Protein Data Bank: 100,000 structures

Wednesday May 14, 2014 00:00 UTC (Coordinated Universal Time) Spanning the
globe from the US, UK, and Japan, the Worldwide Protein Data Bank (wwPDB)
organization announces that the Protein Data Bank archive now contains more
than 100,000 entries.

Established in 1971, this central, public archive of experimentally-determined
protein and nucleic acid structures has reached a critical milestone thanks to the
efforts of structural biologists throughout the world.

Four wwPDB data centers support online access to three-dimensional structures
of biological macromolecules that help researchers understand many facets of
biomedicine, agriculture, and ecology, from protein synthesis to health and
disease to biological energy.

Function follows form
In the 1950s, scientists had their first direct look at the structures of proteins and DNA at the atomic
level. Determination of these early three-dimensional structures by X-ray crystallography ushered in a
new era in biology—one driven by the intimate link between form and biological function. As the value
of archiving and sharing these data were quickly recognized by the scientific community, the Protein
Data Bank (PDB) was established as the first open access digital resource in all of biology by an
international collaboration in 1971 with data centers located in the US and the UK.

Among the first structures deposited in the PDB were those of myoglobin and hemoglobin, two oxygen-
binding molecules whose structures were elucidated by Chemistry Nobel Laureates John Kendrew and
Max Perutz. With this week's regular update, the PDB welcomes 219 new structures into the archive.
These structures join others vital to drug discovery, bioinformatics, and education, for a total of 100,147
entries.

The PDB is growing rapidly, doubling in size since 2008, and releasing around 200 new structures to the
scientific community every week. The resource is accessed hundreds of millions of times annually by
researchers, students, and educators intent on exploring how different proteins are related to one
another, to clarify fundamental biological mechanisms and discover new medicines.

“The PDB is a critical resource for the international community of working scientists which includes
everyone from geneticists to pharmaceutical companies interested in drug targets,” said Nobel Laureate
Venki Ramakrishnan of the MRC Laboratory of Molecular Biology in Cambridge, UK.
A growing community
Since its inception, the PDB has been a community-driven enterprise, evolving into a mission critical international resource for biological research. Since 2003 the Worldwide PDB (wwPDB) organization, a collaboration involving four PDB data centers in the US, UK, and Japan, has ensured that these valuable data are securely stored, expertly managed, and made freely available for the benefit of scientists and educators around the globe. wwPDB data centers work closely with community experts to define deposition and annotation policies, resolve data representation issues, and implement community validation standards. In addition, the wwPDB works to raise the profile of structural biology with increasingly broad audiences.

Each structure submitted to the archive is carefully curated by wwPDB staff before release. New depositions are checked and enhanced with value-added annotations and linked with other important biological data to ensure that PDB structures are discoverable and interpretable by users with a wide range of backgrounds and interests.

Future challenges
The scientific community eagerly awaits the next 100,000 structures and the invaluable knowledge these new data will bring. However, the increasing number, size and complexity of biological data being deposited in the PDB and the emergence of hybrid structure determination methods, which use a variety of biophysical, biochemical, and modelling techniques to determine the shapes of biologically relevant molecules, constitute major challenges for the management and representation of structural data. wwPDB will continue to work with the community to meet these challenges and ensure that the archive maintains the highest possible standards of quality, integrity, and consistency.

About the wwPDB
The wwPDB (http://wwpdb.org) is the international partnership of four data centers that manage the PDB archive. Its mission is to maintain a single archive of macromolecular structural data that is freely and publicly available to the global community. It consists of the Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB; http://rcsb.org) at Rutgers, The State University of New Jersey and the San Diego Supercomputer Center (SDSC) and Skaggs School of Pharmacy and Pharmaceutical Sciences at the University of California San Diego and BioMagResBank (BMRB; http://bmrb.wisc.edu) at the University of Wisconsin in the USA, the Protein Data Bank in Europe (PDBe; http://pdbe.org) at the EMBL European Bioinformatics Institute, and the Protein Data Bank Japan (PDBj; http://pdbj.org) at Osaka University.

The RCSB PDB receives funds from the NSF, NIH, and DOE. BioMagResBank is funded by the NLM. The PDBe receives funding from EMBL, the Wellcome Trust, NIH, EU, BBSRC, and MRC. PDBj is funded by the Japan Science and Technology Agency.
Contacts:
US: Christine Zardecki, RCSB Protein Data Bank, Rutgers, The State University of New Jersey; 848 445 0103, info@rcsb.org

UK: Gary Battle, Protein Data Bank in Europe, European Bioinformatics Institute (EMBL-EBI), Cambridge (UK); +44 1223 49-4654, battle@ebi.ac.uk

Japan: Nahoko Haruki, Protein Data Bank Japan, Institute for Protein Research, Osaka University; +81 6 6879 4311, nahokoh@protein.osaka-u.ac.jp

Number of structures available in the PDB per year through May 14, 2014, with selected examples. Early structures included myoglobin (1; PDB ID 1mbn), the first structure solved by X-ray crystallography, and small enzymes (2; top: 4pti, bottom right: 2cha, bottom left: 3cpa). As technologies developed, the archive grew to host examples of tRNA (3; 6tna), viruses (4; 4rhv), antibodies (5; 1igt), protein-DNA complexes (6; top to bottom, 1j59, 1tro, 2bop, 1aoi), ribosomes (7; 1fjg, 1fka, 1ffk), and chaperones (8; 1aon).

Figure references