

Reporting on the future of integrative structural biology ORAU workshop

George L Hamilton,¹ Joshua Alper,^{1,2,3} Hugo Sanabria¹

¹Department of Physics and Astronomy, 118 Kinard Laboratory, Clemson University, Clemson, SC 29631, ²Department of Biological Sciences, 132 Long Hall, Clemson University, Clemson, SC 29631, ³Eukaryotic Pathogen Innovations Center, Life Sciences Building, Clemson University, Clemson, SC 29631

TABLE OF CONTENTS

1. Abstract
2. Introduction
3. Challenges in integrative structural biology
4. Summary of speaker presentations
 - 4.1. Electron Microscopy
 - 4.2. Fluorescence microscopy imaging
 - 4.3. Label-based methods
 - 4.4. X-ray crystallography
 - 4.5. Computational methods
 - 4.6. Data archiving
5. Roundtable discussions
 - 5.1. Data integration and representations
 - 5.2. Time evolution and biological networks
 - 5.3. The future ahead
6. A Path forward: the “virtual cell”
7. Give tangible form to the ultimate goal of structural biology
8. Acknowledgments
9. References

1. ABSTRACT

Integrative and hybrid methods have the potential to bridge long-standing knowledge gaps in structural biology. These methods will have a prominent role in the future of the field as we make advances toward a complete, unified representation of biology that spans the molecular and cellular scales. The Department of Physics and Astronomy at Clemson University hosted The Future of Integrative Structural Biology workshop on April 29, 2017 and partially sponsored by partially sponsored by a program of the Oak Ridge Associated Universities (ORAU). The workshop brought experts from multiple structural biology disciplines together to discuss near-term steps toward the goal of a molecular atlas of the

cell. The discussion focused on the types of structural data that should be represented, how this data should be represented, and how the time domain might be incorporated into such an atlas. The consensus was that an explorable, map-like Virtual Cell, containing both spatial and temporal data bridging the atomic and cellular length scales obtained by multiple experimental methods, represents the best path toward a complete atlas of the cell.

2. INTRODUCTION

A dogma of structural biology is that the function of biological macromolecules is linked to

Table 1. Nobel prizes awarded in structural biology.

Year	Field	Recipient	Subject
1946	Chemistry ¹	James Batcheller Sumner	Discovery that enzymes can be crystallized
1962	Medicine	Francis Harry Compton Crick, James Dewey Watson, Maurice Hugh Frederick Wilkins	Discoveries concerning the molecular structure of nucleic acids and its significance for information transfer in living material
1962	Chemistry	Max Ferdinand Perutz, John Cowdery Kendrew	Studies of the structures of globular proteins
1964	Chemistry	Dorothy Crowfoot Hodgkin	Determinations by X-ray techniques of the structures of important biochemical substances
1972	Chemistry ¹	Christian B. Anfinsen	Work on ribonuclease, especially concerning the connection between the amino acid sequence and the biologically active conformation
1982	Chemistry	Aaron Klug	Development of crystallographic electron microscopy and structural elucidation of biologically important nucleic acid-protein complexes
1988	Chemistry	Johann Deisenhofer, Robert Huber, Hartmut Michel	Determination of the three-dimensional structure of a photosynthetic reaction center
1991	Chemistry	Richard R. Ernst	Contributions to the development of the methodology of high resolution nuclear magnetic resonance (NMR) spectroscopy
1997	Chemistry ²	John E. Walker	Elucidation of the enzymatic mechanism underlying the synthesis of adenosine triphosphate (ATP)
2002	Chemistry ¹	Kurt Wüthrich	Development of nuclear magnetic resonance spectroscopy for determining the three-dimensional structure of biological macromolecules in solution
2003	Chemistry ¹	Roderick MacKinnon	Structural and mechanistic studies of ion channels
2006	Chemistry	Roger D. Kornberg	Studies of the molecular basis of eukaryotic transcription
2009	Chemistry	Venkatraman Ramakrishnan, Thomas A. Steitz, Ada E. Yonath	Studies of the structure and function of the ribosome
2013	Chemistry	Michael Levitt, Martin Karplus, Arieh Warshel	Dramatically advanced the field of structural biology by developing sophisticated computer algorithms to build models of complex biological molecules
2017	Chemistry	Jacques Dubochet, Joachim Frank, Richard Henderson	Developing cryoelectron microscopy for the high-resolution structure determination of biomolecules in solution
Descriptions of the work for all winners can be found at www.nobelprize.org . ¹ Recipient was awarded half of the prize for the given year.			
² Recipient was awarded one quarter of the prize for the given year.			

their three-dimensional structure down to the atomic scale (1-3). More recently, macromolecular dynamics has been considered as fundamental to biological function as structure (4, 5). Therefore, with structural information as a function of time, it is now possible to relate macromolecules' functions (i.e., metabolite binding, target recognition, signal transduction, catalytic activity) to their role in living systems. Integrative structural biology aims to combine complementary data from multiple experimental and theoretical methodologies to bridge critical gaps in knowledge and to resolve conflicts left by individual methodologies at every level of structural biology (6). Unifying traditionally separate techniques can provide the insight needed to build a complete, functional atlas of life that bridges scales from the cell-level down to atomic resolution. This atlas would be an invaluable resource for the understanding of life now and as it evolved, the education of the next

generations of biomedical researchers, the development of therapeutics including personalized medicines, and more. Already, the pursuit of knowledge necessary for such an atlas drove the development of a considerable body of work spanning more than half a century and resulting in multiple Nobel laureates awarded for their work in structural biology, though most of these approaches have been comprised of individual techniques (Table 1).

Integrative structural biology is a promising approach for advancing biomedical research and overcoming obstacles that traditional methods face. Integration of structural information at multiple biological scales and from multiple techniques makes a common, descriptive, "zoomable" representation of life bridging the atomic and cell scales a real possibility (2). The promises of integrative structural

biology have led to rapid developments by researchers working in a broad range of disciplines, including both novel techniques and methods for integrating existing techniques. For example, the integration of cryo-electron tomography (cryo-ET) with fluorescence imaging and X-ray crystallography bridges subcellular scales (7-11). The combination of Förster resonance energy transfer (FRET) with multiparameter fluorescence detection (MFD), fluorescence correlation spectroscopy (FCS), and computational techniques allows the probing of protein conformational dynamics at high resolutions in both space and time (12-16). Integration of stimulated emission depletion (STED) and fluorescence emission difference (FED) capabilities allows researchers to probe biological systems below diffraction-limited resolutions (17-19). Further, the combination of STED and FCS allows probing the dynamics in living cells below the diffraction limit with high time resolution (20, 21). The development of modeling tools that integrate constraints from electron paramagnetic resonance (EPR), nuclear magnetic resonance (NMR), electron spin echo envelope modulation (ESEEM), double electron-electron resonance (DEER) and FRET takes advantage of each technique's strength and enables accurate determination of macromolecular structures that otherwise are difficult to characterize (9, 22-24). Additionally, novel data science techniques, like machine learning algorithms, are promising tools that might play key roles in informing every aspect of integrative structural biology research from experimental design to model generation using complementary datasets (25). The examples given here only serve to briefly highlight few advances in the field of integrative methods and they are far from constituting a comprehensive review of all examples that have applied these tools. There are many other outstanding examples of integrative techniques making significant scientific advances in our understanding of biological systems, and the rapidly-evolving nature of integrative structural biology means we can expect many more to be added each year (26-34).

Owing to the rapid increase in interest in this field, on April 29, 2017, Clemson University held its first workshop on the Future of Integrative Structural Biology (FISB), hosting scientists

representing multiple structural biology subfields. The goal of this workshop was to address two key questions: (1) What does the future of integrative structural biology look like in its approach toward a complete molecular atlas of the cell? (2) What specific efforts will enable structural biology to reach this future? In this report on the FISB workshop, we summarize discussions of some key challenges faced by integrative structural biology, summarize the presentations given by keynote speakers, and discuss their relevance to integrative approaches. We concluded the workshop with round-table discussion sessions regarding the future path of the field.

3. CHALLENGES IN INTEGRATIVE STRUCTURAL BIOLOGY

Based on the presentations and discussions at the FISB, it became clear that multiple critical challenges arise while integrating distinct structural techniques and associated with their diverse primary data types, visualization schemes, and interpretations. Therefore, the path towards integrative structural biology entails challenges concerning the compatibility of data produced by multiple techniques, the discrepancies generated by alternative approaches, and the dissemination of information between specializations and to the broader public. While these challenges result from standards that are not problematic themselves, integrative structural biology inherently requires a degree of compatibility between the approaches of the various subfields. Here we briefly discuss some of these challenges, as addressing them will be critical to increasing the impact of structural biology research, communicating biological findings to researchers and the public, and ultimately developing a unified and complete picture of structural biology.

Online databases, including the Protein Data Bank (PDB, www.rcsb.org) and Structural Classification of Proteins (SCOP, <http://scop.mrc-lmb.cam.ac.uk/scop/>), are ubiquitous in structural biology due to the large volume of data that is both currently available and produced every year (35-38) (Figure 1). However, only recently have large-scale efforts been undertaken to incorporate the findings of integrative studies at various scales or from different

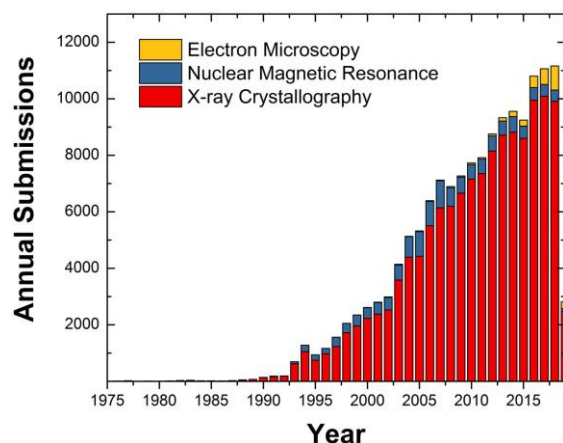


Figure 1. The annual number of structural models archived in the PDB through 2018. Structure totals for each technique as of the beginning of 2019 are: X-ray crystallography – 131959, NMR – 12478, EM – 2732. Data taken from rcsb.org.

methods in one place, such as with PDB-Dev (39). Further, each database and journal has its own set of requirements for deposited data because each tends to cater to a subset of techniques (e.g., PDB is dominated by submissions from X-ray crystallography, NMR, and EM studies (40-42) (Figure 1)). This separate treatment of data from different techniques has led to inconsistencies in nomenclature, representations, validation methods, and other aspects of results between (and sometimes even within) subfields that are otherwise analogous. For example, the term “domain” is ubiquitous in descriptions of the functional residue sequences of proteins, but the terms “clusters” (43) and “sectors” (44) are both used to describe coevolving sequences. This problem further extends to the naming of specific research subjects, which may be referred to by different names in different articles or databases (e.g., postsynaptic density protein 95 (45), disks large homolog 4, and synapse-associated protein 90 (46) are the same protein). The types of data that are provided in publications have become a concern because primary data often either is left to less-diligently reviewed supplementary materials or is absent, depending on the standards of different mediums, and causing a “reproducibility crisis” amongst many scientists (47-52). While these issues apply to many fields, including structural

biology, it is especially relevant to integrative methods because they incorporate and represent data from many sources, leading to increased complexity. These inconsistencies have led to the obfuscation of discoveries between subfields, essentially providing an additional barrier to utilizing those results, while also presenting the challenge of how to properly synthesize such results.

