J|A|C|S
ARTICLES
Published on Web 03/20/2002

How Much Do Enzymes Really Gain by Restraining Their Reacting Fragments?

A. Shurki,[†] M. Štrajbl,[†] J. Villà,[‡] and A. Warshel*,[†]

Contribution from the Department of Chemistry, University of Southern California, Los Angeles, California 90089-1062

Received September 24, 2001

Abstract: The steric effect, exerted by enzymes on their reacting substrates, has been considered as a major factor in enzyme catalysis. In particular, it has been proposed that enzymes catalyze their reactions by pushing their reacting fragments to a catalytic configuration which is sometimes called near attack configuration (NAC). This work uses computer simulation approaches to determine the relative importance of the steric contribution to enzyme catalysis. The steric proposal is expressed in terms of well defined thermodynamic cycles that compare the reaction in the enzyme to the corresponding reaction in water. The S_N2 reaction of haloalkane dehalogenase from *Xanthobacter autotrophicus* GJ10, which was used in previous studies to support the strain concept is chosen as a test case for this proposal. The empirical valence bond (EVB) method provides the reaction potential surfaces in our studies. The reliability and efficiency of this method make it possible to obtain stable results for the steric free energy. Two independent strategies are used to evaluate the actual magnitude of the steric effect. The first applies restraints on the substrate coordinates in water in a way that mimics the steric effect of the protein active site. These restraints are then released and the free energy associated with the release process provides the desired estimate of the steric effect. The second approach eliminates the electrostatic interactions between the substrate and the surrounding in the enzyme and in water, and compares the corresponding reaction profiles. The difference between the resulting profiles provides a direct estimate of the nonelectrostatic contribution to catalysis and the corresponding steric effect. It is found that the nonelectrostatic contribution is about −0.7 kcal/mol while the full "apparent steric contribution" is about −2.2 kcal/mol. The apparent steric effect includes about −1.5 kcal/mol electrostatic contribution. The total electrostatic contribution is found to account for almost all the observed catalytic effect (∼−6.1 kcal/mol of the −6.8 calculated total catalytic effect). Thus, it is concluded that the steric effect is not the major source of the catalytic power of haloalkane dehalogenase. Furthermore, it is found that the largest component of the apparent steric effect is associated with the solvent reorganization energy. This solvent-induced effect is quite different from the traditional picture of balance between the repulsive interaction of the reactive fragments and the steric force of the protein.

1. Introduction

The molecular origin of enzyme catalysis is a problem of major fundamental and practical importance. Biochemical and structural studies have provided the groundwork for tackling this problem (e.g., ref 1). Yet, discrimination between different proposals for the source of enzyme catalysis still requires quantitative structure−function correlation studies (e.g., ref 2). Furthermore, even the advance of computer simulation approaches for studies of enzymatic reactions (for recent review, see ref 3) has not yet provided a consensus in the field.

Here we address the proposal that enzymes exert on their reactants some form of reactants state (RS) strain, and thus "compress" or "mold" the reacting fragments to a configuration that resembles the transition state, TS, (e.g., refs 4−9). Unfortunately, the processes that have been studied do not appear to be directly relevant to the strain proposal. Moliner et al.[8] for example, compared the QM/MM structure of the TS in the

* To whom correspondence should be addressed. E-mail: warshel@usc.edu.
[†] University of Southern California.
[‡] Present address: Grup de Recerca en Informàtica Biomèdica IMIM/UPF C/Doctor Aiguader, 80 08003 Barcelona, Spain.

(1) Fersht, A. *Structure and Mechanism in Protein Science. A Guide to Enzyme Catalysis and Protein Folding*, 2nd ed.; W. H. Freeman and Company: New York, 1999.
(2) Warshel, A. *Computer Modeling of Chemical Reactions in Enzymes and Solutions*; John Wiley & Sons: New York, 1991.
(3) Villà, J.; Warshel, A. *J. Phys. Chem. B* **2001**, *105*, 7887−7907.
(4) Ford, L. O.; Johnson, L. N.; Machin, P. A.; Phillips, D. C.; Tjian, R. *J. Mol. Biol.* **1974**, *88*, 349−371.
(5) Khanjin, N. A.; Snyder, J. P.; Menger, F. M. *J. Am. Chem. Soc.* **1999**, *121*, 11831−11846.
(6) Tapia, O.; Andrés, J.; Safont, V. S. *J. Chem. Soc., Faraday Trans.* **1994**, *90*, 2365−2374.
(7) Castillo, R.; Andrés, J.; Moliner, V. *J. Am. Chem. Soc.* **1999**, *121*, 12140−12147.
(8) Moliner, V.; Andrés, J.; Oliva, M.; Safont, V. S.; Tapia, O. *Theor. Chem. Acc.* **1999**, *101*, 228−233.
(9) Martí, S.; Andrés, J.; Moliner, V.; Silla, E.; Tuñón, I.; Bertrán, J. *J. Phys. Chem. B* **2000**, *104*, 11308−11315.

carboxylation step in ribulose-1,5-bisphosphate carboxylase/oxygenase, (rubisco), to the corresponding TS in gas phase, and found them to be similar. However, they did not explore the RS structure nor the effect of the protein on this state. Another study[7] proposed, on the basis of QM/MM calculations, that dihydrofolate reductase (DHFR) compresses the substrate to configurations which are close to the TS structure. The authors found the $C_6\cdots C_4$ distance to be ~3.3 Å or 3.8 Å in the RS, (depending on the method used), compared to ~2.7 Å in the TS. However, it is not clear that the 3.3 Å distance represents a strained ground state structure.

More specifically, refs. 4−9 and related studies have invoked the proposal that the enzyme "molds" the reacting system to the TS structure, but did not examine whether the enzyme can deform the encounter complex significantly from its reactants state configuration in solution, nor consider the energy cost of such deformation. Computer modeling studies[10] and free energy calculations[11] have indicated that enzymes are quite flexible and cannot exert the very large strain needed to push the substrate significantly toward its TS structure.

A more compelling version of the strain concept is the so-called near attack conformation (NAC) hypothesis, which was suggested by Bruice and co-workers.[12−17] The NAC hypothesis implies that the enzyme reduces the activation barrier by restricting the configurational space of the substrate in the reactants state. This catalytic effect might reflect either enthalpic or entropic contributions. The NAC hypothesis has not been related rigorously to the corresponding difference between the activation barrier of the enzyme and solution reaction, thus making it hard to examine its validity in a quantitative way (see ref 18 for a discussion of the difficulties with the current definition).

Bruice and co-workers have made an important attempt to establish the NAC hypothesis by MD studies of the reactant states in formate dehydrogenase,[19] haloalkane dehalogenase[13,16,20] and catechol *O*-methyltransferase,[21] and have concluded that the enzyme helps in increasing the population of the NAC state. Attempts to evaluate the energy cost of the NAC effect have been restricted to a gas-phase study of haloalkane dehalogenase.[16] This study compared the ab initio energies of gas-phase optimized RS and TS geometries to the corresponding gas-phase energies of an estimated NAC configuration. However, the calculations did not use the same computational model for the two systems, and the energy difference is expected to decrease significantly in solution (see ref 22 for discussion). Studies of Kollman and coworkers (e.g. ref. 26) that seem to support the NAC proposal has been based on inconsistent thermodynamic analysis which will be considered in the discussion section (see also refs. 3 and 30).

Other attempts to support the NAC and related proposals have been based on model compounds.[5,12,14,27,28] One study forced the reacting atoms to be at a close contact and demonstrated that a strong compression of the critical C···C distance leads to a drastic reduction of the activation barrier.[5] However, as noted above, it is unlikely that enzymes can apply strong strain effects.[10,11] The relationship between chain length and rate constant in cyclic anhydride formation was used frequently to support the idea of steric or entropic effects in enzyme catalysis.[12−14,27] Clearly, quantitative analysis of the behavior of such model compounds is a very important tool for validation of computational models. However, the relationship of nonenzymatic reactions of model compounds to enzymatic reactions is far from obvious.[2] Thus, the catalytic effect derived from these studies is not simply related to enzyme catalysis, and it is essential to examine this strain effect in the actual enzyme active site. Apparently, despite the interest in the NAC effect and related steric proposals, there has been no quantitative assessment of the magnitude of the NAC effect by either experimental or theoretical studies.

The present work attempts to obtain a quantitative estimate of the NAC effect, and to reach a general conclusion about the role of configurational restrictions on the substrate. The enzyme chosen for this purpose is haloalkane dehalogenase, (DhlA), from *Xanthobacter autotrophicus* GJ10. This enzyme, whose structure was solved by Verschueren et al. (PDB entry code 2DHD,[29]), was used by Bruice and co-workers in the above-mentioned studies of the NAC effect.[13,16] The involvement of DhlA in a simple S_N2 reaction makes it an excellent candidate for quantitative studies.

