Published in final edited form as:

J Struct Biol. 2011 March ; 173(3): 483–496. doi:10.1016/j.jsb.2010.11.011.

Exploring the Spatial and Temporal Organization of a Cell's Proteome

Martin Beck¹, Maya Topf², Zachary Frazier³, Harianto Tjong³, Min Xu³, Shihua Zhang³, and Frank Alber^{3,*}

¹European Molecular Biology Laboratory, Meyerhofstr. 1, 69117 Heidelberg, Germany

²Molecular Biology, Crystallography, Department of Biological Sciences, Birkbeck College, University of London, London, UK

³Program in Molecular and Computational Biology, University of Southern California, 1050 Childs Way, RRI 413E, Los Angeles, CA 90068, USA

Abstract

To increase our current understanding of cellular processes, such as cell signaling and division, knowledge is needed about the spatial and temporal organization of the proteome at different organizational levels. These levels cover a wide range of length and time scales: from the atomic structures of macromolecules for inferring their molecular function, to the quantitative description of their abundance, and distribution in the cell. Emerging new experimental technologies are greatly increasing the availability of such spatial information on the molecular organization in living cells. This review addresses three fields that have significantly contributed to our understanding of the proteome's spatial and temporal organization: first, methods for the structure determination of individual macromolecular assemblies, specifically the fitting of atomic structures into density maps generated from electron microscopy techniques; second, research that visualizes the spatial distributions of these complexes within the cellular context using cryo electron tomography techniques combined with computational image processing; and third, methods for the spatial modeling of the dynamic organization of the proteome, specifically those methods for simulating reaction and diffusion of proteins and complexes in crowded intracellular fluids. The long-term goal is to integrate the varied data about a proteome's organization into a spatially explicit, predictive model of cellular processes.

Keywords

Macromolecular assemblies; Visual proteomics; Cryo electron tomography; Cryo electron microscopy; Density fitting; Brownian dynamics simulations; Reaction-diffusion modeling

2. Introduction

A vast amount of information on a cell's proteome is being generated by a variety of experimental methods, revealing multiple levels of spatial organization in a cell (Figure 1) [1–6]. The molecular structures of individual macromolecules and their assemblies are

© 2010 Elsevier Inc. All rights reserved.

*correspondence should be addressed to: Frank Alber, alber@usc.edu, phone +1 213 740 0778, fax +1 213 740 2437.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

routinely determined by X-ray crystallography, NMR spectroscopy and electron microscopy (EM). These structures are necessary to determine the molecular functions of macromolecules and their complexes, and provide insights into their evolution [1, 7]. Insights into the proteome organization at a cellular level can be determined by fluorescence imaging [8, 9] and cryo electron tomography (cryoET) [10–12]. CryoET can be applied to a pleomorphic biological specimen in its near native state, and can not only determine a cell's ultra-structure (*ie*, the membrane bound compartmentalization in the cell) [10, 13, 14], but also can increasingly produce 3D snapshots of the distributions of large complexes inside the cell [3, 15, 16].

At lower levels of organization, new genomics and proteomic technologies are documenting what genes are transcribed and when [2, 17]. Quantitative mass spectrometry can provide information about which types of proteins are active at a single point in time, and can also reveal the relative and absolute abundance of these proteins and their assemblies in cells and organelles [3, 18–20]. It is also possible to obtain quantitative information about the association rates of reversible protein interactions and the rates of enzymatic reactions in many process pathways [21–23].

A compelling long-term goal of biology is to combine all available spatial and quantitative information about a cell's constituent parts into a predictive spatially explicit model of cell processes, metabolism and behavior [14, 21, 24, 25]. Ideally such models would include atomic details, relevant to the molecular function of macromolecular assemblies as well as information about the cell's ultra-structure and the higher-order spatial organization of these assemblies at a cellular level. Both scales are necessary to understand key biochemical processes. For instance spatial gradients, irregularities and discontinuities of macromolecular distributions in a cell are known to play an active role in biological processes such as cell division, genome segregation, gene regulation, cell morphogenesis, and shape maintenance. Many biological processes related to cell signaling, cell cycle, and protein transport are also modulated by the precise spatial organization of molecules in space and time [23, 26, 27] [28]. For instance, in the so-called “push-pull” signaling networks, variations in the spatial distribution of two antagonistic enzymes that covalently modify a substrate can either dramatically weaken or amplify the corresponding signal transduction [29]. Two other examples are the chemotaxis pathway in *E. coli* [30, 31], and the mitogen-activated protein kinase (MAPK) cascade processes [22, 23]. Even the simplest cell is quite heterogeneous in their cellular proteome organization [32], but so far relatively little is known about the organization of the cellular proteome at molecular level and how this organization dynamically changes over time.

For a better understanding of cellular processes, it is necessary to expand the scope of structural biology from the molecular functions of the isolated macromolecular assemblies to the concerted roles that such assemblies play in their cellular context [10, 33, 34]. Three fields that in recent years have had a significant contribution in this context are addressed in this review: first, computational methods for calculating pseudo-atomic models of large macromolecular assemblies by combining EM density maps and atomic models of the isolated components - such models provide insights into the molecular functions of the key players in cell processes. These structures provide also a starting point for detecting their localization in the cellular context. Second, cryoET techniques that are used for characterizing the spatial distributions of these assemblies within the cell using computational image processing; and third, spatial modeling of the dynamic organization of the proteome, in particular particle-based reaction-diffusion techniques that simulate the trajectories of these macromolecules and their assemblies over time during reaction processes in their cellular context.

3. Modeling structures of macromolecular assemblies by cryoEM density fitting

High-resolution structures of macromolecular assemblies are key for understanding biological processes at the molecular level [35]. They can serve as templates in the identification of complexes in whole cell tomograms (section 4) and are useful to predict realistic association rates [36] and diffusion properties in particle-based simulations of cellular mixtures (section 5). Although X-ray crystallography has provided spectacular insights into the structure and function of such assemblies, (e.g., the ribosome [37] and the exosome [38]), most of the structural data on abundant large and stable assemblies has been derived in recent years by electron microscopy (EM) methods [39, 40]. EM produces 3D density maps of large assemblies typically at intermediate to low levels of resolution (~5–25 Å). The number of density maps in the public domain is growing rapidly, with over 360 structures in the EM databank alone [41]. Recent advances in cryoEM data acquisition and image processing have made it possible to determine a rapidly increasing number of structures at sub-nanometer resolution, allowing in many cases the identification of secondary structure elements and even the tracing of the backbone [42–45]. A number of structures were even solved at near-atomic resolutions (~3.8–4.5 Å), where many detailed structural features such as turns and deep grooves of helices, strand separation in beta-sheets, densities for loops and bulky amino acids side chains could be resolved [43]. Such structures have so far been obtained in highly symmetrical objects such as the bacterial flagellar filament [46], various viruses (the cytoplasmic polyhedrosis virus [47], rotavirus inner capsid particle [48], and the Epsilon 15 bacteriophage [49]), and the GroEL chaperonin [50]. For most cryoEM maps, however, the level of resolution is still not sufficient to directly determine the structure at atomic detail. Indeed, the number of medium- to low-resolution density maps is increasing exponentially while the increase of high-resolution maps grows at a smaller rate. At the intermediate- to low- resolution range, pseudo-atomic models can be obtained by fitting atomic structures and models of individual components (such as whole proteins, domains and nucleic acid chains) into the EM density maps [45, 51, 52].

In this section we review computational methods for rigid fitting, flexible fitting and multiple-component fitting into EM density maps of macromolecular assemblies. We give examples of successful applications of computational methods that produced pseudo-atomic structures of macromolecular assemblies.

3.1. Rigid fitting

Rigid fitting methods have been very successful in providing many pseudo-atomic models of macromolecular assemblies (Table 1) [52–55]. In most of these methods, the quality-of-fit measure for the placement of a probe structure in a density map is the cross-correlation coefficient (or its variants), calculated between the EM density map and the density of the probe structure (computed by convolving the atomic structure using a point-spread function). To find the highest cross-correlation some of these methods (e.g. COLORES [56]) perform an exhaustive search over all possible probe map positions and orientations relative to the reference EM map. Since performing such a search is computationally intensive, they usually increase the sampling efficiency by applying a Fast Fourier Transform (FFT)-accelerated translational search based on the convolution theorem [57]. Sampling efficiency can also be accelerated by a fast rotational search based on spherical harmonics (SH) (e.g., as implemented in ADP_EM [58]). In contrast to the exhaustive search, some programs can perform a heuristic search, for example by using a Monte Carlo-based search algorithm [59, 60].

Other methods increase the computational efficiency by reducing the complexity of density maps to a small set of feature points (so-called code-book vectors) whose positions best reproduce the density map's gross features, such as its shape and mass distribution (Table 1). The optimal positions of feature points can be determined by the vector quantization (VQ) technique [61]. The fitting problem then effectively reduces into a common point-set matching problem. This matching has been implemented in the SITUS package by an exhaustive search method for single-molecule fitting [62] and a heuristic anchor-point registration method for component fitting into an assembly map [63]. The latter method uses a hierarchical alignment of the point sets and reduces the search-space complexity by an integrated tree-pruning technique.

There are also a number of programs that allow a local search (e.g., Mod-EM [60] and Chimera/Fit_in_Map [64]), that is, maximizing the quality-of-fit measure by searching only in the immediate neighborhood of the initial position of the probe structure. This option is rather useful because often searching for a correct fit of a component in the entire assembly map may lead to an incorrect fit. If knowledge about the approximate position of the component in the map exists (e.g., based on the position of other components or labeling experiments) local fitting may help avoiding this problem. Another way to solve this problem is to use simultaneous multiple component fitting (see below).

3.2. Flexible Fitting

A common problem in EM density fitting is that the isolated component structure may be in a different conformation than in the assembly density map. These conformational differences can originate from the varied experimental conditions under which components and assembly structures were determined as well as from distortions due to crystal packing and experimental noise. In addition, due to sample heterogeneity single-particle cryoEM and image processing often results in a number of maps describing different functional conformations of the entire complex [65]. Common conformational differences are shear and hinge movements of domains and secondary structure elements, as well as loop distortions. When a component structure is generated by a structure prediction method [66] additional errors can be introduced due to misassignment of secondary structure elements to incorrect sequence regions resulting in their shifts in space [35, 67].

