Protein Data Bank Contents Guide:

Atomic Coordinate Entry Format Description

Version 3.1, July 19, 2007
Version 3.1, February 11, 2008

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

Version 3.1 of the PDB file format introduces a small number of changes and extensions supporting the annotation practices adopted by the wwPDB beginning in August 2007. These annotation practices are described in detail in the documentation section of the wwPDB website. Format version 3.1 also incorporates all of the revisions of format version 3.0 which was used by the wwPDB to integrate uniformity and remediation data into a single set of archival data files. This document describes the small number of differences between version 3.1 and the preceding version 2.3 formats. All changes of the changes described in the version 3.0 Format Guide are incorporated into this version 3.1 format description. The complete details of the PDB format version 2.3 can be found at http://www.wwpdb.org/docs.html.

Version 3.1, July 19, 2007   Initial version
Version 3.1, February 11, 2008   minor changes in remark 350 and revdat, add cutoff limits for close contacts

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. The changes in records in this section are described below.

REVDAT

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        Modification detail.
47 - 52     LString(6)    record        Modification detail.
54 - 59     LString(6)    record        Modification detail.
61 - 66     LString(6)    record        Modification detail.
```

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   Other modification.

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

Modification details are typically PDB record names such as SOURCE, TITLE, or COMPND. A special modification detail VERSN indicates that the file has undergone a change in version.   The current version will be specified in REMARK 4.

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

In the case of a version revision, the current will be specified in REMARK 4.

Template

```
         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
REVDAT   2   15-OCT-99 1ABC    1       REMARK
REVDAT   1   09-JAN-89 1ABC    0


         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
REVDAT   2   11-MAR-08 2ABC    1       JRNL VERSN
REVDAT   1   09-DEC-03 2ABC    0
```

REMARK 0

REMARK 0 identifies entries in which a re-refinement has been performed using the data from existing entry.

Template

```
REMARK   0
REMARK   0 THIS ENTRY yyyy REFLECTS AN ALTERNATIVE MODELING OF THE
REMARK   0 STRUCTURAL DATA IN RxxxxSF (or xxxx.MR).
REMARK   0 IN PDB ENTRY YYYY INFORMATION IN REMARK 200 (or 210)
REMARK   0 SERIES IS BASED ON THE EXPERIMENT DESCRIBED IN PDB ENTRY xxxx
REMARK   0 ORIGINAL DATA DETERMINED BY AUTHOR:
REMARK   0 AUTHOR INITIALS, AUTHOR LAST NAME …
```

Note: In entries where REMARK 0 is included as described above, remarks REMARK 1, REMARK 200/210 and REMARK 900 will also reflect the reuse of existing experimental data.

REMARK 4

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

Version 3.1 files include a version remark like the following:

```
          1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
REMARK   4
REMARK   4 1ABC COMPLIES WITH FORMAT V. 3.1, 1-AUG-2007
```

Version 3.0 files include a version remark like the following:

```
REMARK   4
REMARK   4 1ABC COMPLIES WITH FORMAT V. 3.0, 1-DEC-2006
REMARK   4
REMARK   4 THIS IS THE REMEDIATED VERSION OF THIS PDB ENTRY.
REMARK   4 REMEDIATED DATA FILE REVISION 3.XXX (YYYY-MM-DD)
```

REMARK 40

Pre-submission validation and similar software used before deposition, may be listed in a REMARK 40. This remark will list the software name, authors and function of the program. Results of these software methods will not be listed. This free text remark is optional.

Template

```
REMARK 40 FREE TEXT DESCRIPTION
```

REMARKs 102-199 Nucleic Acids

The text remarks for nucleic acids will reflect the standardization of nomenclature for the polymer nucleotides described in later sections. In particular, the polymeric deoxyribonucleotides are represented by 2-letter codes DC, DG, DA, and DT to distinguish these from their ribonucletide counterparts. The asterisk character in the in saccharide atom names is replaced by the single prime character. The text of REMARK 105 is correspondingly changed as follows.

REMARK 105

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

Template

```
          1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
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 200, X-ray Diffraction Experimental Details