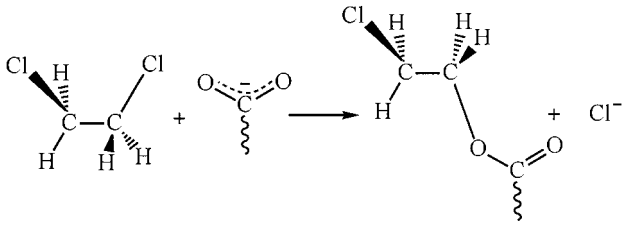
Section 2 defines the steric proposal in a way that makes it amenable to quantitative studies. Section 3 describes our approaches for evaluating the steric contribution to enzyme

(10) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227−249.
(11) Chapter 6 of ref 2.
(12) Lightstone, F. C.; Bruice, T. C. *J. Am. Chem. Soc.* **1996**, *118*, 2595−2605.
(13) Lightstone, F. C.; Zheng, Y. J.; Bruice, T. C. *J. Am. Chem. Soc.* **1998**, *120*, 5611−5621.
(14) Bruice, T. C.; Lightstone, F. C. *Acc. Chem. Res.* **1999**, *32*, 127−136.
(15) Benkovic, S. J.; Bruice, T. C. *Biochemistry* **2000**, *39*, 6267−6274.
(16) Lau, E. Y.; Kahn, K.; Bash, P. A.; Bruice, T. C. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 9937−9942.
(17) Bruice, T. C. *Acc. Chem. Res.* **2002**, *35*, (ASAP).
(18) The NAC effect has not been expressed by rigorous formulation, which is directly related to the relevant activation barriers (as is done in the present work). The definition of ref 17 implies that "NACs are characterized as having reacting atoms within 3.2 Å and an approach angle for reaction of 15° of the bonding angle in the transition state." Although this is a quantitative concept, the way the parameters (distance/angle) were chosen is somewhat arbitrary. Similarly, other points along the reaction coordinate could have been chosen. The closer the chosen NAC structure is to the TS the larger the energy difference between the enzyme and the water reaction will be. Thus, one can obtain a very large NAC effect by choosing the NAC configuration as the geometry of the reacting fragments in the TS of the enzyme. However, this will not tell us whether the resulting NAC effect is just a reflection of the TS stabilization, due to electrostatic effects, or GS destabilization due to steric effects. It should be noted that we did not present this example as a polemic exercise, but in order to emphasize the need for a definition that relates the NAC to transition-state theory.
(19) Torres, R. A.; Schiøtt, B.; Bruice, T. C. *J. Am. Chem. Soc.* **1999**, *121*, 8164−8173.
(20) Lightstone, F.; Zheng, Y.-J.; Maulitz, A.; Bruice, T. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 8417−8420.
(21) Lau, E. Y.; Bruice, T. C. *J. Am. Chem. Soc.* **2000**, *122*, 7165−7171.
(22) Ab initio calculations that we performed using B3LYP/6-31G(d) while comparing the RS energies of the optimized gas-phase configuration to a configuration similar to that in the enzyme (in both cases a 3 Å O···C distance was used) gave a 7 kcal/mol difference. However, in solution this difference dropped almost to zero with both PCM[23] and Langevin dipoles solvation model.[24,25] Furthermore, a more consistent analysis should describe the potential surface of the substrate by the same computational model in the gas phase and in the enzyme. The structure used in that study for the NAC configuration in the ab initio gas-phase calculation was obtained by MD simulations.[16,13] However, an ab initio QM/MM search for a NAC configuration would most likely result in a configuration very different from that obtained in ref 16 and the configuration used in ref 13 may be unaccessible on this surface.
(23) Miertus, S.; Scrocco, E.; Tomasi, J. *J. Chem. Phys.* **1981**, *55*, 117−129.
(24) Florián, J.; Warshel, A. *ChemSol*, Version 2.1; 1997.
(25) Florián, J.; Warshel, A. *J. Phys. Chem B* **1997**, *101*, 5583−5595.
(26) Kollman, P. A.; Kuhn, B.; Perokyla, M. *J. Phys. Chem. B* **2002**, *106*, 1537−1542.
(27) Jencks, W. P. *Catalysis in Chemistry and Enzymology*; Dover: New York, 1987.
(28) Lightstone, F. C.; Bruice, T. C. *J. Am. Chem. Soc.* **1997**, *119*, 9103−9113.
(29) Verschueren, K. H.; Seljee, F.; Rozeboom, H. J.; Kalk, K. H.; Dijkstra, B. W. *Nature* **1993**, *363*, 693−698.

catalysis. Section 4 describes the results of our calculations and section 5 discusses the implication of our findings.

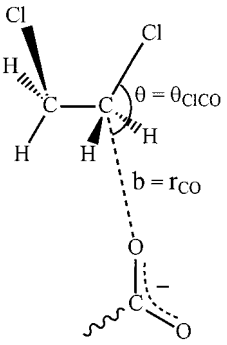
2. Defining the Steric Effect

To illustrate the NAC proposal we consider the S_N2 reaction of DhlA (see **1**) in both water and the protein.



1

Molecular dynamics (MD) trajectories were propagated on the RS and TS to examine the behavior of the C···O bond distance, $r_{C\cdots O}$, and the Cl−C−O angle θ_{ClCO}, depicted in **2**.



2

Figure 1 presents the RS and TS configurations of the system obtained from these trajectories in water (**a**) and in the protein (**b**), along the $r_{C\cdots O}$ and θ_{ClCO} coordinates, (see below for more details). These coordinates are expected to exhibit a major change when moving from the RS to the TS along the reaction coordinate, and hence, should reveal the RS's similarity to the TS. For convenience the same scale was used for both systems.

Looking at the figure it is easy to see that in the water system the area covered by the RS trajectory is far from that of the TS, and is widely distributed. On the other hand, in the protein the two areas are more confined and closer to each other. The difference is especially pronounced in the geometry distribution of the RS structures in water relative to that of the protein, having larger ranges of both the C···O distance and θ_{ClCO}. Based on the figure, it appears that the protein system shows greater similarity of the RS to the TS. A qualitative interpretation of this geometric similarity has led to the hypothesis that a large steric effect controls the reaction and may be responsible for catalysis.[13,16,20] However, the distribution shown in the figure does not provide a quantitative assessment of that effect, and a detailed analysis of the relevant energetics is needed.

Any attempt to examine a catalytic proposal quantitatively requires one to express the proposal in terms of clear thermodynamic concepts (unless the proposal deals with pure dynamical effects). In the present case we are trying to determine the effect of the restricted enzyme active site on the rate constant

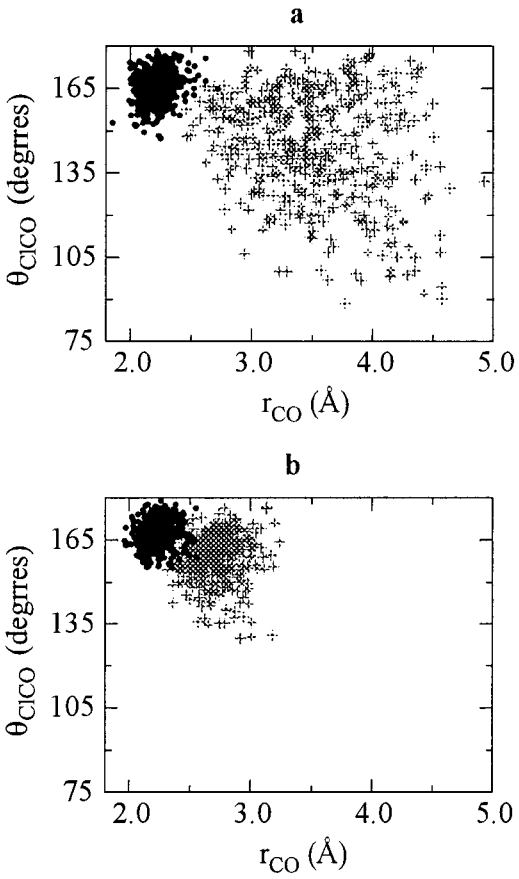


Figure 1. The distribution of $r_{C\cdots O}$ and θ_{ClCO} during MD trajectories in the RS (light plus marks) and TS (black dots). (a) In water and (b) in the protein.

of a reaction. To examine this effect we should consider the steric effects on the same reaction in the enzyme and in solution. This can be done by examining what would be the rate acceleration of the solution reaction if it was conducted in a site that imposes the same steric restrictions as the enzyme does. In other words, we can consider the energetics depicted in Figure 2, where we compare the reaction in water and in the enzyme active site. The overall catalytic effect is given by:

$$\Delta\Delta g^{\ddagger}_{cat} = \Delta g^{\ddagger}_{cat} - \Delta g^{\ddagger}_{w} = \Delta g^{\ddagger}_{cat} - \Delta g(v_0 \rightarrow v_{cage}) - \Delta g^{\ddagger}_{cage} \quad (1)$$

where for convenience we divide the reaction in water into two steps. The first step involves the process of bringing the reactants from a molar volume, v_0, to a cage (where the reactants are assumed to be in a contact distance) with a volume v_{cage}.[30] The second part is the chemical step, whose free energy is designated by $\Delta g^{\ddagger}_{cage}$. The question has always been whether the enzymes provide more catalysis than the trivial effect of $\Delta g(v_o \rightarrow v_{cage})$, which corresponds approximately to the change of effective concentration from 1 M to 55 M. Thus, we will focus here on the remaining contributions of eq 1, namely the difference between $\Delta g^{\ddagger}_{cage}$ and $\Delta g^{\ddagger}_{cat}$:

$$\Delta\Delta g^{\ddagger}_{cage\rightarrow cat} = \Delta g^{\ddagger}_{cat} - \Delta g^{\ddagger}_{cage} \quad (2)$$

Note in this respect that we will use a rigorous approach with a special cage restraint (see next section) that will allow us to obtain $\Delta g^{\ddagger}_{w}$, regardless of the cage definition.

