Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

Visualizing everything, everywhere, all at once: Cryo-EM and the new field of structureomics

Bronwyn A. Lucas[1,2,3]

Abstract

Twenty years ago, the release of the first draft of the human genome sequence instigated a paradigm shift in genomics and molecular biology. Arguably, structural biology is entering an analogous era, with availability of an experimentally determined or predicted molecular model for almost every protein-coding gene from many genomes—producing a reference "structureome". Structural predictions require experimental validation and not all proteins conform to a single structure, making any reference structureome necessarily incomplete. Despite these limitations, a reference structureome can be used to characterize cell state in more detail than by quantifying sequence or expression levels alone. Cryogenic electron microscopy (cryo-EM) is a method that can generate atomic resolution views of molecules and cells frozen in place. In this perspective I consider how emerging cryo-EM methods are contributing to the new field of structureomics.

Addresses

[1] RNA Therapeutics Institute, University of Massachusetts Chan Medical School, Worcester, MA, USA
[2] Department of Molecular and Cell Biology, University of California Berkeley, Berkeley, CA, USA
[3] Center for Computational Biology, University of California Berkeley, Berkeley, CA, USA

Corresponding author: Lucas, Bronwyn A. (bronwynlucas@berkeley.edu)
(Lucas B.A.)

Current Opinion in Structural Biology 2023, **81**:102620

This review comes from a themed issue on **Cryo-electron microscopy (2023)**

Edited by **Alexis Rohou** and **Peijun Zhang**

For a complete overview see the Issue and the Editorial

Available online xxx

https://doi.org/10.1016/j.sbi.2023.102620

The structureome

The term "structure-ome" was initially coined by Allison Doerr [1] with reference to limiting digestion proteolysis and mass-spectrometry (LiP-MS) developed in the lab of Paula Picotti [2]. The last two years have seen a dramatic expansion in the number of available protein structural models driven by improvements in machine learning algorithms [3,4]. The resulting predicted "structureome" has yielded, for the first time, at least one prediction for the 3D conformation of most proteins in the human proteome, as well as the proteomes of an expanding list of other organisms [5,6]. In this perspective I use the term "structureome" to denote a reference collection of experimentally determined and computationally predicted molecular models spanning the proteomes of organisms.

Expanding the structureome

The availability of predicted structures has already accelerated existing pipelines for structure determination. A reference structure can be used as a starting point for structural biologists to build models into density from cryo-EM or crystallographic maps, to indicate candidate residues to mutate in functional studies and as constraints in simulations when investigating molecule motion [7]. Recently, algorithms have been developed to update Alphafold protein structure predictions using the fit into densities derived from cryo-EM [8] or X-ray crystallography [9] experiments as constraints to constrain the prediction. Approaches such as these can be used to update the reference structureome and improve future prediction models to build a progressively more complete parts list of the cell.

Proteins rarely function alone but form part of dynamic and heterogeneous complexes that often perform different functions depending on their composition. Possible protein-protein interactions (PPIs) depend on the structure of the interaction interface and are therefore encoded in the amino acid sequences of the component proteins. Modified deep learning based protein folding predictions have been adapted and trained on complexes to predict PPIs [10,11]. There are orders of magnitude more possible PPIs than exist in nature, therefore characterizing cellular complexes requires experiments in combination with predictions. Cross-linking mass-spectrometry (CLMS), in which interactions between and within proteins are captured by covalent cross-linking before cell lysis, can validate predicted interactions and provide additional constraints on the prediction of higher order complexes [12].

Unsupervised cryo-EM particle classification tools can identify distinct complexes in a mixture. This has been used to examine crude lysate fractions and identify the

complexes by assessing their structures. The structureome provides additional constraints that can be used to narrow the range of possibilities and build models of protein complexes [13]. This may make it possible to identify proteins in less abundant complexes because lower resolution information will sufficiently narrow the range of possible structures.

