
RCSB Protein Data Bank Advisory Committee Meeting

September 14, 2012



Overview

Helen Berman



When Last We Met...



257 participants
35 travel awards
93 posters



PDB40 Speakers

TRIBUNA

40 años reuniendo proteínas

El banco de datos internacional, que almacena 77.000 estructuras moleculares, celebra su origen en Cold Spring Harbor

GRÁFICO: Tabla de contenidos

CELEBRADO ZAPATERO | 20 DIC 2011 - 13:19 CET

Archivado en: Nueva York, Biología molecular, Estados Unidos, Laboratorios, Aniversario, Centros de investigación, Eventos, Investigación científica, Biología, Ciencias naturales, Ciencia, Sociedad

El cronista: Michael G. Rossmann (en el centro) 40 años después de iniciar el Banco de Datos de Proteínas, celebra el aniversario en Cold Spring Harbor. / CHRISTIANE BARRAZZINI



RICHARD L. MCCORMICK
President, Rutgers, The State University of New Jersey

KENNETH BRESLAUER
Dean of Life Sciences,
Vice President for Health Science Partnerships

**TUESDAY, DECEMBER 6, 2011
11:00 a.m.**

174 Frelinghuysen Road
Busch Campus
Piscataway, New Jersey

Buffet lunch to follow

RSVP BY DECEMBER 1
to Stephanie Kenyon at 732-932-7396
or eventsvp@winants.rutgers.edu

Invite you to the dedication of the
**CENTER FOR INTEGRATIVE
PROTEOMICS RESEARCH**

The \$47 million, 75,000-square-foot building will integrate academic and research activity related to proteomics - the study of the structures, functions, and interactions of proteins - at Rutgers. This research can provide a fuller understanding of the molecular basis of health and disease, including drug development and discovery.

The Center for Integrative Proteomics Research (CIPR) will house programs and resources key to worldwide biomedical research, most notably the Protein Data Bank (PDB), repository of information about the 3D structures of biological molecules. As

part of the overarching Proteomics Center, CIPR will also house the BioMaPS Institute for Quantitative Biology that promotes research and education at the interface of biology, mathematics, and physical biology, High Field Bio-NMR Programs, and High Performance Computing. The Center will host core facilities for nuclear magnetic resonance mass spectrometry, x-ray crystallography, and cryo-electron microscopy, and state-of-the-art facilities for computational biology. Center laboratories focusing on various aspects of proteomics research will be sited within the building.





Worldwide Protein Data Bank Foundation

- Established to support specific wwPDB activities
 - Advisory committee meetings
 - Outreach and education activities, including seminars and workshops
 - October 2012: Symposium in Osaka, Japan
- 501(c)3 organization
 - American, tax-exempt association dedicated to scientific, literary, charitable, and educational purposes
- Fundraising on-going

13:30-14:00
中村泰伸 (大阪大学・蛋白質研究所)
「生命科学における蛋白質構造データベースの役割」

14:00-15:00
Stephen Kevin Bailey (カリフォルニア大学サンディエゴ校)
「創薬へのPDBデータベースのインパクト」 (日本語版提供)

15:00-15:20 休憩

15:20-16:20
藤田博一 (大阪大学・生命情報研究所)
「生命を支える生体分子機械 - ナノテクノロジーによる分子認識の最新動向」

16:20-16:30 質疑応答、閉会

一般講演会
**PDB データバンク：
生命科学・創薬の情報基盤**

2012年10月13日(土)
13:30-16:30 (13:00 受付開始)

会場 ハートンホール毎日新聞ビル B 1
〒530-0001 大阪市北区梅田 3-4-5

主催 wwPDB Foundation
(国際蛋白質構造データベース財団)

協賛 科学技術振興機構
バイオサイエンスデータベースセンター
大阪大学蛋白質研究所
大阪大学生命情報研究科
日本生物物理学会
日本蛋白質科学会
大阪商薬品協会

WORLDWIDE
PDB
PROTEIN DATA BANK

Organizational Updates

- New wwPDB Advisory Committee Chair
 - Soichi Wakatsuki (Institute of Materials Structure Science-KEK, Japan)
- New RCSB PDB Advisory Committee Chair
 - Cynthia Wolberger (Johns Hopkins)

Response to Major 2011 Recommendations

- Continue ongoing efforts to better represent ligands
 - New tools and dictionaries developed for complex ligands
- Estimate impact of D&A on annotation speed and efficiency to help project load balancing requirements
 - Initial load balancing outline drafted
 - Impact testing to follow previous efforts
- Improve mobile device access to outreach and education resources
 - RCSB PDB *Mobile*

Funding: Strategy for Sustainability?

- RCSB PDB competitive renewal funded by NSF
 - January 2009 - December 2013
 - Renewal due 2013
- PDBe competitive grant from Wellcome Trust
 - January 2010 - December 2014
- PDBj competitive renewal funded by JST (Japan Science & Technology Agency)
 - April 2011 - March 2014
- BMRB competitive renewal funded from the National Library of Medicine
 - NLM will no longer fund BMRB after 2014

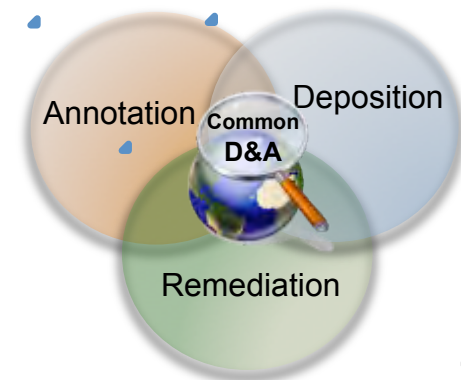
Vision

To provide a global resource for the advancement of research and education in biology and medicine by curating, integrating, and disseminating biological macromolecular structural information in the context of function, biological processes, evolution, pathways and disease states.

We will implement standards, and anticipate and develop appropriate technologies to support evolving science.

Data In

- Improved tools for deposition
- Improved data processing efficiency
- Remediation review
- Biologically Interesting molecule Reference Dictionary (BIRD)
- New and improved Common Tool modules
- wwPDB Task Forces



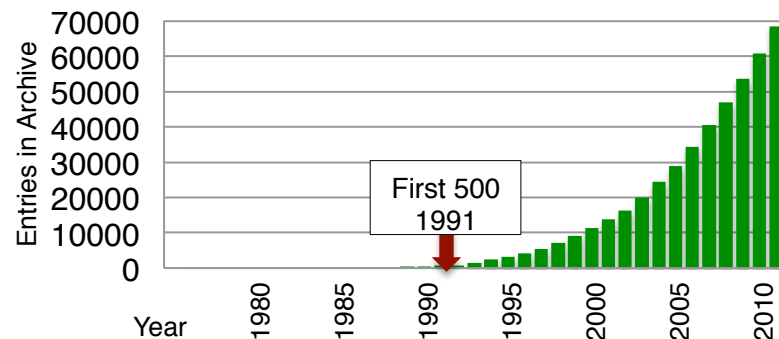
Depositions

Year	Total Depositions	Deposited To			Processed By		
		RCSB PDB	PDBj	PDBe	RCSB PDB	PDBj	PDBe
2000	2983	2445	10	528	2297	158	528
2001	3287	2673	118	496	2408	383	496
2002	3565	2769	289	507	2401	657	507
2003	4830	3488	673	669	3135	1026	669
2004	5508	3796	900	812	3082	1614	812
2005	6678	4507	1166	1005	3563	2110	1005
2006	7282	5145	1052	1085	4252	1945	1085
2007	8130	5399	1603	1128	4703	2299	1128
2008	7073	5452	648	973	4106	1994	973
2009	8300	6715	527	1058	5069	2173	1058
2010	8878	6912	593	1373	5464	2041	1373
2011	9250	7172	582	1496	5938	1816	1496
2012	6728	5273	344	1111	4522	1094	1111
TOTAL	82492	61746	8505	12241	50940	19310	12241

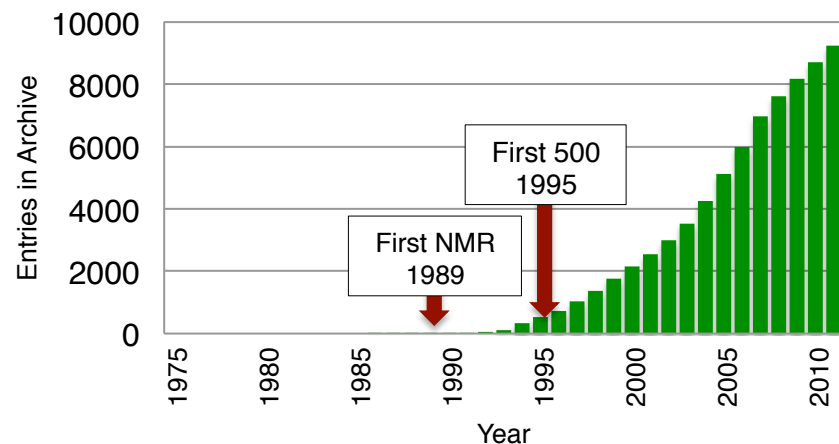
Last Updated: 5 Sep 2012

9965 depositions predicted for 2012

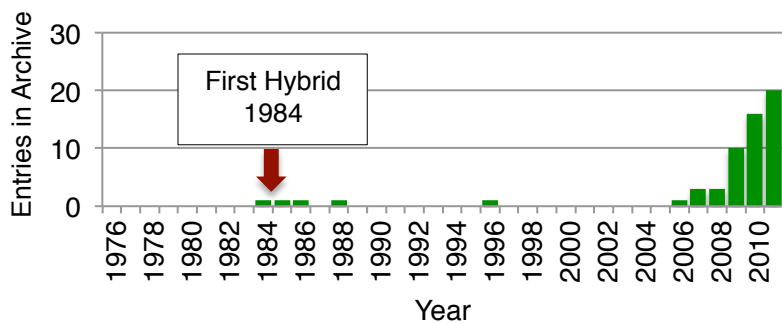
X-ray entries



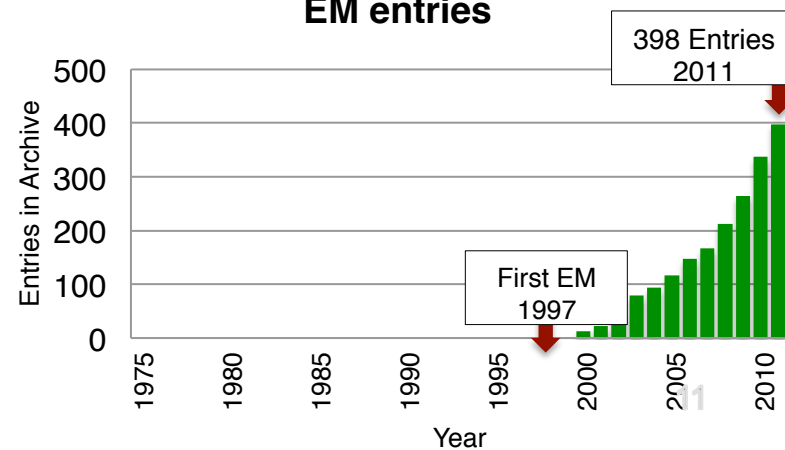
NMR entries



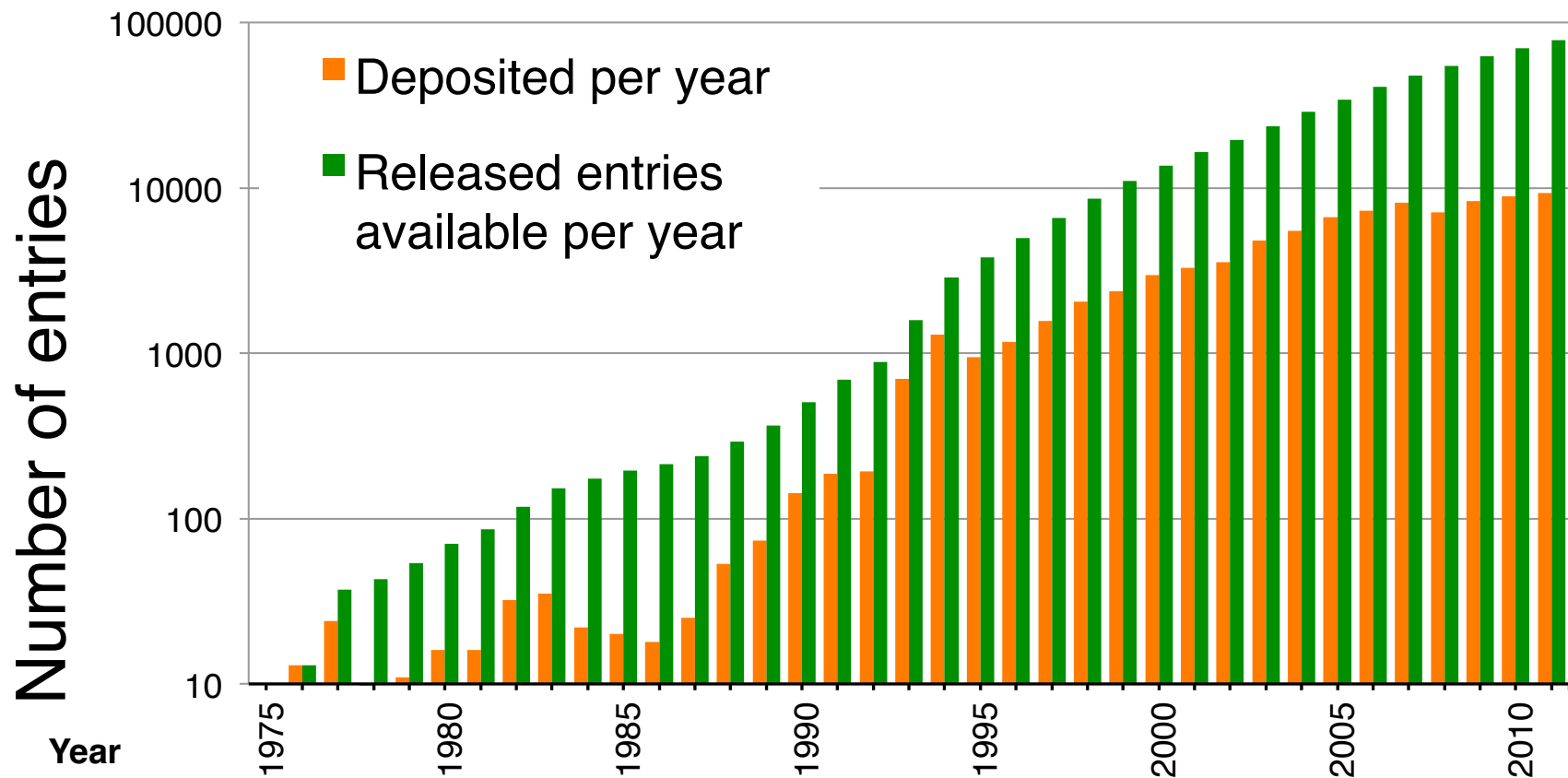
Hybrid Entries



EM entries



Entries per year (Logarithmic Scale)



Common Deposition and Annotation Tool

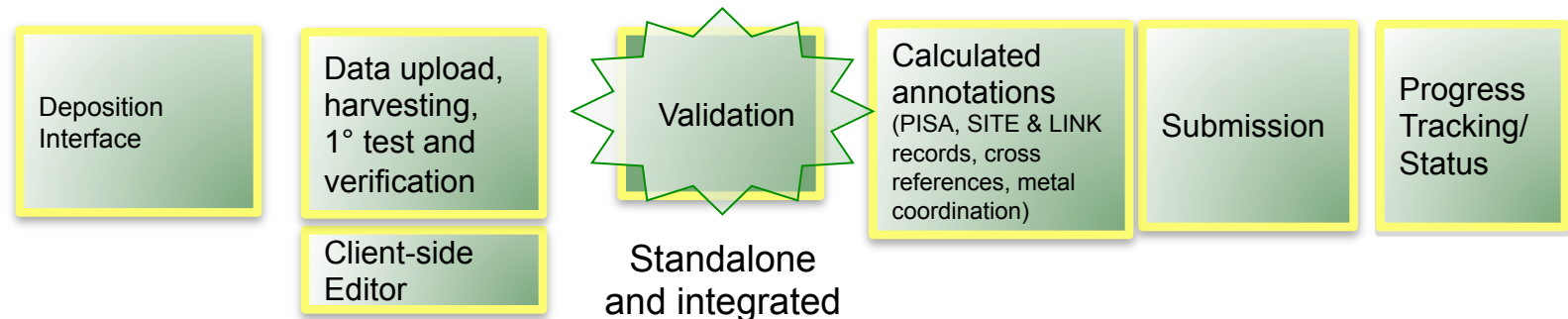


The goal is to implement a set of common deposition and annotation processes and tools that will enable the wwPDB to deliver a resource of increasingly high quality and dependability over the next 10 years.

- Addresses the increase in complexity and experimental variety of submissions and the increase in deposition throughput
- Maximizes the efficiency and effectiveness of data handling and support for the scientific community

Common Deposition and Annotation Pipeline

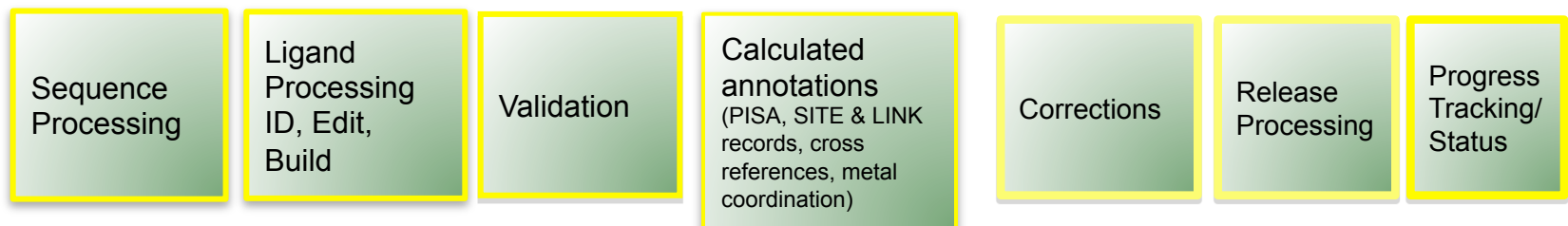
Deposition Pipeline



Communication System

Workflow-Automation System

Annotation Pipeline



Task Forces

To collect recommendations and develop consensus on method-specific issues, including validation checks that should be performed and identification of validation software applications.

X-ray Validation

- 2008 Workshop
- 2011 *Structure* publication
- Chair: Randy J. Read (University of Cambridge)

3DEM Validation

- 2010 Meeting
- Chairs: Richard Henderson (Maps, MRC-LMB), Andrej Sali (Models, UCSF)
- 2012 *Structure* publication

NMR Validation

- 2009, 2011 meetings
- Chairs: Gaetano Montelione (Rutgers), Michael Nilges (Institut Pasteur)
- Report in progress

Small-Angle Scattering

- July 2012 Meeting
- Chair: Jill Trewhella (University of Sydney)
- Report in progress



3DEM VTF Recommendations

Maps

- Develop standards for assessing map resolution and reliability
- Report map resolution in accordance with visible features
- Collect annotations specific to each map type
- Validate map symmetry

Map-derived models

- Assess models using X-ray VTF-recommended criteria
- Develop criteria for assessing model fit to map
- Annotate sequences of all map components
- Capability to archive coarse-grained model representations

Additional recommendations

- Establish benchmarks for methods development
- Establish deposition guidelines for publication of 3DEM structures
- Expand EMDataBank's role to facilitate development of validation and data standards

Outcome of the First Electron Microscopy Validation Task Force Meeting

Richard Henderson,¹ Andrej Sali,² Matthew L. Baker,³ Bridget Carragher,⁴ Bebal Devkota,⁵ Kenneth H. Downing,⁶ Edward H. Egelman,⁷ Zhiqiang Feng,⁸ Joachim Frank,⁹ Melissa Grigorieff,¹⁰ Wei Jiang,¹¹ Steven J. Ludtke,¹² Oded Medalia,¹³ P. Penczek,¹⁴ Peter B. Rosenthal,¹⁵ M. G. Rossmann,¹⁶ Stephen F. Subramanian,¹⁷ Doreen P. Suck,¹⁸ Alexander C. Steven,¹⁹ David L. Stokes,²⁰ John D. Westbrook,²¹ Wally Wriggers,²² Huaming Yang,²³ Lawrence Young,²⁴ Klaus M. Berman,²⁵ W. Chiu,²⁶ David J. Kleywegt,²⁷ and Catherine L. Lawson²⁸

¹MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 0Q, UK
²Department of Bioengineering and Materials Science, Department of Pharmaceutical Chemistry, California Institute for Quantitative Biosciences, Room 308B, University of California at San Francisco, 1700 4th Street, San Francisco, CA 94158, USA
³National Center for Microstructural Imaging, Xiang and Mann Miller, Department of Biochemistry and Molecular Biology, Baylor College of Medicine, Houston, TX 77030, USA
⁴National Institute for Advanced Biomedical Research, Department of Cell Biology, The Scripps Research Institute, La Jolla, CA 92037, USA
⁵Department of Chemistry and Chemical Biology and Howard Crosby Laboratory for Structural Bioinformatics, Rutgers, The State University of New Jersey, 810 Taylor Road, Piscataway, NJ 08854, USA
⁶Life Sciences Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Drive, Berkeley, CA 94720, USA
⁷Department of Biology, University of California at San Diego, 1001 Shiley Avenue, San Diego, CA 92161, USA
⁸Department of Chemistry, University of California at Berkeley, 1160 Chemistry Building, Berkeley, CA 94720, USA
⁹Department of Biology, University of California at Berkeley, 1160 Chemistry Building, Berkeley, CA 94720, USA
¹⁰Department of Biology, University of California at Berkeley, 1160 Chemistry Building, Berkeley, CA 94720, USA
¹¹Department of Chemistry, University of California at Berkeley, 1160 Chemistry Building, Berkeley, CA 94720, USA
¹²Department of Biology, University of California at Berkeley, 1160 Chemistry Building, Berkeley, CA 94720, USA
¹³Department of Chemistry, University of California at Berkeley, 1160 Chemistry Building, Berkeley, CA 94720, USA
¹⁴Department of Chemistry, University of California at Berkeley, 1160 Chemistry Building, Berkeley, CA 94720, USA
¹⁵Department of Chemistry, University of California at Berkeley, 1160 Chemistry Building, Berkeley, CA 94720, USA
¹⁶Department of Chemistry, University of California at Berkeley, 1160 Chemistry Building, Berkeley, CA 94720, USA
¹⁷Department of Chemistry, University of California at Berkeley, 1160 Chemistry Building, Berkeley, CA 94720, USA
¹⁸Department of Chemistry, University of California at Berkeley, 1160 Chemistry Building, Berkeley, CA 94720, USA
¹⁹Department of Chemistry, University of California at Berkeley, 1160 Chemistry Building, Berkeley, CA 94720, USA
²⁰Department of Chemistry, University of California at Berkeley, 1160 Chemistry Building, Berkeley, CA 94720, USA
²¹Department of Chemistry, University of California at Berkeley, 1160 Chemistry Building, Berkeley, CA 94720, USA
²²Department of Chemistry, University of California at Berkeley, 1160 Chemistry Building, Berkeley, CA 94720, USA
²³Department of Chemistry, University of California at Berkeley, 1160 Chemistry Building, Berkeley, CA 94720, USA
²⁴Department of Chemistry, University of California at Berkeley, 1160 Chemistry Building, Berkeley, CA 94720, USA
²⁵Department of Chemistry, University of California at Berkeley, 1160 Chemistry Building, Berkeley, CA 94720, USA
²⁶Department of Chemistry, University of California at Berkeley, 1160 Chemistry Building, Berkeley, CA 94720, USA
²⁷Department of Chemistry, University of California at Berkeley, 1160 Chemistry Building, Berkeley, CA 94720, USA
²⁸Department of Chemistry, University of California at Berkeley, 1160 Chemistry Building, Berkeley, CA 94720, USA

(2012) *Structure* 20: 205-214
 R. Henderson, A. Sali, M.L. Baker, B. Carragher, B. Devkota, K.H. Downing, E.H. Egelman, Z. Feng, J. Frank, N. Grigorieff, W. Jiang, S.J. Ludtke, O. Medalia, P.A. Penczek, P.B. Rosenthal, M.G. Rossmann, M.F. Schmid, G.F. Schröder, A.C. Steven, D.L. Stokes, J.D. Westbrook, W. Wriggers, H. Yang, J. Young, H.M. Berman, W. Chiu, G.J. Kleywegt, C.L. Lawson

This Meeting Report is available at www.jstor.org/stable/2345678. For more information on the work of the EMDataBank, visit www.emdatabank.org.
 © 2012 EMDataBank. All rights reserved.

