

UCSF

UC San Francisco Previously Published Works

Title

From integrative structural biology to cell biology

Permalink

<https://escholarship.org/uc/item/20k8p372>

Author

Sali, Andrej

Publication Date

2021

DOI

10.1016/j.jbc.2021.100743

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

From integrative structural biology to cell biology

Received for publication, February 5, 2021, and in revised form, April 9, 2021. Published, Papers in Press, May 4, 2021.
<https://doi.org/10.1016/j.jbc.2021.100743>

Andrej Sali*¹

From the Research Collaboratory for Structural Bioinformatics Protein Data Bank, the Department of Bioengineering and Therapeutic Sciences, the Quantitative Biosciences Institute (QBI), and the Department of Pharmaceutical Chemistry, University of California, San Francisco, San Francisco, California, USA

Edited by Karen Fleming

Integrative modeling is an increasingly important tool in structural biology, providing structures by combining data from varied experimental methods and prior information. As a result, molecular architectures of large, heterogeneous, and dynamic systems, such as the ~52-MDa Nuclear Pore Complex, can be mapped with useful accuracy, precision, and completeness. Key challenges in improving integrative modeling include expanding model representations, increasing the variety of input data and prior information, quantifying a match between input information and a model in a Bayesian fashion, inventing more efficient structural sampling, as well as developing better model validation, analysis, and visualization. In addition, two community-level challenges in integrative modeling are being addressed under the auspices of the Worldwide Protein Data Bank (wwPDB). First, the impact of integrative structures is maximized by PDB-Development, a prototype wwPDB repository for archiving, validating, visualizing, and disseminating integrative structures. Second, the scope of structural biology is expanded by linking the wwPDB resource for integrative structures with archives of data that have not been generally used for structure determination but are increasingly important for computing integrative structures, such as data from various types of mass spectrometry, spectroscopy, optical microscopy, proteomics, and genetics. To address the largest of modeling problems, a type of integrative modeling called metamodeling is being developed; metamodeling combines different types of input models as opposed to different types of data to compute an output model. Collectively, these developments will facilitate the structural biology mindset in cell biology and underpin spatiotemporal mapping of the entire cell.

Progress of structural biology: From single molecule to cell

Introduction

Structural biology was born in the 1950s, with the atomic structure of the myoglobin molecule determined by modeling based primarily on data from X-ray crystallography (1, 2) (Fig. 1). The toolbox and scope of structural biology have been

expanding ever since. In steady progress, crystallography was joined by nuclear magnetic resonance (NMR) spectroscopy and 3D electron microscopy (3DEM), resulting in molecular architectures of 183,234 biomolecular systems as of February 3, 2021 (3). Some of these structurally defined systems contain as many as hundreds of protein subunits, as exemplified by the Nuclear Pore Complex (NPC) (4). Further progress toward mapping spatiotemporal organization of even larger modules of the cell, and eventually the entire cell, is inevitable. Such complex depictions are unlikely to be obtained using data from any single method. Instead, they are expected to be computed by integrative (hybrid) modeling that combines data from multiple experimental methods and prior models of parts of the entire system (5, 6).

Outline

Here, we review the integrative approach and some of the key roles that the Worldwide Protein Data Bank (wwPDB; <https://wwpdb.org>) (7) has played in its development. In particular, we describe the goals and workflow of integrative structure modeling (5), illustrated by its application to determining the molecular architecture of the NPC. We then outline our open-source *Integrative Modeling Platform* (IMP) program for computing and validating integrative structure models (8, 9). Next, we discuss a number of key challenges in the further development of integrative modeling. Some of these challenges are being addressed under the auspices of the wwPDB, including the expansion of the PDB to archive and validate integrative models and the data on which they are based (10, 11). These activities illustrate the key role of the PDB in nucleating the structural biology community and expanding its scope. A particularly ambitious goal is discussed last, namely, the application of a new class of integrative modeling to the cell mapping problem, in principle fully integrating structural biology and cell biology (12, 13).

Integrative structure modeling

Integrative structures

Some of the very first models of biomolecular structures, including the model of the DNA double helix (14), were in fact integrative models based on a multitude of considerations. Modern integrative structure modeling is inspired by the early fitting of crystallographic subunit structures into a 3DEM

* For correspondence: Andrej Sali, sali@salilab.org.



Andrej Sali, Professor of Bioengineering and Therapeutic Sciences and Associate Dean of Research of School of Pharmacy, University of California, San Francisco, has been a pioneer in the development and application of integrative structure modeling.

density map of the actin–myosin complex (15). Many examples of recent integrative structures were reviewed (5), ranging from small proteins and nucleic acids to the entire genome.

Definition of integrative modeling

Integrative modeling is motivated by the desire to utilize all available information about the modeled system, in an effort to maximize the accuracy, precision, and completeness of the resulting model (5, 6, 8, 10, 11, 16–28). A large variety of experimental and computational methods can provide input information for integrative modeling of biomolecular structures (5). Examples include X-ray crystallography, NMR spectroscopy, 3DEM, small-angle scattering (SAS), chemical cross-linking with mass spectrometry, affinity copurification, quantitative genetic interaction mapping, molecular mechanics force fields, statistical potentials, comparative protein structure modeling, and sequence covariation. Information used for integrative modeling can be at either high or low resolution (*e.g.*, nuclear Overhauser effect and affinity copurification, respectively). It can also be dense or sparse (*e.g.*, a typical X-ray diffraction dataset and an SAS profile, respectively). Given input information, integrative modeling then aims to find “all” models whose properties match the input information within an acceptable tolerance. This description in fact applies to all structure determination methods. For example, a crystallographic structure is computed by finding a set of atomic coordinates whose computed diffraction pattern

in a given crystal arrangement reproduces the observed diffractions while also satisfying stereochemistry rules; a 3DEM structure is computed by finding a set of atomic coordinates whose shape matches that of the observed 3DEM map and satisfies stereochemistry; and an *ab initio* model is computed by finding a set of atomic coordinates that come as close as possible to ideal values for distances, angles, and other spatial features as specified by an energy function. The only distinction of integrative modeling is that it aims to use explicitly all available information of any type.

Integrative modeling workflow

The integrative modeling workflow iterates through four stages that convert input information into an output model (Fig. 2): (i) gathering all available experimental data and prior information (physical theories, statistical analyses, and other prior models); (ii) translating information into representations of model components and a scoring function for ranking alternative models; (iii) sampling models; and (iv) validating the model. In this four-stage scheme, input information can contribute toward a model 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 (acceptable) models; (iv) filtering models based on the input information; and (v) validating the resulting models. We now discuss each of these ways in turn, using integrative structure modeling of the yeast NPC as an example (Fig. 3) (4).

Nuclear pore complex

The yeast NPC is a ~52-MDa complex that consists of ~550 protein subunits of ~30 different types. It is embedded in the nuclear envelope and plays a central role in cell biology by mediating nucleocytoplasmic transport of proteins and RNA, *via* cognate transport factors. The sheer size and

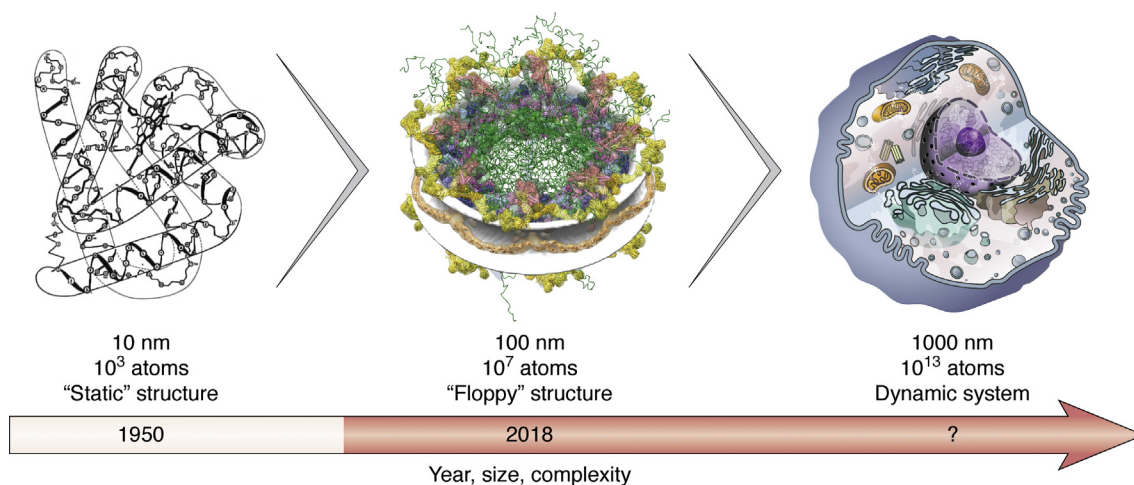


Figure 1. Progress of structural biology. *Left*, the myoglobin structure determined by X-ray crystallography at atomic resolution (1, 2). *Center*, the molecular architecture of the yeast NPC determined by an integrative approach at nanometer resolution (4). *Right*, a visualization of the eukaryotic cell based on the cell biology literature (drawing by Yekaterina Kadyshvskaya). This review focuses on the role and challenges of integrative modeling of complex biomolecular systems (red arrow).

