QRB Discovery

www.cambridge.org/qrd

Frontiers in integrative structural modeling of macromolecular assemblies

Kartik Majila, Shreyas Arvindekar, Muskaan Jindal and Shruthi Viswanath

National Centre for Biological Sciences, Tata Institute of Fundamental Research, Bangalore, India

Perspective

Cite this article: Majila K, Arvindekar S, Jindal M, Viswanath S (2025). Frontiers in integrative structural modeling of macromolecular assemblies. *QRB Discovery*, **6**: e3, 1–13
https://doi.org/10.1017/qrd.2024.15.

Received: 29 June 2024
Revised: 13 September 2024
Accepted: 04 October 2024

Keywords:
Conformational ensembles; Electron cryo-tomography; Generative modeling; Integrative modeling; Intrinsically disordered proteins; Macromolecular assemblies; Protein language models

Corresponding author:
Shruthi Viswanath;
Email: shruthiv@ncbs.res.in

K.M. and S.A. authors have contributed equally.

Abstract

Integrative modeling enables structure determination for large macromolecular assemblies by combining data from multiple experiments with theoretical and computational predictions. Recent advancements in AI-based structure prediction and cryo electron-microscopy have sparked renewed enthusiasm for integrative modeling; structures from AI-based methods can be integrated with *in situ* maps to characterize large assemblies. This approach previously allowed us and others to determine the architectures of diverse macromolecular assemblies, such as nuclear pore complexes, chromatin remodelers, and cell–cell junctions. Experimental data spanning several scales was used in these studies, ranging from high-resolution data, such as X-ray crystallography and AlphaFold structure, to low-resolution data, such as cryo-electron tomography maps and data from co-immunoprecipitation experiments. Two recurrent modeling challenges emerged across a range of studies. First, these assemblies contained significant fractions of disordered regions, necessitating the development of new methods for modeling disordered regions in the context of ordered regions. Second, methods needed to be developed to utilize the information from cryo-electron tomography, a timely challenge as structural biology is increasingly moving towards *in situ* characterization. Here, we recapitulate recent developments in the modeling of disordered proteins and the analysis of cryo-electron tomography data and highlight other opportunities for method development in the context of integrative modeling.

Introduction

Integrative structural modeling is an approach for determining macromolecular structures that are challenging to determine experimentally (Alber et al., 2007; Sali, Glaeser, Earnest, & Baumeister, 2003). Data from multiple experiments is combined with physical principles, statistics of previous structures, and prior models for structure determination. This approach overcomes the limitations of individual techniques for structure determination and maximizes the accuracy, precision, completeness, and efficiency of structure determination (Rout & Sali, 2019; Sali, 2021).

Recent advancements in both computational and experimental domains have prompted a resurgence of interest in integrative modeling (Beck, Covino, Hänelt, & Müller-McNicoll, 2024; McCafferty et al., 2024). On the one hand, AI-based predictions of structures of proteins and their complexes with other proteins and nucleic acids have significantly advanced structural biology of late (Abramson et al., 2024; Akdel et al., 2022; Jumper et al., 2021). This has spurred the development of numerous methods that aim to integrate AI-based structures with diverse types of experimental data, including electron diffraction data from X-ray crystallography, electron density maps from electron cryo-microscopy, and chemical crosslinks from mass spectrometry (Chang et al., 2022; Stahl et al., 2024; Stahl, Graziadei, Dau, Brock, & Rappsilber, 2023; Terwilliger et al., 2022; Terwilliger et al., 2023; Zhang et al., 2023). These methods integrate the data in various ways, ranging from using the data to validate AI-based predictions, to using the data as additional inputs in the deep learning method, to encoding the data in the loss functions, resulting in structure predictions that are consistent with the data (O'Reilly et al., 2023; Stahl et al., 2023, 2024; Terwilliger et al., 2022, 2023; Zhang, Haghighatlari, et al., 2023). On the other hand, experimental techniques for *in situ* structure determination of assemblies are also rapidly advancing, with advancements in both hardware and software for imaging cells using cryo-electron tomography (Beck et al., 2024; McCafferty et al., 2024). This has led to an increase in tomography data, concurrent with an increase in the number and resolution of structures solved using tomography. Together, integrative methods using cryo-electron tomography maps along with AI-based structure predictions have resulted in significant advancements in structure determination, for example for nuclear pore complexes and ciliary complexes (Chen et al., 2023; Fontana et al., 2022; Hesketh, Mukhopadhyay, Nakamura, Toropova, & Roberts, 2022; McCafferty et al., 2024; Mosalaganti et al., 2022; Zhu et al., 2022).

Nonetheless, there is immense potential for advancing integrative modeling methods for macromolecular assemblies. Here, we provide our perspective on two areas warranting immediate method development in the context of integrative modeling: methods for modeling intrinsically disordered regions (IDRs) of proteins and approaches for leveraging *in situ* data. First, unlike ordered proteins, intrinsically disordered proteins (IDPs) comprise a dynamic ensemble of conformations that are best characterized in statistical terms rather than as static structures (Baul, Chakraborty, Mugnai, Straub, & Thirumalai, 2019). They comprise a significant fraction of the eukaryotic proteome and are involved in critical cellular processes (Oldfield & Dunker, 2014). They are found in several macromolecular assemblies, for example, the FG-Nups in the nuclear pore complex (Fontana et al., 2022; Zhu et al., 2022). However, their intrinsic disorder makes their characterization in these assemblies challenging. Improved representations for IDPs and methods for generating realistic IDP ensembles are crucial for understanding their functions. Second, the structural characterization of macromolecules using *in situ* data relies on accurate particle annotations on the tomograms (de Teresa-Trueba et al., 2023; Rice et al., 2023). However, owing to the low signal-to-noise ratio of the acquired tilt images, the missing wedge effect, and the inherent heterogeneity in the sample, the localization and identification of macromolecules in tomograms is time-consuming, laborious, and often challenging (de Teresa-Trueba et al., 2023; Moebel et al., 2021). Advances in deep learning methods and integrative approaches for combining data from other experimental and computational methods with cryo-electron tomograms can facilitate high throughput *in situ* structural characterization of macromolecular species.

In this Perspective, we first briefly review the existing integrative modeling methods and recent examples of macromolecular assemblies characterized using integrative modeling. Then, we discuss methods developed and opportunities for modeling disordered regions and leveraging *in situ* data. Finally, we end with an outlook summarizing other open problems in integrative modeling.

Integrative modeling methods

Several methods have been developed for integrative structure determination (Table 1). A subset of these including Integrative Modeling Platform (IMP), High Ambiguity Driven DOCKing (HADDOCK), and Assembline (Alber et al., 2007; Dominguez, Boelens, & Bonvin, 2003; Honorato et al., 2024; Rantos, Karius, & Kosinski, 2022; Russel et al., 2012) are discussed here. IMP is a framework for Bayesian integrative modeling that facilitates structure determination of macromolecular ensembles at multiple resolutions (multi-scale) and multiple states (multi-state) (Alber et al., 2007; Russel et al., 2012). A wide array of experimental data can be combined using IMP, for example in vivo genetic interactions, co-immunoprecipitation, FRET (Förster Resonance Energy Transfer), SAXS (small angle X-ray scattering), XLMS (chemical crosslinks from mass spectrometry), density maps from cryo electron-microscopy, and atomic structures from X-ray crystallography, NMR (Nuclear Magnetic Resonance), and AI-based predictions (Rout & Sali, 2019; Sali, 2021). The Bayesian inference framework allows for data from multiple sources to be integrated while considering the uncertainty in the data (Schneidman-Duhovny, Pellarin, & Sali, 2014). The modular design of IMP facilitates the mixing and matching of scoring functions and sampling algorithms. It has been used in the modeling of several large assemblies, most notably the nuclear pore complex (Akey et al., 2022; Alber et al., 2007; Rout & Sali, 2019; Sali, 2021; Singh et al., 2024). Recent advancements in IMP include Bayesian scoring functions for *in vivo* genetic interactions (Braberg et al., 2020), Bayesian model selection for optimizing model representation (Arvindekar, Pathak, Majila, & Viswanath, 2024), automated choice of sampling parameters (Pasani & Viswanath, 2021), and annotation of precision for model regions (Ullanat, Kasukurthi, & Viswanath, 2022).

Assembline is a protocol for integrative modeling that builds upon IMP, combining Xlink Analyzer, UCSF Chimera, and IMP to model large assemblies (Rantos et al., 2022). It is applicable for systems for which medium-resolution EM maps and a large number of atomic structures of subunits are available. It improves upon IMP by using pre-computed rigid body fits to EM maps to make the sampling more efficient. HADDOCK is a method for atomistic integrative modeling of protein complexes (Dominguez et al., 2003; Honorato et al., 2024). Experimental data from NMR, SAXS, XLMS, and mutagenesis studies are encoded as Ambiguous Interaction Restraints (AIR). Recent improvements to HADDOCK include the ability to model complexes of up to 20 macromolecules, new restraints based on cryo-EM maps, coarse-grained representations for efficient sampling, customizable pre- and post-processing steps, and a user-friendly web server for integrative modeling (Honorato et al., 2024).

Other than these, several methods allow fitting known protein structures into medium to low-resolution density maps, including MDFF and TEMPy-REFF (Beton, Mulvaney, Cragnolini, & Topf, 2024; Trabuco, Villa, Mitra, Frank, & Schulten, 2008). MDFF (Molecular dynamics flexible fitting) utilizes MD simulations for fitting structures into density maps by biasing the simulation using an additional potential derived from the density map (Trabuco et al., 2008). TEMPy-REFF (Responsibility-based Flexible-Fitting) refines an initial structure within a density map iteratively using the Expectation-Maximization algorithm (Beton et al., 2024).

Recent examples in integrative modeling: focus on nuclear and cell adhesion complexes

Integrative modeling has shed light on diverse cellular processes by determining the structures of assemblies associated with them. A list of representative recently characterized integrative structures is presented (Table 2). Here, we discuss examples of recent integrative structural biology studies in nuclear trafficking, gene expression regulation, and cell–cell adhesion. These studies not only provide novel insights into the structure and function of these assemblies but also highlight areas for future applications and method development.

The nuclear pore complex (NPC) is a large macromolecular assembly in the nuclear envelope that connects the nucleus and cytoplasm and plays an important role in nuclear trafficking (Akey et al., 2022; Alber et al., 2007). Several recent studies have improved our understanding of the components of the NPC (Bley et al., 2022; Fontana et al., 2022; Singh et al., 2024; Yu et al., 2023; Zhu et al., 2022). Some of these studies involve the fitting of AlphaFold and experimentally determined structures into medium-resolution cryo-EM maps and cryo-electron tomograms (Bley et al., 2022; Fontana et al., 2022; Petrovic et al., 2022; Zhu et al., 2022). Other studies additionally incorporate biochemical data including chemical crosslinks (Singh et al., 2024). Together these studies have been used to characterize the structures of the cytoplasmic face, cytoplasmic ring, the linker-scaffold network, and the nuclear basket of the NPC. The resulting structures enabled the identification of

https://doi.org/10.1017/qrd.2024.15 Published online by Cambridge University Press

Table 1. Integrative modeling software

Software	Authors	Reference	URL
ISD	Rieping, Habeck, & Nilges, (2005)	Rieping et al. (2005)	N/A
IMP	Russel et al. (2012)	Russel et al. (2012)	integrativemodeling.org
HADDOCK	Dominguez et al. (2003), Honorato et al. (2024)	Dominguez et al. (2003), Honorato et al. (2024)	rascar.science.uu.nl/haddock2.4
Assembline	Rantos et al. (2022)	Rantos et al. (2022)	embl-hamburg.de/Assembline/
PLUMED-ISDB	Bonomi & Camilloni, (2017)	Bonomi & Camilloni, (2017)	plumed.org
BioEn	Köfinger et al. (2019)	Köfinger et al. (2019)	github.com/bio-phys/BioEn
Rosetta	Simons, Kooperberg, Huang, & Baker, (1997), Leman et al. (2020)	Leman et al., (2020), Simons et al. (1997)	rosettacommons.org
CombFold	Shor & Schneidman-Duhovny, (2024b)	Shor & Schneidman-Duhovny, (2024a)	github.com/dina-lab3D/CombFold
CombDock	Inbar, Benyamini, Nussinov, & Wolfson, (2005), Schneidman-Duhovny & Wolfson, (2020)	Inbar et al. (2005), Schneidman-Duhovny & Wolfson, (2020)	bioinfo3d.cs.tau.ac.il/CombDock/download/

