

# **Protein Data Bank Contents Guide: Atomic Coordinate Entry Format Description**

**Version 2.0, February 21, 1996**

# Protein Data Bank Contents Guide:

## Atomic Coordinate Entry Format Description

Version 2.0, February 21, 1996

---

---

### Preface

The Protein Data Bank (PDB) is an archive of experimentally determined three-dimensional structures of biological macromolecules, serving a global community of researchers, educators, and students. The archives contain atomic coordinates, bibliographic citations, primary and secondary structure information, as well as crystallographic structure factors and NMR experimental data.

Entries conforming to this format description have the following remark within them:

```
REMARK    4 XXXX COMPLIES WITH FORMAT V. 2.0, 21-FEB-1996
```

Entries released after April 15, 1996 will comply with this format. Conversion of older entries to this format will begin in the fall of 1996.

This Contents Guide was prepared through the efforts of all PDB staff members: J. Callaway, M. Cummings, B. Deroski, P. Esposito, A. Forman, P. Langdon, M. Libeson, J. McCarthy, J. Sikora, D. Xue; and especially E. Abola, F. Bernstein, N. Manning, R. Shea, D. Stampf, and J. Sussman. This document also included significant contributions from the scientific community whose members continually send us suggestions and comments regarding the contents and format of PDB entries.

Please send any comments or suggestions on this Contents Guide to the PDB Help Desk.

Protein Data Bank Help Desk  
Chemistry Department, Bldg. 555  
Brookhaven National Laboratory  
P.O. Box 5000  
Upton, NY 11973-5000 USA

E-mail: [pdhelp@bnl.gov](mailto:pdhelp@bnl.gov)  
World Wide Web: <http://www.pdb.bnl.gov>  
TEL: 1 516-344-3629  
FAX: 1 516-344-5751

The PDB is supported by a combination of Federal Government Agency funds and user fees. Support is provided by the U.S. National Science Foundation, the U.S. Public Health Service, National Institutes of Health, National Center for Research Resources, National Institutes of General Medical Sciences, National Library of Medicine, and the U.S. Department of Energy under contract DE-AC02-76CH00016.

---

---

# Table of Contents

Preface	1
Table of Contents	2
1. Introduction	5
Purpose of this Document	5
What's New	5
Changes to PDB Format and to the Contents Guide	6
Basic Notions of the Format Description	7
Record Format	8
Types of Records	8
Order of Records	11
Field Formats	14
2. Title Section	16
HEADER	16
OBSLTE	17
TITLE	18
CAVEAT	19
COMPND	19
SOURCE	22
KEYWDS	25
EXPDTA	26
AUTHOR	27
REVDAT	29
SPRSDE	30
JRNL	31
REMARK	38
REMARK 1	38
REMARK 2	43
REMARK 3	44
REMARK 4 - 999	56
3. Primary Structure Section	70
MODRES	70
DBREF	71
SEQADV	74
SEQRES	75

4. Heterogen Section	78
HET	78
HETNAM	79
HETSYN	80
FORMUL	81
5. Secondary Structure Section	83
HELIX	83
SHEET	84
TURN	87
6. Connectivity Annotation Section	89
SSBOND	89
LINK	90
HYDBND	91
SLTBRG	93
CISPEP	94
7. Miscellaneous Features Section	96
SITE	96
8. Crystallographic and Coordinate Transformation Section	98
CRYST1	98
ORIGXn	99
SCALEn	100
MTRIXn	101
TVECT	102
9. Coordinate Section	104
MODEL	104
ATOM	105
SIGATM	109
ANISOU	110
SIGUIJ	112
TER	112
HETATM	114
ENDMDL	116
10. Connectivity Section	118
CONNECT	118

11. Bookkeeping Section	120
MASTER	120
END	121
Appendix 1: Symmetry Operations	122
Appendix 2: Coordinate Systems and Transformations	124
Appendix 3: Atom Names	125
Amino Acids	125
Nucleic Acids	125
Appendix 4: Standard Residue Names and Abbreviations	126
Amino Acids	126
Nucleic Acids	126
Appendix 5: Formulas and Molecular Weights For Standard Residues	127
Amino Acids	127
Nucleotides	127
Miscellaneous	
Appendix 6: Field Formats	128
Appendix 7: Order of Records	130

---

# 1. Introduction

## Purpose of this Document

The PDB Contents Guide gives a complete and concise description of the contents of PDB coordinate entry files. This document will be helpful to several communities, assisting depositors in preparing their entries for deposition, guiding software and information resource developers, and helping users of PDB to understand the contents of coordinate entries. Finally, this format description is crucial in the effort to produce CIF-compliant data files from PDB entries.

## What's New

List of important changes/enhancements to PDB format as found in Contents Guide Version 2.0.

### Changes to ATOM/HETATM Records

- \* Columns 71 - 80 now contain data. They previously contained the PDB ID code and record serial number. These items may be generated using scripts available from the PDB.
- \* A segment identifier has been added to the coordinate records in columns 73 - 76. This allows unambiguous identification of regions of the chains and the relationship between them by specifying segments of molecules.
- \* The element symbol and charge now appear in columns 77 - 80 of the coordinate records.
- \* When temperature factors are provided, the tempFactor field (columns 61 - 66) always contains the isotropic B value, even when ANISOU records are provided.
- \* Insertion codes (column 27) are now defined as being alphabetic only.

### Changes to Other Records

- \* HELIX records now contain the length of the helix in columns 72 - 76.
- \* Footnotes (FTNOTE) have been dropped.
- \* In CRYST1 records:
  - The full international Hermann-Mauguin symbol is used, e.g., P 1 21 1 instead of P 21.
  - For a rhombohedral space group in the hexagonal setting, the lattice type symbol used is H.
- \* A number of record types which previously contained free text have been restructured as follows:
  - "Keyword: value" pairs have been introduced in certain records such as COMPND and SOURCE to allow easier parsing.
  - EXPDTA has been expanded and now appears in every PDB coordinate entry.
  - REMARK records have been restructured to allow easier parsing and to bring more organization to these records.

### New Record Types Added

- \* TITLE
- \* CAVEAT
- \* KEYWDS
- \* MODRES
- \* DBREF

- \* SEQADV
- \* HETNAM
- \* HETSYN
- \* LINK
- \* HYDBND
- \* SLTBRG
- \* CISPEP

For details on each of these changes, see the section of the associated record type in this document.

## Changes to PDB Format and to the Contents Guide

When a change is made to PDB format, the format version number, as found in the entry and in this Contents Guide, will be incremented to the next whole number. Changes to the format of PDB coordinate entry files will follow the Format Change Policy presented below and will be detailed in this Contents Guide. Beginning January 1997, the format of all PDB entries will be compliant with the current version of this Contents Guide.

Changes to the Contents Guide will be listed at the beginning in the What's New section and denoted by a fractional increase in the document version number. These changes may be of the following kind.

- \* Correction of typographical errors.
- \* Changes to the language for clarity.
- \* Addition or changes to the examples for better representation of format issues.
- \* Addition of new rules (these do not change the format but help to clarify the semantics).
- \* Addition of tokens to specification lists, such as in COMPND and SOURCE records, that are needed to more fully describe the structure and its biological source.
- \* Enhancements to the refinement and experimental details templates in the REMARK records. These remarks are currently being reviewed by several people in the community, and PDB expects to increase the level of detail archived, such as for NMR studies.
- \* Addition of new sections that enhance and expand the document (these may include topics such as PDB to mmCIF cross references or insertion of relevant sections from the PDB Deposition Form).

## Format Change Policy

The PDB will use the following protocol in making changes to the way PDB coordinate entries are represented and archived. The purpose of the new policy is to allow ample time for everyone to understand these changes and to assess their impact on existing programs. These modifications are necessary to address the changing needs of our users as well as the changing nature of the data that is archived.

1. Comments and suggestions will be solicited from the community on specific problems and data representation issues as they arise.
2. Proposed format changes will be disseminated through the PDB Listserver ([pdb-l@pdb.bnl.gov](mailto:pdb-l@pdb.bnl.gov)) and PDB's Internet sites (WWW, FTP, and Gopher). They will also be summarized in the PDB Quarterly Newsletter.
3. A sixty-day discussion period will follow the announcement of proposed changes. Comments and suggestions must be received within this time period. Major changes which are not upwardly compatible will be allotted up to twice the standard amount of discussion time.
4. This sixty-day discussion period will be followed by a thirty-day period in which the PDB staff, the PDB Advisory Board, and the User Group Chair will evaluate and reconcile all suggestions. The final

decision pertaining to the format change, which lies with the Advisory Board Chair, will then be officially announced via the PDB Listserver and PDB's Internet sites (WWW, FTP, and Gopher).

5. Implementation will follow official announcement of the format change. Major changes will not appear in PDB files earlier than sixty days after the announcement, allowing sufficient time to modify files and programs.

6. Changes will be released no more than twice a year, unless extraordinary circumstances require action. This will be done only in consultation with the Advisory Board and following the usual ninety-day discussion and evaluation period.

The PDB format has been in use since the late 1970's. A number of groups including the mmCIF Committee have been looking at ways to upgrade both the file content and the interchange format used by PDB. This is clearly needed due to changes in the data that PDB archives, the size of the database itself, and finally, to allow PDB to use more up-to-date methods for representing and storing biological data.

The PDB plans to be prudent and deliberate in making changes to the current PDB files in order to minimize the need to change existing programs. In particular, we will explore ways and means of ensuring that programs which read the current ATOM/HETATM records can continue to do so in the foreseeable future.

The PDB wishes to acknowledge Dr. Gerald Selzer of the National Science Foundation who urged us to formulate this policy.

## Basic Notions of the Format Description

### Character Set

Only non-control ASCII characters, as well as the space and end-of-line indicator, appear in a PDB coordinate entry file. Namely:

abcdefghijklmnopqrstuvwxyzABCDEFGHIJKLMNOPQRSTUVWXYZ

1234567890

` - = [ ] \ ; ' , . / ~ ! @ # \$ % ^ & \* ( ) \_ + { } | : " < > ?

the space, and end-of-line. The end-of-line indicator is system-specific. Unix uses a line feed character; other systems may use a carriage return followed by a line feed.

### Special Characters

Greek letters are spelled out, i.e., alpha, beta, gamma, etc.

Bullets are represented as (DOT).

Right arrow is represented as -->.

Left arrow is represented as <--.

Superscripts are initiated and terminated by double equal signs, e.g., S==2+==.

Subscripts are initiated and terminated by single equal signs, e.g., F=c=.

If "=" is surrounded by at least one space on each side, then it is assumed to be an equal sign, e.g., 2 + 4 = 6.

Commas and semi-colons are used as list delimiters in certain record types. Any occurrence of either of these characters in another context within fields where it is used as a list delimiter must be preceded with a "\".



## Record Format

Every PDB file may be broken into a number of lines terminated by an end-of-line indicator. Each line in the PDB entry file consists of 80 columns. The last character in each PDB entry should be an end-of-line indicator.

Each line in the PDB file is self-identifying. The first six columns of every line contain a record name, left-justified and blank-filled. This must be an exact match to one of the stated record names.

The PDB file may also be viewed as a collection of record types. Each record type consists of one or more lines.

Each record type is further divided into fields.

Each record type is detailed in this document. The description of each record type includes the following sections:

- \* Overview
- \* Record Format
- \* Details
- \* Verification/Validation/Value Authority Control
- \* Relationship to Other Record Types
- \* Example
- \* Known Problems

For records that are fully described in fixed column format, columns not assigned to fields *must be left blank*.

## Types of Records

It is possible to group records into categories based upon how often the record type appears in an entry.

### Single

There are records which may only appear one time (without continuations) in a file. Listed alphabetically, these are:

RECORD TYPE	DESCRIPTION
CRYST1	Unit cell parameters, space group, and Z.
END	Last record in the file.
HEADER	First line of the entry, contains PDB ID code, classification, and date of deposition.
MASTER	Control record for bookkeeping.
ORIGXn	Transformation from orthogonal coordinates to the submitted coordinates (n = 1, 2, or 3).
SCALEn	Transformation from orthogonal coordinates to fractional crystallographic coordinates (n = 1, 2, or 3).

It is an error for a duplicate of any of these records to appear in an entry.

**Single Continued**

There are records that conceptually exist only once in an entry, but the information content may exceed the number of columns available. These records are therefore continued on subsequent lines. Listed alphabetically, these are:

RECORD TYPE	DESCRIPTION
AUTHOR	List of contributors.
CAVEAT	Severe error indicator. Entries with this record must be used with care.
COMPND	Description of macromolecular contents of the entry.
EXPDTA	Experimental technique used for the structure determination.
KEYWDS	List of keywords describing the macromolecule.
OBSLTE	Statement that the entry has been removed from distribution and list of the ID code(s) which replaced it.
SOURCE	Biological source of macromolecules in the entry.
SPRSDE	List of entries withdrawn from release and replaced by current entry.
TITLE	Description of the experiment represented in the entry.

The second and subsequent lines contain a continuation field which is a right-justified integer. This number increments by one for each additional line of the record, and is followed by a blank character.

**Multiple**

Most record types appear multiple times, often in groups where the information is not logically concatenated but is presented in the form of a list. Many of these record types have a custom serialization that may be used not only to order the records, but also to connect to other record types. Listed alphabetically, these are:

RECORD TYPE	DESCRIPTION
ANISOU	Anisotropic temperature factors.
ATOM	Atomic coordinate records for standard groups.
CISPEP	Identification of peptide residues in cis conformation.
CONNECT	Connectivity records.
DBREF	Reference to the entry in the sequence database(s).
HELIX	Identification of helical substructures.
HET	Identification of non-standard groups or residues (heterogens)
HETSYN	Synonymous compound names for heterogens.
HYDBND	Identification of hydrogen bonds.
LINK	Identification of inter-residue bonds.

MODRES	Identification of modifications to standard residues.
MTRIXn	Transformations expressing non-crystallographic symmetry (n = 1, 2, or 3). There may be multiple sets of these records.
REVDAT	Revision date and related information.
SEQADV	Identification of conflicts between PDB and the named sequence database.
SEQRES	Primary sequence of backbone residues.
SHEET	Identification of sheet substructures.
SIGATM	Standard deviations of atomic parameters.
SIGUIJ	Standard deviations of anisotropic temperature factors.
SITE	Identification of groups comprising important sites.
SLTBRG	Identification of salt bridges
SSBOND	Identification of disulfide bonds.
TURN	Identification of turns.
TVECT	Translation vector for infinite covalently connected structures.

### Multiple Continued

There are records that conceptually exist multiple times in an entry, but the information content may exceed the number of columns available. These records are therefore continued on subsequent lines. Listed alphabetically, these are:

FORMUL	Chemical formula of non-standard groups.
HETATM	Atomic coordinate records for heterogens.
HETNAM	Compound name of the heterogens.

The second and subsequent lines contain a continuation field which is a right-justified integer. This number increments by one for each additional line of the record, and is followed by a blank character.

### Grouping

There are three record types used to group other records. Listed alphabetically, these are:

RECORD TYPE	DESCRIPTION
ENDMDL	End-of-model record for multiple structures in a single coordinate entry.
MODEL	Specification of model number for multiple structures in a single coordinate entry.
TER	Chain terminator.

The MODEL/ENDMDL records surround groups of ATOM, HETATM, SIGATM, ANISOU, SIGUIJ, and TER records. TER records indicate the end of a chain.

### Other

The remaining record types have a detailed inner structure. Listed alphabetically, these are:

RECORD TYPE	DESCRIPTION
JRNL	Literature citation that defines the coordinate set.
REMARK	General remarks, some are structured and some are free form.

## Order of Records

All records in a PDB coordinate entry must appear in a defined order. Mandatory record types are present in all entries. When mandatory data are not provided, the record name must appear in the entry with a NULL indicator. Optional items become mandatory when certain conditions exist. Record order and existence are described in the following table:

RECORD TYPE	EXISTENCE	CONDITIONS IF OPTIONAL
HEADER	Mandatory	
OBSLTE	Optional	Mandatory in withdrawn entries.
TITLE	Mandatory	
CAVEAT	Optional	Mandatory if structure is deemed incorrect by an outside editorial board.
COMPND	Mandatory	
SOURCE	Mandatory	
KEYWDS	Mandatory	
EXPDTA	Mandatory	
AUTHOR	Mandatory	
REVDAT	Mandatory	
SPRSDE	Optional	Mandatory if a replacement entry.
JRNL	Optional	Mandatory if a publication describes the experiment.
REMARK 1	Optional	
REMARK 2	Mandatory	
REMARK 3	Mandatory	
REMARK N	Optional	
MODRES	Optional	Mandatory if modified group exists within the coordinates.

DBREF	Mandatory	Mandatory for each peptide chain with a length greater than ten (10) residues, and for nucleic acid entries that exist in the Nucleic Acid Database (NDB).
SEQADV	Optional	Mandatory if sequence conflict exists.
SEQRES	Optional	Mandatory if ATOM records exist.
HET	Optional	Mandatory if non-standard group other than water appears in the entry.
HETNAM	Optional	Mandatory if non-standard group other than water appears appears in the entry.
HETSYN	Optional	
FORMUL	Optional	Mandatory if non-standard group or water appears.
HELIX	Optional	
SHEET	Optional	
TURN	Optional	
SSBOND	Optional	Mandatory if disulfide bond is present.
LINK	Optional	
HYDBND	Optional	
SLTBRG	Optional	
CISPEP	Optional	
SITE	Optional	
CRYST1	Mandatory	
ORIGX1 ORIGX2 ORIGX3	Mandatory	
SCALE1 SCALE2 SCALE3	Mandatory	
MTRIX1 MTRIX2 MTRIX3	Optional	Mandatory if the complete asymmetric unit must be generated from the given coordinates using non-crystallographic symmetry.
TVECT	Optional	
MODEL	Optional	Mandatory if more than one model is present in the entry.
ATOM	Optional	Mandatory if standard residues exist.
SIGATM	Optional	
ANISOU	Optional	

SIGUIJ	Optional	
TER	Optional	Mandatory if ATOM records exist.
HETATM	Optional	Mandatory if non-standard group appears.
ENDMDL	Optional	Mandatory if MODEL appears.
CONECT	Optional	Mandatory if non-standard group appears.
MASTER	Mandatory	
END	Mandatory	

Note that a PDB file existing outside of the PDB official release may contain locally-defined records beginning with "USER". The PDB reserves the right to add new record types (not beginning with "USER"), so programs which read PDB entries should be prepared to read (and ignore) other record types. PDB will follow standard procedures whenever format changes are proposed.

### Sections of an Entry

The following table lists the various sections of a PDB coordinate entry and the records comprising them:

SECTION	DESCRIPTION	RECORD TYPE
Title	Summary descriptive remarks	HEADER, OBSLTE, TITLE, CAVEAT, COMPND, SOURCE, KEYWDS, EXPDTA, AUTHOR, REVDAT, SPRSDE, JRNL
Remark	Bibliography, refinement, annotations	REMARKs 1, 2, 3 and others
Primary structure	Peptide and/or nucleotide sequence and the relationship between the PDB sequence and that found in the sequence database(s)	MODRES, DBREF, SEQADV, SEQRES
Heterogen	Description of non-standard groups	HET, HETNAM, HETSYN, FORMUL
Secondary structure	Description of secondary structure	HELIX, SHEET, TURN
Connectivity annotation	Chemical connectivity	SSBOND, LINK, HYDBND, SLTBRG, CISPEP
Miscellaneous features	Features within the macromolecule	SITE
Crystallographic	Description of the crystallographic cell	CRYST1
Coordinate transformation	Coordinate transformation operators	ORIGXn, SCALEn, MTRIXn, TVECT

Coordinate	Atomic coordinate data	MODEL, ATOM, SIGATM, ANISOU, SIGUIJ, TER, HETATM, ENDMDL
Connectivity	Chemical connectivity	CONNECT
Bookkeeping	Summary information, ~ end-of-file marker	MASTER, END

The above information on Order of Records is repeated as Appendix 7.

## Field Formats

Each record type is presented in a table which contains the division of the records into fields by column number, defined data type, field name or a quoted string which must appear in the field, and field definition. Any column not specified must be left blank.

Each field contains an identified data type which can be validated by a program. These are:

DATA TYPE	DESCRIPTION
-----	-----
AChar	An alphabetic character (A-Z, a-z).
Atom	Atom name which follow the naming rules in Appendix 3.
Character	Any non-control character in the ASCII character set or a space.
Continuation	A two-character field that is either blank (for the first record of a set) or contains a two digit number right-justified and blank-filled which counts continuation records starting with 2. The continuation number must be followed by a blank.
Date	A 9 character string in the form dd-mmm-yy where DD is the day of the month, zero-filled on the left (e.g., 04); MMM is the common English 3-letter abbreviation of the month; and YY is a year in the 20th century. This must represent a valid date.
IDcode	A PDB identification code which consists of 4 characters, the first of which is a digit in the range 0 - 9; the remaining 3 are alpha-numeric, and letters are upper case only. Entries with a 0 as the first character do not contain coordinate data.
Integer	Right-justified blank-filled integer value.
Token	A sequence of non-space characters followed by a colon and a space.
List	A String that is composed of text separated with commas.
LString	A literal string of characters. All spacing is significant and must be preserved.
LString(n)	An LString with exactly n characters.

Real(n,m)	Real (floating point) number in the FORTRAN format Fn.m.
Record name	The name of the record: 6 characters, left-justified and blank-filled.
Residue name	One of the standard amino acid or nucleic acids, as listed below, or the non-standard group designation as defined in the HET dictionary. Field is right-justified.
SList	A String that is composed of text separated with semi-colons.
Specification	A String composed of a token and its associated value separated by a colon.
Specification list	A sequence of Specifications, separated by semi-colons.
String	A sequence of characters. These characters may have arbitrary spacing, but should be interpreted as directed below.
String(n)	A String with exactly n characters.
SymOP	An integer field of from 4 to 6 digits, right-justified, of the form nnnMMM where nnn is the symmetry operator number and MMM is the translation vector. See details in Appendix 1.

To interpret a String, concatenate the contents of all continued fields together, collapse all sequences of multiple blanks to a single blank, and remove any leading and trailing blanks. This permits very long strings to be properly reconstructed.

The above information about field formats is repeated as Appendix 6.

### Residue Names

Standard residue names used in PDB entries:

RESIDUE TYPE	RESIDUE NAME
Amino acids	ALA, ARG, ASN, ASP, CYS, GLN, GLU, GLY, HIS, ILE, LEU, LYS, MET, PHE, PRO, SER, THR, TRP, TYR, VAL, ASX, GLX
Nucleic acids	A, C, G, T, U, I, +A, +C, +G, +T, +U, +I
Other	UNK (unknown)

See Appendix 4 for more information on the standard residue names and abbreviations, and Appendix 5 for their chemical formulas and molecular weights.



## 2. Title Section

This section contains records used to describe the experiment and the biological macromolecules present in the entry: HEADER, OBLSTE, TITLE, CAVEAT, COMPND, SOURCE, KEYWDS, EXPDTA, AUTHOR, REVDAT, SPRSDE, JRNL, and REMARK records.

---

### HEADER

#### Overview

The HEADER record uniquely identifies a PDB entry through the idCode field. This record also provides a classification for the entry. Finally, it contains the date the coordinates were deposited at the PDB.

#### Record Format

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"HEADER"	
11 - 50	String(40)	classification	Classifies the molecule(s)
51 - 59	Date	depDate	Deposition date. This is the date the coordinates were received by the PDB
63 - 66	IDcode	idCode	This identifier is unique within PDB

#### Details

\* The classification string is left-justified and exactly matches one of a collection of strings. See the class list available from the WWW site. In the case of macromolecular complexes, the classification field must present a class for each macromolecule present. Due to the limited length of the classification field, strings must sometimes be abbreviated. In these cases, the full terms are given in KEYWDS.

\* Classification may be based on function, metabolic role, molecule type, cellular location, etc. In the case of a molecule having a dual function, both may be presented here. A list of valid terms that may be used as the classification appears on PDB's Web server.

#### Verification/Validation/Value Authority Control

The verification program checks that the deposition date is a legitimate date and that the ID code is well-formed. PDB coordinate entry ID codes do not begin with 0, as this is used to identify the NOC files which are bibliographic only, not structural entries. The status and deposition date of an entry are checked against the PDB SYBASE tables, which provide a definitive list of existing ID codes.

#### Relationships to Other Record Types

The KEYWDS record also provides terms that may be used to classify the entry. Even though terms must sometimes be abbreviated in the HEADER record, none are abbreviated in KEYWDS.

**Example**

	1	2	3	4	5	6	7
1234567890123456789012345678901234567890123456789012345678901234567890							
HEADER	MUSCLE PROTEIN				02-JUN-93		1MYS
HEADER	HYDROLASE (CARBOXYLIC ESTER)				08-APR-93		2PHI
HEADER	COMPLEX (LECTIN/TRANSFERRIN)				07-JAN-94		1LGB

**OBSLTE****Overview**

OBSLTE appears in entries which have been withdrawn from distribution.

This record acts as a flag in an entry which has been withdrawn from the PDB's full release. It indicates which, if any, new entries have replaced the withdrawn entry.

The format allows for the case of multiple new entries replacing one existing entry.

**Record Format**

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"OBSLTE"	
9 - 10	Continuation	continuation	Allows concatenation of multiple records.
12 - 20	Date	repDate	Date that this entry was replaced.
22 - 25	IDcode	idCode	ID code of this entry.
32 - 35	IDcode	rIdCode	ID code of entry that replaced this one.
37 - 40	IDcode	rIdCode	ID code of entry that replaced this one.
42 - 45	IDcode	rIdCode	ID code of entry that replaced this one.
47 - 50	IDcode	rIdCode	ID code of entry that replaced this one.
52 - 55	IDcode	rIdCode	ID code of entry that replaced this one.
57 - 60	IDcode	rIdCode	ID code of entry that replaced this one.
62 - 65	IDcode	rIdCode	ID code of entry that replaced this one.
67 - 70	IDcode	rIdCode	ID code of entry that replaced this one.

**Details**

\* It is PDB policy that only the primary author who submitted an entry has the authority to withdraw it. All withdrawn entries are available for research purposes. PDB should be contacted in cases where the withdrawn data are desired.

**Verification/Validation/Value Authority Control**

PDB staff add this record at the time an entry is removed from release.

**Relationships to Other Record Types**

None.

**Example**

```

      1          2          3          4          5          6          7
123456789012345678901234567890123456789012345678901234567890
OBSLTE      31-JAN-94 1MBP      2MBP

```

---

**TITLE****Overview**

The TITLE record contains a title for the experiment or analysis that is represented in the entry. It should identify an entry in the PDB in the same way that a title identifies a paper.

**Record Format**

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"TITLE "	
9 - 10	Continuation	continuation	Allows concatenation of multiple records.
11 - 70	String	title	Title of the experiment.

**Details**

\* The title of the entry is free text and should describe the contents of the entry and any procedures or conditions that distinguish this entry from similar entries. It presents an opportunity for the depositor to emphasize the underlying purpose of this particular experiment.

\* Some items that may be included in TITLE are:

- Experiment type.
- Description of the mutation.
- The fact that only alpha carbon coordinates have been provided in the entry.

**Verification/Validation/Value Authority Control**

This record is free text so no verification of format is required. The title is supplied by the depositor, but PDB staff may exercise editorial judgment in consultation with depositors in assigning the title.

**Relationships to Other Record Types**

COMPND, SOURCE, EXPDTA, and REMARKs provide information that may also be found in TITLE. You may think of the title as describing the experiment, and the compound record as describing the molecule(s).

**Example**

```

1      2      3      4      5      6      7
123456789012345678901234567890123456789012345678901234567890
TITLE      RHIZOPUSPEPSIN COMPLEXED WITH REDUCED PEPTIDE INHIBITOR

TITLE      BETA-GLUCOSYLTRANSFERASE, ALPHA CARBON COORDINATES ONLY

TITLE      NMR STUDY OF OXIDIZED THIOREDOXIN MUTANT (C62A,C69A,C73A)
TITLE      2 MINIMIZED AVERAGE STRUCTURE

```

---

**CAVEAT****Overview**

CAVEAT warns of severe errors in an entry. Use caution when using an entry containing this record.

**Record Format**

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"CAVEAT"	
9 - 10	Continuation	continuation	Allows concatenation of multiple records.
12 - 15	IDcode	idCode	PDB ID code of this entry.
20 - 70	String	comment	Free text giving the reason for the CAVEAT.

**Details**

\* PDB will add this record to incorrect entries that are not withdrawn from the set of released entries. This record will be used sparingly, and only after an external review has been made.

\* Please note the CAVEAT will also be included in cases where PDB is unable to verify the transformation back to the crystallographic cell. In these cases, the molecular structure may still be correct.

**Verification/Validation/Value Authority Control**

CAVEAT will be added by the PDB to entries known to be incorrect.

**Relationships to Other Record Types**

REMARK 5 repeats the comment field of the CAVEAT record.

**Example**

```

1      2      3      4      5      6      7
123456789012345678901234567890123456789012345678901234567890
CAVEAT      1ABC      THE CRYSTAL TRANSFORMATION IS IN ERROR BUT IS
CAVEAT      2 1ABC      UNCORRECTABLE AT THIS TIME

```

---

**COMPND****Overview**

The COMPND record describes the macromolecular contents of an entry. Each macromolecule found in the entry is described by a set of token: value pairs, and is referred to as a COMPND record component. Since the concept of

a molecule is difficult to specify exactly, PDB staff may exercise editorial judgement in consultation with depositors in assigning these names.

For each macromolecular component, the molecule name, synonyms, number assigned by the Enzyme Commission (EC), and other relevant details are specified.

### Record Format

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"COMPND"	
9 - 10	Continuation	continuation	Allows concatenation of multiple records.
11 - 70	Specification list	compound	Description of the molecular components.

