

World Wide Web for Macromolecular Crystallography

Eugene Krissinel

CCP4, STFC Research Complex at Harwell
Didcot, United Kingdom

eugene.krissinel@stfc.ac.uk

Macromolecular Crystallography School "From data processing to structure refinement and beyond"
April 6-16, 2015, Montevideo, Uruguay



Research Complex at Harwell

WWW in Essence

- World Wide Web is, in essence, all about 2 things:
 - *information and means to work with it (including communication)*
 - *services and control*

Both are relevant to Macromolecular Crystallography.

- “Information” aspect includes (examples)
 - *various databases and corresponding web-services (PDB, EBI, NCBI)*
 - *educational resources (e.g. university courses, CCP4/PDB tutorials)*
 - *relevant forums (e.g. CCP4 BB, PHENIX BB, COOT BB)*
- “Services and control” aspect includes (examples)
 - *remote data collection at synchrotrons (just the reality nowadays)*
 - *cloud computing (check with your IT support :))*
 - *specialised web-based resources for MX computations (CCP4)*



Research Complex at Harwell

Databases

- 📌 Given to us in plenty, with confusing variety and redundancy
 - *these days, it is often cheaper to conduct certain experiments than to find the information needed. This is a genuine problem*
- 📌 We will concentrate on few resources relevant to MX
 - *experiment management databases*
 - *sequence databases*
 - *macromolecular structure databases*
 - *small molecule databases*
 - *databases with derived data, useful for structure solution and analysis*
- 📌 No hyperlinks in this presentation, just acronyms

Google is your friend! so google up.



Research Complex at Harwell

Experiment Management Databases

- Would be useful for keeping all experimental details such as
 - *construct design*
 - *protein expression*
 - *crystallisation protocols*
- No universal or commonly accepted repository exists, despite many attempts to establish one
 - *kept local/proprietary to projects/labs*
 - *contents and used solutions vary enormously*
 - *usually not highly appreciated by researchers (“it’s a burden”)*
 - *CCP4 invested in PIMS, now installed in 18 locations*
 - *keeping experimental information and making it accessible remains a not fully resolved and a challenging problem*

PIMS

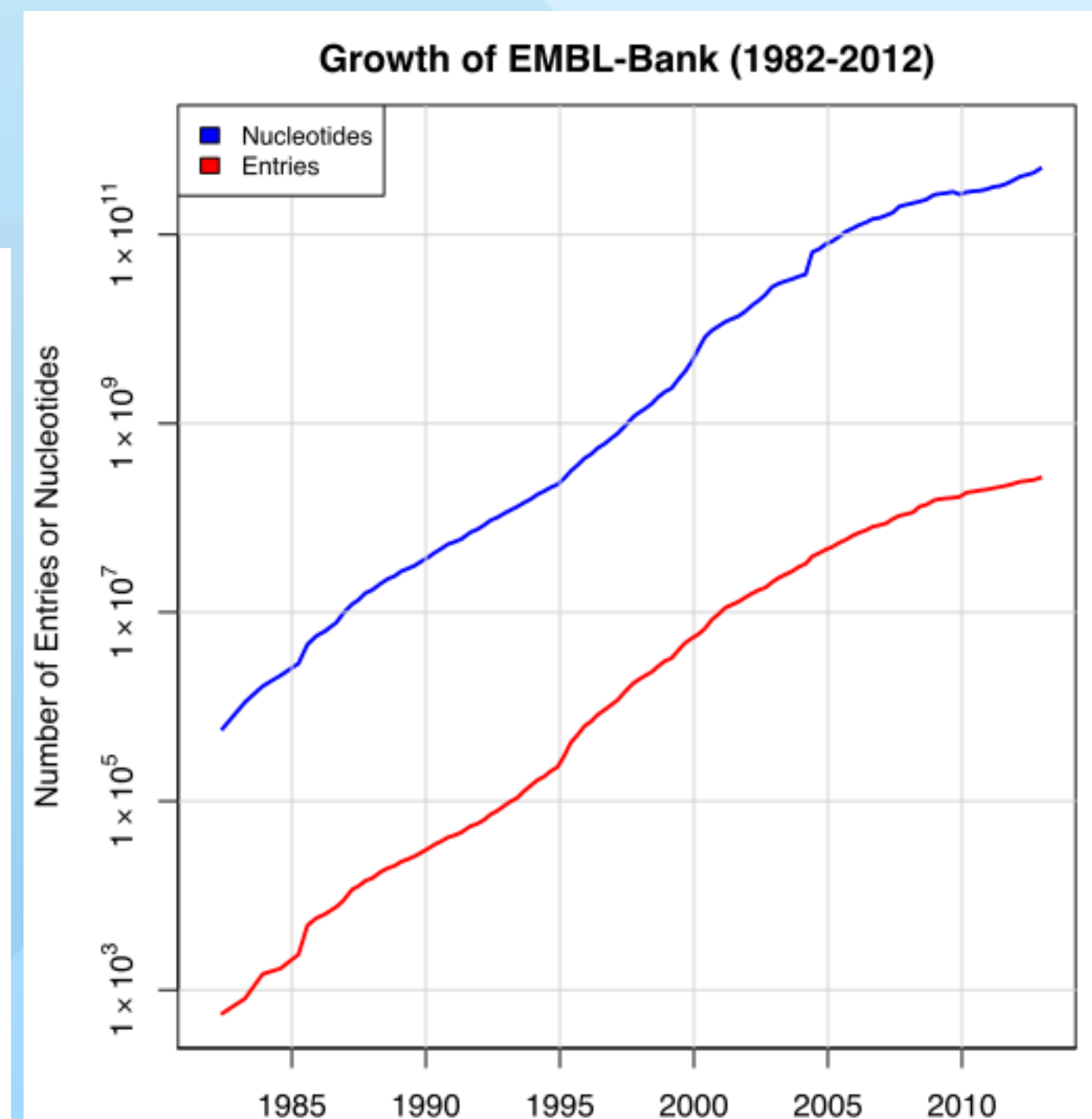
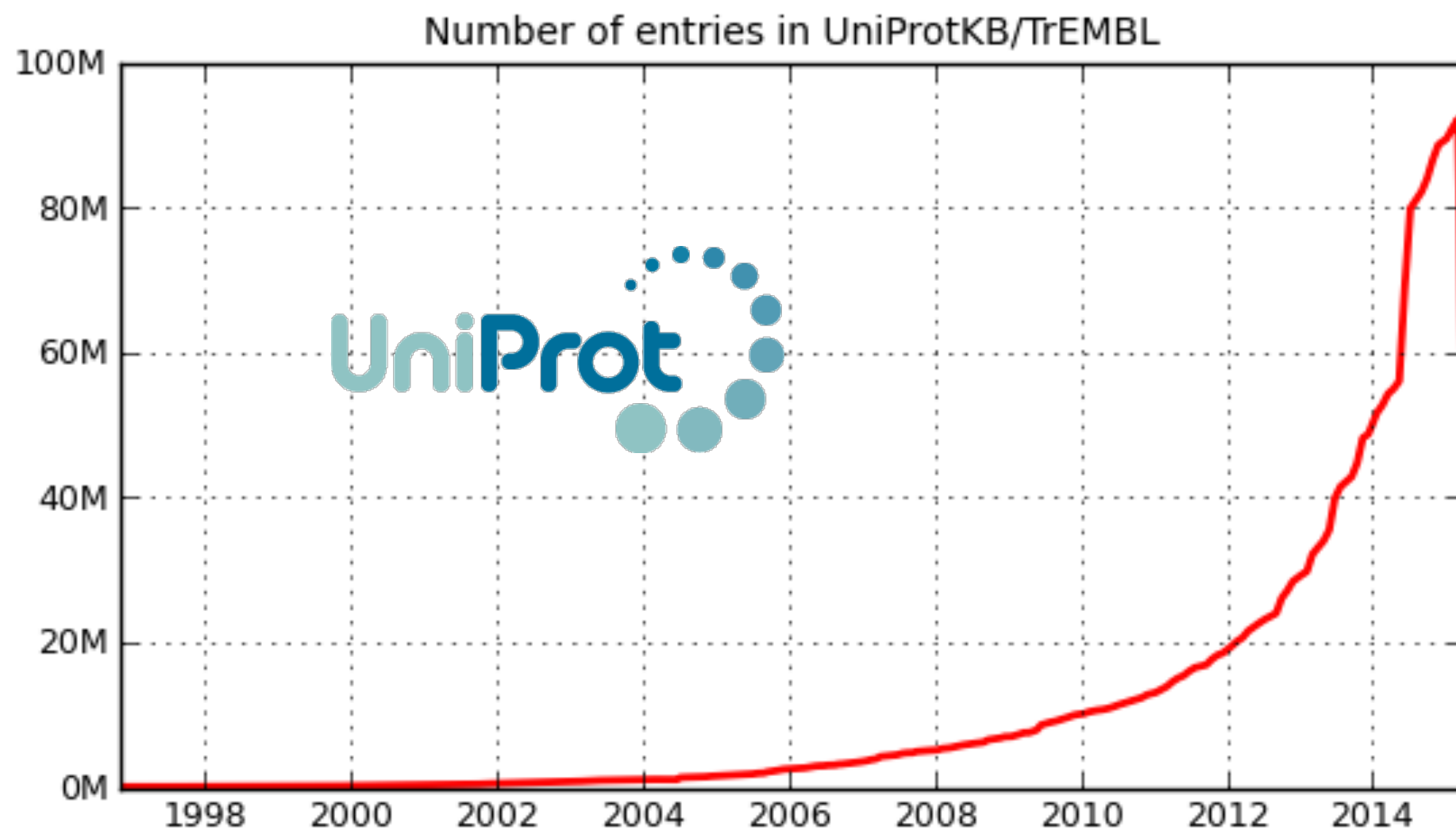


