2017 wwPDB AC Meeting

Stephen K. Burley, Genji Kurisu,
John L. Markley, and Sameer Velankar

WORLDWIDE PDB
PROTEIN DATA BANK

wwpdb.org

Introductions and State of the wwPDB

Stephen K. Burley

WORLDWIDE PDB
PROTEIN DATA BANK

wwpdb.org

Welcome

wwPDB Advisory Committee Chair and
ICMRBS Representative

- R. Andrew Byrd

wwPDB Advisory Committee Members

- RCSB PDB: Paul Adams and Cynthia Wolberger
- PDBe: David Brown and Sarah Butcher
- PDBj: Tsuyoshi Inoue and Kei Yura
- BMRB: Gaetano Montelione and Arthur Edison

Welcome (cont.)

Regional Representatives

- China: Jianping Ding
- India: Manju Bansal

IUCr Representative

- Edward Baker

Macromolecular EM Community Representative

- Wah Chiu

Logistics/Support

In Case of Fire

- Emergency exits at both ends of corridor
- Rally point in front of medical school tower

Restrooms Across the Hall

RCSB PDB Administration Room 110

- Nicole Oorbeek (nicole.oorbeek@rcsb.org)
  Tel: 848-445-4903; Mobile: 732-859-9040

wwPDB Vision

Our Vision is to

Sustain a freely accessible, single global archive of experimentally determined structure data for biological macromolecules as an enduring public good.

wwPDB Mission

Our Mission is to

1. Ensure open access to public domain experimentally determined structural biology data.
2. Provide expert deposition, validation, and biocuration services at no charge to Data Depositors.
3. Enable universal access for expert and non-expert Data Consumers with no limitations on usage.
4. Manage the PDB archive as a public good according to the *FAIR* Principles.
5. Lead the world in structural biology data representation, exchange, and visualization.

Developments since 2016 Meeting I

- Genji Kurisu succeeded Nakamura as PDBj Head
- Continued enhancement of the OneDep system for MX, NMR, 3DEM, and now SAS
- Rapid growth in 3DEM structure depositions and engagement with the 3DEM community
- Continued engagement with the NMR community

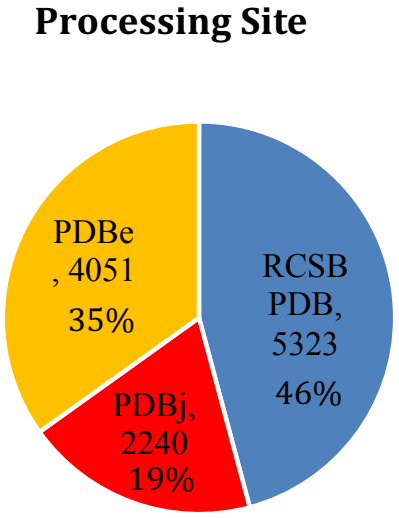
Developments since 2016 Meeting II

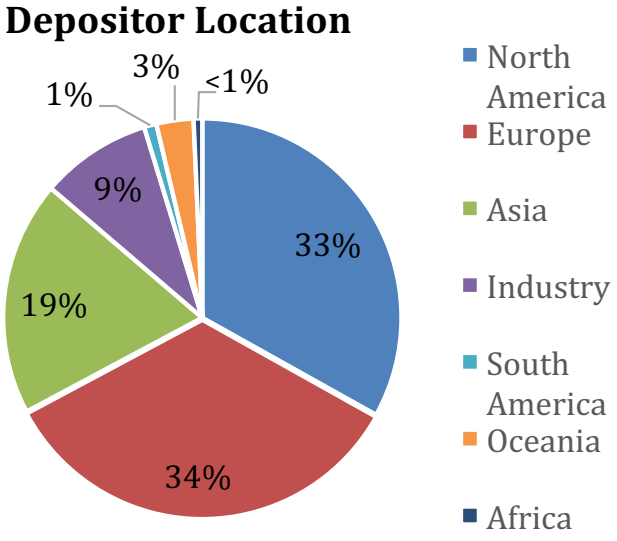
- Implementation of PDB archive versioning initiated
- Prototyping of a loosely coupled Federation of Structural Biology Data Archives with SASBDB
- Archival content and management improvement
- Collection of ligand of interest and experimental support of multimeric assemblies

2016 PDB Deposition Statistics

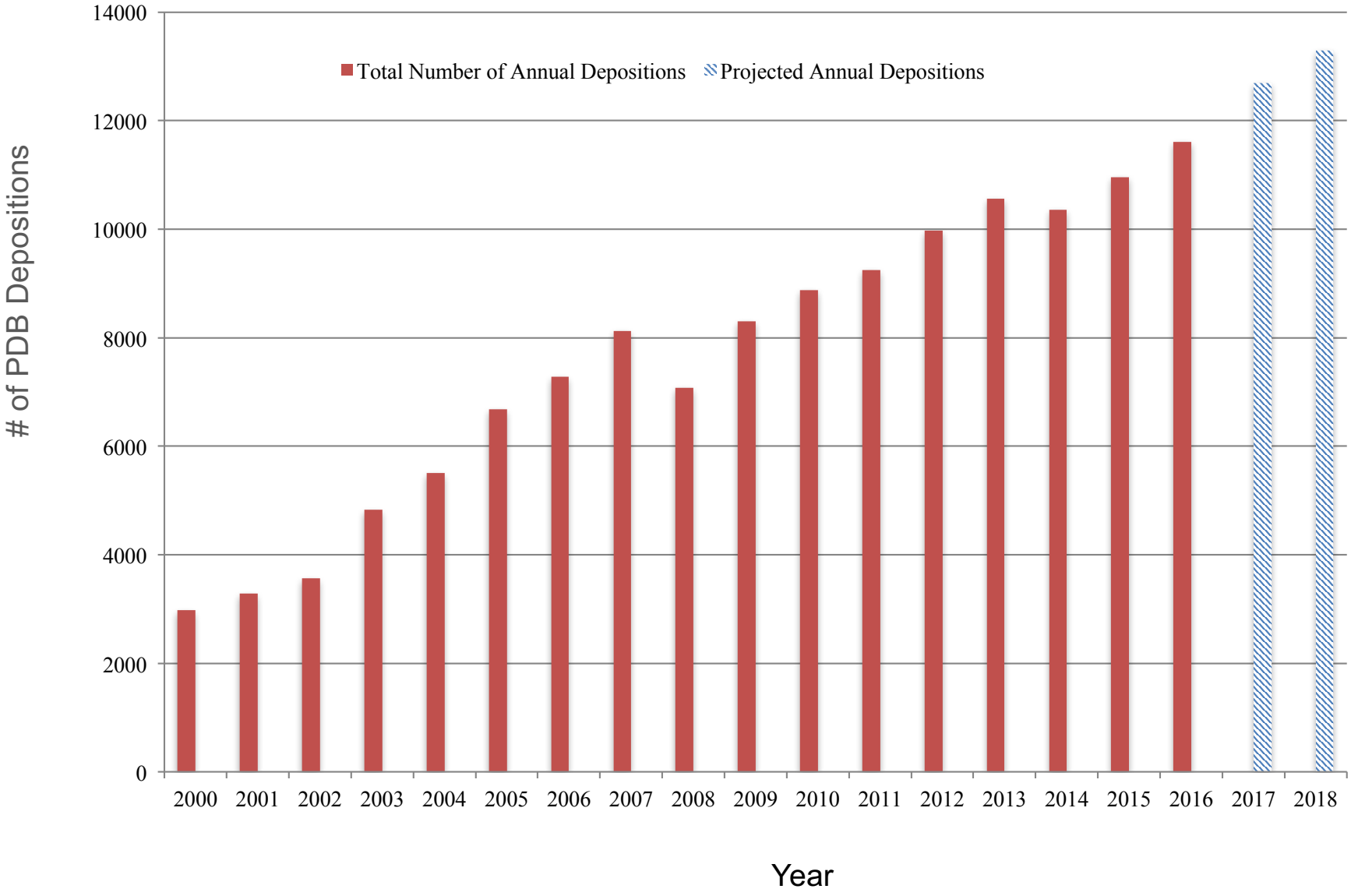
- On track for >12,000 depositions in 2017
- More 3DEM structures deposited in 2016 *versus* NMR structures
- On track for repetition of 3DEM>NMR in 2017

Method	2016 Depositions	2015 Depositions
MX	10583 (91%)	10167
NMR	473 (4%)	510
3DEM	531 (4.6%)	255
Other	27 (.2%)	25

