Worldwide Protein Data Bank Advisory Committee Meeting

October 2nd 2015

WORLDWIDE PDB
PROTEIN DATA BANK

wwpdb.org

Agenda

9:00 am	Welcome	
9:05	Overview/State of the wwPDB and Outreach	Haruki Nakamura
9:45	Milestones, Improvements, and Impact of the Common Deposition & Annotation (D&A) Tool on X-ray Depositions	Stephen Burley
10:35	*Break*	
10:50	NMR: PDB – BMRB	John Markley
11:40	EM: PDB – EMDB – EMPIAR	Sameer Velankar
12:30 pm	*Lunch*	
2:00	D&A V2.0 progress	
2:30	Looking Ahead, Questions, and Discussion	
3:45	*Group photo, followed by break*	
4:00	Executive Session & Feedback	
5:00	*Adjourn*	

Welcome

Haruki Nakamura

WORLDWIDE PDB
PROTEIN DATA BANK

wwpdb.org

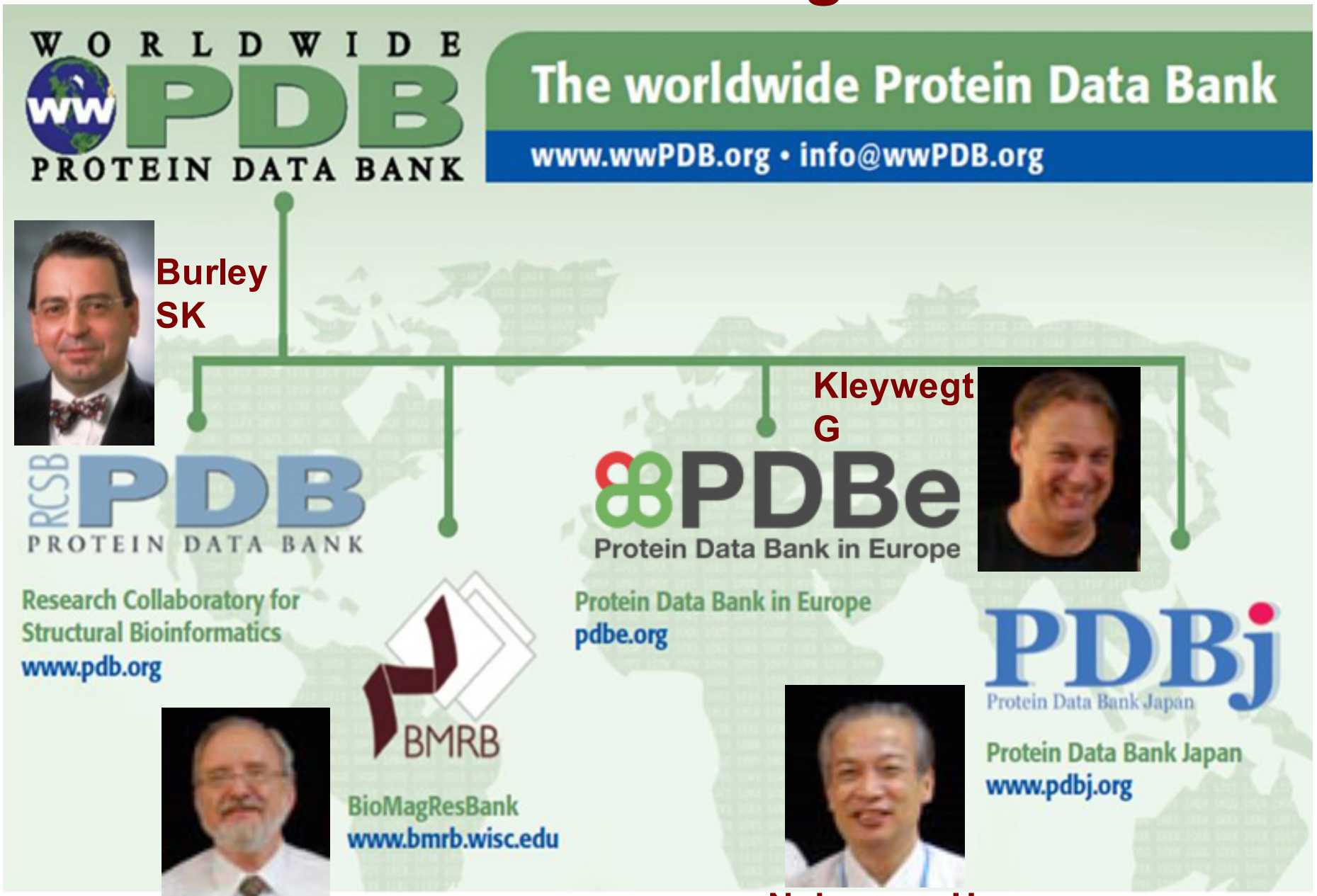
Overview/State of the wwPDB and Outreach

Haruki Nakamura

WORLDWIDE PDB
PROTEIN DATA BANK

wwpdb.org

wwPDB.org



wwPDB partners and their heads

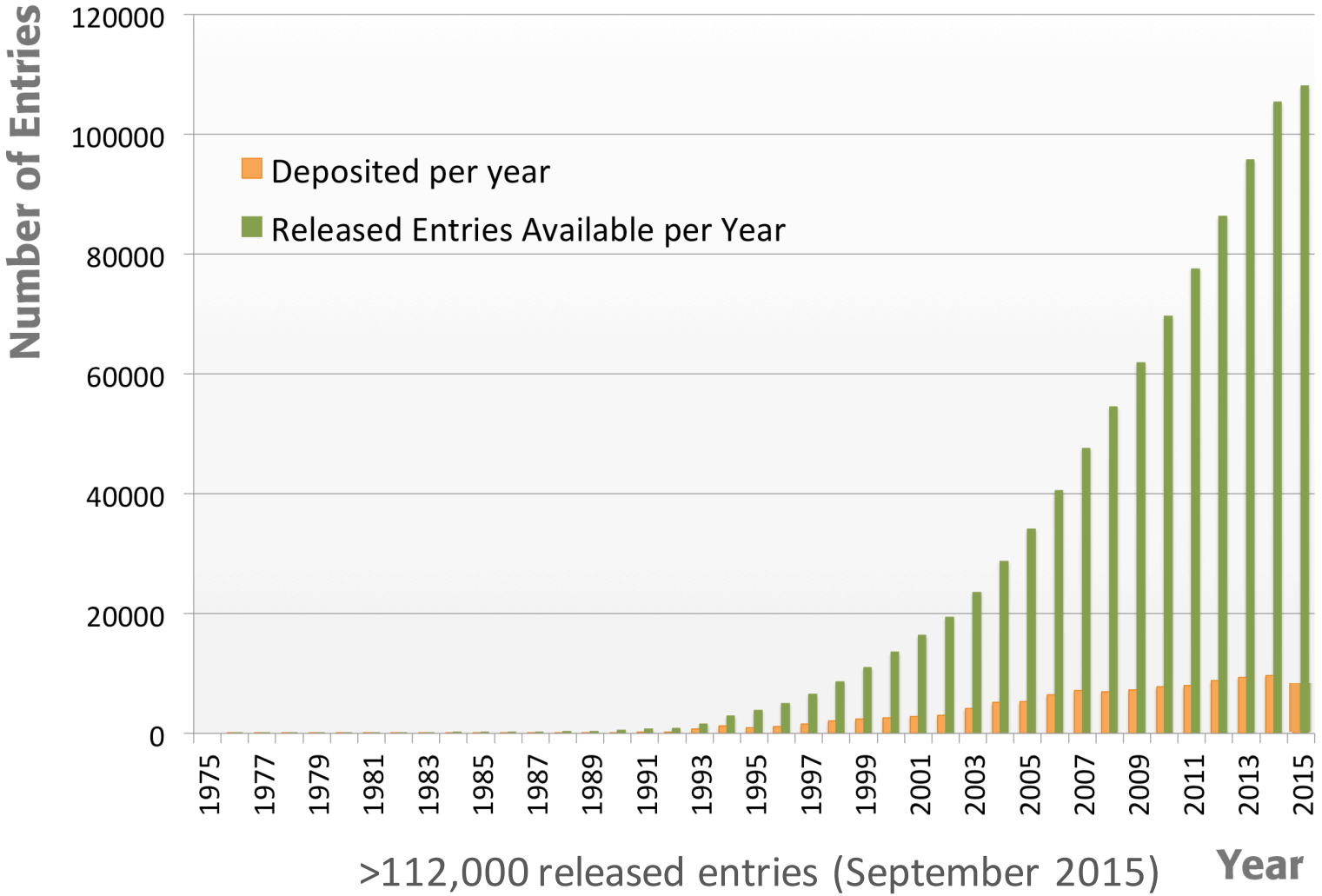
wwPDB
October 2014 - September 2015

- Continued growth of archive: more than 112,000 structures
- Increased use of data
- New wwPDB Web pages
- Release of sequence/bound ligand information with crystal pH prior to full data release
- Data correspondences between the PDB and CSD archives
- More and more large structures released as PDBx files (337 entries as of Sept. 23)
- Common Deposition & Annotation System improved
- Proposals: New Author Identifiers, Federation of Data Resources, Versioning, Inclusion of Editorial Comments in the PDB Archive
- Workshops and meetings
  - Hybrid Methods Task Force

    Support letter to Helen M. Berman and Andrej Sali for NSF application

    "Methods for Validation and Annotation of Integrative/Hybrid Models"
  - NMR Restraints Format Meeting and Working Group
  - Ligand Validation Workshop
- Funding
- Outreach
- wwPDB Foundation activity

Growth of Total PDB entries



>112,000 released entries (September 2015)

>10,000 new entries projected for 2015

PDB Depositions

- As of 2015 region-based processing of D&A-deposited entries:
  - RCSB PDB: Americas & Oceania
  - PDBe: Europe & Africa
  - PDBj: Asia

Year	Total Depositions	Processed By		
		RCSB PDB	PDBj	PDBe
2000	2983	2297	158	528
2001	3287	2408	383	496
2002	3565	2401	657	507
2003	4830	3135	1026	669
2004	5508	3082	1614	812
2005	6678	3563	2110	1005
2006	7282	4252	1945	1085
2007	8130	4703	2299	1128
2008	7073	4106	1994	973
2009	8300	5069	2173	1058
2010	8878	5464	2041	1373
2011	9250	5938	1816	1496
2012	9972	6408	1888	1676
2013	10566	6652	2128	1786
2014	10364	6040	1779	2545
2015	7575	3492	1336	2748
TOTAL	114241	69010	25347	19885

Last Updated: 15 September 2015

RCSB PDB PROTEIN DATA BANK PDBe Protein Data Bank in Europe PDBj Protein Data Bank Japan

FTP + Rsync Entry Downloads

RCSB PDB

2014: 347 million
2013: 312 million
2012: 298 million
2011: 282 million
2010: 159 million

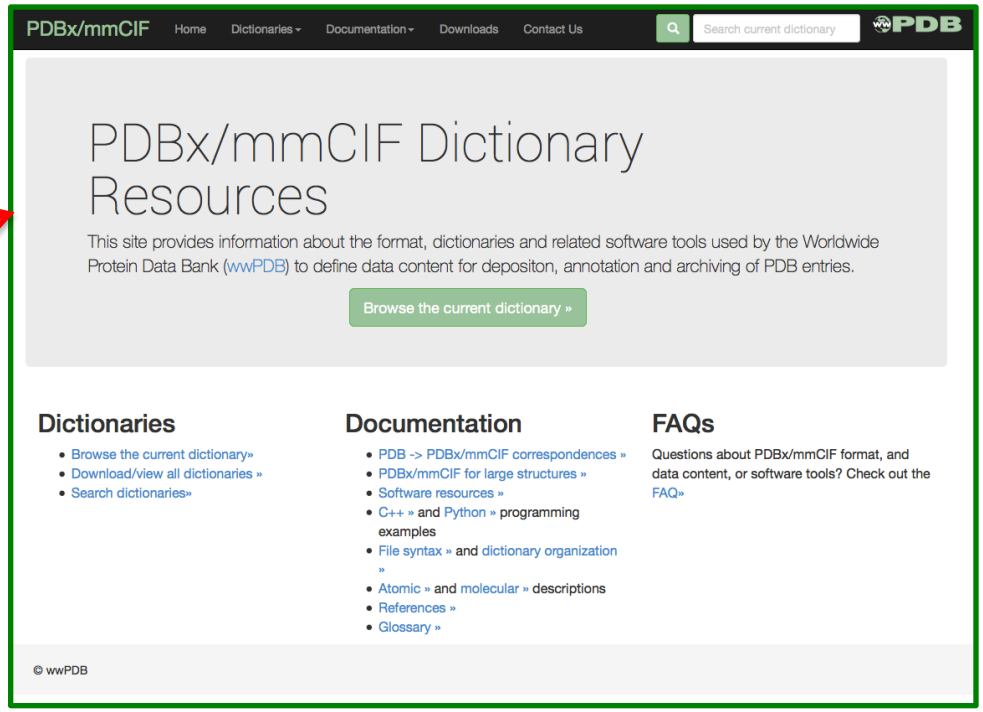
PDBe

2014: 100 million
2013: 81 million
2012: 46 million
2011: 59 million
2010: 34 million

PDBj

2014: 57 million
2013: 40 million
2012: 21 million
2011: 38 million
2010: 16 million

Updating the wwPDB web page



wwpdb.org

Release of sequence and bound ligand information prior to release of structure

- Phase I: Every Saturday by 3:00 UTC, for every new entry, the wwPDB website provides: sequence(s) (amino acid or nucleotide) for each distinct polymer and, where appropriate, the InChI string(s) for each distinct ligand and the crystallization pH value(s)

- Phase II: Every Wednesday by 00:00 UTC, all new and modified data entries will be updated at each of the wwPDB FTP sites.

(Started in April 2015. Crystallization pH information was added in August 2015)

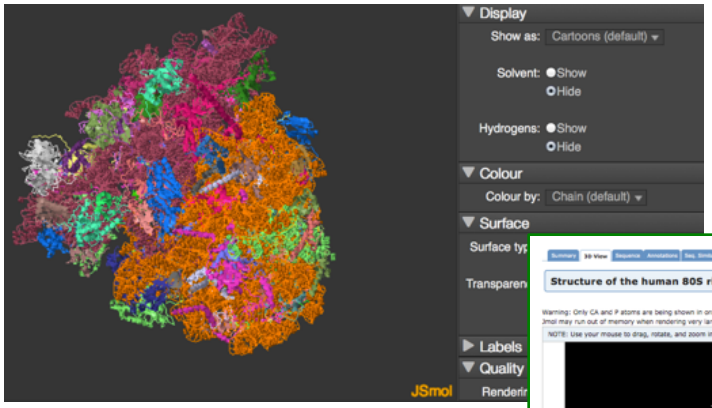
Data correspondences between the PDB and CSD archives

- A new data resource is available containing correspondences between the biopolymer components and ligand molecules found in the PDB Chemical Component Dictionary (CCD) that exactly match small-molecule X-ray structures in the Cambridge Structural Database (CSD) archive

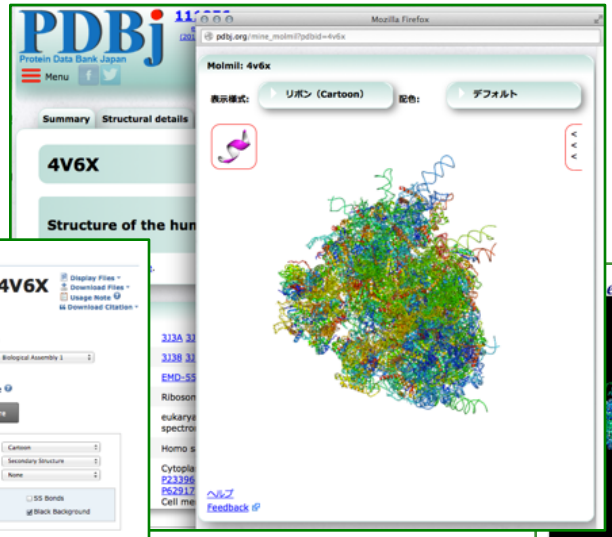
- The new PDB Chemical Component Model data file complements information in the PDB by providing the following CSD information for matching molecular entries: accession code correspondences, Cartesian coordinates and R-value, data-collection temperature and a disorder flag, SMILES and InChI descriptors, and a Digital Object Identifier (DOI) for the citation associated with the CSD entry.

Large Structures

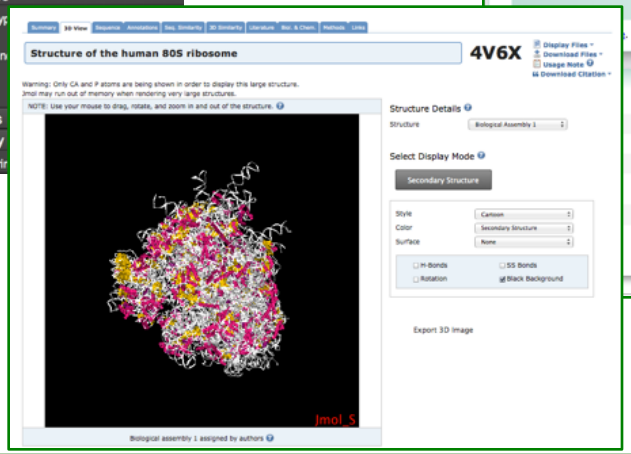
- As of December 2014, large structures are no longer split into multiple PDB entries but are released as single PDBx and PDBML files
  - Existing split entries were consolidated into single entries
  - A bundle of best-effort PDB files (PDB bundle) is available for every large structure through the wwPDB FTP sites
- Each wwPDB member provides tools to view large structures

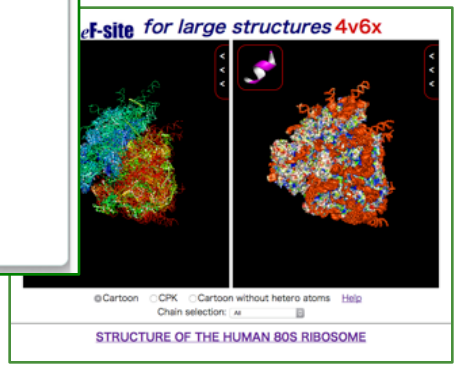


PDBe



PDBj

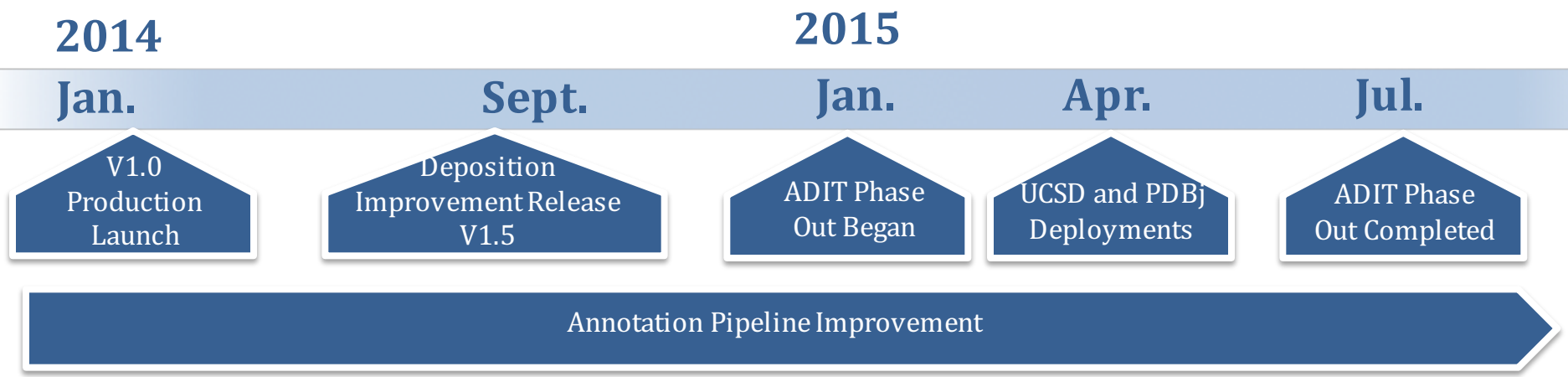




RCSB-PDB

4v6x: structure of human 80S ribosome, 89 chains, 4 MDa

X-ray D&A Milestones



- January 2014: X-ray production release (V1.0)
- Fall 2014: X-ray improvement release (V1.5)
- January 2015: RCSB PDB and PDBj ADIT systems phase out began
- April 2015: System deployments at UCSD and PDBj
- July 2015: RCSB PDB and PDBj ADIT systems phase out concluded
- 2015: Annotation pipeline improvement throughout year

Announcement of ADIT system phasing out for Asia-Pacific region in January 2015

- The PDBj constructed the Asian language pages for introduction and tutorial of the wwPDB Deposition tools in Japanese, Korean, and Traditional/Simplified Chinese

- The PDBj sent mails of the announcement to:

  Pinak Chakraborty (President of AsCA) (India)

  Se Won Suh (Korea), David Hsiao (Taiwan)

  Jianping Ding (wwPDB-AC member) (China)

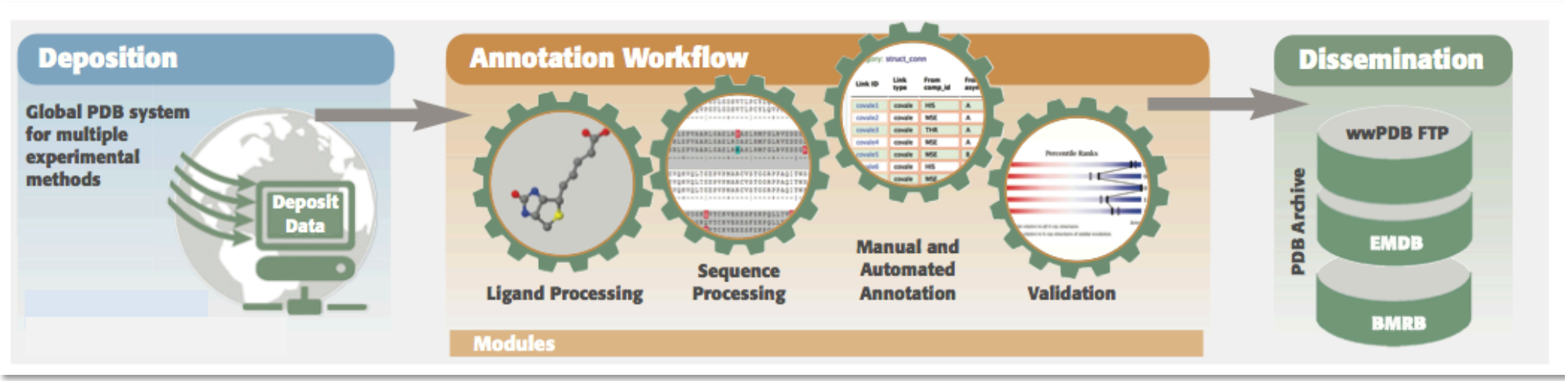
  Jenny Martin (vice-president of AsCA) (Australia)

  Edward Baker (wwPDB-AC member) (New Zealand)

  Robert Robinson (Singapore),

  Jame R. Ketudat-Cairns (Thailand)

2015 Annotation Improvements



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

D&A V2.0 not yet in production
Roadmap of V2.0 is now updated

New Author Identifiers: ORCID

- Rationale

The Open Researcher and Contributor ID (ORCID; orcid.org) is a global digital identifier that can be used to uniquely identify researchers and link them to their research work product

With increasing numbers of depositions coming from China and India, it is not always possible for the wwPDB to distinguish distinct deposition contact authors with identical name spellings in English

- Proposed Solution

With definitive deployment of D&A V2.1, the wwPDB proposes to capture an ORCID identifier for each contact author at the time of deposition

This step will allow us to begin the process of creating a unique list of PDB contributors

A placeholder definition for the ORCID is already available in the PDBx/mmCIF V5 dictionary, *_pdbx_contact_author.identifier_ORCID*

In the future we may take further advantage of the public ORCID resource by retrieving metadata to simplify depositor's input of personal information

Federation of Data Resources for integrative/hybrid 3D structural models

- Following-up on the Outcome of the Hybrid/Integrative Methods Task Force Workshop

  The wwPDB aims to provide a Deposition and Annotation System for collecting 3D structure model coordinates plus primary data and associated metadata (sufficient to validate the 3D structural model) that encompasses currently supported methods (X-ray, Neutron, NMR, and EM) plus additional methods specific to Federated Data Resources

- Data Representation and Exchange Protocols

  The wwPDB and each Federated Data Resource will agree on data content for exchange and a common exchange protocol based on the PDBx/mmCIF dictionary for method specific primary data and associated metadata (sufficient for model validation) between the wwPDB Deposition and Annotation System and the Federated Data Resource

- Data Storage Formatting

  PDBx/mmCIF is the principal archival format for storage of 3D structure model coordinates, associated metadata, and selected primary data supporting model validation in the PDB Archive.

