pH-Dependent Unfolding of Aspergillopepsin II Studied by Small-Angle X-ray Scattering†

Masaki Kojima,‡ Masaru Tanokura,§ Masahiro Maeda,‡ Kazumoto Kimura,|| Yoshiyuki Amemiya,⊥ Hiroshi Kihara,▽ and Kenji Takahashi*,‡

School of Life Science, Tokyo University of Pharmacy and Life Science, Hachioji, Tokyo 192-0392, Japan, Graduate School of Agricultural and Life Sciences, University of Tokyo, Bunkyo-ku, Tokyo 113-8657, Japan, Center of Medical Informatics, Dokkyo University School of Medicine, Mibu, Tochigi 321-0293, Japan, Graduate School of Engineering, University of Tokyo, Bunkyo-ku, Tokyo 113-8656, Japan, and Physics Laboratory, Kansai Medical University, Hirakata, Osaka 573-1136, Japan

Received July 8, 1999; Revised Manuscript Received October 28, 1999

ABSTRACT: Aspergillopepsin II (EC 3.4.23.6) secreted from the fungus *Aspergillus niger* var. *macrosporus* is a non-pepsin-type acid proteinase. It consists of two polypeptide chains (i.e., a heavy chain and a light chain), which are bound noncovalently to each other. The pH titration analysis using small-angle X-ray scattering (SAXS) as well as circular dichroism (CD) and gel filtration indicated that the enzyme was unfolded around a neutral pH with concomitant dissociation of the two chains. Detailed analyses showed that the midpoint pH values for the unfolding are not coincident with one another (pH 6.1 in circular dichroism and gel filtration, pH 6.4 in zero-angle intensity of SAXS, pH 6.8 in radius of gyration). The difference between these values suggested the existence of an intermediate state during the unfolding. Further analyses of the SAXS data showed that the heavy chain just after the dissociation still kept molecular compactness and that it gradually increased its dimensions as the pH was further raised. Noncoincidence of the two phenomena (i.e., chain dissociation and swelling) led to elucidation of a novel intermediate state during unfolding, which was confirmed by the subsequent singular value decomposition (SVD) analysis.

Proteinase A secreted from the fungus *Aspergillus niger* var. *macrosporus* (aspergillopepsin II) (EC 3.4.23.6) is a non-pepsin-type acid proteinase. It shows clear differences in various enzymatic properties from the ordinary pepsin-type aspartic proteinases. The enzyme consists of two polypeptide chains, a light chain of 39 residues and a heavy chain of 173 residues (*1*). The light chain has no cysteine residues, whereas the heavy chain has four cysteine residues that form two intrachain disufide bonds. In the native state, the two chains are bound noncovalently to each other. According to the X-ray crystal structure, the enzyme has much β-sheet and little helical structure (H. Sasaki et al., unpublished results). It is a highly acidic protein since about 20% (38 residues) of the total of 212 residues are glutamic acid or aspartic acid residues. As expected from the amino acid composition, aspergillopepsin II was unfolded in the higher pH region (above neutral pH), presumably due to the electrostatic repulsion in the molecule, accompanied by dissociation of the two chains. Further pH titration analysis may clarify this unfolding process in more detail.

Unfolding and/or refolding studies have been extensively done with various proteins in order to elucidate the protein folding mechanism and/or the architectural features of the native proteins. The unfolding of aspergillopepsin II above neutral pH, however, has the following characteristic aspects. (i) pH-dependent unfolding: Compared with unfolding by denaturants or at high temperature, this type of unfolding may be caused mainly by the change in the molecular charge distribution by pH changes. Since the charge distribution in the molecule can be changed significantly in a narrow pH range, pH-dependent unfolding often presents drastic behavior. In this kind of unfolding, acid denaturation has been studied for many proteins, such as cytochrome *c* (*2*−*4*). On the other hand, unfolding in neutral or alkaline pH regions has been less studied, except for pepsin (*5*). For pepsin, however, its detailed unfolding mechanism is still unclear. The investigation of the unfolding of aspergillopepsin II may elucidate these types of unfolding occurring in neutral or alkaline pH regions. (ii) Unfolding of a helixless protein: Many proteins studied so far have a higher content of helical structures. Since aspergillopepsin II has few helices, its unfolding process may show different behavior from those of other proteins. (iii) Unfolding of a multiple-polypeptide-chain molecule: Aspergillopepsin II consists of two polypeptide chains, which are bound noncovalently in the native conformation. The dissociation of the two chains is accompanied by the unfolding at neutral pH. The analysis of this type of protein is expected to be more complicated than that of single-chain molecules. The major aim of the present

† This work was supported in part by Grants-in-Aid from the Ministry of Education, Science, Sports and Culture of Japan.

* To whom correspondence should be addressed: Tel +81 426 76 7146; Fax +81 426 76 7149; E-mail kenjitak@ls.toyaku.ac.jp.

‡ Tokyo University of Pharmacy and Life Science.

§ Graduate School of Agricultural and Life Sciences, University of Tokyo.

|| Dokkyo University School of Medicine.

⊥ Graduate School of Engineering, University of Tokyo.

▽ Kansai Medical University.

10.1021/bi991584o CCC: $19.00
Published on Web 01/19/2000

study is to elucidate the relationship between the dissociation of the two polypeptide chains and the conformational change of each chain.

Small-angle X-ray scattering (SAXS)[1] is a very effective method for investigating the protein unfolding/refolding process (*6*, *7*), because it provides information about the molecular size and shape. Various SAXS parameters such as radius of gyration (R_g), zero-angle intensity [$I(0)$], Kratky plot, and integral intensity provide useful information about the overall structural properties of molecules. In addition, recent advances in synchrotron radiation enable us to measure SAXS data with a good signal-to-noise ratio.

In this paper, we report equilibrium studies on pH-dependent unfolding of aspergillopepsin II by SAXS and other experimental methods including circular dichroism (CD) and gel filtration. The results showed that the heavy chain just after the dissociation still kept a globular shape and that it becomes gradually swollen as the pH was raised further. Noncoincidence of the two phenomena (i.e., dissociation of the two chains and swelling of the heavy chain) suggested a novel intermediate state during unfolding of aspergillopepsin II. The structural properties of this novel intermediate state are discussed in contrast with molten globule states.

MATERIALS AND METHODS

Materials. Aspergillopepsin II was purified from the crude extract powder prepared from the broth filtrate of *Aspergillus niger* var. *macrosporus*. The purification was achieved by two-step column chromatography on DEAE-Toyopearl 650S according to the method previously reported (*8*). All the chemical reagents used were of the purest grade available commercially. The concentration of aspergillopepsin II was determined spectrophotometrically using an extinction coefficient, $\epsilon_{1cm}^{1\%} = 16.3$ at 280 nm. Enzymatic activity of aspergillopepsin II was determined with denatured bovine hemoglobin as a substrate as described (*8*), with slight modification.

