

# Prediction of Organic Reaction Products: Determining the Best Reaction Conditions

Marco Durante and Guido Sello\*

Dipartimento di Chimica Organica e Industriale, Università degli Studi, via Venezian 21, 20133 Milano, Italy

Received June 11, 1999

We describe some of our most recent achievements concerning the selection of the best set of reaction conditions for a specific reaction. In particular, we will concentrate on the selection of the best solvent to minimize side reactions. The solvent should favor a kinetically controlled reaction if it is a good solvent for the transition state and a bad solvent for the ground state, decreasing the activation energy. Consequently, we need the descriptions of the transition state and of the state solvation. We generally apply the principle of “similarity in solvation”; i.e., we calculate an approximate similarity between reactants, transition states, and solvent molecules. The more the former are similar to the latter, the more they are solvated. This permits the selection and the ordering of the solvents. We considered other aspects of reaction conditions that will be briefly commented on. Some results will be presented to illustrate the power of the method.

## INTRODUCTION

Experimental chemists are still the best inference engines when designing a synthesis route for a desired target structure; they perceive which substructures can be considered building blocks and from what starting materials they can be prepared; they reckon what reactions are most appropriate and what reaction conditions can better make the reaction proceed. However, in recent years much effort has been put into the search for a rational synthesis and reaction design, mainly to improve understanding but also to aid chemists in their choices. Some reaction prediction systems have been thus realized, and their use has proved interesting.<sup>1–5</sup> Our contribution is named Beppe;<sup>6</sup> it is a rational approach to reaction prediction, working free of a priori constraints and grounding its predictions on calculated physicochemical descriptors. Its result is a set of products and a set of reaction descriptors that, in principle, should allow an estimation of the feasibility of the diverse alternatives. This model was applied to both simple and complex compounds and showed good response.

As the next development we would like to ease the understanding of the proposed results to permit a better exploitation of the system. The final goal should be, in our hopes, a system that can effectively aid chemists in mostly favoring the desired reaction while decreasing the yield of any byproduct of side reactions. In this direction we can place the selection of the best reaction conditions.

The choice of the best reaction conditions for performing a reaction is a challenge and a wish at the same time. Very often, chemists work hard in the attempt of increasing the yield of the desired product and decreasing that of unwanted byproducts, and one of the shortcuts is certainly represented by the refinement of the reaction conditions: solvent, temperature, and catalyst. This part of the game is still a kind of magic affair where expertise, knowledge, intuition, and luck all together conspire to the goal. Few and sparse

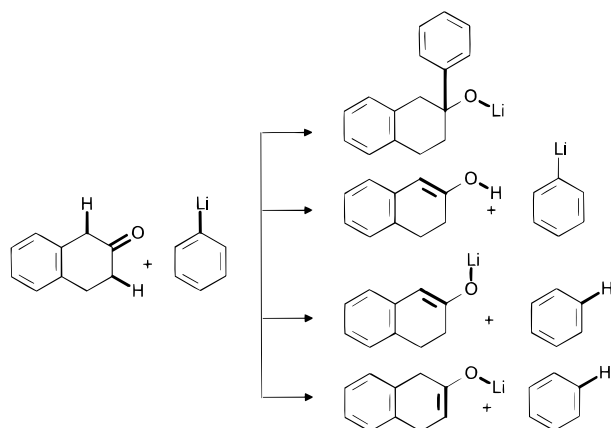
rationale has been developed over the years, mostly because of the giant complexity of the task. In addition, good experimental results have limited the interest of scientists.

A different situation is however present in the field of reaction modeling. Here, the need of the model development strictly requires a rationale, more or less sustainable, behind each activity; and if the rationale is not in the current knowledge, it must be made from scratch. The combined difficulties added by the problem complexity and by the lack of a theory have confined the offer of solutions. Essentially, there are two main schemes: the intelligent search through databases<sup>7</sup> and the elaboration of data into knowledge bases used by prediction programs.<sup>1,2,8</sup> Of course, there are many approaches to the modeling of the solvent in other fields (e.g., molecular dynamics, ab initio simulations), but they are not directed to reaction optimization.<sup>9</sup>

The use of knowledge bases can be seen as the approach most proximate to ours. They are used in programs that operate in the same field, and they attempt the solution of the same problem. The idea behind them is the more or less automatic abstraction of reaction conditions from experimental data, followed by a second level analysis in order to make the connections between these data and the predicted reactions as clever as possible.

In this paper we describe some of our most recent achievements concerning the selection of the best set of reaction conditions for a specific reaction. In particular, we will concentrate on the selection of the best solvent, in the sense of the solvent that minimizes side reactions. The approach considers one specific reaction as the principal one and all the others as side reactions and determines the best set of reaction conditions to favor the former. Presently, the principal variable of our reaction conditions is the solvent in which we perform the transformation. Our idea is that a solvent should favor a kinetically controlled reaction if it is a good solvent for the transition state and a bad solvent for the ground state, decreasing the activation energy. A prerequisite of the analysis is its speed; because we are going to analyze complex synthesis plans, we cannot accept any

\* To whom correspondence should be addressed. Phone: 39-02-2363469. Fax: 39-02-2364369. E-mail: sellogui@icil64.cilea.it.

**Scheme 1.** Products Proposed by Beppe for the Reaction between 2-Decalone and Phenyllithium

time requiring description of the interactions between the reactants and the products and the solvent. Thus, we adopt simple descriptions of both aspects.

Transition states depend on the specific reaction; however, they share a common property: they are either more or less polar than ground states. Considering that the molecules present in both states are similar in skeletal and functional aspects, the main discriminant is exactly the change in polarity determined by the partial bond breaking and forming. Solvation is a complex concurrence of many interactions that occur locally. Their description is out of question in our perspective; nevertheless, we need a measure of state solvation. We generally apply the principle of “similarity in solvation”; i.e., we calculate an approximate similarity between reactants, transition states, and solvent molecules. The more the former are similar to the latter, the more they are solvated. This permits the selection and the ordering of the solvents. We considered other aspects of reaction conditions that will be briefly commented on.

## BACKGROUND

Throughout the paper we will refer to our reaction prediction system Beppe.<sup>6</sup> It is thus worthwhile to review its principles and its applications. Beppe is a completely theoretical model of reactivity; i.e., it does not resort to a priori classification or schematization of compound interactions. It is based on a very simple scheme: after a perception phase that locates the reactive atom pairs, it assembles products by bond breaking and forming as in Scheme 1.

It is evident that both the selection of reactive bonds and the acquisition of the products is regulated by the calculation of physicochemical descriptors. They are mainly as follows: the bond polarity<sup>10</sup> and the bond hardness<sup>11</sup> for the selection of reactive bonds; the molecular electronic energy,<sup>12</sup> the imaginary intermediate (see below) electronic energy, the electrostatic potential, the molecular hardness, for the selection of the products. In addition, the program includes a second analysis level that, after an accurate analysis of the first level products, selects those suitable for a further reaction and applies the procedure again. In principle, the same action can be repeated at will, but the present option is to stop after two levels and to point to products that can react further (Scheme 2).

One of the characteristics of Beppe that will be of interest in the following is its modeling of the reaction through the

use of an imaginary intermediate (BII). This is realized by forming the bond between the two most reactive atoms of the two pairs, charging them, and allowing the reequilibration of the electrons through the usual mechanism (Scheme 3).

The BII has a fundamental role because it can be interpreted as a model, highly approximated, of the transition state, whose energy is believed to govern the reaction kinetics. However, the use of the BII energy is completely misleading if used as the only control mechanism in our model. More often, it is preferable to consider all the descriptors and to weigh them depending on the current reaction. For example, a hierarchical use of the BII energy, the electrostatic potential, and the hardness gain allowed the prediction of the regioselectivity in some Diels–Alder reactions.<sup>13</sup>

Beppe is at the same time the user and the data provider of the reaction condition model.

## MODEL DESCRIPTION

A major problem is to define what we mean by “reaction conditions” (RC); in general RC include all the factors involved in the reaction but the reactants. Thus, they include many factors such as the shape of the vessel where the reaction takes place, the type of stirring, and so on. We decided to limit the research to the following conditions: solvent, temperature, eventually required catalyst, order of reactant addition and their concentration ratio, and external conditions (humidity or air). We only took into consideration reactions under kinetic control, leaving apart the reaction thermodynamics (from our point of view a reaction in the absence of other undesired reactions can work in any condition).

Now we show how the best RC for the desired reaction are defined. Two general reactants may react in several ways to make products; but normally only one of these is exactly the wanted product; the efforts strictly depend on the attempt to improve the yield of the right reaction. As a consequence, we divided the research of the conditions in two main steps: (a) search for the best conditions of the desired reaction; (b) search for the best conditions, among the ones found above, which at the same time minimize all the side reactions.

The operations done are as follows:

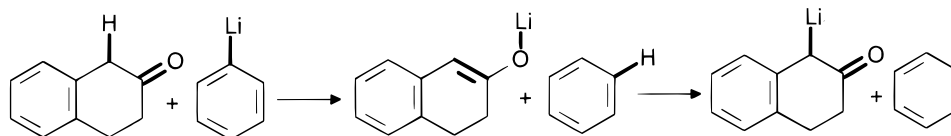
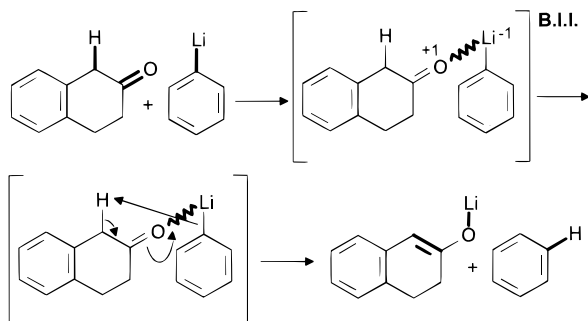
(1) *Input of the reactant molecules.* This is the only operation required to the user.

