Multicoordinate Driven Method for Approximating Enzymatic Reaction Paths: Automatic Definition of the Reaction Coordinate Using a Subset of Chemical Coordinates†

Imre Berente*,‡ and Gábor Náray-Szabó‡,§

Department of Theoretical Chemistry, Eötvös Loránd University, and Protein Modeling Group, Hungarian Academy of Sciences - Eötvös Loránd University, H-1117 Budapest, Pázmány Péter st. 1A, Hungary

Received: July 26, 2005; In Final Form: October 17, 2005

We present a generalization of the reaction coordinate driven method to find reaction paths and transition states for complicated chemical processes, especially enzymatic reactions. The method is based on the definition of a subset of chemical coordinates; it is simple, robust, and suitable to calculate one or more alternative pathways, intermediate minima, and transition-state geometries. Though the results are approximate and the computational cost is relatively high, the method works for large systems, where others often fail. It also works when a certain reaction path competes with others having a lower energy barrier. Accordingly, the procedure is appropriate to test hypothetical reaction mechanisms for complicated systems and provides good initial guesses for more accurate methods. We present tests on a number of simple reactions and on several complicated chemical transformations and compare the results with those obtained by other methods. Calculation of the reaction path for the enzymatic hydrolysis of the substrate by dUTPase for an active-site model with 85 atoms, including several loosely bound water molecules, indicates that the method is feasible for the study of enzyme mechanisms.

Introduction

The multidimensional potential energy surface (PES) of a molecule contains important information on geometries and relative energies of its locally stable structures, as well as reaction paths between these, which are simplest characterized by transition states and internal minima.[1] The transition states (TS) are saddle points on the PES, so they have all gradients zero, just like minima, but just one of their force constants is negative. This means that they are minima for all coordinates except for one, the reaction coordinate. While locating minima is straightforward (move down with steepest gradient), the location of a saddle point is not simple, since there are an infinite number of different paths to move up. Theoretically, the eigenvector belonging to the only negative eigenvalue would be the solution, but the convergence region where the Hessian has the proper structure is small. For large systems, like enzymes, there are several "soft" coordinates, like the rotation of a methyl group at the end of some side chain or the displacement of loosely bound water molecules. Though these coordinates are irrelevant for the chemist, they also may provide mathematically proper saddle points on the PES; therefore, an effective method must be able to distinguish the true reaction path from the irrelevant ones.

The simplest algorithm to determine a reaction path is to perform a relaxed PES scan using an internal coordinate (or a fixed combination of coordinates), which overlaps well with the reaction coordinate.[2] During the scan, we change the value of this "reaction coordinate" stepwise and optimize the geometry in every step keeping the active coordinate frozen. This procedure is called the reaction coordinate driven (RCD) method,[3] which has a serious drawback, because it fails if the overlap between the chosen and real reaction coordinates is not large enough. For complicated multistep reactions, finding an acceptable reaction coordinate is often impossible. There is also another problem with the RCD method extensively analyzed by Williams and Maggiora,[4] since they found that it does not work for every PES. If the saddle point is at the end of a valley, then it can be found by the RCD method; however, if it is located elsewhere, the RCD algorithm cannot bring the system out of the valley (Figure 1). If we start from the reactant, the calculated path is in the reactant valley (R−B line). Point B is an inflection of the surface; thus, the valley ends, and the optimizer finds point C next. So, the calculated reaction path will be R−B−C−P. The opposite direction will result in the P−D-A−R path, which is also far from the desired one. Since

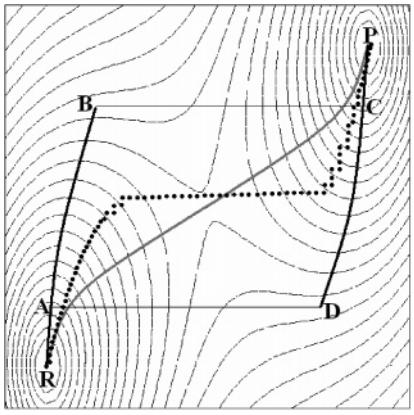

Figure 1. A two-dimensional PES where the RCD method fails. The method yields either R−B−C−P or P−D-A−R, while the MEP between the reactants (R) and products (P) is represented by the gray line. Dots have been calculated by the method described in this paper (for more details, see the Results and Discussion section.

† Part of the special issue "Donald G. Truhlar Festschrift".

* Corresponding author. E-mail: imre.berente@gmail.com.
‡ Department of Theoretical Chemistry.
§ Protein Modeling Group, Hungarian Academy of Sciences.

10.1021/jp054116z CCC: $33.50
Published on Web 11/23/2005

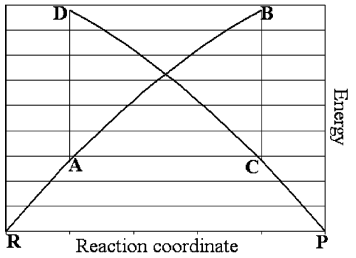

Figure 2. Plot of the energy vs the RCD path for the PES of Figure 1.

no further points can be located between B and C, a discontinuity appears on the energy plot (Figure 2), and the energy of the true saddle point cannot be found.

This problem also exists in higher dimensions; therefore, if we do not have information on the PES in advance, the RCD method may lead to useless results.

