IMOMM: A New Integrated *Ab Initio* + Molecular Mechanics Geometry Optimization Scheme of Equilibrium Structures and Transition States

FELIU MASERAS

Cherry L. Emerson Center for Scientific Computation and Department of Chemistry, Emory University, Atlanta, Georgia 30322 and Unitat de Química Física, Departament de Química, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Catalonia

KEIJI MOROKUMA*

Cherry L. Emerson Center for Scientific Computation and Department of Chemistry, Emory University, Atlanta, Georgia 30322

ABSTRACT

A new computational scheme integrating *ab initio* and molecular mechanics descriptions in different parts of the same molecule is presented. In contrast with previous approaches, this method is especially designed to allow the introduction of molecular mechanics corrections in full geometry optimizations concerning problems usually studied through *ab initio* calculations on model systems. The scheme proposed in this article intends to solve some of the systematic error associated with modeling through the use of molecular mechanics corrections. This method, which does not require any new parameter, evaluates explicitly the energy derivatives with respect to geometrical parameters and therefore has a straightforward application to geometry optimization. Examples of its performance on two simple cases are provided: the equilibrium geometry of cyclopropene and the energy barriers on S_N2 reactions of alkyl chloride systems. Results are in satisfactory agreement with those of full *ab initio* calculations in both cases. © 1995 by John Wiley & Sons, Inc.

---

* Author to whom all correspondence should be addressed.

Introduction

When one faces the theoretical study of a system involving very large molecules (the full computation of which is out of reach of *ab initio* methods), two major possible approaches can be taken. One of them is modeling the real molecule as a simpler one and carrying out *ab initio* calculations on the model system. Cases especially suited for this *ab initio* approach are those in which complex interatomic interactions, not necessarily well known *a priori*, have to be evaluated; these complex interactions are concentrated in a particular region of the molecule. The other major approach is force fields or molecular mechanics. This will be especially appropriate when the force field reproduces accurately all the atomic interactions relevant to the problem under study.

The chemistry of transition metal complexes has been traditionally studied through the first approach, *ab initio* study of a model system.[1,2] Bulky ligands, with large, generally organic, substituents, are commonplace. Those away from the metal atom are usually replaced in the model system by hydrogen atoms. This way, any phosphine ligand PR_3 becomes PH_3, and any cyclopentadienyl ligand η^5-C_5R_5 becomes η^5-C_5H_5. The reason why this approach works is that the metal–ligand interaction remains essentially well described, and as far as this is the leading factor influencing the properties under study, the results are satisfactory. The problem appears when the ligand–ligand interactions (corresponding roughly to what traditionally have been called steric interactions) are important. Since they are conspicuously absent in this simplified model, some other computational scheme is required for their introduction.

Molecular mechanics (MM) or molecular force fields methods have found some of their most successful applications in fields in which very large molecules are involved, like biochemistry,[3] or in which a large number of different calculations are required, like molecular dynamics.[4] But the method finds a main limitation in the relatively small set of different atoms (or types of interatomic interactions) which can be considered, and an extension requires a delicate and time-consuming parameterization process. Although recent progress has been done toward the definition of a unified force field able to encompass all atoms,[5] it seems that there is still a long way to go until MM

results can compete in terms of sheer quality with *ab initio* results. Another major shortcoming of MM is its extension to the computation of transition states; this is by no means trivial.[6]

One possible method to overcome the shortcomings of both approaches is to mix quantum mechanics and MM descriptions in different parts of the same molecule. In this way, the difficult parts of the molecule (those containing nonparameterized atoms, uncommon interactions, or atoms directly participating in the transition state of a reaction) can be computed with the expensive quantum mechanical methodology, while easy parts of the molecule, containing well-parameterized atoms and simple interactions, can be computed in a cheaper way with molecular mechanics. There are a number of examples of applications of this concept in the literature,[7–13] the most elaborate of them in the field of molecular mechanics[8,9] and connected in particular with the problem of transition state modeling, another critical problem of force field methods.[6]

Methods derived from the molecular mechanics school carry ideas which are especially suited for the kind of problems to which molecular mechanics is usually applied. For example, to keep computational cost in check, the semiempirical approach is the first choice for the quantum mechanical part.[9] In the same way, these methods are not shy in the introduction of new parameters, and no special effort is invested in reducing the number of semiempirical quantum mechanical calculations required in the course of an energy optimization. Finally, the application of different descriptions to the same parts of a single molecule is regarded largely as a subcase of a more general problem of considering different descriptions for different parts of the same system, which can be composed of different molecules without a strong chemical bond between them.

