

## CAMEO: a program for the logical prediction of the products of organic reactions

William L. Jorgensen, Ellen R. Laird, Alan J. Gushurst, Jan M. Fleischer,  
Scott A. Gothe, Harold E. Helson, Genevieve D. Paderes, and Shenna Sinclair

Sterling Chemistry Laboratory, Yale University, New Haven, CT 06511, USA

**Abstract** - An interactive computer program, CAMEO, has been developed to predict products of organic reactions given the starting materials and conditions. The analyses primarily feature mechanistic reasoning and have been extended to cover most of the major classes of organic reactions. An important aspect of the project has been to clearly define the logic behind accurate prediction of products and to simultaneously search for organizing principles governing organic reactivity. The current capabilities of the program are summarized in this review and illustrated with comparisons to reported experimental reaction sequences.

### INTRODUCTION

Efficient and elegant syntheses require innovative coordination of reaction sequences, thus ensuring that organic synthesis remains one of the most challenging areas in chemistry. The immense structural variety of desirable target molecules is countered by the vast array of available chemical transformations. Woodward was among the first to address the problem of systematic synthetic design (ref. 1). His efforts expanded the range of target molecules and helped spark a revolution in synthetic planning. The analytical approach was formalized with Corey's development of OCSS and LHASA, the first programs for computer-aided organic synthesis (CAOS) (ref. 2). The number of programs for strategic synthetic planning has expanded dramatically over the past two decades, and several reviews are available (ref. 3). Beyond their evident utility for designing routes toward complex target molecules, these programs coordinate the logic involved in efficient product construction.

Two fundamental protocols exist in the development of computer programs for synthesis planning:

- Given a target molecule, determine strategic sites that could be constructed from more accessible molecules (retrosynthetic analysis).
- Given a substrate molecule, evaluate reaction possibilities under supplied conditions (synthetic analysis).

Virtually all CAOS programs operate retrosynthetically, and are usually dependent upon reaction libraries of known transformations (ref. 4). Program users supply the target molecule, and the programs respond with a series of suggested precursors. Repeating this process results in growth of a "tree" comprising possible routes to the target. The predictions are, however, limited to the known reactions stored in the database. The natural complement to programs that generate precursors is a system for testing the feasibility of reactions for converting those precursors to the desired target.

The interactive CAMEO program is designed for this purpose, and is characterized by its operation in the synthetic direction and its use of mechanistic reasoning in formulating its predictions. Given graphical input of starting materials and conditions, the program arrives at its conclusions by application of a series of rules designed to consider structural features for the determination of reactivity. The approach avoids the use of large databases and topological simplifications; reaction modules correspond to and are designed for the evaluation of reactions primarily by the type of intermediate involved. The program is presently capable of analyzing base-catalyzed and nucleophilic (ref. 5), acid-catalyzed and electrophilic (ref. 6), pericyclic (ref. 7), oxidative and reductive (ref. 8), free radical (ref. 9), and carbenoid (ref. 10) reactions.

The goals of the project are twofold. From a practical standpoint, the utility of a program for analyzing the feasibility of a proposed synthetic pathway is evident. Awareness of potential side products is invaluable to the practicing organic chemist, and increasing structural complexity escalates the difficulty of assessing relative reactivity.

As an academic project, development of the program necessitates the search for and implementation of fundamental principles that govern organic reactivity. Refinements are implemented such that trends are understood in terms of overall structure-reactivity relationships. As a result, the project has stimulated the discovery of relationships for the prediction of, for example,  $pK_a$ s, the reactivity of aromatic rings toward electrophilic substitution, and the rates of pericyclic reactions. Accordingly, unprecedented reactions can be considered by the program's mimicking of mechanistic logic.

This paper provides an overview of the CAMEO program with emphasis on the implementation and status of reaction analyses. Particular attention is given to reaction modules that best espouse the philosophy of applying mechanistic logic to the problems of product prediction.

## REACTION PROCESSING

There are three principal segments of the CAMEO program: a graphical interface, structural perception, and reaction evaluation. An overview of the processing is presented in Figure 1. The user enters structures in a sketch menu plotting box with a mouse and imparts structural features through the perimeter menus (Figure 2). Thus, the molecular framework is "drawn" in much the same way one draws a structure on paper, and heteroatoms, stereochemistry, and charges are added as appropriate. Alternatively, structures may be entered via accessing the "names" feature; an input IUPAC or acceptable trivial name results in placement of the corresponding structure in the plotting box. Existing structures, either drawn or retrieved from a menu, may be elaborated upon by specifying the position number and the desired functionality.

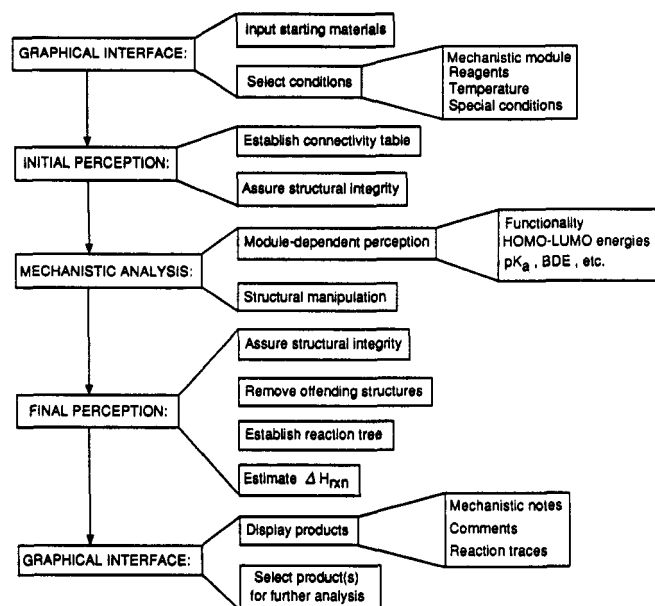


Fig. 1. Processing phases for the CAMEO program.

Options for scaling, rotating, and redrawing (ref. 11) molecules permit creation of aesthetically pleasing structures. Predictions for physical data are also available, and provide convenient access to estimates of  $pK_a$  (ref. 12), BDE (ref. 9), and  $\Delta H_f$  (ref. 11). Much of this information is used internally by the program to determine reactive sites and to evaluate competitions. Commonly used reagents and special conditions (*e.g.*, temperature and reaction medium) are available through menus associated with each mechanistic class. Extensive on-line help features regarding each menu provide convenient assistance during program operation.

