Protein Data Bank Atomic Coordinate Entry Format Description: An Annotated Reference Manual

Version 2.23, January 2, 1996

Protein Data Bank Atomic Coordinate Entry Format Description: An Annotated Reference Manual

Version 2.23, January 2, 1996

```
Changes from V. 2.22:
    Working version for processors' final meetings on document.
    Added reference to NDB in DBREF, along with example of same.
    Removed references to Electronic Deposition Form.
    Removed references to FTNOTEs.
    Changed EXPDTA.
    Removed superflous html commands.
    Removed REMARK 3 for now


Changes from V. 2.21:
    Fixed reference to U(eq) to B(eq) in HETATM and ANISOU records.
    Made some spelling corrections.
```

//INTRODUCTION==========

Introduction - Purpose of this Document

This Format Description gives a complete and concise description of the contents of PDB coordinate entry files. This document will be helpful to several communities, assisting depositors to prepare 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 needed to facilitate future conversion of PDB into CIF.

Several new record types are being proposed, as well as some changes to existing records. The corresponding sections of the PDB Electronic Deposition Form are included for each record type and examples of the records are provided.

Basic Notions

Character Set

The characters permitted to appear in a PDB file are all non-control ASCII characters as well as the space and end-of-line indicator. 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).

The '=' character appears infrequently in PDB files. It has been used in the past to represent superscripts and subscripts. Any text surrounded by '==' is assumed to be superscripted. Similarly, any text surrounded by '=' is assumed to be subscripted. If '=' is surrounded by at least one space on each side, then it is assumed to be an equal sign. No other typesetting codes are used in PDB entries.

Commas and semi-colons are used as list delimiters in certain record types. All appearances of these characters in other context within those fields must be preceded with a "\".

Record Format

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

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

Each record type will be described in detail in this document.

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 five categories based upon how often the record type will appear in an entry.

Single

There are six 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
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.

Continued

There are nine 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	Enumeration of molecular contents of the entry
EXPDTA	Experimental technique of structure determination
KEYWDS	List of keywords describing the macromolecule
OBSLTE	Statement that the entry has been replaced and list of the ID code(s) which replaced it
SOURCE	Biological source of macromolecules found in the entry
SPRSDE	List of entries withdrawn from release and replaced by current entry
TITLE	Description of the experiment represented in the entry

Columns 7 and 8 of each of these records are left blank. The second and subsequent lines contain a right-justified integer in columns 9 and 10 and a blank in column 11. The integer increments by one for each additional line of the record.

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
CONECT	Connectivity records
DBREF	Reference to the sequence database(s)
FORMUL	Chemical formula of non-standard groups
HELIX	Identification of helical substructures
HET	Identification of non-standard groups or residues (heterogens)
HETATM	Atomic coordinate records for heterogens
HETNAM	Compound name of the 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
REVDAT	Revision date and related information
SEQADV	Identification of conflicts between PDB and the 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

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. The beginning of a chain is implicit with the first ATOM/HETATM record that appears in a coordinate entry or model, or with a change of chain identifier.

Other

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

Record Type	Description
JRNL	Literature citation that defines the coordinate set
MTRIXn	Transformations expressing non-crystallographic symmetry (n = 1, 2, or 3). There may be multiple sets of these records
REMARK	General remarks, some structured and some free form

Order of Records

All records in a PDB coordinate entry must appear in a defined order. Mandatory records are present in all entries. When mandatory data are not provided (usually occurring in older entries), the record name must appear in the entry with a NULL indicator. Optional items are required 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	Mandatory	
REMARK 2	Mandatory	
REMARK 3	Mandatory	
REMARK N	Optional	
MODRES	Optional	Mandatory if modified group exists
DBREF	Mandatory	Mandatory for each peptide chain with a l 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	Mandatory	
HET	Optional	Mandatory if non-standard group appears
HETNAM	Optional	Mandatory if non-standard group appears
HETSYN	Optional	Mandatory if non-standard group appears
FORMUL	Optional	Mandatory if non-standard group or water appears
HELIX	Optional	Mandatory if applicable
SHEET	Optional	Mandatory if applicable
TURN	Optional	Mandatory if applicable
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	Mandatory for infinite covalently-connected structures
MODEL	Optional	Mandatory if more than one structure is present in the entry
ATOM	Mandatory	
SIGATM	Optional	
ANISOU	Optional	
SIGUIJ	Optional	
TER	Mandatory	
HETATM	Optional	Mandatory if non-standard group or water appears
ENDMDL	Optional	Mandatory if MODEL appears
CONECT	Optional	Mandatory if non-standard group appears
MASTER	Mandatory	
END	Mandatory	

* ATOM records for proteins are listed from amino to carboxyl terminus.

* Nucleic acid residues are listed from the 5' to the 3' end.

* No ordering is specified for polysaccharides.

* Chains are residue sequences terminated by a TER record.

* If more than one structure is present in the entry, each model is delimited by MODEL and ENDMDL records.

A BNF (Backus-Naur Form) for the order of records appears in Appendix 1. It gives the definitive rule to follow on record ordering.

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. See the statement on PDB's WWW server for the procedure.

PDB files have some underlying structure, as indicated in Appendix 1. 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 and annotations including experimental details	REMARK
Primary structure	Peptide and/or nucleotide sequence and the relationship between the PDB sequence and that found in the sequence database	MODRES, DBREF, SEQADV, SEQRES
Heterogen	Description of non-standard groups	HET, HETNAM, HETSYN, FORMUL
Secondary structure	Description of secondary structure	HELIX, SHEET, TURN
Local	Features within the macromolecule	CISPEP, SITE
Crystal data	Description of the crystallographic cell	CRYST1
Coordinate	Atomic coordinate data	MODEL, ATOM, SIGATM, ANISOU, SIGUIJ, TER, HETATM, ENDMDL
Coordinate transformation	Coordinate transformation operators	ORIGXn, SCALEn, MTRIXn, TVECT
Connectivity	Chemical connectivity	SSBOND, LINK, HYDBND, SLTBRG, CONECT
Bookkeeping	Summary information, end-of-file marker	MASTER, END

Field Formats

Each record type will be detailed in this document. The description of each record type will include the following sections:

* Overview

* Record Format

* Details

* Verification/Validation/Value Control Authority

* Relationship to Other Record Types

* Corresponding section of the Electronic Deposition Form

* Examples

* Known Problems

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
Atom	Atom name which follow the naming rules in Appendix 2
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 from 2
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
Element	Symbol for an element, including the charge
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
Keyword	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 Keyword: and its associated value
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

Standard residues 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, +A, +C, +G, +T, +U
Other	UNK (unknown)

See Appendix 3 for more information on the standard residue names and abbreviations, and Appendix 4 for their chemical formulas and molecular weights.

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 represented.

//TITLE SECTION==

Title Section

This section contains the records used to describe the entry: HEADER, OBSLTE, TITLE, CAVEAT, COMPND, SOURCE, KEYWDS, EXPDTA, AUTHOR, REVDAT, SPRSDE, and JRNL.

HEADER==========

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	Classification field of the molecule.

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 and is assigned randomly to an entry

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. Classification may present a complex of biological macromolecules or a dual function. Due to the length of the classification field, some of the strings must be abbreviated terms. Full terms are given in KEYWDS.

Verification/Validation/Value Control Authority

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.

A mechanism for maintaining the class list file has not yet been defined. At the present time, it is maintained by PDB staff and will be extended as appropriate.

Relationships to Other Record Types

HEADER records contain a classification field which must also appear as the first KEYWDS term. The KEYWDS record provides terms that may be used to classify the entry. Terms abbreviated in the HEADER record are not abbreviated in KEYWDS.

Deposition Form Section and Prompt

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
```

Known Problems

Forty characters for the classification are not sufficient in a number of cases. PDB has in the past eliminated most spacing and used abbreviations. The new record type KEYWDS will be used to list functional or structural classification terms. The classification in HEADER will be repeated in KEYWDS, unabbreviated.

//OBSLTE==========

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 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 Control Authority

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

Relationships to Other Record Types

None.

Deposition Form

Not applicable.

Example

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
OBSLTE     31-JAN-94 1MBP       2MBP

//TITLE============================================================
```

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. TITLE may include mutations. Entries consisting of only alpha carbon atoms should state this fact in the title. Experiment type may also appear here.

Verification/Validation/Value Control Authority

This record is free text so no verification of format is required. The title is supplied by the depositor, but PDB staff may edit the title for various reasons.

Relationships to Other Record Types

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

Deposition Form Section and Prompt

Example

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
```

```
TITLE     RHIZOPUSPEPSIN WITH REDUCED PEPTIDE INHIBITOR

TITLE     CYCLOMALTODEXTRIN GLYCOSYLTRANSFERASE MUTANT (Y195W)

TITLE     ALPHA CARBON COORDINATES FOR BETA-GLUCOSYLTRANSFERASE

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

//CAVEAT==========

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.

Verification/Validation/Value Control Authority

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

Relationships to Other Record Types

None.

Deposition Form Section and Prompt

Example

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
CAVEAT     1ABC    THE CRYSTAL TRANSFORMATION IS IN ERROR BUT IS
CAVEAT   2 1ABC    UNCORRECTABLE AT THIS TIME
```

//COMPND==========

COMPND

Overview

The COMPND record describes the molecular contents of an entry. Each macromolecule found in the entry is described by a set of keyword: value pairs, and is referred to as a COMPND record component. Heterogen information will appear in the TITLE, HET, HETNAM, HETSYN, FORMUL, and REMARK records instead of in COMPND as has been done in the past.

For each component, the molecule name, synonyms, Enzyme Commission (EC) number, 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 components

Details

The compound record is a Specification list, one set of specifications per component. The Keywords that may be used are listed below. The Keyword ENGINEERED may be followed by an optional comment.

KEYWORD	DEFINITION
MOL_ID	Numbers each molecule; also appears in SOURCE
MOLECULE	Name of the macromolecule
CHAIN	Chain identifier(s) of the given molecule. If the molecule is a subunit of a single chain, then the chain and the residues which comprise this molecule may be listed (e.g., CHAIN: A, B or CHAIN: A (47 - 221)). NULL is used if no chain identifier.
DOMAIN	Specifies a domain or region of the molecule
SYNONYM	Comma-separated list of synonyms for the MOLECULE. No more than two synonyms are suggested
EC	The Enzyme Commission number associated with the molecule
ENGINEERED	Indicates that the molecule was produced using recombinant technology or by purely chemical synthetic methods
MUTATION	Describes the mutations present
BIOLOGICAL_UNIT	If the MOLECULE functions as part of a larger biological unit, the entire functional unit may be described
OTHER_DETAILS	Additional comments

No specific rules apply to the ordering of the Keywords, except that the occurrence of MOL_ID and MOLECULE indicates that the following Keywords are related to that specific molecule. Physical layout of these items may be altered by PDB staff to improve human readability of the COMPND record.

ENGINEERED is followed either by "YES" or by a comment.

Asterisks in nucleic acid names are for ease of reading.

Verification/Validation/Value Control Authority

CHAIN must match the chain identification(s) of the molecules. As heterogens are being adequately described elsewhere in the entry (they must appear in the HET and FORMUL records), PDB proposes to remove heterogens from the COMPND record. EC numbers are checked against the Enzyme Data Bank.

Relationships to Other Record Types

See verification above.

Deposition Form Section and Prompt

Example

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
COMPND    MOL_ID: 1;
COMPND   2 MOLECULE: ALKALINE PHOSPHATASE;
COMPND   3 CHAIN: A, B;
COMPND   4 EC: 3.1.3.1;
COMPND   5 ENGINEERED: YES;
COMPND   6 MUTATION: D153G;
COMPND   7 OTHER_DETAILS: MUTANT WITH 4-FOLD INCREASED ACTIVITY AND
COMPND   8 WEAKER MG BINDING

         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
COMPND    MOL_ID: 1;
COMPND   2 MOLECULE: CYTOCHROME C PEROXIDASE;
COMPND   3 CHAIN: NULL;
COMPND   4 SYNONYM: CCP-MKT;
COMPND   5 EC: 1.11.1.5;
COMPND   6 ENGINEERED: YES;
COMPND   7 MUTATION: HIS 175 REPLACED BY GLY

         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
COMPND    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
```

//SOURCE==========

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 scientific names, genus and species. Strain and/or cell-line for immortalized cells are given when they help in uniquely identifying 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 Keyword: value format

Details

* 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 and MOLECULE indicate subsequent specifications are related to that 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 Keyword ENGINEERED.

* When multiple components 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 that molecule.

* Hybrid molecules prepared by fusion of genes are treated as multi-component systems for the purpose of specifying the source.

* 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.

KEYWORD	DEFINITION
MOL_ID	Numbers each molecule. Same as appears in COMPND
SYNTHETIC	Indicates a chemically-synthesized source
DOMAIN	A domain or region 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
ORGANELLE	Organized structure within a cell
CELL	Identifies the particular cell type
CELLULAR_LOCATION	Identifies the location within (or without) the cell
PLASMID	Identifies the plasmid containing the gene
GENE	Identifies the gene
EXPRESSION_SYSTEM	System used to produce recombinant macromolecules
EXPRESSION_SYSTEM_PLASMID	Plasmid used in the recombinant experiment
EXPRESSION_SYSTEM_GENE	Name of the gene used in recombinant experiment
EXPRESSION_SYSTEM_XXX	Strain, organ, tissue, cell, etc. of the expression system will be given in this manner
OTHER_DETAILS	Additional comments

Verification/Validation/Value Control Authority

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. This list is available from the PDB.

Relationships to Other Record Types

Each macromolecule listed in COMPND must have a corresponding source.

