Title

Principles for Integrative Structural Biology Studies

Permalink

https://escholarship.org/uc/item/8sr1026k

Journal

Cell, 177(6)

ISSN

0092-8674

Authors

Rout, Michael P
Sali, Andrej

Publication Date

2019-05-01

DOI

10.1016/j.cell.2019.05.016

Peer reviewed

Conducting an Integrative Structural Biology Study

Michael P. Rout[1], **Andrej Sali**[2]

[1].Laboratory of Cellular and Structural Biology, The Rockefeller University, New York, NY 10065, USA.

[2].Department of Bioengineering and Therapeutic Sciences, Department of Pharmaceutical Chemistry, California Institute for Quantitative Biosciences, Byers Hall, 1700 4th Street, Suite 503B, University of California, San Francisco, San Francisco, CA 94158, USA.

Abstract

Integrative structure determination is a powerful approach to mapping the structures of biological systems, with implications for our understanding of cellular biology and drug discovery. Here, we provide a Primer to introduce the researcher to the theory and methods of integrative approaches, relying on examples to illustrate the practice and challenges involved. These guidelines are intended to aid the researcher in applying integrative structural methods to systems of their interest and thus take advantage of this rapidly evolving field.

Introduction

Our understanding of biological macromolecular systems comes from gathering sufficient information about them from experiments and prior models. Depictions of the spatial and temporal arrangements of these systems are especially helpful in formulating hypotheses about their function and evolution. This mindset is often summarized by two quotes: "structure without function is a corpse; function without structure is a ghost" (Vogel and Wainwright, 1969); and, "nothing in biology makes sense except in the light of evolution" (Dobzhansky, 1973).

Historically, X-ray crystallography and NMR spectroscopy allowed us to determine atomic structures of smaller systems, such as single proteins. Larger systems were depicted at a correspondingly coarser granularity, commensurate to the data used (*eg*, electron or light microscopy images). Now, we are trying to map systems consisting of hundreds of macromolecules (*eg*, nuclear pore complexes, centrosomes), which nevertheless need to be depicted at a high level of detail. Moreover, we wish to describe the dynamics of these systems as they assemble, disassemble, function, and undergo regulation *via* interactions with other such systems. These descriptions also have to be sufficiently informative to allow us to modulate their functions, both to further study their mechanisms and for therapeutic interventions. It is here that traditional structure determination methods can fall short, thus creating a need for different approaches.

rout@rockvax.rockefeller.edu.

Fortunately, one such approach already exists and has a long track record of success: integrative structural biology (Figure 1). In integrative approaches, disparate information, potentially at different scales, is synthesized into a common view of a system. The motivation behind the integrative approach is deceptively simple: namely, any system is described best (*ie*, most accurately, precisely, completely, and efficiently) by using all available information about it; in other words, if information about a system is available, why not use it! The integrative approach constructs a depiction of a system by simultaneously combining information from multiple sources, including varied experimental methods (Table 1) and prior models (physical theories, statistical analyses, and other models).

Integrative approaches actually date back to the very beginning of structural biology, and in a spectacular fashion: one of the first integrative structural models was that of the double helix of DNA (Franklin and Gosling, 1953; Watson and Crick, 1953). It was possible to generate a model of DNA that elegantly explained how genetic information is stored and propagated from one generation to the next, by combining information about its chemical composition, its stoichiometry, the complementarity of its component nucleotides, and X-ray fiber diffraction data about its helical geometry. None of these individual considerations were sufficient on their own; only together did they result in an informative model. The concept of integrating different types of data then moved through a series of methodological milestones towards the current formalization, as reviewed previously (Alber et al., 2007a; Alber et al., 2008; Joseph et al., 2017; Sali et al., 2015).

Integrative modeling as an optimization problem

To introduce how integrative structure determination methods work, it is helpful to first describe modeling approaches in general ("while it may be hard to live with generalization, it is inconceivable to live without it" (Gay, 2002)); these modeling approaches include all structure determinations based primarily on experimental data (such as X-ray crystallography), computational predictions (such as comparative modeling), and even manual models (such as sketching of schematic diagrams). A "model" in this sense is a depiction of a system or process of interest that is useful for rationalizing the existing information and for making predictions about outcomes of future experiments. Thus, modeling is the process of converting input information about a system into a model of the system. All modeling methods share a common design principle: among all possible models, they aim to find those models whose computed properties match the input information (*eg*, structures whose interatomic distances and dihedral angles match those determined by NMR spectroscopy). Critically, modeling should also include the propagation of the uncertainty of the input information and modeling into the uncertainty of the model; this goal is achieved directly and robustly by producing a sample of all single models sufficiently consistent with the input information, not only the "best" single model (*eg*, the ensemble of structures found in a Protein Data Bank (PDB) entry for an NMR-derived structure, each one of which sufficiently satisfies the original data).

Modeling in general is best seen as an optimization in which input information can be used in five different ways, guided by maximizing the accuracy and precision of the model while

remaining computationally feasible: (i) representing components of a model with some variables, (ii) scoring a model for its consistency with input information, (iii) searching for good-scoring models, (iv) filtering models based on input information, and (v) validating the resulting models. We now discuss each one of these ways in turn.

First, information can be used to define the representation of a model (Box 1). The representation specifies the variables whose values will be determined by modeling, based on input information. For an artist, this is the medium of art - whether to paint, sculpt, or photograph. For a structural biologist, the representation of a model first specifies the components of the system (*eg*, atoms, coarse-grained particles, and subunits in a complex), including their copy numbers. Next, it specifies the component coordinates that are fit to the input information (*eg*, their positions, orientations, and conformations). It also specifies if multiple structural states (*eg*, for heterogeneous samples) or trajectories (*eg*, for dynamic systems) need to be modeled. Finally, the representation of a model can also include auxiliary variables that are fit to the input information (*eg*, weights of different types of data and other parameters of the scoring function). The representation is generally selected based on the amount and type of information available (*eg*, a ~30 Å resolution EM density alone does not justify using an atomic representation), the purpose of the model (*eg*, questions about enzymatic mechanism generally require atomic structures), and computational feasibility (*eg*, a rigid representation of subunits in molecular docking enables a systematic search for the binding mode). For difficult modeling problems, a decision about the representation is often critical and can present a fall at the first hurdle, such as trying to squeeze out atomic positions from a low-resolution electron microscopy map and a dash of optimism.

Second, information can be used to construct and compute a value of a scoring function. The scoring function quantifies the degree of a match between a tested model and the input information, for example whether a distance in a model satisfies input information that we actually have about the distance, such as a Nuclear Overhauser Effect or chemical cross-link observation. The most common scoring function is a weighted sum of spatial restraints; each restraint is a function of the deviation of the computed property of a model from its measurement. Consequently, the greater this deviation, the less consistent is the model with the input information - and the worse the score. Optimization of the score then produces models that satisfy the encoded information as well as possible; a good-scoring model is a model that sufficiently satisfies input information by some definition (see "good-scoring models" in Box 3). Examples of spatial restraints include a potential energy function from a molecular mechanics force field, upper distance bounds in NMR spectroscopy, target functions in X-ray crystallography, and a correlation coefficient between a model and an electron microscopy map. The most objective scoring function is a Bayesian posterior model density, in which data likelihoods act as spatial restraints and noise models are effectively their weights (Box 2).

Third, information can be used to constrain the model search space. Given information that your keys were most likely lost in your house, you can focus your search to the house, without completely excluding other areas, just in case. Although rarely computationally feasible, the best search is a systematic enumeration of a defined search space, going

through every possible model one by one with sufficient granularity. In practice, other methods, such as stochastic sampling *via* a Monte Carlo scheme (Allen and Tildesley, 1989; Metropolis, 1953), are often used. As an example, when modeling the quaternary structure of a complex, information that a certain domain spans the membrane can be used to constrain that domain's position only to the membrane during sampling (Alber et al., 2007a; Alber et al., 2007b).

Fourth, some information can be used for filtering good-scoring models after they are produced by searching. Such use is often the case for information that is computationally expensive to incorporate into a scoring function, which is commonly evaluated thousands or millions of times during sampling. An example is using a negative-stain electron microscopy class of a binary complex to filter out all but those molecular docking solutions whose 2D projections match the negative-stain electron microscopy class (Fernandez-Martinez et al., 2016; Fernandez-Martinez et al., 2012; Shi et al., 2014; Velazquez-Muriel et al., 2012).

Fifth, some information can be used only to validate the good-scoring models, without changing or filtering them. Just like scoring and filtering, validation also depends on assessing a degree of consistency between a model and some information not used to compute the model. An example is testing whether or not a site-directed mutagenesis phenotype is consistent with a model (*eg*, whether or not a mutation in a predicted catalytic site actually kills the function in an experiment).

An occasional criticism of integrative structural modeling is that it produces "only a model, and I don't even know what it means". But this judgement is rash, because *every* structure is a model, each one of which is computed based on some information as outlined above. In other words, if it is not understood how a structure is determined, there is a tendency to call it a model rather than a structure; it also tends to be called a model when the expected uncertainty is relatively high or even unknown (*eg*, when the uncertainty of data is not known). Hence, this Primer: it makes no fundamental difference if the molecular model is computed "only" from X-ray crystallography data, "only" from electron microscopy particle images, or from some combination of varied data, so long as the uncertainty of the model is properly quantified and taken into account when interpreting the model (Schneidman-Duhovny et al., 2014). If anything, because integrative modeling can take all the available information into account, integrative structures are in principle more accurate, precise, and complete than structures based on only a subset of information (Lasker et al., 2010; Lasker et al., 2009). Every piece of data, regardless of its precision, is useful if it is not over-interpreted.

An Illustrative Example: Integrative structure determination of the nuclear pore complex

The existing toolbox (Table 2) has already successfully produced integrative structures for a large number of complex systems, all of which were refractory to traditional methods of structural biology (Figure 1). For convenience, however, we focus mainly on one illustrative example: the yeast nuclear pore complex (NPC), structures for which have been solved using integrative structural techniques. Biologically, the NPC encapsulates many of the challenges presented individually by other assemblies. The NPC is a large (50 – 100 MDa) octagonally symmetric cylindrical macromolecular assembly, consisting in yeast of ~500 copies of 30

different structured and intrinsically disordered proteins collectively termed nucleoporins (Alber et al., 2007a; Alber et al., 2007b; Beck and Hurt, 2017; Knockenhauer and Schwartz, 2016). Embedded in the nuclear envelope, it is the only known conduit for trafficking between the nucleoplasm and cytoplasm, mediating the active exchange of a large range of select proteins and RNAs. As such, the NPC interfaces with the nucleoplasm, cytoplasm, and both the membrane and perinuclear cisterna of the nuclear envelope. Thus, it directly interacts with enormously diverse macromolecules, including transmembrane and lumenal nuclear envelope proteins, cytoplasmic proteins, chromatin and nuclear proteins, and ribonucleoproteins. These associations can exist in a large dynamic range, from ultrafast (such as with transported macromolecules) to ultrastable (such as between scaffold components in the NPC) (Baade and Kehlenbach, 2018; Beck and Hurt, 2017; De Magistris and Antonin, 2018; Knockenhauer and Schwartz, 2016; Raices and D'Angelo, 2012). It is this diversity that also presents a string of formidable challenges to traditional structure determination approaches, as the NPC is by nature huge, flexible, heterogeneous in shape and composition, and highly dynamic (Beck and Hurt, 2017; Knockenhauer and Schwartz, 2016). Thus, we chose by necessity to solve structures for its subcomplexes and the entire NPC assembly using integrative approaches.

The five ways of converting input information into a model, outlined above, are conveniently described as an iterative four stage process (Figure 2): (i) gathering input information; (ii) converting input information into system representation and spatial restraints; (iii) searching for models that satisfy the restraints; and (iv) validating the model. Next, we describe these four stages, as applied primarily to the NPC.

Stage 1: Information gathering

Ideally, we aim to collect all the kinds of information, at a sufficient depth and granularity, necessary to solve our structure at the highest precision (*ie*, smallest uncertainty). Practically speaking though, and particularly for difficult biological problems, methodological limitations mean that we often do not have the luxury of using the data we would like to have, but only the data that we can actually obtain. Nevertheless, there is still some flexibility available, in terms of deciding between which methods will give the most "bang for buck", that is, the most useful information for modeling per unit effort.

For NPCs, the ideal information might be an X-ray crystallographic dataset for an entire native purified or reconstituted assembly. However, as indicated above, the nature of the assembly precludes gathering such information, at least for the moment. So, what information can we collect that would be most useful? This decision should be a continuing dialog between the experimentalists and modelers. In our first effort to solve an NPC structure, almost two decades ago, the available technologies were significantly more limiting than today in terms of both the amount and precision of the information; cryo-electron tomography maps had resolutions of ~100 Å (Akey and Radermacher, 1993; Beck et al., 2004; Hinshaw et al., 1992; Yang et al., 1998), and few atomic structures of nucleoporins were available (Brohawn et al., 2009). These limitations in turn set a bound on the precision of the first structure published in 2007 (Figure 3) (Alber et al., 2007b).

An important benefit of integrative structure determination is that it facilitates the use of information from experiments that have generally not been used for structure determination. As a case in point, for our first coarse NPC structure, one such class of data were affinity capture experiments. In these assays, nucleoporins are genomically tagged with an affinity handle, allowing them to be copurified with subsets of other nucleoporins whose identity is then usually determined using mass spectrometry. Such experiments had been used with great success to identify other nucleoporins and even infer nearest neighbors (Grandi et al., 1993; Grandi et al., 1995a; Grandi et al., 1995b; Siniossoglou et al., 1996), but they had not been interpreted as formal spatial restraints that could be used to compute a structure. Nevertheless, these data could in fact be used as restraints; each affinity capture result, which we termed a "composite", defined the composition of a sample of one or more complexes that share the tagged protein. Thus, a model can be scored for consistency with this data by ascertaining whether or not it contains at least one of the possible complexes implied by the composite (Alber et al., 2007a; Alber et al., 2008). However, because each composite carries relatively little spatial information, the experimentalists were challenged to produce a large number of different composites, densely covering all the nucleoporins. In other examples of input information, combined experimental and bioinformatic information defined transmembrane regions in three nucleoporins, allowing us to restrain those regions to the NPC's pore membrane; sequence-based definition of domains and analytical ultracentrifugation of the nucleoporins informed the degree to which they were spherical *versus* elongated, giving us an approximate shape and size for every nucleoporin; and immunoelectron microscopy provided axial and radial distributions of the nucleoporins, albeit with a high uncertainty corresponding to approximately a third of the size of the NPC. Finally, once enough information had been gathered, integrative modeling allowed us to convert it into the molecular architecture of the complete assembly (Alber et al., 2007b). These examples illustrate how integrative modeling facilitates generating and using more information, and thus solving harder structure determination problems (Alber et al., 2007a; Alber et al., 2008; Sali et al., 2015).

