5081

Computer-Assisted Mechanistic Evaluation of Organic Reactions. 22. The Generation and Use of Three-Dimensional Structures

Scott A. Gothe,[†] Harold E. Helson, Iordanis Houdaverdis, Ingvar Lagerstedt, Shenna Sinclair, and William L. Jorgensen[*]

Department of Chemistry, Yale University, New Haven, Connecticut 06511

The CAMEO computer program, which predicts the products of organic reactions, has been enhanced to include the generation and analysis of three-dimensional molecular structures. This required the development of a new interface for molecular visualization and manipulation. The conversion of two-dimensional atom and bond tables to three-dimensional representations has been accomplished by fast molecular modeling techniques originally implemented in the STRFIT program of Saunders. Along with providing structural information directly to the user, the module generates perceptual data based on spatial analysis of molecules. This information can then be used in the refined evaluation of the likelihood of reactions. Examples are provided for carbocation rearrangements and reductions of ketones and C-C double bonds.

I. Introduction

CAMEO is an interactive computer program which has been developed to predict the products of organic reactions given starting materials and reaction conditions as input.[1] Chemical transformations are typically processed by using mechanistic rules and logic to determine the course of a reaction *a priori*, rather than employing large databases of possible known reactions.[2] The development of concise mechanistic rules which map the vast landscape of organic reactivity has been an ongoing task within the CAMEO project. These rules are implemented in mechanistic "modules" covering ranges of reactions, including base-catalyzed and nucleophilic,[3] acid-catalyzed and electrophilic,[4] heterocyclic,[5] pericyclic,[6] free radical,[7] and redox[8] chemistry.

Enhanced understanding of organic reactivity requires the consideration of molecular structure in three-dimensional space. For any chemical transformation, the mere presence of a reactive site or functional group in a molecule is not usually sufficient in itself for the organic chemist to determine if a reaction will indeed occur at that site. A site needs to be sterically accessible to an incoming reagent, or a reaction may occur only if certain molecular conformations are attainable (*cf.* electrocyclic rearrangements). For many years, chemists have employed physical molecular models such as Dreiding, ball-and-stick, and CPK to examine such structural features of reactants. This paper addresses the development of a new module for the CAMEO program which allows the user both to explore the three-dimensional features of molecules on the computer screen and to include optionally this information in mechanistic processing. To begin, the current state of steric analysis within CAMEO and the incorporation of three-dimensional information into the reaction processing scheme are discussed. The implementation and operation of the graphic interface are outlined next, together with the development of algorithms to generate three-dimensional structures and enhanced perception information. Then, several examples of the use of this information to refine and develop new mechanistic rules are covered.

II. Background

The treatment of steric effects on reactivity in the CAMEO program has largely been limited to simple connectivity information and analysis of bond types and stereochemistry along paths between atoms forming a ring.[3c] A recent exception has been the development of stereochemical analysis within the nucleophilic module. Here, a general implementation[9] of Cram's rule[10] has required substituents to be ranked according to their relative sizes with an algorithm that predicts Taft's steric parameters, E_s.[11] The use of E_s values is now spreading to other mechanistic modules of CAMEO for quantification of steric effects.

However, general treatment of conformation-dependent stereochemical features requires the ability to create and manipulate three-dimensional molecular models. First, a graphic interface is required to visualize and manipulate molecules in three dimensions. Secondly, a method to

† Current Address: TRIPOS Associates, Inc., 1699 South Hanley Road, Suite 303, St. Louis, MO 63144.

(1) For a review, see: Jorgensen, W. L.; Laird, E. R.; Gushurst, A. J.; Fleischer, J. M.; Gothe, S. A.; Helson, H. E.; Paderes, G. D.; Sinclair, S. *Pure Appl. Chem.* 1990, 62, 1921.

(2) For a review, see: Barone, R.; Chanon, M. In *Computer Aids to Chemistry*; Vernin, G., Chanon, M., Eds.; Ellis Horwood Ltd.: West Sussex, England, 1986; p 19.

(3) (a) Salatin, T. D.; Jorgensen, W. L. *J. Org. Chem.* 1980, 45, 2043. (b) Salatin, T. D.; McLaughlin, D. R.; Jorgensen, W. L. *J. Org. Chem.* 1981, 46, 5284. (c) Peishoff, C. E.; Jorgensen, W. L. *J. Org. Chem.* 1985, 50, 1056. (d) Gushurst, A. J.; Jorgensen, W. L. *J. Org. Chem.* 1986, 51, 3513. (e) Metivier, P.; Gushurst, A. J.; Jorgensen, W. L. *J. Org. Chem.* 1987, 52, 3724. (f) Gushurst, A. J.; Jorgensen, W. L. *J. Org. Chem.* 1988, 53, 3397.

(4) (a) McLaughlin, D. R. Ph.D. Dissertation, Purdue University, 1983. (b) Bures, M. G.; Roos-Kozel, B. L.; Jorgensen, W. L. *J. Org. Chem.* 1985, 50, 4490.

(5) Bures, M. G.; Jorgensen, W. L. *J. Org. Chem.* 1988, 53, 2504.

(6) (a) Burnier, J. S.; Jorgensen, W. L. *J. Org. Chem.* 1983, 48, 3923. (b) Burnier, J. S.; Jorgensen, W. L. *J. Org. Chem.* 1984, 49, 3001. (c) Paderes, G. D.; Jorgensen, W. L. *J. Org. Chem.* 1992, 57, 1904.

(7) (a) Laird, E. R.; Jorgensen, W. L. *J. Org. Chem.* 1990, 55, 9. (b) Laird, E. R.; Jorgensen, W. L. *J. Chem. Inf. Comput. Sci.* 1990, 30, 458.

(8) (a) Paderes, G. D.; Jorgensen, W. L. *J. Org. Chem.* 1989, 54, 2058. (b) Paderes, G. D.; Metivier, P.; Jorgensen, W. L. *J. Org. Chem.* 1991, 56, 4718.

(9) Fleischer, J. M.; Gushurst, A. J.; Jorgensen, W. L. Manuscript in preparation.

(10) (a) Cram, D. J.; Abd Elhafez, F. A. *J. Am. Chem. Soc.* 1952, 74, 5828. For similar rules, see: (b) Karabatsos, G. J. *J. Am. Chem. Soc.* 1967, 89, 1367. (c) Boyd, D. R.; McKervey, M. A. Q. *Rev. Chem. Soc.* 1968, 22, 95. (d) Chérest, M.; Felkin, H.; Prudent, N. *Tetrahedron Lett.* 1968, 2199.

(11) Taft, R. W. In *Steric Effects in Organic Chemistry*; Newman, M. S., Ed.; Wiley: New York, 1956; Chapter 13 and references cited therein.
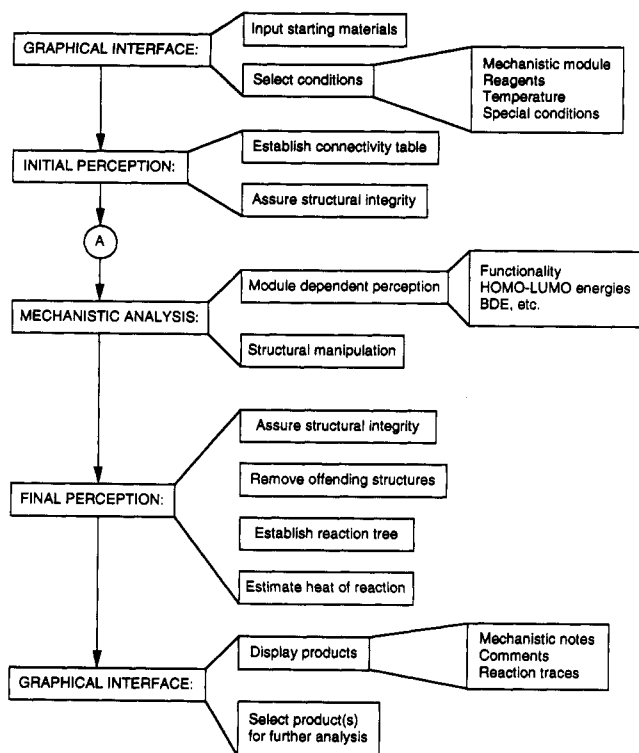
Figure 1. Previous CAMEO reaction processing diagram.

transform existing two-dimensional drawings of structures into adequate three-dimensional (3D) representations is needed. This must operate quickly with a minimum of user involvement in order to maintain the interactive environment of CAMEO. These first two stages represent a "virtual molecular modeling kit". Subsequent additions are planned to provide QSAR data based on the 3D structures and to expand the input/output options so that data from calculational programs such as Sybyl[12] or MacroModel[13] or reaction library programs like REACCS[14] may be shared and incorporated into CAMEO. The last element, creating three-dimensional "perception sets", is the most important for current development of CAMEO. This information is the basis for refinements of the rules for predicting reactivity and selectivity.

Reaction processing in CAMEO can be divided into three principal segments (Figure 1). The first is a graphical interface that allows the user to enter structures and conditions necessary for the reaction. During the structure perception phase, a connectivity table is established for the substrate, and structural information is created for use by the mechanistic modules. Finally, in the mechanistic phase, modules evaluate the structure and form products based on suites of precedent-based rules. Products are reperceived to assure structural integrity and are displayed by the graphical interface. They may then be subjected to further reactions to create full synthetic sequences.

As shown in Figure 2, selection of three-dimensional processing is currently optional. It is intended to supplement the two-dimensional reaction-processing scheme.
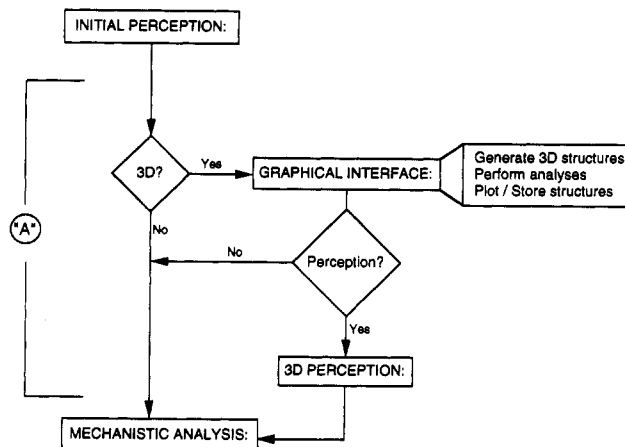

Figure 2. Three-dimensional reaction processing segment.

