

WIZARD: Applications of Expert System Techniques to Conformational Analysis. 1. The Basic Algorithms Exemplified on Simple Hydrocarbons

DANIEL P. DOLATA* and ROBERT E. CARTER

Organic Chemistry 2, Chemical Center, S-221 00 Lund, Sweden

Received April 17, 1986

WIZARD is a program that assists in the exploration of the conformational space of molecules and is based on the artificial intelligence techniques known as expert system methods. The purpose of WIZARD is to provide preliminary estimates of the geometries and energies of likely conformers. These conformers can then be minimized by using any existing molecular mechanics program. The feasibility of this approach was first demonstrated on saturated acyclic hydrocarbons. It has shown that, within this limited domain, the combination of WIZARD and Allinger's molecular mechanics program (MM2) leads to results similar to those obtained by the proper utilization of the bond-driver option in MM2. However, unlike the bond-driving technique, WIZARD can handle up to seven degrees of freedom with relative ease and much greater speed. In addition, this method is general and is being extended to include unsaturation, cycles, and heteroatoms. Deficiencies and possibilities of the current program are discussed.

INTRODUCTION

Current techniques for understanding the conformational freedom of a flexible molecule have some strong limitations. If the chemist attempts to generate the possible conformers "by hand" and then minimize them using an existing molecular mechanics (MM) program, it is both time consuming and difficult to feel confident that all of the local minima have been explored. The process can be automated by driving one or more dihedrals,¹ but the computer time required increases exponentially with the number of degrees of freedom and thus restricts the use of this technique to molecules with a small number of degrees of freedom. For example, to exhaustively explore the conformational space of a molecule with six degrees of freedom by 20° increments would require the evaluation of over 34 million conformations. In addition, the dihedral-driving method has some inherent problems,^{1,2} which can lead to erroneous results in the hands of a naive user. Some programs have been written to explore the conformational space for special types of molecules,³ especially for rings,^{3a,b} but these programs do not provide a general solution. A more general approach has become popular recently, as exemplified by Cohen's elegant SCRIPT program⁴ and De Clercq's SCA program,⁵ but these programs are not totally general or exhaustive. At least one program⁶ has attempted to provide information about the rigidity of a conformer, but only in a limited domain and for a specialized application. Thus, it seemed there was a need for a general program that could address these tasks. This paper reports on our pilot project, which is known as the WIZARD program. Specifically, we will discuss two separate versions of WIZARD, the first based on standard expert system techniques and the second on a more sophisticated approach.

Observation of the interactions between experienced conformational analysts and a molecular mechanics program coupled with a familiarity with artificial intelligence from another domain led us to believe that the application of expert systems techniques to conformational analysis might prove fruitful. For example, one of the projects at the University of Lund involves the analysis and synthesis of moth pheromones and pheromone mimics.⁷ One step in the analysis of these compounds was to construct different compounds whose lowest energy conformers had similar geometrical relationships between functional groups. Since pheromones are highly flexible molecules with a large number of degrees of freedom, exhaustive conformational analysis by existing methods would require prohibitive amounts of computer time. However, it was found that an experienced analyst could guess the geometries of the low-energy conformers well enough so that a

molecular mechanics program would produce a final conformation of similar shape upon minimization. The chemist presumably did this by applying general rules of conformational analysis to the molecule at hand. Since we were interested in the final conformations and strain energies of the conformations, we felt that the combination of a program that utilized such rules to provide a good set of such guesses and the use of an accepted molecular mechanics program to refine these guesses would provide us with the answers we needed. This approach had the advantage that it would allow us to focus our energies on the first program, and we would not have to expend the effort required to recreate the existing MM programs. Since MM2⁸ was available in-house with considerable experience in its use and construction, this was chosen as our default force field program.

The desired program would thus be one that, given some representation of the molecule, would produce conformations sufficiently close to the local minima so that MM2 could easily bring each one to its respective minimum. Since we had seen how successful human chemists were at making such reasoned guesses, we decided to follow this technique and to use logic and high-level reasoning to infer the structures of these conformers instead of some strictly numerical method. This is the approach of the so-called expert system (ES) programs that have recently been developed by researchers in the field of artificial intelligence (AI).⁹ The WIZARD program is the result of this work, and it currently reasons within the domain of saturated acyclic hydrocarbons. Although cycles and heteroatoms may be included, WIZARD does not have the knowledge to manipulate these entities properly as yet and so leaves these sections of the molecule in the input conformation. Before discussing the results or design of WIZARD, it will be instructive to briefly describe how the user sees WIZARD.

OVERVIEW OF WIZARD

WIZARD works as an intermediate between the chemist's input and a conformational analysis program (exemplified by MM2), as shown in Figure 1. WIZARD itself has no graphical input, but relies on data files prepared by use of some other graphical input program. For this purpose, we have employed MOLBUILD¹⁰ and RINGS,¹¹ both of which are part of the MIMIC system.^{10,11} The force field program commonly used is Allinger's MM2 program. WIZARD communicates with MM2 via data files that contain the Cartesian coordinates, connectivity, and atom types for each conformer. While it is possible to directly run MM2 as a subtask of WIZARD, the programs have been left separate to allow the chemist more control over time

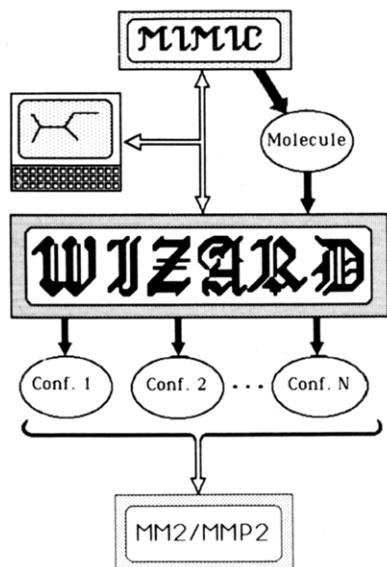


Figure 1. Schematic representation of the relationship of WIZARD to existing programs.

and resource allocation. In a similar fashion, either of the input programs may be run as a subprocess of WIZARD itself.

Once WIZARD has been initialized, it is given the name of a molecule file. This is then read in, patterns of atoms and bonds of which WIZARD has knowledge are recognized, and the molecule is then displayed on a PLOT10-compatible terminal.¹² The user then selects the type of conformational analysis that is to be performed. The user may choose an exhaustive analysis in which even *undesirable conformers* (vide infra) will be created and displayed, or a restricted analysis, which detects these undesirable conformers during generation and eliminates them as early as possible, may be chosen. As will be discussed later, the choice depends on the solution desired and the type of molecule.

WIZARD then proceeds to generate each conformer and attempts to provide an estimate of the relative strain energy. The chemist may choose to see each conformer as it is generated, or viewing may be deferred until after all conformers have been generated. In the latter case, WIZARD presents a simple histogram of the distribution of the conformers and their energies to help the chemist decide how many, and which, conformers are to be displayed. As each conformer is displayed, the chemist has the option of manipulating the display to examine the conformer, saving the conformer for molecular mechanics, or continuing the analysis. The manipulation includes the ability to introduce 3D cueing, to rotate the molecule, or to request a PLUTO plot.¹³

Once an analysis is completed, the chemist may review the conformers generated or may request that certain conformers be *resolved*. This process is described in detail later in the paper but may briefly be described as the process of allowing strained conformers to achieve geometries with lower energies. Since these conformers are generally of significantly higher energy than nonproblematic conformers, the chemist may not wish to perform resolution if such problem-free conformers exist. It is for this reason that resolution is not performed automatically.

GENERAL DESIGN PRINCIPLES FOR WIZARD

As noted above, it seemed that the generation of conformers for refinement by a MM program would be tractable by the use of ES techniques. The definition of an ES seems to be rather flexible at present, but ideally an ES is a program that performs like an expert in a given task domain. All ES's are examples of knowledge-engineering tasks. The program

contains a body of knowledge extracted from some "expert" source and a set of tools for querying and utilizing this knowledge. The difference between a data-base program and an ES program is that the knowledge in an ES is presented in such a fashion that it can be applied to previously unknown problems. Thus, any program that attempts to predict results for the analysis of an as yet unknown system, or one which the program's knowledge base does not explicitly contain, is some sort of an ES if it does as well as a human expert would. However, an additional constraint seems to be that ES's use symbolic reasoning instead of numerical methods. Quantum mechanics programs clearly outshine human experts at giving the energies of complex systems but nevertheless would not be classified as ES programs. Since the construction of an ES is generally divided into two main steps, the acquisition of knowledge and its representation in a program, we will discuss these two steps in that order.

The information that was to be encoded into our knowledge base was obtained in several ways. Perhaps one of the most important has been through interaction with experts in the field of conformational analysis. Thus, early in the implementation of WIZARD, one of the authors spent several hours with Prof. J. Edgar Anderson of University College, London, and attempted to derive a set of rules for simple acyclic hydrocarbons. This session resulted in several pages of tentative rules and many questions that needed to be resolved by reference to a molecular mechanics program. One of the original rules that was proposed was "Place everything on a diamond lattice, and see if it fits without problems. If it does, then quit, otherwise continue." Of course, rules had to be formulated as to what constituted a problem. The well-known pentane rule, i.e., that a C₅ chain with a g+/g- pair (dihedral angles of $\pm 60^\circ$) is much less stable than a conformer lacking such a pair, was set forth. Other rules were proposed as to what actions might be taken if the molecules could not lie along the diamond lattice without problems.