Combining these approaches with single cell proteomic approaches, or examining lysates of single cells using microfluidics could enable characterization of compositional heterogeneity at the single cell scale.

The structureome as a reference to study biological heterogeneity

Individual complexes function together with other complexes and may have different functions in different cellular contexts. Moreover, the higher order spatial organization of these complexes within the cell can affect the outcome of biological pathways. Therefore, many important biological questions are impossible to answer by examining proteins or complexes in isolation. How does the cell produce distinct responses to different stimuli using a common signaling cascade? How do membraneless organelles form? How are molecules organized within organelles? Are the enzymes in metabolic pathways spatially coordinated? What are the conformational changes that accompany changes in cell state that are not reflected in changes to abundance or post-translational modifications? Each of these questions cannot be addressed by examining proteins in isolation but require examining the structure, location, and interactions of multiple proteins at once.

Beyond accelerating structure determination, a reference structureome will change the types of questions that biologists can ask, using molecular models as a starting point to investigate new biological questions at atomistic detail across scales to characterize cellular and pathological states.

Visualizing structures in cells with *in situ* cryo-ET and subtomogram averaging

The local sub-cellular environment affects the composition and conformation of proteins and their complexes. Mass spectrometry can provide information on sequence and abundance of polypeptide fragments and can analyze thousands of fragments from the same sample, making it a powerful proteomic tool [14]. LiP-MS and CLMS are two approaches that have been used to characterize protein structure and interactions proteome-wide [15,16]. With further development in the structural interpretation of barcodes, LiP-MS could ultimately contribute to building a cell-type specific structureome. When combined with CLMS eventually, it may become possible to generate cell-wide interaction networks of protein complexes and examine the changes in these networks in response to physiological and pathological changes. However, mass-spectrometry approaches currently require relatively large quantities to generate high-resolution information and therefore do not capture subcellular differences.

Cryogenic electron microscopy (cryo-EM) can generate images of cells at atomic resolution. This makes cryo-EM a potentially powerful structureomic tool. By collecting cryo-EM images of a tilt series cryogenic electron tomography (cryo-ET) generates a 3D view of the cell at nanometer resolution [17]. However, the sensitivity of biological macromolecules to radiation damage and the difficulty to reliably recreate a 3D model of the cell from the tilt series makes it challenging to visualize and identify individual molecular structures at high resolution directly [17]. For highly repetitive and/or abundant structures, subtomogram averaging of multiple copies of the same complex can be used to reveal additional details that can help identify the protein components [18]. With sufficiently structurally homogenous particles subtomogram averages can achieve resolutions comparable to single particle averages [19,20] and therefore could be used to identify and model the component polypeptides de novo, a task that is increasingly facilitated by machine learning based tools [21].

Most proteins are not present at sufficiently high concentration to generate reconstructions at sufficiently high resolution by subtomogram averaging to read out their protein sequence. For these complexes, the structureome provides a rich dataset that can use existing molecular models to identify the component proteins in lower resolution subtomogram averages. This approach was recently used to build models of the massive nuclear pore complex that would have been challenging in the absence of structural models of the component proteins [22,23]. This approach has also been applied in cases wherein the component proteins were unknown. As one example, a model of the proteins and interactions that form the sperm axoneme was built by comparing the fit of all structures in the predicted structureome into an intermediate resolution subtomogram average [24]. To identify the most likely candidates, the authors used prior knowledge from expression and mass-spectrometry data. In the future, more extensive databases that capture splice variants, alternate conformational states, protein−protein interactions and higher order complexes could be used to characterize molecular complexes in more detail.

Further development of tomography to generate tomographic reconstructions at improved resolution will help in directly identifying the most likely structures that account for the density, without the need to perform subtomogram averaging. Combining cryo-ET with other approaches, such as CLMS and molecular dynamics—as was used to characterize the *Mycoplasma pneumoniae*

expressome [25]—as well as machine learning tools to improve classification of densities, can aid in the interpretation of tomograms and potentially provide a more detailed 3D view of the cellular structureome.

