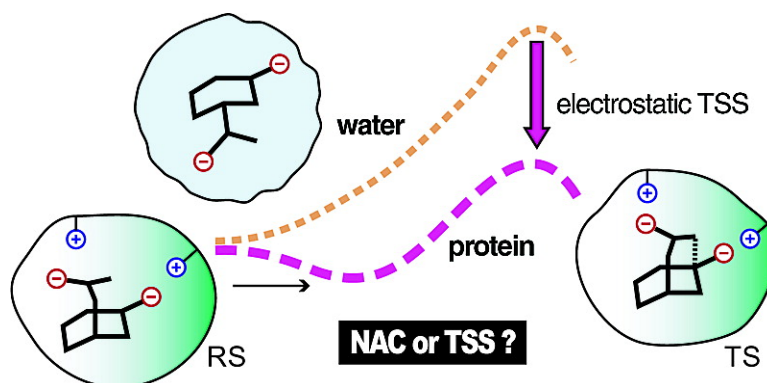## Article

Apparent NAC Effect in Chorismate Mutase
Reflects Electrostatic Transition State Stabilization

Marek trajbl, Avital Shurki, Mitsunori Kato, and Arieh Warshel

More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Links to the 9 articles that cite this article, as of the time of this article download
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML

Apparent NAC Effect in Chorismate Mutase Reflects Electrostatic Transition State Stabilization

Marek Štrajbl, Avital Shurki, Mitsunori Kato, and Arieh Warshel*

Contribution from the Department of Chemistry, University of Southern California, Los Angeles, California 90098-1062

Received April 15, 2003; E-mail: warshel@usc.edu

Abstract: The catalytic reaction of chorismate mutase (CM) has been the subject of major current attention. Nevertheless, the origin of the catalytic power of CM remains an open question. In particular, it has not been clear whether the enzyme works by providing electrostatic transition state stabilization (TSS), by applying steric strain, or by populating near attack conformation (NAC). The present work explores this issue by a systematic quantitative analysis. The overall catalytic effect is reproduced by the empirical valence bond (EVB) method. In addition, the binding free energy of the ground state and the transition state is evaluated, demonstrating that the enzyme works by TSS. Furthermore, the evaluation of the electrostatic contribution to the reduction of the activation energy establishes that the TSS results from electrostatic effects. It is also found that the apparent NAC effect is not the reason for the catalytic effect but the result of the TSS. It is concluded that in CM as in other enzymes the key catalytic effect is electrostatic TSS. However, since the charge distribution of the transition state and the reactant state is similar, the stabilization of the transition state leads to reduction in the distance between the reacting atoms in the reactant state.

1. Introduction

Chorismate mutase (CM) catalyses the Claisen rearrangement of chorismate to prephenate. This reaction is a key step in the shikimate pathway for biosynthesis of phenylalanine and tyrosine in bacteria, fungi, and higher plants.[1] This enzymatic rearrangement has been the focus of major effort in recent years, including analysis of its relationship to catalytic antibodies that catalyze the same reaction (e.g., refs 2–7) and extensive simulation studies.[8–16] One of the appealing features of this enzyme is the fact that the chemical reaction only involves the substrate without participation of the enzyme groups in the actual chemistry. This allows for a convenient analysis of the

environmental effect of the enzyme and makes it an ideal candidate for combined quantum mechanical/molecular mechanics (QM/MM) studies (for general references on QM/MM approaches, see refs 17–27). In fact, the reaction of CM has become an important benchmark for QM/MM calculations.[10–13] It should be mentioned, however, that many of the reported studies did not provide stable results or complete configurational sampling, and none of the reported studies involved ab initio QM/MM calculations with complete configurational sampling. Thus, the study of CM still presents an interesting computational challenge.

The nature of the catalytic effect of CM has been the subject of recent studies[12,14] that considered it as a prototype for the steric strain proposal,[28–31] where it is assumed that the enzyme is designed to be geometrically complementary to the transition state.[12] While the pure steric strain proposal is highly problem-

(1) Haslam, E. *Shikimic Acid: Metabolism and Metabolites*; John Wiley & Sons: New York, 1993.
(2) Weist, O.; Houk, K. *J. Am. Chem. Soc.* **1995**, *117*, 11628–11639.
(3) Hilvert, D. *Annu. Rev. Biochem.* **2000**, *69*, 751–793.
(4) Mader, M. M.; Bartlett, P. A. *Chem. Rev.* **1997**, *97*, 1281–1301.
(5) Cload, S. T.; Liu, D. R.; Pastor, R. M.; Schultz, P. G. *J. Am. Chem. Soc.* **1996**, *118*, 1787–1788.
(6) Kienhofer, A.; Kast, P.; Hilvert, D. *J. Am. Chem. Soc.* **2003**, *125*, 3206–3207.
(7) Barbany, M.; Gutierrez-de-Teran, H.; Sanz, F.; Villà-Freixa, J.; Warshel, A. *ChemBioChem* **2003**, *4*, 277–285.
(8) Copley, S. D.; Knowles, J. R. *J. Am. Chem. Soc.* **1987**, *109*, 5008–5013.
(9) Lee, A.; Stewart, J. D.; Clardy, J.; Ganem, B. *Chem. Biol.* **1995**, *2*, 195–203.
(10) Lyne, P. D.; Mulholland, A. J.; Richards, W. G. *J. Am. Chem. Soc.* **1995**, *117*, 11345–11350.
(11) Guo, H.; Cui, Q.; Lipscomb, W. N.; Karplus, M. *Proc. Natl. Acad. Sci.* **2001**, *98*, 9032–9037.
(12) Martí, S.; Andrés, J.; Moliner, V.; Silla, E.; Tunon, I.; Bertran, J. *J. Phys. Chem. B* **2000**, *104*, 11308–11315.
(13) Martí, S.; Andrés, J.; Moliner, V.; Silla, E.; Tunon, I.; Bertran, J. *Theor. Chem. Acc.* **2001**, *105*, 207–212.
(14) Khanjin, N. A.; Snyder, J. P.; Menger, F. M. *J. Am. Chem. Soc.* **1999**, *121*, 11831–11846.
(15) Hur, S.; Bruice, T. C. *Proc. Natl. Acad. Sci.* **2002**, *99*, 1176–1181.
(16) Hur, S.; Bruice, T. C. *J. Am. Chem. Soc.* **2003**, *125*, 1472–1473.
(17) Bentzien, J.; Muller, R. P.; Florian, J.; Warshel, A. *J. Phys. Chem. B* **1998**, *102*, 2293–2301.
(18) Muller, R. P.; Warshel, A. *J. Phys. Chem.* **1995**, *99*, 17516–17524.
(19) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227–249.
(20) Théry, V.; Rinaldi, D.; Rivail, J.-L.; Maigret, B.; Ferenczy, G. G. *J. Comput. Chem.* **1994**, *15*, 269–282.
(21) Zhang, Y.; Liu, H.; Yang, W. *J. Chem. Phys.* **2000**, *112*, 3483–3492.
(22) Gao, J. *Acc. Chem. Res.* **1996**, *29*, 298–305.
(23) Bakowies, D.; Thiel, W. *J. Phys. Chem.* **1996**, *100*, 10580–10594.
(24) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700–733.
(25) Friesner, R.; Beachy, M. D. *Curr. Opin. Struct. Biol.* **1998**, *8*, 257–262.
(26) Monard, G.; Merz, K. M. *Acc. Chem. Res.* **1999**, *32*, 904–911.
(27) Field, M. *J. Comput. Chem.* **2002**, *23*, 48–58.
(28) Tapia, O.; Andrés, J.; Safront, V. S. *J. Chem. Soc., Faraday Trans. 1* **1994**, *90*, 2365–2374.
(29) Phillips, D. C. *Sci. Am.* **1966**, *215*, 78–90.
(30) Blake, C. C. F.; Johnson, L. N.; Mair, G. A.; North, A. C. T.; Phillips, D. C.; Sarma, V. R. *Proc. R. Soc. Ser. B* **1967**, *167*, 378–388.
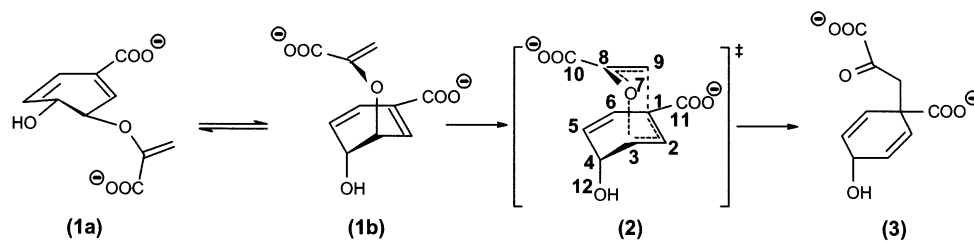(31) Moliner, V.; Andrés, J.; Oliva, M.; Safont, V. S.; Tapia, O. *Theor. Chem. Acc.* **1999**, *101*, 228–233.

