RCSB Protein Data Bank
Advisory Committee Meeting

September 14, 2012

RCSB PDB
PROTEIN DATA BANK

Overview

Helen Berman

When Last We Met…

257 participants
35 travel awards
93 posters

PDB40 Speakers

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

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

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

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.

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 crecimiento

CELE ABAD-ZAPATERO | 20 DIC 2011 - 13:19 CET

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

El cristalógrafo Michael G. Rossmann (en el centro) 40 años después de inicio al Banco de Datos de Proteínas, celebra el aniversario en Cold Spring Harbor. / CONTANCE BUNDSCHUH

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

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

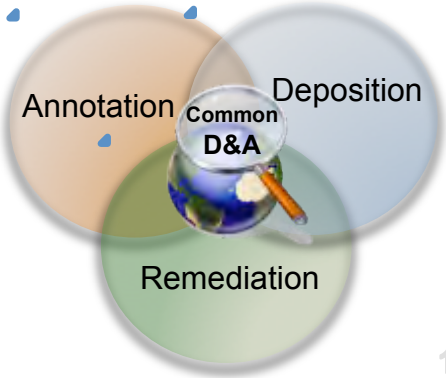
Vision

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

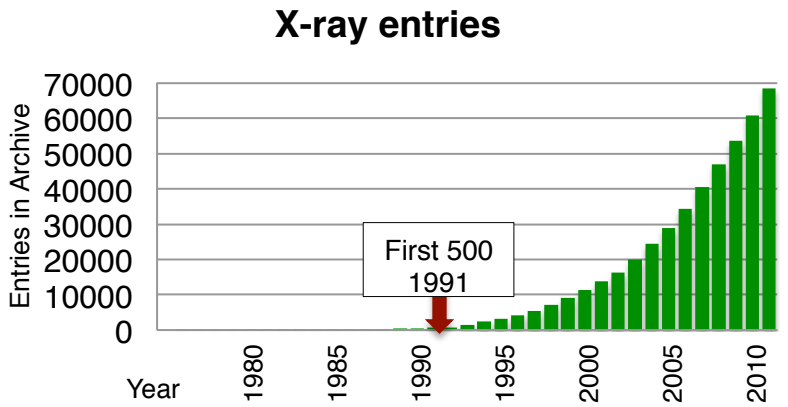


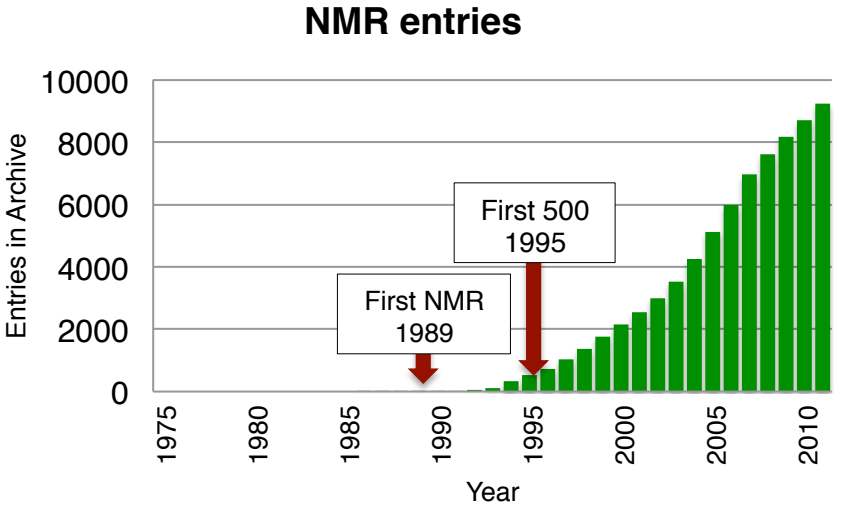
Depositions

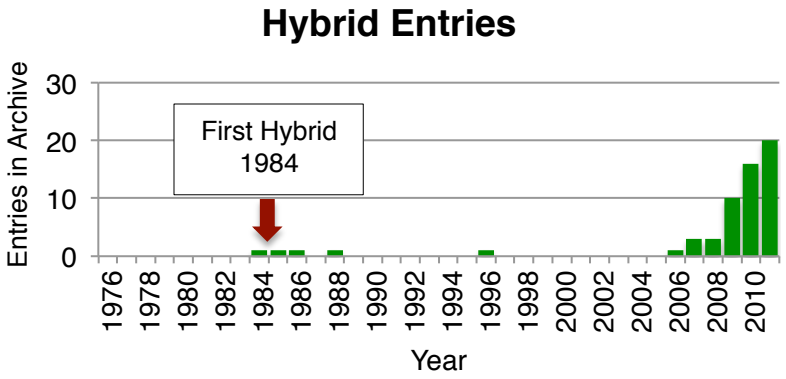
Year	Total Depositions	Deposited To			Processed By		
		RCSB PDB	PDBj	PDBe	RCSB PDB	PDBj	PDBe
2000	2983	2445	10	528	2297	158	528
2001	3287	2673	118	496	2408	383	496
2002	3565	2769	289	507	2401	657	507
2003	4830	3488	673	669	3135	1026	669
2004	5508	3796	900	812	3082	1614	812
2005	6678	4507	1166	1005	3563	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

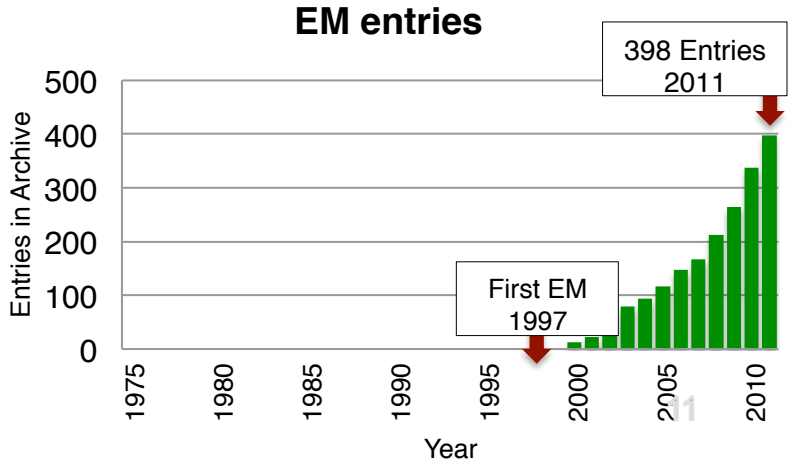
Last Updated: 5 Sep 2012

9965 depositions predicted for 2012

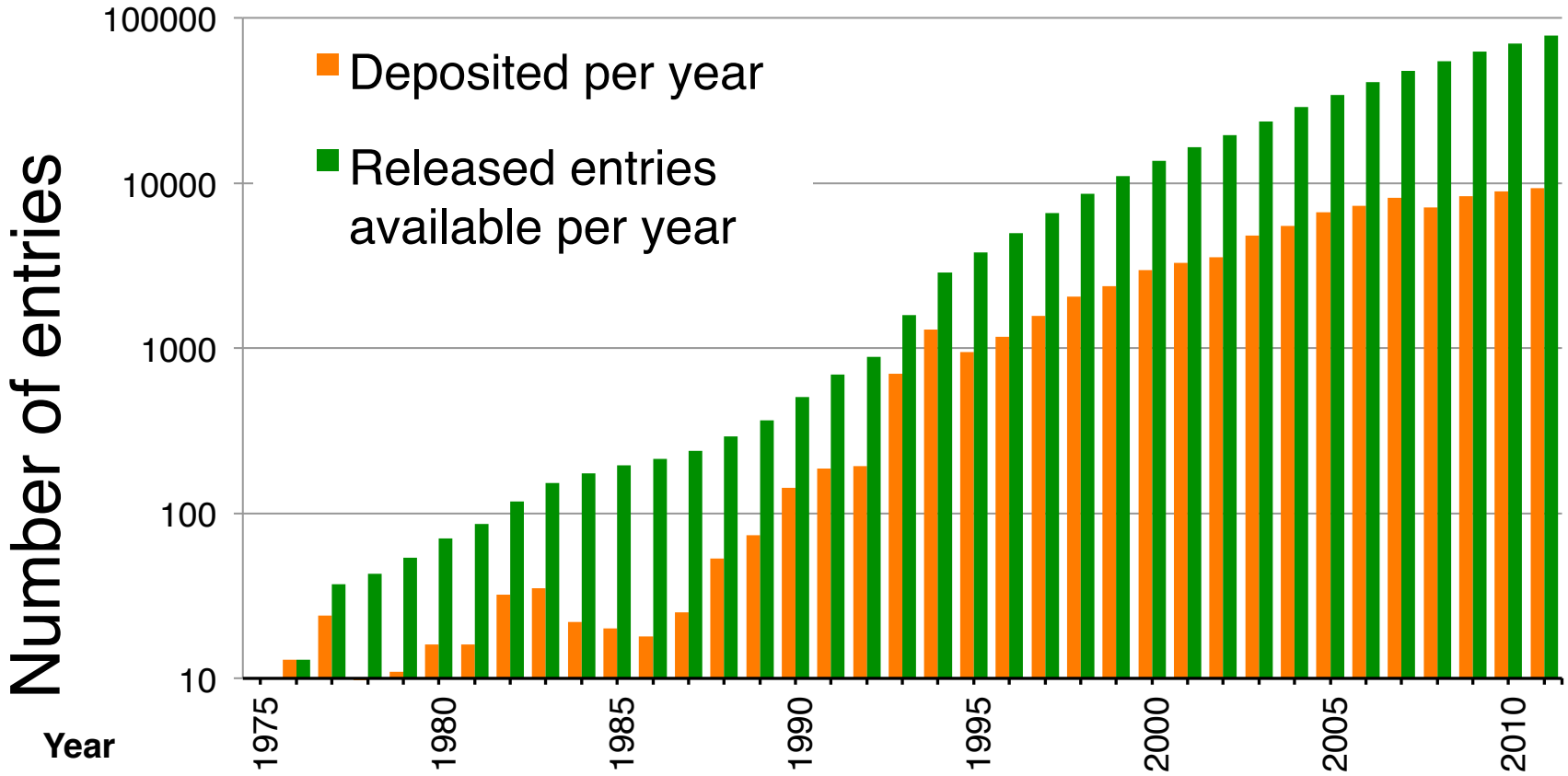








Entries per year (Logarithmic Scale)

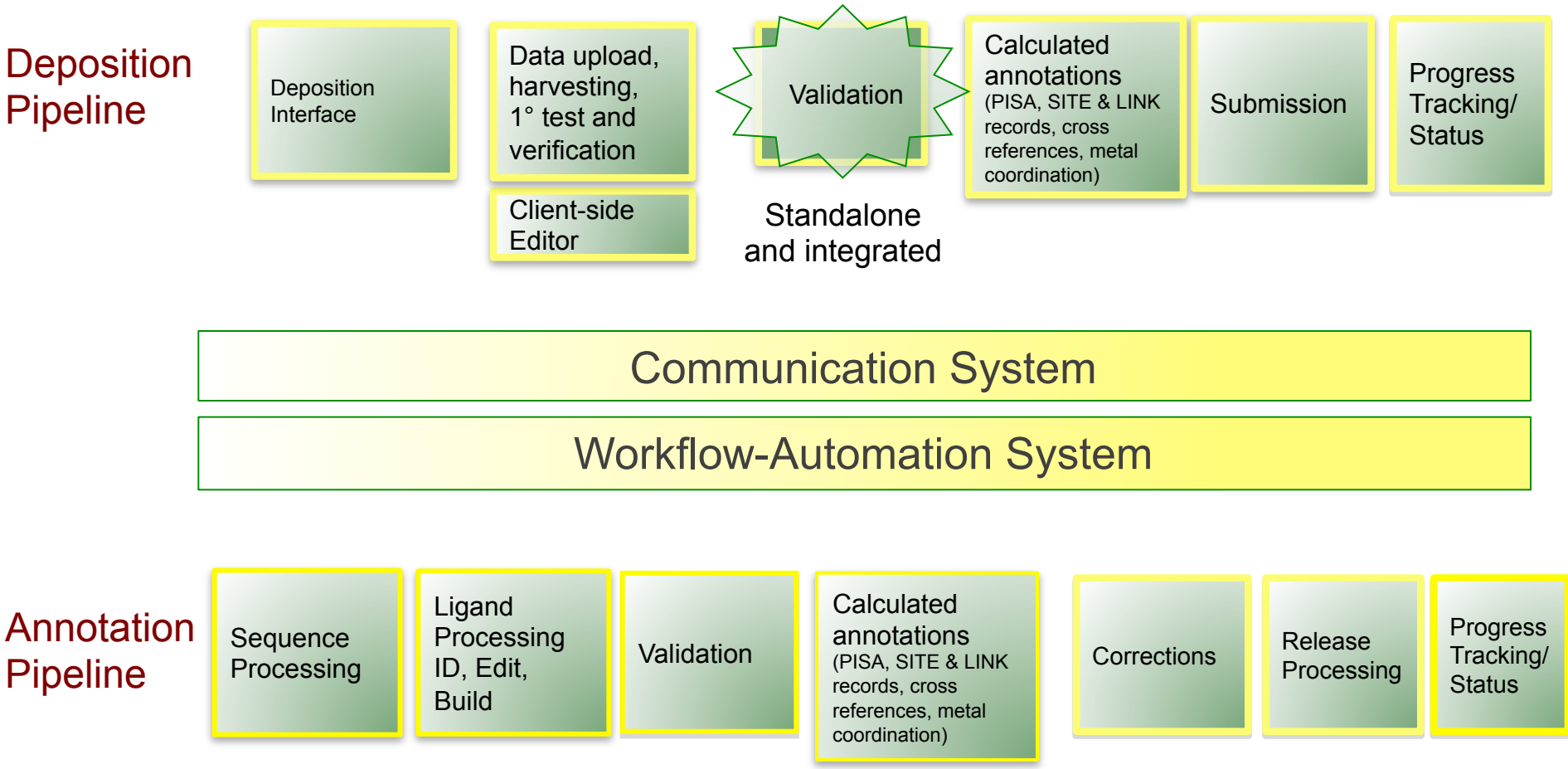


Common Deposition and Annotation Tool

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



Task Forces

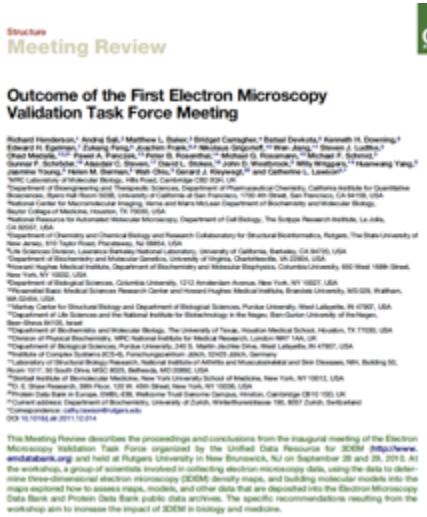
To collect recommendations and develop consensus on method-specific issues, including validation checks that should be performed and identification of validation software applications.

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


Structure
Meeting Review

Outcome of the First Electron Microscopy Validation Task Force Meeting

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

July 2012

3DEM VTF Recommendations

Structure
Meeting Review

Outcome of the First Electron Microscopy Validation Task Force Meeting

(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

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

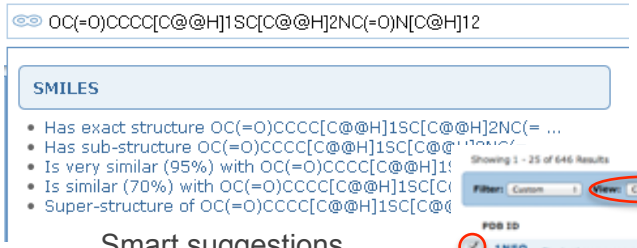
- Assess models using X-ray VTF-recommended criteria
- Develop criteria for assessing model fit to map
- Annotate sequences of all map components
- Capability to archive coarse-grained model representations

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

Data Out

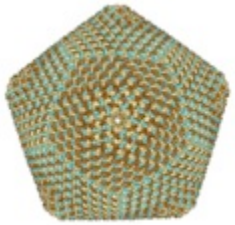
- Improved searching and browsing of results
- Visualization of molecular surfaces
- Improved sequence annotations
- Domain-based structural alignments
- New Web Services

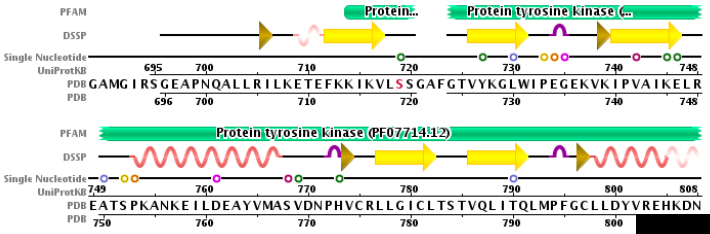


Smart suggestions

Condensed query results view (Fall 2012)

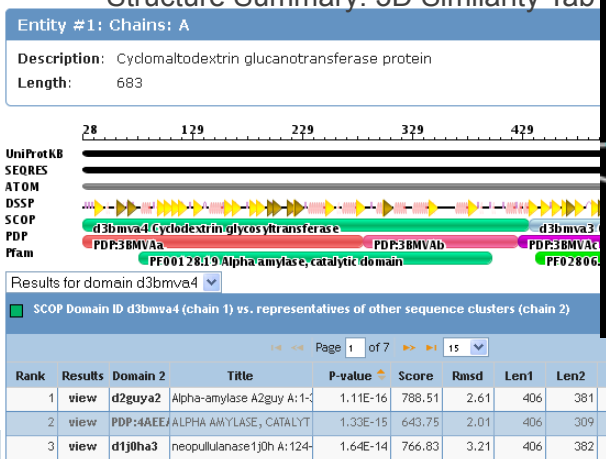
Protein Workshop

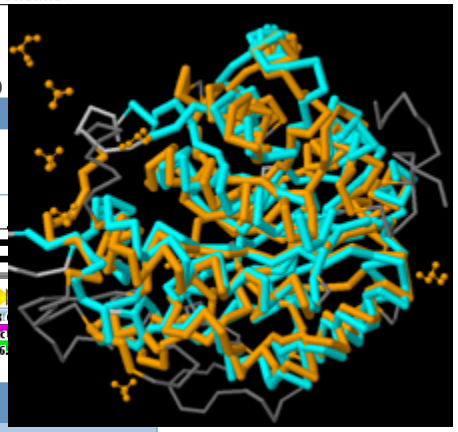




Structure Summary: Sequence Tab

Structure Summary: 3D Similarity Tab





Rank	Results	Domain 2	Title	P-value	Score	Rmsd	Len1	Len2	%ID	%Cov1	%Cov2
1	view	d2guya2	Alpha-amylase A2guy A:1-	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

