Theoretical Correlation of Structure and Energetics in the Catalytic Reaction of Trypsin

Arieh Warshel* and Stephen Russell

Contribution from the Department of Chemistry, University of Southern California, Los Angeles, Los Angeles, California 90089-0482. *Received May 8, 1986*

Abstract: The empirical valence bond (EVB) approach is used in correlating the catalytic activity of trypsin with its observed X-ray structures. This analysis indicates that the most important factor in the catalysis of serine proteases is the stabilization of the negatively charged tetrahedral intermediate (t^-) by the main-chain N–H dipoles of the "oxyanion hole" and by bound water molecules. An analysis of the popular "charge-relay" mechanism indicates that this mechanism is unlikely to be important for the catalytic activity of serine proteases; apparently a transfer of a proton from His-57 to Asp-102 involves an increase of about 10 kcal/mol in the energy of the transition state. Our study indicates that the catalytic role of Asp-102 might be smaller than previously thought. The reliability of the present study is confirmed by evaluating the pK_a's of the relevant groups and by examining the sensitivity of the results to various factors.

I. Introduction

The mode of operation of serine proteases has been the subject of more than 30 years of intensive study. The detailed structures of these enzymes have been explored, by X-ray crystallography, to a greater extent than any other class of enzymes (for reviews see ref 1–4). It is now well established that the different enzymes in this class possess the same active center even when parts of their three-dimensional structure are entirely unrelated. The invariant active center includes the triad of Asp-102, His-57, and Ser-195, which act in a general-base-catalyzed reaction, involving nucleophilic attack by Ser-195 on the amide or ester substrates. However, despite the wealth of information about this reaction and its detailed description in many textbooks, it is still not entirely clear what the most important catalytic factors are. It is not clear, despite early proposals,[2,3] what the catalytic role of Asp-102 is. It is also not clear how such related systems as chymotrypsin and chymotrypsinogen differ by 7 orders of magnitude in their activity and why serine proteases catalyze related substrates (e.g., L- and D-amides) quite differently. In order to attack this problem it is important to correlate the detailed X-ray structures of the different active sites with their catalytic action. Such a correlation should be based on evaluation of the changes in interactions between the substrate and the enzyme during the reaction. Unfortunately it is hard to "measure" these interactions by experimental methods (although the emerging method of genetic engineering might change the situation). Thus is is important to try to "crack" the problem by theoretical approaches.

A quantitative correlation of the structures of serine proteases with their activity requires a reliable evaluation of the effect of the active site on the reaction potential surfaces. Steps in this direction have been reported in molecular orbital studies.[6-14] These studies, however, have not reached the quantitative level required for a unique discrimination between different alternative mechanisms and for consistent reproduction of the activation energy of the rate of limiting steps. That is, the catalytic effect of the enzyme is due to the interactions between the active site and the reacting system.[15] Crucial, therefore, is the effect of the active site on the reacting system and not so much the self-energy of the reacting system. Although some recent studies of serine proteases have included part of the electrostatic interaction between the enzyme and the reacting system, none (with the exception of ref 21) included the effect of the induced dipoles of the protein and the surrounding water molecules. These factors can contribute as much as 40 kcal/mol to the stabilization of ionic resonance forms.[15,20] Moreover, most studies (with the exception of ref 21 and 14) did not evaluate the energetics of the reference reaction in solution. Such an evaluation is important for identifying the actual contributions of different catalytic factors.[15] As a rule of thumb, since the energetics of enzyme catalysis is related to local pK_a's, it is important to use methods capable of evaluating pK_a's in proteins to within a few units in studying catalytic reactions.

A semiquantitative theoretical analysis of the reactions of serine proteases can be carried out by the empirical valence bond (EVB)[15,18] method used in some of our previous studies of enzymatic reactions.[15,18-20] This method allows one to focus on the *difference* between the energetics of the reaction in water and in the enzyme active site and to obtain a reliable estimate of the important catalytic factors. A recent preliminary study[21] used the EVB method and presented the first semiquantitative analysis of the catalytic reactions of serine proteases. This preliminary study is extended in the present work, which represents a much more complete description of the calculation and discussion of their implications.

The present study considers the rate-determining step in the acylation of amides by serine proteases. Such a study can be accomplished now (with the emergence of supercomputers) by a combination of the EVB and an umbrella sampling approach.[22] In fact, a study along this line on trypsin has been accomplished recently.[46] However, in the present work we would like to explore what can be learned about catalysis using static X-ray structures

(1) Bender, M. L.; Killheffer, J. V. *CRC Crit. Rev. Biochem.* **1973**, *1*, 149–199.
(2) (a) Blow, D. M.; Birktoft, J. J.; Hartley, B. S. *Nature (London)* **1969**, *221*, 337–340. (b) Blow, D. M. *Acc. Chem. Res.* **1976**, *9*, 145–152.
(3) Kraut, J. *Annu. Rev. Biochem.* **1977**, *46*, 331–358.
(4) Stroud, R. M.; Kossiakoff, A. A.; Chambers, J. L. *Annu. Rev. Biophys. Bioeng.* **1977**, *6*, 177–193.
(5) Huber, R.; Bode, W. *Acc. Chem. Res.* **1978**, *11*, 114–122.
(6) (a) Umeyama, H.; Imamura, A.; Nagata, C.; Hanano, M. *J. Theor. Biol.* **1973**, *41*, 485–502. (b) Umeyama, H.: Nakagawa, S.; Kudo, T. *J. Mol. Biol.* **1981**, *150*, 409–421.
(7) (a) Scheiner, S.; Kleier, D. A.; Lipscomb, W. N. *Proc. Natl. Acad. Sci. U.S.A.* **1975**, *72*, 2606–2610. (b) Scheiner, S.; Lipscomb, W. N. *Proc. Natl. Acad. Sci. U.S.A.* **1976**, *73*, 432–436.
(8) Kollman, P. A.; Hayes, D. M. *J. Am. Chem. Soc.* **1981**, *103*, 2955–2961.
(9) Van Duijnen, P. Th.; Thole, B. Th.; Hol, W. G. J. *Biophys. Chem.* **1979**, *9*, 273–280.
(10) Clementi, E. *Int. J. Quantum Chem.* **1980**, *17*, 651–660.
(11) Allen, L. C. *Ann. N. Y. Acad. Sci.* **1981**, *367*, 383–406.
(12) (a) Naray-Szabo, G.; Bleha, T. *Progress in Theoretical Organic Chemistry*; Csizmadia, T. G., Ed.; Elsevier: Amsterdam, 1982; Vol. 3, pp 267–336. (b) Angyan, J.; Naray-Szabo, G. *J. Theor. Biol.* **1983**, *103*, 349–356.
(13) Dewar, M.; Storch, D. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 2225–2229.
(14) Weiner, S.; Singh, U. C.; Kollman, P. *J. Am. Chem. Soc.* **1985**, *107*, 2219–2229.
(15) Warshel, A. *Biochemistry* **1981**, *20*, 3167–3177.
(16) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227–249.
(17) Warshel, A. *J. Phys. Chem.* **1979**, *83*, 1640–1652.
(18) Warshel, A.; Weiss, R. M. *J. Am. Chem. Soc.* **1980**, *102*, 6218–6226.
(19) Warshel, A. *Acc. Chem. Res.* **1981**, *14*, 284–290.
(20) (a) Warshel, A.; Russel, S. T. *Q. Rev. Biophys.* **1984**, *17*, 283–422. (b) Russell, S. T.; Warshel, A. *J. Mol. Biol.* **1985**, *185*, 385–404.
(21) Warshel, A.; Russell, S. T.; Weiss, R. M. *Biomimetic Chemistry and Transition State Analogs*; Green, B. S., Ashani, V., Chipman, D., Eds.; Elsevier: Amsterdam, 1982; pp 267–273.
(22) Warshel, A. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81*, 444–448.

0002-7863/86/1508-6569$01.50/0

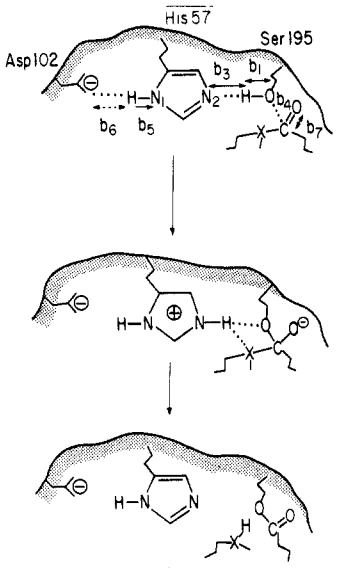

Figure 1. Schematic description of the acylation step in the catalytic reaction of trypsin. The bond notation used in eq 4 is indicated in the figure.

only. If such a simple structure–function correlation is of a semiquantitative value, then it should reproduce the trend in the observed catalytic effect.

The rate constant for amide hydrolysis in the trypsin active site, k_{cat}, is about 7 orders of magnitude larger than the rate constant for the corresponding reference reaction where all the reactants are in the same solvent cage, k_{cage} (for more details, see section III). The rate constants k_{cat} and k_{cage} are related to the corresponding activation free energies of eq 1, where k_b and h are the

$$k_{cat} = (k_bT/h)\exp\{-\Delta G^{\ddagger}_{cat}/k_bT\} \quad (1a)$$

$$k_{cage} = (k_bT/h)\exp\{-\Delta G^{\ddagger}_{cage}/k_bT\} \quad (1b)$$

Boltzmann and Planck constants, respectively. From this relation we obtain a difference of about 6 kcal/mol between $\Delta G^{\ddagger}_{cage}$ and $\Delta G^{\ddagger}_{cat}$. The origin of this difference (which is the key question in determining the catalytic role of the enzyme active site[15,19,20]) is analyzed in this paper.

Section II describes the EVB method and the detailed evaluation of the potential surfaces for serine proteases. Section III describes the results of the calculations. This includes comparison of the calculated and observed pK_a's of the catalytic groups (as a test of reliability for the method), analysis of the energetics of the proton-transfer stage, and examination of the stabilization of the transition state and the tetrahedral intermediate. The reliability of the calculations and their sensitivity to the uncertainty in the X-ray structures and the point charges used are analyzed in section IV.

The calculations presented in this paper indicate that the electrostatic stabilization of the tetrahedral intermediate by the oxyanion hole is the most important catalytic factor and that the charge-relay mechanism[2,3] has no catalytic advantage. It is argued that the role of the His-57 Asp-102 pair might be quite different than previously proposed.

II. How To Construct EVB Potential Surfaces for Serine Proteases

a. The Secular Equation. The EVB method represents the reaction as a simple crossing between covalent and ionic resonance forms, where the ground-state potential surface is obtained by solving the secular equation of these resonance forms. We demonstrate here the construction of the EVB secular equation for the reaction of serine proteases. This reaction can be classified as a general-base-catalyzed nucleophilic attack (by the hydroxyl oxygen of Ser-195) on the carbonyl carbon of the substrate. We concentrate here on the formation of the tetrahedral intermediate (R′O(CO$^-$)(R)X) in the acylation step of the hydrolysis of amides and esters (eq 2), where X is the O–R″ group of an ester or the

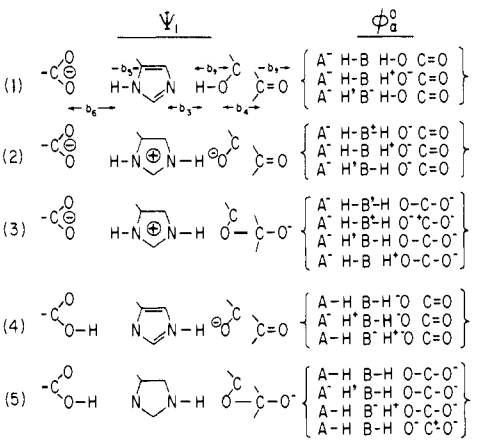

Figure 2. Mixed resonance forms (ψ_i) used to describe the reaction and the zero-order basis set of pure resonance forms (ϕ^0) used to obtain the mixing between the ψ_i.

