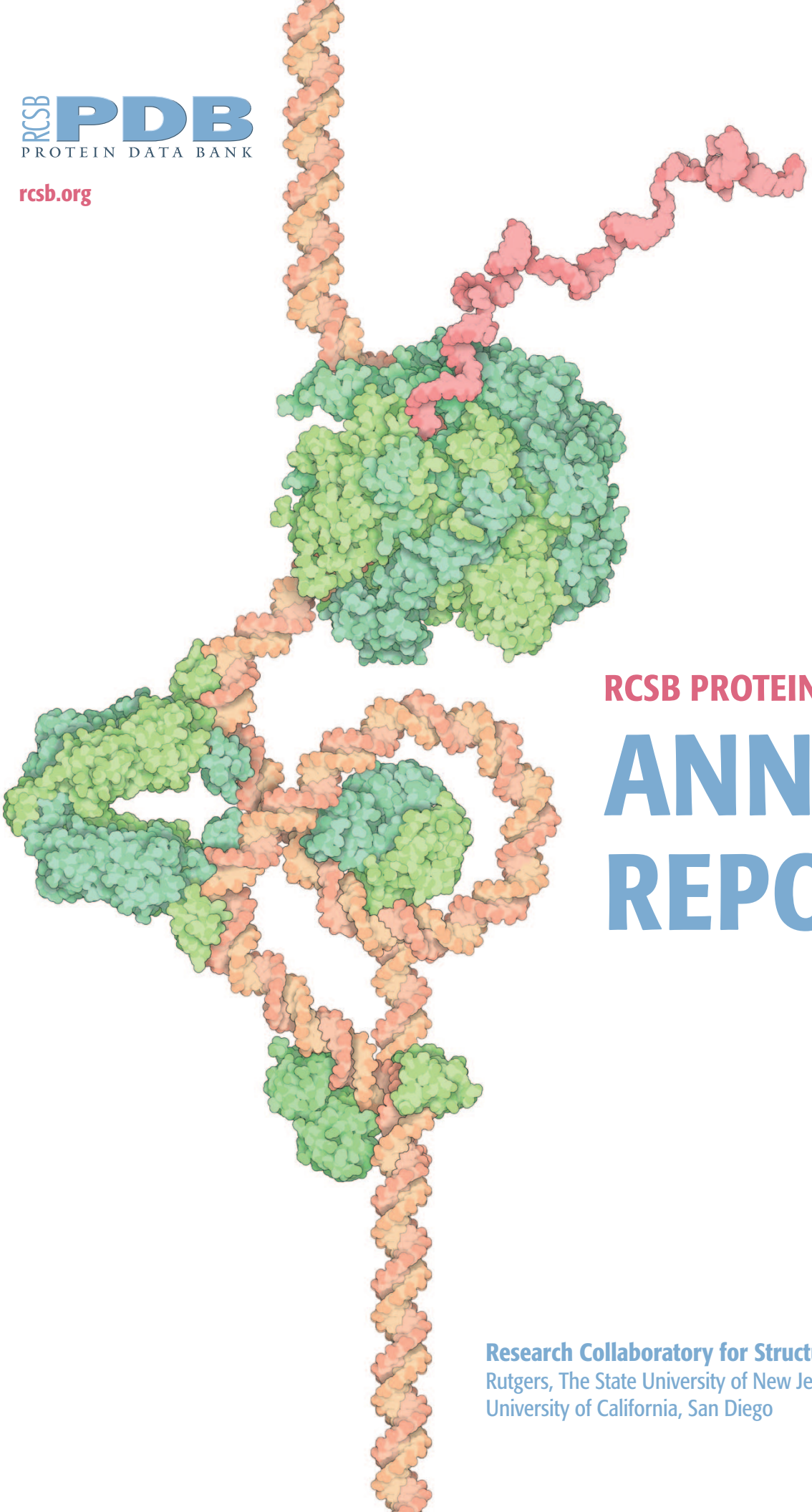
RCSB PDB
PROTEIN DATA BANK

rcsb.org



RCSB PROTEIN DATA BANK

ANNUAL REPORT 2014

Research Collaboratory for Structural Bioinformatics
Rutgers, The State University of New Jersey
University of California, San Diego

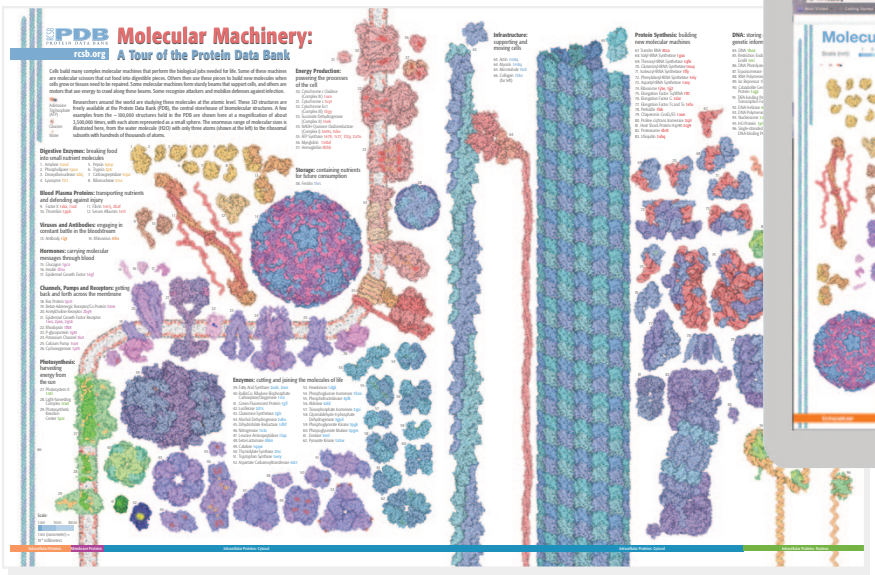
ABOUT THE COVER

When the PDB archive was established in 1971, it contained seven structures. In 2014, the PDB passed the milestone of making more than 100,000 structures publicly available for scientific research and education.

To commemorate this special milestone, RCSB PDB released an updated version of the iconic poster *Molecular Machinery: A Tour of the Protein Data Bank*. This cover highlights a portion of this poster.

The original image, published in 2002 and created by *Molecule of the Month* author David S. Goodsell, featured 75 select PDB structures with relative sizes at a scale of three million to one. Made available as a poster and handout, *Molecular Machinery* had been distributed to thousands of PDB users, teachers, and students over the years.

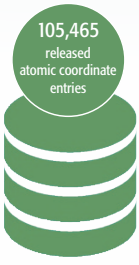
A new version of the poster, published in 2014 uses 96 structures to illustrate the vast range of molecular shapes and sizes in the PDB. Structures are depicted relative to the cellular membrane and organized in categories related to function. A scale bar provides a sense of molecular size in nanometers. This new *Molecular Machinery* poster is available as an interactive website (mm.rcsb.org) and as a poster and flyer (rcsb.org/pdb-101).

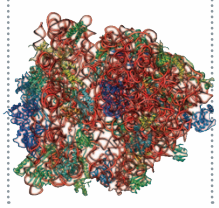


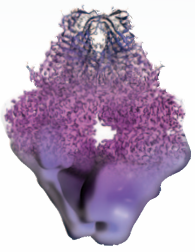
Download at rcsb.org/pdb-101

Explore at mm.rcsb.org

TABLE OF CONTENTS







MESSAGE FROM THE RCSB PDB

2014 was the International Year of Crystallography (IYCr2014), jointly designated by the United Nations Educational, Scientific and Cultural Organization (UNESCO) and the International Union of Crystallography. It was a year replete with milestones for the Protein Data Bank archive and the RCSB PDB.

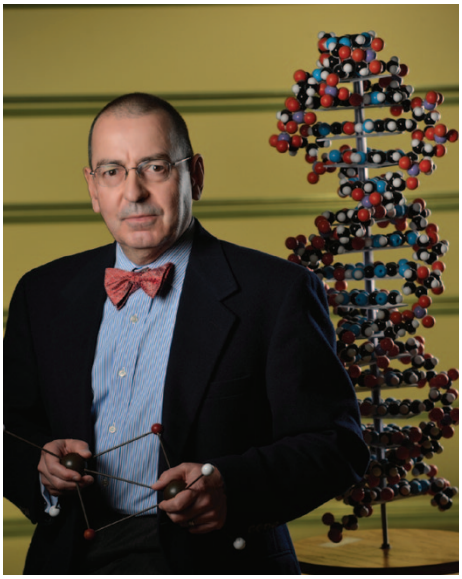
More than forty years after the first seven PDB structures were made freely available, the 100,000th archival entry was released on May 13th. Earlier in the year, the RCSB PDB and our wwPDB partners began using the new Deposition & Annotation System, which was designed to support the next generation of PDB structure data. Large and complex structures, historically split across multiple PDB IDs, were reunited into single entries designated by just one ID. Along with a new look and feel, additional resources were launched at rcsb.org, including tools for analyzing sequence and gene information for proteins.

Leadership transitions also occurred in 2014. In July, Dr. Stephen K. Burley was named Director of the RCSB Protein Data Bank. Burley is an internationally recognized physician-scientist and an authority on structural biology, drug discovery, clinical medicine, and oncology with considerable experience in both academe and industry.