RCSB PDB *Mobile*

Available on the App Store

- 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

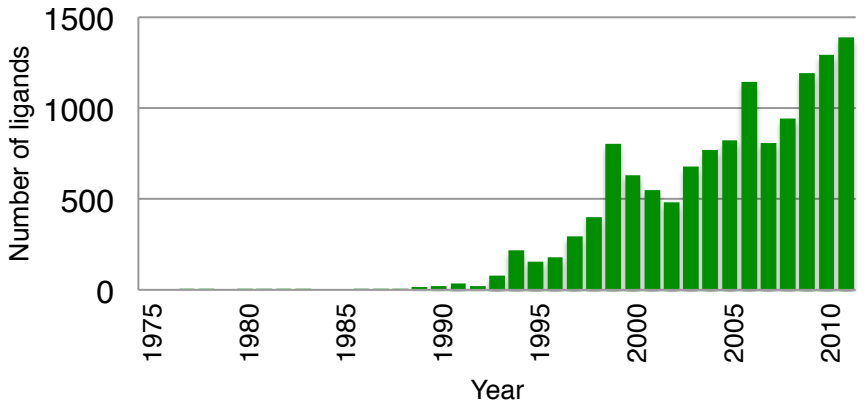
Structure
Meeting Review

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

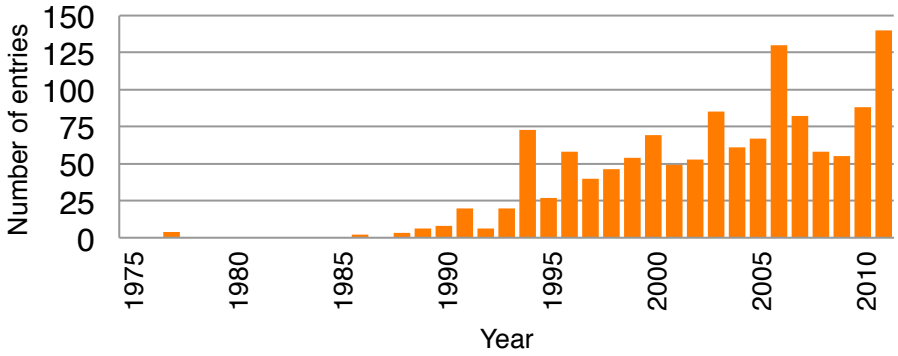
PDB past and present at PDB40

Increasing Complexity

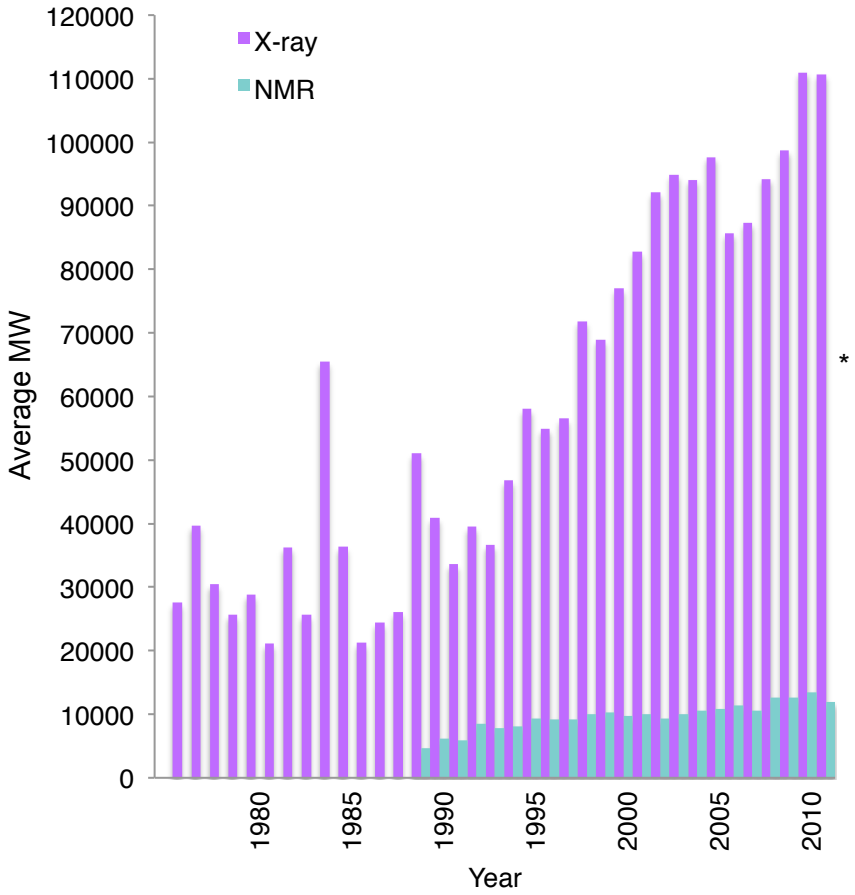
Number of ligands released per year



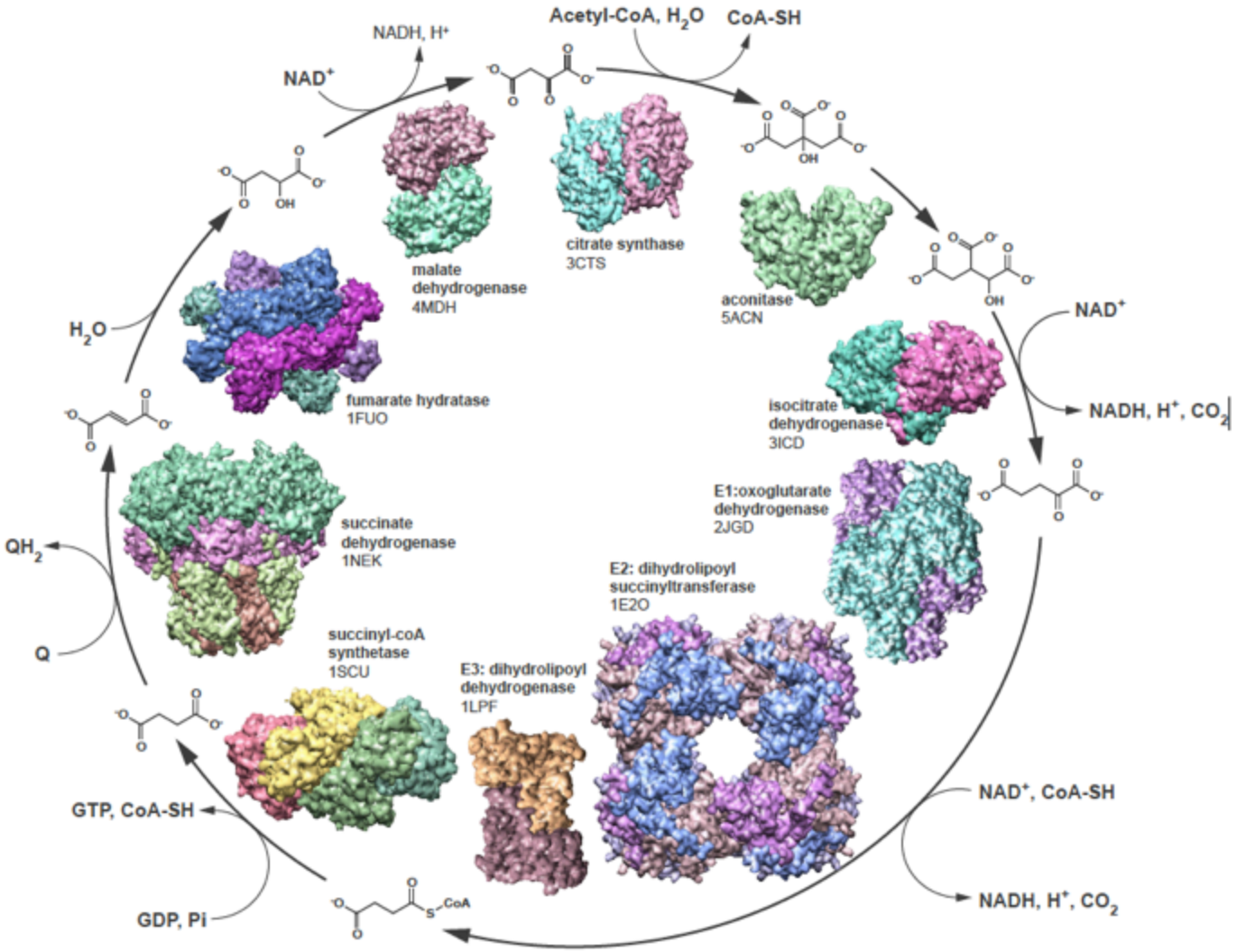
Number of entries with peptide-like inhibitors/antibiotics released per year



Average Molecular Weight for Entries 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)

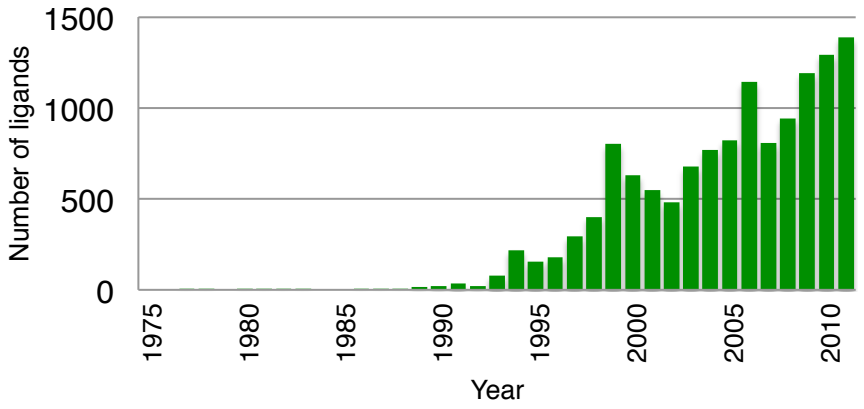
Year	Total Depositions	Deposited To			Processed By		
		RCSB PDB	PDBj	PDBe	RCSB PDB	PDBj	PDBe
2000	2983	2445	10	528	2297	158	528
2001	3287	2673	118	496	2408	383	496
2002	3565	2769	289	507	2401	657	507
2003	4830	3488	673	669	3135	1026	669
2004	5508	3796	900	812	3082	1614	812
2005	6678	4507	1166	1005	3563	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

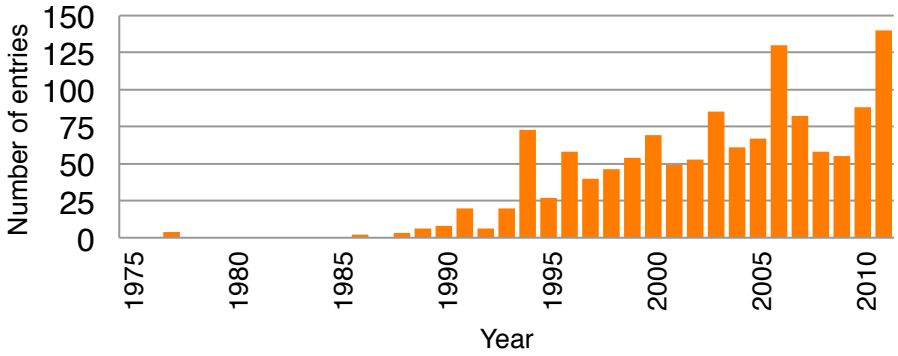
***9956 depositions projected for 2012**

Increasing Complexity

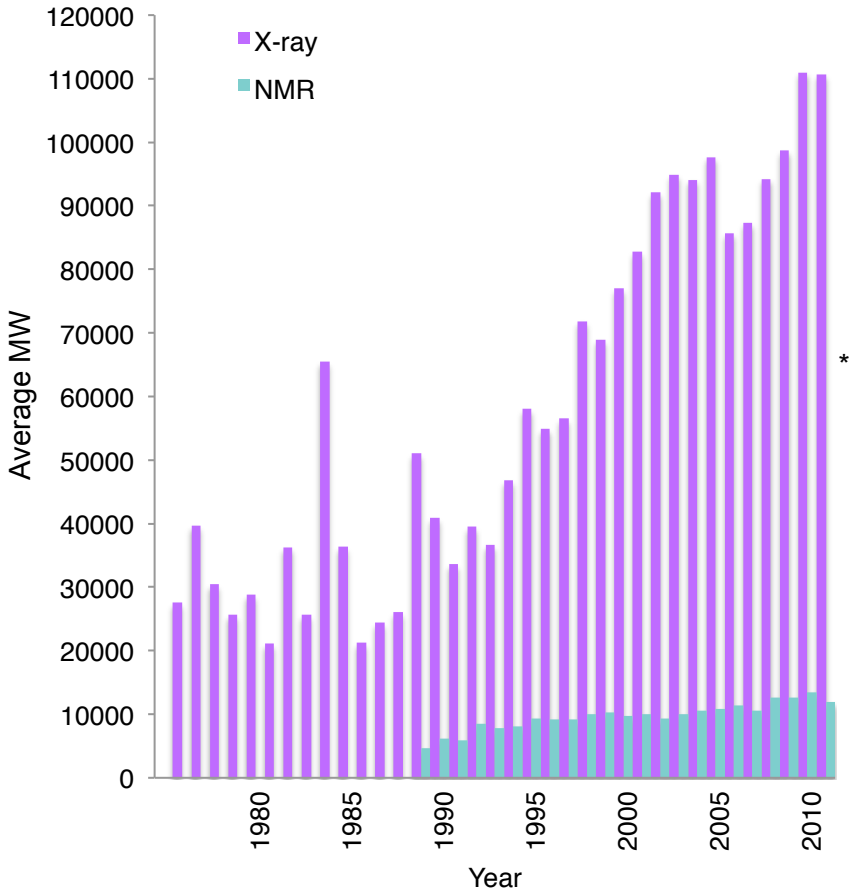
Number of ligands released per year



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



Average Molecular Weight for Entries 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

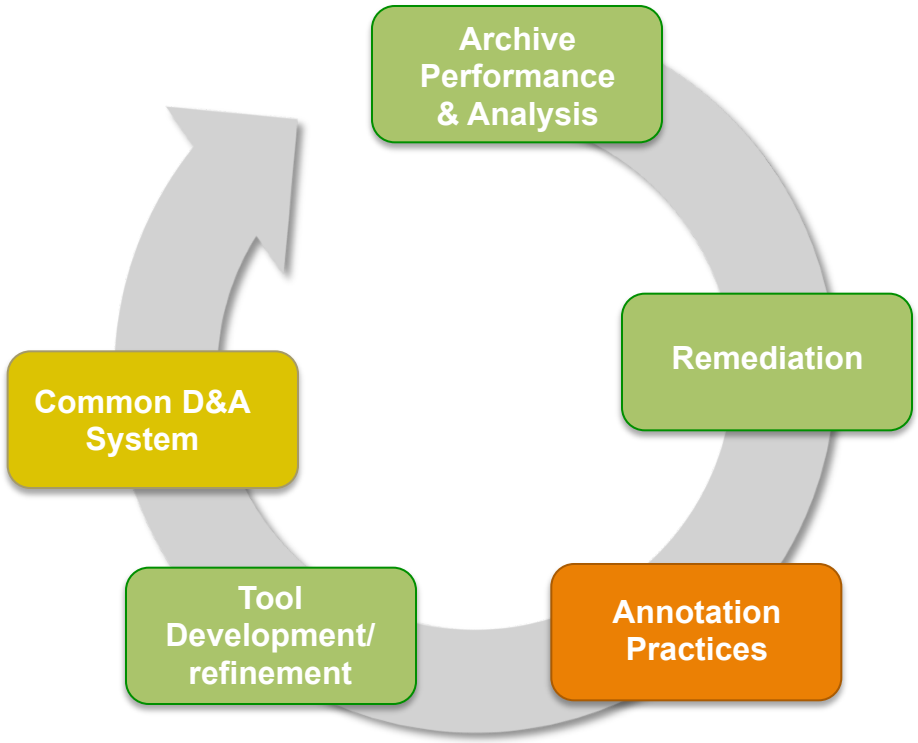
- Annotation pipeline integrated with D&A Ligand Module
- Chemical component dictionary and searching tools
- Infrastructure for small molecules & peptides
  - *Biologically Interesting molecule Reference Dictionary* (BIRD) and searching tools
  - Annotation system

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

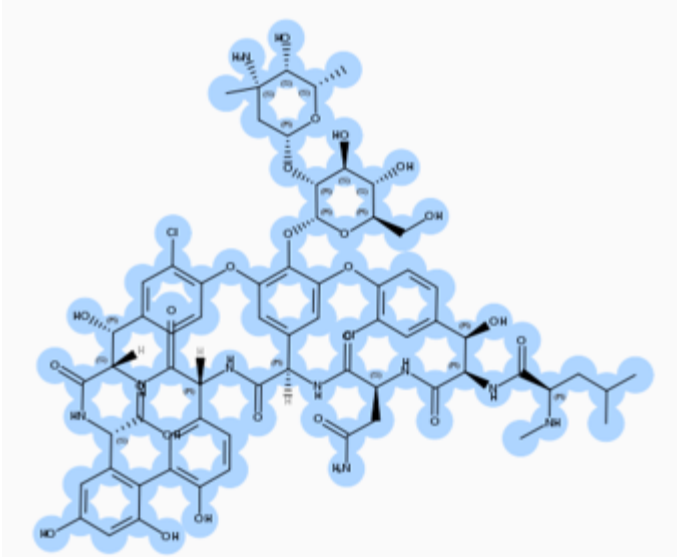
WORLDWIDE PDB
PROTEIN DATA BANK

Better Annotation of Biologically Interesting Molecules

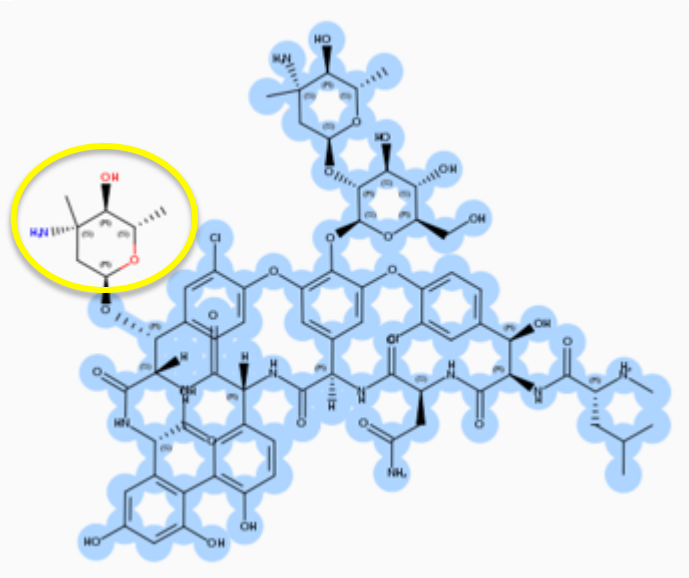
2011 remediation of inhibitors and antibiotics informed the development of an annotation system that supports

- 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

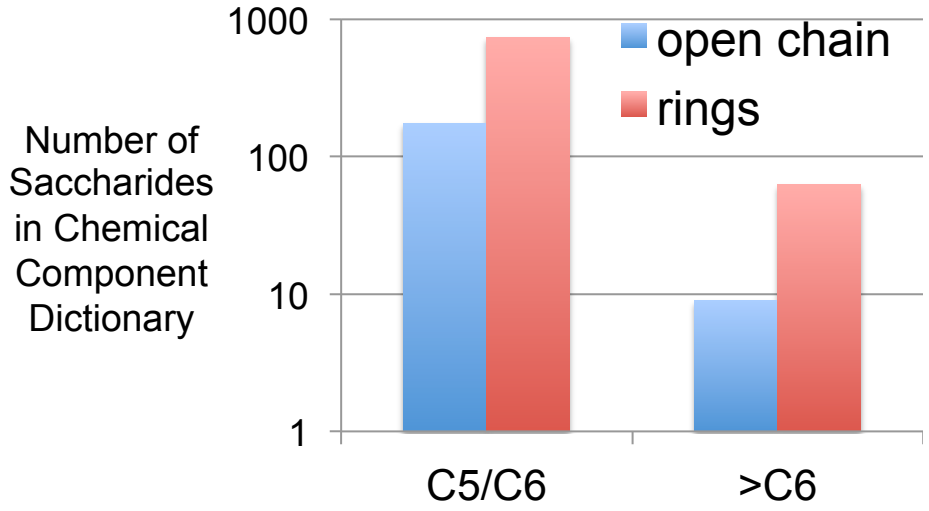
Goal

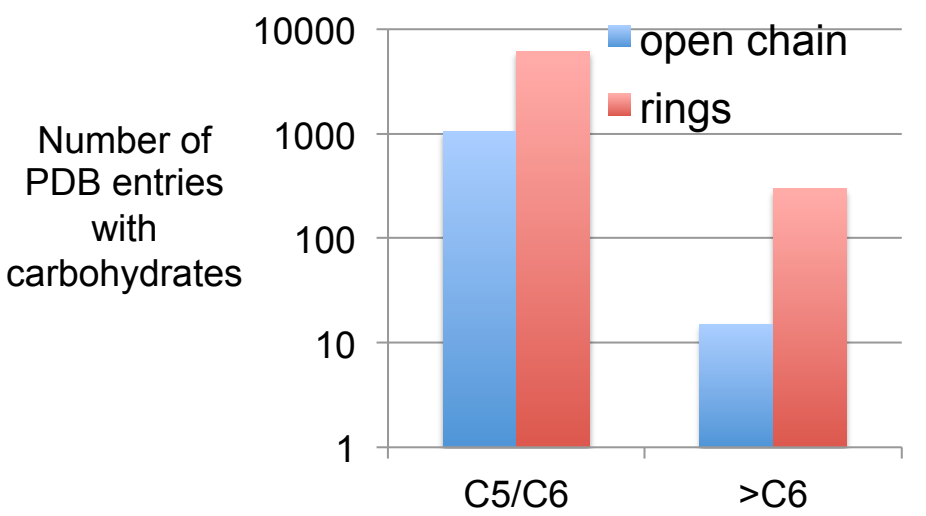
- Represent data consistently within the archive, in agreement with glycobiology community standards
- Enable searches for carbohydrates in the PDB archive

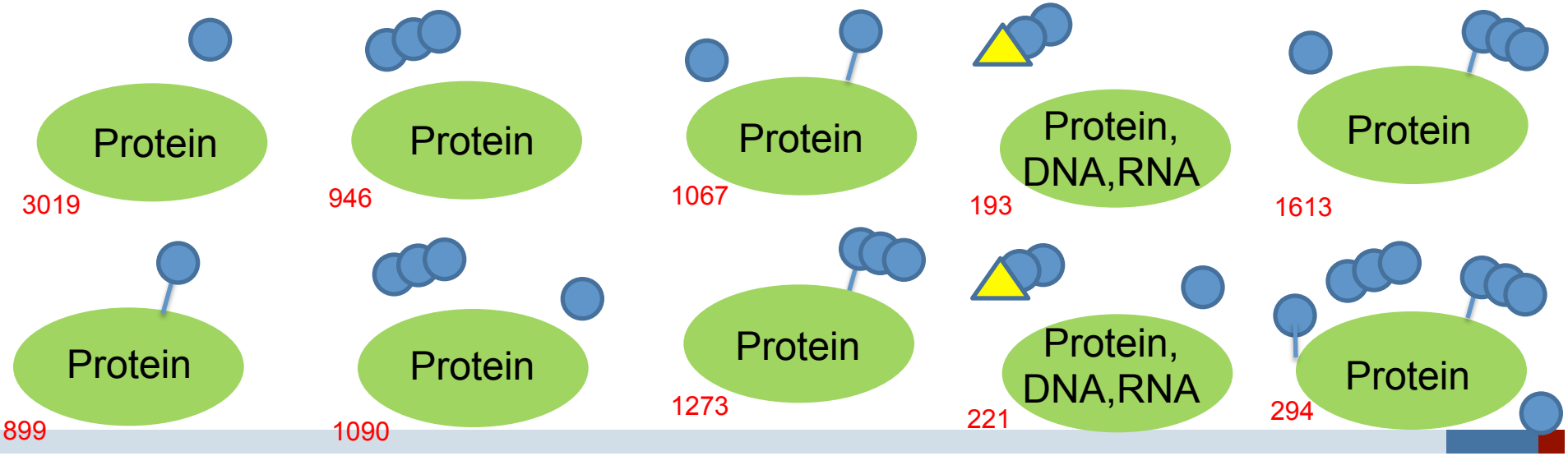
Plan

- 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

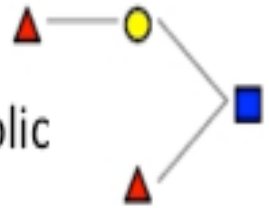
- Represent carbohydrate molecules as polymers of monosaccharides as appropriate
- Adopt glycobiology community standard nomenclature (LINUCS and IUPAC)

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

IUPAC

```
a-L-Fucp-(1-2)-b-D-Galp-(1-4)+
                            |
                       b-D-GlcpNAc
                            |
                  a-L-Fucp-(1-3)+
```