(2) *Prediction of the reaction products and classification of the reactions.* From the data given by Beppe, all the possible reactions resulting from the reactants are identified. Then, the program makes a classification on which the reaction conditions strictly depend, as shown later.

(3) *Calculation of the reaction rate.* To decide which reactions are competitive, the program computes a rate value using the data obtained from the Beppe program which is independent from the conditions. We call this value the *imaginary rate* (IR).

(4) *Identification of the desired reaction (the principal reaction).* Only for this reaction are the conditions improved. This is, by default, the reaction with the highest IR; but, the user may define any reaction as the principal one.

(5) *Classification of the other reactions into undesired competing and noncompeting reactions.* According to the IR, there is a selection between the reactions which consider-

**Scheme 2.** Example of Second Level Product As Proposed by Beppe: Oxygen–Carbon Anions Exchange**Scheme 3.** From Reactants to Beppe Imaginary Intermediate (BII) to Bond and Electron Movements to Products

ably reduce the yield of the principal one (competing undesired reactions) and the reactions which do not reduce it (noncompeting).

(6) *Search for the best solvents for each identified reaction.*

At this point the program makes a list combining the most suitable solvents which improve each reaction.

(7) *Selection of the best solvents for the main reaction.*

After a comparison between the solvent lists of the principal reaction and of the undesired competing reactions, the program selects those solvents which are both the most suitable for the principal reaction and the worst for the undesired competing reactions.

(8) *Search for the best catalyst of the main reaction.* After the selection we also eliminate the solvents which can give undesired reactions with the selected catalysts.

(9) *Search for the best range of temperature, for the addition order of the reactants, and for their concentration.*

(10) *Search for other conditions.* According to the reactants, the program recommends the anhydrous or anoxic environment.

## REACTION DEFINITION

**Classification.** As previously mentioned, the base of the model is the definition of an II for each reaction. But, the II of this part will be different from BII and must be defined in advance. Consequently, a classification of the reactions is necessary; then, each class can be assigned its own II. The reaction classes here considered are reported in Table 1.

It is evident that the chosen classification scheme is analogous to that commonly used in organic chemistry. The choice is due to the need for simplification in the gathering of the literature data, where this classification is currently used. It is worth noting that monomolecular reactions, like elimination reactions, are excluded; this limit is related to the present program structure, but it could be easily overcome in the near future. Deprotonation reactions are a special class because they are usually very fast reactions that, from a kinetic viewpoint, can very often represent highly competitive side reactions. To limit their excessive influence, they are considered as special cases, particularly when selecting the solvent list.

**Table 1.** Reaction Classification<sup>a</sup>

class	subclass
nucleophilic substitution	$S_N1$ $S_N2$ allyl–benzyl $S_N$ acyl– $S_N$
nucleophilic substitution at C–C double bond	b
nucleophilic addition to C–E double bond	
electrophilic addition to C–C double bond	c
aromatic electrophilic substitution	
deprotonation	

<sup>a</sup> The separation between charged and neutral nucleophile (b)/electrophile (c) is required because the reaction kinetics of the two cases are different; this will have fundamental importance in the calculation of the solvation that, being essentially based on the atom charges, strictly depends on the more or less intense dispersion of the charges going from the ground state to the II.

**Product Prediction and IR Calculation.** As previously mentioned, the imaginary reaction rate is dependent on the values of three descriptors calculated by Beppe:<sup>6</sup> (1) *elstatic* (electrostatic contribution of the reaction); (2) *transit* (energy level of the II); (3) *hardness* (hardness contribution of the reaction).

These three descriptors are used as a means for determining the influence of the corresponding physical effects on the reaction control. More exactly, elstatic depends on the electrostatic control, transit on the covalent control, and hardness on the hard–soft, or polarizability, control. Each of them has a different unit of measure (potential, energy, and charge, respectively), but all of them are proportional to the corresponding energy contribution. In fact, transit is already of the right dimension, elstatic is the electrostatic potential that multiplied by the interacting charge gives the corresponding energy, and hardness is the charge polarizability that multiplied by the interacting electric field gives the corresponding potential and subsequently the energy. Their direct combination is thus forbidden; however, their percentage variations with respect to a reference are absolute numbers that can be combined. In principle, any mathematical combination is allowed; we chose a linear combination after an accurate analysis of different alternatives with respect to the better or worse representation of experimental behaviors. A complete description of their calculation can be found in previous papers,<sup>6,11,12</sup> but, to permit a qualitative understanding, we will briefly describe the equations used for their calculation.

$$\text{elstatic} = \left[ \sum_i (Q_i Q_j / (R_{\text{cov}_i} + R_{\text{cov}_j})) \right]_P - \left[ \sum_i (Q_i Q_j / (R_{\text{cov}_i} + R_{\text{cov}_j})) \right]_R$$

where  $Q_i$  and  $Q_j$  are atomic charges,  $R_{\text{cov}_i}$  and  $R_{\text{cov}_j}$  are covalent radii, and the sum concerns all the bonded atom pairs; the two terms are evaluated using the charge distribution<sup>14</sup> on the products and on the reactants, respectively.

$$\text{transit} = \sum_i E_i$$

where  $E_i$  is the energy of each atom calculated using the formula<sup>12</sup>

$$E_i = k_3(A + B + C) - k_2N_3$$

$$k_3 = -k_1/((Z^0)^*\text{Rcov}^0)$$

$$A = (N^2 + aN - 2NN_1 - 2bNN_2 + N_1^2 + 2bN_1N_2 - aN_1 + b^2N_2^2 - abN_2)N_3$$

$$B = 0.5(-2aN + 2aN_1 + 2abN_2 - a^2)N_3^2$$

$$C = (a^{2/3})N_3^3$$

where  $k_1$  and  $k_2$  are constants depending on the atom type,  $(Z^0)^*$  is the effective nuclear charge of the isolated atom for a complete electron shielding, and  $\text{Rcov}^0$  is the covalent radius of the isolated atom.  $N$  is the atomic number;  $N_1$ ,  $N_2$ , and  $N_3$  are the shell occupation numbers. These last three numbers are calculated for a structure that has a bond between the interacting atoms already formed and the interacting atoms partially charged, regardless of the reaction mechanism.

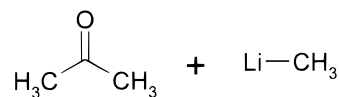
$$\text{hardness} = [\sum_i (H_i - H_j)]_P - [\sum_i (H_i - H_j)]_R$$

where  $H_i$  and  $H_j$  are the hardnesses<sup>11</sup> of each atom in each bonded atom pairs, either in the products or in the reactants.

In the model they concur with the reaction rate figure: the rate increases in correspondence with an increase of elstatic and transit (due to Beppe convention, the higher transit is, the more stable is the II) and to a decrease of hardness. Unfortunately, the exact connection between the descriptor values and the rate is not possible (see Beppe reference). In order to guess an acceptable relation, we analyzed many experimental examples and we formulated the following rate equation:

$$\text{IR} = \left( \frac{\text{elstatic} \times 100}{\text{elstatic}} - 100 \right) + \left( \frac{\text{transit} \times 100}{\text{transit}} - 100 \right) - \left( \frac{\text{hardness} \times 100}{\text{hardness}} - 100 \right)$$

Elstatic, transit, hardness are the arithmetic mean of the three descriptors taken on all the reactions coming from Beppe. The IR is then calculated by summing the average of the percent differences. It is clear that this is not a usual kinetic rate; consequently, it has been called an imaginary rate (IR), stressing its correlation to the II. The usual kinetic rate is directly proportional to the energy difference between the transition state and the ground state; it can be calculated by several theoretical energies obtained using different quantum-mechanical approaches. However, their use is out of the question in the present development. Our IR cannot be compared to the real kinetic rate; it is an attempt to calculate a number that, containing contributions from the three main physical effects that govern a reaction, can be used to compare the feasibility of competitive reactions generated by the same reactants.



**Figure 1.** Reaction between acetone and methyllithium.

**Table 2.** Reactions from Beppe and Their Descriptors

$N$	reacn	elstatic transit hardness
1	$\rightarrow \text{H}_3\text{C}-\text{C}(\text{OH})(\text{CH}_3)-\text{CH}_2\text{Li}$	-0.99 113.00 6.20
2	$\rightarrow \text{H}_2\text{C}(\text{Li})-\text{C}(=\text{O})\text{CH}_3 + \text{CH}_4$	-0.95 118.64 5.31
3	$\rightarrow \text{H}_3\text{C}-\text{C}(\text{OLi})(\text{CH}_3)_2$	-1.29 114.95 6.00
4	$\rightarrow \text{H}_3\text{C}-\text{C}^+(\text{O})\text{CH}_3 + \text{Li}-\text{CH}_3$	-0.72 112.49 3.56
5	$\rightarrow \text{H}_3\text{C}-\text{C}(\text{OCH}_3)(\text{CH}_3)-\text{Li}$	0.00 116.40 5.96

**Table 3.** Mean Values

descriptor	mean value
elstatic	-0.79
transit	115.10
hardness	5.41

**Table 4.** Percent Difference of Descriptors of Eliminated Reactions

ID	descriptor	value	difference (%)
4	hardness	3.56	-34.2
5	elstatic	0.00	-100.0

Given the reactants sketched in Figure 1, Beppe suggests the reactions in Table 2. Given the values reported in Table 2, we can calculate the mean values of elstatic, transit, and hardness (Table 3).

