Simulation of Organic Reactions: From the Degradation of Chemicals to Combinatorial Synthesis

Robert Höllering, Johann Gasteiger,* Larissa Steinhauer, Klaus-Peter Schulz, and Achim Herwig

Computer-Chemie-Centrum, Institut für Organische Chemie, Universität Erlangen-Nürnberg, Nägelsbachstrasse 25, D-91052 Erlangen, Germany

Received June 26, 1999

Organic reactions can be run under a variety of conditions, from laboratory experiments, through technical processes, to combinatorial chemistry. The scope is further extended when the metabolism of compounds and the reactions in the mass spectrometer are included. We present here several concepts: reactors, phases, and modes, which, together with a kinetic modeling, allow the treatment of such a broad scope of organic reactions. These concepts have been implemented in a knowledge-based system, EROS. Several applications of this system to the wide world of organic reactions are given.

1. INTRODUCTION

Most organic bench chemists have become so used to performing all their reactions in a three-necked flask with a stirrer that they take it for granted that this is the way a reaction has to be run. The advent of combinatorial chemistry and parallel synthesis, however, has made it quite clear that there are also ways of running chemical reactions other than those performed in a stirred flask. Clearly, industrial chemists working in large scale production plants always had a more diversified approach to running their processes, working with different types of reactors in batch or continuously. The conversion of a batch laboratory reaction into a continuous industrial process involves a large amount of tedious work by chemists and chemical engineers emphasizing the different paradigms for running a chemical reaction.

Biochemical pathways, such as the metabolism of nutrients, demonstrate quite different ways for performing reactions having rather low concentrations of the substrates and strong feedback between the individual steps of the entire metabolic system. Again different from this is the way chemicals undergo degradation in the environment. Furthermore, the processes in the mass spectrometer are chemical reactions being performed under reaction conditions quite different from those previously mentioned.

As these different ways of running reactions are performed and investigated in different disciplines of chemistry, the need for a unified treatment of chemical reactions has not been urgently felt. However, we believe that the introduction of combinatorial chemistry into the daily life of many organic chemists shows the need to understand and model such a wide variety of performing reactions.

We will introduce here a general scheme that allows the modeling of the different ways of running a reaction and then show how this scheme has been implemented in a computer program system. In this endeavor we could build on our long-standing experience in modeling chemical reactions through the development of the EROS (*e*laboration of *r*eactions for *o*rganic *s*ynthesis) system.[1−4] We will then show how this new version of the EROS system incorporating the ideas introduced here can be applied to a wide variety of running reactions and give examples for the application of the EROS system.

It has to be emphasized from the very beginning that the knowledge base of EROS is presently not yet developed far enough to provide an in-depth coverage of the entire range of organic chemistry. Rather, we have developed a framework that allows chemists or information specialists to add rules to EROS for the reaction types they are specifically interested in. We will show with several examples in the applications section (section 4) how such rules can be developed and how EROS can be taught more chemistry.

2. REACTORS, PHASES, MODES, AND KINETICS

There is an extensive repertoire for running a reaction, and therefore, many factors have to be considered in the simulation of reactions. Reactions can be performed in a homogeneous or heterogeneous environment; transfer of matter can occur between phases when several phases are involved in a reaction sequence.

Reactions can be run as a batch or continuous process. In the course of a reaction, compounds or solvent can be added or removed. Multistep processes can be performed as one-pot reactions or by individually running each reaction step, isolating the products, and then proceeding to the next step.

The course of a reaction is strongly influenced by the concentration of the reaction partners and by reaction times. The modeling of a reaction has therefore to consider the kinetic order of a reaction and the reaction rates, too.

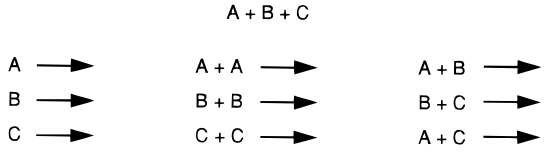
To account for this wide variety of performing reactions, several concepts have been identified and were introduced into the design of the new version of EROS. We understand that the use of these terms in the context of EROS may sometimes be different from their common use. However, we define these terms quite clearly and thus construct a framework for their use in modeling chemical reactions.

2.1. Reactors. A reactor defines the environment for processes that occur at *the same time*. This definition based

* Corresponding author. E-mail: Gasteiger@CCC.Chemie.Uni-Erlangen.DE.

10.1021/ci990433p CCC: $19.00
Published on Web 01/28/2000

Scheme 1. Combinations of Starting Materials A, B, and C That Will Be Explored in Reaction Generation When the Mode *mix* Is Chosen



on time was found more appropriate for modeling reactions than any one based on the physical nature of a reaction vessel. Thus, a reaction being performed in one and the same vessel but consisting, first, of running the reaction and, then, of hydrolyzing the intermediates to isolate the products is treated as two reactors, the first one for performing the reaction and the second reactor for the hydrolysis.

2.2. Phases. A reactor can consist of one or several phases. A phase is a *place* comprising compounds at certain concentrations and being physically separated from other such places, but simultaneously having communication (e.g., exchange of compounds) with other phases. Thus, a phase is physically separated from another phase but exists at the same time and therefore can interact with another phase.

2.3. Mode. The mode of reaction generation in a certain phase is dependent on the *concentration* of the partners. Specification of the mode allows the specification of the type of concentration of the starting materials and, in a general manner, the kinetic order of a reaction. If the concentration of the starting materials is high enough, each compound can react with any other compound and also with another molecule of the same compound (mode = *mix*). Thus, in a mixture of compounds A + B + C, the reactions shown in Scheme 1 can be anticipated.

Quite often, interest is only in reactions between molecules of different structure. Then the (mode = *mix_no_A+A*) has to be chosen. In a mixture of A + B + C the reactions A + A, B + B, and C + C are then not considered any more. On further dilution, the starting materials of a reaction will usually not react with another starting material but with materials that are present in a rather high concentration such as the solvent (e.g., in a hydrolysis). Thus, only monomolecular and pseudo-monomolecular reactions occur (mode = *monomolecular*).

When a reactor consists of two or more phases, there are situations where reactions might occur only between compounds from different phases (mode = *interface*).

2.4. Kinetics. EROS allows the specification of models for the calculation of reaction rates for individual reactions or for reaction types. Thus, the decrease of the starting materials and the appearance of products can be followed over time by integration of the differential equations for all reactions in a reactor. Different methods such as the Gear algorithm,[5,6] the Runge−Kutta[7,8] method, or the Runge−Kutta−Merson[7,8] method can be used. In cases where no reaction rates can be estimated but only probabilities can be given for a certain reaction step to occur (such as in the simulation of mass spectra), the network of parallel and sequential reaction steps can be evaluated by probability theory.

3. IMPLEMENTATION OF THE EROS SYSTEM

A detailed discussion of the implementation of the EROS system for reaction simulation is not intended in this publication but will be given in a more technical paper. Only a brief outline of the EROS system will therefore be presented here. A manual of the EROS system can be found on the Internet.[9]

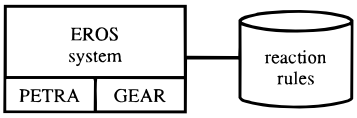


Figure 1. Basic architecture of the EROS system with the subsystems PETRA for the calculation of the physicochemical properties and the Gear algorithm for the integration of differential equations.

The EROS system has access to a variety of methods for the calculation of physicochemical effects that are collected in the PETRA package (*p*arameter *e*stimation for the *t*reatment of *r*eactivity *a*pplications). These methods are so fast that by default all structures generated in an EROS run are processed by these methods, providing a host of physicochemical decriptors that can be used in the evaluation and modeling of chemical reactions and reactivity (see section 3.3).[10]

A major feature of the implementation of the EROS system is that the knowledge base on chemical reactions is kept separately from the system (Figure 1). The knowledge on chemical reactions, i.e., a description of the types of reactions that EROS can be applied to, and mechanisms for the evaluation of these reaction types are kept as reaction rules in external files. An appendix gives an overview of the structure and contents of a knowledge base of EROS.

Keeping the knowledge on reactions in an external file allows easy extension of the scope of EROS to novel reaction types without having to change the code of the EROS system. The EROS system is based on a novel representation of chemical structures[11] that also permits the handling of radical cations and thus the use of EROS for mass spectra simulation. When evaluations of reaction rates are contained in the reaction rules, the integration of differential equations on kinetics is necessary. This is performed by the Gear algorithm or Runge−Kutta methods.

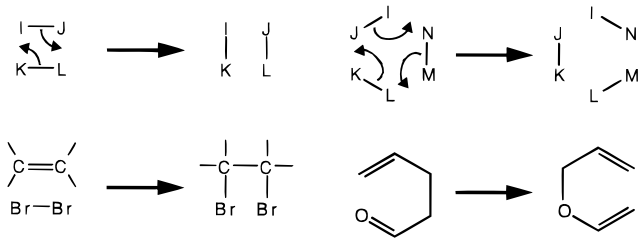
EROS is coded in C++, the subsystems PETRA and GEAR in FORTRAN and C. The knowledge base on chemical reactions is written in the interpreted language *Tcl*.

3.1. Specification of Phases, Reactors, and Modes. At the header of a set of reaction rules in the knowledge base, information has to be given on how reactions are run: the number of reactors, the number of phases within the reactors, and the mode within a phase and between phases. Furthermore, the initial concentration for each starting material has to be specified when a kinetic modeling of the reaction network is desired.