Symbolic



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

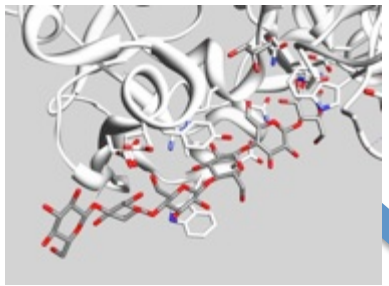
- To identify, classify, and represent all natural protein modifications consistently within the PDB archive and mutually mapped to UniProtKB

Scope

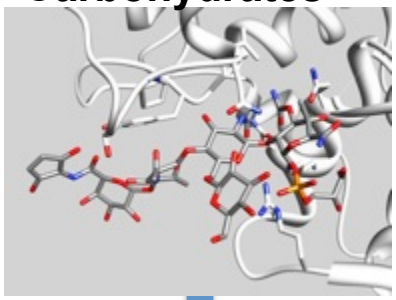
- Chemically modified ribosomal protein including post-translation

Remediated Data

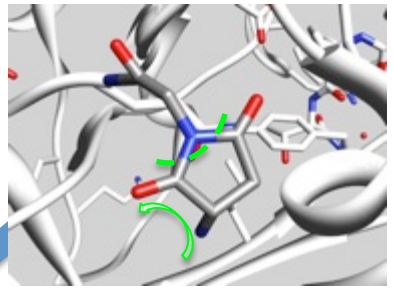
Carbohydrates



Carbohydrates



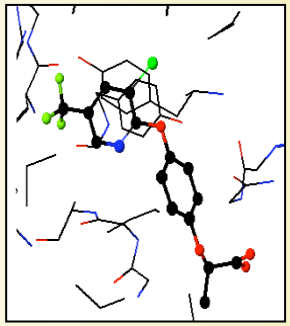
Protein modifications



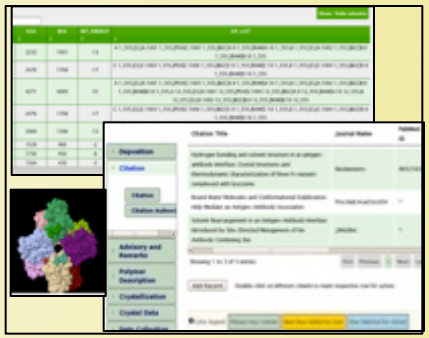
wwPDB Common D&A System

Annotation Pipeline

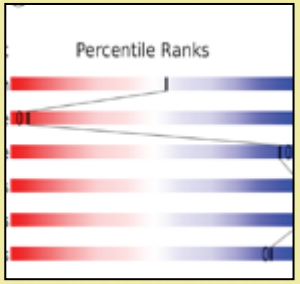
Sequence Processing



Ligand Processing



Manual and Automated Annotation



Validation

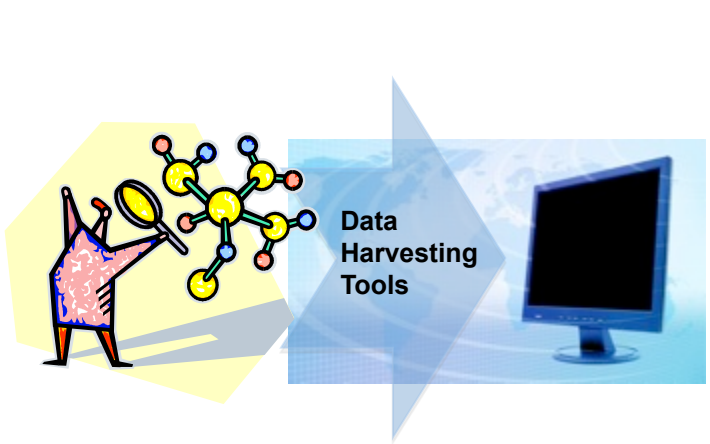
Common Deposition & Annotation Tool

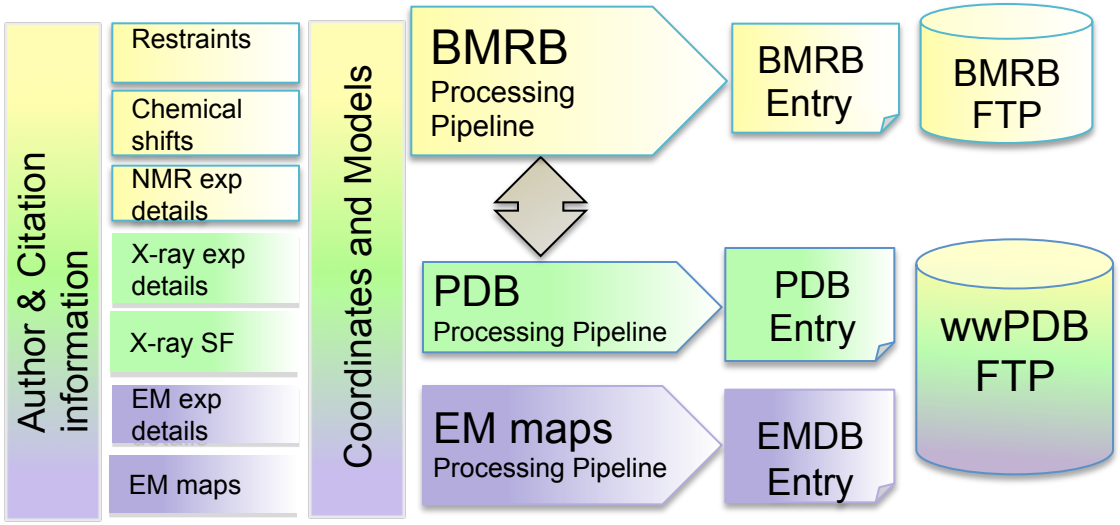
Martha Quesada

WORLDWIDE PDB
PROTEIN DATA BANK

wwpdb.org

The Vision





Integrated Data Capture

Deliverables: Standardization, Quality and Efficiency

Supporting

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

WORLDWIDE PDB
PROTEIN DATA BANK

wwPDB Common Deposition and Annotation Pipeline

Deposition Pipeline

Common Deposition Interface

X-Ray specific

EM specific

NMR specific

Data upload, Harvesting

Client side Editor

Calculated annotations (PISA, SITE & LINK records, cross references, metal coordinates)

Validation

Submission

Progress Tracking/ Status

Common for coordinate processing
Method specific as appropriate

Communication System

Workflow-Automation System

Annotation Pipeline

Sequence Processing

Ligand Processing ID, Edit, Build

Validation

Calculated annotations (PISA, SITE & LINK records, cross references, metal coordinates)

Manual Annotation

Release Processing

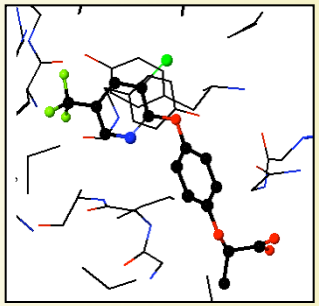
Green and yellow components are common for X-ray, NMR, & EM processing

WORLDWIDE PDB PROTEIN DATA BANK

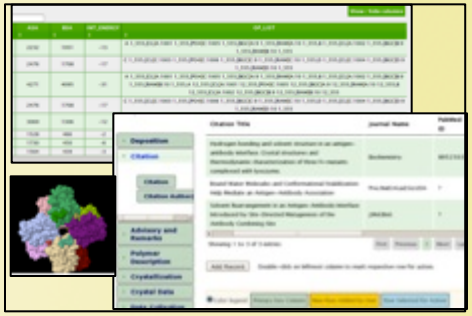
D&A Annotation Pipeline Deliverables

Annotation Pipeline

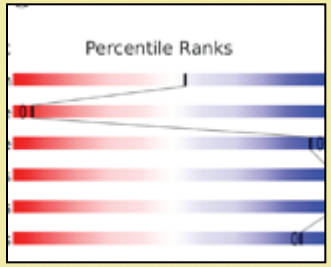
Sequence Processing



Ligand Processing



Manual and Automated Annotation

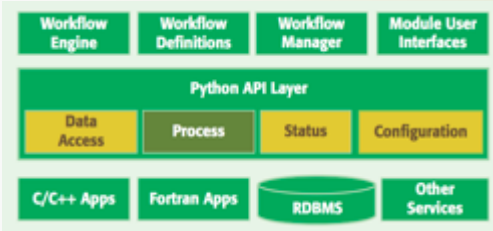


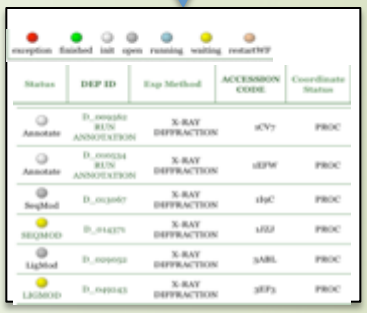
Validation

Infrastructure

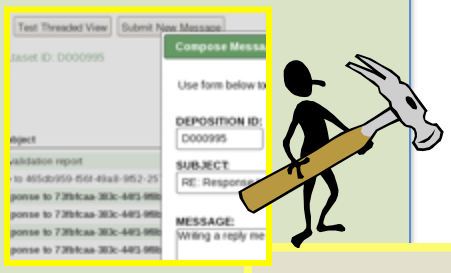
PDBx

NetApp Snap Mirror Data Exchange





Work Flow Manager



Communication

Release Processing

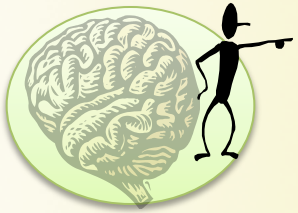
WORLDWIDE PDB
PROTEIN DATA BANK

D&A Deposition Pipeline Deliverables

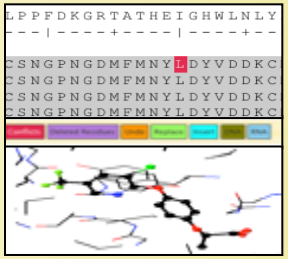
Deposition Pipeline

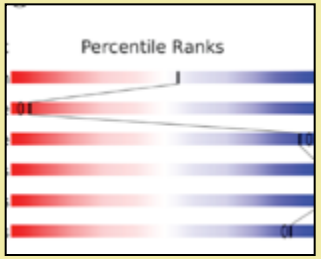
PDB EXTRACT

Data Harvesting



Deposition User Interface



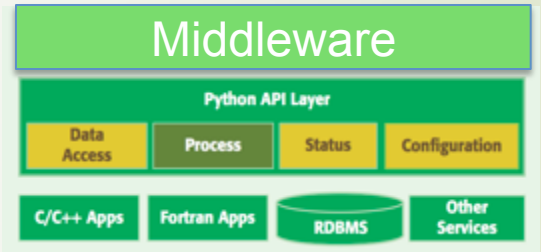


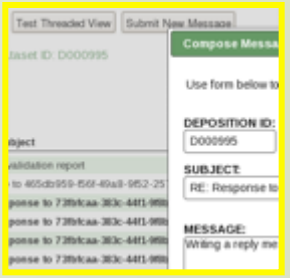
Functional Components

Infrastructure

PDBx

NetApp Snap Mirror Data Exchange





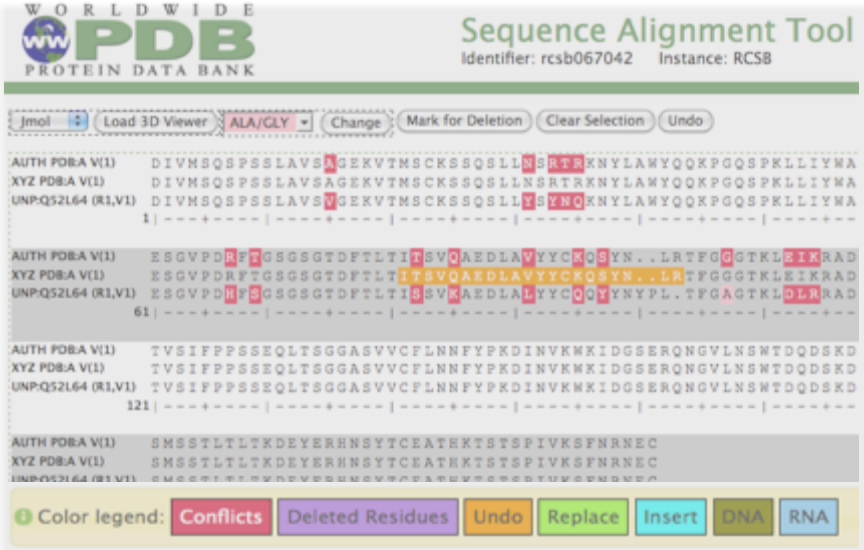
Communication

Release Processing

WORLDWIDE PDB PROTEIN DATA BANK

Processing Modules

Sequence



2012 Enhancements

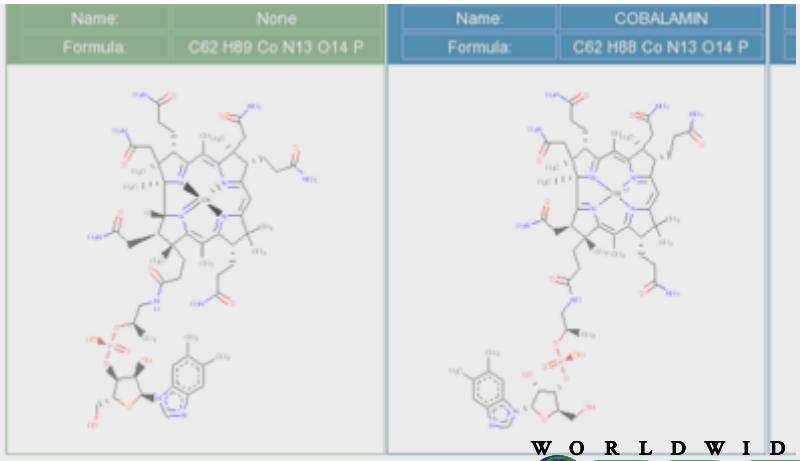
- Chimeric proteins supported
- Sequence match sorting ability
- Create new chemical definitions by splitting or merging existing definitions

➢ Unit and integration testing at all sites

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

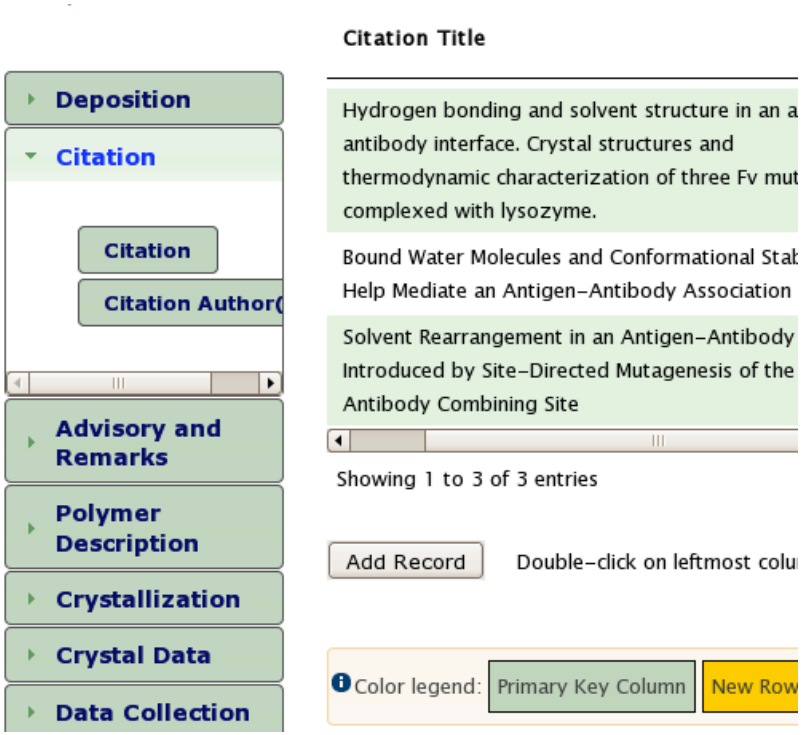


RCSB PDB developers: Zukang Feng, John Westbrook, Raship Shah, Raul Sala, Dimitris Dimitropoulos

WORLDWIDE PDB
PROTEIN DATA BANK

Annotation Pipeline Deliverables

Web form data entry and editing



Automated Annotations

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

RCSB PDB Developers: Raul Sala, John Westbrook, Zukang Feng

WORLDWIDE PDB PROTEIN DATA BANK

Validation Module

Requirements

X-Ray VTF

Functional Components

External Software
Protein Geometry - MolProbity (Richardson lab)
Ligand Geometry - Mogul (CCDC)
Structure Factor - Xtriage, Phenix (P. Adams et al)

Internal Software
Model Validation - EDS (PDBe)
DNA Validation - RCSB PDB
Crystallographic sym. Clashes - RCSB PDB
Ligand stereochemistry and assignment - RCSB PDB
Sequence Validation - RCSB PDB
R factor comparisons from common refinement programs - bundled at RCSB PDB

Report PDF - PDBe

Technical Components

D&A Workflow Manager - PDBe

API (wrapper) for module integration - RCSB PDB

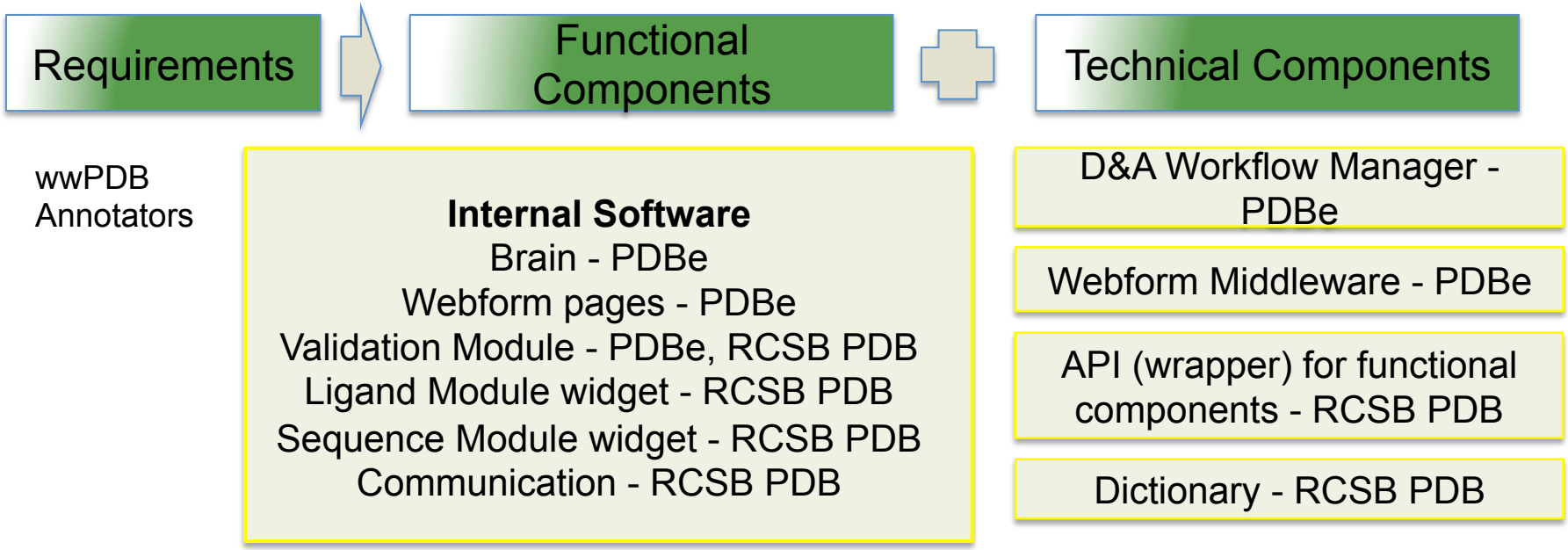
Pipeline - PDBe

Dictionary - RCSB PDB

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

WORLDWIDE PDB
PROTEIN DATA BANK

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

WORLDWIDE PDB PROTEIN DATA BANK

Workload Distribution

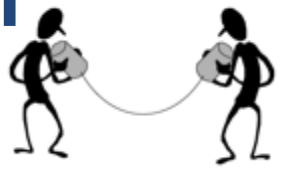
Distribution will take into account

- 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



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

WORLDWIDE PDB PROTEIN DATA BANK

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

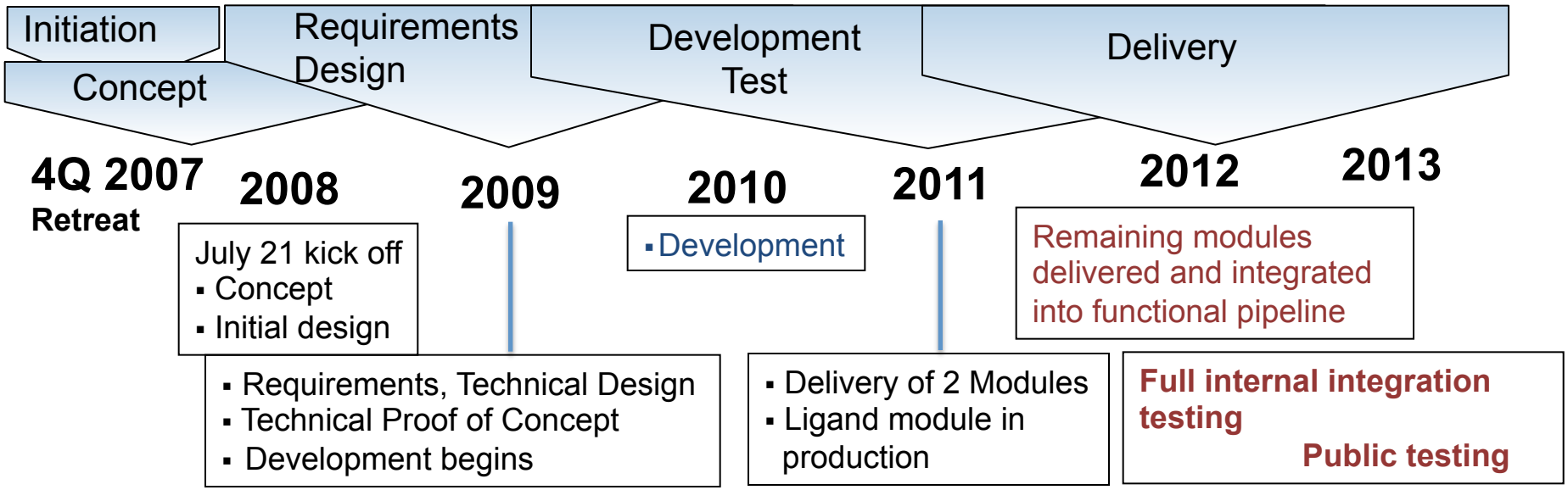
- Additional visualization, data harvesting
- Validation requirements from EM VTF to be supported

