

RCSB Protein Data Bank Advisory Committee Meeting

October 30th 2014



Overview

Stephen K. Burley



Leadership Transition

- Stephen K. Burley, Director
 - Overall direction of RCSB PDB as of July 2014
 - Formally responsible for Rutgers and UCSD
- Helen M. Berman, Associate Director
 - External Partnerships related to hybrid or integrative structure determination methods
 - Unchanged roles in SBKB/NDB/EMDataBank
- Philip E. Bourne—exited early in 2014
 - Associate Director for Data Science, NIH



Advisory Groups

Advisory Committees

- **RCSB PDB AC:** Cynthia Wolberger (Johns Hopkins)
- **wwPDB AC:** Soichi Wakatsuki (Stanford/SLAC)

Working Group

- **PDBx/mmCIF:** Paul Adams (LBL)

wwPDB Task Forces and Workshops

- **X-ray Validation:** Randy Read (Univ of Cambridge)
- **3DEM Validation:** Richard Henderson (MRC-LMB), Andrej Sali (UC San Francisco)
- **NMR Validation:** Gaetano Montelione (Rutgers), Michael Nilges (Institut Pasteur)
- **Small-Angle Scattering:** Jill Trehwella (Univ Sydney)
- **Hybrid Methods:** Andrej Sali (UC San Francisco), Torsten Schwede (Univ Basel), Jill Trehwella (Univ Sydney)

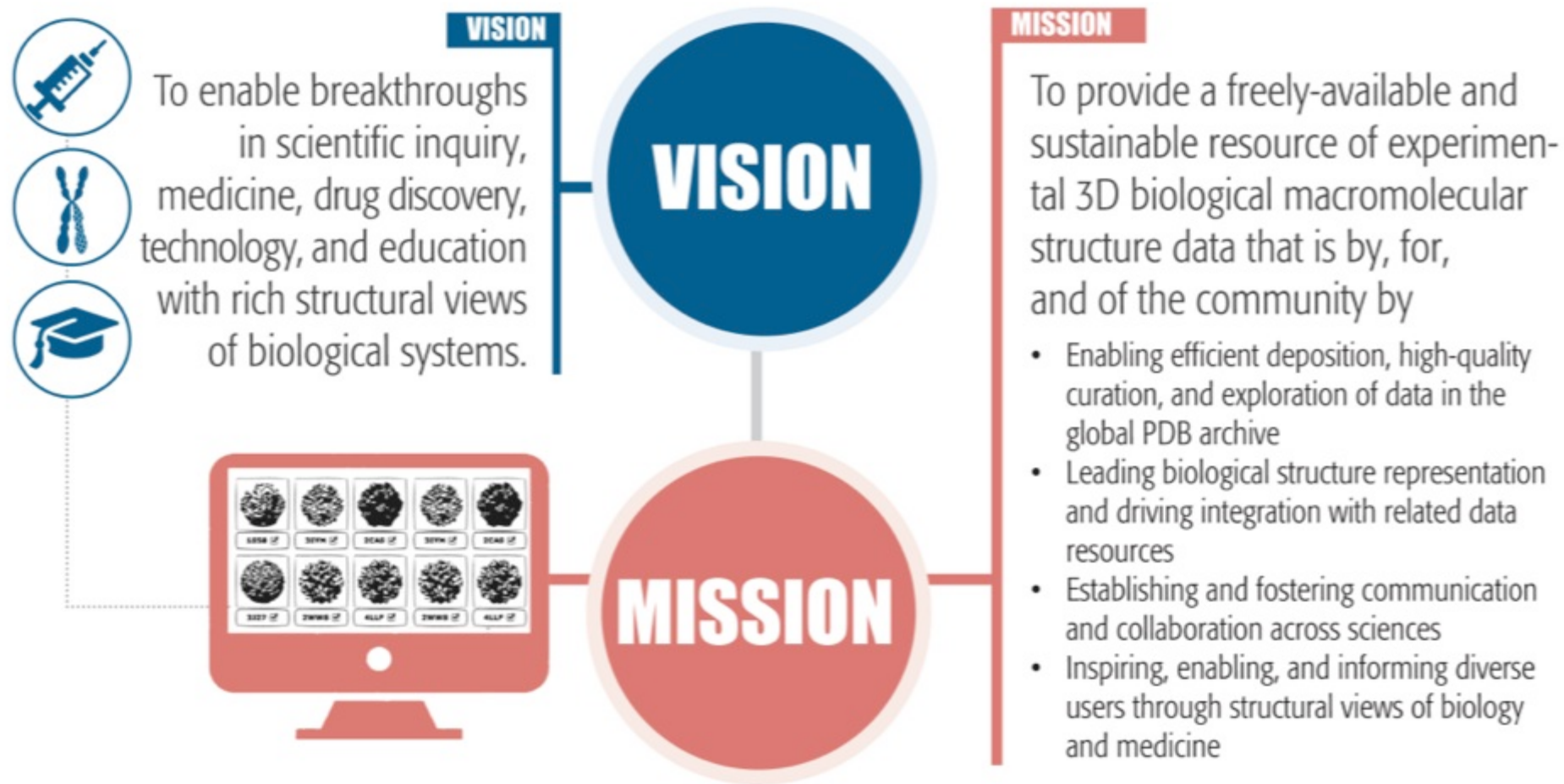


2014 Strategic Planning

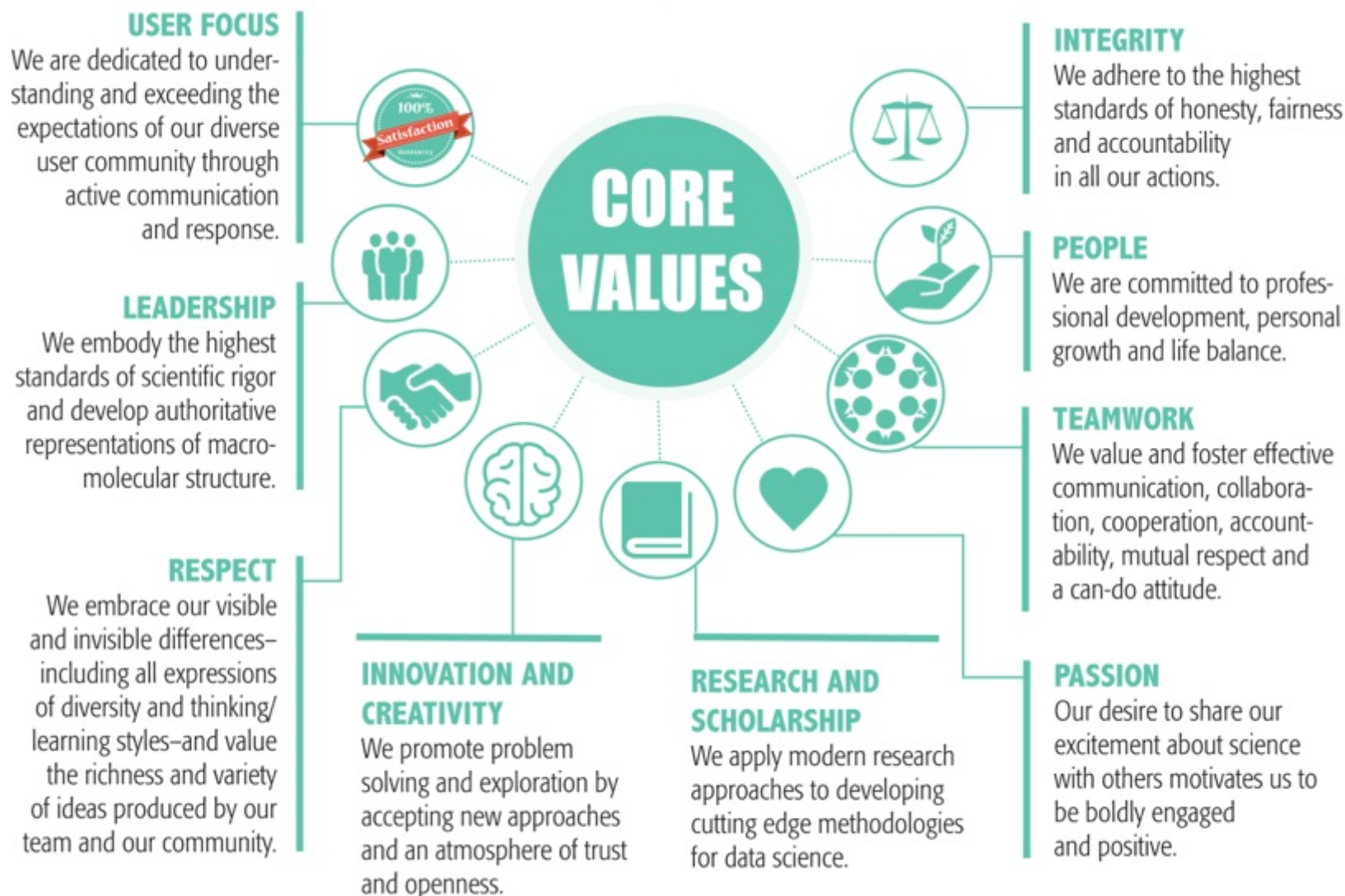
- July meeting with Leadership Team to draft Core Values, Vision, and Mission
- Reviewed broadly → additional refinements
- Informed drafting of project, team, and individual objectives



Vision & Mission



Core Values



Response to 2013 Recommendations

AC: Integration of KB and RCSB PDB

Response: **Deferred until the fate of KB determined**

AC: Review PDB-101 materials for effectiveness, and outreach efforts for cost effectiveness

Response: **Fall 2014 survey informing redesign efforts**

AC: Development of online courses with external funding

Response: **Submitted proposal under review**

AC: Major redesign of **rccb.org** website

Response: **In progress (Phase 1); Addressed in Data Out**

AC: Suppression of structure title in unreleased entries

Response: **Implemented as explicit depositor option in the new system**



PDB Depositors

~850 new entries/month



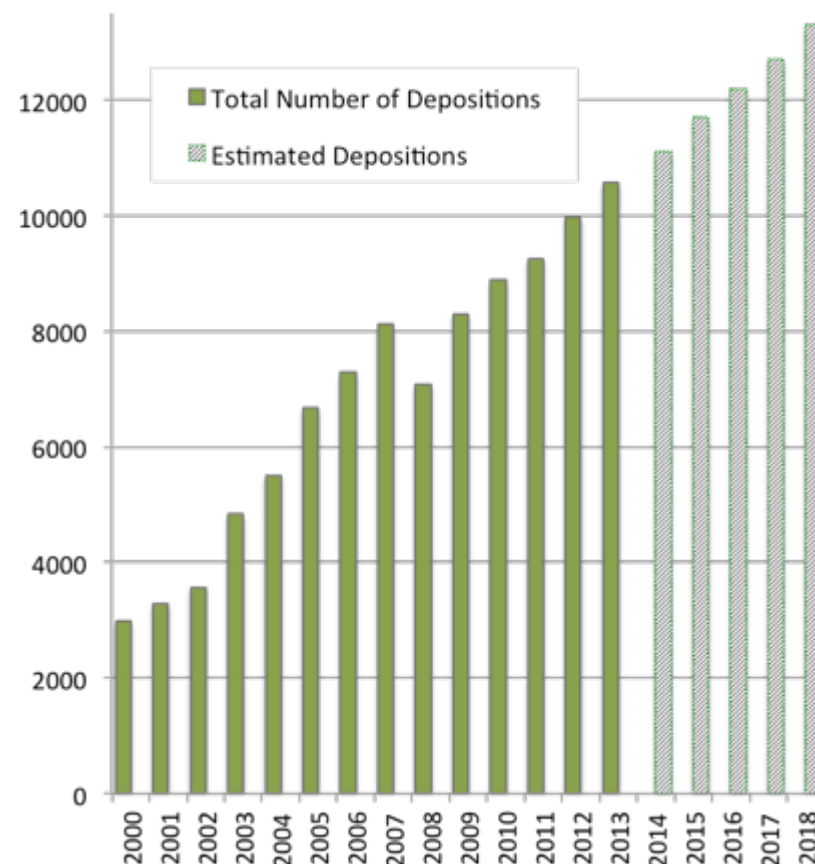
PDB User Access

2013 FTP and RSYNC Download Traffic
433 million downloads



■ RCSB PDB: 312 million ■ PDBe: 81 million ■ PDBj: 40 million

Growing Number of PDB Depositions



2013: 10566 depositions
49% increase in the number of
global depositions since 2008



Ensuring Data Are Freely and Globally Available

- Partners
 - RCSB PDB (Research Collaboratory for Structural Bioinformatics, Rutgers/UCSD)
 - PDBe (Osaka University)
 - PDBe (EMBL-EBI)
 - BioMagResBank (University of Wisconsin, Madison)
- Collaborate on the guiding policies for unified data processing (Data In) and archive management
- Each partner website offers diverse services and views of the data (Data Out)



wwPDB Milestones; September 2013-October 2014

- Archive Growth: 100,000 structure milestone
- PDBx implemented in X-ray software packages
- Large structures released as “non-split” PDBx files
- Stand-alone X-ray validation server in production
- Validation reports released for all X-ray entries
- Common Deposition & Annotation System (D&A) in production for X-ray (NMR/EM pending)
- Workshops and Meetings
- Funding (stable for now, but slightly ↓)



wwPDB AC Closeout; October 10, 2014 EBI, Hinxton, UK

- Overall: continuing growth of archiving, significant progress on X-ray D&A, strongly recommend switching to X-ray D&A ASAP, implementation of PDBx very positive, continuous improvements
- Advisors encourage the wwPDB to meet the projected timelines of NMR & EM D&A
- Collaboration of the 4 PDB centers continues to be excellent
- Recommend stronger representation of EM as part of wwPDB given the increased importance of the cryoEM field, in particular higher resolution (ca. 3Å) structures using single particle imaging method
- RCSB PDB Leadership Transition: effective, well planned and executed
- Achieved the global workload balance owing to D&A implementation
- Funding efforts
 - RCSB: non-competitive NSF funding (2014 - 2018)
 - PDBe: Wellcome Trust and others
 - PDBj: Oct 2013 review, FY2014 new scheme due to started
 - BMRB: NIGMS R01 funding started for 5 years (2014 - 2019)



RCSB PDB Milestones; September 2013-October 2014

- Improved site usability and new structure annotations
 - Integration of membrane protein annotation
 - Mapping PDB structures to human genes
 - Beta site of redesigned home page
 - Mobile web support and visualization
 - Infrastructure and process improvements



RCSB PDB Milestones; September 2013-October 2014

- IYCr calendar and events
- Return of *Art of Science* exhibit
- Video Challenge
- High school curriculum development
- World AIDS Day Symposium
- Targeted outreach



Agenda

Overview & Management

Stephen Burley

Outreach & Education

Christine Zardecki
Shuchismita Dutta

Data In

Martha Quesada
John Westbrook
Jasmine Young

Data Out

Peter Rose
Andreas Prlić



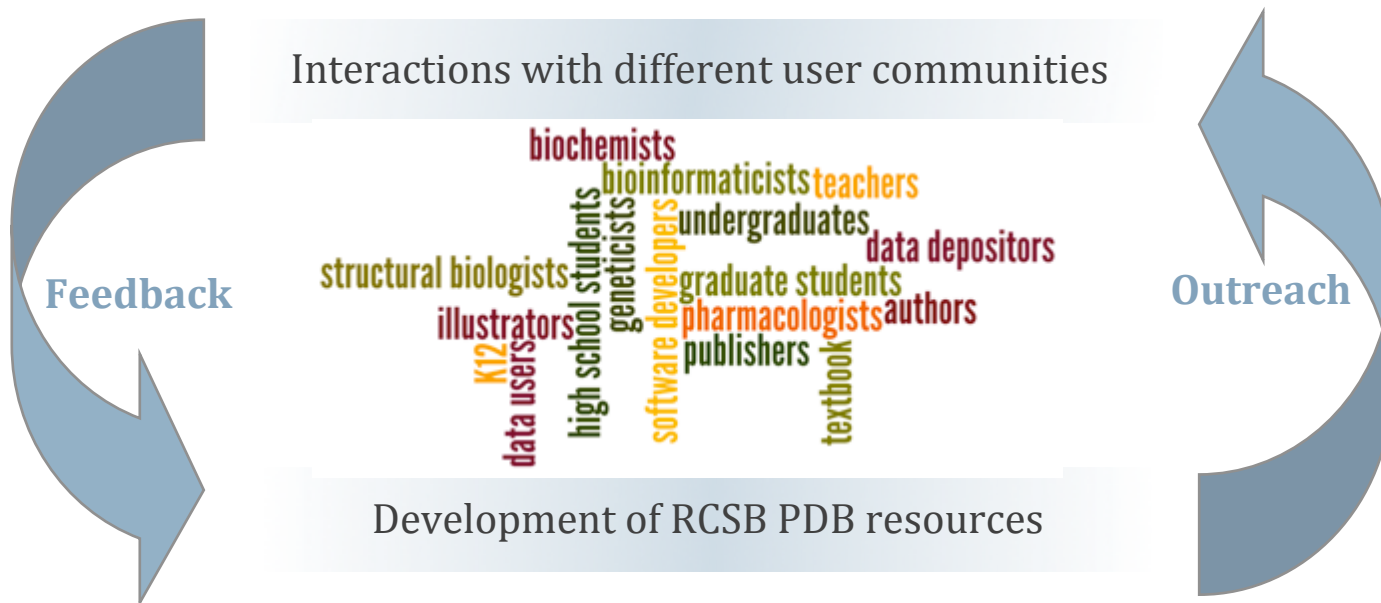
Outreach and Education

Christine Zardecki

Shuchismita Dutta

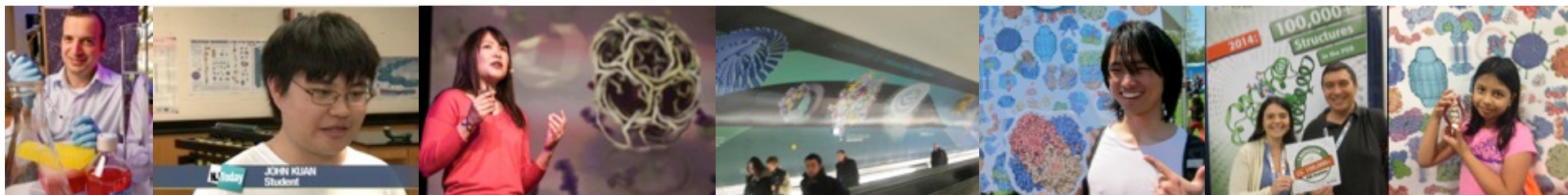
Goals

- To facilitate RCSB PDB Vision & Mission in the interests of science, medicine, and education
- To inspire, enable, and inform diverse users through structural views of biology and medicine



Diverse User Communities

Who are our users?	What are they using?	How do we know?
Biologists: structural biology, biophysics, biochemistry, genetics, Immunology, pharmacology, cell and molecular biology, ...	RCSB PDB website, deposition tools, data	Publication requests, website usage, annotator and Help Desk requests, community outreach, surveys
Other scientists: bioinformatics, software developers, ...	Web Services, search engines, data	Publication requests, website usage, Help Desk requests, community outreach, surveys
Students & teachers	PDB-101	Increase in web hits, email, meeting interactions, specialized workshops/events
Media: Writers, textbook authors, patient advocacy groups, ...	Images, data, information, outreach material, e.g., posters	Publications, image requests
General public: Curious/interested individuals, artists, sculptors, ...	Images, <i>Molecule of the Month</i> , external media	Concerts, media, Wikipedia



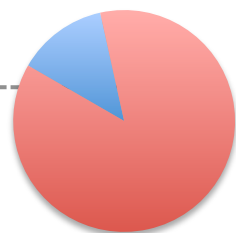
2014 Survey: Who Uses PDB-101?

Based on
~700
Responses

Demographics

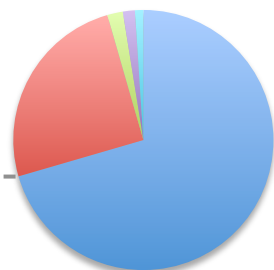


Ethnicity



- Hispanic or Latino (13%)
- Other (87%)

Race



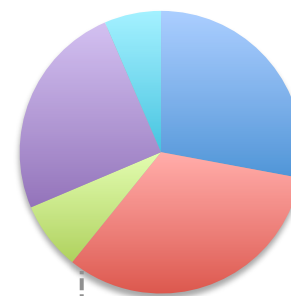
- White (72%)
- Asian (25%)
- American Indian or Alaska Native (2%)
- Black or African American (<2%)
- Native Hawaiian or Other Pacific Islander (1%)

83% have
never
deposited a
structure

Where do they work?

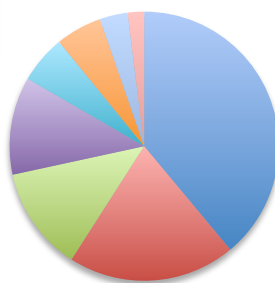
- College/University (74%)
- Research Institute (12%)
- Government (2%)
- Pharma/drug discovery/biotech (3%)
- K12 (8%)

Breakdown



- Undergrad (28%)
- Graduate (32%)
- Postdoc (8%)
- Faculty (25%)
- Staff (6%)

Research Area



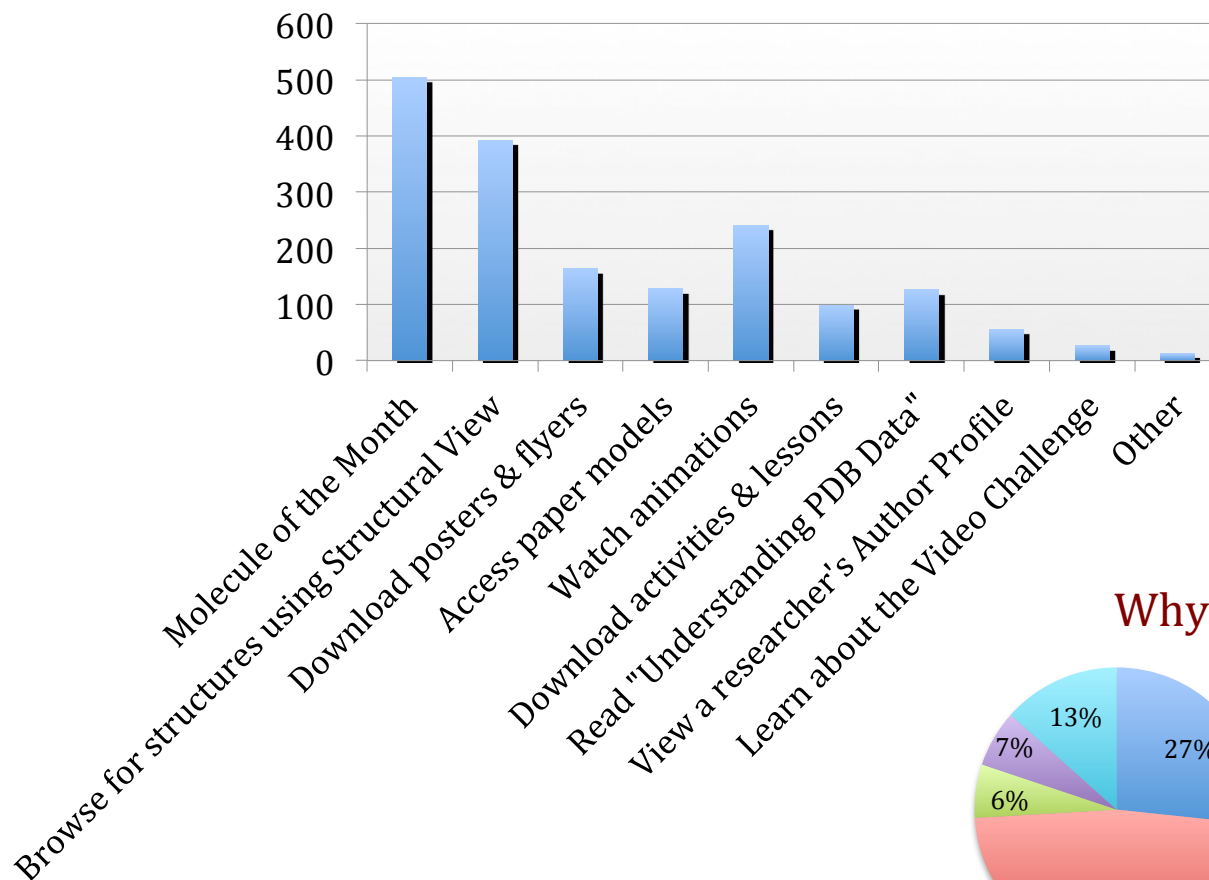
- Life Sciences (39%)
- Chemistry (20%)
- Medical Sciences (12%)
- Computational Sciences (12%)
- Pharmacology (6%)
- Other (6%)
- Physics (3%)
- Math/Statistics (2%)

College/ University Type

- University (409)
- 4-year college (56)
- Historically Black College/University (3)
- Women's college (2)

What Do PDB-101 Users Do?

