wwPDB Processing Procedures and Policies Document

Section B: wwPDB Policies

Authored by the wwPDB annotation staff

May 2016 version 3.8

Table of Contents

1 PDB Entry Requirements 4
2 Entry Authorship 7
3 Release of PDB Entries 8
4 Assignment of PDB IDs and Ligand codes 11
5 Changes to Entries 11

Preface

This document outlines the annotation procedures and policies of the wwPDB. Given the complex nature of some of the issues that can arise during processing, exceptions to policy are considered on a case-by-case basis by the wwPDB leaders.

The two sections in the complete document are:

A: wwPDB processing procedures

B: wwPDB policies

Further information about these sections is available in the introduction to each section.

1 PDB Entry Requirements

What are the requirements of acceptance of an entry to the PDB?

The wwPDB will accept all experimentally determined structures of biological macromolecules that meet the minimum requirements. These requirements include: three-dimensional atomic coordinates, information about the composition of the structure (sequence, chemistry, etc.), information about the experiment performed, details of the structure determination steps and author contact information are also necessary for the deposition. In addition, structure factor or intensity data are required for X-ray submissions, restraints and chemical shifts are required for NMR submissions, and electron density maps are required for EM submissions.

On occasion, the wwPDB receives a request to deposit a structure that was determined before deposition of experimental data became mandatory, for which the experimental data are no longer available. It is difficult to validate such structures without experimental data.

In such cases, the wwPDB Directors will determine if the structure can be deposited to the PDB. Criteria for accepting structures without experimental data that were determined by X-ray or NMR experimental methods are as follows: there is a peer-reviewed publication prior to January 1st 2008 describing the corresponding structure(s) and either the polymer sequence is not represented in the PDB archive or the deposition includes one or more ligand(s) not currently represented in the PDB Chemical Component Dictionary.

Which type of experimentally determined structures are accepted by the PDB?

Since October 15, 2006, PDB depositions are restricted to atomic coordinates that are substantially determined by experimental measurements on actual sample specimens containing biological macromolecules[1]. Currently, coordinate sets produced by X-ray crystallography, NMR, electron microscopy, neutron diffraction, powder diffraction, fiber diffraction, and solution scattering can be deposited to the PDB, provided the molecule studied meets the minimum size requirement.

For structures generated using exact symmetry operations (e.g., strict helical, point, or non-crystallographic symmetry (NCS)), the authors should deposit only those chains that were fitted/refined and supply PDB with the operators (matrix transformations) that can be used to generate the complete assembly. For crystal structures, authors should also supply the NCS operators that generate the crystal asymmetric unit. Example: in PDB entry 2wbh, the complete MS2 bacteriophage capsid biological assembly is generated from the three deposited chains (A, B, C) by applying the 60 operators provided in _pdbx_struct_oper_list (REMARK 350) records. The crystal asymmetric unit of 2wbh, which corresponds to 1/3rd of the complete

[1] H.M. Berman, S.K. Burley, W. Chiu, A. Sali, A. Adzhubei, P.E. Bourne, S.H. Bryant, J. Roland L. Dunbrack, K. Fidelis, J. Frank, A. Godzik, K. Henrick, A. Joachimiak, B. Heymann, D. Jones, J.L. Markley, J. Moult, G.T. Montelione, C. Orengo, M.G. Rossmann, B. Rost, H. Saibil, T. Schwede, D.M. Standley, and J.D. Westbrook (2006) Outcome of a workshop on archiving structural models of biological macromolecules. *Structure*. **14**: 1211-1217

capsid, is generated from the three deposited chains by applying the 20 NCS operations provided in the _struct_ncs_oper (MTRIX) records.

ftp://ftp.wwpdb.org/pub/pdb/data/structures/divided/pdb/wb/pdb2wbh.ent.gz
ftp://ftp.wwpdb.org/pub/pdb/data/structures/divided/mmCIF/wb/2wbh.cif.gz

Theoretical model depositions determined purely *in silico* using, for example, homology or *ab initio* methods, are no longer accepted.

Theoretical models that have been previously released or those that were deposited before October 15, 2006 will continue to be publicly available via the historical models archive at ftp://ftp.wwpdb.org/pub/pdb/data/structures/models/.