NH–R″ group of an amide and Im and R′–OH designate His-57 and Ser-195 of the enzyme (see Figure 1). Other steps in the reaction can be studied by the same approach.

$$\begin{aligned}\text{Im} + \text{R}'\text{—OH} + \text{X—C(=O)R} \rightarrow \\ \text{Im}^+\text{—H} + \text{R}'\text{—O}^- + \text{X—C(=O)R} \rightarrow \\ \text{Im}^+\text{—H} + \text{R}'\text{O(CO}^-\text{)(R)X} \rightarrow \text{Im} + \text{R}'\text{—C(=O)R} + \text{XH}\end{aligned} \quad (2)$$

The reaction in eq 2 can be represented in terms of the resonance forms of eq 3, where A, Im, H—O, and C=O indicate, respectively, Asp-102, His-57, Ser-195, and the carbonyl of the substrate. The notation Im′ indicates that the N_1 nitrogen of the histidine is unprotonated. Here our selection of resonance forms

$$\psi_1 = (\text{A}^-\ \text{Im H—O C=O}) \quad (3a)$$

$$\psi_2 = (\text{A}^-\ \text{Im}^+\text{—H O}^-\ \text{C=O}) \quad (3b)$$

$$\psi_3 = (\text{A}^-\ \text{Im}^+\text{—H O—C—O}^-) \quad (3c)$$

$$\psi_4 = (\text{A—H Im}'\ \text{O}^-\ \text{C=O}) \quad (3d)$$

$$\psi_5 = (\text{A—H Im}'\ \text{O—C—O}^-) \quad (3e)$$

is somewhat different than in the original EVB formulation.[18] That is, in order to simplify the parametrization, the comparison to experimental information, and the discussion, we consider the H–O, C–O, A–H, and H–Im bonds as real bonds, with small ionic character, rather than pure covalent bonds. Thus, for example, the O–H bond in the ψ_1 resonance form includes mixing of the pure covalent resonance form and the high-energy ionic (H^+O^-) resonance form. The relation between the resonance forms of eq 3 and the pure resonance forms is described in Figure 2 and discussed in the Appendix. As explained in the Appendix, the high-energy resonance forms (H^+ O^-), (O^- C^+), A^- H^+), and (H^+ Im^-) are included implicitly in the mixing terms of the low-energy states. The gas-phase energies of the various resonance forms are given by eq 4, where g indicates the gas phase and b_1,

$$E_1^g = \Delta M(b_1) + \Delta M(b_5) + \Delta M(b_7) + V^{(1)}_{nb} + V^{(1)}_{QQ} + V^{(1)}_{strain} \quad (4a)$$

$$E_2^g = \alpha_2^g + \Delta M(b_3) + \Delta M(b_5) + \Delta M(b_7) + V^{(2)}_{nb} + V^{(2)}_{QQ} + V^{(2)}_{strain} \quad (4b)$$

$$E_3^g = \alpha_3^g + \Delta M(b_3) + \Delta M(b_5) + \Delta M(b_4) + V^{(3)}_{nb} + V^{(3)}_{QQ} + V^{(3)}_{strain} \quad (4c)$$

$$E_4^g = \alpha_4^g + \Delta M(b_3) + \Delta M(b_6) + \Delta M(b_7) + V^{(4)}_{nb} + V^{(4)}_{QQ} + V^{(4)}_{strain} \quad (4d)$$

$$E_5^g = \alpha_5^g + \Delta M(b_3) + \Delta M(b_6) + \Delta M(b_4) + V^{(5)}_{nb} + V^{(5)}_{QQ} + V^{(5)}_{strain} \quad (4e)$$

Table I. Bonded and Nonbonded Parameters:[a] (1) Morse Potentials and (2) Nonbonded Potential Functions[e]

bond	$\bar{D}$[c]	D	$\bar{a}$[b]	b_0
O—H	75	102	2.35[b]	0.96[b]
O^+—H	75	97	2.35[b]	0.96[b]
C_a—O	67	92[b]	2.06[b]	1.43[b]
$(C_b{=}O)_\sigma$	67	92[b]	2.06[b]	1.43[b]
$(C_b{=}O)_\pi$	54[d]	82[d]	2.06[d]	1.00[d]
N—H	78	93	2.35	1.00
N^+—H	78	93	2.35	1.00

atom	A	a	α	r^*	ϵ^*
H^+O^-	3470[b]	2.5[b]	0.8		
$C_a{}^+$ O^-	5288[b]	2.5[b]			
C_b O^-	5288	2.5			
$(C_b{}^+{-}O^-)_\pi$	1800[f]	2.5			
$(C_b{-}O^-)_\pi$	1800[f]	2.5			
H^+O	1500[b]	2.5[b]			
H O^-	65[b]	2.5[b]			
H O	65[b]	2.5[b]			
H^+N	1500	2.5			
H N	150	2.5			
O O				3.56	0.085
N O				3.10	0.10

[a] Energies are in kcal/mol, distances in Å. [b] Taken from ref 18. [c] D and $\bar{D}$ are the dissociation energies for the real and pure covalent bonds, respectively. The $\bar{D}$ are evaluated by the geometric mean relation used in ref 18 with values of 105, 51, 88, and 58 kcal/mol taken from bond dissociation energies of H_2, H-O-O-H, C_2H_2, and H_2NN-H_2.[35] The $\bar{D}$ for O^+–H and N^+–H are taken as the $\bar{D}$ of O–H and N–H, respectively. The D for the π contribution of the C=O bond is evaluated from the D values of 50 and 60 kcal/mol for the π contributions to the C=C and O=O bonds. [d] C_a and C_b designate, respectively, an sp^3 carbon and a carbonyl carbon. The C=O bond is described as composed of σ and π bonds. $D_{(C-O)_\pi}$ is chosen so that $D_{(C-O)_\sigma} + D_{(C-O)_\pi}$ gives the observed C=O bond energy of CH_2O (174 kcal/mol[35]). b_0 is chosen so that the minimum of $M_{(C-O)_\sigma} + M_{(C-O)_\pi}$ will coincide with the observed C=O bond length of 1.23 Å.[23] [e] The nonbonded interaction is given by $A\exp(-ar) - 166\alpha/r^4 + \epsilon((r/r)^{12} - (r^*/r)^6)$. Atoms which are not bonded in any resonance form interact by the nonbonded parameters of ref 16 (with the exception of the O N interaction). [f] The parameters for the nonbonded interactions between the carbon and oxygen in the ionic state of the C=O bond were determined by requiring that the minimum of $M_{(C-O)_\pi}$ (which is obtained by the mixing of the ionic and covalent contributions to the π bond) will be around 1.1 Å.

b_3, b_4, b_5, b_6, and b_7 are, respectively, the H—O, Im—H, C—O, H—Im, A—H, and C=O bond lengths (see Figure 1). $\Delta M(b)$ is the Morse potential with its reference energy taken as the equilibrium length of the given bond ($\Delta M(b) = M(b) + D$). The α^g values are the gas-phase energies of forming the indicated resonance forms at infinite separation between their fragments. $V_{nb}^{(i)}$ is the nonbonded interaction (excluding electrostatic interaction) between the fragments of the ith resonance form, and $V_{QQ}^{(i)}$ is the electrostatic interaction between these fragments. V_{QQ} is given in kcal/mol by eq 5, where m runs over fragments, μ runs

$$V_{QQ}^{i} = 332 \sum_{mm'} \sum_{\mu\mu'} Q_{m\mu}^{(i)} Q_{m'\mu'}^{(i)} / R_{m\mu,m'\mu'} \qquad (5)$$

over atoms, Q is atomic charge, and R is distance in Å. The term $V_{strain}^{(i)}$ in eq 4 represents the strain energy associated with the formation of the tetrahedral intermediate (see below). The bonding and nonbonding parameters needed for evaluation of eq 4 are given in Table I. The α^g values are given in Table II. Some of these parameters could have been evaluated from gas-phase experiments. However, since solution experiments are much more readily available, we prefer to calibrate them by using solution experiments and calculated solution energies (see below). The charges, Q, used for the active-site region are given in Table III.

The ground-state potential surface for the gas-phase reaction is obtained by finding the lowest solution, E_G, of the secular eq 6, where the diagonal elements of the matrix **H** are the E_i of eq

$$\mathbf{H}^g\mathbf{C}_G = \mathbf{E}_G{}^g\mathbf{C}_G \qquad (6)$$

4 ($H_{ii} = E_i$) and the mixing terms H_{ij}^g, which represent the interaction between the different resonance forms, and are given in the Appendix.

Table II. Gas-Phase Energies of the Pure Ionic and Covalent States[a]

index	resonance form	expression used	$\bar{\alpha}_\mu^g - \bar{\alpha}_1^g$
a	(N H—O C=O)		0
b	(N H^+O^- C=O)	$I_H - EA_0 + \bar{D}_{O-H}$	347
c	(N^+—H O^- C=O)	$\alpha_2{}^w - \Delta\Delta G_{sol}^{(2)}(\infty)$	154
d	(N^+–H O—C—O^-)	$\alpha_3{}^w - \Delta\Delta G_{sol}^{(3)}(\infty) + B^{(3)}$	120
e	(N^+—H O^- C^+—O^-)	$\bar{\alpha}_d{}^g + (D_{C-O} + I_C - EA_0)$	366
f	(N H—O C^+—O^-)	$\bar{\alpha}_e{}^g - \bar{\alpha}_c{}^g$	212
g	(N H^+ O—C—O^-)	$\bar{\alpha}_b{}^g + (\bar{\alpha}_d{}^g - \bar{\alpha}_c{}^g)$	308
h	(N H^+O^- C^+—O^-)	$\bar{\alpha}_b{}^g + (\bar{\alpha}_e{}^g - \bar{\alpha}_c{}^g)$	559

[a] Energies are in kcal/mol. The gas-phase α's ($\bar{\alpha}^g$) are obtained by using the indicated expression. I and EA are ionization potentials and electron affinities. The ($I_H - EA_0$) for $\bar{\alpha}_b$ is taken from ref 18, and $\alpha_2{}^w$ and $\alpha_3{}^w$ are taken from Table V. The term $B^{(3)}$ (34 kcal/mol) is needed since, in contrast to the situation in other bonds, the electrons of the broken π bond are kept at a close distance by the σ bonds, and the asymptotic energy is not evaluated at an infinite separation. $B^{(3)}$ is evaluated simply by requiring that the calculated minimum of $E_3{}^g$ will be equal to the estimate of Table V (147 kcal/mol). The value of α_e is determined by considering the energy associated with breaking the O_1–C bond of the O_1–C–O_2^- system and forming the O_1^- C^+ state. Since the electrostatic energy between C and O_2 is included explicitly in V_e, we obtain the result listed in the table. The $\Delta\Delta G_{sol}^{(i)}$ term is given by $\Delta G_{sol}^{(i)} - \Delta G_{sol}^{(1)}$, where the corresponding ΔG_{sol} are listed in the legends of Table V.