  Federated Data Resources are free to store their method specific primary data and associated metadata in whatever way they choose to do so.

Versioning of PDB entries

- Policy Proposal
  - Every accession code includes the version number after the PDB code
  - The version number is updated, only when the group of the original author(s) with the same PI revises the atomic coordinates, in addition to change ligand codes or any other metadata
  - When new experimental data are used to change the atomic coordinates, a new PDB code is adopted instead of updating the version number
  - When another group contributes an extension or reinterpretation without any additional experimental data, such a structure can be deposited with a new PDB code, only after publication from a peer-reviewed journal
  - Incorporates the archive name (PDB) explicitly within the accession code to simplify recognition in scientific literature

[Examples of proposed versioning]

```
Template:  PDB-<pdb_code>-<version_number>
Example:   PDB-1ABC-1  (original structure)
           PDB-1ABC-2  (updated structure)
Regex:     PDB-[0-9][A-Z0-9]{3}-[0-9]{1,n}   (n~3)
```

Inclusion of Editorial Comments in the PDB Archive

- Rationale

On occasion, current wwPDB Annotation Practices result in depositions containing apparently incorrect annotations provided by the contact author with no correct alternative annotation provided by automated software tools (e.g., Quaternary Structure)

- Proposed Solution

The wwPDB annotators will include a wwPDB annotation marked as such that provides the missing correct alternative annotation

wwPDB Task Forces

Method-specific (Validation) Task Forces have been convened to collect recommendations and develop consensus on additional validation that should be performed, and to identify software applications to perform validation tasks, and to discuss archival needs and opportunities for non-traditional techniques.

Task Force	Meeting/ Workshop	Chair(s)/Membership	Outcme
X-ray Validation Task Force	2008 (2015)	Randy Read (Univ of Cambridge) 17 members	(2011) *Structure* 19: 1395-1412
NMR Validation Task Force	2009, 2011, 2013 (x2), 2015	Gaetano Montelione (Rutgers) Michael Nilges (Institut Pasteur) 10 members	(2013) *Structure*, 21: 1563-1570
3DEM Validation Task Force	2010	Richard Henderson (MRC-LMB) Andrej Sali (UCSF) 21 members	(2012) *Structure* 20: 205-214
Small-Angle Scattering Task Force	2012, 2014	Jill Trewhella (Univ Sydney) 6 members	(2013) *Structure* 21: 875-881
Hybrid Methods Task Force	2014	Andrej Sali (UCSF), Torsten Schwede (Univ Basel), Jill Trewhella (Univ Sydney) 27 members	(2015) *Structure* 23: 1156-1167

wwPDB NMR Validation Task Force

- Meetings in 2009 (Paris), 2011 (NJ), 2013 (NJ, VT and EMBL-EBI) and 2015 (NJ)
  - GT Montelione, M Nilges, A Bax, P Güntert, T Herrmann, JS Richardson, C Schwieters, WF Vranken, GW Vuister, DS Wishart
- New members joined in 2015 to provide expertise on solid-state NMR and nucleic acids: RA Byrd, T Polenova, S Butcher

  Montelione et al., Structure 21, 1562 (2013)
- Still to do:
  - Make the validation reports public
  - Provide as stand-alone server
  - Incorporate restraint validation and eventually peak lists

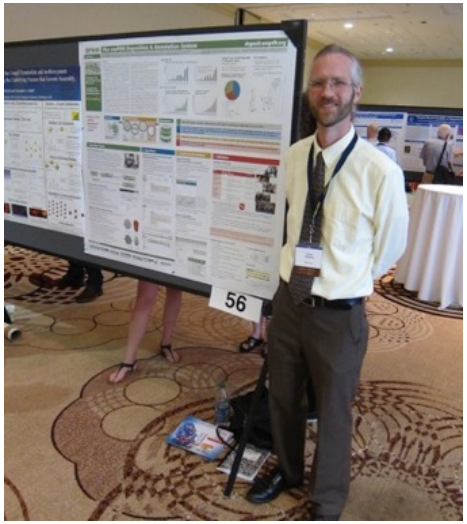
wwPDB/CCDC/D3R Ligand Validation Workshop

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

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

wwPDB Outreach

Annotator Meeting
PDBj hosts RCSB PDB's Jasmine Young, April 2015

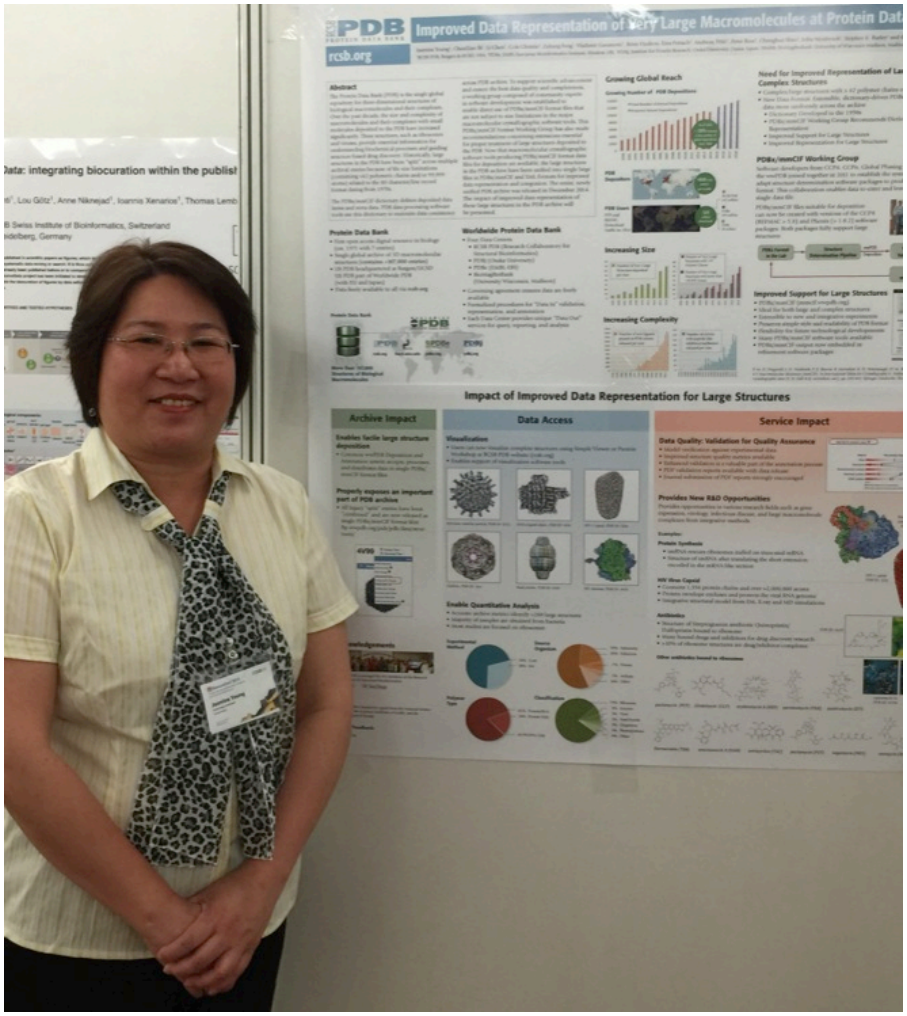


The wwPDB Deposition & Annotation System, **American Crystallographic Association, July 2015**

Recent wwPDB Developments
NY Structural Biology Group, Jan 2015

The wwPDB Symposium October 2015

Outreach: RCSB PDB



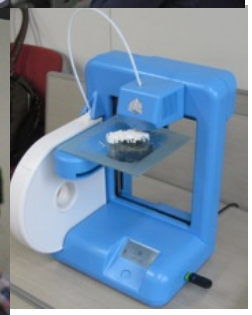
Improved Data Representation of Large Macromolecules at PDB, **Biocuration, April 2015**

Science Olympiad**, Winter 2015**

Society for Advancement of Chicanos/Hispanics and Native Americans in Science (SACNAS)**, October 2014**

Outreach: PDBj & PDBe

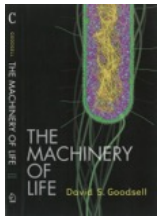
Workshop for young students using 3D printer, Osaka Univ. May 2015

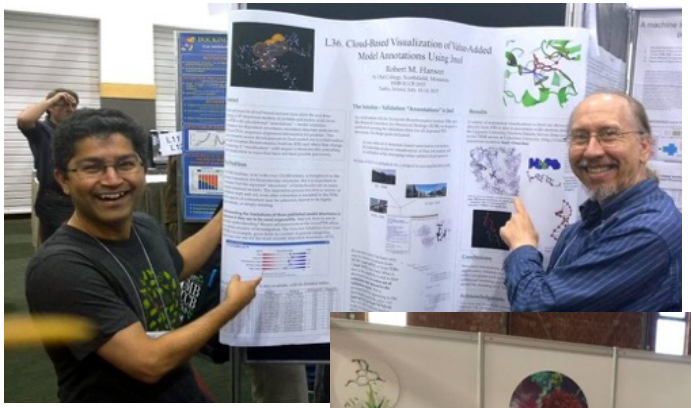


API workshop for programmers

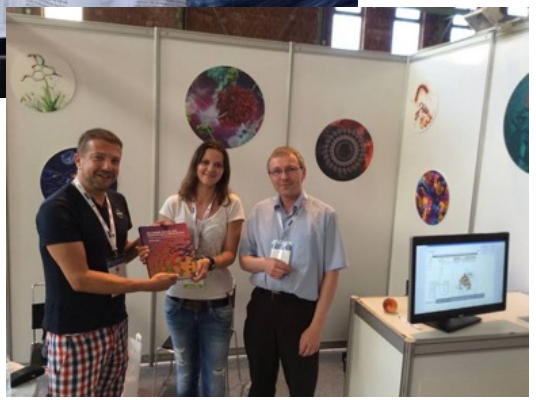
Book Publication
The Machinery of Life
By D.S. Goodsell

Translation for Japanese readers
March 2015





ISMB



ECM

Outreach: BMRB

NMRFAM beginner's workshop
(May 31 – June 5, 2015)

NMRFAM Advanced Workshop
(August 23 – 26, 2015)

Funding

- RCSB PDB non-competitive renewal funded by NSF (2014-2018)
- PDBe main funding from EMBL and Wellcome Trust
  - EMBL-EBI: core funding for ~13 posts
  - Wellcome Trust: competitive Biomedical Resource grant (4 posts, 2015-2019)
  - Additional new project grants from BBSRC and EU
  - Quadrennial review by EMBL SAC – "*outstanding*"
- PDBj competitive renewal funded by JST (Japan Science & Technology Agency) for April 2014 - March 2017
- BMRB is now funded by NIH-NIGMS (2014-2019)

Publications

- A. Sali *et al*. (2015) Outcome of the First wwPDB Hybrid/Integrative Methods Task Force Workshop. *Structure* **23**: 1156–1167, doi: 10.1016/j.str.2015.05.013
- A. Gutmanas *et al*. (2015) NMR Exchange Format: a unified and open standard for representation of NMR restraint data. *Nature Structural & Molecular Biology* **22:** 433–434, doi: 10.1038/nsmb.3041
- J. D. Westbrook *et al*. (2015) The chemical component dictionary: complete descriptions of constituent molecules in experimentally determined 3D macromolecules in the Protein Data Bank. *Bioinformatics* **31:** 1274-1278, doi: 10.1093/bioinformatics/btu789
- S. Sen *et al*. (2014) Small molecule annotation for the Protein Data Bank. *Database* **2014**: doi: 10.1093/database/bau116

Worldwide
Protein Data Bank
Foundation

- Board of Directors (BoD) Membership
  - Chair: Anthony Nicholls
  - President: Helen Berman
  - Vice Presidents: Stephen Burley, Gerard Kleywegt, John Markley, Haruki Nakamura
  - Executive Director/Secretary/Treasurer: Christine Zardecki
- Fundraising on-going
- 2015: Supported wwPDB/CCDC/D3R Ligand Validation Workshop
- 2015: Support wwPDB Symposium for Hybrid Methods
- 2015: Support wwPDB X-ray VTF
- 2016: wwPDB Symposium on Data Mining/Statistics

Milestones, Improvements, and Impact of the Common Deposition & Annotation (D&A) Tool on X-ray Depositions

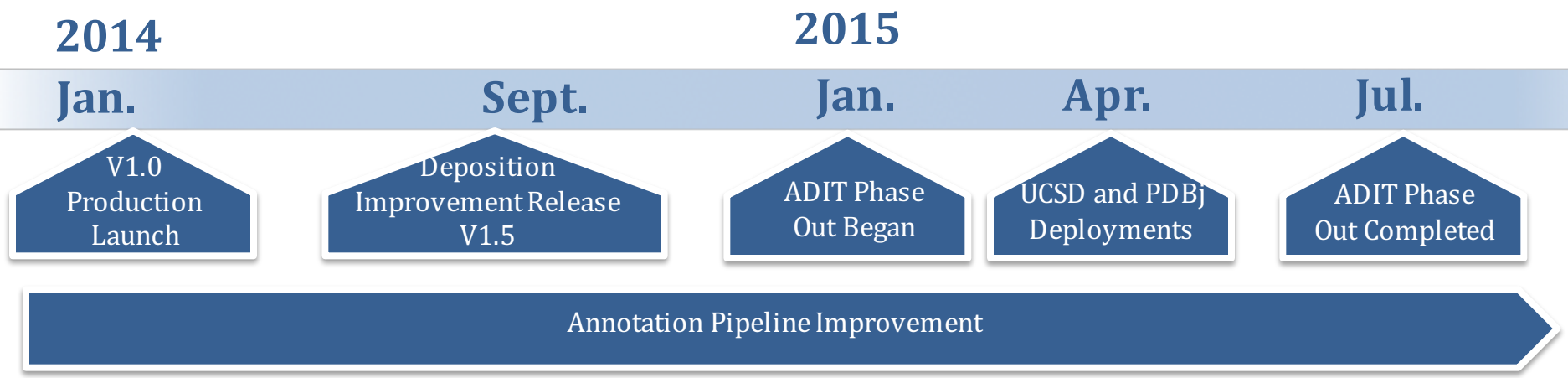
Stephen K. Burley

wwpdb.org

Agenda

- X-ray D&A milestones
  - System deployments at UCSD and PDBj
  - ADIT legacy systems phase out
  - System improvements
- Improvements and impact of D&A on X-ray
  - Utilization metrics
  - Increased efficiency
  - Increased data quality
  - User feedback

X-ray D&A Milestones



- January 2014: X-ray production release (V1.0)
- Fall 2014: X-ray improvement release (V1.5)
- January 2015: RCSB PDB and PDBj ADIT systems phase out began
- April 2015: System deployments at UCSD and PDBj
- July 2015: RCSB PDB and PDBj ADIT systems phase out concluded
- 2015: Annotation pipeline improvement throughout year

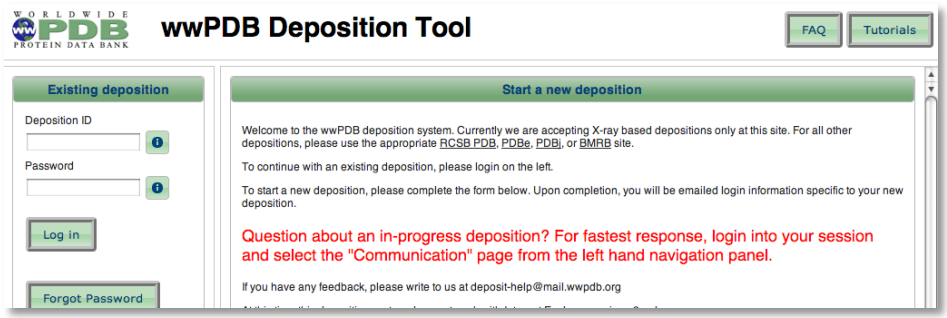
Deployments at UCSD and PDBj

RCSB PDB System at UCSD

- Hardware/software commissioned
- Multiple entries fully processed and released
- Standby for warm failover

PDBj system

- Hardware/software commissioned
- > 150 entries fully processed
- > 40 entries released



deposit-rcsb-west.wwpdb.org/deposition

deposit-pdbj.wwpdb.org/deposition/

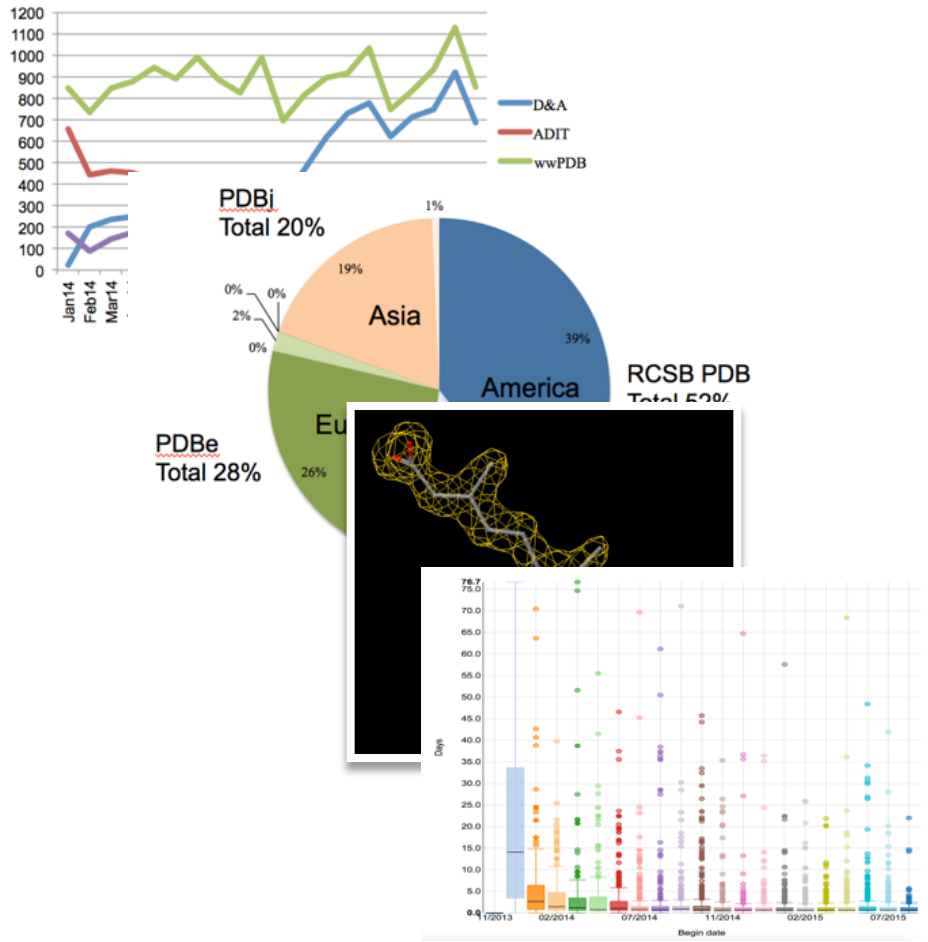
RCSB PDB/PDBj ADIT System Phase Out

- January 27 2015: Acceptance of new X-ray depositions terminated
- January-July 2015: Active sessions finalized
- July 19 2015: Full “retirement”

