Evolution of Rate-Promoting Oscillations in a Model Enzyme

G. S. Blair Williams, Aftab M. Hossain, David E. Kranbuehl, and Carey K. Bagdassarian*

Department of Chemistry, The College of William and Mary, PO Box 8795, Williamsburg, Virginia 23187-8795

Received: May 30, 2003

This work explores dynamical couplings between global enzymatic fluctuation patterns and the catalytic fitness of a model enzyme system. A genetic algorithm drives the evolution of a population of enzyme molecules toward greater catalytic fitness by modifying via recombination and mutation events the interaction strengths between enzymatic subunits. We've identified, through a combination of molecular dynamics simulation and an approximate normal-mode analysis, two features that are tuned for increased catalytic efficacy throughout the evolutionary process. First, the average distance between substrate and the enzymatic subunit responsible for the chemistry is optimized. Second, the fraction of rate-promoting oscillations at the reaction coordinate is maximized. In other words, enzyme evolution in our model system proceeds by tuning geometric constraints and effectively eliminating catalytically nonproductive fluctuations. This modulation at the reaction coordinate results from global fluctuation patterns that are established by the spatial mix of loose and stiff enzymatic domains. The optimized geometry and appearance of rate-promoting oscillations at the reaction coordinate are emergent properties of the fluctuating enzyme system: they are not built into the model or selected for by the genetic algorithm. As is the case for actual enzyme molecules, catalytic fitness in our model system is sensitive to single-point mutations that affect global collective motions.

Introduction

Since Careri posed the idea of a fluctuating enzyme molecule nearly 30 years ago,[1] a great deal of research has been devoted to understanding the nature of protein conformational motions. The existence and importance of glass transitions, conformational substates, associated energy landscapes, and of conformational fluctuations in modulating both ligand binding and potential energy barriers to chemical reaction have been central to protein science.[2−5] As early as 1983, Karplus and McCammon[6] had implicated protein motions in facilitating chemical events in the active site. Early experimental evidence, for example by Farnum et al.,[7] suggested that motions in a flexible loop region of dihydrofolate reductase affect hydride transfer. Some provocative recent results linking dynamics and catalysis, once substrate is sequestered in the active site, are as follows. Benkovic, Hammes-Schiffer, and co-workers have reported groups of amino acid residues exhibiting coupled dynamic motions in dihydrofolate reductase.[8,9] Intriguingly, even amino acid residues distant from the enzyme's active site were found to move in a concerted fashion to contribute to progress along a collective reaction coordinate. Site-directed mutagenesis on amino acid pairs has shown synergistic effects on the rate of hydride transfer: the decrease in catalytic rate for the double mutants is greater than the summed decreases caused by single mutations. Radkiewicz and Brooks have shown, for dihydrofolate reductase complexed with substrate, the presence of strongly coupled motions in groups of amino acids.[10] That these correlated motions disappear in the enzyme−product complex indicates a possible link between dynamics and catalysis. Indeed, catalytically deleterious mutations occur in precisely those regions of the enzyme exhibiting the correlations. The work of Kohen et al., showing that protein motions in a thermophilic alcohol dehydrogenase are important for hydrogen-tunneling events,[11] has motivated investigations into other enzyme systems such as the flavoprotein and quinoprotein amine dehydrogenases.[12] Schwartz and co-workers have identified protein promoting vibrations in horse liver alcohol dehydrogenase: enzymatic thermal fluctuations are symmetrically coupled to the reaction coordinate so that the acceptor−donor distance is modulated for facilitated hydride-tunneling events.[13−15] Such modulation is manifested as a change in tunneling-barrier geometry. Recent quantum dynamical calculations on liver alcohol dehydrogenase have shown that the free energy barrier to catalysis is affected by the enzyme's atomic-scale conformation.[16,17] Experimentally, EPR and time-resolved fluorescence are used to address links between conformational motions in proteins and function.[18] These dynamic catalytic strategies augment other recently proposed catalytic mechanisms such as near attack conformers,[19,20] whose ground-state substrate molecular geometries resemble that of the transition state for easy and rapid progress over the reaction barrier. Furthermore, speculations that enzyme molecules are evolutionarily architectured so that "stray" or nonproductive fluctuations are minimized are attractive.[21,22] Though the above examples mainly address connections between conformational motions and catalysis, the possibility that conformational fluctuations exhibit correlated behavior is under active investigation in other areas of protein science as well.[23−27]

Our group has presented results on the simulated evolution of a model enzyme system that achieves high degrees of catalytic fitness.[28] In the enzyme−substrate complex, a catalytic event is scored when two neighboring enzymatic residues surmount a potential energy barrier between them. These two residues, labeled "S" and "C" for short, represent respectively the substrate molecule and the amino acid involved in the chemistry. We showed that the number of chemical events, or the catalytic fitness, can be maximized by varying the number and spatial distribution of conformationally flexible and stiff

* To whom correspondence should be addressed. Phone: (757) 221-2556. Fax: (757) 221-2715. E-mail: ckbagd@wm.edu.

10.1021/jp0306794 CCC: $25.00
Published on Web 10/09/2003

domains within the body of the enzyme until the "right" mix is achieved. In other words, catalytic efficacy is intimately related to the details of the enzyme molecule's fluctuation patterns.

