Protein Data Bank Contents Guide:

Atomic Coordinate Entry Format Description

Version 2.3, July 9, 1998

1. Introduction

The Protein Data Bank (PDB) is an archive of experimentally determined three-dimensional structures of biological macromolecules that serves a global community of researchers, educators, and students. The data contained in the archive include atomic coordinates, bibliographic citations, primary and secondary structure, information, and crystallographic structure factors and NMR experimental data.

This guide describes the “PDB format” used by the members of the worldwide Protein Data Bank (Berman, H.M., Henrick, K. and Nakamura, H. (2003) Announcing the worldwide Protein Data Bank. *Nat Struct Biol*, **10**, 980). Questions should be sent to info@wwpdb.org.

This version of the PDB file format has been in use since July 9, 1998. Please note that as of July 1, 2002, models are available from a directory separate from the main archive at ftp://ftp.rcsb.org/pub/pdb/data/structures/models/current/. As of October 15, 2006, theoretical models are no longer accepted for deposition.

Basic Notions of the Format Description

Character Set

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

abcdefghijklmnopqrstuvwxyzABCDEFGHIJKLMNOPQRSTUVWXYZ

1234567890

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

We discourage use of punctuation characters in the place of alphanumeric characters.

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

Special Characters

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

Bullets are represented as (DOT).

Right arrow is represented as -->.

Left arrow is represented as <--.

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

Commas, colons, and semi-colons are used as list delimiters in records that have one of the following data types:

List

SList

Specification List

Specification

If a comma, colon, or semi-colon is used in any context other than as a delimiting character, then the character must be escaped, i.e., immediately preceded by a backslash, "\". Examples of this use are found in line 4 of each of the following:

```
COMPND        MOL_ID: 1;
COMPND      2 MOLECULE: GLUTATHIONE SYNTHETASE;
COMPND      3 CHAIN: A;
COMPND      4 SYNONYM: GAMMA-L-GLUTAMYL-L-CYSTEINE\:GLYCINE LIGASE
COMPND      5 (ADP-FORMING);
COMPND      6 EC: 6.3.2.3;
COMPND      7 ENGINEERED: YES


COMPND        MOL_ID: 1;
COMPND      2 MOLECULE: S-ADENOSYLMETHIONINE SYNTHETASE;
COMPND      3 CHAIN: A, B;
COMPND      4 SYNONYM: MAT, ATP\:L-METHIONINE S-ADENOSYLTRANSFERASE;
COMPND      5 EC: 2.5.1.6;
COMPND      6 ENGINEERED: YES;
COMPND      7 BIOLOGICAL_UNIT: TETRAMER;
COMPND      8 OTHER_DETAILS: TETRAGONAL MODIFICATION
```

Record Format

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

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

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

Each record type is further divided into fields.

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

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

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

Types of Records

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

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

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

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

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

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

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

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

RECORD TYPE	DESCRIPTION
ANISOU	Anisotropic temperature factors.
ATOM	Atomic coordinate records for standard groups.
CISPEP	Identification of peptide residues in cis conformation.
CONECT	Connectivity records.
DBREF	Reference to the entry in the sequence database(s).
HELIX	Identification of helical substructures.
HET	Identification of non-standard groups or residues (heterogens)
HETSYN	Synonymous compound names for heterogens.
LINK	Identification of inter-residue bonds.
MODRES	Identification of modifications to standard residues.
MTRIXn	Transformations expressing non-crystallographic symmetry (n = 1, 2, or 3). There may be multiple sets of these records.
REVDAT	Revision date and related information.
SEQADV	Identification of conflicts between PDB and the named Sequence database.
SEQRES	Primary sequence of backbone residues.
SHEET	Identification of sheet substructures.
SIGATM	Standard deviations of atomic parameters.
SIGUIJ	Standard deviations of anisotropic temperature factors.
SITE	Identification of groups comprising important sites.
SSBOND	Identification of disulfide bonds.
TVECT	Translation vector for infinite covalently connected structures.

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

subsequent lines. Listed alphabetically, these are:

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

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

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

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

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

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

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

PDB Format Change Policy

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

1. Comments and suggestions will be solicited from the community on specific problems and data representation issues as they arise.

2. Proposed format changes will be disseminated through pdb-l@rcsb.org and www.pdb.org.

3. A sixty-day discussion period will follow the announcement of proposed changes. Comments and suggestions must be received within this time period. Major changes that are not upwardly compatible will be allotted up to twice the standard amount of discussion time.

4. The PDB will then work in consultation with the wwPDB Advisory Committee and the equivalent partner Scientific Advisory Committees to evaluate and reconcile all suggestions. The final decision will be officially announced via pdb-l@rcsb.org and www.pdb.org.

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

Order of Records

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

RECORD TYPE	EXISTENCE	CONDITIONS IF OPTIONAL
HEADER	Mandatory	
OBSLTE	Optional	Mandatory in withdrawn entries.
TITLE	Mandatory	
CAVEAT	Optional	Typically included if there are chirality errors
COMPND	Mandatory	
SOURCE	Mandatory	
KEYWDS	Mandatory	
EXPDTA	Mandatory	
AUTHOR	Mandatory	
REVDAT	Mandatory	
SPRSDE	Optional	Mandatory if a replacement entry.
JRNL	Optional	Mandatory if a publication Describes the experiment.
REMARK 1	Optional	
REMARK 2	Mandatory	
REMARK 3	Mandatory	
REMARK N	Optional	Mandatory under certain conditions, as noted in the remark descriptions.
DBREF	Optional	Mandatory for each peptide chain with a length greater than ten (10) residues, and for nucleic acid entries that exist in the NDB.
SEQADV	Optional	Mandatory if sequence conflict exists.
SEQRES	Optional	Mandatory if ATOM records exist.
MODRES	Optional	Mandatory if modified group exists within the coordinates.
HET	Optional	Mandatory if non-standard group other

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

MASTER	Mandatory
END	Mandatory

Sections of an Entry

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

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

Field Formats

Each record type is presented in a table which contains the division of the records into fields by column number, defined data type, field name or a quoted string which must appear in the field,

and field definition. Any column not specified must be left blank.

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

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

String	A sequence of characters. These characters may have arbitrary spacing, but should be interpreted as directed below.
String(n)	A String with exactly n characters.
SymOP	An integer field of from 4 to 6 digits, right-justified, of the form nnnMMM where nnn is the symmetry operator number and MMM is the translation vector.

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

2. Title Section

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

HEADER

Overview

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

Record Format

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

Details

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

* Classification may be based on function, metabolic role, molecule type, cellular location, etc. In the case of a molecule having a dual function, both may be presented here.

Verification/Validation/Value Authority Control

The verification program checks that the deposition date is a legitimate date and that the ID code is well-formed.

PDB coordinate entry ID codes do not begin with 0, as this is used to identify the NOC (“no coordinates) files that are bibliographic only, not structural entries.

Relationships to Other Record Types

The classification found in HEADER also appears in KEYWDS, unabbreviated and in no strict order.

Example

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
HEADER    MUSCLE PROTEIN                          02-JUN-93   1MYS

HEADER    HYDROLASE (CARBOXYLIC ESTER)            08-APR-93   2PHI

HEADER    COMPLEX (LECTIN/TRANSFERRIN)            07-JAN-94   1LGB
```

OBSLTE

Overview

OBSLTE appears in entries that have been withdrawn from distribution.

This record acts as a flag in an entry that has been withdrawn from the PDB's full release. It indicates which, if any, new entries have replaced the withdrawn entry. The format allows for the case of multiple new entries replacing one existing entry.

Record Format

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

Details

It is PDB policy that only the primary author who submitted an entry has the authority to obsolete

it. All OBSLTE entries are available from the PDB archive.

Verification/Validation/Value Authority Control

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

Relationships to Other Record Types

None.

Example

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

TITLE

Overview

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

Record Format

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

Details

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

* Some items that may be included in TITLE are:

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

Verification/Validation/Value Authority Control

This record is free text so no verification of format is required. The title is supplied by the

depositor, but PDB staff may exercise editorial judgment in consultation with depositors in assigning the title.

Relationships to Other Record Types

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

Example

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
TITLE     RHIZOPUSPEPSIN COMPLEXED WITH REDUCED PEPTIDE INHIBITOR

TITLE     BETA-GLUCOSYLTRANSFERASE, ALPHA CARBON COORDINATES ONLY

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

CAVEAT

Overview

CAVEAT warns of chirality errors in an entry.

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

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

Verification/Validation/Value Authority Control

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

COMPND

Overview

The COMPND record describes the macromolecular contents of an entry. Each macromolecule found in the entry is described by a set of token: value pairs, and is referred to as a COMPND record component. Since the concept of a molecule is difficult to specify exactly, PDB staff may exercise editorial judgment in consultation with depositors in assigning these names.

Record Format

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

Details

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

TOKEN	VALUE DEFINITION
MOL_ID	Numbers each component; also used in SOURCE to associate the information.
MOLECULE	Name of the macromolecule.
CHAIN	Comma-separated list of chain identifier(s).
FRAGMENT	Specifies a domain or region of the molecule.
SYNONYM	Comma-separated list of synonyms for the MOLECULE.
EC	The Enzyme Commission number associated with the molecule. If there is more than one EC number, they are presented as a comma-separated list.
ENGINEERED	Indicates that the molecule was produced using recombinant technology or by purely chemical synthesis.
MUTATION	Indicates if there is a mutation.
OTHER_DETAILS	Additional comments.

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

* In the case of synthetic molecules, e. g., hybrids, the depositor will provide the description.

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

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

* When insertion codes are given as part of the residue name, they must be given within square brackets, i.e., H57[A]N. This might occur when listing residues in FRAGMENT or OTHER_DETAILS.

* For multi-chain molecules, e.g., the hemoglobin tetramer, a comma-separated list of CHAIN identifiers is used.

* When non-blank chain identifiers occur in the entry, they must be specified.

Verification/Validation/Value Authority Control

CHAIN must match the chain identifiers(s) of the molecule(s). EC numbers are also checked

Relationships to Other Record Types

In the case of mutations, the SEQADV records will present differences from the reference molecule.
REMARK records may further describe the contents of the entry. Also see verification above.

Example

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
COMPND    MOL_ID: 1;
COMPND   2 MOLECULE: HEMOGLOBIN;
COMPND   3 CHAIN: A, B, C, D;
COMPND   4 ENGINEERED: YES;
COMPND   5 MUTATION: YES
COMPND   6 OTHER_DETAILS: DEOXY FORM

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

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

SOURCE

Overview

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

Record Format

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

Details

TOKEN	VALUE DEFINITION
MOL_ID	Numbers each molecule. Same as appears in COMPND.
SYNTHETIC	Indicates a chemically-synthesized source.
FRAGMENT	A domain or fragment of the molecule may be specified.
ORGANISM_SCIENTIFIC	Scientific name of the organism.
ORGANISM_COMMON	Common name of the organism.
STRAIN	Identifies the strain.
VARIANT	Identifies the variant.
CELL_LINE	The specific line of cells used in the experiment.
ATCC	American Type Culture Collection tissue culture number.
ORGAN	Organized group of tissues that carries on a specialized function.
TISSUE	Organized group of cells with a common function and structure.
CELL	Identifies the particular cell type.
ORGANELLE	Organized structure within a cell.
SECRETION	Identifies the secretion, such as saliva,

	urine, or venom, from which the molecule was isolated.
CELLULAR_LOCATION	Identifies the location inside (or outside) the cell.
PLASMID	Identifies the plasmid containing the gene.
GENE	Identifies the gene.
EXPRESSION_SYSTEM	System used to express recombinant macromolecules.
EXPRESSION_SYSTEM_STRAIN	Strain of the organism in which the molecule was expressed.
EXPRESSION_SYSTEM_VARIANT	Variant of the organism used as the expression system.
EXPRESSION_SYSTEM_CELL_LINE	The specific line of cells used as the expression system.
EXPRESSION_SYSTEM_ATCC_NUMBER	Identifies the ATCC number of the expression system
EXPRESSION_SYSTEM_ORGAN	Specific organ which expressed the molecule.
EXPRESSION_SYSTEM_TISSUE	Specific tissue which expressed the molecule.
EXPRESSION_SYSTEM_CELL	Specific cell type which expressed the molecule.
EXPRESSION_SYSTEM_ORGANELLE	Specific organelle which expressed the molecule.
EXPRESSION_SYSTEM_CELLULAR_LOCATION	Identifies the location inside or outside the cell which expressed the molecule.
EXPRESSION_SYSTEM_VECTOR_TYPE	Identifies the type of vector used, i.e., plasmid, virus, or cosmid.
EXPRESSION_SYSTEM_VECTOR	Identifies the vector used.
EXPRESSION_SYSTEM_PLASMID	Plasmid used in the recombinant experiment.
EXPRESSION_SYSTEM_GENE	Name of the gene used in recombinant experiment.
OTHER_DETAILS	Used to present information on the source which is not given elsewhere.

* The srcName is a list of token: value pairs describing each biological component of the entry.

* As in COMPND, the order is not specified except that MOL_ID or FRAGMENT indicates subsequent specifications are related to that molecule or fragment of the molecule.

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

* Only the relevant tokens need to appear in an entry.

* 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. ENGINEERED must appear in the COMPND record.

* In the case of a chemically synthesized molecule using a biologically functional sequence (nucleic or amino acid), SOURCE reflects the biological origin of the sequence and COMPND reflects its synthetic nature by inclusion of the token ENGINEERED. The token SYNTHETIC appears in SOURCE.

* If made from a synthetic gene, ENGINEERED appears in COMPND and the expression system is described in SOURCE (SYNTHETIC does NOT appear in SOURCE).

* If the molecule was made using recombinant techniques, ENGINEERED appears in COMPND and the system is described in SOURCE.

* When multiple macromolecules appear in the entry, each MOL_ID, as given in the COMPND record, must be repeated in the SOURCE record along with the source information for the corresponding molecule.

* Hybrid molecules prepared by fusion of genes are treated as multi-molecular systems for the purpose of specifying the source. The token FRAGMENT is used to associate the source with its corresponding fragment.

- When necessary to fully describe hybrid molecules, tokens may appear more than once for a given MOL_ID.

- All relevant token: value pairs that taken together fully describe each fragment are grouped following the appropriate FRAGMENT.

- Descriptors relative to the full system appear before the FRAGMENT (see Example 3 below).

* ORGANISM_SCIENTIFIC provides the Latin genus and species. Virus names are listed as the scientific name.

* 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.

* For a theoretical modeling experiment, SOURCE describes the modelled compound just as though it were an experimental study.

* New tokens may be added by the PDB.

Verification/Validation/Value Authority Control

The biological source is compared to that found in the sequence databases.

Relationships to Other Record Types

Each macromolecule listed in COMPND must have a corresponding source.

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

KEYWDS

Overview

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

Record Format

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

Details

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

- Functional classification.

- Metabolic role.

- Known biological or chemical activity.

- Structural classification.

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

*Note that the terms in the KEYWDS record duplicate those found in the classification field of the HEADER record. Terms abbreviated in the HEADER record are unabbreviated in KEYWDS, and the parentheses used in HEADER are optional in KEYWDS.

Verification/Validation/Value Authority Control

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

Relationships to Other Record Types

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

Example

```
         1         2         3         4         5         6         7
12345678901234567890123456789012345678901234567890123456789012345678901234567890
KEYWDS    LYASE, TRICARBOXYLIC ACID CYCLE, MITOCHONDRION, OXIDATIVE
KEYWDS   2 METABOLISM
```

EXPDTA

Overview

The EXPDTA record presents information about the experiment.

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

ELECTRON DIFFRACTION
ELECTRON MICROSCOPY
CRYO-ELECTRON MICROSCOPY
SOLUTION SCATTERING, THEORETICAL MODEL
FIBER DIFFRACTION
FLUORESCENCE TRANSFER
NEUTRON DIFFRACTION
NMR (may have a qualifier e.g. number of models see examples below)
SOLUTION SCATTERING
THEORETICAL MODEL*
X-RAY DIFFRACTION

**Note: As of July 1, 2002, models are available from a directory separate from the main archive at* ftp://ftp.rcsb.org/pub/pdb/data/structures/models/current/. *As of October 15, 2006, theoretical models are no longer accepted for deposition.*

Record Format

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

Details

* EXPDTA is mandatory and appears in all entries.

* The technique must match one of the permitted values. See above.

* If more than one model appears in the entry, the number of models included must be stated.

* If only one model appears in the entry, its significance must be stated, such as it being a

minimized average or regularized mean structure.

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

Verification/Validation/Value Authority Control

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

Relationships to Other Record Types

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

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

Example

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

EXPDTA    NEUTRON DIFFRACTION; X-RAY DIFFRACTION

EXPDTA    NMR, 32 STRUCTURES

EXPDTA    NMR, REGULARIZED MEAN STRUCTURE

EXPDTA    FIBER DIFFRACTION
```

AUTHOR

Overview

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

Record Format

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

Details

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

* Representation of personal names:
  - First and middle names are indicated by initials, each followed by a period, and precede the surname.

  - Only the surname (family or last name) of the author is given in full.

  - Hyphens can be used if they are part of the author's name.

  - Apostrophes are allowed in surnames.

  - Umlauts and other character modifiers are not given.

* Structure of personal names:

  - There is no space after any initial and its following period.

  - Blank spaces are used in a name only if properly part of the surname (e.g., J.VAN DORN), or between surname and Junior, II, or III.

  - Abbreviations that are part of a surname, such as St. or Ste., are followed by a period and a space before the next part of the surname.

* Representation of corporate names:

  - Group names used for one or all of the authors should be spelled out in full.

  - The name of the larger group comes before the name of a subdivision, e.g., University of Somewhere Department of Chemistry.

* Structure of list:

  - Line breaks between multiple lines in the authorList occur only after a comma.

  - Personal names are not split across two lines.

* Special cases:

  - Names are given in English if there is an accepted English version; otherwise in the native language, transliterated if necessary.

  - "ET AL." may be used when all authors are not individually listed.

Verification/Validation/Value Authority Control

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

Relationships to Other Record Types

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

Example

```
         1         2         3         4         5         6         7
12345678901234567890123456789012345678901234567890123456789012345678901234567890
AUTHOR    M.B.BERRY,B.MEADOR,T.BILDERBACK,P.LIANG,M.GLASER,
AUTHOR   2 G.N.PHILLIPS JUNIOR,T.L.ST. STEVENS
```

REVDAT

Overview

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

Record Format

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

Details

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

0	Initial released entry.
1	Miscellaneous - mostly typographical.
2	Modification of a CONECT record.
3	Modification to coordinates or transformations.

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

Verification/Validation/Value Authority Control

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

Relationships to Other Record Types

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

Example

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

SPRSDE

Overview

The SPRSDE records contain a list of the ID codes of entries that were made obsolete by the given coordinate entry and withdrawn from the PDB release set. One entry may replace many.

It is PDB policy that only the principal investigator of a structure has the authority to withdraw it.

Record Format

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"SPRSDE"	
9 - 10	Continuation	continuation	Allows for multiple ID codes.
12 - 20	Date	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 ID code list is terminated by the first blank sIDcode field.

Verification/Validation/Value Authority Control

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

Relationships to Other Record Types

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

Example

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


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

JRNL

Overview

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

Record Format

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

Details

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

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

1. AUTH

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

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

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

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

```
COLUMNS        DATA TYPE       FIELD          DEFINITION
--------------------------------------------------------------------------------
 1 -  6        Record name     "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.
```

2. TITL

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

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

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

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

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

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

3. EDIT

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

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

4. REF

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

4a. If the reference has not been published yet, 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"	

* Publication name (first item in pubName field):

- If the publication is a serial (i.e., a journal, an annual, or other non-book or non-monographic item issued in parts and intended to be continued indefinitely), use the abbreviated name of the publication as listed in PubMed and with periods.

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

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

* Supplement (follows publication name in pubName field):

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

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

be abbreviated as "PT.", and the number or letter should follow without an intervening space.

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

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

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

* Reconstruction of publication name:

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

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

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

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

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"JRNL "	
13 - 16	LString(3)	"REF"	
17 - 18	Continuation continuation		Allows long publication names.
20 - 47	LString	pubName	Name of the publication including section or series designation. This is the only field of this sub-record which may be continued on successive sub-records.
50 - 51	LString(2)	"V."	Appears in the first sub-record only, and only if column 55 is non-blank.

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

5. PUBL

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

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

* Place of publication:

- Give the place of publication. If the name of the country, state, province, etc. is considered necessary to distinguish the place of publication from others of the same name, or for identification, then follow the city with a comma, a space, and the name of the larger geographic area.

- If there is more than one place of publication, only the first listed will be used. If an online catalog record is used to verify the item, the first place listed there will be used, omitting any brackets. Preference will be given to the cataloging done by the Library of Congress, the National Library of Medicine, and the British Library, in that order.

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

- Give the name of the publisher in the shortest form in which it can be understood and identified internationally, according to AACR2R rule 1.4D.

- If there is more than one publisher listed in the publication, only the first will be used in the PDB file. If an online catalog record is used to verify the item, the first place listed there will be used for the name of the publisher. Preference will be given to the cataloging of the Library of Congress, the National Library of Medicine, and the British Library, in that order.

* Ph.D. and other theses:

- Theses are presented in the PUBL record if the degree has been granted and the thesis made available for public consultation by the degree-granting institution.

- The name of the degree-granting institution (the issuing agency) is followed by a space and "(THESIS)".

* Reconstruction of place and publisher:

- The PUBL sub-record type can be reconstructed by removing all trailing blanks in the pub field and concatenating all of the pub fields from the continuation lines with an intervening space.

Continued lines do not begin with a space.

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

6. REFN

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

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

* If more than one ISBN is known, select one that matches the individual volume cited (if it happens to be in a set that also has an ISBN for the set). If the reason for multiple ISBNs is that the publication is issued in more than one country, use the ISBN for the country of the first listed place of publication. If there are hardcover and paperback ISBN numbers, use the ISBN for the hardbound version.

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

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

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"JRNL "	
13 - 16	LString(4)	"REFN"	

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

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"JRNL "	
13 - 16	LString(4)	"REFN"	
20 - 23	LString(4)	"ASTM"	
25 - 30	LString(6)	astm	ASTM devised coden.
33 - 34	LString(2)	country	Country of publication code as defined in the OCLC/MARC cataloging format (optional).
36 - 39	LString(4)	"ISBN" "ISSN" or “ESSN”	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).

Verification/Validation/Value Authority Control

PDB verifies that this record is correctly formatted.