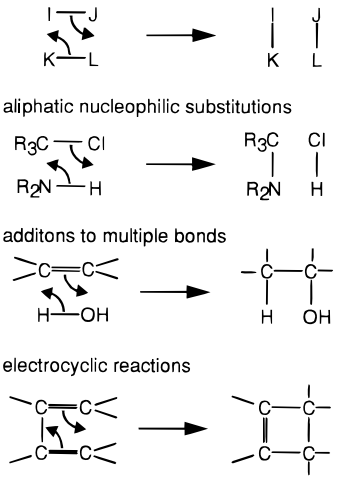
3.2. Formal Reaction Schemes. From the outset of the development of the EROS system[1] we have always relied on a formal representation of chemical reactions, treating reactions as bond and electron shifting schemes in much the same way as chemists draw bond shifting arrows. Scheme 2 shows two such general reaction schemes which might comprise both concerted and multistep reactions.

Use of such formal reaction types provides a general framework for the treatment of the wide variety of organic reactions. Thus, the first reaction scheme in Scheme 2

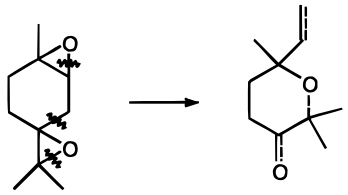
Scheme 2. Two General Reaction Schemes with a Specific Example for Each Scheme



Scheme 3. Reactions Where Two Bonds Are Broken and Two Bonds Are Made



Scheme 4. Example for the Second Reaction Scheme of Scheme 2 with the Three Bonds That Are Broken and the Three Bonds That Are Made Indicated by Wavy or Broken Lines, Respectively



comprises about 50% of all organic reactions, some of them shown in Scheme 3.

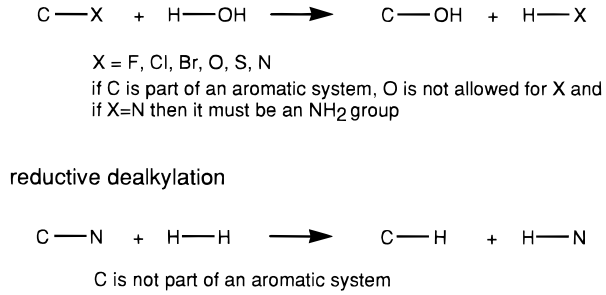
Thus, application of this reaction scheme of breaking two bonds and making two bonds onto the appropriate bonds of the starting materials allows the generation of all conceivable reactions, both known and also novel ones.

To stress the point that the use of such formal reaction schemes can encompass basically all reactions, even some more unusual ones, Scheme 4 shows an instance of the second reaction scheme of Scheme 2. This reaction[12] has been studied with EROS quite some time ago.[2]

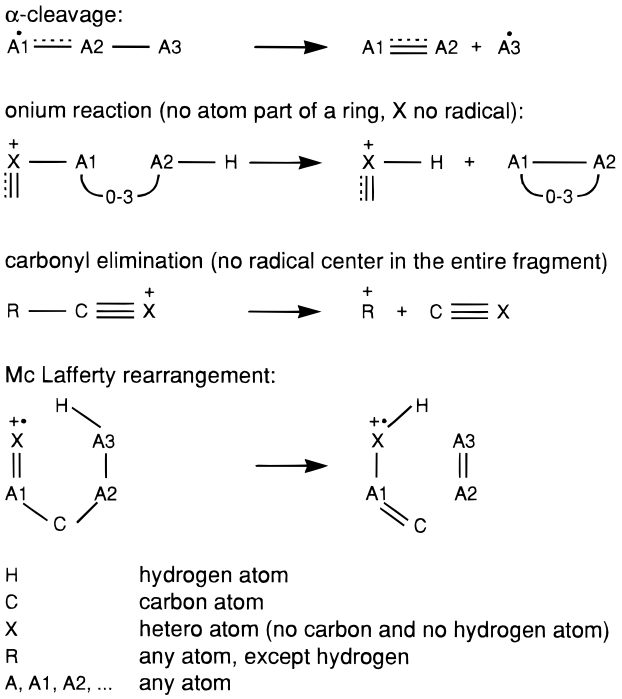
Specific reaction types are obtained by restricting the nature of the atoms I, J, K, L, M, and N in Scheme 2 and the bonds between them to specific instances, such as the addition of bromine to olefins or the oxy−Cope rearrangement as shown in this scheme.

These restrictions are exactly the kind of information contained in the reaction rules. As an example, the first reaction scheme of Scheme 2 can be narrowed down to the two reaction types, hydrolysis and reductive dealkylation, contained in Scheme 5. In the section on examples it will be shown how these two reaction types can model the anaerobic degradation of *s*-triazine herbicides in soil.

Scheme 5. Reaction Types for the Degradation of *s*-Triazines (Text Shows the Constraints Imposed onto the Atoms)



Scheme 6. Reaction Types Used for the Simulation of Mass Spectra



Another set of reaction types is shown in Scheme 6, reactions that are important for the modeling of fragmentations and rearrangements of organic compounds in the mass spectrometer which are used in the examples section for the simulation of mass spectra.[13,14]

Work is under way to automatically extract reaction types from the information contained in reaction databases.[15]

3.3. Hierarchy in Reaction Evaluation. Clearly, the application of formal reaction schemes always runs the danger of producing much too many reactions, many reactions that might formally be conceivable but are chemically not feasible. In this situation, the evaluation of reactions becomes of great importance in order to sift out the chemically reasonable reactions from among the formally conceivable ones. EROS allows the specification of a wide range of evaluation and selection mechanisms in the rule file in order to find the reactions that make chemical sense, or to focus on the reaction that one is interested in. These evaluation procedures form a hierarchy from no selection at all to the calculation of absolute rate constants.

In designing and developing evaluation procedures for reaction types to be stored in the reaction rules, the chemist can build on the physicochemical effects that are automatically calculated by the methods contained in the PETRA package for the molecules generated in an EROS run.

Table 1. Physicochemical Descriptors Calculated by the Methods in PETRA

physicochemical descriptor		ref
atomic properties		
charges	q_σ, q_π, q_{tot}	16, 17
electronegativities	χ_σ, χ_π, χ_{LP}	18
polarizabilities	α	19
bond properties		
charge differences	Δq	16, 17
electronegativity differences	$\Delta\chi$	18
polarizabilities	α_b	19
stabilization by delocalization	D^+, D^-	17
bond dissociation energies	BDE	20
molecular properties		
heats of formation	ΔH_f°	20
mean molecular polarizabilities	$\bar{\alpha}$	19

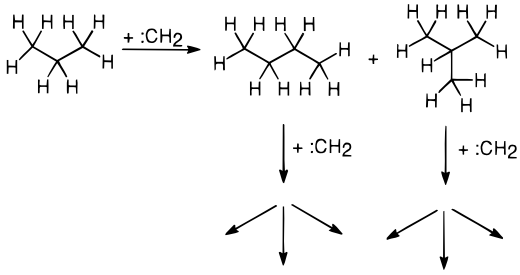


Figure 2. Generation of all isomeric alkanes.

These methods are all empirical in nature having been designed to be rather rapid in order to allow the calculation of many molecules in a short amount of time. Most of them have been published, including those for the calculation of charge distribution,[16,17] inductive,[18] resonance,[18] and polarizability effect,[19] as well as for heats of formation.[20]

Table 1 gives an overview of the more important physicochemical effects calculated by the methods in PETRA.

3.3.1. No Evaluation. When no evaluation is performed, EROS exhaustively applies a given reaction type contained in the rule file. This may be of advantage when, e.g., all isomers should be generated. As an example, the generation of all alkane isomers by the formal carbene insertion into C−H bonds is presented in Figure 2.

Clearly, for small molecules this can be done with pencil and paper. However, when it comes to the generation of all isomeric tetradecanes, even the most accurate and tireless chemist matches his/her limitations. Due to the increase in the number of isomers with growing carbon number, even a computer has to put up with a doubly exponential increase in calculation time. Data files holding the structures of the isomers (as SD-files) mentioned in Table 2 can be found on the Internet.[21] The numbers of isomers obtained in the runs match those calculated by Henze et al.[22]

3.3.2. Constraints on the Atoms and Bonds of the Reaction Site. In section 3.2, we discussed how constraints on the types of atoms and bonds can be imposed in order to focus a reaction scheme onto a certain reaction type. These constraints can be taken one step further in terms of chemical significance by specifying limits or ranges of certain physicochemical properties of the atoms and bonds at the reaction site as obtained by the methods in PETRA.[10]

We will show the use of constraints on physicochemical effects with a simple reaction, the hydrolysis of methyl halides (Scheme 7). The rate of this aliphatic nucleophilic displacement depends on the ease of attack of the nucleophile and on the ability of the halide ion to leave. The attack of the nucleophile should be facilitated by the more positively charged carbon atom of the methyl halides.

Table 2. Generation of Alkane Isomers

generated alkane, C_nH_{2n+2}	no. of isomers	computation time[a] (s)
1	1	
2	1	1.02
3	1	1.06
4	2	1.10
5	3	1.36
6	5	1.85
7	9	2.97
8	18	5.92
9	35	13.71
10	75	34.32
11	159	92.64
12	355	286.95
13	802	1495.71
14	1858	17 936.42

[a] Process user time (as determined by the POSIX "time" command) on a personal computer with an AMD K6-II 300 MHz processor running Linux.