Research Complex at Harwell

Protein Sequence Databases

Well established, many but with few main resources available

- *GenBank (NCBI): nucleotide sequences and their protein translations*
- *European Nucleotide Archive (EMBL): DNA and RNA sequences*
- *DNA Bank of Japan (National Institute of Genetics)*
- **UNIPROT (EBI+SIB+PIR). This is probably the major resource**



UniProt

- *Text Search*
- *BLAST Search*
- *Sequence Alignment Service*
- *Sequence Retrieve*
- *Data Download*
- *Data Deposition*
- *Statistics*
- *only 1% curated*



The screenshot shows the UniProt website homepage. At the top is the UniProt logo and a search bar with a dropdown menu set to 'UniProtKB'. Below the search bar is a navigation bar with links for 'BLAST', 'Align', 'Retrieve/ID mapping', 'Help', and 'Contact'. The main content area features a mission statement: 'The mission of UniProt is to provide the scientific community with a comprehensive, high-quality and freely accessible resource of protein sequence and functional information.' Below this are four colored boxes: 'UniProtKB' (Swiss-Prot 548,208, manually annotated and reviewed; TrEMBL 46,714,516, automatically annotated and not reviewed), 'UniRef' (Sequence clusters), 'UniParc' (Sequence archive), and 'Proteomes'. A 'Supporting data' section includes links for Literature citations, Cross-ref. databases, Taxonomy, Diseases, Subcellular locations, and Keywords. On the right, a 'News' section features social media icons and a list of recent updates, including 'Of CAT tails and protein translation by-products' and 'Regulation of translation initiation through folding'. At the bottom, there are four columns: 'Getting started' with links for Text search, BLAST, and Sequence alignments; 'UniProt data' with links for Download latest release, Statistics, and Forthcoming changes; 'Protein spotlight' featuring a 'Taming Genes' article from March 2015; and a 'News archive' link.

UniProt

UniProtKB

Advanced Search

BLAST Align Retrieve/ID mapping Help Contact

The mission of UniProt is to provide the scientific community with a comprehensive, high-quality and freely accessible resource of protein sequence and functional information.

UniProtKB

Swiss-Prot (548,208)

Manually annotated and reviewed.

TrEMBL (46,714,516)

Automatically annotated and not reviewed.

UniRef

Sequence clusters

UniParc

Sequence archive

Proteomes

Supporting data

Literature citations

Cross-ref. databases

Taxonomy

Diseases

Subcellular locations

Keywords

News

Of CAT tails and protein translation by-products | Reducing redundancy in proteomes | Retirement of UniProt Metagenomic and Environmental...
UniProt release 2015_04

Regulation of translation initiation through folding | New proteomics mapping files | New FTP repository for reference

News archive

Getting started

Text search

Our basic text search allows you to search all the resources available

BLAST

Find regions of similarity between your sequences

Sequence alignments

UniProt data

Download latest release

Get the UniProt data

Statistics

View Swiss-Prot and TrEMBL statistics

Forthcoming changes

Planned changes for the UniProt

Protein spotlight

Taming Genes

March 2015

Protein Structure Databases

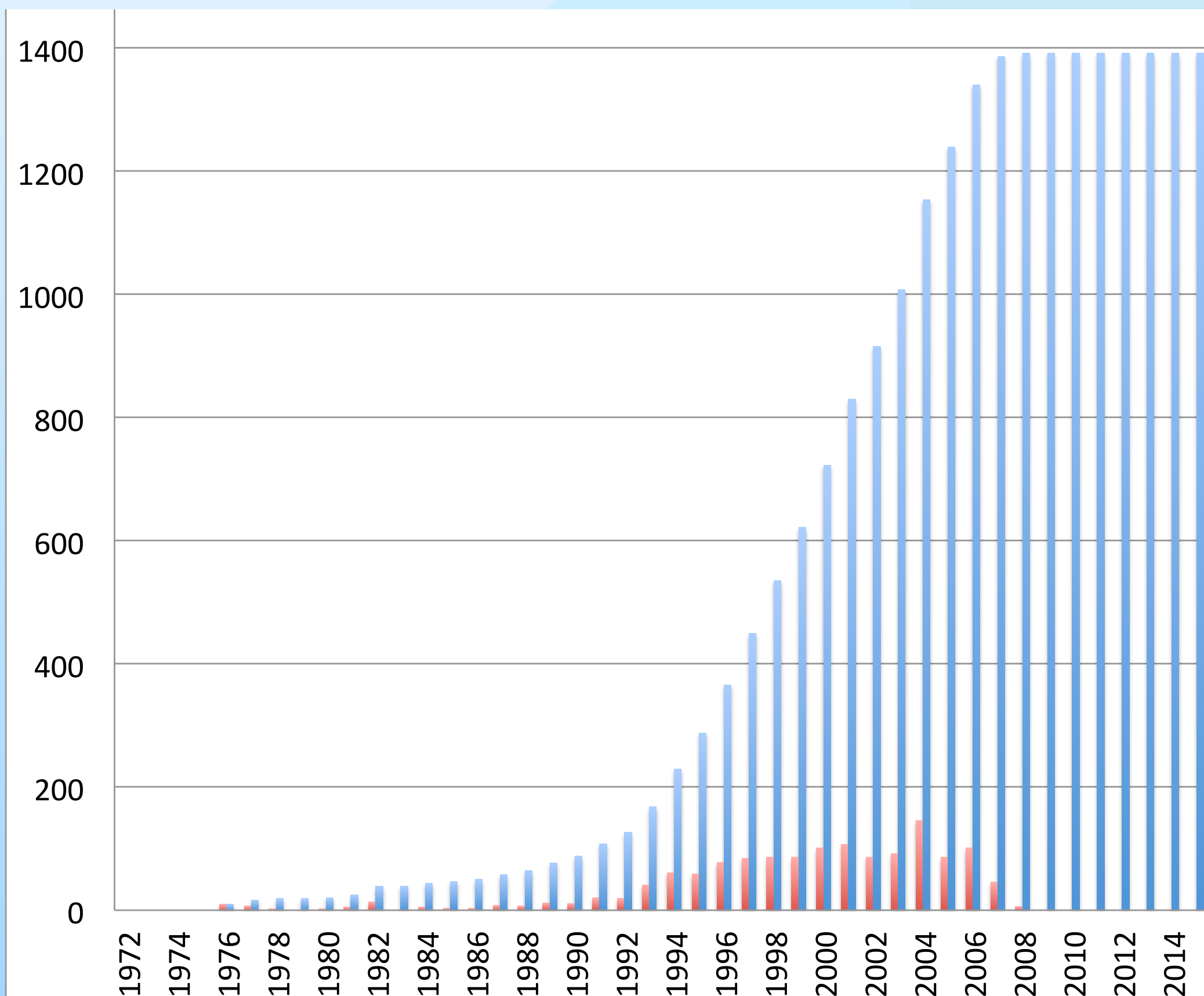
- 📌 Actually a single resource, the wwPDB, with 3 gateways:
 - *PDB @ RCSB (Piscataway, NJ) Head branch*
 - *PDBe @ EBI (Cambridge, UK) European branch*
 - *PDBj @ IPR (Osaka, Japan) East-Asian branch*
- 📌 All PDB entries are curated, this is a major effort
- 📌 Precious
 - \$4B estimated recovery cost if all data is lost*
- 📌 Now has a common deposition system across all branches
- 📌 A few derived resources are there (not a complete list)
 - *MMDB from NCBI - computed domains and literature links*
 - *SCOP from MRC/LMB - curated hierarchical structural domains*
 - *Proteopedia from Weizmann Institute, Israel - the PDB Wiki*



