Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

Structural biology *in cellulo*: Minding the gap between conceptualization and realization

Fotis L. Kyrilis[1], Jason K. K. Low[2], Joel P. Mackay[2] and Panagiotis L. Kastritis[1,3,4,5]

Abstract

Recent technological advances have deepened our perception of cellular structure. However, most structural data doesn't originate from intact cells, limiting our understanding of cellular processes. Here, we discuss current and future developments that will bring us towards a structural picture of the cell. Electron cryotomography is the standard bearer, with its ability to provide *in cellulo* snapshots. Single-particle electron microscopy (of purified biomolecules and of complex mixtures) and covalent crosslinking combined with mass spectrometry also have significant roles to play, as do artificial intelligence algorithms in their many guises. To integrate these multiple approaches, data curation and standardisation will be critical – as is the need to expand efforts beyond our current protein-centric view to the other (macro)molecules that sustain life.

Addresses

[1] Institute of Chemical Biology, National Hellenic Research Foundation, Athens, Greece

[2] School of Life and Environmental Sciences, University of Sydney, Sydney, NSW 2006, Australia

[3] Interdisciplinary Research Center HALOmem, Charles Tanford Protein Center, Martin Luther University Halle-Wittenberg, Kurt-Mothes-Straße 3a, Halle/Saale, Germany

[4] Institute of Biochemistry and Biotechnology, Martin Luther University Halle-Wittenberg, Kurt-Mothes-Straße 3, Halle/Saale, Germany

[5] Biozentrum, Martin Luther University Halle-Wittenberg, Weinbergweg 22, Halle/Saale, Germany

Corresponding authors: Kastritis, Panagiotis L. (panagiotis.kastritis@bct.uni-halle.de); Mackay, Joel P. (joel.mackay@sydney.edu.au)

(Kyrilis F.L.), (Mackay J.P.), (Kastritis P.L.)

Current Opinion in Structural Biology 2024, **87**:102843

This review comes from a themed issue on **Biophysical Methods (2024)**

Edited by **Gabriela Chiosis** and **Elizabeth A. Komives**

For complete overview of the section, please refer the article collection - Biophysical Methods (2024)

Available online 23 May 2024

https://doi.org/10.1016/j.sbi.2024.102843

Introduction

The pioneering work of Hodgkin, Pauling, Corey, Astbury and others in the 1930s and 1940s marked the start of a new era for biology: one that promised a molecular understanding of life itself. The ensuing 80 years have seen a succession of ground-breaking technical advances in X-ray crystallography, NMR spectroscopy and most recently single-particle electron cryo-microscopy that have made huge inroads on that early promise. The dramatic appearance of artificial intelligence (AI) in the form of AlphaFold and related software has also begun to reshape the landscape of structural biology and will undoubtedly evolve rapidly in the near future.

By necessity, the overwhelming majority of this work has been carried out on isolated macromolecules that are trapped in the confines of a crystal, an NMR tube, an EM grid or—in the case of molecular dynamics — a small box of water. Furthermore, most studies have examined proteins that are truncated and lacking their biological binding partners (be they other proteins, membranes, other macromolecules, cofactors …), although we now appreciate that intrinsically disordered regions and the formation of macromolecular complexes can be essential for proper cellular function. In this regard, it is notable that AlphaFold has almost in passing highlighted the prevalence of unstructured regions across the protein universe, particularly to those who are not structural biologists.

Although these approaches have yielded enormous molecular insight across nearly every field of biology, it has not gone unappreciated by the field that the lack of cellular context for most macromolecular structures represents a yawning chasm that must be crossed if we are to bring the fields of structural biology and cell biology to a common ground. Put most simply, what is the structure of a cell?

Fortunately, work to bridge this chasm has been underway for a number of years and in this review, we'll

explore recent advances in the field now recognized as structural cell biology or visual proteomics, discussing the associated opportunities, challenges, and prospects that underpin the ongoing pursuit of this — perhaps unobtainable — goal.

Advances in cellular tomography open doors to visualizing the cellular milieu in unprecedented detail

Conceptually, perhaps the simplest approach to obtaining an atomic resolution snapshot of a cell is to use electron cryo-tomography (cryoET). In cryoET, micrographs are recorded of a thin ($<\sim$500 nm) slice of a cell or any other sample tilted at many angles and the images are computationally combined to create a three-dimensional image (a tomogram). This concept was introduced in 1968 in three independent reports [1–3] but only recently was it substantially advanced by Baumeister and others for high-resolution in-cell structural analysis [4]. Furthermore, if multiple copies of a structure are present in a single tomogram, these can be computationally combined (sub-tomogram averaging) to improve both contrast and resolution of that entity, as in the case of the 80 S ribosome, reconstructed at 3.1 Å, from inside human cells [5].

The unique power of cryoET is that structures are observed in their native cellular environment (albeit at temperatures lower than −155 °C), and this approach has yielded exquisite images and reconstructions of large cellular structures such as nuclear pore complexes (NPCs) [6], flagella and injectisomes [7], as well as diverse macromolecular complexes including ribosomes [5,8,9], proteasomes [10], actin [11] and tubulin [12] structures, and calcium channels [13]. These studies offer new perspectives on both structure and function, as well as informing on spatial relationships between cellular structures. For example, cryoET was used to elucidate the molecular architecture of native cardiac sarcomeres [14,15], revealing the organization of thick filament proteins such as myosin, titin, and myosin-binding protein C; these proteins have specialized roles in strain susceptibility, force generation, and length-dependent activation. This pioneering study also lays the groundwork for understanding muscle disorders associated with sarcomeric components (Figure 1).

Recent work has demonstrated that atomic resolution (2.3 Å) can be achieved by cryoET for purified proteins [9], and sub-4 Å reconstructions for in-cell ribosomes [9] — though more typical values are in the range 10–20 Å. Resolution is limited by several factors, including sample thickness, low contrast, number of subtomograms, radiation damage, and the missing 'wedge' (only $\sim$120° of rotation is typically possible). These resolutions mean that structures cannot be built directly from the reconstructed maps, and models therefore must be built from complementary data, including high-resolution structures or models of subunits. Machine learning has begun to have a significant impact in this space, as demonstrated by AI-driven models derived for the human NPC and its ATP-dependent large-scale dilation properties that integrated cryoET data with decades of biochemical and structural knowledge [6,16]. Battles to improve resolution are likely to be won by combining several different strategies. Cryo-focused ion beam milling methods like plasma milling [17] will provide controllable, cleaner and more uniform samples and will also improve throughput to furnish significantly larger datasets, as will emerging methods for parallel acquisition of multiple tilt series [18]. Mirroring cryoEM, advances in microscope hardware, including detectors and phase plates, will certainly have an impact, as will better algorithms for image alignment and reconstruction.

