[转载]PDB 文件格式

http://blog.sciencenet.cn/blog-468005-426638.html

[转自：]http://www.cgl.ucsf.edu/chimera/docs/UsersGuide//tutorials/framepdbintro.htmlIntroduction to Protein Data Bank Format

Protein Data Bank (PDB) format is a standard for files containing atomic coordinates. It is used for structures in the Protein Data Bank and is read and written by many programs. While this short description will suffice for many users, those in need of further details should consult the definitive description. The complete PDB file specification provides for a wealth of information, including authors, literature references, and the method of structure determination.

PDB format consists of lines of information in a text file. Each line of information in the file is called a record. A PDB file generally contains several different types of records, arranged in a specific order to describe a structure.

Selected Protein Data Bank Record Types	
Record Type	Data Provided by Record
ATOM	atomic coordinate record containing the X,Y,Z orthogonal Å coordinates for atoms in standard residues (amino acids and nucleic acids).
HETATM	atomic coordinate record containing the X,Y,Z orthogonal Å coordinates for atoms in nonstandard residues. Nonstandard residues include inhibitors, cofactors, ions, and solvent. The only functional difference from ATOM records is that HETATM residues are by default not connected to other residues. Note that water residues should be in HETATM records.
TER	indicates the end of a chain of residues. For example, a hemoglobin molecule consists of four subunit chains that are not connected. TER indicates the end of a chain and prevents the display of a connection to the next chain.
HELIX	indicates the location and type (right-handed alpha, *etc.*) of helices. One record per helix.
SHEET	indicates the location, sense (anti-parallel, *etc.*) and registration with respect to the previous strand in the sheet (if any) of each strand in the model. One record per strand.
SSBOND	defines disulfide bond linkages between cysteine residues.

The formats of these record types are given in the tables below. Older PDB files may not adhere completely to the specifications. Some differences between older and newer files occur in the fields following the temperature factor in ATOM and HETATM records; these fields are omitted from the examples. Some fields are frequently blank, such as the alternate location indicator when an atom does not have alternate locations.

Protein Data Bank Format:
Coordinate Section

Record Type	Columns	Data	Justification	Data Type
ATOM	1-4	"ATOM"		character
	7-11#	Atom serial number	right	integer
	13-16	Atom name	left*	character
	17	Alternate location indicator		character
	18-20§	Residue name	right	character
	22	Chain identifier		character
	23-26	Residue sequence number	right	integer
	27	Code for insertions of residues		character
	31-38	X orthogonal Å coordinate	right	real (8.3)
	39-46	Y orthogonal Å coordinate	right	real (8.3)
	47-54	Z orthogonal Å coordinate	right	real (8.3)
	55-60	Occupancy	right	real (6.2)
	61-66	Temperature factor	right	real (6.2)
	77-78	Element symbol	right	character
	79-80	Charge		character
HETATM	1-6	"HETATM"		character
	7-80	same as ATOM records		
TER	1-3	"TER"		character
	7-11#	Serial number	right	integer
	18-20§	Residue name	right	character
	22	Chain identifier		character
	23-26	Residue sequence number	right	integer
	27	Code for insertions of residues		character

#Chimera allows (nonstandard) use of columns 6-11 for the integer atom serial number in ATOM records, and in TER records, only the "TER" is required.

*Atom names start with element symbols right-justified in columns 13-14 as permitted by the length of the name. For example, the symbol FE for iron appears in columns 13-14, whereas the symbol C for carbon appears in column 14 (see Misaligned Atom Names). If an atom name has four characters, however, it must start in column 13 even if the element symbol is a single character (for example, see Hydrogen Atoms).

§Chimera allows (nonstandard) use of four-character residue names occupying an additional column to the right.

