Detection of Intermediate Species in the Refolding of Bovine Trypsinogen†

Albert Light* and Jeffrey N. Higaki‡
Department of Chemistry, Purdue University, West Lafayette, Indiana 47907
Received August 25, 1986; Revised Manuscript Received April 28, 1987

ABSTRACT: The mixed disulfide of bovine trypsinogen and glutathione was refolded at pH 8.6 and 4 °C with a mixture of 3 mM cysteine and 1 mM cystine catalyzing disulfide interchange. The folding process was monitored by analysis of quenched samples with isoelectric focusing and size-exclusion chromatography. Isoelectric focusing showed a progressive change from a p*I* of 5.2 for the mixed disulfide derivative to a p*I* of 9.3 for native trypsinogen. A number of principal intermediates were detected as a function of the refolding time. These intermediates were also separated and further characterized by size-exclusion chromatography on columns of TSK G2000 SW operated in the high-performance liquid chromatographic mode. Rechromatography of a series of sequential fractions taken from the parental peak was necessary to resolve and characterize the principal intermediates. The loss of glutathione moieties produced a partly folded structure with an apparent hydrodynamic volume (Stokes radius, R_s) of 33.9 Å. These structures became compact with time, and more intermediates were detected between 33.9 and 29.2 Å. Finally, a change in conformation, resembling a two-state transition, changed the molecules of R_s 29.2 to the compact structure of native trypsinogen (22.4 Å). The rate of formation of the native structure was determined from the progress curves derived from isoelectric focusing and size-exclusion chromatography. The $t_{1/2}$ values calculated by these methods of analysis were consistent with a $t_{1/2}$ of 58 min as reported earlier from activity measurements and quantitative polyacrylamide gel electrophoresis [Odorzynski, T. W., & Light, A. (1979) *J. Biol. Chem. 254*, 4291-4295]. Size-exclusion chromatography and the rechromatography of consecutively collected peak fractions identified molecules with a single conformation volume if the rechromatographed fractions had the same retention time as the parental peak. If a mixture of conformations was present in the parental peak, rechromatography gave differing retention times across the peak. Size-exclusion chromatography can be helpful in judging protein purity (a single hydrodynamic volume) and in observing the changes in hydrodynamic volume during the refolding process. The large number of intermediate species seen early in the folding of bovine trypsinogen suggests that multiple pathways were followed rather than an ordered sequential pathway.

The in vitro refolding of protein molecules has been investigated with physical probes in order to detect intermediate species in the transition from the unfolded to the folded state (Kim & Baldwin, 1982). To accomplish this, denaturation–renaturation studies of well-characterized proteins were monitored with UV (Labhardt, 1984; Tsong et al., 1972; Hagerman & Baldwin, 1976), circular dichroism (Galat et al., 1985), fluorescence (Kelley et al., 1986; Zuniga, 1983), NMR (Brems & Baldwin, 1984; Osterhout et al., 1985; Blum et al., 1979), gel electrophoresis (Goldenberg & Creighton, 1984), and calorimetry (Privalov, 1979; Pfeil et al., 1986) and and by amide proton exchange (Schmid & Baldwin, 1979). These studies showed that the denaturation–renaturation process was not always a two-state transition; intermediate species could be detected in the transition zone.

The refolding of disulfide-containing proteins begins with the fully reduced molecule and progresses through a series of intermediate species until the globular structure of the native protein is reached (Anfinsen, 1973). During the folding process, the disulfide bonds stabilize partly folded molecules and increase the likelihood that very early as well as later intermediate species can be detected and isolated. Creighton (1978) used ion-exchange chromatography to separate the disulfide-containing intermediates found in refolding bovine pancreatic trypsin inhibitor. He proposed a folding pathway with an initial formation of one-disulfide-containing intermediates, followed by the formation of two-disulfide-containing intermediates, and finally the formation of native molecules. A similar approach with bovine ribonuclease showed that native disulfide pairings appeared late in the folding process and only after the refolded globular structure resembled the conformation of the native enzyme (Creighton, 1979).

Bovine trypsinogen has 229 amino acid residues and six disulfides (Dayhoff, 1972). The physical, chemical, and enzymatic properties of the zymogen are well documented (Walsh, 1970; Keil, 1971). In previous studies, we characterized the refolding of bovine trypsinogen as the mixed disulfide of glutathione (Odorzynski & Light, 1979). We found that intermediate species were produced as the refolding proceeded (Odorzynski, 1978). These studies used gel filtration chromatography (molecular sieving) and gave only a limited resolution of the large number of components that were present.[1] We were unable to characterize the intermediate species and their rates of appearance and disappearance or

† This investigation was supported by the National Institutes of Health (GM-22261) and by funds from the Department of Chemistry. This work was taken from a thesis (1986) by J.N.H. submitted to Purdue University in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

* Author to whom correspondence should be addressed.
‡ Present address: Department of Pharmaceutical Chemistry, University of California, San Francisco, CA 94143.

[1] Samples that were removed as a function of time were quenched with iodoacetate and separated on columns of Sephadex G-100 SF. Analysis on a Du Pont curve resolver showed that a large number of species were present between the hydrodynamic volume of the unfolded protein and that of native trypsinogen. The changes in the areas of the intermediate species as a function of the time of refolding showed increases and decreases in components and agreed with the expected precursor–product relationships.

0006-2960/87/0426-5556$01.50/0

to obtain the fractions in a highly purified state. Recently, Corbett and Roche (1984) reported differences in the Stokes radius of a protein molecule undergoing denaturation, using size-exclusion chromatography (molecular sieving) in the HPLC mode, and Brems et al. (1985) used the same technique in following the equilibrium denaturation of bovine growth hormone. Corbett and Roche (1984) suggested that size-exclusion chromatography should be applicable to volume changes associated with the unfolding and folding of globular proteins and permit the detection of stable intermediates.

We now report the use of size-exclusion chromatography to monitor the folding of trypsinogen. Although the resolving power was insufficient to separate the intermediate species in a single chromatographic step, a better resolution was obtained on rechromatography of small fractions obtained in the initial separation. It was thus possible to detect and characterize a number of intermediate species that appeared early, in the middle, and late in the folding process. The principal intermediate species were also detected by isoelectric focusing (Bouet et al., 1982; Trexler & Patthy, 1983); it showed a progressive increase in the isoelectric points with the time of folding.

Experimental Procedures

Materials. Bovine trypsinogen (1× crystallized), bovine chymotrypsinogen A (5× crystallized), and hen egg white lysozyme were purchased from Worthington. Bovine pancreatic ribonuclease (5× crystallized) was from Mann, bovine β-lactoglobulin (3× crystallized) was from Nutritional Biochemicals, human hemoglobin was from Schwarz/Mann, and yeast hexokinase was from Sigma. Crystalline bovine serum albumin was from Miles Laboratories. Bovine pancreatic trypsin inhibitor (Trasylol[2]) was a gift from Dr. E. Truescheit of the Bayer Werk Co. Chicken ovomucoid was a gift from Prof. M. Laskowski of Purdue University. Bovine enterokinase was partially purified on (diethylaminoethyl)cellulose by the procedure of Liepnieks and Light (1979). Dithioerythritol, iodoacetic acid, iodoacetamide, Gdn-HCl,[3] Tos-Arg-OMe, and glutathione were from Sigma. Ampholytes (pH 3.5–5.0, pH 4–6, pH 3.5–10) were purchased from LKB. Sephadex G25 (fine) was from Pharmacia. The Toya Soda TSK-G2000SW columns and GSWP guard column were from Varian.