As a result, these valuable contributions from the MM field have been largely ignored whenever MM corrections have been introduced in the kind of systems usually studied with *ab initio* methods.[11–13] The procedure in all these applications is to carry out an *ab initio* geometry optimization on the model system and then an MM calculations with the real ligands on the frozen *ab initio* skeleton. While this method allows one to quantify the effect of the ligand–ligand interactions on a reaction profile, it is unable to consider changes in the same profile induced by these interactions because the *ab initio* geometry is always frozen. In other words, this approach works only

when the ligand–ligand interactions are a minor perturbation to the metal–ligand interactions, which must be dominant.

The method we propose in this article tries to fill the gap between the approaches described in the two previous paragraphs. The scheme, which is especially geared for application to *ab initio* problems, avoids introduction of new parameters and takes care of reducing the number of geometrical variables as much as possible. The loss of accuracy with respect to algorithms derived from a more pure MM point of view is compensated by its ready application to any molecule, without further parameterizations. On the other hand, if there is an error, it is deemed to be smaller than that associated with the traditional modeling schemes used in pure *ab initio* calculations. Results of previous attempts to introduce MM corrections in *ab initio* model systems[11-13] would also be improved because of the superior elaboration of the current scheme, which attempts to give the same weight to metal–ligand and ligand–ligand interactions in the course of the geometry optimization.

General Outline

At the risk of repeating work done by other authors,[8,9] we present a full description of our scheme starting from simple ideas. Differences with other methods are highlighted in selected points. This section gives a global picture of our scheme, while the more specific aspects will be addressed in detail in the next section. The general idea is to add in a consistent way the *ab initio* and MM results both at the energy and gradient levels (and, if necessary, higher order derivative levels as well). If this is achieved, an *ab initio* + MM potential hypersurface will be automatically defined, with its corresponding stationary points satisfying a zero-gradient condition.

First, some nomenclature has to be defined. A hypothetical molecule, depicted in Scheme 1, is used as example. The "real" system is $M(P(CH_3)_3)_2$, and the "model" system in the *ab initio* calculation is $M(PH_3)_2$. We divide the atoms in the model and real systems in four different sets, each of them with its corresponding coordinates:

1. *Set 1.* Atoms present as such both in the "model" *ab initio* and "real" MM part. This would be the M and P atoms in this example.
2. *Set 2.* Atoms present in the *ab initio* part but

Model system

Real system
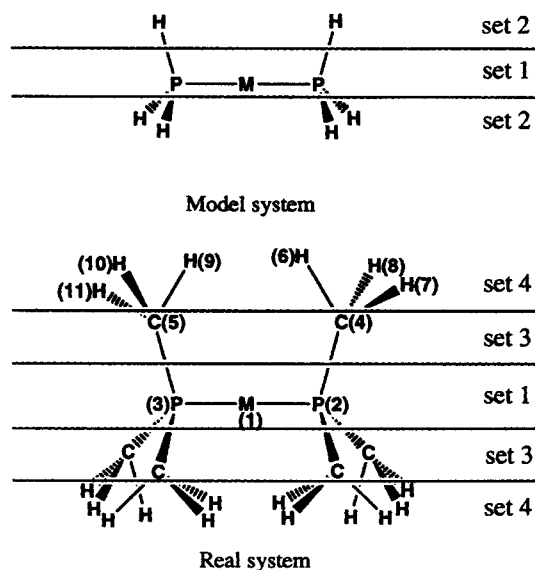
SCHEME 1.

substituted on the MM calculations by the real (different) atoms. These would be the hydrogen atoms of the *ab initio* calculation in the example. This set has been referred to as "junction dummy atoms"[8] or "link atoms"[9] by other authors.

3. *Set 3.* Atoms present as such in the MM calculations, but each atom in set 3 corresponds to a simpler atom (generally hydrogen) in set 2 in the *ab initio* calculation. The carbon atoms represent this set in the example.
4. *Set 4.* Atoms present only in the MM calculation and completely absent in the *ab initio* calculation. These would be the "real" hydrogen atoms in the example.

The coordinates of atoms in set, 3, $\vec{R}_3$, will be a function of those of atoms in sets of 1 and 2, $\vec{R}_1$ and $\vec{R}_2$:

$$\vec{R}_3 = \vec{R}_3\left(\vec{R}_1, \vec{R}_2\right) \tag{1}$$

Taking this into account, it is straightforward to write equations for energies and gradients for the combined *ab initio* + MM description as a function of those of the independent *ab initio* and MM descriptions. For the energy, we will have

$$E_{abin} = E_{abin}\left(\vec{R}_1, \vec{R}_2\right) \tag{2}$$

$$E_{MM} = E_{MM}\left(\vec{R}_1, \vec{R}_3, \vec{R}_4\right) = E_{MM}\left(\vec{R}_1, \vec{R}_2, \vec{R}_4\right) \tag{3}$$

$$E = E_{abin} + E_{MM} = E\left(\vec{R}_1, \vec{R}_2, \vec{R}_4\right) \tag{4}$$

Equation (3) does not imply that $\vec{R}_3$ is equal to $\vec{R}_2$. It just states that $\vec{R}_3$ can be expressed as a function of $\vec{R}_1$ and $\vec{R}_2$, as shown in eq. (1).

