A Hybrid Potential Reaction Path and Free Energy Study of the Chorismate Mutase Reaction

Sergio Martí,[†] Juan Andrés,[†] Vicent Moliner,*[,†] Estanislao Silla,[‡] Iñaki Tuñón,*[,‡] Juan Bertrán,[§] and Martin J. Field[||]

Contribution from the Departament de Ciències Experimentals, Universitat Jaume I, Box 224, 12080 Castellón, Spain, Departament de Química Física, Universidad de Valencia, 46100 Burjasot, Valencia, Spain, Departament de Química, Universitat Autónoma de Barcelona, 08193 Bellaterra, Barcelona, Spain, and Laboratoire de Dynamique Moléculaire, Institut de Biologie Structurale - Jean-Pierre Ebel, 38027 Grenoble Cedex 1, France

Received September 28, 2000. Revised Manuscript Received November 2, 2000

Abstract: We present a combination of two techniques—QM/MM statistical simulation methods and QM/MM internal energy minimizations—to get a deeper insight into the reaction catalyzed by the enzyme chorismate mutase. Structures, internal energies and free energies, taken from the paths of the reaction in solution and in the enzyme have been analyzed in order to estimate the relative importance of the reorganization and preorganization effects. The results we obtain for this reaction are in good agreement with experiment and show that chorismate mutase achieves its catalytic efficiency in two ways; first, it preferentially binds the active conformer of the substrate and, second, it reduces the free energy of activation for the reaction relative to that in solution by providing an environment which stabilizes the transition state.

The complexity of enzyme reactions makes the modeling of these huge systems extremely difficult. As a result, over the years, a variety of computational techniques have been applied and several working hypotheses have been proposed to explain the selectivity and the efficiency of enzyme reactions as compared to the corresponding reactions in solution. Thus, for example, computational studies by Warshel et al.[1−3] have emphasized the idea that the enzyme lowers the free energy of activation for the formation of the transition state (TS) from the Michaelis complex by providing an environment in which stabilizing dipoles in the enzyme are preoriented in the Michaelis complex toward those that are appropriate to the TS structure. This reduces the free energy barrier because the reorganization of the environment is much less expensive in the enzyme than in solution. In contrast, Kollman et al.[4,5] have highlighted the importance of preorganization effects in which the substrate binds to the enzyme in a configuration that undergoes reaction more readily. These phenomena are not exclusive and rate-constant enhancements can come from both these effects, reorganization and preorganization, as well as from others.

Theoretical insights into enzyme catalysis require both quantum chemical and statistical thermodynamical approaches, and in this paper we employ a combination of both types of techniques. Our strategy has been to identify saddle point structures, to compute the intrinsic reaction coordinate (IRC) path in both directions from these structures, thus obtaining a minimum energy path, and then to use this information to help with the determination of the reaction free energies using molecular dynamics simulation techniques.[6,7] To find the saddle points and IRCs we employ a recently developed methodology that was designed for use with fully flexible hybrid quantum mechanical and molecular mechanical (QM/MM) potentials and that is suitable for use on systems, such as enzymes, that contain a large number of degrees of freedom.[8] Of course, as has been pointed out many times (see, for example, refs 8b and 9), a myriad of saddle points exist with very similar solute/substrate properties differing mainly in the conformation of the solvent/enzyme. Nevertheless, we feel that the identification of even a small number of pathways obtained in the presence of the protein environment (and not in the gas phase) can provide useful information and can act as good starting points for the subsequent free-energy calculations.

The reaction that we study here is that catalyzed by the enzyme chorismate mutase. We have analyzed the structures and their energies taken from the paths of the reaction in solution and in the enzyme and have estimated the relative importance of the reorganization and preorganization effects of the enzyme that were investigated by Warshel et al.[1−3] and by Kollman et al.[4,5]

