ACKNOWLEDGMENT

We thank C. J. Bridgeman, D. A. Faulkner, and S. R. Morgan for their assistance.

REFERENCES AND NOTES

(1) J. E. Ash and E. Hyde, Eds., "Chemical Information Systems", Chichester, Ellis Horwood, 1975.
(2) L. A. Evans, M. F. Lynch, and P. Willett, "Structural Search Codes for Online Compound Registration", *J. Chem. Inf. Comput. Sci.*, **18**, 146-149 (1978).
(3) J. H. R. Bragg, M. F. Lynch, and W. G. Town, "The Use of Molecular Formula Distribution Statistics in the Design of Chemical Structure Registry Systems", *J. Chem. Doc.*, **10**, 125-128 (1970).
(4) W. J. Howe and T. R. Hagadone, "Progress towards an Online Chemical and Biological Information System at the Upjohn Company" in "Retrieval of Medicinal Chemical Information", *A.C.S. Symp. Ser.*, **84**, 107-131 (1978).
(5) R. G. Freeland, S. A. Funk, L. J. O'Korn, and G. A. Wilson, "The Chemical Abstracts Service Registry System. II. Augmented Connectivity Molecular Formula", *J. Chem. Inf. Comput. Sci.*, **19**, 94-98 (1979).
(6) M. Randic, "On Characterization of Molecular Branching", *J. Am. Chem. Soc.*, **97**, 6609-6615 (1975).
(7) H. L. Morgan, "The Generation of a Unique Machine Description for Chemical Structures—a Technique Developed at Chemical Abstracts Service", *J. Chem. Doc.*, **5**, 107-113 (1965).
(8) J. M. Dalton, "Evaluation of the Application of a Topological Index to Compound Registration at Pfizer Central Research", M.Sc. thesis, University of Sheffield, 1979.
(9) L. B. Kier and L. H. Hall, "Molecular Connectivity in Chemistry and Drug Research", Academic Press, New York, 1976.

Quantum Chemistry Literature Data Base

Y. OSAMURA[†]

Department of Chemistry, Osaka City University, Sumiyoshi-ku, Osaka 558, Japan

S. YAMABE

Faculty of Education, Nara University of Education, Nara 630, Japan

F. HIROTA

Department of Chemistry, Shizuoka University, Ohya, Shizuoka 442, Japan

H. HOSOYA

Department of Chemistry, Ochanomizu University, Bunkyo-ku, Tokyo 112, Japan

S. IWATA

Institute of Physical and Chemical Research, Wako, Saitama 351, Japan

H. KASHIWAGI and K. MOROKUMA

Institute for Molecular Science, Myodaiji, Okazaki 444, Japan

M. TOGASI,[‡] S. OBARA,[§] K. TANAKA, and K. OHNO*

Department of Chemistry, Hokkaido University, Sapporo 060, Japan

Received August 1, 1980

The quantum chemistry literature data base (QCLDB) contains literature concerning ab initio computations of atomic and molecular electronic structures. Approximately 2000 literature references published from Jan 1977 to June 1979 have been collected from 19 internationally well-known core journals. Keys to references are computational methods, basis sets, and calculated properties and printout is by author and compound indexes.

INTRODUCTION

With many ab initio computations of atomic and molecular electronic structure appearing in many journals, chemists, experimental and theoretical, who would like to know and utilize the results of such calculations often have a difficult time in finding proper references. Of late, computer-based information retrieval systems are becoming available, but it is not easy to satisfy both specialists and nonspecialists who want to make an overall survey as well as obtain some specific information on the available calculations for compounds of interest. Richards' famous book series[1-3] is quite useful in that each molecule is treated separately. Molecules are ordered according to their size, and by looking up a particular molecule, one can find all ab initio calculations published up to a certain date as well as additional information on the geometry, type of calculation, energy, and computed properties (except in the last book of the series[3]). In spite of these merits, it takes at least a year for the bibliography to be published in a book form, and therefore current references cannot be covered.

In this project of the Quantum Chemistry Literature Data Base (QCLDB), a major emphasis is placed on current awareness as well as exhaustiveness. The potential for computer searching is desirable. Therefore, the information in QCLDB should be created in a computer-readable form.

