requirements. A block-cutpoint tree (BCT) is used to represent chemical structures. The algorithm to construct a BCT was written in APL and implemented on the IBM 5100. Input data to this algorithm are in the form of an adjacency matrix of compounds. The compound file and block dictionary have been designed. The compound file is a set of BCT representations of compounds and the topological relationship among blocks is described. Further structural informations, such as atomic identification, bond types, steric relations, etc., are described in a block dictionary. Complicated structures need another file to describe substituted positions (i.e., cutpoints), symmetry, steric relations, and so on. These will be utilized in application systems, such as substructure search, automatic analysis of spectra, and molecular design, which are being developed in our laboratory.

ACKNOWLEDGMENT

The authors wish to thank Dr. Masahiro Uchino for helpful discussions.

REFERENCES AND NOTES

(1) J. E. Ash and E. Hyde, "Chemical Information Systems", Ellis Horwood, Chichester, England, 1975, Chapter 11.
(2) D. J. Gluck, "A Chemical Structure Storage and Search System Developed at DuPont", *J. Chem. Doc.*, **5**, 43 (1965).
(3) J. E. Rush, "Status of Notation and Topological Systems and Potential Future Trends", *J. Chem. Inf. Comput. Sci.*, **16**, 202 (1976).
(4) S. Rossler and A. Kolb, "The GREMAS Systems, an Integral Part of the IDC System for Chemical Documentation", *J. Chem. Doc.*, **10**, 128 (1970).
(5) E. G. Smith, "The Wiswesser Line-Formula Chemical Notation", McGraw-Hill, New York, 1968.
(6) IUPAC "Rules for IUPAC Notation for Organic Compounds", Wiley, New York, 1961.
(7) H. Skolnik, "A New Linear Notation Based on Combination of Carbon and Hydrogen", *J. Chem. Doc.*, **6**, 689 (1969).
(8) L. Quadnelli, V. Bareggi, and S. Spiga, "A New Linear Representation of Chemical Structures", *J. Chem. Inf. Comput. Sci.*, **18**, 37 (1978).
(9) Chi-Hsiung Lin, "SEFLIN-Separate Feature Linear Notation System for Chemical Compounds", *J. Chem. Inf. Comput. Sci.*, **18**, 41 (1978).
(10) R. G. Dromey, "A Simple Tree-Structured Line Formula Notation for Representing Molecular Topology", *J. Chem. Inf. Comput. Sci.*, **18**, 225 (1978).
(11) J. E. Dubois, "Principles of the DARC Topological System", *Entropie*, **25**, 5-13 (1969).
(12) Nicos Christofides, "Graph Theory, An Algorithmic Approach", Academic Press, New York, 1975.
(13) Frank Harary, "Graph Theory", Addison-Wesley, Reading, Mass., 1969.
(14) J. Edmonds, *J. Res. Natl. Bur. Std. B*, 69 (1965).
(15) Masahiro Uchino, "Array-Theoretic Approach to Basic Problems in Chemical Information: Unique Coding, Substructure Search and Perception of Topochromic Symmetry", The ACS/CSJ Chemical Congress, 1979.
(16) Ian C. Ross and Frank Harary, "Identification of the Liaison Person of an Organization Using the Structure Matrix", *Manage. Sci.*, **1** (1955).
(17) C. Berge, "Graphes et hypergraphes", Dunod, Paris, 1970.

MOLY—An Interactive System for Molecular Analysis

THOMAS M. DYOTT, A. J. STUPER,* and G. S. ZANDER
Rohm and Haas Company, Spring House, Pennsylvania 19477

Received July 10, 1979

This paper describes an interactive computer system, which is used to assist in the design of biologically active compounds. The system, called MOLY, is capable of: providing rapid construction and manipulation of three-dimensional, chemically correct molecular images; providing visual comparisons between different molecules; performing detailed conformational analysis; estimating a molecule's lipophilicity; and providing a limited description of a molecule's electronic properties. The architecture of MOLY is such that additions and enhancements can be easily implemented.

INTRODUCTION

The last decade has seen considerable progress toward understanding the relationships between chemical structure and biological activity. A number of advances in this area have resulted from the simultaneous growth in sophistication of the models which describe activity and the computers which perform these calculations.

A major advance in modeling has been the development of relations which predict changes in transport properties on the basis of structural variations occurring within a homologous series.[1] The ability to incorporate certain electronic and steric factors into the model has made development of Quantitative Structure Activity Relationships (QSAR) a powerful tool for rationalizing the effects of structural alteration upon biological activity. Unfortunately, such models are only applicable to changes within a homologous series. They cannot be used to indicate the structural requirements for compounds which deviate significantly from the original lead.

The problem of developing new classes of compounds which exhibit a desired activity remains the subject of intensive investigation. Much work remains to be done in understanding the effect structural alterations have on (1) chemical reactivity, (2) physiochemical properties, (3) electronic properties, and (4) steric interactions.

In recent years a great deal of effort has gone into developing theoretical tools which can shed some light on these aspects of the structure–activity puzzle.[2] Unfortunately, knowing any one of these properties alone is insufficient for developing active compounds. All of these properties work synergistically to cause a specific effect or effects.

This paper describes our efforts to develop a computer system which provides an arsenal of sophisticated techniques with which to attack problems in molecular design. This is not a simple collection of programs, but rather an integrated computer system capable of providing sophisticated forms of molecular analysis. The system, called MOLY, is capable of evolving as more and better methods for structure–activity analyses are developed.

Currently MOLY is capable of providing (1) construction and manipulation of three-dimensional chemically correct molecular images, (2) visual comparisons between different molecules, (3) detailed conformational analysis, (4) a limited description of a molecule's electronic properties, and (5) an estimate of a molecule's lipophilicity. The remaining sections of this paper detail the organization of MOLY and our plans for future growth.

