Proc. Natl. Acad. Sci. USA
Vol. 83, pp. 3806–3810, June 1986
Biophysics

Toward computer-aided site-directed mutagenesis of enzymes

(enzyme catalysis/electrostatic energies/solvation energies in proteins/free energy perturbation methods)

ARIEH WARSHEL AND FREDY SUSSMAN

Department of Chemistry, University of Southern California, Los Angeles, CA 90089-0482

Communicated by Sidney W. Benson, January 21, 1986

ABSTRACT **A preliminary attempt to simulate the observed effect of a site-directed mutagenesis of rat trypsin gives encouraging results. The calculations reproduce in a semiquantitative way the observed change in the activation barrier of the rate-limiting step of amide hydrolysis. This result, which did not require any adjustable parameters, indicates that our method may provide a reliable basis for computer-aided enzyme design. In addition to the potentially practical value of the calculations, they provide important mechanistic information—that is, the change in the catalytic effect in trypsin appears to be almost exclusively due to the change in the electrostatic stabilization of the ionic configurations. This supports the view that electrostatic effects are the major factor in enzyme catalysis.**

The manipulation of active sites of enzymes by site-directed mutagenesis has become a reality in recent years. Several experiments have demonstrated the exciting potential of "engineering" modified enzymes (1–5). Apart from the obvious practical value of these approaches, they offer new breakthroughs in the fundamental understanding of enzyme activity. Changing systematically specific residues can identify those groups that are crucial for the action of an enzyme and determine how much they contribute to the overall catalysis. However, despite the great potential of studying enzymes by systematic replacement of particular amino acid residues, there are still many remaining problems. For example, in some cases the catalytic effect might be due to several environmental factors that cannot be traced directly to single groups. Moreover, since the options for sequence changes are so large, it is hard to locate the optimal mutations that will bring the desired improvement in catalysis. In this respect it is very important to have a detailed model that can tell us *how* the change of a given group is translated into changes in the activation energies of the given reaction.

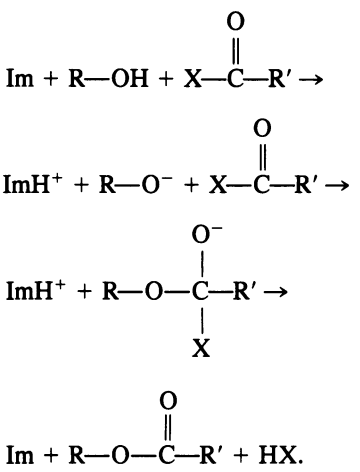
In this work we present some very encouraging results of a preliminary computer simulation of site-directed mutagenesis. As a test we take the site-directed mutagenesis performed recently on rat trypsin by Craik *et al.* (5). This work demonstrated that the replacement of Gly-216 and Gly-226 by alanine produces three orders of magnitude reduction in the catalytic rate constant, k_{cat}, for amide hydrolysis. The same mutation produces a rather small change in the binding constant, K_m (a factor of <20). To analyze these experimental findings, we applied our recently developed simulation methods that are based on a combination of the empirical valence bond (EVB) method and a free energy perturbation method. We found that the experimentally observed change in rate constant can be reproduced in a semiquantitative way by our simulation method. The present study indicates that the key to computer-aided enzyme design is not so much in calculating structural changes but in converting them into catalytic free energies.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Simulation Methods

Our approach for simulation of enzymatic reactions is described in detail elsewhere (6, 7). Here we only consider key points about the simulation of the reaction of serine proteases and specific aspects of the present study.

To simulate an enzymatic reaction one needs a potential surface that relates the energy of the reacting system to its structure. This is obtained here by the convenient and reliable EVB approach (6, 7). This approach describes the given reaction in terms of relevant resonance structures. In the present case we are interested in the acylation of amides by serine proteases (8, 9).