Structure 20, February 8, 2012 0010-0010 Elsevier Ltd. All rights reserved. 205

Data Out

- Improved searching and browsing of results
- Visualization of molecular surfaces
- Improved sequence annotations
- Domain-based structural alignments
- New Web Services

OC(=O)CCCC[C@@H]1SC[C@@H]2NC(=O)N[C@@H]12

SMILES

- Has exact structure OC(=O)CCCC[C@@H]1SC[C@@H]2NC(=O)N[C@@H]12
- Has sub-structure OC(=O)CCCC[C@@H]1SC[C@@H]1
- Is very similar (95%) with OC(=O)CCCC[C@@H]1
- Is similar (70%) with OC(=O)CCCC[C@@H]1SC[C@@H]1
- Super-structure of OC(=O)CCCC[C@@H]1SC[C@@H]1

Showing 1 - 25 of 646 Results


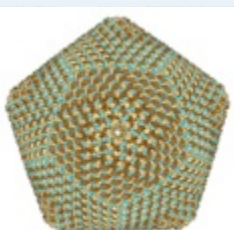
Filters: Custom | View: Condensed | Reports: Select one... | Sort: Relevance

PDB ID	Title	Macromolecule Name
1NS0	Structural consequences of a cancer-causing BRCA1-BRCT missense mutation	BREAST CANCER TYPE 1 SUSCEPTIBILITY...
1JNX	Crystal structure of the BRCT repeat region from the breast cancer associated protein...	BREAST CANCER TYPE 1 SUSCEPTIBILITY...
2KW6	Solution NMR Structure of Cyclin-dependent kinase 2-associated protein 1 (CDK2-associated)	Cyclin-dependent kinase 2-associat...
3HG1	Gemitter-governed recognition of a cancer epitope by an intratumorant human T cell	MHC class I antigen, Beta-2-microg...
2HB0	NMR structures of SAR domain of Deleted in Liver Cancer 2 (DLC2)	SAR-related lipid transfer protein...
2JW2	Validation of inter-helical orientation of the sixth alpha-motif of human deleted in liver	SAR-related lipid transfer protein...
2T38	Solution Structure of a Human Cancer-Related Nucleoside Triphosphatase	Human Cancer-Related NTPase
1T15	Crystal Structure of the Breast BRCT Domains in Complex with the Phosphorylated	BREAST CANCER TYPE 1 SUSCEPTIBILITY...

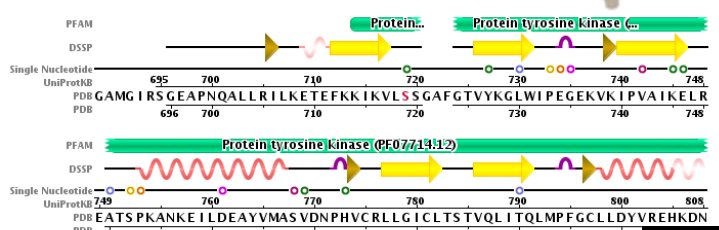
Smart suggestions

Condensed query results view (Fall 2012)

Protein Workshop

Structure Summary: Sequence Tab




Structure Summary: 3D Similarity Tab

Entity #1: Chains: A

Description: Cyclodextrin glucanotransferase protein

Length: 683



UniProtKB
SEQRES
ATOM
DSSP
SCOP
PDP
Pfam

Results for domain d3bmva4

SCOP Domain ID d3bmva4 (chain 1) vs. representatives of other sequence clusters (chain 2)

Rank	Results	Domain 2	Title	P-value	Score	Rmsd	Len1	Len2	%ID	%Cov1	%Cov2
1	view	d2guya2	Alpha-amylose A2guy A:1-...	1.11E-16	788.51	2.61	406	381	22	91	97
2	view	PDP:4AEEJ	ALPHA AMYLASE, CATALYT	1.33E-15	643.75	2.01	406	309	20	74	97
3	view	d1j0ha3	neopullulanase 1j0h A:124-	1.64E-14	766.83	3.21	406	382	22	89	94

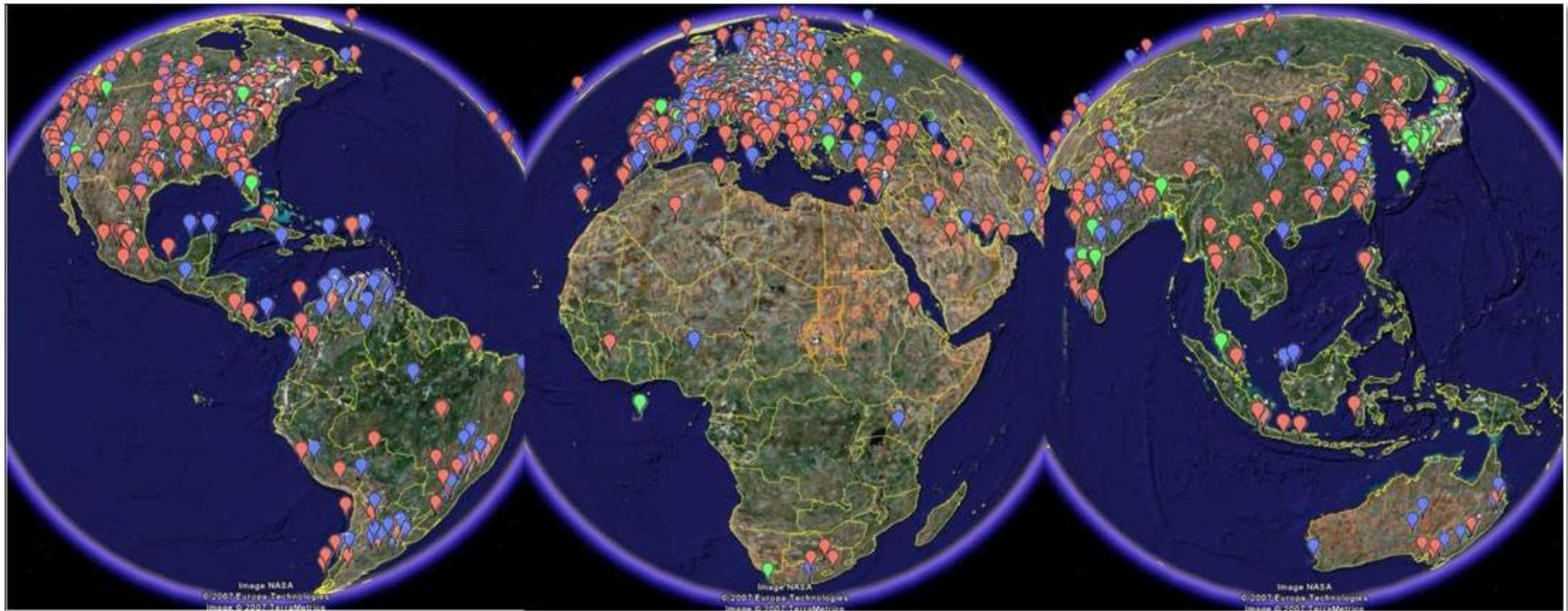
RCSB PDB *Mobile*



- Search the PDB, view the latest structures, access your MyPDB account, view *Molecule of the Month* articles, and more
- Apple iOS platform (iPad/iPhone/iPod)
- Android version in development



2011 FTP & Rsync Data Downloads



RCSB PDB
2011: 282 million
2010: 159 million

PDBe
2011: 59 million
2010: 34 million

PDBj
2011: 38 million
2010: 16 million

Outreach

Educational Communities

- PDB-101 packages together RCSB PDB resources of interest to teachers and students
- Meetings and events
- Molecular Anatomy Project



Virus Structures
handout

Research Communities

- Task Force Meetings
- Professional society meetings
- Publications
- Online resources
- PDB40

Structure
Meeting Review



The Protein Data Bank at 40: Reflecting on the Past to Prepare for the Future

Helen M. Berman,^{1,*} Gerard J. Kleywegt,² Haruki Nakamura,³ and John L. Murray⁴
¹RCSB PDB, Department of Chemistry and Chemical Biology, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA
²Protein Data Bank in Europe, European Bioinformatics Institute, Hinxton, Cambridge CB10 1SD, United Kingdom
³Protein Data Bank Japan, Institute for Protein Research, Osaka University, 3-2 Yamadaoka, Suita, Osaka, 565-0871, Japan
⁴Stoughton/Flare, University of Wisconsin-Madison, Madison, WI 53706, USA
*Correspondence: berman@rcsb.rutgers.edu
DOI: 10.1016/j.str.2012.01.010

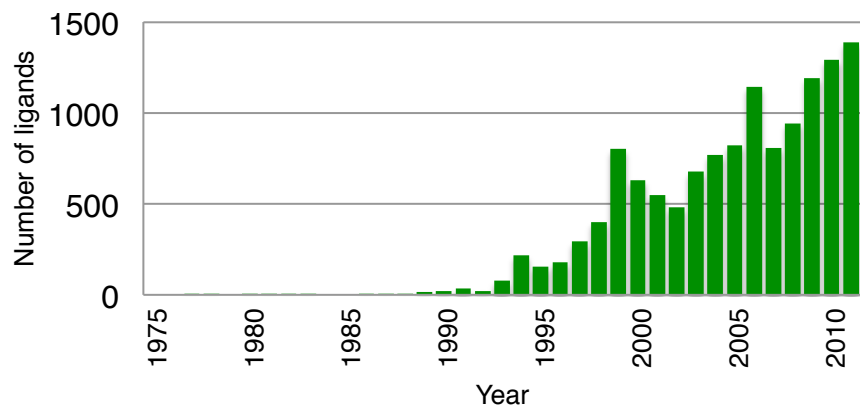
A symposium celebrating the 40th anniversary of the Protein Data Bank archive (PDB), organized by the Worldwide Protein Data Bank, was held at Cold Spring Harbor Laboratory (CSHL) October 28-30, 2011. PDB40's distinguished speakers highlighted four decades of innovation in structural biology, from the early days of structural determination to future directions for the field.



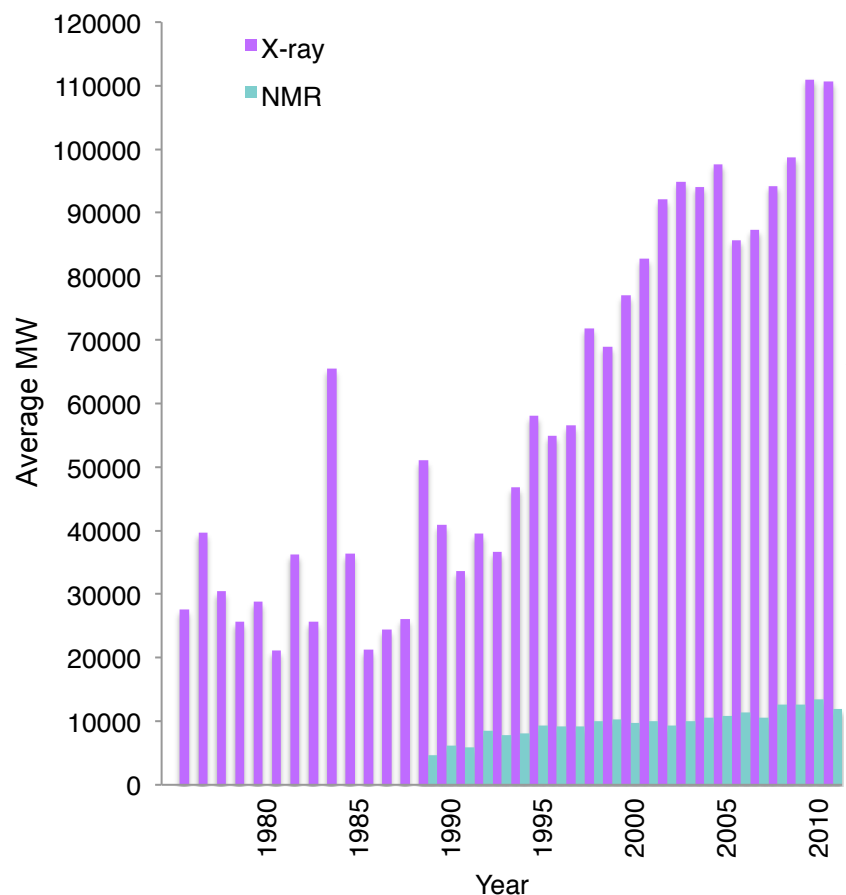
PDB past and present at PDB40

Increasing Complexity

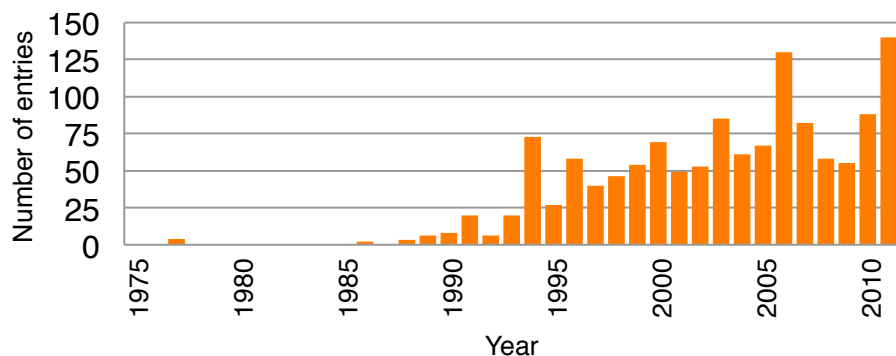
Number of ligands released per year



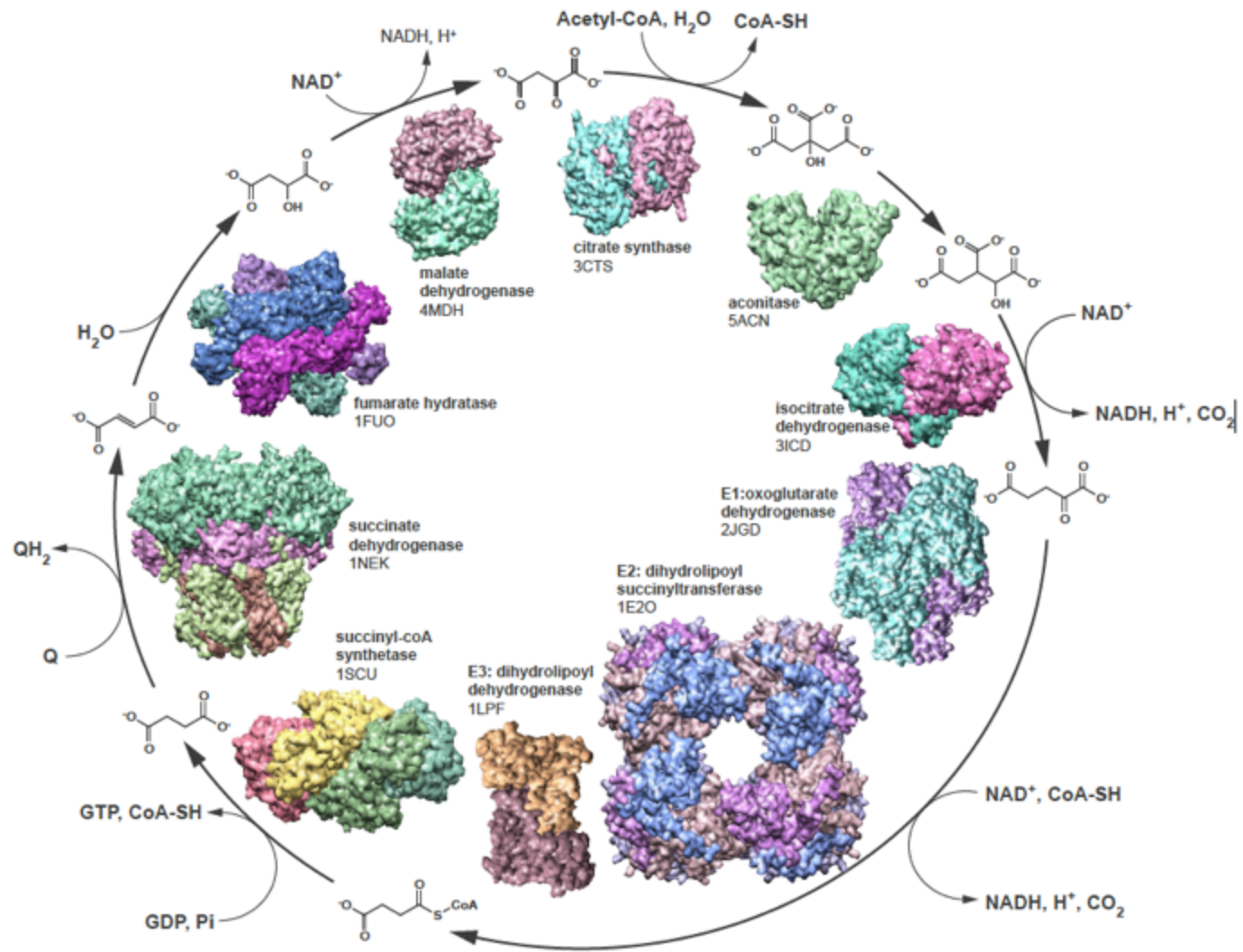
Average Molecular Weight for Entries Released per Year



Number of entries with peptide-like inhibitors/antibiotics released per year



The Krebs Cycle



Deposition, Annotation and Remediation

Jasmine Young

Annotation Team



Data In

Goal: Enable research and discovery in structural biology and biomedical research through

- Capturing experimental data that defines structures of macromolecules
- Maximizing the quality and completeness of data
- Representing data consistently

2012: RCSB PDB Annotating 69% of all Depositions (3% Increase from 2011)

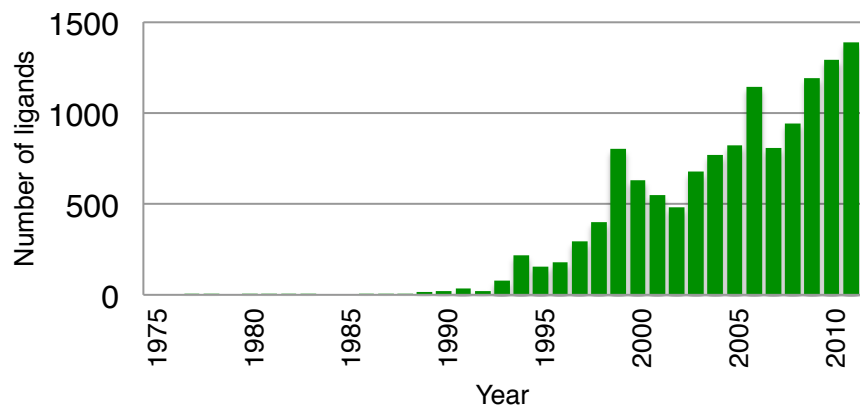
Year	Total Depositions	Deposited To			Processed By		
		RCSB PDB	PDBj	PDBe	RCSB PDB	PDBj	PDBe
2000	2983	2445	10	528	2297	158	528
2001	3287	2673	118	496	2408	383	496
2002	3565	2769	289	507	2401	657	507
2003	4830	3488	673	669	3135	1026	669
2004	5508	3796	900	812	3082	1614	812
2005	6678	4507	1166	1005	3563	2110	1005
2006	7282	5145	1052	1085	4252	1945	1085
2007	8130	5399	1603	1128	4703	2299	1128
2008	7073	5452	648	973	4106	1994	973
2009	8300	6715	527	1058	5069	2173	1058
2010	8878	6912	593	1373	5464	2041	1373
2011	9250	7172	582	1496	5938	1816	1496
2012	6728*	5273	344	1111	4522	1094	1111
TOTAL	82492	61746	8505	12241	50940	19310	12241

Last Updated: 5 Sep 2012

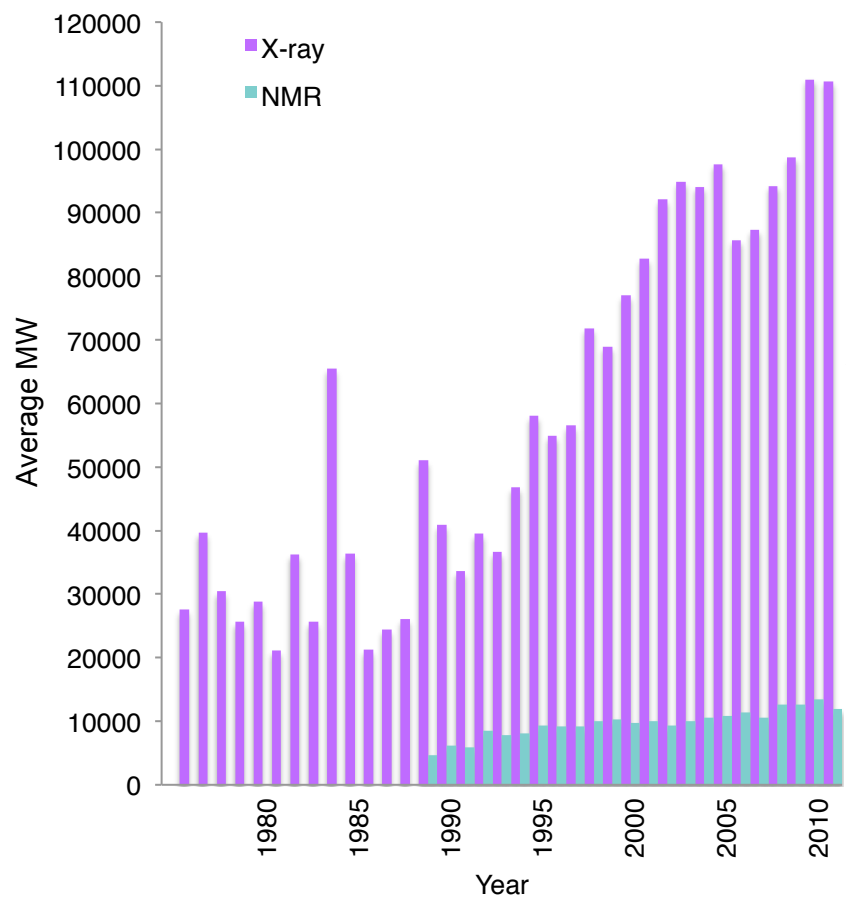
***9956 depositions projected for 2012**