SYSTEM OVERVIEW

MOLY is an interactive computer graphics system with the following characteristics: (1) large (≃800K bytes), (2)

0095-2338/80/1620-0028$01.00/0

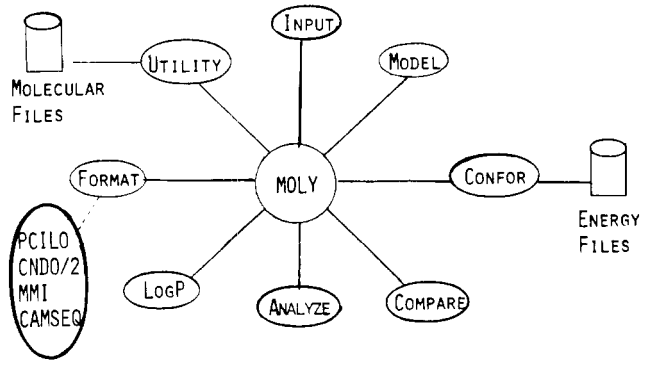


Figure 1. Overview of the MOLY system.

modular, (3) command driven, (4) considerable user prompting, and (5) highly graphics oriented.

Figure 1 shows MOLY's general architecture. The main program, referred to as the MOLY driver, deciphers the user's directions and calls the appropriate modules which then interact with the user to perform a specific task. There are seven major modules whose primary functions are:

1. INPUT – enter molecules by drawing them in via the graphics terminal
2. MODEL – build a reasonable three-dimensional model of a molecule
3. CONFOR – perform conformational analysis
4. ANALYZ – prepare contour maps of energy and population as a function of rotation about one or two bonds
5. LOG P – calculate octanol/water partition coefficient
6. FORMAT – interface with batch programs for CNDO/2 and PCILO quantum mechanical calculations
7. COMPARE – map one molecule onto another using nonlinear least-squares regression

In addition, MOLY contains a large number of utility routines which provide for molecular storage, retrieval, orientation, and display. The largest of these routines, the display routine, is described in a separate section. The other utilities are summarized in Table I.

Two molecules may be in MOLY at any given instant, one (molecule A) in an "active" state and the other (molecule B) in an "inactive" state. The COMPARE and DISPLAY routines will operate on both these molecules; however, all other modules operate solely on the active molecule.

As shown in Figure 1, there are two types of files in the MOLY system: (1) molecular files, which contain molecules in the form of a compact connection table and Cartesian coordinates; (2) conformational energy files, which contain the detailed results of conformational analyses.

MOLY is written entirely in FORTRAN ($\simeq$15 000 lines of code) and is run on an IBM 370/158 under the MVS operating system and the TSO time-sharing system. The user interacts with MOLY via a Tektronix 4006 or 4010 graphics terminal. Hard copy capabilities are provided via Tektronix photocopiers as well as a Calcomp digital plotter.

MOLECULAR INPUT

MOLY provides a convenient mechanism for entering chemical structures for subsequent display and analysis. The scientist draws the structure on the graphics terminal much as one would draw on a blackboard or piece of paper. When entering a structure, the screen is divided into a drawing page (enclosed in a box) and a command menu as shown in Figure 2. Structures are drawn or commands selected from the menu by means of a cursor moved about using thumb wheels or a joystick. Commands, or modes, are selected by pointing at them with the cursor.

Table I. Miscellaneous Utility Functions

MOLY provides the user with a number of additional capabilities which facilitate use of the system. The user can give MOLY the following commands:[a]

STORE	stores a molecule on a file for recall at a later date
GET	retrieves a previously stored molecule from a file
DELM	deletes a molecule from a file
ORNT*	orients the molecule in space according to user specifications
MROT*	rotates the molecule about the *X*, *Y*, and /or *Z* axes
BROT*	change the conformation of a molecule by performing rotations about acyclic bonds
SETC*	sets the molecule into a conformation specified in terms of torsional angles
REFG*	allows the user to modify the geometry by specifying new bond lengths, bond angles, and torsional angles
INFO*	answers user questions about interactomic distances, angles, and torsional angles
ADDH	adds hydrogen to the molecule in the proper positions
DELH	deletes all hydrogens from the molecule
MIRR	reflects the molecule through the *XY* plane to create its mirror image (enantiomer)
INVE	inverts specified stereocenters to create different stereoisomers
SWAP	exchanges the "active" and "inactive" molecules
CPU*	reports the amount of computer CPU time used
DPGM*	displays the cartesian and bond lengths and angles geometry tables
DPCT*	displays the connection table

[a] Entries marked with an asterisk denote that these commands are also available to the user when in the MODEL and/or CONFOR modules.

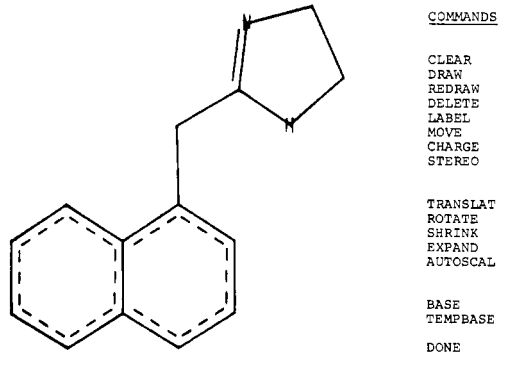


Figure 2. Molecular input "drawing page".

The normal input sequence is: (1) clear the drawing page of any unwanted structure (unless the scientist wants to modify the current molecule); (2) enter the DRAW mode and draw the molecular skeleton, indicating the appropriate bond types (i.e., 1 = single, 2 = double, 3 = triple, 4 = aromatic); (3) enter LABEL mode and specify the heteroatoms by pointing at them with the cursor and typing in the appropriate chemical symbol (all 103 elements are recognized); (4) if necessary, enter CHARGE mode to specify formal charges. Hydrogens can be explicitly entered, but are normally omitted since a utility routine (ADDH) can be used to automatically add them.

Mistakes can be easily corrected. Atom types, formal charges, and configurations can simply be relabeled correctly. Bond types can be changed by redrawing the bond and specifying the correct type. If necessary, the DELETE mode will delete bonds and/or atoms by pointing at them with the cursor. Additionally, individual atoms may be moved by entering MOVE mode, pointing at an atom, and then pointing at a new position.