Scheme 7. Constraints on the Partial Charge, q, and the Polarizability, α, of the Atoms of Methyl Halide for Subjecting Them to Hydrolysis

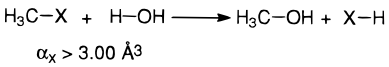


Table 3. Physicochemical Properties of CH_3-X

	X				
	F	Cl	Br	I	ref
q_C (e)	0.080	0.038	0.021	−0.012	16
q_X (e)	−0.253	−0.176	−0.145	−0.089	16
α_X (Å^3)	2.03	3.99	5.14	7.43	19

On the other hand, the propensity of the halide ion to leave increases with increasing polarizability. Table 3 shows the obtained values. Methyl bromide and methyl iodide have half-life times of 20 and 110 days, respectively, at pH 7 and 298 K; methyl chloride has a half-life of about 1 year, whereas the half-life of methyl fluoride is about 30 years.[23]

Thus, in modeling the hydrolysis of methyl halides one might want to allow hydrolysis of the chloro, bromo, and iodo compounds, whereas, because of the slow rate, hydrolysis of the fluoro compound should be suppressed.

This can be achieved by defining a rule requiring, first, the polarity of the C−X bond to be larger than 0.06 e and, furthermore, the polarizability of the leaving group, X, to have a value of at least 3.0 Å^3. These restrictions would allow the generation of hydrolysis of alkyl chlorides, bromides, and iodides, but would prevent the generation of hydrolysis of alkyl fluorides.

3.3.3. Calculation of Heats of Reaction. EROS also has access to an empirical method for the calculation of heats of formation,[20] and, using these values, can calculate heats of reaction. On this basis, reactions that are more exothermic can be given higher priority or endothermic reactions may be suppressed. In other words, EROS can be used to simulate thermodynamic product control.

Figure 3 shows the two conceivable pathways of an intramolecular aldol condensation. As the formation of the enolate is an equilibrium reaction, the reaction product is determined by the viability of the subsequent ring closures.

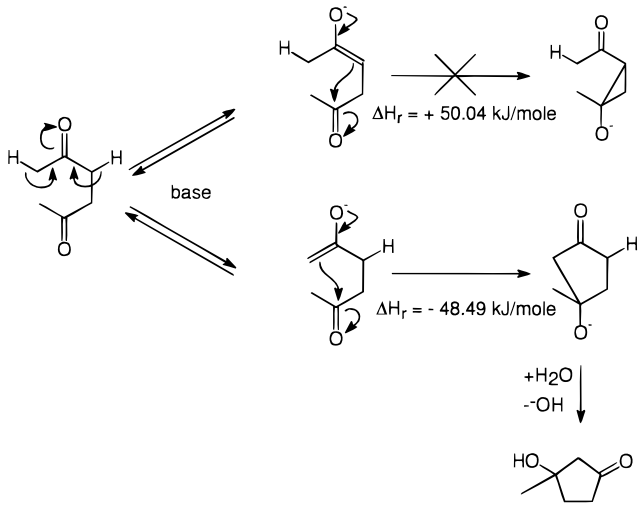


Figure 3. Thermodynamic control of the cyclization of 2,5-hexadione.

On the basis of the heats of formation of the enolates and of the aldol anion to be formed, a reaction enthalpy can be calculated. While the ΔH_f values of anions are highly dependent on the solvent used, their differences are not as much subject to solvent effects and can be used as such. On the basis of this calculation, only the compound with the five-membered ring is generated by EROS.

3.3.4. Calculation of Reaction Rates. The knowledge on chemical reactions for EROS is derived from experimental data and information. Data from kinetic measurements such as reaction rates or activation energies of a series of instances of a certain reaction type can be analyzed with statistical methods or neural networks using physicochemical descriptors of the atoms and/or bonds at the reaction site.

Thus, it has been shown[4] that data on amide hydrolysis, under both acid and base catalysis, can be correlated with quantitative measures on charge distribution and the resonance effects as calculated by the PETRA suite of programs.[10] The equations thus derived were then used to predict the various reaction pathways in the hydrolysis of different benzoylphenylureas, including several agrochemicals.[4]

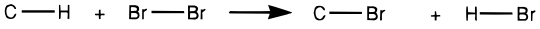
In several of the examples of reactions to be modeled by EROS in the following section, we will attempt to derive methods for the calculation of reaction rates. However, it is quite clear that for many reaction types not enough information is available to allow the establishment of a quantitative model for calculating reaction rates. Then, less demanding evaluation mechanisms, as specified in the previous sections, can be built into the reaction rules, still providing quite efficient methods for the selection of the chemically significant reactions.

4. DIFFERENT TYPES OF APPLICATIONS

In order to further an understanding of the concepts developed here and how they allow a broad range of reactions to be modeled with EROS, specific examples will be given.

4.1. Bromination of Phenol. This reaction is run in a single vessel; the concentration of the starting materials is such that multiple reactions between the different starting materials might occur but not among starting materials of the same sort. The following specifications are made: one reactor, one phase, mode *mix_no_A+A*.

Scheme 8. Reaction Type for the Bromination of Phenol



Next, a reaction rule for the bromination of phenol had to be developed and stored in the knowledge base of the EROS system. The reaction center, i.e., the bonds broken and made in the reaction, was specified as shown in Scheme 8.

The following restrictions were imposed onto the carbon atom: it has to be part of an aromatic system, and an oxygen or nitrogen atom has to be in conjugation to this carbon with a distance of two or four bonds, shifting its lone pair partly into the aromatic system $q_{\pi(O,N)} > 0$. Deactivating substituents are not considered. These constraints are valid for carbocyclic aromatic systems. To also use the reaction rule for heteroaromatic systems the constraints would have to be adapted.

In order to make quantitative predictions, mechanisms for the estimation of the relative rates of bromination at the various positions of phenol have to be given. The following observations were used: Bromination of phenol gives 80% *p*-bromophenol and 20% *o*-bromophenol,[24] allowing the conclusion that bromination in the para position is 8 times faster than bromination in an ortho position (there are two ortho positions!). As the rate of bromination is very much dependent on the solvent, we have chosen an average value for bromination in an ortho position, setting it to 0.01 L/s·mol. This gives for the para position a value of 0.08 L/s·mol. In order to account for the influence of a bromine substituent on the rate of further bromination, recourse was made to the following observation: The rate of nitration of bromobenzene is 3% the rate of nitration of benzene.[25] It was therefore assumed that with each bromine substituent also the rate of bromination drops to 3% of the rate without this additional bromine substituent.

With these rate constants the integration of the differential equations in the kinetic modeling was performed by the Gear algorithm. Figure 4 shows the sequence of reaction products obtained in this reaction modeling. Figure 5 reproduces the results of the kinetic modeling of this system of reactions.

4.2. Degradation of *s*-Triazine Herbicides in the Environment. The major degradation reactions of *s*-triazine herbicides such as atrazine or prometon under anaerobic conditions in soil are reductive dealkylation and hydrolysis.[26] The concentration of the herbicide is generally rather low so that no reaction between triazine molecules will occur. Thus, no reactions of this chemical with other molecules of its kind have to be considered but only reactions with chemicals having high concentration in the environment such as water or reduction equivalents. The general specifications for modeling these reactions were therefore one reactor, with one phase having the mode *monomolecular* (which, in this case, corresponds to a pseudo-monomolecular process).

The two reaction types shown in Scheme 5 were included in the knowledge base. The sequence of degradation reactions obtained for prometon with these two reaction types, and two additional ones for the hydrolysis of cyanuric acid, and for the decarboxylation of a carbaminic acid is shown in Figure 6.

In order to be able to estimate reaction rates for the two reaction types contained in Scheme 5, recourse was made to half-life times of *s*-triazine herbicides reported in the literature. Figure 7 shows half-life times for four such compounds.[27]

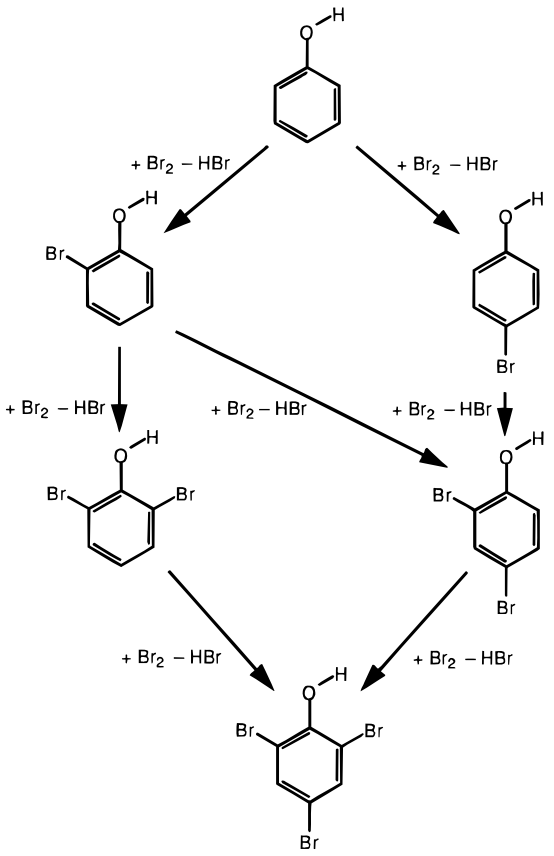


Figure 4. Sequence of reaction products generated for the bromination of phenol.