Then, the IRs are normalized by setting the value of the slowest reaction equal to zero and that of the fastest equal to 100, intending by "slowest" and "fastest" the two reactions that have extreme values of the descriptors regardless of their recognition by the program as "true" reactions. At this level all the reactions that have at least one value lower than 80% of the mean values are eliminated from the analysis; in this way, we eliminate all the reactions that are highly improbable because they have one or more parameters used in the calculation of the corresponding rate so low that the resulting rate would be negligible (in the previous example, reactions 4 and 5) (Table 4).

The IR of the remaining reactions are now calculated (Table 5).

The IR values can be subsequently modified by the steric congestion of the reaction centers. The steric congestion is calculated by the program Viper;<sup>15</sup> it assigns a congestion class to each reaction, and, on this basis, the corresponding



**Table 5.** Descriptor Contribution to Imaginary Rate

<i>N</i>	elstatic	transit	hardness	IR
1	25.91	-1.82	-14.71	9.4
2	20.07	3.08	+1.83	24.9
3	62.86	-0.13	-11.08	51.6

**Table 6.** Congestion, Congestion Classes, and Coefficients

congestion	congestion class	<i>k</i>
Hin < 15	low	1
15 ≤ Hin ≤ 50	medium-low	0.8
50 ≤ Hin ≤ 100	medium-high	0.5
100 ≤ Hin < 300	high	0.2
Hin ≥ 300	very high	0

rate is modified using fixed constants (Table 6). The final IR are thus obtained. Hin derives from the sum of the steric congestion of the two bonds of the reactants that are going to react.

$$\text{Hin} = \text{Hin}_{\text{R1}} + \text{Hin}_{\text{R2}}$$

Each of them is calculated by the program Viper that uses the volumetric congestion given by the atoms in the first three connection spheres of each atom in the bond pair. The calculated values range from 0 (e.g., MeLi plus acetone) to more than 300 (e.g., tBuLi plus tBuI, 333.4). The coefficients in Table 6 come from the analysis of several different situations and permits the application of a corrective factor to classes of sterically congested compounds. It is important to understand that this factor is not experimentally very important in the selection of the reaction conditions; consequently, it is introduced only to distinguish well-defined steric selectivity. In the example of Figure 1 the IR are not modified.

Using the IR values, the reactions are divided into the principal reaction and the set of the side reactions. By default, the principal reaction is that showing the highest IR; however, the user can define any reaction as the principal reaction. Then, the program further divides the side reactions into the sets of the interfering and noninterfering reactions, as mentioned above. Presently, we define as interfering all the reactions that have an IR equal to or greater than 50% of the rate of the principal reaction. This division will allow in the following for the optimization of the solvent choice of the principal reaction. A pertinent example is reported in Scheme 4.

These rates are, after normalization, as follows: (1) IR = 100.0 and (2) IR = 91.0.

In the reference paper, the yield of reaction 1 is equal to 81%. From consideration of very similar reaction conditions for both transformations, the yield is mainly related to the existence of a greater steric congestion disfavoring reaction 2. The calculated steric congestions are as follows: (1) Hin<sub>1</sub> = 295.00, *k* = 0.2; (2) Hin<sub>2</sub> = 317.94, *k* = 0.

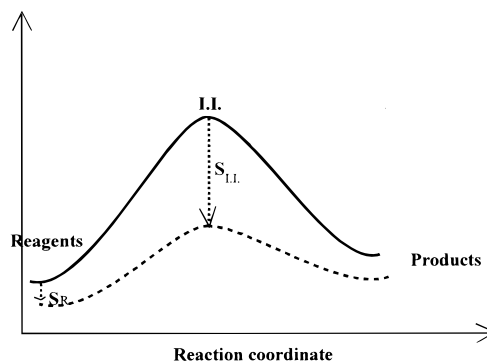
Consequently, the final normalized IR are (1) IR<sub>fin</sub> = 20.0 and (2) IR<sub>fin</sub> = 0.0, in better agreement with the experimental yield, which reports the formation of only one product.

## REACTION CONDITIONS

**Solvent Selection. (a) Solvation Scale.** In order to select the best solvents of the principal reaction, we will again use the idea of the II. We assume that the best solvent of a

**Scheme 4.** Descriptor Contribution to the Imaginary Rate (IR)

Reaction	Elastic	Transit	Hardness	I.R.
1)	-9.411	0.597	-3.245	-12.058
2)	-10.491	-0.024	-3.247	-13.762

**Figure 2.** Solvent effect on reactants and imaginary intermediate.

particular reaction is that better solvating the II in comparison to the reactants, still ensuring the solvation of the latter (Figure 2).

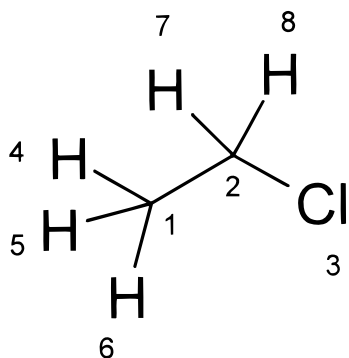
Therefore, given a solvent, the overall stabilization that it gives to a reaction can be assumed to be equal to the difference between the solvation of the two states:

$$\Delta S = S_{\text{II}} - S_{\text{R}}$$

*S*<sub>II</sub> is the solvation of the II and *S*<sub>R</sub> the solvation of the reactants. The greater  $\Delta S$  is, the greater the solvent will stabilize the II and, as a consequence, will facilitate the reaction.

In our approach, the calculation of the solvation degree of a compound is based on the similarity between the solvent and the compound. The needed similarity has been defined on purpose, as described in the following. The first operation applied to all the molecules participating in the reaction consists of the calculation of a coefficient ( $\gamma$ ) correlated with the charge of each atom. The coefficient is calculated for the atoms of both the reactants and all the solvents.  $\gamma$  derives from the combination of two parts: one controlled by the Pauling electronegativities of the atoms ( $\chi$ ) and the other dependent on the ionic or covalent radii of the atoms obtained from literature data,<sup>16</sup> values that are assigned to each atom at perception time:

$$\gamma_i = f(\chi)_i + f(\chi, r)_i$$

**Figure 3.** Calculation of  $\gamma$  coefficients of ethyl chloride.**Table 7.** Calculation of  $f(\chi)$  of Atom 2

$\chi_2$	bonded atoms	$\chi_k$	$\chi_k - \chi_2$
2.5	1	2.5	0.0
	3	3.0	0.5
	7	2.1	-0.4
	8	2.1	-0.4

$$f(\chi)_2 = [0 + 0.5 + (-0.4) + (-0.4)]/4 = -0.07250$$

Given a standard atom  $i$ ,  $\gamma_i$  is the value of  $\gamma$ ;  $f(\chi)_i$  is the part calculated using the electronegativities of the atom and of its neighbors, and  $f(\chi, r)_i$  is the part derived from its ionic or covalent radius. The function  $f(\chi)_i$  is the mean of the electronegativity differences between the atom and its neighbors:

$$f(\chi)_i = \frac{1}{\text{Val}_i} \sum_{k=1}^{\text{Val}_i} \chi_k - \chi_i$$

where  $\text{Val}_i$  is the number of the atoms bonded to atom  $i$ ,  $\chi_i$  is its electronegativity, and  $\chi_k$  is the electronegativity of its neighbors. We consider the molecule in Figure 3. Considering atom 2,  $f(\chi)_2$  is calculated as given in Table 7.

The function  $f(\chi, r)_i$  is the ratio of  $f(\chi)_i$  and the ionic or covalent radius multiplied by a constant:

$$f(\chi, r)_i = \frac{1}{100} \frac{f(\chi)_i}{r}$$

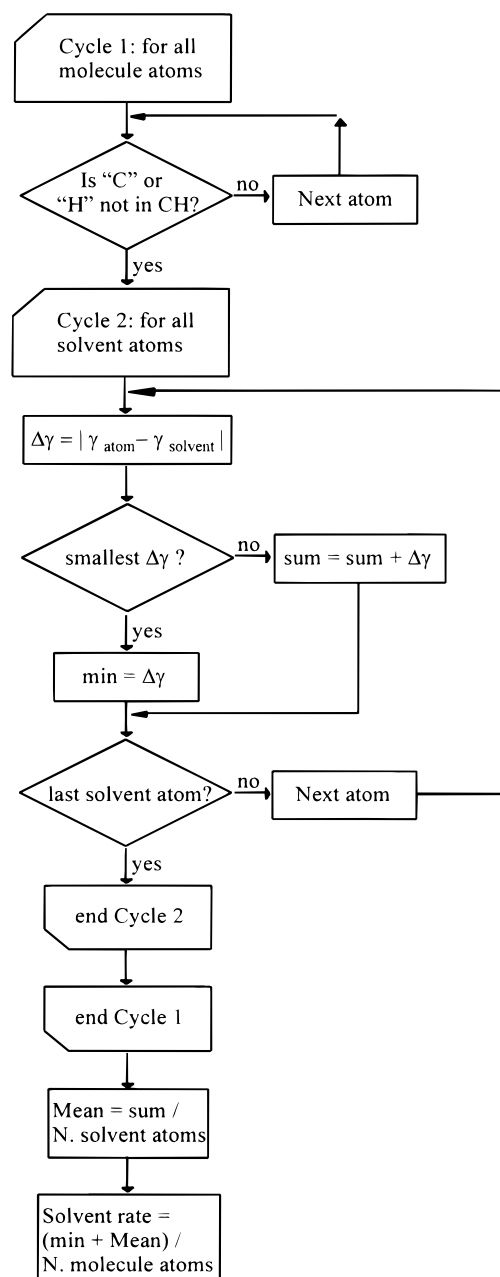
$f(\chi, r)_i$  represents the sensitivity of the atom to the electron distribution and can be associated with its polarizability. In the previous example, again considering atom 2,  $f(\chi, r)_2$  is

$$f(\chi, r)_2 = \frac{1}{100} \left( -\frac{0.07250}{0.77} \right) = -9.4 \times 10^{-4}$$

