Classification:

Major category: Biological Sciences

Minor category: Biophysics

On the relationship between folding and chemical landscapes in enzyme catalysis

Maite Roca[1], Benjamin Messer[1], Donald Hilvert[2] and Arieh Warshel[1]*

[1]Department of Chemistry, University of Southern California, 418 SGM Building, 3620 McClintock Avenue, Los Angeles, CA, 90089-1062, USA

[2]Laboratory of Organic Chemistry, ETH Zürich, Hönggerberg HCI F 339, CH-8093 Zürich, Switzerland

* To whom correspondence should be address.

Arieh Warshel, E-mail: warshel@usc.edu

Phone: +1 213 740 4114; Fax +1 213 740 2701

Manuscript information:

Number of text pages: 17

Number of figures: 6

Abbreviations: CM, Chorismate Mutase; TSA, transition state analog; RS, reactant state; EVB, empirical valence bond; FEP/US, free energy perturbation/umbrella sampling; Rg, radius of gyration; CO, contact order; RMSD, root mean square deviation

Abstract

Elucidating the relationship between the free energy landscape of enzymes and their catalytic power has been one of the challenges of modern enzymology. The present work explores this issue by using a simplified folding model to generate the free energy landscape of an enzyme and then evaluating the activation barriers for the chemical step in different regions of the folding landscape. This approach is used to investigate the recent finding that an engineered monomeric chorismate mutase (CM) exhibits catalytic efficiency similar to the naturally occur dimer even though it exhibits the properties of an intrinsically disordered molten globule. It is found that the molten globule becomes more confined than its native-like counterpart upon ligand binding but still retains a somewhat wider catalytic region. Although the overall rate acceleration is still determined by the reduction of the reorganization energy, the detailed contribution of different barriers provides a more complex picture for the chemical process than that of a single path. This study provides the first systematic study of the relationship between folding landscapes and catalysis. The computational approach employed here may also provide a powerful strategy for modeling single molecule experiments and for enzyme design.

I. Introduction

Although many proposals have been put forward to rationalize the enormous catalytic power of enzymes (see ref (1, 2) for reviews) almost all of these proposals invoke a rather precise orientation of active site groups. However, protein free energy landscapes are very complex (3) and similar complexity may also applies to the landscape of activation barriers for the chemical step (2, 4-7). This raises the intriguing possibility that protein catalytic power may reflect the nature of this landscape. In fact, realization that the chemical landscape is complex originally motivated our approach of averaging calculated activation barriers in studies of enzyme catalysis (e.g. (8)). A closely related experimental observation has been provided by a recent study of Hilvert and coworkers (9, 10) who demonstrated that intrinsically disordered proteins can achieve large catalytic effects. These researchers converted a dimeric chorismate mutase (CM) from *Methanococcus jannaschii* into a highly active monomer (mMjCM). Surprisingly, despite providing essentially the same catalytic power as the native enzyme, the engineered catalyst behaves like a molten globule, a dynamic ensemble of poorly packed and rapidly interconverting conformers. This finding seems to challenge the conventional view that efficient catalysis requires an exquisitely preorganized active site structure.

The current work explores the relationship between folding and catalytic landscapes by using a simplified folding model to generate the folding landscape and then by evaluating the activation barriers for the chemical step in different regions of this landscape. The nature of the catalytic effect in both the engineered monomer and the dimeric wild-type CM, which catalyzes the conversion of chorismate to prephenate (see Fig. 1 and ref (11, 12)) in the biosynthesis of L-tyrosine (Tyr) and L-phenylalanine (Phe), is considered. Our study reproduces the observed experimental trends and reveals an interesting situation where the flat landscape of the monomer allows this system to reach preorganized catalytic configurations without paying significant preorganization energy. In addition to constituting what is probably the first systematic computational study of the landscape for enzyme catalysis, the approach used here provides a fundamental new insight into the relationship between folding and catalysis and could be an effective tool for computer aided enzyme design.

II. Simulating the landscape for folding and catalysis

In the first step of our study we used a simplified folding model to explore the free energy landscape of the monomer and dimer systems. The two systems are the homodimeric chorismate mutase from *Escherichia coli* (EcCM) (13) and the monomeric chorismate mutase mMjCM obtained by Hilvert and coworkers (6) by topological redesign of the thermostable EcCM homologue from *M. jannaschii* (MjCM).[§] The coordinates for the EcCM and mMjCM structures were obtained from the Protein Data Bank, Brookhaven National Laboratory, with PDB access codes 1ECM and 2GTV, respectively. Both enzymes adopt a helix-bundle structure (see Fig. 2) and they contain an endo-oxabicyclic dicarboxylic acid inhibitor that mimics the transition state of the CM reaction (transition state analog, TSA) at their active site.

We started by exploring the free energy of the protein configurations as a function of two well-defined parameters, specifically the radius of gyration (Rg) and the contact order (CO) (see Methods). The resulting free energy surface (which is referred to as a landscape) reflects the probability of finding the protein in different configurations, ranging from fully folded to partially unfolded. The landscapes for the monomer and dimer in the absence of TSA are shown in Fig. 3(a). As seen from the figure, the two surfaces are very different. Specifically, the monomer surface is much more extended along the CO axis than the dimer. This feature is consistent with the corresponding experimental observation that the monomer behaves like a molten globule (9). Fig. 3(b) depicts the folding landscapes in the presence of the TSA. The low energy region of the landscape becomes more confined than in the absence of TSA for both the monomer and the dimer, although this effect is much more pronounced in the case of the monomer, again in agreement with the experimental finding of a drastic reduction in molten globule character upon TSA binding (10).