Protein Data Bank Format:
Protein Secondary Structure and Disulfides

Record Type	Columns	Data	Justification	Data Type
HELIX	1-5	"HELIX"		character
	8-10	Helix serial number	right	integer
	12-14	Helix identifier	right	character
	16-18§	Initial residue name	right	character
	20	Chain identifier		character
	22-25	Residue sequence number	right	integer
	26	Code for insertions of residues		character
	28-30§	Terminal residue name	right	character
	32	Chain identifier		character
	34-37	Residue sequence number	right	integer
	38	Code for insertions of residues		character
	39-40	Type of helix†	right	integer
	41-70	Comment	left	character
	72-76	Length of helix	right	integer
SHEET	1-5	"SHEET"		character
	8-10	Strand number (in current sheet)	right	integer
	12-14	Sheet identifier	right	character
	15-16	Number of strands (in current sheet)	right	integer
	18-20§	Initial residue name	right	character
	22	Chain identifier		character
	23-26	Residue sequence number	right	integer
	27	Code for insertions of residues		character
	29-31§	Terminal residue name	right	character
	33	Chain identifier		character
	34-37	Residue sequence number	right	integer
	38	Code for insertions of residues		character
	39-40	Strand sense with respect to previous‡	right	integer

接下页表格

	The following fields identify two atoms involved in a hydrogen bond, the first in the current strand and the second in the previous strand. These fields should be blank for strand 1 (the first strand in a sheet).			
	42-45	Atom name (as per ATOM record)	left	character
	46-48§	Residue name	right	character
	50	Chain identifier		character
	51-54	Residue sequence number	right	integer
	55	Code for insertions of residues		character
	57-60	Atom name (as per ATOM record)	left	character
	61-63§	Residue name	right	character
	65	Chain identifier		character
	66-69	Residue sequence number	right	integer
	70	Code for insertions of residues		character
SSBOND	1-6	"SSBOND"		character
	8-10	Serial number	right	integer
	12-14	Residue name ("CYS")	right	character
	16	Chain identifier		character
	18-21	Residue sequence number	right	integer
	22	Code for insertions of residues		character
	26-28	Residue name ("CYS")	right	character
	30	Chain identifier		character
	32-35	Residue sequence number	right	integer
	36	Code for insertions of residues		character
	60-65	Symmetry operator for first residue	right	integer
	67-72	Symmetry operator for second residue	right	integer
	74-78	Length of disulfide bond	right	real (5.2)

†Helix types:

1	Right-handed alpha (default)	6	Left-handed alpha
2	Right-handed omega	7	Left-handed omega
3	Right-handed pi	8	Left-handed gamma
4	Right-handed gamma	9	2/7 ribbon/helix
5	Right-handed 3/10	10	Polyproline

‡Sense is 0 for strand 1 (the first strand in a sheet), 1 for parallel, and –1 for antiparallel.

For those who are familiar with the FORTRAN programming language, the following format descriptions will be meaningful. Those unfamiliar with FORTRAN should ignore this gibberish:

ATOM **HETATM**	Format (A6,I5,1X,A4,A1,A3,1X,A1,I4,A1,3X,3F8.3,2F6.2,10X,A2,A2)
HELIX	Format (A6,1X,I3,1X,A3,2(1X,A3,1X,A1,1X,I4,A1),I2,A30,1X,I5)
SHEET	Format (A6,1X,I3,1X,A3,I2,2(1X,A3,1X,A1,I4,A1),I2,2(1X,A4,A3,1X,A1,I4,A1))
SSBOND	Format (A6,1X,I3,1X,A3,1X,A1,1X,I4,A1,3X,A3,1X,A1,1X,I4,A1,23X,2(2I3,1X),F5.2)