Atom 2 is an  $\text{sp}^3$  hybridized carbon atom, and its covalent radius is equal to 0.77. From  $f(\chi, r)_i$  and  $f(\chi)_i$  we can calculate  $\gamma$ , that is, their sum. In the example,  $\gamma$  of atom 2 is

$$\gamma_2 = -0.07250 + (-0.00094) = -0.07344$$

Using the  $\gamma$  of all the atoms, we can calculate the solvation value of each solvent with respect to each molecule. The realized system is based on the similarity between the  $\gamma$ s of the atoms of the molecule and of the solvent. The procedure for the calculation of the solvation is sketched in Scheme 5. Having calculated the  $\gamma$  values, we locate, for each atom of the molecule excluding the hydrogen atoms bonded to carbon atoms, the value of  $\gamma$  most similar to a corresponding atom

**Scheme 5.** Flow of the Calculation of the Similarity between Solvents and Guest Compounds

of the solvent; i.e., for each atom of the molecule

$$\Delta\gamma_{\min} = \min_{k=1 \dots M} |\gamma - \gamma_{S_k}|$$

where  $\Delta\gamma_{\min}$  is the absolute difference between the two most similar  $\gamma$ s,  $M$  is the number of the solvent atoms,  $\gamma$  is the value of the molecule atom, and  $\gamma_{S_k}$  is the value of solvent atom  $k$ . Then, the  $\Delta\gamma_{\min}$  are summed by obtaining a first similarity value, which only considers the most similar atoms:

$$S_{\max} = \sum_{i=1}^N \Delta\gamma_{\min i}$$

where  $S_{\max}$  is the value of the *highest similarity* between the molecule and the solvent; i.e., the value that we should have if the solvent is made only by the atoms most similar to those of the molecule.  $N$  is the number of the molecule atoms.

The next step is the calculation of the *mean similarity* between the molecule and the solvent; this value is obtained by summing all the  $\Delta\gamma$ s resulting from the comparison of all the  $\gamma$ s of the molecule and of the solvent and dividing the result by the number of the solvent atoms:

$$\bar{S} = \frac{1}{M} \sum_{j=1}^N \sum_{k=1}^M |\gamma_j - \gamma_{s_k}|$$

Finally, we can calculate the *similarity of each molecule atom to each solvent atom*. It is obtained by summing  $S_{\max}$  and  $\bar{S}$  and dividing by the number of the molecule atoms:

$$S = 1/N(S_{\max} + \bar{S})$$

A simple example can help in understanding the procedure. We consider again ethyl chloride (Figure 3) and two solvents: ethyl ether and ethyl acetate. The first step is the calculation of the  $\gamma$ s of all the atoms; for the non-H atoms, they are as follows: ethyl chloride,  $-0.30$ ,  $-0.07$ , and  $-0.52$ ; ethyl ether,  $-0.30$ ,  $+0.05$ , and  $-1.02$ ; ethyl acetate,  $-0.30$ ,  $+0.84$ ,  $-1.52$ ,  $-1.02$ ,  $+0.05$ , and  $-0.30$ .

The corresponding  $\Delta\gamma_{\min}$  are as follows: ethyl ether,  $0.0$ ,  $0.12$ , and  $0.21$ ; ethyl acetate,  $0.0$ ,  $0.12$ , and  $0.21$ .

Thus,  $S_{\max}$  is  $0.33$  for both solvents. In this case,  $S_{\max}$  is unable to distinguish the two solvents. We calculate  $\bar{S}$ : ethyl ether,  $\bar{S} = 1/15(8.60 + 6.53 + 11.34) = 1.77$ ; ethyl acetate,  $\bar{S} = 1/14(9.18 + 7.80 + 11.29) = 2.02$ . The numbers in parentheses are calculated by summing the  $\Delta\gamma$ s of all the atoms of the solvent with respect to the atoms of the substrate.

The final  $S$  values are as follows: ethyl ether,  $S = (1.77 + 0.33)/3 = 0.70$ ; ethyl acetate,  $S = (2.02 + 0.33)/3 = 0.78$ . They represent the similarity values and are considered an index of the solvation of ethyl chloride in the two solvents.

The procedure is applied to both reactants, before the reaction starts; the  $S$  values of the two molecules for each solvent are then summed by giving the total solvation of the reactants:

$$S_R = S_1 + S_2$$

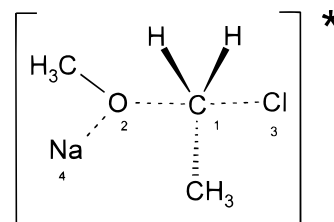
where  $S_R$  is the total solvation in a particular solvent and  $S_1$  and  $S_2$  are the solvations of each reactant.

Repeating the calculation of  $S_R$  for all the considered solvents, we obtain an initial scale of their goodness. The standard set of solvents is reported in Table 8; they have been selected among those most commonly used in laboratories. Nevertheless, the user can add any desired solvent to the list, introducing it directly from the outside. Therefore, the solvent set is practically unlimited.

To this point, the calculations are done only once for each set because the results are valid for all the transformations given by that set. On the contrary, all the operations performed from here on will concern a single transformation, and they will be repeated for each reaction previously defined by the procedure. Differently from the Beppe style, now we will not build the structure of the *imaginary intermediate* from the structures of the reactants and products, but we ascribe a special II to each reaction class. For example, a reaction classified as Aliphatic nucleophilic substitution— $S_N2$  has an II of reaction as represented in Figure 4.

**Table 8.** Standard Solvent Set

acetic acid	hexane
acetone	methanol
acetonitrile	methyl acetate
benzene	methylene chloride
carbon tetrachloride	nitrobenzene
chloroform	nitromethane
1,2-dimethoxyethane	phenol
<i>N,N</i> -dimethylformamide (DMF)	propanol
dimethyl sulfoxide (DMSO)	2-propanol
<i>p</i> -dioxane	pyridine
dyglime	sulfolane
ethanol	tetrahydrofuran (THF)
ethyl acetate	toluene
ethyl ether	triethylamine
ethylenediamine	trifluoroacetic acid
ethylene glycol	water
hexamethylphosphorous triamide (HMPTA)	



**Figure 4.**  $S_N2$  imaginary intermediate.

The program builds its approximated representation by the following operations: (a) the bonds between interested atoms are properly varied, creating new bonds between previously unbonded atoms; (b) new correct  $\gamma$  are obtained in order to calculate the solvation.

In the example of Figure 4, a new bond is made between the oxygen atom and the central carbon atom. The bond length is not important in the present calculation, because the only considered variation concerns the change in the  $\gamma$ s. In this case, the  $\gamma$ s are varied in the following way:

$$\gamma_1 = \gamma_1^0 - \gamma_2^0/2 + \gamma_3^0/2 = -0.08 \quad (\gamma_1^0 = -0.73)$$

$$\gamma_2 = \gamma_2^0/2 = -0.91 \quad (\gamma_2^0 = -1.83)$$

$$\gamma_3 = \gamma_3^0/2 = -0.26 \quad (\gamma_3^0 = -0.52)$$

$$\gamma_4 = \gamma_4^0 = 2.63$$

The 0 index indicates the initial  $\gamma$ s, reported in parentheses. It is clear that the calculated values are a rough approximation of the charge distribution in a “true” transition state, but, at the present level, they are sufficient to allow the calculation of the solvation of our II.

The compound II is again processed exactly in the same way as the reactants. Of course, the new solvation value will be different, being different from the compound compared to the solvents. It is worth noting that the II is always a single structure; then  $S_{II}$  is not composed of two factors as was the situation of the reactants. Finally, the total solvation difference is calculated,

$$\Delta S = S_{II} - S_R$$

To this point, the solvent order of a reaction is only based on the difference of solvation between the II and the starting compounds, while the presence of competing reactions is

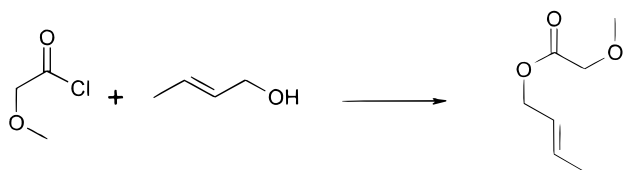
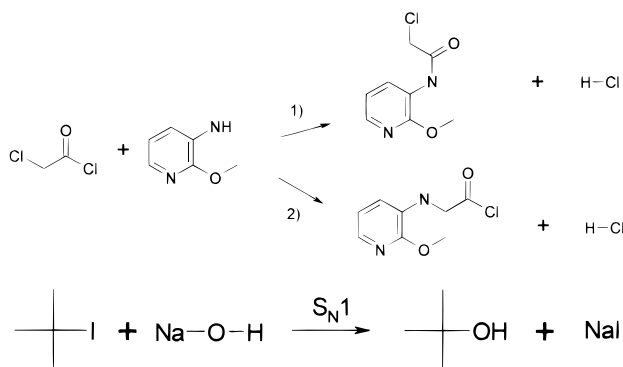
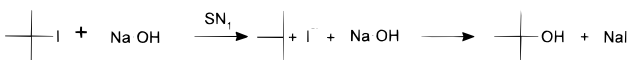


Figure 5. Example reaction for the reactive solvent elimination.