Next we turned to the task of evaluating the catalytic power of the monomer and the dimer in different regions of the folding landscape. This was accomplished by calculating full Empirical Valence Bond (EVB) surfaces for: (a) explicit structures derived from the x-ray and NMR structures of EcCM and mMjCM, respectively (Region I); (b) explicit

structures corresponding to the minimum energy region of the simplified models for both enzymes, using a K'= 5 kcal/mol·Å^2 potential (see the Methods section) to constrain the distance between key catalytic residues and bound ligand to be near the corresponding native distance (Region II); and (c) explicit structures generated from a region far from the minimum of the simplified model, again with K'= 5 kcal/mol·Å^2 (Region III). The sections of the landscape corresponding to Regions II and III for the monomer and the dimer are shown in Figure S1 in supporting information (SI). As can be seen from Table 1 in SI, the procedure used for region I allows us to sample structures that are in the immediate neighborhood of the native proteins (RMSD < 1.0Å). In contrast, the approach for regions II and III provides access to structures that are further away (RMSD < 4.0 Å and > 4.8 Å, respectively). Without a simplified model it would be difficult to generate the latter structures with reasonable statistics.

The structures used in the calculations of the barriers in region I were generated by running 200 ps molecular dynamics (MD) simulations on the relaxed native structure and saving structural files each 5 ps, thus generating a total of 40 different starting conformations. In order to generate the structures for region II, we started by taking randomly simplified structures from the lowest energy portion of the folding landscape (labeled A in Fig. S1 in SI). Next we added the side chains to these simplified structures, while minimizing the distance between the simplified side chain center and the new explicit side chain center. After that, minimization and relaxation of the side chains were performed with the explicit model. Finally, we replaced the TSA by the substrate for the catalytic reaction (chorismate) and evaluated the free energy barriers. The same procedure was followed in the treatment of structures from region III (labeled B in Fig. S1 in SI).

The experimentally observed activation barriers for the monomer and dimer are approximately 16.9 and 16.3 kcal/mol (14), respectively. These can be compared with the calculated activation barriers and the probability of finding them within the population of the sample region, which are depicted as histograms in Fig. 4. As seen from the figure, the lowest barriers are found in the native region (region I) for the monomer and the

dimer, but we also find low barriers in region II (generated by the simplified model). Interestingly, it appears that the monomer has a larger region with catalytic configurations than the dimer. That is, in the monomer we find a large probability of having barriers in the 16-18 kcal/mol range in Region II, whereas the probability of encountering low barriers in the analogous region for the dimer is quite low. In region III, the probability of obtaining low activation barriers is zero in both systems.

In order to further understand the nature of the catalytic effect in the different regions of the landscape we evaluated the reorganization energies (λ) that determine the activation free energy ($\Delta g^{\ddagger}$) (see ref (8) for discussion) in regions that give low $\Delta g^{\ddagger}$ and high $\Delta g^{\ddagger}$ for both systems. The correlation between the $\Delta g^{\ddagger}$ and λ values is given in Fig. 5 for both the monomer (a) and dimer (b). As seen from the figure the regions with low catalytic efficiency involve large reorganization energy in the direction of the reaction coordinate (the calculated reorganization energy is evaluated along the reaction coordinate). This means that although we have a large accessible landscape only a small part of it provides the needed small reorganization energy.

Finally, we considered the overall nature of the catalytic landscape by sorting the activation barriers according to the root mean square deviation (RMSD) of the atomic positions from the corresponding positions in the native structure (this is possible because all the activation barriers were calculated using an explicit model regardless of the way the initial configuration was generated). The results of this analysis are depicted in Fig. 6. This figure arranges the free energy profiles with arbitrary equal spacing and thus can only be considered as a qualitative description of the actual catalytic landscape (more quantitative ordering is given in Table 1 in SI). The height of the different configurations in the reactant states (RS) should be similar, although a more quantitative treatment will be needed to evaluate the actual free energy surface in the RS. This will require adding to the free energy change in the simplified model the free energy of moving from the simplified to the explicit model (the probability of being at different regions of the explicit landscape should be determined in principle by the perturbation approach of ref. (15) that involves perturbation from the simplified to the explicit model). We have not

determined the barrier for motion on the TS ridge between the different configurations considered in Fig. 6, and this challenging task is left for subsequent studies. However, our tentative landscape is clearly instructive. For example, as seen from the figure, the catalytic configurations (with low barriers) for the wild-type EcCM dimer are confined to the native region, whereas the catalytic landscape of the monomer is more extended with some catalytic configurations in region II. That is, in the case of the monomer we have several low barriers in region II, while in the case of the dimer all the barriers in the second region are significantly higher than those in the native region and thus cannot help in the catalytic process. The implications of the present finding are discussed below.

III. Discussion

Because the relatively flat folding landscape found for the monomer appears to contradict the idea of optimized preorganization, we examined this observation from several perspectives. First, we evaluated the reorganization energy in several regions and showed that it is small only in regions with small activation barriers. We also showed in preliminary calculations (2) that the electrostatic contribution to protein stability is minimal in the regions with the largest catalytic effect, which is consistent with the idea that the protein pays in folding energy to obtain minimal reorganization energy (16). The electrostatic reorganization, which is so crucial for enzyme catalysis, can be obtained even in the case of the CM monomer. It seems, however, that in the case of mMjCM the region with low reorganization energy is wider than in the case of the dimer.