The simplest approach to incorporate conformational variability is to divide the probe structure into rigid bodies, such as domains, and fit each of them independently into the map [68, 69]. Clearly, this approach is problematic because the mechanical properties of the probe structure are not reflected in the deformation of the structure. A more objective approach is to generate multiple "valid" conformations of the probe structure and select the top ranking conformation based on its fit in the density. This approach, which essentially separates sampling and scoring, is based on the principle that a correlation exists between the quality of a structural model (e.g. in terms of the RMSD from the native conformation) and its cross correlation with the density map [60]. An alternative to this approach are methods that simultaneously optimize the conformation of the probe and its position and orientation in the cryoEM map [54]. Such fitting methods are similar to crystallographic refinement programs, except that they generally refine groups of atoms held together instead of individual atoms. Common to all the automated methods is the limited sampling of conformational degrees of freedom. Therefore, they are usually applied to components that are first placed into the density map by rigid fitting.

Fitting multiple conformations—In order to generate candidate conformations for fitting one can use normal model analysis [70–72], comparative protein structure modeling (Moulder-EM [73] and the CHOYCE webserver [74]), *ab initio* modeling (Rosetta/

Foldhunter [75]), and exploit the structural variability of protein domains within a given superfamily (S-flexfit [76]). Fitting of multiple conformations is particularly useful when crystal structures of the components are not available and must be modeled by comparative modeling or by *ab initio* protein structure prediction methods. In an example of this approach, most eukaryotic ribosomal proteins were modeled using comparative modeling and fitting based on a density map at ~ 9 Å resolution [77, 78] (Figure 2). A similar approach was implemented in the Moulder-EM protocol. This protocol optimizes a comparative model by iterating over sequence alignment, model building, and model assessment [73]. The model assessment is based on a combined score including the cross-correlation score between the fitted model and the density map and an atomic statistical potential score [73, 79]. Using this protocol, three structural domains were modeled at the N-terminal region of the skeletal muscle Ca^{2+} release channel (RyR1), based on remote template structures ($<20\%$ sequence identity) and a 9.6 Å resolution cryoEM map [80]. 11 of the 15 disease-related residues for these domains were mapped to the surface of these models. In another example, *de novo* protein structure predictions (Rosetta [81]) were assessed by cryoEM density fitting (Foldhunter [82], leading to the discovery of a novel fold for the herpesvirus VP26 core domain [75].

Refinement methods—Real-space refinement methods (e.g., RSRef [83], Flex-EM [67], MDFF [84], Rosetta [85], and EM-IMO [86]) use the density to guide the conformational changes in the probe structure. Generally these methods optimize the fit of the probe structure in the density while maintaining its mechanical properties. In Flex-EM, for example, the scoring function includes a cross-correlation term as well as stereochemical and non-bonded interaction terms [67]. A heuristic optimization that relies on a Monte Carlo search, a conjugate-gradients minimization, and simulated annealing molecular dynamics (MD) is applied to a series of subdivisions of the structure into progressively smaller rigid bodies. The method has been used to refine a number of crystal structures and homology models in cryoEM maps, among them the structure of the elongation factor EF4 in the ~ 11 Å resolution cryoEM map of EF4 bound to the *E. coli* ribosome (Figure 3) [87] and the structure of the apoptosome-procaspase-9 CARD complex at 9.5 Å [88]. In YUP.SCX [89] intra-molecular forces are approximated by a Gaussian Network Model (GNM) and the optimization protocol uses simulated annealing MD. Flex-EM and YUP.SCX were applied to refine pseudo-atomic models of proteins and RNA using the density of a eukaryotic ribosome at ~ 9 Å resolution [78] (Figure 2). In MDFF [84]), external forces proportional to the gradient of the density map are combined with an MD protocol to optimize the conformations of the component structures. This method was recently applied to studies on the mechanisms of protein synthesis by the ribosome [90] and the dynamics of protein translocation [91].

In many refinement methods the conformation is optimized for groups of atoms, either by defining them as rigid bodies [67] or by applying restraints on secondary structure elements [84]. Rigid bodies can be selected manually or by automated methods, such as those based graph theory (e.g., FIRST/FRODA [92]). However, the definition of rigid bodies can limit the conformational degrees of freedom of the structure: if the number of rigid bodies is too small the optimization may not reach the global minimum because a more detailed modification of the conformation is needed. If an all-atom representation is chosen the computational efficiency is largely reduced and the system is likely to get trapped in local minima. A way to avoid this problem can be the use of coarse-grained (CG) or reduced structure representation such as the C β level (e.g., as in DireX which combines an elastic network model with random walk displacements and distance restraints [93]). A recent method applied a Go-model with cross-correlation biased MD without imposing any restraints on the secondary structure elements [94]. A reduced representation can also be used in conjunction with interpolation methods, as an alternative to the constrained MD

[95]. Finally, conformational changes in a structure can also be guided by normal mode analysis (NMA) [70–72]. In NMFF for example, a linear combination of low-frequency normal modes is used to iteratively deform the structure to conform to the low-resolution density map [96]. One of the advantages of this method is that it allows for large conformational changes as observed in the anthrax complex [97] or GroEL [98]. However, NMA is often limited in describing smaller-scale conformational changes [99] that are often reflected in intermediate-resolution cryoEM density map.

3.3. Simultaneous fitting of multiple components

A considerable challenge is the fitting of multiple components into the density map of an assembly if no *a priori* knowledge about the location of the components is available. Sequential fitting of components often fails when they cannot be unambiguously placed in the density map as is often the case for maps of intermediate to low resolution and for assemblies with a large number of components. In such cases, all the components must be fitted simultaneously into the map to identify the global optimum of the quality-of-fit measure. However, simultaneous fitting is difficult as the large search space makes an exhaustive search protocol (that uniformly samples over all degrees of freedom) computationally unfeasible.

One method (MultiFit [100]) uses discrete sampling in combination with an inference optimizer and expands the quality-of-fit measure by additional information such as shape complementarities between interacting components and the protrusion of components from the map envelope. Other methods, such as GMFIT [101] and IQP [102], simplify the search problem by reducing the complexity of the assembly density map and structures [62, 63]. In GMFIT the initial density distribution of assembly and components are approximated by a small set of Gaussian functions to efficiently use gradient-based optimization methods for the structural optimization of component orientations (GMFIT). In the IQP method integer quadratic programming is used for matching two point sets that represent the reduced density distribution of several components and the assembly. IQP is able to match multiple component point sets simultaneously while considering both information about the geometric architecture of the point distributions as well as the consistency of the density map in the immediate neighborhood of the points. Moreover the method allows straightforward integration of additional information about the assembly. Reducing the complexity of the density information to point sets is accompanied by an inevitable loss in accuracy in the fitting process. To overcome this challenge, typically a large number of independent point set matches must be performed, which generates an ensemble of candidate solutions. This ensemble is then assessed using an independent scoring function that measures the quality-of-fit between the components structures and the assembly map using the cross-correlation function. Finally, the best scoring structures are refined to locally optimize the fit between assembly and component density maps.

Although progress has been achieved, both simultaneous and flexible fitting of assembly components are still a challenge, in particular for cryoEM maps at resolution worse than ~10 Å (the majority of maps). There is currently a need to develop faster scoring functions that can help discriminate between good and bad fits as well as optimization methods that can explore both conformational and configurational space more efficiently.

3.4. Comprehensive data integration

For the fitting of low-resolution density maps the sampling can be improved if additional information is available about the assembly structure, for instance about protein interactions, subcomplex composition, or shape information from mass spectrometry and SAXS, respectively. Recently, a systematic computational framework was introduced (IMP) to

determine the structures of large macromolecular assemblies by comprehensive integration of diverse data sources [35, 103–106]. The method can integrate structural information that can vary greatly in terms of their resolution and precision, for instance, data from X-ray crystallography, NMR spectroscopy, electron microscopy, chemical cross-linking, affinity purification, yeast two-hybrid experiments, and computational docking [103, 105, 106]. The approach attempts to compute structures that are consistent with all available information about the assembly regarding the composition and structure of the complex. The structure determination process is formulated as an optimization problem where structures that are consistent with all input information are found by minimizing a scoring function. To comprehensively sample the solution space, an ensemble is generated that contains independently calculated structures, each satisfying all of the input restraints.

By applying this framework, the structure of the 450-protein containing nuclear pore complex (NPC) was determined based on the integration of data from seven independent experimental and theoretical sources [103, 104]. The averaged structural NPC model has led to functional insights and evolutionary understanding of the yeast NPC [104]. Another recent study determined the pseudo-atomic model AAA-ATPase/20S core particle sub-complex of the 26S proteasome [107, 108]. The structure was determined based on a cryoEM map of the 26S proteasome, structures of homologs, and physical protein–protein interaction data.

4. Determining the spatial organization of the cellular proteome

High-resolution structures of macromolecular complexes provide major insights in their molecular mechanism. But to fully understand their function in the context of cellular processes, knowledge of their spatial distribution within living cells is crucial. We currently only begin to understand that the spatial organization of the cell goes far beyond membrane confined compartments and solely diffusion controlled reactions. Cooperating functional modules can function more effectively in local proximity and their reactive sites might be specifically orientated towards each other. For example, recent work demonstrated that certain mRNAs can be locally translated [109] and that polysomal ribosomes specifically orient along the mRNA trace [110]. Such polysomal ribosome arrangements occur in protrusions of Glia cells [111], which might form subcompartments that could allow chaperones to easily access the emerging polypeptide chain and effectively shielding it from the crowded environment until folding is complete. These findings illustrate the need for techniques that reveal the spatial position and orientation of macromolecules at sufficient resolution on a proteome-wide scale.

Classical proteomics approaches determine the protein content of cells, however, any spatial information is lost and cell specific properties are averaged over the population of lysed cells. Recent advances in technology will allow a much better identification and localization of these assemblies in the cell. Among these methods are (i) high-resolution fluorescence imaging, which can provide spatial and dynamic information about the localization, abundance and diffusion rates of macromolecules and their assemblies [112] and (ii) cryo electron tomography (cryoET), which not only can provide insights into the ultra-structure of cells but increasingly even into the cellular distribution of assemblies [16, 113–115]. With these methods it may be possible to gain information about the cellular environment of macromolecular assemblies, and gain insights into the positions of all their instances in a cell, during different time points in its lifetime. They also promise to reveal differences in the distribution of macromolecules between individual cells and their contribution to the behavior of the entire population. Such information will help bridging the knowledge about the molecular function of macromolecules derived from their atomic structure with their

cellular function at the cellular systems level. In the following section we focus on methods and applications of cryoET.