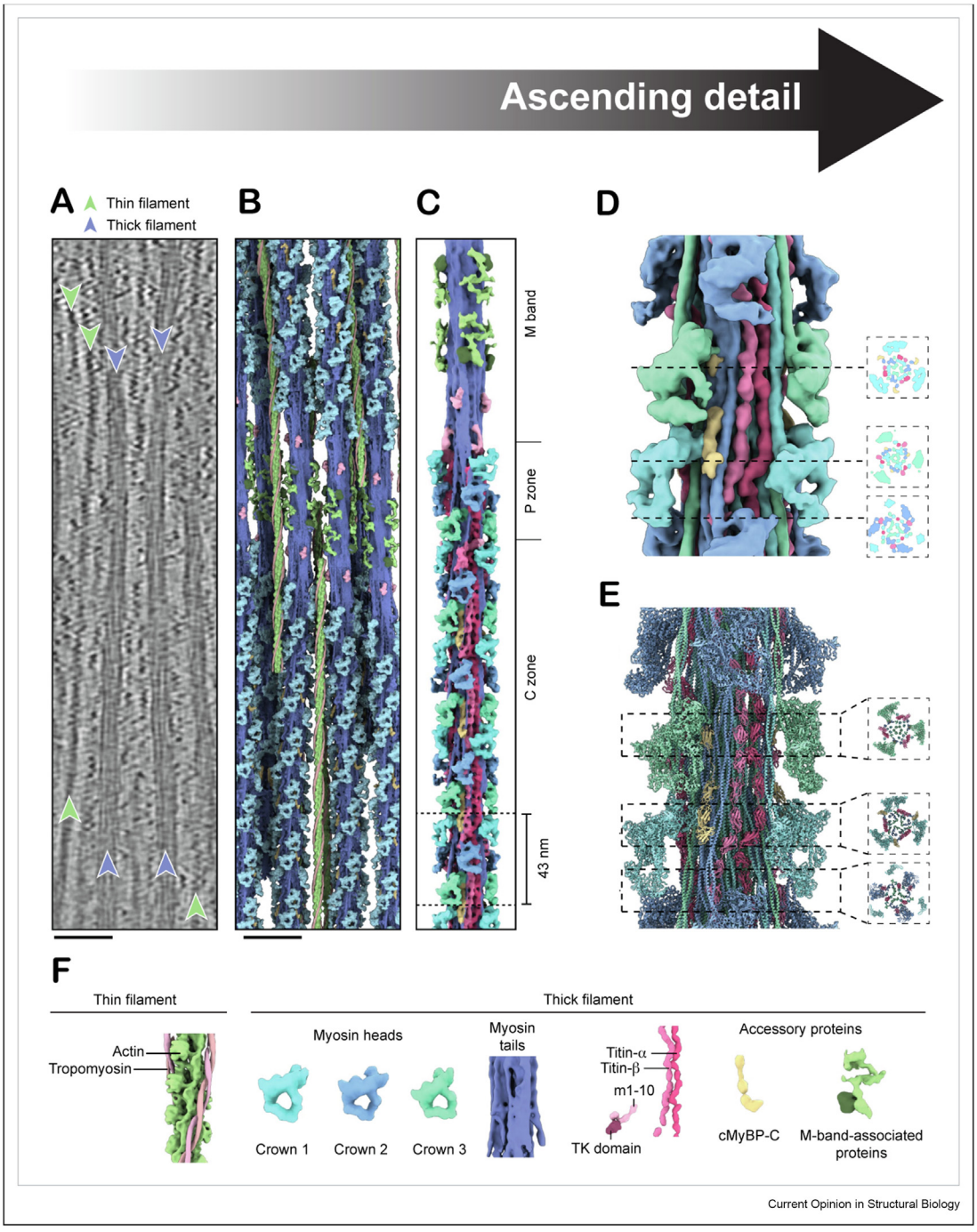
The biggest challenge, however, likely lies in the identification and analysis of small and/or less abundant macromolecules and assemblies, where conventional template matching approaches fail. Two recently described algorithms based on deep learning allowed the *de novo* extraction of 'structural signatures' for unknown complexes [19,20]. Structural signatures are single-particle 2D projections, 2D class averages, or subtomograms that correspond to a specific, clear view of a macromolecular structure — and provide a starting point for more detailed structural analysis. Overall, the efficacy of neural networks in recovering structural signatures from electron tomograms has grown, enabling this task to be accomplished without reliance on pre-trained data. In particular, combinations of machine learning models are currently being explored to allow an unbiased, comprehensive search of structural signatures within tomograms [19,20].

Thus, drawing a parallel to AlphaFold's success in predicting protein structures via leveraging experimentally derived structural data, one can envisage a similar algorithmic approach that would mine structural signatures from an extensive repository of cellular cryo-electron tomograms, consequently unveiling biomolecular signatures and their interrelationships. The success of such strategies will rely heavily on the public availability of extremely large cryoET datasets (by analogy to the relationship between AlphaFold and the PDB). Towards this goal, the open-access EMPIAR database (https://www.ebi.ac.uk/empiar/) must play a leading role in centralizing such data for the decades to come.

Cellular electron microscopy: high contrast, low dose, and high-resolution content

A significant issue in cryoET is the high electron dose and consequent sample damage that is accumulated

Figure 1



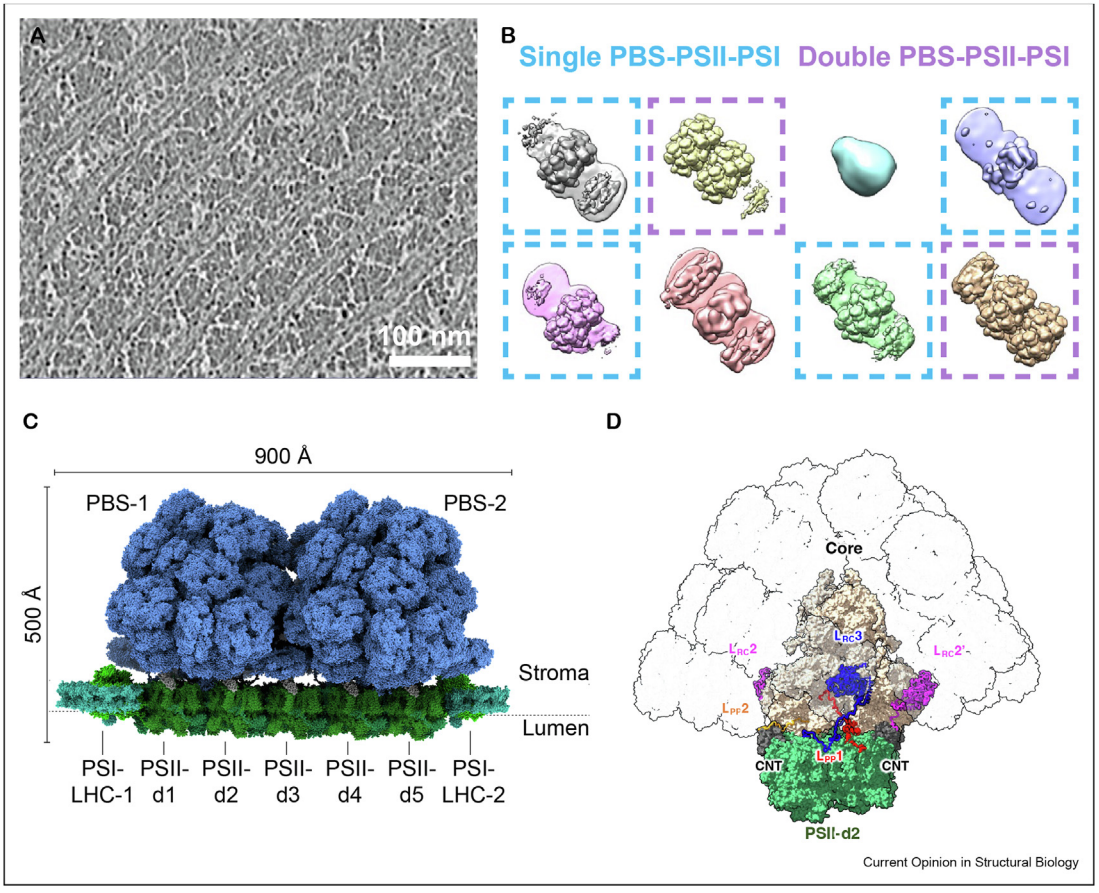
(a) Tomographic slice of a cardiac sarcomere, centered on the M band, depicting thick and thin filaments. Scale bar, 50 nm. **(b)** Reconstructed thick and thin filaments mapped into a tomogram. Thin filaments obstructing the view on the thick filament were removed for clarity. Scale bar, 50 nm. **(c)** Structure of the thick filament from the M band to the C zone. For clarity, only the first four cMyBP-C stripes are shown **(c, d)**. The 3D reconstruction **(d)** and atomic model **(e)** of the C zone, from cMyBP-C stripe no. 4 to stripe no. 9. The volume is colored according to its atomic model. The dotted boxes on the right side depict Z-ward views of cross-sections of the map **(d)** and model **(e)**, providing a more detailed view of the arrangement within the core of the thick filament. **(f)** Illustration of the various sarcomere components and their color code, which is maintained throughout this figure. Data reproduced and adapted from Ref. [15].