A genetic algorithm driver was designed to evolve tens of thousands of enzyme molecules, each with a different distribution of "loose" and "stiff" domains, which were analyzed by molecular dynamics simulation for catalytic fitness. Because of the huge demands on computational power, we work with a simplified model enzyme−substrate complex which encapsulates the richness of the "fluctuating enzyme."

Each enzyme with its particular distribution of domains is an equilibrium entity: Conceptually, our evolutionary process creates structurally and energetically relaxed enzyme molecules whose equilibrium fluctuations are tuned for increasing chemical reaction rates. The dynamical focus of the work is on tuned motions for the enzymatic residues leading to catalytic efficacy.

In this present work, the evolutionary scheme is coupled to an approximate normal-mode analysis to explore the means by which catalytic fitness is increased. Two crucial features are identified. First, geometrical constraints evolve ensuring that the average distance between the two reacting residues is optimized. Second, enzymatic conformational fluctuations are tuned throughout the evolutionary process to decrease catalytically nonproductive fluctuations of the two residues relative to each other. That is, those low-frequency motions in which both the C and S residues move in the same direction in a correlated fashion are minimized, thereby increasing the probability of surmounting the chemical reaction barrier. What emerges for the catalytically fittest enzyme molecules are anti-correlated and, therefore, reaction-promoting oscillations for the C−S pair.

Importantly, these rate-promoting vibrations are truly emergent features of the evolutionary scheme: we do not build them into the model. Furthermore, such local rate-enhancing enzymatic motions are globally organized by the distribution and number of all stiff and loose domains throughout the body of the enzyme.

Enzyme Model, Reaction-Promoting Oscillations, and Methods

A full description of the model and assessment of its performance can be found by Williams et al. in ref 28. Here, we begin with a self-contained summary of the simplified model enzyme system.

Features of the Enzyme Model. As stated in the Introduction, the evolutionary algorithm creates tens of thousands of different enzyme molecules, each of which is analyzed through molecular dynamics (MD) simulation for catalytic fitness. Consequently, a simple model, nonetheless possessing sufficient complexity to exhibit realistic and interesting fluctuation behavior, is required for computational feasibility. An enzyme molecule consists of 450 residues. These residues, with the exception of the C−S pair whose interaction defines the chemical event, are coupled with each other through harmonic forces. This harmonic interaction between nearest-neighbor residues serves as a conceptual coarse-graining of all atomic-scale interactions (electrostatic, steric, etc.) occurring between amino acids in a real protein. In a previous publication,[22] we showed that several physical descriptors can be used to predict the average crystallographic temperature factor of a given amino acid residue in the body of cytidine deaminase. This work quantified and modeled the long appreciated fact that amino acid residues enjoy different degrees of conformational freedom in different regions of the enzyme molecule.[29] In the present model, this feature is the only input describing residue−residue

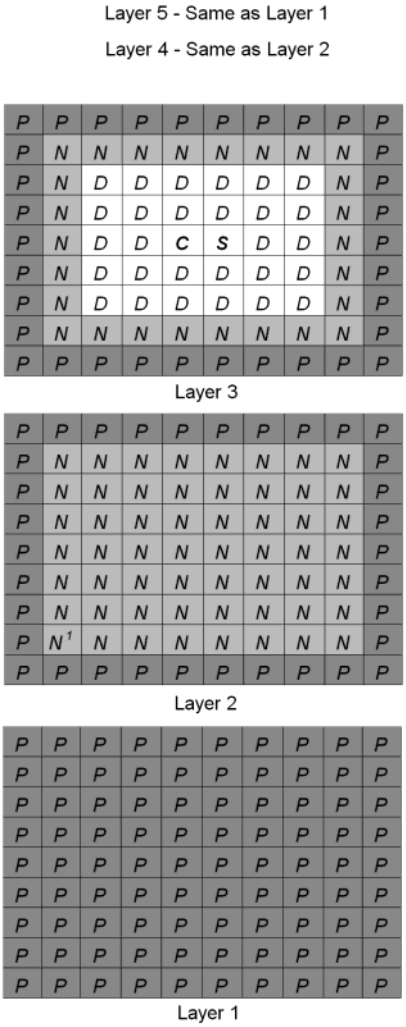

Figure 1. Model enzyme system with 450 subunits. The layers are stacked and each subunit interacts with its six nearest neighbors. Subunits labeled D, C, and S are directly acted upon by the genetic algorithm. C and S must surmount a reaction barrier to score a chemical event. See text for details.

interactions. The actual spatial distribution of force constants, defining regions of loose and stiff interactions in the enzyme body, is determined through genetic-algorithm-driven evolution.

Figure 1 shows a schematic of the lattice model from which the fluctuating system is constructed. It is easiest to imagine the enzyme to be constructed from five interacting layers of residues stacked upon each other. Each amino acid residue interacts harmonically with its six nearest neighbors. The first and fifth layers are constructed entirely of "phantom" or P residues which are fixed in place to preserve the overall structure of the enzyme as it fluctuates. The second and fourth layers are comprised by an outer shell of phantom residues and N or "neutral" amino acids. Though the neutrals do fluctuate through the molecular dynamics algorithm, they are not, as will become clear shortly, directly altered by the genetic algorithm. The middle layer has, in addition to shells of P and N residues, an inner core of 30 dynamic or D residues. Two of these define the chemistry: C, the catalytic residue, must surmount a chemical reaction barrier to interact with substrate S. The genetic algorithm acts upon the D subunits.

