The ROBIA Program for Predicting Organic Reactivity

Ingrid M. Socorro and Jonathan M. Goodman*

Unilever Centre for Molecular Science Informatics, University Chemical Laboratory, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, U. K.

Received September 6, 2005

A program to predict organic reactions, ROBIA, has been developed. It achieves reaction prediction on the basis of coded rules and molecular modeling calculations, generating possible transition states, intermediates, and products given the starting material and reaction conditions. The program generates all possible reaction pathways, on the basis of the selected transformations within its database, and evaluates them selecting the most feasible ones. The program has been successfully tested against several examples.

1. INTRODUCTION

The design and preparation of novel chemical compounds with specified characteristics have important roles in discovery, the construction of new materials, agrochemicals, and so forth. Synthetic organic chemistry is an indispensable tool for this, but organic chemistry is full of interesting but also surprising reactions. The prediction of organic reactions is a challenging task that requires deep insight into chemical reactivity. Software systems can assist chemists in solving the problems met in organic synthesis by helping them to anticipate, analyze, and understand their results. Programs[1,2] have been developed for this purpose since the first approach to computer-aided organic synthesis[3] (CAOS) in the late 1960s. Two different strands exist in the development of programs for CAOS. One, based on the retrosynthetic analysis, leads to synthesis design systems. After determination of strategic sites on a given target molecule, they propose some of many possible synthesis routes. Therefore, their output is dependent on their own retrosynthesis strategies and can vary from one program to another. The second approach is based on the forward direction of a reaction. This leads to reaction prediction programs, which evaluate possible reactions given the reactants and specific reaction conditions. They, in contrast, should give the same answer.

These two approaches to CAOS have been used over the past decades to develop programs for synthesis planning[4−8] and reaction prediction[9−12] with the aim of assisting chemists to do their jobs more effectively. Programs for the prediction of metabolic routes[13−15] have been developed too. They all involve the analysis of chemical transformations. A computational method for the determination and analysis of chemical reactivity should be a useful tool in synthesis planning and drug design.

We describe in this paper an algorithm for reaction prediction based on organic chemistry rules and molecular modeling. This approach is, to our knowledge, the first attempt for using such a hybrid system, which combines rule-based techniques along with molecular mechanics and quantum chemistry, in reaction prediction. This new program, ROBIA (Reaction Outcome By Informatics Analysis), operates in the forward direction predicting the outcome of organic reactions given the starting materials and conditions.

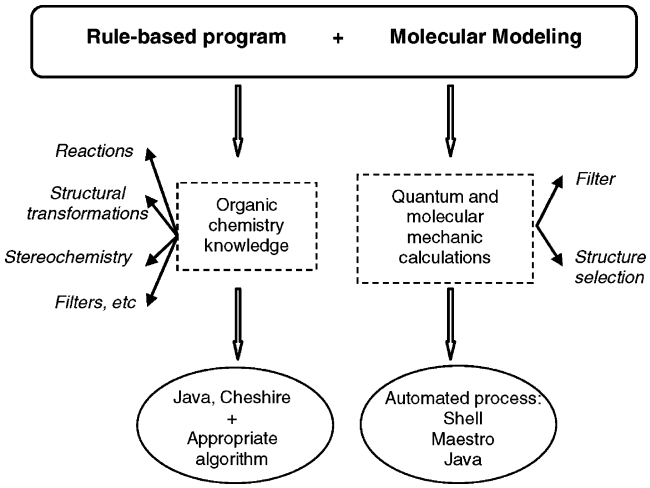


Figure 1. Approach to reaction prediction.

The system arrives at its conclusions by application of a series of rules designed to consider different features in molecules for the determination of their reactivity. In this way, the program determines reactive sites and which bonds are to be broken or formed. It also contains filters for selecting the best possible products based on rules and molecular and quantum mechanics calculations. Thus, it should be possible to predict the reactivity of substrates for which there is no direct experimental data.

2. METHOD

1. Approach to Reaction Prediction. Chemical reactivity is simultaneously influenced by many factors including the reactants (e.g., structural features, electronic properties, etc.) and the reaction conditions (e.g., temperature, etc.). A chemical reaction occurs as a result of complicated interactions between them. To achieve an accurate reaction prediction, ROBIA aims to consider, in each case, the factors responsible for the driving forces of a reaction. For this, our approach to reaction prediction (Figure 1) is based on organic chemistry rules and molecular modeling. This leads to a rule-based program in which chemical reaction knowledge (reactive sites, chemical transformations, stereochemistry, etc.) is coded derived from the collection of a set of organic

10.1021/ci050379e CCC: $33.50
Published on Web 01/21/2006

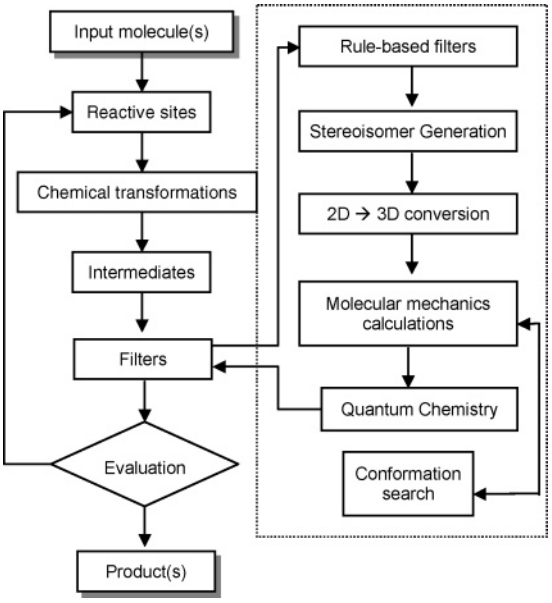


Figure 2. Algorithm for ROBIA.

chemistry rules. Thus, the program generates all possible structures according to its rules. Structure selection is then made by the use of coded filters together with molecular and quantum mechanics calculations. Molecular modeling has been integrated with the rule-based program for this purpose.