during the recording of a tilt series that might comprise >100 individual images. If only single images are recorded — an approach termed single-particle electron microscopy — the consequent reduction in sample damage leads directly to higher resolution data, though at the cost of the loss of information on spatial organization. The Grigorieff laboratory has shown that recovery of structural signatures from high-resolution cryo-electron micrographs of cells is feasible [21,22], allowing the high-throughput accumulation of these signatures that can be utilized in subsequent structural analysis.

Furthermore, a recent tour-de-force analysis of photosystem supercomplexes from intact cells showed that single particle analysis (SPA) can provide atomic resolution structures of gigantic complexes at unprecedented resolution — even within the cellular context [23]. Figure 2 showcases the incredible power of cryoET and *in situ* single-particle analysis demonstrated in this study. The authors determined several near-atomic-resolution structures of the monomeric and dimeric PBS–PSII–PSI–LHC megacomplex from the red alga *Porphyridium purpureum*. The megacomplex contained 1,792 proteins made from more than 300,000 residues and featured 4,500 ligands. Remarkable insights into the assembly of the megacomplex were gained in this work, including the nature and arrangement of the pigment network and the mechanisms governing energy transfer and distribution across the components of the complex (see Figure 2).

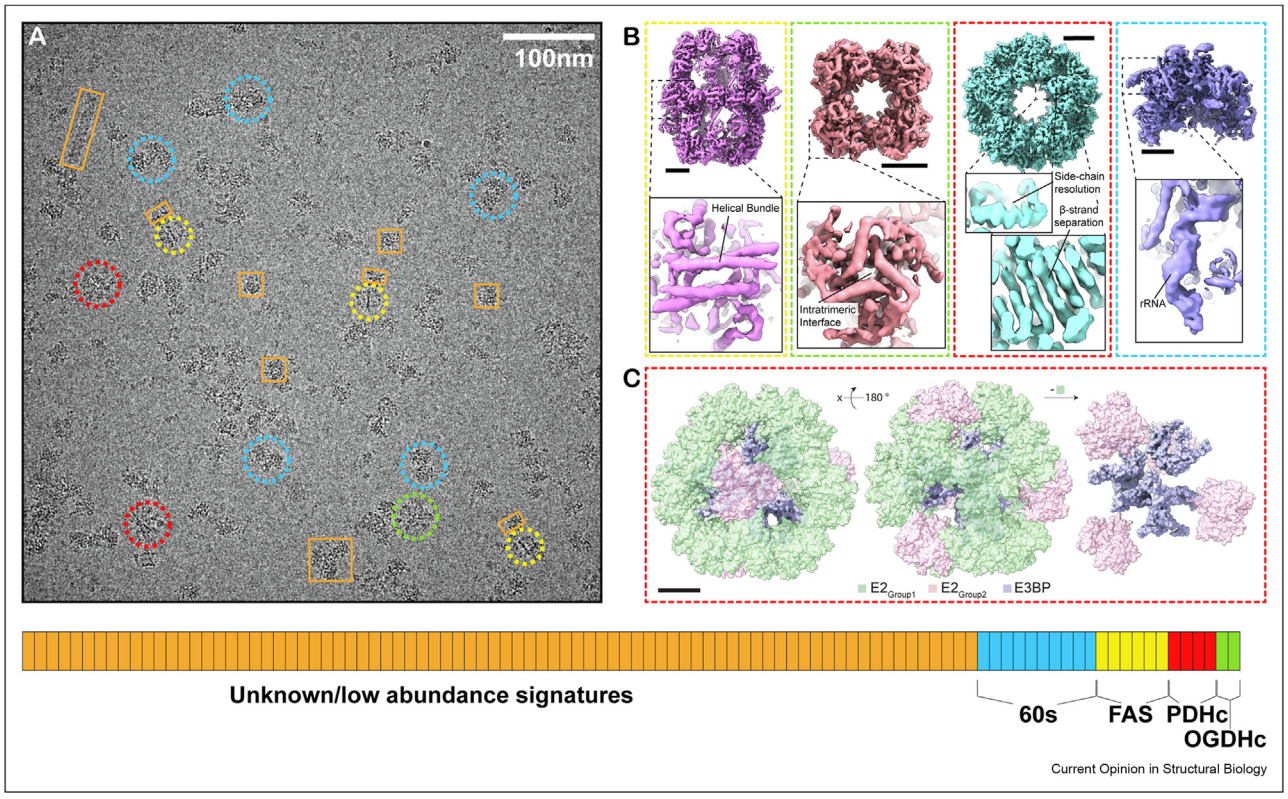
This study shows clearly what is possible in the context of SPA of intact cells. We envision a future in which large databases of high-resolution cellular single-particle cryoEM images not only provide stand-alone structural information on protein complexes in their native environment, but can be leveraged for the interpretation of cryoET data. For this future to become a reality, an enabling step will be a community-wide commitment to open data access according to FAIR (Findable, Accessible, Identifiable, Reproducible) principles.

Figure 2



(a) Representative tomogram of the phycobilisome (PBS)–PSII–PSI–LHC megacomplex from *P. purpureum* [23] **(b)** Representative 3D classes of single (light blue boxes) and double (purple boxes) PBS-PSII-PSI. **(c)** The overall structures of the double PBS–PSII–PSI–LHC assembly is shown as a surface representation; 2.5 million atoms were modeled in the displayed densities. PBSs are corn flower colored whereas PSII dimers and PSI–LHCs are colored in dark and lighter green respectively **(d)** Organization of PBS and PSII mediated by linker proteins (LRC2, LRC3, LPP1 and LPP2). Data for **(a, b)**, and **(d)** panels were kindly provided by the authors [23].

Figure 3



(a) Representative cryoEM micrograph of fractionated MDa-size complexes. Insets highlight the pluralistic abundance of structural signatures found within such fractions. Yellow dotted circles enclose fatty acid synthase (FAS), green oxoglutarate dehydrogenase complex (OGDHc), light blue ribosomal subunits and red color depicts pyruvate dehydrogenase complex (PDHc). **(b)** Reconstructions of signatures highlighted in panel A: FAS complex with α-helical bundles (yellow box), reconstruction of the OGDHc E2 core, where the intra-trimeric interfaces at the edge of the core are recapitulated (green box), reconstruction of the PDHc E2 core, highlighting high-resolution structural features, such as side-chain densities and β-strand separation (red box), reconstruction of the pre-60 S ribosomal subunit with densities belonging to the rRNA structural elements being visible (light blue box), **(c)** PDHc E2 core in proximity to the E3 binding protein (E3BP) with proximal E2 monomers colored green, and distal colored pink. Scale bars in B and C: 5 nm. Bottom bar illustrates the complexity of cell extracts with uncharacterized structural signatures (orange) representing to the majority (Modified from Refs. [27,28,30]).

