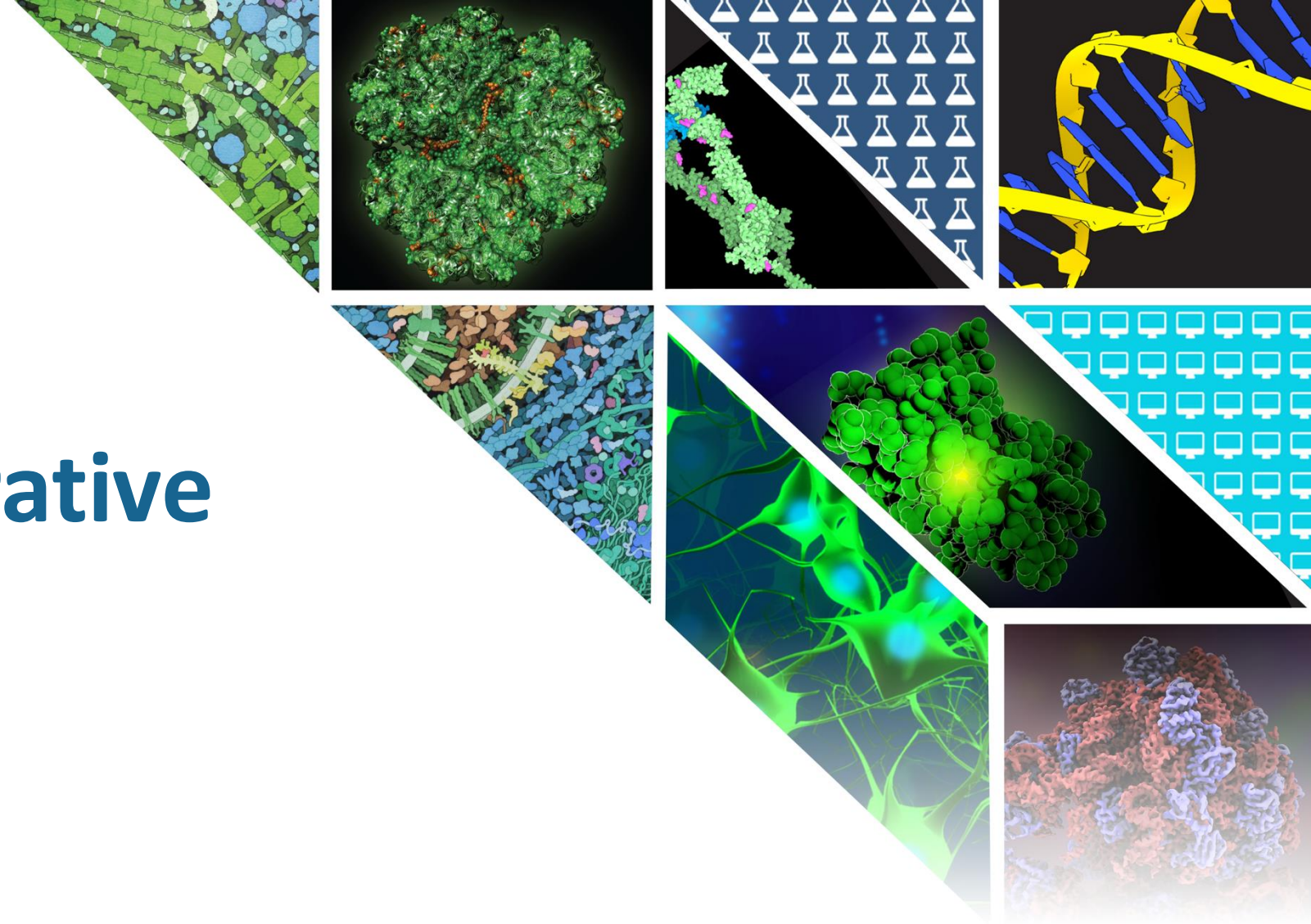
RCSB PDB
PROTEIN DATA BANK

rcsb.org



Exploring Integrative Structures with RCSB.org

Brinda Vallat

RCSB Protein Data Bank

Institute for Quantitative Biomedicine

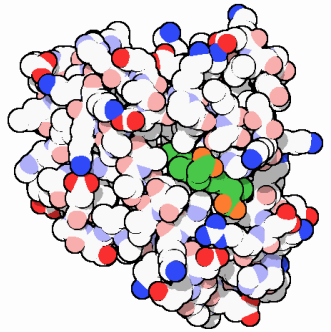
Rutgers, The State University of New Jersey, USA

RUTGERS

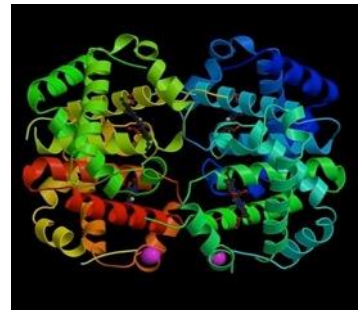
The Protein Data Bank

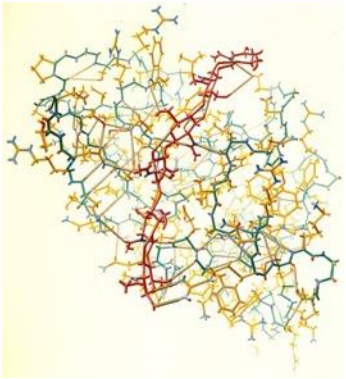
- Single global repository for experimentally determined structures of biological macromolecules and their complexes
- Established in 1971 with seven structures, currently holds ~250,000 structures
- First open access digital resource in biological sciences
- FAIR resource: *Findable, Accessible, Interoperable, and Reusable* data
- wwPDB collaboration: RCSB PDB (US), PDBe (Europe), PDBj (Japan), PDBc (China), BMRB, and EMDB

Myoglobin



Hemoglobin





Lysozyme

Ribonuclease

Structures that Inspired Launch of the PDB

WORLDWIDE PDB PROTEIN DATA BANK

wwpdb.org

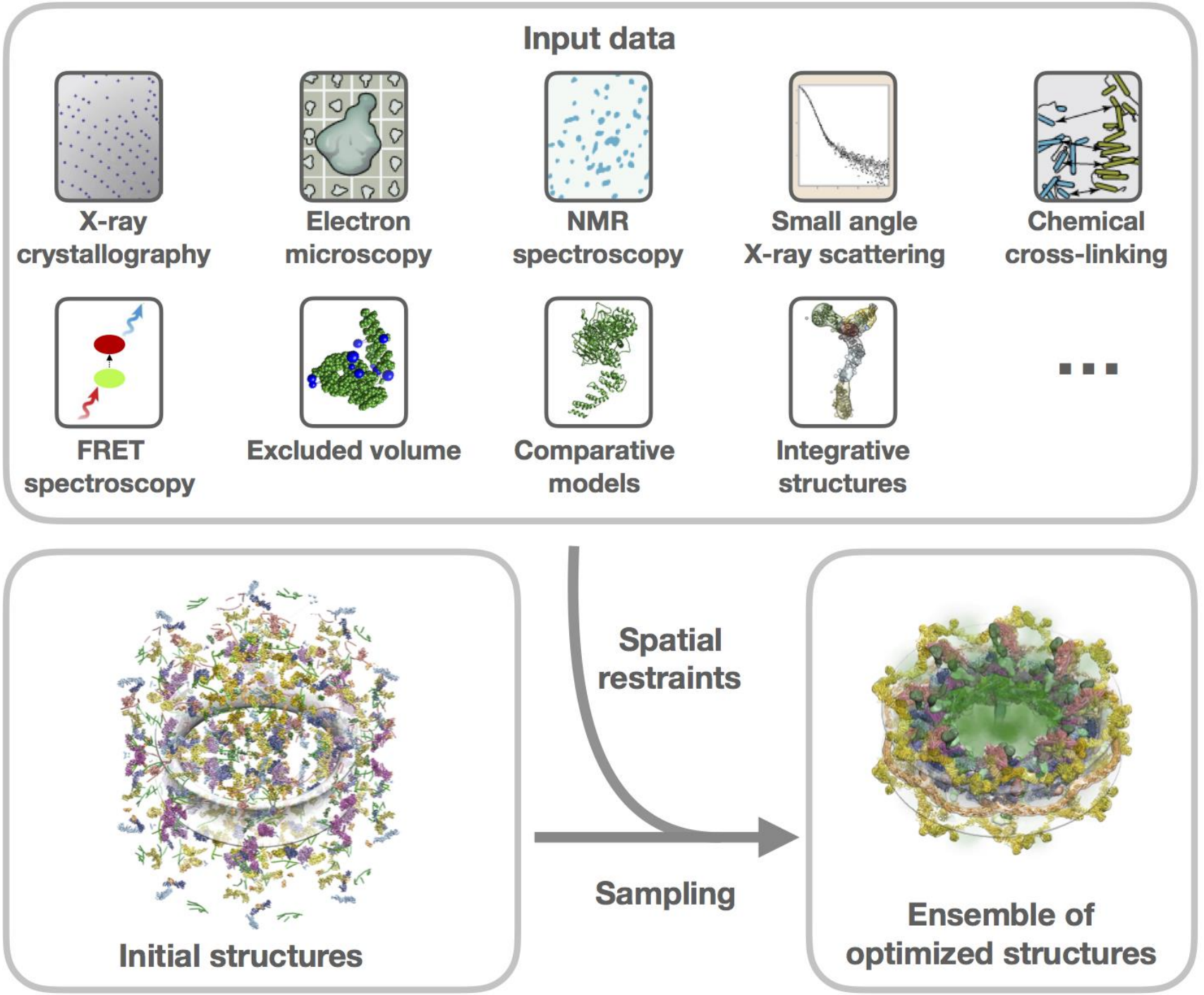
wwPDB consortium. 2019. Nucleic Acids Res. 47, D520-D528

PDB Archives Structures of Macromolecules

- Structures from X-ray Crystallography, Nuclear Magnetic Resonance (NMR) Spectroscopy, and Three-Dimensional Electron Microscopy (3DEM)
- Atomic coordinates and metadata
- Multi-method structures based on X-ray, NMR, and 3DEM
  - X-ray + Neutron Diffraction
  - X-ray + Solution NMR
  - X-ray + Solid-state NMR
  - Solution NMR + Small angle scattering
  - 3DEM + Solid-state NMR
  - 3DEM + Small angle scattering

