Estimating the Degree of Expansion in the Transition State for Protein Unfolding: Analysis of the pH Dependence of the Rate Constant for Caricain Denaturation†

Leticia López-Arenas, Silvia Solís-Mendiola, and Andrés Hernández-Arana*

Area de Biofisicoquímica, Departamento de Química, Universidad Autónoma Metropolitana-Iztapalapa, Apartado Postal 55−534, Iztapalapa D.F. 09340, México

Received July 19, 1999; Revised Manuscript Received September 15, 1999

ABSTRACT: The thermal denaturation of caricain (the most alkaline of papain-related proteinases) was studied in acid media. Under all conditions tested, caricain denatured irreversibly following a single first-order reaction that involves simultaneous loss of secondary and tertiary structures. Besides, variation of the rate constant with temperature gave linear Eyring's plots. Thus, despite its irreversibility, this process resembles the kinetics of reversible protein unfolding. Due to the basicity of caricain, all of the carboxylates in the native protein interact with nearby positively charged groups. Then, it may be thought that p*K* values of titratable sites are mainly influenced by interactions of this type. Accordingly, we set up a simple electrostatic perturbation model, based on charge−charge interactions at distances not larger than 10 Å, which reproduces reasonably well the titration curve of native caricain. Because the pH dependence of the activation free energy for unfolding ($\Delta G^{\ddagger}$) can be related to differences in the protonation behavior of the native (**N**) state with respect to the transition (**TS**) state, the model was further used to analyze the experimental $\Delta G^{\ddagger}$ vs pH curve. Results from this analysis suggest that there is an increase of about 3 Å in the average ion-pair distance when **N** globally expands to form **TS**. Alternatively, if the expansion were restricted to only one molecular domain, the structure of this domain in **TS** would be highly disordered. In either case, it is probable that the solvent-accessible area augments significantly during the expansion.

The role of ion pairs in the stability of proteins has recently been the subject of numerous investigations. Theoretical calculations have led to the proposal that electrostatic interactions produce a net destabilization of native proteins (*1*, *2*). More specifically, calculations by Hendsch and Tidor (*2*) suggest that substitution of both charged partners in a salt bridge for their hydrophobic isosteres (hypothetical hydrophobic side chains of the same shape and size) can result in proteins with increased stability. That is, on the average, an ion pair destabilizes the folded structure by 14.6 kJ mol^{-1}. In contrast, experiments of single and double mutation of particular ion pairs in T4 lysozyme and barnase indicate that this type of interactions actually stabilize the native state by 4 to17 kJ mol^{-1} (*3*, *4*). Moreover, in some cases, the contribution of an ion pair to the stability agrees well with that expected from the observed shifts in p*K* of the intervening ionizable groups (*3*, *5*).

What has long been recognized is that the dependence of stability on pH is a reflection of differences in the p*K* of ionizable groups between the native (**N**) and unfolded states (*3*, *5*−*7*). Oliveberg and Fersht (*8*) have extended this idea to the study of the kinetic stability of barnase, obtaining qualitative information about the structures of transient conformations in the folding pathway. Their results suggest that in the major transition state (**TS**) some carboxylate groups retain the highly anomalous p*K* values characteristic of the native protein; however, some electrostatic interactions are lost, predominantly from the molecular surface. These observations seem to be consistent with a representation of **TS** as an expanded but nativelike structure, in agreement with previous ideas (*9*−*11*). Recently (*12*), it has been reported that the effect of pH on the kinetics of chymopapain denaturation is consistent with large p*K* shifts for carboxylates during the transition from **N** to **TS**. Because most of the carboxyl groups in this protein are involved in ion pairs, it was proposed that these interactions impart significant kinetic stability to the native molecule; i.e., that the free-energy change necessary to weaken or disrupt ion pairs rises the free-energy barrier for unfolding. A similar stabilizing role has been ascribed to ion pairs in the irreversible unfolding of rubredoxin from the hyperthermophile *Pyrococcus furiosus* (*13*).

In the present work, we propose a more detailed approach to the analysis of pH effects on the kinetic stability of proteins. By means of this approach, it is possible to obtain quantitative information on the expansion experienced by the macromolecule when its transition state for unfolding is reached. The protein used in this study was caricain, which is the most basic of all papain-related proteinases. Its three-dimensional structure reveals that all its acidic side chains are engaged in multiple interactions with positively charged groups. As has previously been found for other members of this enzyme family (*12*, *14*), caricain denatured irreversibly according to a first-order reaction over wide ranges of pH and temperature. This process encompasses the loss of both secondary and tertiary structures. It is shown that variation of the activation free energy for unfolding with pH (at

† This work was supported in part by CONACYT, México (convenio no. 473100-5-3771N).

* To whom correspondence should be addressed.

10.1021/bi991658w CCC: $18.00
Published on Web 11/05/1999

constant temperature) is compatible with an important expansion of ion-pair distances in the transition state.

EXPERIMENTAL PROCEDURES

Materials. Partially purified chymopapain was purchased from Sigma (St. Louis, MO). The most abundant enzyme form of caricain contained in this preparation was separated and purified to homogeneity by cation-exchange liquid chromatography as described previously (*15*). To avoid the risk of autolysis during denaturation, the active-site thiol group in caricain was irreversibly blocked with iodoacetamide (*16*). Protein concentration was determined by means of the specific absorbance, $A_{278nm} = 1.83$ mL mg^{-1} cm^{-1} (*17*). All other chemicals were of the highest purity grade.

Circular Dichroism. Circular dichroism (CD)[1] measurements were recorded in JASCO J-500 or J-715 spectropolarimeters calibrated with (+)-10-camphor sulfonic acid (*18*). The J-500 instrument was equipped with a water-jacketed cell holder for temperature control. The J-715 used a Peltier-type temperature control system. Actual temperatures within the cell were measured with a thermistor probe. Protein solutions were prepared in 0.05 M glycine-HCl buffer adjusted to the desired pH. Ionic strength was kept constant at 0.1 M by addition of NaCl. Protein concentration was approximately 0.15 mg mL^{-1} in the far ultraviolet (200−250 nm) and ca. 0.5 mg mL^{-1} in the aromatic absorption region (250−320 nm). Spectral data are expressed as mean residue ellipticities, [θ]. The thermal denaturation transition of caricain was monitored by the ellipticity change, at fixed wavelength, as a function of temperature. Temperature inside the cell was continuously varied at selected heating rates in the range (0.1−1.5) ± 0.05 °C min^{-1}.