If three-dimensional analysis is desired, the chosen substrate molecules are displayed by the new three-dimensional graphical interface following the initial perception. At this stage, a conformer can be readily generated by the user, and three-dimensional perception sets are created. A different conformer can then be considered in a subsequent pass through the program. The reactants are sent for mechanistic analysis with this additional information. If, however, this information is not desired, the user may return to the initial graphical interface, the sketch menu, and continue with normal processing.

III. Implementation

A. Interface and Graphics. The general appearace and function of the 3D interface (Figure 3) is patterned after the sketch menu in CAMEO.[1] While options exist in the three-dimensional module plotting box for the translation and manipulation of the spatial coordinates of the molecule, no provision is made to permanently change the CAMEO atom and bond (connection) table for the molecule. In fact, all processing operates on a virtual structure that is derived from the current atom and bond table in the reaction tree. The virtual structure is deleted upon exit from the module; however, any perceptual information obtained is applied to the current molecule and augments the information derived from two-dimensional premechanistic perception.

The menu options are grouped in several categories indicated by the bold lettering in Figure 3. The buttons are activated by a graphical pointing device.[15] Their purposes are generally obvious, though a few need some clarification. The Push and Pull buttons move individual atoms by a set amount (0.5 Å) along a designated principal coordinate axis. This function is used to guide an optimization towards a desired conformer. The addition and deletion of implicit hydrogen atoms and electron lone pairs for the molecule in the plotting box are accomplished *via* the H Add and H Del buttons. These options only affect the 3D perception sets being manipulated. The

(12) Available from TRIPOS Associates, Inc. 1699 S. Hanley Road, Suite 303, St. Louis, MO, 63144.

(13) Mohamadi, F.; Richards, N. G. J.; Guida, W. C.; Liskamp, R.; Lipton, M.; Caufield, C.; Chang, G.; Hendrickson, T.; Still, W. C. *J. Comput. Chem.* **1990**, *11*, 440.

(14) Available from Molecular Design Limited, 2132 Farallon Drive, San Leandro, CA, 94577.

(15) CAMEO is written in FORTRAN and is developed on Digital Equipment Corp. computers using the VMS operating system. Versions have also been created for use on platforms which support UNIX or AIX operating environments. The graphics interface requires a Textronix 4010, 4208, or 4100 series terminal or emulation. All graphics developed must, however, conform to 4010 conventions, which obviates the use of native graphic polygon and fill patterns available on higher series terminals for any hidden line or surface algorithms.
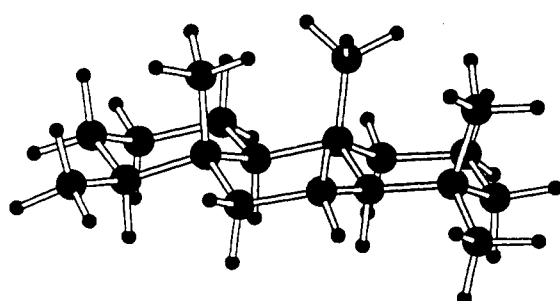
Plotting Box		
	Molecule Translation	Rotate X Y Z Push Pull Scale
	Display Modes	Wireframe Ball & Stick SpaceFill Label Stereo Hide Lone Pairs
	Molecular Information	Distance Bond Angle Dihedral Angle Annote
	Molecular Manipulation	H Add H Del Optimize 3D Perception Help
Volume Area Log P	Undo Opt. Redo Opt.	Reset Wipe Get Refresh ▷Plot Store Sketch.React Stop
QSAR	**Screen Functions**	**I/O** **Program Control**

Figure 3. 3D menu and general function areas.

newly added hydrogen atoms are not stored in the 2D atom and bond table. Implicit hydrogens and lone pairs do, however, influence molecular geometry calculations and are necessary for generating proper 3D perception set information. Molecular geometry calculations to generate an initial 3D structure or to refine a current 3D structure are performed by selecting the Optimize button. The 3D Perception button causes display of favored steric approach vectors, charts predicting the relative reactivities of various reagents, and also raw values for steric congestion and torsion congestion (*vide infra*). The Input/Output section of the menu allows a structure to be saved or retrieved as either a three-dimensional CAMEO storage structure (.3CSS), a file for use by an in-house plotting program, MindTool (.MND),[16] a Brookhaven Protein Data Bank file (.PDB),[17] or a Molecular Design MOLFile (.MOL).[18] Finally, selection of the React button sends

Table I. Programs for 2D to 3D Structure Conversions

program name	method employed	ref
MOLBUL	force field	19
BUILD3D	FF/distance geometry	20
STRFIT	FF/distance geometry	21
WIZARD	template	22
COBRA	template	23
AIMB	template	24
CONCORD	FF/template	25
MOLGEO	FF/distance geometry	26

the reactants on for reaction analysis with the 3D perception sets.

While in the three-dimensional module, two different groups of coordinates and molecular information are required to describe the molecule in the plotting box. The first set is a common block of arrays which contains only the Cartesian coordinates of the atomic centers as double precision real variables (64 bits), the atom type descriptors, and connectivity arrays necessary for all numerical calculations. The second set of arrays is used by the display routines and contains the same coordinate information for the atomic centers as the first, though as single precision real numbers. This "display set" also includes arrays for the coordinates of bond and atom segments for the depiction of the molecule in a ball and stick representation.

B. Structure Conversion to Three Dimensions. Several programs have been reported in the literature which create reasonable three-dimensional representations of molecules from two-dimensional structural formulas (Table I). Each program relies on either structural templates or force field calculations to perform the transformation. The template approach examines a user-supplied two-dimensional structure, or portions of a structure, and maps it as closely as possible onto a database of previously generated three-dimensional representations

(16) Tirado-Rives, J.; Blake, J. F.; Jorgensen, W. L., Yale University, New Haven, CT 06511.

(17) (a) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. *J. Mol. Biol.* 1977, *112*, 535. (b) Abola, E. E.; Bernstein, F. C.; Bryant, S. H.; Koetzle, T. F.; Weng, J. In *Crystallographic Databases—Information Content, Software Systems, Scientific Applications*; Allen, F. H., Bergerhoff, G., Sievers, R., Eds.; Data Commission of the International Union of Crystallography: Bonn/Cambridge/Chester, 1987; p 107.

(18) Dalby, A.; Nourse, J. G.; Hounshell, W. D.; Gushurst, A. K. I.; Grier, D. L.; Leland, B. A.; Laufer, J. *J. Chem. Inf. Comput. Sci.* 1992, *32*, 244.

(19) Kao, J.; Eyermann, C.; Watt, L.; Maher, R.; Leister, D. *J. Chem. Inf. Comput. Sci.* 1985, *25*, 400.

(20) Wenger, J. C.; Smith, D. H. *J. Chem. Inf. Comput. Sci.* 1982, *22*, 29.

(21) Saunders, M.; Jarret, R. M. *J. Comput. Chem.* 1986, *7*, 578.

(22) Dolata, D. P.; Leach, A. R.; Prout, K. *J. Comput.-Aided Mol. Design* 1987, *1*, 73.

(23) Leach, A. R.; Prout, K. *J. Comput. Chem.* 1990, *11*, 1193.

(24) Wipke, W. T.; Hahn, M. A. *ACS Symp. Ser.* 1986, *306*, 136.

(25) Pearlman, R. S. *Chem. Design Auto. News* 1987, *2*, 1.

(26) Gordeeva, E. V.; Katrizky, A. R.; Shcherbukhin, V. V.; Zefirov, N. S. *J. Chem. Inf. Comput. Sci.* 1993, *33*, 102.

or templates. The templates which best fit the input structure are then assembled into a final structure which is output. This pattern matching and assembly approach is computationally very fast; however, the required template databases may be incomplete and require storage. Force-field routes employ molecular mechanics, which requires parameterization of the potential energy equations to reproduce the essential features of molecular geometry. These parameters, which are usually assigned by atom or bond type and hybridization, can be applied repeatedly to each atom or bond of similar type throughout a molecule. If appropriate templates are not included in the database, force-field methods permit molecules that may not be treatable by a template method to be studied. Though the generality of the force-field approach is desirable for the purposes of CAMEO, traditional force-field methods such as MM2[27a] or MM3[27b] are computationally intensive and may not be suitable in an interactive environment unless the molecules are very simple. Fortunately, Saunders has developed a simplified molecular mechanics approach that rapidly generates reasonable molecular structures.[21] The associated program, STRFIT, was originally reported to generate molecular conformations which are very close to MM2 results at computation speeds between 2 and 16 times faster than MM2. Though not explicitly designed for the purpose of converting a molecular representation from two to three dimensions, the STRFIT algorithm was found in our laboratory to produce accurate molecular conformations from many planar input structures of simple organic molecules that were randomly selected from Saunders' original paper.[21] This combination of generality, speed, and accuracy is suitable for our purposes.

The conversion of the STRFIT algorithm into subroutines for the CAMEO program was accomplished in the following manner. The original "turtle" drawing algorithm, STR,[28] was removed, and preprocessing code that translates CAMEO atom and bond tables to the appropriate information needed for optimization was substituted. This preprocessor, SETUP3D, first defines the atomic hybridization sets from connectivity information that is used to define the atom types for the optimization routine. The atom types, SATYPE, that are currently supported are shown in Table II. Appropriate connectivity arrays are then generated, and initial three-dimensional coordinates are assigned by translation and scaling of the existing screen (pixel) coordinates. Lastly, SETUP3D calls several subroutines to generate display information and assign atomic radii for the QSAR calculation phase.

Normally, a CAMEO structure does not have all hydrogens explicitly represented. Since all hydrogens and electron lone pairs need to be explicitly added and placed at reasonable locations in space for refined molecular mechanics calculations, they are generated for the molecule and placed in the 3D atom and bond tables. With these additions, CAMEO's normal 64 atom limit has been extended to 200 atoms in the 3D module. For a molecule read from an external file (PDB or MOLFile), CAMEO decides which hydrogens should be regarded as implicit, and those implicit hydrogens are then deleted from the structure when the user leaves the 3D module.