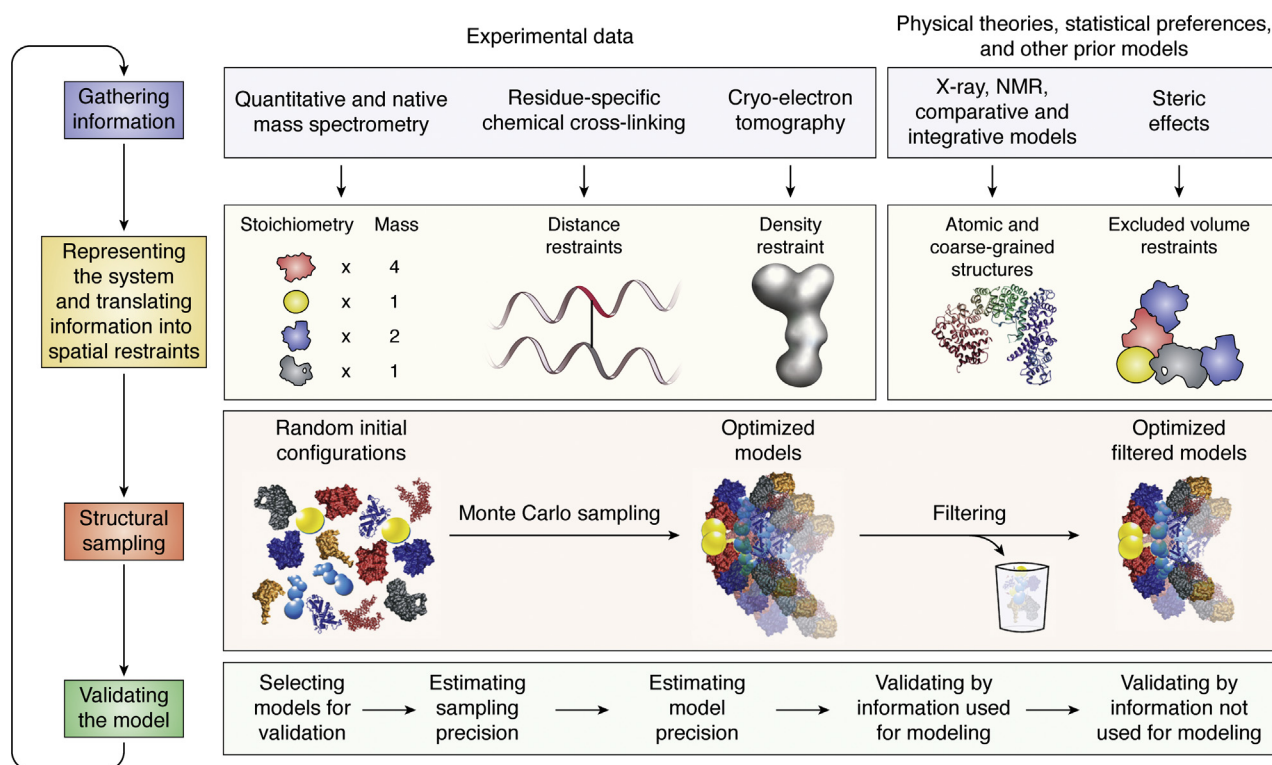


Figure 2. Description of the iterative integrative modeling workflow (5). In this example, representations of the components of a complex are based on models of its components. Some component representations are coarse-grained by using spherical beads corresponding to multiple amino acid residues to reflect the lack of information and/or to increase the efficiency of structural sampling. The scoring function consists of spatial restraints that are obtained from chemical cross-linking with mass spectrometry (CX-MS) cross-links and a cryoelectron 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 tolerances. The sampling is then assessed for its precision, followed by clustering and evaluating the models by the degree to which they satisfy the input information used to and not used to construct them. The protocol can iterate through the four stages until the models are judged to be satisfactory, most often on the basis of their precision and the degree to which they satisfy the data.

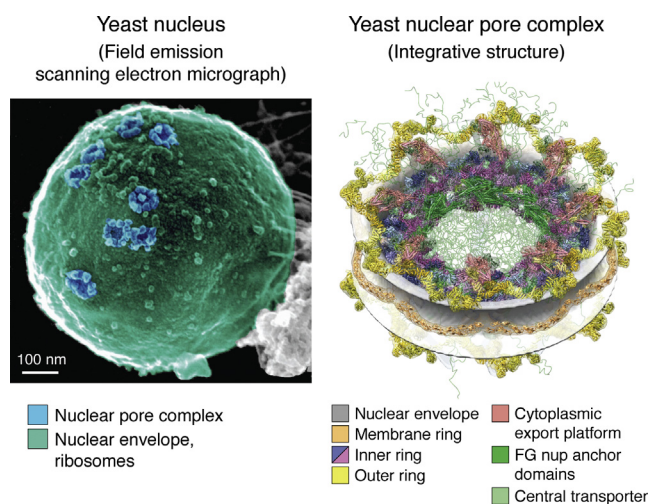


Figure 3. The yeast NPC. *Left*, field emission scanning electron micrograph of the yeast nucleus (119). Blue pseudocoloring highlights the NPCs; green pseudocoloring highlights the nuclear envelope together with the attached ribosomes. Scale bar represents 100 nm. *Right*, integrative structure of the yeast NPC (4). The 8-fold symmetric modular organization of the NPC is indicated, in the context of the nuclear envelope (gray): membrane ring (salmon), inner ring (blue and purple), outer rings (yellow), cytoplasmic export platform (red), FG nup anchor domains (green), and central transporter (light green). The comparison between the two panels illustrates the power of integrative structural biology to provide detailed descriptions of complex biomolecular systems. NPC, Nuclear Pore Complex.

flexibility of the NPC makes it all but impossible to solve its molecular architecture by conventional atomic resolution techniques, such as X-ray crystallography. However, integrating information from multiple sources, including stoichiometry from protein quantification, residue distances from chemical cross-linking with mass spectrometry, and the overall NPC shape from cryoelectron tomography, resulted in a relatively precise model.

Model representation

First, input information can be used to define model representation (5). The representation specifies the variables whose values will be determined by modeling. Thus, it specifies the components of the system, such as atoms, coarse-grained particles, and subunits in a complex, including their copy numbers. It also specifies the type of component coordinates, such as positions, orientations, and conformations, and potentially other aspects of a model (below). For example, in crystallography, model representation commonly includes Cartesian coordinates and an isotropic temperature factor for each atom in the molecule. In the NPC modeling, quantitative MS and *in vivo* calibrated imaging determined the copy numbers of the constituent NPC proteins. The NPC model was represented by the positions and orientations of its parts. These parts included (i)

rigid bodies corresponding to individual protein domains, subunits, or even subcomplexes for which the structure was determined previously by crystallography, comparative protein structure modeling, or integrative modeling and (ii) flexible regions otherwise. Representations were coarse-grained by using spherical beads corresponding to multiple amino acid residues to reflect the lack of information and/or to increase the efficiency of searching for acceptable models.

Scoring function

Second, information can be used to construct a scoring function and compute its value. The scoring function quantifies the degree of a match between a tested model and the input information. 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. An acceptable model is a model that sufficiently satisfies input information by some definition. Spatial restraints on the NPC structure included a correlation coefficient between a model and a cryoelectron tomography map of the entire complex at 28-Å resolution, 3077 distance restraints on pairs of residues spanning chemical cross-links, the rigid shape of the nuclear envelope pore, an excluded volume penalty, and others.

Searching for models

Third, information can be used to constrain the model search space. 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 (29). In practice, other methods, such as stochastic sampling *via* a Monte Carlo scheme (30), are often used. The demands on sampling increase with the number of degrees of freedom spanning the model, which in turn depends on the size of the modeled system and the detail with which it is represented. Sampling efficiency also depends crucially on the shape of the scoring landscape (cf, for many samplers, a funnel landscape is easier to sample than a golf course landscape) and the efficiency of evaluating the scoring function for a given model. The search for acceptable structural models of the NPC relied on replica exchange Gibbs sampling, based on the Metropolis Monte Carlo algorithm (31–33), starting with random initial structures in the context of the nuclear envelope. In addition, to increase the efficiency of sampling, only one of the eight symmetry units of the NPC was sampled, while still evaluating the scoring function for the complete NPC *via* appropriate consideration of its C_8 -symmetry. Hundreds of thousands of independent Monte Carlo runs resulted in a set of similar models that sufficiently satisfied the input information.

Filtering models

Fourth, some information can be used for filtering acceptable 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 EM 2D class of a complex to find all those molecular docking solutions whose 2D projections match the class (33–36). Integrative modeling of the NPC did not rely on any filtering.

Validating models

Finally, a subset of information can be set aside to validate a model. Validation of a model is essential to avoid its over-interpretation (21). 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. The NPC model was tested, for example, by comparison with previously published lower-resolution data, including affinity copurifications and immuno-EM localizations of tagged protein components (17). In addition to assessing the acceptable models by comparison against information not used for modeling, the validation also includes an assessment of the thoroughness of structural sampling (sampling precision (37)), quantification of the agreement between the model and information used for modeling, and estimation of model precision. The model precision is defined by the variability among the acceptable models, provided sufficient sampling was performed (9, 37). In fact, it is this set of superposed acceptable models that can be considered the final model, equivalently to the ensemble of structures computed based on NMR data. The yeast NPC map localizes the 452 constituent proteins with an average precision of ~ 1 nm. Independently determined NPC structures are in broad agreement with each other (4, 38–43), with differences between the models resulting from the heterogeneity of the NPC architecture, different conditions used to purify and characterize physical samples, variation among the species, and uncertainty in the models themselves.

Workflow iteration

The entire four-stage modeling process is generally performed iteratively until a satisfactory model is obtained. In the early iterations, modeling can pinpoint inconsistencies in the input data and modeling assumptions. For example, it may not be possible to find a model of an assembly of multiple components that satisfies reliable chemical cross-link restraints for given component copy numbers, prompting a re-examination of stoichiometry measurements (44). Similarly, the satisfaction of available data may require using a flexible instead of a rigid representation of subunits in a complex (45). Another example is the validation of a crystallographic interface between two components observed in a crystallographic study of the dimer, by its approximate reproduction in a model based on less precise data (35); this validation can then be followed up by imposing the validated crystallographic interface as a constraint to obtain a more precise model. Thus, integrative modeling can contribute continually during a combined experimental and computational study of a system and is best initiated early in a structure determination effort. Integrative studies are distinctly multidisciplinary and consequently often require a team of collaborators, including both experimentalists and modelers.

Using integrative structures

An integrative approach frequently results in a coarse model depicting the molecular architecture instead of a detailed atomic structure. Nevertheless, such lower-resolution depictions still have many applications (5). For example, the NPC map has revealed insights about its architecture, function, and evolution (4, 34, 46, 47). It is also conceivable that the structure will facilitate studying how the NPC assembles and disassembles during the cell cycle; how it interacts with other key systems in the cell, such as chromatin and the spindle pole body; and how its function can be modulated to study the basic biology of the cell as well as for therapeutic purposes.

Software for integrative modeling

A number of computer programs have been used for integrative structure modeling, including HADDOCK (48, 49), Rosetta (50, 51), PHENIX (52), BCL (53), XPLOR-NIH (54), TADbit (55, 56), PGS (57), iSPOT (58), FPS (59), PatchDock (60), BioEn (61), and others (5, 10). Next, we review our own open-source IMP software package that aims to provide comprehensive support for implementing and distributing integrative modeling protocols (Fig. 4) (<https://integrativemodeling.org>) (8, 62). IMP's relative strengths include a large variety of model representations, scoring functions based on different data, and sampling schemes, all of which can be mixed and matched relatively easily with each other to

