On the Generation of Catalytic Antibodies by Transition State Analogues

Montserrat Barbany,[a] Hugo Gutiérrez-de-Terán,[a] Ferran Sanz,*[a] Jordi Villà-Freixa,*[a] and Arieh Warshel*[b]

The effective design of catalytic antibodies represents a major conceptual and practical challenge. It is implicitly assumed that a proper transition state analogue (TSA) can elicit a catalytic antibody (CA) that will catalyze the given reaction in a similar way to an enzyme that would evolve (or was evolved) to catalyze this reaction. However, in most cases it was found that the TSA used produced CAs with relatively low rate enhancement as compared to the corresponding enzymes, when these exist. The present work explores the origin of this problem, by developing two approaches that examine the similarity of the TSA and the corresponding transition state (TS). These analyses are used to assess the proficiency of the CA generated by the given TSA. Both approaches focus on electrostatic effects that have been found to play a major role in enzymatic reactions. The first method uses molecular interaction potentials to look for the similarity between the TSA and the TS and, in principle, to help in designing new haptens by using 3D quantitative struture–activity relationships. The second and more quantitative approach generates a grid of Langevin dipoles, which are polarized by the TSA, and then uses the grid to bind the TS. Comparison of the resulting binding energy with the binding energy of the TS to the grid that was polarized by the TS provides an estimate of the proficiency of the given CA. Our methods are used in examining the origin of the difference between the catalytic power of the 1F7 CA and chorismate mutase. It is demonstrated that the relatively small changes in charge and structure between the TS and TSA are sufficient to account for the difference in proficiency between the CA and the enzyme. Apparently the environment that was preorganized to stabilize the TSA charge distribution does not provide a sufficient stabilization to the TS. The general implications of our findings and the difficulties in designing a perfect TSA are discussed. Finally, the possible use of our approach in screening for an optimal TSA is pointed out.

KEYWORDS:

catalytic antibodies · Langevin dipoles · molecular interaction potentials · structure–activity relationships · transition states

1. Introduction

The idea of using antibodies to catalyze chemical reactions can be traced back to Jencks[1] and, in some respects, to the catalytic concept of Pauling.[2] The practical implementation of this idea by the generation of catalytic antibodies (CAs) has opened a major research field (see, for example, refs. [3–10]) and led to major excitement in the chemical and biochemical communities. It appears that (at least in principle) the CA approach can offer a highly efficient and specific way for organic synthesis and that the development of CAs will provide a powerful tool for understanding enzyme catalysis. The crucial point in the CA hypothesis is the assumption that it should be possible to design a transition state analogue (TSA) for the reaction to be catalyzed. This TSA can then be used as a hapten to elicit the CA.

It seems reasonable to assume and to hope that CAs can be as effective as the corresponding enzymes. This is based on the idea that it is possible to elicit antibodies with optimal binding to specific haptens (see, however, ref. [11]). Thus, if we can find a "perfect" TSA we should, in principle, be able to elicit a CA that will bind the true transition state (TS) almost as strongly as the enzyme does. Since the binding energy of the TS is given by k_{cat}/K_M (see, for example, ref. [12]; k_{cat} = rate constant for the reaction step in the enzyme, K_M =) we should be able to optimize $(k_{cat}/K_M)/k_{non}$ (which has been defined by Radzicka and Wolfenden[13] as the enzyme "proficiency"; k_{non} = rate constant of the uncatalyzed reaction). Now, in principle one would like to optimize the value of k_{cat} (as is the case in many enzymes) and not only k_{cat}/K_M. Unfortunately, this requires ground state (GS) destabilization and it is hard to "teach" the CA to destabilize the GS by using a TSA. Thus, we can only hope that the stabilization of the TS will result in an optimal k_{cat} value. Note in this respect that, in many enzymatic systems, the optimization of the k_{cat}/K_M value frequently results also in the optimization of the k_{cat} value.[14] At any rate, despite the hope that CAs will be able to produce k_{cat}/K_M values as large as the corresponding enzymes, it

[a] *Prof. F. Sanz, Dr. J. Villà-Freixa, M. Barbany, H. Gutiérrez-de-Terán*
Computational Structural Biology Laboratory
Research Group on Biomedical Informatics (GRIB)—IMIM/UPF
Passeig Marítim de la Barceloneta 37–49
08003 Barcelona (Spain)
Fax: (+34) 93-224-0875
E-mail: fsanz@imim.es, jvilla@imim.es

[b] *Prof. A. Warshel*
Department of Chemistry
University of Southern California
Los Angeles, CA 90089-1062 (USA)
Fax: (+1) 213-740-2701
E-mail: warshel@usc.edu

  1439-4227/03/04/04 $ 20.00+.50/0

is commonly found that the proficiency of the CAs is relatively low when compared to the enhancement obtained by enzymes.[9] Thus, it is important to find out why the TSA concept has not satisfied the early expectations and what can be done (in principle) to improve the situation. Possible reasons for the difficulties with the TSAs-elicit-CAs (TSA/CA) concept have been eloquently considered before.[7, 9] The reasons mentioned include difficulties in designing the proper haptens, limitations on the biological process that produces antibodies, and breakdown in the relationship between substrate binding and catalysis. While these general points cover a wide range of possibilities there are no quantitative demonstrations of the importance of different limitations and thus no clear way to estimate a priori the limits on the rate enhancement expected from a given TSA.

The purpose of the present work is to examine the fundamental problems with the search for haptens, which we believe to be the most serious challenge in the TSA/CA concept. In fact, we have suggested before[15] that it is impossible to design an optimal TSA for most enzymatic reactions since these reactions involve several TSs with similar activation barriers, and no single TSA can elicit CAs that will stabilize several TSs in a similar way. Here, however, we will focus on the simple problem of systems with a single high-energy TS.

A previous attempt to address the search for optimal haptens was reported by Tantillo and Houk,[16] who addressed the similarity of the structures and charge distributions of TSs and TSAs for ester hydrolysis reactions. This study evaluated the molecular electrostatic potentials (MESPs) on the van der Waals surfaces of the TSA and the TSs of the hydrolysis reaction. It was pointed out that the phosphonate haptens used do not reproduce a perfect mimic of the TSs and that a better CA may be obtained by hapten redesign.