Examples of PDB Format

Glucagon is a small protein of 29 amino acids in a single chain. The first residue is the amino-terminal amino acid, histidine, which is followed by a serine residue and then a glutamine. The coordinate information (entry 1gcn) starts with:

```
ATOM  1   N    HIS  A  1 49.668 24.248 10.436 1.00 25.00 N
ATOM  2   CA   HIS  A  1 50.197 25.578 10.784 1.00 16.00 C
ATOM  3   C    HIS  A  1 49.169 26.701 10.917 1.00 16.00 C
ATOM  4   O    HIS  A  1 48.241 26.524 11.749 1.00 16.00 O
ATOM  5   CB   HIS  A  1 51.312 26.048 9.843  1.00 16.00 C
ATOM  6   CG   HIS  A  1 50.958 26.068 8.340  1.00 16.00 C
ATOM  7   ND1  HIS  A  1 49.636 26.144 7.860  1.00 16.00 N
ATOM  8   CD2  HIS  A  1 51.797 26.043 7.286  1.00 16.00 C
ATOM  9   CE1  HIS  A  1 49.691 26.152 6.454  1.00 17.00 C
ATOM  10  NE2  HIS  A  1 51.046 26.090 6.098  1.00 17.00 N
ATOM  11  N    SER  A  2 49.788 27.850 10.784 1.00 16.00 N
ATOM  12  CA   SER  A  2 49.138 29.147 10.620 1.00 15.00 C
ATOM  13  C    SER  A  2 47.713 29.006 10.110 1.00 15.00 C
ATOM  14  O    SER  A  2 46.740 29.251 10.864 1.00 15.00 O
ATOM  15  CB   SER  A  2 49.875 29.930 9.569  1.00 16.00 C
ATOM  16  OG   SER  A  2 49.145 31.057 9.176  1.00 19.00 O
ATOM  17  N    GLN  A  3 47.620 28.367 8.973  1.00 15.00 N
ATOM  18  CA   GLN  A  3 46.287 28.193 8.308  1.00 14.00 C
ATOM  19  C    GLN  A  3 45.406 27.172 8.963  1.00 14.00 C
```

Notice that each line or record begins with the record type ATOM. The atom serial number is the next item in each record.

The atom name is the third item in the record. Notice that the first one or two characters of the atom name consists of the chemical symbol for the atom type. All the atom names beginning with C are carbon atoms; N indicates a nitrogen and O indicates oxygen. In amino acid residues, the next character is the remoteness indicator code, which is transliterated according to:

α	A
β	B
γ	G
δ	D
ε	E
ζ	Z
η	H

The next character of the atom name is a branch indicator, if required.

The next data field is the residue type. Notice that each record contains the residue type. In this example, the first residue in the chain is HIS (histidine) and the second residue is a SER (serine).

The next data field contains the chain identifier, in this case A.

The next data field contains the residue sequence number. Notice that as the residue changes from histidine to serine, the residue number changes from 1 to 2. Two like residues may be adjacent to one another, so the residue number is important for distinguishing between them.

The next three data fields contain the X, Y, and Z coordinate values, respectively. The last three fields shown are the occupancy, temperature factor (B-factor), and element symbol.

The spacing of the data fields is crucial. If a data field does not apply, it should be left blank.

The glucagon data file continues in this manner until the final residue is reached:

```
ATOM 239 N  THR A 29 3.391 19.940 12.762 1.00 21.00  N
ATOM 240 CA THR A 29 2.014 19.761 13.283 1.00 21.00  C
ATOM 241 C  THR A 29 0.826 19.943 12.332 1.00 23.00  C
ATOM 242 O  THR A 29 0.932 19.600 11.133 1.00 30.00  O
```

```
ATOM 243 CB  THR A 29 1.845 20.667 14.505 1.00 21.00   C
ATOM 244 OG1 THR A 29 1.214 21.893 14.153 1.00 21.00   O
ATOM 245 CG2 THR A 29 3.180 20.968 15.185 1.00 21.00   C
ATOM 246 OXT THR A 29 -0.317 20.109 12.824 1.00 25.00  O
TER  247     THR A 29
```

Note that this residue includes the extra oxygen atom OXT on the terminal carboxyl group. Other than OXT and the rarely seen HXT, atoms in standard nucleotides and amino acids in version 3.0 PDB files are named according to the IUPAC recommendations (Pure Appl Chem 70: 117 (1998) [abstract] [PDF]). The TER record terminates the amino acid chain.

A more complicated protein, hemoglobin, consists of four amino acid chains, each with an associated heme group. There are two alpha chains (identifiers A and C) and two beta chains (identifiers B and D). The first ten lines of coordinates for this molecule (entry 3hhb) are:

```
ATOM 1  N   VAL A 1  6.452 16.459 4.843 7.00 47.38 N
ATOM 2  CA  VAL A 1  7.060 17.792 4.760 6.00 48.47  C
ATOM 3  C   VAL A 1  8.561 17.703 5.038 6.00 37.13  C
ATOM 4  O   VAL A 1  8.992 17.182 6.072 8.00 36.25  O
ATOM 5  CB  VAL A 1  6.342 18.738 5.727 6.00 55.13  C
ATOM 6  CG1 VAL A 1  7.114 20.033 5.993 6.00 54.30  C
ATOM 7  CG2 VAL A 1  4.924 19.032 5.232 6.00 64.75  C
ATOM 8  N    LEU A 2  9.333 18.209 4.095 7.00 30.18  N
ATOM 9  CA   LEU A 2 10.785 18.159 4.237 6.00 35.60 C
ATOM 10 C    LEU A 2 11.247 19.305 5.133 6.00 35.47 C
```

