Guidelines for Publications in Molecular Modeling Related to Medicinal Chemistry

Peter Gund,[†,⊥] David C. Barry,[‡,⊥] Jeffrey M. Blaney,[§,⊥] and N. Claude Cohen[∥,⊥]

Merck Sharp & Dohme Research Laboratories, Rahway, New Jersey 07065, ICI Pharmaceuticals Division, Alderley Park, Cheshire, U.K., du Pont de Nemours & Company, Wilmington, Delaware 19898, and Ciba-Geigy Ltd. Pharmaceutical Division, Basel, Switzerland. *Received July 14, 1988*

I. Introduction

a. Molecular modeling studies should be subject to the same RIGOROUS SCIENTIFIC STANDARD as other types of experiments, in particular, reproducibility by independent investigators. This requires providing sufficient documentation of the details of the calculation. Conclusions drawn from a study are traditionally considered tentative until the results are independently confirmed, or at least until experiments designed to test the conclusions are successful.

b. If the molecular modeling study is considered a scientific experiment, then there should be an EXPERIMENTAL SECTION, or METHODS SECTION, which should list the computer program(s) used. If those programs are generally available, then merely the specific options utilized, and/or nonstandard aspects of the calculation, should be included. If any of the programs used are not generally available, then a description of the algorithm(s) programmed, with references, and details of extensions over previously published methods should be included. If a program is a reimplementation of a previously published method, ideally the test systems used to verify performance of the new program should be listed. If a program which was used is not generally available, authors should provide either (1) a statement that the program is available from the author or a distributor (if the author is willing that it be distributed) or (2) a listing of the key algorithm(s), in a common programming language or in pseudocode. A statement of the programming language and computer system used to develop the program, and approximate cpu time of the algorithm for the problem discussed, is also helpful to other scientists attempting to evaluate the reported approach.

c. KEY DATA generated by the study should be presented in the body of the publication or in supplementary material. Key data include coordinates of considered molecular structures, which are preferably provided in computer-readable form (see Section VIIe).

d. Papers primarily presenting NEW ALGORITHMS for performing calculations related to molecular modeling would not be published in medicinal chemical journals unless there was included an application of the method to a problem of medicinal chemical interest. In such cases, recommended format would explain the algorithm briefly in the Experimental or Results sections of the article, with a formal expression of the algorithm (or computer listing) given in an appendix or in the supplementary microfilmed material.

e. Similarly, except for some specialty journals, METHODOLOGY DEVELOPMENT is most publishable if it is immediately applied to a problem of medicinal chemical interest. While it is important for methodology to be developed in a generally useful way, authors are encouraged to focus on the utility of the new method for at least one specific medicinal chemical problem.

f. Since one purpose of a research article is generally to provide sufficient information to allow independent verification of the results, it is appropriate to cite the COMMERCIALLY AVAILABLE SOFTWARE package (and its version number) and specific routines used in the study, unless the technique used is so common, and so likely to give essentially the same result using other available programs, that citation is unnecessary. When such COMMERCIALLY AVAILABLE PROGRAMS are proprietary algorithms, proprietary force fields, etc., there may be difficulty in judging the scientific validity of the results. In such cases the commercializing companies have the obligation to submit for publication articles giving at least general descriptions of algorithms and databases used, with sample calculation results to aid calibration and evaluation. Failing that, the company should publish manuals or other program descriptions containing such details.

g. As previously recommended by the IUPAC with respect to QSAR studies, authors should REFRAIN FROM PUBLISHING PREDICTED ACTIVITIES OF SPECIFIC UNKNOWN STRUCTURES, since that may compromise the patentability of such structures if they are active, so there is reduced incentive to synthesize these materials.

II. Generation of Molecular Structures

The source of the three-dimensional coordinates of each molecular structure used must be given. Modeled structures of any kind must be carefully distinguished from those directly derived from experiment (such as crystallographic structures). Possible sources of structures include the following:

a. An experimental (X-ray, electron diffraction, etc.) structure; the reference or, if unpublished, the experimental details of structure derivation should be given.

b. A partially modeled, partially experimental structure (e.g., NMR data, other spectroscopic data used to constrain the modeling); modeling methods used should be given, as well as a measure of how well the model "explains" the experimental observations.

c. A modified experimental structure; details of the modifications should be given.

d. A structure derived from standard fragments or standard bonds and angles; the source of the standard geometries used should be stated. It is also imperative that the structure be checked for bad nonbonded contacts, and the results reported, for specified van der Waals radii.

e. A structure generated from a two-dimensional representation by some published method (e.g., Approximate Model Builder, Distance Geometry); the method used and its accuracy for such structures (if known) should be indicated. Unpublished methods should be documented as in Section Ib.