### Details

\* The compound record is a Specification list. The specifications, or tokens, that may be used are listed below:

TOKEN	VALUE DEFINITION
MOL_ID	Numbers each component; also used in SOURCE to associate the information.
MOLECULE	Name of the macromolecule.
CHAIN	Comma-separated list of chain identifier(s). "NULL" is used to indicate a blank chain identifier.
FRAGMENT	Specifies a domain or region of the molecule.
SYNONYM	Comma-separated list of synonyms for the MOLECULE.
EC	The Enzyme Commission number associated with the molecule. If there is more than one EC number, they are presented as a comma-separated list.
ENGINEERED	Indicates that the molecule was produced using recombinant technology or by purely chemical synthesis.
MUTATION	Describes mutations from the wild type molecule.
BIOLOGICAL_UNIT	If the MOLECULE functions as part of a larger biological unit, the entire functional unit may be described.
OTHER_DETAILS	Additional comments.

\* In the general case the PDB tends to reflect the biological/functional view of the molecule. For example, the hetero-tetramer hemoglobin molecule is treated as a discrete component in COMPND.

\* In the case of synthetic molecules, e. g., hybrids, the description will be provided by the depositor.

\* No specific rules apply to the ordering of the tokens, except that the occurrence of MOL\_ID or FRAGMENT indicates that the subsequent tokens are related to that specific molecule or fragment of the molecule.

\* Physical layout of these items may be altered by PDB staff to improve human readability of the COMPND record.

\* Asterisks in nucleic acid names (in MOLECULE) are for ease of reading.

- \* When insertion codes are given as part of the residue name, they must be given within square brackets, i.e., H57[A]N. This might occur when listing residues in FRAGMENT, MUTATION, or OTHER\_DETAILS.
- \* For multi-chain molecules, e.g., the hemoglobin tetramer, a comma-separated list of CHAIN identifiers is used.
- \* When non-blank chain identifiers occur in the entry, they must be specified.
- \* NULL is used to indicate blank chain identifiers. E.g., CHAIN: NULL, CHAIN: NULL, B, C.
- \* For enzymes, if no EC number has been assigned, "EC: NOT ASSIGNED" is used.
- \* ENGINEERED is followed either by "YES" or by a comment.
- \* For the token MUTATION, the following set of examples illustrate the conventions used by PDB to represent various types of mutations.

MUTATION TYPE	DESCRIPTION	FORM
Simple substitution	His 57 replaced by Asn	H57N
	His 57A replaced by Asn, in chain C only	Chain C, H57[A]N
Insertion	His and Pro inserted before Lys 48	INS(HP-K48)
Deletion	Arg 141 of chains A and C deleted, not deleted in chain B	Chain A, C, DEL(R141)
	His 23 through ARG 26 deleted	DEL(23-26)
	His 23C and Arg 26 deleted from chain B only	Chain B, DEL(H23[C],R26)

- \* When there are more than ten mutations:
  - All the mutations are listed in the SEQADV record.
  - Some mutations may be listed in MUTATION in COMPND to highlight the most important ones, at the depositor's discretion.
- \* New tokens may be added by the PDB as needed.

#### Verification/Validation/Value Authority Control

CHAIN must match the chain identifiers(s) of the molecule(s). EC numbers are checked against the Enzyme Data Bank.

#### Relationships to Other Record Types

Each molecule given a MOL\_ID in COMPND must be listed and given the biological source information in SOURCE. In the case of mutations, the SEQADV records will present differences from the reference molecule. REMARK record may further describe the contents of the entry. Also see verification above.

**Example**

```

123456789012345678901234567890123456789012345678901234567890
COMPND      1  MOL_ID: 1;
COMPND      2  MOLECULE: HEMOGLOBIN;
COMPND      3  CHAIN: A, B, C, D;
COMPND      4  ENGINEERED: YES;
COMPND      5  MUTATION: CHAIN B, D, V1A;
COMPND      6  BIOLOGICAL_UNIT: HEMOGLOBIN EXISTS AS AN A1B1/A2B2 TETRAMER;
COMPND      7  OTHER_DETAILS: DEOXY FORM

COMPND      1  MOL_ID: 1;
COMPND      2  MOLECULE: COWPEA CHLOROTIC MOTTLE VIRUS;
COMPND      3  CHAIN: A, B, C;
COMPND      4  SYNONYM: CCMV;
COMPND      5  MOL_ID: 2;
COMPND      6  MOLECULE: RNA (5'-( *AP*UP*AP*U)-3' );
COMPND      7  CHAIN: D, F;
COMPND      8  ENGINEERED: YES;
COMPND      9  MOL_ID: 3;
COMPND     10  MOLECULE: RNA (5'-( *AP*U)-3' );
COMPND     11  CHAIN: E;
COMPND     12  ENGINEERED: YES

COMPND      1  MOL_ID: 1;
COMPND      2  MOLECULE: HEVAMINE A;
COMPND      3  CHAIN: NULL;
COMPND      4  EC: 3.2.1.14, 3.2.1.17;
COMPND      5  OTHER_DETAILS: PLANT.ENDOCHITINASE/LYSOZYME

```

---

**SOURCE****Overview**

The SOURCE record specifies the biological and/or chemical source of each biological molecule in the entry. Sources are described by both the common name and the scientific name, e.g., genus and species. Strain and/or cell-line for immortalized cells are given when they help to uniquely identify the biological entity studied.

**Record Format**

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"SOURCE"	
9 - 10	Continuation	continuation	Allows concatenation of multiple records.
11 - 70	Specification list	srcName	Identifies the source of the macromolecule in a token: value format.

**Details**

TOKEN	VALUE DEFINITION
MOL_ID	Numbers each molecule. Same as appears in COMPND.
SYNTHETIC	Indicates a chemically-synthesized source.
FRAGMENT	A domain or fragment of the molecule may be specified.
ORGANISM_SCIENTIFIC	Scientific name of the organism.
ORGANISM_COMMON	Common name of the organism.
STRAIN	Identifies the strain.
VARIANT	Identifies the variant.
CELL_LINE	The specific line of cells used in the experiment.
ATCC	American Type Culture Collection tissue culture number.
ORGAN	Organized group of tissues that carries on a specialized function.
TISSUE	Organized group of cells with a common function and structure.
CELL	Identifies the particular cell type.
ORGANELLE	Organized structure within a cell.
CELLULAR_LOCATION	Identifies the location inside (or outside) the cell.
PLASMID	Identifies the plasmid containing the gene.
GENE	Identifies the gene.
EXPRESSION_SYSTEM	System used to produce recombinant macromolecules.
EXPRESSION_SYSTEM_STRAIN	Strain of the organism in which the molecule was expressed.
EXPRESSION_SYSTEM_PLASMID	Plasmid used in the recombinant experiment.
EXPRESSION_SYSTEM_GENE	Name of the gene used in recombinant experiment.



- \* The srcName is a list of token: value pairs describing each biological component of the entry.
- \* As in COMPND, the order is not specified except that MOL\_ID or FRAGMENT indicates subsequent specifications are related to that molecule or fragment of the molecule.
- \* Physical layout of these items may be altered by PDB staff to improve human readability of the SOURCE record.
- \* Molecules prepared by purely chemical synthetic methods are described by the specification SYNTHETIC followed by "YES" or an optional value, such as NON-BIOLOGICAL SOURCE or BASED ON THE NATURAL SEQUENCE.
- \* In the case of a chemically synthesized molecule using a biologically functional sequence (nucleic or amino acid), SOURCE reflects the biological origin of the sequence and COMPND reflects its synthetic nature by inclusion of the token ENGINEERED.
- \* When multiple macromolecules appear in the entry, each MOL\_ID, as given in the COMPND record, must be repeated in the SOURCE record along with the source information for the corresponding molecule.
- \* Hybrid molecules prepared by fusion of genes are treated as multi-molecular systems for the purpose of specifying the source. The token FRAGMENT is used to associate the source with its corresponding fragment.
  - When necessary to fully describe hybrid molecules, tokens may appear more than once for a given MOL\_ID.
  - All relevant token: value pairs that taken together fully describe each fragment are grouped following the appropriate FRAGMENT.
  - Descriptors relative to the full system appear before the FRAGMENT (see Example 3 below).
- \* Cellular origin is described by giving cellular compartment, organelle, cell, tissue, organ, or body part from which the molecule was isolated.
- \* CELLULAR\_LOCATION may be used to indicate where in the organism the compound was found. Examples are: extracellular, periplasmic, cytosol.
- \* Entries containing molecules prepared by recombinant techniques are described as follows:
  - The expression system is described.
  - The organism and cell location given are for the source of the gene used in the cloning experiment.
  - Transgenic organisms, such as mouse producing human proteins, are treated as expression systems.
- \* New tokens may be added by the PDB.

#### Verification/Validation/Value Authority Control

The biological source is compared to that found in the sequence database. Common and scientific names are checked against the "Annotated Classification of Source Organisms: PIR-International Protein Sequence Database" compiled by Andrzej Elzanowski and available from the PDB.

#### Relationships to Other Record Types

Each macromolecule listed in COMPND must have a corresponding source.

**Example**

```

123456789012345678901234567890123456789012345678901234567890
SOURCE      1  MOL_ID: 1;
SOURCE      2  ORGANISM_SCIENTIFIC: AVIAN SARCOMA VIRUS;
SOURCE      3  STRAIN: SCHMIDT-RUPPIN B;
SOURCE      4  EXPRESSION_SYSTEM: ESCHERICHIA COLI;
SOURCE      5  EXPRESSION_SYSTEM_PLASMID: PRC23IN
SOURCE      1  MOL_ID: 1;
SOURCE      2  ORGANISM_SCIENTIFIC: GALLUS GALLUS;
SOURCE      3  ORGANISM_COMMON: CHICKEN;
SOURCE      4  ORGAN: HEART;
SOURCE      5  TISSUE: MUSCLE
SOURCE      1  MOL_ID: 1;
SOURCE      2  EXPRESSION_SYSTEM: ESCHERICHIA COLI;
SOURCE      3  EXPRESSION_SYSTEM_STRAIN: BE167;
SOURCE      4  FRAGMENT: RESIDUES 1-16;
SOURCE      5  ORGANISM_SCIENTIFIC: BACILLUS AMYLOLIQUEFACIENS;
SOURCE      6  EXPRESSION_SYSTEM: ESCHERICHIA COLI;
SOURCE      7  FRAGMENT: RESIDUES 17-214;
SOURCE      8  ORGANISM_SCIENTIFIC: BACILLUS MACERANS

```

---

**KEYWDS****Overview**

The KEYWDS record contains a set of terms relevant to the entry. Terms in the KEYWDS record provide a simple means of categorizing entries and may be used to generate index files. This record addresses some of the limitations found in the classification field of the HEADER record. It provides the opportunity to add further annotation to the entry in a concise and computer-searchable fashion.

**Record Format**

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"KEYWDS"	
9 - 10	Continuation	continuation	Allows concatenation of records if necessary.
11 - 70	List	keywds	Comma-separated list of keywords relevant to the entry.

**Details**

\* The KEYWDS record contains a list of terms relevant to the entry, similar to that found in journal articles. A phrase may be used if it presents a single concept (e.g., reaction center). Terms provided in this record may include those that describe the following:

- Functional classification.
- Metabolic role.
- Known biological or chemical activity.
- Structural classification.

\*Other classifying terms may be used. No ordering is required for these terms. A number of PDB entries contain complexes of macromolecules. In these cases, all terms applicable to each molecule should be provided.

\*Note that the terms in the KEYWDS record duplicate those found in the classification field of the HEADER record. Terms abbreviated in the HEADER record are unabbreviated in KEYWDS.

#### Verification/Validation/Value Authority Control

Terms used in the KEYWDS record are subject to scientific and editorial review. A list of terms, definitions, and synonyms will be maintained at the PDB. Every attempt will be made to provide some level of consistency with keywords used in other biological databases.

#### Relationships to Other Record Types

HEADER records contain a classification term which must also appear in KEYWDS. Scientific judgement will dictate when terms used in one entry to describe a molecule should be included in other entries with the same or similar molecules.

#### Example

```

      1           2           3           4           5           6           7
123456789012345678901234567890123456789012345678901234567890
KEYWDS      LYASE, TRICARBOXYLIC ACID CYCLE, MITOCHONDRION, OXIDATIVE
KEYWDS      2 METABOLISM

```

## EXPDTA

### Overview

The EXPDTA record presents information about the experiment.

The EXPDTA record identifies the experimental technique used. This may refer to the type of radiation and sample, or include the spectroscopic or modeling technique. Permitted values include:

```

ELECTRON DIFFRACTION
FIBER DIFFRACTION
FLUORESCENCE TRANSFER
NEUTRON DIFFRACTION
NMR
THEORETICAL MODEL
X-RAY DIFFRACTION

```

### Record Format

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"EXPDTA"	
9 - 10	Continuation	continuation	Allows concatenation of multiple records.
11 - 70	SList	technique	The experimental technique(s) with optional comment describing the sample or experiment.

### Details

- \* EXPDTA is mandatory and appears in all entries.
- \* The technique must match one of the permitted values. See above.
- \* If more than one model appears in the entry, the number of models included must be stated.
- \* If only one model appears in the entry, its significance must be stated, such as it being a minimized average structure.

\* If more than one technique was used for the structure determination and is being represented in the entry, EXPDTA presents the techniques as a semi-colon separated list. Each technique may have a comment, which appears before the semi-colon.

#### Verification/Validation/Value Authority Control

The verification program checks that the EXPDTA record appears in the entry and that the technique matches one of the allowed values. It also checks that the relevant standard REMARK is added in the case of NMR, fiber, or theoretical modeling studies, and that the correct CRYST1 and SCALE are used in these cases. If an entry contains multiple models, the verification program checks for the correct number of matching MODEL/ENDMDL records.

#### Relationships to Other Record Types

If the experiment is an NMR, fiber, or theoretical modeling study, this may be stated in the TITLE, and the appropriate EXPDTA and REMARK records should appear. Specific details of the data collection and experiment appear in the REMARKs.

In the case of a polycrystalline fiber diffraction study, CRYST1 and SCALE contain the normal unit cell data.

#### Example

```

      1           2           3           4           5           6           7
123456789012345678901234567890123456789012345678901234567890
EXPDTA      X-RAY DIFFRACTION

EXPDTA      NEUTRON DIFFRACTION; X-RAY DIFFRACTION

EXPDTA      NMR, 32 STRUCTURES

EXPDTA      THEORETICAL MODEL

EXPDTA      FIBER DIFFRACTION, FIBER

EXPDTA      FIBER DIFFRACTION, POLYCRYSTALLINE SAMPLE

```

---

## AUTHOR

### Overview

The AUTHOR record contains the names of the people responsible for the contents of the entry.

### Record Format

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"AUTHOR"	
9 - 10	Continuation	continuation	Allows concatenation of multiple records.
11 - 70	List	authorList	List of the author names, separated by commas.

### Details

\* The authorList field lists author names separated by commas with no subsequent spaces.

\* Representation of personal names:

- First and middle names are indicated by initials, each followed by a period, and precede the surname.
- Only the surname (family or last name) of the author is given in full.

- Hyphens can be used if they are part of the author's name.
- Apostrophes are allowed in surnames.
- The word Junior is not abbreviated.
- Umlauts and other character modifiers are not given.

**\* Structure of personal names:**

- There is no space after any initial and its following period.
- Blank spaces are used in a name only if properly part of the surname (e.g., J.VAN DORN), or between surname and Junior, II, or III.
- Abbreviations that are part of a surname, such as St. or Ste., are followed by a period and a space before the next part of the surname.

**\* Representation of corporate names:**

- Group names used for one or all of the authors should be spelled out in full.
- The name of the larger group comes before the name of a subdivision, e.g., University of Somewhere, Department of Chemistry.

**\* Structure of list:**

- Line breaks between multiple lines in the authorList occur only after a comma.
- Personal names are not split across two lines.

**\* Special cases:**

- Names are given in English if there is an accepted English version; otherwise in the native language, transliterated if necessary.
- "ET AL." may be used when all authors are not individually listed.

**Verification/Validation/Value Authority Control**

The verification program checks that the authorList field is correctly formatted. It does not perform any spelling checks or name verification.

**Relationships to Other Record Types**

The format of the names in the AUTHOR record is the same as in JRNL and REMARK 1 references.

**Example**

```

1      2      3      4      5      6      7
123456789012345678901234567890123456789012345678901234567890
AUTHOR   M.B.BERRY,B.MEADOR,T.BILDERBACK,P.LIANG,M.GLASER,
AUTHOR   2 G.N.PHILLIPS JUNIOR,T.L.ST. STEVENS

AUTHOR   C.-I.BRANDEN,C.J.BIRKETT-CLEWS,L.RIVA DI SANSAVERINO

```

---

## REVDAT

### Overview

REVDAT records contain a history of the modifications made to an entry since its release.

### Record Format

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"REVDAT"	
8 - 10	Integer	modNum	Modification number.
11 - 12	Continuation	continuation	Allows concatenation of multiple records.
14 - 22	Date	modDate	Date of modification (or release for new entries). This is not repeated on continuation lines.
24 - 28	String(5)	modId	Identifies this particular modification. It links to the archive used internally by PDB. This is not repeated on continuation lines.
32	Integer	modType	An integer identifying the type of modification. In case of revisions with more than one possible modType, the highest value applicable will be assigned.
40 - 45	LString(6)	record	Name of the modified record.
47 - 52	LString(6)	record	Name of the modified record.
54 - 59	LString(6)	record	Name of the modified record.
61 - 66	LString(6)	record	Name of the modified record.

### Details

\* Each time revisions are made to the entry, a modification number is assigned in increasing (by 1) numerical order. REVDAT records appear in descending order (most recent modification appears first). New entries have a REVDAT record with modNum equal to 1 and modType equal to 0. Allowed modTypes are:

- 0 Initial released entry.
- 1 Miscellaneous - mostly typographical.
- 2 Modification of a CONECT record.
- 3 Modification to coordinates or transformations.
- 4 - 9 Not defined.

\* Each revision may have more than one REVDAT record, and each revision has a separate continuation field.

### Verification/Validation/Value Authority Control

The modType must be one of the defined types, and the given record type must be valid. If modType is 0, the modId must match the entry's ID code in the HEADER record.

**Relationships to Other Record Types**

REMARK 860 presents the correction or change that is made to an entry. Also, see verification above.

**Example**

	1	2	3	4	5	6	7
123456789012345678901234567890123456789012345678901234567890							
REVDAT	3	15-OCT-89	1PRCB	1	REMARK		
REVDAT	2	19-APR-89	1PRCA	2	CONNECT		
REVDAT	1	09-JAN-89	1PRC	0			

---

**SPRSDE****Overview**

The SPRSDE records contain a list of the ID codes of entries that were made obsolete by the given coordinate entry and withdrawn from the PDB release set. One entry may replace many. It is PDB policy that only the principal investigator of a structure has the authority to withdraw it.

**Record Format**

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"SPRSDE"	
9 - 10	Continuation	continuation	Allows for multiple ID codes.
12 - 20	Date	sprsdDate	Date this entry superseded the listed entries. This field is not copied on continuations.
22 - 25	IDcode	idCode	ID code of this entry. This field is not copied on continuations.
32 - 35	IDcode	sIdCode	ID code of a superseded entry.
37 - 40	IDcode	sIdCode	ID code of a superseded entry.
42 - 45	IDcode	sIdCode	ID code of a superseded entry.
47 - 50	IDcode	sIdCode	ID code of a superseded entry.
52 - 55	IDcode	sIdCode	ID code of a superseded entry.
57 - 60	IDcode	sIdCode	ID code of a superseded entry.
62 - 65	IDcode	sIdCode	ID code of a superseded entry.
67 - 70	IDcode	sIdCode	ID code of a superseded entry.

**Details**

\* The ID code list is terminated by the first blank sIDcode field.

**Verification/Validation/Value Authority Control**

PDB checks that the superseded entries have actually been withdrawn from release.

**Relationships to Other Record Types**

The sprsdeDate is usually the date the entry is released, and therefore matches the date in the REVDAT 1 record. The ID code found in the idCode field must be the same as one found in the idCode field of the HEADER record.

**Example**

```

      1      2      3      4      5      6      7
123456789012345678901234567890123456789012345678901234567890
SPRSDE      17-JUL-84 4HHB      1HHB

SPRSDE      27-FEB-95 1GDJ      1LH4 2LH4

```

**JRNL****Overview**

The JRNL record contains the primary literature citation that describes the experiment which resulted in the deposited coordinate set. There is at most one JRNL reference per entry. If there is no primary reference, then there is no JRNL reference. Other references are given in REMARK 1.

PDB is in the process of linking and/or adding all references to CitDB, the literature database used by the Genome Data Base (available at URL <http://gdbwww.gdb.org/gdb-bin/genera/genera/citation/Citation>).

**Record Format**

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"JRNL "	
13 - 70	LString	text	See Details below.

**Details**

\* The following tables are used to describe the sub-record types of the JRNL record.

\* The AUTH sub-record is mandatory in JRNL. This is followed by TITL, EDIT, REF, PUBL, and REFN sub-record types. REF and REFN are also mandatory in JRNL. EDIT and PUBL may appear only if the reference is to a non-journal.

\* If the JRNL reference is in the MEDLINE database the information in the MEDLINE reference will be used to supply information for the sub-record types.

\* When a MEDLINE reference is used, the abbreviation of the journal will be converted to the CASSI abbreviation as listed in the coden list used jointly by the Cambridge Crystallographic Data Centre (CCDC) and the PDB.

**1. AUTH**

\* AUTH contains the list of authors associated with the cited article or contribution to a larger work (i.e., AUTH is not used for the editor of a book).

\* The author list is formatted similarly to the AUTHOR record. It is a comma-separated list of names. Spaces at the end of a sub-record are not significant; all other spaces are significant. See the AUTHOR record for full details.

\* The authorList field of continuation sub-records in JRNL differs from that in AUTHOR by leaving no leading blank in column 20 of any continuation lines.

\* One author's name, consisting of the initials and family name, cannot be split across two lines. If there are continuation sub-records, then all but the last sub-record must end in a comma.



COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"JRNL "	
13 - 16	LString(4)	"AUTH"	Appears on all continuation records.
17 - 18	Continuation	continuation	Allows concatenation of multiple records.
20 - 70	List	authorList	List of the authors.

**2. TITL**

\* TITL specifies the title of the reference. This is used for the title of a journal article, chapter, or part of a book. The TITL line is omitted if the author(s) listed in authorList wrote the entire book (or other work) listed in REF and no section of the book is being cited.

\* If an article is in a language other than English and is printed with an alternate title in English, the English language title is given, followed by a space and then the name of the language (in its English form, in square brackets) in which the article is written.

\* If the title of an article is in a non-Roman alphabet the title is transliterated.

\* The actual title cited is reconstructed in a manner identical to other continued records, i.e., trailing blanks are discarded and the continuation line is concatenated with a space inserted.

\* A line cannot end with a hyphen. A compound term (two elements connected by a hyphen) or chemical names which include a hyphen must appear on a single line, unless they are too long to fit on one line, in which case the split is made at a normally-occurring hyphen. An individual word cannot be hyphenated at the end of a line and put on two lines. An exception is when there is a repeating compound term where the second element is omitted, e.g., "DOUBLE- AND TRIPLE-RESONANCE". In such a case the non-completed word "DOUBLE-" could end a line and not alter reconstruction of the title.

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"JRNL "	
13 - 16	LString(4)	"TITL"	Appears on all continuation lines.
17 - 18	Continuation	continuation	Permits long titles.
20 - 70	LString	title	Title of the article.

**3. EDIT**

\* EDIT appears if editors are associated with a non-journal reference. The editor list is formatted and concatenated in the same way that author lists are.

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"JRNL "	
13 - 16	LString(4)	"EDIT"	Appears on all continuation records.
17 - 18	Continuation	continuation	Allows a long list of editors.
20 - 70	List	editorList	List of the editors.

**4. REF**

\* REF is a group of fields which contains either the publication status or the name of the publication (and any supplement and/or report information), volume, page, and year. There are two forms of this sub-record group, depending upon the citation's publication status.

**4a. If the reference has not yet been published, the sub-record type group has the form:**

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"JRNL "	
13 - 16	LString(3)	"REF"	
20 - 34	LString(15)	"TO BE PUBLISHED"	

\* At the present time, there is no formal mechanism in place for monitoring the subsequent publication of such referenced papers. PDB relies upon the depositor to provide reference update information since preliminary information can change by the time of actual publication.

**4b.** If the reference has been **published**, then the REF sub-record type contains information about the name of the publication, supplement, report, volume, page, and year in the appropriate fields. These fields are detailed below.

\* Publication name (first item in pubName field):

- If the publication is a serial (i.e., a journal, an annual, or other non-book or non-monographic item issued in parts and intended to be continued indefinitely), use the abbreviated name of the publication as listed in American Chemical Society (A.C.S.) publications such as CAS Source Index (CASSI) or Chemical Abstracts. (The A.C.S. abbreviation is based on the International Standards Organization's standard ISO 4-1984[E].) If the A.C.S. has not yet established an abbreviation for the publication, the name is given in full.

- If the publication is a book, monograph, or other non-serial item, use its full name according to the Anglo-American Cataloging Rules, 2nd Ed., 1988 revision (AACR2R). (Non-serial items include theses, videos, computer programs, and anything that is complete in one or a finite number of parts.) If there is a sub-title, and the item is verified in an online catalog, it will be included using the same punctuation as in the source of verification. Preference will be given to verification using cataloging of the Library of Congress, the National Library of Medicine, and the British Library, in that order.

- If a book is part of a monographic series: the full name of the book (according to AACR2R) is listed first, followed by the name of the series in which it was published. The series information is given within parentheses and the series name is preceded by "IN:" and a space. If the series has an A.C.S. abbreviation, that abbreviation should be used; otherwise the series name should be listed in full. If applicable, the series name should be followed, after a comma and a space, by a volume (V.) and/or number (NO.) and/or part (PT.) indicator and the relevant characters to indicate its number and/or letter in the series.

\* Supplement (follows publication name in pubName field):

- If a reference is in a supplement to the volume listed, or if information about a "part" is needed to distinguish multiple parts with the same page numbering, such information should be put in the REF sub-record.

- A supplement indication should follow the name of the publication and should be preceded by a comma and a space. Supplement should be abbreviated as "SUPPL." If there is a supplement number or letter, it should follow "SUPPL." without an intervening space. A part indication should also follow the name of

the publication and be preceded by a comma and a space. A part should be abbreviated as "PT.", and the number or letter should follow without an intervening space.

- If there is both a supplement and a part, their order should reflect the order printed on the work itself.

\* Report (follows publication name and any supplement or part information in pubName field):

- If a book has a report designation, the report information should follow the title and precede series information. The name and number of the report is given in parentheses, and the name is preceded by "REPORT:" and a space.

\* Reconstruction of publication name:

- The name of the publication is reconstructed by removing any trailing blanks in the pubName field, and concatenating all of the pubName fields from the continuation lines with an intervening space. There are two conditions where no intervening space is added between lines: when the pubName field on a line ends with a hyphen or a period, or when the line ends with a hyphen (-). When the line ends with a period (.), add a space if this is the only period in the entire pubName field; do not add a space if there are two or more periods throughout the pubName field, excluding any periods after the designations "SUPPL", "V", "NO", or "PT".

\* Volume, page, and year (volume, page, year fields respectively):

- The REF sub-record type group also contains information about volume, page, and year when applicable.

- In the case of a monograph with multiple volumes which is also in a numbered series, the number in the volume field represents the volume number of the book, not the series. (The volume number of the series is in parentheses with the name of the series, as described above under publication name.)

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"JRNL "	
13 - 16	LString(3)	"REF"	
17 - 18	Continuation	continuation	Allows long publication names.
20 - 47	LString	pubName	Name of the publication including section or series designation. This is the only field of this sub-record which may be continued on successive sub-records.
50 - 51	LString(2)	"V."	Appears in the first sub-record only, and only if column 55 is non-blank.
52 - 55	String	volume	Right-justified blank-filled volume information; appears in the first sub-record only.
57 - 61	String	page	First page of the article; appears in the first sub-record only.
63 - 66	Integer	year	Year of publication; first sub-record only.

## 5. PUBL

\* PUBL contains the name of the publisher and place of publication if the reference is to a book or other non-journal publication. If the non-journal has not yet been published or released, this sub-record is absent.

\* The place of publication is listed first, followed by a space, a colon, another space, and then the name of the publisher/issuer. This arrangement is based on the ISBD(M) International Standard Bibliographic Description for Monographic Publications (Rev.Ed., 1987) and AACR2R and is used in public online catalogs in libraries. Details on the contents of PUBL are given below.

**\* Place of publication:**

- Give the place of publication. If the name of the country, state, province, etc. is considered necessary to distinguish the place of publication from others of the same name, or for identification, then follow the city with a comma, a space, and the name of the larger geographic area.
- If there is more than one place of publication, only the first listed will be used. If an online catalog record is used to verify the item, the first place listed there will be used, omitting any brackets. Preference will be given to the cataloging done by the Library of Congress, the National Library of Medicine, and the British Library, in that order.

**\* Publisher's name (or name of other issuing entity):**

- Give the name of the publisher in the shortest form in which it can be understood and identified internationally, according to AACR2R rule 1.4D.
- If there is more than one publisher listed in the publication, only the first will be used in the PDB file. If an online catalog record is used to verify the item, the first place listed there will be used for the name of the publisher. Preference will be given to the cataloging of the Library of Congress, the National Library of Medicine, and the British Library, in that order.

**\* Ph.D. and other theses:**

- Theses are presented in the PUBL record if the degree has been granted and the thesis made available for public consultation by the degree-granting institution.
- The name of the degree-granting institution (the issuing agency) is followed by a space and "(THESIS)".

**\* Reconstruction of place and publisher:**

- The PUBL sub-record type can be reconstructed by removing all trailing blanks in the pub field and concatenating all of the pub fields from the continuation lines with an intervening space. Continued lines do not begin with a space.

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"JRNL"	
13 - 16	LString(4)	"PUBL"	
17 - 18	Continuation	continuation	Allows long publisher and place names.
20 - 70	LString	pub	City of publication and name of the publisher/institution.