While the approach of Tantillo and Houk is insightful, it does not provide a quantitative measure of the difference between the TS and TSA nor an estimate of the proficiency of the CA that will be elicited by the given TSA. In order to quantify the TSA/CA concept we need a model that will allow one to compare the proficiency of an ideal CA (the CA that was fully evolved to provide the best binding to a TSA) to the proficiency of a hypothetical enzyme that was evolved to stabilize the corresponding TS. This should be done for the common case where we do not have information about the CA. Our search for such a model focuses on electrostatic and structural complementarity for several reasons. First, it appears from accumulating simulation studies that electrostatic effects are the primary source of enzymatic catalysis (see, for example, refs. [12, 14, 15]). Other factors such as entropic effects might still contribute to increase the rate of catalysis, but this effect is usually overestimated.[17] Furthermore, entropic effects cannot help in increasing the k_{cat}/K_M value relative to the uncatalyzed reaction (the motion is more restricted in the TS of the CA than in the TS in water). Thus, it appears that the largest catalytic effects are related to the electrostatic complementarity of the preorganized polar active site.[15, 18] Second, although the binding step usually involves hydrophobic effects, these effects cannot help in a significant way in the chemical step, which involves mainly changes in the charge distribution of the reactant region. Here we assume that a perfect CA binds the nonreacting parts of the TS of the substrate as well as the corresponding enzyme does for the TS. Thus we can focus on the difference in binding of the reacting part of the TS and the corresponding parts of the TSA.

In this work we follow the above considerations and develop two general approaches that should provide measures of the maximum proficiency of an optimal CA that would be elicited by a given TSA. The first approach looks for an optimal "haptenophore" in analogy to what is done in quantitative structure–activity relationships (QSAR) with pharmacophores (see, for example, ref. [19]). That is, we try to find the difference between the TS and TSA in terms of the corresponding steric and electrostatic features. This approach should allow one, in principle, to search for a haptenophore that will lead to a TSA able to elicit the optimal CA. The second approach is based on building a "virtual" active site that provides the best solvation to the TSA and then trying to evaluate the solvation of the real TS in this active site. The two approaches allow us to "predict" the reduced proficiency of the CA and, thus, should provide a way of assessing the quality of different TSA candidates and for elucidating fundamental difficulties with the TSA/CA concept. Section 2 describe the methods used, Section 3 presents the results obtained on a test case of the Claisen rearrangement reaction of chorismate to prephenate, and Section 4 discusses the relevance of these results in the generation of TSAs for CA generation.

2. Methods

Our first approach is based on the assumption that the electrostatic potential generated by the TSA will lead to a CA site with a complementary field. Thus, the actual TS will try to align itself by optimizing its interaction with the field of the CA generated from the TSA. Obviously, the closer the potentials of the TSA and the TS are, the more effective the CA will be. With this in mind, we try to quantify the similarities between the electrostatic potentials from the TS and TSA. Complementary information from classical molecular interaction potentials will also be used to rationalize the steric and hydrophobic differences between TS and TSA. We describe below the computational procedures used to assess the similarity between the TS and TSA electrostatic potentials.

As a starting point for our calculations we generate the structures of the TS and TSA by gas-phase ab initio calculations. This is done here at the HF/6-31+G* level of theory as implemented in the GAUSSIAN 98 set of programs.[20] One can improve the structures by including solvation effects but these effects are not so crucial in cases of compact structures (such as those considered here) as in studies of more "loose" TS/TSA systems.[21]

The generation of a haptenophore, as is the case of a pharmacophore in quantitative structure–activity relationships, involves the utilization of descriptors that represent the physicochemical properties of the molecules we are seeking. Our approach for obtaining such descriptors follows the procedures in 3D-QSAR studies. We first create a box around the TS and build a grid of points where we will evaluate the

molecular field, by using both quantum and classical mechanics, and represent molecular interaction potentials (MIP) with a given probe. In particular, the MESP at a particular position **r**, equivalent to a proton probe in that position, is calculated from the electronic distribution, $\rho(\mathbf{r})$, and from the *N* nuclei according to Equation (1), where Z_α is the charge on nucleus α, located at $\mathbf{R}_\alpha$.

$$\phi_{\mathrm{MESP}}(\mathbf{r}) = \sum_{\alpha}^{N} \frac{Z_\alpha}{|\mathbf{r}-\mathbf{R}_\alpha|} - \int \frac{\rho(\mathbf{r}')d^3\mathbf{r}'}{|\mathbf{r}-\mathbf{r}'|} \quad (1)$$

The integral in [Eq. (1)] runs over all space. The MESP is calculated with the quantum package GAMESS.[22] In addition, several classical probes are used to evaluate MIPs with the program GRID.[23] The two programs are interfaced by the program MIPSIM,[24] a computational package designed to analyze and compare 3D distributions of the MIPs of series of biomolecules. In particular, MIPSIM can obtain similarity indices and calculate superpositions of molecules based on a single MIP or a combination of them.

Two types of calculations are carried out. In the first, a grid of $27 \times 27 \times 24$ points (17 496 points), spaced 0.5 Å apart, is created around both the TS and the TSA, and the MIP (either quantum or classical) is evaluated on them. The classical MIPs are evaluated with probes OH and DRY, available in the program GRID. These probes reproduce the potentials of, respectively, a phenolic hydroxy group and a hydrophobic molecule. The MESP, obtained at the HF/6-31 + G* level of theory, is used as the quantum MIP. Since the exploration of the MESP requires a spread electronic description around the charged carboxylate groups, diffuse functions should be included in the basis set. A conjugated gradient optimization is performed by the module MIPMin of MIPSIM in order to find the minima of the MIP, so additional MIP values are computed as requested by the algorithm. MIP analysis is also performed on the regions of interest for reactivity, namely along the breaking and forming bonds in the TS and their counterparts in the TSA structure.

A second set of calculations involves the superposition of molecules based on their MIPs. The three probes (MESP, OH, and DRY) may be combined in different ways within MIPSIM (see below) in order to generate superpositions of the two systems (TS and TSA). The optimal superposition of the TS and the TSA is obtained as follows. A grid of $23 \times 23 \times 23$ points (12 167 points), spaced 0.7 Å apart, is created around the two molecules, and the MIP is evaluated on them. One of the molecules (the actual TS) is considered as *fixed* and the other (the TSA) as *mobile*, being allowed to rotate and translate until a good superposition, based on similarity indexes obtained from the molecular fields, is achieved. In MIPSIM the potentials around the molecules are calculated only once at the beginning of the calculation, and thus, we need some procedure to avoid noncoincident box problems during the superposition process. This is done by first selecting some points of the MIP that will be used for the superposition. It is, in principle, possible to select only points that are in the van der Waals surface of the molecule;[25] however, in MIPSIM we make use of user-defined regions, and in particular, we can use the whole grid of points. At every orientation of the *mobile* molecule and for every MIP *k*, the similarity index is calculated by a Gaussian coefficient of the form given in Equation (2), where V_i^X is the *i*th potential value for molecule X and r_{ij} is the distance between the two points.[24, 25]

$$s_k = \frac{\sum_{i=1}^{nA}\sum_{j=1}^{nB} V_i^A V_j^B \exp(-\alpha r_{ij}^2)}{\sqrt{\sum_{i=1}^{nA}\sum_{j=1}^{nA} V_i^A V_j^A \exp(-\alpha r_{ij}^2)}\sqrt{\sum_{i=1}^{nB}\sum_{j=1}^{nB} V_i^B V_j^B \exp(-\alpha r_{ij}^2)}} \quad (2)$$