Increasing Complexity

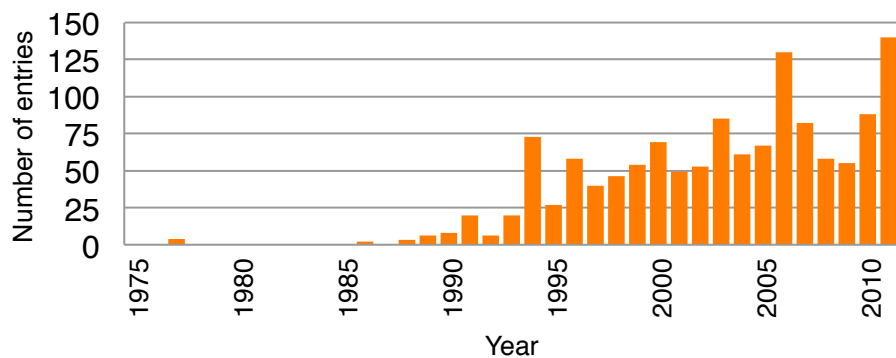
Number of ligands released per year



Average Molecular Weight for Entries Released per Year



Number of entries with peptide-like inhibitors/antibiotics released per year



Data In Systems

Current Deposition

- ADIT: deposition tool
- *pdb_extract*: data harvesting tool
- Validation Server
- *SF-Tool*: converts and validates structure factor data
- Ligand Expo: search and browse released ligands

Current Annotation

- Annotation pipeline integrated with *D&A Ligand Module*
- *Chemical component dictionary* and searching tools
- Infrastructure for small molecules & peptides
 - *Biologically Interesting molecule Reference Dictionary (BIRD)* and searching tools
 - Annotation system

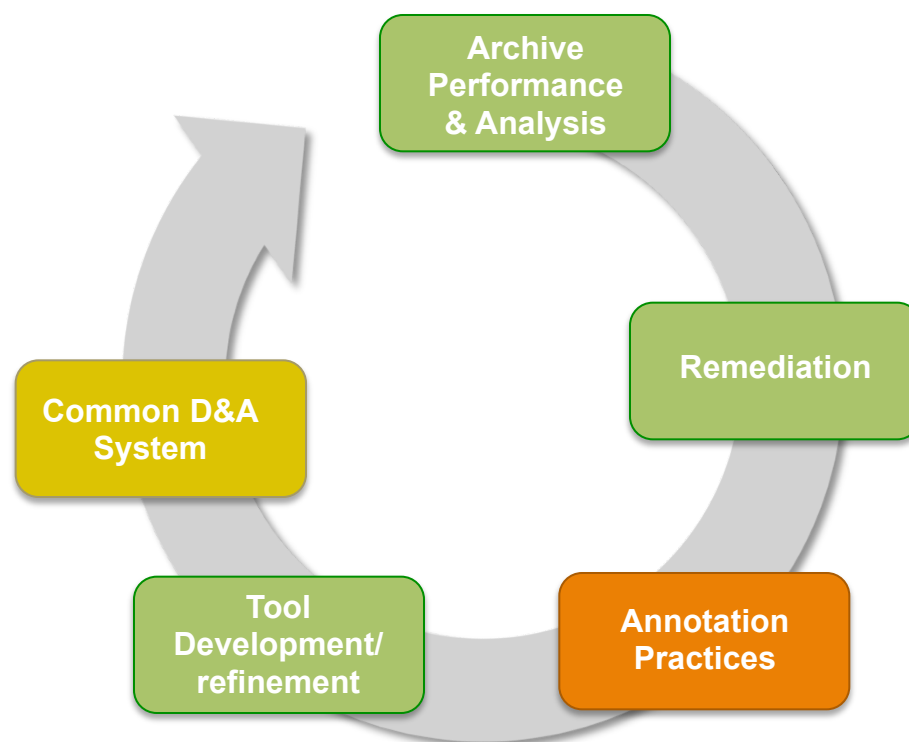
NEAR future: wwPDB Common D&A System

Remediation

- Informs all processes
- Improves consistency in entry and archive annotation
- Enhances chemistry representation



Better query capability



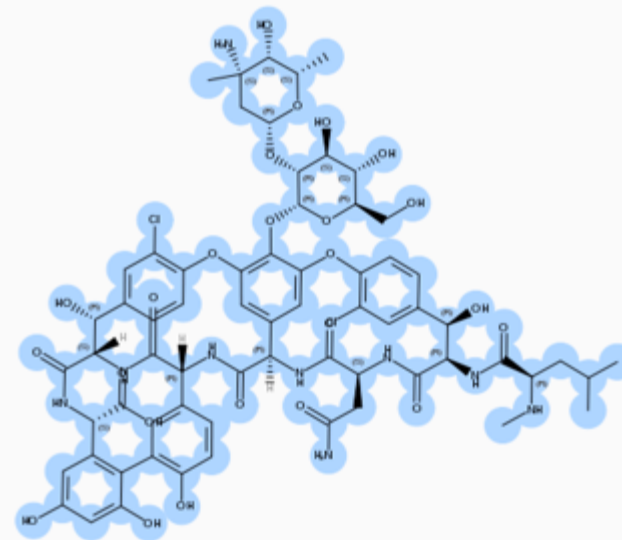
Last remediation release: July 2011

Better Annotation of Biologically Interesting Molecules

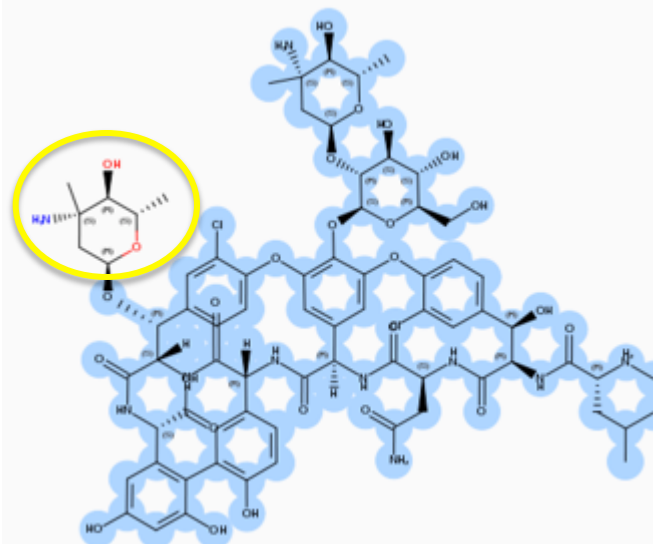
2011 remediation of inhibitors and antibiotics informed the development of an annotation system that supports

- Searches of small molecules and peptides against the new *Biologically Interesting molecule Reference Dictionary* (BIRD)
- 2D and 3D views
- Comparative analysis of structures
- Building new BIRD definitions
- Use of existing templates to maintain consistency in the data presentation

Target



Hit



Future Remediation 2012 - 2014

- **Carbohydrates (RCSB PDB)**
 - Data analysis completed
- **Protein modifications (PDBe)**
 - Data analysis completed
- Metal-containing ligands
- Translation of non-standard crystal frame
- Recalculation of B factors
- Translation of dissociated assemblies
- X-ray multiple models

Carbohydrate Remediation (RCSB PDB)

Issues

- Multiple representations in naming and linking
- Non-standard nomenclature and incomplete linkages
- Representation of branched polymers

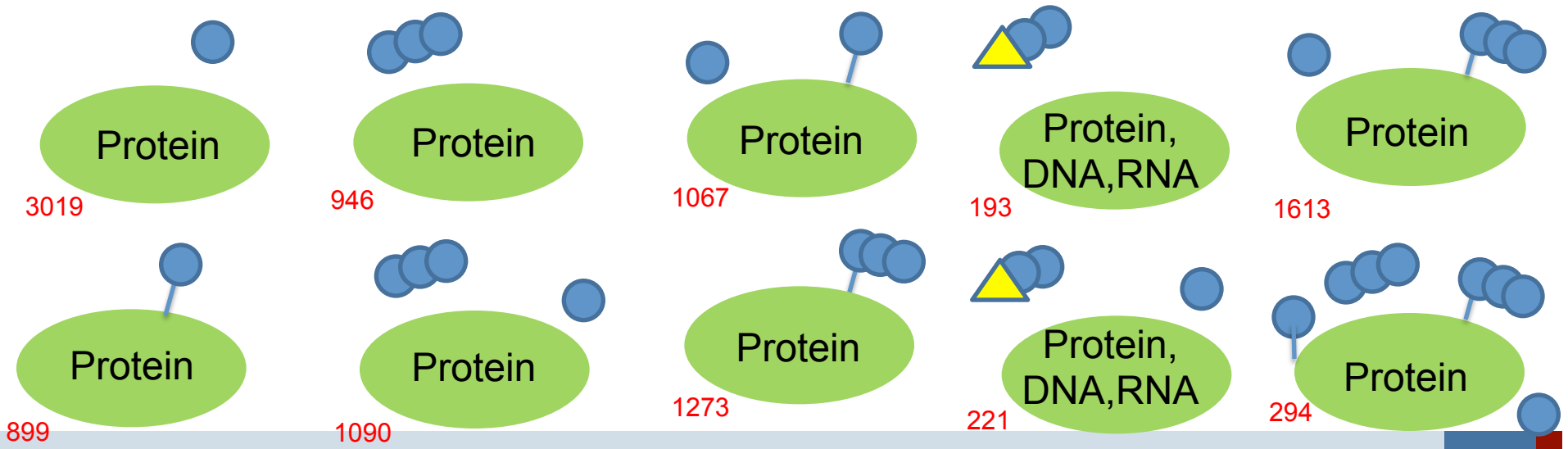
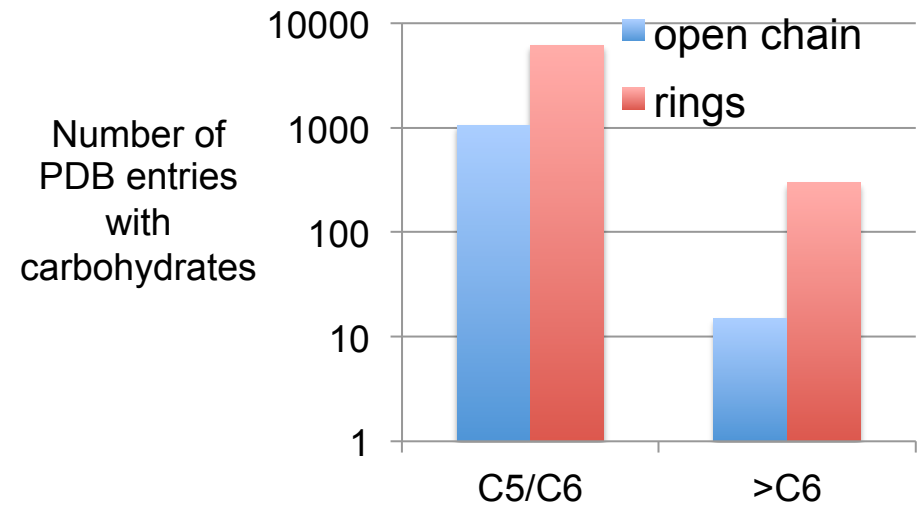
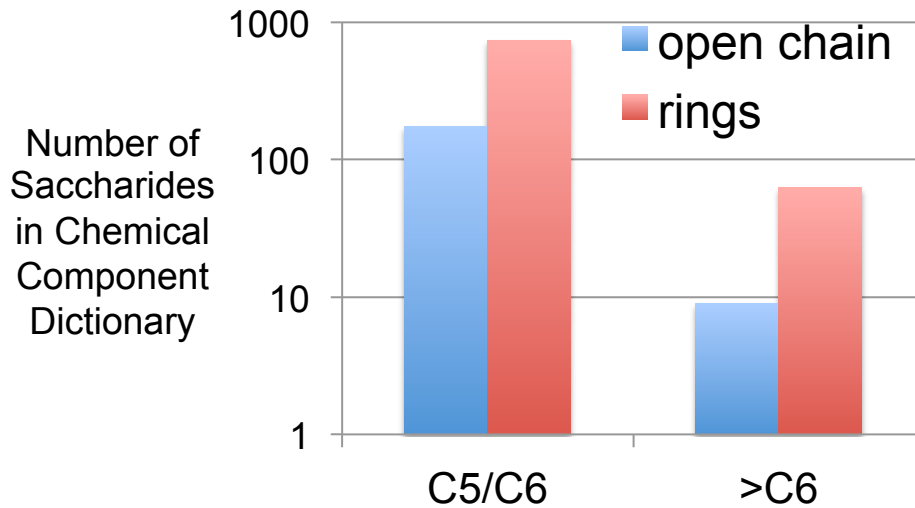
Goal

- Represent data consistently within the archive, in agreement with glycobiology community standards
- Enable searches for carbohydrates in the PDB archive

Plan

- Identify and analyze carbohydrate-containing entries
- Create standard representation for branched polymers
- Incorporate standard nomenclature
- Create a strategy for remediation

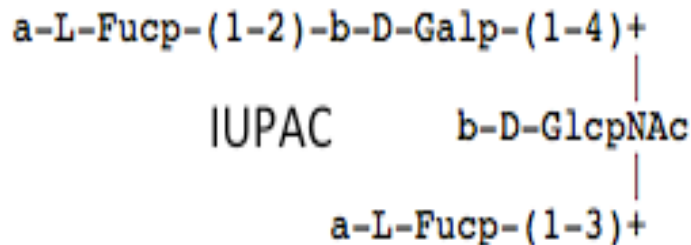
Carbohydrate Remediation Scope



Carbohydrate Remediation Plan

- Represent carbohydrate molecules as polymers of monosaccharides as appropriate
- Adopt glycobiology community standard nomenclature (LINUCS and IUPAC)

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



PDB ID: 2wmg

Protein Modifications (PDBe)

Problem

- Inconsistent annotation of PMs in the archival files results in the inability to search for these important structures

Goal

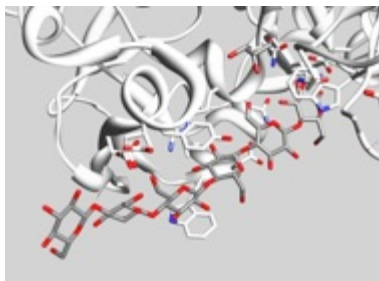
- To identify, classify, and represent all natural protein modifications consistently within the PDB archive and mutually mapped to UniProtKB

Scope

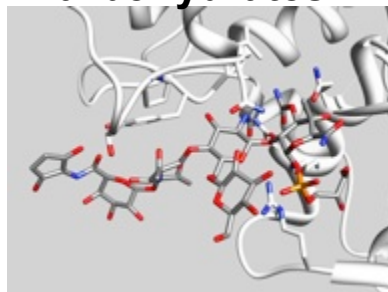
- Chemically modified ribosomal protein including post-translation
-

Remediated Data

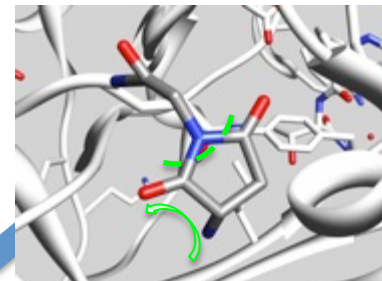
Carbohydrates



Carbohydrates



Protein modifications

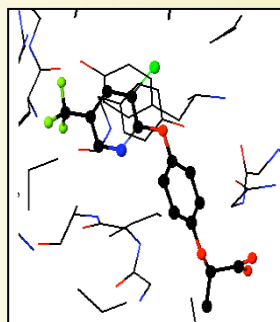


wwPDB Common D&A System

Annotation Pipeline

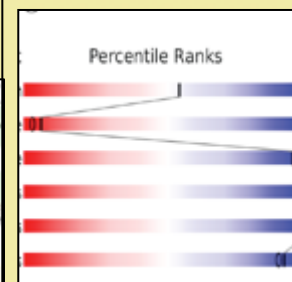
L	P	P	F	D	K	G	R	T	A	T	H	E	I	G	H	W	L	N	L	Y
L	P	P	F	D	K	G	R	T	A	T	H	E	I	G	H	W	L	N	L	Y
L	P	P	F	D	K	G	R	T	A	T	H	E	I	G	H	W	L	N	L	Y
L	P	P	F	D	K	G	R	T	A	T	H	E	I	G	H	W	L	N	L	Y
L	P	P	F	D	K	G	R	T	A	T	H	E	I	G	H	W	L	N	L	Y
L	P	P	F	D	K	G	R	T	A	T	H	E	I	G	H	W	L	N	L	Y
L	P	P	F	D	K	G	R	T	A	T	H	E	I	G	H	W	L	N	L	Y
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
C	S	N	G	P	N	G	D	M	F	M	N	Y	L	D	Y	V	D	D	K	C
C	S	N	G	P	N	G	D	M	F	M	N	Y	L	D	Y	V	D	D	K	C
C	S	N	G	P	N	G	D	M	F	M	N	Y	L	D	Y	V	D	D	K	C
C	S	N	G	P	N	G	D	M	F	M	N	Y	L	D	Y	V	D	D	K	C

Sequence Processing



Ligand Processing

Manual and Automated Annotation



Validation

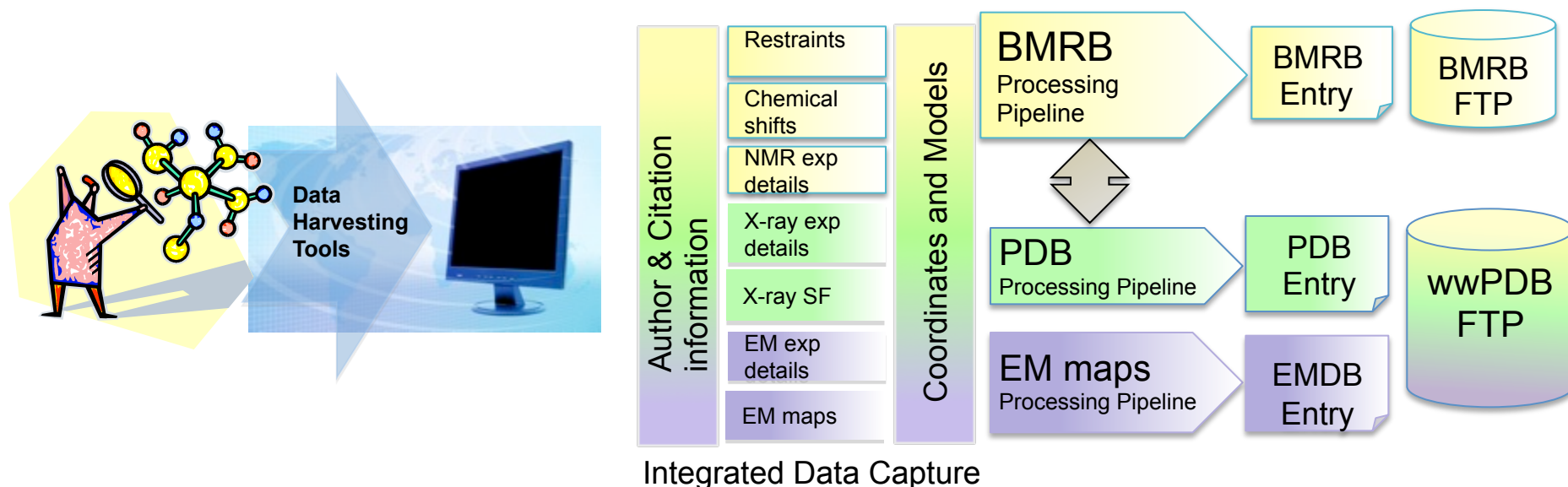
Common Deposition & Annotation Tool

Martha Quesada



wwpdb.org

The Vision



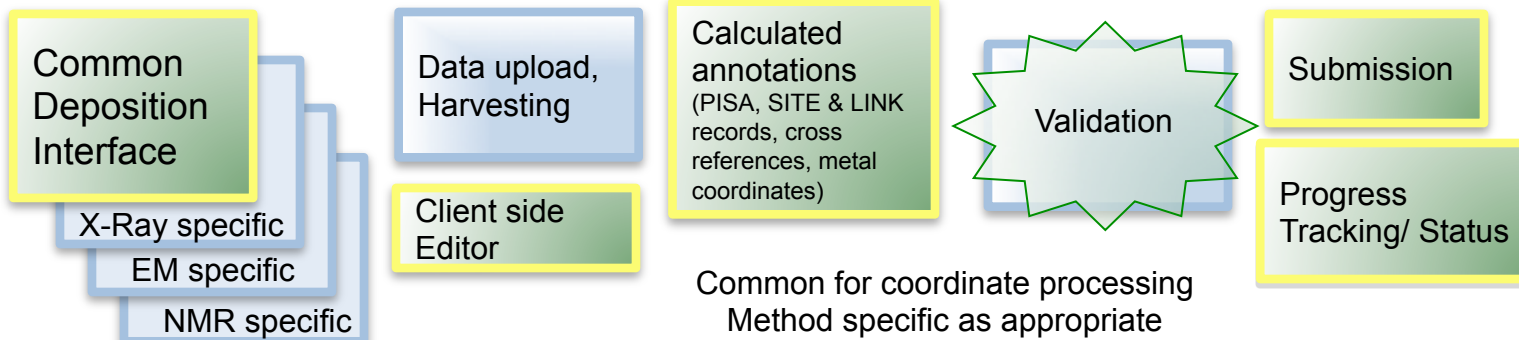
Deliverables: Standardization, Quality and Efficiency