GSSG was prepared as follows: 12.6 g (0.04 mmol) of glutathione (GSH) was suspended in 75 mL of water at pH 8.6. A 12.8-mL aliquot of 6% hydrogen peroxide (10% excess over the sulfhydryl concentration) was added, and the mixture was stirred at room temperature for 5 min. The product (<0.01% sulfhydryl) was concentrated to a viscous mass, redissolved in water, and evaporated again. The last steps were repeated 4 times, and finally the sample was lyophilized to dryness. The disulfide content (Zahler & Cleland, 1968) of the product was approximately 90% of theory.

Preparation of Trypsinogen-S-SG. The procedure was a modification of the description of Odorzynski and Light (1979). A 60-mg sample of trypsinogen was dissolved in 4.0 mL of 6 M Gdn-HCl, pH 3.0, containing 0.01 M DTE, and kept anaerobic at room temperature for 10 min. The pH was then raised to 8.5 with 150 μL of 2 M Tris, and the reduction of disulfides was carried out at 37 °C for 2 h under a nitrogen atmosphere. The mixture was added to 40 mL of 0.5 M Tris, pH 8.6, containing 6 M Gdn-HCl and 0.1 M GSSG, and kept at room temperature for 3 h with magnetic stirring. The pH was lowered to 3.0 with formic acid (88%) and the sample desalted on a 2.5 × 100 cm Sephadex G25 column equilibrated with 0.15 M acetic acid.

Refolding of Trypsinogen-S-SG. A 60-mg sample of trypsinogen-S-SG in 5 mL of 6 M Gdn-HCl, pH 3.0, was diluted with 0.05 M Tris, pH 8.6, containing 0.05 M calcium chloride, 3 mM cysteine, and 1 mM cystine (Odorzynski & Light, 1979). The final protein concentration was 20 μg/mL, and the concentration of Gdn-HCl was reduced to 0.01 M. The mixture was kept anaerobic under a nitrogen atmosphere at 4 °C. Samples of the reaction mixture were quenched with 2 M iodoacetic acid (25× molar excess over the thiol concentration) in 0.5 M Tris, pH 8.6. The mixture was stirred and kept at room temperature in the dark for 1 h. Samples were dialyzed at 4 °C against 1 mM HCl for 48 h and then lyophilized. The dried samples were dissolved in 1 mM HCl, containing 2 M urea, to a final concentration of 15 mg/mL. Any insoluble material was removed by centrifugation.

Activation of Zymogens. Samples of refolded trypsinogen, which had been previously dialyzed against 0.01 M Tris, pH 8.1, containing 0.02 M calcium chloride, were activated with bovine enterokinase (30 units of enzyme/50 μg of protein) at 37 °C for 90 min (Liepnieks & Light, 1979). Trypsin activity toward 0.01 M Tos-Arg-OMe was determined spectrophotometrically according to the procedure of Hummel (1959). A Gilford Model 252 spectrophotometer equipped with a Model 6051 chart recorder was used.

SE-HPLC. Size-exclusion chromatography was performed on Toya Soda TSK G2000SW columns with a Varian Model 5000 HPLC unit. Separations were monitored at 280 nm with a Gilson Model HM Holochrome detector placed in series in the effluent pathway. The retention time was recorded for the peak concentration of the eluting protein or intermediate species.

Samples (1–2.0 mg) of native trypsinogen, trypsinogen-S-SG, and protein standards were separated on a single 0.75 × 50 cm TSK-G2000SW column protected by a 0.75 × 7.5 cm GSWP guard column. The protein standards used were bovine serum albumin (M_r 68 000, R_s = 33.9), human hemoglobin (M_r 64 000, R_s = 33.2), chicken ovalbumin (M_r 45 000, R_s = 31.2), bovine trypsinogen (M_r 24 000, R_s = 22.4), pancreatic ribonuclease (M_r 13 700, R_s = 19.3), and Kunitz pancreatic trypsin inhibitor (M_r 6000). The R_s values (Stokes radii) were reported by Laurent and Killander (1964) and Corbett and Roche (1984) or estimated for other samples from extrapolation of a graphical representation of the R_s vs. M_r values. The solvent system was 0.1 or 0.5 M ammonium formate, pH 4.0, with or without the addition of 2 M urea. The retention time was the same at both buffer concentrations, indicating that ionic interactions were not a factor in the separation.

Refolded samples (1.5 mg) were chromatographed on two TSK-G2000SW columns placed in series, with 0.1 M ammonium formate, pH 4.0, containing 2 M urea. The urea was necessary to improve the recovery of partially refolded intermediates. Twenty consecutive 0.5-mL fractions of the parental peak (initial separation) were rechromatographed under the same conditions as above and with the sensitivity of the UV monitor increased 25-fold. Areas of the peaks were measured with a planimeter.

Isoelectric Focusing and SDS Gel Electrophoresis. Polyacrylamide isoelectric focusing gels (0.7 mm thick; pH

[2] Trasylol is a registered trademark of Bayer AG.

[3] Abbreviations: trypsinogen-S-SG, trypsinogen–glutathione mixed disulfide derivative; GSSG, glutathione disulfide; SDS, sodium dodecyl sulfate; DTE, dithioerythritol; Gdn-HCl, guanidine hydrochloride; SE-HPLC, size-exclusion high-performance liquid chromatography; Tos-Arg-OMe, N^{α}-tosyl-L-arginine methyl ester; R_s, Stokes radius in angstroms; Tris, tris(hydroxymethyl)aminomethane; CM, carboxymethyl.

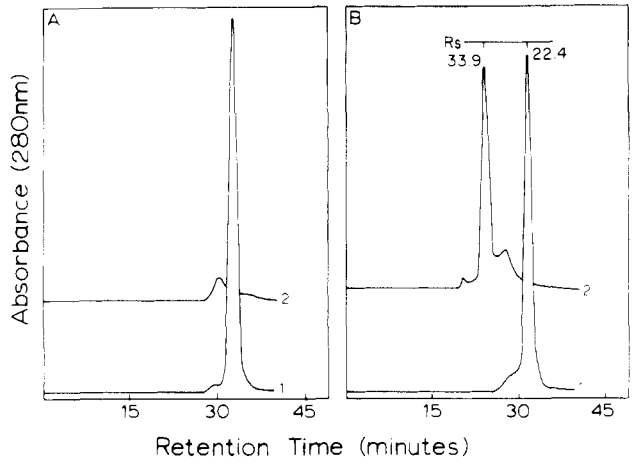

FIGURE 1: Recovery of trypsinogen and trypsinogen-S-SG on size-exclusion chromatography on columns of TSK-G2000SW. A 0.2-mg sample of trypsinogen and of trypsinogen-S-SG was applied to a 0.75 × 50 cm TSK-G2000SW column. Panel A, elution with 0.1 M ammonium formate, pH 4.0. Panel B, elution with 0.1 M ammonium formate, pH 4.0, containing 2 M urea. Flow rate was 0.5 mL/min. Curve 1, trypsinogen; curve 2, trypsinogen-S-SG. The calculated values of the hydrodynamic volumes (R_s) for trypsinogen-S-SG and trypsinogen are shown in panel B.