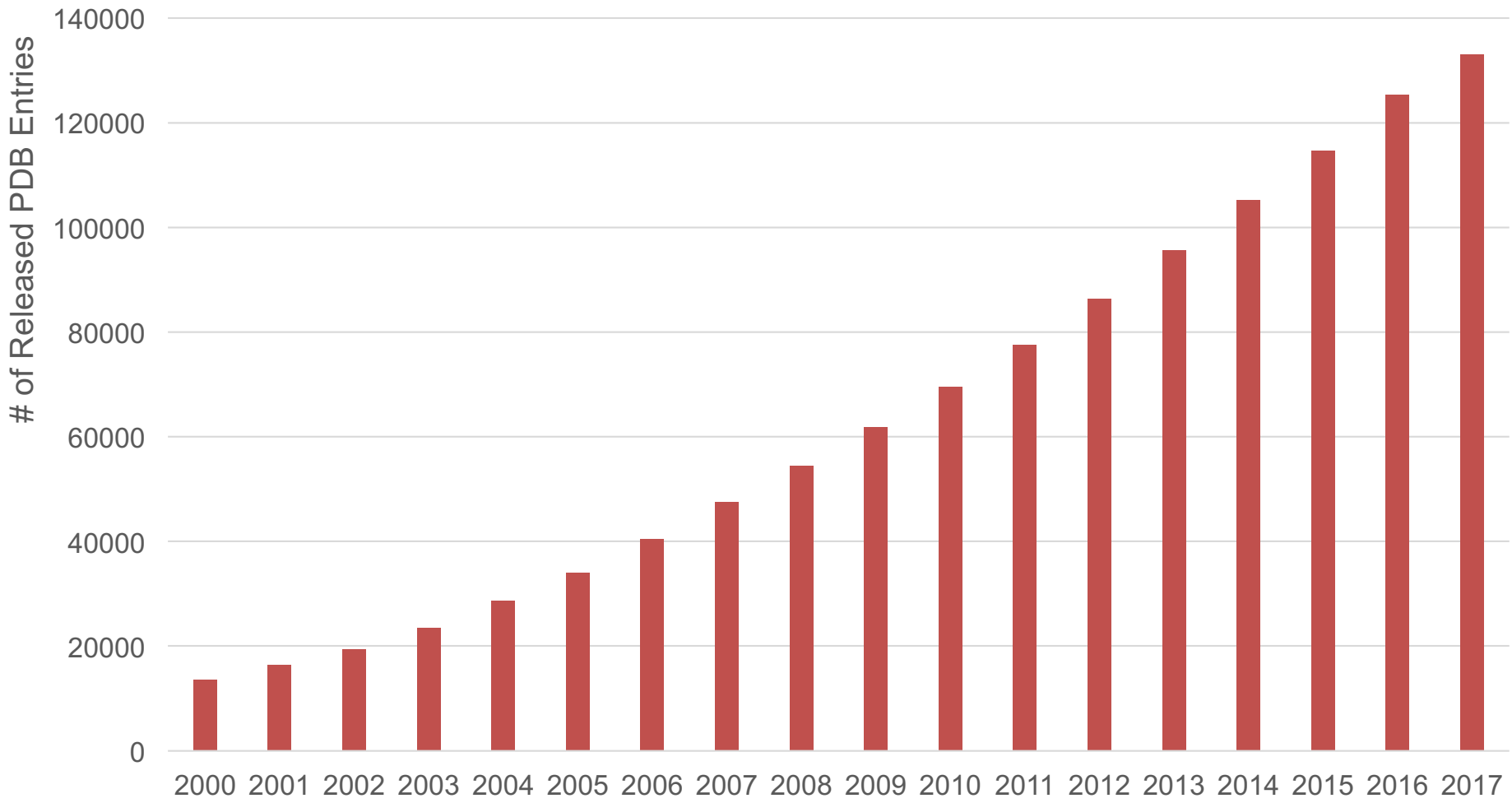




Growth in Annual PDB Depositions



Growth of the PDB Archive



More than 1 billion atoms

PDB Data Download Statistics

Year	Total	Total FTP Archive	Total Website	RCSB PDB FTP Archive	RCSB PDB Website	PDBe FTP Archive	PDBe Website	PDBj FTP Archive	PDBj Website
2009	328,362,536	271,116,934	57,245,602	222,984,760	53,507,785	30,141,339	1,475,116	17,990,835	2,262,701
2010	294,326,976	213,180,966	81,146,010	159,248,214	64,569,658	34,383,219	14,017,349	19,549,533	2,559,003
2011	383,131,048	276,952,286	106,178,762	204,939,406	81,560,098	40,960,368	18,515,245	31,052,512	6,103,419
2012	376,944,070	255,837,735	121,106,335	213,510,347	90,438,501	21,601,103	23,982,801	20,726,285	6,685,033
2013	441,262,210	296,176,290	145,085,920	215,331,908	97,549,580	43,684,850	37,762,496	37,159,532	9,773,844
2014	512,227,251	339,193,721	173,033,530	237,168,615	110,115,316	52,362,370	48,031,414	49,662,736	14,886,800
2015	534,339,871	368,244,766	166,095,105	255,346,630	111,802,897	48,544,330	41,127,219	64,353,806	13,164,989
2016	591,876,087	366,677,897	225,198,190	293,648,366	161,208,456	30,274,284	44,432,830	42,755,247	19,556,904

More than 1.5 million/day

Geographic origins of FTP downloads, 2012-2015

Funding Status Update

- RCSB PDB: NSF/NIH/DOE funding through 12/31/2018 (Competing renewal process)
- BMRB: NIH NIGMS funding through 03/31/2019 (Competing renewal process)
- PDBe: EMBL-EBI, Wellcome Trust through 01/01/2020 (Competing renewal process)
- PDBj: NBDC-JST and AMED funding through 03/31/2022 (Competing renewal process)

wwPDB Collaboration Staffing Commitments Oct 2017-Sep 2018

wwPDB Partner	Software Development	Production Maintenance/ Production Management	Requirements Setting/ Testing	Archive Keeping/ Outreach	Biocuration/ Remediation	Total FTE Commitments
RCSB PDB	3.2	1.5	1.0	2.1	7.0	14.8
PDBe	2.9	1.0	0.5	0.1	4.0	8.5
PDBj	0.1	0.9	0.5	0.5	4.3	6.3
BMRB	1.25	--	--	--	--	1.25
wwPDB	7.45	3.4	2.0	2.7	15.3	30.85

OneDep 2016/2017 Progress *vs*. Goals

	Projects	Timeline
		2016 Q4 | 2017 Q1 Q2 Q3 Q4 | 2018 Q1 Q2 Q3 Q4
Archive content improvement: V5	Phase 1: Migrate and remediate legacy EM entries to V5 files	
	Phase 2: Migrate and remediate legacy NMR and X-ray entries to V5	
	Phase 3: Update FTP archive with remediated V5 files	
Archive content improvement: Author of record file versioning (Parallel versioned ftp tree)	Phase 0: Provide better history revision record in the archive files	
	Phase 1: Provide versioned FTP tree for model files	
	Phase 2: Enable depositor initiated coordinate replacement	
Archive content improvement: Carbohydrate and Protein modification remediation	Phase 1: Finalize representation requirements	
Backend stabilization	Managing the lifecycle of inactive un-submitted sessions	
	Ensure all traffic is encrypted (https)	
	Separate user account from DepUI to support distributed access	
	Enable WF to use external computing resources	
OneDep public facing	Extend collection of ORCID ID	
	Better collection of exp. evidence for depositor's assemblies	
	Collection of ligand info that the focus of depositor's research	
	Inclusion of SAXS/NMR hybrid	
	Re-use previous annotation during coordinate replacement post submission	
	Enable upload of NEF format file for NMR restraint data	
Validation report	Recalculation of archive reports w/ new stats	
	Implement EM MAP validation	
Validation report: Improve ligand validation	Phase 1: Improve ligand handling in the preliminary validation	
	Phase 2: Provide quality indicator in ligand 2D depiction	
	Phase 3: Provide ligand density map in the validation report	
Improve biocuration pipeline	Provide method specific view for CIF Editor	
	Re-use previous annotation for re-processing	
	Improve WF to increase efficiency on processing large structures	
Publication	Validation paper	
	Biocuration paper	

* bold items were re-forecast to 2018

OneDep 2017/2018 Goal Setting

	Projects	Estimated Timeline				
		2017	2018			
		Q4	Q1	Q2	Q3	Q4
Archive content improvement: Author of record file versioning (Parallel versioned ftp tree)	Phase 2: Enable depositor initiated coordinate replacement	■	■			
	Phase 3: Expand versioned FTP tree to include remaining content			■	■	■
Archive content improvement: Carbohydrate and Protein modification remediation	Phase 2: Software development and testing		■	■		
	Phase 3: Produce test data set for internal and friendly testing				■	■
Archive content improvement	Remediate existing XFEL entries according to new mmCIF schema				■	
	Plan update of archival SF files with new data organization				■	■
Backend stabilization	Provide more automated testing (re-usable session test cases)	■				
	Separate user account from DepUI to support distributed access	■	■			
	Validation: Develop wrapper for Mogul API	■				
	Enable WF to use external computing resources		■	■		
	Validation: Modularize reporting, wraper for RDKit/OpenEye			■	■	■
OneDep public facing	Re-use previous annotation during coordinate replacement post submission	■				
	Make collection of ORCID ID mandatory		■			
	Enable upload of NEF format file for NMR restraint data		■	■		
	Support XFEL- collect data related to XFEL technique			■		
Validation	Ligand Phase 2: Provide geometry quality in 2D depiction	■				
	Ligand Phase 3: Provide ligand density map in the validation report		■	■		
	Ligand Phase 4: Provide Mogul standards in the CCD files		■	■		
	Ligand Phase 5: Use Mogul standards in the CCD files for validation			■	■	
	Recalculation of archive reports w/ new stats		■			
	Implement EM map validation			■	■	■
	Implement NMR restraint validation		■	■	■	
	Implement NMR/SAS validation			■	■	■
Improve biocuration pipeline	Re-use previous annotation for re-processing	■				
	Improve miscellaneous UI fixes in Sequence Module	■				
	Increase efficiency on processing large structures		■	■		
	Re-use previous annotation for batch processing				■	■
Publication	Versioning paper			■	■	
	NMR and/or EM validation paper				■	■

* Timeline will be refined further after wwPDB Developer Summit in November 2017

wwPDB Policy Proposals/Discussion

- Resolution of PDB Entry DOIs (Appendix 1)
- wwPDB AC Chair/Co-Chair Restructuring (Appendix 2)
- Invitation to EMDB to Join the wwPDB (Appendix 3)
- Addition of wwPDB Regional Partners (Appendix 4)
- Provision of Mogul Geometry for Ligands (Appendix 5)
- Individual wwPDB Partner AC Reports (Appendix 6)
- Any Requests for Additional Discussion Topics?