A list of commonly used integrative modeling software for large protein complexes. Each of these combines information from three or more experimental and/or computational sources. For a comprehensive overview, see (Bonomi et al., 2017; Habeck, 2023; Rout & Sali, 2019)

Table 2. A table summarizing a representative subset of recent integrative modeling studies

Macromolecular assembly	Subcellular location	Software for integrative modeling	Data used	Authors and year
A3G-CRL5-Vif complex	Nucleus	IMP	Data from XLMS, and structures from X-ray crystallography and solution NMR	Kaake et al. (2021)
Apo-GAFab complex	Plasma membrane	IMP	Data from XLMS, and structures from X-ray crystallography	Gupta et al. (2020)
Bovine adenylyl cyclase 8 in complex with the G protein heterodimer	Plasma membrane	HADDOCK	Data from XLMS, and structures from X-ray crystallography and cryo-EM maps	Khanppnavar, B (2024)
CLOCK-BMAL1 bound to a nucleosome	Nucleus	Rosetta	Data from XLMS, and structures from X-ray crystallography and cryo-EM maps	Michael et al. (2023)
Desmosomal outer dense plaque	Plasma membrane	IMP	Data from cryo-ET, immuno-EM, yeast two-hybrid experiments, co-immunoprecipitation, in vitro overlay, in vivo co-localization assays, in silico sequence-based predictions of transmembrane and disordered regions, and structures from X-ray crystallography and homology modeling	Pasani et al. (2024)
Doublecortin-microtubule complex	Cytoplasm	IMP	Data from cryo-EM and XLMS, and structures obtained from cryo-EM	Rafiei et al. (2022)
gammaTuSC-Spc110 dimer complex	Nuclear membrane	IMP	Data from XLMS, and structures from cryo-EM maps	Brilot et al. (2021)
Human LINE–1 ORF2p	Nucleus	IMP	Data from cryo-EM and XLMS	Baldwin et al. (2024)
Intraflagellar transport - A (IFT-A) complex	Flagella	IMP	Data from XLMS cryo-ET, and AlphaFold structure predictions	McCafferty et al. (2022)
Mis18 Complex Assembly	Centromere	CombDock	Data from NS-EM and XLMS, and structures from X-ray crystallography	Thamkachy et al. (2024)
Mycobacterial ESX–5 type VII secretion system pore complex	Plasma membrane	IMP	Data from cryo-EM and XLMS, and structures obtained using X-ray crystallography and homology modeling	Beckham et al. (2021)

https://doi.org/10.1017/qrd.2024.15 Published online by Cambridge University Press

Table 2 *Continued*

Macromolecular assembly	Subcellular location	Software for integrative modeling	Data used	Authors and year
Nexin-dynein regulatory complex	Cilia	Assembline	Data from XLMS and Alphafold structure predictions.	Ghanaeian et al. (2023)
Nuclear Basket of NPC	Nuclear membrane	IMP	Data from quantitative mass spectrometry, XLMS, cryo-ET, immuno-EM, biochemical studies, and bioinformatics predictions, and prior integrative models	Singh et al. (2024)
Nuclear Pore Complex (NPC)	Nuclear membrane	IMP	Data from cryo-ET, cryo-EM, XLMS, quantitative fluorescence imaging, and biochemical studies, and Alphafold structure predictions	Akey et al. (2023)
NuRD subcomplexes	Nucleus	IMP	Data from SEC-MALLS, DIA-MS, XLMS, negative-stain EM, and structures from X-ray crystallography, NMR spectroscopy, secondary structure predictions, and homology models	Arvindekar et al. (2022)
SARS-CoV2 Nsp1, Nsp2 and nucleocapsid proteins	Host cytoplasm and viral nucleocapsid	CombDock	Data from XLMS and structures from AlphaFold2 and homology modeling	Slavin et al. (2021)
SMC5/6 complex	Nucleus	IMP	Data from NS-EM and XLMS, and structures obtained using X-ray crystallography, cryo-EM, comparative modeling, and coiled-coil predictions	Yu et al. (2021)
Transglutaminase 2 in complex with plasma fibronectin type III modules 14 and 15	Extracellular matrix	HADDOCK	Data from XLMS, and structures from X-ray crystallography	Selcuk et al. (2024)
Type III Secretion System	Plasma membrane, cell wall	Assembline	Data from XLMS and structures from cryo-EM and NMR spectroscopy	Flacht et al. (2023)
WDR76—SPIN1—nucleosome complex	Nucleus	HADDOCK, IMP	Data from XLMS and structures from X-ray crystallography and I-TASSER structure predictions	Liu et al. (2024)

Abbreviations: DIA-MS, Data independent acquisition mass spectrometry; EM, Electron microscopy; ET, Electron tomography; NMR, Nuclear magnetic resonance; NS, Negative staining; SEC-MALLS, Size exclusion chromatography—multi-angle laser light scattering; XLMS, Crosslinking coupled with mass spectrometry.

novel interfaces between disordered nucleoporins (Nups) (Fontana et al., 2022; Zhu et al., 2022), elucidated the function of nucleoporins—Nup38 and the Cytoplasmic Filament Nucleoporin (CFNC) (Bley et al., 2022), delineated the role of Mlp/Trp in assisting mRNP transport (Bley et al., 2022; Fontana et al., 2022; Singh et al., 2024; Yu et al., 2023; Zhu et al., 2022), and revealed the plasticity and robustness of the inner ring (Petrovic et al., 2022). Finally, another study determined the distribution of intrinsically disordered nucleoporins in the NPC and their motion in the central channel using fluorescence lifetime imaging of fluorescence resonance energy transfer (FLIM-FRET) and coarse-grained molecular dynamic (MD) simulations (Yu et al., 2023).

Whereas the above studies are on components of the NPC, (Akey et al., 2022, 2023; Mosalaganti et al., 2022) determined comprehensive integrative structures of the entire NPC. These studies integrate *in situ* cryo-electron tomography data with AlphaFold or experimentally determined structures (Mosalaganti et al., 2022), and additionally cryo-EM maps, chemical crosslinks, and data from quantitative fluorescence imaging and biochemical studies to determine comprehensive structures of NPCs (Akey et al., 2022, 2023). The structures revealed distinct dilated and constricted states of the complex and characterized the plasticity of the pore (Akey et al., 2022, 2023; Mosalaganti et al., 2022). Additionally, they localized precise anchoring sites for the intrinsically disordered Nups (Mosalaganti et al., 2022) and delineated the function of Pom153 in ring dilation (Akey et al., 2023).

The Nucleosome Remodeling and Deacetylase (NuRD) complex is a chromatin remodifying assembly that plays an important role in several cellular processes including transcriptional regulation, cell cycle progression, and cellular differentiation (Arvindekar et al., 2022). It consists of chromatin remodeling and deacetylase modules, connected by MBD and GATAD2 proteins. The structures of three subcomplexes of NuRD were determined by integrating data from negative-stain and low-resolution cryo-EM maps, X-ray crystallography, XLMS, SEC-MALS, DIA-MS, NMR spectroscopy, homology modeling, secondary structure predictions, and physical principles (Arvindekar et al., 2022). The integrative structures depict MBD in two states in NuRD and elucidate the role of the intrinsically disordered region of MBD in bridging the chromatin remodeling and deacetylase modules of NuRD.

https://doi.org/10.1017/qrd.2024.15 Published online by Cambridge University Press

Desmosomes are intercellular junctions that tether the intermediate filaments of adjacent cells in tissues under mechanical stress (Pasani, Menon, & Viswanath, 2024). The integrative structure of the desmosomal outer dense plaque (ODP) was determined by combining data from cryo-electron tomography, X-ray crystallography, immuno-electron microscopy, in vitro overlay, *in vivo* co-localization assays, Yeast Two-Hybrid (Y2H), co-immuno precipitation, *in-silico* sequence-based predictions of transmembrane and disordered regions, homology modeling, and stereochemistry (Pasani et al., 2024). The structure enabled the localization of disordered regions of Plakophilin (PKP) and Plakoglobin (PG) and the identification of novel protein–protein interfaces associated with them, leading to hypotheses about the functions of these disordered regions.

Two elements emerge as common across the aforementioned studies: they leverage *in situ* cryo-electron tomography data and the characterized systems contain significant fractions of disordered regions (Figure 1). This highlights two areas of immediate interest for method development: modeling with intrinsically disordered proteins (IDP) and utilizing data from cryo-electron tomography (cryo-TM), discussed in the following sections.

Integrative modeling of intrinsically disordered proteins

Intrinsically disordered proteins (IDPs) are a class of proteins that lack a well-defined ordered structure in their monomeric state. Rather, they exist as an ensemble of interconverting conformers in equilibrium and hence are structurally heterogeneous (Baul et al., 2019; Lindorff-Larsen & Kragelund, 2021). This heterogeneity of IDPs also makes it challenging to characterize them both experimentally and computationally (Beck et al., 2024).

Learning Representations for IDPs

Recently, protein language models (pLMs) have emerged as powerful tools for learning context-aware representations, providing a compact and informative approach to characterize the structural and functional properties of proteins (Bepler & Berger, 2021; Rives et al., 2021). pLMs enhance the performance of models on downstream tasks via transfer learning, eliminating the need to train a neural network from end to end. This approach is particularly beneficial while training models with small datasets.

Using pLMs for IDPs presents several challenges. First, pLMs trained only on sequences may not be able to capture the conformational heterogeneity of IDPs. Second, the databases used to train pLMs are dominated by ordered protein sequences, leading to a bias in the learned representations. Third, IDPs often function through transient interactions and context-dependent conformations, i.e., the same IDP may adopt different conformations with different binding partners. The state-of-the-art pLMs do not account for the environmental context and interacting partners and thus may not capture these transient interactions. Finally, the lack of structural data representative of IDP conformations poses a significant challenge in training models.

Advances in representation learning techniques are required for accurately characterizing the behavior of IDPs. Representations for IDPs could be improved by fine-tuning existing pLMs on IDP-specific tasks and/or by incorporating additional data on IDPs. Sequence alone might not be sufficient to capture the properties of IDPs; incorporating structural information or physics-based priors might allow pLMs to capture the complex dynamics of IDPs (Wang, Wang, Evans, & Tiwary, 2024). Structure-aware pLMs have been recently developed (Peñaherrera & Koes, 2024; Sun & Shen, 2023; Wang et al., 2024). The same approach can be extended to IDPs. There is a need to obtain more structural data for IDPs (Jahn, Marquet, Heinzinger, & Rost, 2024). Whereas, experimental structural data remains important, acquiring it might be tedious and time-consuming. Computational approaches for generating realistic IDP conformational ensembles, such as MD simulations and generative models, would provide valuable experimental-like structural data. In the next section, we discuss methods for generating IDP ensembles.

Generating IDP ensembles

Determining the conformational ensembles of IDPs is essential for understanding their functions. MD simulations are widely used for generating conformational ensembles. However, their reliability depends on the accuracy of force fields and the ergodicity of sampling (Bonomi, Heller, Camilloni, & Vendruscolo, 2017; Robustelli, Piana, & Shaw, 2018). Force fields typically used for folded proteins often fail to accurately capture the conformations of IDPs when compared with experimental data. Efforts for improving the force fields for IDPs focus on either refining the protein force field (Baul et al., 2019; Huang et al., 2017; Joseph et al., 2021), or accurately accounting for protein-water interactions (Best, Zheng, & Mittal, 2014; Nerenberg, Jo, So, Tripathy, & Head-Gordon, 2012; Robustelli et al., 2018; Vitalis & Pappu, 2009). Coarse-grained models that improve sampling by reducing the degrees of freedom have also been developed (Baratam & Srivastava, 2024; Baul et al., 2019; Joseph et al., 2021; Marrink, Risselada, Yefimov, Tieleman, & de Vries, 2007; Thomasen, Pesce, Roesgaard, Tesei, & Lindorff-Larsen, 2022).

