Characterization of Folding Intermediates of Human Carbonic Anhydrase II: Probing Substructure by Chemical Labeling of SH Groups Introduced by Site-Directed Mutagenesis†

Lars-Göran Mårtensson and Bengt-Harald Jonsson*

Department of Biochemistry, Umeå University, S-901 87 Umeå, Sweden

Per-Ola Freskgård, Anne Kihlgren, Magdalena Svensson, and Uno Carlsson

IFM-Department of Chemistry, Linköping University, S-581 83 Linköping, Sweden

Received July 6, 1992; Revised Manuscript Received October 20, 1992

ABSTRACT: By measurement of UV absorbance, CD spectra, and enzyme activity, we have shown that human carbonic anhydrase II forms a stable and compact folding intermediate at a moderate concentration of guanidine hydrochloride. The major aim of this study was to map the intermediate structure. For that reason, site-directed mutagenesis was used to introduce cysteine residues in various parts of the central β-structure to give in each case a single cysteine residue. Thereafter, the accessibility of the introduced SH group to specific chemical labeling was used to probe the stability and compactness of the area surrounding each cysteine residue. Our results indicate that the folding intermediate has an ordered native-like secondary structure in the central part of the β-sheet, whereas the peripheral part of the β-sheet seems to be less ordered. A large hydrophobic cluster situated between the central β-sheet core and secondary structure elements on the surface appears to be intact in the intermediate and is remarkably stable even at high GuHCl concentrations (>5 M). This unusually stable substructure might function as a "seed" during the initiation of the folding process.

To function in the cell, proteins must adopt a specific and highly ordered folded structure, a process which is accomplished spontaneously by the linear polypeptide chain immediately after biosynthesis. The difference in stability between unfolded states and the folded state is marginal (5–15 kcal/mol). To understand the factors that determine the stability of the native state, it is useful to characterize not only the native three-dimensional conformation but also intermediate forms that are found when folding kinetics are studied and at equilibrium under various denaturing conditions. Knowledge of the properties of these folding intermediates, which include only a subset of the native interactions, is of fundamental importance for the understanding of the hierarchy of interactions that stabilize the native state. Intermediates with properties resembling a model state referred to as a "molten globule", that is, a compact molecule with a high content of secondary structure and fluctuating tertiary structure, have been described: two of the best characterized examples are α-lactalbumin (Kuwajima et al., 1976; Baum et al., 1989) and myoglobin (Hughson et al., 1990, 1991). It has previously been shown that bovine carbonic anhydrase II has a stable intermediate conformation formed when it is denatured in guanidine hydrochloride (GuHCl) (Henkens et al., 1982). Ptitsyn and co-workers have further demonstrated that this enzyme has a collapsed structure at low pH and in moderate concentrations of denaturant, thereby forming a molten-globule state (Dolgikh et al., 1984). Studies in Henkens' group (Henkens & Oleksiak, 1991) indicate that some structural elements of bovine carbonic anhydrase II are stable even in 6.2 M GuHCl. In the present study, we used human carbonic anhydrase II (HCAII), a globular protein that has a molecular weight of 29 300 and is most prevalent in red blood cells. Some of the folding properties of HCAII have been characterized earlier (Carlsson et al., 1973, 1975), and its X-ray structure has been resolved to 2.0-Å resolution (Eriksson et al., 1988). The protein is composed of a single domain, with partial helix structure and a dominating β-structure that extends throughout the entire molecule. The gene for HCAII has been cloned and can be expressed in *Escherichia coli* (Forsman et al., 1988).

By measuring CD spectra, enzymatic activity, and absorption of UV light at 292 nm in various concentrations of GuHCl, we found that HCAII adopts a stable and compact intermediate state at moderate GuHCl levels. Stability parameters for the native and the intermediate state were calculated for cloned HCAII and for several mutants thereof.

We also used a recently developed approach (Freskgård et al., 1991; Minard et al., 1989) to characterize the detected intermediate state in more detail and to determine whether residual structures exist under more strongly denaturing conditions. This approach involves use of site-directed mutagenesis to introduce cysteine residues as local, chemically reactive probes. The residues are introduced one at a time to numerous positions throughout the protein structure, giving a repertoire of protein variants, each with only a single cysteine residue. Considering the known three-dimensional structure of HCAII, we chose mutagenesis positions in the β-sheet structure, in an aromatic cluster, and in the C-terminal part of the protein; these positions were selected so that the cysteine side chain would be buried in the properly folded protein. In the present experiments, the accessibility of the SH group for reaction with an alkylating reagent was used as a measure of the stability and compactness of the surroundings of the cysteine residue at various denaturing conditions. The specific

† This work was supported by grants from "Stiftelsen Bengt Lundqvists Minne" (P.-O.F. and M.S.), "Swedish National Board for Technical Development" (U.C. and B.-H.J.), "Swedish Natural Science Research Council" (B.-H.J.), "Carl Tryggers Stiftelse för vetenskaplig forskning" (U.C.), and "Sven och Lilly Lawskis fond" (L.-G.M.).