Table 9. Reactive Solvent Excluded

solvent	rate	exclusion ground
water	100.0	nucleophile; addition to multiple bonds
ethylene glycol	91.9	nucleophile
trifluoroacetic acid	88.2	addition to multiple bonds
acetic acid	83.3	addition to multiple bonds

Scheme 6. Two Alternative Reactions That Can Be Pushed by Solvent Choice

Figure 6. Catalyst selection: the  $S_N1$  case.Figure 7.  $S_N1$  reaction mechanism.

not yet considered. On the contrary, some factors are considered in order to exclude some selected solvents. The reasons for exclusion are (1)  $pK_a$ ,<sup>17</sup> (2) *reactant solubility*, and (3) *side reactions between solvent and reactant molecules*. Namely, all those solvents that could interact with the reactants, or that could be too bad solvents for the reactants, are eliminated from the list. Considering the reaction in Figure 5, the solvents eliminated from the set are shown in Table 9.

Here, some short remarks on the reactant solubility aspect are required. First of all, it must be absolutely clear that we are not calculating a real solubility of compounds into solvents. This is a problem that has been studied by many groups for a very long time and that is still waiting for a solution. But, we are simply interested in the elimination from the solvent list of those solvents that cannot presumably guarantee a minimal availability of a compound. Consequently, we are going to discard only a few very bad solvents and, more importantly, we calculate a rough solubility power only to this goal. The solubility analysis is mainly based on the presence of heteroatoms in the reactant compound; the

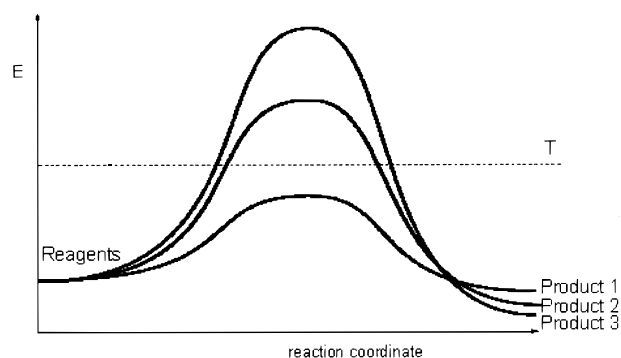


Figure 8. Temperature effect on reaction rates.

factors used are the following: (1) the number of carbon atoms; (2) the number of bonds ("heavy" bonds) between highly electronegative heteroatoms and low electronegative atoms; (3) the number of bonds ("light" bonds) between heteroatoms and atoms with similar electronegativity; (4) the number of metallic ions or atom pairs with  $|\Delta\gamma| \geq 1.5$ .

The solvents of an entire class are eliminated from the list if the circumstances reported in Table 10 are fulfilled. From the table it is clear that the exclusion of a solvent class for solubility reasons is very rare and is limited to the two classes at the extremes of the scale. Further refinement is possible, but we are not presently taking this aspect into more consideration.

**(b) Reaction Comparison.** Now, we are going to compare the solvent ordered lists of the principal reaction and of the secondary interfering reactions. The goal is to cut off the solvent list of the principal reaction all those solvents that also favor the secondary reactions; this is simply achieved by comparing the lists and eliminating all the solvents that are in the first five positions, both of the principal and of the secondary lists. In other words, a solvent that is in positions 1 to 5 of the list of a secondary reaction and, at the same time, in positions 1 to 5 of the list of the principal reaction is eliminated. An example, not referred to any real reaction, is shown in Table 11.

The solvent in the first position of the list of the principal reaction is *ethyl ether*; the same solvent is in position 3 in the list of the secondary reaction, and consequently it is canceled.

The remaining solvents are thus (2) tetrahydrofuran (THF), (3) *p*-dioxane, (4) dichloromethane, and (5) pyridine.

The comparison proceeds to the 12th position of the starting list of the principal reaction. In fact, the remaining solvents below that position are probably too bad for the principal reaction even in the absence of an interfering secondary reaction. This mechanism permits one to favor different reactions using different solvents. An example is sketched in Scheme 6.

The normalized IR are (1) IR = 100 and (2) IR = 70.1.

The yield of reaction 1 is 100% in THF/pyridine.<sup>18</sup> The lists of the best solvents of the two reactions are in Table

Table 10. Relation between Solubility Consideration and Solvent Class Exclusion

fact	effect
more than 10 carbon atoms for each heavy bond	elimination of protic solvents
more than 5 carbon atoms for each light bond AND no heavy bond	elimination of protic solvents
less than 5 carbon atoms for each light bond AND at least 1 heavy bond	elimination of apolar solvents (benzene, hexane; but not ethyl ether)
at least 1 ionic bond	elimination of apolar solvents (benzene, hexane; but not ethyl ether)



**Table 11.** Ordered List of the First Five Solvents of Two Imaginary Reactions

Principal Reaction
(1) ethyl ether
(2) tetrahydrofuran
(3) <i>p</i> -dioxane
(4) dichloromethane
(5) pyridine
Secondary Reaction
(1) <i>N,N</i> -dimethylformamide
(2) acetonitrile
(3) ethyl ether
(4) sulfolane
(5) 2-propanol

**Table 12.** Ordered List of the First Five Solvents of the Reactions of Scheme 6

Principal Reaction 1
(1) pyridine
(2) dichloromethane
(3) ethyl ether
(4) ethylenediamine
(5) nitromethane
Secondary Reaction 2
(1) methanol
(2) ethanol
(3) 2-propanol
(4) propanol
(5) phenol

**Table 13.** Standard Catalyst List

AlCl <sub>3</sub>	HBr(conc)	NaOH
FeCl <sub>3</sub>	HI(conc)	KOH
FeBr <sub>3</sub>	H <sub>2</sub> SO <sub>4</sub> (weak)	AgNO <sub>3</sub>
SnCl <sub>4</sub>	H <sub>2</sub> SO <sub>4</sub> (conc)	Hg <sub>2</sub> Cl <sub>2</sub>
ZnCl <sub>2</sub>	CH <sub>3</sub> COOH(weak)	dimethoxyethane
SbCl <sub>5</sub>	HNO <sub>3</sub> (weak)	Al <sub>2</sub> O <sub>3</sub>
HCl(weak)	N(Bu) <sub>4</sub> Cl	P <sub>2</sub> O <sub>5</sub>
HBr(weak)	N(Bu) <sub>4</sub> Br	pyridine
HI(weak)	18crown6 ether	BF <sub>3</sub>
HCl(conc)	EtONa	

12. Looking at the lists, it is clear that, by an accurate solvent selection, it is possible to give a preference to one reaction. Pyridine is in one of the last positions in the list of reaction 2, in perfect agreement with the experimentally high yield of reaction 1 in this solvent. On the other hand, we cannot affirm that, for example, methanol would make reaction 2 prevailing over reaction 1 (whose list does not have methanol in the first five positions), because we must always take in due account the relative IR values. Nevertheless, the result suggests that polar solvents could to some extent favor reaction 2 with respect to reaction 1. It is obvious that protic solvents are incompatible with reaction 1 and cannot be used; also, this remark is reported in the output file.

**Catalyst Search.** Catalyst selection is performed on the basis of the reaction classification carried out during their perception. The set of the operations that take care of this point is sketched in the following two actions: (1) *designation of a list of catalysts matching the class to which the principal reaction belongs*; (2) *removal from the list of all those catalysts that can give unwanted interactions with the reactants*.

First of all, a set of catalysts has been chosen (Table 13).

During the program development a sublist of catalysts has been assigned to each reaction class known to the program.

**Table 14.** S<sub>N</sub>1 Catalysts

HCl(conc)	H <sub>2</sub> SO <sub>4</sub> (conc)
HCl(aq)	H <sub>2</sub> SO <sub>4</sub> (aq)
HBr(conc)	CH <sub>3</sub> COOH(aq)
HCl(aq)	HNO <sub>3</sub> (aq)
HI(conc)	AgNO <sub>3</sub>
HI(aq)	Hg <sub>2</sub> Cl <sub>2</sub>

**Table 15.** Reactive Excluded Catalysts

catalyst	exclusion ground
HCl(conc)	reaction with OH <sup>-</sup>
HCl(weak)	reaction with OH <sup>-</sup>
HBr(conc)	reaction with OH <sup>-</sup>
HCl(weak)	reaction with OH <sup>-</sup>
HI(conc)	reaction with OH <sup>-</sup>
	common ion effect with the leaving group
HI(weak)	reaction with OH <sup>-</sup>
H <sub>2</sub> SO <sub>4</sub> (conc)	reaction with OH <sup>-</sup>
	dehydration of the formed alcohol
H <sub>2</sub> SO <sub>4</sub> (weak)	reaction with OH <sup>-</sup>
CH <sub>3</sub> COOH(weak)	reaction with OH <sup>-</sup>
HNO <sub>3</sub> (weak)	reaction with OH <sup>-</sup>

For example, a reaction classified as S<sub>N</sub>1 (Figure 6) can potentially use the set of catalysts given in Table 14. The effect of these catalysts is the increase of the fugacity of the leaving group and, consequently, of the yield of an intermediate of the type in Figure 7.