[1]

Here Im and R—OH designate the His-57 and Ser-195 groups of the enzyme. The most important resonance structures involved in this reaction are drawn in Fig. 1 and represented schematically in Eq. 2.

$$\begin{aligned}\psi_1 &= (\mathrm{Im}\quad \mathrm{H{-}O}\quad \mathrm{C{=}O})\\ \psi_2 &= (\mathrm{Im^+{-}H}\quad \mathrm{O^-}\quad \mathrm{C{=}O})\\ \psi_3 &= (\mathrm{Im^+{-}H}\quad \mathrm{O{-}C{-}O^-}).\end{aligned} \qquad [2]$$

Here Im, H—O, and C=O indicate His-57, Ser-195, and the substrate's, carbonyl. Other resonance structures that represent the possibility of "charge-relay" mechanism (where the proton is transferred to Asp-102) (8, 10) are not included here. These resonance structures were found to be of high energy in all of the relevant steps of the reaction (11). The reaction potential surface, E_G, is obtained by considering the mixing of the three resonance structures by solving the appropriate secular equation (6, 11).

The mixing terms are considered elsewhere (11) and our key point is that the protein substrate interaction is introduced through the energies ε_i of the different resonance structures. These energies can be considered as the "force

Abbreviation: EVB, empirical valence bond.

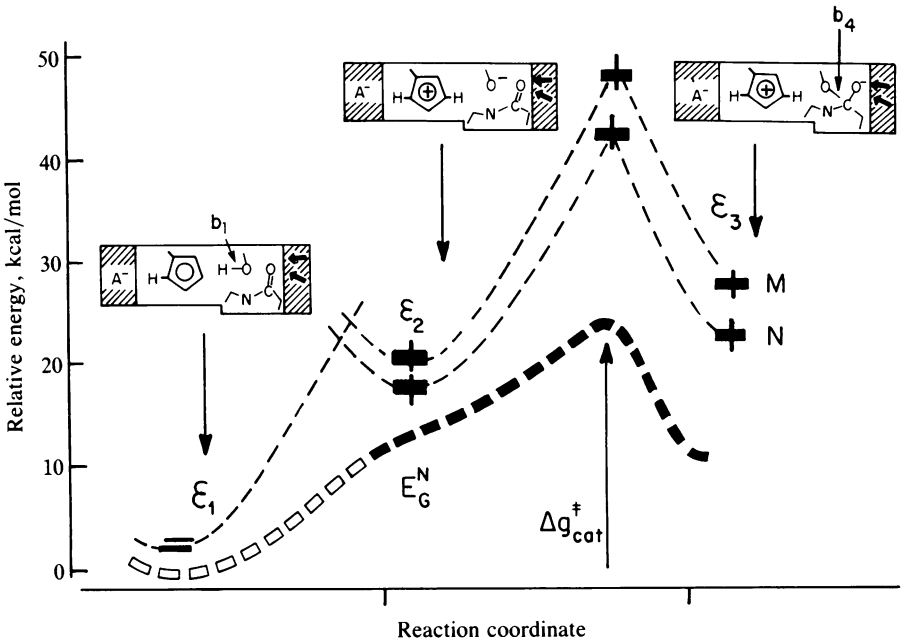


FIG. 1. Schematic representation of the energetics of the main resonance structures that determine the rate-limiting step in the catalytic reaction of trypsin. The bars represent the actual calculated free energies of the indicated resonance structures for the native (N) and the mutant (M). These free energies were obtained by an umbrella sampling study (see text). The barriers at the intersection of ε_2 and ε_3 correspond to a constraint potential with equal contributions from ε_2 and ε_3. (One calorie = 4.184 J.)

fields" of the corresponding resonance structures—that is, we use

$$\varepsilon_i = V_p + \sum_m \Delta M^{(i)}(b_m) + \sum_m K_\theta^{(i)}[\theta_m - \theta_{o,m}^{(i)}]^2 + V_{nb}^{(i)} + \alpha^{(i)} + V_{QQ}^{(i)} + \Delta V_{p,R}^{(i)}, \quad [3]$$

where V_p is the protein force field, represented by empirical functions (see, for example, ref. 21). This potential describes the interactions between those atoms of the substrate–protein–water system not included in the reaction region. $\Delta M^{(i)}(b_m)$ is the change in the value of the Morse potential of the indicated bond, b_m, in the ith resonance structure. The K_θ and V_{nb} terms describe the angle bending and nonbonded interaction terms, respectively, for the given resonance structure. The V_{QQ} term represents the charge–charge interaction between the fragments of the given resonance structure. The $\alpha^{(i)}$ values are the energies involved in the formation of the indicated resonance structure in solution at infinite separation between its fragments. These energies are obtained from experimental information about the reaction in solution (11). $\Delta V_{p,R}$ is the interaction between the reacting atoms of the ith resonance structure and the rest of the protein–water system. This interaction is taken relative to the free energy of the given resonance structure where the relevant fragments are at infinite separation in water.