Deep generative models offer a computationally efficient means for sampling conformations from a learned data distribution. Latent space embeddings from variational autoencoder (VAE) trained on IDP sequences (Mansoor, Baek, Park, Lee, & Baker, 2024), conditional generative adversarial networks (GAN) (Janson, Valdes-Garcia, Heo, & Feig, 2023), denoising diffusion probabilistic models (DDPM) (Janson & Feig, 2024; Zhu et al., 2024) have been used for generating all-atom and Cα coarse-grained ensembles of IDPs. More sophisticated approaches such as flow matching may also be employed for generating ensembles of IDPs. Notably, these aforementioned generative models leverage MD-generated ensembles for training.

Recent studies demonstrate the combined use of MD simulations and machine learning approaches to generate IDP conformers with the aim of predicting the biophysical properties of IDPs and designing IDP sequences (Lotthammer, Ginell, Griffith, Emenecker, & Holehouse, 2024; Pesce et al., 2024; Tesei et al., 2024). For example, the ALBATROSS deep learning model was developed for predicting the biophysical properties of IDPs, such as the radius of gyration, by training on IDP ensembles generated via the MPIPI-GG model (Lotthammer et al., 2024). Similarly, support vector regression models were trained to predict chain compaction for IDP sequences using IDP ensembles generated by the CALVADOS model (Tesei et al., 2024). Lastly, a method for designing IDP sequences with pre-defined conformational properties was developed by combining ensemble generation using CALVADOS with alchemical free-energy calculations within a Markov Chain Monte Carlo (MCMC) optimization framework (Pesce et al., 2024).

https://doi.org/10.1017/qrd.2024.15 Published online by Cambridge University Press

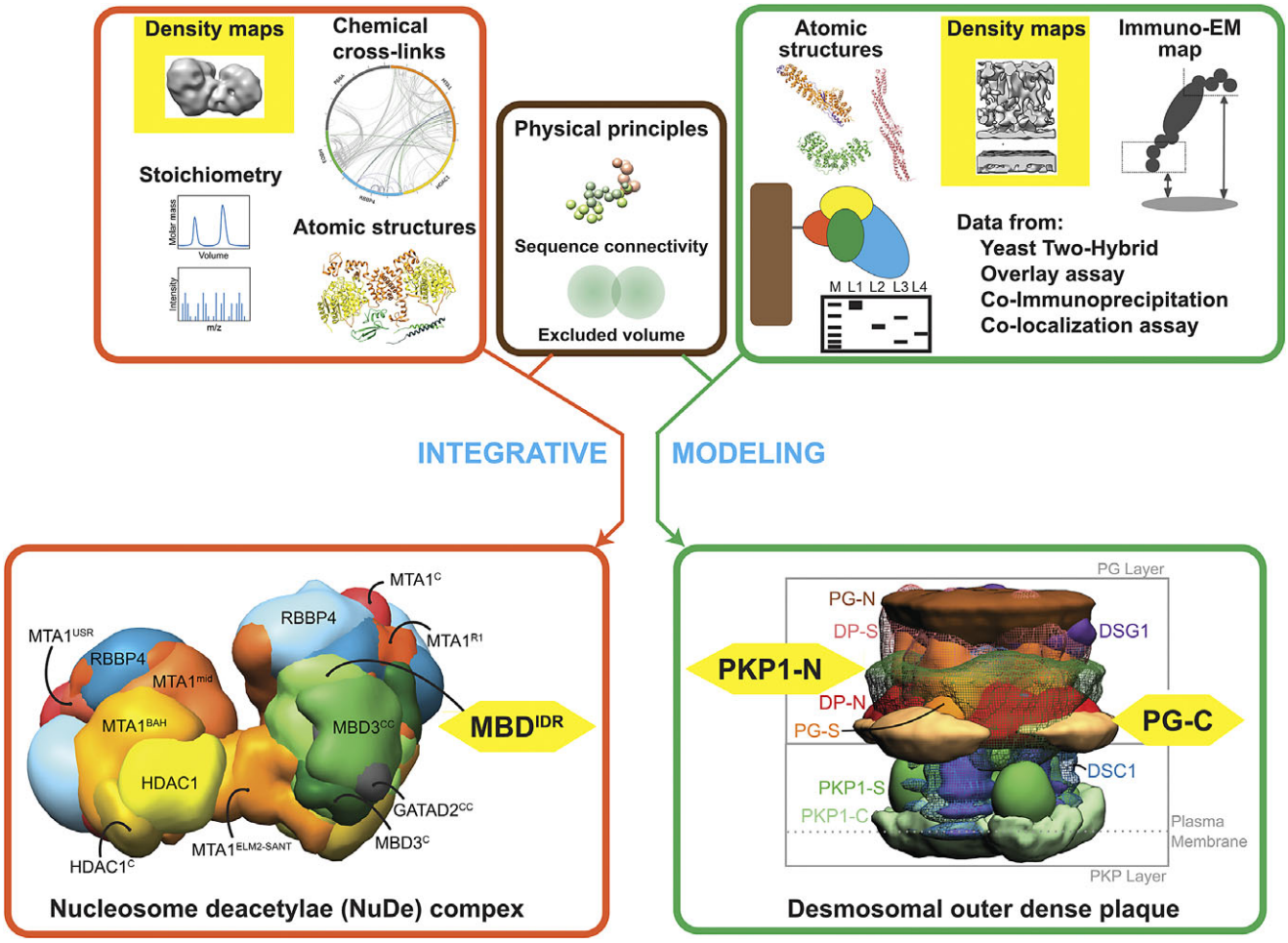

Figure 1. Frontiers in integrative structure determination. Schematic describing integrative structure determination for the nucleosome remodeling and deacetylase complex (orange box) and the desmosomal outer dense plaque (green box) combining data from multiple sources. Input low-resolution cryo-EM and cryo-ET maps and intrinsically disordered regions in both complexes are highlighted in yellow.

Integrating experimental data for generating IDP ensembles

Broadly, experimental data can be utilized for modeling IDPs in several ways: validation of generated ensembles, reweighting generated ensembles using experimental data, incorporating experimental data as restraints for sampling conformations, or using experimental data to improve existing force fields (Bernetti & Bussi, 2023; Chan-Yao-Chong, Durand, & Ha-Duong, 2019; Fisher & Stultz, 2011). A comprehensive list of methods can be found in reviews on this topic (Bonomi et al., 2017; Habeck, 2023).

First, ensemble validation involves generating realistic ensembles of IDPs and validating the results with experimental data (Chan-Yao-Chong et al., 2019). Due to their ability to capture the dynamics of IDPs, NMR, and SAS data are most commonly used for validating the generated ensembles for IDPs (Baratam & Srivastava, 2024; Shrestha, Smith, & Petridis, 2021). Second, ensemble weighting involves using experimental data to refine an existing ensemble, to minimize deviation of the ensemble from the observed data (Chan-Yao-Chong et al., 2019). This can be achieved by maximum parsimony (SES Berlin et al., 2013) or maximum entropy (Pitera & Chodera, 2012; Roux & Weare, 2013; Cavalli, Camilloni, & Vendruscolo, 2013) (EROS Różycki, Kim, & Hummer, 2011, (BioEn Hummer & Köfinger, 2015), and ABSURD (Salvi, Abyzov, & Blackledge, 2016). Bayesian inference methods allow consideration of uncertainty in data (Fisher, Ullman, & Stultz, 2013; Lincoff et al., 2020). Combining Bayesian inference and maximum entropy methods helps overcome the limitations of each (Crehuet, Buigues, Salvatella, & Lindorff-Larsen, 2019; Fröhlking, Bernetti, & Bussi, 2023). Deep learning models in combination with Bayesian and maximum entropy methods can also be used for refining an initial pool of conformations (DynamICE: Zhang, Haghighatlari, et al., 2023). Third, experimental data can also be used as restraints to guide simulations (Chan-Yao-Chong et al., 2019). Metainference uses Bayesian inference for incorporating noisy, ensemble-averaged experimental data using replica-averaged modeling (Bonomi, Camilloni, Cavalli, & Vendruscolo, 2016; Bonomi, Camilloni, & Vendruscolo, 2016). Similarly, parallel replica ensemble restraints based on SAXS data were used in MD simulations of IDPs (Hermann & Hub, 2019). Finally, experimental data can also be used for improving existing force fields on the fly using a Maximum Entropy approach (Cesari, Gil-Ley, & Bussi, 2016).

A holistic understanding of the dynamic behavior of IDPs requires realistic conformational ensembles that can be generated using MD simulations and deep generative models. MD simulations can provide experimental-like ensembles for training deep generative models; the latter may aid in improving force fields, enhancing sampling of IDP conformations, and analyzing the

https://doi.org/10.1017/qrd.2024.15 Published online by Cambridge University Press

ensemble generated via MD. Thus, an integrated approach would enable overcoming the limitations of each and improving our understanding of the dynamic nature of IDPs.

Integrative structure determination using *in situ* data

Cryo-electron tomography (cryo-ET) is a cryo-EM imaging technique that enables structural characterization of macromolecular species (macromolecules, their complexes, and assemblies), in their native cellular environment at nanometer resolution (Gubins et al., 2020; Lamm et al., 2022). High-throughput localization and identification of macromolecular species within a tomogram can provide insights into their conformational heterogeneity, potential interactors, counts, and distributions within the cell (Arvindekar, Majila, & Viswanath, 2024; Beck et al., 2024; Förster, Han, & Beck, 2010; McCafferty et al., 2024). Integrating cryo-ET data along with complementary data from experiments such as XLMS, Y2H, cryo-EM Single Particle Analysis (SPA), FRET, AI-based structure predictions, and prior structural models can help build a comprehensive structural atlas of the cell (Beck et al., 2024; Förster et al., 2010; McCafferty et al., 2024). However, the intracellular crowding, compositional heterogeneity and low copy numbers of macromolecular species, the low signal-to-noise ratio, and the missing wedge in the tomography data pose significant challenges for localizing and identifying macromolecules in the tomograms (Moebel et al., 2021; Pyle & Zanetti, 2021).

Localization and identification of macromolecular species with known structures

Macromolecular species with known structures are often annotated in tomograms either manually or by template matching. Manual particle annotation, however, is time-consuming, laborious, error-prone, and not suitable for high-throughput workflows (Lamm et al., 2022). Template matching involves using a low-pass filtered template of the known structure of a target macromolecule to localize similar densities in the tomogram (Frangakis et al., 2002). Methods for template matching are under active development (Cruz-León et al., 2024; Maurer, Siggel, & Kosinski, 2024). For example, the use of high-resolution information and template-specific search parameter optimization for objective, comprehensive, and high-confidence localization and identification of macromolecular species in tomograms was recently proposed (Cruz-León et al., 2024).

In addition to template matching, several supervised learning methods have also been recently developed. Two such deep learning-based methods, DeepFinder and DeePiCt, utilize convolutional neural networks (CNNs) for simultaneous localization and identification of macromolecular species (de Teresa-Trueba et al., 2023; Moebel et al., 2021). Another deep learning-based object detection method, MemBrain, was developed for estimating the localizations and orientations of membrane-embedded macromolecules (Lamm et al., 2022, 2024). These approaches have been shown to outperform template matching for localizing macromolecules (de Teresa-Trueba et al., 2023; Gubins et al., 2020; Lamm et al., 2022; Moebel et al., 2021). However, similar to manual annotation and template matching, these supervised learning approaches are limited to macromolecules with known structures. They are not suitable for high-throughput workflows and *de novo* structural characterization of macromolecular species (de Teresa-Trueba et al., 2023; Gubins et al., 2020; Lamm et al., 2022; Moebel et al., 2021).

de novo localization and identification of species

For de novo structural characterization of macromolecular species with unknown structures, deep metric learning-based approaches, such as TomoTwin, and unsupervised learning approaches, such as Multi-Pattern Pursuit (MPP) and Deep Iterative Subtomogram Clustering Approach (DISCA) were recently developed (Rice et al., 2023; Xu et al., 2019; Zeng et al., 2023). These approaches aim to cluster subtomograms based on their structural similarity. Subtomogram averaging on the clustered subtomograms can aid in the structural characterization of macromolecular species at 10–20 Å resolutions (Rice et al., 2023; Zeng et al., 2023). These approaches are currently sensitive to noise in the tomograms and the size and abundance of the macromolecular species. However, they hold great promise for de novo high-throughput structural characterization of macromolecular species using tomographic data.