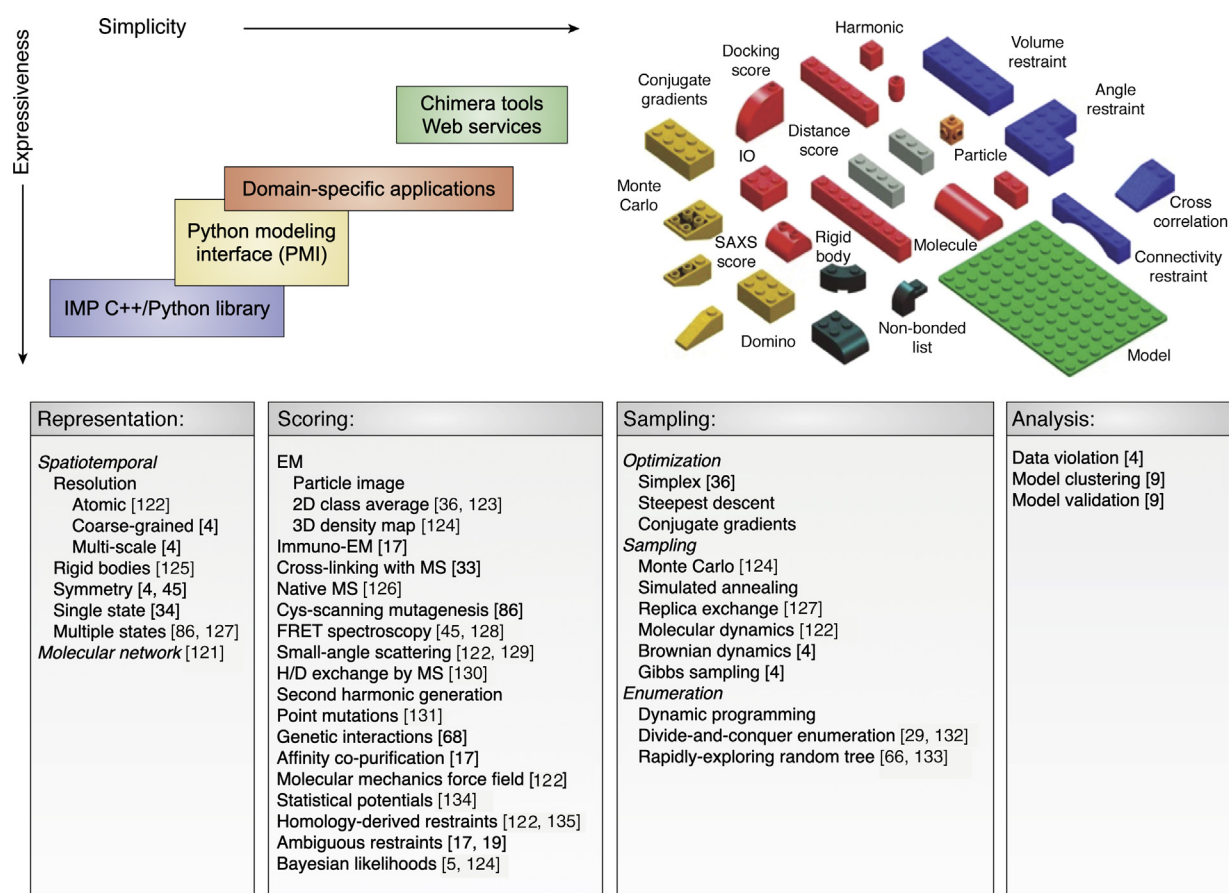


Figure 4. IMP architecture. *Top*, IMP has a multitiered architecture, with the kernel written in C++, with progressively simpler levels that facilitate easier access to IMP functionality, but at a cost of losing some flexibility. The simplest access to IMP is enabled through its interfaces to web servers and to the molecular visualization program ChimeraX (71, 120). A more flexible layer is the *Python Modeling Interface* (PMI) that allows users to script the representation, scoring, sampling, and analysis protocols in Python (62). The Lego blocks indicate the modularity of IMP and the ability to mix and match different model representations, scoring functions, and sampling methods. *Bottom*, IMP implements various model representations, scoring function terms, sampling schemes, and analyses; publications exemplifying the use of various functionalities are cited. The model can be represented at different resolutions, with some parts rigid while others are flexible. A model can also specify multiple states of the system as well as a single state. In addition to spatiotemporal models, IMP has been used to compute models of molecular networks by satisfaction of network restraints (121). The agreement between input information and a model is quantified by a scoring function, which is constructed with the aid of a library of different functional forms (e.g., harmonic and harmonic upper bound) that can operate on different spatial features of the modeled system (e.g., positions, distances, and shape), based on a variety of input information. In ambiguous restraints, the assignment of a restraint to specific particles can itself be a variable, as is needed for chemical cross-links in a system with multiple copies of the cross-linked protein. Increasingly, the scoring function terms correspond to Bayesian data likelihoods. Data-based restraints can be supplemented by a molecular mechanics force field, homology-derived restraints, and various statistical potentials. A number of schemes for searching for good-scoring models are available, including local refinement methods and stochastic sampling methods in the continuous space of Cartesian coordinates as well as enumeration schemes in the discrete spaces of both Cartesian and internal coordinates. Finally, IMP implements protocols for analyzing and validating models, including methods for model clustering based on their structural similarity (4, 5, 9, 17, 19, 29, 33, 34, 36, 45, 66, 68, 86, 121–135). IMP, *Integrative Modeling Platform*.

facilitate integrative structure modeling. Another distinction is an increasingly Bayesian perspective on uncertainties in input information, model representations, and scoring functions.

Integrative Modeling Platform

The development of IMP has been guided by the requirement for general, flexible, and modular software for integrative modeling. To maximize the applicability of integrative modeling to new problems, the software needs to be readily available, well documented, supported, and easy to learn. IMP was designed to appeal to both method developers, who can contribute their own IMP modules while benefiting from the rest of IMP, and users, who prefer simple access to the existing IMP functionality. A key design feature of IMP is that it allows mixing and matching of different model representations, scoring functions, and model sampling schemes, followed by executing various protocols for model validation and analysis. This modularity has greatly facilitated our application of IMP to difficult modeling problems, where multiple iterations through defining model representation and spatial restraints, performing structural sampling, and analyzing the results are needed. IMP has been used mainly for computing architectures of macromolecular complexes by assembling subunits of known structure based on 3DEM maps and chemical cross-links, as exemplified by the integrative structure determination of the NPC (above).

Challenges in integrative modeling

There are a number of interrelated challenges in improving integrative modeling. Addressing these challenges will improve the accuracy and precision of integrative models as well as expand the types of systems that can be modeled to include larger and more complex, heterogeneous, and dynamic systems. We briefly outline several challenges as follows.

Optimizing model representation

In difficult modeling cases, model representation is the first and possibly most important modeling decision. The choice of model representation generally balances the accuracy of translating input information into restraints on models with the efficiency of sampling alternative models (e.g., rigid *versus* flexible subunit structure, single *versus* multiple state models). For example, when only chemical cross-links between pairs of residues are available to guide the fitting of subunits into a medium-resolution 3DEM map, it is often sufficient to represent the individual subunits as rigid bodies at a single residue-per-bead resolution, as opposed to flexible structures at atomic resolution. The representation is generally selected ad hoc, based on experience. Therefore, objective methods for selecting an optimal representation given available input information and sampling power are needed (63). Such methods could in principle be guided by Bayesian model selection or even Bayesian model averaging (64, 65).

Expanding the variety of model representations

Model representation determines the type of model that is computed. Thus, expanding the set of possible representations

of integrative models expands the scope of integrative modeling. Currently, atomic and various coarse-grained models of systems in single and a few discrete static states can be produced relatively routinely (Fig. 3). In contrast, much work remains to be done to compute integrative models of heterogeneous systems existing in many states (61, 66). Similarly uncharted is integrative modeling of spatiotemporal processes, so that the resulting spatiotemporal models are by construction consistent with available experimental data (4, 47). In contrast, various flavors of molecular dynamics simulations compute models based on physical principles alone, followed by validation of the resulting trajectories based on experimental data only after simulation. Additional opportunities for expansion of integrative modeling include modeling of molecular and other types of networks (e.g., metabolic and signaling networks), energy landscapes (defining structures, thermodynamics, and kinetics), reaction fluxes, and diffusion processes. For systems whose salient features include both structural and network degrees of freedom, such as spatially organized metabolic networks (e.g., polyketide synthases (67)), it may be especially attractive to couple structural and network variables into a single unified representation informed by both structural and network data, thus hopefully maximizing the accuracy, precision, and completeness of the resulting models.

Expanding the variety of input data and prior information

Increasing the variety of input data and prior information used for modeling will expand the scope of structural biology. For example, an upper bound on a distance between two point mutations in a protein complex can be inferred from quantitative genetic interaction mapping of point mutations in the background of gene knockouts (68). Such data can be obtained by measuring the size of yeast colonies across a set of mutant strains, therefore allowing structure characterization without purifying a sample for traditional structural biology methods. Likewise, protocols need to be developed for including into integrative modeling other types of data, such as electron and X-ray tomography maps, *in vivo* chemical cross-links, and super-resolution optical microscopy images. Including some of these data may require developing new model representations first.

Bayesian scoring function

Quantifying a match between input information and a model in a fully Bayesian fashion increases the accuracy and optimizes the precision of the resulting models (31). A Bayesian posterior model density is proportional to the individual data likelihoods and priors (5, 69). Data likelihood quantifies the probability of observing the data given the model (or some of its aspects), whereas a prior quantifies the prior probability of the model (or some of its aspects). Thus, a Bayesian scoring function also facilitates the modularity of integrative modeling development, as different data likelihoods and priors for separate sources of information can be independently developed by experts in relevant fields and then simply multiplied to obtain the complete posterior model density.

Model sampling

It is necessary that all models consistent with input information are found, not only a subset of them (*i.e.*, overfitting must be avoided) (37). Sufficient sampling, as opposed to optimization, is especially required for analyzing a Bayesian posterior model density (70). If a sufficient model sample is in hand, the uncertainty of a model can be directly estimated based on the variability in the sample (9). Thus, efficient structural sampling algorithms and computing hardware are needed.

Model analysis

Once a model is in hand, it needs to be validated and interpreted. Although rudimentary protocols for estimating the uncertainty of integrative structure models do exist (4), it is desirable that not only the scoring function but also the validation of an integrative structure is formulated in a Bayesian fashion (10, 11). Of importance, such a formulation may facilitate deconvoluting total model uncertainty into uncertainty arising from lack of information *versus* actual heterogeneity in the samples used to collect the data. In addition, interpretation of any model, and especially a structural model, is generally facilitated by its visualization. Because the representation of integrative structures is often nonatomic, new visualization programs, such as ChimeraX (71) and Mol* (72), were developed. These and other programs need to be continually updated to be able to handle the growing variety of integrative models and the data on which they are based.

Integrative structural biology community is nucleated by the wwPDB

Protein Data Bank

In a visionary achievement, the PDB was founded in 1971 when few protein structures were available (73), followed by the creation of a federation of the US, European, and Japanese sites in 2003 (7). Ever since its founding, the PDB has catalyzed and shaped structural biology. The impact of a freely accessible and comprehensive repository for all published biomolecular structures and corresponding experimental data cannot be overstated. In addition to archival and dissemination, wwPDB has also established *de facto* standards for validation and publication of X-ray (74), NMR (75), 3DEM (76), SAS (77, 78), and integrative structures and data (10, 11) as well as comparative protein structure models (79, 80). In the process, the PDB has vastly improved the efficiency, quality, and scope of structural biology. For example, it is inconceivable to imagine the field of structural bioinformatics without the PDB. The PDB inspired many new methods for analysis and prediction of protein structures, provided input data for these methods, and thus facilitated the discovery of numerous protein sequence–structure–function principles; for example, the PDB provides inputs for sequence and structure comparison, studying the impact of sequence and structure changes on function, computing sidechain rotamer libraries and other statistical potentials, threading sequences through structures, comparative modeling, molecular docking, and integrative modeling. In a positive feedback loop, some of these

advancements, in turn, facilitated improving methods for structure determination, contributing to additional growth of the PDB and increased accuracy of biomolecular structures. Moreover, the impact of the PDB extends beyond structural biology, as it facilitates using structural information in other fields as well, including perhaps most importantly in cell biology and drug discovery (81). Finally, the PDB is a beacon for community-building efforts in any field, clearly illustrating the beneficial impact of a freely accessible and comprehensive resource for the data and models defining the field. Indeed, our vision for mapping the cell is based in part on this insight (below).