Vallat B et al. 2019. J Biomol NMR. 73, 385-398

Integrative Structure Determination



Vallat B et al. 2018. Structure. 26, 894-904

Rout MP & Sali A. 2019. Cell. 177, 1384-1403

Challenges in Archiving Integrative Structures

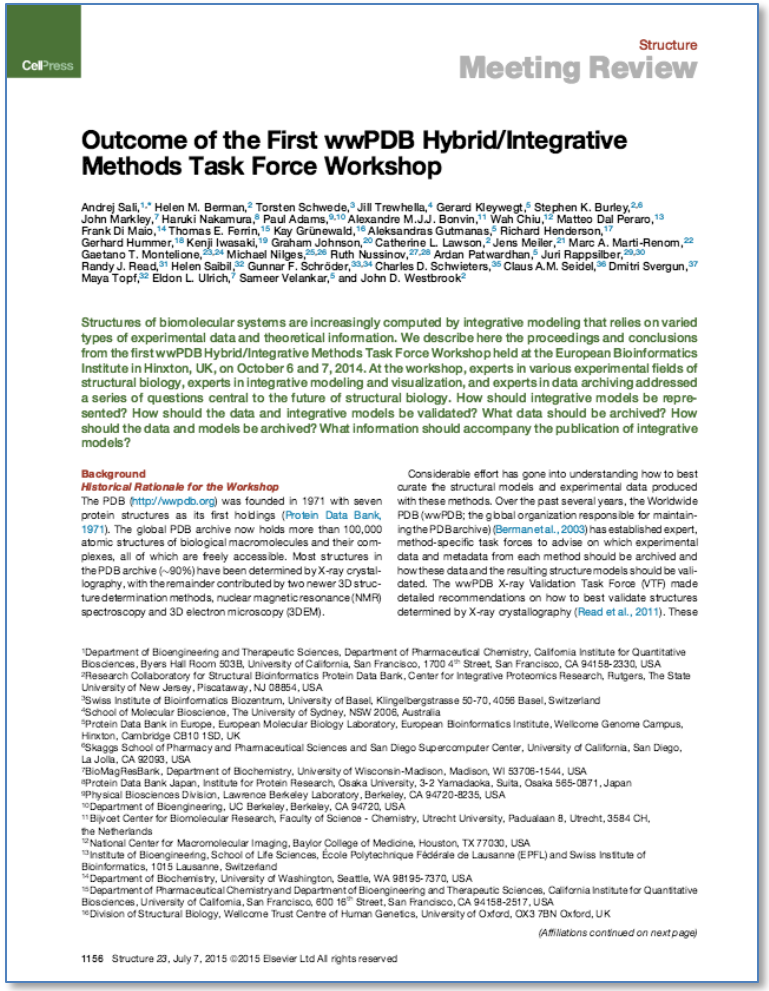
- Restraints from different experiments (Crosslinking, Hydrogen-Deuterium Exchange (HDX), Mass Spectrometry (MS), Small Angle Scattering (SAS), Forster Resonance Energy Transfer (FRET), etc.)
- Multi-scale, multi-state, and ordered models
- Conformational ensembles and models with heterogeneous composition
- Starting models from experimental or computational methods
- Different modeling methodologies and software applications

wwPDB Integrative/Hybrid Methods Task Force

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

Summary of Recommendations

- Archive structural models, data, metadata, and workflows
- Adopt flexible model representation
- Assess structural model uncertainty
- Federate archives for structural models and experimental data
- Establish publication standards



CellPress

Structure

Meeting Review

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

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

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

Background

Historical Rationale for the Workshop

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

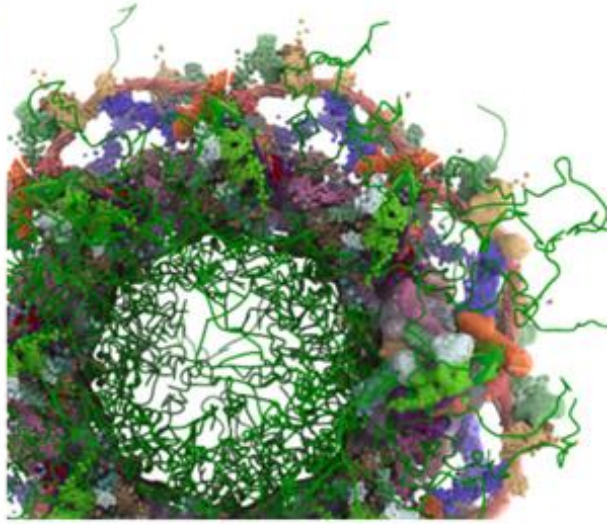
Considerable effort has gone into understanding how to best curate the structural models and experimental data produced with these methods. Over the past several years, the Worldwide PDB (wwPDB; the global organization responsible for maintaining the PDB archive) (Berman et al., 2003) has established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

[1]Department of Bioengineering and Therapeutic Sciences, Department of Pharmaceutical Chemistry, California Institute for Quantitative Biosciences, Byers Hall Room 503B, University of California, San Francisco, 1700 4th Street, San Francisco, CA 94158-2330, USA
[2]Research Collaboratory for Structural Bioinformatics Protein Data Bank, Center for Integrative Proteomics Research, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA
[3]Swiss Institute of Bioinformatics Biozentrum, University of Basel, Klingelbergstrasse 50-70, 4056 Basel, Switzerland
[4]School of Molecular Bioscience, The University of Sydney, NSW 2006, Australia
[5]Protein Data Bank in Europe, European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SD, UK
[6]Skaggs School of Pharmacy and Pharmaceutical Sciences and San Diego Supercomputer Center, University of California, San Diego, La Jolla, CA 92093, USA
[7]BioMagResBank, Department of Biochemistry, University of Wisconsin-Madison, Madison, WI 53706-1544, USA
[8]Protein Data Bank Japan, Institute for Protein Research, Osaka University, 3-2 Yamadaoka, Suita, Osaka 565-0871, Japan
[9]Physical Biosciences Division, Lawrence Berkeley Laboratory, Berkeley, CA 94720-8235, USA
[10]Department of Bioengineering, UC Berkeley, Berkeley, CA 94720, USA
[11]Bijvoet Center for Biomolecular Research, Faculty of Science - Chemistry, Utrecht University, Padualaan 8, Utrecht, 3584 CH, the Netherlands
[12]National Center for Macromolecular Imaging, Baylor College of Medicine, Houston, TX 77030, USA
[13]Institute of Bioengineering, School of Life Sciences, École Polytechnique Fédérale de Lausanne (EPFL) and Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland
[14]Department of Biochemistry, University of Washington, Seattle, WA 98195-7370, USA
[15]Department of Pharmaceutical Chemistry and Department of Bioengineering and Therapeutic Sciences, California Institute for Quantitative Biosciences, University of California, San Francisco, 600 16th Street, San Francisco, CA 94158-2517, USA
[16]Division of Structural Biology, Wellcome Trust Centre of Human Genetics, University of Oxford, OX3 7BN Oxford, UK

(Affiliations continued on next page)

Sali A et al. 2015. Structure. 23, 1156-1167

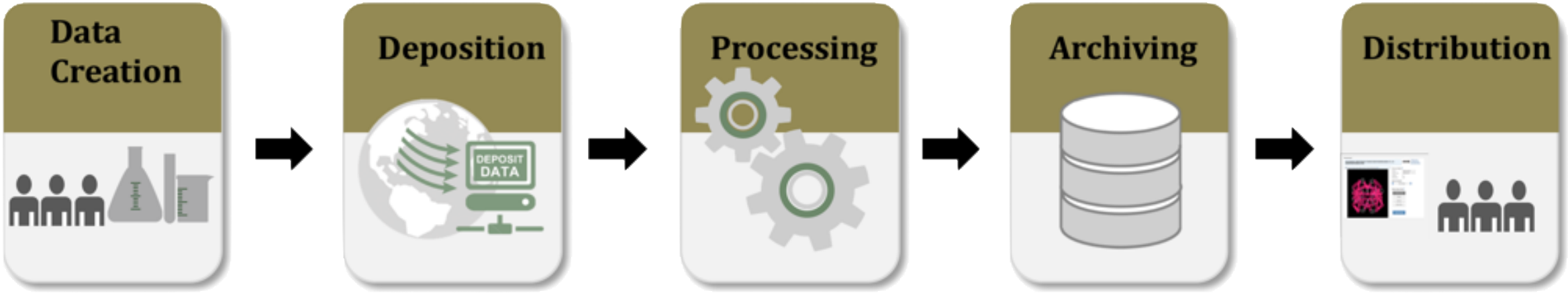
http://www.wwpdb.org/task/hybrid



Archiving Integrative Structures

Archiving Requirements

- Data standard
  - Primary requirement that enables automated data processing and dissemination of data in a standard form
- Software tools that support the data standards
- Data pipeline for deposition, processing, archiving, distribution



Data Standards
Deposition Tools

Biocuration
Validation

Vallat B et al. 2021. Acta D. 77, 1486-1496
Berman et al. 2018. Q Rev Biophys. 51, e8
Vallat B et al. 2018. Structure. 26, 894-904

Data Standard: PDBx/mmCIF

- Data standard for the PDB used for archiving experimental structures of macromolecules
- Standard definitions for
  - Chemistry of polymeric macromolecules and small molecules
  - Atomic model coordinates
  - Supporting metadata: source organisms, samples, authors, citations, software
  - Reference sequence and chemical information
- Extensible to support evolving methods
- Software tools to read, write, and check compliance
- http://mmcif.wwpdb.org

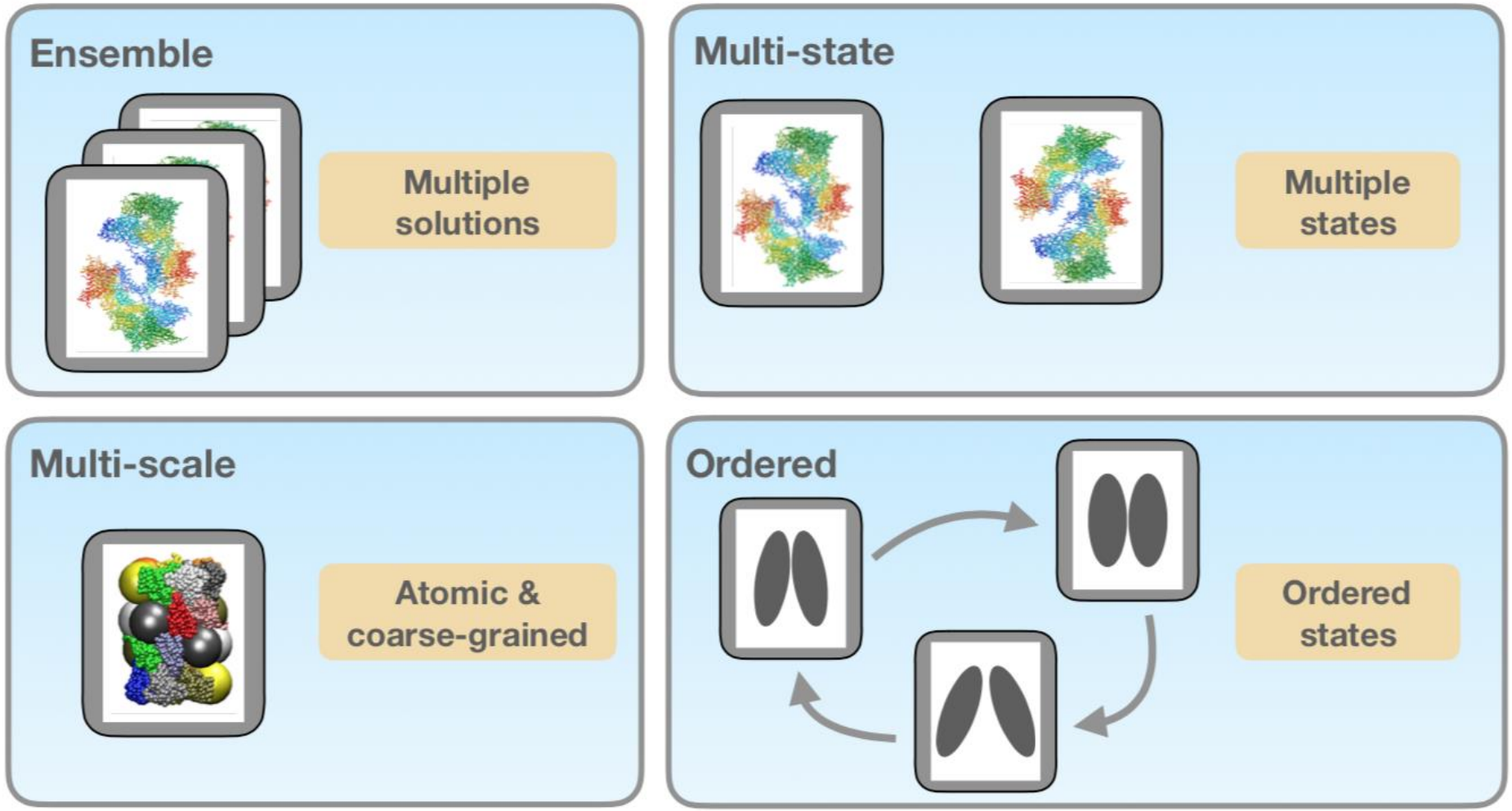
Westbrook et al. 2022. J. Mol. Biol. 434, 167599
Westbrook J et al. 2005. International tables for crystallography. 195-198
Westbrook JD & Fitzgerald PMD. 2009. Structural bioinformatics. 71-291