Visual proteomics

Visual proteomics is an approach that aims to build molecular atlases that encapsulate structural descriptions of macromolecules within the cell using methods such as cryo-ET (Beck et al., 2024; Förster et al., 2010; McCafferty et al., 2024). This approach is inherently integrative. Given a tomogram, large macromolecular species with known atomic structures can be localized and identified within it using methods like template matching. Densities with unknown macromolecular identities can be obtained using the de novo approaches described above. The in situ structures of these uncharacterized macromolecular species can then be determined using an integrative approach by rigid fitting of structures obtained using cryo-EM SPA, X-ray crystallography, and AI-based structure predictions along with data from orthogonal experiments such as fluorescence microscopy and XLMS (Beck et al., 2024; Förster et al., 2010; McCafferty et al., 2024). For example, recent studies used integrative approaches to combine data from cryo-ET, SPA with cryo-EM, mass spectrometry, and predictions from AlphaFold to understand the molecular architecture of the human IFT-A and IFT-B complexes (Hesketh et al., 2022) and microtubule doublets in mouse sperm cells (Chen et al., 2023). In summary, utilizing cryo-ET data in an integrative approach can provide insights into interactors of a macromolecular species, associated protein communities, and larger cellular neighborhoods (Beck et al., 2024; Förster et al., 2010; McCafferty et al., 2024).

Outlook

Integrative modeling has progressed significantly in the past decade, as evidenced by the increasing number, size, and precision of structures deposited to the PDB-Dev and integrated into the PDB (https://pdb-dev.wwpdb.org) (Saltzberg et al., 2021; Vallat et al., 2021). Integrative structural biology plays a crucial role in the era of AI-based structure predictions. Experimental data from rapidly advancing techniques such as cryo-electron tomography, and AI-based predictions can complement each other within an integrative framework (Arvindekar, Majila, & Viswanath, 2024; Beck et al., 2024; McCafferty et al., 2024; Shor & Schneidman-Duhovny, 2024b). This approach has proved powerful for several systems

https://doi.org/10.1017/qrd.2024.15 Published online by Cambridge University Press

such as ciliary complexes and nuclear pore complexes (Chen et al., 2023; Fontana et al., 2022; Hesketh et al., 2022; McCafferty et al., 2024; Mosalaganti et al., 2022; Zhu et al., 2022). Alphafold and similar AI-based prediction methods can increasingly solve structures for larger and more complex systems (Abramson et al., 2024). However, their applicability to solving entire structures of large assemblies remains an open question as they are limited by the GPU memory as well as the availability of training data. For example, membrane proteins and IDPs are under-represented in the training data (Carugo & Djinović-Carugo, 2023; Dobson et al., 2023). The low-pLDDT regions in Alphafold structures often coincide with IDRs, suggesting that Alphafold may be used to predict these regions (Wilson, Choy, & Karttunen, 2022). In contrast, in cases where Alphafold predicts structures of IDPs with high confidence, these regions typically represent the folded conformations of the IDPs, indicating a disorder-to-order transition in the presence of a partner (Alderson, Pritišanac, Kolarić, Moses, & Forman-Kay, 2023; Wilson et al., 2022). Nonetheless, the static structures from Alphafold are not an accurate representation of the dynamic behavior of IDPs, characterized by an ensemble of conformations (Ruff & Pappu, 2021).

In this Perspective, we highlighted two emerging frontiers for method development in integrative modeling: modeling disordered regions and modeling with data from cryo-electron tomography. Here, we briefly point to other open areas in integrative modeling that are the subject of current studies and/or may benefit from timely method development. First, a lack of knowledge about the system stoichiometry is one of the challenges for starting integrative modeling. Methods to estimate the stoichiometry based on the confidence of AI-based predictions are only beginning to be developed and are not yet generalizable (Chim & Elofsson, 2024; Shor & Schneidman-Duhovny, 2024b, 2024a). Second, methods for incorporating in vivo data in modeling are required. Recently, in vivo genetic interaction measurements were encoded as Bayesian distance restraints for integrative modeling of assemblies (Braberg et al., 2020). Similarly, methods for integrating other in vivo data such as data from super-resolution microscopy may also be developed to model larger cellular neighborhoods. Third, on the model representation front, it would be beneficial to determine system representation using objective measures instead of fixing them ad hoc (Arvindekar, Pathak, et al., 2024; Viswanath & Sali, 2019). Current methods for optimizing representations are limited to assessing a small number of candidate representations (Arvindekar, Pathak, et al., 2024; Viswanath & Sali, 2019). Methods that enable sampling and assessing a large number of representations, for example by dynamically varying the model representations during sampling, would benefit integrative modeling (Viswanath & Sali, 2019). Fourth, methods for integrative modeling of dynamic systems with multiple discrete states and/or a continuum of states are also continually advancing (Habeck, 2023; Hoff, Thomasen, Lindorff-Larsen, & Bonomi, 2024; Hoff, Zinke, Izadi-Pruneyre, & Bonomi, 2024; Lincoff et al., 2020; Potrzebowski, Trewhella, & Andre, 2018). Fifth, sampling procedures in integrative modeling may be improved by leveraging the recent advances in deep learning, particularly in generative modeling. Specifically, recent generative modeling methods for protein structure prediction may be extended to incorporate experimental data, potentially leading to more efficient sampling procedures than the current stochastic sampling methods (Jing, Berger, & Jaakkola, 2024; Watson et al., 2023; Wu et al., 2024; Zheng et al., 2024). Finally, methods for comprehensive validation of integrative models, including assessment of model uncertainty and Bayesian assessment of fit to different kinds of input data are also necessary and are under development (Sali et al., 2015; Vallat et al., 2021). In all, these efforts will facilitate faster, more accurate, and more precise characterization of larger assemblies (Sali, 2021). The grand challenge in the field is to construct spatiotemporal models of entire cells. Integrative models of assemblies can contribute directly to this effort via metamodeling efforts that involve the integration of models at different scales to address the grand challenge (Raveh et al., 2021).

Open peer review. To view the open peer review materials for this article, please visit http://doi.org/10.1017/qrd.2024.15.

Acknowledgments. Molecular graphics images were produced using the UCSF Chimera and UCSF ChimeraX packages from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIH P41 RR001081, NIH R01-GM129325, and National Institute of Allergy and Infectious Diseases).

Author contribution. K.M., S.A., and M.J.: reading and synthesis. K.M., S.A., M.J., and S.V.: writing: original draft, writing: revision. K.M.: visualization. S.V.: supervision, funding.

Funding. This work has been supported by the following grants: Department of Atomic Energy (DAE) TIFR grant RTI 4006, Department of Science and Technology (DST) SERB grant SPG/2020/000475, and Department of Biotechnology (DBT) BT/PR40323/BTIS/137/78/2023 from the Government of India to S.V.

Competing interest. None declared.

References

Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A. J., Bambrick, J., Bodenstein, S. W., Evans, D. A., Hung, C.-C., O'Neill, M., Reiman, D., Tunyasuvunakool, K., Wu, Z., Žemgulytė, A., Arvaniti, E., & Jumper, J. M. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, **630**(8016), 493–500. https://doi.org/10.1038/s41586-024-07487-w

Akdel, M., Pires, D. E. V., Pardo, E. P., Jänes, J., Zalevsky, A. O., Mészáros, B., Bryant, P., Good, L. L., Laskowski, R. A., Pozzati, G., Shenoy, A., Zhu, W., Kundrotas, P., Serra, V. R., Rodrigues, C. H. M., Dunham, A. S., Burke, D., Borkakoti, N., Velankar, S., & Beltrao, P. (2022). A structural biology community assessment of AlphaFold2 applications. *Nature Structural & Molecular Biology*, **29**(11), 1056–1067. https://doi.org/10.1038/s41594-022-00849-w

Akey, C. W., Echeverria, I., Ouch, C., Nudelman, I., Shi, Y., Wang, J., Chait, B. T., Sali, A., Fernandez-Martinez, J., & Rout, M. P. (2023). Implications of a multiscale structure of the yeast nuclear pore complex. *Molecular Cell*, **83**(18), 3283–3302.e5. https://doi.org/10.1016/j.molcel.2023.08.025

Akey, C. W., Singh, D., Ouch, C., Echeverria, I., Nudelman, I., Varberg, J. M., Yu, Z., Fang, F., Shi, Y., Wang, J., Salzberg, D., Song, K., Xu, C., Gumbart, J. C., Suslov, S., Unruh, J., Jaspersen, S. L., Chait, B. T., Sali, A., & Rout, M. P. (2022). Comprehensive structure and functional adaptations of the yeast nuclear pore complex. *Cell*, **185**(2), 361–378.e25. https://doi.org/10.1016/j.cell.2021.12.015

Alber, F., Dokudovskaya, S., Veenhoff, L. M., Zhang, W., Kipper, J., Devos, D., Suprapto, A., Karni-Schmidt, O., Williams, R., Chait, B. T., Rout, M. P., & Sali, A. (2007). Determining the architectures of macromolecular assemblies. *Nature*, **450**(7170), 683–694. https://doi.org/10.1038/nature06404

Alderson, T. R., Pritišanac, I., Kolarić, Đ., Moses, A. M., & Forman-Kay, J. D. (2023). Systematic identification of conditionally folded intrinsically disordered regions by AlphaFold2. *Proceedings of the National Academy of Sciences*, **120**(44), e2304302120. https://doi.org/10.1073/pnas.2304302120

Arvindekar, S., Jackman, M. J., Low, J. K. K., Landsberg, M. J., Mackay, J. P., & Viswanath, S. (2022). Molecular architecture of nucleosome remodeling and deacetylase sub-complexes by integrative structure determination. *Protein Science*, **31**(9), e4387. https://doi.org/10.1002/pro.4387

Arvindekar, S., **Majila, K.**, & **Viswanath, S.** (2024). Recent methods from statistical inference and machine learning to improve integrative modeling of macromolecular assemblies (Version 4). arXiv. https://doi.org/10.48550/ARXIV.2401.17894

Arvindekar, S., **Pathak, A. S.**, **Majila, K.**, & **Viswanath, S.** (2024). Optimizing representations for integrative structural modeling using Bayesian model selection. *Bioinformatics*, **40**(3), btae106. https://doi.org/10.1093/bioinformatics/btae106

Baldwin, E. T., **Van Eeuwen, T.**, **Hoyos, D.**, **Zalevsky, A.**, **Tchesnokov, E. P.**, **Sánchez, R.**, **Miller, B. D.**, **Di Stefano, L. H.**, **Ruiz, F. X.**, **Hancock, M.**, **Işik, E.**, **Mendez-Dorantes, C.**, **Walpole, T.**, **Nichols, C.**, **Wan, P.**, **Riento, K.**, **Halls-Kass, R.**, **Augustin, M.**, **Lammens, A.**, & **Taylor, M. S.** (2024). Structures, functions and adaptations of the human LINE-1 ORF2 protein. *Nature*, **626**(7997), 194–206. https://doi.org/10.1038/s41586-023-06947-z

Baratam, K., & **Srivastava, A.** (2024). SOP-MULTI: A self-organized polymer based coarse-grained model for multi-domain and intrinsically disordered proteins with conformation ensemble consistent with experimental scattering data. *Journal of Chemical Theory and Computation*, **20**(22), 10179–10198. https://doi.org/10.1101/2024.04.29.591764

Baul, U., **Chakraborty, D.**, **Mugnai, M. L.**, **Straub, J. E.**, & **Thirumalai, D.** (2019). Sequence effects on size, shape, and structural heterogeneity in intrinsically disordered proteins. *The Journal of Physical Chemistry. B*, **123**(16), 3462–3474. https://doi.org/10.1021/acs.jpcb.9b02575

Beck, M., **Covino, R.**, **Hänelt, I.**, & **Müller-McNicoll, M.** (2024). Understanding the cell: future views of structural biology. *Cell*, **187**(3), 545–562. https://doi.org/10.1016/j.cell.2023.12.017