4.1. Visual proteomics using cryo electron tomography

CryoET provides 3D snapshots of cells under close-to-live conditions (Figure 4, 5). However, assigning the protein identity to the observed shapes and distribution of complexes is not a trivial task due to a number of technical challenges: a tomogram typically has low signal-to-noise ratio, non-isotropic resolution and straight-forward *in vivo* labeling techniques, analogous to GFP-tags used in fluorescence imaging, are not available for cryoEM [10, 14, 33, 116]. For the proportion of protein assemblies whose structures are known, visual proteomics methods have been developed that attempt to generate molecular atlases containing the position and angular orientation of protein complexes in cryo electron tomograms of cells by template matching (or pattern recognition) [12, 33, 114, 117, 118]. Thereby at first a template library is built that contains the reference structures of the protein complexes to be localized. Subsequently, an exhaustive search is performed that scans the tomographic volumes for matches with the structural templates by calculating a cross correlation score. Finally, the observed distribution of the cross correlation score is translated into a position list using peak extraction and statistical methods. In 2000, the first proof-of-concept for template matching in cryo electron tomograms using simulations and a first experimental application was provided [118]. The ultimate goal is to fit high-resolution structures into the cellular context, which would allow for observing protein interactions at work. Nevertheless, the unambiguous detection of macromolecular assemblies in cryo electron tomograms of intact cells remains a considerable challenge [16, 114], not only due to the biological complexity. At the currently achievable resolution only large protein complexes have a chance of being identified and the low signal to noise ratio peculiar to cryo electron tomograms hampers a robust and reliable detection.

Template libraries—Visual proteomics approaches are so far dependent on the availability of template structures, or, with other words, are biased by their own template libraries. In order to generate libraries of template structures that best account for the signal observed in a cryo electron tomogram, the imaging process has to be simulated as realistically as possible. First, the imaged quantity needs to be calculated from the coordinates and identities of the atoms specified in the PDB. Subsequently the density is convoluted with the contrast transfer function, which describes the imaging in the transmission electron microscopy in linear approximation [114, 119]. Beyond these technicalities, there are a few biological aspects to be taken into account when template libraries are being assembled. Can the template sufficiently account for the structural state of the targeted molecule within the cell? A recent structural-proteomic survey of the largest and most abundant protein complexes in *Desulfovibrio vulgaris* revealed a high structural diversity between different microorganisms [120]. The selection of reference structures solved from other species has therefore to be done with caution. Furthermore, the protein complex might exist in diverse oligomeric states in the cell that might differ from the species observed during structure determination. For this reason, visual proteomics approaches so far were limited to highly stable and conserved targets. What is the abundance of the targeted protein complexes in the cell and to what extent does the template library account for all large protein complexes that exist in the cell? This is an important question as well, because high abundant protein complexes of similar structure might compete for assignments and thereby have a negative effect on the performance. Targeted and directed mass spectrometric measurements can provide the required auxiliary information about protein abundances on a proteome-wide scale [121].

Assessment of template matching—Once the template library is built, matches are identified by calculating a cross correlation score between reference structures and the subvolume of a tomogram. Foerster et al implemented a local, constrained correlation function (CCF) for template matching to account for missing wedge effects (due to incomplete tomogram sampling), [122–125]. Recently, a scoring function was introduced that not only evaluates the cross-correlation value for the template, but also the cross-correlation values of competing templates and decoys at the same position in the tomogram [16]. This scoring scheme outperformed the classical workflow particularly for competing protein complexes of similar structural signature. For visualization, the template is rotated and positioned at the determined coordinates and the corresponding angles that lead to the maximal CC values. The major challenge thereby is to discriminate true positive from false positive detections (Figure 4A). The performance not only depends on tomogram-specific parameters such as acquisition settings, specimen thickness, and molecular crowding but also target-specific parameters, such as molecular weights, cellular abundance, and the cellular abundance of protein assemblies competing for assignments. Therefore, the performance can vary for each template contained in the library. The false positive discovery rates of all matches have to be estimated using a statistical model [16, 126]. The performance of template matching can also be assessed based on *a priori* knowledge about the spatial distribution of the template structures, such as the cellular localization or orientation of protein assemblies, which can provide clues about the expected false positive discovery rates. For example, some membrane associated protein complexes exhibit specific positioning and orientation relative to the membrane [127]. Also matching with a mirrored template of ‘non-native’ handedness can provide complementary information about the distribution of cross correlation values. In addition, simulated tomograms can serve as test cases to estimate the performances. Because the position and orientation of all protein complexes in the artificial tomograms are known, the false positive rate can be calculated for specific template complexes under the specific experimental conditions (Figure 4B).

4.2. Applications

In a recent study, namely the visual proteomics project for the pathogen *Leptospira interrogans* [16, 114, 115, 121], targeted proteomics experiments detecting the identity and concentrations of cellular proteins were combined with cryoET based template matching to detect the spatial localizations of a set of protein complexes (Figure 4). To identify template structures that might be suitable for the template matching, 26 candidate assemblies of a certain minimal size (> 250 kDa) and a certain sequence conservation were retrieved from structural databases [16, 114]. Targeted proteomics on a proteome-wide scale determined that the cellular protein abundance in *L. interrogans* ranged over 3.5 orders of magnitude [121]. Only ten other protein complexes of sufficient size for template matching were found with abundances of at least 100 copies per cell. Tomograms of *L. interrogans* cells were acquired and subjected to template matching and scoring (Figure 4 C, D). The local concentration of the targeted protein complexes varied within and across data sets: the cells displayed an average ribosome concentration of ~20 μM (~40 mg/ml) in the cytoplasm, but the local concentration ranged from 5–30 μM (~10–65 mg/ml). The local fluctuations in case of total GroEL together with (GroEL-ES) were larger and ranged from ~8–100 μM (~0.5–6.5 mg/ml). This study showed that: (i) Background noise reduces the performance depending on the MTF of the particular CCD camera; (ii) the missing wedge can introduce angular bias to the discovery rate of templates with anisotropic signal content (such as the elongated ATP-synthase); (iii) the specificity increases with the template abundance and (iv) decreases with the degree of molecular crowding [16, 121]. The last two effects are generally more pronounced for smaller than for larger templates. The conclusion was that the specificity achieved for high abundant Megadalton complexes is satisfactory, however, true-positive discovery rates higher than 50% are difficult to achieve when protein

complexes of smaller molecular weights are targeted. For complexes of very low cellular abundance it is quite a challenge to obtain robust statistical models [16, 121].

Recently, the proteome organization of the small model bacteria *M. Pneumoniae* was analyzed, revealing its proteome composition and organization, as well as its transcriptional and metabolic regulation [2, 3, 17]. Proteomics based TAP-mass spectrometry data were complemented with structural modeling and cryoET to provide insights into the structural anatomy of the model bacteria [3]. The study identified more than 170 protein complexes. Four of these protein complexes whose structures have been known were localized in the tomogram of an intact *M. pneumoniae* cell (Figure 5). This ratio of detected to structurally-characterized protein complexes demonstrates current throughput limitations of cryoEM arising from: (i) the stability of protein complexes during the cryo preparation; (ii) their orientation on the EM-grid and; (iii) the fact that the data analysis by image processing is time consuming and requires intensive user interference. Nevertheless, the study demonstrates that mapping of macromolecular structures into entire-cell tomograms is a powerful strategy when combined with unbiased large-scale complex purification techniques. The template library in this study was in part built from single-particle EM structures of *M. pneumoniae* protein complexes. This considerable effort overcomes the difficulty of structural conservation across species discussed above.

CryoET combined with image averaging provides unique opportunities for structure determination of membrane-bound macromolecular assemblies in intact cells [128, 129]. For instance, a recent study investigated receptor clustering in bacterial chemotaxis and found a universal architecture in a wide range of bacteria [13, 130].

In addition cryoET can provide snapshots of individual time points along a process pathway, as has been demonstrated for cytoskeletal-driven processes that are involved in cell movements [131] and membrane fusion in herpes simplex virus 1 entry [132].

4.3. Outlook

A number of limitations to visual proteomics might be overcome by further technical developments. Currently the most critical limitation in cryoET is the moderate signal-to-noise ratio: Since the dosage of electrons that can be applied to a biological specimen ultimately limits the resolution of cryoET, visual proteomics currently remains restricted to complexes of a certain minimal molecular weight that can be targeted by template matching. However, the signal in visual proteomics primarily arises from the contrast between the targeted protein complex and the surrounding solvent. Therefore, improving contrast, e.g. through phase plates [133], is likely to improve the performance in the future. Furthermore, specimen thinning techniques, e.g. through focused ion beams [134] will be crucial for improving SNR and imaging larger cell types, such as eukaryotic cells, which exceed 500 nm in diameter; thus, application of visual proteomics to eukaryotic cells depends on progress in this field. Finally, direct electron detection systems promise to largely improve SNR of cameras [135], which will be highly beneficial for techniques that target pleomorphic structures, such as cryoET, in which the SNR cannot be improved by imaging the multiple copies of the same specimen.

However, some limitations will remain: cryoET provides static snapshots of single cells without temporal resolution and the spatial resolution of live cell imaging techniques is comparably low. However, cryoET information can be complemented by computational modeling of the proteome diffusion and reaction processes. Such combined approaches hold great potential to reveal insights into the dynamic spatial organization of the proteome and the biological processes in cells.

5. Modeling the dynamic proteome organization

Novel experimental methods have greatly increased the availability of information on the molecular organization of living cells. As described in the previous sections information is increasingly available about the membrane confined compartments of cells and the identity, abundance and spatial distributions of the macromolecular components in the cell. There is a pressing need to integrate these varied data into spatially explicit, predictive models of specific biological processes such as the selective nucleo-cytoplasmic transport, signal transduction, genome separation and mitosis. This goal not only requires knowledge about the atomic structures and molecular functions of macromolecular assemblies involved, but also the quantitative rates of reactions these assemblies are involved as well as their dynamic diffusion properties, which dependent on their spatial environment in the cell. For instance, macromolecular crowding can influence protein diffusion and thus the thermodynamic and statistical properties of important reaction processes. In the following sections, we review computational methods that simulate reaction processes while incorporating the dynamic diffusion behavior of proteins and their assemblies in crowded intracellular fluids. We begin by briefly mentioning the major applications of network models for process simulations, and then list the major particle-based reaction diffusion modeling approaches, which are the main focus of this section.