Table III. Region I Charges Used for the Active Sites of Trypsin[a]

		ionized	unionized	tetrahedral intermediate[c]
Asp-102	C	0.70	0.70	
	O	-0.85	-0.35	
	O	-0.85	-0.35	
His-57[b]	C	0.164	0.035	
	N_1	-0.161	0.013	
	H	0.187	0.187	
	C	0.545	0.110	
	H	0.075	0.077	
	N_2	-0.161	-0.520	
	H	0.187		
	C	0.096	0.028	
	H	0.068	0.070	
Ser-195	O	-1.000	-0.427	0.000
	H		0.427	0.000
substrate	C		0.392	0.200
	O		-0.392	-1.200

[a] Charges are given in atomic charge units. The histidine charges were obtained by the QCFF/PI method (ref 23). [b] The numbering system of the imidazole rings begins with the δ carbon and proceeds in a counterclockwise direction as represented in Figure 1. N_1 is the nitrogen on the interior side of the imidazole (next to Asp-102), and N_2 is the nitrogen which receives a proton from Ser-195 in the first step of the catalytic reaction of trypsin. [c] The model tetrahedral intermediate is formed by modifying the charges of Ser-195 and the substrate as shown in this column.

The EVB Hamiltonian for the reaction in water, H^w, is obtained by retaining the gas-phase mixing term ($H_{ij}{}^w = H_{ij}{}^g$) and modifying the diagonal elements by the addition of the corresponding solvation energies, $\Delta G^w{}_{sol}$. That is, we use relation 7.[15,18] This

$$E_i{}^w = E_i{}^g + \Delta G^w{}_{sol,i}(R) \qquad (7)$$

can be expressed for convenience as eq 8, where $\Delta G_{sol}(\infty)$ is the solvation energy of the ith resonance form at infinite separation between its fragments. The ($E_i{}^g - \alpha_i{}^g$) term includes all the terms

$$E_i{}^w = (E_i{}^g - \alpha_i{}^g) + (\alpha_i{}^g + \Delta G^w{}_{sol,i}(\infty)) + (\Delta G^w{}_{sol,i}(R) - \Delta G^w{}_{sol,i}(\infty)) \qquad (8a)$$

$$E_i{}^w = (E_i{}^g - \alpha_i{}^g) + \alpha_i{}^w + \Delta\Delta G^w{}_{sol,i} \qquad (8b)$$

of eq 4 except the contribution of $\alpha_i{}^g$. $\alpha_i{}^w$ is the energy of forming the indicated resonance form at infinite separation between its fragments, relative to the minimum value of E_1. As will be shown

Table IV. Calculated and Observed Solvation Energies of the Fragments Involved in the Reaction[a,b]

fragment	$(\Delta G^{Q}_{sol})_{calcd}$	$(\Delta G^{0}_{sol})_{calcd}$	$(\Delta\Delta G^{w}_{sol})^{0\rightarrow Q}_{calcd}$	$(\Delta\Delta G^{s}_{sol})^{0\rightarrow Q}_{obsd}$
HCOOH	−80 ± 5	−10 ± 2	−70 ± 5	−70 ± 3
CH_3OH	−94 ± 5	−7 ± 2	−87 ± 5	−89 ± 3
CH_3CH_2OH	−92 ± 5	−6 ± 2	−85 ± 5	−83 ± 3
imidazole	−62 ± 5	−4 ± 2	−59 ± 5	−55 ± 10
C-O-CN(C)OH	−85 ± 5	−6 ± 2	−79 ± 5	

[a] Energies are in kcal/mol. ΔG^{Q}_{sol} and ΔG^{0}_{sol} are the solvation energies of the charged and uncharged fragment, evaluated by the SCSSD method.[17] $\Delta\Delta G^{0\rightarrow Q}_{sol}$ is the difference in solvation energy of the charged and uncharged fragment (e.g., for imidazole it represents $\Delta G_{sol}(ImH^+) - \Delta G_{sol}(Im)$. [b] The observed solvation free energies are estimated from comparison of gas-phase and solution free energies for proton-transfer processes. This is done by using the relation $(\Delta G^{w}_{PT}(AH + H_2O \rightarrow A^- + H_3O^+) \simeq 2.3RT(pK_a(AH) - pK_a(H_3O^+)) \simeq \Delta H^{g}_{PT}(AH + H_2O \rightarrow A^- + H_3O^+) + \Delta\Delta G^{w}_{sol}(H_2O \rightarrow H_3O^+) + \Delta\Delta G^{w}_{sol}(AH \rightarrow A^-)$. The ΔH_{PT} for HCOOH and CH_3OH are taken from ref 17 while for CH_3CH_2OH and $CH_2OC(C)(C)OH$ they are estimated from ref 36. The observed ΔH_{PT} for imidazole is estimated by taking the proton affinity of pyridine[44] (225 kcal/mol).

below, the α^w can be evaluated directly from experiments in solution (e.g., α_2^w is determined from the energy of proton transfer from serine to histidine at infinite separation and is obtained from the corresponding differences in pK_a's). In view of the simplicity of the evaluation of α^w from readily available experiments, we replace the parameter α^g (which appears in the solution Hamiltonian only in the off-diagonal terms) by the corresponding estimated values in eq 9, using the observed α^w and the calculated solvation energies, ΔG^{w}_{sol}, rather than those found in direct gas-phase experiments.

$$\bar{\alpha}_i^g = \alpha_i^w - \Delta G^{w}_{sol}(\infty) \quad (9)$$

The EVB Hamiltonian for the reaction in the protein active site is simply obtained by replacing the solvation energies, ΔG^{w}_{sol}, of the various resonance forms by the corresponding interactions ("solvations") of these resonance forms with the enzyme active site, ΔG^{p}_{sol}. That is

$$H_{ii}^p = E_i^p = E_i^g + \Delta G^{p}_{sol,i}(R) \quad (10a)$$

The off-diagonal matrix elements are kept at their gas-phase values, so that

$$H_{ij}^p = H_{ij}^g \quad (10b)$$

The diagonal elements can be expressed by eq 11. Thus the entire difference between the reaction in the solution and in the protein is expressed in terms of differences in solvation energies.

$$H_{ii}^p = E_i^w + (\Delta G^{p}_{sol,i}(R) - \Delta G^{w}_{sol,i}(R)) \quad (11)$$

b. Calculation of Solvation Energies in Solutions and in Proteins. Reliable evaluation of the solvation energies, ΔG_{sol}, of the ionic resonance forms in polar solvents and in protein active sites is essential for meaningful comparison of solution and enzymatic reactions. Here we calculate the ΔG_{sol} by the previously developed microscopic dielectric models.[17-20] The solvation free energies in solution, ΔG^{w}_{sol}'s, are estimated by the surface constrained soft sphere dipoles (SCSSD) approach,[17,20] which represents the water molecules as point dipoles attached to the centers of soft spheres. The details of this approach and its adaption to the EVB method are described in ref 17 and 18. It is important to mention that the SCSSD is *calibrated* to reproduce the observed solvation free energies of different ions at 300°. In fact our new molecular dynamics version (the surface constraint all atoms (SCAAS) model[47]) gives similar results by much more expensive calculations. It is also important to mention that the actual calculations of the solvation energies and their changes during the reaction are verified by comparison to powerful experimental constraints (section IV).

The calculated solvation energies of the fragments involved in the reaction are summarized in Table IV. The stabilization (solvation) energies of the ionic resonance forms in enzyme active sites are calculated by evaluating the interactions between the charges of these resonance forms and their environment (the partial charges and induced dipoles of the enzyme atoms and the surrounding water molecules). This is done by the protein dipoles Langevin dipoles (PDLD) method which is considered in Figure 3 and described in detail in ref 20. These calculated solvation energies and their verification (using observed pK_a values) are considered in section III.

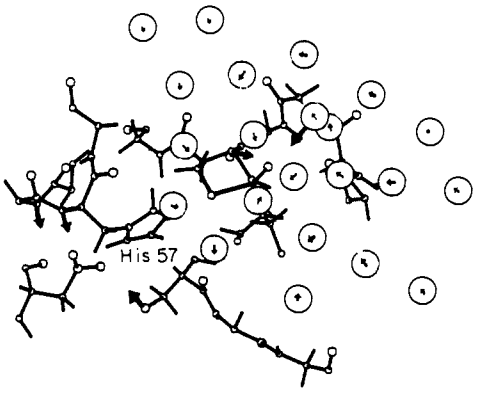

Figure 3. Schematic description of the PDLD model for the active site of trypsin and the surrounding water molecules. The figure shows the active site surrounded by Langevin dipoles (which represent the water molecules). The model considers, in addition to the Langevin dipoles, the protein permanent dipoles, the protein-induced dipoles, and the bulk around the Langevin dipoles. For more details, see ref 20b.

c. Treating the Effect of Intramolecular Strain. The formation of the tetrahedral intermediate involves a significant change in the geometry around the hydrolysis center. This effect is taken into account in a reliable way by the EVB treatment using empirical force fields to evaluate the deformation energies of the various resonance forms. The strain contributions are given by eq 12, where χ and $\theta_i^{(1)}$ are the out-of-plane angle and the bond

$$V^{(1)}_{strain} = V^{(2)}_{strain} = V^{(4)}_{strain} = \tfrac{1}{2}K_\chi(\chi - \chi_0)^2 + \sum_i \tfrac{1}{2}K_\theta^{(1)}(\theta_i^{(1)} - \theta_0^{(1)})^2 \quad (12a)$$

$$V^{(3)}_{strain} = V^{(5)}_{strain} = \sum_i \tfrac{1}{2}K_\theta^{(3)}(\theta_i^{(3)} - \theta_0^{(3)})^2 \quad (12b)$$

angles of the sp^2 hybridized C=O system. The corresponding force constants are taken from ref 23. The $\theta_i^{(3)}$ are the bond angles of the sp^3 system of the tetrahedral intermediate, $\theta_0^{(3)}$ is taken as the tetrahedral angle, and K_θ is the force constant for bending a C–C–C angle.[23] The strain energy terms in E_1, E_2, and E_4 express the energy associated with out-of-plane deformation of the C=O system from its equilibrium sp^2 geometry. The strain energy term for E_3 and E_5 expresses the energy associated with deformation of the sp^3 tetrahedral intermediate.

This type of treatment allows one to explore in a reliable way complicated sterical problems. For example, it can be used to explore the feasibility of the proposed deformation of the C=O system of the BPTI inhibitor by the active site of trypsin.[5]

d. Calibrated Potential Surface for Base-Catalyzed Amide Hydrolysis. A crucial element of the EVB model is the calibration and/or examination of the calculated potential surface of the reference reaction in solution, using relevant experimental information. In fact, when the alternative method of ab initio calculations will reach the stage of proper incorporation of the solvent molecules, it will have to be examined and probably calibrated by using the same experimental information. The

(23) Warshel, A.; Lappicirella, A. *J. Am. Chem. Soc.* **1981**, *103*, 4664–4673.

Table V. Determination of the α^w Parameters[a,b]

index	resonance form	expression used	α^w	$\bar{\alpha}^g - \bar{\alpha}_1^g$
1	(A^- Im H—O C=O)		0 (96)	0
2	(A^- Im^+—H O^- C=O)	$2.3RT(pK_a(Ser) - pK_a(ImH^+))$	12	154
3	(A^- Im^+—H O—C—O^-)	eq 15 + α_2^w	12	147
4	(AH Im′ O^- C=O)	$\alpha_2^w + 2.3RT(pK_a(ImH^+) - pK_a(Asp))$	17	31
5	(AH Im′ O—C—O^-)	$\alpha_3^w + 2.3RT(pK_a(ImH^+) - pK_a(Asp))$	17	20

[a] Energies are in kcal/mol. The α^w are the free energies of formation of the indicated resonance forms in solution at infinite separation between their fragments. All the α^w are determined from experimental information by using the indicated expressions (see text also). [b] The formation energies of the ionic resonance forms in the gas phase, $\bar{\alpha}^g$, are needed only for determination of the mixing terms (see Appendix). These parameters are determined by the relation $\bar{\alpha}^g = \alpha^w - \Delta G_{sol}(\infty)$ (eq 9) where $\Delta G_{sol}(\infty)$ is the sum of the solvation energies of the relevant fragments at infinite separation. Here we use $\Delta G_{sol}(\infty)$ values of −80, −90, −62, and −85 kcal/mol for Asp^-, Ser^-, ImH^+, and tet^- and values of −10, −6, −4, and −6 kcal/mol for the corresponding unionized fragments. The value of α_1^g, which is taken as a reference, is 96 kcal/mol.