[†] Department of Chemistry, University of California, Berkeley, CA 94720.
[‡] Department of Physics.
[§] Institute for Molecular Science, Myodaiji, Okazaki 444, Japan

0095-2338/81/1621-0086$01.25/0

Literature citations contained in QCLDB are restricted to those concerning ab initio computations of atomic and molecular electronic structure. The first reason for this restriction is simply practical. If semiempirical MO calculations were included, the number of references would be too large for our present group to handle. The second results from the high quality and uniqueness of ab initio calculations. For some physical properties, ab initio calculations have become very reliable, worth comparing directly with experiments. Herzberg's recent book[4] on the structure of diatomic molecules lists results of ab initio calculations as well as experimental results. Even for those properties for which it is harder to reliably calculate theoretically, methods of systematically improving the calculation are available.

CHARACTER AND CONTENT OF QCLDB

Our guiding principles are as follows:

(1) QCLDB is for everybody. In other words, it should be designed not only to respond to the needs of specialists in quantum chemistry calculations but also to give brief and standarized information to nonspecialists.

(2) To achieve this high standard, the data collection should be carried out by quantum chemists themselves. Data should be drawn directly from the content of a paper, not simply from the abstract or the title.

(3) QCLDB should be sustained for a fairly long period and updated regularly and frequently. Therefore, the collection and the maintenance of the system must be as simple as possible.

On the basis of these criteria, we have decided to include the following items in QCLDB.

(1) Author(s)
(2) Journal name, volume, page, and year
(3) Compounds
(4) Methods of calculation
(5) Basis sets
(6) Calculated properties
(7) Any comment the abstractor cares to add

A reference concerned only with the theory without any computation would be included if it is judged to have some impact on molecular ab initio calculations. An atomic calculation is also treated similarly. In the next section, the rules and formats of generating the input data from journals are stated.

Nineteen internationally well-known journals, listed in Table I, were selected as source journals. The reason for selecting only these journals is that they contain the bulk of publications on ab initio calculations and because, for the present manpower of the group, it is impossible to include more journals. We hope that, as the situation improves, we would be able to add more journals to the roster.

Table I. Journal and CODEN Names for QCLDB

journal name	CODEN
Bull. Chem. Soc. Jpn.	BCSJA8
Can. J. Chem.	CJCHAG
Chem. Phys.	CMPHC2
Chem. Phys. Lett.	CHPLBC
Int. J. Quantum Chem.	IJQCB2
Int. J. Quantum Chem. Symp.[a]	IJQSAF
Inorg. Chem.	INOCAJ
J. Am. Chem. Soc.	JACSAT
J. Chem. Phys.	JCPSA6
J. Chem. Soc., Faraday Trans.	JCFTBS
J. Mol. Spectrosc.	JMOSA3
J. Mol. Struct.	JMOSB4
J. Phys. Chem.[a]	JPCHAX
Mol. Phys.	MOPHAM
Nouv. J. Chim.[a]	NJCHD4
Phys. Rev.	PLRAAN
Tetrahedron	TETRAB
Tetrahedron Lett.[a]	TELEAY
Theor. Chim. Acta	TCHAAM

[a] These journals were added in 1979.

RULES OF THE FORMAT OF THE QCLDB INPUT DATA

In the card-image input data, items 1–7 are sequentially arranged. Each item is terminated by a slash (/). A distinction between upper and the lower case letters is not made, for easy input from a key punch and easy output on a line printer.

(1) Author(s). The last names of three authors are listed. For papers with more than three authors, the three listed are followed by one or two commas (,). The first of the three has to be the first author of the paper, but the choice of the other two listed is made by the abstractor.

(2) Journal. Name (CODEN, six letters), volume no., page, and year (last two digits, e.g., 79), in this order with a comma in between, are listed. CODEN may be replaced by some familiar abbreviation name such as JCP for JCPSA6. A processing program QCHECK introduced in the next section recovers CODEN internally.

