Coordinate File Description (PDB Format)

General Information

The following describes the minimum coordinate specification in PDB format that is required by the RCSB validation and deposition software. The PDB record formats for coordinate data are reproduced here for your convenience; however, the validation check and deposition may only require the specification of a few of these records. In many cases, a file with the following format is all that is required.

- Each line is 80 columns wide and is terminated by an end-of-line indicator.
- The first six columns of every line contain a "record name". This must be an exact match to one of the stated record names described in detail below.
- The list of ATOM records in each polymer chain must be terminated by a TER record.
- ATOM records for polymer atoms must include non-blank chain ID fields.
- To use the automatic validation check, the coordinate file must include a complete CRYST1 record defining the unit cell and space group information.
- If an alternate setting is being used for the space group symmetry then the orthogonal to fractional transformation must be specified in SCALE records.
- Each file should terminate with a line containing only the word END.
- If atom names or HETATM residue names are unrecognized in the dictionaries, validation results will not be optimal. Definitions for any new HETATM groups sent to deposit@deposit.rcsb.org will be included in our dictionaries.

Coordinate Record Descriptions

CRYST1 | ORIGXn | SCALEn | MTRIXn | TVECT | MODEL
ATOM | ANISOU | TER | HETATM | ENDMDL

Record: CRYST1

Contains: unit cell parameters, space group, and Z value

Notes:

- If the structure was not determined by crystallographic n
(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 wit
e.g., P 21 21 2 and using the full symbol, e.g., C 1 2 1 instea
- The screw axis is described as a two digit number.
- For a rhombohedral space group in the hexagonal settir
- The Z value is the number of polymeric chains in a unit
Z is the number of occurrences of the most populous chain.
- In the case of a polycrystalline fiber diffraction study,
CRYST1 and SCALE contain the normal unit cell data.
- The unit cell parameters are used to calculate SCALE.

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

Example:

```
         1         2         3         4         5
12345678901234567890123456789012345678901234567890123456
CRYST1  117.000   15.000   39.000  90.00  90.00  90.00
```

Record: ORIGXn

Contains: the transformation from the orthogonal coordinates containe
in the database entry to the submitted coordinates

Notes: If the original submitted coordinates are Xsub, Ysub, Zsub a

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
```

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

Example:

```
         1         2         3         4         5
12345678901234567890123456789012345678901234567890123456
ORIGX1      0.963457  0.136613  0.230424       16.6100
ORIGX2     -0.158977  0.983924  0.081383       13.7200
ORIGX3     -0.215598 -0.115048  0.969683       37.6500
```

Record: SCALEn

Contains: the transformation from the orthogonal coordinates containe crystallographic coordinates

Notes:

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

```
xfrac = S11X + S12Y + S13Z + U1
yfrac = S21X + S22Y + S23Z + U2
zfrac = S31X + S32Y + S33Z + U3
```

- For NMR and fiber diffraction submissions, SCALE is gi matrix with no translation.

COLUMNS	DATA TYPE	CONTENTS
1 - 6	Record name	"SCALEn" (n=1, 2, or 3)
11 - 20	Real(10.6)	s[n][1]
21 - 30	Real(10.6)	s[n][2]
31 - 40	Real(10.6)	s[n][3]

```
46 - 55        Real(10.5)      u[n]

Example:

         1         2         3         4         5
12345678901234567890123456789012345678901234567890123456789012345
SCALE1      0.019231  0.000000  0.000000         0.0000
SCALE2      0.000000  0.017065  0.000000         0.0000
SCALE3      0.000000  0.000000  0.016155         0.0000
```

Record: MTRIXn

Contains: the transformations expressing non-crystallographic symmet

Notes: The MTRIX transformations operate on the coordinates in th equivalent representations of the molecule in the same coor trio of MTRIX records with a constant serial number is given non-crystallographic symmetry operation defined.