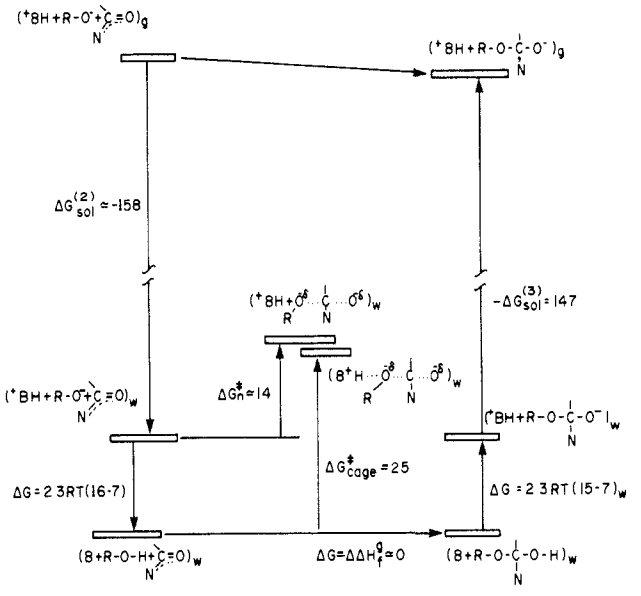

Figure 4. Thermodynamic cycle used to determine the energy of the transition state and the oxyanion configuration (see text for more details).

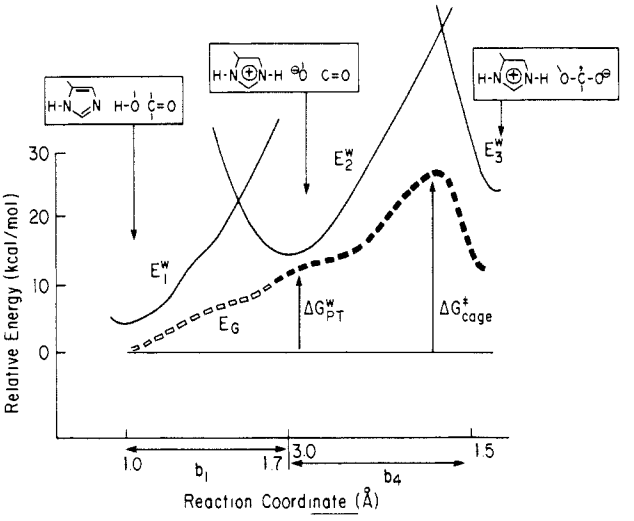

Figure 5. Calibrated EVB potential surface for general-base hydrolysis of amides in a solvent cage. The figure describes the energies E_1, E_2, and E_3 and the ground-state energy, E_G, obtained from their mixing along the reaction coordinate that involves proton transfer (b_1 = 1.0 Å → b_1 = 1.7 Å) and displacement of the serine oxygen toward the peptide carbonyl (b_4 = 2.7 Å → b_4 = 1.5 Å). ΔG^w_{PT} is the free energy for proton transfer in solution.

reference reaction in the present case is simulated by considering formic acid, imidazole, methanol, and dimethylamide in a solvent cage. The calibration procedure involves the following experimental information (see also Figure 4 and Table V). (i) The energetics for a transfer of a proton from serine to imidazole, at infinite separation, is calibrated by eq 13. (ii) The energy of

$$\Delta G^w_{PT} = 2.3RT(pK_a(Ser) - pK_a(ImH^+)) \quad (13)$$

forming the tetrahedral intermediate from the ionized serine and the amide is evaluated by considering the thermodynamic cycle 14 (Figure 4). The value of ΔG_2 is estimated by using standard

$$\Delta G(R{-}O^- + \underset{|}{\overset{|}{C}}{=}O \rightarrow R{-}O{-}\underset{|}{\overset{|}{C}}{-}O^-) \approx$$
$$\Delta G_1(O^- + \underset{|}{\overset{|}{C}}{=}O + H^+ \rightarrow O{-}H + \underset{|}{\overset{|}{C}}{=}O) +$$
$$\Delta G_2(O{-}H + \underset{|}{\overset{|}{C}}{=}O \rightarrow O{-}\underset{|}{\overset{|}{C}}{-}OH) +$$
$$\Delta G_3(O{-}\underset{|}{\overset{|}{C}}{-}OH - O{-}\underset{|}{\overset{|}{C}}{-}O^- + H^+) \quad (14)$$

groups contributions[35b] for the relevant enthalpies of formation, while ΔG_1 and ΔG_2 are evaluated by using the relevant pK_a's. This gives eq 15 (see Figure 4), which gives a ΔG of about 0. A related

$$\Delta G \simeq \Delta\Delta H_f(R{-}OH + \underset{|}{\overset{|}{C}}{=}O \rightarrow R{-}O{-}\underset{|}{\overset{|}{C}}{-}OH)$$
$$- 2.3RT(pK_a(R{-}OH) - pK_a(O{-}\underset{|}{\overset{|}{C}}{-}OH)) \quad (15)$$

gas-phase cycle (the upper part of Figure 4) gives a similar estimate. (iii) The activation barrier for the nucleophilic attack is estimated as follows: The analysis[49] of the hydrolysis of urea by concentrated OH^- gives a pseudo-first-order rate constant of 3×10^{-8} s^{-1} for 1 M OH^-. Interpolation to 55 M (which guarantees the presence of the OH^- in the solvent cage of the amide) gives a rate constant of about 1.5×10^{-6} s^{-1} and an activation free energy of about 26 kcal/mol. However, we are interested in the hydrolysis of amides by methoxy and not by an OH^- ion. Using an analysis of the type presented in eq 15 one obtains a ΔG of about 12 kcal/mol for a nucleophilic attack by OH^-, as compared to a ΔG of about 0 for an attack by a methoxy ion. Assuming a linear free energy relationship between $\Delta G^‡$ and ΔG gives $\Delta G^‡ \simeq 26 - 12 = 14$ for hydrolysis of an amide by a methoxy ion in a solvent cage. Including the proton-transfer step (from serine to histidine) gives an estimated activation free energy, $\Delta G^‡_{cage}$, of ~25 kcal/mol for our reference reaction in a solvent cage. The above arguments can be clarified by inspection of Figure 4.

The effect of the ionized acid (A^-) is included only for the purpose of examining the so-called charge-relay mechanism.[2] Experiments in aqueous solutions[24] indicated that the presence of an ionized group near the imidazole does not lead to a large catalytic effect ($\Delta G^‡_{cage}$ is reduced by about 1 kcal/mol). This effect is considered in section IIId.

The calculated EVB surface for our reference reaction in solution is presented in Figure 5. The corresponding potential surface for the enzymatic reaction will be considered in the next section.

III. Calculations of the Energetics of the Catalytic Reaction of Serine Proteases

a. pK_a's of the Catalytic Groups. Before potential surfaces are evaluated for enzymatic reactions it is crucial to examine the

(24) Rogers, G. A.; Bruice, T. C. *J. Am. Chem. Soc.* **1974**, *96*, 2473–2780.

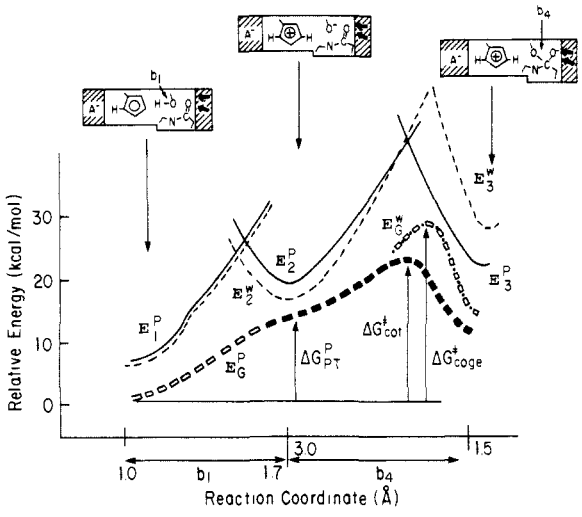

Figure 6. Comparison of the EVB potential surface for formation of the tetrahedral intermediate (t^-) in water (w) and in the active site of the protein (p). Notation as in Figure 5. The figure illustrates how the stabilization of ψ_3 leads to a reduction of $\Delta G^{\ddagger}_{cat}$ relative to $\Delta G^{\ddagger}_{cage}$.

reliability (and the relevance) of the calculation by comparing the calculated and observed pK_a's of the catalytic groups; large disagreement means that the method used cannot be applied to studies of stabilization of ionic transition states and therefore cannot be used for evaluation of the energetics of the catalytic reaction.

Our calculations of the pK_a's in proteins are based on expression 16,[15,19,20] where p and w indicate protein and aqueous solution, respectively. The evaluation of this simple expression involves

$$pK_a^p = pK_a^w + 2.3RT(\Delta G^p_{sol}(AH \rightarrow A^-) - \Delta G^w_{sol}(AH \rightarrow A^-)) \quad (16)$$

only calculations of solvation energies (with an error bound of ±5 kcal/mol)[20] and *does not* require any quantum mechanical calculations. The calculations of the pK_a's of the catalytic groups of trypsin (summarized in Tables VI and VII) were performed using the coordinates of the trypsin + BPTI complex.[32] These calculations are discussed below.

The assignment of the pK_a's of His-57 and Asp-102 in serine proteases has long been controversial.[25-28] However, recent neutron diffraction experiments[30] (that located protons of His-57 in MIP-inhibited trypsin) and careful NMR studies[29] have indicated that the pK_a of His-57 is around 7 and that the pK_a of Asp-102 is around 3. The present calculations give pK_a's of ~7 and ~0 for His-57 and Asp-102, respectively. Even a possible error bound of 5 pK_a units in the calculations leaves His-57 as the only catalytic group with a pK_a of ~7. The electrostatic stabilization of Asp-102 by three specially oriented dipoles (see Figure 7) is larger (more than 5 kcal/mol) than the corresponding stabilization of an ionized carboxylate in aqueous solution. This guarantees that Asp-102 will stay ionized at all the relevant catalytic configurations (see part d of this section).

