Cell
Leading Edge

CellPress
OPEN ACCESS

Perspective

Understanding the cell: Future views of structural biology

Martin Beck,[1,3,4,5,*] Roberto Covino,[2,4,5,*] Inga Hänelt,[3,4,5,*] and Michaela Müller-McNicoll[3,4,5,*]
[1]Max Planck Institute of Biophysics, Max-von-Laue-Straße 3, 60438 Frankfurt am Main, Germany
[2]Frankfurt Institute for Advanced Studies, Ruth-Moufang-Straße 1, 60438 Frankfurt am Main, Germany
[3]Goethe University Frankfurt, Frankfurt, Germany
[4]Senior author
[5]These authors contributed equally
*Correspondence: martin.beck@biophys.mpg.de (M.B.), covino@fias.uni-frankfurt.de (R.C.), haenelt@biochem.uni-frankfurt.de (I.H.), mueller-mcnicoll@bio.uni-frankfurt.de (M.M.-M.)
https://doi.org/10.1016/j.cell.2023.12.017

SUMMARY

Determining the structure and mechanisms of all individual functional modules of cells at high molecular detail has often been seen as equal to understanding how cells work. Recent technical advances have led to a flush of high-resolution structures of various macromolecular machines, but despite this wealth of detailed information, our understanding of cellular function remains incomplete. Here, we discuss present-day limitations of structural biology and highlight novel technologies that may enable us to analyze molecular functions directly inside cells. We predict that the progression toward structural cell biology will involve a shift toward conceptualizing a 4D virtual reality of cells using digital twins. These will capture cellular segments in a highly enriched molecular detail, include dynamic changes, and facilitate simulations of molecular processes, leading to novel and experimentally testable predictions. Transferring biological questions into algorithms that learn from the existing wealth of data and explore novel solutions may ultimately unveil how cells work.

INTRODUCTION

Structural biology is an attempt to answer the question "what are we made of?" This attempt follows the reductionist approach, which aims to identify the most fundamental constituents of matter and study their properties. It led us to discover a hierarchy of structures, from molecules through atoms all the way down to fundamental particles, such as quarks and electrons. Cells are the minimal units of life and are made of billions of distinct molecules. Although this answers part of the question of what we are made of, it does not answer a key question of cell biology—how do cellular functions spontaneously emerge from the interaction of these billions of molecules? Cell biology usually lacks the structural resolution to understand the role of individual molecules and the choreography that organizes them in functional units, which ultimately distinguishes a living cell from an inanimate object. To gain this understanding, the integration of structural and cellular biology is an outstanding challenge.

With the discovery of the DNA double-helix and the first protein structures, a structure-function paradigm emerged, underpinning the implicit assumption of structural biology: by knowing the detailed structures of biomolecules, one will understand their function, and the sum of all individual structure-function relationships will enable us to explain how cells work. This approach has been immensely successful because it led to an atomistic picture of many molecular machines and for many molecules set the foundation of our present understanding of their function. However, with increasing coverage and in-depth characterization of the cell's constituents, challenges to this assumption are emerging.

The first challenge stems from the realization that all biomolecules are inherently dynamic. Thermal fluctuations can transmit energy to molecules from their environment. In response, these molecules will experience spontaneous conformational changes, ranging from the local flipping of a side chain to global folding processes. Instead of considering a biomolecule as a single well-defined static structure, we must think of it as a structural ensemble, i.e., a large collection of conformations, each populated with different probabilities.[1] The molecule will stochastically interconvert between different conformations. For some molecules, there will be few conformations overwhelmingly more probable than others, such as the globular protein serum albumin; but for others, the ensemble will be very heterogeneous, consisting of many conformations, all nearly equally probable, such as in the case of disordered proteins. Increasing evidence points to the fact that the entire conformational ensemble, including rare conformations, determines the function of a biomolecule.[2,3] Such an ensemble view implies that the probability of populating the different alternative conformations can be modulated by thermodynamic parameters, interactions with other biomolecules, post-translational modifications

This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

(PTMs), or the local physiochemical features of the environment, such as ion concentrations. This view is well established in biophysics; rooted in the statistical mechanics describing molecular systems; and explored by molecular dynamics (MD) and biophysical approaches, such as nuclear magnetic resonance (NMR), electron paramagnetic resonance (EPR), Förster resonance energy transfer (FRET), hydrogen-deuterium exchange (HDX) mass spectrometry (MS), and single-molecule force-spectroscopy. A conformational ensemble view provides a comprehensive instrument to understand how a dynamic environment can modulate a molecular conformational ensemble and its function (Figure 1).

The second, more profound challenge is related to the limits of using the reductionist approach to understand complex systems. In 1972, physicist P. Anderson published the very influential article "More is different."[4] Anderson claimed that the reductionist hypothesis does not imply a constructionist one. In other words, knowing the elementary constituents of complex systems is not enough to understand how these systems actually work. The reason for this is that new properties and effective laws emerge in complex systems that are very difficult to predict from their fundamental description. For example, even if we knew the precise chemical structure of a lipid molecule, it would be very challenging to predict that many lipids together with water and in the presence of thermal fluctuations will spontaneously self-assemble into complex structures, such as lipid bilayers. To describe an isolated lipid molecule, we use concepts such as configurations, dihedral angles, and chemical bonds. Yet, these are useless to describe a bilayer, where concepts coming from materials physics, such as phase behavior, mechanical moduli, and lateral organization, are instead more pertinent. Similarly, polymer physics offers important tools to conceptualize the properties of membrane-less organelles or chromatin, while characterization of the individual components relies on different concepts, such as chemical bonds and their rotation, charge distribution, the scaling law or multivalent binding. However, predicting their behavior as polymers *in situ* remains challenging. Molecules in a cell constantly interact with each other in a self-organizing manner, creating a dynamic subcellular organization with higher-order macromolecular assemblies in astonishing complexity in a tightly regulated choreography. This molecular sociology of cells[5] is not well understood. Hence, we need a new conceptual framework, language, and technical tools to bridge the gap from single molecules to the entirety of the cell.

In this perspective on the future of structural cell biology, we review the limits of present-day approaches. We discuss the concept of cellular self-organization, which is crucial for cellular function and includes phenomena such as local confinement, molecular rulers, and self-organization mechanisms of membranes. We identify high-yield targets for technological developments and sketch our vision of the next generation of experiments, structural models, and scientific conceptualization and how this may enable us to extend structural biology from studying isolated macromolecular assemblies toward understanding how these assemblies self-organize into the complex structures and pathways that we observe inside cells. Overcoming this challenge will require the integration of structural and cell biology, biophysics, and computational sciences.

STATE OF THE ART OF STRUCTURAL BIOLOGY: ADVANCES AND LIMITATIONS

As of today, a considerable part of the molecular repertoire of cells has been elucidated at a high resolution. Structural analysis in combination with *in vitro* reconstitution and advanced biochemical techniques uncovered the mechanisms of an increasing number of molecular machines at a very high molecular detail. This wealth of data enabled training of a new generation of AI-based prediction tools that in turn accelerated structure determination efforts further.[6] AI-based analyses revealed a great number of previously unknown domain folds, alternative isoforms, and protein interfaces and even allowed the design of new proteins.[7–10] Integrated structural biology has made very large cellular assemblies amenable to structural analysis, which had seemed far out of reach only 10 years ago. Finally, *in situ* structural biology techniques have been used to provide the first glimpse of how molecular machines operate inside cells. This includes detailed insights into key processes of life, such as DNA replication, transcription, chromatin remodeling, RNA processing, cellular transport, translation, autophagy, membrane remodeling, and many more, as exemplified in work by Greenan et al.,[11] Deguchi et al.,[12] O'Reilly et al.,[13] Greber et al.,[14] Wagner et al.,[15] and Li et al.[16] The progress in charting the structured territories of the cellular interior has, however, also made it very apparent that some regions of our cellular maps continue to remain white spots, despite the new technologies and advances noted above. In the following section, we will cover several current examples, where present-day approaches have reached their technical limits.

Molecules within cells are inherently dynamic, but present-day structural biology is mostly blind to conformational dynamics. Traditional *in vitro* structural biology techniques, such as single-particle cryo-electron microscopy (cryo-EM) or X-ray crystallography, have difficulties in capturing low-abundant or dynamic species and transition states. They usually select for high resolution based on the averaging of large numbers of homogeneous particles. Poorly sampled states, short-lived intermediates, or disordered regions will thus be missed. To some extent, applying X-ray free-electron laser (XFEL) pulses to protein nanocrystals[17] or performing single-particle cryo-EM under turnover conditions, where the biological process under scrutiny is snap frozen,[18,19] can enrich the conformational ensemble of protein complexes. NMR and EPR spectroscopy are capable of complementing these shortcomings[20–22] because they enable sampling of conformational dynamics of biomolecules, or at least distinct atoms, in biomolecules. Single-molecule FRET[23,24] and high-speed atomic force microscopy (hsAFM)[25–27] even provide temporal resolution of the dynamics of individual molecules. Finally, HDX-MS quantifies solvent accessibility and thus reveals binding sites of other molecules and interaction partners or conformational changes.[28] However, all these techniques require rather large amounts of the respective biomolecules or site-specific labeling; they are in part limited in the size of the studied molecule and are mostly applied *in vitro*. As a consequence of these limitations, we often understand little about structural transitions and dynamics, which are key to understanding how molecules carry out their biological functions.

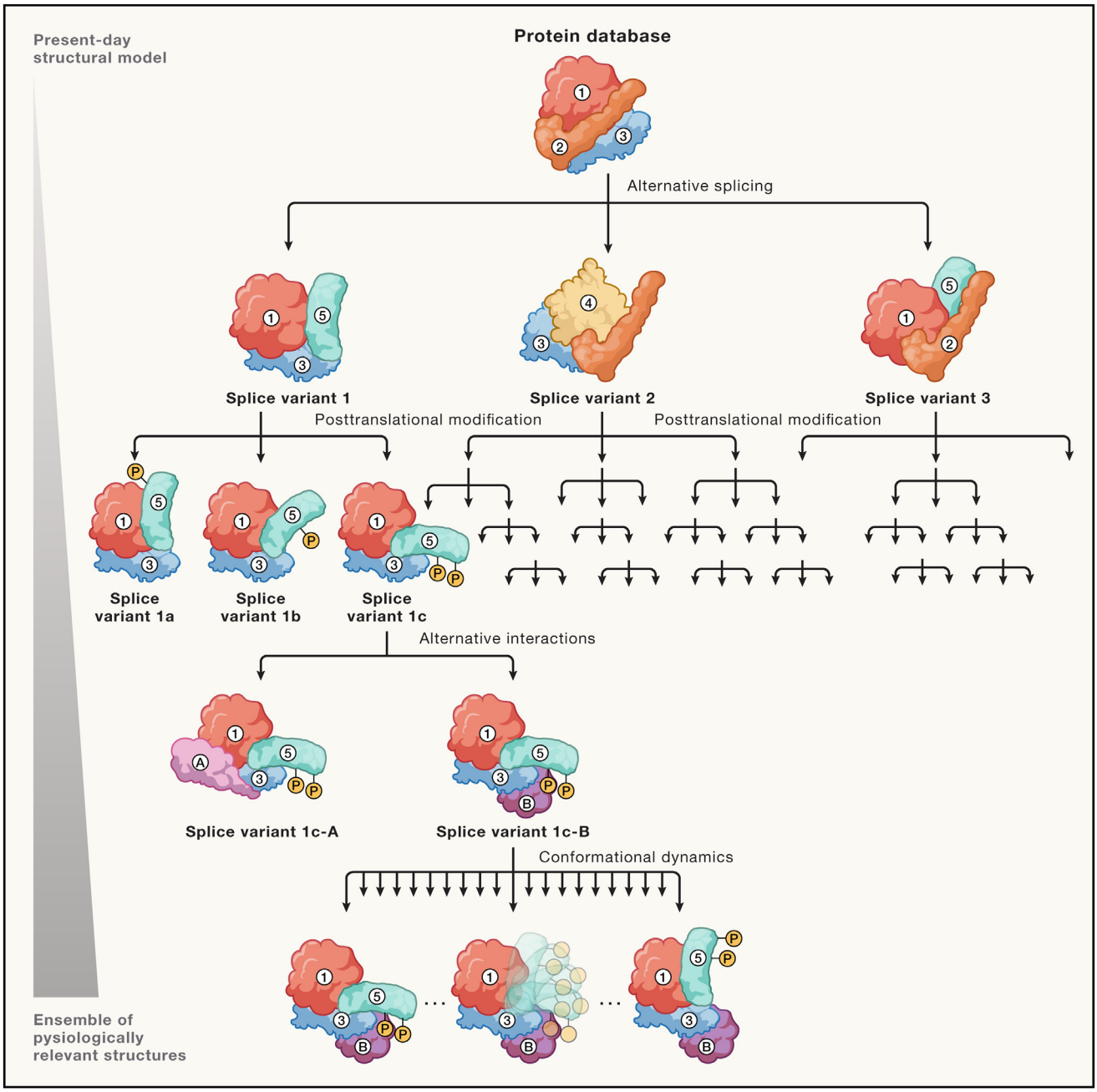

Figure 1. Scheme illustrating the extent to which the structure of protein complexes may be modulated in eukaryotes

A single structure of a given protein complex represented by its PDB entry should be conceived as a collapsed conceptualization of various splice isoforms, post-translational modifications (PTMs), alternative interactions, and conformations that modulate the respective structural ensemble inside of cells.

Thus, characterizing structural dynamics experimentally, albeit being essential, remains challenging, and novel or further refined methods are urgently required.