Cell extracts: an intermediate route to untangling molecular complexity

The direct imaging of cellular sections (lamellae) by cryoET or cryoEM offers the most native-like incarnation of structural cell biology, but also the one that requires deconvolution of the highest level of complexity. There are likely to be of the order of a billion individual protein molecules in a human cell (a number still being challenged [24]) and those molecules are packed cheek by jowl into a confined space, a situation that presents significant contrast challenges for particle picking.

At the opposite end of the complexity scale lies the widespread practice of expressing and purifying individual proteins and protein complexes recombinantly for structural analysis. Such structures are and will continue to be invaluable for the interpretation of cryoET data. In this incarnation, dynamics and other phenomena can complicate analysis but at least the identity of the study subject is unambiguous. Of course, the set of proteins (and other molecules) required for such experiments can only be determined following a substantial (often decades) investment in biochemistry. In some cases, a shortcut can be taken by purifying complexes from their native source — and of course, this was the primary strategy taken in the first decades of structural biology. During that time, only abundant and impeccably behaved proteins could readily be targeted. In contrast, the modern availability of gene editing tools allows the insertion of affinity tags at endogenous gene loci, greatly facilitating the purification of less abundant complexes, even without full prior knowledge of their composition.

Between these two extremes lies a Goldilocks strategy for sample preparation: the use of unpurified cell extracts from which single-particle cryoEM data can simultaneously be obtained on multiple proteins and complexes that are likely to be present in close-to-native states (Figure 3). This approach promises to bridge *in vitro* and *in cellulo* structural biology, accelerating the

accumulation of reference data for the interpretation of *in cellulo* structural data [25]. For example, the cryoEM and integrative structural analysis of keto-acid dehydrogenase complex family members (i.e., the oxoglutarate [26] and pyruvate dehydrogenase [27,28] complexes) revealed the presence and locations of several previously unseen subunits. Notably, these studies revealed differences in stoichiometry and symmetry for several subunits, compared to a recent high-resolution structure of a related subcomplex [29]. This observation demonstrates how data from 'raw' cell lysates can add value beyond what can be obtained from purified recombinant complexes and brings us closer to the interpretation of *in cellulo* images.

Extracts can thus be used as a resource to derive structures of biomolecular complexes in a parallel manner without purification [30–35]. Considering that thousands of proteins are retrieved in such extracts [28], a possibility exists to resolve hundreds of complexes within the extracts. It is also likely that the addition of a modest level of fractionation would simplify image analysis, particularly if combined with the use of mass spectrometry to identify the composition of each fraction. This approach may prove particularly valuable for rarer or substoichiometric complexes, which will provide significant challenges for structural biology *in cellulo*.

Along similar lines, Jijumon et al. combined light microscopy, cryoEM, and functional analysis of microtubules in lysates to systematically characterize 45 microtubule associated proteins (MAPs), some of which had not been observed previously [36]. This study developed a novel approach to analyze microtubule-associated proteins using lysates of mammalian cells, uncovering activities of MAPs that lead to unique microtubule behaviors such as coiling, hook formation, and liquid–liquid phase separation (which initiates microtubule branching). Such methodological breakthroughs underscore the value of working with lysates, and the potential for higher resolution cryoEM models derived from such samples to be placed into the context of the intact parent cells is clear.

Overall, lysates strike a balance between throughput, achievable resolution and sample complexity that make them a valuable staging post between the highest resolution cryoEM of purified proteins and the fully native state represented in cryoET.

Progress in the identification of structural signatures

In all the approaches discussed above, the molecular identities of the captured structural signatures are unknown. Unfortunately, there is currently no mass spectrometry method to combine with cryo-ET or cellular data so that each protein or complex in an image can be identified. Indeed, it is even a considerable image processing challenge to group distinct views of chemically similar particles when only a small number of such particles exist in a sample. In cell extracts, cross-correlation of mass spectrometry data with structural signatures can be performed [28], but again, unambiguously identifying the observed structural signatures is not trivial and has been covered elsewhere [37].

To begin to address these challenges, new algorithms that integrate artificial intelligence for the analysis of cryoEM maps have been reported. Software such as DeepTracer [38,39], ModelAngelo [40] and findMySequence [41] allow tracing of a protein backbone onto high-resolution density maps with impressive results. For example, ModelAngelo was able to build >100,000 residues (of the total of 158,000 deposited in the PDB for the monomeric form) into the 3.3 Å map of the algal phycobilisome structure described above. Currently, however, these methods typically require high-resolution data, which are challenging to obtain from *in situ* studies or studies of cell extracts. Related methods that are better suited to lower resolution data are also being reported, such as the systematic fitting and scoring of AlphaFold models from the organism's proteome [42,43], or the tracing of backbones combined with the varying of amino acid identities, as described for findMySequence [41] with cryo-EM maps derived from native cell extracts [30]. Overall, a range of new experimental and computational tools are being actively developed to allow the community to interpret density maps across different resolutions.

However, with advances in spatially resolved mass spectrometry, it is conceivable that this hurdle might be addressed.

Integration of crosslinking for structural analysis of in-cell complexes

Of all of the other experimental methods that could be brought to bear on the field of structural cell biology, crosslinking combined with mass spectrometry (XLMS) is perhaps one of the most likely to provide data that are complementary to that contributed by electron microscopy or X-ray crystallography in a closer-to-native environment.

XLMS has come to prominence for its ability to provide residue-specific proximity information for individual proteins or purified protein complexes. Such information is particularly valuable for complexes that have not proven amenable to traditional structural approaches [44], such as the nucleosome remodelling and deacetylase complex [45], but XLMS has also been used to assist in the placement of complex subunits or domains into moderate-resolution cryoEM maps [46–48] or to inform on complex stoichiometry and protein conformation and dynamics [49,50]. The modest sample

requirements and high-throughput nature of the method has inevitably led to applications with an eye to system-wide discovery and again the advent of AlphaFold has opened new horizons in the field. One study described XLMS analysis of a set of four organellar lysates, which yielded ∼28,000 crosslinks across 4,000 proteins, identifying >2,000 unique protein interactions [51]. Concurrent AlphaFold analysis demonstrated the power of XLMS to discover new interactions and to corroborate AlphaFold predictions [51], a strategy that should have a significant impact in the interpretation of the complex structural datasets described above. It was also suggested that data of this type could also be directly integrated into AlphaFold predictions: an idea that was quickly realized with the development of AlphaLink [52]. Modelling with restraints can be performed with traditional methods, such as HADDOCK [53], but, recent, AI-only algorithms open a new avenue for proteome-wide predictions [54,55].

Whereas the study from Bartolec et al. closely aligns with the cell extract cryoEM approach, other groups have pushed XLMS closer to the cryoET realm by carrying out the initial chemical crosslinking step in living cells [56,57]. For example, the Rappsilber and Stülke laboratories integrated experiments and AlphaFold to gain insight into bacterial protein−protein interactions (PPIs) [58], directly accessing the proteome in its native context. Recently, XLMS identified transcription elongation factors NusG and NusA which are at the interface of the ribosome and the RNA polymerase in the *in situ* characterized expressome complex [59]. These XLMS data were visualized within the *in situ* cryoET structure of the expressome from the human pathogen *Mycoplasma pneumoniae*.

The differential abundance of proteins and complexes in the cell represents a significant challenge for XLMS. For example, ∼60% of the ∼14,000 XLs observed in Wheat et al. mapped to 3 protein interaction networks, including the histone, chaperonin and ubiquitin-proteasomal networks [57]. In the future, depletion of such complexes (*e.g.*, by the use of antibodies or the introduction of endogenous epitope tags) could increase the depth of coverage achievable in proteome wide studies.