The calculations of the energy required to ionize Ser-195 by forming the (A^- Im Ser$^-$ C=O) configuration [rather than the (A^- ImH$^+$ Ser$^-$ C=O) configuration] are quite instructive. The

Table VI. Comparison of the Energy of Various Configurations in Trypsin and in Solution[a]

configuration	V_{QQ}	$\Delta\Delta G^p_{Q\mu}$	$\Delta\Delta G^p_{Q\alpha}$	$\Delta\Delta G^p_{Qw}$	$\Delta\Delta G^p_{sol}$	$\Delta\Delta G^w_{sol}$	$\Delta\Delta G^w_{tot}$	$\Delta\Delta G^{p,calcd}_{tot}$	$\Delta\Delta G^{p,obsd}_{tot}$
(1) (A^- Im H—O C=O)	-23.2	-35.2	14.0	-22.8	-44.0	-47.0	0.0	0.0 (3.0)	0.0
(2) (A^- Im$^+$—H O$^-$ C=O)	-153.2	-32.8	0.8	-21.1	-63.1	-59.0	11.0	14 (17)	
(3) (A^- Im$^+$—H O—C—O$^-$)	-119.6	-54.2	-3.4	-26.2	-83.8	-82.6	12.0	7 (10)	
(4) (A^- Im$^+$—H (O···C···O)$^-$)	-111.9	-41.4	6.6	-18.0	-52.7	-48.0	26.0	21 (24)	18[c]
(5) (A—H Im′O$^-$ C=O)	15.0	-12.7	-21.9	-36.2	-70.8	-96.0	17.0	40 (43)	
(6) (A^- Im$^+$—H H—O—C—O$^-$)	-103.0	-19.6	-6.2	-9.0	-22.4	-25.0	1.38 (8 - pH)	1.38 (8 - pH)	1.38 (8 - pH)
(7) (A—H Im$^+$—H H—O C=O)	-6.2	-15.5	-30.1	-30.2	-45.0	-52.0	1.33 (pH - 4) + α_6^w	1.33 (pH - 0) + α_6^w	1.38 (pH - 3) + α_6^w
(8) (A^- Im O$^-$ C=O)	-45.8	-55.1	-39.0	-94.7	-188.2	-214.0	1.38 (16 - pH)	1.38 (33 - pH)	

[a] Energies (in kcal/mol) are evaluated at the minimum-energy geometry of the indicated configuration, where the rest of the protein is fixed at the given X-ray structure. ΔV_{QQ}, $\Delta\Delta G_{Q\mu}$, $\Delta\Delta G_{Q\alpha}$, and $\Delta\Delta G_{Qw}$ are, respectively, the electrostatic interaction between the fragments of the indicated configurations, the interaction of these configurations and the enzyme point charges (dipoles), the interaction with the protein-induced dipoles, and the interaction with the surrounding water molecules (Langevin dipoles). ΔG^p_{sol} and ΔG^w_{sol} are the total solvation energies of the indicated configuration in protein and in water. The notation $\Delta\Delta G$ indicates that the given value is taken relative to that of the fully neutral configuration (A—H ImH H—O C=O) which has values of -1.8, -10.5, -808.2, -34.0, and -852.9 kcal/mol for ΔV_{QQ}, $\Delta G_{Q\mu}$, $\Delta G_{Q\alpha}$, ΔG_{Qw}) and ΔG^p_{sol}, respectively. The $\Delta\Delta G^w_{tot}$ values are the energies (relative to the energy of the first configuration) of the various resonance forms in solution, estimated from observed pK_a values and bond energies (see Table V). $\Delta\Delta G^p_{tot}$ is the relative total energy of the indicated resonance form in the protein active site, $\Delta\Delta G^p_{tot} = \Delta\Delta G^w_{tot} + (\Delta\Delta G^p_{sol} - \Delta\Delta G^w_{sol})$. This energy is taken relative to the energy of the (A^- H—Im H—O C=O) configuration. The $\Delta\Delta G^p_{tot}$ in parentheses are taken relative to the $\Delta\Delta G^w_{sol}$ of the first configuration. [c] See text for the estimate of this observed value.

(25) Hunkapillar, M. W.; Smallcombe, S. H.; Whitaker, D. R.; Richards, J. H. *Biochemistry* **1973**, *12*, 4732-4743.
(26) Koeppe, R. E.; Stroud, R. M. *Biochemistry* **1976**, *15*, 3450-3458.
(27) Robillard, G.; Shulman, R. *J. Mol. Biol.* **1972**, *71*, 507-511.
(28) (a) Bachouchin, W. W.; Roberts, J. D. *J. Am. Chem. Soc.* **1978**, *100*, 8041-8047. (b) Kanamori, K.; Roberts, J. D. *Acc. Chem. Res.* **1983**, *16*, 35-41.
(29) Markley, J. L.; Ibanez, I. B. *Biochemistry* **1978**, *17*, 4627-4640.
(30) (a) Kossiakoff, A. A.; Spencer, S. A. *Biochemistry* **1981**, *20*, 6462-6474. (b) Kossiakoff, A. A.; Spencer, S. A. *Nature (London)* **1980**, *288*, 414-416.
(31) Birktoff, J. J.; Blow, D. M. *J. Mol. Biol.* **1972**, *68*, 187-240.

Table VII. Calculated and Observed pK_a's of the Catalytic Groups in Trypsin[a]

process	$\Delta\Delta G^p_{Q\mu}$	$\Delta\Delta G^p_{Q\alpha}$	$\Delta\Delta G^p_{Qw}$	$\Delta\Delta G^p_{sol}$	$\Delta\Delta G^w_{sol}$	$(pK_a^w)_{obsd}$	$(pK_a^p)_{calcd}$	$(pK_a^p)_{obsd}$
ionization of His-57 $A^- Im^+ \rightarrow A^- ImH^+$ (1 → 6)	+15.7	−7.8	−13.8	21.7	22 ± 5	8	7 ± 3	8[b]
ionization of Asp-102 AH $ImH^+ \rightarrow A^- ImH^+$ (7 → 6)	−34.9	36.2	21.2	22.6	27 ± 5	4	0 ± 3	3[b]
ionization of Ser-195 Im Ser → Im Ser^- (1 → 8)	−19.9	−53.0	−71.9	−144.2	−169 ± 5	16	≤33 ± 5[c]	

[a] The calculations are obtained by using Table VI; energies are in kcal/mol. Notation of the various energy contributions is as in Table VI. The ionization states of the relevant groups are indicated in parentheses, where A, Im, and Ser indicate Asp-102, His-57, and Ser-195, respectively. The pK_a values are obtained by taking the energies of the indicated configurations from Table VI. [b] Taken from a study of chymotrypsin (ref 29). [c] The inequality sign indicates that this estimate is for a rigid protein and that the protein will undergo some unfolding to give better stabilization of the ionized Ser^-.

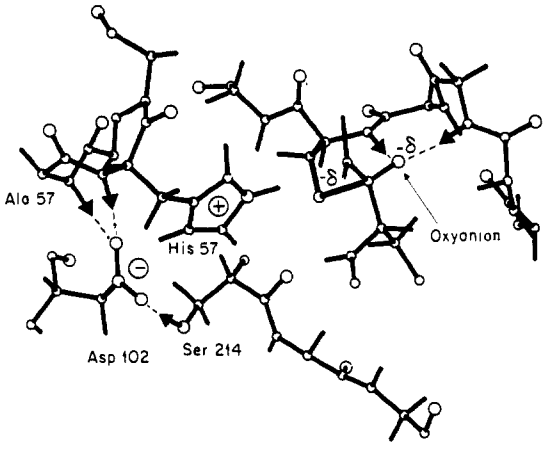

Figure 7. Dipoles responsible for the stabilization of the (− + −) configuration in serine proteases. The negatively charged oxygen of t^- is stabilized by the oxyanion hole (Ser-195 and Gly-193) while ionized Asp-102 is stabilized by the main-chain dipoles of Ala-56 and His-57 and by the hydroxyl dipole of Ser-214.

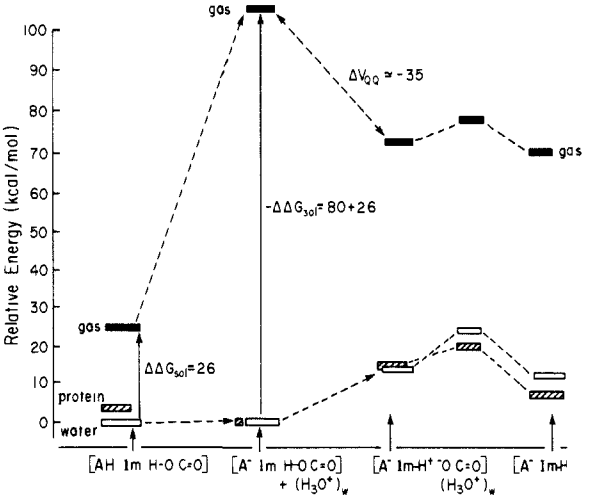

Figure 8. Energetics of the catalytic reaction in the gas phase, in the solution, and in the active site of trypsin. The figure demonstrates that the reaction in protein is not related at all to the corresponding gas-phase reaction.

calculated pK_a^p for this process is about 33. Such a high calculated value indicates that the native X-ray structure of the protein *cannot* stabilize the negatively charged Ser without the protonated imidazole, which does not exist, however, at high pH. Thus the protein–water structure around the ionized Ser-195 must be quite different than the native structure. The protein must either reorient its dipoles or undergo partial unfolding to allow for larger stabilization of the ionized serine by the surrounding water molecules (see discussion in ref 15 and 20).

The comparison of the calculated and observed pK_a^p's in Table VII gives an error bound of 3 pK_a units (±5 kcal/mol). This can be considered as the error bound on the calculated potential surface for the reaction (next section); the EVB error range is determined by the same electrostatic calculations used for the pK_a calculations.

b. Energetics of the Proton-Transfer Process. The mechanism of action of serine proteases involves the transfer of a proton from Ser-195 to His-57 (see Figure 1). The energetics of this process is considered in Figure 6 and Table VI. The calculated free energies for proton transfer in the protein (ΔG^p_{PT}) and in solution (ΔG^w_{PT}) are ~13 and ~10 kcal/mol, respectively. These results can be analyzed without the EVB calculations, using relation 17,[15]

$$\Delta G^p_{PT}(R) \simeq \Delta G^w_{PT}(R) + (\Delta G^p_{sol}(R) - \Delta G^w_{sol}(R)) \quad (17)$$

where R is the distance between the proton donor and acceptor and the $\Delta G^p_{sol} - \Delta G^w_{sol}$ term is the difference between the solvation energy of the Ser^- ImH^+ ion pair in the protein and in solution. The free energy of proton transfer in solution is given to a very good approximation by eq 18,[15] where we used pK_a values of ~16

$$\Delta G^w_{PT}(R) \simeq 2.3RT(pK_a(Ser) - pK_a(ImH^+)) \pm 3 = 12 \pm 3 \text{ kcal/mol} \quad (18)$$

and ~7 for serine and histidine, respectively. From eq 17 and 18 one obtains eq 19. The evaluation of this expression does not

$$\Delta G^p_{PT}(R) \simeq 12 + (\Delta G^p_{sol}(R) - \Delta G^w_{sol}(R)) \pm 3 \text{ kcal/mol} \quad (19)$$

require *any* quantum mechanical calculations and depends only on the reliability of the calculations of the solvation free energies. ΔG^p_{PT} evaluated from eq 19 is 14 kcal/mol as compared to 13 kcal/mol obtained from the complete EVB calculations. Both values are larger than ΔG^w_{PT} since $\Delta G^p_{sol} - \Delta G^w_{sol}$ is positive, indicating that the enzyme does not catalyze this step (which is not the rate-limiting step).