Research Complex at Harwell

Protein Data Bank

- Currently over 107,000 structures (90,000 X-ray structures)
- For an X-ray crystallographer, a treasury of MR models



Chothia C: **One thousand families for the molecular biologist. Nature 1992, 357: 543-544.**

- MR is the dominating phasing method in today's protein crystallography



Protein Data Bank Tools

RCSB
PDB
PROTEIN DATA BANK

An Information Portal to
107754 Biological
Macromolecular Structures

Search by PDB ID, author, macromolecule, sequence, or ligands

Go

[Advanced Search](#) | [Browse by Annotations](#)

RCSB PDB-101

WORLDWIDE
PDB
PROTEIN DATA BANK

EMDataBank
Unified Data Resource for 3DEM

ndb
NUCLEIC ACID
DATABASE

StructuralBiology
Knowledgebase



Welcome

Deposit

Search

Visualize

Analyze

Download

Learn

A Structural View of Biology

This resource is powered by the Protein Data Bank archive—information about the 3D shapes of proteins, nucleic acids, and complex assemblies that helps students and researchers understand all aspects of biomedicine and agriculture, from protein synthesis to health and disease.

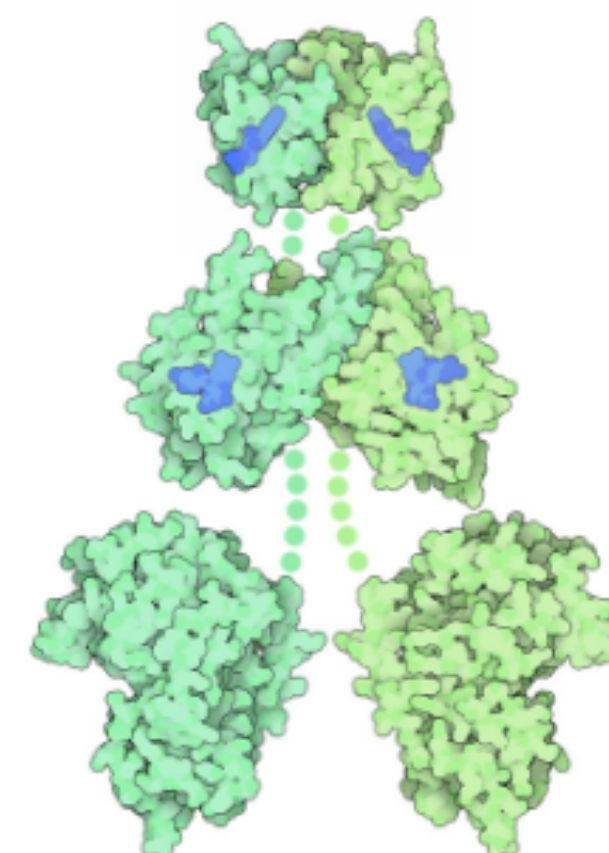
As a member of the wwPDB, the RCSB PDB curates and annotates PDB data.

The RCSB PDB builds upon the data by creating tools and resources for research and education in molecular biology, structural biology, computational biology, and beyond.

Structure and Health Focus: HIV



March Molecule of the Month



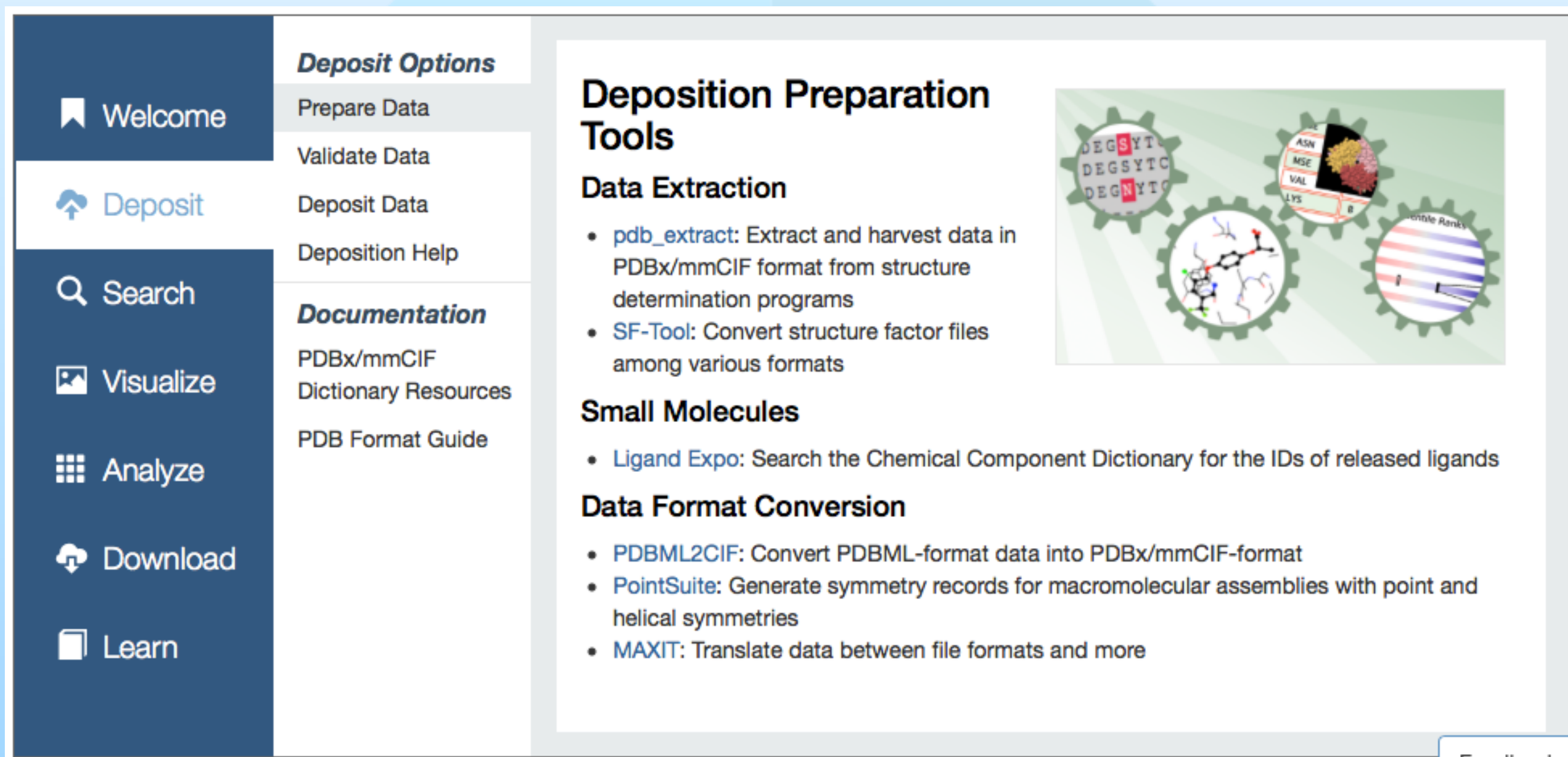
Phototropin



Protein Data Bank Deposition

Step 1: Data preparation

- *harvest data from CCP4 Projects (will be available in CCP4i2)*
- *prepare structure factors in mmCIF format*
- *check Chemical Compound Dictionary for IDs of released ligands*



The screenshot shows the PDB Deposition Preparation Tools website. On the left is a dark blue sidebar with navigation links: Welcome, Deposit, Search, Visualize, Analyze, Download, and Learn. The main content area has a white background with a left-hand menu containing 'Deposit Options' (Prepare Data, Validate Data, Deposit Data, Deposition Help) and 'Documentation' (PDBx/mmCIF, Dictionary Resources, PDB Format Guide). The main content area is titled 'Deposition Preparation Tools' and includes sections for 'Data Extraction' (with links to [pdb_extract](#) and [SF-Tool](#)), 'Small Molecules' (with a link to [Ligand Expo](#)), and 'Data Format Conversion' (with links to [PDBML2CIF](#), [PointSuite](#), and [MAXIT](#)). An illustration of interlocking gears with molecular and data-related icons is positioned to the right of the 'Data Extraction' section. A 'Feedback' button is located in the bottom right corner.