The following example illustrates the how REMARK 200 will be used in cases in which multiple data collections are described. In this example, data items corresponding to different data collection sessions are separated by semi-colons. Multiple data values within a single session (e.g wavelength) are separated by commas.

```
REMARK 200
REMARK 200 EXPERIMENTAL DETAILS
REMARK 200  EXPERIMENT TYPE                : X-RAY DIFFRACTION
REMARK 200  DATE OF DATA COLLECTION         : 17-MAR-2002; 17-MAR-2002
REMARK 200  TEMPERATURE          (KELVIN) : 100; 100
REMARK 200  PH                               : 8.00
REMARK 200  NUMBER OF CRYSTALS USED      : 2
REMARK 200
REMARK 200  SYNCHROTRON             (Y/N) : Y; Y
REMARK 200  RADIATION SOURCE                : APS ; APS
REMARK 200  BEAMLINE                         : 17ID; 17ID
REMARK 200  X-RAY GENERATOR MODEL        : NULL
REMARK 200  MONOCHROMATIC OR LAUE    (M/L) : M; M
REMARK 200  WAVELENGTH OR RANGE        (A) : 1.5545; 1.0720, 1.0723,
REMARK 200                                  1.0543
REMARK 200  MONOCHROMATOR                   : SI (111); SI (111)
REMARK 200  OPTICS                          : NULL
REMARK 200
REMARK 200  DETECTOR TYPE                   : CCD; CCD
REMARK 200  DETECTOR MANUFACTURER           : ADSC QUANTUM 210; ADSC
REMARK 200                                  QUANTUM 210
REMARK 200  INTENSITY-INTEGRATION SOFTWARE : DENZO
REMARK 200  DATA SCALING SOFTWARE          : HKL
REMARK 200
REMARK 200  NUMBER OF UNIQUE REFLECTIONS   : 29132
REMARK 200  RESOLUTION RANGE HIGH      (A) : 1.900
REMARK 200  RESOLUTION RANGE LOW       (A) : 30.000
REMARK 200  REJECTION CRITERIA  (SIGMA(I)) : 0.000
REMARK 200
REMARK 200 OVERALL.
REMARK 200  COMPLETENESS FOR RANGE     (%) : 98.3
REMARK 200  DATA REDUNDANCY                 : 19.800
REMARK 200  R MERGE                    (I) : NULL
REMARK 200  R SYM                      (I) : 0.07500
REMARK 200  <I/SIGMA(I)> FOR THE DATA SET  : 17.0000
REMARK 200
REMARK 200 IN THE HIGHEST RESOLUTION SHELL.
REMARK 200  HIGHEST RESOLUTION SHELL, RANGE HIGH (A) : 1.90
REMARK 200  HIGHEST RESOLUTION SHELL, RANGE LOW  (A) : 1.97
REMARK 200  COMPLETENESS FOR SHELL     (%) : 83.4
REMARK 200  DATA REDUNDANCY IN SHELL       : 3.00
REMARK 200  R MERGE FOR SHELL          (I) : NULL
REMARK 200  R SYM FOR SHELL            (I) : 0.65000
REMARK 200  <I/SIGMA(I)> FOR SHELL         : 1.500
REMARK 200
REMARK 200 DIFFRACTION PROTOCOL: SINGLE WAVELENGTH; MAD
REMARK 200 METHOD USED TO DETERMINE THE STRUCTURE: MAD
REMARK 200 SOFTWARE USED: SOLVE 2.02
REMARK 200 STARTING MODEL: NULL
REMARK 200
REMARK 200 REMARK: NULL
```

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: 1
REMARK 300 SEE REMARK 350 FOR THE AUTHOR PROVIDED AND/OR PROGRAM
REMARK 300 GENERATED ASSEMBLY INFORMATION FOR THE STRUCTURE IN
REMARK 300 THIS ENTRY.  THE REMARK MAY ALSO PROVIDE INFORMATION ON
REMARK 300 BURIED SURFACE AREA.
REMARK 300 free text …
```

Example - Icosahedral virus