Figure 1. Rearrangement of chorismate (**1**) to prephenate (**3**). The presumed transition state is shown (**2**). This [3,3]-pericyclic process is formally analogous to a Claisen rearrangement. The chorismate molecule involves the diequatorial and diaxial classes of configurations, which are represented schematically by **1a** and **1b**, respectively.

atic (see ref 32 and also this paper), CM seems to display an apparent steric effect that can be considered as the result of a confinement by the active site. That is, recent MD studies of Hur and Bruice[15,16] suggested that the enzyme helps in bringing the reacting atoms of the substrate to a typical distance, which they defined as near attack conformation (NAC), and that this distance is rarely attained in water. Thus, they considered this NAC effect as the major reason for the catalytic power of CM. As clarified in our previous work,[32] the NAC effect and the NAC distance are poorly defined and cannot be uniquely related to the catalytic effect of the given enzyme. Nevertheless, the MD studies of Hur and Bruice[15,16] indicate that the $C_9 \cdots C_1$ distance of the substrate (Figure 1 and also Figure 1S in the Supporting Information) can easily reach a distance of 3.7 Å, while it costs significant free energy ($\sim$8 kcal/mol according to ref 16) to reach this distance in water. The question that we would like to address is what is the meaning of the apparent NAC effect? In other words, we would like to find out whether the NAC represents a genuine reason for catalysis or whether it merely reflects the result of electrostatic transition state stabilization.

Here, we address this issue by first defining clearly the concepts of reactant state destabilization (RSD) and transition state stabilization (TSS). Next, we examine which of these effects is operating in CM, using the empirical valence bond (EVB) and the linear response approximation (LRA) approaches for the calculations. It is found that CM operates by electrostatic TSS, and the apparent NAC effect is simply the result of that TSS.

2. Defining the Catalytic Effect

To clarify that we are dealing with a fundamental problem and not with a semantic problem, it is crucial to define what is meant by enzyme catalysis. As discussed in many previous works (e.g., refs 33 and 34), the key issue has always been the origin of the reduction of $\Delta g_{cat}^{\ddagger}$ relative to $\Delta g_{w}^{\ddagger}$, where $\Delta g_{cat}^{\ddagger}$ and $\Delta g_{w}^{\ddagger}$ are, respectively, the activation barriers associated with k_{cat} and the rate constant for the uncatalyzed reaction in water (k_w). More specifically, the fact that the binding free energy of the TS (which corresponds to k_{cat}/K_M) is negative was clearly understood (e.g., ref 35) since large changes upon moving the substrate from water to the enzyme site are easy to rationalize. It was much harder to understand and rationalize
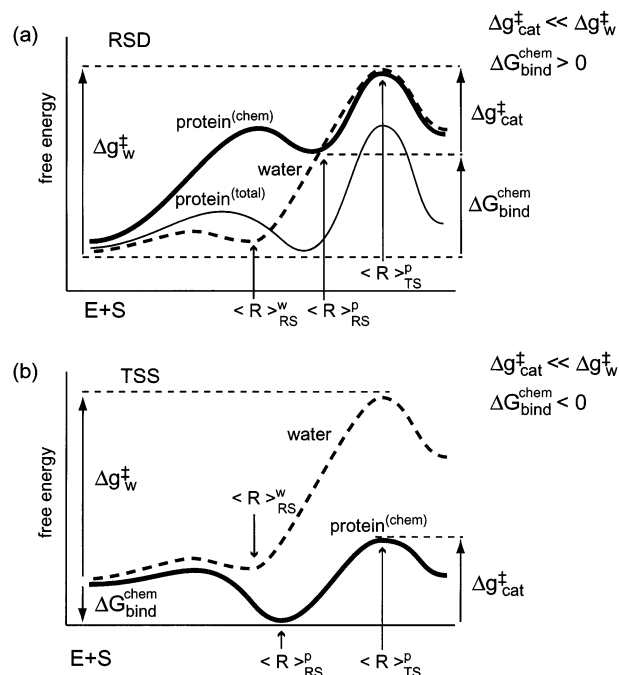

Figure 2. Schematic description of the free energy profiles in protein (bold and plain lines) and in water (dashed line) for the limiting cases of RSD and TSS, which are shown in the upper and lower panels, respectively. The figure focuses for simplicity on the profiles for the "chemical part" of the substrate (bold line) and describes in panel (a) also the profile for the chemical plus nonreactive part of the substrate (see text). For simplicity, we consider a case where the binding of the nonreactive part is along a coordinate orthogonal to the reaction coordinate.

correctly what the origin is of the difference in the binding energy of the TS and RS since both are confined to the same active site, which is quite flexible and thus can complement the geometries of both states. In any case, there are two limiting generic ways for reducing $\Delta g_{cat}^{\ddagger}$, which are shown in Figure 2. The first option (Figure 2a) involves a reduction of $\Delta g_{cat}^{\ddagger}$ because of RSD. The generic RSD model appears in many catalytic hypotheses including the strain proposal[12,14,29,30] and the desolvation proposal (e.g., refs 36 and 37). In the RSD case, it is assumed implicitly that the binding to distant nonreactive parts allows the enzyme to strain the reacting parts.[38] In doing so, it is assumed that the substrate can be divided into reacting fragments (the "chemical part") and distant fragments that do not contribute to the activation free energy. To simplify our

(32) Shurki, A.; Štrajbl, M.; Villa, J.; Warshel, A. *J. Am. Chem. Soc.* **2002**, *124*, 4097–4107.
(33) Warshel, A. *Computer Modeling of Chemical Reactions in Enzymes and Solutions*; John Wiley & Sons: New York, 1991.
(34) Warshel, A. *J. Biol. Chem.* **1998**, *273*, 27035–27038.
(35) Wolfenden, R.; Snider, M. J. *Acc. Chem. Res.* **2001**, *34*, 938–945.

(36) Crosby, J.; Stone, R.; Lienhard, G. E. *J. Am. Chem. Soc.* **1970**, *92*, 2891–2900.
(37) Dewar, M. J. S.; Storch, D. M. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 2225–2229.
(38) Jencks, W. P. *Catalysis in Chemistry and Enzymology*; Dover Publication: New York, 1986.
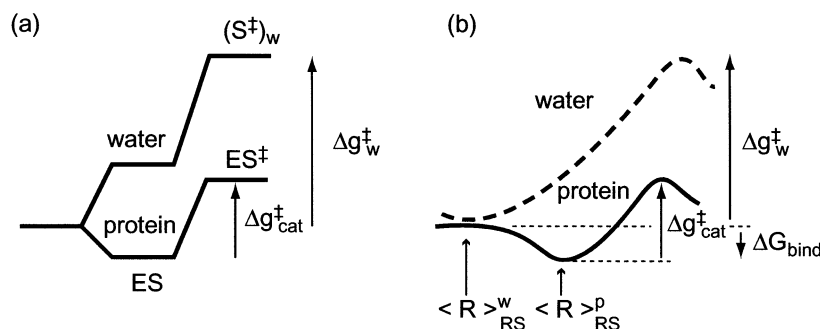
Figure 3. Illustrating the nature of recent diagrams that were used to describe the NAC effect.[16,43,44] Panel (a) shows an energy diagram of the type used in refs 16, 43, and 44 to describe the NAC proposal. Panel (b) clarifies that the position and nature of the assumed NAC plateau does not change any of the relevant activation barriers so that we simply have a regular TSS situation such as in Figure 2b. In other words, the position and height of the NAC plateau has no effect on the difference between $\Delta g_{cat}^{\ddagger}$ and $\Delta g_w^{\ddagger}$ and thus no effect on catalysis. This is not meant to suggest that a properly defined NAC effect is a TSS. That is, at least in the cases considered by ref 43 the NAC was assumed to represent entropic effects, which clearly lead to RSD and to the situation of Figure 2a (as discussed in the text, the diagram of the type used in ref 43 is simply invalid if the NAC reflects entropic or steric effects). If the NAC reflects electrostatic TS stabilization we can use Figure 3a, but in this case we have a standard TSS case, and the NAC is not the reason for catalysis.

discussion, Figure 2 focuses (bold line) on the reaction profile and the corresponding binding energy of the reacting part (ΔG_{bind}^{chem}). Adding the binding of a distant nonreactive part (plain line in Figure 2a) leads to a more complex discussion and analysis[39], but the conclusions of the present work will stay the same. It is important to note that Figure 2a does not contradict any physical law even in the absence of distant groups (see footnote 40). Also, note that in the special case of CM it is not justified to talk on the significant contribution to binding from distant parts of the molecule since the two ionized groups of the substrate are covalently bound to the atoms that form the chemical bond and determine the reaction profile (see also the Conclusion). At any rate, in the RSD mechanism the enzyme is supposed to push the reacting fragments in the reactant state (RS) toward the transition state (TS) direction. Thus, the average distance along the reaction coordinate in the RS ($\langle R \rangle_{RS}$ in Figure 2) will be closer to $\langle R \rangle_{TS}$ in the protein than in water. The second option (Figure 2b) is the TSS mechanism, where the enzyme stabilizes both the RS and the TS more than water does. However, the stabilization is larger in the TS than in the RS, so that $\Delta\Delta G_{TS}^{w \rightarrow p}$ is larger than $\Delta\Delta G_{RS}^{w \rightarrow p}$ (where w and p designate water and protein, respectively). The two options considered in Figure 2 correspond to these two extreme limits of the two different proposals.