Deposition Form Section and Prompt Example

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
SOURCE     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     MOL_ID: 1;
SOURCE    2 ORGANISM_SCIENTIFIC: GALLUS GALLUS;
SOURCE    3 ORGANISM_COMMON: CHICKEN;
SOURCE    4 ORGAN: HEART;
SOURCE    5 TISSUE: MUSCLE

SOURCE     MOL_ID: 1;
SOURCE    2  DOMAIN: RESIDUES 1-16;
SOURCE    3  ORGANISM_SCIENTIFIC: BACILLUS AMYLOLIQUEFACIENS;
SOURCE    4  EXPRESSION_SYSTEM: ESCHERICHIA COLI;
SOURCE    5 MOL_ID: 2;
SOURCE    6  DOMAIN: RESIDUES 17-214;
SOURCE    7  ORGANISM_SCIENTIFIC: BACILLUS MACERANS;
SOURCE    8  EXPRESSION_SYSTEM: ESCHERICHIA COLI
SOURCE    9  EXPRESSION_SYSTEM_STRAIN: BE167
```

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	List of key words 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 Control Authority

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 key words 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.

Deposition Form Section and Prompt

Example

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
KEYWDS    LYASE, TRICARBOXYLIC ACID CYCLE, MITOCHONDRION, OXIDATIVE
KEYWDS   2 METABOLISM
```

//EXPDTA==========

EXPDTA

Overview

EXPDTA, or experimental data, is mandatory and appears in all entries.

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 records if necessary
11 - 70	Slist	technique	The experimental technique(s) with optional comment describing the sample or experiment

Details

The technique must match one of the allowed strings. If an NMR experiment, the number of models included in the entry must be given. 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 were used for the structure determination and are being presented in the entry, EXPDTA will present them as a semi-colon separated list.

Verification/Validation/Value Control Authority

The verification program checks that the EXPDTA record appears in the entry and that the technique string matches one of the allowed strings. It also checks that the standard REMARK is added in the case of NMR, fiber, or modeling studies, and that the correct CRYST1 and SCALE are used in these cases. If an NMR entry contains multiple conformations, 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 modeling study, this should be stated in the TITLE, with the appropriate EXPDTA and REMARK records. Specific details of the data collection and experiment will appear in REMARK 18.

Deposition Form Section and Prompt

Example

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
EXPDTA    X-RAY DIFFRACTION

EXPDTA    NEUTRON DIFFRACTION; X-RAY DIFFRACTION

EXPDTA    NMR, MINIMIZED AVERAGE STRUCTURE

EXPDTA    NMR, 32 STRUCTURES, SOLID STATE

EXPDTA    THEORETICAL MODEL
```

//AUTHOR==========

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.

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 initials.

* Blank spaces are used in a name only if properly part of the surname, 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.

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 will occur only after a comma.

* Personal names will not be 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 Control Authority

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.

Deposition Form Section and Prompt

Example

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
AUTHOR     M.B.BERRY,B.MEADOR,T.BILDERBACK,P.LIANG,M.GLASER,
AUTHOR    2 G.N.PHILLIPS JUNIOR

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

//REVDAT============================================================
```

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 inclusion 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 Control Authority

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

See verification above.

Deposition Form Section and Prompt

These records are inserted by the PDB.

Example

Entry: 1PRC

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
REVDAT   3   15-OCT-89 1PRCB    1       REMARK
REVDAT   2   19-APR-89 1PRCA    2       CONECT
REVDAT   1   09-JAN-89 1PRC     0
```

//SPRSDE==========

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	sprsdeDate	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 sIDCode list is terminated by the first blank IDcode field.

Verification/Validation/Value Control Authority

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 idCode must be the same as the ID code found in the HEADER of this entry.

Deposition Form Section and Prompt

Example

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
SPRSDE     17-JUL-84 4HHB       1HHB

SPRSDE     18-JAN-83 3PTB       2PTB
```

//JRNL==========

JRNL

Overview

The JRNL record contains the literature citation that describes the experiment which resulted in the deposited coordinate set. There is only one JRNL record per entry. If there is no such publication, then there will be no JRNL record. Other references are given REMARK 1.

PDB is in the process of linking and/or adding all references to CitDB (the literature database that serves the Human Genome Project). This will increase the accuracy and consistency of references.

Record Format

Columns	Data Type	Field	Definition
1-6	Record name	"JRNL"	
11-70	LString		see Details below

Details

The JRNL record is fairly complex and has a detailed inner structure requiring many sub-records. The following tables are used to describe these sub-record types.

The AUTH sub-record is always present in JRNL. This is followed by TITL, EDIT, REF, PUBL, and REFN sub-record types. EDIT and PUBL appear only if the reference is to a book.

1. AUTH

The first sub-record type, AUTH, contains the list of authors. 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 names of the authors cited are those associated with the article or other contribution to a larger work (i.e., AUTH is not used for the editor of a book).

There is no leading blank in the authorList field of continuation sub-records. One 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

The second sub-record type, TITL, specifies the title of the reference. Use this 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.

There is no leading blank in the title field of continuation sub-records.

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.

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

The third sub-record type, 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

The next sub-record type, 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 it 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.

Publication name (first item in pubName field) is listed according to one of the following:

* 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 will be within parentheses and the series name will be 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 will be in parentheses, and the name will be preceded by "REPORT:".

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 filed 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 pubNarne 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. (Note: When a series name is listed in addition to a book name, the volume on the first REF line represents the volume number of the item whose title begins on that first REF line, i.e., the volume number of the book if there is one; the volume number of the series is listed within the parentheses that have the name of the series. This allows for multi-volume books within numbered series.)

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 filled in
52 - 55	Integer	volume	Right-justified blank-filled integer ; appears in the first sub-record only
57 - 61	Integer	page	First page of the article;first sub-record only
63 - 66	Integer	year	Year of publication; first sub-record only

5. PUBL

The next sub-record type, 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. 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.

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 is followed by a space and "(THESIS)".

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	Name of the publisher and the city of publication

6. REFN

The final sub-record type, REFN, is a group of fields which contain encoded references to the citation. No continuation lines are possible or necessary. 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 should be 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.

6a. There are two forms of this sub-record type group, depending upon the publication status. 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	astm	American Society for Testing and Materials devised CODEN assigned by Chemical Abstracts Service
33 - 34	LString	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 Cambridge Crystallographic
                                        Data Centre CODEN list or assigned by
                                        PDB
```

6b. 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 PDB CODEN for unpublished works
```

Verification/Validation/Value Control Authority

PDB verifies that this record is correctly formatted.

PDB will use 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 will require following the National Library of Medicine rules for the transcription of names and titles. Articles in non-MEDLINE journals will be 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.

Relationship to Other Records:

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

Deposition Form Section and Prompt

Example

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
JRNL        AUTH   N.THANKI,J.K.M.RAO,S.I.FOUNDLING,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 JMOBAK  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, theses, series, and reports. Some of the points below provide details about the variations and/or lumping 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 on the REF lines in columns 20-47. This sometimes means that the title coden has several matches for REF title information.

* Many published books are in series which have their own titles. Until now, only the series title has been presented in PDB entries.

* 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 REF's 28-column limit for title information.

* When book series information is presented it is added to REF lines. This requires a separate coden for each book in order to avoid several matches for a single coden for the series. Putting the book title on the TITL line in these cases would mean that some books were listed in TITL and others in REF. This would require more complicated rules that would in turn require a higher level of human intervention, hampering database interoperability.

* 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 that are issued as a multi-volume set have a separate title and a separate matching ISBN for each volume in the set, as well as an additional ISBN for the whole set of volumes. In the past, references had one coden for the set only; sub-titles and ISBN numbers were altered as each volume in the set was cited. There is no place for both applicable ISBN numbers to be listed.

* Many books are issued in more than one country. The publisher will have 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.

* If a report is cited in the entry, this information is usually tacked onto the PUBL line in parentheses rather than with other title information on the TITL or REF line. This is similar to the series title problem listed above.

* Pagination is limited to the beginning page.

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

The PDB proposes an expansion of the reference records JRNL and REMARK 1 in order to achieve the following goals:

* To achieve clarity

* To allow for varied types of publications which have not been well-handled in the past

* To include additional information needed for proper representation of all types of publications

* To facilitate interchange among databases

PDB will start contributing to and using CitDB to maintain all its references within the next six months. Since the reference information is more complete in CitDB than in PDB, the additional publication information would readily feed into the expanded reference records being proposed. In fact, the CitDB fields served as a model for creating the proposed expanded references for PDB.

//REMARK==========

REMARK

Overview

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

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

Record Format

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"REMARK"	
8 - 10	Integer	remarkNum	Remark number. Remarks 1, 2, and 3 are standardized. It is not an error for remark n to exist if remark n-1 does not
12 - 70	LString	text	Body of the remark

Details

The REMARK records provide descriptive information not given elsewhere about the experiment and the PDB entry.

Extracting the contents of REMARK records is more difficult than for other records. REMARK records often have embedded formatting (e.g., tables) or keyword: value pairs. Therefore, many remarks have their own rules for concatenation.

All REMARKs start with a blank REMARK N line. REMARK 1, while mandatory within coordinate entries, may be empty and therefore have no other REMARK 1 lines after it.

REMARK 1

REMARK 1 lists important publications related to the structure depicted in the entry. These citations are chosen by the depositor. They are listed in reverse-chronological order. However, if a particular article (or series of articles) is considered to be a definitive description, it may appear first regardless of date. Citations are not repeated from the JRNL records (this is not automatically checked). The format for REMARK 1 (after the first blank record and the REFERENCE sub-record) is completely analogous to the JRNL records. See the JRNL section for details.

PDB is in the process of linking and/or adding all references to CitDB (the literature database that serves the Human Genome Project). This will increase the accuracy and consistency of references.

The REMARK 1 records are fairly complex and have a detailed inner structure. Therefore, the following tables are used to describe these sub-record types. For details, see the JRNL section.

1. A blank REMARK 1 line is always present in a PDB entry, even when there are no references in REMARK 1.

Columns	Data Type	Field	Definition
1 - 6	Record name	"REMARK"	
10	LString(1)	"1"	

2. Each reference is preceded by a line indicating the reference number in the entry. This is the first sub-record type for each reference.

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

3. The second sub-record type, AUTH, contains the list of authors.

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

4. The third sub-record type, 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

5. The fourth sub-record type, 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

6. The next sub-record type,REF, is a group of fields which contains the name of the publication.

6a. 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.

6b. 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	Integer	volume	Right-justified blank-filled integer, appears in the first record only
57 - 61	Integer	page	First page of the article, first record only
63 - 66	Integer	year	Year of publication, first record only

7. The next sub-record type, 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

8. The final sub-record type is a group of fields which contains encoded references to the citation.

8a. 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	Code from Cambridge Crystallographic Data Center CODEN list, or assigned by the PDB

8b. 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

Verification/Validation/Value Control Authority

PDB verifies that this record is correctly formatted.

PDB will use 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 will require following the National Library of Medicine rules for the transcription of names and titles. Articles in non-MEDLINE journals will be 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 also appear in JRNL.

Relationship to Other Records:

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

Deposition Form Section and Prompt

Example

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
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 PSFGEY  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                                               2011

         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
REMARK   1
REMARK   1 REFERENCE 1
REMARK   1  AUTH   G.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 resolution, in Angstroms, of the experiment. As for all the remarks, the first REMARK 2 record is blank. The second REMARK 2 record has one of the following 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.

REMARK 2 when a diffraction experiment:

Columns	Data Type	Field	Definition
1 - 6	Record name	"REMARK"	
10	LString(1)	"2"	
12 - 23	LString(11)	"RESOLUTION."	
24 - 27	Real(4.2)	resolution	Resolution
28 - 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 - 39	LString(28)	"RESOLUTION. NOT APPLICABLE."	
39 - 70	String	comment	Comment

A few PDB entries have REMARK 2 records which are non-parsable. If the format does not match the first two, then the resolution is considered to be unknown. 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 - 23	LString(11)	"RESOLUTION."	
24 - 70	String	comment	Comment

Deposition Form Section and Prompt

Example

```
         1         2         3         4         5         6         7
12345678901234567890123456789012345678901234567890123456789012345678901234567890
REMARK   2
REMARK   2 RESOLUTION. 1.74 ANGSTROMS.

REMARK   2
REMARK   2 RESOLUTION.  NOT APPLICABLE.
```

REMARK 3

REMARK 3 --- text omitted for now --

Other Remarks

The remaining remarks fall into one of two categories. They are either pure text or template style. For all the remarks, the first record is blank. The following paragraphs identify the templates currently in use. Note that N's and X'x are used to represent variables. See the file called **format/remark.txt** which has

the text of each of these canned remarks.

PDB is moving toward assigning standard remark numbers and text where possible . See Appendix 5 for information on standard remarks under consideration.

Remark 16

Remark 16 will replace one of the most often-used footnotes, that which names peptide bonds that deviate from trans conformation. This is generated by the processing program.

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
REMARK  16
REMARK  16     XXX X   NNN  - XXX X   NNN               OMEGA = NNN.NN
REMARK  16 PEPTIDE BOND DEVIATES SIGNIFICANTLY FROM TRANS CONFORMATION
```

Remark 17

Remark 17 will present the details of the NMR experiment. This is being reserved for use.

Remark 18

Remark 18 presents the details of the diffraction experiment in the token: value format. If there is more than one experiment represented by the coordinate entry, the section is repeated using N replaced by 1, 2, etc. When only one experiment is represented in the entry, "EXPERIMENT N." is omitted. Tokens will not appear in an entry if the item is not supplied by the depositor.

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890

REMARK  18
REMARK  18 EXPERIMENTAL DETAILS (X-RAY).  EXPERIMENT N.
REMARK  18  DATE OF DATA COLLECTION        :
REMARK  18  MONOCHROMATIC/LAUE (M/L)       :
REMARK  18  WAVELENGTH OR RANGE (A)        :
REMARK  18  MONOCHROMATOR                  :
REMARK  18  OPTICS                         :
REMARK  18  DETECTOR TYPE                  :
REMARK  18  DETECTOR MANUFACTURER          :
REMARK  18  INTENSITY-INTEGRATION SOFTWARE :
REMARK  18
REMARK  18  RESOLUTION RANGE               :
REMARK  18  OBSERVATIONS COLLECTED         :
REMARK  18  UNIQUE DATA COLLECTED          :
REMARK  18  DATA REDUNDANCY                :
REMARK  18  MERGING R VALUE (INTENSITY)    :
REMARK  18  COMPLETENESS OF DATA           :
REMARK  18  REJECTION CRITERIA  SIGMA(I)   :
REMARK  18
REMARK  18  WILSON B VALUE                 :
```

Remark 19

Remark 19 presents some crystal information.