Table II. LENGTH3D Atomic Types and CAMEO Equivalents

SATYPE	name	currently supported	CAMEO atom type	CAMEO elm. no.
1	C (sp³)	yes	C	4
2	H	yes	H	1
3	O (sp³)	yes	O	2
4	N (sp³)	yes	N	3
5	Cl	yes	Cl	6
6	B	yes	B	9
7	Br	yes	Br	7
8	F	yes	F	5
9	I	yes	I	8
10	S (bo = 2)	yes	S	12
11	P	yes	P	11
12	Si	yes	Si	10
13	C (sp²)	yes	C	4
14	C (sp)	yes	C	4
15	O (sp²)	yes	O	2
16	N (sp²)	yes	N	3
17	N (sp)	yes	N	3
18	S (sp²)	yes	S	12
19	P (sp²)	yes	P	11
20	S (bo = 4)	yes	S	12
21	S (bo = 6)	yes	S	12
22	B (trigonal)	no		
23	lone pairs "O"	yes	LP	64
24	C+	yes	C	4
25	Si (sp²)	yes	Si	10
26	C (c-prop)	yes	C	4
27	lone pairs "N"	yes	LP	64
28	N+ (quat)	yes	N	3

The final stage of STRFIT conversion involved adaptation of the LENGTH3D subroutine from Saunders' structure optimization code. The current algorithm is shown in Figure 4. The routine was modularized to easily interface the calculational results with the three-dimensional graphic routines. The LENGTH3D routine has undergone some significant changes since the original version of STRFIT was described.[21] Foremost, the dependency of the routine on a table of approximate covalent radii to generate "ideal" bond lengths and 1,3-distances has been eliminated. Instead, a table of ideal distances is already present in the subroutine. If a value for two atom types does not exist in the table, default values for bond lengths and angles are substituted. Other changes to the original algorithm recently made by Saunders include the removal of both gradient acceleration step analysis in the structure optimization loop and intermediate pseudoenergy evaluations.[29] We have also imposed bounds on the absolute magnitude of atomic displacements for each optimization step.

The new LENGTH3D routine was first tested to determine its execution speed for 26 simple molecules (see Table III). These molecules were individually input to the 3D module by selecting 3D on the sketch menu. Geometry optimization was first performed only on the non-hydrogen atoms by selecting the Optimize button. Implicit hydrogen atoms and lone pairs of electrons were then added via the H Add button, and the structure was reoptimized. The timing data in Table III are for a VaxStation 3500.[30] The final geometries resulting from each optimization step were determined to be within reasonable limits by examination of all bond lengths, angles, and dihedral angles.[21] However, both cyclobutane and cyclopentane exhibited planar structures after both

(27) (a) Allinger, N. L. J. Am. Chem. Soc. 1977, 99, 8127. (b) Allinger, N. L.; Chen, K.; Rahman, M.; Pathiaseril, A. J. Am. Chem. Soc. 1991, 113, 4505.

(28) Available as QCPE no. 488, Quantum Chemistry Program Exchange, Indiana University, Bloomington, IN 47405.

(29) We thank Prof. M. Saunders for the updated version of the optimization routine.

(30) Tests performed on a VaxStation 3500 with 16 megabytes of memory and measured using embedded VAX-FORTRAN timer calls.
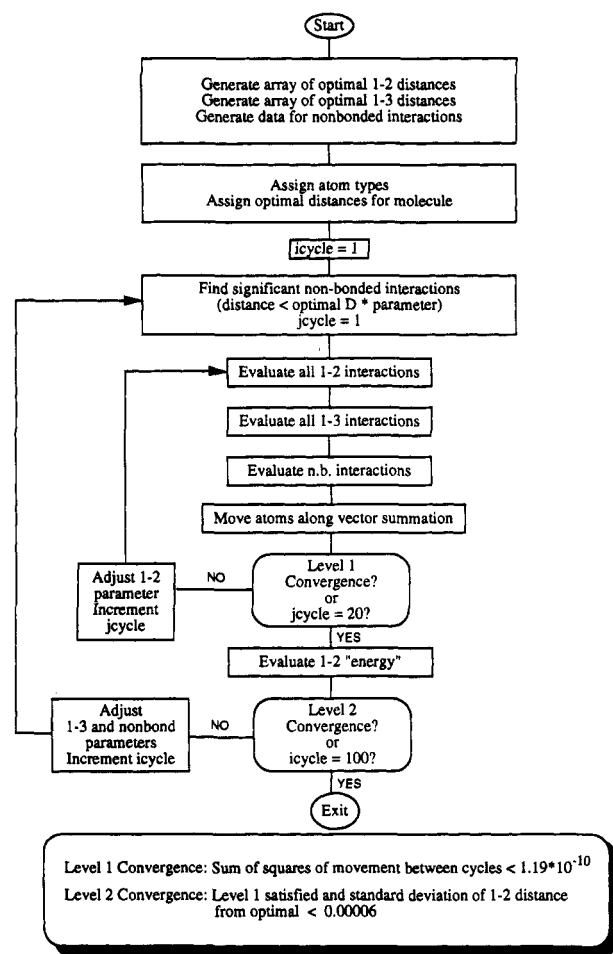
Figure 4. Algorithm of LENGTH3D.

Table III. Timing Data for the LENGTH3D Algorithm on a VaxStation 3500

compd	heavy atoms CPU time[a,b] (s)	all atoms CPU time[a] (s)	total CPU time[a] (s)
propane	0.04	1.84	1.88
butane	0.09	2.76	2.85
pentane	0.17	2.76	2.93
isopentane	0.33	6.51	6.84
neopentane	0.42	6.88	7.30
hexane	0.20	4.02	4.22
cyclopropane	0.03	1.21	1.24
cyclobutane[c]	0.24	3.43	3.67
cyclopentane[c]	0.31	3.43	3.74
cyclohexane	0.71	6.30	7.01
cycloheptane	1.15	11.96	13.11
cyclooctane	0.35	12.45	12.80
cis-2-butene	0.34	4.73	5.07
trans-2-butene	0.16	4.46	4.62
cyclohexene	1.14	10.65	11.79
1,3-cyclohexadiene	1.29	5.53	6.82
1,4-cyclohexadiene	1.57	7.05	8.62
trans-decalin	1.30	11.98	13.28
linear triquinane	2.49	16.39	18.88
angular triquinane	2.99	17.77	20.76
tetrahydropyran	0.70	10.40	11.10
4-methylcyclohexanone	1.88	15.22	17.10
propanoic acid	0.50	7.42	7.92
propylamine	0.09	2.65	2.74
propanamide	0.99	10.02	11.01
2-iminoylbutane	0.56	5.83	6.39

[a] Reference 30. [b] The first time the LENGTH3D module is invoked in each CAMEO session, approximately an additional 3 CPU-s is used to allocate memory. [c] Exhibit planar structures after optimization.

optimization phases, and attempts to obtain bent or envelope conformations of these structures by deforming

the initially minimized structure toward the desired conformation and reoptimizing failed.

In general, the CPU times required for optimization of a molecule from a planar input structure are within practical limits for use in the interactive environment of CAMEO. A large molecule, like a steroid, takes 5–10 min to optimize. These times could feasibly be shortened by starting the optimization on a nonplanar input structure which lies closer in geometry to the final desired conformer. For example, a flat cyclohexane ring may be converted into a rough chair or boat conformation through the Push and Pull options, though such operations can be tedious for large structures. To further shorten optimization time, future plans for the three-dimensional module include incorporating templates for common ring systems in the preprocessing phase. Many faster workstations than the VaxStation 3500 are also now available.

To further explore the accuracy of the optimization routine, a more complete test of the LENGTH3D algorithm was performed using the molecules shown in Figure 5. The accuracy of these optimizations was gauged by computing the root mean square (rms) deviation between structures generated by the LENGTH3D routine and those generated by MM3 or Amber calculations as implemented in MacroModel.[13] The structures were reoptimized a few times to ensure convergence for the LENGTH3D-generated geometries, and the resulting structure was saved as a PDB file. This file was subsequently used as an input file for MacroModel. Minimization of the structure to the nearest local energy minimum was then performed using either the MM3 or Amber force field, and the resulting structure was saved. These minimized structures were then compared to the original structure employing the match option in Midas.[31] Average rms deviations in Å over all atom positions were then calculated for LENGTH3D vs MM3, LENGTH3D vs Amber, and MM3 vs Amber (Table IV). The LENGTH3D routine provides a geometrically acceptable three-dimensional structure in most cases with an rms deviation less than 0.2 Å. Since LENGTH3D predicts less puckered geometries for compounds containing four-membered rings or nonbridged five-membered rings, these systems (16, 20, 21, 34, 48–51) show larger deviations between LENGTH3D and the other force fields. There are also deviations for some of the cyclohexene and methylene cyclohexane compounds (43–47) where the LENGTH3D algorithm again predicts less puckered rings. Nevertheless, as shown in the following section, the optimization routine generates molecular conformations that serve as reasonable models for spatial analysis and mechanistic processing. Refinements of the present algorithm will continue to be pursued. One step that improved the quality of the structures generated by the LENGTH3D routine was to increase the reference bond angle for sp²-carbons from 120 to 130°. This forces ketones in ring-strained cyclic compounds to become more planar, which improves the prediction below for the correct stereochemistry in reduction and addition reactions.

The storage and retrieval of the 3D structures is normally accomplished by the .3CSS files. The information in these files is shown in Table V.