Structures determined by methods new to the PDB will be reviewed in consultation with community of experts to determine if structures determined by the method should in principle be accepted by the PDB. Once a determination is made a new template for the PDB entries from this method will be developed.

How and where are experimental data submitted?

The deposition sites for all experimental methods are available at the following wwPDB sites:

RCSB http://deposit.rcsb.org/

PDBe http://pdbe.org/deposit

PDBj http://pdbdep.protein.osaka-u.ac.jp/

For NMR model coordinates and experimental data an additional access point is located at:

BMRB http://deposit.bmrb.wisc.edu/bmrb-adit/

PDBj-BMRB http://nmradit.protein.osaka-u.ac.jp/bmrb-adit/

PDBe http://www.ebi.ac.uk/pdbe-xdep/autodep/index.jsp

For EM model coordinates and maps data an additional access point is located at:

EMDB http://emdatabank.org/?q=deposition..

What are the format requirements for deposition?

To ensure that all the deposition tools work with minimal error, the format requirements for depositing structures are:

Coordinates and meta data

PDBx format Deposition can be prepared in PDB mmCIF exchange format (PDBx). Definitions and dictionary are available in HTML, ASCII and XML format (see http://mmcif.pdb.org/ for details).

PDB format Definitions and format content guide are available in PDF and HTML format (see http://www.wwpdb.org/docs.html for details).

- Each polymer must be assigned a unique alphanumeric chain ID (A-Z, 0-9, a-z) in column 22.
- Each residue must be assigned a unique residue number within that chain ID.
- If there are alternate conformations in the structure, the alternate conformation indicator must be provided in column 17.
- Atoms with the same coordinates may not be present in the same MODEL.
- All polymer chains should have a TER card at the end. No TER cards should be included at the end of non-polymer residues (such as ions, ligands, waters).
- There should be only one END card at the end of the file.
- Each record must start in the first column.
- Record types (such as COMPND, ATOM, SEQRES) must be capitalized.
- If the entry contains ANISOU, SIGATM, or SIGUIJ lines, there must be a corresponding ATOM record.
- Depositions with multiple models should include MODEL and ENDMDL cards. The models should be listed sequentially in columns 11-14.
- Chain IDs used in the SEQRES records should correspond to those in the coordinates.

Structure factors

- The structure factor file can either be in mmCIF, ASCII, or mtz formats and should include h, k, l, F, SigmaF (and/or I and SigmaI) and test flags.
- Definitions and format are available at http://www.wwpdb.org/documentation/docs.html.

Restraint Data

- Currently there are no restrictions on the format.
- The common format is NMR-STAR (see http://www.bmrb.wisc.edu/formats.html for details).

What types of structure can be deposited to the PDB?

Biomolecular polymers including polypeptides, polynucleotides, polysaccharides, and their complexes that meet the following criteria are accepted:

- For polypeptide structures
  -Gene products
  -Naturally-occurring peptides that are non-ribosomal in origin
  -Peptidic repeat units of larger polymers (such as fibrous and amyloid polymers)
  -Synthetic peptides of at least 24 residues within a polymer chain (smaller, biologically-relevant oligopeptides are also acceptable)
- Polynucleotide structures
  -Of four or more residues
- Polysaccharide structures
  -Of four or more residues

Crystal structures of peptides with fewer than 24 residues within any polymer chain that do not meet the criteria above should be deposited at the Cambridge Crystallographic Data Centre (CCDC, http://www.ccdc.cam.ac.uk/products/csd/deposit/). NMR structures of such molecules should be submitted to Biological Magnetic Resonance Data Bank (BMRB) through the Small

Molecule Structure Deposition (SMSdep, http://smsdep.protein.osaka-u.ac.jp/bmrb-adit/) system.

Smaller oligonucleotides (dinucleotides and trinucleotides) can be deposited at the Nucleic Acid Database (NDB, http://ndbserver.rutgers.edu).

Molecules that do not conform to these guidelines but have been previously deposited in the PDB will not be removed.

Can a re-refined structure be deposited to the PDB?

A re-refined structure based on the data from a different research group or lab can only be deposited to the PDB if there is an associated publication available describing the details of the re-refined structure. A re-refined entry may be deposited prior to publication but will not be processed (will have REFI status) or released until the associated publication has become publicly available. The depositor must provide the relevant publication details to the PDB and allow for extra time required for the processing and release of these entries.

