Designing molecules and crystals by computer

by A. Koide

An in-depth overview of computer-aided chemical design is presented through a discussion of three systems that we developed: the Molecular Design Support System, MolWorld, and the Molecular Orbital Graphics System. The first is considered as an example of the kernel of a simulation system for industrial research and development. The chemical formula interpreter and three-dimensional molecular geometry generator of MolWorld are discussed as a compact realization of intelligence. Finally, the use of visualized molecular electronic structures in relation to chemical reactions is considered in the discussion of the Molecular Orbital Graphics System.

The design of molecules by means of computer graphics and simulation was started just ten years ago[1,2] by a small number of researchers as a new approach to developing drugs. Now, this approach attracts many industrial researchers as a very promising method for accelerating the research on and development of new chemicals and materials.[3–6] The cost and time of the experimental trial-and-error process will be reduced dramatically if it is possible to construct new hypothetical molecules and crystals easily at a graphics terminal, simulate their chemical and physical properties reasonably accurately by computer, and obtain information on how to synthesize or process them.

Considering how important this reduction would be, the IBM Tokyo Scientific Center and other IBM divisions[7–15] have been researching and developing such systems and their related technology since 1983. In this paper, I describe the basic features of computer-aided chemical design, the technology we have developed, and the future direction of research in this area by referring to the following systems:

- Molecular Design Support System
- MolWorld
- Molecular Orbital Graphics System

Work done at the Tokyo Scientific Center on related topics, namely techniques for rewriting simulation programs for new computational environments such as Expanded Storage and the Vector Facility, can be found in a previously published paper.[16]

The Molecular Design Support System[17] is an integrated total system of chemical simulations, graphics, and database on the IBM 9370, 43XX, and 30XX mainframe computers. The system was developed in Japan as a country-implemented program by IBM customers, IBM's industrial marketing group, and the Tokyo Scientific Center. I will focus my discussion on the general outline of the simulation system for industrial research and development.

MolWorld[18] is a program product developed by the Tokyo Scientific Center on the IBM Personal System/55, or PS/55 (a personal computer announced

in May 1987 and marketed by IBM Japan). MolWorld offers a highly intelligent visual molecular model-building environment. For example, a user can build a molecular model by entering a chemical formula such as

N(CH3)3CH2PO2CH2COOH
or n(ch3)3ch2po2ch2cooh

from the keyboard. The system analyzes and converts this keyboard string into a chemical graph and then estimates and displays its stable three-dimensional molecular structure. This process is performed in a fraction of a second. The user can then manipulate the structure on the screen by using a mouse.

The Molecular Orbital Graphics System is an experimental system designed to run on an IBM mainframe and the IBM 5080 Graphics System, which offers a highly interactive visualization environment for molecular electronic structures. I will discuss how the interactive graphics helps in the simulation of chemical reactions.

We are now developing systems similar to MolWorld and the Molecular Orbital Graphics System for use in the field of solids, and we will briefly discuss these projects too.

Designing a molecular design support system

From the 1960s to the 1970s, computer systems in industrial research were concerned mostly with the retrieval of chemical literature and the statistical analysis of experimental data. Today's computer-aided chemical design systems are quite different: from simulations at the atomic and subatomic levels, they help researchers to create their own working hypotheses of how the desired chemical and physical properties can be achieved from particular three-dimensional atomic configurations.

This new approach has been accelerated by the accumulation and distribution of simulation programs by university and industrial[7–15] researchers and by the accumulation of experimental knowledge at atomic and semi-macro levels, as well as by the growth of computer power and by lower costs. For example, in the area of biology, the advance of the electronic scanning microscope revealed that various proteins are built into the cell membrane of lipid molecules.[19] Now, the researchers are studying the mechanism by which these proteins selectively exchange chemical substances inside and outside the membrane at the atomic and subatomic level. This research is very important for designing the drugs that penetrate membranes well or that block penetration and also for designing artificial membranes that have effectively the same function as biological ones. The invention of the scanning tunneling microscope will put material researchers in a similar position.

Characterization of simulation system for industrial research. A computer system for industrial research is not a simple collection of simulation programs or

Our system visualizes all of the available functions, commands, and data as screen objects in multiple windows.

three-dimensional molecular graphics. First, it should allow a strategic combination of simulation programs. For example, programs based on the molecular mechanics method run faster than those based on the molecular orbital method, but the former method is less informative about chemical reactions. A researcher may use the former to estimate stable molecular geometries and then the latter to obtain information on the chemical reactivity from these geometries. Generally, different simulation programs approximate the real world in different ways and calculate different levels of information. The tradeoffs among the programs are the computational time involved and the depth of the resulting information. A strategic combination of simulation programs is economical and sometimes necessary in order to treat problems realistically.

Second, the system should present an easy-to-use or easy-to-learn user interface and not be designed only for a small number of specialists in computational chemistry or physics. The real world is a complicated interacting collection of molecules and crystal fragments. Therefore, the system will be used mostly as a guide and for screening and interpreting experiments, and it should be assumed that some of the users are experimental researchers.

Third, the system should be flexible in allowing new simulation programs, because the advancement of science gives rise to new simulation programs every year.

Finally, it should organize the input and output data of simulation programs as a flexible database. The data may be organized according to their chemical meanings or research project units.

The kernel of the molecular design support system. To meet the requirements above, we adapted an object-oriented-like approach in designing the kernel of the Molecular Design Support System. In principle, our system visualizes all of the available functions, commands, and data as screen objects in multiple windows. The functions are the molecular structure editor, the molecular shape analyzer,[20] various simulation input generators, the batch machine, and so on. They are symbolically represented by icons. The data are chemical graphs, conformers,[21] amino acid sequences, documents, numerical data sets, control command sheets, and so on, and are represented symbolically by icons or in a straightforward manner. For example, the numerical data set can be seen in the editable table of numbers or in multidimensional diagrams in the straightforward representation. A user can perform simulation research by visually manipulating these screen objects with a mouse.

The system organizes the data in a tree structure. The top of the tree is called the "cabinet," and other nodes of the tree are called "holders." Any holder can be stored as an object of another holder unless the link relation destroys the tree structure. The cabinets and holders are visualized by icons. The user can create and manipulate this tree structure by copying or moving the icons and other screen objects with a mouse. Since the cabinets and holders function as the chemical database, the user can directly access his or her data by chemical semantics from all of their descendants. That is to say, the user can see a list of conformers or retrieve the chemical graphs that contain the subgraphs specified.

The user can submit a job to execute a simulation program by following the procedure below:

1. Open the icon of a chosen simulation input generator.
2. Copy or move the data there from the private holder or public cabinets. The data can be chemical graphs, conformers, or command sheets.
3. Edit the individual copy or the default command sheet if necessary.
4. Copy or move the icon of the simulation input generator to the batch machine.

Now the user can begin another interactive job because the simulation starts in the external batch-computing environment, which may be CMS-batch or another real MVS machine, according to the machine configuration. When the simulation finishes, the system stores the result in the batch machine. The user can display it and store it in his or her holder. Thus, users can easily enjoy the combinatorial use of simulation programs by using "copy" or "move" commands on the icons.

MolWorld, a highly intelligent molecular builder

MolWorld is intended to create an intelligent visual environment where users can easily and rapidly build meaningful three-dimensional molecular geometries for further simulation study. Here, "easily" means that

- Commands are unnecessary or easy to learn.
- Visualized data are directly manipulable.
- Required operations are very few.