(3) Compounds. At most, ten compounds can be listed. If there are more than ten compounds, ten are chosen and listed with one or two following commas. Carbon atoms should come first, hydrogen atoms next, and then other atoms in alphabetical order. Although in the output each molecule is expressed in this format, in the input data any standard formula (without parentheses) may be used. For acetic acid, for example, CH3COOH is allowed as well as C2H4O2. QCHECK converts CH3COOH to C2H4O2. Elements expressed by more than one letter must be followed by a period (e.g., CO. for cobalt and HE. for helium). Charged species (cation and anion) are signified with the charge number. For one atom and one charge in the compound, the 1 may be neglected (e.g., not H4N1+1 but H4N+ for NH_4^+). Thus, chemical substances are coded in the style of the formula index of *Chemical Abstracts* except for the expression of two-letter elements and the assignment of charge. For molecular complexes such as the hydrogen-bonded system, both the supermolecule and its constituent monomers are listed. For polymers and crystals (e.g., NaCl crystal), only the smallest unit (CL.NA.) is given. In item 7 (Comment), it is distinguished from the small molecule (diatomic molecule, sodium chloride). An example is shown below.

molecular formula	output format
$C_5H_{12}NO_3$	C5H12NO3
$Al_2C_2H_4^{2-}$	C2H4Al.2-2
$CoCuO_{36}Zn_2H_{12}^{11+}$	H12CO.CU.O36ZN.2+11

(4) Method of Calculation.

(5) Basis Set.

(6) Physical Property.

Code numbers shown in Table II are used for items 4–6. As many codes as applicable can be listed, e.g.,

/135/290/1/

(4) (5) (6)

(7) Comment. Any important information deemed necessary by the abstractor is included (free format). Usually the main objectives are listed. Chemical formulas more familiar than those used in item 3 should be added. Blanks are allowed, but are condensed as one blank in the output. At most two continuation cards, signified by an asterisk (*) in column 1, are allowed. For a genuine methodological study related to the ab initio calculation, items 6 and 7 are usually absent. Such an omission of an item is specified by a null entry, but slashes cannot be omitted. In Figure 1, the style of QCLDB

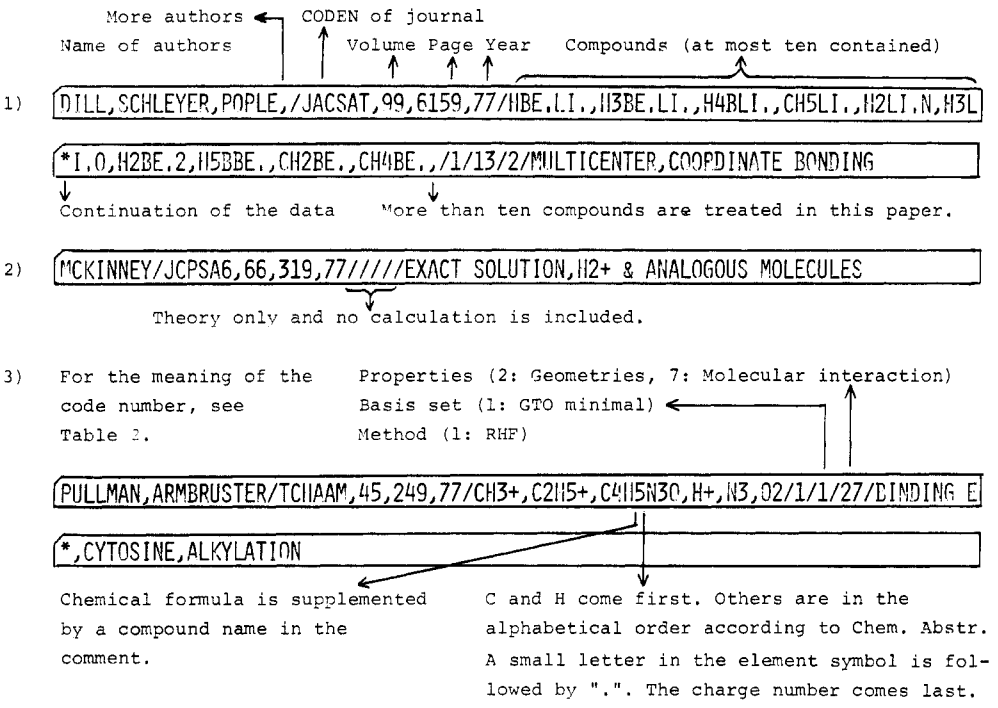

Figure 1. Examples of QCLDB input data.