Beckham, K. S. H., **Ritter, C.**, **Chojnowski, G.**, **Ziemianowicz, D. S.**, **Mullapudi, E.**, **Rettel, M.**, **Savitski, M. M.**, **Mortensen, S. A.**, **Kosinski, J.**, & **Wilmanns, M.** (2021). Structure of the mycobacterial ESX-5 type VII secretion system pore complex. *Science Advances*, **7**(26), eabg9923. https://doi.org/10.1126/sciadv.abg9923

Bepler, T., & **Berger, B.** (2021). Learning the protein language: evolution, structure, and function. *Cell Systems*, **12**(6), 654–669.e3. https://doi.org/10.1016/j.cels.2021.05.017

Berlin, K., **Castañeda, C. A.**, **Schneidman-Duhovny, D.**, **Sali, A.**, **Nava-Tudela, A.**, & **Fushman, D.** (2013). Recovering a representative conformational ensemble from underdetermined macromolecular structural data. *Journal of the American Chemical Society*, **135**(44), 16595–16609. https://doi.org/10.1021/ja4083717

Bernetti, M., & **Bussi, G.** (2023). Integrating experimental data with molecular simulations to investigate RNA structural dynamics. *Current Opinion in Structural Biology*, **78**, 102503. https://doi.org/10.1016/j.sbi.2022.102503

Best, R. B., **Zheng, W.**, & **Mittal, J.** (2014). Balanced protein-water interactions improve properties of disordered proteins and non-specific protein association. *Journal of Chemical Theory and Computation*, **10**(11), 5113–5124. https://doi.org/10.1021/ct500569b

Beton, J. G., **Mulvaney, T.**, **Cragnolini, T.**, & **Topf, M.** (2024). Cryo-EM structure and B-factor refinement with ensemble representation. *Nature Communications*, **15**(1), Article 1. https://doi.org/10.1038/s41467-023-44593-1

Bley, C. J., **Nie, S.**, **Mobbs, G. W.**, **Petrovic, S.**, **Gres, A. T.**, **Liu, X.**, **Mukherjee, S.**, **Harvey, S.**, **Huber, F. M.**, **Lin, D. H.**, **Brown, B.**, **Tang, A. W.**, **Rundlet, E. J.**, **Correia, A. R.**, **Chen, S.**, **Regmi, S. G.**, **Stevens, T. A.**, **Jette, C. A.**, **Dasso, M.**, & **Hoelz, A.** (2022). Architecture of the cytoplasmic face of the nuclear pore. *Science*, **376**(6598), eabm9129. https://doi.org/10.1126/science.abm9129

Bonomi, M., & **Camilloni, C.** (2017). Integrative structural and dynamical biology with PLUMED-ISDB. *Bioinformatics*, **33**(24), 3999–4000. https://doi.org/10.1093/bioinformatics/btx529

Bonomi, M., **Camilloni, C.**, **Cavalli, A.**, & **Vendruscolo, M.** (2016). Metainference: a Bayesian inference method for heterogeneous systems. *Science Advances*, **2**(1), e1501177. https://doi.org/10.1126/sciadv.1501177

Bonomi, M., **Camilloni, C.**, & **Vendruscolo, M.** (2016). Metadynamic metainference: enhanced sampling of the metainference ensemble using metadynamics. *Scientific Reports*, **6**(1), 31232. https://doi.org/10.1038/srep31232

Bonomi, M., **Heller, G. T.**, **Camilloni, C.**, & **Vendruscolo, M.** (2017). Principles of protein structural ensemble determination. *Current Opinion in Structural Biology*, **42**, 106–116. https://doi.org/10.1016/j.sbi.2016.12.004

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.**, **Mavor, D.**, **Pellarin, R.**, **Schneidman, D.**, **Bader, J. S.**, & **Krogan, N. J.** (2020). Genetic interaction mapping informs integrative structure determination of protein complexes. *Science*, **370**(6522), eaaz4910. https://doi.org/10.1126/science.aaz4910

Brilot, A. F., **Lyon, A. S.**, **Zelter, A.**, **Viswanath, S.**, **Maxwell, A.**, **MacCoss, M. J.**, **Muller, E. G.**, **Sali, A.**, **Davis, T. N.**, & **Agard, D. A.** (2021). CM1-driven assembly and activation of yeast γ-tubulin small complex underlies microtubule nucleation. *eLife*, **10**, e65168. https://doi.org/10.7554/eLife.65168

Carugo, O., & **Djinović-Carugo, K.** (2023). Structural biology: a golden era. *PLOS Biology*, **21**(6), e3002187. https://doi.org/10.1371/journal.pbio.3002187

Cavalli, A., **Camilloni, C.**, & **Vendruscolo, M.** (2013). Molecular dynamics simulations with replica-averaged structural restraints generate structural ensembles according to the maximum entropy principle. *The Journal of Chemical Physics*, **138**(9), 094112. https://doi.org/10.1063/1.4793625

Cesari, A., **Gil-Ley, A.**, & **Bussi, G.** (2016). Combining simulations and solution experiments as a paradigm for RNA force field refinement. *Journal of Chemical Theory and Computation*, **12**(12), 6192–6200. https://doi.org/10.1021/acs.jctc.6b00944

Chang, L., **Wang, F.**, **Connolly, K.**, **Meng, H.**, **Su, Z.**, **Cvirkaite-Krupovic, V.**, **Krupovic, M.**, **Egelman, E. H.**, & **Si, D.** (2022). DeepTracer-ID: De novo protein identification from cryo-EM maps. *Biophysical Journal*, **121**(15), 2840–2848. https://doi.org/10.1016/j.bpj.2022.06.025

Chan-Yao-Chong, M., **Durand, D.**, & **Ha-Duong, T.** (2019). Molecular dynamics simulations combined with nuclear magnetic resonance and/or small-angle x-ray scattering data for characterizing intrinsically disordered protein conformational ensembles. *Journal of Chemical Information and Modeling*, **59**(5), 1743–1758. https://doi.org/10.1021/acs.jcim.8b00928

Chen, Z., **Shiozaki, M.**, **Haas, K. M.**, **Skinner, W. M.**, **Zhao, S.**, **Guo, C.**, **Polacco, B. J.**, **Yu, Z.**, **Krogan, N. J.**, **Lishko, P. V.**, **Kaake, R. M.**, **Vale, R. D.**, & **Agard, D. A.** (2023). De novo protein identification in mammalian sperm using in situ cryoelectron tomography and AlphaFold2 docking. *Cell*, **186**(23), 5041–5053.e19. https://doi.org/10.1016/j.cell.2023.09.017

Chim, H. Y., & **Elofsson, A.** (2024). MoLPC2: improved prediction of large protein complex structures and stoichiometry using Monte Carlo Tree Search and AlphaFold2. *Bioinformatics*, **40**(6), btae329. https://doi.org/10.1093/bioinformatics/btae329

Crehuet, R., **Buigues, P. J.**, **Salvatella, X.**, & **Lindorff-Larsen, K.** (2019). Bayesian-maximum-entropy reweighting of IDP ensembles based on NMR chemical shifts. *Entropy*, **21**(9), 898. https://doi.org/10.3390/e21090898

Cruz-León, S., **Majtner, T.**, **Hoffmann, P. C.**, **Kreysing, J. P.**, **Kehl, S.**, **Tuijtel, M. W.**, **Schaefer, S. L.**, **Geißler, K.**, **Beck, M.**, **Turoňová, B.**, & **Hummer, G.** (2024). High-confidence 3D template matching for cryo-electron tomography. *Nature Communications*, **15**(1), 3992. https://doi.org/10.1038/s41467-024-47839-8

de Teresa-Trueba, I., **Goetz, S. K.**, **Mattausch, A.**, **Stojanovska, F.**, **Zimmerli, C. E.**, **Toro-Nahuelpan, M.**, **Cheng, D. W. C.**, **Tollervey, F.**, **Pape, C.**, **Beck, M.**, **Diz-Muñoz, A.**, **Kreshuk, A.**, **Mahamid, J.**, & **Zaugg, J. B.** (2023). Convolutional networks for supervised mining of molecular patterns within cellular context. *Nature Methods*, **20**(2), 2. https://doi.org/10.1038/s41592-022-01746-2

Dobson, L., **Szekeres, L. I.**, **Gerdán, C.**, **Langó, T.**, **Zeke, A.**, & **Tusnády, G. E.** (2023). TmAlphaFold database: membrane localization and evaluation of AlphaFold2 predicted alpha-helical transmembrane protein structures. *Nucleic Acids Research*, **51**(D1), D517–D522. https://doi.org/10.1093/nar/gkac928

Dominguez, C., **Boelens, R.**, & **Bonvin, A. M. J. J.** (2003). HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *Journal of the American Chemical Society*, **125**(7), 1731–1737. https://doi.org/10.1021/ja026939x

Fisher, C. K., & **Stultz, C. M.** (2011). Constructing ensembles for intrinsically disordered proteins. *Current Opinion in Structural Biology*, **21**(3), 426–431. https://doi.org/10.1016/j.sbi.2011.04.001

Fisher, C. K., **Ullman, O.**, & **Stultz, C. M.** (2013). Comparative studies of disordered proteins with similar sequences: application to Aβ40 and Aβ42. *Biophysical Journal*, **104**(7), 1546–1555. https://doi.org/10.1016/j.bpj.2013.02.023

Flacht, L., **Lunelli, M.**, **Kaszuba, K.**, **Chen, Z. A.**, **Reilly, F. J. O.**, **Rappsilber, J.**, **Kosinski, J.**, & **Kolbe, M.** (2023). Integrative structural analysis of the type III secretion system needle complex from *Shigella flexneri*. *Protein Science*, **32**(4), e4595. https://doi.org/10.1002/pro.4595

Fontana, P., **Dong, Y.**, **Pi, X.**, **Tong, A. B.**, **Hecksel, C. W.**, **Wang, L.**, **Fu, T.-M.**, **Bustamante, C.**, & **Wu, H.** (2022). Structure of cytoplasmic ring of nuclear pore complex by integrative cryo-EM and AlphaFold. *Science*, **376**(6598), eabm9326. https://doi.org/10.1126/science.abm9326

Förster, F., **Han, B.-G.**, & **Beck, M.** (2010). Chapter Eleven—Visual Proteomics. In G. J. Jensen, *Methods in Enzymology*. **483**, pp. 215–243. Academic Press. https://doi.org/10.1016/S0076-6879(10)83011-3

Frangakis, A. S., **Böhm, J.**, **Förster, F.**, **Nickell, S.**, **Nicastro, D.**, **Typke, D.**, **Hegerl, R.**, & **Baumeister, W.** (2002). Identification of macromolecular complexes in cryoelectron tomograms of phantom cells. *Proceedings of the National Academy of Sciences*, **99**(22), 14153–14158. https://doi.org/10.1073/pnas.172520299

Fröhlking, T., **Bernetti, M.**, & **Bussi, G.** (2023). Simultaneous refinement of molecular dynamics ensembles and forward models using experimental data. *The Journal of Chemical Physics*, **158**(21), 214120. https://doi.org/10.1063/5.0151163

Ghanaeian, A., **Majhi, S.**, **McCafferty, C. L.**, **Nami, B.**, **Black, C. S.**, **Yang, S. K.**, **Legal, T.**, **Papoulas, O.**, **Janowska, M.**, **Valente-Paterno, M.**, **Marcotte, E. M.**, **Wloga, D.**, & **Bui, K. H.** (2023). Integrated modeling of the Nexin-dynein regulatory complex reveals its regulatory mechanism. *Nature Communications*, **14**(1), 5741. https://doi.org/10.1038/s41467-023-41480-7

Gubins, I., **Chaillet, M. L.**, **van der Schot, G.**, **Veltkamp, R. C.**, **Förster, F.**, **Hao, Y.**, **Wan, X.**, **Cui, X.**, **Zhang, F.**, **Moebel, E.**, **Wang, X.**, **Kihara, D.**, **Zeng, X.**, **Xu, M.**, **Nguyen, N. P.**, **White, T.**, & **Bunyak, F.** (2020). SHREC 2020: classification in cryo-electron tomograms. *Computers & Graphics*, **91**, 279–289. https://doi.org/10.1016/j.cag.2020.07.010

