 'Imaging Life at the Molecular Scale using Electrons'. **Hoofdstuk 1** biedt een inleiding in methoden om structuur en functie in de biologie op moleculaire schaal te bestuderen, en in de ontwikkeling van cryo-EM sinds de oprichting. Een uitgebreide review van eerdere structurele werken over gasblaasjes volgt.

Technologische verbeteringen in elektronendetectoren hebben de kwaliteit van de verkrijgbare eiwitbeelden aanzienlijk verbeterd. Het voorspellen van het succes van cryo-EM structuurbepaling blijft echter een uitdaging, aangezien dit sterk afhankelijk is van de unieke eigenschappen van het monster en het gedrag tijdens het voorbereidingsproces van het monster. Eiwitten worden typisch in een dunne vloeistoflaag gedwongen, die snel wordt bevroren door onderdompeling in een cryogeen. De grote lucht-water interface kan ervoor zorgen dat eiwitten aggregeren, denatureren of een voorkeursoriëntatie aannemen, wat de structuurbepaling hindert. De dikte van de laag is ook moeilijk te beheersen - te dikke lagen leiden tot slecht beeldcontrast, te dunne lagen beschadigen het eiwit. In **Hoofdstuk 2** beschrijven we een nieuwe methode voor monsterbereiding met nanovloeistofkanalen, die de lucht-water interface vervangt door een lucht-siliciumnitride interface en de ijsdikte controleert door de inherente kanaalgeometrie. We testen deze nieuwe methode met drie verschillende testmonsters en tonen in een proof-of-concept studie aan dat structuurbepaling tot 3 Å mogelijk is met deze apparaten.

Tijdens de monsterbereiding wordt het eiwit snel in zijn huidige conformatie geïmmobiliseerd in minder dan een milliseconde. Deze snelle arrestatie kan worden benut voor het uitvoeren van tijdopgeloste EM (time-resolved EM, trEM) studies. **Hoofdstuk 3** beschrijft het ontwerp en de fabricage van een trEM-opstelling, die gebruik maakt van een snelle lineaire motor voor snelle monsterbevriezing en LED's om reacties te initiëren door middel van flash photolysis.

In **Hoofdstuk 4** gebruiken we cryo-EM om een eiwitmontage bekend als een gasblaasje te onderzoeken. Deze organellen, in wezen gasgevulde eiwitschalen, worden door microben geproduceerd om drijfvermogen te produceren, en functioneren als ingewikkelde bewegingsmechanismen. De atomaire structuur van gasblaasjes bleek te complex om te worden opgelost door structurele biologiemethoden zoals NMR of röntgenkristallografie, waardoor hun ongrijpbaarheid behouden bleef. Door recente technische vooruitgang in cryo-EM te benutten en een nieuwe strategie toe te passen die zich richt op stabielere en dunnere gasblaasjes, zijn we er voor het eerst in geslaagd de atomaire structuur van deze blaasjes te ontrafelen. De structuur verklaart intrigerende eigenschappen van gasblaasjes,

zoals hun stabiliteit, gasdoorlaatbaarheid en het vermogen om te groeien terwijl water wordt uitgesloten.

We borduren voort op dit werk in **Hoofdstuk 5**, dat zich richt op een tweede structureel eiwit in gasblaasjes dat zich aan de buitenkant van de schaal bindt. Cryo-EM data, bio-informatica-analyse en experimentele mutatieanalyse worden gecombineerd om de geometrie en bindingsplaats van de GvpC-gasblaasje interactie te onthullen.

Hoofdstuk 6 maakt gebruik van structurele gegevens uit eerdere thesis hoofdstukken om educatieve modellen te creëren door middel van 3D-printen. Deze modellen demonstreren (1) het concept van puntgroepsymmetrie in eiwitten, (2) ons voorgestelde groeimechanisme van gasblaasjes en (3) de algemene atomaire structuur en vouwing van eiwitten.

Acknowledgments

Hello most esteemed reader. I trust that this is not the only section of the thesis that you will look at! If you want to read something light and fun, maybe try **Chapter 6** about 3D printed educational models as a start. If you want to have a giggle about my attempt to translate the work into Dutch, you can also have a go at the summary chapter with great words such as 'gasblaasjes', 'voorkeursoriëntatie' and 'gasdoorlaatbaarheid'. If you really just want to find your name in the acknowledgments, keep going, and watch out for the **bold** words.

I start with a thank you to my scientific home over the last five years, the **Bionanoscience department** and the **Kavli Institute of Nanoscience**: borrels, pub quizzes, the Thursday croissant, Quo Vadis outings, Ski trips, Sailing trips, Kavli days and our Monday Forums were always a fun occasion to learn from one another and hear about interesting science going on.