Deposition Preparation Tools

Data Extraction

- [pdb_extract](#): Extract and harvest data in PDBx/mmCIF format from structure determination programs
- [SF-Tool](#): Convert structure factor files among various formats

Small Molecules

- [Ligand Expo](#): Search the Chemical Component Dictionary for the IDs of released ligands

Data Format Conversion

- [PDBML2CIF](#): Convert PDBML-format data into PDBx/mmCIF-format
- [PointSuite](#): Generate symmetry records for macromolecular assemblies with point and helical symmetries
- [MAXIT](#): Translate data between file formats and more

Protein Data Bank Deposition

Step 2: Data validation

- *automatic service created as a result of VTF initiative*
- *assesses quality of structure and correspondence with the experimental data*
- *checks for many obvious flaws; extremely useful*

Welcome

Deposit

Search

Visualize

Analyze

Download

Learn

Deposit Options

Prepare Data

Validate Data

Deposit Data

Deposition Help

Documentation

PDBx/mmCIF

Dictionary Resources

PDB Format Guide

Validation Services





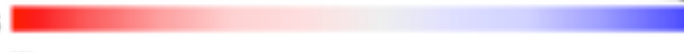
Validation reports contain an assessment of the quality of a structure and highlight specific concerns by considering the coordinates of the model, the experimental data and the fit between the two. Easily interpretable summary information that compares the quality of a model with that of other models in the archive will help users of PDB data to critically assess archived entries and to select the most appropriate structural models for their needs. These reports are developed using the recommendations of the [wwPDB Validation Task Forces](#).

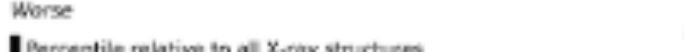
Reports for released entries are available from Structure Summary pages.



Validation reports for manuscript reviewers are created during annotation of deposited structures.

Information and [example Validation Reports \(at wwpdb.org\)](#).

[Check your X-ray crystal structure before depositing \(standalone server\)](#).

Metric	Percentile Ranks	Value
Rfree		0.189
Clashscore		4
Ramachandran outliers		0
Sidechain outliers		2.4%
RSRZ outliers		0

Worse  Better

 Percentile relative to all X-ray structures
 Percentile relative to X-ray structures of similar resolution

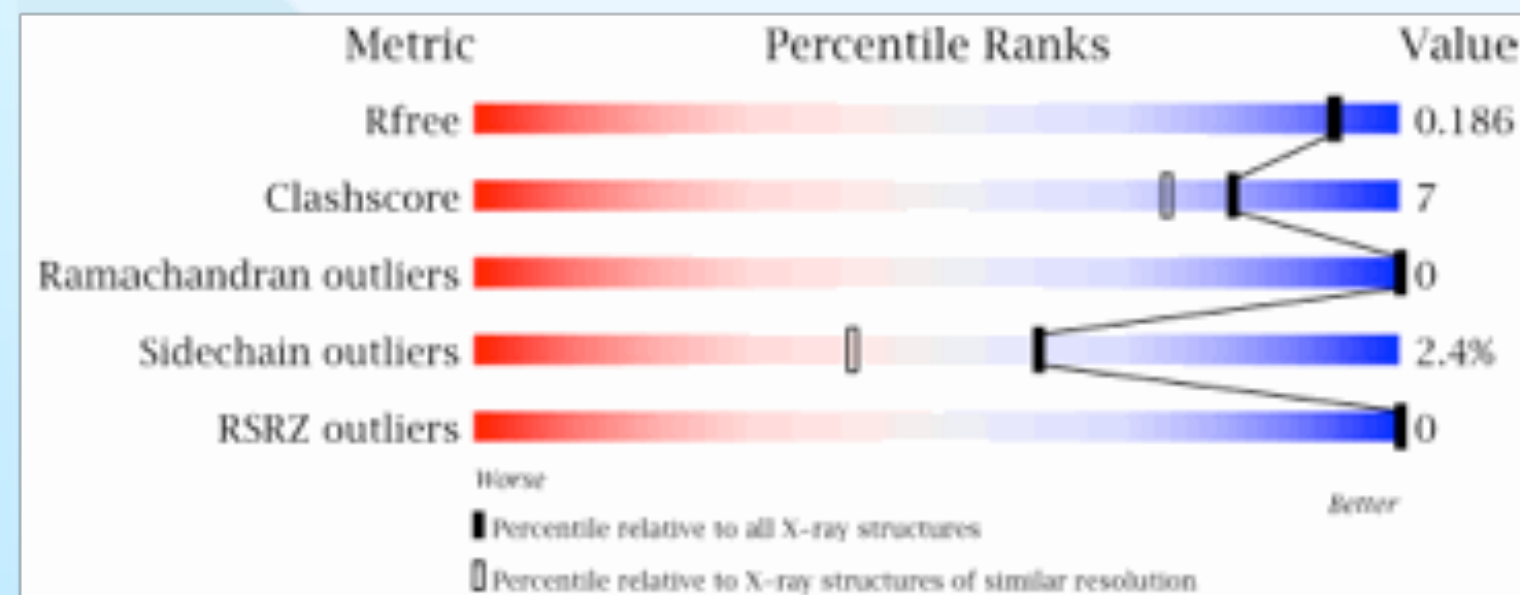
Feedback

VTF Validation Report

(excerpts)

1. Overall quality

- *data for each slider is detailed in individual report sections*



2. Chain quality

- *coloured representation of outliers*

Mol	Chain	Length	Quality of chain
1	A	207	
1	B	207	

3. Compound fit quality

- *wrong geometry or poor fit in density*

Mol	Type	Chain	Res	Geometry	Electron density
6	GOL	A	381	-	X
4	SO4	A	391	-	X
5	OGA	A	370	X	-

4. Extended outlier statistics

- *bond length, angles, chirality, planarity*
- *close contacts*
- *protein backbone (Ramachandran)*
- *backbone torsions*
- *sugar packing*



VTF Validation Report

1. Overall quality

- *summary of key quality indicators to spot possible problems quickly*

2. Entry composition

- *summary of the number of unique molecules and how they have been modelled*

3. Residue-property plots

- *per-residue summary plots of quality information for protein, RNA and DNA molecules*

4. Data and refinement statistics

- *Space Group, Cell, Completeness, R_{merge} , R_{free} , Resolution, Twinning tests etc etc*

5. Model quality

- *summary on geometrical quality, chirality, close contacts etc*

6. Fit of model and data


- *analysis of the fit of the molecule in the entry to experimental data (based on Real Space R-value)*