Data Standard: IHMCIF

- IHMCIF: Extension of PDBx/mmCIF
  - Data standard for integrative structures
  - Existing definitions from PDBx/mmCIF are used
  - New definitions required for integrative modeling are in IHMCIF
    - Multi-scale, multi-state, ordered states, ensembles
    - Compositionally heterogeneous assemblies
    - Experimental restraints, starting models, modeling workflows
    - Information regarding related data residing in external resources
  - https://github.com/ihmwg/IHMCIF
- Python-IHM: Software tool that supports IHMCIF
  - Provides a mechanism to programmatically generate mmCIF files compliant with IHMCIF
  - https://github.com/ihmwg/python-ihm

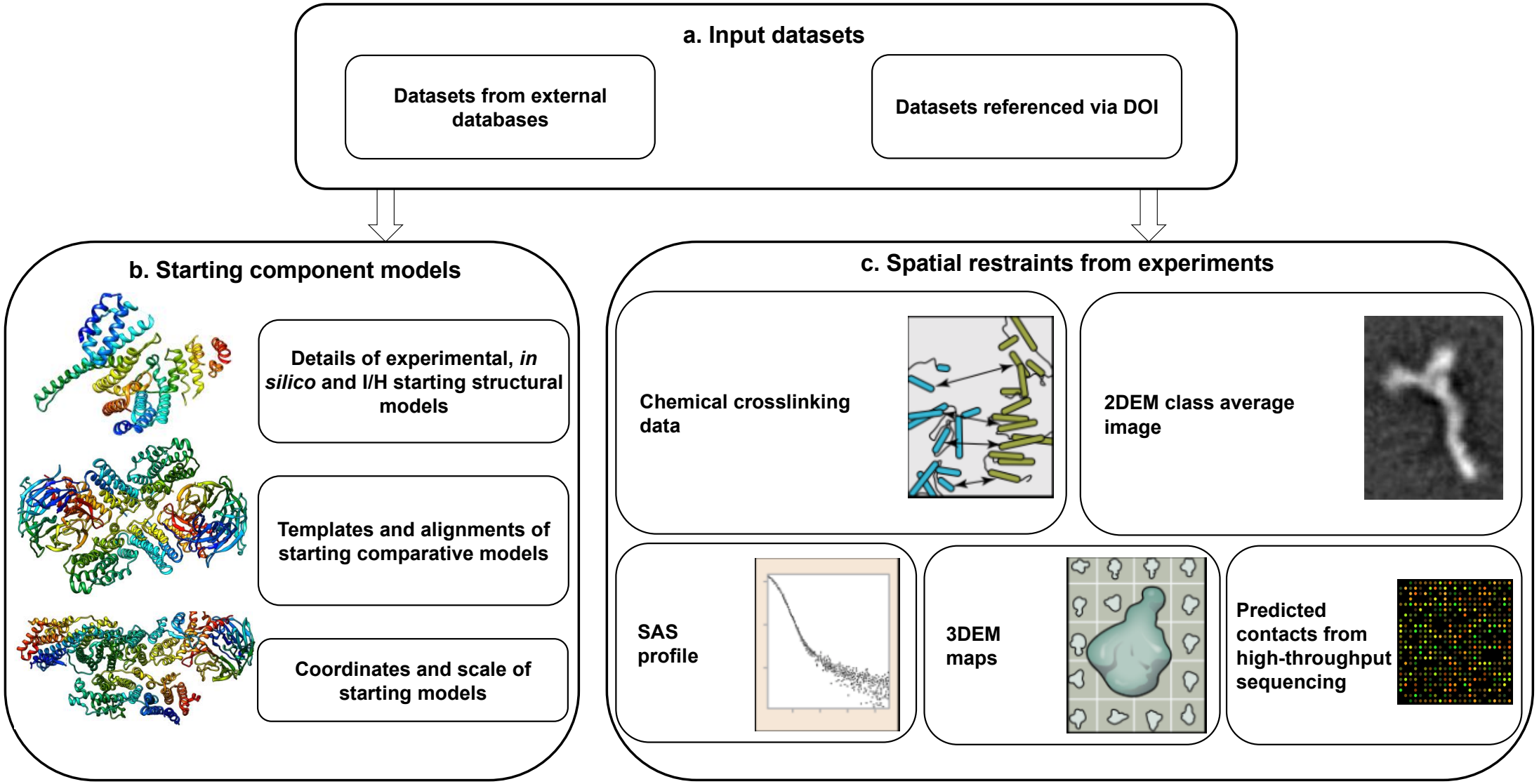
Vallat B et al. 2024. J. Mol. Biol. 436, 168546
Vallat B et al. 2018. Structure. 26, 894-904
Vallat B et al. 2019. J Biomol NMR. 73, 385-398

Multi-scale, Multi-state, Ordered-state, Ensembles

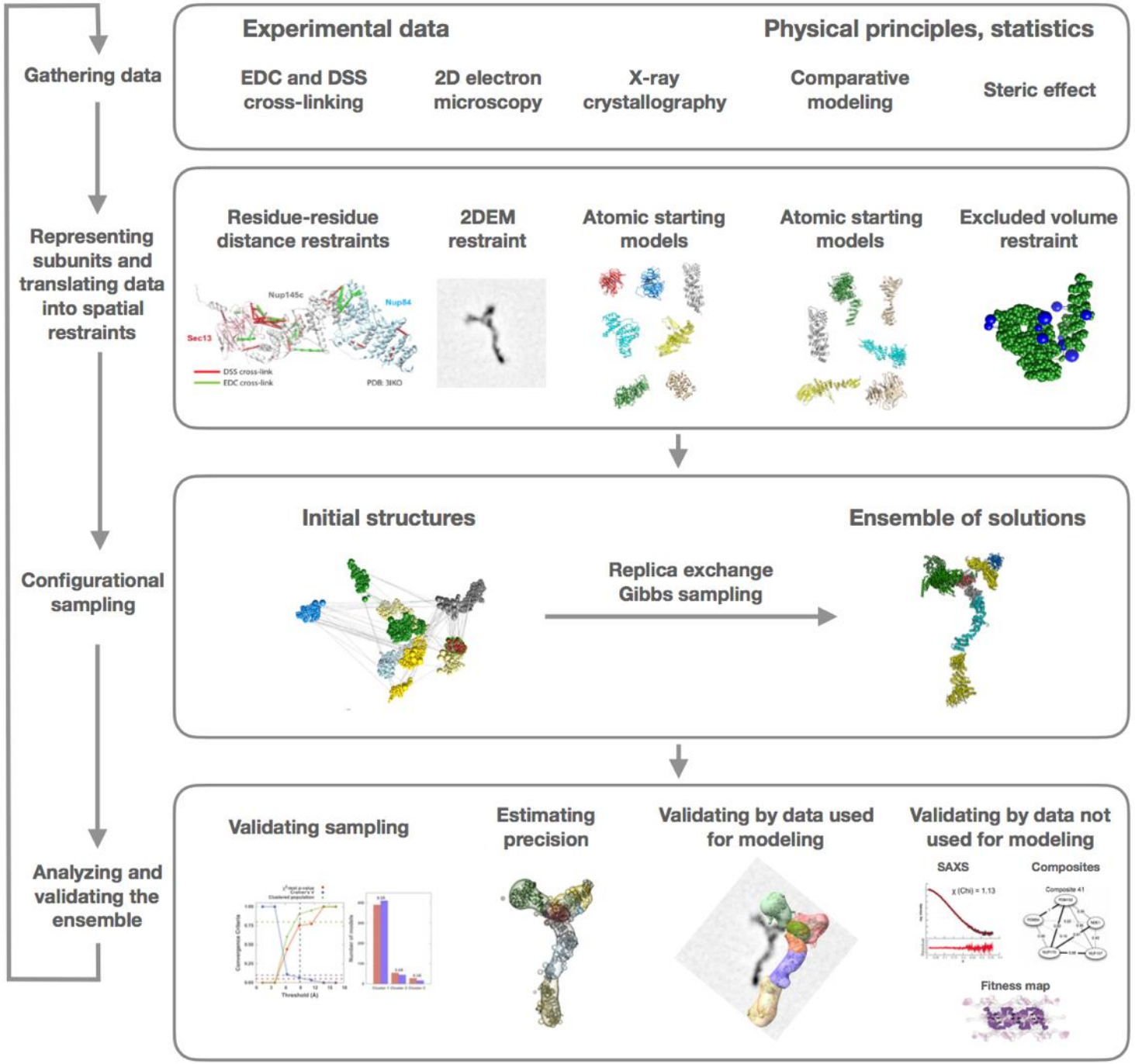


Vallat B et al. 2025. J. Mol. Biol. 437, 168963
Vallat B et al. 2024. J. Mol. Biol. 436, 168546

Spatial Restraints and Starting Models



Vallat B et al. 2025. J. Mol. Biol. 437, 168963; Vallat B et al. 2024. J. Mol. Biol. 436, 168546