"Rapidly" means that the program generates the molecular geometries at the speed that makes users feel as if they were in a molecular world. "Meaningful" means that the three-dimensional molecular geometry is energetically stable.

In a rigorous sense, it is difficult for any simulation program to accurately compute a stable three-dimensional molecular geometry at such a speed, even on a supercomputer. However, since MolWorld assumes further simulation study, it is sufficient if the calculated geometries are so close to stable geometries that they greatly reduce the cost of geometric optimization in a succeeding simulation program. Therefore, we chose the approach of programming empirical chemical knowledge with the critical reformulation from molecular orbital theory.

The problem of offering an "easy" environment is that previous menu systems or direct-manipulation systems, which were intended to be an environment of unnecessary-to-learn or easy-to-learn commands, frequently increased the number of required operations. A conventional approach[22] to this problem is for system designers to consider the user interfaces separately for inexperienced and experienced users.

Table 1 The conformation of the alanine molecule

Atomic Identifiers	Atomic Types	Atomic Coordinates x	y	z
1	N	0.0000	1.0360	−1.0457
2	H	0.8276	0.9344	−1.6220
3	H	−0.8276	0.9344	−1.6220
4	C	0.0000	0.0000	0.0000
5	H	0.0000	−0.9812	−0.4631
6	C	1.2427	0.1526	0.8654
7	H	1.2427	−0.6111	1.6361
8	H	2.1286	0.0438	0.2485
9	H	1.2427	1.1338	1.3285
10	C	−1.2427	0.1526	0.8654
11	O	−1.4310	−0.5936	1.7730
12	O	−2.1182	1.1200	0.6072
13	H	−2.8361	1.0775	1.2389

Table 2 The conventional connection table of the alanine molecule

Identifiers of Bonded Atoms From	To	From	To
1	4	6	7
1	2	6	8
1	3	6	9
4	10	10	12
4	6	10	11
4	5	12	13

Table 3 Our connection table of the alanine molecule

Identifiers of Entry Atom	Ordered Set of Identifiers of Atoms Bonded to the Entry Atom
1	4 2 3
2	1
3	1
4	10 1 6 5
5	4
6	4 7 8 9
7	6
8	6
9	6
10	12 4 11
11	10
12	10 13
13	12

More important to us in designing a specific application program were

- Careful and natural assignments of the functions to the keyboard, PF (program function) keys, and mouse operations, from a chemical viewpoint
- Ability of the system to guess from situations and common sense the user's intentions, from a short input sequence

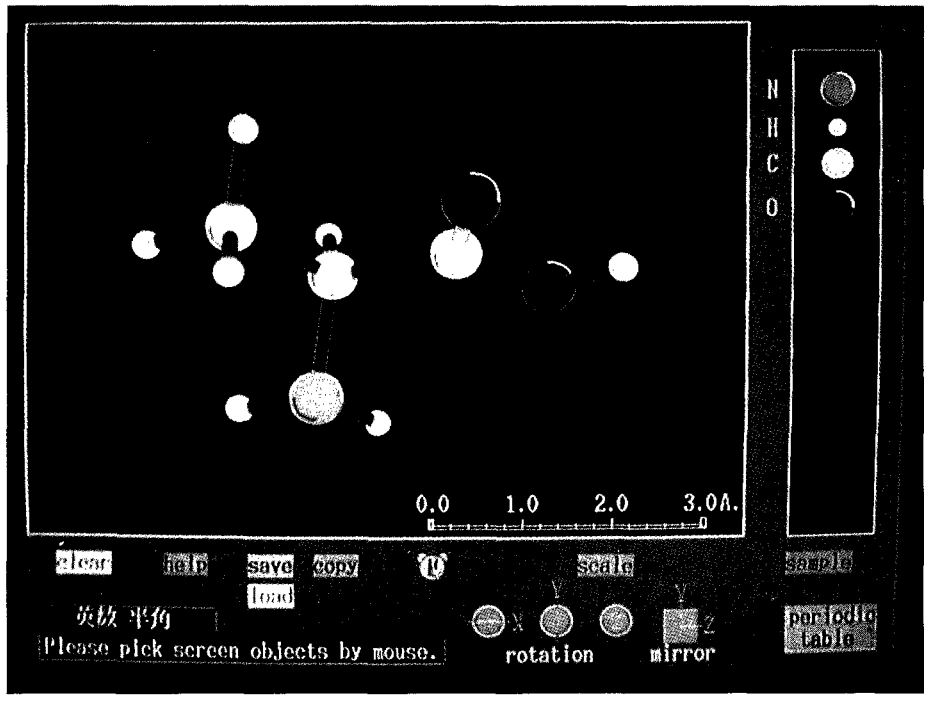
Our chemical formula input method and block manipulation method in MolWorld were developed according to this design philosophy, as was the ability of MolWorld to selectively display as the screen objects the "available" commands from information such as the preceding operations and the displayed subworld and objects. The user of MolWorld first selects atoms and then selects a command with the mouse from the displayed commands to manipulate displayed molecular structures.

To describe how we achieved this intelligence compactly for the personal computer, I must first mention our data model on molecular structures.[23] Then I will describe our block manipulation, chemical formula interpreter, three-dimensional molecular geometry generator, and bonding model.[24]

Data model of molecular structures in MolWorld. Molecules and ions can be expressed for use on computers by molecular charges, collections of the atomic types and coordinates, and collections of bonds. Each bond can be expressed by a pair of atoms (atom-identifiers). A collection of atomic types and coordinates is called a conformation. A collection of bonds is called a connection table. Tables 1 and 2 show, as examples, the conformation and connection table of the alanine molecule $NH_2CH(CH_3)COOH$. The connection table given in Table 2 is sometimes called the "from-to list." Figure 1 shows the display image of this molecule in MolWorld, where the atoms and bonds are visualized as colored spheres and cylinders, and the colors denote their types. The user can change their radii and colors, since this display is based simply on the convention, not on physical fact.

A collection of atomic types and bonds is called a chemical graph and plays a central role in our MolWorld. When we consider the atoms as vertices and the bonds as the edges, they form a graph in the mathematical sense. The difference is that each vertex of the chemical graph has an atomic type as attribute. Since the bonds indicate the pairs of strongly interacting atoms, we can consider the bonds as a user's request to place the specified atoms at a short distance. Therefore, we designed the system so that the creation or manipulation of a chemical graph requests MolWorld to generate a conformation that satisfies this constraint. A user can put data into the chemical graph in one of the following ways:

Figure 1 The alanine molecule $NH_2CH(CH_3)COOH$, generated by MolWorld



- Chemical formula input
- Screen manipulation of displayed molecules
- File input of the connection table

For example, when the user types a chemical formula and presses the ENTER key, MolWorld interprets it as a chemical graph and issues a request to the submodule for coordinate generation.