PDB for integrative modeling

The wwPDB has played several important roles in the development of the integrative structural biology community. The leadership of wwPDB organized and catalyzed meetings of scientists interested in various aspects of integrative structural biology, focusing on representation, validation, archival, and dissemination of integrative structure models and data (10, 11). At the first meeting in 2014, the participants agreed to establish an Integrative/Hybrid Methods (IHM) Task Force to work on a common set of evolving standards (11). Recommendations on several key issues were made, followed by their implementation that culminated in the nascent PDB-Development (PDB-Dev) archive for integrative structures and data (<https://pdb-dev.wwpdb.org>) (82, 83). PDB-Dev is a standalone prototype system for collecting, curating, validating, archiving, and disseminating integrative structures and data. To facilitate an agile development platform, PDB-Dev is implemented separately from the PDB, with the plan to integrate it into PDB in the coming years. This integration will be achieved by making the wwPDB OneDep deposition system and other aspects of the PDB data pipeline (84) fully supportive of integrative structures and data. PDB-Dev is enabled by an expanded set of data standards for representing integrative structures and data; a software library that supports these data standards; a data harvesting system for collecting heterogeneous data from diverse experimental techniques; methods for curating, validating, and visualizing integrative structures; and web services for distributing archived data. Through the PDB-Dev website, scientists interested in integrative structures can search and retrieve archived structures, visualize multiscale and multistate structures, gather information regarding the input data and methods used in modeling, and download the archived data for further research. PDB-Dev already contains 55 entries as of February 1, 2021, even though the number of depositions has not been maximized during the current development stage. Integrative structures can be presently deposited only in PDB-Dev, not PDB. One reason is that PDB-Dev offers a much richer set of model representations than the standard atomic representation used by the PDB. The editors of scientific journals are increasingly requesting that authors deposit their integrative structures and data into PDB-Dev, for the benefit of the authors and the community. Next, we discuss the representation,

validation, and archival of integrative structures in PDB-Dev in more detail.

Model representation

Traditional structural biology methods usually produce a single atomic coordinate set. In contrast, integrative models tend to be more complex in at least four respects (11). First, a model can be multiscale, coarse-graining different levels of structural detail by a collection of geometrical primitives (e.g., points, spheres, tubes, Gaussians, and probability densities) (85). Thus, the same part of the system can be described with multiple representations and different parts of the system can be represented differently. Second, a model can be multistate, specifying multiple discrete states of the system that are needed simultaneously to explain the input information (each state might differ in structure and/or composition) (86, 87). Third, a model can also specify the order of states. This feature allows a representation of a multistep biological process, a functional cycle (88), a kinetic network (89), or a time evolution of a modeled system (e.g., a molecular dynamics trajectory) (90). 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. 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 (21, 28). Thus, the generalized representation allows us to encode an ensemble of multiscale, multistate, and ordered models (82, 83). This expanded molecular representation is implemented *via* an extension of the PDBx/mmCIF dictionary (82, 91).

Model validation

Assessment of both an integrative structure and the data on which it is based is of critical importance for guiding structural interpretation. This assessment is a major research challenge owing to the diverse types of input experimental data and computational methods used in integrative modeling. Correspondingly, the IHM Task Force put forward recommendations for creating methods to validate integrative structures (10, 11). There are currently four main categories of assessment: (i) quantifying the quality of experimental data used to compute and/or assess an integrative structure; (ii) application of extant PDB criteria to assess atomic integrative structures (74); (iii) evaluating the fit of a structure to experimental data and other information, whether or not this information was used to compute the structure; and (iv) estimating the uncertainty (precision) of the structure. These recommendations are being pursued by the PDB-Dev team, with the benefit of contributions from members of the wwPDB IHM Task Force and others. For example, proposed experimental data quality criteria are based on the respective community practices (92–94) and a number of model validation criteria are taken from IMP (4, 9). The validation pipeline will also leverage existing software developed by

the structural biology community (e.g., wwPDB (95), MolProbity (96), BMRB (97), EMDB (98–100), SASBDB (101), PHENIX (52), and PDBStat (102)). Standardized validation of integrative structures will ultimately be part of deposition into the PDB, as is already the case for structures derived using traditional methods (74–78, 95). The validation report will facilitate reviewing, publishing, disseminating, and using the results of integrative structural biology studies.

Federated archive

To support the archival, validation, and dissemination of integrative structures and data, wwPDB initiated a federated network of interoperating structural biology data resources (Fig. 5) (103, 104), as recommended by the IHM Task Force (10, 11). This effort involves collaboration with teams working on other repositories of experimental data, such as BMRB for NMR data (97), EMDB for EM data (98), SASBDB for SAS data (101), and PRIDE for MS data (105). The goal is to create standards and tools for automated data exchange between the PDB and member repositories. Discussions with experts in additional methods, including Foerster resonance energy transfer spectroscopy and microscopy (106), hydrogen/deuterium exchange by MS, and genome modeling, are in progress. More are expected in the future. Thus, integrative structural biology efforts are expanding the structural biology community by connecting it with other communities generating the types of data not previously used for determining protein

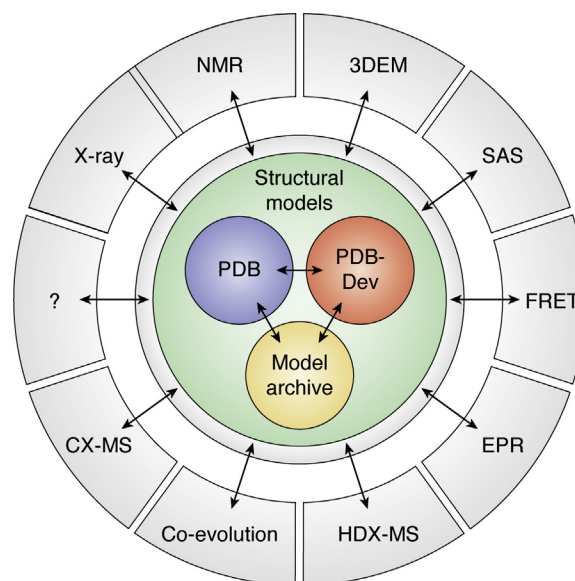


Figure 5. Federated wwPDB archive for integrative structures and data. The scheme (courtesy of Helen Berman) shows the current organization of the archive. The atomic structures produced by traditional methods are stored in the PDB, integrative structures in the PDB-Dev, and comparative protein structure models in the Model Archive (79). The data on which the structures are based are stored in separate databases, often constructed by experts in the community that is generating the data. All resources are interlinked with each other. Many additional data archives are expected to be added in the future, reflecting and catalyzing progress in integrative structural biology. 3DEM, three-dimensional electron microscopy; EPR, electron paramagnetic resonance; FRET, Foerster resonance energy transfer; PDB-Dev, PDB-Development; SAS, small-angle scattering; wwPDB, Worldwide Protein Data Bank.

structures. Conversely, these other communities stand to benefit more directly from structural biology than has been the case so far. In a reductionist view of cell biology, biomolecular structures underlie all systems and processes studied in cell biology. Thus, it is fitting that a structural archive nucleates connections with archives that store other types of data collected in cell biology. The federated wwPDB archive will advance scientific research by promoting efficient data sharing and making research data easily accessible to everyone, impacting both structural biologists and users of structural biology data, including cell biologists and others.

Software development

wwPDB may also play an important role in the development of software for integrative modeling in at least two ways. First, wwPDB could provide benchmarks for new modeling methods, including experimental datasets, reference structures, and criteria for assessing the quality of the methods, perhaps in conjunction with other community benchmarking efforts, such as EM Challenge (http://challenges.emdatabank.org/?q=model_challenge) and Critical Assessment of Protein Structure Prediction (107). Second, wwPDB could catalyze standard interfaces between key molecular modeling operations. Such standards would facilitate mixing and matching of functionalities across molecular modeling programs, not only within one program (*cf.*, Fig. 4). As a result, the efficiency, quality, and scope of structural studies would be improved.

Taking integrative modeling to the next level: Metamodeling of the cell

Cell modeling

Modeling the cell (108) is one of the grand challenges in cell biology (12). Most of the current cell modeling approaches rely on a single type of representation of the cell; for example, spatiotemporal (109), ordinary differential equation (110), and flux balance analysis representations (111). In addition to cell models, there are a myriad of models of different parts of the cell, too numerous to review here. These models may provide a useful starting point for cell modeling, owing to their encoding of expertise, data, and computing used to produce them. However, no general approach yet exists for combining different kinds of models, although steps in this direction have been made (112–115).

Cell model

Despite these efforts, it is not yet clear what a cell model should look like. More precisely, it is not yet clear what the representation of the cell model should be. Nevertheless, it is likely that such a model will need to have a number of desirable attributes, as follows. First, the cell is organized spatially and changes with time; thus, the cell model should be spatiotemporal, defining the positions of its parts in space and time. Second, the cell is modular and hierarchical, with the hierarchy spanning atoms, molecules, organelles, and the cell; thus, the model should be multiscale, including all representations that are useful for reflecting available information onto

the model and for interpreting the model, including nonspatial representations such as molecular networks represented by graphs. Third, the depictions of the cell using these varied representations should clearly be in harmony with each other; for example, when two molecules are interacting in a molecular network representation of a signaling pathway in the cytosol, they should also have sufficient propensity to be in proximity of each other in a Brownian dynamics simulation of the cytosol and *vice versa*. Fourth, the model needs to be as useful as possible; thus, it needs to be maximally accurate, precise, complete, and general. Fifth, the cell is a complex object with many parts and degrees of freedom, requiring a large amount of information to specify; thus, the model should be integrative, based on all available information. Sixth, the model will be imprecise; thus, its uncertainty should be specified. Seventh, in addition to being able to rationalize known facts, the model should also be testable and useful in guiding future experiments; thus, it should allow predicting outcomes of future experiments. Finally, the model will require a large community effort on data collection over a long period of time, in turn requiring iterative and objective improvement of the model; thus, it should be computed from input information automatically and efficiently. A model with all these attributes would clearly be of immense value to both cell biology and drug discovery.

Integrative modeling for cell mapping

It is tempting to adopt integrative modeling for cell modeling. In a brute force approach, model representation would be expanded to include all degrees of freedom of interest, informed by all experimental data and prior information. In practice, however, such a generalization is clearly unlikely to succeed any time soon, owing to insufficient data, prior information, and computing power as well as limitations of existing integrative modeling methods. Thus, substantive development of integrative modeling is needed. We outline one such recent development here, called metamodeling (13).