Remaining Agenda Items

- Partner Meetings and Outreach: Genji Kurisu
- Macromolecular Crystallography: Stephen K. Burley
- 3D Electron Microscopy: Sameer Velankar
- NMR Spectroscopy: John L. Markley
- Looking Ahead: Sameer Velankar
- Lunch and Executive Session (Noon-1:00pm)
- Questions for the AC: Stephen K. Burley
- Executive Session
- Departure for Dinner in New Brunswick (6:00pm)

Partner Meetings and Outreach

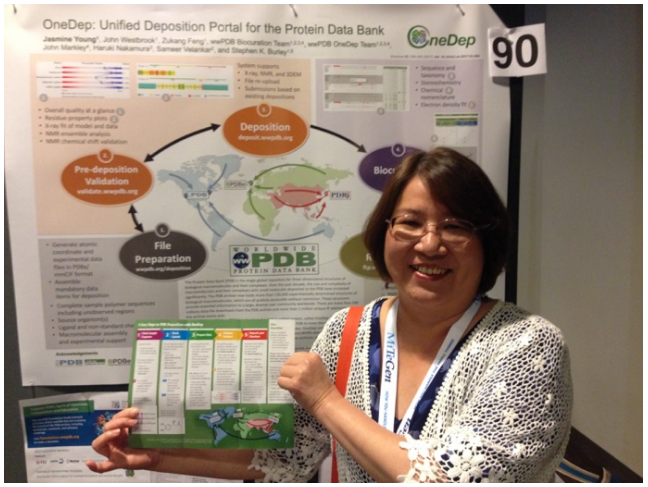
Genji Kurisu

WORLDWIDE PDB
PROTEIN DATA BANK

wwpdb.org

Partner Meetings and Outreach

wwPDB Summit: May EMBL-EBI



OneDep Posters: ACA, ECM, AsCA, APPA, Biocuration Society Meetings

wwPDB PDBx/mmCIF Meeting: July EMBL-EBI

wwPDB Booth: IUCr India

I/HM Book Progress

- Title: "Integrative Structural Biology with Hybrid Methods”
- Publisher: Springer Japan
- Series: Advances in Experimental Medicine and Biology
- Manuscript Submission Deadline: Late 2017

Recent Publications

Chapter 26

Protein Data Bank (PDB): The Single Global Macromolecular Structure Archive

Stephen K. Burley, Helen M. Berman, Gerard J. Kleywegt, John L. Markley, Haruki Nakamura, and Sameer Velankar

Abstract

The Protein Data Bank (PDB)—the single global repository of experimentally determined 3D structures of biological macromolecules and their complexes—was established in 1971, becoming the first open-access digital resource in the biological sciences. The PDB archive currently houses ~130,000 entries (May 2017). It is managed by the Worldwide Protein Data Bank organization (wwPDB; wwpdb.org), which includes the RCSB Protein Data Bank (RCSB PDB; rcsb.org), the Protein Data Bank Japan (PDBj; pdbj.org), the Protein Data Bank in Europe (PDBe; pdbe.org), and BioMagResBank (BMRB; www.bmrb.wisc.edu). The four wwPDB partners operate a unified global software system that enforces community-agreed data standards and supports data Deposition, Biocuration, and Validation of ~11,000 new PDB entries annually (deposit.wwpdb.org). The RCSB PDB currently acts as the archive keeper, ensuring disaster recovery of PDB data and coordinating weekly updates. wwPDB partners disseminate the same archival data from multiple FTP sites, while operating complementary websites that provide their own views of PDB data with selected value-added information and links to related data resources. At present, the PDB archives experimental data, associated metadata, and 3D-atomic level structural models derived from three well-established methods: crystallography, nuclear magnetic resonance spectroscopy (NMR), and electron microscopy (3DEM). wwPDB partners are working closely with experts in related experimental areas (small-angle scattering, chemical cross-linking/mass spectrometry, Forster energy resonance transfer or FRET, etc.) to establish a federation of data resources that will support sustainable archiving and validation of 3D structural models and experimental data derived from integrative or hybrid methods.

Key words Protein Data Bank, PDB, Worldwide Protein Data Bank, wwPDB, PDBx/mmCIF, Chemical Component Dictionary, Crystallography, NMR spectroscopy, NMR-STAR, NMR Exchange Format, NEF, 3D electron microscopy, Integrative or hybrid methods

1 Evolution of Data Sharing and Data Archiving in Structural Biology

The Protein Data Bank (PDB) was established in 1971 with fewer than ten X-ray crystallographic structures of proteins, becoming the first open access digital data resource in the biological sciences [1]. Soon after X-ray structures of myoglobin [2, 3] and hemoglobin [4, 5] were published, the structural biology community

Alexander Wlodawer et al. (eds.), *Protein Crystallography: Methods and Protocols*, Methods in Molecular Biology, vol. 1607,
DOI 10.1007/978-1-4939-7000-1_26, © Springer Science+Business Media LLC 2017

Structure
Resource

OneDep: Unified wwPDB System for Deposition, Biocuration, and Validation of Macromolecular Structures in the PDB Archive

Jasmine Y. Young,[1,12,*] John D. Westbrook,[1] Zukang Feng,[1] Raul Sala,[1] Ezra Peisach,[1] Thomas J. Oldfield,[2,9] Sanchayita Sen,[2] Aleksandras Gutmanas,[2] David R. Armstrong,[2] John M. Berrisford,[2] Li Chen,[1] Minyu Chen,[3] Luigi Di Costanzo,[1] Dimitris Dimitropoulos,[1,10] Guanghua Gao,[1] Sutapa Ghosh,[1] Swanand Gore,[2] Vladimir Guranovic,[1] Pieter M.S. Hendrickx,[2] Brian P. Hudson,[1] Reiko Igarashi,[3] Yasuyo Ikegawa,[3] Naohiro Kobayashi,[3] Catherine L. Lawson,[1] Yuhe Liang,[1] Steve Mading,[4] Lora Mak,[2] M. Saqib Mir,[2] Abhik Mukhopadhyay,[2] Ardan Patwardhan,[2] Irina Persikova,[1] Luana Rinaldi,[2] Eduardo Sanz-Garcia,[2] Monica R. Sekharan,[1] Chenghua Shao,[1] G. Jawahar Swaminathan,[2,11] Lihua Tan,[1] Eldon L. Ulrich,[4] Glen van Ginkel,[2] Reiko Yamashita,[3] Huanwang Yang,[1] Marina A. Zhuravleva,[1] Martha Quesada,[1] Gerard J. Kleywegt,[2] Helen M. Berman,[1] John L. Markley,[1] Haruki Nakamura,[3] Sameer Velankar,[2] and Stephen K. Burley[1,5,6,7,8]

[1]RCSB Protein Data Bank, Center for Integrative Proteomics Research, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA
[2]Protein Data Bank in Europe (PDBe), European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, Cambridgeshire, CB10 1SD, UK
[3]PDBj, Institute for Protein Research, Osaka University, Osaka, 565-0871, Japan
[4]BMRB, BioMagResBank, University of Wisconsin-Madison, Madison, WI 53706, USA
[5]RCSB Protein Data Bank, San Diego Supercomputer Center
[6]Skaggs School of Pharmacy and Pharmaceutical Sciences
University of California San Diego, La Jolla, CA 92093, USA
[7]Institute for Quantitative Biomedicine, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA
[8]Rutgers Cancer Institute of New Jersey, Rutgers, The State University of New Jersey, New Brunswick, NJ 08903, USA
[9]Present address: Dotmatics, The Old Monastery, Windhill, Bishops Stortford, Herts CM23 2ND, UK
[10]Present address: Google, Irvine, CA 92612, USA
[11]Present address: Illumina Cambridge Ltd., Chesterford Research Park, Little Chesterford CB10 1XL, UK
[12]Lead Contact
*Correspondence: jasmine.young@rcsb.org
http://dx.doi.org/10.1016/j.str.2017.01.004

SUMMARY

OneDep, a unified system for deposition, biocuration, and validation of experimentally determined structures of biological macromolecules to the PDB archive, has been developed as a global collaboration by the worldwide PDB (wwPDB) partners. This new system was designed to ensure that the wwPDB could meet the evolving archiving requirements of the scientific community over the coming decades. OneDep unifies deposition, biocuration, and validation pipelines across all wwPDB, EMDB, and BMRB deposition sites with improved focus on data quality and completeness in these archives, while supporting growth in the number of depositions and increases in their average size and complexity. In this paper, we describe the design, functional operation, and supporting infrastructure of the OneDep system, and provide initial performance assessments.

INTRODUCTION

The PDB, which was established in 1971 with just seven X-ray crystal structures of proteins, became the first open-access digital primary data resource in biology (Protein Data Bank, 1971). Today, the PDB archive serves as the single global repository for more than 120,000 experimentally determined atomic-level structures of biological macromolecules (protein, DNA, RNA) and their complexes. The worldwide PDB (wwPDB) partnership, the international collaboration that manages the PDB archive, supports deposition, biocuration, validation, and distribution of PDB data (Berman et al., 2003). This partnership was established in 2003 by three founding members: Research Collaboratory for Structural Bioinformatics PDB or RCSB PDB (Berman et al., 2000), the PDB in Europe or PDBe (Velankar et al., 2016), and the PDB Japan or PDBj (Kinjo et al., 2012). Subsequently, a specialist nuclear magnetic resonance (NMR) spectroscopic data resource, the Biological Magnetic Resonance Data Bank or BMRB (Ulrich et al., 2008), joined the wwPDB. The mission of the wwPDB organization is to ensure that the PDB archive will continue in perpetuity as a high-quality, open-access digital data resource with no limitations on usage (Berman et al., 2013).