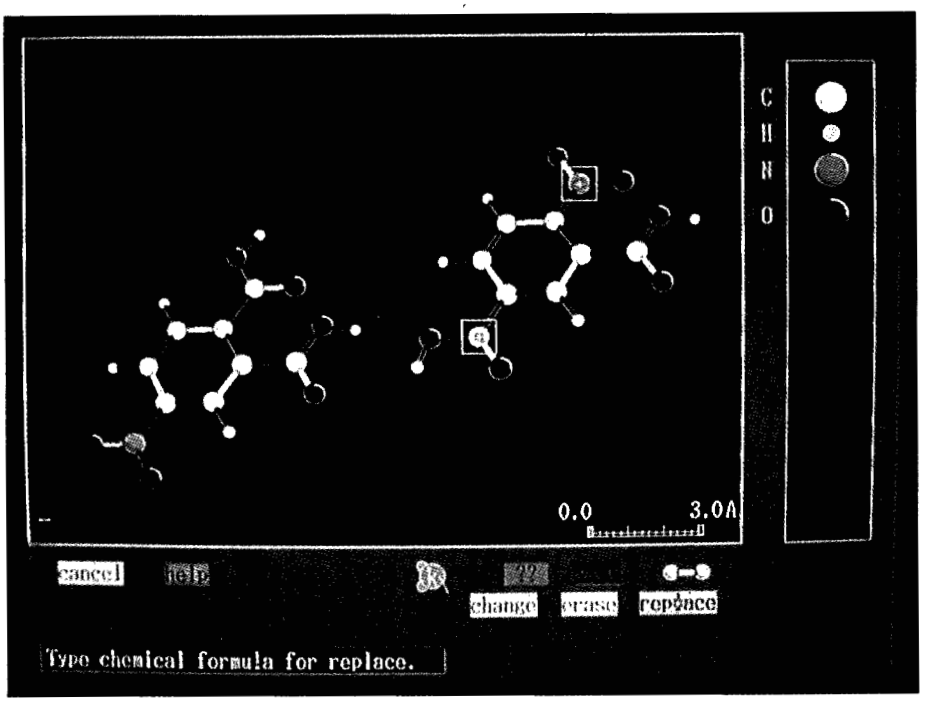
This design stance requires some extension of the chemical graph, because the connection table in Table 2 cannot distinguish optical isomers. In MolWorld, the connection table is expressed as in Table 3; internally it is expressed by the list structure.[23,25] In the connection table of Table 3, each row is the ordered collection of the atom identifiers that are bonded to the entry atom. If we neglect the order of the identifiers in the rows, this representation is equivalent to that of Table 2. In MolWorld, the order is meaningful and its allowed permutations are determined by the atomic type and orbital hybridization type of the entry atom. For example, the odd and even permutations are distinguished for the carbon atom C with the identifier number 4 in Table 3. This treatment is important for generating the optical isomers of amino acids and sugars contained in bio-material.

Block manipulation of molecular structures. Atom-by-atom manipulation of molecular structures is neither efficient nor natural. Since the basic laws of nature are invariant under the rotation and translation of the coordinate system, users want to manipulate structures block by block, keeping the relative geometry of the atomic coordinates of each block. A block can be defined from the chemical semantics and the syntax of the chemical graph. Here I describe how MolWorld dynamically determines blocks from the syntax according to the user's mouse operation.

For this purpose, MolWorld uses a very simple idea: connected components of chemical graphs. In this case, a connected component is the maximum subgraph where any two vertices can be connected by traversing the edges. It is known that the Depth-First-Search (DFS) algorithm[25] rapidly decomposes the graph into connected components, in proportion to the number of vertices and edges, if we use the representation of Table 3 for the connection table.

When a user selects the same atom with the mouse twice in succession, MolWorld defines the block for manipulation by the connected component containing the selected atom. We call this block the molecule in MolWorld. The user can file and erase it and calculate its geometrical properties.

Figure 2 External trees specified and exchanged by selecting two atoms with mouse; before-manipulation and after-manipulation displayed for comparison



When a user selects only one atom once in the operation sequence, MolWorld defines the "trees" attached to this atom as follows. If it is temporarily removed from the graph, its connected component is decomposed into new connected components. We call the resulting components trees. The user can efficiently create isomers by selecting some atom and the "change" command with the mouse; MolWorld permutes the order of identifiers and rotates the atomic coordinates, keeping the relative geometry of each block.

When a user selects two distinct atoms of the same molecule, MolWorld defines the "external trees" as follows. Let us first denote two atoms as *a* and *b*. As mentioned in the previous paragraph, atom *a* is the one that defines the trees attached to it. We define the external tree associated with atom *a* by the union of it and the trees that do not contain atom *b*. In the same manner, we can define the external tree associated with atom *b*. Thus the user can efficiently exchange the external trees by selecting two atoms and the "change" command. Figure 2 shows an example of this block manipulation. If two selected distinct atoms are bonded, MolWorld rotates one of the external trees around the bond in response to the same "change" command. In this case, the rotational angle is determined by the atomic types and bond type.

Chemical formula input method. A chemical formula is a sequence of atomic symbols and numbers. By using a chemical formula, chemists can frequently communicate the same chemical graph to each other. If a machine (i.e., computer) can convert a chemical formula into the same chemical graph as that imagined by the chemists, the chemical formula can be used as the efficient method of entering the chemical graph into the machine. From this reasoning, we developed the chemical formula interpretation method for MolWorld.

For example, a user can input a fragment of the polymer Kevlar-49 by typing (-COC6H4NH-)20 at the keyboard. MolWorld immediately generates the 280 atoms and 289 bonds of the chemical graph from the 14 characters of this string and displays the estimated stable geometry of the molecule. The chemical formula can also be entered in lowercase characters, i.e., (-coc6h4nh-)20. As this string is entered, MolWorld converts it into a sequence of the correct atomic symbols, using the preprogrammed

statistical data on atomic symbol occurrence in chemical formulae. The chemical formula input method can also be used in substituting a block of atoms for the selected atoms. The user selects with a mouse the atoms to be replaced and types a chemical formula at the keyboard.

Our chemical formula interpretation method is based on the strategic rewriting of the input string to go into the chemical graph by using rules and a word dictionary. Let us consider the chemical formula

NH2CH(CH3)COOH .

When we divide this formula into the substrings

NH2 / CH / (CH3) / CO / OH

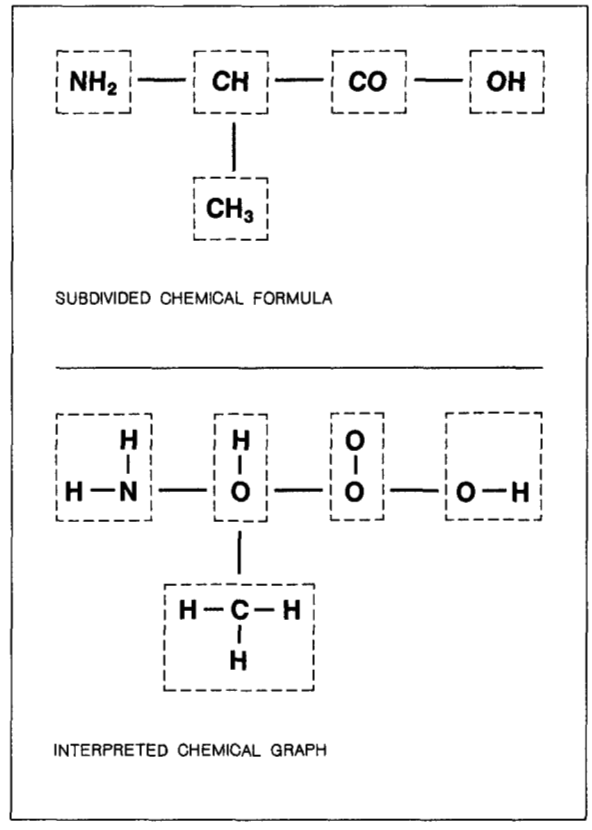
we observe that the atoms are attached to the first atom in each substring, and the substrings of the atomic groups are connected to each other by the first atoms of the substrings, as illustrated in Figure 3. This indicates that the proper subdivision of the string simplifies the interpretation of chemical formulae. By analyzing this example, we know the string can be divided

- Before syntax symbols such as an opening parenthesis "("
- Before specific types of atoms, such as "C" and "N", to which many other atoms can be attached
- Between repeated atomic symbols such as "OO"

By collecting and analyzing chemical formulae found in the literature, we derived various subdivision rules as well as other types of rules such as skeleton-generation rules, rules for the selection of atoms connecting substrings, and rules for the transposition of strings. In our previous example, the chemical graph was essentially in the form of a chain, but it is not generally true. The skeleton-generation rules determine whether a substring forms a chain or a ring structure by considering the valence model. Here the valence model is the hypothesis that the maximum number of bonds going out from an atom is inherent in its atomic type.