Another emerging fact is that many unstructured molecules are important for cellular function. For example, proteins containing intrinsically disordered regions (IDRs) are very abundant in higher eukaryotes and play essential roles in various biological processes,[29,30] but we lack information on their structure or conformation. Such proteins contribute to the formation of membrane-less organelles, define the biophysical properties of subcellular microenvironments, bend membranes, extend the interaction repertoire of protein complexes beyond folded interfaces, and even confine small molecules.[31–35] However, all these physiological functions remain ill-defined at the molecular level due to the absence of structural data.

Box 1. Structural understanding of mRNPs

Although many of the key enzymes such as RNA polymerases, the spliceosome, RNA degraders, helicases, and the ribosome are understood in considerable detail,[36–40] we know little about the actual structure of messenger ribonucleoproteins (mRNPs). This is largely because mRNPs show a large compositional complexity, have a high conformational flexibility, and are constantly remodeled throughout their life cycle. Moreover, a large number of RNA-binding proteins feature extended IDRs and form biomolecular condensates. The different regions of an mRNA (UTRs, coding region, and poly(A)-tail) are occupied by different RNA-binding proteins with different IDRs and different condensate properties. Within mRNPs also RNA structure and RNA:RNA interactions have to be considered. This diversity allows the spatial and temporal separation of RNA processing or editing events; prevents RNA modification, tangling, and degradation; and promotes proper packaging, sorting, and transport of mRNAs.[41–43] At the same time, it makes *in vitro* structural approaches nearly impossible, which complicates the molecular conceptualization of mRNPs.

Structural biology also largely neglects post-transcriptional modifications and PTMs. For example, alternative splicing variants that govern cell-type specificity in eukaryotes are not commonly considered in structural models, and consensus gene models are used instead, e.g., for the production of recombinant proteins. Many PTMs are highly dynamic and added to proteins in a stochastic manner. Thus, they often are lost during averaging techniques, while omics techniques also struggle to provide a comprehensive picture of all PTMs. Box 1 exemplifies how these shortcomings limit our understanding of messenger RNA-protein particles (mRNPs) and their processing.

Functional importance of high-resolution conformational states

Structural analyses in combination with *in vitro* reconstitution have elucidated the fine details of many molecular processes. However, *in vitro* structural biology often operates under the assumption that the structural states that converge to high resolution, e.g., by averaging-based techniques, are also the states that are functionally important. But is this a valid assumption? First, production and assembly of the respective molecular ingredients in a test tube limit which and how many structural states can be captured. More importantly, such *in vitro* approaches are blind to the cellular context, such as local concentrations, exclusion effects, PTMs, or alternatively spliced transcript isoforms. Inarguably, one can infer constructive hypotheses about how given structures look or work *in situ*; yet, local folding and fine details are likely different inside cells, and conformational equilibria will be shifted.

Thus, favoring high-resolution structures imposes a bias. For example, during cryo-EM single-particle classification and sorting, a considerable fraction of molecules that do not contribute to high resolution is often considered "junk" or damaged and is removed from the analysis. The pitfalls of this procedure were recently demonstrated for ribosomes: during *in situ* structural analysis of translation inside of intact cells, where molecular damage can be largely ruled out, many of the analyzed ribosomes showed clear-cut features of native activity, i.e., the presence of tRNAs during translation elongation, but still did not converge to high resolution.[44–46] These particles thus likely represent conformationally variable states or transitions between structurally more defined intermediates within the ensemble of all states (Figure 1). This does not mean that they are functionally less important. To value a given structure based solely on the resolution that has been achieved may therefore be misleading. We should embrace the realization that medium-to-high-resolution structures are often sufficient to draw functionally important conclusions and that coupled with AI-generated structural ensembles, they allow for an accurate understanding of processes in cells that could not be captured solely by *in vitro* high-resolution structural analysis.

Consequently, the question of whether the resolution of a structure is high enough only makes sense in the context of a scientific question that the structure should answer. To assess a large conformational change or to understand the overall architecture of a large macromolecular assembly, medium-range resolutions at around 10–20 Å, often achieved in integrative structural biology, are sufficient. In such cases, it does not negatively affect the overall conclusions if some side-chain dihedrals are wrong. In contrast, understanding the mechanism of ion coordination requires a much higher resolution of <3 Å, possibly also resolving water molecules around the site, which is typical for X-ray crystallography or single-particle EM.

Advancing structural biology by combining *in vitro* and *in situ* data

Much of our understanding of molecular activities at high resolution is inferred from *in vitro* reconstitution techniques. These were traditionally used to analyze complex cellular phenomena experimentally because it is very difficult to observe molecules at work inside of cells. But in almost every single case when *in situ* structural analysis has been successfully carried out, the results challenged previous dogmas or shifted scientific concepts considerably. One example is the barrel-shaped 26S proteasome, composed of the 20S core and 19S caps, which is the major degradation chamber of cells. While the 19S cap unfolds proteins, the 20S core particle chops them into peptides. For decades, biochemists have tried to purify double-capped 26S proteasomes, which had been conceived as notoriously unstable *ex cellulo*, falling apart into the 20S core and 19S cap particles. This quest came to an end when *in situ* structural analysis revealed that the majority of proteasomes are actually singly capped in cells.[47] Other examples are ribosomes, which translate the genetic message encoded in mRNAs into proteins by cycling through various functional states that have been carefully characterized in isolation. Initially, quantitative assessment of the distribution of elongation cycle intermediates inside of cells was inferred from analyses of complexes that were rapidly isolated from active cells.[48] However, the exact distribution of elongation states turned out to be different once such analyses became possible inside of cells.[44,45] Similarly, the architecture of the bacterial expressosome that consists of a leading RNA polymerase and the trailing ribosome differs inside of cells compared with previous *in vitro* analyses.[13]

Another striking example is the nuclear pore complex (NPC), a 120-MDa assembly encoded by ~30 different

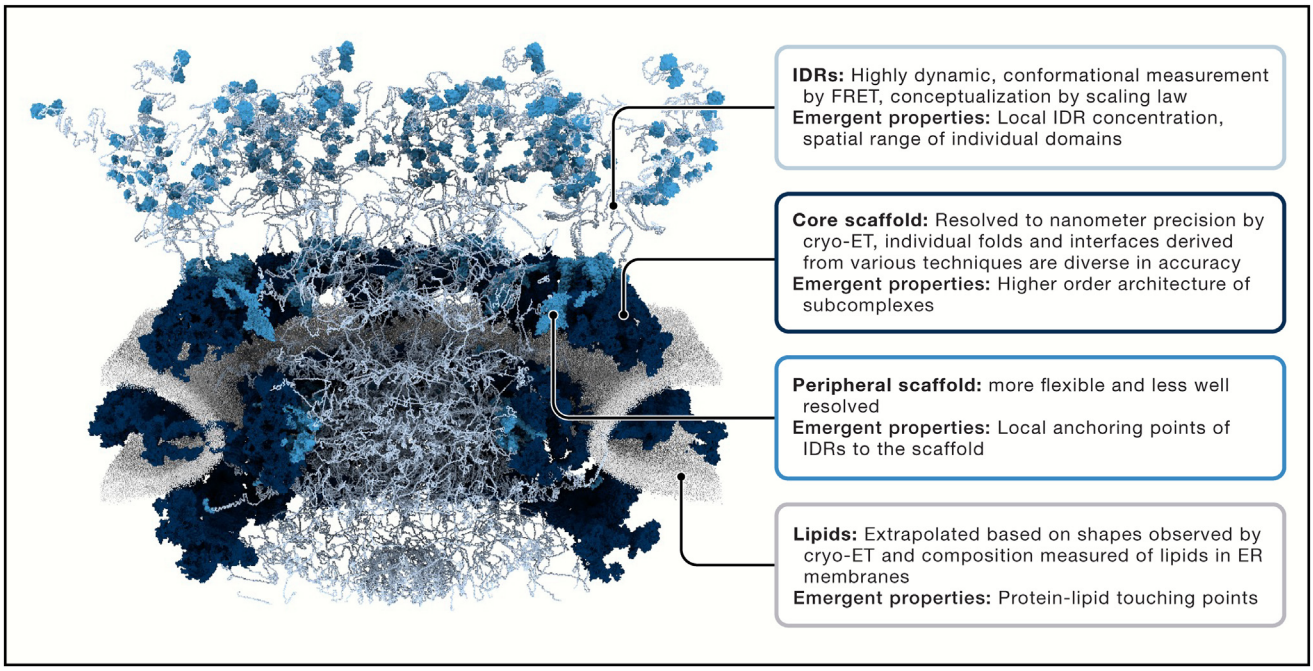

Figure 2. Integrated structural model of the nuclear pore
Different types of data were used to model specific parts of the displayed structural model. The resulting high-fidelity model is heterogeneous in terms of accuracy but a prerequisite for molecular dynamics simulations. The subtypes of components, their accuracy, and emergent properties of the model are indicated in boxes.

genes. *In vitro* structural analysis of nucleoporins by X-ray crystallography revealed the respective folds and their subcomplexes, which were essential to understand NPC architecture. However, the structural ensemble of individual components was insufficient to understand how NPCs oligomerize into fully assembled nuclear pores, which was revealed only in combination with *in situ* structural analysis[49] (Figure 2).

However, the weakness of *in vitro* structural biology is also its strength. Although biological complexity inside test tubes remains limited, the ability to precisely control the ingredients of a given structural analysis provides a strong advantage, as according to Feynman "What I cannot create, I do not understand." *In vitro* structural biology will undoubtedly continue to be important because it is not limited by cellular abundances or the molecular weight of the target molecules. It is crucial when high-throughput is required, for instance, to elucidate how various small molecules bind to a drug target.[50] Recent developments in multi-dataset crystallographic analyses have even improved and accelerated the identification of ligand binding and structural events.[51] In combination with functional assays, *in vitro* structural studies were and continue to be essential to understand the molecular details of a process. They, for example, provide high-resolution information on active sites of an enzyme and allow to directly study the consequences of mutations or modulatory ligands. Structures obtained *in vitro* are also often a prerequisite for *in situ* structural analysis. This is particularly relevant for template matching, a method to identify structural signatures of known assemblies inside of cells (see below).

UNDERSTANDING MOLECULAR ACTIVITIES INSIDE OF CELLS

To ultimately understand how cells work, we need techniques that monitor molecular activities inside cells in a way that enables us to quantify structural dynamics and local concentrations. For any class of molecules, it will be important to quantify cellular content with a very high spatial and temporal resolution. Developing such techniques remains challenging although recent imaging and omics techniques have brought us closer to this goal.

Quantitative and spatial measurements by cryo-ET: Moving away from averaging

Cryo-electron tomography (cryo-ET) is a versatile method to visualize the molecular interior of cells. Here, the fraction of a cell contained in each tomogram is revealed in its entirety, primarily as a 3D distribution of electron optical density. The respective data can be analyzed in manifold ways, first and foremost, by segmentation of easily identifiable features such as ribosomes or membranes. Such analysis techniques are well suited to, e.g., analyze the local curvature and thickness of membranes, to trace the trajectory of mRNA molecules decorated with ribosomes or to identify microtubules or actin filaments. In combination with AI-based image analysis techniques, segmentation becomes objective and quantitative.[52–56] Further minimization of false negative detections and human interference with data analysis will be essential to continuously increase the accuracy of capturing cellular processes.

Subtomogram averaging in cryo-ET is based on correlation averaging and classification of structural features observed

repetitively in tomograms, thereby enabling the high-resolution structure determination of large macromolecular complexes and their functional states. Recently, averaging in close to atomic detail has become possible, enabling high-resolution structural biology inside of cells. This, for example, allowed the elucidation of the substrate-processing species distribution of 26S proteasomes[47] or of the functional states of the ribosome during the translation elongation cycle.[45,46,57] This approach, however, requires that enough particles populate a given functional state.

Therefore, it will be crucial to analyze tomographic content beyond averaging techniques. One possible solution is template matching, which is a method that systematically scans tomographic content for the signature of reference structures. The resulting cross-correlation volume can be statistically analyzed to assess whether the respective reference structure has been observed at a given position and orientation in a cellular tomogram,[58] thus providing information on conformational state and location at a single-molecule level in cells. Among the attempts to experimentally implement this so-called visual proteomics concept, which aims to generate molecular atlases that describe the cellular content in high detail, template matching has been considered a potent approach. However, thus far, template matching has been suffering from size and abundance limitations[59,60] and has rarely been applied to particles other than ribosomes or proteasomes. Conceptually, there is no strict physical limit that would prevent the recognition of smaller and less abundant features, such as the molecular weight limit of single-particle EM, which has been pushed to ever smaller entities. A recent preprint reports systematically optimized template-matching parameters for tomographic data generated with the latest generation of hardware that was able to detect very subtle conformational changes and particles as small as individual tubulin segments.[61] In the future, the template-free identification of recurrent structural features[62] may facilitate an entirely unbiased annotation of cryo-electron tomograms. Also, upcoming machine learning (ML) methods may address this issue.[55] At a first glance, the resulting molecular atlases may appear descriptive. They, however, can be mined to deduce concepts of functional organization in space, which otherwise would not be apparent, and may allow to formulate new types of hypotheses. Examples for this could include the orientation of the catalytic center of different macromolecular complexes toward each other, alignment of molecules at membrane deformations or aggregates, local clustering of conformational states, and many more.

Regardless of all these strengths, the cryo-ET technology also suffers a few limitations. Genetic labeling remains challenging. Although some solutions that rely, e.g., on the fusion of recognizable shapes or clusters of dense material were proposed, they are not generically applicable to proteins of any size or expression level.[63] The fact that the fraction of a cell that can be analyzed by cryo-ET at once is inherently small also limits the analyses. Finally, the inability to resolve biological processes directly in time, simply because samples have to be flash frozen, is a severe drawback. Moving forward, visualizing the molecular sociology of cells comprehensively and in high detail will require a combination of cryo-ET with complementary techniques that will allow to, e.g., indirectly stage biological processes in time to trigger rare events or to identify molecular content. Those will include, but not be limited to, fluorescence-based super-resolution imaging, subcellular omics, single-molecule approaches, AI-based targeting of rare subcellular features, optogenetic control of transient biological events or local confinement, and microfluidic sorting of cellular populations.