Upon submission of the structure for processing, a generalized perception phase establishes a connectivity table of atom numbering, type, charge, coordinates, and bonding and stereochemical information. Erroneous structures are returned to the user for correction before processing may continue. In addition, advanced structural information is assessed, *i.e.*, ring ensemble, aromaticity, principal functionality, and topological symmetry information is garnered for use by the reaction evaluation modules.

Following successful perception, the structure is submitted to the chosen reaction module for analysis and product formation. Products issued by the program are perceived to assure structural integrity, and offending products are transformed or removed. The heat of reaction is also estimated at this point (ref. 11). Products are displayed in a "starting material  $\rightarrow$  products" format, and are accompanied by brief mechanistic notes describing the transformations, as well as the previously specified conditions. A product ranking scheme designates products as "major", "minor", or "disfavored" to provide information concerning the relative feasibility of competing processes. A sample of typical program output is provided in Figure 3. Several mechanistic steps are often involved in formation of a final product. In the illustrated case, the two equivalents of *n*-BuLi resulted in proton transfer and halogen/metal exchange followed by the ring forming addition and  $S_N2$  reactions (ref. 13). Additionally, a "tree" menu displays the genealogical relationships between products and allows selection of individual structures for examination or further reaction. This menu also provides access to detailed comments concerning the decisions made by the mechanistic phase. Comments provide insightful reasons for product rejection or relative favorability, and often cite literature references to pertinent studies.

Fig. 2. The CAMEO sketch menu for structure entry and manipulation.

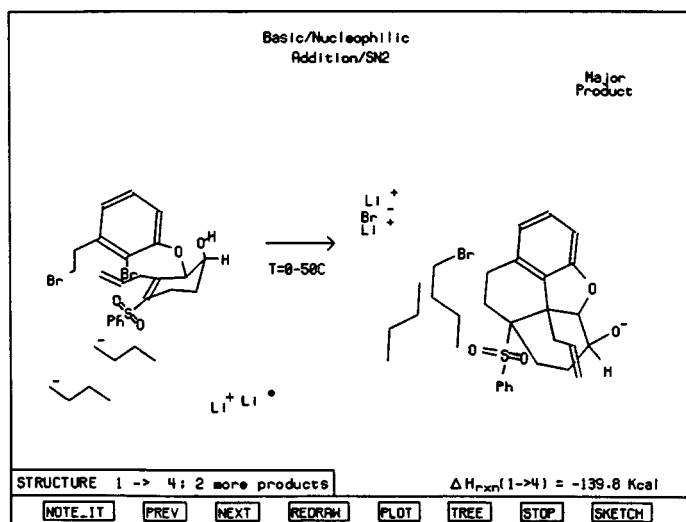
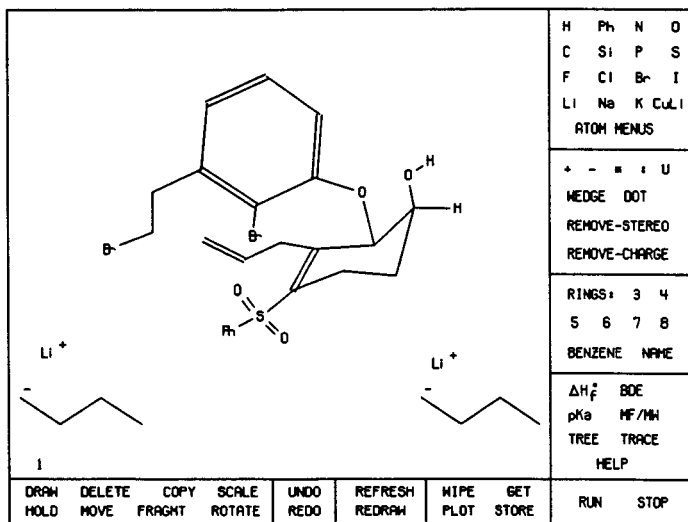


Fig. 3. Sample display of CAMEO output.

## REACTION EVALUATION

Most reactions are evaluated in CAMEO using mechanistic logic familiar to organic chemists. Many reactions can be reduced to series of fundamental steps, *e.g.*, proton transfer, addition, and elimination. By thoroughly modeling these discrete steps and determining when they are applicable, whole reactions can be accurately evaluated by combinations of steps. This approach permits generalization, hence the need for numerous data tables is circumvented. By definition, it also affords the prediction of unprecedented reactions that are

mechanistically reasonable. The six areas of chemistry noted in the introduction are covered in eight mechanistic modules which form the backbone of the program and are discussed in the remainder of this paper. More extensive presentations on specific topics are available in the original publications (refs. 5-10).

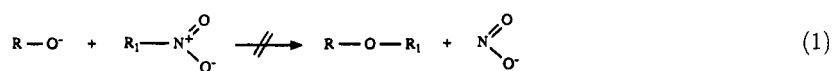
### Basic/nucleophilic module

The general scheme for analysis of base-catalyzed and nucleophilic reactions has been recently reviewed (ref. 5e). Some specific capabilities include the treatment of organometallic reactions involving Li, Mg, and CuLi counterions (ref. 5b), the novel chemistry of organosilicon (ref. 5c) and organophosphorus compounds, reactions of unsaturated systems including nucleophilic aromatic substitution (ref. 5d), and the reactions of phosphorus and sulfur ylides (ref. 5f). The basic/nucleophilic module is divided into four principle segments which focus on evaluating interactions between various nucleophilic and electrophilic sites. These segments involve:

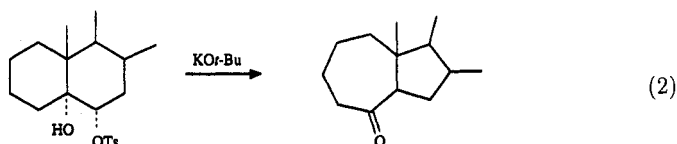
1. perception of nucleophilic and electrophilic sites in the input reactants,
2. determination of applicable mechanisms for a given nucleophilic/electrophilic site pairing,
3. evaluation of the best mechanisms for a given nucleophilic site, and
4. general comparison and screening of all reaction pathways.