Network models

Network models are based on a mathematical description of the law of mass action and rate equations and often treat the cell as a well-mixed reactor where all components of the reaction network have spatially uniform concentrations [21, 25]. However, many cellular processes, such as cell division and nucleo-cytoplasmic transport, are either spatially constrained or segregated [136]. These activities actively exploit spatial gradients in the macromolecular distributions. *Spatial* network models (such as Virtual Cell [137, 138], MesoRD [139], SmartCell [140]) incorporate such spatial effects by partitioning the cell volume into microdomains [24, 25]. One example of this approach is the Virtual Cell model [137, 138], which incorporates experimentally determined cell geometries into a network model of reaction processes.

When biomolecules are present in relatively low copy numbers, their local concentrations can fluctuate widely, which can cause stochastic effects in reaction processes. The effective behavior of such molecules may be very different from their behavior under a constant distribution; furthermore, in such cases the law of mass action for reaction kinetics no longer applies. It has been shown that such stochastic particle behavior can significantly influence gene regulation [141, 142], signal transduction [143] and many other processes [25]. To model stochastic behavior several network models (including SmartCell [140] and MesoRD [139]) replace the law of mass action with a chemical master equation [25, 144]. Master equations are also simulated by the Stochastic Simulation Algorithm (SSA) [145], and its variants such as the tau-leaping method [146], which can account for low copy numbers in cellular network simulations.

Particle-based reaction-diffusion simulations

Particle-based simulations naturally incorporate the concepts of space, crowding and stochasticity, so often provide a more realistic treatment than network models. A protein or other reactants are often treated as single particles. Every particle in the reaction system is treated explicitly, and the particle positions are sampled at discrete time intervals. To simulate particle diffusion, these simulations use a method known as Brownian dynamics (BD) [25, 147]: the net force experienced by a particle contains a random element in addition to contributions from interactions with other particles. The random element is an

explicit approximation to the statistical properties of Brownian forces, due to the effects of collisions with solvent molecules, which are not explicitly modeled. More specifically, the particles are displaced from their position at each time interval by a random vector whose norm is chosen from a probability distribution function that is a solution to the Einstein diffusion equation. As a consequence, spatial gradients and localized interactions naturally occur. A number of reaction-diffusion algorithms incorporating Brownian dynamics have been developed [29, 148–151]. Interactions and reactions occur upon contact between particles according to specific probabilities, which are chosen to reproduce the correct reaction kinetics [29, 148–150]. The particle-based simulation scheme, described above, has been implemented by several packages, including GridCell [149], MCell [152], Smoldyn [150], and the molecular simulators described by Ridgway et al. [148] and Morelli et al. [29]. In GridCell, particles move only to occupy vertices of cubic grid structures. In MCell and Smoldyn the molecules move freely and are treated as point particles without exclusion volume. Thus the latter two methods do not explicitly resolve particle collisions. In Smoldyn, reactions are accepted if two particles pass within a certain distance chosen such that the overall reaction rate is reproduced correctly. As the particles lack volume, macromolecular crowding in the cell interior is incorporated indirectly by adding inert stationary, impenetrable blocks to the cytoplasm [30, 150]. Smoldyn can be applied with relatively long time steps, allowing simulations at biologically relevant time scales. However, it does not reproduce the distribution function of particle diffusion with high accuracy. Smoldyn was used to study the effects of cellular architecture on signal transduction in *Escherichia coli* chemotaxis [30, 153] (Figure 6A). Simulations were performed for the diffusion of CheYp from the cluster of receptors to the flagellar motors, first under control conditions and then in response to attractant and repellent stimuli [30]. Smoldyn has also been used to study the signal amplification in a lattice of coupled protein kinases [154]. MCell is based on Monte Carlo simulations, and can incorporate specific shapes and cell features derived from cryoET and other experimental methods. MCell has been used to study presynaptic calcium dynamics and neurotransmitter release at a neuronal synapse [152]. Another study simulated macromolecular transport through the nuclear pore complex to interpret the experimentally determined cargo distribution in light of existing models for nuclear import [155].

The intracellular space is a highly crowded and non-uniform with an occupied volume of typically up to 50% in eukaryotes and 30–40% in bacteria [25, 156, 157] (Figure 6BC). In order to accurately simulate crowding effects, a number of reaction-diffusion methods take into account the physical dimensions of proteins and complexes [148, 151]. Most of them model the proteins and complexes as hard, elastic spheres [148, 151, 158]. Crowding can have several profound effects on the dynamics within the cell, such as anomalous subdiffusion of macromolecules [159, 160]. These effects in turn have a direct impact on the thermodynamics and kinetics of biological processes, including protein interactions, diffusion, stability, folding, and biochemical reactions rates [157, 161]. Crowding slows rates of diffusion-limited reactions [158, 161, 162], but accelerates reactions with low association rates (i.e., very small reaction probabilities per collision), because it forces colliding particles to remain longer in closer proximity. However, simply giving the particles an excluded volume is not sufficient to accurately reproduce crowding effects [148, 151, 163]. Significantly better results can be obtained by combining the elastic collision method for hard spheres [151] with a mean field description of hydrodynamic interactions [163, 164] and a Yukawa-like interaction potential for electrostatic interactions [165, 166]. The mean-field approach updates the diffusion constant for each particle at every time step according to its local volume fraction [164, 167]. Other advanced approaches have been proposed for the inclusion of hydrodynamic and electrostatic interactions [168–171]. Another hard-sphere algorithm has been developed that rigorously obeys the detailed balance requirement of equilibrium reactions, allowing for an accurate description of

equilibrium properties in biochemical networks [29]. The method was used to study the effect of spatial fluctuations in a push-pull model of two antagonistic enzymes covalently modifying a substrate. The results demonstrate that spatial density fluctuations of the components can strongly reduce the gain of a response in a biochemical network [29], and once again highlight the importance of the spatial dimension. Recently a BD simulation explored macromolecular diffusion at atomic resolution in the crowded *E. coli* cytoplasmic environment (Figure 6C). Snapshots of the simulation trajectories were used to compute the cytoplasm's effects on the thermodynamics of protein folding, association and aggregation [172, 173].

A disadvantage of time-driven BD schemes is that they require a small time step, which limits the effective simulation time. Event-driven simulations, such as the Green's function reaction dynamics (GRFD) scheme [174, 175], on the other hand, make large jumps in time and space when the particles are far apart from each other [174, 175]. When particle concentrations are low, these methods are much more efficient than time-driven methods. GRFD was used to study a mitogen-activated-protein kinase (MAPK) pathway. This research found that rapid enzyme-substrate rebindings can turn a distributive mechanism into a processive mechanism. Furthermore, the response of the network differed dramatically from that obtained by a mean-field analysis of the chemical rate equations [174]. Another event-driven approach is the first-passage Monte Carlo algorithm [176].

The difficulty of reaching long simulation times is a drawback of all particle-based methods. This limitation is particularly onerous when dealing with crowded intracellular fluids, which require short time steps to accurately simulate protein diffusion. Biological processes occur on a wide range of time scales. Some proteins may encounter each other in fractions of a millisecond, while others take hours. DNA transcription and translation are completed in a few minutes, but these processes are made up of thousands of small diffusion-limited reactions, which take fractions of a nanosecond. It is an ongoing challenge to develop particle-based simulations that can cover a wide range of time scales while accurately reproducing the properties of diffusion and reaction networks. In addition a more realistic description of the cellular environment is required, which includes the spatial organization of the cell's membrane bound compartments and the knowledge of the types, abundance and spatial distributions of all the constituent macromolecules.

6. Conclusions

Understanding biological processes will require knowledge about the molecular functions of the cellular components as well as quantitative data about their abundance, interactions and spatio-temporal distributions in the confined geometry of the cell. Emerging new experimental technologies are greatly increasing the availability of such spatial information about the molecular organization in living cells. In this paper we have addressed three distinct fields that have contributed to our understanding of the proteome's spatial and temporal organization. CryoEM density fitting methods can provide pseudo-atomic structures of macromolecular assemblies. Such structures provide insights into the molecular functions of assemblies and can also serve as templates for visual proteomics approaches, which can potentially identify their abundance and distributions in the cell and therefore provide a three-dimensional view of the cellular proteomes and interactomes. Although the method is currently limited to small cells and large macromolecular assemblies, a number of limitations to visual proteomics might be overcome by further technical developments, which will eventually allow to expand this exciting technology to larger eukaryotic cells. Finally, ongoing research is reviewed on the dynamic modeling of cellular reaction processes. Over the last few years several particle-based reaction diffusion methods have been developed that accurately reproduce protein diffusion and reactions in highly crowded

cellular environment. However, it is an ongoing challenge to extend such methods to cover the wide range of time scales seen in biological processes, while maintaining high accuracy in reproducing the reaction properties. The long-term goal will remain to integrate information derived at the various levels of spatial proteome organization, into a predictive spatial model of cell processes. Such integration will be facilitated by the continued extension of structural biology methods from the molecular to the cellular level.

Acknowledgments

The authors acknowledge financial support from the Human Frontier Science Program (RGY0079/2009-C to F.A. and M.T.), Alfred P. Sloan Research foundation (to F.A.); F.A. is a Pew Scholar in Biomedical Sciences, supported by the Pew Charitable Trusts; Medical Research Council Career Development Award (G0600084) (to M.T.) and the 'Special Presidential Prize and Excellent PhD Thesis Award - Scientific Research Foundation of the Chinese Academy of Sciences (to S.Z.). This review is partially based on refs. [16, 35, 114]. The authors also acknowledge M.S. Madhusudhan and Ben Mathiesen for assistance in preparing the manuscript and useful discussions.