Simultaneous representation of multiple data sources is another challenge in integrative structural biology. For example, many biological macromolecules have been studied functionally, (53) evolutionarily, (54, 55) and structurally (56), but little has been done to represent the results of these approaches simultaneously. Yet, these features of biological macromolecules are intrinsically, closely related. Thus, a unified structural/functional representation would fill a prominent gap in biological research reporting and facilitate identification of missing information about biological systems, rational drug design (57), and understanding of larger organization schemes such as metabolic pathways in ways that representations of individual molecules may not capture (58). Evaluating which data must be represented and how to do so in integrative studies will require a case-by-case approach due to the diverse nature of these studies. However, incorporating the results of these studies into a unified database may require a generally applicable approach to evaluating submitted data.

Some efforts have been made to tackle these issues. For example, PDB has dictionaries and standards of representation that provide internal consistency between all the techniques used (42). Further, the wwPDB Integrative/Hybrid Methods Task Force hosted a workshop to address the issue of validation and archiving of structure determination by hybrid techniques (39). That task force developed PDB-Dev, and the associated integrative/hybrid methods dictionary, as a repository for macromolecule structures obtained through integrative/hybrid methods (59). Other communities standardized nomenclatures for their subfields by providing a dictionary of terms based on bioinformatics analyses (60) (61). Additionally, resources like UniProt attempt to alleviate the issue of nomenclature without prescribing a single

approach by cataloguing equivalent names for proteins. (62) Per the plans of the wwPDB small angle scattering task force (63) the Small Angle Scattering Biological Data Bank was developed to be a repository for data from scattering techniques and hybrid models developed from scattering data in conjunction with other methods (64). As more researchers recognize these issues and attempt to find solutions to them, conferences and workshops that survey the field and additional task forces that address the issues will be essential. Such coordinated collaboration will be necessary if the goal of a molecular atlas of the cell is ever to be realized because this goal will require input from a diverse range of techniques that span scales. Realization of this unified picture entails essential questions that must be answered: What information must be represented? Who is the core audience for such a representation? What steps need to be taken to achieve this representation? These were some of the questions at the core of Clemson's 2017 FISB workshop.

4. SUMMARY OF SPEAKER PRESENTATIONS

Experts representing multiple subdisciplines of structural biology participated in Clemson's 2017 workshop on the Future of Integrative Structural Biology. Among the represented methodologies were X-ray crystallography, electron microscopy, label-based approaches, super-resolution microscopy, and computational methods. Invited speakers discussed the challenged mentioned above with the overarching theme of building a molecular atlas of the cell bridging methodologies, length scales, and timescales. Here, we briefly report on their contributions.

4.1. Electron microscopy

The invention of the electron microscope in 1931 by Ernst Ruska and Maximillion Knoll (65) opened the doors to imaging below the resolution limit of light microscopy. More recently, cryo-EM has emerged as an exciting technique (66-69) for the study of structural biology that can bridge the cellular and atomic scales thanks to its lack of a need to

crystallize samples, its high precision (reaching even sub-Ångstrom precision position determination) (69, 70), and the ability to image wide fields of view. Furthermore, cryo-EM is ripe for use with other, complementary techniques thanks to this broad range of applicability.

Dr. Daniela Nicastro (University of Texas Southwestern Medical Center) discussed her group's work using cryo-EM tomography to study the structure and organization of protein complexes within cells, including cilia and flagella. She presented their approach for integrating high-resolution single-particle cryo-EM with lower resolution wide-field cryo-electron tomography to obtain tomography-guided 3D reconstructions of subcellular structures (TYGRESS). Combined with comparative genetics, biochemical methods, and EM-visible labeling, these approaches showcase the power of integrative structural biology at the scale of macromolecular complexes, allowing for the visualization of 3D structures of intact complexes in different states and providing insight into both their structure and function. Dr. Nicastro showed preliminary results of what turned out to be a beautiful demonstration of the power of cryo-EM in determining the spatial distribution of dynein in cilia and its importance in motility (Figure 2).

Dr. Elizabeth R. Wright (Emory University at the time of FISB and currently at University of Wisconsin - Madison) discussed her group's work applying an integrative cryo-EM and molecular biology approach to structural virology and cell biology, including studies of pleomorphic enveloped viruses. Further, she has pioneered numerous techniques, including affinity grid methods for EM, enhanced phase-contrast EM, and cryo-correlative light and electron microscopy (cryo-CLEM). Such advances in techniques are crucial to integrative structural biology as improvements in spatiotemporal measurement capabilities help provide a more detailed picture at the molecular scale. She presented cryo-EM structural investigations of pleomorphic enveloped viruses, namely respiratory syncytial virus (RSV) and measles virus, as well as results of cryo-EM methods development as applied to studies of viruses and cells (Figure 3).

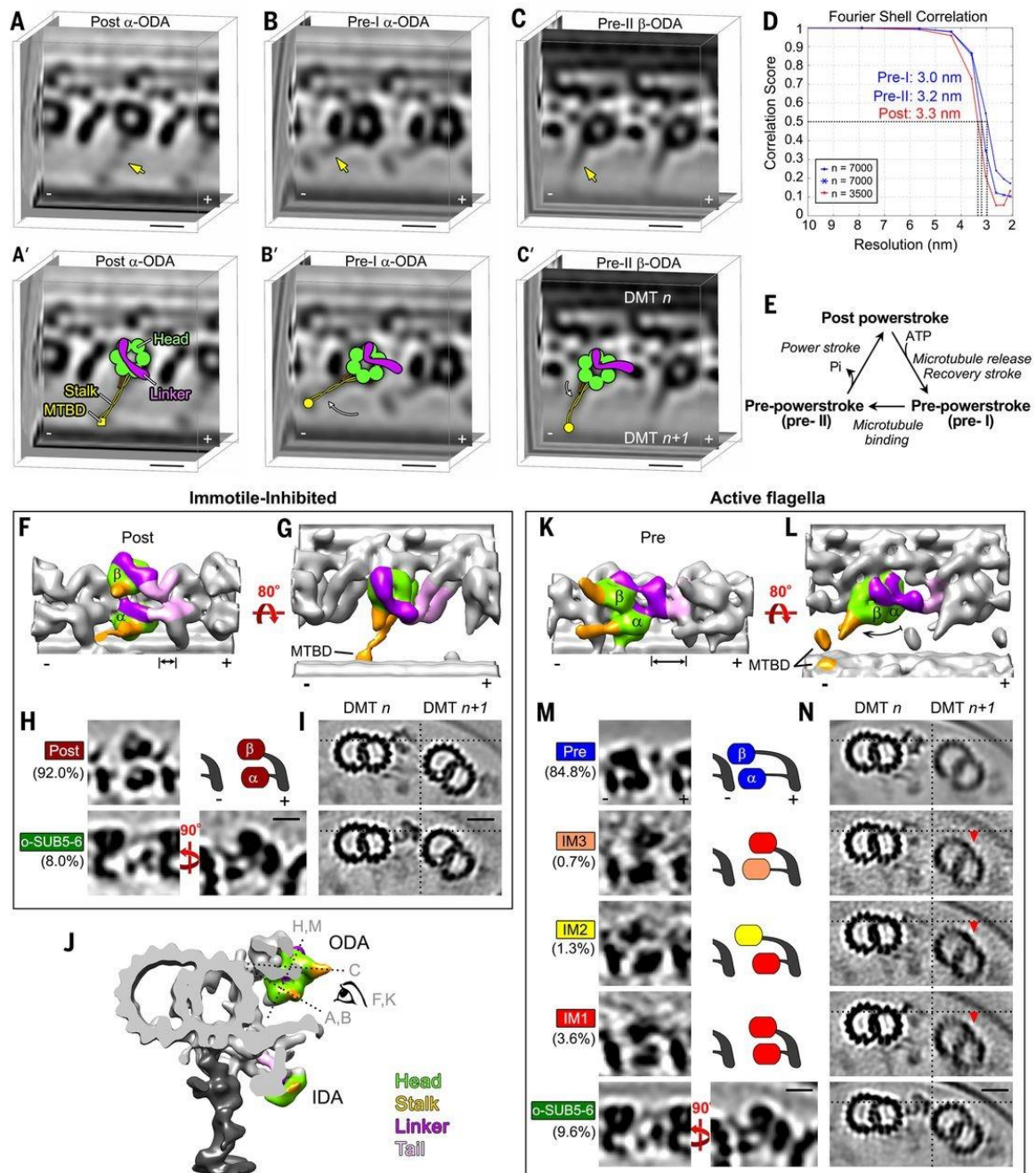


Figure 2. From (96). Reprinted with permission from AAAS. Various tomographic slices resolve different states of outer-arm dynein in immotile-inhibited and active flagella. 3D isosurface renderings and schematic models corresponding to these states are also shown. The insight provided by such images into the function of dynein in cilia and flagella demonstrates the power that comparative cryo-ET has to elucidate the molecular mechanisms of cell-level functions.

Dr. Vera B.S. Chan (a postdoctoral fellow at Clemson University under the supervision of Dr. Andrew Mount at the time and currently at the Institut

Français de Recherche pour l'Exploitation de la Mer), also presented on the topic of correlative microscopy with another approach. She showed that chitin fibers

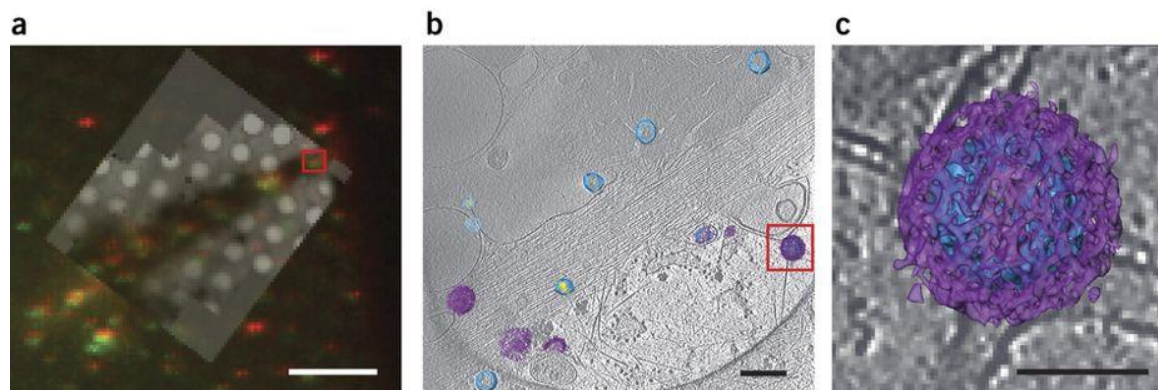


Figure 3. Reprinted by permission from Springer Nature Customer Service Centre GmbH: Springer Nature, Nature Protocols, Correlated fluorescence microscopy and cryo-electron tomography of virus-infected or transfected mammalian cells, Cheri M. Hampton et al., Copyright 2017 (97). Cryo-CLEM imaging of double-labeled HIV-1 particles pseudotyped with avian sarcoma and leukosis virus Env glycoprotein. These images showcase the capabilities of cryo-CLEM in obtaining direct, physiological information about cell-level structure down to individual viral particles.

in eastern oysters are essential components of the molluscan shell and likely significantly contribute to the mechanical strength of those shells. She presented results combining information from confocal laser scanning microscopy (CLSM) images obtained using a chitin-specific fluorescent probe (Figure 4) and results from scanning electron microscopy. This innovative technique allowed Dr. Chan to study nuclei positions and colocalization of chitin within the cell membrane throughout the process of shell repair.

4.2. Fluorescence microscopy imaging

Fluorescence microscopy imaging techniques, and particularly super-resolution imaging that overcomes the diffraction limit (nanoscopy), provide a direct view into the mechanisms of life at the nanometer scale (71, 72). These approaches enable the study of structure and dynamics of macromolecular structures and cells in various experimental conditions, including *in vivo* (73). Further, the field of fluorescence imaging is rapidly evolving, with new techniques such as STORM (74), stepwise optical saturation (75), STED (17), PAINT (76), RESOLFT (77), MINIFLUX (78), and others being introduced rapidly to push past resolution and color limitations. Super-resolution imaging provides intuitive information that helps bridge the scales of cellular and subcellular biology, complementing the many other techniques crucial to integrative structural biology.