I am grateful to **Arjen Jakobi** for many things throughout my PhD journey. I have fond memories of 'environmental sampling', convincing you to buy a 3D printer, PhD movie production, staring at proteins at the microscope, experimenting with 5 people at a time on cryoChips, and many more. Your ability to create an environment where we could follow our scientific curiosity, explore ideas, and have access to all the necessary tools while always counting on your support if necessary — no matter how busy you were — has been quite amazing. I am very happy to have come to your lab for my PhD. Your enthusiasm for both scientific discovery and good scientific practice has been a constant source of learning for me. Whether writing reviews, papers, or discussing results and ideas, I always enjoyed our time together in your office. As I prepare to leave Delft, the bar will be set quite high for my next place to live up to this, as I have (in quite funny instances) experienced during my job search journey. Thank you, Arjen, for everything!

Continuing to my paranymph, **Clémence** (with é): first of all, sorry for having to be a paranymph again after such a short time span, and thank you so much for accepting the job. I hope you delegated all the movie work to our successors in the lab! I have fond memories of our own paranymphing team work, last minute movie production, acting performances and gift brainstormings. It was always fun to have you around, design elefant tweezer holders, to learn about horrifying viruses in the ML-2, making a tiny little bit of fun of a word you pronounced, and truly being very impressed by all the cell biology experiments you design.
Miss paranymph **Tanja**, thank you so much for accepting the role, even from far away in Germany, and I hope it will not be too much hassle (ass'ole) to do remote-paranymphing. So, this time it appears that I am not following you to your new employer... or wait... what is the name of your new lab again? Just kidding. It was fun times for many many years,

starting 2017 in Carstens lab in Heidelberg, and now aparently 7 years later we do not work in the same place anymore. Thank you for always being the positive, social center in any of those environments.

I'd like to thank **Roeland Huijink**, **Edin Sarajlic** and **Tim Horstink**. Prior to my PhD, you had been brewing up ideas with **Arjen** and **Wiel** on using your nanofacrication wizardry to make sample preparation methods for cryo-EM. I joined somewhere in the middle of this process and had the opportunity to help test your amazing prototypes with water, gold particles, and finally luckily also with some actual protein samples. Every time you visited, I was quite blown away by a new sample you etched/deposited/lithographed. I am a bit sorry in the end that I got so side-tracked with other projects and did not spend more time on your amazing chips. Thank you for this cool collaboration!

Wiel Evers is our group's master tinkerer who can seemingly assemble and fix anything from a cryo plunger to an electron microscope. Not sure what we would do in those moments when things break and we think "Wieeeeeel" and you would be not around. I hope I may have picked up a trick or two from you during my time in Delft. Thank you to **Michal Shemesh**, for the amazing organisation of the Winter Quo Vadis (+kids day) in the department. I had lots of fun setting up one of the activity stations for the kids.

A number of Bachelor and Master students worked with me on their thesis projects throughout the years. **Rhodé van den Dool** did her BEP project on computational methods for initial model generation in cryo-EM by repurposing software from the SAXS field. I was impressed how self-organised you were - doing the entire work remotely in Covid times and just meeting from time to time to discuss. My second BEP student was **Kristie Tjokro**. Thanks to Alok for taking on thesis supervision together with me. Kristie worked with the time-resolved plunging setup, did testing, simulations and validation as well as temperature measurements during plunging. Sorry we got you into all the soldering and physics theory, when you actually came to our group to escape physics (?). You did a great job! And finally, **Flip Jansen** came to us for a MEP project. Big thank you to Maarten for co-supervision. Flip worked with us on automated AI image analysis of large cryo-EM datasets. I learned a lot from Flip on how to organise efficient progress meetings. I wish you all the best, no matter if in a startup, in research or getting rich with high-speed trading. I am sure you will do great in any of these.

Cecilia, you were always my first go-to point for questions about TU Delft bureocracy, ordering stuff and how/where stuff is in the lab. Thank you for your help over the years. "Ühhhhhh" **Natasha**, it was always great to have a chat in the office, me and you complaining how difficult everything is. I wish you all the best for your next step. **Lennart**, I hope very much that your helices will start behaving, and I had great fun to look into your puzzles from time to time. Thank you **Alok** for co-supervising Kristie, discussing physics and all the astronomy stories. **Maarten**, I was impressed by your passion for good teaching and I hope I learned a thing or two from you during Flip's MEP project. **Matteo**, I hope your tetramer will work out in the end and you had a good time with us! **Birgit**, it was a joy to be your office neighbor. I was impressed by your attitude to always show

initiative to learn new things and ask questions, and I hope you will be happy with the decision of what to do next. **Sofia**, you were a pleasure to have around, and probably you were the only extrovert person present in most of our meetings and lunches. Thank you for contributing your positive energy. A shoutout to more friendly faces: **Mart**, **Christos**, **Sam**, **Daniel**, **Sophie**, **Jaïrus**, **Beatriz**, **Ramon**, **Leanid**, **Nemo**, **Carlos**, **Ashmiani**.

I want to thank **Dion Terwiel** for our engaging discussions on gas vesicle biology, applications (and on how to surf in the Netherlands). The *A. flos-aquae* GV samples you prepared were quite stunning under the microscope, especially from the eyes of a structural biologist who likes to look at helical structures. Your sample initiated a whole journey of structural investigation into this crazy structure. My best wishes to you for your engineering endeavours and for finishing your PhD journey.