Thus, our method is analogous to the machine translation of natural languages; the input data is interpreted by the rules and the word dictionary (a collection of the stored substrings with their chemical graphs). The difference is that the chemical formula interpretation requires a smaller dictionary, which can be completely loaded in the computer before execution. Needing only a smaller dictionary comes from the fact that actual molecules satisfy the valence model relatively well, which makes the program more compact and faster than the usual machine translation.

Figure 3 Interpretation of the string $NH_2CH(CH_3)COOH$ by subdivision



Generation of three-dimensional molecular structure. The previous methods for generating three-dimensional molecular geometries from chemical graphs can be classified into the following two approaches. One approach optimizes the atomic coordinates by some model potential from the two-dimensional coordinates drawn by a user on the screen. The other approach constructs three-dimensional geometries by combining the prepared three-dimensional parts that have chemical graphs identical to subgraphs of the given chemical graph. The problem of the former approach is that

- A two-dimensional chemical graph is always required as input.

- The optimization process is time-consuming.
- The geometry is sometimes optimized to a meaningless local energy minimum.

The problem of the latter approach is that

- Too many parts are required for handling all kinds of possible input.
- The subdivision way of a chemical graph is ambiguous.

The approach in MolWorld is different from both of these approaches. It is closer to chemists' reasoning process in molecular geometry and is therefore very flexible. First, MolWorld qualitatively estimates electron configurations from chemical graphs, as described in the next subsection. At this stage, the bond types and formal atomic charges are determined. Then, MolWorld estimates the local geometry around each atom from the estimated electron configuration. The estimated local geometries are bond lengths and angles. Finally, MolWorld strategically searches for the molecular geometry that satisfies the local geometries.

MolWorld decomposes the chemical graphs into chain and ring parts.

In the last step, the construction of the global geometry, MolWorld decomposes the chemical graphs into chain and ring parts. In the chain part, three-dimensional atomic coordinates are generated in the order of the DFS algorithm,[25] applying empirical rules for the torsional angles. For example, the two bonded atoms of sp^3 prefer anti-gosh geometry for the torsional angle, the two bonded atoms of sp^3 and sp^2 prefer the staggered geometry, the double bond prefers the planar shape for the delocalization, the adjacent double bonds transmit bond planes orthogonally, and so on. MolWorld decomposes the ring part into a collection of single cycles by the Breadth-First-Search algorithm.[25] It then generates the atomic coordinates to close each ring, respecting local geometries in the order of the bond lengths, bond angles, and torsional angles.

Formulation of bonding model. Bond types can be considered as a kind of qualitative description of the electronic structures of molecules and ions. They are classified by chemists into ionic bonds, single, double, and triple covalent bonds, and coordinated bonds. MolWorld currently handles covalent bonds and coordinated bonds.

In the molecular orbital theory, the bonding of atoms is explained as follows:

- Each electron occupies either an atomic orbital or a molecular orbital. Each orbital can accommodate a maximum of two electrons.
- Molecular orbitals are formed by the linear combination of atomic orbitals, so that the total electronic energy decreases. The electrons occupy the lower-energy orbitals first.
- If the formation of molecular orbitals decreases the energy, bonds are formed.

MolWorld largely simplifies the above interpretation of bonding to formulate the problem as the optimization of integer variables. We assume that

- Molecular orbitals are formed by the overlap of the atomic orbitals of the "two" atoms and are localized around the bonds.
- Molecular orbitals are separated by their energies into bonding orbitals and antibonding orbitals.
- The change in the total electronic energy can be qualitatively given by the sum of the bond orders.

Here we define the bond order for each bond by

$$B = \frac{1}{2}(N_b - N_a) \tag{1}$$

where N_b and N_a are the numbers of electrons in the bonding orbitals and the antibonding orbitals, respectively. From this definition, the bond orders are integers or half-integers. If a bond order is greater than zero, we say a bond is formed.

Since our model does not consider quantitatively the energy gap between bonding and antibonding orbitals, we limit the atomic orbitals and electrons that contribute the bonding and call them valence orbitals and electrons. Let A_i and N_i be the numbers of valence orbitals and electrons of the atom i. We assume that the valence set (A_i,N_i) is inherent to atomic types of the atom i but is not unique. For example, the nitrogen atom N has the single valence set (4,5), whereas the phosphorus atom P has two valence sets (4,5) and (5,5). For most heavy atoms, we assigned two sets, which we distinguish by calling

one set primary and the other secondary. This treatment is based on the assumption that s- and p-orbitals primarily make the main contribution to the bonding, and in some cases the limited number of d-orbitals additionally contribute to the bonding.

Our bonding model is now formulated as follows. For a given chemical graph, we solve nonnegative symmetric matrices A_{ij} and N_{ij}, which optimize the sum of bond orders under the constraints

$$A_i = \sum_j A_{ij} \tag{2}$$

$$N = \sum_i \sum_j N_{ij} \tag{3}$$

and

$$A_{ij} - |A_{ij} - N_{ij}| \geq 0 \tag{4}$$

Here A_{ii} and A_{ij} are the number of unused valence orbitals of the atom i and the number of the used valence orbitals for the bond between the atoms i and j. N_{ii} and N_{ij} are the number of electrons accommodated in the atomic orbitals A_{ii} and half the number of the electrons accommodated in the molecular orbitals localized to the bond between the atoms i and j. The effective total number of electrons N in Equation 3 is given by

$$N = \sum_i N_i - I \tag{5}$$

where I is the molecular charge. For different i and j, the left side of Equation 4 is equal to the bond order. They should be positive if the atoms i and j are bonded in the given chemical graph; otherwise they should be zero.

Now we can find out how to use primary and secondary valence sets. The secondary sets are used only when the primary sets fail to give a solution under the constraints from Equations 2 to 4, or when some bond order takes the value 1/2. If even the secondary sets fail to give a solution, MolWorld estimates the molecule to be unstable. If even the secondary sets yield the value as 1/2 of a bond order on some bond, MolWorld estimates the molecule to be weakly bonded.

In our actual system, more constraints are added to our described bonding model so that the maximization on the sum of bond orders may yield a unique solution. Even so, MolWorld sometimes yields multiple solutions for a given chemical graph. In this case, MolWorld considers that the molecular orbitals are delocalized; in other words, the bond types are in resonance. The typical example is the molecule Benzene. MolWorld uses information on this delocalization to determine bond lengths, bond angles, and torsional angles.