In this way, ROBIA is able to generate all possible reaction pathways within its rule set and to automatically assess the most favorable route.

The approach to reaction prediction is based on a mechanistic approach. The program goes from the reactants to the products through a sequence of intermediate steps in which all possible intermediate structures are formed, then filtered, and selected according to the coded rules and the reaction conditions. The algorithm of the program is shown in Figure 2. ROBIA has been coded using the Java[16] programming language and MDL Cheshire[17] chemical scripting language. An overview of the program's code is shown in Figure 3. Shell scripts have been developed for the integration of molecular modeling with the rule-based program.

The shell scripts allow the whole process of reaction prediction to be automated. They are responsible for launching the programs used for the calculations (Maestro,[18] MacroModel,[19] Jaguar,[20] GAMESS,[21] and MOPAC[22]) and Java programs that have been developed to prepare input files needed for the calculations and to analyze output files. The user interface is MDL Isis/Draw, where the structures of the reactants are entered and the products as well as any other structures generated during the process are displayed. The steps of the algorithm, grouped into those performed by the rule-based program and those involved in the calculations, are described in detail in the following subsections.

1.1. Rule-Based Program. This is the part of the system that analyzes the chemical structures and determines their reactivity by application of a series of coded rules (see Table 1) designed to consider different features in them. Thus, it contains the code for the determination of the reactive sites in a molecule within the *reactive sites block* and the instructions to perform any of the reactions contained in the *chemical transformations block*. It also includes rules (see Table 2) for structure filtering and other code needed to prepare the structures for the calculations as part of the *filters block*. Two programming languages have been used for its development: Java and Cheshire. The Java code controls the program since it first starts reading the structures of the reactants, making use of Cheshire scripts when needed to perform particular tasks on them or any other structure generated. Cheshire scripts cannot run as stand-alone pro-

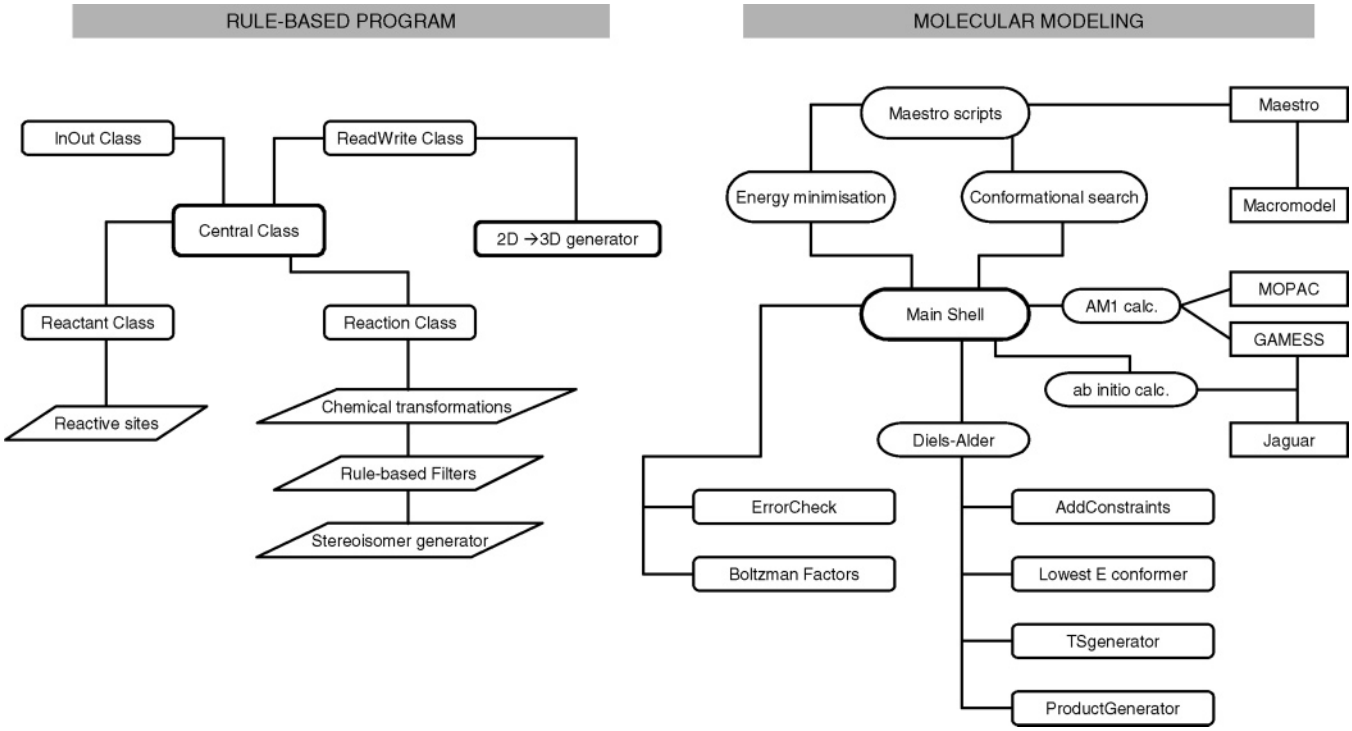
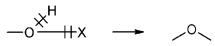
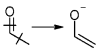
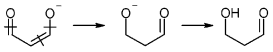
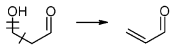
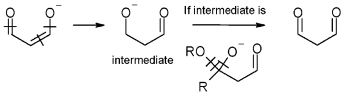
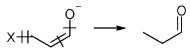
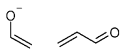
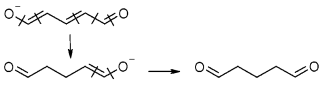
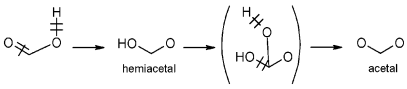
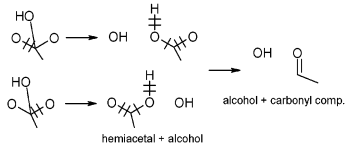
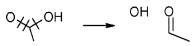
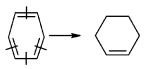


Figure 3. ROBIA's structure.