## 6. REFN

\* REFN is a group of fields which contains encoded references to the citation. No continuation lines are possible. Each piece of coded information has a designated field.

\* The American Society for Testing and Materials (ASTM) number is an encoded reference to the journal title. New ASTM codens are assigned by the Chemical Abstracts Service and appear in CASSI and its supplements.

\* The country field is blank if the reference was published in more than one country.

\* If more than one ISBN is known, select one that matches the individual volume cited (if it happens to be in a set that also has an ISBN for the set). If the reason for multiple ISBNs is that the publication is issued in more than

one country, use the ISBN for the country of the first listed place of publication. If there are hardcover and paperback ISBN numbers, use the the ISBN for the hardbound version.

\* Because some publications do not have an ASTM coden, an ISSN number, or an ISBN number, each publication is assigned a number. This list of numbers, or codens, was established by the Cambridge Crystallographic Data Center (CCDC) and new numbers are assigned by both CCDC and PDB as new publications are added to their respective databases.

\* There are two forms of this sub-record type group, depending upon the publication status.

**6a.** This form of the REFN sub-record type group is used if the citation has not been published.

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"JRNL "	
13 - 16	LString(4)	"REFN"	
67 - 70	LString(4)	"0353"	This is the CCDC/PDB coden for unpublished works.

**6b.** This form of the REFN sub-record type group is used if the citation has been published.

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"JRNL "	
13 - 16	LString(4)	"REFN"	
20 - 23	LString(4)	"ASTM"	
25 - 30	LString(6)	astm	ASTM devised coden.
33 - 34	LString(2)	country	Country of publication code as defined in the OCLC/MARC cataloging format (optional).
36 - 39	LString(4)	"ISBN" or "ISSN"	International Standard Book Number or International Standard Serial Number.
41 - 65	LString	isbn	ISSN or ISBN number (final digit may be a letter and may contain one or more dashes).
67 - 70	LString(4)	coden	Code from CCDC/PDB coden list.

#### Verification/Validation/Value Authority Control

PDB verifies that this record is correctly formatted.

PDB uses MEDLINE to verify the accuracy of references and to obtain information required for CitDB that is not required by the PDB listing. The process of using MEDLINE requires following the National Library of Medicine rules for the transcription of names and titles. Articles in non-MEDLINE journals are verified through other online databases or with the reprint in hand. Verification of book references is done using online cooperative or individual library catalogs.

Citations appearing in JRNL may not also appear in REMARK 1.

**Relationships to Other Record Types**

The publication cited as the JRNL record may not be repeated in REMARK 1.

**Example**

	1	2	3	4	5	6	7
1234567890123456789012345678901234567890123456789012345678901234567890							
JRNL	AUTH	N.THANKI,J.K.M.RAO,S.I.FOURLING,W.J.HOWE,					
JRNL	AUTH 2	A.G.TOMASSELLI,R.L.HEINRIKSON,S.THAISRIVONGS,					
JRNL	AUTH 3	A.WLODAWER					
JRNL	TITL	CRYSTAL STRUCTURE OF A COMPLEX OF HIV-1 PROTEASE					
JRNL	TITL 2	WITH A DIHYDROETHYLENE-CONTAINING INHIBITOR:					
JRNL	TITL 3	COMPARISONS WITH MOLECULAR MODELING					
JRNL	REF	TO BE PUBLISHED					
JRNL	REFN	0353					
JRNL	AUTH	G.FERMI,M.F.PERUTZ,B.SHAANAN,R.FOURME					
JRNL	TITL	THE CRYSTAL STRUCTURE OF HUMAN DEOXYHAEMOGLOBIN AT					
JRNL	TITL 2	1.74 ANGSTROMS RESOLUTION					
JRNL	REF	J.MOL.BIOL. V. 175 159 1984					
JRNL	REFN	ASTM JMOBAC UK ISSN 0022-2836 0070					

**Known Problems**

\* Interchange of bibliographic information and linking with other databases is hampered by the lack of labels or specific locations for certain types of information or by more than one type of information being in a particular location. This is most likely to occur with books, series, and reports. Some of the points below provide details about the variations and/or blending of information.

\* Titles of the publications that require more than 28 characters on the REF line must be continued on subsequent lines. There is some awkwardness due to volume, page, and year appearing on the first REF line, thereby splitting up the title.

\* Information about a supplement and its number/letter is presented in the publication's title field (on the REF lines in columns 20 - 47). This sometimes means that the publication's coden has several versions of REF title information.

\* When series information for a book is presented, it is added to the REF line. The number of REF lines can become large in some cases because of the 28-column limit for title information in REF.

\* There is often an ISBN for a book title and a separate ISSN for the series in which it was published. There is no way to present more than one of these.

\* Books that are issued in more than one series are not accommodated.

\* Many books are issued in more than one country. The publisher has a separate ISBN number in each country. There is no place to put any additional applicable ISBN numbers, which would be useful in an international database such as the PDB.

\* The country code prefix of the ISBN may not match the country of the place of publication that is listed on the PUBL line when a book is published in more than one country.

\* Pagination is limited to the beginning page.

\* There is no place for listing a reference's accession number in another database.

\* MEDLINE truncates the author list after the tenth name.

## REMARK

### Overview

REMARK records present experimental details, annotations, comments, and information not included in other records. In a number of cases, REMARKs are used to expand the contents of other record types. A new level of structure is being used for some REMARK records. This is expected to facilitate searching and will assist in the conversion to a relational database.

The very first line of every set of REMARK records is used as a spacer to aid in reading.

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"REMARK"	
8 - 10	Integer	remarkNum	Remark number. It is not an error for remark n to exist in an entry when remark n-1 does not.
12 - 70	LString	empty	Left as white space in first line of each new remark.

REMARK 1, 2, and 3, detailed below, are specific for references, resolution, and refinement, respectively.

## REMARK 1

REMARK 1 lists important publications related to the structure presented in the entry. These citations are chosen by the depositor. They are listed in reverse-chronological order. Citations are not repeated from the JRNL records. After the first blank record and the REFERENCE sub-record, the sub-record types for REMARK 1 are the same as in the JRNL sub-record types. For details, see the JRNL section.

PDB is in the process of linking and/or adding references to CitDB, the literature database of the Genome Data Base (available at URL <http://gdbwww.gdb.org/gdb-bin/genera/genera/citation/Citation>).

### Record Format and Details

As with all other remarks, the first line is empty and is used as a spacer.

The following tables are used to describe the sub-record types of REMARK 1.

### 1. REFERENCE

Each reference is preceded by a line indicating the reference number in the entry.

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"REMARK"	
10	LString(1)	"1"	
12 - 20	LString(9)	"REFERENCE"	
22 - 70	Integer	refNum	Reference number. Starts with 1 and increments by 1.

**2. AUTH**

AUTH contains the list of authors of the reference.

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"REMARK"	
10	LString(1)	"1"	
13 - 16	LString(4)	"AUTH"	Appears on all continuation records.
17 - 18	Continuation	continuation	Allows a long list of authors.
20 - 70	List	authorList	List of the authors.

See JRNL AUTH for details.

**3. TITL**

TITL specifies the title of the reference.

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"REMARK"	
10	LString(1)	"1"	
13 - 16	LString(4)	"TITL"	Appears on all continuation records.
17 - 18	Continuation	continuation	Permits long titles.
20 - 70	LString	title	Title of the article.

See JRNL TITL for details.

**4. EDIT**

EDIT appears if editors are associated with a non-journal reference.

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"REMARK"	
10	LString(1)	"1"	
13 - 16	LString(4)	"EDIT"	Appears on all continuation records.
17 - 18	Continuation	continuation	Permits long list of editors.
20 - 70	LString	editorList	List of the editors.

See JRNL EDIT for details.



**5. REF**

REF is a group of fields which contains the name of the publication.

**5a. If it has not yet been published, the REF sub-record type has the form:**

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"REMARK"	
10	LString(1)	"1"	
13 - 16	LString(3)	"REF"	
20 - 34	LString(15)	"TO BE PUBLISHED"	

At the present time, there is no formal mechanism in place for monitoring the subsequent publication of referenced papers. PDB relies upon the depositor to provide reference update information since preliminary information can change by the time of actual publication.

**5b. If the reference has been published, then the REF sub-record type group contains information about the name of the publication, supplement, report, volume, page, and year, in the appropriate fields.**

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"REMARK"	
10	LString(1)	"1"	
13 - 16	LString(3)	"REF"	
17 - 18	Continuation	continuation	Permits long publication names.
20 - 47	LString	pubName	Name of the publication including section or series designation. This is the only field of this record which may be continued on successive records.
50 - 51	LString(2)	"V."	Appears in the first record only, and only if column 55 is filled in.
52 - 55	String	volume	Right-justified blank-filled volume information; appears in the first sub-record only.
57 - 61	String	page	First page of the article; appears in the first sub-record only.
63 - 66	Integer	year	Year of publication, first record only.

See JRNL REF for details.

**6. PUBL**

PUBL contains the name of the publisher and place of publication if the reference is to a book or other non-journal publication. If the reference has not yet been published or released, this sub-record is absent.

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"REMARK"	
10	LString(1)	"1"	
13 - 16	LString(4)	"PUBL"	
17 - 18	Continuation	continuation	Permits long publisher and city information.
20 - 70	LString	pub	Name of the publisher and city of publication.

See JRNL PUBL for details.

**7. REFN**

REFN is a group of fields which contains encoded references to the citation.

**7a. If the citation has not been published, this form of the REFN sub-record type group is used.**

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"REMARK"	
10	LString(1)	"1"	
13 - 16	LString(4)	"REFN"	
67 - 70	LString(4)	"0353"	This is the PDB coden for unpublished works.

7b. If the citation has been published, this form of the REFN sub-record type group is used.

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"REMARK"	
10	LString(1)	"1"	
13 - 16	LString(4)	"REFN"	
20 - 23	LString(4)	"ASTM"	Blank if reference is not serialized.
25 - 30	LString	astm	Code from the ASTM file.
33 - 34	LString	country	2-digit abbreviation for country of publication.
36 - 39	LString(4)	"ISBN" or "ISSN"	
41 - 65	LString	isbn	ISSN or ISBN number.
68 - 70	LString(4)	coden	Number from Cambridge Crystallographic Data Center coden list, or assigned by the PDB.

See JRNL REFN for details.

#### Verification/Validation/Value Authority Control

PDB verifies that this record is correctly formatted.

PDB uses MEDLINE to verify the accuracy of references and to obtain information required for CitDB that is not required by the PDB listing. The process of using MEDLINE requires following the National Library of Medicine rules for the transcription of names and titles. Articles in non-MEDLINE journals are verified through other online databases or with the reprint in hand. Verification of book references is done using online cooperative or individual library catalogs.

Citations appearing in REMARK 1 may not appear in JRNL.

#### Relationships to Other Record Types

Citations appearing in REMARK 1 may not appear in JRNL.

#### Example

```

123456789012345678901234567890123456789012345678901234567890
REMARK 1
REMARK 1 REFERENCE 1
REMARK 1 AUTH A.M.J.J.BONVIN,J.A.C.RULLMANN,R.M.J.N.LAMERICHs,
REMARK 1 AUTH 2 R.BOELENS,R.KAPTEIN
REMARK 1 TITL "ENSEMBLE" ITERATIVE RELAXATION MATRIX APPROACH:
REMARK 1 TITL 2 A NEW NMR REFINEMENT PROTOCOL APPLIED TO THE
REMARK 1 TITL 3 SOLUTION STRUCTURE OF CRAMBIN
REMARK 1 REF PROTEINS: STRUCT.,FUNCT., V. 15 385 1993
REMARK 1 REF 2 GENET.
REMARK 1 REFN ASTM PSFGY US ISSN 0887-3585 0867

```

```

REMARK 1 REFERENCE 2
REMARK 1 AUTH J.A.C.RULLMANN,A.M.J.J.BONVIN,R.BOELENS,R.KAPTEIN
REMARK 1 TITL STRUCTURE DETERMINATION BY NMR - APPLICATION TO
REMARK 1 TITL 2 CRAMBIN
REMARK 1 EDIT D.M.SOUMPASIS,T.M.JOVIN
REMARK 1 REF COMPUTATION OF BIOMOLECULAR 1 1992
REMARK 1 REF 2 STRUCTURES; ACHIEVEMENTS,
REMARK 1 REF 3 PROBLEMS, AND PERSPECTIVES
REMARK 1 PUBL BERLIN : SPRINGER-VERLAG
REMARK 1 REFN GW ISBN 3540559515 2010
REMARK 1 REFERENCE 3
REMARK 1 AUTH R.M.J.M.LAMERICHs
REMARK 1 REF 2D NMR STUDIES OF 1989
REMARK 1 REF 2 BIOMOLECULES: PROTEIN
REMARK 1 REF 3 STRUCTURE AND PROTEIN-DNA
REMARK 1 REF 4 INTERACTIONS
REMARK 1 PUBL UTRECHT : UNIVERSITY OF UTRECHT (THESIS)
REMARK 1 REFN NE 2011

REMARK 1
REMARK 1 REFERENCE 1
REMARK 1 AUTH G.F.FERMI,M.F.PERUTZ
REMARK 1 REF HAEMOGLOBIN AND MYOGLOBIN 1981
REMARK 1 REF 2 (IN: ATLAS OF MOLECULAR
REMARK 1 REF 3 STRUCTURES IN BIOLOGY, V.2)
REMARK 1 PUBL OXFORD : CLARENDON PRESS
REMARK 1 REFN ISBN 0-19-854706-4 0986

```

### Known Problems

See JRNL for a listing of problems associated with references.

---

## REMARK 2

REMARK 2 states the highest resolution, in Angstroms, that was used in building the model. As with all the remarks, the first REMARK 2 record is empty and is used as a spacer.

### Record Format and Details

\* The second REMARK 2 record has one of two formats. The first is used for diffraction studies, the second for other types of experiments in which resolution is not relevant, e.g., NMR and theoretical modeling.

\* For diffraction experiments:

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"REMARK"	
10	LString(1)	"2"	
12 - 22	LString(11)	"RESOLUTION."	
23 - 27	Real(5.2)	resolution	Resolution.
29 - 38	LString(10)	"ANGSTROMS."	

REMARK 2 when not a diffraction experiment:

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"REMARK"	
10	LString(1)	"2"	
12 - 38	LString(28)	"RESOLUTION. NOT APPLICABLE."	
41 - 70	String	comment	Comment.

\* Additional explanatory text may be included starting with the third line of the REMARK 2 record. For example, depositors may wish to qualify the resolution value provided due to unusual experimental conditions.

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"REMARK"	
10	LString(1)	"2"	
12 - 22	LString(11)	"RESOLUTION."	
24 - 70	String	comment	Comment.

#### Example

```

1      2      3      4      5      6      7
123456789012345678901234567890123456789012345678901234567890
REMARK  2
REMARK  2 RESOLUTION. 1.74 ANGSTROMS.

REMARK  2
REMARK  2 RESOLUTION. NOT APPLICABLE.

REMARK  2
REMARK  2 RESOLUTION. NOT APPLICABLE.
REMARK  2 THIS EXPERIMENT WAS CARRIED OUT USING FLUORESCENCE TRANSFER
REMARK  2 AND THEREFORE NO RESOLUTION CAN BE CALCULATED.
```

## REMARK 3

### Overview

REMARK 3 presents information on refinement program(s) used and the related statistics. For non-diffraction studies, REMARK 3 is used to describe any refinement done, but its format in those cases is mostly free text.

If more than one refinement package was used, the PROGRAM and AUTHORS tokens may be repeated in order to show the programs used. The final program listed is the one which produced the output statistics.

Refinement packages are being enhanced to output PDB REMARK 3. A token: value template style facilitates parsing. Spacer REMARK 3 lines are interspersed for visually organizing the information.

The templates below have been adopted in consultation with program authors. PDB is continuing this dialogue with program authors, and expects the library of PDB records output by the programs to greatly increase in the near future.

Instead of providing a **Record Format** table, each template is given as it appears in PDB entries.

**Details**

\* The value "NULL" is given when there is no data available for a particular token.

**Refinement using X-PLOR**

This remark will be output by X-PLOR 3.8 for direct submission to PDB. Structures done using earlier versions of X-PLOR will contain the same template, but with many of the data items containing "NULL". X.X is replaced by the version number.

**Template**

```

REMARK      3
REMARK      3 REFINEMENT.
REMARK      3   PROGRAM      : X-PLOR X.X
REMARK      3   AUTHORS      : BRUNGER
REMARK      3
REMARK      3 DATA USED IN REFINEMENT.
REMARK      3   RESOLUTION RANGE HIGH (ANGSTROMS) :
REMARK      3   RESOLUTION RANGE LOW  (ANGSTROMS) :
REMARK      3   DATA CUTOFF              (SIGMA(F)) :
REMARK      3   COMPLETENESS FOR RANGE          (%) :
REMARK      3   NUMBER OF REFLECTIONS              :
REMARK      3
REMARK      3 FIT TO DATA USED IN REFINEMENT.
REMARK      3   CROSS-VALIDATION METHOD              :
REMARK      3   FREE R VALUE TEST SET SELECTION        :
REMARK      3   R VALUE              (WORKING SET)      :
REMARK      3   FREE R VALUE                                :
REMARK      3   FREE R VALUE TEST SET SIZE              (%) :
REMARK      3   FREE R VALUE TEST SET COUNT              :
REMARK      3   ESTIMATED ERROR OF FREE R VALUE          :
REMARK      3
REMARK      3 FIT IN THE HIGHEST RESOLUTION BIN.
REMARK      3   BIN RESOLUTION RANGE HIGH              (A) :
REMARK      3   BIN RESOLUTION RANGE LOW                (A) :
REMARK      3   COMPLETENESS IN THIS BIN                (%) :
REMARK      3   REFLECTIONS IN BIN              (WORKING SET) :
REMARK      3   BIN R VALUE              (WORKING SET)      :
REMARK      3   BIN FREE R VALUE                                :
REMARK      3   BIN FREE R VALUE TEST SET SIZE          (%) :
REMARK      3   BIN FREE R VALUE TEST SET COUNT          :
REMARK      3   ESTIMATED ERROR OF BIN FREE R VALUE      :
REMARK      3
REMARK      3 NUMBER OF NON-HYDROGEN ATOMS USED IN REFINEMENT.
REMARK      3   PROTEIN ATOMS                        :
REMARK      3   NUCLEIC ACID ATOMS                   :
REMARK      3   HETEROGEN ATOMS                      :
REMARK      3   SOLVENT ATOMS                       :
REMARK      3
REMARK      3 B VALUES.
REMARK      3   FROM WILSON PLOT              (A**2) :
REMARK      3   MEAN B VALUE              (OVERALL, A**2) :
REMARK      3   OVERALL ANISOTROPIC B VALUE.
REMARK      3   B11 (A**2) :
REMARK      3   B22 (A**2) :
REMARK      3   B33 (A**2) :
REMARK      3   B12 (A**2) :
REMARK      3   B13 (A**2) :
REMARK      3   B23 (A**2) :

```

```

REMARK 3
REMARK 3 ESTIMATED COORDINATE ERROR:
REMARK 3 ESD FROM LUZZATI PLOT (A) :
REMARK 3 ESD FROM SIGMA (A) :
REMARK 3 LOW RESOLUTION CUTOFF (A) :
REMARK 3
REMARK 3 CROSS-VALIDATED ESTIMATED COORDINATE ERROR.
REMARK 3 ESD FROM C-V LUZZATI PLOT (A) :
REMARK 3
REMARK 3 RMS DEVIATIONS FROM IDEAL VALUES.
REMARK 3 BOND LENGTHS (A) :
REMARK 3 BOND ANGLES (DEGREES) :
REMARK 3 DIHEDRAL ANGLES (DEGREES) :
REMARK 3 IMPROPER ANGLES (DEGREES) :
REMARK 3
REMARK 3 ISOTROPIC THERMAL FACTOR RESTRAINTS. RMS SIGMA
REMARK 3 MAIN-CHAIN BOND (A**2) : ;
REMARK 3 MAIN-CHAIN ANGLE (A**2) : ;
REMARK 3 SIDE-CHAIN BOND (A**2) : ;
REMARK 3 SIDE-CHAIN ANGLE (A**2) : ;
REMARK 3
REMARK 3 NCS MODEL :
REMARK 3
REMARK 3 NCS RESTRAINTS. RMS SIGMA/WEIGHT
REMARK 3 GROUP 0 POSITIONAL (A) : ;
REMARK 3 GROUP 0 B-FACTOR (A**2) : ;
REMARK 3
REMARK 3 PARAMETER FILE 1 :
REMARK 3 PARAMETER FILE 2 :
REMARK 3 TOPOLOGY FILE 1 :
REMARK 3 TOPOLOGY FILE 2 :
REMARK 3
REMARK 3 OTHER REFINEMENT REMARKS:

```

## Refinement using NUCLSQ

## Template

```

REMARK 3
REMARK 3 REFINEMENT.
REMARK 3 PROGRAM : NUCLSQ
REMARK 3 AUTHORS : WESTHOF,DUMAS,MORAS
REMARK 3
REMARK 3 DATA USED IN REFINEMENT.
REMARK 3 RESOLUTION RANGE HIGH (ANGSTROMS) :
REMARK 3 RESOLUTION RANGE LOW (ANGSTROMS) :
REMARK 3 DATA CUTOFF (SIGMA(F)) :
REMARK 3 COMPLETENESS FOR RANGE (%) :
REMARK 3 NUMBER OF REFLECTIONS :
REMARK 3
REMARK 3 FIT TO DATA USED IN REFINEMENT.
REMARK 3 CROSS-VALIDATION METHOD :
REMARK 3 FREE R VALUE TEST SET SELECTION :
REMARK 3 R VALUE (WORKING + TEST SET) :
REMARK 3 R VALUE (WORKING SET) :
REMARK 3 FREE R VALUE :
REMARK 3 FREE R VALUE TEST SET SIZE (%) :
REMARK 3 FREE R VALUE TEST SET COUNT :
REMARK 3
REMARK 3 FIT/AGREEMENT OF MODEL WITH ALL DATA.
REMARK 3 R VALUE (WORKING + TEST SET, NO CUTOFF) :
REMARK 3 R VALUE (WORKING SET, NO CUTOFF) :
REMARK 3 FREE R VALUE (NO CUTOFF) :
REMARK 3 FREE R VALUE TEST SET SIZE (% , NO CUTOFF) :
REMARK 3 FREE R VALUE TEST SET COUNT (NO CUTOFF) :
REMARK 3 TOTAL NUMBER OF REFLECTIONS (NO CUTOFF) :
REMARK 3
REMARK 3 NUMBER OF NON-HYDROGEN ATOMS USED IN REFINEMENT.
REMARK 3 PROTEIN ATOMS :
REMARK 3 NUCLEIC ACID ATOMS :
REMARK 3 HETEROGEN ATOMS :
REMARK 3 SOLVENT ATOMS :
REMARK 3
REMARK 3 B VALUES.
REMARK 3 FROM WILSON PLOT (A**2) :
REMARK 3 MEAN B VALUE (OVERALL, A**2) :
REMARK 3 OVERALL ANISOTROPIC B VALUE.
REMARK 3 B11 (A**2) :
REMARK 3 B22 (A**2) :
REMARK 3 B33 (A**2) :
REMARK 3 B12 (A**2) :
REMARK 3 B13 (A**2) :
REMARK 3 B23 (A**2) :
REMARK 3
REMARK 3 ESTIMATED COORDINATE ERROR.
REMARK 3 ESD FROM LUZZATI PLOT (A) :
REMARK 3 ESD FROM SIGMAA (A) :
REMARK 3 LOW RESOLUTION CUTOFF (A) :
REMARK 3
REMARK 3 RMS DEVIATIONS FROM IDEAL VALUES.
REMARK 3 DISTANCE RESTRAINTS. RMS SIGMA
REMARK 3 SUGAR-BASE BOND DISTANCE (A) : ;
REMARK 3 SUGAR-BASE BOND ANGLE DISTANCE (A) : ;

```



```
REMARK 3 PHOSPHATE BONDS DISTANCE (A) : ;
REMARK 3 PHOSPHATE BOND ANGLE, H-BOND (A) : ;
REMARK 3
REMARK 3 PLANE RESTRAINT (A) : ;
REMARK 3 CHIRAL-CENTER RESTRAINT (A**3) : ;
REMARK 3
REMARK 3 NON-BONDED CONTACT RESTRAINTS.
REMARK 3 SINGLE TORSION CONTACT (A) : ;
REMARK 3 MULTIPLE TORSION CONTACT (A) : ;
REMARK 3
REMARK 3 ISOTROPIC THERMAL FACTOR RESTRAINTS. RMS SIGMA
REMARK 3 SUGAR-BASE BONDS (A**2) : ;
REMARK 3 SUGAR-BASE ANGLES (A**2) : ;
REMARK 3 PHOSPHATE BONDS (A**2) : ;
REMARK 3 PHOSPHATE BOND ANGLE, H-BOND (A**2) : ;
REMARK 3
REMARK 3 OTHER REFINEMENT REMARKS:
```

## Refinement using PROLSQ

## Template

```

REMARK 3
REMARK 3 REFINEMENT.
REMARK 3 PROGRAM : PROLSQ
REMARK 3 AUTHORS : KONNERT, HENDRICKSON
REMARK 3
REMARK 3 DATA USED IN REFINEMENT.
REMARK 3 RESOLUTION RANGE HIGH (ANGSTROMS) :
REMARK 3 RESOLUTION RANGE LOW (ANGSTROMS) :
REMARK 3 DATA CUTOFF (SIGMA(F)) :
REMARK 3 COMPLETENESS FOR RANGE (%) :
REMARK 3 NUMBER OF REFLECTIONS :
REMARK 3
REMARK 3 FIT TO DATA USED IN REFINEMENT.
REMARK 3 CROSS-VALIDATION METHOD :
REMARK 3 FREE R VALUE TEST SET SELECTION :
REMARK 3 R VALUE (WORKING + TEST SET) :
REMARK 3 R VALUE (WORKING SET) :
REMARK 3 FREE R VALUE :
REMARK 3 FREE R VALUE TEST SET SIZE (%) :
REMARK 3 FREE R VALUE TEST SET COUNT :
REMARK 3
REMARK 3 FIT/AGREEMENT OF MODEL WITH ALL DATA.
REMARK 3 R VALUE (WORKING + TEST SET, NO CUTOFF) :
REMARK 3 R VALUE (WORKING SET, NO CUTOFF) :
REMARK 3 FREE R VALUE (NO CUTOFF) :
REMARK 3 FREE R VALUE TEST SET SIZE (% , NO CUTOFF) :
REMARK 3 FREE R VALUE TEST SET COUNT (NO CUTOFF) :
REMARK 3 TOTAL NUMBER OF REFLECTIONS (NO CUTOFF) :
REMARK 3
REMARK 3 NUMBER OF NON-HYDROGEN ATOMS USED IN REFINEMENT.
REMARK 3 PROTEIN ATOMS :
REMARK 3 NUCLEIC ACID ATOMS :
REMARK 3 HETEROGEN ATOMS :
REMARK 3 SOLVENT ATOMS :
REMARK 3
REMARK 3 B VALUES.
REMARK 3 FROM WILSON PLOT (A**2) :
REMARK 3 MEAN B VALUE (OVERALL, A**2) :
REMARK 3 OVERALL ANISOTROPIC B VALUE.
REMARK 3 B11 (A**2) :
REMARK 3 B22 (A**2) :
REMARK 3 B33 (A**2) :
REMARK 3 B12 (A**2) :
REMARK 3 B13 (A**2) :
REMARK 3 B23 (A**2) :
REMARK 3
REMARK 3 ESTIMATED COORDINATE ERROR.
REMARK 3 ESD FROM LUZZATI PLOT (A) :
REMARK 3 ESD FROM SIGMAA (A) :
REMARK 3 LOW RESOLUTION CUTOFF (A) :
REMARK 3
REMARK 3 RMS DEVIATIONS FROM IDEAL VALUES.
REMARK 3 DISTANCE RESTRAINTS. RMS SIGMA
REMARK 3 BOND LENGTH (A) : ;
REMARK 3 ANGLE DISTANCE (A) : ;

```

```

REMARK 3    INTRAPLANAR 1-4 DISTANCE      (A) :      ;
REMARK 3    H-BOND OR METAL COORDINATION  (A) :      ;
REMARK 3
REMARK 3    PLANE RESTRAINT                (A) :      ;
REMARK 3    CHIRAL-CENTER RESTRAINT       (A**3) :     ;
REMARK 3
REMARK 3    NON-BONDED CONTACT RESTRAINTS. •
REMARK 3    SINGLE TORSION                 (A) :      ;
REMARK 3    MULTIPLE TORSION              (A) :      ;
REMARK 3    H-BOND (X...Y)                 (A) :      ;
REMARK 3    H-BOND (X-H...Y)              (A) :      ;
REMARK 3
REMARK 3    CONFORMATIONAL TORSION ANGLE RESTRAINTS.
REMARK 3    SPECIFIED                      (DEGREES) :    ;
REMARK 3    PLANAR                        (DEGREES) :    ;
REMARK 3    STAGGERED                     (DEGREES) :    ;
REMARK 3    TRANSVERSE                    (DEGREES) :    ;
REMARK 3
REMARK 3    ISOTROPIC THERMAL FACTOR RESTRAINTS.    RMS    SIGMA
REMARK 3    MAIN-CHAIN BOND                (A**2) :      ;
REMARK 3    MAIN-CHAIN ANGLE               (A**2) :      ;
REMARK 3    SIDE-CHAIN BOND               (A**2) :      ;
REMARK 3    SIDE-CHAIN ANGLE              (A**2) :      ;
REMARK 3
REMARK 3    OTHER REFINEMENT REMARKS:

```

**Refinement using SHELXL-96**

This remark will be output by SHELXL-96 for direct submission to PDB. Structures done using earlier versions of SHELX will contain the same template, but with many of the data items containing "NULL".  
XX refers to the version used.