References

1. Robinson CV, Sali A, Baumeister W. The molecular sociology of the cell. *Nature*. 2007; 450(7172): 973–82. [PubMed: 18075576]
2. Yus E, et al. Impact of genome reduction on bacterial metabolism and its regulation. *Science*. 2009; 326(5957):1263–8. [PubMed: 19965476]
3. Kühner S, et al. Proteome organization in a genome-reduced bacterium. *Science*. 2009; 326(5957): 1235–40. [PubMed: 19965468]
4. Kiel C, Serrano L. Cell type-specific importance of ras-c-raf complex association rate constants for MAPK signaling. *Science signaling*. 2009; 2(81):ra38. [PubMed: 19638615]
5. Collins SR, et al. Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map. *Nature*. 2007; 446(7137):806–10. [PubMed: 17314980]
6. Lippincott-Schwartz J, Manley S. Putting super-resolution fluorescence microscopy to work. *Nat Methods*. 2009; 6(1):21–3. [PubMed: 19116610]
7. Sali A, et al. From words to literature in structural proteomics. *Nature*. 2003; 422(6928):216–25. [PubMed: 12634795]
8. Betzig E, et al. Imaging intracellular fluorescent proteins at nanometer resolution. *Science*. 2006; 313(5793):1642–5. [PubMed: 16902090]
9. Manley S, et al. High-density mapping of single-molecule trajectories with photoactivated localization microscopy. *Nat Methods*. 2008; 5(2):155–7. [PubMed: 18193054]
10. Leis A, et al. Visualizing cells at the nanoscale. *Trends Biochem Sci*. 2009; 34(2):60–70. [PubMed: 19101147]
11. Li Z, Jensen GJ. Electron cryotomography: a new view into microbial ultrastructure. *Curr Opin Microbiol*. 2009; 12(3):333–40. [PubMed: 19427259]
12. Baumeister W. Electron tomography: towards visualizing the molecular organization of the cytoplasm. *Curr Opin Struct Biol*. 2002; 12(5):679–84. [PubMed: 12464323]
13. Briegel A, et al. Universal architecture of bacterial chemoreceptor arrays. *Proc Natl Acad Sci USA*. 2009; 106(40):17181–6. [PubMed: 19805102]
14. Morris DM, Jensen GJ. Toward a biomechanical understanding of whole bacterial cells. *Annu Rev Biochem*. 2008; 77:583–613. [PubMed: 18355161]
15. Medalia O, et al. Macromolecular architecture in eukaryotic cells visualized by cryoelectron tomography. *Science*. 2002; 298(5596):1209–13. [PubMed: 12424373]
16. Beck M, et al. Visual proteomics of the human pathogen *Leptospira interrogans*. *Nat Methods*. 2009; 6(11):817–23. [PubMed: 19838170]
17. Güell M, et al. Transcriptome complexity in a genome-reduced bacterium. *Science*. 2009; 326(5957):1268–71. [PubMed: 19965477]
18. Malmström J, et al. Proteome-wide cellular protein concentrations of the human pathogen *Leptospira interrogans*. *Nature*. 2009; 460(7256):762–5. [PubMed: 19606093]

19. Ranish JA, et al. The study of macromolecular complexes by quantitative proteomics. *Nat Genet.* 2003; 33(3):349–55. [PubMed: 12590263]
20. Aebersold R, Mann M. Mass spectrometry-based proteomics. *Nature*, March. 2003; 422(6928): 198–207.
21. Kinkhabwala A, Bastiaens PI. Spatial aspects of intracellular information processing. *Curr Opin Genet Dev.* 2010; 20(1):31–40. [PubMed: 20096560]
22. Maeder CI, et al. Spatial regulation of Fus3 MAP kinase activity through a reaction-diffusion mechanism in yeast pheromone signalling. *Nat Cell Biol.* 2007; 9(11):1319–26. [PubMed: 17952059]
23. Dehmelt L, Bastiaens PI. Spatial organization of intracellular communication: insights from imaging. *Nat Rev Mol Cell Biol.* 2010; 11(6):440–52. [PubMed: 20485292]
24. Lemerle C, Di Ventura B, Serrano L. Space as the final frontier in stochastic simulations of biological systems. *FEBS Lett.* 2005; 579(8):1789–94. [PubMed: 15763553]
25. Ridgway D, Broderick G, Ellison MJ. Accommodating space, time and randomness in network simulation. *Curr Opin Biotechnol.* 2006; 17(5):493–8. [PubMed: 16962764]
26. Gavin AC, et al. Proteome survey reveals modularity of the yeast cell machinery. *Nature.* 2006; 440(7084):631–6. [PubMed: 16429126]
27. Scott JD, Pawson T. Cell signaling in space and time: where proteins come together and when they're apart. *Science.* 2009; 326(5957):1220–4. [PubMed: 19965465]
28. Bork P, Serrano L. Towards cellular systems in 4D. *Cell.* 2005; 121(4):507–9. [PubMed: 15907464]
29. Morelli MJ, ten Wolde PR. Reaction Brownian dynamics and the effect of spatial fluctuations on the gain of a push-pull network. *The Journal of chemical physics.* 2008; 129(5):054112. [PubMed: 18698893]
30. Lipkow K, Andrews SS, Bray D. Simulated diffusion of phosphorylated CheY through the cytoplasm of *Escherichia coli*. *J Bacteriol.* 2005; 187(1):45–53. [PubMed: 15601687]
31. Bray D. Bacterial chemotaxis and the question of gain. *Proc Natl Acad Sci USA.* 2002; 99(1):7–9. [PubMed: 11782543]
32. Shapiro L, McAdams HH, Losick R. Why and how bacteria localize proteins. *Science.* 2009; 326(5957):1225–8. [PubMed: 19965466]
33. Nickell S, et al. A visual approach to proteomics. *Nat Rev Mol Cell Biol.* 2006; 7(3):225–30. [PubMed: 16482091]
34. Shtengel G, et al. Interferometric fluorescent super-resolution microscopy resolves 3D cellular ultrastructure. *Proc Natl Acad Sci U S A.* 2009; 106(9):3125–30. [PubMed: 19202073]
35. Alber F, et al. Integrating diverse data for structure determination of macromolecular assemblies. *Annu Rev Biochem.* 2008; 77:443–77. [PubMed: 18318657]
36. Schlosshauer M, Baker D. Realistic protein-protein association rates from a simple diffusional model neglecting long-range interactions, free energy barriers, and landscape ruggedness. *Protein Sci.* 2004; 13(6):1660–9. [PubMed: 15133165]
37. Rodnina MV, Wintermeyer W. The ribosome goes Nobel. *Trends Biochem Sci.* 35(1):1–5. [PubMed: 19962317]
38. Lorentzen E, Basquin J, Conti E. Structural organization of the RNA-degrading exosome. *Curr Opin Struct Biol.* 2008; 18(6):709–13. [PubMed: 18955140]
39. Chiu W, Baker ML, Almo SC. Structural biology of cellular machines. *Trends Cell Biol.* 2006; 16(3):144–50. [PubMed: 16459078]
40. Frank J. Single-particle reconstruction of biological macromolecules in electron microscopy--30 years. *Q Rev Biophys.* 2009; 42(3):139–58. [PubMed: 20025794]
41. Henrick K, et al. EMDep: a web-based system for the deposition and validation of high-resolution electron microscopy macromolecular structural information. *Journal of Structural Biology.* 2003; 144(1–2):228–237. [PubMed: 14643225]
42. Chiu W, et al. Electron cryomicroscopy of biological machines at subnanometer resolution. *Structure (Camb).* 2005; 13(3):363–72. [PubMed: 15766537]

43. Zhou ZH. Towards atomic resolution structural determination by single-particle cryo-electron microscopy. *Curr Opin Struct Biol.* 2008; 18(2):218–28. [PubMed: 18403197]
44. Lindert S, Stewart PL, Meiler J. Hybrid approaches: applying computational methods in cryo-electron microscopy. *Curr Opin Struct Biol.* 2009; 19(2):218–25. [PubMed: 19339173]
45. Stahlberg H, Walz T. Molecular electron microscopy: state of the art and current challenges. *ACS Chem Biol.* 2008; 3(5):268–81. [PubMed: 18484707]
46. Yonekura K, Maki-Yonekura S, Namba K. Complete atomic model of the bacterial flagellar filament by electron cryomicroscopy. *Nature.* 2003; 424(6949):643–650. [PubMed: 12904785]
47. Yu X, Jin L, Zhou ZH. 3.88 Å structure of cytoplasmic polyhedrosis virus by cryo-electron microscopy. *Nature.* 2008; 453(7193):415–9. [PubMed: 18449192]
48. Zhang X, et al. Near-atomic resolution using electron cryomicroscopy and single-particle reconstruction. *Proc Natl Acad Sci U S A.* 2008; 105(6):1867–72. [PubMed: 18238898]
49. Jiang W, et al. Backbone structure of the infectious epsilon15 virus capsid revealed by electron cryomicroscopy. *Nature.* 2008; 451(7182):1130–4. [PubMed: 18305544]
50. Ludtke SJ, et al. De novo backbone trace of GroEL from single particle electron cryomicroscopy. *Structure.* 2008; 16(3):441–8. [PubMed: 18334219]
51. Orlova EV, Saibil HR. Structure determination of macromolecular assemblies by single-particle analysis of cryo-electron micrographs. *Curr Opin Struct Biol.* 2004; 14(5):584–90. [PubMed: 15465319]
52. Rossmann MG, et al. Combining X-ray crystallography and electron microscopy. *Structure (Camb).* 2005; 13(3):355–62. [PubMed: 15766536]
53. Wriggers W, Chacon P. Modeling tricks and fitting techniques for multiresolution structures. *Structure (Camb).* 2001; 9(9):779–88. [PubMed: 11566128]
54. Fabiola F, Chapman MS. Fitting of high-resolution structures into electron microscopy reconstruction images. *Structure (Camb).* 2005; 13(3):389–400. [PubMed: 15766540]
55. Topf M, Sali A. Combining electron microscopy and comparative protein structure modeling. *Curr Opin Struct Biol.* 2005; 15(5):578–85. [PubMed: 16118050]
56. Chacon P, Wriggers W. Multi-resolution contour-based fitting of macromolecular structures. *J Mol Biol.* 2002; 317(3):375–84. [PubMed: 11922671]
57. Katchalski-Katzir E, et al. Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques. *Proc Natl Acad Sci U S A.* 1992; 89(6):2195–9. [PubMed: 1549581]
58. Garzon JI, et al. ADP_EM: fast exhaustive multi-resolution docking for high-throughput coverage. *Bioinformatics.* 2007; 23(4):427–33. [PubMed: 17150992]
59. Wu X, et al. A core-weighted fitting method for docking atomic structures into low-resolution maps: application to cryo-electron microscopy. *J Struct Biol.* 2003; 141(1):63–76. [PubMed: 12576021]
60. Topf M, et al. Structural characterization of components of protein assemblies by comparative modeling and electron cryo-microscopy. *J Struct Biol.* 2005; 149(2):191–203. [PubMed: 15681235]
61. Martinetz TM, Berkovich SG, Schulten KJ. Neural-gas' network for vector quantization and its application to time-series prediction. *IEEE Trans Neural Netw.* 1993; 4(4):558–69. [PubMed: 18267757]
62. Wriggers W, Milligan RA, McCammon JA. Situs: A package for docking crystal structures into low-resolution maps from electron microscopy. *J Struct Biol.* 1999; 125(2–3):185–95. [PubMed: 10222274]
63. Birmanns S, Wriggers W. Multi-resolution anchor-point registration of biomolecular assemblies and their components. *J Struct Biol.* 2007; 157(1):271–80. [PubMed: 17029847]
64. Goddard TD, Huang CC, Ferrin TE. Visualizing density maps with UCSF Chimera. *J Struct Biol.* 2007; 157(1):281–287. [PubMed: 16963278]
65. Spahn CM, Penczek PA. Exploring conformational modes of macromolecular assemblies by multiparticle cryo-EM. *Curr Opin Struct Biol.* 2009; 19(5):623–31. [PubMed: 19767196]