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

Relationships to Other Record Types

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

Example

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
JRNL        AUTH   G.FERMI,M.F.PERUTZ,B.SHAANAN,R.FOURME
JRNL        TITL   THE CRYSTAL STRUCTURE OF HUMAN DEOXYHAEMOGLOBIN AT
JRNL        TITL 2 1.74 A RESOLUTION
JRNL        REF    J.MOL.BIOL.                   V. 175   159 1984
JRNL        REFN   ASTM JMOBAK  UK ISSN 0022-2836
```

Known Problems

* Interchange of bibliographic information and linking with other databases is hampered by the lack of labels or specific locations for certain types of information or by more than one type of information being in a particular location. This is most likely to occur with books, series, and

reports. Some of the points below provide details about the variations and/or blending of information.

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

* Information about a supplement and its number/letter is presented in the publication's title field (on the REF lines in columns 20 - 47).

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

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

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

* Many books are issued in more than one country. The publisher has a separate ISBN number in each country. There is no place to put any additional applicable ISBN numbers.

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

* Pagination is limited to the beginning page.

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

REMARK

Overview

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

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

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

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

REMARK 1

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

Record Format and Details

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

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

1. REFERENCE

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

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

2. AUTH

AUTH contains the list of authors of the reference.

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

See JRNL AUTH for details.

3. TITL

TITL specifies the title of the reference.

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

See JRNL TITL for details.

4. EDIT

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

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

See JRNL EDIT for details.

5. REF

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

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

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

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

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

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

See JRNL REF for details.

6. PUBL

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

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

See JRNL PUBL for details.

7. REFN

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

7a. If the citation hasnot 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"	

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

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"REMARK"	
10	LString(1)	"1"	
13 - 16	LString(4)	"REFN"	
20 - 23	LString(4)	"ASTM"	Blank if reference is not serialized.
25 - 30	LString	astm	Code from the ASTM file.
33 - 34	LString	country	2-digit abbreviation for country of publication.
36 - 39	LString(4)	"ISBN" "ISSN" or “ESSN”	
41 - 65	LString	isbn	ISSN or ISBN number.

See JRNL REFN for details.

Verification/Validation/Value Authority Control

PDB verifies that this record is correctly formatted.

Citations appearing in REMARK 1 may not appear in JRNL.

Relationships to Other Record Types

Citations appearing in REMARK 1 may not appear in JRNL.

Example

```
         1         2         3         4         5         6         7
12345678901234567890123456789012345678901234567890123456789012345678901234567890
REMARK   1
REMARK   1 REFERENCE 1
REMARK   1  AUTH   A.M.BONVIN,J.A.RULLMANN,R.M.LAMERICHS,R.BOELENS,
REMARK   1  AUTH 2 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
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
REMARK   1 REFERENCE 3
REMARK   1  AUTH   R.M.J.M.LAMERICHS
REMARK   1  REF    2D NMR STUDIES OF                          1989
REMARK   1  REF  2 BIOMOLECULES: PROTEIN
REMARK   1  REF  3 STRUCTURE AND PROTEIN-DNA
REMARK   1  REF  4 INTERACTIONS
REMARK   1  PUBL   UTRECHT : UNIVERSITY OF UTRECHT (THESIS)
REMARK   1  REFN               NE
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
```

Known Problems

See JRNL for a listing of problems associated with references.

REMARK 2

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

Record Format and Details

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

* For diffraction experiments:

```
COLUMNS              DATA TYPE            FIELD                            DEFINITION
-------------------------------------------------------------------------------
 1 - 6               Record name          "REMARK"

10                   LString(1)               "2"

12 - 22              LString(11)          "RESOLUTION."

23 - 27              Real(5.2)            resolution                       Resolution.

29 - 38              LString(10)          "ANGSTROMS."
```

* REMARK 2 when not a diffraction experiment:

```
COLUMNS              DATA TYPE            FIELD                           DEFINITION
-------------------------------------------------------------------------------
 1 - 6               Record name          "REMARK"

10                   LString(1)           "2"

12 - 38              LString(28)    "RESOLUTION. NOT APPLICABLE."

41 - 70              String               comment                         Comment.
```

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

```
COLUMNS       DATA TYPE            FIELD             DEFINITION
-------------------------------------------------------------------------------
 1 - 6       Record name          "REMARK"

10            LString(1)      "2"
```

```
12 - 22        LString(11)     "RESOLUTION."


24 - 70        String          comment             Comment.
```

Example

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
REMARK   2
REMARK   2 RESOLUTION. 1.74 ANGSTROMS.
REMARK   2
REMARK   2 RESOLUTION. NOT APPLICABLE.

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

REMARK 3

Overview

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

If more than one refinement package was used, they may be named in "OTHER REFINEMENT REMARKS". However, Remark 3 statistics are given for the final refinement run.

The format of this remark changes with the evolution of refinement software. Selected representative templates or examples are provided here.

Details

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

Refinement using X-PLOR

Template

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

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

Refinement using CNS

Template

```
REMARK   3
REMARK   3 REFINEMENT.
REMARK   3   PROGRAM     : CNS
REMARK   3   AUTHORS     : BRUNGER,ADAMS,CLORE,DELANO,GROS,GROSSE-
REMARK   3               : KUNSTLEVE,JIANG,KUSZEWSKI,NILGES, PANNU,
REMARK   3               : READ,RICE,SIMONSON,WARREN
REMARK   3
REMARK   3  REFINEMENT TARGET :
REMARK   3
REMARK   3  DATA USED IN REFINEMENT.
REMARK   3   RESOLUTION RANGE HIGH (ANGSTROMS) :
REMARK   3   RESOLUTION RANGE LOW  (ANGSTROMS) :
REMARK   3   DATA CUTOFF            (SIGMA(F)) :
REMARK   3   DATA CUTOFF HIGH         (ABS(F)) :
REMARK   3   DATA CUTOFF LOW          (ABS(F)) :
REMARK   3   OUTLIER CUTOFF HIGH (RMS(ABS(F))) :
REMARK   3   COMPLETENESS FOR RANGE        (%) :
REMARK   3   COMPLETENESS (WORKING+TEST)   (%) :
REMARK   3   NUMBER OF REFLECTIONS             :
REMARK   3
REMARK   3  FIT TO DATA USED IN REFINEMENT.
REMARK   3   CROSS-VALIDATION METHOD          :
REMARK   3   FREE R VALUE TEST SET SELECTION  :
REMARK   3   R VALUE            (WORKING SET) :
REMARK   3   FREE R VALUE                     :
REMARK   3   FREE R VALUE TEST SET SIZE   (%) :
REMARK   3   FREE R VALUE TEST SET COUNT      :
REMARK   3   ESTIMATED ERROR OF FREE R VALUE  :
REMARK   3
REMARK   3  FIT IN THE HIGHEST RESOLUTION BIN.
REMARK   3   TOTAL NUMBER OF BINS USED           :
REMARK   3   BIN RESOLUTION RANGE HIGH       (A) :
REMARK   3   BIN RESOLUTION RANGE LOW        (A) :
REMARK   3   BIN COMPLETENESS (WORKING+TEST) (%) :
REMARK   3   REFLECTIONS IN BIN    (WORKING SET) :
REMARK   3   BIN R VALUE           (WORKING SET) :
REMARK   3   BIN FREE R VALUE                    :
REMARK   3   BIN FREE R VALUE TEST SET SIZE  (%) :
REMARK   3   BIN FREE R VALUE TEST SET COUNT     :
REMARK   3   ESTIMATED ERROR OF BIN FREE R VALUE :
REMARK   3
REMARK   3  NUMBER OF NON-HYDROGEN ATOMS USED IN REFINEMENT.
REMARK   3   PROTEIN ATOMS            :
REMARK   3   NUCLEIC ACID ATOMS       :
REMARK   3   HETEROGEN ATOMS          :
REMARK   3   SOLVENT ATOMS            :
REMARK   3
REMARK   3  B VALUES.
REMARK   3   FROM WILSON PLOT           (A**2) :
REMARK   3   MEAN B VALUE      (OVERALL, A**2) :
REMARK   3   OVERALL ANISOTROPIC B VALUE.
REMARK   3    B11 (A**2) :
REMARK   3    B22 (A**2) :
REMARK   3    B33 (A**2) :
REMARK   3    B12 (A**2) :
REMARK   3    B13 (A**2) :
```

```
REMARK   3    B23 (A**2) :
REMARK   3
REMARK   3  ESTIMATED COORDINATE ERROR.
REMARK   3   ESD FROM LUZZATI PLOT        (A) :
REMARK   3   ESD FROM SIGMAA              (A) :
REMARK   3   LOW RESOLUTION CUTOFF        (A) :
REMARK   3
REMARK   3  CROSS-VALIDATED ESTIMATED COORDINATE ERROR.
REMARK   3   ESD FROM C-V LUZZATI PLOT    (A) :
REMARK   3   ESD FROM C-V SIGMAA          (A) :
REMARK   3
REMARK   3  RMS DEVIATIONS FROM IDEAL VALUES.
REMARK   3   BOND LENGTHS                 (A) :
REMARK   3   BOND ANGLES            (DEGREES) :
REMARK   3   DIHEDRAL ANGLES        (DEGREES) :
REMARK   3   IMPROPER ANGLES        (DEGREES) :
REMARK   3
REMARK   3  ISOTROPIC THERMAL MODEL :
REMARK   3
REMARK   3  ISOTROPIC THERMAL FACTOR RESTRAINTS.    RMS    SIGMA
REMARK   3   MAIN-CHAIN BOND              (A**2) :       ;
REMARK   3   MAIN-CHAIN ANGLE             (A**2) :       ;
REMARK   3   SIDE-CHAIN BOND              (A**2) :       ;
REMARK   3   SIDE-CHAIN ANGLE             (A**2) :       ;
REMARK   3
REMARK   3  BULK SOLVENT MODELING.
REMARK   3   METHOD USED :
REMARK   3   KSOL        :
REMARK   3   BSOL        :
REMARK   3
REMARK   3  NCS MODEL :
REMARK   3
REMARK   3  NCS RESTRAINTS.                         RMS   SIGMA/WEIGHT
REMARK   3   GROUP     POSITIONAL           (A) :       ;
REMARK   3   GROUP     B-FACTOR          (A**2) :       ;
REMARK   3
REMARK   3  PARAMETER FILE     :
REMARK   3  TOPOLOGY FILE      :
REMARK   3
REMARK   3  OTHER REFINEMENT REMARKS:
```

Refinement using CNX

Template

```
REMARK   3
REMARK   3 REFINEMENT.
REMARK   3   PROGRAM     : CNX
REMARK   3   AUTHORS     : BRUNGER,ADAMS,CLORE,DELANO,GROS,GROSSE-
REMARK   3               : KUNSTLEVE,JIANG,KUSZEWSKI,NILGES, PANNU,
REMARK   3               : READ,RICE,SIMONSON,WARREN
REMARK   3
REMARK   3  DATA USED IN REFINEMENT.
REMARK   3   RESOLUTION RANGE HIGH (ANGSTROMS) :
REMARK   3   RESOLUTION RANGE LOW  (ANGSTROMS) :
REMARK   3   DATA CUTOFF            (SIGMA(F)) :
REMARK   3   DATA CUTOFF HIGH         (ABS(F)) :
REMARK   3   DATA CUTOFF LOW          (ABS(F)) :
REMARK   3   COMPLETENESS (WORKING+TEST)   (%) :
REMARK   3   NUMBER OF REFLECTIONS             :
REMARK   3
REMARK   3  FIT TO DATA USED IN REFINEMENT.
REMARK   3   CROSS-VALIDATION METHOD          :
REMARK   3   FREE R VALUE TEST SET SELECTION  :
REMARK   3   R VALUE     (WORKING + TEST SET) :
REMARK   3   R VALUE           (WORKING SET) :
REMARK   3   FREE R VALUE                     :
REMARK   3   FREE R VALUE TEST SET SIZE   (%) :
REMARK   3   FREE R VALUE TEST SET COUNT      :
REMARK   3   ESTIMATED ERROR OF FREE R VALUE  :
REMARK   3
REMARK   3  FIT/AGREEMENT OF MODEL WITH ALL DATA.
REMARK   3   R VALUE     (WORKING + TEST SET, NO CUTOFF) :
REMARK   3   R VALUE           (WORKING SET, NO CUTOFF) :
REMARK   3   FREE R VALUE                    (NO CUTOFF) :
REMARK   3   FREE R VALUE TEST SET SIZE   (%, NO CUTOFF) :
REMARK   3   FREE R VALUE TEST SET COUNT     (NO CUTOFF) :
REMARK   3   ESTIMATED ERROR OF FREE R VALUE (NO CUTOFF) :
REMARK   3   TOTAL NUMBER OF REFLECTIONS     (NO CUTOFF) :
REMARK   3
REMARK   3  FIT IN THE HIGHEST RESOLUTION BIN.
REMARK   3   TOTAL NUMBER OF BINS USED           :
REMARK   3   BIN RESOLUTION RANGE HIGH       (A) :
REMARK   3   BIN RESOLUTION RANGE LOW        (A) :
REMARK   3   BIN COMPLETENESS (WORKING+TEST) (%) :
REMARK   3   REFLECTIONS IN BIN    (WORKING SET) :
REMARK   3   BIN R VALUE           (WORKING SET) :
REMARK   3   BIN FREE R VALUE                    :
REMARK   3   BIN FREE R VALUE TEST SET SIZE  (%) :
REMARK   3   BIN FREE R VALUE TEST SET COUNT     :
REMARK   3   ESTIMATED ERROR OF BIN FREE R VALUE :
REMARK   3
REMARK   3  NUMBER OF NON-HYDROGEN ATOMS USED IN REFINEMENT.
REMARK   3   PROTEIN ATOMS            :
REMARK   3   NUCLEIC ACID ATOMS       :
REMARK   3   HETEROGEN ATOMS          :
REMARK   3   SOLVENT ATOMS            :
REMARK   3
REMARK   3  B VALUES.
REMARK   3   FROM WILSON PLOT           (A**2) :
REMARK   3   MEAN B VALUE      (OVERALL, A**2) :
```

```
REMARK   3   OVERALL ANISOTROPIC B VALUE.
REMARK   3    B11 (A**2) :
REMARK   3    B22 (A**2) :
REMARK   3    B33 (A**2) :
REMARK   3    B12 (A**2) :
REMARK   3    B13 (A**2) :
REMARK   3    B23 (A**2) :
REMARK   3
REMARK   3  ESTIMATED COORDINATE ERROR.
REMARK   3   ESD FROM LUZZATI PLOT        (A) :
REMARK   3   ESD FROM SIGMAA              (A) :
REMARK   3   LOW RESOLUTION CUTOFF        (A) :
REMARK   3
REMARK   3  CROSS-VALIDATED ESTIMATED COORDINATE ERROR.
REMARK   3   ESD FROM C-V LUZZATI PLOT    (A) :
REMARK   3   ESD FROM C-V SIGMAA          (A) :
REMARK   3
REMARK   3  RMS DEVIATIONS FROM IDEAL VALUES.
REMARK   3   BOND LENGTHS                 (A) :
REMARK   3   BOND ANGLES            (DEGREES) :
REMARK   3   DIHEDRAL ANGLES        (DEGREES) :
REMARK   3   IMPROPER ANGLES        (DEGREES) :
REMARK   3
REMARK   3  ISOTROPIC THERMAL MODEL :
REMARK   3
REMARK   3  ISOTROPIC THERMAL FACTOR RESTRAINTS.    RMS    SIGMA
REMARK   3   MAIN-CHAIN BOND              (A**2) :       ;
REMARK   3   MAIN-CHAIN ANGLE             (A**2) :       ;
REMARK   3   SIDE-CHAIN BOND              (A**2) :       ;
REMARK   3   SIDE-CHAIN ANGLE             (A**2) :       ;
REMARK   3
REMARK   3  NCS MODEL :
REMARK   3
REMARK   3  NCS RESTRAINTS.                         RMS   SIGMA/WEIGHT
REMARK   3   GROUP     POSITIONAL           (A) :       ;
REMARK   3   GROUP     B-FACTOR          (A**2) :       ;
REMARK   3
REMARK   3  PARAMETER FILE     :
REMARK   3  TOPOLOGY FILE      :
REMARK   3
REMARK   3  OTHER REFINEMENT REMARKS:
```

Refinement using REFMAC

Template

```
REMARK   3
REMARK   3 REFINEMENT.
REMARK   3   PROGRAM     : REFMAC 5.X
REMARK   3   AUTHORS     : MURSHUDOV,VAGIN,DODSON
REMARK   3
REMARK   3    REFINEMENT TARGET :
REMARK   3
REMARK   3  DATA USED IN REFINEMENT.
REMARK   3   RESOLUTION RANGE HIGH (ANGSTROMS) :
REMARK   3   RESOLUTION RANGE LOW  (ANGSTROMS) :
REMARK   3   DATA CUTOFF            (SIGMA(F)) :
REMARK   3   COMPLETENESS FOR RANGE        (%) :
REMARK   3   NUMBER OF REFLECTIONS             :
REMARK   3
REMARK   3  FIT TO DATA USED IN REFINEMENT.
REMARK   3   CROSS-VALIDATION METHOD          :
REMARK   3   FREE R VALUE TEST SET SELECTION  :
REMARK   3   R VALUE     (WORKING + TEST SET) :
REMARK   3   R VALUE           (WORKING SET) :
REMARK   3   FREE R VALUE                     :
REMARK   3   FREE R VALUE TEST SET SIZE   (%) :
REMARK   3   FREE R VALUE TEST SET COUNT      :
REMARK   3
REMARK   3  FIT IN THE HIGHEST RESOLUTION BIN.
REMARK   3   TOTAL NUMBER OF BINS USED           :
REMARK   3   BIN RESOLUTION RANGE HIGH           :
REMARK   3   BIN RESOLUTION RANGE LOW            :
REMARK   3   REFLECTION IN BIN     (WORKING SET) :
REMARK   3   BIN COMPLETENESS (WORKING+TEST) (%) :
REMARK   3   BIN R VALUE           (WORKING SET) :
REMARK   3   BIN FREE R VALUE SET COUNT          :
REMARK   3   BIN FREE R VALUE                    :
REMARK   3
REMARK   3  NUMBER OF NON-HYDROGEN ATOMS USED IN REFINEMENT.
REMARK   3   ALL ATOMS                :
REMARK   3
REMARK   3  NUMBER OF NON-HYDROGEN ATOMS USED IN REFINEMENT.
REMARK   3   PROTEIN ATOMS            :
REMARK   3   NUCLEIC ACID ATOMS       :
REMARK   3   HETEROGEN ATOMS          :
REMARK   3   SOLVENT ATOMS            :
REMARK   3
REMARK   3  B VALUES.
REMARK   3   FROM WILSON PLOT           (A**2) :
REMARK   3   MEAN B VALUE      (OVERALL, A**2) :
REMARK   3   OVERALL ANISOTROPIC B VALUE.
REMARK   3    B11 (A**2) :
REMARK   3    B22 (A**2) :
REMARK   3    B33 (A**2) :
REMARK   3    B12 (A**2) :
REMARK   3    B13 (A**2) :
REMARK   3    B23 (A**2) :
REMARK   3
REMARK   3  ESTIMATED OVERALL COORDINATE ERROR.
REMARK   3   ESU BASED ON R VALUE                            (A):
REMARK   3   ESU BASED ON FREE R VALUE                       (A):
```

```
REMARK   3   ESU BASED ON MAXIMUM LIKELIHOOD                 (A):
REMARK   3   ESU FOR B VALUES BASED ON MAXIMUM LIKELIHOOD (A**2):
REMARK   3
REMARK   3 CORRELATION COEFFICIENTS.
REMARK   3   CORRELATION COEFFICIENT FO-FC      :
REMARK   3   CORRELATION COEFFICIENT FO-FC FREE :
REMARK   3
REMARK   3  RMS DEVIATIONS FROM IDEAL VALUES        COUNT    RMS    WEIGHT
REMARK   3   BOND LENGTHS REFINED ATOMS        (A):       ;       ;
REMARK   3   BOND LENGTHS OTHERS               (A):       ;       ;
REMARK   3   BOND ANGLES REFINED ATOMS   (DEGREES):       ;       ;
REMARK   3   BOND ANGLES OTHERS          (DEGREES):       ;       ;
REMARK   3   TORSION ANGLES, PERIOD 1    (DEGREES):       ;       ;
REMARK   3   TORSION ANGLES, PERIOD 2    (DEGREES):       ;       ;
REMARK   3   TORSION ANGLES, PERIOD 3    (DEGREES):       ;       ;
REMARK   3   TORSION ANGLES, PERIOD 4    (DEGREES):       ;       ;
REMARK   3   CHIRAL-CENTER RESTRAINTS       (A**3):       ;       ;
REMARK   3   GENERAL PLANES REFINED ATOMS      (A):       ;       ;
REMARK   3   GENERAL PLANES OTHERS             (A):       ;       ;
REMARK   3   NON-BONDED CONTACTS REFINED ATOMS (A):       ;       ;
REMARK   3   NON-BONDED CONTACTS OTHERS        (A):       ;       ;
REMARK   3   NON-BONDED TORSION REFINED ATOMS  (A):       ;       ;
REMARK   3   NON-BONDED TORSION OTHERS         (A):       ;       ;
REMARK   3   H-BOND (X...Y) REFINED ATOMS      (A):       ;       ;
REMARK   3   H-BOND (X...Y) OTHERS             (A):       ;       ;
REMARK   3   POTENTIAL METAL-ION REFINED ATOMS (A):       ;       ;
REMARK   3   POTENTIAL METAL-ION OTHERS        (A):       ;       ;
REMARK   3   SYMMETRY VDW REFINED ATOMS        (A):       ;       ;
REMARK   3   SYMMETRY VDW OTHERS               (A):       ;       ;
REMARK   3   SYMMETRY H-BOND REFINED ATOMS     (A):       ;       ;
REMARK   3   SYMMETRY H-BOND OTHERS            (A):       ;       ;
REMARK   3   SYMMETRY METAL-ION REFINED ATOMS  (A):       ;       ;
REMARK   3   SYMMETRY METAL-ION OTHERS         (A):       ;       ;
REMARK   3
REMARK   3  ISOTROPIC THERMAL FACTOR RESTRAINTS.     COUNT   RMS    WEIGHT
REMARK   3   MAIN-CHAIN BOND REFINED ATOMS  (A**2):       ;       ;
REMARK   3   MAIN-CHAIN BOND OTHER ATOMS    (A**2):       ;       ;
REMARK   3   MAIN-CHAIN ANGLE REFINED ATOMS (A**2):       ;       ;
REMARK   3   SIDE-CHAIN BOND REFINED ATOMS  (A**2):       ;       ;
REMARK   3   SIDE-CHAIN ANGLE REFINED ATOMS (A**2):       ;       ;
REMARK   3
REMARK   3 ANISOTROPIC THERMAL FACTOR RESTRAINTS.    COUNT   RMS   WEIGHT
REMARK   3   RIGID-BOND RESTRAINTS          (A**2):       ;       ;
REMARK   3   SPHERICITY; FREE ATOMS         (A**2):       ;       ;
REMARK   3   SPHERICITY; BONDED ATOMS       (A**2):       ;       ;
REMARK   3
REMARK   3  NCS RESTRAINTS STATISTICS
REMARK   3   NUMBER OF DIFFERENT NCS GROUPS :
REMARK   3
REMARK   3     CHAIN NAMES                    :
REMARK   3     NUMBER OF COMPONENTS NCS GROUP :
REMARK   3       COMPONENT C  SSSEQI  TO  C   SSSEQI   CODE
REMARK   3
REMARK   3                    GROUP CHAIN        COUNT   RMS     WEIGHT
REMARK   3   MEDIUM POSITIONAL       A    (A):        ;       ;
REMARK   3   LOOSE POSITIONAL        A    (A):        ;       ;
REMARK   3   MEDIUM THERMAL          A (A**2):        ;       ;
REMARK   3   LOOSE THERMAL           A (A**2):        ;       ;
REMARK   3
REMARK   3  TLS DETAILS
REMARK   3   NUMBER OF TLS GROUPS  :
REMARK   3
REMARK   3   TLS GROUP :
```