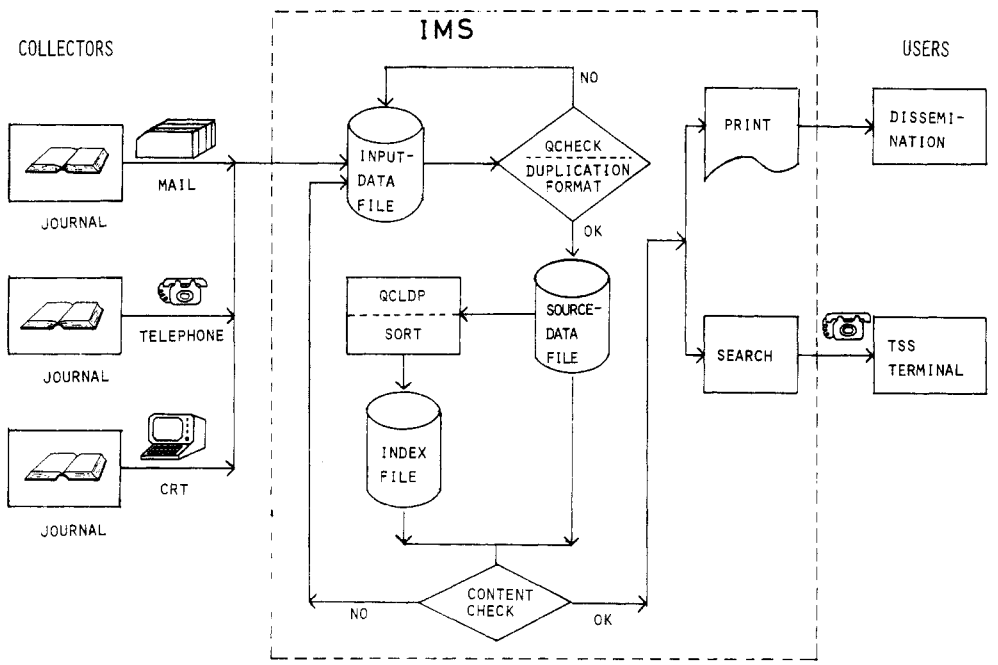

Figure 2. Flow chart of QCLDB system.

input data is illustrated with some explanations.

PROCESSING THE INPUT DATA

In Figure 2, the flow chart from the data collection to output generation is shown together with the necessary data files. Two programs for dealing with the input data have been written in PL/I on the basis of two criteria: (i) the procedure for making the input data should be simple and (ii) the linkage and application to other systems should be easy. One program (QCHECK) has the role of sorting and checking the input data. The other (QCLDP) creates indexes and gives the final printout. After the format errors of items 1–7 are removed in the input-data file, QCHECK creates the source-data file which is kept and used as the source for QCLDP.

All the data sets in the input/source file are in 80-column card-image format throughout the system. The use of such sequential data sets simplifies the structure of the processing programs (modular structure) and makes the application and the development of programs easy. In case the QCLDB system is linked with some well established and sophisticated data-base management system, the card-image treatment provides compatibility.

OUTPUT EXAMPLES

QCLDP gives a listing of source data free from format errors. In Figure 3a, the source data itself is listed. For each reference, a consecutive reference number (at the left edge) is given. To see the listing more clearly, the source data are rearranged into a tabular form (Figure 3b), which we call the master list. After sorting the data in terms of the author index, QCLDP gives the table of author names in alphabetical order in Figure 3c. Similarly, the compound index is shown in Figure 3d.

(a)

```
REF.NO.  MEMBER NAME  VOL79

          /*IWATA/CHPLBC,64(1-3),79/RIKEN(0484-62-1111-3563)/

    41    SELLERS,SCHAFER/CHPLBC,63,609,79/C3H7NO2/1/2/2/ALANINE, PULAY'S FORCE METHOD FOR
          * GEOMETRY OPTIMIZATION

    42    PERDEW/CHPLBC,64,127,79/H,HE.,LI.,BE./9/0/1/ORBITAL FUNCTIONAL FOR EXCHANGE AND
          *CORRELATION

    43    SETH,BAERENDS/CHPLBC,64,165,79/CL.LI./9/0/1/COMPTON PROFILES

    44    MELIUS,BLINT/CHPLBC,64,183,79/HO2/34/3/2/POTENTIAL ENERGY SURFACE H+O2 AND HO+O.

    45    BOUMA,RADOM/CHPLBC,64,216,79/H3FO,CH2FO2-,HF2-/1/2/2/HYDROGEN BONDS OR NOT? H2O.
          *.H..F, (HCOO..H..F)-, (F..H..F)
```

(b)

```
.....  QUATUM CHEMISTRY LITERATURE DATA BASE SYSTEM .....  MASTER LIST .....  80/01/22   19:59.42  * PAGE    4 *
```