Dr. Mark Bathe discussed his group's work programming structured nucleic acid assemblies to engineer synthetic viral capsid mimics for high-resolution imaging, metallic nanoparticle synthesis, and therapeutic delivery (79). He also presented on the application of nucleic acids to multi-scale confocal and super-resolution fluorescence imaging of neuronal synapse proteins to overcome the four-color spectral limit of fluorescence imaging (Figure 5) (80). Hybrid computational-experimental approaches and innovative approaches to experiments on complex systems, like those used by Dr. Bathe, are at the leading edge of integrative biology and will be crucial to the development of new techniques for studies of the molecular structure of the cell.

4.3. Label-based methods

Label-based methods, including FRET, spin-labeling EPR, and DEER, represent a versatile array of techniques for probing the structure and function of biological macromolecules at high resolution in both the spatial and temporal domains. Also, they provide the ability to target specific structural elements and interactions between molecules with site-specific labeling techniques. Along with their potential for integration with other techniques, this leads to their importance in integrative structural biology.

Dr. Mark Bowen discussed his group's work on validating the viability of FRET measurement as a

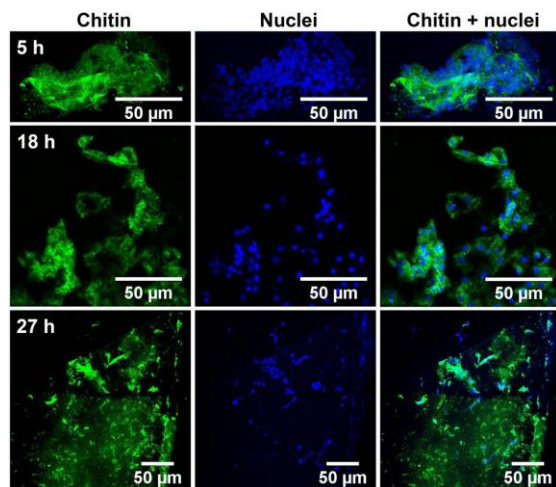


Figure 4. Reproduced with permission from, (98), Copyright 2018 Chan, Johnstone, Wheeler, and Mount. Maximum intensity projections of CLSM images show the succession of shell repair on glass implants in *Crassostrea virginica* shells at various times after implantation. CLSM identified colocalization of chitin with the cell membrane during shell repair, suggesting an important role of cells located on the implants in chitin fibril formation.

molecular ruler, showing that sub-nanometer precision can be achieved if enough constraints are given using multiple FRET pairs. Further, he showed that FRET can accurately measure *in vivo* molecular distances by confirming the constancy of fluorescent protein FRET in live cells. The high-precision nature of FRET and its ability to measure distances *in vivo* positions it as both a powerful standalone technique and one that is ripe for integration with other techniques to gain insight into molecular structures and interactions.

As further validation of FRET's accuracy and precision as a quantitative tool, a multi-laboratory, blind benchmark study was recently completed comparing the FRET distances obtained by different techniques for labeled DNA duplexes (81). This precision recently allowed the identification of weak interactions that give rise to transiently occupied conformational states in the PDZ-2 tandem of PSD95 by combining FRET, discrete molecular dynamics (DMD) simulations, and biochemical methods, thus resolving conflicts between structures obtained in prior studies (Figure 6) (82).

Dr. Tatyana Smirnova discussed her group's work using DEER and FRET in conjunction to measure distance distributions in both disordered

and folded synaptosomal-associated protein 25. The overlapping measurement ranges of FRET and DEER, while rarely utilized, make them ideal for direct comparison and validation by increasing the confidence level in the results obtained through redundant measurements. Additionally, she reported results using DEER to refine the solid-state NMR structure of a heptahelical integral membrane protein, *Anabaena* sensory rhodopsin (ASR), reconstituted in a lipid environment, where the addition of even two DEER distances to NMR constraints significantly improved the resolved ASR trimer structure and revealed a more compact packing of helices and side chains at the intermonomer interface than previously observed (Figure 7) (83). Development of new techniques for integrative structural biology will rely on comparisons and validations of overlapping measurements made with multiple techniques, like Dr. Smirnova's comparison of DEER and FRET. Such overlaps particularly important when evaluating the viability of techniques for measurements in various experimental conditions, such as in live cells or at extreme pH conditions.

4.4. X-Ray crystallography

Diffraction methods such as X-ray crystallography have long dominated the field of atomic structure determination for biological macromolecules. Most structures in the PDB were determined with X-ray crystallography (Figure 1) because it provides the mean position of every atom in a biomacromolecule crystal at Angstrom resolution. However, some limitations of the technique include the need to crystallize the sample, i.e., not all biomolecules can be crystallized, the inability to resolve flexible or disordered regions of proteins, the effects of averaging multiple possible structural states of the molecule, and the non-physiological constraints on the molecules in the crystal. Its benefits and limitations make X-ray crystallography ripe for integration with other techniques.

Dr. Daniel Keedy (previously a postdoctoral fellow at the University of California, San Francisco under the supervision of Dr. James Fraser and currently an Assistant Professor at the City University

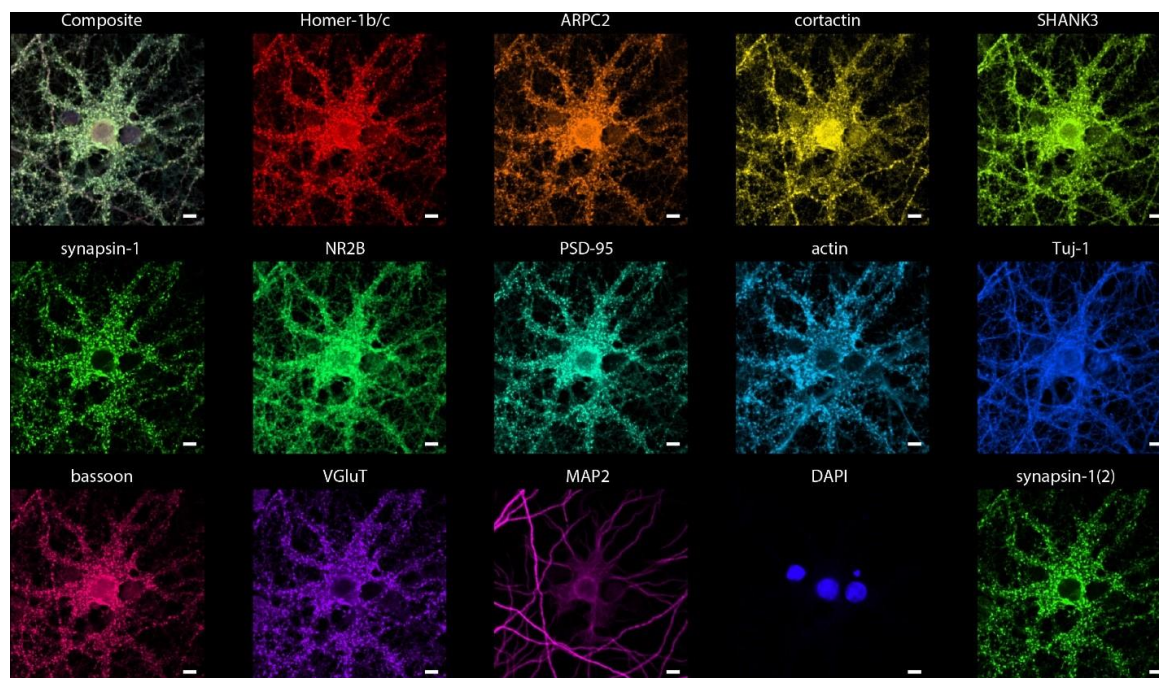


Figure 5. Used with permission from Dr. Mark Bathe. 13-channel image of 13 targeted components of rat hippocampal neuronal synapses obtained via locked nucleic acid probe-based imaging for sequential multiplexing (LNA-PRISM). The composite of all 13 is shown in the top-left. Below are zoomed in views of the individual dendrite indicated in the composite. The high affinity of LNA probes allows specific targeting and simultaneous imaging of many proteins in the same environment. Combined with DNA-PAINT(99), this study showed the power of PRISM by resolving nanometer-scale structural organization and heterogeneity of individual synapses.

of New York) discussed his work utilizing multiple-temperature X-ray crystallography experiments, structure determination from hundreds of individual small-molecule fragment soaks, and biochemical techniques to identify allosteric networks and binding sites in dynamic proteins. Specifically, he showed that small molecules binding at an allosteric site alter the conformational state and inhibit the enzymatic activity of human protein tyrosine phosphatase 1B (PTP1B). Thus, he revealed how perturbations can bias the conformational ensemble of a protein to control its function by integrating experimental and computational approaches (Figure 8). This approach allows for observation of “hidden” or low-occupancy conformational states for protein and ligands.

4.5. Computational methods

Computational methods are used in structural biology for everything from structure prediction and assessment to the modeling of protein dynamics and interactions. Further, computational methods provide the means of integrating data

obtained from multiple techniques to determine structure on scales ranging from individual proteins to macromolecular assemblies (84). Computational approaches include molecular dynamics (85) simulations, sequence alignments (86), and many more modeling, simulation, and bioinformatics techniques (85, 87-89). The significant potential and broad applicability of computational techniques leave them poised to efficiently and effectively bridge the gaps between the many subfields of structural biology. As such, computation has a role in conjunction with all experimental techniques, solidifying its place as a key to advances in integrative structural biology.

Dr. Sichun Yang discussed his group's integration of scattering, footprinting, and docking simulation (iSPOT) for modeling of multi-protein complexes. He demonstrated iSPOT's power for the HNF-4 α homodimer (Figure 9), as an example (90). Additionally, Dr. Yang's group has developed data analysis algorithms for small-angle X-ray scattering and protein footprinting techniques. They showed

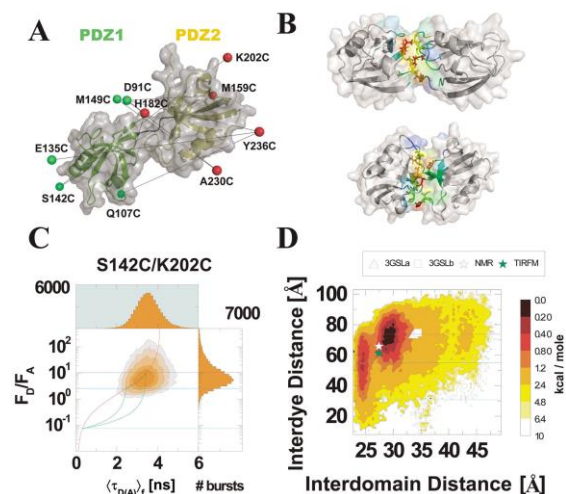


Figure 6. Reproduced with permission from Yanez-Orozco (2018)(82). FRET combined with DMD simulations identified two states in the N-terminal PDZ tandem of PSD95. The two rapidly-exchanging, compact states are stabilized by weak interactions. Biochemical methods and additional FRET experiments verified both the contact interfaces and the stabilizing mechanisms for these states. The combination of these approaches allowed direct validation of results which goes beyond what the individual techniques in previous studies or this study could provide alone.

that simulations, in conjunction with X-ray scattering, are well-suited to the study of structural dynamics and biomolecular complexes and in fostering collaborations across subfields. Such collaborations will be crucial in nearly every facet of integrating data generated by the diverse set of available structural biology tools.

Multiple other speakers highlighted their integration of computational methods into their work, including Dr. Bathe's hybrid computational-experimental approach for the design of nucleic acid assemblies and Dr. Keedy's use of multi-conformer structural modeling.