Previously, Dr. Helen M. Berman had led all aspects of the project since it came to Rutgers in 1999, for which she received many national and international awards and prizes recognizing her vision for the PDB resource and commitment to our diverse user communities. Since the leadership transition, she has continued her more than four-decade involvement with the PDB in a new role as RCSB PDB Associate Director. With a focus on External Partnerships, Helen has been leading activities related to hybrid/integrative structure determination methods, including the Task Force held in UK in October. Berman, a Board of Governors Professor of Chemistry and Chemical Biology at Rutgers, also continues her leadership roles in the Nucleic Acid Database, the Structural Biology Knowledgebase, and the EMDataBank.

Burley joined Rutgers in January 2013 to become a Distinguished Professor of Chemistry and Chemical Biology, Director of the Center for Integrative Proteomics Research, and a Member of the Cancer Institute of New Jersey. For the RCSB PDB, Burley led outreach efforts as an Associate Director, and assumed leadership of the RCSB PDB team at UCSD earlier in the year, when former Associate Director Phil Bourne became Associate Director for Data Science at the National Institutes of Health. Bourne was part of the original team led by Berman that brought management of the PDB to the RCSB more than 15 years ago. As part of the RCSB PDB, Bourne oversaw many important projects, including database development, literature integration, and visualization tools.

In 2015, RCSB PDB continues to build upon the milestones and leadership foundations established in 2014 to develop and enhance structural views of biology that will inspire, enable, and inform our diverse community of users worldwide.



Director Stephen K. Burley

Stephen K. Burley with Helen M. Berman

OVERVIEW: ENABLING BREAKTHROUGHS

VISION: RCSB PDB provides rich structural views of biological systems to enable breakthroughs in scientific inquiry, medicine, drug discovery, technology, and education.[1,2]

These views are part of a freely-available resource that is built on top of experimental 3D biological macromolecular structure data. This resource is supported by its ability to:

- **ENABLE** efficient deposition, high-quality curation, and exploration of data in the global PDB archive;
- **LEAD** biological structure representation and drive integration with related data resources;
- **ESTABLISH** and **FOSTER** communication and collaboration across sciences; and
- **INSPIRE** and **INFORM** diverse users through structural views of biology and medicine.

Growth of the PDB

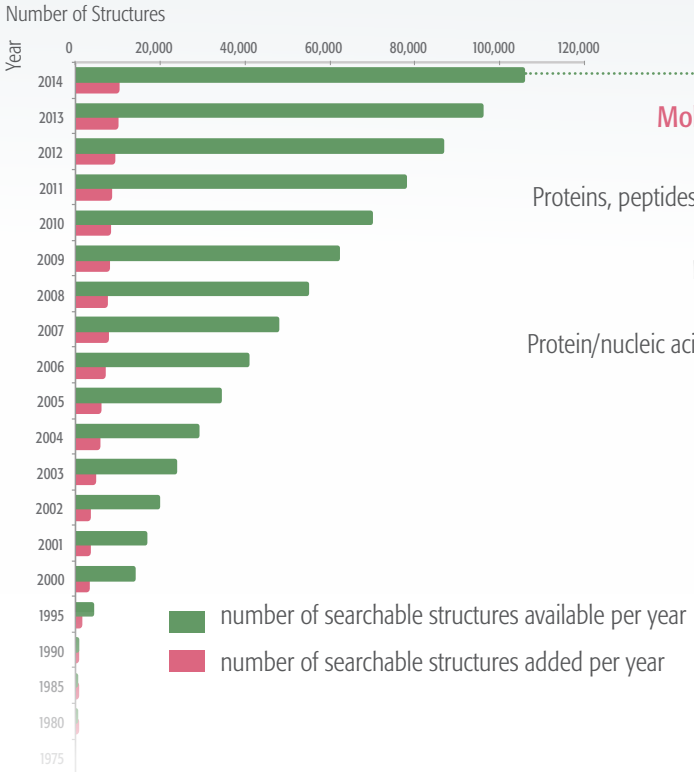
In the 1950s, scientists had their first direct look at the structures of proteins and DNA at the atomic level. Determination of these early three-dimensional structures by X-ray crystallography ushered in a new era in biology-one driven by the intimate link between form and biological function. As the value of archiving and sharing these data were quickly recognized by the scientific community, the Protein Data Bank (PDB) was established in 1971 as the first open access digital resource in biology.[3-5]

Today, the PDB archive is managed by the Worldwide PDB organization (wwPDB, wwpdb.org),[6] a collaboration involving four PDB data centers in the US,[1,7] UK,[8] and Japan.[9] wwPDB ensures 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.

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, clarifying fundamental biological mechanisms, and discovering new medicines.

As a member, RCSB PDB collaborates with the wwPDB partners on matters relating to data deposition, annotation, and validation, while independently developing tools and resources in the form of the RCSB PDB website, tools, and other resources that enable the access and analysis of PDB data.

GROWTH OF THE PDB ARCHIVE



PDB ARCHIVE CONTENTS ON JANUARY 1, 2015

105,465 released atomic coordinate entries

Molecule Type

97,988 Proteins, peptides, and viruses

2,736 Nucleic acids

4,716 Protein/nucleic acid complexes

25 Other

Experimental Technique

93,707 X-ray Crystallography

10,789 NMR

711 Electron Microscopy

74 Hybrid Methods

184 Other

93,928 Related Experimental Data Files

83,281 Structure factors

8,104 NMR restraints

1,861 Chemical shifts

682 3DEM map files

OVERVIEW: ORGANIZATION

RCSB PDB is jointly managed at Rutgers, The State University of New Jersey and the University of California, San Diego (San Diego Supercomputer Center and the Skaggs School of Pharmacy and Pharmaceutical Sciences).

RCSB PDB advisory board in September 2014

Team members regularly demonstrate their varied expertise in computer science, biology, chemistry, and education. Many members of the staff have been with the RCSB PDB since its inception in 1998. The Senior Staff alone represents over a century of this experience. All members co-author scientific papers, present and participate at meetings, and attend and organize workshops. RCSB PDB members also serve as tutors, teachers, and mentors to students of all ages, and frequently participate in activities that promote science to younger students.

An advisory board of experts in X-ray crystallography, nuclear magnetic resonance (NMR), 3D electron microscopy (3DEM), bioinformatics, and education provides guidance to the project.

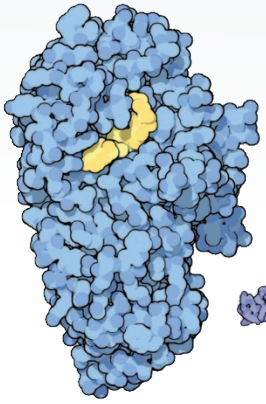
Funding

The RCSB PDB is funded by the National Science Foundation (DBI-1338415), National Institutes of Health, and Department of Energy.

MOLECULAR MACHINES

Cells build many complex molecular machines that perform the biological jobs needed for life.

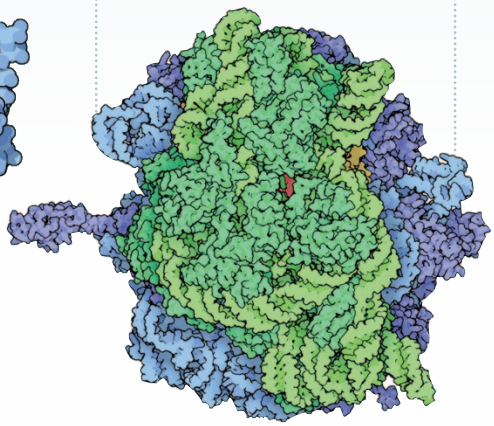
Some of these machines are molecular scissors that cut food into digestible pieces.



Alpha-amylase

Image from February 2006
Molecule of the Month

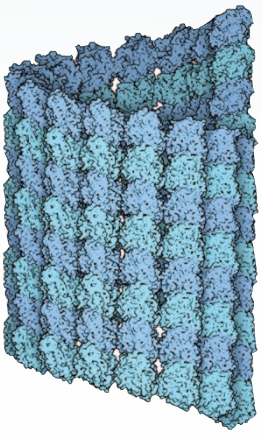
Others then use these pieces to build new molecules when cells grow or tissues need to be repaired.



Ribosome

Image from January 2010
Molecule of the Month

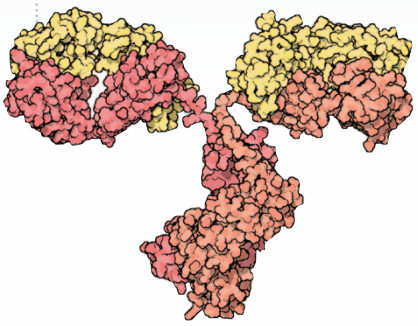
Some molecular machines form sturdy beams that support cells, and others are motors that use energy to crawl along these beams.



Microtubules

Image from June 2014
Molecule of the Month

Some recognize attackers and mobilize defenses against infection.



Antibodies

Image from September 2001
Molecule of the Month

Researchers around the world are studying these molecules at the atomic level. These 3D structures of nucleic acids, proteins, and molecular machines are freely available at the Protein Data Bank (PDB), the central storehouse of biomolecular structures since 1971.