In situ structure characterization using template matching

Another potentially powerful approach for visualization of protein complexes in cells that will benefit greatly from the available structureome is template matching. In template matching, a molecular model is used to generate a template representing a hypothesis of how the molecule would appear in different orientations given the imaging parameters. This template is then cross correlated with the data, and the best matching orientation identified at each position searched. This was initially demonstrated in tomograms [26—28] and led to the concept of visual proteomics, whereby templates spanning the proteome could be used to annotate tomograms and build a visual representation of the arrangement of proteins in the cell [29]. Currently, subtomogram averaging is required to validate the targets identified by template matching in cryo-tomograms. To use template matching to interpret tomograms without averaging, new algorithms are needed to assign probabilities to single instances of particles to indicate the confidence of their classification by taking into account the uncertainties inherent in the tomographic reconstruction.

When models are available at high resolution template matching can also be used to locate macromolecular complexes in 2D cryo-EM images of molecules, cells or lamellae without the need to generate tomograms using an approach called high resolution or 2D-template matching (2DTM) [30—33]. Since a noise model is used to exclude false positives, validation does not require averaging, therefore 2DTM can identify single instances of low abundance complexes [30]. 2DTM can also be used to assess the relative similarity of a set of templates to a given population of molecules in the cell [33]. This implies that 2DTM could be used to identify the most likely structure from a set of hypothetical structures that reflects the cellular population without the need to generate 3D averages. The relative similarity of a given model to the cellular population could provide a metric that could be used to update or refine structural models, analogous to approaches developed for comparing models to 3D densities [8]. This could be used for de novo structure determination, improvement or validation of predicted models. A similar strategy has been successfully used to identify the most likely conformation of a single instance from molecular dynamics simulations in single particle cryo-EM [34].

A predicted structureome greatly expands the number of models available as potential templates. However, current template matching approaches in both 2D cryo-EM images and 3D cryo-EM tomograms only confidently distinguish complexes of >300 kDa from the cellular milieu and require the *in situ* structure to match the prediction at the resolution of the search [31]. Further developments to prediction algorithms that integrate machine learning with experimental data from cryo-EM and mass-spectrometry would generate larger templates that would expand the utility of these approaches. Incorporating molecular dynamics to generate additional candidate templates could improve the detection and classification of complexes *in situ*. Achieving this is limited by the high computational cost of exhaustive template matching. GPU acceleration and improvements in available hardware increased the speed of 2DTM by ~30-fold, to ~45 GPU hours per image per template, depending on the hardware used [32]. Exhaustive template matching in tomograms is even more costly because of the need for a fine search in three dimensions. Further developments that combine 2DTM with tomography on the same sample [32], context based priors and machine learning to narrow the search space and more sophisticated methods to separate true from false detections could increase the sensitivity of this approach. Ultimately, it could be possible to visualize the location, orientation conformation and interactions of all structured proteins and complexes in the cell of > 50 kDa simultaneously using this approach [53].

Illuminating the "dark" structureome

One of the most striking features of the predicted structureome is the abundance of protein sequences for which no fold could be predicted [35]. Of the amino acid sequences 30 residues or longer in the Alphafold2 database ~30% have confidence scores of <50 and are predicted to be unfolded [6,35]. This figure is consistent with prior estimates for the frequency of internally disordered regions (IDRs) in the proteome [35]. IDRs, by definition, represent a structural ensemble, and thus cannot be characterized by a single static structure, making them more challenging to characterize. A more complete structureome will aid the characterization of IDRs by (1) providing a prediction of the location of IDRs structureome wide, (2) aiding the interpretation of biophysical studies such as NMR by providing a structure for ordered protein regions of the protein, and (3) making the sheer abundance of IDRs in the structureome prominent.