3.5–10.0) were prepared according to the *Hoefer Scientific Instruments Catalog* (1980). The gels were polymerized with 7.5% acrylamide, 0.2% bis(acrylamide), 0.15% pH 9–11 ampholytes, 2% pH 3.5–10.0 ampholytes, 5% glycerol, 2 M urea, and 0.03% ammonium persulfate. The urea was included to increase the solubility of partially folded intermediates. The buffer for the anode chamber was 1.0 M phosphoric acid, and the buffer at the cathode was 1.0 M sodium hydroxide. The gels were prefocused on a Hoefer HE 900 horizontal electrofocusing unit at 17 °C for 1 h at 10 W constant power programmed to an upper voltage limit of 80 V/cm. Refolded samples (7.5 μg) were focused at 1000 V for 1.5–2.0 h. Gels were stained with Coomassie Blue R-250. The isoelectric focusing gels were calibrated with a mixture of proteins of known isoelectric points (Pharmacia broad-range calibration kit). The gels were scanned with an E-C Apparatus densitometer, and the absorbance was recorded with a Hewlett-Packard Model 3392A integrator. Areas of the peaks were measured with a planimeter.

SDS–polyacrylamide gel electrophoresis was performed according to the procedure of Laemmli (1970).

Results

In the urea-induced unfolding of sperm whale myoglobin, Corbett and Roche (1984) showed that a displacement toward shorter elution times was observed on size-exclusion chromatography with increasing urea concentrations. We selected size-exclusion chromatography to follow the folding of trypsinogen-S-SG. In using a TSK-G2000SW column for our studies, we found a significant difference in the retention times of native trypsinogen and unfolded trypsinogen-S-SG (Figure 1A). However, only 20% of the trypsinogen-S-SG was recovered from the column (curve 2) unless 2 M urea was present in the eluting solvent; then the sample was recovered quantitatively (Figure 1B, curve 2). The native trypsinogen had the same retention time with or without urea present. Clearly, trypsinogen maintained its hydrodynamic volume and native conformation in 2 M urea, in agreement with the conclusions of Harris (1956) and Delaage and Lazdunski (1968), who showed that bovine trypsin and trypsinogen were stable in low concentrations of urea, including the 2 M urea used in these studies. The difference in retention times of the

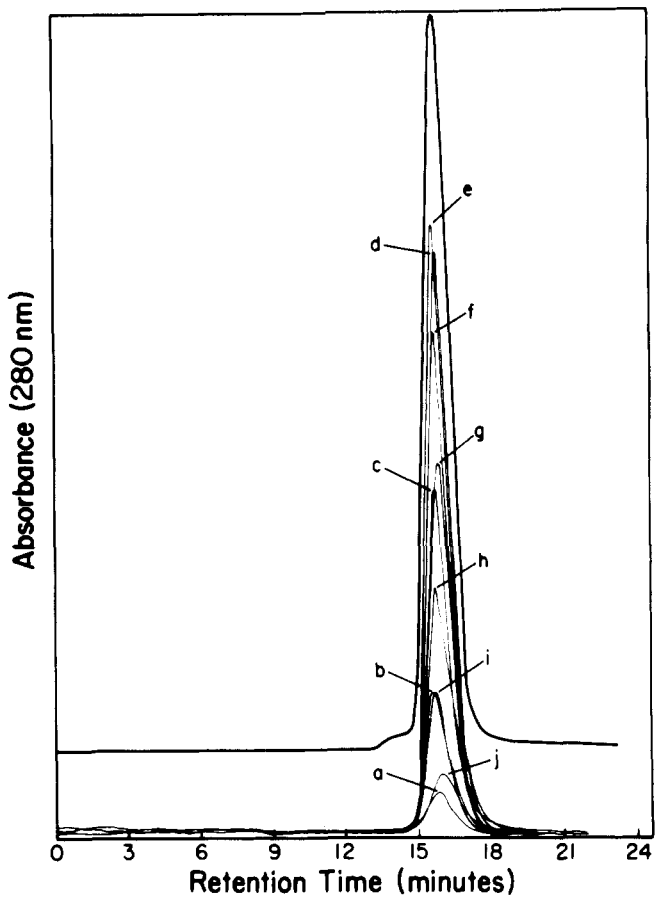

FIGURE 2: Size-exclusion chromatography of trypsinogen. A 0.5-mg sample of native trypsinogen was chromatographed on a column of TSK-G2000SW (0.75 × 50 cm) in the presence of 0.1 M ammonium formate, pH 4.0, containing 0.1 M sodium chloride. The flow rate was 1 mL/min. The parental peak (top profile) was divided into ten 0.5-mL fractions, and 100 μL of each fraction was rechromatographed under the same conditions with the full-scale absorbance sensitivity increased 10-fold (profiles a–j).

unfolded and folded proteins was sufficiently large to ensure that intermediate species with hydrodynamic volumes between the two would be detected on the TSK-G2000SW columns; this was also noted for the smaller sperm whale myoglobin in the studies of Corbett and Roche (1984).

A series of preliminary experiments were performed with the SE-HPLC system to examine the behavior of selected proteins that differed in size. Several samples were homogeneous and others were heterogeneous on analysis with SDS–polyacrylamide gel electrophoresis. A homogeneous preparation of bovine trypsinogen on a 50-cm column gave a single peak with a retention time of 15.7 min, at a flow rate of 1 mL/min (Figure 2). We divided the parental peak[4] into 10 equal fractions (fractions a–j); fraction a was the first fraction with the shorter retention time of the parental peak and fraction j had the largest retention time. These were individually rechromatographed under the same conditions used in the initial separation. Each rechromatographed fraction had the same retention time as the parental peak (Figure 2). With a homogeneous protein, the fractions across the peak had a constant retention time. In contrast, if the sample was heterogeneous, fractions across the parental peak each differed in their retention times on rechromatography (Figure 3).

[4] The original peak from the initial separation will be referred to as the parental peak.

Table I: Size-Exclusion Chromatography of Proteins and Rechromatography of Peak Fractions[a]

sample	P^b (min)	A^b (min)	SD[b] (min)
chymotrypsinogen	16.1	16.1	0.06
trypsinogen	15.9	15.9	0.07
ribonuclease	16.6	16.6	0.07
pancreatic trypsin inhibitor	18.2	18.6	0.18
ovomucoid	13.6	13.6	0.21
lysozyme	18.4	19.2	0.26
β-lactoglobulin	14.1	14.8	0.46
hemoglobin	15.4	15.1	0.54
hexokinase	12.8	12.8	0.93

[a] A 1–2-mg sample in 0.1 M ammonium formate, pH 4.0, was separated on a TSK-G2000SW column (0.75 × 50 cm) at a flow rate of 1 mL/min. The parental peak was divided into 10–20 sequential fractions and rechromatographed at a flow rate of 1 mL/min. [b] Column 2 is the retention time of the parental peak (P), and column 3 is the average retention time (A) of the rechromatographed fractions taken from the parental peak, after correction for the difference in the flow rate. Column 4 is the standard deviation (SD) from the mean (column 3).

Table I summarizes the properties of a few selected proteins after consecutive fractions of the parental peak were rechromatographed on columns of TSK-G2000SW. The retention times of the parental peak (P) are listed in column 2. The retention times from each set of rechromatographed fractions were averaged (A) (column 3), and their standard deviation from the mean (SD) was calculated and listed in column 4. The standard deviation was a quantitative measure of the spread of the retention times and was an expression of the complexity of the sample. The highly purified bovine chymotrypsinogen, trypsinogen, and ribonuclease had standard deviations of 0.059–0.070. In contrast, bovine pancreatic trypsin inhibitor, chicken ovomucoid, hen egg white lysozyme, β-lactoglobulin, and human hemoglobin had standard deviations of 0.179–0.455. The group of proteins behaved as if they were heterogeneous samples.[5] The higher values of the standard deviation reflected a greater complexity of the sample. The standard deviation (0.93) found for the sample of hexokinase used in these experiments is in agreement with the observed heterogeneity of the sample on SDS gel electrophoresis and in the complex elution profile in the initial separation with SE-HPLC (Figure 3).