Allinger's MM2 program was chosen as another expert. About 100 molecules, ranging from butane to hexane and from hexamethylethane to tetra-*tert*-butylethane, were subjected to MM2 analysis. Most of them were analyzed by both the torsion angle driving method and later by complete relaxation for selected conformations. The results of these analyses were then digested and used to verify or modify the rules suggested by Anderson and others. This resulted in a preliminary set of rules which were then formalized as logical axioms that were expressed in PROLOG and used in WIZARD.

Having obtained some of the knowledge needed to build a preliminary conformational analysis ES, we then needed to choose a method of representation for this knowledge. One method which seemed to be applicable was that used by a well-publicized ES, the Stanford MYCIN program.¹⁴ In such an approach, knowledge is stored as rules, each of which draws one conclusion about a subject, and is dependent on one or more terms. These terms may be observed or may themselves be the conclusions of yet other rules. For example, the bond angle at an atom could be inferred to be increased IF it was between two bulky groups (observation) AND if these groups showed steric interactions (inferred). To decide about the possibility of increased bond angles at a certain atom, the system would *backward chain* through the rules until all inferred terms were reduced to their component observational terms. These would then be answered yes/no, and the appropriate conclusions drawn.

At first it seemed that such an approach mirrored exactly the sort of decision making that our conformational analysis experts used in their own thinking. So the first version of WIZARD was based on this paradigm. The language chosen for this project was a mixture of mostly PROLOG with

FORTRAN support routines (see below for an explanation for the mixture). Thus, knowledge about what constitutes a stable conformation could be expressed directly in PROLOG. The knowledge was expressed in a loosely hierarchical fashion with high-level goals, such as "stable conformation", dependent upon medium-level goals, such as "stable bond conformation" and "pentane rule", which in turn rested upon concepts such as "gauche" and "anti", all of which sat upon a basis of very low level concepts such as "atom of bond" and "alpha". The chemist would begin the analysis by requesting that the top-level term "stable conformation" be inferred. The rules that infer about the molecular geometries in a stable conformation would automatically be found and examined, and the terms in those rules similarly inferred. This chaining proceeded until the very lowest level observational terms were examined, and the results propagated upward to the highest level. Interactions between the rules at any one level were kept to a minimum, and no circularity was allowed. To handle nonstrained molecules, fewer than 50 rules specific to conformational analysis were needed, and about 200 rules total were used.

However, as we stated above, not all procedures are encoded in PROLOG. This is due to the differing abilities of PROLOG and FORTRAN. The methods that the programmer uses to encode the program's knowledge can have strong effects on the program's performance and extensibility. Knowledge engineering has so far recognized two different ways of encoding knowledge, in either a procedural or declarative fashion. In a procedural fashion, the knowledge is directly encoded into some programming language, such as FORTRAN. The source code is compiled into machine-executable instructions and linked into a large executable image. This is the most common form of program.

In a declarative system, the knowledge is stored as declarations of the relationships between objects. These declarations are then interpreted by a second program, such as the PROLOG interpreter. The interpreter performs tasks such as backward chaining, file handling, input and output, and maintenance of the data base. This can free the programmer from having to pay attention to these details. The vocabulary used in describing the relationships can focus upon the problem at hand instead of worrying about variables, files, etc. However, since the interpreting system has to be relatively general, a procedural approach can often be much faster or more direct. For example, screen graphics, or geometrical calculations, are generally much faster in a procedural language than in a declarative language.

It was for these reasons that WIZARD was designed to be a combination of declarative and procedural programming. For those processes that are "naturally" procedural, such as the graphical I/O or calculating Cartesian coordinates from internal coordinates, FORTRAN was chosen as a simple, low-level language. However, PROLOG was chosen for the inference of the conformations of the molecules due to its ability to express and manipulate high-level concepts. This ability not only included the automatic handling of backward chaining but also allowed for the use of logic and reasoning to check the consistency of the axioms.

PROLOG is based on the Predicate Calculus¹⁵ (PC), a mathematical theory of reasoning that formalizes argument and inference by manipulating symbolic formulas. The syntax of this calculus is independent of the domain and thus forms a general framework for such reasoning. To express knowledge in the PC one begins by assigning semantic associations to symbols. Then, following the rules of syntax of the PC, one expresses facts about objects and relations between objects. These expressions are called premisses and axioms, respectively. Then, by applying rules of inference, one can create a possibly infinite set of statements about the domain. This

is known as an axiomatic theory, and the set of statements derivable from the axioms and premisses is known as a model. A successful axiomatic theory may be defined as one whose model predicts the behavior of the real world.

This is the underlying principle of WIZARD. In the work reported in this paper we have associated our symbols with terms such as atom, bond, gauche, pentane rule, and stable conformation. In future papers we will report on work that extends our axiom base to cover terms such as bond order, heteroatoms, rings, and fused ring systems. Our axioms are bits of knowledge extracted from experts in the field of conformational analysis. The premisses are simple statements about the molecule to be considered, such as its connectivity and atom types. Together they form a *model* of conformational analysis for that molecule.

The use of such an underlying system of logic has several advantages. First, because the model logically follows from the axioms and premisses, if it does not agree with reality, we know that the axioms or premisses must be revised. This simplifies the writing of a large program because it automatically focuses attention on the facts rather than upon the procedures. As any person who has ever tried to write a large FORTRAN program can attest, getting a finished procedure to correctly manipulate the knowledge even after the algorithm has been properly designed can be as difficult as first creating the algorithm. Such debugging often takes longer than the design and writing stages. Since PROLOG uses a simple control structure and operates directly on clausal forms of PC axioms, this minimizes the chance that bugs will arise due to programming and focuses attention on the knowledge instead. (Of course, typographical and similar errors can still wreak havoc.)

The second advantage is that the axioms, and thus the whole knowledge base of the program, can be used in generalized proof procedures to discover errors and new relationships. By the use of standard algorithms for theorem proving, one can test the axiom base for inconsistent clauses. It is hard to imagine performing this kind of checking if the data base is either written in a standard procedural language (such as FORTRAN or Pascal) or expressed in a language specially devised to represent chemical facts.

DESIGN OF THE FIRST VERSION

Having obtained a body of knowledge that we wished to build into an expert system and having chosen the language and method to express them in, we began to formalize the knowledge into logical axioms. These axioms could be broadly classified into three types: general axioms applicable to problem solving as such; axioms about conformational analysis; and general axioms that provide definitions about molecules. The first type of axioms are often used by ES programs in many dissimilar domains. For example, Anderson suggested that the conformers be created by first attempting to lay the molecule along the diamond lattice, checking to see if problems existed, and if there were to try to fix them. This can be shown by contrasting *n*-octane, which fits nicely on the diamond lattice, and 1,1,2,2-tetra-*tert*-butylethane, which cannot lie on the diamond lattice without severe van der Waals repulsions and so must adopt a higher energy conformation with increased bond angles.¹⁶ However, it can be seen that the same underlying problem-solving technique was used by Sacerdoti in his program NOAH, which helps robots plan how to perform desired actions.¹⁷ A generalized plan is created and criticized for problems, which are resolved if present, and the plan is then refined. As an example, a robot which plans to go through a doorway might find that the door is locked and thus must include the step of obtaining the key as a resolution to this problem. This is an advanced variant of the generate-

The analysis of a Molecule is done IF we have succeeded in:
 reading in Molecule, and recognizing patterns, AND
 performed a basic inference which affords a Suggestion_list, AND
 criticized this Suggestion_list to give a Criticism_list, AND
 used the Criticism_list to resolve problems, giving a
 Resolved_list, AND
 combined all of the entries on the Suggestion_list and Resolved_list
 to give a Final_list, AND
 created disk files of all the molecules on the Final_list.

```
analysis(Molecule) :-
  read_in_and_recognize_patterns(Molecule),
  basic_inference(Molecule, Suggestion_list),
  criticize(Suggestion_list, Criticism_list),
  resolve(Suggestion_list, Criticism_list, Resolved_list),
  combine(Suggestion_list, Resolved_list, Final_list),
  save_molecule_files(Final_list).
```

Figure 2. English and PROLOG versions of the fundamental algorithm of WIZARD.

and-test method.¹⁸ Since this approach seems to be general, we decided to incorporate it into WIZARD.

This level of metacontrol is expressed in the structure of WIZARD's knowledge base. The axioms about conformational analysis are divided into major high-level groups which describe what it means to infer about the basic suggestion, to criticize such a suggestion, to resolve a conformer's problems, and to refine a conformation. A completely automatic version of WIZARD could then be described by the statement shown in English and in PROLOG in Figure 2.

However, this is not a particularly efficient way to proceed. As will be shown in the Results section, it is not always desirable to exhaustively analyze the molecule, to resolve each (or perhaps any) of the strained conformers, or to save each possible suggestion. Thus, this statement has been replaced with a set of more complex statements that allow for significant interactions between the chemist/user and the program. The user may want to decide for each individual conformer if it is to be resolved, or saved. On the other hand, the user may not want to see any strained conformers, in which case the basic inference and the criticism steps are interleaved. But, for the sake of this discussion of the design of WIZARD, we can use the statement as if it were the actual driving statement of WIZARD.