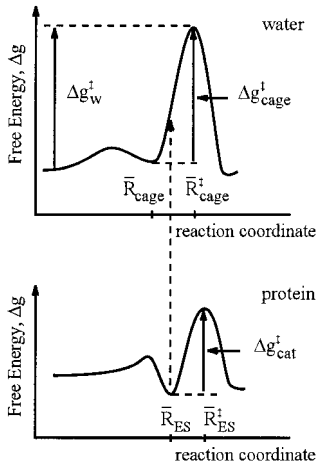


Figure 2. Schematic representation of the reaction profile in water (upper part) compared to the profile of the same reaction in the protein (lower part). $\bar{R}$ and $\bar{R}^{\ddagger}$ are the average distances at the RS and TS of the designated systems. The figure corresponding to a hypothetical case with a large NAC effect.

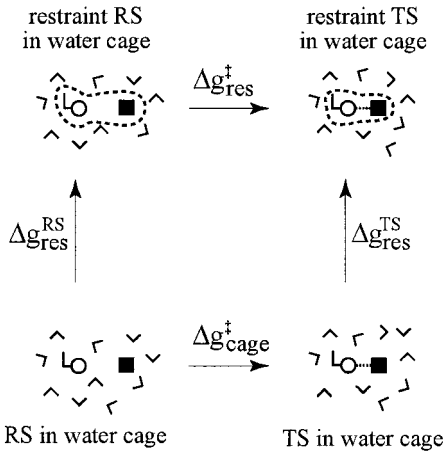


Figure 3. A thermodynamic cycle that defines the NAC effect in terms of steric restraints. The cycle considers the transformation from the unrestrained potential surfaces to the restrained surfaces in the RS and TS of the reaction in the water cage. The effect of the restraint is represented schematically as a dashed line around the reacting fragments.

To analyze the results of our calculations we will have to classify the corresponding energy contributions. In general it is reasonable to separate the activation energy into the internal free energy of the solvated substrate and the interaction of this substrate with its surrounding (enzyme or solution). We can further try to separate the interaction with the surrounding to electrostatic and nonelectrostatic components, writing:

$$\Delta g^{\ddagger}_{cat} \cong \Delta g^{\ddagger}_{cage} + \Delta\Delta g^{\ddagger}_{elec} + \Delta\Delta g^{\ddagger}_{nonelec} \qquad (3)$$

where $\Delta\Delta g^{\ddagger}_{nonelec}$ and $\Delta\Delta g^{\ddagger}_{elec}$ can be considered as the catalytic effects due to the protein steric restraint and electrostatic contributions, respectively.

The nonelectrostatic contribution is usually defined as the sum of the repulsive van der Waals nonbonded interactions and the stretching, bending and torsional energy terms. The electrostatic contributions are defined as the effects of the charge−charge charge−dipole and dipole−dipole interactions, as well as the effects of induced dipoles.[10] Charge-transfer effects are frequently included implicitly in the van der Waals parameters. Thus, we can say that the electrostatic effects include the gas-phase electrostatic interactions and the compensating "solvation" contributions, while the rest of the interactions may be classified as strain forces. Of course, there may be a coupling between the electrostatic and nonelectrostatic contributions. At any rate, with a separation of these contributions effects we may write:

$$\Delta\Delta g^{\ddagger}_{cage\rightarrow cat} \cong \Delta\Delta g^{\ddagger}_{elec} + \Delta\Delta g^{\ddagger}_{nonelec} \qquad (4)$$

The largest contribution to $\Delta\Delta g^{\ddagger}_{elec}$ comes usually from the difference between the reorganization energies of the protein and water systems (see refs 31−33 and section 5). The contributions to $\Delta\Delta g^{\ddagger}_{nonelec}$ may come from the possible distortion of the reacting fragments as well as from the repulsion between the reacting fragments and the entropic effect associated with the steric confinement of these fragments.

The NAC proposal implies that enzymes catalyze reactions by bringing the reacting fragments closer to their TS structure. Since this proposal has not been formulated in a rigorous way (e.g., no clear thermodynamic cycle), we provide here a definition which is amenable to computational verification. The simplest option is to use the magnitude of the steric effect as a quantitative measure of the NAC effect. However, the NAC steric confinement effect might also reflect some electrostatic contributions. Thus, we will have to consider several alternative definitions. In doing so we will consider the NAC contribution as an "apparent steric effect" to indicate the possible contributions from electrostatic factors.

To assess the steric effect, consider the thermodynamic cycle of Figure 3. In this cycle we restrain the reacting fragments in water to the same volume they have in the enzyme and assess the corresponding restraint effect by:

$$\Delta\Delta g^{\ddagger}_{res} = \Delta g^{TS}_{res} - \Delta g^{RS}_{res} \qquad (5)$$

where Δg^{RS}_{res} and Δg^{TS}_{res} are the free energies required to confine the reactants in water to the active site volume in the RS and TS respectively, (see Figure 3). This restraint energy can serve as an estimate of the NAC effect, using:

$$\Delta\Delta g^{\ddagger}_{res} = \Delta\Delta g^{\ddagger}_{NAC} = \Delta\Delta g^{\ddagger}_{steric} \cong \Delta\Delta g^{\ddagger}_{nonelec} \qquad (6)$$

As stated above, we identify the NAC effect with the apparent effect of the steric confinement by the protein, $\Delta\Delta g^{\ddagger}_{steric}$. The approximated relationship between the steric contribution and $\Delta\Delta g^{\ddagger}_{nonelec}$ will be valid if the electrostatic contributions to the NAC effect are small. Our evaluation of the restraint energy involves two steps. First we find a restraint potential that forces the probability distribution in water to be like that of the protein in both the RS and in the TS. Then we evaluate the free energy of releasing the restraint. This free energy provides the desired $\Delta\Delta g^{\ddagger}_{res}$. Alternatively, we calculate the activation free energy with and without this restraint and the difference, again, provides the restraint energy.

(30) Štrajbl, M.; Florián, J.; Warshel, A. *J. Phys. Chem. B* **2001**, *105*, 4471−4484.
(31) Hwang, J.-K.; King, G.; Creighton, S.; Warshel, A. *J. Am. Chem. Soc.* **1988**, *110*, 5297−5311.
(32) Warshel, A. *J. Biol. Chem.* **1998**, *273*, 27035−27038.
(33) Warshel, A. *Proc. Natl. Acad. Sci. U.S.A.* **1978**, *75*, 5250−5254.

In addition to the above approach, we use a complementary strategy which considers the effect of the protein active site without the protein−substrate electrostatic interactions. This approach provides a useful direct estimate of the protein nonelectrostatic effect and the corresponding $\Delta\Delta g^{\ddagger}_{\text{nonelec}}$.

At this point it is useful to clarify the difference between the present restraint release approach to the approach used in our previous calculations of entropic contributions to enzyme catalysis.[34] In studying the NAC effect we are interested in the full free energy contribution of the steric effect, rather than only with the corresponding entropic effect. Thus, we do not have to minimize the restraint release free energy with regards to the restraint coordinates, as is done in the entropic calculations. Furthermore, in studies of entropic contribution to catalysis we release a very large restraint (which forces the system to have zero entropic contribution) in both the protein and the water systems. This is needed to obtain the activation entropies for the reaction in both systems. However,in the present study we have to calculate only the effect of releasing the protein restraint in water.

3. Simulation methods

As stated above we would like to determine $\Delta\Delta g^{\ddagger}_{\text{steric}}$ and the first approach is based on the cycle presented in Figure 3. The simplest way to accomplish this is to add a restraint potential to the potential surface of the water reaction in a way that will force the reacting fragments to behave as if they experience the steric constraint of the protein. Such a treatment forces the reactants in water to behave as if they are in the protein active site. In other words, we are trying to satisfy the relationship:

$$\rho(\mathbf{r})_{V'_w} \cong \rho(\mathbf{r})_{V_p} \tag{7}$$

where ρ is the probability distribution of the system, V_p is the potential surface for the reactants in the protein active site[35] and V'_w is a modified potential surface of the reaction in water that produces a protein-like distribution and satisfies eq 7. V'_w can be written as:

$$V'^{\alpha}_w = V^{\alpha}_w + V^{\alpha}_{res} \cong V^{\alpha}_p \tag{8}$$

Here, α designates RS or TS, V^{α}_w is the undisturbed potential surface in water and V^{α}_{res} is an extra potential added to the water system, enabling it to have a potential that mimics the protein steric potential. The simplest way to obtain V^{α}_{res} is to use a potential that is a function of the solute coordinates and to adjust it until eq 7 is satisfied. Here we use a quadratic function of the form:

$$V^{\alpha}_{\text{res}} = \sum_{i=1}^{n} \frac{1}{2} k^{\alpha}(x_i - x^{\alpha}_{i,0})^2 \tag{9}$$

$\mathbf{x}$ is a vector of the n degrees of freedom of the reactants, which are subjected to the protein steric effect, $\mathbf{x}^{\alpha}_0$ are the restraint coordinates and k^{α} is the restraint force constant. For example, a simple restraint potential can be obtained by using the bond length, b, and bond angle, θ of **2**. In this case we will have:

$$V^{\alpha}_{\text{res}(\theta,b)} = \frac{1}{2} k^{\alpha}_{\theta}(\theta - \theta_0)^2 + \frac{1}{2} k^{\alpha}_{b}(b - b_0)^2 \tag{10}$$

Our goal is to find the constraint coordinates $\mathbf{x}^{\alpha}_0$ and force constants k^{α} that will best fit the probability distribution obtained with V'^{α}_w to that calculated in the protein. Thus, we try to minimize the distance, d^{α}, ($d^{\alpha} = |\bar{\mathbf{x}}^{\alpha}_p - \bar{\mathbf{x}}^{\alpha}_w|$), between the average coordinates, $\bar{\mathbf{x}}^{\alpha}$, obtained with V^{α}_p and V'^{α}_w, respectively, by finding the best $\mathbf{x}^{\alpha}_0$.