Brian Hudson
RCSB PDB

Improvements and Impact of D&A on X-ray Deposition

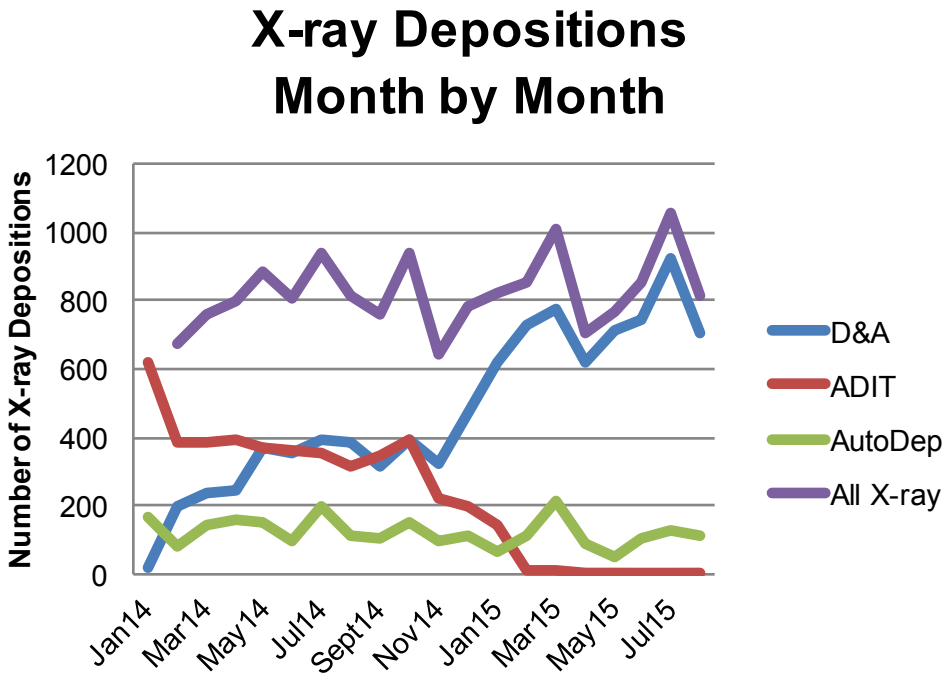
- Utilization metrics
- Improved data quality
- Improved efficiency
- User Feedback

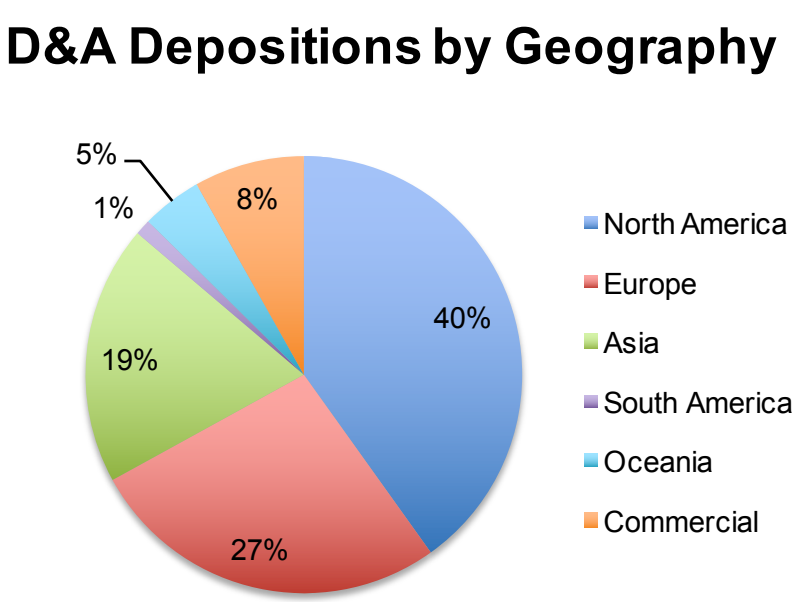


Utilization Metrics:

January 27 2014 – August 31 2015

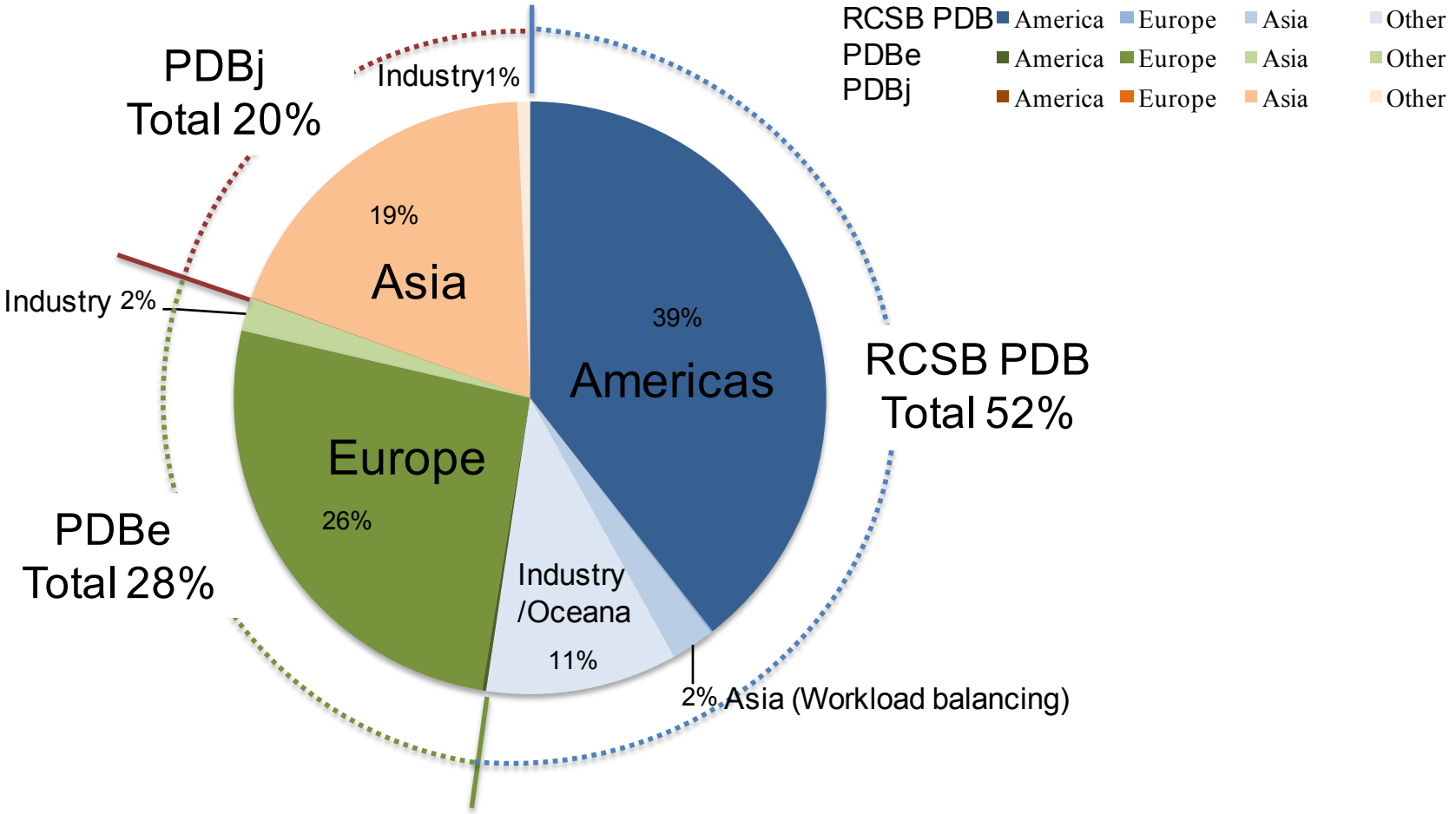
- More than 9500 structures deposited/annotated/validated
- More than 4800 entries released
- 56% of total global depositions during period





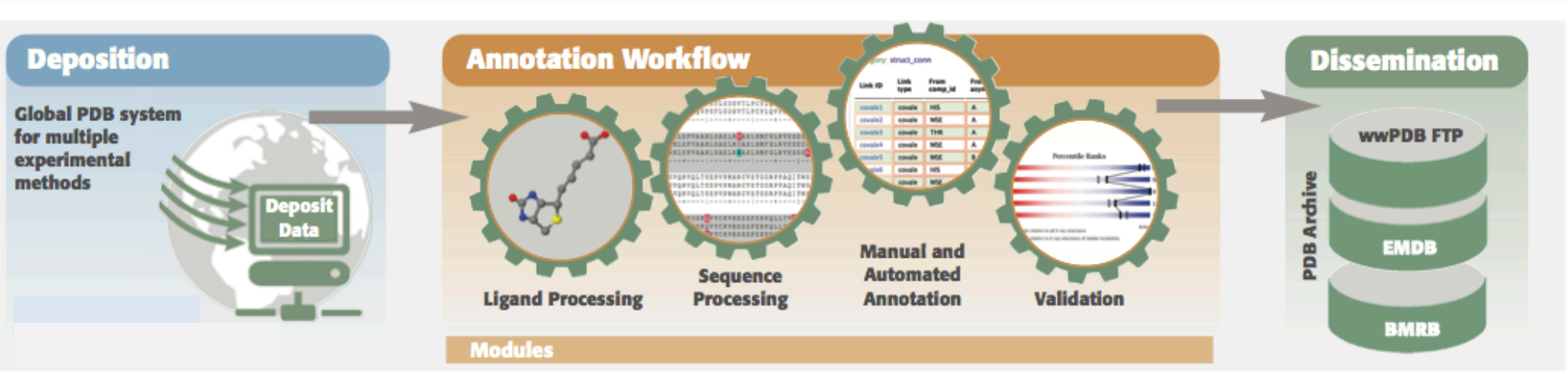
Workload Balancing in Action

September 2014 - August 2015



D&A Annotation Distribution by Processing Site and Geography

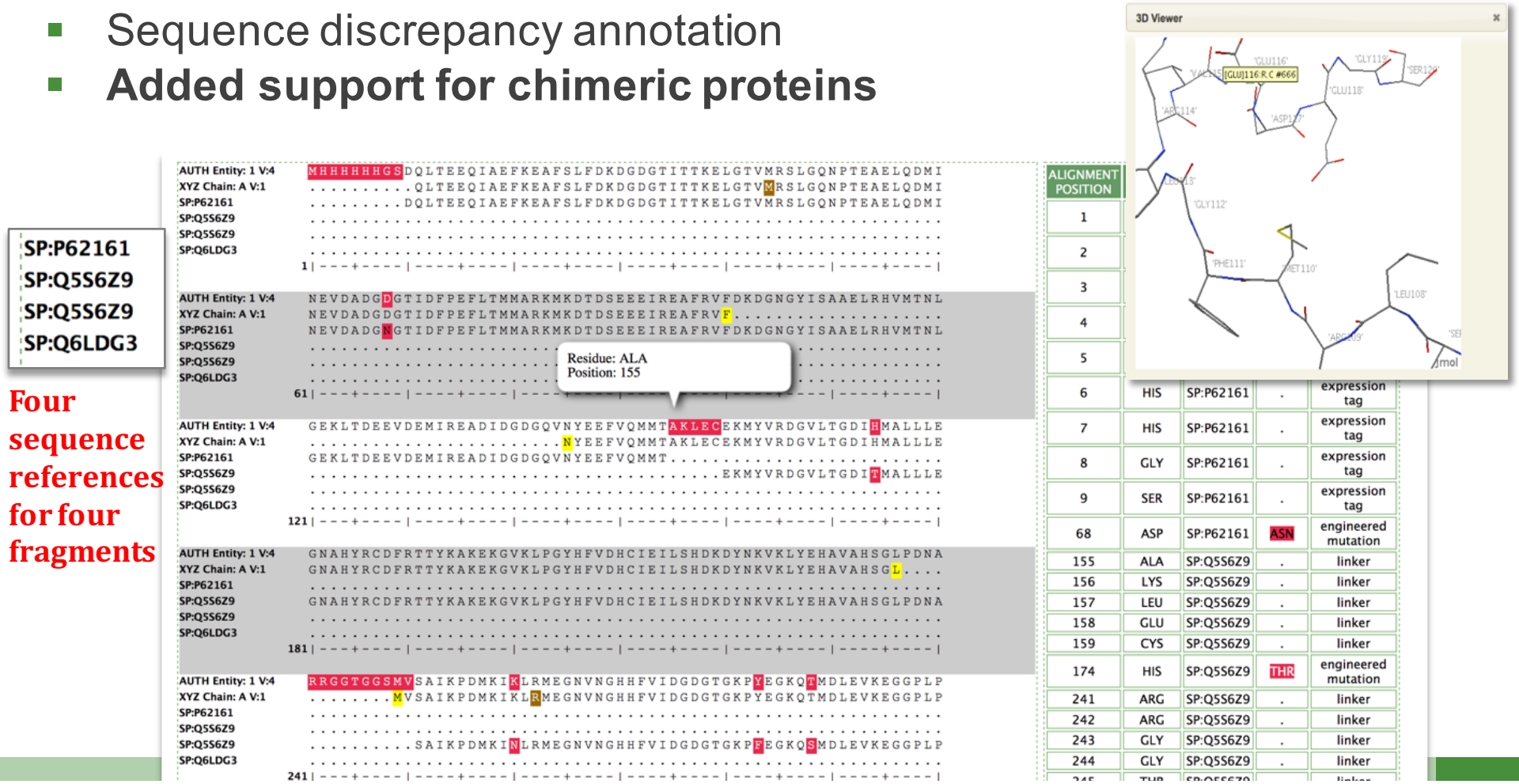
2015 Annotation Improvements



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

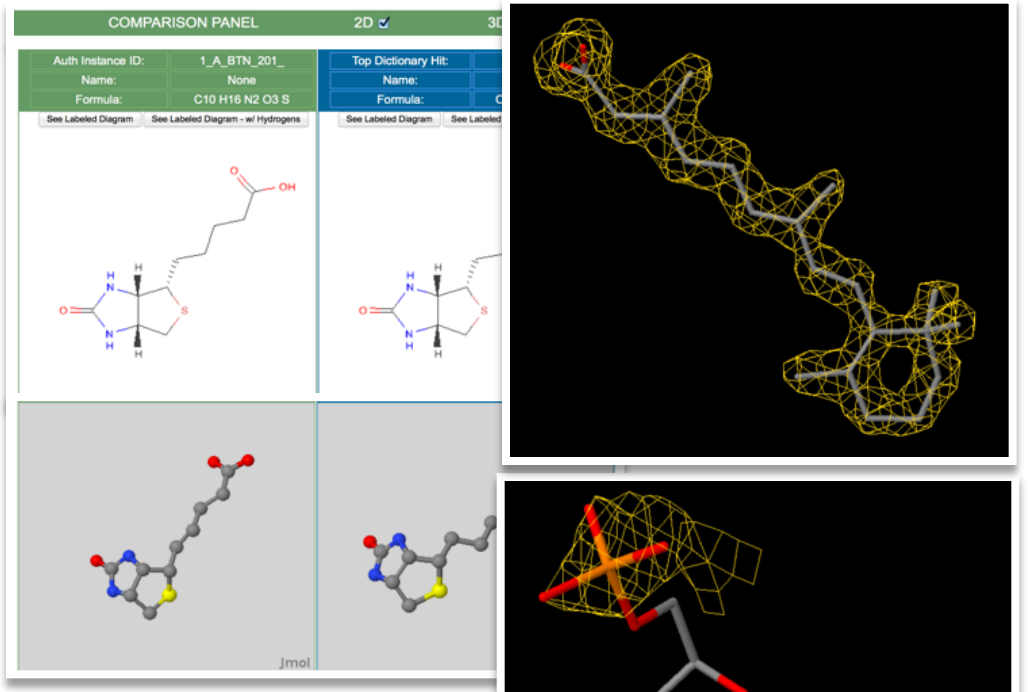
Improved Sequence Annotation

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



Improved Ligand Annotation

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

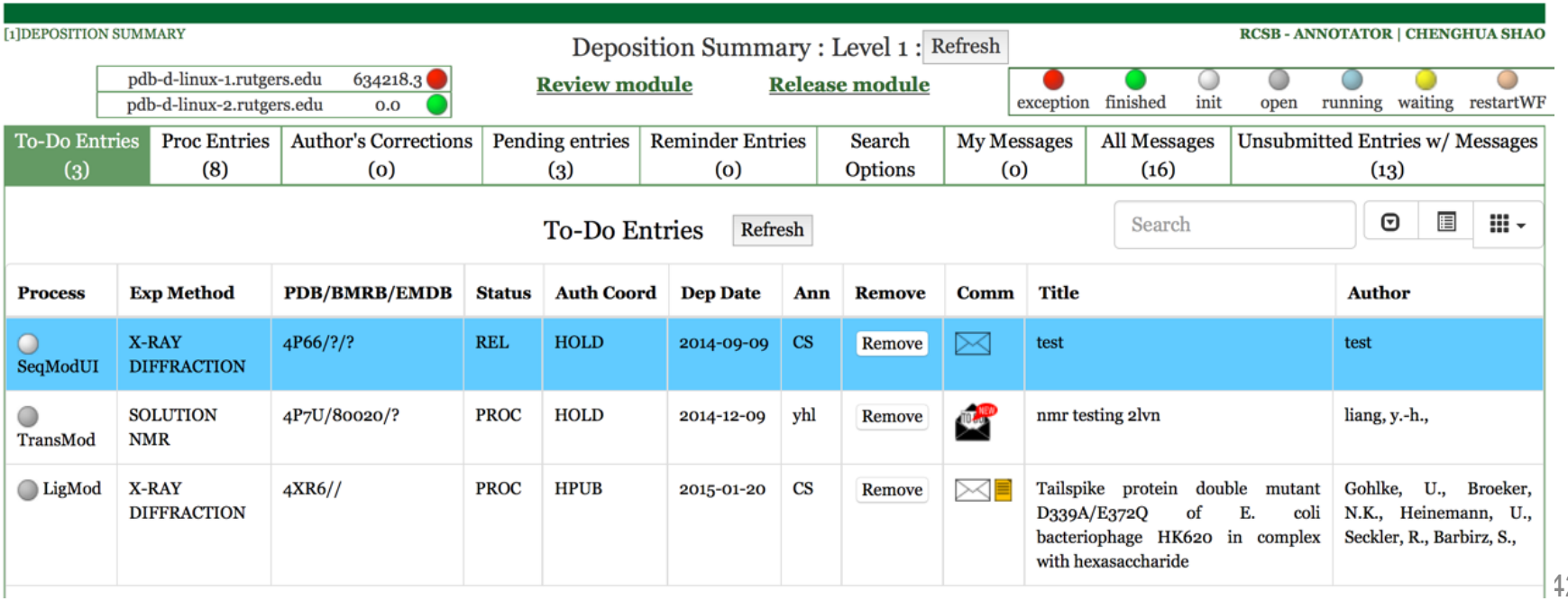


Deposited instance from coordinates (left) and the closest match in the dictionary (right)

Local ligand density display (1.5 sigma omit map)
Top: REA in entry 1CBS with LLDF=1.31 (RSR=0.10, CC=0.95)
Bottom: TMP in entry 3HW4 with LLDF=6.77 (RSR=0.41, CC=0.70)

Improved Workflow Management

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



Improved Efficiency and Advances

Efficiency

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

Advances

- Improved support for large structures
- Workflow ensures completeness of annotation
- Enhanced validation
- Enabled tracking

Improved Throughput

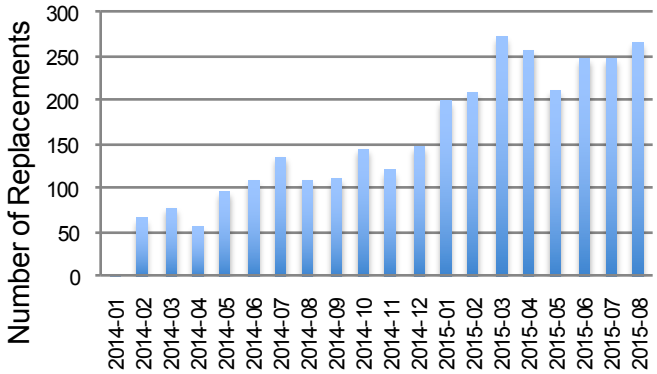
Pros:

- Improved throughput with system enhancements
  - Median: 16.5 days→1.6 days
  - Majority completed in ~2 days
  - ADIT median ~14 days

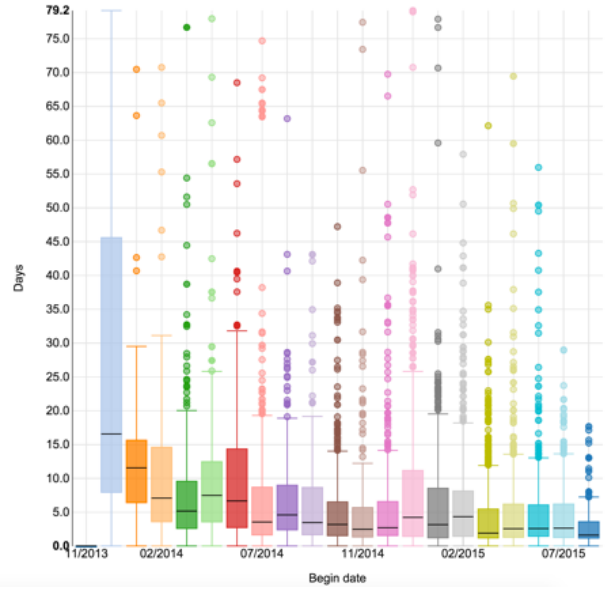
Con:

- Increased coordinate replacements as depositors react to Validation report post deposition

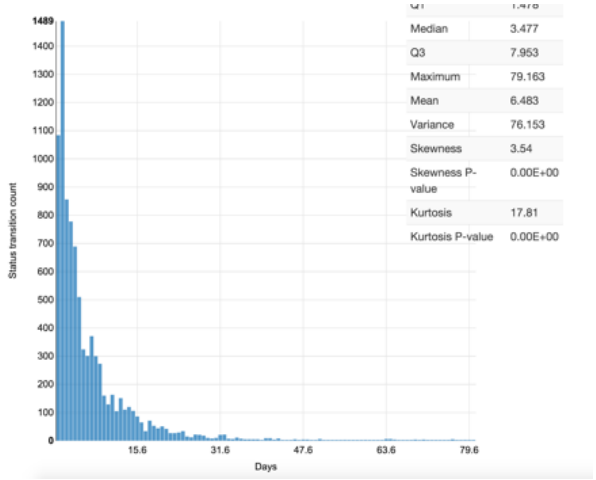
Number of Coordinate Replacements by Month