Refolding Monitored with SE-HPLC. We followed the progress of the folding of trypsinogen-S-SG by removing samples as a function of time. Further folding was prevented by inhibiting disulfide interchange on converting the thiol catalyst to the inactive *S*-carboxymethyl derivative (Odorzynski & Light, 1979; Creighton, 1984). Quenched samples representing folding times from 5 s to 180 min were applied to two 50-cm TSK-G2000SW columns arranged in series (Figure 4). Native trypsinogen had a retention time of 54.4 min, and the size of the peak increased as folding continued. Trypsinogen-S-SG had a retention time of 42.5 min, and the molecule rapidly decreased in hydrodynamic volume in the first few minutes of folding. We also noted aggregates or insoluble particles at a retention time of 34.5 min, and these increased with time. The components eluting between 42.5 min (unfolded molecules) and 54.4 min (native trypsinogen) corresponded to partially folded species. These structures were stable in the buffer system containing 2 M urea since rechromatography of the fractions gave indentical retention times

[5] The apparent heterogeneity of the sample may not reflect the presence of contaminants but may be caused by multiple components differing in hydrodynamic volumes because of altered conformations under the conditions used in the separations.

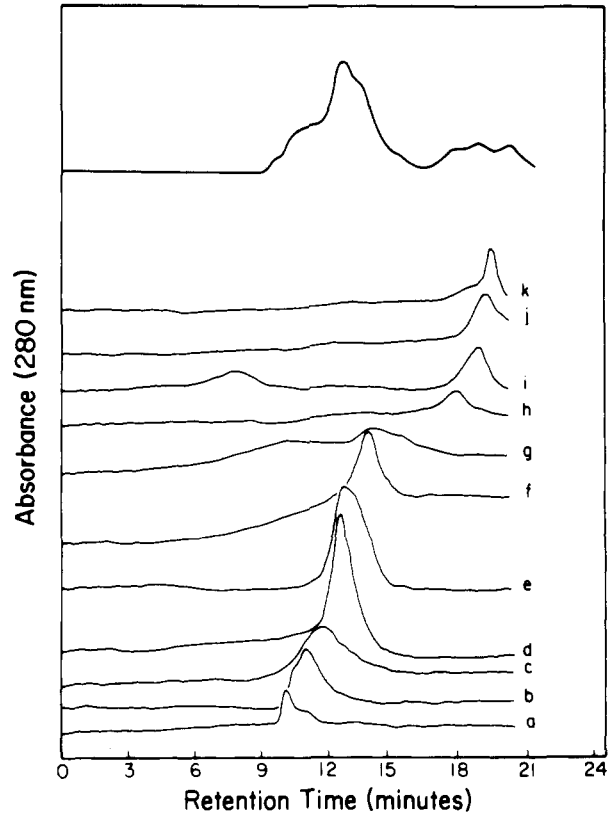

FIGURE 3: Size-exclusion chromatography of yeast hexokinase. A 0.5-mg sample of hexokinase was chromatographed on TSK-G2000SW (0.75 × 50 cm) in the presence of 0.1 M ammonium formate, pH 4.0, containing 0.1 M sodium chloride. The flow rate was 1 mL/min. The parental peak was divided into eleven 0.5-mL fractions, and 100 μL was rechromatographed under the same conditions (profiles a–k).

with the values they had in the parental peak (see below). This is the same behavior as noted before for native trypsinogen which was stable in 2 M urea and had the same retention time on rechromatography (Figure 1).

Each parental peak from 5 s to 5 h of folding was divided into a number of consecutive fractions for rechromatography. Only a few time periods will be described in detail. The rechromatography of samples taken at 10 min (an early time in the folding process) and at 180 min (late time period, at three half-lives) is shown in Figures 5 and 6, respectively. The rechromatographed fractions taken from the 10-min refolded sample each differed in their retention times, but each had the same retention time as it had in the parental peak. This suggested that the 10-min sample was a complex mixture containing a population of different refolded structures with hydrodynamic volumes ranging from the denatured to the compact native state. Even after 10 min of folding, native trypsinogen (fraction 16) was present in small amounts. Fraction 15 was collected from that region of the parental peak that was in the valley between the peak for fraction 14 (R_s of 29.2 Å) and native trypsinogen (fraction 16). On rechromatography, fraction 15 was a broad peak and appeared to be a mixture of equal amounts of the R_s = 29.2 Å component and native trypsinogen (R_s of 22.4 Å). The SE-HPLC of samples refolded longer than 10 min also showed a range of components differing in hydrodynamic volumes, but the species with large R_s values were in small amounts and more compact structures were the dominant species. The size of the native trypsinogen fraction continually increased.

Only a few components were detected in fractions taken from a sample refolded for 180 min (Figure 6). Fractions

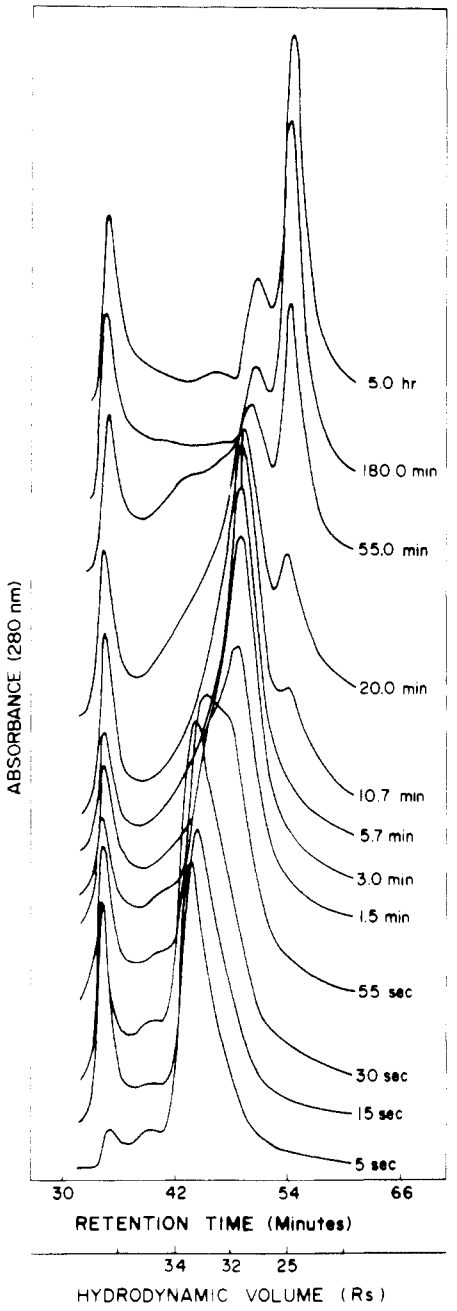

FIGURE 4: Size-exclusion chromatography of folded trypsinogen-S-SG. A 1.5-mg sample at the indicated times was separated on two TSK-G2000SW columns run in series. The buffer was 0.1 M ammonium formate, pH 4.0, containing 2 M urea. The flow rate was 0.5 mL/min. The calculated values of the hydrodynamic volumes are shown on the scale at the bottom.