Denaturation Kinetics. The time course of caricain denaturation was monitored by the change in ellipticity at selected wavelengths. Cells, filled up to 95−99% of its volume with buffer solution, were placed in the spectropolarimeter and allowed to equilibrate at the temperature of the experiment. A concentrated solution of caricain was then injected to complete the cell volume; the sample was injected, employing long-needle syringes to facilitate homogenization of solutions inside the cell. In conventional cylindrical cells, complete mixing and thermal equilibrium were reached in approximately 3 min, after which recording of the CD signal was started. For the fastest reactions, the Peltier system with magnetic stirrer was used; in this case, the dead time of experiments was reduced to less than 10 s. Kinetic data were fitted to exponential decay curves of the form $\theta_t = \theta_f + (\theta_0 - \theta_f) \exp(-kt)$, where θ_t is the ellipticity measured at time t, θ_f is the final ellipticity value, and θ_0 represents the corresponding value at zero time.

Binding of Protons to Native Caricain. The number of protons bound to the macromolecule was determined from potentiometric titrations according to procedures described elsewhere (*19*).

RESULTS

Spectral Features of Native and Heat-Denatured Caricain. Figure 1 shows the CD spectra of native and heat-denatured caricain at pH 3.9 and 2.0. As can be noted from spectra in the far ultraviolet, at 25 °C, the secondary structure of the native protein is fairly well conserved within the studied pH region. On the other hand, in the aromatic region the spectrum at pH 2.0 lost some magnitude with respect to that at pH 3.9. Spectra taken at various pH values (not shown) indicated that this loss of intensity in the aromatic CD signal of native caricain appeared below pH 2.5. Spectral characteristics of the heat-denatured protein were found to be rather insensitive to pH variation. Loss of tertiary structure upon denaturation is reflected in the large decrease in magnitude of the ellipticity in the aromatic-absorption region; spectral changes in the far ultraviolet are, on the other hand, attributable to the loss of secondary structure. It is important to mention that in this latter spectral region the shape and magnitude of the caricain spectrum at high temperature are similar to those of other thermally unfolded proteins (*20*−*22*). Furthermore, we have observed that the thermally unfolded forms of lysozyme and trypsin display similar spectral shapes, i.e., a negative broad shoulder around 220 nm and a negative extremum ca. 200 nm (results not shown).

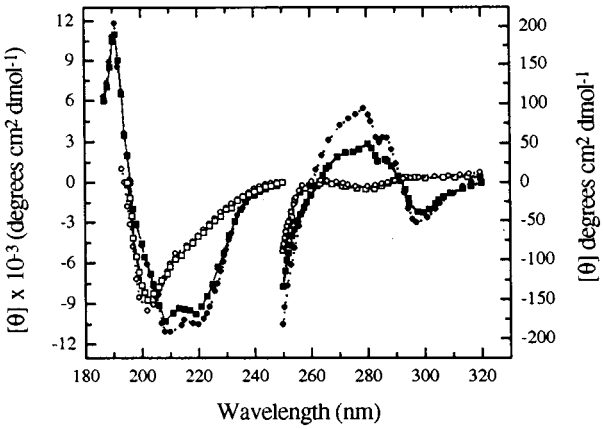

FIGURE 1: CD spectra of native and heat-denatured caricain. Spectra were recorded in 0.05 M glycine-HCl buffers (ionic strength, 0.1 M) under different conditions of pH and temperature: pH 3.9 (dotted lines), at 25 °C (●) and 80 °C (○); pH 2.0 (solid lines), at 25 °C (■), and 70 °C (□).

Characteristics of the Thermal Transition. The thermal denaturation of caricain at pH 2.5 was monitored continuously at 220 nm, using different heating rates (Figure 2). In this figure, f_D is the fraction of denatured protein, which was calculated according to the relationship

$$f_D = (\theta - \theta_N)/(\theta_D - \theta_N) \quad (1)$$

where θ is the sample ellipticity at a particular temperature, θ_D and θ_N are the values of θ for the denatured and native states, extrapolated to the same temperature. As can be seen in Figure 2, the appearance of the thermal transition is strongly influenced by the heating rate, which is typical for processes under kinetic control (*14*, *23*). When denaturation was followed at 279 nm, the observed transition curve was practically superimposable on that obtained at 220 nm at the same heating rate (data not shown). This indicates that caricain denaturation is a global process involving simultaneous losses of secondary and tertiary structure. This process was completely irreversible as judged from the failure to observe a reversal of ellipticity changes when the sample was cooled to 25 °C. Moreover, we found that rapid cooling (∼15 °C min^{-1}) of caricain at the middle of the transition

[1] Abbreviations: CD, circular dichroism.

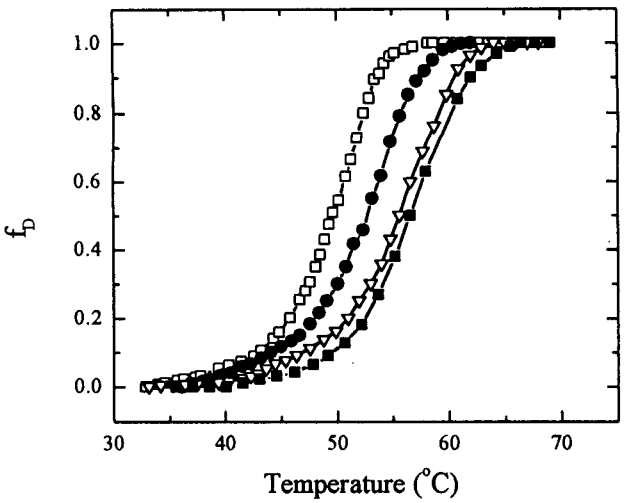

FIGURE 2: Thermal transition curves for caricain, obtained at different heating rates: 1.5 (■), 1.0 (▽), 0.3 (●), and 0.1 °C/min (□). The fraction of denatured protein was calculated according to eq 1 from ellipticity measurements at 220 nm.