Average Processing Time by Month



Distribution of Processing Time



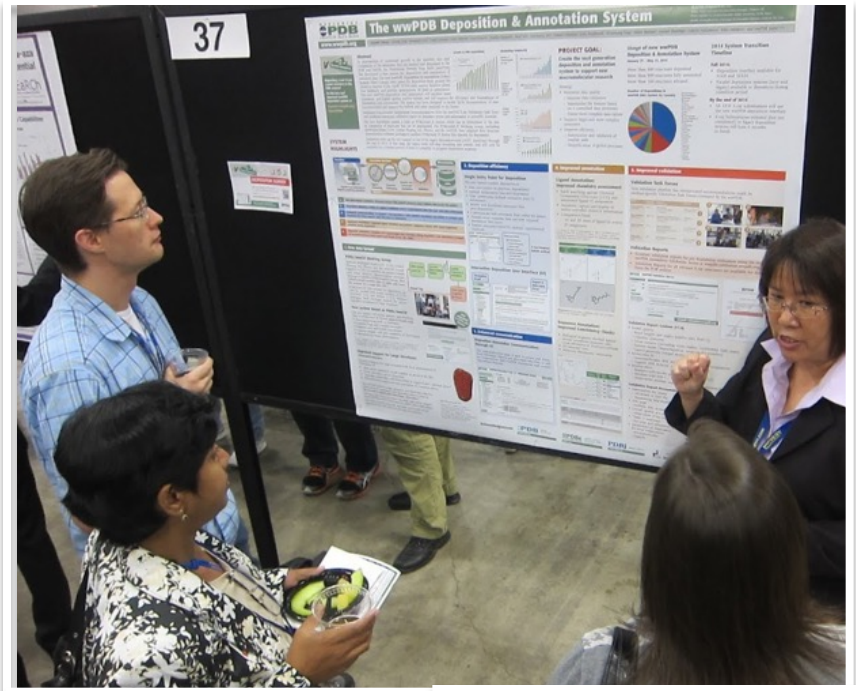
User Feedback

External Users (Depositors)

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

Internal Users (Annotators)

- Combined >20 annotator years of production experience
- Continuous testing and improvement
- Weekly cross-site reviews of issues



Common External Feedback

What they like most

- Easy navigation
- Straightforward input forms

Suggestions under discussion

- Support for simultaneous depositions
- Support for batch depositions
- Ability to access depositions by *Author* rather than session ID

Common Internal Feedback

What they like most

- Ligand annotation module
- Sequence annotation module
- Processing of multiple entries in parallel
- Automated workflows

Suggestions under discussion

- Automation of model coordinate re-submission
- Incorporate ligand chirality check at deposition and for stand-alone validation server
- Optimization for very large structures
- Improve annotation and communication efficiency for related entries

wwPDB/CCDC/D3R Ligand Validation Workshop

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

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

LVW White Paper

- White paper describing recommendations re deposition/validation and editorial/refereeing/publication standards to be drafted and submitted to *Structure* (ideally no later than October 31st 2015)
- Recommendations to be reviewed by the wwPDB X-ray VTF (November 16-17 2015 at EBI)

NMR: PDB – BMRB

John Markley

BMRB Staffing

- Eldon Ulrich, one of the founders of BMRB and its Director for over 20 years, retired as of August 1, 2015
  - Eldon has been granted “emeritus” status
  - He continues to volunteer part time at BMRB and remains a great source of knowledge and inspiration
- Pedro Romero, has assumed the position of BMRB Director
- Other staff members
  - Head annotator: Hongyang Yao
  - Systems manager (75%): Dmitri Maziuk
  - Programmer (75%): Kent Wenger
  - Assistant Scientist: Kumaran Baskaran
  - Postdoc: Vincent Chen (CIBM training grant)
  - Undergraduates
- Computer science advisors
  - Miron Livny (Univ. of Wisconsin Madison)
  - Yannis Ioannidis (Univ. of Athens, Greece)

BMRB External Advisory Board

- Most recent meeting: March 28, 2015
- Membership

  Martin Blackledge, Grenoble (2010-2015)

  Valérie Copié, Bozeman MT (2011-2016)

  Art Edison, Athens GA (2015-2019)

  Mei Hong, Cambridge MA (2011-2016)

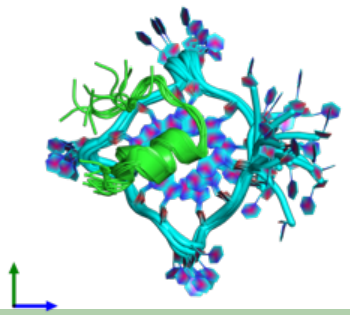
  Peter Tompa, Brussels (2015-2019)

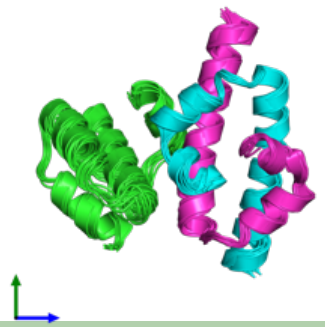
  Representative from PDBj-BMRB

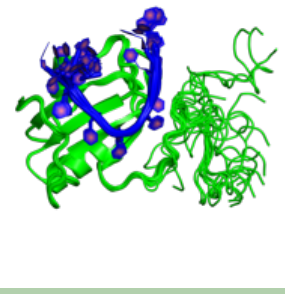
  Naohiro Kobayashi, Osaka

NMR Depositions (9/1/14 – 8/31/15)

Site	NMR structures (NMR data sets annotated)	Experimental NMR data without structures (annotated)	Total (number annotated)
BMRB	390 (410)	323 (285)	713 (684)
PDBe	21	-	21
PDBj-BMRB	16 (17)	10 (48)	26 (65)
Total	427	333	759







NMR Depositions (past four years)

Date	NMR structures	Experimental NMR data without structures	Total
2012	556	272	828
2013	486	308	714
2014	506	240	746
2015	427	333	760

Sept 2015: 12,113 NMR entries in PDB (~11%)

J Biomol NMR
DOI 10.1007/s10858-015-9969-9

ARTICLE

MolProbity for the masses–of data

Vincent B. Chen[1] · Jonathan R. Wedell[2] · R. Kent Wenger[2] · Eldon L. Ulrich[2] · John L. Markley[1]

- Version that supports distributed processing (Condor)
- Enabled MolProbity analysis of all structures in the PDB
- Validation comparisons across PDB

MolProbity Analysis of NMR and X-ray Structures

Table 1 Average and standard deviation for five year periods of various MolProbity quality measures (lower averages indicate better quality)

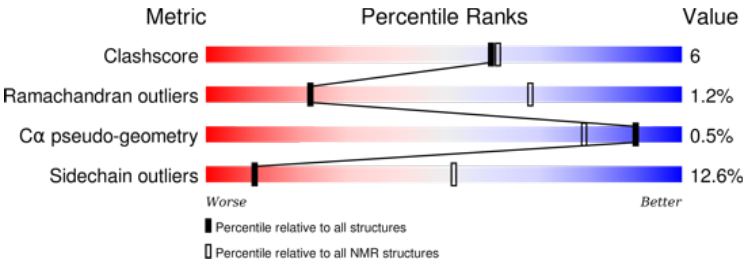
	Clashscore	% Rama outliers	% Rotamer outliers	% Bad bonds	% Bad angles	% RNA pperp outliers
NMR-core						
1994–1998	56.5 (58.4)	5.8 (5.9)	20.3 (13.2)	0.05 (0.49)	0.22 (1.38)	n/a
1999–2003	58.1 (60.0)	5.2 (6.8)	22.0 (14.1)	0.03 (0.60)	0.10 (0.72)	n/a
2004–2008	35.4 (35.3)	3.1 (5.3)	13.8 (12.4)	0.02 (0.47)	0.06 (0.59)	n/a
2009–2013	27.7 (30.8)	2.2 (3.9)	12.7 (11.7)	0.03 (0.67)	0.05 (0.35)	n/a
NMR-full						
1994–1998	51.2 (54.6)	7.1 (6.1)	20.5 (13.5)	0.18 (1.47)	0.66 (2.40)	5.9 (9.3)
1999–2003	51.8 (55.3)	6.6 (6.9)	22.1 (13.7)	0.14 (1.46)	0.41 (1.73)	7.5 (10.9)
2004–2008	32.7 (33.3)	4.2 (4.9)	14.1 (12.3)	0.08 (0.94)	0.18 (1.08)	8.0 (10.6)
2009–2013	26.6 (29.6)	3.2 (4.3)	12.7 (11.3)	0.09 (1.24)	0.18 (0.99)	8.5 (12.8)
X-ray						
1994–1998	14.8 (16.9)	1.1 (2.0)	7.5 (6.1)	0.34 (2.24)	0.89 (2.64)	3.7 (9.5)
1999–2003	14.2 (14.6)	1.0 (2.0)	4.9 (4.8)	0.12 (1.38)	0.33 (1.27)	5.1 (9.5)
2004–2008	12.6 (17.2)	0.9 (1.8)	4.4 (4.3)	0.07 (0.86)	0.17 (0.80)	4.8 (8.6)
2009–2013	9.2 (12.0)	0.6 (1.5)	3.6 (3.9)	0.05 (0.67)	0.10 (0.75)	3.6 (6.6)

wwPDB NMR Validation Task Force

- Meetings in 2009 (Paris), 2011 (NJ), 2013 (NJ, VT and EMBL-EBI) and 2015 (NJ)
  - GT Montelione, M Nilges, A Bax, P Güntert, T Herrmann, JS Richardson, C Schwieters, WF Vranken, GW Vuister, DS Wishart
  - New members joined in 2015 to provide expertise on solid-state NMR and nucleic acids: RA Byrd, T Polenova, S Butcher
  - Recommendations published in 2013
  - Reports reviewed in 2015
  - Most of the feedback is now incorporated

Montelione et al., Structure 21, 1562 (2013)

NMR Validation Pipeline

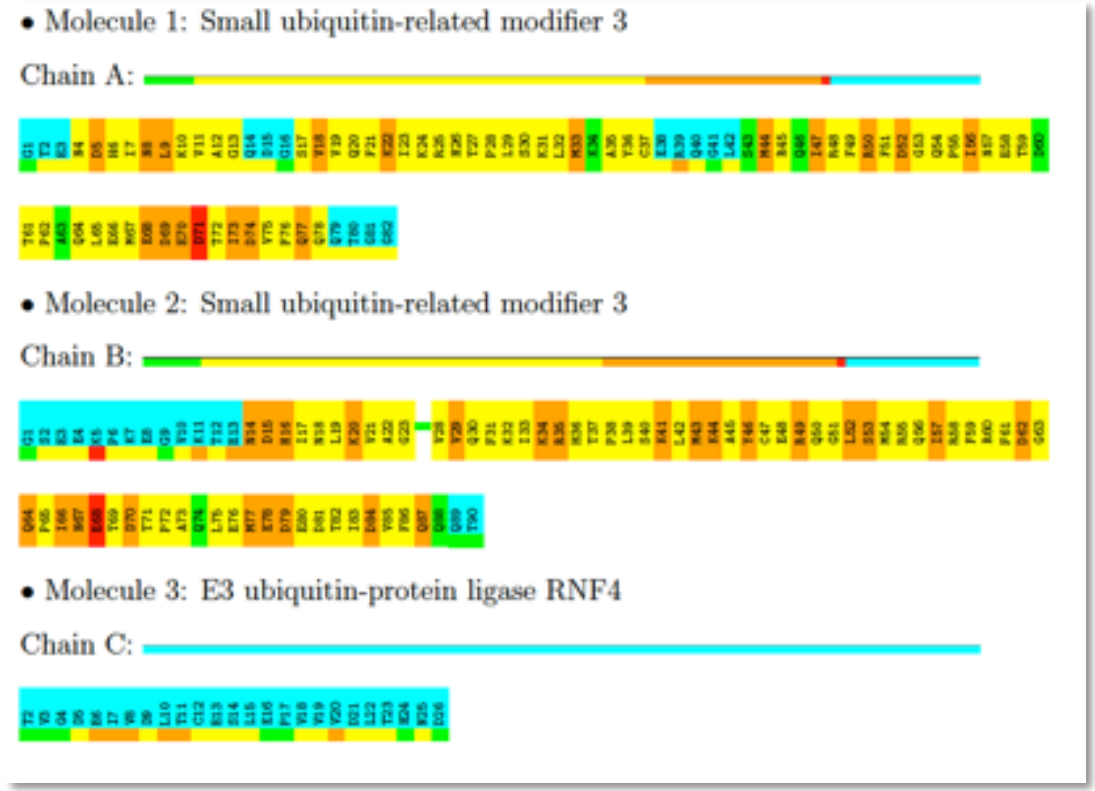


- Reports include
  - Model validation – updated following VTF feedback
  - Chemical-shift validation
  - Annotation information
- Incorporated in deposition and annotation modules of D&A
- Applied to all NMR entries in PDB
- Still to do:
  - Make the validation reports public
  - Provide as stand-alone server
  - Incorporate restraint validation and eventually peak lists

NMR Validation Reports – new features

Mol	Chain	Length	Quality of chain
1	A	82	
2	B	90	
3	C	25	

- Clearer distinction of well-defined and ill-defined (cyan) regions
- Residue plots for average scores and for representative ("medoid") model
- Plots for all models in the full report



NMR Validation Reports – new features

2 Ensemble composition and analysis

This entry contains 10 models. Model 7 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *lowest energy*.

The following well-defined residues (cores) were identified:

Well-defined core	Residues		Backbone RMSD (Å)	Medoid model
1	67	A:4-A:13, A:17-A:37, A:43-A:78	0.21	2
2	75	B:14-B:88	0.31	7

Well-defined cores, RMSDs and representative models

6.9.2 Cα pseudo-geometry

The following table provides a summary of the Cα pseudo-geometry for proteins. The "Percentiles" column shows the percent Cα pseudo-geometry outliers as a percentile score with respect to all PDB entries, followed by that with respect to all NMR entries. The "Analysed" column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
1	A	67/82 (82%)	49±2 (72±3%)	17±2 (26±3%)	1±1 (2±1%)	65 79
2	B	75/90 (83%)	47±2 (63±3%)	26±3 (35±4%)	2±1 (2±2%)	43 68
3	C	0	-	-	-	-
All	All	1420/1970 (72%)	955 (67%)	437 (31%)	28 (2%)	52 73

Pseudo-Ramachandran analysis

5 of 10 unique Cα pseudo-geometry outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
1	A	52	ASP	7
2	B	35	ARG	7
2	B	53	SER	3

NMR Validation Reports – chemical shifts

7.1 Chemical shift list *assigned_chem_shift_list_1*

File name: 2mp2_cs.str

The following table shows the results of parsing the chemical shift list and reports the number of nuclei with statistically unusual chemical shifts.

Total number of shifts	1516
Number of shifts mapped to atoms	1516
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	17

Bookkeeping

Referencing corrections

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# measurements	Correction and precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	90	-0.29 ± 0.16	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	89	-0.28 ± 0.12	None needed (< 0.5 ppm)
$^{13}C'$	89	-0.00 ± 0.11	None needed (< 0.5 ppm)
^{15}N	88	0.68 ± 0.24	Should be applied

	Total	H	C	N
Backbone	455/969 (47%)	188/386 (49%)	179/394 (45%)	88/189 (47%)
Sidechain	806/1346 (60%)	521/786 (66%)	260/492 (53%)	25/68 (37%)
Aromatic	66/128 (52%)	34/68 (50%)	32/50 (64%)	0/10 (0%)
Overall	1327/2443 (54%)	743/1240 (60%)	471/936 (50%)	113/267 (42%)

Completeness of chemical shift assignment

NMR Validation Reports – chemical shifts

The following table lists the statistically unusual chemical shifts.

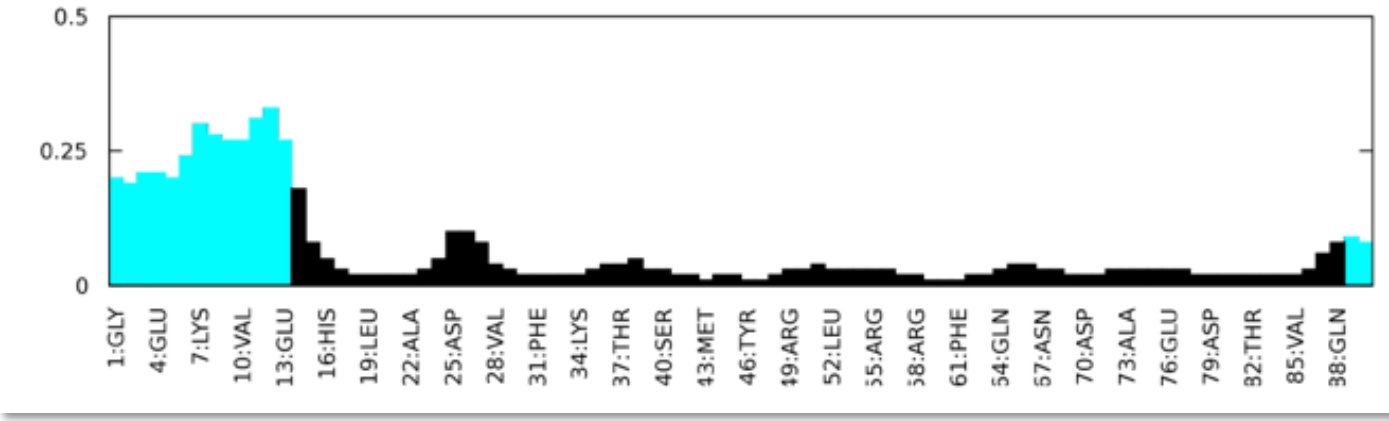
Mol	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	76	PHE	CE2	61.40	124.71 – 136.81	-57.3
1	A	76	PHE	CE1	61.40	123.42 – 137.92	-47.8
1	A	67	MET	CE	88.20	7.37 – 26.97	36.2
1	A	33	MET	CE	87.60	7.37 – 26.97	35.9
1	A	45	ARG	NE	108.37	76.73 – 92.63	14.9
1	A	25	ARG	NE	107.11	76.73 – 92.63	14.1
2	B	87	GLN	HE21	2.39	4.93 – 9.53	-10.5

Chemical shift outliers

RCI: the higher the bar, the more disorder is predicted

Cyan: ill-defined regions

Random coil index (RCI) for chain B:

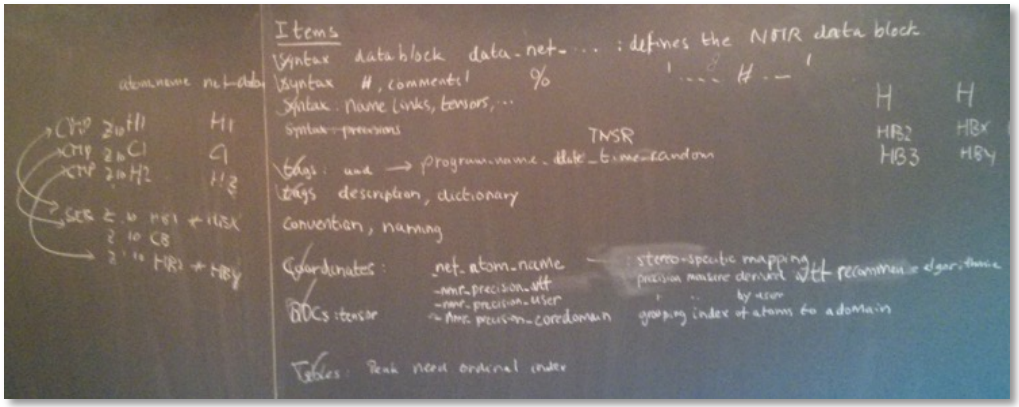


Workshops on Representation of NMR Restraints

- Convened by PDBe on behalf of wwPDB with support from U Leicester, at EMBL-EBI, November 2013
- Follow-on workshop at Rutgers, January 2015
- All major NMR structure determination, refinement and validation packages represented, mostly by PIs:
  - D. Case (AMBER), P. Güntert (CYANA), T. Herrmann (UNIO), O. Lange (CS-Rosetta), M. Nilges (CNS/ARIA), Ch. Schwieters (Xplor-NIH), W. Vranken (CCPN), G. Vuister (CCPN/CING), D. Wishart (PROSESS/PANAV/SHIFTX2), G. Montelione (Autostructure/PSVS)
- Observers
- wwPDB representatives

Workshop Agreements (Nov-2013)

- mmCIF/STAR-like format
  - Polymer sequence
  - Chemical shifts
  - Restraints lists
  - Peak lists, if available



- Commitment from all participants to write and read such files
- Ambitious timeline of 1 year
- Dictionary versioning and namespaces
- Working group to develop detailed proposal
- Publish brief workshop report first and an authoritative paper once format specified and implemented

NMR Exchange Format (NEF)