Gupta, R., **Liu, Y.**, **Wang, H.**, **Nordyke, C. T.**, **Puterbaugh, R. Z.**, **Cui, W.**, **Varga, K.**, **Chu, F.**, **Ke, H.**, **Vashisth, H.**, & **Cote, R. H.** (2020). Structural analysis of the regulatory GAF domains of cGMP phosphodiesterase elucidates the allosteric communication pathway. *Journal of Molecular Biology*, **432**(21), 5765–5783. https://doi.org/10.1016/j.jmb.2020.08.026

Habeck, M. (2023). Bayesian methods in integrative structure modeling. *Biological Chemistry*, **404**(8–9), 741–754. https://doi.org/10.1515/hsz-2023-0145

Hermann, M. R., & **Hub, J. S.** (2019). SAXS-restrained ensemble simulations of intrinsically disordered proteins with commitment to the principle of maximum entropy. *Journal of Chemical Theory and Computation*, **15**(9), 5103–5115. https://doi.org/10.1021/acs.jctc.9b00338

Hesketh, S. J., **Mukhopadhyay, A. G.**, **Nakamura, D.**, **Toropova, K.**, & **Roberts, A. J.** (2022). IFT-A structure reveals carriages for membrane protein transport into cilia. *Cell*, **185**(26), 4971–4985.e16. https://doi.org/10.1016/j.cell.2022.11.010

Hoff, S. E., **Thomasen, F. E.**, **Lindorff-Larsen, K.**, & **Bonomi, M.** (2024). Accurate model and ensemble refinement using cryo-electron microscopy maps and Bayesian inference. *PLOS Computational Biology*, **20**(7), e1012180. https://doi.org/10.1371/journal.pcbi.1012180

Hoff, S. E., **Zinke, M.**, **Izadi-Pruneyre, N.**, & **Bonomi, M.** (2024). Bonds and bytes: the odyssey of structural biology. *Current Opinion in Structural Biology*, **84**, 102746. https://doi.org/10.1016/j.sbi.2023.102746

Honorato, R. V., **Trellet, M. E.**, **Jiménez-García, B.**, **Schaarschmidt, J. J.**, **Giulini, M.**, **Reys, V.**, **Koukos, P. I.**, **Rodrigues, J. P. G. L. M.**, **Karaca, E.**, **Van Zundert, G. C. P.**, **Roel-Touris, J.**, **Van Noort, C. W.**, **Jandová, Z.**, **Melquiond, A. S. J.**, & **Bonvin, A. M. J. J.** (2024). The HADDOCK2.4 web server for integrative modeling of biomolecular complexes. *Nature Protocols*. https://doi.org/10.1038/s41596-024-01011-0

Huang, J., **Rauscher, S.**, **Nawrocki, G.**, **Ran, T.**, **Feig, M.**, **de Groot, B. L.**, **Grubmüller, H.**, & **MacKerell, A. D.** (2017). CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nature Methods*, **14**(1), 71–73. https://doi.org/10.1038/nmeth.4067

Hummer, G., & **Köfinger, J.** (2015). Bayesian ensemble refinement by replica simulations and reweighting. *The Journal of Chemical Physics*, **143**(24), 243150. https://doi.org/10.1063/1.4937786

Inbar, Y., **Benyamini, H.**, **Nussinov, R.**, & **Wolfson, H. J.** (2005). Combinatorial docking approach for structure prediction of large proteins and multimolecular assemblies. *Physical Biology*, **2**(4), S156–S165. https://doi.org/10.1088/1478-3975/2/4/S10

Jahn, L. R., **Marquet, C.**, **Heinzinger, M.**, & **Rost, B.** (2024). Protein embeddings predict binding residues in disordered regions. *Scientific Reports*, **14**(1), 13566. https://doi.org/10.1038/s41598-024-64211-4

Janson, G., & **Feig, M.** (2024). Transferable deep generative modeling of intrinsically disordered protein conformations. *PLoS Computational Biology*, **20**(5), e1012144. https://doi.org/10.1371/journal.pcbi.1012144

Janson, G., **Valdes-Garcia, G.**, **Heo, L.**, & **Feig, M.** (2023). Direct generation of protein conformational ensembles via machine learning. *Nature Communications*, **14**(1), 774. https://doi.org/10.1038/s41467-023-36443-x

Jing, B., **Berger, B.**, & **Jaakkola, T.** (2024). AlphaFold meets flow matching for generating protein ensembles. *Proceedings of the 41st International Conference on Machine Learning*, 22277–22303. https://proceedings.mlr.press/v235/jing24a.html

Joseph, J. A., **Reinhardt, A.**, **Aguirre, A.**, **Chew, P. Y.**, **Russell, K. O.**, **Espinosa, J. R.**, **Garaizar, A.**, & **Collepardo-Guevara, R.** (2021). Physics-driven coarse-grained model for biomolecular phase separation with near-quantitative accuracy. *Nature Computational Science*, **1**(11), Article 11. https://doi.org/10.1038/s43588-021-00155-3

Jumper, J., **Evans, R.**, **Pritzel, A.**, **Green, T.**, **Figurnov, M.**, **Ronneberger, O.**, **Tunyasuvunakool, K.**, **Bates, R.**, **Žídek, A.**, **Potapenko, A.**, **Bridgland, A.**, **Meyer, C.**, **Kohl, S. A. A.**, **Ballard, A. J.**, **Cowie, A.**, **Romera-Paredes, B.**, **Nikolov, S.**, **Jain, R.**, **Adler, J.**, & **Hassabis, D.** (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, **596**(7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2

Kaake, R. M., **Echeverria, I.**, **Kim, S. J.**, **Von Dollen, J.**, **Chesarino, N. M.**, **Feng, Y.**, **Yu, C.**, **Ta, H.**, **Chelico, L.**, **Huang, L.**, **Gross, J.**, **Sali, A.**, & **Krogan, N. J.** (2021). Characterization of an A3G-VifHIV-1-CRL5-CBFβ structure using a cross-linking mass spectrometry pipeline for integrative modeling of host–pathogen complexes. *Molecular & Cellular Proteomics*, **20**, 100132. https://doi.org/10.1016/j.mcpro.2021.100132

Khanppnavar, B., **Schuster, D.**, **Lavriha, P.**, **Uliana, F.**, **Özel, M.**, **Mehta, V.**, **Leitner, A.**, **Picotti, P.**, & **Korkhov, V. M.** (2024). Regulatory sites of CaM-sensitive adenylyl cyclase AC8 revealed by cryo-EM and structural proteomics. *EMBO Reports*, **25**(3), 1513–1540. https://doi.org/10.1038/s44319-024-00076-y

Köfinger, J., **Stelzl, L. S.**, **Reuter, K.**, **Allande, C.**, **Reichel, K.**, & **Hummer, G.** (2019). Efficient ensemble refinement by reweighting. *Journal of Chemical Theory and Computation*, **15**(5), 3390–3401. https://doi.org/10.1021/acs.jctc.8b01231

Lamm, L., **Righetto, R. D.**, **Wietrzynski, W.**, **Pöge, M.**, **Martinez-Sanchez, A.**, **Peng, T.**, & **Engel, B. D.** (2022). MemBrain: A deep learning-aided pipeline for detection of membrane proteins in Cryo-electron tomograms. *Computer Methods and Programs in Biomedicine*, **224**, 106990. https://doi.org/10.1016/j.cmpb.2022.106990

Lamm, L., **Zufferey, S.**, **Righetto, R. D.**, **Wietrzynski, W.**, **Yamauchi, K. A.**, **Burt, A.**, **Liu, Y.**, **Zhang, H.**, **Martinez-Sanchez, A.**, **Ziegler, S.**, **Isensee, F.**, **Schnabel, J. A.**, **Engel, B. D.**, & **Peng, T.** (2024). MemBrain v2: An end-to-end tool for the analysis of membranes in cryo-electron tomography (p. 2024.01.05.574336). bioRxiv. https://doi.org/10.1101/2024.01.05.574336

Leman, J. K., **Weitzner, B. D.**, **Lewis, S. M.**, **Adolf-Bryfogle, J.**, **Alam, N.**, **Alford, R. F.**, **Aprahamian, M.**, **Baker, D.**, **Barlow, K. A.**, **Barth, P.**, **Basanta, B.**, **Bender, B. J.**, **Blacklock, K.**, **Bonet, J.**, **Boyken, S. E.**, **Bradley, P.**, **Bystroff, C.**, **Conway, P.**, **Cooper, S.**, & **Bonneau, R.** (2020). Macromolecular modeling and design in Rosetta: Recent methods and frameworks. *Nature Methods*, **17**(7), 665–680. https://doi.org/10.1038/s41592-020-0848-2

Lincoff, J., **Haghighatlari, M.**, **Krzeminski, M.**, **Teixeira, J. M. C.**, **Gomes, G.-N. W.**, **Gradinaru, C. C.**, **Forman-Kay, J. D.**, & **Head-Gordon, T.** (2020). Extended experimental inferential structure determination method in determining the structural ensembles of disordered protein states. *Communications Chemistry*, **3**(1), Article 1. https://doi.org/10.1038/s42004-020-0323-0

Lindorff-Larsen, K., & **Kragelund, B. B.** (2021). On the potential of machine learning to examine the relationship between sequence, structure, dynamics and function of intrinsically disordered proteins. *Journal of Molecular Biology*, **433**(20), 167196. https://doi.org/10.1016/j.jmb.2021.167196

Liu, X., **Zhang, Y.**, **Wen, Z.**, **Hao, Y.**, **Banks, C. A. S.**, **Cesare, J.**, **Bhattacharya, S.**, **Arvindekar, S.**, **Lange, J. J.**, **Xie, Y.**, **Garcia, B. A.**, **Slaughter, B. D.**, **Unruh, J. R.**, **Viswanath, S.**, **Florens, L.**, **Workman, J. L.**, & **Washburn, M. P.** (2024). An integrated structural model of the DNA damage-responsive

https://doi.org/10.1017/qrd.2024.15 Published online by Cambridge University Press

H3K4me3 binding WDR76:SPIN1 complex with the nucleosome. *Proceedings of the National Academy of Sciences*, **121**(33), e2318601121. https://doi.org/10.1073/pnas.2318601121

Lotthammer, J. M., **Ginell, G. M.**, **Griffith, D.**, **Emenecker, R. J.**, & **Holehouse, A. S.** (2024). Direct prediction of intrinsically disordered protein conformational properties from sequence. *Nature Methods*, **21**(3), 465–476. https://doi.org/10.1038/s41592-023-02159-5

Mansoor, S., **Baek, M.**, **Park, H.**, **Lee, G. R.**, & **Baker, D.** (2024). Protein ensemble generation through variational autoencoder latent space sampling. *Journal of Chemical Theory and Computation*, **20**(7), 2689–2695. https://doi.org/10.1021/acs.jctc.3c01057

Marrink, S. J., **Risselada, H. J.**, **Yefimov, S.**, **Tieleman, D. P.**, & **de Vries, A. H.** (2007). The MARTINI force field: Coarse grained model for biomolecular simulations. *The Journal of Physical Chemistry B*, **111**(27), 7812–7824. https://doi.org/10.1021/jp071097f

Maurer, V. J., **Siggel, M.**, & **Kosinski, J.** (2024). PyTME (Python Template Matching Engine): A fast, flexible, and multi-purpose template matching library for cryogenic electron microscopy data. *SoftwareX*, **25**, 101636. https://doi.org/10.1016/j.softx.2024.101636

McCafferty, C. L., **Klumpe, S.**, **Amaro, R. E.**, **Kukulski, W.**, **Collinson, L.**, & **Engel, B. D.** (2024). Integrating cellular electron microscopy with multimodal data to explore biology across space and time. *Cell*, **187**(3), 563–584. https://doi.org/10.1016/j.cell.2024.01.005

McCafferty, C. L., **Papoulas, O.**, **Jordan, M. A.**, **Hoogerbrugge, G.**, **Nichols, C.**, **Pigino, G.**, **Taylor, D. W.**, **Wallingford, J. B.**, & **Marcotte, E. M.** (2022). Integrative modeling reveals the molecular architecture of the intraflagellar transport A (IFT-A) complex. *eLife*, **11**, e81977. https://doi.org/10.7554/eLife.81977