```
COLUMNS        DATA TYPE       CONTENTS
--------------------------------------------------------
 1 -  6        Record name     "MTRIXn" (n=1, 2, or 3)

 8 - 10        Integer         Serial number

11 - 20        Real(10.6)      m[n][1]

21 - 30        Real(10.6)      m[n][2]

31 - 40        Real(10.6)      m[n][3]

46 - 55        Real(10.5)      v[n]

60             Integer         1 if coordinates for the
                               otherwise, blank.

Example:

         1         2         3         4         5
12345678901234567890123456789012345678901234567890123456789012345
MTRIX1   1 -1.000000  0.000000 -0.000000         0.0000
MTRIX2   1 -0.000000  1.000000  0.000000         0.0000
MTRIX3   1  0.000000 -0.000000 -1.000000         0.0000
```

Record: TVECT

Record: TVECT

Contains: the translation vector which have infinite covalent connectior

Notes: For structures not comprised of discrete molecules (e.g., infi polysaccharide chains), the entry contains a fragment which the full structure by the simple translation vectors of TVECT

COLUMNS	DATA TYPE	CONTENTS
1 - 6	Record name	"TVECT "
8 - 10	Integer	Serial number
11 - 20	Real(10.5)	t[1]
21 - 30	Real(10.5)	t[2]
31 - 40	Real(10.5)	t[3]
41 - 70	String	Text comment

Example:

```
         1         2         3         4         5
12345678901234567890123456789012345678901234567890123456789012345
TVECT    1   0.00000   0.00000  28.30000
```

Record: MODEL

Contains: the model serial number when a single coordinate entry cont

Notes:
- Models are numbered sequentially beginning with 1.
- If an entry contains more than 99,999 total atoms, then it must be divided among multiple models.
- Each MODEL must have a corresponding ENDMDL rec
- In the case of an NMR entry the EXPDTA record states that are present in the individual entry.

COLUMNS	DATA TYPE	CONTENTS
1 - 6	Record name	"MODEL "

```
11 - 14        Integer         Model serial number


Example:

         1         2         3         4         5
12345678901234567890123456789012345678901234567890123456
MODEL        1
ATOM      1  N   ALA     1      11.104   6.134  -6.504
ATOM      2  CA  ALA     1      11.639   6.071  -5.147
...
...
ATOM    293 1HG  GLU    18     -14.861  -4.847   0.361
ATOM    294 2HG  GLU    18     -13.518  -3.769   0.084
TER     295      GLU    18
ENDMDL
MODEL        2
ATOM      1  N   ALA     1      11.304   6.234  -6.104
ATOM      2  CA  ALA     1      11.239   6.371  -5.247
...
...
ATOM    293 1HG  GLU    18     -14.752  -4.948   0.461
ATOM    294 2HG  GLU    18     -13.630  -3.769   0.160
TER     295      GLU    18
ENDMDL
```

Record: ATOM

Contains: the atomic coordinates for standard residues and the occupa

Notes:

- ATOM records for proteins are listed from amino to carb
- Nucleic acid residues are listed from the 5' to the 3' term
- No ordering is specified for polysaccharides.
- The list of ATOM records in a chain is terminated by a T
- If an atom is provided in more than one position, then a

location indicator must be used. Within a residue, all atoms c conformation are assigned the same alternate position indica

- Additional atoms (modifying group) to side chains of sta

which is assigned its own residue name. The chainID, seque number, and insertion code assigned to the HET group is tha standard residue to which it is attached.

- In some entries, the occupancy and temperature factor f
- The segment identifier is a string of up to four (4) alphar

include a space, e.g., CH86, A 1, NASE.

COLUMNS	DATA TYPE	CONTENTS
1 - 6	Record name	"ATOM "
7 - 11	Integer	Atom serial number.
13 - 16	Atom	Atom name.
17	Character	Alternate location indi
18 - 20	Residue name	Residue name.
22	Character	Chain identifier.
23 - 26	Integer	Residue sequence number
27	AChar	Code for insertion of r
31 - 38	Real(8.3)	Orthogonal coordinates
39 - 46	Real(8.3)	Orthogonal coordinates
47 - 54	Real(8.3)	Orthogonal coordinates
55 - 60	Real(6.2)	Occupancy.
61 - 66	Real(6.2)	Temperature factor (Def
73 - 76	LString(4)	Segment identifier, lef
77 - 78	LString(2)	Element symbol, right-j
79 - 80	LString(2)	Charge on the atom.

Example:

```
         1         2         3         4         5
12345678901234567890123456789012345678901234567890123
ATOM    145  N   VAL A  25      32.433  16.336  57.540
ATOM    146  CA  VAL A  25      31.132  16.439  58.160
ATOM    147  C   VAL A  25      30.447  15.105  58.363
ATOM    148  O   VAL A  25      29.520  15.059  59.174
ATOM    149  CB AVAL A  25      30.385  17.437  57.230
ATOM    150  CB BVAL A  25      30.166  17.399  57.373
```

```
ATOM    151  CG1AVAL A  25      28.870  17.401  57.336
ATOM    152  CG1BVAL A  25      30.805  18.788  57.449
ATOM    153  CG2AVAL A  25      30.835  18.826  57.661
ATOM    154  CG2BVAL A  25      29.909  16.996  55.922
```

Record: ANISOU

Contains: the anisotropic temperature factors

Notes:
- Columns 7 - 27 and 73 - 80 are identical to the correspo
- The anisotropic temperature factors (columns 29 - 70) a

(Angstroms**2) and are presented as integers.
- The anisotropic temperature factors are stored in the sa

atomic coordinate records.
- ANISOU values are listed only if they have been provide

COLUMNS	DATA TYPE	CONTENTS
1 - 6	Record name	"ANISOU"
7 - 11	Integer	Atom serial number
13 - 16	Atom	Atom name
17	Character	Alternate location ind
18 - 20	Residue name	Residue name
22	Character	Chain identifier
23 - 26	Integer	Residue sequence numbe
27	AChar	Insertion code
29 - 35	Integer	u[1][1]
36 - 42	Integer	u[2][2]
43 - 49	Integer	u[3][3]
50 - 56	Integer	u[1][2]
57 - 63	Integer	u[1][3]
64 - 70	Integer	u[2][3]

```
73 - 76        LString(4)      Segment identifier, lef

77 - 78        LString(2)      Element symbol, right-j

79 - 80        LString(2)      Charge on the atom
```

Example:

```
         1         2         3         4         5
12345678901234567890123456789012345678901234567890123456789012345
ATOM    107  N   GLY    13      12.681  37.302 -25.211
ANISOU  107  N   GLY    13     2406   1892   1614    1
ATOM    108  CA  GLY    13      11.982  37.996 -26.241
ANISOU  108  CA  GLY    13     2748   2004   1679    -
ATOM    109  C   GLY    13      11.678  39.447 -26.008
ANISOU  109  C   GLY    13     2555   1955   1468
ATOM    110  O   GLY    13      11.444  40.201 -26.971
ANISOU  110  O   GLY    13     3837   2505   1611    1
ATOM    111  N   ASN    14      11.608  39.863 -24.755
ANISOU  111  N   ASN    14     2059   1674   1462
```

Record: TER

Contains: indicates the end of a list of ATOM/HETATM records for a ch

Notes:

- The TER records occur in the coordinate section of the
the last residue presented for each polypeptide and/or nucle
which there are coordinates. For proteins, the residue define
record is the carboxy-terminal residue; for nucleic acids it is
3'-terminal residue.
- For a cyclic molecule, the choice of termini is arbitrary.
- Terminal oxygen atoms are presented as OXT for protei
- The TER record has the same residue name, chain ider
and insertion code as the terminal residue. The serial numbe
record is one number greater than the serial number of the A
preceding the TER.
- For chains with gaps due to disorder, it is recommended
atoms be labeled O and OXT.
- The residue name appearing on the TER record must b
of the immediately preceding ATOM or non-water HETATM r

COLUMNS	DATA TYPE	CONTENTS
1 - 6	Record name	"TER "

7 - 11	Integer	Serial number
18 - 20	Residue name	Residue name
22	Character	Chain identifier
23 - 26	Integer	Residue sequence num
27	AChar	Insertion code

Example:

```
         1         2         3         4         5
12345678901234567890123456789012345678901234567890123
ATOM   4150  H   ALA A 431       8.674  16.036  12.858
TER    4151      ALA A 431

ATOM   1403  O   PRO P  22      12.701  33.564  15.827
ATOM   1404  CB  PRO P  22      13.512  32.617  18.642
ATOM   1405  CG  PRO P  22      12.828  33.382  19.740
ATOM   1406  CD  PRO P  22      12.324  34.603  18.985
HETATM 1407  CA  BLE P   1      14.625  32.240  14.151
HETATM 1408  CB  BLE P   1      15.610  33.091  13.297
HETATM 1409  CG  BLE P   1      15.558  34.629  13.373
HETATM 1410  CD1 BLE P   1      16.601  35.208  12.440
HETATM 1411  CD2 BLE P   1      14.209  35.160  12.930
HETATM 1412  N   BLE P   1      14.777  32.703  15.531
HETATM 1413  B   BLE P   1      14.921  30.655  14.194
HETATM 1414  O1  BLE P   1      14.852  30.178  12.832
HETATM 1415  O2  BLE P   1      13.775  30.147  14.862
TER    1416      BLE P   1
```

Record: HETATM

Contains: the atomic coordinate records for atoms within "non-standar groups. These records are used for water molecules and atc groups.

Notes:
- Insertion codes, segment id, and element naming are fu ATOM section of this document.
- Disordered solvents may be represented by the residue
- No ordering is specified for polysaccharides.
- HETATM records must have corresponding HET, HETN and CONECT records, except for waters.

COLUMNS	DATA TYPE	CONTENTS
1 - 6	Record name	"HETATM"
7 - 11	Integer	Atom serial number.
13 - 16	Atom	Atom name
17	Character	Alternate location indi
18 - 20	Residue name	Residue name
22	Character	Chain identifier
23 - 26	Integer	Residue sequence number
27	AChar	Code for insertion of r
31 - 38	Real(8.3)	Orthogonal coordinates
39 - 46	Real(8.3)	Orthogonal coordinates
47 - 54	Real(8.3)	Orthogonal coordinates
55 - 60	Real(6.2)	Occupancy
61 - 66	Real(6.2)	Temperature factor
73 - 76	LString(4)	Segment identifier, lef
77 - 78	LString(2)	Element symbol, right-j
79 - 80	LString(2)	Charge on the atom

Example:

```
          1         2         3         4         5
12345678901234567890123456789012345678901234567890123456789012345678901234
HETATM 1357 MG    MG   168       4.669  34.118  19.123
HETATM 3835 FE   HEM     1      17.140   3.115  15.066
```

Record: ENDMDL

Contains: these records are paired with MODEL records to group indiv

Notes:
- MODEL/ENDMDL records are used only when more tha
is presented in the entry, or if there are more than 99,999 at
- Every MODEL record has an associated ENDMDL reco

COLUMNS	DATA TYPE	CONTENTS
1 - 6	Record name	"ENDMDL"

Example:

```
         1         2         3         4         5
12345678901234567890123456789012345678901234567890123456789012345
...
...
ATOM  14550 1HG  GLU   122     -14.364  14.787 -14.258
ATOM  14551 2HG  GLU   122     -13.794  13.738 -12.961
TER   14552      GLU   122
ENDMDL
MODEL        9
ATOM  14553  N   SER     1     -28.280   1.567  12.004
ATOM  14554  CA  SER     1     -27.749   0.392  11.256
...
...
ATOM  16369 1HG  GLU   122      -3.757  18.546  -8.439
ATOM  16370 2HG  GLU   122      -3.066  17.166  -7.584
TER   16371      GLU   122
ENDMDL
MODEL       10
ATOM  16372  N   SER     1     -22.285   7.041  10.003
ATOM  16373  CA  SER     1     -23.026   6.872   8.720
...
...
ATOM  18188 1HG  GLU   122      -1.467  18.282 -17.144
ATOM  18189 2HG  GLU   122      -2.711  18.067 -15.913
TER   18190      GLU   122
ENDMDL
```