C. CAMEO Perception in Three Dimensions. Given a reasonable molecular geometry for a substrate, the

(31) Ferrin, T. E.; Huang, C. C.; Jarvis, L. E.; Langridge, R. *J. Mol. Graphics* **1988**, *6*, 13.
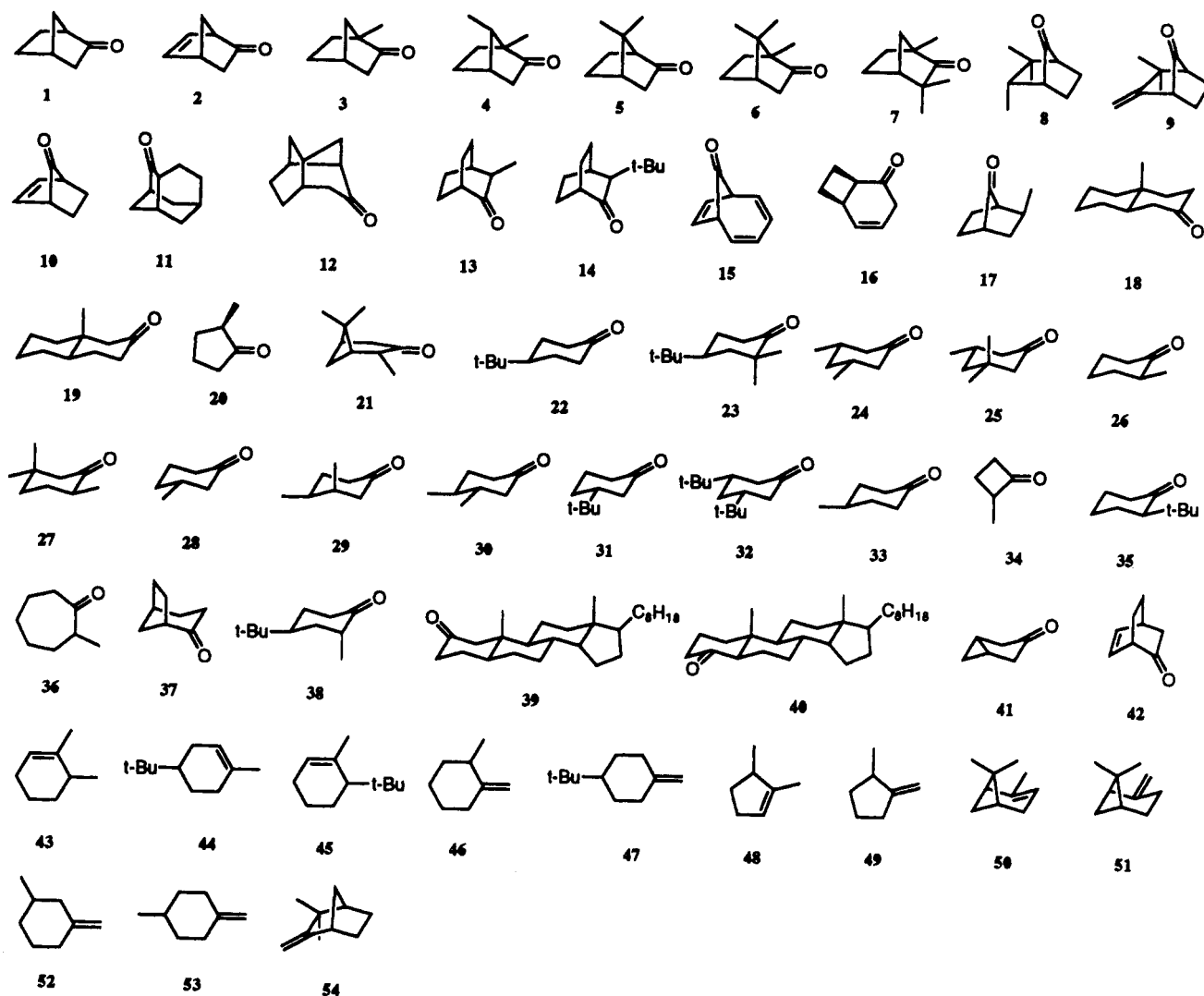
Figure 5. Molecules studied for reductions.

determination of which three-dimensional structural features are required to describe the reactivity of that substrate is an important problem in computer-assisted organic synthesis. Three-dimensional features are made available to the mechanistic modules of CAMEO by the same manner employed in evaluating molecular topology, namely that of perception sets. A CAMEO perception set contains 64 bits, with bits turned on for the indices of the atoms or bonds in the substrate(s) which display the property of the set.[3a] Two-dimensional perception sets based primarily on topology and connectivity are used throughout CAMEO to describe functional groups, rings, stereochemistry, and reactive sites.[1] Unlike these two-dimensional perception sets, the creation of three-dimensional perception sets must consider conformationally dependent molecular features. All spatial perception is performed within the PRCP3D subroutine, which operates by first identifying specific molecular features by topological analysis and then evaluating the spatial environment of those features. Three-dimensional perception sets are stored within a single common block, PRCP3D, which can be queried by the reaction modules. Such perception sets are either module-specific or may be employed generally by different modules. The initial development of perception sets for use by the Acidic and Redox modules is outlined here.

Scheme I. Wagner–Meerwein 1,2-Shift of the Cation Generated from Isoborneol

Isoborneol   Camphene
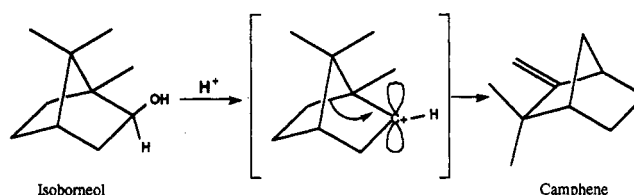
In the first example, module-specific spatial perception sets are used to assist in the evaluation of migratory aptitudes in the Wagner–Meerwein 1,2-shift, (*e.g.*, Scheme I). In these reactions, greater overlap of the filled σ orbital of a β-alkyl group (or hydrogen) with the empty 2p orbital of the cation facilitates rearrangement. The optimal orientation for the migrating substituent is therefore either syn- or antiperiplanar to the 2p orbital, *i.e.*, making a dihedral angle of ca. 0° ± 30° or 180° ± 30° for syn and anti migrations, respectively. Rearrangements may occur over a somewhat larger range of dihedral angles, but substituents oriented between 60° and 120° (240° and 300°) are essentially orthogonal to the orbital of the cation, and these migrations are disfavored. For reaction analysis in CAMEO, a group of three-dimensional perception sets have been created containing those atoms whose dihedral

Table IV. Comparison of Structural Results from Molecular Mechanics Calculations[a]

compd	LENGTH3D-MM3	LENGTH3D-Amber	MM3-Amber	compd	LENGTH3D-MM3	LENGTH3D-Amber	MM3-Amber
1	0.06	0.08	0.04	29	0.09	0.09	0.04
2	0.09	0.08	0.03	30	0.13	0.12	0.04
3	0.06	0.08	0.04	31	0.16	0.08	0.11
4	0.08	0.09	0.04	32	0.21	0.10	0.14
5	0.13	0.14	0.04	33	0.06	0.07	0.04
6	0.11	0.13	0.04	34	0.12	0.14	0.06
7	0.12	0.12	0.04	35	0.20	0.21	0.06
8	0.09	0.09	0.06	36	0.07	0.08	0.05
9	0.16	0.17	0.09	37	0.10	0.14	0.05
10	0.07	0.05	0.03	38	0.40	0.37	0.05
11	0.04	0.05	0.03	39	0.18	0.18	0.14
12	0.08	0.10	0.04	40	0.19	0.21	0.15
13	0.18	0.17	0.04	41	0.18	0.09	0.17
14	0.15	0.16	0.07	42	0.06	0.09	0.03
15	0.41	0.40	0.04	43	0.67	0.65	0.04
16	0.83	0.79	0.09	44	0.79	0.83	0.10
17	0.09	0.10	0.04	45	0.59	0.36	0.45
18	0.08	0.10	0.04	46	0.33	0.27	0.09
19	0.11	0.12	0.05	47	0.29	0.26	0.08
20	0.43	0.40	0.05	48	0.36	0.32	0.07
21	0.31	0.29	0.05	49	0.48	0.41	0.43
22	0.19	0.18	0.05	50	0.19	0.20	0.05
23	0.16	0.13	0.06	51	0.33	0.26	0.12
24	0.07	0.07	0.04	52	0.14	0.11	0.07
25	0.12	0.12	0.03	53	0.17	0.13	0.07
26	0.09	0.10	0.04	54	0.15	0.14	0.06
27	0.14	0.14	0.04	av	0.21	0.19	0.08
28	0.06	0.07	0.04				

[a] Given as rms values between three different pairs of force fields. LENGTH3D geometries generated within CAMEO. MM3 and Amber geometries calculated in MacroModel with the prcg search algorithm. The LENGTH3D geometries were used as starting points for the MM3 and Amber calculations. Comparison of the structures was done in Midas.

Table V. Entries in .3CSS Storage Files

variable	description	dimension
NATOMS	no. of atoms in table	integer
NBOND_3D	no. of bonds in table	integer
BATOMS	no. of atoms for the corresponding 2D table	integer
NBONDS	no. of bonds for the corresponding 2D table	integer
ATOMSA	set of all BATOMS	real*8(set)
BONDSA	set of all NBONDS	real*8(set)
ATOMX_3D	X-coordinate (pixels)	integer*2
ATOMY_3D	Y-coordinate (pixels)	integer*2
ATMCHG_3D	charge on atom	integer*2
ATMTYP_3D	type of atom	integer*2
BNDAT1_3D	first atom in bond	integer*2
BNDAT2_3D	second atom in bond	integer*2
BNDORD_3D	bond order	integer*2
BNDSTR_3D	stereocode of bond	integer*2
X3D	X-coordinate in 3-space	real*8
Y3D	Y-coordinate in 3-space	real*8
Z3D	Z-coordinate in 3-space	real*8
SFACT	bond factor for 3D	real*8

angles with respect to an incipient cation fall within the limits for migration.

The algorithm for assigning migratory aptitudes appears in Figure 6. After finding a cationic site in the molecule, a pseudoatom is placed along a vector normal to the trigonal plane and the dihedral angles between this atom and all possible migrating substituents are then calculated. Any atom which can migrate is placed in one of three sets, WAG30, WAG60, and WAG90. Cutoff ranges are made in increments of 30° such that dihedral angles with the 2p orbital of 0–30° yield the set of best possible migrators, WAG30; 31–61° is the next best, WAG60; and from 61 to 90° is disfavored, WAG90. No provision is made to favor antarafacial over suprafacial migrations. The success of this simple analysis and the application of this information to mechanistic rule development are covered in the next section.

General three-dimensional perception sets are those which can be applied for more than one mechanistic use. These sets may contain general spatial information about the molecule or the algorithms used in their creation may be seminal to the development of information which can be employed in more than one way. An example is the development of perception sets which are used by the Reduction module to evaluate the stereochemical results of nucleophilic or electrophilic attack at an unsaturated site. If such a reaction proceeds under steric approach control,[32,33] the more favored face for attack is less hindered.[34] Before perception sets are created, the unsaturation and its two stereofacial sides must be identified in three-dimensional space. The steric environments for each side of the unsaturation must be evaluated and compared to decide which face is more accessible. This problem is not new in computational modeling, and several viable procedures have previously been described.[35]

The algorithm for treating this problem in CAMEO appears in Figure 7. Following identification of an

(32) Dauben, W. G.; Fonkin, G. J.; Noyce, D. S. *J. Am. Chem. Soc.* **1956**, *78*, 2579.

(33) Dauben, W. G.; Blantz, E. J.; Jiu, J.; Micheli, R. A. *J. Am. Chem. Soc.* **1956**, *78*, 3752.