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
REMARK  19
REMARK  19 SOLVENT CONTENT (VS)       : XX.X %
REMARK  19 MATTHEWS COEFFICIENT (VM) : X.XX ANGSTROMS**3/DA
```

Remark 50

Remark specific to NMR entries.

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
REMARK  50 THESE COORDINATES WERE GENERATED FROM SOLUTION NMR DATA.
REMARK  50 PROTEIN DATA BANK CONVENTIONS REQUIRE THAT CRYST1 AND
REMARK  50 SCALE RECORDS BE INCLUDED, BUT THE VALUES OF THESE
REMARK  50 RECORDS ARE MEANINGLESS.
```

Remark 51

If the data in this entry were derived by NMR techniques and if an NMR restraint file has been deposited, then this REMARK appears and xxxx is replaced by the ID code of this entry.

```
REMARK  51 THE LIST OF EXPERIMENTAL RESTRAINTS IS AVAILABLE AS PDB
REMARK  51 ENTRY RXXXXMR.
```

Remark 110

```
REMARK 110
REMARK 110 THE PROTEIN DATA BANK HAS ADOPTED THE SACCHARIDE CHEMISTS
REMARK 110 NOMENCLATURE FOR ATOMS OF THE DEOXYRIBOSE MOIETY RATHER
REMARK 110 THAN THAT OF THE NUCLEOSIDE CHEMISTS.  THE RING OXYGEN
REMARK 110 ATOM IS LABELLED O4* INSTEAD OF O1*.
```

Remark 111

```
REMARK 111
REMARK 111 THE HYDROGEN BONDS BETWEEN BASE PAIRS IN THIS ENTRY FOLLOW
REMARK 111 THE CONVENTIONAL WATSON-CRICK HYDROGEN BONDING PATTERN.
REMARK 111 THEY HAVE NOT BEEN PRESENTED ON *CONECT* RECORDS IN THIS
REMARK 111 ENTRY.
```

Remark 112

```
REMARK 112
REMARK 112 THE TRANSFORMATION PRESENTED ON MTRIX RECORDS BELOW WILL
REMARK 112 YIELD APPROXIMATE COORDINATES FOR CHAIN *X* WHEN APPLIED TO
REMARK 112 CHAIN *Y*.
```

Remark 113

```
REMARK 113
REMARK 113 ATOMS WITH THERMAL FACTORS WHICH CALCULATE LESS THAN 2.00
REMARK 113 ARE ASSIGNED THIS VALUE.  THIS IS THE LOWEST VALUE ALLOWED
```

```
REMARK 113 BY THE REFINEMENT PROGRAM.
```

Remark 114

```
REMARK 114
REMARK 114 HYDROGEN AND DEUTERIUM ATOMS IN THIS ENTRY HAVE BEEN
REMARK 114 ASSIGNED NAMES CONSISTENT WITH THE RECOMMENDATIONS OF THE
REMARK 114 IUPAC-IUB COMMISSION ON BIOCHEMICAL NOMENCLATURE (E.G.,
REMARK 114 J.MOL.BIOL. (1970) VOL.52, PP 1-17).  THE PROTEIN DATA BANK
REMARK 114 HAS FOLLOWED RULE 4.4 OF THE RECOMMENDATIONS WITH THE
REMARK 114 FOLLOWING MODIFICATION - WHEN MORE THAN ONE HYDROGEN ATOM
REMARK 114 IS BONDED TO A SINGLE NON-HYDROGEN ATOM, THE HYDROGEN ATOM
REMARK 114 NUMBER DESIGNATION IS GIVEN AS THE FIRST CHARACTER OF THE
REMARK 114 ATOM NAME RATHER THAN AS THE LAST CHARACTER (E.G., H BETA 1
REMARK 114 IS DENOTED AS 1HB).
```

Remark 115

```
REMARK 115
REMARK 115 THE SHEET STRUCTURE OF THIS MOLECULE IS BIFURCATED.  IN
REMARK 115 ORDER TO REPRESENT THIS FEATURE IN THE SHEET RECORDS BELOW,
REMARK 115 TWO SHEETS ARE DEFINED.  STRANDS N1, N2, N3 AND N4 OF XX1 AND
REMARK 115 XX2 ARE IDENTICAL.
```

Remark 116

```
REMARK 116
REMARK 116 THERE ARE SEVERAL BIFURCATED SHEETS IN THIS STRUCTURE.
REMARK 116 THESE ARE REPRESENTED BY TWO SHEETS WHICH HAVE ONE OR MORE
REMARK 116 IDENTICAL STRANDS.  SHEETS XXX AND XXX REPRESENT ONE
REMARK 116 BIFURCATED SHEET.  SHEETS XXX AND XXX REPRESENT ONE
REMARK 116 BIFURCATED SHEET.  SHEETS XXX AND XXX REPRESENT ONE
REMARK 116 BIFURCATED SHEET.
```

Remark 117

```
REMARK 117
REMARK 117 THE SHEET PRESENTED AS XXX ON SHEET RECORDS BELOW IS
REMARK 117 ACTUALLY AN N-STRANDED BETA-BARREL.  THIS IS
REMARK 117 REPRESENTED BY A N+1-STRANDED SHEET IN WHICH THE FIRST AND
REMARK 117 LAST STRANDS ARE IDENTICAL.
```

Remark 118

```
REMARK 118
REMARK 118 STRUCTURE FACTORS CORRESPONDING TO THIS ENTRY ARE AVAILABLE
REMARK 118 FROM THE PROTEIN DATA BANK AS A SEPARATE ENTRY.
```

Remark 999

Remark 999 is used for remarks related to the sequence. The text is varied and free-formatted.

```
REMARK 999
REMARK 999 THE SEQUENCE PRESENTED IN THIS ENTRY WAS DETERMINED IN THE
REMARK 999 AUTHOR'S OWN LAB AND HAS BEEN DEPOSITED IN SWISS-PROT.
```

All other REMARKs have to be viewed as unstructured at this time.

Deposition Form Section and Prompt

Example

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
REMARK  16
REMARK  16     GLU B   578  - PRO B   579               OMEGA = 145.65
REMARK  16 PEPTIDE BOND DEVIATES SIGNIFICANTLY FROM TRANS CONFORMATION

         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
REMARK  18
REMARK  18 EXPERIMENTAL DETAILS (X-RAY).  EXPERIMENT N.
REMARK  18  DATE OF DATA COLLECTION        :
REMARK  18  MONOCHROMATIC OR LAUE (M/L)    :
REMARK  18  WAVELENGTH OR RANGE (A)        :
REMARK  18  MONOCHROMATOR                  :
REMARK  18  OPTICS                         :
REMARK  18  SOURCE                         :
REMARK  18  BEAMLINE                       :
REMARK  18  DETECTOR TYPE                  :
REMARK  18  DETECTOR MANUFACTURER          :
REMARK  18  INTENSITY-INTEGRATION SOFTWARE :
REMARK  18
REMARK  18 REMARK:

         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
REMARK  19
REMARK  19 SOLVENT CONTENT (VS)      : XX.X %
REMARK  19 MATTHEWS COEFFICIENT (VM) : X.XX ANGSTROMS**3/DA
```

Known Problems

The refinement information presented in REMARK 3 and the format in which it is presented has evolved as PDB has responded to the changing needs of our user community. The design of the token: value layout for refinement information is meant to resolve the problem of poor parsability.

PDB acknowledges the drawbacks for computer parsing of the REMARK records which should for the most part be viewed as free text. Therefore we are considering adding new structure to these records. Appendix 5 presents one systematic approach to structuring the remarks. We would like comments on the format put forth in Appendix 5, and encourage other suggestions.

//Primary structure section==========

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==========

MODRES

Overview

The MODRES record provides descriptions of modifications (e.g., chemical or post-translational) to protein and nucleic residues. Included are a mapping between residue names given in a PDB entry and the standard 20 protein and 5 nucleic acid 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	Character	insertCode	Insertion code
25 - 27	Residue name	stdRes	Standard residue name
30 - 70	String	comment	Description of the residue modification. A list of values appears on the PDB FTP server as the file named /format/modres.lis

Details

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

Examples of modification descriptions:

Glycosylation site

Post-translational modification

Designed chemical modification

Phosphorylation site

Blocked N-terminus

Aminated C-terminus

D-configuration

Verification/Validation/Value Control Authority

MODRES will be generated by the PDB.

Relationships to Other Record Types

MODRES maps ATOM and HETATM records to the standard residue names as found in SEQRES.

Deposition Form Section and Prompt

Example

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
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
```

//DBREF==========

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, and for nucleic acid entries that exist in the Nucleic Acid Database (NDB).

Record Format

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 5	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	Character	insertBegin	Initial insertion code of the PDB sequence segment
21 - 24	Integer	seqEnd	Ending sequence number of the PDB sequence segment
25	Character	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 seqment
61	Character	insBegpdb	Insertion code of initial residue of the segment, if PDB is the referenc
63 - 67	Integer	seqEnd	Ending sequence number of the database segment
68	Character	insEndpdb	Insertion code of the ending residue of the segment, if PDB is the referenc

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 will be linked to different sequence database entries. In some cases, such as those PDB entries containing immunoglobulin Fab fragments, each chain will be linked to two different PIR, GenBank, and/or SWISS-PROT 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 and those describing sequence motifs (e.g., PROSITE, BLOCKS).

DBREF records have been developed to identify sequence correlations between PDB ATOM records and their corresponding PIR, GenBank, or SWISS-PROT entries. PDB entries containing chains for which residues are missing primarily due to disorder will 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 will be given as the reference.

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

Verification/Validation/Value Control Authority

This record will be generated by PDB. The sequence database entry found during our search will be compared to that provided by the depositor and any differences will be resolved.

Relationships to Other Record Types

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

Deposition Form Section and Prompt

Example

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
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  3HSF       1    92  SWS     P22121   HSF_KLULA        193    284

DBREF  223D A     1     6  NDB     123456                      1      6
DBREF  223D B     7    12  NDB     123456                      7     12
```

SEQADV==========

SEQADV

Overview

The SEQADV record identifies conflicts between sequence information in the PDB entry and the sequence database.

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	Character	insertCode	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, will now be reported in SEQADV.

SEQADV describes conflicts between residue sequences given by PDB ATOM 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 will be given. References will be given in the form YY-VOL-PAGE-CSDCODEN where YY is year of publication, VOL is the journal volume number, PAGE is the starting page and CSDCODEN is the 4-digit code assigned to journals by PDB and the Cambridge Structural Database (CSD). 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" will be 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.

Verification/Validation/Value Control Authority

SEQADV records will be automatically generated by the PDB.

Relationships to Other Record Types

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

Deposition Form Section and Prompt

Example

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
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=========================================================
```

SEQRES

Overview

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

Record Format

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"SEQRES"	
9 - 10	Integer	serialNum	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

* DetailsDB files 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 three-character HET name 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.

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

* If the residue sequence is unknown, the serialNum 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".

Verification/Validation/Value Control Authority

SEQRES is compared to the ATOM records during processing, and both are checked against the sequence database. All discrepancies will either be 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 states all discrepancies between the entry's sequence and that referenced in the database. MODRES describes modifications to the standard residue.

Deposition Form Section and Prompt

Known Problems

PDB plans to address representation of the branched polysaccharides in the near future.

//HET==========

HET

Overview

HET records are used to describe non-standard residues, such as prosthetic groups, inhibitors, solvent molecules, and ions. Basically, all groups are considered HET except for the 20 standard amino acids and the 5 nucleic acids C, G, A, T, and U.

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	Character	insertCode	Insertion code
21 - 25	Integer	numAtoms	Number of HETATM records for the group
28 - 70	String	Text	Text describing Het group

Details

* The HET group is assigned an ID, unique within the PDB, of not more than three characters (right-justified). The sequence number, chain identifier, insertion code, and number of atom 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.

* Rather than provide a separate HET record for each occurrence, if the same group occurs more than 15 times in an entry, a value of -999 is placed in the seqNum field.

* The Text field is for descriptive material. The keyword PART_OF followed by a value indicates that the HET group is part of a larger group which has been represented by its separate components (e.g., PART_OF: actinomycin).

* The PDB has prepared a heterogen group dictionary which may be found on the PDB FTP server as the file named /format/het.dict.

Verification/Validation/Value Control Authority

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.

No other compound will have the same hetID.

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.

Deposition Form Section and Prompt

Example

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
HET    TRS    975        8
HET    CH3    211        1  ATTACHED TO VAL 211
```

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 whole molecule. A few small groups are represented separately as ions, groups, and molecules.

Current PDB entries do not have HET records for water molecules; we would like to begin including them to be consistent with the format for heterogens.

PDB should include CAS numbers. Perhaps a new record could be introduced which will contain the CAS registry number and CSD entry number, if they exist. Please comment.

//HETNAM==========

HETNAM

Overview

This record gives the chemical name of the compound with the given hetID. Depositors will be consulted as to which name is to be used when there are several variation s.

Record Format

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

Details

The chemical name will be unique to the hetID. Other names for the group will be given in HETSYN records. PDB follows IUPAC/IUB naming conventions to describe groups systematically. Other acceptable names or names in use are presented as HETSYN or HETNAM. Continuation of chemical names onto subsequent records is allowed. Only one set of HETSYN records will be included for a given hetID, even if the same hetID appears on more than one HET record.

Verification/Validation/Value Control Authority

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

No other compound will have the same hetID.

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.

Deposition Form Section and Prompt

See the HET section of this document.

Example

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
HETNAM     GLC GLUCOSE
```

```
HETNAM     SAD BETA-METHYLENE SELENAZOLE-4-CARBOXAMIDE ADENINE
HETNAM   2 SAD DINUCLEOTIDE
```

//HETSYN==========

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"	
8 - 9	Continuation	continuation	
11 - 13	LString(3)	hetID	Het identifier, right-justified
15 - 70	Slist		List of synonyms

Details

This list is not guaranteed to be complete, but will be uniform across the PDB. 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 will be included for the hetID.