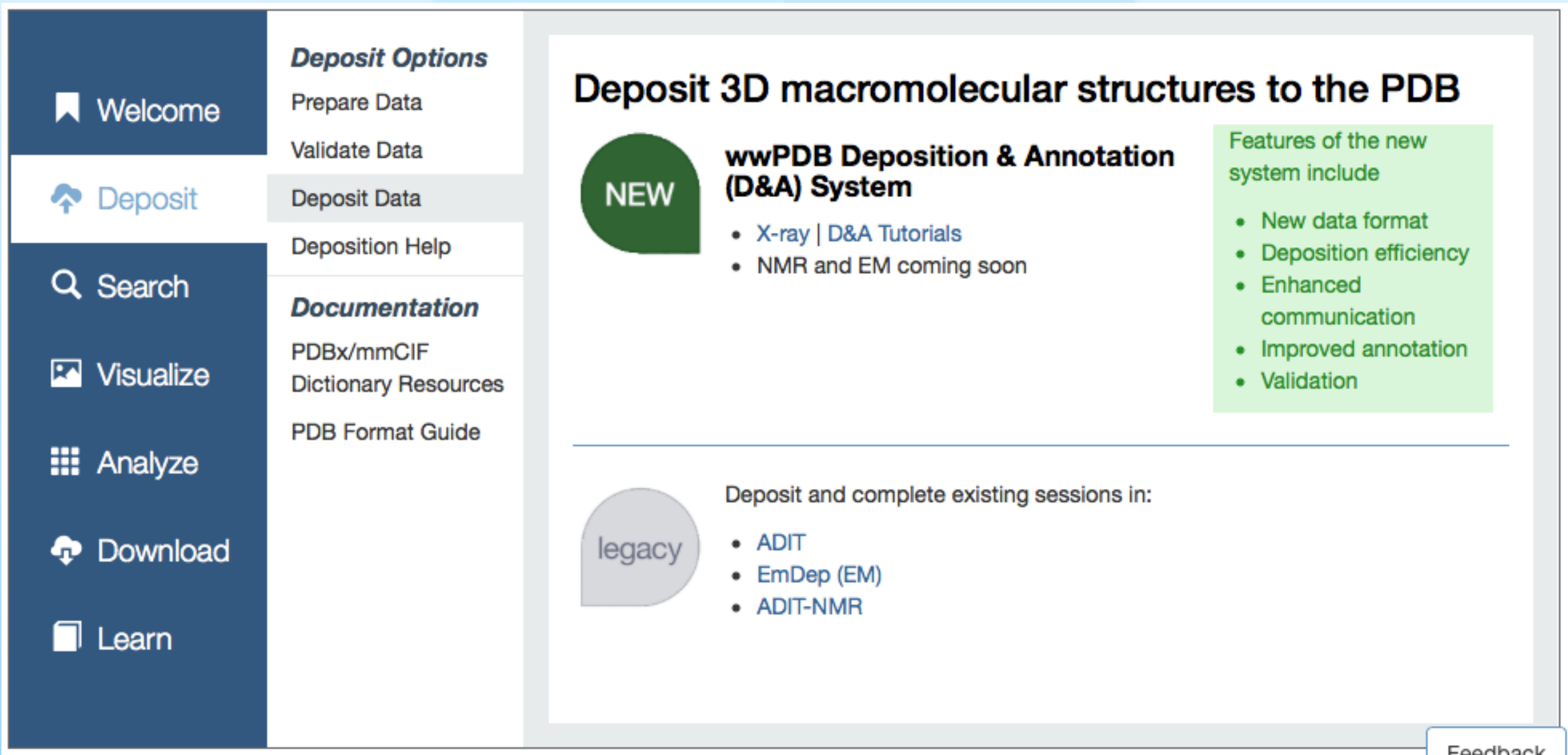
R. J. Read, P. D. Adams, W. B. Arendall III, A. T. Brunger, P. Emsley, R. P. Joosten, G. J. Kleywegt, E. B. Krissinel, T. Lütke, Z. Otwinowski, A. Perrakis, J. S. Richardson, W. H. Sheffler, J. L. Smith, I. J. Tickle, G. Vriend and P. H. Zwart. (2011) Structure 19: 1395-1412.



Research Complex at Harwell

Protein Data Bank Deposition

-  **Step 3: Data deposition**
- *a multi-stage procedure of giving all the necessary details about your structure*
 - *most details are taken from harvest files*
 - *well-documented and equipped with video tutorials*



The screenshot displays the wwPDB Deposition & Annotation (D&A) System interface. On the left is a dark blue sidebar with navigation links: Welcome, Deposit (highlighted with a cloud icon), Search, Visualize, Analyze, Download, and Learn. The main content area has a white background. A top navigation bar contains 'Deposit Options' (Prepare Data, Validate Data, Deposit Data, Deposition Help) and 'Documentation' (PDBx/mmCIF, Dictionary Resources, PDB Format Guide). The main heading is 'Deposit 3D macromolecular structures to the PDB'. Below this, a green 'NEW' badge is next to the title 'wwPDB Deposition & Annotation (D&A) System'. A list of features includes 'X-ray | D&A Tutorials' and 'NMR and EM coming soon'. To the right, a green box lists 'Features of the new system include': New data format, Deposition efficiency, Enhanced communication, Improved annotation, and Validation. At the bottom, a 'legacy' badge is next to the text 'Deposit and complete existing sessions in:', followed by a list: ADIT, EmDep (EM), and ADIT-NMR. A 'Feedback' button is in the bottom right corner.

Deposit 3D macromolecular structures to the PDB

NEW

wwPDB Deposition & Annotation (D&A) System

- X-ray | D&A Tutorials
- NMR and EM coming soon

Features of the new system include

- New data format
- Deposition efficiency
- Enhanced communication
- Improved annotation
- Validation

legacy

Deposit and complete existing sessions in:

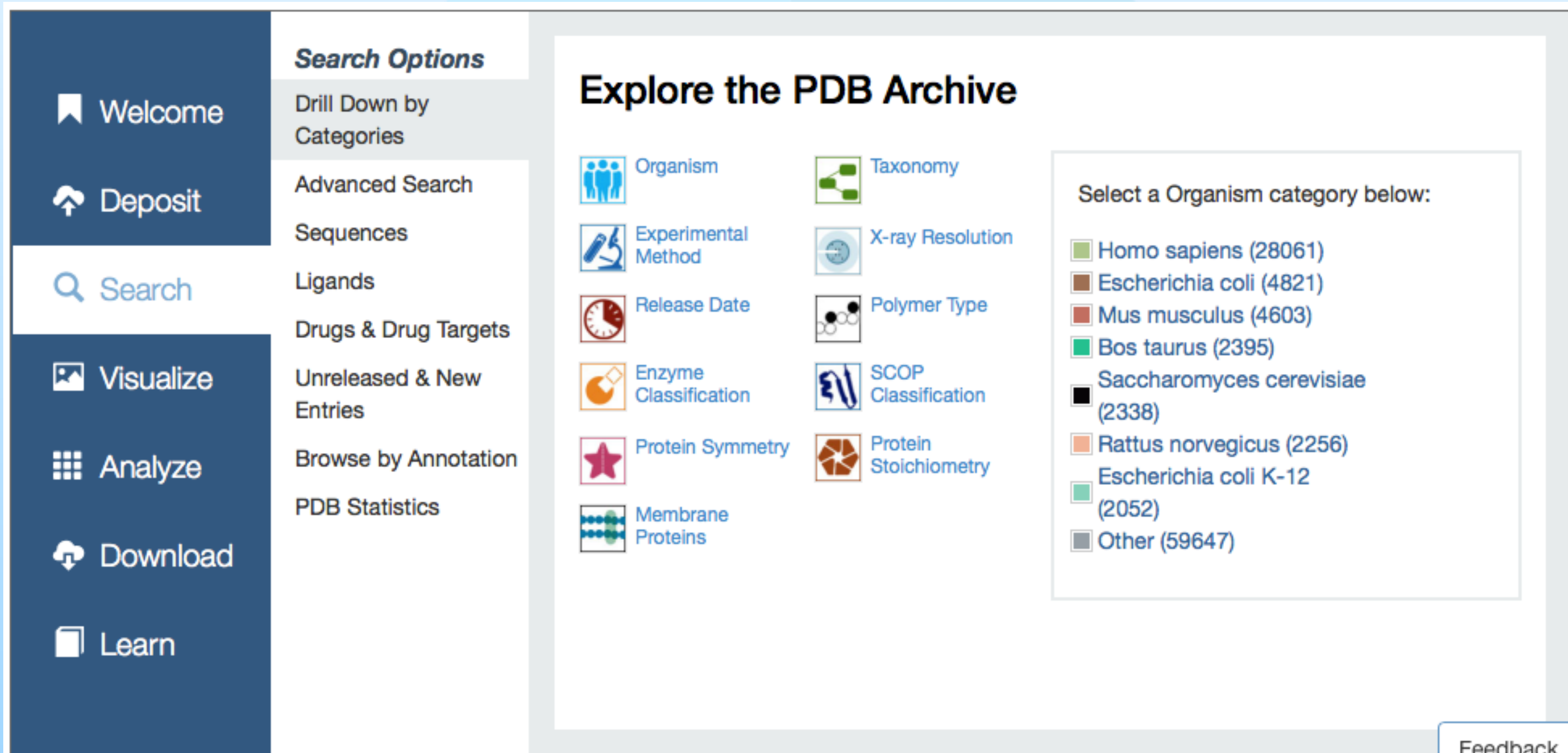
- ADIT
- EmDep (EM)
- ADIT-NMR

Feedback

Protein Data Bank Searches

 Includes keyword, annotation, sequence and ligand structure searches

- *has both “search” and “explore” options*
- *self-explicable*
- *accompanied with pre-computed statistics*