4.6. Data archiving

Data archives, especially online databases like the PDB (40), SCOP (56), and UniProt (62), are essential to structural biological research. They serve as a primary resource for depositing and accessing a wealth of biological data from atomistic structural models of proteins to functional networks. Further, this data provides initial models for computational studies, informs the design and interpretation of experiments and their results, and generally make

the results of previous studies accessible to the scientific community.

Dr. Catherine Lawson (Rutgers University) discussed practices and policies for X-ray crystallography, NMR and 3D-EM data deposition to the PDB as well as the Electron Microscopy Data Bank (EMDB) (91). Further, standards for curation were discussed. The first wwPDB Hybrid/Integrative Methods Task Force Workshop concluded that there is a need for the incorporation of diverse experimental data for structural modeling and that databases should accommodate these data (39). Following the task force recommendations, PDB-Dev was created to enable extension of the scope of the PDB archive to include additional experimental data used in integrative approaches, including FRET, EPR, small angle X-ray scattering (SAXS), and others (59).

5. ROUNDTABLE DISCUSSIONS

At the core of the FISB, 2017 workshop was a series of roundtable discussions concerning the future of integrative structural biology. The purpose of these discussions was to generate ideas of how a fully functional description of biology that spans length scales from atomic to cellular and timescales from thermal fluctuations to cell divisions might be represented. Further, we considered the purpose of such a representation and what its benefits would be. Questions raised during the discussions included what the nature of the representation should be, how data from different methods will be integrated and cross-validated, what types of representations should be prioritized, and who the key stakeholders will be.

The FISB workshop's roundtable discussions were structured as follows: participants formed small groups for two sessions focused on how and what information might be represented in a full, scale-spanning model of structural biology. Following each session, the groups presented summaries of their discussions to the larger audience for further discussion. The outcomes of these discussions are summarized below.

5.1. Data integration and representations

A molecular atlas of the cell with atomic resolution would contain an unreasonable amount of

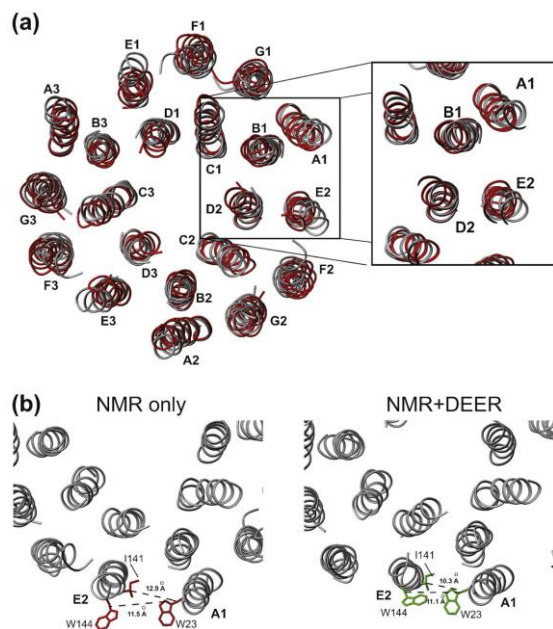


Figure 7. Reprinted from Journal of Molecular Biology, 429 (12), S. Milikisiyants et al, Oligomeric Structure of Anabaena Sensory Rhodopsin in a Lipid Bilayer Environment by Combining Solid-State NMR and Long-range DEER Constraints, Copyright 2017, with permission from Elsevier (83). Refinement of the *Anabaena* sensory rhodopsin (ASR) structure through inclusion of DEER distance constraints in structure calculation. In A, NMR structures are shown in grey, while structures obtained via NMR with DEER constraints are in red. In this study, utilization of DEER distances to constrain NMR modeling allowed significant refinement of the structure of a multi-protein complex as compared with NMR alone.

data for most users to handle. Nonetheless, a set of structural models that bridge the atomic and molecular complex scales would be useful in understanding the molecular machinery and the interactions that make it work. Further, one of the promises of integrative structural biology is its ability to bridge length scales from individual molecules to continuum models of sub-cellular macromolecular complexes. Thus, determining to what degree continuum and molecular-scale models, as well as the quality of those models, should be represented in a joint representation is vital in approaching such a goal.

On this topic, the discussion groups found consensus regarding several important points. First, data from multiple sources should be incorporated into the final representation, including electrical, chemical, and mechanical properties of structures and their environments. Beyond these physical

parameters, evolutionary traits, cellular localization, and functional network integration, e.g., interactome, data should also be included. Second, effective communication of these parameters will require either standardization or increased understanding of terms and representations across subfields. Third, the final representation should include the ability to adjust the parameters of one component of a structure, e.g., the pH or ionic strength of the environment, and have the effects of that change propagate throughout the structure. This adjustability would enable one to observe or predict the response of several parameters to changes in one.

Thoughts on the exact nature of representations were varied, with multiple suggestions for specific implementations. One suggested implementation was to use an “intuitive” 3D map, analogous to projects like Google Earth (<https://www.google.com/earth/>) or SpaceEngine (<http://spaceengine.org/>). Such a representation would rely on a coarse-graining approach, one in which “zooming out” would enable the user to see a simplified model of the larger system. For instance, when zooming out from within an organelle, one would see the organelle represented as a whole as opposed to, say, at atomic resolution. One challenge associated with this approach is how to balance the clarity of representing molecular details with the accuracy of representing protein population densities and crowding within a cell. This situation could be resolved either with visual accuracy, where finding specific data would become difficult in protein-dense regions or by implementing parameters that indicate local numerical densities of specific molecules. Additionally, issues around representing data for multiple conformational states of proteins, for hard-to-visualize parameters, and for unobserved but theorized conformations were raised during the discussion. The suggestions for addressing these issues centered on a randomized or averaged default representations, even though these may be unphysical, and providing additional representation options via submenus. Alternatively, it was suggested that using interpolation algorithms to generate intermediates in continuous parameters might be useful. Other

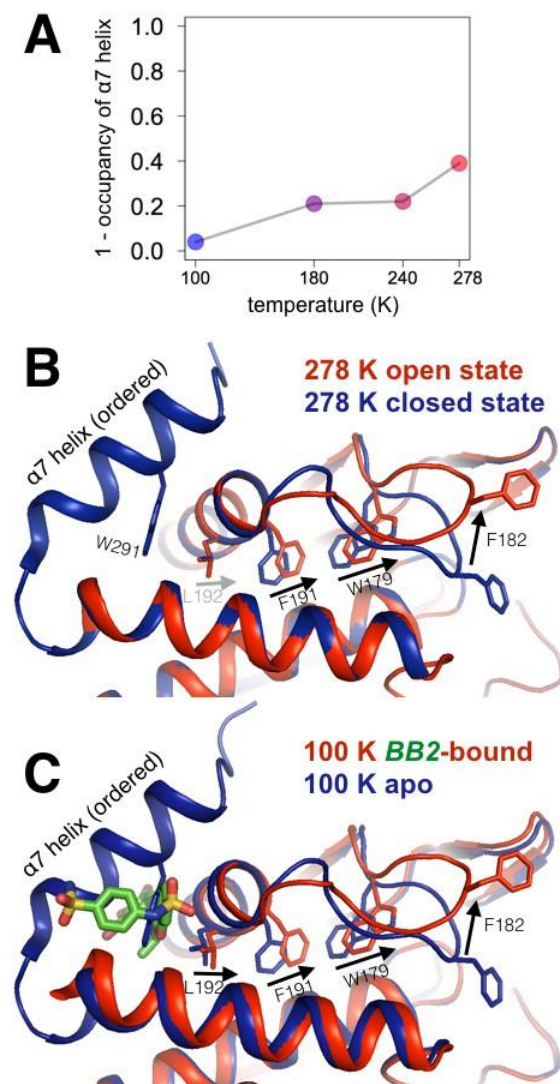


Figure 8. Reproduced with permission from, (100), Copyright 2018 Keedy et al. Multitemperature crystallography of apo protein tyrosine phosphatase 1B (PTP1B) recapitulates an allosteric mechanism. The $\alpha 7$ helix of PTP1B was found to become more disordered with increases in temperature. Further, residues that allosterically link $\alpha 7$ and the active-site WPD loop undergo shifts with increased temperature toward the state of PTP1B when trapped by an allosteric inhibitor. Identification of low-occupancy states of the protein with new modeling techniques helps shed light on this and other allosteric mechanisms in PTP1B.

suggested implementations were less graphical, consisting of multidimensional numerical representations, density maps and contour plots, and traditional database structures of branching topics. The broader group discussion found most consensus with some variation of a graphical 3D map, be it voxel-based, continuous, or similar.

Regarding validation, the discussion turned toward what kinds of validation are needed and how best to execute and display these. All groups agreed that experimental verification should be the final say in validation, using several different techniques wherever possible. However, there were multiple ideas about how specific representations of uncertainty should be handled. Suggestions included developing a unified, cross-subdisciplinary metric of uncertainty in conjunction with maintaining separate, subdisciplinary-specific metrics to avoid prescribing overly restrictive metrics. Finding ways to represent each of these separate metrics will present new challenges, but this approach could allow flexibility in representation by adopting the various standards for differing fields. For example, the R-factor, which is used to represent the extent to which the crystallographic model agrees with the experimental crystallographic X-ray diffraction data, and a standard deviation, which is used to represent variations in FRET-determined distance measurements, could both be inserted into the final representation. The potentially cluttering nature of this approach lends itself to a less-visual representation, though, perhaps such as text-based “tags.” Further, this would avoid potential difficulties in determining a consistent visual scheme because specific metrics may be ill-suited to a single visual representation, such as relative populations of different protein conformations or structures from different methods that disagree. The approach of a unified metric would allow for more visual representations, where, given some unified metric, color coding or different voxel shapes could be applied to represent uncertainties, including differences in structural uncertainties between separate domains of individual proteins, for example. However, this approach does not account for all user’s potential needs due to the various constraints of different techniques.

5.2. Time evolution and biological networks

Another promising aspect of integrative structural biology is its ability to bridge static and dynamic models and data. However, representing both equilibrium and non-equilibrium conditions presents additional challenges. Determining which aspects of equilibrium and non-equilibrium effects and how best to represent each will be necessary for

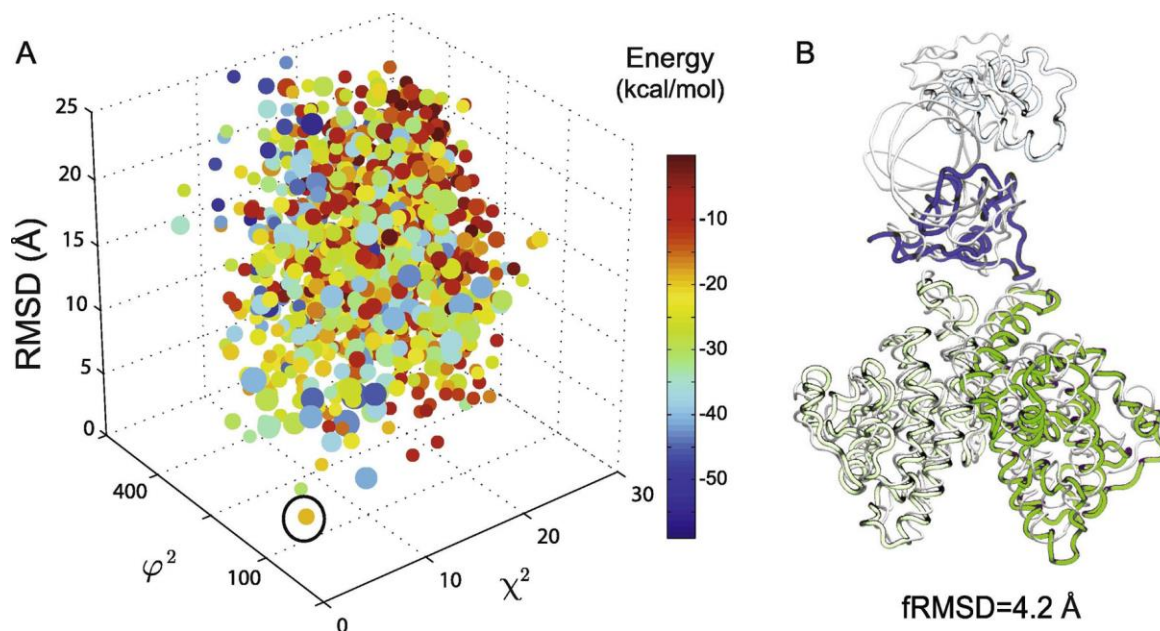


Figure 9. Reprinted from Journal of Structural Biology, 196 (3), W. Huang et al., Theoretical modeling of multiprotein complexes by iSPOT: Integration of small-angle X-ray scattering, hydroxyl radical footprinting, and computational docking, Copyright 2016, with permission from Elsevier (90). iSPOT accurately predicts the target crystal structure of the HNF-4α homodimer. Integration of SAXS, footprinting, and computational docking simulations with iSPOT can reduce ambiguity in structure determination compared to any of these techniques used individually, including for large protein complexes.

constructing such a representation.