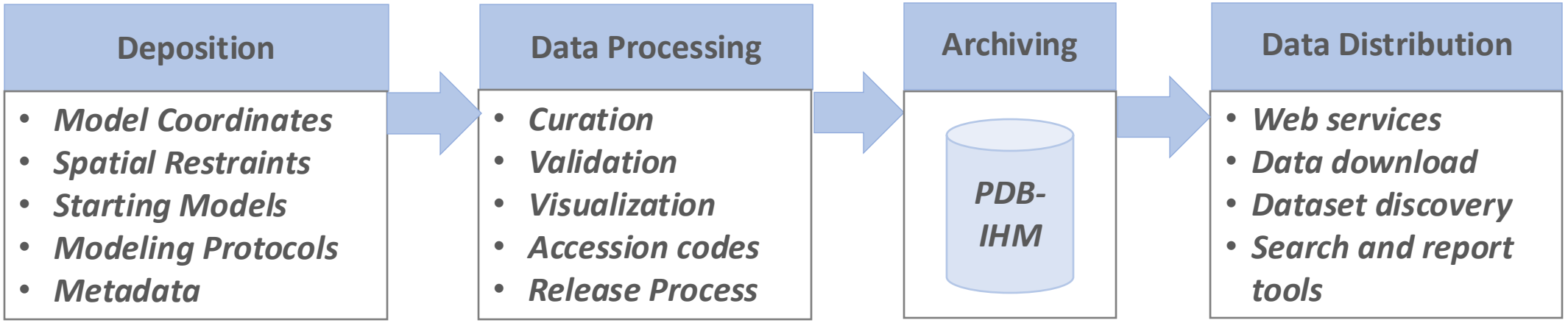


Modeling Workflow

Vallat B et al. 2025. J. Mol. Biol. 437, 168963
Vallat B et al. 2024. J. Mol. Biol. 436, 168546

PDB-IHM System for Archiving Integrative Structures

A system for deposition, processing, archiving, and distribution of integrative structures available at https://data.pdb-ihm.org/



- The data standards for PDB-IHM are provided by IHMCIF
- Data deposition and processing system based on DERIVA scientific asset management platform (http://isrd.isi.edu/deriva)

Vallat B et al. 2025. J. Mol. Biol. 437, 168963
Vallat B et al. 2021. Acta D. 77, 1486-1496
Burley SK et al. 2017. Structure. 25, 1317-1318
Schuler RE et al. 2016. IEEE eScience, 31-40

Deposition Guidelines

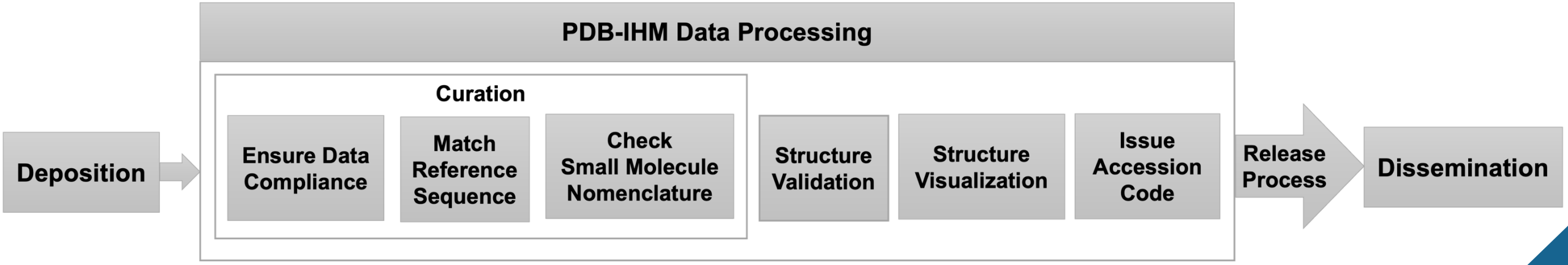
- IHMCIF compliant files
  - Molecular system representation and model coordinates
  - Spatial restraints, starting models, modeling protocols
  - External references: Sequences, chemical moieties, related data
  - Metadata: Authors, citations, software
- Compliant files can be created prior to deposition (e.g., using python-ihm) or using the data harvesting section of the deposition workflow
- Deposition API supports submission of a *collection* of structures belonging to a single investigation or publication

https://data.pdb-ihm.org

helpdesk@pdb-ihm.org

Biocuration Pipeline

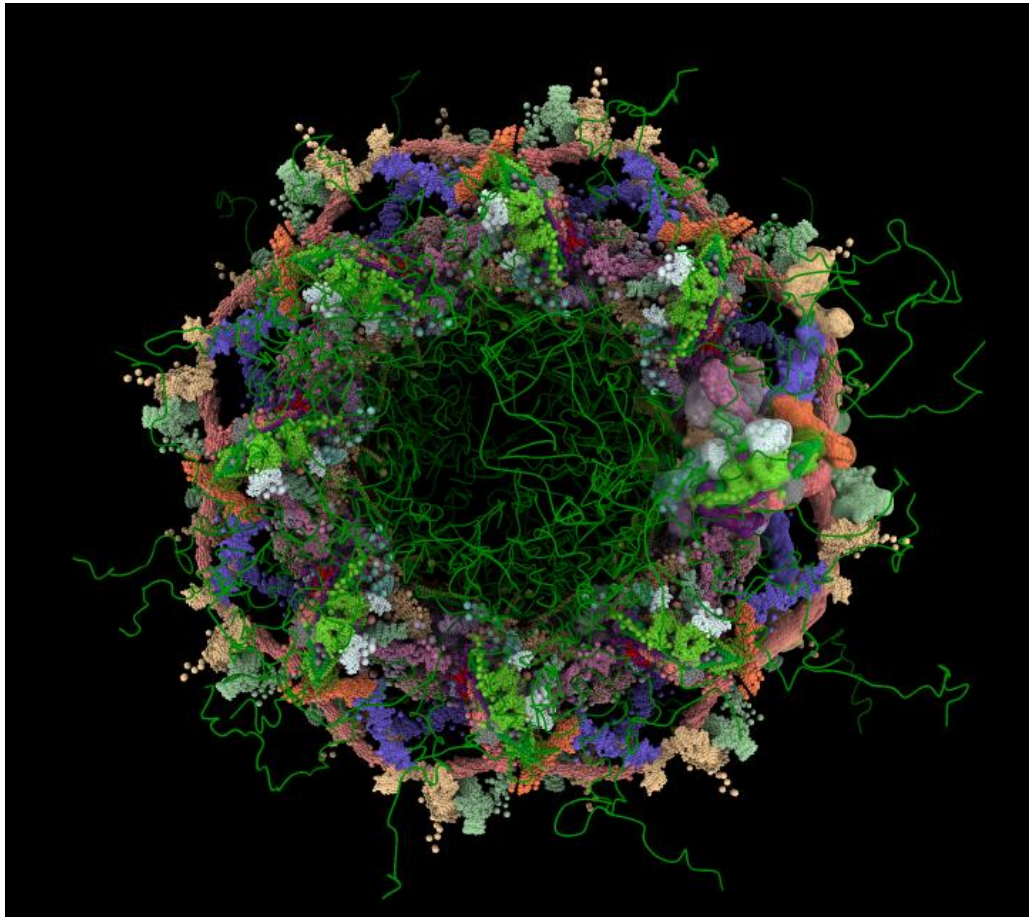
- Data consistency and completeness
- Compliance with PDBx/mmCIF and IHMCIF
- Reference sequence and small molecule information
- Model visualization
- Accession code issue
- Validation report generation



Vallat B et al. 2025. J. Mol. Biol. 437, 168963; Vallat B et al. 2021. Acta D. 77, 1486-1496

Visualization of Integrative Structures

8ZZC: Integrative structure of the nuclear pore complex from budding yeast



- Desktop application
  - ChimeraX
  - https://www.rbvi.ucsf.edu/chimerax/
  - Use “open filename.cif format ihm”
- Web application
  - Mol*
  - https://molstar.org

Meng EC et al. 2023. Protein Sci. 32(11), e4792
Sehnal D et al. 2021. Nucleic Acids Res. 49, W431-W437
Kim SJ et al. 2018. Nature. 555(7697), 475-482

Validation of Integrative Structures

- Based on guidelines and recommendations from the community
- Assessment categories
  - Model composition
  - Data quality assessments
  - Model quality assessments
  - Fit to data used to build the model
- Input Data types
  - Small angle solution scattering
  - Crosslinking-MS
  - 3DEM
- Software integrated with PDB-IHM deposition and curation pipeline
  - Validation reports can be obtained after deposition and used for manuscript review process
  - Summary and full pdf reports are available
- Standalone validation server
  - https://validate.pdb-ihm.org/

Integrative Structure Validation Report

May 26, 2022 - 08:45 AM PDT

The following software was used in the production of this report:

Molprobity Version 4.4
ATSAS Version 3.0.3
Integrative Modeling Validation Version 1.0
Integrative Modeling Platform Version 2.15.0

PDB ID	PDBDEV_00000009
Structure Name	Structure of the human Rev7 dimer
Publication Title	Rev7 dimerization is important for assembly and function of the Rev1/Polζ translesion synthesis complex
Authors	Alessandro A. Rizzo; Faye-Marie Vassel; Nimrat Chatterjee; Sanjay D'Souza; Yunfeng Li; Bing Hao; Michael T. Hemann; Graham C. Walker; Dmitry M. Korzhnev

This is a PDB-Dev IM Structure Validation Report for a publicly released PDB-Dev entry.

We welcome your comments at pdb-dev@mail.wwpdb.org

A user guide is available at https://pdb-dev-beta.wwpdb.org/validation_help.html with specific help available everywhere you see the ? symbol.

List of references used to build this report is available here.

Overall quality

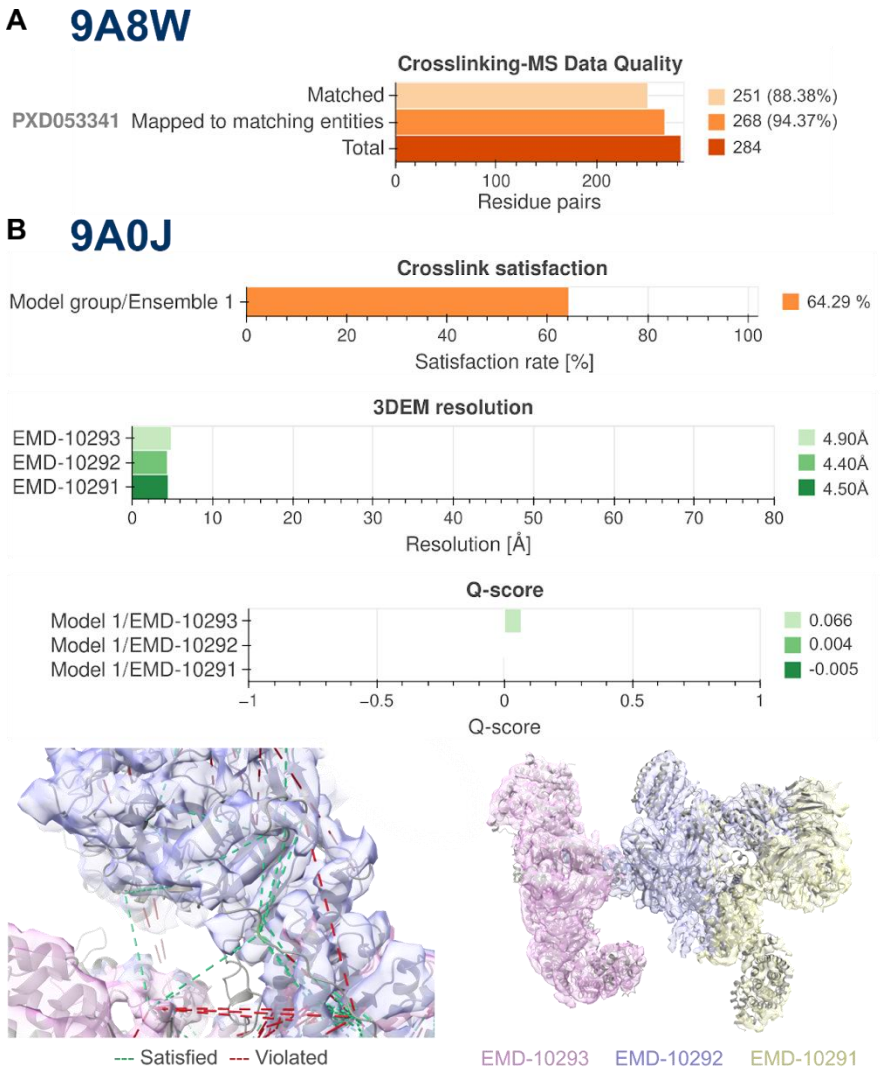
This validation report contains model quality assessments for all structures, data quality assessment for SAS datasets and fit to model assessments for SAS datasets. Data quality and fit to model assessments for other datasets and model uncertainty are under development.