Table 1. Contents of the Chemical Transformation Block[a]

type of transformation	rule (i) – substructure(s) to be found in target molecule(s)	rule (ii) – conditions required	transformation[b]
Nucleophilic substitution		base	
Enolate formation reaction		acid/base	
Aldol reaction		acid/base	
Retro-aldol reaction		acid/base	
Dehydration of aldol product		acid/base	
Claisen condensation		acid/base	
Alkylation of enolates		acid/base	
Michael addition		acid/base	
Hemiacetal / Acetal formation		acid/base for hemiacetal acid for acetal	
Acetal decomposition		acid for hemiacetal acid/base for carbonyl compound	
Hemiacetal decomposition		acid/base	
Diels-Alder reaction		---	

[a] X, any halogen. [b] Marks on bonds: discontinuous line for a bond to be formed, a single line on a bond for change in bond order, a double line on a bond for bond to be broken.

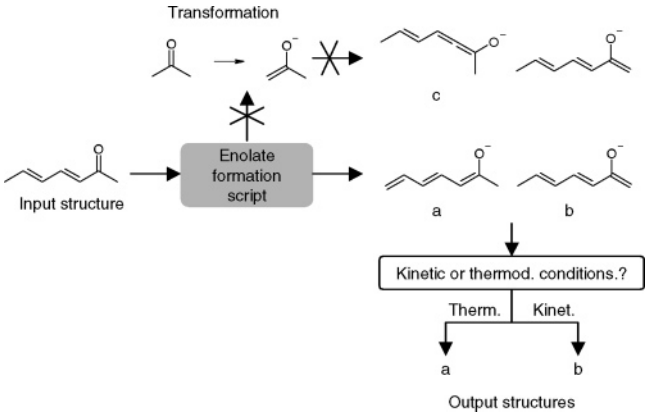
Table 2. Rule-Based Filters

rule-based filter	filter action
ring size	option to exclude three-/four-member rings if five/six possible
identical structures	remove structure if it has already been formed in the same step of the process

grams and have to be called from a Java environment. This way, once the input structures are ready and the program is executed, the main class in Java (central class) starts running (Figure 3). This class contains the instructions to (i) read and analyze the input structures, (ii) perform chemical transformations on them, (iii) filter the generated structures, and (iv) generate all possible stereoisomers from the remaining structures. These instructions are executed in a sequential order from i−iv. Once each instruction is reached, a cascade of other instructions contained in the *reactive sites*, *chemical transformations*, and *filters blocks* is unleashed.

1.1.1. Reactive Sites Block. The information about a structure entered in the reactive sites block (the molecule drawn by the user or any intermediate structure generated by the program) is available to the program as a connection table within its molfile[23] (molfiles are MDL's structure description files). Cheshire contains chemically intelligent methods for the perception of the structure able to detect the stereochemistry of double bonds, aromaticity, the configuration of chiral carbon atoms, and so forth, which can be used by the program whenever this type of information is required. ROBIA uses a method[24] based on substructure searching for the recognition of functional groups. For this, the program maps between the structure under study and a list of functional groups. So far, the program recognizes 16 functional groups: alcohols, carboxylic acids, aldehydes, ketones, esters, ethers, acetals, hemiacetals, acyl chlorides, alkylhalides, amides, nitroalkanes, amines, nitriles, epoxides, and acid anhydrides. Each functional group is created as a

Scheme 1. Simple Example of a Transformation



pattern formed by a group of atoms and bonds using MDL line notation (a string that represents a group of atoms and bonds linked together). The program has an algorithm to map these patterns in the target molecule and determine the functional groups contained in the molecule within its list of functional groups. This list could be extended in the future when new reactions involving other functional groups are included in the program.

1.1.2. Chemical Transformations Block. This part contains a list of reactions available to the program (Table 1) from which one is selected at a time on the basis of the information obtained previously and the reaction conditions (further development of the automated reaction selection is under study). It is also linked to methods that are able to interact with the user, for example, to prompt the user about the reaction conditions.

Once a reaction has been selected within its data set, the program uses Cheshire scripts to perform the selected transformation. It is our aim to increase ROBIA's predictive power by extending the content of the transformation block to more reactions and improving the algorithms of the transformations already contained in ROBIA, making them more general to cover a larger number of situations for each reaction type. The methods for chemical transformations explore the reactive site in a molecule and its environments, determining which atoms and bonds are involved in the transformation. This way, the program generates all possible structures according to its coded rules. A simple example is shown in Scheme 1. An enolate transformation script is applied to the input structure. The program recognizes the ketone group and also its conjugation with the double bonds, thus, forming products **a** and **b**. It does not use a ketone−enolate transformation without considering the rest of the molecule, generating the high-energy structure **c**. Finally, the program selects **a** or **b** as the output structure taking into account the reaction conditions.

1.1.3. Rule-Based Filters. After a reaction is completed, the program uses its *filters block* on all the generated molecules to filter those that are unlikely to be formed as a result of ring strain, high-energy structure, and so forth. Two kinds of filters are applied for this. One is described here, formed by coded rules (rule-based filters, see Table 2), and another is based on molecular modeling results (see subsection 1.2). Two rule-based filters can be used by the program at present. It is our aim to add extensions in the future to improve the efficiency of this part of the process. One of these filters contains a method to filter identical structures that are formed during the same step. This method makes a comparison of the molfiles of the molecules produced after a chemical transformation, discarding those that represent the same structure. The program also filters structures by ring size. The other rule-based filter has a method that filters structures in which three- or four-membered rings have been formed during the reaction. This method first collects all the atoms and bonds linked by a ring to the new bond(s) formed in the molecule after a reaction; then, it creates query rings of increasing size starting from three-membered rings up to six, and it maps them into the previous collection. This way, if the size(s) of the new ring(s) formed is comprised between a three- and six-membered ring, the program gets information about the ring size and is then able to discard three- or four-membered ring structures. Otherwise, molecules with rings bigger than six are not filtered and continue to the next step of the process. This method is reaction specific and can be switched off; for example, this filter would not be applied to a three-membered ring structure formed after an epoxidation.