The first clause in the rule for analysis of a molecule is to recognize patterns and assign higher level names to known patterns. The pattern recognition is broken into two parts. The first section performs the low-level pattern recognition, such as creating the basic connection table. The higher level pattern recognition classifies bonds according to their types and identifies other higher level objects, such as rings, cage structures, and the like. Although WIZARD does not currently infer about some of these structures, we are currently enlarging the knowledge base to encompass these other classes of objects.

The first step in the generation of acyclic conformers is to lay the molecule along the diamond lattice in all possible permutations. However, for several reasons this was not done by simply nesting several loops that assign angles of 60°, -60°, and 180° to each bond. This sort of programming would significantly hinder the further extension of WIZARD. For example, when olefinic and ring bonds are included in WIZARD, the program would be forced to recognize these bond classes as special cases. In addition, as explained below, assigning values that deviate from the diamond lattice can provide better results at the expense of some computational simplicity.

Instead, each object in the molecule is assigned a class, and each object-class has an associated frame of knowledge which includes a set of *natural subconformations* assigned to it. For all objects that are bonds of the class sp^3-sp^3 , the natural angles lie upon the diamond lattice. This technique also allows one to associate relative energies with each bond class and its

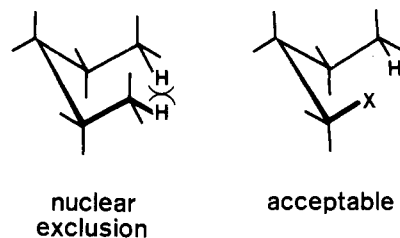


Figure 3. Unacceptable and acceptable 1,3 parallel interactions.

individual subconformations. Thus, a simple butane bond is assigned to be approximately 0.8 kcal less stable in the gauche conformation than in the anti conformation. Since methyl groups are rotationally symmetrical, only one torsional angle is examined. The energies for anti and gauche contributions are differentiated only for the "butyl" and "isopentyl" bonds, since in the more highly branched hydrocarbons the long-range van der Waals effects become too complex for simple prediction (*vide infra*). Further adjustments to these energies can be made for factors such as pentane-rule violations and others (*vide infra*).

When a conformer is constructed from these subconformations, an estimation of the strain energy is obtained by simple addition. It was never expected that this method would approach any realistic force field in calculating energies. However, a comparison of these results with energies obtained from MM2 is quite encouraging. Where these estimations fail is in accounting for long-range effects and the ability of the conformers to find pathways for strain relief. Consequently, attempts to improve them at this stage would not succeed. The first time the molecule is examined as a whole is in the criticism stage. Adjustment of the energies would have to be done at that point.

The criticism axioms currently examine the molecule for two major sorts of problems, viz., pentane-rule violations and nuclear exclusions. The first critic examines the molecule to see if the chain adopts the unfavorable g+/g- substructure. The second critic examines the molecule to see if a chain has folded back upon itself so that two nuclei occupy the same region of space. Future extensions will include dipole interactions and ring fusion strain among others.

These two critics can be collapsed into variations on the same theme. A standard phrasing of the pentane rule states that the problem arises due to a 1,3 parallel interaction. While this may be true to a certain extent, there are counterexamples involving heteroatoms where the g+/g- conformer is present in significant amounts.¹⁹ In addition, one of the design decisions behind WIZARD was that questionable cases should be decided by MM2 as long as the number of ultimately rejected suggestions was not overwhelming. Thus, we have assigned the pentane rule as a specific case of nuclear exclusion, as shown in Figure 3. This then allows the sort of situations represented by, for example, 1,3 parallel methyl-halogen structures (stabilized g+/g- conformers) to slip through for final adjudication by a mechanics program.

The critic knowledge base is currently being used in two major fashions. In the first, a completed conformation that has all bond and dihedral angles assigned is criticized. However, the critics can also be used even as the conformation is being built to indicate if the conformation can be seen as already having a problem. This technique works surprisingly well. (For a more conventional treatment, cf. ref 3c, where the MOLY system's CONFOR module is described.) We were afraid that since a successful conformation would be criticized many times, this method might be slower on molecules which would have few poor conformers. However, due to the simplicity of the current critics, relying as they do on simple addition instead of complex vector geometry, this

Table I. English Paraphrasing of Some Simple Rules for Strain Relief

If there is a pentane violation, it can be relieved by increasing the dihedral angle of either of the central dihedrals by about 25°.

If such a pentane-rule-violation dihedral is being opened, it can open 5° less if the central carbon has a substituent on the side toward which the bond opens.

If such a pentane-rule-violation dihedral is being opened, it should open about 5° more if the carbon it is retreating from has two or more substituents or if it has a single gauche substituent.

If a nuclear exclusion occurs, then each bond in the pseudoring can be treated as a pentane-rule-violation bond.

technique is faster for all except the smallest molecules.

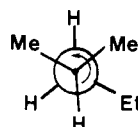
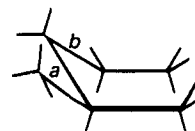
This led us to explore the possibility that interleaving criticism and resolution might lead to similar success. However, the resolution process needs to look at larger portions of the molecule than the critics. Once a pentane violation has been discovered in a partially assigned molecule, no assignment of the rest of the molecule will remove this violation. However, the resolution of such a violation depends on whether it is a single isolated violation or part of a multiple violation system. The differences between these types of systems leads to significantly different types of strain relaxation processes. Thus, to choose the proper resolution procedure, we must build and criticize the entire conformation first.

The major differences between the first and second versions of WIZARD are centered in the resolving stages. Since the differences greatly extended the utility of WIZARD and are a departure from classical ES techniques, we will describe both implementations in some depth. To help the reader understand why we felt it necessary to change our approach, we will first discuss the initial implementation of WIZARD and its results. Since the scope and limitations of these results demonstrate why a new version was required, we will postpone the discussion of the design of the second version of WIZARD until after they have been discussed.

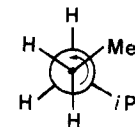
In the first version, we used standard ES techniques, and we obtained information from experts as to how such strain could be resolved. This led initially to a small number of basic postulates, some examples of which are paraphrased in Table I. These axioms are of a simple localized situation/response type. They state that if a certain situation occurs, then the molecule will show a certain response. For example, the first axiom simply states that if a pentane-rule violation is found, then each of the central dihedrals could be opened to relieve strain. However, this rule is too general and ignores substituent effects. If one considers 2-methylpentane, as shown in Figure 4, it can be seen that increasing the dihedral angle of the trisubstituted bond, bond a, will probably be a higher energy process than opening the disubstituted bond, bond b. The former process will bring two methyls into greater eclipse with α -hydrogens, compared to only one methyl in the latter. However, since there are hundreds of different substitution patterns available for such pentane chains, we decided to use a few general axioms instead of hundreds of specific cases. This was done in order to keep in line with our design decision to oversuggest as long as the user is not swamped by doubtful inferences. This can be done since WIZARD uses MM2 as the final arbiter and refinement procedure, and thus all that is necessary for WIZARD to do is to place the suggestion within the energy well for a local minimum. By use of extremely general axioms, the suggested conformers were more likely to cover the surrounding space.

RESULTS

In this paper we present the results of applying WIZARD to linear and branched saturated hydrocarbons. For nonstrained



bond a



bond b

Figure 4. Rotation around different torsional angles will be energetically different.**Table II.** Complete Results of WIZARD's Analysis of 2,3-Dimethylpentane^a

conf ID	est SE, ^b kcal	MM2 SE, ^b kcal	MM2 CPU, s	fit, Å	notes
c1	7.12	7.96	21	0.063	
c2	9.12	10.05	37	0.305	gave r1, r2
c3	9.12	9.60	48	0.326	gave r3, r4
c4	9.78	9.87	57	0.294	gave r5, r6
c5	9.78	9.10	46	0.278	gave r7, r8
c6	7.78	7.75	27	0.092	
c7	7.12	8.20	21	0.065	
c8	7.12	7.75	20	0.110	
c9	7.12	7.74	33	0.142	
r1	9.12	10.05	46	0.162	same result as c2
r2	9.12	10.01	34	0.129	
r3	9.12	9.60	31	0.142	same result as c3
r4	9.12	10.14	35	0.162	
r5	9.78	9.87	30	0.095	same result as c4
r6	9.78	9.43	26	0.138	
r7	9.78	10.14	149	0.157	same result as r4
r8	9.78	9.10	30	0.166	same result as c5

^a WIZARD 49 CPU s, 17 conformers. ^b Steric energy.

molecules and conformers, the "standard ES" version of WIZARD produced excellent results. In this section, the results of some representative analyses will be presented to give an idea of the scope and limitations of the first version of WIZARD. We will begin with small molecules, so as to better be able to fully show how WIZARD proceeds and to show the strengths and deficiencies in our approach. However, WIZARD is capable of handling larger molecules, and a sampling of these results will be presented to demonstrate this. Due to the complexity of these analyses (e.g., up to 100 conformers), these larger molecules will not be analyzed in detail. None of the following examples belongs to the training set, molecules examined by MM2 and used to build WIZARD's knowledge base. Molecules are referred to by name or number, but individual conformers will be referred to by either the WIZARD conformer number preceded by a "c" for an unresolved conformer (e.g., c4), or by an "r" for a resolved suggestion (e.g., r2). These numbers identify the entry in the table pertaining to the appropriate molecule.