It is important to emphasize at this point that calculations of proton transfer from Ser-195 to His-57 must include the stabilization of the Ser^- His^+ ion pair by the protein; otherwise, the results can be in error by about 30 kcal/mol regardless of the method used. That is, experimental analysis of the stabilization of ion pairs in solutions gives relation 20,[20] where V_{QQ} is the

$$\Delta G^w_{sol}(R) \simeq \Delta G^w_{sol}(\infty) - V_{QQ}(R) \pm 3 \text{ kcal/mol} \quad (20)$$

interaction between ImH^+ and Ser^- in vacuum and $\Delta G^w_{sol}(\infty)$ is the solvation energy of ImH^+ and Ser^- at infinite separation, which is estimated to be (−58 − 85) ± 5 kcal/mol by the calibrated estimate of the SCSSD model (see Table IV). With a V_{QQ} of −110 ± 10 (depending on R), we obtain a $\Delta G^w_{sol}(R)$ of −30 ± 10 kcal/mol. Since the solvation energy by the protein, ΔG^p_{sol}, cannot be larger than ΔG^w_{sol} by more than 10 kcal/mol (otherwise the protein will unfold and water molecules will give the proper stabilization[15,20]), we find that ΔG^p_{sol} = −30 ± 20 kcal/mol. Neglecting ΔG^p_{sol} will result in an error of about 30 kcal/mol. This result is based on *experimental* considerations and is independent of the method used for calculating the energy of the ImH^+ Ser^- ion pair.

One can argue that including the interaction between the ImH^+ Ser^- ion pair and ionized Asp-102 can give a very large electrostatic stabilization even in the absence of the protein or any other solvent molecules.[6,7,13] However, such an argument overlooks the fact that the formation of Asp^- from Asp–H under vacuum costs about 345 kcal/mol.[18] This process costs about 70 kcal/mol when the proton is transferred to aqueous solution.[15,20] Since the electrostatic interaction, V_{QQ}, between Asp^- and the ion pair is about 35 kcal/mol, the $(Asp^- ImH^+ Ser^-)_g + (H^+)_{aq}$ configuration is *less* stable by about 35 kcal/mol (70 − 35 kcal/mol) than the $(Asp–H\cdots ImH^+ Ser^-)_g$ configuration. In-

spection of Figure 8 might make the above argument somewhat clearer.

c. Stabilization of the Transition State. The rate-determining step, in the general-base-catalyzed hydrolysis of amides by serine proteases, involves a transition state with a large contribution from the oxyanion configuration, ψ_3. The enzyme increases the rate constant, k_{cat}, of this step by reducing the activation free energy, $\Delta G^{\ddagger}_{cat}$. In Table VI and Figure 5 we analyze the energetics associated with the reduction of $\Delta G^{\ddagger}_{cat}$. The table presents the calculations for trypsin (using the coordinates of the trypsin + PTI complex[32]). The calculations for trypsin and the corresponding solution reactions are summarized in Figure 6 and Table VI. Figure 6 shows the dependence of E_1, E_2, and E_3 on the reaction coordinate and the potential surface obtained by their mixing. The figure shows that the enzyme stabilizes the ionic resonance form (A^- ImH^+ t^-) (where t^- designates the negatively charged tetrahedral intermediate) by about 5 kcal/mol more than water does. This results in a reduction of the activation energy in the enzyme $\Delta G^{\ddagger}_{cat}$ by $\sim$5 kcal/mol relative to the corresponding activation energy in a solvent cage, $\Delta G^{\ddagger}_{cage}$. In considering the corresponding experimental reduction, we compare our estimate of $\Delta G^{\ddagger}_{cage} \simeq 25$ kcal/mol (section IIb) to $\Delta G^{\ddagger}_{cat}$, which is estimated to be around 18 kcal/mol (using k_{cat} in the range 0.1–10 s^{-1} for bad and good substrates, respectively[51]). Thus our rough estimate of the observed difference between $\Delta G^{\ddagger}_{cage}$ and $\Delta G^{\ddagger}_{cat}$ is about 7 kcal/mol.

The above considerations indicate that the entire rate acceleration of amide hydrolysis serine proteases can be accounted for in terms of the difference between electrostatic stabilization (solvation energy) of the ionic resonance forms in solution and in the enzyme active site. This finding is consistent with general considerations[19,20] and previous studies of the catalytic reaction of lysozyme.[15,16,20]

The structural elements responsible for the stabilization of the transition state are described in Figure 7. The most important factors are the main-chain dipoles of Ser-195 and Gly-193 and water molecules that stabilize the negative charge of the tetrahedral intermediate (these dipoles form the so-called oxyanion hole).[38,39] An additional key factor is introduced by the three dipoles that stabilizes the ionized Asp-102. This type of arrangement is invariant in different serine proteases.[3,39]

d. What about the Charge-Relay System? It has been widely accepted in the literature (see, for example, ref 2 and most biochemical textbooks) that the catalytic reaction of serine proteases involves stabilization of the tetrahedral intermediate by a proton transfer from His-57 and Asp-102. This so-called charge-relay mechanism was accepted not only because of supporting experimental information[25,26] (which could have been interpreted differently[27]) but because of a common belief that uncharged intermediates are always more stable than charged intermediates. This "rule" misses a key point in bioenergetics, that enzyme active sites are even more polar than aqueous solutions[15,19,20] and that they can stabilize polar transition states in a very effective way. Early theoretical studies supported the charge-relay mechanism,[6a,7] while more recent calculations,[6b,8] indicated that this mechanism might be incorrect. Although we agree with the more recent studies, we feel that they still could not provide conclusive evidence against this mechanism. That is, as will be shown below, the charge-relay mechanism is only 6 kcal/mol unfavorable in solutions, and the protein can make it favorable. Thus, it is essential to use methods that evaluate all the contributions to $\Delta G^{\ddagger}_{sol}$ in order to reach an error range smaller than the energy difference in question.

The EVB representation allows one to focus on the key aspects of the charge-relay mechanism in a simple quantitative way, relating it to the relative energies of ψ_5 and ψ_3. If the charge-relay mechanism is important, then the minimum value of E_5 (complete proton transfer to Asp-102) should be lower than the minimum of E_3 (where the proton is attached to His-57). The energy difference between E_5 and E_3 in solution is given by eq 21. The

$$E_5^w(b_6 = 1.0) - E_3^w(b_5 = 1.0) \simeq \{1.33(pK_a^w(ImH^+) - pK_a^w(AH))\} + \{(V_{QQ}^{(5)} + \Delta G_{sol}^{(5)}) - (V_{QQ}^{(3)} + \Delta G_{sol}^{(3)})\} \quad (21)$$

first term represents the reversible work of proton transfer from ImH^+ to A^- at infinite separation, and the second term represents the electrostatic work of bringing A^- and (ImH^+ t^-) to their interaction distance in the (A^- ImH^+ t^-) configuration. The pK_a term gives $\sim$4 kcal/mol, and the second term gives $\sim$2 kcal/mol. Thus, it is clear that the charge-relay mechanism is not a useful catalytic channel in polar solvents (see the experimental study of ref 24). The key question, however, is what about this mechanism in the enzyme? This question can be examined simply by evaluating the difference between E_5 and E_3 in the protein with eq 22, where p and w indicate protein and solution, respectively,

$$E_5^p - E_3^p = E_5^w - E_3^w + (\Delta\Delta G^{w-p_{sol}})_5 - (\Delta\Delta G^{w-p_{sol}})_3 \quad (22)$$

and $\Delta\Delta G^{w-p_{sol}} = \Delta G^{p}_{sol} - \Delta G^{w}_{sol}$. The calculated values of $(\Delta\Delta G^{w-p_{sol}})_5$ and $(\Delta\Delta G^{w-p_{sol}})_3$ are 1.4 and −1.2 kcal/mol, respectively. This gives $E_5^p - E_3^p = 6.0 + 2.6 \simeq 9$ kcal/mol, placing the energy of the charge-relay configuration, E_5, above E_3 and indicating that the protonation of Asp-102 has no significant catalytic advantage. Note (Table VII) that our calculations gave $pK_a \simeq 0$ for Asp-102 in ψ_1 and ψ_3, making its protonation harder in the enzyme than in solution.

IV. Reliability of the Calculations

Before model calculations are used for making a mechanistic judgment, it is essential to establish the error associated with the given approach. The present approach is designed to minimize the error in the calculations by concentrating on *differences* between related quantities:[18] solvation of the same resonance forms by the enzyme (ΔG^{p}_{sol}) and by aqueous solutions (ΔG^{w}_{sol}). Thus the error limit of the calculation is determined only by the errors in ΔG^{w}_{sol} and ΔG^{p}_{sol} and not by the reliability of any gas-phase calculations or gas-phase experiments. As far as the solvation free energies in solution are concerned, we note that the SCSSD method gives reasonable free energies for the isolated fragments of the relevant resonance structures (see Table IV). More importantly, by estimating the solvation energies of the isolated fragments we obtain quite reliable estimates of the energies of the relevant resonance structures at any configuration, R. That is, as we noted some time ago,[15] the energy of a given ionic resonance structure is given approximately by eq 23, where V_{QQ}

(32) Huber, R.; Kukla, D.; Bode, W.; Schwager, P.; Bartles, K.; Deisenhofer, J.; Steigemann, W. *J. Mol. Biol.* **1974**, *89*, 73–101.
(33) Epand, R. M.; Wilson, I. B. *J. Biol. Chem.* **1965**, *240*, 1104–1107.
(34) Warshel, A. *Proc. Natl. Acad. Sci. U.S.A.* **1978**, *75*, 5250–5254.
(35) (a) Benson, S. W. *J. Chem. Educ.* **1965**, *42*, 502–518. (b) Benson, S. W. *Thermochemical Kinetics*; Wiley-Interscience: New York, 1976.
(36) Dawson, R. M. C., Elliot, D. C., Elliot, W. H., Jones, K. M., Eds. *Data for Biochemical Research*, 2nd ed.; Oxford, Clarendon, 1974.
(37) Freer, S. T.; Kraut, J.; Robertus, J. D.; Wright, H. T.; Xuong, N.-H. *Biochemistry* **1970**, *9*, 1997–2009.
(38) Henderson, R. *J. Mol. Biol.* **1970**, *54*, 341–354.
(39) Robertus, J. D.; Kraut, J.; Alden, R. A.; Birktoff, J. *Biochemistry* **1972**, *11*, 4293–4303.
(40) Brayer, G. D.; Delbaere, L. T. J.; James, M. N. G. *J. Mol. Biol.* **1978**, *124*, 261–282.
(41) Kebarle, P. *Annu. Rev. Phys. Chem.* **1977**, *28*, 445–476.
(42) Bartmess, J. E.; McIver, R. T. In *Gas Phase Ion Chemistry 2*; Bowers, M. T., Ed.; Academic: New York, 1979; pp 87–121.
(43) Henderson, R. *Biochem. J.* **1971**, *124*, 13–18.
(44) Taagepera, M.; Henderson, W. G.; Brownlee, R. T. C.; Beauchamp, J. L.; Holtz, D.; Taft, R. W. *J. Am. Chem. Soc.* **1972**, *94*, 1369–1373.
(45) Warshel, A.; Weiss, R. M. *J. Am. Chem. Soc.* **1981**, *103*, 446–451.
(46) Warshel, A.; Sussman, F. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 3806–3810.
(47) Warshel, A.; King, G. *Chem. Phys. Lett.* **1985**, *121*, 124–129.
(48) Wang, D.; Bode, W.; Huber, R. *J. Mol. Biol.* **1985**, *185*, 595–624.
(49) Lynn, K. R. *J. Phys. Chem.* **1966**, *96*, 687–689.
(50) Steitz, T. A.; Shulman, R. G. *Annu. Rev. Biophys. Bioeng.* **1982**, *11*, 419–444.
(51) Fersht, A. R. *Enzyme Structure and Mechanism*; W. H. Freeman: San Francisco, 1977.

$$\Delta G^{w}(R) = V_{QQ}(R) + \Delta\Delta G^{w}_{sol}(R) + \Delta G^{w}(\infty) \simeq V_{QQ}(R)/\epsilon(R) + \Delta G^{w}(\infty) \quad (23)$$

is the vacuum electrostatic energy (eq 5), $\epsilon(R)$ is the effective dielectric constant for charge–charge interaction, and $\Delta\Delta G^{w}_{sol} = \Delta G_{sol}(R) - \Delta G_{sol}(\infty)$. Since $\epsilon(R)$ is between 30 and 80 for electrostatic interactions in polar solvents[20] for $R \geq 3$ Å, one obtains quite reliable estimates of $\Delta G^{w}_{sol}(R)$. The error associated with the solvation by the protein was subjected to *verification* by comparing the calculated and observed pK_a's of His-57 which gave an error limit of less than 3 kcal/mol. A more systematic study has been conducted recently by calculation the pK_a's of the ionized acids of BPTI.[20b] This study gave an error bound of 5 kcal/mol for the evaluation of solvation energies in proteins.

