

INTERNATIONAL UNION OF PURE AND APPLIED CHEMISTRY

ORGANIC CHEMISTRY DIVISION
COMMISSION ON PHYSICAL ORGANIC CHEMISTRY*
WORKING PARTY ON QUANTUM CHEMICAL AND COMPUTATIONAL DATA†

GUIDELINES FOR THE PRESENTATION OF QUANTUM MECHANICAL COMPUTATIONAL DATA IN ORGANIC CHEMISTRY

(Technical Report)

Prepared for publication by

T. A. ALBRIGHT¹ and E. A. HALEVI²

¹Department of Chemistry, University of Houston, Houston, Texas 77004, USA

²Department of Chemistry, Technion-Israel Institute of Technology, Haifa 32000, Israel

*Membership of the Commission during the period (1985–91) the report was prepared was as follows:

Chairman: 1985–91 P. Müller (Switzerland); *Secretary:* 1985–87 R. A. Y. Jones (UK); 1987–91 M. P. Doyle (USA); *Titular Members:* P. N. I. Ahlberg (Sweden, 1987–89); E. M. Arnett (USA, 1985–87); E. A. Halevi (Israel, 1985–89); J. M. McBride (USA, 1987–91); O. M. Nefedov (USSR, 1985–91); M. Ōki (Japan, 1985–91); J. Shorter (UK, 1989–91); *Associate Members:* T. A. Albright (USA, 1987–91); M. P. Doyle (USA, 1985–87); W. Drenth (Netherlands, 1987–91); R. D. Guthrie (USA, 1985–87); E. A. Halevi (Israel, 1989–91); R. A. Y. Jones (UK, 1987–89); J. S. Littler (UK, 1985–87); A. K. Maltsev (USSR, 1985–87); J. March (USA, 1985–87); V. I. Minkin (USSR, 1987–91); K. Schwetlick (GDR, 1985–89); Y. Takeuchi (Japan, 1987–91); P. Van Brandt (Belgium, 1985–91); J. R. Zdysiewicz (Australia, 1989–91); *National Representatives:* P. N. I. Ahlberg (Sweden, 1985–87); T. A. Albright (USA, 1986–87); E. Baciocchi (Italy, 1989–91); A. T. Balaban (Romania, 1985–87); M. V. Bhatt (India, 1989–91); W. Drenth (Netherlands, 1985–87); X. Jiang (Chinese Chemical Society, 1985–91); P. Laszlo (Belgium, 1985–91); D. J. McLennan (New Zealand, 1985–91); M. Nógrádi (Hungary, 1985–87); Z. Rappoport (Israel, 1988–91); R. Sabbah (France, 1989–91); B. E. Smart (USA, 1987–91); J. Suh (Republic of Korea, 1989–91); O. Tarhan (Turkey, 1988–91); M. Tisler (Yugoslavia, 1986–91); J. Závada (Czechoslovakia, 1985–89); J. R. Zdysiewicz (Australia, 1986–89).

†Membership of the Working Party: 1985–91

Joint Chairmen: T. A. Albright (USA) & E. A. Halevi (Israel); *Members:* I. Apeloig (Israel); F. Bernardi (Italy); M. Gordon (USA); J. Hay (USA); P. C. Hiberty (France); R. Janoschek (Austria); V. I. Minkin (USSR); K. Morokuma (Japan); L. Radom (Australia); M. A. Robb (UK); P. Schleyer (FRG); W. Thiel (FRG); R. Zahradnik (Czechoslovakia); J. R. Zdysiewicz (Australia).

Republication of this report is permitted without the need for formal IUPAC permission on condition that an acknowledgement, with full reference together with IUPAC copyright symbol (© 1992 IUPAC), is printed. Publication of a translation into another language is subject to the additional condition of prior approval from the relevant IUPAC National Adhering Organization.

Guidelines for the presentation of quantum mechanical computational data in organic chemistry (Technical Report)

Abstract Guidelines are offered concerning information that should be reported in the primary chemical literature relating to the techniques and results of quantum mechanical calculations. The goal of these Guidelines is to outline the information required to allow an informed reader to reproduce the results and make an objective evaluation of the quality of the work.

I. INTRODUCTION

The primary objective of this document is to offer a set of guidelines concerning what information should be reported in the primary chemical literature relating to the techniques and results of quantum mechanical calculations. This information should be comprehensive to the point that an informed reader can reproduce the results and make an objective evaluation of the quality of the work.

II. GENERAL GUIDELINES

Specification of non-default parameters is generally desirable and is described for common methods in the following sections. The publication of results computed with a standard program should include its name, version number and source (e.g., Quantum Chemistry Program Exchange number), the date of its latest revision (where appropriate) and a reference to the most definitive publication in which that version is described. Frequently, the program allows a freedom of choice among basis sets, convergence criteria, etc. It is suggested that authors specify these details in a section of the paper entitled "Computational Details." When presenting results, extensive tabulations of eigenfunctions, eigenvalues, net atomic charges, orbital populations, and contour plots of orbitals should be included only if they are used to support a specific argument in the paper. Quantum chemistry is replete with acronyms. Care should be exercised with definitions when less common acronyms are used. Symbols for physical quantities should be a single letter which may be modified by subscripts, see *Pure and Applied Chem.* 1979, 51, 1-41 for details and examples.