1.1.4. Stereoisomers Generator. After a chemical transformation, new chiral centers can be generated in the product(s), which have an uncertain configuration unless specified in the reaction script. Also, the input molecule(s) drawn by the user might have uncertain stereocenters. However, the stereochemistry of a molecule has to be known before being entered into any of the molecular modeling packages. A script has been developed in ROBIA to generate all possible stereoisomer molecules by assigning a configuration to the undefined chiral centers while maintaining the configuration of the rest. For this, the program generates a matrix of 0 and 1 elements in which each row represents a binary number starting from the zero binary number to the total number of stereoisomers that could be generated (calculated from the number of uncertain chiral centers in the molecule). The 0 and 1 elements are then associated with marking a bond "up wedge" or "down wedge", respectively (the fourth column of the bond block of a molfile represents the stereochemistry of a bond, numbers 1 and 6 being assigned for up and down wedges respectively). In an early version of this script (used in Example 1), the bond chosen by the program to be marked could be any of the bonds formed between the chiral carbon considered and its substituents. Some rules stating a priority on the bond to mark have been recently added to this script to obtain a better representation of the generated stereoisomers: a bond between the chiral carbon and any of its substituents in a chain is preferred over a carbon−hydrogen bond (for this, the program first attaches the hydrogen atom to the asymmetric carbon and then marks the bond, as hydrogen atoms are not explicitly represented in molfiles), which is preferred over a chiral carbon−substituent in-ring bond. This way, the program is able to form every stereoisomer from each row, as each row represents one of the possible up−down wedge combinations (the whole matrix covering all combinations of R−S configurations).

1.1.5. 2D to 3D Conversion. The input and output structures of the rule-based program are 2D structures contained in molfiles [hydrogen atoms are not represented unless the stereoisomers generator explicitly makes a hydrogen atom (or more) appear in the molfile]. However, the 3D structures are needed for the calculations. To obtain 3D

structures from 2D, the program follows a sequence of steps (the molecules enter the molecular modeling section for steps ii and iii, see next section for further details): *(i) Reveal the stereochemistry.* Although the molfile format has information about the stereochemistry of a molecule, it represents flat molecules since its *z* coordinate is formed by a column of zero numbers. ROBIA contains a program that reveals the stereochemistry of the molecules by adding a number to the *z* coordinate of their molfile, thus obtaining the 3D structure. For this, the program first looks for the chiral carbons in the molecule (marked with 1 or 2 in the seventh column of the atom block of a molfile); then, it looks for the bonds with stereochemistry (marked with 1 or 6 for "up" and "down" in the fourth column of the bond block of a molfile) attached to them. Finally, it adds +1.0000 to the *z* coordinate of the correspondent atom with a 1 marked in the bond stereo column or −1.0000 if marked with a 6. *(ii) Energy minimization.* The program uses a Maestro script that adds the hydrogen atoms to the molecule and optimizes it by doing an energy minimization on it. In cases where the molecule has no stereochemistry, it enters the molecular modeling section flat. The addition of the hydrogen atoms by using this Maestro script gives the 3D structure, which is improved with the energy minimization. *(iii) Conformational search.* Finally, the minimized structure is optimized by performing a conformational search on it, from which the lowest energy structure obtained is chosen.

This way of obtaining 3D structures has given good results so far, although it might present some problems with complex polycyclic structures. There are a number of programs available for 2D-to-3D conversion, including CORINA[25] and CONCORD.[26] The aim of these programs is to generate a single good 3D conformation as rapidly as possible. ROBIA has an easier task to do, as the conformations generated will be submitted to a conformation search. Even if the structures the script generates are rather strained, low-energy structures will be found in the subsequent conformation analysis. This process takes much more time than CORINA or CONCORD but should generate all the low-energy conformations accessible to the molecule, instead of just one good conformation.

1.2. Molecular Modeling. Molecular modeling has been integrated within the rule-based program because of its important contribution in the process of structure filtering and selection during reaction prediction. Energy studies on the transition state structures make possible the selection of the kinetic products, while thermodynamic processes are considered when analyzing the energies of the products. ROBIA is able to generate both kinetic and thermodynamic products, and in some cases, it can select between them giving a specific output for a kinetic or thermodynamic reaction (Scheme 1). However, in the majority of cases, the coded rules are not enough to filter or select structures and energy calculations are required.

After the input structures (reactants or intermediates) have been analyzed, transformed by application of a coded reaction, and filtered, all the possible stereoisomers are formed from the remaining structures and converted into 3D molecules ready for the calculations.

It is our aim to make the whole calculation process automated. Shell and Maestro scripts together with some Java programs have been developed for this purpose. The Shell scripts run the programs used for the calculations. They are also able to distribute the work among a network of machines if required and organize the generated data. The Maestro scripts are used for energy minimizations and conformational searches. They are able to prepare the input structures (e.g., addition of hydrogen atoms) and to set the parameters needed for the calculations (e.g., number of steps, force field, etc.). They also use the "perform automatic setup" option in Maestro for conformational searches. This way, the torsion angles to check, the ring-opening locations, the bonds to rotate, and so forth are automatically selected. Some Java programs have been designed to analyze calculation output files (e.g., a conformational search is repeated on the lowest energy conformer if it has not been found at least three times during the conformational search), calculate Boltzmann factors, edit calculation input files, generate transition state structures, and so forth. Thus, there are scripts designed to do energy minimizations and Monte Carlo[27] searches in MacroModel and through calling Maestro scripts. There are also scripts to do AM1 calculations in MOPAC and GAMESS and to run ab initio calculations using Jaguar.