```
REMARK 300
REMARK 300 BIOMOLECULE: 1
REMARK 300 SEE REMARK 350 FOR THE AUTHOR PROVIDED AND/OR PROGRAM
REMARK 300 GENERATED ASSEMBLY INFORMATION FOR THE STRUCTURE IN
REMARK 300 THIS ENTRY.  THE REMARK MAY ALSO PROVIDE INFORMATION ON
REMARK 300 BURIED SURFACE AREA.
REMARK 300 DETAILS: THE ASSEMBLY REPRESENTED IN THIS ENTRY HAS REGULAR
REMARK 300 ICOSAHEDRAL POINT SYMMETRY (SCHOENFLIES SYMBOL = I).
```

Example - Helical viruses

```
REMARK 300 BIOMOLECULE: 1
REMARK 300 SEE REMARK 350 FOR THE AUTHOR PROVIDED AND/OR PROGRAM
REMARK 300 GENERATED ASSEMBLY INFORMATION FOR THE STRUCTURE IN
REMARK 300 THIS ENTRY.  THE REMARK MAY ALSO PROVIDE INFORMATION ON
REMARK 300 BURIED SURFACE AREA.
REMARK 300 DETAILS: THE ASSEMBLY REPRESENTED IN THIS ENTRY HAS REGULAR
REMARK 300 HELICAL SYMMETRY WITH THE FOLLOWING PARAMETERS:
REMARK 300 ROTATION PER SUBUNIT (TWIST) = -33.23 DEGREES
REMARK 300 RISE PER SUBUNIT (HEIGHT) = 16.00 ANGSTROMS
REMARK 300 IN ADDITION, THERE IS 5-FOLD CIRCULAR
REMARK 300 SYMMETRY AROUND THE HELIX AXIS
```

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.   Both author and computational descriptions of assemblies are provided if these are available.

Example – Author and computed assembly predictions agree

```
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 AUTHOR DETERMINED BIOLOGICAL UNIT: DODECAMERIC
REMARK 350 SOFTWARE DETERMINED QUATERNARY STRUCTURE: DODECAMERIC
REMARK 350 SOFTWARE USED: PISA
REMARK 350 TOTAL BURIED SURFACE AREA: 2990 ANGSTROM**2
REMARK 350 TOTAL SURFACE AREA FOR THE COMPLEX: 9330 ANGSTROM**2
REMARK 350 GAIN IN SOLVENT FREE ENERGY: -40 KCAL/MOL
REMARK 350 APPLY THE FOLLOWING TO CHAINS: A, B, C, D, E, F, G, H, I,
REMARK 350                                   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
```

Note: The value for the average buried surface area will be round to the nearest 10.

Example – Author and computed assembly predictions differ

```
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 AUTHOR DETERMINED BIOLOGICAL UNIT: HEXAMERIC
REMARK 350 APPLY THE FOLLOWING TO CHAINS: A, B, C, 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
REMARK 350 BIOMOLECULE: 2
REMARK 350 AUTHOR DETERMINED BIOLOGICAL UNIT: HEXAMERIC
REMARK 350 APPLY THE FOLLOWING TO CHAINS: G, H, 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
REMARK 350 BIOMOLECULE: 3
REMARK 350 SOFTWARE DETERMINED QUATERNARY STRUCTURE: DODECAMERIC
REMARK 350 SOFTWARE USED: PISA
REMARK 350 TOTAL BURIED SURFACE AREA: 2990 ANGSTROM**2
REMARK 350 TOTAL SURFACE AREA FOR THE COMPLEX: 9330 ANGSTROM**2
REMARK 350 GAIN IN SOLVENT FREE ENERGY: -40 KCAL/MOL
REMARK 350 APPLY THE FOLLOWING TO CHAINS: A, B, C, D, E, F, G, H, I,
REMARK 350                                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
```

Example – No experimental information about assembly provided

```
REMARK 350 COORDINATES FOR A COMPLETE MULTIMER REPRESENTING
REMARK 350 OLIGOMERIZATION STATE(S) 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 SOFTWARE DETERMINED QUATERNARY STRUCTURE: MONOMERIC
REMARK 350 SOFTWARE USED: PISA
REMARK 350 APPLY THE FOLLOWING TO CHAINS: A
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
```

REMARK 475, Residues modeled with zero occupancy

REMARK 475 enumerates residues modeled with zero occupancy.

Example –