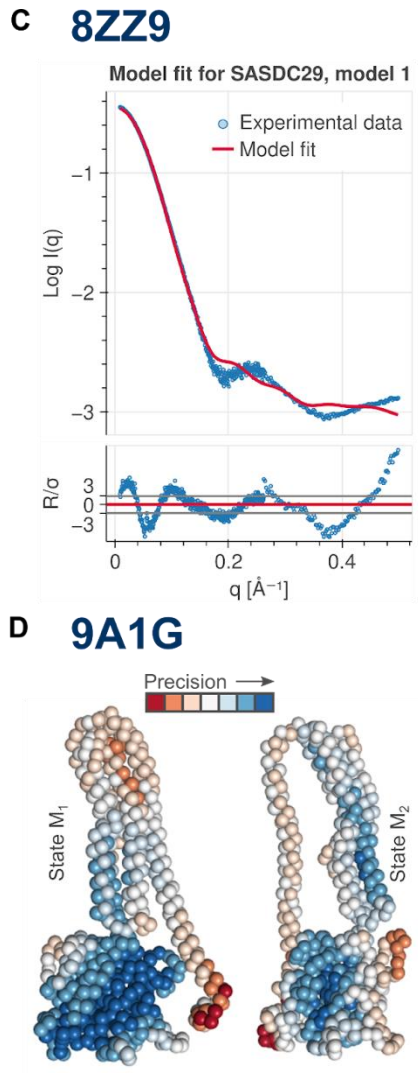
Model Quality: Molprobity Analysis

IM Structure Validation Report

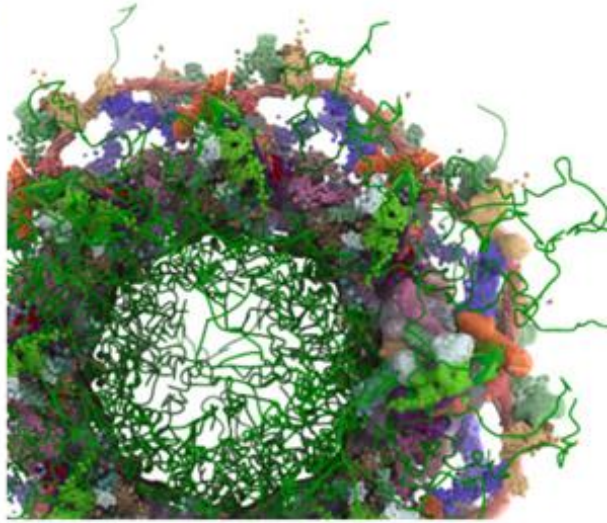
Berman HM et al. 2019. Structure. 27, 1745-1759
Trewhella J et al. 2017. Acta D. 73, 710-728

Validation Methods: Community Engagement



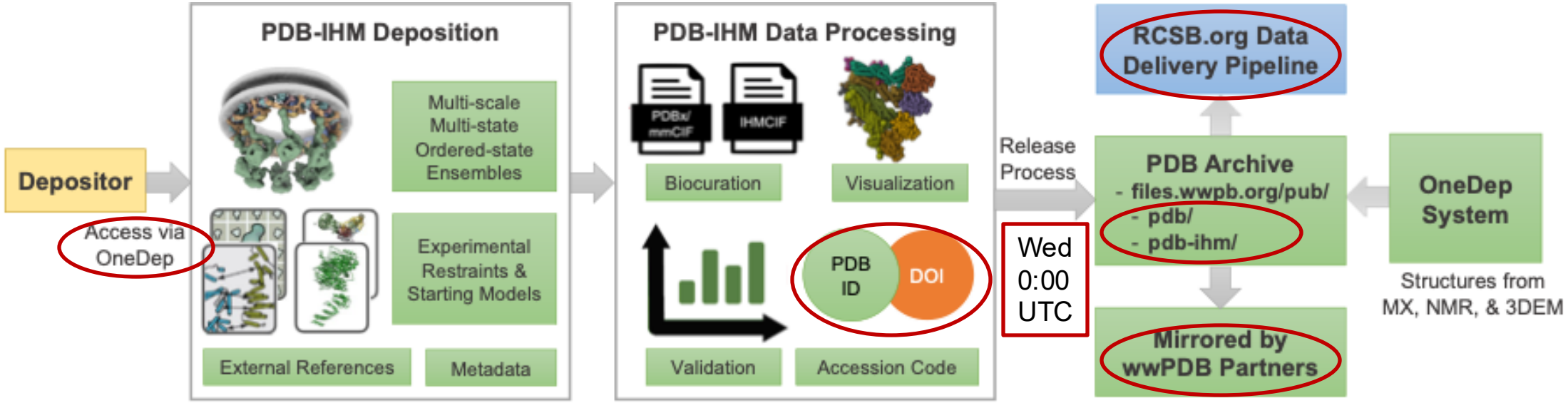


- Support for small angle scattering (SAS), crosslinking-MS, and 3DEM
- 3DEM validation is based on wwPDB standards
- Data exchange with SASBDB and PRIDE for accessing SAS and crosslinking-MS data
- PRIDE Crosslinking supports submission of complete crosslinking-MS datasets
- Collaborations with experts in SAS and crosslinking-MS communities



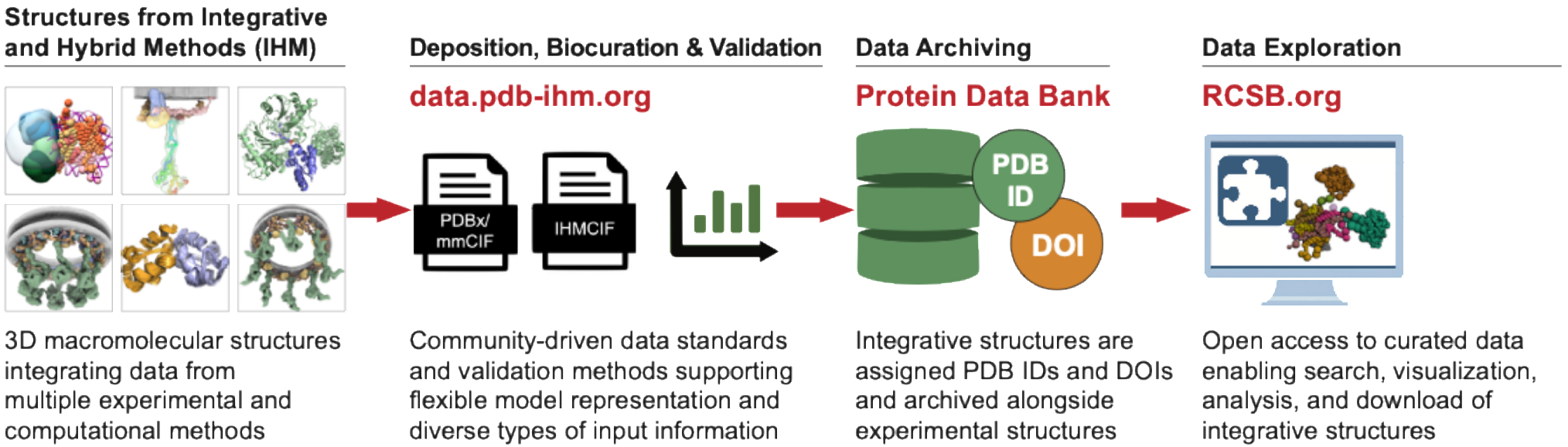
PDB-IHM Unification with the PDB Archive and Dissemination on RCSB.org

PDB-IHM: Unification with PDB Archive



- Integrative structures archived alongside experimental structures in the PDB archive
  - wwPDB partners mirror PDB-IHM
  - mmCIF files, validation reports, holdings files
  - ~100K file downloads
- New and existing integrative structures are issued PDB accession codes
- Synchronized weekly update process
- Registration of PDB DOIs and creation of wwPDB DOI landing pages
  - Online citations cross-linked to PDB entries *via* DOIs
- RCSB.org disseminates integrative structures

PDB-IHM: Integrated with RCSB.org



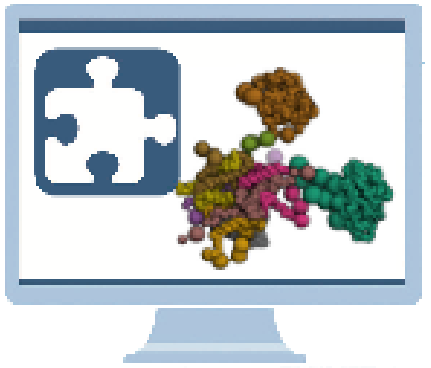
- RCSB.org now supports search, access, analysis, and visualization of integrative structures
- PDB-IHM integration enabled by IHMCIF extension of PDBx/mmCIF data standard

PDB-IHM at RCSB.org

- Primary and computed data for representative model
- API support across all services
- Attribute-based and scientific searches
- Integrated with Search Results and Structure Summary Pages
- Enabled 3D visualization and annotations, including sequence annotations in 3D
- Incorporated IHM-specific information (*e.g.*, multi-scaling, multi-state modeling, input experimental data and starting models used)