Though the subunits here represent, however abstractly, amino acid residues, the physical parameters of the problem are chosen so that MD simulations of the model enzymes are no different than simulations of any molecular system. Therefore, the mass

of each subunit is that of a carbon atom and the equilibrium length of the connecting springs is that of a C−C single bond.

Rules for the harmonic interactions between subunits (except between C and S) are as follows. N and P interactions are characterized by a spring constant with a value (in units of J/m^2) of 2. This relatively small value allows for the possibility of a good deal of thermal play in the relative motions between N and P subunits (mimicking surface amino acid residue motions). Nearest-neighbor N−N interactions are characterized by a spring constant of 25.

The D subunits can be either of type 0 or type 1. The genetic algorithm modifies, through recombination and mutation events as defined below, the distribution of these two types of subunits in the D layer. Two nearest-neighbor 0 subunits interact with a spring constant with value 50. Interaction between a 0 and 1 subunit is through a spring constant of value 25. The most loose interaction is between nearest-neighbor 1 subunits, with a constant of 2. Finally, type 0−N interactions have a spring constant of 50, and a 1−N interaction is through a value of 2.

The distribution and number of spring constants with values of 2, 25, and 50 allow for the body of the model enzyme to feature regions differing in thermal conformational freedom. The genetic algorithm drives the evolution of the system through the space of all possible combinations of loose, intermediate, and stiff enzymatic domains in search of the catalytically fittest enzymes.

An enzyme molecule's catalytic fitness is defined by the number of chemical events occurring during the time of a molecular dynamics run. The catalytic subunit C and the substrate S are not connected by a spring. Instead, their potential energy of interaction is set to zero until their spatial separation decreases to the equilibrium bond length of a carbon−carbon bond. The potential energy then increases linearly until the C−S separation decreases to 0.7 times the length of the C−C bond. For smaller subunit separations, the potential energy is constant. In this way, we model the fact that the chemical reaction coordinate must surmount a transition state barrier. Subsequent potential energy relaxation to the product state is not included because we are not focusing on product release: we are concerned only with collective enzyme motions facilitating access to transition state. The force between the subunits over the separation range 1.0−0.7 times the bond length is fixed at 2×10^{-10} J/A. During MD simulation of a given enzyme, with a given number and distribution of 0 and 1 subunits, a chemical event is scored whenever the potential energy ramp is surmounted. Once the chemical "hit" is scored, the ramp is turned off and the subunits fluctuate away from each other through the equations of motion: the enzyme is reset for another chemical event. The anharmonic "on−off" nature of the C−S interaction precludes study of the system's full dynamics through normal-mode analysis.[30] Indeed, the average equilibrium position of the C−S pair cannot be known a priori because it depends on the repulsive barrier and on the distribution of force constants. Without MD simulation, the effects of the repulsive force on the average C−S separation and on the number of catalytic events, which results in an important emergent feature of the model, would remain inaccessible.

MD Simulation, Genetic Algorithm, and Approximate Normal-Mode Analysis. Evolution and analysis of enzymatic catalytic fitness involves (a) generating an increasing population of enzyme molecules through a genetic algorithm (GA) and assessing the catalytic fitness of each individual enzyme via MD simulation and (b) using an approximate scheme to explore the vibrational normal modes available to each individual as a function of catalytic fitness.

For any given enzyme molecule, that is for any number and distribution of 0 and 1 subunits in the D layer, the equations of motion are integrated through a velocity Verlet algorithm with a half-step calculation.[30] During the equilibration phase of the simulation, the temperature is brought to 298 K through velocity scaling. After the total energy of the enzyme system equilibrates, the MD run continues for data collection. The number of chemical events or the catalytic fitness, i.e., the number of times the reaction barrier is surmounted by the C−S pair, is recorded. We have shown[28] that data collection over 1.3×10^5 time steps (with each time step of duration 1×10^{-14} s) is sufficient to ensure that subunit fluctuations from average positions are Gaussian distributed, that root-mean-square fluctuation magnitudes are similar for analogous positions in the enzyme, and that, for a given enzyme, the number of chemical events is independent of initial conditions to a desired tolerance. The standard deviation in catalytic fitness due to different initial conditions is about 7%.

Several unique features, including what we call a site-specific catalytic coefficient and judicious choice of starting enzymes, have been built into our genetic algorithm (GA) to increase the efficiency of the search for catalytically fit enzymes. These are discussed in detail in Williams et al.;[28] here a general summary of the GA suffices for clarity.

An individual refers specifically to an enzyme molecule with a given distribution and number of type 0 and 1 subunits. From a small starting population of enzymes, the genetic algorithm selects 10 individuals at random. The two catalytically fittest enzymes in this subpopulation, as ascertained through MD simulation, are selected for mating. The two parents mate by switching, at a single site within the D subunits, one parent's subunit for the other's. The crossover point is selected randomly, whereas the probability of an actual crossover at that point is determined through the catalytic coefficient as previously described. In addition, the offspring individuals thus generated can undergo point mutations (a switch from a 0 to a 1 or vice versa) at randomly chosen sites with some small probability. The new offspring are scored for catalytic fitness and introduced into the now growing population of individuals. The next step of GA iteration begins with a new selection of 10 random individuals and identification of the two fittest individuals therein for mating. The GA repeats these steps in its search for increasingly fit catalysts. We call the evolutionary scheme as described the "forward" GA in that it evolves more fit enzyme molecules. Starting from the same initial population of individuals, a "reverse" GA is also run to select for less fit individuals.