```
REMARK 475
REMARK 475 ZERO OCCUPANCY RESIDUES
REMARK 475 THE FOLLOWING RESIDUES WERE MODELED WITH ZERO OCCUPANCY.
REMARK 475 THE LOCATION AND PROPERTIES OF THESE RESIDUES MAY NOT
REMARK 475 BE RELIABLE. (M=MODEL NUMBER; RES=RESIDUE NAME;
REMARK 475 C=CHAIN IDENTIFIER; SSEQ=SEQUENCE NUMBER; I=INSERTION CODE)
REMARK 475   M RES C SSEQI
REMARK 475     DG D    4
```

REMARK 480, Polymer atoms modeled with zero occupancy

REMARK 480 enumerates atoms in residues modeled with zero occupancy.

Example –

```
REMARK 480
REMARK 480 ZERO OCCUPANCY ATOM
REMARK 480 THE FOLLOWING RESIDUES HAVE ATOMS MODELED WITH ZERO
REMARK 480 OCCUPANCY. THE LOCATION AND PROPERTIES OF THESE ATOMS
REMARK 480 MAY NOT BE RELIABLE. (M=MODEL NUMBER; RES=RESIDUE NAME;
REMARK 480 C=CHAIN IDENTIFIER; SSEQ=SEQUENCE NUMBER; I=INSERTION CODE):
REMARK 480   M RES C SSEQI ATOMS
REMARK 480     DC D   3   C4'  O4'  C1'  C3'  O3'
```

REMARK 500, Geometry and Stereochemistry

Remark 500 provides further details on the stereochemistry of the structure. This remark is generated by PDB and may incorporate comments provided by the author. This remark is currently divided into the subtopics:

    CLOSE CONTACTS IN SAME ASYMMETRIC UNIT,
    CLOSE CONTACTS,
    COVALENT BOND LENGTHS,
    COVALENT BOND ANGLES,
    TORSION ANGLES,
    NON-CIS & NON-TRANS,
    PLANAR GROUPS,
    MAIN CHAIN PLANARITY,
    CHIRAL CENTERS.

Additional subtopics may be added as needed. For close contacts, the cutoff limit is 2.2 angstrom for non hydrogen atoms and is 1.6 angstrom for H and D atoms. These distances are listed in the REMARK 500 for close contacts symmetry.

The calculation of bond and angle deviations for protein entries will be based on the updated Engh & Huber amino acid target values[1]. For nucleic acids, the Parkinson et al., statistics will be used for these calculations[2]. All bonds and angles that deviate more than 6 times from their standard target values will be flagged as a deviation. The PHI/PSI values are based on Kleywegt's calculations[3].

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.
```

---

[1] Structure quality and target parameters. R. A. Engh and R. Huber. International Tables for Crystallography (2006). Vol. F, ch. 18.3, pp. 382-392

[2] "New Parameters for the Refinement of Nucleic Acid Containing Structures." Gary Parkinson, Jaroslav Vojtechovsky, Lester Clowney, Axel Brunger*, and Helen M. Berman. (1996) Acta Cryst. D 52, 57-64

[3] "PHI/PSI- Chology: Ramachandran revisited. " GJ Kleywegt and TA Jones (1996) Structure 4, 1395-1400.

Example – Close Contacts in the 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   N    PHE 1    8  -  OD2  ASP 1    31               2.17
REMARK 500   OD2  ASP 1   31  -  N    PHE 1     8               2.17
REMARK 500
REMARK 500 THIS ENTRY HAS     104 CLOSE CONTACTS
REMARK 500
REMARK 500 REMARK: NULL
```

Example – Close Contacts