(34) In addition to ref 10, many other steric and torsional models have been posited in the literature. See: (a) Richer, J. C. *J. Org. Chem.* **1965**, *30*, 324. (b) Marshall, J. A.; Carroll, R. D. *J. Org. Chem.* **1965**, *30*, 2748. (c) Chérest, M.; Felkin, H. *Tetrahedron Lett.* **1968**, *18*, 2205. (d) Pasto, D. J.; Gontarz, J. A. *J. Am. Chem. Soc.* **1971**, *93*, 6909. (e) Ashby, E. C.; Yu, S. H.; Roling, P. V. *J. Org. Chem.* **1972**, *37*, 1918. (f) Klein, J. *Tetrahedron Lett.* **1973**, *44*, 4307. (g) Klein, J. *Tetrahedron* **1974**, *30*, 3349. (h) Ashby, E. C.; Laemmle, J. T. *Chem. Rev.* **1975**, *75*, 521. (i) Ashby, E. C.; Boone, J. R. *J. Org. Chem.* **1976**, *41*, 2890. (j) Anh, N. T. *Top. Curr. Chem.* **1980**, *88*, 145. (k) Cieplak, A. S. *J. Am. Chem. Soc.* **1981**, *103*, 4540. (l) Cieplak, A. S.; Tait, B. D.; Johnson, C. R. *J. Am. Chem. Soc.* **1989**, *111*, 8447. (m) Paddon-Row, M. N.; Rondan, N. G.; Houk, K. N. *J. Am. Chem. Soc.* **1982**, *104*, 7162.

(35) (a) Wipke, W. T.; Gund, P. *J. Am. Chem. Soc.* **1974**, *96*, 299. (b) Wipke, W. T.; Gund, P. *J. Am. Chem. Soc.* **1976**, *98*, 8107. (c) Orsini, F.; Sello, G. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 451 and references cited therein. (d) Lagerstedt, I. C.; Olsson, T., *J. Chem. Inf. Comput. Sci.*, in press.
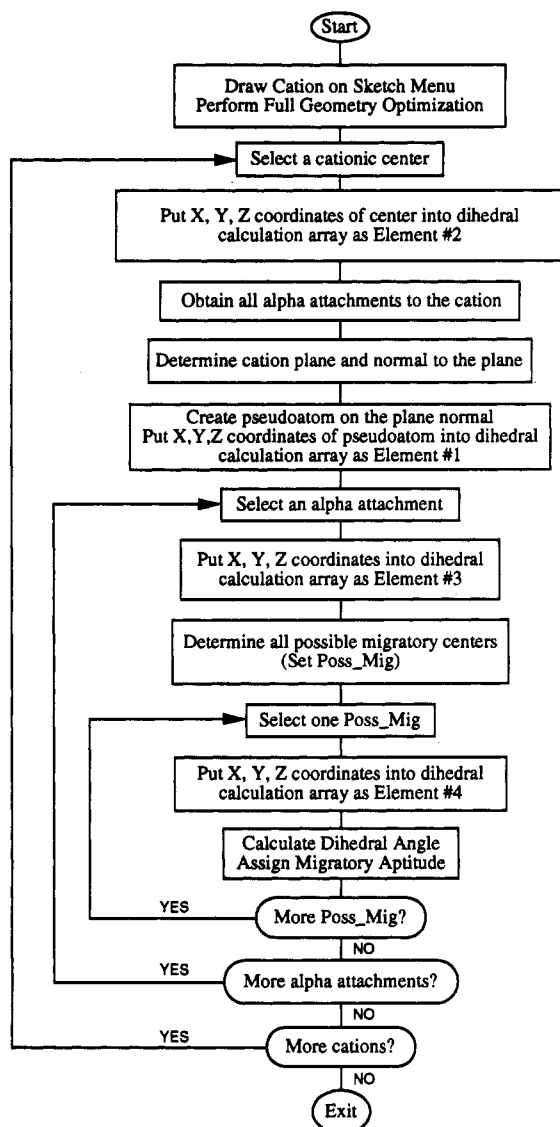
Figure 6. Algorithm for Wagner–Meerwein evaluation.

unsaturation, the best-fit plane through that unsaturation is determined based on a least-squares fit to the general equation of a plane described by the coordinates of the termini, and any atoms directly attached to the two termini, of the unsaturation. If any of the attached atoms are lone pairs of electrons, ATMTYP = 64, they are excluded from this determination. Therefore, as few as three atoms (*e.g.*, an aldehyde with no hydrogens considered) or as many as six (*e.g.*, a fully substituted olefin) may be involved in the calculation. The positive and negative normals to this plane are then calculated by recasting the general equation of the resulting plane to its corresponding normal equation. The accessibility of each face is then evaluated by the INTERFERENCE subroutine. In addition to the steric accessibility, a torsional congestion factor[35a,b] is calculated for both nucleophilic and electrophilic attack. The INTERFERENCE subroutine returns one numerical value proportional to the steric blockage at the site of the unsaturation and one value proportional to the torsional strain for carbon–heteroatom double bonds. This value is not related to the size of any specified nucleophilic or electrophilic reagent which may approach the unsaturation.[34e,35d,36] It is defined to be an intrinsic property of the molecule dependent only on a given conformer.
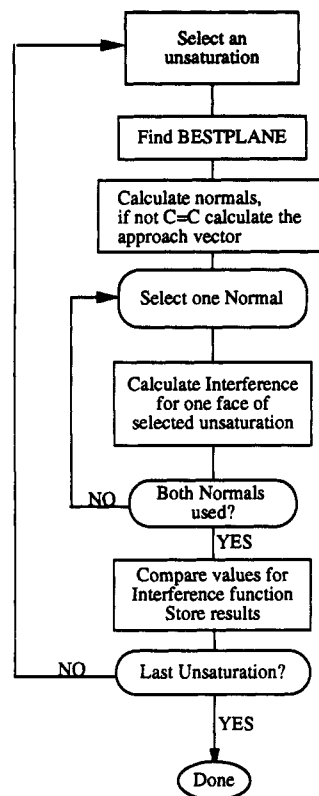

Figure 7. General reduction processing algorithm in PRCP3D.

Table VI. Parameters Used in Eqs 2 and 3

reagent	k_1	k_2	k_3
LiAlH$_4$	−1.93	−1.44	
NaBH$_4$	−2.05	−1.26	
LiAl(OMe)$_3$H	−4.08	−1.67	
LiAl(O-t-Bu)$_3$H	−3.45	−1.82	
AlH$_3$	−5.35	−1.73	
BH$_3$	−3.29	−1.25	
H$_2$/Pt			−1.69

The algorithm of the INTERFERENCE subroutine is based on a simplified version of the steric approach vectors originally described by Bürgi and Dunitz.[37] With a carbonyl group, electrophiles tend to approach the carbon atom along a vector described by the π-electron density of the unsaturation, nearly perpendicular (85° O=C—E+) to the carbonyl group (Figure 8). For nucleophilic attack, the preferred vector of approach is along the π* unoccupied orbital, at an O=C—Nu angle of approximately 100°. The approach of the reagent in three-dimensional space can therefore be described by an inverted cone, centered on the preferred reaction path to the carbon–heteroatom plane with its apex at the carbon atom. The magnitude of the cone angle (α) is determined by the nature of the reaction. The cone angle calculations were done with 5° steps, and the cone angles given here are those that gave the best results for the parameterization of k_1, k_2, and k_3 (Table VI, *vide infra*). For electrophilic reactions, the cone angle (Figures 8 and 9) is small (15°). A slightly larger cone angle of approximately 20° best describes nucleophilic attack. Presently, the following carbon–

(36) Meakins, G. D.; Percy, R. K.; Richards, E. E.; Young, R. N. *J. Chem. Soc. C.* **1968**, 1106.

(37) (a) Bürgi, H. B.; Dunitz, J. D.; Shefter, E. *J. Am. Chem. Soc.* **1973**, *95*, 5065. (b) Bürgi, H. B.; Dunitz, J. D.; Lehn, J. M.; Wipff, G. *Tetrahedron* **1974**, *30*, 1563. (c) Bürgi, H. B. *Angew. Chem., Int. Ed. Engl.* **1975**, *14*, 460.
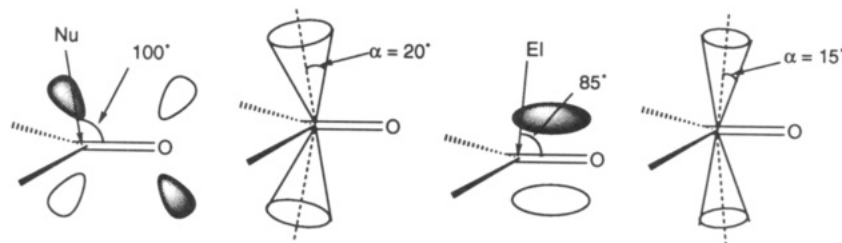
Figure 8. Approach vectors in three-dimensions, nucleophilic attack to the left and electrophilic attack to the right.

α is the cone angle
β is the angle between normal and atom R
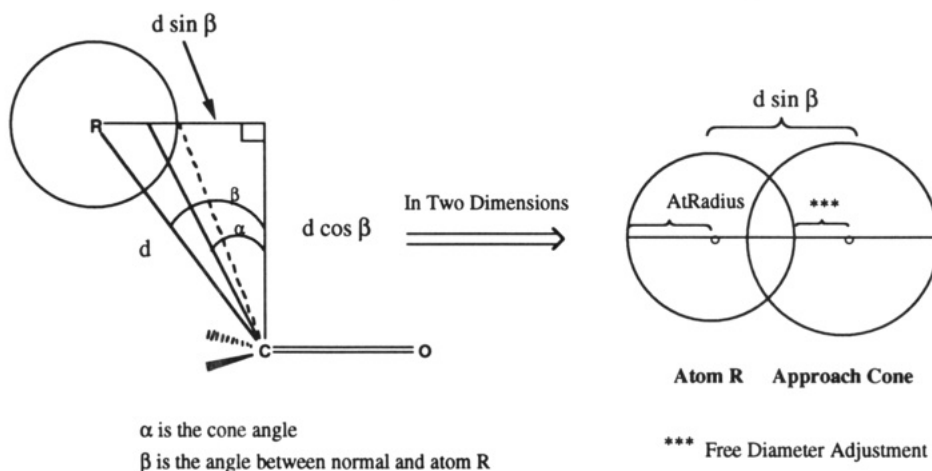
*** Free Diameter Adjustment

Figure 9. Two-dimensional translation of overlap.

heteroatom double bonds can be evaluated in the interference subroutine: C=O, C=N, and C=S. The cone for carbon–carbon double bonds is taken as centered at the midpoint between the two carbons, and the cone angle is 45°.[38]

If one side of the carbonyl plane is sterically hindered, as represented by atom R in Figure 11, the volume of an ideal approach cone is reduced by the volume element which results from overlap of the van der Waals (vdW) sphere of atom R with the cone. The vdW sphere of an atom normally extends beyond the center of any atom bonded to it. If two atoms bonded to each other both intersect the approach cone, the use of vdW spheres leads to the same space being included twice. To diminish this problem, the covalent radius is used for saturated carbons; however, the vdW radius is used for π-bonded carbons. Ideally, unsaturated atoms should be treated as unsymmetrical entities, but the above compromise gives reasonable results within the model presented here. Summation over all such intersections results in an absolute volume reduction of the cone. This process can be repeated for the cone on the opposite face, and the difference between the resultant volumes of the two ideal cones can be used as a measure of which side is more sterically hindered. Unfortunately, the calculation of intersecting volumes in 3-space is a computationally intensive problem which can only be alleviated if the dimensionality of the problem is reduced. A procedure to reduce the problem to two dimensions and calculate normalized intersection areas has been previously reported by Wipke.[35a,b]

In the INTERFERENCE algorithm, the two-dimensional approach is further reduced to a one-dimensional analog. In two dimensions, the overlap of the ideal approach cone and atom R is described by the area of intersection of two circles. In one-dimension, this overlap results in the reduction in the diameter of the approach cone circle due to encroachment by the atom's van der Waals radius. This is outlined in Figures 9 and 10. The overlap between the approach cone circle and the atom circle is dependent only on the separation of the two centers of the circles and their respective radii. The radius of the cone is calculated "on the fly" from the value of the cone angle α, the atom angle β, and the distance d between the carbonyl carbon and the infringing atom R. Depending on the separation distance and the two radii, the amount of overlap can fall into one of the four possible manifolds shown in Figure 10. The steric hindrance in itself is not sufficient to describe the face selectivity for carbonyls.[34,35] We have augmented the analysis with the torsional congestion correction method developed by Wipke and Gund (*vide infra*).[35a,b] They parameterized the torsional congestion term only for carbonyls, though we are currently using it for imines or thiocarbonyls as well.