In addition, a dedicated remark (REMARK 0) will be added to the PDB file along with the primary citation of the original PDB entry (under REMARK 1). There will be no data collection and processing information (REMARK 200 for Xray or REMARK 210 for NMR) in the PDB file.

Details on the annotation of a re-refined PDB entry can be found at http://www.wwpdb.org/procedure.html.

2 Entry Authorship

There are 3 types of authorship associated with a PDB entry: Entry Author, Contact Author, and Citation Author.

The supervisor of the research group where the structural determination work began, known as the Principal Investigator (PI) or Team Leader equivalent, is responsible for the authorship represented in the final PDB entry. If more than one PI or Team Leader equivalent is responsible for the entry, they will need to come to a mutual decision on all issues.

Contact Authors

The Contact Authors indicated at the time of deposition are responsible for depositing the structure, responding to any queries from the wwPDB during processing, and indicating when entries can be released.

At least one Contact Author should be designated "responsible for correspondence" including data submission and responses to questions from the wwPDB. The PI or Team Leader equivalent must be listed as a Contact Author and will be copied on all communications. In some cases, the PI or Team Leader equivalent may be contacted with questions directly. It is the responsibility of the depositor to label author roles correctly.

All Contact Authors will be notified of any changes or requests for changing/obsoleting/removing entries. In the case of a conflict between Contact Authors, the PI makes the final decision.

Entry Authors

The PI or Team Leader equivalent should be included. In addition, it is recommended that all who contributed to the structural determination as identified by the PI or Team Leader equivalent be designated as Entry Authors. Commercial entities should include the company name along with any other relevant names.

Entry Authors can be the same as those listed in the primary citation, or a subset of Citation Authors. Alternatively, there may be more Entry Authors listed than there are Citation Authors. At least one Entry Author should be included in the author list for the primary citation.

It is the responsibility of the PI/Team Leader equivalent to ensure appropriate listing of Entry Authors, and that all listed Entry Authors have approved the final version of the data and have agreed to PDB submission.

Citation Authors

Citation Authors are those listed on the primary publication describing the entry. The Citation Author list may be different from the Entry Author list as described above.

If an entry is to be obsoleted, it is the PI's or Team Leader equivalent's responsibility to notify the corresponding author of the paper.

Authorship and Re-refined Entries

A re-refinement of data available in the PDB must acknowledge the original data set by citing the PDB entry (and corresponding citation, if available) in the re-refined PDB entry. This information can be noted at the time of deposition. A re-refined entry may be deposited prior to publication but will not be processed (will have REFI status) or released until the associated publication has become publicly available. See the wwPDB Processing Procedures Document for further information (Section A.9).

3 Release of PDB Entries

A journal policy of *release upon publication* takes precedence over the 6-month or 1-year hold policy.

What are the status codes for PDB entries?

REL entries are processed and released as soon as authors have approved the processed files. If we do not hear from the authors within three weeks from the mailing of the annotation report and assuming there are no major issues with the submission, we will consider this entry to be author approved. The entry will then be released. If the corresponding paper has been published and there are outstanding issues[2], the entry will be released by wwPDB staff with

[2] Note: Any issues that arise during annotation that prevent the standard processing of submissions are considered to be outstanding issues. These could include, for example, unusual geometry and stereochemistry, sequence-related problems, solvent structures, to name but a few.

CAVEAT record. If the corresponding paper has not been published and the entry has outstanding issues which have not been addressed, the entry will be withdrawn by the wwPDB at the end of the 12-month period after deposition.

Entries can be released without citation information and updated with the complete citation information later.

HPUB entries released upon publication. The wwPDB receives publication dates and citation information from the authors, some journals as well as user community. In addition, the wwPDB scans the literature for publications. While the wwPDB makes every effort to track citations and release files accordingly, it is ultimately the responsibility of the depositors to notify the wwPDB when the citation has been published.

HPUB structures are released when they have been published, either in electronic or paper publication, whichever is sooner. If the manuscript is first released electronically, the entry will be released at that time. The author cannot request to delay the release of the entry until paper publication of the manuscript. The author cannot withdraw an entry once the paper of the corresponding structure is published. It may be obsoleted with replacement coordinates.

Normally authors approve before entries are released. If the contact author does not reply and assuming there are no major issues with the submission, we will consider this entry to be author approved. The entry will be automatically released when the corresponding paper has been published. If the paper has been published and there are outstanding problems with the entry, it will be released with CAVEAT record.