DATA DEPOSITION, ANNOTATION, AND VALIDATION

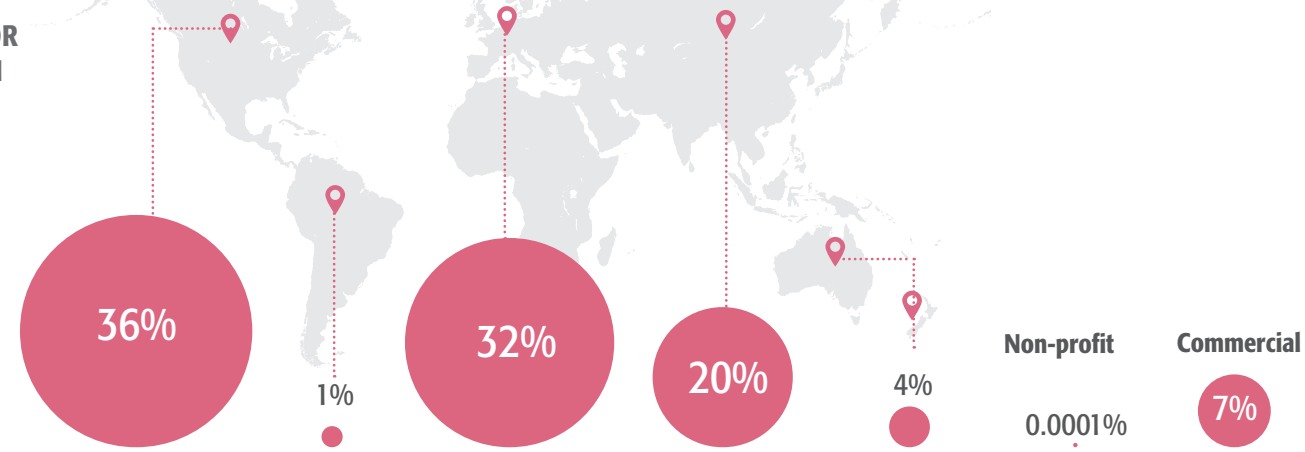
10,368 entries were deposited to the PDB in 2014. wwPDB biocurators carefully review and annotate each structure, and run various validation checks to report back to the depositor in order to finalize the entry for release.

The PDB archive is updated each week with new and modified entries.

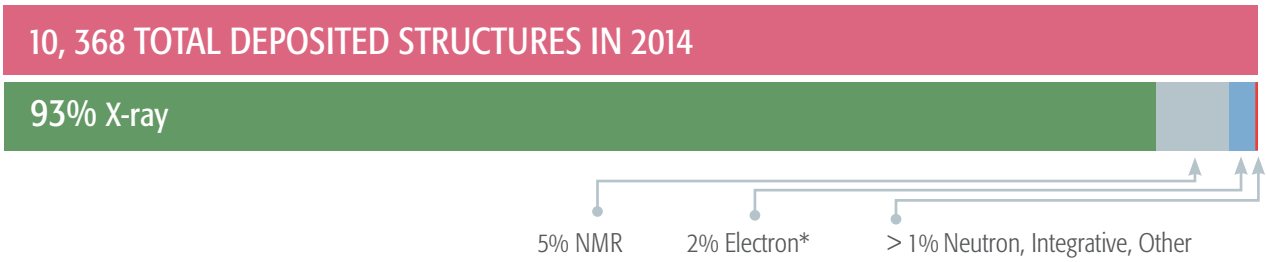
Key to data annotation is the use of the PDBx/macromolecular Crystallographic Information File (mmCIF) data dictionaries in standardizing the macromolecular structures[10,11] and small chemical components[12,13] found in PDB entries.

This careful curation ensures that each entry is an accurate representation of both the structure and experiment. The annotation team reviews polymer sequences, small molecule chemistry, cross-references to other databases, experimental details, correspondence of coordinates with primary data, protein and nucleic acid geometry (*e.g.* Ramachandran plots), biological assemblies, and crystal packing following wwPDB annotation guidelines. This process also ensures that the data are uniformly represented across the archive in order to enable meaningful analysis across the entire archive.

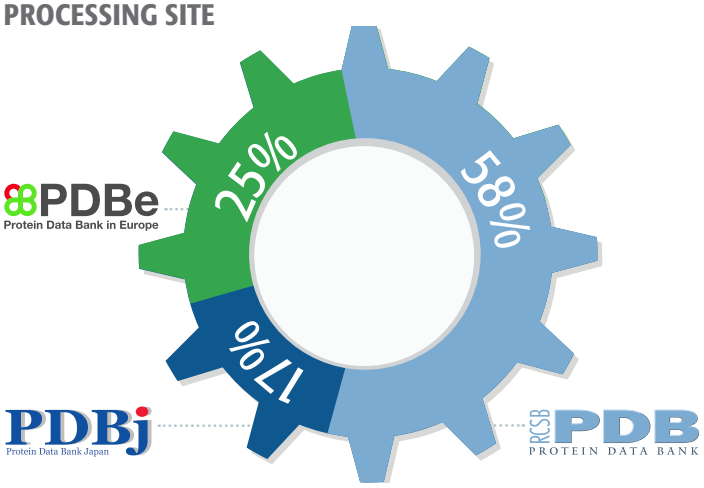
DEPOSITOR LOCATION



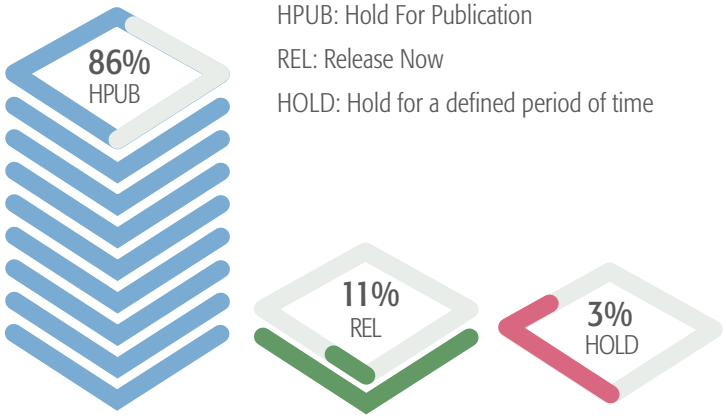
EXPERIMENTAL SOURCE



PROCESSING SITE



AUTHOR-DEFINED RELEASE STATUS



HPUB: Hold For Publication

REL: Release Now

HOLD: Hold for a defined period of time

*Electron category includes Electron Microscopy, Electron Crystallography, and Cryo-electron Microscopy

wwPDB Deposition & Annotation System

A new wwPDB Common Deposition & Annotation (D&A) system has been under development for the last four years. The system was created to handle the anticipated increase in the number, complexity, and size of structures submitted to the PDB archive over the next decade. The system has been developed jointly with our wwPDB partners and provides a uniform deposition and annotation system that will help ensure better data quality and allow for improved load sharing with our partners. The new deposition system for structures determined using X-ray crystallography was released to the public in January 2014. To date, more than 7,000 entries have been submitted using this new system. More than 3,100 of these structures have been released into the archive. Improvements to this system are regularly incorporated based on user feedback. Deposition tools for NMR and 3DEM are being developed by the wwPDB.

As a result of this successful release, RCSB PDB and PDBj are in the process of phasing out the legacy deposition tool ADIT for X-ray crystallographic submissions. ADIT will continue to accept depositions from other experimental methods.

Features of the new system include the use of the PDBx/mmCIF data format, which produces more uniform data; the ability to replace data files pre- and post-deposition; enhanced communication; improved annotation; and geometric and experimental data checking based on recommendations from expert task forces.

Validation Task Forces

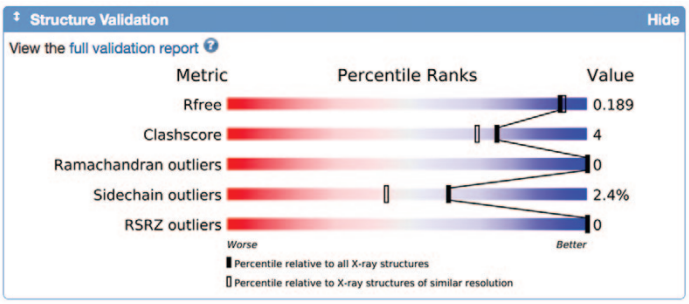
The careful annotation and validation of structural data is also used to improve the quality of the data made available in the archive. To this end, method-specific Validation Task Forces have been convened since 2008 to collect recommendations and develop consensus on additional validation that should be performed, and to identify software applications to perform validation tasks.[14-16] As validation is performed prior to publication and PDB release, the quality of data available to the public is significantly improved.