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



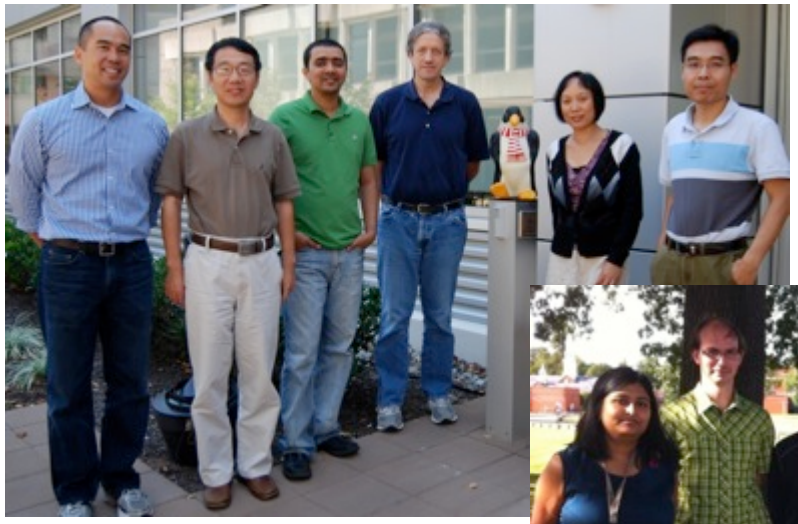
2012 Deliverable: wwPDB Common D&A System Version 1.0

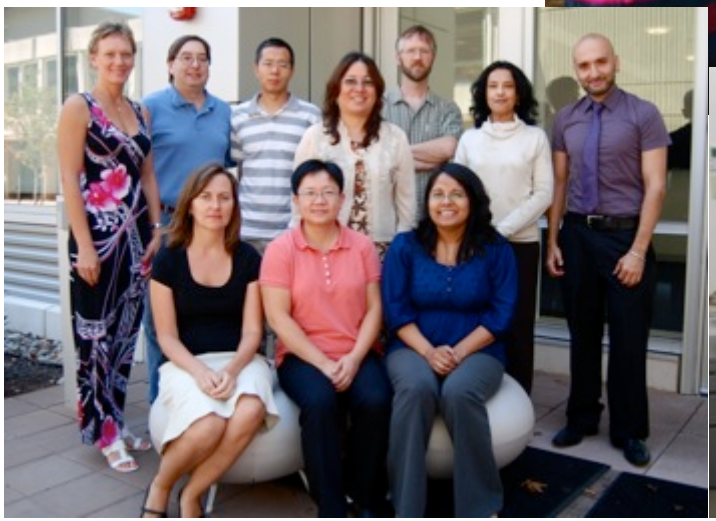
- Able to process a file from deposition through annotation
- Supports all existing D&A processes and procedures
- Offers enhanced User Interfaces for functional modules and deposition process (within reason)
- Provides a “Work Flow” infrastructure that enables task tracking and automation

WORLDWIDE PDB
PROTEIN DATA BANK

Common D&A Project Team

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





WORLDWIDE PDB PROTEIN DATA BANK

RCSB PDB PROTEIN DATA BANK

PDBe PROTEIN DATA BANK EUROPE

PDBj Protein Data Bank Japan

BMRB

Data Out: Website Features

Peter Rose, Philip Bourne

Pharmaceutical Sciences

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

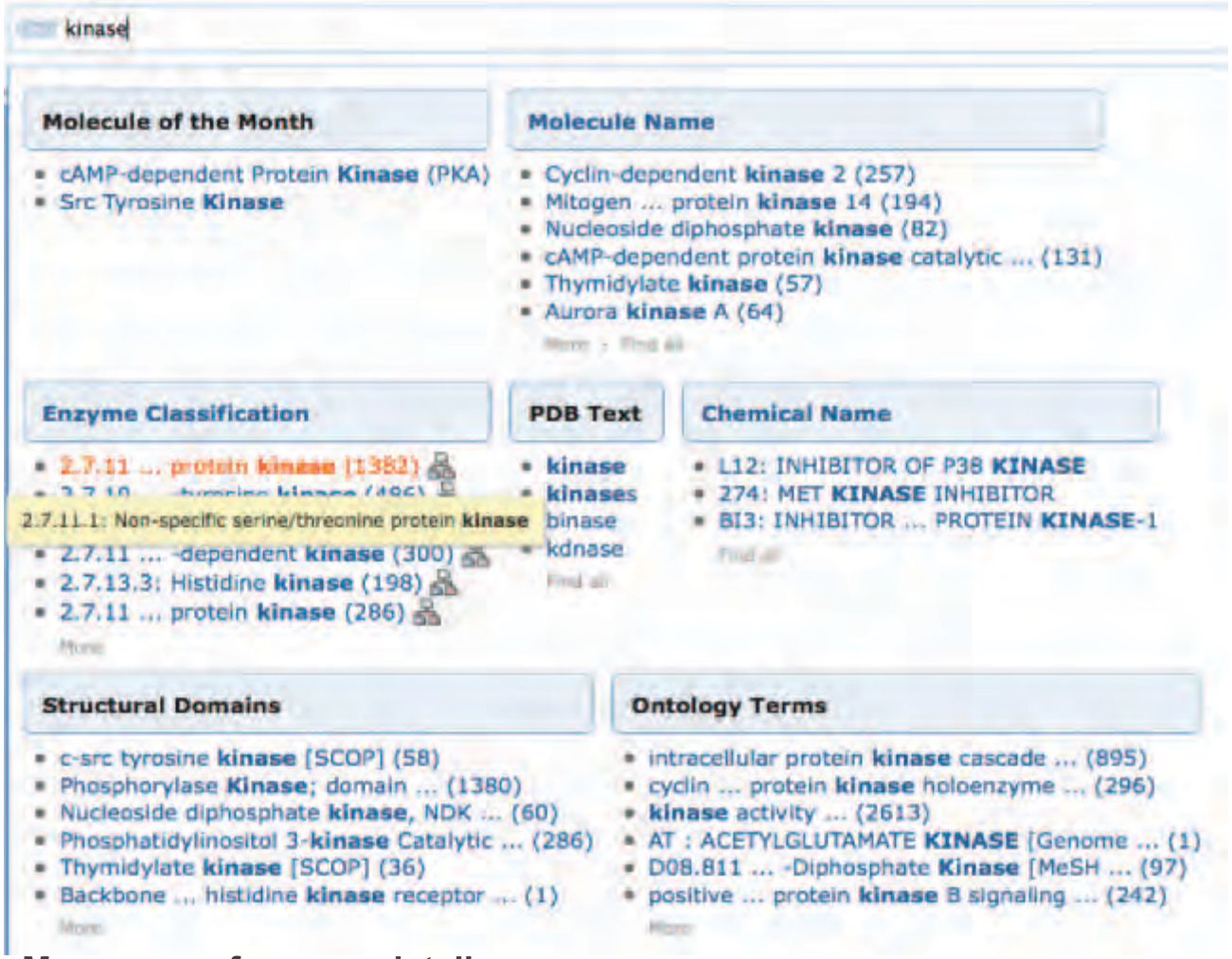
- Searching and Browsing the Archive
- Tools and Services
- Website Usage Statistics
- Hardware Renewal
- Plans

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

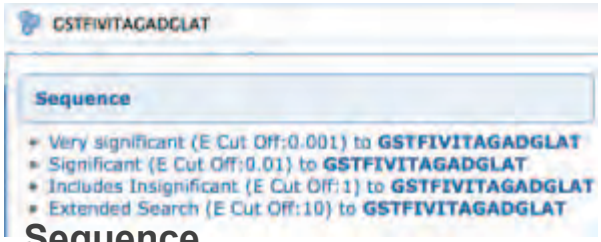
Guiding Users to Specific Results

Search Suggestions

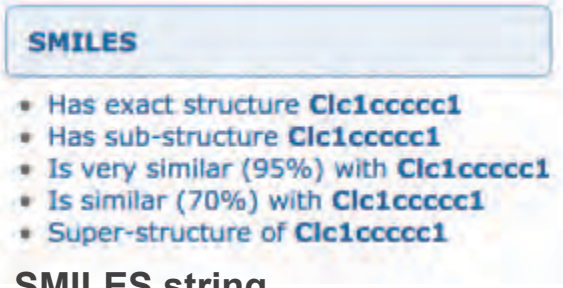


Mouse over for more details

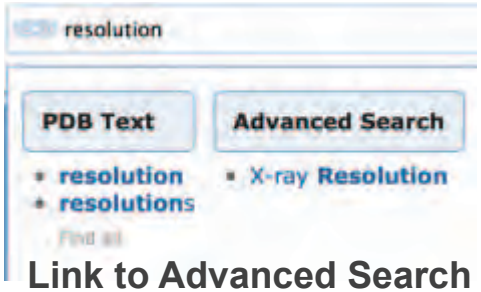
Smart Suggestions



Sequence

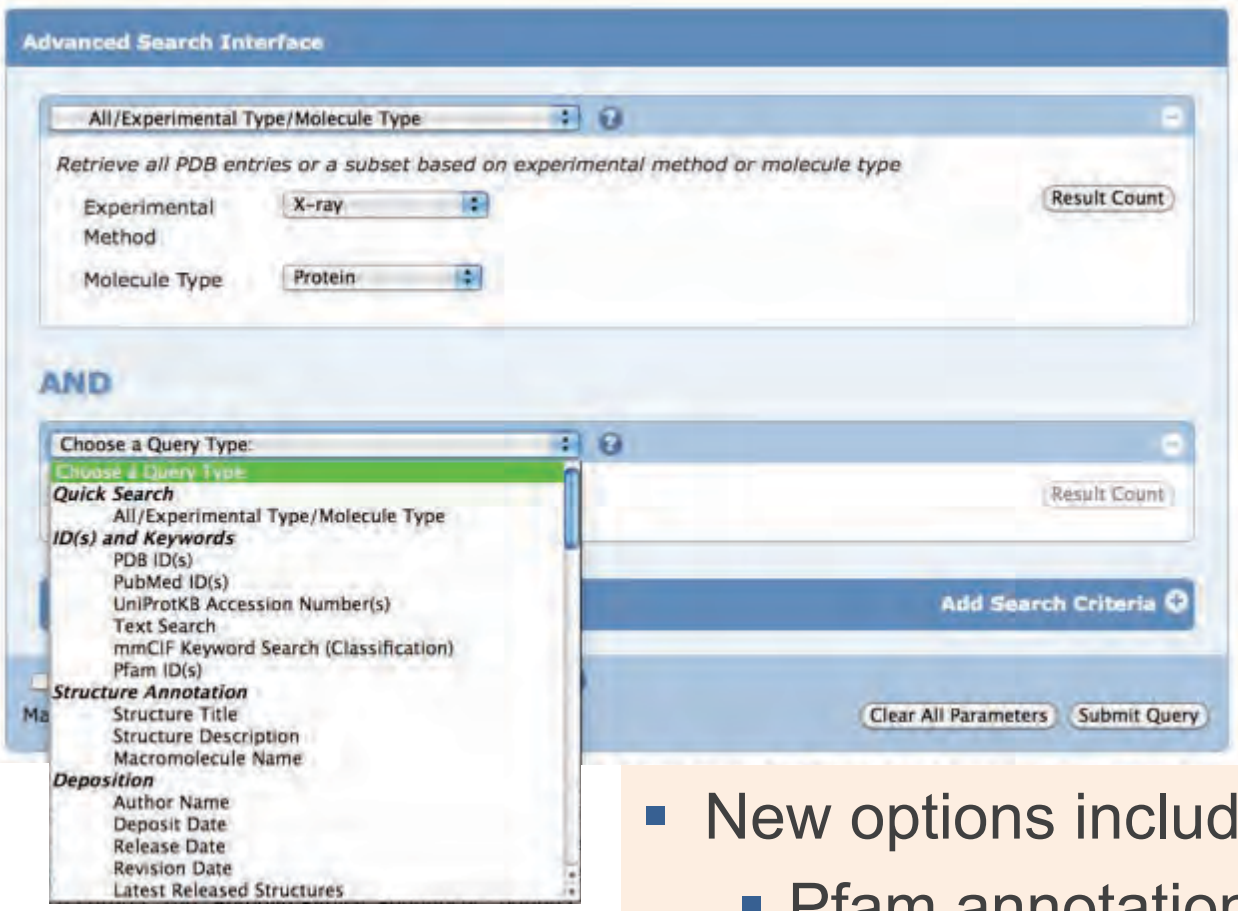


SMILES string



Link to Advanced Search

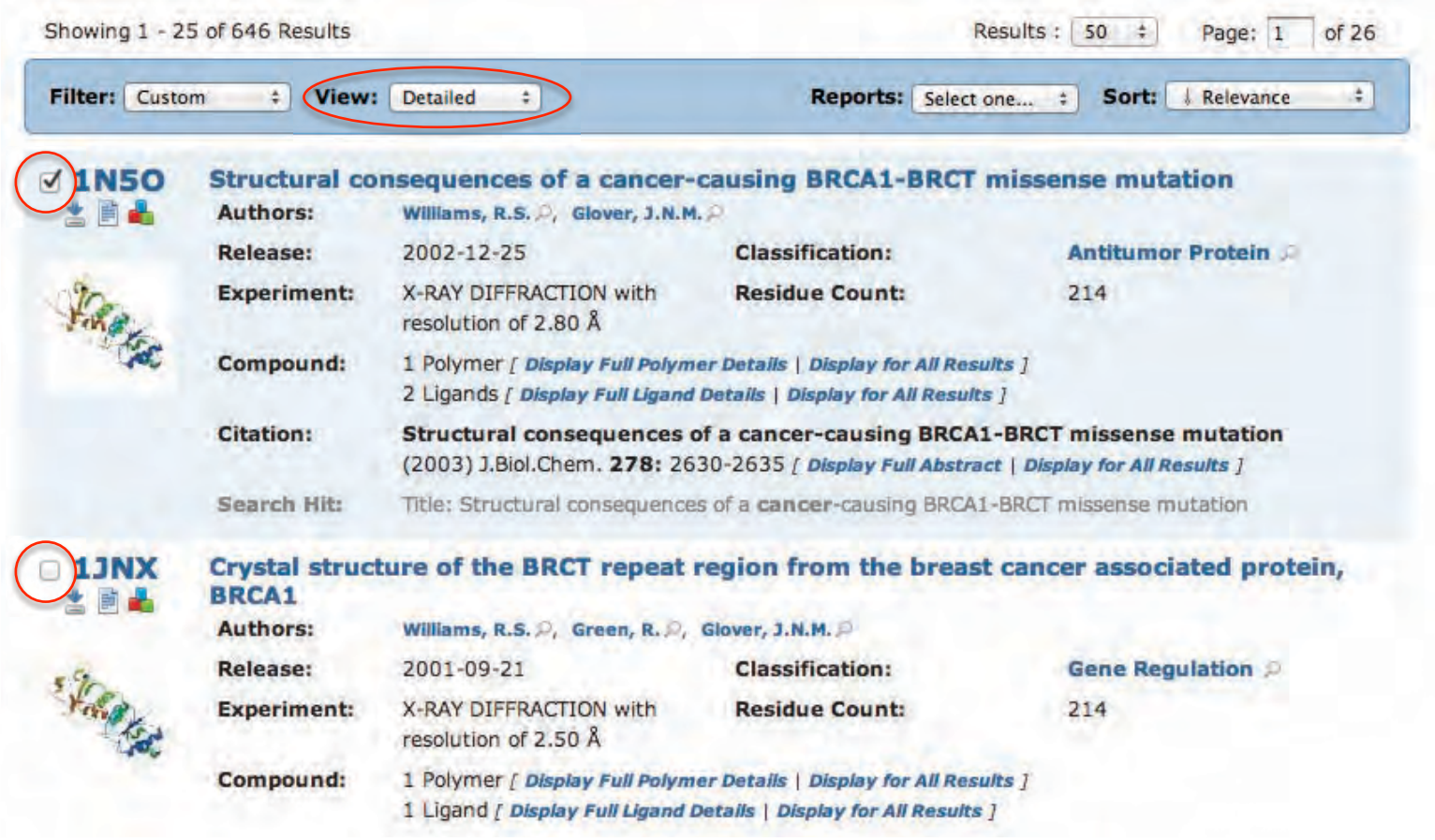
Advanced Search



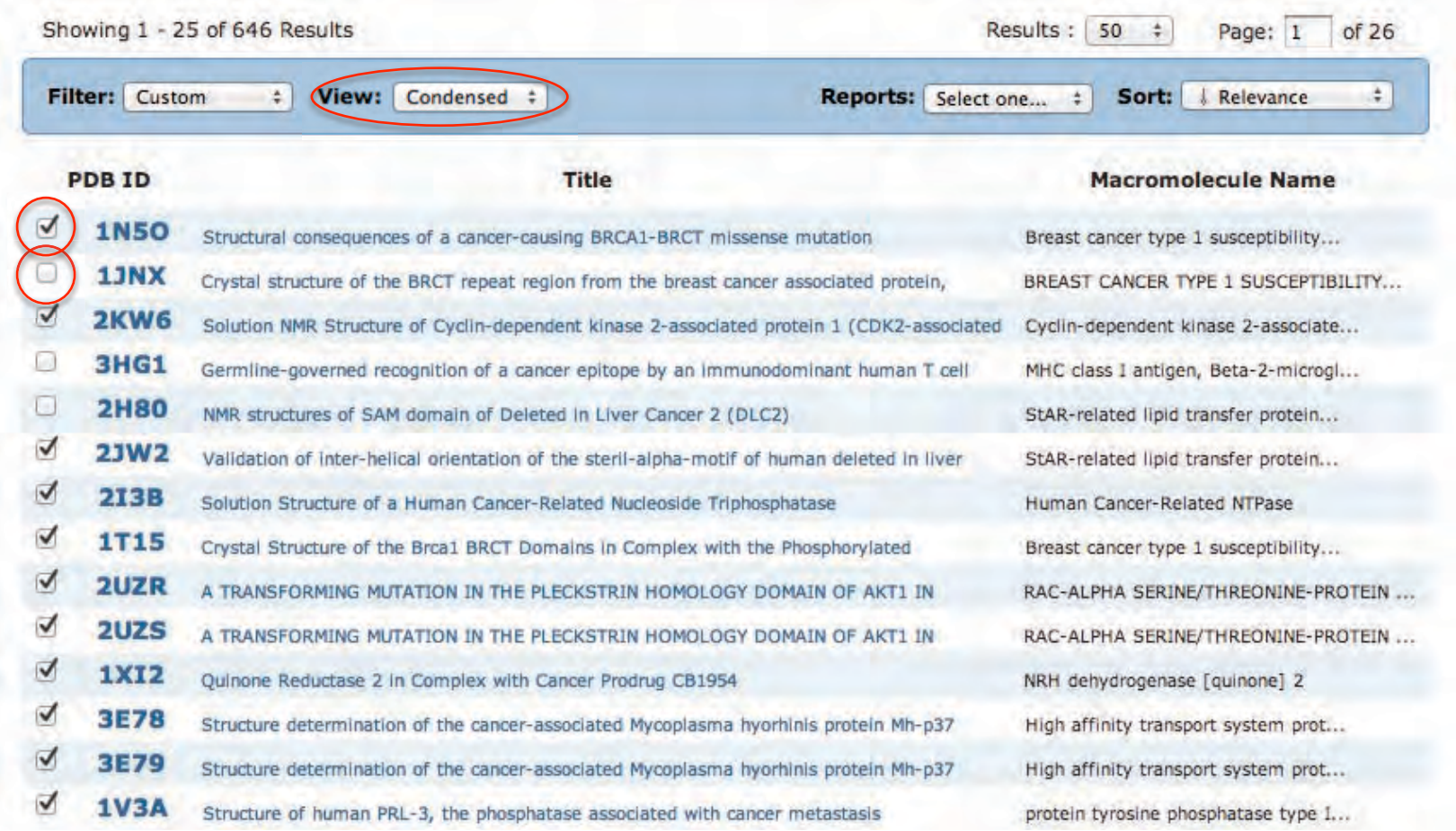
- Search data items in a variety of categories
- Combine queries using AND / OR
- Contextual help

