



Happy anniversary, PDB!

We celebrate the 50th anniversary of the Protein Data Bank together with our colleagues at *Nature Methods* with a special collection that showcases key achievements in structural biology and views of its future.

In October 1971, the establishment of a central open repository for macromolecular structure data, the Protein Data Bank (PDB), was announced in *Nature New Biology* (*Nat. New Biol.* **233**, 223 (1971)). This new repository, which at the time contained seven structures, was the culmination of grassroots efforts led by a cadre of protein crystallographers who were keenly aware of the value of archiving and sharing X-ray crystallography data, including atomic coordinates, structure factors and electron density maps. Initially, it was operated by Brookhaven National Laboratory and the Cambridge Crystallographic Data Centre. Since 2003, the Worldwide PDB ([wwPDB](#)) has been distributed from centers located on three continents: the Research Collaboratory of Structural Bioinformatics (RCSB) PDB in the US, the EMBL-EBI's PDB in Europe (PDBe) and PDB Japan (PDBj). Together, these centers maintain a single archive of macromolecular structure data that is freely and publicly available to the global community (*Nat. Struct. Mol. Biol.* **10**, 980 (2003)).

PDB data are used by researchers worldwide to explore fundamental questions about the mechanism and function of biological macromolecules and how they can be manipulated to combat disease. As a prime example of the latter, specially prepared PDB resources provide access to all available SARS-CoV-2 PDB structures and highlight their important structural features to support structure-based drug design and vaccine development in the fight against the COVID-19 pandemic.

PDB resources are also used beyond the research community. [PDB-101](#), the PDB's education portal, provides teaching materials and introduces students and the wider public to the beauty of the molecular basis of life.

The visionaries who set up the PDB pioneered data-sharing practices at a time when this was still foreign to many other scientific disciplines. The creation

of the PDB was also instrumental in the development of software tools that allow the visualization, validation, analysis and storage of protein structure data. In the early days, facilitating data access was far from trivial. The process that was created to allow remote computers to connect and to search data stored at Brookhaven National Laboratory was a forerunner to the internet-based systems that we do not think twice about today. As a testament to its success, the PDB now hosts more than 175,000 structures. The vast majority have been obtained by X-ray crystallography, but the PDB is also the home for atomic models obtained by NMR spectroscopy (via the Biological Magnetic Resonance DataBank, or [BMRB](#)) and by 3D electron microscopy (via the Electron Microscopy Data Bank, or [EMDB](#)).

As Helen Berman describes in her Comment "[Synergies between the Protein Data Bank and the community](#)", such an accomplishment would not have been possible if the endeavor had not been a community initiative. It required extensive effort, appeals and painstaking consensus building to bring together the heterogeneous research communities served by the PDB, with their different interests and needs, to develop common standards for data sharing and to ensure the success that the PDB now enjoys.

The PDB has played a central role in shaping how we see the molecular world. Our ability to visualize the diverse shapes of proteins and nucleic acids, and thus to virtually build cells from molecules and organisms from cells, helps us to understand life at a fundamental level. Artistic techniques are essential for this visualization, and art can contribute substantially as a means of scientific discovery, as elaborated on in the Comment "[Art as a tool for science](#)" by David Goodsell. Creating illustrations of concepts and models can reveal discrepancies and gaps in our knowledge and inspire new hypotheses and questions.

Structural biologists have traditionally used a 'divide-and-conquer' approach to address these complex questions, analyzing single proteins or protein domains. This tactic has in part been necessary due to technical limitations, but reconstituting a biological system from its individual components in vitro is also an important pathway toward new knowledge. As Richard Feynman aptly put it, "What I cannot create, I do not understand."

Despite the success of such reductionist approaches, molecules do not act in isolation. It is essential to analyze biological macromolecules in their physiological context to appreciate their spatial relationships and functional interactions. Recent technological developments, for example in cryo-electron tomography, have made it possible to view molecular complexes in their native environment and to start understanding the molecular architecture of tissues and organisms. Moreover, integrative or hybrid modelling has enabled structural models to be built using a combination of experimental and computational techniques; these models can currently be deposited and shared via the prototype archive [PDB-Dev](#).

These and other advances are discussed in a series of Comments in *Nature Methods*, introduced in the journal's [Editorial](#) for the May issue, in which researchers from diverse areas of structural biology share their views on the challenges and opportunities that lay ahead.

The fact is, no single technique will allow us to obtain a mechanistic understanding of complete biological systems; only the integration of multidisciplinary approaches and new synergies between cellular and molecular biology will achieve this feat. We look forward to seeing these developments come to light, both in our pages and elsewhere. □

Published online: 7 May 2021
<https://doi.org/10.1038/s41594-021-00598-2>