Michael, A. K., **Stoos, L.**, **Crosby, P.**, **Eggers, N.**, **Nie, X. Y.**, **Makasheva, K.**, **Minnich, M.**, **Healy, K. L.**, **Weiss, J.**, **Kempf, G.**, **Cavadini, S.**, **Kater, L.**, **Seebacher, J.**, **Vecchia, L.**, **Chakraborty, D.**, **Isbel, L.**, **Grand, R. S.**, **Andersch, F.**, **Fribourgh, J. L.**, & **Thomä, N. H.** (2023). Cooperation between bHLH transcription factors and histones for DNA access. *Nature*, **619**(7969), 385–393. https://doi.org/10.1038/s41586-023-06282-3

Moebel, E., **Martinez-Sanchez, A.**, **Lamm, L.**, **Righetto, R. D.**, **Wietrzynski, W.**, **Albert, S.**, **Larivière, D.**, **Fourmentin, E.**, **Pfeffer, S.**, **Ortiz, J.**, **Baumeister, W.**, **Peng, T.**, **Engel, B. D.**, & **Kervrann, C.** (2021). Deep learning improves macromolecule identification in 3D cellular cryo-electron tomograms. *Nature Methods*, **18**(11), Article 11. https://doi.org/10.1038/s41592-021-01275-4

Mosalaganti, S., **Obarska-Kosinska, A.**, **Siggel, M.**, **Taniguchi, R.**, **Turoňová, B.**, **Zimmerli, C. E.**, **Buczak, K.**, **Schmidt, F. H.**, **Margiotta, E.**, **Mackmull, M.-T.**, **Hagen, W. J. H.**, **Hummer, G.**, **Kosinski, J.**, & **Beck, M.** (2022). AI-based structure prediction empowers integrative structural analysis of human nuclear pores. *Science*, **376**(6598), eabm9506. https://doi.org/10.1126/science.abm9506

Nerenberg, P. S., **Jo, B.**, **So, C.**, **Tripathy, A.**, & **Head-Gordon, T.** (2012). Optimizing solute-water van der Waals interactions to reproduce solvation free energies. *The Journal of Physical Chemistry. B*, **116**(15), 4524–4534. https://doi.org/10.1021/jp2118373

O'Reilly, F. J., **Graziadei, A.**, **Forbrig, C.**, **Bremenkamp, R.**, **Charles, K.**, **Lenz, S.**, **Elfmann, C.**, **Fischer, L.**, **Stülke, J.**, & **Rappsilber, J.** (2023). Protein complexes in cells by AI-assisted structural proteomics. *Molecular Systems Biology*, **19**(4), e11544. https://doi.org/10.15252/msb.202311544

Oldfield, C. J., & **Dunker, A. K.** (2014). Intrinsically disordered proteins and intrinsically disordered protein regions. *Annual Review of Biochemistry*, **83**, 553–584. https://doi.org/10.1146/annurev-biochem-072711-164947

Pasani, S., **Menon, K. S.**, & **Viswanath, S.** (2024). The molecular architecture of the desmosomal outer dense plaque by integrative structural modeling. *Protein Science*, **33**(12), e5217. https://doi.org/10.1002/pro.5217

Pasani, S., & **Viswanath, S.** (2021). A framework for stochastic optimization of parameters for integrative modeling of macromolecular assemblies. *Life*, **11**(11), Article 11. https://doi.org/10.3390/life11111183

Peñaherrera, D., & **Koes, D. R.** (2024). Structure-Infused Protein Language Models. *bioRxiv*, https://doi.org/10.1101/2023.12.13.571525

Pesce, F., **Bremer, A.**, **Tesei, G.**, **Hopkins, J. B.**, **Grace, C. R.**, **Mittag, T.**, & **Lindorff-Larsen, K.** (2024). Design of intrinsically disordered protein variants with diverse structural properties. *Science Advances*, **10**(35), eadm9926. https://doi.org/10.1126/sciadv.adm9926

Petrovic, S., **Samanta, D.**, **Perriches, T.**, **Bley, C. J.**, **Thierbach, K.**, **Brown, B.**, **Nie, S.**, **Mobbs, G. W.**, **Stevens, T. A.**, **Liu, X.**, **Tomaleri, G. P.**, **Schaus, L.**, & **Hoelz, A.** (2022). Architecture of the linker-scaffold in the nuclear pore. *Science*, **376**(6598), eabm9798. https://doi.org/10.1126/science.abm9798

Pitera, J. W., & **Chodera, J. D.** (2012). On the use of experimental observations to bias simulated ensembles. *Journal of Chemical Theory and Computation*, **8**(10), 3445–3451. https://doi.org/10.1021/ct300112v

Potrzebowski, W., **Trewhella, J.**, & **Andre, I.** (2018). Bayesian inference of protein conformational ensembles from limited structural data. *PLOS Computational Biology*, **14**(12), e1006641. https://doi.org/10.1371/journal.pcbi.1006641

Pyle, E., & **Zanetti, G.** (2021). Current data processing strategies for cryo-electron tomography and subtomogram averaging. *Biochemical Journal*, **478**(10), 1827–1845. https://doi.org/10.1042/BCJ20200715

Rafiei, A., **Cruz Tetlalmatzi, S.**, **Edrington, C. H.**, **Lee, L.**, **Crowder, D. A.**, **Saltzberg, D. J.**, **Sali, A.**, **Brouhard, G.**, & **Schriemer, D. C.** (2022). Doublecortin engages the microtubule lattice through a cooperative binding mode involving its C-terminal domain. *eLife*, **11**, e66975. https://doi.org/10.7554/eLife.66975

Rantos, V., **Karius, K.**, & **Kosinski, J.** (2022). Integrative structural modeling of macromolecular complexes using assembline. *Nature Protocols*, **17**(1), Article 1. https://doi.org/10.1038/s41596-021-00640-z

Raveh, B., **Sun, L.**, **White, K. L.**, **Sanyal, T.**, **Tempkin, J.**, **Zheng, D.**, **Bharath, K.**, **Singla, J.**, **Wang, C.**, **Zhao, J.**, **Li, A.**, **Graham, N. A.**, **Kesselman, C.**, **Stevens, R. C.**, & **Sali, A.** (2021). Bayesian metamodeling of complex biological systems across varying representations. *Proceedings of the National Academy of Sciences*, **118**(35), e2104559118. https://doi.org/10.1073/pnas.2104559118

Rice, G., **Wagner, T.**, **Stabrin, M.**, **Sitsel, O.**, **Prumbaum, D.**, & **Raunser, S.** (2023). TomoTwin: Generalized 3D localization of macromolecules in cryo-electron tomograms with structural data mining. *Nature Methods*, **20**(6), Article 6. https://doi.org/10.1038/s41592-023-01878-z

Rieping, W., **Habeck, M.**, & **Nilges, M.** (2005). Inferential structure determination. *Science*, **309**(5732), 303–306. https://doi.org/10.1126/science.1110428

Rives, A., **Meier, J.**, **Sercu, T.**, **Goyal, S.**, **Lin, Z.**, **Liu, J.**, **Guo, D.**, **Ott, M.**, **Zitnick, C. L.**, **Ma, J.**, & **Fergus, R.** (2021). Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. *Proceedings of the National Academy of Sciences*, **118**(15), e2016239118. https://doi.org/10.1073/pnas.2016239118

Robustelli, P., **Piana, S.**, & **Shaw, D. E.** (2018). Developing a molecular dynamics force field for both folded and disordered protein states. *Proceedings of the National Academy of Sciences of the United States of America*, **115**(21), E4758–E4766. https://doi.org/10.1073/pnas.1800690115

Rout, M. P., & **Sali, A.** (2019). Principles for integrative structural biology studies. *Cell*, **177**(6), 1384–1403. https://doi.org/10.1016/j.cell.2019.05.016

Roux, B., & **Weare, J.** (2013). On the statistical equivalence of restrained-ensemble simulations with the maximum entropy method. *The Journal of Chemical Physics*, **138**(8), 084107. https://doi.org/10.1063/1.4792208

Różycki, B., **Kim, Y. C.**, & **Hummer, G.** (2011). SAXS ensemble refinement of ESCRT-III CHMP3 conformational transitions. *Structure (London, England: 1993)*, **19**(1), 109–116. https://doi.org/10.1016/j.str.2010.10.006

Ruff, K. M., & **Pappu, R. V.** (2021). AlphaFold and implications for intrinsically disordered proteins. *Journal of Molecular Biology*, **433**(20), 167208. https://doi.org/10.1016/j.jmb.2021.167208

Russel, D., **Lasker, K.**, **Webb, B.**, **Velázquez-Muriel, J.**, **Tjioe, E.**, **Schneidman-Duhovny, D.**, **Peterson, B.**, & **Sali, A.** (2012). Putting the pieces together: integrative modeling platform software for structure determination of macromolecular assemblies. *PLoS Biology*, **10**(1), e1001244. https://doi.org/10.1371/journal.pbio.1001244

Sali, A. (2021). From integrative structural biology to cell biology. *Journal of Biological Chemistry*, **296**, 100743. https://doi.org/10.1016/j.jbc.2021.100743

Sali, A., **Berman, H. M.**, **Schwede, T.**, **Trewhella, J.**, **Kleywegt, G.**, **Burley, S. K.**, **Markley, J.**, **Nakamura, H.**, **Adams, P.**, **Bonvin, A. M. J. J.**, **Chiu, W.**, **Peraro, M. D.**, **Di Maio, F.**, **Ferrin, T. E.**, **Grünewald, K.**, **Gutmanas, A.**, **Henderson, R.**, **Hummer, G.**, **Iwasaki, K.**, & **Westbrook, J. D.** (2015). Outcome of the First wwPDB Hybrid/Integrative methods task force workshop. *Structure*, **23**(7), 1156–1167. https://doi.org/10.1016/j.str.2015.05.013

https://doi.org/10.1017/qrd.2024.15 Published online by Cambridge University Press

Sali, A., **Glaeser, R.**, **Earnest, T.**, & **Baumeister, W.** (2003). From words to literature in structural proteomics. *Nature*, **422**(6928), 216–225. https://doi.org/10.1038/nature01513

Saltzberg, D. J., **Viswanath, S.**, **Echeverria, I.**, **Chemmama, I. E.**, **Webb, B.**, & **Sali, A.** (2021). Using integrative modeling platform to compute, validate, and archive a model of a protein complex structure. *Protein Science*, **30**(1), 250–261. https://doi.org/10.1002/pro.3995

Salvi, N., **Abyzov, A.**, & **Blackledge, M.** (2016). Multi-timescale dynamics in intrinsically disordered proteins from NMR relaxation and molecular simulation. *The Journal of Physical Chemistry Letters*, **7**(13), 2483–2489. https://doi.org/10.1021/acs.jpclett.6b00885

Schneidman-Duhovny, D., **Pellarin, R.**, & **Sali, A.** (2014). Uncertainty in integrative structural modeling. *Current Opinion in Structural Biology*, **28**, 96–104. https://doi.org/10.1016/j.sbi.2014.08.001

Schneidman-Duhovny, D., & **Wolfson, H. J.** (2020). Modeling of Multimolecular Complexes. In Z. Gáspári, *Structural Bioinformatics*. **2112**, pp. 163–174. Springer US. https://doi.org/10.1007/978-1-0716-0270-6_12

Selcuk, K., **Leitner, A.**, **Braun, L.**, **Le Blanc, F.**, **Pacak, P.**, **Pot, S.**, & **Vogel, V.** (2024). Transglutaminase 2 has higher affinity for relaxed than for stretched fibronectin fibers. *Matrix Biology*, **125**, 113–132. https://doi.org/10.1016/j.matbio.2023.12.006

Shor, B., & **Schneidman-Duhovny, D.** (2024a). CombFold: Predicting structures of large protein assemblies using a combinatorial assembly algorithm and AlphaFold2. *Nature Methods*, **21**(3), 477–487. https://doi.org/10.1038/s41592-024-02174-0

Shor, B., & **Schneidman-Duhovny, D.** (2024b). Integrative modeling meets deep learning: Recent advances in modeling protein assemblies. *Current Opinion in Structural Biology*, **87**, 102841. https://doi.org/10.1016/j.sbi.2024.102841