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

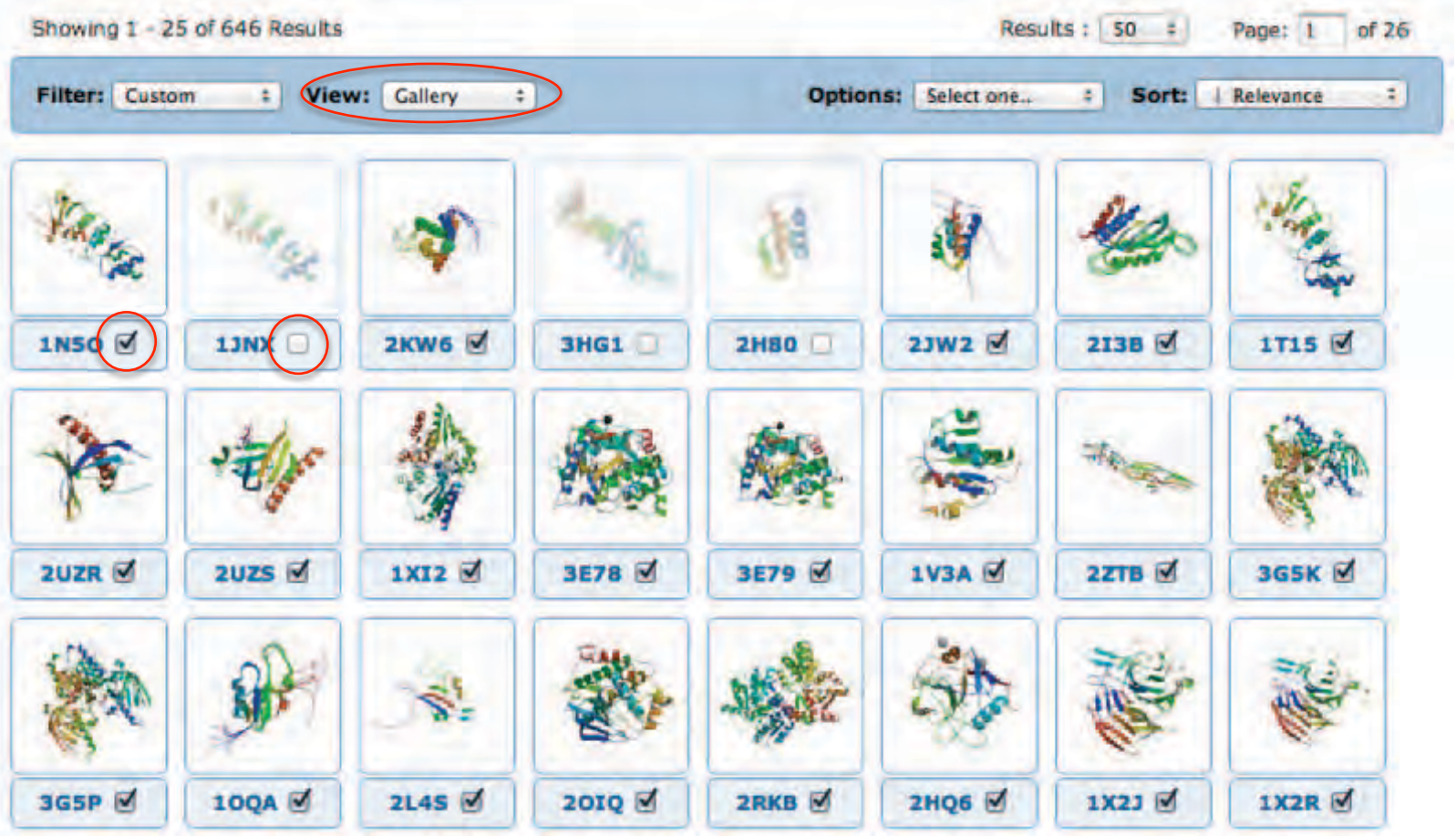
Fall 2012: Views with Synchronized "Shopping Cart"



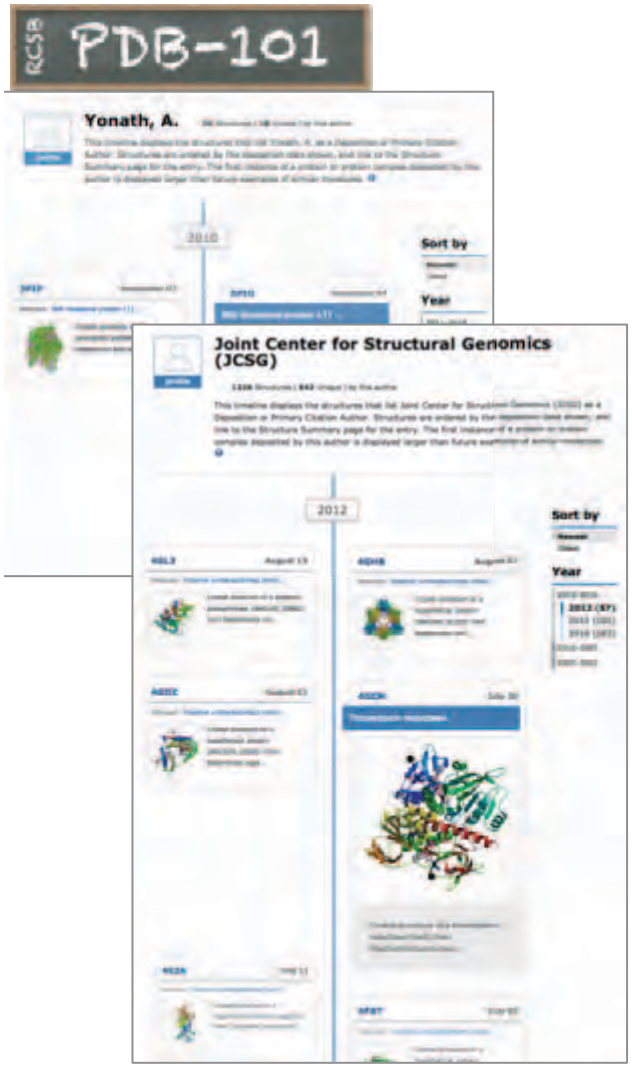
Fall 2012: Condensed View for Rapid Browsing

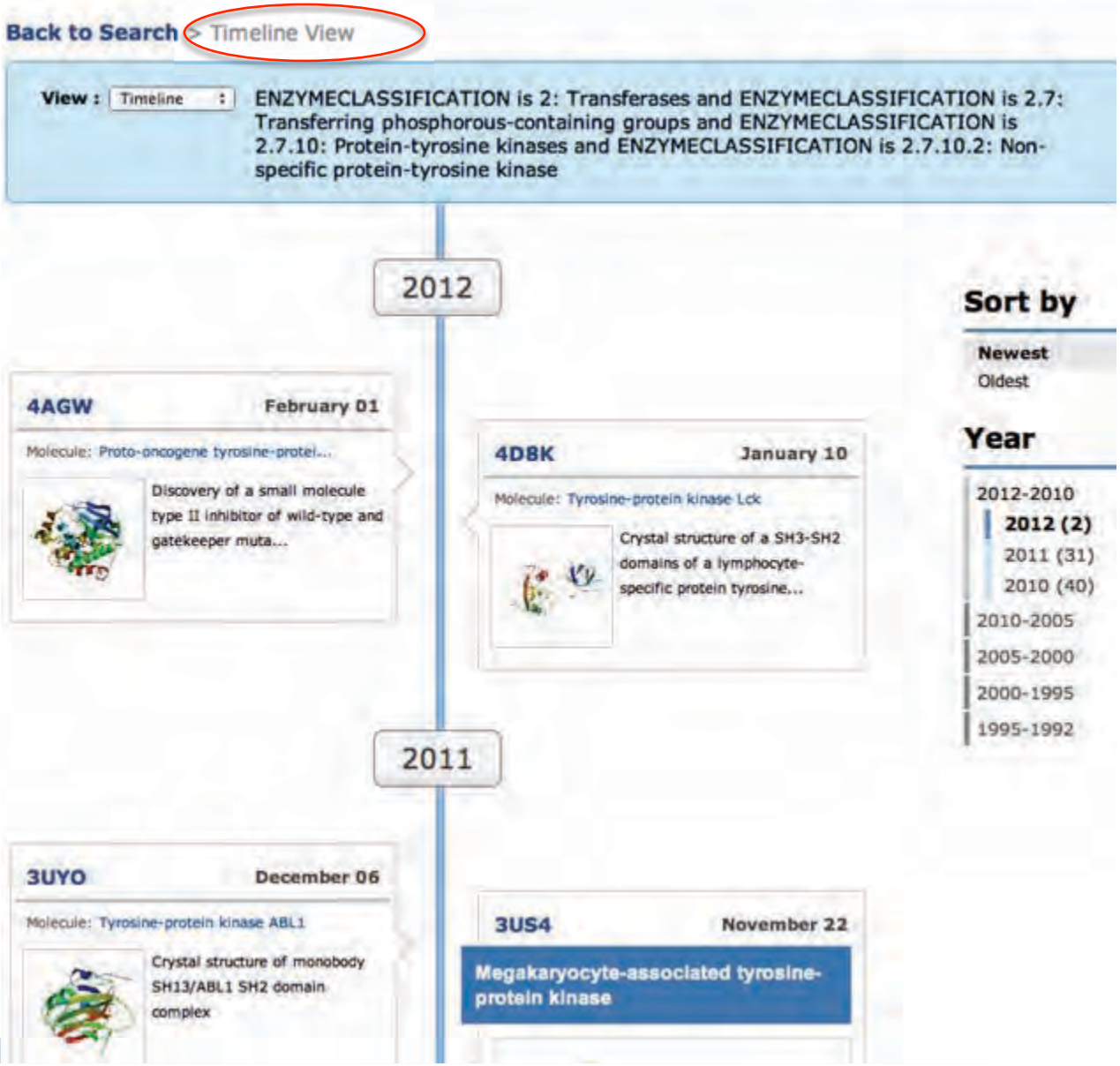


Fall 2012: Gallery Browsing and Selection



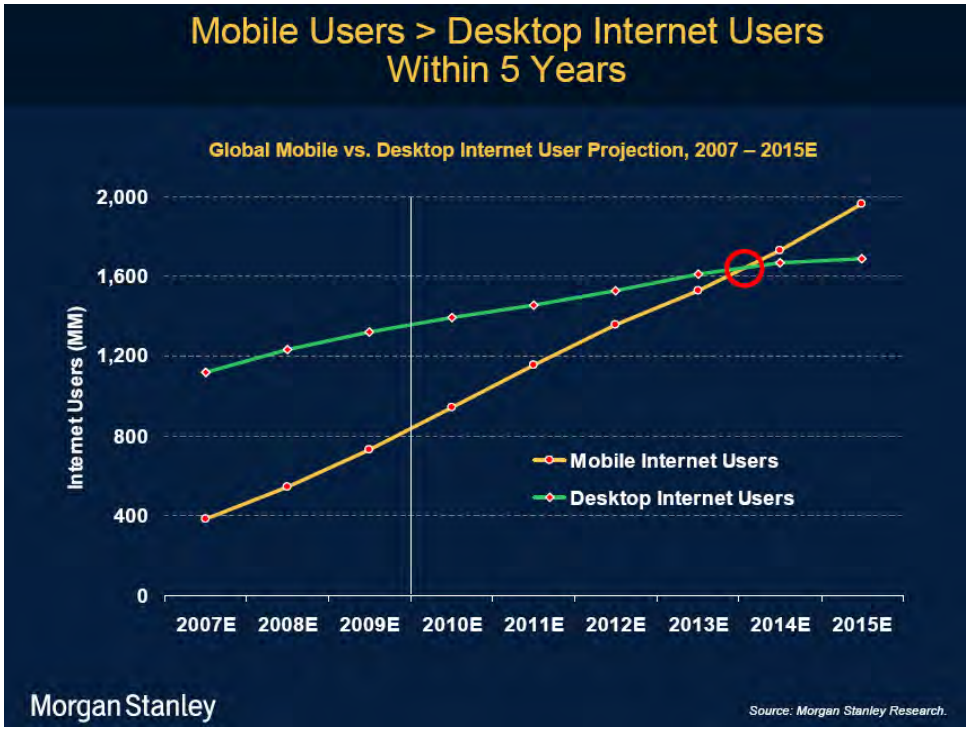
Fall 2012: Timeline View





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



http://www.morganstanley.com/institutional/techresearch/pdfs/Internet_Trends_041210.pdf

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

Available on the App Store

Browse *Molecule of the Month* articles

Simple search

iPhone iPod

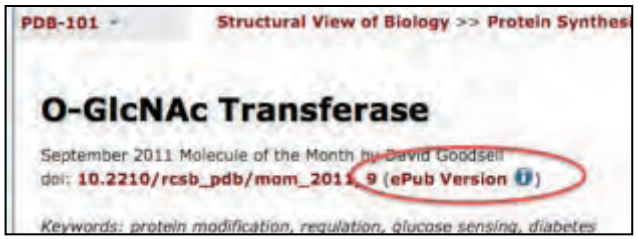
Interactive 3D viewer*

Browse search results

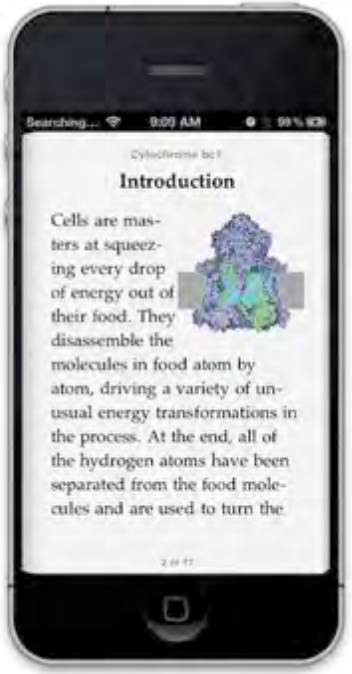
iPad

*NDKMol, Takanori Nakane, Kyoto University

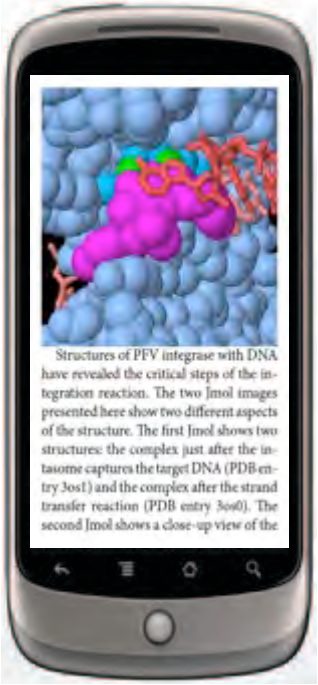
Molecule of the Month in ePub Format



iPad



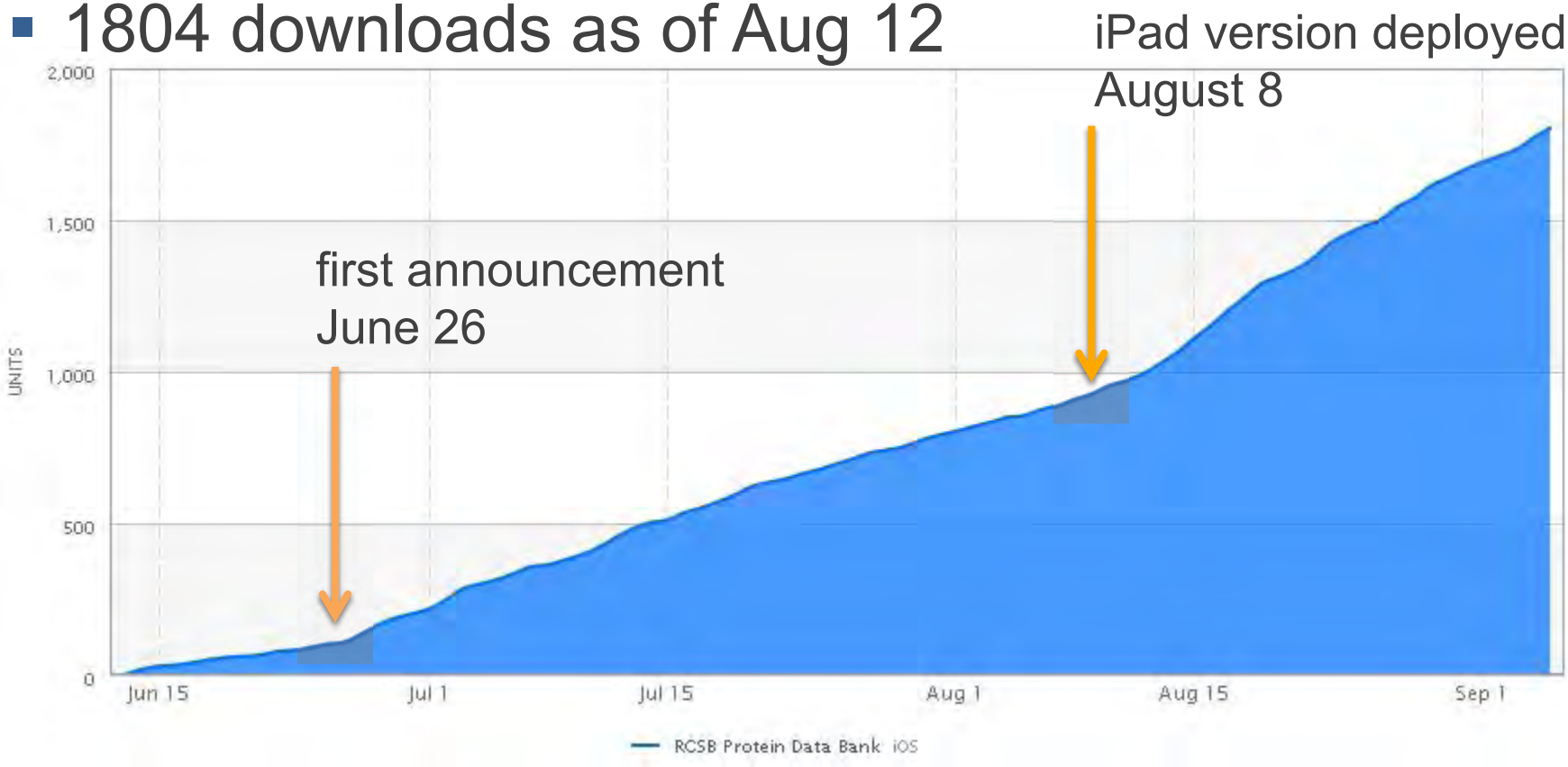
iPhone/iPod



Android

RCSB PDB *Mobile* App Downloads

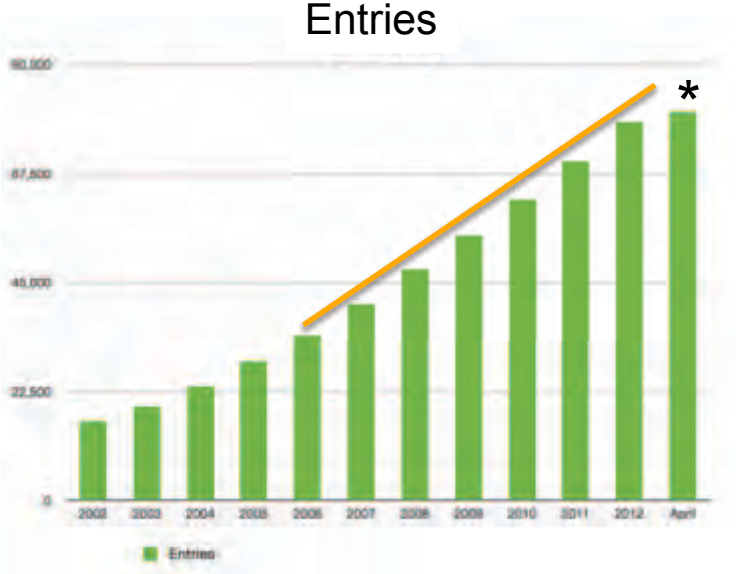
- Deployed in iTunes store (free)
  - iPhone/iPod: May 29
  - iPad: Aug 8
- 1804 downloads as of Aug 12