The sensitivity of the calculations to the residual charges used for the protein groups was also examined in the BPTI test case. It was found that the calculated energies are not extremely sensitive to the parameters used, provided the water molecules around the protein are *included* in the calculations. For sets of smaller charges and smaller $\Delta G_{Q\mu}$ we obtain compensating changes in the contributions of the protein-induced dipoles and the surrounding water molecules ($\Delta G_{Q\alpha}$ and ΔG_{Qw}).

The sensitivity of the calculations to the coordinate used is quite instructive. The same calculations presented in Table VI were performed for α-chymotrypsin, using the old coordinate set of the tosyl–α-chymotrypsin system. For this system we could obtain a reasonable pK_a for His-57 only after minimizing the energy of the system at the native configuration. On the other hand, the trypsin coordinate set gave a reliable pK_a for His-57 without any energy minimization. In general one finds[20b] that the results of the PDLD calculations are insensitive to small geometrical changes. In particular it appears that when the relevant groups are close to the surface the rearrangement of the solvent compensates for changes in the interactions between the protein and the relevant groups. For groups that are not close to the surface, one should either use energy minimization or have the X-ray structure of the protein for the ionized and unionized forms of the relevant groups. In the case of the t^- configuration, we obtain a significant contribution from the nearby water molecules (~30 kcal/mol). This contribution is larger than the corresponding contribution for calculations that allow the oxyanion hole to be polarized toward the negatively charged t^-.[46] In this way the overpolarization of the solvent provides a compensation for using a rigid oxyanion hole. Moreover, the PDLD model which was calibrated by using fixed protein structures may already include some of the effect of the protein reorganization. Further studies that involve a careful assessment of the effect of the protein reorganization are clearly needed, and some effort along this line is now under way.

In view of the above discussion, as well as our experience with related problems, we feel that the present approach has an error limit of about 5 kcal/mol for evaluations of the *absolute* value of the catalytic free energy. However, a smaller error range is expected for studies of closely related systems (e.g., comparing the activation free energies of trypsin and trypsinogen). Thus we expect to obtain quite reliable results in studying the effect of site-directed mutagenesis.[46]

V. Discussion

This work represents a semiquantitative analysis of the energetics of the catalytic reaction of serine proteases. It suggests that the most important catalytic role of the active site is the electrostatic stabilization of the (A^- ImH^+ t^-) resonance form at the transition state. This supports the idea that enzymes can be viewed as "super solvents" that stabilize ionic resonance forms at the transition state.[19] It also demonstrates that the (A^- ImH^+ t^-) system is much more stable than the (AH Im t^-) system. This indicates that the charge-relay mechanism[2,3] (popular, because of the common belief that uncharged intermediates are more stable than charged intermediates) is not an important catalytic factor.

The structural elements that catalyze the reaction by stabilizing the (A^- ImH^+ t^-) resonance form are shown in Figure 7. These elements include the oxyanion hole[3,5,38,39] (the main-chain dipoles of Ser-195 and Gly-193 and water molecules (see Figure 3) that stabilize the negatively charged tetrahedral intermediate) and the dipoles of Ala-56, His-57, and Ser-214 (that stabilize the ionized Asp-102). The calculations presented in Table VI indicate that the stabilization of t^-, by the oxyanion hole and bound water molecules, is the key catalytic factor in serine proteases. This finding is supported by the following points. (i) One of the largest structural changes that occurs upon transition from chymotrypsinogen to chymotrypsin involves residues 191–194 with form the oxyanion hole.[48] This is consistent with the fact that chymotrypsin catalyzes reactions 10^4–10^7 times faster than chymotrypsinogen.[50] A quantitative analysis of this point is now under way in our laboratory. (ii) The spectral red shift of such chromophores as indole–acryloyl that are bound to chymotrypsin (see section V4 of ref 20) can be correlated with the electrostatic interaction between the dipoles of the oxyanion hole and the excited chromophore (negative charge is being transferred to the acryloyl oxygen upon the $\pi \rightarrow \pi^*$ transition). Such a correlation gives clear experimental evidence for the strong electric field created by the oxyanion hole. (iii) The difference in rate of the hydrolysis of L- and D-peptides is around 7 orders of magnitude. This is consistent with the fact that the oxygen of D-peptides cannot align itself properly into the oxyanion hole.

It has been argued recently[13] that trypsin catalyzes amide hydrolysis by providing a nonaqueous environment that makes the reaction similar to the corresponding gas-phase reaction. The problem with this and related proposals can be realized most clearly by inspection of Figure 8; that is, the activation barrier for the gas-phase reaction depends on the selection of the reactant states. The barrier is small only if the reactant state involves an ionized acid or an ionized serine. However, forming the ionized group in the gas phase requires a huge energy. In fact, starting with an ionized group in the gas phase amounts to dealing with an excited-state reaction. Apparently,[34,20] enzymes act by solvating the transition state better (relative to the ground state) than water does. As indicated by Figure 8; the enzyme is not an intermediate between the gas-phase and the solution reaction but rather a better solvent than water or other polar solvents.

If, as argued above, the oxyanion hole and the bound water molecule can account for the major part of the entire rate acceleration, then one has to resolve a major problem: what is the role of the Asp-102? As indicated by this work, the charge-relay mechanism (which requires the Asp His pair) does not provide a major catalytic advantage. Yet, Asp-102 is common to all serine proteases, and it most probably fulfills an important function. In order to examine the role of this group it is important to explore the relative merits of the catalytic triad over alternative catalytic sites. One might consider an alternative site with hydrophobic groups at the Asp-102 side of the histidine. The positively charged histidine will be unstable in this site and will either be pulled toward the solvent or force a small local unfolding and penetration of water molecule to the site. Both options will involve significant solvent reorganization energy and increase the activation barrier for the proton-transfer step. If the site will be sufficiently rigid to prevent penetration of water, then the pK_a^P of the histidine in the (ImH^+ Ser^-) configuration will be reduced, increasing the energy of ψ_2 and slowing the reaction.

The stabilizing role of Asp-102 could have been provided by protein dipoles. However, it might be simpler for the protein to stabilize by its dipoles the ionized Asp-102 which in turn stabilizes ImH^+. Note, in this respect, that the (Asp^- ImH^+ t^-) system which is stabilized by the protein hydrogen bonds is a (− + −) system of the type considered in Figure 5 and Figure 2 of ref 34. Such a system can be stabilized by the protein dipoles without investing a large amount of folding energy.

A possible advantage of Asp-102 might be associated with fixing His-57 and increasing the *specificity* of the enzyme. That is, only certain orientations of the bound substrate which are not possible for some substrates (e.g., D-amino acids) are expected to allow for simultaneous stabilization of the O^- of the tetrahedral intermediate (which is fixed by its bond to Ser-195) and for proton transfer from His-57 to X. Nonspecific substrates which cannot

be catalyzed by the enzyme cannot be catalyzed by OH^- and H_3O^+ since the bound His-57 prevents penetration of these ions. The specificity associated with a relatively fixed His-57 might be even more important if the last step in eq 2 is the rate-limiting step. In this case the transition state would be stabilized simultaneously by both the oxyanion hole and the nitrogen of His-57.

It should also be noted that the fixed His-57 gives the option of slowing the reaction by lowering the external pH. This might be important for some biological functions of serine protease or related enzymes.

The transition state of the reaction of serine proteases is usually described as a tetrahedral intermediate rather than a negatively charged intermediate. It seems to us that the tetrahedral geometry is a simple consequence of the sp^3 bonding arrangement, which does not provide any significant catalytic advantage. That is, the tetrahedral geometry is expected to be the same in solution and in the enzyme since the very small atomic displacements associated with the transition of the carbonyl carbon of the substrate from the sp^2 to the sp^3 hybridization can be accommodated easily by a small relaxation of the enzyme (see in ref 16 a related discussion about strain in lysozyme).

The EVB method provides a powerful insight into the factors that control the transition from the planar to the tetrahedral geometry. Since the substrate tends to be planar in the (Im H—O C=O) resonance form and tetrahedral in the (ImH^+ t^-) resonance form, one can use the relative energies of these resonance forms in determining the substrate geometry. Thus we find by inspection of Figure 5 that as long as the O---C (b_4) bond length is greater than 2.4 Å, the (ImH^+ O^- C=O) resonance form and the (Im H—O C=O) resonance form are more stable than the (ImH^+ t^-) resonance form. In this case the substrate geometry can be evaluated by using the force constants of V_{strain} (eq 12) of the (Im H—O C=O) resonance form. Since the out-of-plane force constant K_χ of the carbonyl system is large, it is not obvious how a trypsin molecule can deform significantly the C=O angle of a bound pancreatic trypsin inhibitor (PTI), toward the sp^3 geometry, as long as the O···C=O distance is larger than 2.4 Å. Thus the observation of a tetrahedral deformation of the C=O of Lys-15 in the trypsin–PTI complex[32] is unlikely to be associated with the proximity of the serine oxygen, which is observed to be 2.6 Å from the carbonyl carbon. It is still possible that the actual deformation is smaller than that estimated by the X-ray study. It is also possible that the protein bends the C—C—N angle, which is easier to deform (by pushing long chains) than the C=O out-of-plane angle. This interesting problem should be explored by actual EVB calculations of the equilibrium geometry of the trypsin–PTI complex.

In view of the length of the above discussion it might be useful to summarize the main conclusions of the present work. (1) The most important catalytic factor in serine proteases is the electrostatic stabilization of t^- by the oxyanion hole. (2) The stabilization by the oxyanion hole includes interaction between bound water molecules and t^-. Removal of these water molecules (as is the case in the trypsin–PTI complex) might slow the rate significantly. (3) The charge-relay mechanism is not an important catalytic factor. (4) The stabilization of the transition state by Asp-101 can be provided in part by polar groups or water. The presence of Asp-102 might, however, be important for fixing His-57 (which is important for increasing the specificity of the system). (5) The geometry around the carbonyl of Lys-15 in the trypsin–PTI complex is not likely to be affected significantly by the presence of the serine oxygen at a distance larger than 2.5 Å.

Acknowledgment. This work was supported by Grant GM-24492 from the National Institutes of Health.

Appendix. Evaluation of the Mixing Terms

The discussion in this paper is simplified considerably by describing the reaction in terms of resonance forms that involve real bonds (eq 3) rather than in terms of the pure ionic and covalent resonance forms of the type used in ref 18. The relation between the two representations is described below for a (B H—O C=O) system (where B is the histidine base).

