
RCSB Protein Data Bank Advisory Committee Meeting

November 3rd 2015



Agenda

State of the RCSB PDB Stephen K. Burley

Integrative/Hybrid Methods Helen M. Berman

Data In John Westbrook
Jasmine Young

Lunch and Demonstrations

Data Out Peter Rose
Andreas Prlić

Education Plan Shuchismita Dutta
PDB-101 David Goodsell

Management Stephen K. Burley

Telephonic Discussion Peter McCartney, NSF

Matters Arising & General Discussion



State of the RCSB PDB

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



RCSB PDB Vision & Mission



VISION

To enable breakthroughs in scientific inquiry, medicine, drug discovery, technology, and education with rich structural views of biological systems.



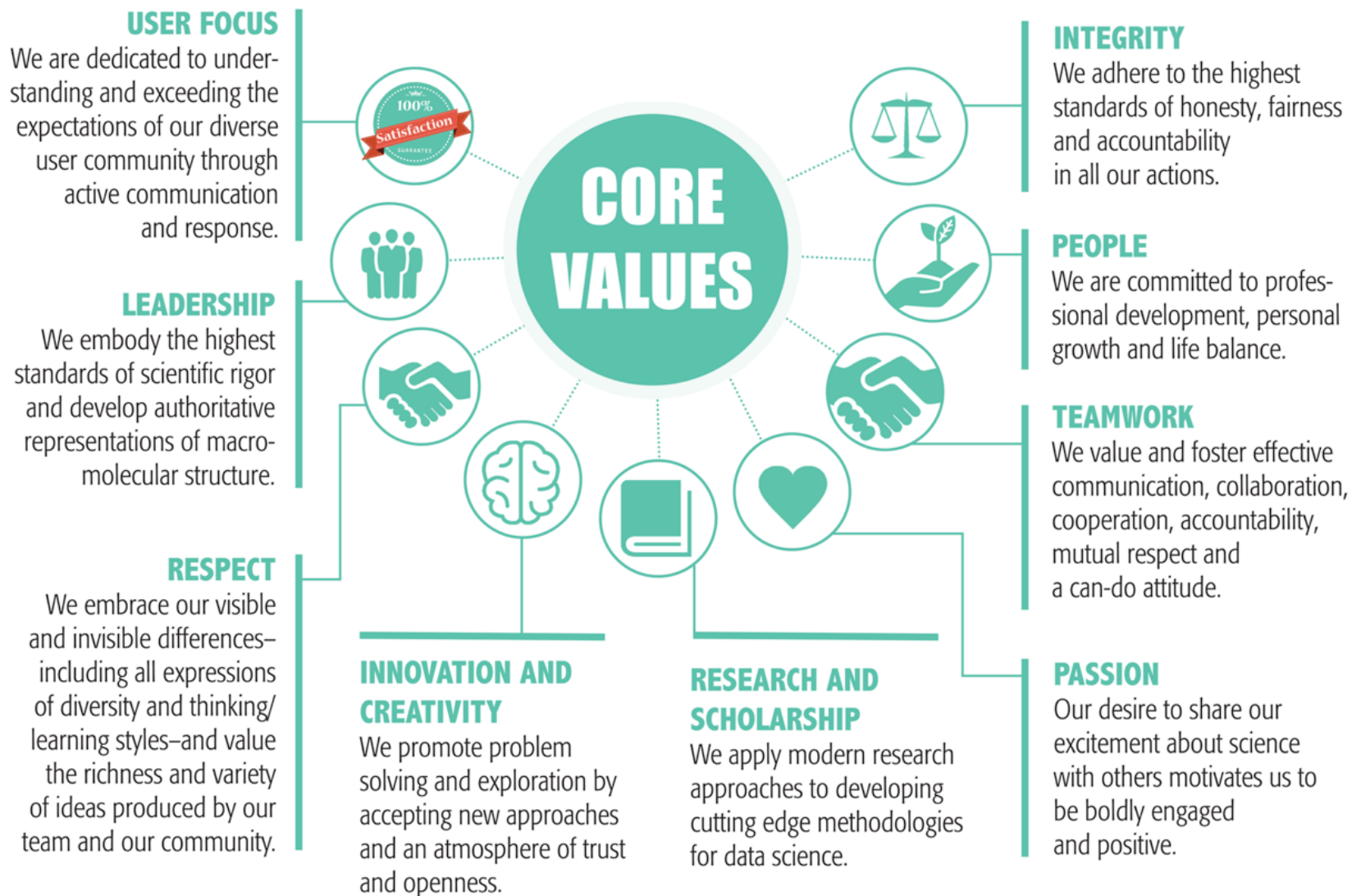
MISSION

To provide a freely-available and sustainable resource of experimental 3D biological macromolecular structure data that is by, for, and of the community by

- Enabling efficient deposition, high-quality curation, and exploration of data in the global PDB archive
- Leading 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



RCSB PDB Core Values



RCSB PDB is a Member of the wwPDB

- Ensures security of the PDB Archive and unrestricted global access to PDB data
- Regional Data Centers
 - RCSB PDB
 - PDBj (Osaka University)
 - PDBe (EMBL-EBI)
 - BioMagResBank (University Wisconsin, Madison)
- Institutional agreement in place
- Formalized procedures for deposition, validation, metadata representation, and annotation
- Regional Data Centers collaborate on Data In and operate independent Data Out services



RCSB PDB Advisory Committee

- Responsible for providing independent advice to the RCSB PDB Director and staff on current and pending issues of policy, operations, technical implementation, and project performance
- Updated Terms of Reference
- Members appointed for 3 years (renewable)
- Service concluded by Jack Chirikjian, Andrzej Joachimiak, Ann Palmenberg, David Searls, Cathy Wu
- New members to enhance input from Industry, cryoEM, Small Angle Scattering, and Integrative/Hybrid Methods (Bridget Carragher to join in 2016)
- Cynthia Wolberger to serve as Chair through 2019



Global Advisory Mechanisms

- Worldwide PDB (wwPDB) Advisory Committee
 - Chair: Byrd (NCI Frederick)
- Task Forces, Working Groups, Workshops
 - X-ray Validation: Read (Cambridge)
 - 3DEM Validation: Henderson (MRC-LMB), Sali (UCSF)
 - NMR Validation: Montelione (Rutgers), Nilges (Pasteur)
 - PDBx/mmCIF: Adams (LBL)
 - Small-Angle Scattering: Trewhella (Sydney/Utah)
 - Integrative/Hybrid Methods: Sali (UCSF), Schwede (Basel), Trewhella (Sydney/Utah)
 - Ligand Validation: wwPDB/CCDC/D3R



Response to 2014 Report

AC Recommendation		Response
Pursue funding to develop approaches for supporting data from Integrative/Hybrid Methods	➔	Proposals submitted; described in next presentation
Terminate the legacy deposition system (ADIT)	➔	ADIT retired on July 19, 2015 for structures determined <i>via</i> X-ray crystallography
Continue to provide mobile-friendly services	➔	Recent redesign efforts (Structure Summary and PDB-101) are responsive to display device form factor
Develop a focused Education Plan	➔	Plan described in afternoon presentation
Make more information available on unpublished structures	➔	Needs discussion with wwPDB and community stakeholders



Data In Milestones: Oct 2014-Sept 2015

- Archive Growth
 - 8,386 entries deposited in 2015
 - >112,000 structures available in the archive
- PDBx/mmCIF format implemented in all major X-ray software packages
- Large structures fully integrated with PDB Archive as single PDBx/mmCIF file entries
- wwPDB D&A X-ray Data In Pipeline
 - More than 10,200 structures deposited since deployment
 - Improvement in biocuration time: Median: 16.5 → 1.6 days
 - Infrastructure improvements
- New content provided to enable blinded docking competition
- Workshops and Meetings



Data Out Milestones: Oct 2014-Sept 2015

- Visualization and query support for large structures represented in PDBx/mmCIF
- Validation and mutation information mapped to sequence (Protein Feature View)
- Improved usability of Structure Summary pages now available for all PDB entries
- Infrastructure and process improvements
- Workshops and Meetings



Education Milestones: Oct 2014-Sept 2015

- Developed focused Education Plan
- Pilot testing of HIV/AIDS curriculum
- Modular Curriculum materials created for
 - Biomolecular Structures and Models
 - Molecular Immunology
 - HIV/AIDS

The screenshot displays the PDB Education website interface. On the left, a sidebar menu for 'Biomolecular Structures and Models' includes links for 'Getting Started', 'Learning Objectives', 'Learning Materials', 'Activities', and 'Evaluation'. The main content area is titled 'Getting Started' and features a large image of various molecular structures with a 'BEGIN EXPLORING' button. To the right, a text box instructs users to launch an interactive view of PDB's molecular machines, select a structure, access the entry in 3D, read a brief summary of the molecule's biological role, and access the corresponding PDB entry and Molecule of the Month column.

Below the 'Getting Started' section is a 'How To' table with the following content:

How To	
Read a Scientific Paper	Guide Quick review
Make sense of the Information Out There	Guide and exercise Teaching Notes Additional Reading
Explore a Topic of Interest at a Molecular Level	Guide Quick Review Worksheet Learning about Asthma
Find Structures Relevant to a Topic of Interest	Guide Quick Review Selected structures related to Asthma
Visualize, Analyze and Compare Selected Structures	Cheat Sheet Basics Menus Selections Structure Analysis Structure Comparisons
Make Publication Quality Images and Movies	Guide Part1 Part2



Outreach Milestones: Oct 2014-Sept 2015

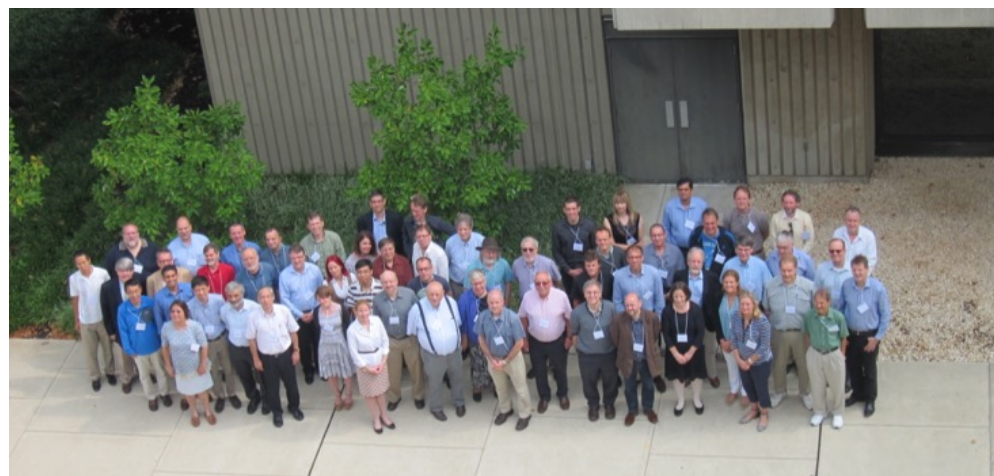
- PDB-101 website redesign project
- Collaboration with HHMI to establish Geis Digital Archive begins
- Collaborative development of Science Olympiad events in protein modeling across US
- 7 peer-reviewed publications
- Targeted participation at professional society meetings and other events
- Cross-site retreats for outreach and software development



wwPDB/CCDC/D3R Ligand Validation Workshop

Meeting Objectives: To bring together co-crystal structure determination experts from Academe and Industry with X-ray Crystallography and Computational Chemistry Software Developers to discuss, develop, and recommend:

- Best practices for PDB archive deposition/validation of co-crystal structures
- Editorial/Refereeing/Publication standards for co-crystal structures
- Improvements in ligand representation across the PDB Archive



Rutgers July 30-31, 2015

Outcome of the first wwPDB/CCDC/D3R Ligand Validation Workshop

Paul D. Adams¹, Kathleen Aertgeerts², Cary Bauer³, Jeffrey A. Bell⁴, Helen M. Berman^{5,6}, Talapady N. Bhat⁷, Jeff Blaney⁸, Evan Bolton⁹, Gerard Bricogne¹⁰, David Brown¹¹, Stephen K. Burley^{5,6,12,*}, David A. Case⁶, Kirk L. Clark¹³, Thomas Darden¹⁴, Paul Emsley¹⁵, Victoria Feher^{16,*}, Zukang Feng^{5,6}, Colin R. Groom^{17,*}, Seth F. Harris⁸, Jorg Hendle¹⁸, Thomas Holder⁴, Andrzej Joachimiak¹⁹, Gerard Kleywegt^{20,*}, Tobias Krojer²¹, Joseph Marcotrigiano^{6,22}, Alan E. Mark²³, John L. Markley^{24,*}, Matthew Miller²², Wladek Minor²⁵, Gaetano T. Montelione^{22,26}, Garib Murshudov¹⁵, Atsushi Nakagawa²⁷, Haruki Nakamura^{27,*}, Anthony Nichols¹⁴, Marc Nicklaus²⁸, Robert Nolte²⁹, Anil K. Padyana³⁰, Catherine E. Peishoff²⁹, Susan Pieniazek³¹, Randy J. Read³², Chenghua Shao⁵, Steven Sheriff³³, Oliver Smart²⁰, Stephen Soisson³⁴, John Spurlino³⁵, Terry Stouch³⁶, Radka Svobodova³⁷, Wolfram Tempel³⁸, Tom Terwilliger³⁹, Dale Tronrud⁴⁰, Sameer Velankar²⁰, Suzanna Ward¹⁷, Greg Warren¹⁴, John D. Westbrook^{5,6}, Pamela Williams⁴¹, Huanwang Yang^{5,6}, and Jasmine Young^{5,6}

Mission Critical Project: wwPDB Deposition & Annotation System (D&A)

Goals: Create the next generation deposition and annotation system to support PDB data from X-ray, NMR, and 3DEM that will:

- Maximize data quality
 - Improve data validation
 - Standardize file formats
 - Ensure more complete data capture
- Support larger and more complex structures
- Improve efficiency and consistency
 - Use of same system across wwPDB to enable load balancing and expansion
 - Automation and validation of routine tasks

WORLDWIDE
PDB
PROTEIN DATA BANK



2007 Initial Discussions



2010 D&A Team Meeting



2014 wwPDB Biocurator Summit



2015 wwPDB Advisory Committee Meeting

October 2nd in Osaka, Japan

- wwPDB PIs reported D&A V2.0 (All Methods) September completion deadline missed
- Initial wwPDB AC Recommendations
 - Release of D&A Version 2.0 ASAP; No later than Jan 8, 2016
STATUS: On track for Nov 15 beta testing
Jan 8 deployment requires resolving issues with backwards compatibility with deposition system
 - Appoint global project manager ASAP and provide monthly project updates to wwPDB AC Chair
STATUS: Jasmine Young named Global Project Lead
1st update shared with wwPDB AC Chair
- Written wwPDB AC report pending



Agenda

State of the RCSB PDB Stephen K. Burley

Integrative/Hybrid Methods Helen M. Berman

Data In John Westbrook
Jasmine Young

Lunch and Demonstrations

Data Out Peter Rose
Andreas Prlić

Education Plan Shuchismita Dutta
PDB-101 David Goodsell

Management Stephen K. Burley

Telephonic Discussion Peter McCartney, NSF

Matters Arising & General Discussion

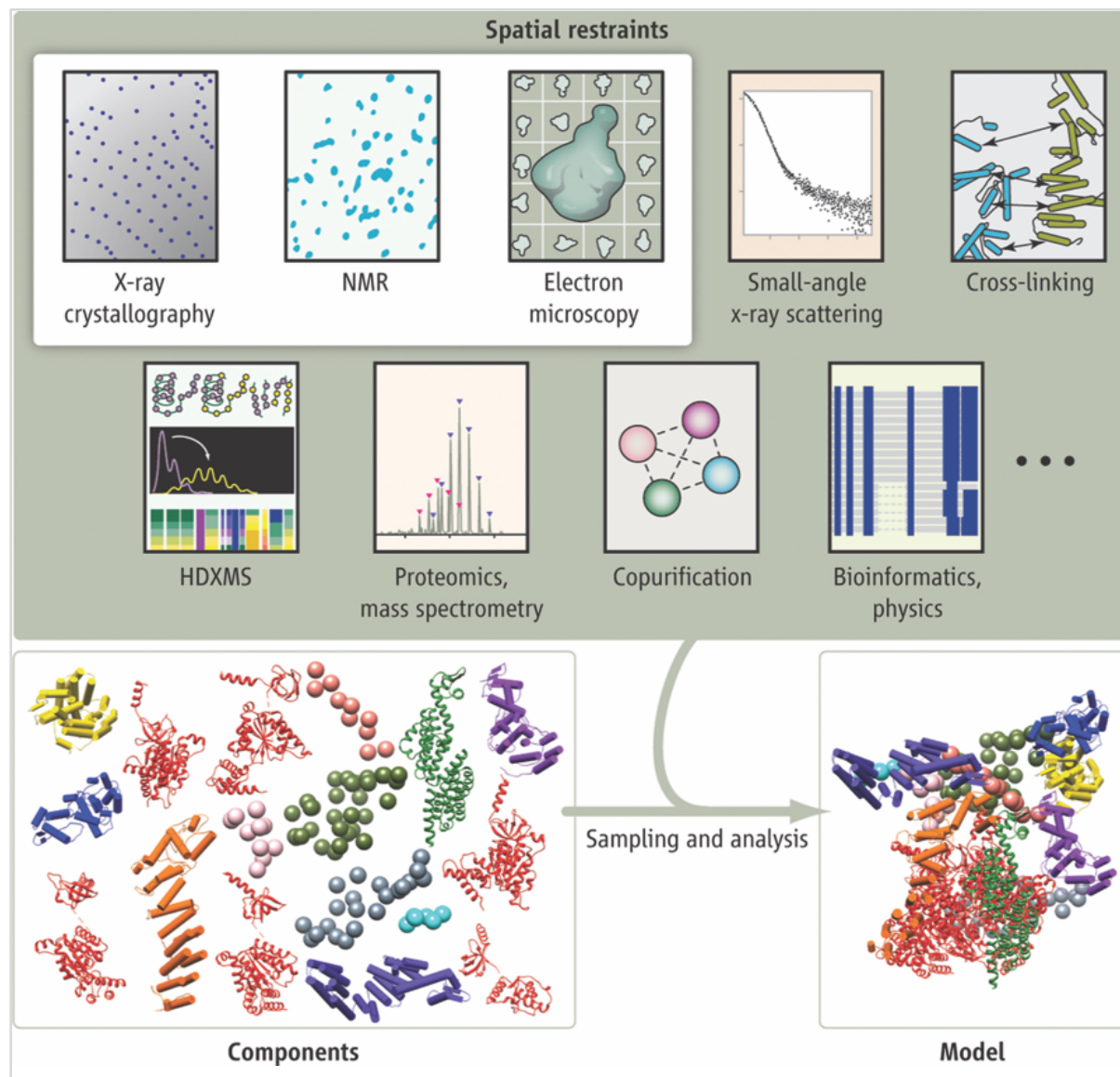


Addressing the Challenges of Integrative/Hybrid Methods

Helen M. Berman



Integrative Structural Biology

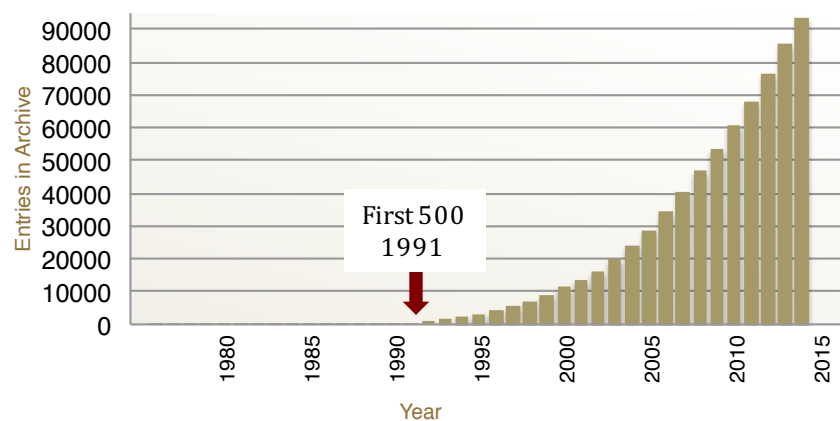


Ward *et al.* (2013)
Science **339**, 913-915

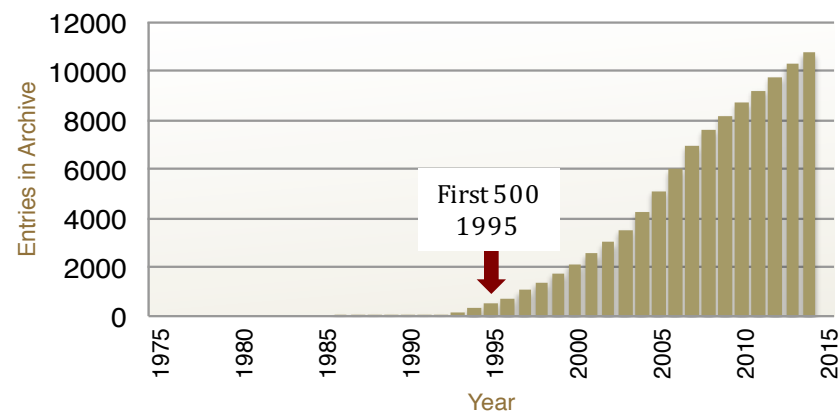


Experimental Methods Used for Structure Determination

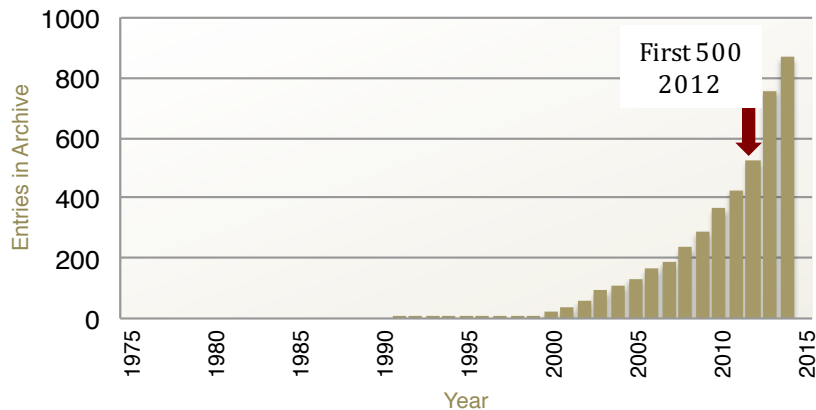
X-ray entries



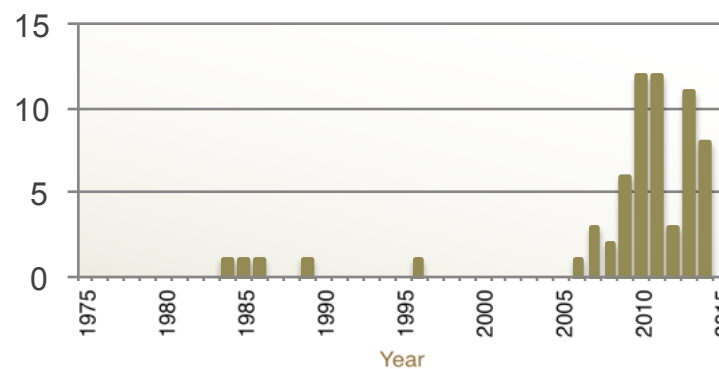
NMR entries



EM entries



Integrative/Hybrid entries



There is No Process to Handle Integrative/Hybrid Models

UCSF



SIB
Swiss Institute of
Bioinformatics

THE UNIVERSITY OF
SYDNEY

The evaluation and archiving of “integrative structural models” that draw on data from multiple experimental techniques and novel modelling approaches presents a number of challenges. To address these challenges, the Worldwide Protein Data Bank (wwPDB; wwpdb.org) has recently formed the Hybrid/Integrative Methods Task Force

[imaging-ups-its-game-1.16196](#)) and a report summarising the Task Force recommendations is being

The Task Force made a strong recommendation that an archive for integrative models be established. Integrative models that have already been submitted to the PDB are currently on-hold awaiting a policy decision. The Task Force recommends that such models should not be processed or archived in the PDB at this point in time. Instead, they should be transferred to and processed by the future integrative data and model archive once it has been established.