The discussion session suggested that the first information that would need to be portrayed are the structural factors that differentiate equilibrium and non-equilibrium. Therefore, establishing a standard set of physiologically relevant non-equilibrium conditions would be important, as encompassing all possible non-equilibrium conditions would be difficult or impossible. These non-equilibrium conditions will need to be determined by the cell's efforts to maintain homeostasis. For instance, they would include steady-state, but non-equilibrium, localized concentrations of chemical fuels (e.g., ATP, GTP, and simple sugars), ions (e.g., Mg^{2+} , Ca^{2+} , and K^+) and pH, to name a few. Also, the role of user-provided input would need to be decided because implementing user-controlled deviations from equilibrium for simulation would be informative and useful, but computationally expensive. Furthermore, in addition to the parameters mentioned in the first roundtable discussion, the time domain becomes important in discussing non-equilibrium kinetics and dynamics. This data could come from time-resolved

experimental and computational sources, including single-molecule FCS, NMR, and molecular dynamics (MD) simulations. Extrapolation of data from experimental techniques that do not directly probe the time domain, such as molecular conformational states identified with crystallographic and cryo-EM methods, will also become important in representing dynamics. However, this presents the additional challenge of properly representing the intermediate states. Finally, information on various equilibration and kinetic pathways, particularly in the context of biological function, should be given because macromolecular conformations do not follow the same paths through state space in every condition, e.g., heat-shock proteins and other chaperones alter the folding energy landscapes of some proteins. Getting some information on distributions of states and changes of state, especially in non-equilibrium conditions, will require tested standards and statistical techniques for extrapolation of data.

Regarding the representation of equilibrium and non-equilibrium states, the overall sentiment was that visual representations are the

best path to take. For equilibrium states, these would be equivalent to the default, static representation. However, numerous implementations were suggested for non-equilibrium dynamics. One suggestion included integrating animations of small subsystems using simplified representations of individual protein conformational changes and interactions between small systems of functionally-related proteins. These visualizations should be able to run independently of others and reset to a fixed point. Alternatively, ensemble representations could be used for further simplification of larger systems, particularly because the difficulty of animating detailed systems scales with their size. Another suggestion was the implementation of sub-menus, where clicking on a specific protein would bring up information regarding its functional network, dynamic timescales, and possibly representative animations of this data. Such an implementation would be more conducive to piecewise addition of new information and animations. Visually, this would separate functional and dynamic representations from the static, but still have them located at the same source. Other, more ambitious suggestions included the implementation of real-time simulations and even virtual reality, though this would likely only become feasible with dramatically improved computational power, simulation force fields, and dynamics algorithms. Again, interpolation between static data points for which there are no known intermediates was mentioned several times, as this becomes especially important in the time domain, such as in displaying transitions between two protein conformations found through X-ray scattering. Without such interpolation, the animations suggested would not be feasible or would display unphysical characteristics.

As for the validation of dynamic models, the consensus was again that integrating various experimental methods would be critical. In the case of changing local environments, however, simulations may be relied on to fill in the gaps between direct measurements. Representation of error would likely take the same form as in the static case, but perhaps with the displayed error changing along with the object it pertains to in real-time. These changes could be visualized through

synchronized bar charts or similar displays.

5.3. The future ahead

Finally, the implementation of a unified representation of structural biology spanning the atomic and cellular scales will likely require coordination between several parties. Identifying these key stakeholders will be important in the realization of such a project, as it would require the commitment of considerable time and resources. Further, deliberate steps will need to be taken by these stakeholders if any progress is to be made.

Various suggestions came from the discussion of key stakeholders and steps forward. One that was emphasized as potentially fruitful was an increased emphasis on science in politics because the scope of a project to build a dynamic atlas of the cell is large enough to suggest a governmental approach, similar to the Human Genome Project (92), the Cancer Genome Atlas (93), and the Molecular Libraries and Imaging project

(<https://commonfund.nih.gov/molecularlibraries/index>), all funded through NIH. Such an approach would be aided not only by impressing upon government officials and the general public the transformative potential of integrative structural biology but also by advocating for increased involvement of scientists in government. A goal of increasing the focus on science at all levels of government would be to open the door to the establishment of additional funding and training opportunities, new forums for interdisciplinary communication, and task-forces for accomplishing a project as large as this one. Furthermore, specific improvements within the field of structural biology were suggested, such as the establishment of interdisciplinary training programs to further unite the field, review of old literature for combination with more current studies, and improvements to individual and integrative methods to better measure dynamics and structure simultaneously. However, these tasks will require diverse, extensive resources and, therefore, the involvement of many groups.

Who the key stakeholders will be comes

The future of integrative structural biology

down to who stands to gain from a unified representation of structural biology. Due to the inherently interdisciplinary nature of structural biology, including biology, physics, chemistry, computer science, and medicine, the group consensus was that essentially everybody is a stakeholder in this project. Notably, key groups to have involved include current and future scientists, universities and academia, technology companies with broader investment interests such as Google, Microsoft, IBM, and Facebook, pharmaceutical and biotech companies, and governmental and research institutions such as NIH, NSF, and the national labs. Having these groups on board will be essential to obtaining and establishing the resources necessary for such an enormous undertaking. Involvement of current databases such as the PDB (40), PDB-Dev (39), SCOP (56), and UniProt (62) would also be instrumental as sources of current and past data.

One important note in this discussion was that the barrier for individuals to contribute to a project of this scale should be set as low as possible to make it attractive to a diverse group of potential stakeholders. Thus, it is likely that an approach including interdisciplinary training programs and a dictionary of terms for integrative structural biology would be more fruitful than an approach prescribing a single standard set of terms and modeling framework for submission, though this may make the curation of submitted data and models more difficult.

6. A PATH FORWARD: THE “VIRTUAL CELL”

The overall conclusion from the FISB discussion sessions was that the best path forward for a unified, complete atlas of a cell is to strive toward the “Virtual Cell” concept. Functionally, this atlas would contain both the spatial and temporal domains of entire cells from various species and cell types, down to the atomic scale of individual macromolecules. Furthermore, it would contain information on environmental conditions, functional roles and networks, and information on model validation such as error. It would contain links to the relevant publications for each molecule or system. Such an atlas would be useful both to the scientific community and to the broader public.

A Virtual Cell will:

1. Serve as a useful database for functional and structural models of all macromolecular components of the cell. As such, it will become an authoritative method for disseminating experimental data and results within and between disciplines, including data that might otherwise be unpublished or reserved for supplemental material.
2. Serve as one site for integrating these data into comprehensive models and visual representations.
3. Facilitate identifying the gaps in knowledge through both literal visualization of these gaps in the representation and the grouping of relevant publications.
4. Create an interface that engages and educates the public using structural biology, exciting future generations of scientists while capturing the imaginations of the general public to help sway public opinion in favor of science.
5. Provide an attractive prospect with which to attract funding into structural biology research from governmental, institutional, and private sources, including tech companies that do not typically invest in basic biological science but may find the hybrid technological and scientific approach alluring.

7. GIVE TANGIBLE FORM TO THE ULTIMATE GOAL OF STRUCTURAL BIOLOGY

The consensus of the group was to take several steps to effectively launch a project to build a single representation that spans scale from atomic to cellular while incorporating dynamics. First, we propose to establish a task force charged with setting specific goals and standards, planning and hosting training workshops, serving as intermediaries for funding opportunities, and generally organizing the effort. A core, dedicated task force is likely essential to ultimate success, as other projects have shown, such as the Human Genome Project’s genetic testing

task force (92) and the PDB's validation and Hybrid/Integrative Methods Task Force (39). The task force's work would include coordinating with existing groups that are well-aligned with the goal of building an atlas of the cell that bridges scale, like the wwPDB Hybrid/Integrative Methods Task Force (39), which is focused on molecular-scale integrative structural biology, the Human Cell Atlas (94), which aims to create comprehensive maps for all cells in the human body, and the Cell Image Library (<http://www.cellimagelibrary.org/home>), which focuses on high resolution images, videos, and animations of cell-level processes. Close coordination is particularly important because the Virtual Cell will need to fill the gaps in scale between the systems treated by these projects. A clear example of similar initiatives is the Pancreatic Beta Cell Consortium of the Bridge Institute (<https://dornsife.usc.edu/bridge-institute/pancreatic-beta-cell-consortium/>).

Second, the task force will have to decide on a set of test cases. We suggest that these test cases represent information that is both important, to demonstrate potential impact, and easiest to incorporate into the Virtual Cell, to demonstrate feasibility and establish a reasonable workflow. We suggest that this could be accomplished by starting at one "end" of the scale spectrum (either the cell-level or molecular level) and working toward the other. The most realistic starting point would be to incorporate representative, static structures to be followed by the introduction of information on dynamics, function, to begin with. This approach will leverage the tools and databases that currently exist, thanks to traditional techniques. Much of the structural data from these techniques is already archived and represented visually in databases, e.g., PDB, that can serve as a "springboard" from which a Virtual Cell could be launched.

Third, the task force will have to determine an appropriate set of cells for which cellular atlases could be constructed most completely and accurately. These candidates must be ubiquitous, well-studied, and of high potential impact so that they can serve as models in the establishment of the methods, procedures, and protocols that will be used in designing the atlas. Much of the data for such an

undertaking already exists, but it will take considerable effort to synthesize it and obtain the necessary missing information, even for the simplest of model systems.

Finally, it will be necessary that the bulk of the actual legwork is done by a larger group than just the core task force. The effort will require a significant contingent of structural biologists interested in integrative methods. Data from multiple sources must be synthesized and complementary information spanning scales in both space and time will have to be reconciled into cohesive models while simultaneously providing a good understanding of the uncertainties in the conclusions that can be drawn. Further, this must be done in a robust and repeatable manner. Efforts that have already been performed to develop specific methods for integrating structural information across scales will be leveraged. For instance, a method for structural modeling from the atomic to cellular scales through the integration of diverse datasets, including X-ray, EM, proteomics, and label-based methods and fluorescence techniques represents a good starting point (95). However, there is much work to be done in undertaking this project.