The parameter α is taken as 0.5. Following this procedure we can evaluate a different similarity index for every "probe" we are interested in. Equation (3) computes the final similarity index in the general case in MIPSIM.

$$S = \frac{\sum_{k=1}^{N_{\mathrm{probes}}} w_k s_k}{\sum_{k=1}^{N_{\mathrm{probes}}} w_k} \quad (3)$$

In Equation (3), w_k is the weight of every particular similarity index s_k. However, in this work we have used weights of 0 for all probes except for the MESP, as we will discuss in the next section. The optimal relative orientation is obtained by a conjugated gradients optimization of *S* as a function of the three rotations and three translations of the *mobile* molecule. The flexibility of the molecules is not considered in the present treatment. The superposition calculations were done using the MIPCOMP module in MIPSIM.

While the above method provides a powerful way of analyzing the differences between the TS and TSA, our aim is to quantify the difference between the proficiency of the CA and the corresponding enzyme. In order to address this issue we developed a second approach that generated an actual complementary environment to the TSA. This generation is done by using the Langevin dipoles (LD) approach (see, for example, ref. [12]) to reproduce the solvation free energy of the TS in different preorganized environments. The LD approach generates a grid of dipoles around different solutes and estimates the corresponding solvation free energy. Usually the dipoles are allowed to reorient upon introduction of a given solute and this involves investment of reorganization energy, as is the situation in water. However, we can also force the grid to be preorganized to a given charge distribution and, thus, simulate a perfect preorganized enzyme. With this in mind, we can express the activation free energies of the reactions in the CA and in the perfect enzyme by Equations (4a) and (4b), where $\Delta G_{\mathrm{sol}}(\mathrm{X})_{\mathrm{y}}$ is the solvation energy of a solute, X, in an environment, Y, that was polarized (preorganized) to solvate the solute Y.

$$\Delta g^{\ddagger}_{\mathrm{CA}} \approx \Delta G_{\mathrm{sol}}(\mathrm{TS})_{\mathrm{TSA}} + \Delta - \Delta G_{\mathrm{sol}}(\mathrm{S})_w \quad (4a)$$

$$\Delta g^{\ddagger}_{\mathrm{perfect}} \approx \Delta G_{\mathrm{sol}}(\mathrm{TS})_{\mathrm{TS}} + \Delta - \Delta G_{\mathrm{sol}}(\mathrm{S})_w \quad (4b)$$

Δ is the activation free energy in the gas phase (assuming for simplicity that the activation energy can be obtained by adding the solvation energies of the TS and the reactant state (RS) to the

gas-phase activation barrier). Finally, $\Delta G_{sol}(S)_w$ is the solvation free energy of the substrate (S) in water. Note that we view the "solvation energy" here as a formal way of expression of the free energy of moving the given solute from the gas phase to the indicated site. A similar cycle could have been obtained by moving the TS from water to the protein site. At any rate, the difference between the proficiency of the CA and the corresponding enzyme is given now by Equation (5).

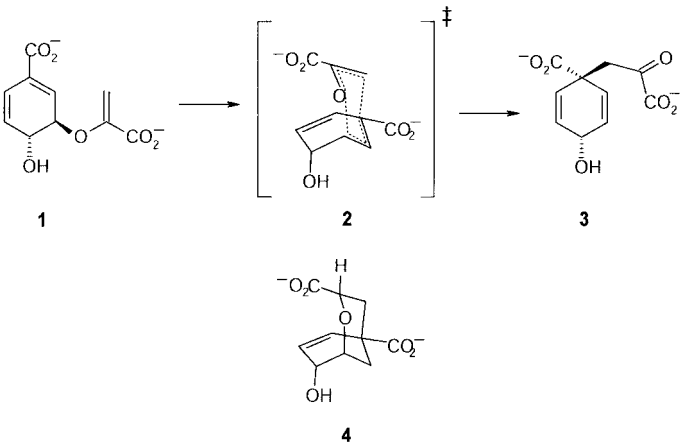
$$\Delta g^{\ddagger}_{CA} - \Delta g^{\ddagger}_{perfect} \approx \Delta G_{sol}(TS)_{TSA} - \Delta G_{sol}(TS)_{TS} \quad (5)$$

Although this model cannot reproduce the details of the actual enzyme and CA, it does capture key physical requirements of the CA concept. For example, it is clear that if the TSA will have the same charge distribution and shape as the TS, we will have a perfect CA. Conversely, if the TSA charge distribution is significantly different from the charge distribution of the TS we will have a poor CA.

In principle we should include in $\Delta G_{sol}(TS)_{TSA}$ the steric interaction between the TS atoms and the TSA-generated grid or allow the TS to find an optimal orientation in this grid. In the present work, however, we use a simplified procedure and place the TS residual charges at the positions of the corresponding TSA atoms in the evaluation of $\Delta G_{sol}(TS)_{TSA}$.

3. Results

In order to examine our approach we chose the Claisen rearrangement of chorismate to prephenate. This extensively studied reaction (see, for example, ref. [9]) involves a single TS and thus can serve as a clear test case. Scheme 1 shows the relevant structures in the mechanism of the Claisen reaction. This reaction is both enzyme catalyzed[26–28] and antibody catalyzed.[29–31] The structures of both the enzyme chorismate mutase (CM) in several organisms (for example, ref. [32–34]) and the CA (1F7)[35] have been solved. For this reaction, $\Delta g^{\ddagger}_{w} = 24.2\ kcal\,mol^{-1}$,[26] $\Delta g^{\ddagger}_{1F7} = 15.4\ kcal\,mol^{-1}$,[30] and $\Delta g^{\ddagger}_{CM} = 9.9\ kcal\,mol^{-1}$ (for CM in *Bacillus subtilis*),[36a] where $\Delta g^{\ddagger}$ in both proteins corresponds to the k_{cat}/K_M value. The determination of the 3D structure of 1F7 provided an insight into the differences between the active sites of the enzyme and the CA and, thus, offers a chance to understand the difference between the proficiencies of the two systems.[35] The CA 1F7 was elicited against the endo oxabicyclic TSA (**4** in Scheme 1),[30] which has actually been proven to be a potent inhibitor for chorismate mutase. The interesting feature of this TSA is its close structural similarity to the TS; it is only disrupted by two additional hydrogen atoms in **4** and different hybridization of several carbon atoms in the central moiety. Thus, although there are qualitative indications that the TS and the TSA are different,[37] there is no rationalization for the fact that the CA is not as proficient as the corresponding enzyme. This important issue has not been resolved despite insightful theoretical studies of the enzyme and the CA (for example, refs. [38–40]).