A question that hovers over every high-throughput analysis is that of false positives and XLMS is likely no exception to this concern. How many XLs identified in a proteome-wide XLMS experiment report on 'true' interactions? It is difficult to precisely estimate this number, given that the experiment has been set up explicitly to identify novel interactions. Nevertheless, >90% of the XLs observed in our recent XLMS analysis of organelle lysates connected proteins already known to interact, meaning that a maximum of 10% (likely much less) might be false positives. Additional confidence could be obtained through the creation of a searchable database that houses all XLMS data; such a database could also harvest the ever-increasing range of experimental (*e.g.*, high-throughput gene knockout and IPMS datasets) and computational data (*e.g.*, gene ontology analysis, AlphaFold) to provide probability estimates for a newly identified interaction. Resources such as BioGRID already organize much of this information and are accessible through an API.

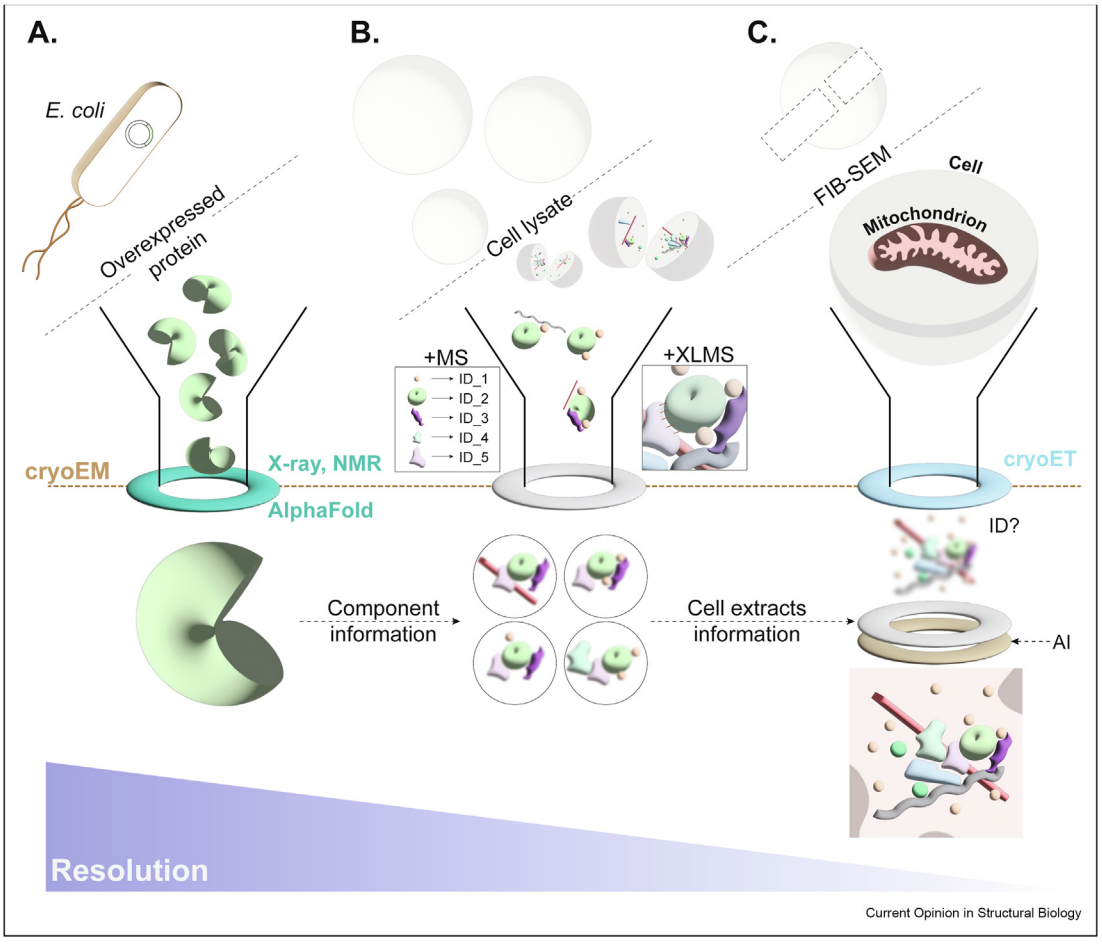
The more complex a sample becomes, the more challenging the data analysis, and XLMS is no exception. Coverage in the studies above was typically only one or a handful of XLs per interaction, enough to corroborate an AlphaFold prediction or experimental data, but not to provide deep interrogation of the interactome (for example, the Human Reference Interactome map contains ∼53,000 interactions) [60]. However, the relentless improvements in MS hardware combined with combinations of more sophisticated crosslinkers, *e.g.*, photocrosslinkers, and data analysis packages will continue to expand the boundaries of what can be achieved with this approach.

Beyond understanding proteins and protein complexes

Although the primary focus of this article has been proteins and protein complexes, it is of course the case that every cell contains a plethora of molecules ranging from individual ions to chromosomes with molecular weights of 100 billion Daltons. These molecules are significantly underrepresented in the known landscape of structural cell biology − in many cases because their conformations are not as well-defined as those of globular proteins. Even within the protein realm, membrane proteins still pose a challenge (despite advances in cryo-EM), and advances in solubilizing agents would accelerate structural cell biology of membrane proteins. For example, native lipid-bilayer nanodiscs are critical for retrieving endogenous protein complexes [61] but improvements are still needed − *e.g.*, co-polymer nanodiscs of specific diameters that may span 10−500 nm−as is expansion of their utilization, *e.g.*, in capturing membrane-associated proteins as well. As noted above, another difficult issue is that of intrinsically disordered proteins, which will likely continue to be largely recalcitrant to crystallographic and EM analysis. Although NMR spectroscopy is perhaps the best suited method to probe such polypeptides, the method is limited in the complexity of the system that is amenable to detailed analysis. This is perhaps an area where computational approaches such as coarse-grained molecular dynamics can make a significant contribution in combination with experimental (such as XLMS) and bioinformatic restraints.

Available implementations of AI structure prediction programs have also largely focused on proteins,

Figure 4



(a) Integration of multi-scale structural data – from (A) *in vitro* to (C) *in cellulo*. **(b)** Cell extracts from the native source can act as a bridge to connect observations across scales. Note, that even if these methods are seamlessly integrated, cellular dynamics, protein complex dynamics, non-polypeptide chemical species, and disorder, are still out of reach to experimentally probe and integrate.

although a first implementation for the prediction protein-DNA interactions has recently been described in RoseTTAFoldNA [62]. Otherwise, co-factors and hydration layers have so far proven extremely challenging to define or predict at large scale, even though AlphaFill can effectively model co-factors by homology [63]. To date, AI algorithms that strive to predict structures beyond individual polypeptide chains have not performed at the same high level, but this is clearly an area of intense interest given the success of the protein-only programs. The training of AIs for the prediction of protein-small molecule complexes is hampered by the much smaller training set available in the PDB and the vast chemical space of small molecules; in this regard, it would be enormously valuable for the entire community to have access to the (most likely) hundreds of thousands of structures of protein-small molecule complexes that sit quietly behind the walls of commercial organizations. Therefore, defining protein complexes and their higher order assemblies with other "molecules of life" currently remains a somewhat distant goal even if Alphafold 3 [64] is reported to predict those with relative success.

Conclusions

What are the prospects for a structural description of a cell? If we consider this question from the perspective of the first people who stared down a microscope and saw individual cells, one could argue that we are there already. Through the combination of the dazzling array of cellular, molecular and computational techniques that have been developed in the past 80 or so years, only some of which we have touched on here, we have developed an incredible wealth of knowledge regarding the compositions, locations, quantities and structures that together make a cell what it is.

