

wwPDB Processing Procedures and Policies Document

Section B: wwPDB Policies

Authored by the wwPDB annotation staff

July 2016 version 4.0

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 Directors/Heads.

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. Map volume deposition to EMDB is mandatory for PDB depositions of 3DEM models.

For structures determined by X-ray crystallography, all atoms must have full B factors. If TLS was used during refinement, the residual B factors must be converted to full B factors. All atoms described by TLS records must have associated ANISOU records.

On occasion, the wwPDB is asked to archive a structure that was determined before deposition of experimental data became mandatory and the experimental data are no longer available. It is difficult to validate such structures without experimental data.

In such cases, the wwPDB Directors/Heads will determine if the structure can be deposited to the PDB. Criteria for accepting structures determined by experimental methods but without experimental data 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.

What types of experimentally-determined structures are accepted by the wwPDB?

Since October 15, 2006, PDB depositions have been restricted to atomic coordinates that are substantially determined by experimental measurements on actual sample specimens containing biological macromolecules¹. Currently, coordinate sets produced by X-ray crystallography, NMR, electron microscopy, neutron diffraction, powder diffraction, and fiber diffraction can be deposited to the PDB, provided the molecule studied meets the minimum size requirement.

Use of non-crystallographic symmetry (NCS):

For crystal structures, coordinates for the complete asymmetric unit should be provided, even if non-crystallographic symmetry (NCS) is used. The only exceptions are models that produce highly symmetric assemblies (e.g., viruses, helical symmetry, etc.) in which only a portion of the asymmetric unit is used in refinement. Depositors must provide the model in the standard crystal frame along with the NCS matrices.

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

For microscopy methods, if symmetry operators are required to create the complete biological assembly from the modeled coordinates, such operators must be provided.

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 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 not currently supported by 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 such a determination is made, a new template for PDB entries derived 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.wwpdb.org/deposition/>

PDBe <https://deposit-pdbe.wwpdb.org/deposition/>

PDBj <http://deposit-pdbj.wwpdb.org/deposition/>

For deposition of additional NMR experimental data, an access point is located at:

BMRB <http://deposit-beta.bmrb.wisc.edu/>

PDBj-BMRB <http://deposit.bmrb.pdbj.org/>

What are the format requirements for deposition?

To ensure that all wwPDB and related deposition tools function with minimal loss of data fidelity, depositors should output PDBx/mmCIF format files from their refinement program, if supported. A [PDBx/mmCIF preparation guide](#) is available. The format requirements for depositing structures are as follows:

Coordinates and meta-data

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

PDB format: Definitions and format content guide are available in PDF and HTML format (see <http://wwpdb.org/documentation/file-format> for details).

Structure factors

- The structure factor file can either be in mmCIF 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://wwpdb.org/documentation/procedure#toc_appendicsB.

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.

Structures of 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 generated by a research group or laboratory different from the contributors can only be deposited to the PDB if there is an associated peer-reviewed

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 peer-reviewed publication has become publicly available. The depositor must provide the relevant publication details to the wwPDB and allow for extra time required for the processing and release of these entries.

In addition, a dedicated remark (`_pdbx_database_remark` with `id = 0`) will be added to the PDBx/mmCIF file along with the primary citation of the original PDB entry (under `_citation` with `id = original_data_1`). There will be no data collection and processing information (e.g. experimental details or sample condition and diffraction source and detectors or spectrometer) in the PDBx/mmCIF file.

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

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/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/Team Leader equivalent must be listed as a Contact Author and will be copied on all communications. In some cases, the PI/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 wwPDB will follow requests from the PI/Team Leader equivalent who ultimately makes the final decision. The PI/Team Leader equivalent is the individual(s) specified as PI/Team Leader in the PDB entry that was approved by the contact authors.

Entry Authors

The PI/Team Leader equivalent should be included as Entry Author. In addition, it is recommended that all who contributed to the structural determination as identified by the

PI/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.

It is the responsibility of the PI/Team Leader equivalent to ensure that the listing of Entry Authors is appropriate 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 responsibility of the PI/Team Leader equivalent 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

Problem structures, as identified on the basis of the wwPDB validation report, 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 cannot be resolved will be withdrawn upon expiration of the one-year hold, unless a corresponding paper has been published. In that case, the entry will be released by wwPDB staff with a CAVEAT record (database_PDB_caveat). If a publication appears in the literature that cites a withdrawn entry, the withdrawn status of an entry may be reversed by the PDB (as determined by the wwPDB staff)

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 time when the annotation report is made available and assuming there are no major issues with the submission, the entry will be deemed to be author approved. The entry will then be released. If the corresponding paper has been published and there are outstanding issues², the entry will be released by wwPDB staff

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.

with a CAVEAT record (database_PDB_caveat). 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 at a later date.