Therapeutic Sciences
Dept. of Pharmaceutical Chemistry
California Institute for Quantitative
Biosciences (QB3)
University of California
San Francisco, CA 94158-2330
USA

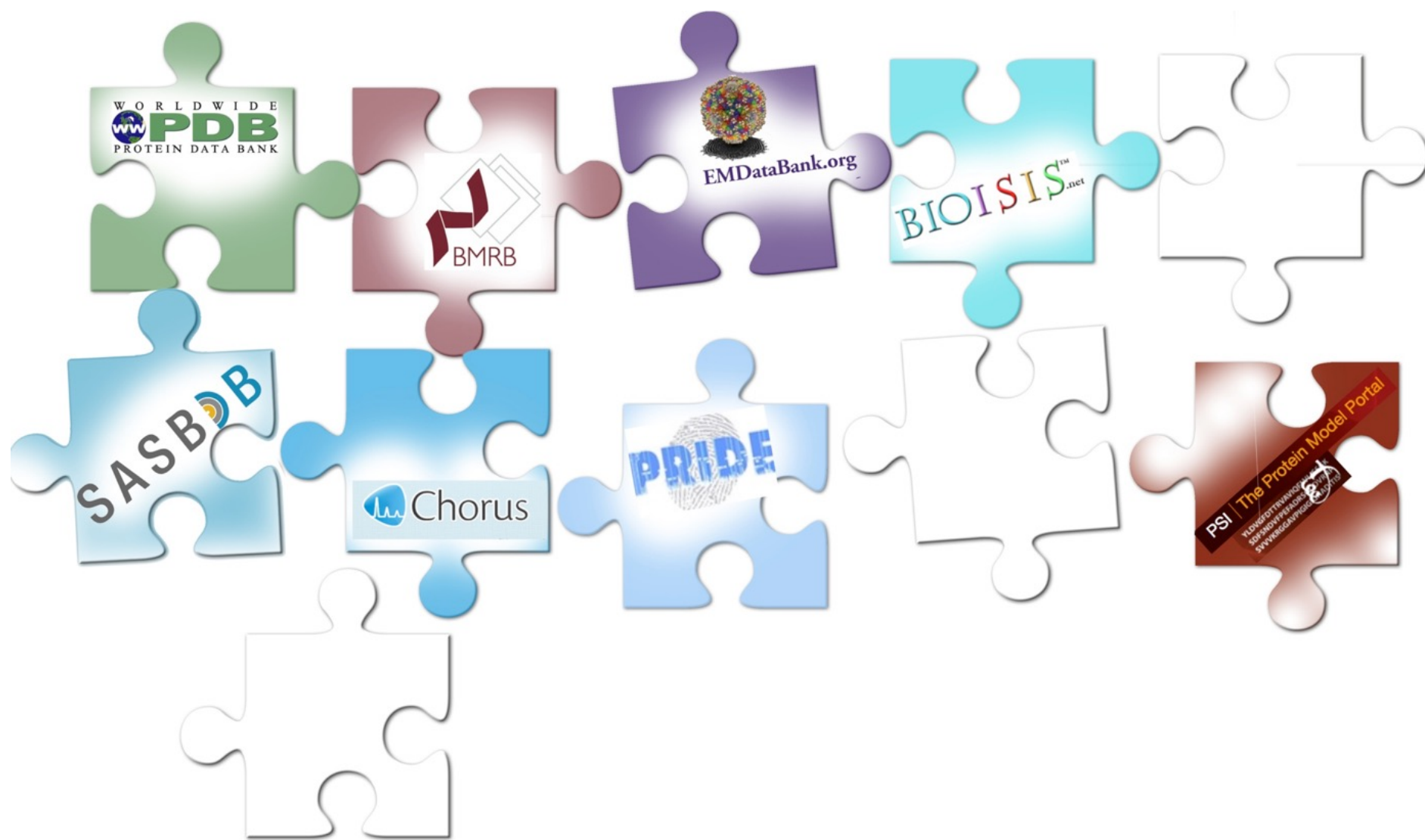
biozentrum, University of Basel
CH-4056 Basel
Switzerland

The University of Sydney
NSW, 2008
Australia

January 24, 2015



How Do We Assemble the Pieces of the Puzzle?



wwPDB Hybrid Methods Task Force



October 6-7, 2014 at EMBL-EBI, Hinxton



Discussion Points

- What experimental data (beyond X-ray, NMR, EM) should be archived, where, and how should it be validated?
- What kinds of non-atomistic models can we expect and how should they be validated?
- What are the criteria for deciding where models should be archived?
- How should non-atomistic and mixed atomistic/non-atomistic models be archived?
- Should there be an archive for integrative (mixed) models (and data)?
- Should we establish a federated system of data and model archives to support integrative structural biology?



Summary of Recommendations

- Archive structural models, data/metadata, and work flows
- Adopt flexible structure representation
- Assess structural model uncertainty
- Federate archives for structural models, and data/metadata, and work flow
- Establish publication standards

CellPress structure
Meeting Review

Outcome of the First wwPDB Hybrid/Integrative Methods Task Force Workshop

Andrej Sali,^{1,*} Helen M. Berman,² Torsten Schwede,³ Jill Trehella,⁴ Gerard Kleynwegt,⁵ Stephen K. Burley,^{2,6} John Markley,⁷ Haruki Nakamura,⁸ Paul Adams,^{9,10} Alexandre M.J.J. Bonvin,¹¹ Wah Chiu,¹² Matteo Dal Peraro,¹³ Frank Di Maio,¹⁴ Thomas E. Ferrin,¹⁵ Kay Grünewald,¹⁶ Aleksandras Gutmanas,⁵ Richard Henderson,¹⁷ Gerhard Hummer,¹⁸ Kenji Iwasaki,¹⁹ Graham Johnson,²⁰ Catherine L. Lawson,² Jens Meiler,²¹ Marc A. Marti-Renom,²² Gaetano T. Montellone,^{23,24} Michael Nilges,^{25,26} Ruth Nussinov,^{27,28} Ardan Patwardhan,⁵ Juri Rappsilber,^{29,30} Randy J. Read,³¹ Helen Saibil,³² Gunnar F. Schröder,^{33,34} Charles D. Schwieters,³⁵ Claus A.M. Seidel,³⁶ Dmitri Svergun,³⁷ Maya Topf,³² Eldon L. Ulrich,³⁸ Sameer Velankar,⁵ and John D. Westbrook²

Structures of biomolecular systems are increasingly computed by integrative modeling that relies on varied types of experimental data and theoretical information. We describe here the proceedings and conclusions from the first wwPDB Hybrid/Integrative Methods Task Force Workshop held at the European Bioinformatics Institute in Hinxton, UK, on October 6 and 7, 2014. At the workshop, experts in various experimental fields of structural biology, experts in integrative modeling and visualization, and experts in data archiving addressed a series of questions central to the future of structural biology. How should integrative models be represented? How should the data and integrative models be validated? What data should be archived? How should the data and models be archived? What information should accompany the publication of integrative models?

Background
Historical Rationale for the Workshop

The PDB (<http://wwpdb.org>) was founded in 1971 with seven protein structures as its first holdings (Protein Data Bank, 1971). The global PDB archive now holds more than 100,000 atomic structures of biological macromolecules and their complexes, all of which are freely accessible. Most structures in the PDB archive (~90%) have been determined by X-ray crystallography, with the remainder contributed by two newer 3D structure determination methods, nuclear magnetic resonance (NMR) spectroscopy and 3D electron microscopy (3DEM).

Considerable effort has gone into understanding how to best curate the structural models and experimental data produced with these methods. Over the past several years, the Worldwide PDB (wwPDB; the global organization responsible for maintaining the PDB archive) (Gorman et al., 2003) has established expert, method-specific task forces to advise on which experimental data and metadata from each method should be archived and how these data and the resulting structure models should be validated. The wwPDB X-ray Validation Task Force (VTF) made detailed recommendations on how to best validate structures determined by X-ray crystallography (Read et al., 2011). These

¹Department of Bioengineering and Therapeutic Sciences, Department of Pharmaceutical Chemistry, California Institute for Quantitative Biosciences, Byers Hall Room 503B, University of California, San Francisco, 1700 4th Street, San Francisco, CA 94158-2330, USA
²Research Collaboratory for Structural Bioinformatics Protein Data Bank, Center for Integrative Proteomics Research, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA
³Swiss Institute of Bioinformatics Biozentrum, University of Basel, Klingelbergstrasse 50-70, 4056 Basel, Switzerland
⁴School of Molecular Bioscience, The University of Sydney, NSW 2006, Australia
⁵Protein Data Bank in Europe, European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SD, UK
⁶Skaggs School of Pharmacy and Pharmaceutical Sciences and San Diego Supercomputer Center, University of California, San Diego, La Jolla, CA 92093, USA
⁷BioMagResBank, Department of Biochemistry, University of Wisconsin-Madison, Madison, WI 53706-1544, USA
⁸Protein Data Bank Japan, Institute for Protein Research, Osaka University, 3-2 Yamadaoka, Suita, Osaka 565-0871, Japan
⁹Physical Biosciences Division, Lawrence Berkeley Laboratory, Berkeley, CA 94720-8235, USA
¹⁰Department of Bioengineering, UC Berkeley, Berkeley, CA 94720, USA
¹¹Bijvoet Center for Biomolecular Research, Faculty of Science - Chemistry, Utrecht University, Padualaan 8, Utrecht, 3584 CA, the Netherlands
¹²National Center for Macromolecular Imaging, Baylor College of Medicine, Houston, TX 77030, USA
¹³Institute of Bioengineering, School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL) and Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland
¹⁴Department of Biochemistry, University of Washington, Seattle, WA 98195-7370, USA
¹⁵Department of Pharmaceutical Chemistry and Department of Bioengineering and Therapeutic Sciences, California Institute for Quantitative Biosciences, University of California, San Francisco, 800 16th Street, San Francisco, CA 94158-2517, USA
¹⁶Division of Structural Biology, Wellcome Trust Centre of Human Genetics, University of Oxford, OX3 7BN Oxford, UK

(Affiliations continued on next page)



Post-Meeting Activities

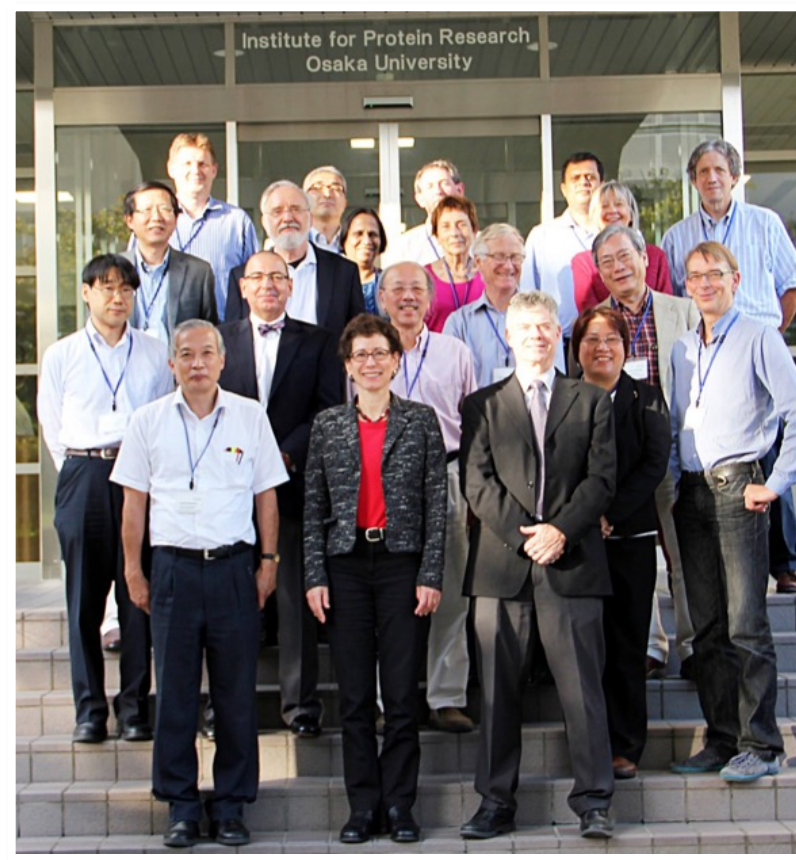
- Task Force Co-Chairs/Working Group Leadership
 - Jill Trewhella, Andrej Sali, Torsten Schwede, Helen Berman
 - Monthly Skype
- Subteams formed
 - Federation (JT and HB)
 - Validation (AS and TS)
- Proposals
 - Deposition and representation (EAGER funded by NSF)
 - Validation (submitted to NSF)
 - Research Collaborative Network (under preparation)



Federation Plan Endorsed by wwPDB AC

Principles

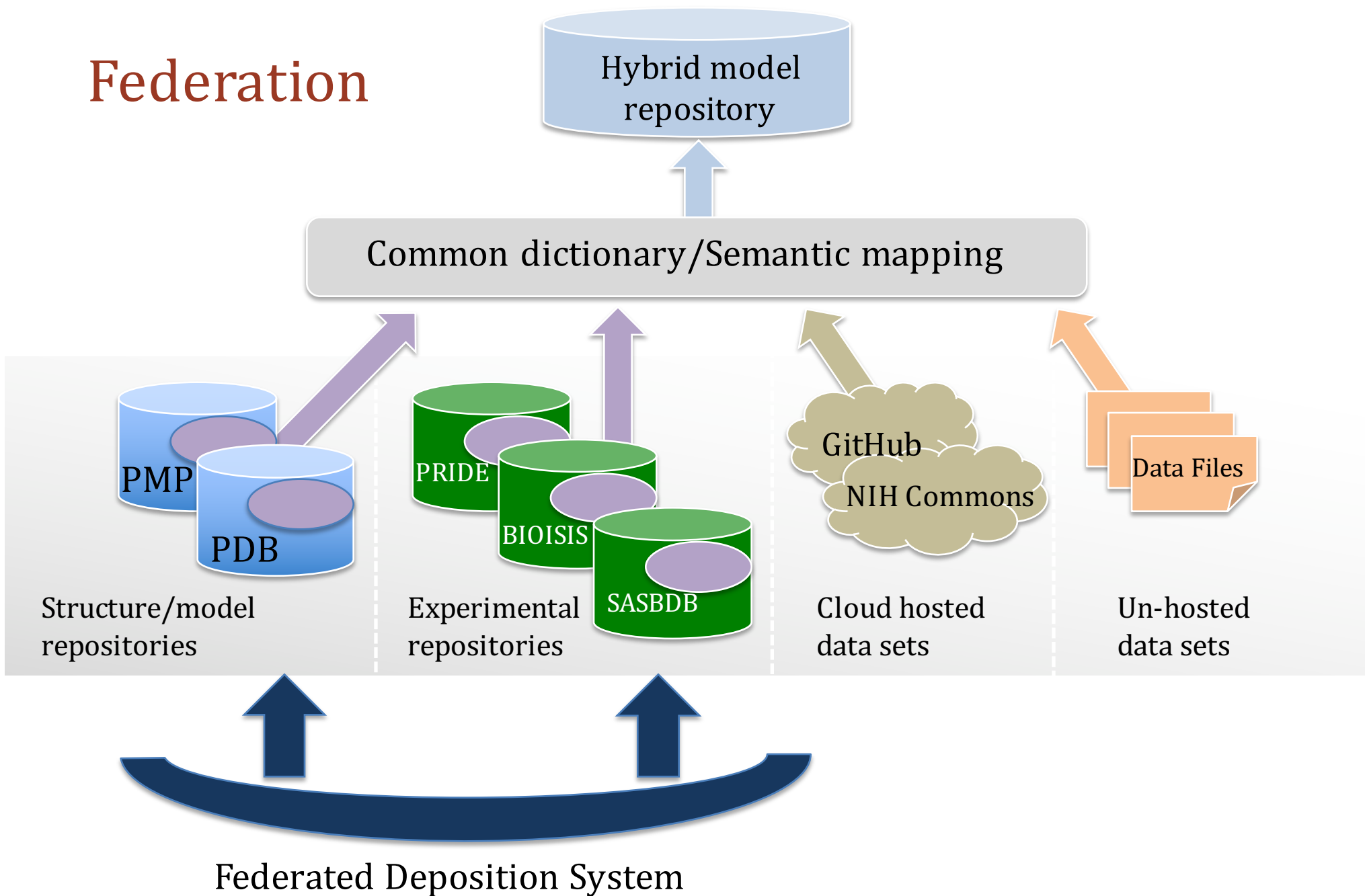
- Sample was made/used and some experimental data was collected
- Model → PDB
- Data → Federated Databases
- Development of validation standards required



October 2, 2015, Osaka, Japan



Federation



Small Angle Scattering Biological Data Bank (SASBDB)

- 49 SAS-only entries deposited to PDB between 2010-2012
- SAS Task Force recommended data transfer to SAS data and model archive
 - SASBDB hosted at EMBL-Hamburg
 - Using the sasCIF extensions to PDBx/mmCIF dictionary
- New PDB status code TRSF will indicate an entry has been transferred to another other archive center
 - Status: TRSF Title: XXXXXXXX, transferred to SASBDB
- Data transferred and should be available for depositor review by year end

SASBDB
Small Angle Scattering Biological Data Bank

Advanced search E.g. SASDAU4, Lyz, Nucleic Acids Res

Home Browse Submit data About SASBDB Help

Curated repository for small angle scattering data and models

Small angle scattering (SAS) of X-ray and neutrons provides structural information on biological macromolecules in solution at a resolution of 1-2 nm. SASBDB is a fully searchable curated repository of freely accessible and downloadable experimental data, which are deposited together with the relevant experimental conditions, sample details, derived models and their fits to the data.

SASBDB currently contains:
138 experimental data sets
234 models
97 experimental data sets on hold
116 models on hold

Browse the contents according to:

- Organism: Human, E. coli, Mouse, Other
- Macromolecule type: Protein
- Model type: Hybrid, Ab initio only
- Dissemination type: Published

Recent depositions:

SASDAC9 – Varkud Satellite (VS) ribozyme in Tris

Sample: Varkud Satellite (VS) ribozyme dimer, Neurospora RNA
Buffer: Tris, pH: 7.5
Experiment: SAXS data collected at BL12.3.1 SIBYLS, ALS 2012-Mar-06

Crystal structure of the Varkud satellite ribozyme.
Nature Chemical Biology doi:10.1038/nchembio.1929.
Suslov NB, DasGupta S, Huang H, Fuller JR, Lilley DMJ, Rice PA, & Piccirilli JA.

R_g Guinier 3.9 nm
 D_{max} 16.4 nm
Volume_{Provd} 240 nm³

SASDAB6 – Protein Interacti... SASDA58 – UL26N of pseud... SASDA57 – mouse olfactom... SASDAG7 – CD44 HABD scF... SASDAB7 – Complex ComE...

Structure of Dimeric and Tetra... Dimerization-Induced Allost... Olfactomedin-1 has a V-shape... Molecular mechanism for the... Modeling the ComD/ComE/Com...

To cite SASBDB refer to: Valentini E, Kikhney AG, Previtali G, Jeffries CM, Svergun DI, SASBDB, a repository for biological small-angle scattering data. Nucleic Acids Res. 2015 Jan 28;43:D357-63.
© BioSAXS group 2014.

www.sasbdb.org

Requirements for an Effective Federation

- Methods for data harvesting and deposition
 - Interoperating deposition sites
- Methods for curation and validation
 - Appropriate domain expertise
- Methods for archiving
 - What goes where?
- Methods for data exchange
 - Dependent on commitment/adherence to agreed standards
- Methods for data distribution

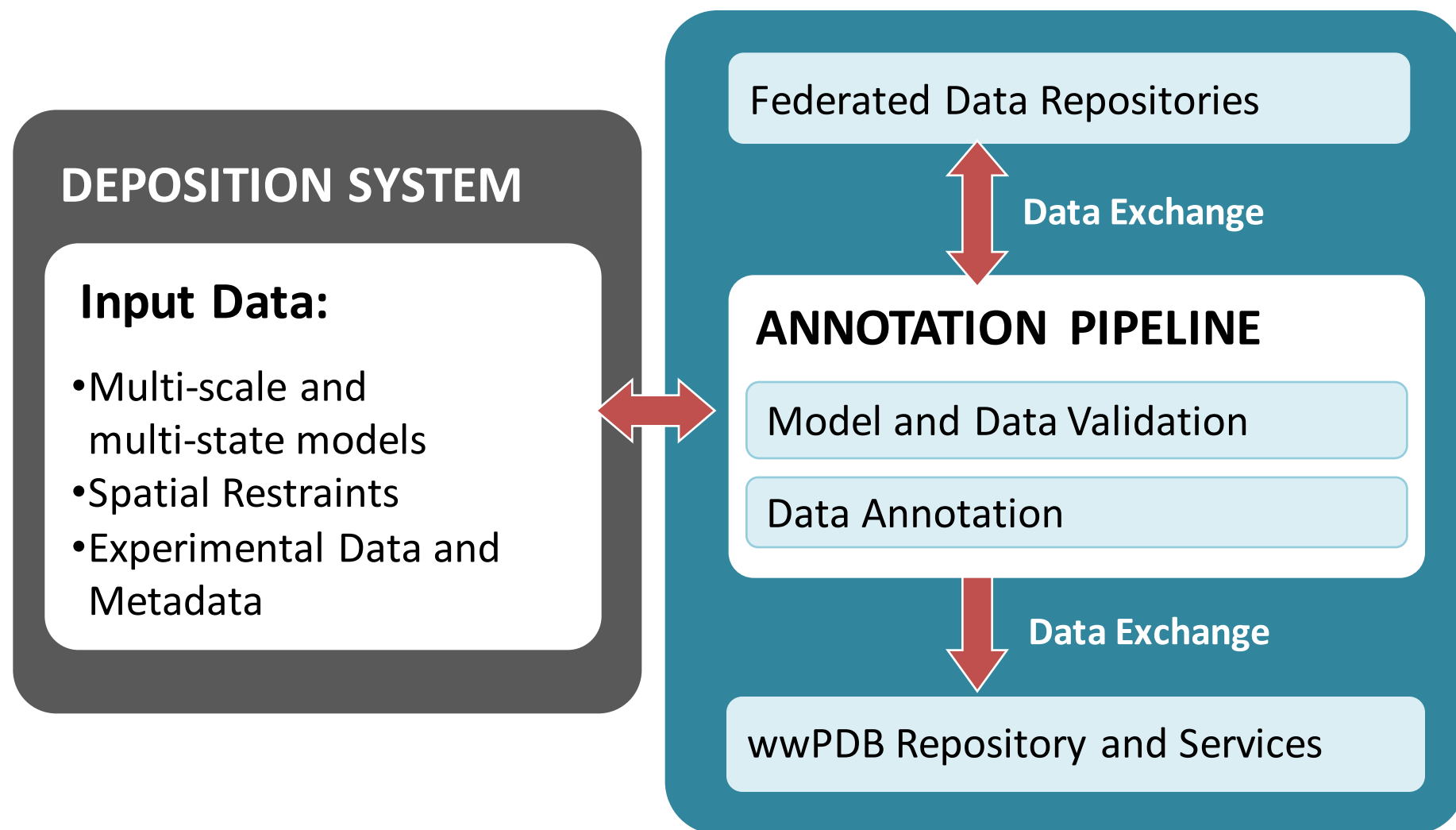


Creating a Research Collaboration Network (RCN)

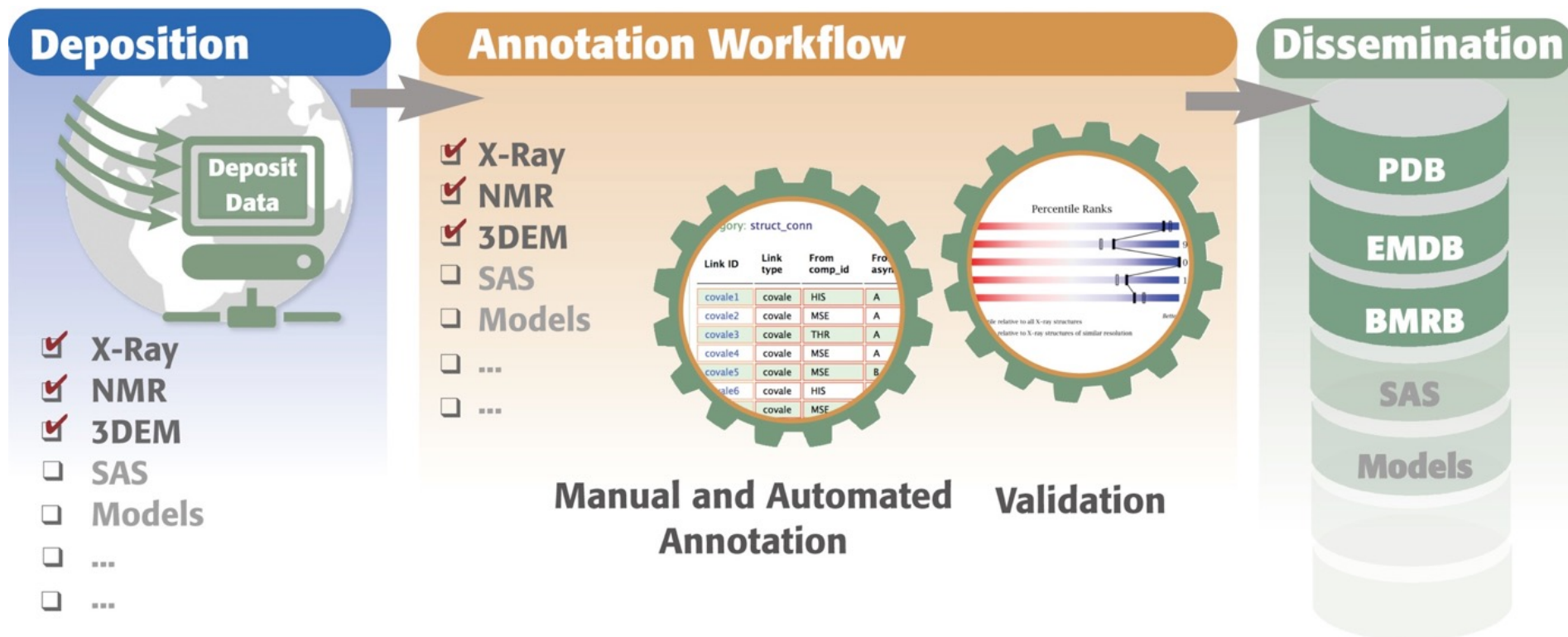
- Bring together representatives of key methods: X-ray, NMR, EM, SAS, MS/crosslinking, FRET, ...
- Chaired by Jill Trewhella
- Identify technical and community challenges
- First meeting November 30, 2015
- Proposal under preparation



Conceptual Model for a Data Pipeline



Build on the wwPDB D&A System



Validation of Integrative Models

- Preliminary standards and methods for estimating and visualizing model uncertainty will be developed
- Benchmark preliminary protocols for assessing model uncertainty
 - Clustering of models
 - Convergence of sampling
 - Fit to input information
 - Estimation of model precision based on variability in the ensemble of good-scoring models
 - Self-consistency of the experimental data
 - Validation models by using random subsets of experimental data
- Development of tools for visualizing model uncertainty
 - Plots to show fit to input information
 - Plots to describe localization density for model ensemble
 - 3D representations of model uncertainty



I/HM Working Group Leadership

- Jill Trehwella, Andrej Sali, Torsten Schwede, Helen Berman



Data In: Deposition and Annotation

John Westbrook, Ph.D.

Jasmine Young, Ph.D.