Figure 4 shows the geometries of the ions IO_4^-, $IO_2F_2^-$, and F_4^- generated by MolWorld. MolWorld estimates the local geometry of molecules by the hybridization types, not by the number of outgoing bonds. Figure 5 illustrates their electronic configurations obtained by MolWorld. The dots denote valence electrons, and the loops denote valence orbitals or bonding orbitals. The ions $IO_2F_2^-$ and IF_4^- require one d-orbital and two d-orbitals, respectively, because of the extra valence electrons. Thus, our bonding model provides a good explanation of the geometries of molecules and ions containing heavy atoms. A user of MolWorld can change the supplied valence sets and bond lengths by editing his or her profile.

Interactive molecular orbital graphics

The simulation programs for molecules can be classified into molecular orbital, molecular mechanics, and molecular dynamics methods. The molecular mechanics and molecular dynamics methods calculate molecular geometries and properties by the use of interatomic model potentials. The molecular orbital method can just calculate the interatomic potentials and can give information on chemical reactivity and optical processes by the quantum mechanical treatment of electrons.[26] However, since this method is very expensive and since quantum mechanics is not very intuitive, an elaborate visual simulation environment is required in order to utilize it efficiently.

In this section, I first describe the method of visualizing electronic structures in our Molecular Orbital Graphics System and then discuss how to use it in the study of chemical reactions.

Method of visualizing electronic structures. The electronic structure of a molecule is theoretically defined by the wavefunction which satisfies the Schroedinger equation and whose arguments are the coordinates and spins of all of the electrons in the molecule. Most computational methods, such as SCF,[27] MCSCF,[28] the MP-method,[29] and the MBPT-method,[30] construct approximate solutions of wavefunctions by the products or the sum of the products of the molecular orbitals that are spatial functions of a single electron.

Figure 4 Display of ions IO_4^-, $IO_2F_2^-$, and IF_4^- from the left on the screen

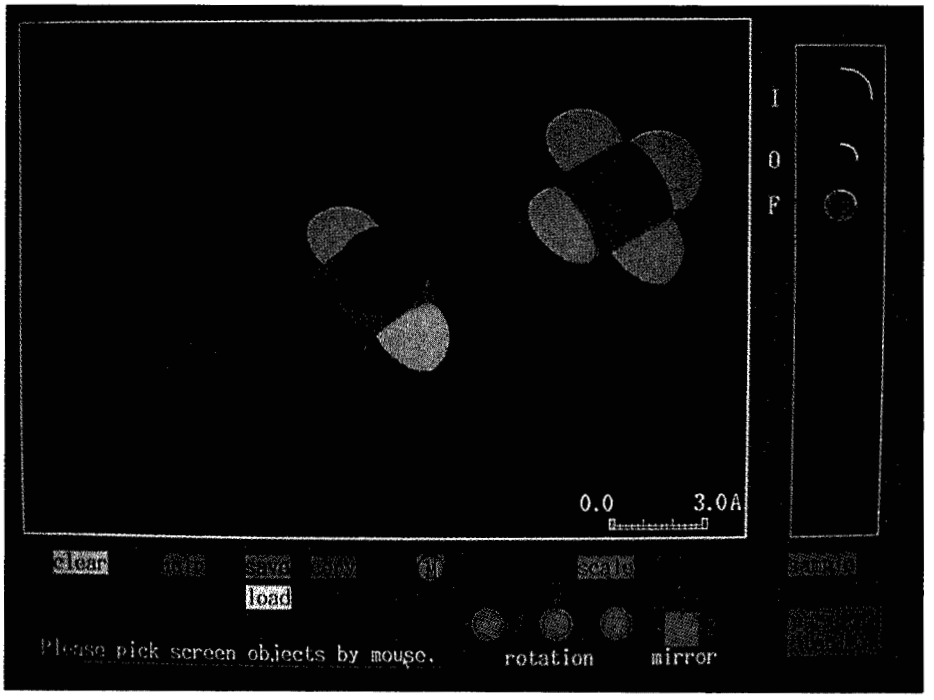
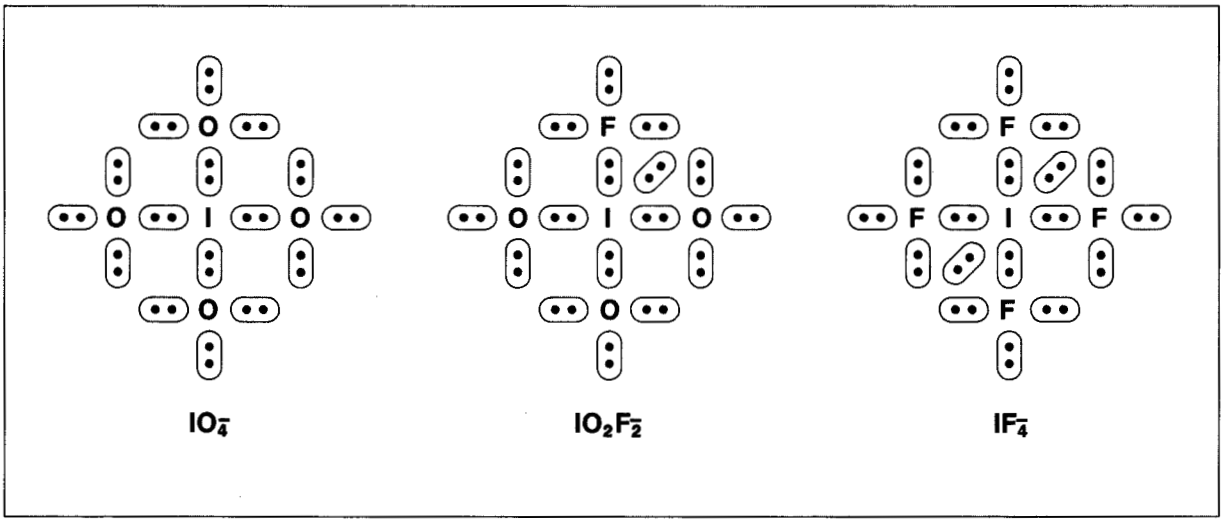


Figure 5 The electron configurations obtained by the bonding model of MolWorld

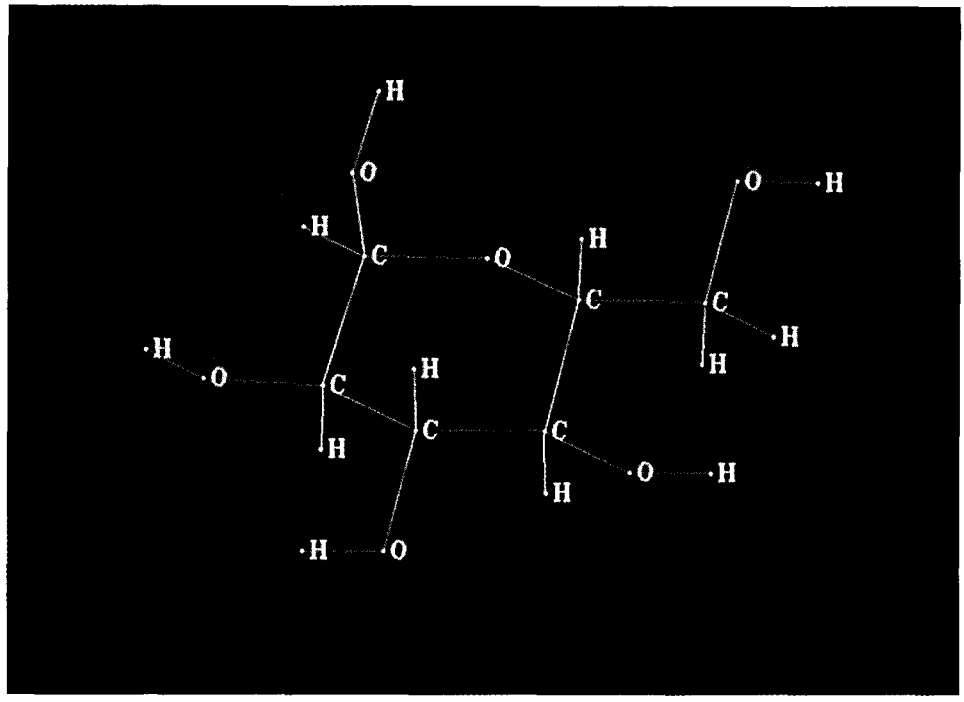


For this reason they are frequently referred to collectively as the molecular orbital method.