```
REMARK   3    NUMBER OF COMPONENTS GROUP :
REMARK   3    COMPONENTS        C SSSEQI   TO  C SSSEQI
REMARK   3    RESIDUE RANGE :
REMARK   3    ORIGIN FOR THE GROUP (A):
REMARK   3    T TENSOR
REMARK   3      T11:          T22:
REMARK   3      T33:          T12:
REMARK   3      T13:          T23:
REMARK   3    L TENSOR
REMARK   3      L11:          L22:
REMARK   3      L33:          L12:
REMARK   3      L13:          L23:
REMARK   3    S TENSOR
REMARK   3      S11:          S12:          S13:
REMARK   3      S21:          S22:          S23:
REMARK   3      S31:          S32:          S33:
REMARK   3
REMARK   3  BULK SOLVENT MODELLING.
REMARK   3   METHOD USED :
REMARK   3   PARAMETERS FOR MASK CALCULATION
REMARK   3   VDW PROBE RADIUS   :
REMARK   3   ION PROBE RADIUS   :
REMARK   3   SHRINKAGE RADIUS   :
REMARK   3
REMARK   3  OTHER REFINEMENT REMARKS:
```

Refinement using NUCLSQ

Template

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

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

Refinement using PROLSQ, CCP4, PROFFT, GPRLSA, and related programs

Template

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

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

Refinement using SHELXL

Template

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

```
REMARK   3
REMARK   3 STEREOCHEMISTRY TARGET VALUES :
REMARK   3  SPECIAL CASE:
REMARK   3
REMARK   3 OTHER REFINEMENT REMARKS:
```

Refinement using TNT/BUSTER

Template

```
REMARK   3
REMARK   3 REFINEMENT.
REMARK   3   PROGRAM     : BUSTER/TNT
REMARK   3   AUTHORS     : BLANC,ROVERSI,VONRHEIN,BRICOGNE,TRONRUD,
REMARK   3               : TEN EYCK,MATTHEWS
REMARK   3
REMARK   3  DATA USED IN REFINEMENT.
REMARK   3   RESOLUTION RANGE HIGH (ANGSTROMS) :
REMARK   3   RESOLUTION RANGE LOW  (ANGSTROMS) :
REMARK   3   DATA CUTOFF            (SIGMA(F)) :
REMARK   3   COMPLETENESS FOR RANGE        (%) :
REMARK   3   NUMBER OF REFLECTIONS             :
REMARK   3
REMARK   3  USING DATA ABOVE SIGMA CUTOFF.
REMARK   3   CROSS-VALIDATION METHOD          :
REMARK   3   FREE R VALUE TEST SET SELECTION  :
REMARK   3   R VALUE     (WORKING + TEST SET) :
REMARK   3   R VALUE           (WORKING SET) :
REMARK   3   FREE R VALUE                     :
REMARK   3   FREE R VALUE TEST SET SIZE   (%) :
REMARK   3   FREE R VALUE TEST SET COUNT      :
REMARK   3
REMARK   3  FIT IN THE HIGHEST RESOLUTION BIN.
REMARK   3   TOTAL NUMBER OF BINS USED               :
REMARK   3   BIN RESOLUTION RANGE HIGH   (ANGSTROMS) :
REMARK   3   BIN RESOLUTION RANGE LOW    (ANGSTROMS) :
REMARK   3   BIN COMPLETENESS (WORKING+TEST)     (%) :
REMARK   3   REFLECTIONS IN BIN (WORKING + TEST SET) :
REMARK   3   BIN R VALUE        (WORKING + TEST SET) :
REMARK   3   REFLECTIONS IN BIN        (WORKING SET) :
REMARK   3   BIN R VALUE               (WORKING SET) :
REMARK   3   BIN FREE R VALUE                        :
REMARK   3   BIN FREE R VALUE TEST SET SIZE      (%) :
REMARK   3   BIN FREE R VALUE TEST SET COUNT         :
REMARK   3   ESTIMATED ERROR OF FREE R VALUE         :
REMARK   3
REMARK   3  NUMBER OF NON-HYDROGEN ATOMS USED IN REFINEMENT.
REMARK   3   PROTEIN ATOMS            :
REMARK   3   NUCLEIC ACID ATOMS       :
REMARK   3   HETEROGEN ATOMS          :
REMARK   3   SOLVENT ATOMS            :
REMARK   3
REMARK   3  B VALUES.
REMARK   3   FROM WILSON PLOT           (A**2) :
REMARK   3   MEAN B VALUE      (OVERALL, A**2) :
REMARK   3   OVERALL ANISOTROPIC B VALUE.
REMARK   3    B11 (A**2) :
REMARK   3    B22 (A**2) :
REMARK   3    B33 (A**2) :
REMARK   3    B12 (A**2) :
REMARK   3    B13 (A**2) :
REMARK   3    B23 (A**2) :
REMARK   3
REMARK   3  ESTIMATED COORDINATE ERROR.
REMARK   3   ESD FROM LUZZATI PLOT        (A) :
REMARK   3
```

```
REMARK   3 CORRELATION COEFFICIENTS.
REMARK   3   CORRELATION COEFFICIENT FO-FC      :
REMARK   3   CORRELATION COEFFICIENT FO-FC FREE :
REMARK   3
REMARK   3  RMS DEVIATIONS FROM IDEAL VALUES        RMS   WEIGHT   COUNT
REMARK   3   BOND LENGTHS                 (A) :        ;       ;
REMARK   3   BOND ANGLES            (DEGREES) :        ;       ;
REMARK   3   TORSION ANGLES         (DEGREES) :        ;       ;
REMARK   3   PSEUDOROTATION ANGLES  (DEGREES) :        ;       ;
REMARK   3   TRIGONAL CARBON PLANES       (A) :        ;       ;
REMARK   3   GENERAL PLANES               (A) :        ;       ;
REMARK   3   ISOTROPIC THERMAL FACTORS (A**2) :        ;       ;
REMARK   3   NON-BONDED CONTACTS          (A) :        ;       ;
REMARK   3
REMARK   3  INCORRECT CHIRAL-CENTERS (COUNT) :
REMARK   3
REMARK   3  OTHER REFINEMENT REMARKS:
```

Refinement using Cryo-Electron Microscopy

Template/Example

```
REMARK   3
REMARK   3 REFINEMENT
REMARK   3   SOFTWARE PACKAGES      : SIMPLEX, PYPFT, EMFIT, O, XPLOR
REMARK   3   RECONSTRUCTION SCHEMA  : ICOSAHEDRAL
REMARK   3
REMARK   3 EM MAP-MODEL FITTING AND REFINEMENT
REMARK   3   PDB ENTRY                    : PDB ID 1HX6
REMARK   3   REFINEMENT SPACE             : RECIPROCAL
REMARK   3   REFINEMENT PROTOCOL          : RIGID BODY REFINEMENT
REMARK   3   REFINEMENT TARGET            : R-FACTOR
REMARK   3   OVERALL ANISOTROPIC B VALUE  : NULL
REMARK   3
REMARK   3 FITTING PROCEDURE : THE CRYSTAL STRUCTURE OF THE MAJOR COAT
REMARK   3 PROTEIN P3 (PDB FILE 1HX6) WAS PLACED INTO THE CRYO-EM
REMARK   3 DENSITY MAP. THE CAPSID PROTEIN WAS FIRST MANUALLY
REMARK   3 POSITIONED INTO THE CRYO-EM DENSITY CORRESPONDING TO
REMARK   3 POSITIONS OF THE FOUR INDEPENDENT TRIMERS IN THE
REMARK   3 ICOSAHEDRAL ASYMMETRIC UNIT. THESE POSITIONS WERE THEN
REMARK   3 REFINED BY RIGID BODY REFINEMENT IN RECIPROCAL SPACE WITH
REMARK   3 THE PROGRAM XPLOR.
REMARK   3 QUALITY OF THE FIT R-FACTOR= 0.339, CROSS-CORRELATI0N
REMARK   3 COEFFICIENT 0.915, ATOMS OUTSIDE DENSITY PER ICOSAHEDRAL
REMARK   3 ASYMMETRIC UNIT 527 (1.5%),ATOM CLASHES PER ICOSAHEDRAL
REMARK   3 ASYMMETRIC UNIT 115 (0.3%)
REMARK   3
REMARK   3 EM IMAGE RECONSTRUCTION STATISTICS
REMARK   3   NOMINAL PIXEL SIZE (ANGSTROMS)    : 3.68
REMARK   3   ACTUAL PIXEL SIZE  (ANGSTROMS)    : 3.44
REMARK   3   EFFECTIVE RESOLUTION (ANGSTROMS)  : 14.0
REMARK   3   NUMBER OF PARTICLES               : 1800
REMARK   3   CTF CORRECTION METHOD             : NULL
REMARK   3
REMARK   3 CRYO-EM RECONSTRUCTION MAGNIFICATION CALIBRATION: THE PIXEL
REMARK   3  SIZE OF THE CRYO-EM MAP WAS OBTAINED USING THE X-RAY
REMARK   3  STRUCTURE OF THE P3 TRIMER AS A REFERENCE. AFTER AN INITIAL
REMARK   3  FITTING USING THE NOMINAL PIXEL SIZE, THE P3 TRIMERS IN THE
REMARK   3  ICOSAHEDRAL ASYMMETRIC UNIT WERE GRADUALLY TRANSLATED TOWARDS
REMARK   3  THE CENTER OF THE PARTICLE UNTIL THE CRYSTALLOGRAPHIC R-FACTOR
REMARK   3  WAS MINIMISED.
REMARK   3
REMARK   3 OTHER DETAILS: THE ORIENTATIONS WERE REFINED BY THE CROSS
REMARK   3  COMMON LINES LINES METHOD (SIMPLEX) AND THE POLAR FOURIER
REMARK   3  TRANSFORM METHOD. MODEL-BASED, POLAR-FOURIER-TRANSFORM
REMARK   3  (FULLER ET AL. 1996, J.STRUC.BIOL. 116, 48-55; BAKER AND
REMARK   3  CHENG, 1996, J.STRUC.BIOL. 116, 120-130) MODEL-BASED CROSS
REMARK   3  COMMON LINES SEARCH AND REFINEMENT (CROWTHER ET AL. 1970,
REMARK   3  NATURE (LONDON) 226, 421-425; FULLER ET AL. 1996,
REMARK   3  J.STRUC.BIOL. 116, 48-55;FERLENGHI ET AL. 1998, J.MOL.BIOL.
REMARK   3  283, 71-81). THE EFFECTIVE RESOLUTION OF THE FINAL
REMARK   3  RECONSTRUCTED DENSITY WAS DETERMINED TO BE AT LEAST 25
REMARK   3  ANGSTROMS, AS MEASURED BY RANDOMLY SPLITTING THE PARTICLES
REMARK   3  INTO TWO SETS AND CALCULATING THE FOURIER SHELL CORRELATION
REMARK   3  OBTAINED FROM SEPARATE RECONSTRUCTIONS (HARAUZ AND VAN HEEL
REMARK   3  1986, OPTIK 73, 146-156). THE EIGENVALUE SPECTRUM GAVE AN
REMARK   3  INDICATION OF THE RANDOMNESS OF THE DATA THAT WAS INCLUDED
REMARK   3  IN THE RECONSTRUCTION. THE COMPLETENESS OF THE DATA WAS
```

```
REMARK   3  VERIFIED IN THAT ALL EIGENVALUES EXCEEDED 100. THE COORDINATES
REMARK   3  ARE IN THE P, Q, R FRAME IN ANGSTROM UNITS AND CORRESPOND
REMARK   3  TO ICOSAHEDRAL SYMMETRY AXES. THE ORIGIN IS CHOSEN AT THE
REMARK   3  CENTER OF THE VIRUS WITH P, Q AND R ALONG MUTUALLY
REMARK   3  PERPENDICULAR TWO-FOLD AXES OF THE ICOSAHEDRON. THEY SHOULD
REMARK   3  REMAIN IN THAT FRAME FOR THE EASE OF THE USER IN CREATING
REMARK   3  THE BIOLOGICALLY SIGNIFICANT VIRAL COMPLEX PARTICLE USING
REMARK   3  THE 60 ICOSAHEDRAL SYMMETRY OPERATORS. RESIDUES NOT VISIBLE
REMARK   3  IN THE ORIGINAL CRYSTAL STRUCTURES ARE NOT INCLUDED IN THE
REMARK   3  CRYO-EM STRUCTURE MODEL.
```

Example for Solution Scatter

```
REMARK   3
REMARK   3 REFINEMENT.
REMARK   3   PROGRAM     : INSIGHT II 98.0
REMARK   3   AUTHORS     : MSI
REMARK   3
REMARK   3  NUMBER OF NON-HYDROGEN ATOMS USED IN REFINEMENT.
REMARK   3   PROTEIN ATOMS            : 1213
REMARK   3   NUCLEIC ACID ATOMS       : 0
REMARK   3   HETEROGEN ATOMS          : 0
REMARK   3   SOLVENT ATOMS            : 0
REMARK   3
REMARK   3  OTHER REFINEMENT REMARKS: DISCOVER WAS USED FOR ENERGY
REMARK   3   MINIMISATION
```

Non-diffraction studies

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

Template

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
REMARK   3
REMARK   3 REFINEMENT.
REMARK   3   PROGRAM     :
REMARK   3   AUTHORS     :
REMARK   3
REMARK   3 FREE TEXT
```

Example

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
REMARK   3
REMARK   3 REFINEMENT.
REMARK   3   PROGRAM     : X-PLOR 3.1
REMARK   3   AUTHORS     : BRUNGER
REMARK   3
REMARK   3 STRUCTURAL STATISTICS:
REMARK   3                                    25 SA
REMARK   3                                    STRUCTURES  SAAVEMIN
REMARK   3  RMS DEVIATIONS FROM EXP. RESTRAINTS[A]
REMARK   3   NOE DISTANCE RESTRAINTS (1430)   0.0451 A       0.044 A
REMARK   3   DIHEDRAL ANGLE RESTRAINTS (130)  0.551 DEG      0.660 DEG
REMARK   3  DEVIATIONS FROM IDEAL GEOMETRY
REMARK   3   BONDS                            0.004  A       0.004 A
REMARK   3   ANGLES                           0.661 DEG      0.650 DEG
REMARK   3   IMPROPERS                       0.371 DEG      0.380 DEG
REMARK   3  X-PLOR ENERGIES (IN KCAL MOL-1)[B]
REMARK   3   ENOE                            167            158
REMARK   3   ECDIH                           2.6            3.4
REMARK   3   ENCS                            0.01           0.01
REMARK   3   EREPEL                          54             50
REMARK   3   EBOND                          36             33
REMARK   3   EANGLE                         263            256
REMARK   3   EIMPROPER                      22             23
REMARK   3   ETOTAL                         545            523
REMARK   3  ATOMIC RMS DIFFERENCES[C]
REMARK   3  BACKBONE(N, CA, C') + LIGAND ATOMS   0.53+/-0.09 A
REMARK   3  ALL HEAVY ATOMS                      0.91+/-0.08 A
```

REMARK 4 - 999

Overview

REMARKs following the refinement remark consist of free text annotation, predefined boilerplate remarks, and token: value pair styled templates. Presented here are examples of remark sections in PDB files

Record Format and Details

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

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

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

REMARK 4, Format

Remark 4 indicates the version of the PDB file format used to generate the file.

Template

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
REMARK   4
REMARK   4 XXXX COMPLIES WITH FORMAT V. 2.3, DD-MMM-YYYY
```

XXXX refers to the ID code of the entry.

N.M refers to the version number.

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

Example

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
REMARK   4
REMARK   4 1ABC COMPLIES WITH FORMAT V. 2.3, 09-JUL-1998
```

REMARKs 5-99, Not in use

REMARK 100, Deposition or Processing Site

This remark indicates the processing site: RCSB, MSD-EBI, PDBj, or NDB.

Example

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
REMARK 100
REMARK 100 THIS ENTRY HAS BEEN PROCESSED BY RCSB ON 29-NOV-2006.
REMARK 100 THE RCSB ID CODE IS RCSB040554.

REMARK 100
REMARK 100 THIS ENTRY HAS BEEN PROCESSED BY EBI  ON  7-NOV-2005.
REMARK 100 THE EBI ID CODE IS EBI-26270.

REMARK 100
REMARK 100 THIS ENTRY HAS BEEN PROCESSED BY PDBJ ON 21-DEC-2005.
REMARK 100 THE RCSB ID CODE IS RCSB025208.

REMARK 100
REMARK 100 THIS ENTRY HAS BEEN PROCESSED BY THE NUCLEIC ACID DATABASE
REMARK 100 ON 08-DEC-2006.
REMARK 100 THE NDB ID CODE IS PH0029
```

REMARKs 102 - 199, Nucleic acids

REMARK 102, For base mispairings

Remark 102 is mandatory if mispaired bases exist and Watson-Crick H-bonding is present.

Template

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
REMARK 102
REMARK 102 BASES   A B  NN AND   X Y  ZZ ARE MISPAIRED.
REMARK 102 BASES   A B  NN AND   X Y  ZZ ARE MISPAIRED.
REMARK 102 ALL OTHER HYDROGEN BONDS BETWEEN BASE PAIRS IN THIS ENTRY
REMARK 102 FOLLOW THE CONVENTIONAL WATSON-CRICK HYDROGEN BONDING
REMARK 102 PATTERN AND THEY HAVE NOT BEEN PRESENTED ON *CONECT*
REMARK 102 RECORDS IN THIS ENTRY.
```

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

Example

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
REMARK 102
REMARK 102 BASES   G A   4 AND   A B  21 ARE MISPAIRED.
REMARK 102 BASES   A A   9 AND   G B  16 ARE MISPAIRED.
REMARK 102 ALL OTHER HYDROGEN BONDS BETWEEN BASE PAIRS IN THIS ENTRY
```

```
REMARK 102 FOLLOW THE CONVENTIONAL WATSON-CRICK HYDROGEN BONDING
REMARK 102 PATTERN AND THEY HAVE NOT BEEN PRESENTED ON *CONECT*
REMARK 102 RECORDS IN THIS ENTRY.
```

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

REMARK 103

Remark 103 is mandatory if non-Watson-Crick H-bonding is present for specific interactions.

Template

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
REMARK 103
REMARK 103 THERE ARE NON-WATSON-CRICK HYDROGEN BONDS BETWEEN THE
REMARK 103 FOLLOWING ATOMS:
REMARK 103  AB   I X   N   AND  AB   Z X  NN
REMARK 103  AB   I X   N   AND  AB   Z X  NN
REMARK 103 ALL OTHER HYDROGEN BONDS BETWEEN BASE PAIRS IN THIS ENTRY
REMARK 103 FOLLOW THE CONVENTIONAL WATSON-CRICK HYDROGEN BONDING
REMARK 103 PATTERN AND THEY HAVE NOT BEEN PRESENTED ON *CONECT*
REMARK 103 RECORDS IN THIS ENTRY.
```

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

REMARK 104

Remark 104 is mandatory if inosine exists.

Template

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
REMARK 104
REMARK 104 RESIDUE I X   N IS INOSINE.
REMARK 104 RESIDUE I X   N IS INOSINE.
```

X is the chain identifier and N the sequence number.