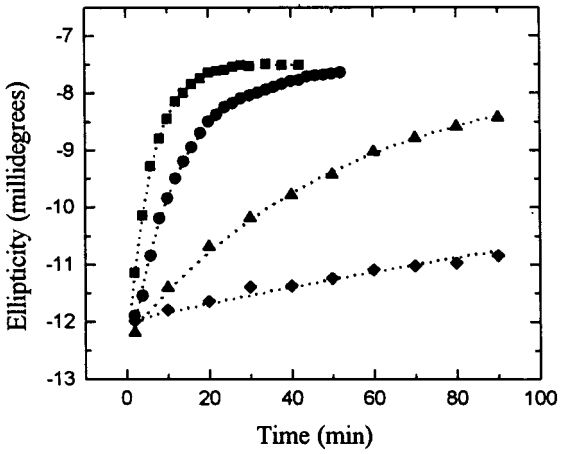

FIGURE 3: Time course of the thermal denaturation of native caricain (pH 2.5) at several temperatures. Data shown correspond to the following temperatures: 55.9 (■), 51 (●), 47 (▲), and 44 °C (◆). Ellipticity was measured at 220 nm. Lines are single-exponential decay curves fitted to experimental data.

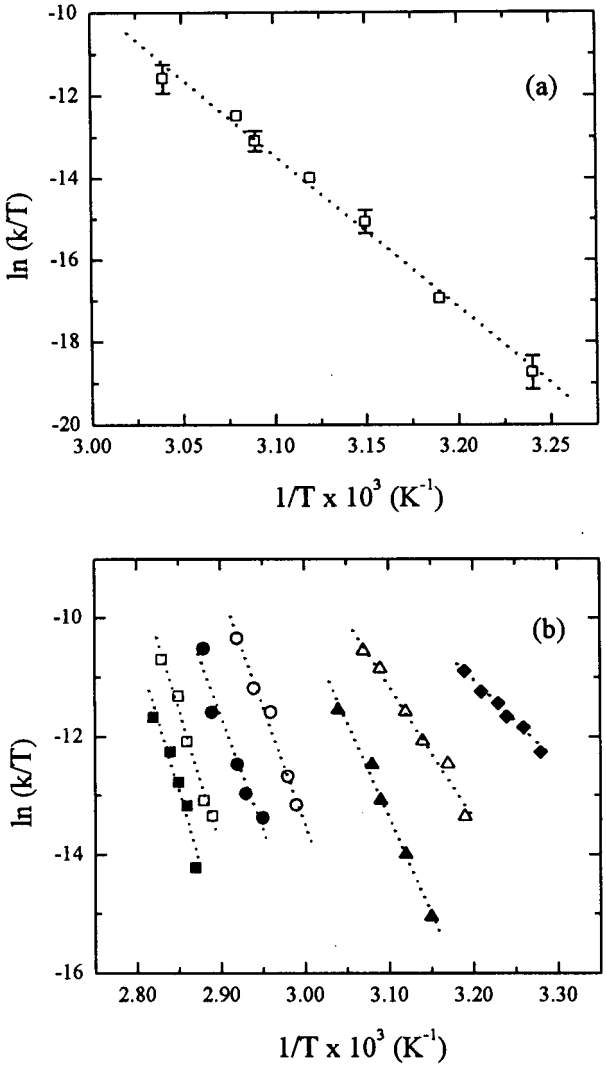

FIGURE 4: Eyring's plots for the denaturation rate-constant of caricain. Data shown in panel a correspond to pH 2.5, and cover a temperature range of 20 °C. Shown in panel b are data obtained at different pH values: 3.9 (■), 3.5 (□), 3.2 (●), 2.9 (○), 2.5 (▲), 2.2 (△), and 2.0 (◆). Lines are linear least-squares regressions to experimental data.

produced no return of the CD signal toward the characteristic ellipticity of the native protein. That is, no intermediates capable to refold could be detected, indicating that caricain denaturation can be described as a simple two-state, irreversible process,

$$N \xrightarrow{k} D \qquad (2)$$

The lack of reversibility is apparently not due to aggregation of the denatured protein because no significant increase in light scattering was observed upon denaturation. In addition, electrophoretic analysis indicated that the polypeptide chain suffered no fragmentation when heated to high temperatures.

Denaturation Kinetics. The kinetics of caricain denaturation at constant temperature was monitored by CD at 220 or 279 nm. Some typical results obtained at pH 2.5 are illustrated in Figure 3, where it can be seen that the loss of secondary structure follows a simple first-order reaction. Extrapolation of data to time zero gave ellipticities very similar to those expected for the native protein at the same temperature. This indicates that no fast kinetic phase occurs during the mixing dead time of experiments. Changes in tertiary structure (observed at 279 nm) followed a time course that paralleled the loss of secondary structure, i.e., rate constants determined by monitoring ellipticities at 220 and 279 nm were essentially identical. In all cases, the changes observed were irreversible. Kinetic experiments performed at different pH values always gave irreversible, single-exponential decay curves such as those in Figure 3. Thus, all the evidence obtained from kinetic studies suggests that caricain denatures according to the simple two-state model **N** → **D**, in agreement with the above proposal.