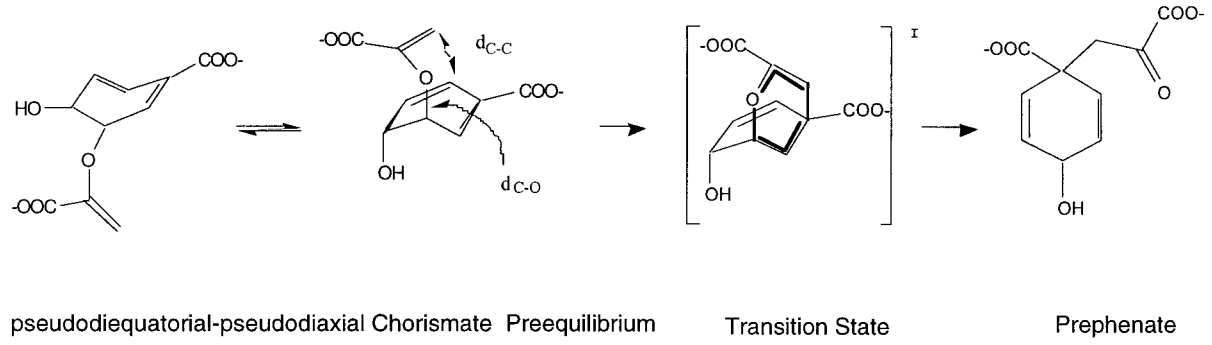
The conversion of (−)-chorismate to prephenate catalyzed by chorismate mutase[10] is shown in Scheme 1. Extensive theoretical analysis of this rearrangement is available in the

[†] Universitat Jaume I.
[‡] Universidad de Valencia.
[§] Universitat Autónoma de Barcelona.
[||] Institut de Biologie Structurale - Jean-Pierre Ebel.

(1) Warshel, A.; Levitt, H. *J. Mol. Biol.* **1976**, *103*, 227−249.
(2) Yadav, A.; Jackson, R. M.; Holbrook, J.; Warshel, A. *J. Am. Chem. Soc.* **1991**, *113*, 4800−4805.
(3) Warshel, A. *Computer Modeling of Chemical Reactions in Enzymes and Solutions*; John Wiley & Sons: New York, 1991.
(4) Stanton, R. V.; Peräkylä, M.; Bakowies, D. B.; Kollman, P. A. *J. Am. Chem. Soc.* **1998**, *120*, 3448−3457.
(5) Kuhn, B.; Kollman, P. A. *J. Am. Chem. Soc* **2000**, *122*, 3909−3916.
(6) Field, M. J. *A Practical Introduction to the Simulation of Molecular Systems*; Cambridge University Press: Cambridge, 1999.
(7) Kollman, P. A. *Chem. Rev.* **1993**, *93*, 2395−2417.
(8) (a) Moliner, V.; Turner, A. J.; Williams, I. H. *J. Chem. Soc. Chem. Comun.* **1997**, 1271−1272. (b) Turner, A. J.; Moliner, V.; Williams, I. H. *J. Phys. Chem. Chem. Phys.* **1999**, *1*, 1323−1331.
(9) Cramer, C. J.; Truhlar, D. G. *Chem. Rev.* **1999**, *99*, 2161−2200.
(10) Haslam, E. *Shikimic Acid: Metabolism and Metabolites*; John Wiley & Sons: New York, 1993.

10.1021/ja003522n CCC: $20.00
Published on Web 02/03/2001

Scheme 1



recent literature (see, for example, ref 11 and references therein) but to our knowledge no free-energy calculations have yet been reported. From a computational point of view, as we intend to use a QM/MM hybrid method to model the process, the chorismate mutase system has the advantage that the enzyme does not take part chemically in the reaction and thus no enzyme atoms have to be included in the QM region. As well as reducing the size of the QM region, this has the advantage that there are no covalent bonds between the QM and the MM regions. Although novel and promising methods have been developed to treat such bonds,[12] their absence at the frontier of the QM and MM regions certainly simplifies the construction of the model system.[13] Another advantage of the chorismate mutase-catalyzed reaction is that it has an observable noncatalyzed solution counterpart[14,15] which makes comparisons between the reaction in solution and in the presence of the protein environment possible.