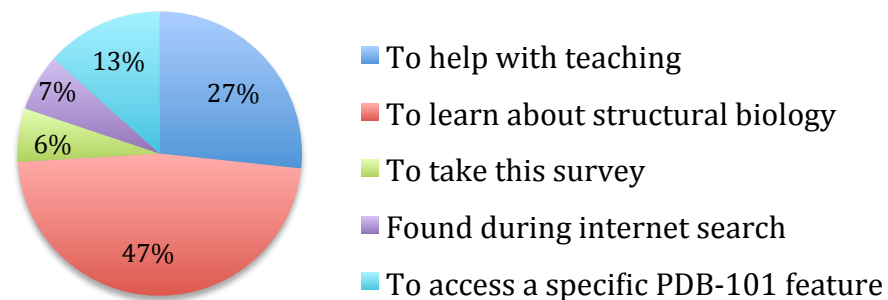
Why do they use PDB-101?



Users visit...

PDB-101 (2014)	rcsb.org (2012)
Daily (6%)	Daily (17%)
Weekly (27%)	Weekly (38%)
Monthly (20%)	Monthly (16%)
Occasionally (39%)	Occasionally (22%)
Never (8%)	Never (7%)

Why did they visit PDB-101 today?



Outreach Activities: Recent Highlights

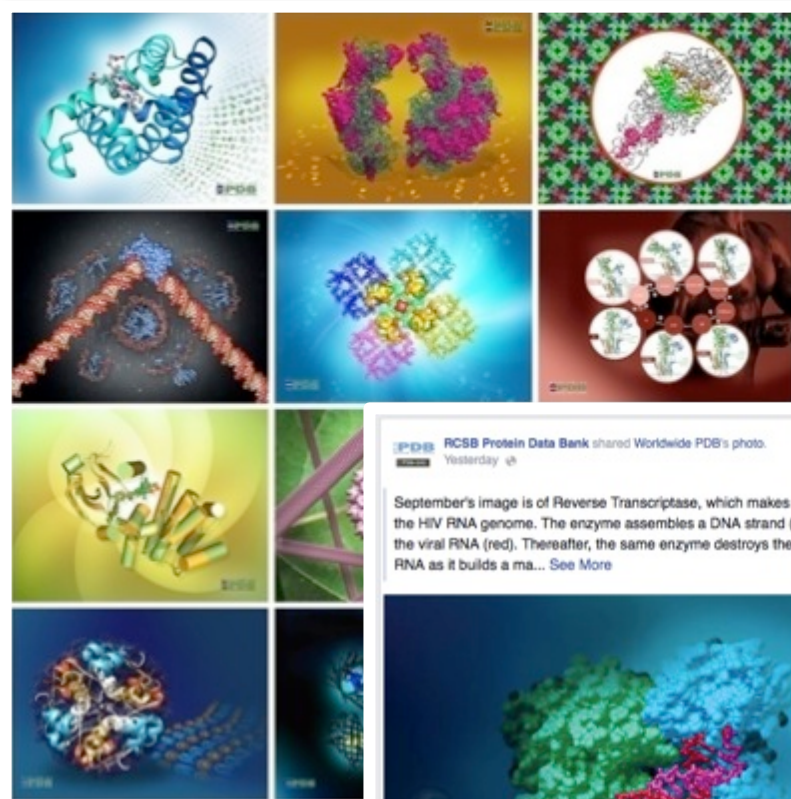


International Year of Crystallography



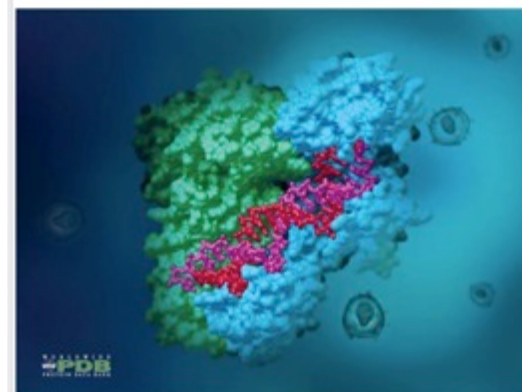
~5500 wwPDB calendars were distributed to classrooms and at scientific and educational meetings worldwide, including the IYCr Opening Ceremony, and promoted throughout the year.

Calendar images are available for use without restrictions at wwpdb.org



RCSB Protein Data Bank shared Worldwide PDB's photo.
Yesterday · 48

September's image is of Reverse Transcriptase, which makes a DNA copy of the HIV RNA genome. The enzyme assembles a DNA strand (magenta) from the viral RNA (red). Thereafter, the same enzyme destroys the original viral RNA as it builds a ma... See More



Like · Comment · Share

Yael Benhar, Takahiro Kudou, Zarbara Babel and 4 others like this.



IYCr and Science Festivals

USA Science & Engineering Festival (April 25-27, Washington DC)

- Exhibit booth
 - Photo booth
 - Materials from IUCr/USNCCr
 - RCSB PDB materials
- ~100 visitors per hour
 - Mostly pre-high school students and their parents
 - Built ~600 DNA structures and close to 1000 viruses

ACA Education Workshop (May 24, Albuquerque, NM)

- Virus/DNA model building in the New Mexico Museum of Natural History & Science
- Parallel with teacher workshop, crystallization station, and public lecture by Dr. Doris Schattschneider on MC Escher



IYCr and *Art of Science*

- Paris, France in May 2014
 - Villejuif Bio Park and Sup'Biotech Open House
- Small ACA grant for IYCr-related outreach will support show sent to
 - Applications and Research Laboratory High School (Ellicott City, MD), to be scheduled
 - McPherson College (McPherson, KS), November
 - Delbarton School (Morristown, NJ), November 2014 – January 2015
- Spring 2015: Rutgers Art History Library



Outreach to Diverse Communities

Annual Biomedical Research Conference for Minority Students (ABRCMS)

November 13-16, 2013, Nashville, TN

- Exhibit booth visited by ~ 250 students and administrators
 - Distributed materials, including ~200 IYCr calendars
 - Many very familiar with PDB structures
 - Students enjoyed the photobooth, and voted for GFP as their favorite protein (pictures on Facebook), the DNA paper model, and the Krebs Cycle flyer

Society for Advancement of Hispanics/Chicanos and Native Americans in Science (SACNAS)

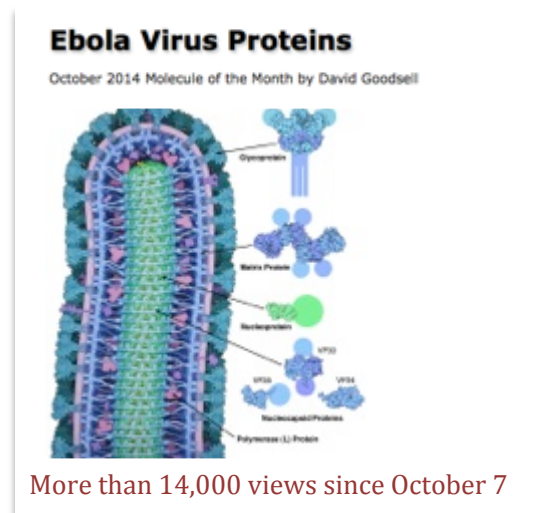
October 16-18, 2014, Los Angeles, CA

- PDB-related symposium with HMB speaking attended by ~130



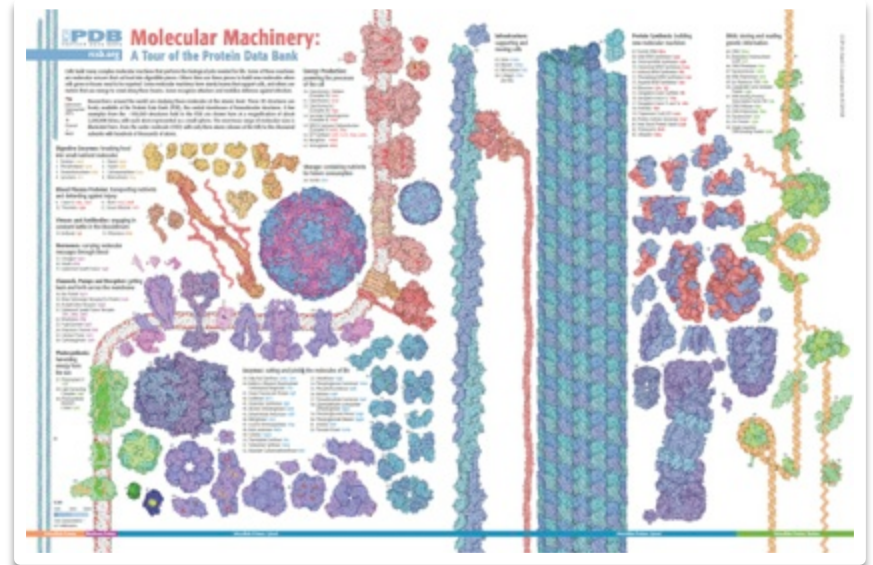
PDB-101: *Molecule of the Month*

- Popular entry point to PDB-101
- Topics are now strategically planned to correspond with related RCSB PDB activities, topical events
- Incorporated with social media efforts
- Frequently used by textbooks and other media
- Increased integration with educational activities



PDB-101: *Molecular Machinery*

- Original poster
 - Released in 2002
 - Printed twice
 - Distributed thousands
 - Consistently requested
- 2014
 - Released when we hit 100K structures in May
 - Downloaded ~1000 times already
 - Interactive version will link to *Molecule of the Month* and more



Recent Publications

- RCSB PDB *Mobile*: iOS and Android mobile apps to provide data access and visualization to the RCSB Protein Data Bank (2014) *Bioinformatics* doi: 10.1093/bioinformatics/btu596
- The Protein Data Bank archive as an open data resource (2014) *Journal of Computer-Aided Molecular Design* doi:10.1007/s10822-014-9770-y
- 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
- 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
- Establishing the Next Generation of the Protein Data Bank (2014) *The Winnower* doi: 10.15200/winn.140076.68556
- Chemical annotation of small and peptide-like molecules at the Protein Data Bank (2013) *Database* doi:10.1093/database/bat079

Original article

Chemical annotation of small and peptide-like molecules at the Protein Data Bank

Jasmine Y. Young¹, Zukang Feng¹, Dimitris Dimitropoulos², Raul Sala¹, John Westbrook¹, Marina Zhuravleva¹, Chenghua Shao¹, Martha Quesada¹, Ezra Peisach¹ and Helen M. Berman^{1,*}

Improving the Representation of Peptide-Like Inhibitor and Antibiotic Molecules in the Protein Data Bank

Shuchismita Dutta,¹ Dimitris Dimitropoulos,² Zukang Feng,¹ Irina Persikova,¹ Sanchayita Sen,³ Chenghua Shao,¹ John Westbrook,¹ Jasmine Young,¹ Marina A. Zhuravleva,¹ Gerard J. Kleywegt,³ Helen M. Berman¹

J Comput Aided Mol Des
DOI 10.1007/s10822-014-9770-y

The Protein Data Bank archive as an open data resource

Helen M. Berman · Gerard J. Kleywegt ·
Haruki Nakamura · John L. Markley

Bioinformatics Advance Access published September 2, 2014

Structural bioinformatics

RCSB PDB *Mobile*: iOS and Android mobile apps to provide data access and visualization to the RCSB Protein Data Bank

Gregory B Quinn¹, Chunxiao Bi¹, Cole H Christie¹, Kyle Pang¹, Andreas Prlić¹, Takanori Nakane², Christine Zardecki³, Maria Voigt³, Helen M Berman³, Philip E Bourne^{1,4,5}, and Peter W Rose^{1,*}

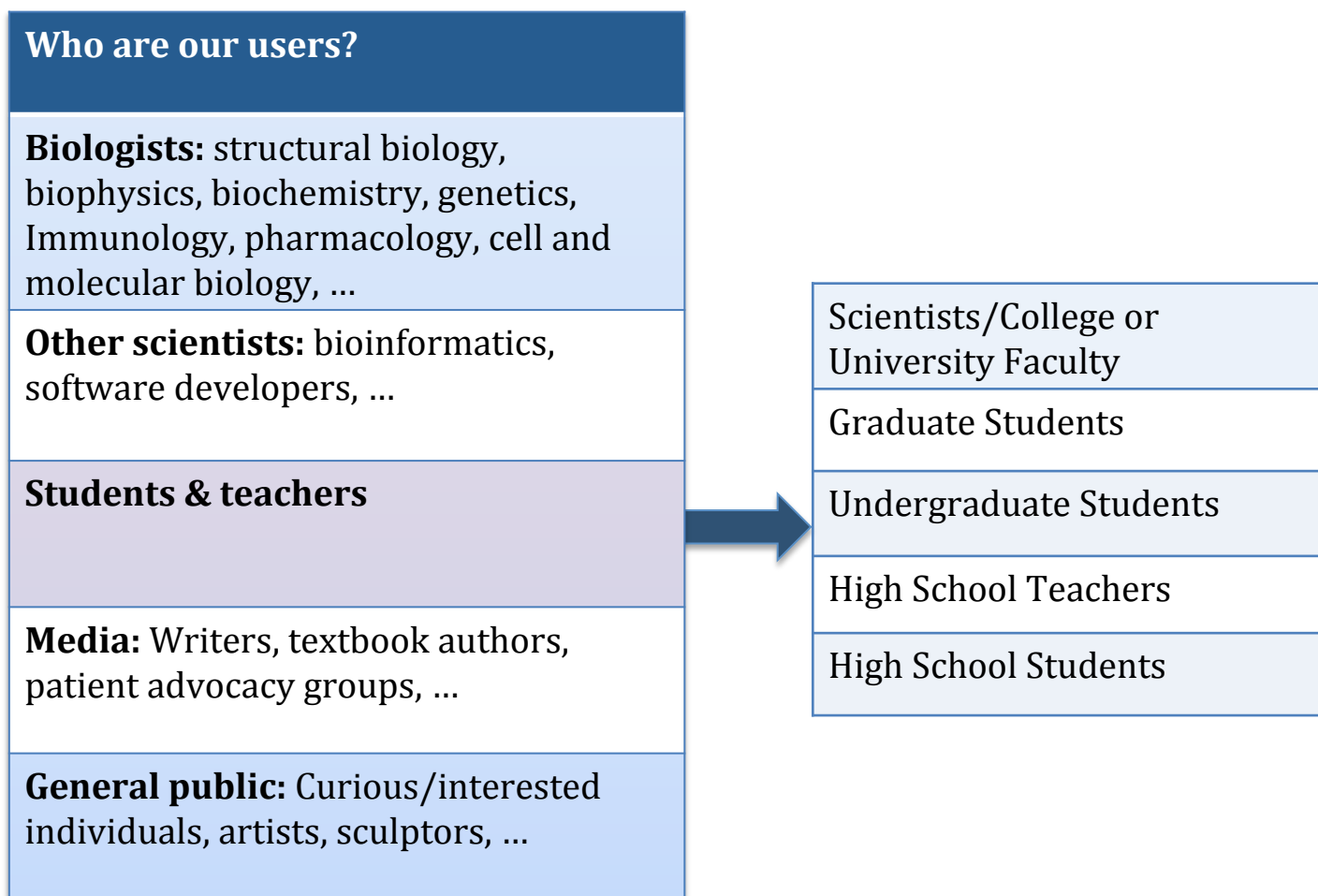


Educational Experiences

UG-research
Collaborative-seminar-courses wk-to-g-teacher-wkshop
HS-Science-Olympiad Scientist-workshop Grad-courses
MOOC HS-activity UG-seminar HS-interns
UG-interns



Training For Specific Audiences



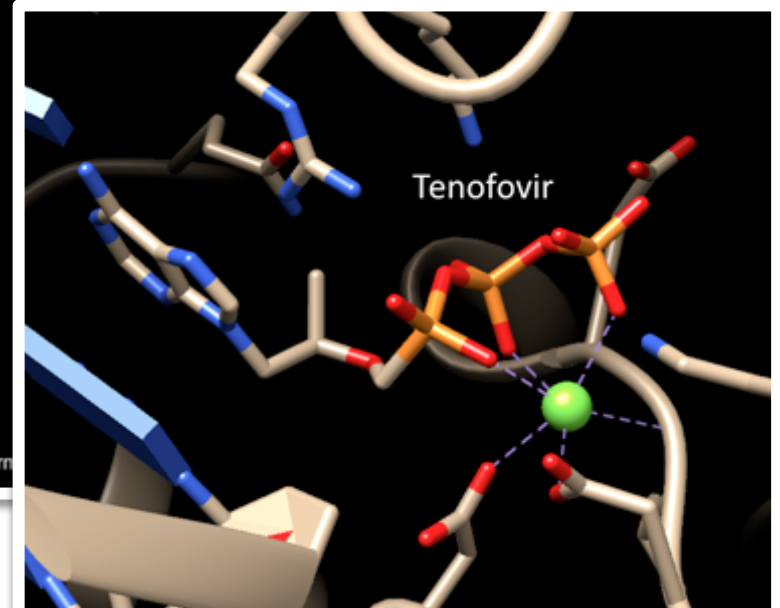
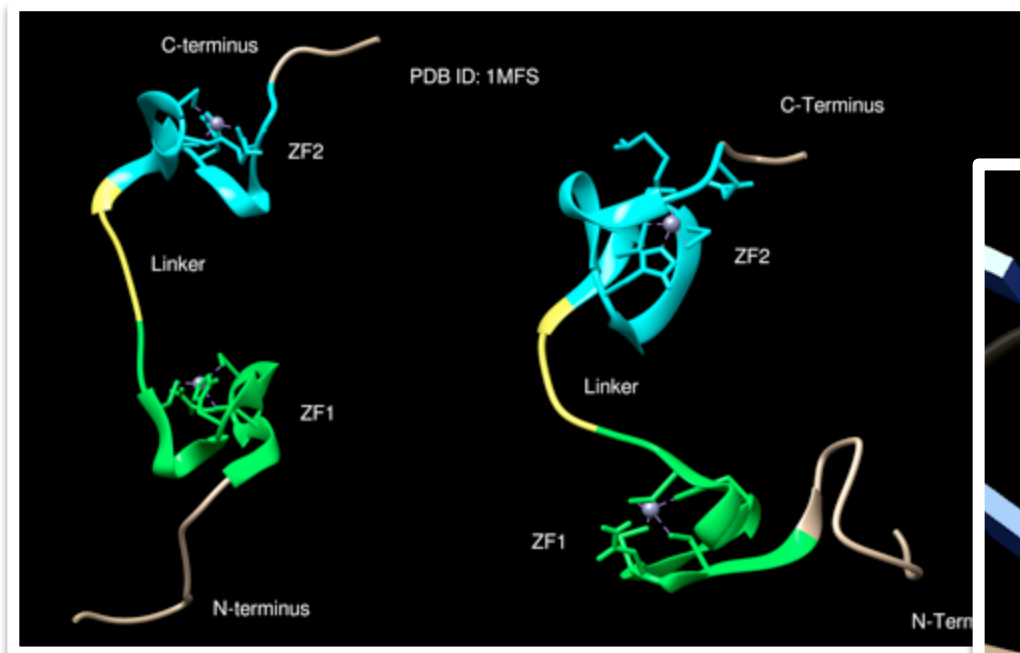
Interdisciplinary Boot Camp: Training Biologists, Physicists, and more

- What?
 - 2 weeks in January 2014
 - Focused on Fluorescent Proteins
 - Exposure to experimental and computational approaches
- Who?
 - >100 registrants
 - Students, Postdocs, Faculty
- Outcome
 - Very well received!
 - SAS Investment!
 - Repeat Boot Camp in January 2015
 - Nucleic Acids focused Boot Camp planned for 2016
 - Alternation thereafter



Undergraduate Training

- MAP Course → Structural View of Biology
- Theme – HIV/AIDS



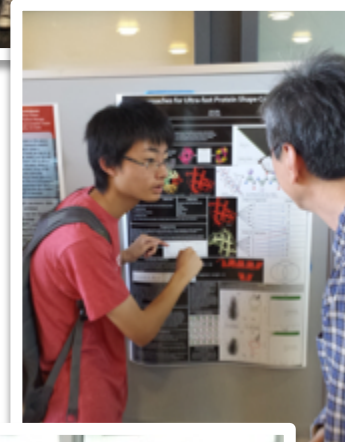
From student reports, Rutgers Univ., 2014



Research Projects

At UCSD

- **SDSC 2014 Research Experience for High School Students**
 - Ezra Kosviner (Computational Model of Naeglariapore A)
 - Kevin Wu (Fast structure alignments)
- **UCSD 2014 STARS Program**
 - Henry Truong (Multiscale rendering of large assemblies)
- **UCSD CSE 199 Special Studies Course**
 - Varun Sharma (Visualization with WebGL)
- **UCSD Undergraduate Research (Fast structure alignments)**
 - Alan Yeung, Albert Chang, Arvind Rao
- **Volunteer from SDSU**
 - Roshni Bhattacharya (Analyzing genetic variation data in PDB)



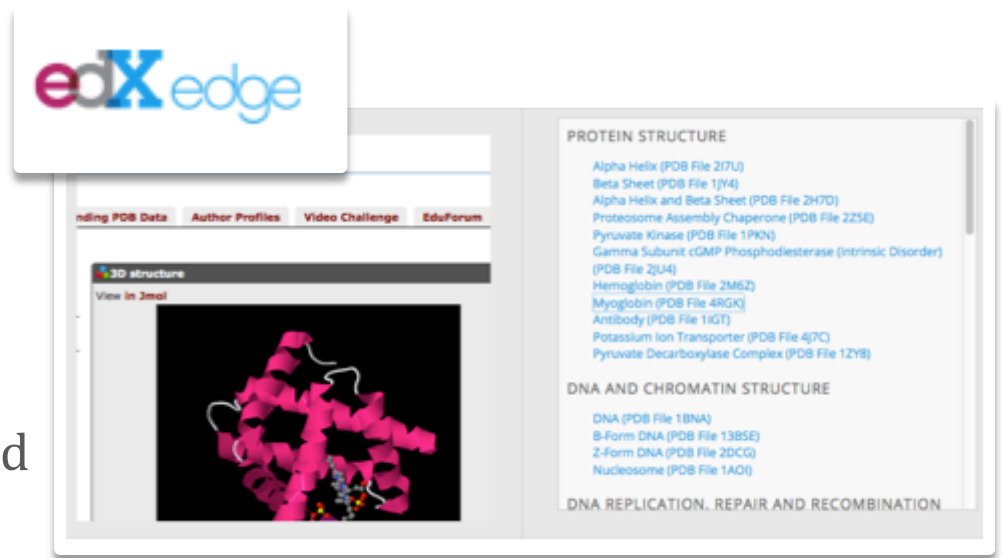
At RU

- **Independent study**
 - Zaeem Billah (Developing a Disease view: HIV/AIDS)
 - Lauren Benzinger (Developing a Disease view: Diabetes)
- **Volunteer from Rutgers**
 - Alex Nazarro (Preparing for Curriculum Development Program)



e-Learning

- Massive Open Online Courses (MOOCs)
 - BIOC 300 (June 2014 ROC) Paradigms in Biochemistry and Cell Biology
 - Upcoming courses starting in January 2015
 - Collaborating to include questions/exercises so students can use various RCSB PDB resources
- Develop materials and lesson plans introducing fundamentals of structural biology for use in
 - RU courses (e.g. MAP course 2015, Quantitative Biology Boot Camp)
 - HS classes (e.g. Collaborative Curriculum Development Program on HIV/AIDS)