66. Baker D, Sali A. Protein structure prediction and structural genomics. *Science*. 2001; 294(5540): 93–6. [PubMed: 11588250]
67. Topf M, et al. Protein structure fitting and refinement guided by cryo-EM density. *Structure*. 2008; 16(2):295–307. [PubMed: 18275820]
68. Wendt T, et al. Three-dimensional image reconstruction of dephosphorylated smooth muscle heavy meromyosin reveals asymmetry in the interaction between myosin heads and placement of subfragment 2. *Proc Natl Acad Sci U S A*. 2001; 98(8):4361–6. [PubMed: 11287639]
69. Volkman N, et al. Evidence for cleft closure in actomyosin upon ADP release. *Nat Struct Biol*. 2000; 7(12):1147–55. [PubMed: 11101898]
70. Tama F, Wriggers W, Brooks CL 3rd. Exploring global distortions of biological macromolecules and assemblies from low-resolution structural information and elastic network theory. *J Mol Biol*. 2002; 321(2):297–305. [PubMed: 12144786]
71. Ma J. Usefulness and limitations of normal mode analysis in modeling dynamics of biomolecular complexes. *Structure (Camb)*. 2005; 13(3):373–80. [PubMed: 15766538]
72. Suhre K, Navaza J, Sanejouand YH. NORMA: a tool for flexible fitting of high-resolution protein structures into low-resolution electron-microscopy-derived density maps. *Acta Crystallogr D Biol Crystallogr*. 2006; 62(Pt 9):1098–100. [PubMed: 16929111]
73. Topf M, et al. Refinement of protein structures by iterative comparative modeling and CryoEM density fitting. *J Mol Biol*. 2006; 357(5):1655–68. [PubMed: 16490207]
74. Rawi R, Whitmore L, Topf M. CHOYCE - A web server for constrained homology modelling with cryoEM maps. *Bioinformatics*.
75. Baker ML, et al. Ab initio modeling of the herpesvirus VP26 core domain assessed by CryoEM density. *PLoS Comput Biol*. 2006; 2(10):e146. [PubMed: 17069457]
76. Velazquez-Muriel JA, et al. Flexible fitting in 3D-EM guided by the structural variability of protein superfamilies. *Structure*. 2006; 14(7):1115–26. [PubMed: 16843893]
77. Chandramouli P, et al. Structure of the mammalian 80S ribosome at 8.7 Å resolution. *Structure*. 2008; 16(4):535–48. [PubMed: 18400176]
78. Taylor DJ, et al. Comprehensive molecular structure of the eukaryotic ribosome. *Structure*. 2009; 17(12):1591–604. [PubMed: 20004163]
79. John B, Sali A. Comparative protein structure modeling by iterative alignment, model building and model assessment. *Nucleic Acids Res*. 2003; 31(14):3982–92. [PubMed: 12853614]
80. Serysheva II, et al. Subnanometer-resolution electron cryomicroscopy-based domain models for the cytoplasmic region of skeletal muscle RyR channel. *Proc Natl Acad Sci U S A*. 2008; 105(28): 9610–5. [PubMed: 18621707]
81. Rohl CA, et al. Protein structure prediction using Rosetta. *Methods Enzymol*. 2004; 383:66–93. [PubMed: 15063647]
82. Jiang W, et al. Bridging the information gap: computational tools for intermediate resolution structure interpretation. *J Mol Biol*. 2001; 308(5):1033–44. [PubMed: 11352589]
83. Chen JZ, et al. Low-resolution structure refinement in electron microscopy. *J Struct Biol*. 2003; 144(1–2):144–51. [PubMed: 14643217]
84. Trabuco LG, et al. Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. *Structure*. 2008; 16(5):673–83. [PubMed: 18462672]
85. DiMaio F, et al. Refinement of protein structures into low-resolution density maps using rosetta. *J Mol Biol*. 2009; 392(1):181–90. [PubMed: 19596339]
86. Zhu J, et al. Building and refining protein models within cryo-electron microscopy density maps based on homology modeling and multiscale structure refinement. *J Mol Biol*. 397(3):835–51. [PubMed: 20109465]
87. Connell SR, et al. A new tRNA intermediate revealed on the ribosome during EF4-mediated back-translocation. *Nat Struct Mol Biol*. 2008; 15(9):910–5. [PubMed: 19172743]
88. Yuan S, et al. Structure of an apoptosome-procaspase-9 CARD complex. *Structure*. 18(5):571–83. [PubMed: 20462491]
89. Tan RK, Devkota B, Harvey SC. YUP.SCX: coaxing atomic models into medium resolution electron density maps. *J Struct Biol*. 2008; 163(2):163–74. [PubMed: 18572416]

90. Seidelt B, et al. Structural insight into nascent polypeptide chain-mediated translational stalling. *Science*. 2009; 326(5958):1412–5. [PubMed: 19933110]
91. Becker T, et al. Structure of monomeric yeast and mammalian Sec61 complexes interacting with the translating ribosome. *Science*. 2009; 326(5958):1369–73. [PubMed: 19933108]
92. Jolley CC, et al. Fitting low-resolution cryo-EM maps of proteins using constrained geometric simulations. *Biophys J*. 2008; 94(5):1613–21. [PubMed: 17993504]
93. Schroder GF, Brunger AT, Levitt M. Combining efficient conformational sampling with a deformable elastic network model facilitates structure refinement at low resolution. *Structure*. 2007; 15(12):1630–41. [PubMed: 18073112]
94. Grubisic I, et al. Biased coarse-grained molecular dynamics simulation approach for flexible fitting of X-ray structure into cryo electron microscopy maps. *J Struct Biol*. 169(1):95–105. [PubMed: 19800974]
95. Rusu M, Birmanns S, Wriggers W. Biomolecular pleiomorphism probed by spatial interpolation of coarse models. *Bioinformatics*. 2008; 24(21):2460–6. [PubMed: 18757874]
96. Tama F, Miyashita O, Brooks CL 3rd. Flexible multi-scale fitting of atomic structures into low-resolution electron density maps with elastic network normal mode analysis. *J Mol Biol*. 2004; 337(4):985–99. [PubMed: 15033365]
97. Tama F, et al. Model of the toxic complex of anthrax: responsive conformational changes in both the lethal factor and the protective antigen heptamer. *Protein Sci*. 2006; 15(9):2190–200. [PubMed: 16943448]
98. Falke S, et al. The 13 angstroms structure of a chaperonin GroEL-protein substrate complex by cryo-electron microscopy. *J Mol Biol*. 2005; 348(1):219–30. [PubMed: 15808865]
99. Tama F, Sanejouand YH. Conformational change of proteins arising from normal mode calculations. *Protein Eng*. 2001; 14(1):1–6. [PubMed: 11287673]
100. Lasker K, et al. Inferential optimization for simultaneous fitting of multiple components into a CryoEM map of their assembly. *J Mol Biol*. 2009; 388(1):180–94. [PubMed: 19233204]
101. Kawabata T. Multiple subunit fitting into a low-resolution density map of a macromolecular complex using a gaussian mixture model. *Biophys J*. 2008; 95(10):4643–58. [PubMed: 18708469]
102. Zhang S, et al. A fast mathematical programming procedure for simultaneous fitting of assembly components into cryoEM density maps. *Bioinformatics (ISMB)*. 2010 in press.
103. Alber F, et al. Determining the architectures of macromolecular assemblies. *Nature*. 2007; 450(7170):683–94. [PubMed: 18046405]
104. Alber F, et al. The molecular architecture of the nuclear pore complex. *Nature*. 2007; 450(7170):695–701. [PubMed: 18046406]
105. Förster F, et al. An atomic model AAA-ATPase/20S core particle sub-complex of the 26S proteasome. *Biochem Biophys Res Commun*. 2009; 388(2):228–33. [PubMed: 19653995]
106. Lasker K, et al. Integrative structure modeling of macromolecular assemblies from proteomics data. *Mol Cell Proteomics*. 2010; 1(1):1.
107. Nickell S, et al. Insights into the molecular architecture of the 26S proteasome. *Proc Natl Acad Sci U S A*. 2009; 106(29):11943–7. [PubMed: 19581588]
108. Forster F, et al. An atomic model AAA-ATPase/20S core particle sub-complex of the 26S proteasome. *Biochem Biophys Res Commun*. 2009; 388(2):228–33. [PubMed: 19653995]
109. Lin AC, Holt CE. Function and regulation of local axonal translation. *Curr Opin Neurobiol*. 2008; 18(1):60–8. [PubMed: 18508259]
110. Brandt F, et al. The native 3D organization of bacterial polysomes. *Cell*. 2009; 136(2):261–71. [PubMed: 19167328]
111. Brandt F, et al. The three-dimensional organization of polyribosomes in intact human cells. *Mol Cell*. 39(4):560–9. [PubMed: 20797628]
112. Werner JN, et al. Quantitative genome-scale analysis of protein localization in an asymmetric bacterium. *Proc Natl Acad Sci USA*. 2009; 106(19):7858–63. [PubMed: 19416866]
113. Ben-Harush K, et al. The supramolecular organization of the *C. elegans* nuclear lamin filament. *J Mol Biol*. 2009; 386(5):1392–402. [PubMed: 19109977]