In October, the wwPDB held a Task Force meeting to explore issues surrounding archiving of structures being studied by hybrid, or "integrative" methods. The meeting was chaired by Andrej Sali (UCSF), Torsten Schwede (University Basel), Jill Trewhella (University Sydney), and attended by attended by a total of 27 experts in integrative modelling, traditional structural biology methods, and archiving. A growing number of integrative structures are being submitted to the PDB, but no mechanisms exist for comprehensive, sustainable archival storage due to the challenges they pose to current practices.[17] A wwPDB proposal for creation of a federated system of interoperable yet separate model and data archives (including the PDB) was unanimously accepted by the Task Force. A white paper from the meeting was published in *Structure*.[18]

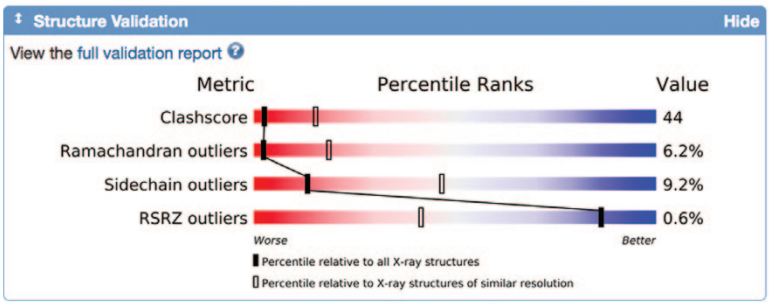
wwPDB Validation Reports

Validation Reports for all X-ray crystal structures in the PDB archive were made available during this report period. These reports implement recommendations from the X-ray Validation Task Force[15] and have been developed in the context of the new wwPDB Deposition and Annotation system.[19] Each validation report contains an assessment of the quality of a structure and highlights specific concerns by considering the coordinates of the model, the experimental data and the fit between the two. An overall quality summary comparing the structure with other crystallographic entries in the archive helps users of PDB data to critically assess archived entries and to select the most appropriate structural models for their needs.

The validation report's quality summary image is displayed on RCSB PDB Structure Summary pages for each entry, along with a link to the full Validation Report PDF and to Ramachandran plots created by MolProbity.[20]



Quality summary image for entry 1CBS, a 1.8Å structure with entry with better overall quality relative to all X-ray structures [G. J. Kleywegt et al. *(1994)* Structure *2: 1241-1258].*



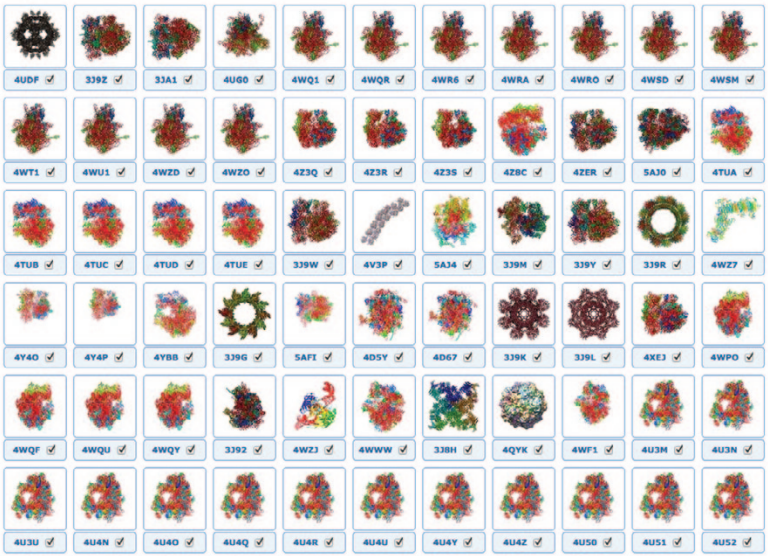
Quality summary image for entry 1FCC, a 3.2Å structure with worse overall quality relative to all X-ray structures [A. E. Sauer-Eriksson et al. *(1995)* Structure *3: 265-278].*

DATA DEPOSITION, ANNOTATION, AND VALIDATION

A major milestone in 2014 was the improvement of support for **LARGE STRUCTURES** in the archive.

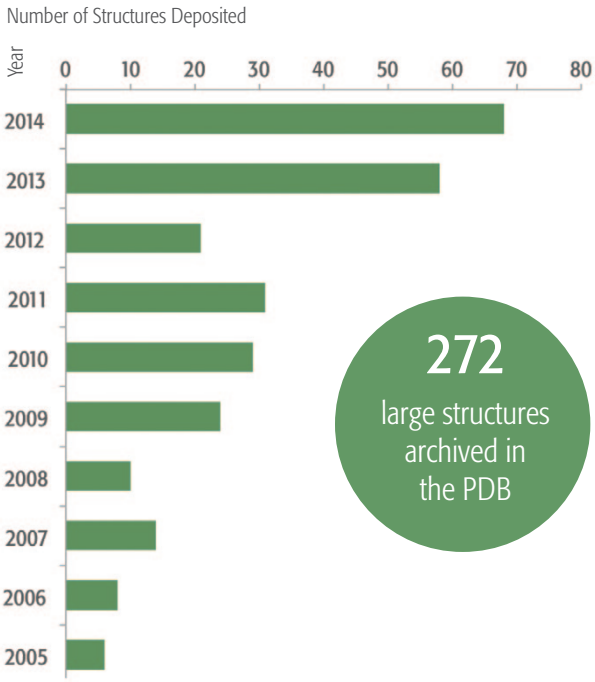
In 2011, the wwPDB established the PDBx/mmCIF Working Group of software developers to enable direct use of PDBx/mmCIF in the major macromolecular crystallographic refinement and analysis tools. PDBx/mmCIF files suitable for deposition can be created with CCP4[21] (REFMAC22 5.8) and Phenix[23] (1.8.2) software packages. The Working Group also made recommendations about essential extensions required for large structures that have been incorporated.

These extensions enabled the release of 272 large structures as single files in PDBx/mmCIF and PDBML formats at the end of 2014. Since then, more than 30 new large structures have been released as single entries in the archive. To support software tools that rely on the PDB file format, a separate directory in the PDB FTP archive contains a TAR file that includes a collection of minimal PDB format files for large structures.



Large structures (containing >62 chains and/or 99999 ATOM lines) have been released as single PDBx/mmCIF files that represent the entire structure. Users can now visualize complete structures using Jmol, Simple Viewer, or Protein Workshop.

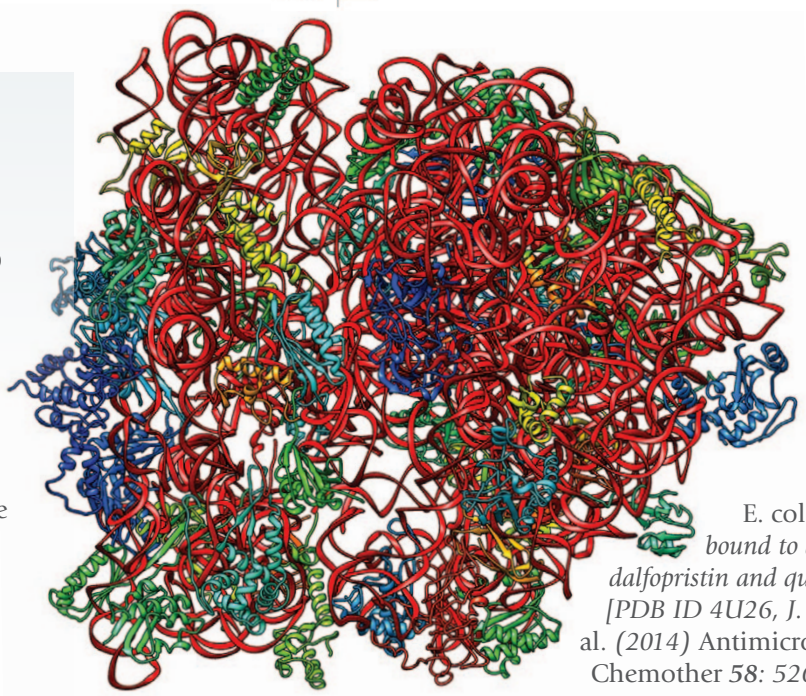
DEPOSITION OF LARGE STRUCTURES



272 large structures archived in the PDB

2014 ANNOTATION PUBLICATIONS

- The Chemical Component Dictionary: complete descriptions of constituent molecules in experimentally determined 3D macromolecules in the Protein Data Bank (2014) *Bioinformatics* doi:**10.1093/bioinformatics/btu789**
- Small molecule annotation for the Protein Data Bank (2014) *Database* doi:**10.1093/database/bau116**
- Improving the representation of peptide-like inhibitor and antibiotic molecules in the Protein Data Bank (2014) *Biopolymers* **101**:659-668 doi:**10.1002/bip.22434**



E. coli *ribosome bound to antibiotics dalfopristin and quinupristin [PDB ID 4U26, J. Noeske* et al. *(2014)* Antimicrob Agents Chemother *58: 5269-5279].*

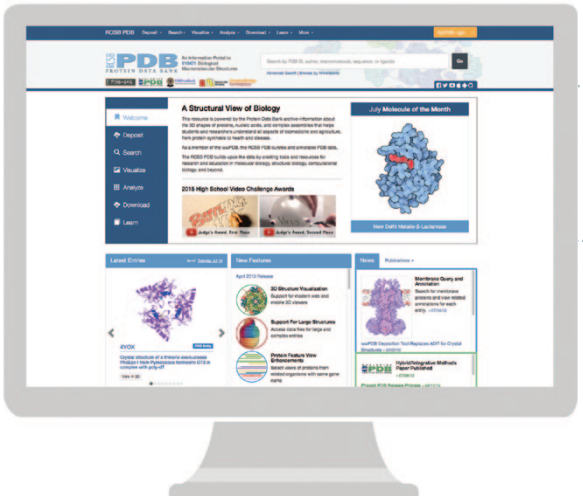
ACCESS AND EXPLORATION

RCSB PDB services and PDB data are freely-available online.