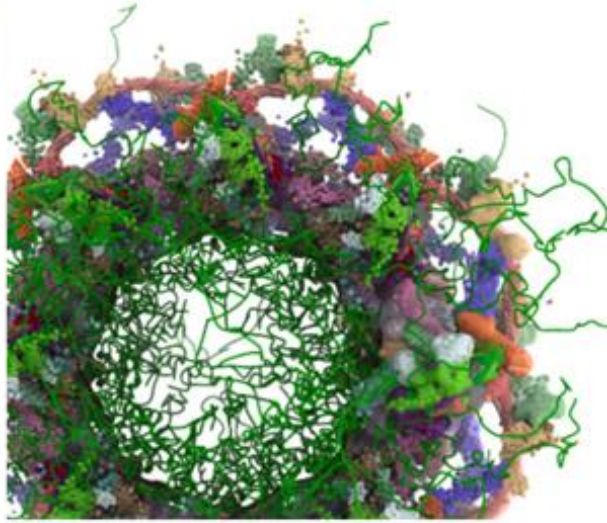
Data Exploration

RCSB.org

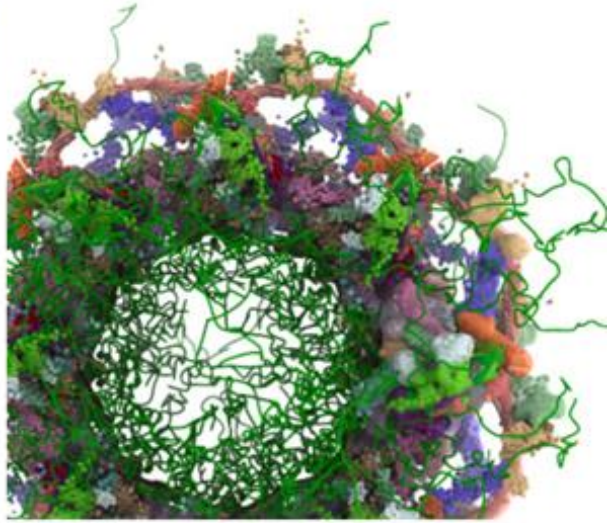


Open access to curated data enabling search, visualization, analysis, and download of integrative structures

https://www.rcsb.org/docs/general-help/integrative-structures-on-rcsborg
https://www.rcsb.org/news/682e14b39fcd08dd08b7de58

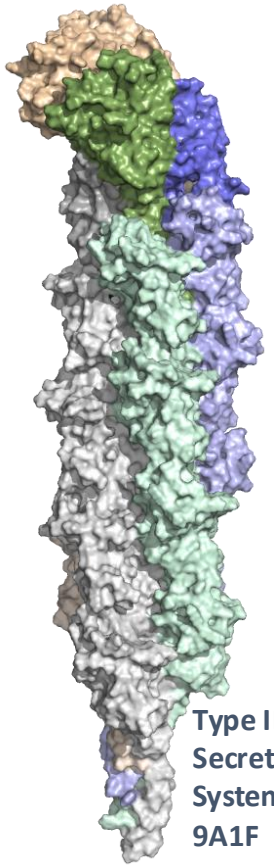


RCSB.org DEMO

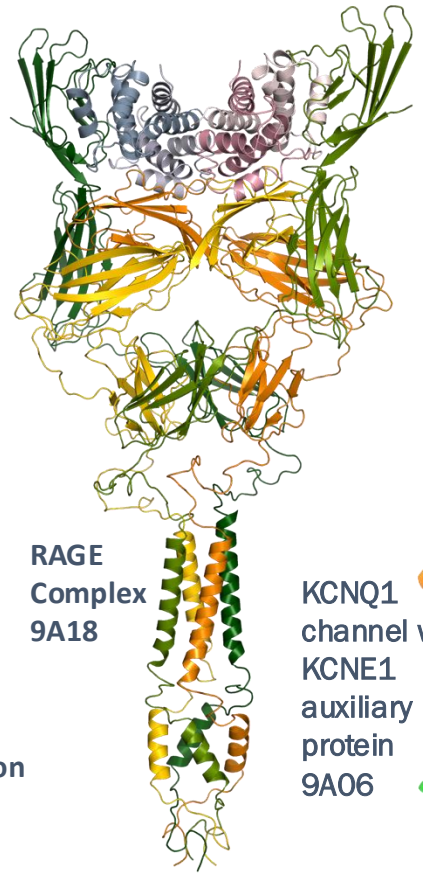


PDB-IHM: Current Status

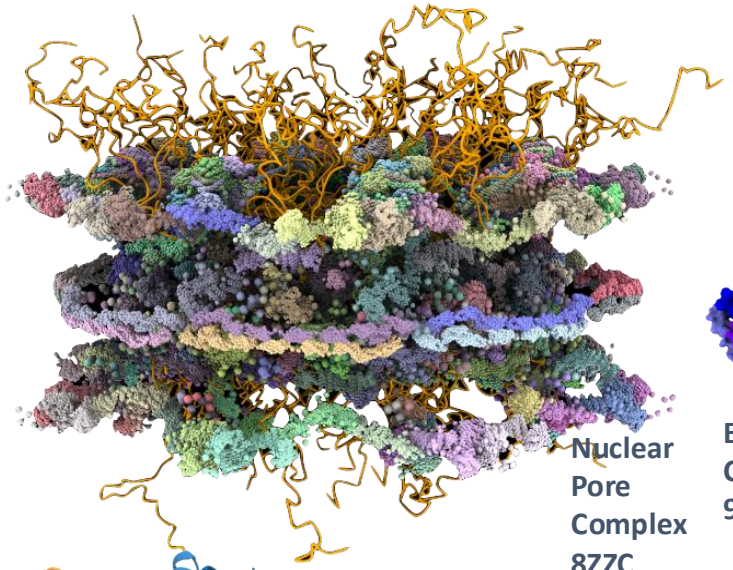
Structures in PDB-IHM



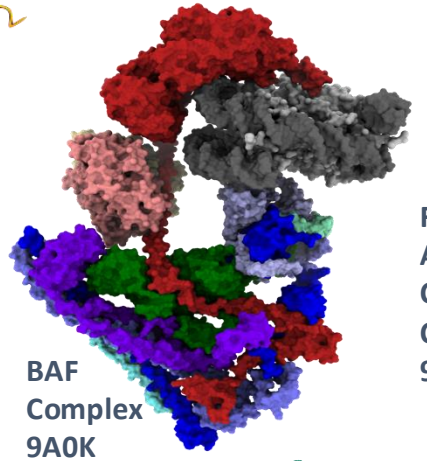
Type II Secretion System 9A1F



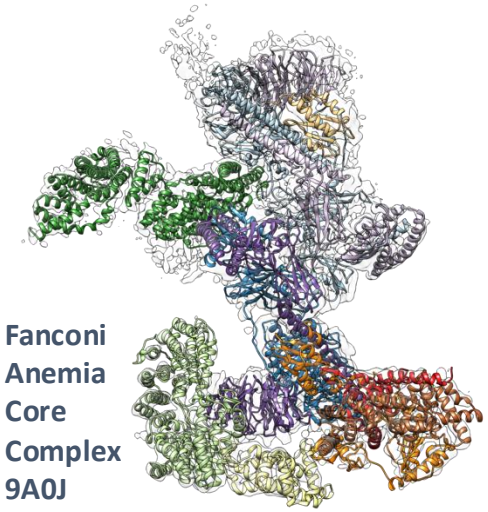
RAGE Complex 9A18



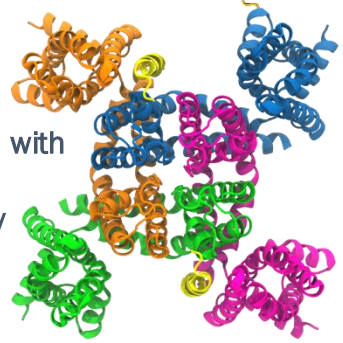
Nuclear Pore Complex 8ZZC



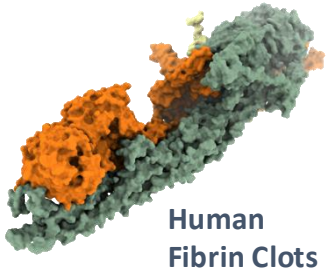
BAF Complex 9A0K



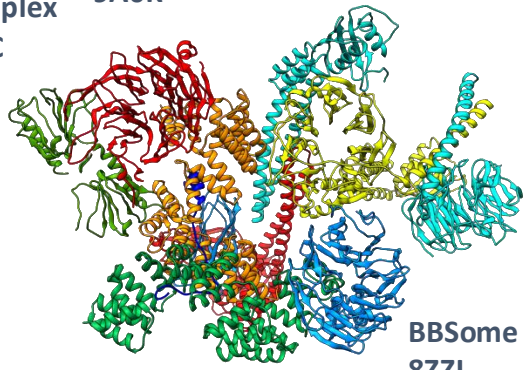
Fanconi Anemia Core Complex 9A0J



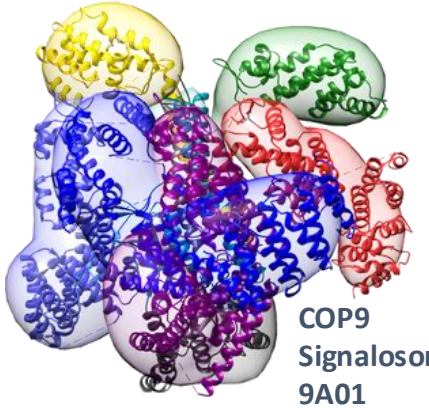
KCNQ1 channel with KCNE1 auxiliary protein 9A06