In the case of HPUB for which there is no publication, there is a one-year hold limit. Entries cannot be held for more than one year past the date of deposition. If there are outstanding problems, further reminder will be sent. By the end of 10 months after deposition, a letter will be sent to the authors of the deposition asking if they wish to release or withdraw the entry by the 1 year anniversary of the deposition date. If the e-mail sent to the contact author(s) bounces or the author does not reply, the entry will be automatically released at 12 months after deposition if there are no outstanding issues. If there are outstanding problems with the entry and it has not been published, it will be withdrawn.

HOLD entries are placed on hold for one year from the date of deposition. They may be released earlier on the date specified by the author. When the corresponding electronic or paper publication occurs, the entry must be released if the journal's policy requires release upon publication. Once the paper is published the entry cannot be withdrawn. It can be obsoleted with replacement coordinates.

Normally, authors approve entries before they are released If the author does not reply to annotation correspondence however, and assuming there are no major issues with the submission, we will consider it to be author approved. The entry will be automatically released when it reaches end of its HOLD period. If there are outstanding problems with the entry, further reminders will be sent.

By the end of 10 months from deposition, a letter will be sent to the authors asking if they wish to release or withdraw the entry at the one year anniversary of the deposition date. If authors do not respond, it will be released irrespective of the availability of a publication at the end of the one year HOLD period.

In the case that there are outstanding issue with the entry at the end of the one year HOLD period,, it will be releasedwith a CAVEAT record if the corresponding paper has been published.

If no publication is found, and issues with the entry prevent complete annotation, then wwPDB may withdraw the entry.

Entries cannot be held for more than one year past the date of deposition

WDRN (withdrawn) Authors may withdraw their unreleased entries at any time as long as the paper has not been published, following the same deadlines as written in the release section. Withdrawn entries will appear in the unreleased entries search as withdrawn.

Problem structures will be discussed with the depositors to resolve issues such as unusual structural chemistry, many distant waters, long/short covalent bonds, many sequence mismatches or other conflicts. Entries for which these issues can not be resolved (as determined by the wwPDB staff) will be withdrawn if the paper has not been published. The withdrawn status of an entry may be reversed by the PDB (as determined by the wwPDB staff) if there is now a publication referring to this withdrawn previously approved entry.

Can the experimental data be released separately from the coordinate file?

The depositor-assigned release status (REL, HPUB, or HOLD) has to be the same for experimental and coordinate data. Coordinate and experimental data files must be released at the same time.

What are the deadlines for requesting release of entries?

PDB entries are processed by the members of the wwPDB (RCSB PDB, PDBe and PDBj). They are either released immediately (REL), when the corresponding paper is published (HPUB), or on a particular date (HOLD).

Each week, all files scheduled for release or modification are checked and validated one final time. Authors may be contacted to resolve any issues that may arise while preparing the entries for release.

When the release of HPUB structures is requested, the PDB staff routinely confirms the primary citation. If this is not accomplished within that release cycle, the entry may be scheduled to be released in a later update.

To be included in the next update, any required author correspondence should be sent to the appropriate wwPDB member by 12:00 hrs noon on Thursday (local time at processing site). Requests received after these cutoff times will be processed during future update cycle.

The email addresses to contact the wwPDB centers are:

- RCSB (deposit@deposit.rcsb.org)
- PDBe (pdbdep@ebi.ac.uk)
- PDBj (adit@adit.protein.osaka-u.ac.jp)

All entries due for release are transferred to the RCSB for final packaging into the master PDB ftp archive. Data entries are added to the PDB archive on a weekly schedule synchronized among FTP sites at RCSB PDB, PDBe, and PDBj.

The process for weekly PDB archive data release with the advice and concurrence of the Advisory Committee to the Worldwide Protein Data Bank follows:

Phase I: Every Saturday by 3:00 UTC, for every new entry, the following will be provided from the wwPDB website: 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.

REVDAT dates The REVDAT date indicates the date of release of the entry. Entries processed from the Wednesday after the last release to the Tuesday of the current release have Wednesday as the REVDAT date.

Who has access to unreleased data?