In the following, we will give a brief overview on the various accurate methods applied to the calculation of transition states and reaction paths; for a more detailed review, see refs 5 and 6. They can be divided in two groups, the surface walking and the two-end methods.

The surface walking methods need an initial guess and look for the TS as a saddle point on the PES. The eigenvector-following (EF) methods[7−9] are able to find minima and first- or higher-order saddle points. When used for searching a transition state, they need an initial guess for the geometry, then the Hessian is constructed analytically, numerically, or by approximation, and finally, the appropriate eigenvectors are generated. If the Hessian has the correct number of negative eigenvalues, then a rational function optimization (RFO)[10] step is made. If the energy in the new point differs significantly from the previous estimate or the new Hessian has a wrong structure, the trust radius (the maximum size of the step) should be reduced. The main drawback of this method is that the initial guess should be in the near vicinity of the TS where the curvature of the PES is similar to that at the TS. If we want to calculate a reaction path containing intermediate structures, which are represented by local minima on the PES, we need initial geometry guesses for each transition state. Since we use the eigenvectors of the Hessian, this needs to be accurate, and other problems may also arise.

The Reduced PES Path Search[11] method solves the problem of the soft coordinates by partitioning the coordinates into two subsets. The coordinates which are important for the reaction form the active subset; the rest form the nonactive one. The TS is searched on the reduced PES, which has as many dimensions as active coordinates. The rest of the coordinates is optimized. In this subset, an "image quadratic function" is defined and minimized in order to find the next point on the path, which was found[12] to be one of the reduced gradient paths.

The two-end methods use information from both the reactants and the products of a reaction and locate the TS between them. Peng and Schlegel[13] proposed a two-end TS search algorithm, called the synchronous transit-guided quasi-Newton (STQN) method, and implemented in the *Gaussian 03* program package[14] under the QST keyword. This method is fast, stable, and does not need analytical Hessians but uses updated approximations; however, it has some drawbacks. First, it can only find one saddle point between two neighboring deep minima and fails to converge if a shallow intermediate minimum is located between them. Second, it highly depends on the difference between the initial guess used for the synchronous transit method[15] and the real TS. If the guess is wrong, than the STQN method converges slowly or even fails to converge.

A path optimization algorithm[17] was also implemented to *Gaussian 03*.[14] In this, the path is represented by several geometries. During refinement, these geometries are changed to reduce the force components perpendicular to the path, and then, the points are redistributed, keeping a similar distance from each other and therefore serving as milestones on the path. With large systems, we face another problem, namely, during the redistribution step, the distance of the points may depend on the chemically irrelevant soft coordinates. This problem can be solved[18] by weighting the coordinates by their chemical importance, so a coordinate change referring to a forming or breaking bond counts much more than the rotation of some loosely bound water molecule.

While the Ayala−Schlegel path optimization method performs microiterations to combine the path relaxation with the point redistribution, the nudged elastic band (NEB) method[19−21] uses a quadratic "spring potential" to prevent the path points from falling back to the ending minima instead of the appropriate point of the reaction path. However, overly strong spring forces can cut down corners of the path. To avoid such kinks (artifacts) and the points falling down to low-energy regions, the tangent of the path is estimated, and the real forces parallel as well as the spring forces perpendicular to the path are discarded. This method can calculate pathways with internal minima and multiple saddle points; it needs initial guesses, which are usually created by linear interpolation, and converges slowly if these are inappropriate. For large systems, it may fail to converge[21] even if the spring force is appropriate, because the root-mean-square (rms) difference referring to two adjacent points can also decrease if changing irrelevant soft coordinates. This problem can be solved[21] if the spring forces relate only to important coordinates.

The string method[23] uses a similar concept as the Ayala−Schlegel path optimization method, but a different algorithm is used to obtain the points of the reaction path. The growing string (GS) method[24] is a further augmentation. It makes predictor and corrector steps by growing two strings from the endpoints with a user set step length until these ends meet, forming a string. Unlike all other methods, this one does not need an initial guess for the geometry; therefore, it is robust, but due to its complexity, it converges very slowly.

The reduced gradient following (RGF)[25,26] method uses the fact that in minima and transition states the gradients are zero. It seeks lines on the PES between such neighboring points with all gradients zero except for the one which serves as the driving coordinate. The RGF projector is found to be applicable[27] for string methods, too.

Molecular dynamics can also be used to examine reactions, but since bond breaking and bond forming cannot be studied by molecular mechanics, we must use ab initio molecular dynamics, which is extremely expensive. There are methods available[28] which concentrate on MD samples between reactant and product structures; therefore, they are much more cost-effective, but still very expensive and should be used only if we need more than just a path, like reaction rates or the ratio of competitive paths.

Method

We propose a new and simple method, which needs optimized reactant and product structures and a rough guess on the mechanism in order to calculate reaction paths. It is quite robust, but still converges at an acceptable rate for a number of

elementary chemical reactions. A special advantage is allowing the user to make a calculation for the hypothetical mechanism, even if it does not yield the minimum energy path or if it involves the synchronous movement of several atoms. After describing the algorithm, we will present some test calculations and compare our results with those obtained by other methods.