[†] Merck Sharp & Dohme Laboratories.
[‡] ICI Pharmaceuticals Division.
[§] du Pont de Nemours & Company.
[∥] Ciba-Geigy Ltd. Pharmaceutical Division.
[⊥] Members of the Working Party on Computer Assisted Molecular Modeling, Provisional Section Committee on Medicinal Chemistry of IUPAC.

0022-2623/88/1831-2230$01.50/0

f. An empirical energy or quantum mechanics energy optimized structure; the basis for the optimization should be given (see Section IIIf).

g. Conformation search, with or without experimental constraints; see below.

h. Macromolecular structures may be generated from secondary structure considerations (α-helix, β-sheet, turns, etc. for proteins; A-, B-, C-helix, etc. for nucleic acids; etc.), by modification of protein crystallographic coordinates and other methods. Energy calculations are more approximate, and more assumptions are required, when dealing with macromolecules; the assumptions and limitations of such macromolecular studies should be clearly indicated. If the study is based on one or more crystal structures, then some discussion of the quality of the original structures is required (with citations to the original literature). When a modeled structure of a protein is published, then it must be carefully differentiated from any protein crystal structures used in the modeling.

i. In the interests of allowing verification or refinement of published results, authors are encouraged to deposit crystallographic results in the public repositories (Brookhaven Protein Data Bank, Cambridge Crystallographic Data Centre). Modeled structures may be deposited with those collections, or provided as supplementary material in the publication, or both (see below).

III. Choosing Molecular Conformations

For most molecules of medicinal interest, more than one conformation is at least theoretically achievable; in fact, different conformations may exist in vacuo, in solution, in the crystal, and for effecting various biological activities. A common fault of publications reporting molecular modeling results on biological systems is that unwarranted or unreported assumptions are made in choosing a conformation for calculations. Conformational space may searched in several ways:

a. If conformations are searched by rigid rotation about key rotatable bonds, then step size (or other means of generating local rotations) and the potential energy function(s) used should be given. If rigid rotation is used without further "all-atom" refinement, then relative energy of conformations should not be overinterpreted.

b. If "torsion driving" such as provided in the MM2 program is used, a similar summary of conformational minima and their energies should be given. Known difficulties associated with this method should be recognized and surmounted if possible.

c. If rigid rotation followed by all-atom geometry optimization is done, a similar listing of conformations, energies, and optimization method should be supplied. The global minimum energy conformation should be labeled if it can be located.

d. If ring structures are involved, ring conformations may be taken from a library, generated systematically by sequential torsion movements, or generated by distance geometry approaches. An indication of the number of ring conformations which were considered should be given; note that methods which depend crucially on choice of parameters (such as ring closure bond tolerance) should not be termed "systematic" conformation search unless it can be shown to have found all possible solutions.

e. Dynamics and Monte Carlo calculations have also been used to generate molecular conformations; but it is hard to assure thorough exploration of conformation space by these methods. Calibration against model systems, where the answer is known, is recommended.

f. If conformation search is done with experimental constraints (NMR, etc.), the constraints used should be indicated.

g. Relative enthalpy of conformations provides a rough guide to probability of occurrence, but is not expected to be rigorous because of the importance of entropy effects, solvation effects, etc. Ideally a Boltzmann distribution of conformations would be computed.

h. When nonenergetic criteria are used during conformation generation (experimental results, presence of a pharmacophoric pattern, fitting to an enzyme active site, etc.), those criteria should be explicitly spelled out.