The first molecule presented is 2,3-dimethylpentane, and the results are shown in Table II. WIZARD recognized that of the six degrees of freedom in this molecule four are simple methyl rotations, and so simply assigned a fixed dihedral angle to these bonds until the resolution step. Thus, this molecule has only two degrees of real freedom. In the rest of this paper, degrees of freedom will only refer to such carbon-chain dihedral angles.

Table III. Dihedral (D) and Bond (A) Angles^a in Conformers of 2,3-Dimethylpentane Found by WIZARD

conf ID	D1, 3,1,2,6	D2, 1,2,6,12	A1, 3,1,2	A2, 1,2,8	A3, 1,2,6	A4, 2,6,12	A5, 12,6,14
c1	59.19	53.88	115.14	109.97	114.25	112.38	108.47
c2	62.21	-83.04	116.97	111.42	117.26	115.60	110.44
c3	60.85	143.89	116.25	111.80	113.89	111.07	108.57
c4	-59.58	93.35	115.88	107.42	113.12	111.70	107.26
c5	-63.13	-42.55	115.72	108.49	116.80	114.04	110.68
c6	-56.63	174.17	114.59	109.07	114.14	112.39	109.67
c7	-179.24	58.03	115.07	110.49	111.81	112.71	107.99
c8	-171.11	-58.09	113.92	111.27	113.65	113.95	109.77
c9	-166.79	166.69	111.18	113.99	111.56	112.36	109.69
r1	62.21	-83.04	116.97	111.42	117.26	115.60	110.44
r2	89.46	-55.95	116.70	112.23	115.78	114.73	109.55
r3	60.85	143.89	116.25	111.80	113.89	111.07	108.57
r4	93.05	173.49	116.52	112.24	113.73	112.18	109.50
r5	-59.58	93.35	115.88	107.42	113.12	111.70	107.26
r6	-97.80	55.25	113.90	109.07	113.17	112.92	108.17
r7	93.05	173.49	116.52	112.24	113.73	112.18	109.50
r8	-63.13	-42.55	115.72	108.49	116.80	114.04	110.68

^aFor numbering, see Figure 5.

An exhaustive analysis provided nine basic conformers. If the user wanted only the most stable conformers, those likely to be present in a mixture at room temperature and pressure, only the noncriticized conformers c1 and c6-c9 would be analyzed. For the sake of the explanation, the four conformers that contained pentane violations, c2-c5, were resolved. This produced a further eight suggestions, labeled r1-r8. The total amount of CPU time required by WIZARD for the generation of these 17 suggestions was 49 s. Each of these suggestions was saved on a separate disk file and analyzed with MM2.

The five suggestions that did not contain pentane-rule violations, numbers c1 and c6-c9, all fell nicely into a nearby local minimum. A least-squares fit between the MM2 final conformations and the WIZARD suggestions was performed with the molecular comparison program in the MIMIC system.²⁰ The results are given for each conformer, and the root mean square differences in the positions of the atoms range from 0.063 to 0.142 Å. An example of a plot showing the superposition of the WIZARD suggestion for c1 and MM2's final result is shown in Figure 5. The four strained conformer suggestions were also minimized and resulted in four different strained conformers. The eight suggestions from the resolution process, r1-r8, were also minimized. Four of these suggestions (r1, r3, r5, and r8) fell into the same local minimum as did the strained but unresolved conformers c2-c5. These conformations thus correspond to those that would be found by the process of aligning the molecule on the diamond lattice and allowing it to relax.

However, the other four suggestions, when minimized, produced three new conformers (two suggestions collapsed to the same conformer). These three new conformers were significantly different from any conformations discovered by minimizing molecules with classical diamond lattice geometry. This can be seen by examining the list of dihedral angles presented in Table III. Thus, a total of three new high-energy conformations was discovered by this process. A superposition of the first resolved suggestion, r1, and the MM2 result is shown in Figure 6. It can be seen that WIZARD's suggestion was quite close to the final minimum-energy conformation found by MM2. Thus, modeling programs that simply assign acyclic portions of the molecule to lie on the classical g⁺, g⁻, or anti torsion angles would fail to find these conformations.

This analysis of 2,3-dimethylpentane shows several interesting features. For example, it is interesting to note that MM2 required a total of 188 CPU s to minimize the four suggestions 2-5, but only required 215.5 CPU s to minimize the eight resolved suggestions derived from them. As can be seen by looking at the initial energies and fits, the resolved suggestions were substantially closer to the local minima than the non-

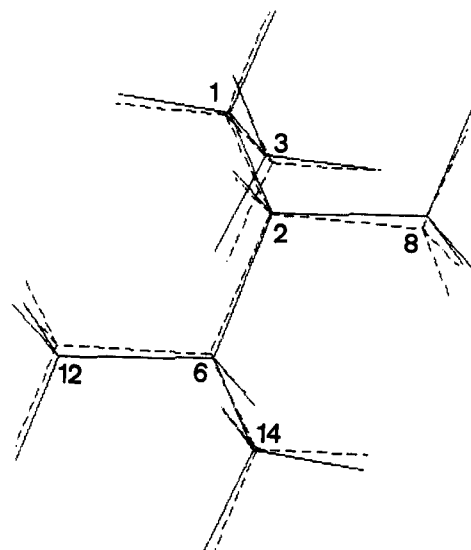


Figure 5. Superimposition of WIZARD's suggestion (dashed lines) and MM2's minimized result (solid lines) for the nonstrained conformer c1 of dimethylpentane (see text).

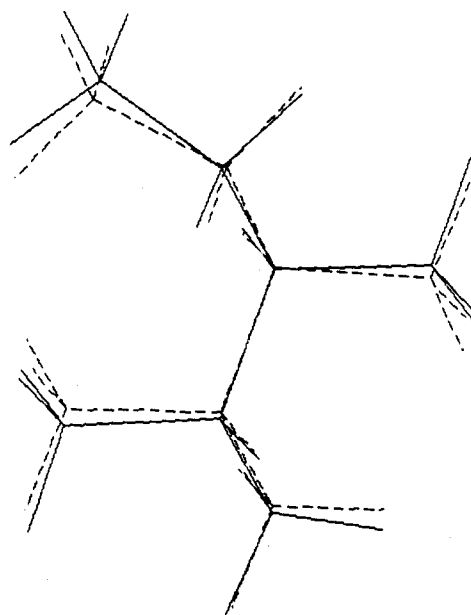


Figure 6. Superimposition of WIZARD's suggestion (dashed lines) and MM2's minimized result (solid lines) for the resolved strained conformer r1 of dimethylpentane (see text).

Table IV. Results of the Naive Use of MM2's Driver Option To Obtain Results Similar to Those of WIZARD's

ID ^a	dihedral angles		steric energy, kcal		excess, %	order ^b
	D1	D2	driver	single		
j1 (r8)	-60	-60	9.921	9.109	8.91	6,6
j2 (r6)	60	-60	11.553	9.433	22.47	8,7
j3 (c9)	180	-60	7.818	7.730	1.14	1,1
j4 (r1)	-60	60	11.155	10.054	10.95	7,8
j5 (c1)	60	60	8.083	7.965	1.48	3,4
j6 (r7)	180	60	12.628	10.146	24.44	9,9
j7 (c8)	-60	180	7.906	7.753	1.97	2,3
j8 (c7)	60	180	8.217	8.208	0.11	4,5
j9 (c6)	180	180	8.617	7.750	11.18	5,2

^aCorresponding WIZARD suggestion in parentheses. ^bRelative order, driver and unrestricted.

resolved conformations. This led to a substantial reduction in the amount of effort MM2 required. This can be seen by comparing the MM2 CPU times for each of the resolved conformers with the time required to minimize each nonresolved parent suggestion.

The fact that r7 did not correspond to an actual separate local minimum was explored. The driver option of MM2 was used at 5-deg intervals to explore the local region around this suggestion. The results indicate that there is no local minimum with sufficient depth to prevent the molecule from falling into one of the other already known minima. (It is possible that there is a minimum of less than 5° width with a few hundredths of a kilocalorie of depth.) The fact that WIZARD predicted a conformer where none exists may be regarded as a fault at first glance. However, this is actually desirable compared to the alternate behavior, i.e., missing out on some conformers. As long as the combination of WIZARD and a MM program takes less time and effort than the use of the MM program with a driver option, WIZARD is time effective. With this molecule, this is definitely the case.

The bond driver option of MM2 was also used to explore the same molecule by stepping the bond driver in 120° increments, as shown in Table IV. While we knew this to be a naive and oversimplified approach, we thought that it might prove interesting to compare the results. While MM2 produced results similar to those produced by WIZARD, the conformers were not properly minimized due to the unnatural locking of the driven torsion angles. Each conformer produced by the torsion driver method was then re-minimized with no restrictions. This showed that the driven analysis resulted in an average overestimation of the strain energy for the stable conformers of about 1/4 kcal/mol. The energies of the strained conformers were overestimated by an average of over 1.5 kcal, about 17%. Perhaps more important than the slight energy difference is the fact that the relative orders of stability of the conformers were different between the driven and totally free MM2 runs in six of the nine conformers.

Considering the popularity of generating conformational space energy maps by use of the bond-driving technique, these results are of great interest. They suggest that great care needs to be exercised in the use of this technique. If possible, each locally stable conformer should be reanalyzed with no restrictions. This will, however, greatly increase the need for human intervention and CPU time, unless a program such as WIZARD is utilized.