Bruice and co-workers suggested that CM catalyzes its reaction by enabling the $C_9 \cdots C_1$ ($C_5 \cdots C_{16}$ in Bruice's notation) to reach a NAC distance (3.7 Å), which is energetically costly in water. Now, from a rigorous point of view, the NAC concept is arbitrary since one can choose any distance on the way to the TS (including the TS distance itself). Furthermore, the rate constant is determined by the difference between the probability of being at the TS and the probability of being in the entire reactant state region (see Figure 3), which is approximately the probability of being at the minimum of the free energy profile at the RS side. Thus, considering only the probability of being at specific points other than $\langle R \rangle_{RS}$ (e.g., NAC) does not tell us much about $\Delta g_{cat}^{\ddagger}$. Nevertheless, as explained before[32], it is

possible to formulate what is probably meant by the NAC concept in a clear (and verifiable) way by the restraint release approach of ref 32. Similar results can be obtained by a simpler approach focusing on $\langle R \rangle_{RS}^p$ and $\langle R \rangle_{RS}^w$ (R is the solute contribution to the reaction coordinate). That is (assuming that the TS structure is similar in water and in the protein), we may simply ask how much energy it would cost to push R to $\langle R \rangle_{RS}^p$ in water and approximate the NAC free energy by

$$\Delta G_{NAC} \approx \Delta G(\langle R \rangle_{RS}^p)_w - \Delta G(\langle R \rangle_{RS}^w)_w \qquad (1)$$

where $\Delta G(\langle R \rangle)_w$ is the free energy in water when R is taken as the designated $\langle R \rangle$. In the above approximation, $\langle R \rangle_{RS}^p$ provides a proper definition for the NAC distance. However, this analysis does not tell us what the origin is of the NAC effect (e.g., electrostatic, steric, entropic, etc.). Here, we can move back to Figure 2. In the RSD case (Figure 2a), we have a real confinement effect where the enzyme pushes the reactants toward the TS, and this leads to a reduction of $\Delta g_{cat}^{\ddagger}$. Now in the TSS case of Figure 2b, we can also have a compression effect where $\langle R \rangle_{RS}^p$ is closer to $\langle R \rangle_{TS}$ than $\langle R \rangle_{RS}^w$. However, this is simply a reflection of the fact that the TSS flattens the potential surface in the enzyme; thus, $\langle R \rangle_{RS}^p$ becomes closer to $\langle R \rangle_{TS}^p$ (Figure 2b).

In view of a referee comment and at the risk of repeating some of the points of our previous work,[32] it is important to try to relate Figure 2 to other descriptions of the NAC proposal. The change of $\langle R \rangle$ in the protein, which has been the main definition of the NAC effect, seems to us as an RSD hypothesis (the protein pushes the reacting fragments to closer distances than in water). Our view, of what can be understood by the NAC proposal, is strengthened by the fact that the NAC proposal has been frequently related to the rate acceleration in cyclic anhydride formation.[41] The rate acceleration in anhydride with a shorter chain is thought by some (e.g., ref 14) to reflect steric RSD or entropic effects. Now, to clarify what is actually meant by different versions of the NAC proposal, it is essential to try to formulate these proposals in terms of the corresponding free energies. Here, we believe that a consistent definition and analysis of the energetics associated with the structural changes has been provided in ref 32. It is also useful to formulate NAC

(39) Warshel, A.; Villà, J.; Štrajbl, M.; Florián, J. *Biochemistry* **2000**, *39*, 14728–14738.

(40) The fact that profiles of the form of Figure 2a are unlikely to exist in most enzymes does not present any conceptual or fundamental problem. This is exactly the profile expected from a protein with RSD and without sufficient binding by distant residues. Of course, the fact that proteins with positive binding energies are not found in most cases, provides a major support for the argument against the RSD proposals.[34]

(41) Lightstone, F. C.; Bruice, T. C. *J. Am. Chem. Soc.* **1996**, *118*, 2595–2605.

proposals by comparing the energies of the reaction in the protein and in solution, as is done in Figure 2. Nevertheless, it is important to consider other representations of the NAC effects. Unfortunately, the very limited alternative descriptions of the energetics of the NAC effect are inconsistent. That is, the first attempt to quantify the energetics of the NAC effect in terms of the difference between the enzyme and the solution profiles was provided by Kollman and co-workers.[42−44] This description used diagrams of the type shown in Figure 3a (see Figure 5 of ref 43 and Figure 4 of ref 42), which look at a first sight like a TSS proposal (see caption of Figure 3a). However, these workers actually described using their TSS diagram a destabilization of the RS by entropic effects (for a clear definition and relation to the assumed NAC effect, see footnote 45). Now, a correct description of the entropic proposal would involve an increase in the height of the protein free energy profile, leading to a RSD profile of the type shown in Figure 2a (see ref 46). At any rate, since Kollman and co-workers attributed the NAC effect to entropic RSD (see footnote 45), we establish here that at least some workers considered the NAC effect in a way that should have been described by Figure 2a. The same type of diagram as the one used by Kollman and co-workers was also used recently by Hur and Bruice (Figure 1B, ref 16) in describing the NAC effect in CM. Although in this case the NAC effect was not attributed to clear RSD effects, such as entropy or steric effects, there is still an inconsistency. That is, the diagrams shown in Figure 3a here and in Figure 1B of ref 16 involve the same $\Delta g_w^{\ddagger}$ and $\Delta g_{cat}^{\ddagger}$ as in the diagram of Figure 3b, and thus, they represent a clear TSS case; as clarified in the caption of Figure 3, $\Delta g_w^{\ddagger}$ and $\Delta g_{cat}^{\ddagger}$ are independent of the shape of the surface or the proposed NAC energy and position. There is nothing wrong with Figure 1B of ref 16, but since it simply represents a TSS proposal, it seems to contradict statements of Bruice and Hur who argued against the TSS proposal (e.g., refs 15 and 16).

In general, it is hard to describe the NAC effect as a new proposal, which is different than our early electrostatic TSS proposal,[47] unless it is considered a RSD proposal of the type presented in Figure 2b. If the NAC is thought of as a TSS proposal, then we are back to the fundamental question of what the origin is of the TSS, which is obviously not strain or confinement effects (these are RSD effects). In this case, if the NAC proposal is considered as an electrostatic TSS proposal, then the question boils down to whether the NAC is the reason for the TSS or its result. At any rate, it is essential to conduct a careful analysis to explore whether the catalytic effect in CM reflects RSD or TSS.
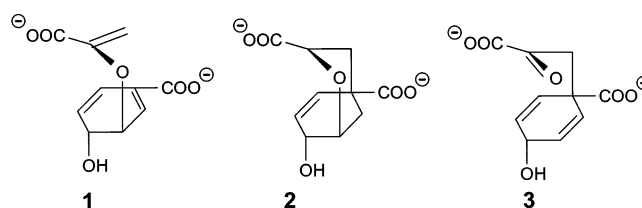

Figure 4. Valence bond resonance structures used in the EVB description of the reaction of Figure 1. Note that the VB diagram designates bonding patterns rather than a specific geometrical structure.

Table 1. Calculated and Observed Activation Free Energies for the Reaction of CM and the Reference Reaction in Water[a]

	$\Delta g_{calc}^{\ddagger}$	$\Delta g_{obs}^{\ddagger}$
water[b]	25.1	24.5[c]
protein	14.8	15.4[d]

[a] Energies in kcal/mol. The calculated values were obtained by the EVB approach with the parameters given in the Supporting Information. [b] Corresponds to 1 M concentration of the substrate. Note, in this respect, that since the reacting fragments are connected covalently to each other through the ring skeleton, we do not need to evaluate the activation barrier at a solvent cage, $\Delta g_{cage}^{\ddagger}$ (e.g., as was done in ref 32). $\Delta g_{calc}^{\ddagger}$ refers to calculated $\Delta g_w^{\ddagger}$ and $\Delta g_{cat}^{\ddagger}$ in the water and protein reaction, respectively. [c] Obtained from refs 8 and 63. [d] Obtained from ref 67.

