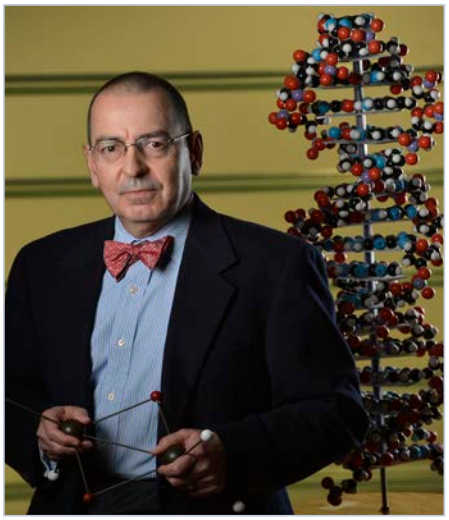
2024 ANNUAL REPORT

RCSB PDB
PROTEIN DATA BANK

RCSB.ORG A LIVING DIGITAL DATA RESOURCE THAT ENABLES SCIENTIFIC BREAKTHROUGHS

DIRECTOR'S MESSAGE



In 2024, the RCSB Protein Data Bank celebrated its 25th year serving many millions of users with state-of-the-art data analysis, visualization, and download tools for exploring information about proteins and nucleic acids.

Through our web portal at **RCSB.org**, researchers and students utilize the rigorously-validated, expertly-biocurated, three-dimensional (3D) biostructure information stored in the Protein Data Bank (PDB) archive alongside Computed Structure Models (CSMs) made possible by modern protein prediction methods.

Since 2000, RCSB PDB has served as the US data center for the global PDB archive of 3D structure data for large biomolecules (proteins, DNA, and RNA) essential for research and education in fundamental biology, health, energy, and biotechnology.

The PDB is managed by an international collaboration among regionally-funded data centers, called the Worldwide Protein Data Bank (wwPDB). RCSB PDB is the global archive keeper and the US data center in this collaboration. PDB data are made available worldwide at no charge and with no limitations on usage.

The PDB archive was established in 1971 at Brookhaven National Laboratory with just seven protein crystal structures. The U.S. National Science Foundation (NSF) was an early supporter of the archive, providing crucial support to the growing resource that has continued for more than five decades.

Today, RCSB PDB Core Operations are funded by the U.S. National Science Foundation (NSF), the US Department of Energy (DOE), and the National Institutes of Health (NIH) under 5-year grants that run through 2028.

RCSB PDB safeguards structural biology data with an estimated replacement cost that exceeds US$23 billion.

RCSB PDB also supports US research by empowering a heterogeneous STEM workforce through Training activities, advancing the frontiers of basic research and discovery through strategic initiatives (*e.g.*, incorporating CSMs at RCSB.org), and enabling applied research in academe and industry with data and tools for exploring new technologies (*e.g.*, protein engineering, de novo protein design, molecular nanotechnology). Training, Outreach, and Education materials are available on our dedicated PDB-101 website ("101", as in an entry-level course, **PDB101.RCSB.org**).

All feedback on RCSB PDB activities is welcomed at RCSB.org and at *info@rcsb.org*.

Sincerely,

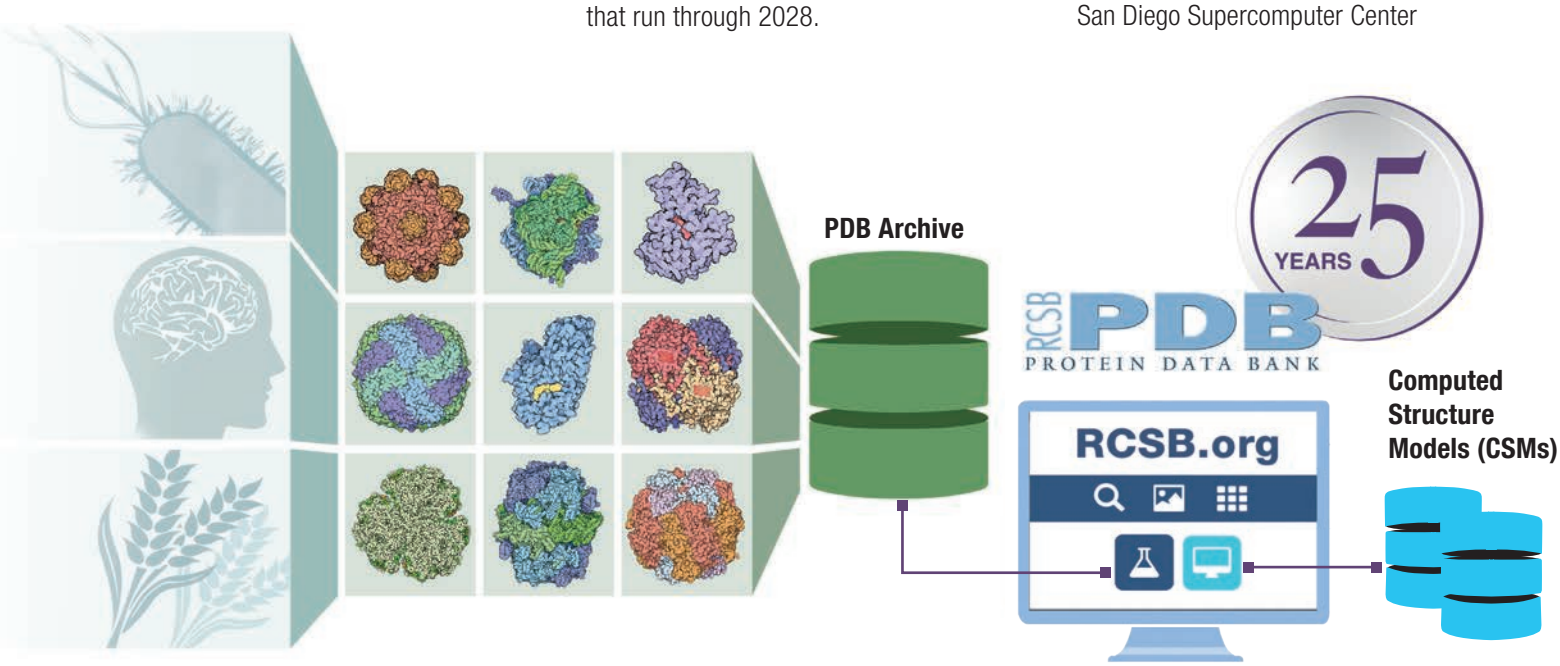
Stephen K. Burley, M.D., D.Phil.