The increase in CPU time necessary to reanalyze each minimum-energy conformation leads to the next point, mainly that driving at 120° intervals was not sufficient to discover the lowest point of the minima that were discovered and to find all of the local minima for this simple molecule. Several different driving increments were tested, and it was found that to discover the other local minima which WIZARD had suggested it was necessary to reduce the driving increment to less

Table V. Partial Results of WIZARD's Analysis of *n*-Octane^a

conf	est SE, ^b kcal	MM2 SE, ^b kcal	MM2 CPU, s	fit, Å
c1	4.5	4.74	25	0.049
c6	5.2	5.66	48	0.147
c13	5.2	5.64	45	0.161
c16	6.0	6.31	48	0.162
c37	6.0	6.50	41	0.125
c77	6.0	6.53	48	0.158
c17	6.8	6.98	37	0.131
c44	6.8	7.68	41	0.141
c66	6.8	7.17	37	0.138
c46	7.5	7.99	44	0.140
c72	7.5	7.72	51	0.147
c97	7.5	7.60	137	0.181
c75	8.3	8.52	51	0.179

^a131 CPU s, 99 conformers. ^bSteric energy.

Table VI. Partial Results of WIZARD's Analysis of 3-Isopropylhexane^a

conf ID	est strain	MM2 CPU, s	init SE, ^b kcal	final SE, ^b kcal	fit, Å
4	58	37.1	26.6	10.68	0.112
49	58	40.8	28.5	10.51	0.118
51	68	37.2	30.9	11.07	0.126
55	50	41.0	28.4	10.79	0.119
56	60	40.9	31.0	11.41	0.138
70	42	33.2	28.5	10.52	0.117
71	52	32.6	31.0	11.19	0.112
61	54 ^c				
r1		44.8	44.1	13.06	0.143
r2		72.0	43.8	12.50	0.151

^aPartial analysis: 57 CPU s, 7 conformers. Exhaustive analysis: 591 CPU s, 83 conformers (2 resolved) (not all conformers shown).

^bSteric energy. ^cIncludes a "penalty" of 12 kcal/mol for excessive strain; this was the only penalty-conformer with a sum less than that of some of the nonpenalty conformers. Resolved to r1, r2.

than 15°. Had this been done in a full analysis, it would have required several CPU hours. Since WIZARD had indicated in which regions to search, we were able to discern the minimum increment with restricted searches, but the amount of human intervention went up significantly. Thus, it can be seen that the combination of WIZARD and MM2 has fulfilled the requirement that it should be more efficient than a search with MM2 alone.

A restricted analysis (see above) of *n*-octane afforded 99 conformers in 131 CPU s. To test the quality of the suggestions without having to minimize all 99, a randomly selected assortment was minimized with MM2, and the results are shown in Table V. The least-squares fits between WIZARD's suggestions and the conformers obtained by MM2 are shown, and the root mean square differences for the atom positions average 0.148 Å for all conformers.

The results for 3-(1-methylethyl)hexane (3-isopropylhexane) are as shown in Table VI. The restricted analysis required 57 CPU s to produce seven suggestions. An exhaustive analysis was performed and required 550 CPU s to generate 81 basic suggestions, 7 of which were violation free. WIZARD suggested that one of these strained conformers might have an equivalent energy to the other seven nonstrained suggestions, so it was resolved and submitted to MM2 as well. However, the results show that this conformer was in fact much more strained. Comparison plots for a nonstrained and a strained conformer are shown in Figures 7 and 8. A restricted analysis of 3,4,6-trimethyloctane produced 16 problem-free conformers, requiring 93 CPU s. All 16 suggestions were minimized, and the results are shown in Table VII.

The correlation between WIZARD's energy estimates and the actual strain energies computed by MM2 is shown in Figure 9. The estimates for octane are quite good, and the estimates for isopropylhexane are acceptable. But in the case of tri-

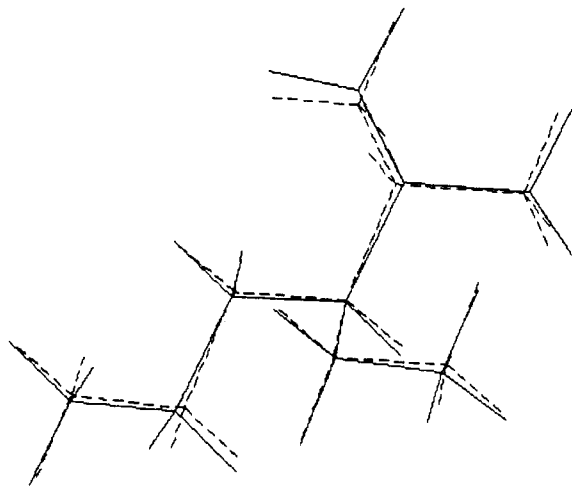


Figure 7. Superimposition of WIZARD's suggestion (dashed lines) and MM2's minimized result (solid lines) for the nonstrained conformer c1 of isopropylhexane (see text).

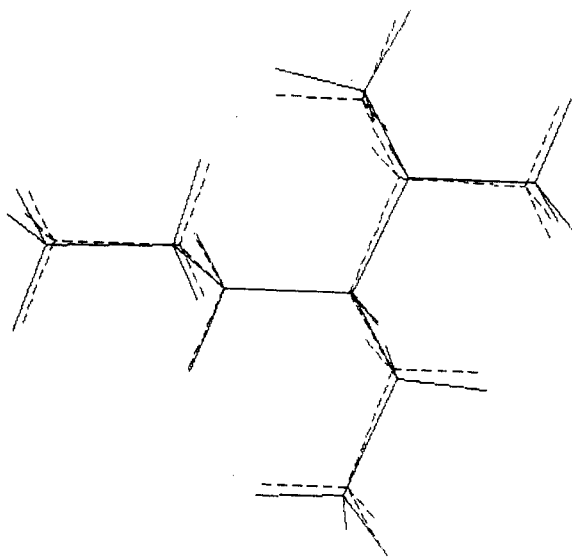


Figure 8. Superimposition of WIZARD's suggestion (dashed lines) and MM2's minimized result (solid lines) for the resolved strained conformer r1 of isopropylhexane (see text).

methyloctane the correlation between the final energies and WIZARD's estimates is quite poor. This is because WIZARD currently attempts to estimate the final strain energy by simply adding up contributions based on the α -environment for each bond. This oversimplified model actually works surprisingly well for many hydrocarbons and has been used to successfully predict conformer populations to explain ^{13}C NMR data.²¹ However, for highly branched compounds, the effects of the substitution can extend to more distant bonds through complex interactions (a mixture of van der Waals and through-bond effects). This has been seen experimentally in NMR studies of highly substituted ethanes.²² Apparently trimethyloctane is sufficiently complex to show these effects.

Another effect that arose when larger molecules were analyzed can be demonstrated by the results of an analysis of 3,6-diethyl-4-isopropyl-5-methyloctane. WIZARD generated 63 conformers in 460 CPU s. A semirandom sampling of these conformers was minimized, and, as expected, the energy estimates were not good; however, the initial geometries were sufficiently good to lead MM2 to find the expected minima. The average root mean square deviations between the suggested and minimized conformers were not as good as in previous analyses, however: the average deviation for the chosen conformers was about 0.25 Å. Upon examination, it could

Table VII. Complete Results of WIZARD's Analysis of 3,4,6-Trimethyloctane

conf ID	est SE, ^a kcal	MM2 SE, ^a kcal	MM2 CPU, s	fit, Å
c1	12.1	14.08	166	0.190
c2	10.98	14.50	174	0.160
c3	10.98	13.34	54	0.145
c4	12.1	13.55	166	0.152
c5	10.98	13.66	170	0.223
c6	12.1	13.81	179	0.138
c7	10.98	14.53	92	0.133
c8	10.98	13.35	48	0.141
c9	12.1	13.54	70	0.142
c10	10.98	13.51	129	0.207
c11	13.22	13.48	69	0.151
c12	12.1	13.72	116	0.155
c13	14.3	12.69	69	0.155
c14	13.22	13.48	91	0.173
c15	13.22	12.87	75	0.173
c16	12.1	13.09	70	0.183

^a Steric energy.

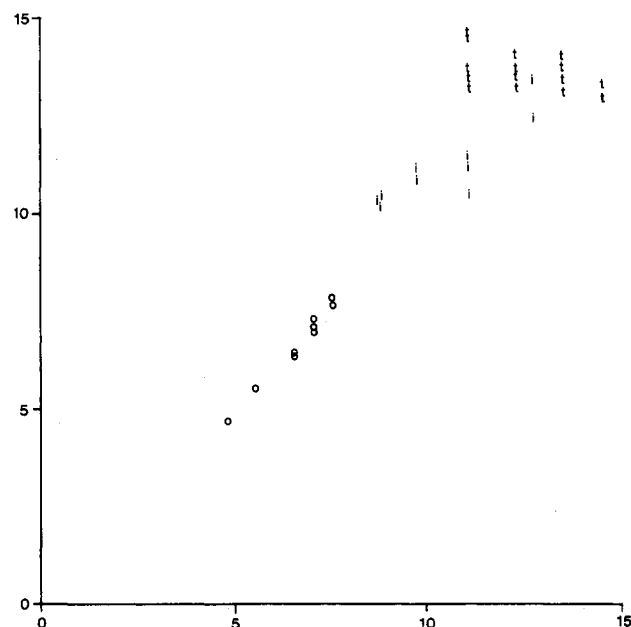


Figure 9. Correlation between WIZARD's estimates and MM2's calculations: (o) *n*-octane; (i) isopropylhexane; (t) 3,4,6-trimethyloctane.

be seen that if either half of the molecule were best fit, the deviation was substantially less, about 0.15 Å. As the molecules become longer, small deviations between the central bond angles in the suggestions and the final minimized bond angles produce a pronounced "lever arm" effect. In an attempt to see if this effect could be reduced, we changed the natural angles in the assignment frames from pure diamond lattice angles to more realistic angles. For example, gauche butane-like bonds were assigned a value of $\pm 66^\circ$ instead of exactly $\pm 60^\circ$. While this improved the fit substantially by suggesting more realistic starting points, the final MM2 results were the same. Additionally, the checks for nuclear exclusion, pentane-rule violations, and other geometrical considerations became more difficult. Instead of having three symbolic angles to manipulate, we now had a continuum of angles, and so explicit calculations became necessary. Since this added complexity purchased no real gains, we returned to the simple expedient of assigning diamond lattice angles to nonstrained suggestions. However, we have allowed the user the option of *refining* the output structure in this fashion.