Metamodeling

Metamodeling is a divide-and-conquer modeling approach that aims to integrate varied input models into a metamodel (Fig. 6). Thus, metamodeling can be seen as a special case of integrative modeling in which the focus is on integrating prior models instead of data. The large problem of computing an integrative model of the cell is broken into a number of smaller modeling problems corresponding to computing models of some aspects of some parts of the cell. Each such input model may be informed by different subsets of available data, relying on its distinct model representation at any scale and level of granularity. Metamodeling then proceeds by assembling and harmonizing the input models into a complete map of the cell. A simple example of metamodeling is the flexible docking of two protein structures into a complex. In this example, the input models are the two apo structures that are refined and juxtaposed by docking (metamodeling) to model the induced fit and the final binding mode (metamodel). Although

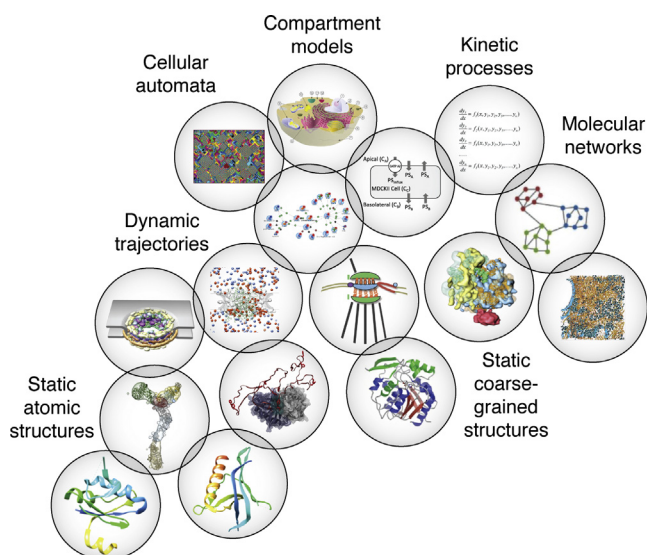


Figure 6. Metamodeling. The scheme conveys the aim of metamodeling, which is to combine input models of different types (*circles*) to obtain a metamodel. The process involves harmonizing input models with each other, as indicated by the overlaps between the circles. Different types of models shown in the scheme include static structures of biomolecules at atomic and coarse-grained resolutions, dynamics of biomolecules represented by molecular dynamics trajectories, processes involving biomolecules represented by Brownian dynamics trajectories, molecular networks represented by graphs, density maps, diffusion, kinetic processes quantified by systems of ordinary differential equations, a cellular automaton, and a compartment model of the cell. The coupling between models, which allows for their harmonization, is indicated by overlapping *circles*.

combining the two input models in this special case is familiar, combining models of different types is generally challenging.

Bayesian metamodeling

Bayesian metamodeling is a specific implementation of metamodeling in which input models are harmonized through a Bayesian statistical model of their relations with each other and/or the physical world. This Bayesian approach enables us to update our “beliefs” in the distribution of model variables (including best single-value estimates and their uncertainties), given information provided by all input models. Bayesian metamodeling proceeds through the following three stages: the input models are (i) converted to a standardized statistical representation relying on Probabilistic Graphical Models, (ii) coupled by modeling their mutual relations with the physical world, and (iii) finally harmonized with respect to each other *via* backpropagation. Bayesian metamodeling was illustrated by a proof-of-principle metamodel of glucose-stimulated insulin secretion by human pancreatic β -cells. The input models included a coarse-grained spatiotemporal simulation of insulin vesicle trafficking, docking, and exocytosis; a molecular network model of glucose-stimulated insulin secretion signaling; a network model of insulin metabolism; a structural model of glucagon-like peptide-1 receptor activation; a linear model of a pancreatic cell population; and ordinary differential equations for systemic postprandial insulin response. The coupling of the input models increased their accuracy and reproduced the behavior of the β -cell not represented by any individual input model, such as the incretin effect.

Metamodeling facilitates community collaboration

Metamodeling benefits from decentralized computing, while often producing a more accurate, precise, and complete model that contextualizes input models as well as resolves conflicting information. By shifting the focus from data integration to model integration, metamodeling facilitates the sharing of data, computational resources, expertise in diverse fields, and already existing models of the cell and its parts. Thus, by construction, metamodeling may be more useful for the large effort of cell mapping than centralized approaches to data integration. At its core, metamodeling is rooted in collaboration and appreciation for the details of disparate data, methods, and models, which cannot be achieved by any individual scientist, research group, or institution. Acting on this premise, the Pancreatic β -Cell Consortium (12) is creating cyberinfrastructure for archiving and disseminating experimental data and models that will include metamodeling (Fig. 7). Thus, each time a data set is determined or a model is computed, the metamodel can be updated automatically. Together with the many community databases of data and models (7, 111, 116–118), metamodeling may help create an effective sociotechnical ecosystem that is no doubt required for successful mapping of the cell.

Conclusions

Integrative modeling is a general approach for computing any kind of a model based on varied types of data and prior information. It can already produce structures of biomolecules that are recalcitrant to traditional structural biology methods, as exemplified by the integrative structure of the NPC. Moreover, its applicability to large, complex, and dynamic systems is increasing, as current shortcomings are being addressed. wwPDB activities, including leadership and service, are contributing significantly to the integrative structural biology community. For example, wwPDB maximizes the impact of integrative structural biology on cell biology by archiving and disseminating integrative structures and the data on which they are based. A special case of integrative modeling, called metamodeling, aims to combine multiple types of models instead of multiple datasets into an output model, potentially providing a practical means toward mapping the entire cell. With the progress of integrative structural

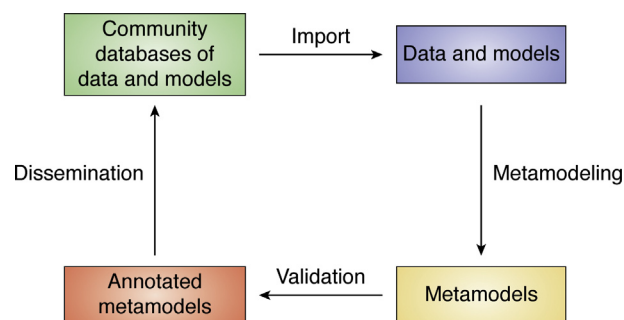


Figure 7. Vision for the contribution of metamodeling to mapping the cell. The scheme illustrates how iterative metamodeling could contribute toward the mapping of the cell.

biology, many more cell biologists will also become structural biologists, for their own gain.

Acknowledgments—I am grateful to Helen Berman for her energy and inspiration about community-oriented efforts in structural biology. I acknowledge the long-term collaboration with Michael P. Rout and Brian Chait on structural and functional characterization of the NPC, which challenged and demonstrated our development of integrative modeling. I am also grateful to the leaders of the wwPDB and the members of the Integrative Methods Task Force of wwPDB for their support of PDB-Dev. I very much appreciate the forward-looking collaboration on cell mapping with Raymond C. Stevens, Kate White, Helen Berman, Jitin Singla, Brinda Vallat, Carl Kesselman, Barak Raveh, and Liping Sun in the context of the Pancreatic β -Cell Consortium. I am grateful to Yekaterina Kadyshevskaya for contributing a panel in Figure 1 and to Helen Berman for Figure 5. I thank Jitin Singla, Brinda Vallat, Seth Axen, and Stephen K. Burley for comments on the manuscript. Finally, I am most grateful to past and present members of our research group at UCSF who contributed to the development and applications of IMP, in particular, Ben Webb who has been curating the software from its beginnings in 2007 and also helped with editing the manuscript. This review is based on refs. (5, 10, 13).

Funding and additional information—Current key grants funding our work include NIGMS, National Institutes of Health R01GM083960; NIGMS, National Institutes of Health P41GM109824; NIGMS, National Institutes of Health P50AI150476; NIAID, National Institutes of Health U19AI135990; NIA, National Institutes of Health P01AG002132; NIGMS, National Institutes of Health P01GM118303; NIDDK, National Institutes of Health U54DK107981; and National Science Foundation DBI-1756250. The RCSB PDB is jointly funded by National Science Foundation DBI-1832184, US Department of Energy DE-SC0019749, and NIGMS/NIAID/NCI, National Institutes of Health R01GM133198. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflict of interest—The author declares that he has no conflicts of interest with the contents of this article.

Abbreviations—The abbreviations used are: 3DEM, three-dimensional electron microscopy; IHM, Integrative/Hybrid Methods; IMP, *Integrative Modeling Platform*; NPC, Nuclear Pore Complex; PDB, Protein Data Bank; PDB-Dev, PDB-Development; SAS, small-angle scattering; wwPDB, Worldwide Protein Data Bank.

References

- Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R. G., Wyckoff, H., and Phillips, D. C. (1958) A three-dimensional model of the myoglobin molecule obtained by X-ray analysis. *Nature* **181**, 662–666
- Kendrew, J. C., Dickerson, R. E., Strandberg, B. E., Hart, R. G., Davies, D. R., Phillips, D. C., and Shore, V. C. (1960) Structure of myoglobin: A three-dimensional Fourier synthesis at 2 Å resolution. *Nature* **185**, 422–427
- Protein Data Bank: The single global archive for 3D macromolecular structure data. *Nucleic Acids Res.* **47**, (2019), D520–D528
- Kim, S. J., Fernandez-Martinez, J., Nudelman, I., Shi, Y., Zhang, W., Raveh, B., Herricks, T., Slaughter, B. D., Hogan, J. A., Upla, P., Chemmama, I. E., Pellarin, R., Echeverria, I., Shivaraju, M., Chaudhury, A. S., et al. (2018) Integrative structure and functional anatomy of a nuclear pore complex. *Nature* **555**, 475–482
- Rout, M. P., and Sali, A. (2019) Principles for integrative structural biology studies. *Cell* **177**, 1384–1403
- Sali, A., Glaeser, R., Earnest, T., and Baumeister, W. (2003) From words to literature in structural proteomics. *Nature* **422**, 216–225
- Berman, H., Henrick, K., and Nakamura, H. (2003) Announcing the worldwide Protein Data Bank. *Nat. Struct. Biol.* **10**, 980
- 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
- Saltzberg, D. J., Viswanath, S., Echeverria, I., Chemmama, I. E., Webb, B., and Sali, A. (2021) Using Integrative Modeling Platform to compute, validate, and archive a model of a protein complex structure. *Protein Sci.* **30**, 250–261
- Berman, H. M., Adams, P. D., Bonvin, A. A., Burley, S. K., Carragher, B., Chiu, W., DiMaio, F., Ferrin, T. E., Gabanyi, M. J., Goddard, T. D., Griffin, P. R., Haas, J., Hanke, C. A., Hoch, J. C., Hummer, G., et al. (2019) Federating structural models and data: Outcomes from a workshop on archiving integrative structures. *Structure* **27**, 1745–1759
- Sali, A., Berman, H. M., Schwede, T., Trewhella, J., Kleywegt, G., Burley, S. K., Markley, J., Nakamura, H., Adams, P., Bonvin, A. M., Chiu, W., Peraro, M. D., Di Maio, F., Ferrin, T. E., Grünwald, K., et al. (2015) Outcome of the first wwPDB hybrid/integrative methods task force workshop. *Structure* **23**, 1156–1167
- Singla, J., McClary, K. M., White, K. L., Alber, F., Sali, A., and Stevens, R. C. (2018) Opportunities and challenges in building a spatiotemporal multi-scale model of the human pancreatic β -cell. *Cell* **173**, 11–19
- Raveh, B., Sun, L., White, K. L., Sanyal, T., Tempkin, J., Zheng, D., Bharat, K., Singla, J., Wang, C., Zhao, J., Li, A., Graham, N. A., Kesselman, C., Stevens, R. C., and Sali, A. B. Bayesian metamodeling of complex biological systems across varying representations. *Proc. Natl. Acad. Sci. U. S. A.*, In revision.
- Watson, J. D., and Crick, F. H. (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171**, 737–738
- Rayment, I., Holden, H. M., Whittaker, M., Yohn, C. B., Lorenz, M., Holmes, K. C., and Milligan, R. A. (1993) Structure of the actin-myosin complex and its implications for muscle contraction. *Science* **261**, 58–65
- Sali, A., Overington, J. P., Johnson, M. S., and Blundell, T. L. (1990) From comparisons of protein sequences and structures to protein modelling and design. *Trends Biochem. Sci.* **15**, 235–240
- Alber, F., Dokudovskaya, S., Veenhoff, L. M., Zhang, W., Kipper, J., Devos, D., Suprpto, A., Karni-Schmidt, O., Williams, R., Chait, B. T., Rout, M. P., and Sali, A. (2007) Determining the architectures of macromolecular assemblies. *Nature* **450**, 683–694
- Robinson, C. V., Sali, A., and Baumeister, W. (2007) The molecular sociology of the cell. *Nature* **450**, 973–982
- Alber, F., Forster, F., Korke, D., Topf, M., and Sali, A. (2008) Integrating diverse data for structure determination of macromolecular assemblies. *Annu. Rev. Biochem.* **77**, 443–477
- Ward, A. B., Sali, A., and Wilson, I. A. (2013) Integrative structural biology. *Science* **339**, 913–915
- Schneidman-Duhovny, D., Pellarin, R., and Sali, A. (2014) Uncertainty in integrative structural modeling. *Curr. Opin. Struct. Biol.* **28**, 96–104
- Braitbard, M., Schneidman-Duhovny, D., and Kalisman, N. (2019) Integrative structure modeling: Overview and assessment. *Annu. Rev. Biochem.* **88**, 113–135
- Koukos, P. I., and Bonvin, A. M. J. J. (2020) Integrative modelling of biomolecular complexes. *J. Mol. Biol.* **432**, 2861–2881
- Srivastava, A., Tiwari, S. P., Miyashita, O., and Tama, F. (2020) Integrative/hybrid modeling approaches for studying biomolecules. *J. Mol. Biol.* **432**, 2846–2860
- Kaptein, R., and Wagner, G. (2019) Integrative methods in structural biology. *J. Biomol. NMR* **73**, 261–263
- Ziegler, S. J., Mallinson, S. J. B., St John, P. C., and Bomble, Y. J. (2021) Advances in integrative structural biology: Towards understanding