**Template**

```

REMARK      3  REFINEMENT.
REMARK      3  PROGRAM      : SHELXL-XX
REMARK      3  AUTHORS      : G.M.SHELDRICK
REMARK      3
REMARK      3  DATA USED IN REFINEMENT.
REMARK      3  RESOLUTION RANGE HIGH (ANGSTROMS) :
REMARK      3  RESOLUTION RANGE LOW  (ANGSTROMS) :
REMARK      3  DATA CUTOFF          (SIGMA(F)) :
REMARK      3  COMPLETENESS FOR RANGE          (%) :
REMARK      3  CROSS-VALIDATION METHOD          :
REMARK      3  FREE R VALUE TEST SET SELECTION :
REMARK      3
REMARK      3  FIT TO DATA USED IN REFINEMENT (NO CUTOFF) .
REMARK      3  R VALUE      (WORKING + TEST SET, NO CUTOFF) :
REMARK      3  R VALUE      (WORKING SET, NO CUTOFF) :
REMARK      3  FREE R VALUE      (NO CUTOFF) :
REMARK      3  FREE R VALUE TEST SET SIZE (% , NO CUTOFF) :
REMARK      3  FREE R VALUE TEST SET COUNT (NO CUTOFF) :
REMARK      3  TOTAL NUMBER OF REFLECTIONS (NO CUTOFF) :
REMARK      3
REMARK      3  FIT/AGREEMENT OF MODEL FOR DATA WITH F>4SIG(F) .
REMARK      3  R VALUE      (WORKING + TEST SET, F>4SIG(F)) :
REMARK      3  R VALUE      (WORKING SET, F>4SIG(F)) :
REMARK      3  FREE R VALUE      (F>4SIG(F)) :
REMARK      3  FREE R VALUE TEST SET SIZE (% , F>4SIG(F)) :
REMARK      3  FREE R VALUE TEST SET COUNT (F>4SIG(F)) :
REMARK      3  TOTAL NUMBER OF REFLECTIONS (F>4SIG(F)) :
REMARK      3
REMARK      3  NUMBER OF NON-HYDROGEN ATOMS USED IN REFINEMENT.
REMARK      3  PROTEIN ATOMS :
REMARK      3  NUCLEIC ACID ATOMS :
REMARK      3  HETEROGEN ATOMS :
REMARK      3  SOLVENT ATOMS :
REMARK      3
REMARK      3  MODEL REFINEMENT.
REMARK      3  OCCUPANCY SUM OF NON-HYDROGEN ATOMS :
REMARK      3  OCCUPANCY SUM OF HYDROGEN ATOMS :
REMARK      3  NUMBER OF DISCRETELY DISORDERED RESIDUES :
REMARK      3  NUMBER OF LEAST-SQUARES PARAMETERS :
REMARK      3  NUMBER OF RESTRAINTS :
REMARK      3
REMARK      3  RMS DEVIATIONS FROM RESTRAINT TARGET VALUES.
REMARK      3  BOND LENGTHS (A) :
REMARK      3  ANGLE DISTANCES (A) :
REMARK      3  SIMILAR DISTANCES (NO TARGET VALUES) (A) :
REMARK      3  DISTANCES FROM RESTRAINT PLANES (A) :
REMARK      3  ZERO CHIRAL VOLUMES (A**3) :
REMARK      3  NON-ZERO CHIRAL VOLUMES (A**3) :
REMARK      3  ANTI-BUMPING DISTANCE RESTRAINTS (A) :
REMARK      3  RIGID-BOND ADP COMPONENTS (A**2) :
REMARK      3  SIMILAR ADP COMPONENTS (A**2) :

```

REMARK 3 APPROXIMATELY ISOTROPIC ADPS (A\*\*2) :  
REMARK 3  
REMARK 3 BULK SOLVENT MODELING.  
REMARK 3 METHOD USED:  
REMARK 3  
REMARK 3 STEREOCHEMISTRY TARGET VALUES :  
REMARK 3 SPECIAL CASE:  
REMARK 3  
REMARK 3 OTHER REFINEMENT REMARKS:

**Refinement using TNT**

XXX refers to the release of TNT used for refinement.

**Template**

```

REMARK      3
REMARK      3 REFINEMENT.
REMARK      3   PROGRAM      : TNT XXX
REMARK      3   AUTHORS      : TRONRUD,TEN EYCK,MATTHEWS
REMARK      3
REMARK      3 DATA USED IN REFINEMENT.
REMARK      3 RESOLUTION RANGE HIGH (ANGSTROMS) :
REMARK      3 RESOLUTION RANGE LOW  (ANGSTROMS) :
REMARK      3 DATA CUTOFF              (SIGMA(F)) :
REMARK      3 COMPLETENESS FOR RANGE          (%) :
REMARK      3 NUMBER OF REFLECTIONS              :
REMARK      3
REMARK      3 USING DATA ABOVE SIGMA CUTOFF.
REMARK      3 CROSS-VALIDATION METHOD              :
REMARK      3 FREE R VALUE TEST SET SELECTION      :
REMARK      3 R VALUE          (WORKING + TEST SET) :
REMARK      3 R VALUE          (WORKING SET)       :
REMARK      3 FREE R VALUE                          :
REMARK      3 FREE R VALUE TEST SET SIZE          (%) :
REMARK      3 FREE R VALUE TEST SET COUNT          :
REMARK      3
REMARK      3 USING ALL DATA, NO SIGMA CUTOFF.
REMARK      3 R VALUE      (WORKING + TEST SET, NO CUTOFF) :
REMARK      3 R VALUE      (WORKING SET, NO CUTOFF)       :
REMARK      3 FREE R VALUE                          (NO CUTOFF) :
REMARK      3 FREE R VALUE TEST SET SIZE (% , NO CUTOFF) :
REMARK      3 FREE R VALUE TEST SET COUNT (NO CUTOFF) :
REMARK      3 TOTAL NUMBER OF REFLECTIONS (NO CUTOFF) :
REMARK      3
REMARK      3 NUMBER OF NON-HYDROGEN ATOMS USED IN REFINEMENT.
REMARK      3 PROTEIN ATOMS              :
REMARK      3 NUCLEIC ACID ATOMS         :
REMARK      3 OTHER ATOMS                :
REMARK      3
REMARK      3 WILSON B VALUE (FROM FCALC, A**2) :
REMARK      3
REMARK      3 RMS DEVIATIONS FROM IDEAL VALUES.      RMS      WEIGHT  COUNT
REMARK      3 BOND LENGTHS              (A) :      ;      ;
REMARK      3 BOND ANGLES                (DEGREES) :      ;      ;
REMARK      3 TORSION ANGLES             (DEGREES) :      ;      ;
REMARK      3 PSEUDOROTATION ANGLES    (DEGREES) :      ;      ;
REMARK      3 TRIGONAL CARBON PLANES    (A) :      ;      ;
REMARK      3 GENERAL PLANES            (A) :      ;      ;
REMARK      3 ISOTROPIC THERMAL FACTORS (A**2) :      ;      ;
REMARK      3 NON-BONDED CONTACTS       (A) :      ;      ;
REMARK      3
REMARK      3 INCORRECT CHIRAL-CENTERS (COUNT) :
REMARK      3
REMARK      3 BULK SOLVENT MODELING.
REMARK      3 METHOD USED :
REMARK      3 KSOL      :
REMARK      3 BSOL      :
REMARK      3
REMARK      3 RESTRAINT LIBRARIES.

```

```

REMARK 3 STEREOCHEMISTRY :
REMARK 3 ISOTROPIC THERMAL FACTOR RESTRAINTS :
REMARK 3
REMARK 3 OTHER REFINEMENT REMARKS:

```

**Example, more than one refinement programs used**

	1	2	3	4	5	6	7
	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890
REMARK	3						
REMARK	3	REFINEMENT.					
REMARK	3	PROGRAM	:	X-PLOR 3.1			
REMARK	3	AUTHORS	:	BRUNGER			
REMARK	3	PROGRAM	:	TNT 5E			
REMARK	3	AUTHORS	:	TRONRUD, TEN EYCK, MATTHEWS			
REMARK	3						
REMARK	3	DATA USED IN REFINEMENT.					
REMARK	3	[TNT REFINEMENT DETAILS HERE]					

**Non-diffraction studies**

Until standard refinement remarks are adopted for non-diffraction studies, their refinement details are given in REMARK 3, but its format will consist totally of free text beginning on the sixth line of the remark.

**Template**

```

1         2         3         4         5         6         7
123456789012345678901234567890123456789012345678901234567890
REMARK    3
REMARK    3 REFINEMENT.
REMARK    3 PROGRAM      :
REMARK    3 AUTHORS      :
REMARK    3
REMARK    3 FREE TEXT

```

**Example**

```

1         2         3         4         5         6         7
123456789012345678901234567890123456789012345678901234567890
REMARK    3
REMARK    3 REFINEMENT.
REMARK    3 PROGRAM      : X-PLOR 3.1
REMARK    3 AUTHORS      : BRUNGER
REMARK    3
REMARK    3 STRUCTURAL STATISTICS:
REMARK    3
REMARK    3                25 SA
REMARK    3                STRUCTURES  SAAVEMIN
REMARK    3 RMS DEVIATIONS FROM EXP. RESTRAINTS [A]
REMARK    3 NOE DISTANCE RESTRAINTS (1430)  0.0451 A      0.044 A
REMARK    3 DIHEDRAL ANGLE RESTRAINTS (130)  0.551 DEG     0.660 DEG
REMARK    3 DEVIATIONS FROM IDEAL GEOMETRY
REMARK    3 BONDS                0.004 A      0.004 A
REMARK    3 ANGLES              0.661 DEG     0.650 DEG
REMARK    3 IMPROPER            0.371 DEG     0.380 DEG
REMARK    3 X-PLOR ENERGIES (IN KCAL MOL-1) [B]
REMARK    3 ENOE                167          158
REMARK    3 ECDIH               2.6           3.4
REMARK    3 ENCS               0.01          0.01
REMARK    3 EREPEL              54            50
REMARK    3 EBOND              36            33
REMARK    3 EANGLE             263           256
REMARK    3 EIMPROPER            22            23
REMARK    3 ETOTAL             545           523
REMARK    3 ATOMIC RMS DIFFERENCES [C]
REMARK    3 BACKBONE (N, CA, C') + LIGAND ATOMS  0.53+/-0.09 A
REMARK    3 ALL HEAVY ATOMS                0.91+/-0.08 A

```



## REMARK 4 - 999

### Overview

REMARKs following the refinement remark consist of free text annotation, predefined boilerplate remarks, and token: value pair styled templates. PDB is beginning to organize the most often used remarks, and assign numbers and topics to them.

Presented here is the scheme being followed in the remark section of PDB files. The PDB expects to continue to adopt standard text or tables for certain remarks, as details are worked out.

### Record Format and Details

\* Non-standard remark annotations, or those with no clearly-defined topic or assigned remark number, appear with remark number 6 or greater, but less than remark number 100.

\* Note that A, B, N, X, Y, and Z are used to represent variables in the following examples.

\* As with all other remarks, the first line of each remark is empty and is used as a spacer.

## REMARK 4, Format

Entries released after April 15, 1996 will comply with Format Version 2.0, described in this document. Conversion of older entries to this format will begin in the fall of 1996.

Entries conforming to the format described in this or future PDB Contents Guides will have a remark of the following form within them:

### Template

```

1           2           3           4           5           6           7
123456789012345678901234567890123456789012345678901234567890
REMARK      4
REMARK      4 XXXX COMPLIES WITH FORMAT V. N.M, DD-MMM-YYYY
```

XXXX refers to the ID code of the entry.

N.M refers to the version number.

DD-MMM-YYYY refers to the release date of that version of the format. DD is a number 01 through 31, MMM is a 3 letter abbreviation for the month, and YYYY is the year.

### Example

```

1           2           3           4           5           6           7
123456789012345678901234567890123456789012345678901234567890
REMARK      4
REMARK      4 1ABC COMPLIES WITH FORMAT V. 2.0, 16-FEB-1996
```

## REMARK 5, Warning

Remark 5 repeats information presented on the CAVEAT record, which warns of severe errors in an entry. It also presents depositors' remarks of a cautionary nature, such as noting regions of poorly defined density.

**Template**

```

      1           2           3           4           5           6           7
123456789012345678901234567890123456789012345678901234567890
REMARK      5
REMARK      5 WARNING
REMARK      5 XXXX: FREE TEXT GOES HERE.

```

XXXX refers to the ID code of the entry.

**Example**

```

      1           2           3           4           5           6           7
123456789012345678901234567890123456789012345678901234567890
REMARK      5
REMARK      5 WARNING
REMARK      5 XXXX: THE CRYSTAL TRANSFORMATION IS IN ERROR BUT IS
REMARK      5 UNCORRECTABLE AT THIS TIME.

```

**REMARK 6 - 99, not assigned**

Non-standard remark annotations, or those with no clearly defined topic or assigned remark number appear with remark number 6 or greater, but less than remark number 100.

**REMARK 100 - 199, Nucleic acids**

These remarks are used in nucleic acid structures.

**For modified residues****Template**

```

      1           2           3           4           5           6           7
123456789012345678901234567890123456789012345678901234567890
REMARK 101
REMARK 101 RESIDUE   X Y   N HAS XXX   BONDED TO AB.
REMARK 101 RESIDUE   X Y   N HAS XXX   BONDED TO AB.

```

X is the modified residue name, Y is the chain identifier, N is the sequence number, XXX is the name of the modifier, A is the atom name and B the sequence number of the atom carrying the modifier.

**Example**

```

      1           2           3           4           5           6           7
123456789012345678901234567890123456789012345678901234567890
REMARK 101
REMARK 101 RESIDUE   G A   4 HAS CH3   BONDED TO O6.
REMARK 101 RESIDUE   G B  16 HAS CH3   BONDED TO O6.

```

**For base mispairings****Template**

```

      1           2           3           4           5           6           7
123456789012345678901234567890123456789012345678901234567890
REMARK 102
REMARK 102 BASES    A B NN AND   X Y ZZ ARE MISPAIRED.
REMARK 102 BASES    A B NN AND   X Y ZZ ARE MISPAIRED.
REMARK 102 ALL OTHER HYDROGEN BONDS BETWEEN BASE PAIRS IN THIS ENTRY
REMARK 102 FOLLOW THE CONVENTIONAL WATSON-CRICK HYDROGEN BONDING
REMARK 102 PATTERN AND THEY HAVE NOT BEEN PRESENTED ON *CONNECT*
REMARK 102 RECORDS IN THIS ENTRY.

```

A is the residue name, B the chain identifier, and NN the sequence number of first base, X is the residue name, Y the chain id, and ZZ the sequence number of the second base.

**Example**

```

      1      2      3      4      5      6      7
123456789012345678901234567890123456789012345678901234567890
REMARK 102
REMARK 102 BASES   G A   4 AND   A B  21 ARE MISPAIRED.
REMARK 102 BASES   A A   9 AND   G B  16 ARE MISPAIRED.
REMARK 102 ALL OTHER HYDROGEN BONDS BETWEEN BASE PAIRS IN THIS ENTRY
REMARK 102 FOLLOW THE CONVENTIONAL WATSON-CRICK HYDROGEN BONDING
REMARK 102 PATTERN AND THEY HAVE NOT BEEN PRESENTED ON *CONECT*
REMARK 102 RECORDS IN THIS ENTRY.

```

**For structures containing inosine**

Inosine is treated like a standard residue, however, entries containing inosine also include remarks 103 and 104.

**Template**

```

      1      2      3      4      5      6      7
123456789012345678901234567890123456789012345678901234567890
REMARK 103
REMARK 103 THERE ARE NON-WATSON-CRICK HYDROGEN BONDS BETWEEN THE
REMARK 103 FOLLOWING ATOMS:
REMARK 103 AB   I X   N   AND AB   Z X  NN
REMARK 103 AB   I X   N   AND AB   Z X  NN
REMARK 103 ALL OTHER HYDROGEN BONDS BETWEEN BASE PAIRS IN THIS ENTRY
REMARK 103 FOLLOW THE CONVENTIONAL WATSON-CRICK HYDROGEN BONDING
REMARK 103 PATTERN AND THEY HAVE NOT BEEN PRESENTED ON *CONECT*
REMARK 103 RECORDS IN THIS ENTRY.

```

AB is the atom name, I the residue name inosine, X the chain identifier, and N the sequence number of inosine, and AB is the atom name, Z the residue name, X the chain identifier, and NN the sequence number of the base which is paired with inosine.

**Template**

```

      1      2      3      4      5      6      7
123456789012345678901234567890123456789012345678901234567890
REMARK 104
REMARK 104 RESIDUE I X   N IS INOSINE.
REMARK 104 RESIDUE I X   N IS INOSINE.

```

X is the chain identifier and N the sequence number.

**Example**

```

      1      2      3      4      5      6      7
123456789012345678901234567890123456789012345678901234567890
REMARK 103
REMARK 103 THERE ARE NON-WATSON-CRICK HYDROGEN BONDS BETWEEN THE
REMARK 103 FOLLOWING ATOMS:
REMARK 103 N1   I A   1   AND N3   C B  16
REMARK 103 O6   I A   1   AND N4   C B  16
REMARK 103 N1   I A   3   AND N3   C B  14
REMARK 103 O6   I A   3   AND N4   C B  14
REMARK 103 ALL OTHER HYDROGEN BONDS BETWEEN BASE PAIRS IN THIS ENTRY
REMARK 103 FOLLOW THE CONVENTIONAL WATSON-CRICK HYDROGEN BONDING

```

REMARK 103 PATTERN AND THEY HAVE NOT BEEN PRESENTED ON CONECT  
 REMARK 103 RECORDS IN THIS ENTRY.  
 REMARK 104  
 REMARK 104 RESIDUE I A 1 IS INOSINE.  
 REMARK 104 RESIDUE I A 3 IS INOSINE.

### For non-mismatched structures

#### Template

1	2	3	4	5	6	7
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890
REMARK 106						
REMARK 106 THE HYDROGEN BONDS BETWEEN BASE PAIRS IN THIS ENTRY FOLLOW						
REMARK 106 THE CONVENTIONAL WATSON-CRICK HYDROGEN BONDING PATTERN.						
REMARK 106 THEY HAVE NOT BEEN PRESENTED ON CONECT RECORDS IN THIS						
REMARK 106 ENTRY.						

### REMARK 200-250, Experimental Details

Remarks in this range present the data collection details for the data which resulted in the refinement statistics of REMARK 3. They provide information on the structure determination experiment, which may have been done by diffraction, NMR, theoretical modelling, or some other technique.

The "NULL" value will be used if the data for a token is not supplied by the depositor.

#### REMARK 200, X-ray Diffraction Experimental Details

To be used for single crystal, fiber, or polycrystalline X-ray diffraction experiments.

#### Template

1	2	3	4	5	6	7
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890
REMARK 200						
REMARK 200 EXPERIMENTAL DETAILS.						
REMARK 200 EXPERIMENT TYPE : X-RAY DIFFRACTION						
REMARK 200 DATE OF DATA COLLECTION :						
REMARK 200 TEMPERATURE (KELVIN) :						
REMARK 200 PH :						
REMARK 200 NUMBER OF CRYSTALS USED :						
REMARK 200						
REMARK 200 SYNCHROTRON (Y/N) :						
REMARK 200 RADIATION SOURCE :						
REMARK 200 BEAMLINE :						
REMARK 200 X-RAY GENERATOR MODEL :						
REMARK 200 MONOCHROMATIC OR LAUE (M/L) :						
REMARK 200 WAVELENGTH OR RANGE (A) :						
REMARK 200 MONOCHROMATOR :						
REMARK 200 OPTICS :						
REMARK 200						
REMARK 200 DETECTOR TYPE :						
REMARK 200 DETECTOR MANUFACTURER :						
REMARK 200 INTENSITY-INTEGRATION SOFTWARE :						
REMARK 200 DATA SCALING SOFTWARE :						
REMARK 200						
REMARK 200 NUMBER OF UNIQUE REFLECTIONS :						
REMARK 200 RESOLUTION RANGE HIGH (A) :						
REMARK 200 RESOLUTION RANGE LOW (A) :						

```

REMARK 200 REJECTION CRITERIA :
REMARK 200
REMARK 200 OVERALL.
REMARK 200 COMPLETENESS FOR RANGE (%) :
REMARK 200 DATA REDUNDANCY :
REMARK 200 MERGING R VALUE (I) :
REMARK 200 <I/SIGMA(I)> FOR THE DATA SET :
REMARK 200
REMARK 200 IN THE HIGHEST RESOLUTION SHELL.
REMARK 200 HIGHEST RESOLUTION SHELL, RANGE HIGH (A) :
REMARK 200 HIGHEST RESOLUTION SHELL, RANGE LOW (A) :
REMARK 200 COMPLETENESS FOR SHELL (%) :
REMARK 200 DATA REDUNDANCY IN SHELL :
REMARK 200 MERGING R VALUE FOR SHELL (I) :
REMARK 200 <I/SIGMA(I)> FOR SHELL :
REMARK 200
REMARK 200 METHOD USED TO DETERMINE THE STRUCTURE:
REMARK 200 SOFTWARE USED :
REMARK 200 STARTING MODEL FOR MOLECULAR REPLACEMENT:
REMARK 200
REMARK 200 REMARK:

```

## Remark 205, Fiber Diffraction, Fiber Sample Experiment Details

Remark specific to X-ray fiber diffraction on fiber samples only.

### Template

```

1 2 3 4 5 6 7
123456789012345678901234567890123456789012345678901234567890
REMARK 205
REMARK 205 THESE COORDINATES WERE GENERATED FROM FIBER DIFFRACTION
REMARK 205 DATA. PROTEIN DATA BANK CONVENTIONS REQUIRE THAT CRYST1
REMARK 205 AND SCALE RECORDS BE INCLUDED, BUT THE VALUES OF THESE
REMARK 205 RECORDS ARE MEANINGLESS.

```

## Remarks 210 and 215, NMR Experiment Details

Remarks specific to NMR studies.

### Template

```

1 2 3 4 5 6 7
123456789012345678901234567890123456789012345678901234567890
REMARK 210
REMARK 210 EXPERIMENTAL DETAILS.
REMARK 210 EXPERIMENT TYPE : NMR
REMARK 210
REMARK 210 FREE TEXT GOES HERE.

```

### Template

```

1 2 3 4 5 6 7
123456789012345678901234567890123456789012345678901234567890
REMARK 215
REMARK 215 NMR STUDY
REMARK 215 THESE COORDINATES WERE GENERATED FROM SOLUTION NMR DATA.
REMARK 215 PROTEIN DATA BANK CONVENTIONS REQUIRE THAT CRYST1 AND
REMARK 215 SCALE RECORDS BE INCLUDED, BUT THE VALUES OF THESE
REMARK 215 RECORDS ARE MEANINGLESS.

```

## Remarks 220 and 225, Theoretical Modelling Experiment Details

Remark specific to theoretical model study.

### Template

1	2	3	4	5	6	7
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890
REMARK 220						
REMARK 220	EXPERIMENTAL DETAILS.					
REMARK 220	EXPERIMENT TYPE: THEORETICAL MODELLING					
REMARK 220						
REMARK 220	FREE TEXT GOES HERE.					

### Template

1	2	3	4	5	6	7
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890
REMARK 225						
REMARK 225	THE COORDINATES IN THIS ENTRY REPRESENT A MODEL STRUCTURE.					
REMARK 225	PROTEIN DATA BANK CONVENTIONS REQUIRE THAT CRYST1 AND					
REMARK 225	SCALE RECORDS BE INCLUDED, BUT THE VALUES ON THESE					
REMARK 225	RECORDS ARE MEANINGLESS.					

## Remark 230, Neutron Diffraction Experiment Details

Remark specific to neutron diffraction study.

### Template

1	2	3	4	5	6	7
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890
REMARK 230						
REMARK 230	EXPERIMENTAL DETAILS.					
REMARK 230	EXPERIMENT TYPE : NEUTRON DIFFRACTION					
REMARK 230	DATE OF DATA COLLECTION :					
REMARK 230	TEMPERATURE (KELVIN) :					
REMARK 230	PH :					
REMARK 230						
REMARK 230	NEUTRON SOURCE :					
REMARK 230	BEAMLINE :					
REMARK 230	WAVELENGTH OR RANGE (A) :					
REMARK 230	MONOCHROMATOR :					
REMARK 230	OPTICS :					
REMARK 230						
REMARK 230	DETECTOR TYPE :					
REMARK 230	DETECTOR MANUFACTURER :					
REMARK 230	INTENSITY-INTEGRATION SOFTWARE :					
REMARK 230	DATA SCALING SOFTWARE :					
REMARK 230						
REMARK 230	NUMBER OF UNIQUE REFLECTIONS :					
REMARK 230	RESOLUTION RANGE HIGH (A) :					
REMARK 230	RESOLUTION RANGE LOW (A) :					
REMARK 230	REJECTION CRITERIA :					
REMARK 230						
REMARK 230	OVERALL.					
REMARK 230	COMPLETENESS FOR RANGE (%) :					
REMARK 230	DATA REDUNDANCY :					
REMARK 230	MERGING R VALUE (I) :					
REMARK 230	<I/SIGMA(I)> FOR THE DATA SET :					
REMARK 230						

```

REMARK 230 IN THE HIGHEST RESOLUTION SHELL.
REMARK 230 HIGHEST RESOLUTION SHELL, RANGE HIGH (A) :
REMARK 230 HIGHEST RESOLUTION SHELL, RANGE LOW (A) :
REMARK 230 COMPLETENESS FOR SHELL (%) :
REMARK 230 DATA REDUNDANCY IN SHELL :
REMARK 230 MERGING R VALUE FOR SHELL (I) :
REMARK 230 <I/SIGMA(I)> FOR SHELL :
REMARK 230
REMARK 230 METHOD USED TO DETERMINE THE STRUCTURE:
REMARK 230 SOFTWARE USED :
REMARK 230 STARTING MODEL:
REMARK 230
REMARK 230 REMARK:

```

## Remark 240, Electron Diffraction Experiment Details

Remark specific to electron diffraction study.

### Template

	1	2	3	4	5	6	7
123456789012345678901234567890123456789012345678901234567890							
REMARK 240							
REMARK 240 EXPERIMENTAL DETAILS.							
REMARK 240 EXPERIMENT TYPE						:	ELECTRON DIFFRACTION
REMARK 240 DATE OF DATA COLLECTION						:	
REMARK 240							
REMARK 240 FREE TEXT GOES HERE.							

## Remark 250, Other Type of Experiment Details

Remark specific to other kinds of studies, not listed above. XXXX stands for the experimental technique used.

### Template

	1	2	3	4	5	6	7
123456789012345678901234567890123456789012345678901234567890							
REMARK 250							
REMARK 250 EXPERIMENTAL DETAILS.							
REMARK 250 EXPERIMENT TYPE						:	XXXX
REMARK 250 DATE OF DATA COLLECTION						:	
REMARK 250							
REMARK 250 FREE TEXT GOES HERE.							

**REMARK 280, Crystal**

Remark 280 presents crystal information. The solvent content and Matthews coefficient are always provided for protein and polypeptide crystals. The crystallization conditions are free text and optional.

**Template**

1	2	3	4	5	6	7
12345678901	2345678901	2345678901	2345678901	2345678901	2345678901	234567890
REMARK 280						
REMARK 280 CRYSTAL						
REMARK 280 SOLVENT CONTENT, VS (%) :						
REMARK 280 MATTHEWS COEFFICIENT, VM (ANGSTROMS**3/DA) :						
REMARK 280						
REMARK 280 CRYSTALLIZATION CONDITIONS:						

**REMARK 300, Biomolecule**

Description of the biologically functional molecule (biomolecule) in free text.

**Template**

1	2	3	4	5	6	7
12345678901	2345678901	2345678901	2345678901	2345678901	2345678901	234567890
REMARK 300						
REMARK 300 BIOMOLECULE						
REMARK 300 FREE TEXT DESCRIPTION OF THE BIOLOGICALLY FUNCTIONAL						
REMARK 300 MOLECULE.						

**REMARK 320, Non-IUPAC Names**

This remark is used when IUPAC-IUB rules are not strictly followed in naming side-chain atoms.

**Template**

1	2	3	4	5	6	7
12345678901	2345678901	2345678901	2345678901	2345678901	2345678901	234567890
REMARK 320						
REMARK 320 NON-IUPAC						
REMARK 320 FREE TEXT GOES HERE.						

**Example**

1	2	3	4	5	6	7
12345678901	2345678901	2345678901	2345678901	2345678901	2345678901	234567890
REMARK 320						
REMARK 320 NON-IUPAC						
REMARK 320 BY REQUEST OF THE DEPOSITOR, THE PROTEIN DATA BANK HAS NOT						
REMARK 320 APPLIED THE IUPAC-IUB RECOMMENDATIONS REGARDING THE						
REMARK 320 DESIGNATION OF BRANCHES 1 AND 2 OF SIDE-CHAIN ATOMS IN						
REMARK 320 RESIDUES ARG, ASP, GLU, LEU, PHE, TYR, AND VAL TO THIS						
REMARK 320 ENTRY.						



**REMARK 350, Symmetry**

Information about crystallographic and non-crystallographic symmetry. Note that NNNMMM gives the symmetry operations.

For more information about symmetry operators see Appendix 1.