The PDB archive has grown substantially during the past 45 years (Figure 1) and now includes structures determined by crystallography (primarily X-ray), NMR spectroscopy, and 3D electron cryomicroscopy (3DEM). The growth of the PDB archive and the increasing complexity of the structures produced by the structural biology community have pushed the limits of

Structure
Letter to the Editor

PDB-Dev: a Prototype System for Depositing Integrative/Hybrid Structural Models

Stephen K. Burley,[1,2,3,*] Genji Kurisu,[4] John L. Markley,[5] Haruki Nakamura,[4] Sameer Velankar,[6] Helen M. Berman,[1] Andrej Sali,[7] Torsten Schwede,[8] and Jill Trewhella[9,10]

[1]RCSB Protein Data Bank, Institute for Quantitative Biomedicine, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA
[2]Skaggs School of Pharmacy and Pharmaceutical Sciences and San Diego Supercomputer Center, University of California, San Diego, La Jolla, CA 92093, USA
[3]Rutgers Cancer Institute of New Jersey, Rutgers, The State University of New Jersey, New Brunswick, NJ 08903, USA
[4]Protein Data Bank Japan (PDBj), Institute for Protein Research, Osaka University, Osaka 565-0871, Japan
[5]BMRB, BioMagResBank, Biochemistry Department, University of Wisconsin-Madison, Madison, WI 53706, USA
[6]Protein Data Bank in Europe (PDBe), European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, Cambridgeshire CB10 1SD, UK
[7]Department of Bioengineering and Therapeutic Sciences, Department of Pharmaceutical Chemistry, California Institute for Quantitative Biosciences, University of California, San Francisco, UCSF MC 2552, Byers Hall at Mission Bay, 1700 4th Street, Suite 503B, San Francisco, CA 94158, USA
[8]SIB Swiss Institute of Bioinformatics and Biozentrum, University of Basel, Klingelbergstrasse 50-70, 4056 Basel, Switzerland
[9]School of Life and Environmental Sciences, The University of Sydney, Sydney, NSW 2006, Australia
[10]Department of Chemistry, University of Utah, Salt Lake City, UT 84112, USA
*Correspondence: stephen.burley@rcsb.org
http://dx.doi.org/10.1016/j.str.2017.08.001

With this Letter to the Editor, the Worldwide PDB (wwPDB) Partnership (wwpdb.org) and the wwPDB Integrative/Hybrid (I/H) Methods Task Force would like to announce public release of a prototype system for depositing I/H structural models, PDB-Development (or "PDB-Dev") (Vallat et al., 2016c). The URL for PDB-Dev is https://pdb-dev.wwpdb.org.

Essential mechanisms in biology frequently involve large macromolecular assemblies (or machines). In favorable cases, their structures can be determined by X-ray crystallography or nuclear magnetic resonance (NMR) spectroscopy or electron microscopy (3DEM) alone, culminating in deposition of atomic structural models (with x, y, z atomic coordinates) into the global PDB archive (PDB; pdb.org). Many such biological machines are, however, poorly suited to single experimental method approaches. Researchers are increasingly forced to combine various experimental data and information from measurements and computational analyses to generate "hybrid" or "integrative" structural models of macromolecular assemblies (Ward et al., 2013). In addition to X-ray crystallography, NMR, and single-particle 3DEM data, structural information in the form of spatial restraints can be obtained from a multitude of measurements, including small-angle scattering, atomic force microscopy, chemical cross-linking, co-purification, Förster resonance energy transfer, electron paramagnetic resonance, mass spectrometry, hydrogen/deuterium exchange, cryo-electron tomography with sub-tomogram averaging, correlative fluorescent light microscopy, and various proteomics and bioinformatics analyses (Ward et al., 2013). I/H approaches have yielded informative structural models of very large macromolecular assemblies, such as the nuclear pore complex and its sub-complexes, the type III secretion system needle, the proteosomal lid sub-complex, the ESCRT-I complex, and an RNA ribosome-binding element from the turnip crinkle virus genome. Despite great need, there are, at present, no standard mechanisms available to represent, deposit, validate, biocurate, archive, disseminate, and visualize I/H models, their supporting experimental data and metadata, and the protocols used to compute the structural models so that they are freely available to researchers and educators around the world. Moreover, some examples of I/H structures have already been submitted to the PDB resource, but currently remain unprocessed owing to the lack of appropriate infrastructure.

To address this challenge, the wwPDB organization sponsored an I/H Methods Task Force workshop in October 2014 at the EMBL-European Bioinformatics Institute (EMBL-EBI). Participants included 38 researchers from Europe, Asia, and North America with expertise in experimental protein structure characterization, computational modeling, visualization, and data archiving (https://www.wwpdb.org/task/hybrid). These experimental and computational scientists, together with wwPDB representatives, contributed to the workshop. Three breakout groups discussed challenges involved in managing I/H structural models and their supporting experimental data. Five consensus recommendations emerging from the meeting were summarized in a White Paper published in *Structure* (Sali et al., 2015), the most important of these being the urgent need for creation of data standards and establishment of a federated system of data resources to standardize representation, validation, archiving, and dissemination of I/H structural models and supporting data.

Two subgroups have been created within the wwPDB I/H Methods Task Force to begin implementing these recommendations (https://pdb-dev.wwpdb.org/contributors/). First is the Model Validation Subgroup, led by Andrej Sali (UCSF) and Torsten Schwede (SIB), which addresses development of methods for I/H model representation, validation, and visualization; and second is the Federation Subgroup, led by Jill Trewhella (Sydney/Utah) and Helen Berman (Rutgers/RCSB PDB), which focuses on building a network of resource partners that can exchange I/H related experimental and structural information in a concerted, systematic manner.

To create data standards for archiving I/H models, we have developed an I/H methods data dictionary that defines the

OneDep Validation Paper In Press Sept 12, 2017

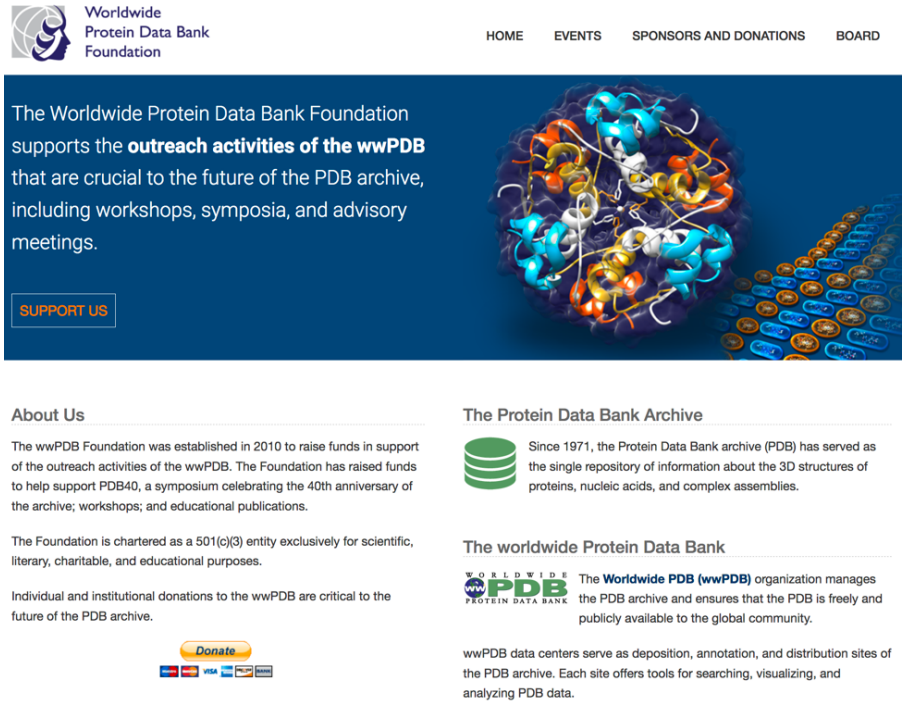
OneDep Biocuration Paper Submitted Oct 11, 2017