This combination of forward and reverse GAs is very effective in evolving individuals with a large range of catalytic fitness. Understanding exactly what changes as catalytic fitness increases requires examination of enzymatic fluctuation patterns throughout the evolutionary process. We stress: *The GA selects only for more or less fit catalysts, the mechanism by which catalytic efficacy changes is an emergent property of the system*. The catalytic fitness of any individual enzyme, a direct result of its dynamics, is determined through MD simulation, whereas an approximate normal-mode analysis uncovers the emerging fluctuation modes responsible for increasingly effective catalysis.

The repulsive interaction between the catalytic subunit and substrate forces the time-averaged equilibrium distance between them to a value greater than the equilibrium distance of a carbon−carbon single bond. On average, then, the interaction force between these subunits is zero (the interaction ramp is on

only when the intersubunit separation is between 0.7 and 1.0 times the C−C bond length). In constructing the matrix of potential energy second derivatives at equilibrium positions, we set those entries involving C−S interactions to zero. This represents the approximate normal-mode analysis.

The coupling between the molecular dynamics simulation, the genetic algorithm driver, and the normal mode scheme works as follows. The genetic algorithm creates new individuals by changing the number and distribution of 0 and 1 subunits in the D layer of subunits. Each new individual is scored for catalytic fitness via MD simulation. Finally, the matrix for normal-mode analysis for each individual is constructed from its distribution of spring constants as established by the GA. The eigenvalues and eigenvectors, whose values and components depend intimately on the number and distribution of stiff, intermediate, and loose interactions, are determined through usual diagonalization of the matrix associated with each individual.

Reaction-Promoting Enzymatic Oscillations. We are interested in dynamical couplings between global enzymatic fluctuation patterns and the reaction coordinate. These global fluctuations, if tuned for effective catalysis, must facilitate relative motions between the catalytic subunit C and substrate S so that the potential energy barrier between them is surmounted with high probability.

The barrier between C and S replaces a harmonic interaction between these two subunits which would have corresponded to a connecting spring in the (on average) *x* direction, as is apparent from Figure 1. Therefore, to understand the relative motions between C and S, it suffices to isolate those components of each eigenvector resulting from the normal-mode analysis that correspond to motions of these two residues in the *x* direction.

We identify rate-promoting motions in the C−S pair as (a) those in which the subunits oscillate in an anti-correlated fashion so that they alternate between moving toward each other and away from each other and (b) those where one of the subunits is stationary while the other oscillates relative to it in the *x* direction. Correlated motions, where both subunits always move in the same *x* direction, are considered detrimental to catalysis. Each eigenvector, or approximate normal mode, is analyzed to determine if it contributes to or hinders catalysis. A given eigenvector is tallied as rate promoting if (a) the *x* components of the C−S pair are of opposite sign and larger than a threshold value of 0.00001 or (b) either the C or S subunit is stationary while the other's component is above threshold. The threshold ensures that our focus is on possibly significant motions in a given mode. The results are robust to a fair range of thresholds. Conversely, an eigenvector is scored as unproductive to catalysis if both subunits' *x* components are above threshold and of the same sign. For each enzymatic individual, we calculate, along with its catalytic fitness, the fraction of its eigenvectors that are rate-promoting.

Results and Discussion

As summarized in Figure 2 from data first presented in Williams et al., the forward GA produces a population of increasingly fit catalysts, whereas the reverse GA produces a population of less fit individuals. In the region in which the forward GA operates, the fitness of the fittest individual in the population and the average population fitness both increase. Similarly, the average fitness of the population created by the reverse GA always decreases, as does the fitness of the least fit individual. Between the forward and reverse GAs, approximately 24 000 different enzyme molecules are generated, each with a different distribution and number of 0 and 1 subunits. The least fit individual found by the reverse GA has 16 catalytic events during MD simulation. The fittest individual boasts 253 catalytic events. We believe, for reasons elaborated upon below, that these two individuals are close to the best and worst catalysts possible for the model enzyme system.

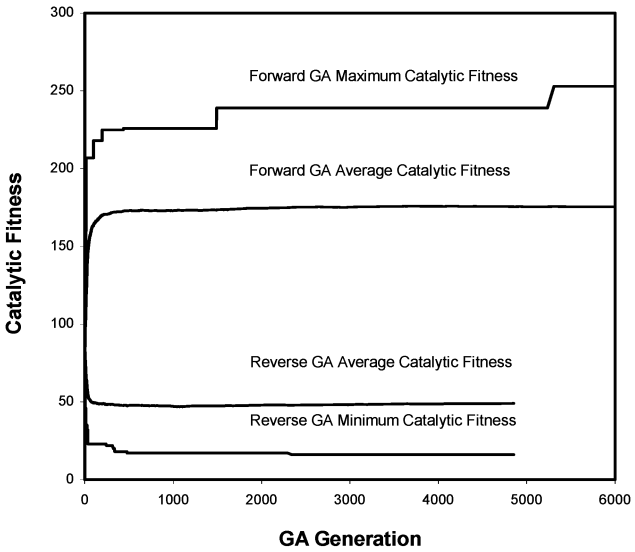

Figure 2. Forward GA increases both average fitness of growing enzyme population and fitness of the overall best enzyme at each generation. Reverse GA decreases average population fitness and fitness of poorest enzyme.

