1.1. Introduction to Protein Data Bank Format

Protein Data Bank (PDB) format is a standard for files containing atomic coordinates. Structures deposited in the Protein Data Bank at the Research Collaboratory for Structural Bioinformatics (RCSB) are written in this standardized format. The short description provided here will suffice for most users. However, those actually creating PDB files should consult the definitive description (see http://www.rcsb.org/pdb/info.html#File_Formats_and_Standards).

The complete PDB file specification provides for a wealth of information, including authors, literature references, and the identification of substructures such as disulfide bonds, helices, sheets, and active sites. Users should bear in mind that modeling programs can be unforgiving of incorrect input formats.

1.1.1. Description

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

Selected Protein Data Bank Record Types

Record Type	
ATOM	atomic coordinate record containing the x,y,z orthogonal Angstrom coordinates for atoms in standard residues (amino acids and nucleic acids).
HETATM	atomic coordinate record containing the x,y,z orthogonal Angstrom 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 which are not connected. TER indicates the end of a chain and prevents the display of a connection to the next chain.
SSBOND	defines disulfide bond linkages between cysteine residues.
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.

The following table describes the format for selected record types. Older PDB files may not adhere completely to the newer format specification. From the standpoint of most users, the most notable differences between older and newer files occur in the fields following the temperature factor in ATOM and HETATM records. These fields are not included in the examples in the subsequent sections. Furthermore, some fields are frequently blank, such as the alternate location indicator when an atom does not have alternate locations.

Protein Data Bank Format

Record Type	Columns	Data	Justification	Data Type
ATOM	1-4	‘‘ATOM’’	left	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 Angstrom coordinate	right	floating
	39-46	Y orthogonal Angstrom coordinate	right	floating
	47-54	Z orthogonal Angstrom coordinate	right	floating
	55-60	Occupancy	right	floating
	61-66	Temperature factor	right	floating
	73-76	Segment identifier (optional)	left	character
	77-78	Element symbol	right	character
	79-80	Charge (optional)		character
HETATM	1-6	‘‘HETATM’’		
	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
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

*The chemical symbols of atoms are right-justified in columns 13-14. For example, the symbol ‘‘FE’’ for iron appears in columns 13-14, whereas the symbol ‘‘C’’ for carbon appears in column 14.

Protein Data Bank Format
(*continued*)

Record Type	Columns	Data	Justification	Data Type
HELIX	1-5	‘‘HELIX’’	left	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
		The following fields identify two atoms, the first in the current strand and the second in the previous strand, which are hydrogen bonded to each other. These fields should be blank for strand 1.		
	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

[†] Helix types are:

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

[‡] Parallel is indicated with ‘‘1,’’ anti-parallel with ‘‘−1.’’ Strand 1 has sense indicator ‘‘0.’’

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

ATOM **HETATM**	Format (A6,I5,1X,A4,A1,A3,1X,A1,I4,A1,3X,3F8.3,2F6.2,6X,A4,A2,A2)
SSBOND	Format (A6,1X,I3,1X,A3,1X,A1,1X,I4,A1,3X,A3,1X,A1,1X,I4,A1,23X,2I3,1X,2I3)
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))

1.1.2. Examples of PDB Format

Fields following the temperature factor in ATOM and HETATM records are not shown in any of the examples.

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 starts with:

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

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. 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 last 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 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 very important for distinguishing between them.

The next three data fields contain the X, Y, and Z coordinate values, respectively. The next data field is the occupancy. The final field shown is the temperature factor (B value).

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

```
ATOM    239  N   THR    29       3.391  19.940  12.762  1.00 21.00
ATOM    240  CA  THR    29       2.014  19.761  13.283  1.00 21.00
ATOM    241  C   THR    29        .826  19.943  12.332  1.00 23.00
ATOM    242  O   THR    29        .932  19.600  11.133  1.00 30.00
ATOM    243  CB  THR    29       1.845  20.667  14.505  1.00 21.00
ATOM    244  OG1 THR    29       1.214  21.893  14.153  1.00 21.00
ATOM    245  CG2 THR    29       3.180  20.968  15.185  1.00 21.00
ATOM    246  OXT THR    29       -.317  20.109  12.824  1.00 25.00
TER     247      THR    29
```

Note that this residue includes the extra oxygen atom, ‘‘OXT,’’ on the terminal carboxyl group. The ‘‘TER’’ record terminates the amino acid chain.