With the coarse molecular architecture of the NPC in hand, we embarked on improving it by gathering additional and higher resolution data for higher resolution representations of nucleoporins and for more accurate and precise modeling of their configuration in the whole assembly (Figure 3A). An impressive repertoire of nucleoporin atomic structures was produced by X-ray crystallography and NMR spectroscopy, *via* the Protein Structure Initiative and sterling efforts from many groups (reviewed in refs. (Brohawn et al., 2009; Knockenhauer and Schwartz, 2016)). In addition, we determined the integrative structures of Pom152, Nup133, the heptameric Nup84 outer ring complex, and the cytoplasmic Nup82 export platform (Fernandez-Martinez et al., 2016; Kim et al., 2014; Shi et al., 2014; Upla et al., 2017). These integrative structures were informed by two additional types of data. First, we required that a good model had a projection whose shape matched 2D negative-stain electron microscopy class averages (Fernandez-Martinez et al., 2012; Velazquez-Muriel et al., 2012). Second, we also required that a good model reproduced the distances implied by chemical cross-links, detected through cross linking - mass spectrometry (XL-MS) (below).

With a more detailed representation of the NPC components in hand, two experimental technological advances enabled us to solve the 3D jigsaw puzzle of how these components

fit together into a much more precise NPC depiction. The first is the dramatic improvements in cryo-electron microscopy, which needs no further explanation here, having revolutionized structural biology in the last few years (Callaway, 2015; Danev and Baumeister, 2017; Murata and Wolf, 2018; Wells and Marsh, 2018). These improvements allowed us to produce a ~28 Å density map of the entire NPC. The second is XL-MS, where substantial improvements in MS sensitivity have allowed investigators to identify large numbers of residues in peptides from a protein or complex that can be chemically cross-linked together, so setting an upper limit on the native distance between those residues based on the length of the cross-linker (Fischer et al., 2013; Gingras et al., 2007; Lauber et al., 2012; Leitner et al., 2012a; Leitner et al., 2012b; Leitner et al., 2010; Rappsilber, 2012; Sinz, 2006; Walzthoeni et al., 2013). These improvements allowed us to generate ~3100 unique cross-links in the entire NPC. We also adopted several complementary experimental methods to unambiguously establish the NPC's composition and stoichiometry. Finally, together with the vastly improved models of NPC components and other information almost entirely distinct from that used for the first coarse structure, integrative modeling was able to produce a significantly more detailed structure of the NPC (Figure 3A) (Kim et al., 2018).

Any output structure will only be as good as its input information ("garbage in, garbage out"). We still needed a physical sample suitable for application of the new technologies discussed above, as previous methods to isolate NPCs were time-consuming and cumbersome, limiting throughput. We thus adapted our affinity capture approaches to rapidly and gently isolate preparations of entire native NPCs that were suitable for higher throughput electron microscopy and XL-MS analyses. The quality of samples for analysis is clearly crucial and integrative (or indeed any other) structural approaches cannot materialize a precise structure from low quality starting material. Even so, there are clear limitations to most samples that must be removed from their native environments for analysis. For the NPC, we remain aware that depletion of chromatin, pore membrane, and a cloud of accessory factors during the purification may have changed the structure compared to its completely native state(s).

More information can generally improve a representation of the system, its model, and the uncertainty estimate. Thus, most structures are a work in progress, especially if initially determined at lower resolution. It is often easy to overlook some valuable information that could have improved the precision of the structure. For example, we did not use our own new NMR structure (Upla et al., 2017) of an Ig domain in the pore membrane nucleoporin Pom152 for assigning the folds of the other Pom152 domains, as publicly available threading web servers could not consider our new structure at the time. This resulted in a lower precision structural model of Pom152 that lacked high resolution representation of a tenth Ig domain (Hao et al., 2018); simply updating our modeling with an atomic representation of this domain can result in an improved model of the entire Pom152 (Figure 3B). Thus, a structure and its validation should be continuously refined to reflect the data and modeling methods that are available at any given moment in time. Importantly, the structure should not be overinterpreted beyond its uncertainty, so that the key functional and evolutionary conclusions survive any adjustments in its depiction as new data and modeling methods become available (*cf*, improving the precision of the Pom152 model has nevertheless not altered our overall interpretation of the NPC structure (Kim et al., 2018)).

Stage 2: Converting input information into system representation and spatial restraints

As mentioned already, information from the first stage can be parsed in two ways at this second stage - into either a system representation or restraints. Deciding how to use input information for modeling is a point of much communication and synergy between the experimentalists and modelers, to make sure that input information is optimally interpreted.

Defining the representation of the modeled system is a non-automated task that depends on the available information, the experience of the researcher, and trial-and-error (Box 1); in particular, the granularity of representation should be commensurate with the amount and precision of information. The representation must also facilitate (i) answering biological questions of interest, (ii) constructing an accurate and efficiently computed scoring function to quantify the consistency of a model with the input information, and (iii) sampling of alternative models (Viswanath and Sali, 2019).

The goal of the first integrative structure determination of the NPC was to map the single static coarse configuration of the component nucleoporins, commensurate with the information available at the time (Alber et al., 2007a; Alber et al., 2007b). Thus, we used a coarse-grained representation of the nucleoporins: each nucleoporin was represented either by a single bead whose radius depended on the number of residues in its sequence or a flexible string of a small number of beads corresponding to individual sequence domains. In the absence of atomic structures or comparative models for most of the constituent nucleoporins, these representations were informed primarily by sequence comparisons to delimit individual domains and ultracentrifugation to inform the globularity of the shape. Subsequently, as X-ray, NMR, and integrative structures of the nucleoporins and their subcomplexes were determined (Knockenhauer and Schwartz, 2016), we were able to use these structures either as rigid or somewhat flexible pre-determined shapes during their computational assembly into the structure of the entire NPC (Kim et al., 2018). In addition, we incorporated information about the nuclear envelope by representing it as a rigid layer of defined shape, size, and thickness that helps to organize the nucleoporins.

Spatial restraints are defined based on a subset of input information, considering the uncertainty in this information as much as possible, and then summed into the scoring function. For the most recent NPC structure, the restraints relied on chemical cross-links, the cryo-electron microscopy density map, immuno-electron microscopy localizations, excluded volume considerations, sequence connectivity, the shape of the pore membrane, and sequence-based membrane localizations (*eg*, the nuclear envelope can only be spanned by transmembrane regions in a fraction of nucleoporins). Definition of the resulting restraints from different types of data was facilitated by the use of a multi-scale representation of the NPC components; for example, chemical cross-links restrained distances between pairs of residues, while membrane localizations constrained entire domains to the membrane (Kim et al., 2018).

An advantage of integrative approaches is that they can include any information about flexible or even completely unstructured regions, such as intrinsically disordered regions (IDRs) in proteins, although they cannot be seen in X-ray and electron microscopy density maps. This advantage is an important consideration, because significant IDRs are found in as

much as a third of eukaryotic proteins and play fundamental and diverse roles in protein interactions and regulation (Oldfield and Dunker, 2014; Sharma et al., 2015; Uversky, 2017). IDRs make up one-fifth of the yeast NPC's mass and are critical to its structural integrity and transport functionality (Fischer et al., 2015; Kim et al., 2018; Knockenhauer and Schwartz, 2016). Thus, being able to depict these regions as flexible strings of beads was another benefit to choosing the integrative structure determination route.

The scoring function allows us to account for uncertainties and mistakes in input information, but the same cannot be said of representation. That is, information that is used for representation must be highly certain, while scoring function can readily incorporate uncertain information. For example, an incorrect stoichiometry or assuming an incorrect rigid protein shape will result in an incorrect representation that is in turn likely to result in incorrect models. In contrast, a cross-linking dataset with some incorrect cross-links can still be used successfully *via* appropriate cross-linking restraints and soft thresholds in the scoring function. Fortunately for the experimentalist, the majority of the input information is usually used to formulate the scoring function, and near perfection is only required for a subset of information going towards the representation.

Potential problems with converting input information into representation and scoring can be illustrated by the considerations needed when modeling a generic, pan-species version of the system based on data about the system from different species. Using data from different species to model a single structure is clearly appropriate only if the differences in composition, stoichiometry, and structure between the species are smaller than the uncertainties in the data. For example, NPCs from different organisms can have very different stoichiometries, morphologies, and even compositions. The average human NPC is approximately twice the mass of that from yeast, and also substantially differs in the composition and arrangement of its more peripheral complexes as well conformations of its components (Alber et al., 2007b; Kim et al., 2018; Kosinski et al., 2016; Mosalaganti et al., 2018; Ori et al., 2013). Exemplifying these pitfalls, a recent generic model meant to represent "the" NPC scaffold, generated by amalgamation of human and fungal data (Lin et al., 2016), is thus not an accurate representation of either a fungal or human NPC.

Stage 3: Searching for models that satisfy the restraints

The purpose of the third stage is to find a sample of all models that are sufficiently consistent with input information, as quantified by the scoring function. If a Bayesian scoring function is used (Box 2), the goal of sampling methods is to accurately sample the posterior model density. The search for good-scoring models is often achieved by a stochastic sampling of alternative structures, avoiding the biases and limitations intrinsic to searches for good-scoring models by humans. The sampling must be done at a precision that is higher than needed for interpreting the models. Potentially, the sampling can be made more efficient by limiting or guiding it based on a subset of input information (*cf*, the "search for keys", above). For example, the search for good-scoring NPC structures relied on a stochastic Monte Carlo scheme that benefited from being constrained to solutions with the 8-fold rotational symmetry of the NPC, an essentially universal feature of the assembly. Optionally, the sampling can be followed by filtering the good-scoring models from

sampling based on some information not used for representation, scoring, or sampling. Such filtering may be useful when using the corresponding information for scoring is computationally costly. Using information only for filtering, however, requires that the sampling generates solutions consistent with the filter even in the absence of the corresponding information from the scoring function considered during sampling. Notably, not all existing modeling methods aim to find a representative sample of all good-scoring solutions, thus vastly overestimating the precision of their solutions. A good example of this pitfall is when modeling is done in one's head, which assumes one can think of all possible models, even though in reality many unimagined models will simply go unexplored. Computer-assisted sampling and estimates of sampling precision can avoid such biases (Viswanath et al., 2017b).

A large variety of sampling methods have been developed. Enumeration of all possible solutions at a sufficiently high precision is an ideal sampling method, as it ensures no solution is missed (Lasker et al., 2012; Lasker et al., 2009), but it is generally not feasible for large systems with many degrees of freedom that need to be sampled finely. Efficient and well-known methods for local refinement of structures include conjugate gradients and steepest descent (Press et al., 2007). Often, however, the structural sampling does not benefit from knowing a starting structure that is close to the correct model, and thus stochastic sampling methods need to be employed. A large variety of such methods have been developed over the years, including many versions of Monte Carlo simulated annealing, replica exchange, Gibbs sampling, and Hamiltonian Monte Carlo (Betancourt, 2017). For stochastic sampling methods, it is imperative that tests of thoroughness of structural sampling are performed (below) (Viswanath et al., 2017b).

Stage 4: Validating a structural model

To avoid overinterpretation of a model, it is essential that a model is validated. Formal model validation can follow five steps: (i) selecting the models for validation; (ii) estimating sampling precision; (iii) estimating model precision; (iv) quantifying the degree to which a model satisfies the information used to compute it; and (v) quantifying the degree to which a model satisfies relevant information not used to compute it. It is anticipated that the nascent worldwide PDB effort on integrative methods will refine and implement a set of specific procedures for these steps and apply them to every integrative structure during its deposition into the PDB (Vallat et al., 2018), as is already the case for traditional atomic structures (Henderson et al., 2012; Montelione et al., 2013; Read et al., 2011; Trewhella et al., 2013; Young et al., 2017).

In the first step, we select sufficiently good-scoring models produced by sampling for validation (the ensemble). For example, a good-scoring model may be defined as a sampled model that satisfies all restraints or sets of restraints within their own uncertainties (*eg*, we may require that the correlation coefficient between a model and the electron microscopy density map is at least 0.8 and that a model violates at most 4% of chemical cross-links). If no such models were produced, we need to reconsider various aspects of modeling: perhaps the input information was not as precise as presumed, representation of the system was not

sufficiently flexible (*eg*, too coarse-grained or too few states in a multi-state model), or sampling was insufficient.

In the second step, the sampling precision can be estimated for stochastic sampling methods by splitting the ensemble of models into two independent sets, followed by quantifying the difference between the two sets. The sampling precision can then be defined as this difference, in a similar fashion to that used for estimating the resolution of electron microscopy density maps (van Heel and Schatz, 2005). It is important that we properly estimate the sampling precision (uncertainty), because clearly only the features of the model larger than the sampling precision are well estimated (Viswanath et al., 2017b), just as traditional microscopy can only map features larger than the resolution of the microscope. When using stochastic sampling methods, sampling precision may often be increased simply by increasing the number of independently computed models. High sampling precision is necessary but not sufficient for exhaustive sampling (Gelman and Rubin, 1992; Viswanath et al., 2017b).