The forward and reverse GAs begin their respective searches for more and less fit enzymes from the same initial population of enzymes. As is implied in Figure 2, the newly created individuals quickly change the average features of the evolving population away from those of the initial starting population. The average fitness of the starting set is approximately 82 chemical events, and within relatively few generations, the forward GA significantly increases the population fitness, whereas the reverse GA decreases it. Because the two GAs cause a rapid divergence from the initial set of enzymes, the fraction of enzyme molecules in its vicinity is very small in the final overall population. The results reported below often involve averages over enzyme individuals with similar characteristics (such as the same number of catalytic events). Overall standard deviations are very good, except in the sparsely populated region around the initial population. For clarity in the results and to stress that this sparse region cannot compare in statistical quality with other more densely populated regions away from the initial set, we omit data points there.

Figure 3 shows the average C−S distance as a function of catalytic fitness or the number of times the catalytic subunit and substrate surmount the potential energy barrier between them. The reported distances are dimensionless as they are scaled by the equilibrium C−C bond length. That the C−S distances are on average longer than a carbon−carbon bond is not surprising: the interaction between the two subunits is repulsive. Two averages are actually involved in the reported C−S distances. First, for each individual in the evolved population, the C−S distance is time-averaged over the duration of the MD simulation. The second averaging deals with the inherent degeneracy of the problem: potentially many different enzyme molecules can have the same catalytic fitness. A bit of care is required here, however. As stated above, if the MD simulation of a given individual is started from different initial conditions, an average 7% error in catalytic fitness arises. In

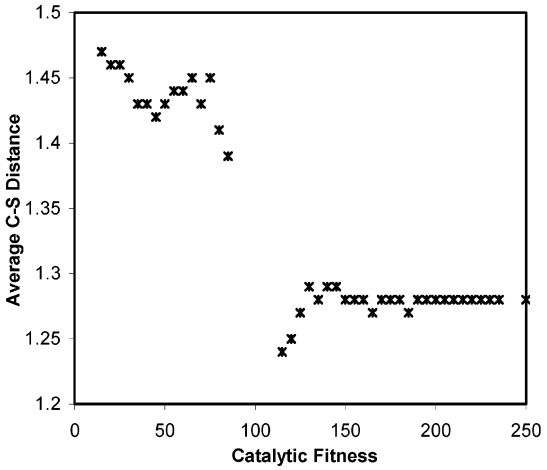

Figure 3. C−S distance as a function of catalytic fitness. The C−S distances reported are doubly averaged as explained in the text.

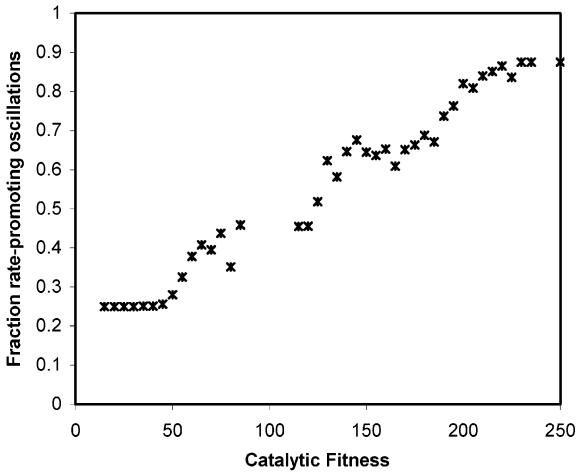

Figure 4. Fraction of eigenvectors contributing to catalytically productive oscillations at the reaction coordinate as a function of catalytic fitness.

this sense, individual enzyme molecules with very similar numbers of catalytic events cannot be distinguished. Figure 3 is created by "binning" together those enzyme molecules generated by the GAs that have fitness values within 5 chemical events. Therefore, the first bin contains all enzymes with 15−19 chemical events, the second is filled with molecules with 20−24 hits, etc. Within each bin, representing a very similar degree of catalytic fitness, the time-averaged C−S distances for the individual enzymes are averaged over the bin population and reported on the *y* axis. Note that the bins labeled with 85−110 catalytic events do not contain any data points; this represents the very sparsely populated region from which the GAs rapidly diverge. The reverse GA created the individuals populating the left-hand-side of the graph, whereas the forward GA generated the increasingly fit individuals on the right. The standard deviation arising from the population averages at each bin are fairly good in both legs with a value of ± 0.03.

As catalytic fitness increases in the regime of the reverse GA, the average C−S distance initially decreases. After initially producing individuals with increasing intersubunit separations (with catalytic fitness from 115 to 140), the forward GA creates individuals with a near constant dimensionless C−S distance. It is clear that increasing catalytic fitness requires that, overall, the two subunits come closer from a starting average separation distance of more than 1.45 to a final distance of 1.28. This final C−S distance, which is not the minimum allowed by the system, represents the optimal geometry for increased catalysis. However, how the forward GA, generating individuals with near constant C−S distance, finds increasingly fit catalysts remains incomprehensible from Figure 3.