CD Spectroscopy and Gel Filtration. CD measurements were performed with Jasco J-600 and J-720 spectropolarimeters. Cuvettes with 1 mm path-length and 10 mm path length were used in far-UV (200−250 nm) and near-UV (250−320 nm) regions, respectively. Temperature was kept at 37 °C, and the sample concentration was 0.05% in 50 mM buffer containing 200 mM NaCl. Sample solutions were prepared by diluting the enzyme stock solution (1.0 mg/mL in 10 mM sodium acetate, pH 4.5) with each buffer solution just before measurement. The buffers used were sodium acetate (pH 4.5−5.25), sodium MES (pH 5.5−6.75), and sodium HEPES (pH 7.0−8.5). For the pH measurements, a Radiometer PHM84 pH meter equipped with a long thin glass electrode from Ingold was used. The secondary structure content was estimated with the program CONTIN (*9*).

Gel filtration was performed at 37 °C using a Tosoh HPLC apparatus CCPM with a column (0.75 × 60 cm) of TSKgel G3000SW. Flow rate was 0.8 mL/min, and elution was monitored at 280 nm. Sample solutions were prepared to a concentration of 0.3 mg/mL in 50 mM eluent buffer containing 200 mM NaCl. An aliquot (10 mL) of the solution was analyzed after preincubation at each pH for 30 min. The buffers used for each pH were the same as in the CD measurements.

SAXS. SAXS measurements were performed at the beamline 15A small-angle installation of the Photon Factory at the National Laboratory for High Energy Physics, Tsukuba, Japan. The sample solution in a sapphire or mica cell with 1 mm path length was irradiated with a monochromatic X-ray beam (1.5 Å). Scattered X-rays were recorded with an argon gas-filled linear position-sensitive proportional counter of 512 channels along the angular region with 0.368 mm width. A camera with a 1.2 m sample-to-detector distance collected data in the range of the scattering vector, h, from 0.002 to 0.3 Å^{-1}, where h is given by $h = 4\pi \sin\theta/\lambda$ (λ, wavelength; 2θ, scattering angle). The exposure time of the beam for equilibrium measurements was 300 s for sapphire cell and 150 s for mica cell. Sample solutions were prepared by 2-fold diluting the enzyme stock solution (10 mg/mL in 10 mM sodium acetate, pH 4.5) with each 100 mM buffer solution containing 200 mM NaCl just before measurements. The same buffers as in the CD measurements were used.

In all the measurements, the temperature of the apparatus was kept at 37 °C. The background data for the buffer solvent were collected before and after data collection for the protein solution. The SAXS data were corrected for the net electron density contrast between solute and solvent as described (*10*). The values of R_g and $I(0)$ were determined from the data in the small-angle region according to the Guinier approximation:

$$I(h) = I(0) \exp(-R_g^2 h^2/3) \quad (1)$$

where $I(h)$ is the scattering intensity.

Data Analyses for pH Titration. All the data were analyzed according to the following proton dissociation scheme:

$$AH_n \rightleftharpoons A^{n-} + nH^+ \quad (2)$$

where AH_n and A^{n-} represent the protonated and deprotonated form, respectively, and H^+ is a proton.

In the present unfolding process, AH_n corresponds to the native (N) species, and A^{n-} to the unfolded (U) species. Considering the cooperativity effect of unfolding, we include the Hill coefficient into the Henderson−Hasselbalch relationship for the analysis (*11*):

$$K = 10^{m(pH - pK_a)} \quad (3)$$

where K represents the equilibrium constant of unfolding ($f_{A^{n-}}/f_{AH_n}$, that is, f_U/f_N), m is the Hill coefficient, pK_a means midpoint pH values, and f represents the mole fraction of each species. The fact that the Hill coefficient was estimated to be m indicates that simultaneous dissociations of at least m protons contributed to the unfolding ($n \geq m$). Optimal values of pK_a and m were estimated by the least-squares minimization from the data in the titration curves.

Singular Value Decomposition. In the SVD analysis, all the SAXS data were arranged in the matrix form so that the kth column of the data matrix D corresponds to the scattering pattern at the kth pH value. The scattering patterns were

[1] Abbreviations: SAXS, small-angle X-ray scattering; R_g, radius of gyration; MES, 2-morpholinoethanesulfonic acid; HEPES, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid; N, native state; U, unfolded state; SVD, singular value decomposition; I, intermediate state; DSC, differential scanning calorimetry.

deconvoluted into the series of orthogonal basis and their weighting factors:

$$D = UWV^{T} \quad (4)$$

where D, U, and V are $M \times N$, $M \times N$, and $N \times N$ matrices, respectively (M and N are the numbers of h and pH values for SAXS data, respectively). The matrix V^{T} represents the transposed matrix of V. W is the $N \times N$ diagonal matrix of the singular values, which are usually rearranged by their magnitudes in descending order. The columns of U are the orthogonal basis scattering curves, and WV^{T} determines the pH-dependent weights for the individual basis. Equation 4 indicates that all the scattering data can be expressed as a linear superposition of these orthogonal bases. In the ideal case without noise, the rank of the matrix D represents the number of mathematically independent states. For real data with noise, however, the number of nonnoise components (L) must be estimated from the SVD results. In this case, bases with $j > L$ are supposed to represent noise. In determining the number L, (1) singular values, (2) autocorrelation of each base (fluctuation of each base function against h), (3) shape of basis functions, and (4) reduced χ^2 values of SVD reconstruction were considered (*12*−*14*). The autocorrelation of the ith column (U_i) of matrix U reflects the signal-to-noise ratio of the column and is expressed as

$$C(U_i) = \sum_{j=1}^{M-1} U_{j,i}U_{j+1,i} \quad (5)$$

where $C(U_i)$ represents the autocorrelation of the column U_i and $U_{j,i}$ is the (j,i) element of matrix U. It is mathematically required that $C(U_i)$ takes a value between −1 and 1 since all the columns of U are normalized. And it is easily indicated that the column which exhibits slow variation from row to row (i.e., signal) has a $C(U_i)$ value close to 1, and the column with rapid row-to-row variation (i.e., noise) has a value much less than 1.

To optimize the signal-to-noise ratio of the basis sets, the original matrix U and V were transformed (rotated) into U' and V' as

$$U' = UR \quad (6a)$$

$$V' = VWR \quad (6b)$$

On the basis of these relationships, the initial data matrix I can be expressed as $D = U'V'^{T}$ instead of eq 4 after the transformation. Elements of the orthogonal matrix R were determined so that a new set of basis (columns of matrix U') has maximum values in autocorrelation $C(U_i)$ by the same procedure described in ref *15*.