Fluorescence-based super-resolution imaging

Recent methods pushed the resolution of fluorescence-based super-resolution microscopy (SRM) techniques to subnanometer precision.[12,64,65] Thus, SRM techniques are well capable of resolving individual macromolecules and their domains, although not with the atomic precision that is achieved with many structural biology techniques. In combination with cryo-ET, these advances are paving the way for structural biology inside cells. The complementary strengths of SRM with respect to cryo-ET are its capability of visualizing unstructured molecules at very high detail, providing quantitative data, such as molecule numbers, stoichiometries, or distances and allowing time-resolved analyses inside living cells.

Minimal fluorescence photon fluxes (MINFLUX) nanoscopy, for example, achieves an unprecedented three-dimensional spatial resolution of 2 to 3 nm, allowing true molecular-scale fluorescence imaging[66] of different subcellular structures, including the NPC[67] and the mitochondrial contact site and cristae organizing system (MICOS) complex.[68] Different adaptations allow multiplexing and quantitative imaging of multiple targets simultaneously,[67,69,70] as well as single-molecule spatiotemporal tracking[71] in living cells.[12,64] Development of MINSTED nanoscopy based on MINFLUX combined with STED (stimulated emission depletion) further increases spatial precision and provides structural information for target macromolecules on the scale of a single amino acid.[72] Using the NPC as a reference yielded an astonishing localization precision of 2.3 Å. This technology also revealed that Mic60 proteins in the mitochondrial inner membrane of human cells form ring-like assemblies at junctions of cristae. Another approach to enhance resolution is resolution enhancement by sequential imaging (RESI),[65] which can separate localizations in very close proximity. This method uses Exchange-PAINT to label neighboring molecules with different photo-switchable tags by orthogonal DNA barcoding.

Despite these advances, major technical challenges of fluorescence nanoscopy remain, e.g., the very small field of view, the long timescales required for image acquisition, and the lack of multiplexing in live cells. Initial attempts to overcome these issues have been made; for example, short-distance self-quenching in fluorophore dimers was used to reduce background fluorescence signal while increasing the photon budget in the bound state by almost 2-fold.[73] Transient adapter-mediated switching for high-throughput 3D DNA-PAINT (FLASH-PAINT) allows imaging of a nearly unlimited number of target sites, revealing the organization of cilia and Golgi in unprecedented detail.[74] To achieve the highest spatial precision, low background signal and endogenous (multiplexed) labeling, smaller fluorophores, and new labeling strategies have to be developed. With those developments, novel approaches such as using RNA-FISH coupled with FLASH-PAINT and MINFLUX could make addressing challenges such as resolving the

ultrastructure of membrane-less organelles or the conformation of mRNPs feasible (Box 1). Moreover, fluorescence lifetime imaging of fluorescence resonance energy transfer (FLIM-FRET) can measure the conformation of IDRs, and when combined with genetic code expansion technologies, FRET pairs can be introduced at multiple sites to generate a conformational map of IDRs directly inside of cells.[75] These and additional approaches, such as fluorescence nanoscopy with expansion microscopy (ExM)[76,77] or SRM with cryo-ET, will further advance the technological possibilities to observe the molecular details of subcellular organization changes in living cells with high quantitative power, thereby bringing us closer to observing molecules at work.

Subcellular and spatial omics

The term omics refers to the investigation of the sum of specific classes of biomolecules, e.g., proteins, lipids, mRNAs, or metabolites within cells. Bulk omics approaches are ideally suited to identify and quantify biomolecules and their variants in specific cellular states and to quantitatively describe molecular processes. However, bulk omics methods cannot capture the dynamic spatiotemporal organization of subcellular architecture because cellular context and spatial information are lost during cell lysis.

More recently developed spatial omics techniques attempt to circumvent this limitation,[78] but a global and precise picture of cellular organization is not yet possible. Current spatial omics applications are either genome-wide or targeted. Genome-wide approaches profile the entire molecular content of single cells within their native context, with spatial information from the location within the tissue, but rarely provide subcellular resolution. Targeted omics approaches, on the other hand, profile the molecular content of a specific cellular segment, a subcellular structure, or organelle *in situ* in great detail but only provide a limited cellular snapshot. The cellular segments have to be biochemically purified or physically separated, e.g., by laser capture microdissection (LCM), followed by RNA sequencing (RNA-seq) or MS in order to identify and quantify their content. Alternatively, if purification is not possible, the content of a specific cellular segment can be obtained by proximity labeling of RNAs or proteins, using, e.g., APEX2 fused to marker proteins of the respective subcellular structures, followed by purification and sequencing or MS of the labeled components.[79] While targeted RNA-seq will capture all expressed RNAs in the respective segment, the detection of local proteomes by MS is very limited due to the lack of signal amplification procedures. Lowly abundant proteins, splice isoforms, cleavage products, and PTMs remain challenging to detect. Hence, further developments of MS technologies to enhance sensitivity and resolution are required for this method to catch up with other omics technologies.

Alternative approaches such as translation imaging in turn allow tracking of global translation and enable the quantification of nascent proteomes in specific subcellular compartments, e.g., at the synapse.[80–82] Cross-linking MS (XL-MS), on the other hand, had originally been limited to *in vitro* systems of limited complexity but was recently applied to more complex systems and in combination with subcellular fractionation and affinity purification of cross-linked peptides, is capable of charting local variations of protein interactions.[83–85] The technique, however, remains less comprehensive in terms of coverage compared with other proteomics techniques.

In contrast, techniques to determine the cellular distribution of different lipid species at high resolution are still entirely lacking. Although imaging MS is capable of resolving local lipid composition, the present resolution is hardly subcellular.[86] Much of our knowledge about local lipid composition is either based on invasive subcellular fractionation experiments or indirectly inferred by perturbation experiments of lipid regulators. A method to locally quantify lipid molecules inside of cells could, however, facilitate novel insights into subcellular organization.

CELLULAR SELF-ORGANIZATION AND ITS IMPACT ON MACROMOLECULAR ASSEMBLY AND ACTIVITY

Which factors modulate the function of macromolecular assemblies in the cellular context? This question has always been central for cell biologists, but due to the lack of appropriate tools, it has been somewhat neglected by structural biologists. In the following section, we will illustrate some of the principles that organize cells and contextually modulate molecular function and structure of macromolecules but that are themselves structurally not well understood. Thus, they comprise high-impact targets toward understanding cellular self-organization. The discussed organizing factors are selected in an exemplifying manner; various others such as the cytoskeleton or PTMs, although equally important, are not covered due to space limitations.

Local confinement

Local confinement facilitates the contextual regulation of cellular function.[87,88] Molecules that form a functional module have to interact with each other at the right time and at the right place. The cell is densely packed with biomolecules, and for each specific protein-protein interaction, many non-specific interactions with lower affinity compete with the respective binding partners. This becomes a considerable challenge with increasing genomic complexity.[89] Thus, the generation of locally confined and specialized molecular communities reduces complexity and establishes a dedicated biophysical environment that is suitable for the proper assembly and function of macromolecules. It may locally expose interaction partners and promote their binding or prevent promiscuous interactions by local exclusion.

We are only beginning to understand the self-organizing principles of local confinement inside cells, and the respective conditions are challenging to recapitulate *in vitro*. Although local confinement by membranes, i.e., in organelles, has long been known, we understand little how these membranes self-organize. In recent years, much attention has been dedicated to membrane-less organelles and whether concepts from polymer physics or multivalent binding based on biochemical entities are well suited to explain their biogenesis.[35,90,91] Despite the still pending final verdict on this issue, they can add to explaining local confinement of functionally dedicated, local molecular communities. To make things even more complex, cells may combine this mode of local confinement with additional

self-organizing principles. For example, alternative splicing factors are sequestered into nuclear bodies for release in response to environmental cues;[92,93] local translation can ensure the biogenesis of selected proteins directly in a suitable environment, e.g., the neuronal synapse,[94,95] while co-translational chaperoning and assembly warrant that nascent chains of proteins already engage in specific interactions before being released from the ribosome.[96] Another self-organizing component that influences local confinement in a complex manner is termed molecular rulers.

Molecular rulers

In their most simple manifestation, molecular rulers define the distance between two molecular entities. This can be realized by two short linear motifs (SLiMs) separated by a linker or, similarly, two protein-binding motifs on an RNA. Molecular rulers can also organize protein complexes. For example, SLiMs in so-called ''linker nucleoporins'' organize scaffold nucleoporins in space, ensure their stoichiometric assembly, and act as sensors for the correct assembly of subcomplexes that form early during the NPC biogenesis process.[97,98] Molecular rulers may furthermore organize co-translational biogenesis,[99] where they define the exact sequence and timing of co-translational interaction events that occur, while the respective domains appear from the exit tunnel of the ribosome.[99] Promiscuously interacting domains, such as coiled coils, may be C-terminally encoded to ensure that the respective native interaction partners have already been recruited to the nascent chain by other more specific assembly motifs. In addition, the codon usage of a molecular ruler may be as equally important as its length because this defines the necessary timing. Another example is the long non-coding RNA that often plays architectural role because these RNAs organize chromatin in 3D, recruit interaction partners, or scaffold nuclear bodies.[100] Architectural RNAs (arcRNAs) may adopt a specific 3D conformation inside membraneless organelles together with bound RNA-binding proteins; for example, the lncRNA *NEAT1* acts as a scaffold for paraspeckles. NEAT1 length and the specific protein-binding sites within it determine the dimensions and the inner organization of the respective condensates in a manner similar to a molecular ruler, in this context, referred to as micellization of block co-polymers.[101] We anticipate that focusing future structural investigations on molecular rulers will be rewarding. This problem is very approachable—one essentially needs to solve structures of the interaction pairs and understand the biophysical behavior of the linker in between.

Self-organization of membranes

Cellular membranes are highly complex systems characterized by a heterogeneous and dynamic composition. One of the mysteries of membrane biology is why thousands of different lipid species exist, and lipid composition is strictly regulated in space and time.[102] Lipids locally influence the biophysical properties of membranes, such as their fluidity, and thus are critical for the function of membrane-bound organelles.[103] Local lipid composition is regulated not only by lipid synthetases and degraders that control the availability of specific lipid species but also by dynamic organelle contacts and lipid transporters that spatially affect lipid availability. Lipids and proteins can self-organize in cell membranes to form nanodomains to concentrate specific proteins and enhance biochemical reactions.[104,105] Recent evidence supports that such mobile nanodomains exist in cell membranes and can help form protein nanoclusters.[106]

Technical limitations force us to study membrane proteins and lipid bilayers separately. Conceptionally this separation is problematic because both entities are interdependent. Lipids and the bilayer can induce conformational changes in proteins or promote the assembly of protein complexes.[107] Vice versa, proteins can lead to extensive membrane-remodeling processes,[108–110] for example, the endomembrane system, mitochondria and chloroplasts in eukaryotes, but also the cell envelope of gram-negative bacteria, can form a very complex membrane architecture that is constantly remodeled and maintained and, in some cases, even formed almost entirely from scratch. Basic design elements such as membrane tubes, budding of trafficking vesicles, or membrane scission are repurposed in manifold ways. An intriguing example is a set of two or more highly parallel membranes termed stacked sheets that are found in mitochondria, the endoplasmic reticulum (ER) and the nuclear envelope, the Golgi apparatus, the rod outer segment, and annulate lamellae. Autophagosomes form membrane sheets from scratch that subsequently bend to engulf cellular debris.[111] In some cells, such membrane sheets even form a helically stacked superstructure.[112] We still do not know the self-assembly mechanisms of these structures.

Until recently, the architecture of such organelles was thought to be shaped mainly by membrane-interacting proteins. Yet, structural analyses of Bin/Amphiphysin/Rvs (BAR) domain, coatomer, clathrin, endosomal sorting complex required for transport (ESCRT), or nucleoporin proteins *in vitro* have not allowed us to predict or manipulate the subcellular outline of the endomembrane system or mitochondria. It has become clear that additional principles of self-organization play a crucial role in shaping organelle architecture.[113–115]

Molecular rulers may define membrane distances, which is the case for nuclear membrane-localized LINC complexes. These complexes consist of SUN and KASH, two transmembrane proteins that shake hands in the lumen of the nuclear envelope, whereby the linker length between transmembrane and interaction domains defines the maximal distance between the inner and outer nuclear membranes.[116] However, the overall organelle shape, more specifically, the surface-to-volume ratio also must be tightly controlled. This could be done by regulating the osmolarity, or similarly, by controlling the number of biomolecules inside of a given organelle. Such concepts appear to be well suited to explain the transformation of a vesicle into a membrane sheet and vice versa, but the respective regulatory circuits remain ill-defined. Together with lipid availability, changing osmolarity regulates membrane tension, the force per unit length acting on the cross-section of a membrane, which defines how much it may be deformed.[117,118] Evidently, these three parameters are intertwined and will be regulated by proteins that make organelle contacts, transport ions and biomolecules across membranes, regulate lipid availability, or deform membranes.

Such a system is in turn capable of controlling protein structure as illustrated by recent work on the NPC, which dilates

and constricts in response to osmotic shock and the consequent changes in membrane tension.[119] The resulting large conformational changes have various functional implications that we are just beginning to understand. Other well-studied examples are mechanosensitive channels such as piezo.[120] Moreover, recent work, e.g., on BAR domains coupled to IDRs, implies that molecular crowding proximal to membrane is a major driver of membrane curvature formation.[121] Similarly, molecular condensation may also be involved in the budding of trafficking vesicles.[122] This appears intuitive when considering that coalescence is coupled to a membrane association activity. The formation of a structured protein coat may thus be a consequence, rather than the driver of membrane curvature.

