

RCSB PROTEIN DATA BANK: SUSTAINING A LIVING DIGITAL DATA RESOURCE THAT ENABLES BREAKTHROUGHS IN SCIENTIFIC RESEARCH AND BIOMEDICAL EDUCATION

Christine Zardecki (christine.zardecki@rcsb.org), Helen M. Berman (helen.berman@rcsb.org), Cole Christie (cole.christie@rcsb.org), Jose M. Duarte (jose.duarte@rcsb.org), Zukang Feng (zukang.feng@rcsb.org), John Westbrook (john.westbrook@rcsb.org) Jasmine Young (jasmine.young@rcsb.org) Stephen K. Burley (stephen.burley@rcsb.org)

The Protein Data Bank (PDB) was established in 1971 as the first open-access digital data resource in biology. Beginning with only seven protein structures, the PDB archive has ballooned to >138,000 structures of proteins, DNA, and RNA (totaling >1 billion atoms). Today, the PDB is universally regarded as a core data science resource of fundamental importance to the wider life-science community and long-term preservation of machine-readable biological data. PDB structures are the molecules of life. Knowledge of 3D structures (shapes) of biomolecules, how they evolve with time, and how they function in nature is essential for understanding critical areas of science. PDB data impacts basic and applied research on health and disease of humans, animals, and plants; production of food and energy; and other research pertaining to global prosperity and environmental sustainability. Structure data are also important to biopharmaceutical and biotechnology companies, accelerating data-driven discovery of new drugs, materials, and devices. Today, powerful pulsed X-ray facilities, cryogenic electron microscopes, and new integrative/hybrid (I/H) methods for structure determination are accelerating biomedical research with functional insights into ever more complex biological systems at the atomic level. Cryo-electron tomography even allows study of molecular machines “caught in the act” inside frozen cells.

The PDB is managed by the Worldwide Protein Data Bank partnership (wwPDB; wwpdb.org). RCSB PDB (RCSB.org) operates the US wwPDB data center, and makes PDB data available at no charge and without limitations. Studies of website usage, bibliometrics, and economics demonstrate the powerful impact of the PDB data on basic and applied research, clinical medicine, education, and the economy. During calendar 2017, ~680 million structure data files were downloaded by *Data Consumers* worldwide. During this same period, RCSB PDB processed >6,200 new atomic level biomolecular structures plus experimental data and metadata coming submitted by *Data Depositors* in the Americas and Oceania. More than >1 million RCSB.org users were served with PDB data integrated with ~40 external resources providing rich structural views of fundamental biology, biomedicine, and energy sciences.

Access to PDB data contribute to patent applications, drug discovery and development, publication of scientific studies, innovations that can lead to new product development and company formation, and STEM education.

RCSB PDB is funded by the NSF (DBI-1338415), NIH, and DOE.