Example

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
REMARK 103
REMARK 103 THERE ARE NON-WATSON-CRICK HYDROGEN BONDS BETWEEN THE
REMARK 103 FOLLOWING ATOMS:
REMARK 103  N1   I A   1   AND  N3   C B  16
REMARK 103  O6   I A   1   AND  N4   C B  16
REMARK 103  N1   I A   3   AND  N3   C B  14
```

```
REMARK 103  O6   I A   3   AND  N4   C B  14
REMARK 103 ALL OTHER HYDROGEN BONDS BETWEEN BASE PAIRS IN THIS ENTRY
REMARK 103 FOLLOW THE CONVENTIONAL WATSON-CRICK HYDROGEN BONDING
REMARK 103 PATTERN AND THEY HAVE NOT BEEN PRESENTED ON CONECT
REMARK 103 RECORDS IN THIS ENTRY.
REMARK 104
REMARK 104 RESIDUE I A   1 IS INOSINE.
REMARK 104 RESIDUE I A   3 IS INOSINE.
```

REMARK 105

Remark 105 is mandatory if nucleic acids exist in an entry.

Template

```
         1         2         3         4         5         6         7
12345678901234567890123456789012345678901234567890123456789012345678901234567890
REMARK 105
REMARK 105 THE PROTEIN DATA BANK HAS ADOPTED THE SACCHARIDE CHEMISTS
REMARK 105 NOMENCLATURE FOR ATOMS OF THE DEOXYRIBOSE/RIBOSE MOIETY
REMARK 105 RATHER THAN THAT OF THE NUCLEOSIDE CHEMISTS.  THE RING
REMARK 105 OXYGEN ATOM IS LABELLED O4* INSTEAD OF O1*.
```

REMARK 106

Remark 106 is mandatory if hydrogen bonding is Watson-Crick.

Template

```
          1         2         3         4         5         6         7
12345678901234567890123456789012345678901234567890123456789012345678901234567890
REMARK 106
REMARK 106 THE HYDROGEN BONDS BETWEEN BASE PAIRS IN THIS ENTRY FOLLOW
REMARK 106 THE CONVENTIONAL WATSON-CRICK HYDROGEN BONDING PATTERN.
REMARK 106 THEY HAVE NOT BEEN PRESENTED ON *CONECT* RECORDS IN THIS
REMARK 106 ENTRY.
```

REMARK 200-250, Experimental Details

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

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

REMARK 200, X-ray Diffraction Experimental Details

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

Remark 200 is mandatory if x-ray.

Template

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
REMARK 200
REMARK 200 EXPERIMENTAL DETAILS
REMARK 200  EXPERIMENT TYPE                : X-RAY DIFFRACTION
REMARK 200  DATE OF DATA COLLECTION        :
REMARK 200  TEMPERATURE           (KELVIN) :
REMARK 200  PH                             :
REMARK 200  NUMBER OF CRYSTALS USED        :
REMARK 200
REMARK 200  SYNCHROTRON              (Y/N) :
REMARK 200  RADIATION SOURCE               :
REMARK 200  BEAMLINE                       :
REMARK 200  X-RAY GENERATOR MODEL          :
REMARK 200  MONOCHROMATIC OR LAUE    (M/L) :
REMARK 200  WAVELENGTH OR RANGE        (A) :
REMARK 200  MONOCHROMATOR                  :
REMARK 200  OPTICS                         :
REMARK 200
REMARK 200  DETECTOR TYPE                  :
REMARK 200  DETECTOR MANUFACTURER          :
REMARK 200  INTENSITY-INTEGRATION SOFTWARE :
REMARK 200  DATA SCALING SOFTWARE          :
REMARK 200
REMARK 200  NUMBER OF UNIQUE REFLECTIONS   :
REMARK 200  RESOLUTION RANGE HIGH      (A) :
REMARK 200  RESOLUTION RANGE LOW       (A) :
REMARK 200  REJECTION CRITERIA  (SIGMA(I)) :
REMARK 200
REMARK 200 OVERALL.
REMARK 200  COMPLETENESS FOR RANGE     (%) :
REMARK 200  DATA REDUNDANCY                :
REMARK 200  R MERGE                    (I) :
REMARK 200  R SYM                      (I) :
REMARK 200  <I/SIGMA(I)> FOR THE DATA SET  :
REMARK 200
REMARK 200 IN THE HIGHEST RESOLUTION SHELL.
REMARK 200  HIGHEST RESOLUTION SHELL, RANGE HIGH (A) :
REMARK 200  HIGHEST RESOLUTION SHELL, RANGE LOW  (A) :
REMARK 200  COMPLETENESS FOR SHELL     (%) :
REMARK 200  DATA REDUNDANCY IN SHELL       :
REMARK 200  R MERGE FOR SHELL          (I) :
REMARK 200  R SYM FOR SHELL            (I) :
REMARK 200  <I/SIGMA(I)> FOR SHELL         :
REMARK 200
REMARK 200 METHOD USED TO DETERMINE THE STRUCTURE:
REMARK 200 SOFTWARE USED:
REMARK 200 STARTING MODEL:
REMARK 200
REMARK 200 REMARK:
```

REMARK 205, Fiber Diffraction, Fiber Sample Experiment Details

Remark 205 is mandatory if fiber diffraction - non-crystalline sample.

Template

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
REMARK 205
REMARK 205 THESE COORDINATES WERE GENERATED FROM FIBER DIFFRACTION
REMARK 205 DATA.    PROTEIN DATA BANK CONVENTIONS REQUIRE THAT CRYST1
REMARK 205 AND SCALE RECORDS BE INCLUDED, BUT THE VALUES OF THESE
REMARK 205 RECORDS ARE MEANINGLESS.
```

REMARKs 210 and 215, NMR Experiment Details

Remark 210 is mandatory if NMR.

Template

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
REMARK 210
REMARK 210 EXPERIMENTAL DETAILS
REMARK 210  EXPERIMENT TYPE                          : NMR
REMARK 210  TEMPERATURE                    (KELVIN) :
REMARK 210  PH                                       :
REMARK 210
REMARK 210  NMR EXPERIMENTS CONDUCTED                :
REMARK 210  SPECTROMETER FIELD STRENGTH              :
REMARK 210  SPECTROMETER MODEL                       :
REMARK 210  SPECTROMETER MANUFACTURER                :
REMARK 210
REMARK 210      STRUCTURE DETERMINATION.
REMARK 210       SOFTWARE USED                         :
REMARK 210       METHOD USED                           :
REMARK 210
REMARK 210 CONFORMERS, NUMBER CALCULATED               :
REMARK 210 CONFORMERS, NUMBER SUBMITTED                :
REMARK 210 CONFORMERS, SELECTION CRITERIA :
REMARK 210
REMARK 210 REMARK:
```

Remark 215 is mandatory if NMR

Template

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
REMARK 215
REMARK 215 NMR STUDY
REMARK 215 THE COORDINATES IN THIS ENTRY WERE GENERATED FROM SOLUTION
REMARK 215 NMR DATA.        PROTEIN DATA BANK CONVENTIONS REQUIRE THAT
REMARK 215 CRYST1 AND SCALE RECORDS BE INCLUDED, BUT THE VALUES ON
REMARK 215 THESE RECORDS ARE MEANINGLESS.
```

REMARK 217, Solid State NMR

This remark will appear in all solid state NMR entries.

```
REMARK 217
REMARK 217 SOLID STATE NMR STUDY
REMARK 217 THE COORDINATES IN THIS ENTRY WERE GENERATED FROM SOLID
REMARK 217 STATE NMR DATA.  PROTEIN DATA BANK CONVENTIONS REQUIRE THAT
REMARK 217 CRYST1 AND SCALE RECORDS BE INCLUDED, BUT THE VALUES ON
REMARK 217 THESE RECORDS ARE MEANINGLESS.
```

REMARKs 220 and 225, Theoretical Modeling Experiment Details*

**Note: As of July 1, 2002, models are available from a directory separate from the main archive at* ftp://ftp.rcsb.org/pub/pdb/data/structures/models/current/. *As of October 15, 2006, theoretical models are no longer accepted for deposition.*

Template

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
REMARK 220
REMARK 220 EXPERIMENTAL DETAILS
REMARK 220  EXPERIMENT TYPE                : THEORETICAL MODELLING
REMARK 220
REMARK 220 REMARK:
```

Remark 225 is mandatory if theoretical model.

Template

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
REMARK 225
REMARK 225 THEORETICAL MODEL
REMARK 225 THE COORDINATES IN THIS ENTRY REPRESENT A MODEL STRUCTURE.
REMARK 225 PROTEIN DATA BANK CONVENTIONS REQUIRE THAT CRYST1 AND
REMARK 225 SCALE RECORDS BE INCLUDED, BUT THE VALUES ON THESE
REMARK 225 RECORDS ARE MEANINGLESS.
```

REMARK 230, Neutron Diffraction Experiment Details

Remark 230 is mandatory if neutron diffraction study.

Template

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
REMARK 230
REMARK 230 EXPERIMENTAL DETAILS
REMARK 230  EXPERIMENT TYPE                : NEUTRON DIFFRACTION
REMARK 230  DATE OF DATA COLLECTION        :
REMARK 230  TEMPERATURE           (KELVIN) :
REMARK 230  PH                             :
REMARK 230  NUMBER OF CRYSTALS USED        :
REMARK 230
REMARK 230  NEUTRON SOURCE                 :
REMARK 230  BEAMLINE                       :
REMARK 230  WAVELENGTH OR RANGE        (A) :
REMARK 230  MONOCHROMATOR                  :
REMARK 230  OPTICS                         :
REMARK 230
REMARK 230  DETECTOR TYPE                  :
REMARK 230  DETECTOR MANUFACTURER          :
REMARK 230  INTENSITY-INTEGRATION SOFTWARE :
REMARK 230  DATA SCALING SOFTWARE          :
REMARK 230
REMARK 230  NUMBER OF UNIQUE REFLECTIONS   :
REMARK 230  RESOLUTION RANGE HIGH      (A) :
REMARK 230  RESOLUTION RANGE LOW       (A) :
REMARK 230  REJECTION CRITERIA  (SIGMA(I)) :
REMARK 230
REMARK 230 OVERALL.
REMARK 230  COMPLETENESS FOR RANGE     (%) :
REMARK 230  DATA REDUNDANCY                :
REMARK 230  R MERGE                    (I) :
REMARK 230  R SYM                      (I) :
REMARK 230  <I/SIGMA(I)> FOR THE DATA SET  :
REMARK 230
REMARK 230 IN THE HIGHEST RESOLUTION SHELL.
REMARK 230  HIGHEST RESOLUTION SHELL, RANGE HIGH (A) :
REMARK 230  HIGHEST RESOLUTION SHELL, RANGE LOW  (A) :
REMARK 230  COMPLETENESS FOR SHELL     (%) :
REMARK 230  DATA REDUNDANCY IN SHELL       :
REMARK 230  R MERGE FOR SHELL          (I) :
REMARK 230  R SYM FOR SHELL            (I) :
REMARK 230  <I/SIGMA(I)> FOR SHELL         :
REMARK 230
REMARK 230 METHOD USED TO DETERMINE THE STRUCTURE:
REMARK 230 SOFTWARE USED :
REMARK 230 STARTING MODEL:
REMARK 230
REMARK 230 REMARK:
```

REMARK 240, Electron Diffraction Experiment Details

Remark 240 is mandatory if electron diffraction study.

Template

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
REMARK 240
REMARK 240 EXPERIMENTAL DETAILS
REMARK 240  EXPERIMENT TYPE                : ELECTRON DIFFRACTION
REMARK 240  DATE OF DATA COLLECTION        :
REMARK 240
REMARK 240 REMARK:
```

REMARK 245, Cryo-Electron Microscopy Experiment Details

Remark 245 is mandatory if cryo-EM study.

Template

```
REMARK 245
REMARK 245 EXPERIMENTAL DETAILS
REMARK 245   EXPERIMENT TYPE : CRYO-ELECTRON MICROSCOPY
REMARK 245
REMARK 245 ELECTRON MICROSCOPE SAMPLE
REMARK 245   SAMPLE AGGREGATION STATE       :
REMARK 245
REMARK 245   NAME OF SAMPLE                 :
REMARK 245
REMARK 245   SAMPLE CONCENTRATION           :
REMARK 245   SAMPLE SUPPORT DETAILS         :
REMARK 245
REMARK 245   SAMPLE VITRIFICATION DETAILS   :
REMARK 245
REMARK 245   SAMPLE BUFFER                  :
REMARK 245
REMARK 245   PH                            :
REMARK 245   SAMPLE DETAILS:
REMARK 245
REMARK 245 DATA ACQUISITION
REMARK 245   DATE OF EXPERIMENT                :
REMARK 245   NUMBER OF MICROGRAPHS-IMAGES      :
REMARK 245   TEMPERATURE (KELVIN)              :
REMARK 245   MICROSCOPE MODEL                  :
REMARK 245   DETECTOR TYPE                    :
REMARK 245   MINIMUM DEFOCUS (NM)             :
REMARK 245   MAXIMUM DEFOCUS (NM)             :
REMARK 245   MINIMUM TILT ANGLE (DEGREES)     :
REMARK 245   MAXIMUM TILT ANGLE (DEGREES)     :
REMARK 245   NOMINAL CS                       :
REMARK 245   IMAGING MODE                     :
REMARK 245   ELECTRON DOSE (ELECTRONS NM**-2)  :
REMARK 245   ILLUMINATION MODE                :
REMARK 245   NOMINAL MAGNIFICATION            :
REMARK 245   CALIBRATED MAGNIFICATION         :
REMARK 245   SOURCE                           :
REMARK 245   ACCELERATION VOLTAGE (KV)        :
REMARK 245   IMAGING DETAILS:
```

Example

```
REMARK 245
REMARK 245 EXPERIMENTAL DETAILS
REMARK 245
REMARK 245 EXPERIMENT TYPE : CRYO-ELECTRON MICROSCOPY
REMARK 245
REMARK 245 ELECTRON MICROSCOPE SAMPLE
REMARK 245   SAMPLE AGGREGATION STATE       : ICOSAHEDRAL
REMARK 245   NAME OF SAMPLE                 : BACTERIOPHAGE PRD1 SUS1 MUTANT
REMARK 245   SAMPLE CONCENTRATION           : NULL
REMARK 245   SAMPLE SUPPORT DETAILS         : HOLEY CARBON
REMARK 245   SAMPLE VITRIFICATION DETAILS   : PLUNGE VITRIFICATION
REMARK 245   SAMPLE BUFFER                  : NULL
REMARK 245   PH                             : 7.2
REMARK 245   SAMPLE DETAILS: THE SAMPLE CONSISTS OF THE ADENOVIRUS-
REMARK 245    RELATED BACTERIOPHAGE PRD1. 400 MESH COPPER GLOW DISCHARGE
REMARK 245    SAMPLES WERE PREPARED AS THIN LAYERS OF VITREOUS ICE.
REMARK 245
REMARK 245 DATA ACQUISITION
REMARK 245   DATE OF EXPERIMENT               : 15 JUNE 1998
REMARK 245   NUMBER OF MICROGRAPHS-IMAGES     : 29
REMARK 245   TEMPERATURE (KELVIN)             : 95
REMARK 245   MICROSCOPE MODEL                 : PHILIPS CM200 FEG
REMARK 245   DETECTOR TYPE                    : SO-163 FILM
REMARK 245   MINIMUM DEFOCUS (NM)             : 1300
REMARK 245   MAXIMUM DEFOCUS (NM)             : 4100
REMARK 245   MINIMUM TILT ANGLE (DEGREES)     : 0
REMARK 245   MAXIMUM TILT ANGLE (DEGREES)     : 0
REMARK 245   NOMINAL CS                       : 2
REMARK 245   IMAGING MODE                     : LOW DOSE
REMARK 245   ELECTRON DOSE (ELECTRONS NM**-2) : 1000
REMARK 245   ILLUMINATION MODE                : BRIGHT FIELD
REMARK 245   NOMINAL MAGNIFICATION           : 36000
REMARK 245   CALIBRATED MAGNIFICATION         : NULL
REMARK 245   SOURCE                           : FIELD EMISSION GUN
REMARK 245   ACCELERATION VOLTAGE (KV)        : 200
REMARK 245   IMAGING DETAILS: SAMPLES WERE MAINTAINED AT LIQUID NITROGEN
REMARK 245    TEMPERATURES IN THE ELECTRON MICROSCOPE WITH A GATAN 626-0300
REMARK 245    CRYOTRANSFER HOLDER.
```

REMARK 247, mandatory if Electron Microscopy

Template

```
         1         2         3         4         5         6         7
12345678901234567890123456789012345678901234567890123456789012345678901234567890
REMARK 247
REMARK 247 ELECTRON MICROSCOPY
REMARK 247   THE COORDINATES IN THIS ENTRY WERE GENERATED FROM
REMARK 247   ELECTRON MICROSCOPY DATA. PROTEIN DATA BANK CONVENTIONS
REMARK 247   REQUIRE THAT CRYST1 AND SCALE RECORDS BE INCLUDED,
REMARK 247   BUT THE VALUES ON THESE RECORDS ARE MEANINGLESS
REMARK 247   EXCEPT FOR THE CALCULATION OF THE STRUCTURE FACTORS
```

REMARK 250, Other Type of Experiment Details

Remark specific to other kinds of studies, not listed above.

Remark 250 is mandatory if other than x-ray, NMR, theoretical model*, neutron, or electron study.

**Note: As of July 1, 2002, models are available from a directory separate from the main archive at* ftp://ftp.rcsb.org/pub/pdb/data/structures/models/current/. *As of October 15, 2006, theoretical models are no longer accepted for deposition.*

Template

```
         1         2         3         4         5         6         7
12345678901234567890123456789012345678901234567890123456789012345678901234567890
REMARK 250
REMARK 250 EXPERIMENTAL DETAILS
REMARK 250  EXPERIMENT TYPE                :
REMARK 250  DATE OF DATA COLLECTION        :
REMARK 250
REMARK 250 REMARK:
```

REMARK 265, Solution Scatter Example Experiment Details

```
REMARK 265
REMARK 265 EXPERIMENTAL DETAILS
REMARK 265
REMARK 265 EXPERIMENT TYPE : X-RAY SOLUTION SCATTERING
REMARK 265  DATA ACQUISITION
REMARK 265   RADIATION/NEUTRON SOURCE                 : SRS BEAMLINE 2.1
REMARK 265   SYNCHROTRON (Y/N)                        : Y
REMARK 265   BEAMLINE                                 : 2.1
REMARK 265   BEAMLINE INSTRUMENT                      : NULL
REMARK 265   DETECTOR TYPE                            : 500-CHANNEL QUADRANT
REMARK 265   DETECTOR MANUFACTURER DETAILS            : NULL
REMARK 265   TEMPERATURE (KELVIN)                     : 288
REMARK 265   PH                                       : NULL
REMARK 265   NUMBER OF TIME FRAMES USED               : 10
REMARK 265   PROTEIN CONCENTRATION RANGE (MG/ML)      : 0.7 - 14
REMARK 265   SAMPLE BUFFER                           : TRIS
REMARK 265   DATA REDUCTION SOFTWARE                  : OTOKO
REMARK 265   DATA ANALYSIS SOFTWARE                   : SCTPL5, GNOM
REMARK 265   GUINIER MEAN RADIUS OF GYRATION (NM)     : 11.1
REMARK 265   SIGMA MEAN RADIUS OF GYRATION           : 0.4
REMARK 265   R(XS-1) MEAN CROSS SECTIONAL RADII (NM)  : 4.4
REMARK 265   R(XS-1) SIGMA MEAN CROSS SECTIONAL RADII : 0.2
REMARK 265   R(XS-2) MEAN CROSS SECTIONAL RADII (NM)  : 1.7
REMARK 265   R(XS-2) SIGMA MEAN CROSS SECTIONAL RADII : 0.1
REMARK 265   P(R) PROTEIN LENGTH (NM)                 : 40
REMARK 265
REMARK 265 EXPERIMENT TYPE : NEUTRON SOLUTION SCATTERING
REMARK 265  DATA ACQUISITION
REMARK 265   RADIATION/NEUTRON SOURCE                 : ILL
REMARK 265   SYNCHROTRON (Y/N)                        : N
REMARK 265   BEAMLINE                                 : NULL
REMARK 265   BEAMLINE INSTRUMENT                      : D11, D22
REMARK 265   DETECTOR TYPE                            : AREA
REMARK 265   DETECTOR MANUFACTURER DETAILS            : NULL
REMARK 265   TEMPERATURE (KELVIN)                     : NULL
REMARK 265   PH                                       : NULL
```

```
REMARK 265   NUMBER OF TIME FRAMES USED               : NULL
REMARK 265   PROTEIN CONCENTRATION RANGE (MG/ML)      : 0.4 - 9.6
REMARK 265   SAMPLE BUFFER                            : PBS IN 99.9% D2O
REMARK 265   DATA REDUCTION SOFTWARE                  : DETEC, RNILS, SPOLLY
REMARK 265   DATA ANALYSIS SOFTWARE                   : SCTPL5, GNOM
REMARK 265   GUINIER MEAN RADIUS OF GYRATION (NM)     : 11.3
REMARK 265   SIGMA MEAN RADIUS OF GYRATION            : 0.4
REMARK 265   R(XS-1) MEAN CROSS SECTIONAL RADII (NM)  : 3.9
REMARK 265   R(XS-1) SIGMA MEAN CROSS SECTIONAL RADII : 0.2
REMARK 265   R(XS-2) MEAN CROSS SECTIONAL RADII (NM)  : 1.51
REMARK 265   R(XS-2) SIGMA MEAN CROSS SECTIONAL RADII : 0.06
REMARK 265   P(R) PROTEIN LENGTH (NM)                 : 37 - 39
REMARK 265
REMARK 265  DATA ACQUISITION
REMARK 265   RADIATION/NEUTRON SOURCE                 : ISIS
REMARK 265   SYNCHROTRON (Y/N)                        : N
REMARK 265   BEAMLINE                                 : PULSED NEUTRON
REMARK 265   BEAMLINE INSTRUMENT                      : LOQ
REMARK 265   DETECTOR TYPE                            : AREA (TIME-OF-FLIGHT)
REMARK 265   DETECTOR MANUFACTURER DETAILS            : NULL
REMARK 265   TEMPERATURE (KELVIN)                     : NULL
REMARK 265   PH                                       : NULL
REMARK 265   NUMBER OF TIME FRAMES USED               : NULL
REMARK 265   PROTEIN CONCENTRATION RANGE (MG/ML)      : 3.7, 6.1
REMARK 265   SAMPLE BUFFER                            : PBS IN 99.9% D2O
REMARK 265   DATA REDUCTION SOFTWARE                  : COLLETTE
REMARK 265   DATA ANALYSIS SOFTWARE                   : SCTPL5, GNOM
REMARK 265   GUINIER MEAN RADIUS OF GYRATION (NM)     : 11.7
REMARK 265   SIGMA MEAN RADIUS OF GYRATION            : 0.5
REMARK 265   R(XS-1) MEAN CROSS SECTIONAL RADII (NM)  : NULL
REMARK 265   R(XS-1) SIGMA MEAN CROSS SECTIONAL RADII : NULL
REMARK 265   R(XS-2) MEAN CROSS SECTIONAL RADII (NM)  : NULL
REMARK 265   R(XS-2) SIGMA MEAN CROSS SECTIONAL RADII : NULL
REMARK 265   P(R) PROTEIN LENGTH (NM)                 : 40
REMARK 265
REMARK 265 EXPERIMENT TYPE: THEORETICAL MODELLING
REMARK 265  METHOD USED TO DETERMINE THE STRUCTURE: CONSTRAINED SCATTERING
REMARK 265                                          FITTING OF HOMOLOGY
REMARK 265                                          MODELS
REMARK 265  SOFTWARE USED    : INSIGHT II, HOMOLOGY, DISCOVERY,
REMARK 265                     BIOPOLYMER, DELPHI
REMARK 265  SOFTWARE AUTHORS : MSI
REMARK 265  STARTING MODEL   : PDB CODE 1HFI, 1HCC, 1HFH, 1VCC
REMARK 265
REMARK 265  EXPERIMENTAL DETAILS: HOMOLOGY MODELS WERE BUILT FOR
REMARK 265   THE 17 SCR DOMAINS AND ENERGY MINIMISATIONS WERE
REMARK 265   PERFORMED TO IMPROVE THE CONNECTIVITY IN THE FH MODEL.
REMARK 265   TRIANTENNARY COMPLEX-TYPE CARBOHYDRATE STRUCTURES
REMARK 265   (MAN3GLCNAC6GAL3FUC3NEUNAC1) WERE ADDED TO EACH OF THE
REMARK 265   N-LINKED GLYCOSYLATION SITES. A LIBRARY OF LINKER PEPTIDE
REMARK 265   CONFORMATIONS WAS USED IN DOMAIN MODELLING CONSTRAINED
REMARK 265   BY THE SOLUTION SCATTERING FITS. MODELLING WITH THE
REMARK 265   SCATTERING DATA WAS ALSO CARRIED OUT BY ROTATIONAL
REMARK 265   SEARCH METHODS. THE X-RAY AND NEUTRON SCATTERING CURVE
REMARK 265   I(Q) WAS CALCULATED ASSUMING A UNIFORM SCATTERING DENSITY
REMARK 265   FOR THE SPHERES USING THE DEBYE EQUATION AS ADAPTED TO
REMARK 265   SPHERES. X-RAY CURVES WERE CALCULATED FROM THE HYDRATED
REMARK 265   SPHERE MODELS WITHOUT CORRECTIONS FOR WAVELENGTH SPREAD OR
REMARK 265   BEAM DIVERGENCE, WHILE THESE CORRECTIONS WERE APPLIED FOR
REMARK 265   THE NEUTRON CURVES BUT NOW USING UNHYDRATED MODELS.
REMARK 265
REMARK 265 CONFORMERS, NUMBER CALCULATED   : 2010
REMARK 265 CONFORMERS, NUMBER SUBMITTED    : 4
```

```
REMARK 265 CONFORMERS, SELECTION CRITERIA  :  THE MODELLED SCATTERING
REMARK 265   CURVES WERE ASSESSED BY CALCULATION OF THE
REMARK 265   RG, RSX-1 AND RXS-2 VALUES IN THE SAME Q RANGES
REMARK 265   USED IN THE EXPERIMENTAL GUINIER FITS. MODELS WERE
REMARK 265   THEN RANKED USING A GOODNESS-OF-FIT R-FACTOR
REMARK 265   DEFINED BY ANALOGY WITH PROTEIN CRYSTALLOGRAPHY
REMARK 265   AND BASED ON THE EXPERIMENTAL CURVES IN THE Q RANGE
REMARK 265   EXTENDING TO 1.4 NM-1.
REMARK 265
```

REMARK 280, Crystal

Remark 280 presents information on the crystal. The solvent content and Matthews coefficient are provided for protein and polypeptide crystals. Crystallization conditions are free text.

Remark 280 is mandatory if single crystal study.

Template

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
REMARK 280
REMARK 280 CRYSTAL
REMARK 280 SOLVENT CONTENT, VS   (%):
REMARK 280 MATTHEWS COEFFICIENT, VM (ANGSTROMS**3/DA):
REMARK 280
REMARK 280 CRYSTALLIZATION CONDITIONS: FREE TEXT GOES HERE.
```

Example

```
REMARK 280 CRYSTAL
REMARK 280 SOLVENT CONTENT, VS  (%): 36.85
REMARK 280 MATTHEWS COEFFICIENT, VM (ANGSTROMS**3/DA): 1.79
REMARK 280
REMARK 280 CRYSTALLIZATION CONDITIONS: 1.4M SODIUM ACETATE,
REMARK 280                              0.1M MES PH 6.5
```

REMARK 285, CRYST1

Remark 285 presents information on the unit cell.