The effect of temperature on the rate constant of denaturation was investigated at various pH values. Results obtained at pH 2.5 are shown in Figure 4a in the form of an Eyring's plot, which is based on the well-known equation

$$\ln(k/T) = \ln(k_B/\hbar) - \Delta G^{\ddagger}/RT \qquad (3)$$

or its alternative form,

$$\ln(k/T) = \ln(k_B/\hbar) + \Delta S^{\ddagger}/R - \Delta H^{\ddagger}/RT \qquad (4)$$

Table 1: Activation Enthalpy, $\Delta H^{\ddagger}$, for the Denaturation Reaction of Caricain at Various pH Values[a]

pH	$\Delta H^{\ddagger}$ (kJ/mol)
3.9	400
3.5	388
3.2	335
2.9	332
2.5	300
2.2	186
2.0	109

[a] $\Delta H^{\ddagger}$ values were calculated from the slopes of lines shown in Figure 4.

where k_B is Boltzmann's constant, $\hbar$ is Planck's constant, and R is the gas constant; $\Delta G^{\ddagger}$, $\Delta H^{\ddagger}$, and $\Delta S^{\ddagger}$ represent the activation free energy, enthalpy, and entropy, respectively. From the slope of the line in Figure 4a, $\Delta H^{\ddagger}$ is calculated as 300 kJ mol^{-1}. It is worth to note the linearity of data displayed in Figure 4a. This data set was obtained over a broad temperature range of 20 °C and encompasses a variation of the rate constant by 3 orders of magnitude. At the lowest temperature (35.5 °C), decay curves had a relaxation time of approximately 5 days, yet they showed reproducible, single-exponential behavior. The linear dependence of ln(k/T) with 1/T indicates that the heat capacity change upon activation is negligible. Similar observations have been reported before for the reversible unfolding of globular proteins (*8*, *9*, *24*−*27*), albeit in this later case studies usually cover a narrow temperature range due to limitations imposed by the reversible nature of the unfolding reaction. Figure 4b shows Eyring's plots corresponding to caricain denaturation at several pH values; there is a noticeable change in the slope of the lines, indicating a large variation of $\Delta H^{\ddagger}$ with pH, especially below pH 2.5 (see Table 1).

Binding of Protons to Native Caricain. Changes in pH directly affect the ionization state of protein molecules. In turn, particular protein differences in the ionization behavior of the transition state with respect to the native state may be considered as the main cause for the variation of its kinetic stability with pH. Thus, we investigated the titration properties of native caricain over a pH region comparable to that of our kinetic studies. Results (Figure 5) showed that at 25 °C about nine protons are bound per protein molecule in the pH range 4.75−2.0. It was not possible to extend titration studies to pH values lower than 2.0 due to the unstableness of the native protein.

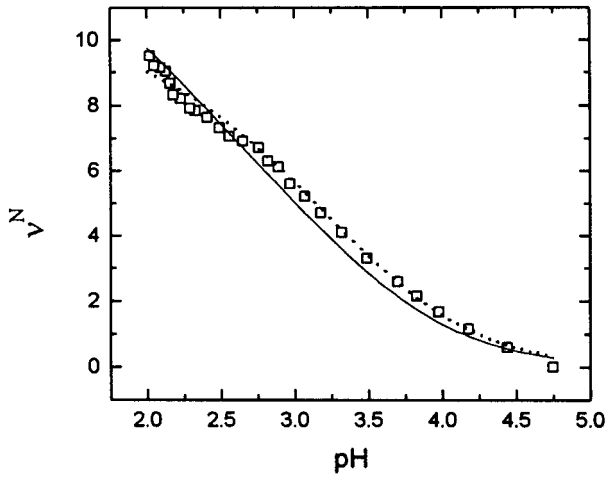

FIGURE 5: Protonation curve of native caricain in water (ionic strength, 0.1 M) at 25 °C. The ordinate represents the number of protons bound per protein molecule, with pH 4.75 taken as the reference state. The solid curve was calculated according to eqs 6−10, under the assumption that there is a unique effective dielectric constant ($D_{mol} = 59$) in the protein interior. Results obtained from calculations that assumed two different dielectric constants are also shown (dotted curve).

DISCUSSION

Results presented in this work indicate that the thermal denaturation of caricain is an irreversible process, as has been reported for other papain-related proteinases (*12*, *14*). It has been proposed that this behavior could be a reflection of the metastable character of the native state in cysteine proteinases (*12*). It must be recalled that what we call native enzyme is actually the mature form derived from the corresponding zymogen by scission of a large proregion of approximately 100 residues. Because the proregion seems to be essential for correct folding of the macromolecule (*28*, *29*), it is likely that the polypeptide chain, once devoid of the proregion, would be unable to refold. In this regard, it should be mentioned that for α-lytic protease the metastability of the mature enzyme form is well documented (*30*); this bacterial serine protease is also synthesized as a precursor with a large proregion which promotes folding of the protease domain (*31*).

As judged from CD spectra, the native, secondary structure of caricain varies little within the studied pH region. Likewise, tertiary structure is well conserved down to pH 2.5. Below this pH, there seems to be a gradual *loosening* of the structure around aromatic residues, which is reflected in the loss of CD intensity. The spectral features of the denatured protein remain practically constant from pH 2.0 to 3.9. It may be thought that the broad shoulder around 220 nm is indicative of some *residual* structure in denatured caricain. Indeed, there is still a controversy regarding the degree of unfolding attained in a thermally denatured protein in comparison with the completely unfolded state existing at high concentrations of denaturing agents (*32*, *33*). It should be remembered, however, that proteins with different folding patterns in their native states exhibit remarkably similar dichroic spectra after thermal denaturation. This argues against the presence of *residual* native structure as the cause of the CD signal at 220 nm. More likely, this spectral signal could be caused by a general property of thermally unfolded proteins such as their high degree of compactness in the absence of chemical denaturants (*33*).

Despite its irreversible character, the kinetics of caricain denaturation resembles that for the reversible unfolding of proteins (in aqueous solution) in two main aspects. First, the time course of denaturation is consistent with a simple two-state process in which secondary and tertiary structures are lost simultaneously. Second, the variation of the rate constant with temperature gives nearly linear Eyring's plots, indicating that $\Delta H^{\ddagger}$ varies little with temperature. This evidence seems to suggest that in caricain the main transition state (or the ensemble of structures representing this state), **TS**, is structurally closer to the native state, as it is thought to occur in other globular proteins (*5*, *9*, *24*−*25*). Thus, kinetic studies on the denaturation of caricain and other cysteine proteinases

represent a good opportunity to explore some characteristics of **TS**, because the kinetics of unfolding can be measured without complications arising from the reverse, folding reaction. The remaining of this discussion focuses on the effect of pH on $\Delta G^{\ddagger}$, attempting to gain some structural information about the transition state.