In the third step, model uncertainty (precision) is estimated. The most explicit description of model uncertainty is provided by the set of all models that are sufficiently consistent with the input information (*ie*, the ensemble). Model precision can be quantified by the variability among the models in the ensemble; in the end, the ensemble can be described by one or more representative models and their uncertainties. For example, good-scoring NPC models cluster in a single cluster with a root-mean-square fluctuation of ~9 Å, quantifying the average degree of uncertainty; importantly, the uncertainty is not distributed evenly across the ensemble, such that some regions were determined at a higher precision than 9 Å and others at a significantly lower precision; thus, even features larger than this estimate should be interpreted with some caution.

It is often convenient if the ensemble structures are clustered based on their structural similarity. As a result, only a structure representative of each major cluster can potentially be used for interpretation. Many clustering methods exist, varying in terms of the criterion used to quantify a similarity between two structures (*eg*, distance root-mean-square deviation between structure coordinates that avoids the need for structure superposition (Koehl, 2001)) as well as a method for converting such pairwise similarities into clusters (Hastie, 2001); the clustering generally also depends on an arbitrary threshold parameter that determines how many clusters are produced. Minor clusters with few structures might be ignored, especially if the scoring function approximates a Bayesian posterior model density (Viswanath et al., 2017a). In our own work (*eg*, (Kim et al., 2018; Viswanath et al., 2017b)), we often rely on a threshold-based clustering, where the threshold specifies the maximum distance between a cluster centroid and a model in the cluster (Daura et al., 1999); the clustering threshold is selected such that the number of major clusters is minimized for parsimony, while also requiring that their precision is worse than the sampling precision yet sufficient for interpreting the models.

The model uncertainty reflects both the actual heterogeneity of the physical sample and the uncertainties in the input information, representation, scoring for sampling, sampling itself, and scoring for filtering. It is difficult to deconvolve from each other the impact of these

different uncertainties on the model uncertainty; in general, only the total model uncertainty is reported. The uncertainty of how to represent the model in particular is often not considered, but can be large and have a major impact on the model uncertainty. For example, it is often possible to explain the small-angle X-ray scattering (SAXS) profile of a protein in solution within its uncertainty by a single or a small number of structures, even when the actual sample is disordered, because the large number of degrees of freedom in the model relatively easily result in a good fit to the data, given the relatively small amount of the data and its relatively large uncertainty (Carter et al., 2015). A mistake in representation is often not recoverable (*eg*, when a protein subunit structure is incorrectly assumed to be rigid during modeling a structure of a complex).

As an aside, the accuracy (error) of a model is defined as the deviation of the model from the truth; the accuracy is therefore not knowable when modeling systems without known structures (in benchmarking, reference answers are of course known by design). In contrast, model precision can be estimated. A conservative assumption is that accuracy is no better than model precision. If model error is larger than its estimated uncertainty, the model is incorrect; correspondingly, a model can be deemed correct if its error is within its uncertainty.

In the fourth step, the model is assessed by quantifying the degree to which it satisfies the information used to compute it. For example, the correlation coefficient between our recent NPC structure and the electron microscopy density map that helped compute the structure is higher than 0.92; less than 10% of chemical cross-links are violated; and less than 5% of bead overlaps are larger than the standard deviation of the harmonic excluded volume restraint (Kim et al., 2018).

The fifth step represents perhaps the most convincing test of model validity. In this step, a model is tested against relevant information that was *not* used to compute it. For example, one can perform a jackknifing test consisting of repetitively omitting a random subset of chemical crosslinks, recomputing the model, and comparing these models against the omitted cross-links, to validate both the model and the cross-links (Brunger et al., 1993). This test is even more powerful when some information is used only for validation. For example, we took advantage of omitting the vast majority of the 2007 NPC data from the most recent structure determination of the NPC, allowing us to use this data as well as the 2007 NPC topology map to validate the 2018 structure (Kim et al., 2018) (Figure 3A).

The integrative approach is unique in providing an especially strong test of model validity: when the structure is consistent with multiple types of data that were collected independently for physically different samples using different methods, the odds of artefacts are reduced compared to structures relying on a single type of information. An example of such an artefact are the false depictions of the NPC's organization that have arisen from reliance on biologically irrelevant intermolecular contacts in protein crystals (Debler et al., 2008; Hsia et al., 2007).

Lastly, even the input data themselves can be validated with respect to each other, *via* a structural model based on these data. A piece of data can be inconsistent with a model, and

thus with other data, when a mistake in model representation, scoring, or sampling is made. More precisely, data can be violated when the model is represented with too few degrees of freedom, data is compared against the model too stringently, or the sampling fails to find an existing good-scoring model. As an example of a representational error, a violation of a chemical cross-link may occur when a rigid subunit in a complex is not allowed to relax, or a single-state instead of multi-state representation is used (as is required when one or more physical samples from which the data are derived contain a mixture of structures). As an example of a scoring error, a violation may also be declared when the data are presumed to be less uncertain (noisy) than they actually are. There are no general protocols for deconvolving possible reasons for a given mismatch between data and a model, although the Bayesian approach shows most promise in this regard (Box 2). Remedies include modeling with alternative representations, scoring functions, and sampling schemes, which in turn often results in a more varied ensemble of good-scoring models and thus an increased estimate of model uncertainty. Most usefully, additional experiments may shed light on the homogeneity of the physical sample(s) and noise in the data. It may be appropriate to discard some data under the explicit assumption that the omitted data applies to non-interesting states or that it is too noisy, although tacitly omitting a subset of data only because a model does not fit it is perhaps one of the worst errors a scientist can make.

Validation is thus key to the iterative nature of the integrative structure determination process (Figure 2), such that the experimentalist and modeler synergize on data and model until consistency among all datasets and the final structure is obtained.

Biological insights from validated structures

The synergistic dialog between experimentalist and modeler continues as validation becomes interpretation, with a subjective but informed consideration of whether or not the structure makes sense in light of current knowledge. If not, the iterative process of information gathering and modeling must continue. After validation is satisfactory, we then move to biological interpretation of our structure.

Ultimately, the true worth of any structure is how informative it is about architectural principles of the modeled system, its assembly and disassembly, its function (*ie*, interactions with other systems), regulation and modulation of its function, and evolutionary relationships. Even though integrative structures are often determined at resolutions lower than atomic, they can still be informative, or at least more informative than the data on which they are based. Once again, each structure must be interpreted biologically only within its precision. For example, being preoccupied by nanometer-scale features of the 100 nm-diameter 2007 map would completely mistake its purpose primarily as a topological representation of the nucleoporin arrangements in the NPC.

An important tool in the interpretation of any structural model is its visualization and manipulation on a computer screen. However, most existing molecular viewers for atomic structures, such as those deposited in the PDB, are of limited utility, because integrative structures are often represented as ensembles of multi-scale models (with atomistic and coarse-grained representations), multi-state models (allowing for simultaneous multiple states), and ordered states (states related by time or other order) (Sali et al., 2015).

Fortunately, the ChimeraX visualization program (Goddard et al., 2018) was developed specifically to support integrative structural models stored in the common mmCIF file format upgraded for integrative models (Vallat et al., 2018). In addition to the models themselves, ChimeraX can also visualize a number of different datasets, such as density maps and chemical crosslinks, thus facilitating an assessment of how well the model fits the data.

Structures of the NPC and its subcomplexes from various organisms (Eibauer et al., 2015; Kim et al., 2018; Kosinski et al., 2016; Mosalaganti et al., 2018; von Appen et al., 2015) have led to a plethora of insights. For example, the overall architecture of the NPC as seen in our recent structure is evocative of the form and function of suspension bridges (Kim et al., 2018) (Figure 3C). In both bridge and NPC, this architecture results in a strong and resilient structure capable of resisting external forces and forces from the enormous transport flux through the pore. Moreover, both structures serve a similar purpose, namely to provide a selective conduit across a barrier. In the NPC, the "roadway" is constructed from anchors that arrange a high density of docking sites lined from cytoplasm to nucleoplasm for cargo-carrying transport factors to follow across the NPC's central channel (Figure 3C). Future directions will add information about the dynamic behavior of these docking sites and the transport factors, to animate the process of nuclear transport and elucidate its detailed mechanisms. Another insight we obtained is that the entire scaffold of the NPC is made of nucleoporins that share their architecture with those of the major scaffold components of vesicle coating complexes, indicating a common evolutionary origin in a primordial "protocoatomer" (Alber et al., 2007b; Devos et al., 2004; Spang et al., 2017). Such coating complexes currently fall into two structurally distinct families (Dacks and Robinson, 2017; Faini et al., 2013). The fact that the NPC is comprised of representatives of both families suggests that these families evolved first, together with an already differentiated internal membrane system. Intriguingly, this pattern in turn implies that the NPC - and the nucleus as a whole as we know it – may have been among the last organelles to evolve on the path of eukaryotic cellular evolution, rather than being among the first as had been previously assumed (Kim et al., 2018).

The NPC is but one of a large number of structures solved by integrative methods that have been biologically highly informative (Figure 1). Highlights include: the complete structure of the mammalian mitochondrial ribosome large subunit, revealing how the 5S ribosomal RNA has become substituted by a tRNA, and showing how insights into unusual aspects of architectural reorganization can be garnered (Greber et al., 2014a); the complete structure of the 26S proteasome, showing how the lid structure is critical for recruiting and partially unfolding the substrate protein for subsequent proteolysis by the 20S core particle, thus showcasing how functional and catalytic insights can be achieved (Lasker et al., 2012); and visualizing how chromosomes are dynamically positioned in the nucleus and revealing the plasticity of genome structures, showing how integrative methods can be applied at different cellular spatial and temporal scales (Kalhor et al., 2011).

The integrative approach is not restricted to a particular granularity or size of the model. Indeed, the integrative structural solution of two smaller subcomplexes from the NPC, Nup82 (Fernandez-Martinez et al., 2016) and Nup84 (Shi et al., 2014), both around 650 kDa

in size, preceded our most recent solution of the entire NPC structure. Similarly, numerous moderately-sized protein structures have also been solved by integration of orthogonal structural data (*cf*, Figure 1). For example, supplementing data from NMR spectroscopy by additional data from SAXS experiments can produce larger and more accurate protein structure models than NMR spectroscopy on its own (Sunnerhagen et al., 1996). Thus, a flexible-domain structure refinement with both NMR and SAXS data allowed the solution of a structure for the 82 kDa Malate Synthase G enzyme (Grishaev et al., 2008). More recently, a combination of data from NMR spectroscopy, SAXS, and small-angle neutron scattering (SANS) was used to determine the structure for a 34 kDa ternary SXL, UNR, and msl-2 mRNA complex (Hennig et al., 2013). Another elegant method uses augmented NMR NOESY-based restraints, which are often insufficient to calculate an accurate model, with evolutionary residue-residue couplings computed from multiple alignments of related protein sequences (Tang et al., 2015). The largest protein solved in this manner was the 41 kDa *E. coli* maltose-binding protein, and the method is applicable to even larger systems (Huang et al., 2019).

Outlook

There is much still to be done to improve all aspects of computing, validating, visualizing, archiving, and disseminating integrative structures (Table 2). This task includes automating as much of the modeling process as possible. It would be particularly helpful to develop better methods for objectively finding optimal representations (Viswanath and Sali, 2019; Wagner et al., 2016), given the available input information, including methods for finding the number of different states in multi-state models and optimal coarse-graining. It is also necessary to formulate all conceivable types of structural information in terms of Bayesian data likelihoods, which will facilitate proper relative consideration of varied information during modeling. Modeling will further benefit from improving the efficiency of sampling methods and computing hardware, resulting in a more thorough search for good-scoring models, especially for large systems with many degrees of freedom. Most importantly, a rigorous and extensive validation pipeline for estimating the uncertainty in integrative structures is essential for their proper interpretation. Finally, the field will benefit from a community-wide set of standards for various aspects of integrative modeling, underpinned by an archive for integrative structures as well as the data on which they are based and the modeling protocols, as spearheaded by the nascent PDB-Development community resource (Burley et al., 2017; Vallat et al., 2018). PDB-Development will further strengthen integrative structural biology, by bringing together specialists in disparate experimental methodologies (Table 1), unified by their intent to iteratively and formally combine their data into as informative models of biomolecular systems as possible.

Improving various aspects of integrative modeling, as outlined above, will further expand its applications. It will become possible to obtain useful models of the larger systems, heterogeneous systems, and dynamic processes that actually typify the organization of cells. A particular strength of integrative modeling is its potential to use all information to compute models represented in any fashion, be it single static structures, mixtures of states, molecular networks, dynamic processes, systems of ordinary differential equations, and others. Indeed, the explicit inclusion of dynamic and state-dependent information into

integrative approaches holds the promise of breathing life and movement into currently mostly static representations, and so visualize the processes that actually drive living cellular systems. As a result, it is conceivable that integrative modeling will play a key role in mapping entire cellular neighborhoods and even whole cells, thus bridging the gap between biophysical methods focused on molecules and optical microscopies focused on the meso-scale organization of the cell.

As the toolbox of integrative structural biology continues to improve, it will be increasingly applied not only to discover the basic principles of biological systems, but also to drug discovery. As a result, it will allow us to rationally target larger systems, in addition to single proteins. While still largely untapped, the potential for using integrative approaches to translate from bench to bedside is surely among the most exciting new future directions open to the biomedical community (Singla et al., 2018).

Acknowledgments

References

Akey CW, and Radermacher M. (1993). Architecture of the Xenopus nuclear pore complex revealed by three-dimensional cryo-electron microscopy. The Journal of cell biology 122, 1–19. [PubMed: 8314837]

Alber F, Dokudovskaya S, Veenhoff L, Zhang W, Kipper J, Devos D, Suprapto A, Karni-Schmidt O, Williams R, Chait B, et al. (2007a). Determining the architectures of macromolecular assemblies. Nature 450, 683–694. [PubMed: 18046405]

Alber F, Dokudovskaya S, Veenhoff LM, Zhang W, Kipper J, Devos D, Suprapto A, Karni-Schmidt O, Williams R, Chait BT, et al. (2007b). The molecular architecture of the nuclear pore complex. Nature 450, 695–701. [PubMed: 18046406]

Alber F, Forster F, Korkin D, Topf M, and Sali A. (2008). Integrating diverse data for structure determination of macromolecular assemblies. Annu Rev Biochem 77, 443–477. [PubMed: 18318657]

Allen MP, and Tildesley DJ (1989). Computer simulation of liquids (Clarendon Press).