Biocurators and Data In Developers



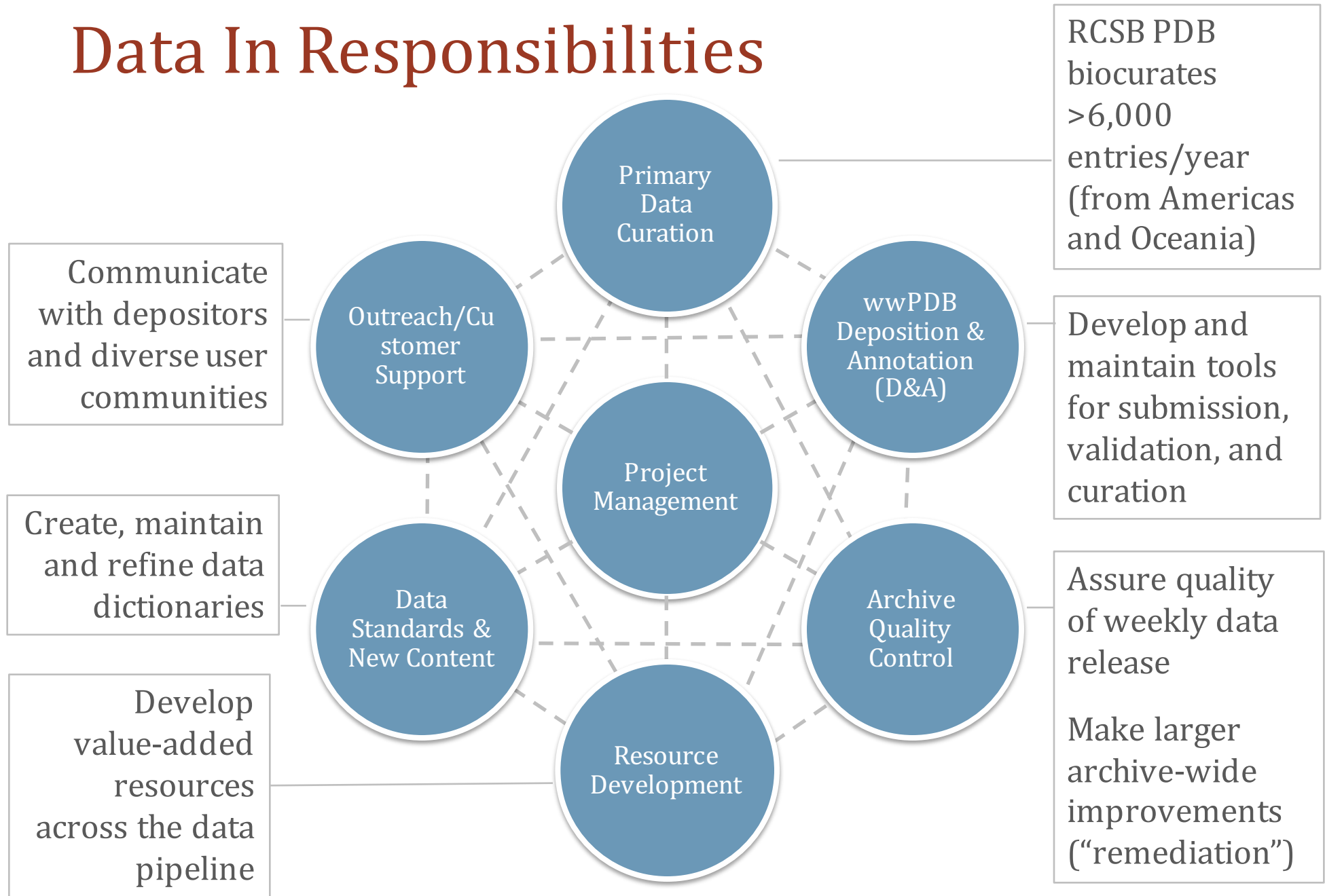
RCSB PDB East, October 2015

Demographics

- 11 scientists, 3 scientific developers, 4 software developers
- 14 Ph.D., 2 M.S., 2 B.S.
- 8 countries, 3 continents



Data In Responsibilities



2015 Goals and Deliverables



■ Primary Data Curation

- Increase productivity and efficiency

■ wwPDB D&A

- Enable wwPDB workload balancing; Support all current methods

■ Archive Quality Control

- Improve representation of large structures; streamlined weekly release

■ Resource Development

- External Reference Files (ERFs)

■ Data Standards and New Content

- NMR Exchange Format (NEF); Support multiple reflection data; Extend EM metadata collection

■ Outreach/ Customer Support

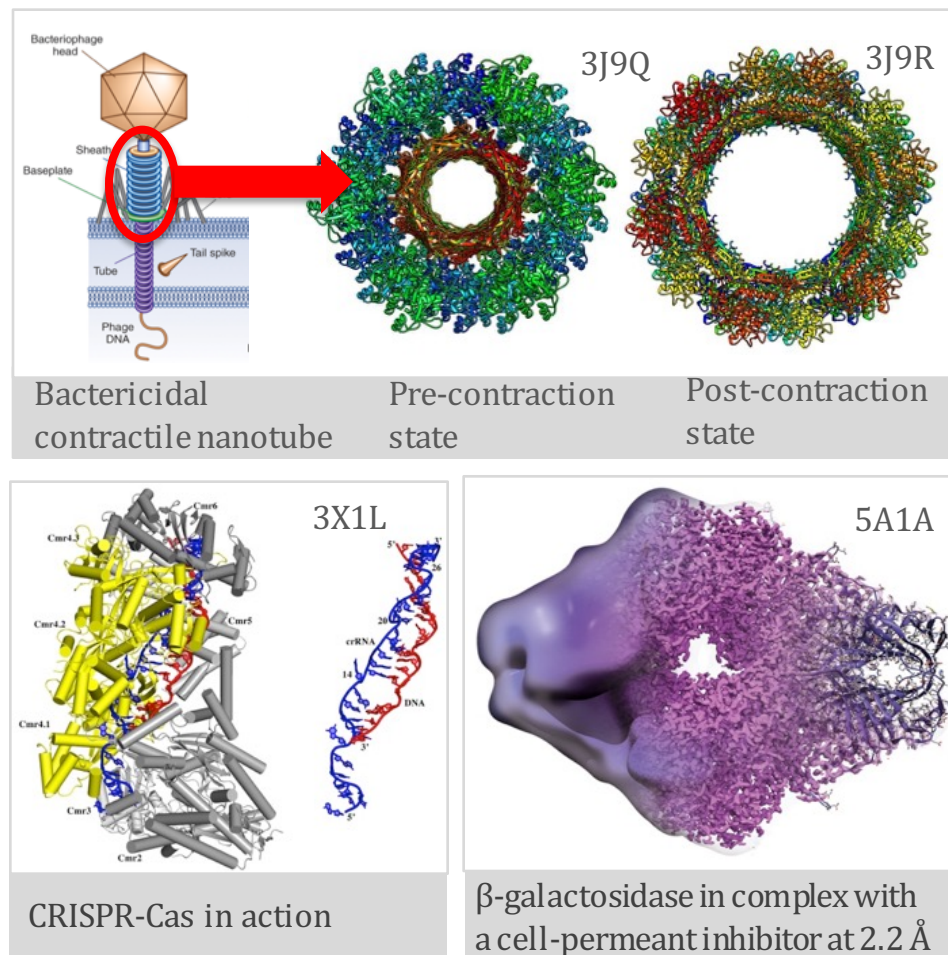
- D&A User feedback: What are they saying?



Primary Data Curation

- 2015: 8.4 FTE
- Domain expertise:
 - Methodologies (X-ray, NMR, EM, *etc.*)
 - Large structures (viruses/ribosomes)
 - Small molecules (drugs, metal-containing ligands, carbohydrates, *etc.*)
 - Software development
 - Statistics
- RCSB PDB processes >6,000 entries/year (from Americas and Oceania)
- Average ~50 entries per month per FTE

2015 Interesting Structures

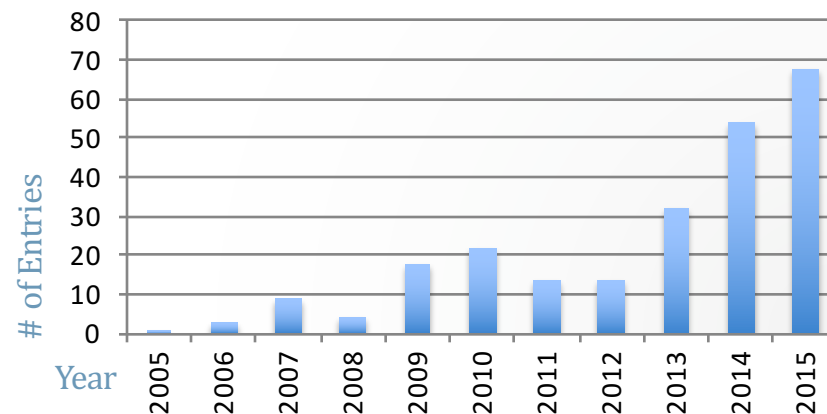


Increasing Size and Complexity

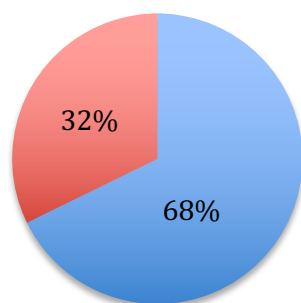
Large structures (containing >62 chains and/or 99999 atoms) cannot be represented in legacy PDB file format

- 2014: 277 large structures combined and re-released
- 2015: 68 new large structures deposited and released

Number of Large Structures Deposited per Year

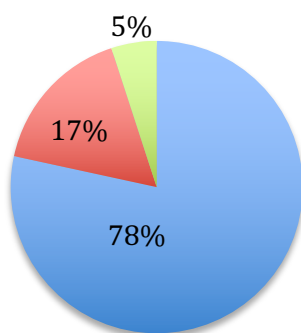


Experimental Method



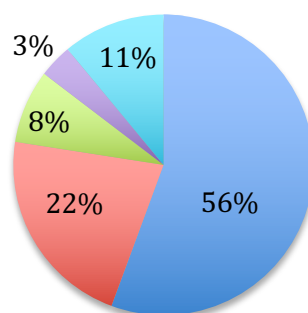
■ X-ray
■ EM

Classification



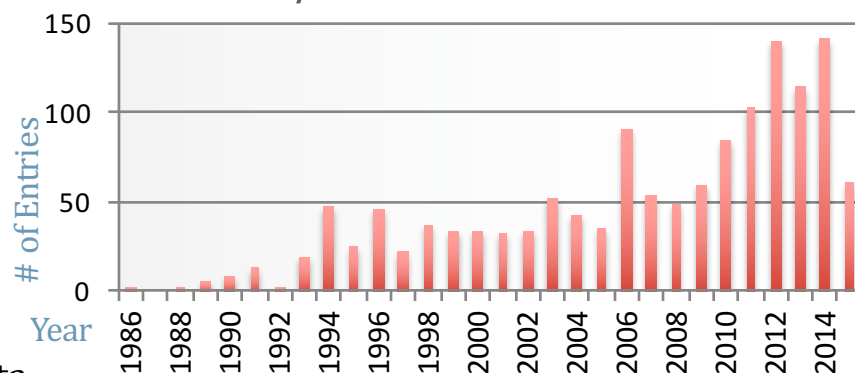
■ Ribosome
■ Virus
■ Other

Source Organism



■ Eubacteria ■ Eukaryota
■ Viruses ■ Archaea
■ Other

Number of Peptide-like Inhibitor/Antibiotic Entries

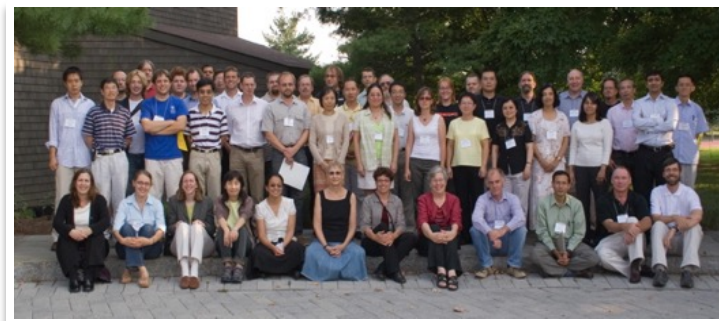


Improving Primary Data Curation with the wwPDB Deposition & Annotation System (D&A)

Goals: Create the next generation deposition and annotation system to support PDB data from X-ray, NMR, and 3DEM that will:

- Maximize data quality
 - Improve data validation
 - Standardize file formats
 - Ensure more complete data capture
- Support larger and more complex structures
- Improve efficiency and consistency
 - Use of same system across wwPDB to enable load balancing and expansion
 - Automation and validation of routine tasks

WORLDWIDE
wwPDB
PROTEIN DATA BANK



2007 Initial Discussions



2010 D&A Team Meeting

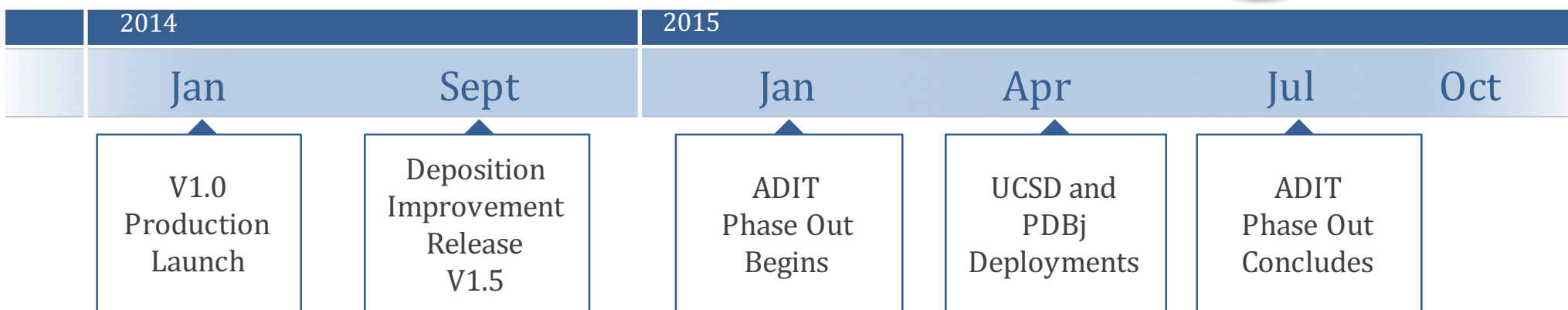
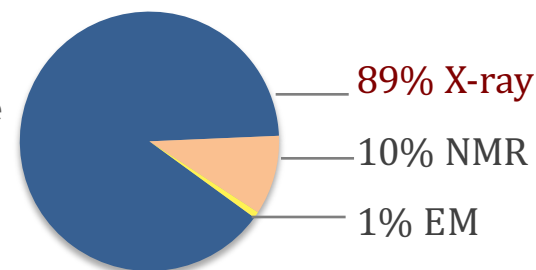


2014 wwPDB Biocurator Summit



X-ray D&A Milestones

PDB Archive



Annotation Pipeline Improvement

- January 2014: X-ray production launch (V1.0)
- Fall 2014: X-ray improvement release (V1.5)
- January 2015: RCSB PDB and PDBj ADIT systems phase out begins
- April 2015: UCSD and PDBj deployments
- July 2015: RCSB PDB and PDBj ADIT systems phase out concludes
- 2015: Annotation pipeline improvement

ADIT: Legacy deposition system developed by RCSB PDB

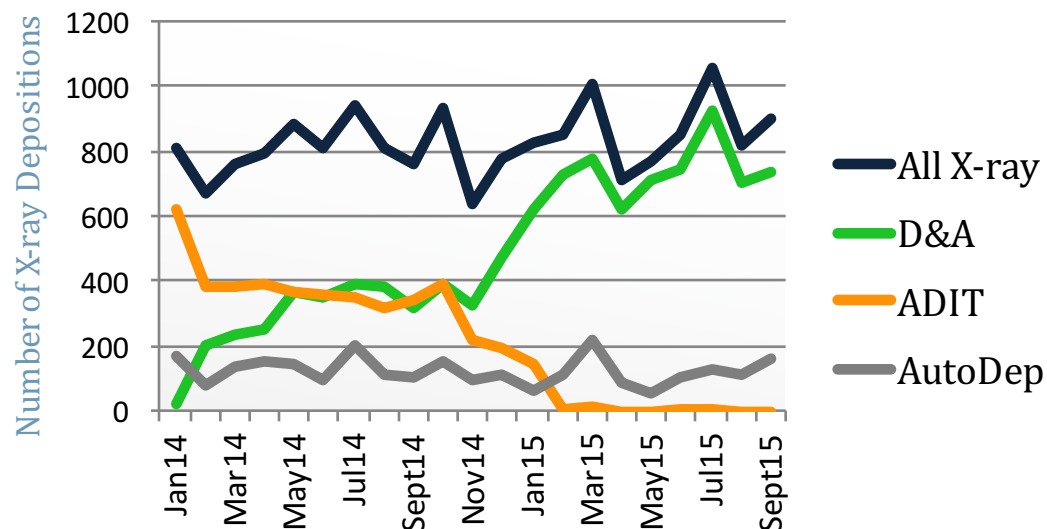


Impact of D&A on X-ray Deposition

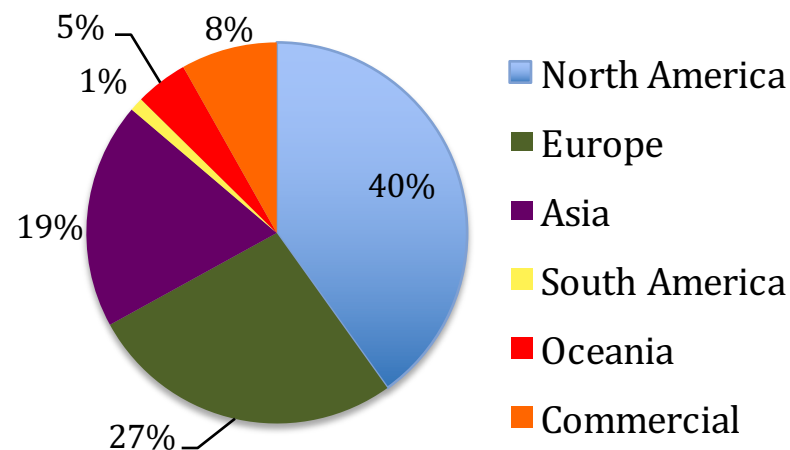
January 27, 2014 – September 30, 2015

- More than 10,200 structures deposited and fully annotated
- More than 5,600 released

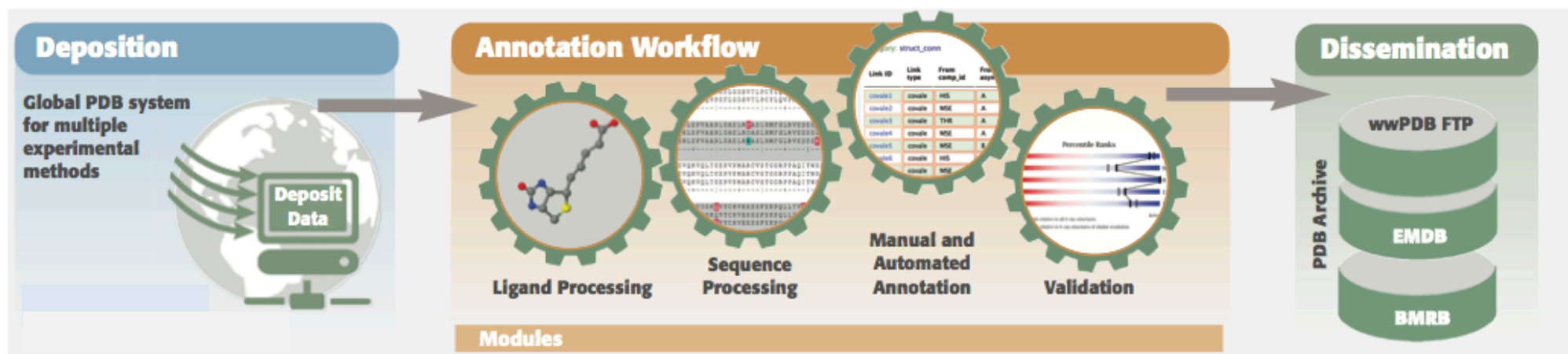
X-ray Depositions by Month



D&A Depositions by Geography



Biocuration Pipeline Improvements



- Enhanced sequence annotation for chimeric proteins
- Enhanced ligand assessment with display of local ligand density fit
- Improved Workflow management
- Increased productivity *via* Workflow Manager multi-processing
- Developed monitoring system



Improved Sequence Annotation

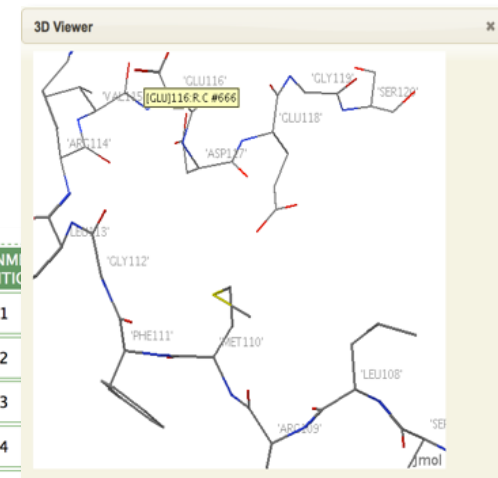
- Biological sequence checked against atomic coordinate sequence and cross-referenced to UniProt/GenBank
- 3D structure view
- Sequence discrepancy annotation
- Added support for chimeric proteins

SP:P62161
SP:Q5S6Z9
SP:Q5S6Z9
SP:Q6LDG3

Four
sequence
references
for four
fragments

AUTH Entity: 1 V:4 XYZ Chain: A V:1 SP:P62161 SP:Q5S6Z9 SP:Q5S6Z9 SP:Q6LDG3	MHHHHHGS	DQLTEEQIAEFKEAFSLFDKDDGGTITTKELGTVMRSLGQNPTEAELQDMI
AUTH Entity: 1 V:4 XYZ Chain: A V:1 SP:P62161 SP:Q5S6Z9 SP:Q5S6Z9 SP:Q6LDG3	NEVDADG	DGTIDFPEFLTMMARKMKDTSDEEEIREAFRVFDKDGNGYISAAELRHVMTNL
AUTH Entity: 1 V:4 XYZ Chain: A V:1 SP:P62161 SP:Q5S6Z9 SP:Q5S6Z9 SP:Q6LDG3	GEKLTDEEVDEMIREADIDGGQVNYEEFVQMMT	AKLECEKMYVRDGLVTGDIHMALLLE
AUTH Entity: 1 V:4 XYZ Chain: A V:1 SP:P62161 SP:Q5S6Z9 SP:Q5S6Z9 SP:Q6LDG3	GNAHYRCDFR	TTYKAKEKGVKLPGYHFVDHCIEILSHDKDYNKVKLYEHAVAHSGLPDNA
AUTH Entity: 1 V:4 XYZ Chain: A V:1 SP:P62161 SP:Q5S6Z9 SP:Q5S6Z9 SP:Q6LDG3	RRGGTGGSMV	SAIKPDMKILRMEGNVNGHHFVIDGGDGTGKPYEGKQTMDELVKEGGPLP

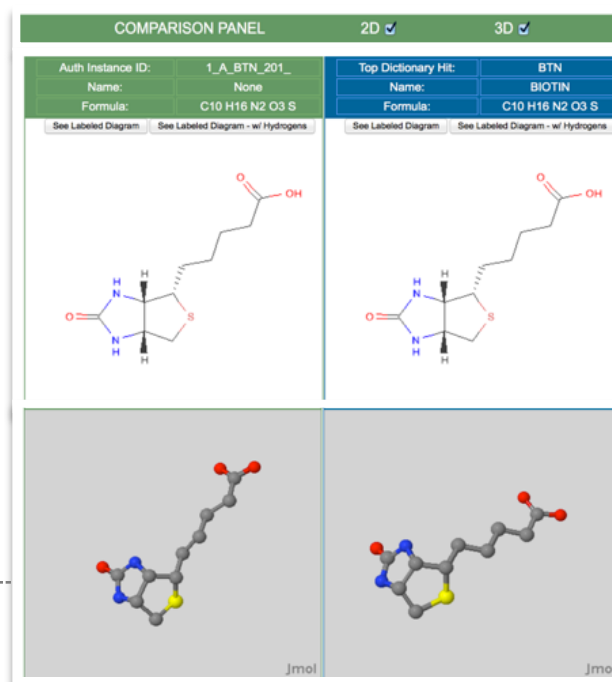
Residue: ALA
Position: 155



ALIGNM POSITION				
1				
2				
3				
4				
5	HIS	SP:P62161	.	expression tag
6	HIS	SP:P62161	.	expression tag
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:Q5S6Z9	.	linker
156	LYS	SP:Q5S6Z9	.	linker
157	LEU	SP:Q5S6Z9	.	linker
158	GLU	SP:Q5S6Z9	.	linker
159	CYS	SP:Q5S6Z9	.	linker
174	HIS	SP:Q5S6Z9	THR	engineered mutation
241	ARG	SP:Q5S6Z9	.	linker
242	ARG	SP:Q5S6Z9	.	linker
243	GLY	SP:Q5S6Z9	.	linker
244	GLY	SP:Q5S6Z9	.	linker
245	THR	SP:Q5S6Z9	.	linker

Improved Ligand Annotation

- Batch search against Chemical Component Dictionary with automated ligand ID assignment
- Captures and displays author-provided chemical information
- Comparison panel
 - 2D and 3D views of ligand for review
 - ID assignment
- Enhanced with display of local ligand electron density fit