Unreleased coordinate sets are distributed only to the authors of the particular entry. Reviewers of the paper may not obtain unreleased coordinate sets from the PDB. If a reviewer wishes to access the validation report, the reviewer should contact the journal editor, who in turn will obtain the validation report from the author and forward them on to the reviewer.

Are Email addresses of authors publicly available? The email addresses of authors who deposit PDB entries are not made publicly available and will not, individually or in bulk, be distributed to those who request them.

What information is available for unreleased entries? The unreleased entries search of the wwPDB web sites contains the title, authorship, status, PDB ID, experimental data status and sequence availability for each entry. Titles may be suppressed at the request of the author, but the authorship, status and PDB ID can not be publicly suppressed.

4 Assignment of PDB IDs and Ligand codes

Can a PDB ID or ligand code be requested? Neither a single PDB ID/ligand code nor a range of PDB IDs/ligand codes can be requested. PDB ID and ligand codes are automatically assigned and do not carry intrinsic information.

When are PDB IDs assigned? PDB IDs are automatically assigned by software when the author has completed his/her deposition (i.e. the author has filled out at least the minimal information for deposition and has pressed the deposit & confirmation buttons.)

5 Changes to Entries

What changes can be made before release? Authors can update the coordinates, structure factors, as well as the related header information any time before release as long as the data is not collected after deposition. If the author has collected new data after deposition and wishes to replace the original deposited data, the author will have to withdraw the old entry and deposit the new entry using the online deposition tools to obtain a new PDB ID. This is because the new data set will be entirely different from the original for data collection, structure factors, refinement, and will need to be completely re-processed. Authors can base the new deposition on the old data in regards to sequence and taxonomy.

If the depositor sends new coordinates for an entry shortly before or at the time of electronic or paper publication, the release of the entry may be delayed because the file must be re-processed.

Once an entry is marked for release, the author has until the deadline time listed above (see Section 3, *Deadline for requesting release of entries*) to submit revisions or to request the entry not to be released.

What changes can be made after release?

Minor changes may be made. These are defined as:

- Update on header section such as citation, author's name, etc.
- Update or change on structure factor or constraint file due to format corrections or addition of data set while coordinates remain unchanged.
- Changes to chain ID, numbering, atom name, ordering of molecules and/or ligands.
- Changes to the terminal sequence- add or remove a region that is unobserved in the coordinates.

A REVDAT appear in the file with a description of the change

Major revisions to coordinates that change the structure's geometry or chemical composition (such as a change in the sequence of the polymers or ligand identity) require the entry to be *obsoleted and superseded* by a new deposition. The major revisions include:

- Replacement in any existing coordinates (the x,y,z values themselves).
- Changes in chemical composition (i.e. from SO4 to PO4 or SER to CYS) which change the identity of the current ligands.
- Changes in sequence that is not adding or removing unobserved terminal region.

Typically, released PDB data (coordinates and experimental data) are obsoleted when the authors have collected new data or re-refined the structure. The obsoleted entry is replaced by a new (superceding) entry that receives a new PDB ID. Obsolete entries remain available to the public through the ftp archive. Users who search for an obsolete structure through the main web search interface will be automatically redirected to the superceding entry. Under no circumstances can a released structure be withdrawn.

There are some rare circumstances in which an obsolete structure is not superceded.

- The publication is retracted. The associated PDB entry will be obsoleted if requested by the journal. If a request has not been received, the wwPDB will do its best to contact the depositor and co-authors, (former) PIs, journal editors, etc. when made aware of the retraction. If the reason(s) for retraction were such that the associated PDB entry needs to be made obsolete, PDB will obsolete the entry. The citation in the obsoleted entry is the published journal retraction.
- The structure is incorrect, and the entry author obsoletes the entry. The entry must contain a statement as to the reason for obsoleting the structure.
- A third-party (such as the employer) requests that the entry is obsoleted (e.g., in case of malfeasance). The citation in the obsoleted entry must be a published explanation and retraction in a peer-reviewed journal.

wwPDB Remediation

The wwPDB reviews the entire archive on a regular basis and remediates the data. The coordinates themselves are not changed but there may be changes in the meta data and nomenclature to assure consistency and uniformity in the files. The nature of the changes are described in a public document on the wwPDB site. In the case of global remediation the

individual authors are not contacted. A version number is assigned. A REMARK with this version number and date is in every file. The older version is maintained as a snapshot on the FTP site.