IV. Availability of Modeled Molecular Coordinates

Whereas crystal structures are considered of enduring value and are available from central repositories (Cambridge Crystallographic Data Centre, Brookhaven Protein Data Bank), modeled three-dimensional structures are generally considered more ephemeral—depending on details of the modeling method used. However, those modeled coordinates are crucial for those who would confirm or extend the published results. Thus, publication of coordinates of carefully modeled structures as supplementary material is encouraged, and several publishers are exploring ways of supplying such material to interested subscribers in computer–readable form. It is also recommended that public repositories such as the Cambridge Crystallographic Data Centre be encouraged to hold such coordinates—clearly labeled, of course, as modeled (not crystallographic) structures. Brookhaven Protein Data Bank, in fact, has already accepted for deposit macromolecular structures derived from modeling.

a. For small molecule crystal structures, the standard file format is that of the Cambridge Crystallographic Data Centre. For modeled small molecules there is no standard format. A recommended general molecule data format is given herein. A more comprehensive Standard Molecular Data (SMD) format has been proposed: Bebak, H.; Buse, C.; Donner, W. T.; Hoever, P.; Jacob, H.; Klaus, H.; Pesch, J.; Roemelt, J.; Schilling, P.; Woost, B.; Zirz, C.; unpublished.

b. For conformational isomers, the conformations may be named descriptively (e.g., synclinal, antiperiplanar, cis-cis-gauche, etc.) or systematically (e.g., consecutive numbers) and identified by Cartesian coordinate lists, or by Cartesian coordinates for one conformer plus lists of key torsion angles (and energy, if computed) of other conformers.

c. For large molecules, information about component parts (residues) is also required. The Brookhaven Protein Databank format is a standard for this area; IUPAC nomenclature for proteins, nucleic acids, and polysaccharides is also well established.

V. Guidelines for Reporting Empirical Force Field Calculations

Empirical force field (molecular mechanics, strain energy) calculations are generally used to optimize molecular geometries and compute conformational energies, intermolecular energies, and various molecular properties. For such calculations the results are no better than the force field parameters for the class of molecules being treated. There are many options and many assumptions in the method, so that results should be interpreted with an appropriate measure of scepticism. Different techniques

and program options will give different results, so the details of studies should be reported with care.

a. The program name (e.g., MM2, BIGSTRN3, AMBER, etc.) and reference should be given, as well as force field used (where more than one can be invoked by the same program) and options (superatoms, all hydrogens, variable dielectric, etc.).

b. Where force field parameters supplied with the program(s) are not sufficient for the molecule(s) at hand, additional parameters must be derived, by analogy to related parameters, by appropriate model calculations, or (preferably) by fitting experimental data for that class of structures. Previously unpublished parameters should be given (with units), and their method of derivation explained, in the publication. If the list is extensive, the parameters may be given as supplementary material. Note that parameters are interdependent, so presenting only a few derived parameters is often insufficient. The parameters must be qualified by giving the equation and units for their use. Nonbonded distance cutoffs and smoothing functions (if used) should be given.

c. If an electrostatic term is used (as it normally is), the source of the partial atomic charges (or bond dipoles, etc.) used should be indicated and their values listed. The model used for the dielectric medium (distance-dependent or constant dielectric, etc.) should be given, as well as any solvation model used.

d. Optimization method used should be indicated, e.g., steepest descent, conjugate gradient, Newton-Raphson, Cartesian vs internal coordinate minimization, convergence criteria. Was the stationary point proved to be a true minimum using second derivative techiques? Compare this minimum to the global minimum, if it has been located. If the software used does not provide adequate information on gradients and the nature of the stationary point, then further tests must be performed before the structure may be described as locally "optimal".

e. Dynamics calculations should include description of starting coordinates, same description of force field model as listed above, integration method (Verlet, Gear, etc.), method of "heating", time steps, equilibration period, length of simulation, constant pressure or constant volume, periodic boundary conditions, effective temperature of calculation, and other pertinent information. Unusual assumptions or methods should be indicated. Results of dynamics runs may be presented as plots of RMS coordinate deviations per frame, as illustrations of typical molecular geometries, etc.

f. Monte Carlo calculations should include technique used, number of trial structures generated, and some statistical measure of the precision of the result.

g. Calculations should be indicated as "in vacuo" or in solvent. Calculations of solvation energy should include a description of the solvation model (and reference if appropriate), type of calculation performed, and robustness of the result.

h. When available from dynamics or Monte Carlo calculations, computed free energy should be indicated as well as enthalpy. However, since the statistical reliability of entropy contributions may be low, such results should not be overinterpreted.

i. The International System Units require energy results to be reported in joules, or ergs (1 joule = 10^7 ergs). However, discussion of energy differences in kcal/mol is permissible (1 kcal = 4.184 kJ).

j. Force field calculations on macromolecules often entail additional approximations, which should be explicitly listed unless they are standard for the program utilized.