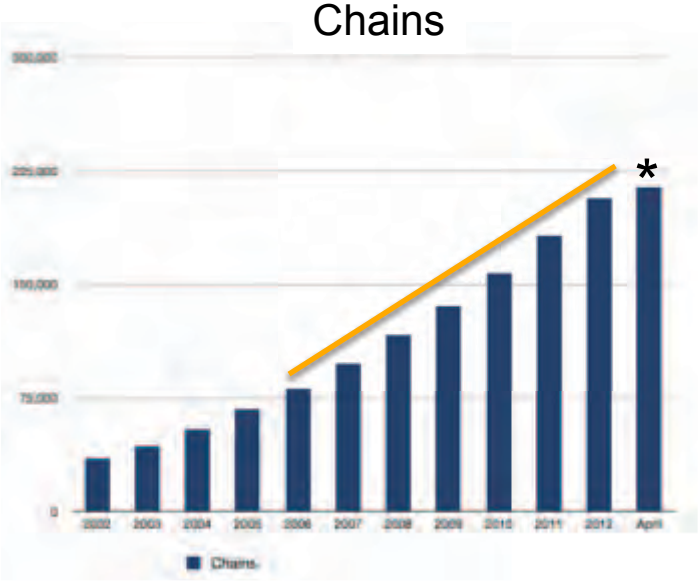
Data Out: Tools and Services

Growth of PDB and Sequence Clusters

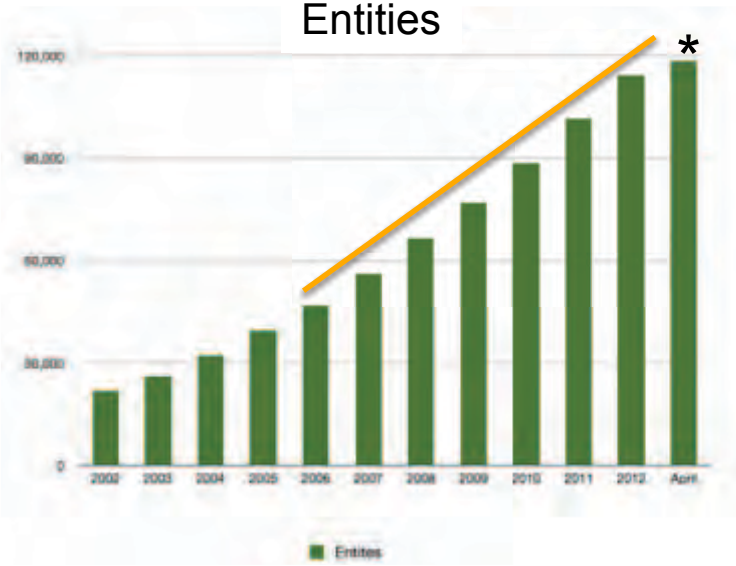
Entries



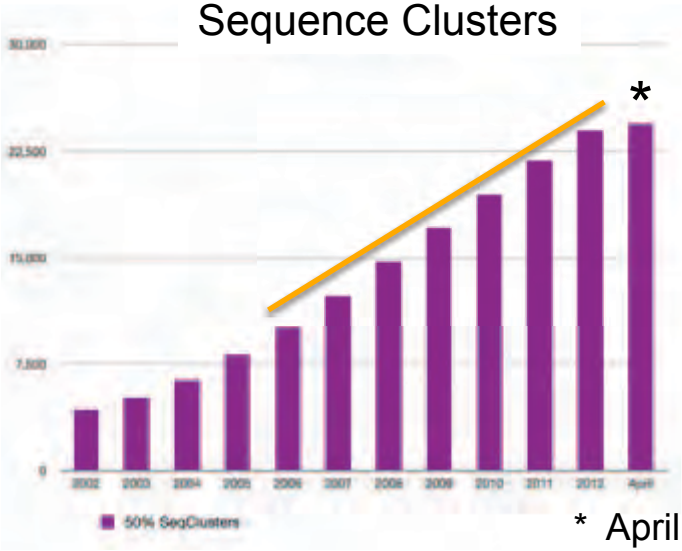
Chains



Entities

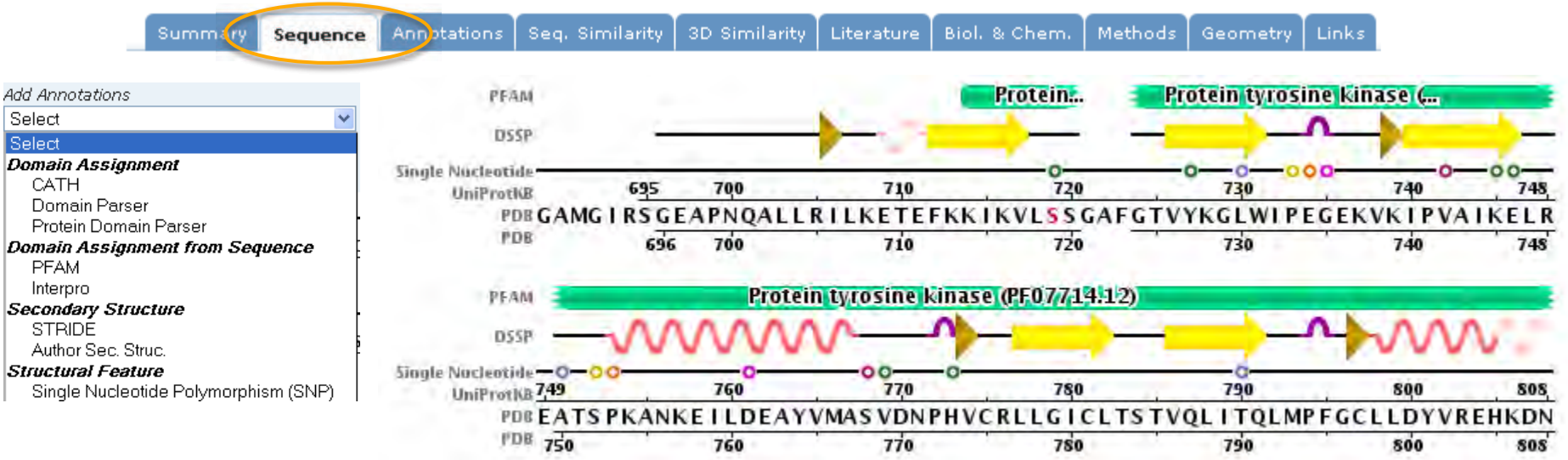


Sequence Clusters



* April 2012

Sequence Annotations

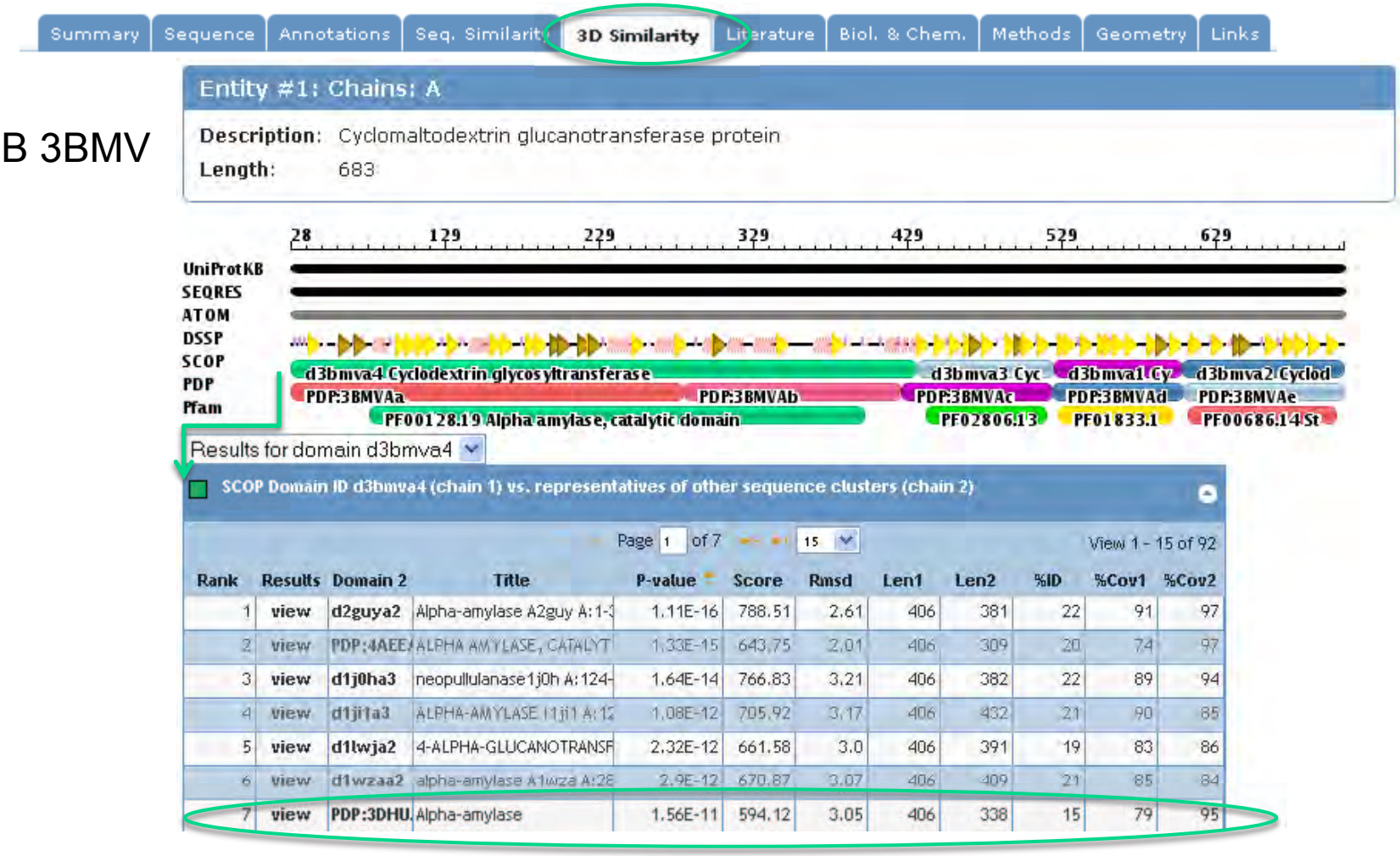


Annotation sources

- 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



Domain-based Alignment Example

Structure Alignment Results

Alignment Details:

P-value: 1.56e-11
Score: 594.12
RMSD: 3.05
%Id: 15.0%

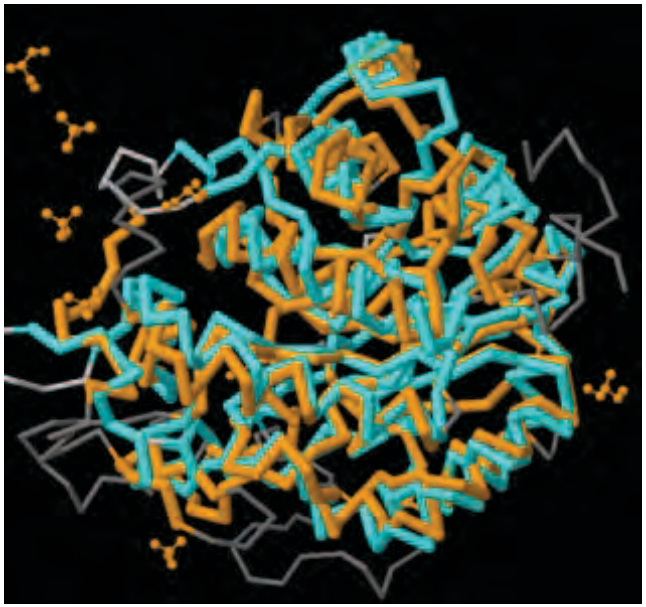
Query: (orange/dark grey)
Cyclomaltodextrin glucanotransferase3bmv A:1-406 c.1.8.1 (Trans)glycosidases

PDB ID:	3BMV
SCOP domain	d3bmva4
Length:	406
Similarity:	79%
EC number:	2.4.1.19

Subject: (cyan/light grey)
Alpha-amylase

PDB ID:	3DHU
PDP Domain:	PDP:3DHUAa
Length:	338
Similarity:	95%
EC number:	3.2.1.1

15% sequence identity



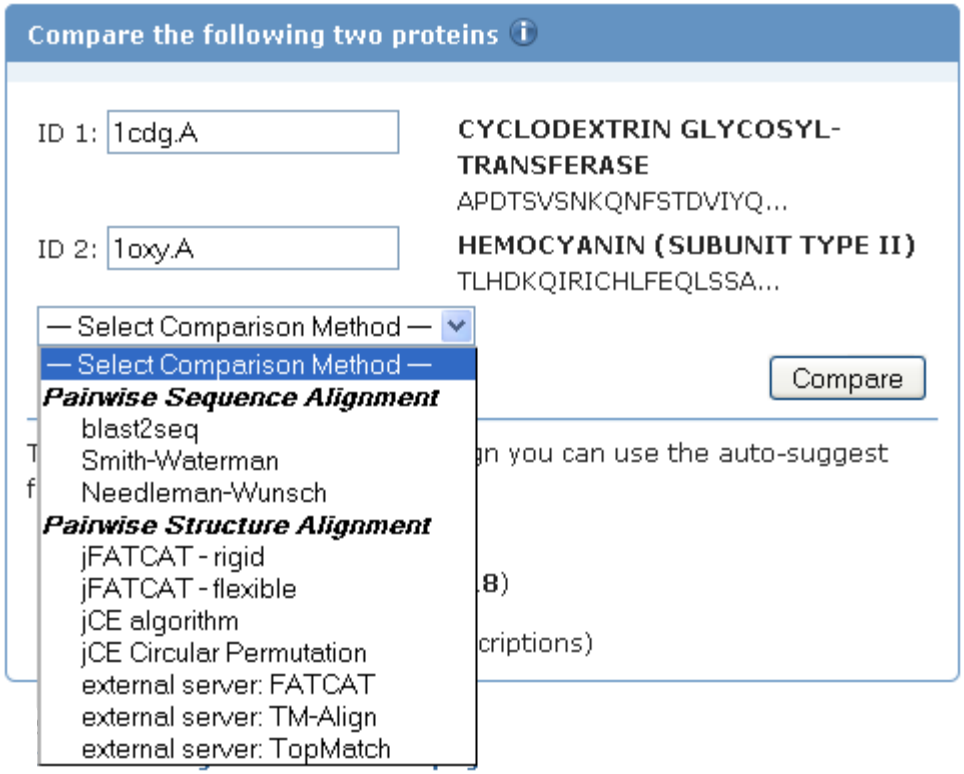
79% and 95% structural overlap

different function:
EC 2.4.1.19 vs. 3.2.1.1

Custom Chain and Domain-based Alignment

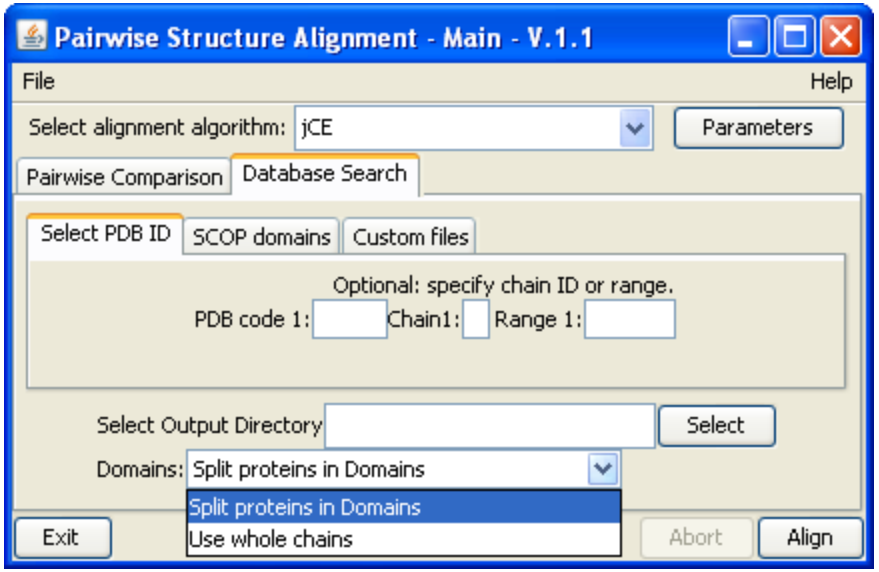
Protein Comparison Tool (website)

- Auto-complete for chain and domain IDs



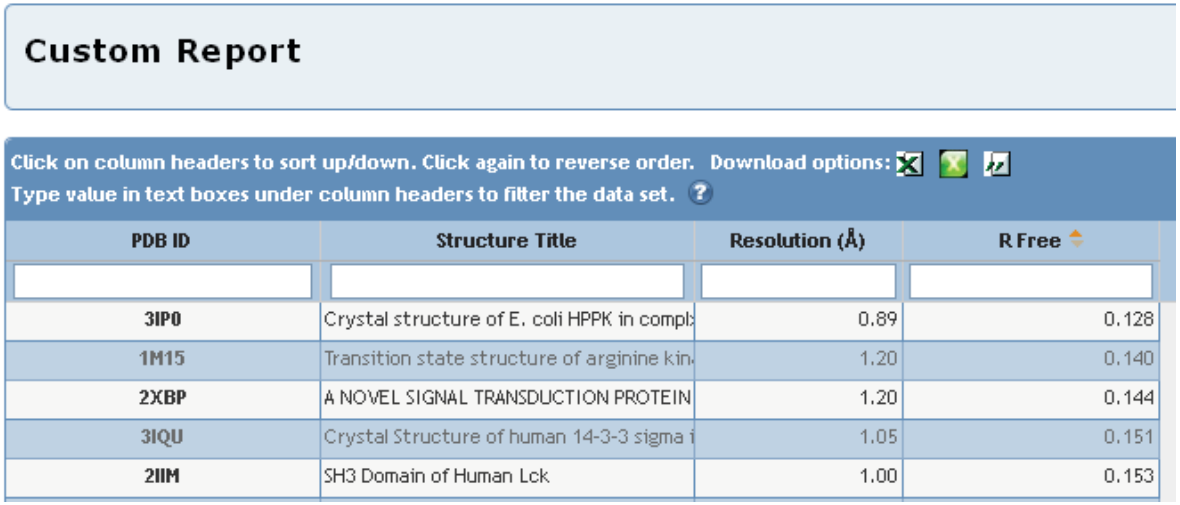
Protein Comparison Tool (download)

- 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
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

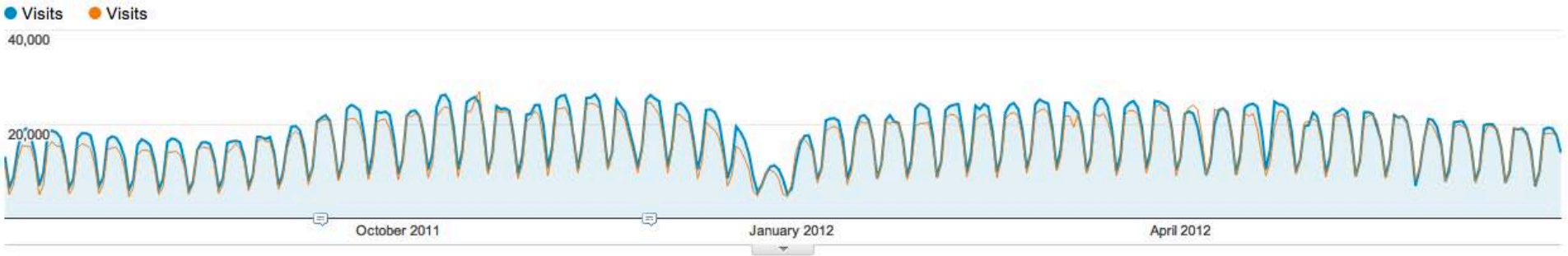
Data Out: Website Usage

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

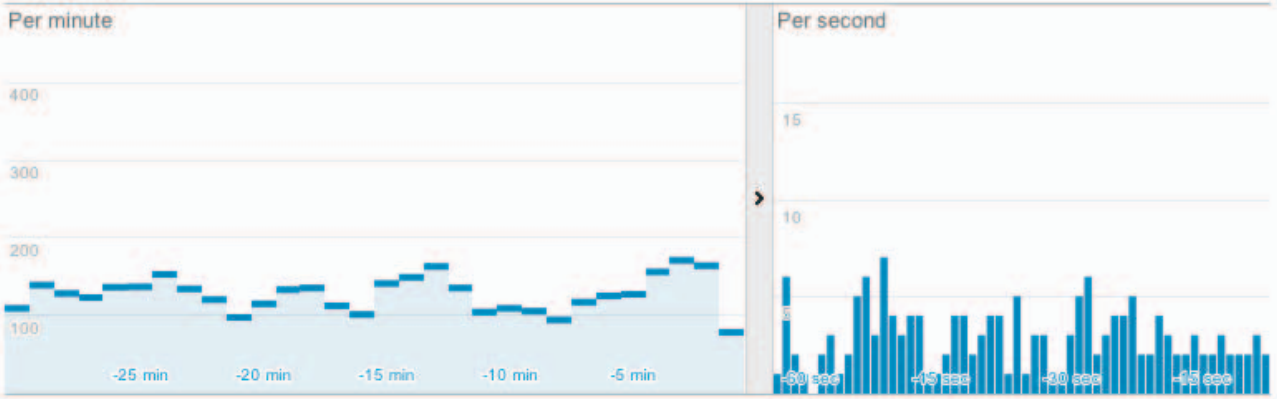
Right now

292

active visitors on site

RETURNING 66% NEW 34%

Pageviews



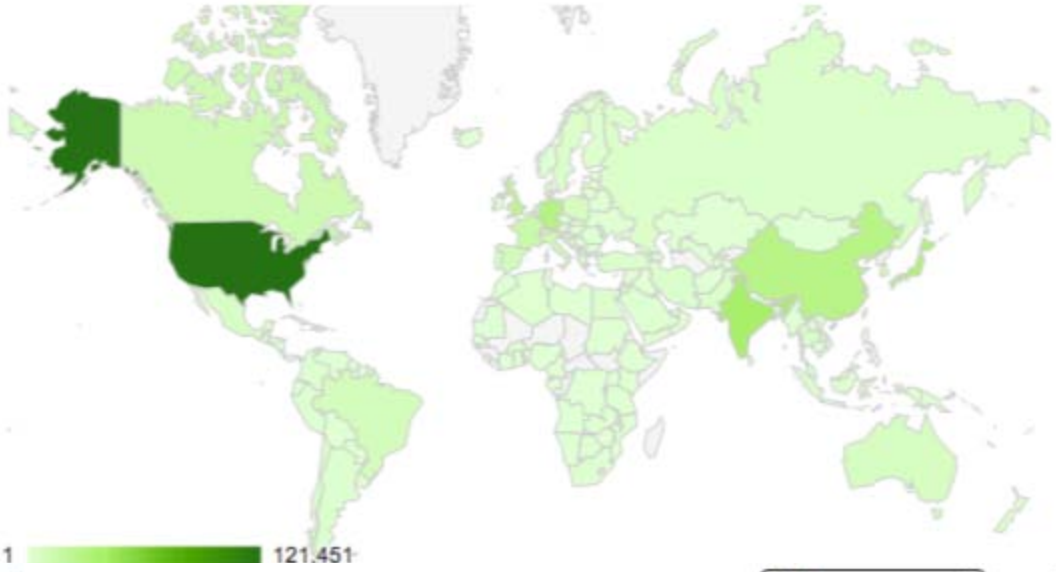
Top Referrals:

	Source	Active Visitors
1.	en.wikipedia.org	8
2.	ncbi.nlm.nih.gov	5
3.	nature.com	3
4.	cbm.msoe.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:

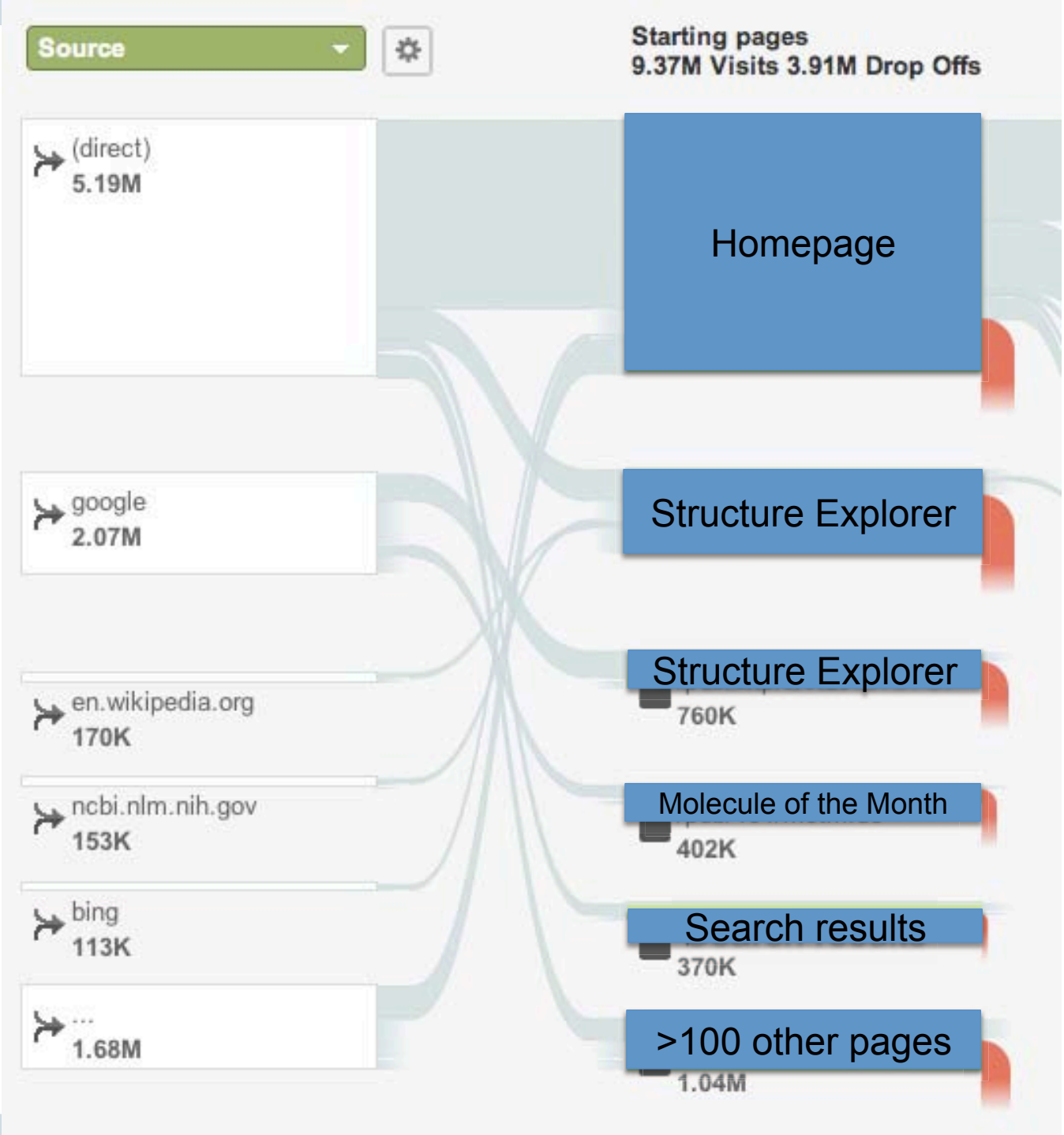
	Active Page	Active Visitors	
1.	/pdb/home/home.do	42	14.38%
2.	/pdb/explore/explore.do?structureId=2H5Z	9	3.08%
3.	/pdb/explore/explore.do?structureId=1KS5	5	1.71%
4.	/pdb/explore/explore.do?structureId=1EMA	3	1.03%
5.	/pdb/explore/explore.do?structureId=2F1C	3	1.03%
6.	/pdb/explore/explore.do?structureId=2h5z	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



Country/Territory	Visits	Visits
1. United States	121,451	26.65%
2. India	38,173	8.38%
3. Japan	31,943	7.01%
4. Germany	29,283	6.43%
5. China	27,651	6.07%
6. United Kingdom	23,521	5.16%
7. France	14,755	3.24%
8. Canada	11,469	2.52%
9. Italy	11,397	2.50%
10. Spain	9,086	1.99%

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

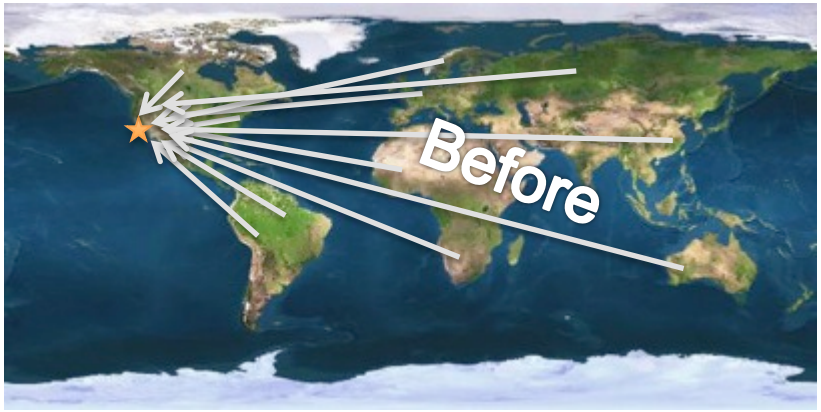
Old

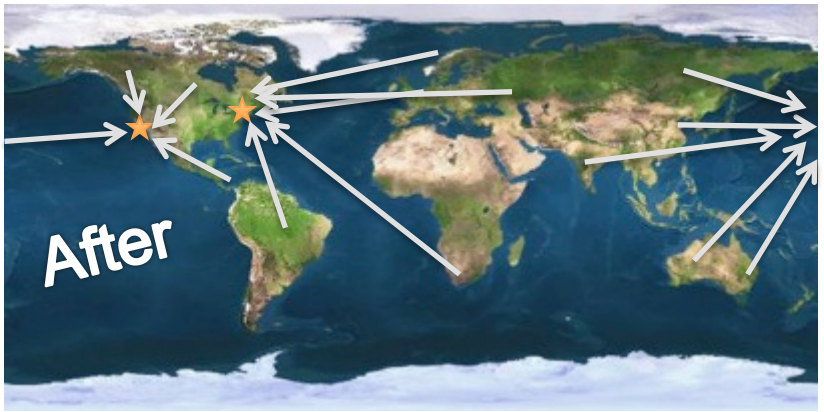
SDSC Rutgers

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

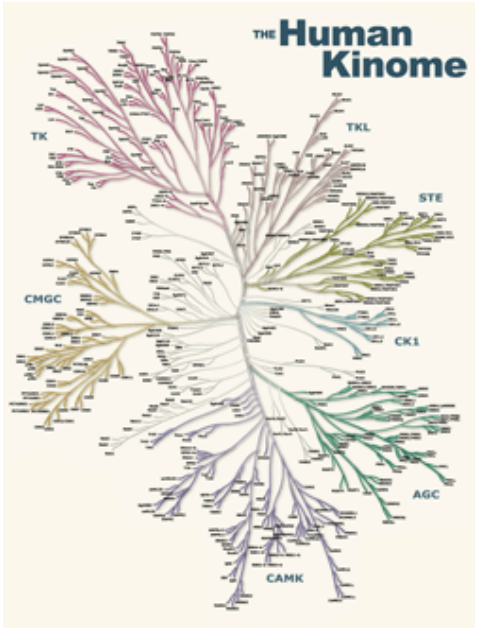
Expand RESTful Web Services

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

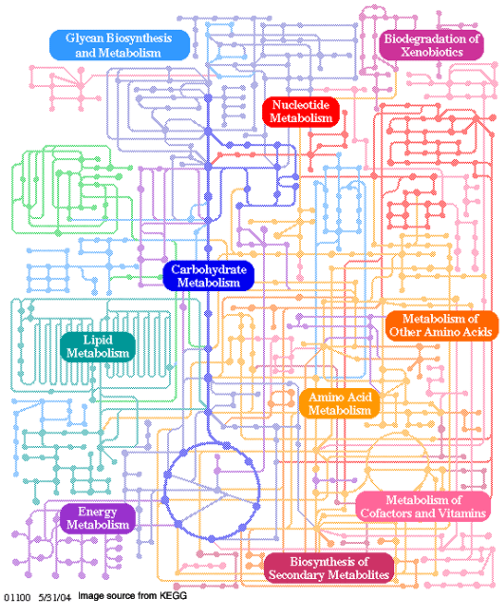
Plans: Structural View of Biology

Map structures to:

Drugs and drug targets



Protein families



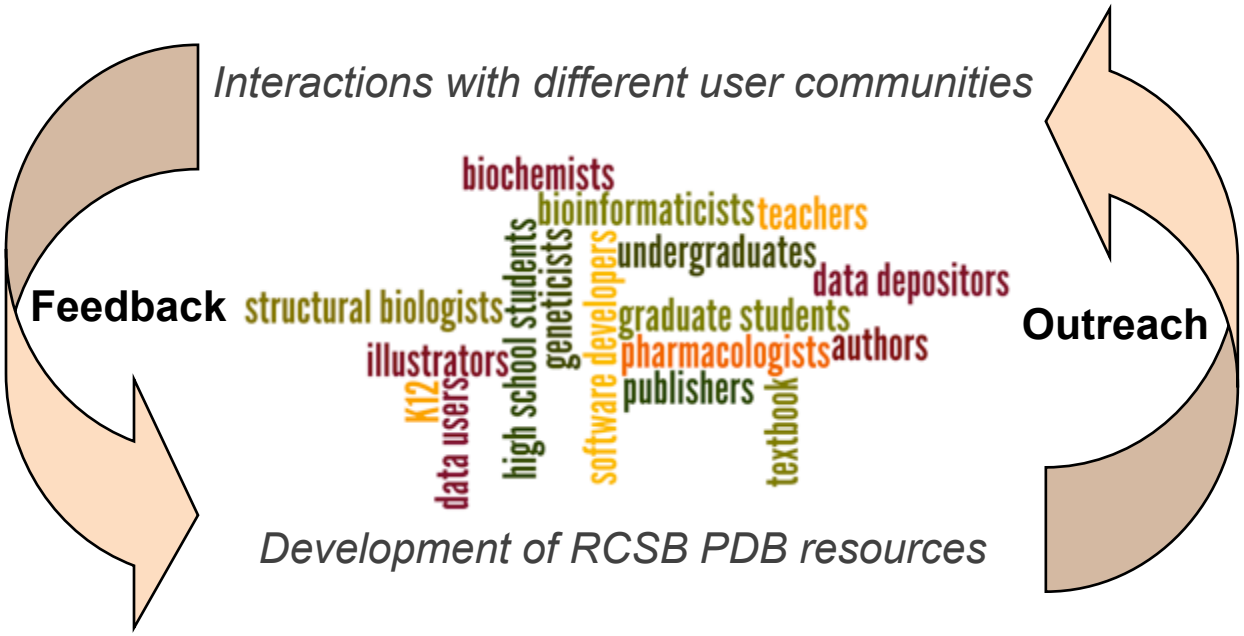
Pathways

Outreach & Education

Christine Zardecki
Shuchismita Dutta

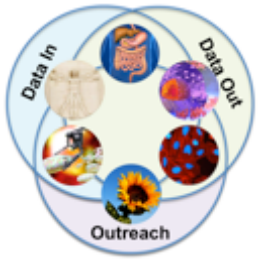
Goals

- 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

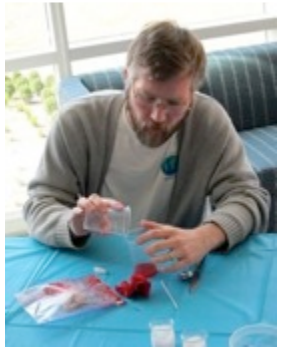
Electronic Support

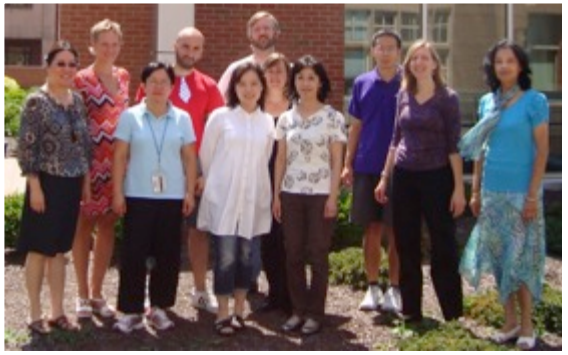
Local Festivals

CIPR Seminars

Educational Materials

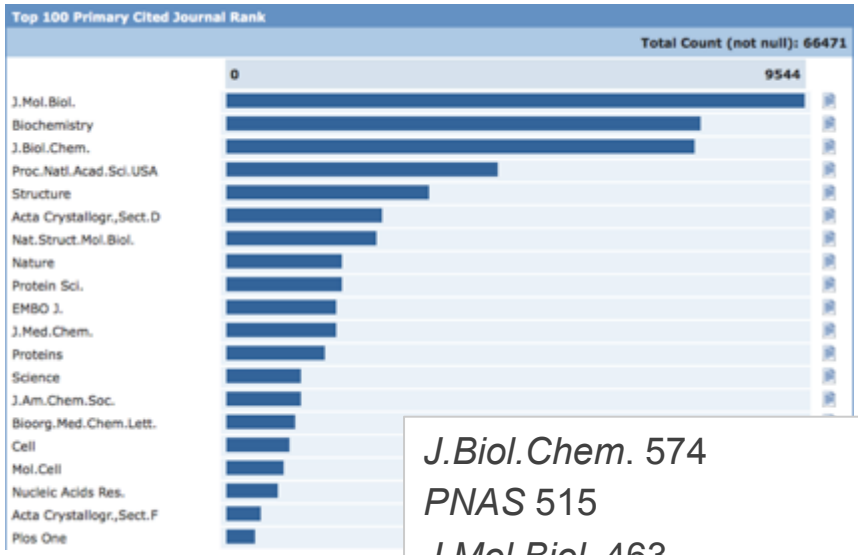
Staff Activities





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

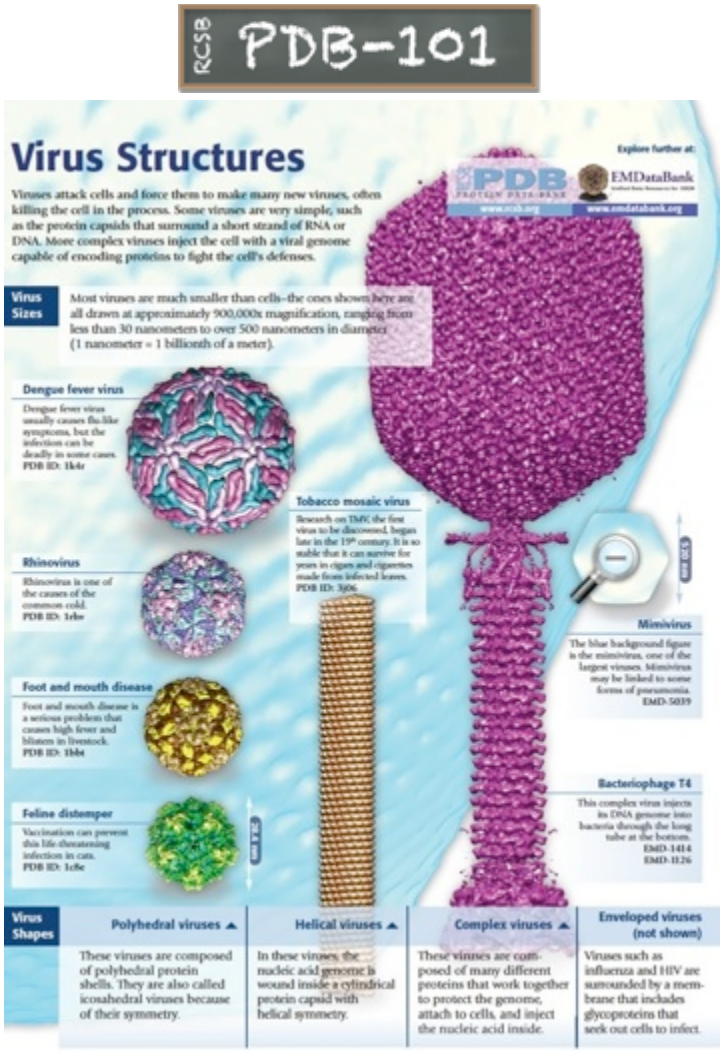
jbc THE JOURNAL OF BIOLOGICAL CHEMISTRY

Future: eLIFE

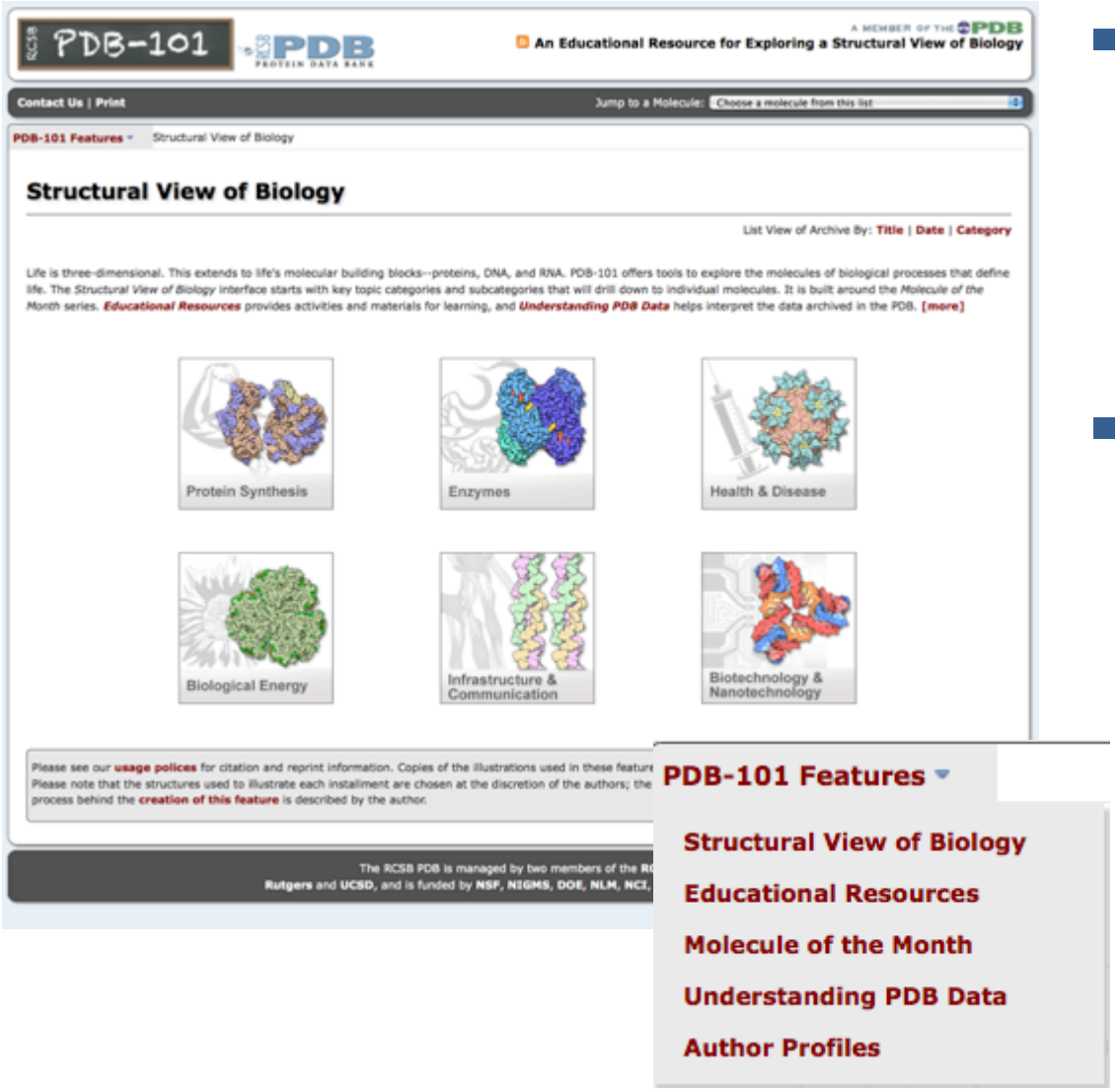
Notification of Structure Publications

- 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



- 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

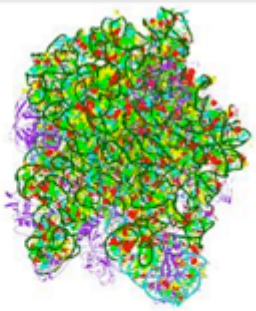
3PIP November 07

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 glucoamylase (BACUNI_03963) from Bacteroides uni...

4GHB August 07

Molecule: Putative uncharacterized prote...

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

4GDZ August 01

Molecule: Putative uncharacterized prote...