If a drawing is made too large, thereby running out of room on the drawing page, five commands (TRANSLATE,

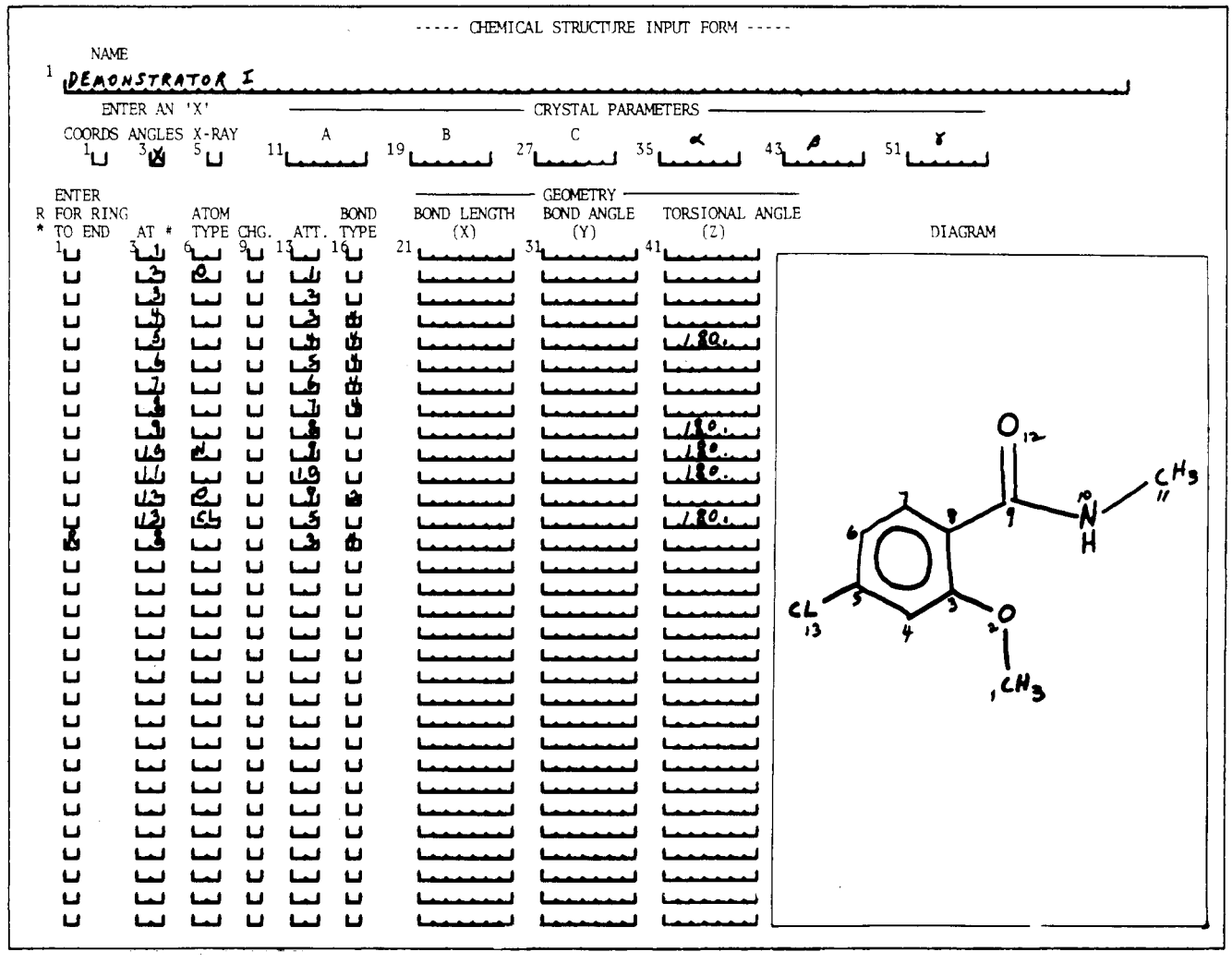


Figure 3. An example of a chemical structure input form.

SHRINK, EXAND, AUTOSCALE, and ROTATE) are available to allow the structure to be readjusted so that drawing may continue within the confines of the page.

An additional convenience feature is available to facilitate the entry of molecules having a common framework. MOLY has two base or template libraries for this purpose. One library is permanent and contains a large number of common bases, e.g., the steroid and diphenyl ether skeletons. The other is temporary, existing only during a particular MOLY session. Any number of bases may be entered into the temporary library by simply drawing each base and pointing at the TEMPBASE command. In order to retrieve a base from either library, one points at the BASE command and types in the name of the desired base. The retrieved base can then be modified to create the desired structure.

When the drawing is complete the DONE command is selected. INPUT inspects the molecule for possible chiral centers. Each such center is arbitrarily placed into either the *R* or *S* configuration and the molecule is redrawn on the screen with the appropriate wedged or slashed bonds. The scientist is forced to inspect each center and either approve or invert the specified configuration.

Upon completion of structural imput, the INPUT module decodes the structure and creates a number of representations: (1) a redundant connection table, (2) a spanning tree, (3) Cartesian coordinates, and (4) bond length, bond angle, and torsional angle representation.

MOLY is capable of interconverting the Cartesian coordinate and bond length, bond angle, and torsional angle representations. This allows structural manipulation to be performed in the coordinate system most appropriate to the task.

Molecular input may also be accomplished through batch mode. An input form (Figure 3) is available for specifying molecular structure to the batch routine. Structures may be specified in terms of (1) *XYZ* coordinates; (2) bond lengths, bond angles, and torsional angles, and (3) unit cell coordinates. If the bond length and bond angle option is selected, any values of bond length and bond angle which are omitted will default to the appropriate standard value. This batch mode is most useful for entering crystal structures into the system.

MOLECULAR MODELING

When a molecule has been entered via the graphical input module it will be accurate in terms of connectivity. The geometry, in terms of bond lengths, bond angles, and torsional angles, is rather crude since it was merely calculated from the sketch. The MODEL module will refine the geometry in order to obtain a three-dimensional model suitable for quantum mechanical calculation and/or conformational analysis. The scientist can make use of any or all of three different mechanisms for specifying geometry.

The first option is to request that standard bond lengths and angles be used. MOLY has an extensive set of standard bond lengths built into it. If, however, it does not have the appropriate standard value, it will use a "best guess" value and inform the scientist of the problem. The scientist can then accept the suggested value or specify another value. The bond

angles used are the "ideal": 109.5° for sp^3 hydridization, 120° for sp^2, and 180° for sp.