Nucleophilic sites can be generated in CAMEO via proton transfer, halide-alkyllithium exchange, or metal insertion. These "mechanistic presteps" are typically faster than substitution, elimination, or addition reactions and are therefore considered during nucleophilic site perception according to a kinetic hierarchy. Organometallic addition can be competitive with these processes and therefore is also considered to determine whether the organometallic base should be kept as a potential nucleophile for further examination.

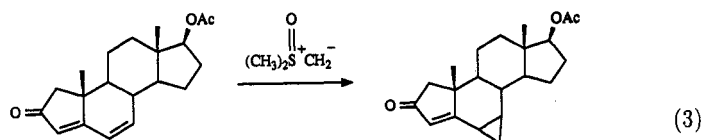
For each mechanistic prestep, the most stable anion and any others within a specified  $pK_a$  window are generated (ref. 14). After all nucleophilic sites have been determined, each is assigned a "nucleophilic qualification value", or NQV, in order to gauge its relative tendency to behave as a nucleophile or as a base. The NQVs thus help the program to distinguish among the predominant mechanisms for a nucleophilic site, *e.g.*,  $S_N2$  vs. E2. The perception of electrophilic sites also utilizes  $pK_a$  values. The initial ranking for these sites is based on the  $pK_a$  of the conjugate acid of the leaving group and is then adjusted to account for special effects, *i.e.*, steric hindrance and relief of ring strain. Electrophilic sites are also categorized according to the reactions they are permitted to undergo. This classification scheme provides an efficient means for avoiding naive chemistry that may have otherwise been considered, *e.g.*,



The applicable mechanisms for each nucleophilic/electrophilic site pairing are then determined by analyzing each potential step individually to assure that fundamental reactivity and structural requirements are satisfied. The reactivity requirements are gauged by a  $\Delta pK_a$  rule, which compares the reactivity of the nucleophilic and electrophilic sites and assesses whether the nucleophilic site is strong enough to displace the leaving group. If so, the substrate is then evaluated on structural grounds. For instance, epoxide formation by an intramolecular  $S_N2$  reaction is considered in eq 2, but is rejected since the tosylate has the wrong stereochemistry for backside displacement. The program instead predicts a 1,2-rearrangement, which corresponds to experimental observation (ref. 15).



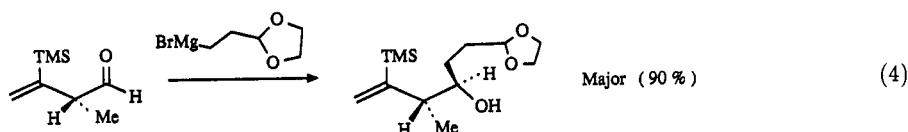
"Mechanistic evaluation" addresses the competitions between the different mechanisms available for *each* nucleophilic site. The established rules assess, for example, 1,2- vs. 1,4-addition, addition-elimination vs. E2, or  $S_N2$  vs. E2 or 1,2-rearrangement (ref. 5e). The competitions for the sulfur ylide in eq 3, for instance, are between 1,2-, 1,4-, and 1,6-addition.



The program predicts the major product to result from initial attack at the 1,6-addition site; reactions at the 1,2-addition sites (the ketone and acetate groups) are rejected since oxosulfonium ylides display a marked preference for conjugate addition over carbonyl addition, while reaction at the 1,4-addition site is rejected due to steric hindrance. This analysis is consistent with the experimental results where the cyclopropanated B ring product was obtained in 57% yield (ref. 16).

Subsequently, "pathway evaluation" addresses competitions between mechanisms issued for *different* nucleophilic sites. It is divided into two sets of rules, known as exclusive and informative. The exclusive rules have been established for obvious competitions such as C *vs.* O silylation with an enolate anion. Tentative products in violation of an exclusive rule are eliminated. The informative rules, in contrast, are for less stringent cases which are often dependent on reaction conditions or on structural changes as in C *vs.* O alkylation with an enolate anion. Violations of informative rules do not typically result in rejection of a product, but do affect the relative yield estimates.

As the oldest reaction processing unit in CAMEO, the basic/nucleophilic module has reached a stage in development appropriate for the implementation of refined rules for asymmetric induction by enolates (ref. 17). Steric hindrance is gauged by an algorithm for estimation of Taft  $E_s$  values (ref. 18), and, consequently, rules for steric approach control can be invoked to predict the stereochemical outcome of reactions such as the asymmetric 1,2 addition shown below (ref. 19).



#### Acidic/electrophilic modules

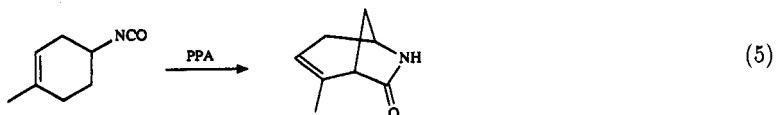
Treatment of reactions that feature cationic intermediates has been divided into three areas: (1) reactions of carbenium ions, (2) electrophilic aromatic substitutions, and (3) a specialized implementation for heterocycle-forming reactions.

**General acid-catalyzed processes** The general acidic/electrophilic module focuses on the generation of carbenium ions and their subsequent addition to multiple bonds, rearrangement, elimination, and quenching by nucleophiles (ref. 6a). The program processes numerous acid-catalyzed functional group interconversions and carbon-carbon bond forming reactions.

Electrophile classification precedes carbenium ion generation. Inorganic electrophiles are classified as acids or the atoms of heteroatom-heteroatom single bonds: protic acids are ranked according to  $pK_a$ , the atoms of heteroatom-heteroatom bonds are rated according to their ability to form cations, and strong Lewis acids are presently presumed to be equivalent in reactivity.

Formation of carbenium ions via addition to carbon-heteroatom multiple bonds is considered first and follows the reactivity hierarchy for electrophilic sites: acyl halides, ketenes > anhydrides > aldehydes, ketones > imines > acids, esters > amides, nitriles. Nucleophiles for quenching electrophiles are classified in three categories; alcohols, amines, and mercaptans are good nucleophiles, carboxylic acids and amides are moderately reactive nucleophiles, and strong acids are poor nucleophiles.