Visualizing dynamic structures such as IDRs will require distinct techniques. Recent developments using in-cell NMR have been able to capture detailed conformational ensembles of small (<30 kDa) proteins over time but lack spatial resolution [36]. Expansion microscopy [37] is another approach that could be used to characterize structural ensembles. By chemical labelling specific amino acids with fluorophores the conformation of some proteins has been visualized at single nanometer resolution [38], which may be applicable to characterize single

instances in cells. A structureome could be used to identify unknown protein complexes by providing a set of references that could be compared to the data using template matching approaches. Since these techniques do not rely on an expected structure, they are suitable for visualizing different states. However, since expansion microscopy requires chemical fixation, and the expansion is not isotropic, it is unclear at what level of detail these conformations will be biologically relevant. Comparison with structures from cryo-fixed cells, such as by cryo-ET or with 2DTM, will be informative.

Higher throughput data collection strategies

To take advantage of structureomic approaches and generate reliable statistics, more systematic cryo-EM data collection strategies are needed that avoid the need for users to select areas of interest to image. Montage data collection strategies that enable cryo-EM data collection over large areas are a promising approach to address this issue [39–41]. DecoLace [39] has the additional benefit that it uses the images in the image series to adjust the defocus locally, avoiding the need to use a sacrificial area for tracking and enabling imaging of entire cell section.

Ultimately, the use of current cryo-EM data collection strategies to visualize the cellular structureome will be limited to the volume imageable with transmission EM. Serial sectioning may provide additional views of single cells or tissues, but current thinning methods using focused ion beams introduce damage that may limit the recovery of information [42,43]. Alternate methods to generate thin electron-transparent samples that permit imaging of multiple sections per cell with minimal damage would be beneficial. Scanning transmission EM (STEM) is an alternate imaging approach wherein the electron beam is focused to a probe that is scanned across the sample. The resolution of the measurement is determined by the probe diameter. STEM can be used to visualize larger volumes [44], and shows promise as a method for structure determination of purified complexes producing sub-nanometer reconstructions [45]. A greater understanding of radiation damage with STEM is required to determine the feasibility of imaging thick samples at high resolution. Further development is needed to evaluate this approach and develop improved models for image formation with this method that could be used to interpret signal from unknown complexes. Imaging of adherent cells is limited by generally exposing the same region of the cell in a lamella. Suspension cells fall onto the grid more randomly, but sample preparation is still biased towards central part of the cell. High-pressure freezing has potential to visualize a broader range of cell regions and cells within tissues but requires further development of workflows and milling approaches to be widely adopted [46,47].

A common format for integrating and viewing structureomic data

To make best use of the data generated by these diverse methods, new tools are required to effectively mine structureomic databases, provide structureomic comparisons, identify and characterize disordered features and to characterize biologically relevant differences within and between cells. Achieving this will require new data formats that enable integration across diverse methods.

The success of molecular biology to integrate RNA concentrations, protein binding, RNA–RNA interactions, higher order genomic architecture, splicing patterns, etc stems in part from the ability to collapse this complex data to a simple 1D format: a linear sequence of bases that can be mapped to the genome. These data can then be viewed as "tracks" on a genome viewer such as the UCSC Genome Browser [48] and multiple data types can be viewed simultaneously. The ability to integrate across diverse data types provides additional biological insight that would not otherwise have been apparent.

What would a structureome viewer look like? Integration and visualization of structural data is more challenging for a few reasons. Firstly, the multidimensional nature of biophysical, structural and cellular data is too complex to directly visualize together. Secondly, the practitioners of different biophysical techniques use different methods that output data in different formats and have different sets of caveats that make them difficult to directly compare. Machine learning approaches that use natural language processing could be used to integrate across different data formats and provide a mechanism for dimensionality reduction of the data and visualization in a new format such as manifold embedding , tSNE plots and/or other representations [49].