The variation of $\Delta G^{\ddagger}$ with pH is easily expressed in terms of differences in the number of protons bound to **TS**, ν^{TS}, and to **N**, ν^{N} (*8*, *12*):

$$-(1/RT)(\partial \Delta G^{\ddagger}/\partial \ln a_H) = \nu^{TS} - \nu^{N} \quad (5)$$

where a_H is the activity of protons. ν^{TS} and ν^{N} are both dependent on a_H and, thus, represent protonation functions for particular macromolecular states. Of course, only ν^{N} is amenable to experimental determination (see Figure 5). On the other hand, ν^{N} can be calculated if one knows the protonation constant, K_i^N, for each of the titratable groups in the native protein; in turn, K_i^N is related to the protonation free energy, $\Delta G_i^N = -RT \ln K_i^N$. Conceptually, ΔG_i^N can be considered as made up of two different contributions: the intrinsic protonation free energy [$\Delta G_{int} = -RT(2.303)pK_{int}$] of a titratable group forming part of a polypeptide chain, but in the absence of any electrostatic perturbation, and the electrostatic perturbation of the group by the molecular environment. Two main perturbation effects usually considered are the hydration energy (or self-energy) of an ion, and the Coulombic energy due to interactions with other charges in the macromolecule (*34*). Thus, for the protonation of a carboxylate group, ΔG_i^N (kJ mol^{-1}) can be expressed as

$$\Delta G_i^N = \Delta G_{int} - \frac{1388 Z_i^2}{2 r_i}\left(\frac{1}{D_{mol}} - \frac{1}{D_{ref}}\right) - 1388 Z_i \sum \frac{Z_j}{D_{mol} r_{ij}} \quad (6)$$

where the second and third terms in the right-hand side represent hydration and Coulombic energies, respectively. Z_i is the charge of the group being protonated and r_i is its radius in angstroms. Z_j stands for any other charge in the molecule; r_{ij} is the distance in angstroms between charges Z_i and Z_j. D_{mol} and D_{ref} refer to effective dielectric constants that include dielectric and screening effects (*34*, *35*). D_{mol} refers to a particular site in the molecule, whereas D_{ref} corresponds to conditions under which ΔG_{int} is experimentally determined.

Equation 6 was used to reproduce the experimental protonation curve of native caricain on the basis of information obtained from the three-dimensional structure of the protein (*36*). However, to keep this task relatively easy to handle, we made the following simplifications: (1) The amino terminus and all Arg, Lys, and His side chains were assumed to be completely protonated in the whole pH region studied. (2) Interactions of a particular carboxylate were restricted to those with charged groups located at 10 Å or closer. As can be seen in Table 2, this cut off criterion leaves each carboxylate with at least two (in most cases, three) interactions with positive charges. In other words, we are assuming that ion pairs, which abound in caricain, constitute the main factor affecting the protonation constant of carboxylates. In addition, there were found groups of two (or three) acid side chains in which carboxyl−carboxyl distances are smaller than 10 Å. Clearly, in such instances, the protonation behavior of an individual carboxylate has to be considered as dependent on the ionization state of neighbor carboxylates. This is why in Table 2 some acid side chains appear grouped together. (3) Calculations were performed with two models. In the simplest model, a unique D_{mol} was applied to all electrostatic interactions. In the second model, two types of D_{mol} were defined according to the average solvent accessibility of charged side chains forming a cluster of interacting charges: D_1 for mean accessible surface area (ASA) larger than 25%, and D_2, for ASA less than 25%. In all cases, effective dielectric constants were taken as adjustable parameters.

Table 2: Estimated p*K* Values for Carboxyl Residues in Native Caricain and in Its Transition State for Unfolding

residue	interactions[a]	pK^b (native state)	pK^c (transition state)
Glu 3	Leu 1, Lys 168, Arg 195	2.7	3.3
Asp 6	Arg 8, Lys 9, Lys 10	1.1	2.0
Glu 187	Arg 8, Arg 192	3.3	3.7
Glu 35	Arg 17, Lys 178, Arg 83	2.5	3.2
Glu 47	Arg 17, Lys 178, Arg 83	3.4	3.7
Glu 50	Arg 17, Lys 178, Arg 83	1.4	2.4
Glu 52	Lys 98, Arg 96, His 81	2.6	3.2
Asp55	Lys 89, Lys 91, Arg 96	2.4	2.9
Glu 57	Arg 58, Arg 59, His 61, Lys 77	2.4	3.1
Glu 73	Arg 59, Lys 77	3.4	3.8
Glu 118	Lys 168, Arg 195	3.0	3.5
Glu 135	Lys 137, Lys 156, Lys 208	3.4	3.8
Asp 158	Lys 137, Lys 156, His 159	1.5	2.2
Glu 150	Lys 172, Lys 194, Arg 192	1.8	2.6
Asn 216 (C-terminus)	Arg 41, Lys 106, Lys 215	2.0	2.6

[a] Positively charged groups located at less than 10 Å from the carboxyl group. [b] Estimated from the fitting to the experimental titration curve at 25 °C (298 K) by using a model that includes a unique value for the effective dielectric constant in the protein ($D_{mol} = 59$); at 330 K, *pK* values in native caricain would be decreased by 0.1 units. [c] Estimated at 330 K from the fitting to data of $\Delta G^{\ddagger}$ vs pH; the value of D_{mol} was 51, and an expansion factor of 1.45 was applied to charge−charge distances to mimic the expansion from the native to the transition state (see text for details).