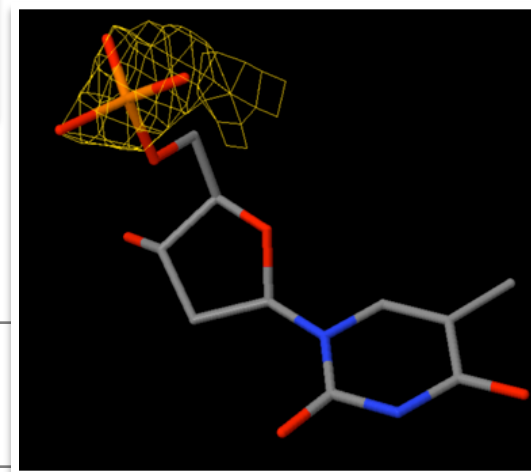
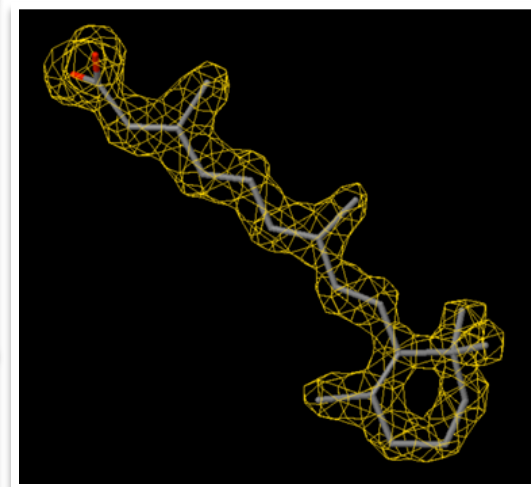


Deposited instance
from coordinates

Closest match in
the dictionary

Local ligand density
display (1.5 sigma
omit map)

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



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

Improved Workflow Management

- Improved messaging with communication tracking
- Enabled task prioritization (tabs, color labels, message flags, *etc.*)
- Streamlined deposition help desk (pre-submission)

[1] DEPOSITION SUMMARY

Deposition Summary : Level 1 : Refresh

RCSB - ANNOTATOR | CHENGHUA SHAO

Review module Release module

exception finished init open running waiting restartWF

pdb-d-linux-1.rutgers.edu 634218.3

pdb-d-linux-2.rutgers.edu 0.0

To-Do Entries (3) Proc Entries (8) Author's Corrections (0) Pending entries (3) Reminder Entries (0) Search Options My Messages (0) All Messages (16) Unsubmitted Entries w/ Messages (13)

To-Do Entries Refresh

Process	Exp Method	PDB/BMRB/EMDB	Status	Auth Coord	Dep Date	Ann	Remove	Comm	Title	Author
SeqModUI	X-RAY DIFFRACTION	4P66/??	REL	HOLD	2014-09-09	CS	Remove		test	test
TransMod	SOLUTION NMR	4P7U/80020/?	PROC	HOLD	2014-12-09	yhl	Remove		nmr testing 2lvn	liang, y.-h.,
LigMod	X-RAY DIFFRACTION	4XR6//	PROC	HPUB	2015-01-20	CS	Remove		Tailspike protein double mutant D339A/E372Q of E. coli bacteriophage HK620 in complex with hexasaccharide	Gohlke, U., Broeker, N.K., Heinemann, U., Seckler, R., Barbirz, S.,

D&A Improved Efficiency and Advances

Efficiency

- Greater efficiency for routine structures
- More automated processing for ligands and polymer sequences
- Processing of multiple entries in parallel
- Processing of large and complex structures without splitting

Advances

- Improved support for large structures
- Workflow ensures completeness of annotation
- Enhanced validation
- Enabled monitoring of system performance



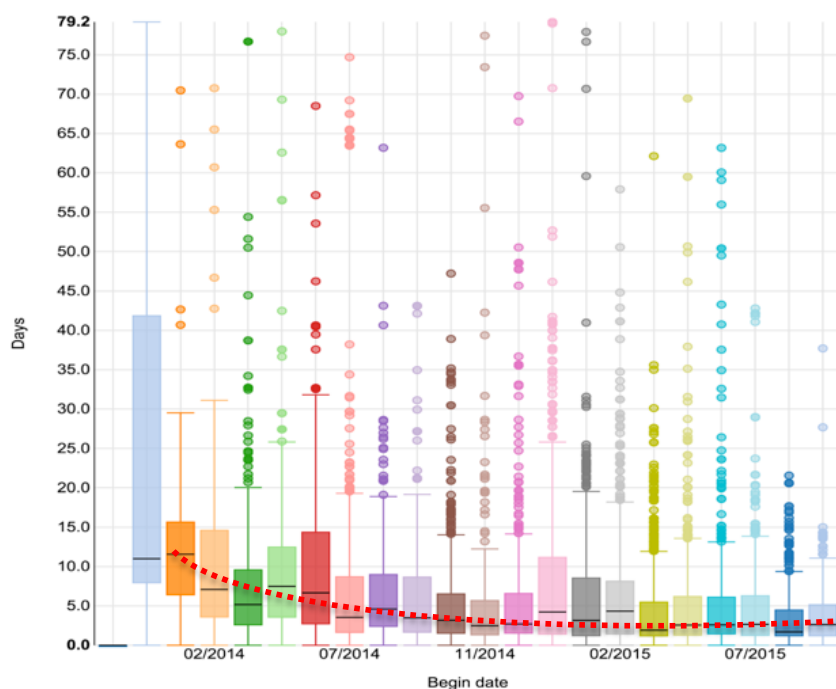
Measuring System Performance

Benefit of new system: Improved throughput with system enhancements

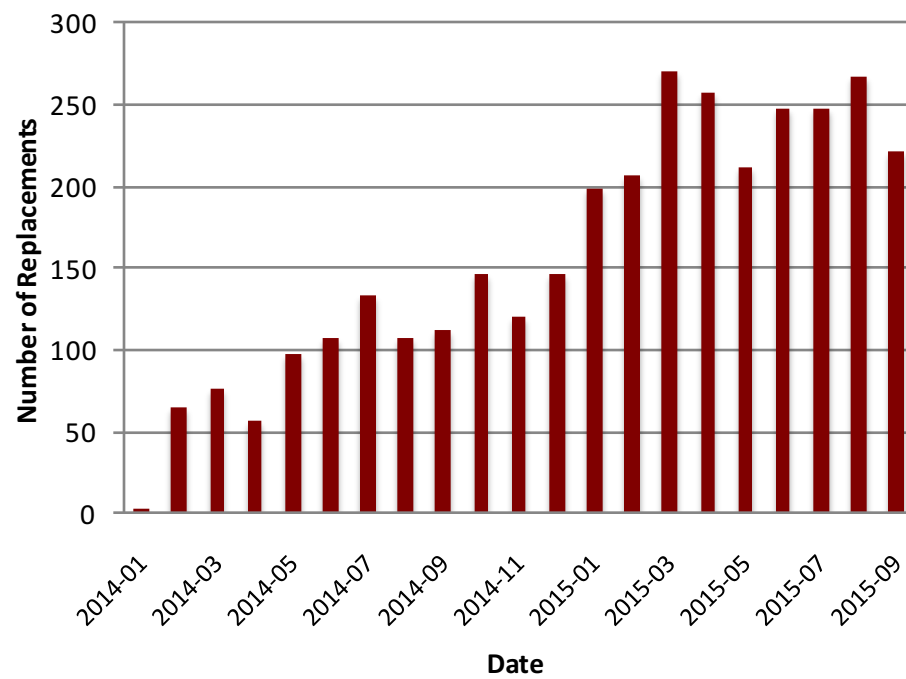
- **Median: 16.5 days → 1.6 days**
- Majority completed within 1.6 days
- ADIT median ~14 days

Unanticipated consequence:
Increased coordinate replacements as
Depositors react to validation reports

Average Processing Time by Month

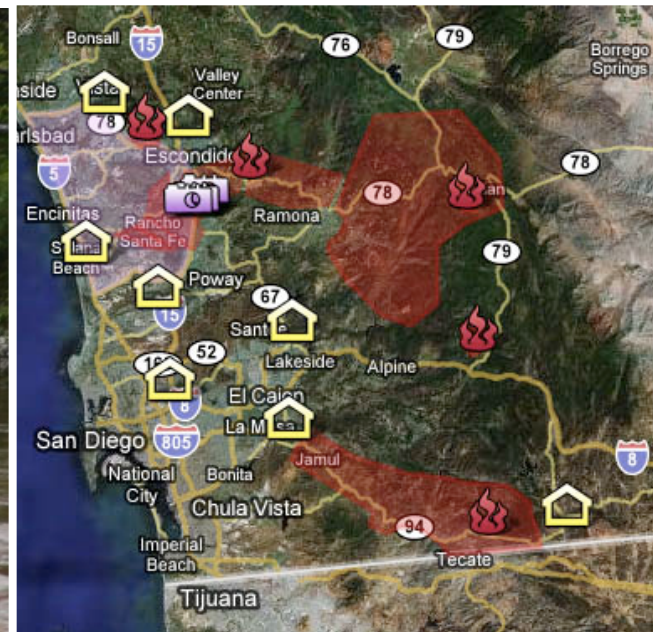


Number of Coordinate Replacements by Month



D&A Deployments and Supporting Infrastructure

- Deployments at Rutgers
- Disaster Recovery and Global Deployment



wwPDB D&A Deployments at Rutgers January 2014-September 2015

- Host multiple D&A platforms
 - Production, alpha/beta test, production staging, development
- Host standalone validation services
- Host annotation services for all partners
- Host archive update operations for legacy and D&A pipelines



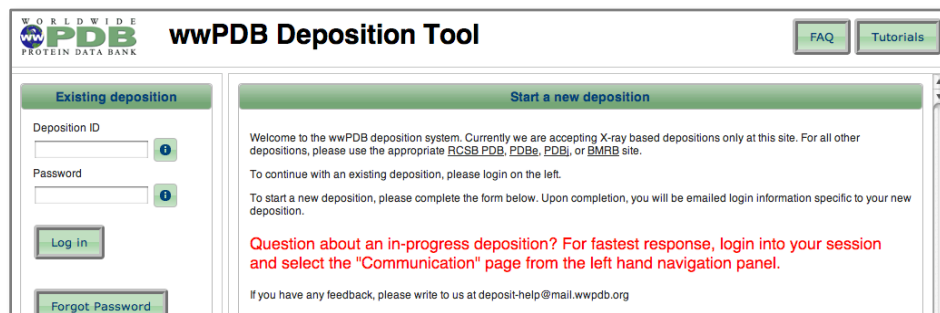
D&A Disaster Recovery and Global Deployment Since April 2015

UCSD

- Full deposition and annotation pipeline
- Hardware/software commissioned
- Multiple entries fully processed and released
- Standby for warm failover

PDBj

- Full deposition and annotation pipeline
- Hardware/software commissioned
- > 170 entries fully processed
- > 60 entries released



deposit-rcsb-west.wwpdb.org/deposition

deposit-pdbj.wwpdb.org/deposition/

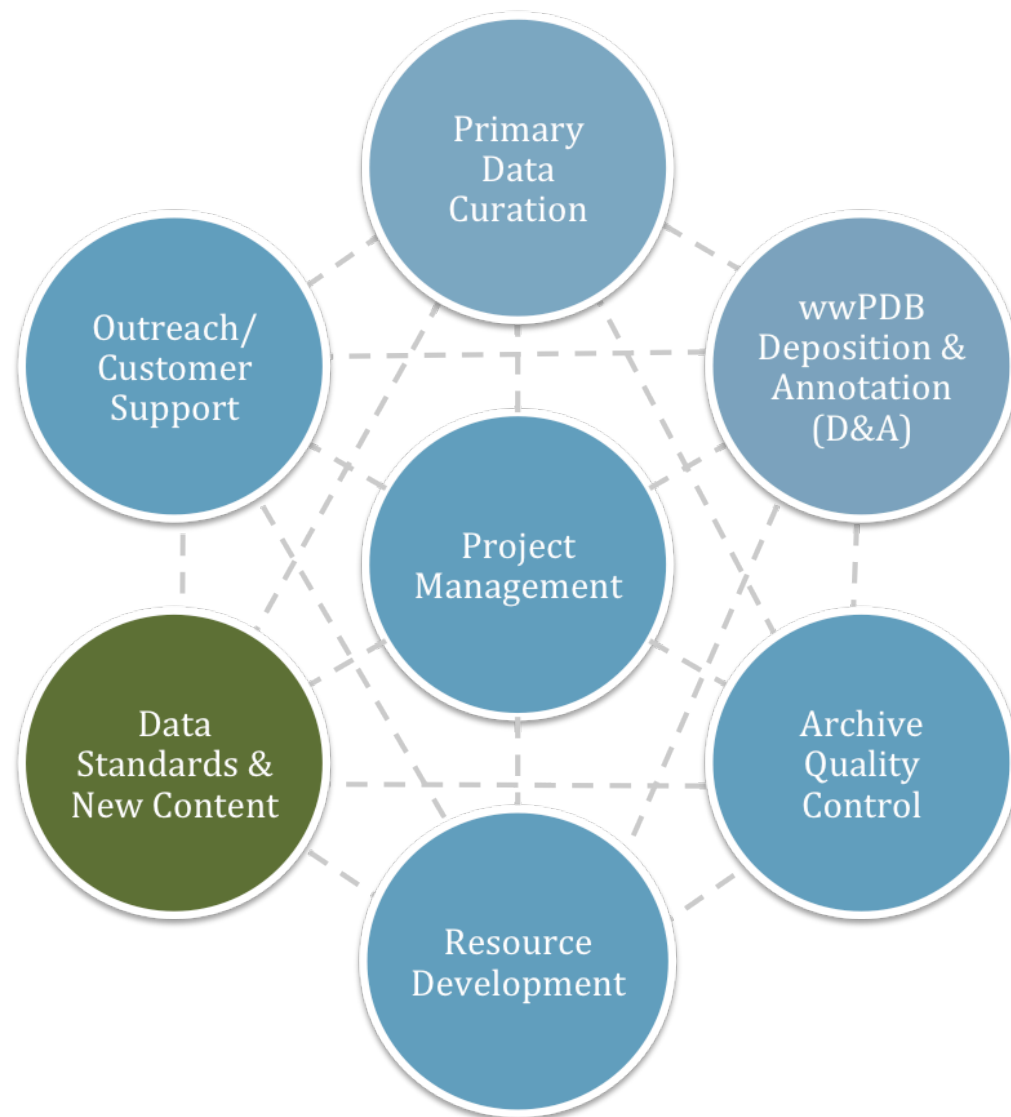


*2015 PDBj
D&A System
Training*



Data Standards and New Content

- PDBx/mmCIF Format
- NEF Format
- ERFs: New Content

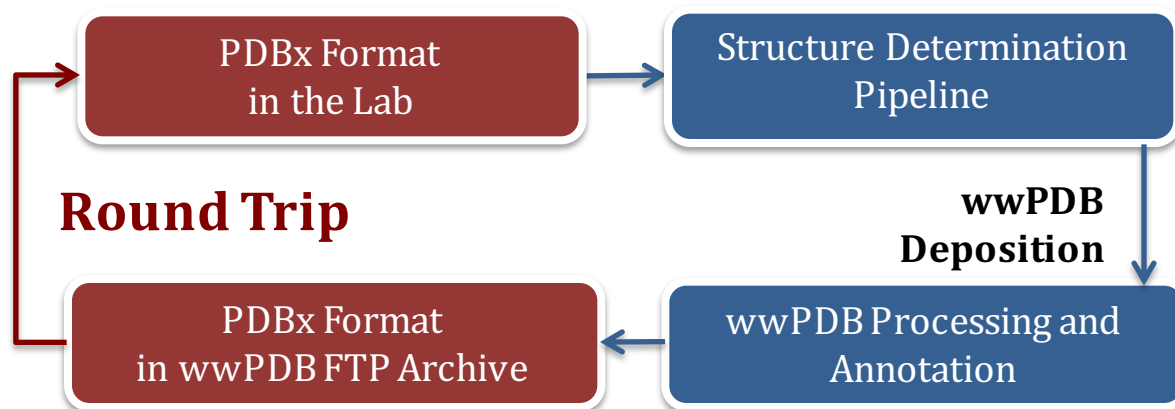


Working Group for PDBx/mmCIF Deposition

- Working Group meets virtually ~monthly
- Currently finalizing recommendations for delivery of complete chemical definitions, chemical restraints, richer X-ray experimental data, and extensions for SFX and XFEL experiments



Workshop Participants, Sept 2011



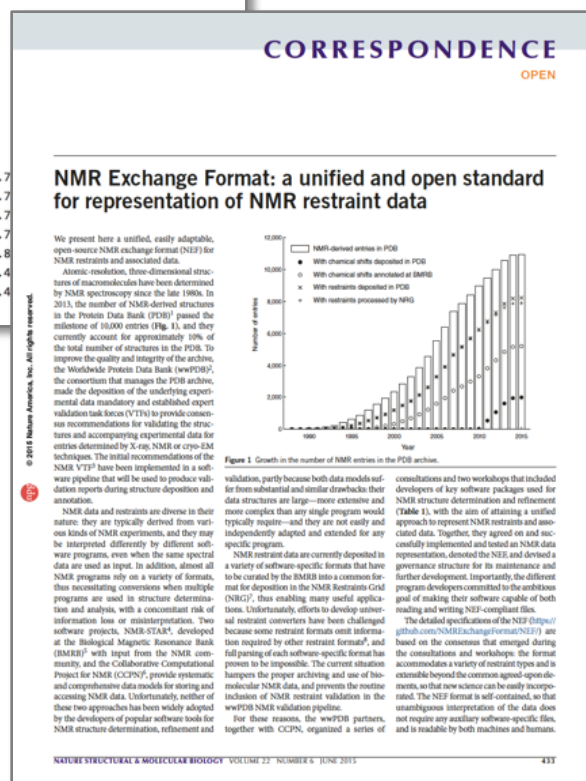
Workshop Participants, Oct 2014

NMR Exchange Format (NEF)

- Format specification and example files developed by working group spearheaded by GW Vuister, U Leicester
- All represented software could write NEF files and read each other's output
- Decisions on outstanding issues (e.g., RDC representation, treatment of stereo-specific assignments)
- Decisions on management and future development
- Roll out NEF-enabled software versions by the end of 2015 (XPLOR-NIH is ready)

```
loop_  
# Mandatory parameters, except for restraint_combination_id  
_nef_distance_restraint.ordinal  
_nef_distance_restraint.restraint_id  
_nef_distance_restraint.restraint_combination_id  
_nef_distance_restraint.chain_code_1  
_nef_distance_restraint.sequence_code_1  
_nef_distance_restraint.residue_type_1  
_nef_distance_restraint.atom_name_1  
_nef_distance_restraint.chain_code_2  
_nef_distance_restraint.sequence_code_2  
_nef_distance_restraint.residue_type_2  
_nef_distance_restraint.atom_name_2  
_nef_distance_restraint.weight  
  
# The following parameters are optional. target_value and target_value_uncertainty  
# should be given whenever a meaningful value is known'  
# Other parameters need be given only if they are defined for the potential_type  
_nef_distance_restraint.target_value  
_nef_distance_restraint.target_value_uncertainty  
_nef_distance_restraint.lower_linear_limit  
_nef_distance_restraint.lower_limit  
_nef_distance_restraint.upper_limit  
_nef_distance_restraint.upper_linear_limit  
  
# key: ordinal  
  
1 1 . A 21 ALA HB% A 17 VAL H 1.00 3.7  
2 1 . A 21 ALA HB% A 18 LEU H 1.00 3.7  
3 1 . A 22 THR HG2% A 17 VAL H 1.00 3.7  
4 1 . A 22 THR HG2% A 18 LEU H 1.00 3.7  
5 5 . A 18 PHE HB2 A 24 ASP HBX 1.00 2.8  
6 8 . A 18 PHE HB2 A 24 ASP HBY 1.00 4.4  
6 8 . E 68 SER HB2 A 24 ASP HBY 1.00 4.4  
stop_
```

[github.com/
NMRExchangeFormat/NEF](https://github.com/NMRExchangeFormat/NEF)



New Chemical Component Model Data File

- Exact small molecules matches between PDB and Cambridge Structural Database (CSD)
- Contains
 - Accession code correspondences
 - Cartesian coordinates and R-value
 - Data-collection temperature and a disorder flag
 - SMILES and InChI descriptors
 - Digital Object Identifier (DOI) for the citation associated with CSD entry
- Distributed *via* wwPDB FTP archive
- Collaboration between wwPDB and CSD

WORLDWIDE
wwPDB
PROTEIN DATA BANK



www.ccdc.cam.ac.uk



Data Out: Data Access and Exploration

Peter Rose, Ph.D.

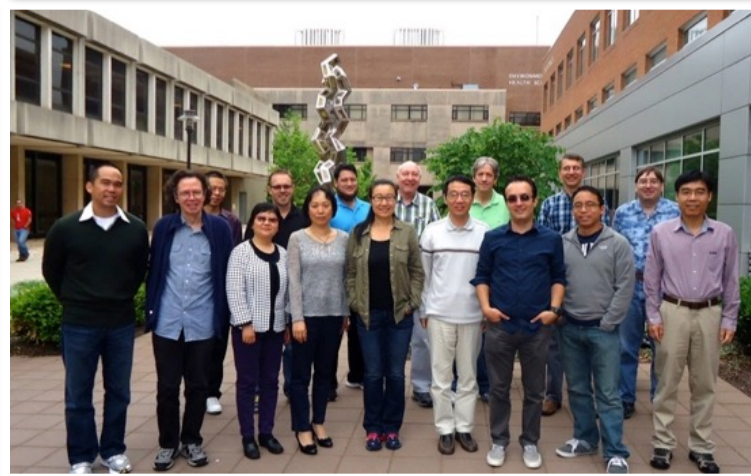
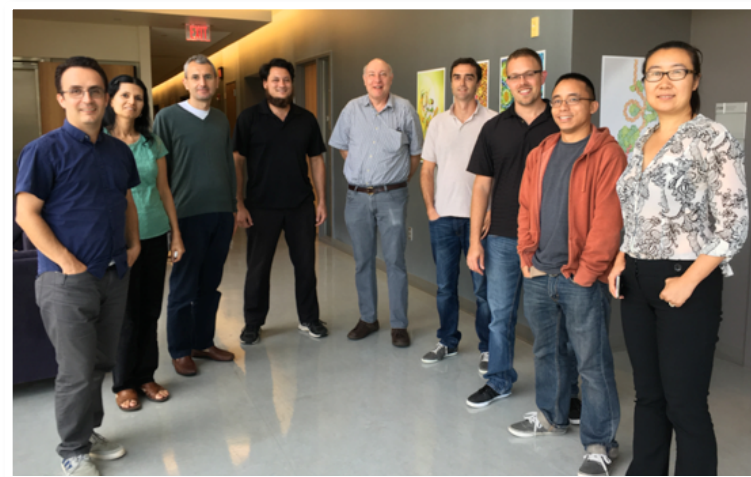
Andreas Prlić, Ph.D.