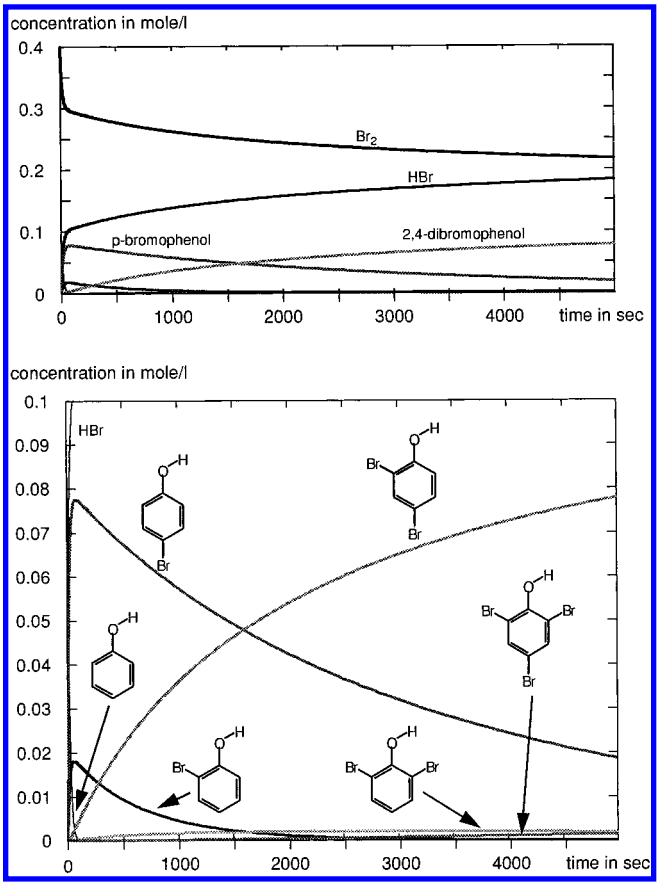


Figure 5. Concentration profile for development of the products in the bromination of phenol (cf. Figure 4). The top figure gives an overview. The bottom shows details.

To derive reaction rates, the following rationale was pursued. First, it is observed that reductive dealkylation is always the major reaction pathway, dominating hydrolysis. As an approximation, it was assumed that the substituents at the amino groups do not influence the rate of hydrolysis of the carbon−chlorine bond. In reality, this will quite certainly not be the case, but in the absence of product ratios in the degradation of these compounds no better analysis can be made. The rate of hydrolysis in all these four compounds was set to 5×10^{-9} s^{-1}. With these assumptions the four half-life times allow the assignment of reaction rates for the reductive cleavage of an ethyl group (3.32×10^{-8} s^{-1}), an isopropyl group (2.65×10^{-8} s^{-1}), and a *tert*-butyl group (2.21×10^{-8} s^{-1}). As an additional assumption, all further incidences of hydrolysis were given a reaction rate that was set to 5×10^{-9} s^{-1}.

With this information, together with an initial concentration of prometon of 0.1 mmol/L, EROS is able to generate individual reaction steps, determine the kinetic order, and compute reaction rates. This information is automatically transferred to the Gear algorithm that sets up a system of differential equations and solves them. This provides the concentration dependence of all compounds on time.

Figure 8 shows these results produced for the sequence of steps shown in Figure 6.

More details will be given in a separate publication.[28]

4.3. Technical Processes. The concept of reactors, phases, and modes allows the modeling of a wide variety of technical processes. Table 4 gives a list of process setups that can be modeled by EROS.

Table 4. Technical Processes That Can Be Modeled by EROS

types of technical processes
batch mode
stirred tank reactor
continuous mode
continuous ideal tank reactor
cascade of tank reactors
laminar tube reactor
ideal tube reactor

4.4. Pharmacokinetics. Phases can also be used for modeling compartments in pharmacokinetics. First, we consider the case of intravenous administering of a drug A into blood serum at regular intervals of 8 h followed by degradation of the drug through a sequence of two metabolites M_1 and M_2. The modeling of this reaction needs only one phase. The following specifications were made: initial concentration of drug A of 0.01 mmol/L in 3 L of blood serum; half-life times of 3.5 h for the conversion of A to M_1 and of 10 h for the conversion of M_1 to M_2. The concentration of the three species A, M_1, and M_2 over time is obtained in the simulation with EROS as shown in Figure 9.

A more complicated—and more realistic—situation is modeled by the system of compartments shown in Figure 10.

The administered drug A_X is contained in a solution (e.g., in the intestinal tract) and is absorbed into the blood serum. The blood transports drug A to the site of action (e.g., to the receptor). Simultaneously, the drug can be stored in the

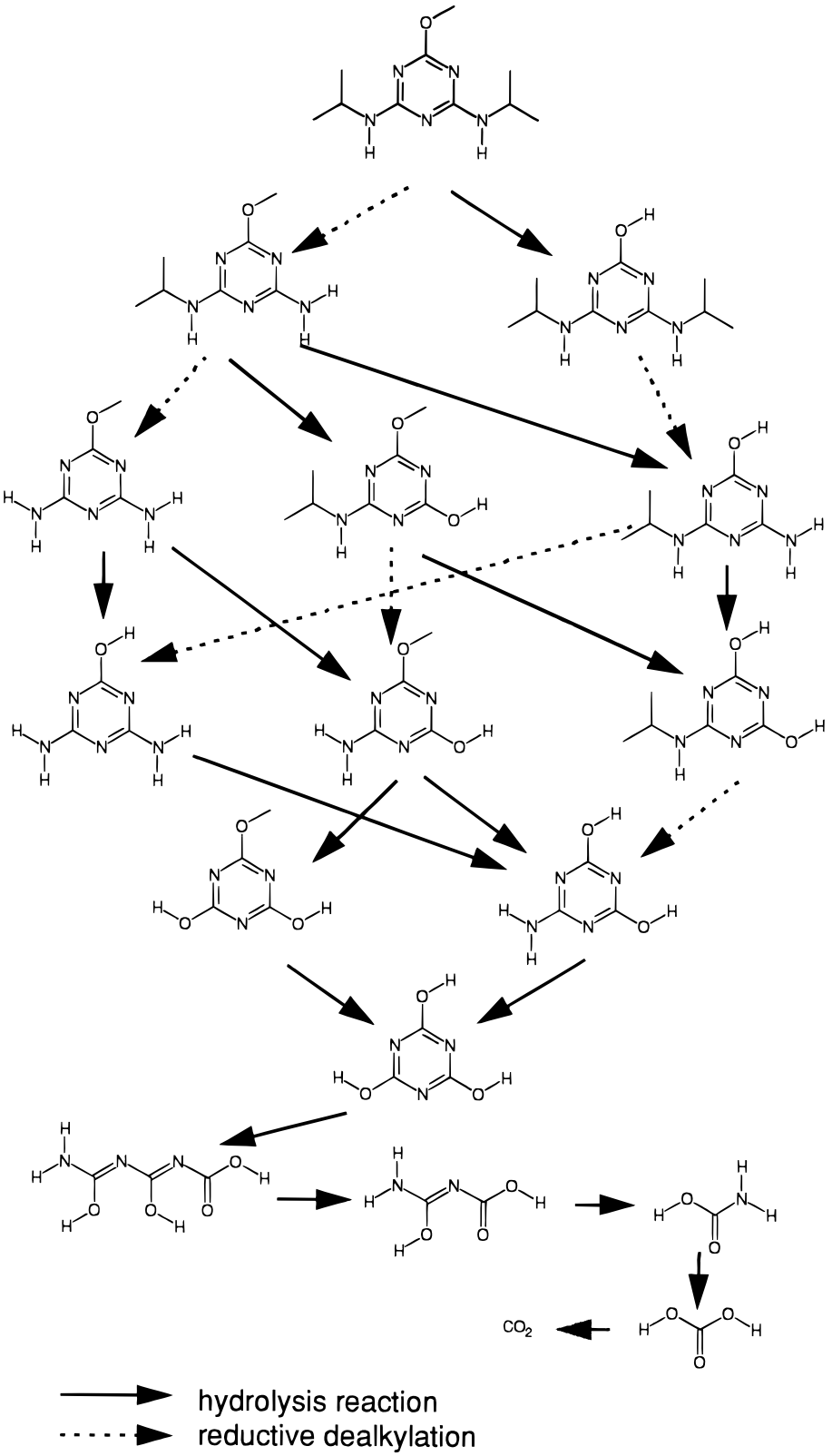


Figure 6. Simulation of the degradation reactions of prometon, obtained with the reaction types shown in Scheme 5, hydrolysis and reductive dealkylation.

tissue from where it can be released. The drug A is broken down to several metabolites, M_1 and M_2, which are excreted. Figure 10 models these different sites as phases and also indicates the exchange of materials between these sites through arrows. Note that all these compartments exist and exchange materials at the same time; these five phases are therefore all part of a single reactor. The concentrations of the drug and its metabolites in the various phases then have a profile as shown in Figure 11.