As our understanding has become more sophisticated, however, the goalposts have shifted. There is clearly a

long way to go if we seek what could be thought of as a PDB entry with the title "HEK293 cell". Nevertheless, the pace of development has increased rather than slowed in the last 10–20 years – perhaps most prominently the convergence of cryo-electron microscopy – and artificial intelligence has dramatically expanded our access to structures of enormous complexity. Combination of deep learning approaches to identify structural patterns in cellular tomograms, cellular micrographs or images from cell extracts, and macromolecular identification methods that can cover different resolution regimes (augmented by AlphaFold and its predicted complexes, *e.g.*, with AlphaPulldown [65]) are expected to transform structural cell biology (Figure 4). This is because the integration of various methods across resolution scales will harvest both spatial information but also the local environment of macromolecular complexes within the cell. Cryo-EM and AI are also benefiting from hardware advances in GPU-based computing, which are greatly expediting the analysis and management of vast amounts of data.

Moreover, integrative structural modeling, a technique that seeks to amalgamate diverse data sources into cohesive models, has substantial promise for bridging the gap between individual structures and cellular assemblies, though there is a need to establish standardized accuracy and quality criteria to gain a firmer foothold. This type of approach is likely to be crucial for developing a systematic approach to visualizing cryo-electron tomograms beyond merely pinpointing well-known and abundant biomolecules. Along the same lines, the future demands a greater availability of standardized formatting and extensive data not only for cryo-electron tomography and related techniques but also for various other types of biomolecules. In this context, open-access databases like PDB, EMDB, PRIDE, Uniprot, EMPIAR, AlphaFold-DB, NAKB, STRING, and similar repositories will play a behind-the-scenes yet pivotal role in eventually enabling the comprehensive visualization of biomacromolecules in their natural context.

And beyond a static snapshot of the cell, any structural representation of the cell must ultimately incorporate motion: protein conformational dynamics from intrinsically disordered regions and corresponding motions from all other molecules in the cell, not to mention diffusion and directed motions (*e.g.*, of motor proteins). Here perhaps more than anywhere, we await gains in processing power – perhaps when quantum computing becomes a reality. There is thus no concern that structural biologists, computer scientists and hardware innovators in this space will be out of business any time soon.

CRediT author statement

Fotis L. Kyrilis: Writing- Reviewing and Editing, Visualization, **Jason K K Low:** Writing- Reviewing and Editing, **Joel P Mackay:** Conceptualization , Writing- Original draft preparation, Funding acquisition, **Panagiotis L Kastritis:** Conceptualization, Writing- Original draft preparation, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgements

We would like to thank all the Kastritis lab members for fruitful discussions. We thank Dr. Xin You and Professor Sen-Fang Sui from the State Key Laboratory of Membrane Biology of Tsinghua University (Beijing, China) for providing data to illustrate their related work in Figure 2 (Panels A, B and D). This work was supported by the European Union through funding of the Horizon Europe ERA Chair "hot4cryo" project number 101086665 (to P.L.K.), the Federal Ministry of Education and Research (BMBF, ZIK program) (Grant nos. 03Z22HN23, 03Z22HI2 and 03COV04 to P.L.K.), the European Regional Development Funds (EFRE) for Saxony-Anhalt (Grant no. ZS/2016/04/78115 to P.L.K.), the German Research Foundation (project number 391498659, RTG 2467), and the Martin-Luther University of Halle-Wittenberg.

References

Papers of particular interest, published within the period of review, have been highlighted as:

* of special interest
** of outstanding interest

1. De Rosier DJ, Klug A: **Reconstruction of three dimensional structures from electron micrographs**. *Nature* 1968, **217**: 130–134.

2. Hart RG: **Electron microscopy of unstained biological material: the polytropic montage**. *Science* 1968, **159**:1464–1467.

3. Hoppe W, Langer R, Knesch G, Poppe C: **Protein crystal structure analysis with electron radiation**. *Naturwissenschaften* 1968, **55**:333–336.

4. Mahamid J, Pfeffer S, Schaffer M, Villa E, Danev R, Cuellar LK, Forster F, Hyman AA, Plitzko JM, Baumeister W: **Visualizing the molecular sociology at the HeLa cell nuclear periphery**. *Science* 2016, **351**:969–972.

5. Xing H, Taniguchi R, Khusainov I, Kreysing JP, Welsch S, Turonova B, Beck M: **Translation dynamics in human cells visualized at high resolution reveal cancer drug action**. *Science* 2023, **381**:70–75.

6. ** Mosalaganti S, Obarska-Kosinska A, Siggel M, Taniguchi R, Turonova B, Zimmerli CE, Buczak K, Schmidt FH, Margiotta E, Mackmull MT, *et al.*: **AI-based structure prediction empowers integrative structural analysis of human nuclear pores**. *Science* 2022, **376**, eabm9506.
This paper combines Alphafold and cryo-electron tomography to propose one of the most complete architectures of the human nuclear pore to date: a 70-MDa atomically resolved model that covers >90% of the scaffold.

7. Berger C, Ravelli RBG, Lopez-Iglesias C, Kudryashev M, Diepold A, Peters PJ: **Structure of the Yersinia injectisome in intracellular host cell phagosomes revealed by cryo FIB electron tomography**. *J Struct Biol* 2021, **213**, 107701.

8. Hoffmann PC, Kreysing JP, Khusainov I, Tuijtel MW, Welsch S, Beck M: **Structures of the eukaryotic ribosome and its translational states in situ**. *Nat Commun* 2022, **13**:7435.

9. Tegunov D, Xue L, Dienemann C, Cramer P, Mahamid J: **Multi-particle cryo-EM refinement with M visualizes ribosome-antibiotic complex at 3.5 A in cells**. *Nat Methods* 2021, **18**: 186–193.

10. Albert S, Schaffer M, Beck F, Mosalaganti S, Asano S, Thomas HF, Plitzko JM, Beck M, Baumeister W, Engel BD: **Proteasomes tether to two distinct sites at the nuclear pore complex**. *Proc Natl Acad Sci U S A* 2017, **114**:13726–13731.

11. * Serwas D, Akamatsu M, Moayed A, Vegesna K, Vasan R, Hill JM, Schoneberg J, Davies KM, Rangamani P, Drubin DG: **Mechanistic insights into actin force generation during vesicle formation from cryo-electron tomography**. *Dev Cell* 2022, **57**: 1132–1145. e1135.

This study uses cryo-ET to unveil the native-state intricacies of force-producing actin networks in endocytosis, highlighting the initiation of branched filament assembly and quantifying the robustness of actin force production during significant network variability.

12. Chen Z, Shiozaki M, Haas KM, Skinner WM, Zhao S, Guo C, Polacco BJ, Yu Z, Krogan NJ, Lishko PV, *et al.*: **De novo protein identification in mammalian sperm using in situ cryoelectron tomography and AlphaFold2 docking**. *Cell* 2023, **186**: 5041–5053. e5019.

13. Zhao Y, Wang H, Wiesehoefer C, Shah NB, Reetz E, Hwang JY, Huang X, Wang TE, Lishko PV, Davies KM, *et al.*: **3D structure and in situ arrangements of CatSper channel in the sperm flagellum**. *Nat Commun* 2022, **13**:3439.

14. ** Dutta D, Nguyen V, Campbell KS, Padron R, Craig R: **Cryo-EM structure of the human cardiac myosin filament**. *Nature* 2023, **623**:853–862.

This paper provides the molecular basis behind cardiac contraction, a groundbreaking paradigm for understanding and potentially treating heart failure caused by mutations in the essential proteins – including at the interaction interfaces that combine to form the structure of the muscle.

15. ** Tamborrini D, Wang Z, Wagner T, Tacke S, Stabrin M, Grange M, Kho AL, Rees M, Bennett P, Gautel M, *et al.*: **Structure of the native myosin filament in the relaxed cardiac sarcomere**. *Nature* 2023, **623**:863–871.