Bi-coastal Developer Team

- **Science:** structural and computational biology, structural bioinformatics, chemoinformatics, genomics, drug design, scientific software development
- **Software:** web design, user experience design, database design, web services, big data, open source development
- **Systems:** hardware, networking, virtualization, cloud computing, data management
- **Diversity**
 - 3 scientists, 4 software developers, 2 systems & infrastructure
 - 3 Ph.D., 5 M.S., 1 B.S.
 - 7 countries, 3 continents

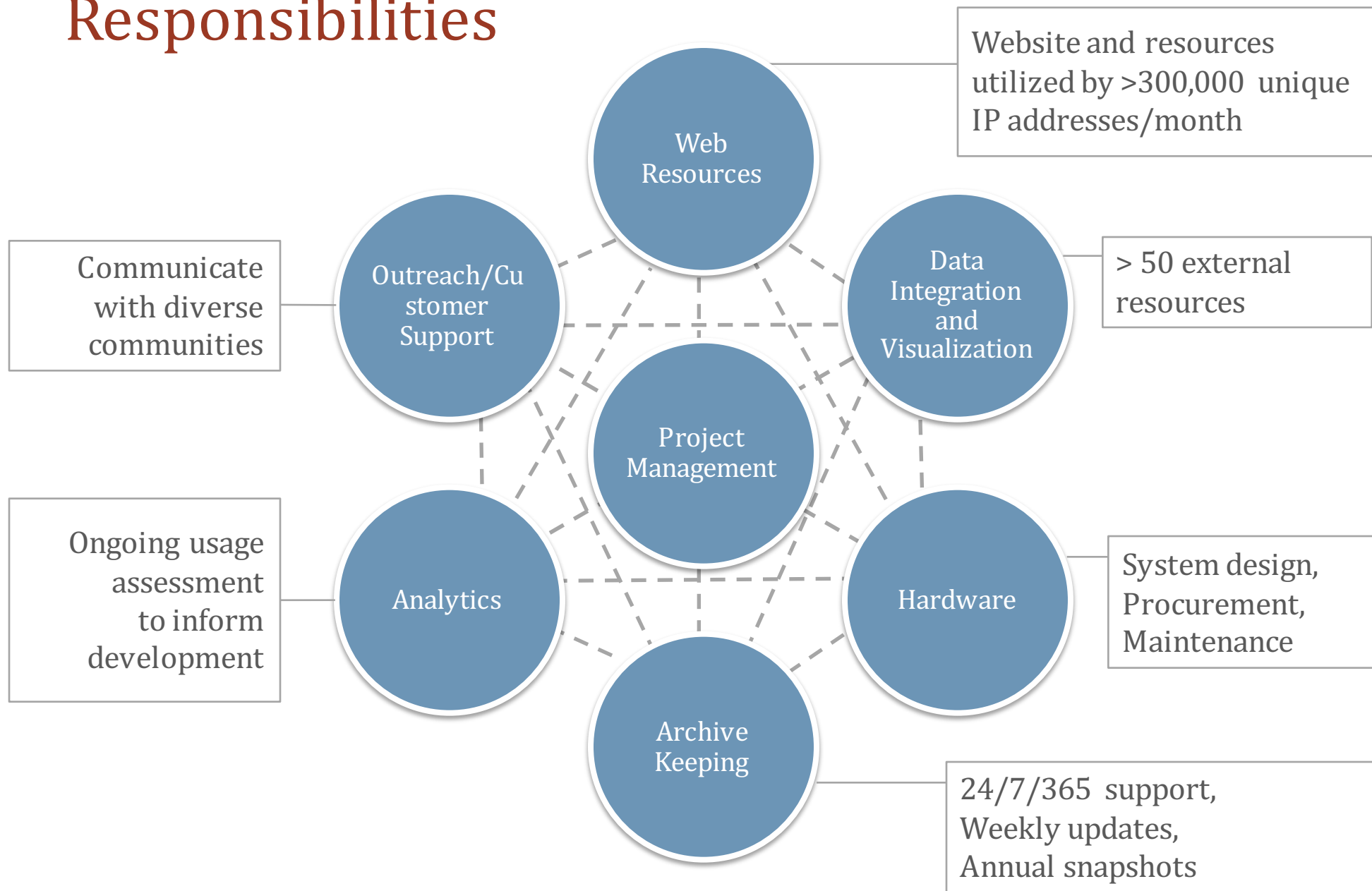
RCSB PDB West, October 2015



Developers Summit at Rutgers, June 2015



Responsibilities

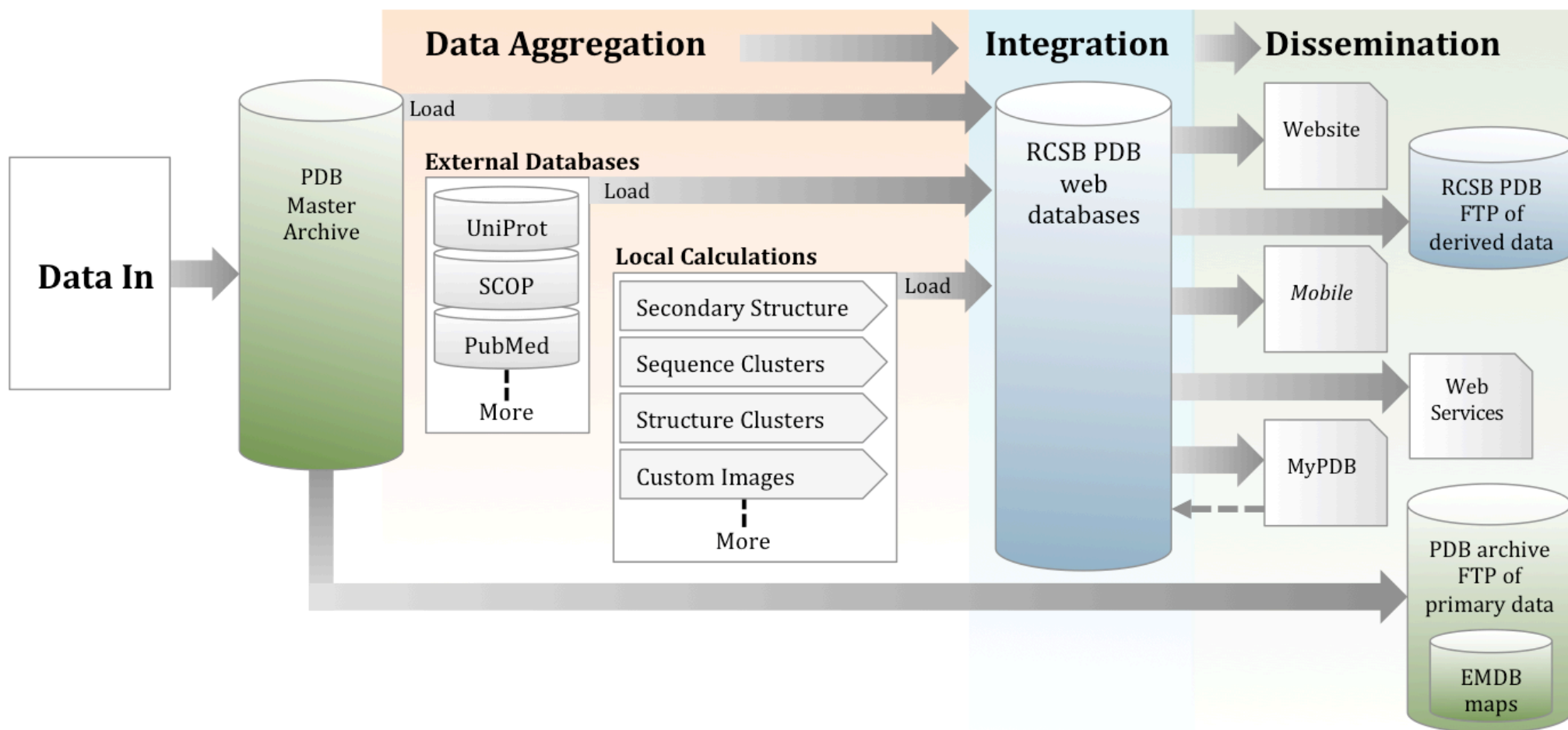


2015 Goals and Deliverables

- **Web Resources:** Enhance user experience
- **Data Integration:** Mutant and validation information, map gene locations
- **Hardware/Infrastructure:** Modularize services to enhance extendibility, scalability, and maintainability
- **Archive Keeping:** Maintain and update a secure and uniform version of the archive
- **Analytics:** Monitor usage trends
- **Outreach:** Professional society and technical meetings, including American Society of Human Genetics, GTC Bio, InChI, OpenStack



Data Flow



Structure Summary Redesign

- Builds on 2014 home page redesign
- Enhanced usability
- Improved content organization

The screenshot shows the RCSB PDB website home page. At the top, there is a navigation bar with links for Deposit, Search, Visualize, Analyze, Download, Learn, and More. Below this is the PDB logo and a search bar. The main content area is divided into several sections: 'Welcome', 'Deposit', 'Search', 'Visualize', 'Analyze', 'Download', and 'Learn'. There are also featured articles and news items, such as 'A Structural View of Biology' and 'October Molecule of the Month'.

The screenshot shows the RCSB PDB Structure Summary page for entry 2RH1. The page is divided into several sections: 'Structure Summary', '3D View', 'Annotations', 'Sequence', 'Sequence Similarity', 'Structure Similarity', 'Experiment', and 'Literature'. The main content area features a 3D ribbon diagram of the protein structure, labeled 'Biological Assembly 1'. To the right of the diagram, there is a detailed summary of the entry, including the title '2RH1 High resolution crystal structure of human B2-adrenergic G protein-coupled receptor.', the DOI, classification, authors, and experimental data. Below the summary, there is a 'wwPDB Validation' section with a bar chart showing various metrics and their percentile ranks. The 'Literature' section at the bottom provides a list of publications related to the entry.

New Responsive Structure Summary Page Layout

The tablet view displays a detailed summary for entry 2RH1. At the top, there are navigation tabs: Deposit, Search, Visualize, Analyze, Learn, and More, along with a MyPDB Login button. The entry title is "2RH1 High resolution crystal structure of human B2-adrenergic G protein-coupled receptor." Below this, the DOI is 10.2210/pdb2rh1/pdb. The classification is MEMBRANE PROTEIN / HYDROLASE. The deposition date is 2007-10-05 and the release date is 2007-10-30. The authors listed are Cherezov, V., Rosenbaum, D.M., Hanson, M.A., Rasmussen, S.G.F., Thian, F.S., Kobilka, T.S., Choi, H.J., Kuhn, P., Weis, W.J., Kobilka, B.K., Stevens, R.C., Accelerated Technologies Center for Gene to 3D Structure, GPCR Network. The organism is Homo sapiens | Enterobacteria phage T4 sensu lato, and the expression system is Spodoptera frugiperda. There are 3 mutations. The Structural Biology Knowledgebase (SBKB.org) shows 5 models and >22 annotations. An Experimental Data Snapshot section shows the method as X-RAY DIFFRACTION with a resolution of 2.4 Å, R-Value Free of 0.232, and R-Value Work of 0.196. A wwPDB Validation chart shows metrics like Rfree (0.231), Clashscore (4), Ramachandran outliers (0), Sidechain outliers (0.3%), and RSRZ outliers (3.5%). A literature section includes a citation to Science 318: 1258-1265 (2007) and a PubMed abstract snippet.

Tablet view

The phone view shows a condensed version of the summary page. It includes the entry title "2RH1 High resolution crystal structure of human B2-adrenergic G protein-coupled receptor.", the DOI, classification, deposition and release dates, authors, organism, and expression system. It also features the SBKB.org statistics and the Experimental Data Snapshot with the method and resolution (2.4 Å).

Phone view

The WebGL-based 3D viewer displays a ribbon representation of the human B2-adrenergic G protein-coupled receptor (2RH1). The structure is shown in a vertical orientation, with the extracellular domain at the top and the intracellular domain at the bottom. The protein is colored in a gradient from blue at the top to green at the bottom. A red stick model of a ligand is visible in the binding pocket. At the bottom of the viewer, there are control buttons for Snapshot, Detach, Full screen, and Reload.

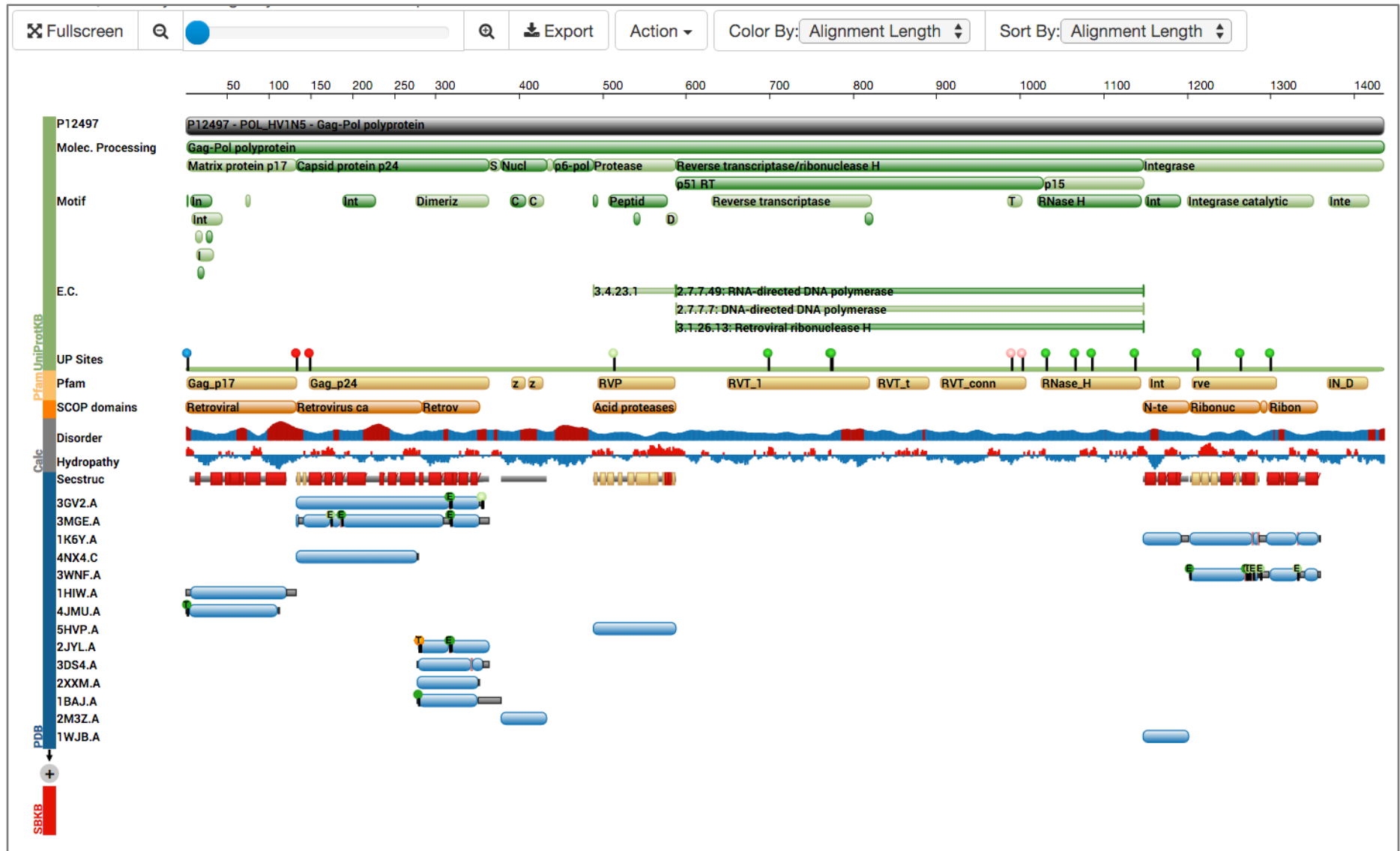
WebGL-based 3D viewers

Data Integration and Visualization

- Protein Feature View: Mutations, Validation
- Gene View: Mapping genome location to structure
- Visualization
 - Protein Viewer (PV) offers mobile-friendly 3D visualization
 - Large structure support
- External Reference Files (ERFs) on FTP Site
 - CSD ligand structures
 - BLASTClust and Cd-hit sequence cluster files
 - Phased release of data to support blinded docking tests

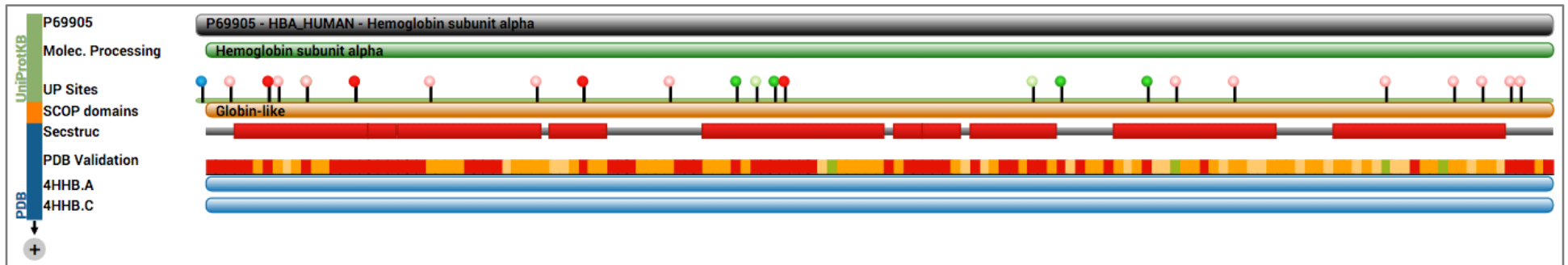


Mutation Information

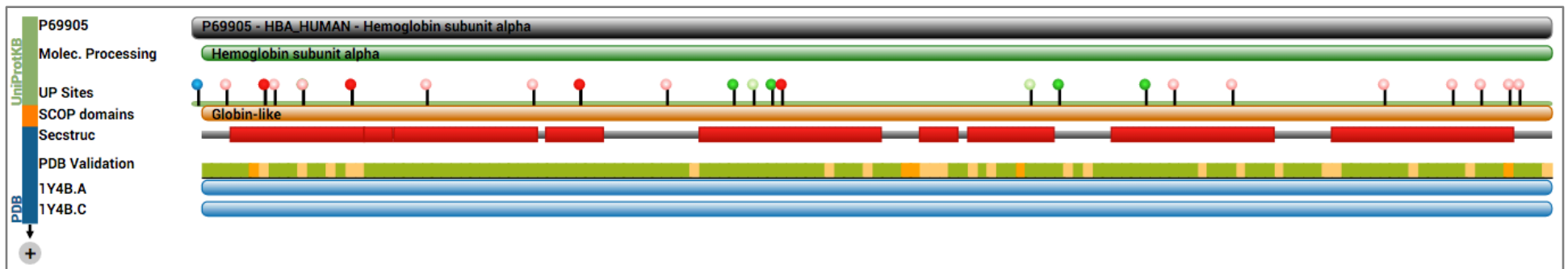


Validation Track

Two hemoglobin subunit A chains



Released: 1984

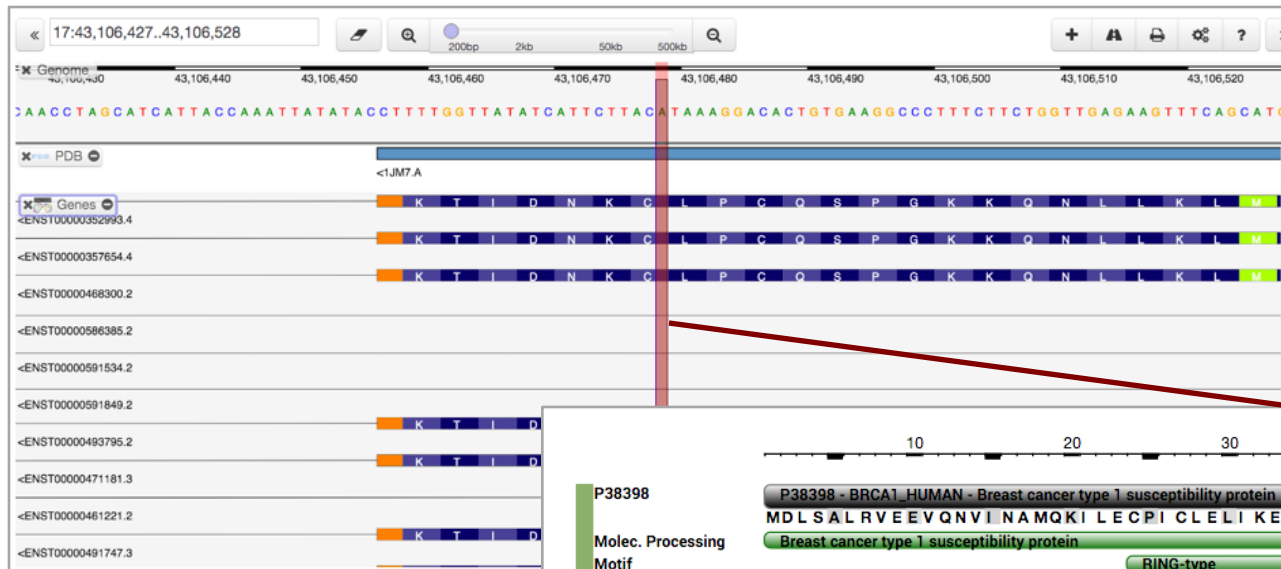


Released: 2005



From Gene to Protein Sequence & 3D

Gene View
biodalliance.org

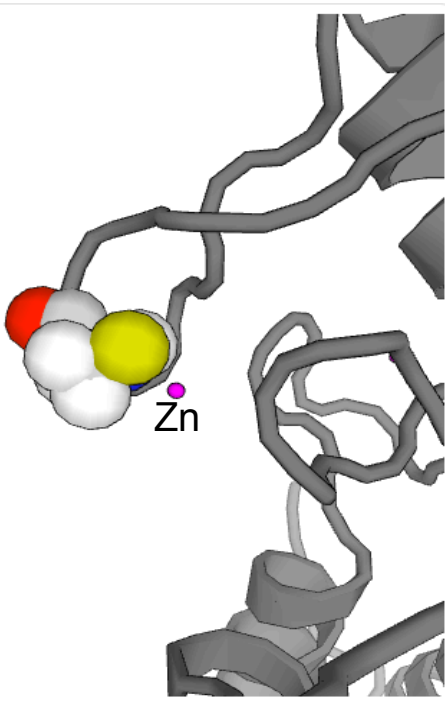


Protein Feature
View



PV (WebGL 3D Viewer)

Note: Use your mouse to drag, rotate, and zoom in and out of the structure.



Snapshot Detach Full screen Reload

Status: Load:

PV is a WebGL based 3D viewer.

Select a Viewer

Structure Details

Symmetry Type	<input type="text" value="Global Symmetry"/> ?
Symmetry	C2
Stoichiometry	A2

Structure Orientation

Select Orientation	<input type="text" value="Front C2 axis"/>
--------------------	--

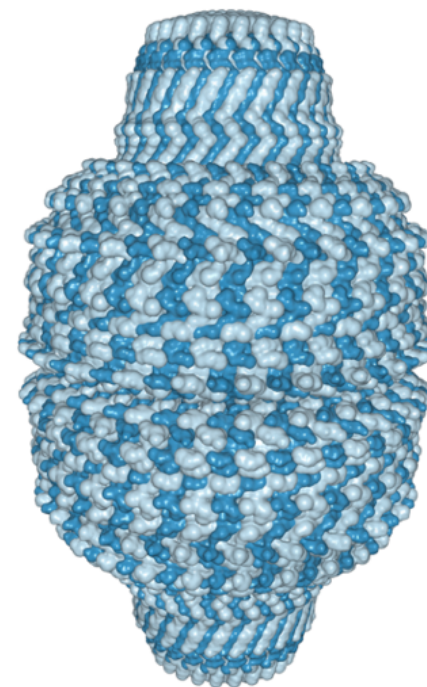
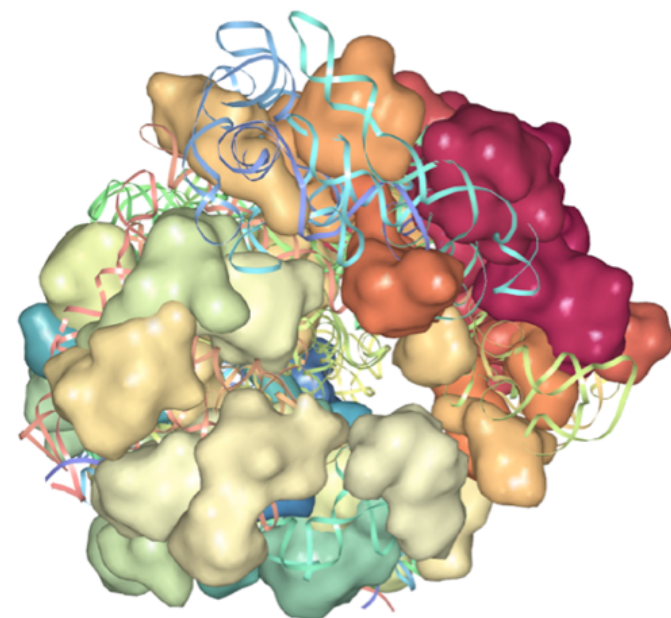
Display Options

Style	<input type="text" value="Cartoon"/>
Color	<input type="text" value="By Chain"/>
Symmetry Axes	<input checked="" type="checkbox"/>
Outline	<input checked="" type="checkbox"/>
Fog	<input checked="" type="checkbox"/>
Animation	<input type="checkbox"/>



Large Structure Support

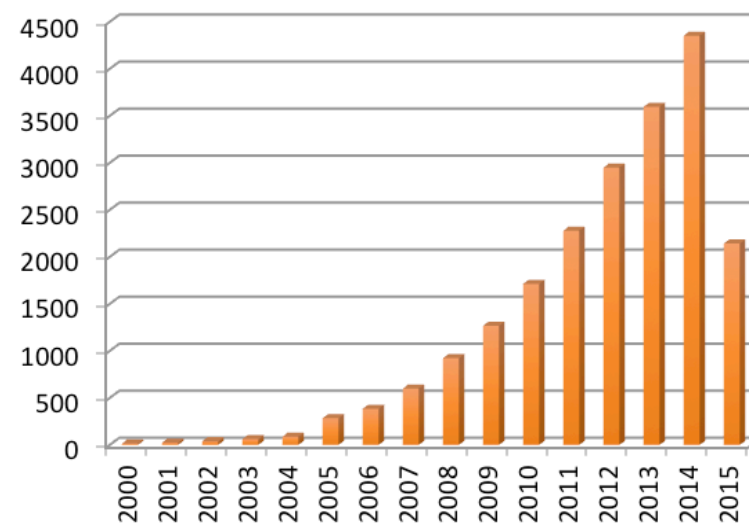
- Large and complex structures, historically split across several entries, merged into single PDBx/mmCIF files (Dec 2014)
- Visualization challenge: slow network transfer, file parsing, and rendering
 - Only represents C α /P backbone
 - Multiscale rendering
 - JavaScript/WebGL based viewers (JSmol, PV)
 - Separate NIH BD2K grant to handle large structure visualization and analysis



Integration with PubMedCentral (PMC)

- How are PDB entries used and mentioned in articles?
- NIH BD2K Data Discovery Index (DDI): [bioCADDIE](#)
- Links between publications and datasets
- Identified PDB ID mentions in full text PubMedCentral Open Access articles (~1 million articles)
- Note: PubMed contains ~25 million articles

PMC Open Access (OA) articles with PDB data mentions per year (Jan. 1, 2000-Aug. 1, 2015)



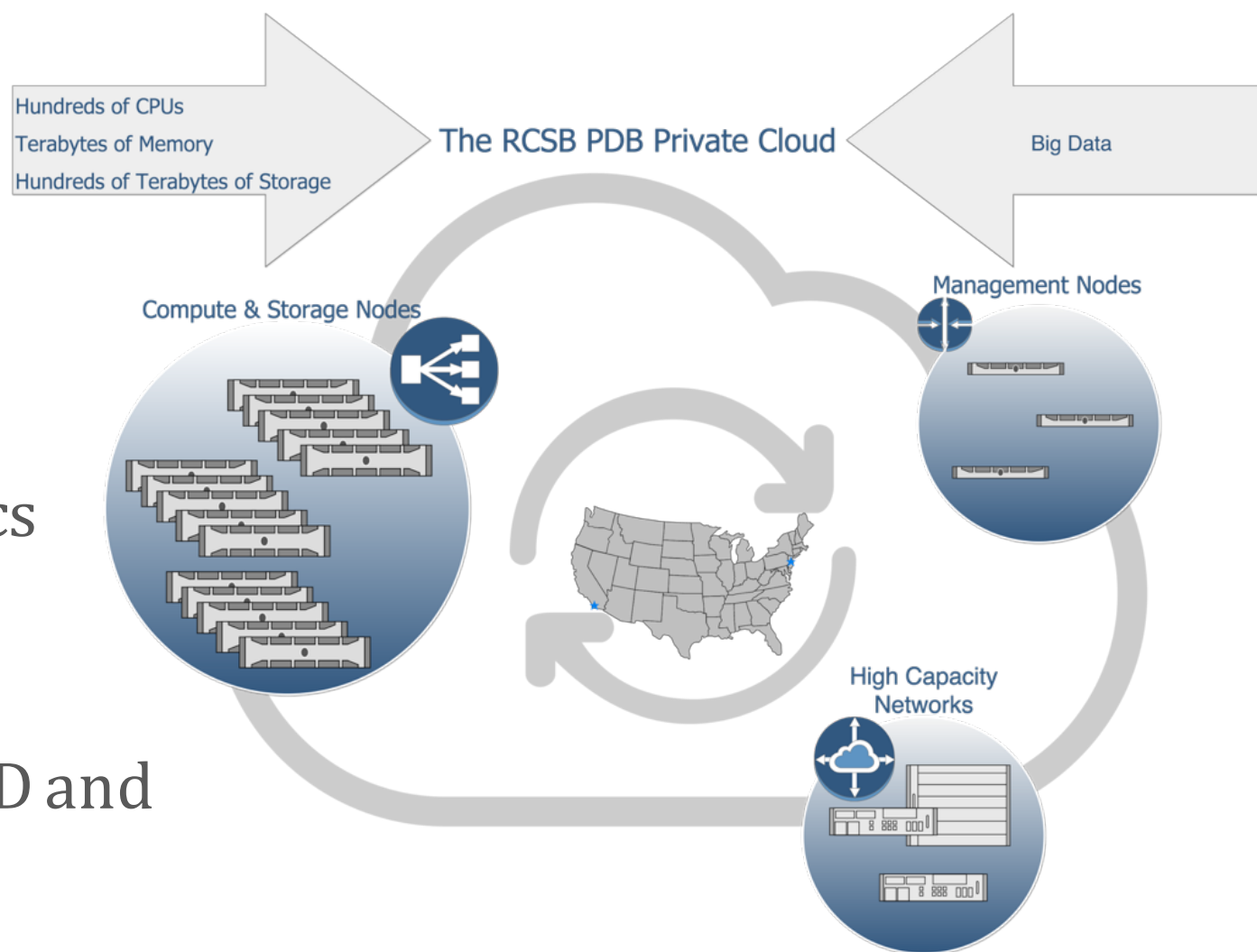
Number of unique PMC OA articles with PDB data mentions	20,731
Number of unique PDB IDs found in data mentions	45,209