4.5. Acetal Formation Monitored by Mass Spectroscopy. The concept of a reactor and its phases is quite general: it can represent many different types of vessels where chemical reactions occur. In this sense, a mass spectrometer is considered as one reactor. In the example

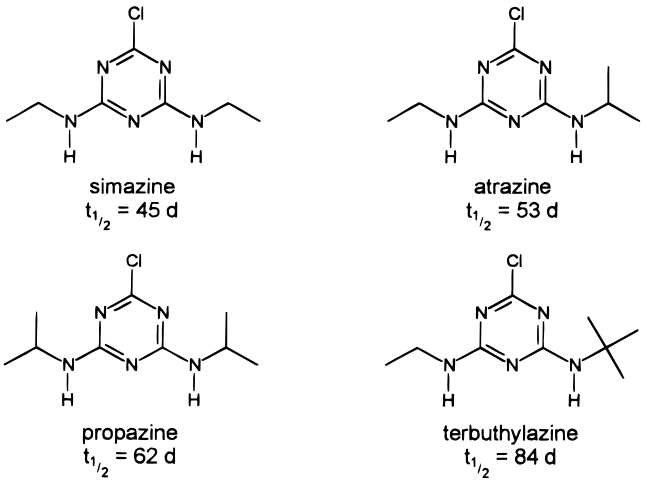


Figure 7. Half-lifes of four *s*-triazines in soil.

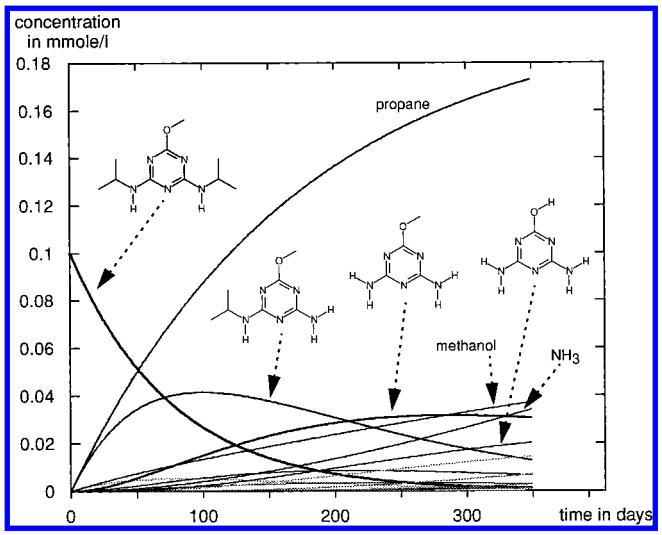


Figure 8. Concentration profile of the degradation reactions of prometon (cf. Figure 6).

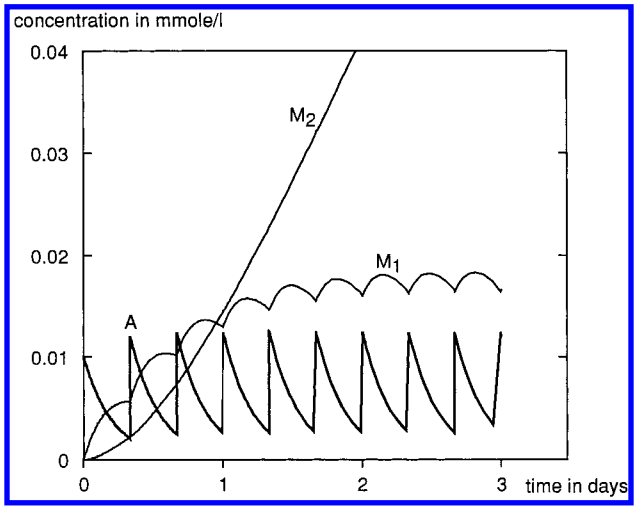


Figure 9. Concentration dependence of a drug A and its two metabolites M_1 and M_2 on time by regular administering.

here, the combination of two reactors is presented: the first one is for performing a laboratory synthesis and the second reactor is a mass spectrometer. Figure 12 schematically shows the setup for modeling the reaction of two starting materials giving products followed by GC/MS. The vessel where the reaction is run is represented by reactor 1, the

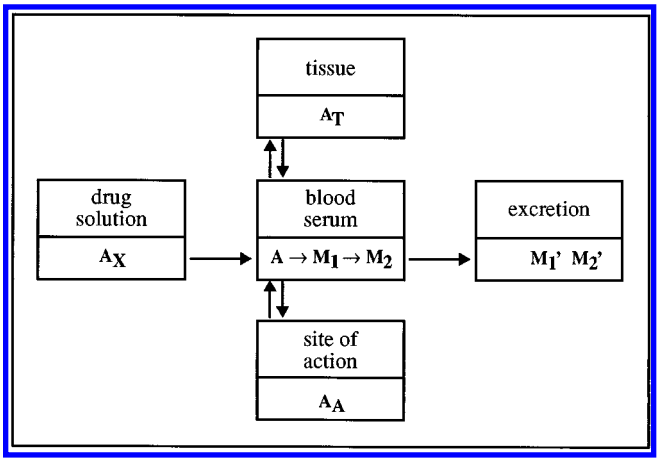


Figure 10. Compartments that are considered in the modeling of the pharmacokinetics of drug A.

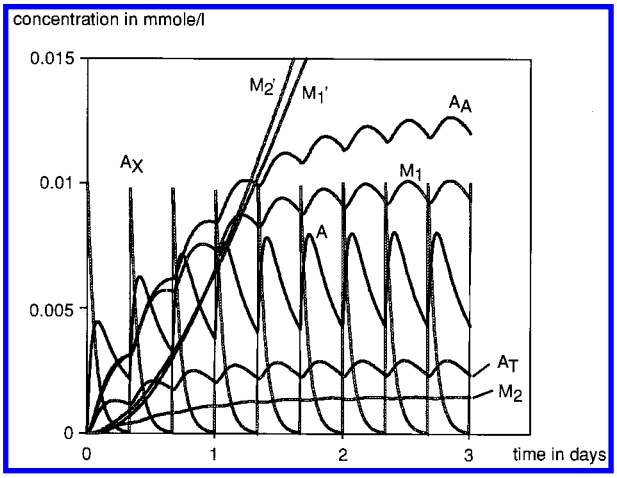


Figure 11. Schematic concentration dependence of materials in the phases of Figure 10.

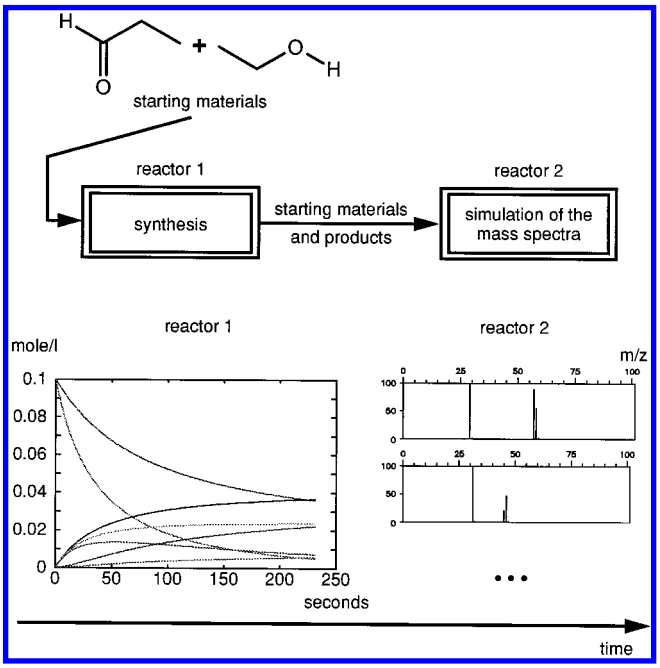


Figure 12. Formal representation of the modeling of a chemical reaction followed by GC/MS to identify the products.

mass spectrometer which ionizes and breaks down the two starting materials, and the products of the reactions is reactor 2.

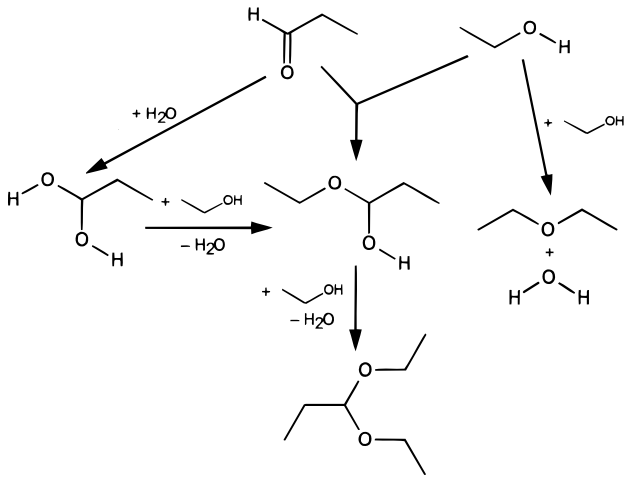
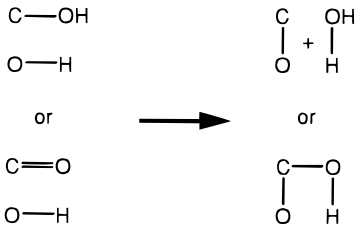


Figure 13. Various reaction steps obtained by applying the reaction scheme of Scheme 9 to propanal and ethanol and to their products. Note that these reactions occur under acid catalysis.

Scheme 9. Reaction Type for Acetal Formation



As a specific example, propanal is allowed to react with ethanol in a stirred flask. Thus, each compound may react with any other compound in the reaction mixture (mode = *mix*). The progress of the reaction in this reactor is monitored by mass spectroscopy. The reaction scheme shown in Scheme 9 was added to the knowledge base for reactor 1.

The CO bond may be a bond to a hydroxy group (single bond) or a carbonyl bond (double bond), but only one oxygen atom is allowed in the neighborhood of the carbon atom. Thus, in effect this CO bond is not allowed to be part of an acid, an ester, a semiacetal, or an acetal. The rate constant for reaction of a C−OH bond was set to 0.05 L/mol·s and that for reaction of a C═O bond to 0.1 L/mol·s. The initial concentrations of propanal and of ethanol were both set to 0.1 mol/L. Figure 13 shows the network of reactions obtained in this EROS run.