As the archive keeper for the wwPDB, the RCSB PDB maintains the PDB archive at ftp.wwpdb.org. Weekly updates of the PDB archive and the RCSB PDB website are coordinated with the other wwPDB sites. 9,414 new PDB structures were released in 2014, accounting for ~9% of the year total holdings of 105,465.

Data and services are accessed via the website, FTP server (supporting FTP and Rsync access), Web Services, and mobile devices using RCSB PDB *Mobile*. Several software libraries developed for the RCSB PDB website are available as open source on GitHub at github.com/rcsb.

RCSB PDB Mobile *is an application for iOS and Android that provides basic search options, 3D visualization, and access to* Molecule of the Month *articles.*



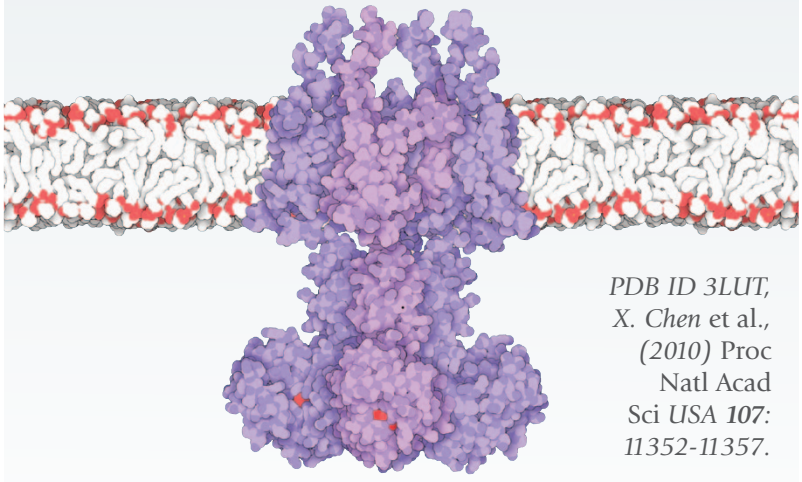
The rcsb.org home page has been redesigned with a clean, streamlined look.

Each month in 2014, rcsb.org was visited 668,000 times on average by 283,000 unique visitors.

A total of 25,033 GB of data were accessed.

Total PDB archive traffic from all wwPDB partners totaled 505,361,092 downloads.

TRANSMEMBRANE PROTEINS



PDB ID 3LUT, X. Chen et al., *(2010)* Proc Natl Acad Sci *USA* ***107****: 11352-11357.*

Transmembrane proteins in the PDB have been identified using the mpstruc database (Stephen White, UC Irvine), sequence clustering, and data derived from UniProt. Users can quickly find all annotated membrane proteins in the PDB by entering "membrane proteins" in the top bar search. The Membrane Protein Browser and the Membrane Proteins drill-down can be used to investigate specific membrane protein classifications and access corresponding structures.

Website Services

The home page has been redesigned with a clean, streamlined look. All information related to high-level topics such as **SEARCH**, **VISUALIZE**, or **ANALYZE** can be browsed from a single panel to offer a 'one stop shopping' experience. New features and enhancements to the website are developed continually.

SEARCH

The website at rcsb.org hosts technologies that support query and reporting of PDB data. Features of the query system include auto-complete search suggestions for text searches; faceted browsing of data based on commonly used characteristics (experimental method, organism, taxonomy, *etc.*); the option to build complex queries using Advanced Search; capability to search for ligand by ID, name, 2D diagram, and more; searching and browsing of drugs and drug targets; and more. Many of these query options are accessible via RESTful Web Services.

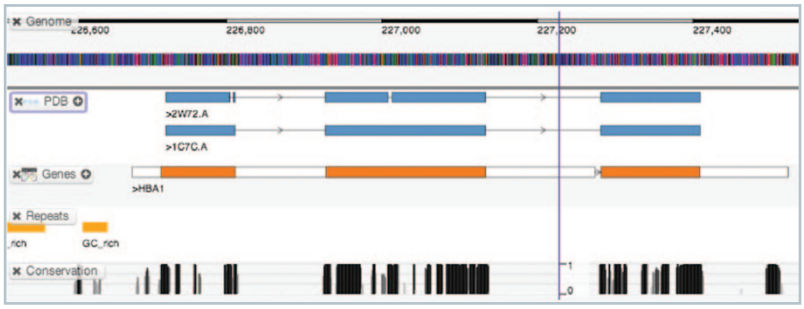
Data in the database are integrated with a variety of external resources, including DrugBank,[24] BindingDB,[25] BindingMOAD,[26] (bindingmoad.org), PDBbind,[27] Transporter Classification Database,[28] MpStruc,[29] RECOORD,[30] PDB-REDO,[31] and SIFTS.[32] These data can be accessed using the search and reporting options.

VISUALIZE

Several molecular viewers, including JSmol/Jmol,[33] Protein Viewer (biasmv.github.io/pv) and the RCSB PDB's Protein Workshop[34] help users interactively view PDB structures. A multi-scale rendering option has been implemented for the efficient display of large structures in Simple Viewer and Protein Workshop.

For a closer look at ligand binding sites visualization, Ligand Explorer has options to turn on the display of interactions including hydrogen bonds, hydrophobic contacts, water mediated hydrogen bonds, and metal interactions. PoseView images, 2D diagrams that highlight a ligand and interacting residues, are also accessible.

RCSB PDB's Protein Feature View visually summarizes how a full-length protein sequence from UniProt[32,35] corresponds to PDB entries. It also loads annotations from external databases (such as Pfam)[36] and homology models from the Protein Model Portal.[37] Annotations display predicted regions of protein disorder, hydrophobic regions, protein function, subunit structure, domains information, and genomic exon structure.



Gene View for gene HBA1: Hemoglobin alpha protein representative chains 2W72.A and 1C7C.A (blue) mapped to eons of HBA1 gene (orange).

A new Gene View illustrates the correspondences between the human genome and 3D structure. All human genes have been mapped to representative PDB protein chains (chosen from sequence clusters at 40% sequence identity) to show which regions of a gene are available in PDB coordinates.

ANALYZE

RCSB PDB's Comparison Tool[38] calculates pairwise sequence (blast2seq,[39] Needleman-Wunsch,[40] and Smith-Waterman[41]) and structure alignments (FATCAT,[42] CE,[43] TM-Align,[44] TopMatch,[45] Dali[46]).

Comparisons can be made for any protein in the PDB archive and for customized or local files not in the PDB. Special features include support for both rigid-body and flexible alignments (*via* jFATCAT) and circular permutations that are difficult for many alignment algorithms to detect (*via* jCE).

To explore Protein Symmetry, the Jmol Symmetry View highlights global, local, and helical symmetry among subunits. The view displays the symmetry axes, a polyhedron that reflects the symmetry, and a color scheme that emphasizes the symmetry.

Mutations in a gene can have profound effects on the function of a protein. The Map Genomic Position to Protein Sequence and 3D Structure analysis tool highlights the location of a gene location (*i.e.*, the site of a SNP).

RCSB PDB's Customer Service

RCSB PDB's Customer Service (info@rcsb.org) collects and answers questions about the website, PDB data, and structural biology. Nearly 1000 unique users initiate electronic conversations with the RCSB PDB each year.

Questions and comments come from students new to structural biology, users involved in the general study of science, and domain experts from the various disciplines that utilize PDB data. Many of these conversations lead to the development of new features at rcsb.org.

2014 DATA QUERY & REPORTING PUBLICATIONS

- Detection of Circular Permutations within Protein Structures using CE-C (2014) *Bioinformatics* **doi:10.1093/bioinformatics/btu823**
- Systematic Detection of Internal Symmetry in Proteins Using CE-Symm (2014) *Journal of Molecular Biology* **426**: 2255-2268 **doi:10.1016/j.jmb.2014.03.010**

EDUCATION AND COMMUNITY OUTREACH

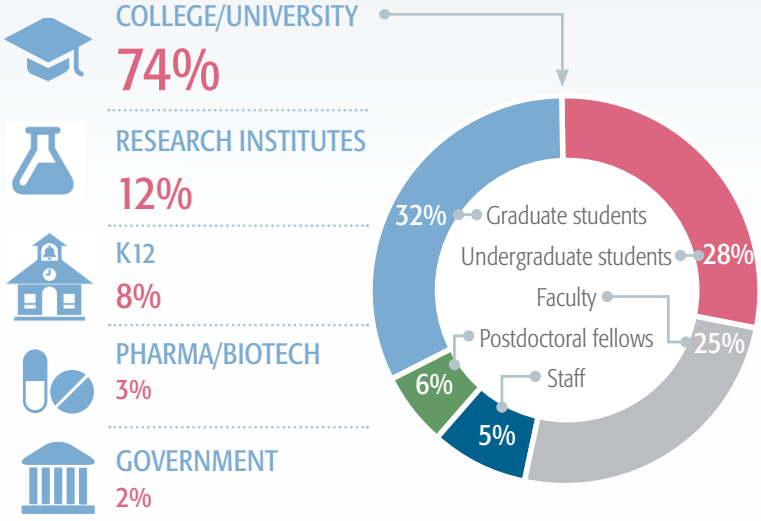
The RCSB PDB supports a variety of users. Our outreach efforts aim to inform different user communities about the RCSB PDB while collecting feedback to help develop a powerful resource for science, medicine, and education.

Users include biologists from a variety of specialties, scientists from other disciplines, students and educators at all levels, authors and illustrators, and the general public.