Baade I, and Kehlenbach RH (2018). The cargo spectrum of nuclear transport receptors. Current opinion in cell biology 58, 1–7. [PubMed: 30530239]

Bau D, Sanyal A, Lajoie BR, Capriotti E, Byron M, Lawrence JB, Dekker J, and Marti-Renom MA (2011). The three-dimensional folding of the alpha-globin gene domain reveals formation of chromatin globules. Nature structural & molecular biology 18, 107–114.

Beck M, Forster F, Ecke M, Plitzko JM, Melchior F, Gerisch G, Baumeister W, and Medalia O. (2004). Nuclear pore complex structure and dynamics revealed by cryoelectron tomography. Science 306, 1387–1390. [PubMed: 15514115]

Beck M, and Hurt E. (2017). The nuclear pore complex: understanding its function through structural insight. Nature reviews Molecular cell biology 18, 73–89. [PubMed: 27999437]

Betancourt M. (2017). A Conceptual Introduction to Hamiltonian Monte Carlo. arXiv:170102434v2 [statME].

Bock LV, Blau C, Schroder GF, Davydov II, Fischer N, Stark H, Rodnina MV, Vaiana AC, and Grubmuller H. (2013). Energy barriers and driving forces in tRNA translocation through the ribosome. Nature structural & molecular biology 20, 1390–1396.

Bonomi M, and Camilloni C. (2017). Integrative structural and dynamical biology with PLUMED-ISDB. Bioinformatics 33, 3999–4000. [PubMed: 28961689]

Boura E, Rozycki B, Herrick DZ, Chung HS, Vecer J, Eaton WA, Cafiso DS, Hummer G, and Hurley JH (2011). Solution structure of the ESCRT-I complex by small-angle X-ray scattering, EPR, and FRET spectroscopy. Proc Natl Acad Sci U S A 108, 9437–9442. [PubMed: 21596998]

Brohawn SG, Partridge JR, Whittle JR, and Schwartz TU (2009). The nuclear pore complex has entered the atomic age. Structure 17, 1156–1168. [PubMed: 19748337]

Brunger AT, Clore GM, Gronenborn AM, Saffrich R, and Nilges M. (1993). Assessing the quality of solution nuclear magnetic resonance structures by complete cross-validation. Science 261, 328–331. [PubMed: 8332897]

Bullock JMA, Sen N, Thalassinos K, and Topf M. (2018a). Modeling Protein Complexes Using Restraints from Crosslinking Mass Spectrometry. Structure.

Bullock JMA, Thalassinos K, and Topf M. (2018b). Jwalk and MNXL web server: model validation using restraints from crosslinking mass spectrometry. Bioinformatics 34, 3584–3585. [PubMed: 29741581]

Burley SK, Kurisu G, Markley JL, Nakamura H, Velankar S, Berman HM, Sali A, Schwede T, and Trewhella J. (2017). PDB-Dev: a Prototype System for Depositing Integrative/Hybrid Structural Models. Structure 25, 1317–1318. [PubMed: 28877501]

Callaway E. (2015). The revolution will not be crystallized: a new method sweeps through structural biology. Nature 525, 172–174. [PubMed: 26354465]

Carter L, Kim SJ, Schneidman-Duhovny D, Stohr J, Poncet-Montange G, Weiss TM, Tsuruta H, Prusiner SB, and Sali A. (2015). Prion Protein-Antibody Complexes Characterized by Chromatography-Coupled Small-Angle X-Ray Scattering. Biophysical journal 109, 793–805. [PubMed: 26287631]

Chen ZA, Jawhari A, Fischer L, Buchen C, Tahir S, Kamenski T, Rasmussen M, Lariviere L, Bukowski-Wills JC, Nilges M, et al. (2010). Architecture of the RNA polymerase II-TFIIF complex revealed by cross-linking and mass spectrometry. Embo J 29, 717–726. [PubMed: 20094031]

Ciferri C, Lander GC, Maiolica A, Herzog F, Aebersold R, and Nogales E. (2012). Molecular architecture of human polycomb repressive complex 2. eLife 1, e00005.

Dacks JB, and Robinson MS (2017). Outerwear through the ages: evolutionary cell biology of vesicle coats. Current opinion in cell biology 47, 108–116. [PubMed: 28622586]

Danev R, and Baumeister W. (2017). Expanding the boundaries of cryo-EM with phase plates. Current opinion in structural biology 46, 87–94. [PubMed: 28675816]

Das R, and Baker D. (2008). Macromolecular modeling with rosetta. Annu Rev Biochem 77, 363–382. [PubMed: 18410248]

Daura X, Gademann K, Jaun B, Seebach D, Van Gunsteren WF, and Mark AE (1999). Peptide folding: when simulation meets experiment. Angewandte Chemie International Edition 38, 236–240.

De Magistris P, and Antonin W. (2018). The Dynamic Nature of the Nuclear Envelope. Current biology : CB 28, R487–r497. [PubMed: 29689232]

de Vries SJ, and Zacharias M. (2012). ATTRACT-EM: a new method for the computational assembly of large molecular machines using cryo-EM maps. PloS one 7, e49733.

Debler EW, Ma Y, Seo HS, Hsia KC, Noriega TR, Blobel G, and Hoelz A. (2008). A fence-like coat for the nuclear pore membrane. Mol Cell 32, 815–826. [PubMed: 19111661]

Degiacomi MT, and Dal Peraro M. (2013). Macromolecular symmetric assembly prediction using swarm intelligence dynamic modeling. Structure 21, 1097–1106. [PubMed: 23810695]

Degiacomi MT, Iacovache I, Pernot L, Chami M, Kudryashev M, Stahlberg H, van der Goot FG, and Dal Peraro M. (2013). Molecular assembly of the aerolysin pore reveals a swirling membrane-insertion mechanism. Nat Chem Biol 9, 623–629. [PubMed: 23912165]

Deshmukh L, Schwieters CD, Grishaev A, Ghirlando R, Baber JL, and Clore GM (2013). Structure and dynamics of full-length HIV-1 capsid protein in solution. J Am Chem Soc 135, 16133–16147. [PubMed: 24066695]

Devos D, Dokudovskaya S, Alber F, Williams R, Chait BT, Sali A, and Rout MP (2004). Components of coated vesicles and nuclear pore complexes share a common molecular architecture. PLoS Biol 2, e380.

Diez M, Zimmermann B, Borsch M, Konig M, Schweinberger E, Steigmiller S, Reuter R, Felekyan S, Kudryavtsev V, Seidel CA, et al. (2004). Proton-powered subunit rotation in single membrane-bound F0F1-ATP synthase. Nature structural & molecular biology 11, 135–141.

Dobzhansky T. (1973). Nothing in Biology Makes Sense except in the Light of Evolution. The American Biology Teacher 35, 125–129.

Dominguez C, Boelens R, and Bonvin AM (2003). HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. J Am Chem Soc 125, 1731–1737. [PubMed: 12580598]

Eibauer M, Pellanda M, Turgay Y, Dubrovsky A, Wild A, and Medalia O. (2015). Structure and gating of the nuclear pore complex. Nature communications 6, 7532.

Erzberger JP, Stengel F, Pellarin R, Zhang S, Schaefer T, Aylett CHS, Cimermancic P, Boehringer D, Sali A, Aebersold R, et al. (2014). Molecular architecture of the 40SeIF1eIF3 translation initiation complex. Cell 158, 1123–1135. [PubMed: 25171412]

Eschweiler JD, Farrugia MA, Dixit SM, Hausinger RP, and Ruotolo BT (2018). A Structural Model of the Urease Activation Complex Derived from Ion Mobility-Mass Spectrometry and Integrative Modeling. Structure 26, 599–606 e593. [PubMed: 29576318]

Faini M, Beck R, Wieland FT, and Briggs JA (2013). Vesicle coats: structure, function, and general principles of assembly. Trends in cell biology 23, 279–288. [PubMed: 23414967]

Feng Y, Zhang Y, and Ebright RH (2016). Structural basis of transcription activation. Science 352, 1330–1333. [PubMed: 27284196]

Fernandez-Martinez J, Kim SJ, Shi Y, Upla P, Pellarin R, Gagnon M, Chemmama IE, Wang J, Nudelman I, Zhang W, et al. (2016). Structure and Function of the Nuclear Pore Complex Cytoplasmic mRNA Export Platform. Cell 167, 1215–1228 e1225. [PubMed: 27839866]

Fernandez-Martinez J, Phillips J, Sekedat MD, Diaz-Avalos R, Velazquez-Muriel J, Franke JD, Williams R, Stokes DL, Chait BT, Sali A, et al. (2012). Structure-function mapping of a heptameric module in the nuclear pore complex. The Journal of cell biology 196, 419–434. [PubMed: 22331846]

Fischer J, Teimer R, Amlacher S, Kunze R, and Hurt E. (2015). Linker Nups connect the nuclear pore complex inner ring with the outer ring and transport channel. Nature structural & molecular biology 22, 774–781.

Fischer L, Chen ZA, and Rappsilber J. (2013). Quantitative cross-linking/mass spectrometry using isotope-labelled cross-linkers. J Proteomics.

Franke D, Petoukhov MV, Konarev PV, Panjkovich A, Tuukkanen A, Mertens HDT, Kikhney AG, Hajizadeh NR, Franklin JM, Jeffries CM, et al. (2017). ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. J Appl Crystallogr 50, 1212–1225. [PubMed: 28808438]

Franklin RE, and Gosling RG (1953). Molecular configuration in sodium thymonucleate. Nature 171, 740–741. [PubMed: 13054694]

Gay P. (2002). Schnitzler's century: the making of middle-class culture, 1815–1914 (WW Norton & Company).

Gelman A, and Rubin DB (1992). Inference from Iterative Simulation Using Multiple Sequences. Statist Sci 7, 457–472.

Gingras AC, Gstaiger M, Raught B, and Aebersold R. (2007). Analysis of protein complexes using mass spectrometry. Nature reviews Molecular cell biology 8, 645–654. [PubMed: 17593931]

Goddard TD, Huang CC, Meng EC, Pettersen EF, Couch GS, Morris JH, and Ferrin TE (2018). UCSF ChimeraX: Meeting modern challenges in visualization and analysis. Protein science : a publication of the Protein Society 27, 14–25. [PubMed: 28710774]

Gong Z, Schwieters CD, and Tang C. (2015). Conjoined use of EM and NMR in RNA structure refinement. PloS one 10, e0120445.

Grandi P, Doye V, and Hurt EC (1993). Purification of NSP1 reveals complex formation with 'GLFG' nucleoporins and a novel nuclear pore protein NIC96. Embo J 12, 3061–3071. [PubMed: 7688296]

Grandi P, Emig S, Weise C, Hucho F, Pohl T, and Hurt EC (1995a). A novel nuclear pore protein Nup82p which specifically binds to a fraction of Nsp1p. The Journal of cell biology 130, 1263–1273. [PubMed: 7559750]

Grandi P, Schlaich N, Tekotte H, and Hurt EC (1995b). Functional interaction of Nic96p with a core nucleoporin complex consisting of Nsp1p, Nup49p and a novel protein Nup57p. Embo J 14, 76–87. [PubMed: 7828598]

Greber BJ, Boehringer D, Leibundgut M, Bieri P, Leitner A, Schmitz N, Aebersold R, and Ban N. (2014a). The complete structure of the large subunit of the mammalian mitochondrial ribosome. Nature 515, 283–286. [PubMed: 25271403]

Greber BJ, Boehringer D, Leitner A, Bieri P, Voigts-Hoffmann F, Erzberger JP, Leibundgut M, Aebersold R, and Ban N. (2014b). Architecture of the large subunit of the mammalian mitochondrial ribosome. Nature 505, 515–519. [PubMed: 24362565]

Grime JM, and Voth GA (2014). Highly Scalable and Memory Efficient Ultra-Coarse-Grained Molecular Dynamics Simulations. J Chem Theory Comput 10, 423–431. [PubMed: 26579921]

Grishaev A, Tugarinov V, Kay LE, Trewhella J, and Bax A. (2008). Refined solution structure of the 82-kDa enzyme malate synthase G from joint NMR and synchrotron SAXS restraints. J Biomol NMR 40, 95–106. [PubMed: 18008171]

Haas J, Roth S, Arnold K, Kiefer F, Schmidt T, Bordoli L, and Schwede T. (2013). The Protein Model Portal-- a comprehensive resource for protein structure and model information. Database (Oxford) 2013, bat031.

Han Y, Luo J, Ranish J, and Hahn S. (2014). Architecture of the Saccharomyces cerevisiae SAGA transcription coactivator complex. Embo J 33, 2534–2546. [PubMed: 25216679]

Hao Q, Zhang B, Yuan K, Shi H, and Blobel G. (2018). Electron microscopy of Chaetomium pom152 shows the assembly of ten-bead string. Cell discovery 4, 56. [PubMed: 30245846]

Harrer N, Schindler CEM, Bruetzel LK, Forne I, Ludwigsen J, Imhof A, Zacharias M, Lipfert J, and Mueller-Planitz F. (2018). Structural Architecture of the Nucleosome Remodeler ISWI Determined from Cross-Linking, Mass Spectrometry, SAXS, and Modeling. Structure 26, 282–294 e286. [PubMed: 29395785]

Hastie T, Tibshirani R, Friedman J. (2001). The Elements of Statistical Learning, Vol 1 (New York: Springer).