The approach exploited in this study provides a computational glimpse into the landscape that governs enzyme catalysis. It includes a statistical analysis of the probability of having small activation barriers in different regions (Fig. 4) and the landscape of activation barriers for randomly selected configurations. A more complete study will be needed to evaluate the barriers for moving between the different configurations used to construct Fig. 6 (see Fig 3 in reference (4) and the tentative barriers in Fig. 6 (a)). Such an analysis will provide a more complete picture of the coupling between the different barriers.

Although the existence of a heterogeneous set of barriers provides an interesting twist to conventional enzyme models, it does not change the basic physics of enzyme catalysis. That is, the lowest barriers still determine the average rate and these barriers are determined by the corresponding reorganization energy. In other words, as long as the barriers between the different configurations in the ground state are lower than the chemical barrier, solution of the multistate rate equation will follow the trend dictated by the lowest activation barriers. Of course, if the chemical barriers are very low (in the range of few k_BT, where k_B is the Botzmann constant and T is the temperature), we will have diffusive type kinetics. However, the chemical barriers in most enzymes are higher than 10 kcal/mol, which represents the diffusion limit, and thus are likely to determine the overall rate.

The finding of a shallow folding landscape might be considered as support for the idea that coupled motions contribute to catalysis (ref. (17) and references given in that work). However, as argued in our recent papers, (2, 16) all reactions involve coupled motions, and properly preorganized active sites have in fact evolved to minimize motions along the reaction coordinate rather than to maximize them. To further explore this idea, we calculated the coordinate vectors for the conformational change along the folding coordinate (from a partially unfolded to a folded structure) as well as the chemical reaction coordinate (evaluated between the reactant and product EVB states). Theses two multidimensional vectors calculated in the monomer are illustrated in Fig. S2 in SI; as seen in the figure, the two vectors are nearly perpendicular when the conformational motion is defined by the vector that takes the system from the native structure to a partially unfolded structure with Rg ≈ 15 Å and %CO ≈ 20 (Fig. S2b in SI). The situation is somewhat different when the protein is almost completely folded (Fig. S2b in SI), indicating that the folding coordinate is not coupled to the chemical reaction coordinate. This finding is also relevant to the idea that motions in the landscape of the monomer constitute dynamical contributions to catalysis. That is, although there are motions on the millisecond timescale in the monomer (the rate constant for conversion of the initial encounter complex between mMjCM and the TSA to give the high affinity complex, k_2 = 5.4 s^{-1} (10), is similar in magnitude to the turnover number for catalysis, k_{cat} = 3.2 s^{-1}), it is hard to see how these motions might be coupled to the chemical step.

It should be emphasized that this study has not explored what happens when the barrier for the binding step is higher than the chemical barrier, since this does not seem to be the case in CM. That is, even with the above k_2 and k_{cat} values, the chemical barrier is not much smaller than the binding barrier. Here we have to realize that there is no evolutionary pressure to reduce the chemical barrier for an enzyme-catalyzed reaction much below the diffusion controlled limit. As a consequence, it is unlikely that the chemical barrier will ever be much lower that the binding barrier. Thus, it is unlikely that the physics of our model will change significantly unless we reach the limit with a chemical barrier much smaller than the binding barrier (a case modeled recently (7)).

There is currently significant interest in the role induced fit plays in catalysis and fidelity (see discussion in ref. (4)) and here we have a clear case of induced fit. However, the induced fit idea does not explain chemical catalysis since chemical catalysis (k_{cat}) is about the barrier for the chemical step in the ES complex and not about the fact that the binding of a substrate might help in preorganizing the active site, which obviously happens in the present case. Now, the issue in the case of the monomer is not the rather obvious finding that a positively charged active site is preorganized upon binding to a negatively charged substrate, but the fact that several configurations are able to provide a similar preorganization. This finding is new and potentially very useful.

IV. Perspectives

This work has explored several fundamental aspects of the relationship between folding and catalytic landscapes, focusing on a comparison of an engineered but intrinsically disordered CM monomer and its native dimeric counterpart and exploring the observation that the molten globule protein can provide as much catalysis as the conventionally folded enzyme. Although the experimental findings may seem puzzling in view of the general assumption that an enzyme active site should be perfectly folded in the enzyme-substrate complex state, the present work demonstrates (in agreement with experiment) that the relatively flat folding landscape for the monomer acquires a deeper

minimum upon binding of the TSA. In contrast, in the case of the dimer, the surface has a clear minimum even in the absence of a ligand.

There is significant interest in the relationship between single molecule experiments and the nature of protein fluctuations and landscapes. A recent study of Prakash and Marcus (19), who focused on the relationship between electrostatic fluctuations and observed dielectric dispersion experiments, is particularly germane to our analysis. They determined the behavior of the autocorrelation function C(t) of the electrostatic energy gap between the reactant and product state, which determines the rate constant (see (19)). The next challenge is to reproduce the relevant information from actual simulations. Now, the behavior of C(t) on short timescales (nanoseconds) can be determined from the electrostatic fluctuations of the EVB energy gap obtained in the simulations of the activation barriers (2). However, the long-term behavior of the autocorrelation function is partially determined by the barriers along the configurational coordinate (the black dash barriers in Fig 6 (a)) and the fluctuations along these barriers. Thus, determining the configurational barrier and defining a "metric" for the distance between the different configurations will be extremely useful for modeling the fluctuations of the chemical barriers. This can provide a more molecular insight in the interpretation of single molecule experiments. In fact, a promising option for determining the behavior of C(t) over long time scales may involve using some Langevin dynamics approach to estimate the slow electrostatic fluctuations due to transfer between different protein configurations.