The second option is to specify the precise bond lengths, bond angles, and torsional angles desired. This provides considerable flexibility, allowing the specification of any "nonideal" geometry desired. This is frequently used, for example, in peptide work where it is known that the bond angles around the amide linkage may differ from the "ideal" 120° by as much as 4–5°.

The third option is to submit all or some portion of the molecule to a classical mechanical model building routine. This is frequently the only reasonable option for handling ring systems. (It is a fortuitous accident of nature that the phenyl ring works out very nicely using just standard bond lengths and angles.)

Our model-building routine is a descendant of the one developed by Wipke et al.[6] The atoms are moved about in order to minimize the total strain in the system. In this way it produces realistic models of strained systems where the bond lengths and angles are reasonable compromises.

Total strain is defined as

$$E_{total} = E_{bond} + E_{angle} + E_{nonbonded} + E_{tor} + E_{config} + E_{hyb}$$

where E_{bond} is bond-length strain, E_{angle} is bond-angle strain, $E_{nonbonded}$ is nonbonded interaction (van der Waals) strain, E_{tor} is torsional strain, E_{config} is configuration strain, and E_{hyb} is atom hydridization strain. The bond-length and bond-angle strain terms are calculated using standard Hooke's law type functions. The nonbonded term is a normal "6–12" Lennard-Jones potential function. The torsional function tries to stagger butane-type interactions, keep double bonds, amides, and esters planar, keep allenes perpendicular, etc. The configuration term makes sure the model has the configuration the scientist specified by making alternative configurations appear to have a very high strain energy. Finally the hybridization term, replacing the usual "flatten" function,[6,7] is used to keep strained carbonyls flat and to automatically increase the angle between the appendages on strained rings.

Such a function, properly parameterized could, and probably should, replace the traditional bond-angle function as well.

The scientist can specify that any of the three techniques be applied to the entire structure or any portion of it. We find that this combination of features generally enables us to build the desired three-dimensional model rapidly.

CONFORMATIONAL ANALYSIS

Conformational analysis is performed by the module CONFOR. An analysis is defined by specifying the bonds to rotate, the increment of rotation, and the total number of degrees to rotate. On the basis of this information, the program calculates the number of conformations which would result. If the problem is too large to be done interactively (>5 CPU min), the problem can be simplified or CONFOR can be used to initiate a background batch job which will perform the analysis. In either event, once an analysis is complete the scientist can ask for the low-energy conformations, look at energy contour maps, place the molecule in various conformations for visual inspection, etc.

Conformational analysis presents a combinatorial problem. An analysis involving 360° rotations in D degree increments about N different bonds requires the examination of $(360/D)^N$ conformations. The number of conformations increases exponentially with the number of bonds being rotated. For example, if D is 30° and N is 2, then only 144 conformations must be examined, but if N is increased to just 5, the number of conformations jumps to 248 832. The sheer magnitude of such a problem rules out even the fastest quantum mechanical technique. Empirical methods are the only practical means of examining such a large number of conformations. Even with an empirical model, we felt we should take a number of steps to assure the efficiency of our algorithm.

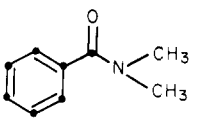
In CONFOR, conformational energy is broken into four terms:

$$E_{conf} = E_{vdw} + E_{coul} + E_{hbond} + E_{conj}$$

where E_{conf} is the relative energy of the conformation; E_{vdw} is the van der Waals term, accounting for steric interactions; E_{coul} is the coulombic interaction term, accounting for partial charge interactions; E_{hbond} is the hydrogen bond term; E_{conj} is the conjugated bond term accounting for torsional preferences due to conjugation.

The exact mathematical form of each energy term is given in Table II. E_{vdw} is the standard "6–12" Lennard-Jones potential function. We have taken our repulsive and attractive constants from the literature.[8] E_{coul} is the classical coulombic potential function. The dielectric constant of the solvent, ϵ, can be specified by the scientist; otherwise it defaults to a value of 3.5. E_{hbond} is the hydrogen bond potential function developed by Scheraga et al.[9] This function has no angular dependence, unlike the function developed by Hopfinger et al.[10] which in our experience is overly restrictive.

The E_{conj} term was developed by us in the course of this work. It is an attempt to take into account the torsional barriers caused by interaction of adjacent π systems and/or lone pairs. Such systems, e.g., amides, benzoic acids, and dienes, show a preference for planar conformations which maximize conjugation. The basic form of the function is the standard cosine relationship frequently used for torsional energy:[11] $E_{conj} = B[1 - \cos(N\theta)]$ where B is half the barrier height, N is a symmetry constant (usually 2), and θ is the torsional angle. E_{conj} has a value of 0.0 for the preferred value of θ and increases in a smooth sinusoidal manner to a maximum of $+2B$ as θ deviates from the preferred values. If, however, either or both of the systems in conjugation are further conjugated as is, for example, a benzamide, where the carbonyl is conjugated with both the amide nitrogen and the phenyl ring, the effective barrier is decreased via two mech-



anisms: (1) The energy gained from conjugation initially is less; i.e., in the case of the benzamide, since the carbonyl is conjugated with the phenyl it cannot conjugate as strongly with the nitrogen. (2) Energy lost by the reduction of conjugation due to rotation about one bond can be partially recovered by increased conjugation with the other conjugated groups.

This effect, however, is dynamic. For example, as the phenyl ring rotates out of conjugation with the amide the torsional barrier about the amide bond returns to full strength. Thus the effective barrier to rotation about a conjugating bond is dependent on: (1) the "normal" barrier for such a bond, (2) any additional conjugated groups, (3) their conjugative strength, and (4) their relative conformation.

CONFOR automatically takes all of these factors into account. It initially finds all conjugating bonds and assigns their "normal" barrier values and symmetry constants. These values are shown to the scientist who can specify any desired alternative values. As the molecule is being stepped through various conformations, the effective torsional barriers of any rotated conjugating bonds are recalculated and then used to calculate E_{conj} via the formulas given in Table II.