In addition, we also try to minimize the difference between the widths of the distributions by finding the optimal k^{α}. This is done by an iterative approach that is a simplified version of the approach used previously in a surface-constrained all-atom solvent model.[36] That is, we start with an initial guess, $\mathbf{x}^{\alpha,1}_0$:

$$\mathbf{x}^{\alpha,1}_0 = \bar{\mathbf{x}}^{\alpha}_p \tag{11}$$

where $\mathbf{x}^{\alpha,1}_0$ is taken as the average structure of the substrate in a given state (RS or TS) in the protein site. We also start with a relatively large force constant $k^{\alpha,1}$. Next we run a MD simulation on the restraint energy surface, V'^{α}_w, and obtain a new probability distribution with an average structure $\bar{\mathbf{x}}^{\alpha,1}_w$. The new distribution is compared to that of the protein and the distance between the averages, d^{α}, is evaluated. Next, in an attempt to minimize d^{α}, we generate a new guess using:

$$\mathbf{x}^{\alpha,n}_0 = \mathbf{x}^{\alpha,n-1}_0 + (\bar{\mathbf{x}}^{\alpha}_p - \bar{\mathbf{x}}^{\alpha,n-1}_w) \tag{12}$$

This iterative procedure is repeated until convergence is achieved.

Finally we need to minimize the difference in the deviations between the two distributions by changing the k^{α}. Thus, we start decreasing or increasing the value of k^{α} gradually (depending on the relative shapes of the distributions) and repeating the calculation until the proper distribution is achieved.

To further clarify the method we present in Figure 4 a sequential representation of the fitting process for the RS distribution. For simplicity we will concentrate only on one coordinate, $r_{\text{C}\cdots\text{O}}$, but the same considerations apply to all the coordinates involved in the fitting (θ_{ClCO} in this case). Figure 4a shows the probability distributions obtained from RS trajectories both in water and in protein when no constraints were applied (the same RS distributions presented in Figure 1). As seen from the figure, $(\bar{r}_{\text{C}\cdots\text{O}})^{\text{RS}}_p = 2.71$ Å. Hence, we started our iterative procedure with $(\bar{r}_{\text{C}\cdots\text{O}})^{\text{RS},1}_0 = 2.71$ Å and $k^{\text{RS}} = 6.0$ kcal mol^{-1} Å^{-2}. The results are presented in Figure 4b. For convenience the protein distribution is shown as well. Comparing (**a**) and (**b**) of Figure 4 it is clear that the application of the restraint both moved the average of the RS C···O distance in water from $(\bar{r}_{\text{C}\cdots\text{O}})^{\text{RS},0}_w = 3.48$ Å (when no restraint is applied[37] see Figure 4a) to $(\bar{r}_{\text{C}\cdots\text{O}})^{\text{RS},1}_w = 2.96$ Å (a value closer to that of the protein) and confined the distribution to a narrower area (more similar to that of the protein). However, as is also evident from the figure, eq 7 is not yet satisfied, and further fitting needs to be carried on. The restraint guess for the next step was chosen following eq 12 to be $(r_{\text{C}\cdots\text{O}})^{\text{RS},2}_0 = 2.46$ Å and the trajectory is calculated. This procedure is repeated until the average C···O distance reproduces that of the protein (Figure 4c). The last step, which included a manipulation the width of the distribution, involved reducing the force constant from 6.0 to 5.0 kcal mol^{-1} Å^{-2}, which was still sufficient to keep the shape shown in Figure 4c.

Once the fitting process is completed and eq 7 is satisfied, it is assumed that the restraint potential provides a reasonable representation of the protein steric effect. Our task is then to evaluate the free energy, $\Delta g^{\alpha}_{\text{res}}$, associated with the release of this restraint. As stated above, the difference between $\Delta g^{\text{TS}}_{\text{res}}$ and $\Delta g^{\text{RS}}_{\text{res}}$ provides our estimate of the NAC effect. The evaluation of $\Delta g^{\alpha}_{\text{res}}$ is done by releasing the constraint gradually, using the free energy perturbation (FEP) method.[38] In this

(34) Villà, J.; Štrajbl, M.; Glennon, T. M.; Sham, Y. Y.; Chu, Z. T.; Warshel, A. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 11899−11904.
(35) V_p includes the effect of the protein relaxation in response to the substrate structural changes.
(36) King, G.; Warshel, A. *J. Chem. Phys.* **1989**, *91*, 3647−3661.
(37) Unless otherwise metioned, the cage constraint was always included with a force constant of 0.4 kcal mol^{-1} Å^{-2} (see text).
(38) Valleau, J. P.; Torrie, G. M. *Modern Theoretical Chemistry*; Plenum Press: New York, 1977; Vol. 5.

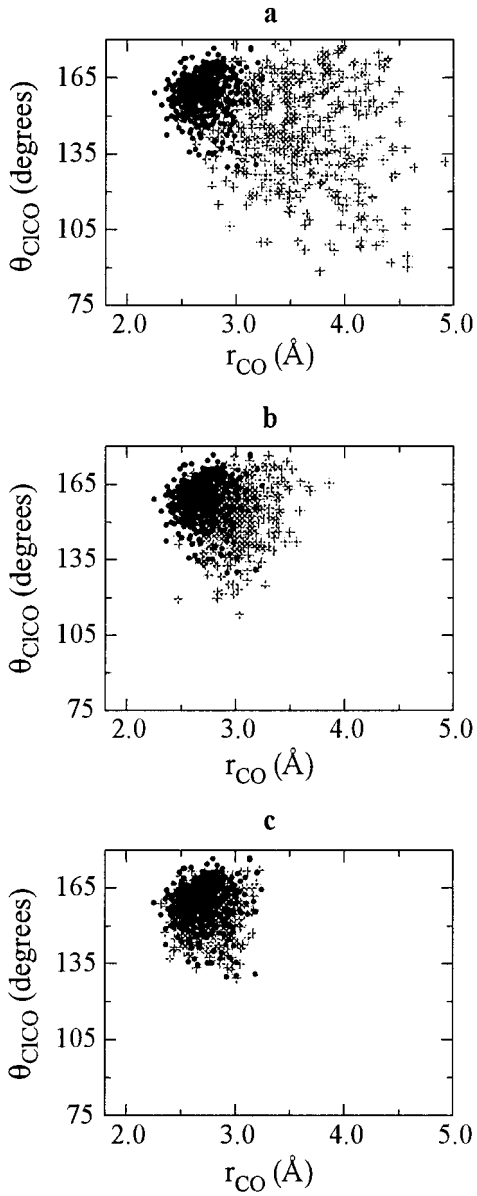


Figure 4. Sequential description of fitting of the RS probability distribution of the reaction in water (light plus signs) to that in the protein (black dots). (a) Both distributions when no restraints are applied.[37] (b) and (c) RS distribution in water when a restraint potential of the form of eq 10 was used. The force constants $k_\theta^{RS} = 6.0$ kcal mol^{-1} rad^{-2} and $k_b^{RS} = 6.0$ kcal mol^{-1} Å^{-2} were used in both b and c, but the restraint bond distance, b_0 and angle, θ_0, used are different: $b_0 = 2.71$ Å and $\theta_0 = 151.0°$ and $b_0 = 2.16$ Å and $\theta_0 = 158.8°$ in b and c, respectively. These values are determined in each step based on the latter, using eq 12. (b) represents the first step and c is the last step.

procedure we use a mapping potential of the form:

$$V_m^\alpha = \lambda_m V_w'^\alpha + (1-\lambda_m)V_w^\alpha = V_w^\alpha + \lambda_m V_{res}^\alpha \qquad (13)$$

where λ_m is a parameter that changes from 1 to 0 in fixed increments ($m = 0, 1, 2, \ldots n$). The free energy, Δg_{res}^α, is then evaluated by:

$$\Delta g_{res}^\alpha = \sum_{m=0}^{n} \Delta\Delta g_m^\alpha \qquad (14)$$

$$\Delta\Delta g_m^\alpha = -\beta^{-1} \ln\langle e^{-\beta(V_{m+1}^\alpha - V_m^\alpha)}\rangle_{V_m^\alpha}$$

where $\langle..\rangle_{V_m^\alpha}$ designate an average over trajectories propagated over the potential surface V_m^α.