Finally, proteins do not represent the full suite of macromolecules in the cell. Protein complexes are integrated into membranes, which can facilitate their interactions or catalytic functions. Nucleic acids, in particular RNA, are present in the cell at high concentrations and can form scaffolds for the generation of large, often highly dynamic RNA-protein complexes [50]. The abundance of metabolites can reflect the activity of different enzymes and the flux through metabolic and catabolic pathways [51]. A more complete view of the cell requires integration of protein structureomic data with other -omic approaches. Achieving this requires a common language for describing and comparing these data.

Concluding remarks

In 2000, just before the release of the draft human genome sequence, Lander and Weinberg surveyed the

history of the study of heredity [52]. They foresaw that the 21st century would see a shift in focus from studying individual components to a more holistic view where the complexity of the cell could be directly studied. The first twenty years of the 21st century have seen this prediction born out in the study of nucleic acid and protein abundance. The next twenty years may see structural biology join this trend and enable the study of 3D structure at the omic scale to directly visualize and quantify a greater component of the complexity of the cell.

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.

Acknowledgments

I thank Nikolaus Grigorieff for helpful discussions and for critical reading of the manuscript. I thank the reviewers for helpful suggestions that improved the manuscript. I gratefully acknowledge funding from the Chan Zuckerberg Initiative, grant # 2021-234617 (5022).

References

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

* of special interest
** of outstanding interest

1. Doerr A: **Probing the protein structure-ome**. *Nat Methods* 2014, **11**:1088–1089.

2. Feng Y, Franceschi GD, Kahraman A, Soste M, Melnik A, Boersema PJ, Laureto PPD, Nikolaev Y, Oliveira AP, Picotti P: **Global analysis of protein structural changes in complex proteomes**. *Nat Biotechnol* 2014, **32**:1036–1044.

3. ** Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, *et al.*: **Highly accurate protein structure prediction with AlphaFold**. *Nature* 2021, **596**:583–589.
Describes the improved machine learning based algorithm Alphafold2 from Deep Mind that won CASP14 and greatly improved protein structure prediction from amino acid sequences.

4. ** Baek M, DiMaio F, Anishchenko I, Dauparas J, Ovchinnikov S, Lee GR, Wang J, Cong Q, Kinch LN, Schaeffer RD, *et al.*: **Accurate prediction of protein structures and interactions using a three-track neural network**. *Science* 2021, **373**:871–876.
An independent group was able to also improve structure predictions using similar strategies as Alphafold2.

5. ** Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, Yuan D, Stroe O, Wood G, Laydon A, *et al.*: **AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models**. *Nucleic Acids Res* 2022, **50**:D439–D444.
Makes Alphafold2 structure predictions available as a searchable database. This made structureomic data more accessible to researchers.

6. Tunyasuvunakool K, Adler J, Wu Z, Green T, Zielinski M, Žídek A, Bridgland A, Cowie A, Meyer C, Laydon A, *et al.*: **Highly accurate protein structure prediction for the human proteome**. *Nature* 2021, **596**:590–596.

7. Akdel M, Pires DEV, Pardo EP, Jänes J, Zalevsky AO, Mészáros B, Bryant P, Good LL, Laskowski RA, Pozzati G, *et al.*: **A structural biology community assessment of AlphaFold2 applications**. *Nat Struct Mol Biol* 2022, **29**:1056–1067.

8. * Terwilliger TC, Poon BK, Afonine PV, Schlicksup CJ, Croll TI, Millán C, Richardson JS, Read RJ, Adams PD: **Improved AlphaFold modeling with implicit experimental information**. *Nat Methods* 2022, **19**:1376–1382.9.
Describes an algorithm that uses the fit of predicted models into densities derived from cryo-EM experiments to update structural predictions using Alphafold2.