Scheme 1. *Claisen rearrangement of chorismate (**1**) to prephenate (**3**). The the transition state (TS) for the reaction **2** and the transition state analogue (TSA) **4** chosen for this study are shown.*

The geometries of the TSA and the TS were obtained from ab initio calculations at the HF/6-31 + G* level and are shown in Figure 1. The figure highlights the lengths obtained for the C–O bond that is being broken and the C–C bond that is being created in the reaction. The TS structure (structure **2**) was characterized as a saddle point by a normal mode analysis. The results are similar to those obtained in previous ab initio calculations[40] and show a generally good structural similarity between the TS and TSA. However, it is apparent that the distances in Figure 1 are different in the TS and the TSA. This will have a significant effect in the generation of the haptenophore, as we will see below.

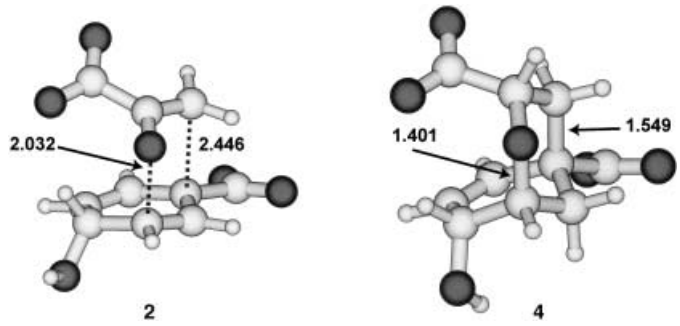


Figure 1. *Relevant geometric parameters obtained at the HF/6–31 + G* level for the TS and the TSA shown in Scheme 1.*

Figure 2 shows the quantum MESP for the TS and TSA and Figure 3 shows the molecular fields for the two GRID probes used in this work (OH and DRY). In evaluating the classical molecular fields we had to assign a specific type for each atom. In the case of the TSA, this was done automatically by GRID, but for the TS we had to add manually the atom types that better describe the atoms involved in the bond-breaking/bond-making process according to the GRID force field (see Figure 1 in the Supporting Information).

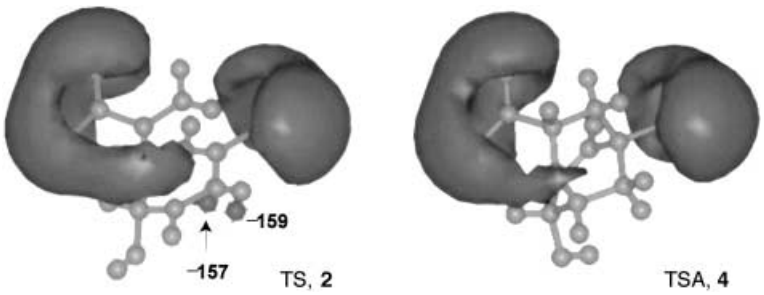


Figure 2. *Quantum molecular electrostatic potential (MESP) for the TS and TSA. The isocontour $\phi_{MESP} = -180$ kcal mol^{-1} is shown. Some minima (in kcal mol^{-1}) that appear in the MESP of the TS and not in the TSA are also depicted in the figure.*

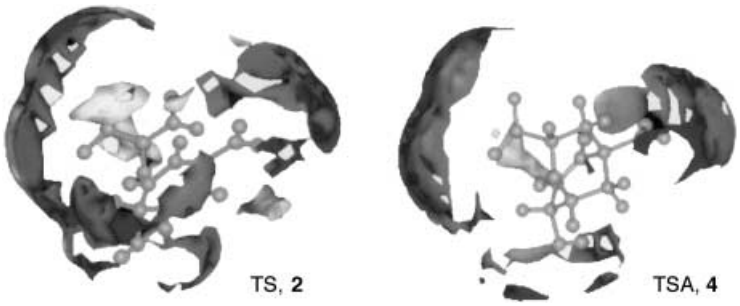


Figure 3. *Classical molecular fields obtained with GRID probes OH (dark gray) and DRY (light gray) for the TS and TSA. The isocontours are shown for $\phi_{OH} = -5$ kcal mol^{-1} and $\phi_{DRY} = -0.2$ kcal mol^{-1}.*

Figures 2 and 3 do not show apparent differences between the molecular fields of the TS and TSA. This indicates that both the shape and the general electrostatic trends of the TS are apparently well reproduced by the suggested TSA. The way 1F7 recognizes the TSA is expected to be similar to the way CM recognizes the TS. Now, despite the similarities, our task is to try to understand and quantify the *difference* between the TS and TSA, and for this we will focus on the respective MESPs.

In order to better describe the differences between the TS and TSA we used the module MIPComp in MIPS$_{IM}$ and superposed the two structures according to the potential generated by these structures. In particular, we used the MESP to align the two molecules based only on how they are recognized by the CA or the enzyme. Thus, in the calculations presented here, we selected for comparison of the two molecular fields only those points in the grid with $\phi_{MESP} < -200$ kcal mol^{-1}. In this way, we ensured that the two structures (TS and TSA) are superposed based on the most apparent features of their electrostatic potential (these are the MESPs generated by the carboxylate groups). These features will be those first recognized by the complementary CA. The search for the optimal alignment was done by the following protocol: first, we randomly generated ten rotations and translations starting from the initial configuration; next, the function in Equation (3) was optimized by the conjugated gradient approach. In our algorithm, trajectories in the rotations – translations space are collected, and when a given optimization crosses a saved trajectory, the current path is rejected. In this way, a final set of five relative orientations (tests) between the TS and TSA was collected and the final similarity index for the optimal orientation (test 1) was 0.991. Figure 4 shows the difference between the two MESPs obtained after superposing the structures of test 1.