At the end of chain A, the heme group records appear:

```
ATOM 1058 N  ARG A 141 -6.466 12.036 -10.348 7.00 19.11 N
ATOM 1059 CA ARG A 141 -7.922 12.248 -10.253 6.00 26.80 C
ATOM 1060 C  ARG A 141 -8.119 13.499 -9.393   6.00 28.93 C
ATOM 1061 O  ARG A 141 -7.112 13.967 -8.853   8.00 28.68 O
ATOM 1062 CB ARG A 141 -8.639 11.005 -9.687   6.00 24.11 C
```

```
ATOM 1063 CG  ARG A 141 -8.153  10.551 -8.308 6.00 19.20 C
ATOM 1064 CD  ARG A 141 -8.914  9.319  -7.796 6.00 21.53 C
ATOM 1065 NE  ARG A 141 -8.517  9.076  -6.403 7.00 20.93 N
ATOM 1066 CZ  ARG A 141 -9.142  8.234  -5.593 6.00 23.56 C
ATOM 1067 NH1 ARG A 141 -10.150 7.487 -6.019 7.00 19.04 N
ATOM 1068 NH2 ARG A 141 -8.725  8.129  -4.343 7.00 25.11 N
ATOM 1069 OXT ARG A 141 -9.233  14.024 -9.296 8.00 40.35 O
TER  1070     ARG A 141
HETATM 1071 FE  HEM A 1 8.128  7.371   -15.022 24.00 16.74 FE
HETATM 1072 CHA HEM A 1 8.617  7.879   -18.361 6.00  17.74 C
HETATM 1073 CHB HEM A 1 10.356 10.005 -14.319 6.00  18.92 C
HETATM 1074 CHC HEM A 1 8.307  6.456   -11.669 6.00  11.00 C
HETATM 1075 CHD HEM A 1 6.928  4.145   -15.725 6.00  13.25 C
```

The last residue in the alpha chain is an ARG (arginine). Again, the extra oxygen atom OXT appears in the terminal carboxyl group. The TER record indicates the end of the peptide chain. It is important to have TER records at the end of peptide chains so a bond is not drawn from the end of one chain to the start of another.

In the example above, the TER record is correct and should be present, but the molecule chain would still be terminated at that point even without a TER record, because HETATM residues are not connected to other residues or to each other. The heme group is a single residue made up of HETATM records.

After the heme group associated with chain A, chain B begins:

```
HETATM 1109 CAD HEM A 1 7.618   5.696   -20.432  6.00 21.38 C
HETATM 1110 CBD HEM A 1 8.947   5.143   -20.947  6.00 29.03 C
HETATM 1111 CGD HEM A 1 9.047   5.155   -22.461  6.00 30.08 C
HETATM 1112 O1D HEM A 1 10.139  5.458   -22.959  8.00 33.72 O
HETATM 1113 O2D HEM A 1 8.096   4.833   -23.177  8.00 33.55 O
ATOM   1114 N   VAL B 1 9.143  -20.582   1.231   7.00 48.92 N
```

```
ATOM 1115 CA   VAL B 1 8.824  -20.084 -0.109 6.00 52.26 C
ATOM 1116 C    VAL B 1 9.440  -20.964 -1.190 6.00 57.72 C
ATOM 1117 O    VAL B 1 9.768  -22.138 -0.985 8.00 55.05 O
ATOM 1118 CB   VAL B 1 9.314  -18.642 -0.302 6.00 58.48 C
ATOM 1119 CG1  VAL B 1 8.269  -17.606  0.113 6.00 59.43 C
ATOM 1120 CG2  VAL B 1 10.683 -18.373 0.331 6.00 45.96 C
```

Here the TER card is implicit in the start of a new chain.