To understand such complex architectural traits, precise quantitative and spatial data are needed. Perturbation and synthetic biology experiments that abolish the respective architectures or generate them from scratch may enlighten the complex circuits of intertwined, self-organizing principles that govern subcellular organization.

NEXT-GENERATION STRUCTURAL CELL BIOLOGY

A common aim should be that next-generation structural cell biology targets complex biological processes in their cellular context. Technological developments alone, as the ones discussed so far, will not be sufficient to discover and conceptualize the principles of molecular self-organization. New theoretical concepts and frameworks that capture dynamics and complexity will also have to be developed and adopted. Structural biology should systematically build on information and concepts not only from other disciplines, such as developmental, systems, and cell biology, but also from physics, information theory, computational sciences, and the science of complex systems. As a notable example, soft-matter physics concepts such as phase transitions, criticality, or scaling exponent are proving necessary to conceptualize the rich phenomenology of biomolecular condensates and IDRs.[31,123–125]

What limits our present-day conceptualization? Without a doubt, cell biological models are context-aware and may capture self-organizing principles of subcellular architecture. However, cell biological processes are often depicted in 2D, e.g., in flowcharts of cellular pathways or interaction maps. While these are useful for identifying individual factors and their functional interactions in given pathways, they neglect the complex spatial context of a crowded cell and thus miss fundamental layers of regulation. They struggle to recapitulate the complex spatial and dynamical context *in situ* or how interactions emerge from the interplay of many multi-valent (many-body) molecular interactions. For example, confinement within organelles, condensates, and lateral domains in membranes lead to varying local concentration of molecules with distinct functional consequences, from polymer packing that controls RNA and DNA accessibility to the formation of cellular barriers that can protect macromolecules by local exclusion.

In structural biology, atomically resolved 3D models of macromolecular complexes are routine. However, despite being 3D, these traditional models are also limited because they only depict distinct functional states of isolated macromolecular complexes and not the subcellular organization that relies on compositional, structural, and dynamic complexity. They also lack information about transitions between states and the influence of neighboring molecules or PTMs and do not integrate functional metadata.

Which new concepts should the next generation of structural cell biologists aim to uncover? To be useful, a principle should serve as a guide to understanding observations, planning experiments, and condensing a rich phenomenology. It should also change the way we think. For example, the laws of General Relativity, the most accurate theory of gravity, can inspire structural biology. It taught us that physical systems do not evolve in a static spatial and temporal background but that the background, the spacetime, is itself an integral part of the physical system and that matter and spacetime are locked in a constant dialog. J.A. Wheeler famously summarized the core principle of the theory by saying, "Spacetime tells matter how to move; matter tells spacetime how to curve."[126] This can serve as an analogy for principles of cellular organizations. Proteins and other biomolecules do not function in a static background. The molecules and their background—be it a membrane, a confined compartment, or a complex solution—are part of the same dynamic system. They are engaged in a dialog. Thinking of membrane-protein interactions, for example, we could be tempted to follow Wheeler in saying, "Membranes tell proteins where to go, and proteins tell membranes how to curve." This sentence provides the gist of many diverse phenomena, and it summarizes the fact that membranes and proteins control each other in an interdependent way. It furthermore gives a useful mental image to understand observations and formulate hypotheses. Of course, not all membrane proteins reshape membranes, but they act on membranes, and all membranes respond in return. The concept of the "fingerprint" of a specific membrane protein[127] is one attempt to quantify how structural features determine the strength of the protein's actions on a membrane and therefore what distinguishes a protein that reshapes a membrane from one that does not. With an increasingly more precise definition of these concepts, we will be able to integrate them into existing theories[128,129] that describe how membranes are organized on an organellar scale as a function of their composition. In this way, we may obtain a quantitative theory—in the form of equations—that links atomistic details in proteins to large reshaping of entire organelles.

Digital twins of cellular segments

How can we progress toward taming cellular complexity? In initial attempts, 3D models of subcellular segments that incorporate all available knowledge from structural, morphological, and omics measurements, as well as biochemical experiments, have been put forward. For example, these have revealed that a synaptic vesicle consists of a similar number of molecular components as a ribosome[130] although both objects are organized in a fundamentally different way. Ribosome architecture relies on folded protein and RNA interfaces; the synaptic vesicles is a result of self-organization, membrane-protein interactions, and molecular sorting. Further steps toward fully representative models have now been made. For instance, recent models of an autophagic vesicles or a nuclear pore[131,132] include explicit

membranes and other dynamic components such as intrinsically disordered proteins (Figure 2). These first comprehensive models contain the molecular content of the respective subcellular segment. They are best conceived as high-fidelity structural models in which the structure of a given protein complex has been used as the basis but has been expanded to the best possible scientific knowledge with additional information of non-regularly structured content. This could be knowledge about biological membranes and lipids, solvents, intrinsically disordered stretches, and interacting surrounding molecules and spatial boundaries. Alternatively, cryo-electron tomograms, which capture the respective subcellular structure, could be fed into the molecular modeling framework, annotated by template matching and enriched with additional molecular detail much like a 3D jigsaw puzzle. Importantly, such high-fidelity structural models can be used as input for MD simulation framework to explore their dynamics.

Such dynamic virtual equivalents of cellular objects, which ideally mirror the respective cellular behavior, are best described by the digital twin concept,[133] which is used in other fields, such as urban planning or construction.[134] A "digital twin" is defined as a virtual representation of a real-world object ("physical twin") that allows the operator to run virtual experiments without real-world constraints, such as object size or gravity. Yet, it can also be more than a static virtual copy and can contain algorithms that allow simulation of its behavior in response to perturbations.

A digital twin could be conceived in different layers of complexity. Atomic models of macromolecular complexes, routinely used in structural biology, are accurate representations of molecular objects that can be explored in a virtual reality; therefore, they could be considered a basic digital twin of a macromolecule. As discussed above, structural models may be further enriched with complementary molecular information, thus resulting in a high-fidelity version of the model that considers cellular context and represents a digital twin of a macromolecular complex with its surrounding environment. This could be done, e.g., by adding membranes to the structure of a membrane protein, by placing all molecular components of a vesicle into a virtual box, or by annotating the cryo-electron tomogram of a subcellular segment with additional molecular detail, e.g., by template matching. Regardless of which specific approach is applied, the meaningful integration of different types of data will be the important common feature. Finally, digital twins should also become dynamic. Their virtual reality should progress over time and respond to external perturbations. We thus envision that a digital twin of a subcellular segment would be an integrated pipeline of datasets, computer simulations, and ML methods to study the system *in silico* to better understand and discover *in situ* phenomena (Figure 3). Digital twins could be constrained by functional meta data to predict cellular behavior more accurately. Predictions made by digital twins can in turn guide experimental design and inspire synthetic biology approaches.

High-fidelity structural models

An important first step toward the digital twin of subcellular segments is the integration of many different technologies. There are different approaches to data integration, whereby the combination of some complementary techniques has become particularly popular. This is apparent for the combination of NMR with small-angle X-ray scattering (SAXS), or X-ray crystallography with cryo-EM, where information about the structure of individual components is combined to obtain the overall assembly.[135–137] Some techniques have turned out to be important facilitators and accelerators of data integration because they elucidate how molecular components interact or bridge across scales. This is true for XL-MS that identifies proximate residues in proteins that are frequently used as spatial restraints during modeling.[138] Cryo-ET, in combination with subtomogram averaging, provides moderately resolved maps of macromolecular assemblies inside cells, which have become popular as a modeling frameworks that define overall shapes and dimensions.[139] Software frameworks such as the integrative modeling platform[140] or Assembline[139] are routinely used. They consider various types of data, including high-resolution structures, shapes, interaction data, and other spatial constraints. They sample an exhaustive ensemble of possible solutions and identify those that explain the given experimental datasets best, whereby very complex structural assemblies can be elucidated. Meanwhile, such frameworks have been successfully applied to various macromolecular assemblies.[141]

Data integration for structural modeling benefits enormously from AI-based structure prediction. New algorithms such as AlphaFold not only unravel protein folds but also predict their interfaces and thus can fill in gaps of structural knowledge, such as by bridging across subcomplexes.[132] AlphaFold predicts 3D configurations of proteins from their amino acid sequences with a high accuracy, matching or surpassing some experimental results.[6,142,143] The algorithm capitalizes on the wealth of publicly available information in both sequence and structural space, in addition to considering fundamental biophysical and chemical principles of how amino acids engage with each other. AI-based structure prediction has achieved remarkable success, but it also has clear limitations. Some of the most important ones are not capturing alternative conformations, dynamics, or the effect of mutations.[144]

The generation of the NPC structure exemplifies the strength of integrative approaches to deal with very large biological systems en route toward generating digital twins of subcellular segments. NPCs consist of 1,000 individual protein building blocks. The elucidation of individual nucleoporin folds by traditional *in vitro* structural biology techniques and AI-based prediction provided the pieces of the puzzle, while tomographic analyses inside cells revealed the overall outline of the nuclear pore. This then served as framework for further computational structural modeling, while interaction studies and XL-MS techniques elucidated interfaces and spatial proximity of the individual components. Finally, when all of these data were jointly analyzed by integrative modeling, the overall molecular picture of the nuclear pore architecture was revealed and could subsequently be refined step by step, analogous to progressively solving a puzzle.[49,145–147] This model has been further complemented with membranes based on compositional knowledge about lipids and membrane shapes.[132] IDR-containing nucleoporins were added based on the knowledge about their anchoring sites to the scaffold,

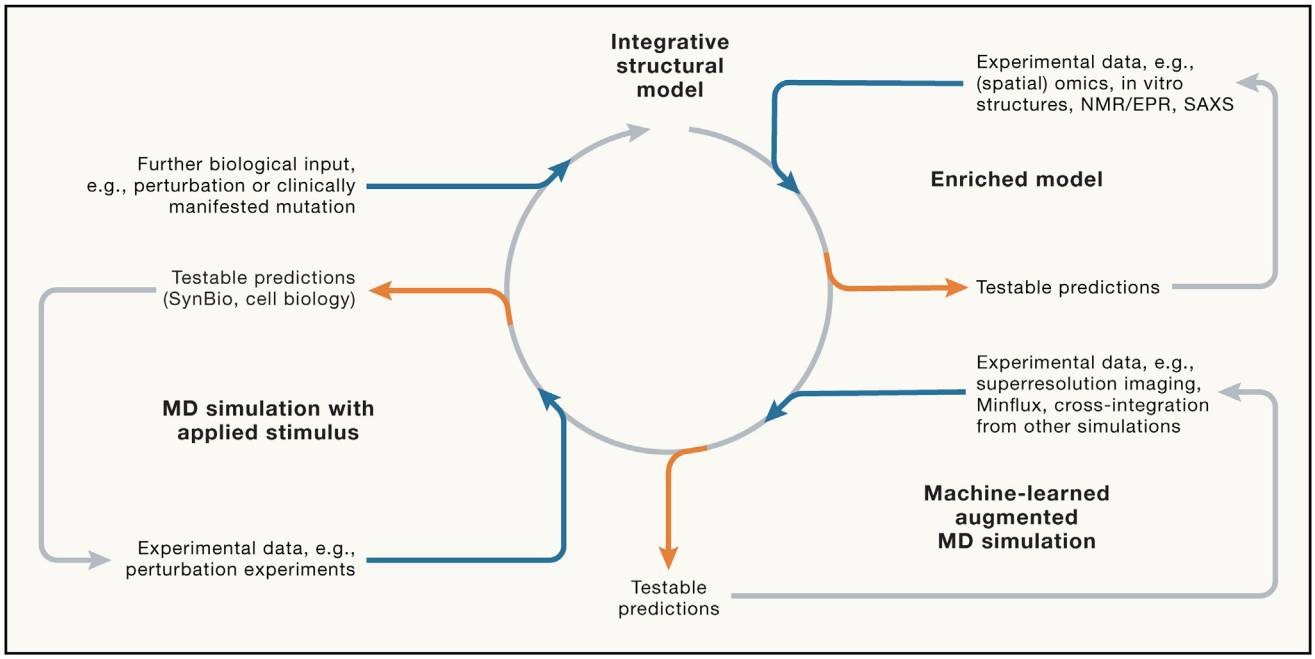

Figure 3. Scheme illustrating a workflow for generating digital twins of subcellular segments

Integrative structural models of complex subcellular segments will be complemented with additional information about cellular context, derived, e.g., from omics or fluorescence-based techniques. The resulting high-fidelity models will serve as input for MD simulation frameworks to explore their physical properties and dynamics. The simulations will be augmented with machine learning methods trained on experimental data. Once stimuli are applied the digital twin will predict a cellular response considering a complex parametric space. These predictions can be experimentally tested.

resulting in a "high-fidelity model." Fluorescence-based methods were used to explore their conformation inside of cells.[75] Today, the respective structural models approach completeness (Figure 2), thus enabling MD simulations of NPCs.[75,132] Such high-fidelity models are heterogeneous in terms of resolution and accuracy. The individual interfaces stem either from high-resolution structural analysis or AI-based prediction and are resolved with differing accuracy, but certainly to the subnanometer level. The exact orientations and positions of the individual proteins are the result of integrative modeling and not precise to the Å level but allow conceptualizing the overall architecture. Thereby, the position and conformation of dynamic components such as lipids or IDRs are not based on traditional structure determination techniques but are added based on the best of present knowledge, stemming, e.g., from omics or FRET measurements. They are still informative, e.g., about the spatial range such dynamic components may have (Figure 2).