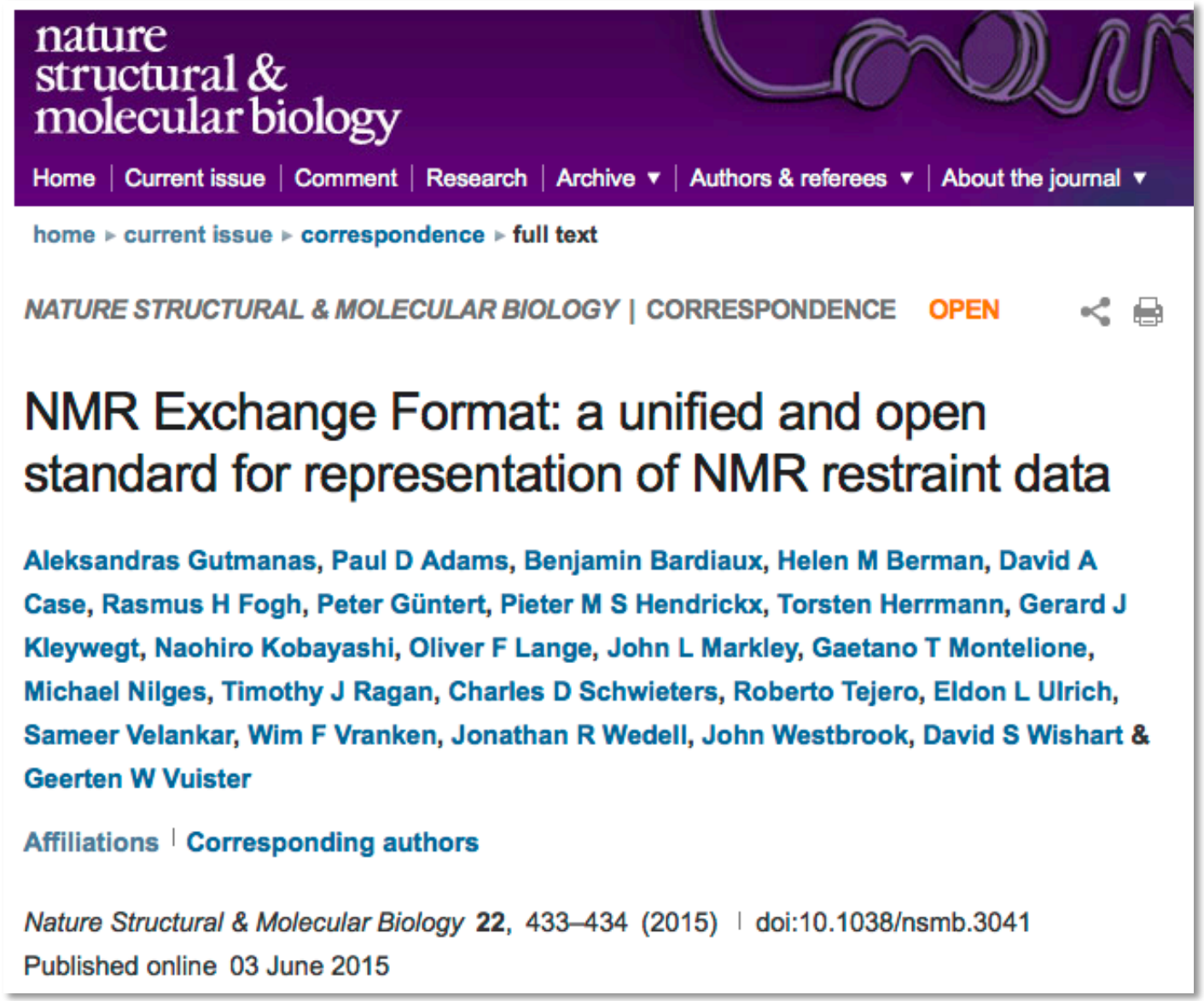
- Format specification and example files developed by working group, spearheaded by U Leicester
- All represented software could write NEF files and read each other's output
- Decisions on outstanding issues (e.g., RDC representation, treatment of stereo-specific assignments)
- Decisions on management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

```
loop_
  # Mandatory parameters, except for restraint_combination_id
  _nef_distance_restraint.ordinal
  _nef_distance_restraint.restraint_id
  _nef_distance_restraint.restraint_combination_id
  _nef_distance_restraint.chain_code_1
  _nef_distance_restraint.sequence_code_1
  _nef_distance_restraint.residue_type_1
  _nef_distance_restraint.atom_name_1
  _nef_distance_restraint.chain_code_2
  _nef_distance_restraint.sequence_code_2
  _nef_distance_restraint.residue_type_2
  _nef_distance_restraint.atom_name_2
  _nef_distance_restraint.weight

# The following parameters are optional. target_value and target_value_uncertainty
# should be given whenever a meaningful value is known'
# Other parameters need be given only if they are defined for the potential_type
  _nef_distance_restraint.target_value
  _nef_distance_restraint.target_value_uncertainty
  _nef_distance_restraint.lower_linear_limit
  _nef_distance_restraint.lower_limit
  _nef_distance_restraint.upper_limit
  _nef_distance_restraint.upper_linear_limit

# key: ordinal

  1  1  .  A 21  ALA  HB%  A 17  VAL  H   1.00 3.7  0.4   2.0  2.5  4.2  4.7
  2  1  .  A 21  ALA  HB%  A 18  LEU  H   1.00 3.7  0.4   2.0  2.5  4.2  4.7
  3  1  .  A 22  THR  HG2% A 17  VAL  H   1.00 3.7  0.4   2.0  2.5  4.2  4.7
  4  1  .  A 22  THR  HG2% A 18  LEU  H   1.00 3.7  0.4   2.0  2.5  4.2  4.7
  5  5  .  A 18  PHE  HB2  A 24  ASP  HBX 1.00 2.8  0.4   1.5  2.0  3.2  3.7
  6  8  .  A 18  PHE  HB2  A 24  ASP  HBY 1.00 4.4  0.4   3.6  4.1  5.2  5.7
  6  8  .  E 6B  SER  HB2  A 24  ASP  HBY 1.00 4.4  0.4   3.6  4.1  5.2  5.7
stop_
```

https://github.com/NMRExchangeFormat/NEF

NEF Announcement, June 2015

Invites the broader community to review and contribute



nature structural & molecular biology

Home | Current issue | Comment | Research | Archive ▾ | Authors & referees ▾ | About the journal ▾

home ▸ current issue ▸ correspondence ▸ full text

NATURE STRUCTURAL & MOLECULAR BIOLOGY | CORRESPONDENCE OPEN

NMR Exchange Format: a unified and open standard for representation of NMR restraint data

Aleksandras Gutmanas, Paul D Adams, Benjamin Bardiaux, Helen M Berman, David A Case, Rasmus H Fogh, Peter Güntert, Pieter M S Hendrickx, Torsten Herrmann, Gerard J Kleywegt, Naohiro Kobayashi, Oliver F Lange, John L Markley, Gaetano T Montelione, Michael Nilges, Timothy J Ragan, Charles D Schwieters, Roberto Tejero, Eldon L Ulrich, Sameer Velankar, Wim F Vranken, Jonathan R Wedell, John Westbrook, David S Wishart & Geerten W Vuister

Affiliations | Corresponding authors

Nature Structural & Molecular Biology **22**, 433–434 (2015) | doi:10.1038/nsmb.3041
Published online 03 June 2015

Examination by BMRB of Current NEF Specifications

- Based on analysis of material in the NEF Github as of 09/01/15 (in the absence of a formal NEF data dictionary)
- Results of analysis of the correspondence between NEF and NMR-STAR
  - 50 tags created by NEF were not present in NMR-STAR and have been added to NMR-STAR for compatibility
    - A few other tags may warrant inclusion in NMR-STAR
  - Other tags have been mapped to existing NMR-STAR
    - Some map directly
    - Some map to old PDB format superseded by PDBx (mmCIF)
    - BMRB has recommended modification of several tags
- BMRB has sent its analysis and recommendations to the NEF organizers (includes a table matching NMR-STAR/PDBx tags)

Examination by BMRB of Current NEF Specifications – conclusions

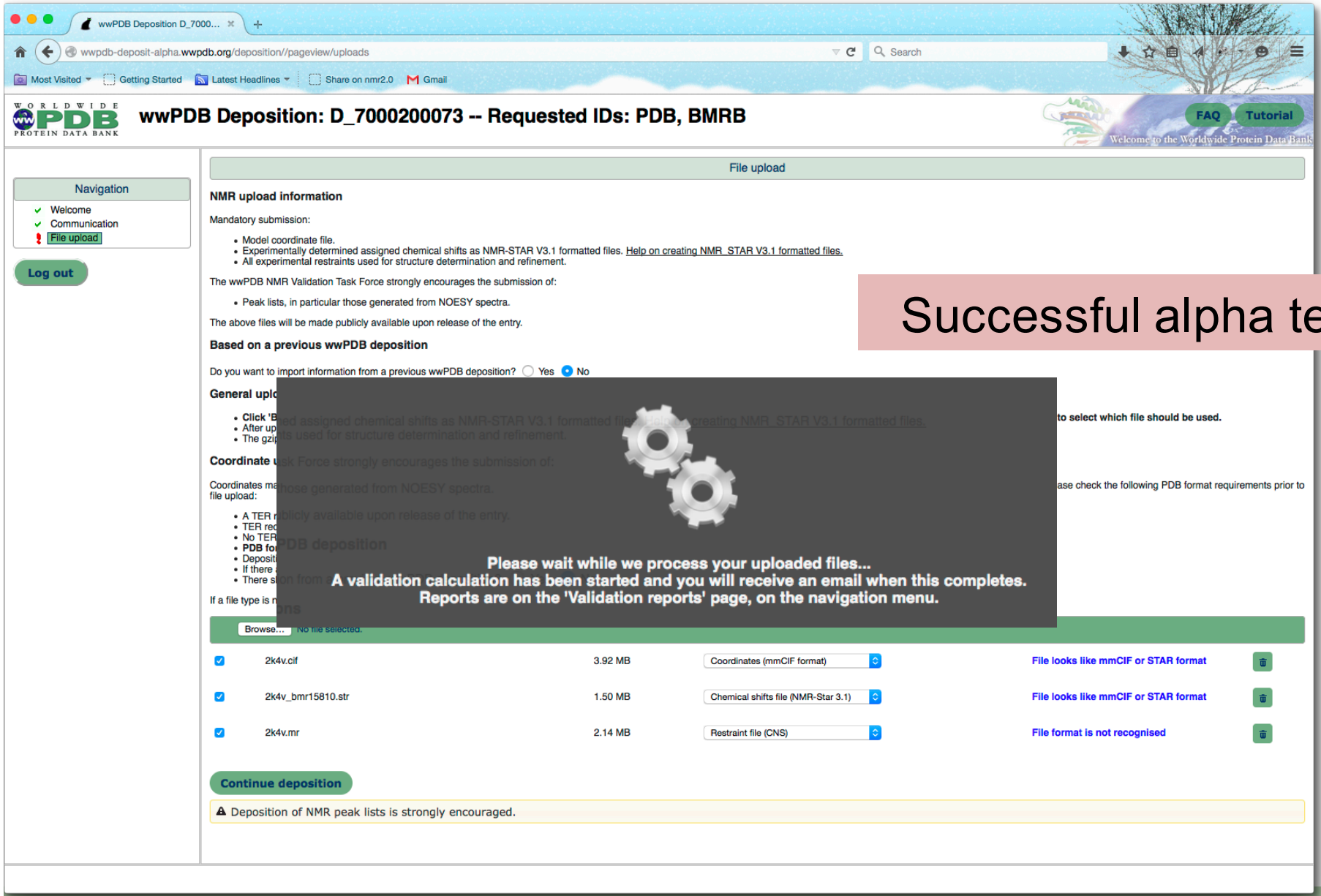
For the NEF to be a deposition format, three main problematic issues found in the current specification need to be resolved:

- The format needs to be brought up to date with the current wwPDB standard, that is, PDBx/mmCIF. This includes being up to date with newer types of NMR restraints and their data.
- Specific and *unambiguous* definitions need to be developed, in which format flexibility is constrained to work within currently accepted parameters.
- The NEF must be fully compatible with the information handled by the public repositories, allowing for unambiguous, direct mapping to the formats used by the public databases (i.e., NMR-STAR for BMRB, PDBx/mmCIF for PDB).

NEF Implications for wwPDB

- NEF could be compatible with NMR-STAR and PDBx
- D&A system to start accepting NEF files
  - Nomenclature checks to be implemented
- Validation pipeline to process NEF files
  - Counts, violations
- Eventually:
  - NEF becomes the standard for NMR restraints at PDB
  - wwPDB D&A system stops accepting legacy formats

NMR in D&A



Successful alpha testing

Validation Reports Must be Viewed Prior to Deposition

List requirements

All items

Mandatory items

Navigation

- Welcome
- Communication
- Re-upload files
- Upload summary
- Admin
  - Contact information
  - Grant information
  - Release status
  - Entry title & author
  - Citation information
- Macromolecules
  - 1) uncharacterized protein PA
- NMR experimental
  - NMR samples
  - NMR data collection
- NMR software
  - NMR software
- NMR data and refinement
  - Peak lists
  - Chemical shift references
  - Chemical shift connection
  - NMR constraints
  - NMR refinement
- Ligands
- Biological assembly
- Validation reports
- Summary & conditions

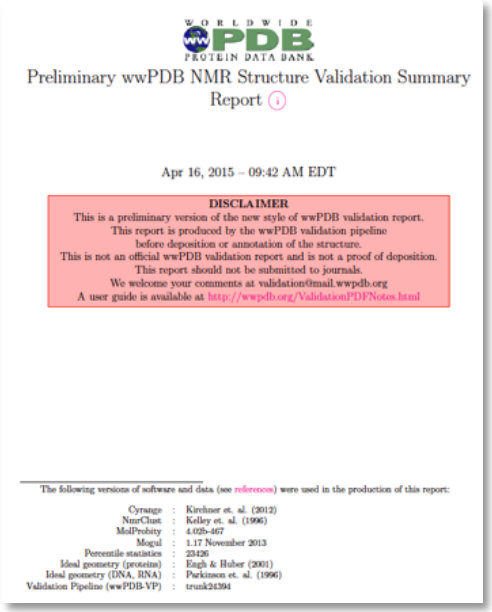
Log out

Download validation reports

The status of the validation calculation is : **finished**

Type	name	Data&Time
PDF File	D_7000200073_val-report-full_P1.pdf.V1	Thu Apr 16 09:43:54 2015
XML File	D_7000200073_val-data_P1.xml.V1	Thu Apr 16 09:43:54 2015
PDF File	D_7000200073_val-report_P1.pdf.V1	Thu Apr 16 09:43:54 2015

Check this box to accept the validation report*: ⚠ No value present for this mandatory item.



WORLDWIDE PDB PROTEIN DATA BANK

Preliminary wwPDB NMR Structure Validation Summary Report

Apr 16, 2015 – 09:42 AM EDT

DISCLAIMER
This is a preliminary version of the new style of wwPDB validation report.
This report is produced by the wwPDB validation pipeline
before deposition or annotation of the structure.
This is not an official wwPDB validation report and is not a proof of deposition.
This report should not be submitted to journals.
We welcome your comments at validation@mail.wwpdb.org
A user guide is available at http://wwpdb.org/ValidationPDFNotes.html

The following versions of software and data (see references) were used in the production of this report:

Cyrange	:	Kirchner et. al. (2012)
NmrClust	:	Kelley et. al. (1996)
MolProbity	:	4.02b-467
Mogul	:	1.17 November 2013
Percentile statistics	:	23426
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk24394

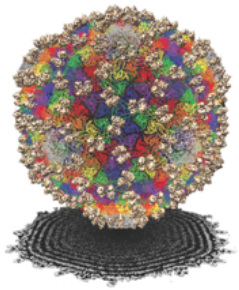
EM: PDB – EMDB – EMPIAR

Gerard Kleywegt

(Sameer Velankar)

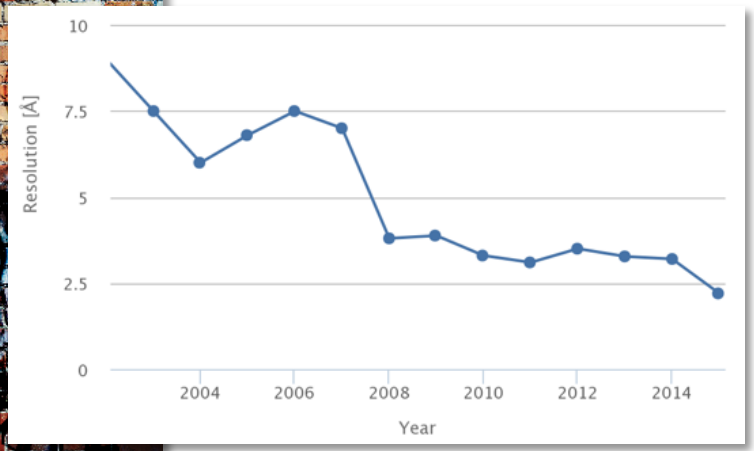
WORLDWIDE PDB
PROTEIN DATA BANK

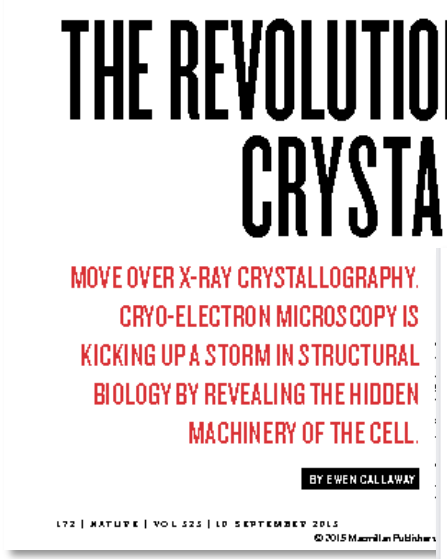
wwpdb.org



EMDataBank
Unified Data Resource for 3DEM

The Revolution in EM





THE REVOLUTION WILL NOT BE CRYSTALLIZED

MOVE OVER X-RAY CRYSTALLOGRAPHY. CRYO-ELECTRON MICROSCOPY IS KICKING UP A STORM IN STRUCTURAL BIOLOGY BY REVEALING THE HIDDEN MACHINERY OF THE CELL.

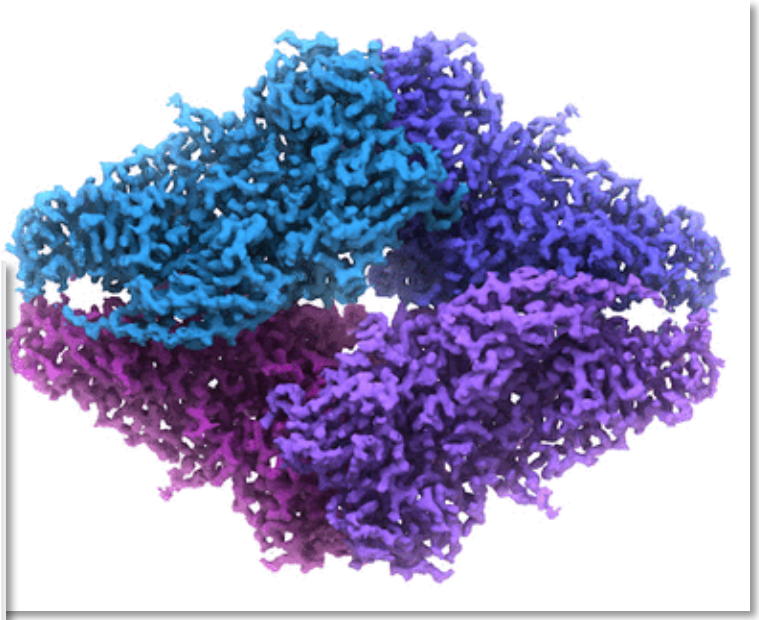
BY EWEN CALLAWAY

172 | NATURE | VOL 525 | 10 SEPTEMBER 2015

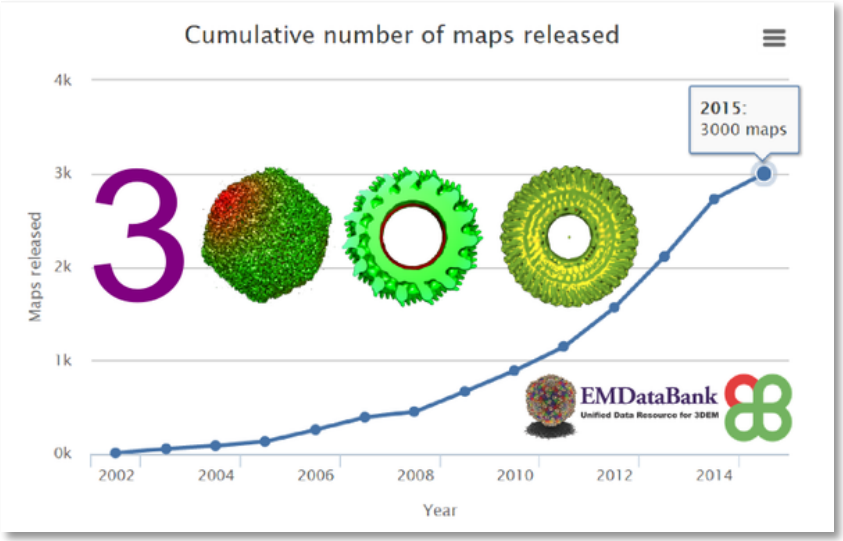
Richard Henderson Werner Kühlbrandt Sriram Subramaniam

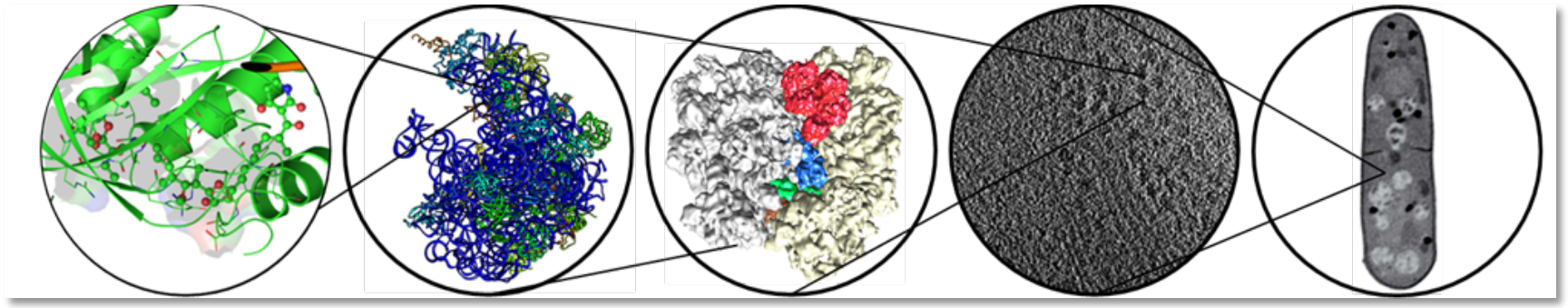
International Union of Crystallography (IUCr)

We are pleased to announce that IUCrJ (http://www.iucrj.org) will start accepting articles in cryo-electron microscopy (cryoEM) within the Biology and Medicine ...