Reconstruction of Scattering Profile. The thermodynamic model for the three-state unfolding can be expressed:

$$N \overset{K_1}{\rightleftharpoons} I \overset{K_2}{\rightleftharpoons} U \quad (7)$$

where N, I, and U correspond to the individual thermodynamic states. K_1 and K_2 represent the equilibrium constants between N and I and between I and U, respectively. Then the fraction of each species can be expressed

$$f_N = 1/(1 + K_1 + K_1K_2) \quad (8a)$$

$$f_I = K_1/(1 + K_1 + K_1K_2) \quad (8b)$$

$$f_U = K_1K_2/(1 + K_1 + K_1K_2) \quad (8c)$$

For a similar reason for eq 3, each equilibrium constant can be represented with the midpoint pH and Hill coefficient as a function of pH.

$$K_i = 10^{m_i(pH - pK_{a_i})} \quad (i = 1, 2) \quad (3a)$$

The subscript i denotes the corresponding transition process. The observed scattering intensity in each state, therefore, can be expressed as

$$I(h, pH) = f_N(pH)I_N(h) + f_I(pH)I_I(h) + f_U(pH)I_U(h) \quad (9)$$

where $I_N(h)$, $I_I(h)$, and $I_U(h)$ are the scattering intensities intrinsic to the states N, I, and U, respectively.

From the SVD results, any scattering patterns including $I_N(h)$, $I_I(h)$, and $I_U(h)$ can be represented as a linear superposition of the basis $U_1(h)$, $U_2(h)$, and $U_3(h)$. For example, in the case of $I_N(h)$,

$$I_N(h) = P_{N,1}U_1(h) + P_{N,2}U_2(h) + P_{N,3}U_3(h) \quad (10)$$

where P is the corresponding weight for each base.

By combining eqs 4, 9, and 10, the WV^{T} elements in SVD can be expressed with f and P as

$$w_iV_{k,i} = f_N(pH)P_{N,i} + f_I(pH)P_{I,i} + f_U(pH)P_{U,i} \quad (i = 1, 2, 3) \quad (11)$$

where $w_iV_{k,i}$ is the (i, k) element of the matrix WV^{T}, which represents the weight for the base U_i at the kth pH value. From eqs 3a, 8a−c, and 11, each value of f and P (and thereby each value of pK_a and m) can be determined by least-squares fitting.

RESULTS

CD Spectra. Figure 1A shows the far-UV CD spectra of aspergillopepsin II at pH 4.5, 6.5, and 7.0. They indicate that at pH 4.5 aspergillopepsin II takes a native structure with a high β-sheet content, and that at pH 7.0 and above the enzyme has no characteristic secondary structures. The contents of helix, β-sheet, and β-turn in the native state, estimated from the CD spectra, were 2%, 57%, and 16%, respectively, at pH 4.5. This estimation was consistent with the results of 2D NMR spectroscopy that most α-proton signals of aspergillopepsin II distributed in the lower magnetic field region than the solvent water signal (*16*). Figure 1B shows the titration curves of $[\theta]_{218}$, $[\theta]_{222}$, and $[\theta]_{229}$ versus various pHs. The midpoint pH values and the Hill coefficients of transitions were estimated to be 6.09 and 4.5, respectively, and are also listed in Table 1.

Figure 2A shows the near-UV CD spectra of the enzyme at various pHs. The spectral patterns at pH 4.5−6.0 and at pH 7.0−7.9 remain almost unchanged and are characteristic of the native and unfolded states, respectively. The titration

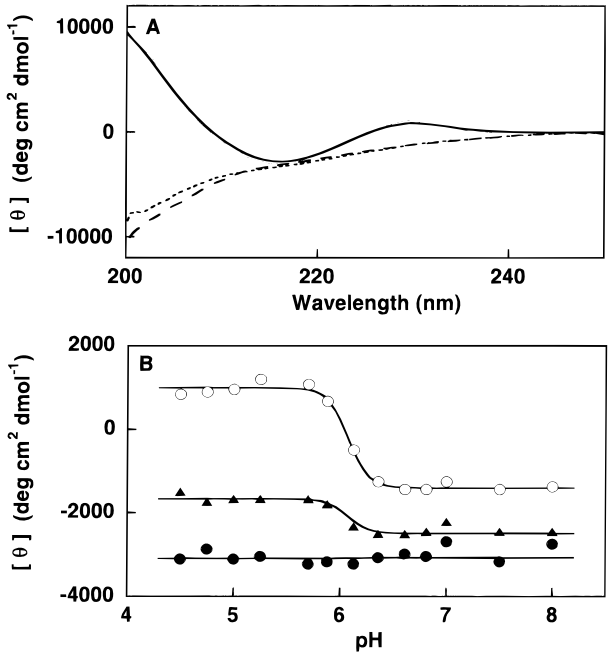

FIGURE 1: Far-UV CD of aspergillopepsin II. (A) Spectra at pH 4.5 (solid line), pH6.5 (dashed line), and pH 7.0 (broken line). (B) pH titration curves of [θ] at 218 nm (●), 222 nm (▲), and 229 nm (○).

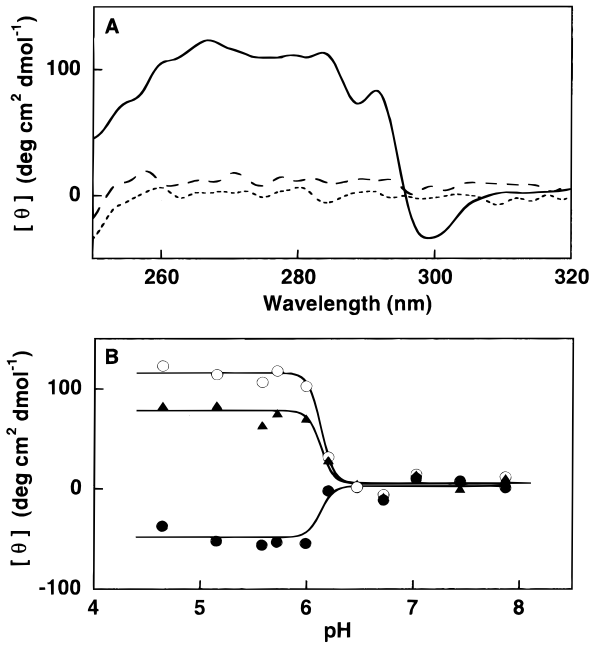

FIGURE 2: Near-UV CD of aspergillopepsin II. (A) Spectra at pH 4.5 (solid line), 6.5 (dashed line), and 7.0 (broken line). (B) pH titration curves of [θ] at 267 nm (○), 291 nm (▲), and 300 nm (●).