In principle, these models can also become multilayered and consider multiple conformations, splicing isoforms, PTMs, local lipid composition, or variable stoichiometries, although this is not yet routinely performed. One challenge thereby is posed by the structural data that constrain a given model and that may have been generated for one isoform of a macromolecule or complex, which, however, may not necessarily be transferable to other forms. Another issue is the meaningful integration of time as an additional axis for biological processes. Overall, however, the integration of data from multiple different technologies will be critical to further define the molecular details of cellular function. Initial options dealing with this challenge have been reported and tested; yet, these will need to be continuously refined and further expanded to incorporate technological advances and novel findings.

MD simulation

Among the existing modeling frameworks, MD simulations may be the closest to the ideal of a digital twin of a cellular segment, because they attempt to comprehensively simulate the molecular content of a given biological system at atomic resolution. This approach has matured from a prototype to a standard toolkit capable of generating hypotheses and making discoveries, as if it were a "computational microscope."[148] In MD simulations, the forces acting between all atoms are modeled, and high-performance supercomputers are used to numerically solve Newton's second law of motion and predict the dynamics of biomolecules.[149] They produce trajectories that sample the molecular structural dynamics and can also show how molecules reorganize between alternative structural organizations (Figure 4A). MD simulations are ultimately based on the physics and chemistry that accurately describe how atoms interact and how these interactions determine the dynamics of molecules.

Yet, the approach has limitations that remain to be addressed. The forces acting between individual atoms are modeled in so-called "force fields." No general force field exists, and the specific system (proteins, lipids, inorganic molecules, etc.) determines which force field to use. The accuracy of MD simulations depends on the accuracy of the force field used. Protein force fields were optimized to describe single-domain globular proteins.[156] Consequently, these force fields were not as accurate when used to

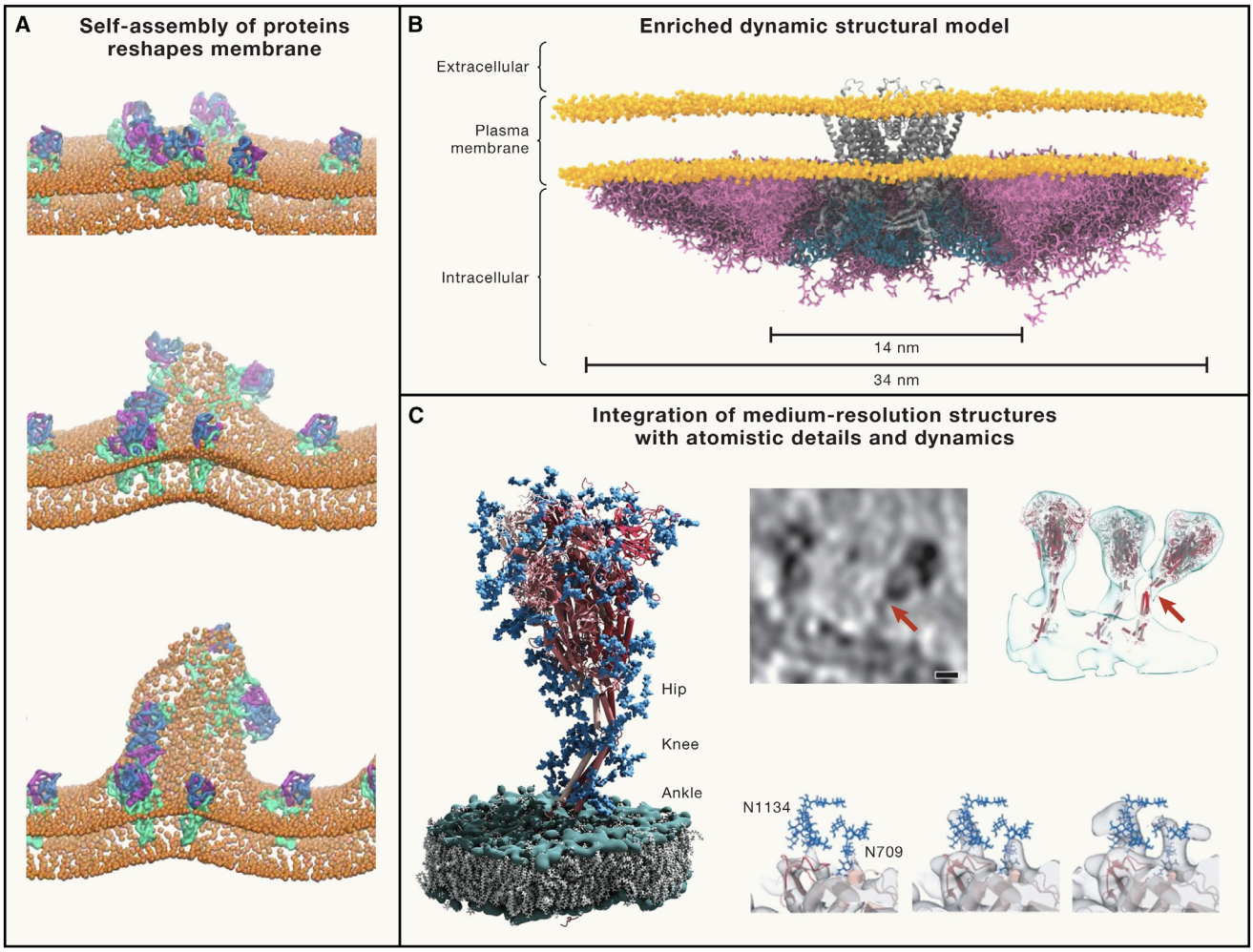

Figure 4. Molecular dynamics simulations of complex cellular systems

The given examples demonstrate the scope that MD simulations of enriched structural models can provide.

(A) Snapshots of an MD simulation of FAM143B in membrane that predicts clustering of FAM143B over time and consequent budding of endoplasmic reticulum membranes during ER-phagy, which was experimentally confirmed.[150,151] Arrowheads indicate progression in time.

(B) MD simulations of an integrative structural model of tetrameric TRPV4 with the previously unresolved four 150 amino-acid-long N-terminal intrinsically disordered regions (IDRs). Superimposed IDR extensions (pink) form a halo around the previously known structured part of TRPV4 (gray and cyan). Coarse-grained MD simulations were carried out within a bilayer (yellow). The simulations uncovered that the IDRs expand the intracellular channel surface from 14 to 34 nm and suggest that the IDRs exert a pulling force on the pore-forming domains through membrane interactions.[152]

(C) Snapshot of an MD simulation of full-length spike protein (red) with explicit glycosylation sites (blue) and bilayer (gray). These simulations predicted flexibility at three hinge regions (hip, knee, and ankle) that were confirmed using cryoelectron tomography of viral particles (top inset). The predicted flexibility of the glycosylation sites was confirmed using subtomogram averaging (bottom inset) and led to the prediction of the surface epitopes approachable by neutralizing antibodies.[153–155]

sample the dynamics of disordered domains and required some recalibration.[157] Similarly, force fields used for simulating DNA and RNA do not yet match the quality of protein force fields. MD simulations are also limited by the spatial and temporal timescales that can be sampled. A virtual simulation box can contain a single protein surrounded by water or a large membrane patch containing many proteins and other molecules. The larger the box, the shorter the covered timescales. Typically, timescales range between micro- and milliseconds, sufficient for observing events such as small protein folding, ligand binding, or short-timescale conformational changes. Yet, longer biological processes, such as large protein folding or protein-protein interactions, which may occur over milliseconds to seconds, are often beyond the reach of traditional MD simulations. Methods to sample larger systems for more extended times are developed by building on physical insight[158] or by trading a smaller spatial resolution for the ability to sample longer timescales.[159] The latter approach is called coarse-graining and has, in recent years, enabled simulations of sub-organellar systems, such as the NPC.[132]

Are these MD simulations precise enough to conceptualize complex biomolecular systems? As discussed for structural models, it depends on the scientific question we are asking.[149] MD simulations can help in characterizing individual proteins at atomic level resolution with sufficient accuracy to enable drug

design. They can estimate the free energy cost to bend or compress a membrane of a given composition—important biophysical parameters—from the lipids dynamics.[160] The results are often within the experimental uncertainties. MD simulations excel in providing mechanistic hypothesis, which can be valuable even if other predictions are not accurate. For instance, a simulation of a conformational change could provide an inaccurate free energy estimate and, at the same time, an accurate sequence of structural intermediates along the conformational change. New ML and AI-based methods such as AlphaFold can synergize with MD simulations and enable us to assess previously uncharacterized molecules. Integrating ML- and physics-based simulations is emerging as a paradigm,[158] thus increasing sampling speed for molecular structures.

MD simulations can be supplemented with experimental data and used as a framework to integrate different types of information. An intriguing example, where IDRs play a crucial role, is the TRP vanilloid channel (TRPV) 4, which is involved in thermo- and osmosensing (Figure 4B). The 150 amino-acid-long N-terminal IDR remained unresolved in traditional structural models; yet, the integrative structural biology approach using NMR, SAXS, tryptophan fluorescence spectroscopy, XL-MS, and HDX-MS combined with atomistic MD simulations finally led to a structural model including the IDRs.[152] This model led to the now testable prediction that the IDR increases the channel dimension at the plasma membrane from a diameter of 14 nm to 34 nm and that the IDR interacts with PIP_2 in the plasma membrane, which provides a pulling force leading to channel sensitization.

Increasingly, MD simulations have become integral parts of interdisciplinary approaches, where experiments, theory, and simulations complement each other to provide more detailed mechanistic and structural insights. The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pandemic was an opportunity to showcase the enabling features of this approach. Only a few months after its onset, the integration of *in situ* cryo-ET and MD simulations provided accurate atomistic models of the SARS-CoV-2 spike protein's structural dynamics[153] (Figure 4C). The integrative approach yielded more than the sum of its parts: the MD simulation could add atomistic details and dynamics to medium-resolution cryo-ET structures, while the structural experiment could validate the MD models and reveal the organization of multiple spikes on the viral capsid at the supramolecular level. Notably, the spike protein of SARS-CoV-2 is highly glycosylated. While it had been known that the positioning of glycosylation sites on viral surface proteins may allow viruses to dodge neutralizing antibodies,[161,162] the MD simulations were instrumental in understanding the exact conformational movements of glycans that create surface patches where the spike is protected from antibodies, thus forming a glycan shield that hinders an immune response. With such striking examples, it has become evident that integrated and complementary experimental studies are required to advance the molecular understanding of how macromolecular complexes function in cells.

MD simulations are moving toward the subcellular and cellular scale. The first pioneering molecular simulations of viruses or entire cells were done, and more work will undoubtedly be undertaken during the next decade.[163,164] However, sampling interesting timescales and using an accurate biochemical description will be challenging. In addition, one outstanding problem remains: how can we obtain a high-fidelity *in silico* representation of subcellular compartments with incomplete knowledge of molecular and cellular biology? We still ignore many details—and maybe more general aspects—of the cellular content, pathways, and mechanisms. An MD simulation that does not contain a crucial component of a cellular pathway will not produce an accurate description. One promising strategy is to systematically augment MD simulations with ML methods trained on experimental data.[165] In other words, one would employ an MD simulation for the known part of the model and use neural network-based approaches to learn the unknown part from experimental data. Ever-increasing computational power, improvements to force fields, and more meaningful coarse-graining, together with an integration of physics-based simulations with data-driven approaches, will make digital twins standard tools to conceptualize complex phenomena and hypothesis generators in cellular biology.

Conclusions

In the past decades, structural biology has evolved, with the focus shifting from high-resolution and atomistic details of single molecules toward the characterization of larger complexes and assemblies and more recently, their analysis in a native environment—*in situ*. A definition of molecular function *in situ* requires taking into account all aspects of the cellular context and integrative approaches combining different technologies that capture additional properties such as structural dynamics or physical forces. As the field progresses in this direction, novel types of models and concepts will be needed to incorporate and visualize the diverse data types.

Once properly built and tuned, digital twins of cellular segments should be able to capture synergistic effects of molecular functions and faithfully predict complex cellular responses. Example problems could be membrane deformation by protein clustering or crowding (Figure 4A), the conformation of mRNPs, the benefit of a local environment for protein folding, how forces act on subcellular architecture, or whether a small molecule will be enriched in membrane-less organelles or not. Importantly, the information gained will not be limited to the wild-type protein in an unperturbed cellular context, but such models could be overlaid with clinically manifested mutations or changing environmental conditions. Casting biological questions into algorithms will be an important skill for future cell and structural biologists, given the wealth of available data and the increasing complexity of cellular modeling frameworks.

The next generation of structural biology investigations are likely to embrace cell biology by accurately quantifying cellular content in space, measuring molecular activities inside of cells, and simulating cellular behavior in a virtual reality to make predictions that are experimentally testable and that motivate synthetic biology approaches. The combined power of these approaches holds great potential to elucidate the grammar that underlies the complex choreography of cellular self-organization. When will we have understood how cells work? Luckily, there will still be a tremendous amount of work waiting for many generations of future biologists, who will study cells with methods that remain to be invented. However, once digital twins of cells predict their

behavior that could not have been conceived without them, an important milestone will have been achieved.

ACKNOWLEDGMENTS

We thank the faculties of the SubCellular Architecture of LifE (SCALE) consortium for discussions and Stefanie Böhm for critical reading of the manuscript. We apologize to those colleagues whose work could not be cited due to space constraints. We thank Agnieszka Obarska-Kosinska for help with preparing the figures. The SubCellular Architecture of LifE (SCALE) consortium has been funded by the Goethe University, Frankfurt, Germany. M.B. acknowledges funding by the Max Planck Society, Germany; R.C. acknowledges the support of the Frankfurt Institute of Advanced Studies, Frankfurt, Germany. I.H. and M.M.-M. acknowledge funding by the Goethe University, Frankfurt, Germany. I.H. acknowledges funding by the DFG, Germany; Heisenberg grant (HA6322/5-1).