9. Terwilliger TC, Afonine PV, Liebschner D, Croll TI, McCoy AJ, Oeffner RD, Williams CJ, Poon BK, Richardson JS, Read RJ, *et al.*: **Accelerating crystal structure determination with iterative AlphaFold prediction**. *Acta Crystallogr Sect Struct Biol* 2023, **79**:234–244.

10. ** Evans R, O'Neill M, Pritzel A, Antropova N, Senior A, Green T, Žídek A, Bates R, Blackwell S, Yim J, *et al.*: **Protein complex prediction with AlphaFold-Multimer**. *bioRxiv* 2022, https://doi.org/10.1101/2021.10.04.463034.
Describes an updates Alphafold algorithm specifically trained to predict protein complexes.

11. ** Humphreys I, Pei J, Baek M, Krishnakumar A, Anishchenko I, Ovchinnikov S, Zhang J, Ness TJ, Banjade S, Bagde SR, *et al.*: **Computed structures of core eukaryotic protein complexes**. *Science* 2021:374.
Describes an AI-based algorithm trained to predict protein complexes.

12. * O'Reilly FJ, Graziadei A, Forbrig C, Bremenkamp R, Charles K, Lenz S, Elfmann C, Fischer L, Stülke J, Rappsilber J: **Protein complexes in cells by AI-assisted structural proteomics**. *Mol Syst Biol* 2023, **19**, e11544.
Describes an integrative CLMS and AI approach to predict and characterize complexes in cell lysates.

13. Skalidis I, Kyrilis FL, Tüting C, Hamdi F, Chojnowski G, Kastritis PL: **Cryo-EM and artificial intelligence visualize endogenous protein community members**. *Structure* 2022, **30**:575–589.e6.

14. Bantscheff M, Lemeer S, Savitski MM, Kuster B: **Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present**. *Anal Bioanal Chem* 2012 4044 2012, **404**:939–965.

15. * Cappelletti V, Hauser T, Piazza I, Pepelnjak M, Malinovska L, Fuhrer T, Li Y, Dörig C, Boersema P, Gillet L, *et al.*: **Dynamic 3D proteomes reveal protein functional alterations at high resolution in situ**. *Cell* 2021, **184**:545–559.e22.
Uses limiting proteolysis mass-spectrometry to characterize changes to the structureome in response to changes in the cellular environment.

16. O'Reilly FJ, Rappsilber J: **Cross-linking mass spectrometry: methods and applications in structural, molecular and systems biology**. *Nat Struct Mol Biol* 2018, **25**:1000–1008.

17. Turk M, Baumeister W: **The promise and the challenges of cryo-electron tomography**. *FEBS Lett* 2020, **594**:3243–3261.

18. Wan W, Briggs JAG: **Cryo-electron tomography and subtomogram averaging**. *Methods Enzymol* 2016, **579**:329–367.

19. ** Tegunov D, Xue L, Dienemann C, Cramer P, Mahamid J: **Multi-particle cryo-EM refinement with M visualizes ribosome-antibiotic complex at 3.5 Å in cells**. *Nat Methods* 2021, **18**:186–193.
Presents a new algorithm that uses machine learning to improve the resolution of subtomogram averages once particles are identified.

20. Himes BA, Zhang P: **emClarity: software for high-resolution cryo-electron tomography and subtomogram averaging**. *Nat Methods* 2018, **15**:955–961.

21. Jamali K, Kimanius D, Sjors, Scheres HW: *A graph neural network approach to automated model building in cryo-EM maps*. 2022, https://doi.org/10.48550/arxiv.2210.00006.

22. ** Fontana P, Dong Y, Pi X, Tong AB, Hecksel CW, Wang L, Fu T-M, Bustamante C, Wu H: **Structure of cytoplasmic ring of nuclear pore complex by integrative cryo-EM and AlphaFold**. *Science* 2022, **376**, eabm9326.
Integrates structureomic predictions with other data including densities from single particle cryo-EM experiments to build a model of the cytoplasmic ring of the *Xenopus laevis* nuclear pore complex.