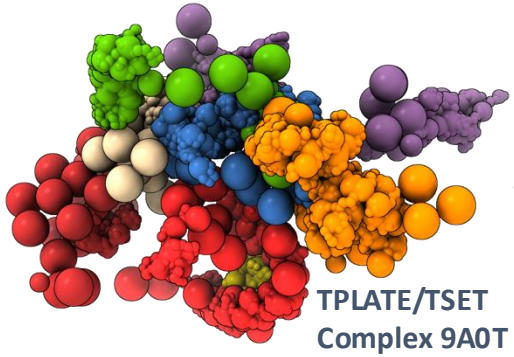
Human Fibrin Clots 8ZZU



BBSome 8ZZI

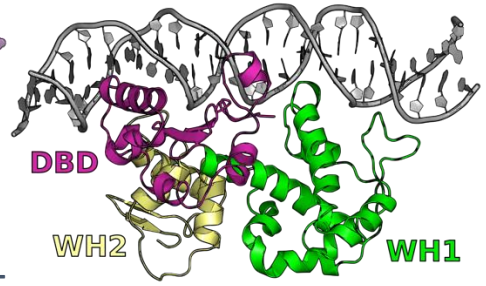


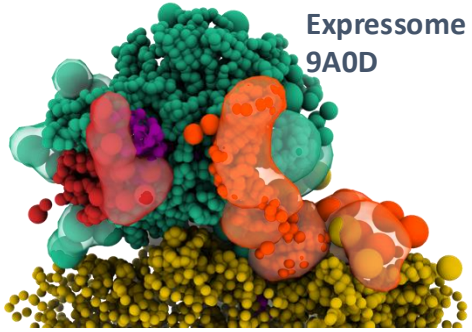
COP9 Signalosome 9A01



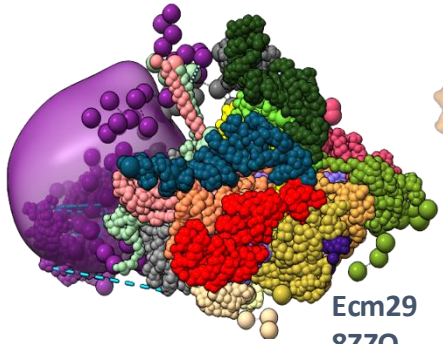
TPLATE/TSET Complex 9A0T

TrfA with dsDNA 9A0W

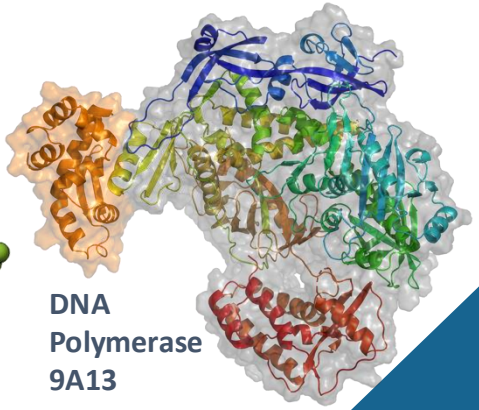




Expressome 9A0D

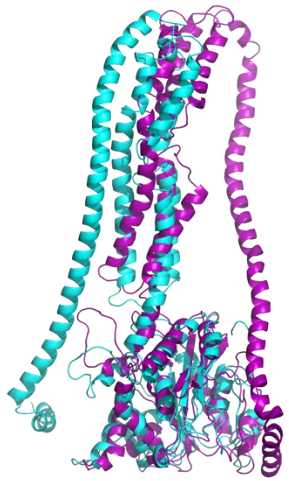


Ecm29 8ZZQ



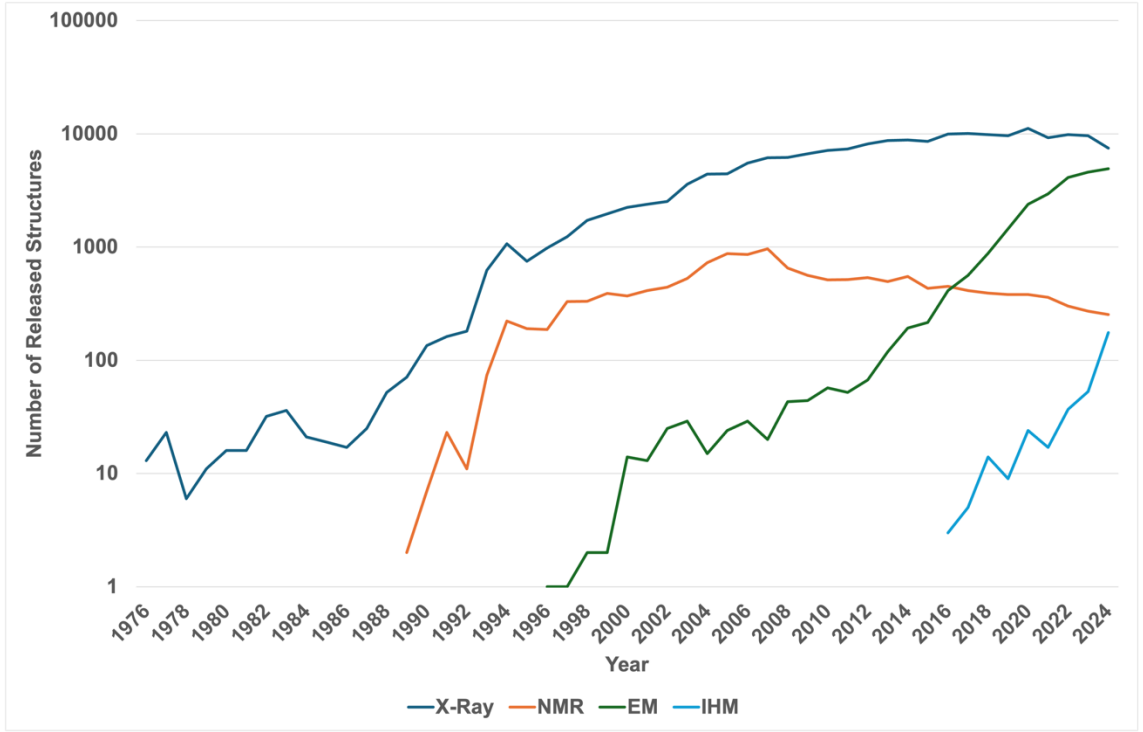
DNA Polymerase 9A13

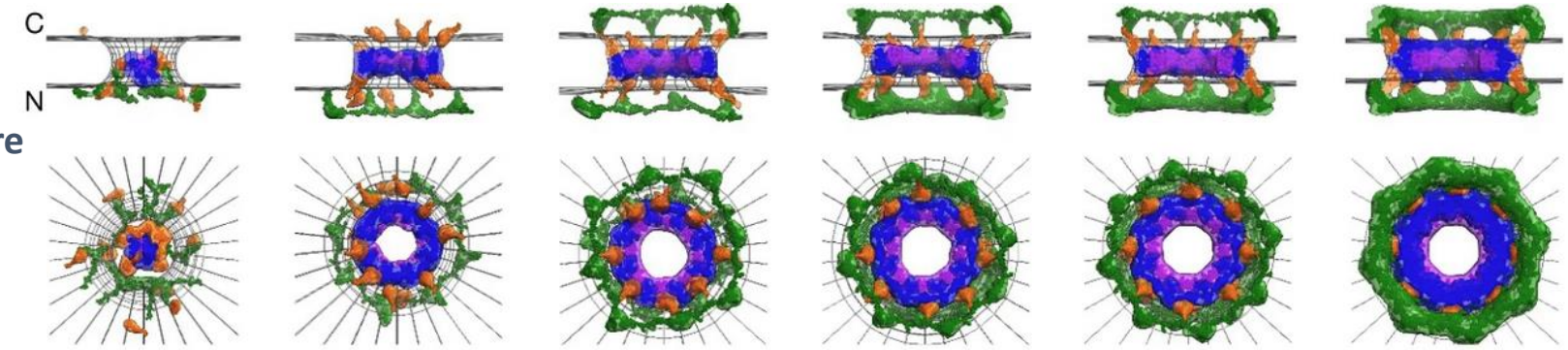
Structures in PDB-IHM



Multiple States of hGBP1 9A1G

CirA by AlphaLink 9A36





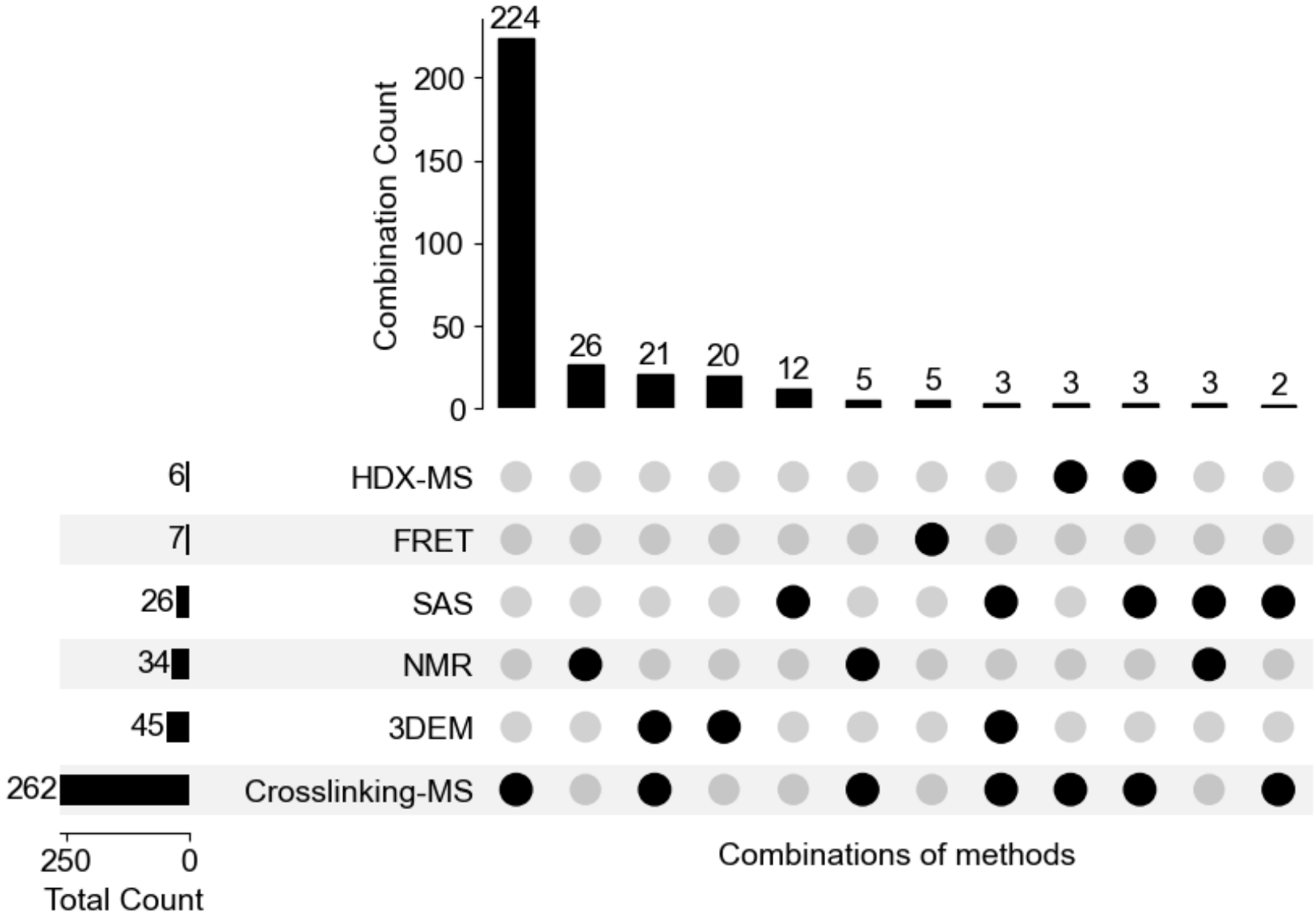
Nuclear Pore Complex 9A25

Otsuka S et al. 2023. Nature. 613, 575-581

Peulen T et al. 2023. Elife, 12, e79565

Stahl K et al. 2023. Nat Biotechnol. 41, 1810-1819

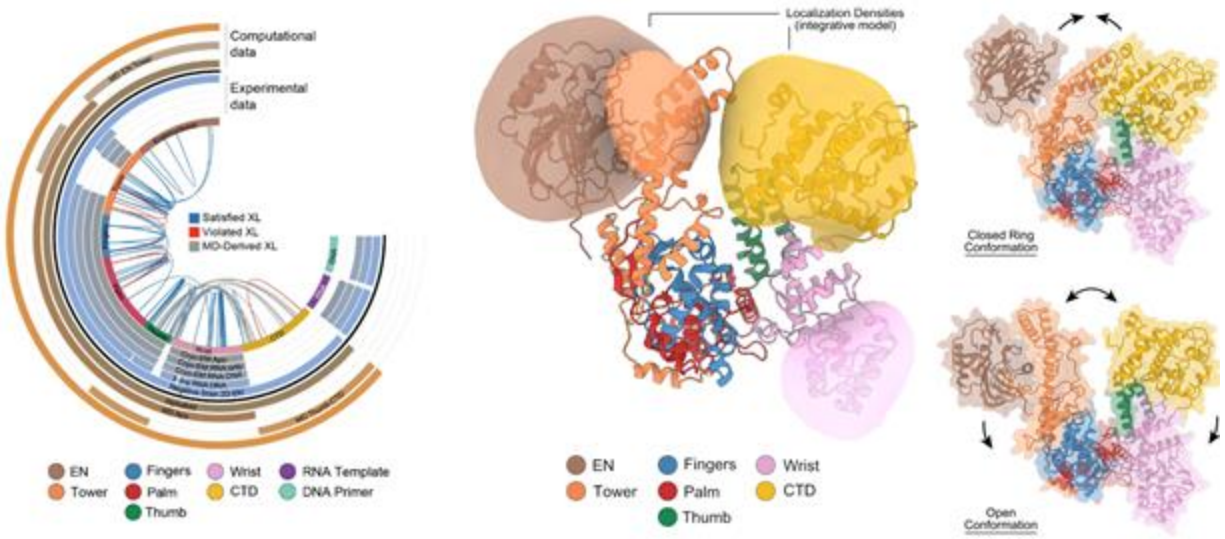
Experimental Methods in PDB-IHM

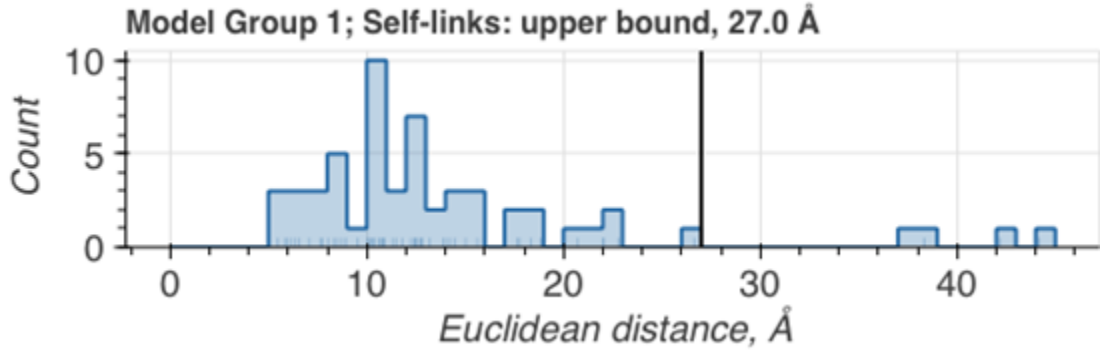


Vallat B et al. 2025.
J. Mol. Biol. 437, 168963

PDB-IHM: Impact on 3DEM-based Structures

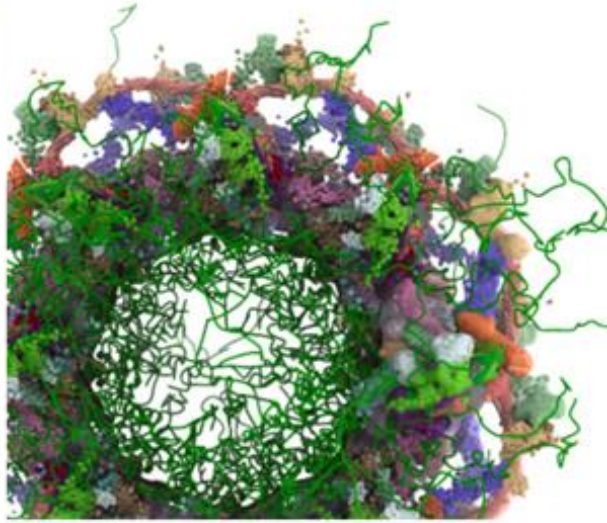
Integrative structure of the
Human LINE-1 ORF2 Protein (9a3q)





- ~25% of low resolution 3DEM structures deposited in the PDB in 2023 are likely integrative structures
- Many of these use crosslinking-MS data and are not annotated appropriately, leading to incomplete data capture
- PDB-IHM provides a robust mechanism for archiving such structures along with complete information
- Enables enhanced structure validation and access to FAIR data

Baldwin ET, et al. 2024. Nature. 626(7997), 194-206



PDB-IHM: Looking Forward

Integrative Structures: Looking Forward

- Community engagement to improve deposition
  - Encourage deposition of integrative structures to PDB-IHM
  - Communicate with journals to recommend structure deposition
  - Enhanced deposition practices for 3DEM structures that use additional experimental restraints
- Continue to support evolving integrative modeling methods
- Continue to develop methods for validation
  - Collaborate with other experimental communities
  - Enhance interoperation with other data resources
  - Deliver validation data in software-amenable mmCIF format
  - Develop Bayesian approach for validation
- Create a federated network of interoperating data resources that contribute to integrative structural biology

Berman HM et al. 2019. Structure. 27, 1745-1759

Transition to Extended PDB IDs and PDBx/mmCIF File Format

By 2028

4-character PDB IDs (*e.g.* **1abc**) will be fully allocated. After that, all new entries will be assigned only extended PDB IDs.