The complete algorithm of the INTERFERENCE subroutine is shown in Figure 11. After all atoms within a 5 Å sphere of the reaction center (which is now the maximum value of d in Figure 9) are selected, those atoms which are on the opposite face of the plane, or are connected α to the unsaturation, are eliminated from the calculation. The summation of overlap over each of the remaining atoms is then calculated, normalized, and output from the routine. While not a mathematically rigorous treatment (*i.e.*, the percentage of "free" diameter of the approach cone is not *directly* proportional to the volume element described above), comparison of the relative
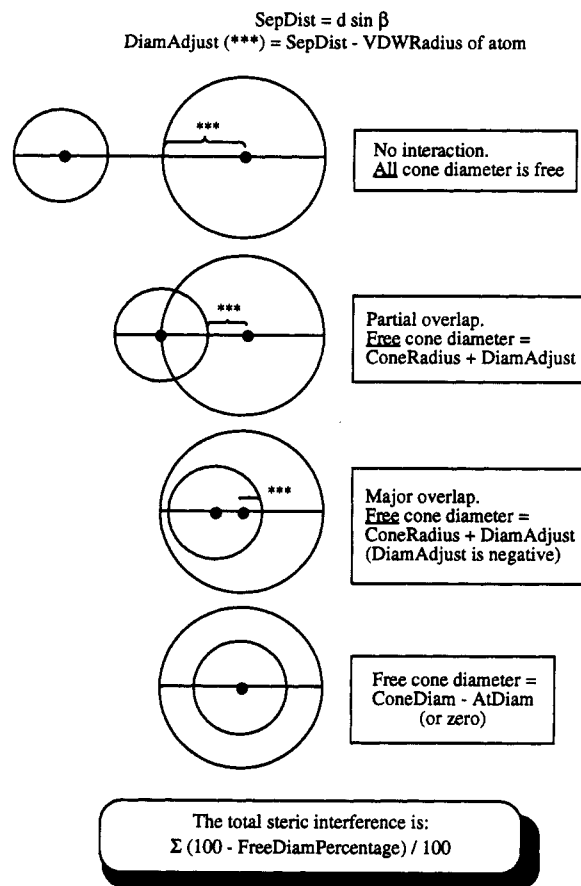
---

(38) This is a bond-centered congestion as opposed to an atom-centered as used for carbon–heteroatom double bonds. If the bond is highly polarized, as in an enol, there is no ambiguity in which atom would be attacked, and an atom-centered congestion would be preferred. Wipke and Gund have mentioned the use of a bond-centered congestion approach for reactions at carbon–carbon double bonds (note 33 in ref 35b), but to our knowledge no details have been published.

SepDist = d sin β
DiamAdjust (***) = SepDist - VDWRadius of atom

No interaction.
All cone diameter is free

Partial overlap.
Free cone diameter =
ConeRadius + DiamAdjust

Major overlap.
Free cone diameter =
ConeRadius + DiamAdjust
(DiamAdjust is negative)

Free cone diameter =
ConeDiam - AtDiam
(or zero)

The total steric interference is:
Σ (100 - FreeDiamPercentage) / 100

Figure 10. Possible overlap manifolds in two dimensions.

magnitudes of the function for each side successfully predicts the more sterically hindered face of the unsaturation.

Torsional strain between the incoming reagent and the substrate is calculated for both faces of a carbon–heteroatom double bond, *i.e.*, C=O, C=N, or C=S. The torsional angles are calculated for the preferred reaction paths of all substituents β to the reaction center. For each torsional angle that is less than 35°, a torsional strain contribution is calculated by eq 1.[35a,b] The difference in

$$\text{torsional strain} = \cos(\text{dihedral angle}(90/35)) \quad (1)$$

reactivity between the two faces is then approximated by eq 2 for C=X bonds and by eq 3 for C=C bonds. *k* is the

$$k = \exp(k_1 \Delta\text{steric interactions} + $$

$$k_2 \Delta\text{torsional strain}) \quad (2)$$

$$k = \exp(k_3 \Delta\text{steric interactions}) \quad (3)$$

quotient between the predicted rates for the two faces of the unsaturated bond. The parameters k_1, k_2, and k_3 for the studied reagents are given in Table VI. The parameters were derived by fitting calculated steric interactions and torsional strains to experimental data.[39] Some statistical data from the parameter fitting calculations are shown in Table VII. This model does not take into account temperature, solvent polarity, dilution, or pressure, factors

(39) The parameters were fitted by a geometric least-squares approach. It is necessary to use a geometric model, as an arithmetric model would give to much weight to the data with very high selectivity.
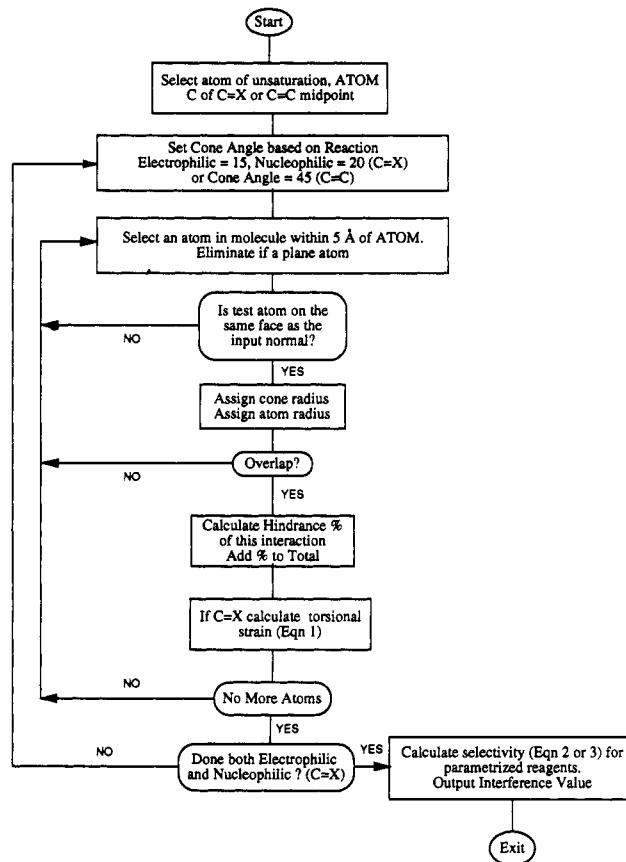

Figure 11. Algorithm of INTERFERENCE subroutine.

Table VII.   Summary of the Results from Tables VIII and X

reagent	no. of entries	r^a	MD^b
LiAlH$_4$	27	0.882	2.18
NaBH$_4$	23	0.951	1.82
LiAl(OMe)$_3$H	14	0.940	2.31
LiAl(O-t-Bu)$_3$H	18	0.931	1.86
AlH$_3$	10	0.960	1.82
BH$_3$	13	0.784	2.45
H$_2$/Pt	12	0.833	1.83

[a] Correlation factor for calculated vs experimental relative ratios for the data in Tables VIII and X. [b] Mean deviation as the average quotient between the calculated and experimental relative ratios for the data in Tables VIII and X.

that often can give a change in the selectivity by a factor of 2 or more. Considering the limitations of the model, the mean deviations in Table VII seem satisfactory. Calculated and available experimental results for the substrates in Figure 6 are given in Tables VIII and X (see section IV B). CAMEO three-dimensional perception sets are created from the calculations, and this information is used to indicate the stereochemical outcome by making the attacked carbon the origin of a wedged or dotted bond to the incoming reagent. Several examples of such predictions in the Redox module are illustrated below.

IV. Examples

A. Cationic Rearrangements. Although thermodynamic stability is the primary factor in predicting the preferred reaction path of a Wagner–Meerwein rearrangement,[40] knowing the correct spatial orientation of the sites that undergo cationic shifts is crucial in determining the