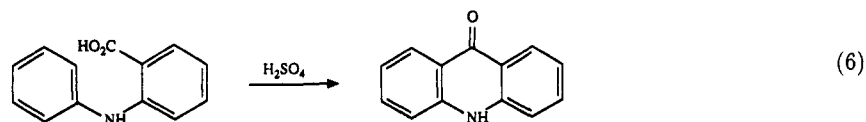
In the absence of a good nucleophile/electrophile pairing, the reactive site is evaluated for addition to other  $\pi$  bonds, *e.g.*, the isocyanate in eq 5 undergoes acid-catalyzed cyclization to form the lactam in 80% yield (ref. 20). At this point, any poor nucleophiles present are permitted to compete. At all reactivity levels, intramolecular quenching is favored over other processes.



Potential solvolytic sites are rated according to (1) stability of the incipient carbenium ion, (2) leaving ability of the departing heteroatom, (3) catalyst, and (4) ionizing ability of the solvent. For addition to carbon-carbon  $\pi$  or cyclopropyl bonds, the most stable ion is formed. Carbenium ions generated in this phase are also considered for Wagner-Meerwein rearrangements to more stable ions. The most readily formed carbenium ions are finally converted to neutral products by quenching with nucleophiles and/or elimination of hydrogen or silyl groups.

**Electrophilic aromatic substitution** The analysis of electrophilic substitutions at sites in aromatic rings incorporates algorithms for gauging the relative reactivities of the available electrophiles, alternative aromatic rings, and individual sites on the aromatic rings (ref. 6b).

Electrophiles are classified according to the type of aromatic system that they can substitute. Strong electrophiles, *i.e.*, those derived from mineral acids, are permitted to substitute any aromatic ring. Moderate electrophiles, such as those involved in Friedel-Crafts alkylations and acylations, are prohibited from reaction with deactivated rings. Poor electrophiles are only capable of substituting strongly activated aromatic systems. All precursors that yield carbenium ions are scanned for possible Wagner-Meerwein rearrangements prior to further processing. When several electrophiles are in competition, those in the most reactive level act to the exclusion of others, unless intramolecular reactions are possible, as in eq 6 (ref. 21). Sulfonation is not predicted in this case even though sulfuric acid is classified as the stronger electrophile.



Following electrophile perception, the set of most reactive rings is determined by calculating a reactivity number (RN) for each ring, where RN is the predicted log of the rate of reaction for the aromatic ring relative to benzene.

The RN estimates account for the effects of heteroatoms, ring fusions, substituents, and reaction medium. The ring(s) with the highest RN plus all others within one unit are retained as the set of most reactive rings, and are then divided into three classes to facilitate perception of reactive sites. For isolated benzenoid systems, the directing effect and steric encumbrance of substituents is assessed. Heteroaromatic rings are classified as  $\pi$ -excessive or  $\pi$ -deficient, in addition to consideration of substituent effects. Finally, polycyclic aromatic hydrocarbons are handled by considering them to be composed of smaller naphthalene, anthracene, or phenanthrene units and applying the rules established for those systems.

**Heterocycle formation** This module is specifically designed for rapid evaluation of possible multistep routes for 3- to 8-membered aromatic and non-aromatic heterocyclic compounds (ref. 6c). For comprehensive analysis, the module not only assesses and performs electrophilic ring formations, but also delegates processing to other mechanistic modules in CAMEO when warranted by the presence of or ability to generate appropriate intermediates.

The evaluation of electrophilic heterocyclic reactions is performed explicitly by the heterocyclic module. Classic examples are the Paal, Knorr, and Hinsberg syntheses. Here, nucleophilic and electrophilic site pairings are assessed with respect to ring-size selectivity rules in addition to inherent reactivity considerations. Thus, sites considered less electrophilic may be selected for reaction if the cyclization pathway is deemed more favorable.

Heterolytic reactions are treated in most CAMEO modules by combining fundamental steps, *e.g.*, proton transfer, substitution, addition, and elimination. Starting materials are permitted to undergo one to three fundamental steps to create a first set of products, which could be resubmitted by the user for further processing. While this method is successful for base-catalyzed reactions, the frequent proton transfer and dehydration steps involved in acid-catalyzed reactions render stepwise treatments cumbersome. For example, enamine formation from a ketone and an amine involves five such steps. The heterocyclic module avoids the proliferation of such intermediates by treating combinations of fundamental steps as one "extended mechanistic step" (ref. 6c). Extended mechanistic steps have been formulated for each of the major types of electrophilic reactions, *i.e.*, addition, addition-elimination, and substitution, thus enabling the user to evaluate complete ring-forming reactions in one pass. For example, Taylor's keto-alkyne cyclization (ref. 22) shown below is processed during one submission cycle in CAMEO. The transformation is composed of two extended steps: addition/dehydration to form an enamine and addition/proton transfer followed by tautomerization.



### Pericyclic module

The mechanistic module for thermal pericyclic reactions performs analyses for 4-, 6-, or 10- $\pi$  electron cycloadditions, electrocyclic processes, sigmatropic rearrangements, and cheletropic reactions (ref. 7). Each reaction class is evaluated separately; cycloadditions, for example, are analyzed with the aid of frontier molecular orbital (FMO) considerations (ref. 7a). Competitions between processes are evaluated following product formation, *i.e.*, all possible pericyclic processes are considered for the input reactants and are compared by an algorithm for the general treatment of periselectivity (ref. 7b).

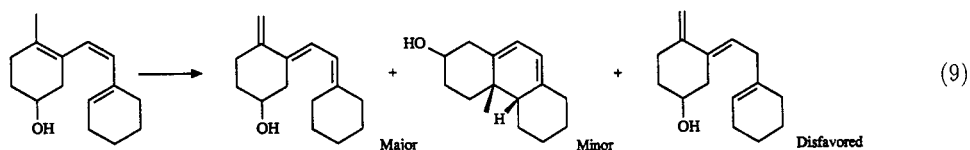
For cycloadditions, general algorithms have been developed for the estimation of the HOMO and LUMO energies of  $\pi$  components, and the polarization of these FMOs. The algorithms accommodate a variety of neutral and charged carbon-carbon, carbon-hetero, and hetero-hetero  $\pi$  systems. The results are used to determine the likelihood and regiochemistry of the cycloaddition under consideration. Endo stereoselectivity is also implemented for Diels-Alder reactions of appropriately substituted reactants. For each cycloaddition, estimated energies of the controlling HOMO-LUMO pair are displayed with the structure. Comparable considerations are made for the other reaction classes with particular emphasis on substituent effects.