114. Foerster F, Han BG, Beck M. Visual Proteomics. Book. 2010; 1(1):1.
115. Grünewald K. Adding a spatial dimension to the proteome. *Nat Methods*. 2009; 6(11):798–800. [PubMed: 19876016]
116. Lucic V, Förster F, Baumeister W. Structural Studies by Electron Tomography: From Cells to Molecules. *Annu Rev Biochem*. 2005; 74:833–65. [PubMed: 15952904]
117. Frangakis AS, et al. Identification of macromolecular complexes in cryoelectron tomograms of phantom cells. *Proc Natl Acad Sci USA*. 2002; 99(22):14153–8. [PubMed: 12391313]
118. Böhm J, et al. Toward detecting and identifying macromolecules in a cellular context: template matching applied to electron tomograms. *Proc Natl Acad Sci U S A*. 2000; 97(26):14245–50. [PubMed: 11087814]
119. Lucic V, Forster F, Baumeister W. Structural studies by electron tomography: from cells to molecules. *Annu Rev Biochem*. 2005; 74:833–65. [PubMed: 15952904]
120. Han BG, et al. Survey of large protein complexes in *D. vulgaris* reveals great structural diversity. *Proc Natl Acad Sci U S A*. 2009; 106(39):16580–5. [PubMed: 19805340]
121. Malmstrom J, et al. Proteome-wide cellular protein concentrations of the human pathogen *Leptospira interrogans*. *Nature*. 2009; 460(7256):762–5. [PubMed: 19606093]
122. Foerster, F. PhD thesis. Vol. 1. Technical University of Munich; 2005. p. 1
123. Forster F, et al. Retrovirus envelope protein complex structure in situ studied by cryo-electron tomography. *Proc Natl Acad Sci U S A*. 2005; 102(13):4729–34. [PubMed: 15774580]
124. Roseman AM. FindEM--a fast, efficient program for automatic selection of particles from electron micrographs. *J Struct Biol*. 2004; 145(1–2):91–9. [PubMed: 15065677]
125. Ortiz JO, et al. Mapping 70S ribosomes in intact cells by cryoelectron tomography and pattern recognition. *J Struct Biol*. 2006; 156(2):334–41. [PubMed: 16857386]
126. Keller A, et al. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal Chem*. 2002; 74(20):5383–92. [PubMed: 12403597]
127. Ortiz J, et al. Mapping 70S ribosomes in intact cells by cryo-electron tomography and pattern recognition. *J Struct Biol*. 2006 in press.
128. Bartesaghi A, Subramaniam S. Membrane protein structure determination using cryo-electron tomography and 3D image averaging. *Curr Opin Struct Biol*. 2009; 19(4):402–7. [PubMed: 19646859]
129. Förster F, Hegerl R. Structure determination in situ by averaging of tomograms. *Methods Cell Biol*. 2007; 79:741–67. [PubMed: 17327182]
130. Briegel A, et al. Location and architecture of the *Caulobacter crescentus* chemoreceptor array. *Mol Microbiol*. 2008; 69(1):30–41. [PubMed: 18363791]
131. Ben-Harush K, et al. Visualizing cellular processes at the molecular level by cryo-electron tomography. *J Cell Sci*. 2010; 123(Pt 1):7–12. [PubMed: 20016061]
132. Maurer UE, Sodeik B, Grünewald K. Native 3D intermediates of membrane fusion in herpes simplex virus 1 entry. *Proc Natl Acad Sci USA*. 2008; 105(30):10559–64. [PubMed: 18653756]
133. Murata K, et al. Zernike phase contrast cryo-electron microscopy and tomography for structure determination at nanometer and subnanometer resolutions. *Structure*. 18(8):903–12. [PubMed: 20696391]
134. Rigort A, et al. Micromachining tools and correlative approaches for cellular cryo-electron tomography. *Journal of Structural Biology*. 2010
135. McMullan G, et al. Detective quantum efficiency of electron area detectors in electron microscopy. *Ultramicroscopy*. 2009; 109(9):1126–43. [PubMed: 19497671]
136. Terry LJ, Shows EB, Wentz SR. Crossing the nuclear envelope: hierarchical regulation of nucleocytoplasmic transport. *Science*. 2007; 318(5855):1412–6. [PubMed: 18048681]
137. Schaff J, et al. A general computational framework for modeling cellular structure and function. *Biophysical Journal*. 1997; 73(3):1135–46. [PubMed: 9284281]
138. Slepchenko BM, et al. Quantitative cell biology with the Virtual Cell. *Trends Cell Biol*. 2003; 13(11):570–6. [PubMed: 14573350]
139. Hattne J, Fange D, Elf J. Stochastic reaction-diffusion simulation with MesoRD. *Bioinformatics*. 2005; 21(12):2923–4. [PubMed: 15817692]

140. Ander M, et al. SmartCell, a framework to simulate cellular processes that combines stochastic approximation with diffusion and localisation: analysis of simple networks. *Syst Biol* (Stevenage). 2004; 1(1):129–38. [PubMed: 17052123]
141. Raser JM, O'Shea EK. Noise in gene expression: origins, consequences, and control. *Science*. 2005; 309(5743):2010–3. [PubMed: 16179466]
142. Cai L, Friedman N, Xie XS. Stochastic protein expression in individual cells at the single molecule level. *Nature*. 2006; 440(7082):358–62. [PubMed: 16541077]
143. Kollmann M, et al. Design principles of a bacterial signalling network. *Nature*. 2005; 438(7067):504–7. [PubMed: 16306993]
144. Turner TE, Schnell S, Burrage K. Stochastic approaches for modelling in vivo reactions. *Comput Biol Chem*. 2004; 28(3):165–78. [PubMed: 15261147]
145. Gillespie DT. Exact Stochastic Simulation of Coupled Chemical Reactions. *The Journal of Physical Chemistry*. 1977; 81(25):2340–2361.
146. Rathinam M, et al. Stiffness in stochastic chemically reacting systems: The implicit tau-leaping method. *Journal of chemical physics*. 2003; 119(24):12784–12794.
147. Ermak DL, McCammon JA. Brownian dynamics with hydrodynamic interactions. *The Journal of chemical physics*. 1978
148. Ridgway D, et al. Coarse-grained molecular simulation of diffusion and reaction kinetics in a crowded virtual cytoplasm. *Biophysical Journal*. 2008; 94(10):3748–59. [PubMed: 18234819]
149. Boulianne L, et al. GridCell: a stochastic particle-based biological. system simulator. *BMC systems biology*. 2008; 2(1):66. [PubMed: 18651956]
150. Andrews SS, Bray D. Stochastic simulation of chemical reactions with spatial resolution and single molecule detail. *Physical biology*. 2004; 1(3–4):137–51. [PubMed: 16204833]
151. Sun J, Weinstein H. Toward realistic modeling of dynamic processes in cell signaling: quantification of macromolecular crowding effects. *The Journal of chemical physics*. 2007; 127(15):155105. [PubMed: 17949221]
152. Coggan JS, et al. Evidence for ectopic neurotransmission at a neuronal synapse. *Science*. 2005; 309(5733):446–51. [PubMed: 16020730]
153. Lipkow K. Changing cellular location of CheZ predicted by molecular simulations. *PLoS Comput Biol*. 2006; 2(4):e39. [PubMed: 16683020]
154. Goldman JP, Levin MD, Bray D. Signal amplification in a lattice of coupled protein kinases. *Mol Biosyst*. 2009
155. Beck M, et al. Snapshots of nuclear pore complexes in action captured by cryo-electron tomography. *Nature*. 2007; 449(7162):611–5. [PubMed: 17851530]
156. Minton AP. Macromolecular crowding. *Curr Biol*. 2006; 16(8):R269–71. [PubMed: 16631567]
157. Ellis RJ, Minton AP. Cell biology: join the crowd. *Nature*. 2003; 425(6953):27–8. [PubMed: 12955122]
158. Kim JS, Yethiraj A. Effect of macromolecular crowding on reaction rates: a computational and theoretical study. *Biophysical Journal*. 2009; 96(4):1333–40. [PubMed: 19217851]
159. Golding I, Cox EC. Physical nature of bacterial cytoplasm. *Phys Rev Lett*. 2006; 96(9):098102. [PubMed: 16606319]
160. Weiss M, et al. Anomalous subdiffusion is a measure for cytoplasmic crowding in living cells. *Biophys J*. 2004; 87(5):3518–24. [PubMed: 15339818]
161. Schreiber G, Haran G, Zhou HX. Fundamental Aspects of Protein-Protein Association Kinetics. *Chemical Reviews*. 2008; 109(3):839–860. [PubMed: 19196002]
162. Wieczorek G, Zielenkiewicz P. Influence of macromolecular crowding on protein-protein association rates--a Brownian dynamics study. *Biophysical Journal*. 2008; 95(11):5030–6. [PubMed: 18757562]
163. Ando T, Skolnick J. Crowding and hydrodynamic interactions likely dominate in vivo macromolecular motion. *Proc Natl Acad Sci U S A*.
164. Urbina-Villalba G, García-Sucre M, Toro-Mendoza J. Average hydrodynamic correction for the Brownian dynamics calculation of flocculation rates. *Physical Review E*. 2003
165. Tokuyama M, Oppenheim I. Dynamics of hard-sphere suspensions. *Physical Review E*. 1994