Supporting

- Larger and more complex biological molecules
- Expanded annotation
- Increased throughput: Automation and validation of routine submissions

wwPDB Common Deposition and Annotation Pipeline

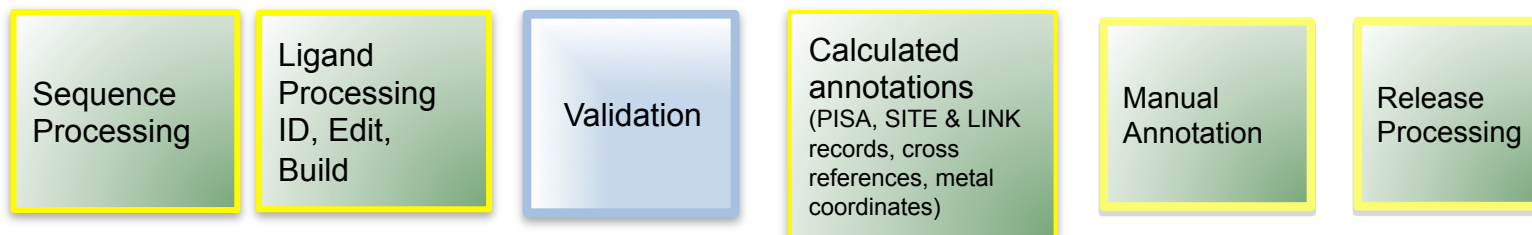
Deposition Pipeline



Communication System

Workflow-Automation System

Annotation Pipeline



Green and yellow components are common for X-ray, NMR, & EM processing

D&A Deposition Pipeline Deliverables

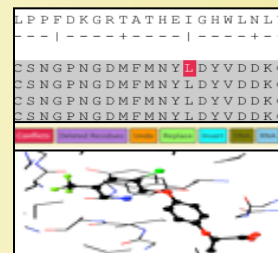
Deposition Pipeline



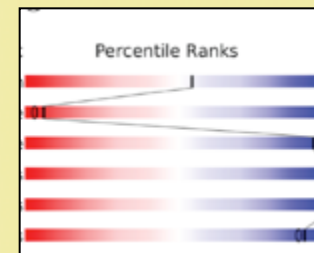
Data Harvesting



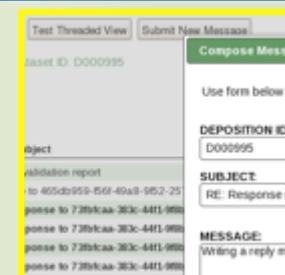
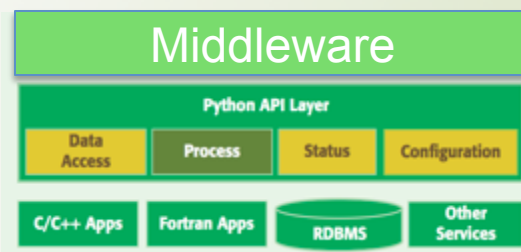
Deposition User Interface



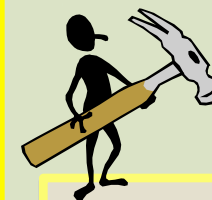
Functional Components



Infrastructure



Communication



Release Processing

Processing Modules

Sequence

WORLDWIDE PDB PROTEIN DATA BANK

Sequence Alignment Tool
Identifier: rcsb067042 Instance: RCSB

jmol Load 3D Viewer ALA/GLY Change Mark for Deletion Clear Selection Undo

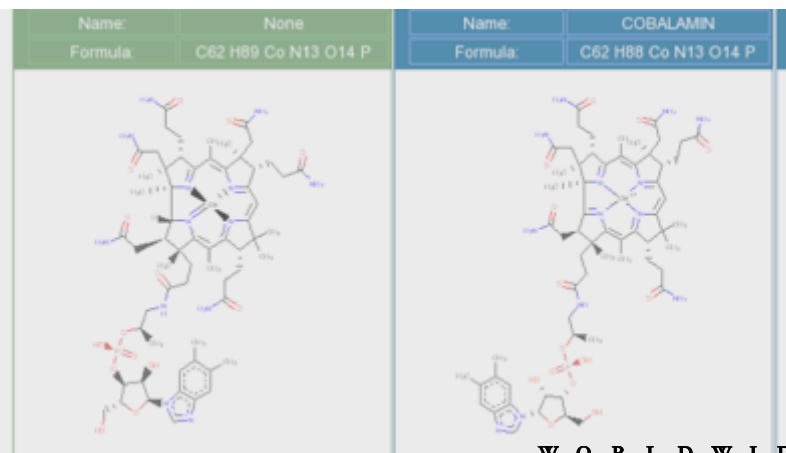
```
AUTH PDB:A V(1) DIVMSQSPSSLAVSAGEKVTMSCKSSQSLLSRRTKXNYLAWYQQKPGQSPKLLIYMA
XYZ PDB:A V(1) DIVMSQSPSSLAVSAGEKVTMSCKSSQSLLSRTRKXNYLAWYQQKPGQSPKLLIYMA
UNP:Q52L64 (R1,V1) DIVMSQSPSSLAVSYGEKVTMSCKSSQSLLSYNOXNYLAWYQQKPGQSPKLLIYMA
1 |---+---|---+---|---+---|---+---|---+---|---+---|---+---|---+---|---+---|---+---|
AUTH PDB:A V(1) ESGVDPDRFTSGSGSDTDFLTIITGVQAEDLAVYYCKQSYN..LRTPGGGKLEIKRAD
XYZ PDB:A V(1) ESGVDPDRFTSGSGSDTDFLTIITSVQAEDLAVYYCKQSYN..LRTPGGGKLEIKRAD
UNP:Q52L64 (R1,V1) ESGVDPDRFTSGSGSDTDFLTIITGVQAEDLAVYYCKQSYNYPLTFPGAGTKLDRRAD
61 |---+---|---+---|---+---|---+---|---+---|---+---|---+---|---+---|---+---|
AUTH PDB:A V(1) TVSIFFPPSSEQLTSGGASVVCFLNNFYPKDINVKWKIDGSEERQNGVLSNNTDQDSKD
XYZ PDB:A V(1) TVSIFFPPSSEQLTSGGASVVCFLNNFYPKDINVKWKIDGSEERQNGVLSNNTDQDSKD
UNP:Q52L64 (R1,V1) TVSIFFPPSSEQLTSGGASVVCFLNNFYPKDINVKWKIDGSEERQNGVLSNNTDQDSKD
121 |---+---|---+---|---+---|---+---|---+---|---+---|---+---|---+---|---+---|
AUTH PDB:A V(1) SMSSTLTTLTKDEYERHNSYTCATHKTSSTSPIVKSFNRNEC
XYZ PDB:A V(1) SMSSTLTTLTKDEYERHNSYTCATHKTSSTSPIVKSFNRNEC
UNP:Q52L64 (R1,V1) SMSSTLTTLTKDEYERHNSYTCATHKTSSTSPIVKSFNRNEC
```

Color legend: Conflicts Deleted Residues Undo Replace Insert DNA RNA

Ligand

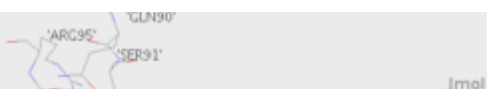
2012 Enhancements

- Create new chemical definitions by splitting or merging existing definitions
 - Integration with Chemical Component Dictionary
 - Data file extractor User Interface
- Component archiving User Interface
 - In production at RCSB PDB & PDBj
 - Unit and integration testing at PDBe



2012 Enhancements

- Chimeric proteins supported
- Sequence match sorting ability
- Create new chemical definitions by splitting or merging existing definitions
- Unit and integration testing at all sites



RCSB PDB developers: Zukang Feng, John Westbrook, Raship Shah, Raul Sala, Dimitris Dimitropoulos

Annotation Pipeline Deliverables

Web form data entry and editing

Automated Annotations

The screenshot shows a web interface for data entry. On the left is a vertical navigation menu with buttons for: Deposition, Citation (selected), Advisory and Remarks, Polymer Description, Crystallization, Crystal Data, and Data Collection. The main content area is titled 'Citation Title' and contains a table with three entries. The first entry is highlighted in green and reads: 'Hydrogen bonding and solvent structure in an antibody interface. Crystal structures and thermodynamic characterization of three Fv mutants complexed with lysozyme.' The second entry is: 'Bound Water Molecules and Conformational States Help Mediate an Antigen-Antibody Association'. The third entry is: 'Solvent Rearrangement in an Antigen-Antibody Complex Introduced by Site-Directed Mutagenesis of the Antibody Combining Site'. Below the table, it says 'Showing 1 to 3 of 3 entries'. There is an 'Add Record' button and a note: 'Double-click on leftmost column'. At the bottom, a color legend indicates that a green box represents a 'Primary Key Column' and a yellow box represents a 'New Row'.

Semi-automated

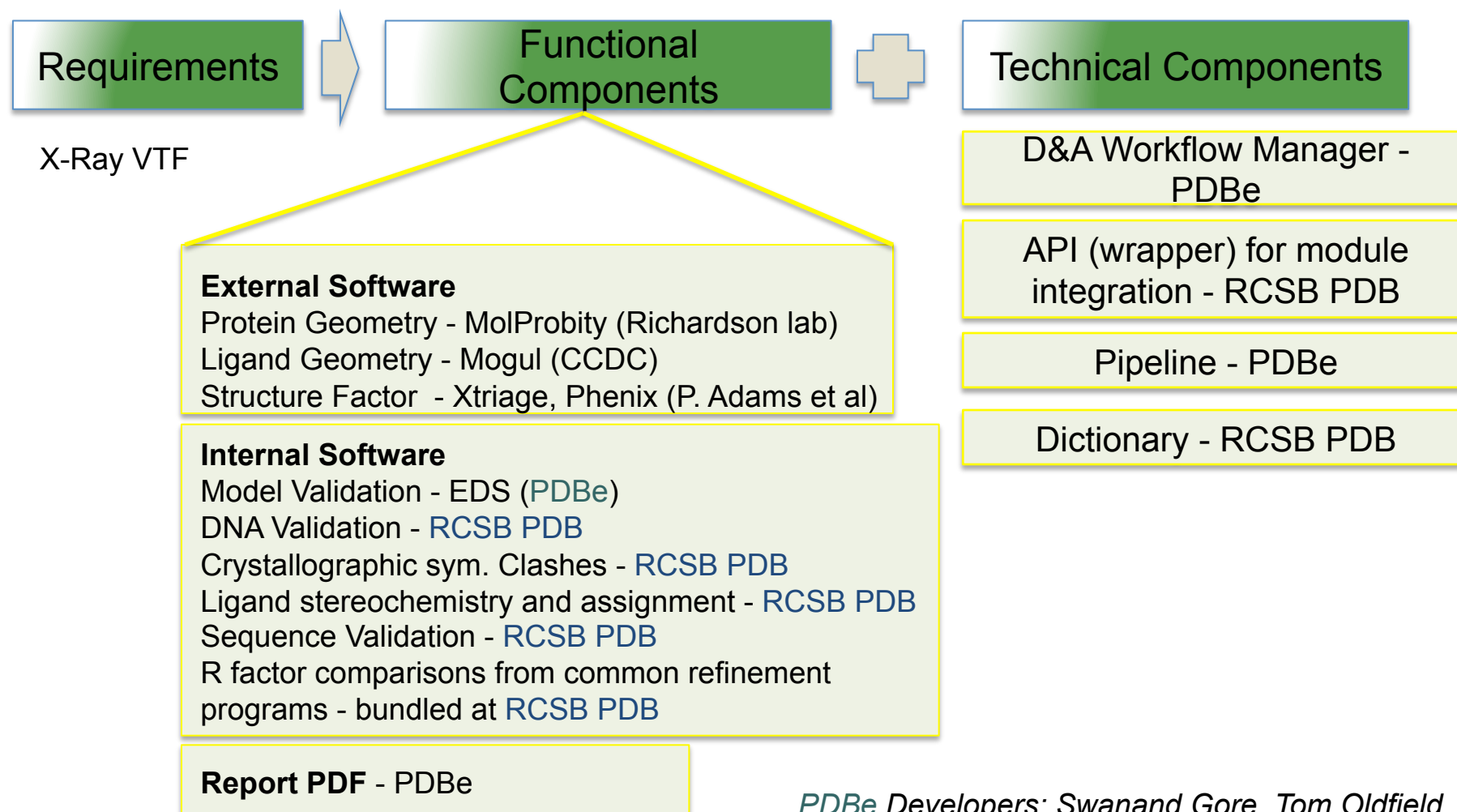
- Biological Assemblies (PISA)

Automated Annotations

- Site environment
- Solvent position
- Linkage review module
- Secondary structure

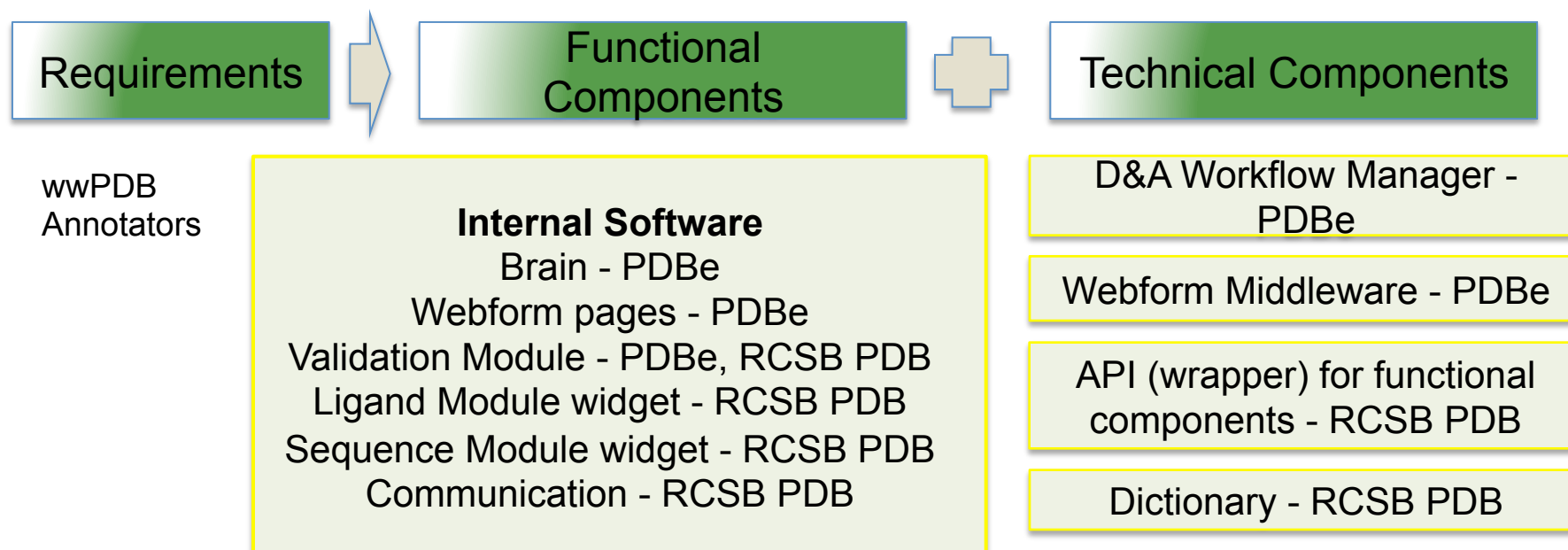
Code was refactored where needed, Middleware & User Interfaces for review and editing were developed

Validation Module



*PDBe Developers: Swanand Gore, Tom Oldfield
RCSB PDB Developers: Zukang Feng, Huanwang Yang, John Westbrook*

Deposition Pipeline



PDBe Developers: Swanand Gore, Tom Oldfield, Robert Hulme, Pieter Hendrick, Saqib Mir

RCSB PDB Developers: Zukang Feng, Huanwang Yang, Raul Sala, Raship Shah, John Westbrook



Workload Distribution

Distribution will take into account

- Restart of a deposition session
- Advisory and funding guidelines
- Time zone, to facilitate “help” and communication
- Load balance based on resource capacity

Single, wwPDB-branded, point of contact for all new depositions (e.g., <http://wwpdb.org/deposit>)

RCSB PDB Developers: Raul Sala, John Westbrook

Communication Interface



WORLDWIDE PDB PROTEIN DATA BANK

wwPDB Messaging Module

Test Conventional View Test Threaded View Submit New Message

Deposition Dataset ID: D000995

Compose Message

Use form below to submit message.

DEPOSITION ID: D000995

SUBJECT: RE: Response to bbadd288

MESSAGE: Writing a reply message here

wwPDB Message

Submit New Message

SUBJECT: Response to bbadd288-f343-41bf-ab88-486

SENDER: Depositor

DATETIME: 2012-06-05 14:51:27

DEPOSITION ID: D000995

All of the C-alpha and C-beta atom coordinates have got reflections that we misplaced earlier. We are hoping that we discover the source of super-linearity in the coordinates on a 4 Angstrom vector. We agree that the real-space R- problem will be resolved when we upgrade the operating

- Look and feel of email
- Linked to web page content

EM V1.0

- Dictionary enhancements complete
- EM specific interfaces to be implemented in October
- Large data file requirements to be supported in the deposition module to be implemented in November

Future

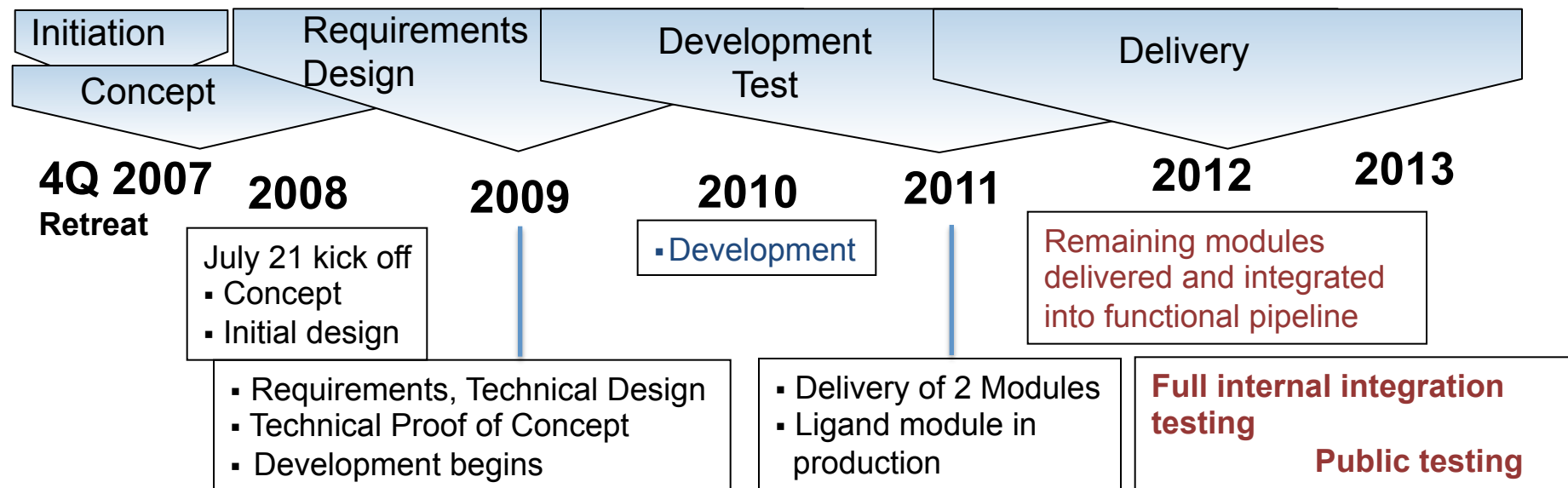
- Additional visualization, data harvesting
- Validation requirements from EM VTF to be supported

NMR V1.0

- Dictionary data items in place
- Data requirements are defined and mapped for chemical shifts
- Integration of software for PDB atom nomenclature correspondence to NMR experimental data in place

- D&A NMR User Interfaces to be implemented in October
- Common D&A and ADIT-NMR data exchange to be implemented

wwPDB Common D&A Tool Project Timeline - Actual

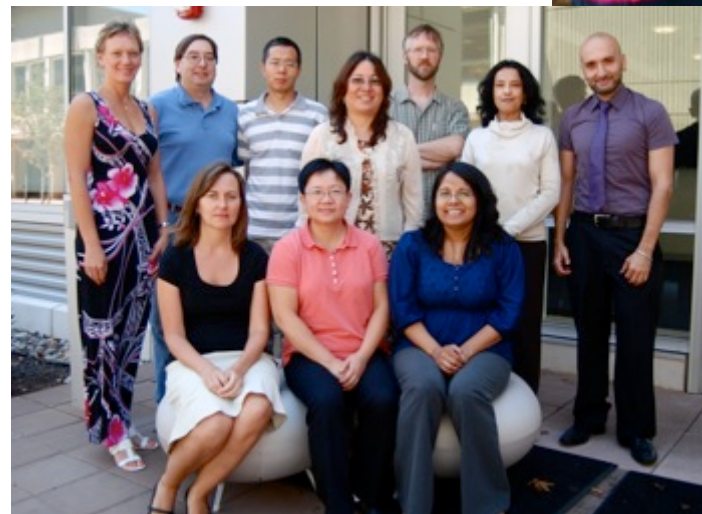


2012 Deliverable: wwPDB Common D&A System Version 1.0

- Able to process a file from deposition through annotation
- Supports all existing D&A processes and procedures
- Offers enhanced User Interfaces for functional modules and deposition process (within reason)
- Provides a “Work Flow” infrastructure that enables task tracking and automation

Common D&A Project Team

*Experience,
Expertise and
Diverse Skills*
representing
the broad
interests of
the wwPDB