The potential surfaces for the chemical reaction in water and in the protein (V_w and V_p, respectively) are obtained by the empirical valence bond (EVB) method,[2] whose implementation for DhlA is described elsewhere[39] (see also supporting information). V^{RS} is obtained by using the ground-state EVB surface at the reactant region. V^{TS}, on the other hand, is obtained by using the combination:

$$V^{TS} = c_1\epsilon_1 + c_2\epsilon_2 - H_{1,2} \qquad (15)$$

where ϵ_1 and ϵ_2 are the potential surface of the reactant and product diabatic states, respectively. The relative weights, c_1 and c_2, are found using the EVB mapping procedure by looking for the combination that keeps the system closest to its TS. The TS location is determined as the highest point along the free energy profile, (see refs 40, 41 for a detailed description). In the present S_N2 reaction we found the weights (c_1, c_2) to be (0.50, 0.50) in water and (0.53, 0.47) in the protein, respectively; in agreement with related studies.[42,43] Our V^{TS} has the same dependence on the coordinates which are perpendicular to the reaction coordinate as the actual EVB ground state at the TS region. However, along the reaction coordinate V^{TS} has a minimum rather than maximum. This keeps the system at the TS region during the simulation.

In addition to the restraint release approach, we found it useful to evaluate the free energy profile for the reaction in water in the presence of the "protein-like" restraint potential. This was done using the EVB FEP/umbrella sampling mapping as described elsewhere (e.g., refs 3 and 44), except that the restraint potential was added to the EVB potential surface.

At this stage, we should address the evaluation of $\Delta g(\upsilon_0 \rightarrow \upsilon_{cage})$ of eq 1. This correction is needed in order to relate $\Delta g_{cat}^{\ddagger}$ to the experimentally estimated $\Delta g_w^{\ddagger}$. As pointed out in section 2 and elsewhere,[45] the catalytic effect of the enzyme is best formulated and analyzed in terms of the difference between $\Delta g_{cat}^{\ddagger}$ and $\Delta g_{cage}^{\ddagger}$. However the direct experimental information is given in terms of $\Delta g_w^{\ddagger}$ and $\Delta g_{cat}^{\ddagger}$. Fortunately, $\Delta g(\upsilon_0 \rightarrow \upsilon_{cage})$ is a trivial correction that is qualitatively given by the well-known effect of moving from 1M to 55M, which corresponds to ~2.4 kcal/mol for two fragments.[41,45] This means that we have an approximate "experimental" value.

$$\Delta g_{cage}^{\ddagger} = \Delta g_w^{\ddagger} - \Delta g(\upsilon_0 \rightarrow \upsilon_{cage}) \cong \Delta g_w^{\ddagger} - 2.4 \qquad (16)$$

A quantitative estimate of $\Delta g(\upsilon_0 \rightarrow \upsilon_{cage})$ can also be obtained by using a small constraint, k_{cage}, (K_{cage} in the notation of ref 41) and evaluating the free energy associated with the release of that constraint.[41] In this case the simulations of $\Delta g_{cage}^{\ddagger}$ are performed while using k_{cage}. Thus, in all the calculations we imposed a cage constraint on the O−C distance, with $k_{cage} = 0.4$ kcal mol^{-1} Å^{-2}. The release of this constraint gave $\Delta g(\upsilon_0 \rightarrow \upsilon_{cage}) \cong 2.6$ kcal/mol. Note, however, that the use of Eq. (10) is an arbitrary devision, for practical and conceptual purpose, and that the same $\Delta g_w^{\ddagger}$ will be obtained for any k_{cage}.

It is important to realize that the requirement of eq 7 is more than an ad hoc intuitive relationship. To illustrate this we consider the simple case where the main difference between the protein and the water systems is their steric potential. In this illustrative case we want to find the free energy associated with changing the water steric potential, V_w^α, to the protein steric potential V_p^α. This can be done using the following relation:[46]

(39) Shurki, A.; Warshel, A. Manuscript in preparation.
(40) Hwang, J.-K.; Warshel, A. *Biochemistry* **1987**, *26*, 2669−2673.
(41) Strajbl, M.; Sham, Y. Y.; Villà, J.; Chu, Z. T.; Warshel, A. *J. Phys. Chem. B* **2000**, *104*, 4578−4584.
(42) Shaik, S.; Ioffe, A.; Reddy, A. C.; Pross, A. *J. Am. Chem. Soc.* **1994**, *116*, 262−273.
(43) Shaik, S.; Reddy, A. C. *J. Chem. Soc., Faraday Trans.* **1994**, *90*, 1631−1642.
(44) Chapter 3.4 and 3.5 of ref 2.
(45) Chapter 5.1 of ref 2.
(46) McQuarrie, D. A. *Statistical Mechanics*; Harper and Row: New York, 1976.

$$\Delta g^{\alpha}_{(w\rightarrow p)} = -\beta^{-1}\ln\left(\frac{Z^{\alpha}_{p}}{Z^{\alpha}_{w}}\right) \quad (17)$$

where Z^{α}_{p} and Z^{α}_{w} are the partition functions of the water and the protein systems, respectively. The classical partition functions are given by:

$$Z^{\alpha}_{s} = \int e^{-\beta V_s^{\alpha}} d\mathbf{r} = \int \rho(\mathbf{r})_{V_s^{\alpha}} d\mathbf{r} \quad (18)$$

where, s designates the particular system (water or protein). As can be seen from eq 18 the partition function of each system corresponds to its probability distribution and the free energy associated with transferring from one probability distribution to the other is given by $\Delta g^{\alpha}_{(w\rightarrow p)}$.

Now, it is easy to use eq 18 and show, by the same approach used in deriving the FEP expression (see ref 47), that:

$$\frac{Z^{\alpha}_{p}}{Z^{\alpha}_{w}} = \langle e^{-\beta(V_p^{\alpha}-V_w^{\alpha})}\rangle_{V_w^{\alpha}} \quad (19)$$

Using this and eq 17, we see that $\Delta g^{\alpha}_{(w\rightarrow p)}$ can be evaluated by a standard FEP approach, which involves the change of V^{α}_{w} to V^{α}_{p}. Usually the one-step expression of eq 17 is evaluated by n steps, as is done in eq 14. Thus, the free energy invested in changing the potential from V^{α}_{w} to V^{α}_{p} (estimated by V'^{α}_{w} in our work, see eq 8) is given by the same FEP approach used in eqs 13 and 14 to satisfy eq 7. In other words, instead of trying to satisfy eq 7, we can consider a more rigorous formulation where we simply ask what is the free energy $\Delta g^{\alpha}_{(w\rightarrow p)}$ that uniquely describes the effect of moving from the regular water cage to a water cage with the protein steric effect ($V^{\alpha}_{w} \rightarrow V'^{\alpha}_{w}$). Using these potentials automatically satisfies eq 7.

The actual implementation of the different methods described above was done with the program ENZYMIX,[48] which is now a part of the modeling package MOLARIS.[49] The average distributions, needed to determine the restraint parameters in the RS and TS of the water and protein systems, were determined, in each case, by a 50 ps MD run. In most cases, we performed the FEP restraint-release calculations and the reaction profile calculations, using 31 frames ($n = 31$ in eq 14), each involving a 5 ps MD run.

4. Results

Before examining the magnitude of the NAC contribution it is important to capture the overall catalysis. This was done by EVB calculations of activation free energies for the S_N2 reaction in water and in the protein. The corresponding EVB parameters are given here as Supporting Information and more details are given in ref 39. Table 1 gives the calculated activation free energies for the S_N2 reaction in water and in protein as well as the corresponding observed values.

As seen from the table the enzyme reduces the activation free energy of the reaction by about 10.7 kcal/mol (calculated: 9.4 kcal/mol) relative to water, or 8.3 kcal/mol (calculated: 6.8 kcal/mol) relative to a water cage. Our goal is to determine how large is the steric contribution to this catalytic effect.

4.1. Quantifying the Steric Effect. The magnitude of the steric effect was evaluated by the restraint release approach outlined in section 3. The nature of this approach can be best realized by considering Figure 3 and going from Figure **4c** to a. Figure 4c shows the probability distributions for a RS protein trajectory (black dots) and a RS trajectory on a restraint $r_{C\cdots O}$, and θ_{ClCO} potential in water (plus signs).[50] The trajectories populate the same area and have the same averages for the two coordinates of interest. Hence, one can say that the substrate in water feels a "protein-like" potential.

Now, releasing the restraint and moving to Figure **4a** gives us the free energy associated with the protein steric effect. The quantitative results of this restraint release process are given in Table 2 for both the RS and the TS. b and θ denote the $r_{C\cdots O}$, and θ_{ClCO} coordinates that were restrained.[50] The results are given for three different initial conditions along with their average (each presented in a separate entry), and appear to be quite stable.