Verification/Validation/Value Control Authority

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

No other compound will have the same hetID.

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.

Deposition Form Section and Prompt

See the HET section of this document.

Example

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
HETSYN     NAD NICOTINAMIDE ADENINE DINUCLEOTIDE
```

```
HETSYN     COA COA

HETSYN     CMP CYCLIC AMP; CYCLIC ADENOSINE MONOPHOSPHATE

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

//FORMUL==========

FORMUL

Overview

The FORMUL record presents the chemical formula and charge of a non-standard group.

Record Format

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"FORMUL"	
9 - 10	Integer		Component number
13 - 15	LString(3)	hetID	Het identifier
18	Continuation	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 (including '1') of each atom type present immediately follows its chemical symbol.

* If the group is charged, the formula is followed by a blank and then the charge of the group, given as the numeral followed by the + or - characters, e.g., 2+, 1-.

* If the group appears more than once in the entry, the number of occurrences precedes the formula and charge which are then enclosed in parentheses.

* The component number is assigned sequentially. Components represented by a set of SEQRES records are counted first, and then each HETgroup is assigned a component number in sequence. If a HET group is contained within a chain represented by a set of SEQRES records, the component number assigned to the chain is used.

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

* The '*' in column 19 is used if the HET group is water, indicating that it should be excluded from the molecular weight calculation. This is maintained for historical reasons, but is being deprecated.

* A continuation field is provided in the event that more space is needed for the formula. Column 18 is

used in order to maintain continuity with the existing format.

Verification/Validation/Value Control Authority

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

No other compound will have the same hetID.

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.

Deposition Form Section and Prompt

See the HET section of this document.

Example

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
FORMUL   3  GLC    C6 H12 O6
FORMUL   2  SO4    2(S1 O4 2-)
```

//HELIX==========

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	Character	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	Character	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 closer to the amino-terminal, when tracing the chain, than the end residue of the helix.

* Helices are classified as follows:

1. Right-handed [[alpha]] (default)

2. Right-handed [[omega]]

3. Right-handed [[pi]]

4. Right-handed [[gamma]]

5. Right-handed 310

6. Left-handed [[alpha]]

7. Left-handed [[omega]]

8. Left-handed [[gamma]]

9. 27 ribbon/helix

10. Polyproline

Verification/Validation/Value Control Authority

HELIX records are now being generated automatically by PDB using DSSP if they are not provided by the depositor. PDB verifies that named residues exist in the ATOM records.

Relationships to Other Record Types

Deposition Form Section and Prompt

Example

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
HELIX    1  HA GLY A   86  GLY A   94  1
HELIX    2  HB GLY B   86  GLY B   94  1
```

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 following suggestions:

* Use of columns 72 - 76 to present the number of residues forming the helix.

* A new record to present the percent helix, sheet, turn, and coil in the structure.

* A new record which would present the various domain types found in the molecule, e.g., Residues 12 --> 120: [[alpha]]/[[beta]].

//SHEET==========

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	Character	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	Character	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	Character	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	Character	prevICode	Registration. Insertion code in previous strand

Details

The initial residue is closer to the amino-terminal, when tracing the chain, than the end residue of the strand. 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 Control Authority

SHEET records are now being generated automatically by PDB using DSSP if they are not provided by the depositor. 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 standard remarks 115, 116, and 117.

Deposition Form Section and Prompt

Example

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
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==========

TURN

Overview

The TURN records identify hairpin turns which do not occur in helices.

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	Character	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	Character	endICode	Insertion code of terminal residue of turn
41 - 70	String	comment	Associated comment

Details

* Turns are those non-helical sets of residues which form beta turns, i.e., have a hydrogen bond linking (C-O)i to (N-H)i+3. Turns which link residue i to i+2 (gamma-bends) may also be included.

* The initial residue is closer to the amino-terminal, when tracing the chain, than the end residue of the turn.

Verification/Validation/Value Control Authority

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

Deposition Form Section and Prompt

Example

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
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
```

//ANNOTATION SECTION==========

Annotation Section

The annotation section allows the depositors to specify the existence and location of disulfide bonds and other linkages, as well as active sites.

//SSBOND==========

SSBOND

Overview

The SSBOND record supplies the identification and location of disulfide bonds by identifying the two residues involved in the bond.

Record Format

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"SSBOND"	
8 - 10	Integer	seqNum	Sequence number
12 - 14	LString(3)	"CYS"	Residue name
16	Character	chainID1	Chain identifier
18 - 21	Integer	seqNum1	Residue sequence number
22	Character	icode1	Insertion code
26 - 28	LString(3)	"CYS"	Residue name
30	Character	chainID2	Chain identifier
32 - 35	Integer	seqNum2	Residue sequence number
36	Character	icode2	Insertion code
41 - 70	String	ssbond	Comment

Details

S-S bonds which are between molecules related by crystallographic symmetry will be listed in the REMARK section. Those appearing within the coordinates presented in the entry are placed in the SSBOND record. Bond distances between the sulfur atoms must be within expected values.

Verification/Validation/Value Control Authority

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

Relationships to Other Record Types

CONECT records are generated for the disulfide bonds.

Deposition Form Section and Prompt

Example

```
         1         2         3         4         5         6         7
```

```
12345678901234567890123456789012345678901234567890123456789012345678901234567890
SSBOND   1 CYS E   48     CYS E   51
SSBOND   2 CYS E  252     CYS E  285
```

//LINK==

LINK

Overview

The LINK records specify connectivity between residues not implied by the primary structure. Connectivity is expressed in terms of the atom names.

Record Format

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"LINK"	
13 - 16	Atom	name	Atom name
18 - 20	Residue name	resName	Residue name
22	Character	chainID	Chain identifier
23 - 26	Integer	resSeq	Residue sequence number
27	Character	iCODE	Insertion code
43 - 46	Atom	name	Atom name
48 - 50	Residue name	resName	Residue name
52	Character	chainID	Chain identifier
53 - 56	Integer	resSeq	Residue sequence number
57	Character	iCode	Insertion code
60	Integer	which	Indicates atom to which symmetry is applied
62 - 65	Character	symNum	Symmetry number. This is found in proposed REMARK 15

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 given.

* Each LINK record specifies one linkage.

* These records do not specify connectivity within a HET group (see CONECT) or disulfide bridges (see SSBOND).

* Hydrogen bonds and salt bridges are described on HYDBND and SLTBRG records, respectively.

* If the "which" field in column 60 equals:

Blank - the corresponding CONECT record appears in the entry

1 - apply symmetry to the first atom to form the bond

2 - apply symmetry to the second atom to form the bond

* The symNum in columns 62 - 65 refers to the symmetry number found in proposed REMARK 15.

Verification/Validation/Value Control Authority

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

Relationships to Other Record Types

CONECT records will be generated from LINKs.

Deposition Form Section and Prompt

Example

```
>
>          1         2         3         4         5         6         7
> 12345678901234567890123456789012345678901234567890123456789012345678901234567890
> LINK         O1  DDA     1                   C3  DDL     2       S1
> LINK        MN    MN   391                   OE2 GLU   217
>

>

//HYDBND=============================================================
```

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	name	Atom name
17	Character	altLoc	Alternate location indicator
18 - 20	Residue name	resName	Residue name
22	Character	Chain	Chain identifier
23 - 27	Integer	resSeq	Residue sequence number
28	Character	ICode	Insertion code
30 - 33	Atom	name	Hydrogen atom name
34	Character	altLoc	Alternate location indicator
36	Character	Chain	Chain identifier
37 - 41	Integer	resSeq	Residue sequence number
42	Character	iCode	Insertion code
44 - 47	Atom	name	Atom name
48	Character	altLoc	Alternate location indicator

```
49 - 51        Residue name    resName        Residue name
53             Character       chainID        Chain identifier
54 - 58        Integer         resSeq         Residue sequence number
59             Character       iCode          Insertion code
60             Integer         which          Indicates atom to which symmetry
                                              is applied
62 - 65        Character       symNum         Symmetry number.  This is found in
                                              proposed REMARK 15
```

Details

* The hydrogen bonds listed will be 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. The residue name is not given as it is already present on the record.

* For nucleic acids, Watson-Crick hydrogen bonds between bases may be listed, but this is optional.

* If "which" in column 60 equals:

Blank - the corresponding CONECT record appears in the entry

1 - apply symmetry to the first atom to form the bond

2 - apply symmetry to the second atom to form the bond

* The symNum in columns 62 - 65 refers to the symmetry number found in proposed REMARK 15.

Verification/Validation/Value Control Authority

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

Relationships to Other Record Types

CONECT records will be generated.

Deposition Form Section and Prompt

Example

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
HYDBND       N   LEU     10                 AO3* NDP    501
```

```
HYDBND       NH2 ARG    111                 OD2 ASP    149     S1
```

//SLTBRG==========

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	atom	First 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	Character	iCode	Insertion code
43 - 46	Atom	atom	Second atom name
47	Character	altLoc	Alternate location indicator
48 - 50	Residue name	resName	Residue name
52	Character	chainID	Chain identifier
53 - 56	Integer	resSeq	Residue sequence number
57	Character	iCode	Insertion code
60	Integer	which	Indicates atom to which symmetry is applied
62 - 65	Character	symNum	Symmetry number. This is found in proposed REMARK 15

Details

* Salt bridges listed will be 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.

* If "which" in column 60 equals:

Blank - the corresponding CONECT record appears in the entry

1 - apply symmetry to the first atom to form the bond

2 - apply symmetry to the second atom to form the bond

* The symNum in columns 62 - 65 refers to the symmetry number found in the proposed REMARK 15.

Verification/Validation/Value Control Authority

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

Relationships to Other Record Types

CONECT records will be generated.

Deposition Form Section and Prompt

Example

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
SLTBRG       O   GLU    10      NZ   LYS   115
SLTBRG       O   GLU    10      NZ   LYS   115                 S2
```

Known Problems

A specification as to which atom is more positively or negatively charged might be useful, but it would be difficult to verify.

//CISPEP==========

CISPEP

Overview

CISPEP records specify the prolines and other peptides found to deviate significantly from trans conformation. This record will replace the use of footnote to describe this conformation for both cis prolines and other cis peptides.

Record Format

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"CISPEP"	
8 - 10	Integer	seqNum	Record sequence number
12 - 14	LString(3)	pep1	Residue name
16	Character	chainID1	Chain identifier
18 - 21	Integer	seqNum1	Residue sequence number
22	Character	icode1	Insertion code
26 - 28	LString(3)	pep2	Residue name
30	Character	chainID2	Chain identifier
32 - 35	Integer	seqNum2	Residue sequence number
36	Character	icode2	Insertion code
44 - 48	LString(5)	"MODEL"	Used for entries with multiple models
50 - 52	Integer	modnum	Identifies the specific model
57 - 61	LString(5)	"OMEGA"	Name the angle
63	LString(1)	"="	Equal sign
65 - 70	Real(6.2)	measure	Measure in degrees of the angle

Details

Each cis peptide is listed on a separate line, with an incrementally ascending sequence number. Due to the occurrence of cis peptides other than prolines, it was decided to make this record reflect the general case. However, in the case of prolines, columns 26 through 70 will remain blank.

Verification/Validation/Value Control Authority

PDB generates these records automatically. If the depositor wishes to supply this information, PDB will compare these to the generated set of cis peptides and notify him/her of any differences.

Relationships to Other Record Types

CISPEP is replacing the footnote which previously contained this information.

Deposition Form Section and Prompt

Example

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
CISPEP   1 PRO A   48
CISPEP   2 PRO B   48

CISPEP   1 THR A   34    VAL A   35        MODEL   3    OMEGA = 143.95
```

//SITE==========

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	Character	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

Crystallographic Section

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

//CRYST1==========

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 will contain a=b=c=1.0, alpha=beta=gamma=90 degrees, space group P 1, and Z=1.

* 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 P2 that has 2 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

Verification/Validation/Value Control Authority

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 or THEORETICAL MODEL, the CRYST1 record is predefined as a=b=c=1.0, alpha=beta=gamma=90 degrees, 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 real CRYST1 record.

Deposition Form Section and Prompt

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


//ORIGX============================================================
```

ORIGXn

Overview

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

Record Format

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"ORIGXn"	
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:

Xsub = O11X + O12Y + O13Z + T1

Ysub = O21X + O22Y + O23Z + T2

Zsub = O31X + O32Y + O33Z + T3

The Electronic Deposition Form details the derivation of the ORIGX coordinate transformation in the Atomic Coordinate Preparation section.

Verification/Validation/Value Control Authority

If the coordinates are submitted in the same orthogonal Angstrom coordinate frame as they appear in the entry (the usual case), then ORIGX should be an identity matrix with translation vector equal to zero. If the transformation is not an identity matrix with 0 translation vector, then applying this transformation to the coordinates in the entry should yield the coordinates in the original deposited file.

Relationships to Other Record Types

If fractional coordinates were submitted, ORIGX will be equal to SCALE. ORIGX relates the coordinates in the ATOM and HETATM records to the coordinates in the submitted file.

Deposition Form Section and Prompt

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

//SCALE=========================================================
```

SCALEn

Overview

The SCALEn (n = 1, 2, or 3) records present the transformation from 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	"SCALE"	
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 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:

xfrac = S11X + S12Y + S13Z + U1

yfrac = S21X + S22Y + S23Z + U2

zfrac = S31X + S32Y + S33Z + U3

For NMR, fiber diffraction, and theoretical model entries, SCALE is given as an identity transformation with no translation.