Table VIII. CAMEO Predictions for Product Ratios in Ketone Reductions

compd[a]	preferred side[b]	LiAlH₄[c]	NaBH₄[c]	LTMA[c,d]	LTBA[c,e]	AlH₃[c]	BH₃[c]
1	exo	6.8 (11.5)	5.9 (6.1)	15.9 (49)	15.2 (19.0)[f]	7.5 (24)[g]	3.7 (49)
2	exo	4.8 (10.1)	4.4 (19.0)		10.1 (9.0)[f]		
3	exo	6.7 (12.7)	5.8 (5.7)	15.4 (24)			
4	exo	6.2 (9.0)		13.1 (49)			
5	endo	4.8 (9.0)	6.7 (3.6)	97 (49)		11.6 (10.1)[h]	
6	endo	5.0 (12.5)	7.0 (6.2)	100 (99)	35 (24)[f]	9.6 (9.0)[h]	3.6 (1.08)
7	exo	6.3 (9.0)		13.9 (32)			
8	exo	3.7 (93)					
9	exo		8.7 (4.3)				
10	exo		1.67 (5.7)				
11	exo	100 (100)					
12	exo	76 (10.1)					
13	trans	1.32 (1.00)			1.74 (3.0)[f]		
14	trans	4.8 (5.7)			17.7 (19.0)[f]		
15	exo		100 (100)				
16	trans		13.1 (11.5)				
17	exo					16.5 (50)[i]	
18	trans		5.1 (0.69)		8.5 (19.0)[f]		
19	trans		3.0 (5.7)				
20	cis	0.59 (3.8)		0.33 (1.28)	0.39 (2.6)[h]	1.24 (3.7)[h]	1.19 (3.0)
21	cis						3.4 (9.0)[j]
22	cis	4.7 (9.0)	3.1 (6.1)	1.79 (1.56)	3.7 (9.0)[f]	11.8 (6.7)[h]	6.0 (9.0)[j]
23	cis	11.7 (19.0)		11.7 (4.0)	18.2 (10.1)[f]		
24	trans	3.9 (4.9)	2.7 (4.0)	1.64 (1.13)[f]	3.1 (5.7)[f]		
25	cis	2.1 (4.0)	3.4 (2.4)	43 (24)	12.5 (24)[f]	10.1 (7.3)[k]	3.4 (1.94)
26	cis	5.0 (4.6)	3.4 (2.7)	2.70 (0.45)	4.8 (2.3)[h]	6.8 (2.8)[h]	4.0 (2.8)
27	trans		3.6 (4.3)				
28	trans	4.7 (5.7)	3.2 (6.7)		4.2 (6.7)[f]		4.8 (7.3)[j]
29	trans		1.26 (6.7)				
30	trans		3.8 (7.3)				
31	trans	5.0 (4.6)	3.4 (4.9)		4.4 (4.3)[f]		5.4 (2.7)
32	trans		3.2 (4.3)				
33	cis	7.5 (5.2)	5.0 (8.1)				5.8 (8.1)[j]
34	cis	0.52 (3.0)					0.65 (1.44)
35	trans	0.37 (1.38)		0.83 (1.78)	0.48 (1.18)[f]		0.37 (3.3)
36	trans	11.4 (2.7)					3.6 (2.8)
37	exo				8.9 (11.5)[f]		
38	trans	7.2 (19.0)		4.9 (15.7)			
39	trans	1.83 (1.44)	2.9 (4.6)			10.2 (5.7)[l]	
40	trans	6.9 (12.9)				98 (100)[l]	
41	trans				4.1 (7.3)[f]		
42	exo				0.51 (1.51)[f]		

[a] See Figure 6. [b] The calculated side of attack is not always the overt side. A value below 1.0 indicates that attack is predicted from the wrong side. [c] Calculated values from eq 2. Experimental values in parentheses. Experimental values are taken from ref 35d, except where noted. [d] Lithium trimethoxyaluminum hydride. [e] Lithium tri-*tert*-butoxyaluminum hydride. [f] Malek, J. *Org. React.* **1985**, *34*, 1. [g] Ashby, E. C.; Boone, J. R. *J. Org. Chem.* **1976**, *41*, 2890. [h] Yoon, N. M.; Brown, H. C. *J. Am. Chem. Soc.* **1968**, *90*, 2927. [i] Ashby, E. C.; Noding, S. A. *J. Am. Chem. Soc.* **1976**, *98*, 2010. [j] Lane, C. F. *Chem. Rev.* **1976**, *76*, 773. [k] Ayres, D. C.; Sawdaye, R. *J. Chem. Soc. B* **1967**, 581. [l] Ayres, D. C.; Kirk, D. N.; Sawdaye, R. *J. Chem. Soc. B* **1970**, 505.

path leading to the final product. As a rule, any substituent attached α to a carbocationic center in a Wagner–Meerwein rearrangement in CAMEO may migrate if the carbocation which results from the migration is of the highest possible order and at least as stable as the initial cation. Such behavior is often observed in the rearrangement chemistry of conformationally labile acyclic cations, and many products can result from both single and multistep rearrangements of one suitably substituted cation. The three-dimensional module of CAMEO is used to evaluate conformationally locked systems that can undergo multistep cationic rearrangements and enhances the simple rule for predicting acyclic cationic rearrangements in the electrophilic module.

If three-dimensional information is available, the above simple rule is modified to first evaluate any explicit hydrogen atoms in the migratory aptitude sets generated by the 3D module (WAG30, WAG60, and WAG90). When no migratory hydrogen atoms are available, all other α

substituents are evaluated. In addition to the stability requirement mentioned above, these migrating groups must be in the appropriate perception sets (WAG30 or WAG60). If no atoms exist in the WAG30 set, only then may atoms in the WAG60 set be considered. The application of these rules is illustrated in the following examples of single and multistep rearrangement chemistry.

In Scheme II, there are four carbon substituents which can migrate, yet only one product, camphene, is experimentally observed.[41] Since a primary carbocation would result, atom 4 would not be considered for migration; however, the other three possible migrations all generate tertiary carbocations. Migration of atom 3 is not competitive, since it is in the WAG90 set and the choice between atoms 1 and 2 is facilitated by the 3D information. Subjection of the isobornyl cation to the dihedral analysis outlined returns only one substituent (atom 1) whose dihedral angle with the unoccupied orbital of the cation

(40) *E.g.*; March, J. *Advanced Organic Chemistry*; John Wiley & Sons; New York, 1985; pp 958–963 and references cited therein.

(41) (a) Simonsen, J. L. *The Terpenes*, 2nd ed.; Cambridge University Press, Cambridge, 1949; Vol. 2, pp 290, 360. (b) Pickard, R. H.; Littlebury, W. O. *J. Chem. Soc.* **1907**, *91*, 1973.

Scheme II. Atoms β to the Cation Are Placed in Different Perception Sets Depending on the Dihedral Angle to a Dummy Atom Located Perpendicular to the Cation
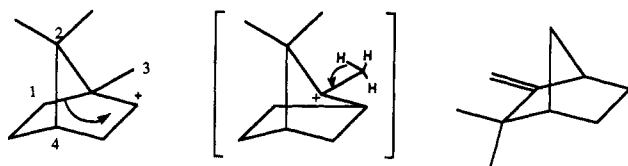
Atom	Dihedral Angle	Perception Set
1	18.34	WAG30
2	56.26	WAG60
3	69.31	WAG90
4	91.20	WAG90

Scheme III. Possible products from Cationic Rearrangement for Cation 55
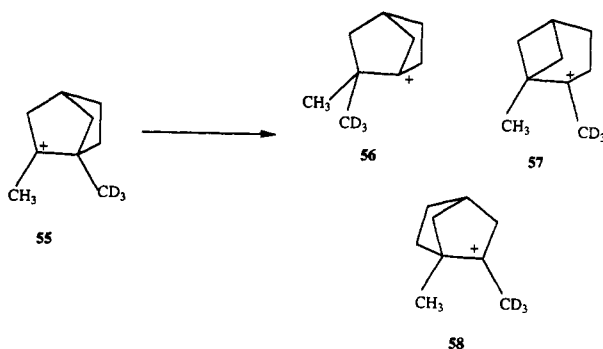
is less than 30°. This atom is placed in the WAG30 perception set. Rearrangement of this center followed by elimination of an implicit hydrogen atom from the methyl group (atom 3) affords camphene as a single product.

A related example is shown in Scheme III. In this case, reported by Saunders et al.,[42] an isotopically labeled bridged system undergoes a degenerate cationic rearrangement. In the absence of spatial perception, CAMEO predicts three products, two of which (56,57) are disfavored due to poor orbital overlap. Inclusion of the 3D analysis in the mechanistic evaluation, though, directs CAMEO to propose only one possible product (58), which is the one corresponding to the degenerate rearrangement observed experimentally.[42]

As a more complex example, even multistep rearrangements can now be considered, such as the rearrangement of 3β-friedelanol (59) to olean-13(18)-ene (60).[43] In this system, rearrangement of the cation 61 proceeds through seven successive submissions to CAMEO to yield cation 62 as the only stereoelectronically allowed product (Scheme IV).

B. Reductions. Metal Hydride Reductions. Most reductions of a carbonyl group to an alcohol are performed with metal hydrides. An important aspect of their synthetic utility is the stereoselectivity of the reduction. While acyclic carbonyls are often so flexible that several conformations may undergo reaction, hydride addition to cyclic carbonyls that are conformationally locked can lead to highly stereoselective attack at the less hindered side

(42) Saunders, M.; Telkowski, L.; Kates, M. R. *J. Am. Chem. Soc.* **1977**, *99*, 8070.

(43) Corey, E. J.; Ursprung, J. J. *J. Am. Chem. Soc.* **1956**, *78*, 5041.

Scheme IV. Seven-Step Rearrangement from the Cation Generated from 3β-Friedelanol to Olean-13(18)-ene and the Corresponding CAMEO Analysis
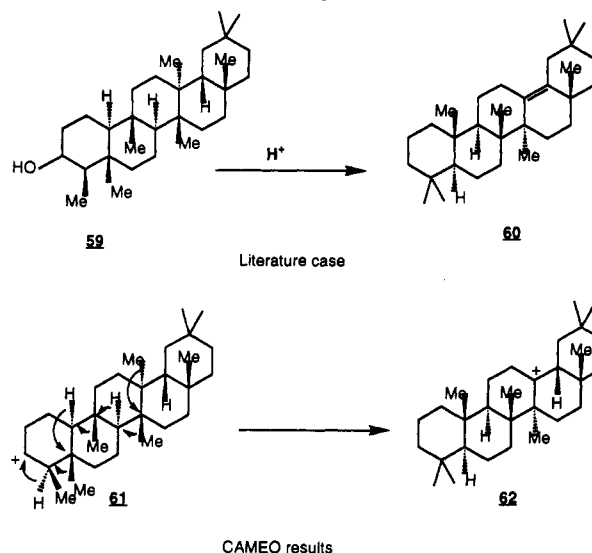
of the carbonyl. It is assumed that the attacking hydride approaches the carbonyl carbon along a line which allows for maximum orbital overlap in the transition state.

Table VIII illustrates the ability of the 3-D module to determine the stereoselectivity of a metal hydride reduction and the stereochemistry of the products. In general, the module is able to evaluate the cooperation or opposition of the steric and torsional factors inherent in the substrate's interaction with a particular hydride reagent. For example, adding bulky substituents to a hydride reagent may lead to even greater stereoselectivities. The nucleophilic or electrophilic nature of the hydride reagent must also be considered. The reduction of camphor, 6, illustrates the ability of the 3D module to accurately differentiate between the stereoselectivities of hydride reagents. The bulky LiAl(OMe)$_3$H and LiAl(O-t-Bu)$_3$H reagents give high experimental selectivities (k = 24 corresponds to a 96:4 mixture). Lithium aluminum hydride, sodium borohydride, and alane all show medium selectivity (k = 9.0 corresponds to a 90:10 mixture), while borane gives a poor selectivity in this case (k = 1.08 corresponds to a 52:48 mixture). Although the exact order between the reagents is not found, the calculated selectivities from eq 2 are in the correct ranges.