Table 1: Unfolding Parameters of Aspergillopepsin II

	midpoint pH (pK_{a1})	Hill coefficient (m_1)	midpoint pH (pK_{a2})	Hill coefficient (m_2)
far-UV CD[a]	6.09 ± 0.03	4.5 ± 1.2		
near-UV CD[b]	6.13 ± 0.00	7.3 ± 1.9		
gel filtration	6.03 ± 0.03	3.9 ± 0.9		
$I(0)$	6.39 ± 0.05	5.6 ± 2.7		
R_g[c]	6.75 ± 0.03	3.8 ± 0.9		
SVD	6.40	5.1	7.02	3.8

[a] The titration curves of $[\theta]_{218}$, $[\theta]_{222}$, and $[\theta]_{229}$ were fitted simultaneously. [b] The titration curves of $[\theta]_{267}$, $[\theta]_{291}$, and $[\theta]_{300}$ were fitted simultaneously. [c] Data were analyzed by $R_g^2 = f_N R_N^2 + f_U R_U^2$, where R_N and R_U represent the R_g values intrinsic to the states of N and U, respectively.

curves of $[\theta]_{267}$, $[\theta]_{291}$, and $[\theta]_{300}$ versus pH are also shown in Figure 2B. The midpoint pH value and the Hill coefficient of the transition were 6.13 and 7.3, respectively, and are also listed in Table 1. These values are in good agreement with those obtained from far-UV CD spectra within estimated errors. These results indicate that the secondary and the tertiary structures (especially around aromatic side chains) of aspergillopepsin II are unfolded simultaneously around pH 6.0.

Gel Filtration. Since the native enzyme consists of two polypeptide chains bound noncovalently, gel-filtration experiments were carried out at various pHs in order to clarify the relationship between the unfolding of the enzyme and the dissociation of the two chains (Figure 3). Although only a single peak showing the native protein was observed at pH 4.5, additional two peaks, corresponding to the heavy chain and the light chain, appeared at pH 6.0 before and after the peak of the native enzyme, respectively. At pH 6.5 and above, the peak of the native protein disappeared, indicating that the two chains were completely dissociated. The fact that the heavy chain was eluted faster than the native

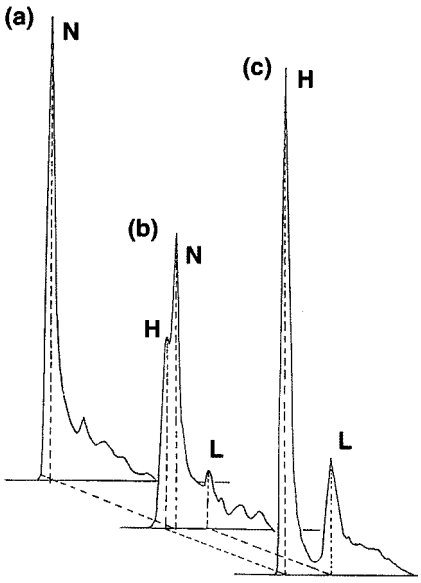

FIGURE 3: Gel filtration patterns of aspergillopepsin II at pH 4.5 (a), pH 6.0 (b), and pH 6.5 (c). N, H, and L represent the peaks of the native enzyme, the dissociated heavy chain, and the dissociated light chain, respectively.

protein is consistent with the SAXS results described below, indicating that the dissociated heavy chain is less compact than the native enzyme. The peak areas, proportional to the amount of the corresponding molecular species, were titrated versus pH, and the midpoint pH values and the Hill coefficients of the transition were estimated (Table 1). The values monitored by gel filtration were practically the same as those obtained by far-UV and near-UV CD spectroscopy, which indicates that the dissociation of the two chains is accompanied with changes in the secondary and tertiary structures in the whole molecule.

SAXS Data for Guinier Plot—R_g and $I(0)$. Since SAXS intensity can be approximated in the Guinier region ($R_g h <$

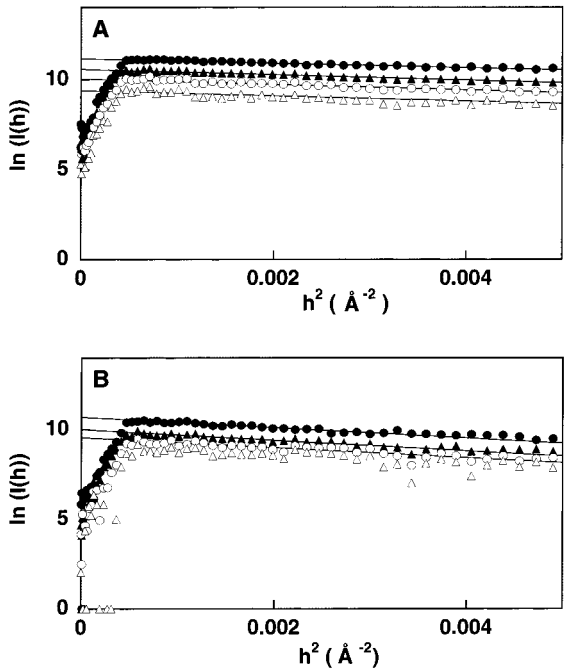

FIGURE 4: Guinier plots of aspergillopepsin II at various protein concentrations and at pH 4.5 (A), and pH 8.1 (B). Sample concentrations are 5.0 mg/mL (●), 2.5 mg/mL (▲), 1.25 mg/mL (○), and 0.625 mg/mL (△). The fitted lines are also presented.

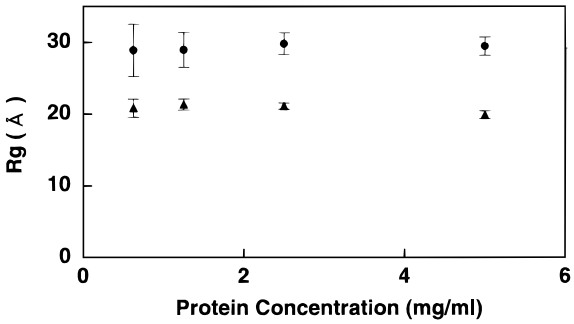

FIGURE 5: Protein concentration dependency of R_g at pH 4.5 (▲) and pH 8.1(●).