Structural formulae should be presented in which the atoms are unambiguously numbered unless a standard numbering system is used; in the latter case, it should be cited. Whenever

the symmetry of an orbital, electronic state or reaction path is specified by a group theoretical notation, the axis convention adopted should be stated.

III. SEMI-EMPIRICAL METHODS

A. Extended Hückel

For the extended Hückel method authors should give at least a literature reference to the basis set used for each atom, and an explicit statement should be given as to whether single zeta values were used throughout or which selected atoms used a double zeta basis. The source of the valence state ionization potentials (H_{ii}) should be given. If these were derived from an iterative extended Hückel (Self-consistent Charge) calculation, then the compound, associated geometry and a literature citation to where the A, B and C parameters have been taken from needs to be specified. Authors should state how the resonance (H_{ij}) integrals have been computed, i.e., Wolfsberg-Helmholz, weighted H_{ij} , Balhausen-Gray, etc. Geometrical details should be presented in the "Computational Details" section.

B. CNDO, INDO, MINDO, MNDO, AM1, PM3

It is essential that there be a reference to the computer program used. In most cases this will be a standard QCPE or other public domain program; therefore, additional specifications concerning SCF convergence and other options need not be given as long as the default values have been used. In cases where the parameter set for a given element or any integral approximation has been altered from the default values, a literature reference or an explicit listing of these changes must be given. For open-shell systems, the SCF method (UHF or half-electron RHF) must be defined.

The calculated total energies (to 10^{-5} hartree) or heats of formation (to 0.5kJ/mol) should be reported for each molecule studied. Details associated with reporting the results of CI calculations and geometry optimizations are given in sections V and VI, respectively.

IV. NONEMPIRICAL METHODS

A. All-electron *ab initio*

The program used should be cited and any use of non-default values, e.g. convergence criteria, integral cutoff threshold, etc., should normally be noted. Particular care is required for the specification of the basis sets used for all atoms in the molecule(s). A literature reference must be given for a standard basis set. If this is modified and/or augmented in any fashion,

then the manner in which this was done, along with a listing of exponents and contraction coefficients, is required. The functional form of the basis functions, e.g., Gaussian Type Orbitals (GTO), Slater Type Orbitals (STO), Floating Spherical Gaussian Orbitals (FSGO), should be indicated in the case of non-standard basis sets. Particularly if the results from basis sets at different levels are discussed in the text, then suitable abbreviations are encouraged, e.g., STO-3G, 3-21G, MIDI, MINI, etc. However, it is strongly recommended that, unless standard, widely-used basis sets are used throughout, the actual construction of the basis sets for each atom should be specified with the notation (abc...,ghi...,lm...,o...). Here, a is the number of primitives (usually Gaussian) in the first contracted s function, b is the number of primitives in the second contracted s function, etc., g is the number of primitives in the first contracted p function, l is the number of primitives in the first contracted d function, and o is the number of primitives in the first contracted f function. Thus, a 6-31+G* basis for carbon would be designated (6311,311,1). An indication should also be given as to whether 5 "pure" or 6 cartesian d functions (7 pure or 10 cartesian f functions) have been used. The SCF procedure should be specified, i.e., restricted, closed-shell Hartree-Fock (RHF), restricted, open-shell Hartree-Fock (ROHF) or unrestricted, open-shell Hartree-Fock (UHF). In the last case, the expectation value of the spin-squared operator, $\langle S^2 \rangle$, should be given for the molecules studied to provide an indication of the extent of spin contamination. If the orbitals were allowed to become complex, this should be noted. Authors are encouraged to provide details about how SCF convergence was achieved for problematic cases. The total energies for, at least, selected molecules must be given (other conformations, etc. may be quoted in terms of relative energies) to 10^{-5} Hartree.

B. Pseudopotential/effective core potential

In addition to relevant aspects arising from section IV A, the form of the potential must be indicated, e.g., numerical, analytical fit, core projections. References should be given to the potentials used or, if they have not been published elsewhere, the parameters associated with the potentials should be tabulated. Finally, it should be stated as to whether or not relativistic corrections have been included.

C. Valence bond methods

In addition to relevant aspects arising from section IV A, it should be indicated which orbitals are constrained to be orthogonal to each other and how the orbitals used to construct the determinants are defined, e.g., atomic orbitals of free atoms, SCF orbitals of molecular fragments, optimized so as to minimize the energy of the total valence bond wavefunction, etc. All valence bond configurations involved in the total wavefunction or the way in which they were generated should be specified.

D. Density functional approaches

In addition to relevant comments in section IV A, the nature of the exchange, correlation potential and approach, i.e., scattered wave or basis set, should be specified. The sphere size and value of α (if applicable), numerical integration scheme and total energy (if applicable) are required.