EM Data Archives





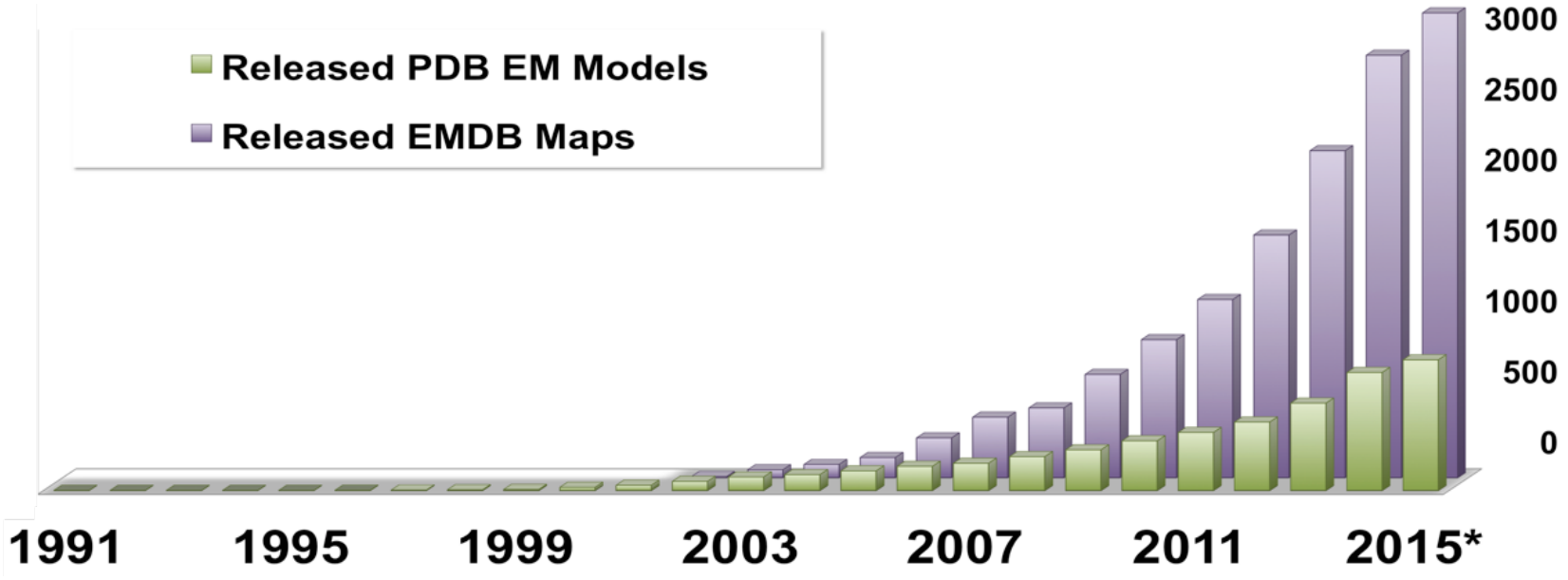
PDB (1971)

EMDB (2002)

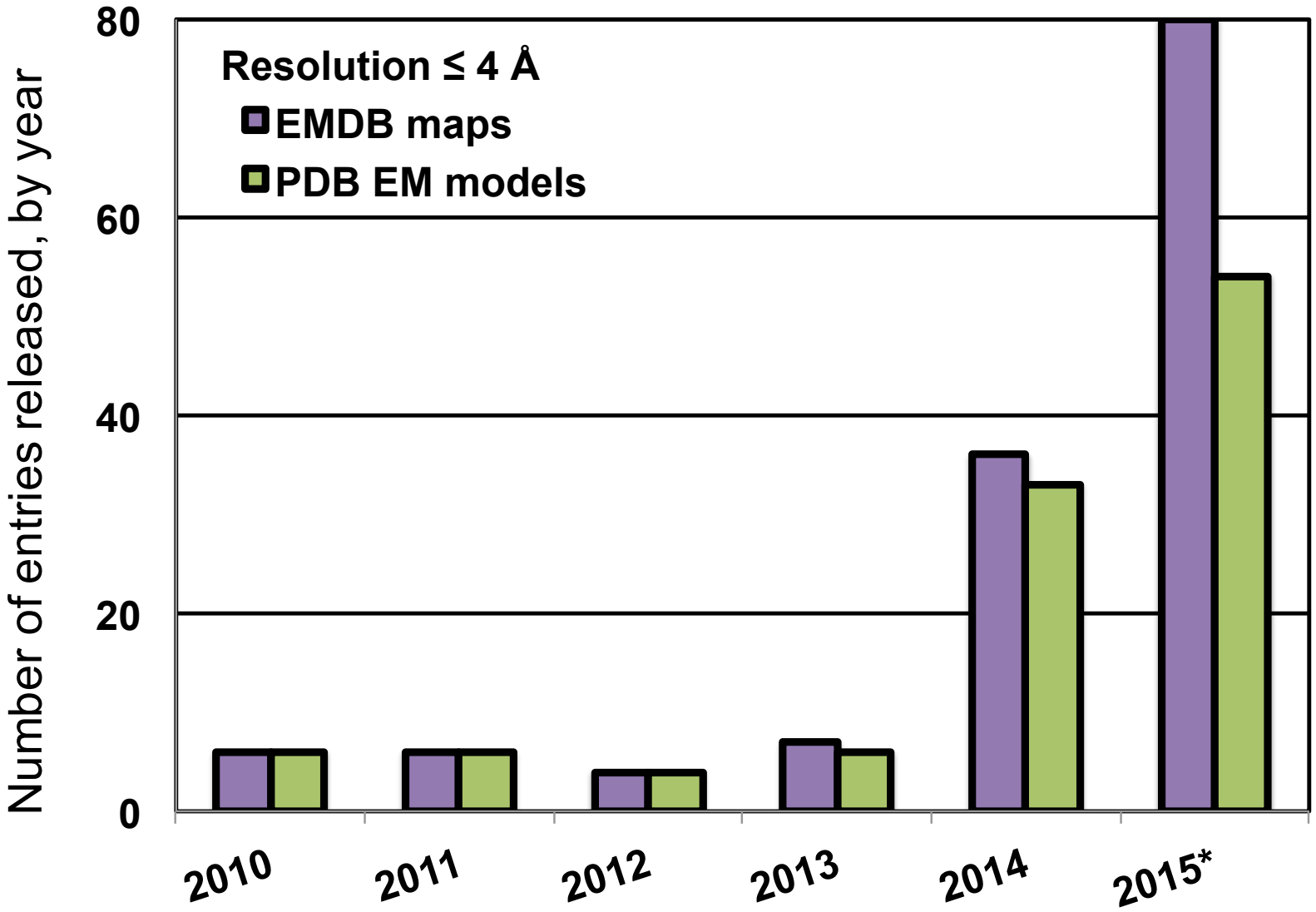
EMPIAR (2014)

EM Data Archives

- 1990s: First EM structures deposited to PDB
- 2002: EM Data Bank (EMDB) map archive established at EBI
- 2004-5: Development workshops with EM community: call for “one-stop shop” for maps and models
- 2007-present: EMDataBank, Unified Resource for 3DEM
- 2014: EMPIAR raw image data archive established at EBI

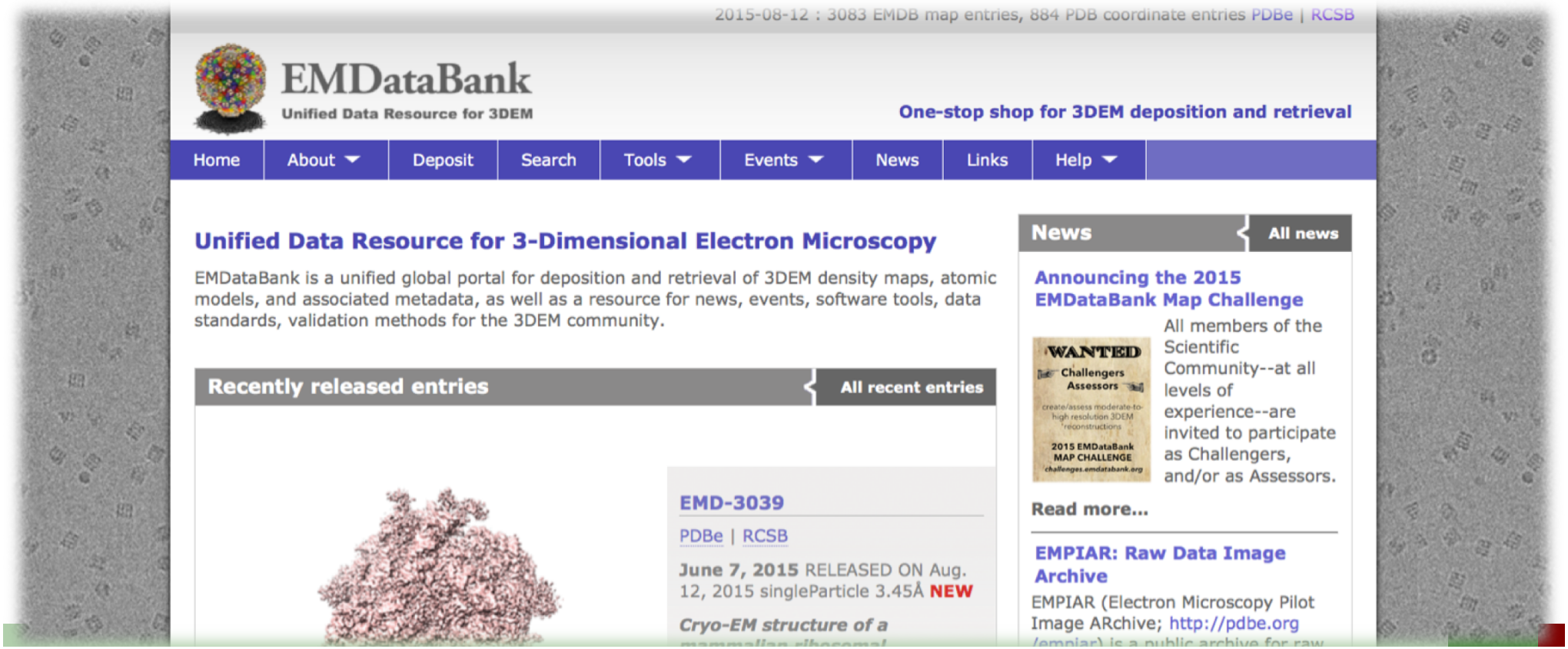


EMDB maps and PDB EM models with resolution 4 Å or better



EMDataBank.org

- Unified global portal for deposition and retrieval of 3DEM volume maps, atomic models, and associated metadata
- Resource for news, events, software tools, data standards, validation methods for the 3DEM community



EMDataBank Project Team

Baylor College of Medicine

Wah Chiu, PI
Steven Ludtke
Corey Hryc
Grigore Pintilie
Matthew Baker
Matthew Dougherty

Rutgers University

Helen Berman, co-PI
Catherine Lawson
Raul Sala
Brian Hudson
John Westbrook

EMBL-European Bioinformatics Institute

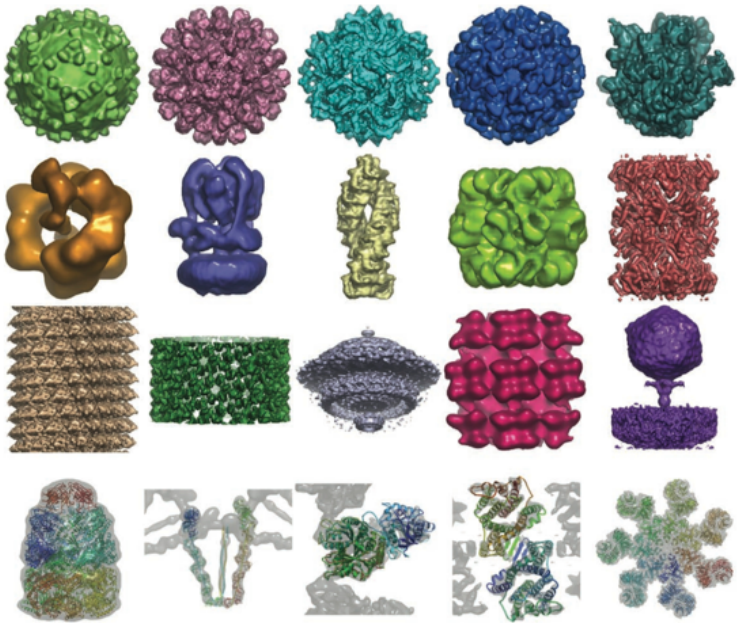
Gerard Kleywegt, co-PI
Ardan Patwardhan
Eduardo Sanz Garcia
Ingvar Lagerstedt
Matthew Conroy

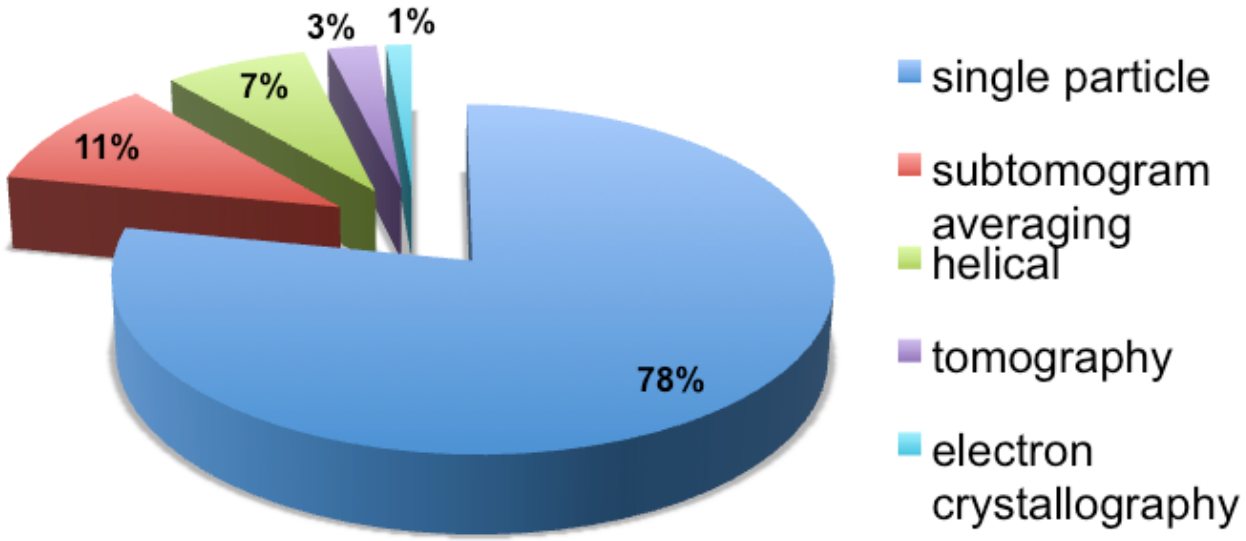
EMDataBank Advisory Committee

Paul Adams (Chair), Richard Henderson, Bram Koster, Maryanne Martone, Andrej Sali

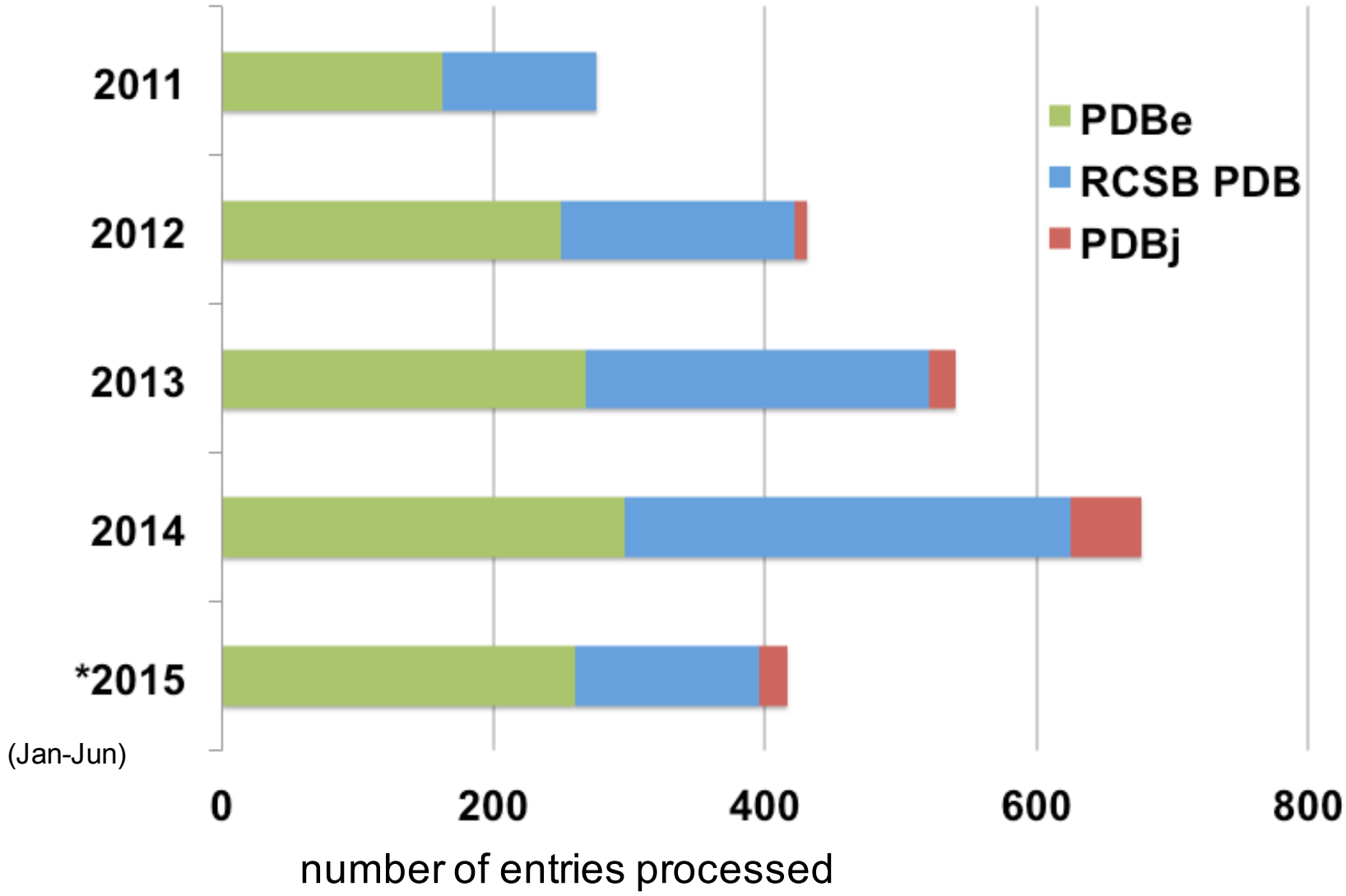
EMDB Content

- Archived maps range from macromolecular complexes to cellular tomograms
- Wide resolution range (100-2 Å)
- ~1/3 of maps have fitted coordinates



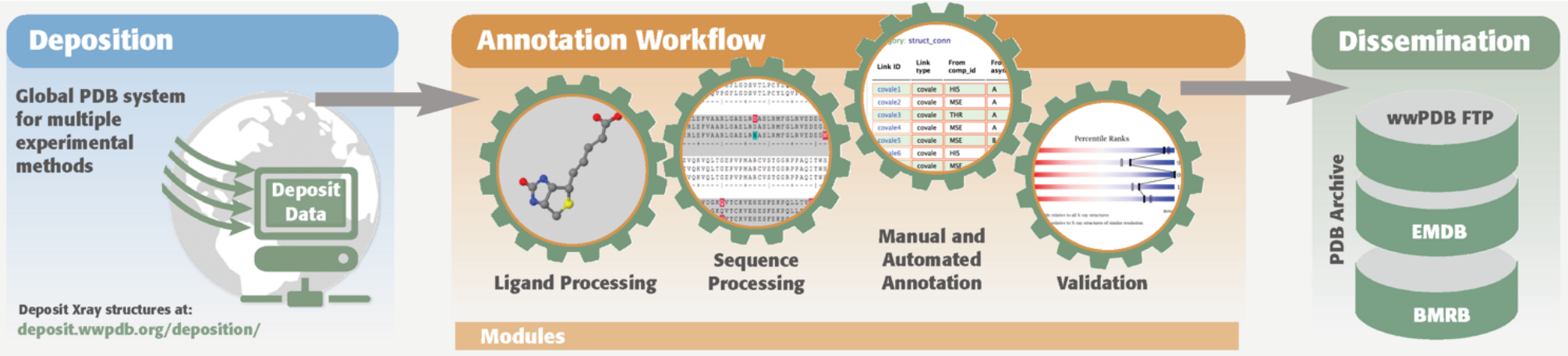


EMDB Map Annotation



EM support in D&A 2.0 - status

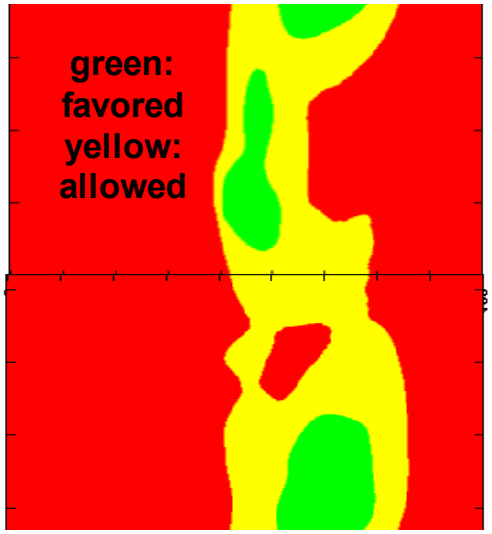
- Support for all EM sub-methods implemented
- Joint map and model deposition, annotation and release
- Basic 3DEM validation reports will be provided to depositors (model validation, “Table 1” about experiment, annotation information)
- Old deposition and annotation systems for maps and models will be retired

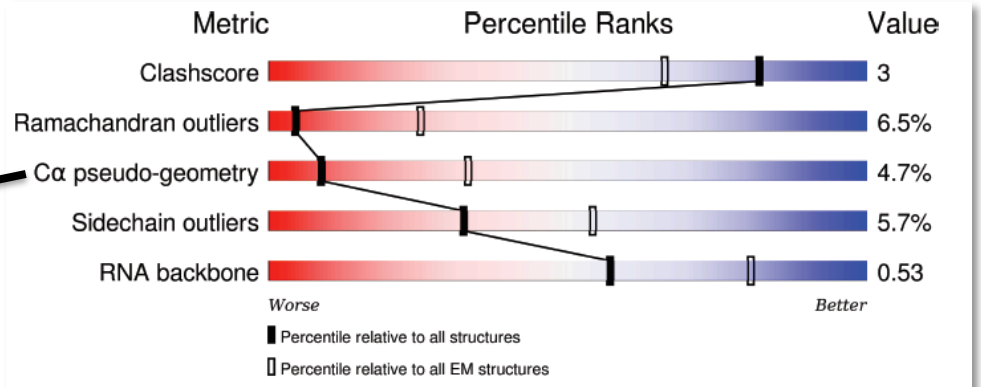


EM Validation Reports

- "Table 1" for EM
- Metrics relevant for EM models

Property	Value	Source
Reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	I	Depositor
Number of images	30000	Depositor
Resolution determination method	FSC 0.143	Depositor
CTF correction method	Not provided	Depositor
Microscope	JEOL 3200FSC	Depositor
Voltage (kV)	300	Depositor
Electron dose $(e^-/^2)$	Not provided	Depositor
Minimum defocus (nm)	500	Depositor
Maximum defocus (nm)	2000	Depositor
Magnification	50000	Depositor
Image detector	DIRECT ELECTRON DE-12 (4k x 3k)	Depositor