3. Methods

The basis of our simulation is the empirical valence bond (EVB) approach,[33,48] which has been used extensively by us, and others (e.g., refs 49−53). Here, we described the Claisen rearrangement reaction by considering the three VB structures described in Figure 4 where **1**, **2**, and **3** correspond to the reactant state (i.e., chorismate), the TS, and the product state (i.e., prephenate), respectively. In the present case, we refined the EVB using ab initio charges as well as experimental information about the activation barrier in solution and for the exothermicity of the reaction in solution (see section 5 for more details). The corresponding parameters are given in Table 1 of the Supporting Information. The free energy of the EVB surface was evaluated by the umbrella sampling/free energy perturbation (UM/FEP) method.[33] The simulations were subjected to the surface constraint all atom solvent (SCAAS) model and its special polarization constraints[54] and long-range effects were treated by the local reaction field (LRF) long-range treatment.[55]

Another feature in the present study is the use of the linear response approximation (LRA) treatment[56] for the analysis of the contribution to the free energy of the reactants, transition states, and products. This approach provides a good estimate for the free energy associated with the change between two potential surfaces (U_1 and U_2) by

$$\Delta G(U_1 \rightarrow U_2) = (1/2)(\langle U_2 - U_1 \rangle_1 + \langle U_2 - U_1 \rangle_2) = (1/2)(\langle \Delta U \rangle_1 + \langle \Delta U \rangle_2) \quad (2)$$

where $\langle \Delta U \rangle_i$ designates an average over trajectories propagated on U_i. The microscopic LRA approach will be used here for several tasks including calculations of binding free energies according to the approach described in ref 57

(42) Stanton, R. V.; Peräkylä, M.; Bakowies, D.; Kollman, P. A. *J. Am. Chem. Soc.* **1998**, *120*, 3448−3457.
(43) Kollman, P. A.; Kuhn, B.; Donini, O.; Perakyla, M.; Stanton, R.; Bakowies, D. *Acc. Chem. Res.* **2001**, *34*, 72−79.
(44) Kollman, P. A.; Kuhn, B.; Perokyla, M. *J. Phys. Chem. B* **2002**, *106*, 1537−1542.
(45) Some readers who would start with ref 43 might be confused by our statement about the description of the NAC effect by Kollman and co-workers. However, one should realize that the analysis of the NAC effect by this group started with the clear analysis of ref 42 where the entropic effect (which they called a cratic effect) was added to the free energy profile of the reaction in water. The same approach was used in all of their explicit calculations and summarized in ref 43. Subsequently, ref 44 used the same diagrams to describe the NAC effect, while still considering the same calculations (see Table 1 of ref 44) that were based on the analysis of entropic effects.
(46) Warshel, A.; Parson, W. W. *Q. Rev. Biophys.* **2001**, *34*, 563−670.
(47) Warshel, A. *Proc. Natl. Acad. Sci. U.S.A.* **1978**, *75*, 5250−5254.

(48) Åqvist, J.; Warshel, A. *Chem. Rev.* **1993**, *93*, 2523−2544.
(49) Bala, P.; Grochowski, P.; Lesyng, B.; McCammon, J. A. *J. Phys. Chem.* **1996**, *100*, 2535−2545.
(50) Chang, Y.-T.; Miller, W. H. *J. Phys. Chem.* **1990**, *94*, 5884−5888.
(51) Kim, Y.; Corchado, J. C.; Villà, J.; Xing, J.; Truhlar, D. G. *J. Chem. Phys.* **2000**, *112*, 2718−2735.
(52) Schmitt, U. W.; Voth, G. A. *J. Phys. Chem. B* **1998**, *102*, 5547−5551.
(53) Vuilleumier, R.; Borgis, D. *Chem. Phys. Lett.* **1998**, *284*, 71−77.
(54) King, G.; Warshel, A. *J. Chem. Phys.* **1989**, *91*, 3647−3661.
(55) Lee, F. S.; Warshel, A. *J. Chem. Phys.* **1992**, *97*, 3100−3107.
(56) Lee, F. S.; Chu, Z. T.; Bolger, M. B.; Warshel, A. *Protein Eng.* **1992**, *5*, 215−228.

In addition to the fully microscopic models, we used the semimacroscopic PDLD/S-LRA approach for calculations of binding free energies. This approach, which is described in detail in several works (e.g., ref 57), provides a reasonable estimate for binding free energies of highly charged systems.

Our study was conducted by using the CM crystal structure from *Bacillus subtilis*, BsCM (PDB 1COM[58]) as a starting point. The EVB and LRA simulations were done using the ENZYMIX and POLARIS modules of the program MOLARIS.[59,60] The SCAAS radius for the explicit region was taken as 22 Å. The Claisen rearrangement reaction was described by mixing the zero order VB states, which are described schematically by the bonding patterns of **1**, **2**, and **3** in Figure 4. The selection of proper VB states and off diagonal terms (see Supporting Information) allows one to generate a ground-state model that reproduces the ab initio charge distribution of the RS and TS. The model was further parametrized to reproduce the observed activation barrier for the reaction in solution. The free energy surface for the reaction was mapped in two steps; first, from state **1** to **2** as a function of the reaction coordinate $\epsilon_2 - \epsilon_1$ and then from state **2** to **3** with the reaction coordinate $\epsilon_3 - \epsilon_2$. The FEP mapping of each step was typically evaluated after 50 ps equilibration time, followed by 21 windows of 5 ps each for moving along the reaction coordinate. The LRA calculations involved 50 ps on each state. All the simulations were done at 300 K with a 1 fs time step. Several initial conditions were used in each cas to ensure the stability of the results and to obtain a proper average over the configurations of the system both in the protein and in the solution.

The simulations considered Arg7, Glu78, Arg90, and Arg116 of chain C to be explicitly ionized. The effect of other ionized residues was considered macroscopically after the microscopic calculations. This was done by using the effective dielectric constant of ref 61 for charge–charge interactions. This dielectric constant is a function of distance and is usually larger than 20. The justification of our approach is given elsewhere.[62]

4. Evaluating the Activation Free Energies for the Reaction of CM and the Corresponding Reference Reaction in Solution

To exploit the power of the EVB method in quantitative studies of enzymatic reactions, it is essential to fit the EVB free energy profile of the reference reaction in water so that it will reproduce key experimental and theoretical information known about this reaction. Here, we used $(\Delta g_w^{\ddagger})_{obs} = 24.5$ kcal/mol based on the experimental information of ref 63. Since the reacting fragments are connected covalently by the ring skeleton, we do not need to define a "cage" reference system for the uncatalyzed reaction (see refs 32 and 64 for discussion of the cage complex). We also note that the equilibrium between the pseudodiequatorial and the pseudodiaxial conformational states of the chorismate molecule does not involve large energy (0.9–1.4 kcal/mol according to the estimate of Copley and Knowles[8]). This finding is supported by recent PMF studies.[65,66]

(57) Sham, Y. Y.; Chu, Z. T.; Tao, H.; Warshel, A. *Proteins: Struct., Funct., Genet.* **2000**, *39*, 393–407.
(58) Chook, Y. M.; Gray, J. V.; M., K. H.; Lipscomb, W. N. *J. Mol. Biol.* **1994**, *240*, 476–500.
(59) Lee, F. S.; Chu, Z. T.; Warshel, A. *J. Comput. Chem.* **1993**, *14*, 161–185.
(60) Chu, Z. T.; Villa, J.; Štrajbl, M.; Schutz, C. N.; Shurki, A.; Warshel, A., in preparation.
(61) Schutz, C. N.; Warshel, A. *Proteins: Struct., Funct., Genet.* **2001**, *44*, 400–417.
(62) Sham, Y. Y.; Chu, Z. T.; Warshel, A. *J. Phys. Chem. B* **1997**, *101*, 4458–4472.
(63) Andrews, P. R.; Smith, G. D.; Young, I. G. *Biochemistry* **1973**, *12*, 3492–3498.
(64) Štrajbl, M.; Florian, J.; Warshel, A. *J. Phys. Chem. B* **2001**, *105*, 4471–4484.
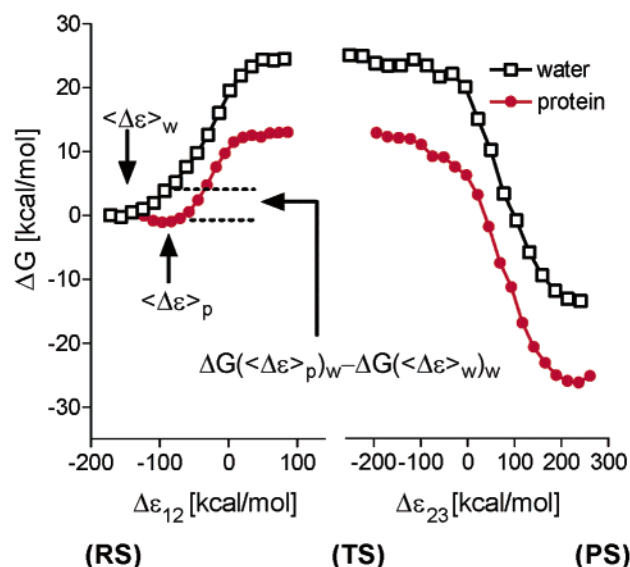
Figure 5. EVB free energy profile for the reaction in water and in CM. The EVB mapping is done from the TS to the reactant and product state, respectively. The two mapping steps are done along two different reaction coordinates ($\Delta\epsilon_{12}$ and $\Delta\epsilon_{23}$, respectively). Notation RS, TS, PS designates the reactant state, transition state, and product state.