While the website serves as the primary tool for outreach, staff interact directly with users at local, national, and international meetings, presentations, festivals, and more. Many efforts are targeted towards developing and supporting high school teachers through conferences, workshops, and other programs.

RCSB PDB hosted a very well-attended session at the 2014 National Conference of the Society for Advancement of Hispanics/Chicanos and Native Americans in Science (SACNAS).

PDB-101 USERS*



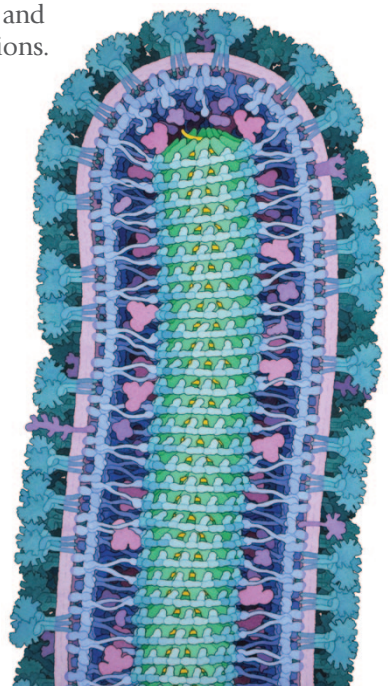
RESEARCH AREAS

*Results of 2014 PDB-101 Survey of 700 respondents

PDB-101

PDB-101 (rcsb.org/pdb-101) is an educational view of the RCSB PDB that packages together resources that promote exploration in the world of proteins and nucleic acids for teachers, students, and the general public. Features include:

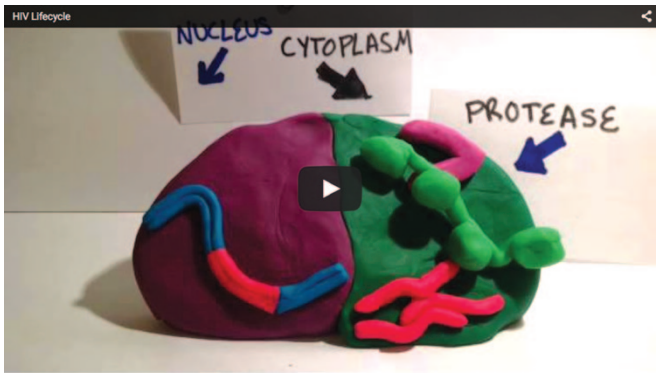
- ***Molecule of the Month.*** Since 2000, the RCSB PDB has published articles that describe the structure and function of a molecule along with interactive views, discussion topics, and links to structure examples. The Ebola Virus Proteins feature was published in October 2014 and has been viewed more than 12,000 times.
- **Structural View of Biology**, a top-down exploration of the PDB that starts with browsable, high-level functional categories that can be explored down to reach *Molecule of the Month* articles and related structural examples.
- **Related Educational Resources** and materials, including posters, animations, and classroom materials. Recent additions include an updated *Molecular Machinery* poster, a paper model of Green Fluorescent Protein, and a video tour of Ebola proteins.
- **Understanding PDB Data**, a reference to help explore and interpret individual PDB entries.
- **HIV/AIDS Curricula** were developed through collaborations and participation of scientists, curriculum design experts, educators, clinicians and local teachers to create authentic, hands-on teaching materials, individual and group activities and assessment suggestions. The curriculum also includes relevant materials and activities related to the topic from existing public resources.



Ebola Virus Proteins as painted by David S. Goodsell for Molecule of the Month *and related animation.*

HIV Video Challenge Awards

RCSB PDB challenged high school students to create short videos that tell a story about HIV/AIDS. A panel of expert judges reviewed the impressive submissions for scientific content, creativity, overall impact, originality, entertainment value, and production quality. An additional award was made for the Viewer's Choice. The winning videos can be seen online at bit.ly/1MfPFbQ.



RCSB PDB Reference in the Top 100

The primary reference for the RCSB PDB was included in *Nature's* 2014 list of the 100 most-cited papers as tracked by the Web of Science

The Protein Data Bank, *Nucleic Acids Research* (2000) **28**: 235-242. **doi: 10.1093/nar/28.1.235**

This article, the first of many publications from the RCSB PDB, announced the new systems put in place for deposition, validation, and exploration of PDB data. The partnerships for data deposition described in the article continue to this day, now formalized with the wwPDB collaboration. Since the time of the article, RCSB PDB systems for data distribution and exploration have grown and expanded to provide access to the more than 100,000 structures that are distributed in the archive and to support our growing and diverse user community.[2,47]

2014 OUTREACH PUBLICATIONS

- The Protein Data Bank archive as an open data resource (2014) *Journal of Computer-Aided Molecular Design* **28**: 1009-1014 **doi:10.1007/s10822-014-9770-y**
- Establishing the Next Generation of the Protein Data Bank (2014) The Winnower **doi:10.15200/winn.140076.68556**

IYCr and the *Art of Science*

The RCSB PDB's *Art of Science* uses the context of an art exhibit to introduce audiences new and old to the beauty of the biomacromolecular structures and the molecular view of biology available in the PDB. In celebration of the International Year of Crystallography (IYCr), two shows were hosted in Paris, France. Following on the success of those events, and thanks to a grant from the American Crystallographic Association, two schools in the US–McPherson College (Kansas) and The Delbarton School (New Jersey)–were awarded permanent editions of the exhibit. Overall, the show was seen by thousands of students in 2014.

Since its beginnings in an art gallery at Rutgers University, the exhibit has been hosted around the world including Texas A&M University, EMBL-Hamburg, Germany; University of Wisconsin-Madison; California State University, Fullerton; Purdue University; and Hyderabad, India. The exhibit primarily toured during 2002-2007, with an iteration in Egypt (2013). The current exhibit includes images from the original *Art of Science* show along with new works inspired by important *Molecule of the Month* topics and IYCr.

Art of Science *at McPherson College (top), Paris (bottom left), and Delbarton School (bottom-right).*

RELATED RESOURCES

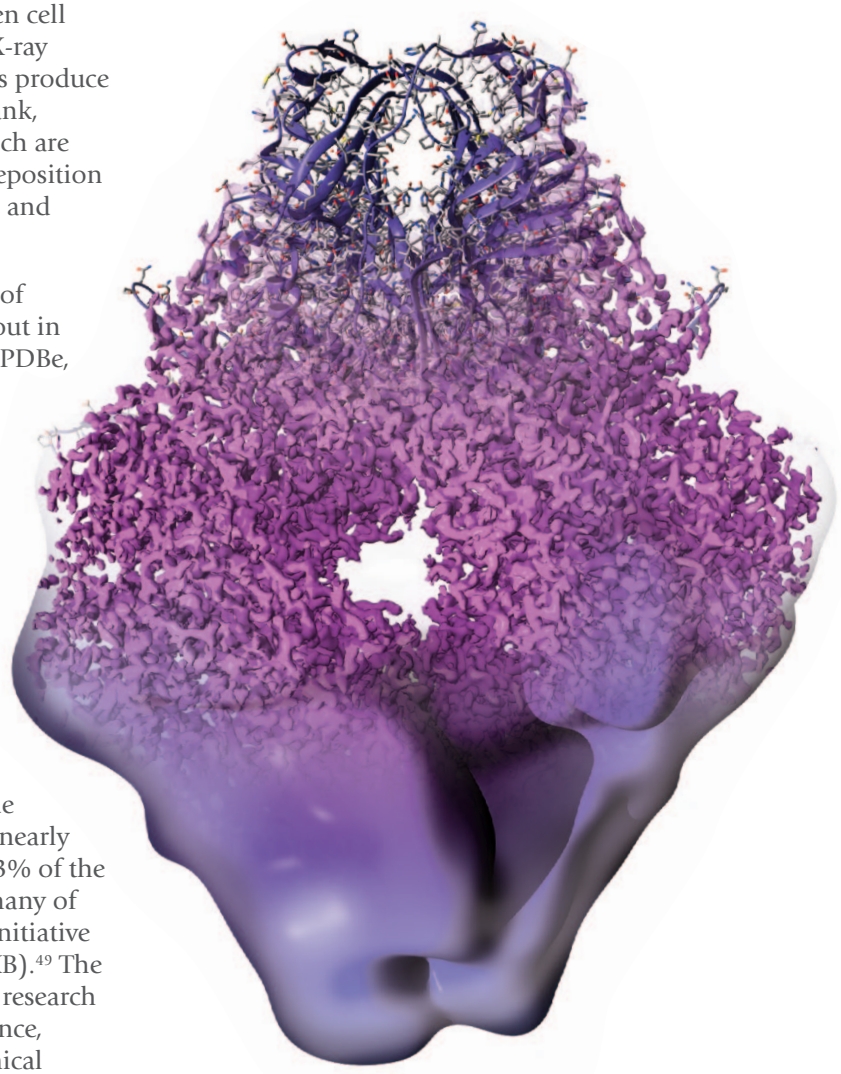
EMDataBank: Unified Data Resource for 3-Dimensional Electron Microscopy

The PDB archives large biological assemblies determined by 3DEM, a maturing methodology in structural biology that bridges the gap between cell biology and the experimental techniques of X-ray crystallography and NMR. 3DEM experiments produce 3D density maps, archived in the EM Data Bank, and often yield fitted coordinate models, which are archived in the PDB. **EMDataBank.org** is a deposition and retrieval network for 3DEM map, model, and associated metadata.[48]

The EMDB map archive is distributed as part of the PDB archive. This work has been carried out in collaboration with the RCSB PDB at Rutgers, PDBe, and the National Center for Macromolecular Imaging at Baylor College of Medicine.