The Electronic Deposition Form details the derivation of the SCALE coordinate transformation in the Atomic Coordinate Preparation section.

Verification/Validation/Value Control Authority

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

Relationships to Other Record Types

SCALE is related to the CRYST1 record, as the inverse of the determinant of the SCALE matrix equals the cell volume. If fractional coordinates were submitted, SCALE will be equal to ORIGX.

Deposition Form Section and Prompt

Example

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
SCALE1      0.019231  0.000000  0.000000        0.00000
SCALE2      0.000000  0.017065  0.000000        0.00000
SCALE3      0.000000  0.000000  0.016155        0.00000

//MTRIX=====================================================
```

MTRIXn

Overview

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

Record Format

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"MTRIXn"	
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 MTRIX transformations operate on the coordinates in the entry to yield equivalent representations of the molecule in the same coordinate frame. One trio of MTRIX 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 transformation in question are contained in the file, the iGiven field is set to 1. Otherwise, this field is blank.

Verification/Validation/Value Control Authority

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) and by checking the packing of the generated molecules.

Relationships to Other Record Types

If MTRIX records occur in a PDB coordinate entry, a corresponding REMARK must appear which describes the transformation.

Deposition Form Section and Prompt

Example

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
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==========

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 will contain a fragment which can be built into the full structure by the simple translation vectors of TVECT records.

Verification/Validation/Value Control Authority

None.

Relationships to Other Record Types

Any entry containing a TVECT record should have a corresponding REMARK describing the structure.

Deposition Form Section and Prompt

None at this time.

Example

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
TVECT    1   0.00000   0.00000  28.30000
```

/MODEL==========

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 - 5          Record name     "MODEL"
11 - 14        Integer         serial         Model serial number
```

Details

This record is used when more than one model, or conformation, 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, ANISOU, and TER records for each model structure will be interspersed as needed between MODEL and ENDMDL records.

Verification/Validation/Value Control Authority

Entries with NMR in the EXPDTA record are checked to verify that MODEL/ENDMDL pairs are present.

Relationships to Other Record Types

Each MODEL must have a corresponding ENDMDL record. TITLE and EXPDTA state that an entry represents an NMR experiment.

Deposition Form Section and Prompt

The MODEL/ENDMDL records appear within the coordinate list.

Example

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
MODEL        1
ENDMDL
MODEL        2
ENDMDL
```

//COORDINATE SECTION==========

Coordinate Section

The Coordinate Section contains the collection of atomic coordinates.

//ATOM==========

ATOM

Overview

The ATOM records present the atomic coordinates for standard residues. They also present the temperature factor and occupancy for each atom. Heterogen coordinates use the HETATM record type.

It is being proposed that segment identification, element symbol, and charge also be included on each ATOM record.

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	Character	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, IUPAC form

Details

* Protein residues are listed in order from the N to the C-terminal, nucleic acids from the 5'-to 3' end. Within each residue, the atoms are ordered as indicated in Appendix 3.

* See Appendix 2 for atom naming conventions.

* Alternate location is used to indicate that there is more than one position assigned for the given atom. May be denoted by A, B, C, etc.

* For residue names, see Appendix 3 and the HET section of this document.

* 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. Researchers may use a common 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 1A-8A and 9B-16B.

* The isotropic B value is given for the temperature factor.

* Standard PDB practice is to store the B equivalent in the tempFactor field when anisotropic temperature factors are provided in ANISOU records.

B(eq) = 8pi**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)] x 10**-4

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

* PDB has proposed the use of columns 73 - 76 to identify specific segments of the molecule. The segment can 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. Entries will begin appearing in this form early in 1996.

* Segment id is defined as a string of at maximum four (4) alphanumeric characters, left-justified, and can include a white space, e.g., CH86, A 1, NASE.

* PDB has also proposed the use of columns 77 - 78 to identify each atom's element type and associated charge. Computer programmers have requested this identification as it will allow unambiguous parsing. For example, in large het groups such as nicotinamide adenine dinucleotide, atoms are named with an A or N depending on which portion of the molecule they appear in, e.g., AC6 or NC6, AN1 or NN1. Now the element will be explicitly indicated in columns 77 - 78. Also, hydrogen naming sometimes conflicted with IUPAC conventions. For example, we have not been able to label a hydrogen HG11, but must label it 1HG1 in order for it not to be confused with mercury. After adopting the format change early in 1996, HG11 will be allowed in columns 13 - 16, as hydrogen will be clearly identified in columns 77 - 78. Columns 13 - 16 will be defined as presenting a unique name for each atom.

* Columns 77 - 78 will contain the element symbol, right justified.

* Columns 79 - 80 will indicate any charge on the atom, e.g., MN2+, O1-, H.

Verification/Validation/Value Control Authority

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 data in the near future. Details on this will be forthcoming.

Relationships to Other Record Types

The ATOM records are compared to SEQRES and 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, except for the Record name field.

Deposition Form Section and Prompt

The Atomic Coordinate Preparation section of the Electronic Deposition Form gives detailed instructions for preparing the coordinates for submission to the PDB.

Example

```
         1         2         3         4         5         6         7         8
12345678901234567890123456789012345678901234567890123456789012345678901234567890
ATOM      1  N   PRO A   1       7.091  29.139  59.073  1.00  0.00      A1   N
ATOM      2  CA  PRO A   1       8.484  28.695  59.356  1.00  0.00      A1   C
ATOM      3  C   PRO A   1       9.187  29.457  60.491  1.00  0.00      A1   C
ATOM      4  O   PRO A   1       8.849  30.652  60.747  1.00  0.00      A1   O
ATOM      5  CB  PRO A   1       9.262  29.160  58.195  1.00  0.00      A1   C
ATOM      6  CG  PRO A   1       8.316  29.998  57.361  1.00  0.00      A1   C
```

Known Problems

The changes being proposed to the ATOM and HETATM records address the need for explicit

identification of atom type and charge.

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 99,000 atoms.

Only 5 digits are available for the atom serial number, but some structures have already been received with more that 99,999 atoms. Changing the field length would make earlier entries incorrect.

//SIGATM==========

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	Character	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, IUPAC form

Details

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

Verification/Validation/Value Control Authority

None.

Relationships to Other Record Types

SIGATM is related to the preceding ATOM and HETATM coordinates.

Deposition Form Section and Prompt

Not in the deposition form. SIGATMs are found in the atom coordinate file.

Example

```
         1         2         3         4         5         6         7
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
SIGATM  237  HA  PRO     15       0.000   0.000   0.000  0.00  0.00
ATOM    238 1HB  PRO     15      22.240  28.540  10.860  1.00 17.70
SIGATM  238 1HB  PRO     15       0.000   0.000   0.000  0.00  0.00
ATOM    239 2HB  PRO     15      21.670  27.240  11.840  1.00 17.70
SIGATM  239 2HB  PRO     15       0.000   0.000   0.000  0.00  0.00
ATOM    240 1HG  PRO     15      20.360  29.240  10.740  1.00 23.90
SIGATM  240 1HG  PRO     15       0.000   0.000   0.000  0.00  0.00
ATOM    241 2HG  PRO     15      19.900  28.120  11.020  1.00 23.90
SIGATM  241 2HG  PRO     15       0.000   0.000   0.000  0.00  0.00
ATOM    242 1HD  PRO     15      19.230  30.160  12.320  1.00 15.50
SIGATM  242 1HD  PRO     15       0.000   0.000   0.000  0.00  0.00
ATOM    243 2HD  PRO     15      19.120  28.600  13.120  1.00 15.50
SIGATM  243 2HD  PRO     15       0.000   0.000   0.000  0.00  0.00
```

//ANISOU==========

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	Character	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)

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).

* The anisotropic temperature factors will be stored in the same coordinate frame as the atomic coordinate records.

Verification/Validation/Value Control Authority

None.

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.

Deposition Form Section and Prompt

Not in the deposition form. ANISOU records are found in the atom coordinate file.

Example

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

//SIGUIJ==========

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	Character	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)

Details

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

Verification/Validation/Value Control Authority

None.

Relationships to Other Record Types

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

Deposition Form Section and Prompt

Not in the deposition form. SIGUIJ records are included in the atom coordinate file.

Example

No example yet exists in the PDB.

//TER==========

TER

Overview

The TER record indicates the end of a chain.

Record Format

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

Details

* Every chain has a TER record.

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

* For cyclic molecules, the choice of termini is arbitrary.

* Present terminal oxygen atoms as atom OXT for proteins, and as O3P or O5* for nucleic acids. The residue name, chain identifier, sequence number and insertion code must be the same as for the terminal residue.

* If the refinement procedure requires that terminal oxygen atoms be used to denote gaps in the chain, remove these atoms before deposition or rename them as the first atom of the next residue.

Verification/Validation/Value Control Authority

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

TER must appear at the end of a chain. The appropriate terminal atom name, OXT, O2P, or O5*, should appear just before the TER record.

Deposition Form Section and Prompt

Example

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
TER     809      GLU A  54
TER    1618      GLU B  54
```

//HETATM==========

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 contained 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	Character	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, IUPAC form

Details

* 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.

* Alternate location is used to indicate that there is more than one position assigned for the given atom. May be denoted by A, B, C, etc.

* The isotropic B value is given for the temperature factor.

* Standard PDB practice is to store the B equivalent in the tempFactor field when anisotropic temperature factors are provided in ANISOU records.

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

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)] x 10**-4

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

* Segment id and element naming are fully described in the ATOM section of this document.

Verification/Validation/Value Control Authority

PDB processing programs check ATOM/HETATM records for PDB format. 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, HETSYN, FORMUL and CONECT records.

Deposition Form Section and Prompt

The Atomic Coordinate Preparation section of the Electronic Deposition Form gives detailed instructions for preparing the coordinates for submission to the PDB.

Example

```
         1         2         3         4         5         6         7         8
12345678901234567890123456789012345678901234567890123456789012345678901234567890

HETATM 1357 MG MG 168 4.669 34.118 19.123 1.00 3.16 MG2+

HETATM 2604  P   PHO I   1      51.711  18.957  -6.294  1.00 15.26           P2+
HETATM 2605  O1P PHO I   1      51.695  20.529  -6.165  1.00 17.35           O
HETATM 2606  O2P PHO I   1      53.151  18.547  -6.825  1.00 12.11           O
HETATM 2607  O3P PHO I   1      50.559  18.333  -7.008  1.00 14.32           O

//ENDMDL=====================================================
```

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 when more than one structure is presented in the entry. These records are generally used in NMR entries. Every MODEL record has an associated ENDMDL record.

Verification/Validation/Value Control Authority

Entries 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. TITLE and EXPDTA must also state that the entry is the result of an NMR study.

Deposition Form Section and Prompt

None; ENDMDL should be included within the coordinate list.

Example

```
          1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
MODEL        1
ENDMDL
MODEL       25
ENDMDL
```

//CONNECTIVITY SECTION==========

Connectivity Section

This section provides information on chemical connectivity. Proposed LINK, HYDBND, and SLTBRG are found in sequence following SSBOND in the Annotation section.

//CONECT==========

CONECT

Overview

The CONECT records specify connectivity between atoms, regardless of bond order. The connectivity is described using the atom serial number as found in the entry. CONECT records are mandatory for HET groups and for other bonds not specified in the standard residue connectivity table which involve atoms in standard residues (see Appendix 3 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

Details

* Connectivity within a non-standard (HET) residue and connections of HET groups to standard groups or to other HET groups are represented on the CONECT records.

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

* Atoms specified in the connectivity are represented 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, a second CONECT record with the same atom serial number in columns 7 - 11 will be used.

* The occurrence of a negative atom serial number denotes that a translationally equivalent copy of the target atom specified is linked to the first atom of the record.

* For nucleic acids, Watson-Crick hydrogen bonds between bases may be listed, but this is optional.

* It is proposed that hydrogen bonds and salt bridges will no longer be included in CONECT but will be given their own records. See HYDBND and SLTBRG.

Verification/Validation/Value Control Authority

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,

Deposition Form Section and Prompt

Example

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
CONECT  461  460  506
CONECT  506  461  505
```

```
CONECT  534  375  533
CONECT  540   25  539
CONECT 1033 1032 2964
CONECT 1274 1273 1549
```

Known Problems

Only five digits are available for the atom serial number, but some structures have already been received with more that 99,999 atoms. Changing the field length would make earlier entries incorrect. Columns beyond 31 have been used for H bonds and salt bridges, and so perhaps with those being moved to new, explicit record types, the additional columns can be used to increase the serial field length.

CONECTs to atoms whose coordinates are not in the entry (e.g., MTRIX related) are not given.

//BOOKKEEPING SECTION==========

Bookkeeping Section

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

//MASTER==========

MASTER

Overview

The MASTER record is a control record for bookkeeping. It lists the number of lines presented in the data entry 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	numFtnote	0, as it used to give the number of FTNOT
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 Control Authority

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.

Deposition Form Section and Prompt

None.

Example

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

//END=============================================================
```

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 Control Authority

END must appear in every coordinate entry

Relationships to Other Record Types

This is the final record in the entry.

Deposition Form Section and Prompt

None.

Example

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
END
```

//APPENDIX 1==========

Appendix 1: BNF for the PDB File

The following is a Backus-Naur Form (BNF) for the standard PDB coordinate entry file.