After the selection of the feasible reaction catalysts, they are again considered to see if they can have unwanted interactions with any of the reactants or, only in some cases, with any of the products; e.g., in the previous reaction, the catalysts in Table 15 are excluded. The remaining catalysts at this point are AgNO<sub>3</sub> and Hg<sub>2</sub>Cl<sub>2</sub>, which represent the final result of the selection.

At the end of the procedure a comparative analysis of the solvent list of the principal reaction is repeated with the aim of eliminating all those solvents that can interact with the selected reaction catalysts. Thus, the very final list of the reaction solvents is compiled. It is worth noting that the solvent list is not optimized in relation to the catalyst but only that the solvents interacting with the catalysts are excluded.

Some points can be raised on this part of the system. It is clear that the procedure for the selection of the catalysts should be further improved in the future; the final goal is to realize an entirely theoretical approach, as presently done for the solvent selection. However, there is an important limitation due to the impossibility of collecting the correct  $\Pi$  of all the catalyzed reactions. In addition, our model takes into consideration only reactions that proceed in one step, while an accurate analysis of the catalysts should also consider multistep reactions (e.g., interaction with the catalysts giving a first intermediate, its interaction with the reaction partner giving a second intermediate, its break-up to the final products).

**Search for Other Conditions.** At the end of the solvent selection and of the best catalyst designation for the main reaction, the program performs the analysis of the other conditions that can affect the reaction. They are (1) *temperature*, (2) *order of reactant addition and their relative amount*, and (3) *necessity of anhydrous or anoxic conditions*. Besides, a prediction of a hypothetical yield is done (Figure 8).

**Table 16.** Temperature Selection Scheme

principal react rate	secondary react rate	T (°C)
low	no react	heat (>25)
	low	room temp
	fair	room temp
	high	cool (0–25)
fair	no react	heat (>25)
	low	room temp
	fair	cool (0–25)
	high	cool (<0)
high	no react	room temp
	low	room temp
	fair	cool (0–25)
	high	cool (<0)

**Table 17.** Addition Order of Compounds and Reaction Rate Control<sup>a</sup>

principal react rate	secondary react rate	addition order	concentration
low	no react or lower	substrate to nucleophile	nucleophile = substrate
fair	equal	no matter	nucleophile = substrate
	higher	substrate to nucleophile	nucleophile > substrate
	no react	nucleophile to substrate	nucleophile = substrate
	lower	nucleophile to substrate	nucleophile = substrate
high	equal	no matter	nucleophile = substrate
	higher	substrate to nucleophile	nucleophile > substrate
	no react	nucleophile to substrate	nucleophile = substrate
	lower	nucleophile to substrate	nucleophile = substrate
	equal or higher	no matter	nucleophile = substrate

<sup>a</sup> Nucleophile is conventionally the attacking compound in nucleophilic reactions. It must be substituted by an electrophile in electrophilic reactions.

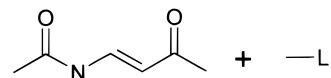
(1) The reaction temperature is inferred by the analysis of the side reactions; the starting assumption is that in the case of a “single” reaction, i.e., a reaction that does not present competing side reactions, the temperature can be as high as possible, thus maximally accelerating the reaction completion. On the contrary, if side reactions of a comparable rate are present, a decrease in the temperature should increase the kinetic control pushing the desired reaction. The possible combinations are in Table 16. The final goal of the temperature setting is to slow as much as possible the secondary reactions but adequately preserve the rate of the principal reaction. The best chance is a temperature that decreases the kinetic barrier just over the energetic top of the main reaction.

(2) Also the addition order of the reactants and their concentration ratio are calculated on the basis of the relative reaction rates; the scheme of the feasible combinations is in Table 17. In the case of electrophilic addition to a carbon–carbon multiple bond or of an electrophilic aromatic substitution, “nucleophile” should be substituted by “electrophile”. The addition order is explicitly suggested only when its definition is possible. The basis of the scheme is the following: when among reactions with similar rates it is still possible to define the fastest one, the nucleophile (or electrophile) will be mainly used by the fastest reaction if it is added stepwise, thus being always in low concentration.

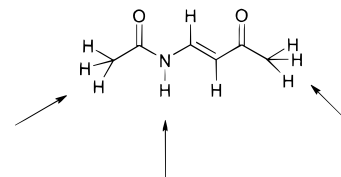
(3) At this point, the program takes into consideration the eventual necessity of excluding humidity or oxygen from the reaction vessel. The decision is taken by just considering the reactivity of the present compounds with water or oxygen, as can be inferred by their  $pK_a$  values. We assume that a compound whose conjugate acid has  $pK_a$  greater than 20 is sensitive to the atmospheric humidity. If the  $pK_a$  is greater than 30 (mostly organometallic derivatives), an inert atmosphere is considered necessary.

**Table 18.** Yield Prediction

principal react rate	secondary react rate	yield
low	no react or lower	fair
	equal	low
fair	higher	low
	no react	high
	lower	fair
	equal	fair
high	higher	low
	no react	high
	lower	high
	equal or higher	fair

**Figure 9.** First example of complete reaction condition selection.**Table 19.** Reactions and Rates for Reactants of Figure 9

	reactn	IR
(1)		82.6
(2)		59.2
(3)		8.5

**Figure 10.** Reactive hydrogen atoms.**Table 20.** Solvent Lists for Reactions of Table 19

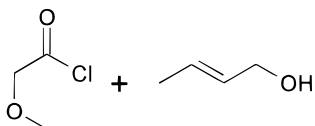
	goodness		goodness
Reaction 1			
tetrahydrofuran	71.2	<i>N,N</i> -dimethylformamide	42.4
ethyl ether	67.9	pyridine	40.9
<i>p</i> -dioxane	57.5	nitrobenzene	35.9
1,2-dimethoxyethane	53.7	triethylamine	33.0
diglyme	53.7	hexamethylphosphorous triamide	0.0
dichloromethane	42.9		
Reaction 2			
nitrobenzene	42.3	1,2-dimethoxyethane	34.1
pyridine	40.5	<i>N,N</i> -dimethylformamide	27.4
<i>p</i> -dioxane	39.8	tetrahydrofuran	17.7
dichloromethane	36.2	triethylamine	14.4
diglyme	34.3	ethyl ether	5.5
hexamethylphosphorous triamide	34.3		
Reaction 3			
dichloromethane	100.0	diglyme	26.1
pyridine	99.2	1,2-dimethoxyethane	25.9
ethyl ether	58.3	triethylamine	17.8
nitrobenzene	34.5	tetrahydrofuran	17.1
<i>p</i> -dioxane	31.3	hexamethylphosphorous triamide	0.0
<i>N,N</i> -dimethylformamide	27.2		

Finally, the yield is predicted on the basis of the reaction rate ratio. There are three classes of yield: *high*, *fair*, and *low*. Table 18 combines the relative rates and suggests the corresponding yield.

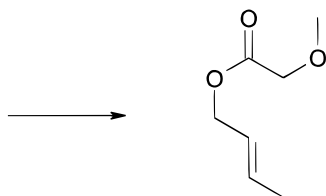
## EXAMPLES

Here, we report two commented examples to show the entire result that can be obtained by the program.

The first reactant pair is in Figure 9.



**Figure 11.** Second example of complete reaction condition selection.



**Figure 12.** Single product of the reaction.

**Table 21.** Excluded Solvents of Reaction of Figure 9

solvent	reason
acetic acid	$pK_a$ ; multiple bond addition
acetone	$pK_a$ ; reactive as electrophile in condensation
acetonitrile	$pK_a$
benzene	solubility
carbon tetrachloride	solubility
chloroform	$pK_a$
dimethyl sulfoxide	$pK_a$
ethanol	$pK_a$
ethyl acetate	$pK_a$ ; reactive as electrophile in condensation and/or acyclic SN
ethylenediamine	$pK_a$
ethylene glycol	$pK_a$
hexane	solubility
methanol	$pK_a$
methyl acetate	$pK_a$ ; reactive as electrophile in condensation and/or acyclic SN
nitromethane	$pK_a$
phenol	$pK_a$
propanol	$pK_a$
2-propanol	$pK_a$
sulfolane	$pK_a$
toluene	solubility
trifluoroacetic acid	$pK_a$ ; multiple bond addition
water	$pK_a$ ; multiple bond addition

**Table 22.** Final Solvent List of Figure 9

tetrahydrofuran	71.2
ethyl ether	68.0
<i>p</i> -dioxane	57.5

**Table 25.** Two Selected Examples of Reaction Condition Prediction

reaction <sup>a</sup>	solvent	other conditions
	sulfolane, DMSO, acetone, methyl acetate	$T$ , 25 °C; yield, high; cat., $(Bu)_4NCl$ , $(Bu)_4NBr$ , dimethoxyethane
	methanol, ethanol, propanol, 2-propanol	$T$ , 25 °C; yield, fair; cat., none

**Table 23.** Solvent List of Reaction of Figure 11

ethyl ether	97.2
methanol	88.7
dichloromethane	86.4
pyridine	85.8

**Table 24.** Solvent List of Reaction of Figure 11 after Catalyst Selection

ethyl ether	97.2
dichloromethane	86.4
pyridine	85.8
<i>p</i> -dioxane	73.1

**First Step: Reaction Location; Descriptor Calculation; IR Calculation; Reaction Ordering.** The reactions found by the program, excluding deprotonation reactions, and the corresponding IR are in Table 19. The deprotonation reactions involve the hydrogen atoms pointed to by the arrows in Figure 10. The greatest IR identifies the principal reaction 1.