The reaction free energy, ΔG_0, was approximated by the corresponding observed reaction enthalpy, $\Delta H_0 = -13.3$ kcal/mol, in the absence of a direct experimental measurement of the corresponding ΔG_0. It is important to note in this respect that none of our conclusions depend on the value of ΔG_0.

The experimental $\Delta g_{cat}^{\ddagger}$ was taken from the corresponding value of CM from *B. subtilis* (BsMC), where $(\Delta g_{cat}^{\ddagger})_{obs} = 15.4$ kcal/mol.[67] The ability to reproduce these values without adjusting any special parameters was then used to validate our approach. The results of our simulations for the reaction in water and in the enzyme active site are summarized in Figure 5 and Table 1. The figure depicts the calculated reaction profiles for both the water and the protein reactions. These EVB free energy surfaces correspond to much more extensive and consistent configurational averaging than that reported in some quantum mechanical studies. The minimum of the free energy profile at the RS in the solvent cage involves a complete configuration sampling on the potential of the ground state EVB surface (the RS region reflects mainly the first VB state of Figure 4 and some mixing with the second state). This minimum does not represent a single geometrical structure but the result of an extensive exploration of the surface in the region populated by the different diaxial configurations. This point is illustrated by the geometry distributions of the RS of both the water and the protein systems in Figure 1S of the Supporting Information. The center of the distribution for the $O_7-C_3-C_4-O_{12}$ torsional angle is similar to that obtained in ref 11. However, the exact details are not so relevant for the NAC issue. Since the EVB/FEP calculations did not explore transitions to the diequatorial configuration, we estimated the free energy for this transition to be around −1 kcal/mol by considering the results of refs 65 and 66 and the NMR results of Copley and Knowles.[8] Note

(65) Repasky, M. P.; Guimarães, C. R. W.; Chandrasekhar, J.; Tirado-Rives, J.; Jorgensen, W. L. *J. Am. Chem. Soc.* **2003**, *125*, 6663–6672.
(66) Martí, S.; Andrés, J.; Moliner, V.; Silla, E.; Tunon, I.; Bertran, J. *Chem.—Eur. J.* **2003**, *9*, 984–991.
(67) Kast, P.; Asif-Ullah, M.; Hilvert, D. *Tetrahedron Lett.* **1996**, *37*, 2691–2694.

that these NMR results are not in conflict with more recent studies[68] since the diaxial state includes several configurations.

The main point that emerges from Figure 5 and Table 1 is the fact that our calculation reproduced the observed activation free energies and the observed catalytic effect. It is useful to note that using other EVB parameters, that gave a narrower TS region, led basically to the same results. Also note that the possible over-stabilization of the product state in the protein does not present any problem in the present analysis (see footnote 69). At any rate, our ability to reproduce quantitatively the observed barrier is a crucial step in our analysis since only methods that can reproduce the overall catalytic effect are expected to provide conclusive results in analyzing the individual contributions to this effect.

Examination of the free energy surfaces of Figure 5 can give an estimate of the apparent NAC effect. Here, the upper limit of eq 1 is obtained using the generalized reaction coordinate instead of only the solute contribution to the reaction coordinate. Using eq 1, we obtain from the figure an upper limit of $\sim$5 kcal/mol for the NAC effect (note that the solvent coordinate undergoes a larger change in water than in the protein active site and thus contributes to part of this overall 5 kcal/mol). This estimate is smaller than the estimate of 8.4 kcal/mol obtained by Hur and Bruice[16] for another form of the enzyme (from *Escherichia coli*). It should be noted in this respect that ref 16 used 30 ns direct MD on the reactant state rather than a free energy perturbation approach. As explained eloquently by Valleau and Torrie,[70] it is very hard to obtain the proper probability of being at a high-energy region by a direct MD. For this purpose, one has to use the umbrella sampling or related approaches (see analysis of the difference between direct MD estimates and proper sampling approaches in ref 71). It is also useful to note that the early calculations of Martí et al.[13] produced a very large jump in the energy of the solution reaction at the RS region. This jump can be interpreted as a NAC effect of about 15 kcal/mol. However, this result is an artifact of the use of a limited energy minimization (rather than a proper configurational sampling) for a solvated system with two strongly interacting charges. Energy minimizations of the type used in standard quantum mechanical studies cannot produce the proper solvent rearrangement in large cluster of water molecules and are unable to reproduce the large dielectric screening of polar solvents. This point can be easily verified by trying to reproduce the energetics of decarboxylic acids in solutions[72] by energy minimization approaches. At any rate, regardless of the possible discrepancies between our results and those of Hur and Bruice (which are much more reasonable than those of ref 13), it clearly seems that we have a considerable

Table 2. Calculated Binding Free Energies for the RS and TS of CM[a]

	LRA	PDLD/S-LRA[b]	observed[c]
ΔG_{bind}^{RS}	−13	−10	−5.6
ΔG_{bind}^{TS}	−27	−15	−14.6

[a] The binding free energies (kcal/mol) correspond to the free energy of moving the substrate from water (in 1 M standard state) to the enzyme active site. The calculated binding energies include an estimated entropic correction of 10.0 kcal/mol (estimated from restraint release entropy calculations of ligands of similar size[57]). Arg7, Glu78, Arg90, and Arg116 of chain C were charged explicitly. The LRA approach for calculations of binding free energies is described in ref 57. [b] The PDLD/S-LRA calculations were performed by the approach outlined in ref 57 and with an entropic correction of 10 kcal/mol (see above). [c] The observed RS binding free energy is calculated from K_m from ref 5, and the TS binding energy is evaluated by $\Delta G_{bind}^{TS} = \Delta G_{bind}^{RS} + \Delta g_{cat}^{\ddagger} - \Delta g_{w}^{\ddagger}$.

apparent NAC effect. Thus, the question is what is the catalytic significance of this effect? This issue will be explored below.

5. Exploring the Apparent NAC Effect

After validating the quantitative nature of our free energy calculations, we can start analyzing the apparent NAC effect. As a first step in this analysis, we performed microscopic LRA and PDLD/S-LRA calculations of the binding free energy of both the RS and the TS. The results are presented in Table 2. It should be noted that the calculated LRA binding energies increase significantly with the number of positively charged groups, which are treated explicitly (e.g., charging Arg63 of chain B, in addition to the other charged residues). This reflects the difficulties of obtaining full screening for charge–charge interactions by microscopic models (e.g., refs 73 and 74). In the present case, a part of this effect might reflect the possibility that Arg 116 of chain C and Arg63 of chain B are not protonated at the same time. Fortunately, the calculated $\Delta g_{cat}^{\ddagger}$ is only weakly dependent on treating Arg63 of chain B explicitly. At any rate, the main point is the fact that both the observed and the calculated binding energies are negative. Thus, the reactant state must be more stable in the enzyme than in water. Furthermore, it is seen that the binding energy of the RS is smaller than that of the TS. This point establishes clearly that CM acts by TSS and not by RSD. The two options of RSD and TSS are also illustrated in Figure 6 in terms of the relative position of the corresponding free energy surfaces. As seen from the figure, the actual situation corresponds to Figure 2b. This demonstrates that the classical steric confinement model is invalid. Furthermore, this establishes that the apparent NAC is not the reason for the catalysis.