Our reaction path searching algorithm is based on the old RCD algorithm outlined in the Introduction; this is why we call it the multicoordinate driven (MCD) method. Like the other two-end methods, it needs both reactant and product structures. First, we need to select internal coordinates, which should change monotonically during the reaction; their number is not limited, unlike in RCD. Let us now define these coordinates for the $F^- + H_3C-F^* \rightarrow F-CH_3 + F^{*-}$ reaction. These are the F−C distance, changing from 2.6 to 1.4 Å, and the C−F* distance, changing from 1.4 to 2.6 Å. Now, locating the minimum energy pathway means finding the lowest-energy way to change the active coordinates from their initial values to the final ones.

At this point, the MCD method follows the reaction path philosophy of the RCD method: steepest descent from the TS to the minimum and shallowest ascent from the minimum to the TS. Three problems are also inherited from the RCD method. At first, the forward path differs from the backward path. This has no physical meaning: it is a natural error of the method. If a more sophisticated path is needed, IRC[16] can provide it using the found TSs. Second, since it applies constraints to the step, turnings are not allowed, so it cuts down the real turns of the path. If the turns are small, than the TS candidate found is not far from the real TS, so it is an acceptable guess. However, where the true reaction path has large turns, our method does not work, just like RCD. Finally, as in the problem mentioned in Figures 1−2, RCD gets stuck in a valley. If the valley is just an artifact, created by projection of the PES into a subspace, MCD solves the problem, since the subspace can be increased. If the valley is real in the full dimensionality of the system, like in Figure 1, the MCD path will also differ from the real; however, it will not become discontinuous and provide points with different energy. In most cases, the highest energy point is a good initial guess, but in rare cases, it can be too far from the real TS, which provides a failure in the subsequent EF calculations. In the case of the problem in Figure 1, the MCD calculated points differ from the right path and even miss the true TS, but are much closer to it than that provided by the RCD method and are within the convergence region of the TS, so it provides a good initial guess.

The energy change between two geometries labeled *a* and *b* is given by eq 1, where $f_i(\underline{x})$ is the force on the *i*th coordinate at its value of $\underline{x}$, an *N* dimensional vector, and summation runs over all independent coordinates

$$\Delta E = \sum_{i=1}^{N} \int_{x_{ia}}^{x_{ib}} - f_i(\underline{x}) \, \mathrm{d}x_i \qquad (1)$$

Equation 1 cannot be implemented, since we cannot follow a coordinate change, which we cannot predict at least approximately. For example, in the $F^- + H_3C-F^* \rightarrow F-CH_3 + F^{*-}$ reaction, the H−C−H angle is 110.9° in both the reactants and the products, but it is 120° in the transition state. This cannot be predicted without knowing the reaction path accurately. Accordingly, we follow only predictable coordinates which are selected by exploiting the user's knowledge of the reaction and chemical intuition. This also allows one to use the popular redundant coordinate systems for the geometry optimizer; since only "active" coordinates have to be linearly independent, the nonactive ones may be redundant. To avoid an energy change due to dropping the rest of the coordinates, we constantly optimize their value, thus keeping the forces along them practically zero. Accordingly, the respective term in eq 1 will vanish, and we obtain the following expression

$$\Delta E = \sum_{i=1}^{N\text{active}} \int_{x_{ia}}^{x_{ib}} - f_i(\underline{x}) \, \mathrm{d}x_i \qquad (2)$$

Integration in eq 2 cannot be performed analytically, so we have to apply a numerical procedure. The nonactive coordinates will vanish only if we keep them continuously optimized along the reaction path. This would mean that the numerical integration must be performed with an infinitely small step size. In case of finite step sizes, the error is acceptable if the change of the nonactive coordinates or the forces along them are small. This can be assured by setting every important coordinate active.

The forces along the active coordinates change continuously, too, leading to errors in the numerical integration with finite step sizes. To reduce this error, instead of using only the force $f_i(x)$ between points *a* and *b*, we use a linearly changing gradient approximated by the Hessian in point *a*, as given in eq 3, where $\Delta x_i = x_i(b) - x_i(a)$ and H_{ija} is the (ij) element of the Hessian in point *a*. This approximation is acceptable if the step size does not exceed the size of the quadratic region of *a*.

$$\Delta E \approx \sum_{i}^{N\text{active}} \left(-f_{ia} + \frac{1}{2} \sum_{j}^{N\text{active}} H_{ija} \Delta x_j \right) \Delta x_i \qquad (3)$$

The MCD method calculates Δx_i values for all active coordinates in every point, by minimizing ΔE of eq 3 with two constraints. The sum of Δx_i values (which is not equal to the length of the vector) is constrained to the step size. The other constraint is that no Δx_i value can change backward, i.e., in the product → reactant direction. This can be avoided if we use internal coordinates, which change monotonically during the reaction. The minimization is carried out in a simple way: change one Δx_i and accept the change if ΔE decreases, then repeat the procedure until convergence. We are aware that more sophisticated minimization methods exist, but the cost of the calculation of Δx_i is irrelevant compared to self-consistent field (SCF) or force calculations, so an improvement would not reduce computational costs. The active coordinates have to be modified by the Δx_i values, and then, the geometry of the new structure should be optimized keeping the active coordinates fixed, just like in the case of the RCD method. This way we can construct the reaction path point by point until all coordinates reach their final values set by the user. If the tentative mechanism was correct, the last point will correspond to the product structure.