The concentration of the various products at a given time can be obtained by integration of the differential equations for the processes in reactor 1 as shown in Figure 12.

Let us now assume that the final reaction mixture is injected into a GC−MS leading to a separation of all compounds. The mass spectra were simulated in reactor 2 by application of the reaction types contained in Scheme 6 to the various starting materials and products obtained in reactor 1 for modeling acetal formation.

Clearly, the reaction types of Scheme 6 are incomplete for formulating the diverse processes occurring in the mass spectrometer. In particular, rearrangements through transfer of hydrogen atoms are conspicuously absent in this scheme. However, these reaction types do comprise the major degradation reactions in these types of compounds (and for many organic molecules in general).

The next task is then an evaluation of the reaction probabilities for these different reaction types in order to estimate the relative importance of the individual fragmentation and rearrangement steps. This is necessary for estimating peak intensities in the mass spectrum. In previous work we have shown how evaluations for the fragmentation and rearrangement processes in the mass spectrometer can be derived from an analysis of experimental mass spectra.[13,14,29]

This analysis had still to be based on a connection table representation for radical cations. It was shown that a connection table has inherent deficiencies when applied to the representation of radical cations, a fact that led us to the development which forms the basis of the new version of the EROS system.[11] However, the evaluation function used for the reaction types of Scheme 6 has not yet been adjusted to this novel structure representation.

With these caveats in mind we now simulate the mass spectra of the species obtained in modeling acetal formation from propanal and ethanol and compare them in Figure 14 with the corresponding experimental mass spectra as contained in the MassLib database.[30,31] This database did not contain mass spectra for the hydrate and the semiacetal of propanal, presumably because of the thermal instability of these compounds.

In view of the reservations on the state of the art of mass spectra simulation mentioned above, the simulated mass spectra are surprisingly good. Clearly, some smaller peaks contained in the experimental mass spectrum are not reproduced. These are mostly the result of inductive cleavages, hydrogen rearrangements followed by cleavages, and hydrogen and methane eliminations. However, the major peaks, and to a large extent their relative intensities, are simulated quite well. This gives promise that the models inherent in our approach to mass spectra simulation can form a sound basis for further work.

The combination of two reactors, one for running a reaction and the second for simulating mass spectra may be particularly valuable for the modeling and analysis of the reactions performed in combinatorial chemistry experiments.

4.6. Parallel Synthesis in Combinatorial Chemistry. In combinatorial chemistry, different sets of starting materials are combined in all conceivable variations to synthesize a wide range of compounds. This can be achieved in a variety of experimental setups such as in parallel synthesis or liquid-phase experiments.[32]

The concept of phases allows a clear-cut handling of combinatorial chemistry synthesis. The starting materials of one sort are kept in one phase, and a second set of starting materials is assigned to a second phase. In order to have no reactions going on within a phase, a phase has to have the mode of *inert* or *interface*. On the other hand, one does want to have reactions between compounds from the first phase with compounds of the second phase. This can be accomplished by assigning the mode of *inert* to phase 1 and *interface* to phase 2 and by defining a contact of phase 2 to phase 1. This has the result that each compound from the first phase is reacted with each starting material of the second phase one after another. Thus, all combinations of products between the two sets of starting materials are generated and are put into the third phase as shown for the synthesis of esters from acid chlorides and alcohols in Figure 15.

To illustrate the application of EROS and, in particular, the concept of phases to combinatorial chemistry, an experi-

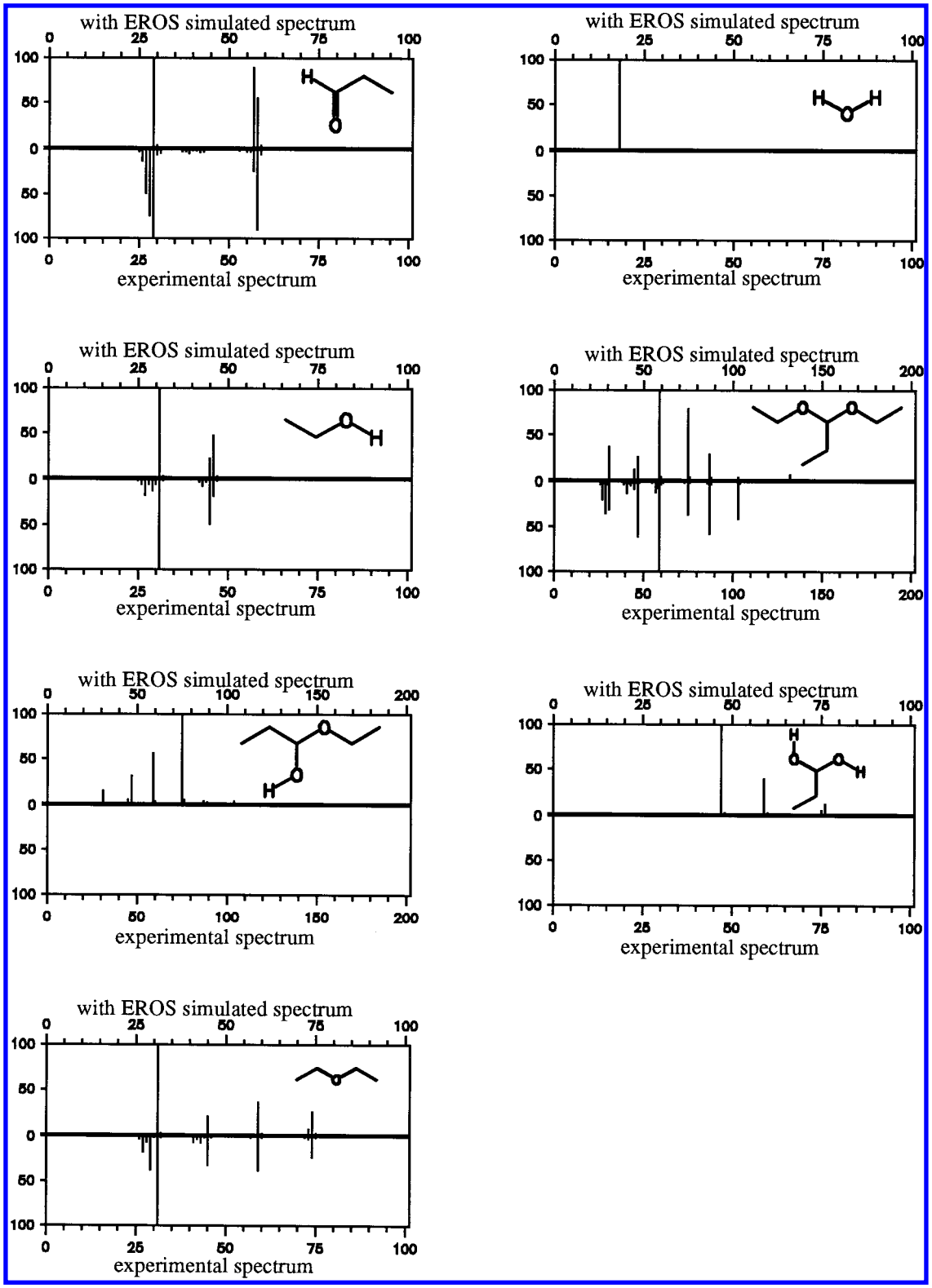


Figure 14. Simulated and experimental mass spectra of the starting materials, intermediates, and products in the acetal formation from propanal and ethanol. (The experimental mass spectra of the hydrate and the semiacetal were not available, presumably because of the instability of these compounds; the mass spectrum of H_2O is of no concern.)

ment by Ellman et al.[33] for the synthesis of 1,4-benzodiazepines (see Figure 16) will be analyzed.

Three sets of starting materials are necessary to perform this sequence: 2-aminobenzophenones, amino acids, and alkylating agents. For modeling this experiment with EROS, the two 2-aminobenzophenones, three amino acids, and three alkylating agents shown in Figure 17 were chosen.

The entire reaction sequence can be combined into two essential reaction steps, one for the synthesis of 1,4-benzodiazepines and one for alkylation at the nitrogen atom N-1. For modeling this reaction sequence we used five phases (see Figure 18): three phases for storing the three different sets of starting materials, one phase for storing the intermediate 1,4-benzodiazepine (not yet alkylated), and one phase

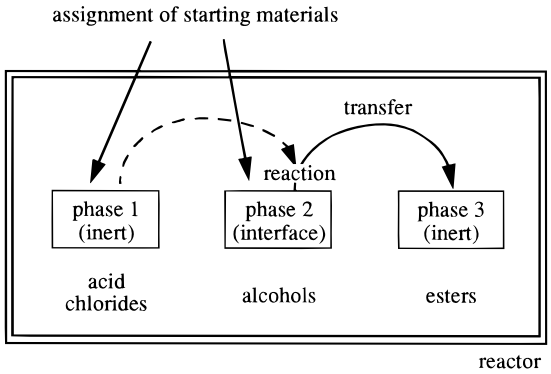


Figure 15. Reaction setup for the modeling of a parallel synthesis in combinatorial chemistry. The reactions are done at the interface of phase 2 with one compound of phase 2 and one compound of phase 1. After the reaction is completed, all the products are transferred to phase 3.