Shrestha, U. R., **Smith, J. C.**, & **Petridis, L.** (2021). Full structural ensembles of intrinsically disordered proteins from unbiased molecular dynamics simulations. *Communications Biology*, **4**(1), 1–8. https://doi.org/10.1038/s42003-021-01759-1

Simons, K. T., **Kooperberg, C.**, **Huang, E.**, & **Baker, D.** (1997). Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and bayesian scoring functions. *Journal of Molecular Biology*, **268**(1), 209–225. https://doi.org/10.1006/jmbi.1997.0959

Singh, D., **Soni, N.**, **Hutchings, J.**, **Echeverria, I.**, **Shaikh, F.**, **Duquette, M.**, **Suslov, S.**, **Li, Z.**, **Van Eeuwen, T.**, **Molloy, K.**, **Shi, Y.**, **Wang, J.**, **Guo, Q.**, **Chait, B. T.**, **Fernandez-Martinez, J.**, **Rout, M. P.**, **Sali, A.**, & **Villa, E.** (2024). The molecular architecture of the nuclear basket. *Cell*, **187**(19), 5267–5281.e13. https://doi.org/10.1016/j.cell.2024.07.020

Slavin, M., **Zamel, J.**, **Zohar, K.**, **Eliyahu, T.**, **Braitbard, M.**, **Brielle, E.**, **Baraz, L.**, **Stolovich-Rain, M.**, **Friedman, A.**, **Wolf, D. G.**, **Rouvinski, A.**, **Linial, M.**, **Schneidman-Duhovny, D.**, & **Kalisman, N.** (2021). Targeted in situ cross-linking mass spectrometry and integrative modeling reveal the architectures of three proteins from SARS-CoV-2. *Proceedings of the National Academy of Sciences*, **118**(34), e2103554118. https://doi.org/10.1073/pnas.2103554118

Stahl, K., **Graziadei, A.**, **Dau, T.**, **Brock, O.**, & **Rappsilber, J.** (2023). Protein structure prediction with in-cell photo-crosslinking mass spectrometry and deep learning. *Nature Biotechnology*, 1–10. https://doi.org/10.1038/s41587-023-01704-z

Stahl, K., **Warneke, R.**, **Demann, L.**, **Bremenkamp, R.**, **Hormes, B.**, **Brock, O.**, **Stülke, J.**, & **Rappsilber, J.** (2024). Modelling protein complexes with crosslinking mass spectrometry and deep learning. *Nature Communications*, **15**(1), 7866. https://doi.org/10.1038/s41467-024-51771-2

Sun, Y., & **Shen, Y.** (2023). Structure-informed protein language models are robust predictors for variant effects. *Research Square, rs.3.rs-3219092*. https://doi.org/10.21203/rs.3.rs-3219092/v1

Terwilliger, T. C., **Afonine, P. V.**, **Liebschner, D.**, **Croll, T. I.**, **McCoy, A. J.**, **Oeffner, R. D.**, **Williams, C. J.**, **Poon, B. K.**, **Richardson, J. S.**, **Read, R. J.**, & **Adams, P. D.** (2023). Accelerating crystal structure determination with iterative AlphaFold prediction. *Acta Crystallographica. Section D, Structural Biology*, **79**(3), 234–244. 10.1107/S205979832300102X

Terwilliger, T. C., **Poon, B. K.**, **Afonine, P. V.**, **Schlicksup, C. J.**, **Croll, T. I.**, **Millán, C.**, **Richardson, J. S.**, **Read, R. J.**, & **Adams, P. D.** (2022). Improved AlphaFold modeling with implicit experimental information. *Nature Methods*, **19**(11), 1376–1382. https://doi.org/10.1038/s41592-022-01645-6

Tesei, G., **Trolle, A. I.**, **Jonsson, N.**, **Betz, J.**, **Knudsen, F. E.**, **Pesce, F.**, **Johansson, K. E.**, & **Lindorff-Larsen, K.** (2024). Conformational ensembles of the human intrinsically disordered proteome. *Nature*, **626**(8000), 897–904. https://doi.org/10.1038/s41586-023-07004-5

Thamkachy, R., **Medina-Pritchard, B.**, **Park, S. H.**, **Chiodi, C. G.**, **Zou, J.**, **De La Torre-Barranco, M.**, **Shimanaka, K.**, **Abad, M. A.**, **Gallego Páramo, C.**, **Feederle, R.**, **Ruksenaite, E.**, **Heun, P.**, **Davies, O. R.**, **Rappsilber, J.**, **Schneidman-Duhovny, D.**, **Cho, U.-S.**, & **Jeyaprakash, A. A.** (2024). Structural basis for Mis18 complex assembly and its implications for centromere maintenance. *EMBO Reports*, **25**(8), 3348–3372. https://doi.org/10.1038/s44319-024-00183-w

Thomasen, F. E., **Pesce, F.**, **Roesgaard, M. A.**, **Tesei, G.**, & **Lindorff-Larsen, K.** (2022). Improving martini 3 for disordered and multidomain proteins. *Journal of Chemical Theory and Computation*, **18**(4), 2033–2041. https://doi.org/10.1021/acs.jctc.1c01042

Trabuco, L. G., **Villa, E.**, **Mitra, K.**, **Frank, J.**, & **Schulten, K.** (2008). Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. *Structure (London, England: 1993)*, **16**(5), 673–683. https://doi.org/10.1016/j.str.2008.03.005

Ullanat, V., **Kasukurthi, N.**, & **Viswanath, S.** (2022). PrISM: Precision for integrative structural models. *Bioinformatics*, **38**(15), 3837–3839. https://doi.org/10.1093/bioinformatics/btac400

Vallat, B., **Webb, B.**, **Fayazi, M.**, **Voinea, S.**, **Tangmunarunkit, H.**, **Ganesan, S. J.**, **Lawson, C. L.**, **Westbrook, J. D.**, **Kesselman, C.**, **Sali, A.**, & **Berman, H. M.** (2021). New system for archiving integrative structures. *Acta Crystallographica Section D Structural Biology*, **77**(12), 1486–1496. https://doi.org/10.1107/S2059798321010871

Viswanath, S., & **Sali, A.** (2019). Optimizing model representation for integrative structure determination of macromolecular assemblies. *Proceedings of the National Academy of Sciences*, **116**(2), 540–545. https://doi.org/10.1073/pnas.1814649116

Vitalis, A., & **Pappu, R. V.** (2009). ABSINTH: A new continuum solvation model for simulations of polypeptides in aqueous solutions. *Journal of Computational Chemistry*, **30**(5), 673–699. https://doi.org/10.1002/jcc.21005

Wang, D., **Wang, Y.**, **Evans, L.**, & **Tiwary, P.** (2024). From latent dynamics to meaningful representations. *Journal of Chemical Theory and Computation*, **20**(9), 3503–3513. https://doi.org/10.1021/acs.jctc.4c00249

Watson, J. L., **Juergens, D.**, **Bennett, N. R.**, **Trippe, B. L.**, **Yim, J.**, **Eisenach, H. E.**, **Ahern, W.**, **Borst, A. J.**, **Ragotte, R. J.**, **Milles, L. F.**, **Wicky, B. I. M.**, **Hanikel, N.**, **Pellock, S. J.**, **Courbet, A.**, **Sheffler, W.**, **Wang, J.**, **Venkatesh, P.**, **Sappington, I.**, **Torres, S. V.**, & **Baker, D.** (2023). De novo design of protein structure and function with RFdiffusion. *Nature*, **620**(7976), Article 7976. https://doi.org/10.1038/s41586-023-06415-8

Wilson, C. J., **Choy, W.-Y.**, & **Karttunen, M.** (2022). AlphaFold2: A role for disordered protein/region prediction? *International Journal of Molecular Sciences*, **23**(9), 4591. https://doi.org/10.3390/ijms23094591

Wu, K. E., **Yang, K. K.**, **van den Berg, R.**, **Alamdari, S.**, **Zou, J. Y.**, **Lu, A. X.**, & **Amini, A. P.** (2024). Protein structure generation via folding diffusion. *Nature Communications*, **15**(1), 1059. https://doi.org/10.1038/s41467-024-45051-2

Xu, M., **Singla, J.**, **Tocheva, E. I.**, **Chang, Y.-W.**, **Stevens, R. C.**, **Jensen, G. J.**, & **Alber, F.** (2019). De novo structural pattern mining in cellular electron cryotomograms. *Structure*, **27**(4), 679–691.e14. https://doi.org/10.1016/j.str.2019.01.005

Yu, M., **Heidari, M.**, **Mikhaleva, S.**, **Tan, P. S.**, **Mingu, S.**, **Ruan, H.**, **Reinkemeier, C. D.**, **Obarska-Kosinska, A.**, **Siggel, M.**, **Beck, M.**, **Hummer, G.**, & **Lemke, E. A.** (2023). Visualizing the disordered nuclear transport machinery in situ. *Nature*, **617**(7959), 162–169. https://doi.org/10.1038/s41586-023-05990-0

Yu, Y., **Li, S.**, **Ser, Z.**, **Sanyal, T.**, **Choi, K.**, **Wan, B.**, **Kuang, H.**, **Sali, A.**, **Kentsis, A.**, **Patel, D. J.**, & **Zhao, X.** (2021). Integrative analysis reveals unique structural and functional features of the Smc5/6 complex. *Proceedings of the National Academy of Sciences*, **118**(19), e2026844118. https://doi.org/10.1073/pnas.2026844118

Zeng, X., **Kahng, A.**, **Xue, L.**, **Mahamid, J.**, **Chang, Y.-W.**, & **Xu, M.** (2023). High-throughput cryo-ET structural pattern mining by unsupervised deep iterative subtomogram clustering. *Proceedings of the National Academy of Sciences*, **120**(15), e2213149120. https://doi.org/10.1073/pnas.2213149120

https://doi.org/10.1017/qrd.2024.15 Published online by Cambridge University Press

Zhang, O., **Haghighatlari, M.**, **Li, J.**, **Liu, Z. H.**, **Namini, A.**, **Teixeira, J. M. C.**, **Forman-Kay, J. D.**, & **Head-Gordon, T.** (2023). Learning to evolve structural ensembles of unfolded and disordered proteins using experimental solution data. *The Journal of Chemical Physics*, **158**(17), 174113. https://doi.org/10.1063/5.0141474

Zhang, Y., **Zhang, Z.**, **Kagaya, Y.**, **Terashi, G.**, **Zhao, B.**, **Xiong, Y.**, & **Kihara, D.** (2023). Distance-AF: Modifying predicted protein structure models by Alphafold2 with user-specified distance constraints (p. 2023.12.01.569498). bioRxiv. https://doi.org/10.1101/2023.12.01.569498

Zheng, S., **He, J.**, **Liu, C.**, **Shi, Y.**, **Lu, Z.**, **Feng, W.**, **Ju, F.**, **Wang, J.**, **Zhu, J.**, **Min, Y.**, **Zhang, H.**, **Tang, S.**, **Hao, H.**, **Jin, P.**, **Chen, C.**, **Noé, F.**, **Liu, H.**, & **Liu, T.-Y.** (2024). Predicting equilibrium distributions for molecular systems with deep learning. *Nature Machine Intelligence*, **6**(5), 558–567. https://doi.org/10.1038/s42256-024-00837-3

Zhu, J., **Li, Z.**, **Zhang, B.**, **Zheng, Z.**, **Zhong, B.**, **Bai, J.**, **Wang, T.**, **Wei, T.**, **Yang, J.**, & **Chen, H.-F.** (2024). Precise generation of conformational ensembles for intrinsically disordered proteins using fine-tuned diffusion models. *bioRxiv* https://doi.org/10.1101/2024.05.05.592611

Zhu, X., **Huang, G.**, **Zeng, C.**, **Zhan, X.**, **Liang, K.**, **Xu, Q.**, **Zhao, Y.**, **Wang, P.**, **Wang, Q.**, **Zhou, Q.**, **Tao, Q.**, **Liu, M.**, **Lei, J.**, **Yan, C.**, & **Shi, Y.** (2022). Structure of the cytoplasmic ring of the *Xenopus laevis* nuclear pore complex. *Science*, **376**(6598), eabl8280. https://doi.org/10.1126/science.abl8280

https://doi.org/10.1017/qrd.2024.15 Published online by Cambridge University Press