

RCSB Protein Data Bank: Integrated Searching and Efficient Access to Macromolecular Structure Data from the PDB Archive

Brian Hudson¹, Yana Rose², Jose M. Duarte³, Robert Lowe⁴, Chunxiao Bi⁵, Charmi Bhikadiya⁶, Li Chen⁷, Sebastian Bittrich⁸, Joan Segura⁹, Stephen Burley¹⁰, John Westbrook¹¹, Alexander S. Rose¹²

¹RCSB Protein Data Bank ²RCSB Protein Data Bank, ³RCSB Protein Data Bank, ⁴RCSB Protein Data Bank, ⁵RCSB Protein Data Bank, ⁶RCSB Protein Data Bank, ⁷RCSB Protein Data Bank, ⁸RCSB Protein Data Bank, ⁹RCSB Protein Data Bank, ¹⁰RCSB Protein Data Bank, Rutgers University, ¹¹No affiliation given, ¹²RCSB Protein Data Bank

hudson@rcsb.rutgers.edu

The US Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB) serves many millions of unique users worldwide by delivering experimentally determined 3D structures of biomolecules integrated with more than 40 external data resources via RCSB.org, application programming interfaces (APIs), and FTP downloads. Herein, we present the architectural redesign of RCSB PDB data delivery services that build on existing PDBx/mmCIF data schemas. New data access APIs (data.rcsb.org) enable efficient delivery of all PDB archive data. A novel GraphQL-based API provides flexible, declarative data retrieval along with a simple-to-use REST API. A powerful new search system (search.rcsb.org) seamlessly integrates heterogeneous types of searches across the PDB archive. Searches may combine text attributes, protein or nucleic acid sequences, small-molecule chemical descriptors, 3D macromolecular shapes, and structure and sequence motifs. The new RCSB.org architecture adheres to the FAIR Principles, empowering users to address a wide array of research problems in fundamental biology, biomedicine, biotechnology, bioengineering, and bioenergy.