~40% of PDB IDs mentioned in PMC OA articles



Hardware

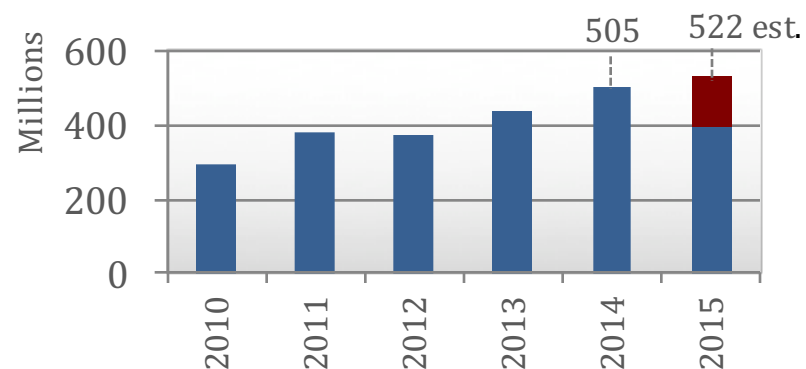
- Expanded capacity
- More elastic
- On-demand resources
- Better analytics
- Bicoastal load balancing
- Hosted at UCSD and Rutgers



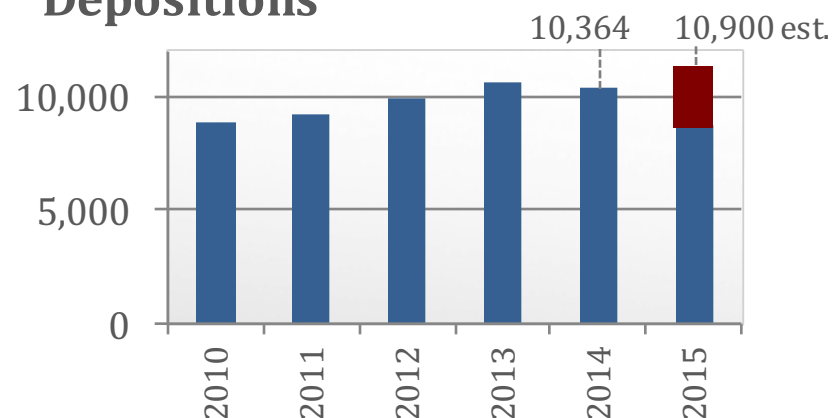
Archive Keeping

- “Archive Keeper” for wwPDB
- Maintain a secure and uniform version of archive
 - Package weekly update
 - Yearly snapshots
- Maintain **wwpdb.org**
- Monitor download statistics

FTP + Web Downloads

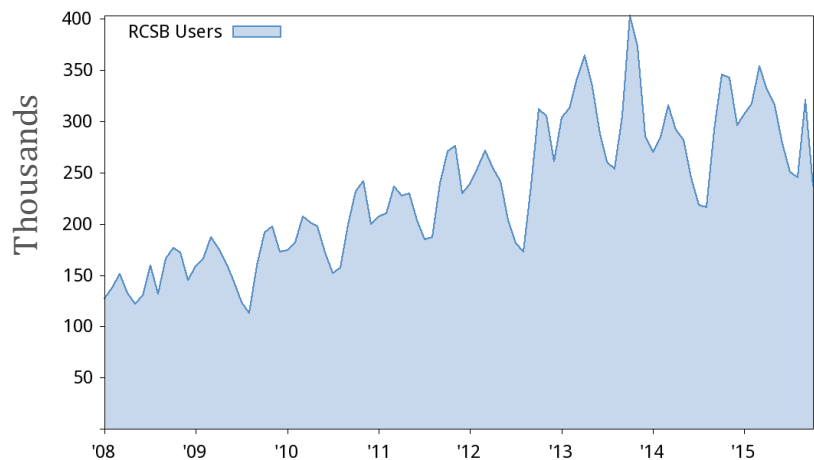


Depositions

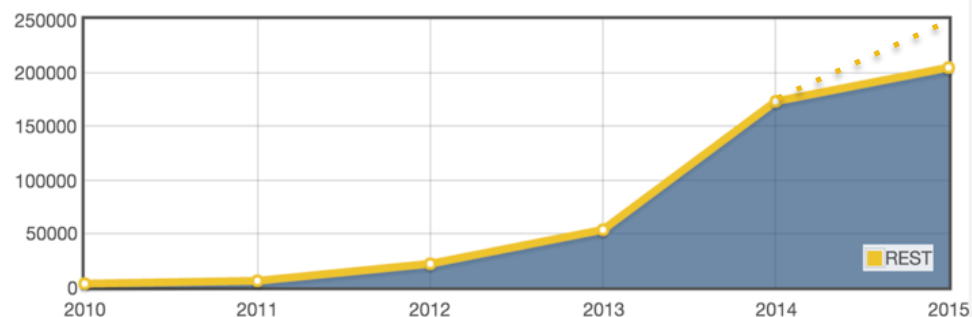


Analytics: Overall Usage

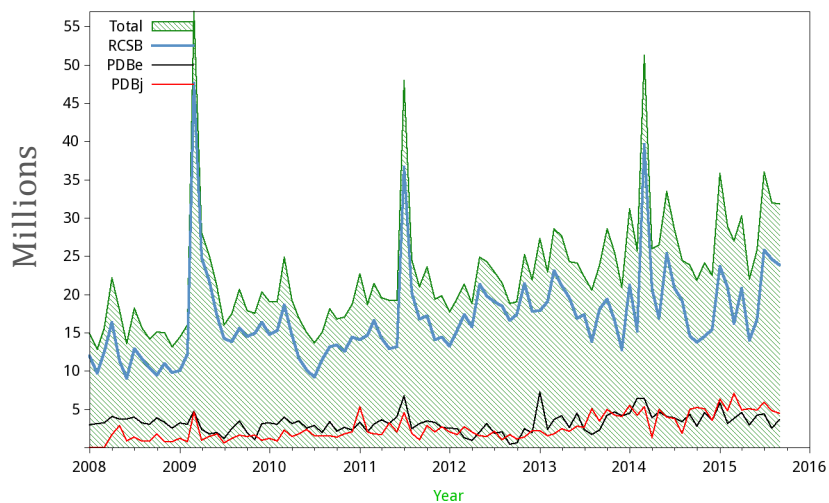
Website Unique IP addresses/month



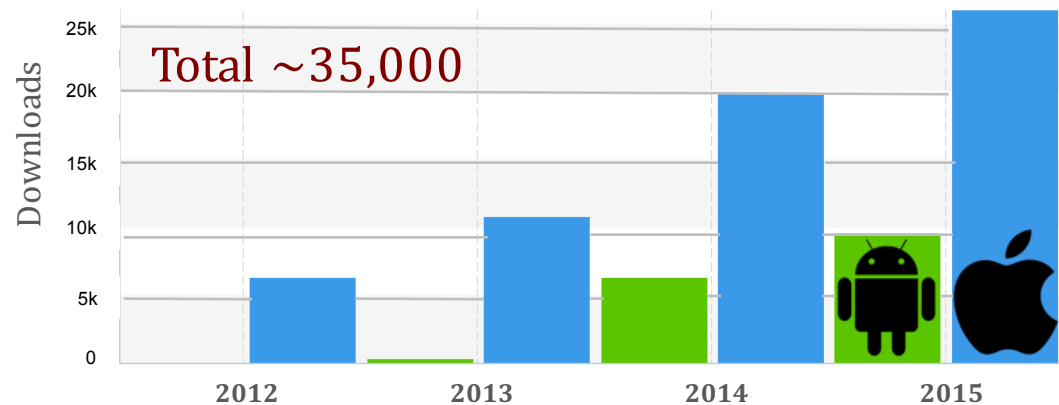
Web Service API Unique IP addresses/month



FTP downloads/month



RCSB PDB Mobile cumulative download

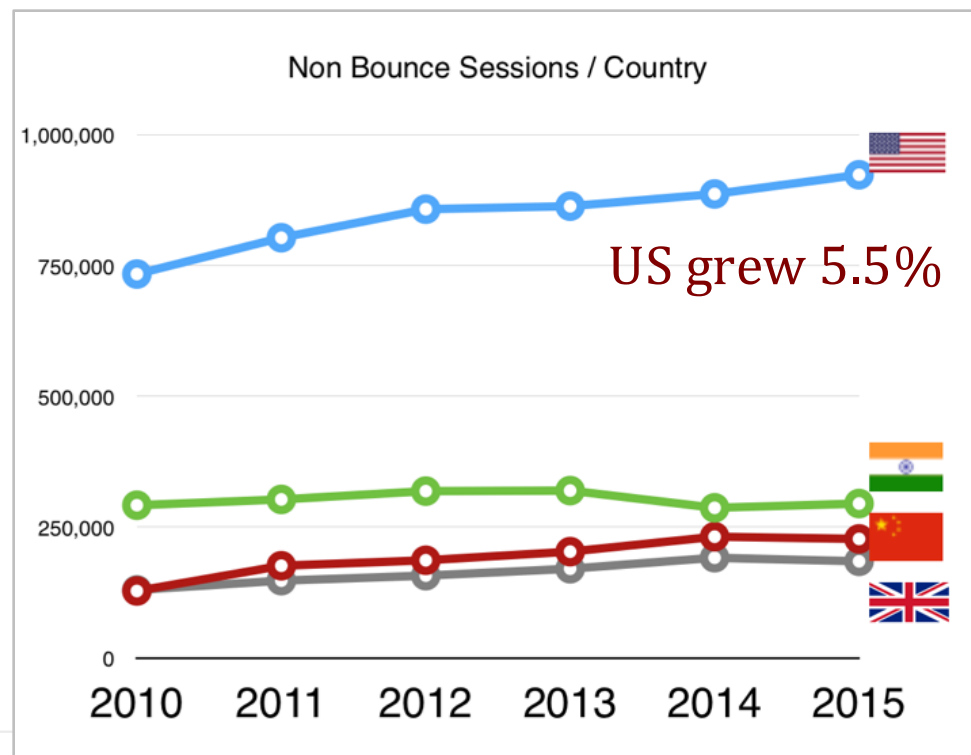


Analytics: Overall Website Usage

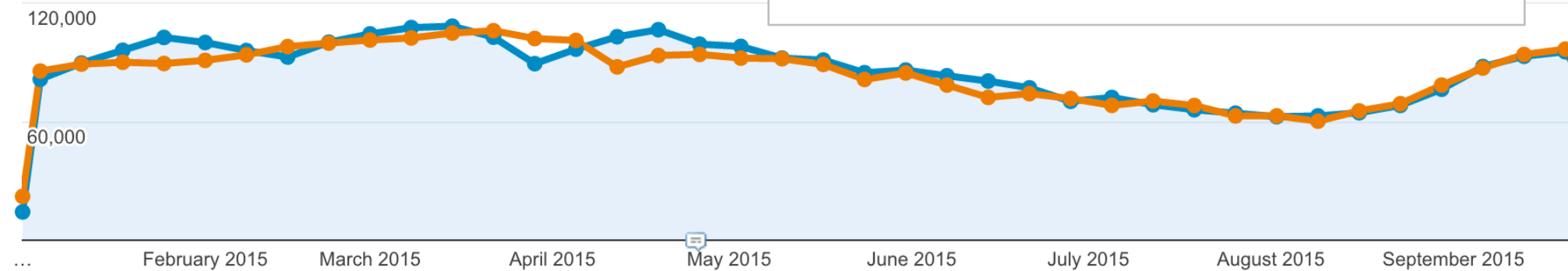
Non-bounce Sessions:

2014 → 2015: ↑ 2%

2013 → 2014: ↑ 2%



Jan 1, 2015 - Sep 30, 2015: ● Sessions
 Jan 2, 2014 - Oct 1, 2014: ● Sessions



Page Views by Category

Structure Summary Page:

2014 → 2015: ↑16%

Protein Feature View (PFV) comes in 2 flavors:

Simplified view on Structure Summary Page

PFV "full" page views:
2014 → 2015: ↑66%

1GAV
BACTERIOPHAGE GA PROTEIN CAPSID
DOI: 10.2210/pdb1gav/pdb

Classification: **Virus**
Deposited: 1997-01-28 Released: 1997-09-04
Deposition author(s): [Tars, K.](#), [Bundulo, M.](#), [Liljas, L.](#)
Organism: [Enterobacteria phage BZ13](#)
Mutation(s): 90

Structural Biology Knowledgebase: 1GAV (1 model >15 annotations) [SBKB.org](#)

Experimental Data Snapshot
Method: X-RAY DIFFRACTION
Resolution: 3.4 Å
R-Value Observed: 0.279

Metric	Percentile Ranks	Value
Clashscore		11
Ramachandran outliers		1.0%
Sidechain outliers		6.1%
RSRZ outliers		0

P04585 - POL_HV1H2 - Gag-Pol polyprotein

Molec. Processing: Gag-Pol polyprotein, Matrix pr, Capid protein p24, Nuc, gp6, Protease, Reverse transcriptase/ribonuclease H, Integrase

Motif: Dimer, Capt, Reverse transcriptase, RNase H, Integrase catalytic

E.C.: 2.7.7.49: RNA-directed DNA polymerase, 2.7.7.7: DNA-directed DNA polymerase, 3.1.26.13: Retroviral ribonuclease H

UP Sites: Gag_p17, Gag_p24, RVP, RVT_1, RVT_2, RVT_con, RNase_H, In, rve, IN

Pfam: Retrovir, Ga, Acid p, DNA/RNA polymerases, Ribonuc, Ribonuc

SCOP domains: Retrovir, Ga, Acid p, DNA/RNA polymerases, Ribonuc, Ribonuc

Secstruc: 2YNG.A, 2YNG.B, 1EX4.A, 1EXQ.A, 101W.A, 1TAM.A, 3KT2.A, 4U7V.A, 2K0D.A, 1ESK.A, 1NCP.N

PDB Validation: 1TIM.A, 1TIM.B

Homology Models: SBKB

P00940 - TPIS_CHICK - Triosephosphate isomerase

Molec. Processing: Triosephosphate isomerase

E.C.: 5.3.1.1: Triose-phosphate isomerase

UP Sites: Triosephosph

SCOP domains: Triosephosphate isomerase (TIM), Triosephosp

Secstruc: Triosephosphate isomerase (TIM), Triosephosp

PDB Validation: 1TIM.A, 1TIM.B

Homology Models: SBKB

Page Views by Category, continued

Searches

2014 → 2015: ↑34%

Advanced Search Interface

UniProt Gene Name Result Count

Find PDB entities linked to a UniProt Gene Name (e.g. HBA1).

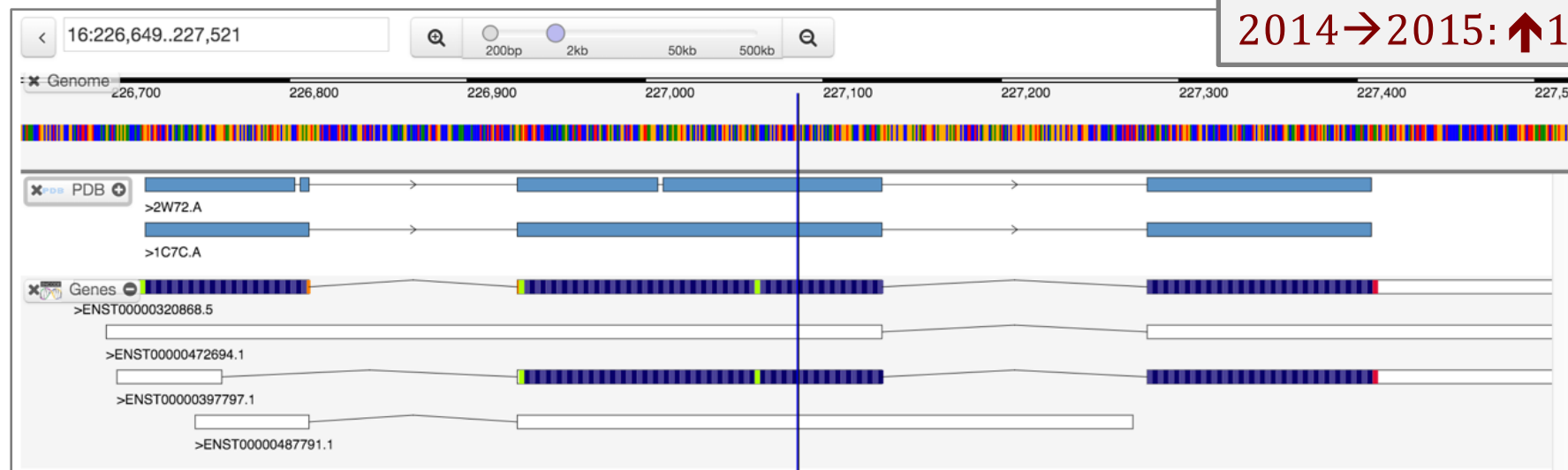
UniProt Gene Name

Add Search Criteria

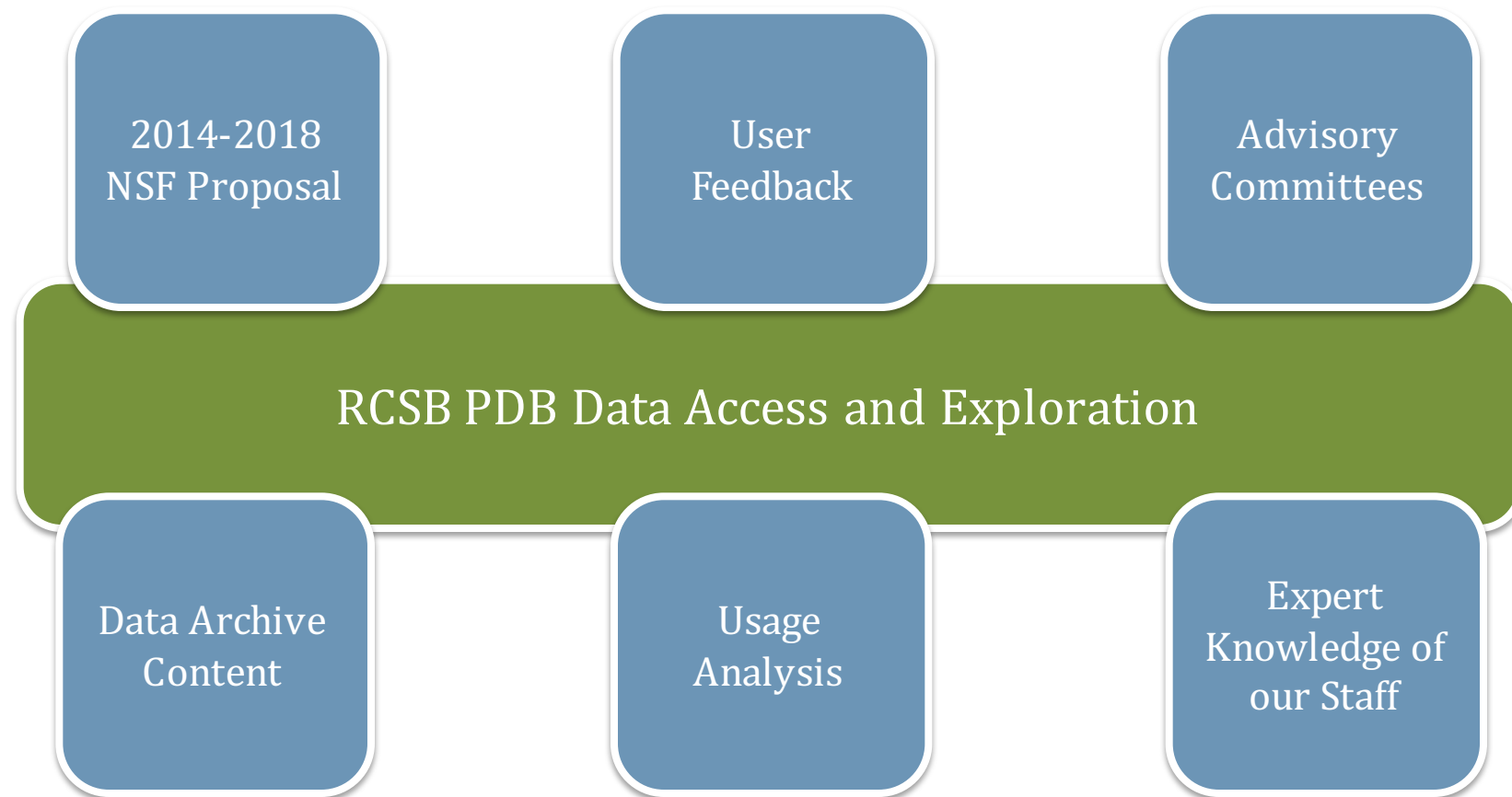
Retrieve only representatives at 90% sequence identity Match all of the above conditions. Clear All Parameters Submit Query

Gene View

2014 → 2015: ↑111%

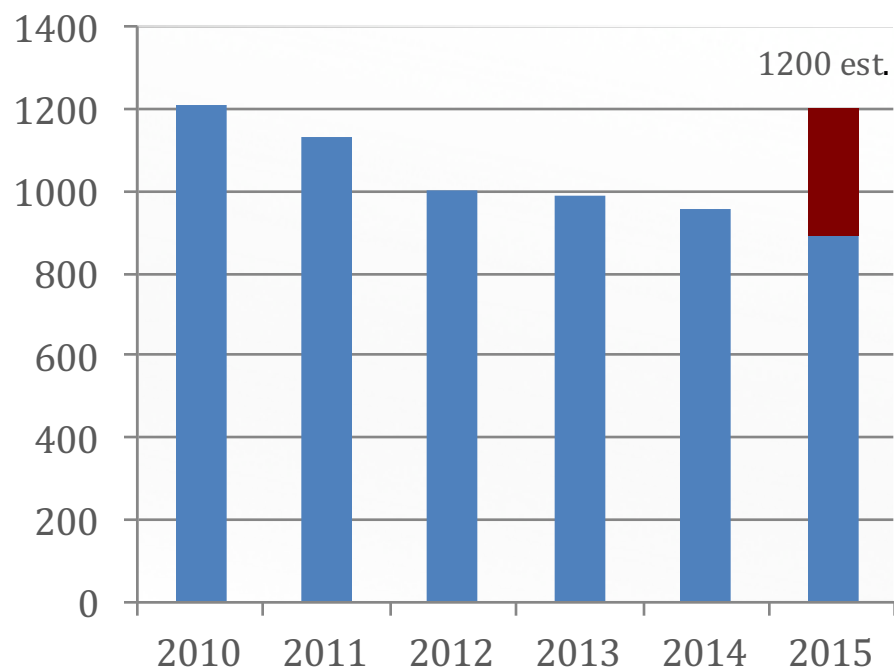


Factors Influencing Development

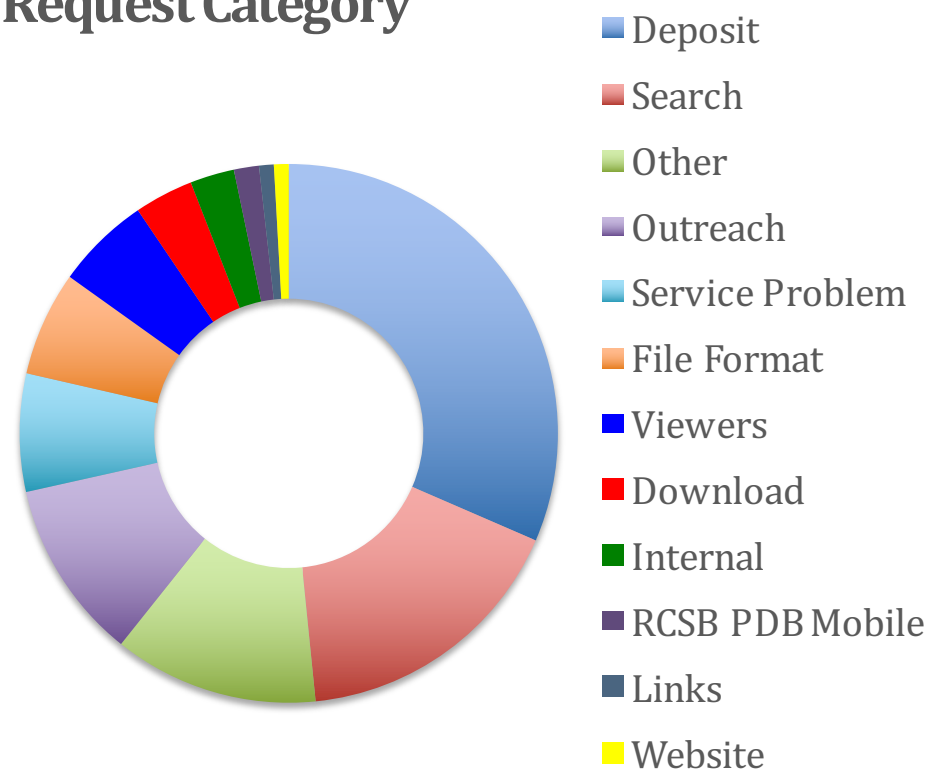


Customer Service: Help Desk (info@rcsb.org)

Requests



Request Category



Education Efforts

Shuchismita Dutta, Ph.D.