Evaluating the EVB energies as a function of the protein geometry provides the reaction potential surface. It is important to note in this respect that the energy of each resonance structure is determined by the interaction between the protein residual charges (dipoles) and the full charges of the given resonance structure. However, the orientation (polarization) of the protein dipoles is determined by the actual ground state charge distribution of the atoms included in the reacting region. This charge distribution is obtained by mixing of the different resonance structures. For example, at the transition state of our reaction, the protein dipoles are polarized by the charge distribution obtained from the mixing of ψ_2 and ψ_3. The actual mixing depends in a consistent way on the polarization of the protein. This appears to provide a much more powerful approach than molecular orbital treatments. In addition, the EVB approach offers a very convenient parameterization by forcing the potential surface to reproduce observed pK_a values for the fragments at infinite separation in water (6, 11).

In evaluating the protein–substrate interaction, $\Delta V^{(i)}_{p,R}$, we consider in addition to the regular nonbonded interactions the electrostatic interactions between the protein and the charges of the reacting region. The parameters used for the protein residual charges and induced dipoles are given in ref. 13. In addition, we include in the present calculations a "shell" of a surface-constrained all-atom solvent model (14) that is built around the reacting region (the present simulation included the water molecules in a sphere of 12 Å around the oxyanion's oxygen).

One of the key bottlenecks in the simulation of enzymatic reactions is the convergence problem (7). Including all of the protein atoms in the simulation is *in principle* more rigorous but might lead to serious convergence problems. In the present case we are not interested in simulating the folding of the complete protein but the effect of limited structural changes. Thus, we follow a philosophy similar to the one used in our early study of hemoglobin (15) and improve the convergence by constraining parts of the protein to vibrate near the observed x-ray structure. The reliability of the results obtained in this way was examined by checking their sensitivity to the size of the constraint region. (The present results were obtained by constraining the protein atoms outside a sphere of 16 Å, centered around the carbonyl group of the substrate.)

With a convenient and reliable potential surface, one faces the challenge of evaluating the relevant activation free energies. This can be accomplished by Monte Carlo or molecular dynamics methods. Here we use a molecular dynamics method, noting, however, that the activation free energy is related to the probability of reaching the transition state and is not a dynamical property (16). To evaluate this probability we use the method of umbrella sampling (12, 16), propagating the trajectories over a constraint potential that drives them toward the transition state. Based on our

previous experience regarding the key role of electrostatic forces, we select a constraint potential of the form

$$E_m = \Theta_1^m \varepsilon_1 + \Theta_2^m \varepsilon_2 + \Theta_3^m \varepsilon_3, \quad [4]$$

where ε_1, ε_2, and ε_3 are the energies of each resonance structure as given by Eq. 3 and the Θ values are variables that vary between zero and 1 ($\Theta_2 + \Theta_2 + \Theta_3 = 1$).

In studies of reactions of small isolated molecules it is reasonable to constrain the length of specific bonds. This, however, provides a rather impractical way for obtaining the activation free energies for enzymatic reactions. Our approach, on the other hand, concentrates on a much more physical reaction coordinate by changing gradually the charge distribution of the reacting region. This gradual "charging" forces the protein to respond in a stable way driving the system toward the transition state. This process, which is quite similar to the familiar charging process in physics textbooks, forces the enzyme–water system to polarize its dipoles toward the developing charges. This effect, which might be the real contender for the name "induced fit" (17), appears to be quite significant in the trypsin system.