As seen from Figure 4, the superposition of the TS and TSA, based on the positions and charge distribution of their carboxylate groups, yields an almost perfect match. However, a part of the impression of perfect superposition is due to the overwhelming effect of the two carboxylates that are common to both the TS and TSA. Thus, we have to look more carefully at the three regions in the isocontours of Figure 4. Region I corresponds to differences in orientation of the hydroxy group, which can be neglected because of the almost free rotation of the C–OH bond. Region II corresponds to the positions of the superimposed carboxylate groups. Finally, region III corresponds to the differential MESP in the central region. The differences are due to the electron density in the bond-breaking/bond-making pattern in the TS with respect to the TSA. In order to avoid problems arising from the wrong superposition of the molecules we show in Figure 5 the main characteristics of the MESP along these two bonds in the TS and the TSA. These differences may account for the difference in proficiency between chorismate mutase and 1F7 in the reaction studied.

While the MEP approach is very useful in providing a visual insight about the difference between the TS and

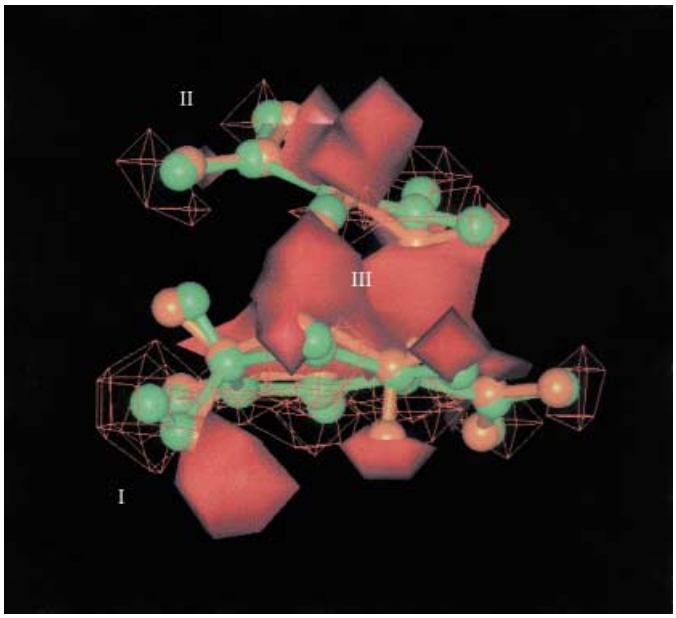


Figure 4. *Superposition of the quantum mechanical MESP grid for the TS and TSA. The superposition was performed with the MIPComp module of the program MIPS$_{IM}$ maximizing the similarity between the potential fields of the two molecules. Only the very negative values of the MESP ($\phi < -200$ kcal mol^{-1}) were selected for the similarity calculations in order to superpose the MESP around the carboxylate groups, our reference anchoring points (see text). Isocontours in the figure represent the differences between the potential fields for the actual TS and the TSA. The isocontours represent regions where $\phi_{TS} - \phi_{TSA} = -100$ kcal mol^{-1} (solid surfaces) and $\phi_{TS} - \phi_{TSA} = 100$ kcal mol^{-1} (wire frame surfaces). The meaning of regions I, II and III is described in the text.*

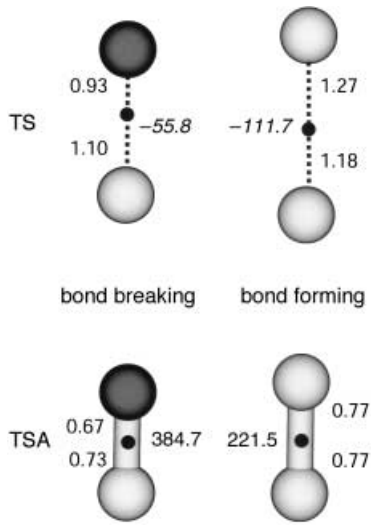


Figure 5. The geometries and MESPs (in kcal mol^{-1}) for critical points[54] for breaking the C–O bond and forming the C–C bond in the TS and the corresponding bonds in the TSA. The bond being broken is represented on the left and the bond being formed is represented on the right. Carbon atoms are light gray and oxygen atoms is dark gray. The critical points (shown in black) are characterized by $\nabla\phi_{MESP} = 0$ and a Hessian matrix of ϕ_{MESP} with two negative eigenvalues[54, 55]. The value of ϕ_{MESP} at the critical point is shown in italics and the distances from it to each atom are shown in roman characters. The calculations were performed with MIPS$_{IM}$.[24]

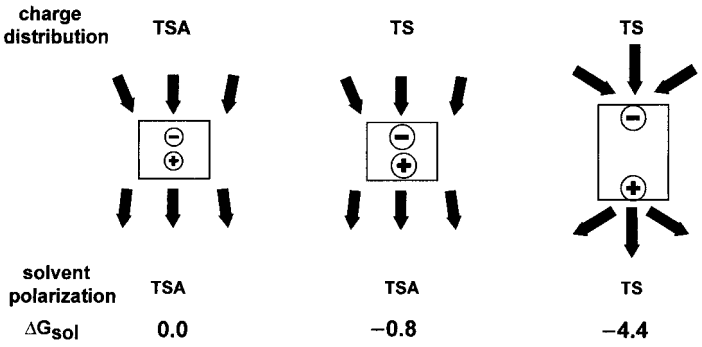


Figure 6. A schematic representation of the results shown in Table 1. The notation charge distribution describes the solute charge distribution. The notation solvent polarization describes the polarization of the corresponding LD (shown schematically by dipoles). This polarization is designated by the solute charge distribution that generated the given LD (TSA or TS). The value of ΔG_{sol} is given in kcal mol^{-1}. The size of the spheres correlates with the magnitude of the hypothetical corresponding charge. The size of the rectangle represents the differences in structure of the TSA (small rectangle) and the TS (large rectangle) and further emphasizes the larger charge separation in the TS than in the TSA.

TSA it is important to try to convert this difference into a more quantitative "scoring" function. This important task was addressed by considering the LD complementarity approach described in Section 2. We started by considering the TS and TSA of Figure 1. Since these systems differ in the number of hydrogen atoms we added the charges on these atoms to the charges of the appropriate heavy atoms. Next we generated relaxed LD grids[12] for the TS and TSA and then used the potential generated by each grid to evaluate the solvation energy of the TS and TSA. The results, summarized in Table 1 and Figure 6, provided the terms needed for the evaluation of [Eq. (5)].

As seen from the table, and as is illustrated schematically in Figure 6, the TS is stabilized more than the TSA. More importantly, the qualitative trend in the difference between the proficiencies of the CA and the enzyme is obtained from [Eq. (5)]. That is, using the difference between the last entry in the last column of the table (-4.4 kcal mol^{-1}) and the second entry in the last column (-0.8 kcal mol^{-1}), we obtain $\Delta g^{\ddagger}_{CA} - \Delta g^{\ddagger}_{perfect} = 3.6$ kcal mol^{-1}. This value is in qualitative agreement with the observed difference in proficiency (6 kcal mol^{-1}). This means that our approach can convert seemingly very small changes in charge distribution and structure between the TSA and the TS into meaningful differences in proficiencies. Apparently, the LD grid that was polarized (and preorganized) to bind the TSA cannot bind the TS as effectively as a grid that was preorganized to bind the TS. The same effect must be true for the CA and the corresponding enzyme.