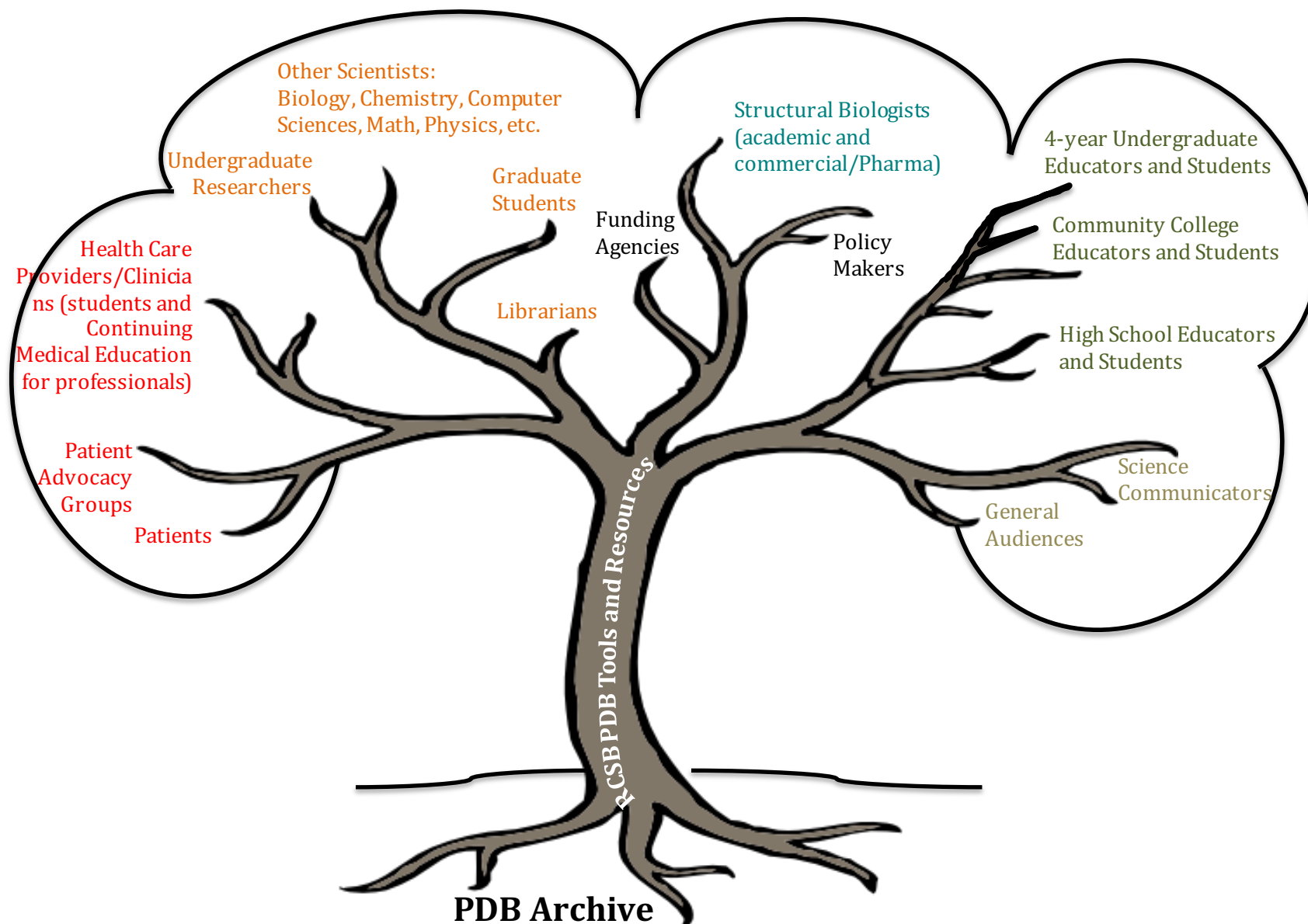


From RCSB PDB AC 2014 Report

While the Committee found the array of PDB outreach and education activities to be truly impressive, there was concern that these are taking too many person-hours and resources during a period in which the budget has continued to shrink. The Committee fully endorses the importance of these activities and also recognizes that broader impacts is a review criterion for NSF proposals. Educational activities at both Rutgers and UCSD are also important for maintaining strong institutional ties. Given the current funding situation, however, the Committee recommends that the RCSB thoroughly evaluate their education plan, articulate a set of goals and draw up a focused education plan that will maximize impact and return. As part of this process, the Committee recommends that the RCSB leadership determine how to assess the impact of different programs and activities, both to make decisions about which programs to continue as well as to periodically evaluate programs in the future. In light of comments from Dr. Peter McCartney of the NSF (see section below), the Committee encourages the RCSB PDB to seek partners who could help in education and outreach efforts, rather than becoming directly involved.



RCSB PDB: A Community Resource



Overarching Principles and Goals

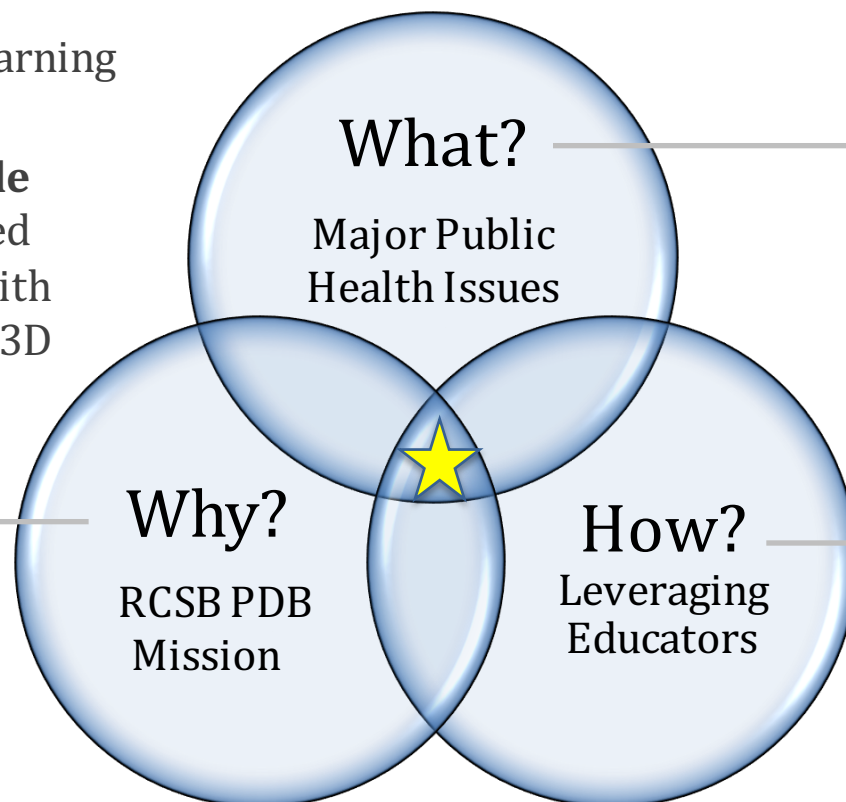
Outreach	Education
<p>Inform and receive feedback about PDB data, RCSB PDB tools, services, resources</p>	<p>Explain/Teach about Biomolecules in general Foundations of PDB Data RCSB PDB resources</p>
<p>Engage users with tools, services, and resources that Meet community needs Are easy to use</p> <p>Inspire with Examples of resource applications</p>	<p>Transform/Enable the application of RCSB PDB resources to Explore biology/medicine-related topics that impact society Promote a Structural View of Biology and Medicine</p>
<p>Leverage community interactions Through known experts, and champions</p>	<p>Collaborate with teaching professionals to develop materials for Undergraduate students High school teachers and students</p> <p>Reuse developed materials Suitably adapted for healthcare professionals and patient advocates</p>



Focusing Education Efforts

Education Goals

- 1. Explain/Teach**
Exploration and learning in STEM fields
- 2. Transform/Enable**
Promoting informed decision making with comprehension in 3D



- HIV/AIDS
- Diabetes
- Other Global Health concerns

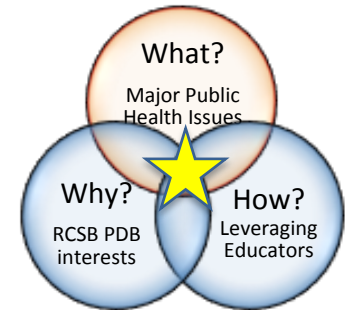
- Promote understanding of biomolecules and the PDB
- Structural View of Biology and Medicine

- Professional Development
- Curricular materials, aligned with current Education Standards

 RCSB PDB Educational Activities



Selecting Educational Themes



HIV/AIDS

Diabetes

Social Impact

- 50,000 new infections/year in US
- Youth rates (age 13-24) very high (26%)
- Lack of understanding and misconceptions

- ~1.5 million new diagnosis/year in US for adults (age 18-79)
- Increasing numbers of Type II in <18 year olds
- Incidence rates higher in individuals with <high school education

Economic Impact

Socio-economic consequences
Rising health costs

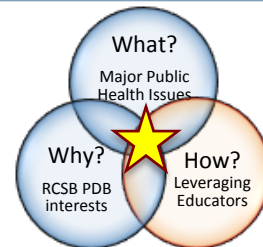
RCSB PDB Mission

- Highlight academic biomed research
- Structure-based Drug Design (SBDD) Success Stories → Drugs used in cART
- Visualize basis for Drug Efficacy and Resistance

- Designer Insulins
- SBDD Success Stories → new drugs
- Understanding molecular mechanisms for action for Type II Diabetes Mellitus drugs



Theme-Based Education Strategy



RCSB PDB Educational Task

Develop/Teach undergraduate Honors course



Develop online RCSB PDB curriculum



Test/Refine curriculum



Promote curricular modules



Possible repackaging for other audiences

In collaboration with

Subject matter-experts

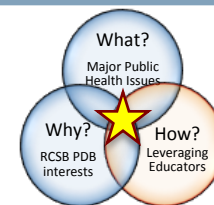
Subject matter-experts
High school teachers

High school teachers
Educational consultants

High school teachers
Educational coordinators
& Related societies

Healthcare professionals
Patient advocates





Theme-Based Education Efforts






















Activity	HIV/AIDS (2014-2016)	Diabetes (2015-2017)
Undergraduate honors course	2014 (Relevant PDB entries and initial research)	2015, 2016 (Relevant PDB entries and initial research)
Promote/develop <i>Molecule of the Month</i> articles	HIV-1 Protease, HIV Reverse Transcriptase, HIV Capsid, HIV Envelope Glycoprotein	Insulin, Leptin, Insulin Receptor, Glucagon, Receptor for Advanced Glycation End Products
Create molecular origami	HIV-1 Capsid, Antibody (related)	Insulin
Develop curricula	V1.0 (2014), V2.0 (2015)	V1.0 (planned 2016)
Test curricula	Pilot (2014-15), Efficacy (2015-16-17)	Efficacy (2016-2017)
Host symposia	NJ Science Convention for teachers (2014) NSTA Regional for teachers (2015) *World AIDS Day Symposia: 2013 (Robert Wood Johnson AIDS Program); 2014 (FXB Center, AIDS Education Training Center National Resource Council)	SACNAS for Chicano, Hispanic, Native American students/faculty (2015) NJ Science Convention for teachers (2015) *World Diabetes Day, planned for 2016
Mentor independent research	Rutgers undergraduates (2014) Rutgers graduate student (2015-2016)	Rutgers undergraduates (2014-2015)
Host Video Challenge	2014: 23 entries (4 states) 2015: 38 entries (5 states)	2016: planned 2017: planned



Collaboratively-Developed HIV/AIDS Curriculum

- Broad Focus
- Lessons directly related to RCSB PDB materials
- Curricular Modules
 - Biomolecular Structures and Models
 - Molecular Immunology
 - Molecular View of HIV/AIDS
 - Includes “Skills box”

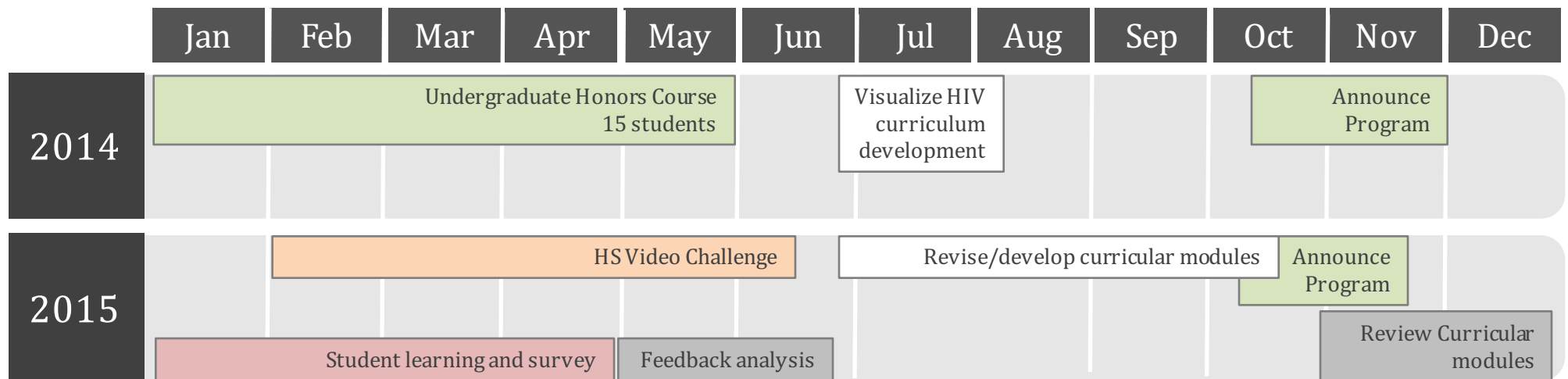
Overview	Biomolecular Structures and Models
Biomolecular Structures and Models	
Molecular Immunology	
Molecular View of HIV/AIDS	
Getting Started >	
Learning Objectives >	

Skills	How To	
	Read a Scientific Paper	Guide  Quick review 
	Make sense of the Information Out There	Guide and exercise  Teaching Notes  Additional Reading 
	Explore a Topic of Interest at a Molecular Level	Guide  Quick Review  Worksheet  Learning about Asthma 
	Find Structures Relevant to a Topic of Interest	Guide  Quick Review  Selected structures related to Asthma 
	Visualize, Analyze and Compare Selected Structures	Cheat Sheet  Basics  Menus  Selections  Structure Analysis  Structure Comparisons 
	Make Publication Quality Images and Movies	Guide  Part1  Part2 



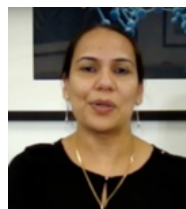
2014-2015 Curriculum Development

- Collaborative Curriculum Development Program (2014)
 - 5 HS teachers (NJ), educators, scientists, clinicians
- Presented to NJ teachers at 2014 convention
- Pilot Testing (2014-2015)
- Professional Development Workshop (2015)
 - 4 HS teachers (NJ), educators
 - Reviewed curricular material → Modules

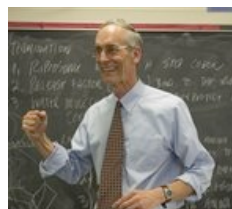


Contributors

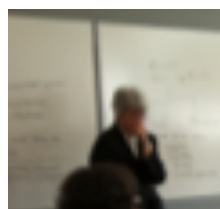
Teachers and Educators



A. Bhatia
(Teacher)



B. Buck
(Teacher)



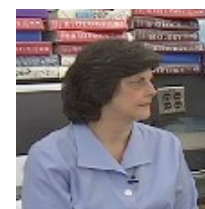
S. Coletta
(Educator)



K. Lucci
(Teacher)



W. Mott
(Teacher)



A. Sanelli
(Teacher)



M. Witkowski
(Teacher)

Subject Matter Experts



T. Bhowmick
(Infectious Disease)



L. Denzin,
(Immunology)

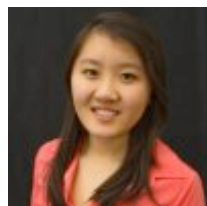


J. Dougherty
(Pharmacology)



J. Nelson
(HIV, Education, Nursing)

Rutgers Interns



B. Lin



A. Nazzaro



K. Shah



D. Goodsell
(Scientist/Artist/
Educator)

RCSB PDB Team



2015 HIV/AIDS Pilot Testing

(25 testers from USA, 18 from rest of world)



Revised curricular modules based on feedback released November 1, 2015. Module review and testing taking place in academic year 2015-2016.

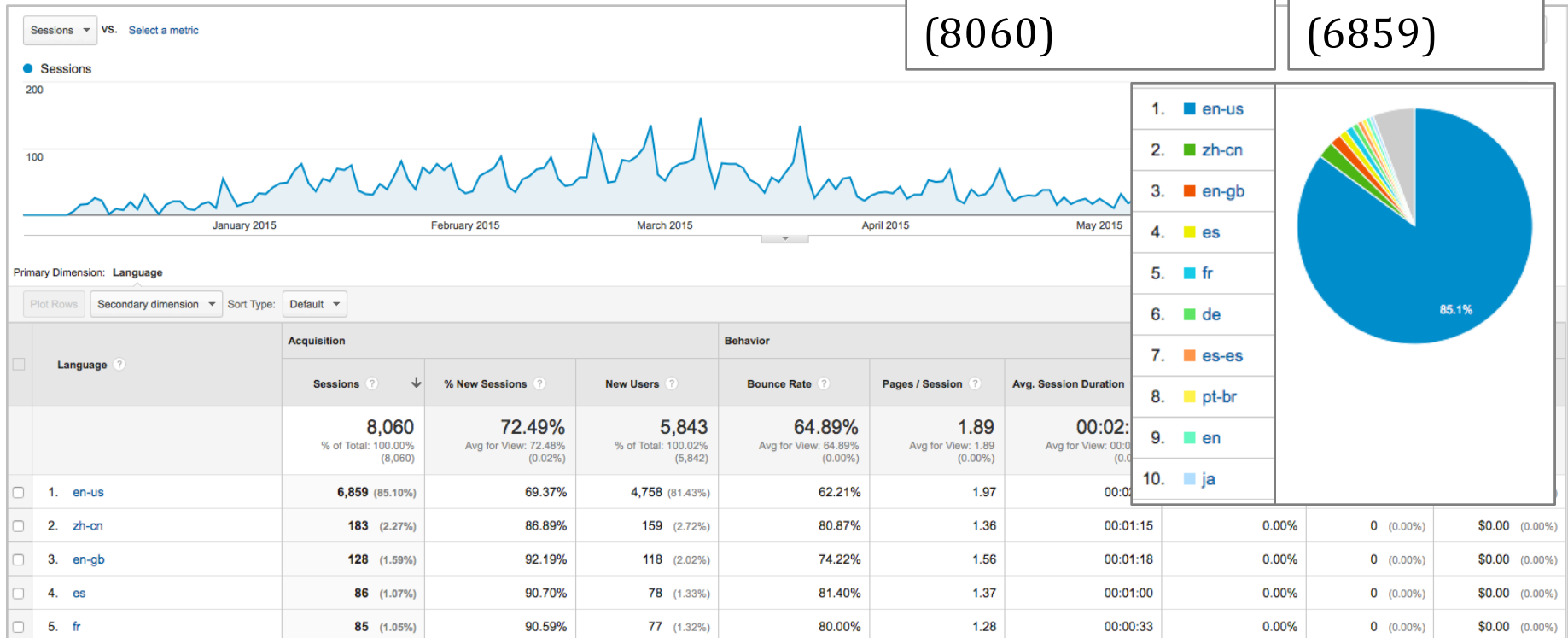


Monitor Usage: Google Analytics

Period: Dec 1, 2014 to Jun 30, 2015

Overall Sessions
(8060)

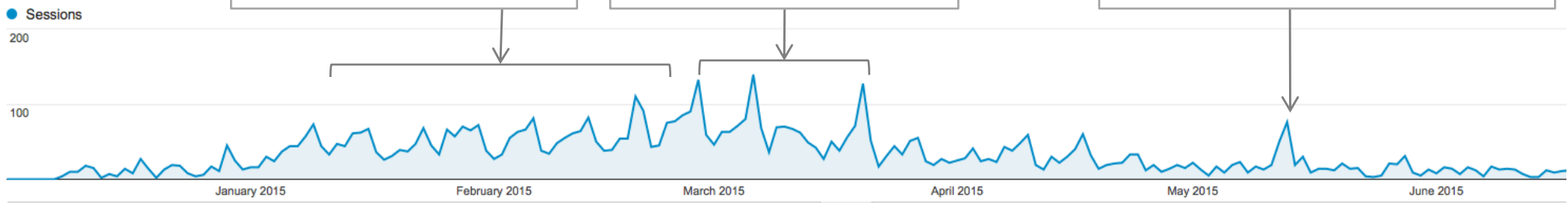
US Sessions
(6859)



1/9 to 2/13: ~75-80
Student Class on Fridays?

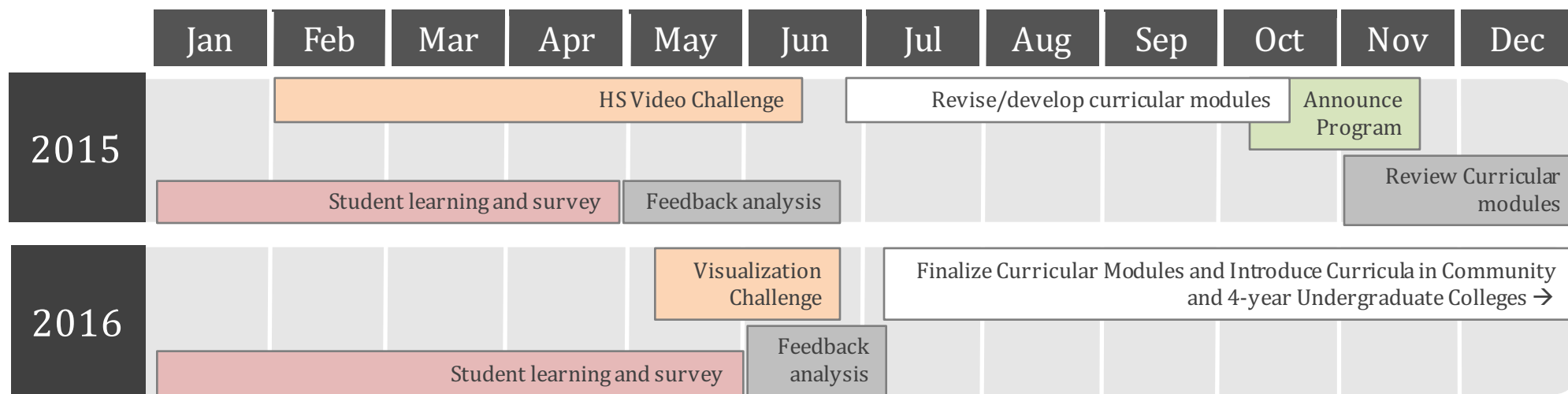
2/27 to 3/20: ~130
Student Class on Fridays?