Protein Data Bank format relies on the concept of residues:

- Each atom in a residue must be uniquely identifiable. Two atoms in the same residue can only have the same name if they have different alternate location identifiers.
- Residue names are a maximum of three characters long§ and uniquely identify the residue type. Thus, all residues of a given name should be the same type of residue and have the same structure (contain the same atoms with the same connectivity).

Common Errors in PDB Format Files

If a data file fails to display correctly, it is sometimes difficult to determine where in the hundreds of lines of data the mistake occurred. This section enumerates some of the most common errors found in PDB files.

Program-Generated PDB Files Spurious Long Bonds

A couple of common errors in program-generated PDB files result in the display of very long bonds between residues:

- Missing TER cards - Either a TER card or a change in the chain ID is needed to mark the end of a chain.
- Improper use of ATOM records instead of HETATM records - HETATM records should be employed for compounds that do not form chains, such as water or heme. The first six columns of the ATOM record should be changed to HETATM so that the remaining columns stay aligned correctly.

Apart from any format errors, Chimera also uses long bonds to indicate the underlying connectivity across chain segments that lack coordinates (e.g., regions of missing density due to

crystallographic disorder). Regardless of their cause, long bonds in Chimera can be hidden with the command ~longbond. Misaligned Atom Names

Incorrectly aligned atom names in PDB records can cause problems. Atom names are composed of an atomic (element) symbol right-justified in columns 13-14, and trailing identifying characters left-justified in columns 15-16. A single-character element symbol should not appear in column 13 unless the atom name has four characters (for example, see Hydrogen Atoms). Many programs simply left-justify all atom names starting in column 13. The difference can be seen clearly in a short segment of hemoglobin (entry 3hhb):

Correct:

```
HETATM 1071 FE  HEM A  1 8.128 7.371   -15.022 24.00 16.74 FE
HETATM 1072 CHA HEM A 1 8.617  7.879   -18.361 6.00  17.74 C
HETATM 1073 CHB HEM A 1 10.356 10.005 -14.319 6.00  18.92 C
HETATM 1074 CHC HEM A 1 8.307  6.456   -11.669 6.00  11.00 C
HETATM 1075 CHD HEM A 1 6.928  4.145   -15.725 6.00  13.25 C
```

Incorrect:

```
HETATM 1071 FE  HEM A 1 8.128  7.371   -15.022 24.00 16.74 FE
HETATM 1072 CHA HEM A 1 8.617  7.879   -18.361 6.00  17.74 C
HETATM 1073 CHB HEM A 1 10.356 10.005 -14.319 6.00  18.92 C
HETATM 1074 CHC HEM A 1 8.307  6.456   -11.669 6.00  11.00 C
HETATM 1075 CHD HEM A 1 6.928  4.145   -15.725 6.00  13.25 C
```

Hand-Edited PDB Files Duplicate Atom Names

One possible editing mistake is the failure to uniquely name all atoms within a given residue. In the following example, two atoms in the same residue are named CA:

```
ATOM 185 N  VAL A 23 13.455 17.883 10.517 1.00 7.00   N
ATOM 186 CA VAL A 23 12.574 17.403 11.589 1.00 7.00   C
ATOM 187 C  VAL A 23 11.283 18.205 11.729 1.00 7.00   C
ATOM 188 O  VAL A 23 10.233 17.600 12.052 1.00 7.00   O
ATOM 189 CA VAL A 23 13.339 17.278 12.906 1.00 10.00 C
```

```
ATOM 190 CG1 VAL A 23 12.441 17.004 14.108 1.00 13.00 C
ATOM 191 CG2 VAL A 23 14.455 16.248 12.794 1.00 13.00 C
ATOM 192 N   GLN A 24 11.255 19.253 10.941 1.00 8.00  N
ATOM 193 CA  GLN A 24 10.082 20.114 10.818 1.00 8.00 C
ATOM 194 C   GLN A 24 9.158  19.638 9.692  1.00 8.00 C
```

Depending on the display program, the residue may be shown with incorrect connectivity, or it may become evident only upon labeling that the residue is missing a CB atom. Residues Out of Sequence