***Table 1.** Estimating the solvation free energies of the TS and TSA in the dipolar grids generated by different charge distributions.[a]*

Charge distribution[b]	Preorganization[c]	Structure[d]	ΔG_{sol}
TSA	TSA	TSA	0.0
TS	TSA	TSA	−0.8
TS	TS	TS	−4.4

[a] In kcal mol^{-1} with respect to $\Delta G_{sol}(TS)_{TSA}$ of the TSA structure (this is, the free energy obtained for "solvating" the TSA structure and charge distribution in a grid of Langevin dipoles generated by the designated structure and charge distribution). [b] The charge distribution of the given solute. [c] The specific preorganization (or polarization) is designed by the charge distribution that polarized the given LD environment. [d] The geometry used for the atoms of the given solute charge distribution. Note that we used the TSA structure in evaluating $\Delta G\ (TS)_{TSA}$. This choice reflects the considerations described in the last paragraph of Section 2.

Before concluding this point it might be useful to comment on the fact that the difference between the activation free energies of the reactions catalyzed by CM and 1F7 is due to the corresponding difference in activation entropy (see, for example, ref. [41]). It is not entirely clear what this trend means, since enzymes usually reduce the activation entropy. Thus, it is possible that the observed entropic effects reflect differences in solvation entropies and in conformational flexibility of the protein.[17] Furthermore, the activation entropy appears to be very different for different enzymes[36] and is likely to be different for different temperatures.[15] This supports our view that because of entropy/enthalpy compensation[42] the total activation free energy is much more relevant than its individual components. Thus our LD modeling of the CA complementarity is aimed at the corresponding "solvation" free energy.

4. Concluding Remarks

This work tries to rationalize why the rate enhancement by a CA elicited from a given TSA does not approach the rate enhancement produced by the enzymatic counterpart when this is available. This is done by indirect pharmacophore-like approaches that only take into account the molecular interaction potentials of each substrate (TS and TSA) without considering

the structural information from the enzyme and CA. The choice of the indirect approach is deliberate. In this way we are able to show a general procedure useful for cases where no structural information of the biomolecules is available.

An initial qualitative study of the relative importance of these energy contributions has been carried out through the classical GRID molecular interaction potentials and through quantum mechanical molecular electrostatic potentials. The GRID potentials show the relative importance of the hydrophobic region of the cyclohexenyl part of the TS, which in the enzyme is occupied by Phe57. This interaction is not present in the TSA, due to the additional hydrogen atom of this structure. This result, however, makes GRID potentials only partially discriminative between the affinities of different haptens to a given protein. This is because GRID potentials, due to their intrinsic classical nature, are unable to identify the changes in polarity during the bond-breaking/bond-making reaction process. In this study, however, we tried to minimize this problem by changing the atom types of the atoms involved in such processes in the TS (see Figure 1 in the Supporting Information). An alternative to the classical potentials is the use of quantum mechanical MESP, which has been extensively used in studies of molecular recognition (see, for example, refs. [16, 25, 43]). The use of the MESP in reactivity studies allows us to take into account the features of the potential in places that are inaccessible by classical probes (that is, bonds that are being broken or formed). This gives an additional insight on the way an enzyme works and provides hints on the design of better inhibitors. This indirect approach can help to overcome the lack of structural information that the organic chemist faces when developing CAs (a related enzyme is rarely available for the reactions of interest). Once the structure of the TS is known for reactions with a single barrier, the GRID potential, in combination with automated procedures,[44] can be used for a fast screening of large databases in search of better TSAs. This approach has been proven to be successful in drug design[45] by focusing on the molecular recognition between enzymes and their substrates. A further refinement step, on the first set of candidates selected in this way, may be introduced for the design of haptens that will elicit optimal CAs. This can be done by quantum MESP calculations like those described here in order to find the best charge distribution within the bonds that are being transformed in the reactive process.

The qualitative nature of the MIPSIM approach is upgraded to a semiquantitative level by using the LD approach of [Eq. (5)]. This approach provides a direct way of testing the implicit assumption of the CA concept, by generating an environment that is fully complementary to the TSA and examining its interaction with the TS. In order to illustrate the LD approach we considered the widely studied case of the Claisen rearrangement and showed that the environment generated to complement the charge distribution of the TSA is not fully complementary to the TS. In particular, we demonstrated that a relatively small difference between the charge distributions of the TSA and TS leads to significant differences in the preorganization (polarization) of the corresponding complementary environment. This makes it hard to develop a hapten that will elicit a perfect CA, since it is usually impossible to find a TSA with identical charge distribution to that of the relevant TS.

Another difficulty in the TSA/CA concept is associated with the fact that even when the TSA and TS charge distributions are relatively similar there is no information in these charge distributions about the RS charge distribution. Thus, there is no direct information about the k_{cat} value (and the corresponding $\Delta g^{\neq}_{cat}$) so that the CA might evolve to give a relatively small k_{cat} value. However, this is probably not the most crucial problem since the environment that will evolve to stabilize the TS is unlikely to over-stabilize the RS.

Perhaps the most serious problem with the TSA/CA concept is due to the fact that many reactions involve several important transition states with similar activation barriers (see, for example, refs. [46, 47]). This situation becomes very common in enzymatic reactions (when an enzyme stabilizes the highest barrier it makes its height similar to that of the lower barrier). Now, while an enzyme active site is able to evolve to stabilize simultaneously the highest transition states, it is simply impossible to find a single TSA that will display at the same time the charge distributions of several different TSs. Thus, it is impossible to use a hapten that will elicit a perfect CA for a multistep reaction with several activation barriers of a similar height.