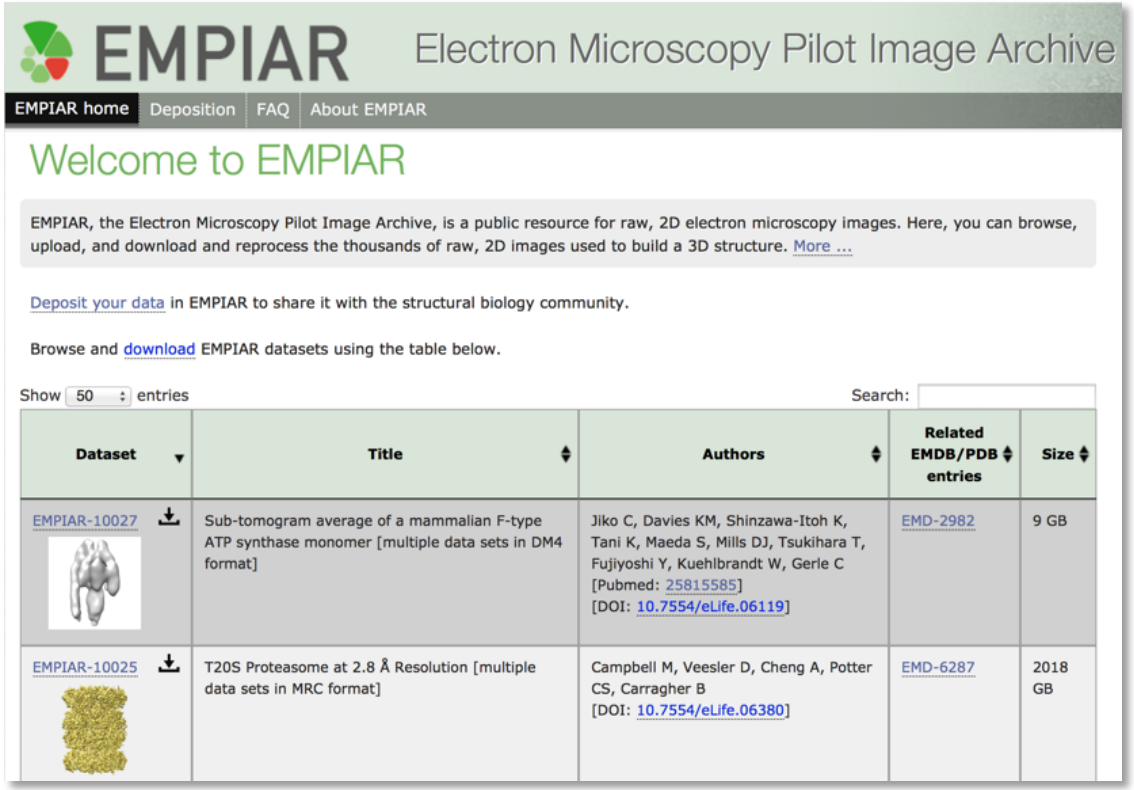




Metric	Whole archive (#Entries)	EM structures (#Entries)
Clashscore	99129	735
Ramachandran outliers	96105	539
Cα geometry	96347	682
Sidechain outliers	96047	526
RNA backbone	2807	214

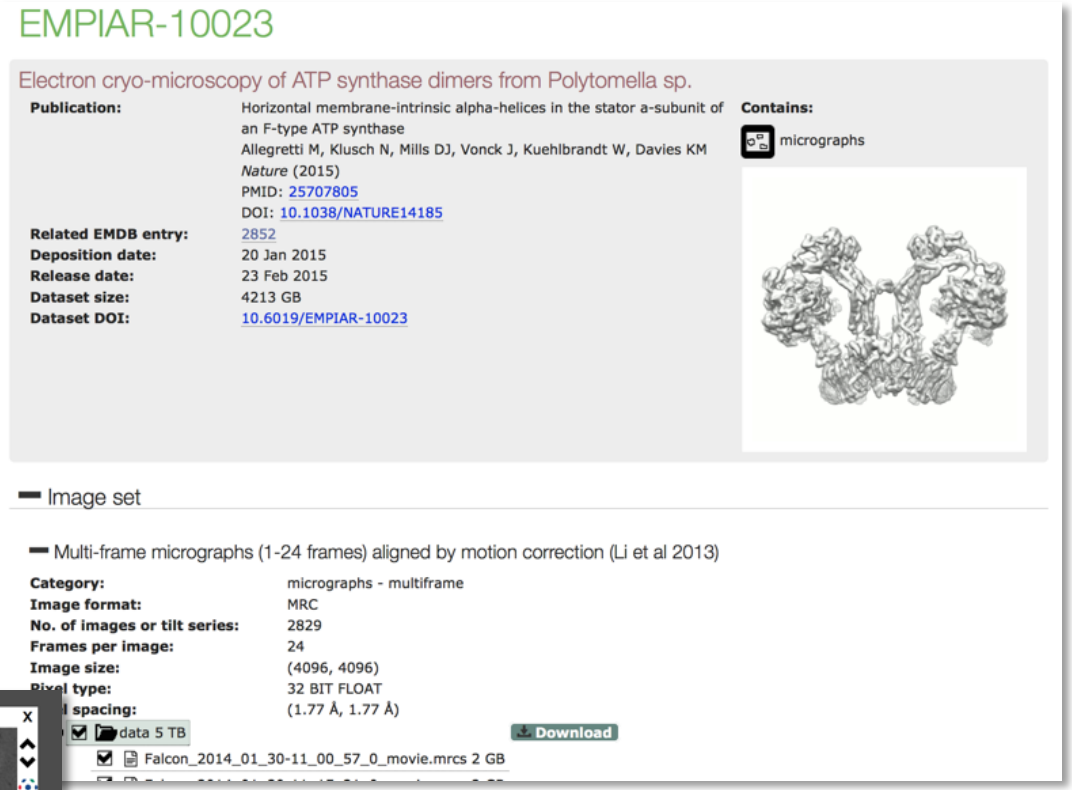
What is EMPIAR?

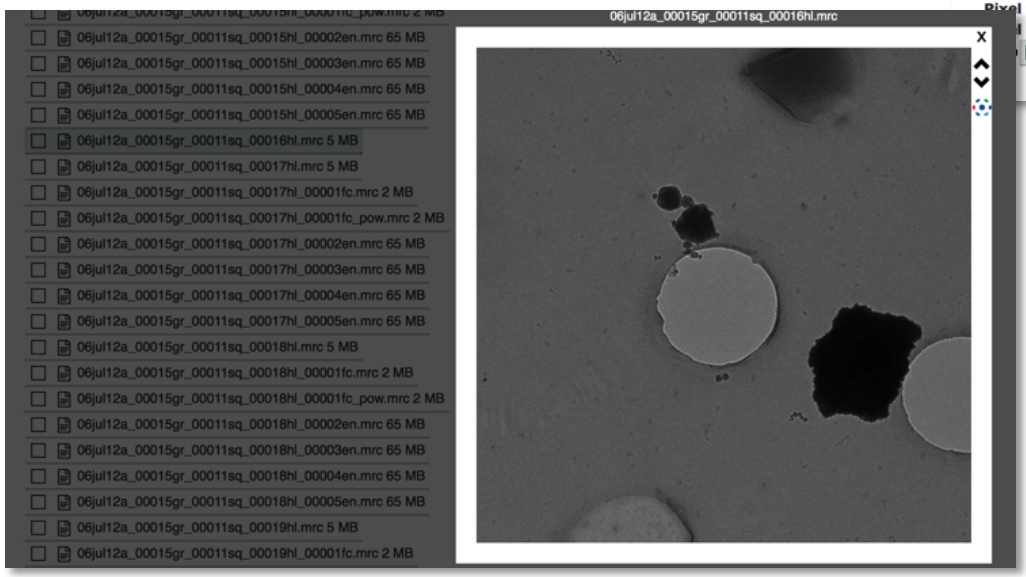
- New pilot archive for raw 2D image data related to EMDB entries
- Provides data for methods development, training, validation and community challenges (including EMDataBank map challenge)
- 32 Entries released to date, from 1.5 GB to 6.5 TB in size (4 entries > 1 TB)
- To be expanded to cover other imaging modalities
- pdbe.org/empiar



EMPIAR features

- EMPIAR entry pages
- Upload/download
  - Mainly Aspera
  - Globus and ftp tarballs also available
- Browser to select files/directories to download
- Every entry has a DOI



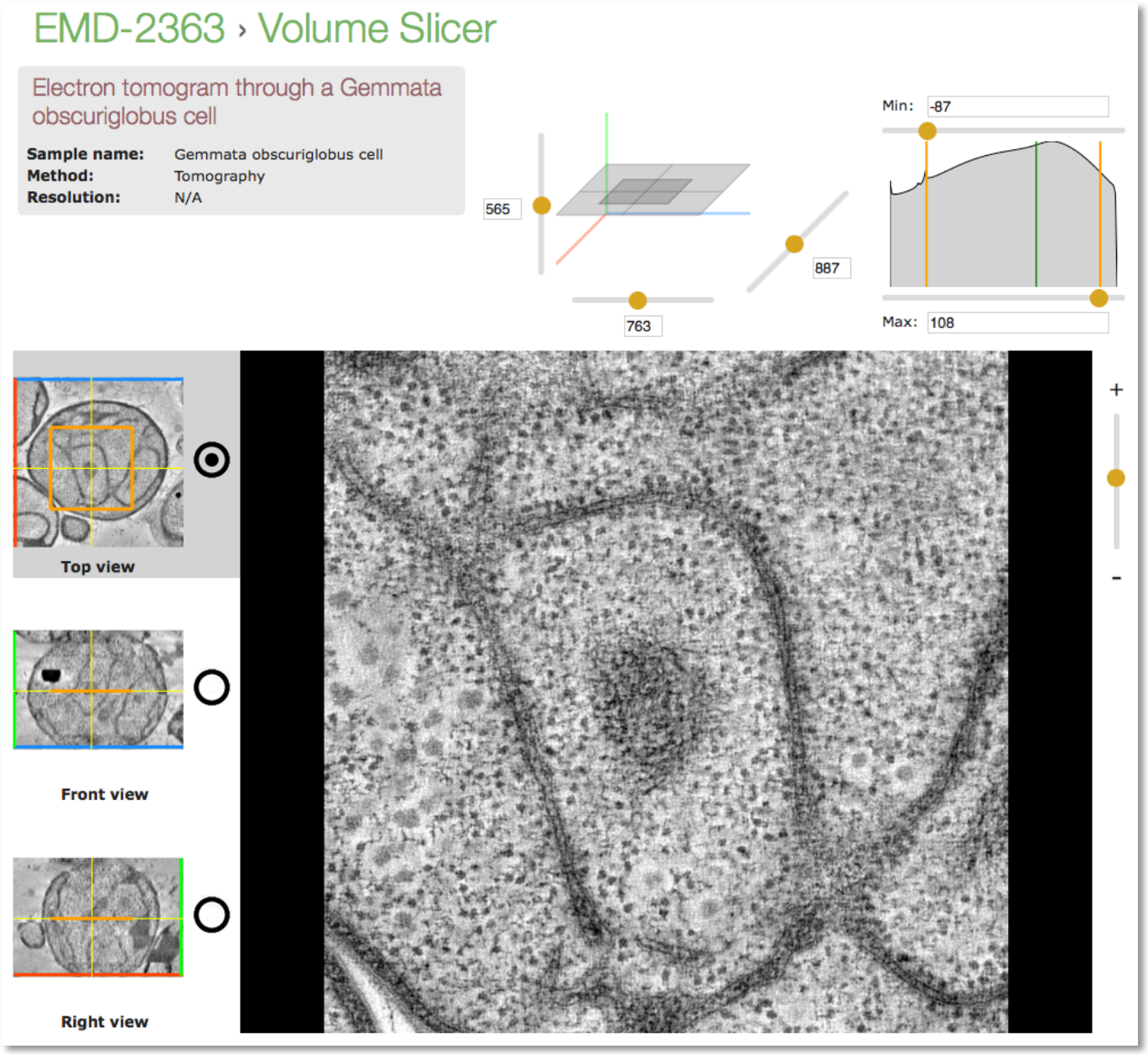


Additional information

Accession codes: Cryo-EM maps and models of BMV have been deposited in EMDB under accession code EMD-6000, and in the Protein Data Bank under accession codes 3J7L, 3J7M and 3J7N. The complete EM data set is available for download through the EMDB Electron Microscopy Pilot Image ARchive (EMPIAR), accession codes 10010 and 10011. This includes both raw micrographs and box coordinates.

EMPIAR features

- New volume slicer
- Soon available for all existing EMDB entries

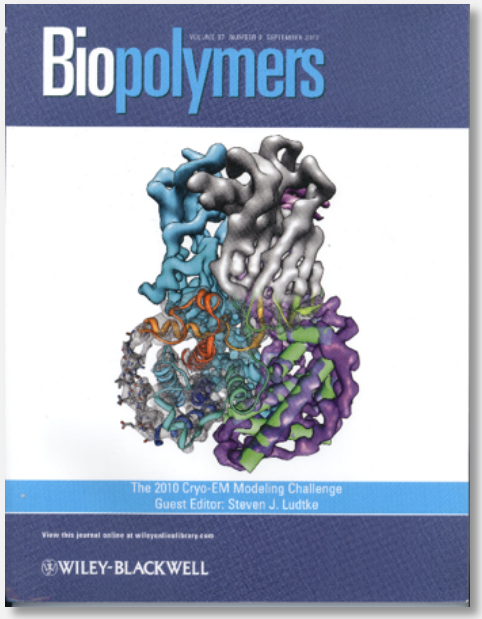


Validation for 3DEM

EM Validation Task Force
Henderson et al. (2012) **Structure 20**, 205-214

Maps: Standards for assessing resolution and accuracy need to be developed

Models: Criteria needed for model only, fit to map, and fit to additional structural data



2010 CryoEM Modeling Challenge
Collected papers in **Biopolymers** special issue September 2012

13 target maps
58 participants
10 research groups
136 submitted models
13 software packages

EMDataBank EM VTF

- Main recommendations for EM maps
  - Standards for assessing resolution and accuracy of a map need to be developed
  - Structural features in a map should be in accordance with the claimed resolution
- Main recommendations for models fitted into EM maps
  - Criteria for assessing models need to be developed
  - Capability to archive coarse-grained representations of models is needed
- More research and development needed!

EMDataBank grant

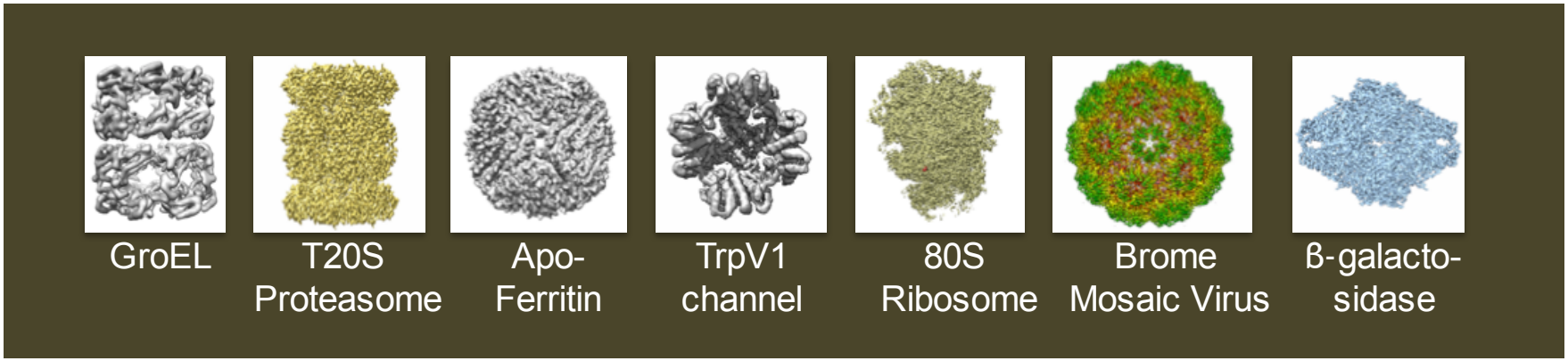
EMDataBank
Unified Data Resource for 3DEM

- NIH R01 grant 2013-2017
  - Wah Chiu PI, Helen Berman, Gerard Kleywegt co-PIs

- Project goals:
  - Establish data-validation methods that can be used in the process of structure determination
  - Define the key indicators of a well-determined 3DEM structure that should accompany every deposition
  - Implement appropriate validation procedures for maps and map-derived models into the deposition pipeline

2015 Map and Model Challenges

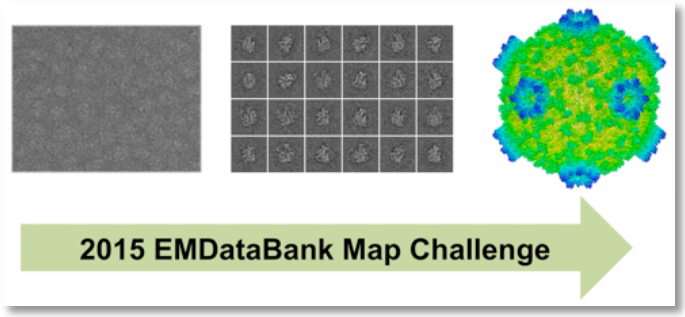
- Goals: Develop benchmarks, encourage development of best practices in **3DEM reconstruction and model fitting**, evolve criteria for validation, compare and contrast different approaches
- Targets are raw image datasets in EMPIAR, maps in EMDB



More info: http://challenges.emdatabank.org

2015 Map Challenge

- Goals:
  - Establish a benchmark set of single particle raw image datasets suitable for high-resolution cryoEM, suitable for both software developers and beginners
  - Encourage developers of 3DEM software packages and biological end users to analyze these datasets and present results using their best practices
  - Evolve criteria for evaluation and validation of the results of the reconstruction and analysis
  - Compare and contrast the various reconstruction approaches in a positive spirit, to achieve high efficiency and accuracy
- Committee: B Carragher (chair), JM Carazo, W Jiang, J Rubinstein, P Rosenthal, F Sun, J Vonck, EMDataBank reps
- 2015: development and challenge phase
- 2016: assessment phase



2015 Model Challenge

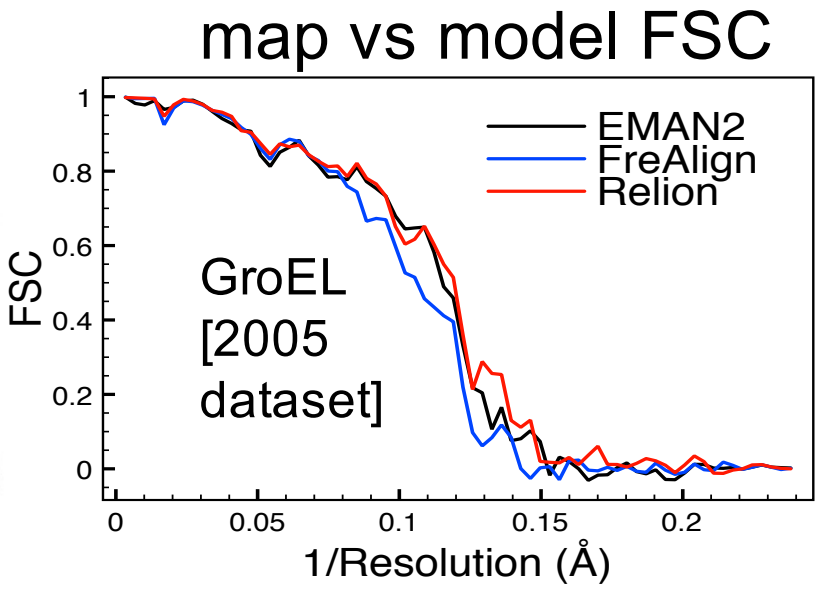
- Goals:
  - Establish a benchmark set of 3DEM maps in the 3.0-4.5 Å resolution range, where significant growth in the number of maps is anticipated over the next few years and where a number of technical challenges exist to map interpretation and fitting
  - Encourage developers of modelling software packages and biological end users to analyze these maps and present modelling results using their best practices
  - Evolve criteria for evaluation and validation of 3DEM map-derived models
  - Compare and contrast the various modelling and analysis approaches in a positive spirit
- Committee: P Adams (chair), A Brunger, R Read, T Schwede, M Topf, EMDataBank reps
- 2015: development
- 2015/2016: challenge
- 2016: assessment

Validation-related work as part of the NIH grant

- At Baylor
  - Map validation
  - Map/model validation
- At RCSB
  - Trace-atom models
  - Nucleic acid models
- At PDBe
  - FSC server
  - Tilt-pair validation server
  - Validation pipeline