DECLARATION OF INTERESTS

M.B. is a member of the advisory board of *Cell*.

REFERENCES

1. Bonomi, M., and Vendruscolo, M. (2019). Determination of protein structural ensembles using cryo-electron microscopy. Curr. Opin. Struct. Biol. *56*, 37–45.
2. Yabukarski, F., Doukov, T., Pinney, M.M., Biel, J.T., Fraser, J.S., and Herschlag, D. (2022). Ensemble-function relationships to dissect mechanisms of enzyme catalysis. Sci. Adv. *8*, eabn7738.
3. Ken, M.L., Roy, R., Geng, A., Ganser, L.R., Manghrani, A., Cullen, B.R., Schulze-Gahmen, U., Herschlag, D., and Al-Hashimi, H.M. (2023). RNA conformational propensities determine cellular activity. Nature *617*, 835–841.
4. Anderson, P.W. (1972). More is different. Science *177*, 393–396.
5. Robinson, C.V., Sali, A., and Baumeister, W. (2007). The molecular sociology of the cell. Nature *450*, 973–982.
6. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature *596*, 583–589.
7. Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D., Stroe, O., Wood, G., Laydon, A., et al. (2022). AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Res. *50*, D439–D444.
8. Humphreys, I.R., Pei, J., Baek, M., Krishnakumar, A., Anishchenko, I., Ovchinnikov, S., Zhang, J., Ness, T.J., Banjade, S., Bagde, S.R., et al. (2021). Computed structures of core eukaryotic protein complexes. Science *374*, eabm4805.
9. Dauparas, J., Anishchenko, I., Bennett, N., Bai, H., Ragotte, R.J., Milles, L.F., Wicky, B.I.M., Courbet, A., de Haas, R.J., Bethel, N., et al. (2022). Robust deep learning-based protein sequence design using ProteinMPNN. Science *378*, 49–56.
10. Anishchenko, I., Pellock, S.J., Chidyausiku, T.M., Ramelot, T.A., Ovchinnikov, S., Hao, J., Bafna, K., Norn, C., Kang, A., Bera, A.K., et al. (2021). De novo protein design by deep network hallucination. Nature *600*, 547–552.
11. Greenan, G.A., Vale, R.D., and Agard, D.A. (2020). Electron cryotomography of intact motile cilia defines the basal body to axoneme transition. J. Cell Biol. *219*, e201907060.
12. Deguchi, T., Iwanski, M.K., Schentarra, E.M., Heidebrecht, C., Schmidt, L., Heck, J., Weihs, T., Schnorrenberg, S., Hoess, P., Liu, S., et al. (2023). Direct observation of motor protein stepping in living cells using MINFLUX. Science *379*, 1010–1015.
13. O'Reilly, F.J., Xue, L., Graziadei, A., Sinn, L., Lenz, S., Tegunov, D., Blötz, C., Singh, N., Hagen, W.J.H., Cramer, P., et al. (2020). In-cell architecture of an actively transcribing-translating expressome. Science *369*, 554–557.
14. Greber, B.J., Toso, D.B., Fang, J., and Nogales, E. (2019). The complete structure of the human TFIIH core complex. Elife *8*, e44771.
15. Wagner, F.R., Dienemann, C., Wang, H., Stützer, A., Tegunov, D., Urlaub, H., and Cramer, P. (2020). Structure of SWI/SNF chromatin remodeller RSC bound to a nucleosome. Nature *579*, 448–451.
16. Li, M., Tripathi-Giesgen, I., Schulman, B.A., Baumeister, W., and Wilfling, F. (2023). In situ snapshots along a mammalian selective autophagy pathway. Proc. Natl. Acad. Sci. USA *120*, e2221712120.
17. Stagno, J.R., Liu, Y., Bhandari, Y.R., Conrad, C.E., Panja, S., Swain, M., Fan, L., Nelson, G., Li, C., Wendel, D.R., et al. (2017). Structures of riboswitch RNA reaction states by mix-and-inject XFEL serial crystallography. Nature *541*, 242–246.
18. Hofmann, S., Januliene, D., Mehdipour, A.R., Thomas, C., Stefan, E., Brüchert, S., Kuhn, B.T., Geertsma, E.R., Hummer, G., Tampé, R., and Moeller, A. (2019). Conformation space of a heterodimeric ABC exporter under turnover conditions. Nature *571*, 580–583.
19. Arkhipova, V., Guskov, A., and Slotboom, D.J. (2020). Structural ensemble of a glutamate transporter homologue in lipid nanodisc environment. Nat. Commun. *11*, 998.
20. Hänsel, R., Luh, L.M., Corbeski, I., Trantirek, L., and Dötsch, V. (2014). In-cell NMR and EPR spectroscopy of biomacromolecules. Angew. Chem. Int. Ed. Engl. *53*, 10300–10314.
21. Hänelt, I., Wunnicke, D., Bordignon, E., Steinhoff, H.J., and Slotboom, D.J. (2013). Conformational heterogeneity of the aspartate transporter Glt(Ph). Nat. Struct. Mol. Biol. *20*, 210–214.
22. Huang, Y., Reddy, K.D., Bracken, C., Qiu, B., Zhan, W., Eliezer, D., and Boudker, O. (2023). Environmentally ultrasensitive fluorine probe to resolve protein conformational ensembles by ^{19}F NMR and cryo-EM. J. Am. Chem. Soc. *145*, 8583–8592.
23. Wang, Z., Campos, L.A., and Muñoz, V. (2016). Single-molecule fluorescence studies of fast protein folding. Methods Enzymol. *581*, 417–459.
24. Erkens, G.B., Hänelt, I., Goudsmits, J.M., Slotboom, D.J., and van Oijen, A.M. (2013). Unsynchronised subunit motion in single trimeric sodium-coupled aspartate transporters. Nature *502*, 119–123.
25. Ando, T., Uchihashi, T., and Kodera, N. (2013). High-speed AFM and applications to biomolecular systems. Annu. Rev. Biophys. *42*, 393–414.
26. Ruan, Y., Miyagi, A., Wang, X., Chami, M., Boudker, O., and Scheuring, S. (2017). Direct visualization of glutamate transporter elevator mechanism by high-speed AFM. Proc. Natl. Acad. Sci. USA *114*, 1584–1588.
27. Matin, T.R., Heath, G.R., Huysmans, G.H.M., Boudker, O., and Scheuring, S. (2020). Millisecond dynamics of an unlabeled amino acid transporter. Nat. Commun. *11*, 5016.
28. Konermann, L., Pan, J., and Liu, Y.H. (2011). Hydrogen exchange mass spectrometry for studying protein structure and dynamics. Chem. Soc. Rev. *40*, 1224–1234.
29. Wright, P.E., and Dyson, H.J. (2015). Intrinsically disordered proteins in cellular signalling and regulation. Nat. Rev. Mol. Cell Biol. *16*, 18–29.
30. Tsang, B., Pritišanac, I., Scherer, S.W., Moses, A.M., and Forman-Kay, J.D. (2020). Phase separation as a missing mechanism for interpretation of disease mutations. Cell *183*, 1742–1756.
31. Hyman, A.A., Weber, C.A., and Jülicher, F. (2014). Liquid-liquid phase separation in biology. Annu. Rev. Cell Dev. Biol. *30*, 39–58.
32. Ryan, V.H., and Fawzi, N.L. (2019). Physiological, pathological, and targetable membraneless organelles in neurons. Trends Neurosci. *42*, 693–708.
33. Su, M., Zhuang, Y., Miao, X., Zeng, Y., Gao, W., Zhao, W., and Wu, M. (2020). Comparative study of curvature sensing mediated by F-BAR and an intrinsically disordered region of FBP17. iScience *23*, 101712.

34. Stachowiak, J.C., Schmid, E.M., Ryan, C.J., Ann, H.S., Sasaki, D.Y., Sherman, M.B., Geissler, P.L., Fletcher, D.A., and Hayden, C.C. (2012). Membrane bending by protein-protein crowding. Nat. Cell Biol. *14*, 944–949.

35. Alberti, S., Gladfelter, A., and Mittag, T. (2019). Considerations and challenges in studying liquid-liquid phase separation and biomolecular condensates. Cell *176*, 419–434.

36. Girbig, M., Misiaszek, A.D., and Müller, C.W. (2022). Structural insights into nuclear transcription by eukaryotic DNA-dependent RNA polymerases. Nat. Rev. Mol. Cell Biol. *23*, 603–622.

37. Vorländer, M.K., Pacheco-Fiallos, B., and Plaschka, C. (2022). Structural basis of mRNA maturation: time to put it together. Curr. Opin. Struct. Biol. *75*, 102431.

38. Will, C.L., and Lührmann, R. (2011). Spliceosome structure and function. Cold Spring Harb. Perspect. Biol. *3*, a003707.

39. Bohnsack, K.E., Yi, S., Venus, S., Jankowsky, E., and Bohnsack, M.T. (2023). Cellular functions of eukaryotic RNA helicases and their links to human diseases. Nat. Rev. Mol. Cell Biol. *24*, 749–769.

40. Jobe, A., Liu, Z., Gutierrez-Vargas, C., and Frank, J. (2019). New insights into ribosome structure and function. Cold Spring Harb. Perspect. Biol. *11*, a032615.

41. Uzonyi, A., Dierks, D., Nir, R., Kwon, O.S., Toth, U., Barbosa, I., Burel, C., Brandis, A., Rossmanith, W., Le Hir, H., et al. (2023). Exclusion of m6A from splice-site proximal regions by the exon junction complex dictates m6A topologies and mRNA stability. Mol. Cell *83*, 237–251.e7.

42. Pacheco-Fiallos, B., Vorländer, M.K., Riabov-Bassat, D., Fin, L., O'Reilly, F.J., Ayala, F.I., Schellhaas, U., Rappsilber, J., and Plaschka, C. (2023). mRNA recognition and packaging by the human transcription-export complex. Nature *616*, 828–835.

43. Bonneau, F., Basquin, J., Steigenberger, B., Schäfer, T., Schäfer, I.B., and Conti, E. (2023). Nuclear mRNPs are compact particles packaged with a network of proteins promoting RNA-RNA interactions. Genes Dev. *37*, 505–517.

44. Hoffmann, P.C., Kreysing, J.P., Khusainov, I., Tuijtel, M.W., Welsch, S., and Beck, M. (2022). Structures of the eukaryotic ribosome and its translational states in situ. Nat. Commun. *13*, 7435.

45. Xing, H., Taniguchi, R., Khusainov, I., Kreysing, J.P., Welsch, S., Turoňová, B., and Beck, M. (2023). Translation dynamics in human cells visualized at high resolution reveal cancer drug action. Science *381*, 70–75.

46. Xue, L., Lenz, S., Zimmermann-Kogadeeva, M., Tegunov, D., Cramer, P., Bork, P., Rappsilber, J., and Mahamid, J. (2022). Visualizing translation dynamics at atomic detail inside a bacterial cell. Nature *610*, 205–211.

47. Asano, S., Fukuda, Y., Beck, F., Aufderheide, A., Förster, F., Danev, R., and Baumeister, W. (2015). Proteasomes. A molecular census of 26S proteasomes in intact neurons. Science *347*, 439–442.

48. Behrmann, E., Loerke, J., Budkevich, T.V., Yamamoto, K., Schmidt, A., Penczek, P.A., Vos, M.R., Bürger, J., Mielke, T., Scheerer, P., and Spahn, C.M. (2015). Structural snapshots of actively translating human ribosomes. Cell *161*, 845–857.

49. Lin, D.H., and Hoelz, A. (2019). The structure of the nuclear pore complex (an update). Annu. Rev. Biochem. *88*, 725–783.

50. Blundell, T.L., and Patel, S. (2004). High-throughput X-ray crystallography for drug discovery. Curr. Opin. Pharmacol. *4*, 490–496.

51. Pearce, N.M., Krojer, T., Bradley, A.R., Collins, P., Nowak, R.P., Talon, R., Marsden, B.D., Kelm, S., Shi, J., Deane, C.M., and von Delft, F. (2017). A multi-crystal method for extracting obscured crystallographic states from conventionally uninterpretable electron density. Nat. Commun. *8*, 15123.

52. Beck, M., and Baumeister, W. (2016). Cryo-electron tomography: can it reveal the molecular sociology of cells in atomic detail? Trends Cell Biol. *26*, 825–837.

53. Baumeister, W. (2022). Cryo-electron tomography: a long journey to the inner space of cells. Cell *185*, 2649–2652.

54. Zeng, X., Kahng, A., Xue, L., Mahamid, J., Chang, Y.W., and Xu, M. (2023). High-throughput cryo-ET structural pattern mining by unsupervised deep iterative subtomogram clustering. Proc. Natl. Acad. Sci. USA *120*, e2213149120.

55. de Teresa-Trueba, I., Goetz, S.K., Mattausch, A., Stojanovska, F., Zimmerli, C.E., Toro-Nahuelpan, M., Cheng, D.W.C., Tollervey, F., Pape, C., Beck, M., et al. (2023). Convolutional networks for supervised mining of molecular patterns within cellular context. Nat. Methods *20*, 284–294.

56. Lamm, L., Righetto, R.D., Wietrzynski, W., Pöge, M., Martinez-Sanchez, A., Peng, T., and Engel, B.D. (2022). MemBrain: a deep learning-aided pipeline for detection of membrane proteins in Cryo-electron tomograms. Comput. Methods Programs Biomed. *224*, 106990.