The corresponding expressions for the gradients will be

$$\frac{\partial E}{\partial \vec{R}_1} = \frac{\partial E_{abin}}{\partial \vec{R}_1} + \frac{\partial E_{MM}}{\partial \vec{R}_1} + \sum_{\vec{R}_3} \frac{\partial E_{MM}}{\partial \vec{R}_3} \cdot \frac{\partial \vec{R}_3}{\partial \vec{R}_1} \quad (5)$$

$$\frac{\partial E}{\partial \vec{R}_2} = \frac{\partial E_{abin}}{\partial \vec{R}_2} + \sum_{\vec{R}_3} \frac{\partial E_{MM}}{\partial \vec{R}_3} \cdot \frac{\partial \vec{R}_3}{\partial \vec{R}_2} \quad (6)$$

$$\frac{\partial E}{\partial \vec{R}_4} = \frac{\partial E_{MM}}{\partial \vec{R}_4} \quad (7)$$

Equations (1) to (7) formally define the problem of geometry optimization as the minimization of a function E with respect to a set of geometrical variables $\vec{R}_1$, $\vec{R}_2$, and $\vec{R}_4$; the value of the function and its gradient are readily available from independent *ab initio* and MM calculations. Any of the standard geometry optimization algorithms available in most *ab initio* and MM packages could be in principle applied to the solution of this problem.

Specific Details

DEFINITION OF $\vec{R}_3$

The general development in the previous section is valid for any definition of $\vec{R}_3$, as far as it is a function of $\vec{R}_1$ and $\vec{R}_2$ [eq. (1)]. But a proper choice of this dependency can simplify substantially eqs. (5) and (6). In particular, the scheme we propose is to choose for each pair of atoms in sets 2 and 3 the same definition in internal coordinates (a bond distance, a bond angle, and a dihedral angle), with the same value for the bond angle and the dihedral angle. In the implementation presented in this article, we simply fix the bond distances in $\vec{R}_3$ at some reasonable values and those in $\vec{R}_2$ also at some (different) reasonable values. These choices eliminate the dependence of $\vec{R}_3$ with respect to $\vec{R}_1$, eliminate the distances in $\vec{R}_3$ and $\vec{R}_2$ from the set of variables, and make dependence of angles and dihedral angles with respect to $\vec{R}_2$ trivial. In mathematical terms, we would have $\partial \vec{R}_3 / \partial \vec{R}_2 = \mathbf{I}$, where $\mathbf{I}$ would be a $2n \times 2n$ unit matrix, n being the number of atoms in set 3 (equal to the number of atoms in set 2). Alternatively, one may assume the difference in the bond distances in model $\vec{R}_2$ and real $\vec{R}_3$ to be

a constant. This alternative scheme would treat these distances still as variables but retain the relatively simple relationship between the sets: $\partial \vec{R}_3 / \partial \vec{R}_2 = \mathbf{I}$, with $\mathbf{I}$ being $3n \times 3n$ in this case. In any case, the bond distance values, which may be taken from experimental data or from some other calculation, will be obviously different for $\vec{R}_2$ and $\vec{R}_3$.

With this particular definition of coordinates, eqs. (5) and (6) become

$$\frac{\partial E}{\partial \vec{R}_1} = \frac{\partial E_{abin}}{\partial \vec{R}_1} + \frac{\partial E_{MM}}{\partial \vec{R}_1} \quad (8)$$

$$\frac{\partial E}{\partial \vec{R}_2} = \frac{\partial E_{abin}}{\partial \vec{R}_2} + \frac{\partial E_{MM}}{\partial \vec{R}_2} \quad (9)$$

This definition of coordinates is certainly the most intuitive. But it is important to remark that eqs. (8) and (9) are only valid for them and would not be valid for other definitions. In particular, they would not work in Cartesian coordinates. Concerning the practical applications, this reasoning does not require every geometry optimization to be carried out in internal coordinates. But it does require that *ab initio* and MM gradients are translated to internal coordinates before being added.

Our scheme differs sharply from other approaches at this point. For instance, Karplus and co-workers[9] let $\vec{R}_2$ optimize independently from $\vec{R}_3$, with the assumption that bond angle and dihedral angles will be similar because of the constraints coming from the relationship of atoms in sets 2 and 3 with atoms in set 1, explicit in both quantum mechanics and molecular mechanics calculations. Although this would not be the case in general, it is likely to be true in the kind of systems they consider.

DOUBLE COUNTING OF CONTRIBUTIONS

If the MM calculation is describing to some extent any of the interactions already introduced in the *ab initio* calculation, these contributions will be included twice in the total energy. Then eq. (4), which is the basis of the validity of the whole method, will be wrong. The solution we have chosen to this problem is to delete selectively the terms in the MM calculation that are assumed to be taken into account in the *ab initio* calculation. This has required some modifications in the code of the program.