166. Dzubiella J, McCammon JA. Substrate concentration dependence of the diffusion-controlled steady-state rate constant. *The Journal of chemical physics*. 2005; 122(18):184902. [PubMed: 15918760]
167. Heyes DM. Brownian dynamics simulations of self and collective diffusion of near hard sphere colloidal liquids: inclusion of many-body hydrodynamics. *Molecular Physics*. 1996; 87(2):287–298.
168. Banchio AJ, Brady JF. Accelerated Stokesian dynamics: Brownian motion. *Journal of Chemical Physics*. 2003; 118:10323–10333.
169. Tanaka H, Araki T. Simulation Method of Colloidal Suspensions with Hydrodynamic Interactions: Fluid Particle Dynamics. *Phys Rev Lett*. 2000; 85(6):1338–1341. [PubMed: 10991546]
170. Tucci K, Kapral R. Mesoscopic model for diffusion-influenced reaction dynamics. *Journal of chemical physics*. 2004; 120(7):8262–8270. [PubMed: 15267746]
171. Hansen JP, Lowen H. Effective interactions between electric double layers. *Annual Review of Physical Chemistry*. 2000; 51(20):209–242.
172. McGuffee SR, Elcock AH. Diffusion, crowding & protein stability in a dynamic molecular model of the bacterial cytoplasm. *PLoS Comput Biol*. 2010; 6(3):e1000694. [PubMed: 20221255]
173. Elcock AH. Models of macromolecular crowding effects and the need for quantitative comparisons with experiment. *Current Opinion in Structural Biology*. 2010; 20(2):196–206. [PubMed: 20167475]
174. Takahashi K, Tanase-Nicola S, ten Wolde PR. Spatio-temporal correlations can drastically change the response of a MAPK pathway. *Proc Natl Acad Sci U S A*. 2009; 107(6):2473–2478. [PubMed: 20133748]
175. van Zon JS, ten Wolde PR. Green's-function reaction dynamics: a particle-based approach for simulating biochemical networks in time and space. *The Journal of chemical physics*. 2005; 123(23):234910. [PubMed: 16392952]
176. Opplestrup T, et al. First-Passage Monte Carlo Algorithm: Diffusion without All the Hops. *Phys Rev Lett*. 2006; 97(23):230602–230606. [PubMed: 17280187]
177. Evans RN, et al. The structure of LepA, the ribosomal back translocase. *Proc Natl Acad Sci U S A*. 2008; 105(12):4673–8. [PubMed: 18362332]
178. Connell SR, et al. Structural basis for interaction of the ribosome with the switch regions of GTP-bound elongation factors. *Mol Cell*. 2007; 25(5):751–64. [PubMed: 17349960]
179. Volkman N, Hanein D. Quantitative fitting of atomic models into observed densities derived by electron microscopy. *J Struct Biol*. 1999; 125(2–3):176–84. [PubMed: 10222273]
180. Rossmann MG. Fitting atomic models into electron-microscopy maps. *Acta Crystallogr D Biol Crystallogr*. 2000; 56(Pt 10):1341–9. [PubMed: 10998631]
181. Roseman AM. Docking structures of domains into maps from cryo-electron microscopy using local correlation. *Acta Crystallogr D Biol Crystallogr*. 2000; 56(Pt 10):1332–40. [PubMed: 10998630]
182. Navaza J, et al. On the fitting of model electron densities into EM reconstructions: a reciprocal-space formulation. *Acta Crystallogr D Biol Crystallogr*. 2002; 58(Pt 10 Pt 2):1820–5. [PubMed: 12351826]
183. Siebert X, Navaza J. UROX 2.0: an interactive tool for fitting atomic models into electron-microscopy reconstructions. *Acta Crystallogr D Biol Crystallogr*. 2009; 65(Pt 7):651–8. [PubMed: 19564685]
184. Kovacs JA, et al. Fast rotational matching of rigid bodies by fast Fourier transform acceleration of five degrees of freedom. *Acta Crystallogr D Biol Crystallogr*. 2003; 59(Pt 8):1371–6. [PubMed: 12876338]
185. Ceulemans H, Russell RB. Fast Fitting of Atomic Structures to Low-resolution Electron Density Maps by Surface Overlap Maximization. *J Mol Biol*. 2004; 338(4):783–93. [PubMed: 15099745]
186. Orzechowski M, Tama F. Flexible fitting of high-resolution x-ray structures into cryoelectron microscopy maps using biased molecular dynamics simulations. *Biophysical Journal*. 2008; 95(12):5692–705. [PubMed: 18849406]

Figure 1.

Bridging molecular and cellular biology. (Upper panels) Structural information gathered at different levels of organization, ranging from atomic structures of macromolecules (left), to the molecular architectures of larger cellular components (middle) and the higher order proteome organization (right). The right scheme shows a model of the intracellular organization of *E. coli* (Reprinted with permission of David S. Goodsell, The Scripps Research Institute, La Jolla, CA.). (Lower panels) Selection of experimental and computational methods that can provide spatial information of the proteome organization at the various scales of resolution. (SAXS: Small-angle X-ray scattering. TEM: Transmission electron microscopy).

Figure 2.

Pseudo-atomic models of the yeast 60S (A) and 40S (B) ribosomal subunits determined by combining cryoEM density fitting (at ~ 9 Å resolution) with RNA and protein homology modeling. YUP.SCX [89] and Flex-EM [67] were applied to refine the models. Ribosomal RNA is represented as tubes. 5S and 5.8S rRNA are shown as magenta and green tubes, respectively. Ribosomal proteins, with known homologs and placement, are shown as orange (A) and pink (B) cartoons. Density accounting for proteins that are unique to eukaryotes with no structural homolog available or that have not been localized within the context of the ribosome is represented in yellow (A) and slate (B). The boundary of each cryoEM map is shown as a gray mesh. (Figure is reproduced from Reference [78]).

Figure 3.

The molecular model of ribosome-bound EF4 (rbEF4) in the process of back-translocating tRNAs. (A) The model of rbEF4 based on Flex-EM refinement of the unbound EF4 crystal structure (PDB 3CB4 [177]) in the cryoEM density (mesh), which was computationally segmented from the density of the entire 70S–EF4 complex (at ~11 Å resolution). (B) The individual domains (corresponding to the EF-G domain nomenclature, [178]) are colored blue (I), green (II), yellow (III), dark orange (V) lighter orange (V') and red (CTD). (B) The EF4 X-ray structure (gray cartoon) superimposed on the rbEF4 structure by aligning domain I to illustrate the domain rearrangement between the two. (C,D) The contrasting interaction of domain IV in rbEF-G [178] (PDB 2OM7) and domain V' in rbEF4 with the tRNA intermediate (A/L-tRNA, purple ladder representation) is shown (rbEF-G superimposed on the cryoEM density of rbEF4). Domain IV and the G' sub-domain of EF-G (orange) have no equivalent in EF4 (d, colored as in a). The absence of domain IV in EF4 is a prerequisite for allowing the tRNA to access the A-site during the back-translocation reaction (from the P-site to the A-site). The contacts with the A/L-tRNA (via sub-domain V' and the CTD), revealed by flexible fitting, allow EF4–GTP to form a ternary complex with tRNA on the ribosome. (Figure is reproduced from Reference [87]).

Figure 4.

Visual proteomics of the human pathogen *Leptospira interrogans*. A. Score distributions obtained from artificial tomograms of the observed (blue), true positive (green) false positive hits (red). The arrowhead marks the threshold that accounts for a specificity of 40%. B. Centered slices through a artificial *Leptospira* cell shown perpendicular (left) and parallel (right) to the electron optical axis, as well as before (top) and after (bottom) simulating the image formation process. Numbers account for protein complexes as described below. C. Centered slice through a tomographic reconstruction, oriented perpendicular to the electron optical axis. D. The templates assigned into this volume are shown surface rendered. 70S ribosomes (1) are shown in dark brown, 2 RNA-polymerase II (2) in green, ATP-synthase (3) in bright brown, GroEL in bright blue, GroEL-ES (4) in dark blue, Hsp (5) in red, cytoplasmic membrane in transparent blue and the cell wall in transparent brown). (Figure is adapted from Reference [16].)

Figure 5.

By a combination of pattern recognition and classification algorithms, the following TAP-identified complexes from *M. pneumoniae*, matching to existing electron microscopy and x-ray and tomogram structures (A), were placed in a whole-cell tomogram (B): the structural core of pyruvate dehydrogenase in blue (~23 nm), the ribosome in yellow (~26 nm), RNA polymerase in purple (~17 nm), and GroEL homo-multimer in red (~20 nm). Cell dimensions are ~300 nm by 700 nm. The cell membrane is shown in light blue. The rod, a prominent structure filling the space of the tip region, is depicted in green. Its major structural elements are HMW2 (Mpn310) in the core and HMW3 (Mpn452) in the periphery, stabilizing the rod (42). The individual complexes (A) are not to scale, but they are shown to scale within the bacterial cell (B). (Figure is reproduced from Reference [3]).

Figure 6.

(A) Smoldyn simulations of the chemotaxis pathway of the bacterium *Escherichia coli*. Each point represents the position of a molecule, the color of the point indicates its specific type. The simulations show differential localization of CheZ and its oligomeric forms. Upper left panel: In the absence of CheA_{long} kinase activity and phosphorylated CheY, all CheZ dimers are unbound and freely diffusing in the cytoplasm. Middle left panel: Upon sudden increase of kinase activity, the level of CheYp rises in the entire cell. Lower left panel: After 1 min at constant kinase activity, the increase of CheYp has led to the formation of oligomeric CheZ₂Yp clusters at the inner face of the polar receptor cluster shown as black circles. Panel A is reproduced from Reference [153].

(B) Hard-sphere representation of the *E. coli* cytoplasm at 50% volume fraction occupied. Spheres represent various macromolecules and their complexes. The particle-size distribution was estimated from the relative abundance, mass, and stoichiometries of protein complexes using an experimentally derived proteome catalog from *E. coli* K12. Panel B is reproduced from reference [148].

(C) Atomically detailed model of the *E. coli* cytoplasmic environment that includes 50 of the most abundant types of macromolecules at experimentally measured concentrations. Rendering shows the cytoplasm model at the end of a Brownian dynamics simulation performed. Panel C is reproduced from Reference [172].

Table 1

Computational methods for fitting of atomic structures into cryo-EM density maps.

Rigid Fitting	<i>Density based fitting</i> COAN [179], EMFit [180], DOCKEM [181], FoldHunter [82], Situs/Colores [56] [62], UROX [182, 183], FRM [184], grid- threading MC [59], 3SOM [185], Mod-EM [60], ADP_EM [58], Chimera/Fit_in_Map [64] <i>Point-matching methods</i> SITUS [62] [63]
Flexible Fitting	<i>Fitting multiple component conformations</i> NMA [70, 71], NORMA[72, 183], Moulder-EM [73], S-flexfit [76], Rosetta/Foldhunter [75], CHOYCE [74]),
	<i>Refinement methods</i> RSRef [83], NMFF[96], Flex-EM [67], MDFF [84], Rosetta [85], YUP.SCX [89], FIRST/FRODA [92], DireX [93], EM-IMO [86], Cross-correlation based MD [186], SITUS [62] [63, 95]
Simultaneous Fitting of Components	GMFIT [101], MultiFit [100], IQP [102]
Comprehensive data integration	IMP [35, 103–106]