As seen from the table, the restraint energy in the RS has an average value of 2.9 kcal/mol. This calculated $\Delta g^{RS}_{res(\theta,b)}$ reflects the steric restriction applied by the protein on the substrate in the RS. Similarly, the corresponding effect in the TS is about 0.1 kcal/mol. The difference between the restraint free energy in the RS and in the TS gives an estimate for the steric effect in the protein (column 4 in the table). Thus, on the basis of these results, one obtains a steric effect which amounts to about −2.8 kcal/mol.

Looking at the cycle in Figure 3 it is seen that the steric effect can also be obtained by using the EVB−FEP/umbrella sampling procedure and calculating the activation free energy of the restrained substrate in water. The resulting $\Delta g^{\ddagger}_{res(\theta,b)}$ can be compared to the activation free energy of the unrestrained substrate in water, $\Delta g^{\ddagger}_{cage}$. The last column of Table 2 gives the $\Delta g^{\ddagger}_{res(\theta,b)}$ values. As seen from the table the different initial conditions result in very similar barriers, with an average of

Table 1. Activation Free Energies[a] for the S_N2 Step of the Reaction of DhlA and the Corresponding Reference Reaction

	$\Delta g^{\ddagger}_{calc}$	$\Delta g^{\ddagger}_{expt}$
water[b]	24.9	26[c]
water cage[d]	22.3	23.6
protein[e]	15.5	15.3[f]

[a] Energies in kcal/mol. The calculated values were obtained by the EVB approach with the parameters given as Supporting Information (for more details see ref 39). [b] Corresponds to 1 M concentration of the substrate namely $\Delta g^{\ddagger}$ refers to $\Delta g^{\ddagger}_{w}$. [c] Obtained by extrapolating from related reactions[65] see ref 39 for more details. [d] Corresponds to 55 M concentration of the substrate namely $\Delta g^{\ddagger}$ refers to $\Delta g^{\ddagger}_{cage}$. [e] $\Delta g^{\ddagger}$ refers to $\Delta g^{\ddagger}_{cat}$ [f] obtained from k_2 in ref. 66 using the transition-state theory.[67,68]

Table 2. NAC Effect[a] Obtained Using the (θ, b) Constraints[b]

entry[c]	$\Delta g^{RS}_{res(\theta,b)}$	$\Delta g^{TS}_{res(\theta,b)}$	$\Delta\Delta g^{\ddagger}_{res(\theta,b)}$ [d]	$\Delta g^{\ddagger}_{res(\theta,b)}$ [e]	$(\Delta\Delta g^{\ddagger}_{res(\theta,b)})'$ [f]
1	3.2	0.1	−3.1	22.3	
2	2.6	0.1	−2.5	19.2	
3	2.8	0.1	−2.7	20.5	
average	2.9	0.1	−2.8	20.7	−1.6

[a] Energies in kcal/mol were obtained using a restraint potential for $r_{C\cdots O}$ and θ_{ClCO}. [b] $k^{\alpha}_{\theta} = 5.0$ kcal mol^{-1} rad^{-2} and $k^{\alpha}_{b} = 5.0$ kcal mol^{-1}Å^{-2} were the force constants in both the RS and the TS. $b_0 = 2.16$ Å and $\theta_0 = 158.8°$ were the restrained bond distance and angle in the RS, respectively. $b_0 = 2.23$ Å and $\theta_0 = 167.0°$ were the restrained bond distance and angle in the TS, respectively (see eq 10). [c] The different entries refer to different initial conditions. [d] $\Delta\Delta g^{\ddagger}_{res(\theta,b)}$ evaluated according to eq 5. [e] $\Delta g^{\ddagger}_{res(\theta,b)}$ is the activation barrier obtained by direct EVB calculations of the activation barrier in water but with the $V_{res(\theta,b)}$ potential. [f] $(\Delta\Delta g^{\ddagger}_{res(\theta,b)})'$ evaluated as difference between the average calculated restrained and unrestrained activation barriers. Thus $(\Delta\Delta g^{\ddagger}_{res(\theta,b)})' \cong \Delta g^{\ddagger}_{res(\theta,b)} - \Delta g^{\ddagger}_{cage(\theta,b)}$.

(47) Chapter 3.3 of ref *2*.
(48) Lee, F. S.; Chu, Z. T.; Warshel, A. *J. Comput. Chem.* **1993**, *14*, 161−185.
(49) Chu, Z. T.; Villà, J.; Schutz, C. N.; Strajbl, M.; Warshel, A. Manuscript in preparation.

(50) The potential described in eq 10 using $k^{\alpha}_{b} = 3.0$ kcal mol^{-1} Å^{-2} and $k^{\alpha}_{\theta} = 3.0$ kcal mol^{-1} rad^{-2}.

20.7 kcal/mol. Comparing these free energies to those obtained in the unrestrained case, (entry 2 in Table 1) it is seen that the differences again are relatively small, (around −1.6 kcal/mol steric contribution). The difference between the two estimates might reflect a difference in convergence and thus we consider the average of the two results (∼−2.2 kcal/mol) as the result of this specific restraint release calculation.

The above calculations only restrained the b and θ coordinates. Thus, one may argue that the NAC effect involves more coordinates and that our study is inconclusive. A seemingly obvious solution is to repeat the calculations while including all the Cartesian coordinates in the restraint of eq 9. Calculations with this "protein-like" restraint potential resulted with an average restraint release effect of ∼−3.5 kcal/mol (see Supporting Information). The finding that the restraint release energy is somewhat larger here than the result obtained while using only two degrees of freedom, (−2.2 kcal/mol), is probably an overestimate. That is, the quadratic steric effect correctly describes restriction of breathing motions of the substrate atoms toward the "protein-like envelope" (i.e., expansion of the substrate). However, by restraining all the substrate's degrees of freedom, this approach also restricts the motion of the substrate in compression and rotation movements, which are not necessarily restricted in the protein.[51] Therefore, this steric restriction leads to an overestimation of the entropic effect of the active site. One way to resolve this problem is to build a cavity with a shape of the enzyme around the reacting fragments in water. Another, simpler approach is to use the actual protein structure as the steric cavity for the water reaction. This procedure is applied in the next section.

4.2. Separating the Electrostatic and Nonelectrostatic Effects. As stated above, we would like to estimate in a more direct way the restrictive effect of the protein. The simplest way to do so is to consider the nonelectrostatic contribution of the actual active site. If the steric effect involves no significant electrostatic contribution then this nonelectrostatic effect gives the best description of the steric effect. Instead of finding a restraint that approximates the effect of the active site, we can use the actual shape of the active site. Thus, we calculated the nonelectrostatic contribution by eliminating the substrate-surrounding electrostatic interactions both in the protein and water systems.

Table 3 gives the activation free energy for the S_N2 reaction when the electrostatic interaction between the substrate and its surrounding (protein and/or water) is either included or set to zero. These free energies are denoted by $\Delta g^{\ddagger}$ and $\Delta g^{\ddagger}_{\text{noelec}}$, respectively. The upper part refers to the reaction in water, whereas the lower part refers to the reaction in the protein. Results for different initial conditions are given in both cases along with the average values (in bold).

Comparing $\Delta g^{\ddagger}$ to $\Delta g^{\ddagger}_{\text{noelec}}$ in water (upper part of Table 3), it is seen that elimination of the electrostatic contribution reduces the activation barrier in water considerably, (by ∼10.5 kcal/mol on the average). This is consistent with the well-known solvent effect on S_N2 reactions (e.g., ref 52), where the solvent

Table 3. Free Energy of Activation[a,b] with and without Electrostatic Contributions

entry	$\Delta g^{\ddagger}$	$\Delta g^{\ddagger}_{\text{noelec}}$
	Water[c]	
1	22.7	9.3
2	20.9	11.2
3	24.5	12.9
4	22.4	14.0
5	21.1	11.7
average	**22.3**	**11.8**
	Protein[c]	
7	15.6	11.4
8	15.4	11.9
9	14.9	11.6
10	15.9	10.5
11	15.7	10.2
average	**15.5**	**11.1**

[a] Energies in kcal/mol. [b] $\Delta g^{\ddagger}$ and $\Delta g^{\ddagger}_{\text{noelec}}$ correspond to calculations with and without consideration of electrostatic interactions between the substrate and its surroundings, respectively. [c] The different entries reffer to different initial conditions.

stabilizes the concentrated RS charge more then the delocalized TS charge. The remaining barrier represents mainly the intrinsic chemical activation energy of the bond-making bond-breaking process, which involves a major reorganization of the electronic structure of the substrate, and is approximately equal to the barrier of this reaction in the gas phase.