Template

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
REMARK 285
REMARK 285 CRYST1
REMARK 285 FREE TEXT GOES HERE.
```

Example

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
REMARK 285
REMARK 285 CRYST1
REMARK 285 TEXT TO EXPLAIN UNUSUAL UNIT-CELL DATA: THE DATA WAS
```

```
REMARK 285 COLLECTED ON TWO-DIMENSIONAL CRYSTALS AND HENCE THE
REMARK 285 C-AXIS REPEAT DOES NOT CORRESPOND TO A REAL REPEAT, BUT
REMARK 285 INSTEAD REFERS TO THE SAMPLING THAT IS USED TO DESCRIBE
REMARK 285 THE CONTINUOUS TRANSFORM. THE C VALUE OF 100.9 IS
REMARK 285 THEREFORE THE VALUE WHICH SHOULD BE USED IN
REMARK 285 INTERPRETING THE MEANING OF THE L INDEX.
```

REMARK 290, Crystallographic Symmetry

Remark 290 is mandatory for crystalline studies. The remark is generated by PDB.

Example

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
REMARK 290
REMARK 290 CRYSTALLOGRAPHIC SYMMETRY
REMARK 290 SYMMETRY OPERATORS FOR SPACE GROUP: P 21 21 21
REMARK 290
REMARK 290        SYMOP      SYMMETRY
REMARK 290     NNNMMM   OPERATOR
REMARK 290       1555   X,Y,Z
REMARK 290       2555   1/2-X,-Y,1/2+Z
REMARK 290       3555   -X,1/2+Y,1/2-Z
REMARK 290       4555   1/2+X,1/2-Y,-Z
REMARK 290
REMARK 290     WHERE NNN -> OPERATOR NUMBER
REMARK 290           MMM -> TRANSLATION VECTOR
REMARK 290
REMARK 290 CRYSTALLOGRAPHIC SYMMETRY TRANSFORMATIONS
REMARK 290 THE FOLLOWING TRANSFORMATIONS OPERATE ON THE ATOM/HETATM
REMARK 290 RECORDS IN THIS ENTRY TO PRODUCE CRYSTALLOGRAPHICALLY
REMARK 290 RELATED MOLECULES.
REMARK 290   SMTRY1   1  1.000000  0.000000  0.000000        0.00000
REMARK 290   SMTRY2   1  0.000000  1.000000  0.000000        0.00000
REMARK 290   SMTRY3   1  0.000000  0.000000  1.000000        0.00000
REMARK 290   SMTRY1   2 -1.000000  0.000000  0.000000       36.30027
REMARK 290   SMTRY2   2  0.000000 -1.000000  0.000000        0.00000
REMARK 290   SMTRY3   2  0.000000  0.000000  1.000000       59.50256
REMARK 290   SMTRY1   3 -1.000000  0.000000  0.000000        0.00000
REMARK 290   SMTRY2   3  0.000000  1.000000  0.000000       46.45545
REMARK 290   SMTRY3   3  0.000000  0.000000 -1.000000       59.50256
REMARK 290   SMTRY1   4  1.000000  0.000000  0.000000       36.30027
REMARK 290   SMTRY2   4  0.000000 -1.000000  0.000000       46.45545
REMARK 290   SMTRY3   4  0.000000  0.000000 -1.000000        0.00000
REMARK 290
REMARK 290 REMARK:
```

REMARK 295, Non-Crystallographic Symmetry

Description of non-crystallographic symmetry. Mandatory when MTRIX records are present.

Template

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
REMARK 295
REMARK 295 NON-CRYSTALLOGRAPHIC SYMMETRY
REMARK 295 THE TRANSFORMATIONS PRESENTED ON THE MTRIX RECORDS BELOW
REMARK 295 DESCRIBE NON-CRYSTALLOGRAPHIC RELATIONSHIPS AMONG ATOMS
REMARK 295 IN THIS ENTRY.  APPLYING THE APPROPRIATE MTRIX
REMARK 295 TRANSFORMATION TO THE RESIDUES LISTED FIRST WILL YIELD
REMARK 295 APPROXIMATE COORDINATES FOR THE RESIDUES LISTED SECOND.
REMARK 295 CHAIN IDENTIFIERS GIVEN AS "?" REFER TO CHAINS FOR WHICH
REMARK 295 ATOMS ARE NOT FOUND IN THIS ENTRY.
REMARK 295
REMARK 295               APPLIED TO          TRANSFORMED TO
REMARK 295   TRANSFORM CHAIN  RESIDUES       CHAIN  RESIDUES     RMSD
REMARK 295     SSS       ?    ? .. ?           ?    ? .. ?       ?
REMARK 295
REMARK 295    WHERE SSS -> COLUMNS 8-10 OF MTRIX RECORDS
REMARK 295
REMARK 295 REMARK:
```

Example

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
REMARK 295
REMARK 295 NON-CRYSTALLOGRAPHIC SYMMETRY
REMARK 295 THE TRANSFORMATIONS PRESENTED ON THE MTRIX RECORDS BELOW
REMARK 295 DESCRIBE NON-CRYSTALLOGRAPHIC RELATIONSHIPS AMONG ATOMS
REMARK 295 IN THIS ENTRY.  APPLYING THE APPROPRIATE MTRIX
REMARK 295 TRANSFORMATION TO THE RESIDUES LISTED FIRST WILL YIELD
REMARK 295 APPROXIMATE COORDINATES FOR THE RESIDUES LISTED SECOND.
REMARK 295 CHAIN IDENTIFIERS GIVEN AS "?" REFER TO CHAINS FOR WHICH
REMARK 295 ATOMS ARE NOT FOUND IN THIS ENTRY.
REMARK 295
REMARK 295               APPLIED TO          TRANSFORMED TO
REMARK 295   TRANSFORM CHAIN  RESIDUES       CHAIN  RESIDUES     RMSD
REMARK 295     SSS
REMARK 295    M  1       A    1 .. 374         C    1 .. 374     0.010
REMARK 295    M  2       B    1 .. 374         D    1 .. 374     0.010
REMARK 295
REMARK 295    WHERE SSS -> COLUMNS 8-10 OF MTRIX RECORDS
REMARK 295
REMARK 295 REMARK:
```

REMARK 300, Biomolecule

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

Remark 300 is mandatory if Remark 350 is provided.

Template

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
REMARK 300
REMARK 300 BIOMOLECULE:
REMARK 300 THIS ENTRY CONTAINS THE UNIQUE NON-CRYSTALLOGRAPHIC VIRAL
REMARK 300 REPEAT UNIT, WHICH CONSISTS OF ? CHAIN(S). SEE REMARK 350 FOR
REMARK 300 INFORMATION ON GENERATING THE BIOLOGICAL MOLECULE(S).
REMARK 300
REMARK 300  free text
```

Example (for Cryo-Electron Microscopy)

```
REMARK 300
REMARK 300 BIOMOLECULE: 1
REMARK 300 THIS ENTRY CONTAINS THE UNIQUE NON-CRYSTALLOGRAPHIC VIRAL
REMARK 300 REPEAT UNIT, WHICH CONSISTS OF 12 CHAIN(S). SEE REMARK 350 FOR
REMARK 300 INFORMATION ON GENERATING THE BIOLOGICAL MOLECULE(S).
REMARK 300
REMARK 300 ASSEMBLY COMPONENTS
REMARK 300   COM_ID: 1;
REMARK 300    NAME: ENTEROBACTERIA PHAGE PRD1 SUS1 MUTANT;
REMARK 300    OTHER_DETAILS: VIRUS;
REMARK 300
REMARK 300 VIRUS PARTICULARS
REMARK 300   COM_ID: 1;
REMARK 300    VIRUS_HOST_CATEGORY:  BACTERIA ;
REMARK 300    VIRUS_HOST_SPECIES: ENTEROBACTERIA ;
REMARK 300    VIRUS_HOST_GROWTH_CELL: NULL ;
REMARK 300    VIRUS_TYPE: VIRUS;
REMARK 300    VIRUS_ISOLATE: SPECIES ;
REMARK 300    ICTVDB_ID: 68.0.1.0.001
REMARK 300
REMARK 300 THE VIRUS PARTICLE HAS AN ICOSAHEDRALLY ARRANGED OUTER PROTEIN
REMARK 300 COAT. EACH CAPSID ICOSAHEDRAL ASYMMETRIC UNIT CONTAINS FOUR
REMARK 300 INDEPENDENT COPIES OF THE P3 TRIMER.
REMARK 300


REMARK 300
REMARK 300 BIOMOLECULE: 1
REMARK 300 THIS ENTRY CONTAINS THE CRYSTALLOGRAPHIC ASYMMETRIC UNIT
REMARK 300 WHICH CONSISTS OF   2 CHAIN(S). SEE REMARK 350 FOR
REMARK 300 INFORMATION ON GENERATING THE BIOLOGICAL MOLECULE(S).
REMARK 300
REMARK 300 QUATERNARY STRUCTURE FOR THIS ENTRY: DIMERIC
REMARK 300
REMARK 300 FOR THE HOMO-ASSEMBLY DESCRIBED BY REMARK 350
REMARK 300 THE DIFFERENCE IN ACCESSIBLE SURFACE AREA PER
REMARK 300 CHAIN BETWEEN THE ISOLATED CHAIN AND THAT FOR
REMARK 300 THE CHAIN IN THE COMPLEX IS     2498.4 ANGSTROM**2
```

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
REMARK 300
REMARK 300 BIOMOLECULE: 1, 2, 3
REMARK 300 THIS ENTRY CONTAINS THE CRYSTALLOGRAPHIC ASYMMETRIC UNIT
REMARK 300 WHICH CONSISTS OF 6 CHAIN(S). SEE REMARK 350 FOR
REMARK 300 INFORMATION ON GENERATING THE BIOLOGICAL MOLECULE(S).
REMARK 300
REMARK 300 THE CATALYTIC SUBUNIT OF LIVER ALCOHOL DEHYDROGENASE FROM
REMARK 300 EQUUS CABALLUS IS A HOMO DIMER.
```

REMARK 350, Generating the Biomolecule

Remark 350 presents all transformations, both crystallographic and non-crystallographic, needed to generate the biomolecule. These transformations operate on the coordinates in the entry.

Remark 350 is mandatory if Remark 300 is provided.

Template

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
REMARK 350
REMARK 350 GENERATING THE BIOMOLECULE
REMARK 350 COORDINATES FOR A COMPLETE MULTIMER REPRESENTING THE KNOWN
REMARK 350 BIOLOGICALLY SIGNIFICANT OLIGOMERIZATION STATE OF THE
REMARK 350 MOLECULE CAN BE GENERATED BY APPLYING BIOMT TRANSFORMATIONS
REMARK 350 GIVEN BELOW.  BOTH NON-CRYSTALLOGRAPHIC AND
REMARK 350 CRYSTALLOGRAPHIC OPERATIONS ARE GIVEN.
REMARK 350
REMARK 350 BIOMOLECULE: ?
REMARK 350 APPLY THE FOLLOWING TO CHAINS: ?, ?...
REMARK 350   BIOMT1   N  N.NNNNNN  N.NNNNNN  N.NNNNNN        N.NNNNN
REMARK 350   BIOMT2   N  N.NNNNNN  N.NNNNNN  N.NNNNNN        N.NNNNN
REMARK 350   BIOMT3   N  N.NNNNNN  N.NNNNNN  N.NNNNNN        N.NNNNN
```

Example

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
REMARK 350
REMARK 350 GENERATING THE BIOMOLECULE
REMARK 350 COORDINATES FOR A COMPLETE MULTIMER REPRESENTING THE KNOWN
REMARK 350 BIOLOGICALLY SIGNIFICANT OLIGOMERIZATION STATE OF THE
REMARK 350 MOLECULE CAN BE GENERATED BY APPLYING BIOMT TRANSFORMATIONS
REMARK 350 GIVEN BELOW.  BOTH NON-CRYSTALLOGRAPHIC AND
REMARK 350 CRYSTALLOGRAPHIC OPERATIONS ARE GIVEN.
REMARK 350
REMARK 350 BIOMOLECULE: 1
REMARK 350 APPLY THE FOLLOWING TO CHAINS: A, B, C
REMARK 350   BIOMT1   1  1.000000  0.000000  0.000000        0.00000
REMARK 350   BIOMT2   1  0.000000  1.000000  0.000000        0.00000
REMARK 350   BIOMT3   1  0.000000  0.000000  1.000000        0.00000
REMARK 350   BIOMT1   2  1.000000  0.000000  0.000000        0.00000
```

```
REMARK 350   BIOMT2   2  0.000000  1.000000  0.000000       60.00000
REMARK 350   BIOMT3   2  0.000000  0.000000  1.000000        0.00000
REMARK 350   BIOMT1   3 -1.000000  0.000000  0.000000        0.00000
REMARK 350   BIOMT2   3  0.000000  1.000000  0.000000     -120.00000
REMARK 350   BIOMT3   3  0.000000  0.000000 -1.000000        0.00000
REMARK 350 APPLY THE FOLLOWING TO CHAINS: D, E, F
REMARK 350   BIOMT1   1  1.000000  0.000000  0.000000        0.00000
REMARK 350   BIOMT2   1  0.000000  1.000000  0.000000        0.00000
REMARK 350   BIOMT3   1  0.000000  0.000000  1.000000        0.00000
REMARK 350   BIOMT1   4  1.000000  0.000000  0.000000        0.00000
REMARK 350   BIOMT2   4  0.000000 -1.000000  0.000000       60.00000
REMARK 350   BIOMT3   4  0.000000  0.000000  1.000000        0.00000
REMARK 350   BIOMT1   5 -1.000000  0.000000  0.000000        0.00000
REMARK 350   BIOMT2   5  0.000000 -1.000000  0.000000     -120.00000
REMARK 350   BIOMT3   5  0.000000  0.000000  1.000000        0.00000
```

(**NOTE** Identity matrix in BIOMT is mandatory)

```
REMARK 350
REMARK 350 GENERATING THE BIOMOLECULE
REMARK 350 COORDINATES FOR A COMPLETE MULTIMER REPRESENTING THE KNOWN
REMARK 350 BIOLOGICALLY SIGNIFICANT OLIGOMERIZATION STATE OF THE
REMARK 350 MOLECULE CAN BE GENERATED BY APPLYING BIOMT TRANSFORMATIONS
REMARK 350 GIVEN BELOW.  BOTH NON-CRYSTALLOGRAPHIC AND
REMARK 350 CRYSTALLOGRAPHIC OPERATIONS ARE GIVEN.
REMARK 350
REMARK 350 APPLY THE FOLLOWING TO CHAINS: A, B, C, D, E, F, G, H
REMARK 350 APPLY THE FOLLOWING TO CHAINS: I, J, K, L
REMARK 350   BIOMT1   1  1.000000  0.000000  0.000000        0.00000
REMARK 350   BIOMT2   1  0.000000  1.000000  0.000000        0.00000
REMARK 350   BIOMT3   1  0.000000  0.000000  1.000000        0.00000
REMARK 350   BIOMT1   2 -0.500000 -0.865983  0.000000        0.00000
REMARK 350   BIOMT2   2  0.866068 -0.500000  0.000000        0.00000
REMARK 350   BIOMT3   2  0.000000  0.000000  1.000000        0.00000
```

REMARK 375, Special Position

Remark 375 specifies atoms that are known to lie in particular locations, related by the symmetry elements, at which objects may be placed if and only if they possess symmetry which coincides with that of the cell.

Template

```
         1         2         3         4         5         6         7
12345678901234567890123456789012345678901234567890123456789012345678901234567890
REMARK 375
REMARK 375 SPECIAL POSITION
REMARK 375 FREE TEXT GOES HERE.
```

Example

```
         1         2         3         4         5         6         7
12345678901234567890123456789012345678901234567890123456789012345678901234567890
REMARK 375
REMARK 375 SPECIAL POSITION
REMARK 375      HOH   301  LIES ON A SPECIAL POSITION.
REMARK 375      HOH    77  LIES ON A SPECIAL POSITION.
REMARK 375
```

```
REMARK 375 SPECIAL POSITION
REMARK 375 MG   MO4 A  10  LIES ON A SPECIAL POSITION.
REMARK 375      HOH A  13  LIES ON A SPECIAL POSITION.
REMARK 375      HOH A  28  LIES ON A SPECIAL POSITION.
REMARK 375      HOH A  36  LIES ON A SPECIAL POSITION.
```

REMARK 400, Compound

Further details on the macromolecular contents of the entry.

Template

```
         1         2         3         4         5         6         7
12345678901234567890123456789012345678901234567890123456789012345678901234567890
REMARK 400
REMARK 400 COMPOUND
REMARK 400 FREE TEXT GOES HERE.
```

Example

```
         1         2         3         4         5         6         7
12345678901234567890123456789012345678901234567890123456789012345678901234567890
REMARK 400
REMARK 400 COMPOUND
REMARK 400 THE PRD1 SUS1 MUTANT LACKS THE PACKAGING PROTEIN P9
REMARK 400 AND PRODUCES ONLY EMPTY PARTICLES, WHICH REPRESENT
REMARK 400 AN ASSEMBLY INTERMEDIATE (MINDICH, L. ET AL.,
REMARK 400 J. VIROL 44, 1013-1020 (1982); MINDICH, L. ET AL.,
REMARK 400 J. VIROL 44, 1021-1030 (1982)).