Calculations in gas phase were carried out with the GAUSSIAN94 package of programs[16] at the AM1[17] semiempirical level and with the B3LYP[18] method, based on density functional theory, using the standard 6-31G* basis set. QM/MM calculations in condensed phase were carried out using the AM1/CHARMM24b2[19] and GRACE[8] programs. The reacting system in solution was the chorismate structure (24 QM atoms) placed in a cavity deleted from a 15 Å radius sphere of water molecules described by the TIP3P[20] empirical potential. As the full system was allowed to move, a solvent boundary potential[21] was applied to the resulting 488 MM water molecules to maintain its structure at the edges of the system. For the QM/MM enzyme calculations the structure of the chorismate mutase enzyme was obtained from *Bacillus subtilis* (BsCM) in the Protein Data Bank (PDB).[22] In these calculations the quantum subsystem was the chorismate, while the enzyme trimer plus crystallization and solvation water molecules formed the MM subsystem (5573 enzyme atoms plus 224 nonrigid TIP3P water molecules). During the enzyme optimizations the QM atoms and the MM atoms lying in a sphere of 18 Å of radius centered on the QM system were allowed to move (a total of 3349 atoms). QM/MM stationary-point location and characterization were guided by means of the GRACE program which uses a Newton−Raphson algorithm with a Hessian matrix of size 72. The IRC reaction paths were traced down to reactant (R) and product (P) valleys from the saddle points as described in a previous paper,[8b] followed by an optimization of the full system.

Once the potential energy profiles were obtained in the gas phase and in the condensed phase (both aqueous and enzymatic media), the potentials of mean force (PMF)[23] were calculated using the umbrella-sampling approach[24] implemented in the Dynamo program.[25] The solution-phase system was a cubic box of side 28 Å containing the solute and 711 water molecules, whereas the enzymatic system consisted of a cubic box of side 57 Å with the enzyme and 3835 water molecules (17183 atoms in total). Periodic boundary conditions were used in both cases. The starting geometries were the saddle point structures previously located in the three different environments, while the distinguished reaction coordinate was taken as the antisymmetric combination of the distances describing the breaking and forming bonds, i.e., $d_{C-C} - d_{C-O}$ (see Scheme 1). This variable represents very closely (or exactly) the IRCs that were found in the enzyme, in the gas phase and in solution and thus its choice is the natural one when determining the PMFs. The procedure for the PMF calculation is straightforward and requires a series of molecular dynamics simulations in which the reaction coordinate variables are constrained about particular values. The values of the variables sampled during the simulations are then pieced together to construct a distribution function from which the PMF is obtained. The full details of the approach are very similar to those in, for example, ref 26 and therefore are not repeated here.

The consequences of the existence of at least two conformational chorismate structures, each one giving rise to a different

(11) Khanjin, N. A.; Snyder, J. P.; Menger, F. M. *J. Am. Chem. Soc.* **1999**, *121*, 11831−11846.
(12) (a) Gao, J.; Amara, P.; Alhambra, C.; Field, M. J. *J. Phys. Chem. A* **1998**, *102*, 4714−4721. (b) Antonczak, S.; Monard, G.; Ruiz-López, M. F.; Rivail, J. L. *J. Am. Chem. Soc.* **1998**, *120*, 8825−8833.
(13) For a recent comparison of different approaches treating frontier bonds in QM/MM methods, see: Reuter, N.; Dejaegere, A.; Maigret, B.; Karplus, M. *J. Phys. Chem. A* **2000**, *104* 1720−1735.
(14) Young, I. G.; Gibson, I.; MacDonald, C. G. *Biochim. Biophys. Acta* **1969**, *192*, 62−72.
(15) Gajewski, J. J.; Jurayi, J.; Kimbrough, D. R.; Gande, M. E.; Ganem, B.; Carpenter, B. K. *J. Am. Chem. Soc.* **1987**, *109*, 1170−1186.
(16) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T.; Peterson, G. A.; Montgomery, J. A.; Raghavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Cioslowski, J.; Stefanov, B. B.; Nanayakkara, A.; Challacombe, M.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. P.; Head-Gordon, M.; Gonzalez, C.; Pople, J. A. *GAUSSIAN94*; Gaussian Inc.: Pittsburgh, PA, 1995.
(17) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902−3909.
(18) (a) Lee, C.; Yang, W.; Parr, R. G.; *Phys. Rev. B.* **1988**, *37*, 785−789. (b) Becke, A. D. *Phys. Rev. A.* **1988**, *38*, 3098−3100.
(19) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.
(20) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.
(21) Brooks, C. L.; Karplus, M. *J. Mol. Biol.* **1989**, *208*, 159.
(22) Chook, Y. M.; Ke, H.; Lipscomb, W. N. Protein Data Bank ID code 1COM.
(23) Roux, B. *Comput. Phys. Commun.* **1995**, *91*, 275−282.
(24) Torrie, G. M.; Valleau, J. P. *J. Comput. Phys.* **1977**, *23*, 187−199.
(25) Field, M. J.; Albe, M.; Bret, C.; Proust-de Martin, F.; Thomas, A. *J. Comput. Chem.*, **2000**, *21*, 1088−1100.
(26) Proust-de Martin, F.; Dumas R.; Field, M. J. *J. Am. Chem. Soc.*, **2000**, *122*, 7688−7697.