**Template**

```

123456789012345678901234567890123456789012345678901234567890
REMARK 350
REMARK 350 SYMMETRY
REMARK 350 NON-CRYSTALLOGRAPHIC SYMMETRY TRANSFORMATIONS: Number
REMARK 350
REMARK 350 THE TRANSFORMATIONS PRESENTED ON THE MTRIX RECORDS BELOW
REMARK 350 DESCRIBE NON-CRYSTALLOGRAPHIC RELATIONSHIPS AMONG THE
REMARK 350 VARIOUS DOMAINS IN THIS ENTRY. APPLYING THE APPROPRIATE
REMARK 350 MTRIX TRANSFORMATION TO THE RESIDUES LISTED FIRST WILL
REMARK 350 YIELD APPROXIMATE COORDINATES FOR THE RESIDUES LISTED
REMARK 350 SECOND. CHAIN IDENTIFIERS GIVEN AS "?" REFER TO CHAINS
REMARK 350 FOR WHICH ATOMS ARE NOT FOUND IN THIS ENTRY.
REMARK 350
REMARK 350          APPLIED TO          TRANSFORMED TO
REMARK 350 TRANSFORM CHAIN RESIDUES      CHAIN RESIDUES      RMSD
REMARK 350 MS          A      1 .. 999      B      1 .. 999      0.32
REMARK 350
REMARK 350      HERE  M -> TRANSFORMATION MATRIX (MTRIX  S)
REMARK 350          S -> MATRIX SERIAL NUMBER
REMARK 350
REMARK 350 CRYSTALLOGRAPHIC SYMMETRY TRANSFORMATIONS: Number
REMARK 350
REMARK 350 THE TRANSFORMATIONS PRESENTED ON THE SYMTRY RECORDS
REMARK 350 LISTED IN THIS REMARK DESCRIBE THE CRYSTALLOGRAPHIC
REMARK 350 SYMMETRY TRANSFORMATIONS NECESSARY TO GENERATE THE
REMARK 350 BIOLOGICALLY SIGNIFICANT OLIGOMERIZATION STATE OF
REMARK 350 THE MOLECULE. APPLYING THE APPROPRIATE SYMTRY
REMARK 350 TRANSFORMATION TO THE RESIDUES LISTED WILL GENERATE
REMARK 350 THE SYMMETRY-RELATED DOMAIN.
REMARK 350
REMARK 350 OPERATOR CRYSTALLOGRAPHIC TRANSLATION FRACTIONAL
REMARK 350 NUMBER SYMMETRY OPERATOR VECTOR TRANSLATIONS
REMARK 350 001 -X,-Y,Z 545 0.0 -1.0 0.0
REMARK 350
REMARK 350 TRANSFORM          APPLIED TO
REMARK 350 NNNMMMS CHAIN(S) RESIDUES
REMARK 350 0015451 A 1 .. 98
REMARK 350
REMARK 350      HERE NNN -> OPERATOR NUMBER
REMARK 350          MMM -> TRANSLATION VECTOR
REMARK 350          S -> SMTRY MATRIX SERIAL NUMBER
REMARK 350
REMARK 350 CORRESPONDING SYMTRY MATRIX FOR TRANSFORM 0015551
REMARK 350
REMARK 350 SMTRY1 1 -1.000000 0.000000 0.000000 0.000000
REMARK 350 SMTRY2 1 0.000000 -1.000000 0.000000 -1.000000
REMARK 350 SMTRY3 1 0.000000 0.000000 1.000000 0.000000

```

**REMARK 400, Compound**

Further details on the macromolecular contents of the entry.

**Template**

```
      1      2      3      4      5      6      7
123456789012345678901234567890123456789012345678901234567890
REMARK 400
REMARK 400 COMPOUND
REMARK 400 FREE TEXT GOES HERE.
```

**REMARK 450, Source**

Further details on the biological source of the macromolecular contents of the entry.

**Template**

```
      1      2      3      4      5      6      7
123456789012345678901234567890123456789012345678901234567890
REMARK 450
REMARK 450 SOURCE
REMARK 450 FREE TEXT GOES HERE.
```

**REMARK 500, Geometry and Stereochemistry**

Further details on the stereochemistry of the structure. Remark 500 may be used to name peptide bonds that deviate from trans conformation by more than 30 degrees. This remark is generated by PDB, but may also be provided by the depositor. Other text may appear here as well.

**Template**

```
      1      2      3      4      5      6      7
123456789012345678901234567890123456789012345678901234567890
REMARK 500
REMARK 500 GEOMETRY AND STEREOCHEMISTRY
REMARK 500 FREE TEXT GOES HERE.
```

**Example**

```
      1      2      3      4      5      6      7
123456789012345678901234567890123456789012345678901234567890
REMARK 500
REMARK 500 STEREOCHEMISTRY AND GEOMETRY
REMARK 500 PEPTIDE BOND DEVIATES SIGNIFICANTLY FROM TRANS CONFORMATION
REMARK 500      VAL A   123  -  GLN A   124                      OMEGA = 221.48

REMARK 500
REMARK 500 STEREOCHEMISTRY AND GEOMETRY
REMARK 500 MANY OF THE WATER MOLECULES APPEAR TO BE ASSOCIATED WITH
REMARK 500 A SYMMETRY-RELATED MOLECULE.
```

**REMARK 550, SEGID**

Description of the segment identifiers used in ATOM/HETATM.

**Template**

```
      1      2      3      4      5      6      7
123456789012345678901234567890123456789012345678901234567890
REMARK 550
REMARK 550 SEGID
REMARK 550 FREE TEXT GOES HERE.
```

**REMARK 600, Heterogens**

Further details on the heterogens in the entry.

**Template**

```
      1      2      3      4      5      6      7
123456789012345678901234567890123456789012345678901234567890
REMARK 600
REMARK 600 HETEROGENS
REMARK 600 FREE TEXT GOES HERE.
```

**REMARK 650, Helix**

Further details on the helix contents of the entry.

**Template**

```
      1      2      3      4      5      6      7
123456789012345678901234567890123456789012345678901234567890
REMARK 650
REMARK 650 HELIX
REMARK 650 FREE TEXT GOES HERE.
```

**REMARK 700, Sheet**

Further details on the sheet contents of the structure. Several boilerplate remarks are included here.

**Template**

```
      1      2      3      4      5      6      7
123456789012345678901234567890123456789012345678901234567890
REMARK 700
REMARK 700 SHEET
REMARK 700 FREE TEXT GOES HERE.

REMARK 700
REMARK 700 SHEET
REMARK 700 THE SHEET STRUCTURE OF THIS MOLECULE IS BIFURCATED.  IN
REMARK 700 ORDER TO REPRESENT THIS FEATURE IN THE SHEET RECORDS BELOW,
REMARK 700 TWO SHEETS ARE DEFINED.  STRANDS N1, N2, N3 AND N4 OF SHEET
REMARK 700 XXX AND XXX ARE IDENTICAL.
```

REMARK 700  
 REMARK 700 SHEET  
 REMARK 700 THE SHEET PRESENTED AS XXX ON SHEET RECORDS BELOW IS  
 REMARK 700 ACTUALLY AN N-STRANDED BETA-BARREL. THIS IS  
 REMARK 700 REPRESENTED BY A N+1-STRANDED SHEET IN WHICH THE FIRST AND  
 REMARK 700 LAST STRANDS ARE IDENTICAL.  
 REMARK 700

REMARK 700 SHEET  
 REMARK 700 THERE ARE SEVERAL BIFURCATED SHEETS IN THIS STRUCTURE.  
 REMARK 700 EACH IS REPRESENTED BY TWO SHEETS WHICH HAVE ONE OR MORE  
 REMARK 700 IDENTICAL STRANDS.  
 REMARK 700 SHEETS XXX AND XXX REPRESENT ONE BIFURCATED SHEET.  
 REMARK 700 SHEETS XXX AND XXX REPRESENT ONE BIFURCATED SHEET.

N1, N2, N3 and N4 represent strand numbers, and XXX represents sheet identifiers.

When the remark for several bifurcated sheets is used, its last line is repeated for the appropriate number of bifurcated sheets, as shown in the last template above.

## REMARK 750, Turn

Further details on the turns.

### Template

1	2	3	4	5	6	7
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890
REMARK 750						
REMARK 750 TURN						
REMARK 750 FREE TEXT GOES HERE.						

## REMARK 800, Site

Further details on the helix contents of the entry.

### Template

1	2	3	4	5	6	7
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890
REMARK 800						
REMARK 800 SITE						
REMARK 800 FREE TEXT GOES HERE.						

## REMARK 850, Revisions to Deposited Coordinates, Before Release

### Template

1	2	3	4	5	6	7
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890
REMARK 850						
REMARK 850 REVISION BEFORE RELEASE						
REMARK 850 ORIGINAL DEPOSITION REVISED PRIOR TO RELEASE						
REMARK 850 TRACKING NUMBER: TNNNN, DATE REVISED: DD-MMM-YYYY						

DD is a number 01 through 31, MMM is a 3 letter abbreviation for the month, and YYYY is the year.

**Example**

1	2	3	4	5	6	7
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890
REMARK 850						
REMARK 850 REVISION BEFORE RELEASE						
REMARK 850 ORIGINAL DEPOSITION REVISED PRIOR TO RELEASE						
REMARK 850 TRACKING NUMBER: T5914, DATE REVISED: 10-MAR-1995						

**REMARK 860, Correction, After Release**

Further details on corrections that have been made to the PDB entry, as referred to in the REVDAT record.

**Template**

1	2	3	4	5	6	7
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890
REMARK 860						
REMARK 860 CORRECTION AFTER RELEASE						
REMARK 860 FREE TEXT GOES HERE.						

**Example**

1	2	3	4	5	6	7
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890
REMARK 860						
REMARK 860 CORRECTION						
REMARK 860 CORRECT RESIDUE IDENTIFICATION ON SITE RECORDS. ADD						
REMARK 860 RESIDUE TO SITE RECORDS. 15-JUL-81.						
REMARK 860						
REMARK 860 CORRECT DATES IN REMARKS 7 AND 16. 15-JAN-82.						
REMARK 860						
REMARK 860 CORRECT ATOM NAME FOR ATOM 6 FROM CG2 TO CG1. 07-MAR-83.						
REMARK 860						
REMARK 860 CHANGE RESIDUE 122 FROM ASN TO ASP. ADD REFERENCE.						
REMARK 860 12-MAY-83.						
REMARK 860						
REMARK 860 INSERT REVDAT RECORDS. 30-SEP-83.						
REMARK 860						
REMARK 860 CORRECT CODEN FOR REFERENCE 1. 27-OCT-83.						

**REMARK 900, Related Entries**

This remark gives ID codes of PDB files related to the entry. These may include coordinate entries deposited as a related set, the structure factor or NMR restraint file related to the entry, or the file containing the biologically functional molecule ("biomolecule") generated by the PDB from symmetry records.

**Template**

1	2	3	4	5	6	7
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890
REMARK 900						
REMARK 900 RELATED ENTRIES						
REMARK 900 FREE TEXT REMARK HERE.						

**Example**

	1	2	3	4	5	6	7
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890
REMARK 900							
REMARK 900 RELATED ENTRIES							
REMARK 900 XXXX AND XXXX ARE RELATED TO THIS ENTRY AS CRYSTAL FORM II							
REMARK 900 AND CRYSTAL FORM III, RESPECTIVELY.							
REMARK 900							
REMARK 900 RELATED ENTRIES							
REMARK 900 THE STRUCTURE FACTORS FOR THIS EXPERIMENT ARE AVAILABLE AS							
REMARK 900 PDB FILE RXXXXSF.ENT.							
REMARK 900							
REMARK 900 RELATED ENTRIES							
REMARK 900 THE LIST OF EXPERIMENTAL RESTRAINTS IS AVAILABLE AS PDB							
REMARK 900 FILE XXXX.MR.							
REMARK 900							
REMARK 900 RELATED ENTRIES							
REMARK 900 THE BIOMOLECULE IS AVAILABLE AS PDB FILE BIOXXXX.PDB.							

XXXX stands for the PDB ID code.

**REMARK 999 Sequence**

Further details on the sequence.

**Template**

	1	2	3	4	5	6	7
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890
REMARK 999							
REMARK 999 SEQUENCE							
REMARK 999 FREE TEXT GOES HERE.							

---

### 3. Primary Structure Section

The primary structure section of a PDB file contains the sequence of residues in each chain of the macromolecule. Embedded in these records are chain identifiers and sequence numbers that allow other records to link into the sequence.

---

#### MODRES

##### Overview

The MODRES record provides descriptions of modifications (e.g., chemical or post-translational) to protein and nucleic acid residues. Included are a mapping between residue names given in a PDB entry and standard residues.

##### Record Format

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"MODRES"	
8 - 11	IDcode	idCode	ID code of this entry.
13 - 15	LString	resName	Residue name used.
17	Character	chainID	Chain identifier.
19 - 22	Integer	seqNum	Sequence number.
23	AChar	iCode	Insertion code.
25 - 27	Residue name	stdRes	Standard residue name.
30 - 70	String	comment	Description of the residue modification.

##### Details

\* Residues modified post-translationally or by design are described in MODRES records. In those cases where PDB has opted to use a non-standard residue name for the residue, MODRES also provides a mapping to the precursor standard residue name.

\* MODRES is mandatory when PDB uses a non-standard residue name for a modified standard group.

\* Examples of some modification descriptions:

- Glycosylation site
- Post-translational modification
- Designed chemical modification
- Phosphorylation site
- Blocked N-terminus
- Aminated C-terminus
- D-configuration
- Reduced peptide bond

\* MODRES is not included when ATOM records are not provided for the modified residue.

\* When a standard residue name is used to describe a modified site, resName (columns 13-15) and stdRES (columns 25-27) contain the same value.

**Verification/Validation/Value Authority Control**

MODRES is generated by the PDB.

**Relationships to Other Record Types**

MODRES maps ATOM and HETATM records to the standard residue names.

**Example**

```

      1      2      3      4      5      6      7
123456789012345678901234567890123456789012345678901234567890
MODRES 1ABC ASN A   22A ASN  GLYCOSYLATION SITE

MODRES 2ABC TTQ A   50A TRP  POST-TRANSLATIONAL MODIFICATION

MODRES 3ABC ALA A   32  ALA  POST-TRANSLATIONAL MODIFICATION, D-ALANINE
MODRES 3ABC ALA B   32  ALA  POST-TRANSLATIONAL MODIFICATION, D-ALANINE

```

**Known Problems**

Mapping between SEQRES and MODRES residue numbers when the numbering is non-sequential has to be constructed using DBREF, SEQRES, and SEQADV.

**DBREF****Overview**

The DBREF record provides cross-reference links between PDB sequences and the corresponding database entry or entries. A cross reference to the sequence database is mandatory for each peptide chain with a length greater than ten (10) residues. For nucleic acid entries a DBREF record pointing to the Nucleic Acid Database (NDB) is mandatory when the corresponding entry exists in NDB.

**Record Format**

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"DBREF "	
8 - 11	IDcode	idCode	ID code of this entry.
13	Character	chainID	Chain identifier.
15 - 18	Integer	seqBegin	Initial sequence number of the PDB sequence segment.
19	AChar	insertBegin	Initial insertion code of the PDB sequence segment.
21 - 24	Integer	seqEnd	Ending sequence number of the PDB sequence segment.
25	AChar	insertEnd	Ending insertion code of the PDB sequence segment.
27 - 32	LString	database	Sequence database name. "PDB" when a corresponding sequence database entry has not been identified.



34 - 41	LString	dbAccession	Sequence database accession code. For GenBank entries, this is the NCBI gi number.
43 - 54	LString	dbIdCode	Sequence database identification code. For GenBank entries, this is the accession code.
56 - 60	Integer	seqBegin	Initial sequence number of the database segment.
61	AChar	insBegpdb	Insertion code of initial residue of the segment, if PDB is the reference
63 - 67	Integer	seqEnd	Ending sequence number of the database segment.
68	AChar	insEndpdb	Insertion code of the ending residue of the segment, if PDB is the reference

**Details**

\* PDB entries contain multi-chain molecules with sequences that may be wild type, variant, or synthetic. Sequences may also have been modified through site-directed mutagenesis experiments (engineered). A number of PDB entries report structures of domains cleaved from larger molecules.

\* The DBREF record was designed to account for these differences by providing explicit correlations between contiguous segments of sequences as given in the PDB ATOM records and the sequence database entry. Several cases are easily represented by means of pointers between the databases using DBREF. PDB entries containing heteropolymers are linked to different sequence database entries. In some cases, such as those PDB entries containing immunoglobulin Fab fragments, each chain is linked to two different SWISS-PROT, PIR, and/or GenBank entries. This facility is needed because these databases represent sequences for the various immunoglobulin domains as separate entries. DBREF also is able to represent molecules engineered by altering the gene (fusing genes, altering sequences, creating chimeras, or circularly permuting sequences). This design has the additional advantage that it will be possible to construct pointers to other relevant databases such as the Nucleic Acid Database, BioMagResBank, and databases describing sequence motifs (e.g., PROSITE, BLOCKS).

\* Database names and their abbreviations as used on DBREF records.

Database name	database (code in columns 27 - 32)
BioMagResBank	BMRB
BLOCKS	BLOCKS
European Molecular Biology Laboratory	EMBL
GenBank	GB
Genome Data Base	GDB
Nucleic Acid Database	NDB
PROSITE	PROSIT
Protein Data Bank	PDB
Protein Identification Resource	PIR
SWISS-PROT	SWS
TREMBL	TREMBL

\* When no sequence numbers are given (columns 19 - 30 and 56 - 80), then the mapping is between database entries rather than segments within an entry. For example, this is normally used to point to the related NDB entry.

\* DBREF records present sequence correlations between PDB ATOM records and corresponding PIR, GenBank, or SWISS-PROT, etc. entries.

\* PDB does not guarantee that all possible references to the listed databases will be provided. In most cases, only one reference to a sequence database will be provided.

\* PDB entries containing chains for which residues are missing primarily due to disorder contain several DBREF records, each linking an observed sequence segment to a sequence database entry.

\* If no reference is found in the sequence databases, then the PDB entry itself is given as the reference.

\* For nucleic acid entries a DBREF record pointing to the Nucleic Acid Database (NDB) is mandatory when the corresponding entry exists in NDB.

\* Selection of the appropriate sequence database entry or entries to be linked to a PDB entry is done on the basis of the sequence and its biological source. Questions on entry assignment that may arise are resolved by consultation with database staff.

#### Verification/Validation/Value Authority Control

The sequence database entry found during PDB's search is compared to that provided by the depositor and any differences are resolved or annotated.

In most cases, only one reference to a sequence database will be provided. PDB does not guarantee that all possible references to the listed databases will be provided.

#### Relationships to Other Record Types

DBREF represents the sequence as found in ATOM and HETATM records.

#### Example

	1	2	3	4	5	6	7
123456789012345678901234567890123456789012345678901234567890							
DBREF	1ABC B	1B 36	PDB		1ABC	1B 36	
DBREF	3AKY	3 220	SWS	P07170	KAD1_YEAST	5 222	
DBREF	1HAN	2 288	GB	397884	X66122	1 287	
DBREF	3HSV A	1 92	SWS	P22121	HSF_KLULA	193 284	
DBREF	3HSV B	1 92	SWS	P22121	HSF_KLULA	193 284	
DBREF	224D		NDB	DDF063			

## SEQADV

### Overview

The SEQADV record identifies conflicts between sequence information in the PDB entry and the sequence database entry given on DBREF. Please note that these records were designed to identify differences and not errors. No assumption is made as to which database contains the correct data. PDB may include REMARK records in the entry that reflect the depositor's view of which database has the correct sequence.

### Record Format

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"SEQADV"	
8 - 11	IDcode	idCode	ID code of this entry.
13 - 15	Residue name	resName	Name of the PDB residue in conflict.
17	Character	chainID	PDB chain identifier.
19 - 22	Integer	seqNum	PDB sequence number.
23	AChar	iCode	PDB insertion code.
25 - 28	LString	database	Sequence database name.
30 - 38	LString	dbIdCode	Sequence database accession number.
40 - 42	Residue name	dbRes	Sequence database residue name.
44 - 48	Integer	dbSeq	Sequence database sequence number.
50 - 70	LString	conflict	Conflict comment.

### Details

\* In a number of cases, conflicts between the sequences found in PDB entries and in PIR or SWISS-PROT entries have been noted. There are several possible reasons for these conflicts, including natural variants or engineered sequences (mutants), polymorphic sequences, or ambiguous or conflicting experimental results. These discrepancies, which were previously described in REMARK records, are now reported in SEQADV.

\* SEQADV describes conflicts between residue sequences given by PDB ATOM/HETATM records and those in the appropriate sequence database entry, such as residues missing due to disorder.

\* Some of the possible conflict comments:

- Cloning artifact
- Conflict
- Engineered
- Disordered
- Gap in PDB entry
- Missing from [database name]
- Variant
- Insertion
- Deletion
- Microheterogeneity

\* When conflicts arise which are not classifiable by these terms, a reference to either a published paper, a PDB entry, or a REMARK within the entry is given. References are given in the form YY-VOL-PAGE-CODEN where YY is year of publication, VOL is the journal volume number, PAGE is the starting page and CODEN is the 4-digit code assigned to journals by PDB and the Cambridge Crystallographic Data Centre (CCDC).

\* When reference is made to a PDB entry, then the form is PDB: 1ABC, where 1ABC is the relevant entry ID code.

\* Finally, the comment "SEE REMARK XXX" is included, where XXX is the remark number within the entry in which the primary explanation of the discrepancy is given.

\* Microheterogeneity is to be represented as a variant with one of the possible residues in the site being selected (arbitrarily) as the primary residue, in which case a SEQADV record must be provided for the alternate residue.

#### Verification/Validation/Value Authority Control

SEQADV records are automatically generated by the PDB.

#### Relationships to Other Record Types

SEQADV refers to the sequence as found in the ATOM and HETATM records, and to the sequence database reference found on DBREF.

#### Example

```

      1      2      3      4      5      6      7
123456789012345678901234567890123456789012345678901234567890
SEQADV 1ABC ASN A 100A SWS P10725 ASP 100 1994-300-1200-0070

SEQADV 2ABC ASN A 100A SWS P10725 ASP 100 PDB: 1ABC

SEQADV 3ABC MET A -1 SWS P10725 CLONING ARTIFACT
SEQADV 3ABC GLY A 50 SWS P10725 VAL 50 ENGINEERED

```

## SEQRES

### Overview

SEQRES records contain the amino acid or nucleic acid sequence of residues in each chain of the macromolecule that was studied.

### Record Format

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"SEQRES"	
9 - 10	Integer	serNum	Serial number of the SEQRES record, for the current chain. Starts at 1 and increments by one each line. Reset to 1 for each chain.
12	Character	chainID	Chain identifier. This may be any single legal character, including a blank which is used if there is only one chain.
14 - 17	Integer	numRes	Number of residues in the chain. This value is repeated on every record.

20 - 22	Residue name	resName	Residue name.
24 - 26	Residue name	resName	Residue name.
28 - 30	Residue name	resName	Residue name.
32 - 34	Residue name	resName	Residue name.
36 - 38	Residue name	resName	Residue name.
40 - 42	Residue name	resName	Residue name.
44 - 46	Residue name	resName	Residue name.
48 - 50	Residue name	resName	Residue name.
52 - 54	Residue name	resName	Residue name.
56 - 58	Residue name	resName	Residue name.
60 - 62	Residue name	resName	Residue name.
64 - 66	Residue name	resName	Residue name.
68 - 70	Residue name	resName	Residue name.

**Details**

\* PDB entries use the three-letter abbreviation for amino acid names and the one letter code for nucleic acids.

\* In the case of non-standard groups, a hetID of up to three (3) alphanumeric characters is used. Common HET names appear in the HET dictionary.

\* Each covalently contiguous sequence of residues (connected via the "backbone" atoms) is represented as an individual chain.

\* Heterogens which are integrated into the backbone of the chain are listed as being part of the chain and are included in the SEQRES records for that chain.

\* Each set of SEQRES records and each HET group is assigned a component number. The component number is assigned serially beginning with 1 for the first set of SEQRES records. This number is given explicitly in the FORMUL record, but only implicitly in the SEQRES record.

\* The SEQRES records must list residues present in the molecule studied, even if the coordinates are not present.

\* C- and N-terminus residues for which no coordinates are provided due to disorder must be listed on SEQRES.

\* All occurrences of standard amino or nucleic acid residues (ATOM records) must be listed on a SEQRES record. This implies that a numRes of 1 is valid.

\* No distinction is made between ribo- and deoxyribonucleotides in the SEQRES records. These residues are identified with the same residue name (i.e., A, C, G, T, U, I).

\* If the entire residue sequence is unknown, the serNum in column 10 is "0", the number of residues thought to comprise the molecule is entered as numRes in columns 14 - 17, and resName in columns 20 - 22 is "UNK".

\* In case of microheterogeneity, only one of the sequences is presented. A REMARK is generated to explain this and a SEQADV is also generated.

**Verification/Validation/Value Authority Control**

The residues presented on the SEQRES records must agree with those found in the ATOM records.

The SEQRES records are checked by PDB using the sequence databases and information provided by the depositor.

SEQRES is compared to the ATOM records during processing, and both are checked against the sequence database. All discrepancies are either resolved or annotated in the entry.

### Relationships to Other Record Types

The residues presented on the SEQRES records must agree with those found in the ATOM records. DBREF refers to the corresponding entry in the sequence databases. SEQADV lists all discrepancies between the entry's sequence for which there are coordinates and that referenced in the sequence database. MODRES describes modifications to a standard residue.

### Example

	1	2	3	4	5	6	7
1234567890123456789012345678901234567890123456789012345678901234567890							
SEQRES	1 A	21	GLY ILE VAL GLU GLN CYS CYS THR SER ILE CYS SER LEU				
SEQRES	2 A	21	TYR GLN LEU GLU ASN TYR CYS ASN				
SEQRES	1 B	30	PHE VAL ASN GLN HIS LEU CYS GLY SER HIS LEU VAL GLU				
SEQRES	2 B	30	ALA LEU TYR LEU VAL CYS GLY GLU ARG GLY PHE PHE TYR				
SEQRES	3 B	30	THR PRO LYS ALA				
SEQRES	1 C	21	GLY ILE VAL GLU GLN CYS CYS THR SER ILE CYS SER LEU				
SEQRES	2 C	21	TYR GLN LEU GLU ASN TYR CYS ASN				
SEQRES	1 D	30	PHE VAL ASN GLN HIS LEU CYS GLY SER HIS LEU VAL GLU				
SEQRES	2 D	30	ALA LEU TYR LEU VAL CYS GLY GLU ARG GLY PHE PHE TYR				
SEQRES	3 D	30	THR PRO LYS ALA				

### Known Problems

Polysaccharides do not lend themselves to being represented in SEQRES.

There is no mechanism provided to describe sequence runs when the exact ordering of the sequence is not known.

For cyclic peptides, PDB arbitrarily assigns a residue as the N-terminus.

For microheterogeneity only one of the possible residues in a given position is provided in SEQRES.

No distinction is made between ribo- and deoxyribonucleotides in the SEQRES records. These residues are identified with the same residue name (i.e., A, C, G, T, U).

---

## 4. Heterogen Section

The heterogen section of a PDB file contains the complete description of non-standard residues in the entry.

### HET

#### Overview

HET records are used to describe non-standard residues, such as prosthetic groups, inhibitors, solvent molecules, and ions for which coordinates are supplied. Groups are considered HET if they are:

- not one of the standard amino acids, and
- not one of the nucleic acids (C, G, A, T, U, and I), and
- not one of the modified versions of nucleic acids (+C, +G, +A, +T, +U, and +I), and
- not an unknown amino acid or nucleic acid where UNK is used to indicate the unknown residue name.

HET records also describe heterogens for which the chemical identity is unknown, in which case the group is assigned the hetID UNK.

#### Record Format

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"HET"	"
8 - 10	LString(3)	hetID	Het identifier, right-justified.
13	Character	ChainID	Chain identifier.
14 - 17	Integer	seqNum	Sequence number.
18	AChar	iCode	Insertion code.
21 - 25	Integer	numHetAtoms	Number of HETATM records for the group present in the entry.
28 - 70	String	Text	Text describing Het group.

#### Details

\* Each HET group is assigned a hetID of not more than three (3) alphanumeric characters. The sequence number, chain identifier, insertion code, and number of coordinate records are given for each occurrence of the HET group in the entry. The chemical name of the HET group is given in the HETNAM record and synonyms for the chemical name are given in the HETSYN records.

\* There is a separate HET record for each occurrence of the HET group in an entry.

\* A particular HET group is represented in the PDB archives with a unique hetID.

\* PDB entries do not have HET records for water molecules.

\* The Text field is for descriptive material. The token PART\_OF followed by a value may be used to indicate that the HET group is part of a larger group which has been represented by its separate components (e.g., PART\_OF: actinomycin). Segment identifiers, columns 73 - 76 of ATOM/HETATM records, may also be used to relate individual components of a large HET group.

#### Verification/Validation/Value Authority Control

For each het group that appears in the entry, PDB checks that the corresponding HET, HETNAM, HETSYN, FORMUL, HETATM, and CONECT records appear, if applicable. The HET record is generated automatically by PDB using the het group dictionary and information from the HETATM records.

Each unique hetID represents a unique molecule.

### Relationships to Other Record Types

For each het group that appears in the entry, the corresponding HET, HETNAM, HETSYN, FORMUL, HETATM, and CONECT records must appear, if applicable. LINK records may also appear.

### Example

```

      1           2           3           4           5           6           7
123456789012345678901234567890123456789012345678901234567890
HET   TRS      975      8
HET   STA  I    4      25  PART_OF: HIV INHIBITOR

```

### Known Problems

Even though groups may be chemically bound to others with loss of atoms (e.g., H, O), the PDB has only one representation for the complete molecule. However, a few small groups are represented separately as ions, groups, and molecules.

PDB does not include CAS registry and Cambridge Structural Database (CSD) accession numbers.

Large het groups are broken into recognizable sub-groups to obviate difficulties associated with the limitations of the atom naming conventions used by the PDB. The description of how to reassemble the full molecule is addressed in a REMARK. The token PART\_OF and use of segment identifiers may help to describe the larger entity.

---

## HETNAM

### Overview

This record gives the chemical name of the compound with the given hetID.

### Record Format

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"HETNAM"	
9 - 10	Continuation	continuation	Allows concatenation of multiple records.
12 - 14	LString(3)	hetID	Het identifier, right-justified.
16 - 70	String		Chemical name.