PDB on edX, Rice University, 2014



Kathleen Matthews
Stewart Memorial Professor
of Biochemistry & Cell Biology
Rice University



Reaching High School Students

Goal:

- Promoting awareness about biomolecular structures

Approach:

- Teach teachers
 - Specialized workshops
 - Year-long programs
- Teachers teach students
- Challenge students to assess learning
 - Video challenge



National HS Teacher Workshop,
Princeton Univ., 2014



HS Teacher Workshop,
Colorado, 2014



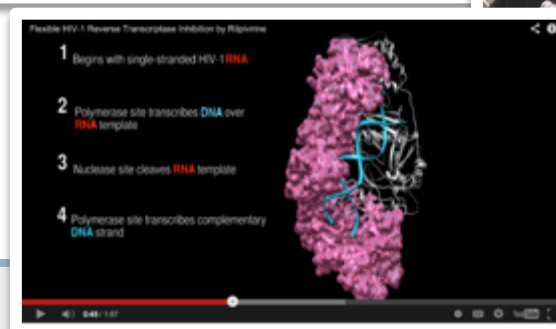
Working Together to Visualize:

A Molecular Structural View of HIV/AIDS

2013-2014

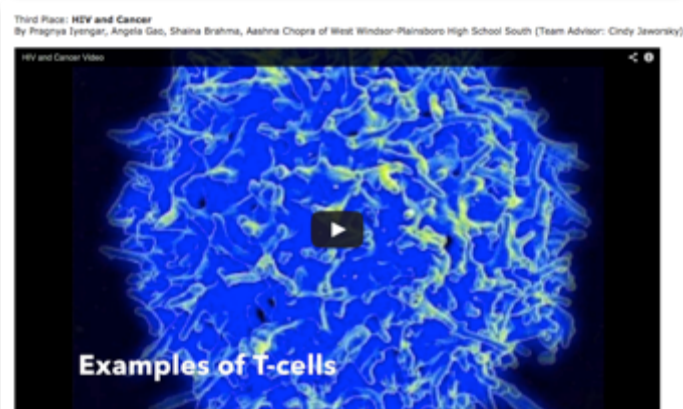
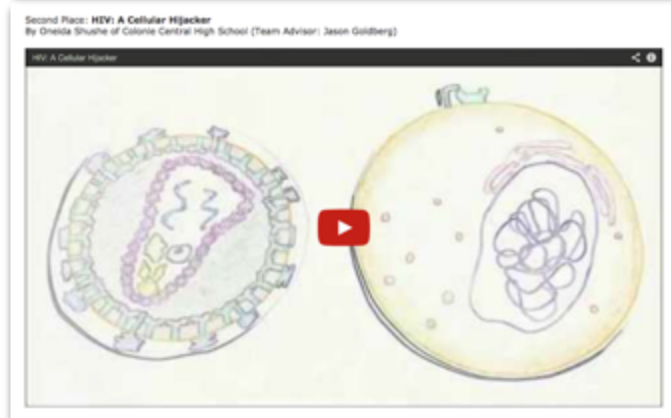
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2013									Announce Program	Teacher workshop/ webinar	World AIDS Day 2014	HS student learning
2014		Video Challenge				Select winners						
	HS student learning											

- Teacher workshops – technical training
- Public symposia – theme based learning
- Video challenge – Assessment



HS Video Challenge 2014

- **Challenge:**
 - Create short videos (two minutes or less) that tell a story about any topic related to structural biology of HIV/AIDS
 - Must include a visualization/animation of two or more PDB structures
- **Resources:**
 - Collection of MOM features
 - Video tutorials
 - Resources for making videos
- **Entries:**
 - 23 Videos
 - 7 schools/4 states
 - First place viewed ~500 times
- **Judges:**
 - 8 (scientists, clinicians, educators or scientific animators)
 - General public
 - AIDS Education Training Center National Resource Center
- **Prizes:**
 - Judges' Award
 - Viewers' Choice Award
 - Service to the Community Award



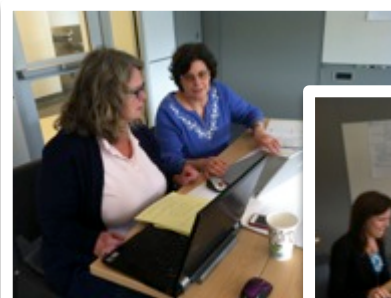
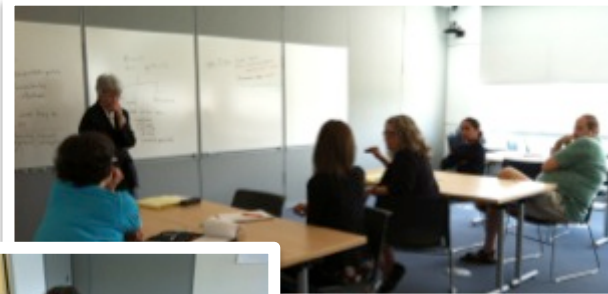
Lessons Learned

Teacher Workshops	Future Plans
Teachers need prepared lesson plans	Collaborative Curriculum Development Program
The discussion on HIV/AIDS needs context e.g. learning about biological molecules, Immune system, infections, treatments, practical awareness	
World AIDS Day Symposium	Future Plans
Patient experiences are valuable	World AIDS Day Symposium December 5 th 2014
General audiences and patients want to learn more	
HS Video Challenge	Future Plans
Repeat same theme in 2015	HS Video Challenge on structural view of HIV/AIDS Spring 2015
May include more focused topics and suggested structures	



Targeting HS Teachers and Students

- Collaborative Curriculum Development Program (2014): on HIV/AIDS
 - 5 HS teachers from NJ HS + educators, scientists, clinicians
 - Developed Curricula
 - General Bio/Honors Bio
 - Advanced Placement Bio
- Presented to teachers (@ NJSC 2014)
- Testing and Participation in Video Challenge (2014-2015)

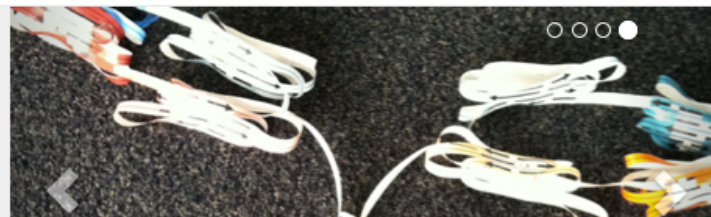


	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2014							Visualize HIV curriculum development			Announce Program		
2015		HS Video Challenge										
	Student learning & survey				Survey & feedback analysis		Fine tuning Visualize HIV curriculum					



RCSB PDB Curricula

The RCSB PDB Curricula were developed through collaborations and participation of scientists, curriculum design experts, educators, clinicians and local teachers to



● RCSB PDB Resource ● Collaborative Curriculum Development Program Resource ● External Resource

Topic	Contents	For General Biology	For AP Biology
Pre-exposure to Curriculum	Objective	Understand students' prior knowledge about biological molecules, immune system, HIV/AIDS	Understand students' prior knowledge about biological molecules, immune system, HIV/AIDS
	Materials	Pre-test question bank with answers ●	Pre-test question bank with answers ●
	Instructions	Suggestions for administering Pre-test	Suggestions for administering Pre-test
Review of Biological Molecules	Learning Objectives	1. Bonding in biological molecules 2. Structure Function Relationships 3. Modeling	1. Bonding in biological molecules 2. Structure Function Relationships 3. Modeling 4. Visualization and analysis
	Slides and notes	Introduction to Biomolecular structures and the Protein Data Bank Download Slides ●	Introduction to Biomolecular structures and the Protein Data Bank Download Slides ●
	Activities	1. Vocabulary map ● - individual or group activity 2. Exploring Insulin - individual activity using Jmol and resources found on the RCSB PDB website 3. Antibody paper model - hands-on group activity Antibody model - Gen Bio level discussion questions	1. Vocabulary map ● - individual or group activity 2. Exploring and visualization of the Antibody structure - individual activity using UCSF Chimera and resources found on the RCSB PDB website 3. Antibody paper model - hands-on group activity Antibody model - AP Bio level discussion questions
	Instructions	Suggestions for how to run the activities provided here	Suggestions for how to run the activities provided here



Science Olympiad: Protein Modeling

- Students build 3D models, answer questions on structure and function
- MSOE CBM - RCSB PDB collaboration
- 2015 Topic: Genome Editing
- Funds:
 - 2015: (NIH) SEPA, SEDAPA, 3DMD and Vertex Pharma
 - 2016: NIDA and 3DMD



Next SO Nationals at University
of Nebraska, Lincoln, NE
in May 2015

Science Olympiad
Leadership meeting,
Orlando FL, May 2014



Educational Experiences: Future

- Targeting High School teachers and students
 - Collaborative Curriculum on HIV/AIDS
 - Testing and improving the curriculum
 - HS Video Challenge 2015 - on HIV/AIDS
 - Reaching out nationwide
 - Science Olympiad 2015 - Genome Editing
- Designing course materials/Question Banks for
 - MOOCs - collaborating with other faculty and scientists
 - Educators and students in a range of institutions
 - Rutgers Boot Camp
- Facilitating a Structural View of Biology for all
 - Preparing for next theme focus (Diabetes)
 - Preparing for 2015 Spring MAP course



Data Deposition and Annotation

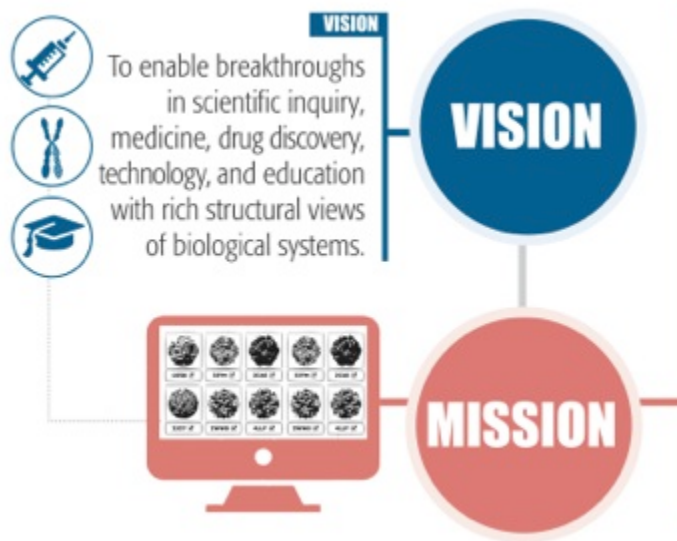
Martha Quesada

John Westbrook

Jasmine Young



Vision, Mission, and Goals



- Enable efficient deposition, high-quality curation and exploration of data in the PDB archive
- Lead biological structure representation and driving integration with related data resources
- Establishing and fostering communication and collaboration across sciences
- Inspiring, enabling and informing diverse users through structural views of biology and medicine

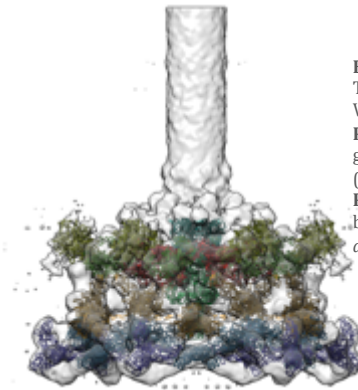


Common Deposition & Annotation System (D&A)

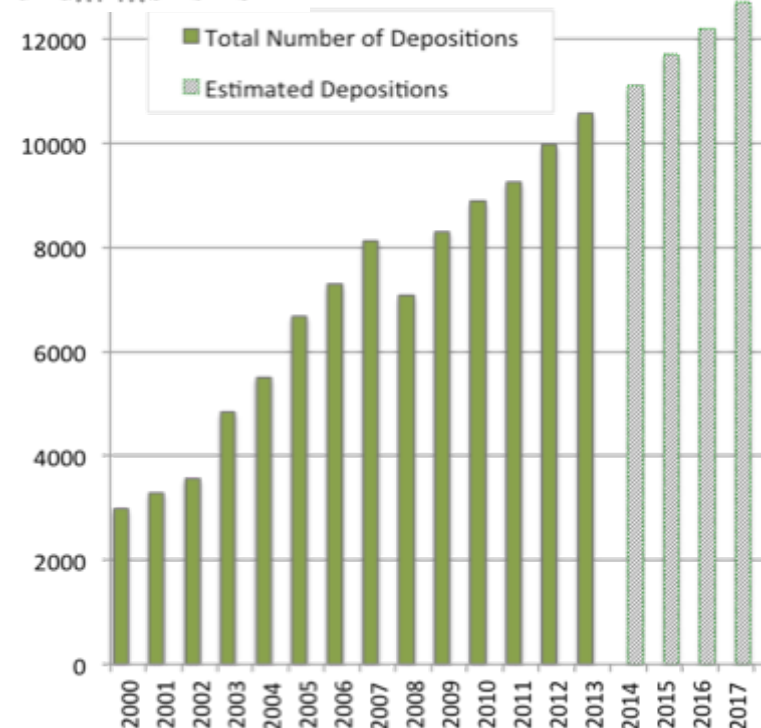
Project goal: Create the next generation deposition and annotation system supporting the PDB Vision

Strategy:

- Maximize data quality
 - Improve data validation
 - Standardize file formats
 - Ensure more complete data capture
- Support larger and more complex structures
- Improve efficiency
 - Automation and validation of routine tasks
 - Simplification of global process

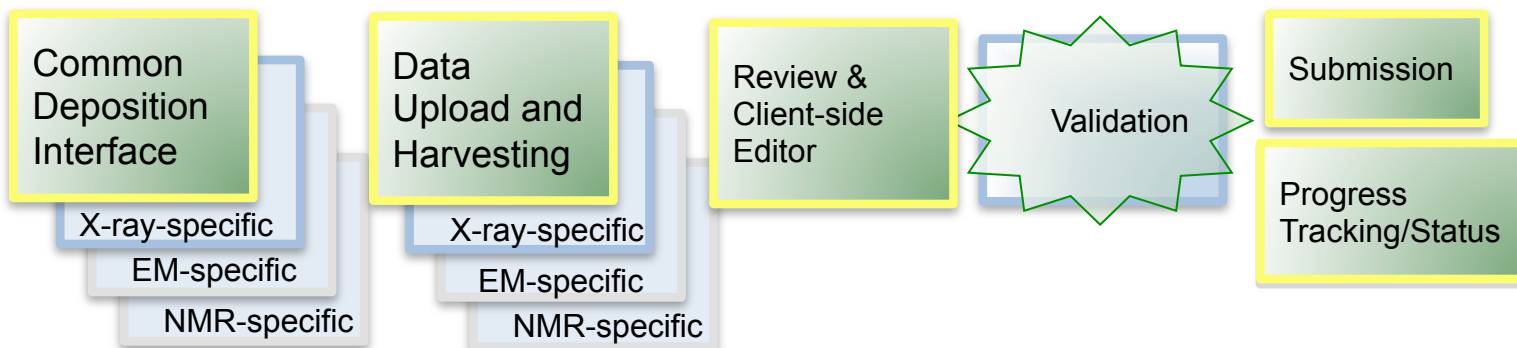


EMD-1048; PDB IDs 1pdf, 1pdi, 1pdj, 1pdl, 1pdm, 1pdp:
Three-dimensional structure of bacteriophage T4 baseplate.
VA Kostyuchenko *et al. Nat. Struct. Biol.* 10, 688-93 (2003);
PDB ID 2fl8: Evolution of bacteriophage tails: Structure of T4
gene product 10. PG Leiman *et al. J. Mol. Biol.* 358, 912-21
(2006);
PDB ID 3h3w: The structure of gene product 6 of
bacteriophage T4, the hinge-pin of the baseplate. AA Aksyuk *et al. Structure* 17, 800-808 (2009)



wwPDB Common D&A Pipeline

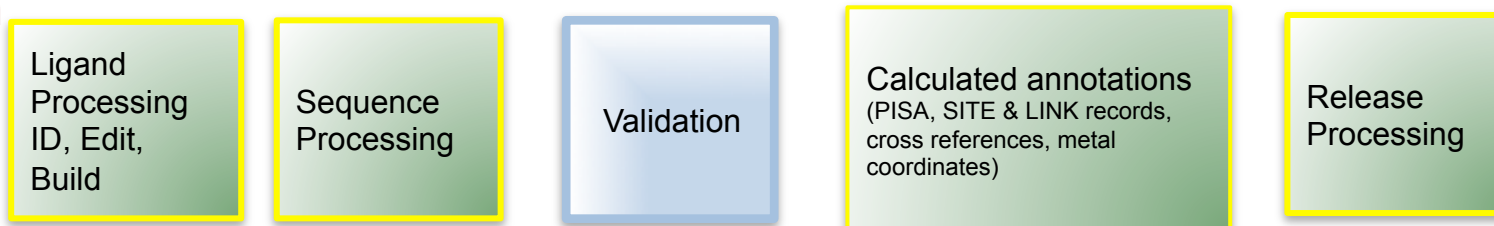
Deposition Pipeline



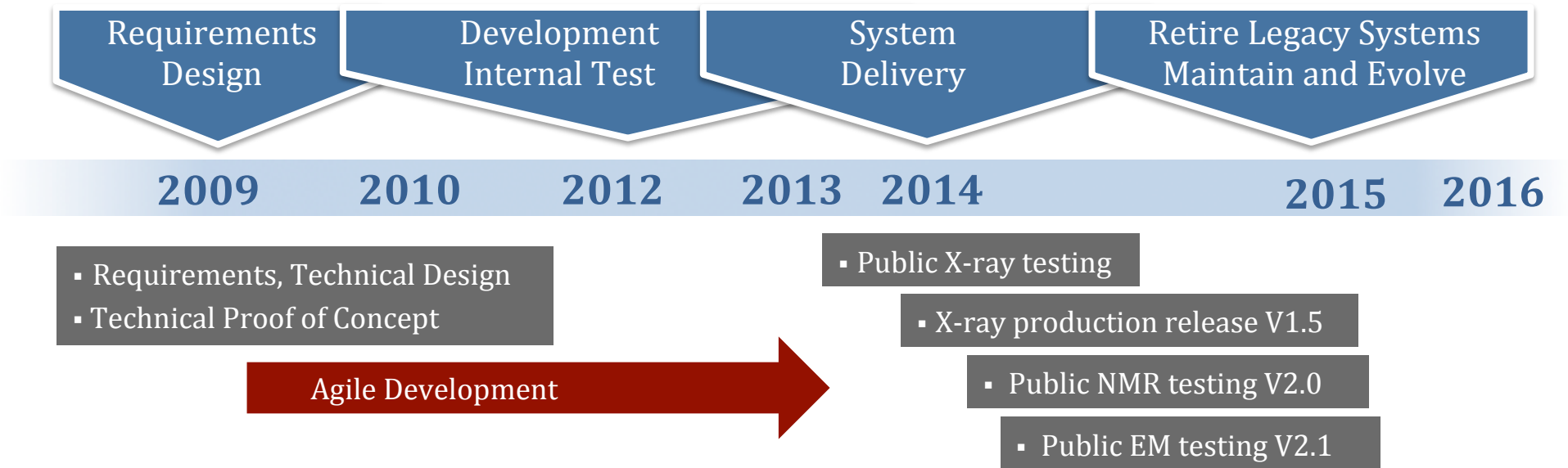
Communication System

Workflow-Automation System

Annotation Pipeline



wwPDB Common D&A Project

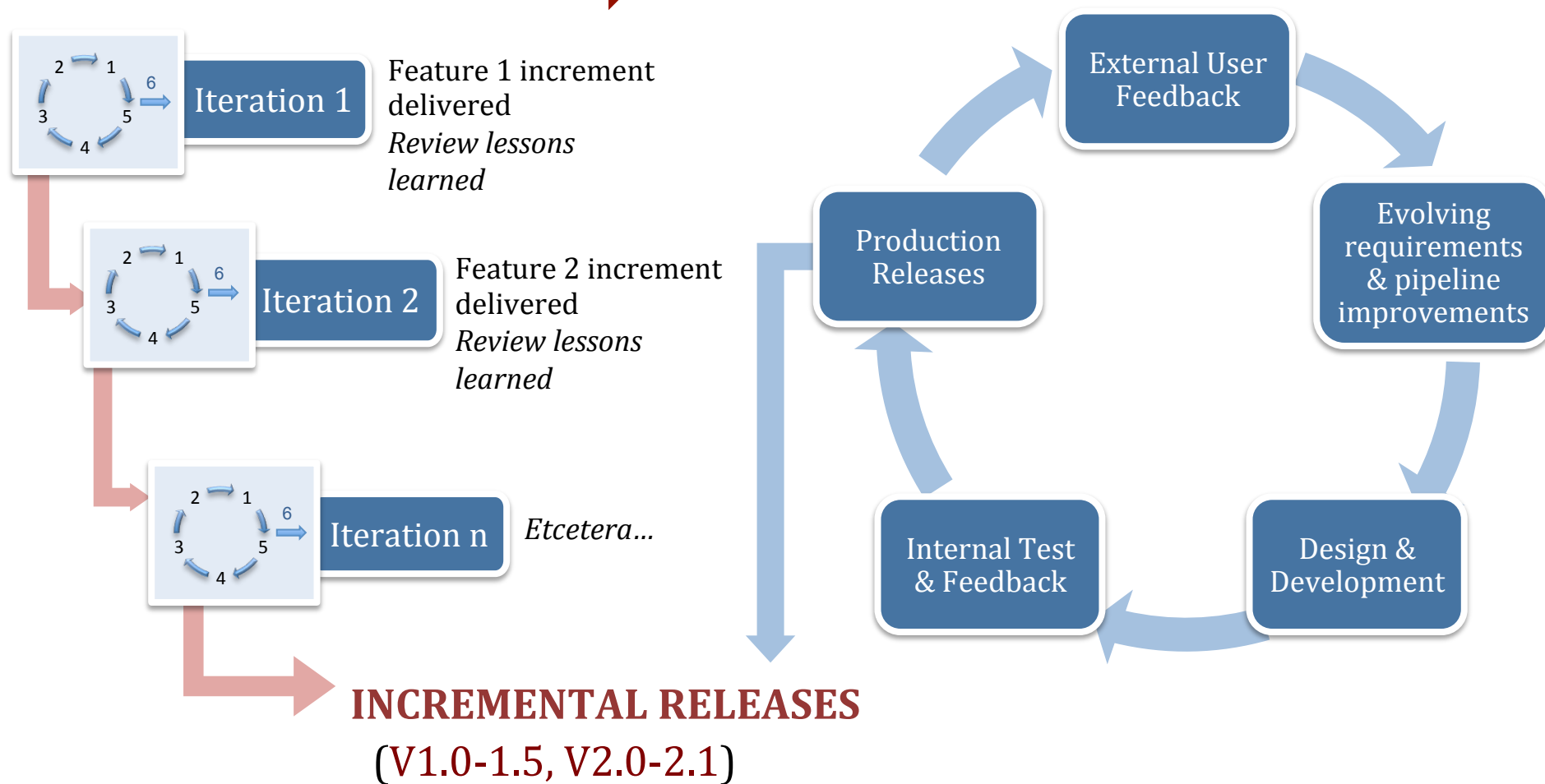


- **January 2014: X-ray production release (V1.0)**
- **Fall 2014: NMR public testing to begin (V2.0)**
- **End of 2014: 3DEM public testing to begin (V2.1)**
- **2015: Q3 Retire legacy systems**
 - Parallel deposition systems (new and legacy) available to depositors during transition period