This way, when considering thermodynamic processes, the set of 3D structures ready for the calculations are first optimized by doing an energy minimization, and then, their energy is computed. Different methods are used, from molecular mechanics to quantum chemistry. Molecular mechanics is usually chosen first to compute the energy of the molecules, since it is usually a less time-consuming method. Semiempirical or ab initio calculations are performed when more accurate results are needed for a better discrimination between the molecules.

For kinetic reactions, we have so far only used a molecular mechanics force field: the force field for Diels−Alder transition states by Brown and co-workers.[28] Thus, ROBIA is able to predict kinetic and thermodynamic products of Diels−Alder reactions (see next section, Example 2).

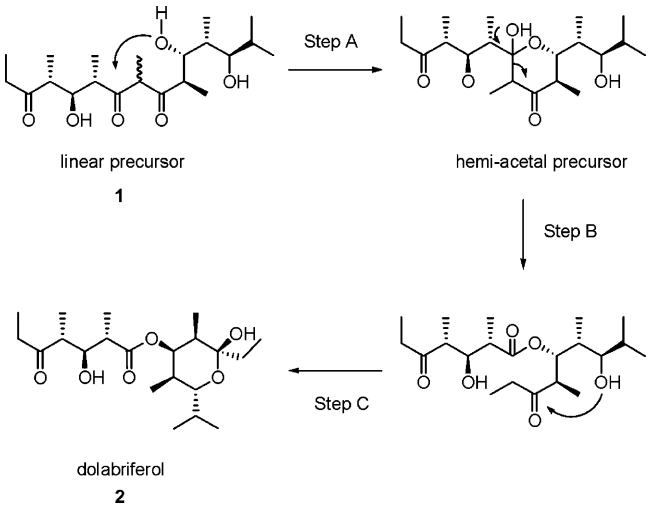
3. RESULTS AND DISCUSSION

Various examples are shown here showing the capabilities of ROBIA.

Example 1. Investigations into the Possible Reaction Pathways to Dolabriferol. Dolabriferol[29] is a complex natural product of uncertain biosynthetic origin. A pathway from the linear precursor (Scheme 2) has been suggested. However, an enormous number of competing reactions are possible if this linear precursor **1** is allowed to form acetals, do aldol reactions, and so on. To analyze the suggestion, it is necessary to consider all of the possible schemes and decide if the one leading to dolabriferol (**10**) is more likely than all the rest. This would be an extremely time-consuming task to do by hand, and mistakes would inevitably occur. ROBIA is able to automate the process.

We applied the program to investigate this possible pathway under thermodynamic conditions.[30] In this example, we compared the reaction scheme generated by the program when the linear precursor **1** (under aqueous acidic conditions) is used as the input molecule to the proposed rearrangement of dolabriferol (Scheme 2). ROBIA applied the hemiacetal/acetal formation reaction for steps A and C and the retro-aldol reaction for Step B. According to its algorithm (see Figure 2), the program first analyzes the input structure **1**,

Scheme 2. Proposed Rearrangement to Form Dolabriferol



recognizing the ketone and alcohol groups present in the molecule. On the basis of this information and the reaction conditions (an aqueous acid), the hemiacetal/acetal formation reaction is chosen from the *chemical transformations block*. This way, the program generates all possible hemiacetals and acetals by finding the substructures required for such reactions and changing them according to the coded transformations (see Table 1). After filtering three- and four-membered ring structures, all the stereoisomers are then generated from each structure and converted into 3D structures, on which energy minimizations followed by conformational searches using MM2 are done. High-energy structures are then filtered on the basis of the MM2 results. Scheme 3 shows this whole cycle for step A (only including the hemiacetals formed for simplicity), which is then repeated for steps B and C, and some of the predicted products. Finally, ROBIA generated 234 structures from which a mixture of dolabriferol **2** and its stereoisomer **3** was selected as the final output on the basis of single-point ab initio calculation results (see ref 28 for more details). The program was able to generate this matrix of possible pathways and to evaluate them all, concluding that it was a feasible pathway, as it was preferred to all of the competing routes.

Example 2. Diels−Alder Reactions. The ROBIA program can be applied to kinetically controlled processes as well as to thermodynamically controlled ones, provided that a force field for the transition state of the reaction pathway is available. This is true for Diels−Alder reactions, and so we have applied ROBIA to two examples of complex intramolecular Diels−Alder processes for which experimental data is available to test whether the ROBIA results match the experimental observations. A Diels−Alder reaction will usually generate up to four new chiral centers, for each of which there are two possible regioisomers. These five choices might lead to 32 possible products. However, the Diels−Alder mechanism restricts the number of combinations of stereocenters that can form, and so only eight products are possible. These correspond to three rather than five choices: exo/endo attack, si/re attack, and two regioisomers. The exhaustive generation of all possible stereoisomers used for dolabriferol cannot be used, as it will give some structures which cannot form through a Diels−Alder process. The Diels−Alder analysis, therefore, takes a modified path. First, four conformational searches are performed on the starting materials to generate conformations corresponding to the transition state geometries for endo/exo and si/re attack. The four possibilities are separated using constraints to enforce the right stereochemistry, and also an s-cis conformation for the diene. From each of these conformations, the two regioisomeric transition states are generated. This process generates the eight possible transition states. If the reactants have more than one possible diene/dienophile pairing, the process is repeated for all the possibilities. Conformation searches are run on all the Diels−Alder transition states, using Houk's MM2-based force field.[28] The kinetic prediction for the outcome of the reaction is generated by comparing the energies of all the conformations of all the competing transition states, to estimate the relative rates of production of the different stereoisomers and regioisomers. The thermodynamic analysis is now generated from the transition states, by mutating the transition states into the corresponding products, and running conformational analyses on each of these. The process ensures that only the subset of stereoisomers which correspond to the Diels−Alder transition states are generated. The energies of the structures are compared to produce a prediction for the selectivity on the basis of the assumption of thermodynamic control, as for the dolabriferol example.