Our Molecular Orbital Graphics System offers an interactive environment for displaying the electron densities and molecular orbitals by using equivalued surfaces. Here, the equivalued surfaces are defined by the set of points satisfying

$$F(x,y,z) = C \tag{6}$$

Figure 6 Equal-electron-density surfaces of alpha-glucose ($C_6H_{12}O_6$) in a semitransparent display



for a given spatial function $F(x,y,z)$ and user-specified surface constant C. Figure 6 shows the equivalued surfaces of the electron density of alpha-glucose in a semitransparent display.[31] The surface constants are 0.0081, 0.03, and 0.09 from the outer surfaces. Thus, equivalued surfaces can be considered as a three-dimensional extension of ordinary contour maps.

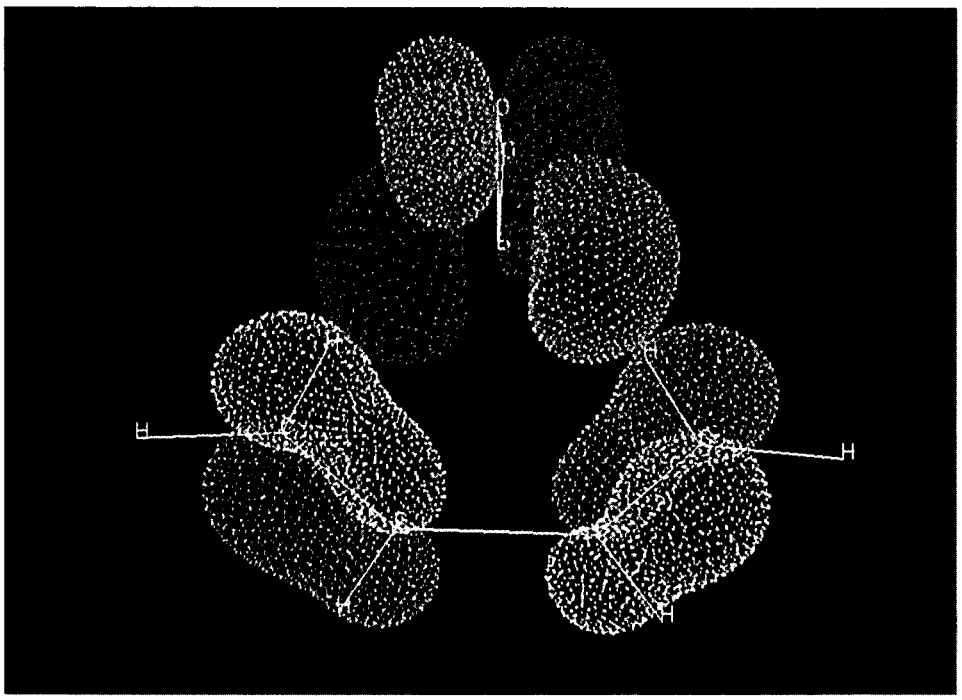
Our system first triangulates equivalued surfaces to polyhedral data and then displays them on the screen. The advantages of this approach are as follows:

- It is economical for interactive graphics, because the re-evaluation of spatial functions is not required for rotation, shift, and zoom once the polyhedral data have been generated.
- Many graphics software and hardware technologies are applicable for rapid image synthesis, because polyhedral data are very standard in computer graphics.

We developed two methods for the triangulation of equivalued surfaces: the propagation method[32] and the modified tetrahedral cell method.[33] Both methods yield the vertex and triangle tables. The vertex table is a collection of the coordinates and surface normals of the vertices. The triangle table is a collection of the ordered sets of three vertex identifiers. The order of identifiers is used to distinguish between the inside and outside of surfaces. The surface normals are used to create smoothly shaded images.[34] Our observation is that the modified tetrahedral cell method is slightly faster than the propagation method and can treat more general types of equivalued surfaces. In contrast, the propagation method generally produces better quality polyhedral data, particularly when equivalued surfaces are near the nuclei, because the modified tetrahedral method uses essentially quadratic interpolation, whereas the derivatives of electron density are generally discontinuous at the nuclei.

Our system visualizes the equivalued surfaces by a polymarker display and a smoothly shaded image display. Figure 7 shows the polymarker display of molecular orbitals. By using the dials for the display, a user can move and rotate each molecule with orbitals continuously. This function is useful in guessing the reaction geometry. The following types of manipulation are supported:

Figure 7 Frontier orbitals, HOMO of C_4H_6 and the LUMO of SO_2, are shown in polymarker display; surface constant of the orbitals is ± 0.09, equivalent to 0.0081 as the probability density



- Selection of molecular orbitals or molecules
- Setting of the display types and colors
- Movement and rotation of molecules
- Reading atomic coordinates of displayed geometries

Currently the continuous movement and rotation of molecules with orbitals is supported only for the polymarker display.

Use of molecular orbital graphics for chemical reaction. Chemical reaction is the reorganization of atomic configurations caused by interatomic potentials. The interatomic potentials and their derivatives can be computed by the molecular orbital method when specific atomic configurations are given. Efficiency in research on chemical reactions is largely dependent on the selection of atomic configurations at which the interatomic potential and derivative are calculated, because the potential is a function of all of the atomic coordinates.

The typical steps for researching reaction mechanics are as follows:

1. Find the locally stable molecular geometries.
2. Find the minimum energy of the paths connecting any two of them.

MolWorld is effective for generating the initial trial geometry for the energy minimization process of step 1. The problem is step 2, since we must find the saddle points of multidimensional potential surfaces. Chemists frequently call the saddle point the transition state, which means the boundary between the reactants and products of the reaction. In any computational method, the convergency to the saddle points is slower than the convergency to local minimum points. Thus, the better initial trial geometries are required to compute the saddle points. This is why the computational study of chemical reactions is time-consuming. Once the saddle point is computed, the minimum energy of the reaction path can be computed relatively easily by solving the Intrinsic Reaction Coordinate[35] by the Runge-Kutta method.[36] Knowing the saddle points is also important for computing reaction rates by the Eyring method.[37]

Our Molecular Orbital Graphics System is a very effective way of searching for saddle points. The basic idea is to use the frontier orbital theory[38] visually.

For two closed shells of molecules, overlapping the electron densities of two molecules increases the energy because of the Pauli exclusion law. But, overlapping an occupied orbital and an unoccupied orbital decreases the energy in the second-order perturbation correction. The smaller the energy gap between the two orbitals and the larger their overlap integral, the more the energy decreases.

Thus, we propose a visual method of searching for reacting configurations as follows:

1. We display a higher occupied molecular orbital for one molecule and a lower unoccupied molecular orbital for the other molecule.
2. We overlap these by moving and rotating the second molecule.
3. We display the electron densities of the two molecules and check that the electron densities are not overlapped too much.
4. We read the atomic coordinates of this reacting configuration and submit it to the program, searching for saddle points by relaxing the forces.