### Details

- \* Each hetID is assigned a unique chemical name for the HETNAM record.
- \* Other names for the group are given on HETSYN records.
- \* PDB follows IUPAC/IUB naming conventions to describe groups systematically.
- \* Continuation of chemical names onto subsequent records is allowed.



\* Only one HETNAM record is included for a given hetID, even if the same hetID appears on more than one HET record.

#### Verification/Validation/Value Authority Control

For each het group that appears in the entry, the corresponding HET, HETNAM, FORMUL, HETATM and CONECT records must appear. The HETNAM record is generated automatically by PDB using the het group dictionary and information from HETATM records.

#### Relationships to Other Record Types

For each het group that appears in the entry, the corresponding HET, HETNAM, FORMUL, HETATM, and CONECT records must appear. HETSYN and LINK records may also appear.

#### Example

```

      1           2           3           4           5           6           7
123456789012345678901234567890123456789012345678901234567890
HETNAM      GLC  GLUCOSE
HETNAM      SAD  BETA-METHYLENE SELENAZOLE-4-CARBOXAMIDE ADENINE
HETNAM      2   SAD  DINUCLEOTIDE

```

---

## HETSYN

### Overview

This record provides synonyms, if any, for the compound in the corresponding (i.e., same hetID) HETNAM record. This is to allow greater flexibility in searching for HET groups.

### Record Format

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"HETSYN"	
9 - 10	Continuation	continuation	Allows concatenation of multiple records.
12 - 14	LString(3)	hetID	Het identifier, right-justified.
16 - 70	SList		List of synonyms.

### Details

\* This is not guaranteed to be a complete list of possible synonyms, but is uniform across the PDB. New synonyms may be added. The list can be continued onto additional HETSYN records. Even if the same hetID appears on more than one HET record, only one set of HETSYN records is included for the hetID.

#### Verification/Validation/Value Authority Control

For each HETSYN record in the entry, the corresponding HET, HETNAM, FORMUL, HETATM and CONECT records must appear.

#### Relationships to Other Record Types

If there is a HETSYN record there must be corresponding HET, HETNAM, FORMUL, HETATM, and CONECT records. LINK records may also appear.

**Example**

```

1      2      3      4      5      6      7
123456789012345678901234567890123456789012345678901234567890
HETSYN      NAD NICOTINAMIDE ADENINE DINUCLEOTIDE
HETSYN      COA COA

HETSYN      CMP CYCLIC AMP; CYCLIC ADENOSINE MONOPHOSPHATE

HETSYN      TRS TRIS BUFFER; TRISAMINE;
HETSYN      2 TRS TRIS (HYDROXYMETHYL) AMINOMETHANE; TRIMETHYLOL
HETSYN      3 TRS AMINOMETHANE

```

**FORMUL****Overview**

The FORMUL record presents the chemical formula and charge of a non-standard group. (The formulas for the standard residues are given in Appendix 5.)

**Record Format**

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"FORMUL"	
9 - 10	Integer	compNum	Component number.
13 - 15	LString(3)	hetID	Het identifier.
17 - 18	Integer	continuation	Continuation number.
19	Character		"*" for water.
20 - 70	String		Chemical formula.

**Details**

\* The elements of the chemical formula are given in the order C, H, N, and O, with other elements following in alphabetical order, each separated by a single blank.

\* The number of each atom type present immediately follows its chemical symbol with no intervening blank.

\* Each set of SEQRES records and each HET group is assigned a component number in an entry. These numbers are assigned serially, beginning with 1 for the first set of SEQRES records. In addition:

- If a HET group is presented on a SEQRES record its FORMUL is assigned the component number of the chain.

- If the HET group occurs more than once and is not presented on SEQRES records, the component number of its first occurrence is used.

\* All occurrences of the HET group within a chain are grouped together with a multiplier. The remaining occurrences are also grouped with a multiplier. The sum of the multipliers is the number equaling the number of times that that HET group appears in the entry.

\* The "\*" in column 19 is used if the HET group is water, indicating that it should be excluded from the molecular weight calculation.

\* A continuation field is provided in the event that more space is needed for the formula. Columns 17 - 18 are used in order to maintain continuity with the existing format.

**Verification/Validation/Value Authority Control**

For each het group that appears in the entry, the corresponding HET, HETNAM, FORMUL, HETATM, and CONECT records must appear. The FORMUL record is generated automatically by PDB processing programs using the het group template file and information from HETATM records.

**Relationships to Other Record Types**

For each het group that appears in the entry, the corresponding HET, HETNAM, FORMUL, HETATM, and CONECT records must appear.

**Example**

	1	2	3	4	5	6	7
123456789012345678901234567890123456789012345678901234567890							
FORMUL	2	SO4	2(O4 S1 2-)				
FORMUL	3	GLC	C6 H12 O6				
FORMUL	3	FOL	2(C19 H17 N7 O6 2-)				
FORMUL	4	CL	2(CL1 1-)				
FORMUL	5	CA	CA1 2+				
FORMUL	6	HOH	*429(H2 O1)				

**Known Problems**

Partially deuterated centers are not well represented in this record.

---

## 5. Secondary Structure Section

The secondary structure section of a PDB file describes helices, sheets, and turns found in protein and polypeptide structures.

---

### HELIX

#### Overview

HELIX records are used to identify the position of helices in the molecule. Helices are both named and numbered. The residues where the helix begins and ends are noted, as well as the total length.

#### Record Format

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"HELIX "	
8 - 10	Integer	serNum	Serial number of the helix. This starts at 1 and increases incrementally.
12 - 14	LString(3)	HelixID	Helix identifier. In addition to a serial number, each helix is given a character helix identifier.
16 - 18	Residue name	initResName	Name of the initial residue.
20	Character	initChainID	Chain identifier for the chain containing this helix.
22 - 25	Integer	initSeqNum	Sequence number of the initial residue.
26	AChar	initICode	Insertion code of the initial residue.
28 - 30	Residue name	endResName	Name of the terminal residue of the helix.
32	Character	endChainID	Chain identifier for the chain containing this helix.
34 - 37	Integer	endSeqNum	Sequence number of the terminal residue.
38	AChar	endICode	Insertion code of the terminal residue.
39 - 40	Integer	class	Helix class (see below).
41 - 70	String	comment	Comment about this helix.
72 - 76	Integer	length	Length of this helix.

#### Details

\* Additional HELIX records with different serial numbers and identifiers occur if more than one helix is present.

\* The initial residue is the N-terminal residue of the helix.

\* Helices are classified as follows:

TYPE OF HELIX	CLASS NUMBER (COLUMNS 39 - 40)
Right-handed alpha (default)	1
Right-handed omega	2
Right-handed pi	3
Right-handed gamma	4
Right-handed 310	5
Left-handed alpha	6
Left-handed omega	7
Left-handed gamma	8
27 ribbon/helix	9
Polyproline	10

#### Verification/Validation/Value Authority Control

HELIX records are now being generated automatically by PDB using the Kabsch and Sander algorithm [Kabsch and Sander, Biopolymers 22: 2577-2637 (1983)], although they may be provided by the depositor instead. PDB verifies that named residues exist in the ATOM records.

#### Relationships to Other Record Types

There may be related information in the REMARKs.

#### Example

```

123456789012345678901234567890123456789012345678901234567890123456
HELIX 1 HA GLY A 86 GLY A 94 1 9
HELIX 2 HB GLY B 86 GLY B 94 1 9

```

#### Known Problems

PDB is considering addition of some new information related to HELIX, in order to present more complete structural information. Please comment on the suggestion of adding a new record which would present the various domain types found in the molecule, e.g., Residues 12 --> 120: alpha/beta.

## SHEET

#### Overview

SHEET records are used to identify the position of sheets in the molecule. Sheets are both named and numbered. The residues where the sheet begins and ends are noted.

#### Record Format

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"SHEET "	
8 - 10	Integer	strand	Strand number which starts at 1 for each strand within a sheet and increases by one.
12 - 14	LString(3)	sheetID	Sheet identifier.
15 - 16	Integer	numStrands	Number of strands in sheet.
18 - 20	Residue name	initResName	Residue name of initial residue.

22	Character	initChainID	Chain identifier of initial residue in strand.
23 - 26	Integer	initSeqNum	Sequence number of initial residue in strand.
27	AChar	initICode	Insertion code of initial residue in strand.
29 - 31	Residue name	endResName	Residue name of terminal residue.
33	Character	endChainID	Chain identifier of terminal residue.
34 - 37	Integer	endSeqNum	Sequence number of terminal residue.
38	AChar	endICode	Insertion code of terminal residue.
39 - 40	Integer	sense	Sense of strand with respect to previous strand in the sheet. 0 if first strand, 1 if parallel, -1 if anti-parallel.
42 - 45	Atom	curAtom	Registration. Atom name in current strand.
46 - 48	Residue name	curResName	Registration. Residue name in current strand.
50	Character	curChainId	Registration. Chain identifier in current strand.
51 - 54	Integer	curResSeq	Registration. Residue sequence number in current strand.
55	AChar	curICode	Registration. Insertion code in current strand.
57 - 60	Atom	prevAtom	Registration. Atom name in previous strand.
61 - 63	Residue name	prevResName	Registration. Residue name in previous strand.
65	Character	prevChainId	Registration. Chain identifier in previous strand.
66 - 69	Integer	prevResSeq	Registration. Residue sequence number in previous strand.
70	AChar	prevICode	Registration. Insertion code in previous strand.

**Details**

\* The initial residue for a strand is its N-terminus. Strand registration information is provided in columns 39 - 70. Strands are listed starting with one edge of the sheet and continuing to the spatially adjacent strand.

\* The sense in columns 39 - 40 indicates whether strand n is parallel (sense = 1) or anti-parallel (sense = -1) to strand n-1. Sense is equal to zero (0) for the first strand of a sheet.

\* The registration (columns 42 - 70) of strand n to strand n-1 may be specified by one hydrogen bond between each such pair of strands. This is done by providing the hydrogen bonding between the current and previous strands. No registration information should be provided for the first strand.

\* For structures which form a closed sheet (beta-barrel), the first strand is repeated as the last strand. An explanatory remark is included in the REMARK section.

\* Split strands, or strands with two or more runs of residues from discontinuous parts of the amino acid sequence, are explicitly listed. Provide a description to be included in the REMARK section.

#### Verification/Validation/Value Authority Control

SHEET records are now being generated automatically by PDB using the Kabsch and Sander algorithm [Kabsch and Sander, Biopolymers 22: 2577-2637 (1983)], although they may be provided by the depositor instead. PDB verifies that named residues exist in the ATOM records.

#### Relationships to Other Record Types

If the entry contains bifurcated sheets or beta-barrels, the relevant REMARK records must be provided. See the REMARK section for details.

#### Example

	1	2	3	4	5	6	7
	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890
SHEET	1	A 5 THR A 107	ARG A 110	0			
SHEET	2	A 5 ILE A 96	THR A 99	-1 N LYS A 98	O THR A 107		
SHEET	3	A 5 ARG A 87	SER A 91	-1 N LEU A 89	O TYR A 97		
SHEET	4	A 5 TRP A 71	ASP A 75	-1 N ALA A 74	O ILE A 88		
SHEET	5	A 5 GLY A 52	PHE A 56	-1 N PHE A 56	O TRP A 71		
SHEET	1	B 5 THR B 107	ARG B 110	0			
SHEET	2	B 5 ILE B 96	THR B 99	-1 N LYS B 98	O THR B 107		
SHEET	3	B 5 ARG B 87	SER B 91	-1 N LEU B 89	O TYR B 97		
SHEET	4	B 5 TRP B 71	ASP B 75	-1 N ALA B 74	O ILE B 88		
SHEET	5	B 5 GLY B 52	ILE B 55	-1 N ASP B 54	O GLU B 73		

The sheet presented as BS1 below is an eight-stranded beta-barrel. This is represented by a nine-stranded sheet in which the first and last strands are identical.

SHEET	1	BS1 9 VAL	13	ILE	17	0				
SHEET	2	BS1 9 ALA	70	ILE	73	1	O TRP	72	N ILE	17
SHEET	3	BS1 9 LYS	127	PHE	132	1	O ILE	129	N ILE	73
SHEET	4	BS1 9 GLY	221	ASP	225	1	O GLY	221	N ILE	130
SHEET	5	BS1 9 VAL	248	GLU	253	1	O PHE	249	N ILE	222
SHEET	6	BS1 9 LEU	276	ASP	278	1	N LEU	277	O GLY	252
SHEET	7	BS1 9 TYR	310	THR	318	1	O VAL	317	N ASP	278
SHEET	8	BS1 9 VAL	351	TYR	356	1	O VAL	351	N THR	318
SHEET	9	BS1 9 VAL	13	ILE	17	1	N VAL	14	O PRO	352

The sheet structure of this example is bifurcated. In order to represent this feature, two sheets are defined. Strands 2 and 3 of BS7 and BS8 are identical.

SHEET	1	BS7 3 HIS	662	THR	665	0				
SHEET	2	BS7 3 LYS	639	LYS	648	-1	N PHE	643	O HIS	662
SHEET	3	BS7 3 ASN	596	VAL	600	-1	N TYR	598	O ILE	646
SHEET	1	BS8 3 ASN	653	TRP	656	0				
SHEET	2	BS8 3 LYS	639	LYS	648	-1	N LYS	647	O THR	655
SHEET	3	BS8 3 ASN	596	VAL	600	-1	N TYR	598	O ILE	646

## TURN

### Overview

The TURN records identify turns and other short loop turns which normally connect other secondary structure segments.

### Record Format

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"TURN "	
8 - 10	Integer	seq	Turn number; starts with 1 and increments by one.
12 - 14	LString(3)	turnId	Turn identifier
16 - 18	Residue name	initResName	Residue name of initial residue in turn.
20	Character	initChainId	Chain identifier for the chain containing this turn.
21 - 24	Integer	initSeqNum	Sequence number of initial residue in turn.
25	AChar	initICode	Insertion code of initial residue in turn.
27 - 29	Residue name	endResName	Residue name of terminal residue of turn.
31	Character	endChainId	Chain identifier for the chain containing this turn.
32 - 35	Integer	endSeqNum	Sequence number of terminal residue of turn.
36	AChar	endICode	Insertion code of terminal residue of turn.
41 - 70	String	comment	Associated comment.

### Details

\* Turns include those sets of residues which form beta turns, i.e., have a hydrogen bond linking (C-O)<sub>i</sub> to (N-H)<sub>i+3</sub>. Turns which link residue *i* to *i*+2 (gamma-bends) may also be included. Others may be also be classified as turns.

\* The initial residue is the N-terminus.

### Verification/Validation/Value Authority Control

The validation program checks the number of residues in the given turn. PDB verifies that named residues exist in the ATOM records.

### Relationships to Other Record Types

There may be related information in the REMARKs.



**Example**

	1	2	3	4	5	6	7
TURN	1 S1A GLY A 16	GLN A 18	SURFACE				
TURN	2 FLA ILE A 50	GLY A 52	FLAP				
TURN	3 S2A ILE A 66	HIS A 69	SURFACE				
TURN	4 S1B GLY B 16	GLN B 18	SURFACE				
TURN	5 FLB ILE B 50	GLY B 52	FLAP				
TURN	6 S2B ILE B 66	HIS B 69	SURFACE				

---

---

## 6. Connectivity Annotation Section

The connectivity annotation section allows the depositors to specify the existence and location of disulfide bonds and other linkages.

### SSBOND

#### Overview

The SSBOND record identifies each disulfide bond in protein and polypeptide structures by identifying the two residues involved in the bond.

#### Record Format

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"SSBOND"	
8 - 10	Integer	serNum	Serial number.
12 - 14	LString(3)	"CYS"	Residue name.
16	Character	chainID1	Chain identifier.
18 - 21	Integer	seqNum1	Residue sequence number.
22	AChar	icode1	Insertion code.
26 - 28	LString(3)	"CYS"	Residue name.
30	Character	chainID2	Chain identifier.
32 - 35	Integer	seqNum2	Residue sequence number.
36	AChar	icode2	Insertion code.
60 - 65	SymOP	sym1	Symmetry operator for 1st residue.
67 - 72	SymOP	sym2	Symmetry operator for 2nd residue.

#### Details

\* Bond distances between the sulfur atoms must be close to expected values.

\* The cysteine closer to the N-terminal is listed first in each intra-chain pair. The cysteine which occurs first in the coordinate entry is listed first for inter-chain pairs.

\* sym1 and sym2 are given as blank when the identity operator (and no cell translation) is to be applied to the residue.

#### Verification/Validation/Value Authority Control

PDB processing programs generate these records automatically. If the depositor supplies these records, they are compared to those generated and the depositor is notified of any differences.

#### Relationships to Other Record Types

CONNECT records are generated for the disulfide bonds when SG atoms of both cysteines are present in the coordinate records. If symmetry operators are given to generate one of the residues involved in the disulfide bond, a REMARK record must appear which defines the symmetry transformation.

**Example**

	1	2	3	4	5	6	7
12345678901234567890123456789012345678901234567890123456789012							
SSBOND	1	CYS E	48	CYS E	51		
SSBOND	2	CYS E	252	CYS E	285	2555	

**Known Problems**

If SG of cysteine is disordered then there are possible alternate linkages. PDB's practice is to put together all possible SSBOND records. This is problematic because the alternate location identifier is not specified in the SSBOND record.

---

**LINK****Overview**

The LINK records specify connectivity between residues that is not implied by the primary structure. Connectivity is expressed in terms of the atom names. This record supplements information given in CONECT records and is provided here for convenience in searching.

**Record Format**

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"LINK "	
13 - 16	Atom	name1	Atom name.
17	Character	altLoc1	Alternate location indicator.
18 - 20	Residue name	resName1	Residue name.
22	Character	chainID1	Chain identifier.
23 - 26	Integer	resSeq1	Residue sequence number.
27	AChar	iCODE1	Insertion code.
43 - 46	Atom	name2	Atom name.
47	Character	altLoc2	Alternate location indicator.
48 - 50	Residue name	resName2	Residue name.
52	Character	chainID2	Chain identifier.
53 - 56	Integer	resSeq2	Residue sequence number.
57	AChar	iCode2	Insertion code.
60 - 65	SymOP	sym1	Symmetry operator for 1st atom.
67 - 72	SymOP	sym2	Symmetry operator for 2nd atom.

**Details**

\* The atoms involved in bonds between HET groups or between a HET group and standard residue are listed.

\* Interresidue linkages not implied by the primary structure are listed (e.g., reduced peptide bond).

- \* Non-standard linkages between residues, e.g., side-chain to side-chain, are listed.
- \* Each LINK record specifies one linkage.
- \* These records do not specify connectivity within a HET group (see CONECT), hydrogen bonds (see HYDBND), or disulfide bridges (see SSBOND).
- \* Hydrogen bonds and salt bridges are described on HYDBND and SLTBRG records, respectively.
- \* sym1 and sym2 are given as blank when the identity operator (and no cell translation) is to be applied to the atom.

#### Verification/Validation/Value Authority Control

The distance between the pair of atoms listed must be consistent with the bonding.

#### Relationships to Other Record Types

CONECT records are generated from LINKs when both atoms are present in the entry. If symmetry operators are given to generate one of the residues involved in the bond, a REMARK record must appear which defines the symmetry transformation.

#### Example

```

      1      2      3      4      5      6      7
12345678901234567890123456789012345678901234567890123456789012
LINK      O1  DDA      1      C3  DDL      2
LINK      MN   MN   391      OE2 GLU   217      2565

```

## HYDBND

### Overview

The HYDBND records specify hydrogen bonds in the entry.

### Record Format

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"HYDBND"	
13 - 16	Atom	name1	Atom name.
17	Character	altLoc1	Alternate location indicator.
18 - 20	Residue name	resName1	Residue name.
22	Character	Chain1	Chain identifier.
23 - 27	Integer	resSeq1	Residue sequence number.
28	AChar	ICode1	Insertion code.
30 - 33	Atom	nameH	Hydrogen atom name.
34	Character	altLocH	Alternate location indicator.
36	Character	ChainH	Chain identifier.
37 - 41	Integer	resSeqH	Residue sequence number.
42	AChar	iCodeH	Insertion code.

44 - 47	Atom	name2	Atom name.
48	Character	altLoc2	Alternate location indicator.
49 - 51	Residue name	resName2	Residue name.
53	Character	chainID2	Chain identifier.
54 - 58	Integer	resSeq2	Residue sequence number.
59	AChar	iCode2	Insertion code.
60 - 65	SymOP	sym1	Symmetry operator for 1st non-hydrogen atom.
67 - 72	SymOP	sym2	Symmetry operator for 2nd non-hydrogen atom.

**Details**

- \* The hydrogen bonds listed normally are those supplied by the depositor.
- \* The atoms forming the hydrogen bond are listed on the HYDBND record.
- \* Each record has place for three atom specifications.
- \* Columns 13 - 28 and 44 - 59 are for the atoms associated with the hydrogen atom of the hydrogen bond.
- \* If the coordinates of the hydrogen atom itself are presented in the entry, that atom is specified in columns 30 - 42.
- \* For nucleic acids, Watson-Crick hydrogen bonds between bases may be listed, but this is optional.
- \* sym1 and sym2 are given as blank when the identity operator (and no cell translation) is to be applied to the atom. For hydrogen atoms use the symmetry operator of the heavy atom to which it is bonded.

**Verification/Validation/Value Authority Control**

The distance between the atoms listed must be consistent with the bonding.

**Relationships to Other Record Types**

CONNECT records are generated consistent with the bond type. If symmetry operators are given to generate one of the residues involved in the hydrogen bond, a REMARK record must appear which defines the symmetry transformation.

**Example**

	1	2	3	4	5	6	7
12345678901234567890123456789012345678901234567890123456789012							
HYDBND	N	LEU	10		AO3*	NDP	501
HYDBND	NH2	ARG	111		OD2	ASP	149 1555

## SLTBRG

### Overview

The SLTBRG records specify salt bridges in the entry.

### Record Format

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"SLTBRG"	
13 - 16	Atom	atom1	First atom name.
17	Character	altLoc1	Alternate location indicator.
18 - 20	Residue name	resName1	Residue name.
22	Character	chainID1	Chain identifier.
23 - 26	Integer	resSeq1	Residue sequence number.
27	AChar	iCode1	Insertion code.
43 - 46	Atom	atom2	Second atom name.
47	Character	altLoc2	Alternate location indicator.
48 - 50	Residue name	resName2	Residue name.
52	Character	chainID2	Chain identifier.
53 - 56	Integer	resSeq2	Residue sequence number.
57	AChar	iCode2	Insertion code.
60 - 65	SymOP	sym1	Symmetry operator for 1st atom.
67 - 72	SymOP	sym2	Symmetry operator for 2nd atom.

### Details

- \* Salt bridges listed normally are those provided by the depositor.
- \* The two atoms forming the salt bridge through their electrostatic interactions are specified.
- \* No distinction is made as to which atom has excess positive or negative charge.
- \* sym1 and sym2 are given as blank when the identity operator (and no cell translation) is to be applied to the atom.

### Verification/Validation/Value Authority Control

The distance between the pair of atoms listed must be consistent with the bonding.

### Relationships to Other Record Types

CONNECT records are generated consistent with the bond type. If symmetry operators are given to generate one of the residues involved in the salt bridge, a REMARK record must appear which defines the symmetry transformation.



**Example**

	1	2	3	4	5	6	7
123456789012345678901234567890123456789012345678901234567890							
CISPEP	1	GLY A	116	GLY A	117	18.50	
CISPEP	2	THR D	92	PRO D	93	359.80	

---

---



## 7. Miscellaneous Features Section

The miscellaneous features section describes features in the molecule such as the active site. Other features may be described in the remarks section but are not given a specific record type so far.

---

### SITE

#### Overview

The SITE records supply the identification of groups comprising important sites in the macromolecule.

#### Record Format

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"SITE "	
8 - 10	Integer	seqNum	Sequence number.
12 - 14	LString(3)	siteID	Site name.
16 - 17	Integer	numRes	Number of residues comprising site.
19 - 21	Residue name	resName1	Residue name for first residue comprising site.
23	Character	chainID1	Chain identifier for first residue comprising site.
24 - 27	Integer	seq1	Residue sequence number for first residue comprising site.
28	AChar	iCode1	Insertion code for first residue comprising site.
30 - 32	Residue name	resName2	Residue name for second residue comprising site.
34	Character	chainID2	Chain identifier for second residue comprising site.
35 - 38	Integer	seq2	Residue sequence number for second residue comprising site.
39	AChar	iCode2	Insertion code for second residue comprising site.
41 - 43	Residue name	resName3	Residue name for third residue comprising site.
45	Character	chainID3	Chain identifier for third residue comprising site.
46 - 49	Integer	seq3	Residue sequence number for third residue comprising site.
50	AChar	iCode3	Insertion code for third residue comprising site.

52 - 54	Residue name	resName4	Residue name for fourth residue comprising site.
56	Character	chainID4	Chain identifier for fourth residue comprising site.
57 - 60	Integer	seq4	Residue sequence number for fourth residue comprising site.
61	AChar	iCode4	Insertion code for fourth residue comprising site.

**Details**

- \* Site records specify residues comprising catalytic, cofactor, anticodon, regulatory or other important sites.
- \* The sequence number (columns 8 - 10) is reset to 1 for each new site.
- \* SITE identifiers (columns 12 - 14) should be fully explained in a remark.
- \* If a site is comprised of more than four residues, these may be specified on additional records bearing the same site identifier.
- \* SITE records can include HET groups.

**Verification/Validation/Value Authority Control**

Every SITE must have a corresponding REMARK that describes it. The numbering of sequential SITE records and format of each one is verified, as well as the existence of each residue in the ATOM records.

**Relationships to Other Record Types**

Each listed SITE needs a corresponding REMARK that details its significance.

**Example**

```

123456789012345678901234567890123456789012345678901234567890
SITE 1 DTA 3 ASP A 25 THR A 26 GLY A 27
SITE 1 DTB 3 ASP B 25 THR B 26 GLY B 27
SITE 1 A 4 U A 44 C A 46 G A 61 U A 118
SITE 1 ZN1 5 CYS A 97 CYS A 100 CYS A 103 CYS 1 111
SITE 2 ZN1 5 ZN A 375

```

## 8. Crystallographic and Coordinate Transformation Section

The Crystallographic Section describes the geometry of the crystallographic experiment and the coordinate system transformations.

---

### CRYST1

#### Overview

The CRYST1 record presents the unit cell parameters, space group, and Z value. If the structure was not determined by crystallographic means, CRYST1 simply defines a unit cube.

#### Record Format

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"CRYST1"	
7 - 15	Real (9.3)	a	a (Angstroms) .
16 - 24	Real (9.3)	b	b (Angstroms) .
25 - 33	Real (9.3)	c	c (Angstroms) .
34 - 40	Real (7.2)	alpha	alpha (degrees) .
41 - 47	Real (7.2)	beta	beta (degrees) .
48 - 54	Real (7.2)	gamma	gamma (degrees) .
56 - 66	LString	sGroup	Space group .
67 - 70	Integer	z	Z value .

#### Details

\* If the coordinate entry describes a structure determined by a technique other than crystallography, CRYST1 contains  $a = b = c = 1.0$ ,  $\alpha = \beta = \gamma = 90$  degrees, space group = P 1, and  $Z = 1$ .

\* The Hermann-Mauguin space group symbol is given without parenthesis, e.g., P 43 21 2. Please note that the screw axis is described as a two digit number.

\* The full international Hermann-Mauguin symbol is used, e.g., P 1 21 1 instead of P 21.

\* For a rhombohedral space group in the hexagonal setting, the lattice type symbol used is H.

\* The Z value is the number of polymeric chains in a unit cell. In the case of heteropolymers, Z is the number of occurrences of the most populous chain.

As an example, given two chains A and B, each with a different sequence, and the space group P 2 that has two equipoints in the standard unit cell, the following table gives the correct Z value.

Asymmetric Unit Content	Z value
A	2
AA	4
AB	2
AAB	4
AABB	4

\* In the case of a polycrystalline fiber diffraction study, CRYST1 and SCALE contain the normal unit cell data.

#### Verification/Validation/Value Authority Control

The given space group and Z values are checked during processing for correctness and internal consistency. The calculated SCALE is compared to that supplied by the depositor. Packing is also computed, and close contacts of symmetry-related molecules are diagnosed.

#### Relationships to Other Record Types

The unit cell parameters are used to calculate SCALE. If the EXPDTA record is NMR, THEORETICAL MODEL, or FIBER DIFFRACTION, FIBER, the CRYST1 record is predefined as a = b = c = 1.0, alpha = beta = gamma = 90 degrees, space group = P 1 and Z = 1. In these cases, an explanatory REMARK must also appear in the entry. Some fiber diffraction structures will be done this way, while others will have a CRYST1 record containing measured values.

#### Example

	1	2	3	4	5	6	7
1234567890123456789012345678901234567890123456789012345678901234567890							
CRYST1	52.000	58.600	61.900	90.00	90.00	90.00 P 21 21 21	8
CRYST1	1.000	1.000	1.000	90.00	90.00	90.00 P 1	1
CRYST1	42.544	69.085	50.950	90.00	95.55	90.00 P 1 21 1	2

#### Known Problems

No standard deviations are given.

## ORIGXn

#### Overview

The ORIGXn (n = 1, 2, or 3) records present the transformation from the orthogonal coordinates contained in the entry to the submitted coordinates.

#### Record Format

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"ORIGXn"	n=1, 2, or 3
11 - 20	Real (10.6)	o[n][1]	On1
21 - 30	Real (10.6)	o[n][2]	On2
31 - 40	Real (10.6)	o[n][3]	On3
46 - 55	Real (10.5)	t[n]	Tn

#### Details

\* The PDB supplies this information even if the transformation is an identity transformation (unit matrix, null vector). See the SCALE section of this document for a definition of the default orthogonal Angstroms system.