In order to reduce the cost of conformational analyses without sacrificing accuracy, we implemented techniques which

Table II. Mathematical Form of Potential Functions

$$E_{vdw} = [B_{ij}/d^6 - A_{ij}]/d^6$$

where B_{ij} = repulsive constant dependent on atoms i, j
A_{ij} = attractive constant dependent on atoms i, j
d = Euclidian distance between atoms i and j

$$E_{coul} = \frac{332.0}{\epsilon}\frac{|q_i \cdot q_j|}{|d|}$$

where ϵ = dielectric constant
q_i = partial atomic charge on atom i
d = Euclidian distance between atoms i and j

$$E_{hbond} = E[r_o/r_{HX}]^{12} - 2.0E[r_o/r_{HX}]^{10}$$

where E = energy constant dependent on type of atoms participating in the hydrogen bond
r_o = internuclear distance parameter dependent on type of atoms
r_{HX} = Euclidian distance between donor and acceptor atoms

$$E_{tor} = B_k'[1 - \cos(N\theta_k)]$$

$$B_k' = (B_k/2)\left[1 - \sum_{\substack{i=1 \\ i\neq k}}^{n_k} (B_i[1 + \cos(N\theta_i)]/\sum_{j=1}^{n_k} B_j)\right]$$

where B_k' = the effective barrier to rotation about the bond of interest
B_k = the nonconjugated barrier to rotation about the bond of interest
B_i = the nonconjugated barrier to rotation about the ith conjugated bond
θ_k = the torsional angle for the bond being rotated
θ_i = the torsional angle for the conjugate bonds
n_k = number of bonds conjugated with bond k
N = a symmetry factor

(1) detect and eliminate a large percentage of high-energy conformations without having to calculate them, (2) eliminate the recalculation of interactions which are unaffected by the various rotational changes, and (3) perform bond rotations in an efficient manner. We succeeded to such a degree that we routinely conduct conformational analyses in a time-sharing mode. For molecules of the size we are normally evaluating (30–70 atoms), CONFOR operates at a speed of 20–80 conformations examined/CPU second (IBM 370/158 MVS). Since CONFOR is able to skip the calculation of a large percentage of all high-energy conformations, its effective speed is usually in the range of 40–400 conformations/CPU second.

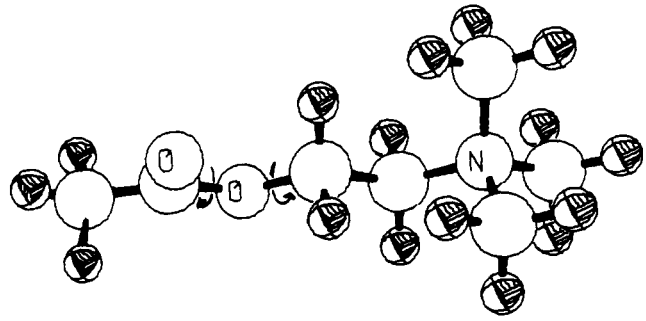


Figure 4. Acetylcholine: bonds with arrows were rotated in 15° increments.

In order to compare CONFOR's speed to that of other programs and techniques, let us consider the example of acetylcholine. If we rotate in 15° increments around the two bonds indicated in Figure 4, there are a total of 576 conformations. Table III contains the comparative information.

Table IV contains some performance information for a number of actual analyses we have run recently. The number of conformations actually calculated per second ranges from 29 to 240. However, because of CONFOR's ability to rule out large numbers of high-energy conformations without calculating them, its effective speed ranged from 37 to 16 589 conformations/second.

A detailed analysis of the accuracy of this conformational analysis module is well beyond the scope of this paper, and, in fact, is the subject of a subsequent paper.[16] However, we would like to indicate that our studies lead us to believe that CONFOR is at least as accurate as any other conformational analysis technique. We have compared CONFOR's results with crystal structures, NMR data, and quantum mechanical

Table III

technique	program	computer	CPU sec	cost at $15/CPU min
quantum mech	GAUSSIAN 70[12]	IBM 370/158	2 073 600	$518 400
quantum mech	CNINDO[13]	IBM 370/158	68 704	17 176
quantum mech	PCILO[14]	IBM 370/158	8 640	2 160
empirical	CAMSEQ[15]	IBM 370/165[a]	203[b]	51
empirical	MOLY	IBM 370/158	8[b]	2

[a] An IBM 370/165 is estimated to be about 1.5–2.0 times as fast as an IBM 370/158. [b] Does not include the time required for initially calculating the partial charges (~120 s).

Table IV. Performance Information from 16 Actual Analyses

no. of atoms	no. of bonds	CPU, s	no. possible	no. poss/s	% skipped	no. calcd	no. calcd/s
22	3	24	6 912	288	24.7	5 207	217
28	3	21	3 888	185	11.9	3 426	163
33	3	155	15 552	100	26.0	11 504	74
35	3	201	11 664	58	20.3	9 300	46
35	6	180	2 985 984	16 589	98.6	43 220	240
36	4	292	20 736	71	44.4	11 529	39
40	5	34	20 736	610	90.9	1 890	56
41	5	1434	125 416	87	59.2	50 716	35
41	4	428	41 472	97	70.2	12 354	29
44	5	2480	248 832	100	69.6	75 758	31
45	4	111	20 736	187	53.8	9 578	86
46	6	70	27 648	395	89.7	2 835	40
49	5	380	746 496	1 964	98.0	15 112	39
49	5	56	82 944	1 487	96.0	3 282	59
50	4	2142	82 944	39	12.7	72 359	34
50	3	186	6 912	37	0.0	6 912	37
min							
22	3	21	3 888	37	0.0	1 890	29
max							
50	6	2480	2 985 984	16 589	98.6	75 758	240

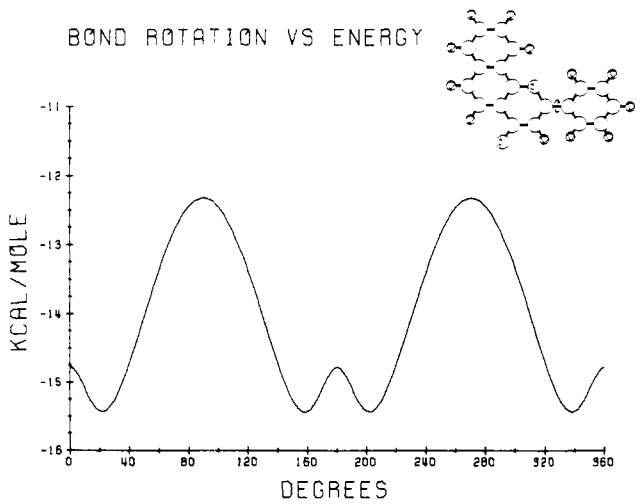


Figure 5. Example of a plot of energy vs. bond rotation.