The screenshot displays the PDB website's navigation and search interface. On the left is a dark blue sidebar with icons and labels for 'Welcome', 'Deposit', 'Search', 'Visualize', 'Analyze', 'Download', and 'Learn'. The main content area is divided into two sections. The left section, titled 'Search Options', lists various search methods: 'Drill Down by Categories', 'Advanced Search', 'Sequences', 'Ligands', 'Drugs & Drug Targets', 'Unreleased & New Entries', 'Browse by Annotation', and 'PDB Statistics'. The right section, titled 'Explore the PDB Archive', features a grid of 12 categories with icons: Organism, Taxonomy, Experimental Method, X-ray Resolution, Release Date, Polymer Type, Enzyme Classification, SCOP Classification, Protein Symmetry, Protein Stoichiometry, Membrane Proteins, and a 'Select a Organism category below:' box. This box contains a list of organisms with their respective PDB entry counts: Homo sapiens (28061), Escherichia coli (4821), Mus musculus (4603), Bos taurus (2395), Saccharomyces cerevisiae (2338), Rattus norvegicus (2256), Escherichia coli K-12 (2052), and Other (59647). A 'Feedback' button is located in the bottom right corner.

Search Options

- Drill Down by Categories
- Advanced Search
- Sequences
- Ligands
- Drugs & Drug Targets
- Unreleased & New Entries
- Browse by Annotation
- PDB Statistics

Explore the PDB Archive

- Organism
- Taxonomy
- Experimental Method
- X-ray Resolution
- Release Date
- Polymer Type
- Enzyme Classification
- SCOP Classification
- Protein Symmetry
- Protein Stoichiometry
- Membrane Proteins

Select a Organism category below:

- Homo sapiens (28061)
- Escherichia coli (4821)
- Mus musculus (4603)
- Bos taurus (2395)
- Saccharomyces cerevisiae (2338)
- Rattus norvegicus (2256)
- Escherichia coli K-12 (2052)
- Other (59647)

Feedback

Protein Data Bank Structure Analysis Services

- Alignment tools (sequence and structure)
- Protein symmetry visualisation
- Structure quality (VTF report)
- Genomic mapping
- Reference to other resources
- Statistics

 Welcome Deposit Search Visualize Analyze Download Learn

Sequence- Structure Alignment

Sequence & Structure
Alignment

Protein Symmetry

Other Featured Tools

Structure Quality

Map Genomic
Location to Protein

Third Party Tools

PDB Statistics

Sequence & Structure Alignment

RCSB PDB's Comparison Tool calculates pairwise sequence (blast2seq, Needleman-Wunsch, and Smith-Waterman) and structure alignments (FATCAT, CE, Mammoth, TM-Align, TopMatch).

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 and detection of circular permutations.



Select Associated Chain ID

...

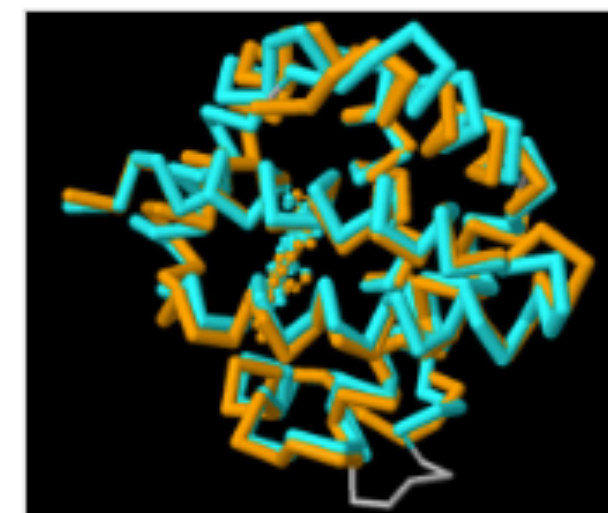
Select Associated Chain ID

...

- Select Comparison Method -

Align

More Options



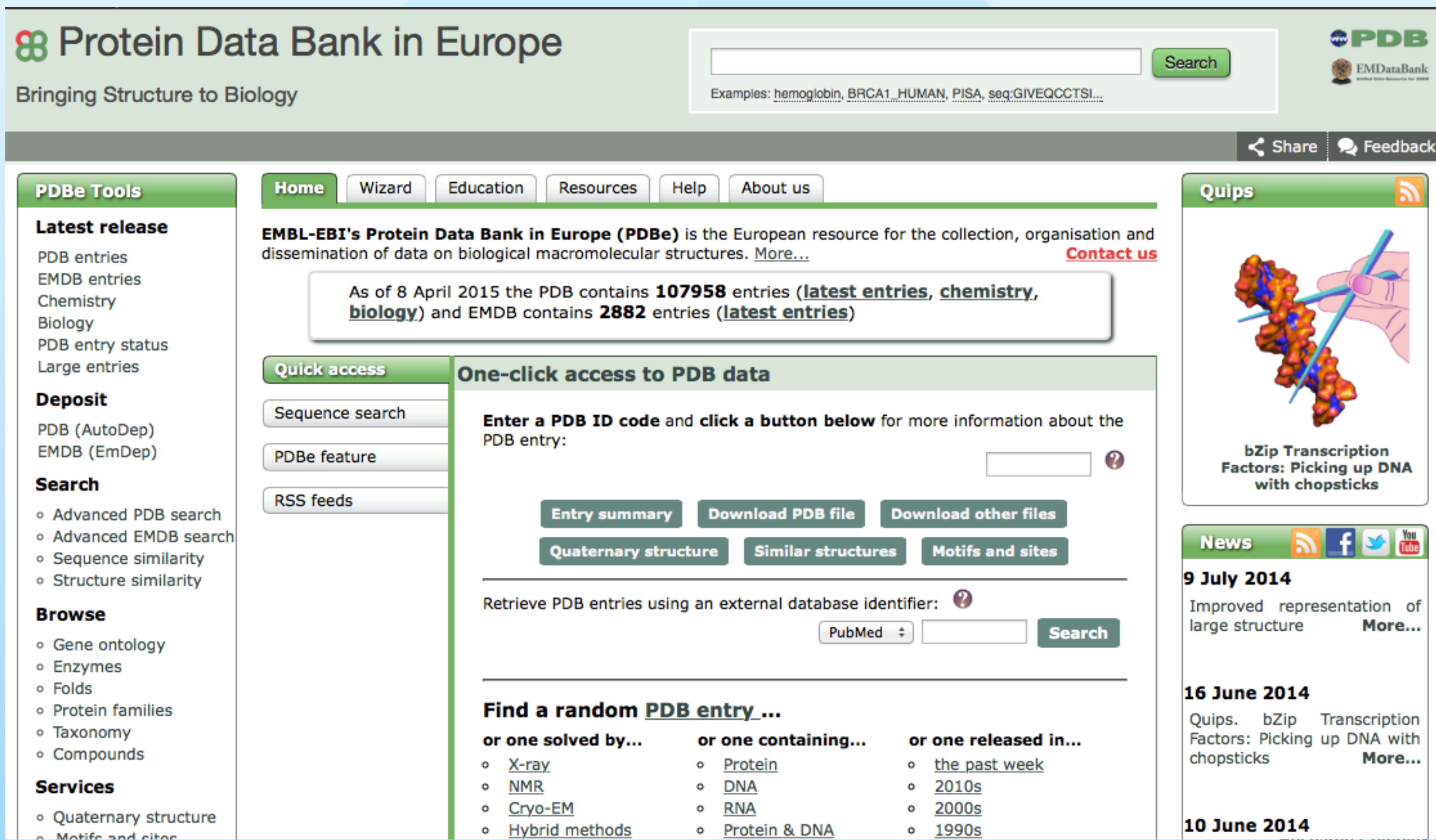
PDBe Gateway



Offers the same deposition services
Different tools for exploring PDB archive in multiple ways



Unique services: PISA and PDBeMotif
Useful services: SSM (PDBeFold) and PDBeChem



Protein Data Bank in Europe
Bringing Structure to Biology

Search: Search
Examples: hemoglobin, BRCA1_HUMAN, PISA, seq:GIVEQCCTSI...