Our method cannot find the exact transition states; it provides only initial guesses. It may lead to discontinuous paths, so if not only the stationary points are needed, the path has to be recalculated by the IRC method[16] using the exact stationary points. It allows the user to find a subset of coordinates which change monotonically during the reaction; therefore, a preliminary guess for the path is needed. The STQN,[13] NEB,[19−21] string,[23,24] reduced PES search,[11] and RGF[25,26] methods are all able to find exact stationary points, as well as the path. Some of them even work as a black box; only a good initial guess for the reactant and product structures is needed, and the rest is done automatically. The reason we propose this method for complicated reactions is its simplicity and its power to provide

good initial guesses for further calculations. In the absence of such guesses, the presently available methods may fail for most complicated reactions; they do not converge or provide false results for the saddle point. All geometry optimizers, transition state, and reaction path calculation methods need the Hessian ($H_{ij} = \partial^2 E/\partial x_i \partial x_j$). However, calculating it explicitly is extremely expensive in terms of both CPU time and storage space. The alternative of the analytical or accurate numerical calculation is to start from some initial guess[29,30] and update it in every geometry optimization step. Several methods exist to do this, like BFGS,[31] which is most popular for minimization, and the algorithms by Bofill[32,33] for transition states. Though these methods provide good approximations for the Hessian, two fundamental problems cannot be circumvented. First, if the initial Hessian was inaccurate, the updates involve a lot of computational work. Second, these methods use a number of N gradients and N displacements to update the Hessian having $N(N + 1)/2$ independent elements; therefore, the accuracy will decrease linearly with the size of the system, N. For geometry optimizations, this only increases the computational cost, since one can make small steepest descent steps with the gradients getting gradually closer to the minimum and providing information for the updated Hessian. However, the vector pointing toward the TS can only be derived from the Hessian; therefore, a poor estimate can even lead to the failure of the search. Our MCD method will also work with low-quality Hessians, since in eq 3, it is multiplied by the second power of the Δx_i steps, which means that a smaller step size decreases its weight. A smaller step size also increases the number of force calculations, providing more information for the updating of the Hessian, which is saved in every point and can be used as an initial guess for the EF method locating of the TS. On the other hand, the exact methods above, which use sophisticated algebra, make several transformations in the Hessian accumulating its errors, which finally leads to a failure in the convergence. Our method is stable; we have never observed a failure in convergence for the test calculations presented here.

The current implementation of the MCD method is a code written by one of the authors (I. B.) in the Pascal program language and compiled with the FreePascal[34] compiler. It reads the output from and creates input for the *Gaussian 03* package,[14] using it for the geometry optimization in each step. The source code, the WIN32 and the LINUX executable programs, as well as a manual, and sample calculations are available for download from the website http://www.chem.elte.hu/departments/elmkem/berente.

Results and Discussion

We tested the MCD method for various chemical reactions of increasing complexity, which we will discuss below in detail. All stationary point geometries were converged to the default values of the *Gaussian 03*:[14] maximum force = 0.00045, rms force = 0.0003, maximum displacement = 0.0018, rms displacement = 0.0012 (all values in atomic units). The existence of the transition states were verified by the IRC[16] method; they are on the reaction path running between the reactants and the products.

Walden Inversion. We calculated the $F^- + H_3C{-}F^* \rightarrow F{-}CH_3 + F^{*-}$ reaction path with the Hartree−Fock molecular orbital method at the RHF/6-31G level with the F−C and C−F* distances as active coordinates using the *Gaussian 03* program package.[14] To test the maximal performance of the method, we set the step size as small as 1 pm. The calculated reaction path is compared with those obtained by the RCD and Ayala−Schlegel methods in Figure 3. As can be seen, our method provides a path, which is remarkably close to the accurate one.

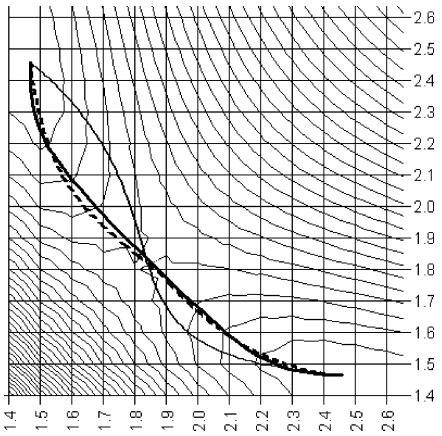

Figure 3. Reaction paths for the Walden inversion as calculated by the MCD (dashed line), RCD (thin line), and Ayala−Schlegel (thick line) methods. Coordinates in angstroms.

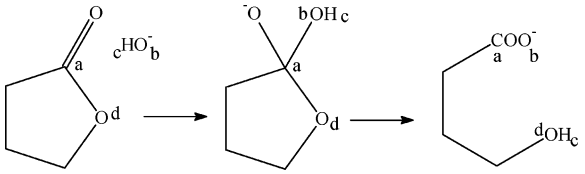

Figure 4. Hydrolysis of γ-butyrolactone.