\* If the original submitted coordinates are Xsub, Ysub, Zsub and the orthogonal Angstroms coordinates contained in the data entry are X, Y, Z, then:

$$X_{\text{sub}} = O11X + O12Y + O13Z + T1$$

$$Y_{\text{sub}} = O21X + O22Y + O23Z + T2$$

$$Z_{\text{sub}} = O31X + O32Y + O33Z + T3$$

\* Appendix 2 details the derivation of the ORIGX coordinate transformation.

#### Verification/Validation/Value Authority Control

If the coordinates are submitted in the same orthogonal Angstrom coordinate frame as they appear in the entry (the usual case), then ORIGX is an identity matrix with a null translation vector. If the transformation is not an identity matrix with a null translation vector, then applying this transformation to the coordinates in the entry yields the coordinates in the original deposited file.

#### Relationships to Other Record Types

ORIGX relates the coordinates in the ATOM and HETATM records to the coordinates in the submitted file.

#### Example

	1	2	3	4	5	6	7
1234567890123456789012345678901234567890123456789012345678901234567890							
ORIGX1	0.963457	0.136613	0.230424		16.61000		
ORIGX2	-0.158977	0.983924	0.081383		13.72000		
ORIGX3	-0.215598	-0.115048	0.969683		37.65000		

## SCALEn

### Overview

The SCALEn (n = 1, 2, or 3) records present the transformation from the orthogonal coordinates as contained in the entry to fractional crystallographic coordinates. Non-standard coordinate systems should be explained in the remarks.

### Record Format

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"SCALEn"	n=1, 2, or 3
11 - 20	Real (10.6)	s[n][1]	Sn1
21 - 30	Real (10.6)	s[n][2]	Sn2
31 - 40	Real (10.6)	s[n][3]	Sn3
46 - 55	Real (10.5)	u[n]	Un

### Details

\* The standard orthogonal Angstroms coordinate system used by the PDB is related to the axial system of the unit cell supplied (CRYST1 record) by the following definition:

\* If vector a, vector b, vector c describe the crystallographic cell edges, and vector A, vector B, vector C are unit cell vectors in the default orthogonal Angstroms system, then vector A, vector B, vector C and vector a, vector b, vector c have the same origin; vector A is parallel to vector a, vector B is parallel to vector C times vector A, and vector C is parallel to vector a times vector b (i.e., vector c\*).

\* If the orthogonal Angstroms coordinates are X, Y, Z, and the fractional cell coordinates are xfrac, yfrac, zfrac, then:

$$x_{\text{frac}} = S11X + S12Y + S13Z + U1$$

$$y_{\text{frac}} = S21X + S22Y + S23Z + U2$$

$$z_{\text{frac}} = S31X + S32Y + S33Z + U3$$

\* For NMR, fiber diffraction - fiber sample, and theoretical model entries, SCALE is given as an identity matrix with no translation.

\* Appendix 2 details the derivation of the SCALE coordinate transformation.

#### Verification/Validation/Value Authority Control

The inverse of the determinant of the SCALE matrix equals the volume of the cell. This volume is calculated and compared to the SCALE matrix supplied by the depositor.

#### Relationships to Other Record Types

The SCALE transformation is related to the CRYST1 record, as the inverse of the determinant of the SCALE matrix equals the cell volume.

#### Example

	1	2	3	4	5	6	7
123456789012345678901234567890123456789012345678901234567890							
SCALE1	0.019231	0.000000	0.000000		0.000000		
SCALE2	0.000000	0.017065	0.000000		0.000000		
SCALE3	0.000000	0.000000	0.016155		0.000000		

## MATRIXn

### Overview

The MATRIXn (n = 1, 2, or 3) records present transformations expressing non-crystallographic symmetry.

### Record Format

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"MATRIXn"	n=1, 2, or 3
8 - 10	Integer	serial	Serial number.
11 - 20	Real(10.6)	m[n][1]	Mn1
21 - 30	Real(10.6)	m[n][2]	Mn2
31 - 40	Real(10.6)	m[n][3]	Mn3
46 - 55	Real(10.5)	v[n]	Vn
60	Integer	iGiven	1 if coordinates for the representations which are approximately related by the transformations of the molecule are contained in the entry. Otherwise, blank.

### Details

\* The MATRIX transformations operate on the coordinates in the entry to yield equivalent representations of the molecule in the same coordinate frame. One trio of MATRIX records with a constant serial number is given for each non-crystallographic symmetry operation defined. If coordinates for the representations which are approximately related by the given transformation are contained in the file, the iGiven field is set to 1. Otherwise, this field is blank.

\* A corresponding REMARK must appear which describes the transformation.

**Verification/Validation/Value Authority Control**

The PDB verifies all MTRIX records by applying the given transformation and determining the RMSD between the calculated and supplied coordinates if iGiven is equal to 1. If iGiven is blank, PDB verifies MTRIX by checking the packing of the generated molecules.

**Relationships to Other Record Types**

A corresponding REMARK must appear which describes the transformation.

**Example**

```

      1      2      3      4      5      6      7
123456789012345678901234567890123456789012345678901234567890
MTRIX1  1 -1.000000  0.000000 -0.000000      0.00001  1
MTRIX2  1 -0.000000  1.000000  0.000000      0.00002  1
MTRIX3  1  0.000000 -0.000000 -1.000000      0.00002  1

```

**TVECT****Overview**

The TVECT records present the translation vector for infinite covalently connected structures.

**Record Format**

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"TVECT "	
8 - 10	Integer	serial	Serial number.
11 - 20	Real(10.5)	t[1]	Components of translation vector.
21 - 30	Real(10.5)	t[2]	Components of translation vector.
31 - 40	Real(10.5)	t[3]	Components of translation vector.
41 - 70	String	text	Comment.

**Details**

\* For structures not comprised of discrete molecules (e.g., infinite polysaccharide chains), the entry contains a fragment which can be built into the full structure by the simple translation vectors of TVECT records.

\* A corresponding REMARK describing the structure must appear.

**Verification/Validation/Value Authority Control**

PDB applies the translation and checks the generated molecule.

**Relationships to Other Record Types**

A corresponding REMARK describing the structure must appear.

**Example**

	1	2	3	4	5	6	7
123456789012345678901234567890123456789012345678901234567890							
TVECT	1	0.00000	0.00000	28.30000			

---

---



## 9. Coordinate Section

The Coordinate Section contains the collection of atomic coordinates as well as the MODEL and ENDMDL records.

---

### MODEL

#### Overview

The MODEL record specifies the model serial number when multiple structures are presented in a single coordinate entry, as is often the case with structures determined by NMR.

#### Record Format

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"MODEL "	
11 - 14	Integer	serial	Model serial number.

#### Details

\* This record is used only when more than one model appears in an entry. Generally, it is employed only for NMR structures. The chemical connectivity should be the same for each model. ATOM, HETATM, SIGATM, SIGUIJ, ANISOU, and TER records for each model structure are interspersed as needed between MODEL and ENDMDL records.

\* The numbering of models is sequential beginning with 1.

\* If a collection contains more than 99,999 total atoms, then more than one entry must be made. In such a case the collection is divided between models (between an ENDMDL and the following MODEL record) and the model numbering is sequential throughout such a set of entries.

#### Verification/Validation/Value Authority Control

Entries with multiple structures in the EXPDTA record are checked for corresponding pairs of MODEL/ENDMDL records, and for consecutively numbered models.

#### Relationships to Other Record Types

Each MODEL must have a corresponding ENDMDL record.

In the case of an NMR entry the EXPDTA record states the number of model structures that are present in the individual entry.

**Example**

```

      1      2      3      4      5      6      7      8
123456789012345678901234567890123456789012345678901234567890
MODEL      1
ATOM      1  N  ALA      1      11.104    6.134   -6.504    1.00    0.00      N
ATOM      2  CA ALA      1      11.639    6.071   -5.147    1.00    0.00      C
...
ATOM     293 1HG  GLU     18     -14.861   -4.847    0.361    1.00    0.00      H
ATOM     294 2HG  GLU     18     -13.518   -3.769    0.084    1.00    0.00      H
TER      295      GLU     18
ENDMDL
MODEL      2
ATOM     296  N  ALA      1      10.883    6.779   -6.464    1.00    0.00      N
ATOM     297  CA ALA      1      11.451    6.531   -5.142    1.00    0.00      C
...
ATOM     588 1HG  GLU     18     -13.363   -4.163   -2.372    1.00    0.00      H
ATOM     589 2HG  GLU     18     -12.634   -3.023   -3.475    1.00    0.00      H
TER      590      GLU     18
ENDMDL

```

**ATOM****Overview**

The ATOM records present the atomic coordinates for standard residues. They also present the occupancy and temperature factor for each atom. Heterogen coordinates use the HETATM record type. The element symbol is always present on each ATOM record; segment identifier and charge are optional.

**Record Format**

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"ATOM "	
7 - 11	Integer	serial	Atom serial number.
13 - 16	Atom	name	Atom name.
17	Character	altLoc	Alternate location indicator.
18 - 20	Residue name	resName	Residue name.
22	Character	chainID	Chain identifier.
23 - 26	Integer	resSeq	Residue sequence number.
27	AChar	iCode	Code for insertion of residues.
31 - 38	Real(8.3)	x	Orthogonal coordinates for X in Angstroms.
39 - 46	Real(8.3)	y	Orthogonal coordinates for Y in Angstroms.
47 - 54	Real(8.3)	z	Orthogonal coordinates for Z in Angstroms.

55 - 60	Real(6.2)	occupancy	Occupancy.
61 - 66	Real(6.2)	tempFactor	Temperature factor.
73 - 76	LString(4)	segID	Segment identifier, left-justified.
77 - 78	LString(2)	element	Element symbol, right-justified.
79 - 80	LString(2)	charge	Charge on the atom.

**Details**

- \* ATOM records for proteins are listed from amino to carboxyl terminus.
- \* Nucleic acid residues are listed from the 5' to the 3' terminus.
- \* No ordering is specified for polysaccharides.
- \* The list of ATOM records in a chain is terminated by a TER record.
- \* If more than one model is present in the entry, each model is delimited by MODEL and ENDMDL records.
- \* For more information on atom naming conventions, see Appendix 3, and for residue names, see Appendix 4 and the HET section of this document
- \* If an atom is provided in more than one position, then a non-blank alternate location indicator must be used as the alternate location indicator for each of the positions. Within a residue all atoms that are associated with each other in a given conformation are assigned the same alternate position indicator.
- \* For atoms that are in alternate sites indicated by the alternate site indicator, sorting of atoms in the ATOM/HETATM list uses the following general rules:
  - In the simple case that involves a few atoms or a few residues with alternate sites, the coordinates occur one after the other in the entry.
  - In the case of a whole macromolecular chain, or significant portion of a chain, having alternate sites, the atoms for each alternate position are listed together. The two conformers are delineated by MODEL/ENDMDL records. In this case each MODEL must represent the entire molecular assemblage, including any heterogen group which is not necessarily disordered. Such is the case when DNA molecules are placed in UP and DOWN positions.
  - In the case of a large heterogen groups which are disordered, the atoms for each conformer are listed together. The two lists are not separated by MODEL/ENDMDL as is done for macromolecular chains.
- \* Addition of atoms to side chains of standard residues are handled as follows:
 

The additional atoms (modifying group) are represented as a HET group which is assigned its own residue name. The chainID, sequence number, and insertion code assigned to the HET group is that of the standard residue to which it is attached.
- \* Chemical modifications of standard residue side chains by addition of new atoms are handled as follows:
  - The new atoms are represented as a HET group. This group is assigned the chain name, sequence number, and insertion code of the standard residue that it modifies.
  - The atoms comprising these het groups are listed as HETATM and are inserted in the ATOM list immediately after the TER record of the chain. These groups are listed in the same order as the standard residue to which they are bonded (i.e. from the N- to C-terminus for polypeptides and from the 5' to 3' end for nucleic acids).

- Modified standard residues and the modifying het group may be assigned the same SEGID to further describe the relationship between the groups. PDB will use this mechanism only if SEGID's were not assigned to these atoms for other purposes.

- Modified standard residues must have a corresponding MODRES record.

\* The insertion code is commonly used in sequence numbering and is described here. In most cases, the amino acids that comprise a protein are numbered sequentially starting with 1. However, there are a number of situations that may give rise to different numbering schemes:

- Homologous proteins can exist in a number of different species. Depositors may use a residue numbering scheme in order to preserve the homology. The reference protein may be numbered sequentially starting with 1, then the homologous protein from another species aligned to it. If residues are not present in the homologous sequence, residue numbers may be skipped so that alignment can be preserved. If additional residues are present relative to the reference protein, they may have a letter, called an insertion code, appended to the sequence number. Negative numbers and zeros are permitted if they are needed to align the N-terminus.

REFERENCE PROTEIN NUMBERING	HOMOLOGOUS PROTEIN NUMBERING
59	59
60	60
61	
62	62

  

REFERENCE PROTEIN NUMBERING	HOMOLOGOUS PROTEIN NUMBERING
85	85
86	86
	86A
	86B
87	87

- The numbering of a proenzyme may be used for the enzyme following cleavage.

- The molecule studied might be a portion of the whole protein. The residue numbering scheme could show the relationship to the intact protein.

- The protein might be a mutant with residues inserted and deleted. As above, the residue numbering of the native protein could be preserved by appropriate use of gaps in the numbering and/or insertion codes.

- The nucleic acid community generally numbers structures sequentially. For double-stranded nucleic acids, entries usually use two different chain identifiers. For example, an octameric duplex would be numbered 1 - 8 for chain A, and 9 - 16 for chain B.

\* If the depositor provides the data, then the isotropic B value is given for the temperature factor.

\* If there is no isotropic B value from the depositor, but there is an ANISOU record with anisotropic temperature factors, then the B equivalent is stored in the tempFactor field, as calculated by:

$$B(\text{eq}) = 8\pi^2 \{ 1/3 [U(1,1) + U(2,2) + U(3,3)] \}$$

- This will obviate the need to check if ANISOU records are present before interpreting the contents of the temperature factor field.

- In some previously released PDB entries with anisotropic temperature factors provided as ANISOU records, the temperature factor field of the corresponding ATOM or HETATM record contained the equivalent U-isotropic [U(eq)] which is calculated by:

$$U(\text{eq}) = 1/3 [U(1,1) + U(2,2) + U(3,3)] \times 10^{-4}$$

\* If there are neither isotropic B values from the depositor, nor anisotropic temperature factors in ANISOU, then the default value of 0.0 is used for the temperature factor.

\* In some entries, the occupancy and temperature factor fields are used for other quantities. In these cases, an explanation is provided in the remarks.

\* Columns 73 - 76 identify specific segments of the molecule. The segment id is a string of up to four (4) alphanumeric characters, left-justified, and may include a space, e.g., CH86, A 1, NASE. The segment itself may consist of a complete chain or a portion of a chain. The importance of this new field can be appreciated if one considers an antibody structure having two molecules in the asymmetric unit. Since each chain must have a unique chain identifier, the two heavy chains and two light chains cannot currently be labeled to indicate their nature. Segment id's of CH, VH1, VH2, VH3, CL, and VL would clearly identify regions of the chains and the relationship between them. Users of X-PLOR will be familiar with SEGID as used in the refinement application of X-PLOR.

\* Columns 77 - 78 contain the atom's element symbol (as given in the periodic table), right-justified. This is especially needed because in some cases it has not been possible to follow the convention that columns 13 - 14 of the atom name contain the element symbol. The most common cases are:

- In large het groups it sometimes is not possible to follow the convention of having the first two characters be the chemical symbol and still use atom names that are meaningful to users. A example is nicotinamide adenine dinucleotide, atom names begin with an A or N, depending on which portion of the molecule they appear in, e.g., AC6 or NC6, AN1 or NN1.

- Hydrogen naming sometimes conflicts with IUPAC conventions. For example, a hydrogen named HG11 in columns 13 - 16 is differentiated from a mercury atom by the element symbol in columns 77 - 78. Columns 13 - 16 present a unique name for each atom.

\* Columns 79 - 80 indicate any charge on the atom, e.g., 2+, 1-. In most cases these are blank.

#### Verification/Validation/Value Authority Control

PDB checks ATOM/HETATM records for PDB format, sequence information, and packing. The PDB reserves the right to return deposited coordinates to the author for transformation into PDB format.

PDB intends to verify the coordinates against the experimental structure factor data in the when available. Details on this will be forthcoming.

#### Relationships to Other Record Types

The ATOM records are compared to the corresponding sequence database. Residue discrepancies appear in the SEQADV record. Missing atoms are annotated in the remarks. HETATM records are formatted in the same way as ATOM records. The sequence implied by ATOM records must be identical to that given in SEQRES, with the exception that residues that have no coordinates, e.g., due to disorder, must appear in SEQRES. Remark 550 is used to describe the meaning assigned to any segment identifiers used.

#### Example

	1	2	3	4	5	6	7	8
1234567890123456789012345678901234567890123456789012345678901234567890								
ATOM	145	N	VAL A	25	32.433	16.336	57.540	1.00 11.92 A1 N
ATOM	146	CA	VAL A	25	31.132	16.439	58.160	1.00 11.85 A1 C
ATOM	147	C	VAL A	25	30.447	15.105	58.363	1.00 12.34 A1 C
ATOM	148	O	VAL A	25	29.520	15.059	59.174	1.00 15.65 A1 O
ATOM	149	CB	AVAL A	25	30.385	17.437	57.230	0.28 13.88 A1 C
ATOM	150	CB	BVAL A	25	30.166	17.399	57.373	0.72 15.41 A1 C
ATOM	151	CG1A	AVAL A	25	28.870	17.401	57.336	0.28 12.64 A1 C
ATOM	152	CG1B	AVAL A	25	30.805	18.788	57.449	0.72 15.11 A1 C
ATOM	153	CG2A	AVAL A	25	30.835	18.826	57.661	0.28 13.58 A1 C
ATOM	154	CG2B	AVAL A	25	29.909	16.996	55.922	0.72 13.25 A1 C

**Known Problems**

Due to the ever-increasing size of protein structures in the PDB, the atom serial number field may soon need to be increased. An increase of one column will allow for cases where entries have more than 99,999 atoms. Only 5 digits are available for the atom serial number, but some structures have already been received with more than 99,999 atoms.

No distinction is made between ribo- and deoxyribonucleotides in the SEQRES records. These residues are identified with the same residue name (i.e., A, C, G, T, U).

---

**SIGATM****Overview**

The SIGATM records present the standard deviation of atomic parameters as they appear in ATOM and HETATM records.

**Record Format**

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"SIGATM"	
7 - 11	Integer	serial	Atom serial number.
13 - 16	Atom	name	Atom name.
17	Character	altLoc	Alternate location indicator.
18 - 20	Residue name	resName	Residue name.
22	Character	chainID	Chain identifier.
23 - 26	Integer	resSeq	Residue sequence number.
27	AChar	iCode	Insertion code.
31 - 38	Real(8.3)	sigX	Standard deviations of the stored coordinates (Angstroms).
39 - 46	Real(8.3)	sigY	Standard deviations of the stored coordinates (Angstroms).
47 - 54	Real(8.3)	sigZ	Standard deviations of the stored coordinates (Angstroms).
55 - 60	Real(6.2)	sigOcc	Standard deviation of occupancy.
61 - 66	Real(6.2)	sigTemp	Standard deviation of temperature factor.
73 - 76	LString(4)	segID	Segment identifier, left-justified.
77 - 78	LString(2)	element	Element symbol, right-justified.
79 - 80	LString(2)	charge	Charge on the atom.

**Details**

\* Columns 7 - 27 are identical to the corresponding ATOM/HETATM record.

\* Each SIGATM record immediately follows the corresponding ATOM/HETATM record.

\* SIGATM is provided only for ATOM/HETATM records for which values are supplied by the depositor and only when the value is not zero (0).

**Verification/Validation/Value Authority Control**

The depositor provides SIGATM records, PDB verifies their format.

**Relationships to Other Record Types**

SIGATM is related to the immediately preceding ATOM/HETATM record.

**Example**

	1	2	3	4	5	6	7	8
1234567890123456789012345678901234567890123456789012345678901234567890								
ATOM	230	N	PRO	15	20.860	29.640	13.460	1.00 12.20
SIGATM	230	N	PRO	15	0.040	0.030	0.030	0.00 0.00
ATOM	231	CA	PRO	15	22.180	29.010	12.960	1.00 14.70
SIGATM	231	CA	PRO	15	0.060	0.040	0.050	0.00 0.00
ATOM	232	C	PRO	15	23.170	30.090	12.670	1.00 19.10
SIGATM	232	C	PRO	15	0.080	0.070	0.060	0.00 0.00
ATOM	233	O	PRO	15	24.360	29.860	12.670	1.00 17.50
SIGATM	233	O	PRO	15	0.040	0.030	0.030	0.00 0.00
ATOM	234	CB	PRO	15	21.710	28.220	11.640	1.00 17.70
SIGATM	234	CB	PRO	15	0.060	0.040	0.050	0.00 0.00
ATOM	235	CG	PRO	15	20.470	28.710	11.590	1.00 23.90
SIGATM	235	CG	PRO	15	0.080	0.060	0.060	0.00 0.00
ATOM	236	CD	PRO	15	19.640	29.320	12.660	1.00 15.50
SIGATM	236	CD	PRO	15	0.060	0.040	0.050	0.00 0.00
ATOM	237	HA	PRO	15	22.630	28.400	13.620	1.00 14.70
ATOM	238	1HB	PRO	15	22.240	28.540	10.860	1.00 17.70
ATOM	239	2HB	PRO	15	21.670	27.240	11.840	1.00 17.70
ATOM	240	1HG	PRO	15	20.360	29.240	10.740	1.00 23.90
ATOM	241	2HG	PRO	15	19.900	28.120	11.020	1.00 23.90
ATOM	242	1HD	PRO	15	19.230	30.160	12.320	1.00 15.50
ATOM	243	2HD	PRO	15	19.120	28.600	13.120	1.00 15.50

**ANISOU****Overview**

The ANISOU records present the anisotropic temperature factors.

**Record Format**

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"ANISOU"	
7 - 11	Integer	serial	Atom serial number.
13 - 16	Atom	name	Atom name.
17	Character	altLoc	Alternate location indicator.

18 - 20	Residue name	resName	Residue name.
22	Character	chainID	Chain identifier.
23 - 26	Integer	resSeq	Residue sequence number.
27	AChar	iCode	Insertion code.
29 - 35	Integer	u[0][0]	U(1,1)
36 - 42	Integer	u[1][1]	U(2,2)
43 - 49	Integer	u[2][2]	U(3,3)
50 - 56	Integer	u[0][1]	U(1,2)
57 - 63	Integer	u[0][2]	U(1,3)
64 - 70	Integer	u[1][2]	U(2,3)
73 - 76	LString(4)	segID	Segment identifier, left-justified.
77 - 78	LString(2)	element	Element symbol, right-justified.
79 - 80	LString(2)	charge	Charge on the atom.

**Details**

\* Columns 7 - 27 are identical to the corresponding ATOM/HETATM record.

\* The anisotropic temperature factors (columns 29 - 70) are scaled by a factor of 10\*\*4 (Angstroms\*\*2) and are presented as integers.

\* The anisotropic temperature factors are stored in the same coordinate frame as the atomic coordinate records.

\* ANISOU values are listed only if they have been provided by the depositor.

**Verification/Validation/Value Authority Control**

The depositor provides ANISOU records, PDB verifies their format.

**Relationships to Other Record Types**

The anisotropic temperature factors are related to the corresponding ATOM/HETATM isotropic temperature factors as B(eq), as described in the ATOM and HETATM sections.

**Example**

	1	2	3	4	5	6	7	8
1234567890123456789012345678901234567890123456789012345678901234567890								
ATOM	107	N	GLY	13	12.681	37.302	-25.211	1.000 15.56 N
ANISOU	107	N	GLY	13	2406	1892	1614	198 519 -328 N
ATOM	108	CA	GLY	13	11.982	37.996	-26.241	1.000 16.92 C
ANISOU	108	CA	GLY	13	2748	2004	1679	-21 155 -419 C
ATOM	109	C	GLY	13	11.678	39.447	-26.008	1.000 15.73 C
ANISOU	109	C	GLY	13	2555	1955	1468	87 357 -109 C
ATOM	110	O	GLY	13	11.444	40.201	-26.971	1.000 20.93 O
ANISOU	110	O	GLY	13	3837	2505	1611	164 -121 189 O
ATOM	111	N	ASN	14	11.608	39.863	-24.755	1.000 13.68 N
ANISOU	111	N	ASN	14	2059	1674	1462	27 244 -96 N



## SIGUIJ

### Overview

The SIGUIJ records present the standard deviations of anisotropic temperature factors scaled by a factor of  $10^{**4}$  (Angstroms\*\*2).

### Record Format

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"SIGUIJ"	
7 - 11	Integer	serial	Atom serial number.
13 - 16	Atom	name	Atom name.
17	Character	altLoc	Alternate location indicator.
18 - 20	Residue name	resName	Residue name.
22	Character	chainID	Chain identifier.
23 - 26	Integer	resSeq	Residue sequence number.
27	AChar	iCode	Insertion code.
29 - 35	Integer	sig[1][1]	Sigma U(1,1)
36 - 42	Integer	sig[2][2]	Sigma U(2,2)
43 - 49	Integer	sig[3][3]	Sigma U(3,3)
50 - 56	Integer	sig[1][2]	Sigma U(1,2)
57 - 63	Integer	sig[1][3]	Sigma U(1,3)
64 - 70	Integer	sig[2][3]	Sigma U(2,3)
73 - 76	LString(4)	segID	Segment identifier, left-justified.
77 - 78	LString(2)	element	Element symbol, right-justified.
79 - 80	LString(2)	charge	Charge on the atom.

### Details

\* Columns 7 - 27 are identical to the corresponding ATOM/HETATM record.

\* SIGUIJ are listed only if they have been provided by the depositor and only if they are not zero (0).

### Verification/Validation/Value Authority Control

The depositor provides SIGUIJ records, PDB verifies their format.

### Relationships to Other Record Types

The standard deviations for the anisotropic temperature factors are related to the corresponding ATOM/HETATM ANISOU temperature factors.

**Example**

	1	2	3	4	5	6	7	8
1234567890123456789012345678901234567890123456789012345678901234567890								
ATOM	107	N	GLY	13	12.681	37.302	-25.211	1.000 15.56 N
ANISOU	107	N	GLY	13	2406	1892	1614	198 519 -328 N
SIGUIJ	107	N	GLY	13	10	10	10	10 10 N
ATOM	108	CA	GLY	13	11.982	37.996	-26.241	1.000 16.92 C
ANISOU	108	CA	GLY	13	2748	2004	1679	-21 155 -419 C
SIGUIJ	108	CA	GLY	13	10	10	10	10 10 C
ATOM	109	C	GLY	13	11.678	39.447	-26.008	1.000 15.73 C
ANISOU	109	C	GLY	13	2555	1955	1468	87 357 -109 C
SIGUIJ	109	C	GLY	13	10	10	10	10 10 C
ATOM	110	O	GLY	13	11.444	40.201	-26.971	1.000 20.93 O
ANISOU	110	O	GLY	13	3837	2505	1611	164 -121 189 O
SIGUIJ	110	O	GLY	13	10	10	10	10 10 O
ATOM	111	N	ASN	14	11.608	39.863	-24.755	1.000 13.68 N
ANISOU	111	N	ASN	14	2059	1674	1462	27 244 -96 N
SIGUIJ	111	N	ASN	14	10	10	10	10 10 N

**TER****Overview**

The TER record indicates the end of a list of ATOM/HETATM records for a chain.

**Record Format**

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"TER "	
7 - 11	Integer	serial	Serial number.
18 - 20	Residue name	resName	Residue name.
22	Character	chainID	Chain identifier.
23 - 26	Integer	resSeq	Residue sequence number.
27	AChar	iCode	Insertion code.

**Details**

\* Every chain of ATOM/HETATM records presented on SEQRES records is terminated with a TER record.

\* The TER records occur in the coordinate section of the entry, and indicate the last residue presented for each polypeptide and/or nucleic acid chain for which there are coordinates. For proteins, the residue defined on the TER record is the carboxy-terminal residue; for nucleic acids it is the 3'-terminal residue.

\* For a cyclic molecule, the choice of termini is arbitrary.

\* Terminal oxygen atoms are presented as OXT for proteins, and as O5T or O3T for nucleic acids.

\* The TER record has the same residue name, chain identifier, sequence number and insertion code as the terminal residue. The serial number of the TER record is one number greater than the serial number of the ATOM/HETATM preceding the TER.

\* For chains with gaps due to disorder, it is recommended that the C-terminus atom be labelled O and OXT, and a REMARK explaining the ambiguity be provided.

**Verification/Validation/Value Authority Control**

TER must appear at the end of a chain. For proteins, there is usually a terminal oxygen, labeled OXT. The validation program checks for the occurrence of TER and OXT records.

**Relationships to Other Record Types**

The residue name appearing on the TER record must be the same as the residue name of the immediately preceding ATOM or non-water HETATM record.

**Example**

	1	2	3	4	5	6	7	8
1234567890123456789012345678901234567890123456789012345678901234567890								
ATOM	4150	H	ALA A 431	8.674	16.036	12.858	1.00	0.00
TER	4151		ALA A 431					H
ATOM	1403	O	PRO P 2	12.701	33.564	15.827	1.09	18.03
ATOM	1404	CB	PRO P 2	13.512	32.617	18.642	1.09	9.32
ATOM	1405	CG	PRO P 2	12.828	33.382	19.740	1.09	12.23
ATOM	1406	CD	PRO P 2	12.324	34.603	18.985	1.09	11.47
HETATM	1407	CA	BLE P 1	14.625	32.240	14.151	1.09	16.76
HETATM	1408	CB	BLE P 1	15.610	33.091	13.297	1.09	16.56
HETATM	1409	CG	BLE P 1	15.558	34.629	13.373	1.09	14.27
HETATM	1410	CD1	BLE P 1	16.601	35.208	12.440	1.09	14.75
HETATM	1411	CD2	BLE P 1	14.209	35.160	12.930	1.09	15.60
HETATM	1412	N	BLE P 1	14.777	32.703	15.531	1.09	14.79
HETATM	1413	B	BLE P 1	14.921	30.655	14.194	1.09	15.56
HETATM	1414	O1	BLE P 1	14.852	30.178	12.832	1.09	16.10
HETATM	1415	O2	BLE P 1	13.775	30.147	14.862	1.09	20.95
TER	1416		BLE P 1					

---

**HETATM****Overview**

The HETATM records present the atomic coordinate records for atoms within "non-standard" groups. These records are used for water molecules and atoms presented in HET groups.