VI. Guidelines for Reporting Quantum Mechanical Calculations

Quantum mechanical calculations are generally ab initio or semiempirical in nature. If sufficient computer power is available, these programs may be used to compute optimized geometries, partial charges, frontier orbitals, and reaction pathways (among other properties). In the METHODS SECTION, the following should be described:

a. The program name (e.g., GAUSS82, MNDO, IBMOL) and a reference. If the program has been modified, the modifications must be explained or referenced.

b. Level of SCF (RHF, UHF, ROHF, etc.) used.

c. Basis set used (e.g., STO-3G, 6-31G** or other notations).

d. Type of correlation correction (MP2, MP3, CI, MCSCF, etc.).

e. Electronic state (singlet, doublet, etc.).

f. Symmetry of molecule and symmetry constraints used in the calculations.

g. Method of characterizing the stationary point in geometry optimizations and optimization procedure used.

h. For semiempirical methods, calibration of the method against ab initio results for relevant model systems is helpful and should be reported if carried out.

i. Although quantum mechanical results are output from most computer programs in units of atomic units (au), results for publication should be converted to International System Units. Conversion factors are as follows:

$$\begin{aligned} 1\ \text{au} &= 27.21161\ \text{eV} \\ &= 627.5098\ \text{kcal/mol} \\ &= 2.625\,501\ \text{kJ/mol} \end{aligned}$$

$$\text{(Bohr radius) } 1\ a_0 = 5.292 \times 10^{-11}\ \text{m}$$

e. Partial atomic charges are recommended to be given as net charge in units of absolute electron charge (i.e., positive represents net excess nuclear charge over average electric charge in the environment of the atom). The method of allocating charge density (among atoms, bonds, etc.) should be carefully described. Electrostatic potential is represented as the energy of a point probe at some position in space; unless otherwise specified, the probe is a unit positive charge.

f. Dipole moments of molecules or fragments are expressed in debye units. Bond dipole moments may be expressed in units of electron-angstrom. The sign convention of dipole vector quantities has not been consistent in the literature; the preferred convention orients the positively signed dipole vector TOWARD the region of predominant positive charge.

VII. General Recommendations

a. Recommended Nomenclature. This field has little unique nomenclature not shared with the disciplines of medicinal, physical, organic, and biological chemistry. The standard IUPAC nomenclature for these fields applies.

b. Accepted Abbreviations. Standard chemical and physical abbreviations are used.

c. Presentation of Molecular Views

i. Stick figures: no reference necessary. If in color, bonds are conventionally split, with each half colored according to atom type of attached atom using CPK colors (nitrogen blue, oxygen red, hydrogen white, sulfur yellow, etc.) except for carbon, which cannot be represented in the conventional black color unless the background is colored. On a black background, recommended colors for carbon are gray or green. Other colors may be used for specific

reasons, but the presenter should be aware that use of dramatic but nonstandard colors may favor illustration at the expense of comprehension.

ii. Stereoscopic pair view: specify relaxed-eye or crossed-eye presentation. If complexity of the figure allows, the centers of the side-by-side stereo images should be no farther than 2 in. (5 cm) apart, to facilitate viewing without stereo viewers. Stereoscopic views should be simplified if possible; complexity interferes with achieving stereopsis. Stereoscopic "depth" should not be overly exaggerated.

iii. Ball-and-stick: specify source of program if applicable (e.g., PLUTO, ORTEP) and reference. If in color, balls take CPK colors except, again, carbon may be gray or green, or black (if background is highlighted). Stick bonds may be white or gray or transparent.

iv. Spacefilling: specify source of program if applicable (e.g., SPACFIL, CPK). Heteroatoms are conventionally represented by fill patterns in black-and-white representations, or by CPK colors as in iii.