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

4GCM July 30

Thioredoxin reductase

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

Sort by

Newest

Oldest

Year

2013-2010

2012 (67)

2011 (102)

2010 (183)

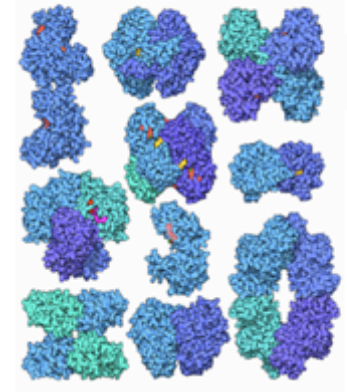
2010-2005

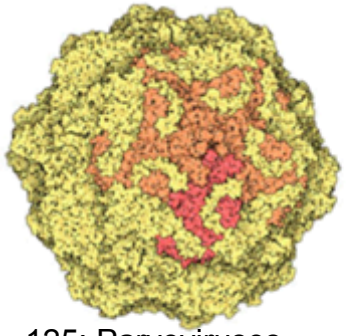
2005-2002

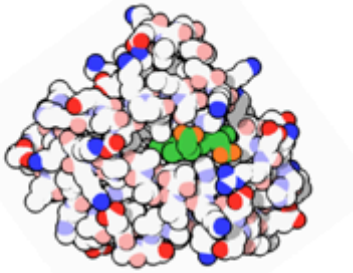
Molecule of the Month

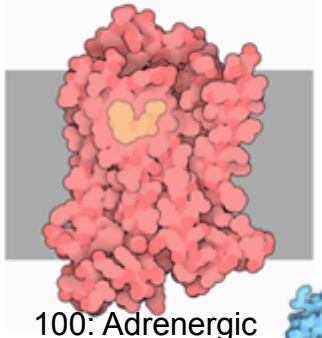
- 150th article published June 2012
- Top accessed articles July 2011–June 2012:

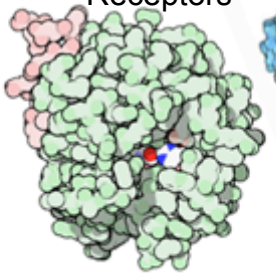
Title	Page views
Hemoglobin	34081
Collagen	21937
GFP	21709
Catalase	19207
Caspases	17755
Alpha-amylase	16597
DNA Polymerase	16392
Insulin	15909
Alcohol Dehydrogenase	15853
Lysozyme	14890

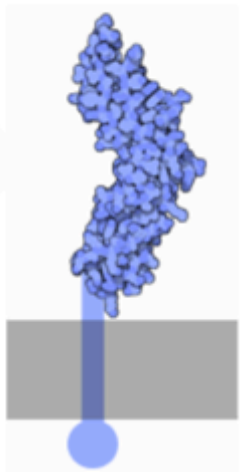

50: Glycolytic Enzymes

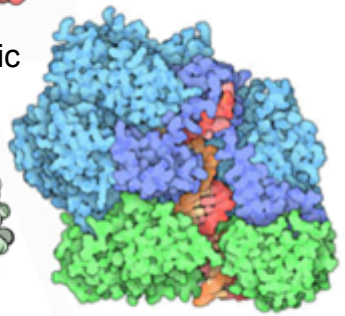

125: Parvoviruses


1: Myoglobin


100: Adrenergic Receptors

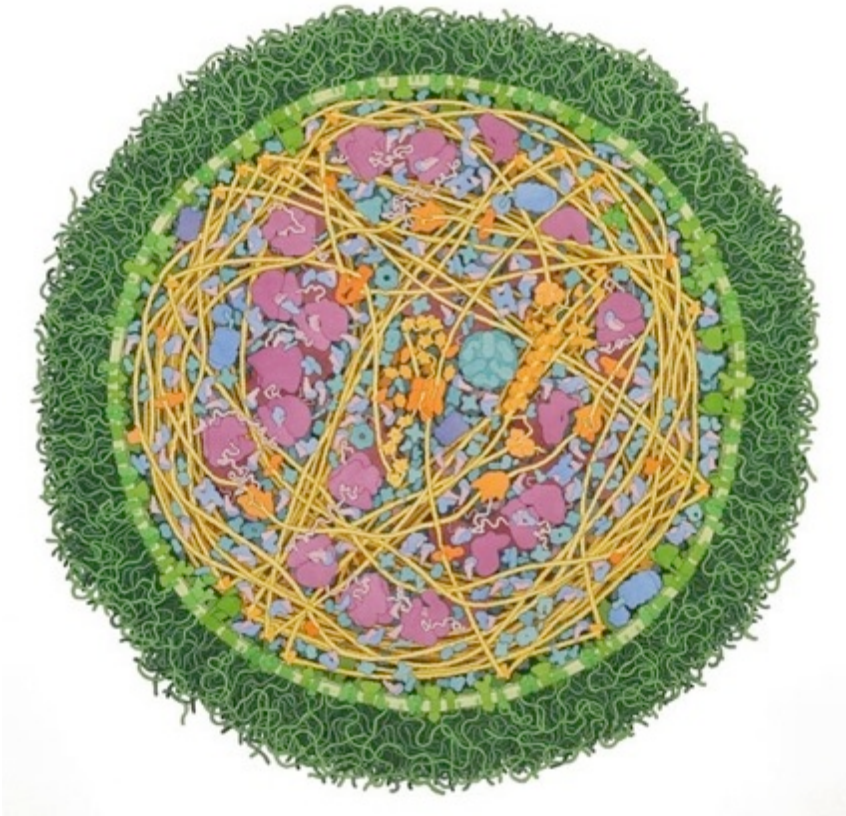

25: Thrombin


75: Tissue Factor


150: Sliding Clamps

Future

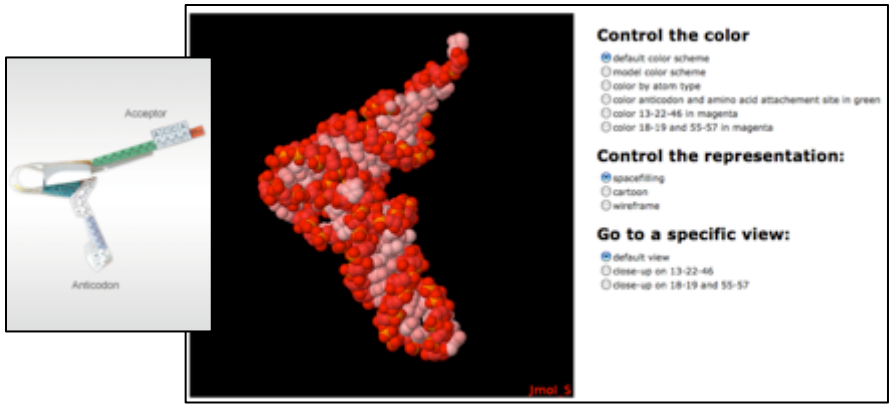
- Improved access and display of different PDB-101 sections
- Redesign of *Molecule of the Month* pages to improve readability
- Creation of cell-based poster and related online display
- 2013 look at *Molecular Machinery*



Mycoplasma cell

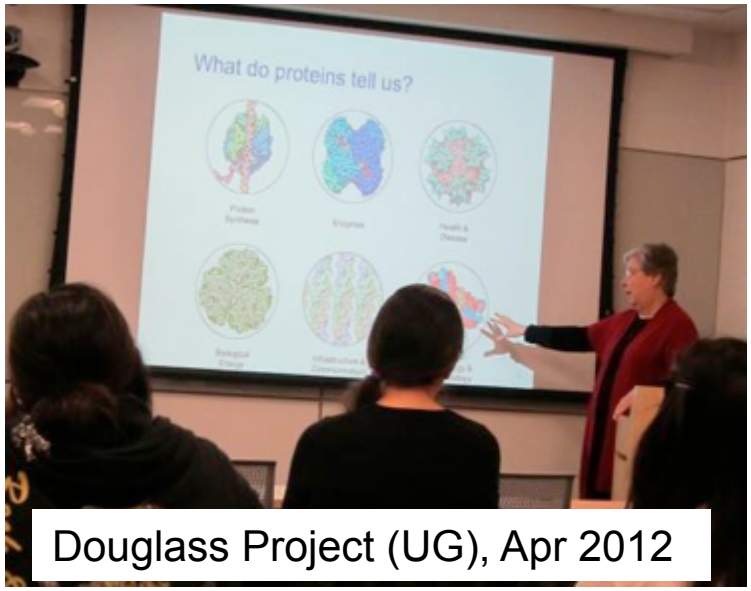
Teaching and Learning: A Structural View of Biology

- Booths, demonstrations at Society meetings
- Online resources

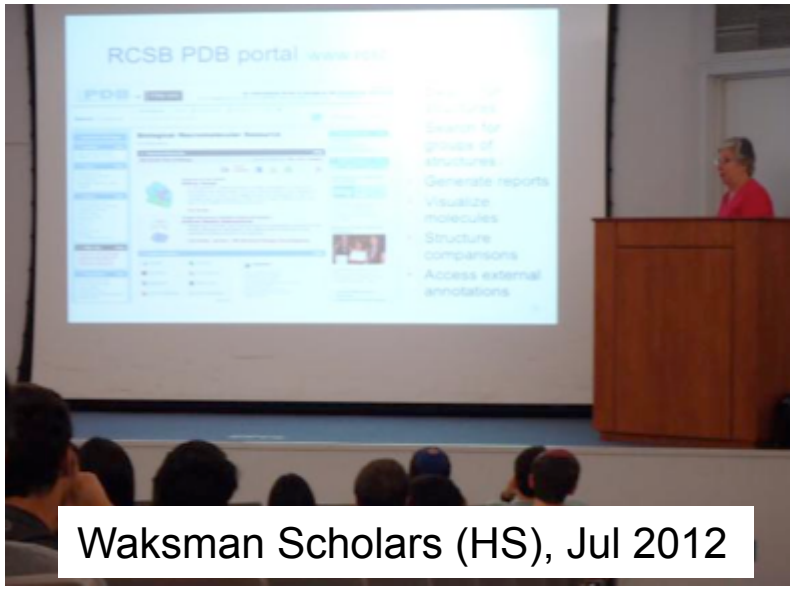


NJSC 2012, Oct. 9-10

- Presentations



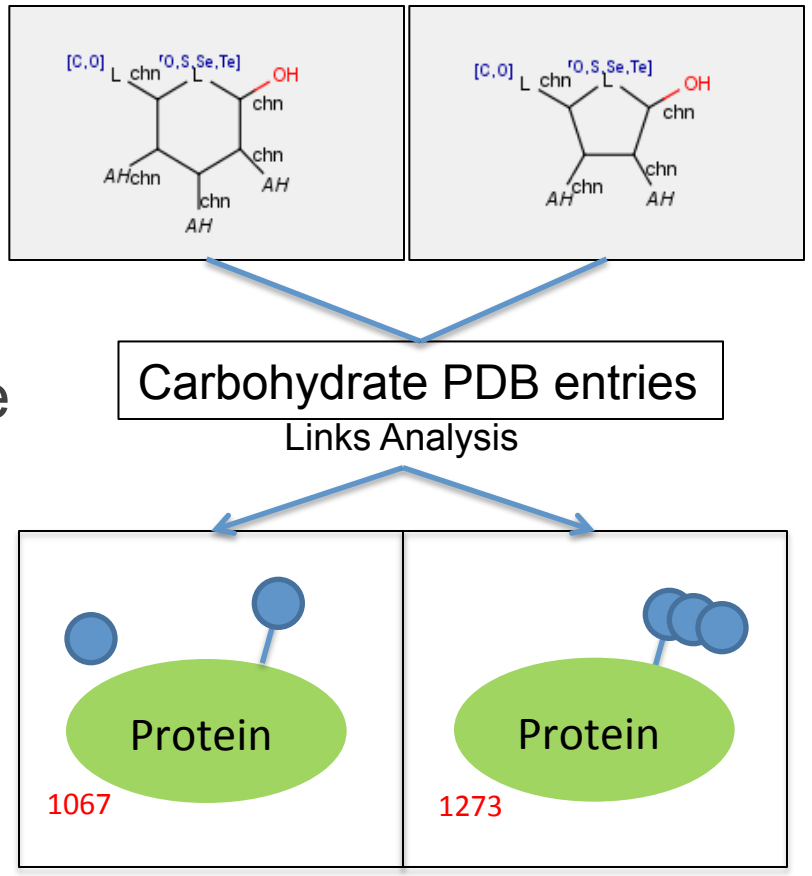
Douglass Project (UG), Apr 2012

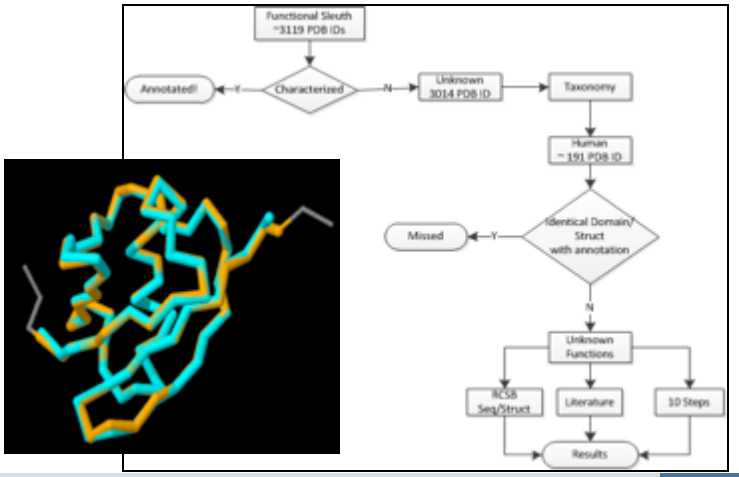


Waksman Scholars (HS), Jul 2012

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

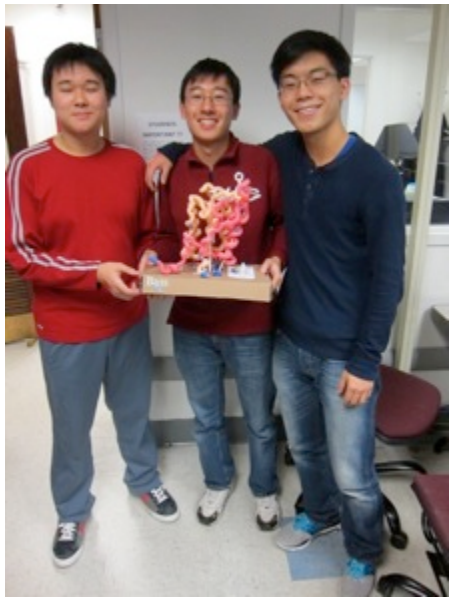


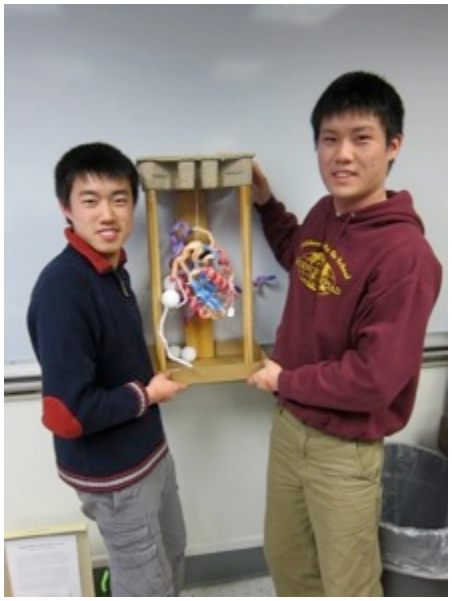


Protein Modeling at the 2012 Science Olympiad

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

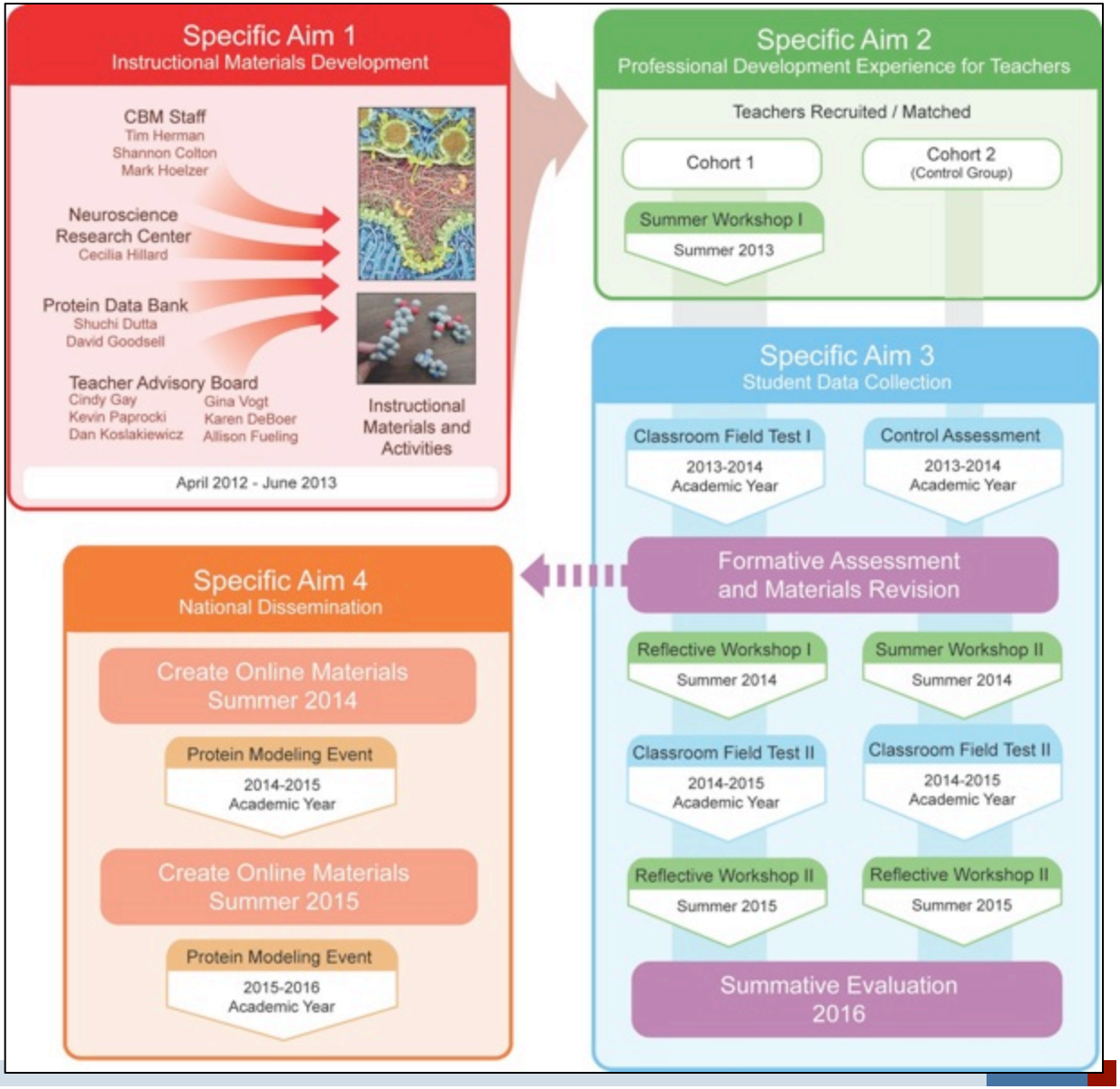
Drugs, Drug Targets and You: Exploring the Molecular Mechanisms of Drug Abuse with Tactile Teaching Tools

NIDA NATIONAL INSTITUTE ON DRUG ABUSE
The Science of Drug Abuse & Addiction

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

MSOE

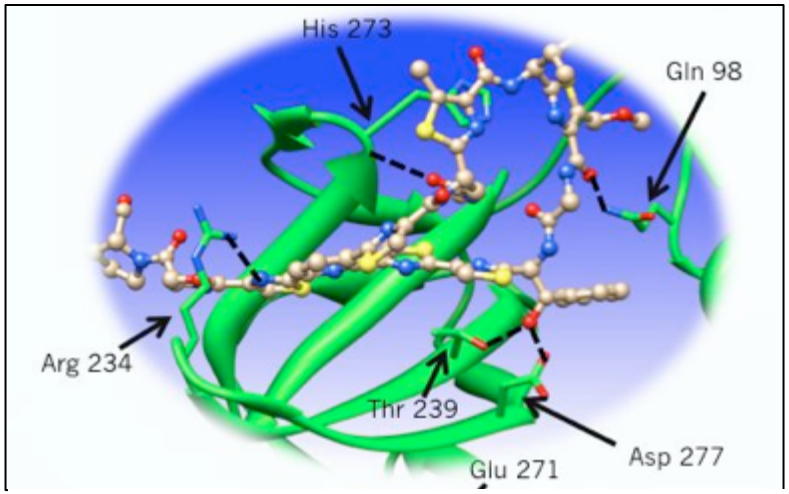
RCSB PDB-101



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

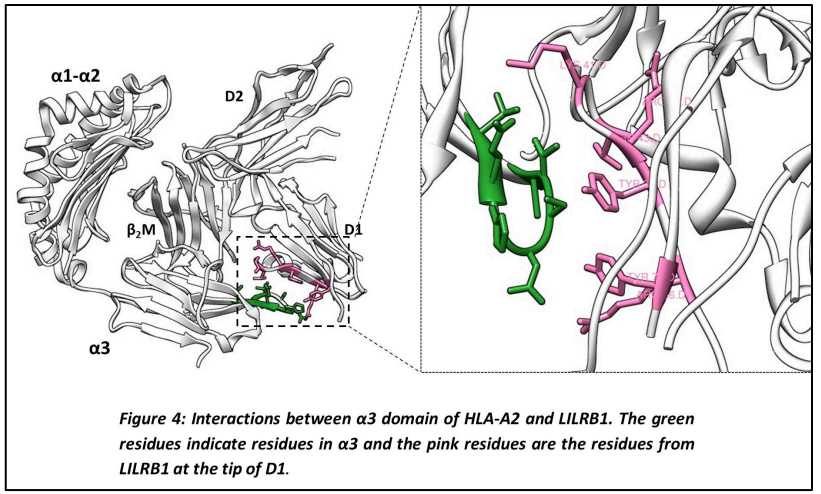


Figure 4: Interactions between α3 domain of HLA-A2 and LILRB1. The green residues indicate residues in α3 and the pink residues are the residues from LILRB1 at the tip of D1.

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

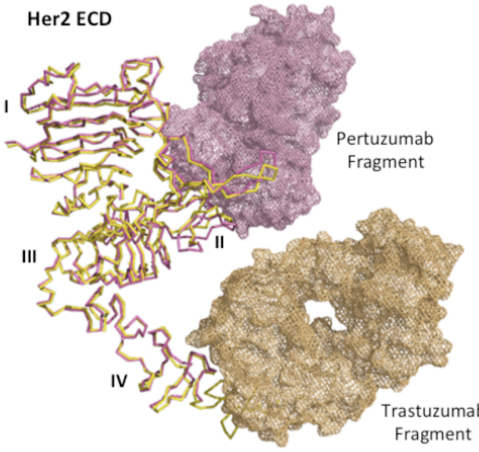


Fig.9. Binding of monoclonal antibodies Pertuzumab and Trastuzumab to human Her2 ECD. Superposition of human Her2 ECD shown in yellow and magenta ribbon show structural alignment. Trastuzumab fab is shown as mustard yellow mesh. Pertuzumab fab is shown in magenta mesh. (PDB 1N8Z, 1S78)

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

RCSB PDB PROTEIN DATA BANK
RUTGERS

BioMaPS Institute
and
Graduate Program in
Computational Biology & Molecular Biophysics

UCSD
SDSC • SKAGGS SCHOOL of PHARMACY and PHARMACEUTICAL SCIENCES