Together with our collaborators, we have established its essential metabolic reactions and kinetic models for genetic information processing, and provided genome-wide gene essentiality and proteomics data[2,3]. The protein products of 155 genes are involved in 174 metabolic reactions that are organized into 9 subsystems. 251 genes participate in the ~2,000 reactions needed for the genetic information processes of DNA replication, transcription, translation, mRNA degradation, tRNA charging and cell growth, leaving approximately 87 genes of unknown function—a number that is steadily decreasing but may represent new biochemical reactions. In addition to reactions, whole-cell, spatially resolved kinetic models require cellular architecture, including spatial distributions of ribosomes and the circular chromosome's conformations. The cellular architecture is reconstructed at the single-cell level directly from cryo-electron tomograms that include ribosome distributions. For each replicate studied, self-avoiding circular chromosome configurations are generated with a resolution of 11.8 bp per monomer embedded in a lattice representation of the entire cell. From the ensembles of simulated chromosome configurations, we derive contact maps of 1,000 bp that are in agreement with preliminary experimental chromosome conformation capture maps.

To create a whole-cell kinetic model of growth behavior, experimental data and theoretical or computational models of the various processes have to be integrated and validated via simulations over a cell cycle (Fig. 1). Validating the balance sheet for energy generation and costs is an important next step that will require the ability to measure certain metabolites and cellular intermediates as a function of time. Because of the large variation in time scales and concentrations, these goals can only be reached by combining hybrid stochastic and deterministic simulations using graphical processing unit (GPU)-based software like our Lattice Microbes program[4]. GPU computer clusters represent yet another revolution necessary for bringing structural biology and its wealth of information to understanding how even a minimal cell functions. Once these goals are achieved, the cell model can be used to make predictions about responses to the environment, mutations, and rules governing the correlations between the various cellular processes. ❐

Zaida Luthey-Schulten ✉
Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, IL, USA.
✉*e-mail: zan@illinois.edu*

Published online: 7 May 2021
https://doi.org/10.1038/s41592-021-01150-2

References
1. Hutchison, C. A. III et al. *Science* **351**, aad6253 (2016).
2. Breuer, M. et al. *eLife* **2019**, e36842 (2019).
3. Thornburg, Z. R. et al. *Front. Mol. Biosci.* **6**, 130 (2019).
4. Hallock, M. J., Stone, J. E., Roberts, E., Fry, C. & Luthey-Schulten, Z. *Parallel Comput.* **40**, 86–99 (2014).

Competing interests
The author declares no competing interests.

50 years of PDB: a catalyst in structural biology

Integrative structural biology, the culmination of experimental and theoretical methods, will provide a holistic view of molecular processes.

Alexandre M. J. J. Bonvin

We are celebrating the fiftieth anniversary of the Protein Data Bank (PDB)[1], a resource that has shaped the way we do structural biology. What started in 1971 as a simple collection of ASCII files containing three-dimensional (3D) atomic coordinates of biological molecules has grown into an indispensable resource, now operated as a single worldwide archive managed by the Worldwide PDB Consortium (wwPDB)[2]. This consortium has evolved over the years to include the experimental data resource BioMagResBank (BMRB)[3], with the cryo-electron microscopy (cryo-EM) database EMDB soon to become the latest addition. What is even more remarkable (but also scary) is that this essential resource has been operating mainly on soft money, relying on successful grant applications to survive.

The field of structural biology and our general knowledge of the molecules of life would not be the same without the PDB, which has acted as a catalyst for developments in both experimental and, more importantly, computational methods. Any computational approach used routinely for structure determination or molecular simulations has been developed and benchmarked using PDB data. The same applies to many force fields and statistical potentials used in computational structural biology. Deep learning, with the success of AlphaFold2 for 3D structure prediction as a recent example, would not have reached such achievements without the rich information available in the PDB.

This catalyst role also applies to experimental methods, especially in the current era of integrative structural biology. While X-ray crystallography and, to a smaller extent, nuclear magnetic resonance have long been the main data contributors to the PDB, cryo-EM has become a major player over the last decade. In recent years we have moved into a new era where many different experimental and bioinformatics approaches are contributing pieces of a complex puzzle. As we are tackling larger and more intricate assemblies and molecular machineries, this complexity is constantly increasing. As such, no single experimental method can provide all answers; only a combination of those together with computational methods can allow us to model those complex systems. For example, mass spectrometry has moved into structural proteomics, contributing valuable information in the form of cross-links and/or hydrogen/deuterium exchange data; various spectroscopic methods allow measurement of very specific and accurate distances; and large-scale mutagenesis and deep mutational scans are providing information about key residues for both structure and interactions. This new era of structural biology also means that the distinction between models and experimental structures is becoming