REMARK 400
REMARK 400 COMPOUND
REMARK 400  COMPONENT OF NAPHTHALENE DIOXYGENASE (NDO)
REMARK 400  MULTICOMPONENT ENZYME SYSTEM WHICH CATALYZES THE INCORPORATION
REMARK 400  OF BOTH ATOMS OF MOLECULAR OXYGEN INTO NAPHTHALENE TO FORM
REMARK 400  CIS-NAPHTHALENE DIHYDRODIOL.
REMARK 400
REMARK 400 ENGINEERED RESIDUE IN CHAIN A, ARG 270 TO GLY
```

REMARK 450, Source

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

Template

```
         1         2         3         4         5         6         7
12345678901234567890123456789012345678901234567890123456789012345678901234567890
REMARK 450
REMARK 450 SOURCE
REMARK 450 FREE TEXT GOES HERE.
```

REMARK 465, Missing residues

Remark 465 lists the residues that are present in the SEQRES records but are not present in the coordinates section.

Template

```
REMARK 465
REMARK 465 MISSING RESIDUES
REMARK 465 THE FOLLOWING RESIDUES WERE NOT LOCATED IN THE
REMARK 465 EXPERIMENT. (M=MODEL NUMBER; RES=RESIDUE NAME; C=CHAIN
REMARK 465 IDENTIFIER; SSSEQ=SEQUENCE NUMBER; I=INSERTION CODE.)
REMARK 465
REMARK 465   M RES C SSSEQI
```

REMARK 470, Missing Atom

Non-hydrogen atoms of standard residues which are missing from the coordinates are listed. Missing HETATMS are not listed here.

Template

```
         1         2         3         4         5         6         7
12345678901234567890123456789012345678901234567890123456789012345678901234567890
REMARK 470
REMARK 470 MISSING ATOM
REMARK 470 THE FOLLOWING RESIDUES HAVE MISSING ATOMS (M=MODEL NUMBER;
REMARK 470 RES=RESIDUE NAME; C=CHAIN IDENTIFIER; SSEQ=SEQUENCE NUMBER;
REMARK 470 I=INSERTION CODE):
REMARK 470   M RES CSSEQI  ATOMS
```

Example

```
         1         2         3         4         5         6         7
12345678901234567890123456789012345678901234567890123456789012345678901234567890
REMARK 470
REMARK 470 MISSING ATOM
REMARK 470 THE FOLLOWING RESIDUES HAVE MISSING ATOMS (M=MODEL NUMBER;
REMARK 470 RES=RESIDUE NAME; C=CHAIN IDENTIFIER; SSEQ=SEQUENCE NUMBER;
REMARK 470 I=INSERTION CODE):
REMARK 470   M RES CSSEQI  ATOMS
REMARK 470     ARG A 412    CG   CD   NE   CZ   NH1  NH2
REMARK 470     ARG A 456    CG   CD   NE   CZ   NH1  NH2
REMARK 470     GLU A 486    CG   CD   OE1  OE2
REMARK 470     GLU A 547    CG   CD   OE1  OE2
REMARK 470     GLU A 548    CG   CD   OE1  OE2
REMARK 470     LYS A 606    CG   CD   CE   NZ
REMARK 470     ARG B 456    CG   CD   NE   CZ   NH1  NH2
REMARK 470     ASP B 484    CG   OD1  OD2
REMARK 470     GLN B 485    CG   CD   OE1  NE2
REMARK 470     GLU B 486    CG   CD   OE1  OE2
REMARK 470     ARG B 490    CG   CD   NE   CZ   NH1  NH2
REMARK 470     GLU B 522    CG   CD   OE1  OE2
REMARK 470     ARG B 576    CG   CD   NE   CZ   NH1  NH2
REMARK 470     ASP B 599    CG   OD1  OD2
```

REMARK 500, Geometry and Stereochemistry

Further details on the stereochemistry of the structure. This remark is generated by PDB, but may also be provided by the depositor. Additional subtopics may be added as needed.

Template

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
REMARK 500
REMARK 500 GEOMETRY AND STEREOCHEMISTRY
REMARK 500 SUBTOPIC:
REMARK 500
REMARK 500 FREE TEXT GOES HERE.
```

Subtopic: Close Contacts

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
REMARK 500
REMARK 500 GEOMETRY AND STEREOCHEMISTRY
REMARK 500 SUBTOPIC: CLOSE CONTACTS
REMARK 500
REMARK 500 THE FOLLOWING ATOMS THAT ARE RELATED BY CRYSTALLOGRAPHIC
REMARK 500 SYMMETRY ARE IN CLOSE CONTACT.  SOME OF THESE MAY BE ATOMS
REMARK 500 LOCATED ON SPECIAL POSITIONS IN THE CELL.
REMARK 500
REMARK 500 DISTANCE CUTOFF: 2.2 ANGSTROMS
REMARK 500
REMARK 500  ATM1  RES C  SSEQI   ATM2  RES C  SSEQI  SSYMOP   DISTANCE
REMARK 500   CB   LEU D    68  -  CE   LYS E    76     1656     2.10
REMARK 500   CB   THR D   173  -  O    HOH    1151     4455     1.73
REMARK 500   O    HOH    1151  -  CB   THR D   173     4566     1.73
REMARK 500   CZ   ARG D    64  -  O    HOH    1422     3656     1.75
```

Subtopic: Close Contacts In Same Asymmetric Unit

```
REMARK 500
REMARK 500 GEOMETRY AND STEREOCHEMISTRY
REMARK 500 SUBTOPIC: CLOSE CONTACTS IN SAME ASYMMETRIC UNIT
REMARK 500
REMARK 500 THE FOLLOWING ATOMS ARE IN CLOSE CONTACT.
REMARK 500
REMARK 500  ATM1  RES C  SSEQI   ATM2  RES C  SSEQI           DISTANCE
REMARK 500   O    HOH     761  -  O    ARG      17             1.89
REMARK 500   O    HOH     806  -  N    ARG      88             1.46
```

Subtopic: Non-Cis, Non-Trans

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
REMARK 500
REMARK 500 GEOMETRY AND STEREOCHEMISTRY
REMARK 500 SUBTOPIC: NON-CIS, NON-TRANS
REMARK 500
REMARK 500 THE FOLLOWING PEPTIDE BONDS DEVIATE SIGNIFICANTLY FROM BOTH
REMARK 500 CIS AND TRANS CONFORMATION.  CIS BONDS, IF ANY, ARE LISTED
REMARK 500 ON CISPEP RECORDS.  TRANS IS DEFINED AS 180 +/- 30 AND
REMARK 500 CIS IS DEFINED AS 0 +/- 30 DEGREES.
REMARK 500                                MODEL     OMEGA
REMARK 500 VAL A  123    GLN A  124          0       221.48
REMARK 500 VAL B  123    GLN B  124          0       222.43
```

Subtopic: Chiral Centers

```
          1             2         3         4       5         6        7
12345678901234567890123456789012345678901234567890123456789012345678901234567890
REMARK 500
REMARK 500 GEOMETRY AND STEREOCHEMISTRY
REMARK 500 SUBTOPIC: CHIRAL CENTERS
REMARK 500
REMARK 500 UNEXPECTED CONFIGURATION OF THE FOLLOWING CHIRAL
REMARK 500 CENTER(S) (M=MODEL NUMBER; RES=RESIDUE NAME; C=CHAIN
REMARK 500 IDENTIFIER; SSEQ=SEQUENCE NUMBER; I=INSERTION CODE).
REMARK 500
REMARK 500 STANDARD TABLE:
REMARK 500 FORMAT: (10X,I3,1X,A3,1X,A1,I4,A1,6X,A12)
REMARK 500
REMARK 500    M RES CSSEQI
REMARK 500    0 GLU       1
REMARK 500    0 GLU       1
REMARK 500    0 GLU       1
```

Subtopic: Covalent Bond Angles

```
         1         2         3         4         5         6         7
12345678901234567890123456789012345678901234567890123456789012345678901234567890
REMARK 500
REMARK 500 GEOMETRY AND STEREOCHEMISTRY
REMARK 500 SUBTOPIC: COVALENT BOND ANGLES
REMARK 500
REMARK 500 THE STEREOCHEMICAL PARAMETERS OF THE FOLLOWING RESIDUES
REMARK 500 HAVE VALUES WHICH DEVIATE FROM EXPECTED VALUES BY MORE
REMARK 500 THAN 6*RMSD (M=MODEL NUMBER; RES=RESIDUE NAME; C=CHAIN
REMARK 500 IDENTIFIER; SSEQ=SEQUENCE NUMBER; I=INSERTION CODE).
REMARK 500
REMARK 500 STANDARD TABLE:
REMARK 500 FORMAT: (10X,I3,1X,A3,1X,A1,I4,A1,3(2X,A4,17X,F5.1)
REMARK 500
REMARK 500 EXPECTED VALUES: ENGH AND HUBER, 1991
REMARK 500
REMARK 500  M RES CSSEQI ATM1   ATM2   ATM3
REMARK 500  0 ASP     3   C-1 -  N   -  CA  ANGL. DEV. =  21.7 DEGREES
```

Subtopic: Torsion Angles

```
         1         2         3         4         5         6         7
12345678901234567890123456789012345678901234567890123456789012345678901234567890
REMARK 500
REMARK 500 GEOMETRY AND STEREOCHEMISTRY
REMARK 500 SUBTOPIC: TORSION ANGLES
REMARK 500
REMARK 500 TORSION ANGLES OUTSIDE THE EXPECTED RAMACHANDRAN REGIONS:
REMARK 500 (M=MODEL NUMBER; RES=RESIDUE NAME; C=CHAIN IDENTIFIER;
REMARK 500 SSEQ=SEQUENCE NUMBER; I=INSERTION CODE).
REMARK 500
REMARK 500 STANDARD TABLE:
REMARK 500 FORMAT:(10X,I3,1X,A3,1X,A1,I4,A1,4X,F7.2,3X,F7.2)
REMARK 500
REMARK 500  M RES CSSEQI        PSI       PHI
REMARK 500  0 VAL    26     -174.85   -134.80
REMARK 500  0 MET    61       46.11   -176.53
```

Subtopic: Covalent Bond Lengths

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

```
REMARK 500
REMARK 500 GEOMETRY AND STEREOCHEMISTRY
REMARK 500 SUBTOPIC: COVALENT BOND LENGTHS
REMARK 500
REMARK 500 THE STEREOCHEMICAL PARAMETERS OF THE FOLLOWING RESIDUES
REMARK 500 HAVE VALUES WHICH DEVIATE FROM EXPECTED VALUES BY MORE
REMARK 500 THAN 6*RMSD AND BY MORE THAN 0.150 ANGSTROMS (M=MODEL
REMARK 500 NUMBER; RES=RESIDUE NAME; C=CHAIN IDENTIFIER; SSEQ=SEQUENCE
REMARK 500 NUMBER; I=INSERTION CODE).
REMARK 500
REMARK 500 STANDARD TABLE:
REMARK 500 FORMAT: (10X,I3,1X,A3,1X,A1,I4,A1,1X,2(A4,A1,3X),12X,F5.3)
REMARK 500
REMARK 500 EXPECTED VALUESS: ENGH AND HUBER, 1991
REMARK 500
REMARK 500  M RES CSSEQI ATM1   RES CSSEQI ATM2   DEVIATION
REMARK 500    LYS A  24  CB     LYS A  24  CG       0.269
REMARK 500    LYS A 109  CD     LYS A 109  CE      -0.251
REMARK 500    GLU B  39  CG     GLU B  39  CD       0.158
REMARK 500
REMARK 500 REMARK: NULL
```

REMARK 525, Solvent

Remarks specific to the solvent molecules of the entry.

Template

```
         1         2         3         4         5         6         7
12345678901234567890123456789012345678901234567890123456789012345678901234567890
REMARK 525
REMARK 525 SOLVENT
REMARK 525 FREE TEXT GOES HERE.
```

Example

```
         1         2         3         4         5         6         7
12345678901234567890123456789012345678901234567890123456789012345678901234567890
REMARK 525
REMARK 525 SOLVENT
REMARK 525 MANY OF THE WATER MOLECULES APPEAR TO BE ASSOCIATED WITH
REMARK 525 A SYMMETRY-RELATED MOLECULE.

         1         2         3         4         5         6         7
12345678901234567890123456789012345678901234567890123456789012345678901234567890
REMARK 525
REMARK 525 SOLVENT
REMARK 525
REMARK 525 THE SOLVENT MOLECULES ARE GIVEN CHAIN IDENTIFIERS TO
REMARK 525 INDICATE THE PROTEIN CHAIN TO WHICH THEY ARE MOST CLOSELY
REMARK 525 ASSOCIATED WITH:
REMARK 525   PROTEIN CHAIN  SOLVENT CHAIN
REMARK 525     A              Z
REMARK 525     B              Y
REMARK 525
REMARK 525 THE FOLLOWING SOLVENT MOLECULES LIE FARTHER THAN EXPECTED
REMARK 525 FROM THE PROTEIN OR NUCLEIC ACID MOLECULE AND MAY BE
REMARK 525 ASSOCIATED WITH A SYMMETRY RELATED MOLECULE (M=MODEL
REMARK 525 NUMBER; RES=RESIDUE NAME; C=CHAIN IDENTIFIER; SSEQ=SEQUENCE
REMARK 525 NUMBER; I=INSERTION CODE):
```

```
REMARK 525
REMARK 525 THESE MOLECULES CAN BE PLACED WITHIN  5.00 ANGSTROM OF THE
REMARK 525 OBSERVED OLIGOMER BY APPLYING THE SYMMETRY TRANSFORMATION
REMARK 525 INDICATED.
REMARK 525
REMARK 525  M RES CSSEQI     ORIGINAL COORDINATES   SYMMETRY TRANS.    DIST.
REMARK 525                    X       Y       Z
REMARK 525  1 HOH Z 197   -23.309 -18.431  27.821    002      456      2.85
REMARK 525  1 HOH Z  64   -18.204 -18.469   0.503    002      455      2.91
REMARK 525  1 HOH Z 236   -18.883 -15.861  -2.863    002      455      2.13
REMARK 525  1 HOH Z 185   -28.011   5.875  32.170    001      554      2.60




REMARK 525
REMARK 525 SOLVENT
REMARK 525 THE FOLLOWING SOLVENT MOLECULES LIE FARTHER THAN EXPECTED
REMARK 525 FROM THE PROTEIN OR NUCLEIC ACID MOLECULE AND MAY BE
REMARK 525 ASSOCIATED WITH A SYMMETRY RELATED MOLECULE (M=MODEL
REMARK 525 NUMBER; RES=RESIDUE NAME; C=CHAIN IDENTIFIER; SSEQ=SEQUENCE
REMARK 525 NUMBER; I=INSERTION CODE):
REMARK 525
REMARK 525  M RES CSSEQI
REMARK 525  0 HOH    561      DISTANCE =  5.07 ANGSTROMS
REMARK 525  0 HOH    791      DISTANCE =  5.08 ANGSTROMS


REMARK 525
REMARK 525 SOLVENT
REMARK 525
REMARK 525 THE SOLVENT MOLECULES ARE GIVEN CHAIN IDENTIFIERS TO
REMARK 525 INDICATE THE PROTEIN CHAIN TO WHICH THEY ARE MOST CLOSELY
REMARK 525 ASSOCIATED WITH:
REMARK 525   PROTEIN CHAIN  SOLVENT CHAIN
REMARK 525     A              Z
REMARK 525     B              Y
```

REMARK 550, SEGID (deprecated)

This record has been deprecated.

REMARK 600, Heterogen

Further details on the heterogens in the entry.

Template

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
REMARK 600
REMARK 600 HETEROGEN
REMARK 600 FREE TEXT GOES HERE.
```

Example

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
REMARK 600
REMARK 600 HETEROGEN
REMARK 600
REMARK 600
REMARK 600 FOR METAL ATOM MG    MG A1192  THE COORDINATION ANGLES ARE:
REMARK 600  1 HOH    157Z  O
REMARK 600  2 ASP    144A  OD2       87.3
REMARK 600  3 ADP   1190A  O3B       94.6  87.9
REMARK 600  4 ASP    173A  OD1       88.0  90.6 176.9
REMARK 600  5 HOH    199Z  O         85.1 172.1  90.7  91.2
REMARK 600  6 ADP   1190A  O1A      167.1 105.4  88.0  89.8  82.3
REMARK 600                             1     2     3     4     5
REMARK 600
REMARK 600 HETGROUPS RENAMED OR RENUMBERED
REMARK 600 PRIMARY PUBLICATION            THIS ENTRY
REMARK 600        CA 1 W                   CA B2326
REMARK 600       GOL 2 W                  GOL A1327
REMARK 600       GOL 3 W                  GOL A1328
REMARK 600       GOL 4 W                  GOL B1326
REMARK 600       GOL 5 W                  GOL B1327
```

REMARK 650, Helix

Further details on the helix contents of the entry.

Template

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
REMARK 650
REMARK 650 HELIX
REMARK 650 FREE TEXT GOES HERE.
```

Example

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
REMARK 650
REMARK 650 HELIX
REMARK 650 DETERMINATION METHOD: KDSSP
REMARK 650 THE MAJOR DOMAINS ARE: "N" FOR N-TERMINAL DOMAIN, "B" FOR
REMARK 650 BETA-BARREL DOMAIN, AND "C" FOR C-TERMINAL DOMAIN. "F"
REMARK 650 REFERS TO THE ACTIVE SITE FLAP.  ALPHA HELICES ARE NAMED
REMARK 650 WITH TWO CHARACTERS, THE FIRST REFERRING TO THE DOMAIN
REMARK 650 IN WHICH THEY OCCUR.
```

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
REMARK 650
REMARK 650 HELIX
REMARK 650 DETERMINATION METHOD: AUTHOR PROVIDED.
```

REMARK 700, Sheet

Further details on the sheet contents of the structure. Several standard templates are shown.

Template

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
REMARK 700
REMARK 700 SHEET
REMARK 700 FREE TEXT GOES HERE.
REMARK 700
REMARK 700 SHEET
REMARK 700 DETERMINATION METHOD:
REMARK 700 THE SHEET STRUCTURE OF THIS MOLECULE IS BIFURCATED.  IN
REMARK 700 ORDER TO REPRESENT THIS FEATURE IN THE SHEET RECORDS BELOW,
REMARK 700 TWO SHEETS ARE DEFINED.  STRANDS N1, N2, N3 AND N4 OF SHEET
REMARK 700 XXX AND XXX ARE IDENTICAL.

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

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

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

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

Example

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
REMARK 700
REMARK 700 SHEET
REMARK 700 THE SHEET STRUCTURE OF THIS MOLECULE IS BIFURCATED.  IN
REMARK 700 ORDER TO REPRESENT THIS FEATURE IN THE SHEET RECORDS BELOW,
REMARK 700 TWO SHEETS are defined.  STRANDS 3, 4, AND 5
REMARK 700 OF SHEET *B2A* AND *B2B* ARE IDENTICAL.  STRANDS 3, 4, AND
REMARK 700 5 OF SHEET *B2C* AND *B2D* ARE IDENTICAL.
REMARK 700
REMARK 700 SHEET
REMARK 700 STRANDS 1 TO 4 OF THE BETA-SHEET HAVE GREEK-KEY TOPOLOGY.
REMARK 700 THE SHEET FORMS A FIVE-STRANDED BETA-BARREL WITH BULGES IN
REMARK 700 STRANDS 3 AND 5.  IN ORDER TO REPRESENT THIS FEATURE IN THE
```

```
REMARK 700 SHEET RECORDS BELOW, TWO SHEETS ARE DEFINED.

REMARK 700
REMARK 700 SHEET
REMARK 700 THE SHEET PRESENTED AS S5 ON SHEET RECORDS BELOW IS
REMARK 700 ACTUALLY A 6-STRANDED BETA-BARREL.  THIS IS
REMARK 700 REPRESENTED BY A 7-STRANDED SHEET IN WHICH THE FIRST AND
REMARK 700 LAST STRANDS ARE IDENTICAL.

REMARK 700
REMARK 700 SHEET
REMARK 700 SHEET
REMARK 700 DETERMINATION METHOD: AUTHOR PROVIDED.
```

REMARK 750, Turn

Further details on the turns.

Template

```
         1         2         3         4         5         6         7
12345678901234567890123456789012345678901234567890123456789012345678901234567890
REMARK 750
REMARK 750 TURN
REMARK 750 FREE TEXT GOES HERE.
```

Example

```
         1         2         3         4         5         6         7
12345678901234567890123456789012345678901234567890123456789012345678901234567890
REMARK 750
REMARK 750 TURN
REMARK 750  TURN_ID: T4, TYPE I (ONE OR MORE OF THE PHI, PSI ANGLES
REMARK 750  DEVIATE BY MORE THAN PLUS,MINUS 45 DEGREES FROM THE IDEAL
REMARK 750  VALUES USED BY WILMOT & THORNTON(1989)).
REMARK 750
REMARK 750  TURN_ID: T10, TYPE I (ONE OR MORE OF THE PHI, PSI ANGLES
REMARK 750  DEVIATE BY MORE THAN PLUS,MINUS 45 DEGREES FROM THE IDEAL
REMARK 750  VALUES USED BY WILMOT & THORNTON(1989)).
REMARK 750
REMARK 750  TURN_ID: T16, TYPE VIII (ONE OR MORE OF THE PHI, PSI
REMARK 750  ANGLES DEVIATE BY MORE THAN PLUS,MINUS 45 DEGREES FROM
REMARK 750  THE IDEAL VALUES USED BY WILMOT & THORNTON(1989)).
```

REMARK 800, Site

Further details on the site contents of the entry. Remark 800 is mandatory if site records exist.

Template

```
         1         2         3         4         5         6         7
12345678901234567890123456789012345678901234567890123456789012345678901234567890
REMARK 800
REMARK 800 SITE
```

```
REMARK 800 SITE_IDENTIFIER: FREE TEXT GOES HERE.
REMARK 800 SITE_DESCRIPTION: FREE TEXT GOES HERE.
```

Example

```
         1         2         3         4         5         6         7
12345678901234567890123456789012345678901234567890123456789012345678901234567890
REMARK 800
REMARK 800 SITE
REMARK 800 SITE_IDENTIFIER: RCA
REMARK 800 SITE_DESCRIPTION: DESIGNATED RECOGNITION REGION IN PRIMARY
REMARK 800 REFERENCE.  PROPOSED TO AFFECT SUBSTRATE SPECIFICITY.
REMARK 800
REMARK 800 SITE_IDENTIFIER: RCB
REMARK 800 SITE_DESCRIPTION: DESIGNATED RECOGNITION REGION IN PRIMARY
REMARK 800  REFERENCE.  PROPOSED TO AFFECT SUBSTRATE SPECIFICITY.

         1         2         3         4         5         6         7
12345678901234567890123456789012345678901234567890123456789012345678901234567890
REMARK 800
REMARK 800 SITE
REMARK 800 SITE_IDENTIFIER: AC1
REMARK 800 SITE_DESCRIPTION: BAT BINDING SITE FOR CHAIN A
REMARK 800
REMARK 800 SITE_IDENTIFIER: AC2
REMARK 800 SITE_DESCRIPTION: CA BINDING SITE FOR CHAIN A
REMARK 800
REMARK 800 SITE_IDENTIFIER: AC3
REMARK 800 SITE_DESCRIPTION: CA BINDING SITE FOR CHAIN A
```

REMARK 860, Correction, After Release

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

Template

```
         1         2         3         4         5         6         7
12345678901234567890123456789012345678901234567890123456789012345678901234567890
REMARK 860
REMARK 860 CORRECTION AFTER RELEASE
REMARK 860 FREE TEXT GOES HERE.
```

Example

```
         1         2         3         4         5         6         7
12345678901234567890123456789012345678901234567890123456789012345678901234567890
REMARK 860
REMARK 860 CORRECTION
REMARK 860 CORRECT RESIDUE IDENTIFICATION ON SITE RECORDS.  ADD
REMARK 860 RESIDUE TO SITE RECORDS.  15-JUL-81.
REMARK 860
REMARK 860 CORRECT DATES IN REMARKS 7 AND 16. 15-JAN-82.
REMARK 860
REMARK 860 CORRECT ATOM NAME FOR ATOM 6 FROM CG2 TO CG1.  07-MAR-83.
REMARK 860
REMARK 860 CHANGE RESIDUE 122 FROM ASN TO ASP.  ADD REFERENCE.
```

```
REMARK 860  12-MAY-83.
REMARK 860
REMARK 860 INSERT REVDAT RECORDS. 30-SEP-83.
REMARK 860
REMARK 860 CORRECT CODEN FOR REFERENCE 1.  27-OCT-83.
```

REMARK 900, Related Entries

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

Template

```
         1         2         3         4         5         6         7
12345678901234567890123456789012345678901234567890123456789012345678901234567890
REMARK 900
REMARK 900 RELATED ENTRIES
REMARK 900 FREE TEXT GOES HERE.
```

Example

```
         1         2         3         4         5         6         7
12345678901234567890123456789012345678901234567890123456789012345678901234567890
REMARK 900
REMARK 900 RELATED ENTRIES
REMARK 900 THE BIOMOLECULE RELATED TO THIS ENTRY HAS BEEN GENERATED
REMARK 900 AND IS AVAILABLE AS PDB FILE BIO1ABC.PDB
REMARK 900
REMARK 900 RELATED ENTRIES
REMARK 900 THE STRUCTURE FACTORS FOR THIS EXPERIMENT ARE AVAILABLE AS
REMARK 900 PDB FILE R1ABCSF.ENT

REMARK 900
REMARK 900 RELATED ENTRIES
REMARK 900 THE LIST OF EXPERIMENTAL RESTRAINTS IS AVAILABLE AS PDB
REMARK 900 FILE 1ABC.MR

REMARK 900
REMARK 900 RELATED ENTRIES
REMARK 900 THE BIOMOLECULE IS AVAILABLE AS PDB FILE BIO1ABC.PDB
REMARK 900
REMARK 900 RELATED ENTRIES
REMARK 900 RELATED ID: 2CKI   RELATED DB: PDB
REMARK 900  STRUCTURE OF ULILYSIN, A MEMBER OF THE
REMARK 900  PAPPALYSIN FAMILY OF METZINCIN
REMARK 900  METALLOENDOPEPTIDASES.
```

REMARK 999 Sequence

Further details on the sequence.