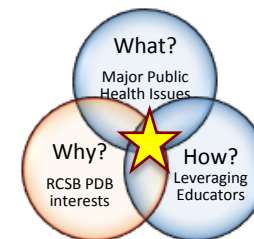
5/13: Prep for Video Challenge
(after AP Bio exam on 5/11/2015)



2015-2016 Curriculum Development

- Present updated materials to teachers
 - Rutgers Waksman Summer Program participants (Oct 2015)
 - National Science Teachers Association (NSTA) Regional meeting (Nov 2015)
 - Advertise curricular modules online, through contacts and *via* social media
- Review, field testing and independent assessment of efficacy (2015-2016)





Measuring Success

Short Term

- **Teacher interest** in Curricular modules
 - *Document teacher participation*
 - *Google Analytics*
- **Assess efficacy** of Curricular modules
 - *Independent evaluation*
- **↑Efficiency** in Curriculum development/and efficacy testing
 - *Time and resources needed to develop curricular modules*

Long Term: ↑

- **Teacher partnerships and utilization**
- **Use of 3D structural data** by teachers and students
 - *Google Analytics*
- **High quality images/animations** in publications/presentations
 - *Google Images*
 - *Reproduction requests*
- **Brand Loyalty:** Experts and general audiences keep coming back
 - *Google Analytics*
 - *Survey*
 - *Internal NSF statistics*
 - *Increased understanding (Improvement in Pre- vs. Post-tests)*



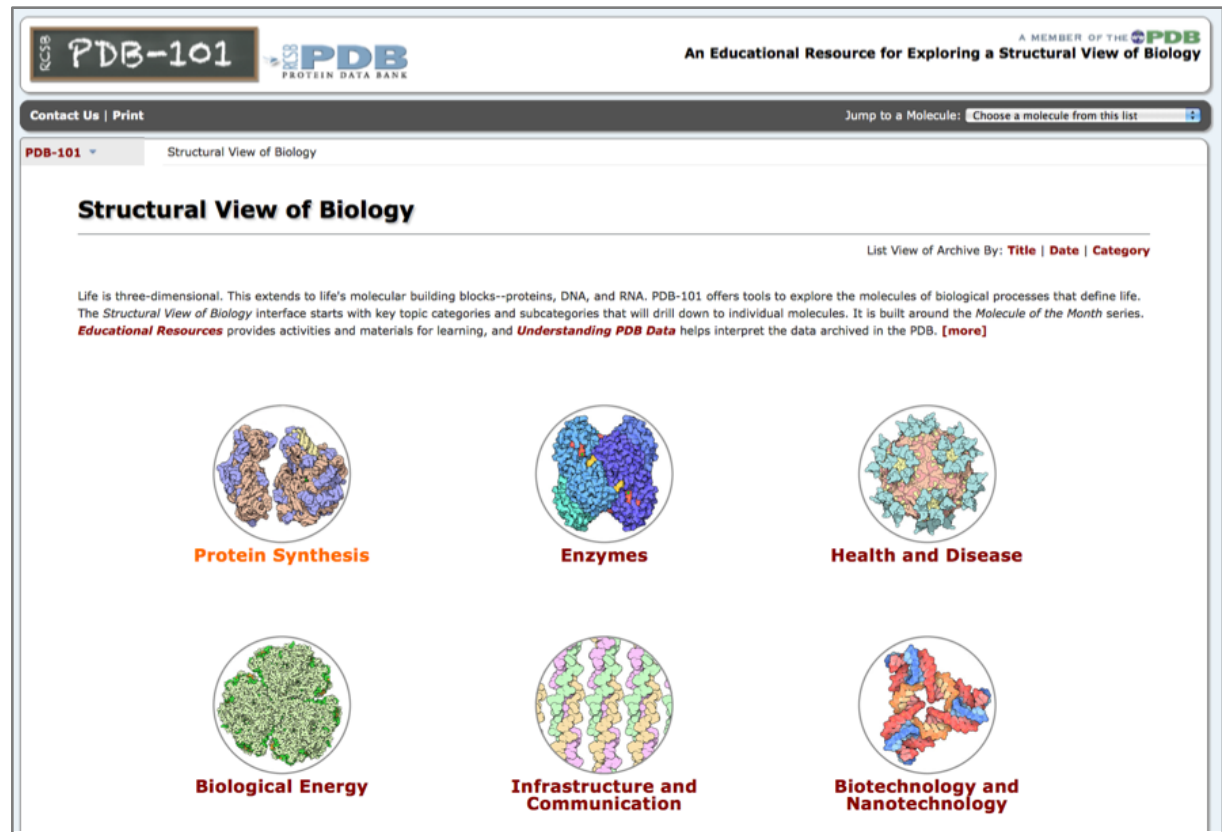
Next Generation PDB-101 Website

David S. Goodsell, Ph.D.



PDB-101: Established 2011

- Online portal for teachers, students, educational developers and the general public
- Integrated the growing number of materials developed by RCSB PDB
- Enabled topic-based exploration of PDB structures through *Molecule of the Month* articles



The screenshot displays the PDB-101 website interface. At the top, it features the RCSB PDB logo and the text 'An Educational Resource for Exploring a Structural View of Biology'. Below the navigation bar, the main heading is 'Structural View of Biology'. A sub-heading reads 'List View of Archive By: Title | Date | Category'. A paragraph of introductory text states: 'Life is three-dimensional. This extends to life's molecular building blocks--proteins, DNA, and RNA. PDB-101 offers tools to explore the molecules of biological processes that define life. The *Structural View of Biology* interface starts with key topic categories and subcategories that will drill down to individual molecules. It is built around the *Molecule of the Month* series. **Educational Resources** provides activities and materials for learning, and **Understanding PDB Data** helps interpret the data archived in the PDB. [more]

Below the text, there are six circular icons, each representing a biological topic with a corresponding molecular model:

- Protein Synthesis**: A model of a ribosome.
- Enzymes**: A model of a protein with a substrate.
- Health and Disease**: A model of a protein structure.
- Biological Energy**: A model of a protein structure.
- Infrastructure and Communication**: A model of a protein structure.
- Biotechnology and Nanotechnology**: A model of a protein structure.



2014 Advisory Committee Report

Following the recommendation of the Committee in 2013, a survey was conducted to assess the effectiveness of PDB-101 and determine the nature of the user community and how this resource is utilized. Of the ~700 respondents, fully 83% had never deposited a structure and 60% were students (28% undergraduates, 32% graduate students), pointing to the importance of PDB-101 as an educational tool. The most popular feature was the Molecule of the Month, which, combined with the popularity of these images in textbooks and other media, reinforces the impact of this particular activity. The Molecule of the Month page on Ebola virus proteins garnered over 10,000 views in just a few weeks in October 2014, reinforcing the importance of PDB-101 as a resource for educating the public about the molecular basis of disease.



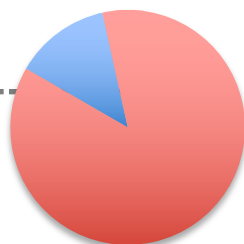
2014 Survey: Who Uses PDB-101?

Based on
~700
Responses

Demographics

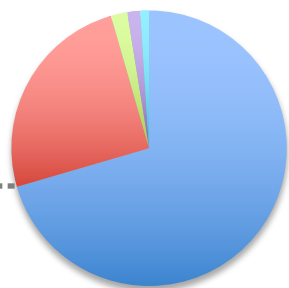


Ethnicity



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

Race



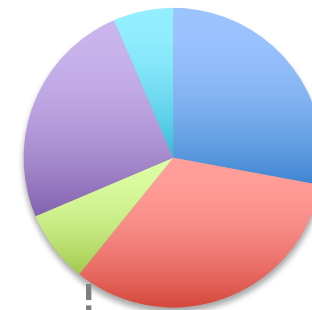
- White (72%)
- Asian (25%)
- American Indian or Alaska Native (2%)
- Black or African American (<2%)
- Native Hawaiian or Other Pacific Islander (1%)

83% have never deposited a structure

Where do they work?

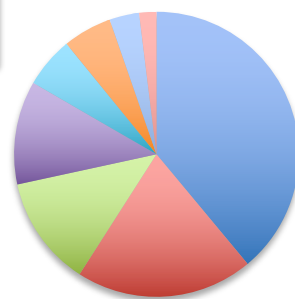
- College/University (74%)
- Research Institute (12%)
- Government (2%)
- Pharma/drug discovery/biotech (3%)
- K12 (8%)

Breakdown



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

Research Area



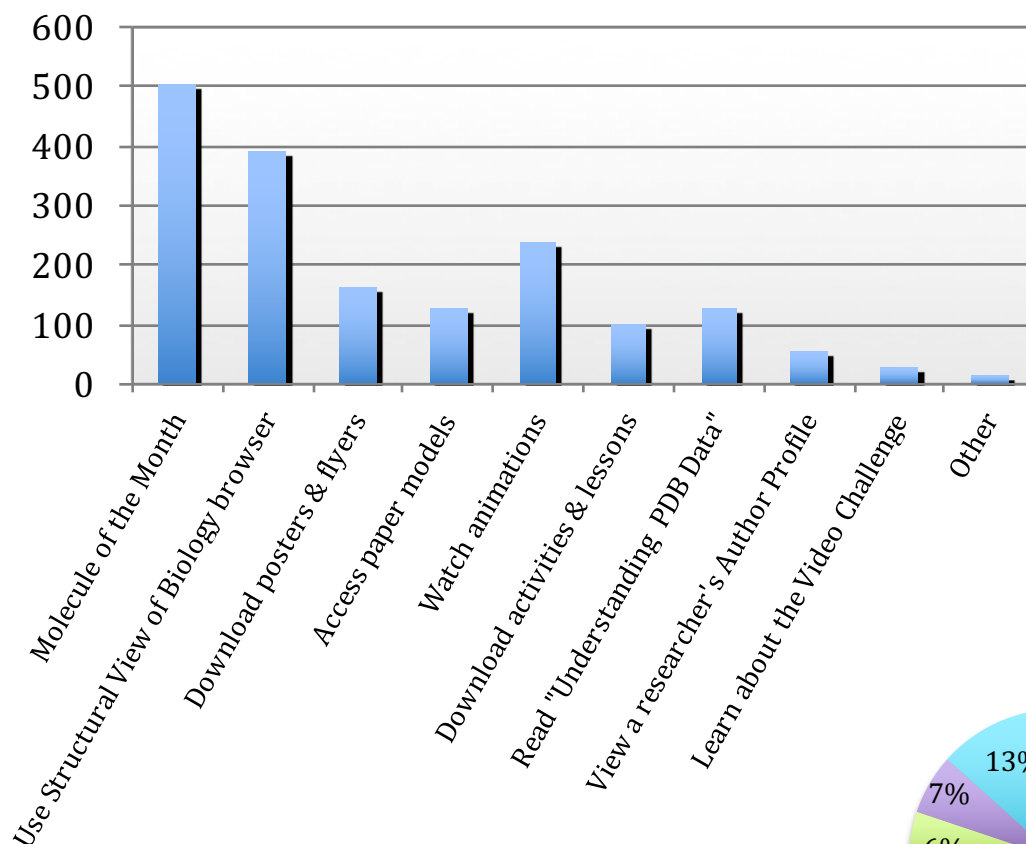
- Life Sciences (39%)
- Chemistry (20%)
- Medical Sciences (12%)
- Computational Sciences (12%)
- Pharmacology (6%)
- Other (6%)
- Physics (3%)
- Math/Statistics (2%)

College/University Type

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

What Do PDB-101 Users Do?

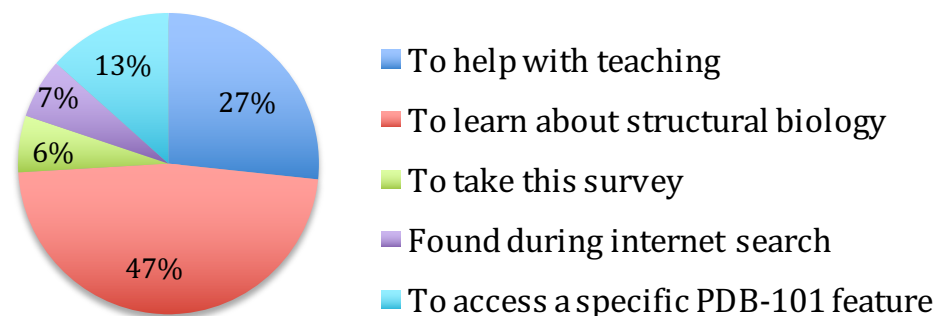
Why do they use PDB-101?



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%)

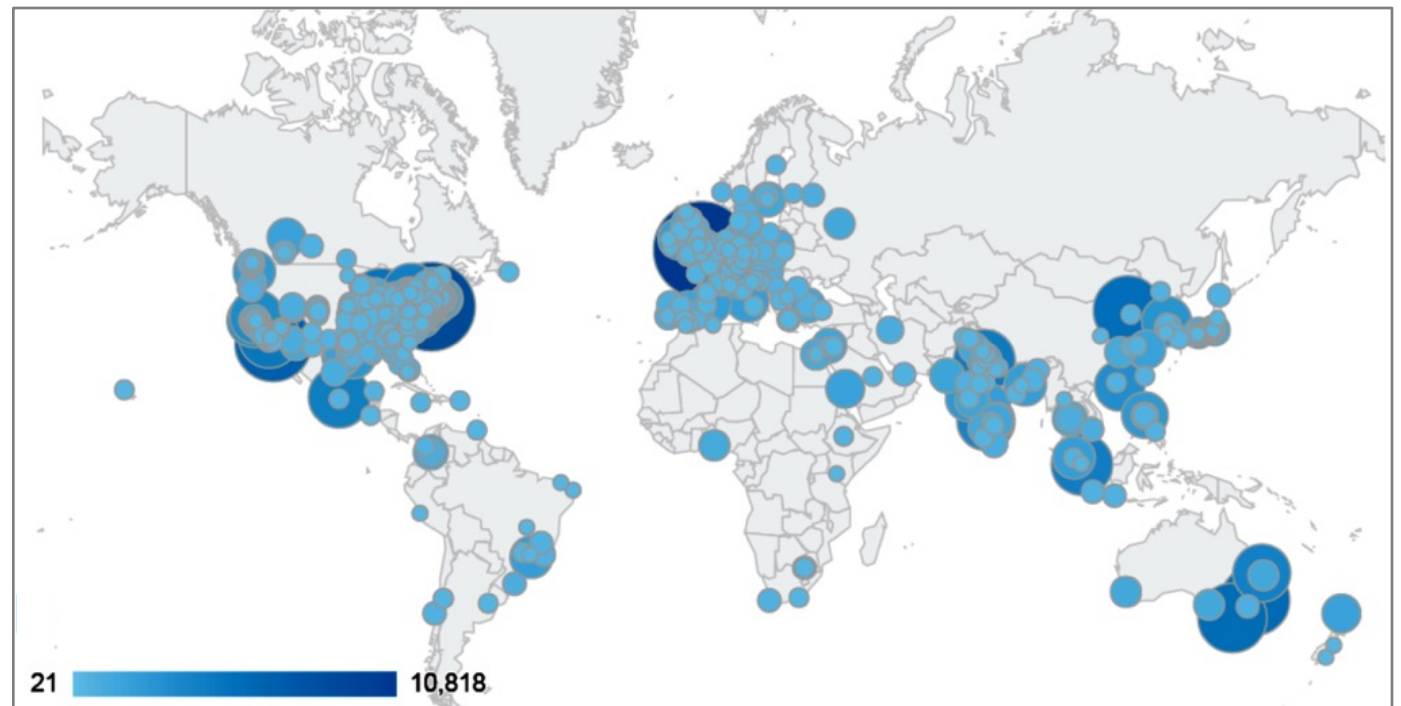
Why did they visit PDB-101 today?



PDB-101 Usage

January 1, 2015 – September 30, 2015

- ~630,000 sessions
- ~12% of total **rCSB.org** traffic
- Top cities
 - London
 - New York
 - New Delhi
 - San Diego
 - Sydney
 - Melbourne
 - Beijing
 - Chicago
 - Los Angeles



Current PDB-101

- Growing content
 - 191 *Molecule of the Month* articles
 - Activities: molecular origami
 - Educational lesson plans
 - Posters, flyers, animations
- Challenges
 - No search functionality
 - Browse-by-topic category only returns *Molecule of the Month* articles
 - Hard to find new content
 - Limited technical design requires complicated update process and affects rcsb.org

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

Protein Synthesis Enzymes Health & Disease
Biological Energy Infrastructure & Communication Biotechnology & Nanotechnology

PDB-101 News
Molecule of the Month:
Two-component Systems

2015 HS Video Challenge Winners
Follow us on **Twitter** and **Facebook**

About PDB-101
Life is three-dimensional. This extends to life's molecular building blocks--proteins, DNA, and RNA. PDB-101 offers tools to explore the molecules of biological processes that define life. The **Structural View of Biology** interface starts with key topic categories and subcategories that drill down to individual molecules. It is built around the **Molecule of the Month** series, which regularly describes the structure and function of a molecule. **Educational Resources** provides activities and materials for learning, and **Understanding PDB Data** helps interpret the data archived in the PDB. **Author Profiles** are a new and unique historical and educational tool that offers a timeline display of all structures associated with a particular researcher. The RCSB PDB develops these resources to support exploration of the structures found in the Protein Data Bank archive of experimentally-determined structures of proteins, nucleic acids, and complex assemblies.

www.rcsb.org/pdb-101

Redesign Goal: Enhance User Experience

- Improved User Interface
 - New keyword search capabilities for *Molecule of the Month*
 - Improved browser
 - New menus
- Highlight all content
 - More features visible and accessible from home page
 - Less clicks needed to access materials for teaching and learning
- Technical improvements
 - Modular, extensible, and independent from **rCSB.org**
 - Easier to update and add content
 - Leverages RCSB PDB software development

Beta test site: pdb101.rcsb.org

The screenshot displays the PDB-101 website interface. At the top, there is a navigation bar with links for 'Molecule of the Month', 'Browse', 'Learn', 'Teach', 'Events', 'Gels Archive', and 'More'. Below this is a search bar for 'PDB-101' articles and resources. The main content area is divided into several sections:

- Molecule of the Month (July 2015):** Features a 3D model of New Delhi Metallo-β-Lactamase. The title is 'New Delhi Metallo-β-Lactamase'. A brief description states: 'Antibiotics can save lives, but antibiotic-resistant strains of bacteria pose a dangerous threat'. There are options for '3D View: 4eyl' with settings for Style (Cartoon, Spheres, Trace) and Color (Rainbow, Chain, Structure). A 'Spin' control is also present.
- Health Focus: Human Immunodeficiency Virus (HIV):** A section highlighting a study of HIV at the molecular level, providing insight into how the virus works and can be stopped. It includes links to explore features on the HIV capsid, related enzymes, and other proteins, as well as an interactive animation and classroom materials.
- 2015 Video Challenge for High School Students:** A section titled 'Detecting & Combating HIV in 3D'. It invites high school students to create short videos that tell the molecular story of health and disease. It includes links for 'Introduction, Dates, & Awards', 'Rules', 'Resources', and 'Participation Guide'. Below this, there are thumbnails for '2015 Winners'.
- News and Events:** A section listing recent events, including 'SACNAS 27 October, 2015' (Society for Advancement of Chicanos/Hispanics and Native Americans in Science National Conference) and 'NJ Science Convention Oct 13-14 30 September, 2015'. It also mentions that RCSB PDB will be connecting and sharing ideas with area science teachers at the NJ Science Convention.
- Learn: Materials for Exploration:** A section titled 'Take an Interactive Tour of the Protein Data Bank'. It describes exploring the beauty and diversity of 3D PDB molecular shapes and sizes with a unique viewer. It includes links for 'Paper Models', 'Posters', 'Videos', and 'Animations'.
- Teach: Curricula:** A section titled 'Biomolecular Structures and Models'. It states that the RCSB PDB Curricula provide authentic, hands-on teaching materials, individual and group activities and assessment suggestions. It includes a link for 'All Curricula'.
- Education Corner: Using PDB Resources:** A section titled 'Issue 66 - July 2015 High School Video Challenge Results'. It includes a thumbnail for the 'HIV' video challenge results and a link for 'All Issues'.



Browser

Browser accessed through home page (shown) or top menu

Each subcategory provides access to all related *Molecule of the Month* columns and Educational Materials

Browse resources by category

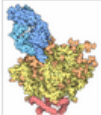
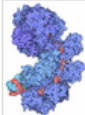
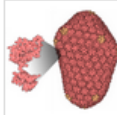
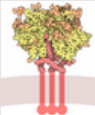
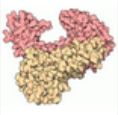
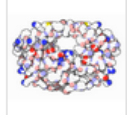
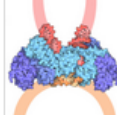
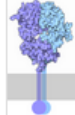
- ▲ Health and Disease
 - You and Your Health >
 - Immune System >
 - HIV and AIDS >
 - Diabetes >
 - Cancer >
 - Viruses >
 - Toxins and Poisons >
 - Drug Action >
 - Drug Resistance >
- ▼ Molecules of Life
- ▼ Biotech and Nanotech
- ▼ Structures and Structure Determination

HIV and AIDS





the virus and how we fight it

The AIDS epidemic began a few decades ago and since then scientists have revealed the atomic structures of most of the molecules in HIV. Using these structures, researchers have designed new treatments for HIV infection, which are now part of effective drug regimens that halt the growth of the virus. Structures of HIV molecules also provide new hope for development of vaccines and methods to rid an infected individual of the latent virus.

▲ Related Molecule of the Month Articles

 Broadly Neutralizing Antibodies	 Cascade and CRISPR	 HIV Capsid	 HIV Envelope Glycoprotein
 HIV Reverse Transcriptase	 HIV-1 Protease	 Integrase	 T-Cell Receptor

▲ Learning Resources

 HIV Capsid PDF ↗	 HIV Capsid Video	 The Structural Biology of HIV Poster, high res. ↗	 The Structural Biology of HIV Poster, low res. ↗
---	---	--	---



Molecule of the Month

HIV Capsid

At the center of HIV, an unusual cone-shaped capsid protects the viral genome and delivers it into infected cells

Viruses
or DNA
complex
and ce
packag
to the e

Fighting Back

TRIM5 is one of the weapons deployed in the ongoing battle being fought between living organisms and viruses. TRIM5 binds to the retroviral capsid and interferes with the

Stayi

The HI
from P
domain
assem
six, an
to form

Break

The fie
capsid
cone-s
over a
electro
with m
micros
and a s
large s
read m

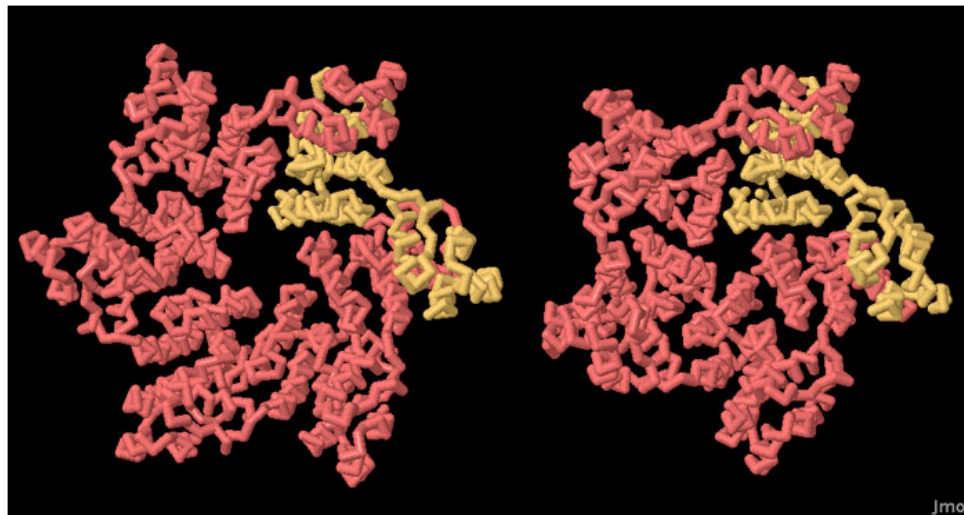
Cellular prot

Download hig

Exploring the Structure

Image

Jmol 1



The capsid protein is able to form hexamers and pentamers by shifting slightly in structure. This is an example of the principle of "quasiequivalence", first proposed by Caspar and Klug in 1962. Quasiequivalence is the way that many viruses build capsids that are much larger than is possible with perfect symmetry, but that still only use a single type of protein chain. In the HIV capsid, the interactions between the many subunits are similar, but are deformed slightly to accommodate the different shapes of the cone-shaped portion and the round caps. To take a closer look at these two structures, PDB entries [3mge](#) and [3p05](#), click on the image for an interactive Jmol.



Learn: Educational Resources

Learn

Paper Models >

Posters, Flyers & Calendars >

Videos >

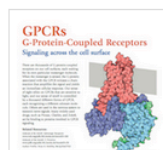
Interactive Animations >

Education Corner >

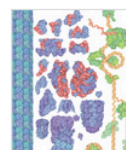
Guide to Understanding PDB Data

Posters, Flyers & Calendars

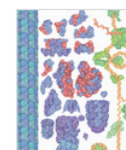
Grid List



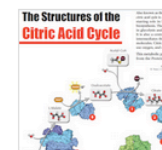
GPCRs
G-Protein-Coupled Receptors
Signaling across the cell surface
Flyer [↗](#)



Molecular Machinery: A Tour of
the Protein Data Bank (2014)
Poster [↗](#)



Molecular Machinery: A Tour of
the Protein Data Bank (2014)
Flyer [↗](#)



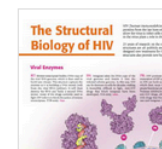
The Structures of the
Citric Acid Cycle
Flyer [↗](#)



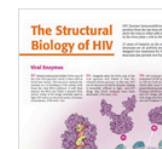
Virus Structures
Flyer [↗](#)



What is a Protein?
Flyer [↗](#)



The Structural Biology of HIV
Poster, high res. [↗](#)



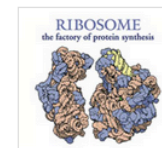
The Structural Biology of HIV
Poster, low res. [↗](#)



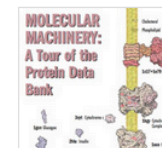
How Do Drugs Work?
Poster, high res. [↗](#)



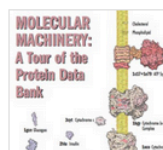
How Do Drugs Work?
Poster, low res. [↗](#)



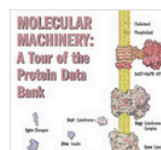
The Ribosome
Flyer [↗](#)



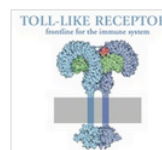
Molecular Machinery: A Tour of
the Protein Data Bank (2002)
Poster, high res. [↗](#)



Molecular Machinery: A Tour of
the Protein Data Bank (2002)
Poster, low res. [↗](#)



Molecular Machinery: A Tour of
the Protein Data Bank (2002)
Flyer [↗](#)



Toll-like Receptors
Flyer [↗](#)



2015 Molecule of the Month
Calendar [↗](#)

Teach

How To	
Read a Scientific Paper	Guide  Quick review 
Make sense of the Information Out There	Guide and exercise  Teaching Notes  Additional Reading 
Explore a Topic of Interest at a Molecular Level	Guide  Quick Review  Worksheet  Learning about Asthma 
Find Structures Relevant to a Topic of Interest	Guide  Quick Review  Selected structures related to Asthma 
Visualize, Analyze and Compare Selected Structures	Cheat Sheet  Basics  Menus  Selections  Structure Analysis  Structure Comparisons 
Make Publication Quality Images and Movies	Guide  Part1  Part2 

“How to” skills

Organization of all relevant materials



Future

- Searching of Educational Resources enabled November 2015
- Internal and external beta testing November 2015
- Initial deployment December 2015
- 2016 Evaluation
 - User surveys
 - Feedback from students and teachers
 - Focus groups
 - Analytics (usage and navigation)
- Continued infrastructure development and improvement
 - Mobile responsiveness
 - Improved search
 - Authoring tools
- Ongoing content development leveraged with Education efforts