```
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.  AN ATOM LOCATED WITHIN 0.15
REMARK 500 ANGSTROMS OF A SYMMETRY RELATED ATOM IS ASSUMED TO BE ON A
REMARK 500 SPECIAL POSITION AND IS, THEREFORE, LISTED IN REMARK 375
REMARK 500 INSTEAD OFREMARK 500.  ATOMS WITH NON-BLANK ALTERNATE
REMARK 500 LOCATION INDICATORS ARE NOT INCLUDED IN THE CALCULATIONS.
REMARK 500
REMARK 500 DISTANCE CUTOFF:
REMARK 500 2.2 ANGSTROMS FOR CONTACTS NOT INVOLVING HYDROGEN ATOMS
REMARK 500 1.6 ANGSTROMS FOR CONTACTS INVOLVING HYDROGEN ATOMS
REMARK 500
REMARK 500  ATM1  RES C  SSEQI   ATM2  RES C  SSEQI  SSYMOP   DISTANCE
REMARK 500
REMARK 500   OH   TYR 1   90     O    HOH     343     1566      2.09
REMARK 500   OE1  GLU 1   134    CB   LYS 2   135     1556      2.18
REMARK 500
REMARK 500 THIS ENTRY HAS      64 SYMMETRY CONTACTS
REMARK 500
REMARK 500 REMARK: NULL
```

Example – Covalent bond lengths

```
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 (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,A1,1X,2(A4,A1,3X),12X,F5.3)
REMARK 500
REMARK 500 EXPECTED VALUES PROTEIN: ENGH AND HUBER, 1999
REMARK 500 EXPECTED VALUES NUCELIC ACID: CLOWNEY ET AL 1996
REMARK 500
REMARK 500  M RES CSSEQI ATM1    RES CSSEQI ATM2   DEVIATION
REMARK 500    ASP B 117   CB     ASP B 117   CG     -0.129
REMARK 500    CYS J  29   CB     CYS J  29   SG     -0.111
REMARK 500
REMARK 500 REMARK: NULL
```

Example – Covalent bond angles

```
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(1X,A4,2X),12X,F5.1)
REMARK 500
REMARK 500 EXPECTED VALUES: ENGH AND HUBER, 1999
REMARK 500
REMARK 500  M RES CSSEQI ATM1    ATM2    ATM3
REMARK 500    VAL A 124   CB  -  CA  -  C   ANGL. DEV. = -12.0 DEGREES
REMARK 500    ARG B  70   NE  -  CZ  -  NH1 ANGL. DEV. =  -3.0 DEGREES
REMARK 500
REMARK 500 REMARK: NULL
```

Example – Torsion angles

```
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 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    ASN A 100      -110.87   -163.72
REMARK 500    ILE A 166       -28.81    -31.64
REMARK 500
REMARK 500 REMARK: NULL
```

Example – Cis/Trans geometry

```
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 ARG A  413      ASP A  414           0        147.84
REMARK 500 ALA B  288      ASN B  289           0        -39.12
REMARK 500
REMARK 500 REMARK: NULL
```

Example – Planar groups

```
REMARK 500
REMARK 500 GEOMETRY AND STEREOCHEMISTRY
REMARK 500 SUBTOPIC: PLANAR GROUPS
REMARK 500
REMARK 500 PLANAR GROUPS IN THE FOLLOWING RESIDUES HAVE A TOTAL
REMARK 500 RMS DISTANCE OF ALL ATOMS FROM THE BEST-FIT PLANE
REMARK 500 BY MORE THAN AN EXPECTED VALUE OF 6*RMSD, WITH AN
REMARK 500 RMSD 0.02 ANGSTROMS, OR AT LEAST ONE ATOM HAS
REMARK 500 AN RMSD GREATER THAN THIS VALUE
REMARK 500 (M=MODEL NUMBER; RES=RESIDUE NAME; C=CHAIN IDENTIFIER;
REMARK 500 SSEQ=SEQUENCE NUMBER; I=INSERTION CODE).
REMARK 500
REMARK 500  M RES CSSEQI        RMS     TYPE
REMARK 500    TYR A  36        0.08    SIDE CHAIN
REMARK 500    TYR A 104        0.08    SIDE CHAIN
REMARK 500
REMARK 500 REMARK: NULL
```

Example – Main chain planarity

```
REMARK 500
REMARK 500 GEOMETRY AND STEREOCHEMISTRY
REMARK 500 SUBTOPIC: MAIN CHAIN PLANARITY
REMARK 500
REMARK 500 THE FOLLOWING RESIDUES HAVE A PSEUDO PLANARITY
REMARK 500 TORSION, C(I) - CA(I) - N(I+1) - O(I), GREATER
REMARK 500 10.0 DEGREES. (M=MODEL NUMBER; RES=RESIDUE NAME;
REMARK 500 C=CHAIN IDENTIFIER; SSEQ=SEQUENCE NUMBER;
REMARK 500 I=INSERTION CODE).
REMARK 500
REMARK 500  M RES CSSEQI        ANGLE
REMARK 500
REMARK 500 REMARK: NULL
```

Example – Chiral centers