Periselectivity is handled in CAMEO by predicting an appropriate temperature range for each reaction: <0, 0-100, 100-200, 200-300, or >300 °C. Naturally, reactions predicted to occur at lower temperatures are favored over those at higher temperatures. For reaction types whose kinetics have been extensively studied, a semiquantitative approach is invoked. The Arrhenius equation (eq 8) has been used to determine limiting energies of activation ( $E_a$ s) to achieve a prescribed rate constant for each of the above five temperature classes. The chosen rate constant standards are  $1.0 \times 10^{-5} \text{ s}^{-1}$  for first order reactions, *e.g.*, electrocyclic and Cope rearrangements, and  $2.2 \times 10^{-5} \text{ L/mol-s}$  for second order reactions, *e.g.*, cycloadditions. These rate constants are consistent with typical experimental conditions and correspond to 75% conversion in approximately 38 hours. Conveniently, the preexponential factor  $A$  is relatively constant among members of a reaction class, and the  $E_a$  range for a given temperature slot is readily estimated.

$$E_a = RT(\ln A - \ln k) \quad (8)$$

Operationally, approximate expressions have been developed to estimate the  $E_a$ s for reactions of each class. The relationship devised for Diels-Alder reactions, for example, accounts for the frontier molecular orbital gap, entropic and energetic features such as the ease with which the diene can achieve *s-cis* coplanarity, intramolecularity and steric effects, and, to a lesser extent, the predicted heat of reaction. The estimated  $E_a$  for a particular reaction is then compared with the stored limiting  $E_a$ s for the reaction type and the appropriate temperature range is selected. For reaction types where the available kinetic data are limited, *e.g.*, [1,3] and [2,3]-sigmatropic rearrangements, a qualitative approach is employed. Various structural features are identified and used to determine the appropriate temperature range. The predicted temperature range and heat of reaction are displayed for each product. The latter is often illuminating in comparisons between products that are predicted to occur in the same temperature range.

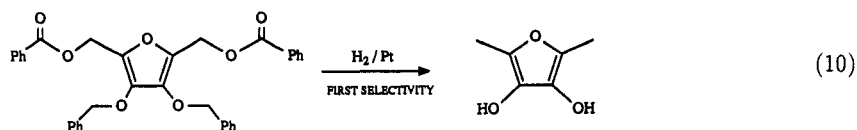
Equation 9 illustrates periselectivity between electrocyclic ring closures and sigmatropic hydrogen shifts. Formation of the [1,7]-hydrogen shift product is predicted to predominate; the predicted temperature range is 0-100 °C with an estimated  $E_a$  of 19 kcal/mol that is comparable to the experimental value of 20.9 kcal/mol (ref. 23). In addition, the program predicts a disrotatory 6 $\pi$  electrocyclic closure in the 100-200 °C range as a minor component, as well as a disfavored [1,5]-sigmatropic hydrogen shift at 200-300 °C.



### Oxidation-reduction module

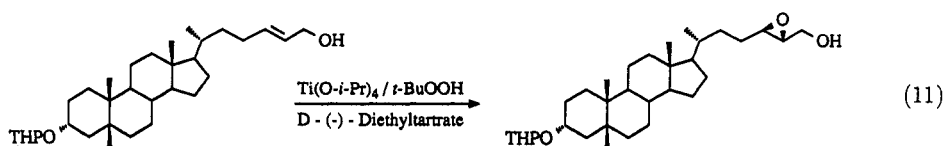
The implementation of a module for the analysis of oxidation and reduction reactions requires a unique approach (ref. 8). The procedure used for mechanistically based modules keys reactivity on substrate functionality; such a technique is not amenable to reactions where (1) the mechanisms are obscure or limited to specific reagent-substrate combinations, (2) the reactions involve a diversity of intermediates, and (3) the reaction pathways are highly dependent on reaction conditions such as temperature, pH, and stoichiometry. Consequently, the approach in this module keys reactivity on the reagent and employs specific reaction conditions to distinguish among the possible competitions. The subprograms comprising this module thus correspond to individual reagents, and within each a reactivity hierarchy is established for determining functional group lability.

Consistent with this unique approach is the development of a reagent menu from which users *must* select reagents and reaction conditions. About 50 commonly used oxidizing and reducing agents are currently considered. Reaction stoichiometry is handled with three choices: 1 EQUIVALENT, FIRST SELECTIVITY, and EXCESS. The default choice, FIRST SELECTIVITY, causes the program to automatically determine the number of equivalents needed to achieve the first selective conversion of the substrate. Hence, in eq 10 all four side chains are reduced with this stoichiometry, whereas choosing 1 EQUIVALENT would have resulted in a mixture of four monohydrogenated products (ref. 24).



Since mechanisms are not known for many transformations, competitions are often assessed by use of reactivity tables derived from experimental observations of reaction conditions and product distributions. When mechanistic data are available, they are used to analyze competitions more generally since intermediates can be considered and duly evaluated. For example, the  $pK_a$ ,  $E_s$ , ion stability, or FMO determining routines discussed previously may be invoked, or transfer to an appropriate mechanistic module may occur to facilitate processing.

The ability to evaluate stereocontrol is a major concern for oxidation-reduction reactions. For example, algorithms for stereorecognition permit formation of the appropriate Sharpless epoxidation product in eq 11 (refs. 8b, 25).



### Free radical module

Processing of organic free radical chain and diradical reactions encompasses four major phases, *i.e.*, radical initiation, primary propagations (specifically, intermolecular abstraction), secondary propagations, and radical trapping. The operational premise of the module is that complex interconversions may be explained and predicted by reducing processes to a series of fundamental steps, and chains are permitted to propagate by automatic resubmission of generated intermediates (ref. 9). Hence, chains are initiated by bond homolysis or cycloaromatization reactions, and each intermediate is subjected to the same detailed analysis to determine its potential for abstraction, addition, or fragmentation. When no new intermediates result, closed-shell products are formed by an implicit chain transfer step; recombinations and disproportionations are also considered during processing of diradicals.