Although the present work has focused on a microscopic catalytic landscape, it illustrates the potential of using a simplified model in studies of enzyme catalysis. For example, we can use the simplified model for fast exploration of the effect of charged mutations (20) and then evaluate the free energy of moving from the simplified to the explicit model at the TS region. This strategy should be useful for more systematic studies of the relationship between the protein folding and catalysis and in computer-aided enzyme design

V. Methods

In order to explore the landscape effects and the probability of being at different configurations, it is important to be able to sample the protein configurational space in an efficient way. At present, it is hard to accomplish this task with all atom models and one viable option involves the use of a simplified protein model of the type used in simulations of protein folding (e.g.(21-26)). The version used in the present work is similar to that described in ref (15, 16). The simplified model is created by replacing the explicit side chain of each residue by an effective unified "atom" and an additional dummy atom. The unified atoms are placed at the center of mass of the corresponding side chains (with a residue dependent charge and van der Waals parameters), and the dummy atoms are placed along the corresponding C_α - C_β vectors and serve as tools for rotational transformations in the process of moving between the simplified and explicit models. The dummy atoms do not have any charge or van der Waals interactions with the rest of the system. The backbone atoms of each residue are treated explicitly, and the interactions between main chain atoms are identical to those used in the explicit model. The potential surface of the simplified model is described elsewhere (16) and is written as:

$$U_{simplified} = U_{main} + U_{main-side} + U_{side-side} + U^{self}_{solvation} \quad (1)$$

U_{main} describes the potential energy for the main chain and is basically the force field used in our explicit simulation, which is a standard part of the MOLARIS software package (27). $U_{side-side}$ describes the interaction between the side chains and is based on an "8-6" potential (as reported in ref(15, 16)). The $U_{main-side}$ term describes the interaction between the effective side chains and the main chain atoms, whereas $U^{self}_{solvation}$ accounts for the change in the solvation energy of each of these groups upon moving from water to its protein site.

This simplified model can be used to determine the free energy of the protein as a function of any given set of coordinates (e.g., contact order, native contacts, or native

hydrogen bonds (28, 29)). In the present case, we evaluated the free energy landscape in terms of two parameters, the radius of gyration (Rg) and the contact order (CO) (28), which are defined by :

$$CO = \frac{1}{LN}\sum^{N}\Delta Z_{i,j} \qquad (2)$$

where N is the total number of contacts in the protein, $\Delta Z_{i,j}$ is the number of residues separating contacts *j* and *i*, and L is the number of residues in the protein.

The free energy surface was evaluated by using the free energy perturbation umbrella sampling (FEP/US) method (30, 31) in the same way as in our previous study (16) as a function of the radius of gyration, Rg, and sorting the results in two dimensions (X_1=Rg, X_2=CO). The starting points for the free energy landscapes were taken as the structure of the simplified model after 200 ps of equilibration. Starting from this structure, we obtained the free energy surfaces, following the above FEP/US method and applying a force constant of 100 kcal/mol·Å^2 by unfolding the systems by increasing their Rg along 21 frames of 60 ps each at 300 K and with 1 fs time step.

A crucial element of our study is the evaluation of the barrier for the chemical step in different protein conformations. The reaction (see Fig. 1) was described by the empirical valence bond (EVB) approach using the same treatment employed in our previous EVB studies of CM (32) with the MOLARIS simulation program (27) using the ENZYMIX force field. The EVB activation barriers were calculated at the configurations selected by using the same free energy perturbation umbrella sampling (FEP/US) approach used in all of our EVB studies. The simulation systems were solvated by the surface constrained all atom solvent (SCAAS) model (see ref. (27)) using a radius for the explicit region of 18 Å , while long-range electrostatic effects were treated by the local reaction field (LRF) method (see ref. (27)). The FEP mapping was evaluated by 21 frames of 20 ps each for moving along the reaction coordinate with our all atom surface constrained spherical model. All the simulations were done at 300 K with a time step of 1 fs. Four catalytic residues that interact with the substrate (the catalytic residues) (Arg 9, Arg 34, Lys 45, Glu 58 in the monomer and Arg 11, Arg 28, Lys 39, Glu 52 in the dimer) were considered ionized.

The problem is, of course, to relate the activation barriers of the chemical steps to the corresponding regions on the free energy landscape. Here we exploited the simplified folding model as a reference potential for studies of the free energy surface of the explicit model (15, 16). This was done by taking points from the simplified landscape of the protein + TSA system, generating from them explicit models, and then calculating the full EVB profile starting from the given relaxed explicit model. Our preliminary exploration of this approach indicated that the lowest free energy region in the simplified model (minimum region) did not produce the best catalytic configurations, because the simplified enzyme substrate model has not been refined sufficiently in terms of protein-substrate interactions. This does not pose a fundamental problem, however, since the simplified model is only used as a reference potential for calculations of the explicit landscape. Thus, any variation of the simplified potential is allowed, provided that one can get the difference between the simplified and explicit potentials (and use it to determine the free energy of moving from the simplified to the explicit model). To that end, we added an additional term (U$^{'}$), the reflected constraint on the distance between the catalytic residues and the substrate, to the simplified potential:

$$U^{'} = K^{'}\sum_{i}(r_i - r_{0,i})^2 \qquad (3)$$

where the $r_{0,i}$ are the distances between key charged residues and the substrate in the simplified model generated from the original X-ray structure. In principle, we could evaluate the free energy landscape of the simplified model in terms of (Rg, CO and U$^{'}$) and then calculate the free energy of moving from the simplified model to the explicit model. However, at this stage we use U$^{'}$ mainly to explore different ranges in the overall landscape.