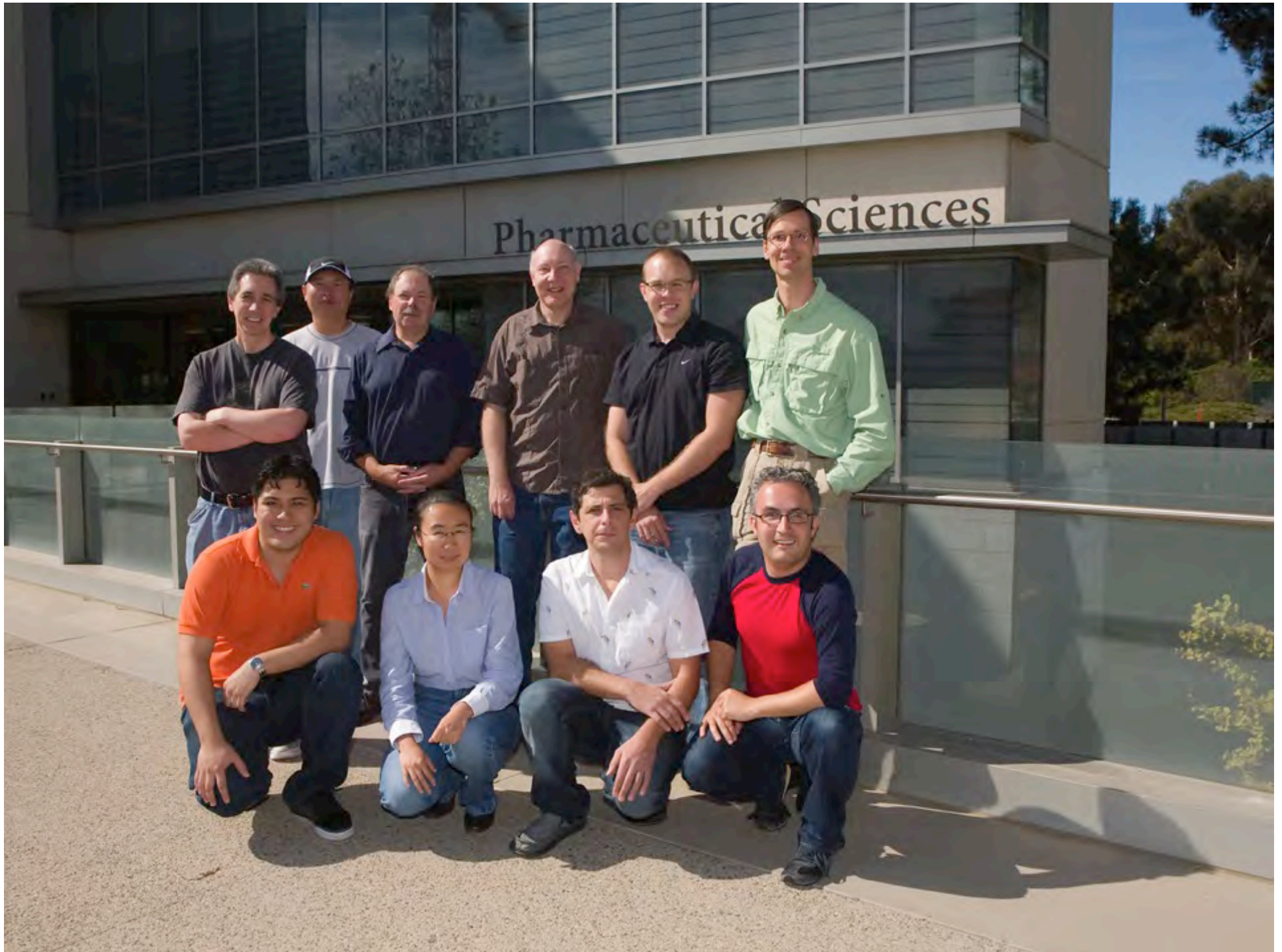




Data Out: Website Features

Peter Rose, Philip Bourne





Data Out

Goal: Enable research and discovery in science, medicine, and education by

- Presenting an accurate, concise and meaningful understanding of structure and structure-function relationships
- Answering broad biological questions where macromolecular structure is key
- Enabling computational analysis involving macromolecular structure

Outline

- Searching and Browsing the Archive
 - Tools and Services
 - Website Usage Statistics
 - Hardware Renewal
 - Plans
-

Searching and Browsing the Archive

- Precise text search results through search suggestions
- Efficient browsing of archive and selection of relevant hits using text and graphics
- Mobile searching and browsing



Guiding Users to Specific Results

Search Suggestions

kinase

Molecule of the Month

- cAMP-dependent Protein **Kinase** (PKA)
- Src Tyrosine **Kinase**

Molecule Name

- Cyclin-dependent **kinase** 2 (257)
- Mitogen ... protein **kinase** 14 (194)
- Nucleoside diphosphate **kinase** (82)
- cAMP-dependent protein **kinase** catalytic ... (131)
- Thymidylate **kinase** (57)
- Aurora **kinase** A (64)

Enzyme Classification

- 2.7.11 ... protein **kinase** (1382)
- 2.7.10 ... **kinase** (496)
- 2.7.11.1: Non-specific serine/threonine protein **kinase**
- 2.7.11 ... -dependent **kinase** (300)
- 2.7.13.3: Histidine **kinase** (198)
- 2.7.11 ... protein **kinase** (286)

PDB Text

- kinase**
- kinases**
- kinase
- kdnase

Chemical Name

- L12: INHIBITOR OF P38 **KINASE**
- 274: MET **KINASE** INHIBITOR
- B13: INHIBITOR ... PROTEIN **KINASE**-1

Structural Domains

- c-src tyrosine **kinase** [SCOP] (58)
- Phosphorylase **Kinase**; domain ... (1380)
- Nucleoside diphosphate **kinase**, NDK ... (60)
- Phosphatidylinositol 3-**kinase** Catalytic ... (286)
- Thymidylate **kinase** [SCOP] (36)
- Backbone ... histidine **kinase** receptor ... (1)

Ontology Terms

- intracellular protein **kinase** cascade ... (895)
- cyclin ... protein **kinase** holoenzyme ... (296)
- kinase** activity ... (2613)
- AT : ACETYLGLUTAMATE **KINASE** [Genome ... (1)
- D08.811 ... -Diphosphate **Kinase** [MeSH ... (97)
- positive ... protein **kinase** B signaling ... (242)

Mouse over for more details

Smart Suggestions

GSTFIVITAGADGLAT

Sequence

- Very significant (E Cut Off:0.001) to **GSTFIVITAGADGLAT**
- Significant (E Cut Off:0.01) to **GSTFIVITAGADGLAT**
- Includes Insignificant (E Cut Off:1) to **GSTFIVITAGADGLAT**
- Extended Search (E Cut Off:10) to **GSTFIVITAGADGLAT**

SMILES

- Has exact structure **C1c1ccccc1**
- Has sub-structure **C1c1ccccc1**
- Is very similar (95%) with **C1c1ccccc1**
- Is similar (70%) with **C1c1ccccc1**
- Super-structure of **C1c1ccccc1**

SMILES string

resolution

PDB Text

- resolution**
- resolutions**

Advanced Search

- X-ray **Resolution**

[Link to Advanced Search](#)



Advanced Search

Advanced Search Interface

All/Experimental Type/Molecule Type ?

Retrieve all PDB entries or a subset based on experimental method or molecule type

Experimental Method: X-ray Result Count

Molecule Type: Protein

AND

Choose a Query Type: ?

Choose a Query Type:

- Quick Search**
 - All/Experimental Type/Molecule Type
- ID(s) and Keywords**
 - PDB ID(s)
 - PubMed ID(s)
 - UniProtKB Accession Number(s)
 - Text Search
 - mmCIF Keyword Search (Classification)
 - Pfam ID(s)
- Structure Annotation**
 - Structure Title
 - Structure Description
 - Macromolecule Name
- Deposition**
 - Author Name
 - Deposit Date
 - Release Date
 - Revision Date
 - Latest Released Structures

Result Count

Add Search Criteria +

Clear All Parameters Submit Query

- Search data items in a variety of categories
- Combine queries using AND / OR
- Contextual help

- New options include
 - Pfam annotations (updated weekly)
 - Quick searches
 - Inter-residue links (LINK records)

Fall 2012: Views with Synchronized “Shopping Cart”

Showing 1 - 25 of 646 Results Results : 50 Page: 1 of 26

Filter: Custom **View:** Detailed **Reports:** Select one... **Sort:** Relevance

1N50 **Structural consequences of a cancer-causing BRCA1-BRCT missense mutation**

Authors: Williams, R.S., Glover, J.N.M.

Release: 2002-12-25 **Classification:** Antitumor Protein

Experiment: X-RAY DIFFRACTION with resolution of 2.80 Å **Residue Count:** 214

Compound: 1 Polymer [[Display Full Polymer Details](#) | [Display for All Results](#)]
2 Ligands [[Display Full Ligand Details](#) | [Display for All Results](#)]

Citation: **Structural consequences of a cancer-causing BRCA1-BRCT missense mutation** (2003) J.Biol.Chem. **278**: 2630-2635 [[Display Full Abstract](#) | [Display for All Results](#)]

Search Hit: Title: Structural consequences of a cancer-causing BRCA1-BRCT missense mutation

1JNX **Crystal structure of the BRCT repeat region from the breast cancer associated protein, BRCA1**

Authors: Williams, R.S., Green, R., Glover, J.N.M.

Release: 2001-09-21 **Classification:** Gene Regulation

Experiment: X-RAY DIFFRACTION with resolution of 2.50 Å **Residue Count:** 214

Compound: 1 Polymer [[Display Full Polymer Details](#) | [Display for All Results](#)]
1 Ligand [[Display Full Ligand Details](#) | [Display for All Results](#)]

Fall 2012: Condensed View for Rapid Browsing

Showing 1 - 25 of 646 Results Results : 50 Page: 1 of 26



Filter: Custom **View:** Condensed **Reports:** Select one... **Sort:** Relevance

PDB ID	Title	Macromolecule Name
<input checked="" type="checkbox"/> 1N5O	Structural consequences of a cancer-causing BRCA1-BRCT missense mutation	Breast cancer type 1 susceptibility...
<input type="checkbox"/> 1JNX	Crystal structure of the BRCT repeat region from the breast cancer associated protein,	BREAST CANCER TYPE 1 SUSCEPTIBILITY...
<input checked="" type="checkbox"/> 2KW6	Solution NMR Structure of Cyclin-dependent kinase 2-associated protein 1 (CDK2-associated	Cyclin-dependent kinase 2-associate...
<input type="checkbox"/> 3HG1	Germline-governed recognition of a cancer epitope by an immunodominant human T cell	MHC class I antigen, Beta-2-microgl...
<input type="checkbox"/> 2H80	NMR structures of SAM domain of Deleted In Liver Cancer 2 (DLC2)	StAR-related lipid transfer protein...
<input checked="" type="checkbox"/> 2JW2	Validation of inter-helical orientation of the steri-alpha-motif of human deleted in liver	StAR-related lipid transfer protein...
<input checked="" type="checkbox"/> 2I3B	Solution Structure of a Human Cancer-Related Nucleoside Triphosphatase	Human Cancer-Related NTPase
<input checked="" type="checkbox"/> 1T15	Crystal Structure of the Brca1 BRCT Domains in Complex with the Phosphorylated	Breast cancer type 1 susceptibility...
<input checked="" type="checkbox"/> 2UZR	A TRANSFORMING MUTATION IN THE PLECKSTRIN HOMOLOGY DOMAIN OF AKT1 IN	RAC-ALPHA SERINE/THREONINE-PROTEIN ...
<input checked="" type="checkbox"/> 2UZS	A TRANSFORMING MUTATION IN THE PLECKSTRIN HOMOLOGY DOMAIN OF AKT1 IN	RAC-ALPHA SERINE/THREONINE-PROTEIN ...
<input checked="" type="checkbox"/> 1XI2	Quinone Reductase 2 In Complex with Cancer Prodrug CB1954	NRH dehydrogenase [quinone] 2
<input checked="" type="checkbox"/> 3E78	Structure determination of the cancer-associated Mycoplasma hyorhinis protein Mh-p37	High affinity transport system prot...
<input checked="" type="checkbox"/> 3E79	Structure determination of the cancer-associated Mycoplasma hyorhinis protein Mh-p37	High affinity transport system prot...
<input checked="" type="checkbox"/> 1V3A	Structure of human PRL-3, the phosphatase associated with cancer metastasis	protein tyrosine phosphatase type 1...

Fall 2012: Gallery Browsing and Selection

Showing 1 - 25 of 646 Results Results : 50 Page: 1 of 26

Filter: Custom **View: Gallery** Options: Select one.. Sort: Relevance

 1N5G <input checked="" type="checkbox"/>	 1JNX <input type="checkbox"/>	 2KW6 <input checked="" type="checkbox"/>	 3HG1 <input type="checkbox"/>	 2H80 <input type="checkbox"/>	 2JW2 <input checked="" type="checkbox"/>	 2I3B <input checked="" type="checkbox"/>	 1T15 <input checked="" type="checkbox"/>
 2UZR <input checked="" type="checkbox"/>	 2UZS <input checked="" type="checkbox"/>	 1XI2 <input checked="" type="checkbox"/>	 3E78 <input checked="" type="checkbox"/>	 3E79 <input checked="" type="checkbox"/>	 1V3A <input checked="" type="checkbox"/>	 2ZTB <input checked="" type="checkbox"/>	 3G5K <input checked="" type="checkbox"/>
 3G5P <input checked="" type="checkbox"/>	 1OQA <input checked="" type="checkbox"/>	 2L4S <input checked="" type="checkbox"/>	 2OIQ <input checked="" type="checkbox"/>	 2RKB <input checked="" type="checkbox"/>	 2HQ6 <input checked="" type="checkbox"/>	 1X2J <input checked="" type="checkbox"/>	 1X2R <input checked="" type="checkbox"/>

Fall 2012: Timeline View

RCSB PDB-101

Yonath, A.
 This timeline displays the structures that Yonath, A. is a Depositor or Primary Citation Author. Structures are ordered by the deposition date shown, and link to the Structure Summary page for the entry. The first instance of a protein or protein complex deposited by this author is displayed larger than future occurrences of similar molecules.

2010

Sort by
 Year

Joint Center for Structural Genomics (JCSG)
 This timeline displays the structures that the Joint Center for Structural Genomics (JCSG) as a Depositor or Primary Citation Author. Structures are ordered by the deposition date shown, and link to the Structure Summary page for the entry. The first instance of a protein or protein complex deposited by this Author is displayed larger than future occurrences of similar molecules.

2012

Sort by
 Newest
 Oldest
 Year

2012 (2)
 2011 (31)
 2010 (40)
 2010-2005
 2005-2000
 2000-1995
 1995-1992

Back to Search > **Timeline View**

View : Timeline : ENZYMECLASSIFICATION is 2: Transferases and ENZYMECLASSIFICATION is 2.7: Transferring phosphorous-containing groups and ENZYMECLASSIFICATION is 2.7.10: Protein-tyrosine kinases and ENZYMECLASSIFICATION is 2.7.10.2: Non-specific protein-tyrosine kinase

2012

4AGW February 01

Molecule: Proto-oncogene tyrosine-prote...



Discovery of a small molecule type II inhibitor of wild-type and gatekeeper muta...

4DBK January 10

Molecule: Tyrosine-protein kinase Lck




Crystal structure of a SH3-SH2 domains of a lymphocyte-specific protein tyrosine...

2011

3UYO December 06

Molecule: Tyrosine-protein kinase ABL1



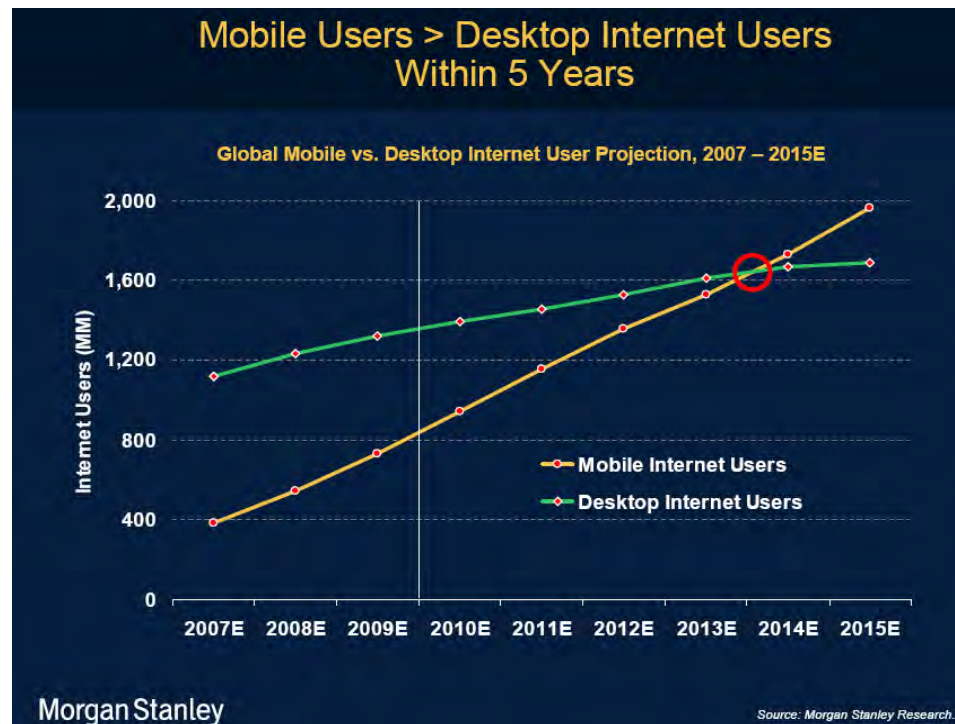
Crystal structure of monobody SH13/ABL1 SH2 domain complex

3US4 November 22

Megakaryocyte-associated tyrosine-protein kinase

Why Support Mobile Devices

- Mobile devices well suited to deploy educational materials (*Molecule of the Month*)
- Mobile device users expected to exceed desktop users by 2014



Explore the RCSB PDB on the Go with RCSB PDB *Mobile*



Browse
*Molecule of
the Month*
articles

Simple
search

**iPhone
iPod**

Interactive
3D viewer*

Browse
search
results

iPad

*NDKMol, Takanori Nakane, Kyoto University

Molecule of the Month in ePub Format

PDB-101 - Structural View of Biology >> Protein Synthesis

O-GlcNAc Transferase

September 2011 Molecule of the Month by David Goodsell
doi: [10.2210/rcsb_pdb/mom_2011_9](https://doi.org/10.2210/rcsb_pdb/mom_2011_9) (ePub Version)

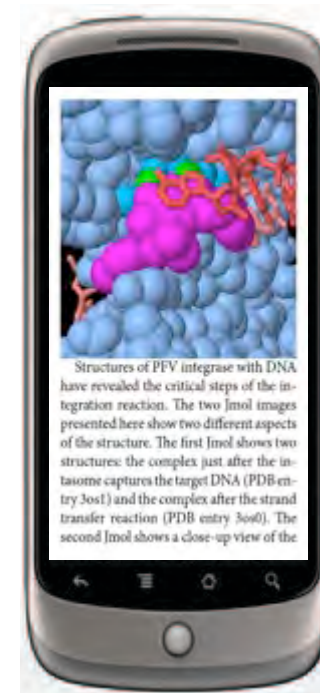
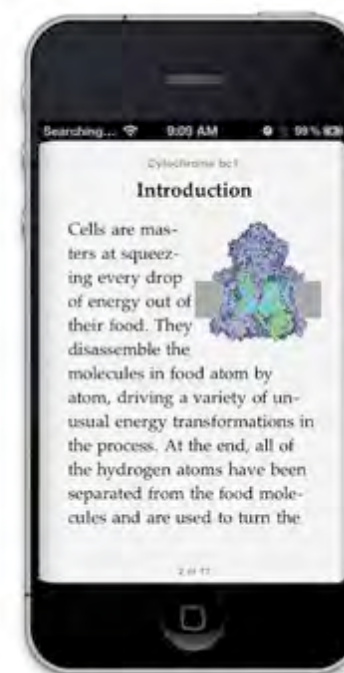
Keywords: protein modification, regulation, glucose sensing, diabetes