Extended PDB ID Format

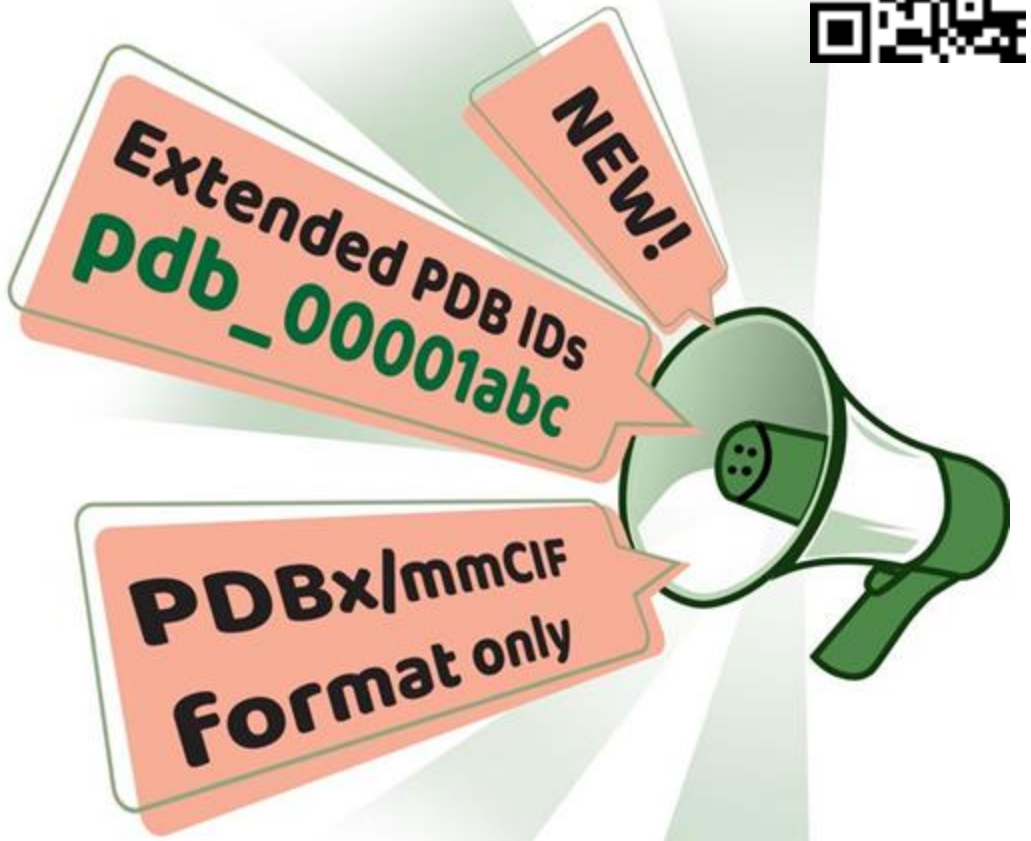
8 alphanumeric characters with prefix *pdb_*

e.g. **pdb_00001abc**

Entries with extended PDB IDs will be available only in PDBx/mmCIF format.

All PDB users, including software developers, and journal editors must transition to this format.

Start using extended IDs and PDBx/mmCIF data files today



PDB-IHM entries will also transition to extended PDB IDs

WORLDWIDE PDB PROTEIN DATA BANK

wwpdb.org/documentation/new-format-for-pdb-ids

RCSB PDB
PROTEIN DATA BANK

RCSB.ORG • info@rcsb.org

Core Operations Funding

US National Science Foundation (DBI-2321666), National Institute of General Medical Sciences, National Institute of Allergy and Infectious Disease, and National Cancer Institute (NIH R01GM157729), and the US Department of Energy (DE-SC0019749)

Management

RUTGERS–NEW BRUNSWICK
Artificial Intelligence and Data Science Collaboratory

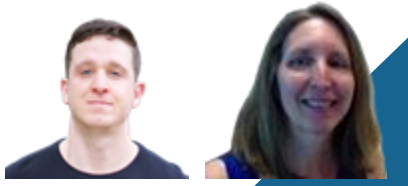
UC San Diego
SDSC SAN DIEGO SUPERCOMPUTER CENTER

UCSF
University of California San Francisco

Follow us

WORLDWIDE PDB
PROTEIN DATA BANK

Member of the Worldwide Protein Data Bank (wwPDB; **wwpdb.org**)



Acknowledgements

Rutgers

Helen Berman

Stephen Burley

John Westbrook

Cathy Lawson

Monica Sekharan

UCSF

Andrej Sali

Benjamin Webb

Arthur Zalevsky

Jared Sagendorf

Sai Ganesan

Ignacia Echeverria

USC

Carl Kesselman

Hongsuda Tangmunarunkit

Serban Voinea

Aref Shafaeibejestan

wwPDB Directors

Stephen Burley, RCSB PDB

Jeffrey Hoch, BMRB

Genji Kurisu, PDBj

Kyle Morris, EMDB

Sameer Velankar, PDBe

PDB-IHM Depositors

ChimeraX and Mol* Developers

Funding

PDB-IHM: NSF awards DBI-1519158, DBI-1756248, DBI-1756250, DBI-2112966, DBI-2112967, DBI-2112968

wwPDB Integrative/Hybrid Methods Task Force

Federation Working Group

Helen Berman

Jill Trewhella

Rudolph Aebersold

Nigel Kirby

Paul Adams

Gaetano Montelione

Wah Chiu

George Phillips Jr.

Bridget Carragher

Juri Rappsilber

Claus Seidel

Geerten Vuister

Dmitri Svergun

Model Working Group

Andrej Sali

Torsten Schwede

Alexandre Bonvin

Frank Dimaio

Gerhard Hummer

Jens Meiler

Emad Tajkhorshid

Task Force Members

Matteo dal Peraro

Kay Grunewald

Richard Henderson

Kenji Iwasaki

Graham Johnson

Marc Marti-Renom

Frank Alber

Michael Nilges

Ruth Nussinov

Randy Read

Helen Saibil

Gunnar Schroder

Charles Schwieters

Maya Topf

Aleks Gutmanas

Catherine Lawson

Ardan Patwardan

Eldon Ulrich

John Westbrook

Gerard Kleywegt

Haruki Nakamura

Thank You