1.3) by eq 1, R_g and $I(0)$ can be determined from the Guinier plot [$ln\ (I(h))$ vs h^2]. Figure 4 represents the Guinier plots at various protein concentrations, and Figure 5 shows the concentration dependency of R_g in the native and unfolded states. It was shown that there was no dependency of the R_g values upon the protein concentration in the present measurements, and the same results were obtained for $I(0)$ values (data not shown). Figure 6 shows the pH dependency of R_g. In the native state aspergillopepsin II has an R_g value of about 20 Å (20.0 ± 0.6 at pH 4.5), which gradually increased to about 30 Å (28.9 ± 1.3 at pH 8.1) as the unfolding proceeded.

On the other hand, $I(0)$ is proportional to $(\Delta\rho v)^2$ ($\Delta\rho$, difference in electron density between the protein molecule and solvent; v, molecular volume). If the solvent molecules are not fixed on the protein surface, the volume is approximately the same in the globular and coiled states. In those cases, $I(0)$ can be used as an effective parameter for molecular association/dissociation states. As shown in Figure 7, the $I(0)$ value decreased to about 65% of the native one above pH 6.5. Considering that $I(0)$ should be proportional to $\sum_i c_i M_{w,i}^2$ (c_i and $M_{w,i}$ are the molar protein concentration and the molecular weight of the ith species, respectively), it can be interpreted that the light and heavy chains may have dissociated above pH 6.5. Furthermore, in the dissociated state the observed R_g^2 was the Z-averaged value of the intrinsic R_g^2 values of the light and heavy chains. Considering the molecular weights of the two chains, most contribution to the R_g^2 value in the dissociated state should be made by the heavy chain.

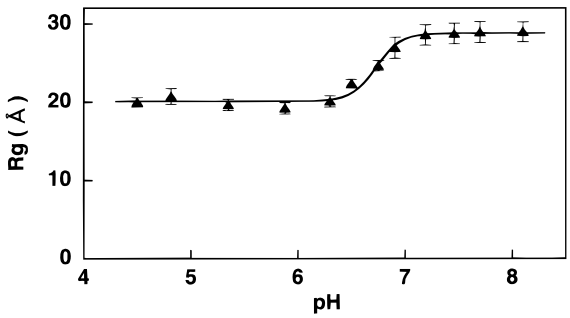

FIGURE 6: pH titration curve of R_g. Curve fitting was carried out for R_g^2 values, and the corresponding R_g values and their errors are presented by solid line and bars, respectively.

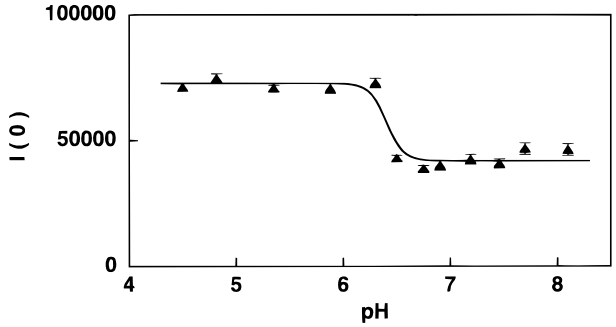

FIGURE 7: pH titration curve of $I(0)$. Fitted lines are also presented.

The optimal values of the midpoint pHs and the Hill coefficients determined from the titration data for both parameters are listed in Table 1. The midpoint pH value for $I(0)$ was apparently lower than that for R_g. Especially, at pH 6.5 the $I(0)$ transition had been mostly completed, although the R_g value was only 10% larger than the native one. This indicates that the heavy chain just after the dissociation still kept molecular compactness, and it gradually swelled as pH was raised further toward the alkaline region. The noncoincidence of the two phenomena (i.e., dissociation of the two chains and swelling of the heavy chain) suggested the existence of an unfolding intermediate species. This point will be described further below.

SAXS Data for Kratky Plot. Information about molecular globularity can be obtained from the entire scattering pattern by plotting $I(h)h^2$ against h (Kratky plot). It is known that the presence of a peak in a Kratky plot indicates a globular shape of a molecule, while a plateau in moderate angles of the plot indicates a coiled structure (*3*). It is readily seen that the value of the Kratky plot maximum for the globular states should be close to $3I(0)/eR_g^2$ at $h = \sqrt{3}/R_g$, which is obtained from the maximum of the function, $I(h)h^2 = h^2I(0)\exp(-R_g^2h^2/3)$, where e denotes the base of the natural logarithm (*6*, *17*). Therefore, in globular states, the peak positions of the maximum reflects the inverse of R_g, and its height depends on $I(0)$ and R_g. As shown in Figure 8, the Kratky plot curves at pH 4.5 indicate globular shapes, and

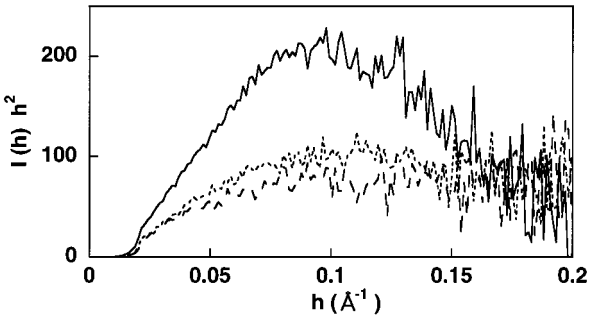

FIGURE 8: Kratky plots of aspergillopepsin II at pH 4.5 (solid line), pH 6.5 (dashed line), and pH 8.1 (broken line).

the curves at pH 8.1 showed coiled structures. The curves in the range of pH 4.5−6.3 and pH 6.75−8.1 were almost unchanged. At pH 6.5, there is a peak in the curve, although the peak height is much smaller than that in the native state (about 60% of the native peak height). A slight peak shift toward lower *h* was also observed. These features of the plot at pH 6.5 can be explained by the fact that at that pH the $I(0)$ value decreased to 60% of that in the native state, while R_g was little larger than that of the native one (Figures 6 and 7). The molecular shape at pH 6.5 (mainly the shape of the dissociated heavy chain) was, therefore, globular. Furthermore, there were no isoscattering points among the scattering curves at pH 4.5, 6.5, and 8.1, which strongly suggests that the folding state at pH 6.5 cannot be explained as an equilibrium mixture of the two states at pH 4.5 (native) and 8.1 (unfolded).

Singular Value Decomposition. The results described above strongly suggest the existence of an unfolding intermediate state around pH 6.5. To confirm this indication, we carried out an SVD analysis (*15*, *18*) of the SAXS data of aspergillopepsin II. Figure 9 shows the SVD results for the SAXS data of aspergillopepsin II. From the results, the number of components was concluded to be 3 or 4 [it should be noted that in the Kratky plot the noise amplitudes are more emphasized at the larger *h* values due to the ordinate of $I(h)h^2$]. Especially upon consideration of the shapes of the basis scattering patterns and their autocorrelation values, the fourth component appeared to be more contributory than the third one. Since the number of bases in the space spanned by the columns of the data matrix *D* (i.e., dimension) is independent of the individual basis sets, another orthogonal basis sets can also be used for analysis. Therefore, we carried out the orthogonal transformation (rotation) (*15*) of the subset of the original basis obtained by SVD, to optimize the signal-to-noise ratio of the retained components. Figure 10 shows the results of the transformation. From this, it was clearly indicated that the number of components (L) was 3, which confirmed the SAXS results for the Guinier and Kratky plots described above. Therefore, we suppose that there are three different species during the unfolding process of aspergillopepsin II.