14–20 (principal components) each had retention times of 54.4 min and therefore could be identified as native trypsinogen. Fractions 12 and 13 (minor fractions) had retention times of 50.7 min, and fractions 9–11 (in very low amount) had retention times of 49.1 min. These retention times suggested that even after a sample was refolded for 180 min, in addition to the high yield of native trypsinogen, species with larger hydrodynamic volumes could still be detected but in minor amount. These minor components may be dead-end structures and intermediates with relatively slow rates of folding.

The types of changes taking place in the conformation of the intermediate species as a function of time were examined by arbitrarily dividing the intermediates into four principal groups, differing in hydrodynamic volumes (Table II). Intermediates corresponding to an apparent R_s of 33.2 Å ap-

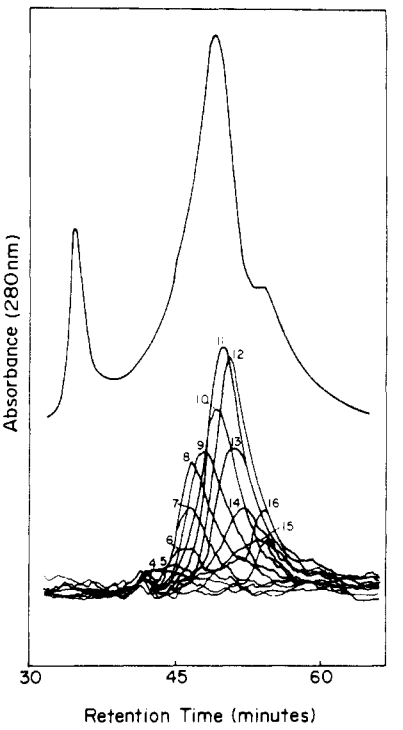

FIGURE 5: Size-exclusion chromatography of refolded trypsinogen after 10 min of folding. The parental peak was divided into 0.5-mL fractions, and 100 μL was rechromatographed. The upper curve is the parental peak, and the rechromatographed fractions are presented as overlapping peaks shown in the lower part of the figure.

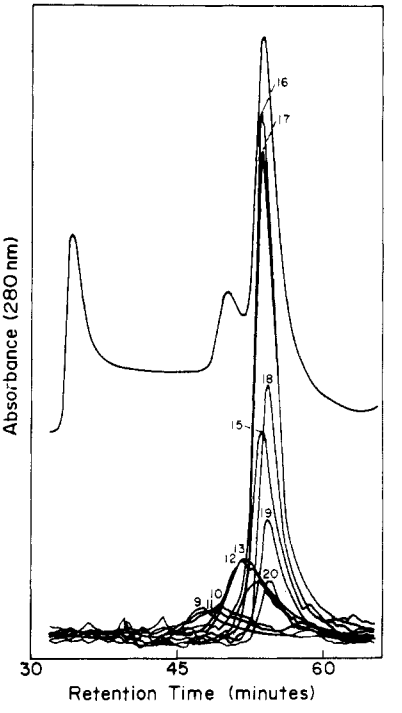

FIGURE 6: Size-exclusion chromatography of refolded trypsinogen after 180 min. A sample was divided into 0.5-mL fractions and 100 μL was rechromatographed. The upper curve is the parental peak, and the rechromatographed fractions are presented as overlapping peaks shown in the lower part of the figure.

peared within the first 5 s and then rapidly decreased in amount in the next 60 s. These initial structures were converted to a group of intermediates with a Stokes radius between 33.2 and 30.8 Å. In the next 10 min, intermediates also appeared with an R_s of 30.0–29.2 Å. Native trypsinogen, with an R_s of 22.4 Å, appeared at low levels after 10 min of folding

Table II: Characterization of the Principal Intermediate Species by Size-Exclusion Chromatography and Isoelectric Focusing[a]

time interval	hydrodynamic vol[b] (app R_s)	IEF band	isoelectric point
0	33.9	U	5.20
0–5 s	33.2	A, B, C	5.20–5.80
15–30 s	33.0	D, E	6.30–6.75
55–90 s	30.8	F, G	7.05–7.75
3–10 min	29.2	H, I	8.10–8.55
10 min–5 h	22.4	J, K, N	8.60–9.30

[a] Trypsinogen-S-SG was folded at a concentration of 20 μg/mL at pH 8.6, in the presence of 2 M urea, 3 mM cysteine, and 1 mM cystine, at 4 °C, under a nitrogen atmosphere. Samples were removed at timed intervals, quenched with 2 M iodoacetate, dialyzed, lyophilized, and analyzed by size-exclusion chromatography and isoelectric focusing. [b] The hydrodynamic volume is given as the Stokes radius (R_s) in angstroms. The R_s was estimated from a calibration curve of proteins of known R_s plotted against their retention times.

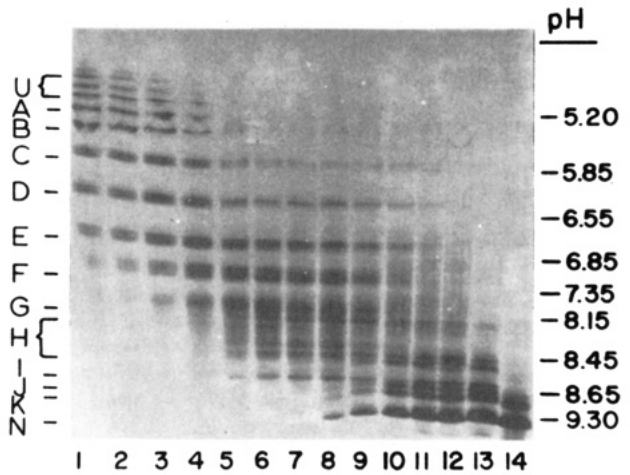

FIGURE 7: Isoelectric focusing of samples of refolded trypsinogen. A 7.5-μg sample was applied to the isoelectric focusing gel (pH 3.5–10.0) containing 2 M urea. The gel was focused at 17 °C for 1.5 h, at 1000 V. Lanes 1–12, samples taken at 5 s, 15 s, 30 s, 55 s, 1.5 min, 3.0 min, 5.7 min, 10.7 min, 20 min, 55 min, 180 min, and 5 h. Lane 13, trypsinogen subjected to the same manipulations of refolding. Lane 14, native trypsinogen.

and then increased to a maximum in 180 min. While intermediates were detected within the R_s range of 33.9–29.2 Å, intermediates were not found between 29.2 and 22.4 Å (Figure 5, fraction 15).

Refolding Monitored by Isoelectric Focusing. We also analyzed samples of refolded trypsinogen by isoelectric focusing (Figure 7). Trypsinogen-S-SG (p*I* 5.2) was well separated from native trypsinogen (p*I* 9.3). A number of components in the p*I* range of 5.2–9.3 were detected as the trypsinogen-S-SG folded. These intermediate species are designated as A–K. U is the position of trypsinogen-S-SG (p*I* 5.2) and N is native trypsinogen (Figure 7). The sample of trypsinogen-S-SG used in these refolding studies had a strong band at a p*I* of 5.2 and weak bands at a slightly higher p*I* (data not shown, but similar to Figure 7, lane 1, bands U). The difference in p*I* of the bands in the starting sample is approximately one negative charge, suggesting that the sample of trypsinogen-S-SG also contained small amounts of the mixed disulfide with 11 and 10 glutathione moieties. In the samples taken at early times of folding (lanes 1–5), the loss of the mixed disulfide was sequential, indicating that the protein disulfides initially formed one by one. Native trypsinogen appeared after 10 min, and the amount increased with time.