In addition to the steps discussed here, a critical factor in the achievability of the Virtual Cell will be interest in the project that extends beyond the researchers directly working in the emerging field of integrative structural biology. Such an undertaking will require a significant amount of time and money, and the commitment of these resources will have to come from various sources. For instance, the Human Genome Project shows us that interest in comprehensive and understandable archiving and representation of human genetic information can be stoked in the government and the public. However, significant efforts will need to be made to impress upon officials the importance of dedicating resources to these projects. Avenues like NSF's Big Ideas (https://www.nsf.gov/news/special_reports/big_ideas/) initiative provide realistic means of kickstarting ambitious projects through direct funding. Moreover, this project will likely span generations of scientists, and stoking interest in the public will be necessary to provide a feedback loop that can inspire the next generation of scientists. Large companies with

interests in research, such as Google or Amazon, may also have a role to play, as they can provide resources beyond just funding, including personnel, technology, and public relations.

The Virtual Cell is the ultimate manifestation of the field of structural biology, and it will only be achievable with a massive, coordinated, integrative effort. However, as the field of integrative structural biology is emerging, the path to this ultimate goal has just become feasible. When complete, the Virtual Cell has the potential to revolutionize the biomedical sciences, just as other massive projects have done. It will provide new insights into the big picture of cellular function as constituted by subcellular machinery. The intuitive structure of the Virtual Cell, combined with the robust conclusions and data of integrative methods, will make clear exactly what is known about a given subcellular system. This, in turn, helps pinpoint gaps in knowledge, paving the way for the development and application of integrative methods that can probe those gaps. Further, the usefulness of such a representation extends beyond academic research, as the intuition provided may prove invaluable for rational drug design, biotech development, and even education in the biological sciences. The proposed project is certainly ambitious, but the successes of other projects, including the Human Genome Project, wwPDB, and the Cell Image Library, are a testament to both its feasibility and potential impact.

8. ACKNOWLEDGMENTS

Funding sources: ORAU, Clemson University (Department of Physics and Astronomy, Department of Genetics and Biochemistry, College of Science, EPIC), NSF (CAREER MCB- 1749778) and US National Institutes of Health (NIH) grants R01MH08192311 and P20GM121342. We thank our invited speakers for their contributed talks, facilitators, participants, and organizers of the workshop. The authors declare no competing interests.

9. REFERENCES

1. P. E. Wright and H. J. Dyson: Intrinsically unstructured proteins: re-assessing the

protein structure-function paradigm. *J Mol Biol*, 293(2), 321-31 (1999)
DOI: 10.1006/jmbi.1999.3110

2. A. Sali, R. Glaeser, T. Earnest and W. Baumeister: From words to literature in structural proteomics. *Nature*, 422(6928), 216 (2003)
3. O. C. Redfern, B. Dessailly and C. A. Orengo: Exploring the structure and function paradigm. *Curr Opin Struct Biol*, 18(3), 394-402 (2008)
DOI: 10.1016/j.sbi.2008.05.007
4. K. Kuwata: An Emerging Concept of Biomolecular Dynamics and Function: Applications of NMR & MRI. *Magnetic Resonance in Medical Sciences*, 1(1), 27-31 (2002)
DOI: 10.2463/mrms.1.27
5. M. Karplus and J. Kuriyan: Molecular dynamics and protein function. *Proceedings of the National Academy of Sciences of the United States of America*, 102(19), 6679 (2005)
DOI: 10.1073/pnas.0408930102
6. A. B. Ward, A. Sali and I. A. Wilson: Biochemistry. Integrative structural biology. *Science*, 339(6122), 913-5 (2013)
DOI: 10.1126/science.1228565
7. A. Doerr: Cryo-electron tomography. *Nature Methods*, 14, 34 (2016)
DOI: 10.1038/nmeth.4115
8. C. M. Oikonomou and G. J. Jensen: Cellular Electron Cryotomography: Toward Structural Biology *In situ*. *Annu Rev Biochem*, 86, 873-896 (2017)
DOI: 10.1146/annurev-biochem-061516-044741
9. E. H. Egelman: Single-particle

- reconstruction from EM images of helical filaments. *Curr Opin Struct Biol*, 17(5), 556-61 (2007)
DOI: 10.1016/j.sbi.2007.07.006
10. I. Patla, T. Volberg, N. Elad, V. Hirschfeld-Warneken, C. Grashoff, R. Fässler, J. P. Spatz, B. Geiger and O. Medalia: Dissecting the molecular architecture of integrin adhesion sites by cryo-electron tomography. *Nature Cell Biology*, 12, 909 (2010)
DOI: 10.1038/ncb2095
<https://www.nature.com/articles/ncb2095#supplementary-information>
11. P. Schellenberger, R. Kaufmann, C. A. Siebert, C. Hagen, H. Wodrich and K. Grünwald: High-precision correlative fluorescence and electron cryo microscopy using two independent alignment markers. *Ultramicroscopy*, 143(100), 41-51 (2014)
DOI: 10.1016/j.ultramic.2013.10.011
12. S. Kalinin, R. Kuhnemuth, H. Vardanyan and C. A. Seidel: Note: a 4 ns hardware photon correlator based on a general-purpose field-programmable gate array development board implemented in a compact setup for fluorescence correlation spectroscopy. *Rev Sci Instrum*, 83(9), 096105 (2012)
DOI: 10.1063/1.4753994
13. S. Felekyan, S. Kalinin, H. Sanabria, A. Valeri and C. A. Seidel: Filtered FCS: species auto- and cross-correlation functions highlight binding and dynamics in biomolecules. *Chemphyschem*, 13(4), 1036-53 (2012)
DOI: 10.1002/cphc.201100897
14. V. Kudryavtsev, M. Sikor, S. Kalinin, D. Mokranjac, C. A. Seidel and D. C. Lamb: Combining MFD and PIE for accurate single-pair Forster resonance energy transfer measurements. *Chemphyschem*, 13(4), 1060-78 (2012)
DOI: 10.1002/cphc.201100822
15. T. O. Peulen, O. Opanasyuk and C. A. M. Seidel: Combining Graphical and Analytical Methods with Molecular Simulations To Analyze Time-Resolved FRET Measurements of Labeled Macromolecules Accurately. *J Phys Chem B*, 121(35), 8211-8241 (2017)
DOI: 10.1021/acs.jpcc.7b03441
16. M. Dimura, T. O. Peulen, C. A. Hanke, A. Prakash, H. Gohlke and C. A. M. Seidel: Quantitative FRET studies and integrative modeling unravel the structure and dynamics of biomolecular systems. *Current Opinion in Structural Biology*, 40, 163-185 (2016)
17. S. W. Hell and J. Wichmann: Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy. *Opt Lett*, 19(11), 780-2 (1994)
18. W. Wang, G. Zhao, C. Kuang, L. Xu, S. Liu, S. Sun, P. Shentu, Y. M. Yang, Y. Xu and X. Liu: Integrated dual-color stimulated emission depletion (STED) microscopy and fluorescence emission difference (FED) microscopy. *Optics Communications*, 423, 167-174 (2018)
19. C. Kuang, S. Li, W. Liu, X. Hao, Z. Gu, Y. Wang, J. Ge, H. Li and X. Liu: Breaking the diffraction barrier using fluorescence emission difference microscopy. *Scientific reports*, 3, 1441 (2013)
20. G. Vicidomini, P. Bianchini and A. Diaspro: STED super-resolved microscopy. *Nat Methods*, 15(3), 173-

- 182 (2018)
DOI: 10.1038/nmeth.4593
21. C. Eggeling, C. Ringemann, R. Medda, G. Schwarzmann, K. Sandhoff, S. Polyakova, V. N. Belov, B. Hein, C. von Middendorff and A. Schönle: Direct observation of the nanoscale dynamics of membrane lipids in a living cell. *Nature*, 457(7233), 1159 (2009)
22. M. T. Lerch, Z. Yang, C. Altenbach and W. L. Hubbell: High-Pressure EPR and Site-Directed Spin Labeling for Mapping Molecular Flexibility in Proteins. *Methods Enzymol*, 564, 29-57 (2015)
DOI: 10.1016/bs.mie.2015.07.004
23. C. Altenbach, S. L. Flitsch, H. G. Khorana and W. L. Hubbell: Structural Studies on Transmembrane Proteins .2. Spin Labeling of Bacteriorhodopsin Mutants at Unique Cysteines. *Biochemistry*, 28(19), 7806-7812 (1989)
DOI: 10.1021/bi00445a042
24. S. Milikisiyants, M. A. Voinov, A. Marek, M. Jafarabadi, J. Liu, R. Han, S. Wang and A. I. Smirnov: Enhancing sensitivity of Double Electron-Electron Resonance (DEER) by using Relaxation-Optimized Acquisition Length Distribution (RELOAD) scheme. *J Magn Reson*, 298, 115-126 (2019)
DOI: 10.1016/j.jmr.2018.12.004
25. C. Mura, E. J. Draizen and P. E. Bourne: Structural biology meets data science: does anything change? *Current opinion in structural biology*, 52, 95-102 (2018)
26. J. Lengyel, E. Hnath, M. Storms and T. Wohlfarth: Towards an integrative structural biology approach: combining Cryo-TEM, X-ray crystallography, and NMR. *J Struct Funct Genomics*, 15(3), 117-24 (2014)
DOI: 10.1007/s10969-014-9179-9
27. M. D. Purdy, B. C. Bennett, W. E. McIntire, A. K. Khan, P. M. Kasson and M. Yeager: Function and dynamics of macromolecular complexes explored by integrative structural and computational biology. *Curr Opin Struct Biol*, 27, 138-48 (2014)
DOI: 10.1016/j.sbi.2014.08.006
28. M. Rey, V. Sarpe, K. M. Burns, J. Buse, C. A. Baker, M. van Dijk, L. Wordeman, A. M. Bonvin and D. C. Schriemer: Mass spec studio for integrative structural biology. *Structure*, 22(10), 1538-48 (2014)
DOI: 10.1016/j.str.2014.08.013
29. A. Politis and A. J. Borysik: Assembling the pieces of macromolecular complexes: Hybrid structural biology approaches. *Proteomics*, 15(16), 2792-803 (2015)
DOI: 10.1002/pmic.201400507
30. H. Van Den Bedem and J. S. Fraser: Integrative, dynamic structural biology at atomic resolution—it's about time. *Nature methods*, 12(4), 307 (2015)
31. M. Faini, F. Stengel and R. Aebersold: The Evolving Contribution of Mass Spectrometry to Integrative Structural Biology. *J Am Soc Mass Spectrom*, 27(6), 966-74 (2016)
DOI: 10.1007/s13361-016-1382-4
32. P. Romano and F. Cordero: NETTAB 2014: From high-throughput structural bioinformatics to integrative systems biology. In: BioMed Central, (2016)
33. F. Forneris and A. Mattevi: Expanding the structural biology toolbox with single-