Share Feedback

PDBe Tools

- Latest release**
 - PDB entries
 - EMDB entries
 - Chemistry
 - Biology
 - PDB entry status
 - Large entries
- Deposit**
 - PDB (AutoDep)
 - EMDB (EmDep)
- Search**
 - Advanced PDB search
 - Advanced EMDb search
 - Sequence similarity
 - Structure similarity
- Browse**
 - Gene ontology
 - Enzymes
 - Folds
 - Protein families
 - Taxonomy
 - Compounds
- Services**
 - Quaternary structure
 - Motifs and sites

Home Wizard Education Resources Help About us

EMBL-EBI's Protein Data Bank in Europe (PDBe) is the European resource for the collection, organisation and dissemination of data on biological macromolecular structures. [More...](#) [Contact us](#)

As of 8 April 2015 the PDB contains **107958** entries ([latest entries](#), [chemistry](#), [biology](#)) and EMDb contains **2882** entries ([latest entries](#))

Quick access

- Sequence search
- PDBe feature
- RSS feeds

One-click access to PDB data

Enter a **PDB ID code** and **click a button below** for more information about the PDB entry:

?

[Entry summary](#)
[Download PDB file](#)
[Download other files](#)

[Quaternary structure](#)
[Similar structures](#)
[Motifs and sites](#)

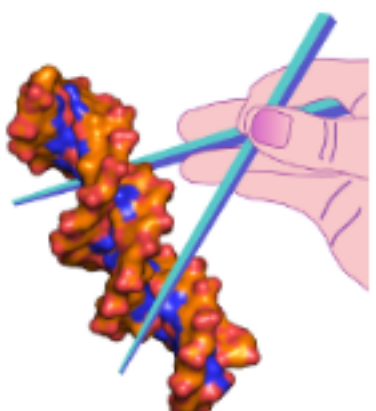
Retrieve PDB entries using an external database identifier: ?

PubMed Search

Find a random PDB entry ...

or one solved by...	or one containing...	or one released in...
<ul style="list-style-type: none"> X-ray NMR Cryo-EM Hybrid methods 	<ul style="list-style-type: none"> Protein DNA RNA Protein & DNA 	<ul style="list-style-type: none"> the past week 2010s 2000s 1990s

Quips



bZip Transcription Factors: Picking up DNA with chopsticks

News




9 July 2014
Improved representation of large structure [More...](#)



16 June 2014
Quips. bZip Transcription Factors: Picking up DNA with chopsticks [More...](#)


10 June 2014

research complex at Harwell

Proteopedia

-  Protein WiKi
-  Editable articles
-  Community-curated

-  Setup in Weizmann Institute of Science, Israel
-  A prototype of future resources?



PROTEOPEDIA
—LIFE IN 3D—

navigation

- Main Page
- Table of Contents
- Structure Index
- Random
- Recent Changes
- Help

search

Go Search

Search

toolbox

- Upload file
- Special pages
- Printable version
- Permanent link

article | discussion | edit this page | history

Welcome to Proteopedia The free, collaborative 3D-encyclopedia of proteins & other molecules *ISSN 2310-6301*

Table of Contents * Structure Index * Function to Structure * About * Video Guide * Help

Featured Article

Green links change the 3D image!
Click and drag on the molecule!



HIV-1 Protease JSmol

toggle spin toggle quality popup

HIV-1 protease
by David Canner

The X-ray structure of HIV-1 protease reveals that it is composed of **two symmetrically related subunits**, each consisting of 99 amino acid residues. The subunits come together in such a way as to **form a tunnel where they meet**. This tunnel is of critical importance because the active site of the protease is located in its interior. The active site consists of two Asn. The Glu

Proteopedia News

- Talking about Proteopedia [on 12/04/14-12/06/14](#), a live online talk organised by DivCHED CCCE: Committee on Computers in Chemical Education
- [Course](#) in Spanish/English on Proteopedia and its uses to study, display and teach macromolecules.
- [How to create fast loading pages](#) in Proteopedia.
- [How to be as safe as possible with Java](#) (a *Must read* for Proteopedia users)
- [Proteopedia on iPads!](#)
- [What version of Jmol is running?](#)
- [Proteopedia status](#)
- [Awards](#)

The Scoreboard

Last 30 Days (Top 10)

Score	Pages	Changes	Username
28	8	110	Mackenzie A. Smith
26	11	69	Mark Meredith
24	7	76	Braden Sciarra

Small Molecule Databases

- 📌 Indispensable resource for finding information about ligands
 - *most of MX research is about protein-ligand interactions*
 - *fitting ligands is often the most valuable and most difficult task*
 - *ligand variety is almost unlimited*
 - *ligand structure is often difficult to compute accurately enough*
- 📌 Chemical structure databases are given in confusing variety
 - *ZincDatabase (UCSF)*
 - *ChEMBL (EBI)*
 - *PDB Compound Library (RCSB)*
 - *Chemspider (RSC)*
 - *and many, many others, Google is your friend*
- 📌 They are not equivalent to each other in what they offer
- 📌 Many are commercialised