Moving now to the protein (lower part of the table), we find that the elimination of the electrostatic contribution reduces the activation barrier by only ∼4.4 kcal/mol. Thus, the electrostatic contributions to the activation barriers are ∼4.4 and ∼10.5 kcal/mol in protein and water, respectively. In other words, the electrostatic contribution to the catalysis is approximately −6.1 kcal/mol. Note that the electrostatic (solvation) effects *increase* the intramolecular barrier for the S_N2 reaction but it does so in a less pronounced way in the enzyme than in water, (see also next section). At any rate, the value of $\Delta g^{\ddagger}_{\text{noelec}}$ is 11.1 and 11.8 kcal/mol in protein and water, respectively. The difference between these values gives a small (∼−0.7 kcal/mol) nonelectrostatic contribution to the NAC effect, indicating that the traditional components of the steric effect (i.e., the strain and other nonelectrostatic factors) provide a very small catalytic contribution. As is clear from the present analysis, the pure nonelectrostatic effect gives a smaller catalytic contribution than the full apparent steric effect, (∼−0.7 and ∼−2.2 kcal/mol, respectively). The difference between these two estimates is due to special electrostatic effects. First, in the RS in water the negatively charged carboxylate is better solvated when it is further away from the neutral substrate. In fact, the presence of such an effect in reactions of model compound has been noted by Peräkylä and Kollman.[53] Second, the solvent contribution to the activation barrier is much larger in water than in the protein. This solvation effect is associated with the change in the solute charge distribution along the reaction coordinate. The reduction in solvation energy upon formation of the TS is much larger in water than in the protein. This makes it harder to push the carboxylate oxygen toward the substrate in water than in the protein. The above effects are quite different than the traditional steric effect since it is not due to a repulsion between the solute fragments but to a distance-dependent change in the solvation energy. This point can be best understood by thinking

(51) Note that in this case we have large restraint release contribution both in the RS and in the TS. This reflects the contributions from coordinates that are free to move both in the RS and TS in water (see ref 34 for a related effect).

(52) Shaik, S. S.; Schlegel, H. B.; Wolfe, S. *Theoretical Aspects of Physical Organic Chemistry. The S_N2 Mechanism*; John Wiley & Sons: New York, 1992.

(53) Peräkylä, M.; Kollman, P. A. *J. Am. Chem. Soc.* **1999**, *103*, 8067−8074.

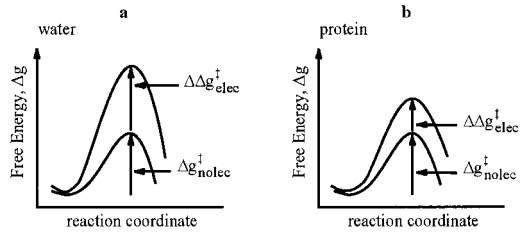


Figure 5. Schematic description of the reaction profile in water (a) and in protein (b), with and without the electrostatic interaction between the substrate and its environment.

on the hypothetical situation where the enzyme reduces to zero the activation barrier by electrostatic stabilization of the TS. In this case the RS and TS will coincide in the enzyme. If the activation barrier is high in water we will need a very strong restraint to force the reactants in water to be in the same RS structure as in the enzyme. Thus, one will conclude that we have a large NAC effect, although we have here a basic case of TS stabilization. At any rate, solvation-induced repulsion leads to ~−1.5 kcal/mol contribution, which is also not a major effect (see also Figure 5).

5. Discussion

The calculations presented in the results section indicate that the nonelectrostatic strain effect does not constitute a major contribution to catalysis in DhlA. This conclusion is in agreement with previous studies of lysozyme[10,11] which showed that strain contribution is not significant. As an additional example, we may consider the case of chorismate mutase which was advanced by two research groups[5,9] as a prototype of a system with a large catalytic steric effect. As pointed in the Introduction, none of these works demonstrated the validity of their proposal by calculating the actual steric effect of the enzyme. In fact, most other studies have concluded that chorismate mutase catalyzes its reactions by electrostatic effects (for review see refs 54−56). Since the strain effect is a RS destabilization effect, we can examine its validity by mutation experiments. That is, if the catalysis is due to RS destabilization, then there are probably some specific residues that lead to this effect. Mutations of these residues will increase the binding (reduce K_m) while leaving k_{cat}/K_m unchanged (see ref. 32). On the other hand, if the catalysis is due to TS stabilization, then mutations will decrease k_{cat}/K_m while leaving K_m unchanged. The limited mutation studies of chorismate mutase are inconsistent with RS destabilization, namely, k_{cat}/K_m decreases and K_m increases.[57]

A recent work of Kollman and coworkers[26] seems to support the NAC hypothesis. However, this work has not defined properly the NAC contribution, nor the corresponding entropic effects. It also drastically overestimated entropic contributions to catalysis by using gas phase vibrational contributions (see discussions in footnotes 90 and 92 of ref. 30) and considering an incomplete thermodynamic cycle (see ref. 3). Note in this respect that in contrast to the misunderstanding of ref. 26, our cage concept is a rigorous mathematical tool (see ref. 30 and the present work) that allows us to evaluate $\Delta g^{\ddagger}_{w}$ in two steps. Thus, our $\Delta g(\nu_o \rightarrow \nu_{cage})$ is not (and never has been) an estimate of the NAC contribution.

The present work finds that the electrostatic contribution in DhlA is responsible for a major part of the catalysis, (~6.1 kcal/mol of the calculated 6.8 kcal/mol $\Delta\Delta g^{\ddagger}_{cage\rightarrow cat}$). Although a more extensive analysis of this electrostatic contribution will be given in a subsequent paper,[39] we would like to clarify several points about the origin of $\Delta\Delta g^{\ddagger}_{elec}$ and its nontrivial nature. We start by clarifying that electrostatic contribution to enzyme catalysis cannot be simply identified by looking at specific interactions in the enzyme. For example, one may argue that the interaction between Asp124 and His289 (in its ionized form) should reduce the catalytic effect of the enzyme by stabilizing the RS. Similarly, one might argue that if the overall charge of the ionized groups near the carboxylate is positive, then the reaction in the protein will be slower than the corresponding reaction in water. However, such a simplified analysis would overlook the fact that the energy of the attacking carboxylate is determined by its overall surrounding. For example, in solution the attacking carboxylate is stabilized by water molecules. This stabilization may be larger than the corresponding stabilization in the protein. Thus, a careful comparison of the overall electrostatic energy in the protein and in the reference reaction in water, for both the TS and RS is needed.

This need for careful considerations can be illustrated by the analysis of the NAC in DhlA. That is, as stated above, the NAC and related concepts (e.g., the desolvation proposal[27,58,59]) assume RS destabilization effects. The finding that in some enzymes the electrostatic interactions do not stabilize the TS more than the RS has been used to support this idea.[15,19,20] While this finding can sometimes be correct, it is not necessarily relevant to catalysis. To explain catalysis it is not enough to look at the absolute stabilization of the RS and the TS in the enzyme. It is also essential to compare the relative energies of the TS and the RS in water and in the protein.[60] Here, it is usually found that the enzyme stabilizes the TS more than water does. The S_N2 reaction of DhlA is a case in point. In water, the solvent stabilizes (solvates) the localized charges of the RS more than the delocalized charges on the TS. The same scenario may be true in the enzyme, that is, the enzyme can stabilize its RS more then its TS. Still, this does not contradict the fact that the TS in the enzyme is likely to be more stable than the TS in water, which leads to catalysis. In other words, the requirement for catalysis is:

$$\Delta g^{\ddagger}_{p} - \Delta g^{\ddagger}_{w} = (\Delta g^{TS}_{p} - \Delta g^{RS}_{p}) - (\Delta g^{TS}_{w} - \Delta g^{RS}_{w}) < 0 \quad (20)$$

Assuming that the intramolecular part of the activation barrier is similar in water and in the enzyme, the requirement of eq 20 becomes:

$$(\Delta g^{TS}_{p,sol} - \Delta g^{RS}_{p,sol}) - (\Delta g^{TS}_{w,sol} - \Delta g^{RS}_{w,sol}) < 0 \quad (21)$$

where Δg^{α}_{sol} designates the "solvation" energy of the reacting fragments in a given environment and we refer to the overall noncovalent interactions as solvation energy. This requirement is usually satisfied by the following relations:[60]

(54) Hilvert, D. *Annu. Rev. Biochem.* **2000**, *69*, 751−793.
(55) Davidson, M. M.; Gould, I. R.; Hillier, I. H. *J. Chem. Soc., Perkin Trans. 2* **1996**, 525−532.
(56) Lyne, P. D.; Mulholland, A. J.; Richards, W. G. *J. Am. Chem. Soc.* **1995**, *117*, 11345−11350.
(57) Cload, S. T.; Liu, D. R.; Pastor, R. M.; Schultz, P. G. *J. Am. Chem. Soc.* **1996**, *118*, 1787−1788.
(58) Cohen, S. G.; Vaidya, V. M.; Schultz, R. M. *Proc. Natl. Acad. Sci. U.S.A.* **1970**, *66*, 249−256.
(59) Dewar, M.; Storch, D. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 2225−2229.
(60) Warshel, A.; Strajbl, M.; Villà, J.; Florián, J. *Biochemistry* **2000**, *39*, 14728−14738.

$$\Delta g^{TS}_{p,sol} < \Delta g^{TS}_{w,sol} \quad (22)$$

$$\Delta g^{RS}_{p,sol} \cong \Delta g^{RS}_{w,sol} \quad (23)$$

That is, the enzyme "solvates" the TS stronger than water does, even if it also "solvates" its RS stronger than its TS ($\Delta g^{RS}_{p,sol} < \Delta g^{TS}_{p,sol}$). However, actual binding energy calculations are needed to quantify this point. Such calculations will be performed on DhlA in future studies.