23. Mosalaganti S, Obarska-Kosinska A, Siggel M, Turonova B, Zimmerli CE, Buczak K, Schmidt FH, Margiotta E, Mackmull M-T, Hagen W, *et al.*: *Artificial intelligence reveals nuclear pore complexity*. 2021, https://doi.org/10.1101/2021.10.26.465776.
** Combines structural predictions with subtomogram averaging of human nuclear pore complexes in different conditions to build multiple models to capture conformational changes.

24. Chen Z, Greenan GA, Shiozaki M, Liu Y, Skinner WM, Zhao X, Zhao S, Yan R, Yu Z, Lishko PV, *et al.*: **In situ cryo-electron tomography reveals the asymmetric architecture of mammalian sperm axonemes**. *Nat Struct Mol Biol* 2023, **30**: 360–369.
** Uses the predicted structureome to identify likely candidates to account for density derived from subtomogram averaging of the sperm axoneme.

25. O'Reilly FJ, Xue L, Graziadei A, Sinn L, Lenz S, Tegunov D, Blötz C, Singh N, Hagen WJH, Cramer P, *et al.*: **In-cell architecture of an actively transcribing-translating expressome**. *Science* 2020, **369**:554–557.
** Combines CLMS and cryo-ET to build a model of coupling of transcription and translation in *Mycoplasma pneumoniae* cells.

26. Böhm J, Frangakis AS, Hegerl R, Nickell S, Typke D, Baumeister W: **Toward detecting and identifying macromolecules in a cellular context: template matching applied to electron tomograms**. *Proc Natl Acad Sci U S A* 2000, **97**:14245–14250.
* Early description of the use of template matching to locate macromolecules in cryo-tomograms of cells using their structures as a probe.

27. Frangakis AS, Böhm J, Förster F, Nickell S, Nicastro D, Typke D, Hegerl R, Baumeister W: **Identification of macromolecular complexes in cryoelectron tomograms of phantom cells**. *Proc Natl Acad Sci USA* 2002, **99**:14153–14158.

28. Nickell S, Förster F, Linaroudis A, Net WD, Beck F, Hegerl R, Baumeister W, Plitzko JM: **TOM software toolbox: acquisition and analysis for electron tomography**. *J Struct Biol* 2005, **149**: 227–234.

29. Nickell S, Kofler C, Leis AP, Baumeister W: **A visual approach to proteomics**. *Nat Rev Mol Cell Biol* 2006, **7**:225–230.

30. Rickgauer JP, Grigorieff N, Denk W: **Single-protein detection in crowded molecular environments in cryo-EM images**. *eLife* 2017, **6**, e25648.

31. Rickgauer JP, Choi H, Lippincott-Schwartz J, Denk W: **Label-free single-instance protein detection in vitrified cells**. *bioRxiv* 2020, https://doi.org/10.1101/2020.04.22.053868.
** Describes the use of high-resolution template matching to visualize ribosomes and characterize translation states in thin extensions of mouse embryonic fibroblasts without 3D averaging.

32. 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**.
** Presents a user friendly, GUI implementation of a GPU-accelerated 2DTM algorithm and compares the results with template matching in tomograms.

33. Lucas BA, Zhang K, Loerch S, Grigorieff N: **In situ single particle classification reveals distinct 60S maturation intermediates in cells**. *eLife* 2022:11.
* Describes a statistical approach to classify cellular populations of related complexes in FIB-milled yeast lamellae by comparison to structureomic references.

34. Cossio P, Rohr D, Baruffa F, Rampp M, Lindenstruth V, Hummer G: *BioEM: GPU-accelerated computing of Bayesian inference of electron microscopy images*. 2018.