The *ab initio* calculation is going to take care of interactions involving exclusively atoms in set 1, which are therefore zeroed out in our MM calculations. And it clearly ignores any contribution involving at least one atom in set 4, which therefore will have to be introduced fully in the MM part. Things are slightly more complicated for atoms in set 3. As mentioned earlier, each atom in set 3 is represented in the *ab initio* calculation by an atom of set 2. We assume that interactions involving atoms in sets 1 and 3 are properly reproduced by the *ab initio* set 1/set 2 interactions. After all (if bonds are frozen as mentioned earlier), these will be mainly bending and torsional interactions, with a minor dependence on the nature of the terminal atom. As a result, set 1/set 3 interactions are also neglected in the MM calculation. On the contrary, interactions within set 3, which are of van der Waals type and are more sensitive to the nature of the atom, are likely to be essentially ignored in the *ab initio* calculation if smaller atoms (typically hydrogen atoms) are used in set 2. Therefore, these intraset 3 interactions are computed in the MM part.

All the criteria discussed in the previous paragraph can be summarized in one general rule, with a single exception. The general rule is to zero out MM energy and gradient contributions involving exclusively atoms of sets 1 and 3 and compute explicitly contributions including any atom of set 4 (even if atoms in 1 and/or 3 are also involved). The exception is that van der Waals interactions involving exclusively atoms in set 3 are included. Use of the example in Scheme 1 can further clarify the procedure. Following the atom numbering in Scheme 1, MM terms concerning, for example, energies and gradients concerning the following interactions would be computed: H(6)—H(9) (within set 4), P(2)—H(8) (between sets 1 and 4), C(4)—H(6), H(6)—C(4)—H(7) (the last two involving sets 3 and 4), P(2)—C(4)—H(6), M(1)—P(2)—C(4)—H(6) (the last two involving the three sets 1, 3, and 4), and C(4)—C(5) (a van der Waals interaction within set 3). Gradient and energy contributions coming from the following MM terms, for example, will be set to zero: M(1)—P(2) (within set 1), P(3)—C(4), M(1)—P(3)—C(5), and P(3)—M(1)—P(2)—C(4) (the last three involving sets 1 and 3). This method to estimate the total energy through cancellation of the "double-counted" contributions may not be the most accurate possible, but the bias it eventually introduces is hopefully outweighed by its computational efficiency.

REDUCTION OF THE NUMBER OF GEOMETRICAL VARIABLES

The last point we want to mention about the implementation of the method does not concern its accuracy but its computational efficiency. Most practical applications will have the model system much smaller than the real system. This will mean that most of the geometrical variables in the system are going to be in the $\vec{R}_4$ set. If one were to optimize directly the total energy E as a function of the whole set of variables, many steps would be required exclusively for the optimization of this $\vec{R}_4$ set. Since these are the geometrical variables present only in the MM calculation, it would be highly desirable to find a procedure to optimize them separately.

Independent optimization of geometrical variables in the $\vec{R}_4$ set can be accomplished through the definition of E'_{MM}, which is same as E_{MM} except that it is a function of only $\vec{R}_1$ and $\vec{R}_2$ and is optimized with respect to variables in $\vec{R}_4$. That is,

$$E'_{MM}\left(\vec{R}_1, \vec{R}_2\right) = E_{MM}\left(\vec{R}_1, \vec{R}_2, \vec{R}_4^0\right) \quad (10)$$

$$\left.\frac{\partial E_{MM}}{\partial \vec{R}_4}\right|_{\vec{R}_4 = \vec{R}_4^0} = 0 \quad (11)$$

A total energy E' can be readily defined from E'_{MM}:

$$E' = E'\left(\vec{R}_1, \vec{R}_2\right) = E_{abin}\left(\vec{R}_1, \vec{R}_2\right) + E'_{MM}\left(\vec{R}_1, \vec{R}_2\right) \quad (12)$$

Things are a little more complicated for the gradient, but only from a formal point of view. For $\vec{R}_1$, there is

$$\frac{\partial E'}{\partial \vec{R}_1} = \frac{\partial E_{abin}}{\partial \vec{R}_1} + \frac{\partial E'_{MM}}{\partial \vec{R}_1} + \sum_{\vec{R}_4}\left.\left(\frac{\partial E_{MM}}{\partial \vec{R}_4} \cdot \frac{\partial \vec{R}_4}{\partial \vec{R}_1}\right)\right|_{\vec{R}_4 = \vec{R}_4^0} \quad (13)$$

where the last term of eq. (13) would account for the change of the optimal $\vec{R}_4$ value $\vec{R}_4^0$ associated with changes in $\vec{R}_1$. But this term is zero because of the definition of E'_{MM} [eq. (11)]. As a result,

$$\frac{\partial E'}{\partial \vec{R}_1} = \frac{\partial E_{abin}}{\partial \vec{R}_1} + \frac{\partial E'_{MM}}{\partial \vec{R}_1} \quad (14)$$