High-throughput Structural Studies and the PSI Structural Biology Knowledgebase

Using high-throughput techniques, worldwide structural genomics centers have determined nearly 13,500 structures to date, which represents 13% of the PDB archive. RCSB PDB works closely with many of these centers and with the Protein Structure Initiative (PSI) Structural Biology Knowledgebase (SBKB).[49] The SBKB (**sbkb.org**) was established to facilitate research design by integrating protein structure, sequence, and functional annotations, along with technical information regarding protein production and structure determination. Researchers can search the SBKB by sequence, PDB ID or UniProt accession code, and receive a current list of matching PDB entries, pre-built theoretical models from the Protein Model Portal,[37] annotations from 100+ open biological resources, structural genomics target histories and protocols from TargetTrack, and ready-to-use DNA clones from DNASU.[50] It is also possible to find structures according to functional relevance (KB-Rank tool)[51] or find related technologies and publications from the PSI Technology[52] and Publications Portals, respectively. Interactive tools such as real-time theoretical modeling and biophysical parameter prediction also enhance understanding of proteins that are not yet well characterized. Seven years worth of research and technical highlights, written by the Nature Publishing Group in partnership with the SBKB, have been collected into a new searchable online E-collection, categorized by biological system and/or lab technique.



Recently, PDB and EMDB released the highest-resolution structure yet determined using 3D cryoelectron microscopy (3DEM). At 2.2 Å resolution, this structure of beta-galactosidase [PDB ID 5A1A, EMDB ID EMD-2984; A. Bartesaghi et al. *(2015)* Science ***348****: 1147-1151] resolves an active-site inhibitor and approximately 800 water molecules, demonstrating the power of the new EM detectors and processing software that are bringing 3DEM into the same resolution regime as X-ray crystallography. The bottom portion of the image shows the same structure captured at 3.2 Å resolution [A. Bartesaghi* et al. *(2014)* Proc Natl Acad Sci USA ***111****: 11709-11714].*