- protein complexes in their cellular context. *Comput. Struct. Biotechnol. J.* **19**, 214–225
27. Cerofolini, L., Fragai, M., Ravera, E., Diebolder, C. A., Renault, L., and Calderone, V. (2019) Integrative approaches in structural biology: A more complete picture from the combination of individual techniques. *Biomolecules* **9**, 370
 28. Schroder, G. F. (2015) Hybrid methods for macromolecular structure determination: Experiment with expectations. *Curr. Opin. Struct. Biol.* **31**, 20–27
 29. 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
 30. Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953) Equation of state calculations by fast computing machines. *J. Chem. Phys.* **21**, 1087–1092
 31. Rieping, W., Habeck, M., and Nilges, M. (2005) Inferential structure determination. *Science* **309**, 303–306
 32. Swendsen, R. H., and Wang, J. S. (1986) Replica Monte Carlo simulation of spin glasses. *Phys. Rev. Lett.* **57**, 2607–2609
 33. Shi, Y., Fernandez-Martinez, J., Tjioe, E., Pellarin, R., Kim, S. J., Williams, R., Schneidman-Duhovny, D., Sali, A., Rout, M. P., and Chait, B. T. (2014) Structural characterization by cross-linking reveals the detailed architecture of a coatomer-related heptameric module from the nuclear pore complex. *Mol. Cell. Proteomics* **13**, 2927–2943
 34. Fernandez-Martinez, J., Kim, S. J., Shi, Y., Upla, P., Pellarin, R., Gagnon, M., Chemmama, I. E., Wang, J., Nudelman, I., Zhang, W., Williams, R., Rice, W. J., Stokes, D. L., Zenklusen, D., Chait, B. T., et al. (2016) Structure and function of the nuclear pore complex cytoplasmic mRNA export platform. *Cell* **167**, 1215–1228.e25
 35. Fernandez-Martinez, J., Phillips, J., Sekedat, M. D., Diaz-Avalos, R., Velazquez-Muriel, J., Franke, J. D., Williams, R., Stokes, D. L., Chait, B. T., Sali, A., and Rout, M. P. (2012) Structure-function mapping of a heptameric module in the nuclear pore complex. *J. Cell Biol.* **196**, 419–434
 36. Velazquez-Muriel, J., Lasker, K., Russel, D., Phillips, J., Webb, B. M., 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
 37. Viswanath, S., Chemmama, I. E., Cimermancic, P., and Sali, A. (2017) Assessing exhaustiveness of stochastic sampling for integrative modeling of macromolecular structures. *Biophys. J.* **113**, 2344–2353
 38. Alber, F., Dokudovskaya, S., Veenhoff, L. M., Zhang, W., Kipper, J., Devos, D., Suprpto, A., Karni-Schmidt, O., Williams, R., Chait, B. T., Sali, A., and Rout, M. P. (2007) The molecular architecture of the nuclear pore complex. *Nature* **450**, 695–701
 39. Mosalaganti, S., Kosinski, J., Albert, S., Schaffer, M., Strenkert, D., Salome, P. A., Merchant, S. S., Plitzko, J. M., Baumeister, W., Engel, B. D., and Beck, M. (2018) In situ architecture of the algal nuclear pore complex. *Nat. Commun.* **9**, 2361
 40. Kosinski, J., Mosalaganti, S., von Appen, A., Teimer, R., DiGiulio, A. L., Wan, W., Bui, K. H., Hagen, W. J., Briggs, J. A., Glavy, J. S., Hurt, E., and Beck, M. (2016) Molecular architecture of the inner ring scaffold of the human nuclear pore complex. *Science* **352**, 363–365
 41. Allegretti, M., Zimmerli, C. E., Rantos, V., Wilfling, F., Ronchi, P., Fung, H. K. H., Lee, C. W., Hagen, W., Turoňová, B., Karius, K., Börmel, M., Zhang, X., Müller, C. W., Schwab, Y., Mahamid, J., et al. (2020) In-cell architecture of the nuclear pore and snapshots of its turnover. *Nature* **586**, 796–800
 42. Eibauer, M., Pellanda, M., Turgay, Y., Dubrovsky, A., Wild, A., and Medalia, O. (2015) Structure and gating of the nuclear pore complex. *Nat. Commun.* **6**, 7532
 43. Zimmerli, C. E., Allegretti, M., Rantos, V., Goetz, S. K., Obarska-Kosinska, A., Zagoriy, I., Halavatyi, A., Mahamid, J., Kosinski, J., and Beck, M. (2020) *Nuclear Pores Constrict Upon Energy Depletion*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
 44. Algret, R., Fernandez-Martinez, J., Shi, Y., Kim, S. J., Pellarin, R., Cimermancic, P., Cochet, E., Sali, A., Chait, B. T., Rout, M. P., and Dokudovskaya, S. (2014) Molecular architecture and function of the SEA complex - a modulator of the TORC1 pathway. *Mol. Cell. Proteomics* **13**, 2855–2870
 45. Viswanath, S., Bonomi, M., Kim, S. J., Klenchin, V. A., Taylor, K. C., Yabut, K. C., Umbreit, N. T., Van Epps, H. A., Meehl, J., Jones, M. H., Russel, D., Velazquez-Muriel, J. A., Winey, M., Rayment, I., Davis, T. N., et al. (2017) The molecular architecture of the yeast spindle pole body core determined by Bayesian integrative modeling. *Mol. Biol. Cell* **28**, 3298–3314
 46. Raveh, B., Karp, J. M., Sparks, S., Dutta, K., Rout, M. P., Sali, A., and Cowburn, D. (2016) Slide-and-exchange mechanism for rapid and selective transport through the nuclear pore complex. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E2489–E2497
 47. Timney, B. L., Raveh, B., Mironska, R., Trivedi, J. M., Kim, S. J., Russel, D., Wentz, S. R., Sali, A., and Rout, M. P. (2016) Simple rules for passive diffusion through the nuclear pore complex. *J. Cell Biol.* **215**, 57–76
 48. Dominguez, C., Boelens, R., and Bonvin, A. M. (2003) HADDOCK: A protein-protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.* **125**, 1731–1737
 49. van Zundert, G. C. P., Rodrigues, J. P. G. L. M., Trellet, M., Schmitz, C., Kastriitis, P. L., Karaca, E., Melquiond, A. S. J., van Dijk, M., de Vries, S. J., and Bonvin, A. M. J. J. (2016) The HADDOCK2.2 web server: User-friendly integrative modeling of biomolecular complexes. *J. Mol. Biol.* **428**, 720–725
 50. Das, R., and Baker, D. (2008) Macromolecular modeling with rosetta. *Annu. Rev. Biochem.* **77**, 363–382
 51. Leaver-Fay, A., Tyka, M., Lewis, S. M., Lange, O. F., Thompson, J., Jacak, R., Kaufman, K., Renfrew, P. D., Smith, C. A., Sheffler, W., Davis, I. W., Cooper, S., Treuille, A., Mandell, D. J., Richter, F., et al. (2011) ROSETTA3: An object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol.* **487**, 545–574
 52. Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., et al. (2010) Phenix: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221
 53. Karakaş, M., Woetzel, N., Staritzbichler, R., Alexander, N., Weiner, B. E., and Meiler, J. (2012) BCL::Fold - de novo prediction of complex and large protein topologies by assembly of secondary structure elements. *PLoS One* **7**, e49240
 54. Schwieters, C. D., Bermejo, G. A., and Clore, G. M. (2018) Xplor-NIH for molecular structure determination from NMR and other data sources. *Protein Sci.* **27**, 26–40
 55. Trussart, M., Serra, F., Bau, D., Junier, I., Serrano, L., and Marti-Renom, M. A. (2015) Assessing the limits of restraint-based 3D modeling of genomes and genomic domains. *Nucleic Acids Res.* **43**, 3465–3477
 56. Serra, F., Bau, D., Goodstadt, M., Castillo, D., Filion, G. J., and Marti-Renom, M. A. (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
 57. Hua, N., Tjong, H., Shin, H., Gong, K., Zhou, X. J., and Alber, F. (2018) Producing genome structure populations with the dynamic and automated PGS software. *Nat. Protoc.* **13**, 915–926
 58. Hsieh, A., Lu, L., Chance, M. R., and Yang, S. (2017) A practical guide to iSPOT modeling: An integrative structural biology platform. *Adv. Exp. Med. Biol.* **1009**, 229–238
 59. Dimura, M., Peulen, T. O., Hanke, C. A., Prakash, A., Gohlke, H., and Seidel, C. A. (2016) Quantitative FRET studies and integrative modeling unravel the structure and dynamics of biomolecular systems. *Curr. Opin. Struct. Biol.* **40**, 163–185
 60. Schneidman-Duhovny, D., Inbar, Y., Nussinov, R., and Wolfson, H. J. (2005) PatchDock and SymmDock: Servers for rigid and symmetric docking. *Nucleic Acids Res.* **33**, W363–W367
 61. Hummer, G., and Köfinger, J. (2015) Bayesian ensemble refinement by replica simulations and reweighting. *J. Chem. Phys.* **143**, 243150
 62. Saltzberg, D., Greenberg, C. H., Viswanath, S., Chemmama, I., Webb, B., Pellarin, R., Echeverria, I., and Sali, A. (2019) Modeling biological complexes using integrative modeling platform. *Methods Mol. Biol.* **2022**, 353–377