The last step required to calculate the titration curve is to relate K_i to the number of protons bound to a titratable site or group of sites, ν_i. A single carboxyl that is not interacting with other acid side chains represents an independent titratable site whose ν_i is given by

$$\nu_i = K_i a_H/(1 + K_i a_H) \quad (7)$$

For a group of two interacting carboxylates, four protonation states should be considered according to the following cycle:

$$\begin{array}{ccc} & K_1 & \\ A & \Leftrightarrow & AH_1 \\ K_2 \Updownarrow & & \Updownarrow K_2' \\ AH_2 & \Leftrightarrow & AH_1H_2 \\ & K_1' & \end{array}$$

where A is the unprotonated state, AH_1 and AH_2 are states in which either site 1 or site 2 is protonated, and AH_1H_2 is the state with both sites protonated. In this case, ν_i can be written in terms of three individual constants, for example,

$$\nu_i = \frac{K_1 a_H + K_2 a_H + 2K_l' K_2 (a_H)^2}{1 + K_1 a_H + K_2 a_H + K_l' K_2 (a_H)^2} \quad (8)$$

where K_1' is the protonation constant of site 1 when site 2 is already protonated. In the case of three interacting carboxyls, the protonation scheme involves eight different protonation states; ν_i can be written in terms of seven individual protonation constants as shown in eq 9

$$\nu_i = [(K_1 + K_2 + K_3)a_H + 2(K_1 K_{l,2}' + K_2 K_{2,3}' + K_1 K_{l,3}')a_H^2 + 3(K_1 K_{l,2}' K_{l,2,3}'')a_H^3]/[1 + (K_1 + K_2 + K_3)a_H + (K_1 K_{l,2}' + K_2 K_{2,3}' + K_1 K_{l,3}')a_H^2 + (K_1 K_{l,2}' K_{l,2,3}'')a_H^3] \quad (9)$$

. In this equation, single-primed constants refer to the protonation of a site when another site is already protonated, and the double-primed constant refers to the protonation of a site when the other two sites have been previously protonated.

Finally, the titration curve for the native protein is given by

$$\nu^N = \sum_i \nu_i \quad (10)$$

where the sum is taken over the nine groups of carboxylates listed in Table 2. In fitting eqs 6−10 to the experimental curve, we used an r_i of 1.82 Å, which is the average for hydrogen-bonded and non-hydrogen-bonded oxygen atoms (*37*). Values of pK_{int} for Glu, Asp, and α-COOH residues were constrained to vary only within ±0.1 of their commonly accepted values (*38*). Best fit (solid curve in Figure 5) was obtained with pK_{Glu} = 4.5, pK_{Asp} = 3.9, and p$K_{\alpha-COOH}$ = 3.7 and with D_{mol} = 59. Although the quality of fit is not impressive, it is apparent that the calculated curve follows the general trend of experimental data. Moreover, the value of D_{mol} does not seem unreasonable at all. Effective dielectric constants between 10 and 50 have been reported to apply for protein interiors (*35*, *39*). For each individual carboxylate, its estimated pK was calculated as the pH at which the site is half-protonated (Table 2). Slightly better fit to titration data (dotted line in Figure 5) was obtained with the model that includes two different values of D_{mol}, but only at the cost of getting an exaggeratedly large value (over 200) for one of the dielectric constants.

To apply our model to the analysis of the pH-variation of $\Delta G^‡$ it is convenient to rewrite eq 5 as its integral form:

$$\Delta G^‡ = \Delta G_0^‡ - RT \sum_i \int (\nu_i^{TS} - \nu_i^N)\,(\mathrm{d}\, a_H/a_H) \quad (11)$$

where = $\Delta G_0^‡$ is the value of $\Delta G^‡$ in the limit $a_H = 0$. In deriving this last equation, it is assumed that an expression analogous to eq 10 holds for ν^{TS}. Given the relationships between ν_i and a_H (eqs 7−9), integrals in eq 11 can be easily performed. For a single titratable site the result is

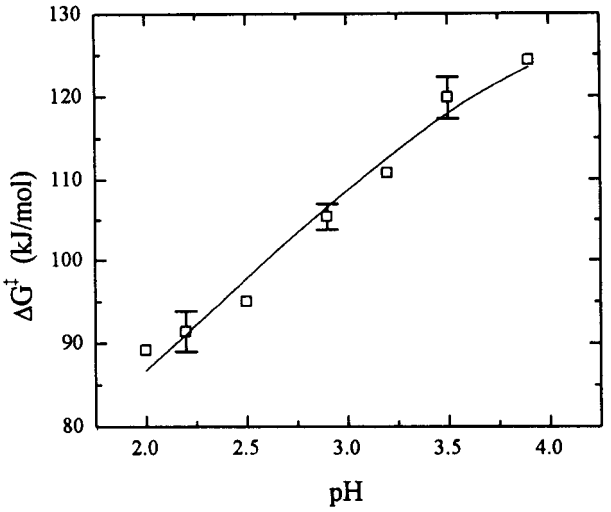

FIGURE 6: pH dependence of the activation free energy ($\Delta G^‡$) for caricain denaturation. $\Delta G^‡$ was determined at 330 K by extrapolation of lines in Figure 4. The solid curve is the best fit of a simple electrostatic model (based on eqs 6−13) to experimental data. An effective dielectric constant of 51 was used for the protein interior; the expansion factor that resulted from the fitting was 1.45 (see text for details).

$$\int (\nu_i^{TS} - \nu_i^N)\,(\mathrm{d}\, a_H/a_H) = \ln(1 + K_i^{TS} a_H) - \ln(1 + K_i^N a_H) \quad (12)$$

For a group of two or three interacting sites, the corresponding integral can be written in the general form

$$\int (\nu_i^{TS} - \nu_i^N)\,(\mathrm{d}\, a_H/a_H) = \ln Q_i^{TS} - \ln Q_i^N \quad (13)$$

where Q_i is a function like that appearing in the divisor of eq 8 or eq 9. Thus, we have a set of equations (eqs 6−13) that can be fitted to experimental data of $\Delta G^‡$ vs pH (Figure 6). $\Delta G^‡$ was calculated from kinetic data, using eq 3, by extrapolation of Eyring's plots (Figure 4) to 330 K (56.9 °C). This temperature was chosen because in this way the extrapolation range was minimized. It must be mentioned that the preexponential factor (k_B/h) in Eyring's original formulation of the theory of reaction rates may not apply to protein unfolding reactions (*40*, *41*). However, if the *correct* factor is assumed to depend mainly on temperature, we would be introducing at most a constant additive error in the determination of $\Delta G^‡$.