v. Surface display: specify source of program used and reference. Parameters used (atomic and probe radii) should be explicitly identified. Surfaces should be identified as follows (see Richard, F. M. *Annu. Rev. Biochem. Bioeng.* **1977**, *6*, 151): VAN DER WAALS SURFACE (envelope of atom spheres having van der Waals atomic radii); ACCESSIBLE SURFACE (envelope of atom spheres having van der Waals plus probe radii); MOLECULAR SURFACE (consisting of "contact surface" patches coinciding with van der Waals surface, and "reentrant surface" patches). Reported areas or volumes should clearly specify any surface definition involved. If coloring is used, the following coloring guidelines should be used unless there are logical reasons for choosing a different scheme (in which case the scheme used should be explained). Electrostatic charge coloring is conventionally blue for strongly positively charged (as nitrogen), light blue for weakly positively charged, pink for weakly negatively charged, red for strongly negatively charged (as oxygen); green for neutral or hydrophobic. Electrostatic potential coloring may follow similar conventions. Electric field arrows conventionally point TOWARD negatively charged regions, following the electric potential gradient. Solvent contact surface may be colored by adjacent atom type (CPK colors), by region, or by other coloring criteria which should be clearly defined.

vi. Superimposed structures: identify by solid and dashed lines, solid and open lines in ORTEP, color, or some other differentiating method. The superposition technique used and some measure of the goodness of superposition (e.g., rms deviation of specified superposed atoms) should be given.

vii. Macromolecular display: designate as Cα, etc.; reference the display program used. Color may be used to designate secondary structure, residue types, subunits, or other information; the coloring convention used should be clearly designated. Indicate chain direction with arrows or residue numbers. Hidden-line rendering or ribbon plots greatly clarify chain overlaps.

viii. Structure designation: molecule should be clearly labeled as derived from crystallography (with reference or crystallographic details), or from molecular modeling, with method of structure generation explained.

d. Use of Supplementary Material

i. However useful molecular coordinates, etc., are for verifying and extending reported results, they take up much space in journals, and are most useful in machine-readable form. These data and related data (discussed below) are profitably published as supplementary material, and journal publishers are encouraged to supply such supplementary material in computer-readable form when appropriate.

ii. Submission of modeled structure coordinates to public repositories also would encourage use of such structures for structure survey type studies such as have been so usefully performed with sets of crystal structures. The public repositories are encouraged to collect such modeled structure coordinates, with appropriate documentation of their modeled nature.

e. Recommended Formats for Supplementary or Archived Molecular Data (a general molecule file format is appended)

i. Small Molecules

(a) Molecule name, CAS Registry Number, submitter, reference.

(b) Source, details of structure creation (modeling, optimization, fitting to pharmacophore, etc.).

(c) For crystal structure, space group information.

(d) For each atom: atom identifier, atom type, atomic charge (optional), *xyz* coordinates (in angstroms, to three decimal places).

(e) Hydrogen atoms are stored if their positions are important; otherwise they are optional.

(f) For each bond: bond identifier, atoms at each end of bond, bond type (single, double, triple, delocalized, partial, dative).

(g) Additional information, such as atomic orbital coefficients; molecular surfaces; symmetry elements.

(h) Derivative information (atoms connected to each atom, color codes, etc.) are easily computed from the above data and need not be stored; however, lists of internal coordinates (bond lengths, bond angles, torsion angles) are quite useful and may be optionally supplied.

(i) Cambridge Crystallographic Data Base or other common formats may be utilized.

ii. Macromolecules: Brookhaven Protein Data Bank format is recommended; however, bonding information between atoms is important and should be recorded, although this is optional for the Brookhaven format. Nonstandard residues should be illustrated with whatever atomic notation has been assigned.

VIII. A General Molecule File Format

For X-ray determinations, either the Cambridge (small molecule) or Brookhaven (macromolecule) formats should be used, or the Standard Crystallographic File Structure: Brown, I. D. *Acta Crystallogr.* **1983**, *A39*, 216.

For molecular modeling studies, the following file format is recommended for reporting results:

1. Title Record (one only)

Cols 1–2	(A2)	"TI" identifies this as a Title Record.
4–80	(A77)	A descriptive title.

2. Comment Records (zero or more)

Cols 1–2	(A2)	"CO" identifies this as a Comment Record.
4–80	(A77)	Descriptive comments. Should indicate the origin of the structure.