Director, RCSB Protein Data Bank

University Professor and
Henry Rutgers Chair
Rutgers, The State University of New Jersey

Interim Director
Rutgers Artificial Intelligence and Data Science (RAD) Collaboratory

Research Scientist
San Diego Supercomputer Center



PDB IMPACT

The PDB archive is central to basic and applied research and education across the biosciences, and provides validated information and analysis tools to the nation's researchers, policy makers, federal, state, and local health authorities, and elected representatives.

PDB data have been contributed by ~60,000 structural biologists (Depositors) working on all inhabited continents using experimental techniques such as macromolecular crystallography, 3D electron microscopy, and nuclear magnetic resonance to study the shapes of important biological macromolecules.

Knowing the 3D structure of a biological macromolecule is essential for understanding its role in human and animal health and disease, its function in plants and food and energy production, and its importance to other topics related to global prosperity and the bioeconomy.

PDB data are used by many millions of PDB Data Consumers (basic and applied researchers, trainees, educators, and students) working or learning in nearly every sovereign country and territory recognized by the United Nations. RCSB PDB tools for analysis and visualization are key for driving research and understanding of fundamental biology, biomedicine, energy sciences, and biotechnology.

2024 NOBEL PRIZE IN CHEMISTRY

Researchers have long used the wealth of experimentally-determined structures available from the open-access PDB archive to understand basic principles of protein architecture. More recently, PDB structures were used as training data for Artificial Intelligence/Machine Learning software tools (*e.g.*, AlphaFold, RoseTTAfold, OpenFold) that can predict three-dimensional structures of proteins with accuracies comparable to lower-resolution experimental methods and drive design of proteins with novel shapes and biochemical functions.

The 2024 Nobel Prize in Chemistry was awarded to David Baker, who was recognized for computational protein design, and to Demis Hassabis and John M. Jumper, who were recognized for protein structure prediction. These achievements also celebrate the work of the PDB depositor community and the PDB archive that underpinned their pioneering achievements.

PDB structures were used as training data for Artificial Intelligence/Machine Learning software tools, which led to the groundbreaking progress in the computational protein design recognized by the 2024 Nobel Prize in Chemistry. RCSB.org provides access to these Computed Structure Models (as symbolized with the computer screen in this image) alongside experimental PDB structures (symbolized with flasks).

BROAD IMPACT ACROSS RESEARCH AND EDUCATION

The inaugural RCSB PDB article (Berman *et al.*, *Nucleic Acids Research* 2000) remains one of the top-cited scientific publications of all time with more than 30,000 citations. These publications report research in subject areas running the gamut from Agriculture to Zoology.

PDB data are re-used by hundreds of external biodata resources. An internal review of the 2025 *Nucleic Acids Research* database collection identified 502 scientific databases that repackage and redistribute PDB information. Copies of PDB archival data are also stored inside for-profit biopharmaceutical and biotechnology company firewalls for use in proprietary research.

IMPACT OF PDB STRUCTURES ON US FDA DRUG APPROVALS

Open access to PDB data facilitates discovery and development of life saving drugs, including

- 100% of the 34 new small-molecule, protein-targeted anti-cancer agents approved by the US FDA from 2019 to 2023[1]
- >90% of the 79 new anti-neoplastic agents approved by US FDA from 2010 to 2018[2]
- 171 new drugs approved by the US Food and Drug Administration (FDA) from 2010 to 2016[3]

PDB IMPACT CONT.

PDB DATA AND RCSB PDB SERVICES SUPPORT FEDERALLY FUNDED US RESEARCHERS AND THE US BIOECONOMY

RCSB PDB safeguards structural biology data with an estimated replacement cost exceeding US$23 billion. Nearly all of this information was generated by structural biologists working in the US, Europe, and Asia, who were locally supported with public monies.

Additionally, RCSB PDB services and PDB data support federally-funded researcher innovation

- Provides key research infrastructure that drives innovation by researchers in the US
- PDB structures have contributed data to >1 million published research papers
- Connects federally-funded research and scientists with worldwide structural biology data from public and private sector research
- RCSB PDB provides facile access to >1 million Computed Structure Models of proteins
- Ensures rigor and reproducibility across research in biomedicine, fundamental biology, and bioenergy
- Training events and online tools promote workforce development in a competitive environment

RCSB PDB SAFEGUARDS STRUCTURAL BIOLOGY DATA SUPPORTING THE OPPORTUNITIES AND GRAND CHALLENGES OF THE NSF

RCSB PDB safeguards structural biology data generated with more than half a billion dollars of NSF funding over the lifetime of the PDB

- RCSB.org tools drive research across all NSF directorates, illuminating biology from Agriculture to Zoology
- Structural biology provides foundational knowledge into the rules that make life work
- Cryo-electron microscopy (cryoEM) resolution revolution and predicted protein structures are advancing the frontiers of basic and applied research
- RCSB PDB accelerates new and data-intensive research through advanced cyberinfrastructure