Deployment and Life Cycle Management

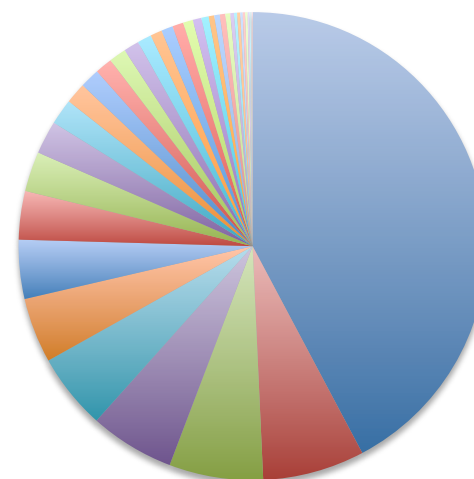
Agile Development



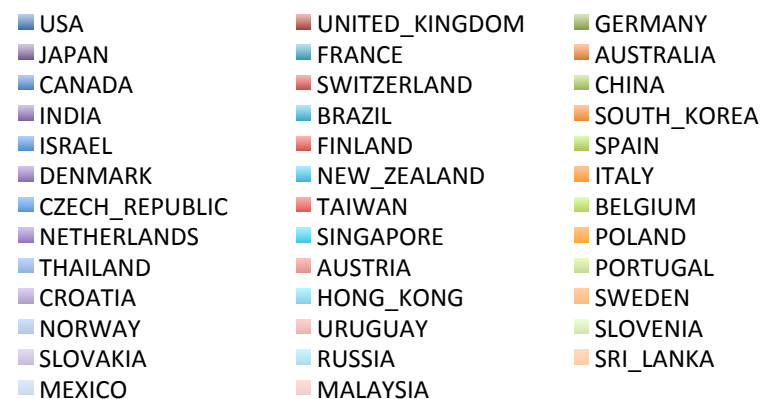
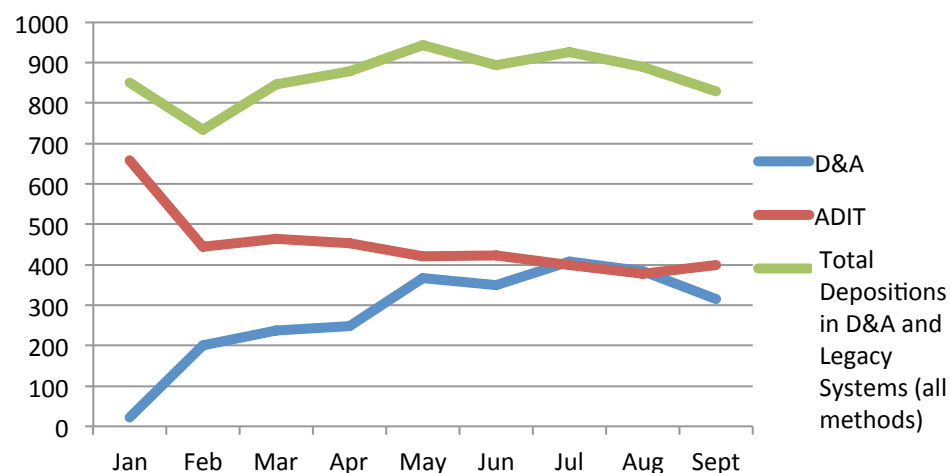
D&A System Usage: January 27 – September 30, 2014

- More than 2600 structures deposited and fully annotated
- More than 900 released
- More than 940 unique PIs

Number of Depositions by Country



Number of depositions per month

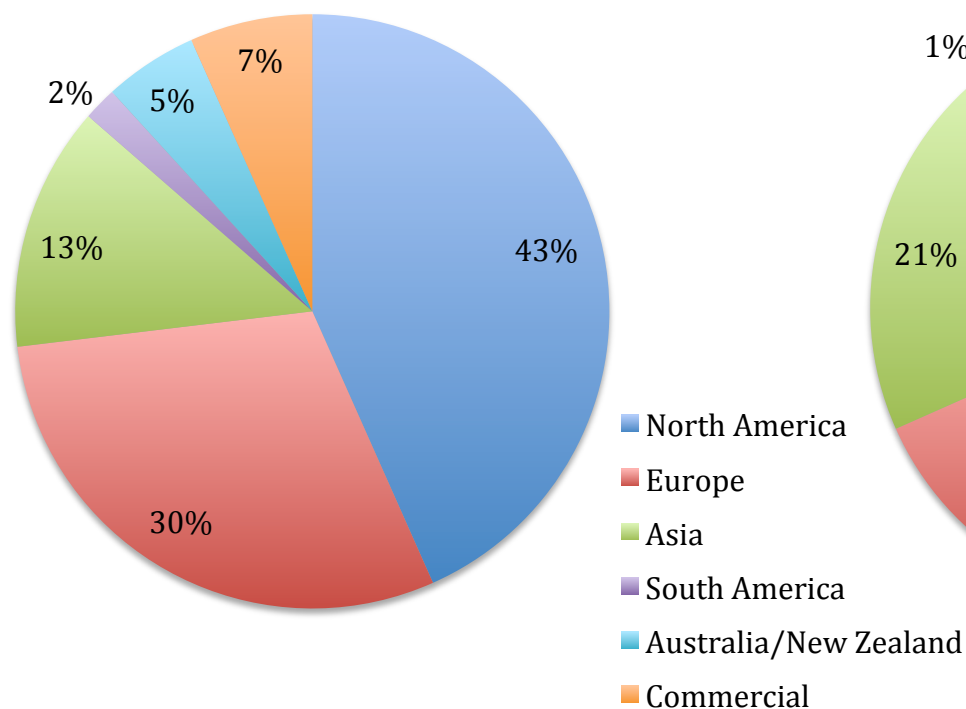


2014 Depositions by Geography

January 1 – September 30, 2014

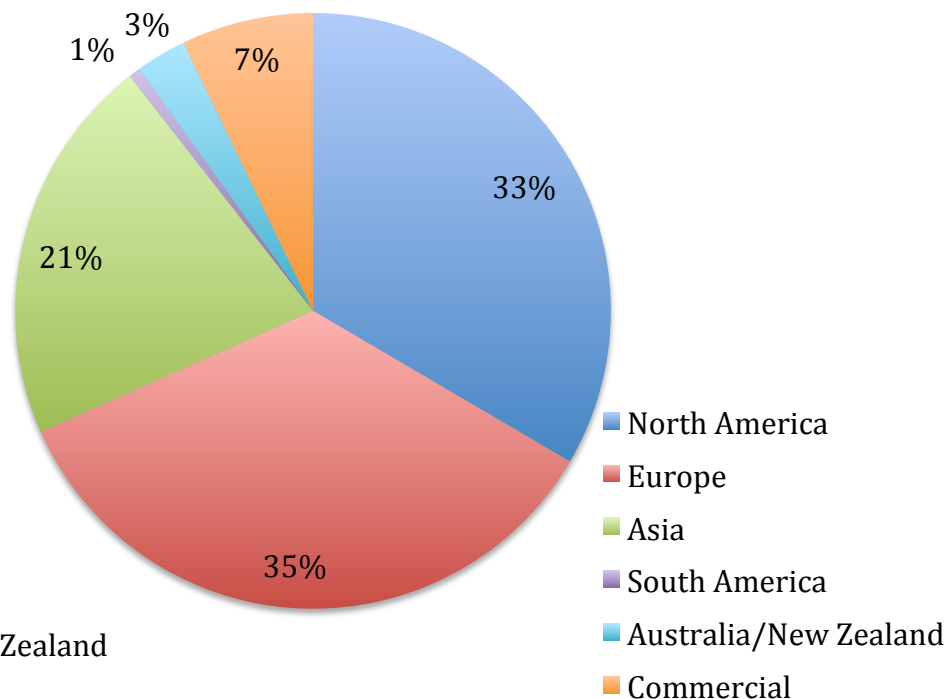
D&A System

(2531 entries, 33%)

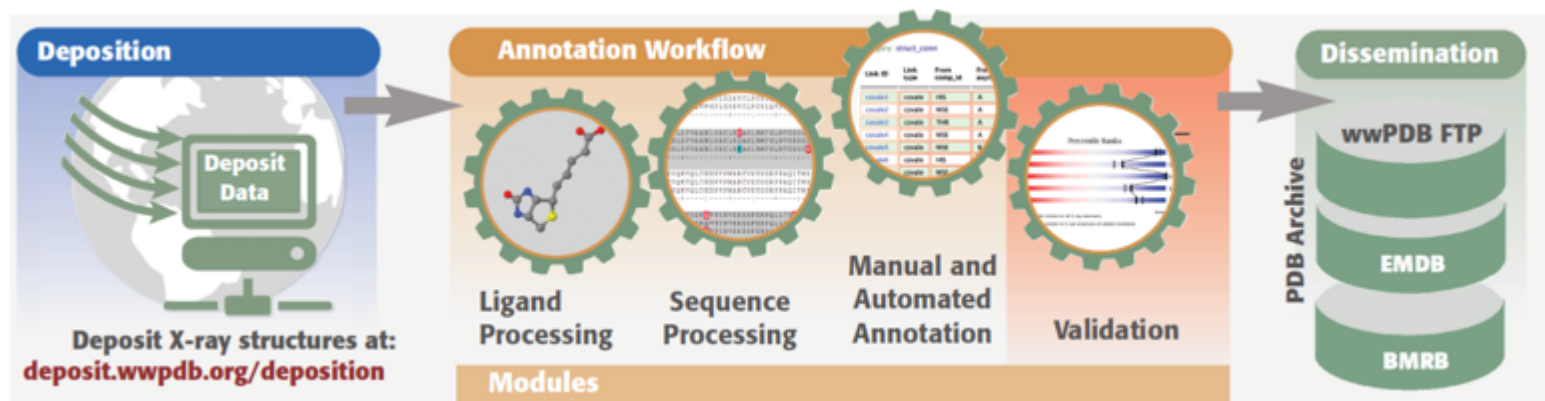


Legacy System

(5260 entries, 67%)



Key Deposition & Annotation Features



- **Data format:** Extensible, dictionary-driven PDBx/mmCIF produces more uniform data across the archive
- **Validation:** Geometry and experimental data checking during deposition and annotation is based on recommendations from expert task forces (X-ray, EM, NMR)
- **Deposition efficiency:** ability to replace coordinate and/or experimental data files pre- and post-submission
- **Enhanced communication:** "In context" correspondence with wwPDB annotators and ability to preview and download the PDBx/mmCIF entry file prior to submission
- **Improved annotation:** Improved ligand chemistry and polymer sequence checks with visual inspection provided during deposition

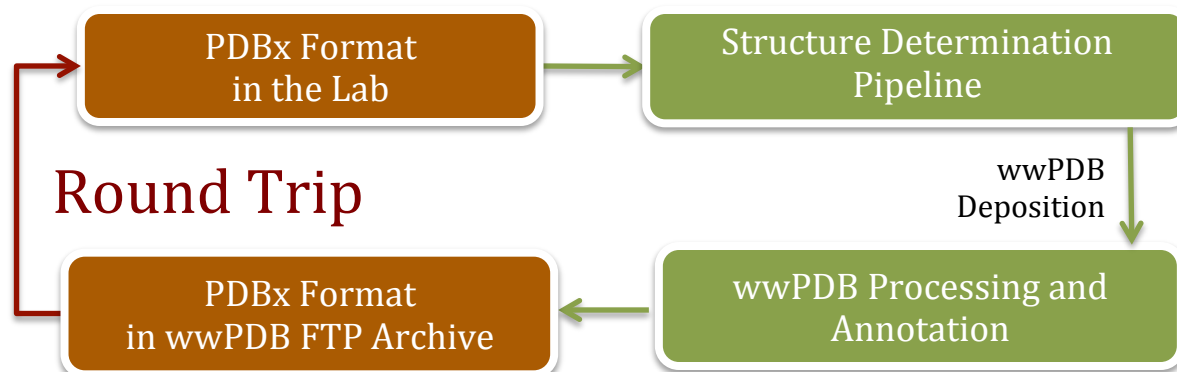


Deposition Data Format; PDBx/mmCIF Format

- PDBx addresses limitations in molecular size and complexity and extensibility of existing PDB format
- PDBx/mmCIF produced by major X-ray packages (CCP4/REFMAC and Phenix)
- PDBx/mmCIF Deposition Working Group recommending new content
- Working Group meets virtually ~monthly
- October 2014 Workshop finalized recommendations for X-ray data files and chemical restraints



Workshop Participants, September 2011



Workshop Participants, October 2014

PDBx and Release of Consolidated Large Structures

- New D&A system accepts, processes, and distributes depositions of any size in PDBx

July 9th 2014

- “Combined” single files representing large structures in both PDBx and PDBML formats released in a transitional FTP area for public review

December 10th 2014

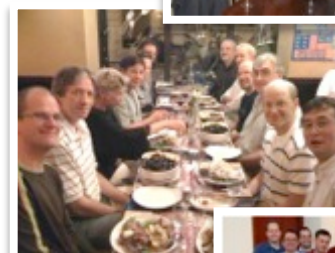
- Single files will be moved to the main FTP archive and split entries obsoleted
 - Best-effort PDB format-like bundles available in separate directory for existing and future large entries
 - Searches for split entries return the new “combined” entry



wwPDB Task Forces

Methods-specific (Validation) Task Forces and Workshops convened to develop recommendations re additional validation measures and archival needs/opportunities for non-traditional techniques

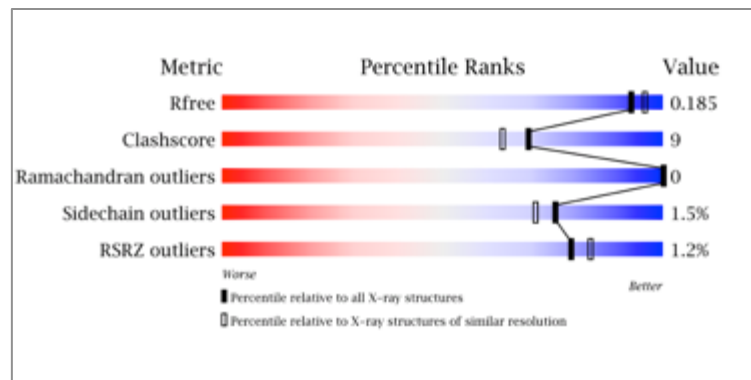
Task Force	Meeting/ Workshop	Chair(s)/Membership	Outcome
X-ray Validation Task Force	2008 (2015)	Randy Read (Univ of Cambridge) 17 members	(2011) <i>Structure</i> 19: 1395-1412
NMR Validation Task Force	2009, 2011, 2013 (x2) (2015)	Gaetano Montelione (Rutgers) Michael Nilges (Institut Pasteur) 10 members	(2013) <i>Structure</i> , 21: 1563-1570
3DEM Validation Task Force	2010	Richard Henderson (MRC- LMB) Andrej Sali (UCSF) 21 members	(2012) <i>Structure</i> 20: 205-214
Small-Angle Scattering Task Force	2012, 2014	Jill Trehwella (Univ Sydney) 6 members	(2013) <i>Structure</i> 21: 875-881
Hybrid Methods Workshop	2014	Andrej Sali (UCSF), Torsten Schwede (Univ Basel), Jill Trehwella (Univ Sydney) 27 members	(To be published)



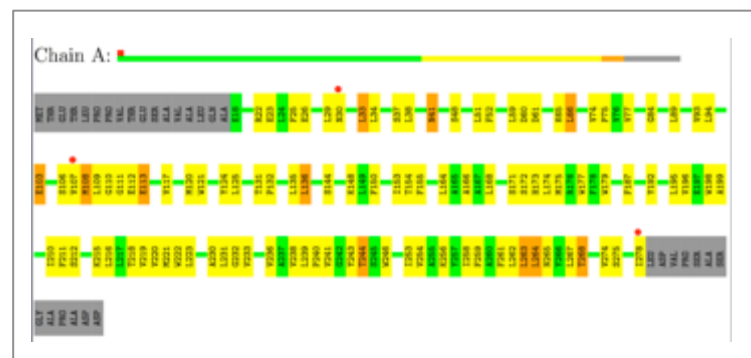
New Validation Reporting (v1.0)

- **Model Quality (all methods)**
 - Bond lengths and angles (outlier info, RMS-Z)
 - Chirality, planarity
 - Close contacts (including worst clashes, MolProbity clash score)
 - Torsion angles (Ramachandran statistics, protein rotamers)
 - Ligand geometry (Mogul analysis)
- **Residue Plots (all methods)**
 - Residues with model-quality outliers (0, 1, 2, >2)
 - Residues with RSR-Z > 5 are highlighted
 - Residues not observed
- **Method Specific**
 - X-ray: Model *versus* Electron Density
 - NMR: Ensemble RMSD in well-defined region
 - EM: model/map fit (in development/testing)

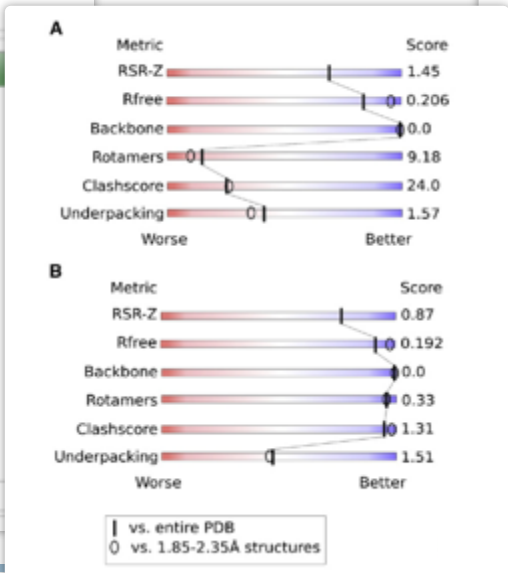
Overall Quality Summary



Residue Plots



- <http://wwpdb-validation.wwpdb.org/>

[illegible]

Deposition Pipeline

Depositors can

- Base new entries on previous depositions
- Upload replacement files mid-deposition (e.g., new coords)
- Provide more complete entries *via* mandatory data items
- View community-defined validation prior to submission
- Review and download annotated files post-deposition
- Communicate with expert annotators during deposition process using web-based tools

The screenshot shows the 'wwPDB Deposition Tool' interface. On the left, there is a section for 'Existing deposition' with fields for 'Deposition ID' and 'Password', and a 'Log in' button. On the right, there is a section for 'New deposition' with fields for 'E-mail' (filled with 'jasmin@rcsb.rutgers.edu'), 'Preferred deposition site' (set to 'RCSB'), and 'Location' (set to 'United States'). Below these fields is a list of 'Experimental Method' options: X-Ray Diffraction (checked), Electron Microscopy, Solution NMR, Neutron Diffraction (checked), Electron Crystallography, Solid-state NMR, and Solution Scattering. A callout box highlights the 'Experimental Method' list, showing the same options with checkboxes.

Experimental Method

- ☒ X-Ray Diffraction
- ☐ Electron Microscopy
- ☐ Solution NMR
- ☒ Neutron Diffraction
- ☐ Electron Crystallography
- ☐ Solid-state NMR
- ☐ Solution Scattering
- ☐ Fiber Diffraction

Multiple methods

Report & Data-entry Panel

Navigation Panel

Communication User Interface

wwPDB Deposition Tool - D_1000200336 (4OZO)

[FAQ](#) [Tutorials](#)

Navigation

- Welcome
- Replacement Upload
- Upload Summary
- Admin
 - Contact information
 - Grant information
 - Release status
 - Entry Title & author
 - Citation information
- Macromolecules
 - Molecule 1
- Data collection
 - X-ray data
- Refinement
 - X-ray refine
 - X-ray TLS
- Ligands
- Biological assembly
- Communication
- Summary & conditions
- View Reports
 - Validation Reports

Log out

Old messages

TimeStamp	Sender	Subject	Content (click to show more)
2014-02-25 21:44:11	MRS	RE: 4OZO entry release	Dear Pierre, Thank you for y...
2014-02-20 20:22:06	depositor	4OZO entry release	Dear Monica, Thank you for th...
2014-02-20 17:10:41	MRS	RE: Answer to the validation report for 4OZO	Dear Pierre, Thank you for y...
2014-02-20 08:38:35	depositor	Answer to the validation report for 4OZO	All 5 five items have been che...
2014-02-19 21:18:02	MRS	Structure 4OZO validation report	Dear Depositors, Please ackn...

Last message from annotator

Time stamp: 2014-02-19 21:18:02
Sender: MRS
Subject: Structure 4OZO validation report

Dear Depositors,
Please acknowledge receipt of this message.

Full Text: This message concerns your structure PDB ID (4OZO) and Deposition ID (D_1000200336) entitled:

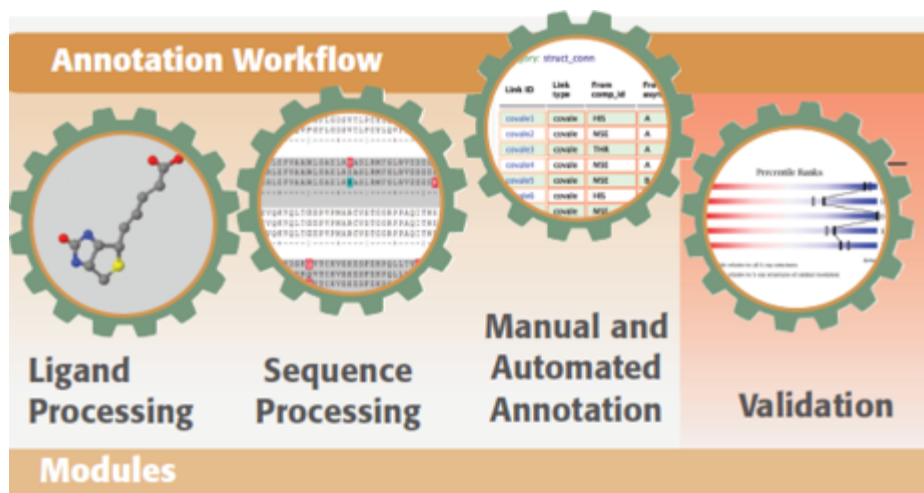
Crystal structure of an α-L-fucosidase GH29 from Bacteroides thetaiotamicron (BT2192) in complex with oNPTG

Attachments: D_1000200336_model-annotate_P1.cif.V1
D_1000200336_model-review_P1.cif.V1
D_1000200336_model-annotate_P1.pdb.V1
D_1000200336_val-annotate_P1.cif.V1
D_1000200336_val-report-annotate_P1.pdf.V1
D_1000200336_val-report-full-annotate_P1.pdf.V1
D_1000200336_val-data-annotate_P1.xml.V1

New message

This page supports review and download of validation reports and processed data files, and all communication with wwPDB annotation staff, including release requests.