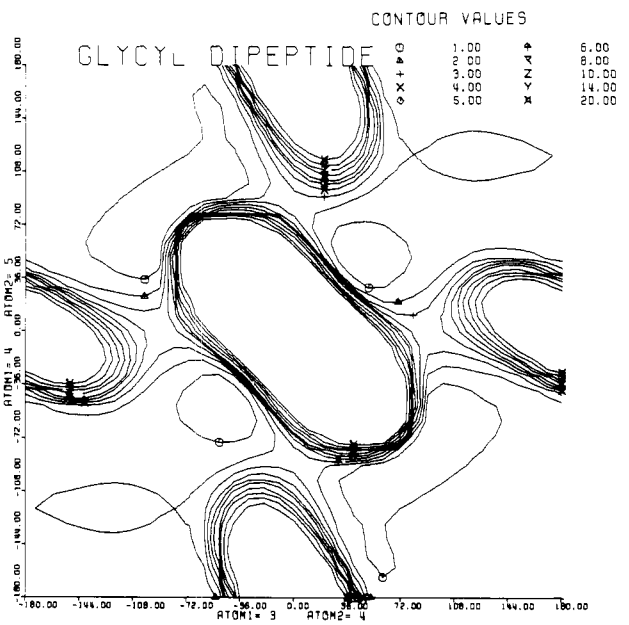


Figure 6. Example of a contour plot of conformational energies.

calculations (primarily PCILO) and found no significant discrepancies.

CONFORMATIONAL ENERGY AND POPULATION CONTOUR MAPS

In order to facilitate the interpretation and presentation of the results of a conformational analysis, MOLY's ANALYZ module can provide (1) lists of the energy of all conformations examined, (2) lists of all conformations within a specified number of kcal/mol of the minimum, (3) graphs of energy as a function of rotation about a specified bond (Figure 5), and (4) contour maps of energy and populations as a function of rotation about two specified bonds (Figures 6 and 7). When requesting these lists, graphs, and maps, the scientist specifies the bond or bonds of interest. Other bonds which were rotated during the analysis are then specified to be either "fixed" at a specific angle or "free" to assume the most favorable orientation. Relative conformational populations are calculated via the Boltzmann distribution; thus the relative population for a given conformation, p_i, is defined as:

$$p_i = \frac{e^{-(E_i - E_0)/RT}}{\sum p_i}$$

where E_i is the energy of conformation i, E_0 is the energy of the lowest energy conformation, R is the gas constant

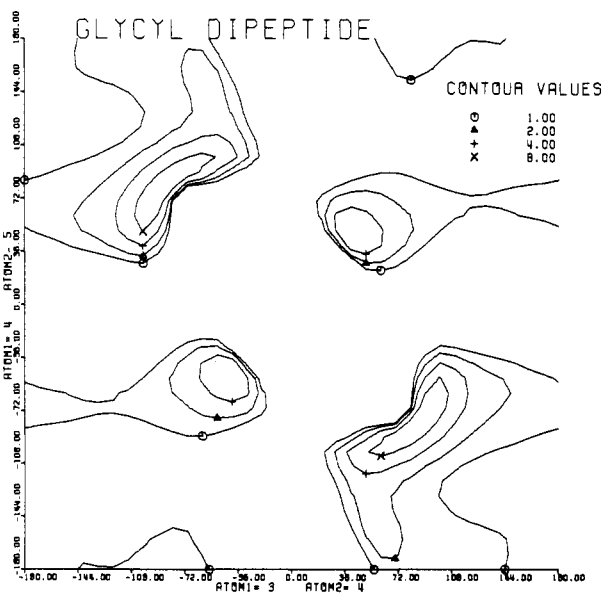


Figure 7. Example of a contour diagram of conformer populations.

(0.0019872 kcal mol^{-1} deg K^{-1}), T is the temperature in degrees K, and $\sum p_i$ is a normalization factor. The populations are then scaled so that they are relative to random; i.e., a value of 2 indicates that a conformation is twice as probable as random. Thus a totally flat energy surface would result in a value of 1 at every conformation, while a highly constrained molecule might have conformational regions where the population is 100–200 times more likely than random.

These aids provide not only a concise summary of an analysis, but also facilitate comparison between a number of analyses.

DISPLAY

The heart of any interactive graphical molecular analysis system is its ability to display molecules. MOLY contains a versatile, easily used routine for displaying molecular structures. Visual analysis provides a means of comparing low-energy conformations of molecules, understanding the spatial arrangement of functional groups, and developing an understanding of the features required for interaction with the receptor site. The DISPLAY module provides a multitude of features which greatly facilitate these processes.

Like the bulk of MOLY modules, DISPLAY is command driven. The type of display is selected by means of keyword options, which remain in effect until explicitly turned off. Five types of options are available to control: (1) the structural model, (2) the degree of structural detail, (3) the viewing mode, (4) the display of either one or two molecules, and (5) the orientation of the molecule(s). Various combinations of the options available in the first three categories alone result in nearly 800 different types of displays.

Three types of structural models are available: simple stick figures, space-filling models, and space-filling with imbedded stick figures. The space-filling models are created by a version of Warme's routine[3] which we modified so the spheres are in proportion to each atom's van der Waals' radius. The spheres can be set to any fraction of the van der Waals' radius.