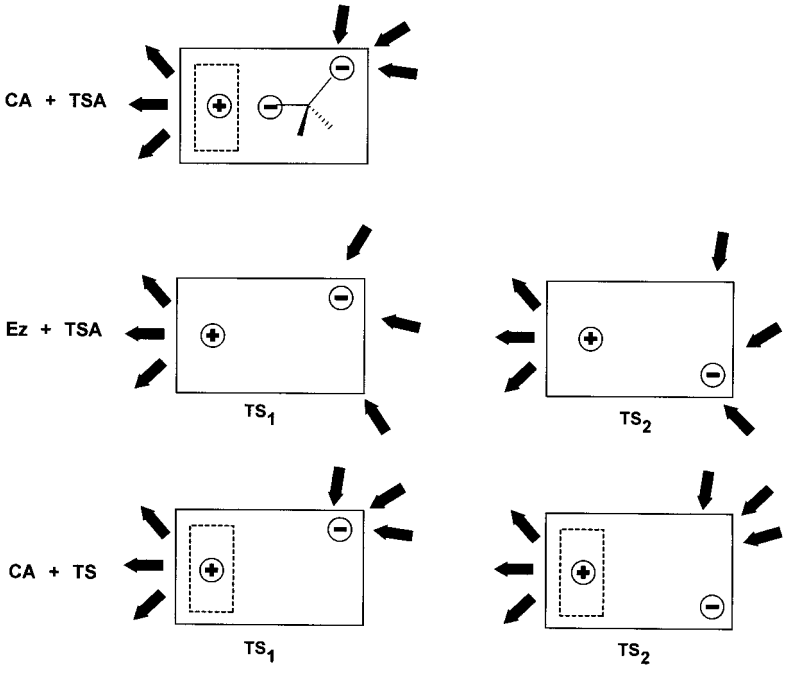


Figure 7. A representation of the difficulties with the "bait and switch" concept and with any case with several TSs. The figure considers a hypothetical attempt to elicit a CA for the reaction of serine protease by the bait and switch approach. The upper figure presents the CA which includes the catalytic histidine (designated by "+"). The second row presents the environment in serine proteases (Ez) which are evolved to stabilize two TSs (for the acylation and deacylation steps). The lower figure shows that the environment elicited in the upper figure is far from being optimal for the two TSs.

A useful way to illustrate the above problem is to consider the so-called "bait and switch" approach.[48] This innovative approach involves the placement of point charges on the hapten to elicit charged functional groups. Now let us consider the use of the approach to design a CA that works like serine proteases. Here one may try to use a hapten of the type presented in Figure 7. This hapten may elicit a histidine as a general base (Figure 7). However, even the best resulting CA will not provide as much stabilization to the TS as the corresponding enzyme. That is, as demonstrated in ref. [12], serine proteases evolved to stabilize the [$His^{+}t^{-}$] TS (where t^{-} is the oxyanion tetrahedral intermediate). On the other hand the CA would evolve to stabilize the very different charge distribution of the [$+ - -$] system presented by the hapten plus the elicited histidine. Thus, the two complementary environments are quite different. Now, an even more serious problem arises when we consider the fact that the same CA should also stabilize the second transition state which presents to the active site a new ion pair with a different charge separation than that in TS_1.[46]

The present study illustrates the difficulties in the use of TSAs to elicit CAs. It was shown that even in the case when the TSAs and TSs are quite similar and have similar binding energies the interaction between the reactive part and its surrounding is different for the CA and the corresponding perfect enzyme. This explains cases where there is no correlation between the high affinity of the hapten and the rate enhancement of the corresponding CA.[8, 49] The difficulty in knowing a priori how effective a given TSA will be is probably the reason why a large part of the search for CAs now involves screenings rather than the original TSA concept.[50, 51] Nevertheless, the search for improved TSAs can be helped by the present approaches. Hopefully it should be possible to screen the feasible synthetic candidates of already existing molecules by the GRID and LD approaches and to assess which candidate will elicit the best complementary environment.

Finally, it seems to us that the present approach should provide an effective way of developing and refining drugs. That is, our general ability to assess the similarity between TSAs and the corresponding TSs can be used to screen different inhibitors of specific enzymes (rather than CAs) and to identify the best TSAs that would be potent inhibitors.[52]

We are grateful to the Fondo de Investigaciones Sanitarias for financial support of this research (Grant no.: FIS 01/1330) and to the Centre de Computació i Comunicacions de Catalunya (C4) for computer time allocation. The work of A.W. was supported in part by the National Institutes of Health (Grant no.: GM24492). J.V.F. acknowledges the Ramón y Cajal program for funding. We also thank Dr. Gregori Ujaque and Fabien Fontaine for helpful discussions.