Calculated Structural Deformations

As pointed out in the Introduction, we have taken as a test case the experiment carried out by Craik *et al.* (5) on rat trypsin. Our simulation studies were performed on bovine trypsin, starting from the coordinates of ref. 18. As a substrate we use an Ala-Lys-Ala peptide. Bovine trypsin has a 74% homology with rat trypsin and, as argued in ref. 5, seems to provide a reasonable approximation for studying rat trypsin. The residues Gly-216 and Gly-266, which were modified by Craik *et al.*, are conserved in bovine and rat trypsin. The mutations, which are located at the substrate binding site, led to a very large change in the catalytic rate constant (around a factor of 2000) and only a relatively small change in the binding constant (a factor of <20).

The problem is, of course, to relate the steric effects of the mutations to the change in the catalytic activity. An inspection of the structure of the native enzyme shows that Gly-216 forms part of a I^{β}-type bend together with residues 217–219–220, and its carbonyl group is turned inward forming a hydrogen bond with the NH group of residue 220. Gly-226 is located in a strand that runs antiparallel to that of Gly-216 in the bottom of the oxyanion hole (see Fig. 3). Replacement

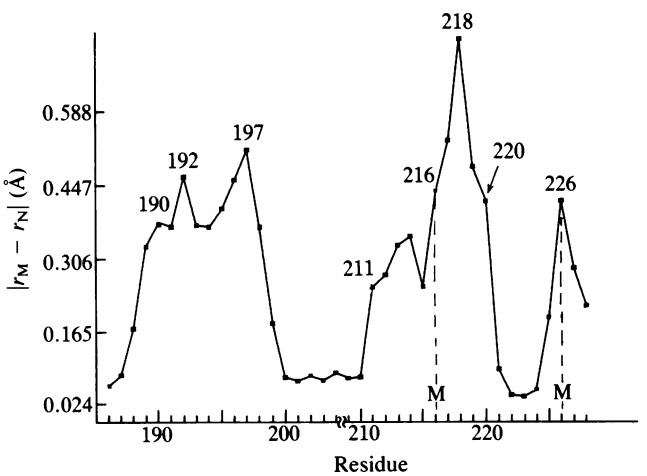


FIG. 2. Calculated average of the difference between the position vectors of the C_α atoms of the native and mutant structures. The calculations correspond to a constraint potential that represents the charge distribution close to that of the oxyanion configuration ($\Theta_1 = 0$, $\Theta_2 = 0.2$, $\Theta_3 = 0.8$).

of Gly-216 and Gly-226 by alanine produces close nonbonded contacts between the C_β atoms of Ala-216 and residue 220 and between the C_β of Ala-226 and residues 189 and 190. The calculated relaxation of these steric effects is described in Fig. 2, which depicts the average displacements of the C_α atoms of the mutant relative to the corresponding position in the native enzyme. The calculated geometrical changes appear to be small and rather complicated. Yet, some insight can be gained by focusing on the electrostatic elements of the system—that is, as seen from Figs. 3 and 4, the hydrogen bond network around the C—O^- of the oxyanion configuration (ψ_3) is somewhat different in the mutant and the native enzymes. This network, that includes water molecules, appears to form somewhat better hydrogen bonds with the O^- in the native enzyme than in the mutant enzyme. Thus, the active site of the mutant cannot provide the same electrostatic stabilization as that of the native enzyme. The calculations indicate that the native and mutant enzymes are both strongly polarized by the O^- of the oxyanion configuration (ψ_3). However, in the ψ_2 configuration (where the substrate is uncharged) the oxyanion hole is polarized to a larger extent in the native enzyme than in the mutant enzyme (see Fig. 4).

It is also instructive to note that residue 216 is close to Ser-214, which forms a hydrogen bond with Asp-102. Increasing the hydrogen bonding (solvation) of Asp-102 in the mutant destabilizes ψ_2 and ψ_3. Also, as seen from Fig. 4, the His^+ Ser^- ion pair has a closer interaction distance in the native than in the mutant enzyme. This might help in reducing the energy of ψ_2.