The same reasoning is also valid for $\vec{R}_2$:

$$\frac{\partial E'}{\partial \vec{R}_2} = \frac{\partial E_{abin}}{\partial \vec{R}_2} + \frac{\partial E'_{MM}}{\partial \vec{R}_2} \qquad (15)$$

When $\vec{R}_1$, $\vec{R}_2$ change, $\vec{R}_4$ changes. Equations (12), (14), and (15) define the problem as the optimization of a function depending only of variables in $\vec{R}_1$ and $\vec{R}_2$ sets. Variables in $\vec{R}_4$ are "hidden" in the definition of E'_{MM} [eqs. (10) and (11)], which requires also a geometry optimization with respect to these variables. Most important, optimization with respect to $\vec{R}_4$ only has to be carried out in the MM part and can be completely neglected in the *ab initio* part. Therefore, the variables considered explicitly in the *ab initio* optimization cycle are only those in the model system ($\vec{R}_1$ and $\vec{R}_2$), and the number of *ab initio* steps in the *ab initio* + MM geometry optimization should be of the same order as that in the model system. Characteristically, this aspect was not addressed in combined quantum mechanics + molecular mechanics schemes derived from MM models in conjunction with semiempirical MO methods.[8,9]

Computational Algorithm

All the ideas presented earlier are put into work in the following computational algorithm for geometry optimization:

1. Input a set of $\vec{R}_1$, $\vec{R}_2$, and $\vec{R}_4$ variables.
2. Compute with the *ab initio* program E_{abin}, $\partial E_{abin}/\partial \vec{R}_1$, and $\partial E_{abin}/\partial \vec{R}_2$.
3. Carry out an MM optimization with respect to $\vec{R}_4$ in a frozen $\vec{R}_1$, $\vec{R}_3$ environment. This will give E'_{MM}, $\partial E'_{MM}/\partial \vec{R}_1$, and $\partial E'_{MM}/\partial \vec{R}_2$ as well as a new value for $\vec{R}_4$. (Of course, steps 2 and 3 are independent and therefore can be carried out in the reverse order.)
4. Calculate E', $\partial E'/\partial \vec{R}_1$, and $\partial E'/\partial \vec{R}_2$ as a sum of *ab initio* and MM contributions.
5. Check for convergence. If not achieved, generate new values for $\vec{R}_1$ and $\vec{R}_2$, and go back to step 2.

Some Test Calculations

To check the validity of the method, we carry out some test calculations on simple systems. We have built a program by taking the *ab initio* part

from the GAUSSIAN 92 program[14] and the MM part from the MM2 program.[15] Both source codes have to be modified, slightly in the case of the *ab initio* program and more substantially in the molecular mechanics part.

The resulting program has a single input file which contains the standard input files for both the GAUSSIAN 92 and MM2 programs, with small modifications in each case. The only changes in the GAUSSIAN input part concern the introduction of some nonstandard input options (IOPs in this particular program language). For the MM2 input part, some more input is necessary. It includes the definition of each of the atomic sets. Information about which are the frozen bond distances between set 1 and set 3 is also included in the MM input part. The program is executed as a normal GAUSSIAN 92 job, with the MM2 program called as a subroutine in some selected places. More details about the execution of the program are provided later in this article with the first example of its application.

Although the method is intended to be applied to large molecules, probably with transition metal atoms in them for our interest, the tests are carried out in rather small organic molecules. In this way, results can be checked against full *ab initio* calculations, which would not be affordable in larger molecules.

GEOMETRY OF CYCLOPENTENE

The first problem we consider is the determination of the geometry of cyclopentene using ethylene as the model system. *Ab initio* calculations are at the restricted Hartree-Fock (RHF) level with the minimal STO-3G basis set. Standard parameters are used in the MM2 program. The frozen distances between atoms in sets 1 and 3 are C—C distances fixed to 1.523 Å, and the two corresponding C—H distances of the *ab initio* calculation are fixed to 1.08 Å.