Alanine Dipeptide Rearrangement. We repeated the test calculation by Peters et al.,[24] a density functional study using the B3LYP parametrization and a 6-31G basis set on the test molecule $HCONH{-}CH(CH_3){-}CONH_2$. The reaction coordinates were the $C{-}N{-}C_\alpha{-}C$ (ϕ) and $N{-}C_\alpha{-}C{-}N$ (ψ) torsion angles; the initial structure was the one called C_5 in ref 24; the step size was set to 0.5 Å (∼50°). The final structure was identical to that called C_{7AX} in ref 24. Calculating the points on the reaction path needed 52 SCF and subsequent gradient calculations. We used the highest-energy point on the path as an initial guess for an RFO saddle point search, which found a saddle point within 18 steps. Our transition-state geometry is identical to that shown in Figure 8 of ref 24 with $\phi = 112.1°$ (107.3°) and $\psi = -140.4°$ (−139.7°); the activation energy is 31.7 (31.8) kJ/mol (values of ref 24 in parentheses). To calculate the full reaction path containing the saddle point, we did 69 SCF and subsequent gradient calculations, while the nudged elastic band[19] and growing string[24] methods need more than 250 and 200 steps, respectively, according to Peters et al.[24]

Hydrolysis of γ-Butyrolactone. This reaction (cf. Figure 4) serves as an appropriate test of the stability of the MCD method in the case of a multistep reaction. We did calculations at the RHF/6-31+G* level and selected the $a{-}b$, $a{-}d$, and $c{-}d$ distances as active coordinates; the step sizes were 0.15, 0.3, 0.45, and 0.6 Å, respectively (see Figure 4). Geometries corresponding to the local minima were optimized, and the highest-energy points served as initial guesses for an EF search. Though the MCD calculations provided more and more inaccurate results with increasing step size, even the worst has provided a good guess for the RFO method to find both transition states. Convergence could be reached after 250 + 49, 142 + 57, 109 + 67, and 82 + 58 MCD + EF gradients for step sizes of 0.15, 0.3, 0.45, and 0.6 Å, respectively (see Figure 5). The STQN calculation obviously failed, since it is unable to handle more than one transition state. However, an STQN calculation between the initial and internal minima converged

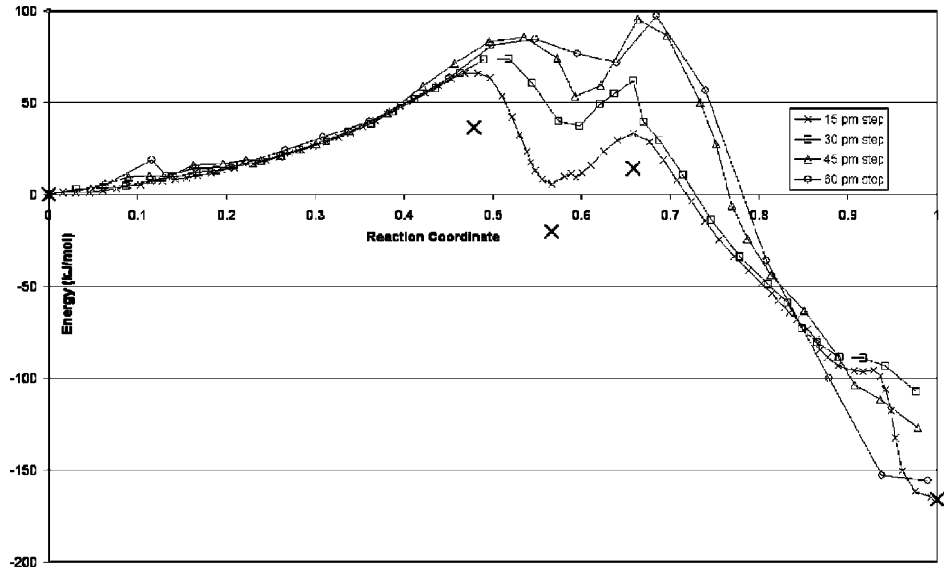

Figure 5. Energy (kJ/mol) vs reaction coordinate plot for butyrolactone hydrolysis. Step sizes are shown in the box; large crosses denote from left to right: reactants, TS_1, internal minimum, TS_2, and products, respectively, each obtained by refining the MCD provided guess with the eigenvector following method. The reaction coordinate is defined as the average progress of the active coordinates between their values for the reactant and the product.

TABLE 1: Comparison of the Performance of the MCD and STQN Methods for the Calculation of Relatively Complex Reactions

reaction	step number (MCD)	step number STQN[14]	energy barrier (kJ/mol)
pyrolytic cracking	116	62	255.7
Grignard reaction	101	66	75.9
butane 1−4 HCl addition	162	41	170.6
hydroxymethylation	282	Failed	131.6
MPV	121	66	115.1
ether condensation	109	Failed	240.2

and provided TS_1. Between the internal minimum and the endpoint, STQN locates a TS, which belongs to a different reaction path with a barrier higher than ours by 2.8 kJ/mol.

We tested the MCD method on six further reactions, for which published results of STQN calculations (*Gaussian 03*,[14] QST2 keyword) are available; the results are summarized in Table 1. All calculations were made at the RHF/3-21G* level; the input structures are available on http://www.chem.elte.hu/departments/elmkem/berente. We used the highest-energy point as an initial guess for the saddle-point search eigenvector-following method. The initial Hessian for EF was calculated analytically. We would like to emphasize that we did not consider whether the mechanisms of the studied reactions are appropriate or not; only the mathematical performance of the method has been considered. For each reaction, we obtained transition-state geometries and activation energies identical with those provided by the STQN method where it was available. We also tried to calculate the path by the Ayala−Schlegel method; however, we observed a failure in the convergence.