Table 3. Energy Value Results of MM2* Calculation for the Transition States and the Products Generated by ROBIA for Examples 2a and 2b

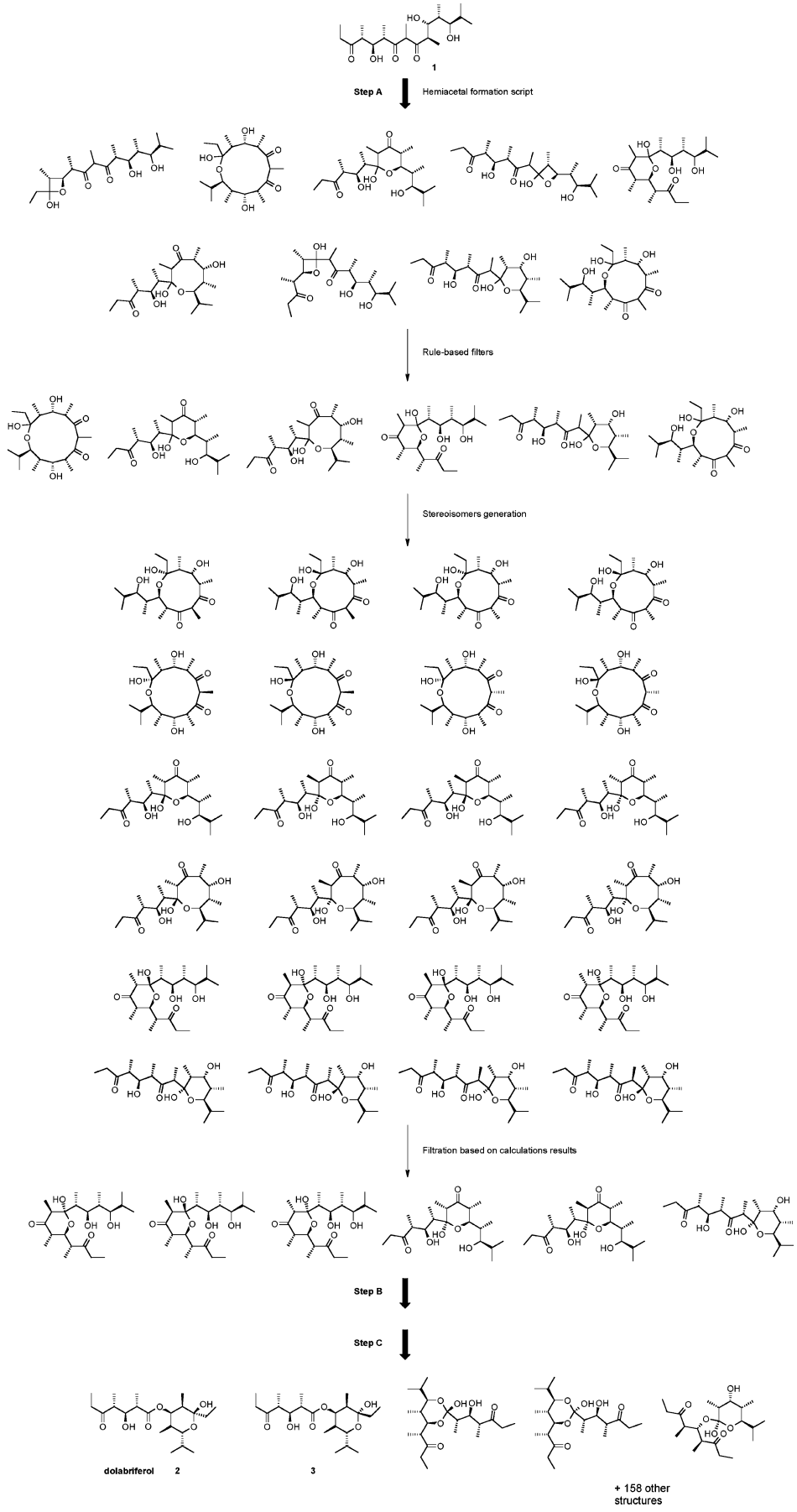
structure	*E* for the transition state (kJ/mol)	*E* for the products (kJ/mol)
a2	−121.1	132.6
a3	−111.3	154.4
a4	−113.7	170.3
a5	−107.5	168.5
a6	80.6	373.3
a7	82.4	333.5
a8	116.6	358.1
a9	70.2	372.5
b2	−85.2	109.8
b3	−71.6	95.3
b4	−61.3	95.6
b5	−68.8	115.8
b6	6.5	308.7
b7	−6.2	286.8
b8	−3.06	291.3
b9	−0.1	304.3

Example 2a. This procedure was applied to the key intramolecular Diels−Alder reaction in Takahashi et al.'s[31] synthesis of the steroid skeleton. Eight possible products may form from the reaction of **a1** (Scheme 4), and it is not obvious which of them would be preferred. Only product **a2** was detected.