Results are presented in Figure 1. Agreement between the fully *ab initio* optimized geometry and the present *ab initio* + MM optimized geometry is excellent. The C(2)—C(1)—C(5) angle, probably the most critical parameter, is 112.96° in the *ab initio* + MM calculation, a value which is close to the 112.81° of the full *ab initio* optimization and very different from the 122.18° associated with the optimization of free ethylene with the same basis set. The largest difference is in the C(1)—C(5)—C(2)—C(11) dihedral angle, which measures the planarity of the ring. Changes in this
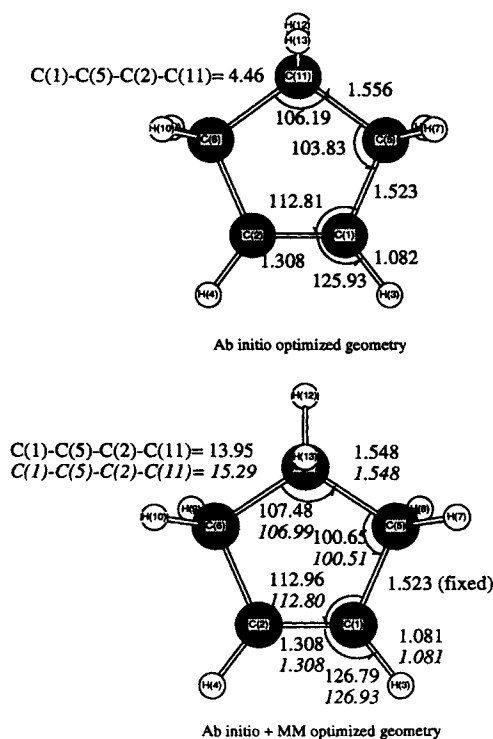
FIGURE 1. Geometry of cyclopentene, optimized with a full *ab initio* RHT STO-3G method (top) and with an *ab initio* + MM method (bottom). In the bottom figure, geometrical values obtained with an MM sp^2 definition of C(1) and C(2) are presented in roman, while those obtained with an sp^3 description are shown in italic.

angle are expected to have a small effect in the energy, and this deviation does not seem critical. Figure 1 also shows the results of an *ab initio* + MM optimization in which the two sp^2 carbons are defined as sp^3 to simulate a hypothetical case in which we did not know about the specific bonding properties of olefinic carbons. Results are very similar. Since bonding and bending contributions centered in these particular atoms (which are in set 1) are zeroed out in our scheme, and the effect of their precise nature in the final result is indeed expected to be small.

Although this particular problem of geometry optimization can be adequately solved through "pure" *ab initio* or "pure" MM methods, this example also shows the power of the combination of both methods. The combined method clearly beats the *ab initio* method in terms of computer time. We are computing cyclopentene (33 basis functions) at the price of ethylene (14 basis functions). The advantage with respect to the MM method is that the combined method does not require any previous knowledge about the nature

of sp^2 carbons. Although this is not relevant in this particular case, it will become a major advantage when more exotic atoms are to be considered.

Once this first example has been discussed, it may be used as an illustration of the particular application we have done with GAUSSIAN 92 and MM2 of the algorithm described in the previous section. In particular, the program is used as a GAUSSIAN job with the following particularities:

1. Input. The geometry of ethylene is read in link L101, as it corresponds to a normal *ab initio* calculation. The C—H bond distance is fixed to 1.08 Å in two of the cases. This same link L101 calls the MM2 program, which at this point reads the geometry of cyclopentene, together with other input MM information. The MM information is saved in the checkpoint file for later use.

2. Computation of *ab initio* contributions. The *ab initio* program follows the normal route and computes the energy and gradient of ethylene.

3. Computation of MM contributions. Once all *ab initio* calculations on this geometry are performed, the MM2 program is called again, now from link L716. Most of its input information is read from the checkpoint file. This notwithstanding, the *ab initio* geometry from ethylene, which defines $\vec{R}_1$ and $\vec{R}_3$, is also transferred to the MM program. $\vec{R}_3$, the coordinates of the C(5) and C(6) carbon atoms, is defined by placing these atoms with the same bond and dihedral angle as the corresponding *ab initio* hydrogen atoms, and with the C(1)—C(5) and C(2)—C(6) distances fixed to 1.523 Å. The MM program freezes $\vec{R}_1$ and $\vec{R}_3$ and optimizes $\vec{R}_4$, the geometry of the H(7), H(8), H(9), C(10), H(11), and H(12) atoms. This gives a total MM energy, together with a set of gradients for $\vec{R}_1$ and $\vec{R}_3$. The gradients are translated to gradients on $\vec{R}_1$ and $\vec{R}_2$ with the simple scheme described earlier.

4. Addition to *ab initio* and MM contributions. The energies and gradients coming from steps 2 and 3 are added in this same link L716.

5. Check for convergence. The GAUSSIAN 92 job continues its route and checks for convergence in link L103, which is left unmodified. If convergence is not achieved, a new set of coordinates is generated and the program goes back to step 2.

STERIC EFFECTS IN S$_N$2 REACTIONS

The influence of steric effects on the energy barrier of the S$_N$2 reaction of different alkyl chloride molecules with chloride anions is evaluated with the method proposed in this article. In particular, the reaction considered is that presented in Scheme 2, with R^1, R^2, and R^3 being H or CH$_3$.