Table II compares the principal components detected by isoelectric focusing and by SE-HPLC. After 5 s, only the

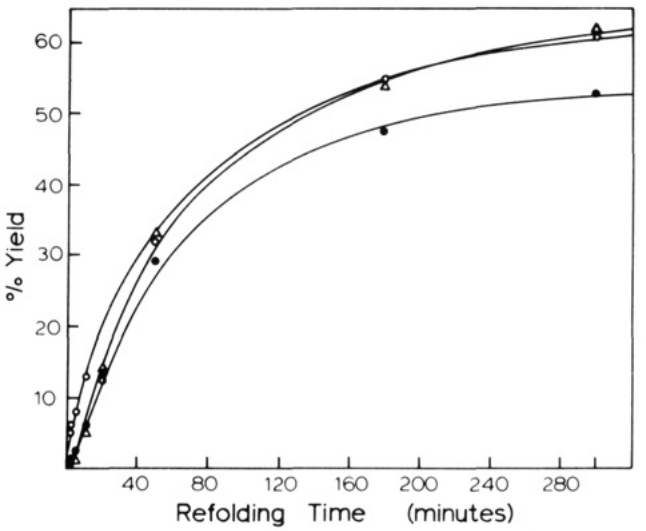

FIGURE 8: Rate of appearance of regenerated trypsinogen from analysis with enzymatic activity, size-exclusion chromatography, and isoelectric focusing. Samples analyzed after activation with enterokinase: tryptic activity measured (●); SE-HPLC (○); isoelectric focusing (△).

unfolded protein (33.9 Å) and bands U, A, B, and C were detected. These species were most likely trypsinogen-S-SG and loosely folded structures with one, two, or more protein disulfides. In the next 90 s, intermediates were found with R_s values varying from 33.0 to 30.8 Å and with p*I* values ranging from 6.30 to 7.75 (bands D–G). These early changes in p*I* were consistent with the loss of negative charges from the glutathione moieties (Odorzynski & Light, 1979). The same intermediate species seen in Figure 7 were observed (data not shown) when iodoacetamide was substituted for iodoacetate to quench the folding process; therefore, the different p*I* values were not caused by alkylation of the protein.

In an earlier study, we detected only low levels of protein thiols during the refolding of trypsinogen-S-SG (Odorzynski, 1978). With [^{14}C]iodoacetate as the alkylating agent, approximately 0.1–0.2 residues of (carboxymethyl)cysteine per molecule of protein were detected. However, we also noted the formation of $(SH)_2$-(179,203)-trypsinogen, which was converted to $(CM)_2$-trypsinogen when iodoacetate was added (Odorzynski & Light, 1979). The origin of the small amount of $(SH)_2$-(179,203)-trypsinogen is caused by the reduction of the exposed disulfide band by cysteine in the native protein. Presumably, the band at p*I* 8.6 (Figure 7, lane 14) may be the bis(carboxymethyl) derivative.

The early intermediates with R_s values of 33.0 and 30.8 Å and p*I*'s of 6.30–7.75 (Figure 7; Table II) were precursors of intermediates with an R_s of 29.2 Å and p*I*'s ranging from 8.10 to 8.55 (Figure 7, bands H and I). Finally, the native structure accumulated with an R_s of 22.4 Å, and bands J and K appeared with p*I*'s of 8.60 and 8.75 and band N with a p*I* of 9.30. The 8.75 component contained Ser-neotrypsinogen,[6] and 9.30 was native trypsinogen.

The progress curves for the regeneration of native trypsinogen calculated from three methods of analysis, namely, enzyme activity, size-exclusion chromatography, and isoelectric

[6] Ser-neotrypsinogen is formed from native trypsinogen on proteolysis of Lys-131–Ser-132. The susceptible peptide bond is in the region of the molecule linking the amino- and carboxy-terminal domains (Higaki & Light, 1985). The two-chain Ser-neotrypsinogen is covalently linked by disulfide bonds. Val-neotrypsinogen is formed on proteolysis of Arg-105–Val-106, and the two-chain Val-neotrypsinogen is also linked by disulfide bonds. The susceptible peptide bond is in the amino-terminal domain (Higaki & Light, 1985).

Table III: Rate of Folding of Trypsinogen-S-SG from Three Methods of Analysis[a]

method of analysis	k (s^{-1})	$t_{1/2}$ (min)
isoelectric focusing	1.88×10^{-4}	61.4
SE-HPLC	1.99×10^{-4}	58.0
enzymatic activity	2.14×10^{-4}	54.0

[a] Rate constants and half-times were determined by linear regression analysis of the first-order rate plots for the appearance of native trypsinogen by isoelectric focusing, SE-HPLC, and activity toward Tos-Arg-OMe.

focusing, are shown in Figure 8. The three are in excellent agreement with one another. Any one of these analytical procedures gave reliable results for monitoring the folding process. Furthermore, the rate constants and half-times (Table III) calculated from the first-order plots of the progress curves also agreed with one another and with the values previously determined (Odorzynski & Light, 1979).

Discussion

The folding of trypsinogen-S-SG was monitored in these studies by the techniques of size-exclusion chromatography and isoelectric focusing. We found size-exclusion chromatography (SE-HPLC) to have a number of distinct advantages compared with other methods used previously to follow protein folding (Creighton, 1978). The principal advantage was that the separation was based on differences in the hydrodynamic volume (Stokes radius). This meant that we immediately could interpret the order of appearance of species on SE-HPLC as the protein folded from the random coil to the native structure. The order ranged from molecules with large molecular radii, appearing early in the folding process, to more compact, globular structures, appearing in the middle time periods, and, lastly, to the compact structure of the native molecule (Figure 4). A quantitative analysis of the folding process established the rates for the disappearance and appearance of denatured trypsinogen and native trypsinogen, respectively, and the rates of formation and disappearance of a number of intermediate species.[7] Intermediates were identified as components that increased and then decreased in concentration with time. We detected these intermediates because they had a distinctive retention time on SE-HPLC. We estimated their hydrodynamic volumes from the retention time found for a group of selected proteins of known Stokes radius on a column of TSK-G2000SW.

The identification and characterization of intermediates on SE-HPLC required rechromatography of small fractions taken from the initial separation of the partially folded sample. Rechromatography was necessary since the initial chromatographic separation (parental peak) only gave a partial resolution (Figure 4) and was too complex a mixture to interpret further (Odorzynski, 1978). After rechromatography, the retention times of the principal intermediates and the amount of each were determined (Figure 5). A large number of intermediate species were identified with R_s values ranging from 33.9 to 29.2 Å (Table II).

The folding process was also followed as a function of time by isoelectric focusing (Figure 7). The partially folded molecules were resolved into a number of bands differing in their isoelectric points (Table II). The species with low pI's appeared early. These were followed by species that showed a progressive increase in pI. Native trypsinogen was the last to form with a pI of 9.3. In addition, we observed the presence of Ser-neotrypsinogen at a pI of 8.75.[6] Both Ser- and Val-neotrypsinogen were present in the samples of trypsinogen used in these studies (Higaki & Light, 1985), and the isoelectric focusing results were our first indication that Ser-neotrypsinogen could be refolded from the mixed disulfide derivative (Light & Higaki, 1986).[8]

Changes in conformation of the globular structure during folding produced the intermediates that we observed on isoelectric focusing. A similar range of pI values was noted previously by Bouet et al. (1982) in detecting intermediates in the refolding of a reduced snake neurotoxin, erabutoxin *b*. Since the pI is primarily dependent upon the distribution of charged groups on the surface of the molecule, changes in conformation would be expected to alter their number. As the local environment changed in the vicinity of the charged groups, the net charge of the molecule, and therefore the isoelectric point, varied.