Table 1. Breaking Bond Distance (d_{C-O} in Å), Forming Bond Distance (d_{C-C} in Å), and the Angle Describing the Orientation of the Hydroxyl Group with Respect to the Ring (ϕ_{OH} in Degrees) Obtained for R and TS in Aqueous Solution and in the Enzymatic Environment by Means of AM1/CHARMM/GRACE[a]

	aqueous solution		enzyme	
	R	TS	R	TS
d_{C-O}	1.447 ⟨1.445⟩	1.874 ⟨1.775⟩	1.443 ⟨1.453⟩	1.869 ⟨1.944⟩
d_{C-C}	3.209 ⟨3.749⟩	2.149 ⟨2.095⟩	3.425 ⟨3.047⟩	2.116 ⟨2.181⟩
ϕ_{OH}	125.8 ⟨172.8⟩	122.6 ⟨296.3⟩	144.0 ⟨187.7⟩	151.1 ⟨171.5⟩

[a] Average values, obtained with Dynamo QM/MM statistical calculations, are listed in brackets.

Table 2. Activation Energies ($\Delta E^{\ddagger}$ and Activation Free Energies ($\Delta G^{\ddagger}$) Obtained by Means of the CHARMM/GRACE and Dynamo Programs, Respectively, Corrected Activation Free Energies ($\Delta G^{\ddagger}_{corr}=\Delta G^{\ddagger} - (\Delta E^{\ddagger}_{B3LYP-vacuo} - \Delta E^{\ddagger}_{AM1-vacuo})$), and Experimental Activation Free Energies ($\Delta G^{\ddagger}_{exp}$)[a]

	$\Delta E^{\ddagger}$	$\Delta G^{\ddagger}$	$\Delta G^{\ddagger}_{corr}$	$\Delta G^{\ddagger}_{exp}$
in vacuo	52.7	48.9	40.2	
aqueous solution	44.1	38.0	29.3	24.5[b]
enzyme environment	35.7	29.3	20.6	15.4[c]

[a] All values are in kcal mol^{-1}. [b] From Andrews, P. R.; Smith, G. D.; Young, I. G. *Biochemistry* **1973**, *12*, 3492−3498. [c] From Kast, P.; Asif-Ullah, M.; Hilvert, D. *Tetrahedron Lett.* **1996**, *37*, 2691−2694.

reaction path leading to prephenate through two different TSs (TS_{OHin} and TS_{OHout}) have been studied.[27a] These structures mainly differ in the orientation of the hydroxyl group with respect to the ring. As the OHout conformation has been demonstrated to be the energetically favorable profile,[27a] our molecular dynamics trajectories were obtained starting from the TS_{OHout} conformational structure. Nevertheless, analysis of the angle describing the relative orientation of the OH group, ϕ_{OH}, obtained from the aqueous medium trajectories reveals values ranging between OH_{IN} and OH_{OUT}, the latter being the most populated structure. This result can be interpreted as being due to an in−out conformational change of the hydroxyl group in aqueous solution with a small energy barrier. Probably due to steric interactions with the protein, this effect was not observed in the enzyme environment, where the OH_{OUT} conformation was maintained during the full trajectory. The breaking and forming bond distances obtained from the QM/MM internal energy minimizations, as well as the ϕ_{OH} dihedral angle are listed in Table 1. Average values obtained from the QM/MM trajectories are also presented in the same table. The distances rendered by the internal energy optimizations are quite similar to the average values. On the other hand, the orientation of the OH group appears to be much more flexible, particularly in solution.