This is the back-to-back paper to Ref. [14], offering insights into the specialized capacities of myosin and titin along the cardiac filament, and providing the molecular basis to understand muscle disorders, particularly familial hypertrophic cardiomyopathy.

16. Zimmerli CE, Allegretti M, Rantos V, Goetz SK, Obarska-Kosinska A, Zagoriy I, Halavatyi A, Hummer G, Mahamid J, Kosinski J, *et al.*: **Nuclear pores dilate and constrict in cellulo**. *Science* 2021, **374**, eabd9776.

17. Berger C, Premaraj N, Ravelli RBG, Knoops K, Lopez-Iglesias C, Peters PJ: **Cryo-electron tomography on focused ion beam lamellae transforms structural cell biology**. *Nat Methods* 2023, **20**:499–511.

18. Eisenstein F, Fukuda Y, Danev R: **Smart parallel automated cryo electron tomography**. *bioRxiv* 2023, https://doi.org/10.1101/2023.12.14.571776:2023.2012.2014.571776.

19. Rice G, Wagner T, Stabrin M, Sitsel O, Prumbaum D, Raunser S: **TomoTwin: generalized 3D localization of macromolecules in cryo-electron tomograms with structural data mining**. *Nat Methods* 2023, **20**:871–880.

20. Zeng X, Kahng A, Xue L, Mahamid J, Chang YW, Xu M: **High-throughput cryo-ET structural pattern mining by unsupervised deep iterative subtomogram clustering**. *Proc Natl Acad Sci U S A* 2023, **120**, e2213149120.

21. Elferich J, Schiroli G, Scadden DT, Grigorieff N: **Defocus Corrected Large Area Cryo-EM (DeCo-LACE) for label-free detection of molecules across entire cell sections**. *Elife* 2022, **11**.

22. Lucas BA, Himes BA, Xue L, Grant T, Mahamid J, Grigorieff N: **Locating macromolecular assemblies in cells by 2D template matching with cisTEM**. *Elife* 2021, **10**.

23. ** You X, Zhang X, Cheng J, Xiao Y, Ma J, Sun S, Zhang X, Wang HW, Sui SF: **In situ structure of the red algal phycobilisome-PSII-PSI-LHC megacomplex**. *Nature* 2023, **616**: 199–206.

The authors combine single-particle cryo-EM within the red algal cell with cryo-electron tomography to understand how light capture and energy transfer occurs in a complete photosynthetic supercomplex. It is staggering when one realizes that authors propose models that include ~1.0–2.5 million atoms of proteins and cofactors.

24. Milo R: **What is the total number of protein molecules per cell volume? A call to rethink some published values**. *Bioessays* 2013, **35**:1050–1055.

25. Kyrilis FL, Meister A, Kastritis PL: **Integrative biology of native cell extracts: a new era for structural characterization of life processes**. *Biol Chem* 2019, **400**:831–846.

26. * Skalidis I, Kyrilis FL, Tuting C, Hamdi F, Trager TK, Belapure J, Hause G, Fratini M, O'Reilly FJ, Heilmann I, *et al.*: **Structural analysis of an endogenous 4-megadalton succinyl-CoA-generating metabolon**. *Commun Biol* 2023, **6**:552.

Here, authors have performed an integrative structural biology approach in native cell extracts to understand the architecture of a functional oxoglutarate dehydrogenase complex. They show that substrate shuttling is regulated by complementary electrostatics and a high abundance of disorder-driven protein flexibility.

27. Tuting C, Kyrilis FL, Muller J, Sorokina M, Skalidis I, Hamdi F, Sadian Y, Kastritis PL: **Cryo-EM snapshots of a native lysate provide structural insights into a metabolon-embedded transacetylase reaction**. *Nat Commun* 2021, **12**:6933.

28. Kyrilis FL, Semchonok DA, Skalidis I, Tuting C, Hamdi F, O'Reilly FJ, Rappsilber J, Kastritis PL: **Integrative structure of a 10-megadalton eukaryotic pyruvate dehydrogenase complex from native cell extracts**. *Cell Rep* 2021, **34**, 108727.

29. Forsberg BO: **The structure and evolutionary diversity of the fungal E3-binding protein**. *Commun Biol* 2023, **6**:480.

30. ** Skalidis I, Kyrilis FL, Tuting C, Hamdi F, Chojnowski G, Kastritis PL: **Cryo-EM and artificial intelligence visualize endogenous protein community members**. *Structure* 2022, **30**: 575–589. e576.

This paper was the first to combine Alphafold models with multiple maps generated from native cell extracts. The authors have performed "blind" protein modelling, where AI tools are driving the modelling procedures.

31. Su CC, Lyu M, Morgan CE, Bolla JR, Robinson CV, Yu EW: **A 'Build and Retrieve' methodology to simultaneously solve cryo-EM structures of membrane proteins**. *Nat Methods* 2021, **18**:69–75.

32. Ho CM, Li X, Lai M, Terwilliger TC, Beck JR, Wohlschlegel J, Goldberg DE, Fitzpatrick AWP, Zhou ZH: **Bottom-up structural proteomics: cryoEM of protein complexes enriched from the cellular milieu**. *Nat Methods* 2020, **17**:79–85.

33. ** Su CC, Lyu M, Zhang Z, Miyagi M, Huang W, Taylor DJ, Yu EW: **High-resolution structural-omics of human liver enzymes**. *Cell Rep* 2023, **42**, 112609.

The authors have provided molecular structures from cell extracts derived from human liver. This is the first application of cryo-EM of endogenous, native cell extracts to a human organ, opening routes to understand holistically disease phenotypes at the molecular level.

34. Morgan CE, Zhang Z, Miyagi M, Golczak M, Yu EW: **Toward structural-omics of the bovine retinal pigment epithelium**. *Cell Rep* 2022, **41**, 111876.

35. Semchonok DA, Kyrilis FL, Hamdi F, Kastritis PL: **Cryo-EM of a heterogeneous biochemical fraction elucidates multiple protein complexes from a multicellular thermophilic eukaryote**. *J Struct Biol X* 2023, **8**, 100094.

36. Jijumon AS, Bodakuntla S, Genova M, Bangera M, Sackett V, Besse L, Maksut F, Henriot V, Magiera MM, Sirajuddin M, *et al.*: **Lysate-based pipeline to characterize microtubule-**

associated proteins uncovers unique microtubule behaviours. *Nat Cell Biol* 2022, **24**:253–267.

37. Kyrilis FL, Belapure J, Kastritis PL: **Detecting protein communities in native cell extracts by machine learning: a structural biologist's perspective**. *Front Mol Biosci* 2021, **8**, 660542.

38. Chang L, Wang F, Connolly K, Meng H, Su Z, Cvirkaite-Krupovic V, Krupovic M, Egelman EH, Si D: **DeepTracer-ID: de novo protein identification from cryo-EM maps**. *Biophys J* 2022, **121**:2840–2848.

39. Pfab J, Phan NM, Si D: **DeepTracer for fast de novo cryo-EM protein structure modeling and special studies on CoV-related complexes**. *Proc Natl Acad Sci U S A* 2021, **118**.

40. ** Jamali K, Kall L, Zhang R, Brown A, Kimanius D, Scheres SHW: **Automated model building and protein identification in cryo-EM maps**. *bioRxiv* 2023, https://doi.org/10.1101/2023.05.16.541002.
Modelangelo has changed the way we perform model building in high-resolution cryo-EM maps; now, AI tools are rapidly becoming valuable assistants in the atomic modelling and significantly reducing the time required to solve a macromolecular structure. Following their application, a few rounds of Coot/Phenix refinement are enough for deriving a high-quality atomic model.