* "|" indicates a logical or

* parentheses () are used for grouping

* curly braces { } are used to indicate options

* "-list" indicates one or more

* upper case refers to record types

* lower case refers to records grouped for expository purposes

PDBFILE --> title remark primary heterogen secondary

local crystallographic coordinate MASTER END

title --> HEADER {OBSLTE-list} TITLE-list {CAVEAT-list}

COMPND-list SOURCE-list KEYWDS-list EXPDTA-list

AUTHOR-list REVDAT-list {SPRSDE-list} {JRNL-list}

remark --> REMARK-list

primary --> {MODRES-list} DBREF-list {SEQADV-list} SEQRES-list

heterogen --> {HET-list} {HETNAM-list} {HETSYN-list} {FORMUL-list}

secondary --> {HELIX-list} {SHEET-list} {TURN-list}

local --> {SSBOND-list} {LINK-list} {HYDBND-list} {SLTBRG-list}

{CISPEP-list} {SITE-list}

crystallographic --> CRYST1 ORIGX1 ORIGX2 ORIGX3 SCALE1 SCALE2

SCALE3 {(MTRIX1 MTRIX2 MTRIX3)-list} {TVECT}

coordinate --> model-list | atom-sequence CONECT-list

model --> MODEL atom-sequence ENDMDL

atom-sequence --> (atom | chain)-list

atom --> (ATOM | HETATM) {SIGATM} {ANISOU} {SIGUIJ}

chain --> atom-list TER

//APPENDIX 2==========

Appendix 2: Atom Names

Amino Acids

* Atom names follow the IUPAC-IUB rules1 except:

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

* Atoms for which some ambiguity exists in the crystallographic results are designated A. This will usually apply 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.

* Four characters (columns) are reserved for atom names, assigned as follows:

```
COLUMN      VALUE
1-2         Chemical symbol - right justified
3           Remoteness indicator (alphabetic)
4           Branch designator (numeric)
```

* For protein coordinate sets containing hydrogen atoms, the IUPAC-IUB rules1 have been followed, except that recommendation rule number 4.4 has been modified as follows:

* when more than one hydrogen atom is bonded to a single non-hydrogen atom, the hydrogen atom number designation is given as the first character of the atom name rather than as the last character (e.g., H[[beta]]2 is denoted as 2HB).

* If PDB adopts the proposed use of element name in columns 77 - 78 of the ATOM/HETATM records (see ATOM section of this document), this modification to IUPAC-IUB hydrogen naming rules will not be needed.

* Exceptions to IUPAC-IUB rules may occur in certain data sets at the depositor's request. Any such exceptions will be denoted clearly in the REMARK records.

1IUPAC-IUB Commission on Biochemical Nomenclature. "Abbreviations and Symbols for the Description of the Conformation of Polypeptide Chains. Tentative Rules (1969)", *J. Biol. Chem. 245*, 6489 (1970).

The 1974 recommendations on the "Nomenclature of [[alpha]]-Amino Acids" *Biochemistry 14*, 449 (1975) provides a scheme based on normal rules for organic compounds, but this scheme will not be used here.

Nucleic Acids

Atom names employed for polynucleotides generally follow the precedent set for mononucleotides. The following points are worthy of note:

* The prime character (') commonly used to denote atoms of the ribose originally was avoided because of non-uniformity of its external representation. An asterisk (*) therefore was used in its place in entries released through January 1992.

* Of the four characters reserved for atom names, the leftmost two are used for the chemical symbol (right justified) and the other two denote the atom's position. See the section on ATOM records for more information.

* Atoms exocyclic to the ring system have the same position identifier as the atom to which they are bonded unless this results in ambiguous atom names. In this case an alphabetic character is used to avoid ambiguity. See the section on ATOM records for more information.

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

* The extra oxygen atom at the free 5' phosphate terminus is designated OXT. This atom is listed first in the coordinate set.

//APPENDIX 3==========

Appendix 3: 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. At that time, only the twenty common amino acids and five nucleic acids will be treated as "standard" with all others being treated as modified residues to be described by MODRES records. Residues that have appeared in PDB files in the past but which will now be treated as nonstandard groups appear here with an asterisk (*) preceding the name.

Amino Acids

Residue	Abbreviation	Synonym
* [[gamma]]-Aminobutyric acid	ABU	

* Acidic unknown	ACD	
Alanine	ALA	A
* [[beta]]-Alanine	ALB	
* Aliphatic unknown	ALI	
Arginine	ARG	R
* Aromatic unknown	ARO	
Asparagine	ASN	N
Aspartic acid	ASP	D
ASP/ASN ambiguous	ASX	B
* Basic unknown	BAS	
Cysteine	CYS	C, CYH, CSH
Glutamine	GLN	Q
Glutamic acid	GLU	E
GLU/GLN ambiguous	GLX	Z
Glycine	GLY	G
Histidine	HIS	H
* Hydroxyproline	HYP	
Isoleucine	ILE	I, ILU
Leucine	LEU	L
Lysine	LYS	K
Methionine	MET	M
* Pyrrolidone carboxylic acid	PCA	PGA(pyroglutamate)
Phenylalanine	PHE	F
Proline	PRO	P, PR0, PRZ
Serine	SER	S
Threonine	THR	T
Tryptophan	TRP	W, TRY
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

//APPENDIX 4==========

Appendix 4: 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 H71/2 N11/2 O31/2	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 H91/2 N11/2 O31/2	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

Miscellaneous

NAME	CODE	FORMULA	MOL. WT.
Acetic Acid	ACE	C2 H4 O2	60.05
Formic Acid	FOR	C1 H2 O2	46.03
Water	HOH	H2 O1	18.015

//APPENDIX 5==========

Appendix 5: Changes to Remarks (under Consideration)

Overview

PDB acknowledges the drawbacks to computer parsing resulting from the presentation of information in

free format, as found in the REMARK section, and is considering a new structure as well as the addition of more standardized remarks. The following system for organizing and formatting remark records is being presented for comment.

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
REMARK   N
REMARK   N /-------------------------------------------------------\
REMARK   N TOPIC:
REMARK   N SUBTOPIC:                                  (sort number)
REMARK   N STANDARD TEXT:
REMARK   N STANDARD TABLE:
REMARK   N VARIABLE TEXT:
REMARK   N VARIABLE TABLE:
REMARK   N \-------------------------------------------------------/
```

A blank REMARK N line separates keywords from text and lists. The text and tables leave columns 11 - 12 blank. The standard remarks for each subject will have subtopics and there will be a standard ordering for the subtopics. Remarks that have variable text only will have SUBTOPIC: VARIABLE and will appear ahead of remarks for that category that include standard text. The STANDARD TABLE consists of lists that are related to the STANDARD TEXT and the VARIABLE TABLE consists of lists that are related to the VARIABLE TEXT. The sort number will be of the form N.M where M is the number of the standard remark within category N. For variable remarks the sort number will be of the form N.0.M where M is a unique number. Remarks will be ordered by this sort number, first sorted on N and then on M. Remarks with numbers of the form N.0.M will precede remarks with numbers of the form N.M.

In general, the contents of STANDARD TEXT and STANDARD TABLE will be generated by the PDB and the contents of VARIABLE TEXT and VARIABLE TABLE will be generated by the depositor, with guidance from PDB staff as needed.

Verification/Validation/Value Control Authority

All STANDARD TEXT fields must be identical to those given in the template. Some of the information in the STANDARD TABLE fields will be generated automatically by PDB processing programs and that information will not need further verification. Other information in the STANDARD TABLE fields can be verified by comparison with other related records. If the information in VARIABLE TABLE fields can be properly identified as residue or atom names, then these can be checked for existence in the entry. If an external authority was used as a source of information, it will be identified in the remark.

Topics

1. REFERENCE

REMARK 1.

2. RESOLUTION

REMARK 2.

3. REFINEMENT

REMARK 3 has traditionally been used for refinement data. Perhaps a different numbered remark should be used to present analogous information for the NMR entries.

6. WARNING

This would include information present on CAVEAT records so that someone looking at the remarks would be aware of the caveat. Also included would be explanations about regions that are poorly defined, indication that this is early data, etc.

5. OTHER

This would include information that does not fit into other categories. On the assumption that we will see patterns and develop standard remarks for some of these, we will allow gaps in the standard numbering that we are here proposing to allow for new categories in the future.

12. MOLECULE

Information on the biologically functional molecule.

15. SYMMETRY

Information about crystallographic and non-crystallographic symmetry.

18. EXPERIMENTAL DETAILS

21. STRUCTURE SOLUTION

This would include the method used to solve the structure.

24. ENTRY CONTENTS

27. HYDROGEN AND/OR DEUTERIUM ATOMS

30. RESIDUE NUMBERING

33. RELATED ENTRIES

This would point to related coordinate entries, structure factor files, NMR restraints files, etc.

36. DIAGNOSTICS

This section includes remarks about bond lengths, bond angles, chirality, remote waters, etc. that are generated by PDB checking programs as well as related remarks that are provided by the depositor.

39. SEQUENCE

Missing atoms, missing residues.

42. NUCLEIC ACIDS

45. BONDS

48. HET

51. HELIX

54. SHEET

57. TURN

60. SITE

63. CORRECTION

Standard remarks:

This section contains a summary that shows each topic and subtopic and the standard remarks that have been defined so far.

```
REMARK   3 TOPIC:
REMARK   3 SUBTOPIC:                                (sort number)

REMARK   6 TOPIC: WARNING
REMARK   6 SUBTOPIC: NON-IUPAC-IUB                  (6.1)

REMARK  15 TOPIC: SYMMETRY
REMARK  15 SUBTOPIC:                                (sort number)

REMARK  18 TOPIC: EXPERIMENTAL TECHNIQUE
REMARK  18 SUBTOPIC: EXPERIMENTAL DATA N            (18.1)
REMARK  18 SUBTOPIC: EXPERIMENTAL DATA N            (18.2)

REMARK  21 TOPIC: STRUCTURE SOLUTION
REMARK  21 SUBTOPIC: NMR                            (21.1)
REMARK  21 SUBTOPIC: MODEL                          (21.2)
REMARK  21 SUBTOPIC: FIBER DIFFRACTION              (21.3)

REMARK  24 TOPIC: ENTRY CONTENTS
REMARK  24 SUBTOPIC: MODIFIED RESIDUE               (24.1)
REMARK  24 SUBTOPIC: INOSINE                        (24.2)
REMARK  24 SUBTOPIC: LONE PAIRS                     (24.3)

REMARK  27 TOPIC: HYDROGEN AND/OR DEUTERIUM ATOMS
REMARK  27 SUBTOPIC: PDB ATOM NAMING                (27.1)

REMARK  33 TOPIC: RELATED ENTRIES
REMARK  33 SUBTOPIC: STRUCTURE FACTORS              (33.1)
REMARK  33 SUBTOPIC: NMR RESTRAINTS                 (33.2)
REMARK  33 SUBTOPIC: RELATED COORDINATE ENTRIES     (33.3)

REMARK  36 TOPIC: STEREOCHEMISTRY
REMARK  36 SUBTOPIC: BOND LENGTHS                   (36.1)
REMARK  36 SUBTOPIC: COVALENT BOND ANGLES           (36.2)
```

```
REMARK  36 SUBTOPIC: CHIRAL CENTERS                    (36.3)

REMARK  42 TOPIC: NUCLEIC ACIDS
REMARK  42 SUBTOPIC: NOMENCLATURE                      (42.1)

REMARK  45 TOPIC: BONDS
REMARK  45 SUBTOPIC: WATSON-CRICK                      (45.1)
REMARK  45 SUBTOPIC: NON-WATSON-CRICK                  (45.2)
REMARK  45 SUBTOPIC: BASE MISPAIRINGS                  (45.3)

REMARK  54 TOPIC: SHEET
REMARK  54 SUBTOPIC: BARREL                            (54.1)
REMARK  54 SUBTOPIC: BIFURCATED                        (54.2)

REMARK  63 TOPIC: CORRECTION
REMARK  63 SUBTOPIC: CORRECTION                        (63.1)

REMARK   N TOPIC:
REMARK   N SUBTOPIC:                                   (sort number)

REMARK   3
REMARK   3 /----------------------------------------------------------\
REMARK   3 TOPIC:
REMARK   3 SUBTOPIC:                                   (sort number)
REMARK   3 STANDARD TEXT:
REMARK   3 STANDARD TABLE:
REMARK   3 VARIABLE TEXT:
REMARK   3 VARIABLE TABLE:
REMARK   3 \----------------------------------------------------------/
REMARK   3
REMARK   3  ATOMS WITH THERMAL FACTORS WHICH CALCULATE LESS THAN 2.00
REMARK   3  ARE ASSIGNED THIS VALUE.  THIS IS THE LOWEST VALUE ALLOWED
REMARK   3  BY THE REFINEMENT PROGRAM.