FAILURE OF THE FIRST VERSION

For relatively simple unstrained and strained molecules, this first version of WIZARD seemed to be living up to its design

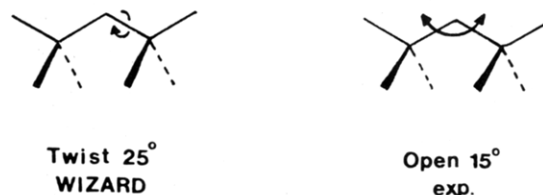


Figure 10. Comparison of suggested processes for relief of strain in 2,2,4,4-tetramethylpentane.

Table VIII. Additional Rules for Strain Relief Added to First Version of WIZARD

If two pentane-rule-violation pathways share the same central dihedrals (e.g., di-*tert*-butylmethane), then dihedral opening would close the other pentane-rule-violation. In this case the central bond angle will undergo 15° of in-plane spreading.

If three pentane-rule-violation pathways share the same central carbon (e.g., tri-*tert*-butylmethane), then the three bonds will bend out of plane by about 5° and the three dihedrals will shear by about 10°.

criteria. Axioms had been found to determine initial suggestions for conformations, and other axioms were used to determine how strained molecules would relieve the strain. For example, conformers that contain pentane-rule violations often relieve the accompanying van der Waals' strain by increasing one of the associated torsional angles. A rule to this effect has already been discussed. However, in certain cases this sort of bond twisting would lead to greater strain. For example, the first such case that we discovered that was not covered by the rules suggested a priori was 2,2,4,4-tetramethylpentane. All conformations of this molecule were criticized, and a partial analysis returned no unstrained conformers in 10 CPU s. Resolution of the strained conformers with the initial set of resolution rules led to the result that all of WIZARD's suggestions were quite wrong. Utilizing the simple rules shown in Table I, WIZARD suggested that this molecule could relieve its strain by the process shown in Figure 10. However, this actually leads to more strain due to the increased proximity of two methyl groups. The actual process, as found by MM2 and confirmed by experimental evidence,²² involves an increase in the bond angles at the central atom.

Another rule (the first in Table VIII) was added to account for this case. However, several other molecules that had such "opposing pentane-rule violations" were discovered upon thought and experiment, and so special-case rules had to be added for these molecules. These rules all decided between torsion angle changes or in-plane bending of bond angles. After about 15 of these rules were added, we attempted the analysis of triisopropylmethane. Once again we were disappointed to see that the current set of rules was insufficient. Since triisopropylmethane contains three such shared pentane-rule violations about a central carbon, the interactions between the relaxation processes are complex. In spite of the fact that the molecule actually behaves as an entity, in this version of WIZARD the resolutions were performed stepwise, without regard to other resolution processes in the molecule, and the opening of the third angle effected the closing of the first. This required the addition of yet another set of rules to account for such interaction problems, including the second rule shown in Table VIII. By this time we were gaining experience at finding counterexamples to our rule set, and it took very little time to find still more molecules that showed some local C₃ symmetry, but were asymmetric overall, which were not tractable by these rules.

Other rules were added to account for these cases. It shortly became evident that one would need a very large set of special-case rules, and the system would never be complete. In

New Critics and Resolvers

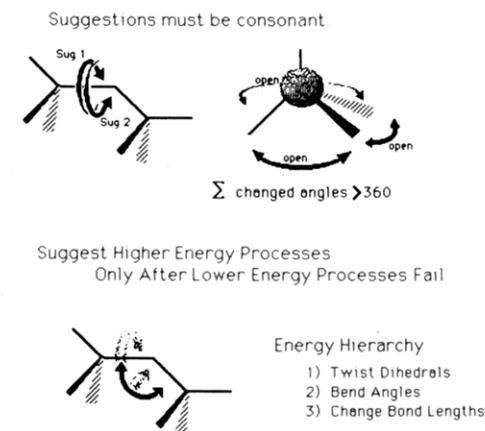


Figure 11. Examples of critics and resolvers in the new version of WIZARD.

order to solve this problem, we stepped away from the standard ES paradigm.

DESIGN OF THE SECOND VERSION OF WIZARD

Since WIZARD was performing well for unstrained conformations and did well in detecting strained conformations, the major area of improvement in the second version of WIZARD was in the area of the resolution of geometries of strained conformers. This was done by using fewer rules, which were more general, and allowing the program to combine them in a fashion that more closely resembles reasoning than does simple backward chaining.

In a normal ES, information is propagated by backward chaining of rules. This creates a tree structure, in which the individual nodes only communicate with the nodes at a higher or lower level in the tree. However, as we pointed out earlier, this is exactly the problem in trying to suggest how molecules such as triisopropylmethane relieve their strain; the molecular processes are not independent but are actually highly dependent and interactive. Furthermore, it was felt that the actual processes available to relieve strain were actually fairly simple: i.e., one could change torsional angles, change bond angles by in- or out-of- plane bending, or change bond lengths. The actual combination of these processes could be very complex, but, nevertheless, relaxation occurred through a combination of these simple processes.

Upon reflection, it seemed obvious that this process was similar to the process of conformer generation. In the main process, conformers are suggested by first making relatively naive suggestions about the subconformations of individual bonds, and then the entire conformer is criticized and problems are resolved. This same process could be applied to discovering how the strained conformers relaxed. We could apply relatively simple suggestions to each source of strain and then collectively criticize these suggestions and resolve problems in our suggestions. Note how the process is the same, even though the actions are quite different. In the main process, criticism involves the discovery of van der Waals strain and nuclear exclusions. In this process, criticism would involve the discovery of strain relief suggestions that were incommensurate. For example, it would be relatively simple to discover when the same bond was being twisted in both a positive and negative fashion simultaneously, as shown in Figure 11. Since the same mechanism of generate, criticize, and resolve was being used, we could utilize the same procedures as before, thus saving program space and human effort and making the program simpler overall.

Table IX. Strain Resolution Suggestions and Critics in the Second Version of WIZARD**Strain Relievers**

If two bonds are the central bonds of a pentane-rule violation, then mark one bond as being twistable by 15° and mark the other bond as being passive.

If a bond pair has been marked as openable and passive but has been criticized, then remove those marks and mark the pair as in-plane angle-openable.

If a bond pair has been marked as angle-openable but has been criticized, then remove those marks and mark each bond as lengthenable.

Critics

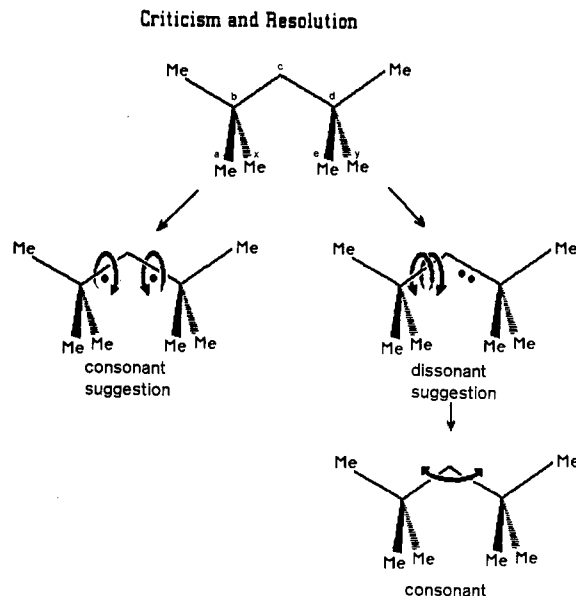
If a bond has been marked as twistable two or more times and if the difference between the angles of twist is more than 5° between any pair of twists, then criticize these markings.

If an atom has three angles that have been marked as in-plane angle-openable, then criticize these markings.

Thus, we discarded all of the special-case rules we had painfully created and substituted a small number of naive "strain resolution suggestions", which are shown in Table IX. These were ordered according to the probable energy required to perform such a process. Torsion angle changes were considered to be easier than bond angle changes, and bond length changes are considered to be the last resort. We also added geometrical and arithmetical critics that check to see if all of the strain resolution procedures are consistent in their action and intent to supplement the existing conformational critics. For example, if several resolvers all suggest that a torsional bond should be twisted in the same fashion, this is not criticized. However, it often turns out that the relief of strain by torsion angle twisting for a conformer which contains two or more sources of van der Waals strain would require simultaneous positive and negative twists about the same bond. An arithmetic critic would note this and mark this solution as bad. Similar critics note other problems such as dissonant angle changes.