iPad



iPhone/iPod



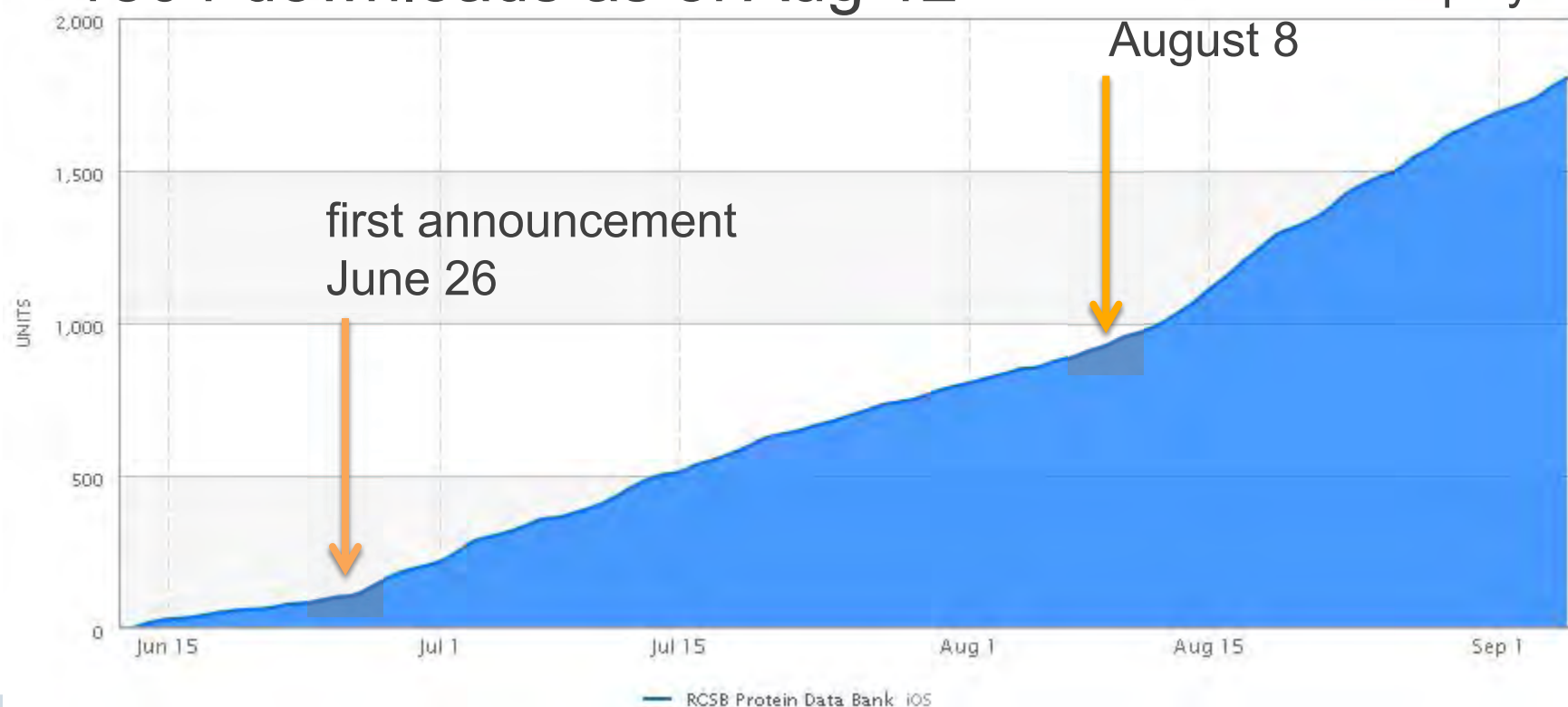
Android

RCSB PDB *Mobile App* Downloads

- Deployed in iTunes store (free)
 - iPhone/iPod: May 29
 - iPad: Aug 8

- 1804 downloads as of Aug 12

iPad version deployed
August 8



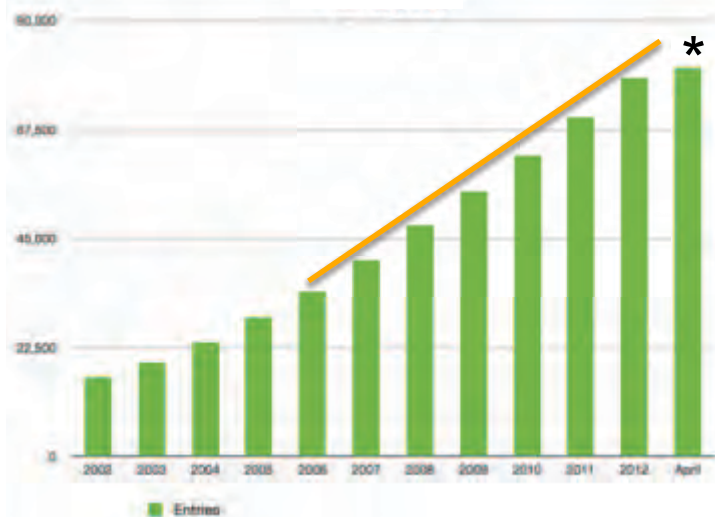


Data Out: Tools and Services

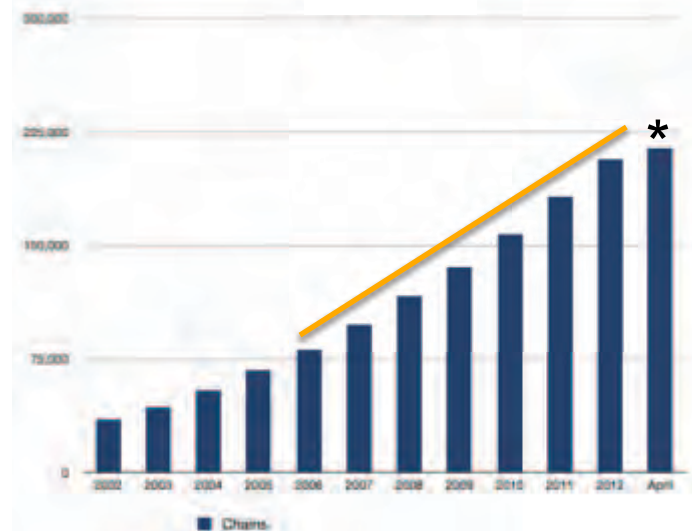


Growth of PDB and Sequence Clusters

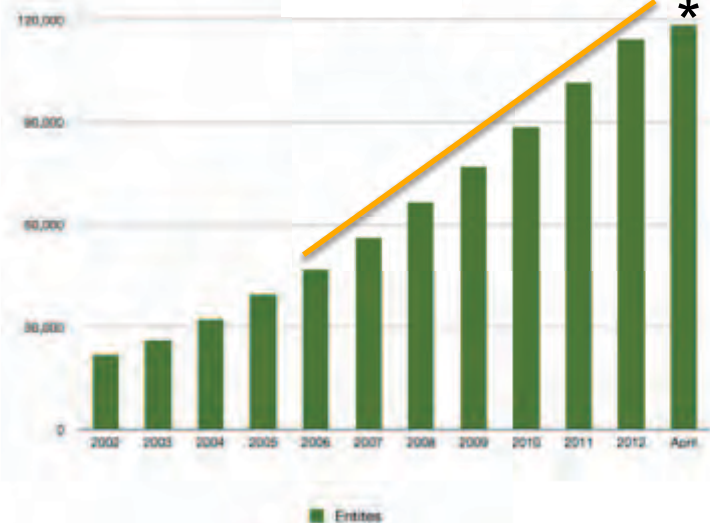
Entries



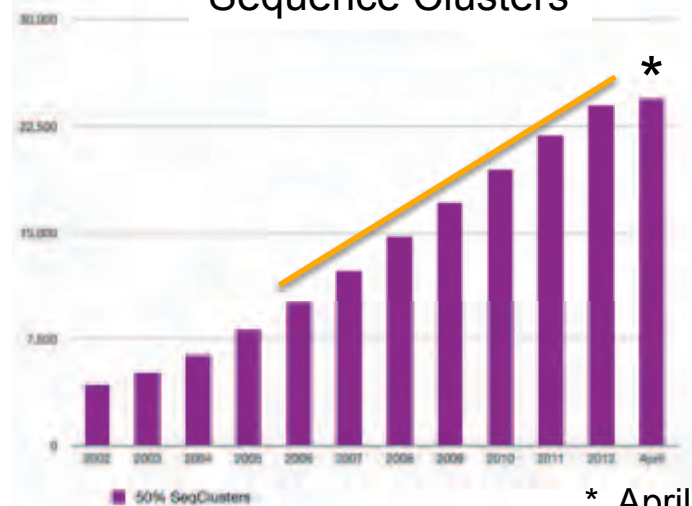
Chains



Entities

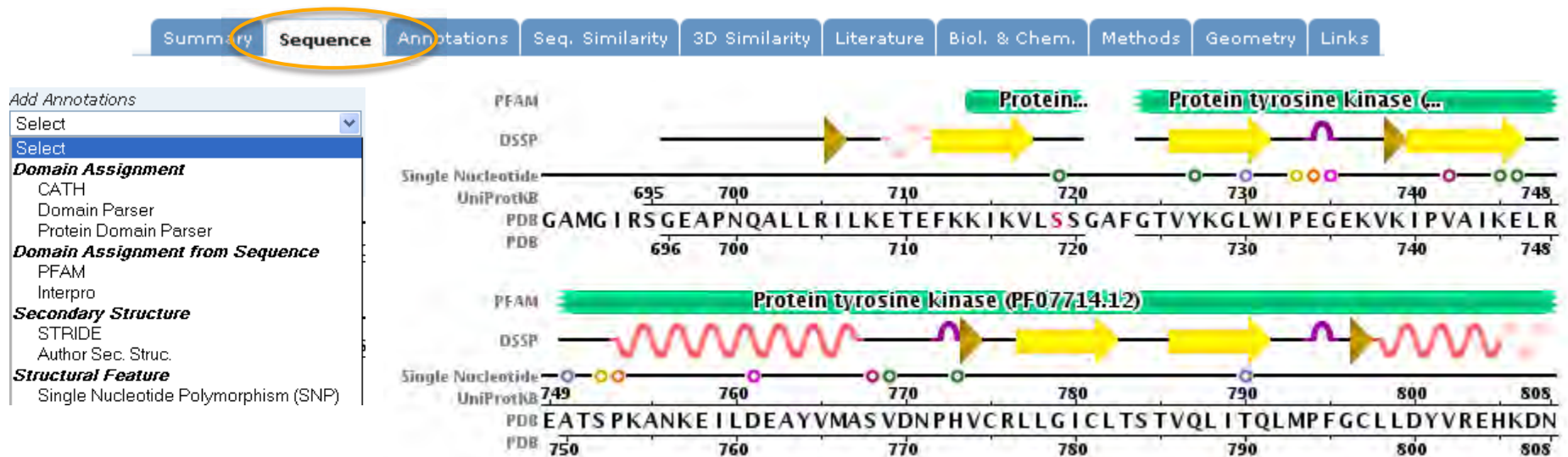


Sequence Clusters



* April 2012

Sequence Annotations



Annotation sources

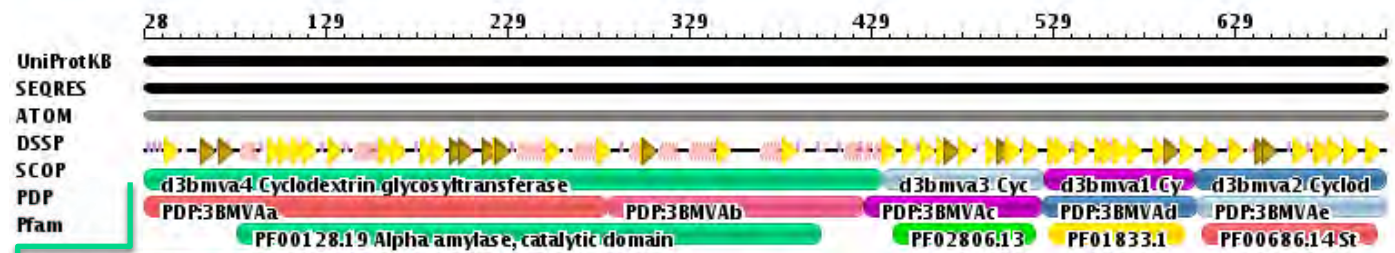
- Pfam: weekly update by running PFAM HMMER Web Service
 - xfam.wordpress.com/2012/05/09/pdb-pfam-mapping/#more703
- UniProt to PDB Mapping: SIFTS project
 - www.ebi.ac.uk/pdbe/docs/sifts
- SNP: LS-SNP
 - ls-snp.icm.jhu.edu
- Binding sites: SITE records in PDB entries

Pre-calculated Domain-based Structural Alignments

PDB 3BMV

- Summary
- Sequence
- Annotations
- Seq. Similarity
- 3D Similarity**
- Literature
- Biol. & Chem.
- Methods
- Geometry
- Links

Entity #1: Chains: A
Description: Cyclomalto-dextrin glucanotransferase protein
Length: 683





Results for domain d3bmva4

SCOP Domain ID d3bmva4 (chain 1) vs. representatives of other sequence clusters (chain 2)

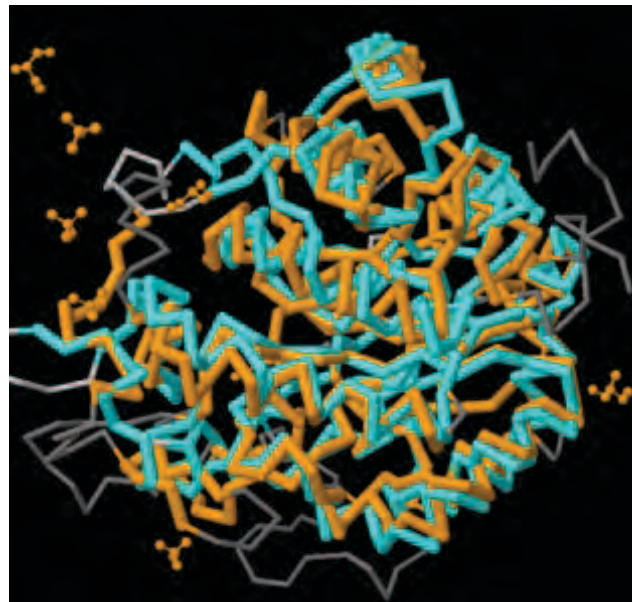
Page 1 of 7 View 1 - 15 of 92

Rank	Results	Domain 2	Title	P-value	Score	Rmsd	Len1	Len2	%ID	%Cov1	%Cov2
1	view	d2guya2	Alpha-amylase A2guy A:1-3	1.11E-16	788.51	2.61	406	381	22	91	97
2	view	PDP:4AEE	ALPHA-AMYLASE, CATALYT	1.33E-15	643.75	2.01	406	309	20	74	97
3	view	d1j0ha3	neopullulanase1j0h A:124-	1.64E-14	766.83	3.21	406	382	22	89	94
4	view	d1ji1a3	ALPHA-AMYLASE I1ji1 A:12	1.08E-12	705.92	3.17	406	432	21	90	85
5	view	d1twja2	4-ALPHA-GLUCANOTRANSF	2.32E-12	661.58	3.0	406	391	19	83	86
6	view	d1wzaa2	alpha-amylase A1wza A:28	2.9E-12	670.87	3.07	406	409	21	85	84
7	view	PDP:3DHU	Alpha-amylase	1.56E-11	594.12	3.05	406	338	15	79	95

Domain-based Alignment Example

Structure Alignment Results			
Alignment Details:	Query: (orange/dark grey) <i>Cyclomaltodextrin glucanotransferase3bmv A:1-406 c.1.8.1 (Trans)glycosidases</i>	Subject: (cyan/light grey) <i>Alpha-amylase</i>	
P-value: 1.56e-11		PDB ID: 3BMV	
Score: 594.12		SCOP domain: d3bmva4	PDB ID: 3DHU
RMSD: 3.05		Length: 406	PDP: PDP:3DHUAa
%Id: 15.0%		Similarity: 79%	Domain:
		EC number: 2.4.1.19	Length: 338
			Similarity: 95%
			EC number: 3.2.1.1

15% sequence identity



79% and 95% structural overlap

different function:
EC 2.4.1.19 vs. 3.2.1.1

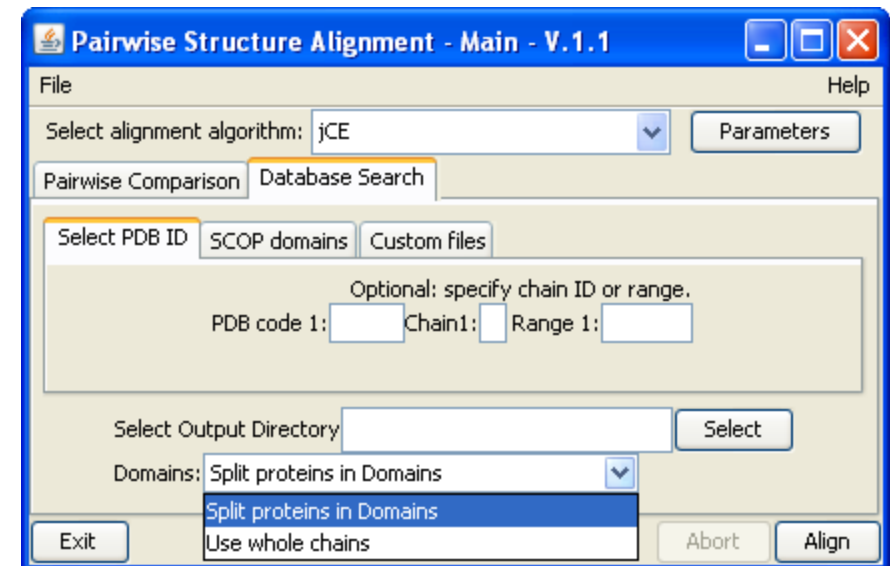
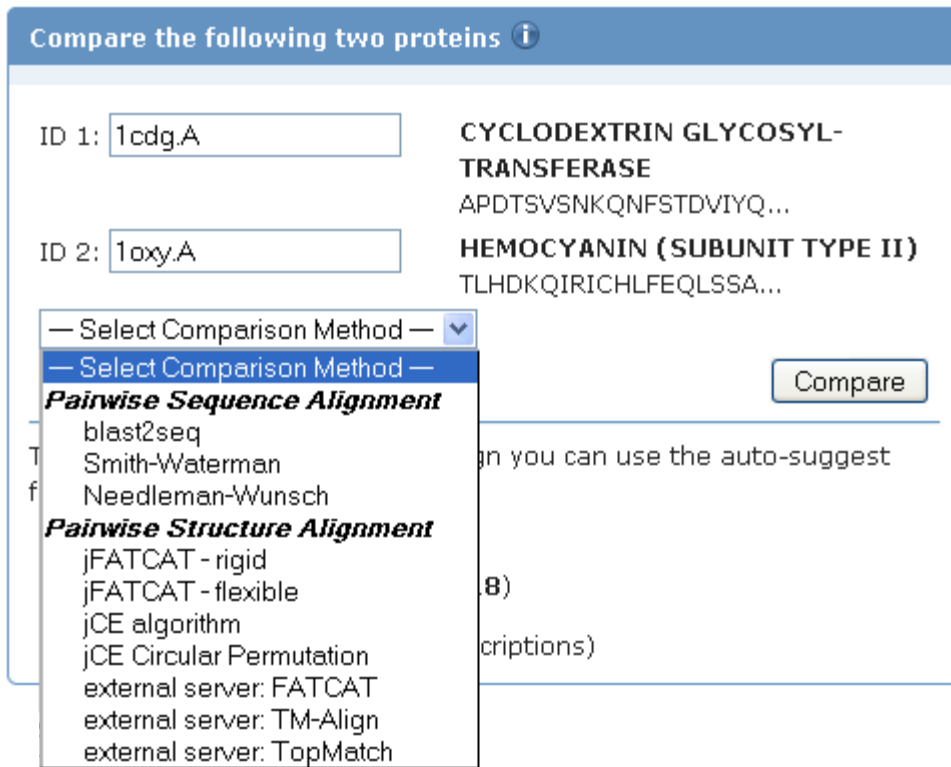
Custom Chain and Domain-based Alignment

Protein Comparison Tool (website)

- Auto-complete for chain and domain IDs

Protein Comparison Tool (download)


- Upload of custom PDB files
- Database search for structural neighbors




New Web Services

- Sequence and structure related annotations
 - Sequence and structure clusters
 - Pfam annotations, updated weekly using HMMer
 - Pre-released sequences (useful for blind predictions)
- General data export
 - Tabular report services (CSV, XML, Excel)
 - > 100 data fields

Custom Report

Click on column headers to sort up/down. Click again to reverse order. Download options:   
Type value in text boxes under column headers to filter the data set. 

PDB ID	Structure Title	Resolution (Å)	R Free 
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
3IP0	Crystal structure of E. coli HPPK in compl	0.89	0.128
1M15	Transition state structure of arginine kin	1.20	0.140
2XBP	A NOVEL SIGNAL TRANSDUCTION PROTEIN	1.20	0.144
3IQU	Crystal Structure of human 14-3-3 sigma i	1.05	0.151
2IIM	SH3 Domain of Human Lck	1.00	0.153



Data Out: Website Usage

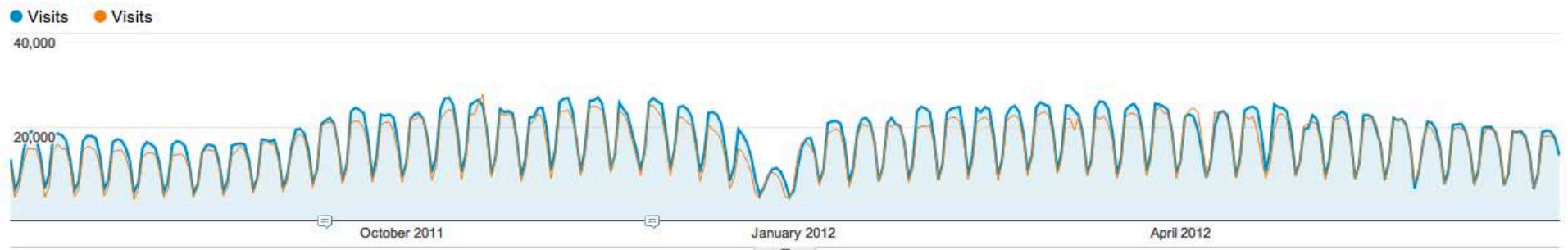


www.rcsb.org: Growth Since Last Year

Increase in Visits: 8.5%

6,316,875 v. 5,820,794

Increase in Unique Visitors: 14%



July 2011- June 2012 in blue

July 2010 - June 2011 in orange

A Live View of the Site: RCSB PDB never sleeps

Overview

Right now
292

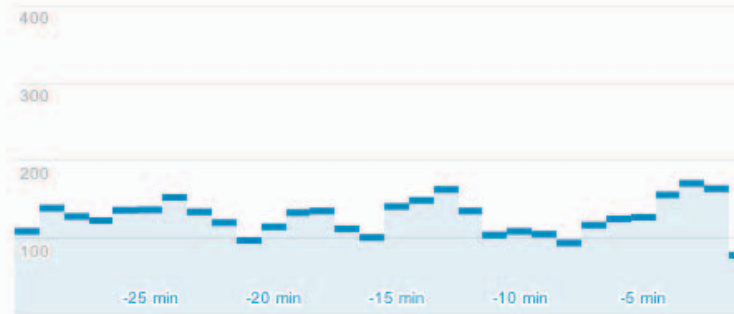
active visitors on site

■ RETURNING ■ NEW



Pageviews

Per minute



Per second



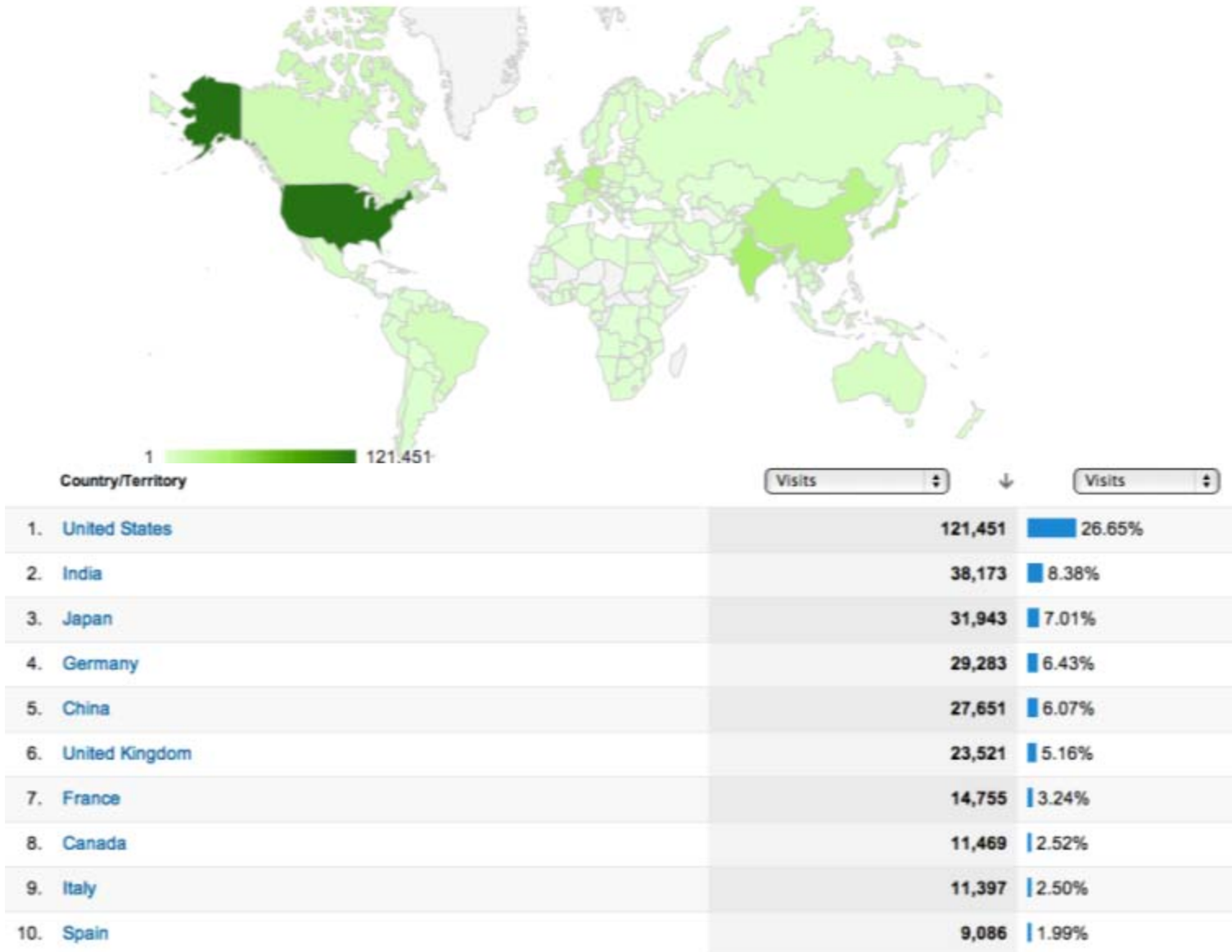
Top Referrals:

	Source	Active Visitors ↓
1.	en.wikipedia.org	8
2.	ncbi.nlm.nih.gov	5
3.	nature.com	3
4.	cbm.msoc.edu	2
5.	moodle3.ehu.es	2
6.	proteinmodelportal.org	2
7.	swissmodel.expasy.org	2
8.	uniprot.org	2
9.	zhanglab.ccmb.med.umich.edu	2
10.	aspergillusgenome.org	1

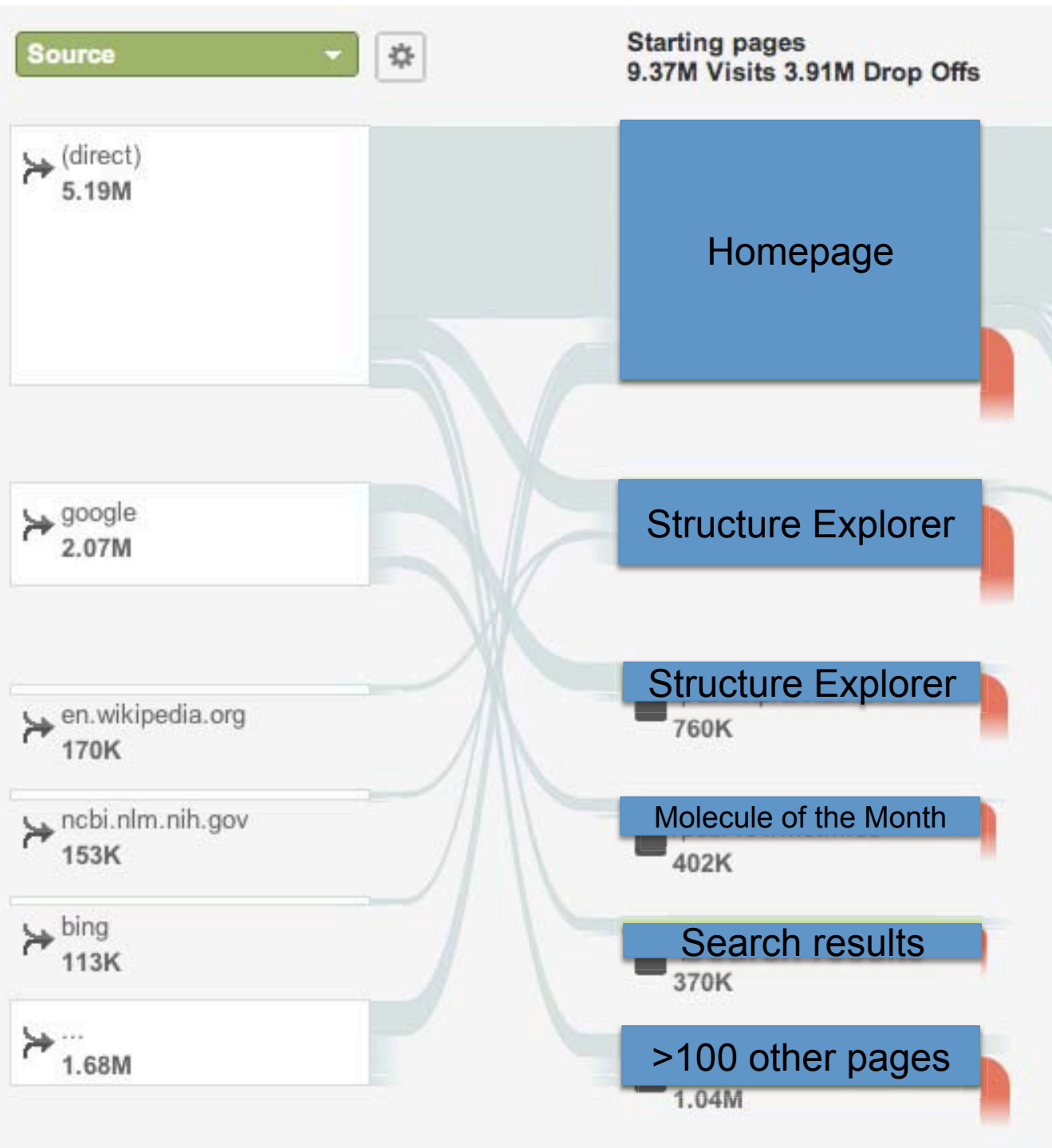
Top Active Pages:

	Active Page	Active Visitors ↓	
1.	/pdb/home/home.do	42	14.38%
2.	/pdb/explore/explore.do?structureId=2H5Z	9	3.08%
3.	/pdb/explore/explore.do?structureId=1KS5	5	1.71%
4.	/pdb/explore/explore.do?structureId=1EMA	3	1.03%
5.	/pdb/explore/explore.do?structureId=2F1C	3	1.03%
6.	/pdb/explore/explore.do?structureId=2h5z	3	1.03%
7.	/pdb/101/motm.do?momID=12	2	0.68%
8.	/pdb/download/downloadApplet.do	2	0.68%
9.	/pdb/explore/explore.do?structureId=1ENY	2	0.68%
10.	/pdb/explore/explore.do?structureId=1FP1	2	0.68%

Website Access by Country



Flow depends on traffic source



Other Key Results

- Education section (PDB-101) growing faster than rest of the site (~18% usage increase)
 - Search engine optimization is important
 - Role of social media is still limited
-



Data Out: Hardware Update



5-Year Hardware Replacement

- New Hardware
 - Deployment on track for November 2012
 - Virtualization of servers offers more flexibility and scalability
- Legacy Hardware (5+ years old)
 - Repurposed for non-production compute-intensive workflows



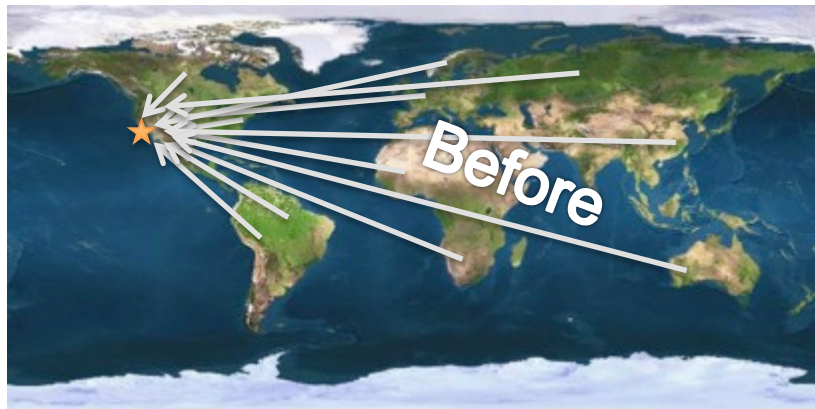
SDSC Rutgers

Old

New

Global Load Balancing

- Why is it important?
 - Enables us to better serve our users by providing increased reliability and quicker results
- How will it be done?
 - By more evenly allocating our resources at Rutgers and UCSD
 - By directing users to the closest site





Data Out: 2012-2013 Plans



Plans: Mobile and Programmatic Access

Mobile access to data and educational materials

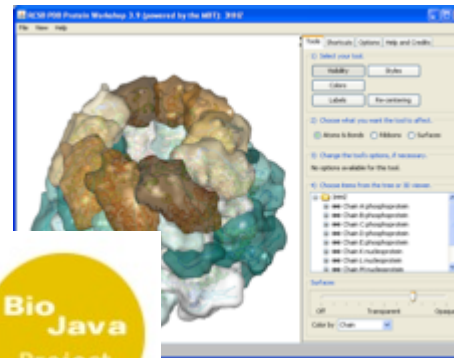


Android support

Bioinformatics Services



Expand RESTful Web Services



APIs and file parsers to access PDB data, including new PDBx format

Plans: Structural View of Biology

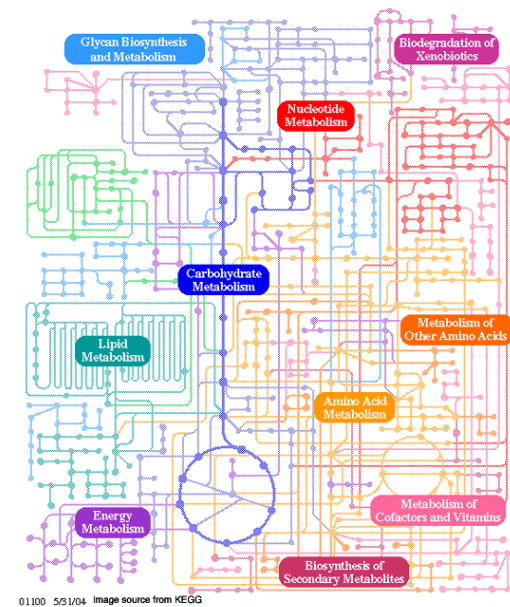
Map structures to:



Drugs and drug targets



Protein families



01100 5/5/04 Image source from KEGG

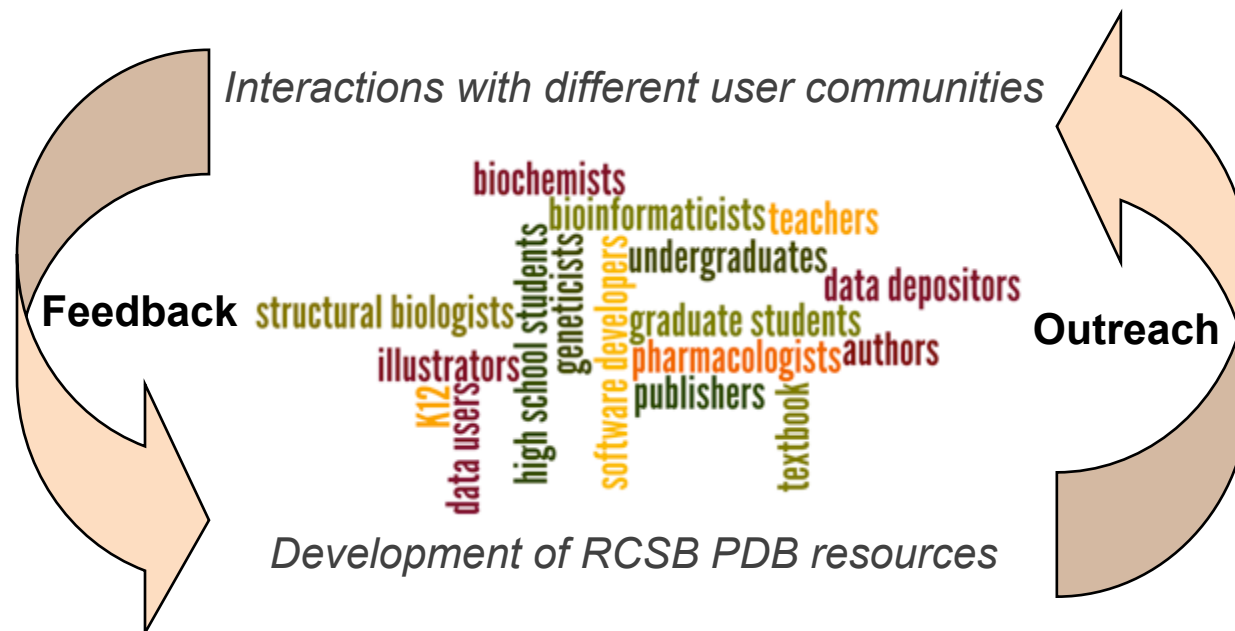
Pathways

Outreach & Education

Christine Zardecki
Shuchismita Dutta

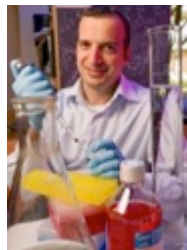
Goals

- RCSB PDB resource should meet its mission in the interest of science, medicine and education
- RCSB PDB is defined by, designed for, and owned by the communities it serves



International User Communities

Who are our users?	What are they using?	How do we know?
Biologists: structural biology, biochemistry, genetics, pharmacology, ...	RCSB PDB website, deposition tools, data	Publication requests, website usage, info@rcsb.org requests, community outreach
Other scientists: bioinformatics, software developers, ...	Web Services, search engines, data	Publications requests, website usage, info@rcsb.org requests, community outreach
Students & teachers	PDB-101	Increase in web hits, email, meeting interactions
Media	Images, data, information	Publications, image requests
General public	Images, Molecule of the Month, information from external media	Concerts, media, Wikipedia





Communicating with Users

Professional Societies



Electronic Support

2012-08-21
Download RCSB PDB Mobile for the iPhone/iPad

Search the PDB, access MyPDB, view molecules in 3D, and more

Questions and Comments

Please use this form to send any questions, bug reports, or suggestions. If you are reporting a bug, please include detailed information about what you were doing, and attach any relevant screenshots, if possible.

First Name*
Last Name*
Email Address*
Confirm Email*
Subject*
Category*
Comments*

Attachments: Attach another file
MSUZ
Can't load this? Load New Image
Verification Code*
*Security required first

Local Festivals



CIPR Seminars



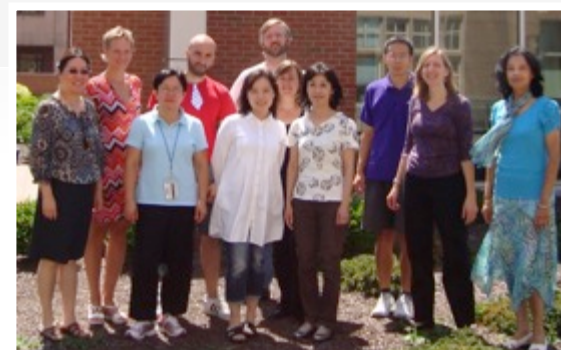
Educational Materials

TOLL-LIKE RECEPTORS
Sentinels for the immune system

PDB ANNUAL REPORT 2011

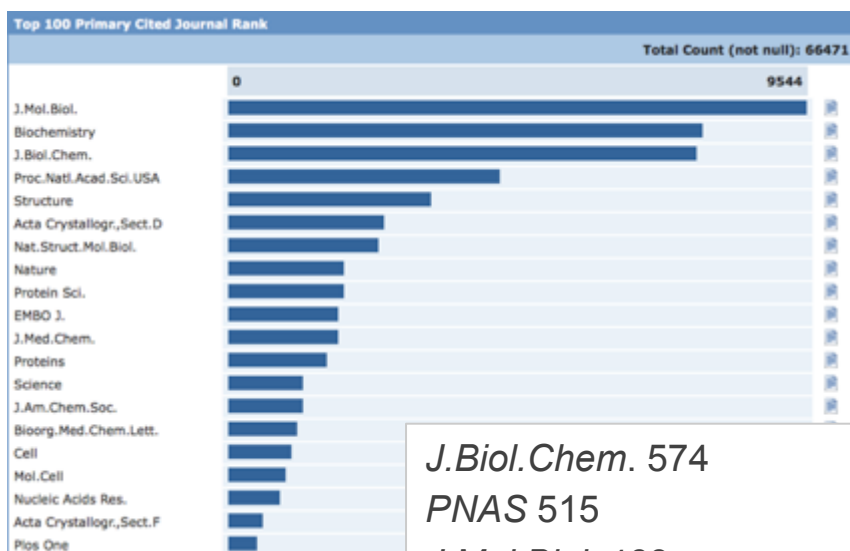
PDB NEWSLETTER
Spring 2011 - Number 51

Staff Activities



Current Journal Interactions

Top Overall



Top in 2011

J. Biol. Chem. 574
PNAS 515
J. Mol. Biol. 463
Biochemistry 402
J. Med. Chem. 220
Bioorg. Med. Chem. Lett. 201
Structure 192
Nat. Struct. Mol. Biol. 183
Nature 163
J. Am. Chem. Soc. 153

Validation Reports

IUCr
Journals

jbc THE JOURNAL OF
BIOLOGICAL CHEMISTRY

Future:  eLIFE

Notification of Structure Publications

- Ongoing: *Acta D&F*, *FEBS J.*, *Journal of Biological Chemistry*, *Journal of Molecular Biology*, *Nature*, *NSB*, *NCB*, *Nucleic Acids Research*, *Proteins*, *PNAS*
- New: *Nature Comm.*, *Science*

Recent Educational Activities

RCSB PDB-101

Virus Structures

Viruses attack cells and force them to make many new viruses, often killing the cell in the process. Some viruses are very simple, such as the protein capsids that surround a short strand of RNA or DNA. More complex viruses inject the cell with a viral genome capable of encoding proteins to fight the cell's defenses.

Explore further at:
www.rcsb.org
www.emdatabank.org

Virus Sizes
 Most viruses are much smaller than cells—the ones shown here are all drawn at approximately 900,000x magnification, ranging from less than 30 nanometers to over 500 nanometers in diameter (1 nanometer = 1 billionth of a meter).

Dengue fever virus
 Dengue fever virus usually causes flu-like symptoms, but the infection can be deadly in some cases. PDB ID: 1bdt

Rhinovirus
 Rhinovirus is one of the causes of the common cold. PDB ID: 1bta

Foot and mouth disease
 Foot and mouth disease is a serious problem that causes high fever and blisters in livestock. PDB ID: 1bba

Feline distemper
 Vaccination can prevent this life-threatening infection in cats. PDB ID: 1bfc

Tobacco mosaic virus
 Research on TMV, the first virus to be discovered, began late in the 19th century. It is so stable that it can survive for years in pipes and cigarettes made from infected leaves. PDB ID: 3j0c

Mimivirus
 The blue background figure is the mimivirus, one of the largest viruses. Mimivirus may be linked to some forms of paramecia. EMD-5039

Bacteriophage T4
 This complex virus injects its DNA genome into bacteria through the long tube at the bottom. EMD-1414 EMD-1126

Virus Shapes	Polyhedral viruses ▲	Helical viruses ▲	Complex viruses ▲	Enveloped viruses (not shown)
	These viruses are composed of polyhedral protein shells. They are also called icosahedral viruses because of their symmetry.	In these viruses, the nucleic acid genome is wound inside a cylindrical protein capsid with helical symmetry.	These viruses are composed of many different proteins that work together to protect the genome, attach to cells, and inject the nucleic acid inside.	Viruses such as influenza and HIV are surrounded by a membrane that includes glycoproteins that seek out cells to infect.

rCSB pdb
 @buildmodels
 Education/outreach for the RCSB PDB (www.rcsb.org), interesting structural biology links, tips and info on the Science Olympiad's protein modeling event.
 Piscataway, NJ - <http://education.pdb.org>

rCSB pdb @buildmodels
 Fold DNA origami using the PDF template at rcsb.org/pdb/education... #DNADay pic.twitter.com/ffcfrgxH
 Hide photo



DNA DAY 12 APRIL 20

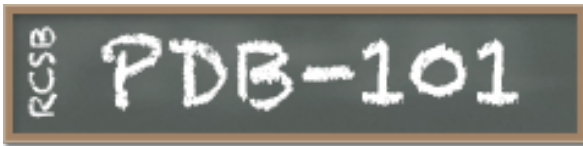
powered by Photobucket

Flag this media

13 RETWEETS 2 FAVORITES

1:52 PM - 20 Apr 12 via web · Details

Reply Delete Favorite



RCSB PDB-101 A MEMBER OF THE PDB PROTEIN DATA BANK An Educational Resource for Exploring a Structural View of Biology


Contact Us | Print Jump to a Molecule: Choose a molecule from this list

PDB-101 Features - Structural View of Biology

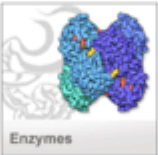
Structural View of Biology

List View of Archive By: Title | Date | Category


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. [\[more\]](#)



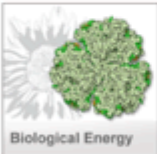
Protein Synthesis



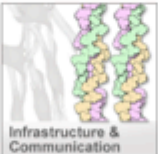
Enzymes




Health & Disease



Biological Energy



Infrastructure & Communication



Biotechnology & Nanotechnology

Please see our [usage policies](#) for citation and reprint information. Copies of the illustrations used in these features are available for educational use. Please note that the structures used to illustrate each installment are chosen at the discretion of the authors; the process behind the [creation of this feature](#) is described by the author.

The RCSB PDB is managed by two members of the Rutgers and UCSF, and is funded by NSF, NIGMS, DOE, NLM, NCI,

PDB-101 Features

- Structural View of Biology
- Educational Resources
- Molecule of the Month
- Understanding PDB Data
- Author Profiles

- Increase in traffic
 - June 2012 stats 18% higher than June 2011
- Home page redesigned to improve appearance on different browsers

Author Profiles

RCSB PDB-101



profile

Yonath, A. 36 Structures (18 Unique) by this author

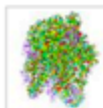
This timeline displays the structures that list Yonath, A. as a Deposition or Primary Citation Author. Structures are ordered by the deposition date shown, and link to the Structure Summary page for the entry. The first instance of a protein or protein complex deposited by this author is displayed larger than future examples of similar molecules. ?

2010

3PIP

November 07

Molecule: 50S ribosomal protein L11 ...

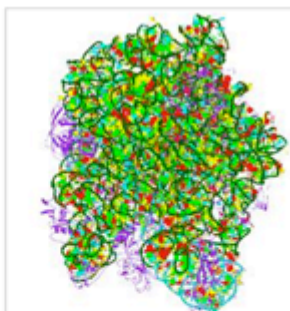


Crystal structure of the synergistic antibiotic pair lankamycin and lankacidin I...

3PIO

November 07

50S ribosomal protein L11 ...



Crystal structure of the synergistic antibiotic pair lankamycin and lankacidin I...

Sort by

Newest
Oldest

Year

2011-2010
2010 (2)
2010-2005
2005-2000
2000-1995
1995-1990
1990-1985
1985-1980
1980-1977

PDB-101 Features ▾

Structural View of Biology

Educational Resources

Molecule of the Month

Understanding PDB Data

Author Profiles



profile

Joint Center for Structural Genomics (JCSG)

1326 Structures (643 Unique) by this author

This timeline displays the structures that list Joint Center for Structural Genomics (JCSG) as a Deposition or Primary Citation Author. Structures are ordered by the deposition date shown, and link to the Structure Summary page for the entry. The first instance of a protein or protein complex deposited by this author is displayed larger than future examples of similar molecules. ?

2012

4GL3

August 13

Molecule: Putative uncharacterized prote...



Crystal structure of a putative glucosylase (BACUNI_03963) from Bacteroides uni...

4GDZ

August 01

Molecule: Putative uncharacterized prote...



Crystal structure of a hypothetical protein (BACEGG_02002) from Bacteroides egge...