Clearly, Figure 3 indicates phase-transition-like behavior in the C−S distance with increasing catalytic fitness. The average number of 0 or "stiff" subunits in the initial population is 15. The forward GA, in its search for more fit enzymes, creates individuals with greater numbers of 0 subunits. These enzymes on average are less flexible than the starting population, and the C−S distance is smaller because the repulsive force cannot push these two subunits as far apart. Indeed, most enzymes with more than 17 0 subunits have very similar C−S distances on average. Conversely, the reverse GA tends to decrease the number of 0 subunits for overall looser enzymes with greater C−S distances. The phase-transition behavior in Figure 3 is explained by these observations.

The coupling between emergent rate-promoting enzymatic fluctuations at the reaction coordinate and increasing catalytic fitness is clearly demonstrated via the approximate normal-mode analysis. Recall that the best enzyme has a catalytic fitness of 253, whereas the least fit displays only 16 chemical events. A study of eigenvalue distributions and their frequency of occurrence cannot distinguish these two enzymes.

For each individual enzyme, we calculate the fraction of rate-promoting modes; that is, we count the number of eigenvectors contributing to effective catalysis relative to the total number of eigenvectors (above threshold, as defined above). This eigenvector analysis reveals a greater fraction of emergent rate-enhancing oscillations in the C−S pair as the enzyme molecules become catalytically more fit. This pattern of increasing rate-enhancing motions with increasing enzymatic fitness is uncovered within those eigenvectors corresponding to lower frequency eigenvalues, that is, for those eigenvalues with magnitudes in the lower half of a given enzyme's frequency spectrum. Eigenvectors corresponding to higher frequencies are not tuned by the evolutionary algorithm for an increased fraction of possible rate-promoting oscillations. Once again, we bin together all individuals with catalytic fitness within 5 chemical events. Within each bin, the average fraction of rate-promoting eigenvectors is calculated over all individuals. Results are shown in Figure 4, and the standard deviations at all points are roughly ± 0.09. Note that until the enzymatic catalytic fitness reaches a value of approximately 50, the fraction of catalytically productive motions in the C−S pair remains nearly constant. A clear increase in the fraction of rate-enhancing motions, even through the sparsely populated region from which the two GAs diverge, is afterward correlated with increasing enzymatic fitness.

It is extremely informative to superimpose Figures 3 and 4 in an attempt to understand fully the evolution of catalytic fitness in our model enzyme system. For the least fit enzymes showing nearly constant fractions of rate-promoting oscillations, Figure 5 makes clear that any increases in enzyme performance are first due to strong decreases in the average distance between the catalytic subunit and the substrate. After this starting regime, however, the system increases catalytic fitness by both an emergent increase in the fraction of rate-enhancing motions and an overall decrease in C−S distance. Most dramatically, after the average C−S distance is optimized and stays nearly constant, the only way to increase catalytic fitness is by tuning enzymatic fluctuations for a continually increasing fraction of rate-promoting oscillations at the reaction coordinate.

Evolution, therefore, proceeds via a two-front attack: (a) optimization of the C−S distance so that the two subunits are

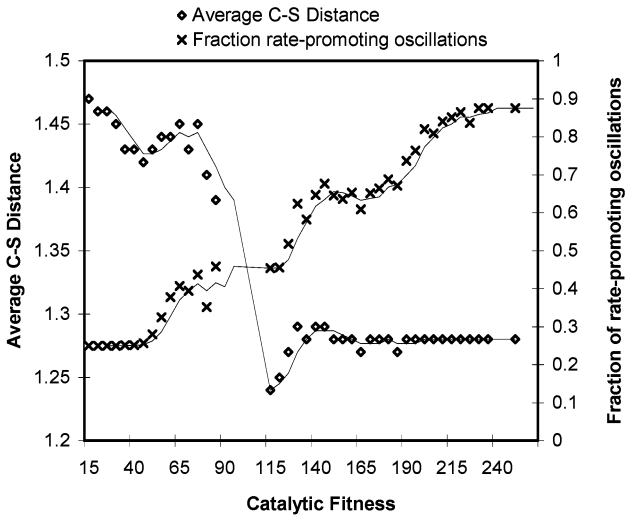

Figure 5. Fraction of eigenvectors contributing to catalytically productive oscillations and average C−S distance vs catalytic fitness.

geometrically well positioned for surmounting the reaction barrier and (b) organization of global fluctuation patterns, as established by the number and distribution of stiff and loose domains, allowing for emergent rate-promoting oscillations at the C−S loci. These two effects arise from tuning the enzymatic equilibrium fluctuations. The optimal and constant average C−S distance highlights the importance of the rate-promoting motions in surmounting the reaction barrier. These motions provide a mechanism by which the C−S pair frequently accesses the shorter intersubunit separations required for a chemical event. Rate-promoting oscillations are a signature of the dynamic motions selected for by evolution to give catalytically fit enzyme molecules.