The case where R^1, R^2, and R^3 are all hydrogen atoms constitutes the *ab initio* model. The choice of this model system, instead of the formally simpler Cl$^-$ + CH$_3$Cl, is due to the possible effect of changes in the C$_1$—C$_2$ distance, a parameter that would be fixed in the simpler model. Full *ab initio* results published by Jensen[16] are used to assess the quality of our results. Consequently, the *ab initio* method applied here is the same as that used for the geometry optimization in that article: full MP2 calculations with a 6-31G* basis set. Standard MM2 parameters are used in the molecular mechanics calculation.[15] The frozen distances for bonds between different sets of atoms are 1.08 Å for C—H (*ab initio* calculations) and 1.523 Å for C—C (MM calculations).

The combination of R^1, R^2, and R^3 being H or CH$_3$ gives rise to four different possible reactions,

corresponding as reactant to ethyl chloride, propyl chloride, isobutyl chloride, and neopentyl chloride. The geometries of the reactant and of the transition state for each one of the four reactions were computed with our method, the ethyl case being the *ab initio* model. The nature of the transition state as such (one single negative eigenvalue in the Hessian) was checked through numerical *ab initio* + MM computation of second derivatives, though the combination of analytical *ab initio* + numerical MM second derivatives would also be possible. The computed energy barriers with respect to the reactants were 8.4, 8.1, 10.2, and 14.3 kcal/mol for the ethyl, propyl, isobutyl, and neopentyl systems, respectively. The corresponding full *ab initio* values[16] are 8.4, 7.1, 9.1, and 10.4 kcal/mol. Although, as expected, agreement is not perfect, the ordering is well reproduced, the lowest barrier corresponding to propyl and the highest to neopentyl. There are a number of possible sources of error which can account for this discrepancy. Among them, limitations in the MM2 force field itself and the neglect of charge contributions in the MM calculations seem the most serious. But problems can also arise from a poor description in the *ab initio* de-
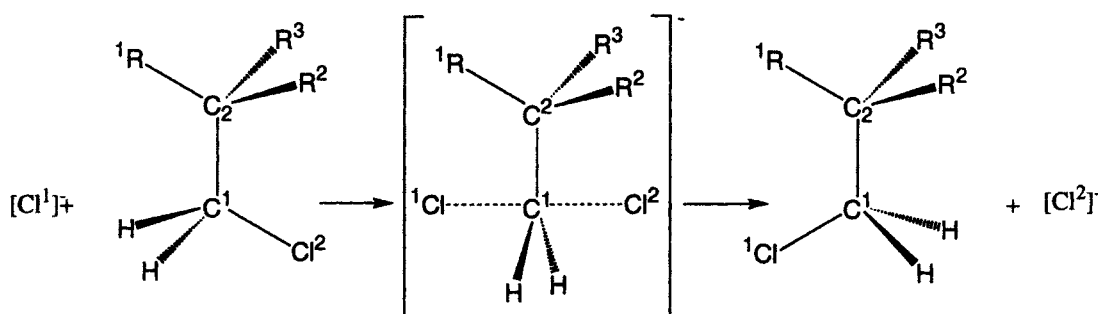
SCHEME 2.

TABLE I.
Selected Geometrical Parameters of the Transition State Structures for the S$_N$2 Reaction between Cl$^-$ and CH$_2$Cl—C(R^1)(R^2)(R^3), as Obtained in *Ab Initio* + MM Geometry Optimizations.

System	C^1—Cl^1	C^1—Cl^2	C^2—C^1—Cl^1	C^2—C^1—Cl^2	C^1—Cl^2—C^2—R^1
Ethyl	(2.342)	(2.369)	(100.0)	(94.6)	(0.0)
Propyl	2.362	2.362	97.7	97.7	89.1
	(2.356)	(2.356)	(97.2)	(97.2)	(~ 90.0)
Isobutyl	2.344	2.410	99.8	100.5	0.0
	(2.355)	(2.391)	(99.9)	(99.6)	(0.0)
Neopentyl	2.383	2.417	107.2	101.2	0.0
	(2.405)	(2.412)	(109.1)	(101.3)	(0.0)

Geometrical definitions follow the labeling in Scheme 2. Values in parentheses from full *ab initio* optimizations are also included for comparison. R^1 corresponds to Me in the case of propyl and to H in the case of isobutyl.

scription of the van der Waals interaction, and substantial changes associated with a basis set improvement point to this direction.[16] In any case, our energetics are compared with those of MP2/6-31G* quality because this is the level at the which the full *ab initio* optimization was carried out. The fact that this is not the best possible method for this particular chemical problem, as demonstrated by single-point calculations at levels up to MP4/6-311 + G(2*d*) in the aforementioned article,[16] is not relevant to the comparison.