World Data System Accreditation Application for PDB Filed Oct 9 2017

- World Data System (WDS) is an Interdisciplinary Body of the International Council for Science (ICSU; http://www.icsu-wds.org)
- Promotes long-term stewardship of, and universal and equitable access to, **quality-assured scientific data and data services, products, and information**
- Copies of WDS application available on side table

wwPDB Foundation Progress

- 501(c)(3) entity exclusively for scientific, literary, charitable, and educational purposes
- Ongoing solicitations of Corporate donations
- Individual Membership program launched with limited success



http://foundation.wwpdb.org/

Macromolecular Crystallography

Stephen K. Burley

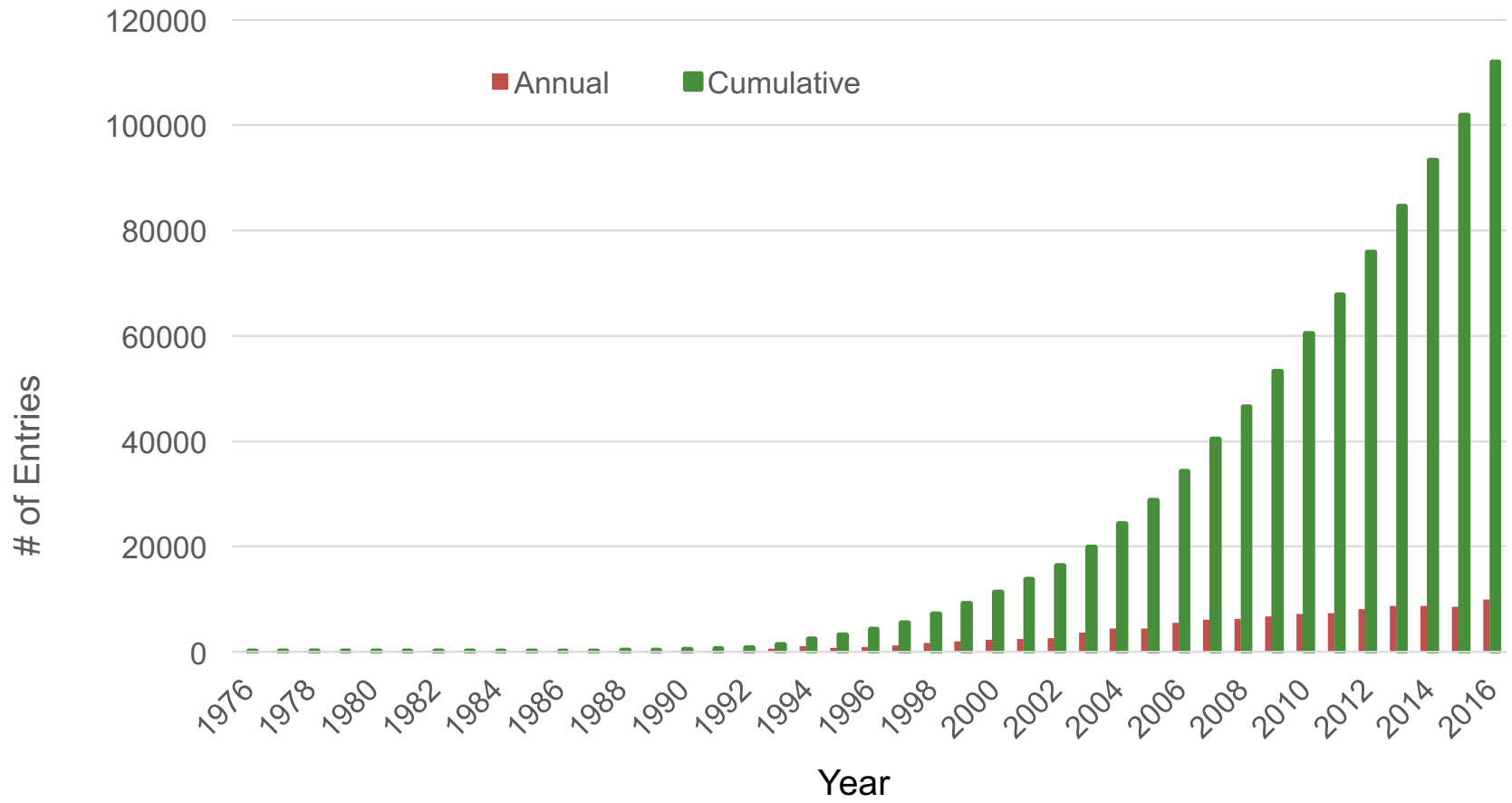
WORLDWIDE PDB
PROTEIN DATA BANK

wwpdb.org

Agenda

- MX Data Deposition Metrics
- 2015 X-ray VTF Meeting Follow Up
- 2017 PDBx/mmCIF Working Group Meeting Outcome

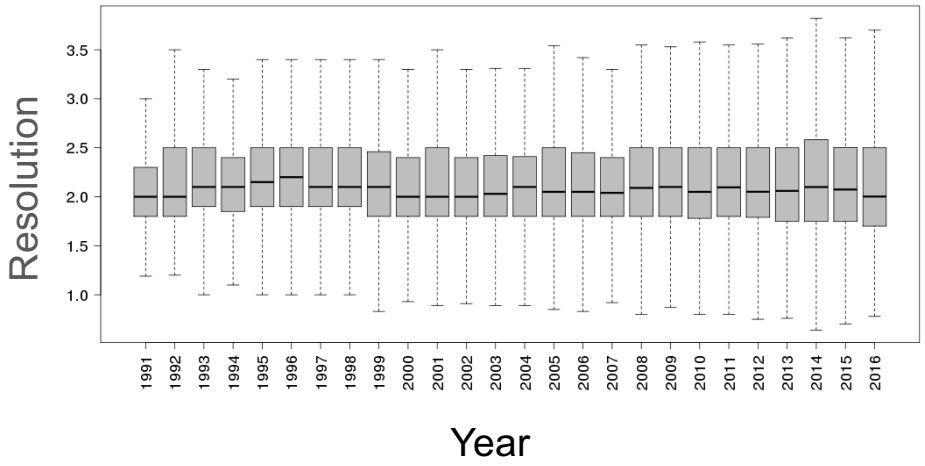
Growth of Released MX Entries



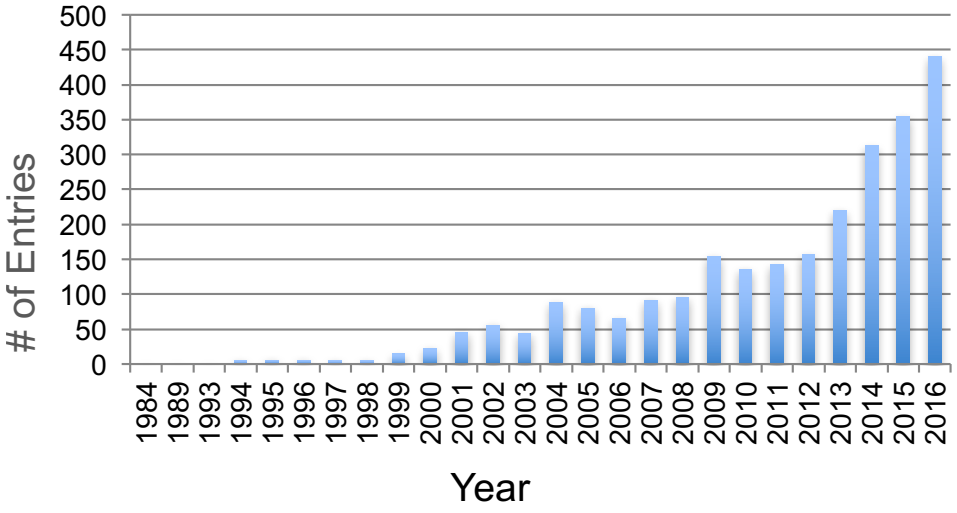
>120,000 Total Released MX Entries Projected for End 2017

MX Deposition Complexity

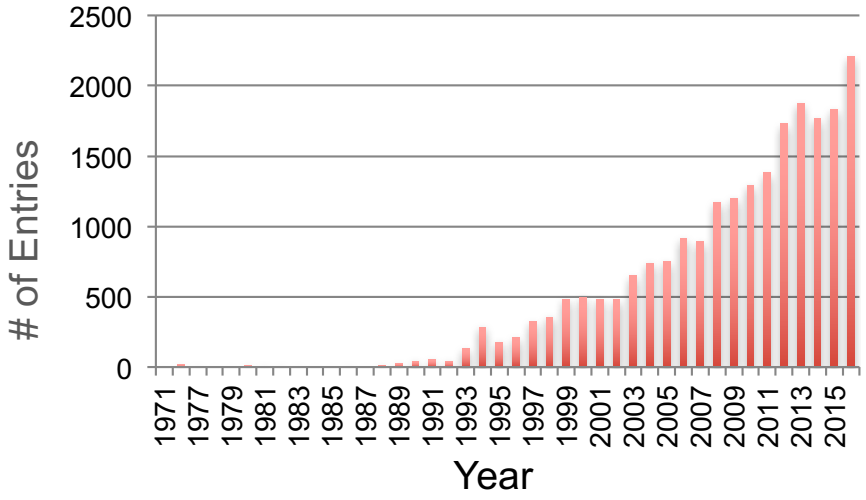
Annual Distribution for High Resolution Limit



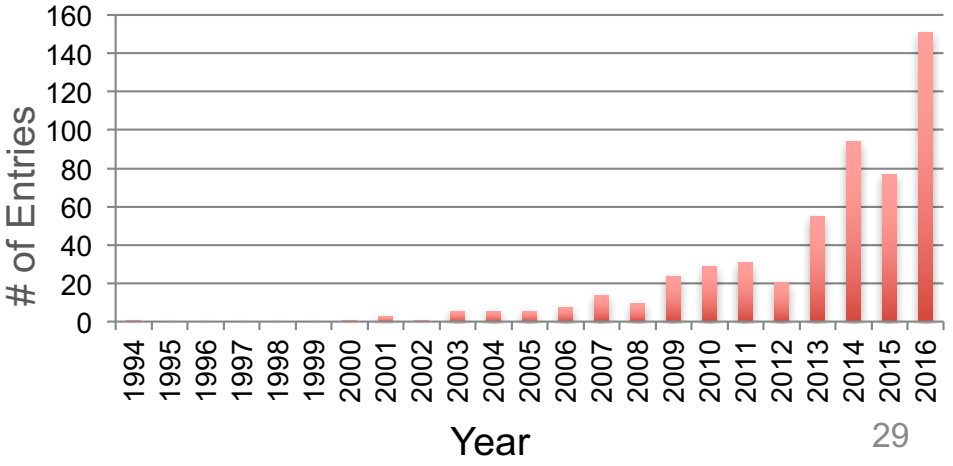
Annual Released Structures With AU MW > 500,000



Total Number of New CCD Entries



Annual Released Large Structures (chains > 62 & atoms > 99999)