SUPPORTING THE NIH TURN DISCOVERY INTO HEALTH

RCSB PDB safeguards 3D biostructure data generated using NIH research funding, NIH-funded synchrotron beamlines, and CryoEM facilities supported by the NIH Common Fund, representing $5.4 Billion worth of NIH data over the lifetime of the PDB

- 2nd most heavily used online data resource after ClinicalTrials.gov for NIH-funded researchers[4]
- Links all relevant PDB structures to NIH Common Fund Resources
- Enables structure-guided discovery of new small-molecule and biologics drugs, diagnostic tools, and technologies

SUPPORTING THE RESEARCH GOALS OF DOE

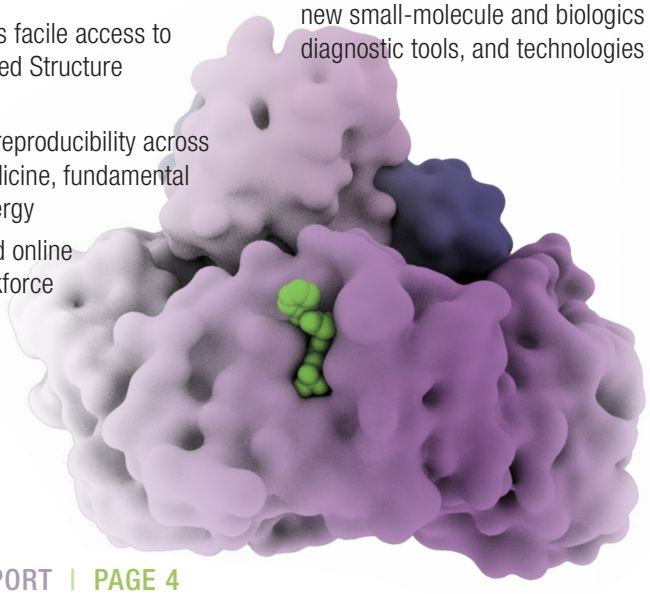
RCSB PDB safeguards structural breakthroughs (Synchrotron, Neutron, and Cryo-EM) generated with DOE funding:

- >$5.7 Billion worth of DOE data over the lifetime of the PDB
- RCSB PDB supports the fundamental research and cryoEM facilities funded by the DOE Biological and Environmental Research program. RCSB PDB also supports scientific user facilities funded by DOE Basic Energy Sciences, including Advanced Photon Source, Advanced Light Source, and National Synchrotron Light Source II; Linac Coherent Light Source-II; and neutron diffraction instruments at Oak Ridge National Laboratory
- DOE Synchrotrons have produced >62,000 PDB structures over their lifetime
- RCSB PDB hosts BioSync (**biosync.rcsb.org**), an online Guide to High Energy Data Collection Facilities that provides up-to-date information on over 130 X-ray beamlines at facilities worldwide
- Integration of RCSB PDB and KBase Resources supports basic and applied research in plant molecular biology and microbial physiology

GROWING INDUSTRY AND ECONOMIC DEVELOPMENT

RCSB PDB services and PDB data heavily impact pharmaceutical and biotechnology companies by enabling:

- Basic and applied research
- Patent applications
- Discovery of lifesaving drugs
- Innovations that can lead to new product development and company formation



*The US FDA approved anticancer drug Selinexor shown here in green bound to its target protein Exportin 1 from entry **PDB_00007L5E**, is an example how the synergy of academic research, PDB data, and private enterprise can lead to the discovery of life saving treatments and promote economic development.[1]*

3D DATA AND RCSB PDB SERVICES

RCSB PDB offers a suite of services to support researchers and educators in accessing, analyzing, and visualizing 3D structure data of biological macromolecules.

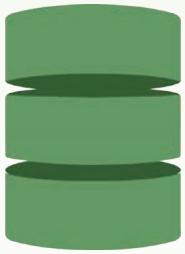
DECEMBER 31, 2024

PDB ARCHIVE CONTENTS

229,564 Released atomic coordinate entries

By Experimental Method

190,619	X-ray
14,431	NMR
24,155	Electron Microscopy
238	Multi Method
84	Neutron Diffraction
37	Other

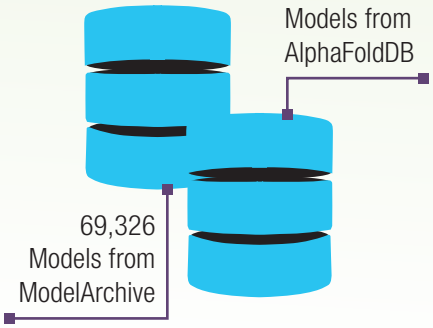


By Molecule Type

198,096	Proteins, peptides, and viruses
4,560	Nucleic acids
13,967	Protein/nucleic acid complexes
12,706	Protein/Oligosaccharide
22	Oligosaccharide (only)
213	Other

CSMs ACCESSIBLE ON RCSB.ORG

1,068,577 Models



SERVICE 0

IT INFRASTRUCTURE

IT Infrastructure supports all RCSB PDB Services and systems by establishing policies and processes to ensure standardized systems configurations and management, redundancy, security, high availability, and disaster recovery.

From data deposition to user training, all RCSB PDB activities are supported by a robust commitment to modern IT Infrastructure.

2024 HIGHLIGHT

All public-facing servers had an uptime of >99.99% in order to provide RCSB PDB services 24 hours a day, 366 days in 2024.