The activation energies, $\Delta E^{\ddagger}$, listed in Table 2 are in agreement with the expected catalytic effect of the water molecules and the *Bacillus subtilis* enzyme compared with the reaction in gas phase. When the statistical contributions are taken into account, the activation barrier, $\Delta G^{\ddagger}$, calculated as the maximum G value found along the reaction path,[28] follows the same trend; that is, the chorismate to prephenate rearrangement is faster in an enzymatic environment than in aqueous solution. At this point, it is important to bear in mind that a semiempirical Hamiltonian has been used to describe the quantum region in all QM/MM calculations. As a qualitative attempt to correct this possible source of error, we have evaluated a correction term as the subtraction between the in vacuo activation energy computed by means of the B3LYP/6-31G* density functional method ($\Delta E^{\ddagger} = 44.0$ kcal·mol^{-1}) and the AM1 semiempirical one ($\Delta E^{\ddagger} = 52.7$ kcal·mol^{-1}). The resulting corrected values, $\Delta G^{\ddagger}_{corr}$, are listed in Table 2. As expected, these values are significantly closer to the experimental data.

Table 3. E_T, $\Delta E^{\ddagger}$, ΔE_{QM}, ΔE_{int}, ΔE_{MM} in Water Environment and Enzyme Active Site [a]

	E_T	$\Delta E^{\ddagger}$	$\Delta E_{QM}{}^{\ddagger}$	ΔE_{int}	ΔE_{MM}
		aqueous solution			
R	−6411.3	0.0	0.0	0.0	0.0
TS	−6367.2	44.1	36.4	2.4	5.3
		⟨54.9⟩	⟨37.4⟩	⟨2.6⟩	⟨14.9⟩
		enzyme			
R	−3547.9	0.0	0.0	0.0	0.0
TS	−3512.2	35.7	38.3	−2.5	−0.1
		⟨28.2⟩	⟨41.6⟩	⟨−17.8⟩	⟨4.4⟩

[a] E_T = total energies of corresponding QM/MM optimized structures. $\Delta E^{\ddagger}$ = barrier energy. $\Delta E_{QM}{}^{\ddagger}$= in vacuo quantum single point energy using the solute/substrate structures optimized in the QM/MM calculations. ΔE_{int} = solvent−solute or enzyme−substrate interaction energy. ΔE_{MM} = solvent−solvent or enzyme−enzyme interaction energy. All values, in kcal mol^{-1}, are relative to reactants. Average values, obtained with Dynamo QM/MM statistical calculations, are listed in brackets.

The comparison of average and optimized geometrical values listed in Table 1 confirms the fact that the OHout conformer obtained from the minimizations is representative of the families obtained by the statistical calculations. Thus, we carried out an analysis of the different energy terms of the unique R and TS molecules obtained from the AM1/CHARMM/GRACE internal energy optimizations. The total QM/MM activation energies obtained in aqueous solution or in the presence of the enzyme environment can be written as the following sum:

$$\Delta E^{\ddagger} = \Delta E_{QM} + \Delta E_{int} + \Delta E_{MM}$$

where ΔE_{QM} is the in vacuo AM1 single point energy calculation relative to R using the solute/substrate structures obtained in the QM/MM calculations, ΔE_{int} is the solvent−solute or enzyme−substrate interaction energy relative to R, and ΔE_{MM} is the MM energy relative to R structure. From the ΔE_{QM} column of Table 3, it is clear that the TS relative energy is slightly more favorable in the water environment than in the enzyme active site. Consequently, the catalytic character of the enzyme comes from the favorable enzyme−substrate interaction energy term (see the ΔE_{int} column) and the smaller energy of deformation for the enzyme on going from R to TS (see the ΔE_{MM} column). From these results we conclude that the enzyme preferentially interacts with the TS structure, while the solvent interacts more strongly with the R structures of the solute, favoring the enzyme reaction by .∼4.9 kcal·mol^{-1}. Furthermore, the reorganization of the environment is much less expensive in the enzyme than in solution by .∼5.4 kcal·mol^{-1}. At this point of the discussion, it is important to emphasize that the previous analysis comes from unique structures (i.e., local minima) and it is possible that allowing the water to reorient during dynamics could reduce the sum of ΔE_{int} and ΔE_{MM} for