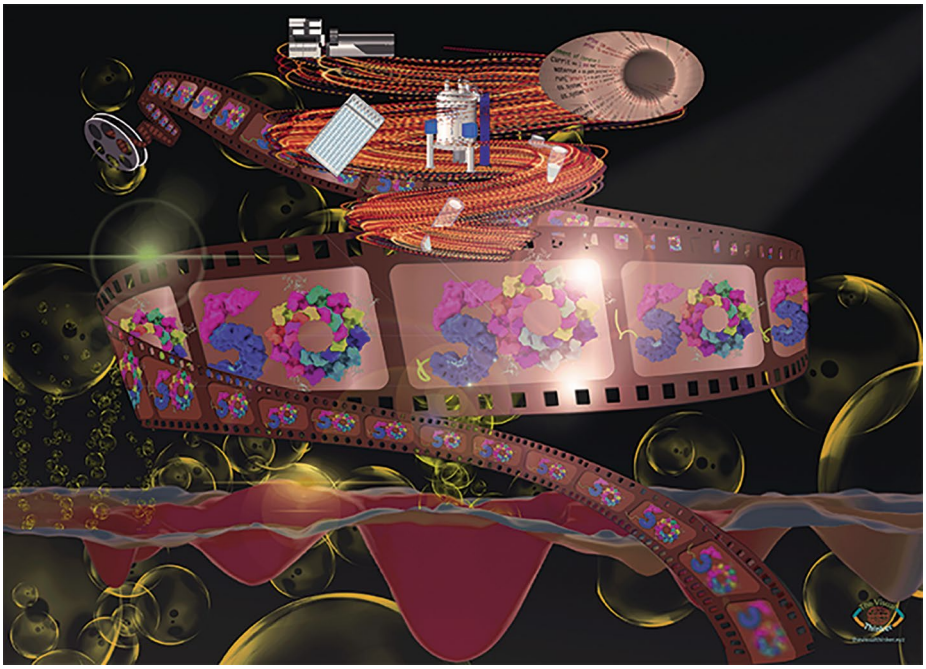

Fig. 1 | Artist's conception of integrative structural biology of dynamical landscapes. Using a variety of experimental techniques together with computational methods, integrative models describing multiple states of an assembly (the "50 years of PDB complex") connected by a timeline (dynamics) are obtained. Image courtesy of Dr. Gloria Fuentes (thevisualthinker.xyz).

less and less clear-cut. To react to such developments, the wwPDB regularly brings together community experts to define new guidelines and standards. The hybrid/integrative methods task force[4] and the integrative models prototype deposition system are recent examples (https://pdb-dev.wwpdb.org).

With our structural biology toolbox expanding, we are also increasingly realizing the complexity of these intricate molecular machines: as the name machine implies, these consist of parts that move at different speeds and that can take different states. As such, the experimental data often reflect this diversity and dynamics. For example, through image classification in cryo-EM one can visualize different states of a molecular assembly, and invisible states can be made visible by NMR, which can also add dynamical information to the descriptions of those machines. Static 3D structures are being replaced by an ensemble view of states and their interconnected dynamics, ideally in the context of the cellular environment. This is even more important when considering the substantial fraction of disordered proteins or regions thereof in the human proteome, many of which are involved in key interactions. This complex landscape itself is dynamical as it is modulated by all kinds of events, such as post-translational modifications, localization within the cell and/or time within the cell cycle.

There is a clear challenge in capturing, properly describing, visualizing and archiving the conformational and temporal datasets that will describe these complex, heterogeneous and dynamical landscapes. As we are moving from integrative structural biology to what I would call the integrative structural biology of dynamical landscapes (Fig. 1), the wwPDB is facing the challenge of collecting all this information. Fifty years from now, when we will be celebrating its 100th anniversary, I hope we can state that it also acted as a catalyst for a fully dynamical description of life at the molecular level. ❐

Alexandre M. J. J. Bonvin ✉
Science for Life, Bijvoet Centre for Biomolecular Research, Faculty of Science, Department of Chemistry, Utrecht University, Utrecht, the Netherlands.
✉*e-mail: a.m.j.j.bonvin@uu.nl*

Published online: 7 May 2021
https://doi.org/10.1038/s41592-021-01138-y

References

1. wwPDB Consortium. *Nucleic Acids Res.* **47**, D520–D528 (2019). (D1).
2. Berman, H., Henrick, K. & Nakamura, H. *Nat. Struct. Biol.* **10**, 980 (2003).
3. Ulrich, E. L. et al. *Nucleic Acids Res.* **36**, D402–D408 (2008).
4. Sali, A. et al. *Structure* **23**, 1156–1167 (2015).

Competing interests

The author is a member of the scientific advisory board of the Molecular & Cellular Structure cluster of the European Bioinformatics Institute (EBI), operating the PDB Europe site (https://www.ebi.ac.uk/pdbe/about/sac).