Annotation System



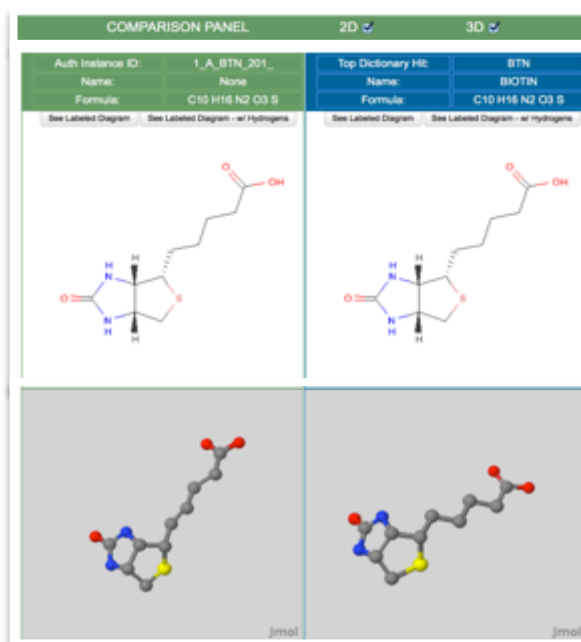
- Enables workload balancing and increased productivity
- Better quality assurance of ligand chemistry and polymer sequences
- Validation suites based on recommendations from expert task forces; X-ray validation pipeline available as a stand-alone server
- System supports all accepted experimental methods
- All data management in PDBx/mmCIF



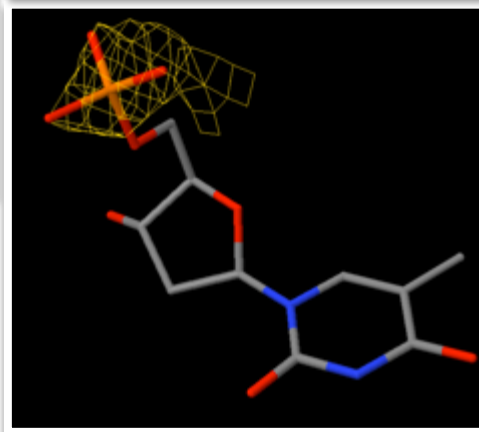
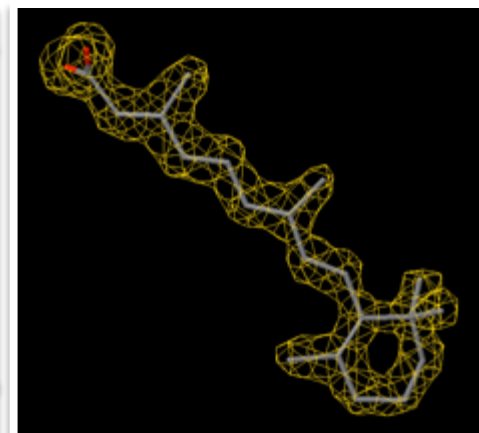
Ligand Annotation: Improved Chemistry Assessment

- Batch searching against Chemical Component Dictionary (CCD) and automated ligand ID assignment
- Automatic capture and display of Author-provided chemical information
- Comparison Panel
 - 2D and 3D views of ligand for review
 - ID assignment
- Enhanced with local ligand density fit
- Paper published

J. Y. Young, Z. Feng, D. Dimitropoulos, R. Sala, J. Westbrook, M. Zhuravleva, C. Shao, M. Quesada, E. Peisach, H. M. Berman. (2013) Chemical annotation of small and peptide-like molecules at the Protein Data Bank. *Database* 2013: bat079.



Deposited instance from coordinates (left) and the closest match in the dictionary (right)



Local ligand density display (1.5 sigma omit map)

Top: REA in entry 1CBS with LLDF=1.31 (RSR=0.10, CC=0.95)

Bottom: TMP in entry 3HW4 with LLDF=6.77 (RSR=0.41, CC=0.70)

Sequence Annotation: ↑ Data Consistency

- Biological sequence checked against atomic coordinate sequence and cross-referenced to UniProt/GenBank
- 3D structure view
- Sequence discrepancy annotation
- Enhanced annotation for chimeric proteins

Four sequence references

SP:P62161
SP:Q556Z9
SP:Q556Z9
SP:Q6LDG3

AUTH Entity: 1 V:4
XYZ Chain: A V:1
SP:P62161
SP:Q556Z9
SP:Q556Z9
SP:Q6LDG3

..... MHHHHHGGDQLTEEQIAEFKEAFSLFDKDDGGTITTTKELGTVMRSLGQNPTEAELQDMI
..... QLTTEEQIAEFKEAFSLFDKDDGGTITTTKELGTVMRSLGQNPTEAELQDMI
..... DQLTEEQIAEFKEAFSLFDKDDGGTITTTKELGTVMRSLGQNPTEAELQDMI

1 |-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|

NEVDADGGDTIDFPEFLTMMARKMKDTSDEEEIREAFRVFDKDGNGYISAAELRHVMTNL
NEVDADGGDTIDFPEFLTMMARKMKDTSDEEEIREAFRV.....YEEFVQMMTAKLECEKMYVRDGVLTGDIHMALLE
NEVDADGGDTIDFPEFLTMMARKMKDTSDEEEIREAFRVFDKDGNGYISAAELRHVMTNL

.....
.....
.....

61 |-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|

GEKLTDEEVDEMIREADIDGGQVNYEEFVQMMTAKLECEKMYVRDGVLTGDIHMALLE
.....YEEFVQMMTAKLECEKMYVRDGVLTGDIHMALLE
GEKLTDEEVDEMIREADIDGGQVNYEEFVQMMT.....EKMYVRDGVLTGDIHMALLE

.....
.....
.....

121 |-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|

GNAHYRCDFRTTYKAKEKGVKLPGYHFVDHCIEILSHDKDYNKVKLYEHAVAHSGLPDNA
GNAHYRCDFRTTYKAKEKGVKLPGYHFVDHCIEILSHDKDYNKVKLYEHAVAHSG.....
GNAHYRCDFRTTYKAKEKGVKLPGYHFVDHCIEILSHDKDYNKVKLYEHAVAHSGLPDNA

.....
.....
.....

181 |-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|

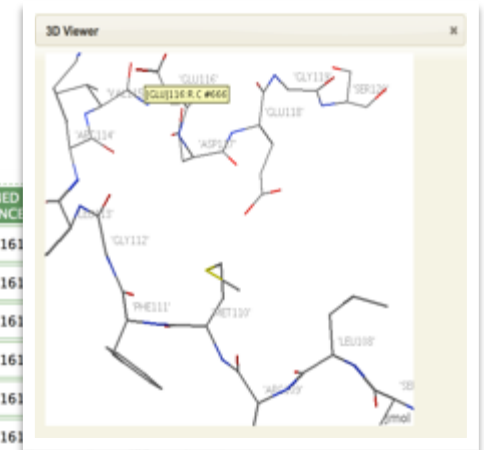
RRGGTGGSHVSAIKPDMKILRMEGNVNGHHFVIDGGTGKPEGKQMDLEVKEGGPLP
.....VSAIKPDMKILRMEGNVNGHHFVIDGGTGKPEGKQMDLEVKEGGPLP
.....SAIKPDMKILRMEGNVNGHHFVIDGGTGKPEGKQMDLEVKEGGPLP

.....
.....
.....

241 |-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|

ALIGNMENT POSITION	AUTH ENTITY:1	ALIGNED SEQUENCE	
1	MET	SP:P62161	
2	HIS	SP:P62161	
3	HIS	SP:P62161	
4	HIS	SP:P62161	
5	HIS	SP:P62161	
6	HIS	SP:P62161	
7	HIS	SP:P62161	- expression tag
8	GLY	SP:P62161	- expression tag
9	SER	SP:P62161	- expression tag
68	ASP	SP:P62161	ASN engineered mutation
155	ALA	SP:Q556Z9	- linker
156	LYS	SP:Q556Z9	- linker
157	LEU	SP:Q556Z9	- linker
158	GLU	SP:Q556Z9	- linker
159	CYS	SP:Q556Z9	- linker
174	HIS	SP:Q556Z9	THR engineered mutation
241	ARG	SP:Q556Z9	- linker
242	ARG	SP:Q556Z9	- linker
243	GLY	SP:Q556Z9	- linker
244	GLY	SP:Q556Z9	- linker
245	THR	SP:Q556Z9	- linker

Residue: ALA
Position: 155



Annotation Training and Transition

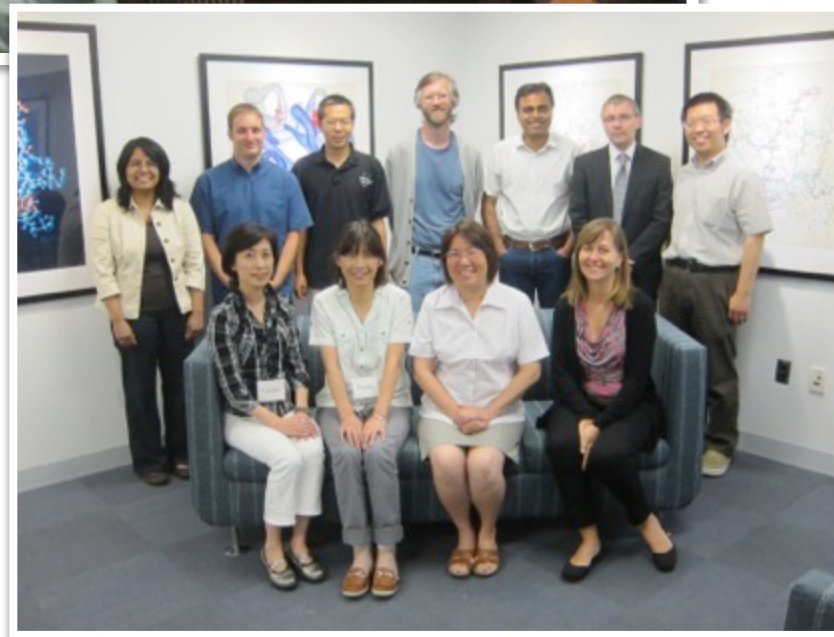
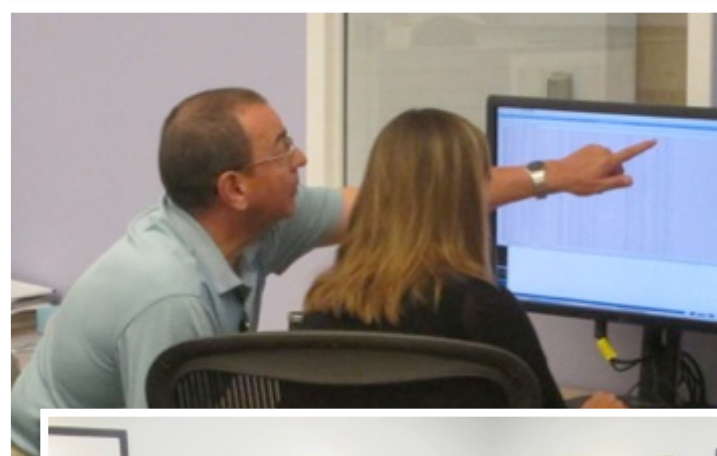
RCSB PDB training

- Staff migrating to new system throughout 2014
 - Testing and training
 - Processing with both new and legacy systems in parallel
- Stephen Burley trained; processed two production entries

wwPDB training

- wwPDB processing guidelines developed
- VTC training carried out
- Annotator Summit in June 2014

RCSB PDB currently hosting production system, serving US, EU, and Japan



Efficiencies/Advances: Current Status

Efficiencies

- Greater efficiency for routine structures
- Evolving processing of large and complex structure
- Workflow flexibility still evolving
- Detailed efficiency measurements to be carried out in 2015

Advances

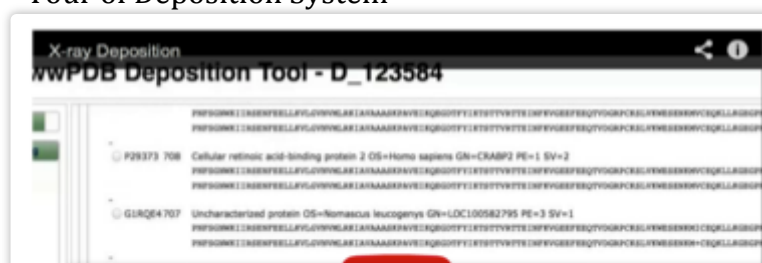
- Improved support for large structures
- Workflow ensures completeness of annotation
- Enhanced validation
- More automated processing for ligands and sequences
- Tracking enabled
- Enable processing multiple entries in parallel



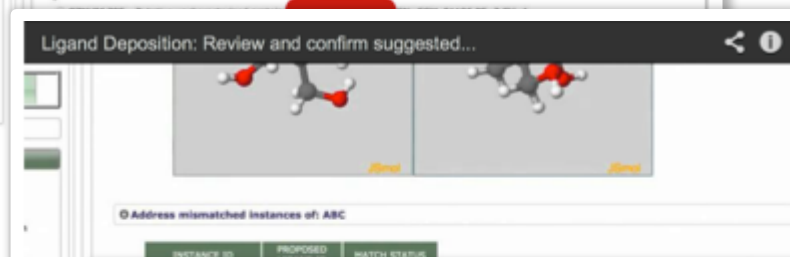
Depositor Resources

- Info at wwpdb.org
 - System overview
 - Step-by-step instructions
 - Video tutorials
 - FAQ
- Live and/or video demos at ACA and IUCr meetings

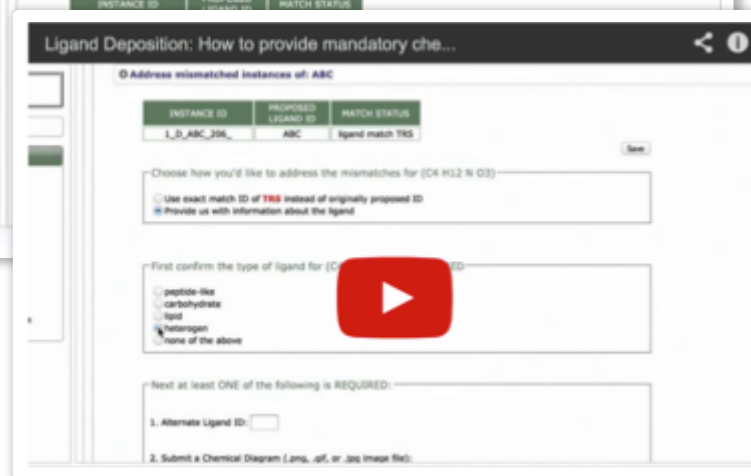
Tour of Deposition System



X-Ray Deposition: file upload, validation, and providing mandatory information



Ligand Deposition: review and confirm suggested matches to deposited ligands



Ligand Deposition: how to provide mandatory chemical information



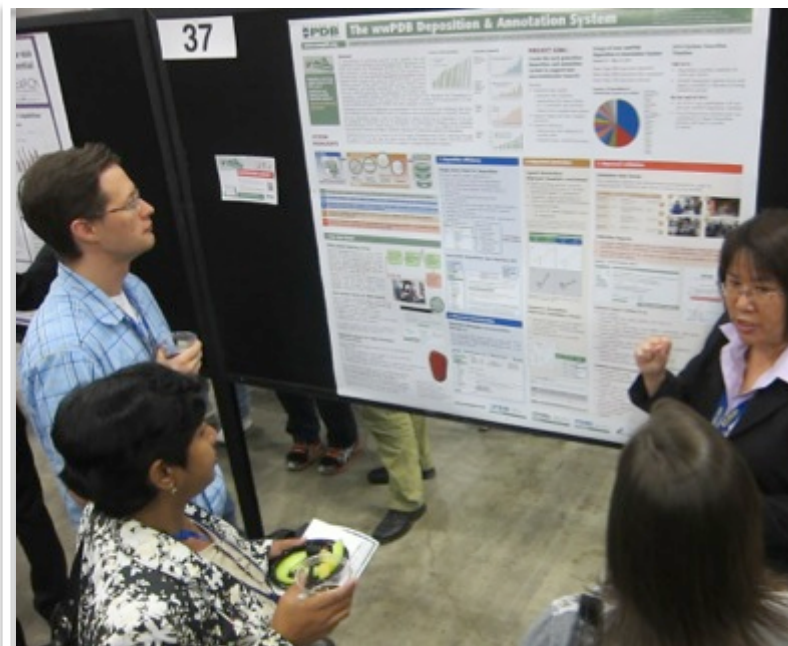
Feedback

External Users (Depositors)

- Daily communication between annotators and depositors
- ACA, IUCr meetings: demonstrations, posters and booth

Internal Users (Annotators)

- Continuous testing
- Internal training
- 9 months of production experience
- Weekly cross-site review of issues



Common Feedback from External Users

What they like most

- Easy navigation
- Straightforward input forms

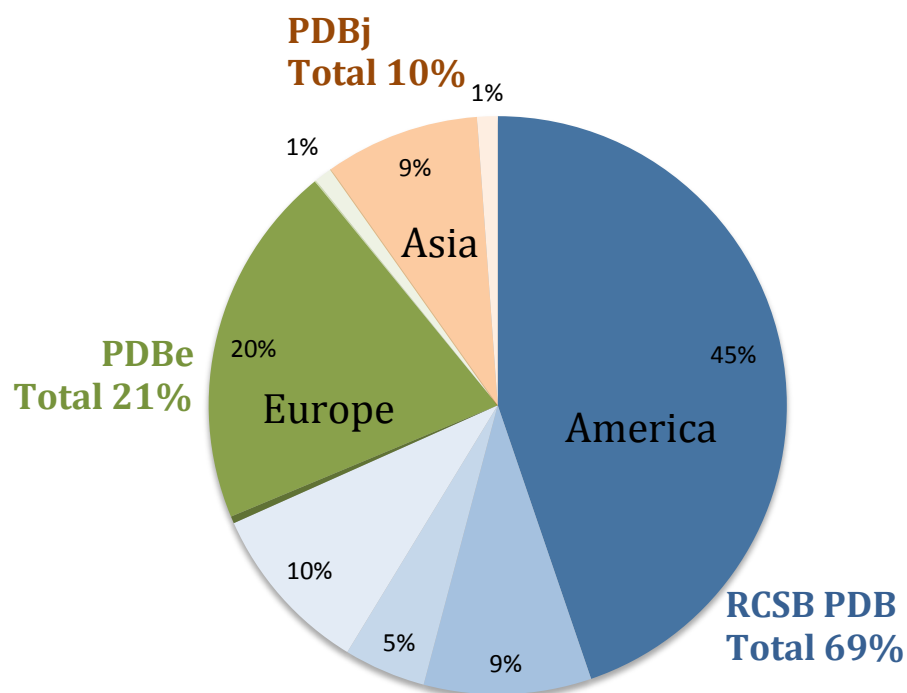
Suggestions under discussion

- Include annotator message content in alerts emailed to Depositors (in progress)
- Support for simultaneous depositions
- Ability to access depositions by Author rather than session ID

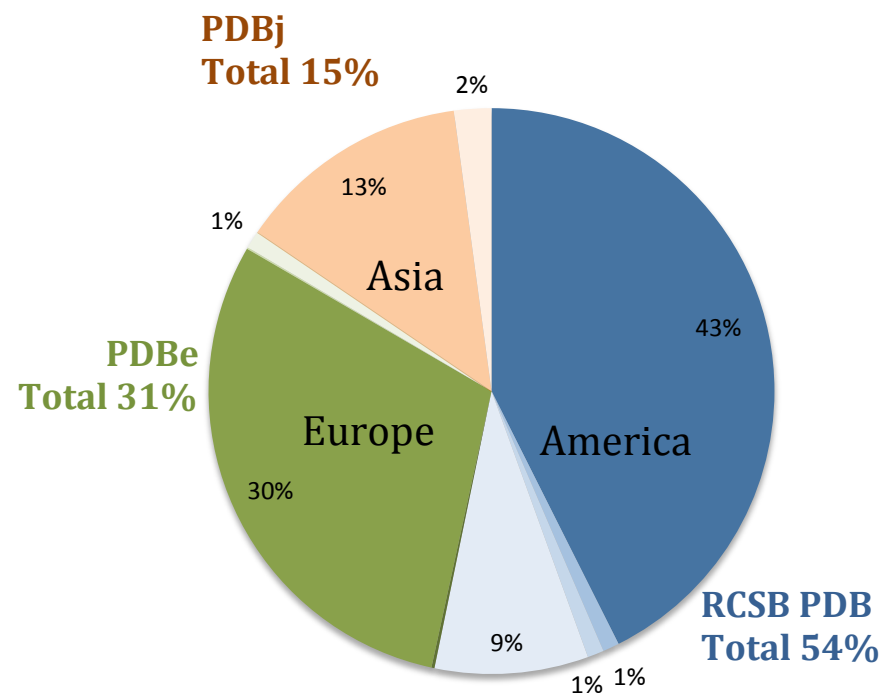


Global Workload Distribution

Launch of the D&A System
January-September 2014



Start of “Workload Share”
June-September 2014



RCSB PDB	■ America	■ Europe	■ Asia	■ Other
PDBe	■ America	■ Europe	■ Asia	■ Other
PDBj	■ America	■ Europe	■ Asia	■ Other



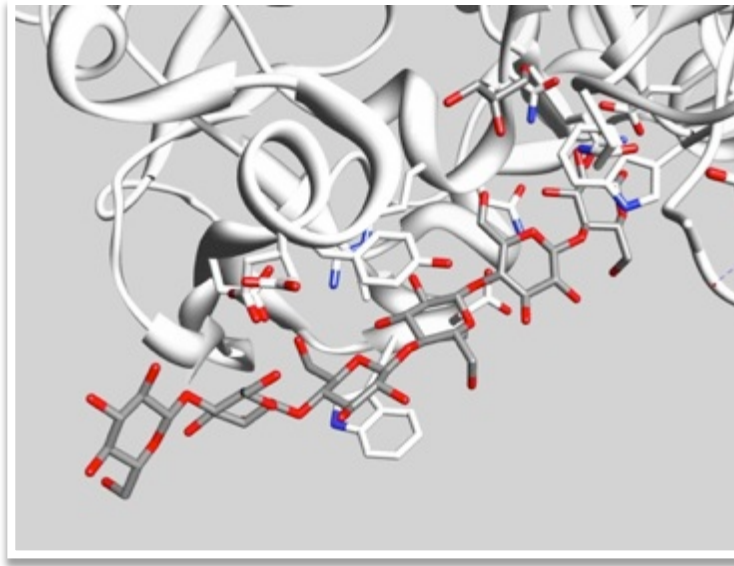
What challenges remain?