Figures 8 and 9 show the geometry of the saddle points in the ring-closing reaction of $C_2H_4 + C_4H_6$. This geometry was easily found by visually overlapping the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO), and then by relaxing the force in an external program. The total electron density is shown in the color gray with the surface constant 0.09. The receptor, a dominant orbital receiving electrons, is shown in the colors blue and green. The donor, a dominant orbital offering electrons, is shown in the colors brown and orange. The surface constant of the orbitals is ±0.09. The sign is distinguished by colors. The activation energy (energy gap between the transition state and reactants) was computed as 35 Kcal/mole and the product was computed as 37 Kcal/mole lower than the reactants in the SCF calculation with a 4-21G basis set. Figures 8 and 9 also show the interacting frontier orbitals computed by Fujimoto's method.[39] This analysis indicates that the frontier orbital theory is effective in this reaction.

Figure 8 Display of the chemical reaction $C_2H_4 + C_4H_6$ with the charge transfer analysis at the transition state; receptor and donor were found close to LUMO of C_2H_4 and HOMO of C_4H_6, respectively

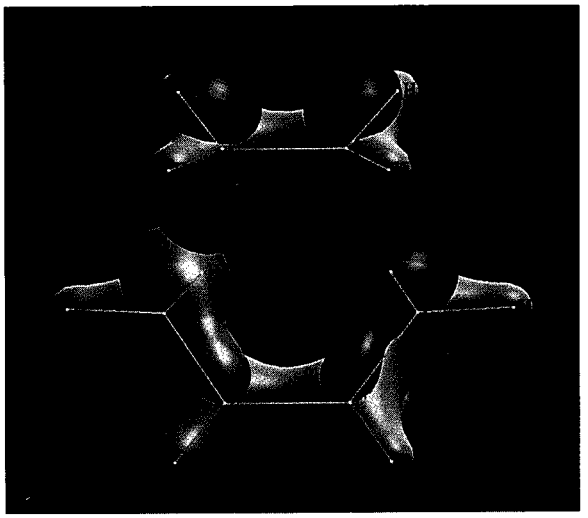
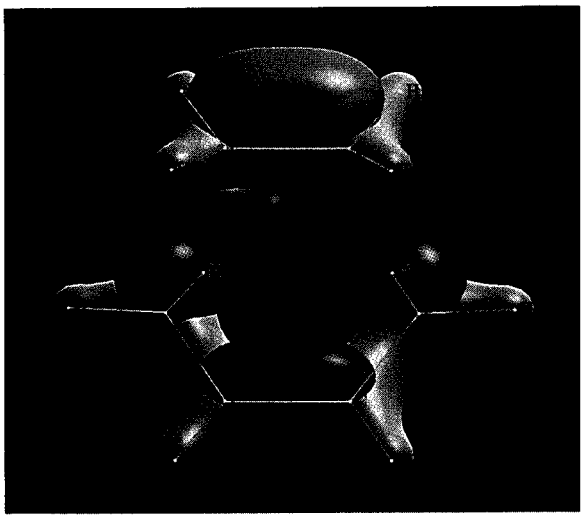


Figure 9 Another donor and receptor shown in the same convention as those in Figure 8; receptor and donor were found close to LUMO of C_4H_6 and HOMO of C_2H_4, respectively



Current projects in solids

For use in the field of solids, Crystal World and the Band Energy Visualization System are currently under development at the Tokyo Scientific Center.

Crystal World is intended to allow the user to build meaningful crystal structures rapidly and easily on a PS/55. For generating crystal structures, metallic and ionic bonds must be considered in addition to covalent bonding. Thus the bonding model of MolWorld should be extended by the use of the chemical concept of electronegativity. As a new type of input

method, the use of crystal symmetry is very promising; the system generates atomic coordinates under the constraint of a user-given crystal symmetry. Conversely, it can rapidly recognize the crystal symmetry from given atomic coordinates and lattice constants.[40]

The Band Energy Visualization System is intended to allow users to create their own working hypotheses on electronic conductivity and optical processes by displaying the energy bands, electron densities, and other related physical quantities of crystals three-dimensionally in reciprocal space. A modulator and polarized glasses are used with the IBM 5080 and an IBM mainframe for a direct stereoscopic view. Since the band energies are computed at only a finite number of sampling points in reciprocal space by a simulation program, the system first constructs the interpolating function from these values under the periodic condition, and then creates their stereoscopic images. The volume-rendering method is currently used for the image synthesis.

Concluding remarks

By describing systems and technologies that we have developed, I discussed what a computer-aided chemical design system should do for making simulation programs more accessible, more valuable, and more economical for industrial research. It can be summarized as follows:

- The system offers users an interactive data communication function among simulation programs for their strategically combined use.
- The system offers users an intelligent and visual environment for rapid creation of meaningful input data and for easy extraction of working hypotheses from simulated data.

A tight coupling of computer graphics, artificial intelligence, and numerically intensive simulation will become more and more important in future systems, because the design of molecules and crystals is essentially a strategic search for a particular atomic configuration having certain desired properties from among an infinity of possible atomic configurations. For this purpose, it is not important whether artificial intelligence is written in artificial languages or not. It is more important to formulate empirical concepts clearly in the frame of mathematics. It is done not only because of the compactness and execution speed of programmed intelligence, but also because it helps to enrich empirical concepts.

I believe that our efforts in computer-aided chemical design systems will open a new era in which computer systems will become a valuable medium for transmitting dynamic knowledge to the scientific and engineering community in the same way that book publishing has contributed to society by popularizing scientific and engineering knowledge.

Acknowledgments

The author is deeply indebted to Emeritus Professor Kenich Fukui at the University of Kyoto for his encouragement in our interactive visual search for reacting configurations by use of frontier orbitals, and to the researchers of Teijin System Technology Ltd. for their helpful discussion on the kernel of the Molecular Design Support System.

Cited references and notes

1. A. J. Stuper, T. M. Dyott, and G. S. Zander, "Conformational analysis," *ACS Symposium Series 112, Computer-Assisted Drug Design* (1979), pp. 383–414.
2. P. Gund et al., "Three-dimensional molecular modelling and drug design," *Science* **208,** 1425–1431 (1980).
3. J. B. Tucker, "Designing molecules by computer," *High Technology*, 52–59 (January 1984).
4. C. H. Hassall, "Computer graphics as an aid to drug design," *Chemistry in Britain* **21,** 39–46 (1985).
5. D. Luther, R. G. Fisher, and E. Swanson, "Molecular modeling for chemical design," *Computer Graphics World* **9,** No. 11, 28–32 (1986).
6. A. Koide, "Computer graphics in chemical CAD," *Johou-Shori, Information Processing Society of Japan* **29,** No. 10, 1182–1189 (1988).
7. S. Chin, D. P. Vercauteren, D. Vanderveken, R. Scateni, and E. Clementi, *Proceedings of the Fourth International Conference on Supercomputing*, Santa Clara, CA (April 30–May 5, 1989), L. P. Kartashev and S. I. Kartashev, Editors, Volume II (1989), p. 147.
8. P. Quarendon, *Proceedings of the Fourth International Conference on Supercomputing*, Santa Clara, CA (April 30–May 5, 1989), L. P. Kartashev and S. I. Kartashev, Editors, Volume II (1989), p. 140.
9. J.-L. Fauchere, P. Quarendon, and L. Kaetterer, "Estimating and representing hydrophobicity potential," *Journal of Molecular Graphics* **6,** 203–206 (1988).
10. M. Re, W. L. Luke, S. Chin, and E. Clementi, *Computer Graphics Tools for Chemistry and Engineering*, IBM Kingston Technical Report KGN-172, IBM Corporation (1988).
11. R. J. Hinde, W. L. Luke, and S. Chin, *Display of Charge Density Surfaces*, IBM Kingston Technical Report KGN-141, IBM Corporation (1988).
12. W. L. Luke and S. Chin, *How to Make a Movie*, IBM Kingston Technical Report KGN-140, IBM Corporation (1987).
13. P. Kowalczyk, W. L. Luke, and S. Chin, *Molecular Display Package (MDP)*, IBM Kingston Technical Report KGN-138, IBM Corporation (1987).
14. J. M. Burridge and S. J. P. Todd, "Protein secondary structural representations using real-time interactive computer graphics," *Journal of Molecular Graphics* **4,** 220 (1988).