REMARK   6
REMARK   6 /----------------------------------------------------------\
REMARK   6 TOPIC: WARNING
REMARK   6
REMARK   6 SUBTOPIC: NON-IUPAC-IUB                     (6.1)
REMARK   6
REMARK   6 STANDARD TEXT:
REMARK   6
REMARK   6  BY REQUEST OF THE DEPOSITOR, THE PROTEIN DATA BANK HAS NOT
REMARK   6  APPLIED THE IUPAC-IUB RECOMMENDATIONS REGARDING THE
REMARK   6  DESIGNATION OF BRANCHES 1 AND 2 OF SIDE-CHAIN ATOMS IN
REMARK   6  RESIDUES ARG, ASP, GLU, LEU, PHE, TYR, AND VAL TO THIS
REMARK   6  ENTRY.
REMARK   6
REMARK   6 STANDARD TABLE:
REMARK   6
REMARK   6 VARIABLE TEXT:
REMARK   6
REMARK   6 VARIABLE TABLE:
REMARK   6
REMARK   6 \----------------------------------------------------------/

REMARK  15
REMARK  15 /----------------------------------------------------------\
REMARK  15 TOPIC: SYMMETRY
REMARK  15
REMARK  15 SUBTOPIC: CRYSTAL SYMMETRY OPERATORS (sort number)
```

```
REMARK  15
REMARK  15 STANDARD TEXT:
REMARK  15
REMARK  15  FOLLOWING ARE THE SYMMETRY OPERATORS USED IN THE
REMARK  15  REFINEMENT.
REMARK  15
REMARK  15 STANDARD TABLE:
REMARK  15
REMARK  15      X, Y, Z       (START IN COLUMN 17, ALIGN THE COMMAS)
REMARK  15  1/2-X, Y, -Z
REMARK  15 \-----------------------------------------------------------/

REMARK  15
REMARK  15 /-----------------------------------------------------------\
REMARK  15 TOPIC: SYMMETRY
REMARK  15
REMARK  15 SUBTOPIC: SPECIAL CRYSTAL SYMMETRY MATRICES (sort number)
REMARK  15
REMARK  15 STANDARD TEXT:
REMARK  15
REMARK  15  THE FOLLOWING MATRICES ARE USED TO OPERATE ON THE
REMARK  15  COORDINATES GIVEN IN THIS ENTRY.  THEY PRODUCE ATOMS
REMARK  15  RELATED BY CRYSTALLOGRAPHIC SYMMETRY.  THESE  ARE
REMARK  15  PROVIDED FOR USE IN GENERATING BIOMOLECULES, DESCRIBING
REMARK  15  BONDING BETWEEN ATOMS, ETC.  THESE OPERATORS FOLLOW THE
REMARK  15  SAME FORMAT AS MTRIX WITH THE EXCEPTION THAT THEY SHOULD
REMARK  15  BE READ STARTING IN COLUMN 13.
REMARK  15
REMARK  15  THESE MATRICES ARE REFERRED TO AS S1, S2, ETC. AND ARE
REMARK  15  DESCRIBED IN REMARK 15.X.
REMARK  15
REMARK  15 STANDARD TABLE:
REMARK  15
REMARK  15  SYMTY1   1  A11      A12       A13              T1
REMARK  15  SYMTY2   1  A21      A22       A23              T2
REMARK  15  SYMTY3   1  A31      A32       A33              T3
REMARK  15  SYMTY1   2  A11      A12       A13              T1
REMARK  15  SYMTY2   2  A21      A22       A23              T2
REMARK  15  SYMTY3   2  A31      A32       A33              T3
REMARK  15 \-----------------------------------------------------------/
REMARK  15
REMARK  15 /-----------------------------------------------------------\
REMARK  15.1
REMARK  15 TOPIC: SYMMETRY
REMARK  15
REMARK  15 SUBTOPIC: USE OF SYMMETRY MATRICES  (sort number)
REMARK  15
REMARK  15 STANDARD TEXT:
REMARK  15
REMARK  15  THE FOLLOWING TABLE DESCRIBES THE USE OF THE SYMMETRY
REMARK  15  MATRICES FOUND IN THIS ENTRY.
REMARK  15
REMARK  15  FOR EXAMPLE, S1 REFERS TO THE FIRST CRYSTALLOGRAPHIC
REMARK  15  SYMMETRY MATRIX GIVEN IN 15.X, AND M1 REFERS TO THE
REMARK  15  FIRST SET OF NON-CRYSTALLOGRAPHIC MTRIX RECORDS BELOW.
REMARK  15
REMARK  15  CHAIN NAMES GIVEN AS "?" REFER TO CHAINS FOR WHICH
REMARK  15  ATOMS ARE NOT FOUND IN THIS ENTRY.
REMARK  15
REMARK  15 STANDARD TABLE:
REMARK  15
```

```
REMARK  15  1 DESCRIPTION OF S1: GENERATES ONE CHAIN OF THE TRIMER
REMARK  15  2 DESCRIPTION OF S2: GENERATES ONE CHAIN OF THE TRIMER
REMARK  15  3 DESCRIPTION OF M1: RELATES MONOMER A TO MONOMER B
REMARK  15  4 DESCRIPTION OF M2: RELATES MONOMER A TO MONOMER C
REMARK  15
REMARK  15            APPLIED TO          TRANSFORMED TO
REMARK  15    MTRIX  CHAIN  RESIDUES       CHAIN  RESIDUES     RMSD
REMARK  15  1   S1     A    1  -  340        ?    1  -  340
REMARK  15  2   S2     A    1  -  340        ?    1  -  340
REMARK  15  3   M1     A    1  -  340        B    1  -  340     0.03
REMARK  15  4   M2     A    1  -  340        C    1  -  340     0.03
REMARK  15 \----------------------------------------------------------/

REMARK  21
REMARK  21 /----------------------------------------------------------\
REMARK  21 TOPIC: EXPERIMENT
REMARK  21
REMARK  21 SUBTOPIC: NMR                                  (21.1)
REMARK  21
REMARK  21 STANDARD TEXT:
REMARK  21
REMARK  21  THESE COORDINATES WERE GENERATED FROM SOLUTION NMR DATA.
REMARK  21  PROTEIN DATA BANK CONVENTIONS REQUIRE THAT CRYST1 AND
REMARK  21  SCALE RECORDS BE INCLUDED, BUT THE VALUES OF THESE
REMARK  21  RECORDS ARE MEANINGLESS.
REMARK  21
REMARK  21 STANDARD TABLE:
REMARK  21
REMARK  21 VARIABLE TEXT:
REMARK  21
REMARK  21 VARIABLE TABLE:
REMARK  21
REMARK  21 \----------------------------------------------------------/

REMARK  21
REMARK  21 /----------------------------------------------------------\
REMARK  21 TOPIC: EXPERIMENT
REMARK  21
REMARK  21 SUBTOPIC: THEORETICAL MODEL                    (21.2)
REMARK  21
REMARK  21 STANDARD TEXT:
REMARK  21
REMARK  21  THE COORDINATES IN THIS ENTRY REPRESENT A MODEL STRUCTURE.
REMARK  21  PROTEIN DATA BANK CONVENTIONS REQUIRE THAT CRYST1 AND
REMARK  21  SCALE RECORDS BE INCLUDED, BUT THE VALUES ON THESE
REMARK  21  RECORDS ARE MEANINGLESS.
REMARK  21
REMARK  21 STANDARD TABLE:
REMARK  21
REMARK  21 VARIABLE TEXT:
REMARK  21
REMARK  21 VARIABLE TABLE:
REMARK  21
REMARK  21 \----------------------------------------------------------/
REMARK  21
REMARK  21 /----------------------------------------------------------\
REMARK  21 TOPIC: EXPERIMENT
REMARK  21
REMARK  21 SUBTOPIC: FIBER DIFFRACTION                    (21.3)
REMARK  21
REMARK  21 STANDARD TEXT:
```

```
REMARK  21
REMARK  21  THESE COORDINATES WERE GENERATED FROM FIBER DIFFRACTION
REMARK  21  DATA.  PROTEIN DATA BANK CONVENTIONS REQUIRE THAT CRYST1
REMARK  21  AND SCALE RECORDS BE INCLUDED, BUT THE VALUES OF THESE
REMARK  21  RECORDS ARE MEANINGLESS.
REMARK  21
REMARK  21 STANDARD TABLE:
REMARK  21
REMARK  21 VARIABLE TEXT:
REMARK  21
REMARK  21 VARIABLE TABLE:
REMARK  21
REMARK  21 \------------------------------------------------------/

REMARK  24
REMARK  24 /------------------------------------------------------\
REMARK  24 TOPIC: ENTRY CONTENTS
REMARK  24
REMARK  24 SUBTOPIC: INOSINE                          (24.2)
REMARK  24
REMARK  24 STANDARD TEXT:
REMARK  24
REMARK  24  THIS ENTRY CONTAINS ONE OR MORE INOSINE RESIDUES.
REMARK  24  (M=MODEL NUMBER; RES=RESIDUE NAME; C=CHAIN IDENTIFIER;
REMARK  24  SSEQ=SEQUENCE NUMBER; I=INSERTION CODE).
REMARK  24
REMARK  24 STANDARD TABLE:
REMARK  24
REMARK  24  M RES CSSEQI
REMARK  24      I A   1
REMARK  24      I A   3
REMARK  24
REMARK  24 VARIABLE TEXT:
REMARK  24
REMARK  24 VARIABLE TABLE:
REMARK  24
REMARK  24 \------------------------------------------------------/

REMARK  24
REMARK  24 /------------------------------------------------------\
REMARK  24 TOPIC: ENTRY CONTENTS
REMARK  24
REMARK  24 SUBTOPIC: LONE PAIRS                       (24.3)
REMARK  24
REMARK  24 STANDARD TEXT:
REMARK  24
REMARK  24  BELOW ARE THE COORDINATES FOR LONE PAIRS OF ELECTRONS
REMARK  24  REPORTED IN THIS NMR ENTRY.  THEY ARE PRESENTED EXACTLY
REMARK  24  AS THEY APPEARED IN THE RAW DATA FILE SENT TO THE
REMARK  24  PROTEIN DATA BANK BY THE DEPOSITOR.
REMARK  24
REMARK  24 STANDARD TABLE:
REMARK  24
REMARK  24  MODEL        3
REMARK  24   CP1 MET    34      21.761   8.679  -8.481
REMARK  24
REMARK  24 VARIABLE TEXT:
REMARK  24
REMARK  24 VARIABLE TABLE:
REMARK  24
REMARK  24 \------------------------------------------------------/
```

```
REMARK  27
REMARK  27 /------------------------------------------------------------\
REMARK  27 TOPIC: HYDROGEN AND/OR DEUTERIUM ATOMS
REMARK  27
REMARK  27 SUBTOPIC: PDB ATOM NAMING                    (27.1)
REMARK  27
REMARK  27 STANDARD TEXT:
REMARK  27
REMARK  27  HYDROGEN AND DEUTERIUM ATOMS IN THIS ENTRY HAVE BEEN
REMARK  27  ASSIGNED NAMES CONSISTENT WITH THE RECOMMENDATIONS OF THE
REMARK  27  IUPAC-IUB COMMISSION ON BIOCHEMICAL NOMENCLATURE
REMARK  27  (J.MOL.BIOL. (1970) VOL. 52, PP 1-17).  THE PROTEIN DATA
REMARK  27  BANK HAS FOLLOWED RULE 4.4 OF THE RECOMMENDATIONS WITH THE
REMARK  27  FOLLOWING MODIFICATION:  WHEN MORE THAN ONE HYDROGEN ATOM
REMARK  27  IS BONDED TO A SINGLE NON-HYDROGEN ATOM, THE HYDROGEN ATOM
REMARK  27  NUMBER DESIGNATION IS GIVEN AS THE FIRST CHARACTER OF THE
REMARK  27  ATOM NAME RATHER THAN AS THE LAST CHARACTER (E.G., H BETA1
REMARK  27  IS DENOTED AS 1HB).
REMARK  27
REMARK  27 STANDARD TABLE:
REMARK  27
REMARK  27 VARIABLE TEXT:
REMARK  27
REMARK  27 VARIABLE TABLE:
REMARK  27
REMARK  27 \------------------------------------------------------------/

REMARK  33
REMARK  33 /------------------------------------------------------------\
REMARK  33 TOPIC: RELATED ENTRIES
REMARK  33
REMARK  33 SUBTOPIC: STRUCTURE FACTORS                  (33.1)
REMARK  33
REMARK  33 STANDARD TEXT:
REMARK  33
REMARK  33  STRUCTURE FACTORS CORRESPONDING TO THIS ENTRY ARE
REMARK  33  AVAILABLE FROM THE PROTEIN DATA BANK AS A SEPARATE ENTRY.
REMARK  33  THESE ARE LISTED ONE PER LINE.
REMARK  33
REMARK  33 STANDARD TABLE:
REMARK  33
REMARK  33  R1ABCSF
REMARK  33
REMARK  33 VARIABLE TEXT:
REMARK  33
REMARK  33 VARIABLE TABLE:
REMARK  33
REMARK  33 \------------------------------------------------------------/

REMARK  33
REMARK  33 /------------------------------------------------------------\
REMARK  33 TOPIC: RELATED ENTRIES
REMARK  33
REMARK  33 SUBTOPIC: NMR RESTRAINTS                     (33.2)
REMARK  33
REMARK  33 STANDARD TEXT:
REMARK  33
REMARK  33  NMR RESTRAINTS CORRESPONDING TO THIS ENTRY ARE
REMARK  33  AVAILABLE FROM THE PROTEIN DATA BANK AS A SEPARATE ENTRY.
REMARK  33  THESE ARE LISTED ONE PER LINE.
REMARK  33
```

```
REMARK  33 STANDARD TABLE:
REMARK  33
REMARK  33  R1ABCMR
REMARK  33
REMARK  33 VARIABLE TEXT:
REMARK  33
REMARK  33 VARIABLE TABLE:
REMARK  33
REMARK  33 \-----------------------------------------------------------/

REMARK  33
REMARK  33 /-----------------------------------------------------------\
REMARK  33 TOPIC: RELATED ENTRIES
REMARK  33
REMARK  33 SUBTOPIC: RELATED COORDINATE ENTRIES     (33.3)
REMARK  33
REMARK  33 STANDARD TEXT:
REMARK  33
REMARK  33  THE FOLLOWING PROTEIN DATA BANK ENTRIES ARE RELATED TO
REMARK  33  THIS ENTRY.  THESE ARE LISTED ONE PER LINE.
REMARK  33
REMARK  33 STANDARD TABLE:
REMARK  33
REMARK  33  1ABC
REMARK  33
REMARK  33 VARIABLE TEXT:
REMARK  33
REMARK  33 VARIABLE TABLE:
REMARK  33
REMARK  33 \-----------------------------------------------------------/