57. Gemmer, M., Chaillet, M.L., van Loenhout, J., Cuevas Arenas, R., Vismpas, D., Gröllers-Mulderij, M., Koh, F.A., Albanese, P., Scheltema, R.A., Howes, S.C., et al. (2023). Visualization of translation and protein biogenesis at the ER membrane. Nature *614*, 160–167.

58. Bohm, J., Frangakis, A.S., Hegerl, R., Nickell, S., Typke, D., and Baumeister, W. (2000). Toward detecting and identifying macromolecules in a cellular context: template matching applied to electron tomograms. Proc. Natl. Acad. Sci. USA *97*, 14245–14250.

59. Beck, M., Malmström, J.A., Lange, V., Schmidt, A., Deutsch, E.W., and Aebersold, R. (2009). Visual proteomics of the human pathogen Leptospira interrogans. Nat. Methods *6*, 817–823.

60. Förster, F., Han, B.G., and Beck, M. (2010). Visual proteomics. Methods Enzymol. *483*, 215–243.

61. Cruz-León, S., Majtner, T., Hoffmann, P.C., Kreysing, J.P., Tuijtel, M.W., Schaefer, S.L., Geißler, K., Beck, M., Turoňová, B., and Hummer, G. (2023). High-confidence 3D template matching for cryo-electron tomography. Preprint at bioRxiv.

62. Xu, M., Beck, M., and Alber, F. (2011). Template-free detection of macromolecular complexes in cryo electron tomograms. Bioinformatics *27*, i69–i76.

63. Fung, H.K.H., Hayashi, Y., Salo, V.T., Babenko, A., Zagoriy, I., Brunner, A., Ellenberg, J., Müller, C.W., Cuylen-Haering, S., and Mahamid, J. (2023). Genetically encoded multimeric tags for subcellular protein localization in cryo-EM. Nat. Methods *20*, 1900–1908.

64. Wolff, J.O., Scheiderer, L., Engelhardt, T., Engelhardt, J., Matthias, J., and Hell, S.W. (2023). MINFLUX dissects the unimpeded walking of kinesin-1. Science *379*, 1004–1010.

65. Reinhardt, S.C.M., Masullo, L.A., Baudrexel, I., Steen, P.R., Kowalewski, R., Eklund, A.S., Strauss, S., Unterauer, E.M., Schlichthaerle, T., Strauss, M.T., et al. (2023). Ångström-resolution fluorescence microscopy. Nature *617*, 711–716.

66. Balzarotti, F., Eilers, Y., Gwosch, K.C., Gynnå, A.H., Westphal, V., Stefani, F.D., Elf, J., and Hell, S.W. (2017). Nanometer resolution imaging and tracking of fluorescent molecules with minimal photon fluxes. Science *355*, 606–612.

67. Gwosch, K.C., Pape, J.K., Balzarotti, F., Hoess, P., Ellenberg, J., Ries, J., and Hell, S.W. (2020). MINFLUX nanoscopy delivers 3D multicolor nanometer resolution in cells. Nat. Methods *17*, 217–224.

68. Stephan, T., Brüser, C., Deckers, M., Steyer, A.M., Balzarotti, F., Barbot, M., Behr, T.S., Heim, G., Hübner, W., Ilgen, P., et al. (2020). MICOS assembly controls mitochondrial inner membrane remodeling and crista junction redistribution to mediate cristae formation. EMBO J. *39*, e104105.

69. Pape, J.K., Stephan, T., Balzarotti, F., Büchner, R., Lange, F., Riedel, D., Jakobs, S., and Hell, S.W. (2020). Multicolor 3D MINFLUX nanoscopy of mitochondrial MICOS proteins. Proc. Natl. Acad. Sci. USA *117*, 20607–20614.

70. Ostersehlt, L.M., Jans, D.C., Wittek, A., Keller-Findeisen, J., Inamdar, K., Sahl, S.J., Hell, S.W., and Jakobs, S. (2022). DNA-PAINT MINFLUX nanoscopy. Nat. Methods *19*, 1072–1075.

71. Schmidt, R., Weihs, T., Wurm, C.A., Jansen, I., Rehman, J., Sahl, S.J., and Hell, S.W. (2021). MINFLUX nanometer-scale 3D imaging and microsecond-range tracking on a common fluorescence microscope. Nat. Commun. *12*, 1478.

72. Weber, M., von der Emde, H., Leutenegger, M., Gunkel, P., Sambandan, S., Khan, T.A., Keller-Findeisen, J., Cordes, V.C., and Hell, S.W. (2023). MINSTED nanoscopy enters the Angstrom localization range. Nat. Biotechnol. *41*, 569–576.

73. Kessler, L.F., Balakrishnan, A., Deußner-Helfmann, N.S., Li, Y., Mantel, M., Glogger, M., Barth, H.D., Dietz, M.S., and Heilemann, M. (2023). Self-quenched fluorophore dimers for DNA-PAINT and STED microscopy. Angew. Chem. Int. Ed. Engl. *62*, e202307538.

74. Schueder, F., Rivera-Molina, F., Su, M., Kidd, P., Rothman, J.E., Toomre, D., and Bewersdorf, J. (2023). Unraveling cellular complexity with unlimited multiplexed super-resolution imaging. Preprint at bioRxiv.

75. Yu, M., Heidari, M., Mikhaleva, S., Tan, P.S., Mingu, S., Ruan, H., Reinkemeier, C.D., Obarska-Kosinska, A., Siggel, M., Beck, M., et al. (2023). Visualizing the disordered nuclear transport machinery in situ. Nature *617*, 162–169.

76. Chen, F., Tillberg, P.W., and Boyden, E.S. (2015). Optical imaging. Expansion microscopy. Science *347*, 543–548.

77. Jurriens, D., van Batenburg, V., Katrukha, E.A., and Kapitein, L.C. (2021). Mapping the neuronal cytoskeleton using expansion microscopy. Methods Cell Biol. *161*, 105–124.

78. Bressan, D., Battistoni, G., and Hannon, G.J. (2023). The dawn of spatial omics. Science *381*, eabq4964.

79. Barutcu, A.R., Wu, M., Braunschweig, U., Dyakov, B.J.A., Luo, Z., Turner, K.M., Durbic, T., Lin, Z.Y., Weatheritt, R.J., Maass, P.G., et al. (2022). Systematic mapping of nuclear domain-associated transcripts reveals speckles and lamina as hubs of functionally distinct retained introns. Mol. Cell *82*, 1035–1052.e9.

80. Dörrbaum, A.R., Alvarez-Castelao, B., Nassim-Assir, B., Langer, J.D., and Schuman, E.M. (2020). Proteome dynamics during homeostatic scaling in cultured neurons. Elife *9*, e52939.

81. Sun, C., Nold, A., Fusco, C.M., Rangaraju, V., Tchumatchenko, T., Heilemann, M., and Schuman, E.M. (2021). The prevalence and specificity of local protein synthesis during neuronal synaptic plasticity. Sci. Adv. *7*, eabj0790.

82. Glock, C., Biever, A., Tushev, G., Nassim-Assir, B., Kao, A., Bartnik, I., Tom Dieck, S., and Schuman, E.M. (2021). The translatome of neuronal cell bodies, dendrites, and axons. Proc. Natl. Acad. Sci. USA *118*, e2113929118.

83. Ruwolt, M., Piazza, I., and Liu, F. (2023). The potential of cross-linking mass spectrometry in the development of protein-protein interaction modulators. Curr. Opin. Struct. Biol. *82*, 102648.

84. Stahl, K., Graziadei, A., Dau, T., Brock, O., and Rappsilber, J. (2023). Protein structure prediction with in-cell photo-crosslinking mass spectrometry and deep learning. Nat. Biotechnol. *41*, 1810–1819.

85. Chen, Z.A., and Rappsilber, J. (2023). Protein structure dynamics by crosslinking mass spectrometry. Curr. Opin. Struct. Biol. *80*, 102599.

86. Alexandrov, T. (2020). Spatial metabolomics and imaging mass spectrometry in the age of artificial intelligence. Annu. Rev. Biomed. Data Sci. *3*, 61–87.

87. Hernández-Vega, A., Braun, M., Scharrel, L., Jahnel, M., Wegmann, S., Hyman, B.T., Alberti, S., Diez, S., and Hyman, A.A. (2017). Local nucleation of microtubule bundles through tubulin concentration into a condensed tau phase. Cell Rep. *20*, 2304–2312.

88. Marzahn, M.R., Marada, S., Lee, J., Nourse, A., Kenrick, S., Zhao, H., Ben-Nissan, G., Kolaitis, R.M., Peters, J.L., Pounds, S., et al. (2016). Higher-order oligomerization promotes localization of SPOP to liquid nuclear speckles. EMBO J. *35*, 1254–1275.

89. Johnson, M.E., and Hummer, G. (2011). Nonspecific binding limits the number of proteins in a cell and shapes their interaction networks. Proc. Natl. Acad. Sci. USA *108*, 603–608.

90. Musacchio, A. (2022). On the role of phase separation in the biogenesis of membraneless compartments. EMBO J. *41*, e109952.

91. Mittag, T., and Pappu, R.V. (2022). A conceptual framework for understanding phase separation and addressing open questions and challenges. Mol. Cell *82*, 2201–2214.

92. de Oliveira Freitas Machado, C., Schafranek, M., Brüggemann, M., Hernández Cañás, M.C., Keller, M., Di Liddo, A., Brezski, A., Blümel, N., Arnold, B., Bremm, A., et al. (2023). Poison cassette exon splicing of SRSF6 regulates nuclear speckle dispersal and the response to hypoxia. Nucleic Acids Res. *51*, 870–890.

93. Königs, V., de Oliveira Freitas Machado, C., Arnold, B., Blümel, N., Solovyeva, A., Löbbert, S., Schafranek, M., Ruiz De Los Mozos, I., Wittig, I., McNicoll, F., et al. (2020). SRSF7 maintains its homeostasis through the expression of Split-ORFs and nuclear body assembly. Nat. Struct. Mol. Biol. *27*, 260–273.

94. Sun, C., and Schuman, E. (2023). A multi-omics view of neuronal subcellular protein synthesis. Curr. Opin. Neurobiol. *80*, 102705.

95. Hafner, A.S., Donlin-Asp, P.G., Leitch, B., Herzog, E., and Schuman, E.M. (2019). Local protein synthesis is a ubiquitous feature of neuronal pre- and postsynaptic compartments. Science *364*, eaau3644.

96. Frydman, J. (2001). Folding of newly translated proteins in vivo: the role of molecular chaperones. Annu. Rev. Biochem. *70*, 603–647.

97. Petrovic, S., Samanta, D., Perriches, T., Bley, C.J., Thierbach, K., Brown, B., Nie, S., Mobbs, G.W., Stevens, T.A., Liu, X., et al. (2022). Architecture of the linker-scaffold in the nuclear pore. Science *376*, eabm9798.

98. Fischer, J., Teimer, R., Amlacher, S., Kunze, R., and Hurt, E. (2015). Linker Nups connect the nuclear pore complex inner ring with the outer ring and transport channel. Nat. Struct. Mol. Biol. *22*, 774–781.

99. Schwarz, A., and Beck, M. (2019). The benefits of cotranslational assembly: a structural perspective. Trends Cell Biol. *29*, 791–803.

100. Hirose, T., Ninomiya, K., Nakagawa, S., and Yamazaki, T. (2023). A guide to membraneless organelles and their various roles in gene regulation. Nat. Rev. Mol. Cell Biol. *24*, 288–304.

101. Yamazaki, T., Yamamoto, T., Yoshino, H., Souquere, S., Nakagawa, S., Pierron, G., and Hirose, T. (2021). Paraspeckles are constructed as block copolymer micelles. EMBO J. *40*, e107270.

102. Harayama, T., and Riezman, H. (2018). Understanding the diversity of membrane lipid composition. Nat. Rev. Mol. Cell Biol. *19*, 281–296.

103. Halbleib, K., Pesek, K., Covino, R., Hofbauer, H.F., Wunnicke, D., Hänelt, I., Hummer, G., and Ernst, R. (2017). Activation of the unfolded protein response by lipid bilayer stress. Mol. Cell *67*, 673–684.e8.

104. Singer, S.J., and Nicolson, G.L. (1972). The fluid mosaic model of the structure of cell membranes. Science *175*, 720–731.

105. Simons, K., and Ikonen, E. (1997). Functional rafts in cell membranes. Nature *387*, 569–572.

106. Levental, I., and Lyman, E. (2023). Regulation of membrane protein structure and function by their lipid nano-environment. Nat. Rev. Mol. Cell Biol. *24*, 107–122.

107. Jensen, M.Ø., and Mouritsen, O.G. (2004). Lipids do influence protein function-the hydrophobic matching hypothesis revisited. Biochim. Biophys. Acta *1666*, 205–226.

108. Rangamani, P., Mandadap, K.K., and Oster, G. (2014). Protein-induced membrane curvature alters local membrane tension. Biophys. J. *107*, 751–762.

109. Bigay, J., Gounon, P., Robineau, S., and Antonny, B. (2003). Lipid packing sensed by ArfGAP1 couples COPI coat disassembly to membrane bilayer curvature. Nature *426*, 563–566.

110. Boucrot, E., Pick, A., Çamdere, G., Liska, N., Evergren, E., McMahon, H.T., and Kozlov, M.M. (2012). Membrane fission is promoted by

insertion of amphipathic helices and is restricted by crescent BAR domains. Cell *149*, 124–136.