(27) (a) Martí, S.; Andrés, J.; Moliner, V.; Silla, E.; Tuñón, I.; Bertrán, J. *Theor. Chem. Acc.*, **2001,** *105*, 207−212. (b) Martí, S.; Andrés, J.; Moliner, V.; Silla, E.; Tuñón, I.; Bertrán, J. *J. Phys. Chem.* **2000**, *104*, 11308−11315.

(28) Making the free energy of activation equivalent to the PMF barrier height along a selected distinguished reaction coordinate is an assumption that is made routinely in applications of standard transition state theory (see, for example: Field, M. J. In *The Simulation of Chemical Reactions*; van Gunsteren, W. F., Weiner, P. K., Wilkinson, A. J., Eds.; Computer Simulation of Biomolecular Systems: Theoretical and Experimental Applications, Vol 2; ESCOM: Leiden, The Netherlands, 1993; pp 82−123 and references therein). More refined estimates of the activation energy necessitate the inclusion of dynamical effects although these often have only a small (energetic) influence on the values obtained (see, for example: Alhambra, C.; Gao, J.; Corchado, J. C.; Villa, J.; Truhlar, D. G. *J. Am. Chem. Soc.* **1999**, *121*, 2253−2258).

the solution case. To gain insight into this point, we computed the average values of these contributions from the QM/MM molecular dynamics trajectories obtained with Dynamo. Applying a harmonic constraint to the antisymmetric combination of $d_{C-C}-d_{C-O}$ distances, as used in the PMF calculations, one hundred conformations were generated for the TS and R structures. The average values of the different energetic terms are listed in brackets in Table 3. These results, although they have to be considered as approximate, indicate that our hypothesis, deduced from single stationary point structures of a potential energy hypersurfase, can be extrapolated to other conformers.

Finally, to investigate the preorganization effect of the enzyme in the formation of the Michaelis complex, we have analyzed the equilibrium between the pseudodiequatorial and pseudodiaxial conformers of chorismate (see Scheme 1).[27b] Obviously, the Michaelis complex corresponds to the pseudodiaxial structure. In gas phase, the equilibrium is completely displaced to the pseudodiequatorial one, whereas Copley and Knowles observed a rather small free energy change for this equilibrium in water (1.2 kcal·mol^{-1}),[29] which leads to an increase in the population of the diaxial structure in solution. Using QM/MM conformational searching minimization techniques, we found that this latter structure, that is, the one that is found in the Michaelis complex, is much more stable than the pseudodiequatorial one in the protein environment (an internal energy difference of 71.7 kcal·mol^{-1} was obtained between the most stable pseudodiaxial and pseudodiequatorial conformers), displacing the equilibrium toward the reactive chorismate conformer.[27b] One of the reasons for this stabilization is the rather small enzyme deformation that is required which can be interpreted as being due to a preorganization of the enzyme favoring this complex. Furthermore, a deformation of the R structure toward the TS can be deduced from the average values appearing in Table 1. While the d_{c-c} distance of the minimum energy R structures is larger in the enzyme than in aqueous solution, the average value is substantially shorter in the former. This result can be interpreted as the enzyme pushing the substrate toward the TS of the reaction to be catalyzed.

The results we obtain for this reaction show that both reorganization and preorganization effects are important in helping chorismate mutase achieve its catalytic activity—thus, the enzyme preferentially binds the reactive pseudodiaxial conformer of chorismate, and second, it reduces the free energy of activation for the reaction relative to that in solution by providing an environment which preferentially stabilizes the transition state**.**

Acknowledgment. We are indebted to DGICYT for Project BQU2000-1425 that supported this research, the Servei d'Informatica of the Universitat Jaume I and the Institut de Biologie Structurale for providing us with computer facilities. S.M. thanks Generalitat Valenciana for a doctoral fellowship and for travel funding.

JA003522N

(29) Copley, S. D.; Knowles J. R. *J. Am. Chem. Soc.* **1987**, *109*, 5008−5013.