For cases where there are gaps in the structure as reflected in missing ATOM records missing N-

terminus and C-terminus residues are delineated in REMARK 999 records, whereas internal structural gaps are represented in SEQADV records. Several cases must be considered when evaluating these REMARK 999 records:

1. The missing N-terminus atoms are not found in the ATOM record as they represent precursor sequence and are not found in the mature protein.

2. The missing N-terminus residues were not found in the density map. Although PDB will attempt to flag these as SEQADV records, we cannot guarantee that they will always be handled uniformly. The primary reason for this inconsistency is that in a number of cases, neither PDB nor the depositors, are certain where chains start and end.

Template

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
REMARK 999
REMARK 999 SEQUENCE
REMARK 999 FREE TEXT GOES HERE.
```

Example

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
REMARK 999
REMARK 999 SEQUENCE
REMARK 999
REMARK 999  REFERENCE
REMARK 999   REFERENCE: PETRA, ET AL., (1971) BIOCHEMISTRY 10, PP
REMARK 999   4023-4025.
REMARK 999
REMARK 999   SHOHAM, G., NECHUSHTAI, R., STEPPUN, J.,NELSON, H.,
REMARK 999   NELSON N., UNPUBLISHED RESULTS.
REMARK 999
REMARK 999   LE HUEROU,I., GUILLOTEAU P., TOULLEC, R., PUIGSERVER, A.,
REMARK 999   WICKER,C., (1991) BIOCHEMICAL, BIOPHYSICAL RESEARCH
REMARK 999   COMM., 175, PP 110 - 116.
REMARK 999
REMARK 999 THE SEQUENCE USED IS THAT PROVIDED BY THE CDNA, WHICH
REMARK 999 CORRECTS SEVERAL ASP/ASN AND GLU/GLN MISASSIGNMENTS.
REMARK 999
REMARK 999 SEQUENCE
REMARK 999 MET A    1  - MET A    1  - MISSING FROM SWS    P10599
REMARK 999
REMARK 999 THR AT POSITION 74 WAS FOUND BY WOLMAN ET AL., JOURNAL OF
REMARK 999 BIOCHEMISTRY 263, 15506 (1988).

REMARK 900
REMARK 900 RELATED ENTRIES
REMARK 900 RELATED ID: 1CJD   RELATED DB: PDB
REMARK 900  THE BACTERIOPHAGE PRD1 COAT PROTEIN, P3, IS
REMARK 900  STRUCTURALLY SIMILAR TO HUMAN ADENOVIRUS HEXON
REMARK 900 RELATED ID: 1HB5   RELATED DB: PDB
REMARK 900  QUASI-ATOMIC RESOLUTION MODEL OF BACTERIOPHAGE
REMARK 900  PRD1 P3-SHELL, OBTAINED BY COMBINED CRYO-
REMARK 900  EM AND X-RAY CRYSTALLOGRAPHY.
```

3. Primary Structure Section

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

DBREF

Overview

The DBREF record provides cross-reference links between PDB sequences and the corresponding database entry or entries.

Record Format

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"DBREF "	
8 - 11	IDcode	idCode	ID code of this entry.
13	Character	chainID	Chain identifier.
15 - 18	Integer	seqBegin	Initial sequence number of the PDB sequence segment.
19	AChar	insertBegin	Initial insertion code of the PDB sequence segment.
21 - 24	Integer	seqEnd	Ending sequence number of the PDB sequence segment.
25	AChar	insertEnd	Ending insertion code of the PDB sequence segment.
27 - 32	LString	database	Sequence database name.
34 - 41	LString	dbAccession	Sequence database accession code.
43 - 54	LString	dbIdCode	Sequence database identification code.
56 - 60	Integer	dbseqBegin	Initial sequence number of the database segment.
61	AChar	idbnsBeg	Insertion code of initial residue of the segment, if PDB is the reference.
63 - 67	Integer	dbseqEnd	Ending sequence number of the database segment.
68	AChar	dbinsEnd	Insertion code of the ending residue of the segment, if PDB is the reference.

Details

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

* The DBREF record was designed to account for these differences by providing explicit correlations between sequences as given in the SEQRES records and the sequence database entry. Several cases are easily represented by means of pointers between the databases using DBREF. PDB entries containing heteropolymers are linked to different sequence database entries.

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

Database name	database (code in columns 27 - 32)
GenBank	GB
Protein Data Bank	PDB
Protein Identification Resource	PIR
SWISS-PROT	SWS
TREMBL	TREMBL
UNIPROT	UNP

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

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

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

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

Verification/Validation/Value Authority Control

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

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

Relationships to Other Record Types

DBREF represents the sequence as found in SEQRES records.

Example

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
DBREF  2J83 A   61   322  UNP    Q8TL28   Q8TL28_METAC    61    322
DBREF  2J83 B   61   322  UNP    Q8TL28   Q8TL28_METAC    61    322
DBREF  1ABC B    1B   36  PDB    1ABC     1ABC             1B    36
DBREF  3AKY      3   220  SWS    P07170   KAD1_YEAST       5    222

DBREF  1HAN      2   288  GB     397884   X66122           1    287

DBREF  3HSV A    1    92  SWS    P22121   HSF_KLULA      193    284
DBREF  3HSV B    1    92  SWS    P22121   HSF_KLULA      193    284

DBREF  1ARL      1   307  SWS    P00730   CBPA_BOVIN     111    417
```

SEQADV

Overview

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

Record Format

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

Details

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

* SEQADV describes conflicts between residue sequences given by SEQRES records and those in the appropriate sequence database entry.

* Some of the possible conflict comments:
    Cloning artifact

Conflict
Engineered
Disordered
Variant
Insertion
Deletion
Microheterogeneity
D-configuration

* When conflicts arise which are not classifiable by these terms, a reference to either a published paper, a PDB entry, or a REMARK within the entry is given.

* Finally, the comment "SEE REMARK 999" is included when the explanation for the conflict is too long to fit the SEQADV record.

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

Verification/Validation/Value Authority Control

SEQADV records are automatically generated by the PDB.

Relationships to Other Record Types

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

REMARK 999 contains text that explains discrepancies when the explanation is too lengthy to fit in SEQADV.

Example

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
SEQADV 2J83 ALA A  269  UNP  Q8TL28    CYS   269 ENGINEERED MUTATION
SEQADV 2J83 ALA B  269  UNP  Q8TL28    CYS   269 ENGINEERED MUTATION

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

SEQRES

Overview

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

Record Format

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"SEQRES"	
9 - 10	Integer	serNum	Serial number of the SEQRES record for the current chain. Starts at 1 and increments by one each line. Reset to 1 for each chain.
12	Character	chainID	Chain identifier. This may be any single legal character, including a blank which is used if there is only one chain.
14 - 17	Integer	numRes	Number of residues in the chain. This value is repeated on every record.
20 - 22	Residue name	resName	Residue name.
24 - 26	Residue name	resName	Residue name.
28 - 30	Residue name	resName	Residue name.
32 - 34	Residue name	resName	Residue name.
36 - 38	Residue name	resName	Residue name.
40 - 42	Residue name	resName	Residue name.
44 - 46	Residue name	resName	Residue name.
48 - 50	Residue name	resName	Residue name.
52 - 54	Residue name	resName	Residue name.
56 - 58	Residue name	resName	Residue name.
60 - 62	Residue name	resName	Residue name.
64 - 66	Residue name	resName	Residue name.
68 - 70	Residue name	resName	Residue name.

Verification/Validation/Value Authority Control

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

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

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

Example

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

Known Problems

Polysaccharides do not lend themselves to being represented in SEQRES.

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

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

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

MODRES

Overview

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

Record Format

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

Details

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

* MODRES is mandatory for when modified standard residues exist in the entry.

* Examples of some modification descriptions:

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

* MODRES is not required if coordinate records are not provided for the modified residue.

* D-amino acids are given their own resName , i.e., DAL for D-alanine. This resName appears in the SEQRES records, and has the associated SEQADV, MODRES, HET, and FORMUL records. The coordinates are given as HETATMs within the ATOM records and occur in the correct order within the chain. This ordering is an exception to the stated Order of Records.

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

Verification/Validation/Value Authority Control

MODRES is generated by the PDB.

Relationships to Other Record Types

MODRES maps ATOM and HETATM records to the standard residue names. SEQADV, HET, and FORMUL may also appear.

Example

```
          1         2         3         4         5         6         7
12345678901234567890123456789012345678901234567890123456789012345678901234567890
MODRES 1ABC ASN A   22A ASN  GLYCOSYLATION SITE
MODRES 2ABC TTQ A   50A TRP  POST-TRANSLATIONAL MODIFICATION
MODRES 3ABC DAL A   32  ALA  POST-TRANSLATIONAL MODIFICATION,D-ALANINE
MODRES 3ABC DAL B   32  ALA  POST-TRANSLATIONAL MODIFICATION,D-ALANINE
```

4. Heterogen Section

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

HET

Overview

HET records are used to describe non-standard residues, such as prosthetic groups, inhibitors, solvent molecules, and ions for which coordinates are supplied. Groups are considered HET if they are not part of a biological polymer described in SEQRES and considered to be a molecule bound to the polymer, or they are a chemical species that constitutes part of a biological polymer that is not one of the following:
* not one of the standard amino acids, and
* not one of the nucleic acids (C, G, A, T, U, and I), and
* not an unknown amino acid or nucleic acid where UNK is used to indicate the unknown residue name.

Het records also describe heterogens for which the chemical identity is unknown, in which case the group is assigned the hetID UNL (Unknown Ligand).

Record Format

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"HET "	
8 - 10	LString(3)	hetID	Het identifier, right-justified.
13	Character	ChainID	Chain identifier.
14 - 17	Integer	seqNum	Sequence number.
18	AChar	iCode	Insertion code.
21 - 25	Integer	numHetAtoms	Number of HETATM records for the group present in the entry.
31 - 70	String	text	Text describing Het group.

Details

* Each HET group is assigned a hetID of not more than three (3) alphanumeric characters. The sequence number, chain identifier, insertion code, and number of coordinate records are given for each occurrence of the HET group in the entry. The chemical name of the HET group is given in the HETNAM record and synonyms for the chemical name are given in the HETSYN records, see http://deposit.pdb.org/public-component-erf.cif.

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

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

* PDB entries do not have HET records for water molecules, deuterated water, or methanol (when used as solvent).

* Unknown atoms or ions will be represented as UNX with the chemical formula X1. Unknown ligands are UNL; unknown amino acids are UNK.

Verification/Validation/Value Authority Control

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

Each unique hetID represents a unique molecule.

Relationships to Other Record Types

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

Example

```
         1         2         3         4         5         6         7
12345678901234567890123456789012345678901234567890123456789012345678901234567890
HET    TRS    975        8
HET    STA  I   4       25     PART_OF: HIV INHIBITOR;

HET    FUC  Y   1       10     PART_OF: NONOATE COMPLEX; L-FUCOSE
HET    GAL  Y   2       11     PART_OF: NONOATE COMPLEX
HET    NAG  Y   3       15
HET    FUC  Y   4       10
HET    NON  Y   5       12

HET    UNX  A 161        1     PSEUDO CARBON ATOM OF UNKNOWN LIGAND
HET    UNX  A 162        1     PSEUDO CARBON ATOM OF UNKNOWN LIGAND
HET    UNX  A 163        1     PSEUDO CARBON ATOM OF UNKNOWN LIGAND
```

HETNAM

Overview

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

Record Format

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

Details

* Each hetID is assigned a unique chemical name for the HETNAM record, see http://deposit.pdb.org/public-component-erf.cif.

* Other names for the group are given on HETSYN records.

* PDB follows IUPAC/IUB naming conventions to describe groups systematically.

* The special character “~” is used to indicate superscript in a heterogen name. For example: N^6 will be listed in the HETNAM section as N~6~, with the ~ character indicating both the start and end of the superscript in the name, e.g.,

```
N-(BENZYLSULFONYL)SERYL-N~1~-{4-[AMINO(IMINO)METHYL]BENZYL}GLYCINAMIDE
```

* Continuation of chemical names onto subsequent records is allowed.

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

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

Relationships to Other Record Types

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

Example

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
HETNAM     GLC GLUCOSE
HETNAM     SAD BETA-METHYLENE SELENAZOLE-4-CARBOXAMIDE ADENINE
HETNAM  2  SAD DINUCLEOTIDE

HETNAM     UNX UNKNOWN ATOM OR ION
HETNAM     UNL UNKNOWN LIGAND

HETNAM     CYE 45-(3-AMINOPROPYL)-5,11,22,28,34-PENTAMETHYL-3,9,15,
HETNAM   2 CYE  20,26,32,38,43-OCTAOXO-2,5,8,14,19,22,25,28,31,34,37,
HETNAM   3 CYE  42,45,48-TETRADECAAZA-11-AZONIAHEPTACYCLO[42.2.1.1~4,
HETNAM   4 CYE  7~.1~10,13~.1~21,24~.1~27,30~.1~33,36~]DOPENTACONTA-
HETNAM   5 CYE  1(46),4(52),6,10(51),12,21(50),23,27(49),29,33(48),35,
HETNAM   6 CYE  44(47)-DODECAENE
```

HETSYN

Overview

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

Record Format

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

Details

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

Verification/Validation/Value Authority Control

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

Relationships to Other Record Types

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

Example

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
HETSYN     NAD NICOTINAMIDE ADENINE DINUCLEOTIDE
HETSYN     COA COA
HETSYN     CMP CYCLIC AMP; CYCLIC ADENOSINE MONOPHOSPHATE

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

FORMUL

Overview

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

Record Format

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

Details

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

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

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

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

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

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

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

Verification/Validation/Value Authority Control

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

Relationships to Other Record Types

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

Example

```
         1         2         3         4         5         6         7
12345678901234567890123456789012345678901234567890123456789012345678901234567890
FORMUL   2  SO4    2(O4 S1 2-)
FORMUL   3  GLC    C6 H12 O6
FORMUL   3  FOL    2(C19 H17 N7 O6 2-)
FORMUL   4   CL    2(CL1 1-)
FORMUL   5   CA    CA1 2+

FORMUL   1  ACE    C2 H3 O1
FORMUL   2  ACE    C2 H3 O1

FORMUL   8  HOH   *463(H2 O1)
```

Known Problems

Partially deuterated centers are not well represented in this record.

5. Secondary Structure Section

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

HELIX

Overview

HELIX records are used to identify the position of helices in the molecule. Helices are 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 an alphanumeric character helix identifier.
16 - 18	Residue name	initResName	Name of the initial residue.
20	Character	initChainID	Chain identifier for the chain containing this helix.
22 - 25	Integer	initSeqNum	Sequence number of the initial residue.
26	AChar	initICode	Insertion code of the initial residue.
28 - 30	Residue name	endResName	Name of the terminal residue of the helix.
32	Character	endChainID	Chain identifier for the chain containing this helix.
34 - 37	Integer	endSeqNum	Sequence number of the terminal residue.
38	AChar	endICode	Insertion code of the terminal residue.
39 - 40	Integer	helixClass	Helix class (see below).
41 - 70	String	comment	Comment about this helix.
72 - 76	Integer	length	Length of this helix.

Details

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

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

* Helices are classified as follows:

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

Relationships to Other Record Types

There may be related information in the REMARKs.

Example

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890123456
HELIX    1  HA GLY A   86  GLY A   94  1                                   9
HELIX    2  HB GLY B   86  GLY B   94  1                                   9
```

SHEET

Overview

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

Record Format

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"SHEET "	
8 - 10	Integer	strand	Strand number which starts at 1 for each strand within a sheet and increases by one.
12 - 14	LString(3)	sheetID	Sheet identifier.
15 - 16	Integer	numStrands	Number of strands in sheet.
18 - 20	Residue name	initResName	Residue name of initial residue.
22	Character	initChainID	Chain identifier of initial residue in strand.
23 - 26	Integer	initSeqNum	Sequence number of initial residue in strand.
27	AChar	initICode	Insertion code of initial residue in strand.
29 - 31	Residue name	endResName	Residue name of terminal residue.
33	Character	endChainID	Chain identifier of terminal residue.
34 - 37	Integer	endSeqNum	Sequence number of terminal residue.
38	AChar	endICode	Insertion code of terminal residue.
39 - 40	Integer	sense	Sense of strand with respect to previous strand in the sheet. 0 if first strand, 1 if parallel, -1 if anti-parallel.
42 - 45	Atom	curAtom	Registration. Atom name in current strand.
46 - 48	Residue name	curResName	Registration. Residue name in current strand.
50	Character	curChainId	Registration. Chain identifier in current strand.
51 - 54	Integer	curResSeq	Registration. Residue sequence number in current strand.

55	AChar	curICode	Registration. Insertion code in current strand.
57 - 60	Atom	prevAtom	Registration. Atom name in previous strand.
61 - 63	Residue name	prevResName	Registration. Residue name in previous strand.
65	Character	prevChainId	Registration. Chain identifier in previous strand.
66 - 69	Integer	prevResSeq	Registration. Residue sequence number in previous strand.
70	AChar	prevICode	Registration. Insertion code in previous strand.

Details

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

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

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

* 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.

Relationships to Other Record Types

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

Example

```
         1         2         3         4         5         6         7
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

Overview

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

Record Format

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

Details

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

* The initial residue is the N-terminus.

Verification/Validation/Value Authority Control

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

residues exist in the ATOM records.

Relationships to Other Record Types

There may be related information in the REMARKs.

Example

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

6. Connectivity Annotation Section

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

SSBOND

Overview

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

Record Format

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

Details

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

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

Verification/Validation/Value Authority Control

PDB processing programs generate these records automatically

Relationships to Other Record Types

CONECT records are generated for the disulfide bonds when SG atoms of both cysteines are present in the coordinate records.

Example

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890123456789012
SSBOND   1 CYS E   48    CYS E   51                          2555
SSBOND   2 CYS E  252    CYS E  285

SSBOND   1 CYS A  250    CYS A  277                          1555   1555
SSBOND   2 CYS B  250    CYS B  277                          1555   1555
```

Known Problems

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

LINK

Overview

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

Record Format

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"LINK "	
13 - 16	Atom	name1	Atom name.
17	Character	altLoc1	Alternate location indicator.
18 - 20	Residue name	resName1	Residue name.
22	Character	chainID1	Chain identifier.
23 - 26	Integer	resSeq1	Residue sequence number.
27	AChar	iCode1	Insertion code.
43 - 46	Atom	name2	Atom name.
47	Character	altLoc2	Alternate location indicator.
48 - 50	Residue name	resName2	Residue name.
52	Character	chainID2	Chain identifier.
53 - 56	Integer	resSeq2	Residue sequence number.
57	AChar	iCode2	Insertion code.
60 - 65	SymOP	sym1	Symmetry operator for 1st atom.
67 - 72	SymOP	sym2	Symmetry operator for 2nd atom.

Details

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

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

* Non-standard linkages between residues, e.g., side-chain to side-chain, are listed.

* Each LINK record specifies one linkage.

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

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

* For NMR entries only one set (or model) of LINK records will be supplied.

Verification/Validation/Value Authority Control

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

Relationships to Other Record Types

CONECT records are generated from LINKs when both atoms are present in the entry. If symmetry operators are given to generate one of the residues involved in the bond, REMARK 290 defines the symmetry transformation.

Example

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890123456789012
LINK         O1  DDA     1                  C3  DDL     2
LINK        MN    MN   391                  OE2 GLU   217             2565

LINK         C   LYS A 296                  N   CME A 297     1555   1555
LINK         C   CME A 297                  N   MET A 298     1555   1555
LINK        CA    CA A 997                  O   GLN A 262     1555   1555
LINK        CA    CA A 997                  O   TRP A 240     1555   1555
LINK        CA    CA A 997                  O   HOH Z 169     1555   1555
LINK        CA    CA A 997                  O   PRO A 249     1555   1555
```

CISPEP

Overview

CISPEP records specify the prolines and other peptides found to be in the cis conformation. This record
replaces the use of footnote records to list cis peptides.

Record Format

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"CISPEP"	
8 - 10	Integer	serNum	Record serial number.
12 - 14	LString(3)	pep1	Residue name.
16	Character	chainID1	Chain identifier.
18 - 21	Integer	seqNum1	Residue sequence number.
22	AChar	icode1	Insertion code.
26 - 28	LString(3)	pep2	Residue name.
30	Character	chainID2	Chain identifier.
32 - 35	Integer	seqNum2	Residue sequence number.
36	AChar	icode2	Insertion code.
44 - 46	Integer	modNum	Identifies the specific model.
54 - 59	Real(6.2)	measure	Measure of the angle in degrees.

Details

* Cis peptides are those with omega angles of 0°±30°. Deviations larger than 30° are listed in REMARK 500.

* Each cis peptide is listed on a separate line, with an incrementally ascending sequence number.

Verification/Validation/Value Authority Control

PDB generates these records automatically.

Relationships to Other Record Types

Peptide bonds which deviate significantly from either cis or trans conformation are annotated in REMARK 500.

Example

```
         1         2         3         4         5         6         7
12345678901234567890123456789012345678901234567890123456789012345678901234567890
CISPEP   1 GLY A  116    GLY A  117          0        18.50
CISPEP   2 THR D   92    PRO D   93          0       359.80
```

7. Miscellaneous Features Section

The miscellaneous features section may describe features in the molecule such as environments surrounding a non-standard residue or an active site. Other features may be described in the remarks section but are not given a specific record type so far.