35. Ruff KM, Pappu RV: **AlphaFold and implications for intrinsically disordered proteins**. *J Mol Biol* 2021:433.

36. Theillet FX, Luchinat E: **In-cell NMR: why and how?** *Prog Nucl Magn Reson Spectrosc* 2022, **132–133**:1–112.

37. Wassie AT, Zhao Y, Boyden ES: **Expansion microscopy: principles and uses in biological research**. *Nat Methods* 2018, **16**:33–41.

38. Shaib AH, Chouaib AA, Chowdhury R, Mihaylov D, Zhang C, Imani V, Georgiev SV, Mougios N, Monga M, Reshetniak S, *et al.*: *Visualizing proteins by expansion microscopy* 2023, https://doi.org/10.1101/2022.08.03.502284.

39. 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.

40. Yang JE, Larson MR, Sibert BS, Kim JY, Parrell D, Sanchez JC, Pappas V, Kumar A, Cai K, Thompson K, *et al.*: **Correlative cryogenic montage electron tomography for comprehensive in-situ whole-cell structural studies**. *bioRxiv* 2022, https://doi.org/10.1101/2021.12.31.474669.

41. Peck A, Carter SD, Mai H, Chen S, Burt A, Jensen GJ: **Montage electron tomography of vitrified specimens**. *J Struct Biol* 2022, **214**, 107860.

42. Lucas BA, Grigorieff N: **Quantification of gallium cryo-FIB milling damage in biological lamella**. *Proc Natl Acad Sci* 2023, **120**, e2301852120.

43. Dumoux M, Glen T, Smith JL, Ho EM, Perdigão LM, Pennington A, Klumpe S, Yee NB, Farmer DA, Lai PY, *et al.*: **Cryo-plasma FIB/SEM volume imaging of biological specimens**. *eLife* 2023, **12**, e83623.

44. Wolf SG, Elbaum M: **CryoSTEM tomography in biology**. *Methods Cell Biol* 2019, **152**:197–215.

45. Lazić I, Wirix M, Leidl ML, de Haas F, Mann D, Beckers M, Pechnikova EV, Müller-Caspary K, Egoavil R, Bosch EGT, *et al.*: **Single-particle cryo-EM structures from iDPC–STEM at near-atomic resolution**. *Nat Methods* 2022, **19**:1126.

46. Kelley K, Raczkowski AM, Klykov O, Jaroenlak P, Bobe D, Kopylov M, Eng ET, Bhabha G, Potter CS, Carragher B, *et al.*: **Waffle Method: a general and flexible approach for improving throughput in FIB-milling**. *Nat Commun* 2022, **13**: 1857.

47. Schaffer M, Pfeffer S, Mahamid J, Kleindiek S, Laugks T, Albert S, Engel BD, Rummel A, Smith AJ, Baumeister W, *et al.*: **A cryo-FIB lift-out technique enables molecular-resolution cryo-ET within native Caenorhabditis elegans tissue**. *Nat Methods* 2019, **16**:757–762.

48. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D: **The human genome browser at UCSC**. *Genome Res* 2002, **12**:996–1006.

49. Bepler T, Berger B: **Learning the protein language: evolution, structure, and function**. *Cell Syst* 2021, **12**:654–669.e3.
* Review presenting a vision for how natural language processing could use structureomic data to infer biologically meaningful information.

50. de Groot NS, Armaos A, Graña-Montes R, Alriquet M, Calloni G, Vabulas RM, Tartaglia GG: **RNA structure drives interaction with proteins**. *Nat Commun* 2019, **10**:1–13.

51. Fiehn O: **Metabolomics — the link between genotypes and phenotypes**. *Funct Genomics* 2002, https://doi.org/10.1007/978-94-010-0448-0_11.

52. Lander ES, Weinberg RA: **Journey to the center of biology**. *Science* 2000, **287**:1777–1782.

53. Russo CJ, Dickerson JL, Naydenova K: **Cryomicroscopy in situ : what is the smallest molecule that can be directly identified without labels in a cell?** *Faraday Discussions* 2022, **240**:277–302.