- Extend and harden existing system for all methods
- Improving efficiency for very large structures
- Enable greater automation of depositor replacement of coordinates post submission
- Improve usability of processing tools for annotation and the flexibility of current workflow
- Incorporate additional data checking to improve data quality and processing efficiency
- Improve support for legacy PDB file format variants



Future Remediation

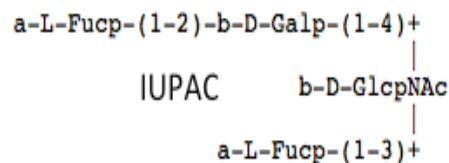
- Residual B factors (2015)
- Non-crystal frame (2015)
- Archival mmCIF update: multiple reflection data sets and V5.0 (2015)
- Representation of Carbohydrates



Carbohydrate Remediation

Funding Proposal	Proposal for Mizutani Foundation (MF) funding submitted September 01, 2014. Waiting for decision by Spring 2015
Goals	- Gather input from community experts on new search and delivery functionalities for the glycosciences community - Enable searches for carbohydrates in PDB archive
Plans Forward	- Identify and analyze carbohydrate-containing entries - Remediate batches of carbohydrate containing files using the Deposition & Annotation pipeline - Create standard representation for branched polymers

LINUCS: `[ ][b-D-GlcpNAc]{[(3+1)][a-L-Fucp]{ }[(4+1)][b-D-Galp]{[(2+1)][a-L-Fucp]{ }}}`



PDB ID: 2wmg MA Higgins *et al. J.Biol.Chem.* 284, 26161-26173 (2009)



Plans for 2014/2015

- 2014
- wwPDB D&A system for all methods in production testing
- 2015
- Assess efficiency/usability and harden new D&A system
- Retire legacy deposition systems at RCSB PDB and PDBe
- Archival mmCIF updates
- Enabling industrial structure deposition
- Secure funding for Carbohydrate Remediation
- Explore the way forward for hybrid structure and data deposition
- Begin discussions re PDB “franchises” in India and China (with wwPDB partners)



RCSB PDB Data Deposition and Annotation Team



wwPDB D&A Team



wwPDB Annotator Summit



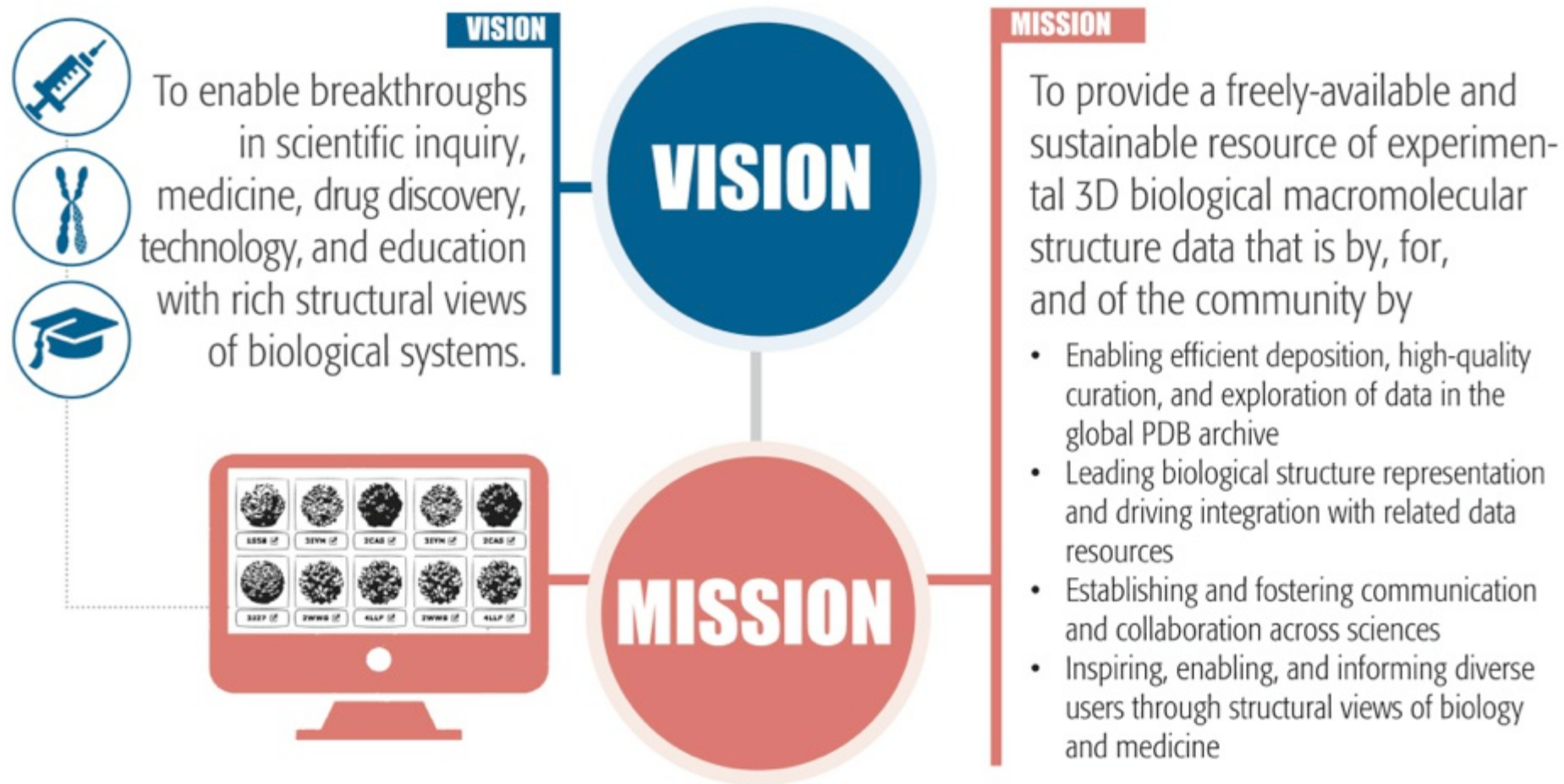
Data Access and Exploration

Peter Rose

Andreas Prlić



Vision and Mission



Primary Goals

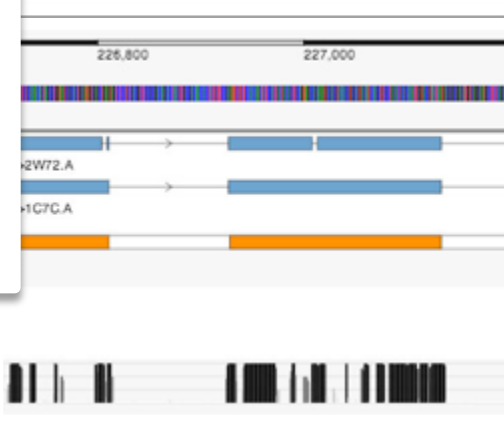
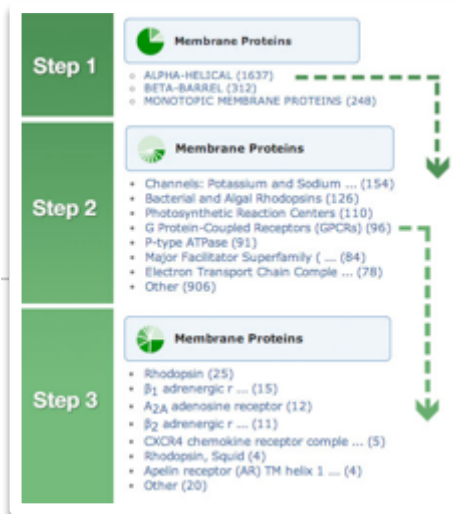
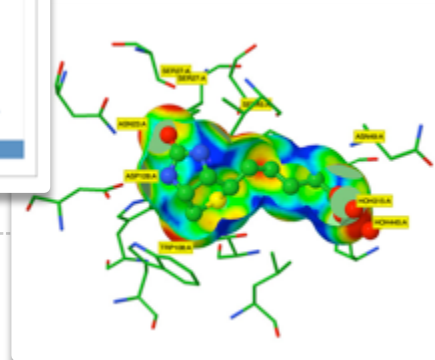
- Enable efficient exploration of the PDB archive
- Enhance User experience
- Design data stores supporting efficient query
- Modularize services to enhance extendibility, scalability, and maintainability
- Develop geometry and interaction query capabilities
- Expand programmatic access *via* Web Services



Recent Accomplishments

Improved website usability, new structure annotations, improved processes

- Home page redesign
- JS for web support and visualization
- Membrane protein annotation/query
- Mapping structures to human genes



Improving Website Usability

- Design focus
 - Reduce information overload and redundancy
 - Support diverse User groups
 - Optimize User workflows
 - Employ modern (yet proven) web technologies
- Scope
 - Phase 1
 - Home Page (November 2014)
 - Search Engine optimization, make website mobile friendly, usage tracking, review User feedback (January 2015)
 - Phase 2 (through September 2015)
 - Redesign of secondary pages, including query results and Structure Summary pages
 - Redesign PDB-101website



Header
accessible
from every
page

Common
tasks

New Features

Static
footer

Topic Panels
change
based on
selected
task

BETA

RCSB PDB

Deposit

Search

Visualize

Analyze

Download

Learn

More

MyPDB Login

RCSB PDB

An Information Portal to 102550 Biological Macromolecular Structures

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

Go

Advanced Search | Browse by Annotations | Search History | Previous Results

PDB-101

EMDataBank

Welcome

Deposit

Search

Visualize

Analyze

Download

Learn

A Structural View of Biology

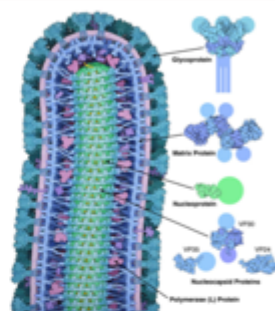
This resource is powered by the Protein Data Bank archive - information about the 3D shapes of proteins, nucleic acids, and complex assemblies that helps students and researchers understand all aspects of biomedicine and agriculture, from protein synthesis to health and disease.

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.

Use this website to access curated and integrated biological macromolecular information in the context of function, biological processes, evolution, pathways, and disease states.

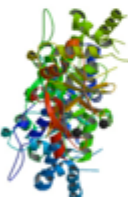
October Molecule of the Month



Ebola Virus Proteins

Latest Entries

As of this Tuesday > See all



3ZJF

A20 OTU domain with irreversibly oxidised Cys103 from 270 min H2O2 soak.

View in 3D

New Features

See August 2014 Release Features



New Top Bar Menu
Providing More Space to Users



Gene View
Gene Structure and Mapping to PDB Entries for Human Genes



Protein Feature View
Enhanced with New Information



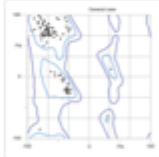
Ramachandran Plots
Enhancing Protein Structure Validation



Average Temperature (B) Factor

News

Publications



Exploring Structure Quality
View a validation "slider" image, full Validation Report, and Ramachandran plots for X-ray structures in the PDB. - 09/16/14

Head Back to School with PDB-101 - 09/09/14

Join the RCSB PDB Team - 09/02/14



Inclusion of Large Structures in the Main PDB Archive
- 09/16/14

wwPDB Events at IUCr (August 5-12) - 08/01/14

Improved Representation of Large Structures in the

PDB at a Glance

34228 Distinct Protein Sequences

26331 Structures of Human Sequences

7621 Nucleic Acid Containing Structures

More Statistics

About

About Us

Citing Us

Publications

Team

Careers

Usage & Privacy

Help

Contact Us

Help System

Website FAQ

Browser Test

Glossary

RCSB Partners

EMDataBank

Nucleic Acid Database

Structural Biology Knowledgebase

The RCSB PDB (citation) is managed by two members of the Research Collaboratory for Structural Bioinformatics:

RUTGERS

UC San Diego

RCSB PDB is a member of the

PDB

EMDataBank

The RCSB PDB is funded by a grant from the National Science Foundation, the National Institutes of Health, and the US Department of Energy.

Live Demo of New Website

- Topic panels and subpanels
- Top menu bar
- New features and news sections
- Context specific feedback form
- Responsive for mobile devices



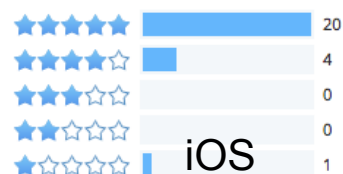
RCSB PDB Mobile

Fully supported on
iOS and Android

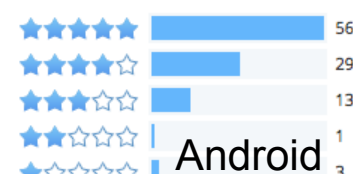


RCSB PDB Mobile: iOS and Android
mobile apps to provide data access
and visualization to the RCSB Protein
Data Bank (2014) *Bioinformatics*
doi:10.1093/bioinformatics/btu596

★★★★☆ 25 ratings



★★★★☆ 102 ratings



Exceeds expectations

by Waynebio on Saturday, September 07 2013



Really surprise by the existence of this app. Even more surprised by the functions it carries! It completely transformed the experience of finding, viewing, and saving structures. I find it much easier to use this app compare to using a web browser.

Awesome!

by helloldon'twanttoshowmyname on Saturday, November 30 2013



RCSB really went above and beyond with this app. The molecule viewer is fantastic, and I'm glad they included the MoM. I am entirely pleased with the search function. One useful feature that is missing is the SCOP classifications of the proteins, but I think I can live without those on my phone. I'd like to be able to browse MoM by category, but it's not so big of a deal.

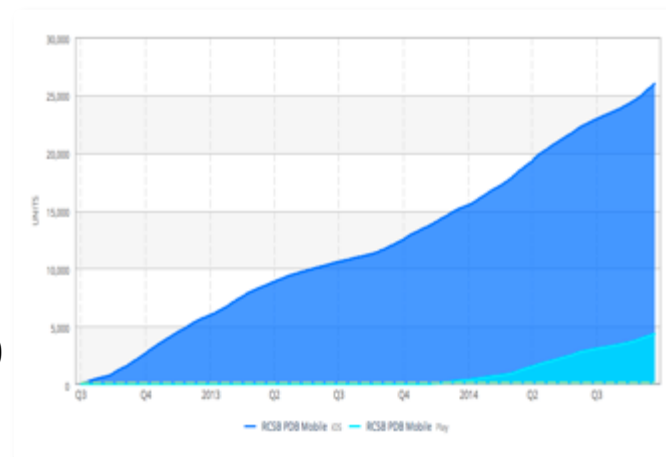
Structures on demand!

by Dale Ang on Tuesday, March 25 2014 version 3.01

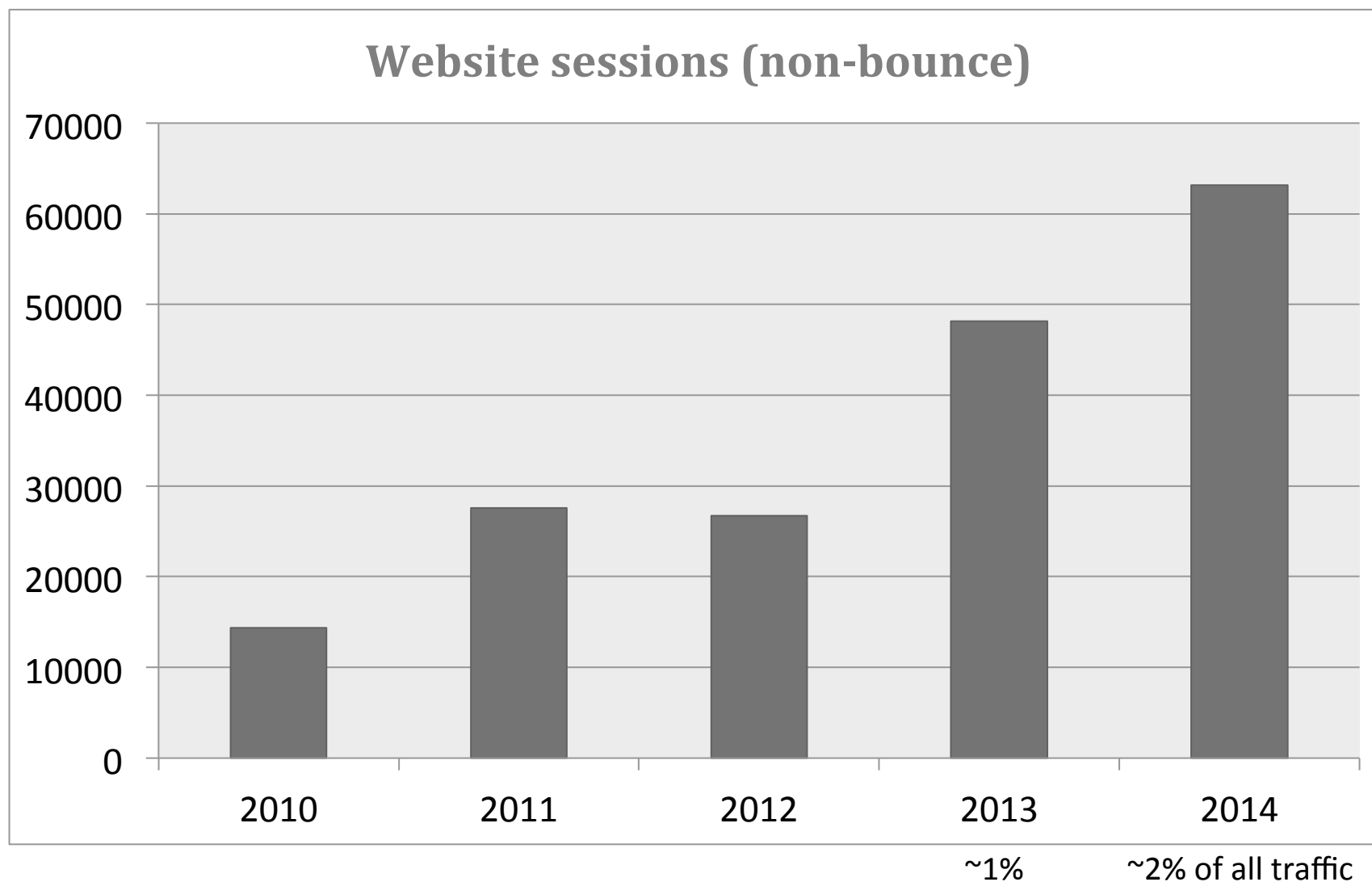
Interface works well and the graphical rendering is very smooth with plenty of style choices for displaying the structures. Great job!

Downloads

iOS: 22,000
Android: 4,500



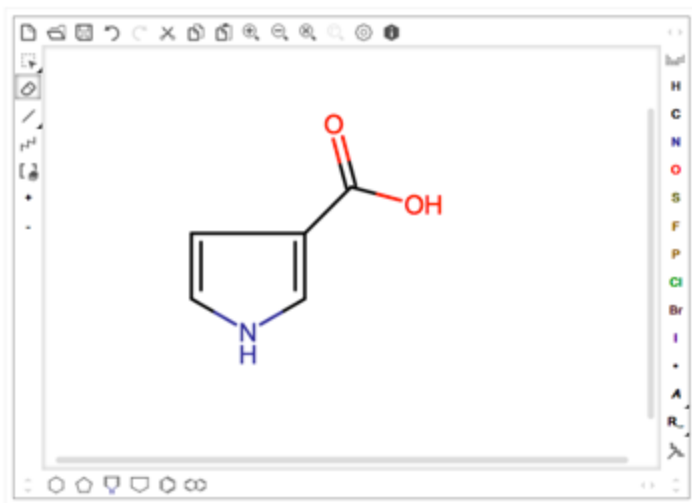
Mobile Use of Website



Mobile Web Support/Visualization

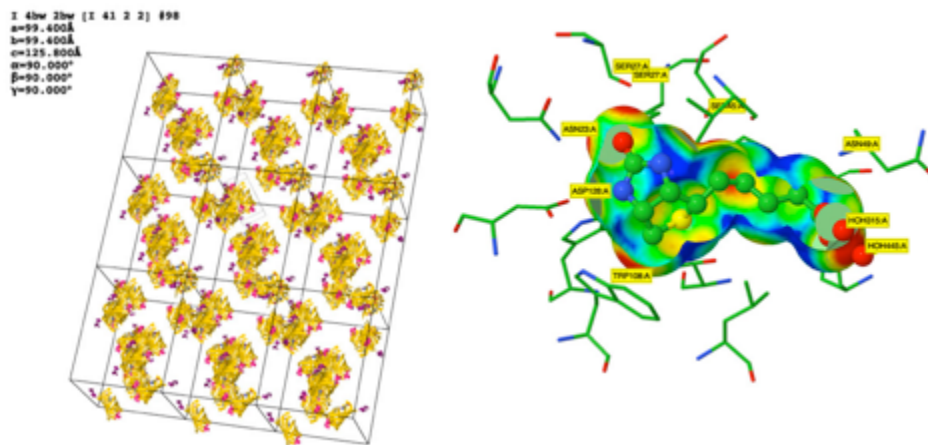
- New JavaScript applications run on devices and web browsers that do not support Java (phones, tablets)

Chemical Structure Searching



MarvinJS (ChemAxon)

3D Visualization



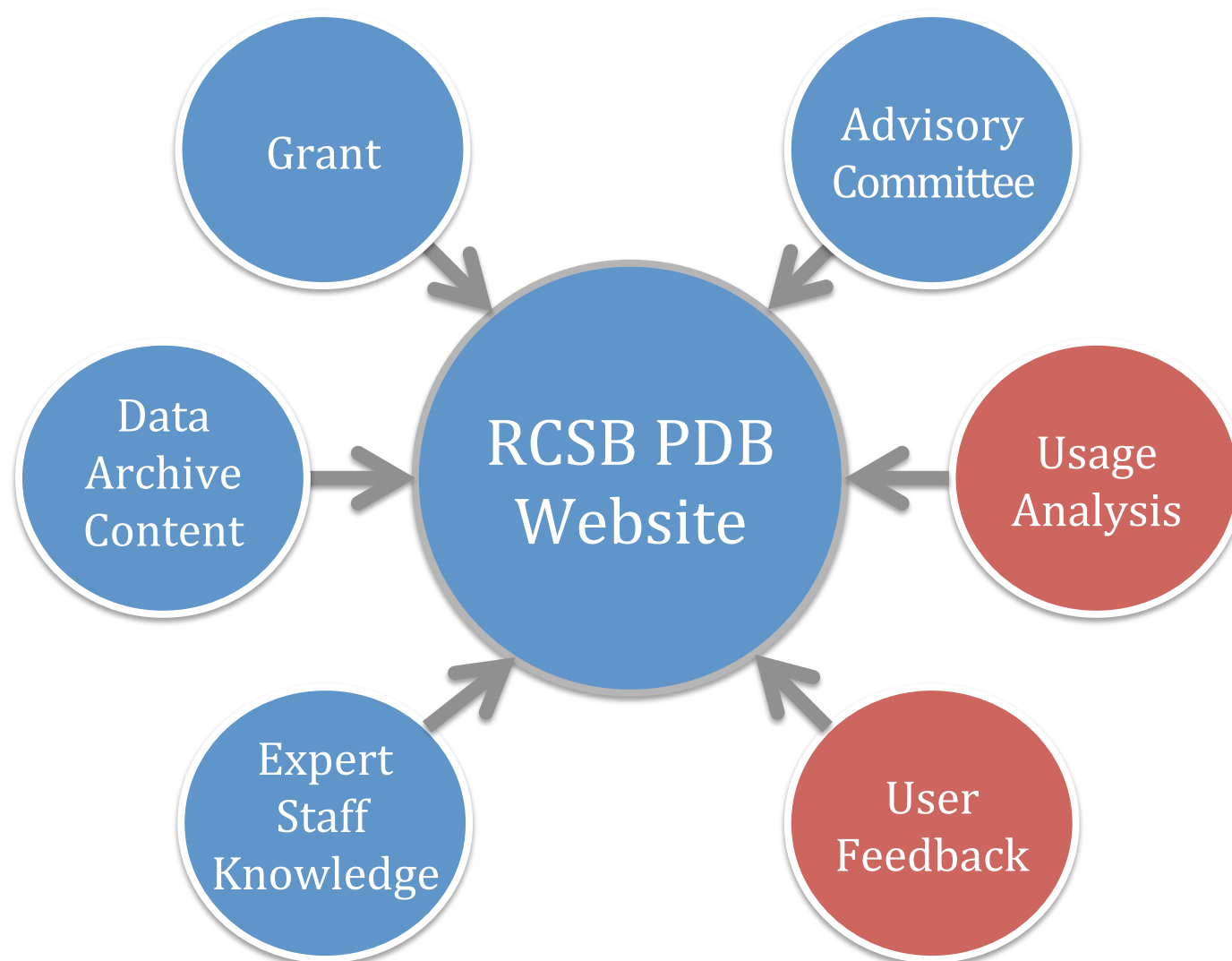
crystal packing

binding site

JSmol (Robert Hanson)



Factors Influencing Our Decisions



Feedback from Help Desk, Analytics

- What membrane proteins are in the PDB?
- What can we learn when we search by gene-identifier?