[1] W. P. Jencks, *Catalysis in Chemistry and Enzymology*, McGraw-Hill, New York, **1969**.
[2] L. Pauling, *Chem. Eng. News* **1946**, *24*, 1375 – 1377.
[3] A. Tramontano, K. D. Janda, R. A. Lerner, *Science* **1986**, *234*, 1566 – 1570.
[4] S. J. Pollack, J. W. Jacobs, P. G. Schultz, *Science* **1986**, *234*, 1570 – 1573.
[5] M. Powell, D. Hansen, *Prot. Eng.* **1989**, *3*, 69 – 75.
[6] N. A. Larsen, A. Heine, L. Crane, B. F. Cravatt, R. A. Lerner, I. A. Wilson, *J. Mol. Biol.* **2001**, *314*, 93 – 102.
[7] M. M. Mader, P. A. Bartlett, *Chem. Rev.* **1997**, *97*, 1281 – 1301.
[8] H. D. Ulrich, P. G. Schultz, *J. Mol. Biol.* **1998**, *275*, 95 – 111.
[9] D. Hilvert, *Annu. Rev. Biochem.* **2000**, *69*, 751 – 793.
[10] G. Pohnert, *ChemBioChem* **2001**, *2*, 873 – 875.
[11] Admittedly, there are examples of such strong binding of TSs by enzymes[13] which are probably unattainable by CAs, even if the perfect TSAs were available. However, most of these examples involve metal ions or special tricks (see ref. [53]) that are almost impossible to elicit in CAs by TSAs. It is also conceivable that the structural constraints on the inmunoglobulin scaffold and the mechanism through which hapten binding evolves prevents an optimal binding of a perfect TSA. At any rate, eliciting antibodies to TSAs that bind very strongly to specific enzymes will help to resolve this issue.
[12] A. Warshel, *Computer Modeling of Chemical Reactions in Enzymes and Solutions*, John Wiley & Sons, New York, **1991**.
[13] A. Radzicka, R. Wolfenden, *Science* **1995**, *267*, 90 – 93.
[14] A. Warshel, *J. Biol. Chem.* **1998**, *273*, 27 035 – 27 038.
[15] J. Villà, A. Warshel, *J. Phys. Chem. B* **2001**, *105*, 7887 – 7907.
[16] D. J. Tantillo, K. N. Houk, *J. Org. Chem.* **1999**, *64*, 3066 – 3076.
[17] J. Villà, M. Štrajbl, T. M. Glennon, Y. Y. Sham, Z. T. Chu, A. Warshel, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 11 899 – 11 904.
[18] A. Warshel, *Proc. Natl. Acad. Sci. USA* **1978**, *75*, 5250 – 5254.
[19] T. L. Blundell, H. Jhoti, C. Abell, *Nat. Rev. Drug Discovery* **2002**, *1*, 45 – 54.
[20] Gaussian 98 (Revision A.7), M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. A. Montgomery, R. E. Stratmann, J. C. Burant, S. Dapprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M. C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K. Morokuma, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. Cioslowski, J. V. Ortiz, A. G. Baboul, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, C. Gonzalez, M. Challacombe, P. M. W. Gill, B. G. Johnson, W. Chen, M. W. Wong, J. L. Andres, M. Head-Gordon, E. S. Replogle, J. A. Pople, Gaussian, Inc., Pittsburgh, PA, **1998**..
[21] E. C. Sherer, G. Yang, G. M. Turner, G. C. Shields, D. W. Landry, *J. Phys. Chem. A* **1997**, *101*, 8526 – 8529.
[22] M. Schmidt, K. Baldridge, J. Boatz, S. Elbert, J. Gordon, M. S. Jensen, S. Koseki, N. Matsunaga, K. Nguyen, S. J. Su, T. Windus, M. Dupuis, J. A. Montgomery, *J. Comput. Chem.* **1993**, *14*, 1347 – 1363.
[23] P. Goodford, *J. Med. Chem.* **1985**, *28*, 849 – 857.
[24] M. de Cáceres, J. Villà, J. J. Lozano, F. Sanz, *Bioinformatics* **2000**, *16*, 568 – 569.
[25] C. K. Bagdassarian, V. L. Schramm, S. D. Schwartz, *J. Am. Chem. Soc.* **1996**, *118*, 8825 – 8836.
[26] P. R. Andrews, G. D. Smith, I. G. Young, *Biochemistry* **1973**, *12*, 3492 – 3498.
[27] E. Haslam, *Shikimic Acid: Metabolism and Metabolites*, John Wiley & Sons, New York, **1993**.
[28] A. Lee, J. Stewart, J. Clardy, B. Ganem, *Chem. Biol.* **1995**, *2*, 195 – 203.
[29] P. G. Schultz, *Science* **1988**, *240*, 426 – 433.
[30] D. Hilvert, S. H. Carpenter, K. D. Nared, M.-T. Auditor, *Proc. Natl. Acad. Sci. USA* **1988**, *85*, 4953 – 4955.
[31] D. Y. Jackson, M. N. Liang, P. A. Bartlett, P. G. Schultz, *Angew. Chem.* **1992**, *104*, 196 – 198; *Angew. Chem. Int. Ed. Engl.* **1992**, *31*, 182 – 183.
[32] Y. M. Chook, H. Ke, W. N. Lipscomb, *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 8600 – 8603.
[33] A. Y. Lee, A. Karplus, B. Ganem, J. Clardy, *J. Am. Chem. Soc.* **1995**, *117*, 3267 – 3628.
[34] N. Stater, G. Schnappauf, G. Braus, W. N. Lipscomb, *Structure* **1997**, *5*, 1437 – 1452.
[35] M. R. Haynes, E. A. Stura, D. Hilvert, I. A. Wilson, *Science* **1994**, *263*, 646 – 652.
[36] a) P. Kast, M. Asif-Ullah, D. Hilvert, *Tetrahedron Lett.* **1996**, *37*, 2691 – 2694; b) C. C. Galopin, S. Zhang, D. B. Wilson, B. Ganem, *Tetrahedron Lett.* **1996**, *37*, 8675 – 8678.
[37] D. J. Gustin, P. Mattei, P. Kast, O. Wiest, L. Lee, W. W. Cleland, D. Hilvert, *J. Am. Chem. Soc.* **1999**, *121*, 1756 – 1757.
[38] Y. S. Lee, M. Hodoscek, B. R. Brooks, P. F. Kador, *Biophys. Chem.* **1998**, *70*, 203 – 216.

[39] S. Martí, J. Andrés, V. Moliner, E. Silla, I. Tuñon, J. Bertrán, M. J. Field, *J. Am. Chem. Soc.* **2001**, *123*, 1709 – 1712.
[40] O. Wiest, K. N. Houk, *J. Am. Chem. Soc.* **1995**, *117*, 11 628 – 11 639.
[41] D. Hilvert, *Acc. Chem. Res.* **1993**, *26*, 552 – 558.
[42] R. M. Levy, E. Gallicchio, *Annu. Rev. Phys. Chem.* **1998**, *49*, 531 – 567.
[43] F. Sanz, F. Manaut, J. Rodríguez, E. Lozoya, E. López de Briñas, *J. Comput.-Aided Mol. Des.* **1993**, *7*, 337 – 347.
[44] M. Pastor, G. Cruciani, I. McLay, S. Pickett, S. Clementi, *J. Med. Chem.* **2000**, *43*, 3233 – 3243.
[45] M. von Itzstein, I. C. Dyason, S. W. Oliver, H. F. White, W. y. Wu, G. B. Kok, M. S. Pegg, *J. Med. chem*. **1996**, 39, 388 – 391.
[46] M. Štrajbl, J. Florián, A. Warshel, *J. Am. Chem. Soc.* **2000**, *122*, 5354 – 5366.
[47] J. Åqvist, M. Fothergill, *J. Biol. Chem.* **1996**, *271*, 10010 – 10016.
[48] K. D. Janda, *Biotechnol. Prog.* **1990**, *6*, 178 – 181.
[49] T. Asada, H. Gouda, P. A. Kollman, *J. Am. Chem. Soc.* **2002**, *124*, 12535 – 12542.
[50] C. Gao, B. J. Lavey, C.-H. L. Lo, A. Datta, P. Wentworth, K. D. Janda, *J. Am. Chem. Soc.* **1997**, *120*, 2211 – 2217.
[51] B. DeSilva, G. Orosz, K. Egodage, R. Carlson, R. Schowen, G. Wilson, *Appl. Biochem. Biotechnol.* **2000**, *83*, 195 – 206.
[52] Supporting Information available: A figure with the atom types used for the GRID calculations on the TS and TSA (structures **2** and **4**) is available.
[53] A. Warshel, M. Štrajbl, J. Villà, J. Florián, *Biochemistry* **2000**, *39*, 14728 – 14738.
[54] C. H. Suresh, N. Koga, *J. Am. Chem. Soc.* **2002**, *124*, 1790 – 1797.
[55] R. F. W. Bader, *Chem. Rev.* **1991**, *91*, 893 – 928.

Received: 13 September, 2002
Revised version: 20 January, 2003 [F 485]