SERVICE 1

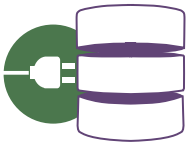
DEPOSITION AND BIOCURATION

RCSB PDB and other members of the Worldwide PDB support tens of thousands of individual data depositors around the globe, ensuring completeness and high quality for the ever-growing body of experimental structure information. **More on page 6.**

SERVICE 2

ARCHIVE MANAGEMENT AND ACCESS

RCSB PDB maintains the PDB archive according to FAIR principles, provides access to the data, and integrates the structural information with other scientific resources. **More on page 8.**



SERVICE 3

DATA EXPLORATION

RCSB PDB develops tools for searching, visualization, and analysis of PDB structures and computed structure models (CSMs). These tools are freely available on **RCSB.org**. **More on page 9.**

SERVICE 4

TRAINING, OUTREACH, AND EDUCATION

RCSB PDB creates resources to support the broad user community in research, education, and training, made freely available on **PDB101.RCSB.org**. **More on page 10.**

SERVICE I

DEPOSITION AND BIOCURATION

Supporting Data Depositors who freely contribute the results of their structural studies of biological macromolecules to the PDB.

All data deposited to PDB undergo expert review. Each structure is examined for self-consistency, standardized using controlled vocabularies, cross-referenced with other biological data resources, and validated for scientific/technical accuracy.

The Worldwide Protein Data Bank (wwPDB) was established in 2003 to manage the single PDB archive of macromolecular structural data and make it freely and publicly available to the global community. It consists of organizations that act as deposition, data processing, and distribution centers for PDB data.

PDB structures contain

- 3D atomic coordinates
- Experimental data
- Mandatory metadata
- Authors (*e.g.*, PI contact information)
- Primary citation
- Sample preparation, data collection, and structure determination details
- Polymer sequence(s) (proteins, DNA, RNA, oligosaccharides)
- Chemical information

wwPDB develops and uses the OneDep system for data validation, deposition, and biocuration. This single system helps to maintain high data quality and completeness within the PDB archive, while supporting growth in the number, size and complexity of the deposited structures.

The OneDep software system is regularly updated to facilitate data submission. In 2024, upgrades improved the file upload process for macromolecular crystallography structure depositions, cryoEM cross-file checks, and the annotation workflow for processing large structures.

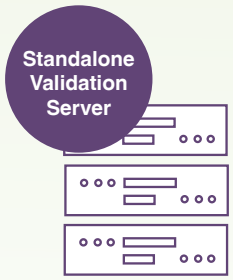
Other software tools are used to prepare data for PDB deposition (pdb_extract, used for ~3800 structures submitted to RCSB PDB in 2024); and for automated depositions of large numbers of structures (GroupDep; used for 1369 structures from 38 research groups in 2024).

Validation is central to ensuring the highest quality data. wwPDB Working Groups and Task Forces represent more than 100 academic and industrial volunteers who make recommendations and contribute

wwPDB VALIDATION

The PDB Validation Process is central to ensuring the highest quality of data and entails multiple steps.

1.Pre-deposition



The wwPDB provides a standalone validation server for Depositors to evaluate their PDBx/mmCIF file anonymously as they prepare for deposition. This tool minimizes the number of revisions and reduces the processing time.

validate.wwpdb.org

2.Deposition

Preliminary Validation Report generated

File revisions/reupload inside the OneDep system

Preliminary report approved by the Depositor

PDB ID issued

deposit.wwpdb.org

3. Biocuration

wwPDB biocurators provide further error checking

Final Validation Report generated

Final report approved by the Depositor

This report can be submitted to journals for quality assesment during peer review.

4. Structure Release

Final Validation Report regenrated and made public.

The public validation reports enable meaningful analysis and comparison across the entire archive.

software used to generate wwPDB Validation Reports that assess the quality and accuracy of every structure stored in the PDB archive.

Before deposition, structural biologists can anonymously assess the quality and accuracy of their structures using an online validation server (60,000 sessions in 2024) and API (used >15,000 times in 2024).

During deposition, a "preliminary" version of the wwPDB validation report identifies issues with the atomic coordinates and/or data upload. The OneDep system supports revisions and upload of replacement files to finalize the submission. Once the preliminary wwPDB validation report is approved by the Depositor, a PDB ID is provided, and the submission is passed to a wwPDB biocurator for annotation and further consistency/error checking.

Following expert biocuration, the official wwPDB validation report is generated for final review and approval. Submission of this report is required by several scientific journals to help peer reviewers assess structure quality and accuracy to ensure the integrity of the scientific literature.

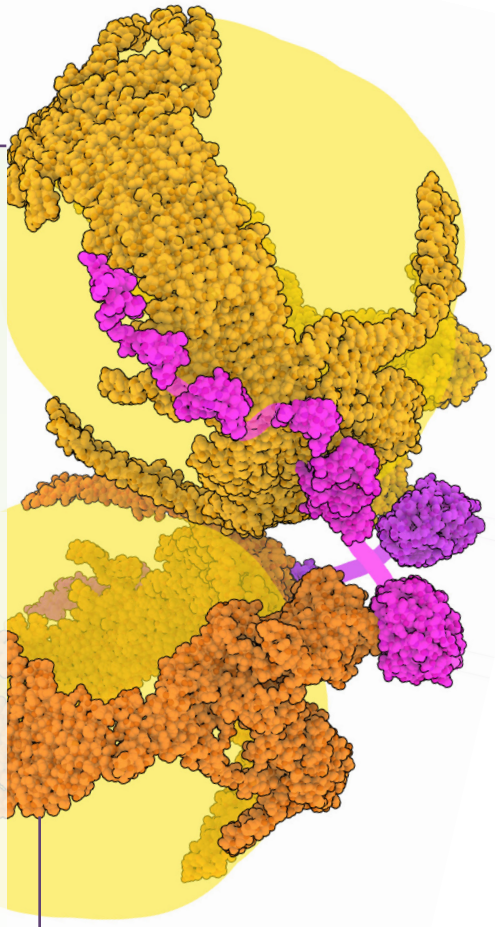
Upon release, the wwPDB validation reports are re-generated and made public at the time of the structure data release. These public validation reports enable meaningful analysis and comparison across the entire archive.