In view of the above conclusions, we should focus on the analysis of the reason for the apparent NAC effect. As a starting point in this analysis we will examine in a quantitative way the nature of the reacting system. As seen from Figure 1, we have a system with two negative charges, which are covalently linked to the atoms that are involved in the bond making process. Thus, we can view the electrostatic contribution to the activation free energy as the free energy of bringing two negative charges together. Since the TS involves the two charges in close proximity, the enzyme stabilizes this configuration by providing

(68) Campbell, A. P.; Tarasow, T. M.; Massefski, W.; Wright, P. E.; Hilvert, D. *Proc. Natl. Acad. Sci.* **1993**, *90*, 8663−8667.
(69) The results of Figure 5 correspond to $\Delta G_0^p \cong -25$ kcal/mol in the enzyme. This might look as a major problem and as a case of product state inhibition. However, the problem is not so serious as it might look. That is, most probably we have here some overestimate of the charge–charge interaction with the protein because of the fact that microscopic simulations do not provide sufficient screening for charge–charge interactions (see section 5). However, the present results are more reasonable than previously reported ΔG_0^p (e.g., ref 10). More importantly, the overall ΔG_0^p reflects in part the very negative ΔG_0^w for the water reaction ($\Delta G_0^w \cong -13.3$ kcal/mol). Thus, the calculated product binding energy is only about 12 kcal/mol. While this is still an overestimate, it does not indicate a major problem with product inhibition.
(70) Valleau, J. P.; Torrie, G. M. *Modern Theoretical Chemistry*; Plenum Press: New York, 1977; Vol. 5.
(71) King, G.; Warshel, A. *J. Chem. Phys.* **1990**, *93*, 8682−8692.
(72) Warshel, A.; Russell, S. T. *Q. Rev. Biophys.* **1984**, *17*, 283−421.
(73) Burykin, A.; Schutz, C. N.; Villa, J.; Warshel, A. *Proteins: Struct., Funct., Genet.* **2002**, *47*, 265−280.
(74) Muegge, I.; Schweins, T.; Langen, R.; Warshel, A. *Structure* **1996**, *4*, 475−489.
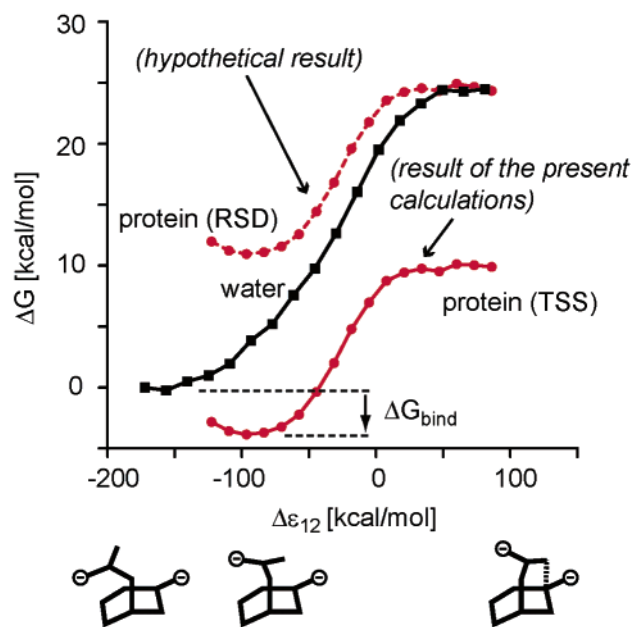
Figure 6. Illustrating the two options (RSD and TSS) for relating the protein and water free energy surfaces. As is found in the present work (and confirmed by the observed binding energy), the reaction of the CM should be described by the TSS option. Since the RSD case (dashed line) corresponds to a classical steric confinement model, we conclude that such a model is invalid. The molecular configurations at the lower part of the figure are only meant as a schematic representation. The open configuration, on the lower left, corresponds to an extensive sampling on all the diaxial conformations.

a complementary electrostatic environment. Figure 7 presents a TS configuration in a stereoview. As is apparent from the figure in the TS structure, the two positively charged groups Arg7 and Arg90 of chain C "bind" one of the carboxylates and the oxygen whose bond with the ring is broken (O_7 in the notation of Figure 1), whereas the other carboxylate is stabilized by Arg116 of chain C. As a result of this stabilization effect, the RS in the enzyme may also involve a closer distance between the charges, and this may lead to the apparent NAC effect.

To quantify the above hypothesis, we separated the electrostatic and steric effects in both CM and the corresponding reference reaction in water. This was done by performing three sets of calculations. The first set involved a full EVB, the second set involved omission of the electrostatic interaction between the two carboxylates and between the carboxylates and their surrounding environment, while in the third set all the charges and the residual charges of the substrate were set to zero. The results of this analysis are shown in Figure 8. As seen from the figure (comparing Figure 8a and 8b), the omission of the carboxylates' electrostatic contribution leads to a disappearance of a major part of the catalytic effect (the difference between the reaction barrier in water and the protein). The rest of the catalytic effect disappears when all residual charges of the substrate are set to zero (Figure 8c). Once all residual charges are set to zero (Figure 8c), the free energy profile for the reaction in CM is similar to that in water. Thus, it can be concluded that a major part of the difference between the reaction in CM and the reaction in water is due to electrostatic effects. This means that the contributions from steric and/or other nonelectrostatic effects are small.

Another way to demonstrate that the electrostatic effect is the origin of catalysis in this case is to examine the nature of

the LRA calculations for charging the RS and TS that were used in the binding calculations of Table 2. Such an analysis is summarized in Table 3. The table considers the average electrostatic interaction between the substrate and its surrounding for trajectories where the environment "sees" the substrate charges, $\langle \Delta U \rangle_Q$, and for trajectories where the environment does not "see" the charges $\langle \Delta U \rangle_0$. Consideration of the results in the framework of eq 1 establishes that the catalytic effect is due to a larger electrostatic stabilization of the TS. A large part of the electrostatic effect is due to the $\langle \Delta U \rangle_0$ term, which reflects the preorganization energy of the enzyme. In fact, as much as the binding is concerned we have a very large electrostatic preorganization contribution $(^1/_2)\langle \Delta U \rangle_0^p$. Furthermore, the preorganization effect also contributes significantly to the overall catalysis $((^1/_2)(\langle \Delta U_{TS} \rangle_0^p - \langle \Delta U_{RS} \rangle_0^p) \cong 4$ kcal/mol). However, in the present case there is also a large electrostatic contribution from the $\langle \Delta U \rangle_Q$ term. This is interesting since usually one expects most of the electrostatic contribution to the catalysis to be due to the $\langle \Delta U \rangle_0$ term. Here, however, since the charges of the RS and TS are quite similar (see, however, the Conclusion) and since $\langle \Delta U \rangle_0$ plays a major role in stabilizing these charges (in the binding of the RS and TS), the differential stabilization that leads to the catalytic effect is due, in part, to the $\langle \Delta U \rangle_Q$ term.

To further explore the apparent NAC effect, we considered the dependence of the free energy surface on the reaction coordinate. These calculations were done by the LRA approach while fixing both the $C_9 \cdots C_1$ distance and the distance between the carboxylates at three characteristic distances $\langle R \rangle_{TS}^p$, $\langle R \rangle_{RS}^p$, and $\langle R_{open} \rangle^w$ (where $\langle R_{open} \rangle^w$ is the average configuration at the extended diaxial configuration in water). We use the LRA approach, rather than the potential of mean force (PMF) approach, since we are interested in the *electrostatic* free energy of the system. The calculated LRA energies are presented in Figure 9 as a function of this approximated reaction coordinate, for both the reaction in water and in the protein active site. Comparing the electrostatic free energies of the water reaction at $\langle R_{open} \rangle^w$ and $\langle R \rangle_{RS}^p$ indicates that the solute contribution to the apparent NAC effect is not so large (around 3 kcal/mol). Also note that most of this increase in the free energy upon changing R^w from $\langle R_{open} \rangle^w$ to $\langle R \rangle_{RS}^p$ is due to the electrostatic effect. As discussed in section 4, we believe that the total apparent NAC effect, that includes the effect of the solvent coordinate (estimated by us to be between 5 and 3 kcal/mol), is significantly smaller than the estimate of ref 16. However, even a larger apparent NAC effect would not change our main conclusion that this effect is the result rather than the reason for catalysis. Thus, we do not find it essential to use the approach of ref 32 and to fully quantify the NAC effect.

6. Concluding Remarks

The present work examines the origin of the catalytic power of CM by systematic EVB and LRA analysis. It was found that the catalytic effect is almost entirely due to TSS by the electrostatic effect of the active site. This electrostatic stabilization leads to a reduction in the average $C_9 \cdots C_1$ distance relative to the corresponding distance in water. Although this change in equilibrium distance can be called a NAC effect, it is simply the *result* rather than the *reason* for the catalytic effect, side effect of TSS. That is, in analyzing enzyme catalysis it is very
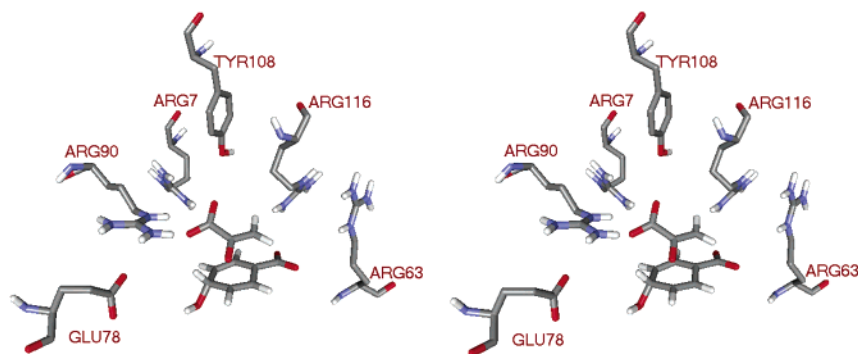
Figure 7. Active site configuration of CM from BsCM in the TS configuration, a cross-eyed stereoview. Note the stabilization of the two carboxylate groups by Arg7, Arg90, and Arg116 of chain C.
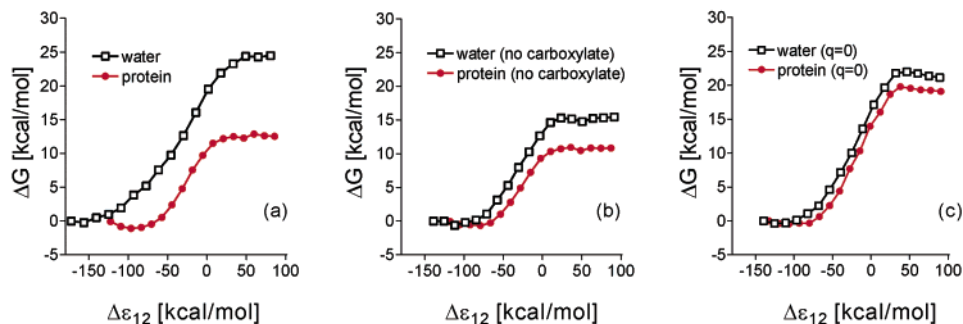

Figure 8. Free energy surfaces for the reaction in water and in CM for cases where (a) the complete system is included, (b) the charges of the carboxylate groups are set to zero, and (c) all the charges and residual charges of the whole substrate are set to zero.