**Second Step: Solvent Selection and Ordering; Solvent Exclusion.** The solvent lists for all the reactions are filled (Table 20). The lists are ordered using the similarity approach described above. Then, the solvents that could interact with the reactants or that are too bad solvents are excluded from the lists. It is interesting to note that the three lists are different even if the starting compounds are always the same.

The solvents excluded from the sets are in Table 21.

Finally, combining the reaction IR and the reaction solvent lists, the solvents chosen for the principal reaction are in Table 22:

**Third Step: Calculation of Qualitative Yield; Selection of Other Conditions. (a) Yield Calculation.** The principal reaction is preferred, but there are slower secondary reactions. The first suggestion is addition of the nucleophile to the substrate in small portions; the second suggestion is a decrease in the reaction temperature: yield, fair;  $T$ , 0–25 °C.

**(b) Other Conditions.** No catalyst is chosen, mainly because of the presence of organometallic reactants. The

Table 26.

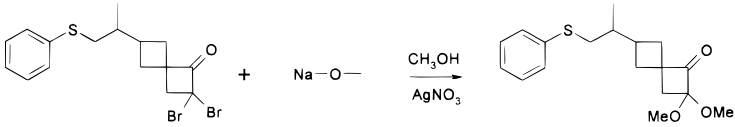
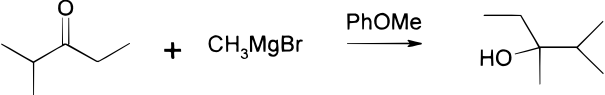
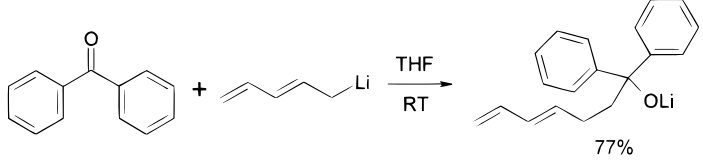
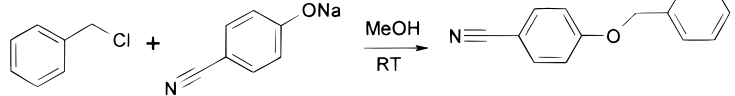
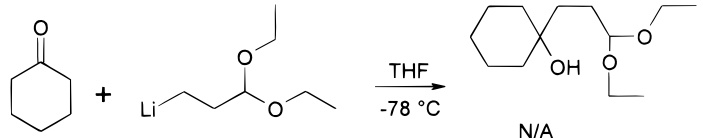
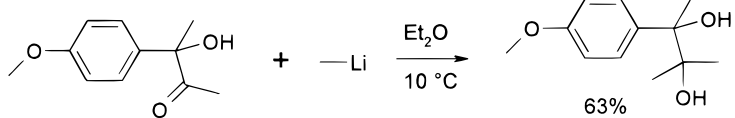
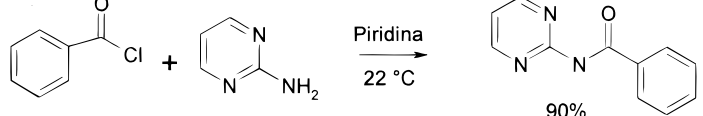
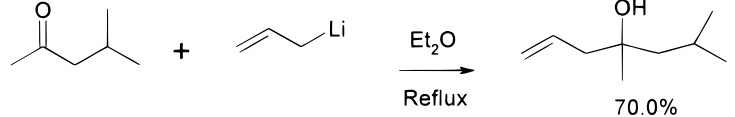
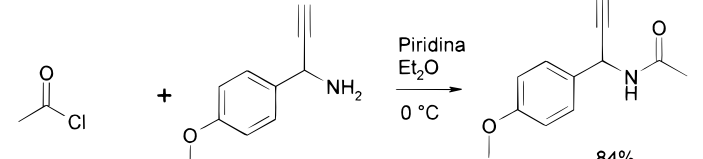
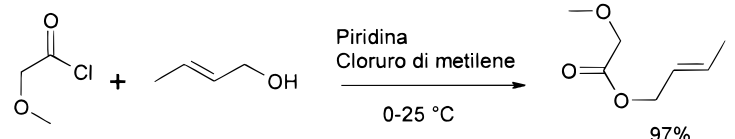
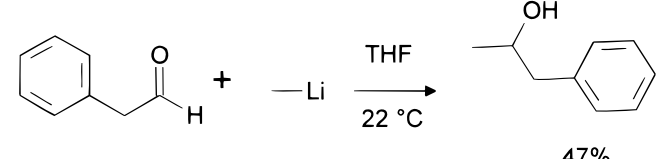
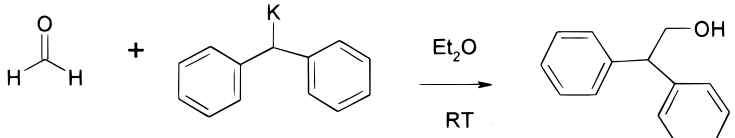
reaction		best solvents	other conditions
		2-propanol, ethanol, <i>methanol</i> , propanol	$T$ , 25 °C; yield, high; cat., $\text{AgNO}_3$ , $\text{Hg}_2\text{Cl}_2$
		tetrahydrofuran, <i>PhOMe</i> , <sup>a</sup> ethyl ether, <i>p</i> -dioxane	$T$ , 25 °C; yield, fair; cat., none
		ethyl ether, <i>tetrahydrofuran</i> , <i>p</i> -dioxane, chloroform	$T$ , 25 °C; yield, fair; cat., none
		<i>methanol</i> , ethanol, 2-propanol, propanol	$T$ , 25 °C; yield, fair; cat., $\text{AgNO}_3$ , $\text{Hg}_2\text{Cl}_2$
		ethyl ether, <i>tetrahydrofuran</i> , <i>p</i> -dioxane, methylene chloride	$T$ , <0 °C; yield, high; cat., none
		tetrahydrofuran, <i>ethyl ether</i> , <i>p</i> -dioxane, 1,2-dimethoxyethane	$T$ , 25 °C; yield, high; cat., none
		methylene chloride, <sup>b</sup> ethyl ether, $\text{CCl}_4$ , chloroform	$T$ , 0–25 °C; yield, fair; cat., <i>pyridine</i>
		<i>ethyl ether</i> , tetrahydrofuran, chloroform, methylene chloride	$T$ , ≥ 25 °C; yield, fair; cat., none
		methylene chloride, <sup>b</sup> <i>ethyl ether</i> , nitromethane, nitrobenzene	$T$ , 25 °C; yield, high; cat., <i>pyridine</i>
		ethyl ether, <sup>b</sup> <i>methylene chloride</i> , <i>p</i> -dioxane, diglyme	$T$ , 25 °C; yield, high; cat., <i>pyridine</i>
		<i>tetrahydrofuran</i> , ethyl ether, <i>p</i> -dioxane, 1,2-dimethoxyethane	$T$ , 25 °C; yield, high; cat., none
		<i>ethyl ether</i> , tetrahydrofuran, methylene chloride, <i>p</i> -dioxane	$T$ , 25 °C; yield, high; cat., none



Table 26 (Continued)

reaction		best solvents	other conditions
		ethyl ether, <i>methylene chloride</i> , sulfolane, ethyl acetate	<i>T</i> , 25 °C; yield, high; cat., none
		CCl <sub>4</sub> , nitrobenzene, nitromethane, acetonitrile, (6) THF <sup>c</sup>	<i>T</i> , 0–25 °C; yield, fair; cat., pyridine
		<i>methylene chloride</i> , <i>ethyl ether</i> , DMF, benzene	<i>T</i> , 0–25 °C; yield, fair; cat., none
		DMF, HMPTA, <i>tetrahydrofuran</i> , <i>ethyl ether</i>	<i>T</i> , 25 °C; yield, high; cat., none
		<i>ethyl ether</i> , <i>tetrahydrofuran</i> , acetonitrile, chloroform	<i>T</i> , 0–25 °C; yield, fair; cat., none
		2-propanol, <sup>d</sup> phenol, CCl <sub>4</sub> , chloroform	<i>T</i> , 25 °C; yield, high; cat., none
		<i>ethyl ether</i> , <i>methylene chloride</i> , hexane, benzene	<i>T</i> , 25 °C; yield, high; cat., none
		<i>tetrahydrofuran</i> , <i>ethyl ether</i> , <i>p</i> -dioxane, 1,2-dimethoxyethane	<i>T</i> , <0 °C; yield, high; cat., none
		<i>ethyl ether</i> , <i>tetrahydrofuran</i> , <i>p</i> -dioxane, 1,2-dimethoxyethane	<i>T</i> , 25 °C; yield, high; cat., none
		<i>ethyl ether</i> , <i>tetrahydrofuran</i> , <i>p</i> -dioxane, 1,2-dimethoxyethane	<i>T</i> , <0 °C; yield, high; cat., none

<sup>a</sup> The solvent PhOMe has been introduced by the user. <sup>b</sup> In these cases pyridine was also chosen as the solvent, placed in one of the top positions, and then discarded because it was too nucleophilic. <sup>c</sup> Tetrahydrofuran is present in the sixth position of the solvent list. <sup>d</sup> The experimental solvent water was in the second position but was discarded because it was too nucleophilic.

presence of a highly reactive compound, due to the bond between a carbon atom and a lithium atom, has been detected. Consequently, the reaction temperature is lowered to  $T \ll 0$  °C, and the use of an inert atmosphere is suggested.