3. Atom Records (one for each atom in the molecule)

Required fields for each atom:

Cols 1–2	(A2)	"AT" identifies this as an Atom Record.
4–7	(I4)	Atom Identifier Number, e.g., "1", "2" Note: must be unique, need not be sequential or contiguous. This number is used in the Bond Records; see below.
9–17	(F9.4)	*x* coordinate, angstroms
19–27	(F9.4)	*y* coordinate

29–37	(F9.4)	*z* coordinate
39–40	(A2)	Atomic Symbol (e.g., "C", "Ca") No distinction is made between upper- and lower-case letters
42–43	(A2)	Formal Charge (e.g, "+", "+2", "–", "–2", ... for charges; ↶ for a radical)

Optional fields for each atom:

45–51	(F7.4)	Calculated Partial Atomic Charge.
53–56	(A4)	Atom type code for the empirical force field program used (e.g., MM2, AMBER, ...)
58–61	(I4)	Residue Sequence Number. Residues occur in order of their residue sequence numbers, which should increase starting from the N-terminal residue in the case of proteins and the 5′-terminal for nucleic acids. Atom Records for a given Residue Sequence Number should be adjacent. Residue Sequence Numbers for a given residue must be unique. The numbers need not be contiguous, e.g., "1005" can immediately follow "1001".
62	(A1)	Residue Insertion Code Letter (e.g., "A", "B"). Allows residues to be inserted without disturbing the residue numbering of the parent structure.
64–67	(A4)	Residue Name. Standard residue names as described for the Brookhaven Data Bank should be used, where possible.
69–72	(A4)	Atom Name. Should adhere to atom name convention defined in Appendix B of Protein Data Bank "Atomic Coordinate Entry Format Description", where possible.
74–77	(A4)	Atom Label. An atom label of your choosing, for display annotation purposes.

4. Bond Records (one for each "origin atom" in the molecule)

Cols 1–2	(A2)	"BD" identifies this as a Bond Record.
4–7	(I4)	Atom Identifier Number of Origin Atom

Atom Identifier Number of Connected Atoms (atoms connected to Origin Atom):

9–12	(I4)	First connected atom, etc.
14–17	(I4)	
19–22	(I4)	
24–27	(I4)	
29–32	(I4)	
34–37	(I4)	
39–42	(I4)	
44–47	(I4)	

Note: Atom Identifier Numbers are identical with those in cols 4–7 of the Atom Records, above. For each Bond Record, a single bond is defined between the Origin Atom (cols 4–7) and each of the following Connected Atoms. A double bond is indicated by including the atom sequence number for the connected atom twice. Similarly, a triple bond is indicated by including the atom sequence number for the connected atom three times.

5. Supplementary Records (Optional) The following information may be useful when performing empirical force field calculations or for visual display of molecular information.

A. Aromatic Ring Records

Cols 1–2	(A2)	"AR" identifies this as an Aromatic Ring Record.

List of Atom Identifier Numbers for Atoms of type "Aromatic":

4–7	(I4)	First aromatic atom.
9–12	(I4)	Second aromatic atom, etc.
14–17	(I4)	
19–22	(I4)	
24–27	(I4)	
29–32	(I4)	
34–37	(I4)	
39–42	(I4)	
44–47	(I4)	
49–52	(I4)	
54–57	(I4)	
59–62	(I4)	
64–67	(I4)	
69–72	(I4)	

Note: The aromatic bonds are defined by the order of atoms tracing the ring, one aromatic ring per record. The ring closure bond is taken to join the first and last atom in the ring. Fused rings will have the common atoms repeated in the records describing both rings.

B. Hydrogen Bond Records (one per hydrogen bond)

Cols 1–2	(A2)	"HB" identifies this record as a Hydrogen Bond Record.
4—7	(I4)	Atom Identifier Number of "donor" atom.
9–13	(I4)	Atom Identifier Number of "acceptor" atom.

Note: In the case of an all-atom force field calculation, the "donor" atom should be the hydrogen atom of the hydrogen bond. In the case of a "united atom" force field calculation in which no hydrogen atoms are explicitly included, the "donor" atom should be the atom to which the hydrogen atom of the hydrogen bond would be singly bonded.

C. Other types of Data

Most other types of information may be accommodated similarly, i.e., in one 80-column record, with a unique 2-character "key" in cols 1–2, and the data keyed to an atom, two atoms (bond/nonbonded data), three atoms (angle data), or four atoms (torsion or out-of-plane data).

6. Termination Record (one per molecular species in the file) Separates individual molecules in the file.

Cols 1–2	(A2)	"$$" identifies this record as Termination Record.

Acknowledgment. Thanks are due to B. L. Bush, D. J. Underwood, and the many other scientists who read and commented on the manuscript at various stages. J. D. Andose provided the recommended General Molecule File Format.