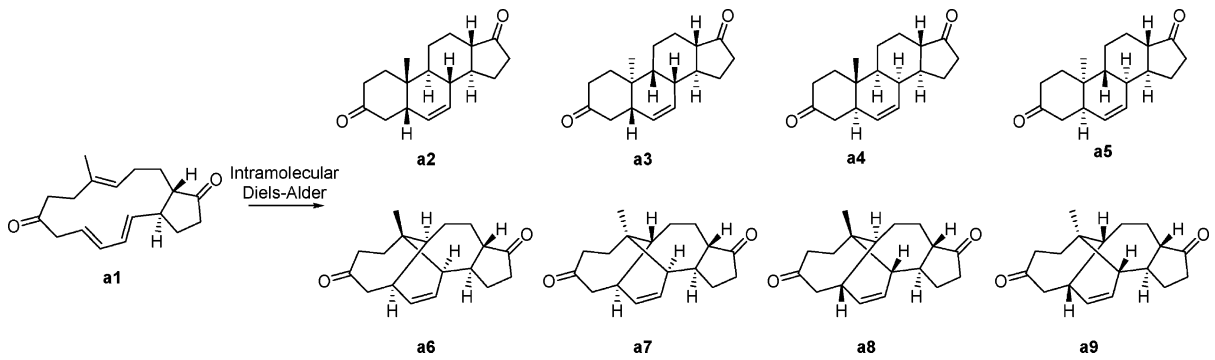
ROBIA was applied to this problem, and it calculated the relative energies of the transition states and the products of the reaction (Table 3). These show that **a2** should be formed preferentially, both under kinetic control and also under thermodynamic control. ROBIA automatically ensures that all of the possibilities are investigated, and not just the ones which look most promising. The computational result is clear, for thermodynamic and kinetic control, and both correspond to the experimental result.

Example 2b. A transannular Diels−Alder reaction was also done on structure **b1** (Scheme 5), and the experimental ratios were reported[32] by Takahashi and co-workers.

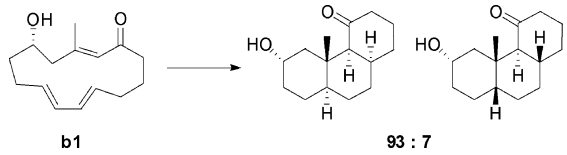
Scheme 3. Some Structures Generated and Filtered in Step A and Some Predicted Products by ROBIA



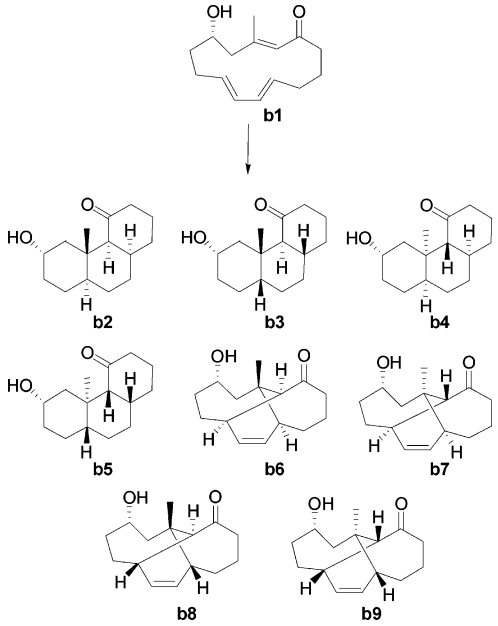
Scheme 4. Products Generated and Evaluated by the Program for Example 2a



Scheme 5. Products Obtained Experimentally for Example 2b



Scheme 6. Eight Products Formed and Analyzed by ROBIA for Example 2b



The program evaluated eight products (Scheme 6), which required the analysis of 909 structures, and predicted **b2** as the major kinetic product (Table 3). The product ratio was calculated to be 99.5:0.5 for **b2**/**b3,** being in close agreement with the experimental ratios, found to be 93:7 for **b2**/**b3**. Additionally, the program predicted **b3** and **b4** as the thermodynamic products in a ratio 52.5:47.5, respectively. Only the former is in agreement with the experiment, so the calculation suggests the process is under kinetic control.

Takahashi et al. reported a force field analysis[32] using a similar force field looking only at transition states and obtained comparable results. ROBIA performs a more complex analysis automatically, replacing an intricate manual analysis with a more complex automated process. This means it is now feasible to tackle much more challenging problems than is feasible by hand.

4. CONCLUSION

ROBIA is a computational tool for reaction prediction which uses coded rules and molecular modeling to achieve its predictions. It uses a mechanistic approach to predict reactions, generating possible intermediate structures from the reactants to the products. It is able to evaluate thermodynamic processes and, in some cases, kinetic reactions. The program has been successfully tested against literature reports. It could be able to anticipate results, making possible the prediction and analysis of unprecedented reactions as well as the investigation of biosynthetic and metabolic pathways. The time required for the examples shown was a few seconds for structure predictions and 1−2 h per structure for Example 1 and 1−2 min per structure for Example 2, for the molecular modeling calculations on a desktop PC (Pentium 4 CPU, 3.2 GHz, 1 GB RAM, Linux), which can easily be distributed onto a cluster. Further uses of the program could include the checking of reaction databases, as a teaching tool in undergraduate courses and as an assistant in synthesis design to test reactions in the forward direction and check for possible side reactions.

ACKNOWLEDGMENT

We thank Elsevier MDL and Unilever for financial support and Dr. Keith T. Taylor for helpful discussions.

REFERENCES AND NOTES

(1) Todd, M. H. Computer-aided organic synthesis. *Chem. Soc. Rev.* **2005**, *34*, 247−266.

(2) Barone, R.; Chanon, M. In *Encyclopedia of Computational Chemistry*; Schleyer, P. V. R., Ed.; Wiley: Chichester, U. K., 1998; Vol. 4, pp 2931−2948.

(3) Corey, E. J. General methods for the construction of complex molecules. *Pure Appl. Chem.* **1967**, *14*, 19−37.

(4) Johnson, A. P.; Marshal, C.; Judson, P. N. Starting Material Oriented Retrosynthetic Analysis in the LHASA Program. 1. General Description. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 411−417.

(5) Johnson, A. P.; Marshal, C. Starting Material Oriented Retrosynthetic Analysis in the LHASA Program. 2. Mapping the SM and Target Structures. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 418−425.