To get meaningful results from the fitting to data in Figure 6, it is convenient to reduce to a minimum the number of adjustable parameters in eq 6 and in the analogous expression for ΔG_i^{TS}. For this purpose, we assumed that r_{ij} values in the native state do not change with temperature. ΔG_{int} for Asp, Glu, and α-COOH can be calculated at 330 K employing the corresponding protonation enthalpies. Given the small magnitudes of these enthalpies (*42*), the necessary correction for ΔG_{int} is negligible. Otherwise, dielectric constants are expected to decrease significantly with temperature. For water, the ratio D_{330}/D_{298} is approximately 0.86; similar values of this ratio are found for liquids less polar than water, like methanol and ethanol (*43*). Using the factor 0.86 together with the value of D_{mol} at 298 K found in the regression of the titration curve of native caricain, we estimated that D_{mol} is equal to 51 at 330 K. In a first fitting attempt, it was supposed that the same value of D_{mol} applies

to both **N** and **TS**; that is, we tried to test the *dry molten globule* model of the transition state (*10*), which assumes that no significant increase in solvent-exposed area occurs on going from **N** to **TS**. To mimic the expansion thought to occur during formation of **TS**, all r_{ij} values were allowed to vary in the same proportion by multiplying them by a unique *expansion factor*, ef. For a uniform expansion of a spherical globule, ef would be the ratio of the expanded radius (R_{exp}) to the original radius of the globule (R_0). Thus, distances between points located at, or close to, the surface of the sphere would increase by a factor ef when the radius increases from R_0 to R_{exp}. Best fit (solid line in Figure 6) was obtained with $\Delta G_0^{\ddagger} = 130 \pm 2$ kJ mol^{-1} and ef $= 1.45 \pm 0.04$. For the nearest ion pairs found in native caricain ($r_{ij} \cong 2.6$ Å), such an expansion represents an increase in r_{ij} of 1.2 Å, whereas the average ion-pair distance changes from 5.9 Å in **N** to 8.6 Å in **TS**. As shown in Table 2, this expansion translates into considerable p*K* shifts for the carboxylates. Remarkably, the model that includes two dielectric constants for the protein gave also a good fit with a not very different ef of 1.62 ± 0.04. These results are certainly incompatible with the *dry molten globule* of **TS**, inasmuch as an increase of 40% in the radius of the globule would double its surface area. In this case, it is likely that several side chains augment their solvent exposure, thus leading to a variation of D_{mol} values in **TS** with respect to those in **N**. To explore this possibility, we studied the effect that varying D_{mol} (in **TS**) has on the quality of curve fitting. It was found that increasing D_{mol} led to a gradual worsening of the goodness of fit and to a decrease of ef. For example, if D_{mol} is increased from 51 to 68 (i.e., the value of D_{water} at 330 K), the root-mean-square of deviations for the fit changes from 1.2 to only 1.5 kJ mol^{-1}, and ef is reduced to 1.32. Given the uncertainties in experimental data, the goodness of fit criterion cannot be used to establish a lower limit for the expansion factor.

Overall, our results suggest that the conversion of **N** into **TS** may be accompanied by a significant increase in the solvent-accessible area of the protein. It must be stressed, however, that the expansion does not need to be uniform, for to explain the variation of $\Delta G^{\ddagger}$ with pH, it is sufficient to consider an expansion of ion-pair distances. It is reasonable to conceive a scenario in which charged, and perhaps also polar, regions at the protein surface expand to a larger extent than hydrophobic regions. In addition, this scenario might offer an explanation for the small change in heat capacity observed upon activation. Indeed, it is generally recognized (*44*, *45*) that in protein unfolding events hydration of apolar regions produces an increase in heat capacity, whereas hydration of polar regions decreases this thermodynamic property; however, the contribution from polar hydration expressed per unit of area exposed is smaller in magnitude than that arising from apolar hydration. Thus, a larger increase in polar vis-à-vis apolar solvent-accessible area might well lead to a cancellation of their opposite effects on the heat capacity.

It must be recalled, however, that results discussed above are based on the assumption that the same expansion factor applies to all r_{ij} values. Alternatively, it is possible that different charged regions in the protein might expand to distinct extents, an extreme situation being that in which only a few ion-pair distances change whereas the rest remain unchanged. We investigated this alternative by keeping the r_{ij} set for certain carboxylates fixed at their values in **N** during the fitting procedure. We found that at least five carboxylates should be allowed to vary their r_{ij} distances in order to preserve a good fitting. Nevertheless, in the later case, expansion factors are around 5−12, which translate into an increase of at least 30 Å in the average r_{ij} and p*K* shifts so large that all carboxyls involved in the expansion would reach their pK_{int} values. This would mean that in the transition state, nativelike and completely disordered regions of the molecule coexist without conflict, a situation that seems to lack physical significance. More meaningful results are obtained when all carboxylates present in one domain of the macromolecule are involved in the expansion. For example, Glu residues 35, 47, 50, 52, 57, 73, Asp 55, and the carboxyl terminus are located in the L-domain (*36*, *46*). Besides, residues 35, 47, and 50 are responsible for some of the interactions across the domain−domain interface. If only these eight carboxylates are included in the expansion, ef would be approximately 2.2, corresponding to an increase in the average r_{ij} of 7 Å. Furthermore, p*K* values for these acid groups in **TS** would still be 0.7 units below their pK_{int} values, on the average. According to these results, the transition state could be viewed as a state in which the whole structure of one domain is highly disordered, yet not completely unfolded.

In conclusion, our analysis suggests that there is a mean increase of the order of 3 Å in ion-pair distances when all charged regions in the caricain molecule expand to reach the transition state on the unfolding pathway. On the other hand, the minimal region to be expanded in this process seems to be a complete domain of the protein, in which case, distances would augment by ∼7 Å. These magnitudes should be taken only as rough estimations due to the simplicity of the method employed and the assumptions made in its application to the analysis of experimental data. Particularly, the proposal that charge−charge distances in **N** do not vary with pH is probably not fulfilled below pH 2.5, where the tertiary structure begins to be gradually loosened. Indeed, it is expected that as carboxylates are being protonated the electrostatic repulsion between positive charges would lead to an expanded form of **N**. If so, the degree of expansion necessary to attain the transition state may decrease at low pH. This would explain, in part, the reduction of the activation enthalpy when the pH is lowered. It is clear, however, that the approach proposed here can be very useful in characterizing transition states in protein unfolding reactions.