In the following example, the second residue in the file is erroneously numbered residue 5. Many display programs will show this residue as connected to residues 1 and 3. If this residue was meant to be connected to residues 4 and 6 instead, it should appear between those residues in the PDB file.

```
ATOM 1  N   HIS A  1 49.668 24.248 10.436 1.00 25.00 N
ATOM 2  CA  HIS A  1 50.197 25.578 10.784 1.00 16.00 C
ATOM 3  C   HIS A  1 49.169 26.701 10.917 1.00 16.00 C
ATOM 4  O   HIS A  1 48.241 26.524 11.749 1.00 16.00 O
ATOM 5  CB  HIS A  1 51.312 26.048 9.843  1.00 16.00 C
ATOM 6  CG  HIS A  1 50.958 26.068 8.340  1.00 16.00 C
ATOM 7  ND1 HIS A  1 49.636 26.144 7.860  1.00 16.00 N
ATOM 8  CD2 HIS A  1 51.797 26.043 7.286  1.00 16.00 C
ATOM 9  CE1 HIS A  1 49.691 26.152 6.454  1.00 17.00 C
ATOM 10 NE2 HIS A  1 51.046 26.090 6.098  1.00 17.00 N
ATOM 11 N   SER A 5 49.788 27.850 10.784 1.00 16.00 N
ATOM 12 CA  SER A 5 49.138 29.147 10.620 1.00 15.00 C
ATOM 13 C   SER A 5 47.713 29.006 10.110 1.00 15.00 C
ATOM 14 O   SER A 5 46.740 29.251 10.864 1.00 15.00 O
ATOM 15 CB  SER A  5 49.875 29.930 9.569  1.00 16.00 C
ATOM 16 OG  SER A  5 49.145 31.057 9.176  1.00 19.00 O
```

```
ATOM 17 N  GLN A 3 47.620 28.367 8.973 1.00 15.00 N
ATOM 18 CA GLN A 3 46.287 28.193 8.308 1.00 14.00 C
```

Common Typos

Sometimes the letter l is accidentally substituted for the number 1. This has different repercussions depending on where in the file the error occurs; a grossly misplaced atom may indicate the presence of such an error in a coordinate field. These errors can be located readily if the text of the data file appears in uppercase, by invoking a text editor to search for all instances of the lowercase letter l.

Hydrogen Atoms

In brief, conventions for hydrogen atoms in version 3.0 PDB format are as follows:

- Hydrogen atom records follow the records of all other atoms of a particular residue.
- A hydrogen atom name starts with H. The next part of the name is based on the name of the connected nonhydrogen atom. For example, in amino acid residues, H is followed by the remoteness indicator (if any) of the connected atom, followed by the branch indicator (if any) of the connected atom; if more than one hydrogen is connected to the same atom, an additional digit is appended so that each hydrogen atom will have a unique name. Hydrogen atoms in standard nucleotides and amino acids (other than the rarely seen HXT) are named according to the IUPAC recommendations (Pure Appl Chem 70:117 (1998) [abstract] [PDF]). Names of hydrogen atoms in HETATM residues are determined in a similar fashion.
- If the name of a hydrogen has four characters, it is left-justified starting in column 13; if it has fewer than four characters, it is left-justified starting in column 14.

In the following excerpt from entry 1vm3, atom H is attached to atom N. Atom HA is attached to atom CA; the remoteness indicator A is the same for these atoms. Two hydrogen atoms are connected to CB, one is connected to CG, three are connected to CD1, and three are connected to CD2.

```
ATOM 10 N   LEU A 2 4.595 6.365 3.756 1.00 0.00 N
ATOM 11 CA  LEU A 2 4.471 5.443 2.633 1.00 0.00 C
ATOM 12 C   LEU A 2 5.841 5.176 2.015 1.00 0.00 C
```

```
ATOM 13 O    LEU A 2 6.205 4.029 1.755 1.00 0.00 O
ATOM 14 CB   LEU A 2 3.526 6.037 1.578 1.00 0.00 C
ATOM 15 CG   LEU A 2 2.790 4.919 0.823 1.00 0.00 C
ATOM 16 CD1  LEU A 2 3.803 3.916 0.262 1.00 0.00 C
ATOM 17 CD2  LEU A 2 1.817 4.196 1.769 1.00 0.00 C
ATOM 18 H    LEU A 2 4.169 7.246 3.704  1.00 0.00 H
ATOM 19 HA   LEU A 2 4.063 4.514 2.992  1.00 0.00 H
ATOM 20 HB2  LEU A 2 2.804 6.675 2.065  1.00 0.00 H
ATOM 21 HB3  LEU A 2 4.099 6.623 0.873  1.00 0.00 H
ATOM 22 HG   LEU A 2 2.234 5.353 0.004  1.00 0.00 H
ATOM 23 HD11 LEU A 2 4.648 4.447 -0.148 1.00 0.00 H
ATOM 24 HD12 LEU A 2 3.334 3.331 -0.516 1.00 0.00 H
ATOM 25 HD13 LEU A 2 4.137 3.260 1.052  1.00 0.00 H
ATOM 26 HD21 LEU A 2 0.941 3.892 1.216  1.00 0.00 H
ATOM 27 HD22 LEU A 2 1.522 4.860 2.568  1.00 0.00 H
ATOM 28 HD23 LEU A 2 2.296 3.323 2.188  1.00 0.00 H
```