The above geometry changes are not sufficiently unique to reach a clear conclusion about their effect on the reactivity of the two systems. Such conclusions should be based on energy calculations (see below). Nevertheless, it is instructive to trace the calculated geometry changes to the close contacts generated by the mutations. Inspection of Fig. 2 shows significant displacements of the C_α atoms of residues 189–198, which include the oxyanion hole (see Fig. 3). Other large displacements occur at the 211–221 and 225–227 stretches, which include the sites of the mutations. The nature of these shifts becomes clearer upon inspection of Figs. 3 and 4. These figures indicate that the steric interaction between Ala-226 and Asp-189 exerts a force on the 193–187 stretch, which is apparently propagated to the oxyanion hole. This converts the strain energy associated with the mutation to a change in the catalytic electrostatic energy. A second effect is associated with residue 216. This residue appears to act by magnifying the effect of residue 226—that is, the steric force that Ala-226 exerts on Asp-189 can be reduced by a motion of Ala-226 away from Asp-189. This alternative relaxation mode is partially blocked when Gly-216 is replaced by the more bulky alanine residue. The above analysis is consistent with the fact that the Gly-216 → alanine mutation leads to a small change in k_{cat} (a factor of <5), the Gly-226 → alanine mutation leads to a somewhat larger change (≈100 and ≈10 for the arginyl and lysyl substrates, respectively), and the largest effect (≈2000) is obtained by the double mutations (216, 226).

The above calculation did not explore the possibility that these mutations might have caused major conformational changes. Such changes could only be studied by a much larger simulation time or by knowing (using structural information) what is the nature of the new conformation. This point is considered in the last section.

Estimating the Change of the Catalytic Energy

The geometry changes associated with the mutation can be evaluated by standard energy minimization or molecular dynamics approaches. What is much harder, and apparently more important, is the conversion of the geometrical changes to reliable activation energies. Such a task cannot be carried