X-ray VTF Meeting Update

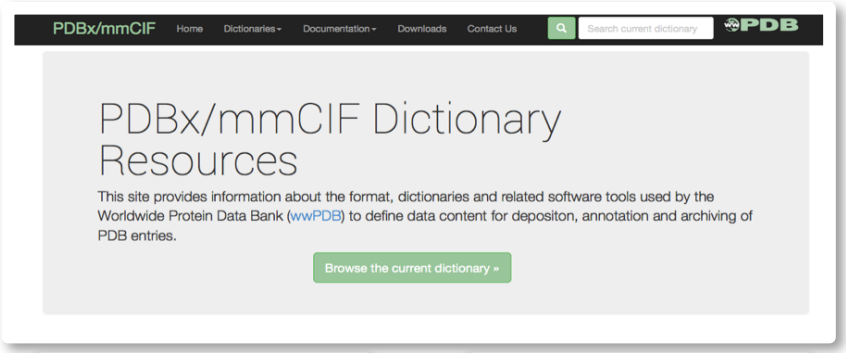
- **Still waiting for 2015 wwPDB VTF Meeting report and recommendations**

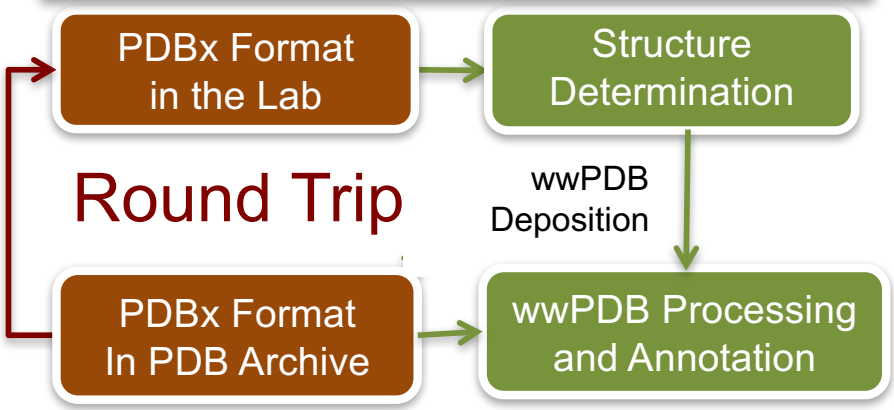
- wwPDB is in the process of implementing recommendations from LVW for ligand validation (verbally endorsed by the wwPDB VTF)

VTF Members: Paul Adams, Gérard Bricogne, Dave Brown, Paul Emsley, Richard Henderson, Nobutoshi Ito, Robbie Joosten, Thomas Lütteke, Michael Nilges, Arwen Pearson, Tassos Perrakis, Randy Read (Chair), Jane Richardson, Janet Smith, Tom Terwilliger, Ian Tickle, Gert Vriend
wwPDB Attendees: Burley, Feng, Gutmanas, Velankar, Westbrook

2017 PDBx/mmCIF Working Group

- PDBx/mmCIF is the archival data standard for the repository
- wwPDB together with the PDBx/mmCIF Working Group of community experts and methods developers oversee the evolution of the standard
- Working Group ensures that the standard is well supported by key community software tools.
- PDB hosts community workshops and maintains mmcif.wwpdb.org serving PDBx/mmCIF data dictionaries, schema and software tools
- **2017 PDBx/mmCIF Working Group meeting finalized new content recommendations for diffraction data and ligand refinement restraint data**





PDBx/mmCIF Workshop Participants, July 2017

3D Electron Microscopy

Sameer Velankar

Agenda

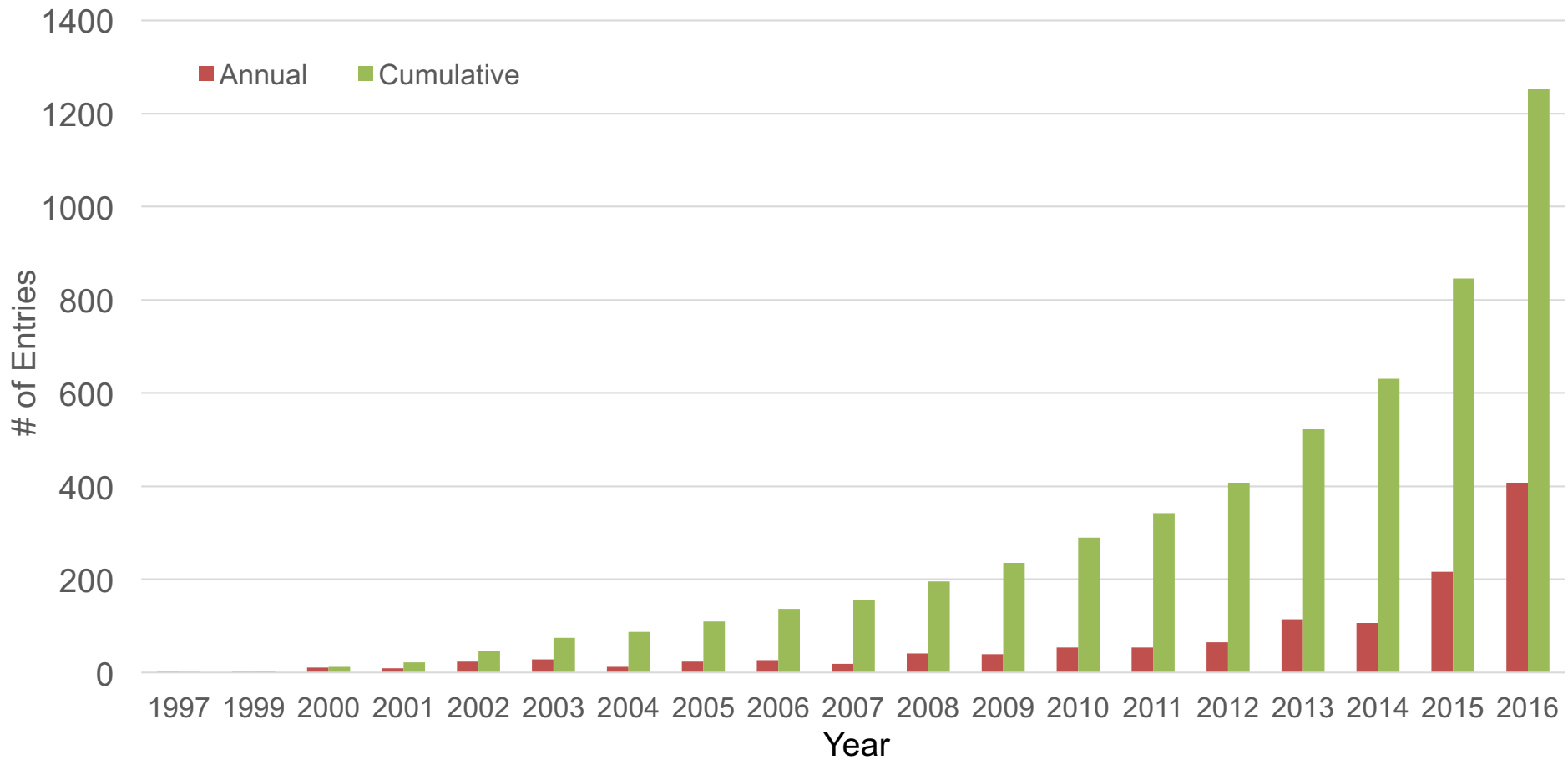
- Archiving 3DEM Data in PDB/EMDB
- 3DEM Data Deposition Metrics
- Status of 3DEM Validation
- Engaging the 3DEM Community and Software Providers

Archiving 3DEM Data in PDB/EMDB

- Integration of 3DEM into OneDep was a coordinated and collaborative effort between wwPDB and EMDataBank
- All 3DEM deposition, biocuration and validation services provided by RCSB PDB, PDBe, and PDBj (OneDep)
- 3DEM Structure Atomic Coordinates archived in PDB
- 3DEM Mass Density Maps archived in EMDB
- EMBL-EBI reorganization yielded independent PDBe (Velankar) and EMDB (Patwardhan) Teams

N.B.: At present, there is no formal wwPDB/EMDB agreement re data sharing, security, and release.