Membrane Protein Annotation/Query

- Integration with mpstruc database (UC Irvine)
 - mpstruc currently annotates ~1500 PDB entries
 - Collaboration with Structural Biology Knowledgebase
- Extended to homologs using RCSB PDB sequence clusters
 - ~2500 membrane proteins annotated



Browse Membrane Proteins



Membrane Proteins

- ALPHA-HELICAL (1865)
- BETA-BARREL (343)
- MONOTOPIC MEMBRANE PROTEINS (317)



Membrane Proteins

- Channels: Potassium, Sodium, & ... (177)
- Bacterial and Algal Rhodopsins (133)
- G Protein-Coupled Receptors (GPCRs) (125)
- Photosynthetic Reaction Centers (117)
- Multi-Drug Efflux Transporters (83)
- P-type ATPase (82)
- Electron Transport Chain Comple ... (80)
- Other (1069)

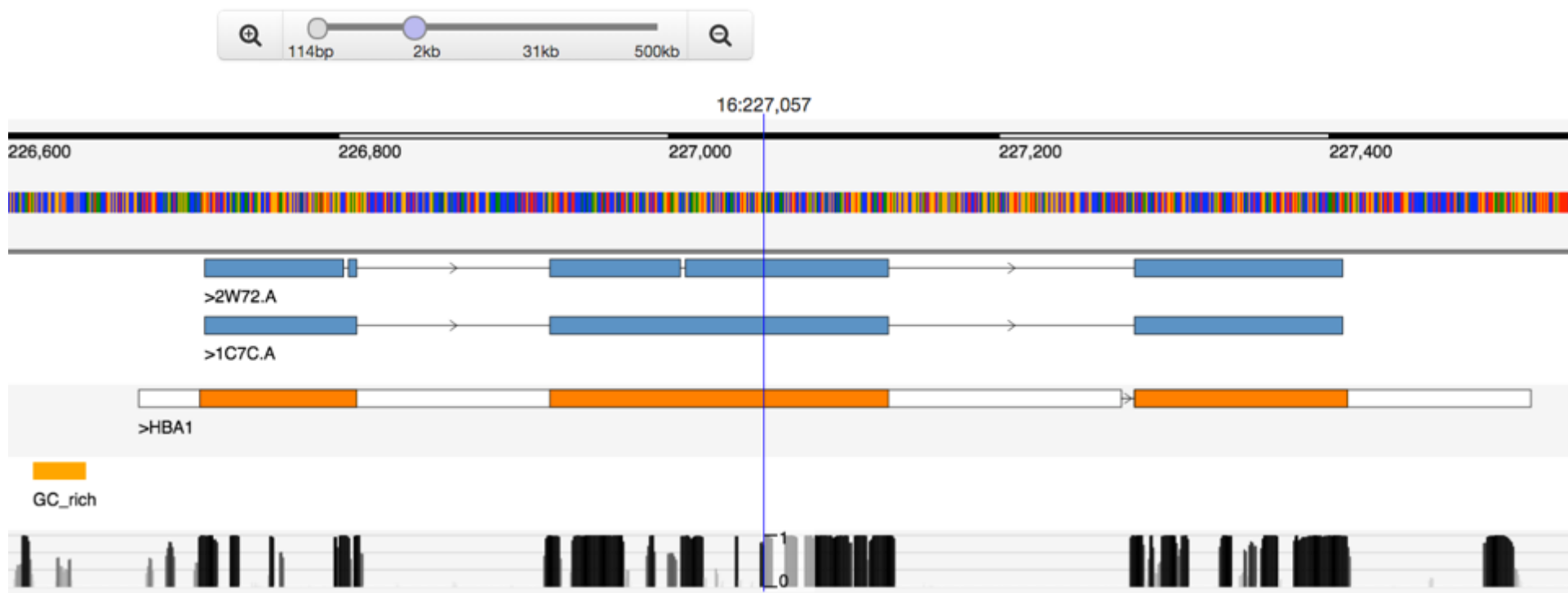
- ALPHA-HELICAL - [1836 Structures]
- BETA-BARREL - [343 Structures]
- Adventitious Membrane Proteins: Beta-sheet Pore-forming Toxins - [23 Structures]
- Beta-Barrel Membrane Proteins: Mitochondrial Outer Membrane - [5 Structures]
- Beta-Barrel Membrane Proteins: Monomeric/Dimeric - [151 Structures]
- Beta-Barrel Membrane Proteins: Porins and Relatives - [91 Structures]
- Lipopolysaccharide (LPS) Transport Proteins - [3 Structures]
- Omp85-TpsB Outer Membrane Transporter Superfamily - [33 Structures]
- Outer Membrane Autotransporters - [19 Structures]
- Outer Membrane Carboxylate Channels (Occ) - [18 Structures]
- MONOTOPIC MEMBRANE PROTEINS - [316 Structures]
- ADP-Ribosylation Factors - [5 Structures]
- Cyclooxygenases - [60 Structures]
- Cytochromes P450 - [2 Structures]
- Dehydrogenases - [12 Structures]
- Dihydroorotate Dehydrogenases (DHODH, class 2) - [62 Structures]
- Glycosyltransferases - [15 Structures]
- Hydrolases - [29 Structures]
- Isomerases - [6 Structures]
- Lipoxygenases - [34 Structures]
- Oxidases - [47 Structures]
- Oxidoreductases (Monotopic) - [12 Structures]
- Peptidases - [11 Structures]
- Polymerases - [5 Structures]
- Squalene-Hopene Cyclases - [16 Structures]



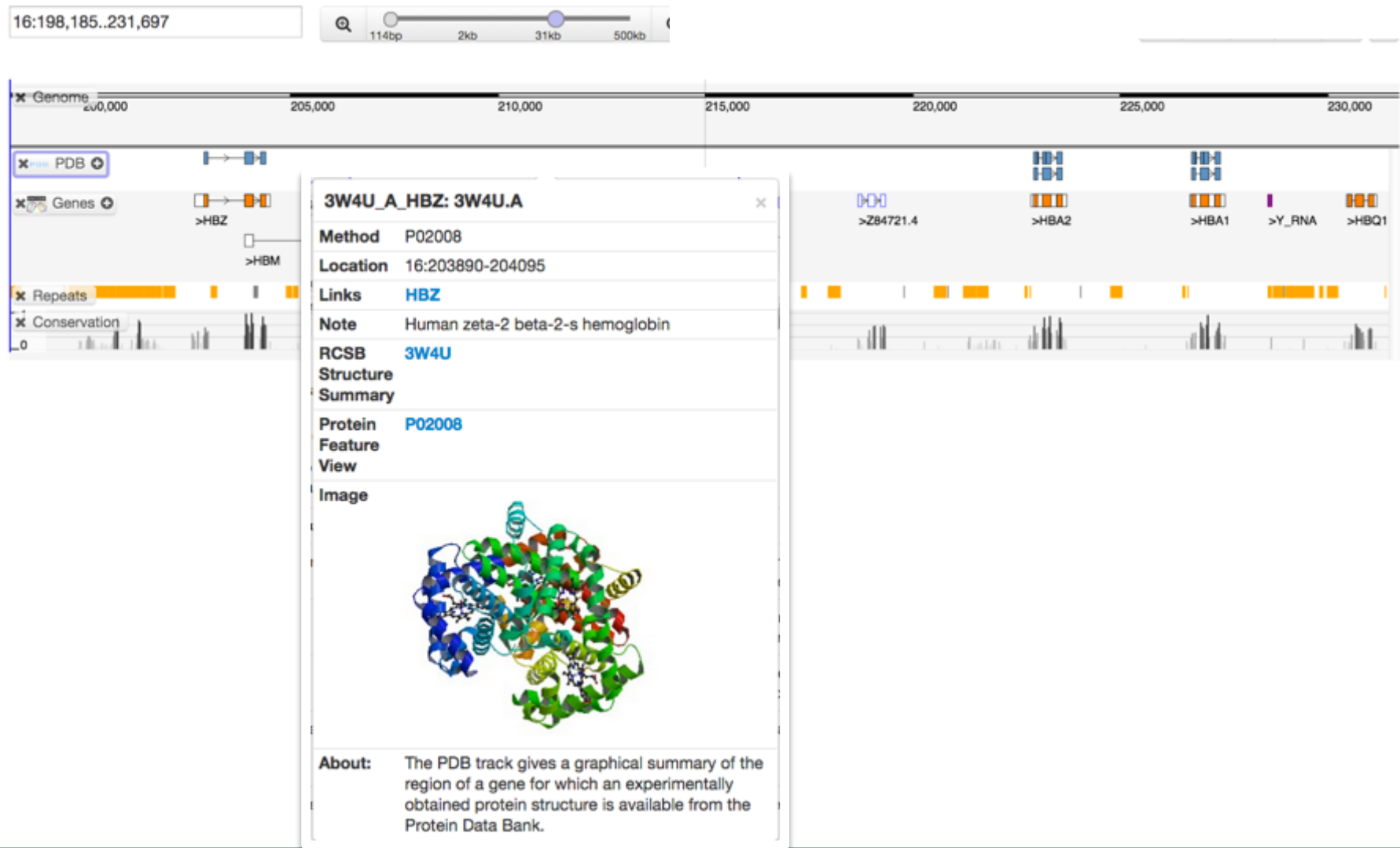
Gene View: Mapping PDB → Genes

HBA1 - hemoglobin, alpha 1

Number of PDB entities (unique chains) for this gene: [225](#)



Gene View → Scroll, Zoom, Interact



Live Demo of Gene View

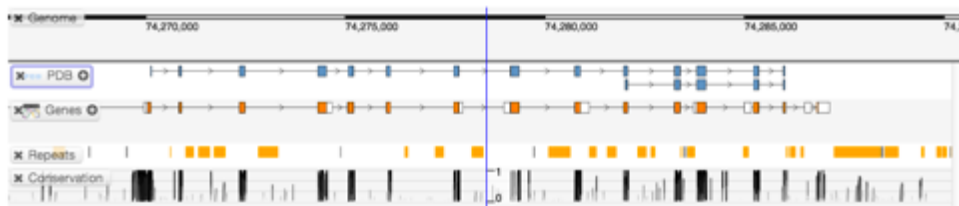
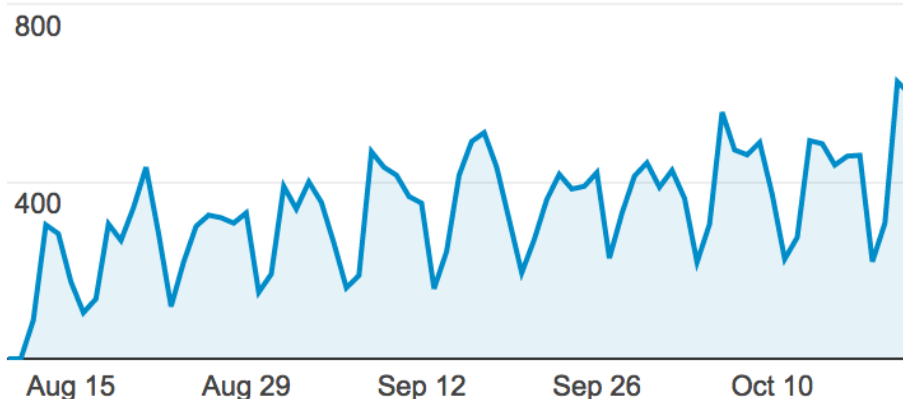
- Explaining page layout
- Viewing genomic context
- Scrolling along the length of a gene or a structure
- Zooming in on particular segments



Gene View → Good Initial Uptake

Pageviews - gene view

● Pageviews



Gene	Pageviews
insulin	269
hemoglobin alpha 1	200
KRAS	169
cytochrome C	146
albumin	115
myoglobin	107
lysozyme	98
hemoglobin beta	90
p53	88
superoxide dismutase 1	67

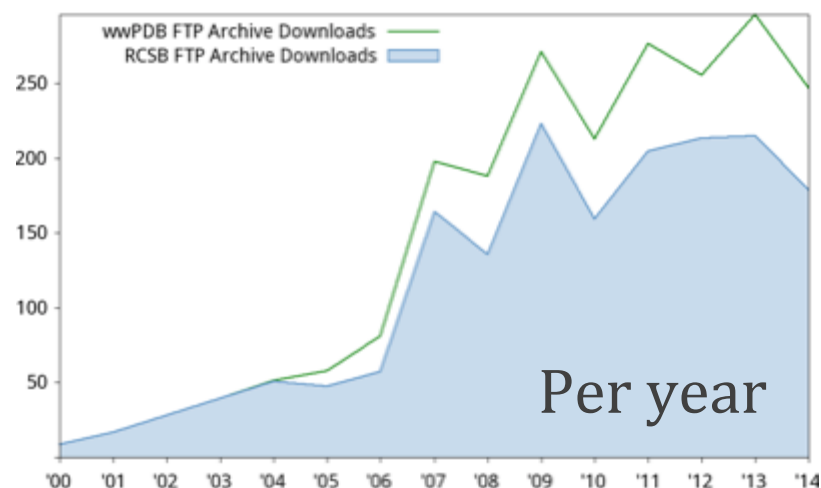


Assessing Impact

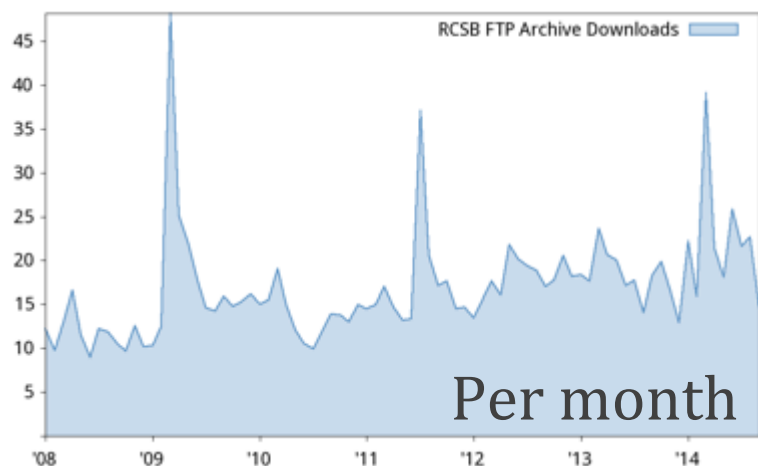
- Usage Metrics
- User Demographics
- Growth Areas
- Growth Opportunities
- Infrastructure Improvements



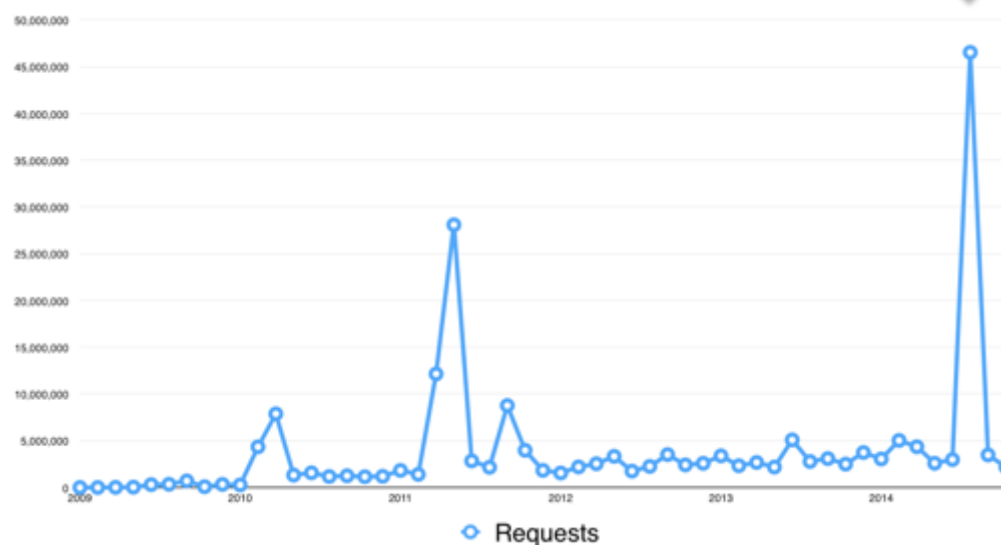
Overall FTP and Web Service Usage



FTP downloads



Argonne National Lab
search API



RESTful Web Services



Overall Website Usage

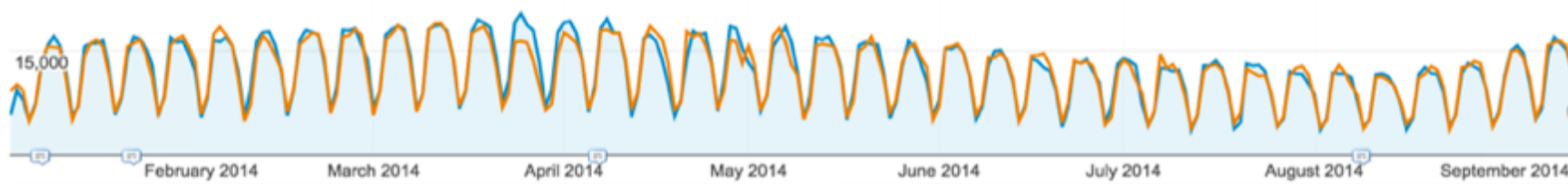
Sessions ▾ VS. [Select a metric](#)

2013→2014: ↑2% (2012→2013: ↑5%)

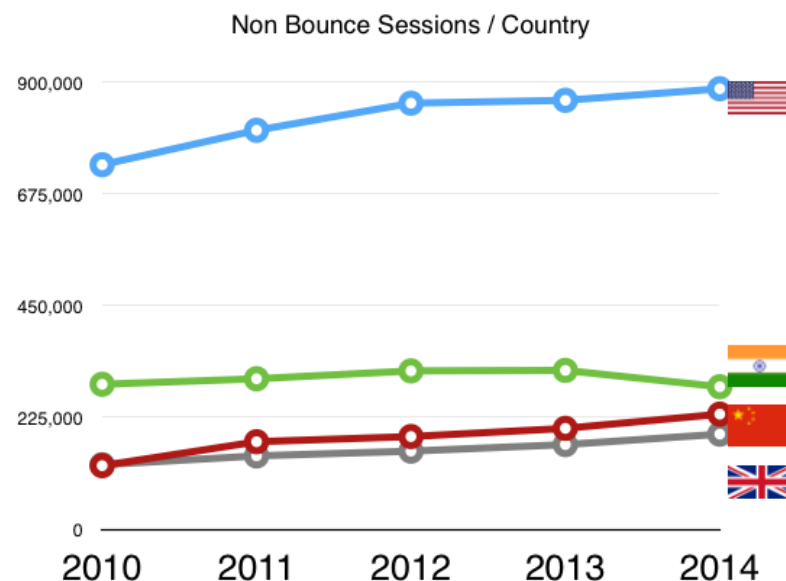
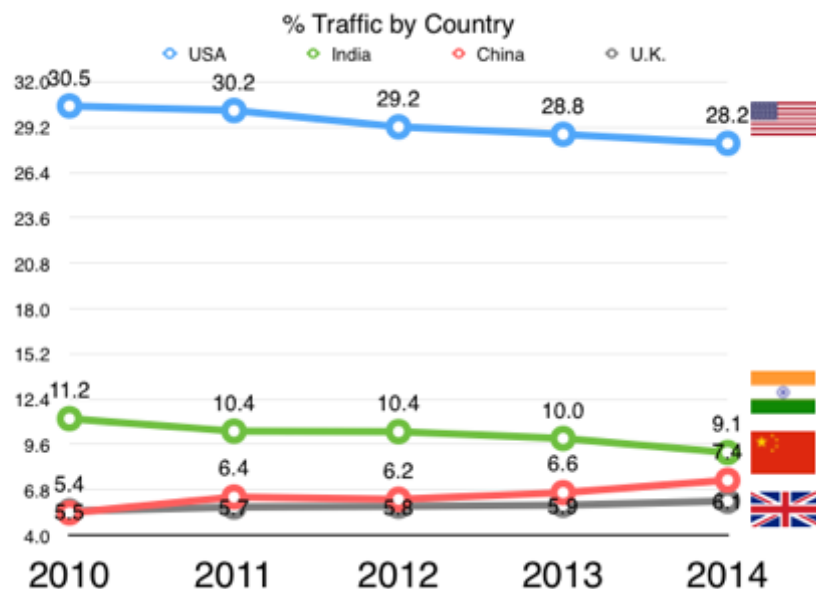
Jan 1, 2014 - Sep 17, 2014: ● Sessions

Jan 2, 2013 - Sep 18, 2013: ● Sessions

30,000

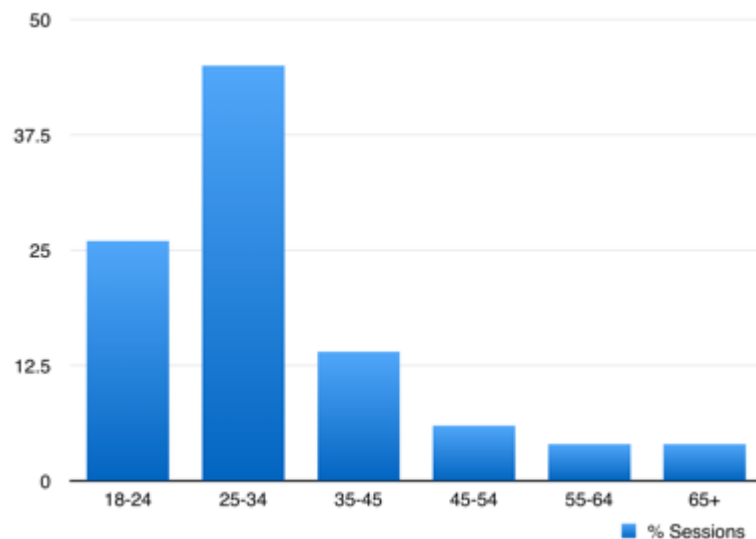


Non-bounce visits

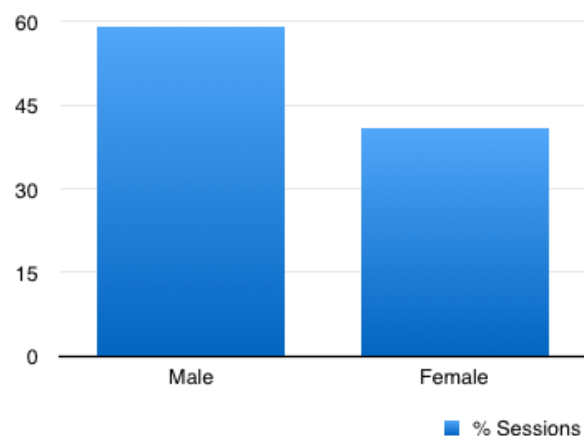


Available User Demographics

Age of Users



Gender



Age of Users

Age	% Sessions
18-24	26
25-34	45
35-45	14
45-54	6
55-64	4
65+	4

Gender

Gender	% Sessions
Male	59
Female	41



Growth Area: Jmol

Pageviews - Jmol

Jan 1, 2014 - Sep 17, 2014

654,903

% of Total: 2.82% (23,212,114)



Jan 2, 2013 - Sep 18, 2013

489,941

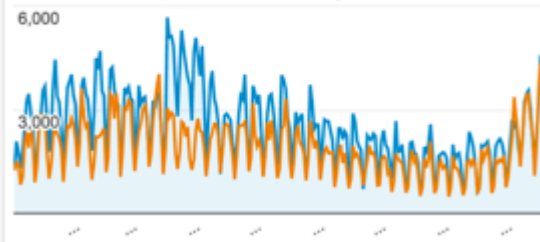
% of Total: 2.07% (23,655,681)



Pageviews - jmol

Jan 1, 2014 - Sep 17, 2014: ● Pageviews

Jan 2, 2013 - Sep 18, 2013: ● Pageviews



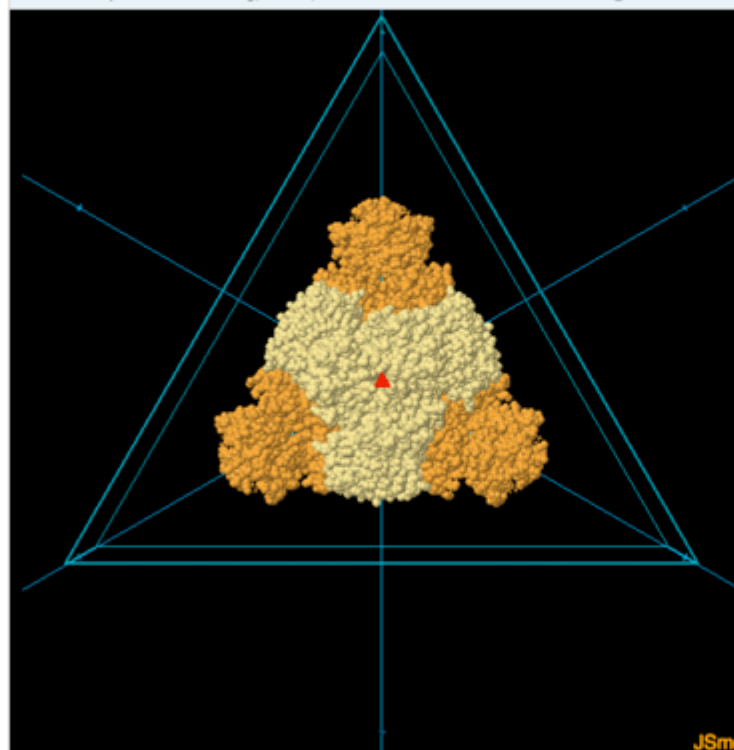
The R state structure of E. coli ATCase with ATP bound

4KGV

[Display Files](#) [Download Files](#)

Only CA atoms are being shown in order to display this large structure in Jmol.