4GHB

August 07

Molecule: Putative uncharacterized prote...



Crystal structure of a hypothetical protein (BACUNI_01323) from Bacteroides unif...

4GCM

July 30

Thioredoxin reductase



Crystal structure of a thioredoxin reductase (trxB) from Staphylococcus aureus ...

Sort by

Newest
Oldest

Year

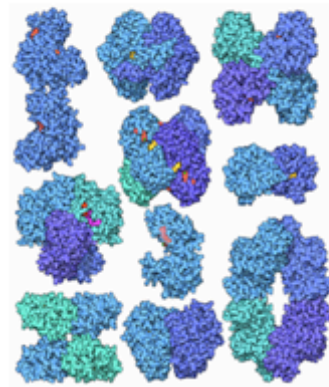
2013-2010
2012 (67)
2011 (102)
2010 (183)
2010-2005
2005-2002

Molecule of the Month

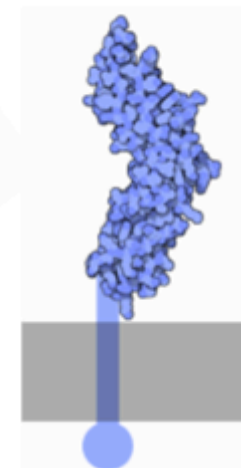
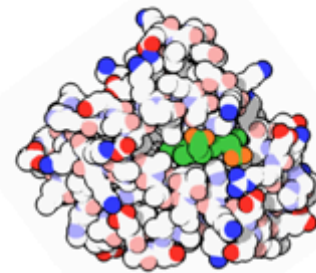
- 150th article published June 2012
- Top accessed articles July 2011–June 2012:

Title	Page views
Hemoglobin	34081
Collagen	21937
GFP	21709
Catalase	19207
Caspases	17755
Alpha-amylase	16597
DNA Polymerase	16392
Insulin	15909
Alcohol Dehydrogenase	15853
Lysozyme	14890

50: Glycolytic Enzymes



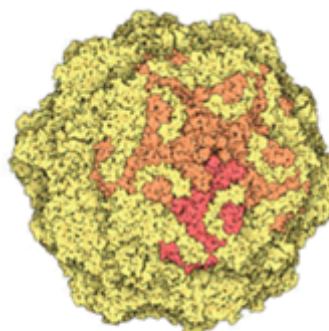
1: Myoglobin



75: Tissue Factor



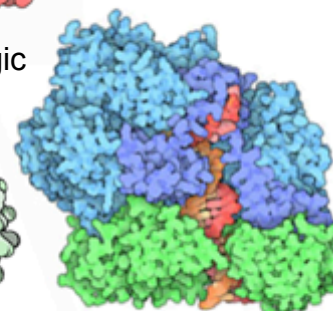
100: Adrenergic Receptors



125: Parvoviruses



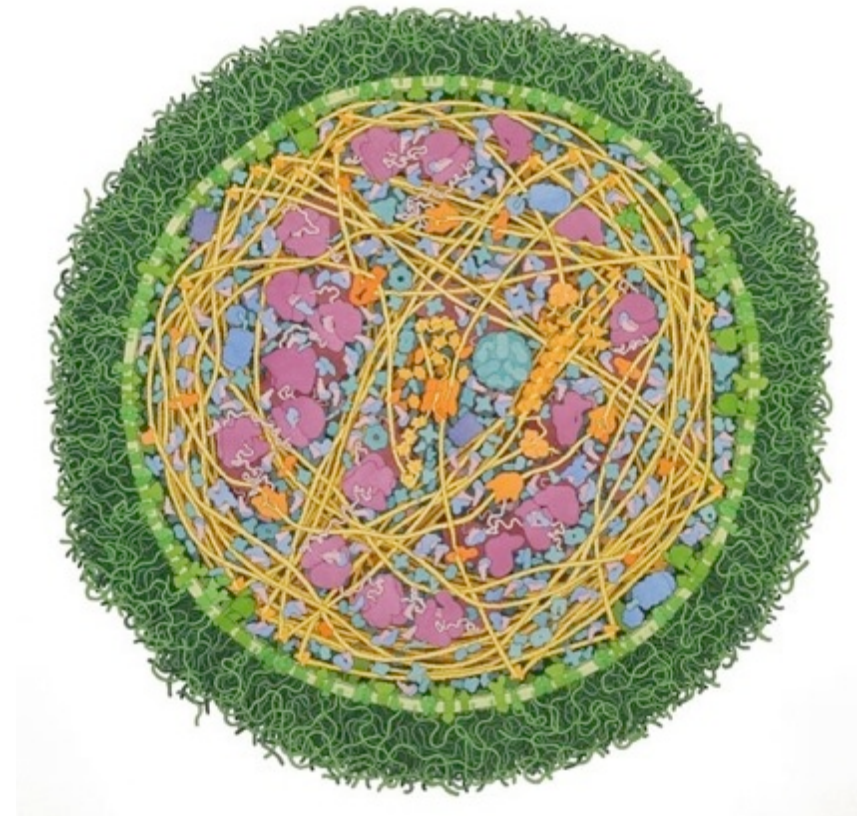
25: Thrombin



150: Sliding Clamps

Future

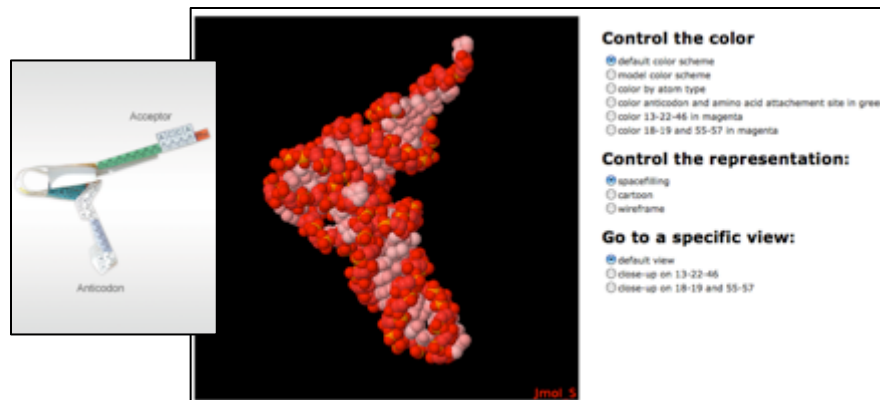
- Improved access and display of different PDB-101 sections
- Redesign of *Molecule of the Month* pages to improve readability
- Creation of cell-based poster and related online display
- 2013 look at *Molecular Machinery*



Mycoplasma cell

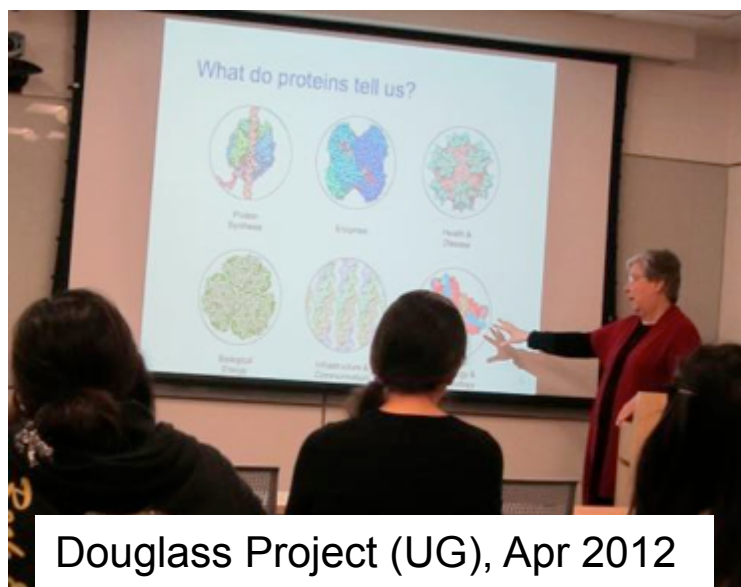
Teaching and Learning: A Structural View of Biology

- Booths, demonstrations at Society meetings
- Online resources



NJSC 2012,
Oct. 9-10

- Presentations



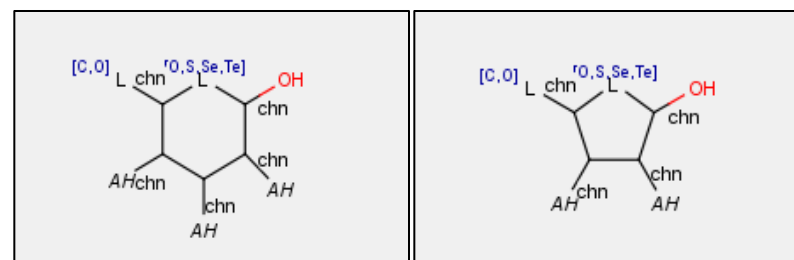
Douglass Project (UG), Apr 2012



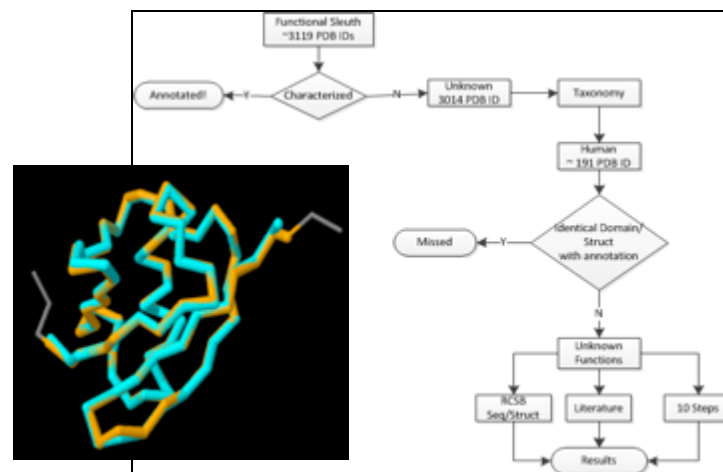
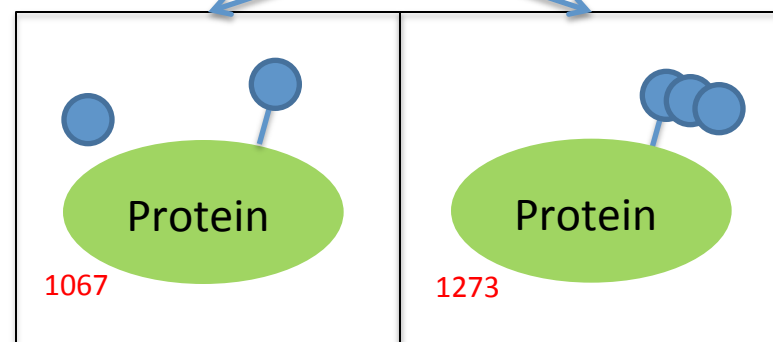
Waksman Scholars (HS), Jul 2012

Summer Programs 2012

- Carbohydrates:
 - Systematic Identification, and Analysis of Carbohydrates in the PDB
 - Amanda Ullmer and Mahd Nishtar
- Use of various open source software libraries and tools
 - Developing examples
 - Forrest Price
- Functional Sleuth
 - Annotation of human proteins
 - Swagata Das



Carbohydrate PDB entries
Links Analysis

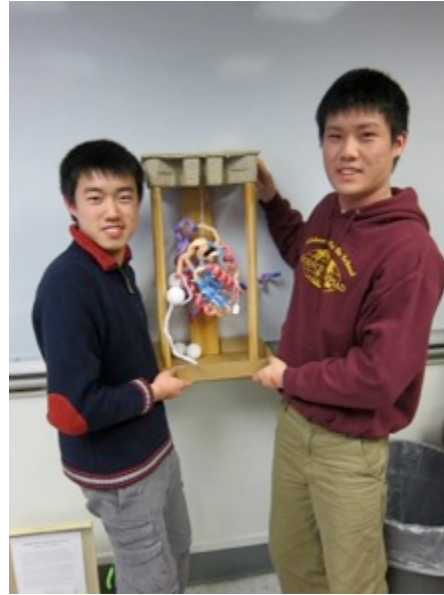
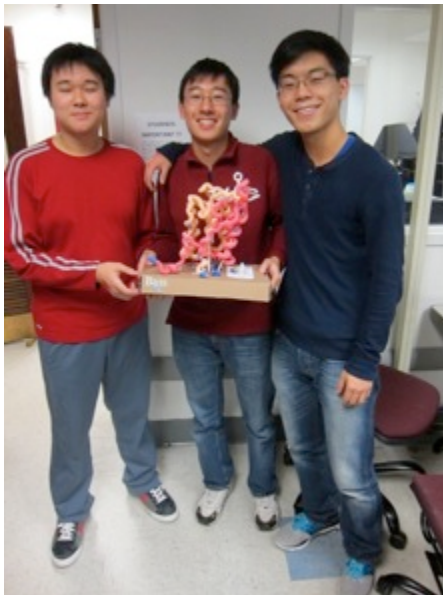


Protein Modeling at the 2012 Science Olympiad



National Champions:
New Trier High School (IL)

- New Jersey
 - 3 regionals
 - State championships
- California
 - Judged in San Diego
 - Hosted workshops



Event now on hiatus for 2 years

New Audiences: K-12 Students

- Take 5 to Visualize
 - Pilot workshops in Jan 2012
 - Princeton HS
- 4H-sponsored summer program
 - 34 students visiting Rutgers
 - From inner city schools in 7 NJ counties



New Audiences: K-12 Teachers

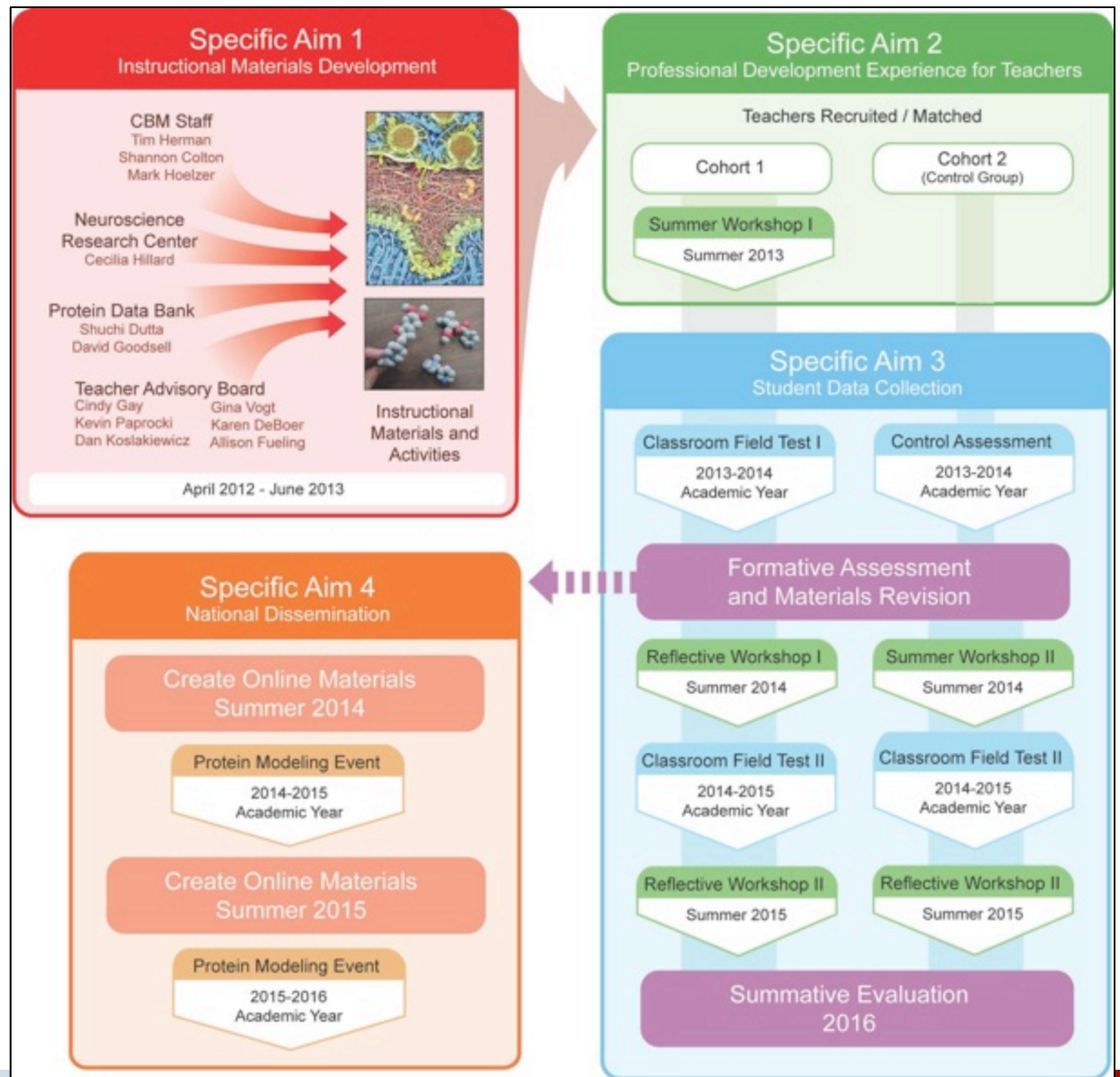
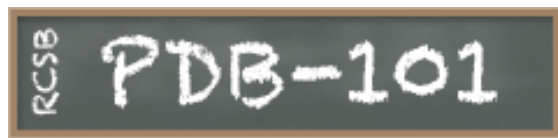
- Princeton Molecular Biology Outreach Program
 - July 16, 2012
 - 19 teachers from all over nation
 - >50% interested in additional workshop
- Working with NJ HS teachers
 - Proposed workshop (October 2012)
 - Contacted >25 Central NJ science coordinators/teachers
- Working with DC HS teachers
 - Proposed workshop pending approval from Chancellor of DC public schools
 - Collaboration with Edvotek



Drugs, Drug Targets and You: Exploring the Molecular Mechanisms of Drug Abuse with Tactile Teaching Tools



- Science Education Drug Abuse Partnership Award
- Funded, Aug 2012

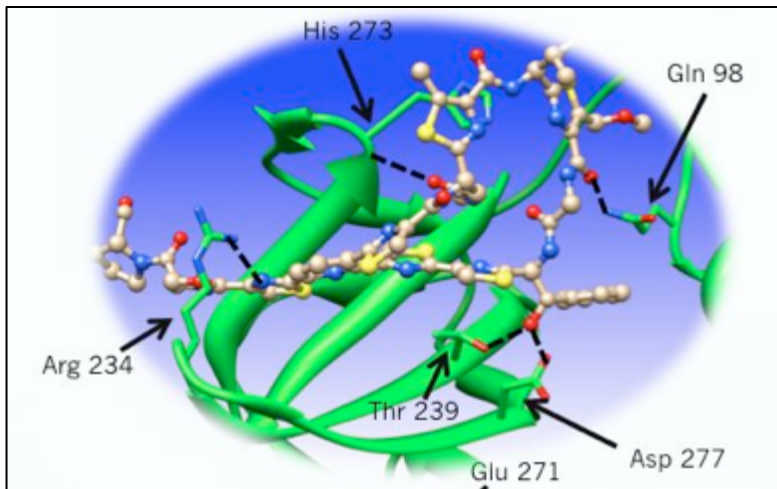


Creating UG Research Opportunities

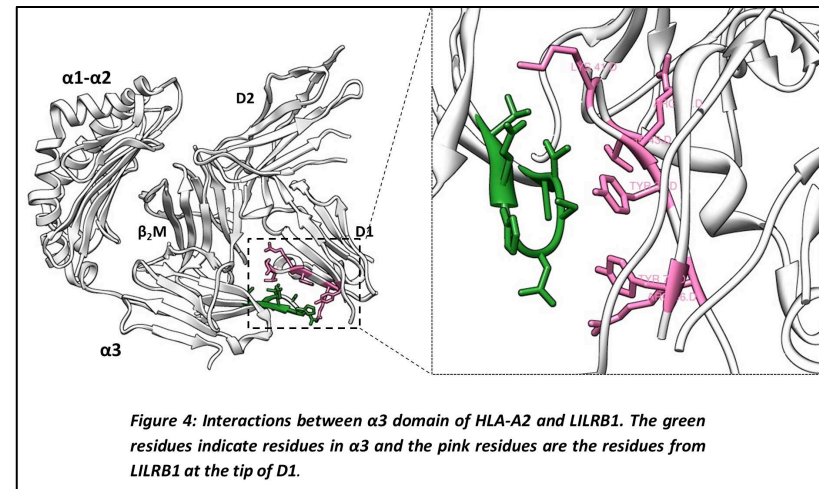
- Capstone Research
 - Structure and Interactions of Aminoglycosides
 - Anna Carleen
- Douglass Chemistry Scholars Project
 - Designed to engage, retain students in Chemistry related subjects
 - Proposed new course - Introduction to Chemistry Research
 - At least 2 research experiences along with various workshops and networking opportunities
 - Submitted to NSF, August 2012



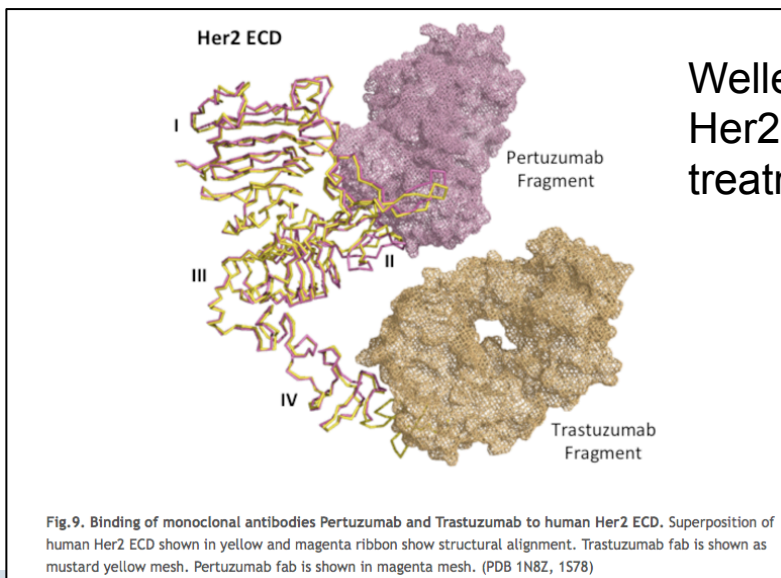
Experiments with MAP-based courses



Rutgers student report on GE2770 and eFTu complex, Spring 2012



Georgetown University student report on a SNP related complex, Spring 2012



Wellesley College student report on Her2 based Breast Cancer treatments, Spring 2012

Spring 2013
Rutgers University: Immune System
Wellesley College: Biochemistry

Proposed Collaborative Summer Programs

- Inspired by MAP based courses
- Proposal submission in 2011 received good reviews but not funded
- Proposal to be re-submitted to NIH Sept. 2012
- Earliest date of offering Summer 2013