Strain-resolved conformers that pass the critics are added to the list of suggestions to be minimized. Conformers that do not pass this level of criticism are returned to the resolver, and the next higher energy level of strain relief processes is suggested. These are then recriticized, and so forth. This process terminates either when all problems have been resolved or when no more suggestions can be made. This technique works surprisingly well. With only 10 total rules for strain relief and arithmetic and geometric criticism, we achieve better results than with the over 50 special-case rules in version one. In addition, since the process of naive suggestion and criticism is general and might even in some fashion mirror the actual molecular process, it has not been found necessary to add more rules as we have extended WIZARD to include more complex molecular types.

For example, let us return to 2,2,4,4-tetramethylpentane and see how the new version of WIZARD handles this molecule, as shown in Figure 12. First, the conformational critics mark this molecule as having two pentane-rule violations, one along atoms a, b, c, d, and e and the other along atoms x, b, c, d, and y. Thus, this conformer is turned over to the resolver. The resolver finds two pentane-rule violations and tries to relieve strain using the lowest energy process, i.e., torsion angle changes. It has been noted by examining the output of MM2 that a molecule twists only one bond in response to such a violation. Since the pentane path has two such candidate bonds and we have not included any selection terms in the current rules, each pentane-rule-violation pathway (a, b, c, d, and e or x, b, c, d, and y) receives two separate suggestions, one for each "central bond". Thus, we obtain four new suggestions,

**Figure 12.** Example of criticism and resolution in 2,2,4,4-tetramethylpentane.

of which two are shown in Figure 12 (the other two are symmetrical to the two shown).

Now the suggestions are given to the critics. The suggestion on the left is self-consistent and thus is passed on to MM2 for final judgment. However, the suggestion on the right is not self-consistent. The resolver has suggested that bond b,c be twisted by both $+15^\circ$ and -15° . This is obviously ridiculous and is criticized by a simple arithmetical critic. Thus, this suggestion is returned to the resolver, which then applies the next higher energy process, bond angle spreading. This passes the critics on the next examination, and so it is given as a conformer to be considered as well. Upon minimization, the first suggestion is shown to be a false minimum and will be discarded.

It might be objected that by improving the naive suggestors this false minimum could have been avoided. By the use of more specific rules this error was avoided in the first implementation of WIZARD. But this leads to the pitfall of a never-ending series of special-case rules. And in practice the naive suggestors do not always lead to undesirable false minima. In the case of tri-*tert*-butylmethane, some combinations of torsion angle changes turned out to be consistent and yielded the actual minimum-energy conformation. WIZARD passively discovered the threefold axis of rotation which leads to the well-known phenomenon of corotation.²³ It could be said that WIZARD applied the general theories of geometry incorporated in the simple critics to discover the associativity needed to allow such a relaxation process. This sort of process underlies some of the recent work in AI on learning machines.²⁴ However, it would not be proper to say that the current version of WIZARD is such a learning machine, but it does show the possibility for future expansion. It is also important to note that all of the special-case rules we had to create for the first implementation WIZARD are passively discovered by the second version, so it is not a question of completeness, but of efficiency.

RELATIVE EFFICIENCY

We have demonstrated that, in a limited domain, WIZARD can help a molecular mechanics program find minimum-energy conformers. In comparison to the torsion angle driving technique, the WIZARD technique provides fully minimized results. However, it is important to assess the cost of this extra step of analysis in comparison to the use of MM2 alone.

In the case of triisopropylmethane, WIZARD suggested a total of 23 potential conformers for MM2 to resolve. This process

required 350 CPU s (vide infra for machine details). MM2 required 2000 CPU s to minimize all 23 suggestions or an average of 87 CPU s each. The results of MM2's analysis showed that WIZARD had suggested nine unique conformers and that 14 suggestions were either high-energy conformers which collapsed to one of the nine lower energy conformers or duplicates by symmetry. Thus, the total CPU time for the analysis was about 2350 CPU s. Compare that to the amount of time that it might take to obtain the same results by the torsion angle driving method. The minimal torsion angle difference between two of the conformers was 33°. Thus, this gives us an estimate for the smallest angle we could increment by and yet hope to find the minima. But even stepping three bonds at this rather coarse interval would still require about 2×10^5 CPU s. And of course it would be necessary to step by a much finer increment, which would send the time up drastically.

Thus it can be seen that the use of WIZARD is cost effective for such molecules. Additionally, this imbalance tips more toward WIZARD's favor as the number of degrees of freedom increases. While WIZARD can make suggestions for molecules with seven or more degrees of freedom, producing a few hundred conformers at most, the amount of time that the torsion angle driving method would take for such a molecule would be at least 10^7 times that of a single analysis.

PROGRAM IMPLEMENTATION

WIZARD was developed on a VAX 11/780 which utilizes the VMS operating system (v 3.7). To run WIZARD itself, it is necessary to run the University of Sussex POPLOG system, which then allows the user to run his or her own programs. The POPLOG system includes three major AI languages, PROLOG, LISP, and POP11, and allows the user to link any VAX-compiled object modules from FORTRAN or Assembler. POPLOG provides methods for facilitating the passing of data between all these types of modules, but some types of data passing can be difficult due to the need to convert data types (e.g., FORTRAN-77 strings to PROLOG). WIZARD has portions written in all of these languages.

The main portions of WIZARD are written in PROLOG. However, the display portions of WIZARD are pieces of code from the local molecular modeling system (MIMIC) that were only slightly modified, as are the molecule data file input and output routines. This assured complete compatibility between WIZARD and the existing MIMIC/molecular mechanics environment at Lund. The pattern recognition routines, which are central to the feature recognition, have been written in PROLOG, LISP, and FORTRAN to facilitate testing the comparative speeds of each language. As expected, the speed increases in that order. However, since the current version of WIZARD is not limited by the speed of pattern recognition, the PROLOG pattern matcher is currently being used to facilitate ease of programming. The routines that manipulate the geometry of the molecule were also taken directly from MIMIC and are written in FORTRAN.

CONCLUSIONS

The work presented in this paper comprises a pilot project to explore the strengths and limitations of this approach to conformational analysis. In brief, we can conclude that WIZARD has lived up to its design goals. The combination of WIZARD and MM2 can explore the conformational space of a saturated acyclic hydrocarbon with greater rapidity than by use of the torsion angle driver of MM2 alone and with greater confidence of complete coverage than by hand generation followed by MM minimization. The use of higher level symbolic reasoning allows us to find otherwise ignored minima and to avoid some poor suggestions. The time required for this

extra ability is not substantial because WIZARD works at a higher level of reasoning where the interactions among many individual atoms and bonds can be handled simultaneously as systems, as in the resolution axioms.

This project has also shown some directions for further research. Of course, extension of the axioms to include multiple bonds, heteroatoms, and rings is important. The success of this pilot project has given us the confidence that this work will provide useful results, and so we are proceeding on this track. Other projects underway or under consideration include the recognition and proper utilization of both global and local symmetry. In addition, the utilization of other forms of higher level reasoning and knowledge is interesting. For example, not all saturated acyclic hydrocarbon bonds have similarly steep potential wells for torsion angle changes. By suitably encoding this knowledge as a measure of "rocking difficulty", we can learn to estimate where molecules will flex and what limits such flexing can have. Another area that WIZARD is being extended into is the estimation of thermodynamic flexibility. Since we can estimate the energy barrier between two conformers that differ by only one torsional bond and can use the criticism axioms to attempt to discover if the bond rotation process is likely to bring atoms into undesirable proximity, it follows that a variant on the traveling-salesman problem can be used to estimate the lowest energy path between two conformers. These are all projects that are simplified by the availability of higher level knowledge and the high level of representation that WIZARD uses. Several of these projects are currently under exploration, and the results will be published in future papers.

ACKNOWLEDGMENT

We thank Prof. J. Edgar Anderson for his time and many valuable comments. We also thank Dr. Tommy Liljefors for allowing us to use large portions of MIMIC in the current work. We thank Dr. C. W. von der Lieth for allowing us to use some of his FORTRAN subroutines originally written for the RINGS program for the display of molecules in WIZARD. We gratefully acknowledge the National Swedish Board for Technical Development (STU) for financial support.