Reconstruction of Scattering Profile. Since the number of independent species is 3, we reconstructed the scattering profile in each state according to the three-state model described in Materials and Methods. In the actual calculation, we used the optimized SVD parameters (U' and V') instead of the original ones (U and wV). The estimated pK_a and m values are listed in Table 1, and the calculated f values (populations of three species N, I, and U at each pH) are

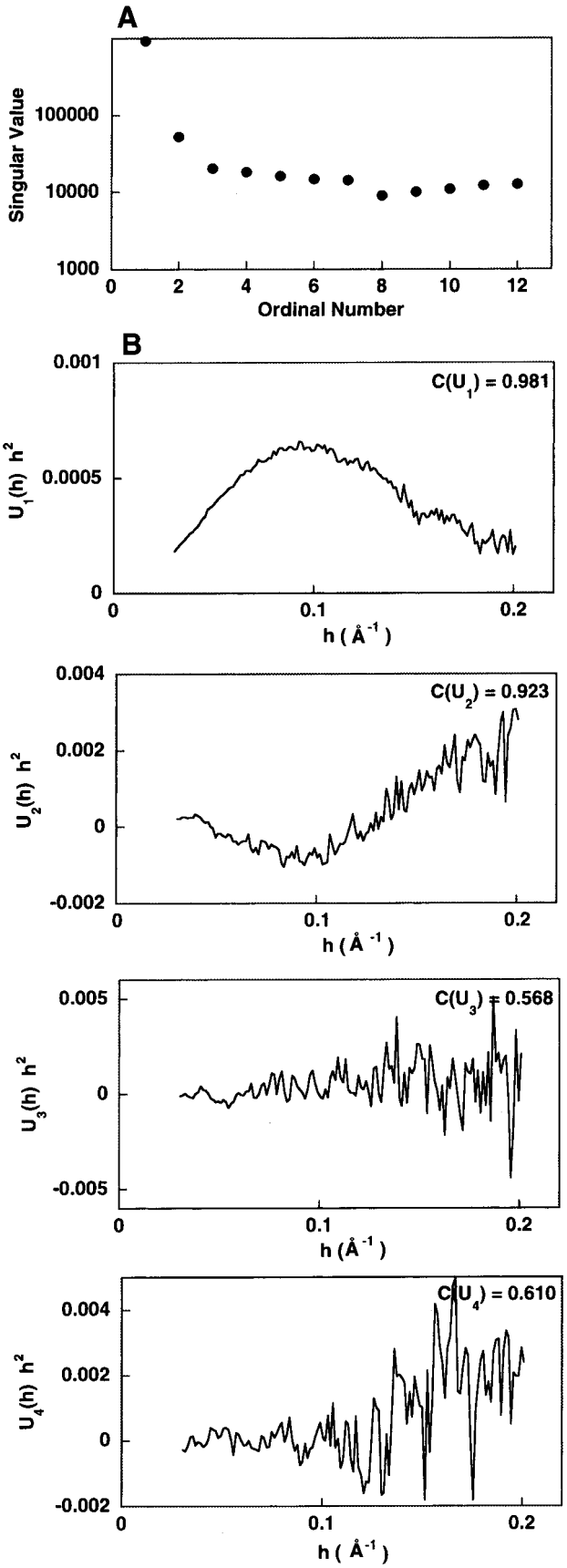

FIGURE 9: Singular value decomposition (SVD) analysis for SAXS data. (A) Logarithm plot of singular values of matrix *D* against ordinal numbers. (B) Kratky plots of the SVD basis scattering curves (columns of matrix *U*). The first four basis scattering curves are shown. In each curve the autocorrelation value is also represented.

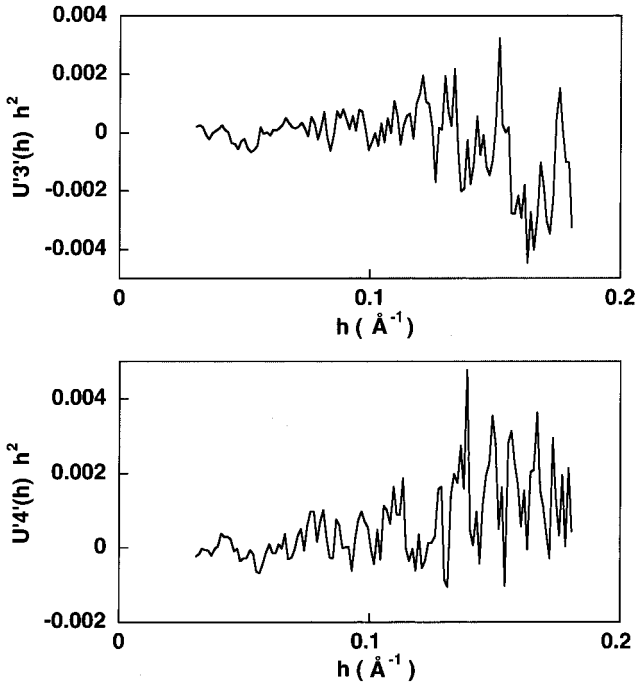

FIGURE 10: Kratky plots of the optimized SVD basis scattering curves. Basis curves 3 through 8 in Figure 10 were optimized by the orthogonal transformation described in the text, and the first two basis curves ($U'_{3'}$ and $U'_{4'}$) are presented.