Henderson R, Sali A, Baker ML, Carragher B, Devkota B, Downing KH, Egelman EH, Feng Z, Frank J, Grigorieff N, et al. (2012). Outcome of the first electron microscopy validation task force meeting. Structure 20, 205–214. [PubMed: 22325770]

Hennig J, Wang I, Sonntag M, Gabel F, and Sattler M. (2013). Combining NMR and small angle X-ray and neutron scattering in the structural analysis of a ternary protein-RNA complex. J Biomol NMR 56, 17–30. [PubMed: 23456097]

Hinshaw JE, Carragher BO, and Milligan RA (1992). Architecture and design of the nuclear pore complex. Cell 69, 1133–1141. [PubMed: 1617726]

Hsia KC, Stavropoulos P, Blobel G, and Hoelz A. (2007). Architecture of a coat for the nuclear pore membrane. Cell 131, 1313–1326. [PubMed: 18160040]

Hsieh A, Lu L, Chance MR, and Yang S. (2017). A Practical Guide to iSPOT Modeling: An Integrative Structural Biology Platform. Advances in experimental medicine and biology 1009, 229–238. [PubMed: 29218563]

Hua N, Tjong H, Shin H, Gong K, Zhou XJ, and Alber F. (2018). Producing genome structure populations with the dynamic and automated PGS software. Nature protocols 13, 915926.

Huang YJ, Brock KP, Ishida Y, Swapna GVT, Inouye M, Marks DS, Sander C, and Montelione GT (2019). Combining Evolutionary Covariance and NMR Data for Protein Structure Determination. Methods Enzymol 614, 363–392. [PubMed: 30611430]

Humphrey W, Dalke A, and Schulten K. (1996). VMD: visual molecular dynamics. J Mol Graph 14, 33–38, 27–38. [PubMed: 8744570]

Joseph AP, Lagerstedt I, Patwardhan A, Topf M, and Winn M. (2017). Improved metrics for comparing structures of macromolecular assemblies determined by 3D electron-microscopy. Journal of structural biology 199, 12–26. [PubMed: 28552721]

Kalhor R, Tjong H, Jayathilaka N, Alber F, and Chen L. (2011). Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. Nat Biotechnol 30, 90–98. [PubMed: 22198700]

Kalinin S, Peulen T, Sindbert S, Rothwell PJ, Berger S, Restle T, Goody RS, Gohlke H, and Seidel CA (2012). A toolkit and benchmark study for FRET-restrained high-precision structural modeling. Nat Methods 9, 1218–1225. [PubMed: 23142871]

Kim SJ, Fernandez-Martinez J, Nudelman I, Shi Y, Zhang W, Raveh B, Herricks T, Slaughter BD, Hogan JA, Upla P, et al. (2018). Integrative structure and functional anatomy of a nuclear pore complex. Nature 555, 475–482. [PubMed: 29539637]

Kim SJ, Fernandez-Martinez J, Sampathkumar P, Martel A, Matsui T, Tsuruta H, Weiss TM, Shi Y, Markina-Inarrairaegui A, Bonanno JB, et al. (2014). Integrative structure-function mapping of the nucleoporin Nup133 suggests a conserved mechanism for membrane anchoring of the nuclear pore complex. Molecular & cellular proteomics : MCP 13, 2911–2926. [PubMed: 25139911]

Knockenhauer KE, and Schwartz TU (2016). The Nuclear Pore Complex as a Flexible and Dynamic Gate. Cell 164, 1162–1171. [PubMed: 26967283]

Knuth KH, Habeck M, Malakar NK, Mubeen AM, and Placek B. (2015). Bayesian evidence and model selection. Digit Signal Process 47, 50–67.

Koehl P. (2001). Protein structure similarities. Current opinion in structural biology 11, 348–353. [PubMed: 11406386]

Kosinski J, Mosalaganti S, von Appen A, Teimer R, DiGuilio AL, Wan W, Bui KH, Hagen WJ, Briggs JA, Glavy JS, et al. (2016). Molecular architecture of the inner ring scaffold of the human nuclear pore complex. Science 352, 363–365. [PubMed: 27081072]

Lasker K, Forster F, Bohn S, Walzthoeni T, Villa E, Unverdorben P, Beck F, Aebersold R, Sali A, and Baumeister W. (2012). Molecular architecture of the 26S proteasome holocomplex determined by an integrative approach. Proc Natl Acad Sci U S A 109, 1380–1387. [PubMed: 22307589]

Lasker K, Sali A, and Wolfson HJ (2010). Determining macromolecular assembly structures by molecular docking and fitting into an electron density map. Proteins 78, 3205–3211. [PubMed: 20827723]

Lasker K, Topf M, Sali A, and Wolfson HJ (2009). Inferential optimization for simultaneous fitting of multiple components into a CryoEM map of their assembly. Journal of molecular biology 388, 180–194. [PubMed: 19233204]

Lauber MA, Rappsilber J, and Reilly JP (2012). Dynamics of ribosomal protein S1 on a bacterial ribosome with cross-linking and mass spectrometry. Molecular & cellular proteomics : MCP 11, 1965–1976. [PubMed: 23033476]

Leitner A, Joachimiak LA, Bracher A, Monkemeyer L, Walzthoeni T, Chen B, Pechmann S, Holmes S, Cong Y, Ma B, et al. (2012a). The molecular architecture of the eukaryotic chaperonin TRiC/CCT. Structure 20, 814–825. [PubMed: 22503819]

Leitner A, Reischl R, Walzthoeni T, Herzog F, Bohn S, Forster F, and Aebersold R. (2012b). Expanding the chemical cross-linking toolbox by the use of multiple proteases and enrichment by size exclusion chromatography. Molecular & cellular proteomics : MCP 11, M111 014126.

Leitner A, Walzthoeni T, Kahraman A, Herzog F, Rinner O, Beck M, and Aebersold R. (2010). Probing native protein structures by chemical cross-linking, mass spectrometry, and bioinformatics. Molecular & cellular proteomics : MCP 9, 1634–1649. [PubMed: 20360032]

Leone V, and Faraldo-Gomez JD (2016). Structure and mechanism of the ATP synthase membrane motor inferred from quantitative integrative modeling. J Gen Physiol 148, 441–457. [PubMed: 27821609]

Li Q, Tjong H, Li X, Gong K, Zhou XJ, Chiolo I, and Alber F. (2017). The threedimensional genome organization of Drosophila melanogaster through data integration. Genome Biol 18, 145. [PubMed: 28760140]

Lin DH, Stuwe T, Schilbach S, Rundlet EJ, Perriches T, Mobbs G, Fan Y, Thierbach K, Huber FM, Collins LN, et al. (2016). Architecture of the symmetric core of the nuclear pore. Science 352, aaf1015.

Loquet A, Sgourakis NG, Gupta R, Giller K, Riedel D, Goosmann C, Griesinger C, Kolbe M, Baker D, Becker S, et al. (2012). Atomic model of the type III secretion system needle. Nature 486, 276–279. [PubMed: 22699623]

Lukoyanova N, Kondos SC, Farabella I, Law RH, Reboul CF, Caradoc-Davies TT, Spicer BA, Kleifeld O, Traore DA, Ekkel SM, et al. (2015). Conformational changes during pore formation by the perforin-related protein pleurotolysin. PLoS Biol 13, e1002049.

Luo J, Cimermancic P, Viswanath S, Ebmeier CC, Kim B, Dehecq M, Raman V, Greenberg CH, Pellarin R, Sali A, et al. (2015). Architecture of the Human and Yeast General Transcription and DNA Repair Factor TFIIH. Mol Cell 59, 794–806. [PubMed: 26340423]

Metropolis NR, A. W; Rosenbluth MN; Teller AH (1953). Equation of State Calculations by Fast Computing Machines. J Chem Phys 21, 1087–1092.

Miyazaki Y, Irobalieva RN, Tolbert BS, Smalls-Mantey A, Iyalla K, Loeliger K, D'Souza V, Khant H, Schmid MF, Garcia EL, et al. (2010). Structure of a conserved retroviral RNA packaging element by NMR spectroscopy and cryo-electron tomography. Journal of molecular biology 404, 751–772. [PubMed: 20933521]

Molnar KS, Bonomi M, Pellarin R, Clinthorne GD, Gonzalez G, Goldberg SD, Goulian M, Sali A, and DeGrado WF (2014). Cys-scanning disulfide crosslinking and bayesian modeling probe the transmembrane signaling mechanism of the histidine kinase, PhoQ. Structure 22, 1239–1251. [PubMed: 25087511]

Montelione GT, Nilges M, Bax A, Guntert P, Herrmann T, Richardson JS, Schwieters CD, Vranken WF, Vuister GW, Wishart DS, et al. (2013). Recommendations of the wwPDB NMR Validation Task Force. Structure 21, 1563–1570. [PubMed: 24010715]

Mosalaganti S, Kosinski J, Albert S, Schaffer M, Strenkert D, Salome PA, Merchant SS, Plitzko JM, Baumeister W, Engel BD, et al. (2018). In situ architecture of the algal nuclear pore complex. Nature communications 9, 2361.

Murakami K, Elmlund H, Kalisman N, Bushnell DA, Adams CM, Azubel M, Elmlund D, Levi-Kalisman Y, Liu X, Gibbons BJ, et al. (2013). Architecture of an RNA polymerase II transcription pre-initiation complex. Science 342, 1238724.

Murata K, and Wolf M. (2018). Cryo-electron microscopy for structural analysis of dynamic biological macromolecules. Biochimica et biophysica acta 1862, 324–334. [PubMed: 28756276]

Nir G, Farabella I, Perez Estrada C, Ebeling CG, Beliveau BJ, Sasaki HM, Lee SD, Nguyen SC, McCole RB, Chattoraj S, et al. (2018). Walking along chromosomes with superresolution imaging, contact maps, and integrative modeling. PLoS genetics 14, e1007872.

Oldfield CJ, and Dunker AK (2014). Intrinsically disordered proteins and intrinsically disordered protein regions. Annu Rev Biochem 83, 553–584. [PubMed: 24606139]

Ori A, Banterle N, Iskar M, Andres-Pons A, Escher C, Khanh Bui H, Sparks L, Solis-Mezarino V, Rinner O, Bork P, et al. (2013). Cell type-specific nuclear pores: a case in point for context-dependent stoichiometry of molecular machines. Molecular systems biology 9, 648. [PubMed: 23511206]

Pelikan M, Hura GL, and Hammel M. (2009). Structure and flexibility within proteins as identified through small angle X-ray scattering. Gen Physiol Biophys 28, 174–189.

Pirchi M, Ziv G, Riven I, Cohen SS, Zohar N, Barak Y, and Haran G. (2011). Singlemolecule fluorescence spectroscopy maps the folding landscape of a large protein. Nature communications 2, 493.

Politis A, Stengel F, Hall Z, Hernandez H, Leitner A, Walzthoeni T, Robinson CV, and Aebersold R. (2014). A mass spectrometry-based hybrid method for structural modeling of protein complexes. Nat Methods 11, 403–406. [PubMed: 24509631]

Press WH, Teukolsky SA, Vetterling WT, and Flannery BP (2007). Numerical Recipes 3rd Edition: The Art of Scientific Computing (Cambridge University Press).

Prischi F, Konarev PV, Iannuzzi C, Pastore C, Adinolfi S, Martin SR, Svergun DI, and Pastore A. (2010). Structural bases for the interaction of frataxin with the central components of iron-sulphur cluster assembly. Nature communications 1, 95.

Raices M, and D'Angelo MA (2012). Nuclear pore complex composition: a new regulator of tissue-specific and developmental functions. Nature reviews Molecular cell biology 13, 687–699. [PubMed: 23090414]

Rappsilber J. (2012). Cross-linking/mass spectrometry as a new field and the proteomics information mountain of tomorrow. Expert Rev Proteomics 9, 485–487. [PubMed: 23194264]

Ratje AH, Loerke J, Mikolajka A, Brunner M, Hildebrand PW, Starosta AL, Donhofer A, Connell SR, Fucini P, Mielke T, et al. (2010). Head swivel on the ribosome facilitates translocation by means of intra-subunit tRNA hybrid sites. Nature 468, 713–716. [PubMed: 21124459]

Read RJ, Adams PD, Arendall WB 3rd, Brunger AT, Emsley P, Joosten RP, Kleywegt GJ, Krissinel EB, Lutteke T, Otwinowski Z, et al. (2011). A new generation of crystallographic validation tools for the protein data bank. Structure 19, 1395–1412. [PubMed: 22000512]

Rieping W, Habeck M, and Nilges M. (2005). Inferential structure determination. Science 309, 303–306. [PubMed: 16002620]

Russel D, Lasker K, Webb B, Velazquez-Muriel J, Tjioe E, Schneidman-Duhovny D, Peterson B, and Sali A. (2012). Putting the pieces together: integrative modeling platform software for structure determination of macromolecular assemblies. PLoS Biol 10, e1001244.

Sailer C, Offensperger F, Julier A, Kammer KM, Walker-Gray R, Gold MG, Scheffner M, and Stengel F. (2018). Structural dynamics of the E6AP/UBE3A-E6-p53 enzyme-substrate complex. Nature communications 9, 4441.

Sali A, Berman HM, Schwede T, Trewhella J, Kleywegt G, Burley SK, Markley J, Nakamura H, Adams P, Bonvin AM, et al. (2015). Outcome of the First wwPDB Hybrid/Integrative Methods Task Force Workshop. Structure 23, 1156–1167. [PubMed: 26095030]

Saunders MG, and Voth GA (2013). Coarse-graining methods for computational biology. Annu Rev Biophys 42, 73–93. [PubMed: 23451897]

Schneidman-Duhovny D, Inbar Y, Nussinov R, and Wolfson HJ (2005). PatchDock and SymmDock: servers for rigid and symmetric docking. Nucleic acids research 33, W363–367. [PubMed: 15980490]

Schneidman-Duhovny D, Pellarin R, and Sali A. (2014). Uncertainty in integrative structural modeling. Current opinion in structural biology 28, 96–104. [PubMed: 25173450]

Schroder GF (2015). Hybrid methods for macromolecular structure determination: experiment with expectations. Current opinion in structural biology 31, 20–27. [PubMed: 25795086]

Schwieters CD, Bermejo GA, and Clore GM (2018). Xplor-NIH for molecular structure determination from NMR and other data sources. Protein science : a publication of the Protein Society 27, 26–40. [PubMed: 28766807]

Serra F, Bau D, Goodstadt M, Castillo D, Filion GJ, and Marti-Renom MA (2017). Automatic analysis and 3D-modelling of Hi-C data using TADbit reveals structural features of the fly chromatin colors. PLoS Comput Biol 13, e1005665.

