The next generation RCSB.org

Justin Flatt[1], Jose M. Duarte[2], Charmi Bhikadiya[3], Chunxiao Bi[4], Sebastian Bittrich[5], Li Chen[6], Shuchismita Dutta[7], Robert Lowe[8], Alexander S. Rose[9], Yana Rose[10], Joan Segura[11], John Westbrook[12], Jasmine Y Young[13]

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

[13]Rutgers University

justin@rcsb.rutgers.edu

The US Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB) provides tools for analysis and visualization of 3D structures of biological macromolecules stored in the PDB archive. A complete redesign of the RCSB.org user interfaces and backend services represents a full overhaul of the software/data management architecture, transforming a monolithic application into a micro-service-oriented and cloud-ready resource.

The new search system allows for queries with arbitrary Boolean logic. The queries can combine specialized bioinformatic search algorithms (sequence search, structure search, small molecule search, structure or sequence motif search) with text search of any attribute present in PDB files or integrated external data resources. The search results can be presented at different PDB data hierarchal levels: entry, polymer entity, assembly and non-polymer entity (selectable with the "Display results as" drop-down menu).

Rapid, interactive 3D visualization of macromolecules is enabled by Mol*, a molecular viewer developed as a collaboration between RCSB PDB, PDBe and the Central European Institute of Technology.

The in-house developed Protein Feature View is based on modern web technologies, allowing facile visualization of sequence features from PDB and annotations from external resources, such as domains or PDB validation data.

Improved carbohydrate data are presented at the structure summary page with 2D SNFG symbols and in Mol* viewer with 3D SNFG symbols. Users can perform advanced searches to find carbohydrate entries with glycosylation in the Polymer Molecular Features, as well as entries with Oligosaccharide Features by searching any of these attributes: linear descriptors, systematic name, monosaccharide composition, etc.

A new pairwise Structure Alignment application offers some of the most popular structure alignment algorithms (FatCat, CE, TMalign) with integrated visualization in the Mol* viewer. Structure motif searches can be carried out using an interface integrated into Mol*. And ultrafast global shape search, which relies on the newly designed BioZernike descriptor, finds similar assemblies in the PDB within milliseconds.