Footnotes

[§]Computations were performed with EcCM rather than MjCM, which is more closely related to the monomer, due to the absence of detailed three-dimensional structural information for MjCM. Because EcCM is mesostable, wherease MjCM is thermostable, differences between the monomer and the EcCM dimer are probably less pronounced than the corresponding differences between the monomer and the MjCM enzyme.

Acknowledgments

This work was supported by NIH grant GM24492 and the Schweizerischer Nationalfonds. We thank Mitsunori Kato for his previous work on CM and the University of Southern California´s High Performance Computing and Communication Center (HPCC) for computer time. M.R. thanks the Generalitat Valenciana from Spain for the postdoctoral fellowship.

References

1. Marti S, Roca M, Andres J, Moliner V, Silla E, Tunon I & Bertran J (2004) Theoretical insights in enzyme catalysis. *Chem. Soc. Rev.* 33**:** 98-107.
2. Warshel A, Sharma PK, Kato M, Xiang Y, Liu HB & Olsson MHM (2006) Electrostatic basis for enzyme catalysis. *Chem. Rev.* 106**:** 3210-3235.
3. Onuchic JN & Wolynes PG (2004) Theory of protein folding. *Curr. Opin. Struct. Biol.* 14**:** 70-75.
4. Xiang Y, Goodman MF, Beard WA, Wilson SH & Warshel A (2008) *Proteins* 70**:** 231-247.
5. Benkovic SJ, Hammes GG & Hammes-Schiffer S (2008) Free-Energy Landscape of Enzyme Catalysis. *Biochemistry* 47**:** 3317-3321.
6. Prakash MK & Marcus RA (2007) An interpretation of fluctuations in enzyme catalysis rate, spectral diffusion, and radiative component of lifetimes in terms of electric field fluctuations. *Proc. Natl. Acad. Sci. USA* 104**:** 15982-15987.
7. Min W, Xie XS & Bagchi B (2008) Two-dimensional reaction free energy surfaces of catalytic reaction: Effects of protein conformational dynamics on enzyme catalysis. *J. Phys. Chem. B* 112**:** 454-466.
8. Liu H & Warshel A (2007) The Catalytic Effect od Dihydrofolate Reductase and Its Mutants Is Determined by Reorganization Energies. *Biochemistry* 46**:** 6011-6025.
9. Vamvaca K, Vogeli B, Kast P, Pervushin K & Hilvert D (2004) An enzymatic molten globule: Efficient coupling of folding and catalysis. *Proc. Natl. Acad. Sci. USA* 101**:** 12860-12864.
10. Pervushin K, Vamvaca K, Vogeli B & Hilvert D (2007) Structure and dynamics of a molten globular enzyme. *Nat. Struct. Mol. Biol.* 14**:** 1202-1206.
11. Copley SD & Knowles JR (1985) The Uncatalyzed Claisen Rearrangement of Chorismate to Prephenate Prefers a Transition-State of Chairlike Geometry. *J. Am. Chem. Soc.* 107**:** 5306-5308.
12. Gajewski JJ, Jurayj J, Kimbrough DR, Gande ME, Ganem B & Carpenter BK (1987) On the Mechanism of Rearrangement of Chorismic Acid and Related-Compounds. *J. Am. Chem. Soc.* 109**:** 1170-1186.
13. Lee AY, Karplus PA, Ganem B & Clardy J (1995) Atomic-Structure of the Buried Catalytic Pocket of Escherichia-Coli Chorismate Mutase. *J. Am. Chem. Soc.* 117**:** 3627-3628.
14. MacBeath G, Kast P & Hilvert D (1998) Redesigning enzyme topology by directed evolution. *Science* 279**:** 1958-1961.
15. Fan ZZ, Hwang JK & Warshel A (1999) Using simplified protein representation as a reference potential for all-atom calculations of folding free energy. *Theor. Chem. Acc.* 103**:** 77-80.
16. Roca M, Liu H, Messer B & Warshel A (2007) On the relationship between thermal stability and catalytic power of enzymes. *Biochemistry* 46**:** 15076-15088.
17. Hammes-Schiffer S & Benkovic SJ (2006) Relating protein motion to catalysis. *Annu. Rev. Biochem.* 75**:** 519-541.
18. Min W, Gopich IV, English BP, Kou SC, Xie XS & Szabo A (2006) When does the Michaelis-Menten equation hold for fluctuating enzymes? *J. Phys. Chem. B* 110**:** 20093-20097.
19. Prakash MK & Marcus RA (2007) An interpretation of fluctuations in enzyme catalysis rate, spectral diffusion, and radiative component of lifetimes in terms of electric field fluctuations. *Proc. Natl. Acad. Sci. USA* 104**:** 15982-15987.
20. Ishikita H & Warshel A (2008) Predicting drug-resistant mutations of HIV protease. *Angew. Chem. Int. Ed.* 47**:** 697-700.