A more complicated protein, fetal hemoglobin, consists of two amino acid chains (alpha and gamma) and two heme groups. The first ten lines of coordinates for this molecule are:

```
ATOM      1  N   VAL A   1       6.280  17.225   4.929  1.00   0.00
ATOM      2  CA  VAL A   1       6.948  18.508   4.671  1.00   0.00
ATOM      3  C   VAL A   1       8.436  18.338   4.977  1.00   0.00
ATOM      4  O   VAL A   1       8.813  17.657   5.941  1.00   0.00
ATOM      5  CB  VAL A   1       6.317  19.598   5.527  1.00   0.00
ATOM      6  CG1 VAL A   1       6.959  20.999   5.376  1.00   0.00
ATOM      7  CG2 VAL A   1       4.819  19.636   5.383  1.00   0.00
ATOM      8  N   LEU A   2       9.259  18.958   4.152  1.00   0.00
ATOM      9  CA  LEU A   2      10.715  18.872   4.330  1.00   0.00
ATOM     10  C   LEU A   2      11.156  20.058   5.187  1.00   0.00
```

This data file appears much the same as the file for glucagon, with the exception that the fifth data field now contains the single-character chain indicator. In this case, the chain indicator is ‘‘A,’’ denoting the alpha chain of the hemoglobin molecule. This field was simply blank in the glucagon example. At the end of chain A, the heme group records appear:

```
ATOM   1058  N   ARG A 141      -6.576  12.834 -10.275  1.00  0.00
ATOM   1059  CA  ARG A 141      -8.044  12.831 -10.214  1.00  0.00
ATOM   1060  C   ARG A 141      -8.186  14.096  -9.365  1.00  0.00
ATOM   1061  O   ARG A 141      -7.591  15.139  -9.671  1.00  0.00
ATOM   1062  CB  ARG A 141      -8.579  11.531  -9.580  1.00  0.00
ATOM   1063  CG  ARG A 141      -8.386  11.441  -8.054  1.00  0.00
ATOM   1064  CD  ARG A 141      -8.727  10.045  -7.568  1.00  0.00
ATOM   1065  NE  ARG A 141      -9.095  10.056  -6.143  1.00  0.00
ATOM   1066  CZ  ARG A 141      -9.268   8.931  -5.414  1.00  0.00
ATOM   1067  NH1 ARG A 141      -8.602   8.795  -4.282  1.00  0.00
ATOM   1068  NH2 ARG A 141     -10.097   7.962  -5.830  1.00  0.00
ATOM   1069  OXT ARG A 141      -8.973  13.984  -8.310  1.00  0.00
TER    1070      ARG A 141
HETATM 1071 FE   HEM A   1       8.133   8.321 -15.014  1.00  0.00
HETATM 1072  CHA HEM A   1       8.863   8.752 -18.417  1.00  0.00
HETATM 1073  CHB HEM A   1      10.362  10.946 -14.389  1.00  0.00
HETATM 1074  CHC HEM A   1       8.482   7.374 -11.743  1.00  0.00
HETATM 1075  CHD HEM A   1       6.982   5.180 -15.773  1.00  0.00
HETATM 1076  N A HEM A   1       9.452   9.545 -16.178  1.00  0.00
```

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.