(6) Johnson, A. P.; Marshal, C. Starting Material Oriented Retrosynthetic Analysis in the LHASA Program. 3. Heuristic Estimation of Synthetic Proximity. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 426−429.

(7) Gasteiger, J.; Pförtner, M.; Sitzmann, M.; Höllering, R.; Sacher, O.; Kostka, T.; Karg, N. Computer-assisted synthesis and reaction planning in combinatorial chemistry. *Persp. Drug Discovery Des.* **2000,** *20*, 245−264.

(8) Satoh, K.; Funatsu, K. A Novel Approach to Retrosynthetic Analysis Using Knowledge Bases Derived from Reaction Databases. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 316−325.

(9) Jorgensen, W. L.; Laird, E. R. Cameo: a program for the logical prediction of the products of organic reactions. *Pure Appl. Chem.* **1990**, *62*, 1921−1932.

(10) Höllering, R.; Gasteiger, J.; Steinhauer, L.; Schulz, K.; Herwig, A. Simulation of Organic Reactions: From the Degradation of Chemicals to Combinatorial Synthesis. *Chem. Inf. Comput. Sci.* **2000**, *40*, 482−494.

(11) Satoh, H.; Funatsu, K. Further Development of a Reaction Generator in the SOPHIA System for Organic Reaction Prediction. Knowledge-Guided Addition of Suitable Atoms and/or Atomic Groups to Product Skeleton. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 173−184.
(12) Sello, G. Reaction Prediction: The Suggestions of the Beppe Program. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 713−717.
(13) Greene, N.; Judson, P. N.; Langowski, J. J.; Marchant, C. A. Knowledge-based Expert Systems for Toxicity and Metabolism Prediction: DEREK, StAR, and METEOR. SAR & QSAR. *Environ. Res.* **1999**, *10*, 299−313.
(14) Klopman, G.; Dimayuga, M.; Talafous, J. META. 1. A Program for the Evaluation of Metabolic Transformation of Chemicals. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1320−1325.
(15) Borodina, Y.; Sadym, A.; Filimonov, D.; Blinova, V.; Dmitriev, A.; Poroikov, V. Predicting Biotransformation Potential from Molecular Structure. *J. Chem. Inf. Comput. Sci*. **2003**, *43*, 1636−1646.
(16) Java Version 1.4.1. http://java.sun.com/ (accessed Sep 2005).
(17) *Cheshire Studio*, version 3.0.0.54.; Elsevier MDL: San Leandro, CA.
(18) *Maestro*, version 5.0.019; Schrodinger Inc.: Portland, Oregon, 2000.
(19) Mohamadi, F.; Richards, N. G. J.; Guida, W. C.; Liskamp, R.; Lipton, M.; Caufield, C.; Chang, G.; Hendrickson, T.; Still, W. C. MacroModel − an integrated software system for modeling organic and bioorganic molecules using molecular mechanics. *J. Comput. Chem*. **1990**, *11*, 440−467.
(20) *Jaguar*, version 4.2; Schrodinger, Inc.: Portland, Oregon, 2000.
(21) *GAMESS*, version 20; Iowa State University: Ames, IA. Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. I.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. General Atomic and Molecular Electronic-Structure System. *J. Comput. Chem.* **1993**, *14*, 1347−1363.
(22) MOPAC: Stewart, J. J. P. MOPAC − A Semiempirical Molecular-Orbital Program. *J. Comput.-Aided Mol. Des.* **1990**, *4*, 1−45.
(23) MDL CTfile Formats. http://www.mdl.com/solutions/white_papers/ctfile_formats.jsp (accessed Sep 2005).
(24) Chen, L.; Nourse, J. G.; Christie, B. D.; Leland, B. A.; Grier, D. L. Over 20 Years of Reaction Access Systems from MDL: A Novel Reaction Substructure Search Algorithm. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1296−1310.
(25) Sadowski, J.; Gasteiger, J. From Atoms and Bonds to Three-dimensional Atomic Coordinates: Automatic Model Builders. *Chem. Rev.* **1993**, *93*, 2567−2581.
(26) Pearlman, R. S. *CONCORD*; Tripos Inc.: St. Louis, MO.
(27) Chang, G.; Guida, W. C.; Still, W. C. An Internal Coordinate Monte-Carlo Method for Searching Conformational Space. *J. Am. Chem. Soc.* **1989**, *111*, 4379−4386.
(28) Raimondi, L.; Brown, F. K.; Gonzalez, J.; Houk, K. N. Empirical Force-Field Models for the Transition States of Intramolecular Diels−Alder Reactions Based upon ab Initio Transition Structures. *J. Am. Chem. Soc.* **1992**, *114*, 4796−4804.
(29) Ciavatta, M. L.; Gavagnin, M.; Puliti, R.; Cimino, G. Dolabriferol: A New Polypropionate from the Skin of the Anaspidean Mollusc Dolabrifera dolabrifera. *Tetrahedron* **1996**, *52*, 12831−12838.
(30) Socorro, I. M.; Goodman, J. M.; Taylor, K. ROBIA: A Reaction Prediction Program. *Org. Lett.* **2005**, *7*, 3541−3544.
(31) Takahashi, T.; Shimizu, K.; Doi, T.; Tsuji, J. Macroring Contraction Methodology. 4. A Novel Route to Steroid A, B, C Rings by the Transannular Diels-Alder Reaction of the 14-Membered (*E*,*E*,*E*)-Macrocyclic Triene. *J. Am. Chem. Soc.* **1988**, *110*, 2674−2676.
(32) Takahashi, T.; Sakamoto, Y.; Doi, T. Macroring Contraction Methodology. 6. Transannular Diels−Alder Reaction of the 14-Membered (*E*,*E*,*E*)-Trienone. *Tetrahedron Lett.* **1992**, *33*, 3519−3522.

CI050379E