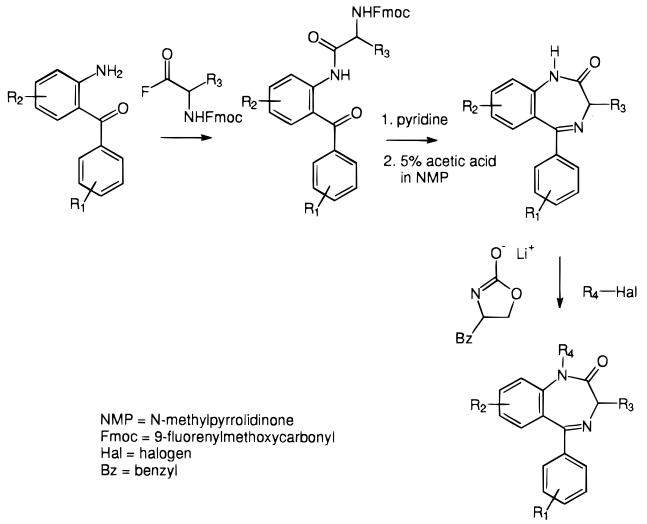


Figure 16. Reaction scheme developed by Ellman et al.[33] for the synthesis of 1,4-benzodiazepines.

for the final product. Two of these phases are performing the two major reaction steps: the combination of the 2-aminobenzophenone with an amino acid followed by cyclization to 1,4-benzodiazepines, and the alkylation step at nitrogen-1 of this ring system.

Figure 19 shows the structures obtained in this combinatorial chemistry experiment. All 18 ($2 \times 3 \times 3$) conceivable structures were obtained in this run.

5. CONCLUSIONS

The concepts introduced here, reactors, phases, modes, and kinetic modeling, are capable of modeling a wide variety of chemical reactions and processes, organic laboratory systems, technical processes, the degradation of chemicals in the environment, metabolic pathways, and pharmacokinetics, mass spectra, and combinatorial chemistry experiments. This has been demonstrated by the implementation of these concepts in the new version of the EROS system.

Clearly, the results obtained are heavily dependent on the amount and depth of knowledge on chemical reactions stored in the reaction rules. This is where most of the future work has to be devoted. The architecture of the system that keeps this knowledge base separate from the system itself provides a firm basis for this endeavor.

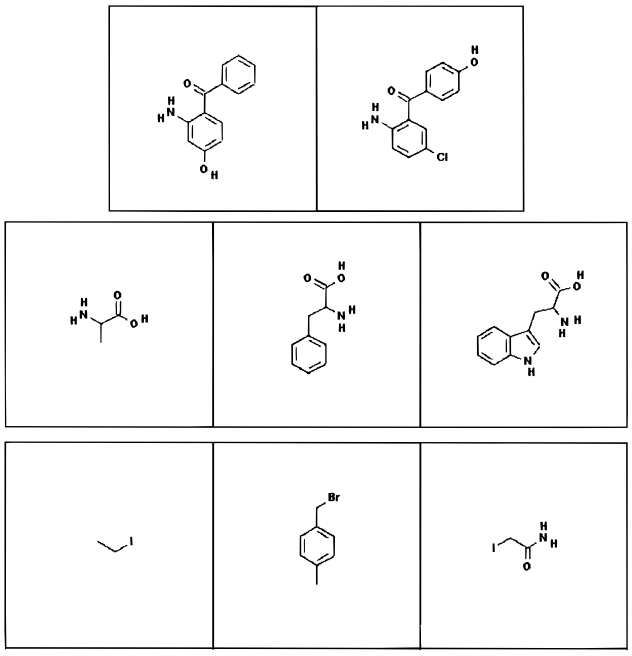


Figure 17. The three sets of compounds used for modeling the synthesis of 1,4-benzodiazepines.

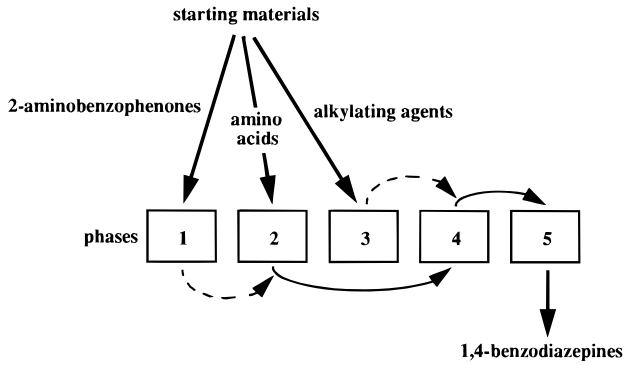


Figure 18. Assembly of five phases to model the synthesis of 1,4-benzodiazepines.

ACKNOWLEDGMENT

We thank the German Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie (bmb+f) and the Deutsche Bundesstiftung Umwelt for their financial support. Additionally, we thank colleagues of our group for providing us with their work: S. Bauerschmidt for the novel chemical data structure,[11] on which EROS is built; T. Kostka for the development of the rules for the degradation of *s*-triazines in soil[28] and the graphical user interface to EROS; W.-D. Ihlenfeldt for the CACTVS system[34] that includes a molecule editor and several tools to display the results; N. Karg for testing the EROS system.

6. APPENDIX: STRUCTURE OF THE KNOWLEDGE BASE

The knowledge base of EROS is organized in so-called "rule files", each comprising the experimental setup of reactors, phases, modes, and rules for the individual reaction types. These reaction rules are further structured into a part describing the rule, another defining the constraints on the educts, and a further section containing the instructions for the bond reorganization in a reaction. The rule file is written

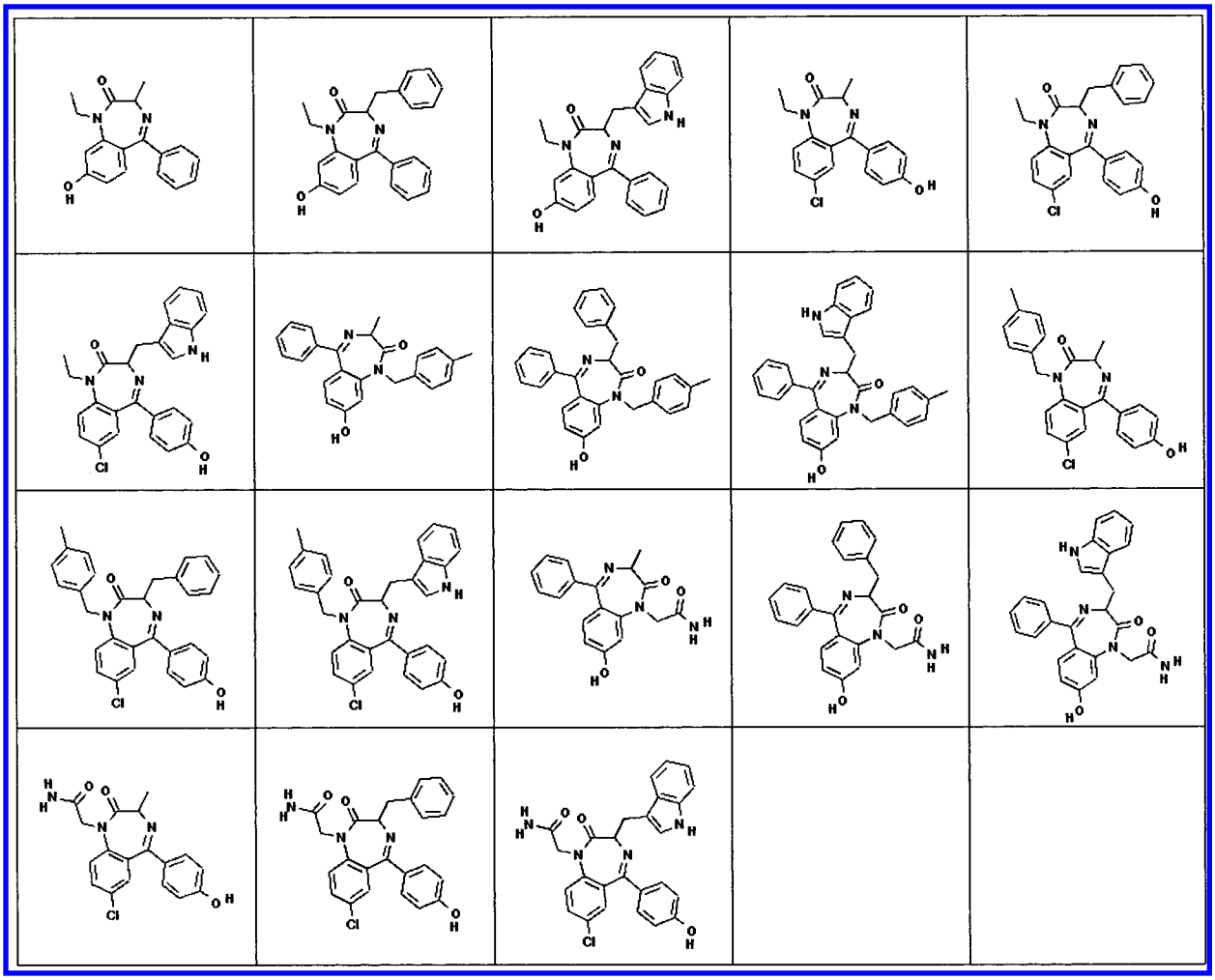


Figure 19. Products obtained in modeling the reaction of the three sets of starting materials shown in Figure 17 by the setup of five phases of Figure 18.

in the interpreted language Tcl, which has been extended by some EROS-specific commands.