- molecule holography. *Proc Natl Acad Sci U S A*, 114(7), 1448-1450 (2017)
DOI: 10.1073/pnas.1620897114
34. S. Olsson, H. Wu, F. Paul, C. Clementi and F. Noe: Combining experimental and simulation data of molecular processes via augmented Markov models. *Proc Natl Acad Sci U S A*, 114(31), 8265-8270 (2017)
DOI: 10.1073/pnas.1704803114
35. C. Morris: The Life Cycle of Structural Biology Data. *Data Science Journal*, 17, 26 (2018)
DOI: 10.5334/dsj-2018-026
36. R. M. Yennamalli: Structural Bioinformatics and Big Data Analytics: A mini-review. *International Journal for Computational Biology (IJCB)*; Vol 6, No 1 (2017) (2017)
37. G. J. Kleywegt, S. Velankar and A. Patwardhan: Structural biology data archiving - where we are and what lies ahead. *FEBS Lett*, 592(12), 2153-2167 (2018)
DOI: 10.1002/1873-3468.13086
38. A. Gutmanas, T. J. Oldfield, A. Patwardhan, S. Sen, S. Velankar and G. J. Kleywegt: The role of structural bioinformatics resources in the era of integrative structural biology. *Acta Crystallogr D Biol Crystallogr*, 69(Pt 5), 710-21 (2013)
DOI: 10.1107/S0907444913001157
39. A. Sali, H. M. Berman, T. Schwede, J. Trehwella, G. Kleywegt, S. K. Burley, J. Markley, H. Nakamura, P. Adams and A. M. J. J. Bonvin: Outcome of the first wwPDB hybrid/integrative methods task force workshop. *Structure*, 23(7), 1156-1167 (2015)
40. H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne: The protein data bank. *Nucleic acids research*, 28(1), 235-242 (2000)
41. H. Berman, K. Henrick and H. Nakamura: Announcing the worldwide Protein Data Bank. *Nat Struct Biol*, 10(12), 980 (2003)
DOI: 10.1038/nsb1203-980
42. H. Berman, K. Henrick, H. Nakamura and J. L. Markley: The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data. *Nucleic Acids Research*, 35, D301-D303 (2007)
DOI: 10.1093/nar/gkl971
43. M. Guharoy and P. Chakrabarti: Conserved residue clusters at protein-protein interfaces and their use in binding site identification. *BMC Bioinformatics*, 11(1), 286 (2010)
DOI: 10.1186/1471-2105-11-286
44. T. Tesileanu, L. J. Colwell and S. Leibler: Protein sectors: statistical coupling analysis versus conservation. *PLoS Comput Biol*, 11(2), e1004091 (2015)
DOI: 10.1371/journal.pcbi.1004091
45. H. C. Kornau, L. T. Schenker, M. B. Kennedy and P. H. Seeburg: Domain Interaction between Nmda Receptor Subunits and the Postsynaptic Density Protein Psd-95. *Science*, 269(5231), 1737-1740 (1995)
DOI: 10.1126/science.7569905
46. U. Kistner, B. M. Wenzel, R. W. Veh, C. Cases-Langhoff, A. M. Garner, U. Appeltauer, B. Voss, E. D. Gundelfinger and C. C. Garner: SAP90, a rat presynaptic protein related to the product of the Drosophila tumor suppressor gene dlg-A. *J Biol Chem*, 268(7), 4580-3 (1993)

47. M. D. Zimmerman, M. Grabowski, M. J. Domagalski, E. M. Maclean, M. Chruszcz and W. Minor: Data management in the modern structural biology and biomedical research environment. *Methods Mol Biol*, 1140, 1-25 (2014)
DOI: 10.1007/978-1-4939-0354-2_1
48. M. Baker: 1,500 scientists lift the lid on reproducibility. *Nature*, 533(7604), 452-4 (2016)
DOI: 10.1038/533452a
49. P. D. Schloss: Identifying and Overcoming Threats to Reproducibility, Replicability, Robustness, and Generalizability in Microbiome Research. *Mbio*, 9(3) (2018)
DOI: 10.1128/mBio.00525-18
50. A. Nekrutenko and J. Taylor: Next-generation sequencing data interpretation: enhancing reproducibility and accessibility. *Nature Reviews Genetics*, 13(9), 667-U93 (2012)
DOI: 10.1038/nrg3305
51. S. R. Horn, M. M. Long, B. W. Nelson, N. B. Allen, P. A. Fisher and M. L. Byrne: Replication and reproducibility issues in the relationship between C-reactive protein and depression: A systematic review and focused meta-analysis. *Brain Behavior and Immunity*, 73, 85-114 (2018)
DOI: 10.1016/j.bbi.2018.06.016
52. M. B. O'Rourke, S. P. Djordjevic and M. P. Padula: The quest for improved reproducibility in MALDI mass spectrometry. *Mass Spectrom Rev*, 37(2), 217-228 (2018)
DOI: 10.1002/mas.21515
53. A. Y. Lau and D. I. Chasman: Functional classification of proteins and protein variants. *Proc Natl Acad Sci U S A*, 101(17), 6576-81 (2004)
DOI: 10.1073/pnas.0305043101
54. M. Manoharan, S. A. Muhammad and R. Sowdhamini: Sequence Analysis and Evolutionary Studies of Reelin Proteins. *Bioinform Biol Insights*, 9, 187-93 (2015)
DOI: 10.4137/BBI.S26530
55. R. J. Najmanovich: Evolutionary studies of ligand binding sites in proteins. *Curr Opin Struct Biol*, 45, 85-90 (2017)
DOI: 10.1016/j.sbi.2016.11.024
56. A. G. Murzin, S. E. Brenner, T. Hubbard and C. Chothia: SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J Mol Biol*, 247(4), 536-40 (1995)
DOI: 10.1006/jmbi.1995.0159
57. S. Mandal, M. Moudgil and S. K. Mandal: Rational drug design. *Eur J Pharmacol*, 625(1-3), 90-100 (2009)
DOI: 10.1016/j.ejphar.2009.06.065
58. M. K. Hellerstein: *In vivo* measurement of fluxes through metabolic pathways: the missing link in functional genomics and pharmaceutical research. *Annu Rev Nutr*, 23, 379-402 (2003)
DOI: 10.1146/annurev.nutr.23.011702.073045
59. B. Vallat, B. Webb, J. D. Westbrook, A. Sali and H. M. Berman: Development of a prototype system for archiving integrative/hybrid structure models of biological macromolecules. *Structure*, 26(6), 894-904 (2018)
60. C. J. Lawrence, R. K. Dawe, K. R. Christie, D. W. Cleveland, S. C. Dawson, S. A. Endow, L. S. Goldstein, H. V.

- Goodson, N. Hirokawa, J. Howard, R. L. Malmberg, J. R. McIntosh, H. Miki, T. J. Mitchison, Y. Okada, A. S. Reddy, W. M. Saxton, M. Schliwa, J. M. Scholey, R. D. Vale, C. E. Walczak and L. Wordeman: A standardized kinesin nomenclature. *J Cell Biol*, 167(1), 19-22 (2004)
DOI: 10.1083/jcb.200408113
61. E. F. Hom, G. B. Witman, E. H. Harris, S. K. Dutcher, R. Kamiya, D. R. Mitchell, G. J. Pazour, M. E. Porter, W. S. Sale, M. Wirschell, T. Yagi and S. M. King: A unified taxonomy for ciliary dyneins. *Cytoskeleton (Hoboken)*, 68(10), 555-65 (2011)
DOI: 10.1002/cm.20533
62. T. UniProt Consortium: UniProt: the universal protein knowledgebase. *Nucleic Acids Res*, 46(5), 2699 (2018)
DOI: 10.1093/nar/gky092
63. J. Trewthella, W. A. Hendrickson, G. J. Kleywegt, A. Sali, M. Sato, T. Schwede, D. I. Svergun, J. A. Tainer, J. Westbrook and H. M. Berman: Report of the wwPDB Small-Angle Scattering Task Force: data requirements for biomolecular modeling and the PDB. *Structure*, 21(6), 875-81 (2013)
DOI: 10.1016/j.str.2013.04.020
64. E. Valentini, A. G. Kikhney, G. Previtali, C. M. Jeffries and D. I. Svergun: SASBDB, a repository for biological small-angle scattering data. *Nucleic Acids Research*, 43(D1), D357-D363 (2015)
DOI: 10.1093/nar/gku1047
65. M. Knoll and E. Ruska: The Electron Microscope. *Zeitschrift Fur Physik*, 78(5-6), 318-339 (1932)
DOI: 10.1007/Bf01342199
66. H. Wang: Cryo-electron microscopy for structural biology: current status and future perspectives. *Sci China Life Sci*, 58(8), 750-6 (2015)
DOI: 10.1007/s11427-015-4851-2
67. J. P. Renaud, A. Chari, C. Ciferri, W. T. Liu, H. W. Remigy, H. Stark and C. Wiesmann: Cryo-EM in drug discovery: achievements, limitations and prospects. *Nature Reviews Drug Discovery*, 17(7), 471-492 (2018)
DOI: 10.1038/nrd.2018.77
68. Y. F. Cheng: Single-Particle Cryo-EM at Crystallographic Resolution. *Cell*, 161(3), 450-457 (2015)
DOI: 10.1016/j.cell.2015.03.049
69. K. Murata and M. Wolf: Cryo-electron microscopy for structural analysis of dynamic biological macromolecules. *Biochimica Et Biophysica Acta-General Subjects*, 1862(2), 324-334 (2018)
DOI: 10.1016/j.bbagen.2017.07.020
70. A. Merk, A. Bartesaghi, S. Banerjee, V. Falconieri, P. Rao, M. I. Davis, R. Pragani, M. B. Boxer, L. A. Earl, J. L. S. Milne and S. Subramaniam: Breaking Cryo-EM Resolution Barriers to Facilitate Drug Discovery. *Cell*, 165(7), 1698-1707 (2016)
DOI: 10.1016/j.cell.2016.05.040
71. B. Huang, M. Bates and X. Zhuang: Super-resolution fluorescence microscopy. *Annu Rev Biochem*, 78, 993-1016 (2009)
DOI: 10.1146/annurev.biochem.77.061906.092014
72. S. J. Sahl, S. W. Hell and S. Jakobs: Fluorescence nanoscopy in cell biology. *Nat Rev Mol Cell Biol*, 18(11), 685-701 (2017)
DOI: 10.1038/nrm.2017.71

73. S. J. Sahl and W. E. Moerner: Super-resolution fluorescence imaging with single molecules. *Curr Opin Struct Biol*, 23(5), 778-87 (2013)
DOI: 10.1016/j.sbi.2013.07.010
74. M. J. Rust, M. Bates and X. Zhuang: Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat Methods*, 3(10), 793-5 (2006)
DOI: 10.1038/nmeth929
75. Y. Zhang, P. D. Nallathamby, G. D. Vigil, A. A. Khan, D. E. Mason, J. D. Boerckel, R. K. Roeder and S. S. Howard: Super-resolution fluorescence microscopy by stepwise optical saturation. *Biomed Opt Express*, 9(4), 1613-1629 (2018)
DOI: 10.1364/BOE.9.001613
76. A. Sharonov and R. M. Hochstrasser: Wide-field subdiffraction imaging by accumulated binding of diffusing probes. *Proc Natl Acad Sci U S A*, 103(50), 18911-6 (2006)
DOI: 10.1073/pnas.0609643104
77. P. Hoyer, G. de Medeiros, B. Balazs, N. Norlin, C. Besir, J. Hanne, H. G. Krausslich, J. Engelhardt, S. J. Sahl, S. W. Hell and L. Hufnagel: Breaking the diffraction limit of light-sheet fluorescence microscopy by RESOLFT. *Proc Natl Acad Sci U S A*, 113(13), 3442-6 (2016)
DOI: 10.1073/pnas.1522292113
78. F. Balzarotti, Y. Eilers, K. C. Gwosch, A. H. Gynna, V. Westphal, F. D. Stefani, J. Elf and S. W. Hell: Nanometer resolution imaging and tracking of fluorescent molecules with minimal photon fluxes. *Science*, 355(6325), 606-612 (2017)
DOI: 10.1126/science.aak9913
79. K. Pan, D. N. Kim, F. Zhang, M. R. Adendorff, H. Yan and M. Bathe: Lattice-free prediction of three-dimensional structure of programmed DNA assemblies. *Nat Commun*, 5, 5578 (2014)
DOI: 10.1038/ncomms6578
80. S.-M. Guo, R. Veneziano, S. Gordonov, L. Li, D. Park, A. B. Kulesa, P. C. Blainey, J. R. Cottrell, E. S. Boyden and M. Bathe: Multiplexed confocal and super-resolution fluorescence imaging of cytoskeletal and neuronal synapse proteins. *bioRxiv*, 111625 (2017)
81. B. Hellenkamp, S. Schmid, O. Doroshenko, O. Opanasyuk, R. Kuhnemuth, S. R. Adariani, B. Ambrose, M. Aznauryan, A. Barth, V. Birkedal, M. E. Bowen, H. T. Chen, T. Cordes, T. Eilert, C. Fijen, C. Gebhardt, M. Gotz, G. Gouridis, E. Gratton, T. Ha, P. Y. Hao, C. A. Hanke, A. Hartmann, J. Hendrix, L. L. Hildebrandt, V. Hirschfeld, J. Hohlbein, B. Y. Hua, C. G. Hubner, E. Kallis, A. N. Kapanidis, J. Y. Kim, G. Krainer, D. C. Lamb, N. K. Lee, E. A. Lemke, B. Levesque, M. Levitus, J. J. McCann, N. Naredi-Rainer, D. Nettels, T. Ngo, R. Y. Qiu, N. C. Robb, C. Rocker, H. Sanabria, M. Schlierf, T. Schroder, B. Schuler, H. Seidel, L. Streit, J. Thurn, P. Tinnefeld, S. Tyagi, N. Vandenberg, A. M. Vera, K. R. Weninger, B. Wunsch, I. S. Yanez-Orozco, J. Michaelis, C. A. M. Seidel, T. D. Craggs and T. Hugel: Precision and accuracy of single-molecule FRET measurements-a multi-laboratory benchmark study (vol 15, pg 984, 2018). *Nature Methods*, 15(11), 984-984 (2018)
DOI: 10.1038/s41592-018-0193-x
82. I. S. Yanez Orozco, F. A. Mindlin, J. Ma, B. Wang, B. Levesque, M. Spencer, S.