It is important to realize that the magnitude of the electrostatic contribution to catalysis cannot even be determined by comparing the electrostatic interactions in the enzyme to the corresponding interaction in water. Such analysis overlooks the electrostatic work invested in reorganizing the solvent (or the protein) toward the charge distribution of the TS and RS. In fact most of the catalytic effect is due to this difference in the reorganization energy. We will present below a partial analysis of the reorganization effect and relate it to the analysis of the NAC effect.

To analyze the electrostatic contribution to catalysis and to relate it to the NAC effect it is useful to find out which coordinates store $\Delta\Delta g^{\ddagger}_{cage\rightarrow cat}$. Considering only the substrate coordinates may be confusing since, the energy stored in the substrate coordinates (which is associated with the NAC effect) is not the major catalytic factor. Apparently, the substrate coordinates are not sufficient to describe reactions in condensed phases, since both the solute and solvent coordinates (Q and S, respectively) must be considered.[31]

The solvent coordinate represents a collective solvation coordinate, which can be represented approximately by the product of the solute dipole and the reaction field induced by its environment (either a solvent or a protein).[31] Probably the best way to reduce the enormous coordinate space of the system into a subspace of solute and solvent coordinates is to use the EVB formulation. In this formulation, the S_N2 reaction can be described by the diabatic states of the reactants and products, whose energies are ϵ_1 and ϵ_2, respectively. The energy gap between the two states, $\Delta\epsilon$, serves as the general reaction coordinate and is the sum of the solvent and solute coordinates. The electrostatic contribution to this gap is the EVB solvent coordinate.[31] This definition of the solvent coordinate provides a rigorous microscopic equivalent of the macroscopic coordinate used in Marcus' theory of electron-transfer reaction (for more details see ref 61).

The free energy surfaces along the solvent coordinate are somewhat flatter than those along the solute coordinate. Furthermore, the solvent coordinate involves many more degrees of freedom than the solute coordinate. Thus, it seems easier to store energy in the solvent rather than in the solute coordinate. Now, let us try to quantify this issue.

In general we can express the activation free energy by a modified Marcus' expression:[62]

$$\Delta g^{\ddagger} \cong \frac{(\Delta G_0 + \lambda)^2}{4\lambda} - H_{ij}(x^{\ddagger}) + \frac{H_{ij}{}^2(x_0)}{(\Delta G_0 + \lambda)} \quad (24)$$

where the first term is the familiar Marcus expression,[63] λ is the reorganization energy and ΔG_0 is the free energy difference between the reactant and product states. $H_{ij}(x^{\ddagger})$ and $H_{ij}(x_0)$ are the matrix element that mix the reactant and the product states at the TS and the reactant geometries, respectively. In the case where $\Delta G_0 \ll \lambda$ and $H_{ij}(x_0) \ll \lambda$, we can write:

$$\Delta g^{\ddagger} \cong \frac{\lambda}{4} - H_{ij}(x^{\ddagger}) \quad (25)$$

The reorganization energy of eq 25 can be expressed in terms of the solute and solvent coordinates using (see ref. 31);

$$\lambda = \lambda_Q + \lambda_S \cong \frac{1}{2}\hbar\omega_Q\Delta_Q{}^2 + \frac{1}{2}\hbar\omega_S\Delta_S{}^2 \quad (26)$$

where Δ_Q and Δ_S are the dimensionless "origin shifts" between the free energy minima of the reactant and product states along the solute and solvent coordinates respectively and ω_Q and ω_S are the effective solute and solvent frequencies, (the frequencies are assumed to be identical in both the reactant and product states). With our EVB mapping approach we can evaluate λ_S and λ_Q. Figure 6 presents the solvent (a) and solute (b) contributions to the free energy of the reactant and product states for both the reaction in water (dashed lines) and in the protein (solid lines). As seen from the figure, the difference between the protein and the solution reaction is manifested mainly in the difference between the solvent reorganization energies, λ_S^w and λ_S^p. This means that the protein catalyzes its reaction primarily by reducing the reorganization energy along the solvent coordinate. In other words, the polar environment of the enzyme active site is already partially preorganized to stabilize the TS relative to the corresponding state in water. Thus, the reorganization energy is smaller in the enzyme than in water.

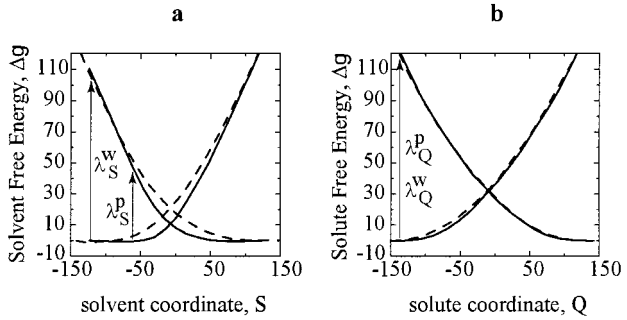


Figure 6. Calculated solvent (a) and solute (b) contributions to the diabatic energies of the reactant and product states for the S_N2 reaction in water (dashed lines) and in the protein (solid lines). For convenience, the difference between the reactants and products minima is forced to be zero in all cases. The figure shows that the largest change in the reorganization energy is associated with the solvent coordinate.

It is pertinent to comment on the relationship between the above environmental preorganization energy concept and a recent study of formate dehydrogenase,[19] in which the interactions of the enzyme with the reacting groups were found to be similar in the RS and TS. The observation that the structure of the enzyme does not change considerably during the reaction was used to support the NAC hypothesis. However, this finding is, in fact, consistent with our environmental preorganization concept, where the enzyme dipoles in the RS are already partially oriented toward their TS configuration (see above and

(61) King, G.; Warshel, A. *J. Chem. Phys.* **1990**, *93*, 8682−8692.
(62) Chapter 3.7 of ref *2*.
(63) Marcus, R. A. *Annu. Rev. Phys. Chem.* **1964**, *15*, 155−196.

in refs 33, 32, 64). This effect corresponds to a restraint in the environment and not, as claimed, to a NAC arrangement of the reacting fragments. As stated above, the preorganization of the enzyme helps to reduce the reorganization energy and is the primary basis of the catalytic effect.

6. Concluding Remarks

This work examines the magnitude of the steric contribution to the catalytic effect of DhlA using two independent approaches based on different thermodynamic cycles. By trying to mimic the protein steric effect in water using the b, θ constraints, we obtained a steric contribution of about −2.2 kcal/mol. A second estimate, obtained by a direct calculation of the nonelectrostatic effect via removal of the electrostatic contributions gave −0.7 kcal/mol. This means that the overall steric effect of around −2.2 kcal/mol contains about −1.5 kcal/mol electrostatic contribution. The corresponding solvation-induced repulsion between the carboxylate and the substrate (see section 4) is clearly not what is usually meant by steric effect. That is, in the traditional steric picture this increase in the repulsion should be associated with the solute coordinates. However, as shown in the previous section the solute contribution to the reorganization energy (λ_Q) is similar in protein and in solution. At any rate, even with the ∼−2.2 kcal/mol contribution of an apparent steric effect (that includes coupling with the electrostatic contributions) we find that the NAC effect is not the primary catalytic effect. Apparently, although the probability distributions of the RS configurations are different in the enzyme and in the solution cases (Figure 1), the free energy equivalence of this difference is not as large as previously proposed.

The conclusions from the present study are different from those obtained for qualitative considerations (e.g., refs 5, 12−16, 19, 21, 27). This fact emphasizes the importance of a rigorous definition of the thermodynamic cycles used and of comparing the reaction in the protein to the corresponding reaction in solution. In addition, our study emphasizes the importance of using the same computational model for the reaction in the protein and in solution. It is also crucial to evaluate the actual free energy associated with the protein steric effect, rather than to use other estimates (e.g., average interaction distance), which cannot provide a quantitative measure of the magnitude of the steric effect. In summary, our work has shown that the steric effects do not account for a major part of the catalysis in DhlA. In addition, we have found that the electrostatic effects are responsible for most of the overall catalytic effect. These findings provide a further support to the concept that enzyme catalysis is a result of TS stabilization caused mainly by electrostatic contributions.[3,60]

Acknowledgment. This work was supported by National Institute of Health Grant GM24492.

Supporting Information Available: Tables of EVB calculations of activation free energies and steric effect using all the Cartesian coordinates of the reacting system (PDF), along with the parameters used to describe the EVB potential energy surface. This material is available free of charge via the Internet at http://pubs.acs.org.

JA012230Z

(64) Cannon, W.; Benkovic, S. *J. Biol. Chem.* **1998**, *273*, 26257−26260.
(65) Swain, G.; Scott, C. B. *J. Am. Chem. Soc.* **1953**, *75*, 141−147.
(66) Schanstra, J. P.; Kingma, J.; Janssen, D. B. *J. Biol. Chem.* **1996**, *271*, 14747−14753.
(67) Eyring, H. *Acc. Chem. Res.* **1935**, *17*, 65−77.
(68) Glasstone, S.; Laidler, K. J.; Eyring, H. *The Theory of Rate Processes*; McGraw-Hill: New York, 1941.