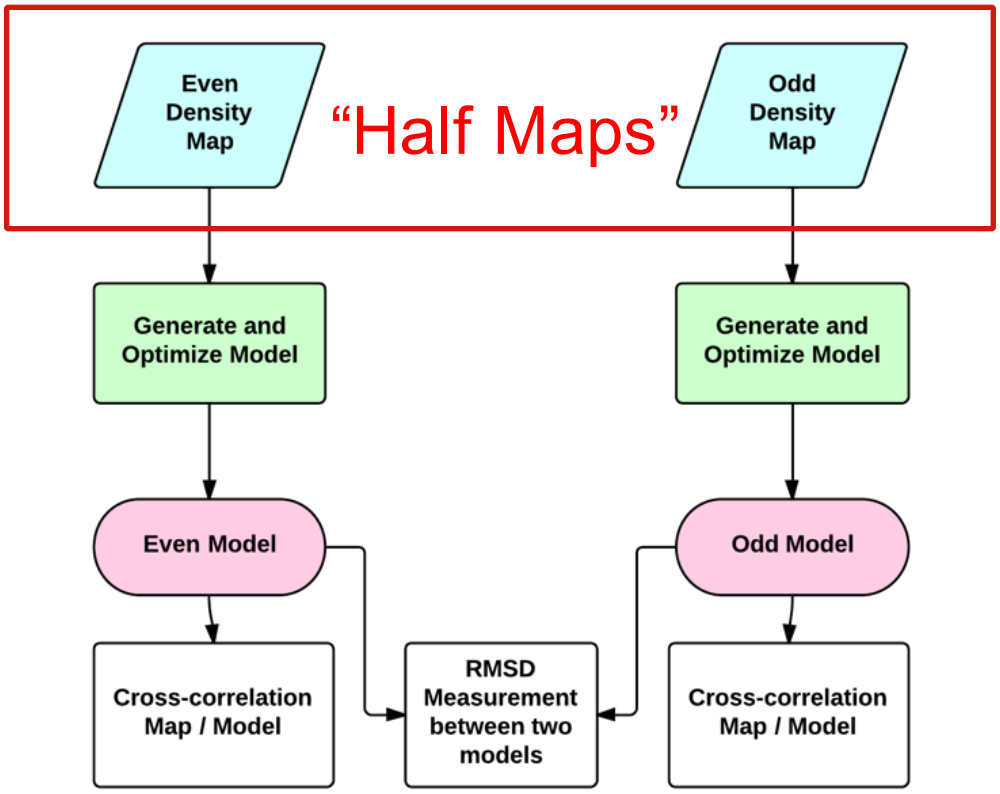
Map Validation

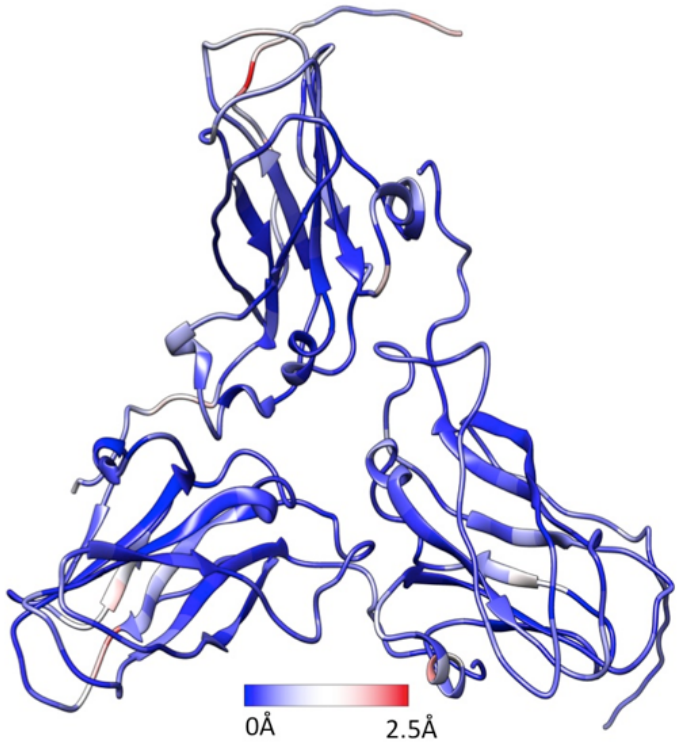
- Development of new strategies for reconstruction overfitting
- Comparisons of different reconstruction software packages/algorithms



Map/Model Validation

- Medium Resolution (~10 Å): Monitoring for overfitting through FSC analysis
- High Resolution (3-5 Å): Model refinement incorporating cross-correlation, atom position uncertainty

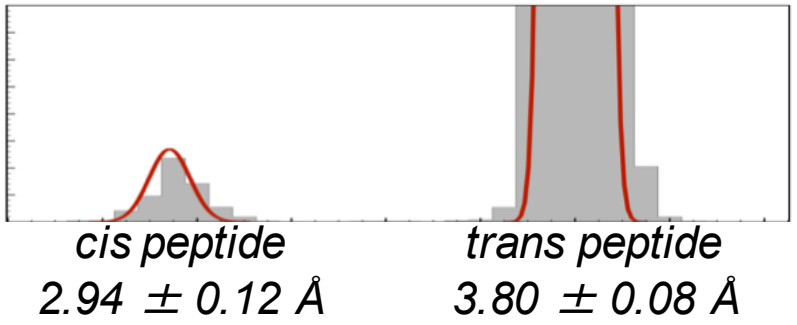


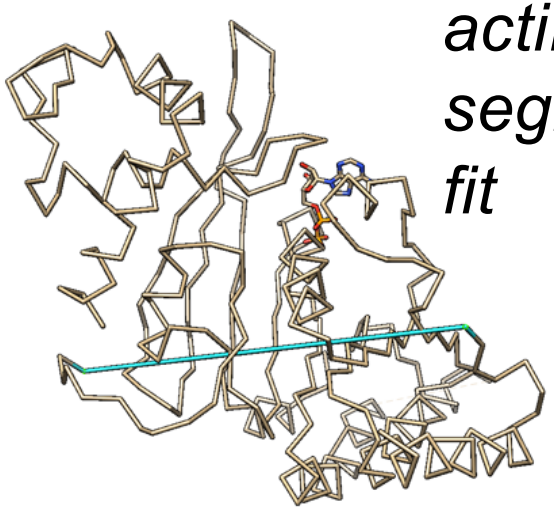


Wang et al., Nat Commun. (2014)

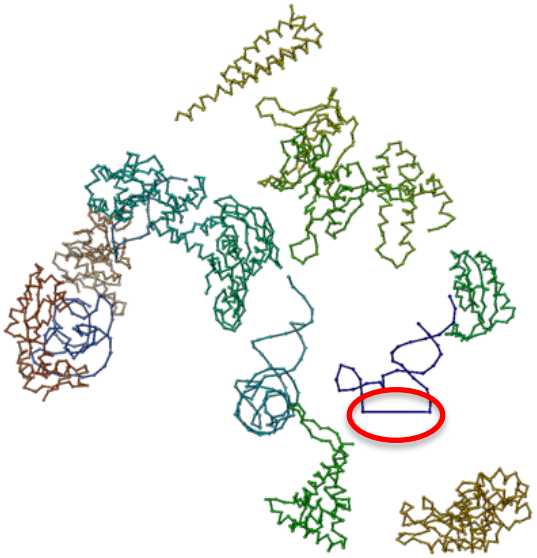
Model-only Validation: Trace Atoms

- Outlier criteria established
  - Cα-Cα distances outside ±3σ
  - P-P distances <4.4 or >8.0 Å





1N03
actin, 2 segment fit



1EGO
ribosome partial model

Model-only Validation: Trace Atoms

5.9.1 Abnormal distances between consecutive trace atoms

The following table provides a summary of the distances observed betwe
(in protein chains) or P atoms (in RNA or DNA chains).

Mol	Chain	Analysed	Outliers
1	A	300/352 (85.2%)	1/300 (0.3%)
1	B	300/352 (85.2%)	1/300 (0.3%)
1	C	300/352 (85.2%)	1/300 (0.3%)
1	D	300/352 (85.2%)	1/300 (0.3%)
1	E	300/352 (85.2%)	1/300 (0.3%)
1	F	300/352 (85.2%)	1/300 (0.3%)
1	G	300/352 (85.2%)	1/300 (0.3%)
All	All	2100/2464 (85.2%)	7/2100 (0.3%)

All (7) distance outliers are listed below:

Mol	Chain	Res	Type	Error type	Distance error
1	A	28	ARG	long	38.47
1	B	28	ARG	long	38.47
1	C	28	ARG	long	38.47
1	D	28	ARG	long	38.47
1	E	28	ARG	long	38.47
1	F	28	ARG	long	38.47
1	G	28	ARG	long	38.47

1N03

5.9 Trace atom model geometry

5.9.1 Abnormal distances between consecutive trace atoms

The following table provides a summary of the distances observed between consecutive Cα atoms (in protein chains) or P atoms (in RNA or DNA chains).

Mol	Chain	Analysed	Outliers
1	I	32/33 (97.0%)	5/32 (15.6%)
10	F	87/89 (97.8%)	0/87 (0.0%)
11	G	88/93 (94.6%)	4/88 (4.5%)
12	H	99/100 (99.0%)	4/99 (4.0%)
13	N	219/229 (95.6%)	1/219 (0.5%)
14	J	163/171 (95.3%)	2/163 (1.2%)
15	K	132/140 (94.3%)	0/132 (0.0%)
2	L	56/57 (98.2%)	0/56 (0.0%)
3	M	25/26 (96.2%)	0/25 (0.0%)
4	O	75/76 (98.7%)	0/75 (0.0%)
5	A	158/159 (99.4%)	0/158 (0.0%)
6	B	144/148 (97.3%)	1/144 (0.7%)
7	C	96/97 (99.0%)	0/96 (0.0%)
8	D	134/146 (91.8%)	0/134 (0.0%)
9	E	135/138 (97.8%)	0/135 (0.0%)
All	All	1643/1702 (96.5%)	17/1643 (1.0%)

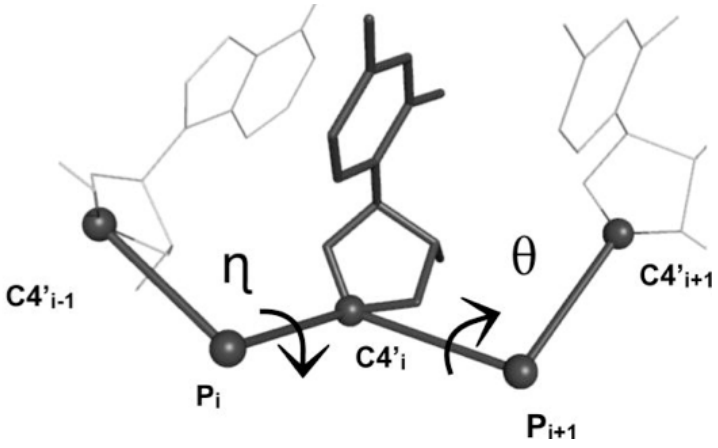
All (17) distance outliers are listed below:

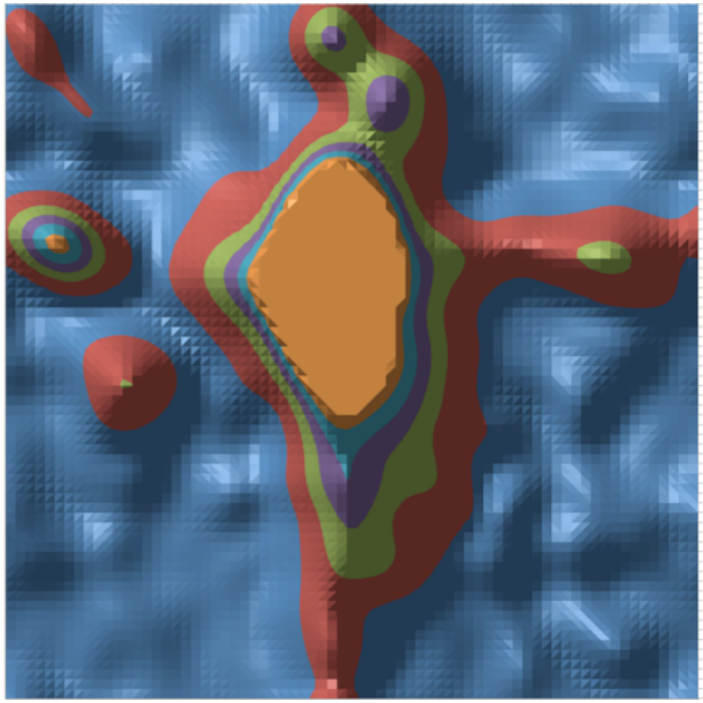
Mol	Chain	Res	Type	Error type	Distance error
1	I	14	N	long	16.52
1	I	3	N	long	3.38
1	I	27	N	long	1.46
12	H	36	UNK	short	1.32
11	G	81	ARG	short	1.10
12	H	58	UNK	short	1.09
14	J	75	GLY	long	1.05
13	N	152	ARG	short	0.64
1	I	29	N	long	0.63
12	H	37	UNK	long	0.61
11	G	53	GLY	long	0.34
1	I	25	N	long	0.24
14	J	74	LYS	long	0.12
11	G	46	GLU	abnormal	0.11
11	G	78	LYS	short	0.07
6	B	26	ARG	long	0.01
12	H	66	UNK	abnormal	0.00

1EGO

Model-only Validation: Nucleic Acids

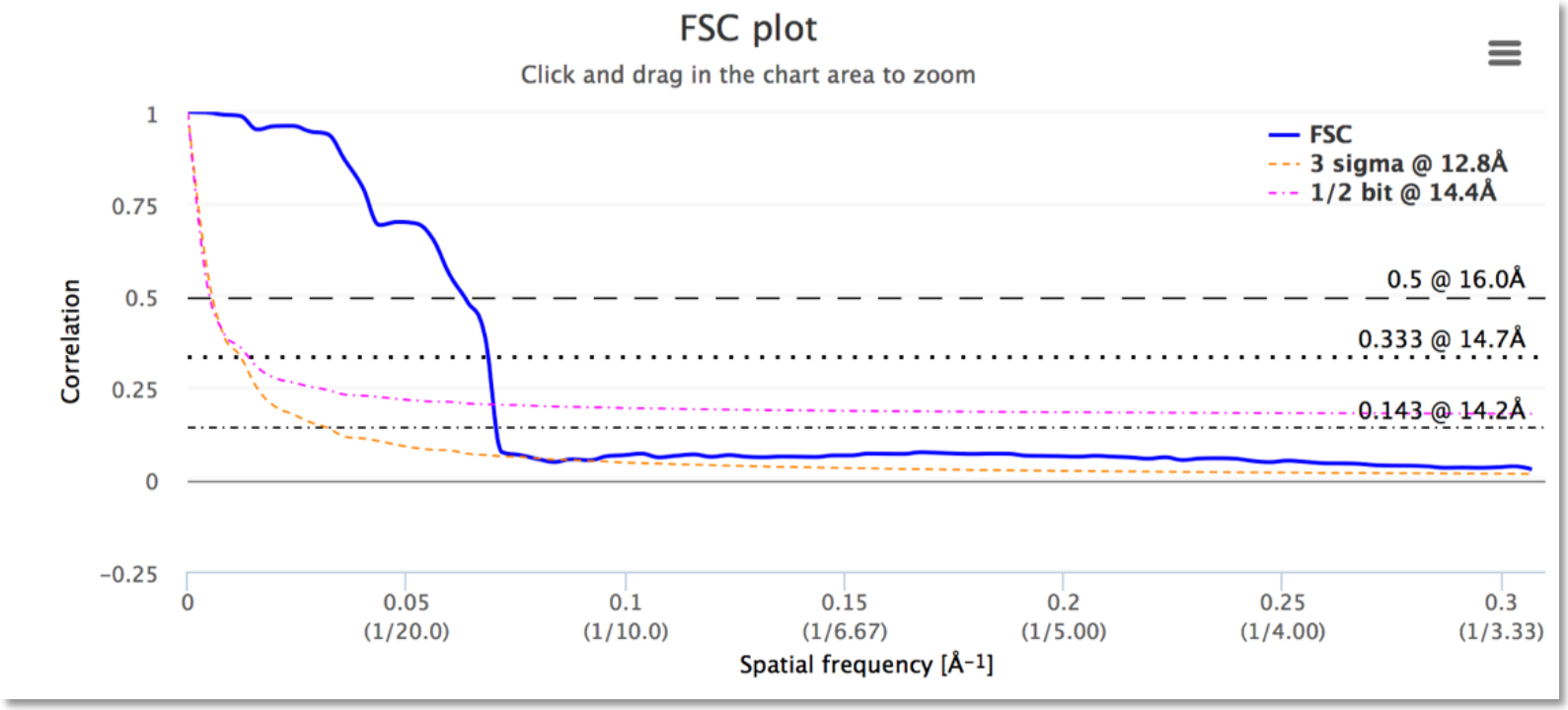
- Looking into use of pseudo-bond plot (Anna Pyle Lab) for new backbone validation criterion
- Peaks correspond to discrete conformations, strong dependence on consecutive base stacking found





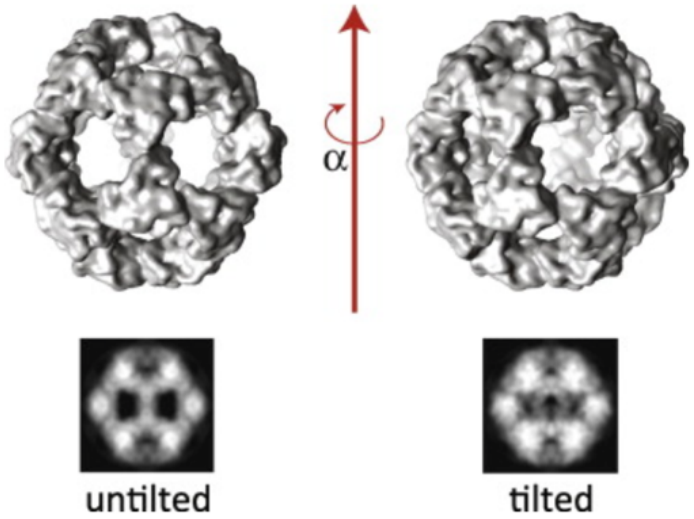
Fourier Shell Correlation (FSC) server

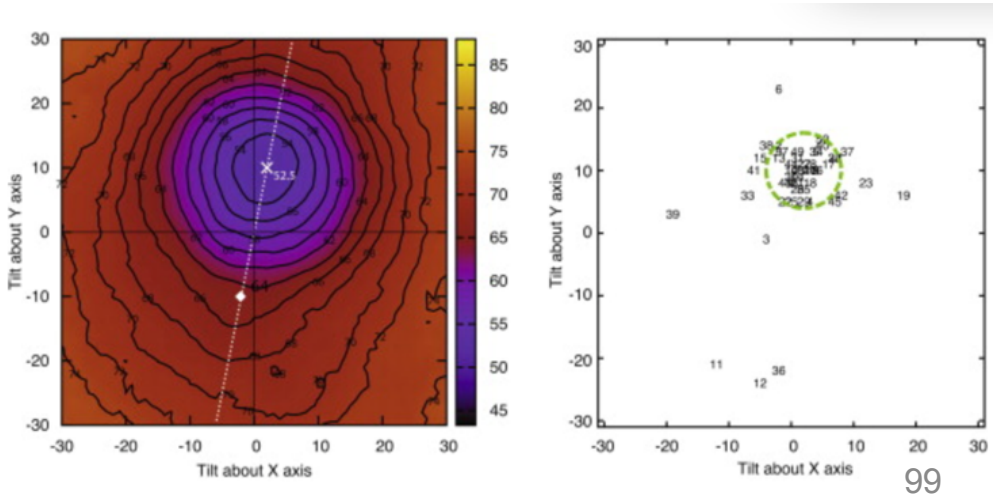
- User uploads two maps - server calculates FSC curve
- FSC curve can be deposited to EMDB
- Automated calculation of thresholds
- Bsoft, Relion and EMAN2 generate FSC curves that can be deposited as well
- pdbe.org/fsc



Tilt-pair validation server

- Validation method to check angle assignment to single-particle projections
- Procedure:
  - Collect two sets of micrograph pairs – one untilted and one tilted
  - Reconstruct untilted set
  - Check if reprojections from 3D with added tilt angle match corresponding images from tilted set
- Code provided by Rosenthal lab
- Deposition of tilt-pair data to EMDB still to be implemented
- pdbe.org/tiltpair





D&A V2.0 and Beyond

D&A Status

- D&A 2.0 not yet in production
- Limitations and assumptions in 1.5 that prevented it to work with NMR and EM
  - File-upload and management – needs to support multiple instances of data files (e.g., EM maps and masks, NMR chemical shifts)
  - Deposition in absence of “molecule” (i.e., no model coordinate file)
  - Move processing away from webservers to support large structures
  - Add support for separate ftp upload of large files

D&A Status

- Progress
  - All EM sub-methods implemented
  - Added support for neutron diffraction
  - NMR validation pipeline
  - EM validation pipeline
  - Improvements to annotation system
  - Issues revealed in “round-trip” testing (dep -> ann -> back to dep) addressed

Roadmap for V2.0

- System ready at Rutgers for internal testing starting 2nd October 2015
- Annotators training for all the wwPDB partner sites
- Tutorials for depositors
- Issues arising to be prioritised and addressed
- Limited public beta testing at Rutgers (real depositions by selected external depositors) in the middle of November 2015
- Issues arising to be prioritised and addressed
- System to be ported to UCSD and PDBj
- Make validation reports etc. available for all archive entries (NMR, EM; updated X-ray reports)
- Public release and publicity possibly in January 2016
- After release, system will be ported to EBI

Looking Ahead, Questions, and Discussion

WORLDWIDE PDB
PROTEIN DATA BANK

wwpdb.org

Plans For the Coming Years

- 2015
  - X-ray VTF 2.0 at EBI on November 16-17
- 2016
  - Phasing out of legacy deposition & annotation systems for NMR and EM
  - Start carbohydrate remediation
  - wwPDB AC meeting at Madison-Wisconsin
  - wwPDB Symposium on Data Mining and Advanced Analytics at Rutgers
- 2017
  - Extend the franchise to appropriately qualified wwPDB partner sites in China and India

BMRB Will Host the 2016 wwPDB AC

- Date: Friday, October 7, 2016

- Location:

  Wisconsin Institute for Discovery
  University of Wisconsin-Madison
  Madison, Wisconsin
  USA

Questions for the wwPDB AC

1. Does the wwPDB AC concur with the recommendation by the wwPDB Partners (RCSB PDB, PDBe, PDBj, and BMRB) that the ORCID system be adopted for definitive identification of deposition contact authors as outlined in Appendix 1?

1. Does the wwPDB AC concur with adoption of the Principles Guiding Federation of Data Resources with the PDB Archive as outlined in Appendix 2?

Questions for the wwPDB AC (cont.)

3. Does the wwPDB AC endorse the wwPDB Partners collaborating with the Integrative/Hybrid Methods Working Group to establish a Federation of Data Resources to enable inclusion of Integrative/Hybrid 3D structural models in the PDB Archive together with certain primary data and metadata (sufficient for validation)?

4. Does the wwPDB AC concur with adoption of the Principles Guiding Versioning of PDB Archival Entries outlined in Appendix 3?

Questions for the wwPDB AC (cont.)

5. Does the wwPDB AC concur with adoption of the Principles Guiding Limited Inclusion of wwPDB Annotations by the wwPDB Partners outlined in Appendix 4?