- Rezaei Adariani, G. Hamilton, F. Ding, M. E. Bowen and H. Sanabria: Identifying weak interdomain interactions that stabilize the supertertiary structure of the N-terminal tandem PDZ domains of PSD-95. *Nat Commun*, 9(1), 3724 (2018)
DOI: 10.1038/s41467-018-06133-0
83. S. Milikisiyants, S. L. Wang, R. A. Munro, M. Donohue, M. E. Ward, D. Bolton, L. S. Brown, T. I. Smirnova, V. Ladizhansky and A. I. Smirnov: Oligomeric Structure of Anabaena Sensory Rhodopsin in a Lipid Bilayer Environment by Combining Solid-State NMR and Long-range DEER Constraints. *Journal of Molecular Biology*, 429(12), 1903-1920 (2017)
DOI: 10.1016/j.jmb.2017.05.005
84. D. Russel, K. Lasker, B. Webb, J. Velazquez-Muriel, E. Tjioe, D. Schneidman-Duhovny, B. Peterson and A. Sali: Putting the pieces together: integrative modeling platform software for structure determination of macromolecular assemblies. *PLoS Biol*, 10(1), e1001244 (2012)
DOI: 10.1371/journal.pbio.1001244
85. A. Hospital, J. R. Goni, M. Orozco and J. L. Gelpi: Molecular dynamics simulations: advances and applications. *Adv Appl Bioinform Chem*, 8, 37-47 (2015)
DOI: 10.2147/AABC.S70333
86. S. F. Altschul, W. Gish, W. Miller, E. W. Myers and D. J. Lipman: Basic Local Alignment Search Tool. *Journal of Molecular Biology*, 215(3), 403-410 (1990)
DOI: 10.1016/S0022-2836(05)80360-2
87. B. Webb and A. Sali: Comparative protein structure modeling using MODELLER. *Current protocols in bioinformatics*, 47(1), 5-6 (2014)
88. R. O. Dror, R. M. Dirks, J. P. Grossman, H. Xu and D. E. Shaw: Biomolecular simulation: a computational microscope for molecular biology. *Annu Rev Biophys*, 41, 429-52 (2012)
DOI: 10.1146/annurev-biophys-042910-155245
89. K. Cowtan: Phase Problem in X-ray Crystallography, and Its Solution. *e LS* (2001)
90. W. Huang, K. M. Ravikumar, M. Parisien and S. C. Yang: Theoretical modeling of multiprotein complexes by iSPOT: Integration of small-angle X-ray scattering, hydroxyl radical footprinting, and computational docking. *Journal of Structural Biology*, 196(3), 340-349 (2016)
DOI: 10.1016/j.jsb.2016.08.001
91. A. Patwardhan and C. L. Lawson: Databases and Archiving for CryoEM. *Methods Enzymol*, 579, 393-412 (2016)
DOI: 10.1016/bs.mie.2016.04.015
92. E. S. Lander, I. H. G. S. Consortium, L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczy, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J. P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, N. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M.

Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J. C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R. H. Waterston, R. K. Wilson, L. W. Hillier, J. D. McPherson, M. A. Marra, E. R. Mardis, L. A. Fulton, A. T. Chinwalla, K. H. Pepin, W. R. Gish, S. L. Chissoe, M. C. Wendl, K. D. Delehaunty, T. L. Miner, A. Delehaunty, J. B. Kramer, L. L. Cook, R. S. Fulton, D. L. Johnson, P. J. Minx, S. W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J. F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, R. A. Gibbs, D. M. Muzny, S. E. Scherer, J. B. Bouck, E. J. Sodergren, K. C. Worley, C. M. Rives, J. H. Gorrell, M. L. Metzker, S. L. Naylor, R. S. Kucherlapati, D. L. Nelson, G. M. Weinstock, Y. Sakaki, A. Fujiyama, M. Hattori, T. Yada, A. Toyoda, T. Itoh, C. Kawagoe, H. Watanabe, Y. Totoki, T. Taylor, J. Weissenbach, R. Heilig, W. Saurin, F. Artiguenave, P. Brottier, T. Bruls, E. Pelletier, C. Robert, P. Wincker, A. Rosenthal, M. Platzer, G. Nyakatura, S. Taudien, A. Rump, H. M. Yang, J. Yu, J. Wang, G. Y. Huang, J. Gu, L. Hood, L. Rowen, A. Madan, S. Z. Qin, R. W. Davis, N. A. Federspiel, A. P. Abola, M. J. Proctor, R. M. Myers, J. Schmutz, M. Dickson, J. Grimwood, D. R. Cox, M. V. Olson, R. Kaul, C. Raymond, N. Shimizu, K. Kawasaki, S. Minoshima, G. A. Evans, M. Athanasiou, R. Schultz, B. A. Roe, F. Chen, H. Q. Pan, J. Ramser, H. Lehrach, R. Reinhardt, W. R. McCombie, M. de la Bastide, N. Dedhia, H. Blocker, K. Hornischer, G. Nordsiek, R. Agarwala, L. Aravind, J. A. Bailey, A. Bateman, S. Batzoglou, E. Birney, P. Bork, D. G. Brown, C. B. Burge, L. Cerutti, H. C. Chen, D. Church, M. Clamp, R. R. Copley, T. Doerks, S. R. Eddy, E. E.

Eichler, T. S. Furey, J. Galagan, J. G. R. Gilbert, C. Harmon, Y. Hayashizaki, D. Haussler, H. Hermjakob, K. Hokamp, W. H. Jang, L. S. Johnson, T. A. Jones, S. Kasif, A. Kasprzyk, S. Kennedy, W. J. Kent, P. Kitts, E. V. Koonin, I. Korf, D. Kulp, D. Lancet, T. M. Lowe, A. McLysaght, T. Mikkelsen, J. V. Moran, N. Mulder, V. J. Pollara, C. P. Ponting, G. Schuler, J. R. Schultz, G. Slater, A. F. A. Smit, E. Stupka, J. Szustakowski, D. Thierry-Mieg, J. Thierry-Mieg, L. Wagner, J. Wallis, R. Wheeler, A. Williams, Y. I. Wolf, K. H. Wolfe, S. P. Yang, R. F. Yeh, F. Collins, M. S. Guyer, J. Peterson, A. Felsenfeld, K. A. Wetterstrand, A. Patrinos, M. J. Morgan and I. H. G. S. Conso: Initial sequencing and analysis of the human genome. *Nature*, 409(6822), 860-921 (2001)
DOI: 10.1038/35057062

93. N. Cancer Genome Atlas Research, J. N. Weinstein, E. A. Collisson, G. B. Mills, K. R. Shaw, B. A. Ozenberger, K. Ellrott, I. Shmulevich, C. Sander and J. M. Stuart: The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet*, 45(10), 1113-20 (2013)
DOI: 10.1038/ng.2764
94. A. Regev, S. A. Teichmann, E. S. Lander, I. Amt, C. Benoist, E. Birney, B. Bodenmiller, P. Campbell, P. Carninci, M. Clatworthy, H. Clevers, B. Deplancke, I. Dunham, J. Eberwine, R. Elis, W. Enard, A. Farmer, L. Fugger, B. Gottgens, N. Hacohen, M. Haniffa, M. Hemberg, S. Kim, P. Klenerman, A. Kriegstein, E. D. Lein, S. Linnarsson, E. Lundberg, J. Lundeberg, P. Majumder, J. C. Marioni, M. Merad, M. Mhlanga, M. Nawijin, M. Netea, G. Nolan, D. Pe'er, A. Phillipakis, C. P. Ponting, S. Quake, W. Reik, O. Rozenblatt-Rosen, J. Sanes, R. Satija, T.

- N. Schumacher, A. Shalek, E. Shapiro, P. Sharma, J. W. Shin, O. Stegle, M. Stratton, M. J. T. Stubbington, F. J. Theis, M. Uhlen, A. Van Oudenaarden, A. Wagner, F. Watt, J. Weissman, B. Wold, R. Xavier, N. Yosef and H. C. A. Meeting: The Human Cell Atlas. *Elife*, 6 (2017)
DOI: 10.7554/eLife.27041
95. F. Alber, F. Forster, D. Korkin, M. Topf and A. Sali: Integrating diverse data for structure determination of macromolecular assemblies. *Annu Rev Biochem*, 77, 443-77 (2008)
DOI: 10.1146/annurev.biochem.77.060407.135530
96. J. F. Lin and D. Nicastro: Asymmetric distribution and spatial switching of dynein activity generates ciliary motility. *Science*, 360(6387) (2018)
DOI: 10.1126/science.aar1968
97. C. M. Hampton, J. D. Strauss, Z. L. Ke, R. S. Dillard, J. E. Hammonds, E. Alonas, T. M. Desai, M. Marin, R. E. Storms, F. Leon, G. B. Melikyan, P. J. Santangelo, P. W. Spearman and E. R. Wright: Correlated fluorescence microscopy and cryo-electron tomography of virus-infected or transfected mammalian cells. *Nature Protocols*, 12(1) (2017)
DOI: 10.1038/nprot.2016.168
98. V. B. S. Chan, M. B. Johnstone, A. P. Wheeler and A. S. Mount: Chitin Facilitated Mineralization in the Eastern Oyster, 5(347) (2018)
DOI: 10.3389/fmars.2018.00347
99. R. Jungmann, M. S. Avendano, J. B. Woehrstein, M. Dai, W. M. Shih and P. Yin: Multiplexed 3D cellular super-resolution imaging with DNA-PAINT and Exchange-PAINT. *Nat Methods*, 11(3), 313-8 (2014)
DOI: 10.1038/nmeth.2835
100. D. A. Keedy, Z. B. Hill, J. T. Biel, E. Kang, T. J. Rettenmaier, J. Brandao-Neto, N. M. Pearce, F. von Delft, J. A. Wells and J. S. Fraser: An expanded allosteric network in PTP1B by multitemperature crystallography, fragment screening, and covalent tethering. *Elife*, 7 (2018)
DOI: 10.7554/eLife.36307.001

Key Words: Integrative Methods, Hybrid Methods, Structural Biology, Workshop, Review

Send correspondence to: Hugo Sanabria, Physics and Astronomy, Clemson University, 214 Kinard Lab, Clemson, S.C. 29634-0978, Tel: 864-656-1749, Fax: 864-656-0805, E-mail: hsanabr@clemson.edu