Table 3. LRA Analysis of the Electrostatic Solvation Contributions to the Binding of the RS and TS in CM[a]

	water		protein	
	RS	TS	RS	TS
$\langle \Delta U \rangle_Q$	−388.8	−403.1	−369.0	−402.8
$\langle \Delta U \rangle_0$	0.0	0.0	−66.2	−74.4
ΔG_{LRA}	−194.4	−201.5	−217.6	−238.6

[a] All energies are given in kcal/mol. $\langle \Delta U \rangle$ designates the average of the electrostatic interaction between the substrate and its surroundings (water and protein). $\langle \Delta U \rangle_Q$ and $\langle \Delta U \rangle_0$ designate the corresponding average over a potential surface that includes fully charged substrate and nonpolar substrate. The calculated energies are converted to "solvation" energies by subtracting the corresponding values of $\langle \Delta U \rangle_0$ in water.

important to determine what factors actually lead to the overall catalytic effect and to distinguish these factors from other effects that result from the catalytic factors. For example, if an enzyme operates by electrostatic stabilization and this electrostatic effect changes the color of the substrate, we will have to recognize the color change as the result and not as the reason for the catalytic effect.

The present work is consistent with previous studies, where it was found that enzymes catalyze their reactions by TSS.[34] Namely, that the reason for the reduction of $\Delta g_{cat}^{\ddagger}$ is the TS electrostatic stabilization by the enzyme.[34,47,75] In some cases, $\langle R \rangle_{RS}^{p} - \langle R \rangle_{RS}^{w}$ may be correlated with the catalytic effects, but this will happen only when the charge distribution of the RS and TS are similar. However, as found in the present work, this will be the result rather than the reason for catalysis.

The case of CM presents an excellent opportunity to explore the nature of enzyme catalysis since the reaction region does not include any of the enzyme groups. This makes it particularly
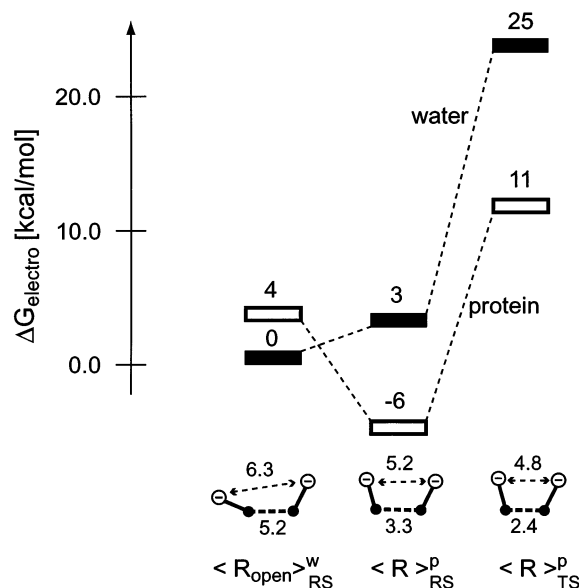
(75) Feierberg, I.; Aqvist, J. *Theor. Chem. Acc.* **2002**, *108*, 71−84.

Figure 9. LRA estimate of the electrostatic free energy for several points along the reaction coordinate. Energies in kcal/mol are indicated over the corresponding bars. Distances in angstroms are given for the separation between the carboxylate charge centers (designated by (−)) and for the distances between C_1 and C_9. The relative position of the protein and water profiles are set in a way that the binding energy at $\langle R \rangle_{RS}^{p}$ will correspond approximately to the observed ΔG_{bind} (−5.6 kcal/mol). The TS free energy in water includes a constant term that reproduces the corresponding observed value (since the LRA electrostatic contribution does not include the intermolecular activation energy). The same constant is used for the protein TS.

simple to determine whether we have RSD or TSS (for more complex cases, see ref 39). Now, early theoretical studies[12,14] implied clearly that CM works by RSD. Other studies[16] suggested that CM works mainly by RS effects, while some

works[76] supported TSS. It should also be noted that experimental studies have supported the TSS proposal.[3] Fortunately, the issue of RSD or TSS is not a semantic issue but a very well-defined problem. That is, if one uses a complete thermodynamic cycle with a clear reference state (e.g., the reactants in water) and actually calculates the binding of the RS, then the problem can be resolved. Indeed, the calculations of the present work established that CM works by electrostatic TSS. Now, since CM does not work by RSD, it is clear that the apparent NAC effect is not similar to the steric effect observed in model compounds[14,41] and is not the reason for the rate acceleration.

The suggestion that steric effects play an important role in the catalytic effect of CM[14,15] has paid particular attention to Val 135 in the *E. coli* enzyme and Leu 115 in the *B. subtilis* enzyme. These residues were assumed to play an important role in pinning the substrate to a correct catalytic configuration[14] or increasing the NAC population without a strong steric strain. The present work shows that the pure nonelectrostatic steric effect does not contribute to catalysis. In reaching this conclusion, we rely on the analysis presented in Figure 8, where we basically obtain the same results for protein and water once we omit the electrostatic interactions. Of course, the nonpolar groups might help in providing the correct arrangement for maximal electrostatic effect.

A recent work of Hilvert and co-workers[6] provided an extremely clear insight about the origin of the catalytic effect in CM. This work demonstrates that the electrostatic effect of Arg90C amounts to about 5.9 kcal/mol and contributes only 0.6 kcal/mol to the binding energy. While the contributions of the different residues are not additive, we have here a very large contribution that influences almost entirely the TS with a very small effect on the RS. This TSS effect reflects stabilizing the increase in charge separation of the actual reacting atoms rather than the direct interaction with the carboxylates. This type of effect is apparent in Figure 8, where we consider the effect of the enzyme on the substrate in the absence of the carboxylates. A similar conclusion about the importance of the stabilization of the nonionized part of the substrate was deduced in our recent study of catalytic antibodies (CAs).[7] It was shown that CAs that are designed to catalyze the reaction of CM are significantly less effective than CM due, in part, to the effect of the nonionized part.

It is important to consider here the idea that enzymes compress or mold the reacting fragments to a configuration that resembles their TS.[12,14,28,31,77,78] The TS is defined by both the solute and the "solvent" coordinates.[79−81] That is, even when some solute coordinates look close to their TS, the system is usually very far from the real TS. Here, one must look on the distance in terms of energy rather than structure. For example, if we define the solute reaction coordinate in CM as the $C_9 \cdots C_1$ distance (a more general definition should include also the distance between the two carboxylates), then $\langle R \rangle_{RS} \approx 3.6$ Å and $\langle R \rangle_{TS} \approx 2.1-2.4$ Å. This small difference of 1.2−1.5 Å is very costly in terms of energy. This is one of the reasons for the importance of using a generalized solute−solvent coordinate ($\Delta\epsilon$ in Figures 5−7) as the reaction coordinate. With this coordinate, it is quite clear that the RS is quite far from the TS. Thus, it is also useful to note that in most cases the main effect of the enzyme is to push the solvent coordinate toward the TS. This solvent preorganization effect[33,47] has nothing to do with the NAC effect, which considers only the solute coordinate. The enzyme electrostatic TSS also leads to the significant changes in the solute contribution to the reaction coordinate only when the solute charges of both the RS and the TS are similar.