NOTE: Use your mouse to drag, rotate, and zoom in and out of the structure. [?](#)



Biological assembly 1 assigned by authors and generated by PISA [?](#)

Structure Details [?](#)

Structure: [Biological Assembly 1](#)
Symmetry Type: [Global Symmetry](#)
Symmetry: D3
Stoichiometry: A6B6

Select Orientation

[Front C3 axis](#)

Select Display Mode [?](#)

Secondary Structure

Subunit

Symmetry

Symmetry View options

☒ Polyhedron ☒ Axes

Custom View



Growth Area: Protein Feature View

Pageviews -Protein feature View

Jan 1, 2014 - Sep 17, 2014

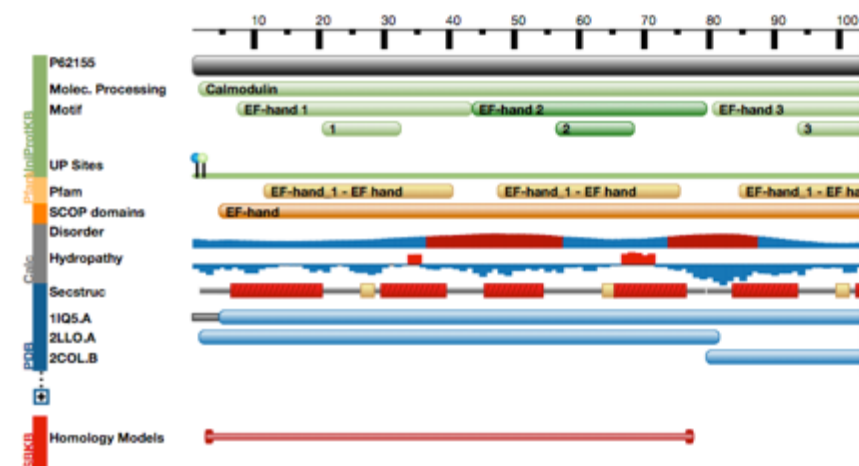
212,295

% of Total: 0.91% (23.212.114)

Jan 2, 2013 - Sep 18, 2013

153,100

% of Total: 0.65% (23,655,681)



Name Pageviews

Calmodulin 1010

Lysozyme C 859

Hemoglobin 856

GFP 700

Insulin 625

P53 534

Myoglobin 485

Endolysin 484

Serum albumin 448



Growth Area: Ligand Summary

Pageviews ligandsymmary.do

Jan 1, 2014 - Sep 17, 2014

194,542

Jan 2, 2013 - Sep 18, 2013

164,304

Visits - ligandSummary.do

Jan 1, 2014 - Sep 17, 2014

29,473

Jan 2, 2013 - Sep 18, 2013

11,778

7-[2-(4-FLUORO-PHENYL)-5-ISOPROPYL-3-PHENYL-4-PHENYLCARBAMOYL-PYRROL-1-YL]-3,5-DIHYDROXY-HEPTANOIC ACID

117 is found in 1 entries

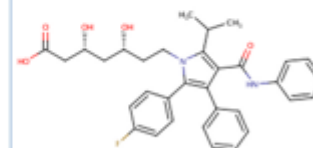
Chemical Component Summary ?		Hide
Name	7-[2-(4-FLUORO-PHENYL)-5-ISOPROPYL-3-PHENYL-4-PHENYLCARBAMOYL-PYRROL-1-YL]-3,5-DIHYDROXY-HEPTANOIC ACID	
Identifiers	(3R,5R)-7-[2-(4-fluorophenyl)-5-(1-methylethyl)-3-phenyl-4-(phenylcarbamoyl)-1H-pyrrol-1-yl]-3,5-dihydroxyheptanoic acid (3R,5R)-7-[2-(4-fluorophenyl)-3-phenyl-4-(phenylcarbamoyl)-5-propan-2-yl-pyrrol-1-yl]-3,5-dihydroxy-heptanoic acid	
Synonyms	ATORVASTATIN	
Formula	C ₃₃ H ₃₅ F N ₂ O ₅	
Molecular Weight	558.64 g/mol	
Type	non-polymer	
Isomeric SMILES	CC(C)C1C(C(=O)Nc2ccccc2)c(c(-c2ccc(F)cc2)n1CC(C@H)(O)C[C@H](O)CC(=O)O)=O)-c1ccccc1	
InChI	InChI=1S/C33H35FN2O5/c1-21(2)31-30(33(41)35-25-11-7-4-8-12-25)29(22-9-5-3-6-10-22)32(23-13-15-24(34)16-14-23)36(31)18-17-26(37)19-27(38)20-28(39)40/h3-16,21,26-27,37-38H,17-20H2,1-2H3,(H,35,41)(H,39,40)/t26-,27-/m1/s1	
InChI key	XUKUURHRXDUEBC-KAYWLYCHSA-N	

Drug Info: DrugBank		Hide
DrugBank ID	DB01076 (Different stereochemistry)	
Name	Atorvastatin	
Groups	approved	
Description	Atorvastatin (Lipitor) is a member of the drug class known as statins. It is used for lowering cholesterol. Atorvastatin is a competitive inhibitor of hydroxymethylglutaryl-coenzyme A (HMG-CoA) reductase, the rate-determining enzyme in cholesterol biosynthesis via the mevalonate pathway. HMG-CoA reductase catalyzes the conversion of HMG-CoA to mevalonate. Atorvastatin acts primarily in the liver. Decreased hepatic cholesterol levels increases hepatic uptake of cholesterol and reduces plasma cholesterol levels.	
Synonyms	Lipovastatinklonal	
Salts	Atorvastatin Calcium	
Brand names	- Atogal - Cardyl - Faboxim - Hipolixan - Lipitor [more]	
Affected organism	Humans and other mammals	
Indication	May be used as primary prevention in individuals with multiple risk factors for	

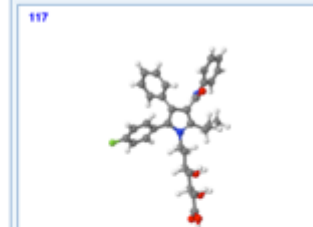
117

Display Files +
Download Files +

Ligand Image



Large image - View atom labels



Spin Hydrogens Labels

Instances in PDB Entries Hide
As free ligands: 1 entries
Ex: 1HWK

Opportunity: Structure Summary

Pageviews - explore.do

Jan 1, 2014 - Sep 17, 2014

6,551,494

Jan 2, 2013 - Sep 18, 2013

6,813,348

Visits - explore.do

Jan 1, 2014 - Sep 17, 2014

991,829

Jan 2, 2013 - Sep 18, 2013

1,245,275

- 2014 Structure Summary Pageviews slightly down from 2013
- Plans Forward:
 - Improve searching to ↑ traffic
 - Monitor impact

THE CRYSTAL STRUCTURE OF HUMAN DEOXYHAEMOGLOBIN AT 1.74 ANGSTROMS RESOLUTION

DOI:10.2210/pdb4hhb/pdb

ENTRY 4HHB SUPERSEDES 1HHB

Primary Citation

The crystal structure of human deoxyhaemoglobin at 1.74 Å resolution.

Fermi, G., Perutz, M.F., Shaanan, B., Fourme, R.

Journal: (1984) J.Mol.Biol. 175: 159-174

PubMed: 6726807

Search Related Articles in PubMed

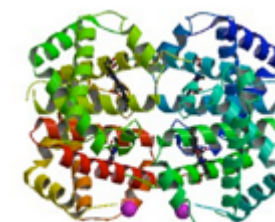
PubMed Abstract:

The structure of human deoxyhaemoglobin was refined at 1.74 Å resolution using data collected on film at room temperature from a synchrotron X-ray source. The crystallographic R-factor is 16.0%. The estimated error in atomic positions is 0.1 Å overall, 0.14 Å for main-chain atoms of internal segments, and 0.05 Å for the iron atoms. The effects of intermolecular contacts on the structure were investigated; such contacts cause only highly localized distortions, as judged from the degree of molecular asymmetry that they induce. The geometry of the iron-nitrogen complex closely resembles that of the deoxymyoglobin structure of Takano (1977) and of the 5-co-ordinated model compounds of Hoard (1975) and Jameson et al. (1980). The distance of the iron from the mean plane of N(porphyrin) is 0.40(5) Å and 0.36(5) Å, respectively, at the alpha and beta haems, in contrast to the corresponding distance of +0.12(8) Å and -0.11(8) Å in oxyhaemoglobin (Shaanan, 1983); the Fe-N epsilon (F8) bond length is 2.12(4) Å and the Fe-N(porphyrin) bond length is 2.06(2) Å; the last is also in good agreement with extended X-ray fluorescence spectroscopy measurements on deoxyhaemoglobin (Eisenberger et al., 1978; Perutz et al., 1982). The haems are domed toward the proximal side; the separation between the mean planes of N(porphyrin) and C(porphyrin) being 0.16(6) Å and 0.10(6) Å, respectively at the alpha and beta haems. At the alpha haems, the normals to the mean

4HHB

Display Files
Download Files

Biological Assembly



3D View

More Images...

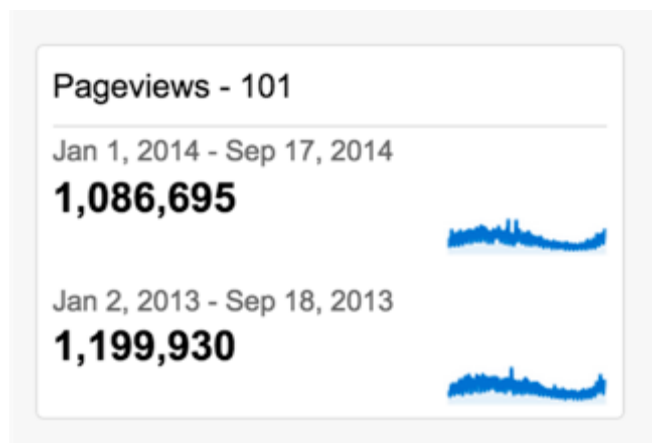
Symmetry: C2 view

Stoichiometry: Hetero 4-mer - A2B2
Biological assembly 1 assigned by authors and generated by PISA (software)

Downloadable viewers:

Simple Viewer Protein Workshop
Kiosk Viewer

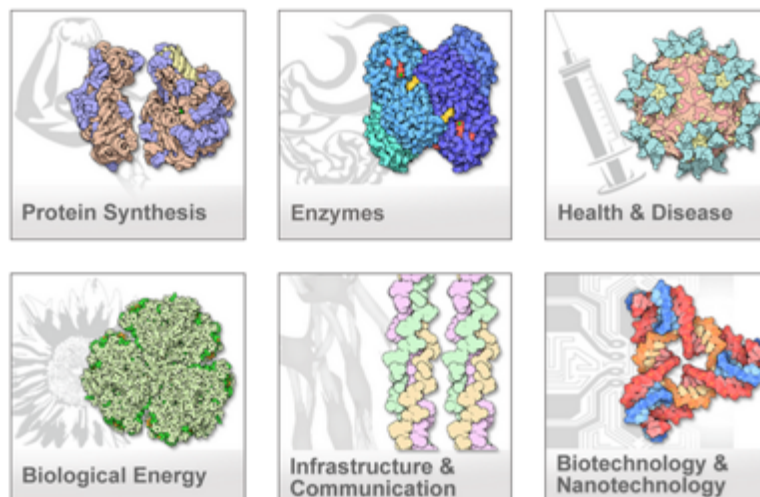
Opportunity: PDB-101



- 2014 page views slightly down from 2013
- Plans Forward
 - Reposition to ↑ traffic
 - Redesign to enhance usability
 - Monitor impact

Structural View of Biology

Select one of the key topics below to start exploring. Each subcategory leads related *Molecule of the Month* articles and examples of proteins and nucleic



Continuous Integration Software Development Process

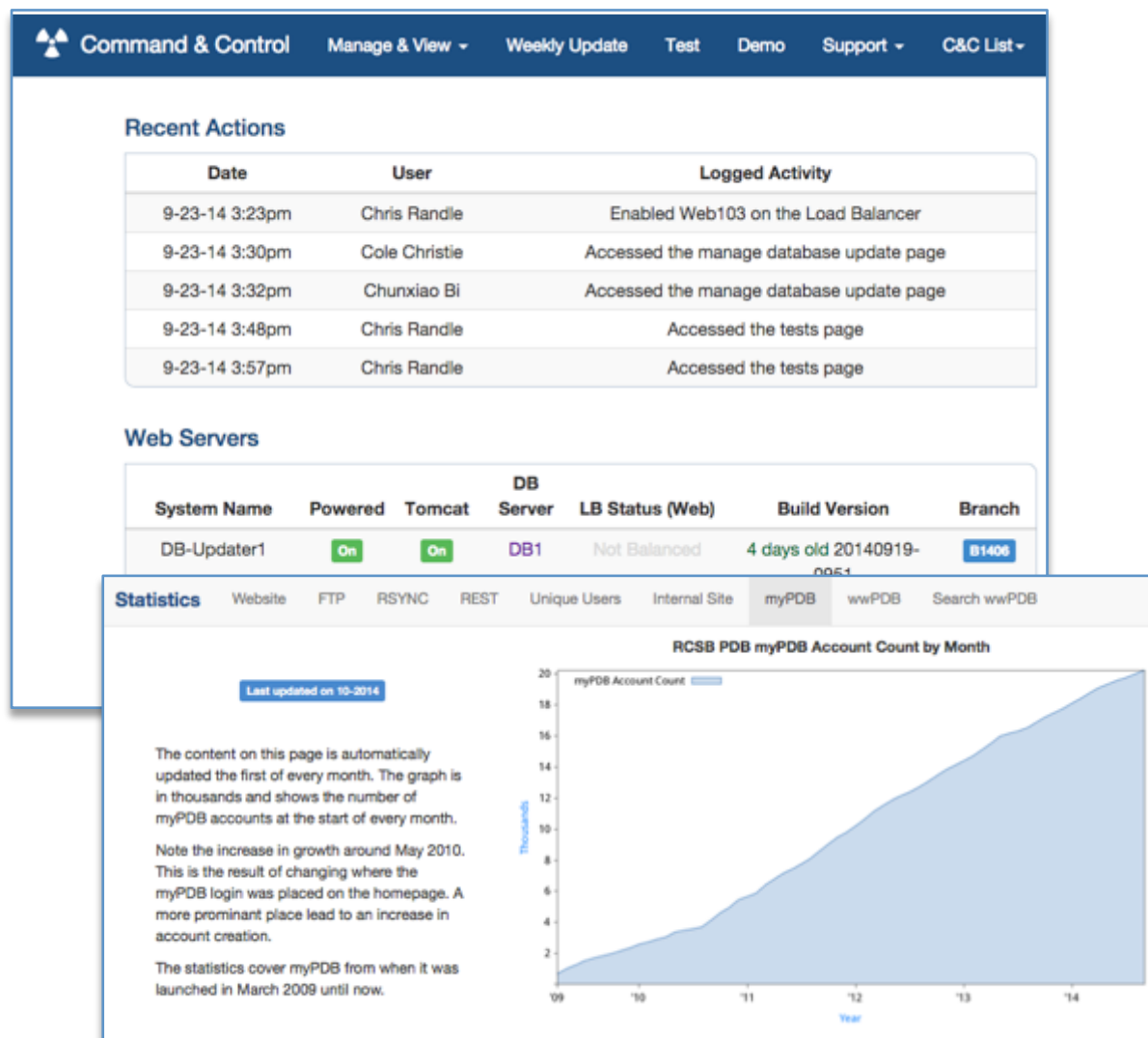


Image source: <http://www.meranetworks.com/services/processes/integration>



Automation Improvements

- Weekly update
- Software deployments
- Monitoring of services
- Additional usage statistics



Plans for 2014/2015

2014

- Deploy redesigned RCSB PDB website (Phase 1)
- Support large structures deployment (December 10th)

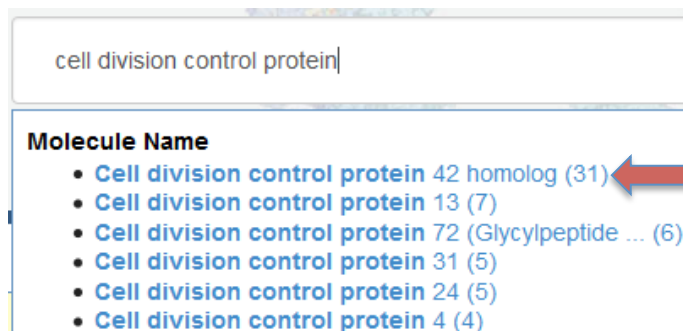
2015

- Scientific projects
 - Expand synonym search using sequence clusters (Q1 2015)
 - Visualization
 - Validation data (geometry, e-density)
 - Membrane protein orientation
 - Structure mapping
 - Protein Kinase, GPCR, and Nuclear Hormone Receptor phylogenetic trees
 - Signal transduction pathways
- Technical projects
 - Website redesign Phase 2, including PDB-101
 - Improve scalability of backend calculations and data integration
 - Quantify PDB impact through citation network analysis

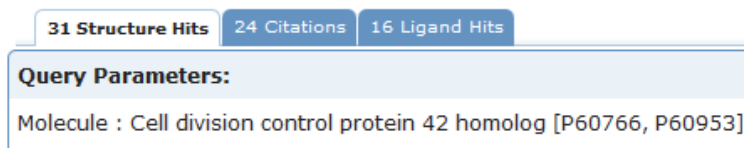


Molecule Name Searching Limitations

- Based on UniProt annotation
 - Consistent and unambiguous naming: www.uniprot.org/docs/nameprot
 - Search suggestion



- Select *Cell division control protein 42 homolog*
 - 1. Find all UniProt entries with a matching name -> P60766, P60953



- 2. Retrieve all structures with UniProt IDs: P60766, P60953
- Limitation: Searches using Depositor-assigned protein name
 - GTP-BINDING RHO-LIKE PROTEIN (Biologists love synonyms!)
 - Some typographical inconsistencies and errors inevitable



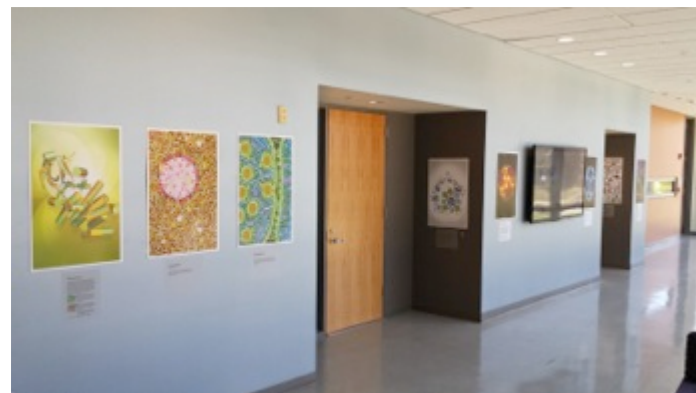
Sequence-based Synonym Searching

- Search for *GTP-BINDING RHO-LIKE PROTEIN* → 1 hit
- Archive contains multiple names for the same protein
 - Example: 90% Sequence Identity Cluster (31 entries, 13 synonyms)
 - Cell division control protein 42 homolog (UniProt name)
 - cell division cycle 42 isoform 1
 - PROTEIN (GTP BINDING PROTEIN (G25K))
 - PROTEIN (GTP BINDING PROTEIN)
 - G25K GTP-binding protein, placental isoform
 - CDC42 HOMOLOG
 - Human Cell Division Cycle 42 (CDC42)
 - GTP-BINDING PROTEIN
 - G25K GTP-BINDING PROTEIN
 - CDC42HS
 - CDC42HS-GDP
 - G25K GTP-BINDING PROTEIN, PLACENTAL ISOFORM (GP),
 - PROTEIN (CDC42 HOMOLOG)
 - GTP-BINDING RHO-LIKE PROTEIN (protein name used for text search)
- These proteins can be found by current text search, but results will be incomplete due to inconsistent and partial annotation
- Not practical to remediate protein names for >100,000 entries
- **Our approach: Attribute all names to all members of each high sequence identity cluster to improve accuracy of text-based searching**

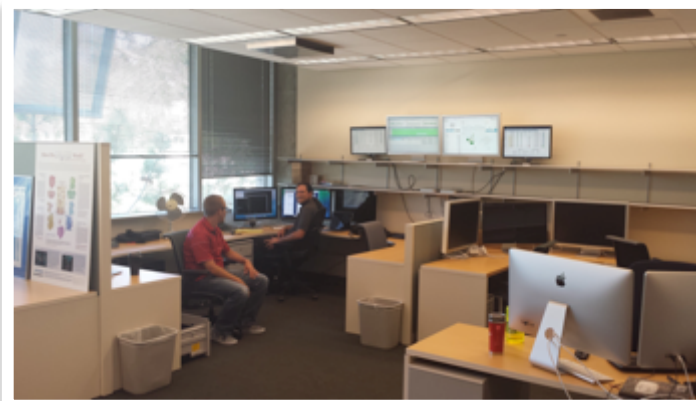


New Office Space at UCSD

- Moved to larger space at 2nd floor of Skaggs School of Pharmacy and Pharmaceutical Sciences



- Provides room for additional staff, students, and postdocs



RCSB PDB UCSD Team



Job Openings at UC San Diego

- Senior Java Web Developer
- Senior Scientist