The zero-order basis set of the (B H—O C=O) system is composed of pure covalent and ionic resonance forms given in eq 24. The gas-phase Hamiltonians of these resonance forms are constructed according to the procedure of ref 18. The diagonal

$$\phi_a^0 = (N\ H—O\ C=O) \tag{24a}$$

$$\phi_b^0 = (N\ H^+\ O^-\ C=O) \tag{24b}$$

$$\phi_c^0 = (N^+—H\ O^-\ C=O) \tag{24c}$$

$$\phi_d^0 = (N^+—H\ O—C—O^-) \tag{24d}$$

$$\phi_e^0 = (N^+—H\ O^-\ C^+—O^-) \tag{24e}$$

$$\phi_f^0 = (N\ H—O\ C^+—O^-) \tag{24f}$$

$$\phi_g^0 = (N\ H^+O—C—O^-) \tag{24g}$$

$$\phi_h^0 = (N\ H^+O^-\ C^+—O^-) \tag{24h}$$

elements are given by eq 25, where g indicates the gas phase, $\Delta\bar{M}$ is the pure covalent Morse potential for the bonds (b_i) in the ν'th resonance form, $V_{nb}^{(\nu)}$ and $V_{QQ}^{(\nu)}$ are the nonbonded and electrostatic interactions of the ν'th resonance form, and α^ν is the energy of forming this resonance form in the gas phase at infinite separation between the fragments. The corresponding energy parameters

$$H_{\nu\nu}^g = \alpha^\nu + \sum_{i(\nu)} \Delta\bar{M}(b_i) + V_{nb}^{(\nu)} + V_{QQ}^{(\nu)} \tag{25}$$

are given in Table I. The off-diagonal matrix elements are evaluated as follows:[18] When the resonance structure ν is formed from the μ'th resonance structure by breaking the bond X–Y and forming the state $X^+ + Y^-$, we approximate $H_{\mu\nu}$ by eq 26, where

$$H_{\mu\nu} = L_{\mu\nu} = \{[\bar{M}(b) - M(b)][E_\infty^\nu(b) - E_\infty^\mu(b) - \bar{D} - M(b)]\} \tag{26}$$

$\bar{M}$ is the Morse potential of the pure covalent bond and E_∞^ν is the energy of the ν'th resonance structure when all of its fragments except the one with the given bond are at infinite separation. This forces the bond-stretching potential obtained from the mixing of ν and μ to reproduce the corresponding observed Morse potential.[18] If two resonance forms (μ and ν) differ by the presence of more than one bond, $H_{\mu\nu}$ is set to 0, except for H_{ac} which is given here by eq 27,[18] where S_{NH} is the overlap integral between the 1s orbital

$$H_{ac} = S_{NH}H_{ab} \tag{27}$$

of the hydrogen and the 2p orbital of the nitrogen. Note that we are not trying to describe all the resonance states within the fragment B but to treat it as an effective atom.

In order to simplify the discussion and the parametrization, we transform the zero-order pure states to the basis set of eq 3 including the mixing with the high-energy states by a perturbation treatment by using eq 28, where N are normalization constants

$$\psi_1 = (\phi_a^0 + \sum_\mu{}' \lambda_{1\mu}\phi_p^\mu)/N_1 \tag{28a}$$

$$\psi_2 = (\phi_c^0 + \sum_\mu{}' \lambda_{2\mu}\phi_\mu^0)/N_2 \tag{28b}$$

$$\psi_3 = (\phi_d^0 + \sum_\mu{}' \lambda_{3\mu}\phi_\mu^0)/N_3 \tag{28c}$$

and ψ_1, ψ_2, and ψ_3 represent resonance forms (with a small polar character) rather than pure ionic and covalent states. The λ_i coefficients are given by eq 29, where the index α is the first index

$$\lambda_{i\mu} = H_{i\mu}^g/(H_{\alpha\alpha}^g - H_{\mu\mu}^g) \tag{29}$$

in the expression for ψ_1 (e.g., a for ψ_1). The off-diagonal matrix elements of the simplified Hamiltonian are obtained by using the expansion coefficients of ψ_i to transform the Hamiltonian. This gives eq 30. The atomic charges (Q') of the new resonance forms

$$\bar{H}_{ij}^g = \sum_{\mu\nu} \lambda_{i\nu}\lambda_{j\mu}H_{\mu\nu}^g/(N_iN_j) \tag{30}$$

can be obtained from the charges of the pure resonance structures

(Q^0) by transformation 31, where m designates the mth atom.

$$Q'_{im} = \sum_{\mu} Q_{\mu m}{}^{0}(\lambda_{im}{}^{2}/N_i{}^{2}) \quad (31)$$

Alternatively, as is done in the present work, one can take the charges Q' from quantum mechanical calculations of the relevant isolated fragments.

The EVB Hamiltonian for the gas-phase reaction is now given by eq 32, where E_i are the energies of the mixed states. The Hamiltonian for the solution reaction is given by eq 33. In the

$$\bar{H}^{g} = \begin{pmatrix} E_1{}^{g} & \bar{H}_{12}{}^{g} & 0 \\ \bar{H}_{12}{}^{g} & E_2{}^{g} & \bar{H}_{23}{}^{g} \\ 0 & \bar{H}_{23}{}^{g} & E_3{}^{g} \end{pmatrix} \quad (32)$$

$$\bar{H}^{w} = \begin{pmatrix} E_1{}^{g} + G_{sol}^{(1)} & \bar{H}_{12}{}^{g} & 0 \\ \bar{H}_{12}{}^{g} & E_2{}^{g} + G_{sol}^{(2)} & \bar{H}_{23}{}^{g} \\ 0 & \bar{H}_{23}{}^{g} & E_3{}^{g} + G_{sol}^{(3)} \end{pmatrix} \quad (33)$$

present approximation we do not introduce the solvent effects in the off-diagonal elements. In this way the ground-state energies obtained by solving the Hamiltonian of the pure states and the restricted Hamiltonian of eq 33 are slightly different. However, since we parametrize the restricted Hamiltonian by choosing the α_i from solution reactions (eq 9), we can use it as a reliable tool for comparing reactions in solutions and in enzymes.

Registry No. L-His, 71-00-1; L-Asp, 56-84-8; L-Ser, 56-45-1; trypsin, 9002-07-7; serine proteinase, 37259-58-8.

Synthesis, Crystal Structure, and Absorption Spectroscopy of a Carbazole–Dinitrofluorene Bichromophore. Identity of Ground-State Molecular Interactions in the Single Crystal and Aggregates in Solution

François Brisse,* Gilles Durocher,* Sylvain Gauthier, Denis Gravel,* Rolande Marquès, Caroll Vergelati, and Bogumil Zelent

Contribution from the Département de Chimie, Université de Montréal, Montréal, Québec, Canada H3C 3J7. *Received October 8, 1985*

Abstract: The synthesis, crystal structure, conformational analysis, and absorption spectroscopy of the bichromophore 9-(γ-(*N*-carbazolyl)propyl)-9-methyl-2,7-dinitrofluorene (**1**) are reported. The torsion angles within the propyl part of the molecule correspond to a gauche–trans–gauche–gauche conformation while the dinitrofluorene and the carbazole groups of atoms are nearly planar and are at 25.6° from each other. The observed conformational behavior has been analyzed with the aid of a conformational analysis program based on empirical methods, and a good agreement between calculated and crystallographic minima has been found. It is shown by the comparison of spectra in polar and nonpolar solvents that no ground-state intramolecular interaction takes place in the bichromophore. All observed spectroscopic changes are due to solute–solute intermolecular interactions or to solute–solvent intermolecular interactions. A new band appearing at 370 nm in a solution of the bichromophore **1** in 3-methylpentane at high concentrations or at low temperatures is assigned to aggregates by comparison with the spectrum of a single crystal. Furthermore, single-crystal data allow evaluation of the energetics of the ground and first excited electronic states of the crystal. The exciton splitting is calculated as being 887 cm^{-1}, but this is the sum of the dipole–dipole interactions along with the charge-resonance interactions in the first excited singlet state which is mainly responsible for the 2380-cm^{-1} bathochromic shift observed in comparing the spectrum of the free molecule with that of the crystal.

The chemistry, spectroscopy, and structural analysis of bichromophoric compounds[1] continue to stimulate much interest in view of the fundamental objectives being pursued. Some recent examples range from elucidating the mechanism of photochemical reactions,[2-5] of electronic energy transfer,[6-9] or of electron transfer[10-13] to eventually establishing the structure and spatial arrangement of interacting chromophores in the ground and excited states.[14,15]

It is clear that the acquisition of basic knowledge concerning these systems can accelerate the solution of important fundamental problems such as the mechanism of the respiratory chain and the photosynthetic process[11] as well as favor the development of photoconducting polymers[16] and molecular electronics.[12,17]

(1) (a) De Schryver, F. C.; Boens, N.; Put, J. *Adv. Photochem.* **1977**, *10*, 359. (b) Morrison, H. *Acc. Chem. Res.* **1979**, *12*, 383.
(2) Holden, D. A.; Gray, J. B.; McEwan, I. *J. Org. Chem.* **1985**, *50*, 866.
(3) Desvergne, J.-P.; Bitit, N.; Castellan, A.; Bouas-Laurent, H. *J. Chem. Soc., Perkin Trans. 2* **1983**, 109.
(4) McCullough, J. J.; MacInnis, W. K.; Lock, C. J. L.; Faggiani, R. *J. Am. Chem. Soc.* **1982**, *104*, 4644.
(5) (a) McIntosh, A. R.; Siemiarczuk, A.; Bolton, J. R.; Stillman, M. J.; Ho, T.-F.; Weedon, A. C. *J. Am. Chem. Soc.* **1983**, *105*, 7215. (b) Siemiarczuk, A.; McIntosh, A. R.; Ho, T.-F.; Stillman, M. J.; Roach, K. J.; Weedon, A. C.; Bolton, J. R.; Connolly, J. S. *J. Am. Chem. Soc.* **1983**, *105*, 7224.
(6) Hassoon, S.; Lustig, H.; Rubin, M. B.; Speiser, S. *J. Phys. Chem.* **1984**, *88*, 6367.
(7) Ebata, T.; Suzuki, Y.; Mikami, N.; Miyashi, T.; Ito, M. *Chem. Phys. Lett.* **1984**, *110*, 597.
(8) Hassoon, S.; Lustig (Richter), H.; Rubin, M. B.; Speiser, S. *Chem. Phys. Lett.* **1983**, *98*, 345.
(9) Speiser, S.; Katriel, J. *Chem. Phys. Lett.* **1983**, *102*, 88.
(10) Mes, G. F.; de Jong, B.; van Ramesdonk, H. J.; Verhoeven, J. W.; Warman, J. M.; de Hass, M. P.; Horsman-van den Dool, L. E. W. *J. Am. Chem. Soc.* **1984**, *106*, 6524.
(11) Mes, G. F.; van Ramesdonk, H. J.; Verhoeven, J. W. *J. Am. Chem. Soc.* **1984**, *106*, 1335.
(12) Mes, G. F.; van Ramesdonk, H. J.; Verhoeven, J. W. *Recl.: J. R. Neth. Chem. Soc.* **1983**, *102*, 55.
(13) Pasman, P.; Koper, N. W.; Verhoeven, J. W. *Recl.: J. R. Neth. Chem. Soc.* **1982**, *101*, 363.
(14) Masuhara, H. *J. Mol. Struct.* **1985**, *126*, 145.
(15) Itoh, M.; Fujiwara, Y. *Bull. Chem. Soc. Jpn.* **1984**, *57*, 2261.
(16) *Electrical Properties of Polymers*; Seanor, D. A., Ed.; Academic: New York, 1982.
(17) Aviram, A.; Ratner, M. A. *Chem. Phys. Lett.* **1974**, *29*, 277.

0002-7863/86/1508-6579$01.50/0