SITE

Overview

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

Record Format

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"SITE "	
8 - 10	Integer	seqNum	Sequence number.
12 - 14	LString(3)	siteID	Site name.
16 - 17	Integer	numRes	Number of residues comprising site.
19 - 21	Residue name	resName1	Residue name for first residue comprising site.
23	Character	chainID1	Chain identifier for first residue comprising site.
24 - 27	Integer	seq1	Residue sequence number for first residue comprising site.
28	AChar	iCode1	Insertion code for first residue comprising site.
30 - 32	Residue name	resName2	Residue name for second residue comprising site.
34	Character	chainID2	Chain identifier for second residue comprising site.
35 - 38	Integer	seq2	Residue sequence number for second residue comprising site.
39	AChar	iCode2	Insertion code for second residue comprising site.
41 - 43	Residue name	resName3	Residue name for third residue comprising site.
45	Character	chainID3	Chain identifier for third residue comprising site.
46 - 49	Integer	seq3	Residue sequence number for third residue comprising site.
50	AChar	iCode3	Insertion code for third residue comprising site.
52 - 54	Residue name	resName4	Residue name for fourth residue comprising site.
56	Character	chainID4	Chain identifier for fourth residue comprising site.
57 - 60	Integer	seq4	Residue sequence number for fourth residue comprising site.
61	AChar	iCode4	Insertion code for fourth residue comprising site.

Details

* Site records specify residues comprising catalytic, cofactor, anticodon, regulatory or other important sites or environments surrounding ligands present in the structure.

* The sequence number (columns 8 - 10) is reset to 1 for each new site.

* SITE identifiers (columns 12 - 14) should be fully explained in a remark.

* If a site is comprised of more than four residues, these may be specified on additional records bearing the same site identifier.

* SITE records can include HET groups.

Verification/Validation/Value Authority Control

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

Relationships to Other Record Types

Each listed SITE needs a corresponding REMARK 800 that details its significance.

Example

```
         1         2         3         4         5         6         7
12345678901234567890123456789012345678901234567890123456789012345678901234567890
SITE     1 DTA  3 ASP A  25  THR A  26  GLY A  27
SITE     1 DTB  3 ASP B  25  THR B  26  GLY B  27
SITE     1   A  4   U A  44    C A  46    G A  61    U A 118
SITE     1 ZN1  5 CYS A  97  CYS A 100  CYS A 103  CYS 1 111
SITE     2 ZN1  5  ZN A 375
```

8. Crystallographic and Coordinate Transformation Section

This section describes the geometry of the crystallographic experiment and the coordinate system transformations.

CRYST1

Overview

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

Record Format

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

Details

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

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

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

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

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

As an example, given two chains A and B, each with a different sequence, and the space group P

2 that has two equipoints in the standard unit cell, the following table gives the correct Z value.

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

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

Verification/Validation/Value Authority Control

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

Relationships to Other Record Types

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

Example

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
CRYST1   52.000   58.600   61.900  90.00  90.00  90.00 P 21 21 21    8
CRYST1    1.000    1.000    1.000  90.00  90.00  90.00 P 1           1

CRYST1   42.544   69.085   50.950  90.00  95.55  90.00 P 1 21 1      2
```

Known Problems

No standard deviations are given.

ORIGXn

Overview

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

Record Format

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

Details

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

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

Xsub = O11X + O12Y + O13Z + T1

Ysub = O21X + O22Y + O23Z + T2

Zsub = O31X + O32Y + O33Z + T3

Verification/Validation/Value Authority Control

If the coordinates are submitted in the same orthogonal Angstrom coordinate frame as they appear in the entry (the usual case), then ORIGX is an identity matrix with a null translation vector.

If the transformation is not an identity matrix with a null translation vector, then applying this transformation to the coordinates in the entry yields the coordinates in the original deposited file.

Relationships to Other Record Types

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

Example

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

SCALEn

Overview

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

Record Format

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

Details

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

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

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

xfrac = S11X + S12Y + S13Z + U1

yfrac = S21X + S22Y + S23Z + U2

zfrac = S31X + S32Y + S33Z + U3

* For NMR, fiber diffraction - fiber sample, and theoretical model entries, SCALE is given as an

identity matrix with no translation.

Verification/Validation/Value Authority Control

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

Relationships to Other Record Types

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

Example

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

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

Details

* The 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 given transformation are contained in the file, the iGiven field is set to 1. Otherwise, this field is blank.

* A corresponding REMARK must appear which describes the transformation.

Verification/Validation/Value Authority Control

The PDB verifies all MTRIX records using records from the author and review.

Relationships to Other Record Types

A corresponding REMARK must appear which describes the transformation.

Example

```
         1         2         3         4         5         6         7
12345678901234567890123456789012345678901234567890123456789012345678890
MTRIX1   1 -1.000000  0.000000  0.000000         0.00000    1
MTRIX2   1  0.000000  1.000000  0.000000         0.00000    1
MTRIX3   1  0.000000  0.000000 -1.000000         0.00000    1
```

TVECT

Overview

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

Record Format

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

Details

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

* A corresponding REMARK describing the structure must appear.

Verification/Validation/Value Authority Control

PDB applies the translation and checks the generated molecule.

Relationships to Other Record Types

A corresponding REMARK describing the structure must appear.

Example

```
         1         2         3         4         5         6         7
12345678901234567890123456789012345678901234567890123456789012345678901234567890
TVECT    1   0.00000   0.00000  28.30000
```

9. Coordinate Section

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

MODEL

Overview

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

Record Format

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

Details

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

* The numbering of models is sequential beginning with 1.

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

Verification/Validation/Value Authority Control

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

Relationships to Other Record Types

Each MODEL must have a corresponding ENDMDL record.

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

Example

```
         1         2         3         4         5         6         7         8
12345678901234567890123456789012345678901234567890123456789012345678901234567890
MODEL        1
ATOM      1  N   ALA     1      11.104   6.134  -6.504  1.00  0.00           N
ATOM      2  CA  ALA     1      11.639   6.071  -5.147  1.00  0.00           C
...
...
ATOM    293 1HG  GLU    18     -14.861  -4.847   0.361  1.00  0.00           H
ATOM    294 2HG  GLU    18     -13.518  -3.769   0.084  1.00  0.00           H
TER     295      GLU    18
ENDMDL
MODEL        2
ATOM    296  N   ALA     1      10.883   6.779  -6.464  1.00  0.00           N
ATOM    297  CA  ALA     1      11.451   6.531  -5.142  1.00  0.00           C
...
...
ATOM    588 1HG  GLU    18     -13.363  -4.163  -2.372  1.00  0.00           H
ATOM    589 2HG  GLU    18     -12.634  -3.023  -3.475  1.00  0.00           H
TER     590      GLU    18
ENDMDL
```

ATOM

Overview

The ATOM records present the atomic coordinates for standard residues (see http://deposit.pdb.org/public-component-erf.cif). They also present the occupancy and temperature factor for each atom. Heterogen coordinates use the HETATM record type. The element symbol is always present on each ATOM record; segment identifier and charge are optional.

Record Format

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

47 - 54	Real(8.3)	z	Orthogonal coordinates for Z in Angstroms
55 - 60	Real(6.2)	occupancy	Occupancy.
61 - 66	Real(6.2)	tempFactor	Temperature factor.
77 - 78	LString(2)	element	Element symbol, right-justified.
79 - 80	LString(2)	charge	Charge on the atom.

Details

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

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

* No ordering is specified for polysaccharides.

* The list of ATOM records in a chain is terminated by a TER record.

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

* If an atom is provided in more than one position, then a non-blank alternate location indicator must be used as the alternate location indicator for each of the positions. Within a residue all atoms that are associated with each other in a given conformation are assigned the same alternate position indicator.

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

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

In the case of a large heterogen groups which are disordered, the atoms for each conformer are listed together.

* The insertion code is commonly used in sequence numbering

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

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

* Columns 77 - 78 contain the atom's element symbol (as given in the periodic table), right-justified.

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

Verification/Validation/Value Authority Control

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

Relationships to Other Record Types

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

Example

```
         1         2         3         4         5         6         7         8
12345678901234567890123456789012345678901234567890123456789012345678901234567890
ATOM    145  N   VAL A  25      32.433  16.336  57.540  1.00 11.92           N
ATOM    146  CA  VAL A  25      31.132  16.439  58.160  1.00 11.85           C
ATOM    147  C   VAL A  25      30.447  15.105  58.363  1.00 12.34           C
ATOM    148  O   VAL A  25      29.520  15.059  59.174  1.00 15.65           O
ATOM    149  CB AVAL A  25      30.385  17.437  57.230  0.28 13.88           C
ATOM    150  CB BVAL A  25      30.166  17.399  57.373  0.72 15.41           C
ATOM    151  CG1AVAL A  25      28.870  17.401  57.336  0.28 12.64           C
ATOM    152  CG1BVAL A  25      30.805  18.788  57.449  0.72 15.11           C
ATOM    153  CG2AVAL A  25      30.835  18.826  57.661  0.28 13.58           C
ATOM    154  CG2BVAL A  25      29.909  16.996  55.922  0.72 13.25           C
```

Known Problems

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

SIGATM

Overview

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

Record Format

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

Details

* Columns 7 - 27 and 73 - 80 are identical to the corresponding ATOM/HETATM record.

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

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

Verification/Validation/Value Authority Control

The depositor provides SIGATM records, PDB verifies their format.

Relationships to Other Record Types

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

Example

```
         1         2         3         4         5         6         7         8
12345678901234567890123456789012345678901234567890123456789012345678901234567890
ATOM    230  N   PRO     15      20.860  29.640  13.460  1.00 12.20           N
SIGATM  230  N   PRO     15       0.040   0.030   0.030  0.00  0.00           N
ATOM    231  CA  PRO     15      22.180  29.010  12.960  1.00 14.70           C
SIGATM  231  CA  PRO     15       0.060   0.040   0.050  0.00  0.00           C
ATOM    232  C   PRO     15      23.170  30.090  12.670  1.00 19.10           C
SIGATM  232  C   PRO     15       0.080   0.070   0.060  0.00  0.00           C
ATOM    233  O   PRO     15      24.360  29.860  12.670  1.00 17.50           O
SIGATM  233  O   PRO     15       0.040   0.030   0.030  0.00  0.00           O
ATOM    234  CB  PRO     15      21.710  28.220  11.640  1.00 17.70           C
SIGATM  234  CB  PRO     15       0.060   0.040   0.050  0.00  0.00           C
ATOM    235  CG  PRO     15      20.470  28.710  11.590  1.00 23.90           C
SIGATM  235  CG  PRO     15       0.080   0.060   0.060  0.00  0.00           C
ATOM    236  CD  PRO     15      19.640  29.320  12.660  1.00 15.50           C
SIGATM  236  CD  PRO     15       0.060   0.040   0.050  0.00  0.00           C
ATOM    237  HA  PRO     15      22.630  28.400  13.620  1.00 14.70           H
ATOM    238 1HB  PRO     15      22.240  28.540  10.860  1.00 17.70           H
ATOM    239 2HB  PRO     15      21.670  27.240  11.840  1.00 17.70           H
ATOM    240 1HG  PRO     15      20.360  29.240  10.740  1.00 23.90           H
ATOM    241 2HG  PRO     15      19.900  28.120  11.020  1.00 23.90           H
ATOM    242 1HD  PRO     15      19.230  30.160  12.320  1.00 15.50           H
ATOM    243 2HD  PRO     15      19.120  28.600  13.120  1.00 15.50           H
```

ANISOU

Overview

The ANISOU records present the anisotropic temperature factors.

Record Format

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"ANISOU"	
7 - 11	Integer	serial	Atom serial number.
13 - 16	Atom	name	Atom name.
17	Character	altLoc	Alternate location indicator
18 - 20	Residue name	resName	Residue name.
22	Character	chainID	Chain identifier.
23 - 26	Integer	resSeq	Residue sequence number.
27	AChar	iCode	Insertion code.
29 - 35	Integer	u[0][0]	U(1,1)
36 - 42	Integer	u[1][1]	U(2,2)
43 - 49	Integer	u[2][2]	U(3,3)
50 - 56	Integer	u[0][1]	U(1,2)
57 - 63	Integer	u[0][2]	U(1,3)
64 - 70	Integer	u[1][2]	U(2,3)
77 - 78	LString(2)	element	Element symbol, right-justified.
79 - 80	LString(2)	charge	Charge on the atom.

Details

* Columns 7 - 27 and 73 - 80 are identical to the corresponding ATOM/HETATM record.

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

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

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

Verification/Validation/Value Authority Control

The depositor provides ANISOU records, and the PDB verifies their format.

Relationships to Other Record Types

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

Example

```
         1         2         3         4         5         6         7         8
12345678901234567890123456789012345678901234567890123456789012345678901234567890
ATOM    107  N   GLY    13      12.681  37.302 -25.211 1.000 15.56           N
ANISOU  107  N   GLY    13     2406   1892   1614    198    519   -328       N
ATOM    108  CA  GLY    13      11.982  37.996 -26.241 1.000 16.92           C
ANISOU  108  CA  GLY    13     2748   2004   1679    -21    155   -419       C
ATOM    109  C   GLY    13      11.678  39.447 -26.008 1.000 15.73           C
ANISOU  109  C   GLY    13     2555   1955   1468     87    357   -109       C
ATOM    110  O   GLY    13      11.444  40.201 -26.971 1.000 20.93           O
ANISOU  110  O   GLY    13     3837   2505   1611    164   -121    189       O
ATOM    111  N   ASN    14      11.608  39.863 -24.755 1.000 13.68           N
ANISOU  111  N   ASN    14     2059   1674   1462     27    244    -96       N
```

SIGUIJ

Overview

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

Record Format

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

Details

* Columns 7 - 27 and 73 - 80 are identical to the corresponding ATOM/HETATM record.

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

Verification/Validation/Value Authority Control

The depositor provides SIGUIJ records, PDB verifies their format.

Relationships to Other Record Types

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

Example

```
         1         2         3         4         5         6         7         8
12345678901234567890123456789012345678901234567890123456789012345678901234567890
ATOM    107  N   GLY    13      12.681  37.302 -25.211 1.000 15.56           N
ANISOU  107  N   GLY    13     2406   1892   1614    198    519   -328       N
SIGUIJ  107  N   GLY    13       10     10     10     10     10     10       N
ATOM    108  CA  GLY    13      11.982  37.996 -26.241 1.000 16.92           C
ANISOU  108  CA  GLY    13     2748   2004   1679    -21    155   -419       C
SIGUIJ  108  CA  GLY    13       10     10     10     10     10     10       C
ATOM    109  C   GLY    13      11.678  39.447 -26.008 1.000 15.73           C
ANISOU  109  C   GLY    13     2555   1955   1468     87    357   -109       C
SIGUIJ  109  C   GLY    13       10     10     10     10     10     10       C
ATOM    110  O   GLY    13      11.444  40.201 -26.971 1.000 20.93           O
ANISOU  110  O   GLY    13     3837   2505   1611    164   -121    189       O
SIGUIJ  110  O   GLY    13       10     10     10     10     10     10       O
ATOM    111  N   ASN    14      11.608  39.863 -24.755 1.000 13.68           N
ANISOU  111  N   ASN    14     2059   1674   1462     27    244    -96       N
SIGUIJ  111  N   ASN    14       10     10     10     10     10     10       N
```

TER

Overview

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

Record Format

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

Details

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

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

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

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

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

Verification/Validation/Value Authority Control

TER must appear at the end carboxy or 3' of a chain. For proteins, there is usually a terminal oxygen, labeled OXT. The validation program checks for the occurrence of TER and OXT records.

Relationships to Other Record Types

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

Example

```
         1         2         3         4         5         6         7         8
12345678901234567890123456789012345678901234567890123456789012345678901234567890
ATOM   4150  H   ALA A 431       8.674  16.036  12.858  1.00  0.00           H
TER    4151      ALA A 431
ATOM   1403  O   PRO P  22      12.701  33.564  15.827  1.09 18.03           O
ATOM   1404  CB  PRO P  22      13.512  32.617  18.642  1.09  9.32           C
ATOM   1405  CG  PRO P  22      12.828  33.382  19.740  1.09 12.23           C
ATOM   1406  CD  PRO P  22      12.324  34.603  18.985  1.09 11.47           C
HETATM 1407  CA  BLE P   1      14.625  32.240  14.151  1.09 16.76           C
HETATM 1408  CB  BLE P   1      15.610  33.091  13.297  1.09 16.56           C
HETATM 1409  CG  BLE P   1      15.558  34.629  13.373  1.09 14.27           C
HETATM 1410  CD1 BLE P   1      16.601  35.208  12.440  1.09 14.75           C
HETATM 1411  CD2 BLE P   1      14.209  35.160  12.930  1.09 15.60           C
HETATM 1412  N   BLE P   1      14.777  32.703  15.531  1.09 14.79           N
HETATM 1413  B   BLE P   1      14.921  30.655  14.194  1.09 15.56           B
HETATM 1414  O1  BLE P   1      14.852  30.178  12.832  1.09 16.10           O
HETATM 1415  O2  BLE P   1      13.775  30.147  14.862  1.09 20.95           O
TER    1416      BLE P   1
```

HETATM

Overview

The HETATM records present the atomic coordinate records for atoms within "non-standard" groups. These records are used for water molecules and atoms presented in HET groups (see http://deposit.pdb.org/public-component-erf.cif).

Record Format

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"HETATM"	
7 - 11	Integer	serial	Atom serial number.
13 - 16	Atom	name	Atom name.
17	Character	altLoc	Alternate location indicator.
18 - 20	Residue name	resName	Residue name.
22	Character	chainID	Chain identifier.
23 - 26	Integer	resSeq	Residue sequence number.
27	AChar	iCode	Code for insertion of residues.
31 - 38	Real(8.3)	x	Orthogonal coordinates for X.
39 - 46	Real(8.3)	y	Orthogonal coordinates for Y.
47 - 54	Real(8.3)	z	Orthogonal coordinates for Z.
55 - 60	Real(6.2)	occupancy	Occupancy.
61 - 66	Real(6.2)	tempFactor	Temperature factor.
77 - 78	LString(2)	element	Element symbol; right-justified.
79 - 80	LString(2)	charge	Charge on the atom.

Details

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

* No ordering is specified for polysaccharides.

* See the HET section of this document regarding naming of heterogens. See the HET dictionary for residue names, formulas, and topology of the HET groups that have appeared so far in the PDB (see http://deposit.pdb.org/public-component-erf.cif).

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

* If there are neither isotropic B values from the depositor, nor anisotropic temperature factors in

ANISOU, then the default value of 0.0 is used for the temperature factor.

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

Verification/Validation/Value Authority Control

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

Relationships to Other Record Types

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

Example

```
         1         2         3         4         5         6         7         8
12345678901234567890123456789012345678901234567890123456789012345678901234567890
HETATM 1357 MG    MG   168       4.669  34.118  19.123  1.00  3.16          MG2+
HETATM 3835 FE   HEM     1      17.140   3.115  15.066  1.00 14.14          FE3+
```

ENDMDL

Overview

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

Record Format

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

Details

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

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

* Every MODEL record has an associated ENDMDL record.

Verification/Validation/Value Authority Control

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

Relationships to Other Record Types

There must be a corresponding MODEL record.

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

Example

```
         1         2         3         4         5         6         7         8
12345678901234567890123456789012345678901234567890123456789012345678901234567890
...
...
ATOM  14550 1HG  GLU   122      -14.364  14.787 -14.258  1.00  0.00           H
ATOM  14551 2HG  GLU   122      -13.794  13.738 -12.961  1.00  0.00           H
TER   14552      GLU   122
ENDMDL
MODEL        9
ATOM  14553  N   SER     1      -28.280   1.567  12.004  1.00  0.00           N
ATOM  14554  CA  SER     1      -27.749   0.392  11.256  1.00  0.00           C
...
...
ATOM  16369 1HG  GLU   122       -3.757  18.546  -8.439  1.00  0.00           H
```

```
ATOM  16370 2HG  GLU   122      -3.066  17.166  -7.584  1.00  0.00           H
TER   16371      GLU   122
ENDMDL
MODEL       10
ATOM  16372  N   SER     1     -22.285   7.041  10.003  1.00  0.00           N
ATOM  16373  CA  SER     1     -23.026   6.872   8.720  1.00  0.00           C
...
...
ATOM  18188 1HG  GLU   122      -1.467  18.282 -17.144  1.00  0.00           H
ATOM  18189 2HG  GLU   122      -2.711  18.067 -15.913  1.00  0.00           H
TER   18190      GLU   122
ENDMDL
```

10. Connectivity Section

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

CONECT

Overview

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

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

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

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

* No differentiation is made between donor and acceptor for hydrogen bonds.

* No differentiation is made between atoms with excess negative or positive charge.

* Atoms specified in the connectivity are presented by their serial numbers as found in the entry.

* All atoms connected to the atom with serial number in columns 7 - 11 are listed in the remaining fields of the record.

* If more than four fields are required for non-hydrogen and nonsalt-bridge bonds, a second CONECT record with the same atom serial number in columns 7 - 11 will be used.

* These CONECT records occur in increasing order of the atom serial numbers they carry in columns 7 - 11. The target-atom serial numbers carried on these records also occur in increasing order.

* The connectivity list given here is redundant in that each bond indicated is given twice, once with each of the two atoms involved specified in columns 7 - 11.

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

* For hydrogen bonds, when the hydrogen atom is present in the coordinates, PDB generates a CONECT record between the hydrogen atom and its acceptor atom.

* For NMR entries, CONECT records for all models are generated describing heterogen connectivity and others for LINK records.

Verification/Validation/Value Authority Control

Connectivity is checked for unusual bond lengths.

Relationships to Other Record Types

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

Example

```
         1         2         3         4         5         6         7
12345678901234567890123456789012345678901234567890123456789012345678901234567890
CONECT 1179  746 1184 1195 1203
CONECT 1179 1211 1222
CONECT 1021  544 1017 1020 1022 1211 1222      1311
```

Known Problems

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

11. Bookkeeping Section

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

MASTER

Overview

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

Record Format

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

Details

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

Verification/Validation/Value Authority Control

The MASTER line is generated by the PDB.

Relationships to Other Record Types

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

Example

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

END

Overview

The END record marks the end of the PDB file.

Record Format

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

Details

* END is the final record of a coordinate entry.

Verification/Validation/Value Authority Control

END must appear in every coordinate entry.

Relationships to Other Record Types

This is the final record in the entry.

Example

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