NO.	AUTHOR	JOURNAL	COMPOUND	METHOD	BASIS	PROPERTY	COMMENT
41	SELLERS SCHAFER	CHPLBC 63,609('79)	C3H7NO2	1	2	2	ALANINE, PULAY'S FORCE METHOD FOR GEOMETRY OPTIMIZATION
42	PERDEW	CHPLBC 64,127('79)	H HE. LI. BE.	9	0	1	ORBITAL FUNCTIONAL FOR EXCHANGE AND CORRELATION
43	SETH BAERENDS	CHPLBC 64,165('79)	CL.LI.	9	0	1	COMPTON PROFILES
44	MELIUS BLINT	CHPLBC 64,183('79)	HO2	34	3	2	POTENTIAL ENERGY SURFACE H+O2 AND HO+O.
45	BOUMA RADOM	CHPLBC 64,216('79)	H3FO CH2FO2- HF2-	1	2	2	HYDROGEN BONDS OR NOT? H2O..H..F, (HCOO. .H..F)-, (F..H..F)

(c)

```
.....  QCLDB SYSTEM  .....          AUTHOR INDEX          80/01/22   20:00.26  * PAGE    6 *
```

HILLIER	193	JOSHI	82	LAZZERETTI	172	MCPHERSON	121	NOELL	54
HILLIER	280	KAIJSER	273	LEE	114	MEHLER	30	NOODLEMAN	128
HILLIER	324	KAO	57	LEE	178	MELIUS	44	NOODLEMAN	221
HINCHLIFFE	267	KAPUR	196	LEE	179	MELIUS	230	NORMAN	128
HINZE	212	KARLAU	193	LEE	191	MESTECHKIN	170	NORMAN	221
HINZE	239	KARPAS	98	LEE	220	METIU	227	NOVARO	11
HOJER	157	KARPFEN	31	LEE	325	MEYERS	127	NOVARO	312
HONJOU	284	KARPFEN	49	LEFEBVRE-BRION	247	MEZEY	196	NUMRICH	203
HORSLEY	119	KARPLUS	87	LEFOUR	133	MEZEY	302	OATES	88
HOSOYA	140	KAY	203	LEHN	295	MICHA	183	OHKAMI	140

(d)

```
.....  QCLDB SYSTEM  .....          COMPOUND INDEX          80/01/22   20:00.28  * PAGE    9 *
```

C2H4	160	C2H6	298	C3H7NO3	302	C5H6N2O2	314	C7H9N4O4	190
C2H4	168	C2H6	299	C3H8	160	C5H6SI.	33	C7H9O	302
C2H4	210	C2H6	308	C3H8	168	C5H8	100	C8H10	106
C2H4	253	C2H6N+	64	C3H8NA.S+	237	C5H8	260	C8H10N6O2	23
C2H4	286	C2H6N2	104	C3LI.4	116	C5H9+	53	C8H10O2	106
C2H4	296	C2H6N2	135	C3N2	189	C5KR.MO.05	122	C8H8	276
C2H4	297	C2H6N2	313	C4+	153	C5MO.05XE.	122	C8H9+	76
C2H4	298	C2H6O2	275	C4H10	65	C6CO.N6-3	4	C9H12O3	106
C2H4	299	C2H6O4P-	85	C4H10NA.O+	237	C6F4O2	9	C9H8	316
C2H4	313	C2H6P+	64	C4H2	88	C6H10	55	F	32
C2H4AG.+	130	C2H7B	118	C4H2N2O2	306	C6H10	133	F	114
C2H4AL.	79	C2H7N3	302	C4H3FS	57	C6H14O4P	188	F	143

Figure 3. Output examples: (a) list of original input data in the source-data file, (b) master list, (c) author index, and (d) compound index.

When a user seeks a reference of interest, he will scan the author index or the compound index and find several related references. Using the reference number, he will pick up the correct one from the master list. It is also possible to carry out such a procedure through the interactive search of the source-data file with the computer. Such a service has become available recently.