REFERENCES

1. Honig, B., and Yang, A. S. (1995) *Adv. Protein Chem. 46*, 27−58.
2. Hendsch, Z. S., and Tidor, B. (1994) *Protein Sci. 3*, 211−226.
3. Anderson, D. E., Becktel, W. J., and Dahlquist, F. W. (1990) *Biochemistry 29*, 2403−2408.
4. Horovitz, A., Serrano, L., Avron, B., Bycroft, M., and Fersht, A. R. (1990) *J. Mol. Biol. 216*, 1031−1044.
5. Oliveberg, M., Arcus, V. L., and Fersht, A. R. (1995) *Biochemistry 34*, 9424−9433.
6. Pfeil, W., and Privalov, P. L. (1976) *Biophys. Chem. 4*, 23−32.

7. Tanford, C. (1970) *Adv. Protein Chem. 24*, 1−95.
8. Oliveberg, M., and Fersht, A. R. (1996) *Biochemistry 35*, 2726−2737.
9. Segawa, S., and Sugihara, M. (1984) *Biopolymers 23*, 2473−2488.
10. Finkelstein, A. V., and Shakhnovich, E. (1989) *Biopolymers 28*, 1681−1694.
11. Serrano, L., Matouschek, A., and Fersht, A. R. (1992) *J. Mol. Biol. 224*, 805−818.
12. Solís-Mendiola, S., Gutiérrez-González, L. H., Arroyo-Reyna, A., Padilla-Zúñiga, J., Rojo-Domínguez, A., and Hernández-Arana, A. (1998) *Biochim. Biophys. Acta 1388*, 363−372.
13. Cavagnero, S., Debe, D. A., Zhou, Z. H., Adams, M. W. W., and Chan, S. I. (1998) *Biochemistry 37*, 3369−3376.
14. Arroyo-Reyna, A., and Hernández-Arana, A. (1995) *Biochim. Biophys. Acta 1248*, 123−128.
15. Solís-Mendiola, S., Arroyo-Reyna, A., and Hernández-Arana, A. (1992) *Biochim. Biophys. Acta 1118*, 288−292.
16. Solís-Mendiola, S., Zubillaga-Luna, R., Rojo-Domínguez, A., and Hernández-Arana, A. (1989) *Biochem. J. 257*, 183−186.
17. Polgár, L. (1981) *Biochim. Biophys. Acta 658*, 262−269.
18. Hennessey, J. P., and Johnson, W. C. (1982) *Anal. Biochem. 125*, 177−188.
19. Nozaki, Y., and Tanford, C. (1967) *Methods Enzymol. 11*, 715−734.
20. Vita, C., Fontana, A., Seeman, J. R., and Chaiken, I. M. (1979) *Biochemistry 18*, 3023−3031.
21. Goto, Y., and Hamaguchi, K. (1987) *Biochemistry 26*, 1879−1884.
22. Labhardt, A. M. (1982) *J. Mol. Biol. 157*, 331−355.
23. Sánchez-Ruiz, J. M., López-Lacomba, J. L., Cortijo, M., and Mateo, P. L. (1988) *Biochemistry 27*, 1648−1652.
24. Schmid, F. X. (1992) in *Protein Folding* (Creighton, T. E., Ed.) pp 197−241, Freeman, New York.
25. Jackson, S. E., and Fersht, A. R. (1991) *Biochemistry 30*, 10436−10443.
26. Pohl, F. M. (1968) *Eur. J. Biochem. 7*, 146−152.
27. Alexander, P., Orban, J., and Bryan, P., (1992) *Biochemistry 31*, 7243−7248.
28. Groves, M. R., Taylor, M. A., Scott, M., Cummings, N. J., Pickersgill, R. W., and Jenkins, R. W. J. A. (1996) *Structure 4*, 1193−1203.
29. Vernet, T., Berti, P. J., de Montigny, C., Musil, R., Tessier, D. C., Ménard, R., Magny, M. C., Storer, A. C., and Thomas, D. Y. (1995) *J. Biol. Chem. 270*, 10838−10846.
30. Sohl, J. L., Jaswal, S. S., and Agard, D. A. (1998) *Nature 395*, 817−819.
31. Anderson, D. E., Peters, R. J., Wilk, B., and Agard, D. A. (1999) *Biochemistry 38*, 4728−4735.
32. Privalov, P. L. (1979) *Adv. Protein Chem. 33*, 167−241.
33. Fink, A. L. (1995) *Annu. Rev. Biophys. Biomol. Struct. 24*, 495−522.
34. García-Moreno E., B. (1995) *Methods Enzymol. 259*, 512−538.
35. Warshel, A., and Åqvist, J. (1991) *Annu. Rev. Biophys. Biophys. Chem. 20*, 267−298.
36. Pickersgill, R. W., Rizkallah, P., Harris, G. W., and Goodenough, P. W. (1991) *Acta Crystallogr., Sect. B 47*, 766−771.
37. Rashin, A. A., and Namboodiri, K. (1987) *J. Phys. Chem. 91*, 6003−6012.
38. Antosiewicz, J., McCammon, J. A., and Gilson, M. K. (1994) *J. Mol. Biol. 238*, 415−436.
39. García-Moreno E., B., Dwyer, J. J., Gittis, A. G., Lattman, E. E., Spencer, D. S., and Stites, W. E. (1997) *Biophys. Chem. 64*, 211−224.
40. Parker, M. J., Lorch, M., Sessions, R. B., and Clarke A. R. (1998) *Biochemistry 37*, 2538−2545.
41. Hagen, S. J., Hofrichter, J., Szabo, A., and Eaton, W. A. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 11615−11617.
42. Pfeil, W., and Privalov, P. L. (1976) *Biophys. Chem. 4*, 41−50.
43. *Handbook of Chemistry and Physics* (1971) CRC Press, Cleveland, OH.
44. Murphy, K. P., and Gill, S. J. (1991) *J. Mol. Biol. 222*, 699−709.
45. Spolar, R. S., Livingstone, J. R., and Record, M. T., Jr. (1992) *Biochemistry 31*, 3947−3955.
46. Baker, E. N., and Drenth, J. (1987) in *Biological Macromolecules and Assemblies*, Vol. 3 (Jurnak, F. A., and McPherson, A., Eds.) pp 313−368, John Wiley & Sons, New York.

BI991658W