Sharma R, Raduly Z, Miskei M, and Fuxreiter M. (2015). Fuzzy complexes: Specific binding without complete folding. FEBS Lett 589, 2533–2542. [PubMed: 26226339]

Sharma S, Ding F, and Dokholyan NV (2008). iFoldRNA: three-dimensional RNA structure prediction and folding. Bioinformatics 24, 1951–1952. [PubMed: 18579566]

Shi Y, Fernandez-Martinez J, Tjioe E, Pellarin R, Kim SJ, Williams R, Schneidman-Duhovny D, Sali A, Rout MP, and Chait BT (2014). Structural characterization by crosslinking reveals the detailed architecture of a coatomer-related heptameric module from the nuclear pore complex. Molecular & cellular proteomics : MCP 13, 2927–2943. [PubMed: 25161197]

Singla J, McClary KM, White KL, Alber F, Sali A, and Stevens RC (2018). Opportunities and Challenges in Building a Spatiotemporal Multi-scale Model of the Human Pancreatic beta Cell. Cell 173, 11–19. [PubMed: 29570991]

Siniossoglou S, Wimmer C, Rieger M, Doye V, Tekotte H, Weise C, Emig S, Segref A, and Hurt EC (1996). A novel complex of nucleoporins, which includes Sec13p and a Sec13p homolog, is essential for normal nuclear pores. Cell 84, 265–275. [PubMed: 8565072]

Sinz A. (2006). Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein-protein interactions. Mass spectrometry reviews 25, 663–682. [PubMed: 16477643]

Snijder J, Burnley RJ, Wiegard A, Melquiond AS, Bonvin AM, Axmann IM, and Heck AJ (2014). Insight into cyanobacterial circadian timing from structural details of the KaiB-KaiC interaction. Proc Natl Acad Sci U S A 111, 1379–1384. [PubMed: 24474762]

Spang A, Caceres EF, and Ettema TJG (2017). Genomic exploration of the diversity, ecology, and evolution of the archaeal domain of life. Science 357.

Sunnerhagen M, Olah GA, Stenflo J, Forsen S, Drakenberg T, and Trewhella J. (1996). The relative orientation of Gla and EGF domains in coagulation factor X is altered by Ca2+ binding to the first EGF domain. A combined NMR-small angle X-ray scattering study. Biochemistry 35, 11547–11559. [PubMed: 8794734]

Tang Y, Huang YJ, Hopf TA, Sander C, Marks DS, and Montelione GT (2015). Protein structure determination by combining sparse NMR data with evolutionary couplings. Nat Methods 12, 751–754. [PubMed: 26121406]

Tosi A, Haas C, Herzog F, Gilmozzi A, Berninghausen O, Ungewickell C, Gerhold CB, Lakomek K, Aebersold R, Beckmann R, et al. (2013). Structure and subunit topology of the INO80 chromatin remodeler and its nucleosome complex. Cell 154, 1207–1219. [PubMed: 24034245]

Trabuco LG, Villa E, Mitra K, Frank J, and Schulten K. (2008). Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. Structure 16, 673–683. [PubMed: 18462672]

Trewhella J, Hendrickson WA, Kleywegt GJ, Sali A, Sato M, Schwede T, Svergun DI, Tainer JA, Westbrook J, and Berman HM (2013). Report of the wwPDB Small-Angle Scattering Task Force: data requirements for biomolecular modeling and the PDB. Structure 21, 875–881. [PubMed: 23747111]

Upla P, Kim SJ, Sampathkumar P, Dutta K, Cahill SM, Chemmama IE, Williams R, Bonanno JB, Rice WJ, Stokes DL, et al. (2017). Molecular Architecture of the Major Membrane Ring Component of the Nuclear Pore Complex. Structure 25, 434–445. [PubMed: 28162953]

Uversky VN (2017). Intrinsic disorder here, there, and everywhere, and nowhere to escape from it. Cellular and molecular life sciences : CMLS 74, 3065–3067. [PubMed: 28589440]

Vallat B, Webb B, Westbrook JD, Sali A, and Berman HM (2018). Development of a Prototype System for Archiving Integrative/Hybrid Structure Models of Biological Macromolecules. Structure 26, 894–904 e892. [PubMed: 29657133]

van Heel M, and Schatz M. (2005). Fourier shell correlation threshold criteria. Journal of structural biology 151, 250–262. [PubMed: 16125414]

Velazquez-Muriel J, Lasker K, Russel D, Phillips J, Webb BM, Schneidman-Duhovny D, and Sali A. (2012). Assembly of macromolecular complexes by satisfaction of spatial restraints from electron microscopy images. Proc Natl Acad Sci U S A 109, 18821–18826. [PubMed: 23112201]

Viswanath S, Bonomi M, Kim SJ, Klenchin VA, Taylor KC, Yabut KC, Umbreit NT, Van Epps HA, Meehl J, Jones MH, et al. (2017a). The molecular architecture of the yeast spindle pole body core determined by Bayesian integrative modeling. Molecular biology of the cell 28, 3298–3314. [PubMed: 28814505]

Viswanath S, Chemmama IE, Cimermancic P, and Sali A. (2017b). Assessing Exhaustiveness of Stochastic Sampling for Integrative Modeling of Macromolecular Structures. Biophysical journal 113, 2344–2353. [PubMed: 29211988]

Viswanath S, and Sali A. (2019). Optimizing model representation for integrative structure determination of macromolecular assemblies. Proc Natl Acad Sci U S A 116, 540–545. [PubMed: 30587581]

Vogel S, and Wainwright SA (1969). A Functional bestiary: Laboratory studies about living systems (Addison-Wesley Publishing Company).

von Appen A, Kosinski J, Sparks L, Ori A, DiGuilio AL, Vollmer B, Mackmull MT, Banterle N, Parca L, Kastritis P, et al. (2015). In situ structural analysis of the human nuclear pore complex. Nature 526, 140–143. [PubMed: 26416747]

Wagner JW, Dama JF, Durumeric AE, and Voth GA (2016). On the representability problem and the physical meaning of coarse-grained models. J Chem Phys 145, 044108.

Walzthoeni T, Leitner A, Stengel F, and Aebersold R. (2013). Mass spectrometry supported determination of protein complex structure. Current opinion in structural biology.

Wang X, Cimermancic P, Yu C, Schweitzer A, Chopra N, Engel JL, Greenberg C, Huszagh AS, Beck F, Sakata E, et al. (2017). Molecular Details Underlying Dynamic Structures and Regulation of the Human 26S Proteasome. Molecular & cellular proteomics : MCP 16, 840–854. [PubMed: 28292943]

Wang Z, and Schroder GF (2012). Real-space refinement with DireX: from global fitting to side-chain improvements. Biopolymers 97, 687–697. [PubMed: 22696405]

Watson JD, and Crick FH (1953). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature 171, 737–738. [PubMed: 13054692]

Wells JN, and Marsh JA (2018). Experimental Characterization of Protein Complex Structure, Dynamics, and Assembly. Methods Mol Biol 1764, 3–27. [PubMed: 29605905]

Whitten AE, Jeffries CM, Harris SP, and Trewhella J. (2008). Cardiac myosin-binding protein C decorates F-actin: implications for cardiac function. Proc Natl Acad Sci U S A 105, 18360–18365. [PubMed: 19011110]

Woetzel N, Lindert S, Stewart PL, and Meiler J. (2011). BCL::EM-Fit: rigid body fitting of atomic structures into density maps using geometric hashing and real space refinement. Journal of structural biology 175, 264–276. [PubMed: 21565271]

Yang Q, Rout MP, and Akey CW (1998). Three-dimensional architecture of the isolated yeast nuclear pore complex: functional and evolutionary implications. Mol Cell 1, 223–234. [PubMed: 9659919]

Young JY, Westbrook JD, Feng Z, Sala R, Peisach E, Oldfield TJ, Sen S, Gutmanas A, Armstrong DR, Berrisford JM, et al. (2017). OneDep: Unified wwPDB System for Deposition, Biocuration, and Validation of Macromolecular Structures in the PDB Archive. Structure 25, 536–545. [PubMed: 28190782]

Zhou CY, Stoddard CI, Johnston JB, Trnka MJ, Echeverria I, Palovcak E, Sali A, Burlingame AL, Cheng Y, and Narlikar GJ (2017). Regulation of Rvb1/Rvb2 by a Domain within the INO8O Chromatin Remodeling Complex Implicates the Yeast Rvbs as Protein Assembly Chaperones. Cell Rep 19, 2033–2044. [PubMed: 28591576]

Box 1.

Molecular representation.

A structural model of a macromolecular assembly is defined by the relative positions and orientations of its components (*eg*, atoms, united atoms, residues, secondary structure elements, domains, subunits, and subcomplexes). Thus, the representation of a system is defined by all the structural variables that need to be determined based on input information, including the assignment of the system components to geometric objects such as points and spheres (Schneidman-Duhovny et al., 2014). An atomic representation can be coarsegrained by assigning unique subsets of atoms to higher level primitives (*eg*, beads and 3D Gaussians). Coarse-grained representations have proven useful, for example, in molecular dynamics simulations of lipid bilayers as well as structured and disordered proteins (Saunders and Voth, 2013). In our experience, selecting an appropriate representation is one of the most important decisions when performing integrative modeling, given the varied sparseness, noise, ambiguity, and resolution of the input datasets (Schneidman-Duhovny et al., 2014). An optimal representation facilitates accurate formulation of spatial restraints as well as efficient and complete sampling of good-scoring models, while still retaining sufficient detail without overfitting, so that the resulting models are maximally useful for subsequent biological analyses (Saunders and Voth, 2013; Schneidman-Duhovny et al., 2014; Viswanath and Sali, 2019).

While traditional structural biology methods usually produce a single atomic coordinate set, integrative models tend to be more complex in at least four respects (Sali et al., 2015). First, a model can be multi-scale, coarse-graining different levels of structural detail by a collection of geometrical primitives (*eg*, points, spheres, tubes, Gaussians, and probability densities) (Grime and Voth, 2014). Thus, the same part of the system can be described with multiple representations or different parts of the system can be represented differently. Second, a model can be multi-state, specifying multiple discrete states of the system that are needed simultaneously to explain the input information (each state may differ in structure and/or composition) (Molnar et al., 2014; Pelikan et al., 2009). Third, a model can also specify the order of states in time. This feature allows a representation of a multi-step biological process, a functional cycle (Diez et al., 2004), a kinetic network (Pirchi et al., 2011), or time evolution of a modeled system (*eg*, a molecular dynamics trajectory) (Bock et al., 2013). Finally, an ensemble of models is often provided to specify the uncertainty in the input information by including each model that on its own satisfies the input information within an acceptable threshold (*eg*, an ensemble of structures computed from a single NMR dataset). This aspect of the representation allows us to describe model uncertainty resulting from the incompleteness of input information; such ensembles are distinct from multiple states that represent actual variations in the structure, as implied by experimental information that cannot be accounted for by a single representative structure (Schneidman-Duhovny et al., 2014; Schroder, 2015). Thus, a generalized representation, already implemented in PDB-Dev (Burley et al., 2017; Vallat et al., 2018), allows us to encode an ensemble of multi-scale, multi-state, and time-ordered models.

Box 2:

Bayesian inference for scoring alternative models.

This solution came from a man with no direct connection to the problems of molecular cell biology. Thomas Bayes was an eighteenth century Presbyterian minister, who in his later life spent a significant amount of his spare time considering the "Doctrine of Chances", or probability theory. In essence, he understood that the probability of a model can be updated by iteratively considering additional information. For example, if "Happy Gallop" has won ten of his last twenty horse races, we tend to be ambivalent as to his chances of trotting to a comfortable win in his next race. However, what if one found out that when a particular jockey had ridden him, the horse won every one of those races - and that this jockey will be riding him in the next race? Then this information modifies upwards our estimate of him being first past the post. The corresponding formalization is Bayesian inference, a method of statistical inference in which Bayes' theorem is used to update the probability for a hypothesis as more information becomes available. As a structural biology exemplar, if we observe a cross-link between two residues, one can take this observation explicitly into account in formulating the likelihood of the structure having a distance between these two residues that is less than the maximal length of the cross-linker (Molnar et al., 2014). When sufficient information is available, the structure can be determined with high precision. An elegant and insightful application of Bayesian inference was described for determining protein structures based on NMR data (Rieping et al., 2005).

Formally, the posterior probability of model M given data D and prior information I is $p(M|D,I) \propto p(D|M,I) \cdot p(M|I)$. The model, M, consists of a structure X and unknown parameters Y, such as noise in the data. The prior $p(M|I)$ is the probability density of model M given I. The prior reflects information such as excluded volume, statistical potentials, and a molecular mechanics force field. The likelihood function $p(D|M,I)$ is the probability density of observing data D given M and I, and can be defined as a product over the individual measurements, $p(D|M,I) = \Pi_{i=1} N(d_i|f_i(X),\sigma_i)$, where $f_i(X)$ is a forward model that predicts the data point d_i in D that would have been observed for structure X in an experiment without noise; $N(d_i|f_i(X),\sigma_i)$ is a noise model that quantifies the deviation between the predicted and observed data points. A Gaussian noise model is often used, $N(d_i|f_i(X),\sigma_i) \propto exp(-[d_i - f_i(X)]^2/2\sigma_i^2)$, where σ_i is the noise parameter in Y that can optionally be determined as part of the model. Finally, a Bayesian scoring function is defined as the negative logarithm of the posterior probability density: $S(M) = -log\ p(M|D,I)$. In the Bayesian view, the output model is in fact best equated to the posterior model density that specifies a distribution of alternative single models M with varying probabilities, not a single model (although single representative or average models can always be proposed based on the posterior model density).