```
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                    OMEGA
REMARK 500    ASP B 405        ALPHA-CARBON 150.48
REMARK 500
REMARK 500 REMARK: NULL
```

REMARK 525 – Distant Solvent Atoms

REMARK 525 lists solvent atoms more than 5 Angstroms from any polymer chain.

Example –

```
REMARK 525
REMARK 525 SOLVENT
REMARK 525
REMARK 525 THE SOLVENT MOLECULES HAVE CHAIN IDENTIFIERS THAT
REMARK 525 INDICATE THE POLYMER CHAIN WITH WHICH THEY ARE MOST
REMARK 525 CLOSELY ASSOCIATED. THE REMARK LISTS ALL THE SOLVENT
REMARK 525 MOLECULES WHICH ARE MORE THAN 5A AWAY FROM THE
REMARK 525 NEAREST POLYMER CHAIN (M=MODEL NUMBER;
REMARK 525 RES=RESIDUE NAME; C=CHAIN IDENTIFIER; SSEQ=SEQUENCE
REMARK 525 NUMBER; I=INSERTION CODE):
REMARK 525
REMARK 525  M RES CSSEQI
REMARK 525    HOH B  89        DISTANCE =  6.29 ANGSTROMS
REMARK 525    HOH B  94        DISTANCE =  5.58 ANGSTROMS
```

REMARK 610 – Non-polymer residues with missing atoms

REMARK 610 enumerates non-polymer residues with missing atoms.

Example -

```
REMARK 610
REMARK 610 MISSING HETEROATOM
REMARK 610 THE FOLLOWING RESIDUES HAVE MISSING ATOMS (M=MODEL NUMBER;
REMARK 610 RES=RESIDUE NAME; C=CHAIN IDENTIFIER; SSEQ=SEQUENCE NUMBER;
REMARK 610 I=INSERTION CODE):
REMARK 610   M RES C SSEQI
REMARK 610     PPI    438
```

REMARK 615 – Non-polymer residues containing atoms with zero occupancy

REMARK 615 enumerates non-polymer residues containing atoms modeled with zero occupancy.

Example –

```
REMARK 615
REMARK 615 ZERO OCCUPANCY ATOM
REMARK 615 THE FOLLOWING RESIDUES HAVE ATOMS MODELED WITH ZERO
REMARK 615 OCCUPANCY. THE LOCATION AND PROPERTIES OF THESE ATOMS
REMARK 615 MAY NOT BE RELIABLE. (M=MODEL NUMBER; RES=RESIDUE NAME;
REMARK 615 C=CHAIN IDENTIFIER; SSEQ=SEQUENCE NUMBER; I=INSERTION CODE):
REMARK 615   M RES C SSEQI
REMARK 615     PPI    438
```

REMARK 620 – Metal Coordination

Details of metal coordination are provided in REMARK 620. Coordination angles for any metal coordination and surrounding residues (if present) will be provided this remark.

Template:

```
REMARK 620
REMARK 620 METAL COORDINATION
REMARK 620   formatted text
```

Example -