63. 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
64. Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T. (1999) Bayesian model averaging: A tutorial (with comments by M. Clyde, David Draper and E. I. George, and a rejoinder by the authors). *Stat. Sci.* **14**, 382–417
65. McElreath, R. (2018) *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*, CRC Press, Boca Raton, FL
66. Carter, L., Kim, S. J., Schneidman-Duhovny, D., Stöhr, J., Poncet-Montange, G., Weiss, T. M., Tsuruta, H., Prusiner, S. B., and Sali, A. (2015) Prion protein-antibody complexes characterized by chromatography-coupled small-angle X-ray scattering. *Biophys. J.* **109**, 793–805
67. Dutta, S., Whicher, J. R., Hansen, D. A., Hale, W. A., Chemler, J. A., Congdon, G. R., Narayan, A. R., Håkansson, K., Sherman, D. H., Smith, J. L., and Skiniotis, G. (2014) Structure of a modular polyketide synthase. *Nature* **510**, 512–517
68. Braberg, H., Echeverria, I., Bohn, S., Cimermancic, P., Shiver, A., Alexander, R., Xu, J., Shales, M., Dronamraju, R., Jiang, S., Dwivedi, G., Bogdanoff, D., Chaung, K. K., Hüttenhain, R., Wang, S., *et al.* (2020) Genetic interaction mapping informs integrative structure determination of protein complexes. *Science* **370**, eaa4910
69. Albert, J. (2020) Review of statistical rethinking: A Bayesian course with examples in R and stan, second edition, by Richard McElreath, Chapman and Hall, 2020. *J. Stat. Educ.* **28**, 248–250
70. Betancourt, M. (2019) The convergence of Markov chain Monte Carlo methods: from the metropolis method to Hamiltonian Monte Carlo. *Annalen der Physik*. <https://doi.org/10.1002/andp.201700214>
71. Goddard, T. D., Huang, C. C., Meng, E. C., Pettersen, E. F., Couch, G. S., Morris, J. H., and Ferrin, T. E. (2018) UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25
72. Sehnal, D., Rose, A., Koca, J., Burley, S., and Velankar, S. (2018) *Mol*: Towards a Common Library and Tools for Web Molecular Graphics*, The Eurographics Association, Geneva, Switzerland
73. Crystallography: Protein Data Bank. *Nat. New Biol.* **233**, (1971), 223
74. Read, R. J., Adams, P. D., Arendall, W. B., 3rd, Brunger, A. T., Emsley, P., Joosten, R. P., Kleywegt, G. J., Krissinel, E. B., Lütke, T., Otwinowski, Z., Perrakis, A., Richardson, J. S., Sheffler, W. H., Smith, J. L., Tickle, I. J., *et al.* (2011) A new generation of crystallographic validation tools for the Protein Data Bank. *Structure* **19**, 1395–1412
75. Montelione, G. T., Nilges, M., Bax, A., Guntert, P., Herrmann, T., Richardson, J. S., Schwieters, C. D., Vranken, W. F., Vuister, G. W., Wishart, D. S., Berman, H. M., Kleywegt, G. J., and Markley, J. L. (2013) Recommendations of the wwPDB NMR validation task force. *Structure* **21**, 1563–1570
76. Henderson, R., Sali, A., Baker, M. L., Carragher, B., Devkota, B., Downing, K. H., Egelman, E. H., Feng, Z., Frank, J., Grigorieff, N., Jiang, W., Ludtke, S. J., Medalia, O., Penczek, P. A., Rosenthal, P. B., *et al.* (2012) Outcome of the first electron microscopy validation task force meeting. *Structure* **20**, 205–214
77. Trehwella, J., Hendrickson, W. A., Kleywegt, G. J., Sali, A., Sato, M., Schwede, T., Svergun, D. I., Tainer, J. A., Westbrook, J., and Berman, H. M. (2013) Report of the wwPDB small-angle scattering task force: Data requirements for biomolecular modeling and the PDB. *Structure* **21**, 875–881
78. Trehwella, J., Duff, A. P., Durand, D., Gabel, F., Guss, J. M., Hendrickson, W. A., Hura, G. L., Jacques, D. A., Kirby, N. M., Kwan, A. H., Pérez, J., Pollack, L., Ryan, T. M., Sali, A., Schneidman-Duhovny, D., *et al.* (2017) 2017 publication guidelines for structural modelling of small-angle scattering data from biomolecules in solution: An update. *Acta Crystallogr. D Struct. Biol.* **73**, 710–728
79. Schwede, T., Sali, A., Honig, B., Levitt, M., Berman, H. M., Jones, D., Brenner, S. E., Burley, S. K., Das, R., Dokholyan, N. V., Dunbrack, R. L., Fidelis, K., Fiser, A., Godzik, A., Huang, Y. J., *et al.* (2009) Outcome of a workshop on applications of protein models in biomedical research. *Structure* **17**, 151–159
80. Berman, H. M., Burley, S. K., Chiu, W., Sali, A., Adzhubei, A., Bourne, P. E., Bryant, S. H., Dunbrack, R. L., Fidelis, K., Frank, J., Godzik, A., Henrick, K., Joachimiak, A., Heymann, B., Jones, D., *et al.* (2006) Outcome of a workshop on archiving structural models of biological macromolecules. *Structure* **14**, 1211–1217
81. Westbrook, J. D., and Burley, S. K. (2019) How structural biologists and the Protein Data Bank contributed to recent FDA new drug approvals. *Structure* **27**, 211–217
82. Vallat, B., Webb, B., Westbrook, J. D., Sali, A., and Berman, H. M. (2018) Development of a prototype system for archiving integrative/hybrid structure models of biological macromolecules. *Structure* **26**, 894–904. e2
83. Burley, S. K., Kurisu, G., Markley, J. L., Nakamura, H., Velankar, S., Berman, H. M., Sali, A., Schwede, T., and Trehwella, J. (2017) PDB-dev: A prototype system for depositing integrative/hybrid structural models. *Structure* **25**, 1317–1318
84. Young, J. Y., Westbrook, J. D., Feng, Z., Sala, R., Peisach, E., Oldfield, T. J., Sen, S., Gutmanas, A., Armstrong, D. R., Berrisford, J. M., Chen, L., Chen, M., Di Costanzo, L., Dimitropoulos, D., Gao, G., *et al.* (2017) OneDep: Unified wwPDB system for deposition, biocuration, and validation of macromolecular structures in the PDB archive. *Structure* **25**, 536–545
85. Grime, J. M. A., and Voth, G. A. (2014) Highly scalable and memory efficient ultra-coarse-grained molecular dynamics simulations. *J. Chem. Theory Comput.* **10**, 423–431
86. Molnar, K. S., Bonomi, M., Pellarin, R., Clinthorne, G. D., Gonzalez, G., Goldberg, S. D., Goulian, M., Sali, A., and DeGrado, W. F. (2014) Cys-scanning disulfide crosslinking and Bayesian modeling probe the transmembrane signaling mechanism of the histidine kinase, PhoQ. *Structure* **22**, 1239–1251
87. Pelikan, M., Hura, G. L., and Hammel, M. (2009) Structure and flexibility within proteins as identified through small angle X-ray scattering. *Gen. Physiol. Biophys.* **28**, 174–189
88. Diez, M., Zimmermann, B., Borsch, M., König, M., Schweinberger, E., Steigmiller, S., Reuter, R., Felekyan, S., Kudryavtsev, V., Seidel, C. A., and Gräber, P. (2004) Proton-powered subunit rotation in single membrane-bound F0F1-ATP synthase. *Nat. Struct. Mol. Biol.* **11**, 135–141
89. Pirchi, M., Ziv, G., Riven, I., Cohen, S. S., Zohar, N., Barak, Y., and Haran, G. (2011) Single-molecule fluorescence spectroscopy maps the folding landscape of a large protein. *Nat. Commun.* **2**, 493
90. Bock, L. V., Blau, C., Schroder, G. F., Davydov, I. I., Fischer, N., Stark, H., Rodnina, M. V., Vaiana, A. C., and Grubmüller, H. (2013) Energy barriers and driving forces in tRNA translocation through the ribosome. *Nat. Struct. Mol. Biol.* **20**, 1390–1396
91. Vallat, B., Webb, B., Westbrook, J., Sali, A., and Berman, H. M. (2019) Archiving and disseminating integrative structure models. *J. Biomol. NMR* **73**, 385–398
92. Leitner, A., Bonvin, A. M. J. J., Borchers, C. H., Chalkley, R. J., Chamot-Rooke, J., Combe, C. W., Cox, J., Dong, M. Q., Fischer, L., Götz, M., Gozzo, F. C., Heck, A. J. R., Hoopmann, M. R., Huang, L., Ishihama, Y., *et al.* (2020) Toward increased reliability, transparency, and accessibility in cross-linking mass spectrometry. *Structure* **28**, 1259–1268
93. Lawson, C. L., Kryshtafovych, A., Adams, P. D., Afonine, P. V., Baker, M. L., Barad, B. A., Bond, P., Burnley, T., Cao, R., Cheng, J., Chojnowski, G., Cowtan, K., Dill, K. A., DiMaio, F., Farrell, D. P., *et al.* (2020) *Outcome of the 2019 EMDDataResource Model Challenge: Validation of Cryo-EM Models at Near-Atomic Resolution*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
94. Lerner, E., Ambrose, B., Barth, A., Birkedal, V., Blanchard, S. C., Borner, R., Cordes, T., Craggs, T. D., Ha, T., Haran, G., Hugel, T., Ingargiola, A., Kapanidis, A., Lamb, D. C., Laurence, T., *et al.* (2020) The FRET-based structural dynamics challenge – community contributions to consistent and open science practices. *arXiv: Biomolecules*, 2006.03091
95. Gore, S., Sanz García, E., Hendrickx, P. M. S., Gutmanas, A., Westbrook, J. D., Yang, H., Feng, Z., Baskaran, K., Berrisford, J. M., Hudson, B. P., Ikegawa, Y., Kobayashi, N., Lawson, C. L., Mading, S., Mak, L., *et al.* (2017) Validation of structures in the Protein Data Bank. *Structure* **25**, 1916–1927
96. Williams, C. J., Headd, J. J., Moriarty, N. W., Prisant, M. G., Videau, L. L., Deis, L. N., Verma, V., Keedy, D. A., Hintze, B. J., Chen, V. B., Jain, S., Lewis, S. M., Arendall, W. B., Snoeyink, J., Adams, P. D., *et al.* (2018)

- MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci.* **27**, 293–315
97. Ulrich, E. L., Akutsu, H., Doreleijers, J. F., Harano, Y., Ioannidis, Y. E., Lin, J., Livny, M., Mading, S., Maziak, D., Miller, Z., Nakatani, E., Schulte, C. F., Tolmie, D. E., Kent Wenger, R., Yao, H., *et al.* (2008) BioMagResBank. *Nucleic Acids Res.* **36**, D402–D408
 98. Tagari, M., Newman, R., Chagoyen, M., Carazo, J. M., and Henrick, K. (2002) New electron microscopy database and deposition system. *Trends Biochem. Sci.* **27**, 589
 99. Lawson, C. L., Patwardhan, A., Baker, M. L., Hryc, C., Garcia, E. S., Hudson, B. P., Lagerstedt, I., Ludtke, S. J., Pintilie, G., Sala, R., Westbrook, J. D., Berman, H. M., Kleywegt, G. J., and Chiu, W. (2016) EMDataBank unified data resource for 3DEM. *Nucleic Acids Res.* **44**, D396–D403
 100. Patwardhan, A., and Lawson, C. L. (2016) Databases and archiving for CryoEM. *Methods Enzymol.* **579**, 393–412
 101. Valentini, E., Kikhney, A. G., Previtali, G., Jeffries, C. M., and Svergun, D. I. (2015) SASBDB, a repository for biological small-angle scattering data. *Nucleic Acids Res.* **43**, D357–D363
 102. Tejero, R., Snyder, D., Mao, B., Aramini, J. M., and Montelione, G. T. (2013) PDBStat: A universal restraint converter and restraint analysis software package for protein NMR. *J. Biomol. NMR* **56**, 337–351
 103. Berman, H. M., Trewhella, J., Vallat, B., and Westbrook, J. D. (2018) Archiving of integrative structural models. *Adv. Exp. Med. Biol.* **1105**, 261–272
 104. Berman, H. M., Lawson, C. L., Vallat, B., and Gabanyi, M. J. (2018) Anticipating innovations in structural biology. *Q. Rev. Biophys.* **51**, e8
 105. Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu, D. J., Inuganti, A., Griss, J., Mayer, G., Eisenacher, M., Pérez, E., Uszkoreit, J., Pfeuffer, J., Sachsenberg, T., Yilmaz, S., *et al.* (2019) The PRIDE database and related tools and resources in 2019: Improving support for quantification data. *Nucleic Acids Res.* **47**, D442–D450
 106. Lerner, E., Barth, A., Hendrix, J., Ambrose, B., Birkedal, V., Blanchard, S. C., Börner, R., Sung Chung, H., Cordes, T., Craggs, T. D., Deniz, A. A., Diao, J., Fei, J., Gonzalez, R. L., Gopich, I. V., *et al.* (2021) FRET-based dynamic structural biology: Challenges, perspectives and an appeal for open-science practices. *Elife* **10**, e60416
 107. Kryshchuk, A., Schwede, T., Topf, M., Fidelis, K., and Moutl, J. (2019) Critical assessment of methods of protein structure prediction (CASP)-round XIII. *Proteins* **87**, 1011–1020
 108. Sadava, D. E., Hillis, D. M., Craig Heller, H., and Berenbaum, M. (2012) *Loose-Leaf Version for Life: The Science of Biology*, W. H. Freeman, New York, NY
 109. McGuffee, S. R., and Elcock, A. H. (2010) Diffusion, crowding & protein stability in a dynamic molecular model of the bacterial cytoplasm. *PLoS Comput. Biol.* **6**, e1000694
 110. Roy, M., and Finley, S. D. (2017) Computational model predicts the effects of targeting cellular metabolism in pancreatic cancer. *Front. Physiol.* **8**, 217
 111. King, Z. A., Lu, J., Dräger, A., Miller, P., Federowicz, S., Lerman, J. A., Ebrahim, A., Palsson, B. O., and Lewis, N. E. (2016) BiGG models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Res.* **44**, D515–D522
 112. Karr, J. R., Sanghvi, J. C., Macklin, D. N., Gutschow, M. V., Jacobs, J. M., Bolival, B., Jr., Assad-Garcia, N., Glass, J. I., and Covert, M. W. (2012) A whole-cell computational model predicts phenotype from genotype. *Cell* **150**, 389–401
 113. Ghaemi, Z., Peterson, J. R., Gruebele, M., and Luthey-Schulten, Z. (2020) An in-silico human cell model reveals the influence of spatial organization on RNA splicing. *PLoS Comput. Biol.* **16**, e1007717
 114. Macklin, D. N., Ahn-Horst, T. A., Choi, H., Ruggero, N. A., Carrera, J., Mason, J. C., Sun, G., Agmon, E., DeFelicis, M. M., Maayan, I., Lane, K., Spangler, R. K., Gillies, T. E., Paull, M. L., Akhter, S., *et al.* (2020) Simultaneous cross-evaluation of heterogeneous E. coli datasets via mechanistic simulation. *Science* **369**, eaav3751
 115. Agmon, E., and Spangler, R. K. (2020) A multi-scale approach to modeling E. coli chemotaxis. *Entropy (Basel)* **22**, 1101
 116. Malik-Sheriff, R. S., Glont, M., Nguyen, T. V. N., Tiwari, K., Roberts, M. G., Xavier, A., Vu, M. T., Men, J., Maire, M., Kananathan, S., Fairbanks, E. L., Meyer, J. P., Arankalle, C., Varusai, T. M., Knight-Schrijver, V., *et al.* (2020) BioModels—15 years of sharing computational models in life science. *Nucleic Acids Res.* **48**, D407–D415
 117. Hucka, M., Finney, A., Bornstein, B. J., Keating, S. M., Shapiro, B. E., Matthews, J., Kovitz, B. L., Schilstra, M. J., Funahashi, A., Doyle, J. C., and Kitano, H. (2004) Evolving a lingua franca and associated software infrastructure for computational systems biology: The systems biology markup language (SBML) project. *Syst. Biol.* **1**, 41–53
 118. Waltemath, D., Karr, J. R., Bergmann, F. T., Chelliah, V., Hucka, M., Krantz, M., Liebermeister, W., Mendes, P., Myers, C. J., Pir, P., Alaybeyoglu, B., Aranganathan, N. K., Baghalian, K., Bittig, A. T., Burke, P. E., *et al.* (2016) Toward community standards and software for whole-cell modeling. *IEEE Trans. Biomed. Eng.* **63**, 2007–2014
 119. Kiseleva, E., Allen, T. D., Rutherford, S., Bucci, M., Wente, S. R., and Goldberg, M. W. (2004) Yeast nuclear pore complexes have a cytoplasmic ring and internal filaments. *J. Struct. Biol.* **145**, 272–288
 120. Yang, Z., Lasker, K., Schneidman-Duhovny, D., Webb, B., Huang, C. C., Pettersen, E. F., Goddard, T. D., Meng, E. C., Sali, A., and Ferrin, T. E. (2012) UCSF chimera, MODELLER, and IMP: An integrated modeling system. *J. Struct. Biol.* **179**, 269–278
 121. Calhoun, S., Korczynska, M., Wichelecki, D. J., San Francisco, B., Zhao, S., Rodionov, D. A., Vetting, M. W., Al-Obaidi, N. F., Lin, H., O'Meara, M. J., Scott, D. A., Morris, J. H., Russel, D., Almo, S. C., Osterman, A. L., *et al.* (2018) Prediction of enzymatic pathways by integrative pathway mapping. *Elife* **7**, e31097
 122. Sampathkumar, P., Gheyi, T., Miller, S. A., Bain, K. T., Dickey, M., Bonanno, J. B., Kim, S. J., Phillips, J., Pieper, U., Fernandez-Martinez, J., Franke, J. D., Martel, A., Tsuruta, H., Atwell, S., Thompson, D. A., *et al.* (2011) Structure of the C-terminal domain of Saccharomyces cerevisiae Nup133, a component of the nuclear pore complex. *Proteins* **79**, 1672–1677
 123. Schneidman-Duhovny, D., Rossi, A., Avila-Sakar, A., Kim, S. J., Velazquez-Muriel, J., Strop, P., Liang, H., Krukenberg, K. A., Liao, M., Kim, H. M., Sobhanifar, S., Dötsch, V., Rajpal, A., Pons, J., Agard, D. A., *et al.* (2012) A method for integrative structure determination of protein-protein complexes. *Bioinformatics* **28**, 3282–3289
 124. Robinson, P. J., Trnka, M. J., Pellarin, R., Greenberg, C. H., Bushnell, D. A., Davis, R., Burlingame, A. L., Sali, A., and Kornberg, R. D. (2015) Molecular architecture of the yeast mediator complex. *Elife* **4**, e08719
 125. Upla, P., Kim, S. J., Sampathkumar, P., Dutta, K., Cahill, S. M., Chemmama, I. E., Williams, R., Bonanno, J. B., Rice, W. J., Stokes, D. L., Cowburn, D., Almo, S. C., Sali, A., Rout, M. P., and Fernandez-Martinez, J. (2017) Molecular architecture of the major membrane ring component of the nuclear pore complex. *Structure* **25**, 434–445
 126. Politis, A., Schmidt, C., Tjioe, E., Sandercock, A. M., Lasker, K., Gordiyenko, Y., Russel, D., Sali, A., and Robinson, C. V. (2015) Topological models of heteromeric protein assemblies from mass spectrometry: Application to the yeast eIF3:eIF5 complex. *Chem. Biol.* **22**, 117–128
 127. Shi, Y., Pellarin, R., Fridy, P. C., Fernandez-Martinez, J., Thompson, M. K., Li, Y., Wang, Q. J., Sali, A., Rout, M. P., and Chait, B. T. (2015) A strategy for dissecting the architectures of native macromolecular assemblies. *Nat. Methods* **12**, 1135–1138
 128. Bonomi, M., Pellarin, R., Kim, S. J., Russel, D., Sundin, B. A., Riffle, M., Jaschob, D., Ramsden, R., Davis, T. N., Muller, E. G., and Sali, A. (2014) Determining protein complex structures based on a Bayesian model of in vivo Förster resonance energy transfer (FRET) data. *Mol. Cell. Proteomics* **13**, 2812–2823
 129. Schneidman-Duhovny, D., Hammel, M., Tainer, J. A., and Sali, A. (2016) FoXS, FoXSDock, and MultiFoXS: Single-state and multi-state structural modeling of proteins and their complexes based on SAXS profiles. *Nucleic Acids Res.* **44**, W424–W429
 130. Saltzberg, D. J., Hepburn, M., Pilla, K. B., Schriemer, D. C., Lees-Miller, S. P., Blundell, T. L., and Sali, A. (2019) SSEThread: Integrative threading of the DNA-PKcs sequence based on data from chemical cross-linking and hydrogen deuterium exchange. *Prog. Biophys. Mol. Biol.* **147**, 92–102

131. Schulze-Gahmen, U., Echeverria, I., Stjepanovic, G., Bai, Y., Lu, H., Schneidman-Duhovny, D., Doudna, J. A., Zhou, Q., Sali, A., and Hurley, J. H. (2016) Insights into HIV-1 proviral transcription from the structure and dynamics of the Tat:AFF4:P-TEFb:TAR complex. *Elife* **5**, e15910
132. Lasker, K., Topf, M., Sali, A., and Wolfson, H. J. (2009) Inferential optimization for simultaneous fitting of multiple components into a CryoEM map of their assembly. *J. Mol. Biol.* **388**, 180–194
133. Wu, H., Saltzberg, D. J., Kratochvil, H. T., Jo, H., Sali, A., and DeGrado, W. F. (2019) Glutamine side chain 13C–18O as a nonperturbative IR probe of amyloid fibril hydration and assembly. *J. Am. Chem. Soc.* **141**, 7320–7326
134. Dong, G. Q., Fan, H., Schneidman-Duhovny, D., Webb, B., and Sali, A. (2013) Optimized atomic statistical potentials: Assessment of protein interfaces and loops. *Bioinformatics* **29**, 3158–3166
135. Sali, A., and Blundell, T. L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779–815