2024 HIGHLIGHT

19,322 structures were deposited, validated, and biocurated by the wwPDB (up from 17,063 in 2023).

3,679 new ligands and 23 new Biologically Interesting Molecule Reference dictionary items were created.

RCSB PDB is responsible for managing depositions from the Americas and Oceania. Of all US structures released in 2024, ~73% reported NIH funding; ~5% reported NSF funding. ~77% of X-ray crystal structures released in 2024 from US labs utilized DOE-supported synchrotron or XFEL facilities for data collection.



The PDB archive reflects the trends and interest in scientific research. Among groundbreaking structures deposited in 2024 is the Apolipoprotein B-100. The protein offers insights into how LDL (the "bad cholesterol") is removed from the bloodstream creating a foundation for new medical treatments. Learn more from April 2025 Molecule of the Month (doi.org/10.2210/rcsb_pdb/mom_2025_4).

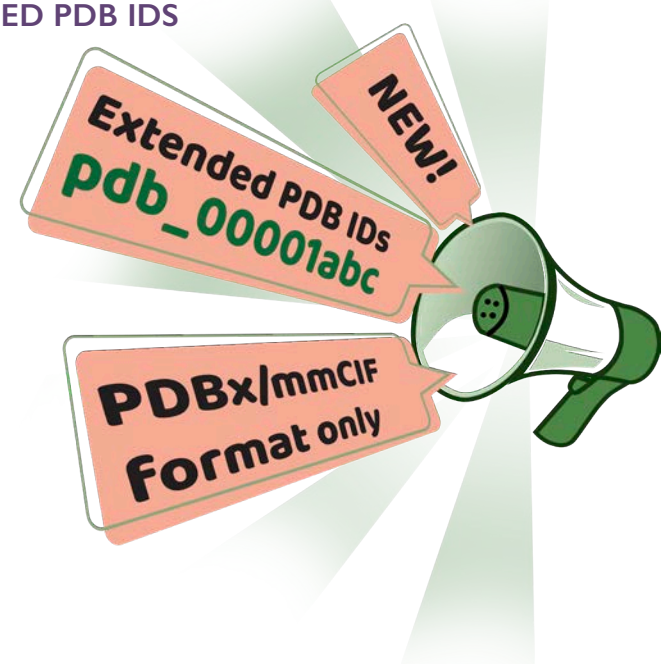
TRANSITIONING TO PDBx/mmCIF AND EXTENDED PDB IDS

When the PDB was established in 1971, PDB IDs were set to contain 4-alphanumberic characters (*e.g.*, ***1ABC***).

With the continuous growth of PDB archive, wwPDB anticipates that all possible four character PDB IDs will be consumed in 2028.

PDB IDs are being extended and prepended with "PDB" (*e.g.*, ***1ABC*** will become ***PDB_00001ABC***). Structures assigned these new extended PDB IDs will not be compatible with the legacy PDB file format.

Tools and documentation are available to help users transition to the PDBx/mmCIF and Extended PDB IDs NOW–visit **wwPDB.org** for more.



SERVICE 2

ARCHIVE MANAGEMENT AND ACCESS

Supporting PDB Data Consumers by maintaining the PDB archive; data dictionary development and standardization, enabling global data delivery and DOI registration, and integrating PDB data with external information.

In its role as wwPDB archive keeper, the RCSB PDB is responsible for safeguarding and maintaining the PDB archive (**files.wwpdb.org**). RCSB PDB coordinates weekly updates of the PDB archive with wwPDB Data Centers in Europe and Asia.

PDB archival format and data standards are defined by the PDBx/mmCIF dictionary (**mmCIF.wwpdb.org**).

At present, PDBx/mmCIF is jointly maintained by RCSB PDB, our wwPDB partners, and the wwPDB PDBx/mmCIF Working Group. Data dictionary terms and definitions are formulated, reviewed, and modified to support remediation of existing data and inclusion of new and rapidly evolving methodologies, from new experimental techniques to Computed Structure Models.

To support RCSB.org resources, large-scale calculations are run weekly to generate clusters of similar sequences and 3D structures to support search and analysis applications. Data are also integrated with 50 trusted external data resources from across the Life Sciences ecosystem to provide the most complete and up-to-date representation of a PDB structure.

PDB data are also delivered through >500 external resources that freely repackage and redistribute PDB information, and through copies of the archival data stored inside for-profit company firewalls for use alongside proprietary 3D biostructure information.

wwPDB continues to standardize and update PDB data. In 2024, PDB structure updates included the addition of extended PDB ID/DOIs; generation of PDBx/mmCIF-formatted validation reports; and inclusion of remediated IHM structures. Additional updates included improved annotation of post-translation modification features (~77,000 entries); improved and re-organized data about starting models used in structure determination (~31,000); source of macromolecules studied (~400); authors information including email, country and PI-designation (~250); standardized software name (~100). Ligand chemical components were updated with added post-translation modification features at the component level (~6,400 CCD entries).

2024 HIGHLIGHT

Throughout 2024, a record 15,314 new PDB structures were released. On January 1, 2025, the PDB archive hosted 229,564 structures.