PQR Variant of PDB Format

Several programs use a modified PDB format called PQR, in which atomic partial charge (Q) and radius (R) fields follow the X,Y,Z coordinate fields in ATOM and HETATM records.

The format is quite loosely defined and varies according to which program is producing or using the format. For example, APBS requires only that all fields be whitespace-delimited.

If an ATOM or HETATM record being read by Chimera is not in PDB format, Chimera next tries to read it as PQR format. In that case, all fields up to and including the coordinates are still expected to adhere to the standard format, but the next two eight-column fields are each expected to contain a floating-point number: charge is read from columns 55-62 and radius is read from columns 63-70. The values are assigned as the atom attributes charge and radius, respectively.

For example, PDB2PQR version 1.6 generates PQR files that can be read as such by Chimera. An excerpt :

```
ATOM 1  N    ALA 1 46.457 12.189 21.556 0.1414   1.8240
ATOM 2  CA   ALA 1 47.614 11.997 22.448 0.0962   1.9080
ATOM 3  C    ALA 1 47.538 12.947 23.645 0.6163   1.9080
ATOM 4  O    ALA 1 46.441 13.476 23.962 -0.5722  1.6612
ATOM 5  CB   ALA 1 48.911 12.134 21.650 -0.0597  1.9080
ATOM 6  H2   ALA 1 45.672 11.684 21.917 0.1997   0.6000
ATOM 7  H3   ALA 1 46.235 13.163 21.506 0.1997   0.6000
ATOM 8  H    ALA 1 46.683 11.849 20.642 0.1997   0.6000
ATOM 9  HA   ALA 1 47.603 11.052 22.786 0.0889   1.1000
ATOM 10 HB1 ALA 1 49.041 11.319 21.087 0.0300   1.4870
ATOM 11 HB3 ALA 1 48.855 12.941 21.064 0.0300   1.4870
ATOM 12 HB2 ALA 1 49.679 12.231 22.281 0.0300   1.4870
ATOM 13 N   ASP  2 48.702 13.128 24.279 -0.5163  1.8240
ATOM 14 CA  ASP  2 48.826 13.956 25.493 0.0381   1.9080
ATOM 15 C   ASP  2 48.614 15.471 25.323 0.5366   1.9080
ATOM 16 O   ASP  2 49.292 16.362 24.807 -0.5819  1.6612
ATOM 17 CB  ASP  2 50.156 13.635 26.226 -0.0303  1.9080
ATOM 18 CG  ASP  2 49.984 12.419 27.136 0.7994   1.9080
ATOM 19 OD1 ASP  2 50.595 12.308 28.221 -0.8014 1.6612
ATOM 20 OD2 ASP  2 49.198 11.502 26.778 -0.8014 1.6612
ATOM 21 H    ASP  2 49.511 12.637 23.845 0.2936  0.6000
ATOM 22 HA  ASP   2 48.104 13.630 26.146 0.0880  1.3870
ATOM 23 HB3  ASP  2 50.392 14.413 26.773 -0.0122 1.4870
ATOM 24 HB2  ASP  2 50.832 13.431 25.545 -0.0122 1.4870
```

Note that PDBPQR strips chain IDs, so residues in a multichain structure written out by that program may not be uniquely Specifiable in Chimera.