```
REMARK 620
REMARK 620 METAL COORDINATION
REMARK 620  (M=MODEL NUMBER; RES=RESIDUE NAME; C=CHAIN IDENTIFIER;
REMARK 620   SSEQ=SEQUENCE NUMBER; I=INSERTION CODE):
REMARK 620
REMARK 620 COORDINATION ANGLES FOR:  M RES CSSEQI METAL
REMARK 620                              F3S A 107  FE1
REMARK 620 N RES CSSEQI ATOM
REMARK 620 1 CYS A  39    SG
REMARK 620 2 F3S A 107   FE3  142.2
REMARK 620 3 F3S A 107   FE4  154.3  59.7
REMARK 620 4 F3S A 107    S1  120.2  53.8   55.7
REMARK 620 5 F3S A 107    S2  113.0 103.5   54.3 106.6
REMARK 620 6 F3S A 107    S3  103.8  53.0 101.7 103.2 109.2
REMARK 620 N                    1     2      3     4      5
REMARK 620
REMARK 620 COORDINATION ANGLES FOR:  M RES CSSEQI METAL
REMARK 620                              F3S A 107  FE3
REMARK 620 N RES CSSEQI ATOM
REMARK 620 1 F3S A 107   FE1
REMARK 620 2 F3S A 107   FE4   59.0
REMARK 620 3 F3S A 107    S1   52.7  55.1
REMARK 620 4 F3S A 107    S3   52.9 101.0 102.1
REMARK 620 5 CYS A  45    SG  146.5 146.2 115.6 112.8
REMARK 620 6 F3S A 107    S4  103.5  54.5 106.3 109.6 110.0
REMARK 620 N                    1     2      3     4      5
```

```
REMARK 620
REMARK 620 COORDINATION ANGLES FOR:  M RES CSSEQI METAL
REMARK 620                              F3S A 107  FE4
REMARK 620 N RES CSSEQI ATOM
REMARK 620 1 F3S A 107   FE1
REMARK 620 2 F3S A 107   FE3    61.3
REMARK 620 3 F3S A 107   S1     53.4  53.9
REMARK 620 4 F3S A 107   S2     54.4 105.0 104.5
REMARK 620 5 CYS A  20   SG    142.7 140.2 109.0 114.5
REMARK 620 6 F3S A 107   S4    105.1  54.1 104.8 111.7 111.6
REMARK 620 N                      1     2     3     4     5
REMARK 620
REMARK 620 COORDINATION ANGLES FOR:  M RES CSSEQI METAL
REMARK 620                              F3S A 108  FE1
REMARK 620 N RES CSSEQI ATOM
REMARK 620 1 F3S A 108   S3
REMARK 620 2 CYS A  16   SG    120.1
REMARK 620 3 F3S A 108   FE3    51.4 145.9
REMARK 620 4 F3S A 108   FE4    54.3 148.5  59.9
REMARK 620 5 F3S A 108   S1     98.3 110.0  50.6 101.5
REMARK 620 6 F3S A 108   S2    104.2 109.4 104.5  53.3 114.7
REMARK 620 N                      1     2     3     4     5
REMARK 620
```

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.

The changes in the records in this section result from the standardization of nomenclature the standard nucleotides and nucleotide modifications.

DBREF

The DBREF record provides cross-reference links between PDB sequences and the corresponding database entry or entries. There are no character/column format changes in the records in this section. A detailed description of the assignment of sequence database references is presented in Section A of the wwPDB Processing Procedures and Policies Document.

SEQADV

The SEQADV record identifies conflicts between sequence information in the SEQRES records of the PDB entry and the sequence database entry given on DBREF. There are no character/column format changes in the records in this section. A detailed description of the assignment of sequence difference records is presented in Section A of the wwPDB Processing Procedures and Policies Document.

SEQRES

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

The ribo- and deoxyribonucleotides in the SEQRES records are now distinguished. The deoxy- forms of these residues are now identified with the residue names DA, DC, DG, DT, and DU. Modified nucleotides in the sequence are now identified by separate 3-letter residue codes. The use of the *plus* character prefix to label modified nucleotides (e.g. +A, +C, +T) is no longer used.

MODRES

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

Modified nucleotides in the sequence are now identified by separate 3-letter residue codes. The use of the *plus* character prefix to label modified nucleotides (e.g. +A, +C, +T) is no longer used.

4. Heterogen Section

The heterogen section of a PDB file contains the complete description of non-standard residues in the entry.   Changes in the detailed chemical descriptions of non-polymer chemical components are described in the PDB Chemical Components dictionary,

http://remediation.wwpdb.org/downloads/Components-rel-alt.cif.

There are no character/column format changes in the records in this section; however, the definition of a PDB HET group is revised owing to the change in nomenclature for the standard deoxyribonucleotides as described in the following section.

HET

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 ribonucleic acids (C, G, A, T, U, and I), and
- not one of the deoxyribonucleic acids (DC, DG, DA, DT, DU and DI)
- not an unknown amino acid or nucleic acid where UNK is used to indicate the unknown residue name.

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

5. Secondary Structure Section

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

There are no changes in the formats of the records in this section.

6. Connectivity Annotation Section

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

SSBOND

The disulfide bond distance is included after the symmetry operations at the end of the SSBOND record.

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 op for residue 1.

67 - 72           SymOP            sym2             Symmetry op for residue 2.

74 – 78           Real(5.2)        Length           Disulfide bond distance
```