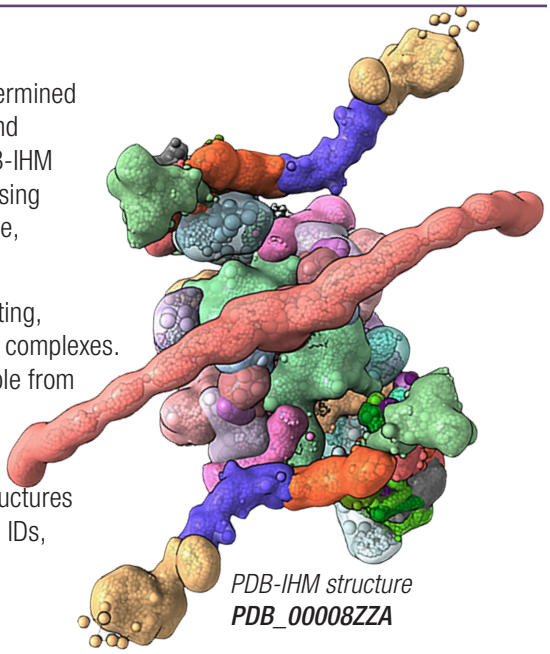
~3 billion data files in various file formats, including structure files, experimental data files, chemical and molecular reference data files, FASTA sequence files, and validation reports, were downloaded and/or viewed from RCSB PDB-hosted FTP and websites. Additional data were downloaded from wwPDB partners PDBe and PDBj, for more than 3.6 billion data file downloads.

PDB-IHM

Structures of a growing number of large macromolecular assemblies are now being determined using integrative approaches, wherein information derived from multiple experimental and computational methods is combined to compute their three-dimensional structures. PDB-IHM (formerly PDB-Dev) is a system for archiving and disseminating structures determined using integrative or hybrid methods (IHM), and making them Findable, Accessible, Interoperable, and Reusable (FAIR).

The PDB-IHM ecosystem provides a robust infrastructure for depositing, curating, validating, visualizing, archiving, and disseminating integrative structures of biomolecules and their complexes. Integrative structures can be deposited through the PDB-IHM deposition portal (accessible from the wwPDB OneDep home page) and processed in parallel to the OneDep system.

Structures processed by PDB-IHM are released synchronously with PDB structures weekly on Wednesdays at 00:00 UTC. IHM structures are now served alongside PDB structures at the PDB archive (**files.wwpdb.org/pub/pdb_ihm**). All IHM structures are issued PDB IDs, and have corresponding PDB DOIs and wwPDB DOI landing webpages (*e.g.*, https://doi.org/10.2210/pdb8zz1/pdb).



PDB-IHM structure ***PDB_00008ZZA***

SERVICE 3
DATA EXPLORATION

Supporting PDB Data Consumers in the US and around the world through our open-access, research-focused web portal RCSB.org that provides tools for finding, visualizing, and analyzing structures.

RCSB PDB services are accessed via our website (**RCSB.org**) and public API endpoints for programmatic access, with two main APIs being for Data (**data.rcsb.org**) and Search (**search.rcsb.org**). Supporting API services include the 1D Coordinate Server API (**1d-coordinates.rcsb.org**) that serves alignments between structural and sequence databases and integrates protein positional features from multiple resources and Alignment API (**alignment.rcsb.org**) that serves as a comprehensive platform for the seamless computation of structure alignments.

The website supports a broad range of skill levels and interests. In addition to finding and retrieving 3D structure data, PDB users access comparative data, and external annotations from resources such as Comprehensive Antibiotic Resistance Database (CARD) and Pharos, a comprehensive knowledge base for the Druggable Genome. Gene Ontology, InterPro and Pharos annotations were also made available for the computed structure models (CSMs) in addition to experimental structures. In 2024, we introduced integration with M-CSA (Mechanism and Catalytic Site Atlas[5]), providing users with detailed insights into enzyme mechanisms and catalytic sites. Additionally, we enriched our Browse Annotations feature by incorporating Disease Ontology, leveraging the Mondo Disease Ontology[6] to improve the exploration of disease associations through PDB sequences.

RCSB PDB services go well beyond the original structure and scientific publication. Each PDB structure is represented by a Structure Summary page that organizes access to important information, including a snapshot of the validation report and other high-level content, annotations, sequence information, sequence similarity clusters, and experimental data. These data are updated weekly, which means that while the corresponding scientific publication remains static, RCSB PDB delivers contemporary views of all structures. Structure Summary pages are continuously enhanced by incorporating additional information and improving visualization options. New features include information about the starting model used for the experimental structure, display of serial crystallography annotations, and enhanced visualization options for glycoproteins.

Our research-focused RCSB.org web portal offers a rich collection of software tools and features that can be used to search, browse, analyze, and visualize PDB data. These include powerful search services as well as interactive analytical and visualization tools such as the Mol* molecular graphics system, sequence annotations view, and specialized tools that provide redundancy-reduced "groups" view of similar data. The Pairwise Structure Alignment Tool can align one or more protein chains to a reference structure in a pairwise manner, for simultaneous analysis and visualization of 3D structure alignments and structure-based 1D sequence alignments. User help guides are available to support all users.

3D structure similarity and motif searches can be executed based on 3D coordinates uploaded by the user. This allows the identification of similarities between uploaded coordinates to experimental structures and CSMs available from RCSB.org.

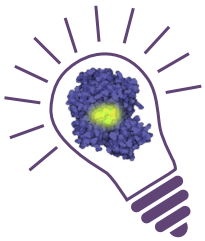
2024 HIGHLIGHT

RCSB PDB is one of the most heavily used biological data resources worldwide. For 2024, internal analytics tracked ~14.5 million unique IP addresses, making ~5.0 billion requests/interactions (for data, web pages, *etc.*). During the same period, Google Analytics estimated that RCSB.org hosted ~8.7 million unique users (~13.5% from USA) viewing ~64 million web pages.

CREATING TOOLS FOR RESEARCHERS AND TRAINING A NEW GENERATION OF SCIENTISTS

Students are trained by team members on software development projects that become part of RCSB PDB projects in production.