21. Levitt M & Warshel A (1975) Computer-Simulation of Protein Folding. *Nature* 253**:** 694-698.
22. Bryngelson JD & Wolynes PG (1987) Spin-Glasses and the Statistical-Mechanics of Protein Folding. *Proc. Natl. Acad. Sci. USA* 84**:** 7524-7528.
23. Hinds DA & Levitt M (1992) A Lattice Model for Protein-Structure Prediction at Low Resolution. *Proc. Natl. Acad. Sci. USA* 89**:** 2536-2540.
24. Shakhnovich E, Abkevich V & Ptitsyn O (1996) Conserved residues and the mechanism of protein folding. *Nature* 379**:** 96-98.
25. Dill KA (1990) *Biochemistry* 29**:** 7133-7155.
26. Olszewski KA, Kolinski A & Skolnick J (1996) Folding simulations and computer redesign of protein A three-helix bundle motifs. *Proteins* 25**:** 286-299.
27. Lee FS, Chu ZT & Warshel A (1993) Microscopic and Semimicroscopic Calculations of Electrostatic Energies in Proteins by the Polaris and Enzymix Programs. *J. Comput. Chem.* 14**:** 161-185.
28. Fersht AR (2000) Transition-state structure as a unifying basis in protein-folding mechanisms: Contact order, chain topology, stability, and the extended nucleus mechanism. *Proc. Natl. Acad. Sci. USA* 97**:** 1525-1529.
29. Shea JE, Onuchic JN & Brooks CL (2002) Probing the folding free energy landscape of the src-SH3 protein domain. *Proc. Natl. Acad. Sci. USA* 99**:** 16064-16068.
30. Warshel A (1991) *Computer Modeling of Chemical Reactions in Enzymes and Solutions* (Wiley Interscience, New York).
31. Hwang JK, King G, Creighton S & Warshel A (1988) Simulation of Free-Energy Relationships and Dynamics of Sn2 Reactions in Aqueous-Solution. *J. Am. Chem. Soc.* 110**:** 5297-5311.
32. Strajbl M, Shurki A, Kato M & Warshel A (2003) Apparent NAC effect in chorismate mutase reflects electrostatic transition state stabilization. *J. Am. Chem. Soc.* 125**:** 10228-10237.

Figure legends

Figure. 1. Rearrangement of chorismate to prephenate via a chair-like transition state.

Figure. 2. Three-dimensional structural representation of dimeric EcCM (left) and monomeric mMjCM (right). The active site is occupied by the transition state analog (TSA), which is represented as a ball-and-stick model in both structures.

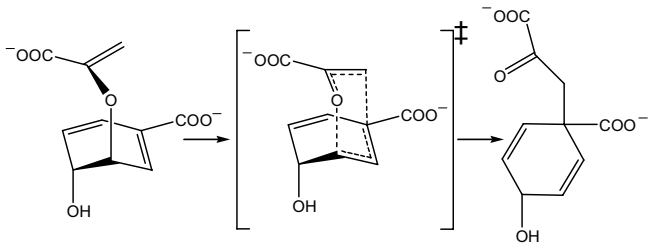
Figure 3. Free energy landscapes for the monomeric (mMjCM) and dimeric (EcCM) enzymes in the absence of TSA (a) and in the presence of TSA (b). The free energy surface is represented in terms of the radius of gyration (Rg) and the percentage contact order (%CO). Energies are expressed in kcal/mol and distances in Å.

Figure. 4. The distribution of activation barriers for the monomer and dimer for different regions of the folding landscape. Region I was generated from the native mMjCM and EcCM structures; region II corresponds to low energy structures obtained by the simplified model (see A in Fig. S1 in SI); and region III corresponds to higher energy structures from the simplified landscape (see Fig. S1 in SI). The average RMSD of structures in these regions are 0.9 Å, 3.6 Å and 5.3 Å, respectively. The figure describes the probability of having a given value of the activation barrier as a function of the value of the activation barrier.

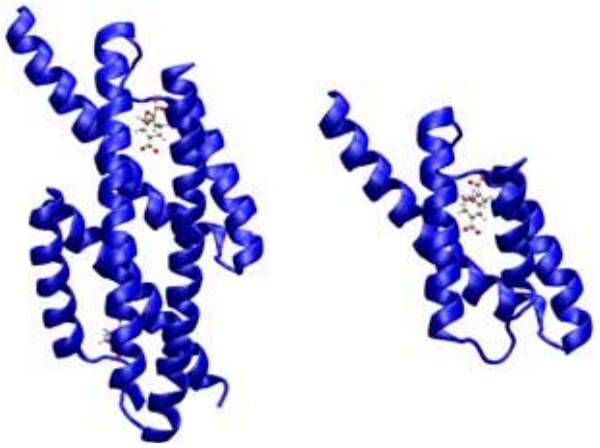
Figure. 5. The correlation between the calculated activation barriers ($\Delta g^{\ddagger}$) and the calculated reorganization energy (λ) for the monomer (a) and the dimer (b).