The geometries, summarized in Table I, exhibit also a reasonable agreement with full *ab initio* values published by Jensen.[16] In particular, global variations in the shape of the transition state for different cases are well reproduced. From the point of view of symmetry, there are two possibilities for the transition state of this reaction, both of them with a C_s symmetry. The first possibility, shown in Scheme 2, has C^1 and the two chlorine atoms in the symmetry plane, with R^1 contained also in this plane (dihedral angle $C^1—Cl^1—C^2—R^1 = 0°$). This is the case for the ethyl, isobutyl, and neopentyl systems. The propyl system responds to another pattern, with the symmetry plane perpendicular to the $C^1—Cl^1$ direction (dihedral angle $C^1—Cl^1—C^2—R^1$ approximately 90°). In the second case, the $C^1—Cl^1$ and $C^1—Cl^2$ must be equal, and they are indeed found to be slightly different in the first case.

In summary, although in this particular case full *ab initio* results are not perfectly reproduced, our *ab initio* + MM approach is accurate enough to give an appropriate qualitative description and reproduce accurately the trends. This is especially satisfactory in this kind of problem, which is not affordable for pure MM methods because it requires the computation of a transition state and is fairly expensive for pure *ab initio* methods because of its size.

Conclusions

A new algorithm has been developed for the introduction of molecular mechanics corrections in *ab initio* geometry optimizations. The method, although based in well-known concepts, is especially designed for *ab initio* problems traditionally dealing with *ab initio* quantum calculations on model systems. Because of this, the introduction of additional parameters is avoided. Special care is also taken to reduce to the minimum the number of

necessary *ab initio* calculations through the separation of geometrical variables in two sets, with the smaller of them being involved directly in the *ab initio* optimization. Its limitations are offset by its straightforward application to a large number of molecules and quantum mechanical approaches. The formal simplicity of this scheme, as well as its unnecesity of additional parameters, makes it application to any force field or *ab initio* method straightforward. For instance, MM3 and UFF (unified force field) can be considered as force fields, while DFT (density functional theory) and highly correlated methods are also available for the *ab initio* part. As for the kind of geometries to be computed, equilibrium geometries and transition states appear readily accesible.

Test applications of this scheme on two different organic systems have also been presented, and the results are in agreement with full *ab initio* results in both cases. Good results in these two particular examples, in which the energy differences involved are small, prompt optimism about the application of the method to larger systems, in which the range of energies involved will be much larger. Calculations in such larger systems are currently under way. Application to a large set of chemical problems involving transition metal systems and biomolecules is also promising.

Acknowledgments

References

1. A. Veillard, *Chem. Rev.*, **91**, 743 (1991).
2. N. Koga and K. Morokuma, *Chem. Rev.*, **91**, 823 (1991).
3. J. Åqvist and A. Warshel, *Chem. Rev.*, **93**, 2523 (1993).
4. P. Kollman, *Chem. Rev.*, **93**, 2395 (1993).
5. A. K. Rappé, C. J. Casewit, K. S. Colwell, and W. A. Goddard III, *J. Am. Chem. Soc.*, **114**, 10024 (1992).
6. J. E. Eksterowicz and K. N. Houk, *Chem. Rev.*, **93**, 2439 (1993).
7. A. Warshel and M. Karplus, *J. Am. Chem. Soc.*, **94**, 5612 (1972).
8. U. C. Singh and P. A. Kollman, *J. Comp. Chem.*, **7**, 718 (1986).
9. M. J. Field, P. A. Bash, and M. Karplus, *J. Comp. Chem.*, **11**, 700 (1990).
10. A. Warshel, *Computer Modeling of Chemical Reactions in Enzymes and Solutions*, John Wiley & Sons, New York, 1991.

11. H. Kawamura-Kuribayashi, N. Koga, and K. Morokuma, *J. Am. Chem. Soc.*, **114**, 8687 (1992).

12. J. Eckert, G. J. Kubas, J. H. Hall, P. J. Hay, and C. M. Boyle, *J. Am. Chem. Soc.*, **112**, 2324 (1990).

13. F. Maseras, N. Koga, and K. Morokuma, *Organometallics*, **13**, 4008 (1994).

14. M. J. Frisch, G. W. Trucks, H. B. Schlegel, P. M. W. Gill, B. G. Johnson, M. W. Wong, J. B. Foresman, M. A. Robb, M. Head-Gordon, E. S. Replogle, R. Gomperts, J. L. Andres, K. Raghavachari, J. S. Binkley, C. Gonzalez, R. L. Martin, D. J. Fox, D. J. Defrees, J. Baker, J. J. P. Stewart and J. A. Pople, Gaussian 92/DFT, Gaussian, Inc., Pittsburgh, PA, 1993.

15. (a) N. L. Allinger and Y. Yuh, *QCPE*, **12** 395 (1980); (b) U. Burkert and N. L. Allinger, *Molecular Mechanics*, ACS, Washington, DC, 1982.

16. F. Jensen, *Chem. Phys. Lett.*, **196**, 368 (1992).