Most recently, students worked with team members to develop a new rcsb-api Python package which enables researchers to easily interact with RCSB.org Search and Data API services. Using this toolkit, users can seamlessly integrate search and data retrieval tasks into computational workflows for custom analyses.[7]



SERVICE 4

TRAINING, OUTREACH, AND EDUCATION

Building and supporting the broad PDB user community with a wide range of resources for understanding 3D biostructures.

PDB-101 (**pdb101.rcsb.org**) is an online portal for exploring the world of proteins and nucleic acids. The diverse shapes and functions of biological macromolecules help explain all aspects of biomedicine and agriculture, from protein synthesis to health and disease to biological energy.

- Training materials, such as the *Guide to Understanding PDB Data* and webinars are available to help graduate students, postdoctoral scholars, and researchers use PDB data and RCSB PDB tools.
- Outreach content, including the *Molecule of the Month* series and molecular animations, demonstrate how PDB data can be used to understand fundamental biology, biomedicine, bioengineering/biotechnology, and energy sciences in 3D by a diverse and multidisciplinary user community.
- Education Materials, such as Curriculum Modules, provide lessons and activities for teaching and learning.

In 2024, the RCSB PDB hosted a variety of virtual training events: *Visualize Biomolecular Structures with Mol*: From Atoms to Movies*; *A Deep Dive into Computed Structure Model Exploration at RCSB.org*; *Understanding PDB Validation: Which experimental structures should I rely on?*; *Python Scripting for Molecular Docking*; *Teaching Enzymology with the Protein Data Bank: From Pandemic to Paxlovid*; *Unlock Rapid Analyses Across the Whole PDB Using BinaryCIF*. Virtual training events are advertised to participants online and through a targeted electronic newsletter. Event recordings and materials are posted on PDB-101 at **pdb101.rcsb.org/train/training-events**.

Additional virtual office hours were held to support users in topics including deposition, RCSB.org, and emerging changes in RCSB PDB and wwPDB offerings.

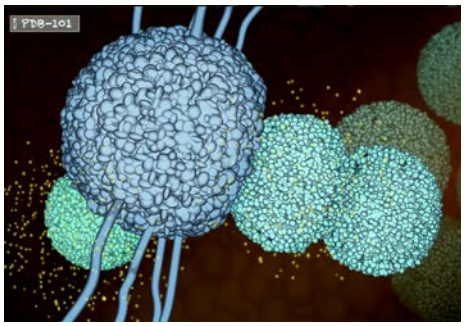
Other PDB-101 content released this year included a 2025 calendar highlighting the structural biology of Peak Performance: *Structural Biology of Vitamins*; new features *Exploring the Structural Biology of Computed Structure Models* and *Exploring the Structural Biology of Evolution*; and a three-part video series exploring the human immune response to cancer focusing on cellular and molecular details of the process. The *Guide to Understanding PDB Data* was reorganized to better highlight PDBx/mmCIF-related content.

'Immunology and Cancer' is a three-part video series released in 2024 exploring how immunotherapies work to amplify the immune response to cancer

2024 HIGHLIGHT

In 2024, 473,000 PDB-101 users accounted for >1.7 million page views (Google Analytics).

PDB-101 molecular animation videos, webinar recordings, and tutorials are hosted on YouTube. The channel (RCSBProteinDataBank) has ~88,000 subscribers; videos were viewed ~750,000 times in 2024.



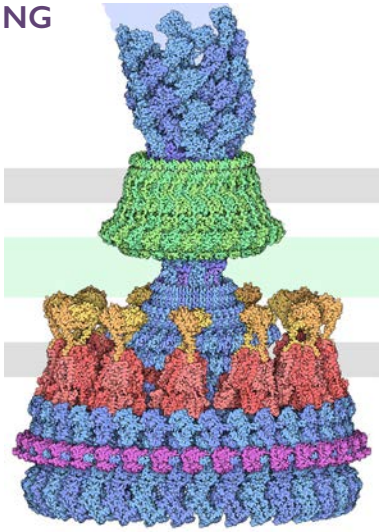
MOLECULE OF THE MONTH: CELEBRATING 25 YEARS OF STORYTELLING AND ANNOUNCING NEW BEGINNINGS

After publishing a remarkable 300 columns, David Goodsell has retired from the *Molecule of the Month*. Janet Iwasa will continue the series, beginning with the January 2025 feature.

David has been a prodigious contributor to PDB-101 over the past 25 years. In addition to establishing the *Molecule of the Month*, he shared his incredible Molecular Landscapes and created a collection of articles that help foster an understanding of PDB Data.

RCSB PDB is grateful for his years of mentoring and collaboration, and wishes him a most enjoyable retirement.

RCSB PDB is excited to continue the *Molecule of the Month* journey with Janet, and for the amazing molecular stories to come.



December 2024 Molecule of the Month on Flagellar Motor (doi.org/10.2210/rcsb_pdb/mom_2024_12) marks the last feature written and illustrated by David S. Goodsell. This illustration shows the clockwise form of the flagellar motor.

SERVICE 4

TRAINING, OUTREACH, AND EDUCATION CONT.

SCIENTIFIC SUPPORT AND USER ENGAGEMENT

RCSB PDB regularly participates in scientific meetings and professional societies, hosts Working Groups, and convenes method-specific Task Forces. Recognized experts in fields, including but not limited to, structural biology, cell and molecular biology, computational biology, information technology, and education serve as advisors to the RCSB PDB.

Depositors and PDB Data Consumers are supported by responsive Help Desks covering all RCSB PDB and wwPDB services. User feedback helps inform prioritization and resource development to meet the needs of diverse research and education communities.