Example –

```
SSBOND   1 CYS A    6    CYS A  127                          1555   1555   2.03
SSBOND   2 CYS A   30    CYS A  115                          1555   1555   2.07
SSBOND   3 CYS A   64    CYS A   80                          1555   1555   2.06
SSBOND   4 CYS A   76    CYS A   94                          1555   1555   2.04
```

LINK

Link records include the distance associated with the each linkage following the symmetry operations at the end of each record.

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 atom 1.

67 - 72             SymOP               sym2            Symmetry operator atom 2.

74 – 78             Real(5.2)           Length          Link distance
```

Example -

```
LINK         O   GLY A  49                NA    NA A6001   1555   1555   2.98
LINK         OG1 THR A  51                NA    NA A6001   1555   1555   2.72
LINK         OD2 ASP A  66                NA    NA A6001   1555   1555   2.72
LINK         NE  ARG A  68                NA    NA A6001   1555   1555   2.93
```

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.

There are no changes in the formats of the records in this section.

8. Crystallographic Coordinate Transformation Section

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

There are no changes in the formats of the records in this section.

9. Coordinate Section

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

ATOM/HETATM

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

The character/column format of the ATOM/HETATM records is not changed. Changes in ATOM/HETATM records result from the standardization atom and residue nomenclature.   This nomenclature is described in electronic form in the PDB Chemical Components Dictionary, which may be downloaded at http://remediation.wwpdb.org/downloads/Components-rel-alt.cif.

A detailed description of the assignment of alternate conformations and the assignment of chain identifiers for non-polymers and solvent is presented in Section A of the wwPDB Processing Procedures and Policies Document.

10. Connectivity Section

This section provides information on chemical connectivity.

There are no changes in the formats of the records in this section.

11. Bookkeeping Section

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

There are no changes in the formats of the records in this section.

12. Nomenclature

Atom and residue nomenclature has been standardized in a variety of ways in PDB version 3.0 data files. All changes in nomenclature are documented in the electronic chemical components dictionary,

http://remediation.wwpdb.org/downloads/Components-rel-alt.cif.

 The changes in nomenclature include:

- **IUPAC nomenclature for standard amino acid and nucleotides.** Atoms names follow the recommendations of described in *Pure & Appl. Chem.,* 70, 117-142, 1998. with the exception of the well-established convention for C-terminal atoms OXT and HXT.   In this and other cases where an atom name has been changed, the previous name is retained in an alternate name in the PDB Chemical Components dictionary.
- **Discrimination of DNA and RNA linking nucleotides and modifications.** Deoxy- and ribose nucleotides now have separate chemical definitions with the DNA forms relabeled as DA, DC, DG, DT, DI and DU. Modified nucleotides formerly identified as using the *"plus-nucleotide"* syntax (e.g. +C, +G) have been relabeled with the particular 3-letter code corresponding to the full chemical description of the modified nucleotide.
- **More conventional atom labeling for non-polymer atoms.**
  In the new chemical definitions the following changes have been made to move the atom naming to a more conventional state.
    - Atom names begin with their element symbol
    - Heavy atom names follow the traditional PDB justification rules in which the atom element symbol is right justified in the second character position of the 4-character atom name. 4-character names for atoms with 1-character element symbols have been compressed to 3 characters.
    - Hydrogen atoms names all begin with "H" and are not subject to the justification rule.
- **Removal of redundant and deprecated ligands.** In cases where the same monomer or ligand had been defined using different identifiers, the most common identifier has been retained and the others have been marked as obsolete. Definitions which were deemed incorrect or better represented in other ways (e.g. metal hydrates) have also been obsoleted.