A key advantage of defining the posterior model density in a Bayesian fashion, compared to traditional least-squares-scoring functions, is that it allows for objective mixing of different types of information (*ie*, balancing varying uncertainties of varying input information), which is an essential requirement for integrative modeling. As a result, the output models tend to be more accurate, with more accurate estimates of their

uncertainty. A Bayesian approach allows us to quantify model uncertainty in a strict sense. Repeated nonlinear least-squares minimization may produce a diverse set of solutions, but its spread will mainly reflect the power of the optimizer. Another advantage is that we know how to deal with the nuisance parameters Y, whereas least-squares minimization needs to invoke additional recipes, such as cross-validation. The Bayesian approach is also relatively robust in terms of the specific parameterization of the representation of M. Finally, multiple choices about model representation and scoring function can in principle be quantified and compared using model selection criteria (Viswanath and Sali, 2019), such as the model evidence (Knuth et al., 2015).

Box 3:

Glossary

Modeling: the process of converting input information into a model and its uncertainty.

Input information: experimental data and prior models used for computing a model.

Prior models: physical theories, statistical preferences, and other models (*eg*, X-ray structures and comparative models of subunits in a complex) used for computing a model.

Spatial restraint: a function that quantifies the degree of consistency between a model and a single piece of information; often expressed as the squared difference between the model and target value of some spatial feature, such as a distance in a traditional least-squares approach, or as a data likelihood or prior in a Bayesian approach.

Scoring function: a function that quantifies the degree of consistency between a model and input information; often expressed as a weighted sum of spatial restraints in a traditional least-squares approach or as a posterior model density in a Bayesian approach.

Model representation: the set of variables whose values are determined by modeling based on the input information.

Representation precision: a descriptor of the detail in the representation of the structural model (*eg*, atomic models consist of atoms).

Sampling precision: granularity of sampling used to find models consistent with input information. Model accuracy or error: the deviation of the model from the "truth".

Good-scoring model: a model that is sufficiently consistent with given information; for example, a model whose score is better than some threshold on the scoring function used for sampling, or a model that is within all error bars on the input data. In a truly Bayesian approach, there is in principle no need to consider only good-scoring models as each sampled model can be weighted by its posterior probability.

Ensemble of models: a sample of sufficiently good-scoring models.

Precision or uncertainty of a model or ensemble of models: a measure of variability of the ensemble of models.

Single-state model: a model that specifies a single structural state and value for any other parameter.

Multi-state model: a model that specifies two or more structural states in the samples used for determining input information and values for any other parameter.
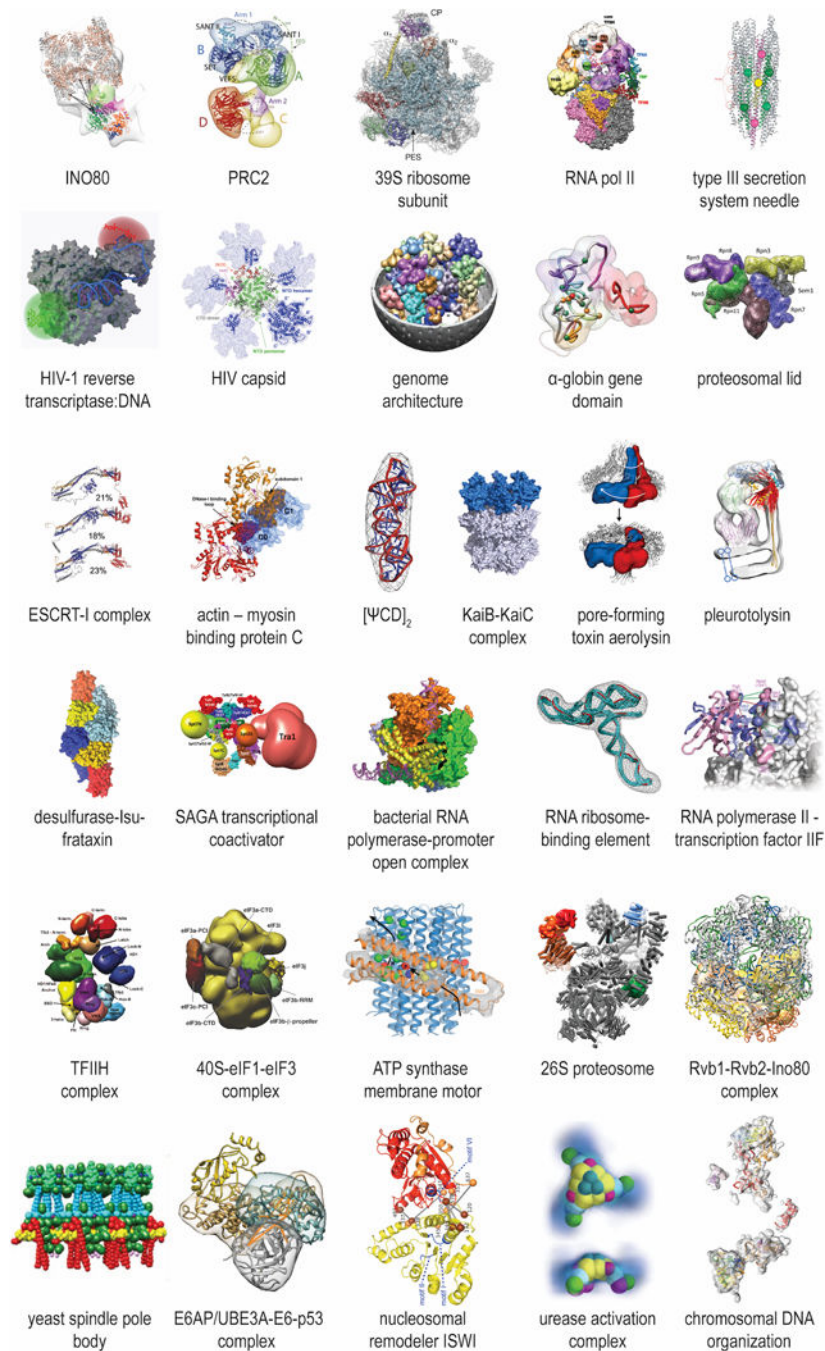
Figure 1: Examples of integrative structures.
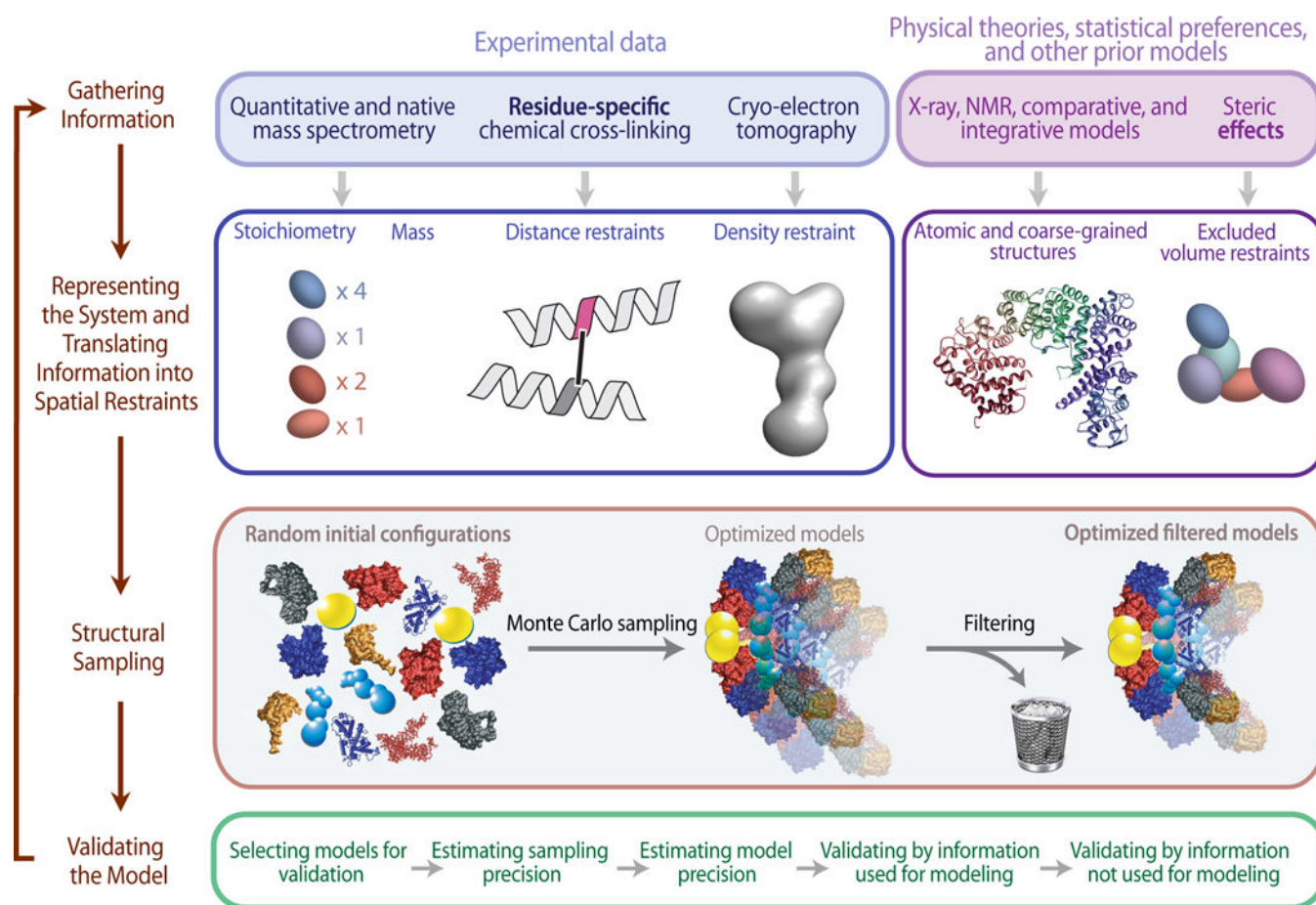See Table 3 for details on each structure.

Figure 2: Description of the iterative integrative modeling workflow.
As illustrated, the four stages include: (1) gathering all available experimental data and prior information; (2) translating information into representations of assembly components and a scoring function for ranking alternative assembly structures; (3) sampling structural models; and (4) validating the model. In this example, representations of the components of a complex are based on models of its components. Some component representations are coarse-grained by using spherical beads corresponding to multiple amino acid residues, to reflect the lack of information and/or to increase efficiency of structural sampling. The scoring function consists of spatial restraints that are obtained from CX-MS experiments and a cryo-electron tomography density map. The sampling explores both the conformations of the components and/or their configuration, searching for those assembly structures that satisfy the spatial restraints as well as possible. The result is an ensemble of many good-scoring models that satisfy the input data within acceptable thresholds. The sampling is then assessed for convergence, models are clustered, and evaluated by the degree to which they satisfy the input information used to construct them as well as omitted information. The protocol can iterate through the four stages until the models are judged to be satisfactory, most often based on their precision and the degree to which they satisfy the data. Finally, the models and data are deposited into PDB-Dev (https://pdb-dev.wwpdb.org) (Burley et al., 2017; Vallat et al., 2018).
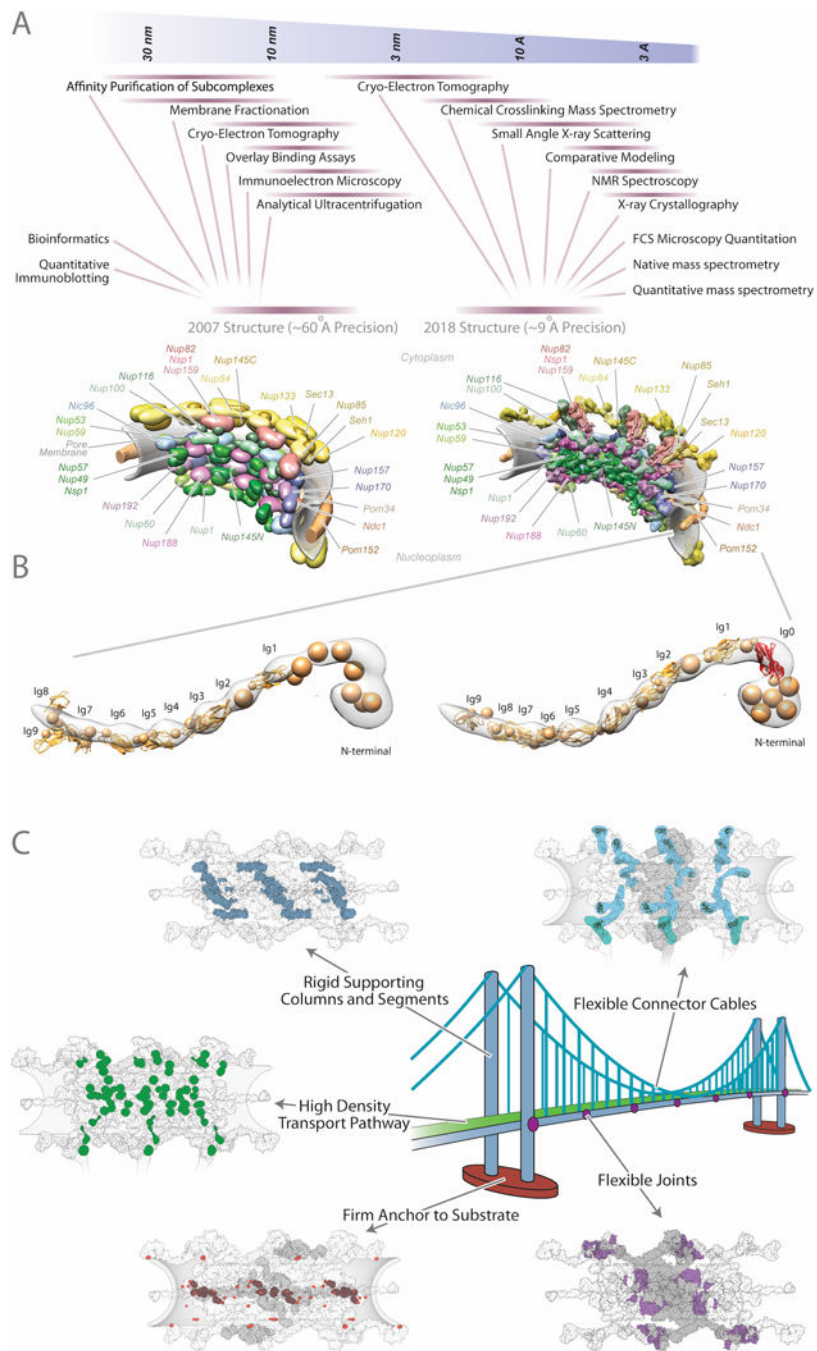
Figure 3: Integrative structures of the yeast NPC.
(A) A comparison of the integrative NPC structures determined in 2007 (Alber et al., 2007b) and 2018 (Kim et al., 2018) illustrates how the integration of a larger amount of more precise data led in turn to a structure with a higher precision. **(B)** A comparison of two representative Pom152 models, without and with an atomic model of the first Ig domain (Hao et al., 2018; Upla et al., 2017), shows how consideration of additional information (*ie*, knowledge of an atomic structure of the first Ig domain (Ig0)) into the representation of a protein improves its model. **(C)** Insights into the architectural principles and functions of the

NPC. Five examples of analogous structural principles shared by an NPC and a suspension bridge are each illustrated by a specific color showing the features sharing these principles, in the NPC by panels showing three of the eight spokes of an NPC viewed from its axial center: (i) a firm anchor to a substrate (brown / orange), (ii) rigid supporting columns and segments (dark blue) that (iii) appear somewhat flexibly jointed to each other (purple), (iv) flexible connector cables that tie together all the structural elements (light blue), (v) collectively forming a transport pathway supporting a high density of trafficking routes (green).