Reductions of monocyclic cyclohexanones, 22–33, 35, 38, usually result in intermediate selectivity. In almost all cases, the 3D module correctly predicts whether axial or equatorial attack is favored. If an axial 3-substituent is not present, the attack is almost always on the axial side. Torsional effects are then the most important factor. If there is an axial 3-substituent, the steric bulk of that substituent determines facial selectivity. Equatorial attack is prevalent, and steric effects are more influential.

The planar ring geometries given for four- and five-membered rings by our LENGTH3D optimizer have resulted in some inconsistencies. Our calculations predict that 2-methylcyclobutanone, 34, and 2-methylcyclopentanone, 20, undergo hydride attack leading to trans product. If MM3-optimized structures are used, the calculations for these molecules predict predominantly cis products, in accordance with the experimental results. Selectivity calculations for the 2-substituted cyclohex-

anones, 26 and 35, are very sensitive to the geometry of the substituent. If geometries calculated from other force fields (MM3 or AM1) are used, the results are more in agreement with experimental data. For ketone 35, the correct side of attack is predicted for most of the hydride reagents, and any errors with the remaining reagents are smaller than those obtained with the LENGTH3D geometries. Experimentally, the LTMA reduction of ketone 26 results in addition to the side opposite that observed with the other reagents. Our calculations indicate a larger amount of trans-attack for LTMA reductions relative to the other reagents, though with the LENGTH3D geometry the calculation still favors cis-attack. Using either AM1- or MM3-generated geometries leads to calculated selectivities near 1.0 for LTMA reductions (1.08 trans-attack for AM1; 1.08 cis-attack for MM3), while the other reagents remain cis-selective.

The poor correspondence between calculated and experimental selectivity for ketone 8 may lie in the experimental value. Only one isomer was found in 92.8% yield. If the unaccounted 7.2% is the other isomer, then the experimental selectivity would be 12.9, and the difference between the two values would not be unreasonably large.
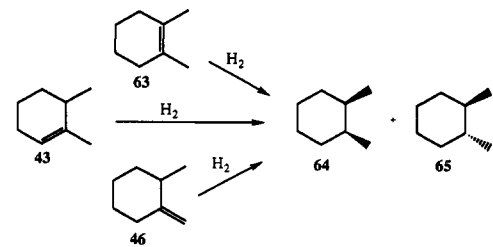
Borane is the reagent with the poorest fit between calculated and experimental selectivities (Table VII). This poor fit may be due to the fact that in some cases the rate-limiting step in the reaction is the nucleophilic attack of a carbonyl oxygen on a hydrogen atom and not the electrophilic attack of boron on a carbonyl carbon.

Catalytic Hydrogenations. Hydrogenations involving heterogeneous metal catalysis (*e.g.*, Lindlar's catalyst, Adam's catalyst, palladium black, *etc.*) are often stereo-selective. In rigid ring systems containing double bonds, high stereoselectivity results from the tendency of the metal surface of the catalyst to complex with the less hindered face of the olefin. Since the components of the hydrogenation reaction are believed to be simultaneously adsorbed onto the surface of the metal catalyst, the complexed face of the olefin will be in closest proximity to the adsorbed hydrogen atoms. Hydrogen should, therefore, add to the less sterically congested face.

Although catalytic hydrogenation is usually the method of choice for reducing an isolated carbon–carbon double bond, the stereochemical results can be complicated by double bond migrations and cis–trans isomerizations. In the catalytic hydrogenation mechanism, a hydrogen atom actually adds to the less hindered carbon of the double bond in a stepwise fashion, resulting in an intermediate, monoadsorbed species. Depending on reaction conditions, this half-hydrogenated species may undergo a configurational change before reduction is complete. Since addition of hydrogen is reversible, the half-hydrogenated species may even be deadsorbed from the catalyst surface before the second hydrogen is added. Not only can this result in a loss of syn selectivity, but random readsorption may lead to hydrogen exchange and double bond shifts. In some cases migration of the double bond is not detectable without tracers. But if a hydrogen shift occurs such that a new, highly substituted double bond formed cannot reach the catalyst surface to be reduced or that a substrate containing prochiral sites is isomerized (Table IX), such migration will affect the expected stereochemistry.

When conditions favor high hydrogen availability at the catalyst surface, the second hydrogen will add cis

Table IX. Stereoproduct Formation in the Hydrogenation of Dimethylcyclohexenes[a]

compd	Pd-on-C/AcOH/25 °C/1 atm		PtO$_2$/AcOH/25 °C/1 atm	
	% cis (64)	% trans (65)	% cis (64)	% trans (65)
63	25–30	70–75	80	20
43	25	75	70–80	20–30
46	30	70	70	30

[a] Bartok, M. *Stereochemistry of Heterogeneous Metal Catalysis*; Wiley: New York, 1985; pp 84, 94, 101 and references cited therein.
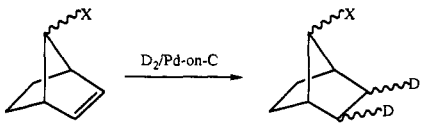
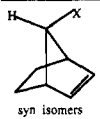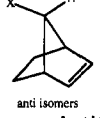**Table X. CAMEO Predictions for Platinum-Catalyzed Hydrogenations**

compd[a]	preferred side of attack	calcd selectivity[b]	expl selectivity[c]
43	trans	1.50	4.0
44	cis	0.34	1.50
45	trans	9.2	19.0
46	trans	1.44	2.6[d]
47	trans	1.81	2.7[d]
48	cis	0.59	2.3
49	trans	1.94	4.0
50	trans	11.4	13.3[e]
51	trans	9.0	5.3[e]
52	cis	1.85	1.86[d]
53	trans	1.74	1.78[d]
54	exo	2.3	3.0[e]

[a] See Figure 6. [b] Calculated by eq 3. [c] From Bartok, M. *Stereochemistry of Heterogenous Metal Catalysis*; Wiley: New York, 1985; unless otherwise noted. [d] Mitsui, S.; Gohke, K.; Saito, H.; Nanbu, A.; Senda, Y. *Tetrahedron* 1973, 29, 1523. [e] van Tamelen, E. E.; Timmons, R. J. *J. Am. Chem. Soc.* 1962, 84, 1067.

relative to the first hydrogen, deadsorption and readsorption will not have an opportunity to occur, and little migration will take place on the surface of the metal catalyst. As shown in Table IX, the metal catalyst used in a reduction has the greatest influence on the stereoselectivity of the reaction. The stereoselectivity is completely reversed for the palladium and platinum hydrogenation catalysts. Palladium catalysts promote double bond migrations, while platinum catalysts do not cause as much migration of the double bond. (In neutral solvents, platinum catalysts do not promote any migration; Raney nickel catalyst, usually not favoring bond migration because of basic reaction media, exhibits an intermediate degree of bond migration). It is difficult to always predict the stereochemical outcome of a reaction based solely on the conformation of an individual substrate, since rearrangement of the starting material may occur prior to or during the hydrogenation process. Reagents and reaction conditions that promote the least isomerization are more appropriate for application to the 3D module.

Since platinum-catalyzed hydrogenations are less prone to isomerizations than most other catalytic hydrogenations, this reagent was parameterized first. The quantitative results from the parameterization of H$_2$/PtO$_2$ are offered in Table X. Most of the reactions in Table X show modest selectivity; however, for those reactions with somewhat higher experimental selectivities (45 and 50), the calculated

Table XI. Stereofacial Selectivity in Reduction of 7-Substituted Norbornenes

orientation of substituent	X	preferred face of attack (CAMEO)	% exo,cis addition (expl)
syn isomers	OH	endo	100
	CH₃	endo	(10)[a]
	OOCCH₃	endo	60
	C(CH₃)₃	endo	20
anti isomers	OH	exo	100
	CH₃	exo	
	OOCCH₃	exo	100
	C(CH₃)₃	exo	75
unsubstituted isomer	H	exo	(90)[b]

[a] Data are for the hydrogenation with platinum catalyst of the 7,7-dimethyl system, 2,7,7-trimethylnorbornene; Brown, H. C.; Kawakami, J. H. *J. Am. Chem. Soc.* **1970**, *92*, 201. [b] Data are for the hydrogenation with platinum catalyst of 2-methylnorbornene; Brown, H. C.; Kawakami, J. H. *J. Am. Chem. Soc.* **1970**, *92*, 201.

selectivities are correspondingly high. For olefins 44 and 48 the side of preferred attack calculated by CAMEO is incorrect, though using geometries from other force fields improves the situation. The correct side of attack is predicted for olefin 44 when geometries generated be either MM3 or AM1 are used. For olefin 48 the calculated selectivity is closer to 1.0; however, the predominant attack is still calculated to be from the wrong side.

For the remaining catalytic hydrogenation reagents in CAMEO, the predictions are only qualitative at present. The preferred face of attack is indicated, but not the degree to which it is less hindered. The bridged ring system shown in Table XI[44] illustrates facial selectivity in a rigid molecule. For an unsubstituted norbornene, exo addition is preferred over endo addition because the exo face is less

hindered by a methylene bridge than the endo face is by an ethylene bridge. A very large substituent on the methylene bridge, however, may effectively block reduction at the exo face. Table XI illustrates the influence of both the size and orientation of a 7-substituent. Because the bulky tertiary butyl group cannot be oriented away from the double bond, the highly hindered *syn-tert*-butylnorbornene preferentially undergoes endo addition of hydrogen. Although endo addition is sterically preferred for *syn*-hydroxynorbornene and *syn*-acetoxynorbornene, the ability of the oxygen atom to anchor onto the catalyst may reverse the stereochemical outcome by directing addition of hydrogen to its side of the molecule. In this case, the chelation or haptophilicity of an oxygen atom favors exo addition. The incorporation of haptophilicity as a future refinement of the computational analysis can be envisioned now that the basic abilities for 3D structure generation have been implemented in CAMEO.

V. Conclusion

Generation and analysis of three-dimensional chemical structures has been implemented in the CAMEO program. This allows the user to elicit spatial information similar to that which can be obtained through the use of physical molecular models. This information has also been made optionally available to the reaction analysis phase of CAMEO, allowing the development of mechanistic rules which take molecular shape and accessibility into account. Several initial examples of rule generation and application have been shown. This module represents a powerful new way of thinking for CAMEO program development and provides a means by which molecular conformation and shape-specific information may be involved in formulating predictions of reactivity.

(44) Baird, W. C.; Surridge, J. H. *J. Org. Chem.* **1972**, *37*, 1182.