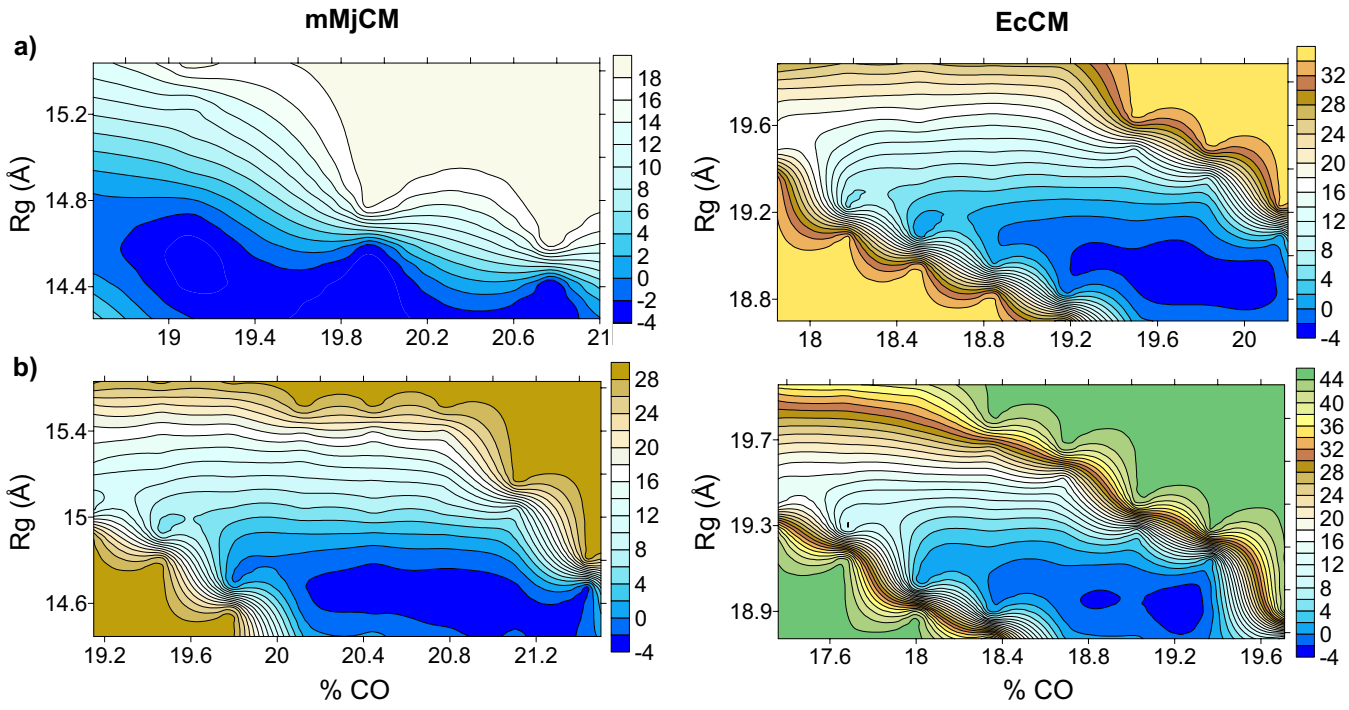
Figure. 6. The landscape for the chemical profiles for the monomer (a) and the dimer (b). The arrangement of the profiles, in equal spacing, is according to the RMSD from the native structure for the three regions (I, II and III). The dash line in orange designates the 16 kcal/mol height that corresponds to reasonably low barriers. This line allows one to see that the monomer has several catalytic configurations in the second region while the dimer does not have any (see also Fig. 4). The figure also represents in black dash lines

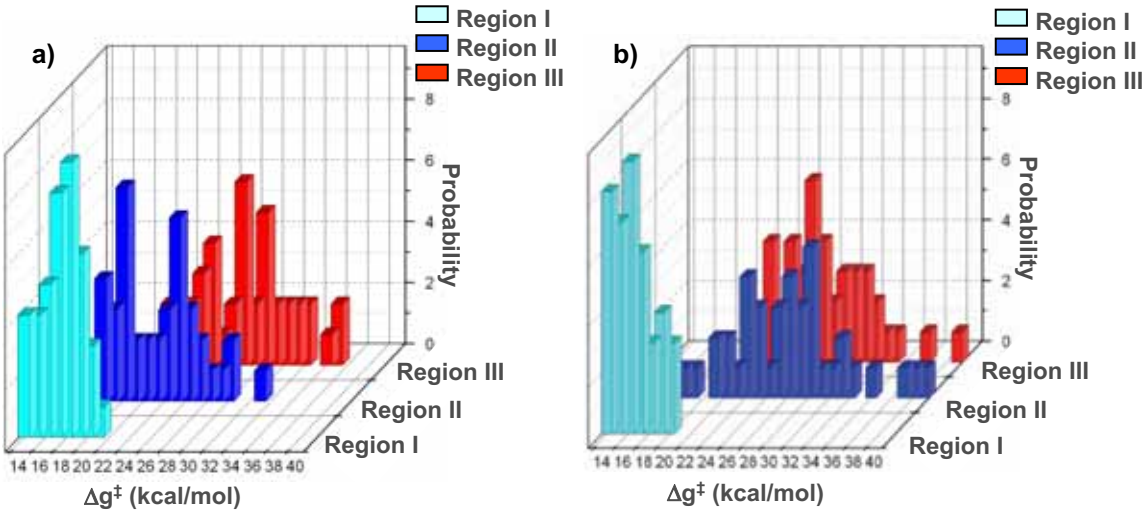
the barriers between the conformational states along the conformational coordinate. RS designates reactant state and TS transition state.