Table 1:

Example methods that are informative about a variety of structural aspects of biomolecular systems are listed.

3DEM, 3D electron microscopy; DEER, double electron-electron resonance; EPR, electron paramagnetic resonance; FRET, Foerster resonance energy transfer; HDex, hydrogen/deuterium exchange; NMR, nuclear magnetic resonance; SAS, small-angle scattering; XL-MS, cross linking - mass spectrometry.

Structural information	Method
Stoichiometry	MS, quantitative fluorescence imaging
Atomic structures of parts of the studied system	X-ray and neutron crystallography, NMR spectroscopy, 3DEM, comparative modeling, and molecular docking
3D maps and 2D images	Electron microscopy and tomography
Atomic and protein distances	NMR, FRET, and other fluorescence techniques; DEER, EPR, and other spectroscopic techniques; XL-MS, disulfide bonds detected by gel electrophoresis
Binding site mapping	NMR spectroscopy, mutagenesis, FRET, and XL-MS
Size, shape, and distributions of pairwise atomic distances	SAS
Shape and size	Atomic force microscopy, ion mobility mass spectrometry, fluorescence correlation spectroscopy, fluorescence anisotropy, analytical ultracentrifugation
Component positions	Super-resolution optical microscopy, FRET imaging, immuno- electron microscopy
Physical proximity	Co-purification, native mass spectrometry, XL-MS, molecular genetic methods, and gene/ protein sequence covariance
Solvent accessibility	Footprinting methods, including HDex assessed by MS or NMR, and even functional consequences of point mutations
Proximity between different genome segments	Chromosome Conformation Capture
Propensities for different interaction modes	Molecular mechanics force fields, potentials of mean force, statistical potentials, and sequence co-variation

Table 2:

Software resources for integrative modeling.

Program	Functionality	Web Site	Reference
ISD	Bayesian modeling based on NMR data		(Rieping et al., 2005)
IMP	Integrative modeling	integrativemodeling.org	(Russel et al., 2012)
Rosetta	Integrative modeling	rosettacommons.org	(Das and Baker, 2008)
ISDB	Integrative modeling	plumed.org	(Bonomi and Camilloni, 2017)
pow^{er}	Integrative modeling	lbm.epfl.ch/resources/	(Degiacomi and Dal Peraro, 2013)
cMNXL, Jwalk/ MNXL	Integrative modeling	topf-group.ismb.lon.ac.uk/Software	(Bullock et al., 2018a; Bullock et al., 2018b)
PyRy3D	Integrative modeling	genesilico.pl/pyry3d/	J. M. Kasprzak, M. Dobrychtop, J. Bujnicki
PGS	Modeling genome structure	github.com/alberlab/PGS	(Hua et al., 2018)
TADBit	Modeling genome structure	sgt.cnag.cat/3dg/tadbit/	(Serra et al., 2017)
MDFF/NAMD	Fitting of molecular models into EM maps using MD simulations	ks.uiuc.edu/Research/mdff	(Trabuco et al., 2008)
ATSAS	Integrative modeling using SAXS	embl-hamburg.de/biosaxs	(Franke et al., 2017)
iFoldRNA	Integrative modeling of RNA	iFoldRNA.dokhlab.org	(Sharma et al., 2008)
HADDOCK	Integrative modeling using docking and data derived restraints	haddock.science.uu.nl	(Dominguez et al., 2003)
ATTRACT-EM	Integrative modeling using docking and EM	attract.ph.tum.de	(de Vries and Zacharias, 2012)
DireX	Flexible fitting of EM maps with data derived distance restraints.	schroderlab.org/software/direx/	(Wang and Schroder, 2012)
MDFit	MD based integrative modeling using EM maps	smog-server.org/SBMextension.html#mdfit	(Ratje et al., 2010)
FPS	Integrative modeling using FRET data	www.mpc.hhu.de/en/software/fps.html	(Kalinin et al., 2012)
XPLOR-NIH	Structure determination using NMR data	nmr.cit.nih.gov/xplor-nih/	(Schwieters et al., 2018)
PatchDock	Molecular docking by shape complementarity	bioinfo3d.cs.tau.ac.il/PatchDock/	(Schneidman-Duhovny et al., 2005)
iSPOT	Structure determination using SAS, footprinting and docking	www.theyanglab.org/ispot/	(Hsieh et al., 2017)
BCL	Various servers for integrative modeling	meilerlab.org/index.php/servers	(Woetzel et al., 2011)
ChimeraX	Model visualization	rbvi.ucsf.edu/chimerax	(Goddard et al., 2018)
VMD	Model visualization	ks.uiuc.edu/research/vmd	(Humphrey et al., 1996)
Protein Model Portal	Portal to atomic models of proteins	proteinmodelportal.org	(Haas et al., 2013)
PDB-Development	Archiving of integrative structures	pdb-dev.wwpdb.org	(Burley et al., 2017)

Table 3:

Examples of integrative structures (shown in Figure 1).

System name	Input data	Accession	Citation
INO80	17-Å resolution cryo-electron microscopy (EM) map, 212 intra-protein and 116 inter-protein cross-links		(Tosi et al., 2013)
Polycomb Repressive Complex 2 (PRC2)	21-Å resolution negative-stain EM map and ~60 intra-protein and inter-protein cross-links		(Ciferri et al., 2012)[1]
Large subunit of the mammalian mitochondrial ribosome (39S)	4.9-Å resolution cryo-EM map and ~70 inter-protein cross-links	4CE4	(Greber et al., 2014b)[2]
RNA polymerase II transcription pre-initiation complex	16-Å resolution cryo-EM map plus 157 intra-protein and 109 inter-protein cross-links		(Murakami et al., 2013)[3]
Type III secretion system needle	19.5-Å resolution cryo-EM map and solid-state nuclear magnetic resonance (NMR) data	2LPZ	(Loquet et al., 2012)[4]
Productive HIV-1 reverse transcriptase:DNA primer-template complex in the open educt state	Foerster resonance energy transfer (FRET) positioning and screening using a known HIV-1 reverse transcriptase structure		(Kalinin et al., 2012)[2]
HIV-1 capsid protein	Residual dipolar couplings and small-angle X-ray scattering (SAXS) data	2M8L, 2M8N, 2M8P	(Deshmukh et al., 2013)[5]
Human genome	Tethered chromosome conformation capture and population-based modeling		(Kalhor et al., 2011)[6]
Drosophila genome	Chromosome conformation capture and lamina DamID		(Li et al., 2017)
α-globin gene domain	Chromosome Conformation Capture Carbon Copy (5C)		(Bau et al., 2011)[7]
Proteosomal lid	Native mass spectrometry and 28 cross-links		(Politis et al., 2014)[8]
ESCRT-I complex	SAXS, double electron-electron transfer, and FRET		(Boura et al., 2011)
Actin together with the cardiac myosin binding protein C	Crystallographic and NMR structures of subunits and domains, with positions and orientations optimized against SAXS and small-angle neutron scattering data to reveal information about the quaternary interactions		(Whitten et al., 2008)[9]
[ΨCD]$_2$	Averaged cryo-electron tomography map, NMR	2L1F	(Miyazaki et al., 2010)
Cyanobacterial circadian timing KaiB-KaiC complex	Hydrogen/deuterium exchange and collision cross-section data from mass spectrometry		(Snijder et al., 2014)
Pre-pore and pore conformations of the pore-forming toxin aerolysin	Cryo-EM data and molecular dynamics simulations		(Degiacomi et al., 2013)[10]
A segment of a pleurotolysin pore map (~11 Å resolution); an ensemble of conformations shows the trajectory of β-sheet opening during pore formation	Cryo-EM, X-ray crystal subunit structures, fluorescence spectroscopy, cross-linking	4V2T	(Lukoyanova et al., 2015)[11]
Ternary complex of the iron-sulfur cluster assembly proteins desulfurase (orange) and scaffold protein Isu (blue) with a bacterial ortholog of frataxin (yellow)	NMR chemical shifts, SAXS, mutagenesis		(Prischi et al., 2010)[12]
SAGA transcription coactivator complex	199 inter- and 240 intra-subunit cross-links, several comparative models based on X-ray crystal structures, and a transcription factor IID core EM map at 31 Å resolution		(Han et al., 2014)[13]

System name	Input data	Accession	Citation
Bacterial (Thermus aquaticus) RNA polymerase-promoter open complex; subsequently validated by a crystal structure (Feng et al., 2016)	FRET		Mekler, 2002 #400}
RNA ribosome-binding element from the turnip crinkle virus genome	NMR, SAXS, EM	DOI: 10.6084/ m9.figshare.1295199	(Gong et al., 2015)[14]
Complex between RNA polymerase II and transcription factor IIF	Deposited crystal structure of RNA polymerase II, comparative models of some domains in transcription factor IIF and 95 intra-protein and 129 inter-protein cross-links		(Chen et al., 2010)[15]
Human and yeast TFIIH	XL-MS data, biochemical analyses, and previously published electron microscopy maps		(Luo et al., 2015)
40S-eIF1-eIF3 translation initiation complex	X-ray crystallography, EM, XL-MS		(Erzberger et al., 2014)
ATP synthase membrane motor	cryo-EM (~7.8 Å resolution), XL-MS, and evolutionary couplings		(Leone and Faraldo-Gomez, 2016)[16]
26S proteasome	67 inter-protein and 26 intra-protein chemical cross-links in combination with EM maps	5LN3	(Wang et al., 2017)
Ino80 insert domain bound to the Rbv1/Rvb2 dodecamer	12-Å resolution cryo-EM map, 226 chemical cross-links		(Zhou et al., 2017)
Core of the yeast spindle pole body (SPB)	in vivo FRET, SAXS, X-ray crystallography, EM, two-hybrid analysis		(Viswanath et al., 2017a)[2]
E6AP/UBE3A-E6-p53 enzyme-substrate complex	XL-MS data of the complex with and without E6	PDBDEV_00000022, PDBDEV_00000023	(Sailer et al., 2018)[17]
Nucleosome remodeler ISWI	XL-MS, SAXS, protein-protein docking		(Harrer et al., 2018)
Urease activation complex	Mobility-mass spectrometry data		(Eschweiler et al., 2018)
Chromosomal DNA organization	Super-resolution microscopy methods OligoSTORM and OligoDNA-PAINT, Hi-C data		(Nir et al., 2018)[18]

[1] Figure panel reprinted from figure 11 of (Ciferri et al., 2012), used under the terms of the Creative Commons Attribution 3.0 license (https://creativecommons.org/licenses/by/3.0/).

[2] Figure panel obtained *via* personal communication and used with permission of the author.

[3] Panel from Figure 5 of (Murakami et al., 2013). Reprinted with permission from AAAS.

[4] Figure 3 reprinted by permission from Springer Nature Terms and Conditions for RightsLink Permissions Springer Nature Customer Service GmbH: Nature "Atomic model of the type III secretion system needle." Loquet A, Sgourakis NG, Gupta R, Giller K, Riedel D, Goosmann C, Griesinger C, Kolbe M, Baker D, Becker S, Lange A. Copyright Springer Nature Publishing AG (2012).

[5] Figure 9 reprinted (adapted) with permission from (Deshmukh et al., 2013). Copyright (2013) American Chemical Society.

[6] Figure 6 reprinted by permission from Springer Nature Terms and Conditions for RightsLink Permissions Springer Nature Customer Service Centre GmbH: Nature Biotechnology "Genome architectures revealed by tethered chromosome conformation capture and population-based modeling." Kalhor R, Tjong H, Jayathilaka N, Alber F, Chen L. Copyright Springer Nature Publishing AG (2012).

[7] Figure 4 reprinted by permission from Springer Nature Terms and Conditions for RightsLink Permissions Springer Nature Customer Service Centre GmbH: Nature Structural & Molecular Biology "The three-dimensional folding of the α-globin gene domain reveals formation of chromatin globules." Baù D, Sanyal A, Lajoie BR, Capriotti E, Byron M, Lawrence JB, Dekker J, Marti-Renom MA. Copyright Springer Nature Publishing AG (2011).

[8] Figure 2 reprinted by permission from Springer Nature Terms and Conditions for RightsLink Permissions Springer Nature Customer Service Centre GmbH: Nature Methods "A mass spectrometry–based hybrid method for structural modeling of protein complexes." Politis A, Stengel F, Hall Z, Hernández H, Leitner A, Walzthoeni T, Robinson CV, Aebersold R. Copyright Springer Nature Publishing AG (2014).