Pyrolytic Cracking of 2-Butyl Acetate (Figure 6). We selected the *a*−*b*, *c*−*d*, and *d*−*e* distances as active coordinates, and the step size was 0.5 Å.

Grignard Reaction (Figure 7). The active coordinates were the *a*−*c* distance and the *b*−*d*−*e* bond angle, and the step size was 0.5 Å. For the bending and torsion coordinates, the program considers 1 rad to be equivalent with 1 bohr (0.529 Å).

1,4-Addition of HCl to Butadiene (Figure 8). We selected the *a*−*b*, *c*−*d*, and *b*−*d* distances as active coordinates, and set the step size at 0.5 Å.

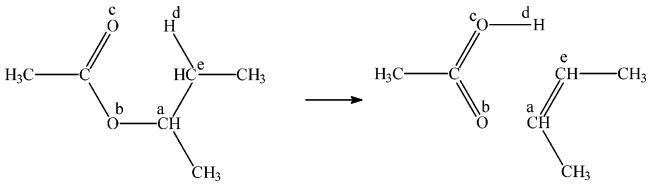

Figure 6. Reactant and products for the pyrolytic cracking reaction of 2-butyl acetate.

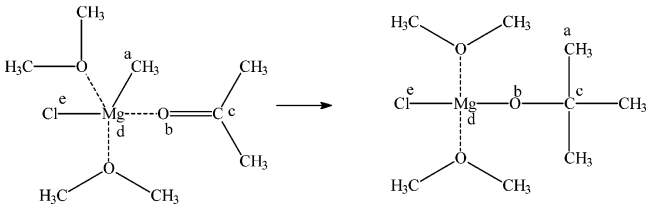

Figure 7. Reactants and products for the Grignard reaction involving dimethyl ether and acetone.

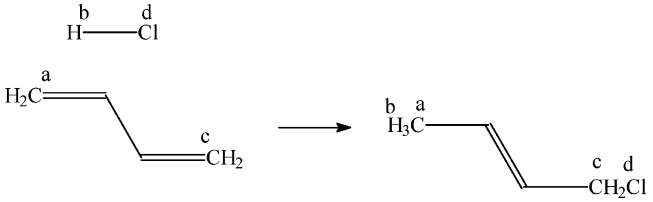

Figure 8. Reactants and products for the HCl addition to butadiene.

Hydroxymethylation of Benzene by formaldehyde with $HCl{\cdot}ZnCl_2$ as a catalyst (Figure 9). The active coordinates were the *a*−*b*, *b*−*f*, *e*−*f*, *c*−*g*, and *c*−*d* distances, and the step size was 0.5 Å.

Meerwein−Ponndorf−Verley (MPV) Reaction with acetaldehyde (Figure 10). The step size was 0.5 Å, and the coordinates were the *a*−*b* and *a*−*c* distances.

Acid-Catalyzed Condensation of propanol and ethanol to yield ethyl propyl ether (Figure 11). The system was protonated, therefore had a net positive charge. The active coordinates were the *a*−*b* and *a*−*c* distances, and the step size was 0.5 Å.

Substrate Hydrolysis by dUTPase. The ubiquitous enzyme dUTPase performs a key role in preventing uracil incorporation

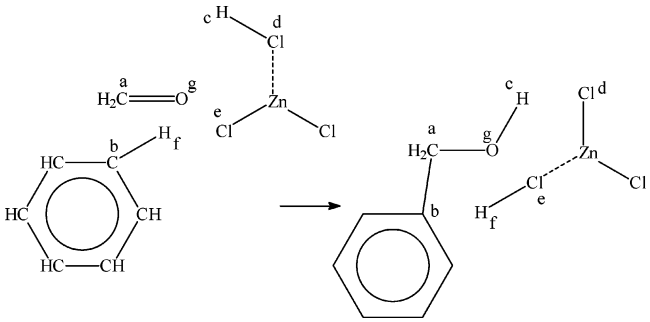

Figure 9. Hydroxymethylation of benzene by formaldehyde with HCl·$ZnCl_2$ as a catalyst.

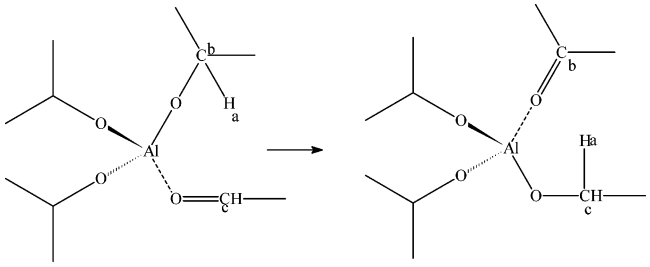

Figure 10. The Meerwein−Ponndorf−Verley (MPV) reaction.

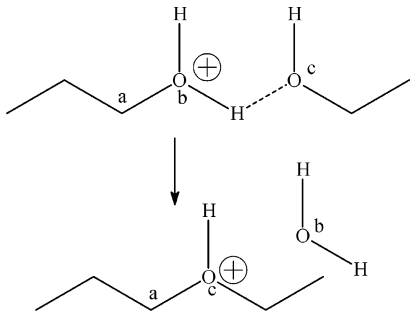

Figure 11. Acid-catalyzed condensation of propanol and ethanol.

into DNA by catalyzing the cleavage of dUTP into dUMP and inorganic pyrophosphate.[35,36] The reaction scheme is depicted in Figure 12.