At the end of the heme group associated with the alpha chain, the gamma chain begins:

```
HETATM 1109  CAD HEM A   1       7.582   6.731 -20.480  1.00  0.00
HETATM 1110  CBD HEM A   1       8.992   6.848 -20.968  1.00  0.00
HETATM 1111  CGD HEM A   1       8.998   6.529 -22.465  1.00  0.00
HETATM 1112  O1D HEM A   1       9.693   5.683 -22.895  1.00  0.00
HETATM 1113  O2D HEM A   1       8.276   7.153 -23.229  1.00  0.00
ATOM   1114  C   ACE G   0       7.896 -18.462  -1.908  1.00  0.00
ATOM   1115  O   ACE G   0       7.246 -18.839   -.922  1.00  0.00
ATOM   1116  CH3 ACE G   0       9.415 -18.301  -1.832  1.00  0.00
ATOM   1117  N   GLY G   1       7.354 -18.174  -3.077  1.00  0.00
ATOM   1118  CA  GLY G   1       5.904 -18.282  -3.283  1.00  0.00
ATOM   1119  C   GLY G   1       7.139 -19.112  -2.930  1.00  0.00
ATOM   1120  O   GLY G   1       7.026 -20.248  -2.448  1.00  0.00
ATOM   1121  N   HIS G   2       8.300 -18.533  -3.176  1.00  0.00
ATOM   1122  CA  HIS G   2       9.565 -19.224  -2.889  1.00  0.00
```

Here the ‘‘TER’’ card is implicit in the start of a new chain. The new chain identifier is ‘‘G.’’ The file continues in the same pattern as before until the entire gamma chain and its associated heme group have been specified.

The spacing of the data fields is crucial. If a data field does not apply, it should be left blank. For example, a protein which consists of a single amino acid chain has no chain identifier, and thus column 22 is blank.

From this example, it is apparent that Protein Data Bank format relies on the concept of *residues*. The rules for residues can be summarized as:

(1) All atoms within a single residue must have unique names. For example, residue ''VAL'' may have only one atom named ''CA.'' Other residues may also have a ''CA'' atom but not more than one ''CA'' may appear in ''VAL.''

(2) Residue names are a maximum of three characters long and uniquely identify the residue type. Thus, all residues of a given name in a file will be the same type of residue and have the same structure. Each occurrence of a particular residue in the Protein Data Bank file should have the same atoms with the same connectivity.

1.1.3. 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.

1.1.4. 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 that should be disconnected.

One such error is the omission of `TER` cards at the end of molecule chains. According to the PDB standard, `TER` cards mark the end of molecule chains. They should be inserted in the file, if missing. Alternatively, all chains could be marked with differing chain IDs.

A second common cause of long bonds is the 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. Many programs generate files that fail to employ `HETATM` records appropriately. The first *six* columns of the `ATOM` record should be changed to `HETATM` so that the remaining columns stay aligned correctly.

Misaligned Atom Names

Incorrectly aligned atom names in PDB records can cause problems. Atom names are composed of an atomic symbol (such as ''C''), *right*-justified in columns 13-14 of `ATOM` and `HETATM` records, and trailing identifying characters (such as ''A'') *left*-justified in columns 15-16. Many programs simply left-justify the entire atom name starting in column 13. The difference can be seen clearly in a short segment of hemoglobin:

correct

```
HETATM  976 FE   HEM     1      12.763  34.157   9.102  1.00  0.00
HETATM  977  CHA HEM     1      16.124  33.461  10.405  1.00  0.00
HETATM  978  CHB HEM     1      11.350  32.580  12.046  1.00  0.00
HETATM  979  CHC HEM     1        9.326  34.709   7.887  1.00  0.00
HETATM  980  CHD HEM     1      14.138  35.379   6.119  1.00  0.00
```

incorrect

```
HETATM  976 FE   HEM     1      12.763  34.157   9.102  1.00  0.00
HETATM  977 CHA  HEM     1      16.124  33.461  10.405  1.00  0.00
HETATM  978 CHB  HEM     1      11.350  32.580  12.046  1.00  0.00
HETATM  979 CHC  HEM     1        9.326  34.709   7.887  1.00  0.00
HETATM  980 CHD  HEM     1      14.138  35.379   6.119  1.00  0.00
```