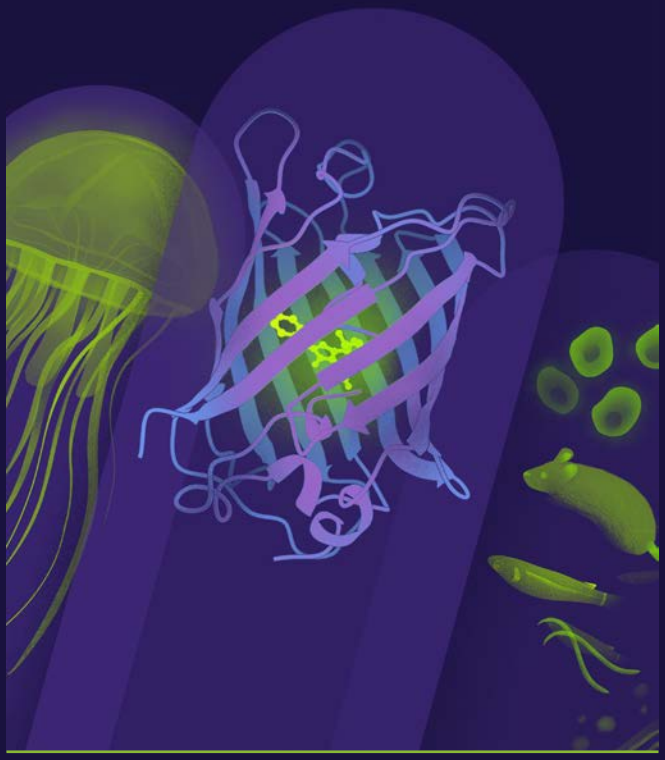
The RCSB PDB Team in November 2024. All members participate in training, outreach and education efforts by participating in scientific meetings, conferences, and professional societies as well as hosting workshops and training events and developing training resources for PDB-101.

REFERENCES

1. S.K. Burley *et al.* (2024) Impact of structural biology and the Protein Data Bank on US FDA new drug approvals of low molecular weight antineoplastic agents 2019–2023 *Oncogene* **43**: 2229-2243 doi: 10.1038/s41388-024-03077-2
2. J.D. Westbrook *et al.* (2020) Impact of the Protein Data Bank on antineoplastic approvals *Drug Discovery Today* **25**: 837-850 doi: 10.1016/j.drudis.2020.02.002
3. J.D. Westbrook and S.K Burley (2019) How Structural Biologists and the Protein Data Bank Contributed to Recent FDA New Drug Approvals *Structure* **27**: 211-217 doi: 10.1016/j.str.2018.11.007
4. K.B. Read *et al.* (2015) Sizing the Problem of Improving Discovery and Access to NIH-Funded Data: A Preliminary Study. *PLoS One* **10**: e0132735 doi: 10.1371/journal.pone.0132735
5. A.J.M. Ribeiro *at al.* (2018) Mechanism and Catalytic Site Atlas (M-CSA): a database of enzyme reaction mechanisms and active sites. *Nucleic Acids Res* **46**: D618-D623 (2018) doi: 10.1093/nar/gkx1012
6. N.A. Vasilevsky *et al.* (2022) Mondo: Unifying diseases for the world, by the world. *medRxiv* 2022.04.13.22273750; doi: 10.1101/2022.04.13.22273750
7. D.W. Piehl *et al.* (2025) *rcsb-api*: Python Toolkit for Streamlining Access to RCSB Protein Data Bank APIs. *J Mol Biol.* doi: 10.1016/j.jmb.2025.168970. Epub ahead of print

CITE RCSB PDB

1. The Protein Data Bank (2000) *Nucleic Acids Research* **28**: 235-242. doi: 10.1093/nar/28.1.235
2. Updated resources for exploring experimentally-determined PDB structures and Computed Structure Models at the RCSB Protein Data Bank (2025) *Nucleic Acids Research* **53**: D564–D574; doi: 10.1093/nar/gkae1091
3. RCSB Protein Data Bank: Tools for visualizing and understanding biological macromolecules in 3D (2022) *Protein Science* **31**: e4482 doi: 10.1002/pro.4482



ABOUT THE COVER

Illustration by Xinyi Christine Zhang

This cover highlights the tiny Green Fluorescent Protein (GFP) from jellyfish that has revolutionized cell biology and biotechnology. GFP is amazingly useful in scientific research, because it allows us to look directly into the inner workings of cells.

GFP was selected in honor of the National Science Foundation, which is celebrating its 75th anniversary in 2025.

FUNDING

RCSB PDB core operations are funded by U.S. National Science Foundation (DBI-2321666), the US Department of Energy (DE-SC0019749), and the National Cancer Institute, National Institute of Allergy and Infectious Diseases, and National Institute of General Medical Sciences of the National Institutes of Health under grant R01GM157729.

RCSB PDB
PROTEIN DATA BANK

RCSB.ORG • INFO@RCSB.ORG

RCSB PDB is managed by the members of the Research Collaboratory for Structural Bioinformatics: Rutgers, UCSD/SDSC, and UCSF

RUTGERS–NEW BRUNSWICK
Artificial Intelligence and Data Science Collaboratory

UC San Diego SDSC

UCSF

WORLDWIDE PDB PROTEIN DATA BANK

RCSB PDB is a member of the wwPDB organization | **wwPDB.org**

CORE TRUST SEAL

PDB is a CoreTrustSeal Board certified Trusted Digital Repository.

GLOBAL CORE BIODATA RESOURCE

PDB is a Global Core Biodata Resource whose long term funding and sustainability is of fundamental importance to biological and life science communities.

FOLLOW US

RCSBPDB

RCSBProteinDataBank

buildmodels

rcsbpdb.bsky.social

rcsb-protein-data-bank

rcsb