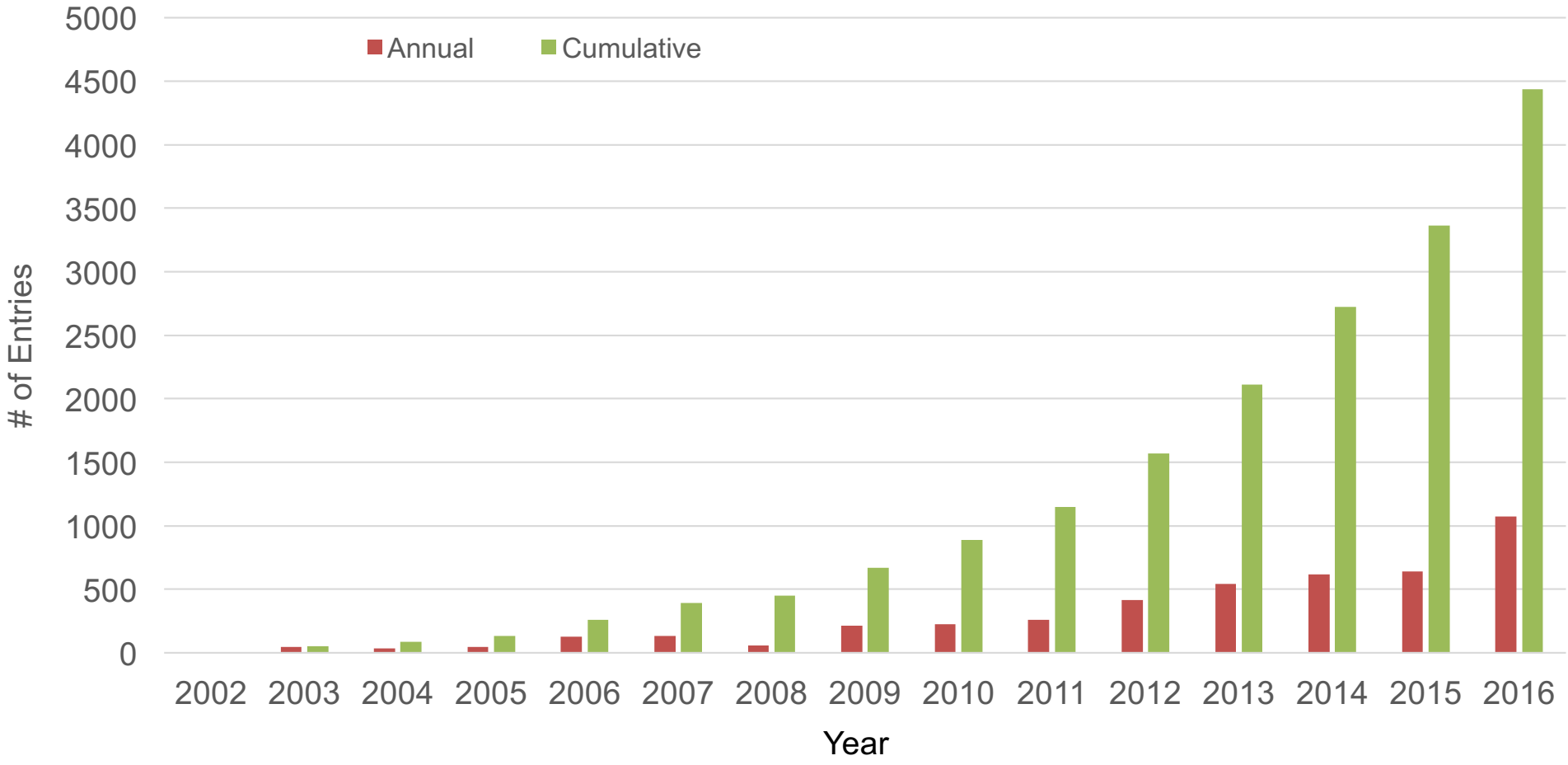
Growth of 3DEM Structures in PDB



As of Oct 1, 2017, >1600 3DEM Structures in PDB Archive

371 new structures released January 1 – September 1, 2017

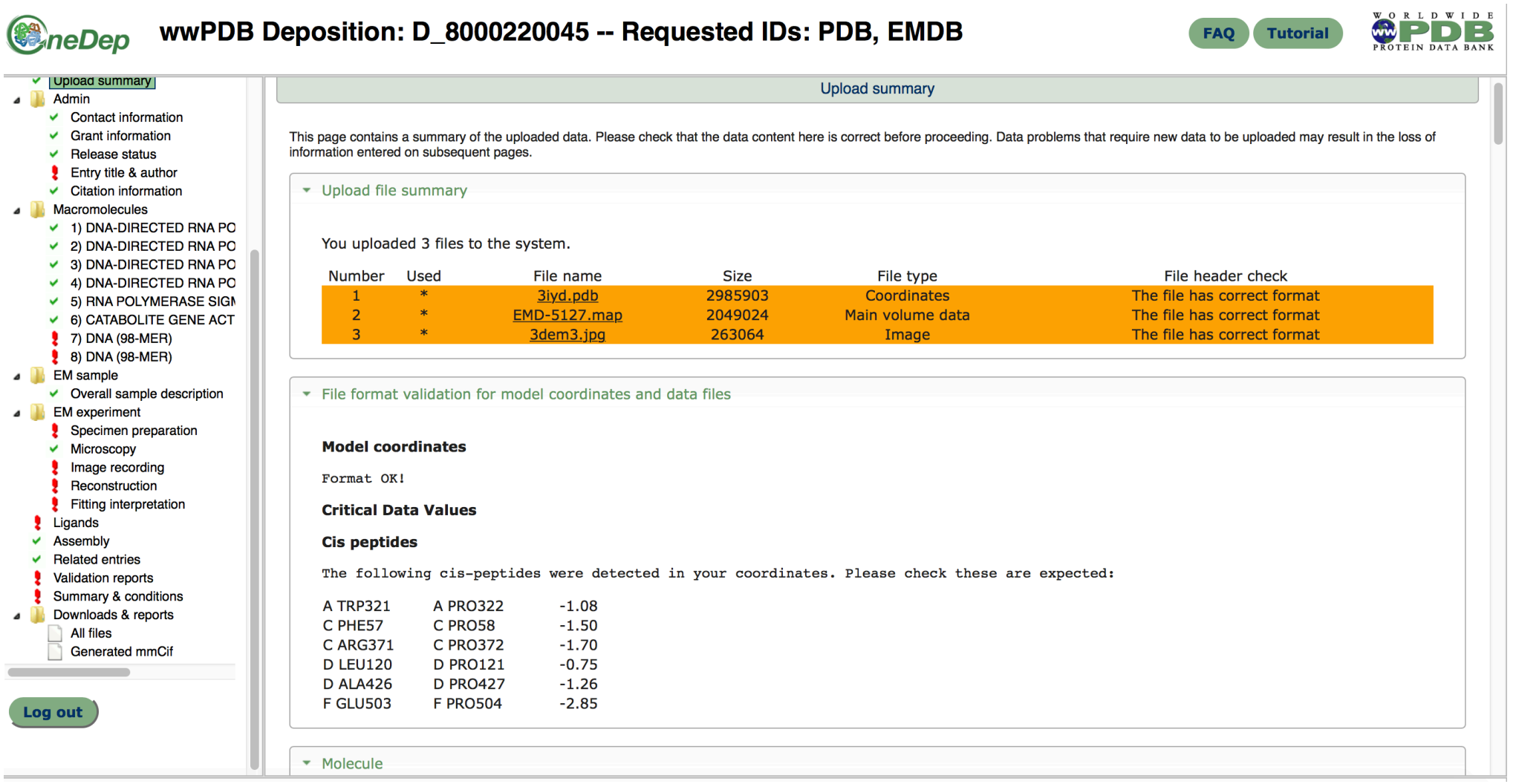
Growth of 3DEM Maps in EMDB



As of Oct 1, 2017, >5100 3DEM Maps in EMDB archive

688 new entries released January 1 – September 1, 2017

3DEM Data Deposition with OneDep



3DEM Deposition Metrics *via* OneDep

Structures	Total Depositions	Processed by		
		RCSB PDB	PDBe	PDBj
2016	513	214	226	73
2017	*389	189	135	65

Maps	Total Depositions	Processed by		
		RCSB PDB	PDBe/ EMDB	PDBj
2016	1097	477	507	113
2017	*633	307	273	53

*As of September 7,2017

Status of 3DEM Validation

- OneDep manages deposition, biocuration, and limited validation (structures) of all incoming 3DEM data
- EMDB performs limited validation (maps, fit of structures to maps) after data transfer from OneDep
- Wrap Up Workshop for current round of EMDataBank Map and Model Challenges was held in early Oct 2017
- Validation methods development funding sources
  - NIGMS – EMDataBank (Stanford/RCSB)
  - WT – broadly collaborative UK effort (incl. EMDB)
  - EU – INSTRUCT collaboration

Engaging the EM Community I

- Pressing need to formalize coordination of deposition, biocuration, and validation efforts between wwPDB and EMDB

- 2017 PDBx/mmCIF Working Group Meeting Day 2
  - Discussion on EM model deposition and validation requirements
  - Participants included – Paul Adams, Maya Topf, Wah Chiu, Corey Hryc, Garib Murshudov, Carsten Sachse, Martyn Winn, Cathy Lawson, John Westbrook, Stephen Burley, Ardan Patwardhan, Sameer Velankar, Gerard Kleywegt

Engaging the EM Community II

- 2017 PDBx/mmCIF Working Group made progress with software developers on 3DEM-related data matters
- The wwPDB PDBx/mmCIF Working Group will include a 3DEM subcommittee going forward
  - Currently recruiting others involved in development of 3DEM model building software
- wwPDB will collaborate with EM stakeholders to reconvene the EM Validation Task Force to review present status and determine consensus path forward

NMR Spectroscopy

John L. Markley

Agenda

- Archiving NMR Data in PDB/BMRB
- NMR Data Deposition Metrics
- BMRB Activities Related to OneDep
- NMR/SAS Hybrid Method Progress

Archiving NMR Data in PDB/BMRB I

- Biomolecular NMR Data Archiving
  - NMR model coordinates and related data are archived in PDB and backed up in BMRB
  - NMR data pertaining to structures, but not archived by PDB, are solicited by BMRB
  - NMR data not pertaining to structures are archived at BMRB: assignments without structure, information on dynamics, conformational changes, ligand binding, solvent accessibility, etc.