The first partly folded structures (R_s = 33.2 Å) that rapidly formed from trypsinogen-S-SG were slightly more compact than trypsinogen-S-SG itself (R_s = 33.9 Å). Most likely, the formation of local regions of secondary structure (α-helices, β-bends, and β-sheets) accounted for the change (Finkelstein & Ptitsyn, 1976; Richardson, 1981; Taylor & Thornton, 1984). The early formation of secondary structure had been noted before; for example, residues 18–35 of bovine pancreatic trypsin inhibitor formed an antiparallel β-structure (Chou et al., 1985).

After the polypeptide chain acquired a partial secondary structure, half-cystine residues would then be physically close to one another and disulfide bonds would form. In earlier studies on the folding of trypsinogen-S-SG, we showed a rapid loss of the mixed disulfide and the early formation of protein disulfides (Odorzynski & Light, 1979). Since early intermediates were loose and unstable structures with large hydrodynamic volumes, nonnative disulfide bonds were most likely the first to form. As folding continued, the changes in hydrodynamic volume and isoelectric point suggested that the disulfide pairing underwent continual changes with alterations in the conformation and surface charges. The molecules then formed more compact structures with higher isoelectric points. Since the rate of these changes was limited by disulfide interchange (Creighton, 1978; Freedman, 1984), clearly, the conformations had to be stabilized by many different disulfide pairings.

The accumulation of the species with an R_s of 29.2 Å late in the folding process and the absence of species with R_s values between 29.2 and 22.4 Å (native trypsinogen) suggested that the globular structure reached a conformation where, in a single rate-limiting step, the 29.2-Å conformation was converted to the native molecule. This conversion resembled a two-state cooperative change in conformation and produced the more compact native trypsinogen.

The large number of species detected suggests that the folding pathway differs from the ordered pathway of bovine pancreatic trypsin inhibitor (Creighton, 1978, 1979). Creighton proposed a folding pathway with the initial formation of one-disulfide-containing intermediates, followed by the formation of two-disulfide intermediates, and finally the formation of native disulfide pairing and native molecules. Similar studies of disulfide pairing in the refolding of bovine ribonuclease did not show a sequential formation of disulfides

[7] A. Light, T. W. Odorzynski, and J. N. Higaki, unpublished results.

[8] Val-neotrypsinogen with a peptide bond cleaved in the amino-terminal domain could not be refolded as the mixed disulfide derivative, and thus, it was not found in refolded samples examined by isoelectric focusing.

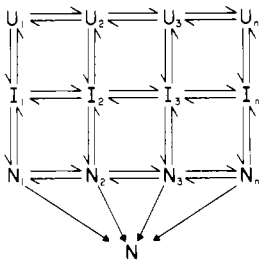

FIGURE 9: Proposed folding pathway of serine proteinases. Initially the random coil structures are partly folded (U_1–U_n) and capable of changing from one conformation to another. These structures continue to fold, and disulfide pairing stabilizes the intermediates (I_1–I_n). These globular intermediates on further folding become more compact with an R_s of 29.2 (N_1–N_n). A final two-step conversion produces the native molecule (N).

(Creighton, 1979). Instead, native disulfide pairings only appeared when the refolded globular structure resembled the conformation of the native enzyme. The pathway for trypsinogen resembles the general features of the folding of ribonuclease (Creighton, 1979) and of hen egg white lysozyme (Acharya & Taniuchi, 1976, 1980, 1982). In the folding process, trypsinogen, ribonuclease, and lysozyme apparently only form native disulfide bonds late in the folding pathway at the time the native molecule is produced.

It is likely that multiple pathways were followed in refolding trypsinogen (Figure 9). This is to say that in the initial folding of the random coil molecules many different initial structures are possible (U_1–U_n) rather than a single preferred nucleation event. These weak interactions are rapidly lost, followed by new interactions that continue to change until a globular structure is produced (I_1–I_n). These changes in structure and the associated hydrodynamic volumes agree with the results of analysis with SE-HPLC and isoelectric focusing. In our previous folding studies on selectively reduced and carboxymethylated trypsinogen (CM_2)-(179,203)-trypsinogen and ($CM)_4$-(122,179,189,203)-trypsinogen), we suggested at that time that alternate pathways must be followed in order to regain the native structure when one or two disulfides could no longer form and stabilize intermediate structures (Odorzynski & Light, 1979; Light & Odorzynski, 1979). The argument for multiple pathways in protein folding has been eloquently described by Harrison and Durbin (1985). The large number of intermediate species detected in the folding of trypsinogen also argues for multiple pathways in the initial folding process. However, at later stages of folding when the molecules become compact, the pathway may be more restrictive as seen in the two-step transition from molecules of R_s = 29.2 Å (N_1–N_n) to those of the native structure (N).

The multiple pathway for the folding of trypsinogen would be more rapid and presumably simpler in the number of structures formed if folding occurred at the domain level. This was a suggestion made much earlier by Wetlaufer (1973), who argued for independent folding of domains in the serine proteinase family. In agreement with this proposal, we showed in recent studies (Light et al., 1986) with neochymotrypsinogen (Duda & Light, 1982) and neotrypsinogen (Higaki & Light, 1986) that the folding does indeed proceed at the domain level.

Registry No. Trypsinogen, 9002-08-8; chymotrypsinogen, 9035-75-0; ribonuclease, 9001-99-4; pancreatic trypsin inhibitor, 9087-70-1; lysozyme, 9001-63-2; hexokinase, 9001-51-8; serine proteinase, 37259-58-8.

REFERENCES

Acharya, A. S., & Taniuchi, H. (1976) *J. Biol. Chem. 251*, 6934–6946.

Acharya, A. S., & Taniuchi, H. (1980) *J. Biol. Chem. 255*, 1905–1911.

Acharya, A. S., & Taniuchi, H. (1982) *Mol. Cell. Biochem. 44*, 129–148.

Anfinsen, C. B. (1973) *Science (Washington, D.C.) 181*, 223–230.

Blum, A. D., Smallcombe, S. H., & Baldwin, R. L. (1978) *J. Mol. Biol. 118*, 305–316.

Bouet, F., Menez, A., Hider, R. C., & Fromageot, P. (1982) *Biochem. J. 201*, 495–499.

Brems, D. N., & Baldwin, R. L. (1984) *J. Mol. Biol. 180*, 1141–1156.

Brems, D. N., Plaisted, S. M., Havel, H. A., Kauffman, E. W., Stodola, J. D., Eaton, L. C., & White, R. D. (1985) *Biochemistry 24*, 7662–7668.

Chou, K. C., Nemethy, G., Pottle, M. S., & Scheraga, H. A. (1985) *Biochemistry 24*, 7948–7953.

Corbett, R. J. T., & Roche, R. S. (1984) *Biochemistry 23*, 1888–1894.

Creighton, T. E. (1978) *Prog. Biophys. Mol. Biol. 33*, 231–297.

Creighton, T. E. (1979) *J. Mol. Biol. 129*, 411–431.

Creighton, T. E. (1984) *Methods Enzymol. 107*, 305–329.

Dayhoff, M. O. (1972) *Atlas of Protein Sequence and Structure*, Vol. 5, National Biomedical Research Foundation, Silver Spring, MD.

Delaage, M., & Lazdunski, M. (1968) *Eur. J. Biochem. 4*, 378–384.

Duda, C. T., & Light, A. (1982) *J. Biol. Chem. 257*, 9866–9871.

Finkelstein, A. V., & Ptitsyn, O. B. (1976) *J. Mol. Biol. 103*, 15–24.