These emergent fluctuation patterns that we've identified are very reminiscent of the promoting protein vibrations found by Schwartz and co-workers for horse liver alcohol dehydrogenase.[13−15] In their work, however, the vibrations modulate the barrier to a quantum hydrogen-tunneling event between donor and acceptor sites.

The interplay between C−S distance and the fraction of rate-enhancing vibrations is highlighted in Table 1 where four classes or types of enzyme molecules are revealed. Type 1 enzymes have a high fraction of their eigenvectors contributing to catalysis and a near optimal C−S distance. This class is exemplified by the fittest enzyme created through the genetic algorithm with 253 catalytic events, a scaled C−S distance of 1.27, and 88% of its eigenvectors contributing to productive catalysis. At the other extreme, type 4 enzymes with small fractions of rate-promoting fluctuations and large C−S distances, lies the worst catalyst found (16 chemical events with a C−S distance of 1.45 and 25% rate-promoting motions). An example of a type 2 enzyme (with 112 hits) has an optimal C−S distance but a very low fraction of productive motions. Type 3 enzymes feature good percentages of properly tuned fluctuations but suffer from poor C−S distances.

The success of the GA scheme in finding very fit enzymes is apparent from the optimal C−S distance and very high fraction of rate-promoting eigenvectors for the enzyme with 253 hits. Indeed, the GA has found what is likely to be one of the best catalysts allowable within the framework of the model. Because the representative type 2 enzyme (Table 1) has a significantly smaller fraction of rate-promoting motions than does the worst enzyme found by the GA, it is conceivable that

TABLE 1: Coupled Relationship between Active-Site Geometry (C−S Distance), Fraction of Catalytically Productive Fluctuations, and Catalytic Fitness

type 1 (best) high fraction rate-promoting oscillations:: good C−S distance	
catalytic fitness	253
percent rate-promoting oscillations	88%
average C−S distance	1.27
type 2 low fraction rate-promoting oscillations :: good C−S distance	
catalytic fitness	112
percent rate-promoting oscillations	8%
average C−S distance	1.25
type 3 high fraction rate-promoting oscillations:: poor C−S distance	
catalytic fitness	92
percent rate-promoting oscillations	92%
average C−S distance	1.45
type 4 (worst) low fraction rate-promoting oscillations :: poor C−S distance	
catalytic fitness	16
percent rate-promoting oscillations	25%
average C−S distance	1.45

the evolutionary scheme would eventually (if run even longer) identify enzymes with fewer than 16 hits.

We've previously reported that catalytic fitness, the number of catalytic events scored during the MD simulations, is not a monotonic function of the number of 0 subunits.[28] The "all stiff" enzyme becomes on average a much better catalyst if it is "peppered" with loose domains. However, if too many loose domains are introduced, average catalytic fitness decreases to its lowest values before recovering somewhat for the "all loose" enzyme. Importantly, the relative degree of enzymatic conformational freedom by itself is not sufficient to generate highly fit catalysts: the spatial distribution of type 0 and 1 subunits is crucial. For example, simple rearrangement of the 19 type 0 subunits in the fittest enzyme can decrease catalytic fitness from 253 chemical events to only 136.

It was stated above that the pattern of increasing fractions of rate-promoting oscillations is best uncovered by focusing on low frequency motions in the evolving enzyme population. This does not imply that individual enzyme molecules that are on the whole "looser" contain greater fractions of rate-promoting motions. Indeed, the "all loose" enzyme, with overall lower vibrational frequencies, does not feature a proper mix of loose and stiff domains to organize the motions at the C−S pair for effective elimination of catalytically unproductive fluctuations. Once again, the number and distribution of loose and stiff domains are crucial for tuning enzymatic conformational fluctuations for effective catalysis; the effects of the tuning are observed in the lower frequency collective vibrational modes, whereas higher frequency motions are not tuned for increased rate-promoting oscillations.

The ability of the C−S pair to surmount the local potential energy barrier is globally determined by the entire enzyme's fluctuation profile. A number of single-point mutations to the fittest enzyme molecule have been show to significantly reduce its fitness, in some cases by nearly 80 catalytic events,[28] and these mutations can be far from the C−S pair. Radkiewicz and Brooks have found, through MD simulations of dihydrofolate deductase,[10] that the enzyme−substrate complex shows strongly correlated motions between groups of amino acids. These collective motions can be destroyed by single-point mutations

far from the active site. Indeed, in our best model enzyme, we have been able to identify single-subunit mutations that decrease the fraction of catalytically productive oscillations without affecting greatly the C−S distance. For example, modifying the N subunit to the left of the C−S pair (see Figure 1, layer 3) so that all interactions with its neighbors are governed through the largest spring constant decreases the percent of rate-promoting oscillations from 88% to 62%. The catalytic fitness drops from 253 chemical events to 157. All intersubunit interactions in the enzyme come into play in establishing rate-promoting oscillations at the reaction coordinate.

In conclusion, the number and distribution of loose, intermediate, and stiff domains in the body of an enzyme molecule can govern fluctuation patterns leading to maximum catalytic fitness. Two emergent local features arise from the global tuning of conformational fluctuations: optimization of the distance between the catalytic subunit and substrate and a near maximal fraction of rate-promoting oscillations between these two subunits. Emergent tuning of either property alone does not suffice to maximize catalytic fitness. Our model system is by necessity not very complex; it simplifies enzyme structure and residue interactions and does not consider the issue of maintaining foldability with changes to the intersubunit interactions. Yet, it evolves to capture several features of real enzyme molecules: the existence of rate-promoting oscillations, catalytic sensitivity to point mutations far from the active site, and disruption of collective motions by single-point mutations, for example those affecting the fraction of rate-promoting oscillations.