Thank you to **Simon Mortensen** for the many phone calls over the years, your help with academic troubles and your scientific advice. The things I learned from you during our time in Heidelberg is still incredibly useful all the time.

Many things I wrote are work-related, so some of you readers might get nervous to not find your name? I cannot but give a shoutout to friends in the Netherlands who I shared a lot of fun with and got support from over the years: **Nicole**, **Stephan**, **Quentin**, **Sona**, **Luca**, **Milan**, **Luigi**, **Constantina**, **Gert-Jan**, **Deepika**, **Deniz**, **Raj**. And to **Nina**: I am happy you said yes to that dinner invitation ;)

Last but not least I want to thank all my committee members who took on this role and the time to engage with the content of the thesis and propositions. With around 1000 defenses per year at TU Delft, I imagine you get asked to be in a committee all the time, even for topics somewhat further away from your interests. Thank you for accepting.

Stefan
Delft, August 2024

Curriculum Vitæ

Stefan HUBER

10.12.1993	**Born in** Gengenbach, Baden-Württemberg, Germany
2004–2012	**Marta-Schanzenbach-Gymnasium Gengenbach** High school diploma (Abitur) in 2012
2012-2015	**University of Heidelberg** Bachelor Degree in Molecular Biotechnology Bioinformatics track
2015	**Riken Center for Developmental Biology** Research Assistant in Kobe, Japan
2015-2018	**University of Heidelberg** Master Degree in Molecular Biotechnology Biophysics Specialization
2017-2019	**EMBL Heidelberg** European Molecular Biology Laboratory Research assistant and Master thesis in the Group of Dr. Carsten Sachse
2019-2024	**TU Delft - Kavli Institute of Nanoscience** PhD candidate in the group of Dr. Arjen Jakobi

List of Publications

1. Huber, S. T., Kuhm, T. & Sachse, C. Automated tracing of helical assemblies from electron cryo-micrographs. *Journal of structural biology* **202,** 1–12 (2018)

2. Bullmann, T., Radivojevic, M., Huber, S. T., Deligkaris, K., Hierlemann, A. & Frey, U. Large-scale mapping of axonal arbors using high-density microelectrode arrays. *Frontiers in cellular neuroscience* **13,** 404 (2019)

3. Huber, S. T., Mostafavi, S., Mortensen, S. A. & Sachse, C. Structure and assembly of ESCRT-III helical Vps24 filaments. *Science advances* **6,** eaba4897 (2020)

4. Jakobi, A. J., Huber, S. T., Mortensen, S. A., Schultz, S. W., Palara, A., Kuhm, T., Shrestha, B. K., Lamark, T., Hagen, W. J., Wilmanns, M., *et al.* Structural basis of p62/SQSTM1 helical filaments and their role in cellular cargo uptake. *Nature communications* **11,** 1–15 (2020)

5. Junglas, B., Huber, S. T., Heidler, T., Schlösser, L., Mann, D., Hennig, R., Clarke, M., Hellmann, N., Schneider, D. & Sachse, C. PspA adopts an ESCRT-III-like fold and remodels bacterial membranes. *Cell* **184,** 3674–3688 (2021)

📄 6. Huber, S. T., Sarajlic, E., Huijink, R., Weis, F., Evers, W. H. & Jakobi, A. J. Nanofluidic chips for cryo-EM structure determination from picoliter sample volumes. *Elife* **11,** e72629 (2022)

> **Editor's evaluation**
> Sample preparation for single-particle electron cryo-microscopy (cryo-EM) remains a bottleneck of this technique. The sample ice thickness cannot be accurately controlled, molecules may display strongly preferred orientations that make more elaborate data collection schemes necessary, and the sample may degrade at the air-water interface before it is finally frozen. In their pioneering work, the authors describe a prototype of a new microfluidic device that addresses some of these problems, including a refreshingly objective and critical discussion about the pros and cons of this novel approach. While some development will be required for this method to become mainstream, it has the potential to become a powerful alternative to the conventional workflow of single-particle cryo-EM, enabling full automation and making sample preparation highly reproducible.

📄 7. Huber, S. T., Terwiel, D., Evers, W. H., Maresca, D. & Jakobi, A. J. Cryo-EM structure of gas vesicles for buoyancy-controlled motility. *Cell* **186,** 975–986 (2023)

8. Kuhm, T. I., Gross, L., de Agrela Pinto, C., Huber, S. T., Taisne, C., Giannopoulou, E. A., Pardon, E., Steyaert, J., Tans, S. J. & Jakobi, A. J. Structural basis of membrane targeting and coatomer assembly by human GBP1. *bioRxiv,* 2023–03 (2023)

📄 9. Huber, S. T., Jakobi, A. J. Structural biology of microbial gas vesicles: Historical milestones and current knowledge. *Biochemical Society Transactions* (2024)

📄 Included in this thesis.

.