REFERENCES

1. H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne. (2000) The Protein Data Bank. *Nucleic Acids Res* **28**: 235-242.
2. P. W. Rose, A. Prlic, C. Bi, W. F. Bluhm, C. H. Christie, S. Dutta, R. K. Green, D. S. Goodsell, J. D. Westbrook, J. Woo, J. Young, C. Zardecki, H. M. Berman, P. E. Bourne, S. K. Burley. (2015) The RCSB Protein Data Bank: views of structural biology for basic and applied research and education. *Nucleic Acids Res* **43**: D345-356.
3. Protein Data Bank. (1971) Protein Data Bank. *Nature New Biol.* **233**: 223.
4. H. M. Berman, G. J. Kleywegt, H. Nakamura, J. L. Markley. (2014) The Protein Data Bank archive as an open data resource. *J Comput Aided Mol Des* **28**: 1009-1014.
5. H. M. Berman, G. J. Kleywegt, H. Nakamura, J. L. Markley. (2013) How community has shaped the Protein Data Bank. *Structure* **21**: 1485-1491.
6. H. M. Berman, K. Henrick, H. Nakamura. (2003) Announcing the worldwide Protein Data Bank. *Nat Struct Biol* **10**: 980.
7. E. L. Ulrich, H. Akutsu, J. F. Doreleijers, Y. Harano, Y. E. Ioannidis, J. Lin, M. Livny, S. Mading, D. Maziuk, Z. Miller, E. Nakatani, C. F. Schulte, D. E. Tolmie, R. Kent Wenger, H. Yao, J. L. Markley. (2008) BioMagResBank. *Nucleic Acids Res* **36**: D402-408.
8. A. Gutmanas, Y. Alhroub, G. M. Battle, J. M. Berrisford, E. Bochet, M. J. Conroy, J. M. Dana, M. A. Fernandez Montecelo, G. van Ginkel, S. P. Gore, P. Haslam, R. Hatherley, P. M. Hendrickx, M. Hirshberg, I. Lagerstedt, S. Mir, A. Mukhopadhyay, T. J. Oldfield, A. Patwardhan, L. Rinaldi, G. Sahni, E. Sanz-Garcia, S. Sen, R. A. Slowley, S. Velankar, M. E. Wainwright, G. J. Kleywegt. (2014) PDBe: Protein Data Bank in Europe. *Nucleic Acids Res* **42**: D285-291.
9. A. R. Kinjo, H. Suzuki, R. Yamashita, Y. Ikegawa, T. Kudou, R. Igarashi, Y. Kengaku, H. Cho, D. M. Standley, A. Nakagawa, H. Nakamura. (2012) Protein Data Bank Japan (PDBj): maintaining a structural data archive and resource description framework format. *Nucleic Acids Res* **40**: D453-460.
10. P. M. D. Fitzgerald, J. D. Westbrook, P. E. Bourne, B. McMahon, K. D. Watenpaugh, H. M. Berman. (2005). 4.5 Macromolecular dictionary (mmCIF). In *International Tables for Crystallography G. Definition and exchange of crystallographic data* (S. R. Hall & B. McMahon, eds.), pp. 295-443. Springer, Dordrecht, The Netherlands.
11. J. Westbrook, K. Henrick, E. L. Ulrich, H. M. Berman. (2005). 3.6.2 The Protein Data Bank exchange data dictionary. In *International Tables for Crystallography* (S. R. Hall & B. McMahon, eds.), Vol. G. Definition and exchange of crystallographic data, pp. 195-198. Springer, Dordrecht, The Netherlands.
12. J. D. Westbrook, C. Shao, Z. Feng, M. Zhuravleva, S. Velankar, J. Young. (2014) The chemical component dictionary: complete descriptions of constituent molecules in experimentally determined 3D macromolecules in the Protein Data Bank. *Bioinformatics* **31**: 1274-1278.
13. S. Dutta, D. Dimitropoulos, Z. Feng, I. Persikova, S. Sen, C. Shao, J. Westbrook, J. Young, M. A. Zhuravleva, G. J. Kleywegt, H. M. Berman. (2014) Improving the representation of peptide-like inhibitor and antibiotic molecules in the Protein Data Bank. *Biopolymers* **101**: 659-668.
14. G. T. Montelione, M. Nilges, A. Bax, P. Guntert, T. Herrmann, J. S. Richardson, C. D. Schwieters, W. F. Vranken, G. W. Vuister, D. S. Wishart, H. M. Berman, G. J. Kleywegt, J. L. Markley. (2013) Recommendations of the wwPDB NMR Validation Task Force. *Structure* **21**: 1563-1570.
15. R. J. Read, P. D. Adams, W. B. Arendall, 3rd, A. T. Brunger, P. Emsley, R. P. Joosten, G. J. Kleywegt, E. B. Krissinel, T. Lutteke, Z. Otwinowski, A. Perrakis, J. S. Richardson, W. H. Sheffler, J. L. Smith, I. J. Tickle, G. Vriend, P. H. Zwart. (2011) A new generation of crystallographic validation tools for the protein data bank. *Structure* **19**: 1395-1412.
16. J. Trewhella, W. A. Hendrickson, G. J. Kleywegt, A. Sali, M. Sato, T. Schwede, D. I. Svergun, J. A. Tainer, J. Westbrook, H. M. Berman. (2013) Report of the wwPDB Small-Angle Scattering Task Force: data requirements for biomolecular modeling and the PDB. *Structure* **21**: 875-881.
17. E. Callaway. (2014) Data bank struggles as protein imaging ups its game. *Nature* **514**: 416.
18. A. Sali, T. Schwede, J. Trewhella, H. M. Berman, G. Kleywegt, S. K. Burley, J. Markley, H. Nakamura, P. D. Adams, A. Bonvin, W. Chiu, M. Dal Peraro, F. Di Maio, T. E. Ferrin, K. Grunewald, A. Gutmanas, R. Henderson, G. Hummer, K. Iwasaki, G. Johnson, C. L. Lawson, J. Meiler, M. A. Marti-Renom, G. T. Montelione, M. Nilges, R. Nussinov, A. Patwardhan, J. Rappsilber, R. J. Read, H. Saibil, G. F. Schroder, C. Schwieters, C. A. M. Seidel, D. Svergun, M. Topf, E. Ulrich, S. Velankar, J. Westbrook. (2015) Outcome of the First wwPDB Hybrid/Integrative Methods Task Force Workshop. *Structure* **23***: 1156-1167*.
19. S. Gore, S. Velankar, G. J. Kleywegt. (2012) Implementing an X-ray validation pipeline for the Protein Data Bank. *Acta Crystallogr D Biol Crystallogr* **68**: 478-483.
20. V. B. Chen, W. B. Arendall, 3rd, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson, D. C. Richardson. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**: 12-21.
21. CCP4. (1994) The Collaborative Computational Project Number 4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* **50:** 760-763.
22. R. A. Nicholls, F. Long, G. N. Murshudov. (2012) Low-resolution refinement tools in REFMAC5. *Acta Crystallogr D Biol Crystallogr* **68**: 404-417.
23. P. D. Adams, R. W. Grosse-Kunstleve, L.-W. Hung, T. R. Ioerger, A. j. McCoy, N. W. Moriarty, R. J. Read, J. C. Sacchettini, N. K. Sauter, T. C. Terwilliger. (2002) PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallographica* **D58**: 1948-1954.
24. V. Law, C. Knox, Y. Djoumbou, T. Jewison, A. C. Guo, Y. Liu, A. Maciejewski, D. Arndt, M. Wilson, V. Neveu, A. Tang, G. Gabriel, C. Ly, S. Adamjee, Z. T. Dame, B. Han, Y. Zhou, D. S. Wishart. (2014) DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Res* **42**: D1091-1097.
25. A. M. Wassermann, J. Bajorath. (2011) BindingDB and ChEMBL: online compound databases for drug discovery. *Expert Opin Drug Discov* **6**: 683-687.
26. L. Hu, M. L. Benson, R. D. Smith, M. G. Lerner, H. A. Carlson. (2005) Binding MOAD (Mother Of All Databases). *Proteins* **60**: 333-340.
27. R. Wang, X. Fang, Y. Lu, C. Y. Yang, S. Wang. (2005) The PDBbind database: methodologies and updates. *J Med Chem* **48**: 4111-4119.
28. M. H. Saier, Jr., V. S. Reddy, D. G. Tamang, A. Vastermark. (2014) The transporter classification database. *Nucleic Acids Res* **42**: D251-258.
29. S. H. White, C. Snider. (2014). Membrane Proteins of Known 3D Structure (MPStruc), Vol. 2014.
30. A. J. Nederveen, J. F. Doreleijers, W. Vranken, Z. Miller, C. A. Spronk, S. B. Nabuurs, P. Guntert, M. Livny, J. L. Markley, M. Nilges, E. L. Ulrich, R. Kaptein, A. M. Bonvin. (2005) RECOORD: a recalculated coordinate database of 500+ proteins from the PDB using restraints from the BioMagResBank. *Proteins* **59**: 662-672.
31. R. P. Joosten, F. Long, G. N. Murshudov, A. Perrakis. (2014) The PDB_REDO server for macromolecular structure model optimization. *IUCrJ* **1**: 213-220.
32. S. Velankar, J. M. Dana, J. Jacobsen, G. van Ginkel, P. J. Gane, J. Luo, T. J. Oldfield, C. O'Donovan, M. J. Martin, G. J. Kleywegt. (2013) SIFTS: Structure Integration with Function, Taxonomy and Sequences resource. *Nucleic Acids Res* **41**: D483-489.
33. Jmol: an open-source Java viewer for chemical structures in 3D. **http://www.jmol.org/**.
34. J. L. Moreland, A. Gramada, O. V. Buzko, Q. Zhang, P. E. Bourne. (2005) The Molecular Biology Toolkit (MBT): a modular platform for developing molecular visualization applications. BMC *Bioinformatics* **6**: 21.
35. The UniProt Consortium. (2014) Activities at the Universal Protein Resource (UniProt). *Nucleic Acids Res* **42**: D191-198.
36. R. D. Finn, A. Bateman, J. Clements, P. Coggill, R. Y. Eberhardt, S. R. Eddy, A. Heger, K. Hetherington, L. Holm, J. Mistry, E. L. Sonnhammer, J. Tate, M. Punta. (2014) Pfam: the protein families database. *Nucleic Acids Res* **42**: D222-230.
37. J. Haas, S. Roth, K. Arnold, F. Kiefer, T. Schmidt, L. Bordoli, T. Schwede. (2013) The Protein Model Portal--a comprehensive resource for protein structure and model information. *Database (Oxford)* 2013: bat031.
38. A. Prlic, S. Bliven, P. W. Rose, W. F. Bluhm, C. Bizon, A. Godzik, P. E. Bourne. (2010) Pre-calculated protein structure alignments at the RCSB PDB website. *Bioinformatics* **26**: 2983-2985.
39. T. A. Tatusova, T. L. Madden. (1999) BLAST 2 Sequences, a new tool for comparing protein and nucleotide sequences. *FEMS Microbiol Lett* **174**: 247-250.
40. S. B. Needleman, C. D. Wunsch. (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol* **48**: 443-453.
41. T. F. Smith, M. S. Waterman. (1981) Identification of common molecular subsequences. *J Mol Biol* **147**: 195-197.
42. Y. Ye, A. Godzik. (2003) Flexible structure alignment by chaining aligned fragment pairs allowing twists. *Bioinformatics* **19**: ii246-255.
43. I. N. Shindyalov, P. E. Bourne. (1998) Protein structure alignment by incremental combinatory extension of the optimum path. *Protein Eng.* **11**: 739-747.
44. Y. Zhang, J. Skolnick. (2005) TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic acids research* **33**: 2302-2309.
45. M. J. Sippl, M. Wiederstein. (2012) Detection of spatial correlations in protein structures and molecular complexes. *Structure* **20**: 718-728.
46. L. Holm, J. Park. (2000) DaliLite workbench for protein structure comparison. *Bioinformatics* **16**: 566-567.
47. P. W. Rose, C. Bi, W. F. Bluhm, C. H. Christie, D. Dimitropoulos, S. Dutta, R. K. Green, D. S. Goodsell, A. Prlic, M. Quesada, G. B. Quinn, A. G. Ramos, J. D. Westbrook, J. Young, C. Zardecki, H. M. Berman, P. E. Bourne. (2013) The RCSB Protein Data Bank: new resources for research and education. *Nucleic Acids Res* **41**: D475-482.
48. C. L. Lawson, M. L. Baker, C. Best, C. Bi, M. Dougherty, P. Feng, G. van Ginkel, B. Devkota, I. Lagerstedt, S. J. Ludtke, R. H. Newman, T. J. Oldfield, I. Rees, G. Sahni, R. Sala, S. Velankar, J. Warren, J. D. Westbrook, K. Henrick, G. J. Kleywegt, H. M. Berman, W. Chiu. (2011) EMDataBank.org: unified data resource for CryoEM. *Nucleic Acids Research* **39**: D456-464.
49. M. J. Gabanyi, P. D. Adams, K. Arnold, L. Bordoli, L. G. Carter, J. Flippen-Andersen, L. Gifford, J. Haas, A. Kouranov, W. A. McLaughlin, D. I. Micallef, W. Minor, R. Shah, T. Schwede, Y. P. Tao, J. D. Westbrook, M. Zimmerman, H. M. Berman. (2011) The Structural Biology Knowledgebase: a portal to protein structures, sequences, functions, and methods. *J Struct Funct Genomics* **12**: 45-54.
50. C. Y. Seiler, J. G. Park, A. Sharma, P. Hunter, P. Surapaneni, C. Sedillo, J. Field, R. Algar, A. Price, J. Steel, A. Throop, M. Fiacco, J. LaBaer. (2014) DNASU plasmid and PSI:Biology-Materials repositories: resources to accelerate biological research. *Nucleic Acids Res* **42**: D1253-1260.
51. E. S. Julfayev, R. J. McLaughlin, Y. P. Tao, W. A. McLaughlin. (2012) KB-Rank: efficient protein structure and functional annotation identification via text query. *Journal of structural and functional genomics* **13**: 101-110.
52. L. K. Gifford, L. G. Carter, M. J. Gabanyi, H. M. Berman, P. D. Adams. (2012) The Protein Structure Initiative Structural Biology Knowledgebase Technology Portal: a structural biology web resource. *Journal of structural and functional genomics* **13**: 57-62.

CREDITS

RCSB PDB Partners

Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB)

RCSB.ORG • INFO@RCSB.ORG

Jointly managed by:

Rutgers, The State University of New Jersey
Center for Integrative Proteomics Research
174 Frelinghuysen Road
Piscataway, NJ 08854-8087

University of California, San Diego
San Diego Supercomputer Center and the Skaggs School of Pharmacy and Pharmaceutical Sciences
9500 Gilman Drive
La Jolla, CA 92093-0743

Worldwie Protein Data Bank (wwPDB)

RCSB PDB is a member of wwPDB
wwpdb.org

Other wwPDB Members:

Protein Data Bank Europe
pdbe.org

Protein Data Bank Japan
pdbj.org

BioMagResBank
www.bmrb.wisc.edu

Find Us Online

facebook.com/RCSBPDB

twitter.com/buildmodels

youtube.com/RCSBProteinDataBank

github.com/rcsb

RCSB PDB Leaders & Leadership Team

DR. STEPHEN K. BURLEY
Director

DR. HELEN M. BERMAN
Associate Director
External Partnerships

CHRISTINE ZARDECKI
Deputy Director

COLE H. CHRISTIE
Development and Operations Manager

DR. SHUCHISMITA DUTTA
Scientific Educational Development Lead

DR. ZUKANG FENG
Principal Scientific Application/Web Developer

DR. DAVID S. GOODSELL
Scientific Outreach Lead

DR. RACHEL KRAMER GREEN
Scientific Support & Customer Service Lead

DR. IRINA PERSIKOVA
Principal Biocurator

DR. ANDREAS PRLIĆ
Technical & Scientific Team Lead

DR. PETER W. ROSE
UCSD Site Head

DR. JOHN WESTBROOK
Data & Software Architect Lead

DR. JASMINE YOUNG
Biocuration Team Lead

rcsb.org • info@rcsb.org