#### COMPARISON WITH EXPERIMENTAL DATA

Experimental data<sup>19</sup> are as follows: tetrahydrofuran,  $T = -78$  °C, no catalyst, yield = 57%. The agreement is very good.

The second reactant pair is given in Figure 11.

**First Step: Reaction Location; Descriptor Calculation; IR Calculation; Reaction Ordering.** The program finds a single reaction (Figure 12). Therefore, the descriptors are not used to calculate the IR.

**Second Step: Solvent Selection and Ordering; Solvent Exclusion.** The solvent list for the reaction is prepared. It is ordered using the similarity approach, and the interfering and bad solvents are excluded. The best solvent list is in Table

23. In this case, the list is not the final one because in the following section the catalyst selection will modify the list.

**Third Step: Calculation of Qualitative Yield; Selection of Other Conditions. (a) Catalyst Selection.** Because the reaction could need a catalyst, the analysis steps to the catalyst selection. The best catalyst is

catalyst	level	action
pyridine	20/20	neutralizes the forming acid (stoichiometric or solvent)

The program perceives that the reaction produces an acid (HCl); consequently, a neutralizing base (pyridine) is selected. Its use can be either as a catalysts only (stoichiometric to the acid) or, being pyridine, also a solvent, as both the catalyst and the solvent. A consequence of the catalyst selection is a change in the solvent list that is shown in Table 24.

**(b) Yield Calculation.** Because secondary reactions are absent, as is steric congestion, both the yield and the temperature are not changed from standard: yield, high;  $T \sim 25^\circ\text{C}$ .

**(c) Other Conditions.** The reaction does not need additional conditions.

#### COMPARISON WITH ADDITIONAL EXPERIMENTAL DATA

The experimental data<sup>20</sup> are as follows:  $\text{CH}_2\text{Cl}_2$ ,  $T = 0-25^\circ\text{C}$ , catalyst = pyridine (2 equiv), yield = 97%. Also in this case the agreement is very good.

The program has been applied to some other examples ( $\sim 300$  reactions, partly from literature data<sup>21</sup> and partly created on purpose), and the results are encouraging. In approximately 80% of the studied cases the experimentally used solvent is among the first five selected of the list, and the other reaction conditions are very similar to those predicted by the program. For the sake of showing that the program can also analyze complex compounds, two examples are shown in Table 25. In addition, in the Appendix there is a list of other reactions with the calculated and the experimental conditions. One comment concerns the solvent selection that is always in agreement with the experiment. A second comment concerns the other conditions and the calculation of the yield. These aspects still need some refinements, but, in general, they are already in line with laboratory experience. A final point concerns the selection of the principal reaction. This can be done manually, as reported above, but it is encouraging that in all the studied cases the principal reaction, automatically selected, is always the reaction that is reported by the experimental data.

#### CONCLUSION

The selection of the reaction conditions more appropriate to improve the yield of a reaction with respect to potential side reactions still requires the skill and the intuition of the chemist. Neither the analysis of the giant amount of experimental data nor the everyday growing power of the theoretical calculations can give conclusive answers. Nevertheless, both the need of completing our approach to reaction product prediction and the challenge presented by the problem have moved us toward this investigation. The

results are encouraging and, despite the apparent simplicity of our approach, they are in good agreement with the experimental data. Further refinements are, however, possible, particularly in the catalyst selection, but the realized system is ready to use even at this stage.

#### ACKNOWLEDGMENT

Partial financial support by the Consiglio Nazionale delle Ricerche and by the Ministero dell'Università e della Ricerca Scientifica e Tecnologica is gratefully acknowledged. M.D. would like to thank the Organizing Committee of the 5th Chemical Structure Conference for the invitation to present part of this paper and The Chemical Structure Association Trust for a bursary to attend the Conference. We would like to thank Prof. J. B. Hendrickson for helpful suggestions.

#### APPENDIX

Examples of program results for reactions from the REACCS database are given in Table 26.

#### REFERENCES AND NOTES

- (1) Ihlenfeldt, W.-D.; Gasteiger, J. Computer-Assisted Planning of Organic Syntheses: The Second Generation Of Programs. *Angew. Chem., Int. Ed. Engl.* **1995**, *34*, 2613-2633.
- (2) Jorgensen, W. L.; Laird, E. R.; Gushurst, A. J.; Fleischer, J. M.; Gothe, S. A.; Helson, H. E.; Paderes, G. D.; Sinclair, S. CAMEO: A Program for the Logical Prediction of the Products of Organic Reactions. *Pure Appl. Chem.* **1990**, *62*, 1921-1932.
- (3) Tratch, S. S.; Zefirov, N. S. Systematic Search for New Types of Chemical Interconversions: Mathematical Models and Some Applications. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 331-348.
- (4) Blurock, E. S. Use of Atomic and Bond Parameters in a Spectral Representation of a Molecule for Physical Property Determination. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 1111-1118.
- (5) Satoh, H.; Funatsu, K. Further Development of s 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.
- (6) Sello, G. Reaction Prediction: The Suggestions of the Beppe Program. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 713-717.
- (7) E.g., the databases REACCS, CASREACT, SYNLIB, ORAC, and CROSSFIRE plus REACTIONS.
- (8) Satoh, H.; Funatsu, K. SOPHIA, a Knowledge Base-Guided Reaction Prediction System-Utilization of a Knowledge Base Derived from a Reaction Database. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 34-44.
- (9) E.g.: Drago, R. S.; Hirsch, M. S.; Ferris, D. C.; Chronister, C. W. A Unified Scale of Solvent Polarities for Specific and Non-specific Interactions. *J. Chem. Soc., Perkin Trans. 2* **1994**, 219-230.
- (10) Baumer, L.; Sala, G.; Sello, G. Computer Assisted Organic Synthesis Planning: Effective Bond Polarity as a Guideline to Reactivity. *J. Am. Chem. Soc.* **1991**, *113*, 2994-2500.
- (11) Baumer, L.; Sala, G.; Sello, G. A New Method for the Calculation of Atomic and Local Hardness. *J. Comput. Chem.* **1990**, *11*, 694-699.
- (12) Baumer, L.; Sello, G. A New Method for the Calculation of Natural Bond Polarity Using Molecular Electronic Energy. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 125-130.
- (13) Sello, G. Reaction Prediction by the "Beppe" Program. The Diels-Alder Cycloaddition. *Theochem.* **1995**, *340*, 29-43.
- (14) (a) Baumer, L.; Sala, G.; Sello, G. Residual Charges on Atoms in Organic Structures: A New Algorithm for Their Calculation. *Tetrahedron Comput. Met.* **1989**, *2*, 37-46. (b) Baumer, L.; Sala, G.; Sello, G. Residual Charges on Atoms in Organic Structures: A New Method for the Identification of Conjugated Systems and the Evaluation of Atomic Charge Distribution on Them. *Tetrahedron Comput. Met.* **1989**, *2*, 93-103. (c) Baumer, L.; Sala, G.; Sello, G. Residual Charges on Atoms in Organic Structures: Molecules Containing Charged and Backdonating Atoms. *Tetrahedron Comput. Met.* **1989**, *2*, 105-118.
- (15) Magni, S. Thesis, Università degli Studi di Milano, 1998.
- (16) Pauling, L. *The Nature of the Chemical Bond*, 3rd ed.; Cornell University: Ithaca, NY, 1960.

- (17) Deprotonation reactions, which are influenced by the  $pK_a$  of the compounds, are also here considered a special case and separated from the other side reactions of point 3.
- (18) Balko, T. W.; Brinkmeyer, R. S. The Demethylation Chemistry of 3-Substituted 2-Methoxypyridines. *J. Heterocycl. Chem.* **1987**, *24*, 901–905.
- (19) Ohta, S.; Sumino, M.; Sasaki, T.; Yamagami, N. Synthesis and Reaction of 3-Hydroxyenamides. *Heterocycles* **1989**, *29*, 1455–1458.
- (20) Burke, S. D.; Fobare, W. F.; Pacofsky, G. J. Chelation Control of Enolate Geometry. Acyclic Diastereoselection via the Enolate Claisen Rearrangement. *J. Org. Chem.* **1983**, *48*, 5221–5228.
- (21) (a) REACCS, v. 8.3; MDL Information System Inc.: San Leandro, CA, 1997. (b) March, J. *Advanced Organic Chemistry*, 3rd ed.; John Wiley & Sons: New York, 1985.
- (22) (a) Wierenga, W. The Total Synthesis of Ionophores. In *The Total Synthesis of Natural Products*; ApSimon, J., Ed.; John Wiley & Sons: New York, 1981; p 297. (b) Wierenga, W.; Evans, B. R.; Woltersom, J. A. Synthesis of New, Noncyclic Ionophores Exhibiting Efficient  $Ca^{2+}$  Transport. *J. Am. Chem. Soc.* **1979**, *101*, 1334–1336.

CI990430C