Acknowledgment. This work was supported by grants from the Jeffress Memorial Trust (J-432) and Research Corporation (CC4937). We thank Professors Vern Anderson, Vern Schramm, and Michael Trosset for fruitful discussions. We also thank two anonymous reviewers for their insights and suggestions.

References and Notes

(1) Careri, G. In *Quantum Statistical Mechanics in the Natural Sciences*; Korsunoglu, B., Mintz, S. L., Widmayer, S. M., Eds.; Plenum: New York, 1974; p 15.

(2) Fraunfelder, H.; Wolynes, P. G.; Austin, R. H. *Rev. Mod. Phys.* **1999**, *71*, S419.

(3) Fraunfelder, H.; Sligar, S. G.; Wolynes, P. G. *Science* **1991**, *254*, 1598.

(4) Parak, F. G. *Rep. Prog. Phys.* **2003**, *66*, 103.

(5) Gammaitoni, L.; Hänggi, P.; Jung, P.; Marchesoni, F. *Rev. Mod. Phys.* **1998**, *70*, 223.

(6) Karplus, M.; McCammon, J. *Annu. Rev. Biochem.* **1983**, *52*, 263.

(7) Farnum, M. F.; Magde, D.; Howell, E. E.; Hirai, J. T.; Warren, M. S.; Grimsley, J. K.; Kraut, J. *Biochemistry* **1991**, *30*, 11567.

(8) Rajagopalan, P. T. R.; Lutz, S.; Benkovic, S. J. *Biochemistry* **2002**, *41*, 12618.

(9) Agarwal, P. K.; Billeter, S. R.; Rajagopalan, P. T. R.; Benkovic, S. J.; Hammes-Schiffer, S. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 2794.

(10) Radkiewicz, J. L.; Brooks, C. L., III. *J. Am. Chem. Soc.* **2000**, *122*, 225.

(11) Kohen, A.; Cannio, R.; Bartolucci, S.; Klinman, J. P. *Nature* **1999**, *399*, 496.

(12) Sutcliffe, M. J.; Scrutton, N. S. *Eur. J. Biochem.* **2002**, *269*, 3096.

(13) Antoniou, D.; Schwartz, S. D. *J. Phys. Chem. B* **2001**, *105*, 5553.

(14) Caratzoulas, S.; Mincer, J. S.; Schwartz, S. D. *J. Am. Chem. Soc.* **2002**, *124*, 3270.

(15) Mincer, J. S.; Schwartz, S. D. *J. Phys. Chem. B* **2003**, *107*, 366.

(16) Billeter, S. R.; Webb, S. P.; Agarwal, P. K.; Iordanov, T.; Hammes-Schiffer, S. *J. Am. Chem. Soc.* **2001**, *123*, 11262.

(17) Alhambra, C.; Corchado, J.; Sánchez, M. L.; Garcia-Viloca, M.; Gao, J.; Truhlar, D. G. *J. Phys. Chem. B* **2001**, *105*, 11326.

(18) Lozinsky, E.; Febbraio, F.; Shames, A. I.; Likhtenshtein, G. I.; Bismuto, E.; Nucci, R. *Protein Sci.* **2002**, *11*, 2535.

(19) Bruice, T. C.; Benkovic, S. J. *Biochemistry* **2000**, *39*, 6267.

(20) Har, S.; Bruice, T. C. *J. Am. Chem. Soc.* **2003**, *125*, 1472.

(21) Alper, K. O.; Singla, M.; Stone, J. L.; Bagdassarian, C. K. *Protein Sci.* **2001**, *10*, 1319.

(22) Noonan, R. C.; Carter, C. W., Jr.; Bagdassarian, C. K. *Protein Sci.* **2002**, *11*, 1424.

(23) Bahar, I.; Jernignan, R. L. *Biochemistry* **1999**, *38*, 3478.

(24) Ota, N.; Agard, D. A. *Protein Sci.* **2001**, *10*, 1403.

(25) Roche, P.; Mouawad, L.; Perahia, D.; Samama, J. P.; Kahn, D. *Protein Sci.* **2002**, *11*, 2622.

(26) Peters, G. H.; Frimurer, T. M.; Andersen, J. N.; Olsen, O. H. *Biophys. J.* **2000**, *78*, 2191.

(27) Idiyatullin, D.; Krushelnitsky, A.; Nesmelova, I.; Blanco, F.; Daragan, V. A.; Serrano, L.; Mayo, K. H. *Protein Sci.* **2000**, *9*, 2118.

(28) Williams, G. S. B.; Hossain, A. M.; Shang, S.; Kranbuehl, D. E.; Bagdassarian, C. K. *J. Theor. Comput. Chem.* in press.

(29) Karplus, M.; Petsko, G. A. *Nature* **1990**, *347*, 631.

(30) Leach, A. R. *Molecular Modelling: Principles and Applications*; Prentice Hall: New York, 2001.