41. Chojnowski G, Simpkin AJ, Leonardo DA, Seifert-Davila W, Vivas-Ruiz DE, Keegan RM, Rigden DJ: **findMySequence: a neural-network-based approach for identification of unknown proteins in X-ray crystallography and cryo-EM**. *IUCrJ* 2022, **9**:86–97.

42. Tuting C, Schmidt L, Skalidis I, Sinz A, Kastritis PL: **Enabling cryo-EM density interpretation from yeast native cell extracts by proteomics data and AlphaFold structures**. *Proteomics* 2023, **23**, e2200096.

43. He J, Lin P, Chen J, Cao H, Huang SY: **Model building of protein complexes from intermediate-resolution cryo-EM maps with deep learning-guided automatic assembly**. *Nat Commun* 2022, **13**:4066.

44. Piersimoni L, Kastritis PL, Arlt C, Sinz A: **Cross-Linking mass spectrometry for investigating protein conformations and protein-protein interactions horizontal line A method for all seasons**. *Chem Rev* 2022, **122**:7500–7531.

45. Low JKK, Silva APG, Sharifi Tabar M, Torrado M, Webb SR, Parker BL, Sana M, Smits C, Schmidberger JW, Brillault L, *et al.*: **The nucleosome remodeling and deacetylase complex has an asymmetric, dynamic, and modular architecture**. *Cell Rep* 2020, **33**, 108450.

46. Zhang Z, Will CL, Bertram K, Dybkov O, Hartmuth K, Agafonov DE, Hofele R, Urlaub H, Kastner B, Luhrmann R, *et al.*: **Molecular architecture of the human 17S U2 snRNP**. *Nature* 2020, **583**:310–313.

47. Flacht L, Lunelli M, Kaszuba K, Chen ZA, Reilly FJO, Rappsilber J, Kosinski J, Kolbe M: **Integrative structural analysis of the type III secretion system needle complex from Shigella flexneri**. *Protein Sci* 2023, **32**, e4595.

48. Ziemianowicz DS, Saltzberg D, Pells T, Crowder DA, Schrader C, Hepburn M, Sali A, Schriemer DC: **IMProv: a resource for cross-link-driven structure modeling that accommodates protein dynamics**. *Mol Cell Proteomics* 2021, **20**, 100139.

49. Yuan S, Xia L, Wang C, Wu F, Zhang B, Pan C, Fan Z, Lei X, Stevens RC, Sali A, *et al.*: **Conformational dynamics of the activated GLP-1 receptor-G(s) complex revealed by cross-linking mass spectrometry and integrative structure modeling**. *ACS Cent Sci* 2023, **9**:992–1007.

50. Felker D, Lee K, Pospiech TH, Morishima Y, Zhang H, Lau M, Southworth DR, Osawa Y: **Mapping interactions of calmodulin and neuronal NO synthase by crosslinking and mass spectrometry**. *J Biol Chem* 2024, **300**, 105464.

51. Bartolec TK, Vazquez-Campos X, Norman A, Luong C, Johnson M, Payne RJ, Wilkins MR, Mackay JP, Low JKK: **Cross-linking mass spectrometry discovers, evaluates, and corroborates structures and protein-protein interactions in the human cell**. *Proc Natl Acad Sci U S A* 2023, **120**, e2219418120.

52. * Stahl K, Graziadei A, Dau T, Brock O, Rappsilber J: **Protein structure prediction with in-cell photo-crosslinking mass spectrometry and deep learning**. *Nat Biotechnol* 2023, **41**:1810–1819.
Alphalink was introduced here and it is the first-in-class algorithm to demonstrate feasibility in combining XLMS data in Alphafold predictions on a proteome scale.

53. van Zundert GCP, Rodrigues J, Trellet M, Schmitz C, Kastritis PL, Karaca E, Melquiond ASJ, van Dijk M, de Vries SJ, Bonvin A: **The HADDOCK2.2 web server: user-friendly integrative modeling of biomolecular complexes**. *J Mol Biol* 2016, **428**:720–725.

54. ** Schweke H, Pacesa M, Levin T, Goverde CA, Kumar P, Duhoo Y, Dornfeld LJ, Dubreuil B, Georgeon S, Ovchinnikov S, *et al.*: **An atlas of protein homo-oligomerization across domains of life**. *Cell* 2024, https://doi.org/10.1016/j.cell.2024.01.022.
Here, the authors predict homo-oligomeric assemblies across proteomes and offer a valuable structural context for interpreting disease mutations. Authors highlight coiled-coils as contributors to the evolution of quaternary protein structures and provide a comprehensive insight into homo-oligomerization.

55. Shor B, Schneidman-Duhovny D: **CombFold: predicting structures of large protein assemblies using a combinatorial assembly algorithm and AlphaFold2**. *Nat Methods* 2024, https://doi.org/10.1038/s41592-024-02174-0.

56. Lee K, O'Reilly FJ: **Cross-linking mass spectrometry for mapping protein complex topologies in situ**. *Essays Biochem* 2023, **67**:215–228.

57. Wheat A, Yu C, Wang X, Burke AM, Chemmama IE, Kaake RM, Baker P, Rychnovsky SD, Yang J, Huang L: **Protein interaction landscapes revealed by advanced in vivo cross-linking-mass spectrometry**. *Proc Natl Acad Sci U S A* 2021, **118**.

58. O'Reilly FJ, Graziadei A, Forbrig C, Bremenkamp R, Charles K, Lenz S, Elfmann C, Fischer L, Stulke J, Rappsilber J: **Protein complexes in cells by AI-assisted structural proteomics**. *Mol Syst Biol* 2023, **19**, e11544.

59. O'Reilly FJ, Xue L, Graziadei A, Sinn L, Lenz S, Tegunov D, Blotz C, Singh N, Hagen WJH, Cramer P, *et al.*: **In-cell architecture of an actively transcribing-translating expressome**. *Science* 2020, **369**:554–557.

60. Luck K, Kim DK, Lambourne L, Spirohn K, Begg BE, Bian W, Brignall R, Cafarelli T, Campos-Laborie FJ, Charloteaux B, *et al.*: **A reference map of the human binary protein interactome**. *Nature* 2020, **580**:402–408.

61. Janson K, Kyrilis FL, Tuting C, Alfes M, Das M, Trager TK, Schmidt C, Hamdi F, Vargas C, Keller S, *et al.*: **Cryo-electron microscopy snapshots of eukaryotic membrane proteins in native lipid-bilayer nanodiscs**. *Biomacromolecules* 2022, **23**:5084–5094.

62. Baek M, McHugh R, Anishchenko I, Jiang H, Baker D, DiMaio F: **Accurate prediction of protein-nucleic acid complexes using RoseTTAFoldNA**. *Nat Methods* 2024, **21**:117–121.

63. Hekkelman ML, de Vries I, Joosten RP, Perrakis A: **AlphaFill: enriching AlphaFold models with ligands and cofactors**. *Nat Methods* 2023, **20**:205–213.

64. ** Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, *et al.*: **Accurate structure prediction of biomolecular interactions with AlphaFold 3**. *Nature* 2024, https://doi.org/10.1038/s41586-024-07487-w.
The authors present the updated architecture of Alphafold 3, that can predict 3D structures of biomolecules; It can also predict post-translational modifications, co-factors and ligands, as well as proteins in complex with other biomolecules, such as DNA and RNA. Currently, a server is available to test basic functions of Alphafold 3, available here: http://alphafoldserver.com.

65. Yu D, Chojnowski G, Rosenthal M, Kosinski J: **AlphaPulldown-a python package for protein-protein interaction screens using AlphaFold-Multimer**. *Bioinformatics* 2023, **39**.