V. POST HARTREE-FOCK TECHNIQUES

There is much flexibility associated with post Hartree-Fock calculations. Therefore, in general, authors should exercise care in describing the particular method and associated details. The most common techniques are covered below. In all cases, the total energy, including the correlation contribution, should be reported (to 10^{-5} Hartree).

A. CI calculations

It is important to specify which orbitals are used as the basis for the CI calculations e.g. orbitals from an RHF, generalized valence bond (GVB), natural orbitals, multi-configurational self-consistent-field (MCSCF), etc. calculation. The reference configurations as well as the method of selecting configurations must be clearly specified. Only a few common notations (e.g., CISD, full CI, full pi-space CI) are universal. Other schemes should be referenced (e.g., POL-CI, first order CI, etc.) or details given as to which occupied and virtual orbitals are active, the excitation level, the maximum allowed number of unpaired electrons, and the total number of configurations. Any extrapolation (higher order correction) procedure must be given, as well as any scheme for reducing the number of configurations (the criteria used as thresholds). Extensive tabulation of CI coefficients, in general, is not encouraged in the main body of the paper but may be given as supplementary material.

B. GVB, MCSCF, Complete-active-space self-consistent-field (CASSCF) calculations

The source of the starting orbitals along with the convergence criteria and the starting configuration(s) must be specified. Specification must be made as to which orbitals are always doubly occupied, which are always empty, and which are active, i.e., have variable occupation. In addition, some rationale as to why these were selected should normally be given. Any subsequent CI procedure should follow the guidelines given in section V A.

C. Many-body perturbation techniques

The source of the orbitals should be cited and any of them, in particular core orbitals, which have been excluded from the perturbation calculation must be clearly documented. The order of perturbation theory used and, in the case of partial treatments, which excitations have been included must also be given; e.g., MP4(SDQ) indicates a fourth order Möller-Plesset treatment with single, double, and quadruple excitations, but not triple excitations. The notation MP_n would indicate a full *n*th order treatment. Other perturbation treatments are less common and some description of the method should be given.

VI. GEOMETRY OPTIMIZATIONS AND FOLLOWING REACTION PATHS

If a standard program is used, the specific optimization technique (Fletcher-Powell, Murtaugh-Sargent, Schlegel, etc.) should be mentioned and the appropriate reference should be given. Any modification of the parameters associated with a standard program must be stated. If a non-standard program is used, then the convergence criteria should be given along with a description of the method.

All geometric constraints, e.g., bond lengths and/or angles arbitrarily held at fixed values, or restriction to a particular molecular symmetry, need to be given. Certain geometric features, such as cyclic bonding, may result in the overspecification of internal coordinates and it should be clearly stated how such redundancies are eliminated. In problematic instances, particularly where several stationary points on a potential energy surface have been found, some indication should be given for how these were found. The phrase "the geometry was fully optimized subject to a (specified) symmetry constraint" should be reserved for computations in which no constraints whatsoever were imposed, other than those imposed by the specified symmetry.

Optimized bond lengths (to 0.1 pm) and angles (to 0.1°) should be included in the manuscript (either in figures or in tabular form). In situations where this is cumbersome, the values of the internal coordinates should be provided as supplementary material, or a statement should be given that this information will be supplied by request to the author.

Particular attention should be paid to the description of stationary points. It should be noted whether the Hessian matrix is an approximate one obtained by the updating scheme of the optimization algorithm or the stationary point has been tested at the optimized geometry by explicit computation of the frequencies (or Hessian matrix) to determine if it is a minimum (all real frequencies, positive eigenvalues), a first order saddle point (one imaginary frequency, negative eigenvalue) or a higher order saddle point (more than one imaginary frequency, negative eigenvalue). It should be noted whether the frequencies have been calculated using analytical first and second derivatives, analytical first and numerical second derivatives, or numerical first and second derivatives. It should be indicated whether or not the harmonic frequencies have been corrected for anharmonicity. Frequencies should normally be quoted to 1 cm^{-1} and their symmetry assignments given. It must be specified whether the reported frequencies and/or corresponding zero-point vibrational energies are scaled and, if so, by what scaling factor. Whenever spurious frequencies appear for translations and rotations as may occur, for example, as a result of optimization with an excessive number of independent coordinates, their magnitude should be given explicitly and the implication with regard to the accuracy of the true frequencies discussed. When vibrational intensities are reported, the method used and precision with which they were computed should be given.

If a schematic potential energy surface (consisting of a series of linked stationary points) is reported, it should be noted whether or not the connections between the stationary points have been explicitly confirmed. In cases where a detailed reaction pathway has been determined, the nature of the procedure used, e.g., intrinsic reaction coordinate (IRC) in mass-weighted cartesian coordinates, etc., and the size of the interval between points along the path, whether this is fixed or variable, and in the latter case the criteria used for varying it, must be specified.