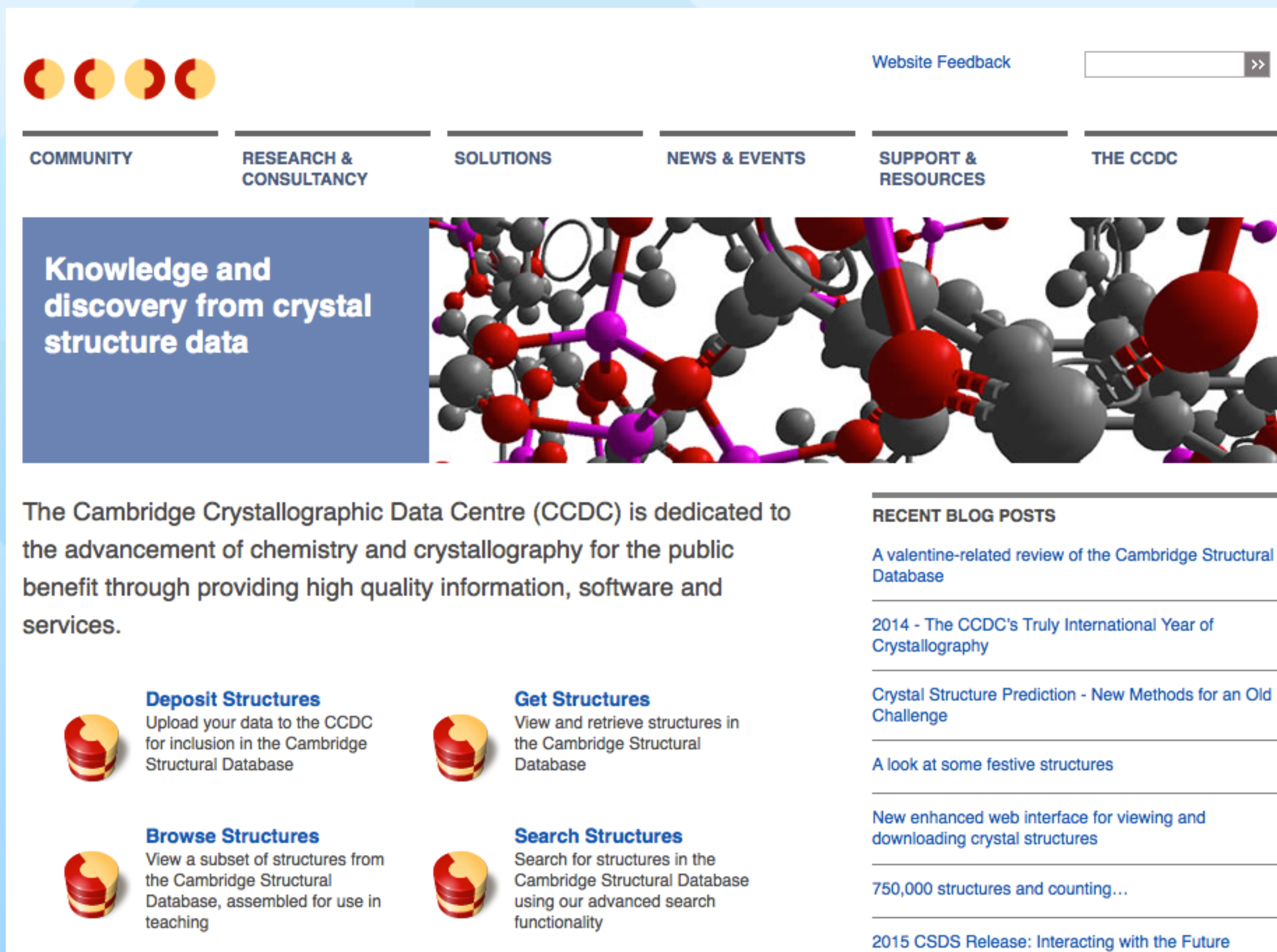


Research Complex at Harwell

Cambridge Structural Database (CSD)

 Maintained by the Cambridge Crystallographic Data Centre (CCDC)

- *effectively a "PDB" for small molecule crystallographers*
- *currently 0.75M structures*
- *extensive services (deposition, search, browse, view etc)*
- *commercialised (basic features available for academics)*



The screenshot shows the Cambridge Structural Database (CSD) website. At the top, there are four orange circular icons and a "Website Feedback" form. Below this is a navigation bar with links: COMMUNITY, RESEARCH & CONSULTANCY, SOLUTIONS, NEWS & EVENTS, SUPPORT & RESOURCES, and THE CCDC. The main content area features a large blue box with the text "Knowledge and discovery from crystal structure data" and a 3D molecular model. Below this, a paragraph states: "The Cambridge Crystallographic Data Centre (CCDC) is dedicated to the advancement of chemistry and crystallography for the public benefit through providing high quality information, software and services." To the right, there is a "RECENT BLOG POSTS" section with several links. At the bottom, there are four service tiles, each with a circular icon and a description: "Deposit Structures" (Upload your data to the CCDC for inclusion in the Cambridge Structural Database), "Get Structures" (View and retrieve structures in the Cambridge Structural Database), "Browse Structures" (View a subset of structures from the Cambridge Structural Database, assembled for use in teaching), and "Search Structures" (Search for structures in the Cambridge Structural Database using our advanced search functionality).

Website Feedback

COMMUNITY RESEARCH & CONSULTANCY SOLUTIONS NEWS & EVENTS SUPPORT & RESOURCES THE CCDC

Knowledge and discovery from crystal structure data

The Cambridge Crystallographic Data Centre (CCDC) is dedicated to the advancement of chemistry and crystallography for the public benefit through providing high quality information, software and services.

Deposit Structures
Upload your data to the CCDC for inclusion in the Cambridge Structural Database

Get Structures
View and retrieve structures in the Cambridge Structural Database

Browse Structures
View a subset of structures from the Cambridge Structural Database, assembled for use in teaching

Search Structures
Search for structures in the Cambridge Structural Database using our advanced search functionality

RECENT BLOG POSTS

- [A valentine-related review of the Cambridge Structural Database](#)
- [2014 - The CCDC's Truly International Year of Crystallography](#)
- [Crystal Structure Prediction - New Methods for an Old Challenge](#)
- [A look at some festive structures](#)
- [New enhanced web interface for viewing and downloading crystal structures](#)
- [750,000 structures and counting...](#)
- [2015 CSDS Release: Interacting with the Future](#)

CSD for crystallographers



Structure visualisation, look-up, validation, restraints generation


- small print: restraint generation through Mogul plugins in Coot, GPhL Grade and PHENIX's eLBOW*

Crystallography	Crystal structure visualisation	Structure look-up	Structure validation	Storage of in-house structures	Creation of restraint data/ligand dictionaries	Structure solution from powder
CSD System						
ConQuest		✓				
IsoStar			✓			
Mercury	✓					
Mogul			✓		✓	
PreQuest				✓		
WebCSD		✓				
Life Sciences						
Relibase+				✓		
Free Software						
CellCheckCSD		✓				
enCIFer			✓			
Mercury	✓					
Powder Diffraction						
DASH						✓


Crystallography Open Database (COD)

 Maintained by Vilnius University of Biotechnology, Lithuania


- *developed in close collaboration with Cambridge University*
- *currently 0.2M structures*
- *basic services*
- *effectively a public resource*
- *was used for AceDRG development in CCP4*



Crystallography Open Database



COD Home

[Home](#)
[What's new?](#) 

Accessing COD Data

[Browse](#)
[Search](#)
[Search by structural formula](#)

Add Your Data

[Deposit your data](#)
[Manage depositions](#)
[Manage/release prepublications](#)

Documentation

[COD Wiki](#)
[Obtaining COD](#)
[Citing COD](#)
[COD Mirrors](#)
[Advices to donators](#)
[Useful links](#)

Open-access collection of crystal structures of organic, inorganic, metal-organic compounds and minerals, excluding biopolymers.

Including data and software from CrystalEye, developed by department of Chemistry, the University of Cambridge and Pe...

All data on this site have been placed in the public domain by the contributors.

COD Advisory Board thanks [The Research Council of Lithuania](#) for their financial support of the "[Crystallography Open Database \(COD\): an open-access collection of crystal structures and platform collaboration](#)",
Nucleic Acids Research. (2012) [PDF version](#)

We thank [Crystal Impact GbR](#) for their financial support of the publication "[Crystallography Open Database - an open-access collection of crystal structures](#)",
Nucleic Acids Research. (2008) [PDF version](#)

ChEMBL

Maintained by the European Bioinformatics Institute (EBI), Cambridge

EMBL-EBI

Services Research Training About us

ChEMBL wellcome

EBI > Databases > Small Molecules > ChEMBL Database > Home

Search ChEMBL...

Compounds Targets Assays Documents Cells ☐ Exact Match [Activity Source Filter](#)

Ligand Search Target Search Browse Targets Browse Drugs Browse Drug Targets Drug Approvals About

CH₃ OH NH O CH₃ PH₂

List Search

☐ SMILES Search ☒ ChEMBL ID Search ☐ Keyword Search

Please enter a list of Compound IDs, keywords, or SMILES separated by newlines

Fetch Compounds

Biologicals Blast Search

Run BLAST

ChEMBL Statistics

- DB: ChEMBL_20
- Targets: 10,774
- Compound records: 1,715,667
- Distinct compounds: 1,463,270
- Activities: 13,520,737
- Publications: 59,610
- [Release Notes](#)

Web Services



GRADE from Global Phasing Ltd

- *generation of crystallographic restraints*

GΦL



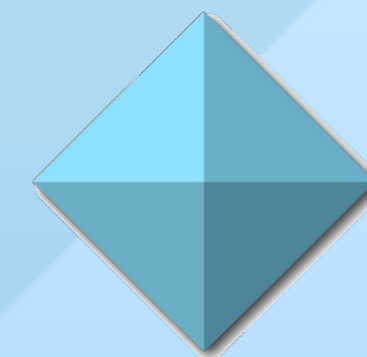
TLS Motion Determination from Ethan Merritt, University of Washington

- *automatic identification of TLS groups*
- *can be used for further refinement in Refmac*



CCP4 on-line

- *BALBES, MrBUMP: automatic MR*
- *ZANUDA: Space Group and Origin validation*
- *jsPISA: Crystal Packing analysis & Complexes*
- *Coming soon: CRANK-2 for Experimental Phasing*
- *Coming soon: AMPLE for ab-initio MR*



CCP4
on-line



CCP4 Cloud

- *new CCP4 initiative for a comprehensive on-line crystallographic solution*



Research Complex at Harwell