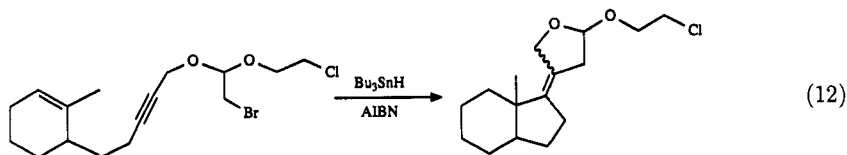
Initiator molecules are sought according to relative decomposition rates. Simple homolysis of the weakest available bond is insufficient for subsequent processing, since initiators may double as chain transfer agents or may act only as catalysts, and relative trapping rates must be known for assessment of competitions between trapping and propagation. Radicals generated at all stages of processing are subjected to brief characterization for determination of radical site location, philicity, and stability. For diradicals, the relative reactivity and relationship between the two radical sites is also determined.

The primary propagation phase segregates intermolecular abstractions from remaining propagations to prevent indiscriminant abstraction by resubmitted intermediates. Identification of a chain-transferring fragment may occur during this phase if the initiating species is recognized as non-catalytic. For example,  $Bu_3SnH$  is generally present in slight excess and may perform the dual role of initiator and chain transfer agent. However, initiators such as peroxides or AIBN are used as catalysts only, and chain transfer is assumed to occur via abstraction from a duplicate of the original abstraction target.

The bulk of radical processing occurs during the secondary propagation phase. Fragmentations, ring closures, additions, and intramolecular abstraction reactions are executed according to the separation of processes into four kinetic subphases. Only processes in the subphase with the highest rate are performed to avoid formation of spurious products, *e.g.*, intermolecular additions are not considered if favorable 5- or 6-membered ring closures are possible. Subroutines for the detailed analysis of fragmentations, cyclizations, additions, and intramolecular abstractions are invoked to determine the relative feasibility of competing processes, and consider structural features such as bond dissociation energies, ring strain, stereoelectronic requirements, and stabilizing substituents. Diradicals are processed by alternate consideration of each radical site, unless disparate stabilities render one site "passive". Diradicals are also subjected to processing for reactions specific to 1,2-, 1,3-, or 1,4-diradicals during this phase.



The self-propagating nature of radical reactions is mimicked by resubmitting intermediates automatically until no new radicals result. Thus, reactions involving several discrete steps may be effected by one pass through the module without loss of mechanistic purity: all intermediates are explicitly formed and are available for a stepwise trace of program decisions. For example, the tandem cyclization in eq 12 involves six steps executed automatically (ref. 26).



Naturally, the algorithms are designed such that propagation is halted when less favorable or structurally infeasible radicals would be generated. When resubmission yields no new intermediates, the final phase of processing generates closed shell products from selected intermediates. For chain reaction sequences, this entails enacting the implicit chain transfer step described previously. If diradicals or radical pairs are present, implicit chain transfer may compete with recombination and/or disproportionation.

### Carbene module

Implementation for the chemistry of carbenoid intermediates is focused on establishing a general framework for evaluating the generation of these reactive intermediates and their participation in subsequent addition, insertion, and rearrangement reactions (ref. 10). The most commonly used methods to form carbenes, *i.e.*,  $\alpha$ -eliminations under basic conditions and photolytic decompositions of diazo precursors, are presently handled. Competitions for available reactions of the intermediates are addressed by perceiving and ranking potentially reactive sites. Reactions are typically performed only with sites in the highest available reactivity levels, although additional considerations are made for ring size selectivity and steric hindrance. Even at this early stage in development, the module is able to properly evaluate such well known reactions as the Skattebol and Wolff rearrangements and the Bamford-Stevens reaction.

## SAMPLE REACTION SEQUENCES

The recent expansions of the CAMEO program to encompass processing of a variety of intermediates permits whole reaction sequences to be examined. Algorithm development and testing must accommodate reported experimental findings. Consequently, comparisons of program output to two prominent experimental reaction sequences are illustrated in the following schemes and are discussed below.

The CAMEO analysis for Curran's recent hirsutene synthesis is presented in Scheme 1 (ref. 27). Selective conversion of ester **1** to the silyl enol ether **2** was accomplished by the basic/nucleophilic module. C-Silylation does not interfere as the program recognizes O-Si affinity (ref. 5c). Submission of **2** to the pericyclic module gave the Claisen rearranged **3** as the major product, although the predicted temperature range (100-200 °C) was slightly higher than the experimental 60 °C. The disfavored [2+2] and retro-ene products were predicted to require significantly higher temperatures (>300 °C).

Submission of **3** with PhSeCl to the acidic module gave **4** exclusively. Conversion to the olefin **6** occurred cleanly by oxidation with H<sub>2</sub>O<sub>2</sub> followed by pyrolysis in the pericyclic module. The high degree of selectivity predicted by the CAMEO program for these reactions is supported by the reported overall yield of 62% for **1** → **6** (ref. 27).

Submission of lactone **6** with the alkyl cuprate to the basic/nucleophilic module resulted in two major products; direct addition/elimination upon the ester provides **7**, while S<sub>N</sub>2' attack gives the reported product **8**. Experimentally, **8** was recovered in yields of 50-75% (ref. 27). Direct S<sub>N</sub>2 displacement of the lactone would involve intermolecular attack at a tertiary site and was thus rejected by the program.