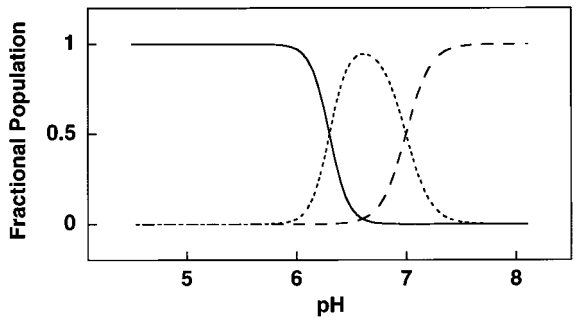

FIGURE 11: Fractional populations of the three states at each pH, N (solid line), I (dashed line), and U (broken line), estimated from the SVD results according to the three-state model.

shown in Figure 11. Because of the arbitrarity of Hill coefficients in the three-state transition, the precise values of the unfolding parameters in both transitions could not be determined simultaneously. For example, if we attempted to determine the values of pK_{a1} and m_1 more precisely, the estimated errors of pK_{a2} and m_2 became larger. Therefore, the least-squares minimization fitting was performed only near the predetermined values from the two-state pH titration analyses described above, and the estimated errors are not presented in Table 1. On the basis of this calculation, the scattering patterns intrinsic to the species N, I, and U were constructed as shown in Figure 12. The scattering patterns of N, I, and U resembled the experimental ones at pH 4.5, 6.5, and 8.1 (in Figure 8), respectively. Guinier analysis of these profiles yielded R_g values of 18.3 ± 0.2, 24.0 ± 0.9, and 29.8 ± 1.1 Å for N, I, and U, respectively. Although the values of I and U were a little larger than those at pH 6.5 and 8.1, respectively, the results were nearly consistent with the pH titration experiments.

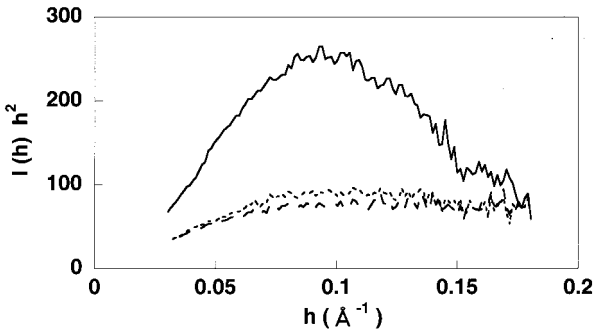

FIGURE 12: Kratky plots intrinsic to the states, N (solid line), I (dashed line), and U (broken line), based on the SVD analysis.

DISCUSSION

All the results described above strongly suggest that around pH 6.5, aspergillopepsin II takes an unfolding intermediate state, which is definitely distinguishable from both the native and the unfolded states. Its features are (1) the heavy chain is dissociated from the light chain, (2) the secondary and tertiary structures in each chain are lost, (3) the molecular size (of the dissociated heavy chain) is a little (by 10%) larger than in the native state, and (4) the molecule has a globular shape. Among the features, molecular compactness is an important characteristic of the folding state of proteins, although its direct observation is difficult. Recently Semisotnov et al. (*6*) have shown the integral SAXS intensity (I_{int}) in the limited region to be an effective parameter of molecular compactness. Therefore, we investigated the pH dependency of I_{int} in order to confirm the above indication. We employed an integral region of h from 0.03 to 0.18 Å^{-1} because the change in Kratky plots in this region was very large between the native and the unfolded states and because the corresponding curves in these two states intersected around h = 0.18 Å^{-1} (Figure 8). I_{int} showed the same behavior as the R_g values against pH (data not shown), which was consistent with the above features at pH 6.5, and indicated that the molecule was a little less compact than the native molecule at that pH.

Among the above features at pH 6.5, all the features except (2) coincide with those of molten globules (*2*). The molten globules have been characterized as having a compact size, a globular shape, pronounced secondary but little rigid tertiary structure, and a hydrophobic core exposed to solvent (*7*). Recently Kataoka et al. (*17*) classified the various types of molten globules from SAXS data. The folding state of aspergillopepsin II at pH 6.5, however, is not a molten globule state because the molecule has no secondary structures. In such a sense, it may be more appropriate to express the folding state at pH 6.5 not as an intermediate state (I) between the native (N) and unfolded (U) states, but as another unfolded state (U_1) different from completely unfolded state (U_2). This state was also observed in the thermal unfolding of aspergillopepsin II. In differential scanning calorimetry (DSC) experiments of aspergillopepsin II, two peaks were observed in the curves above pH 6. Fukada et al. (*19*) assigned the first transition (in the lower temperature region) as the dissociation of the two chains, and the second transition (in the higher temperature region) as the conformational transition of the dissociated heavy chain. Inspection of the DSC data at pH 6.5 indicated that around 37 °C all the molecule had already finished the

FIGURE 13: Schematic views showing the pH-dependent unfolding mechanism of aspergillopepsin II.

dissociation process and that about one-third of the fraction had transited to the final unfolded state. This is consistent with the present results obtained from the pH-dependent unfolding experiments. It is unknown from the DSC data whether the second transition in thermal unfolding (that is, the conformational change of the heavy chain) corresponds to the molecular expansion of the heavy chain in pH-dependent unfolding or not. However, it is feasible that this conformational change may be caused by the decrease in hydrophobic interaction at high temperature and that the main stabilizing factor of the intermediate state may thus be a nonspecific hydrophobic interaction inside the heavy chain. The same deduction may be applicable to the pH-dependent unfolding. We think the main stabilizing factor of the intermediate state in the pH-dependent unfolding (especially the factor stabilizing the molecular compactness) is a nonspecific hydrophobic interaction as in the case of the thermal unfolding. Furthermore, the two disulfide bonds inside the heavy chain may inhibit more or less the molecular expansion both in the intermediate and the more unfolded states. (Even at pH 8.1, the shape of the Kratky plot is a little convex upward, suggesting that the molecule is not completely coiled.) On the other hand, the main stabilizing factor of molten globule structures is assumed to be a nonspecific hydrophobic interaction. Therefore, the thermodynamic properties of the unfolding intermediate of aspergillopepsin II may be similar to those of molten globules. Since the dissociation of the two chains causes the loss of the secondary structures, the overall properties of the above states may seem to be apparently different. Figure 13 shows the schematic description of the pH-dependent unfolding of aspergillopepsin II.

At the molecular level, pH-dependent unfolding is triggered by the change in the electrostatic charge balances among the stabilizing factors supporting the native structure, although thermal unfolding is caused by the increase in the thermal motion. It is interesting to note that aspergillopepsin II shows the same behaviors in the two kinds of unfolding processes caused by different factors. Probably, the dissociation of the two chains may be a crucial step for the architecture of this molecule, and the folding states after the dissociation may have the same structural features irrespective of the applied methods.

Recently Segel et al. (*14*) and Bilgin et al. (*13*) have demonstrated the effectiveness of SVD analysis for discerning the number of subensembles in the individual thermodynamic states. In the present study, we used SVD analysis only to confirm the three-states unfolding model. But SVD not only can minimize the number of independent parameters that must be fitted by reducing the rank of matrix but also can accumulate the widely distributed signal amplitudes into a few retained basis components by the following optimization procedure, which increases signal-to-noise ratio in the reconstructed spectra (as shown for the Kratky plots in Figures 8 and 12). Therefore, for the SAXS data, which had a relatively low signal-to-noise ratio compared with other experimental methods, SVD analysis may be a useful procedure for reducing the noise level in the obtained spectra as well as for characterizing the multiple component system.