We calculated the reaction path in the frame of the Hartree−Fock approximation with a 6-31G* basis set on the model with a zero total charge (Figure 13). In the following, we not only show an example but also present a step-by-step guide to an MCD calculation.

1. On the basis of the assumed reaction path, we set the active coordinates: the H_3-O_4, H_3-O_5, O_5-P_1, and P_1-O_6 bond lengths. We used the usual 50 pm step size. We had no initial guess for the product structure; therefore, we set the target value of the active coordinates to be equal to the values for a different system[37] which was similar to what we expected as the product. The MCD calculation provided the energy versus reaction coordinate plot denoted by squares in Figure 14.

2. The structure that belongs to the point denoted by a in Figure 14 is a local minimum candidate. We optimized it and located structure A. We did not expect an intermediate product during the reaction, but the MCD procedure provided one. In this intermediate, the H_3-O_4 bond is formed and the H_3-O_5 bond is broken. However, the O_5-P_1 bond length is 172 pm, while the P_1-O_6 length is 182 pm; both are too large as compared to realistic values, which are 155 pm for the P═O and 160 pm for the P—O bond lengths. P_1 has a trigonal bipyramid structure. We also noticed that the H_3-O_5 distance is much less than we originally expected: H_3 remains within

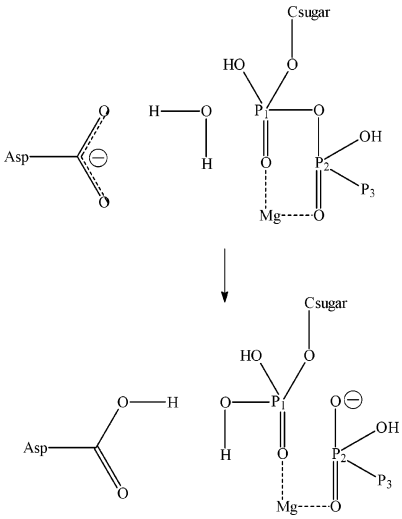

Figure 12. Schematic reaction scheme of the hydrolysis of dUTP by the enzyme dUTPase.

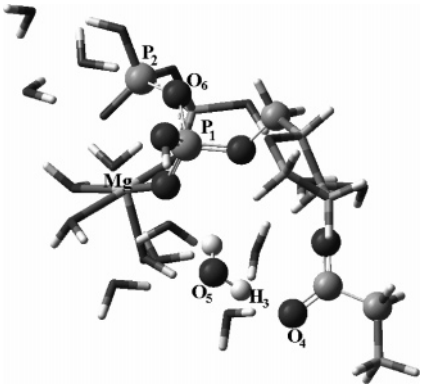

Figure 13. Atomistic resolution model of the active site of the dUTPase−substrate complex.

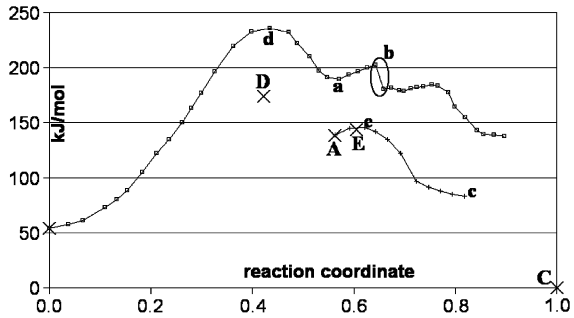

Figure 14. Reaction path for the substrate hydrolysis by dUTPase calculated by the MCD method.

H-bonding distance from O_5. This explains the peak denoted by b in Figure 14. The difference between the two b structures is a H-bond formed between H_3 and one of the oxygen atoms of P_1. However, this is an artifact caused by forcibly removing the H-bonding partner of H_3.

3. To solve the problem explained before, we started a new calculation from the local minimum; the results are denoted by plus signs in Figure 14. It has only one active coordinate, the P_1-O_6 distance.

4. After a few steps, to save time, we terminated the MCD calculation and optimized its last structure, c, and obtained the product structure C.

5. Structures d and e served as initial guesses for transition states. First, we optimized their geometries, keeping the active coordinate frozen and allowing the updater to create an

acceptable Hessian; than, we started an RFO TS search which provided D and E. Both were found to be real saddle points and belong to the reactant−intermediate and intermediate−product reaction paths.

The MCD calculations needed 214 steps; refining the intermediate and the transition states needed 577 steps. Using MCD and evaluating results, we were able to find an unexpected intermediate structure and both transition states for a system with 85 atoms. The existence of such an intermediate is nowadays subject to intensive debate,[38,39] and there is strong evidence that, for another substrate, α,β-imino-dUTP, the intermediate could be located.[36] Our calculations indicate that the presence of a short-lived intermediate cannot be ruled out. Further sophisticated calculations are in progress to address the problem of the α,β-imino-dUTP substrate.

Conclusions

The MCD method is a useful tool to calculate the pathway of any hypothetical reaction, allowing the chemist to follow numerically even the most complicated mechanisms. It determines the path point by point, so the procedure can be terminated if unacceptable results are obtained in a point calculated earlier. It cannot be used as a black box method, and furthermore, it is relatively slow; therefore, where available, the STQN, RGF, Bofill, or even the manual guess + EF methods provide better results. However, for complicated reactions, like enzyme processes, its stability and wide parametrizability make the MCD method useful to obtain an initial guess for the reaction path, especially for the transition-state geometry.