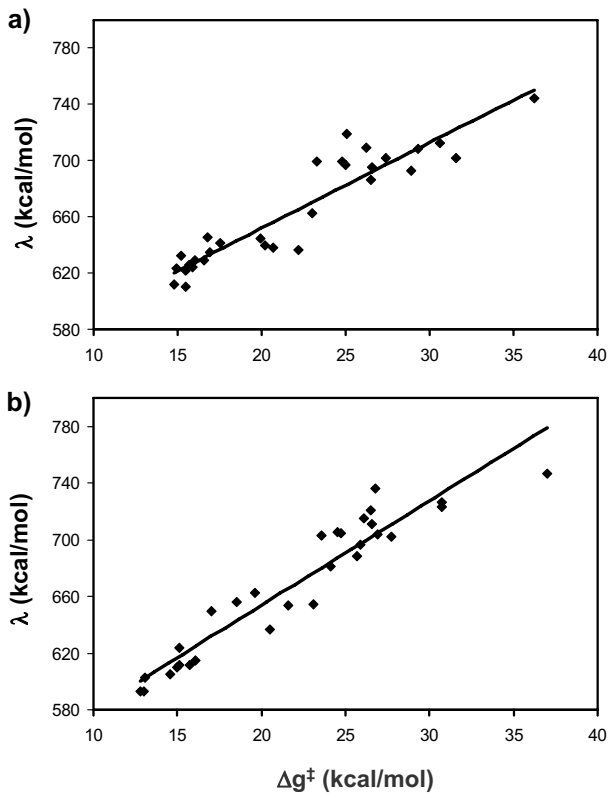


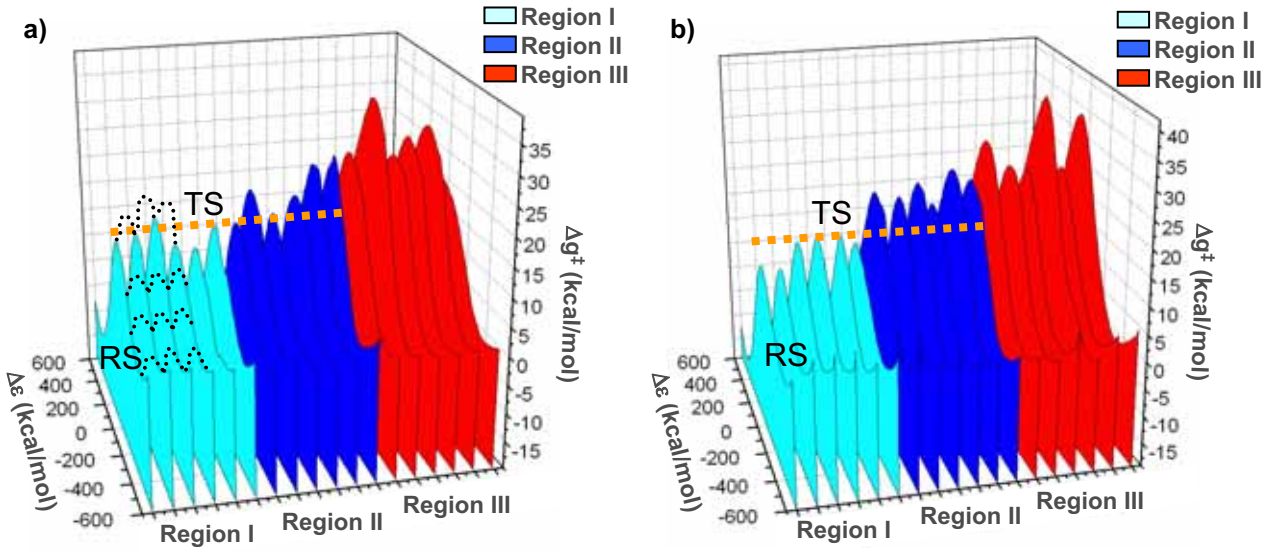
Chorismate Transition State Prephenate




mMjCM
EcCM
a)
b)
Rg (Å)
% CO
15.2
14.8
14.4
19
19.4
19.8
20.2
20.6
21
18
16
14
12
10
8
6
4
2
0
-2
-4
19.6
19.2
18.8
18
18.4
18.8
19.2
19.6
20
32
28
24
20
16
12
8
4
0
-4
15.4
15
14.6
19.2
19.6
20
20.4
20.8
21.2
28
24
20
16
12
8
4
0
-4
19.7
19.3
18.9
17.6
18
18.4
18.8
19.2
19.6
44
40
36
32
28
24
20
16
12
8
4
0
-4


a)
Region I
Region II
Region III
Probability
0
2
4
6
8
Region III
Region II
Region I
14 16 18 20 22 24 26 28 30 32 34 36 38 40
Δg‡ (kcal/mol)
b)
Region I
Region II
Region III
Probability
0
2
4
6
8
Region III
Region II
Region I
14 16 18 20 22 24 26 28 30 32 34 36 38 40
Δg‡ (kcal/mol)


a)
λ (kcal/mol)
780
740
700
660
620
580
10
15
20
25
30
35
40
b)
λ (kcal/mol)
780
740
700
660
620
580
10
15
20
25
30
35
40
Δg‡ (kcal/mol)


a)
Region I
Region II
Region III
TS
RS
Δε (kcal/mol)
600
400
200
0
-200
-400
-600
Region I
Region II
Region III
Δg‡ (kcal/mol)
35
30
25
20
15
10
5
0
-5
-10
-15
b)
Region I
Region II
Region III
TS
RS
Δε (kcal/mol)
600
400
200
0
-200
-400
-600
Region I
Region II
Region III
Δg‡ (kcal/mol)
40
35
30
25
20
15
10
5
0
-5
-10
-15