6.1. The Rule File Header. In the Rule File Header the setup of the reaction to be pursued is defined. This includes the number of reactors, the number of phases, their modes, their connection, and the information of which phase is considered as output phase of the entire reaction sequence. If a modeling of the kinetics is intended, the integration method for the evaluation of the differential equations has to be chosen. Preprocessing of educts, e.g., putting special markers on atoms or molecules, or the definition of starting concentrations can also be accomplished here.

6.2. The Individual Reaction Rules. Each rule describing a single reaction type, e.g., the insertion of carbene into a C−H bond (section 3.3.1), is further divided into three parts (see sections 6.2.1−6.2.3). Special rules for directing the educts into different phases or simple transport rules can be given.

6.2.1. Reaction Rule Information. Three variables have to be set in this section. The first one is the reaction name, which will be written into the output file; the second one is a variable that contains special attributes for the evaluation of kinetics, e.g., enforcing reactions of zero-order or Michaelis−Menten kinetics. The third and most important information is the connectivity of atoms in the reaction substructure. The reaction substructure also provides the basis for the constraints on atoms and bonds (section 6.2.2).

6.2.2. Constraints on Educts. For each atom of the reaction substructure required, properties can be defined. This starts from simple properties like the atomic number but can include the whole set of PETRA[10] parameters. Additionally, molecule or bond properties are available in their full range. As the entire knowledge base is written in a programming language, even complex requirements can be formulated here.

6.2.3. The Transformation. The third part of an individual rule defines the rearrangement of the electrons in the course of a reaction, i.e., the bonds to be broken and to be made.

Furthermore, mechanisms for the evaluation of a reaction can be specified such as the calculation of the heats of formation of educts and products (see section 3.3.3) or the calculation of reaction rates (section 3.3.4).

REFERENCES AND NOTES

(1) Gasteiger, J.; Jochum, C. EROS—A Computer Program for Generating Sequences of Reactions. *Topics Curr. Chem.* **1978**, *74*, 93−126.
(2) Gasteiger, J.; Hutchings, M. G.; Christoph, B.; Gann, L.; Hiller, C.; Löw, P.; Marsili, M.; Saller, H.; Yuki, K. A New Treatment of Chemical Reactivity: Development of EROS, an Expert System for Reaction Prediction and Synthesis Design. *Topics Curr. Chem.* **1987**, *137*, 19−73.
(3) Röse, P.; Gasteiger, J. Automated Derivation of Reaction Rules for the EROS 6.0 System for Reaction Prediction. *Anal. Chim. Acta* **1990**, *235*, 163−168.
(4) Gasteiger, J.; Hondelmann, U.; Röse, P.; Witzenbichler, W. Computer Assisted Prediction of the Degradation of Chemicals: Hydrolysis of Amides and Benzoylphenylureas. *J. Chem. Soc., Perkin Trans. 2* **1995**, 193−204.

(5) Gear, C. W. *Numerical Initial Value Problems in Ordinary Differential Equations*; Prentice Hall: Englewood Cliffs, NJ, 1971.
(6) Gear Algorithm. QCPE Program No. QCMP022.
(7) Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; Vettering, W. T. *Numerical Recipes, The Art of Scientific Computing*; Cambridge University Press: New York, 1989.
(8) Köckler, N. *Numerische Algorithmen in Softwaresystemen*; Teubner: Stuttgart, 1990.
(9) Gasteiger, J.; Bauerschmidt, S.; Höllering, R.; Ihlenfeldt, W.-D.; Karg, N.; Kostka, T.; Sacher, O.; Schulz, K. P.; Steinhauer, L. Elaboration of Reactions for Organic Synthesis. EROS—A Program for Reaction Prediction. http://www2.ccc.uni-erlangen.de/software/eros/index.html.
(10) Gasteiger, J.; Kleinöder, T. Parameter Estimation for the Treatment of Reactivity Applications—The PETRA package. http://www2.ccc.uni-erlangen.de/software/petra/index.html.
(11) Bauerschmidt, S.; Gasteiger, J. Overcoming the Limitations of a Connection Table Description: A Universal Representation of Chemical Species. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 705−714.
(12) Ho, T. L.; Stark, C. J. Novel Rearrangement of 1,2:4,8-Diepoxy-*p*-menthane. *Liebigs Ann. Chem.* **1983**, *8*, 1446−1447.
(13) Gasteiger, J.; Hanebeck, W.; Schulz, K.-P. Prediction of Mass Spectra from Structural Information. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 264−271.
(14) Gasteiger, J.; Hanebeck, W.; Schulz, K.-P.; Bauerschmidt, S.; Höllering, R. Automatic Analysis and Simulation of Mass Spectra. In *Computer-Enhanced Analytical Spectroscopy 4*; Wilkins, C. L., Ed.; Plenum Press: New York, 1993; pp 97−133.
(15) Chen, L.; Gasteiger, J. Knowledge Discovery in Reaction Databases: Landscaping Organic Reactions by a Self-Organizing Neural Network. *J. Am. Chem. Soc.* **1997**, *119*, 4033−4042.
(16) Gasteiger, J.; Marsili, M. Iterative Partial Equalization of Orbital Electronegativity—A Rapid Access to Atomic Charges. *Tetrahedron* **1980**, *36*, 3219−3228.
(17) Gasteiger, J.; Saller, H. Calculation of the Charge Distribution in Conjugated Systems by a Quantification of the Resonance Concept. *Angew. Chem.* **1985**, *97*, 699−701; *Angew. Chem., Int. Ed. Engl.* **1985**, *24*, 687−689.
(18) Hutchings, M. G.; Gasteiger, J. Residual Electronegativity—An Empirical Quantification of Polar Influences and its Application to the Proton Affinity of Amines *Tetrahedron Lett.* **1983**, *24*, 2541−2544.
(19) Gasteiger, J.; Hutchings, M. G. Quantification of Effective Polarisability. Applications to Studies of X-Ray Photoelectron Spectroscopy and Alkylamine Protonation. *J. Chem. Soc., Perkin Trans. 2* **1984**, 559−564.
(20) Gasteiger, J. Automatic Estimation of Heats of Atomization and Heats of Reaction. *Tetrahedron* **1979**, *35*, 1419−1426.
(21) Herwig, A.; Gasteiger, J. Generation of Alkane Isomers using EROS. http://www2.ccc.uni-erlangen.de/research/reaction_prediction/alkane.html.
(22) Henze, H. R.; Blair, C. M. The Number of Isomeric Hydrocarbons of the Methane Series. *J. Am. Chem. Soc.* **1931**, *53*, 3077−3085.
(23) Mabey, W.; Mill, T. Critical Review of Hydrolysis of Organic Compounds in Water Under Environmental Conditions. *J. Phys. Chem. Ref. Data* **1978**, *7*, 383−415.
(24) Streitwieser, Jr., A.; Heathcock, C. H. *Introduction to Organic Chemistry*, 3rd ed.; Macmillan Publishing Company, New York, 1985; p 672.
(25) Bird, M. L.; Ingold, C. K. Influence of Directing Groups on Nuclear Reactivity in Oriented Aromatic Substitutions. Part IV. Nitration of the Halogenbenzenes. *J. Chem. Soc., Part 1* **1938**, 918−929.
(26) Kearney, P. C.; Kaufmann, D. D.; Sheets, T. J. Degradation of Atrazine and Related s-Triazines. *J. Agric. Food Chem.* **1965**, *13*, 369−372.
(27) Burkhard, N.; Guth, J. A. Chemical Hydrolysis of 2-Chloro-4,6-bis-(alkylamino)-1,3,5-triazine Herbicides and Their Breakdown in Soil under the Influence of Adsorption *Pestic. Sci.* **1981**, *12*, 45−52.
(28) Kostka, T.; Selzer, P.; Gasteiger, J. Computer Assisted Method for Structure Elucidation of s-Triazine Herbicides Combining Reaction Prediction and the Simulation of Infrared Spectra. Manuscript in preparation.
(29) Bauerschmidt, S.; Hanebeck, W.; Schulz, K.-P.; Gasteiger, J. Elucidation of Reactions in the Mass Spectrometer. *Anal. Chim. Acta* **1992**, *265*, 169−181.
(30) Henneberg, D.; Weimann, B. Spektrenbibliotheken in der massenspektrometrischen Analytik. In *Software-Entwicklung in der Chemie 2*; Gasteiger, J., Ed.; Springer-Verlag: Heidelberg, 1988; pp 153−157.
(31) MassLib from MPI für Kohlenforschung, Mülheim a. d. Ruhr, Germany.
(32) Warr, W. A. Combinatorial Chemistry. In *Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Clark, T., Gasteiger, J., Kollman, P. A., Schaefer III, H. F., Schreiner, P. R., Allinger, N. L., Eds.; Wiley: Chichester, U.K., 1998; Vol. I, pp 407−417.
(33) Bunin, B. A.; Plunkett, M. J.; Ellman, J. A. Synthesis and Evaluation of Three 1,4-Benzodiazepine Libraries. In *Combinatorial Peptide and Nonpeptide Libraries, A Handbook*; Jung, G., Ed.; VCH: Weinheim, 1996; pp 405−424.
(34) Ihlenfeldt, W. D.; Takahashi, Y.; Abe, H.; Sasaki, S. Computation and Management of Chemical Properties in CACTVS: An Extensible Networked Approach toward Modularity and Flexibility. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 109−116; http://www2.ccc.uni-erlangen.de/software/cactvs/index.html.

CI990433P