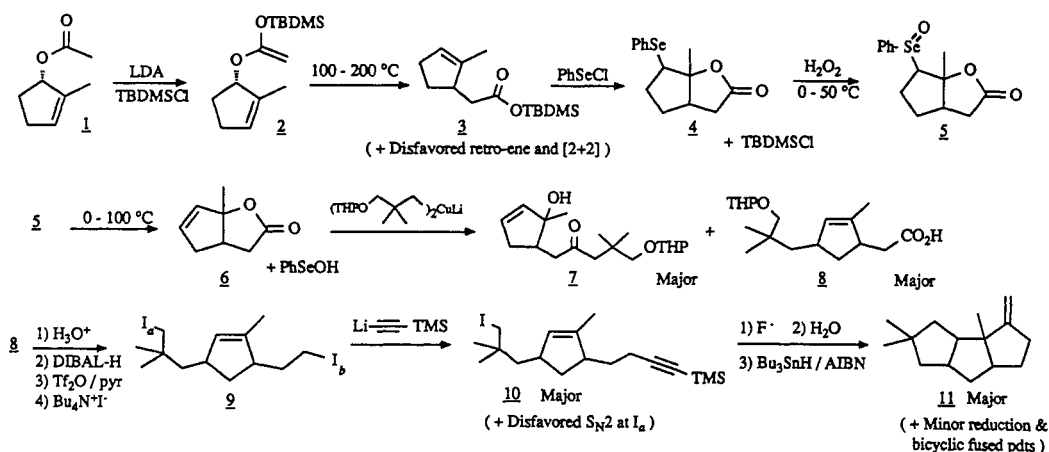
Several functional group interconversions were necessary for transforming **8** to **9**. Acidic hydrolysis of the OTHP group followed by DIBAL-H reduction of the carboxylic acid resulted in the reported diol. Conversion to a ditriflate was carried out by the basic/nucleophilic module. The program suggested that this transformation might occur by initial addition/elimination of Tf<sub>2</sub>O by pyridine, followed by a second addition/elimination by the neutral alcohol sites yielding the di-triflate. Alternatively, the program suggested

that pyridine might act as a base by deprotonating the alcohol sites, thus facilitating formation of the ditiolate. Finally, conversion to the diiodide by treatment with  $\text{Bu}_4\text{N}^+\text{I}^-$  resulted in formation of **9**, although the program suggested that displacement at the neopentyl position should be slow due to its steric encumbrance (ref. 17).

Conversion of **9** to **10** is also controlled by recognition of steric environments. Processing of **9** with lithium trimethylsilylacetylide in the basic/nucleophilic module gave **10** as the major substitution product;  $\text{S}_{\text{N}}2$  reaction at the neopentyl site was displayed with a “disfavored” ranking. Desilylation of **10** with fluoride ion yielded an acetylene poised for tandem cyclization. An intramolecular displacement of iodine by the intermediate acetylide anion was also performed, albeit with a disfavored relative yield estimate.

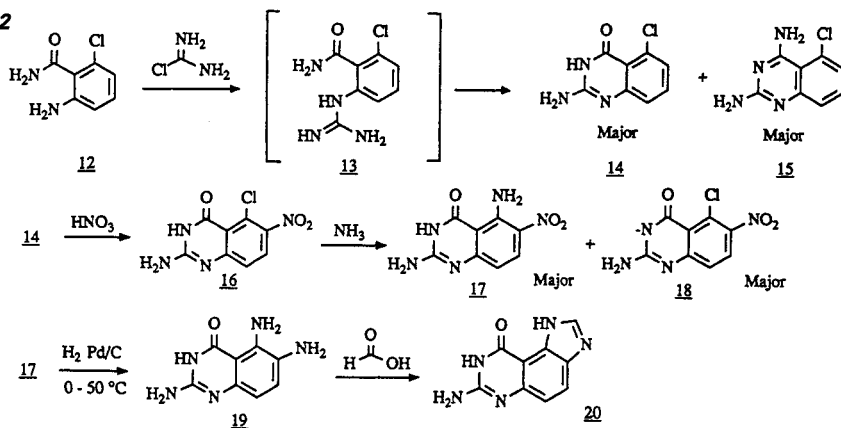
Treatment with  $\text{Bu}_3\text{SnH/AIBN}$  in the free radical module effected the regiospecific tandem cyclization **10**→**11** in one submission to the program. The preference for exo ring closure in both cyclizations reflects rules for (1) the exclusion of bridge formation when fusions are possible, and (2) the stereoelectronic prohibition of endo closure upon acetylenes. Simple reduction at the iodide site and trapping of a monofused product were also predicted as minor components due to the known high trapping rate of stannyl hydrides (ref. 28).

Scheme 1



Scheme 2 summarizes the evaluation of a recent synthesis of proximal-benzohypoxanthine **20** (ref. 29). The heterocyclic module identified the aniline nitrogen and chloroformamidine carbon in **12** to be the most reactive sites. Hence, evaluation lead to the intermediate **13** which the module automatically resubmitted to give **14** via addition/ $\text{NH}_3$  loss and **15** by addition/dehydration. Experimentally, only **14** was reported in quantitative yield (ref. 29). Although the benzenoid ring in **14** was not predicted to be very reactive ( $\text{RN} = -10.1$ ), nitration gave exclusively **16**, corresponding to the only experimentally observed product. Submission of **16** with  $\text{NH}_3$  to the basic/nucleophilic module gave the reported substitution product **17**. This product was accompanied by the proton transfer structure **18** due to the low  $\text{pK}_a$  predicted for the amide nitrogen: 10 in DMSO (ref. 12). Reduction of the nitro group of **17** by the oxidation/reduction module followed by submission of the sole product **19** with formic acid to the heterocyclic module gave exclusively the desired proximal-benzohypoxanthine **20**, consistent with the experimental yield of 58% (ref. 29).

Scheme 2



## COMPUTATIONAL DETAILS

The CAMEO program has been under development since 1975, and presently comprises over 900 FORTRAN subroutines and *ca.* 100,000 lines of code. Development is carried out on Digital Equipment Corporation VAX computers using the VMS operating system with the graphical interface programmed for Tektronix compatible terminals. UNIX and AIX versions have also been created.

CAMEO has capabilities for providing sophisticated analyses of reactions involving a variety of intermediates. In general, algorithms have been implemented to err on the lenient side; rules permitting occasional generation of a less likely product are preferable to restrictive rules that might reject a legitimate possibility. The abbreviated discussions of reaction processing algorithms presented here highlight the general logic applied. The familiar features of mechanistic reasoning are therefore evident, *i.e.*, locate and rank electrophilic and nucleophilic sites, and carry out reactions for each site pairing. The challenge lies in determining what structural features contribute to reactivity, and in striking a subtle balance between competing features. For thorough discussions of algorithms for each module, the interested reader is referred to references 5-10.

## Acknowledgements

Gratitude is expressed to the earlier workers listed in the references whose efforts formed the background for development of the current program, CAMEO/90. Financial support has been provided by the National Science Foundation, Imperial Chemical Industries, Procter and Gamble, Pfizer Ltd., W. R. Grace & Co., Rhone-Poulenc, Hoechst AG, Sandoz AG, E. Merck AG, Eastman Kodak Co., Schering AG, and IBM.