At the present time, the distribution of the listing itself is expected to be quite useful to chemists. The reason is that there are not many references of ab initio calculations in a year

Table II. Code Numbers of Items 4, 5, and 6

(4) Method

1. Restricted Hartree–Fock (RHF)
2. HF other than RHF
3. Multiconfiguration self-consistent field
4. Configuration interaction
5. Valence bond
6. Perturbation
7. Correlation function
8. Many-body treatment
9. Density functional, statistical exchange approximation (Xα)
0. Else

(5) Basis Set

1. GTO[a] minimal including STO[b]-NG
2. GTO DZ[c]
3. GTO DZ + P[d]
4. GTO extended
5. STO minimal
6. STO DZ
7. STO DZ + P
8. STO extended
9. Else

(6) Property

1. Simple one-electron properties: dipole, quadrupole moment, charge density map
2. Properties involving change of geometry: potential energy surface, bond length, bond angle
3. Dynamical properties: polarizability, scattering factor
4. Vibronic interaction: Jahn–Teller effect, Λ doubling, relaxation
5. Excited state (bound state): transition energy, oscillator strength
6. Ionization potential, photoelectron spectra
7. Molecular interaction: dimer, charge-transfer complex
8. Reaction
9. Magnetic properties: g value, coupling constant
0. Else

[a] Gaussian-type orbital. [b] Slater-type orbital. [c] Double zeta. [d] Polarization function.

and consequently alphabetical entry to the reference list (like consultation of a dictionary) is not yet difficult.

SUMMARY AND FUTURE PLAN

Several features of QCLDB are summarized with brief discussions accompanying them.

(1) QCLDB contains a correct and proper description of references. Data are drawn directly from the text, not from the title or the abstract, by a quantum chemist and reexamined by another. The collaboration of many Japanese quantum chemists has supported the high standards of QCLDB.

(2) QCLDB is compact and standarized. Only important information is included briefly. The inclusion of more information would be too time consuming and would be too much of a burden to quantum chemists who would collaborate.

(3) QCLDB is frequently updated, which is of essential importance for active users. It will be updated every half year. Users can find references which have been published in the prior 3–9 months. This timely updating is one of the merits of a computerized data base.

(4) QCLDB is transferable to other computers. All data are compiled in the card-image format. In addition, QCLDB can be easily linked with ready-made information-retrieval supporting systems. An on-line service of QCLDB has been implemented by using ORION, which is a document retrieval system available on a HITAC computer. More detailed description of this on-line system will be published elsewhere.

The service of QCLDB is currently available on the computer at the Institute for Molecular Science (IMS) and will shortly become available on a few more computers in Japan.

Users can obtain a complete output listing of the data and employ the TSS-searching system. Although QCLDB is still on a trial basis and only 2000 literature citations have been collected at this stage, we have already found QCLDB quite useful for searching a particular reference. We are planning to continue the data collection and service as well as to improve the system. We also hope that the service of QCLDB will become available for overseas users through an appropriate channel.

In this paper we have presented the literature data base (QCLDB). We envision for the future a complete Quantum Chemistry Data Base (QCDB) consisting of QCLDB, Quantum Chemistry Numerical Data Base (QCNDB), and Quantum Chemistry Program Data Base (QCPDB). By making full use of these data bases, a researcher who would like to know a calculated value of certain properties for a molecule within a given accuracy will be guided to a desired result. He may find a proper reference in QCLDB or the values themselves in QCNDB. Otherwise, he may actually make a calculation for the molecule with a suitable program in QCPDB.

ACKNOWLEDGMENT

This project has been supported by the grant-in-aid from the Japanese ministry of education. A number of young Japanese quantum chemists have been contributing to the collection of data for the QCLDB. We express our appreciation for their invaluable help. Thanks are also due to the technical staff members, especially F. Nishimoto and S. Yamamoto, of the computer center of IMS for their active assistance.

REFERENCES AND NOTES

(1) W. G. Richards, T. E. H. Walker, and R. K. Hinkley, "A Bibliography of ab initio Molecular Wave Functions", Clarendon Press, Oxford, 1971.
(2) W. G. Richards, T. E. H. Walker, L. Farnell, and P. R. Scott, "Bibliography of ab initio Molecular Wave Functions. Supplement for 1970-1973", Clarendon Press, Oxford, 1974.
(3) W. G. Richards, P. R. Scott, E. A. Colbourn, and A. F. Marchington, "Bibliography of ab initio Molecular Wave Functions: Supplement for 1974-77", Clarendon Press, Oxford, 1978.
(4) K. P. Huber and G. Herzberg, "Molecular Spectra and Molecular Structure IV. Constants of Diatomic Molecule", Van Nostrand Reinhold, New York, 1979.