15. P. Quarendon, "A general approach to surface modelling applied to molecular graphics," *Journal of Molecular Graphics* **2,** 91–95 (1988).
16. M. Sakaki, H. Samukawa, and N. Honjou, "Effective utilization of IBM 3090 large virtual storage in the numerically intensive computations of *ab initio* molecular orbitals," *IBM Systems Journal* **27,** No. 4, 528–540 (1988).
17. Program Numbers 5788-JKF/JKG/JKH/JKJ/JKK/JKL/JKN, IBM Japan.
18. Program Numbers 5600-115 and 5605-115, IBM Corporation.
19. S. Hakomori, "Glycospingolipids," *Scientific American* **254,** No. 5, 44–53 (1986); G. W. Goldstein and A. L. Betz, "The blood-barrier," *Scientific American* **255,** No. 3, 74–83 (1986).
20. The molecular shape analyzer offers the interactive measurement of molecular geometries such as distances, angles, surface areas, and volumes and also offers the interactive logical operations of the molecular surfaces to discover visually the similarity and difference of molecular structures in relation to chemical activity: G. R. Marshall et al., "The conformational parameters in drug design," *ACS Symposium Series 112, Computer-Assisted Drug Design* (1979), pp. 205–226.
21. The conformer is defined here by the collection of the atomic types and coordinates that form a single molecule or several interacting molecules. Thus this is wider than the conventional definition, which limits stable geometries.
22. B. Shneiderman, "Direct manipulation, beyond programming languages," *Computer* **16,** No. 8, 57–69 (1983).
23. A. Koide, "Data representation for 3-dimensional molecular structure manipulation," *Working Group of the Software Foundation of the Information Processing Society of Japan,* **WGSF-2** (September 26, 1986), pp. 1–8.
24. A. Koide, "Molecular modeling on personal computer," *Chemistry Today,* No. 198, 62–65 (September 1987).
25. M. C. Golumbic, *Algorithmic Graph Theory and Perfect Graphs,* Academic Press, New York (1980), pp. 31–42.
26. P. S. Bagus and A. R. Williams, "Electronic Structure Theory," *IBM Journal of Research and Development* **25,** No. 5, 793–809 (September 1981).
27. C. C. J. Roothaan, "New developments in molecular orbital theory," *Reviews of Modern Physics* **23,** 69–89 (1951), *ibid.* **32,** 179 (1960).
28. E. Dalgaard and P. Joergensen, "Optimization of orbitals for multiconfigurational reference states," *Journal of Chemical Physics* **69,** 3833–3844 (1978); H. J. A. Jensen and P. Joergensen, "A direct approach to the second-order MCSCF calculations using a norm extended optimization scheme," *Journal of Chemical Physics* **80,** 1204–1214 (1984).
29. R. Krinshnan, M. J. Frisch, and J. A. Pople, "Contribution of triple substitutions to the electron correlation energy in fourth order perturbation theory," *Journal of Chemical Physics* **72,** 4244–4245 (1980).
30. K. Koch and W. Kutzelnigg, "Comparison of CEPA and CP-MET methods," *Theoretica Chimica Acta* **59,** 387–411 (1981).
31. A. Koide and K. Miyata, "Semi-transparent display of equivalued surfaces," *Information Processing Society of Japan, the 37th Conference* (1988), pp. 1683–1684.
32. A. Koide, A. Doi, and K. Kajioka, "Polyhedral approximation approach to molecular orbital graphics," *Journal of Molecular Graphics* **4,** 149–156, 160 (1986).
33. A. Doi and A. Koide, "New triangulation methods of equivalued surface for interactive 3-D molecular graphics," *IEEE Compint '87* (November 1987); A. Koide and A. Doi, *A Novel Triangulation Method of Equi-Valued Surfaces Based on Tetrahedral Grids,* Tokyo Research Laboratory Research Report TR87-1017, IBM Japan (1987).
34. H. Gouraud, "Continuous shading of curved surface," *IEEE Transactions on Computers* **C-20,** 623–629 (1971).
35. K. Fukui, "A formulation of the reaction coordinate," *Journal of Physical Chemistry* **74,** 4161–4163 (1971); A. Tachibana and K. Fukui, "Intrinsic dynamism in chemically reacting system," *Theoretica Chimica Acta* **51,** 189–206 (1979).
36. J. D. Lambert, *Computational Methods in Ordinary Differential Equations,* John Wiley & Sons, New York (1973), pp. 120, 225–228.
37. K. J. Laider, *Chemical Kinetics,* 2nd ed., McGraw-Hill Book Publishing Co., New York (1965).
38. K. Fukui, "A simple quantum-theoretical interpretation of the chemical reactivity of organic compounds," *Molecular Orbitals in Chemistry, Physics, and Biology,* edited by P. O. Lowdin and B. Pullman, Academic Press, New York (1964), pp. 513–537; K. Fukui, "An MO-theoretical illumination for the principle of stereoselection," *Bulletin of the Chemical Society of Japan* **39,** 498–503 (1966); K. Fukui, "Recognition of stereochemical paths by orbital interaction," *Accounts of Chemical Research* **4,** 57–64 (1971).
39. H. Fujimoto, N. Koga, and I. Hataue, "Orbital transformation analysis, a simplification in the description of charge transfer," *Journal of Physical Chemistry* **88,** 3539–3544 (1984).
40. A. Koide, *A New Fast Classification Algorithm of Bravais Lattices,* Tokyo Research Laboratory Research Report TR 87-1022, IBM Japan (1988).

Akio Koide *IBM Japan Limited, Tokyo Scientific Center, Tokyo Research Laboratory, 5-19 Sanban-cho, Chiyoda-ku, Tokyo 102, Japan.* Dr. Koide received a B.Science in 1970 and a Ph.D in physics in 1975, both from the University of Tokyo. He is currently an advisory researcher in the Tokyo Research Laboratory. Before joining IBM in 1981, he worked as a research associate in the chemical department at the University of Western Ontario, Canada. In IBM, he is researching chemical computer-aided design systems. He developed MolWorld, an intelligent molecular modeling system on the PS/55, which was announced as an IBM program product in 1987. He also joined in the design of the Molecular Design Support System, which was announced as a Country-Implemented-Program from IBM in 1985. Dr. Koide is a member of the Chemical Society of Japan, the Physical Society of Japan, and the Information Processing Society of Japan.

Reprint Order No. G321-5378.