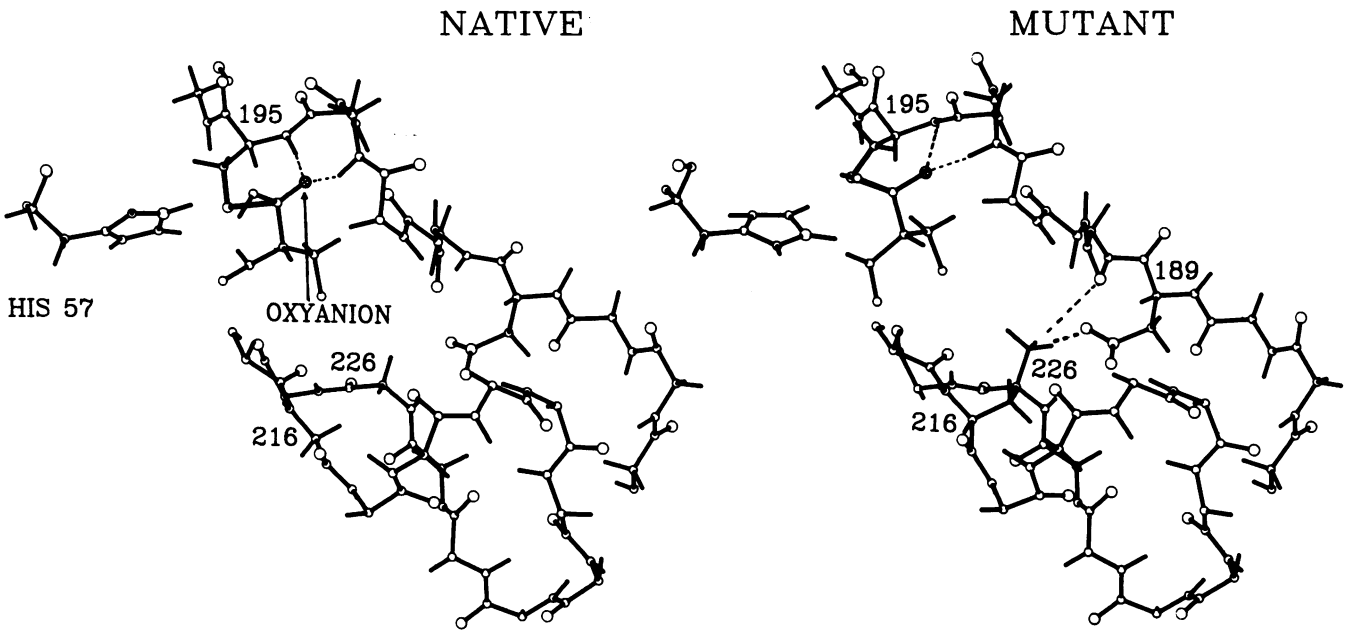


FIG. 3. Geometrical changes involved in the Gly-216/Gly-226 → alanine mutation. Low-energy snapshots obtained by using a constraint potential that corresponds to $\psi_2(\Theta_1 = 0, \Theta_2 = 1, \Theta_3 = 0)$ are compared. The interaction between Ala-226 and the extension of the oxyanion hole bend (residues 190–189) is emphasized. This interaction is involved in deforming the oxyanion hole.

out without modeling the *function* of the system. This can be accomplished by using the EVB and the umbrella sampling methods as outlined above. The present work used the umbrella sampling method only in a preliminary way—that is, previous studies (7, 16, 19) have indicated that the changes in the activation free energies ($\Delta g^{\ddagger}$) are correlated with the changes of the free energies of the relevant resonance structures. Thus, one can use the approximated relation

$$\Delta\Delta g^{\ddagger}_{i\to j} \leq [\Delta G^{(j)} - \Delta G^{(i)}] = \Delta\Delta G_{i\to j}, \qquad [5]$$

where $\Delta G^{(i)}$ is the free energy associated with the potential ε_i of the ith resonance structure. The $\Delta G^{(i)}$ can be evaluated by the relation (12, 16)

$$\exp\{-\Delta G_{m\to m'}/k_bT\} = \langle\exp\{-(E_{m'} - E_m)/k_bT\}\rangle_m, \qquad [6]$$

where k_b is the Boltzmann constant and E_m is the mapping potential of Eq. 4. Summing the $\Delta G_{m\to m'}$, for small changes of Θ_m, gives the free energy associated with changing one resonance structure to another. For example, gradually changing the $\boldsymbol{\Theta}$ vector $[\boldsymbol{\Theta} = (\Theta_1, \Theta_2, \Theta_3)]$ from (1, 0, 0) to (0, 1, 0) gives the free energy associated with changing the potential of the system from ε_1 to ε_2. The calculated free energies of each resonance structure, relative to that of $\Delta G^{(i)}$, are given in Fig. 1. Using these free energies and Eq. 5 we can estimate the change in rate constant by:

$$k^{M}_{cat}/k^{N}_{cat} = \exp\{-(\Delta g^{\ddagger}_{M} - \Delta g^{\ddagger}_{N})/k_bT\}$$
$$\simeq \exp\{-(\Delta G^{M}_{1\to 3} - \Delta G^{N}_{1\to 3})/k_bT\}, \qquad [7]$$

where N and M indicate the native and mutant systems, respectively. The present calculations gave a value of ≈6 kcal/mol for $\Delta G^{M}_{1\to 3} - \Delta G^{N}_{1\to 3}$. This corresponds to a reduction of about 5×10^{-5} in k_{cat} as compared to the observed change of about 5×10^{-4}. Considering the preliminary nature of the calculations and our estimate of the convergence error (≈2 kcal/mol), we find the current results encouraging.