Freedman, R. (1984) *Trends Biochem. Sci. (Pers. Ed.) 9*, 438–441.

Galat, A., Yang, C-C., & Blout, E. R. (1985) *Biochemistry 24*, 5678–5685.

Goldenberg, D. P., & Creighton, T. E. (1984) *Anal. Biochem. 138*, 1–18.

Hagerman, P. J., & Baldwin, R. L. (1976) *Biochemistry 15*, 1462–1473.

Harris, J. I. (1956) *Nature (London) 177*, 471–473.

Harrison, S. C., & Durbin, R. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 4028–4030.

Higaki, J. N., & Light, A. (1985) *Anal. Biochem. 148*, 111–120.

Higaki, J. N., & Light, A. (1986) *J. Biol. Chem. 261*, 10606–10609.

Hoefer Scientific Instruments Catalog (1980) Vol. 1, pp 88–89, Hoefer Scientific Instruments, San Francisco.

Hummel, B. C. W. (1959) *Can. J. Biochem. Physiol. 37*, 1393–1399.

Keil, B. (1971) *Enzymes (3rd Ed.) 3*, 250–275.

Kelley, R. F., Wilson, J., Bryant, C., & Stellwagen, E. (1986) *Biochemistry 25*, 728–732.

Kim, P. S., & Baldwin, R. L. (1982) *Annu. Rev. Biochem. 51*, 459–489.

Labhardt, A. M. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 7674–7678.

Laemmli, U. K. (1979) *Nature (London) 227*, 680–685.

Liepnieks, J. J., & Light, A. (1979) *J. Biol. Chem. 254*, 1677–1683.

Light, A., & Odorzynski, T. W. (1979) *J. Biol. Chem. 254*, 9162–9166.

Light, A., Duda, C. T., Odorzynski, T. W., & Moore, W. G. I. (1986) *J. Cell. Biochem. 26*, 163-170.
Odorzynski, T. W. (1978) Ph.D. Thesis, Purdue University.
Odorzynski, T. W., & Light, A. (1979) *J. Biol. Chem. 254*, 4291-4295.
Osterhout, J. J., Jr., Muthukrishnan, K., & Nall, B. T. (1985) *Biochemistry 24*, 6680-6684.
Pfeil, W., Bychkova, V. E., & Ptitsyn, O. B. (1986) *FEBS Lett. 198*, 287-291.
Privalov, P. L. (1979) *Adv. Protein Chem. 33*, 167-241.
Richardson, J. S. (1981) *Adv. Protein Chem. 34*, 167-339.
Schmid, F. X., & Baldwin, R. L. (1979) *J. Mol. Biol. 135*, 199-215.
Taylor, W. R., & Thornton, J. M. (1984) *J. Mol. Biol. 173*, 487-514.
Trexler, M., & Patthy, L. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 2457-2461.
Tsong, T. Y., Baldwin, R. L., & McPhie, P. (1972) *J. Mol. Biol. 63*, 453-475.
Walsh, K. A. (1970) *Methods Enzymol. 19*, 41-63.
Wetlaufer, D. B. (1973) *Proc. Natl. Acad. Sci. U.S.A. 70*, 697-701.
Zahler, W. L., & Cleland, W. W. (1968) *J. Biol. Chem. 243*, 716-719.
Zuniga, E. H., & Nall, B. T. (1983) *Biochemistry 22*, 1430-1437.

Calcium-Sensitive Thermal Transitions and Domain Structure of Human Complement Subcomponent C1r[†]

T. F. Busby and K. C. Ingham*

Biochemistry Laboratory, American Red Cross Biomedical Research and Development, Rockville, Maryland 20855

Received February 17, 1987; Revised Manuscript Received April 27, 1987

ABSTRACT: Fluorescent probes and other methods have been used to investigate the thermal stability of activated C1r and functionally intact fragments isolated from tryptic digests of the protein. This enzyme exhibits two irreversible transitions that differ with respect to their sensitivity to metal ions. The high-temperature transition occurs with a midpoint near 53 °C in 0.02 M tris(hydroxymethyl)aminomethane buffer and 0.15 M NaCl, pH 7.4. It is relatively insensitive to Ca^{2+} and ionic strength and is accompanied by a loss of catalytic activity. The low-temperature transition is most easily observed in the presence of ethylenediaminetetraacetic acid and is completely abolished by 100 μM Ca^{2+}. Its midpoint varies between 26 °C at low ionic strength and 40 °C in the presence of 0.5 M NaCl. The low-temperature transition results in extensive polymerization of the protein without loss of the esterolytic activity or the ability to react with C1 inhibitor; however, the ability to reconstitute hemolytically active $C\bar{1}$ or even bind to $C1\bar{s}$ in the presence of Ca^{2+} is destroyed. A highly purified N-terminal fragment generated by tryptic digestion of $C1\bar{r}$ in the presence of Ca^{2+} retained its ability to interact with $C1\bar{s}$, disrupting the formation of $C1\bar{s}$ dimers in the presence of Ca^{2+}. In the absence of Ca^{2+}, this fragment displays only a low-temperature transition that is very similar to the one observed with the whole protein and that destroys its ability to bind to $C1\bar{s}$. Addition of Ca^{2+} stabilizes this fragment, shifting the midpoint of its melting transition upward by more than 20 °C. Assignment of the high-temperature transition in $C1\bar{r}$ to the catalytic domain was confirmed by observation of a similar transition in a catalytically active C-terminal fragment obtained from the same tryptic digest. The occurrence of independent thermal transitions in human $C1\bar{r}$ and their assignment to specific regions of the molecule, together with the direct evidence that the N-terminal Ca^{2+}-sensitive domain is the one that interacts with $C1\bar{s}$, provide strong support for current models of the domain structure of these proteins.

The first component of complement C1[1] is a large multiprotein complex comprised of three subcomponents designated C1q, C1r, and C1s [reviewed by Cooper (1985)]. Each of these contains multiple domains that can be visualized in the electron microscope and that appear to be responsible for various functions of the complex. Some domains are involved in the Ca^{2+}-dependent associations that hold the subcomponents together while others are responsible for the interaction of C1 with immune complexes or cellular receptors. Still other domains possess the catalytic sites responsible for the proteolytic action of $C\bar{1}$ and for its interaction with C1 inhibitor.

The occurrence of multiple independent domains within a protein is often associated with the appearance of more than one thermal transition as the temperature of the protein is

[†]Supported by Grant HL21791 from the National Institutes of Health.

*Author to whom correspondence should be addressed at the Holland Laboratory: Biochemistry, American Red Cross, Rockville, MD 20855.

[1] Abbreviations: ANS, 8-anilino-1-naphthalenesulfonate; Z-Gly-Arg-sBzl, *N*-(benzyloxycarbonyl)-L-glycylarginine thiobenzyl ester; IgG, immunoglobulin G; EDTA, ethylenediaminetetraacetic acid; SDS-PAGE, sodium dodecyl sulfate-polyacrylamide gel electrophoresis; Tris, tris(hydroxymethyl)aminomethane; TBS, Tris-buffered saline (0.02 M Tris-HCl, 0.15 M NaCl, pH 7.4); C1-Inh, C1 inhibitor; C1, first component of complement; C1r and C1s, zymogen forms of C1 subcomponents; $C1\bar{r}$ and $C1\bar{s}$, activated forms of C1r and C1s, differing from the zymogens by the presence of single clip in the polypeptide chain; kDa, kilodalton(s).

0006-2960/87/0426-5564$01.50/0