REMARK  36
REMARK  36 /-----------------------------------------------------------\
REMARK  36 TOPIC: STEREOCHEMISTRY
REMARK  36
REMARK  36 SUBTOPIC: BOND LENGTHS                   (36.1)
REMARK  36
REMARK  36 STANDARD TEXT:
REMARK  36
REMARK  36  THE STEREOCHEMICAL PARAMETERS OF THE FOLLOWING RESIDUES
REMARK  36  HAVE VALUES WHICH DEVIATE FROM EXPECTED VALUES BY MORE
REMARK  36  THAN X*RMSD (M=MODEL NUMBER; RES=RESIDUE NAME; C=CHAIN
REMARK  36  IDENTIFIER; SSEQ=SEQUENCE NUMBER; I=INSERTION CODE).
REMARK  36
REMARK  36 STANDARD TABLE:
REMARK  36
REMARK  36  EXPECTED VALUES: ENGH AND HUBER, 1991
REMARK  36  X: 4.0
REMARK  36  M RES CSSEQI ATM1   ATM2       DISTANCE_DEVIATION_IN_ANGS.
REMARK  36  0 TYR A  40   CB     CA        0.08
REMARK  36  0 LYS A 180   C      O         0.08
REMARK  36
REMARK  36 VARIABLE TEXT:
REMARK  36
REMARK  36 VARIABLE TABLE:
REMARK  36
REMARK  36 \-----------------------------------------------------------/

REMARK  36
REMARK  36 /-----------------------------------------------------------\
REMARK  36 TOPIC: STEREOCHEMISTRY
```

```
REMARK  36
REMARK  36 SUBTOPIC: COVALENT BOND ANGLES                (36.2)
REMARK  36
REMARK  36 STANDARD TEXT:
REMARK  36
REMARK  36  THE STEREOCHEMICAL PARAMETERS OF THE FOLLOWING RESIDUES
REMARK  36  HAVE VALUES WHICH DEVIATE FROM EXPECTED VALUES BY MORE
REMARK  36  THAN X*RMSD (M=MODEL NUMBER; RES=RESIDUE NAME; C=CHAIN
REMARK  36  IDENTIFIER; SSEQ=SEQUENCE NUMBER; I=INSERTION CODE;
REMARK  36  DEV=DEVIATION IN DEGREES).
REMARK  36
REMARK  36 STANDARD TABLE:
REMARK  36
REMARK  36  EXPECTED VALUES: ENGH AND HUBER, 1991
REMARK  36  X: 4.0
REMARK  36  M RES CSSEQI ATM1  RES CSSEQI ATM2  RES CSSEQI ATM2  DEV
REMARK  36  0 LEU A  32   CA   LEU A  32  CB    LEU A  32   CG   17.3
REMARK  36  0 THR A  41   N    THR A  41  CA    THR A  41   C    14.0
REMARK  36  0 THR A 148   CA   THR A 148  C     GLY A 149   N    17.3
REMARK  36
REMARK  36 VARIABLE TEXT:
REMARK  36
REMARK  36 VARIABLE TABLE:
REMARK  36
REMARK  36 \----------------------------------------------------------/

REMARK  36
REMARK  36 /----------------------------------------------------------\
REMARK  36 TOPIC: STEREOCHEMISTRY
REMARK  36
REMARK  36 SUBTOPIC: CHIRAL CENTERS                      (36.3)
REMARK  36
REMARK  36 STANDARD TEXT:
REMARK  36
REMARK  36  UNEXPECTED CONFIGURATION OF THE FOLLOWING CHIRAL
REMARK  36  CENTER(S) (M=MODEL NUMBER; RES=RESIDUE NAME; C=CHAIN
REMARK  36  IDENTIFIER; SSEQ=SEQUENCE NUMBER; I=INSERTION CODE).
REMARK  36
REMARK  36 STANDARD TABLE:
REMARK  36
REMARK  36  M RES CSSEQI
REMARK  36
REMARK  36 VARIABLE TEXT:
REMARK  36
REMARK  36 VARIABLE TABLE:
REMARK  36
REMARK  36 \----------------------------------------------------------/

REMARK  42
REMARK  42 /----------------------------------------------------------\
REMARK  42 TOPIC: NUCLEIC ACIDS
REMARK  42
REMARK  42 SUBTOPIC: NOMENCLATURE                        (42.1)
REMARK  42
REMARK  42 STANDARD TEXT:
REMARK  42
REMARK  42  THE PROTEIN DATA BANK HAS ADOPTED THE SACCHARIDE CHEMISTS'
REMARK  42  NOMENCLATURE FOR ATOMS OF THE DEOXYRIBOSE/RIBOSE MOIETY
REMARK  42  RATHER THAN THAT OF THE NUCLEOSIDE CHEMISTS.  THE
REMARK  42  RING-OXYGEN ATOM, ACCORDINGLY, IS LABELED O4* AND NOT
REMARK  42  O1*.
```

```
REMARK  42
REMARK  42 STANDARD TABLE:
REMARK  42
REMARK  42 VARIABLE TEXT:
REMARK  42
REMARK  42 VARIABLE TABLE:
REMARK  42
REMARK  42 \------------------------------------------------------/

REMARK  45
REMARK  45 /------------------------------------------------------\
REMARK  45 TOPIC: BONDS
REMARK  45
REMARK  45 SUBTOPIC: WATSON-CRICK                    (45.1)
REMARK  45
REMARK  45 STANDARD TEXT:
REMARK  45
REMARK  45  THE HYDROGEN BONDS BETWEEN BASE PAIRS IN THIS ENTRY FOLLOW
REMARK  45  THE CONVENTIONAL WATSON-CRICK HYDROGEN BONDING PATTERN.
REMARK  45  THEY HAVE NOT BEEN PRESENTED ON CONECT RECORDS IN THIS
REMARK  45  ENTRY.
REMARK  45
REMARK  45 STANDARD TABLE:
REMARK  45
REMARK  45 VARIABLE TEXT:
REMARK  45
REMARK  45 VARIABLE TABLE:
REMARK  45
REMARK  45 \------------------------------------------------------/

REMARK  45
REMARK  45 /------------------------------------------------------\
REMARK  45 TOPIC: BONDS
REMARK  45
REMARK  45 SUBTOPIC: NON-WATSON-CRICK                (45.2)
REMARK  45
REMARK  45 STANDARD TEXT:
REMARK  45
REMARK  45  THERE ARE NON-WATSON-CRICK HYDROGEN BONDS BETWEEN THE
REMARK  45  ATOMS IN THIS ENTRY AS SHOWN BELOW.  ALL OTHER HYDROGEN
REMARK  45  BONDS BETWEEN BASE PAIRS IN THIS ENTRY FOLLOW THE
REMARK  45  CONVENTIONAL WATSON-CRICK HYDROGEN BONDING PATTERN AND
REMARK  45  THEY HAVE NOT BEEN PRESENTED ON CONECT RECORDS IN THIS
REMARK  45  ENTRY.
REMARK  45
REMARK  45 STANDARD TABLE:
REMARK  45
REMARK  45  M ATM1 RES CSSEQI   ATM2 RES CSSEQI
REMARK  45     N1    I A   1     N3    C B  16
REMA
REMARK  45 VARIABLE TABLE:
REMARK  45
REMARK  45 \------------------------------------------------------/

REMARK  45
REMARK  45 /------------------------------------------------------\
REMARK  45 TOPIC: BONDS
REMARK  45
REMARK  45 SUBTOPIC: BASE MISPAIRINGS                (45.3)
REMARK  45
REMARK  45 STANDARD TEXT:
```

```
REMARK  45
REMARK  45  THERE ARE ONE OR MORE MISPAIRINGS BETWEEN BASES IN THIS
REMARK  45  ENTRY.  ALL OTHER HYDROGEN BONDS BETWEEN BASE PAIRS IN
REMARK  45  THIS ENTRY FOLLOW THE CONVENTIONAL WATSON-CRICK HYDROGEN
REMARK  45  BONDING PATTERN AND THEY HAVE NOT BEEN PRESENTED ON
REMARK  45  CONECT RECORDS IN THIS ENTRY.
REMARK  45
REMARK  45 STANDARD TABLE:
REMARK  45
REMARK  45  M RES CSSEQI   RES CSSEQI
REMARK  45      G A   4      A B  21
REMARK  45      A A   9      G B  16
REMARK  45
REMARK  45 VARIABLE TEXT:
REMARK  45
REMARK  45 VARIABLE TABLE:
REMARK  45
REMARK  45 \----------------------------------------------------------/

REMARK  54
REMARK  54 /----------------------------------------------------------\
REMARK  54 TOPIC: SHEET
REMARK  54
REMARK  54 SUBTOPIC: BARREL                          (54.1)
REMARK  54
REMARK  54 STANDARD TEXT:
REMARK  54
REMARK  54  THIS ENTRY CONTAINS ONE OR MORE BETA-BARRELS.
REMARK  54  A BETA-BARREL IS REPRESENTED BY REPEATING THE FIRST
REMARK  54  STRAND AT THE END OF THE LIST OF STRANDS.  THUS A BARREL
REMARK  54  WITH N STRANDS IS REPRESENTED BY A SHEET WITH N+1 STRANDS.
REMARK  54
REMARK  54 STANDARD TABLE:
REMARK  54
REMARK  54  NAME OF BARREL
REMARK  54      BAR
REMARK  54
REMARK  54 VARIABLE TEXT:
REMARK  54
REMARK  54 VARIABLE TABLE:
REMARK  54
REMARK  54 \----------------------------------------------------------/

REMARK  54
REMARK  54 /----------------------------------------------------------\
REMARK  54 TOPIC: SHEET
REMARK  54
REMARK  54 SUBTOPIC: BIFURCATED                      (54.2)
REMARK  54
REMARK  54 STANDARD TEXT:
REMARK  54
REMARK  54  THIS ENTRY CONTAINS ONE OR MORE BIFURCATED SHEETS.
REMARK  54  A BIFURCATED SHEET IS REPRESENTED BY MULTIPLE SHEETS WHICH
REMARK  54  HAVE ONE OR MORE IDENTICAL STRANDS.  EACH LINE IN THE LIST
REMARK  54  BELOW REPRESENTS THE SHEETS THAT TAKEN TOGETHER REPRESENT
REMARK  54  ONE BIFURCATED SHEET.
REMARK  54
REMARK  54 STANDARD TABLE:
REMARK  54
REMARK  54  S2A S2B
REMARK  54  S3A S3B
```

```
REMARK  54  S4A S4B
REMARK  54
REMARK  54 VARIABLE TEXT:
REMARK  54
REMARK  54 VARIABLE TABLE:
REMARK  54
REMARK  54 \-----------------------------------------------------------/

REMARK  63
REMARK  63 /-----------------------------------------------------------\
REMARK  63 TOPIC: CORRECTION
REMARK  63
REMARK  63 SUBTOPIC: CORRECTION                          (63.1)
REMARK  63
REMARK  63  A CORRECTION HAS BEEN APPLIED TO THIS ENTRY.  CONSULT THE
REMARK  63  REVDAT RECORD(S) ASSOCIATED WITH THIS CORRECTION AND THE
REMARK  63  VARIABLE TEXT FIELD BELOW FOR ADDITIONAL INFORMATION.
REMARK  63
REMARK  63 STANDARD TABLE:
REMARK  63
REMARK  63  DATE:
REMARK  63  REVISION NAME:
REMARK  63
REMARK  63 VARIABLE TEXT:
REMARK  63
REMARK  63 VARIABLE TABLE:
REMARK  63
REMARK  63 \-----------------------------------------------------------/
```

For structures with less than the complete duplex in the asymmetric unit.

```
REMARK 100
REMARK 100 THIS ENTRY CONTAINS THE CRYSTALLOGRAPHIC ASYMMETRIC UNIT
REMARK 100 WHICH CONSISTS OF <number of chains> <chain name:e.g.DNA>
REMARK 100 CHAIN(S) AND <number and name of drug molecules if present>
REMARK 100 MOLECULE(S).
REMARK 100
REMARK 100 TO GENERATE COMPLEX <number of complex that can be generated>
REMARK 100 THE FOLLOWING CRYSTALLOGRAPHIC SYMMETRY OPERATION MUST BE
REMARK 100 APPLIED TO THE COORDINATES FOR THE ATOMS WITH CHAIN ID
REMARK 100 <strand id> OF THIS ENTRY
REMARK 100
REMARK 100   SMTRY1      -1.000000  0.000000  0.000000        0.00000
REMARK 100   SMTRY2       0.000000 -1.000000  0.000000        0.00000
REMARK 100   SMTRY3       0.000000  0.000000  1.000000        0.00000
```

Example

```
REMARK 100
REMARK 100 THIS ENTRY CONTAINS THE CRYSTALLOGRAPHIC ASYMMETRIC UNIT
REMARK 100 WHICH CONSISTS OF 2 INDEPENDENT DNA CHAINS AND 2 HALF
REMARK 100 ACTINOMYCIN D MOLECULES.
REMARK 100
REMARK 100 TO GENERATE COMPLEX 1 THE FOLLOWING CRYSTALLOGRAPHIC
REMARK 100 SYMMETRY OPERATION MUST BE APPLIED TO THE COORDINATES FOR
REMARK 100 THE ATOMS WITH CHAIN IDENTIFIER A OF THIS ENTRY:
REMARK 100
REMARK 100   SMTRY1      -1.000000  0.000000  0.000000        0.00000
REMARK 100   SMTRY2       0.000000 -1.000000  0.000000        0.00000
REMARK 100   SMTRY3       0.000000  0.000000  1.000000        0.00000
REMARK 100
```

```
REMARK 100 TO GENERATE COMPLEX 2 THE FOLLOWING CRYSTALLOGRAPHIC
REMARK 100 SYMMETRY OPERATION MUST BE APPLIED TO THE COORDINATES FOR
REMARK 100 THE ATOMS WITH CHAIN ID B OF THIS ENTRY:
REMARK 100
REMARK 100   SMTRY1     -1.000000  0.000000  0.000000       34.95000
REMARK 100   SMTRY2      0.000000 -1.000000  0.000000       30.70000
REMARK 100   SMTRY3      0.000000  0.000000  1.000000        0.00000
```

==========