It is important to note that the calculated change in ΔG is

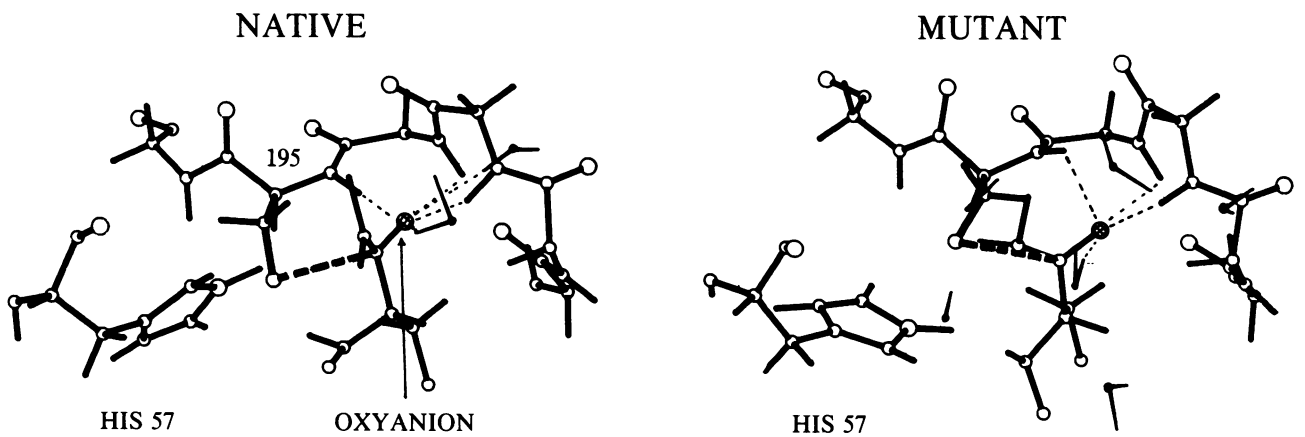


FIG. 4. Comparison of the geometrical changes in the oxyanion hole region for the native and mutant systems. The same snapshots used in Fig. 3 are considered. The hydrogen bond network that stabilizes the oxyanion is indicated by ---. This network includes main chain N—H dipoles and water molecules. The calculated geometrical changes suggest that the electrostatic stabilization of the oxyanion configuration is reduced in the mutant system.

reproduced to a good approximation by considering only the electrostatic energy contributions to the relevant ε_i. This finding, which requires further studies, indicates that the major contribution to the catalytic energy in trypsin is associated with electrostatic effects.

Concluding Remarks

This work presents a posteriori analysis of the effect of site-directed mutagenesis of rat trypsin. The calculations point toward changes in the oxyanion hole as a key effect of the mutation. This finding supports early suggestions (7) that electrostatic effects are the key factor in catalytic reactions. It is interesting to mention in this respect a recent work of Wells and Fersht (2), which identified hydrogen bond stabilization as a key factor in tyrosyl-tRNA-synthetase. This stabilization by hydrogen bonds is one of the key elements of what we call "electrostatic stabilization of transition states"—that is, as we have stressed repeatedly (7), electrostatic stabilization is not just the interaction between opposite charges but the interaction between isolated charges and their polar environment. Viewing the enzyme as the "solvent" for the transition states of its reaction appears to be a powerful general way for analyzing enzymatic reactions.

The present approach evaluates activation free energies rather than activation enthalpies. Thus, it is quite possible that the calculated activation barriers include some entropic contributions, which may constitute a significant part of the catalytic electrostatic energy. Such entropic effects (which are probably associated with the polarization of the dipoles of the oxyanion hole) are an integral part of any electrostatic free energy. In fact, the term "electrostatic energy" is referred to as the reversible work (free energy) needed to polarize the solvent around a given charge.

The nature of the key catalytic factors in enzymatic reactions has been a source of discussion and controversy in the literature. In examining our point of view that electrostatic effects are the key catalytic factor (7), we have evaluated the electrostatic contributions to the *change* in the activation free energy. It was found here that a major part of the calculated change in activation free energy is accounted for by the corresponding electrostatic contributions.

The difference between the catalytic power of the mutant and the native enzyme is related in some respects to the difference in activity of chymotrypsinogen and chymotrypsin—that is, chymotrypsinogen is activated by the cutting of the single bond between amino acids 15 and 16 (see ref. 10). The new amino-terminus at Ile-16 then forms an ionic bond (a "salt-bridge") with Asp-194, altering the orientation of Asp-194. The observed shift of Asp-194 from chymotrypsin to chymotrypsinogen (20) is associated with a large shift of the main chain dipoles of residues 193, 194, and 195. This partial destruction of the oxyanion hole makes it harder to stabilize the O^- of the oxyanion configuration. Apparently, the reduction of the stabilization of the oxyanion configuration (and the corresponding conformational changes) are larger in chymotrypsinogen than in the Gly-216/Gly-226 → alanine mutation and the change in k_{cat} is larger by three orders of magnitude.