It is useful to comment here on the option that the carboxylate groups of the substrate can be considered as the nonreactive part of the substrate in a RSD mechanism, which supposedly uses the binding energy to strain the reactive part.[38] First, as mentioned above, it is unjustified to consider the carboxylates as being separated from the chemical part since carboxylates are practically bonded to the reacting atoms. Second, in the RSD mechanism the binding of the distant groups is supposed to destabilize the binding of the chemical part in the RS. However, our calculations of the binding of the substrate without the charges of the carboxylates gave the same free energies (about −4 kcal/mol) as the increase in binding energy upon making the substrate (without the carboxylates) polar, in the presence of the charged carboxylates. Note that the difference between these two values approximates the effect of the carboxylates on the rest of the system and thus represents their supposed effect as distant groups. Finally, the general RSD idea appeared to be in a clear contrast to the results of the above-mentioned experiments of Hilvert and co-workers.[6] The fact that the mutation of Arg90C only reduced the TS stabilization while leaving the RS unchanged indicates that Arg90C stabilizes the chemical part in the TS rather than destabiling it in the RS.

A recent study of Jorgensen and co-workers that used a QM/MM AM1 model reported a reasonable difference between the PMFs of reaction in CM and the corresponding water reaction.[82] Unfortunately, the analysis presented involved significant inconsistencies. That is, although these workers obtained a large transition state stabilization (enzyme relative to water), they concluded that the TSS plays a secondary role and that the catalysis is due to "conformational compression of the NAC by the enzyme". This seems at odds with the fact that ref 82 found a small energy contribution for the formation of the NAC in water. The confusion may reflect unfamiliarity with the proper thermodynamic cycle in enzyme catalysis and with the definitions of TSS, RSD, and the NAC effect (for problems with the NAC definition, see ref 83). More specifically, ref 82 considered 8.8 kcal/mol gas-phase energy for moving between the minima of the PMF in water (our $\langle R \rangle_{RS}^w$) and in the protein (our $\langle R \rangle_{RS}^p$) as a major contribution to the calculated change in activation barrier upon moving from water to the protein state. Unfortunately, the gas-phase contributions cannot (and should not) be used in studies of the difference between the reaction

(76) Martí, S.; Andrés, J.; Moliner, V.; Silla, E.; Tunon, I.; Bertran, J.; Field, M. J. *J. Am. Chem. Soc.* **2001**, *123*, 1709−1712.
(77) Ford, L. O.; Johnson, L. N.; Machin, P. A.; Phillips, D. C.; Tijian, R. *J. Mol. Biol.* **1974**, *88*, 349−371.
(78) Castillo, R.; Andrés, J.; Moliner, V. *J. Am. Chem. Soc.* **1999**, *121*, 12140−12147.
(79) Hwang, J.-K.; King, G.; Creighton, S.; Warshel, A. *J. Am. Chem. Soc.* **1988**, *110*, 5297−5311.
(80) Kim, H. J.; Hynes, J. T. *J. Am. Chem. Soc.* **1992**, *114*, 10508−10537.
(81) Kurz, J. L.; Kurz, L. C. *Isr. J. Chem.* **1985**, *26*, 339−348.

(82) Guimarães, C. R. W.; Repasky, R. M.; Chandrasekhar, J.; Tirado-Rives, J.; Jorgensen, W. L. *J. Am. Chem. Soc.* **2003**, *125*, 6892−6899.
(83) The work of Jorgensen and co-workers[82] seems to represent a misunderstanding of the NAC proposal. That is, these workers used very different NAC geometries in the protein and in water. These geometries were chosen more or less as the minima of the corresponding PMFs. Actually, the minimum of the protein PMF is the most logical definition of the NAC geometry in the water reaction (see ref 32). Otherwise the NAC proposal has little to do with the effect of the protein.
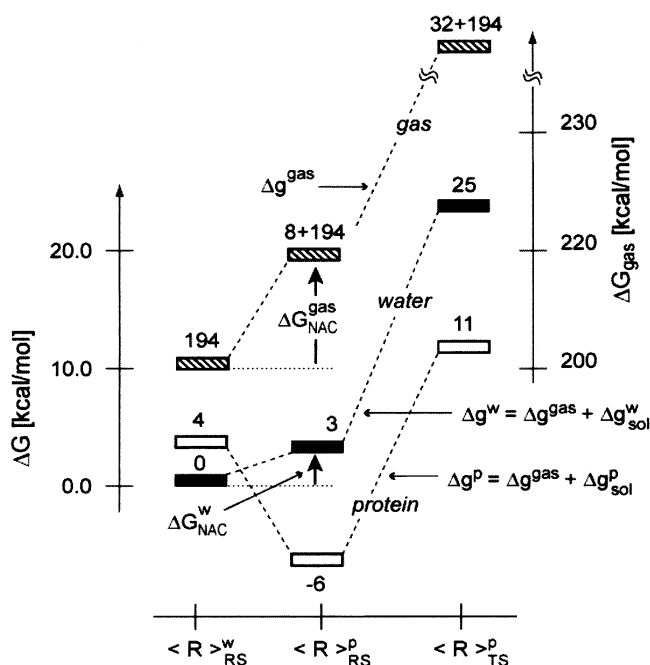
Figure 10. Relationship between the free energy surfaces for the reaction of CM in the gas phase, solution, and protein. The figure describes schematically the relationship between the reaction profile in the gas phase and the corresponding profiles in water and protein. The energy contributions are given in kcal/mol, and the gas-phase surface corresponds to a much higher energy that is designated on the right side of the figure. The value 194 kcal/mol is the negative of the solvation energy in water ($-\Delta G_{sol}^{w}$) at $\langle R \rangle_{RS}^{w}$. As is clear from the scheme, the value of ΔG_{NAC}^{gas} does not contribute to catalysis since it provides an identical contribution in both environments. Note that the possibility of having somewhat different reactions coordinate in different environments is not going to change our conclusions.

in the protein and in solution, and the analysis of ref 82 is simply inconsistent. To clarify the above problems and to prevent a possible confusion by the elaborated free energy diagram of Figure 7 of ref 82, we provide in Figure 10 a clear and consistent analysis of the relevant energetics of the reaction in the gas phase, in solution, and in the protein, along the same coordinate. We note in this respect that the relatively small difference in the reaction coordinate, for the different phases, is neglected here to clarify the main conceptual points. Now, as seen from the Figure 10, there is, indeed, a large energy contribution for moving from $\langle R \rangle_{RS}^{w}$ to $\langle R \rangle_{RS}^{p}$ on the gas-phase surface. However, this energy (ΔG_{NAC}^{gas}), which is mainly due to the electrostatic repulsion between the carboxylates, is largely screened in solution and in the protein; the corresponding contribution (ΔG_{NAC}^{w} in the notation of Figure 10 and ΔG_{NAC} in the notation of eq 1) is around 3 kcal/mol in water according to the PMF calculations of ref 82. Thus, the energy of deforming the substrate in the gas phase cannot be used in analyzing enzyme catalysis, unless one considers the energy of solvating the substrate (this will give the Δg^{w} and Δg^{p} of Figure 10). In fact, as clarified in Figure 10, the same 8 kcal/mol contribution

for moving from $\langle R \rangle_{RS}^{w}$ to $\langle R \rangle_{RS}^{p}$ in the gas phase (ΔG_{NAC}^{gas}) contributes both to water and to protein profiles. Thus, the gas-phase contribution leads to no catalytic effect. The actual catalysis comes from the difference in the interaction between the substrate and its environment in the solution and in the enzyme. In summary, the problem with the conclusions of ref 82 is not in the calculations presented in that paper but with the incorrect analysis of the corresponding results.

The conclusion of a recent instructive theoretical study by Martí et al.[66] appears to be similar to some of the conclusions reached in the present work. However, we would like to clarify that ref 66 did not attempt to evaluate binding energies, and therefore, did not examine the issue of RSD versus TSS. Furthermore, ref 66 did not provide a quantitative decomposition of the catalytic free energy to electrostatic and steric contributions, which is one of the main points of the present work. At any rate, it is instructive to comment here on the conclusion of Martí et al. that "both reorganization and preorganization effects have to be considered as the two faces of the same coin".[66] To clarify this statement, it is important to note that Martí et al. chose to call our enzyme preorganization effect (which reduces the enzyme reorganization energy) a "reorganization effect" and call "preorganization effect" the enzyme effect on the substrate, which includes the NAC effect. The point is that the enzyme preorganization effect and the NAC effect are not two faces of the same coin. The enzyme preorganization is the reason for catalysis, which in some special cases also leads to an apparent NAC effect.

Finally, we would like to make a general point. At present, all consistent studies indicate that enzymes work by TSS rather than by RSD. Studies that suggested RSD mechanisms either involved an incorrect reference state (see, for example, discussion in ref 39) convergence of the calculated binding free energies (see discussion in ref 84). Now since enzyme catalysis involves TSS and since this TSS is due to electrostatic preorganization, it is preferable to calculate the electrostatic stabilization effect and to obtain a clear estimate of the catalytic effect rather than to evaluate the NAC effect that might or might not help in predicting the catalytic effect (see discussion in ref 33).

Supporting Information Available: Table of the EVB parameters, a figure of the VB structures, and a figure of geometry distributions of the RS in water and in protein. This material is available free of charge via the Internet at http://pubs.acs.org.

JA0356481

(84) Warshel, A. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 425−443