The transition observed by $I(0)$ may be reflected in the value of pK_{a1} for SVD, and the pK_a value monitored by R_g is expected to correspond to the transition between N and U. From eqs 8a−c, the apparent equilibrium constant between N and U (f_U/f_N) is expressed as K_1K_2. Therefore, the corresponding pK_a values are estimated from eq 3a as $(m_1pK_{a1} + m_2pK_{a2})/(m_1 + m_2)$. This value was 6.66 for the SVD data and was consistent with that obtained from the titration curve of R_g. In addition, as shown in Figure 12, the reconstructed scattering pattern in the intermediate state was quite similar to the measured scattering pattern at pH 6.5. The population of intermediate states was estimated to be 89.9% at pH 6.5 (Figure 11). This indicates that most species are in the intermediate states, and as a result it verifies the above discussions about the whole ensembles at pH 6.5.

We analyzed the SAXS data according to the three-state unfolding model. Exactly, however, there are four molecular components above pH 6.5, since the light chain will scatter X-rays independently of the heavy chain in the dissociated state. Therefore, we also analyzed the data using four SVD basis functions according to the following model expressed:

$$\mathrm{N} \overset{K_1}{\rightleftharpoons} \mathrm{H}_1 + \mathrm{L} \overset{K_2}{\rightleftharpoons} \mathrm{H}_2 + \mathrm{L} \qquad (12)$$

where H_1 and H_2 correspond to the heavy chain in the intermediate and the unfolded states, respectively, and L represents the light chain in the unfolded state. We could not, however, successfully calculate to extract the appropriate scattering profile of each chain.

Along with the SAXS data, the CD data were also analyzed by SVD, which showed a clear two-state transition behavior (data not shown). This indicates that the intermediate state described above cannot be distinguished from the completely unfolded state by CD despite the difference in the dissociation/association state of the two chains.

In the present study, we estimated the midpoint pH values and the Hill coefficients by various experimental methods (as listed in Table 1). The latter values differ more than the former ones depending on the experimental methods used. Considering the magnitudes of errors, however, we think the Hill coefficients had a constant value of about 5 irrespective of the methods, indicating that the simultaneous deprotonation of at least 5 protons contributed to the unfolding. Combining eq 3 with the relationship $\Delta G = -RT \ln K$, the free energy change of unfolding (ΔG) can be determined as

$$\Delta G = -2.303mRT(\mathrm{pH} - pK_a) \qquad (13)$$

where R and T mean the gas constant and the absolute temperature, respectively. At pH 4.5 and 310 K, the value of ΔG is about 40−50 kJ/mol, indicating that the native state was more stable by that value than the unfolded state. This is a reasonable value for the protein stability and is consistent with the value obtained for the denaturant-induced unfolding of aspergillopepsin II (K. Yokoyama et al., unpublished results).

ACKNOWLEDGMENTS

We thank Dr. Stephan Provencher for use of the CONTIN program. For SAXS data analysis, we used software developed in other laboratories along with our homemade program. We thank Dr. Kunihiro Kuwajima and Dr. Alexander Timchenko for the use of their programs. We are also grateful to Dr. Alexander Timchenko for reading the manuscript. We thank Dr. Harumi Fukada for helpful discussions about thermal unfolding of the enzyme.

REFERENCES

1. Takahashi, K., Inoue, H., Sakai, K., Kohama, T., Kitahara, S., Takishima, K., Tanji, M., Athauda, S. B. P., Takahashi, T., Akanuma, H., Mamiya, G., and Yamasaki, M. (1991) *J. Biol. Chem. 266*, 19480−19483.
2. Ogushi, M., and Wada, A. (1983) *FEBS Lett. 164*, 21−24.
3. Kataoka, M., Hagihara, Y., Mihara, K., and Goto, Y. (1993) *J. Mol. Biol. 229*, 591−596.
4. Goto, Y., and Nishikiori, S. (1991) *J. Mol. Biol. 222*, 679−686.
5. Nakayama, S., Nagashima, Y., Hoshino, M., Moriyama, A., Takahashi, K., Watanabe, T., and Yoshida, M. (1983) *J. Biochem.* (*Tokyo*) *95*, 465−475.
6. Semisotnov, G. V., Kihara, H., Kotova, N. V., Kimura, K., Amemiya, Y., Wakabayashi, K., Serdyuk, I. N., Timchenko, A. A., Chiba, K., Nikaido, K., Ikura, T., and Kuwajima, K. (1996) *J. Mol. Biol. 262*, 559−574.
7. Arai, M., Ikura, T., Semisotnov, G. V., Kihara, H., Amemiya, Y., and Kuwajima, K. (1998) *J. Mol. Biol. 275*, 149−162.
8. Takahashi, K. (1995) *Methods Enzymol. 248*, 146−155.
9. Provencher, S. W., and Glöckner, J. (1981) *Biochemistry 20*, 33−37.
10. Glatter, O., and Kratky, O. (1982) *Small-Angle X-ray Scattering*. Academic Press, New York.
11. Tanokura, M., Tasumi, M., and Miyazawa, T. (1976) *Biopolymers 15*, 393−401.
12. Chen, L., Hodgson, K. O., and Doniach, S. (1996) *J. Mol. Biol. 261*, 658−671.
13. Bilgin, N., Ehrenberg, M., Ebel, C., Zaccai, G., Sayers, Z., Koch, M. H. J., Svergun, D. I., Barberato, C., Volkov, V., Nissen, P., and Nyborg, J. (1998) *Biochemistry 37*, 8163−8172.
14. Segel, D. J., Fink, A. L., Hodgson, K. O., and Doniach, S. (1998) *Biochemistry 37*, 12443−12451.
15. Henry, E. R., and Hofrichter, J. (1992) *Methods Enzymol. 210*, 129−192.
16. Kojima, M., Tanokura, M., Muto, Y., Miyano, H., Suzuki, E., Hamaya, T., Takizawa, T., Kono, T., and Takahashi, K. (1995) in *Aspartic Proteinases*: *Structure, Function, Biology, and Biomedical Implications* (Takahashi, K., Ed.) pp 611−615, Plenum Press, New York.
17. Kataoka, M., Kuwajima, K., Tokunaga, F., and Goto, Y. (1997) *Protein Sci. 6*, 422−430.
18. Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T. (1989) *Numerical Recipes*: *The Art of Scientific Computing*. Cambridge University Press, Cambridge, England.
19. Fukada, H., Takahashi, K., Sorai, M., Kojima, M., Tanokura, M., and Takahashi, K. (1995) *Thermochim. Acta 267*, 373−378.

BI991584O