It is important to note that the conclusions drawn from the present treatment should be verified by structural studies—that is, if the actual effect of a mutation involves very large structural changes [as large as those observed in the recent comparison of chymotrypsin and chymotrypsinogen (20)], then the simulation must be extended to explore the new conformation. A brute force simulation would require a very long time to reach configurations that are very different than the native configurations. On the other hand, when the conformational changes are small (as we believe they are in the present case), then one can consider the predictions of the present approach to be valid.

In addition to the instructive mechanistic conclusions that can be drawn from the present study, it provides encouraging indications of a possible general practical method—that is, the present study reproduces *without any adjustable parameters* the observed trend of the enzyme modification. Clearly, more studies are needed to exclude the possibility of an accidental success. However, it seems that our approach might offer the prospect of a quantitative computer-aided enzyme design. If such an approach is proven reliable, then one can assist experimental effort by simulating the effects of different possible mutations.

This work was supported by Grant GM-24492, National Institutes of Health.

1. Wilkinson, A. J., Fersht, A. R., Blow, D. M., Carter, C. & Winter G. (1984) *Nature (London)* **307**, 187–188.
2. Wells, T. N. C. & Fersht, A. R. (1985) *Nature (London)* **316**, 656–657.
3. Villafranca, J. E., Howell, E. E., Don Voet, D. H., Strobel, M. S., Ogden, R. C., Abelson, J. N. & Kraut, J. (1983) *Science* **222**, 782–788.
4. Dalbodie-McFarland, G., Cohen, L. W., Riggs, A. D., Morin, C., Itakura, K. & Richards, J. H. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 6409–6413.
5. Craik, C. S., Largman, C., Fletcher, T., Roczniak, S., Barr, P. J., Fletterick, R. & Rutter, W. J. (1985) *Science* **228**, 291–297.
6. Warshel, A. & Weiss, R. M. (1980) *J. Am. Chem. Soc.* **102**, 6218–6226.
7. Warshel, A. & Russell, S. T. (1984) *Quart. Rev. Biophys.* **17**, 283–422.
8. Blow, D. M., Birktoft, J. J. & Hartley, S. S. (1969) *Nature (London)* **221**, 337–340.
9. Huber, R. & Bode, W. (1977) *Acc. Chem. Res.* **11**, 114–122.
10. Kraut, J. (1977) *Annu. Rev. Biochem.* **46**, 331–358.
11. Warshel, A., Russell, S. T. & Weiss, R. M. (1982) in *Biomimetic Chemistry and Transition State Analogs*, eds. Green, B. S., Ashani, V. & Chipman, D. (Elsevier, Amsterdam), pp. 267–273.
12. Valleau, J. P. & Torrie, G. M. (1977) in *Modern Theoretical Chemistry*, ed. Berne, B. J. (Plenum, New York), Vol. 5, pp. 169–194.
13. Warshel, A. & Russell, S. T. (1985) *J. Mol. Biol.* **185**, 389–404.
14. Warshel, A. & King, G. (1985) *Chem. Phys. Lett.* **121**, 124–129.
15. Warshel, A. & Weiss, R. M. (1981) *J. Am. Chem. Soc.* **103**, 440–451.
16. Warshel, A. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 444–448.
17. Koshland, D. E., Jr., Nemethy, G. & Filmer, D. (1966) *Biochemistry* **5**, 365.
18. Huber, R., Kukla, D., Bode, W., Schwager, P., Bartels, K., Deisenhofer, J. & Steigemann, W. (1974) *J. Mol. Biol.* **89**, 73–101.
19. Warshel, A. (1984) in *Specificity in Biological Interaction*, eds. Chagas, C. & Pullman, B. (Pontificiae Academiae Scientiarum Scripta Varia, Citta Del Vaticano), Vol. 55, 59–81.
20. Wang, D., Bode, W. & Huber, R. (1985) *J. Mol. Biol.* **185**, 595–624.
21. Levitt, M. (1982) *Annu. Rev. Biophys. Bioeng.* **11**, 251–271.