The degree of structural detail is controlled by a series of options which display or prohibit the display of:

1. hydrogen atoms
2. heteroatom labels (carbons are never explicitly labeled)
3. multiple bonds (═, ≡, and ⇌)
4. stereo bonds (◁ and ···)
5. atom numbers
6. a perspective view of the structure (a view indicating depth of field)

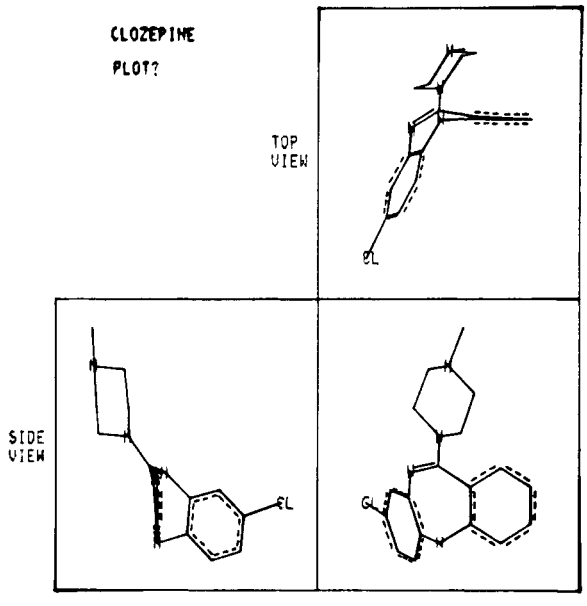


Figure 8. An example of three perpendicular views.

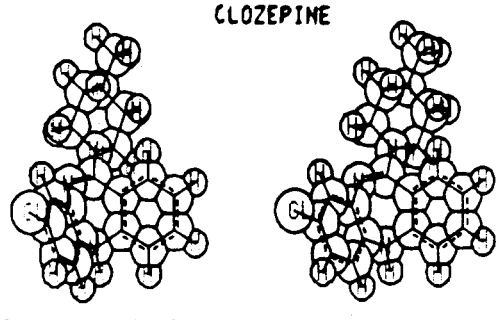


Figure 9. An example of stereopair and space-filling drawings.

Three viewing modes are available: (1) simple *XY* plane projection; (2) three perpendicular (*XY*, *XZ*, and *YZ*) projections (Figure 8); and (3) stereo views for a three-dimensional effect (Figure 9). Two types of stereo views are available: one, the traditional stereo pair, is used with the Taylor-Merchant Stereopticon 707;[4] and the other, a modified stereo pair in which the image on the left has been reflected, is used with the AMSOM stereo box.[5] The depth effect in either case is quite dramatic.

The three viewing modes can be used to display either one or two molecules simultaneously. The molecules can be viewed side by side, or superimposed. An example of superimposing two structures is shown in Figure 10.

The user controls the molecule's orientation by means of rotate, translate, and orient options. The rotate options allows specific *X*, *Y*, or *Z* rotations to be accomplished. The orient option allows repositioning of the molecule through specification of atoms which are to be at the origin, on the *X* axis, and in the *X*, *Y* plane. Appropriate use of the orient and rotate options provides views from any directions or along any bond. Orientation of two molecules may be made separately or in tandem, as the situation warrants.

Copies of the display screen can be made anytime through use of the hard copy device attached to the graphics terminal. High-quality digital plots varying in size from 3.5 in. to 11 in. square can also be provided.

LOG *P* PREDICTION

In order to provide an estimation of a molecule's transport properties, we have incorporated a program which estimates

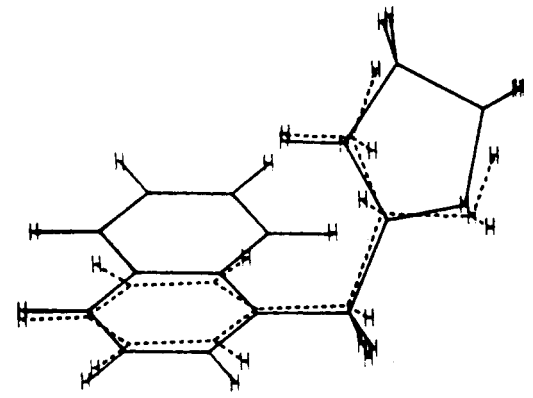


Figure 10. An example of molecular comparisons: amphetamine superimposed on naphazoline. Total distance of 1.23 Å between six pairs of atoms.

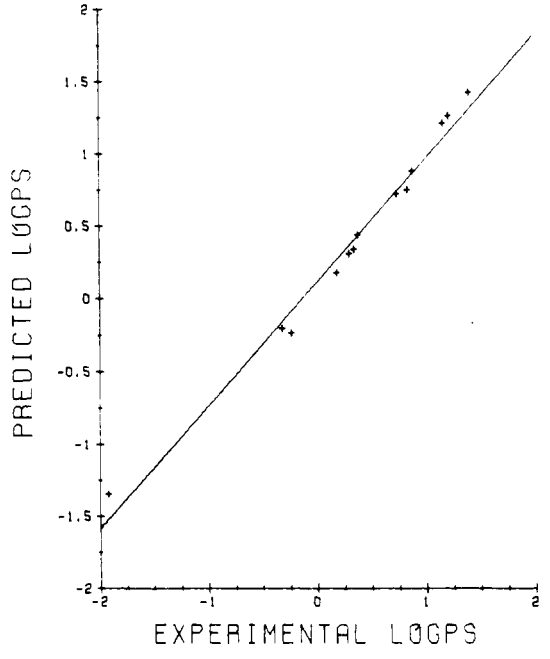


Figure 11. Plot of measured vs. predicted log *P* values for 13 different alcohols, esters, and ethers.

the partition coefficient between 1-octanol and water. The program is a version of the one developed by Chou and Jurs.[17] It develops a prediction of log *P* using the fragment constant method developed by Leo.[18]

The program predicts log *P* directly from the structure of the molecule. Thus, to predict log *P* the scientist need only draw the molecule using the input module of MOLY, or recall the molecule if it has previously been input. Figure 11 shows a graph of the measured vs. predicted values for 13 different alcohols, esters, and ethers. We have found the program to provide reasonable estimates of this parameter for more complex structures with which we deal.