**Record Format**

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"HETATM"	
7 - 11	Integer	serial	Atom serial number.
13 - 16	Atom	name	Atom name.
17	Character	altLoc	Alternate location indicator.
18 - 20	Residue name	resName	Residue name.
22	Character	chainID	Chain identifier.
23 - 26	Integer	resSeq	Residue sequence number.
27	AChar	iCode	Code for insertion of residues.
31 - 38	Real(8.3)	x	Orthogonal coordinates for X.

39 - 46	Real (8.3)	y	Orthogonal coordinates for Y.
47 - 54	Real (8.3)	z	Orthogonal coordinates for Z.
55 - 60	Real (6.2)	occupancy	Occupancy.
61 - 66	Real (6.2)	tempFactor	Temperature factor.
73 - 76	LString(4)	segID	Segment identifier; left-justified.
77 - 78	LString(2)	element	Element symbol; right-justified.
79 - 80	LString(2)	charge	Charge on the atom.

**Details**

\* The x, y, z coordinates are in Angstrom units.

\* Disordered solvents may be represented by the residue name DIS.

\* No ordering is specified for polysaccharides.

\* See the HET section of this document regarding naming of heterogens. See the HET dictionary (soon to be released) for residue names, formulas, and CONECT records of the HET groups that have appeared so far in the PDB.

\* For atoms that are in alternate sites indicated by the alternate site indicator, sorting of atoms in the ATOM/HETATM list uses the following general rules:

- In the simple case that involves a few atoms or a few residues with alternate sites, the coordinates occur one after the other in the entry.

- In the case of a whole macromolecular chain, or significant portion of a chain, having alternate sites, the atoms for each alternate position are listed together. The two conformers are delineated by MODEL/ENDMDL records. In this case each MODEL must represent the entire molecular assemblage, including any heterogen group which is not necessarily disordered. Such is the case when DNA molecules are placed in UP and DOWN positions.

- In the case of a large heterogen groups which are disordered, the atoms for each conformer are listed together. The two lists are not separated by MODEL/ENDMDL as is done for macromolecular chains.

\* If the depositor provides the data, then the isotropic B value is given for the temperature factor.

\* If there is no isotropic B value from the depositor, but there is an ANISOU record with anisotropic temperature factors, then the B equivalent is stored in the tempFactor field, as calculated by:

$$B(eq) = 8\pi^2 \{ 1/3 [U(1,1) + U(2,2) + U(3,3)] \}$$

- This will obviate the need to check if ANISOU records are present before interpreting the contents of the temperature factor field.

- In some previously released PDB entries with anisotropic temperature factors provided as ANISOU records, the temperature factor field of the corresponding ATOM or HETATM record contained the equivalent U-isotropic [U(eq)] which is calculated by:

$$U(eq) = 1/3 [U(1,1) + U(2,2) + U(3,3)] \times 10^{-4}$$

\* If there are neither isotropic B values from the depositor, nor anisotropic temperature factors in ANISOU, then the default value of 0.0 is used for the temperature factor.

\* In some entries, the occupancy and temperature factor fields are often used for other quantities. In these cases, an explanation is provided in the remarks.

\* Insertion codes, segment id, and element naming are fully described in the ATOM section of this document.

### Verification/Validation/Value Authority Control

PDB processing programs check ATOM/HETATM records for PDB format, sequence information, and packing. The PDB reserves the right to return deposited coordinates to the author for transformation into PDB format.

### Relationships to Other Record Types

HETATM records must have corresponding HET, HETNAM, FORMUL and CONECT records, except for waters.

### Example

	1	2	3	4	5	6	7	8
1234567890123456789012345678901234567890123456789012345678901234567890								
HETATM 1357 MG	MG	168	4.669	34.118	19.123	1.00	3.16	MG2+
HETATM 3835 FE	HEM	1	17.140	3.115	15.066	1.00	14.14	FE3+

## ENDMDL

### Overview

The ENDMDL records are paired with MODEL records to group individual structures found in a coordinate entry.

### Record Format

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"ENDMDL"	

### Details

\* MODEL/ENDMDL records are used only when more than one structure is presented in the entry, as is often the case with NMR entries.

\* All the models in a multi-model entry must represent the same structure.

\* Every MODEL record has an associated ENDMDL record.

### Verification/Validation/Value Authority Control

Entries with multiple structures in the EXPDTA record are checked for corresponding pairs of MODEL/ENDMDL records, and for consecutively numbered models.

### Relationships to Other Record Types

There must be a corresponding MODEL record.

In the case of an NMR entry the EXPDTA record states the number of model structures that are present in the individual entry.

## Example

	1	2	3	4	5	6	7	8
1234567890123456789012345678901234567890123456789012345678901234567890								
...								
...								
ATOM	14550	1HG	GLU	122	-14.364	14.787	-14.258	1.00 0.00 H
ATOM	14551	2HG	GLU	122	-13.794	13.738	-12.961	1.00 0.00 H
TER	14552		GLU	122				
ENDMDL								
MODEL	9							
ATOM	14553	N	SER	1	-28.280	1.567	12.004	1.00 0.00 N
ATOM	14554	CA	SER	1	-27.749	0.392	11.256	1.00 0.00 C
...								
...								
ATOM	16369	1HG	GLU	122	-3.757	18.546	-8.439	1.00 0.00 H
ATOM	16370	2HG	GLU	122	-3.066	17.166	-7.584	1.00 0.00 H
TER	16371		GLU	122				
ENDMDL								
MODEL	10							
ATOM	16372	N	SER	1	-22.285	7.041	10.003	1.00 0.00 N
ATOM	16373	CA	SER	1	-23.026	6.872	8.720	1.00 0.00 C
...								
...								
ATOM	18188	1HG	GLU	122	-1.467	18.282	-17.144	1.00 0.00 H
ATOM	18189	2HG	GLU	122	-2.711	18.067	-15.913	1.00 0.00 H
TER	18190		GLU	122				
ENDMDL								

## 10. Connectivity Section

This section provides information on chemical connectivity. LINK, HYDBND, SLTBRG, and CISPEP are found in the Connectivity Annotation section.

---

### CONECT

#### Overview

The CONECT records specify connectivity between atoms for which coordinates are supplied. The connectivity is described using the atom serial number as found in the entry. CONECT records are mandatory for HET groups (excluding water) and for other bonds not specified in the standard residue connectivity table which involve atoms in standard residues (see Appendix 4 for the list of standard residues). These records are generated by the PDB.

#### Record Format

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"CONECT"	
7 - 11	Integer	serial	Atom serial number
12 - 16	Integer	serial	Serial number of bonded atom
17 - 21	Integer	serial	Serial number of bonded atom
22 - 26	Integer	serial	Serial number of bonded atom
27 - 31	Integer	serial	Serial number of bonded atom
32 - 36	Integer	serial	Serial number of hydrogen bonded atom
37 - 41	Integer	serial	Serial number of hydrogen bonded atom
42 - 46	Integer	serial	Serial number of salt bridged atom
47 - 51	Integer	serial	Serial number of hydrogen bonded atom
52 - 56	Integer	serial	Serial number of hydrogen bonded atom
57 - 61	Integer	serial	Serial number of salt bridged atom

#### Details

\* Intra-residue connectivity within non-standard (HET) residues (excluding water) is presented on the CONECT records.

\* Inter-residue connectivity of HET groups to standard groups (including water) or to other HET groups are represented on the CONECT records.

\* Disulfide bridges specified in the SSBOND records have corresponding CONECT records.

\* Hydrogen bonds and salt bridges also have CONECT records.

- \* No differentiation is made between donor and acceptor for hydrogen bonds.
- \* No differentiation is made between atoms with excess negative or positive charge.
- \* Atoms specified in the connectivity are presented by their serial numbers as found in the entry.
- \* All atoms connected to the atom with serial number in columns 7 - 11 are listed in the remaining fields of the record.
- \* If more than four fields are required for non-hydrogen and nonsalt-bridge bonds, a second CONECT record with the same atom serial number in columns 7 - 11 will be used.
- \* These CONECT records occur in increasing order of the atom serial numbers they carry in columns 7 - 11. The target-atom serial numbers carried on these records also occur in increasing order.
- \* The connectivity list given here is redundant in that each bond indicated is given twice, once with each of the two atoms involved specified in columns 7 - 11.
- \* For nucleic acids, Watson-Crick hydrogen bonds between bases may be listed, but this is optional.
- \* For hydrogen bonds, when the hydrogen atom is present in the coordinates, PDB generates a CONECT record between the hydrogen atom and its acceptor atom.

#### Verification/Validation/Value Authority Control

Connectivity is checked for unusual bond lengths.

#### Relationships to Other Record Types

CONECT records must be present in an entry that contains either non-standard groups or disulfide bonds.

#### Example

	1	2	3	4	5	6	7
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890
CONECT	1179	746 1184	1195 1203				
CONECT	1179	1211 1222					
CONECT	1021	544 1017	1020 1022	1211 1222	1311		

#### Known Problems

Only five digits are available for the atom serial number, but some structures have already been received with more than 99,999 atoms. Changing the field length would make earlier entries incorrect.

CONECTs to atoms whose coordinates are not in the entry (e.g., symmetry-generated) are not given.



## 11. Bookkeeping Section

The Bookkeeping Section provides some final information about the file itself.

---

### MASTER

#### Overview

The MASTER record is a control record for bookkeeping. It lists the number of lines in the coordinate entry or file for selected record types.

#### Record Format

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"MASTER"	
11 - 15	Integer	numRemark	Number of REMARK records
16 - 20	Integer	"0"	
21 - 25	Integer	numHet	Number of HET records
26 - 30	Integer	numHelix	Number of HELIX records
31 - 35	Integer	numSheet	Number of SHEET records
36 - 40	Integer	numTurn	Number of TURN records
41 - 45	Integer	numSite	Number of SITE records
46 - 50	Integer	numXform	Number of coordinate transformation records (ORIGX+SCALE+MTRIX)
51 - 55	Integer	numCoord	Number of atomic coordinate records (ATOM+HETATM)
56 - 60	Integer	numTer	Number of TER records
61 - 65	Integer	numConect	Number of CONECT records
66 - 70	Integer	numSeq	Number of SEQRES records

#### Details

\* MASTER gives checksums of the number of records in the entry, for selected record types.

#### Verification/Validation/Value Authority Control

The MASTER line is generated by the PDB.

#### Relationships to Other Record Types

MASTER presents a checksum of the lines present for each of the record types listed above.

**Example**

	1		2		3		4		5		6		7
1234567890123456789012345678901234567890123456789012345678901234567890													
MASTER	40	0	0	0	0	0	0	0	6	2930	2	0	29

**END****Overview**

The END record marks the end of the PDB file.

**Record Format**

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"END"	"

**Details**

\* END is the final record of a coordinate entry.

**Verification/Validation/Value Authority Control**

END must appear in every coordinate entry.

**Relationships to Other Record Types**

This is the final record in the entry.

**Example**

	1		2		3		4		5		6		7
1234567890123456789012345678901234567890123456789012345678901234567890													
END													

# Protein Data Bank Contents Guide:

## Atomic Coordinate Entry Format

### Description:

### Appendices

## Appendix 1: Symmetry Operations

The data type SymOP is used to succinctly describe crystallographic symmetry operations that may be performed on ATOM/HETATM coordinates. Symmetry operators applicable to a given entry are presented in REMARK 350. Each operator is assigned a serial number. The SymOP is a number of up to six (6) digits that indicates the serial number of the symmetry operator and the cell translations along the x, y, and z axes.

The SymOP data type is of the form nnnMMM where 'n' is the serial number of the symmetry operator, and 'MMM' is the concatenated cell translations along x, y, z with respect to the base number 555. Symmetry operators listed in REMARK 350 operate on fractional crystallographic coordinates.

The FORTRAN I3 I3 format statement can be used to interpret nnnMMM.

As an example, the SymOP 2456 indicates that the second symmetry operation as listed in REMARK 350 is applied with translation of -1 on x, and +1 on z. A program will be made available shortly that converts SymOP data into transformations that operate in the coordinate frame used in the entry.

The SymOP data type is used in SSBOND, LINK, HYDBND, SLTBRG and REMARKs.

### Template

```
REMARK 350
REMARK 350 SYMMETRY
REMARK 350 NON-CRYSTALLOGRAPHIC SYMMETRY TRANSFORMATIONS: Number
REMARK 350
REMARK 350 THE TRANSFORMATIONS PRESENTED ON THE MTRIX RECORDS BELOW
REMARK 350 DESCRIBE NON-CRYSTALLOGRAPHIC RELATIONSHIPS AMONG THE
REMARK 350 VARIOUS DOMAINS IN THIS ENTRY. APPLYING THE APPROPRIATE
REMARK 350 MTRIX TRANSFORMATION TO THE RESIDUES LISTED FIRST WILL
REMARK 350 YIELD APPROXIMATE COORDINATES FOR THE RESIDUES LISTED
REMARK 350 SECOND. CHAIN IDENTIFIERS GIVEN AS "?" REFER TO CHAINS
REMARK 350 FOR WHICH ATOMS ARE NOT FOUND IN THIS ENTRY.
REMARK 350
REMARK 350          APPLIED TO          TRANSFORMED TO
REMARK 350 TRANSFORM CHAIN RESIDUES    CHAIN RESIDUES    RMSD
REMARK 350   MS           A      1 .. 999      B      1 .. 999    0.32
REMARK 350
REMARK 350   HERE      M -> TRANSFORMATION MATRIX (MTRIX      S)
REMARK 350           S -> MATRIX SERIAL NUMBER
REMARK 350
REMARK 350 CRYSTALLOGRAPHIC SYMMETRY TRANSFORMATIONS: Number
REMARK 350
REMARK 350 THE TRANSFORMATIONS PRESENTED ON THE SYMTRY RECORDS
REMARK 350 LISTED IN THIS REMARK DESCRIBE THE CRYSTALLOGRAPHIC
REMARK 350 SYMMETRY TRANSFORMATIONS NECESSARY TO GENERATE THE
REMARK 350 BIOLOGICALLY SIGNIFICANT OLIGOMERIZATION STATE OF
```

```
REMARK 350  THE MOLECULE.  APPLYING THE APPROPRIATE SYMTRY
REMARK 350  TRANSFORMATION TO THE RESIDUES LISTED WILL GENERATE
REMARK 350  THE SYMMETRY-RELATED DOMAIN.
REMARK 350
REMARK 350      OPERATOR  CRYSTALLOGRAPHIC  TRANSLATION  FRACTIONAL
REMARK 350      NUMBER    SYMMETRY OPERATOR    VECTOR    TRANSLATIONS
REMARK 350      001      -X,-Y,Z              545      0.0 -1.0  0.0
REMARK 350
REMARK 350  TRANSFORM          APPLIED TO
REMARK 350  NNNMMMS          CHAIN(S)    RESIDUES
REMARK 350  0015451          A            1 .. 98
REMARK 350
REMARK 350      HERE NNN -> OPERATOR NUMBER
REMARK 350      MMM -> TRANSLATION VECTOR
REMARK 350      S -> SMTRY MATRIX SERIAL NUMBER
REMARK 350
REMARK 350  CORRESPONDING SYMTRY MATRIX FOR TRANSFORM 0015551
REMARK 350
REMARK 350  SMTRY1 1  -1.000000  0.000000  0.000000      0.00000
REMARK 350  SMTRY2 1   0.000000 -1.000000  0.000000     -1.00000
REMARK 350  SMTRY3 1   0.000000  0.000000  1.000000      0.00000
```

---

## Appendix 2: Coordinate Systems and Transformations

The coordinates distributed by the Protein Data Bank give the atomic positions measured in Angstroms along three orthogonal directions. Unless otherwise specified, the default axial system detailed below is assumed.

If  $a$ ,  $b$ ,  $c$  describe the crystallographic cell edges and  $A$ ,  $B$ ,  $C$  are unit vectors in the default orthogonal Angstrom system, then the following apply.

$A$ ,  $B$ ,  $C$  and  $a$ ,  $b$ ,  $c$  have the same origin.

$A$  is parallel to  $a$ .

$B$  is parallel to  $(a \times b) \times A$  (cross product between  $C$  and  $A$ ).

$C$  is parallel to  $a \times b$  (i.e.,  $c^*$ ) (cross product between  $a$  and  $b$ ).

The matrix which pre-multiplies the column vector of the fractional crystallographic coordinates to yield the distributed coordinates in the  $A$ ,  $B$ ,  $C$  system is:

$$\begin{array}{lll} a & b(\cos(\gamma)) & c(\cos(\beta)) \\ 0 & b(\sin(\gamma)) & c(\cos(\alpha) - \cos(\beta) \cos(\gamma)) / \sin(\gamma) \\ 0 & 0 & V / (ab \sin(\gamma)) \end{array}$$

$$V = abc(1 - \cos^2(\alpha) - \cos^2(\beta) - \cos^2(\gamma) + 2(\cos(\alpha) \cos(\beta) \cos(\gamma)))^{1/2}$$

The distributed entry will contain the following records.

ORIGX - transformation from the distributed to the submitted coordinates.

SCALE - transformation from the distributed to the fractional coordinates.

## Appendix 3: Atom Names

### Amino Acids

The following rules are used in assigning atom names.

\* Greek letter remoteness codes are transliterated as follows: alpha = A, beta = B, gamma = G, delta = D, epsilon = E, zeta = Z, eta = H, etc.

\* Atoms for which some ambiguity exists in the crystallographic results are designated A. This usually applies only to the terminal atoms of asparagine and glutamine and to the ring atoms of histidine.

\* The extra oxygen atom of the carboxy terminal amino acid is designated OXT.

\* Six characters (columns) are reserved for atom names, assigned as follows.

COLUMN	VALUE
13 - 14	Chemical symbol - right justified, except for hydrogen atoms
15	Remoteness indicator (alphabetic)
16	Branch designator (numeric)
77 - 78	Element symbol, right-justified

\* Columns 73 - 76 identify specific segments of the molecule. The segment may consist of a complete chain or a portion of a chain. The importance of this new field can be appreciated if one considers an antibody structure having two molecules in the asymmetric unit. Since each chain must have a unique chain identifier, the two heavy chains and two light chains cannot currently be labeled to indicate their nature. Segment id's of CH, VH1, VH2, VH3, CL, and VL would clearly identify regions of the chains and the relationship between them. Users of X-PLOR will be familiar with SEGID as used in the refinement application of X-PLOR.

See the ATOM record for more details on atom naming.

### Nucleic Acids

Atom names employed for polynucleotides generally follow the precedent set for mononucleotides. The following points should be noted.

\* The asterisk (\*) is used in place of the prime character (') for naming atoms of the sugar group. The prime was avoided historically because of non-uniformity of its external representation.

\* The ring oxygen of the ribose is denoted O4 rather than O1.

\* The extra oxygen atom at the free 5' and 3' termini are designated O5T and O3T, respectively.

## Appendix 4: Standard Residue Names and Abbreviations

Note that there will be a change to what are considered standard groups due to the adoption of the new PDB Het Group Dictionary. Only the twenty common amino acids and five nucleic acids plus inosine will be treated as "standard" with all others being treated as modified residues to be described by MODRES records.

No distinction is made between ribo- and deoxyribonucleotides in the SEQRES records. These residues are identified with the same residue name (i.e., A, C, G, T, U, I).

### Amino Acids

RESIDUE	ABBREVIATION	SYNONYM
Alanine	ALA	A
Arginine	ARG	R
Asparagine	ASN	N
Aspartic acid	ASP	D
ASP/ASN ambiguous	ASX	B
Cysteine	CYS	C
Glutamine	GLN	Q
Glutamic acid	GLU	E
GLU/GLN ambiguous	GLX	Z
Glycine	GLY	G
Histidine	HIS	H
Isoleucine	ILE	I
Leucine	LEU	L
Lysine	LYS	K
Methionine	MET	M
Phenylalanine	PHE	F
Proline	PRO	P
Serine	SER	S
Threonine	THR	T
Tryptophan	TRP	W
Tyrosine	TYR	Y
Unknown	UNK	
Valine	VAL	V

### Nucleic Acids

RESIDUE	ABBREVIATION
Adenosine	A
Modified adenosine	+A
Cytidine	C
Modified cytidine	+C
Guanosine	G
Modified guanosine	+G
Inosine	I
Modified inosine	+I
Thymidine	T
Modified thymidine	+T
Uridine	U
Modified uridine	+U
Unknown	UNK

Remarks 103 and 104 are included when an entry contains inosine.

## Appendix 5: Formulas and Molecular Weights for Standard Residues

These weights and formulas correspond to the unpolymerized state of the component. The atoms of one water molecule are eliminated for each two components joined.

### Amino Acids

NAME	CODE	FORMULA	MOL. WT.
Alanine	ALA	C3 H7 N1 O2	89.09
Arginine	ARG	C6 H14 N4 O2	174.20
Asparagine	ASN	C4 H8 N2 O3	132.12
Aspartic acid	ASP	C4 H7 N1 O4	133.10
ASP/ASN ambiguous	ASX	C4 H7 1/2 N1 1/2 O3 1/2	132.61
Cysteine	CYS	C3 H7 N1 O2 S1	121.15
Glutamine	GLN	C5 H10 N2 O3	146.15
Glutamic acid	GLU	C5 H9 N1 O4	147.13
GLU/GLN ambiguous	GLX	C5 H9 1/2 N1 1/2 O3 1/2	146.64
Glycine	GLY	C2 H5 N1 O2	75.07
Histidine	HIS	C6 H9 N3 O2	155.16
Isoleucine	ILE	C6 H13 N1 O2	131.17
Leucine	LEU	C6 H13 N1 O2	131.17
Lysine	LYS	C6 H14 N2 O2	146.19
Methionine	MET	C5 H11 N1 O2 S1	149.21
Phenylalanine	PHE	C9 H11 N1 O2	165.19
Proline	PRO	C5 H9 N1 O2	115.13
Serine	SER	C3 H7 N1 O3	105.09
Threonine	THR	C4 H9 N1 O3	119.12
Tryptophan	TRP	C11 H12 N2 O2	204.23
Tyrosine	TYR	C9 H11 N1 O3	181.19
Valine	VAL	C5 H11 N1 O2	117.15
Undetermined	UNK	C5 H6 N1 O3	128.16

### Nucleotides

NAME	CODE	FORMULA	MOL. WT.
Adenosine	A	C10 H14 N5 O7 P1	347.22
Cytidine	C	C9 H14 N3 O8 P1	323.20
Guanosine	G	C10 H14 N5 O8 P1	363.22
Inosine	I	C10 H13 N4 O8 P1	348.21
Thymidine	T	C10 H15 N2 O8 P1	322.21
Uridine	U	C9 H13 N2 O9 P1	324.18



## Appendix 6: Field Formats

(This information is repeated from the Introduction.)

Each record type is presented in a table which contains the division of the records into fields by column number, defined data type, field name or a quoted string which must appear in the field, and field definition. Any column not specified must be left blank.

Each field contains an identified data type which can be validated by a program. These are:

DATA TYPE	DESCRIPTION
AChar	An alphabetic character (A-Z, a-z).
Atom	Atom name which follow the naming rules in Appendix 3.
Character	Any non-control character in the ASCII character set or a space.
Continuation	A two-character field that is either blank (for the first record of a set) or contains a two digit number right-justified and blank-filled which counts continuation records starting with 2. The continuation number must be followed by a blank.
Date	A 9 character string in the form dd-mmm-yy where DD is the day of the month, zero-filled on the left (e.g., 04); MMM is the common English 3-letter abbreviation of the month; and YY is a year in the 20th century. This must represent a valid date.
IDcode	A PDB identification code which consists of 4 characters, the first of which is a digit in the range 0 - 9; the remaining 3 are alpha-numeric, and letters are upper case only. Entries with a 0 as the first character do not contain coordinate data.
Integer	Right-justified blank-filled integer value.
Token	A sequence of non-space characters followed by a colon and a space.
List	A String that is composed of text separated with commas.
LString	A literal string of characters. All spacing is significant and must be preserved.
LString(n)	An LString with exactly n characters.
Real(n,m)	Real (floating point) number in the FORTRAN format Fn.m.
Record name	The name of the record: 6 characters, left-justified and blank-filled.
Residue name	One of the standard amino acid or nucleic acids, as listed below, or the non-standard group designation as defined in the HET dictionary. Field is right-justified.
SList	A String that is composed of text separated with semi-colons.

Specification	A String composed of a token and its associated value separated by a colon.
Specification list	A sequence of Specifications, separated by semi-colons.
String	A sequence of characters. These characters may have arbitrary spacing, but should be interpreted as directed below.
String(n)	A String with exactly n characters.
SymOP	An integer field of from 4 to 6 digits, right-justified, of the form nnnMMM where nnn is the symmetry operator number and MMM is the translation vector. See details in Appendix 1.

To interpret a String, concatenate the contents of all continued fields together, collapse all sequences of multiple blanks to a single blank, and remove any leading and trailing blanks. This permits very long strings to be properly reconstructed.

---

## Appendix 7: Order of Records

(This information is repeated from the Introduction.)

All records in a PDB coordinate entry must appear in a defined order. Mandatory record types are present in all entries. When mandatory data are not provided, the record name must appear in the entry with a NULL indicator. Optional items become mandatory when certain conditions exist. Record order and existence are described in the following table:

RECORD TYPE	EXISTENCE	CONDITIONS IF OPTIONAL
HEADER	Mandatory	
OBSLTE	Optional	Mandatory in withdrawn entries.
TITLE	Mandatory	
CAVEAT	Optional	Mandatory if structure is deemed incorrect by an outside editorial board.
COMPND	Mandatory	
SOURCE	Mandatory	
KEYWDS	Mandatory	
EXPDTA	Mandatory	
AUTHOR	Mandatory	
REVDAT	Mandatory	
SPRSDE	Optional	Mandatory if a replacement entry.
JRNL	Optional	Mandatory if a publication describes the experiment.
REMARK 1	Optional	
REMARK 2	Mandatory	
REMARK 3	Mandatory	
REMARK N	Optional	
MODRES	Optional	Mandatory if modified group exists within the coordinates.
DBREF	Mandatory	Mandatory for each peptide chain with a length greater than ten (10) residues, and for nucleic acid entries that exist in the Nucleic Acid Database (NDB).
SEQADV	Optional	Mandatory if sequence conflict exists.
SEQRES	Optional	Mandatory if ATOM records exist.

HET	Optional	Mandatory if non-standard group other than water appears in the entry.
HETNAM	Optional	Mandatory if non-standard group other than water appears appears in the entry.
HETSYN	Optional	
FORMUL	Optional	Mandatory if non-standard group or water appears.
HELIX	Optional	
SHEET	Optional	
TURN	Optional	
SSBOND	Optional	Mandatory if disulfide bond is present.
LINK	Optional	
HYDBND	Optional	
SLTBRG	Optional	
CISPEP	Optional	
SITE	Optional	
CRYST1	Mandatory	
ORIGX1 ORIGX2 ORIGX3	Mandatory	
SCALE1 SCALE2 SCALE3	Mandatory	
MTRIX1 MTRIX2 MTRIX3	Optional	Mandatory if the complete asymmetric unit must be generated from the given coordinates using non-crystallographic symmetry.
TVECT	Optional	
MODEL	Optional	Mandatory if more than one model is present in the entry.
ATOM	Optional	Mandatory if standard residues exist.
SIGATM	Optional	
ANISOU	Optional	
SIGUIJ	Optional	
TER	Optional	Mandatory if ATOM records exist.
HETATM	Optional	Mandatory if non-standard group appears.
ENDMDL	Optional	Mandatory if MODEL appears.

CONECT	Optional	Mandatory if non-standard group appears.
MASTER	Mandatory	
END	Mandatory	

Note that a PDB file existing outside of the PDB official release may contain locally-defined records beginning with "USER". The PDB reserves the right to add new record types (not beginning with "USER"), so programs which read PDB entries should be prepared to read (and ignore) other record types. PDB will follow standard procedures whenever format changes are proposed.

### Sections of an Entry

The following table lists the various sections of a PDB coordinate entry and the records comprising them:

SECTION	DESCRIPTION	RECORD TYPE
Title	Summary descriptive remarks	HEADER, OBSLTE, TITLE, CAVEAT, COMPND, SOURCE, KEYWDS, EXPDTA, AUTHOR, REVDAT, SPRSDE, JRNL
Remark	Bibliography, refinement, annotations	REMARKS 1, 2, 3 and others
Primary structure	Peptide and/or nucleotide sequence and the relationship between the PDB sequence and that found in the sequence database(s)	MODRES, DBREF, SEQADV, SEQRES
Heterogen	Description of non-standard groups	HET, HETNAM, HETSYN, FORMUL
Secondary structure	Description of secondary structure	HELIX, SHEET, TURN
Connectivity annotation	Chemical connectivity	SSBOND, LINK, HYDBND, SLTBRG, CISPEP
Miscellaneous features	Features within the macromolecule	SITE
Crystallographic	Description of the crystallographic cell	CRYST1
Coordinate transformation	Coordinate transformation operators	ORIGXn, SCALEn, MTRIXn, TVECT
Coordinate	Atomic coordinate data	MODEL, ATOM, SIGATM, ANISOU, SIGUIJ, TER, HETATM, ENDMDL
Connectivity	Chemical connectivity	CONECT
Bookkeeping	Summary information, end-of-file marker	MASTER, END