Archiving NMR Data in PDB/BMRB II

- Biocuration and validation services for NMR structures are provided through the OneDep system by RCSB PDB, PDBe, and PDBj
  - Current data content: coordinates, assigned chemical shifts, peak lists (optional), metadata, restraints in native format
  - Current validation: coordinates as for X-ray, chemical shift outliers (software provided by BMRB)
- Entries deposited via OneDep are sent to BMRB, where they are checked by BMRB software and biocurated as needed

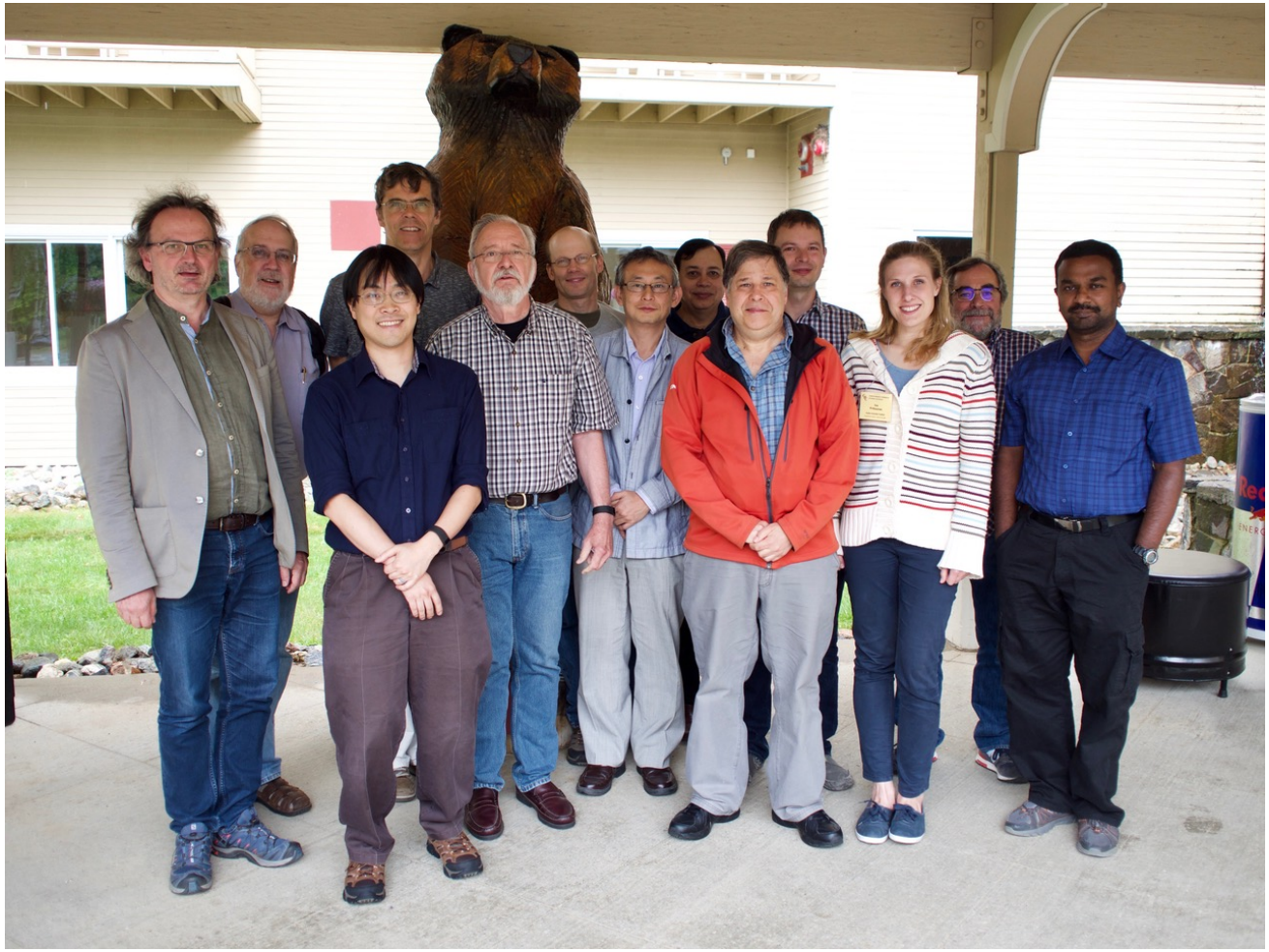
NMR Data Depositions (2014-2017)

Year	NMR Depositions to PDB Archive	Depositions of NMR data without Structures	Total BMRB Depositions
2014	515	240	755
2015	510	333	843
2016	473	276	749
2017 (to 9/1)	280	253	533

BMRB Activities Related to OneDep

- BMRB staff members participate in OneDep data development
- NMR Exchange Format (NEF)
  - NEF was developed as a means for exchanging data among software developers and for restraint deposition
  - NMR-STAR files containing restraints and peak lists are available
  - A NEF to NMR-STAR converter has been implemented by BMRB for incorporation into OneDep
- wwPDB now accepts NEF as an alternative deposition format for structural restraints
- BMRB has worked with NEF developers on NEF v1.0 and incorporated feeback into NMR-STAR
  - NEF restraint data must be accompanied by a NMR-STAR v3.1 or later chemical shift assignment file and mmCIF coordinate file with NEF atom nomenclature mapping

NEF-BMRB Meeting 06/16/17 Sunday River, Maine



- Michael Nilges
- David Case
- Vincent Chen
- Geerten Vuister
- John Markley
- Charles Schwieters
- Naohiro Kobayashi
- Pedro Romero
- Guy Montelione
- Aleks Gutmanas
- Iva Pritisanac
- Roberto Tejero
- Kumaran Baskaran
- (Jill Trewhella)
- (Andy Byrd)

Discussed: NEF progress and next steps (6-month, 1-year); validation as part of OneDep; SAS-NMR validation; CryoEM-NMR validation.

NMR/SAS Hybrid Methods Progress I

- Kumaran Baskaran (BMRB-Wisconsin) participates in the Integrative/Hybrid Methods (I/HM) Model Validation Developer Group
- Biomolecular NMR is itself already a “hybrid method” (i.e., structures are derived from multiple experiments)
- NMR-STAR dictionary accommodates data from a number of experimental methods complementing NMR
  - SAXS
  - H/D exchange
  - FRET
  - MS

NMR/SAS Hybrid Methods Progress II

- SASCIF extensions have been added to PDBx/mmCIF
- SAS data deposition now supported by OneDep *via* API calls to SASBDB Deposition User Interface
- SAS data retrieval from SASBDB currently limited SASBDB accession code (to be expanded)
- Validation of NMR/SAS Hybrid structures pending recommendations from NMR and SAS VTFs

Looking Ahead

Sameer Velankar

WORLDWIDE PDB
PROTEIN DATA BANK

wwpdb.org

Plans for the Coming Years I

2017/2018 (OneDep Team)

- Implement Ligand Validation Workshop recommendations (2D depictions and ED maps)
- Enable Data Depositor initiated coordinate update post release (versioned file)
- Collect new data related to XFEL/SFX techniques
- Automate testing of updated OneDep system
- Increase biocuration efficiency
- Implement NMR restraint validation
- Implement EM map validation

Plans for the Coming Years II

2017/2018

- Remediation work continuing
  - Carbohydrates (Lead: RCSB PDB)
  - Post-translational modifications (Lead: PDBe)
- wwPDB Partnership
  - PDBx/mmCIF dictionary management
  - Weekly release process auditing and automation
  - Invitation to EMDB (subject AC concurrence)
  - Negotiations with new Regional wwPDB Partners
- wwPDB AC meeting at PDBe EMBL-EBI

Plans for the Coming Years III

2018/2019

- wwPDB AC meeting at PDBj Osaka University
  - Friday October 18 2019 *versus*
  - Friday October 25 2019

- Planning for PDB 50th Anniversary in 2021 with celebratory scientific meetings and outreach events by the wwPDB and individual wwPDB partners

PDBe to Host 2018 wwPDB AC

- Date: Friday November 2nd 2018

- Location:

  Madingley Hall Cambridge Conference Center

  University of Cambridge, Madingley

  Cambridge CB23 8AQ

  United Kingdom

Lunch and Executive Session

Questions for the wwPDB AC

Stephen K. Burley

WORLDWIDE PDB
PROTEIN DATA BANK

wwpdb.org

Questions for the wwPDB AC

1. Does the wwPDB AC concur with recommendation from wwPDB Partners re DOI resolution of PDB Entries as outlined in Appendix 1?

2. Does the wwPDB AC concur with recommendation from wwPDB Partners re wwPDB Chair/Co-Chair restructuring as outlined in Appendix 2?

3. Does the wwPDB AC concur with recommendation from wwPDB Partners re inviting EMDB to join the wwPDB organization as outlined in Appendix 3?

Questions for the wwPDB AC (cont.)

4. Does the wwPDB AC concur with recommendation from wwPDB Partners re the process for adding new Regional wwPDB Partners as outlined in Appendix 4?

5. Does the wwPDB AC concur with recommendation from wwPDB Partners re providing Mogul geometry standards for ligands as outlined in Appendix 5?

6. Does the wwPDB AC have any questions or concerns regarding the individual RCSB PDB, PDBe, PDBj, or BMRB Advisory Committee reports provided in Appendix 6?

Acknowledgements and Closing Remarks

Stephen K. Burley and R. Andrew Byrd

WORLDWIDE PDB
PROTEIN DATA BANK

wwpdb.org

Slide Preparation and Organization

- Jasmine Young
- Christine Zardecki
- Nicole Oorbeek (nee Malkiewicz)
- RCSB Team Members
- IQB Support Staff
- Pauline Haslam
- Tomoko Shimizu

Meeting Funding

- NSF/NIH/DoE grant to RCSB PDB
- Institute for Quantitative Biomedicine and Rutgers, The State University of New Jersey
- wwPDB Foundation
- NIGMS grant to BMRB
- EMBL-EBI and Wellcome Trust grant to PDBe
- IPR and JST and AMED grants to PDBj

Closing Remarks

Thank you for your enduring support of the Worldwide Protein Data Bank partnership and for taking the time to attend the annual meeting

Safe travels home