COMPARE

We are frequently interested in comparing the structures of two molecules which possess similar biological activities but different chemical structures. Such comparisons can lead us to a better understanding of the steric requirements of the active site. The COMPARE module provides a mechanism for making such comparisons. COMPARE allows specific atoms of molecule A to be mapped onto specific atoms of molecule B. A nonlinear technique[19] is then used to rotate and translate molecule A in order to minimize the distances between the matched atom pairs. COMPARE reports back

the sum of the distances between the atom pairs.

In order to provide for a better fit, the user may also specify some conformational flexibility on the part of either or both molecules. Up to a total of ten bonds may be specified as having some rotational freedom, by indicating the allowed range of torsional values for each bond. In addition to rotation and translation of the entire molecule, COMPARE will rotate about these specified bonds in an attempt to effect a better fit between the matched atom pairs.

A comparison between two molecules may appear to be very poor because they are in opposite enantiomeric forms. In order to avoid this problem, COMPARE asks if the enantiomer of A should also be compared with B. If that is desired, A is inverted and again compared to B. The enantiomer that results in the better fit is automatically kept. COMPARE actually repositions the molecules so that the comparison can be displayed, as shown in Figure 10.

FORMAT

We attempt to make full use of molecular analysis programs which are available from various outside sources, such as the Quantum Chemistry Program Exchange. However, since these programs were developed by different research groups around the world, two major problems exist: (1) while the programs all deal with molecules, each has its own unique format for specifying them; (2) the input to some of these programs is particularly cumbersome, taking as much as 2 to 3 h to encode a moderately sized (40 atom) molecule.

Several years ago we adopted a standard molecular representation for our internally developed molecular analysis programs, including MOLY. Since then we have developed convenient programs for producing this molecular representation from a variety of inputs (cards, graphics terminal, our internal compound data base, and the Chemical Abstracts Service Registry System). We solved the problems of structure input into these external programs by creating a module, called FORMAT, which would accept molecules in our standard format and produce the appropriately formatted card decks. FORMAT will produce, on command, the appropriate decks for:

1. CNINDO — CNDO/2 and INDO quantum mechanical calculations[13]
2. PCILO — PCILO quantum mechanical calculations[14]
3. CAMSEQ — conformational analysis[15]
4. MINMZ — classical mechanical molecular model builder[6]
5. ORTEP — sophisticated "ball and stick" plots[20]

The FORMAT module can punch out the appropriate card decks or, if desired, initiate a background batch job. FORMAT even estimates the computer time required and checks for a number of common errors, e.g., too many atoms or orbitals or unsupported elements.

CONCLUSION

The molecular analysis system (MOLY) we have developed provides us with a number of new tools for attacking structure–activity studies. MOLY's interactive nature gives the scientist full and immediate control of these tools. MOLY's modular structure will make it relatively easy for us to incorporate additional capabilities as they become available.

REFERENCES AND NOTES

(1) C. Hansch, E. W. Deutsch, and R. N. Smith, *J. Am. Chem. Soc.*, **87**, 2738 (1965).
(2) (a) G. Redl, R. D. Cramer, and C. E. Berkoff, "Quantitative Drug Design", *Chem. Soc. Rev.*, **3**, 273 (1974); (b) "Structure-Activity Relationships", C. J. Cavallito, Ed., Pergamon, Oxford, 1973; (c) Y. C. Martin, "Quantitative Drug Design. A Critical Introduction", Marcel Dekker, New York, 1978; (d) A. J. Stuper, W. E. Brugger, and P. C. Jurs "Computer Assisted Studies of Chemical Structure and Biological Function", Wiley, New York, 1979.
(3) P. K. Warme, *Comput. Biomed. Res.*, **10**, 75 (1977).
(4) Taylor-Merchant Corp., 25 W. 45th St., New York, N.Y. 10036.
(5) Marketed by Tracor Jitco, Inc., 1776 E. Jefferson St., Rockville, Md. 20852.
(6) W. T. Wipke, P. Gund, J. M. Verbalis, and T. M. Dyott, Abstracts, 162nd National Meeting of the American Chemical Society, Washington, D.C., Sept., 1971, No. ORGN-17.
(7) N. L. Allinger, J. T. Sprague, and T. Liljefors, *J. Am. Chem. Soc.*, **96**, 5100 (1974).
(8) A. J. Hopfinger, "Conformational Properties of Macromolecules", Academic Press, New York, 1973.
(9) R. F. McGuire, F. A. Momany, and H. A. Scheraga, *J. Phys. Chem.*, **76**, 375 (1972).
(10) H. J. R. Weintraub and A. J. Hopfinger, *J. Theor. Biol*, **41**, 53 (1973).
(11) J. P. Lowe, *Prog. Phys. Org. Chem.*, **6**, 1 (1968).
(12) W. J. Hehre, W. A. Lathan, R. Ditchfield, M. D. Newton, and J. A. Pople, *QCPE*, **10**, 236 (1973).
(13) P. A. Dobosh, *QCPE*, **10**, 141 (1969).
(14) P. Claverie, J. P. Daudey, S. Diner, Cl. Geissner-Prettre, M. Gilbert, J. Langlet, J. P. Mabrieu, V. Pincelli, and B. Pullman, *QCPE*, **10**, 220 (1972).
(15) R. Potenzone, Jr., E. Cavicchi, H. J. R. Weintraub, and A. J. Hopfinger, *Comput. Chem.*, **1**, 187 (1977).
(16) A. J. Stuper, T. M. Dyott, and G. S. Zander in "Computer Assisted Drug Design", ACS Symposium Series, American Chemical Society, Washington, D.C., 1979.
(17) J. T. Chou and P. C. Jurs, "Computer-Assisted Computation of Partition Coefficients from Molecular Structures Using Fragment Constants", *J. Chem. Inf. Comput. Sci.*, **19**, 172 (1979).
(18) A. Leo, P. Y. C. Jow, C. Silipo, and C. Hansch, "Calculation of Hydrophobic Constant (Log *P*) from π and *f* Constants", *J. Med. Chem.*, **18**, 865 (1975).
(19) J. P. Chandler, *QCPE*, **10**, 307 (1976).
(20) C. K. Johnson, "ORTEP II, Oak Ridge National Laboratory, Oak Ridge, Report ORNL-3704, 1971.