1.1.5. 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 residue VAL are named CA:

```
ATOM      1  N   VAL A   1       6.280  17.225   4.929  1.00  0.00
ATOM      2  CA  VAL A   1       6.948  18.508   4.671  1.00  0.00
ATOM      3  C   VAL A   1       8.436  18.338   4.977  1.00  0.00
ATOM      4  O   VAL A   1       8.813  17.657   5.941  1.00  0.00
ATOM      5  CA  VAL A   1       6.317  19.598   5.527  1.00  0.00
ATOM      6  CG1 VAL A   1       6.959  20.999   5.376  1.00  0.00
ATOM      7  CG2 VAL A   1       4.819  19.636   5.383  1.00  0.00
ATOM      8  N   LEU A   2       9.259  18.958   4.152  1.00  0.00
ATOM      9  CA  LEU A   2      10.715  18.872   4.330  1.00  0.00
ATOM     10  C   LEU A   2      11.156  20.058   5.187  1.00  0.00
```

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 (SER) appearing in the file is erroneously numbered residue 5. Many display programs will show residue 5 as connected to residue 1 and residue 3. This may be correct, but only if it is what was originally intended. If, however, residue number 5 was supposed to appear between residue 4 and residue 6, it should have appeared in that position in the PDB file.

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

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.’’

1.2. Modeling Hydrogen Atoms

The conventions for hydrogen atoms in PDB files are as follows:

(1) Hydrogen atoms appear as ATOM records following the ATOM records of all other atoms of a particular residue.

(2) The name of each hydrogen atom is determined by the name of the atom to which it is connected:

The first space of the name (column 13) is an optional digit to be used if two or more hydrogens are attached to the same atom.

The second column, 14, is used for the chemical symbol, ‘‘H.’’

The next two columns contain the remoteness and branch indicators (one or two characters) of the atom to which the hydrogen is attached.

For example,

```
ATOM      1  N    VAL     1     -13.090   1.966   9.741  1.00  0.00
ATOM      2  CA   VAL     1     -12.852   3.121   8.892  1.00  0.00
ATOM      3  C    VAL     1     -13.047   4.399   9.711  1.00  0.00
ATOM      4  O    VAL     1     -12.143   5.228   9.800  1.00  0.00
ATOM      5  CB   VAL     1     -13.753   3.058   7.658  1.00  0.00
ATOM      6  CG1  VAL     1     -13.930   4.446   7.036  1.00  0.00
ATOM      7  CG2  VAL     1     -13.208   2.063   6.631  1.00  0.00
ATOM      8  H    VAL     1     -13.919   1.449   9.527  1.00  0.00
ATOM      9  HA   VAL     1     -11.816   3.075   8.557  1.00  0.00
ATOM     10  HB   VAL     1     -14.734   2.707   7.977  1.00  0.00
ATOM     11 1HG1  VAL     1     -13.951   4.357   5.950  1.00  0.00
ATOM     12 2HG1  VAL     1     -14.866   4.883   7.384  1.00  0.00
ATOM     13 3HG1  VAL     1     -13.098   5.085   7.333  1.00  0.00
ATOM     14 1HG2  VAL     1     -12.623   1.298   7.142  1.00  0.00
ATOM     15 2HG2  VAL     1     -14.039   1.594   6.104  1.00  0.00
ATOM     16 3HG2  VAL     1     -12.575   2.588   5.917  1.00  0.00
```

Note in this example that:

- All hydrogens appear after the other atoms of the residue.
- Atom 9, ‘‘HA’’ is attached to atom 2, ‘‘CA.’’ The remoteness indicator ‘‘A’’ is the same for these atoms.
- There are three hydrogen atoms connected to ‘‘CG1.’’ They all have the same remoteness and branch indicators, but contain a distinguishing digit in column 13. Thus, each has a unique name.
- It is not necessary to use a digit as a prefix to the atom name when only one hydrogen is attached to a given atom.