## REFERENCES

1. For example, see: R.B. Woodward in Perspectives in Organic Chemistry, A.R. Todd, Ed., p. 155, Interscience, New York (1956).
2. (a) E.J. Corey, Pure Appl. Chem. **14**, 19 (1967). (b) E.J. Corey, Quart. Rev., Chem. Soc. **25**, 455 (1971). (c) E.J. Corey and W.T. Wipke, Science **166**, 178 (1969). (d) E.J. Corey, A.K. Long and S.D. Rubenstein, Science **228**, 408 (1985) and references therein.
3. (a) M. Bersohn and A. Esack, Chem. Rev. **76**, 269 (1976). (b) R. Carlson, C. Albano, L.G. Hammarstrom, E. Johansson, A. Nilsson and R.E. Carter, J. Molecular Science **2**, 1 (1983). (c) A.K. Long, S.D. Rubenstein and L.J. Joncas, Chem. Eng. News **61** (19), 22 (1983). (d) J. Haggin, Chem. Eng. News **61** (19), 7 (1983). (e) R. Barone and M. Chanon, Computer Aids to Chemistry, G. Vernin and M. Chanon, Eds., p. 19, Ellis Horwood Limited, West Sussex, England (1986). (f) H.B. Hendrickson, Acc. Chem. Res. **19**, 274 (1986). (g) N.S. Zefirov and E.V. Gordeeva, Russ. Chem. Rev. **56**, 1002 (1987). (h) M. Marsili, Computer Chemistry, Chapter 7, CRC Press, Boca Raton, Florida (1990).
4. An exception is the program EROS, which considers physical parameters such as charge or polarizability, rather than functionality, to generate bond isomerizations. J. Gasteiger, M. Hutchings, B. Christophe, L. Gann, C. Hiller, P. Low, M. Marsili, H. Saller and K. Yuki, Top. Curr. Chem. **137**, 18 (1987).
5. (a) T.D. Salatin and W.L. Jorgensen, J. Org. Chem. **45**, 2043 (1980). (b) T.D. Salatin, D.R. McLaughlin and W.L. Jorgensen, J. Org. Chem. **46**, 5284 (1981). (c) C.E. Peishoff and W.L. Jorgensen, J. Org. Chem. **48**, 1970 (1983). (d) C.E. Peishoff and W.L. Jorgensen, J. Org. Chem. **50**, 1056 (1985). (e) P. Metivier, A.J. Gushurst and W.L. Jorgensen, J. Org. Chem. **52**, 3724 (1987). (f) A.J. Gushurst and W.L. Jorgensen, J. Org. Chem. **53**, 3397 (1988).
6. (a) D.R. McLaughlin, Ph.D. Thesis, 1983, Purdue University. (b) M.G. Bures, B.L. Roos-Kozel and W.L. Jorgensen, J. Org. Chem. **50**, 4490 (1985). (c) M.G. Bures and W.L. Jorgensen, J. Org. Chem. **53**, 2504 (1988). (d) R.T. Mosley, M.S. Thesis, 1989, Purdue University.
7. (a) J.S. Burnier and W.L. Jorgensen, J. Org. Chem. **48**, 3923 (1983). (b) J.S. Burnier and W.L. Jorgensen, J. Org. Chem. **49**, 3001 (1984). (c) C.H. MacMahon, M.S. Thesis, 1989, Purdue University. (d) G.D. Paderes and W.L. Jorgensen, to be published.
8. (a) G.D. Paderes and W.L. Jorgensen, J. Org. Chem. **54**, 2058 (1989). (b) S. Sinclair, M.S. Thesis, 1990, Purdue University. (c) P. Metivier, G.D. Paderes and W.L. Jorgensen, to be published.
9. (a) E.R. Laird and W.L. Jorgensen, J. Org. Chem. **55**, 9 (1990). (b) E.R. Laird and W.L. Jorgensen, J. Chem. Info. and Comp. Sci. (1990), in press.

10. H. E. Helson and W.L. Jorgensen, to be published.
11. H. E. Helson, Ph.D. Thesis, 1990, Purdue University.
12. A.J. Gushurst and W.L. Jorgensen, J. Org. Chem. **51**, 3513 (1986).
13. J. E. Toth, P.R. Hamann and P.L. Fuchs, J. Org. Chem. **53**, 4694 (1988).
14. The performance of the basic/nucleophilic module is critically dependent on accurate  $pK_a$  predictions; see refs. 5e and 12. The  $pK_a$  window is a function of temperature; the room temperature window of 4  $pK_a$  units is expanded or contracted with higher or lower temperature range selection.
15. Y. Mazur and M. Nussim, J. Am. Chem. Soc. **83**, 3911 (1961).
16. J. Pfister, C. Lehmann and H. Serhri, Helv. Chim. Acta. **51**, 1505 (1968).
17. J.M. Fleischer, A.J. Gushurst and W.L. Jorgensen, to be published.
18. S.H. Unger and C. Hansch, Prog. Phys. Org. Chem. **12**, 91 (1976).
19. S.J. Danishefsky, H.G. Selnick, R.E. Zelle and M.P. DeNinno, J. Am. Chem. Soc. **110**, 4368 (1988).
20. H. Knecht and W. Haeffiger, Tetrahedron Lett. **25**, 281 (1984).
21. S. Patai, The Chemistry of the Quinoid Compounds, Part 1, Wiley, New York, (1979).
22. K.A. Richardson, J. Saunders and R.J.K. Taylor, Tetrahedron Lett. **26**, 1171 (1985).
23. C.W. Spangler, Chem. Rev. **76**, 187 (1976).
24. D.W. Henry and R.M. Silverstein, J. Org. Chem. **33**, 140 (1968).
25. C.K. Lai and M. Gut, J. Org. Chem. **52**, 685 (1987).
26. G. Stork and R. Mook, J. Am. Chem. Soc. **105**, 3720 (1983).
27. D.P. Curran and D.M. Rakiewicz, J. Am. Chem. Soc. **107**, 1448 (1985).
28. A.L.J. Beckwith and G. Moad, J. Chem. Soc., Perkin Trans. 2, 1726 (1975).
29. S.W. Schneller and A.C. Ibay, J. Org. Chem. **51**, 4067 (1986).