111. Boyle, E., and Wilfling, F. (2023). Autophagy as a caretaker of nuclear integrity. FEBS Lett. *597*, 2728–2738.
112. Terasaki, M., Shemesh, T., Kasthuri, N., Klemm, R.W., Schalek, R., Hayworth, K.J., Hand, A.R., Yankova, M., Huber, G., Lichtman, J.W., et al. (2013). Stacked endoplasmic reticulum sheets are connected by helicoidal membrane motifs. Cell *154*, 285–296.
113. von der Malsburg, A., Sapp, G.M., Zuccaro, K.E., von Appen, A., Moss, F.R., 3rd, Kalia, R., Bennett, J.A., Abriata, L.A., Dal Peraro, M., van der Laan, M., et al. (2023). Structural mechanism of mitochondrial membrane remodelling by human OPA1. Nature *620*, 1101–1108.
114. Romanauska, A., and Köhler, A. (2023). Lipid saturation controls nuclear envelope function. Nat. Cell Biol. *25*, 1290–1302.
115. Bharathan, N.K., Giang, W., Hoffman, C.L., Aaron, J.S., Khuon, S., Chew, T.L., Preibisch, S., Trautman, E.T., Heinrich, L., Bogovic, J., et al. (2023). Architecture and dynamics of a desmosome-endoplasmic reticulum complex. Nat. Cell Biol. *25*, 823–835.
116. Ungricht, R., and Kutay, U. (2017). Mechanisms and functions of nuclear envelope remodelling. Nat. Rev. Mol. Cell Biol. *18*, 229–245.
117. Venkova, L., Vishen, A.S., Lembo, S., Srivastava, N., Duchamp, B., Ruppel, A., Williart, A., Vassilopoulos, S., Deslys, A., Garcia Arcos, J.M., et al. (2022). A mechano-osmotic feedback couples cell volume to the rate of cell deformation. Elife *11*, e72381.
118. Roffay, C., Molinard, G., Kim, K., Urbanska, M., Andrade, V., Barbarasa, V., Nowak, P., Mercier, V., García-Calvo, J., Matile, S., et al. (2021). Passive coupling of membrane tension and cell volume during active response of cells to osmosis. Proc. Natl. Acad. Sci. USA *118*, e2103228118.
119. Zimmerli, C.E., Allegretti, M., Rantos, V., Goetz, S.K., Obarska-Kosinska, A., Zagoriy, I., Halavatyi, A., Hummer, G., Mahamid, J., Kosinski, J., and Beck, M. (2021). Nuclear pores dilate and constrict in cellulo. Science *374*, eabd9776.
120. Jiang, Y., Yang, X., Jiang, J., and Xiao, B. (2021). Structural designs and mechanogating mechanisms of the mechanosensitive piezo channels. Trends Biochem. Sci. *46*, 472–488.
121. Yuan, F., Lee, C.T., Sangani, A., Houser, J.R., Wang, L., Lafer, E.M., Rangamani, P., and Stachowiak, J.C. (2023). The ins and outs of membrane bending by intrinsically disordered proteins. Sci. Adv. *9*, eadg3485.
122. Mund, M., Tschanz, A., Wu, Y.L., Frey, F., Mehl, J.L., Kaksonen, M., Avinoam, O., Schwarz, U.S., and Ries, J. (2023). Clathrin coats partially preassemble and subsequently bend during endocytosis. J. Cell Biol. *222*, e202206038.
123. Hubatsch, L., Jawerth, L.M., Love, C., Bauermann, J., Tang, T.D., Bo, S., Hyman, A.A., and Weber, C.A. (2021). Quantitative theory for the diffusive dynamics of liquid condensates. Elife *10*, e68620.
124. Gouveia, B., Kim, Y., Shaevitz, J.W., Petry, S., Stone, H.A., and Brangwynne, C.P. (2022). Capillary forces generated by biomolecular condensates. Nature *609*, 255–264.
125. Fare, C.M., Villani, A., Drake, L.E., and Shorter, J. (2021). Higher-order organization of biomolecular condensates. Open Biol. *11*, 210137.
126. Misner, C.W., Thorne, K.S., Wheeler, J.A., and Kaiser, D. (2017). Gravitation (Princeton University Press).
127. Corradi, V., Mendez-Villuendas, E., Ingólfsson, H.I., Gu, R.X., Siuda, I., Melo, M.N., Moussatova, A., DeGagné, L.J., Sejdiu, B.I., Singh, G., et al. (2018). Lipid-protein interactions are unique fingerprints for membrane proteins. ACS Cent. Sci. *4*, 709–717.
128. Helfrich, W. (1973). Elastic properties of lipid bilayers: theory and possible experiments. Z. Naturforsch. C *28*, 693–703.
129. Andersen, O.S., and Koeppe, R.E., 2nd. (2007). Bilayer thickness and membrane protein function: an energetic perspective. Annu. Rev. Biophys. Biomol. Struct. *36*, 107–130.
130. Takamori, S., Holt, M., Stenius, K., Lemke, E.A., Grønborg, M., Riedel, D., Urlaub, H., Schenck, S., Brügger, B., Ringler, P., et al. (2006). Molecular anatomy of a trafficking organelle. Cell *127*, 831–846.
131. Sawa-Makarska, J., Baumann, V., Coudevylle, N., von Bülow, S., Nogellova, V., Abert, C., Schuschnig, M., Graef, M., Hummer, G., and Martens, S. (2020). Reconstitution of autophagosome nucleation defines Atg9 vesicles as seeds for membrane formation. Science *369*, eaaz7714.
132. Mosalaganti, S., Obarska-Kosinska, A., Siggel, M., Taniguchi, R., Turoňová, B., Zimmerli, C.E., Buczak, K., Schmidt, F.H., Margiotta, E., Mackmull, M.T., et al. (2022). AI-based structure prediction empowers integrative structural analysis of human nuclear pores. Science *376*, eabm9506.
133. Alber, M., Buganza Tepole, A., Cannon, W.R., De, S., Dura-Bernal, S., Garikipati, K., Karniadakis, G., Lytton, W.W., Perdikaris, P., Petzold, L., and Kuhl, E. (2019). Integrating machine learning and multiscale modeling-perspectives, challenges, and opportunities in the biological, biomedical, and behavioral sciences. NPJ Digit. Med. *2*, 115.
134. Grieves, M. (2019). Virtually intelligent product systems: digital and physical twins. In Complex Systems Engineering: Theory and Practice, S. Flumerfelt, K.G. Schwartz, D. Mavris, and S. Briceno, eds. (American Institute of Aeronautics and Astronautics), pp. 175–200.
135. Delhommel, F., Martínez-Lumbreras, S., and Sattler, M. (2023). Combining NMR, SAXS and SANS to characterize the structure and dynamics of protein complexes. Methods Enzymol. *678*, 263–297.
136. Thompson, M.K., Ehlinger, A.C., and Chazin, W.J. (2017). Analysis of functional dynamics of modular multidomain proteins by SAXS and NMR. Methods Enzymol. *592*, 49–76.
137. Trabuco, L.G., Villa, E., Mitra, K., Frank, J., and Schulten, K. (2008). Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. Structure *16*, 673–683.
138. Rappsilber, J. (2011). The beginning of a beautiful friendship: cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes. J. Struct. Biol. *173*, 530–540.
139. Rantos, V., Karius, K., and Kosinski, J. (2022). Integrative structural modeling of macromolecular complexes using assembline. Nat. Protoc. *17*, 152–176.
140. Russel, D., Lasker, K., Webb, B., Velázquez-Muriel, J., Tjioe, E., Schneidman-Duhovny, D., Peterson, B., and Sali, A. (2012). Putting the pieces together: integrative modeling platform software for structure determination of macromolecular assemblies. PLoS Biol. *10*, e1001244.
141. Rout, M.P., and Sali, A. (2019). Principles for integrative structural biology studies. Cell *177*, 1384–1403.
142. Senior, A.W., Evans, R., Jumper, J., Kirkpatrick, J., Sifre, L., Green, T., Qin, C., Žídek, A., Nelson, A.W.R., Bridgland, A., et al. (2020). Improved protein structure prediction using potentials from deep learning. Nature *577*, 706–710.
143. Tunyasuvunakool, K., Adler, J., Wu, Z., Green, T., Zielinski, M., Žídek, A., Bridgland, A., Cowie, A., Meyer, C., Laydon, A., et al. (2021). Highly accurate protein structure prediction for the human proteome. Nature *596*, 590–596.
144. Ruff, K.M., and Pappu, R.V. (2021). AlphaFold and implications for intrinsically disordered proteins. J. Mol. Biol. *433*, 167208.
145. Fernandez-Martinez, J., and Rout, M.P. (2021). One ring to rule them all? Structural and functional diversity in the nuclear pore complex. Trends Biochem. Sci. *46*, 595–607.
146. Schwartz, T.U. (2022). Solving the nuclear pore puzzle. Science *376*, 1158–1159.
147. Hampoelz, B., Andres-Pons, A., Kastritis, P., and Beck, M. (2019). Structure and assembly of the nuclear pore complex. Annu. Rev. Biophys. *48*, 515–536.
148. Dror, R.O., Dirks, R.M., Grossman, J.P., Xu, H., and Shaw, D.E. (2012). Biomolecular simulation: a computational microscope for molecular biology. Annu. Rev. Biophys. *41*, 429–452.

149. Hollingsworth, S.A., and Dror, R.O. (2018). Molecular Dynamics simulation for all. Neuron *99*, 1129–1143.

150. Bhaskara, R.M., Grumati, P., Garcia-Pardo, J., Kalayil, S., Covarrubias-Pinto, A., Chen, W., Kudryashev, M., Dikic, I., and Hummer, G. (2019). Curvature induction and membrane remodeling by FAM134B reticulon homology domain assist selective ER-phagy. Nat. Commun. *10*, 2370.

151. González, A., Covarrubias-Pinto, A., Bhaskara, R.M., Glogger, M., Kuncha, S.K., Xavier, A., Seemann, E., Misra, M., Hoffmann, M.E., Bräuning, B., et al. (2023). Ubiquitination regulates ER-phagy and remodelling of endoplasmic reticulum. Nature *618*, 394–401.

152. Goretzki, B., Wiedemann, C., McCray, B.A., Schäfer, S.L., Jansen, J., Tebbe, F., Mitrovic, S.A., Nöth, J., Cabezudo, A.C., Donohue, J.K., et al. (2023). Crosstalk between regulatory elements in disordered TRPV4 N-terminus modulates lipid-dependent channel activity. Nat. Commun. *14*, 4165.

153. Turoňová, B., Sikora, M., Schürmann, C., Hagen, W.J.H., Welsch, S., Blanc, F.E.C., von Bülow, S., Gecht, M., Bagola, K., Hörner, C., et al. (2020). In situ structural analysis of SARS-CoV-2 spike reveals flexibility mediated by three hinges. Science *370*, 203–208.

154. Sikora, M., von Bülow, S., Blanc, F.E.C., Gecht, M., Covino, R., and Hummer, G. (2021). Computational epitope map of SARS-CoV-2 spike protein. PLoS Comput. Biol. *17*, e1008790.

155. von Bülow, S., Sikora, M., Blanc, F.E.C., Covino, R., and Hummer, G. (2023). Antibody accessibility determines location of spike surface mutations in SARS-CoV-2 variants. PLoS Comput. Biol. *19*, e1010822.

156. Lindorff-Larsen, K., Maragakis, P., Piana, S., Eastwood, M.P., Dror, R.O., and Shaw, D.E. (2012). Systematic validation of protein force fields against experimental data. PLoS One *7*, e32131.

157. Robustelli, P., Piana, S., and Shaw, D.E. (2018). Developing a molecular dynamics force field for both folded and disordered protein states. Proc. Natl. Acad. Sci. USA *115*, E4758–E4766.

158. Jung, H., Covino, R., Arjun, A., Leitold, C., Dellago, C., Bolhuis, P.G., and Hummer, G. (2023). Machine-guided path sampling to discover mechanisms of molecular self-organization. Nat. Comput. Sci. *3*, 334–345.

159. Souza, P.C.T., Alessandri, R., Barnoud, J., Thallmair, S., Faustino, I., Grünewald, F., Patmanidis, I., Abdizadeh, H., Bruininks, B.M.H., Wassenaar, T.A., et al. (2021). Martini 3: a general purpose force field for coarse-grained molecular dynamics. Nat. Methods *18*, 382–388.

160. Venable, R.M., Brown, F.L.H., and Pastor, R.W. (2015). Mechanical properties of lipid bilayers from molecular dynamics simulation. Chem. Phys. Lipids *192*, 60–74.

161. Dacheux, L., Moreau, A., Ataman-Onal, Y., Biron, F., Verrier, B., and Barin, F. (2004). Evolutionary dynamics of the glycan shield of the human immunodeficiency virus envelope during natural infection and implications for exposure of the 2G12 epitope. J. Virol. *78*, 12625–12637.

162. Poon, A.F., Lewis, F.I., Pond, S.L., and Frost, S.D. (2007). Evolutionary interactions between N-linked glycosylation sites in the HIV-1 envelope. PLoS Comput. Biol. *3*, e11.

163. Thornburg, Z.R., Bianchi, D.M., Brier, T.A., Gilbert, B.R., Earnest, T.M., Melo, M.C.R., Safronova, N., Sáenz, J.P., Cook, A.T., Wise, K.S., et al. (2022). Fundamental behaviors emerge from simulations of a living minimal cell. Cell *185*, 345–360.e28.

164. Casalino, L., Dommer, A.C., Gaieb, Z., Barros, E.P., Sztain, T., Ahn, S.-H., Trifan, A., Brace, A., Bogetti, A.T., Clyde, A., et al. (2021). AI-driven multiscale simulations illuminate mechanisms of SARS-CoV-2 spike dynamics. Int. J. High Perform. Comput. Appl. *35*, 432–451.

165. Lavin, A., Krakauer, D., Zenil, H., Gottschlich, J., Mattson, T., Brehmer, J., Anandkumar, A., Choudry, S., Rocki, K., Baydin, A.G., et al. (2022). Simulation intelligence: towards a new generation of scientific methods. Preprint at arXiv.