REFERENCES AND NOTES

- (1) (a) Wiberg, K. B.; Boyd, R. H. "Application of Strain Energy Minimization to the Dynamics of Conformational Changes". *J. Am. Chem. Soc.* **1972**, *94*, 8426. (b) Burkert, U.; Allinger, N. L. *Molecular Mechanics*; ACS Monograph 177; American Chemical Society: Washington, DC, 1982; pp 72-76.
- (2) (a) Anet, F. A. L.; Krane, F. "Strain Energy Calculation of Conformations and Conformational Changes in Cyclooctane". *Tetrahedron Lett.* **1973**, 5029. (b) Burkert, U.; Allinger, N. L. "Pitfalls in the Use of the Torsion Angle Driving Method for the Calculation of Conformational Interconversions". *J. Comput. Chem.* **1982**, *3*, 40. (c) Osawa, E. "Mechanisms of Conformational Chirality Inversion in Bicyclo[4.2.1]nonan-9-one and Bicyclo[4.2.2]decane as Studied in Two-Parametric Torsional Energy Surfaces". *J. Comput. Chem.* **1982**, *3*, 400. (d) Ivanov, P. M.; Osawa, E. "Remarks on the Analysis of Torsional Energy Surfaces of Cycloheptane and Cyclooctane by Molecular Mechanics". *J. Comput. Chem.* **1984**, *5*, 307. (e) van de Graaf, B.; Baas, J. M. A. "The Lagrange Multiplier Method for Manipulating Geometries. Implementation and Applications in Molecular Mechanics". *J. Comput. Chem.* **1984**, *5*, 314.
- (3) (a) Hendrickson, J. B. "Molecular Geometry. VII. Modes of Interconversion in the Medium Rings". *J. Am. Chem. Soc.* **1967**, *89*, 7047 and previous papers in this series. (b) Still, W. C.; Galynker, I. "Chemical Consequences of Conformation in Macrocyclic Compounds". *Tetrahedron* **1981**, *23*, 3981. (c) Dyott, T. M.; Stuper, A. J.; Zander, G. S. "MOLY—An Interactive System for Molecular Analysis". *J. Chem. Inf. Comput. Sci.* **1980**, *20*, 28.
- (4) (a) Cohen, N. C.; Colin, P.; Lemoine, G. "SCRIPT: Interactive Molecular Geometrical Treatments on the Basis of Computer-Drawn Chemical Formula". *Tetrahedron* **1981**, *37*, 1711. (b) Cohen, N. C. "Drug Design in Three Dimensions". *Adv. Drug Res.* **1985**, *14*, 42.
- (5) De Clercq, P. J. "Systematic Conformational Analysis. A Microcomputer Method for the Semiquantitative Evaluation of Polycyclic Systems Containing Five-, Six- and Seven-Membered Rings". *Tetrahedron* **1984**, *40*, 3717, 3729.

- (6) Corey, E. J.; Feiner, N. F. "Computer-Assisted Synthetic Analysis. Rapid Computer Method for the Semiquantitative Assignment of Conformations of Six-Membered Ring Systems". *J. Org. Chem.* **1980**, *45*, 757, 765. The domain is limited to six-membered rings, and the method is used by a synthesis planning program (LHASA) to help decide the course of reactions on such rings.
- (7) (a) Liljefors, T.; Thelin, B.; van der Pers, J. N. C.; Löfstedt, C. "Chain-Elongated Analogues of a Pheromone Component of the Turnip Moth, *Agrotis segetum*. A Structure-Activity Study Using Molecular Mechanics". *J. Chem. Soc., Perkin Trans. 2* **1985**, 1957. (b) Liljefors, T.; Thelin, B.; van der Pers, J. N. C. "Structure-Activity Relationships between Stimulus Molecule and Response of a Pheromone Receptor Cell in Turnip Moth, *Agrotis Segetum*. Modifications of the Acetate Group". *J. Chem. Ecol.* **1984**, *10*, 1661.
- (8) Allinger, N. L.; Yuh, Y. H. "MM2: Molecular Mechanics II". *QCPE* **1980**, *11*, 395.
- (9) For an overview of AI, see: Barr, A., Feigenbaum, E. A., Eds. "The Handbook of Artificial Intelligence". William Kaufmann: Los Altos, CA. Volume II contains reviews of several well-known ES programs. For a recent book on chemical applications of AI, see: Pierce, T. H., Hohne, B. A., Eds. *Artificial Intelligence Applications in Chemistry*; ACS Symposium Series 306; American Chemical Society: Washington, DC, 1986.
- (10) Liljefors, T.; "MOLBUILD—An Interactive Computer Graphics Interface to Molecular Mechanics". *J. Comput. Chem.* **1983**, *1*, 111.
- (11) von der Lieth, C. W.; Carter, R. E.; Dolata, D. P.; Liljefors, T. "RINGS—a General Program to Build Ring Systems". *J. Comput. Chem.* **1984**, *2*, 117.
- (12) We currently use the Visual 500 series (Visual Technology, Inc., 540 Main Street, Tewksbury, MA) as our "standard" terminals. The "selective erase" feature and the possibility to create a scrolling region together with the graphics are the only enhancements beyond full PLOT10 compatibility.
- (13) Motherwell, W. D. S. *PLUTO—a Program for Plotting Molecular and Crystal Structures*; University Chemical Laboratory: Cambridge, 1978.
- (14) See ref 9, Vol. II, Section VIII.B.1, and references therein.
- (15) See, for example: Pospesel, H. *Introduction to Logic: Predicate Logic*; Prentice-Hall: Englewood Cliffs, NJ, 1976.
- (16) (a) Beckhaus, H.-D.; Hellmann, G.; Ruchardt, C. "Tetra-*tert*-butylethane". *Chem. Ber.* **1978**, *111*, 72. (b) Hounshell, W. D.; Dougherty, D. A.; Mislow, K. "On the Role of Newman Projections in Conformational Analysis. Evidence for Nonalternating Ground States in Highly Strained Molecules". *J. Am. Chem. Soc.* **1978**, *100*, 3149. (c) Osawa, E.; Shirahama, H.; Matsumoto, T. "Internal Rotation in Simple to Congested Hydrocarbons Including 2,3-Dimethylbutane, 1,1,2,2-Tetra-*tert*-butylethane, 2,2,4,4,5,5,7,7-Octamethylcyclotetane, and Cholestane". *J. Am. Chem. Soc.* **1979**, *101*, 4824.
- (17) Sacerdoti, E. D. *A Structure for Plans and Behavior*; Elsevier: New York, 1977.
- (18) Charniak, E.; Riesbeck, C. K.; McDermott, D. V. *Artificial Intelligence Programming*; Lawrence Erlbaum: Hillsdale, NJ, 1980.
- (19) In *n*-butyl chloride, the *g*⁺/*g*⁻ conformation makes up as much as 24% of the total: Ukaji, T.; Bonham, R. A. "The Molecular Structure of *n*-Butyl Chloride and *sec*-Butyl Chloride. II. Hindered Rotation". *J. Am. Chem. Soc.* **1962**, *84*, 3631.
- (20) See ref 10.
- (21) Beierbeck, H.; Saunders, J. K. "Conformational and Configurational Analysis of Hydrocarbon Chains Based on Time-Averaged Carbon-13 Chemical Shifts". *Can. J. Chem.* **1977**, *55*, 772.
- (22) Anderson, J. E.; Pearson, H. "A Secondary Effect on Barriers to Rotation in Substituted Ethanes". *J. Am. Chem. Soc.* **1975**, *97*, 764.
- (23) See, for example: Berg, U.; Liljefors, T.; Roussel, C.; Sandström, J. "Steric Interplay between Alkyl Groups Bonded to Planar Frameworks". *Acc. Chem. Res.* **1985**, *18*, 80. Hounshell, W. D.; Iroff, L. D.; Iverson, D. J.; Wroczynski, R. J.; Mislow, K. "Is the Effective Size of an Alkyl Group a Gauge of Dynamic Gearing?" *Isr. J. Chem.* **1980**, *20*, 65.
- (24) See, for example: Michalski, R. S.; Carbonell, J. G.; Mitchell, T. M., Eds. *Machine Learning: An Artificial Intelligence Approach*; Springer: Berlin, 1984; Vol. XI and previous volumes in this series.

COMPUTER SOFTWARE REVIEWS

CHEMBASE

JOHN C. MARSHALL

Professor of Chemistry, St. Olaf College, Northfield, Minnesota 55057

Received November 26, 1986

In the course of the past several years I have examined a great many programs and reviewed quite a number of them for various magazines. Most of the software has been very forgettable, but every now and then when working with a piece of software you realize that you are dealing with something special. CHEMBASE from Molecular Design Ltd. is one of the special ones.

CHEMBASE is a powerful data-base tool for the chemist, a tool that is strongly directed toward the synthetic organic chemist but certainly of use to any chemist charged with keeping track of organic molecules and organic reactions. To enumerate all of the capabilities of this program would require much more space than I am allowed to use for this review. I shall therefore mention only some of the features of the program with the hope that many who read this will take a careful look at this powerful software tool.

In brief, CHEMBASE is a data-base manager with a difference. The difference is that it is visually oriented in terms of chemical structures. Unlike the more traditional data-base managers that allow you to manipulate text and numbers, CHEMBASE allows you to store, manipulate, and search your data base in terms of chemical structures and substructures and even in terms of complete reactions. This means that a

synthetic chemist could search a data base for all entries containing a specified compound, a specified structural unit, a specified functional group conversion, or any combination of the above. In addition, the program also supports more conventional types of data-base interrogation like the retrieval of all compounds available with molecular weights in a certain range, boiling points in a certain range, or combinations connected by Boolean logic. Each file "hit" is returned as a complete frame of information (see, for an example, Figure 1) that has provision for containing physical constants, various registry numbers, alternate names, empirical formulas, and whatever else you or the data-base creator thought to include. Very complete default information templates are supplied, and the user is free to design information templates to suit his or her specific applications. Powerful facilities for manipulating lists of information are also provided. While such power has long existed on large mainframe computers, it is now available for microcomputers with very few important features sacrificed in the transition. Indeed, the added convenience of having such a powerful package available on a microcomputer for convenient interactive use makes the missing features seem quite unimportant. Perhaps the most important feature missing is chirality. While CHEMBASE allows chirality to be repre-