HPUB entries are released upon publication. The wwPDB receives publication dates and citation information from the authors, some journals, and the 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.

It is normal practice for authors to approve entries before they are released. If the contact author does not reply and assuming there are no major issues with the deposition, we will deem 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 a CAVEAT record (database_PDB_caveat).

Entries cannot be held for more than one year beyond the date of deposition. Thus, if no publication occurs within this period the entry must be released. Ten months following deposition, a communication will be sent to the authors of the deposition asking if they wish to release or withdraw the entry before the one year anniversary of the deposition date. If there are outstanding problems, a further reminder will be sent. If the author does not respond, the entry will be automatically released at 12 months after deposition, provided 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 a date specified by the Contact Author. When the corresponding electronic or paper publication occurs, the entry must be released if the journal policy requires release upon publication. Once the corresponding paper is published, the entry cannot be withdrawn. It can be obsoleted with replacement coordinates.

Normally, authors approve entries before they are released. If, however, the author does not reply to annotation correspondence, and provided 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.

Ten months from deposition, a communication will be sent to the authors asking if they wish to release or withdraw the entry on or before 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 issues with the entry at the end of the one-year HOLD period, it will be released with a CAVEAT record (database_PDB_caveat) if the corresponding paper has been published. If no corresponding publication can be found following a good faith effort, and issues with the entry prevent complete annotation, then wwPDB may withdraw the entry.

WDRN (withdrawn) Authors may withdraw their unreleased entries at any time, provided the paper has not been published, following the same deadlines as written in the release section. At the time the entry is withdrawn, the latest version of the processed files will be made available to authors in case they wish to re-deposit in the future. Withdrawn entries will remain in the [status list](#) for unreleased entries at PDB archive.

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

Coordinates and experimental data share the same release status (REL, HPUB, or HOLD). Thus, coordinate and experimental data files can only be released simultaneously.

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 revision are subjected to a final data integrity check. Contact Authors may be asked to resolve issues arising while entries are prepared for release.

When release of HPUB structures is requested, wwPDB staff routinely check for the primary citation. To be included in the upcoming update, any required author correspondence should be sent to the appropriate wwPDB member by 12:00 noon on Thursday (local time at processing site). Occasionally the request cutoff date may be changed under certain circumstances. Requests received after these cutoff times will be processed during a later update cycle.

Depositors should contact the wwPDB through the web communication within the relevant deposition while general PDB users should contact the wwPDB at deposit-help@mail.wwpdb.org regarding publication and/or release.

All entries set for release are transferred to the RCSB-PDB (the current Archive Keeper) for final packaging into the master PDB ftp archive. Data entries are added to the PDB archive on a weekly schedule and 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, is as follows:

Phase I: Every Saturday from 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 from 00:00 UTC, all new and modified data entries will be updated at each of the wwPDB FTP sites.

Revision dates (`database_PDB_rev.date`, `pdbx_version`) The revision date indicates the date of release of the entry. Revision date will be set to the date of scheduled release, which is on Wednesday.

Who has access to unreleased data?

Unreleased coordinate sets are distributed only to the authors of that entry. Reviewers of the journal submissions may not obtain unreleased coordinate sets from the wwPDB. The wwPDB strongly encourages journal editors and referees to request wwPDB validation reports from authors as part of the manuscript submission and review process.

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

What information is available for unreleased entries? Unreleased entries at the PDB archive contain the title, authorship, status, PDB ID, experimental data status and sequence availability. Entry titles and authorship may be suppressed at the request of the Contact Author, but status and PDB ID cannot 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 may be requested. The wwPDB reserves the right to change author's ligand codes. PDB ID and ligand codes are automatically assigned and do not carry identifying information.

When are PDB IDs assigned? PDB IDs are automatically assigned by the deposition software tool, 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, and related header information any time before release, provided the experimental data was not produced after deposition. If the Contact Author has collected new experimental data after deposition and wishes to replace the original deposited data, the author must 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.

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 subject to delay 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 metadata 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 sample sequence- add or remove a region that is unobserved in the coordinates such as terminal or loop region.

A revision record (pdbx_version) 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 (e.g., SO₄→PO₄ or SER→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, the wwPDB 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 PDB data as required. The coordinates themselves are never 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 is 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 and recorded in _pdbx_version mmCIF category in every file. The older version is maintained as a snapshot on the FTP site.