Acknowledgment. The Hungarian National Fund for Scientific Research (grant no. TS044730) supported this work. The authors are indebted to Dr. Ö. Farkas (Budapest) for fruitful discussions and to Dr. O. Barabás (Budapest) for providing the coordinates of the dUTPase−substrate complex.

References and Notes

(1) Mezey, P. G. *Potential Energy Hypersurfaces*; Elsevier: Amsterdam, 1987.
(2) Halgren, T. A.; Lipscomb, W. N. *Chem. Phys. Lett.* **1977**, *49*, 225.
(3) Scharfenberg, P. J. *J. Comput. Chem.* **1982**, *3*, 277.
(4) Williams, I. H.; Maggiora, G. M. *THEOCHEM* **1982**, *89*, 365.
(5) Schlegel, H. B. *J. Comput. Chem.* **2002**, *24*, 1516.
(6) Wales, D. J. *Energy Landscapes*; Cambridge University Press: Cambridge, 2003.
(7) Simons, J.; Banerjee, A.; Adams, N.; Shepard, R. *J. Phys. Chem.* **1985**, *89*, 52.
(8) Baker, J. *J. Comput. Chem.* **1986**, *7*, 385.
(9) Wales, D. J.; Walsh, T. R. *J. Chem. Phys.* **1996**, *105*, 6957.
(10) Simons, J.; Jørgensen, P.; Taylor, H.; Ozment, J. *J. Phys. Chem.* **1983**, *87*, 2745.
(11) Anglada, J. M.; Besalú, E.; Bofill, J. M.; Crehuet, R. *J. Comput. Chem.* **2001**, *22*, 387.
(12) Quapp, W. *J. Comput. Chem.* **2001**, *22*, 537.
(13) Peng, C.; Schlegel, H. B. *Isr. J. Chem.* **1994**, *33*, 449.
(14) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, J.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, revision B.05; Gaussian, Inc.: Pittsburgh, PA, 2003.
(15) Halgren, T. A.; Lipscomb, W. N. *Chem. Phys. Lett.* **1997**, *49*, 225.
(16) Gonzalez, C.; Schlegel, H. B. *J. Chem. Phys.* **1990**, *94*, 5523.
(17) Ayala, P. Y.; Schlegel, H. B. *J. Chem. Phys.* **1997**, *107*, 375.
(18) Liu, H.; Lu, Z.; Cisneros, G. A.; Yang, W. *J. Chem. Phys.* **2004**, *121*, 697.
(19) Henkelman, G.; Jónsson, H. J. *J. Chem. Phys.* **2000**, *113*, 9978.
(20) Trygubenko, S. A.; Wales, D. J. *J. Chem. Phys.* **2004**, *120*, 2082.
(21) Maragakis, P.; Andreev, S. A.; Brumer, Y.; Reichman, D. R. *J. Chem. Phys.* **2002**, *117*, 4651.
(22) Xie, L.; Liu, H. *J. Chem. Phys.* **2004**, *120*, 8039.
(23) Ren, W. Ph.D. Thesis. New York University, 2002.
(24) Peters, B.; Heyden, A.; Bell, A. T.; Chakraborty, A. *J. Chem. Phys.* **2004**, *120*, 7877.
(25) Quapp, W.; Hirsch, M.; Imig, O.; Heidrich, D. *J. Comput. Chem.* **1998**, *19*, 1087.
(26) Hirsch, M.; Quapp, W. *J. Comput. Chem.* **2002**, *23*, 887.
(27) Quapp, W. *J. Comput. Chem.* **2004**, *25*, 1277.
(28) Aktah, D.; Passerone, D.; Parrinello, M. *J. Phys. Chem.* **2004**, *108*, 848.
(29) Fischer, T. H.; Almlof, J. *J. Phys. Chem.* **1992**, *96*, 9768.
(30) Wittbrodt, J. M.; Schlegel, H. B. *THEOCHEM* **1997**, *398*, 55.
(31) Shanno, D. F. *Math. Comput.* **1970**, *24*, 647.
(32) Bofill, J. M. *J. Comput. Chem.* **1994**, *15*, 1.
(33) Bofill, J. M.; Anglada, J. M. *Theor. Chem. Acc.* **2001**, *105*, 463.
(34) http://www.freepascal.org.
(35) Goulian, M.; Bleile, B. M.; Dickey, L. M.; Grafstrom, R. H.; Ingraham, H. A.; Neynaber, S. A.; Peterson, M. S.; Tseng, B. Y. *Adv. Exp. Med. Biol.* **1986**, *195*, B89.
(36) Barabás, O.; Pongrácz, V.; Kovári, J.; Wilmanns, M.; Vértessy, B. G. *J. Biol. Chem.* **2004**, *279*, 42907.
(37) Barabas, O.; Pongrácz, V.; Vértessy, B. G. Protein Data Bank File 2AKV, unpublished.
(38) Mildvan, A. S. *Proteins* **1997**, *29*, 401.
(39) Lahiri, S. D.; Zha, G.; Dunaway-Mariano, D.; Allen, K. N. *Science* **2003**, *299*, 2067.