* To whom correspondence should be addressed.

0006-2960/93/0432-224$04.00/0

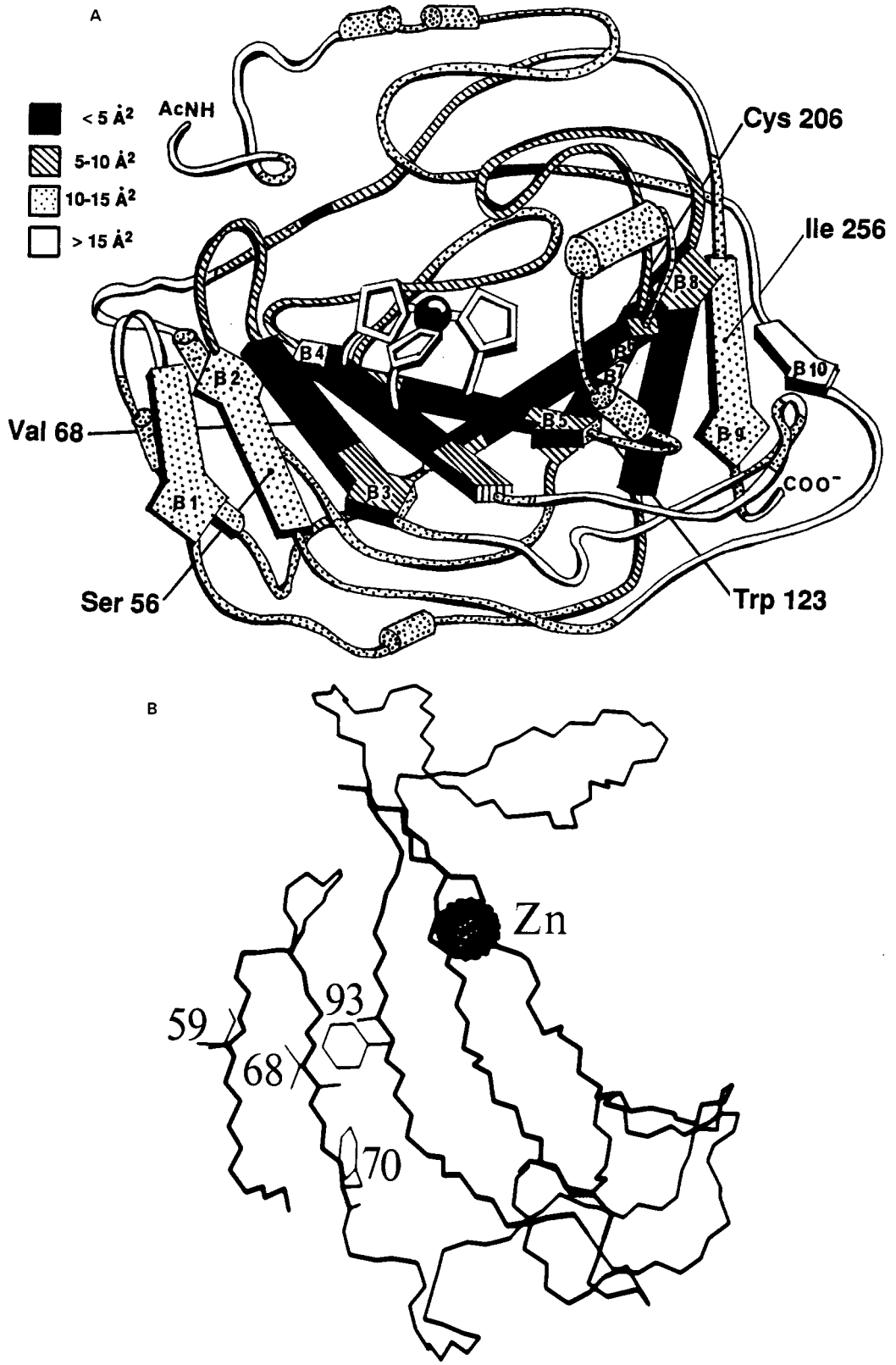

FIGURE 1: (A) Schematic drawing of the polypeptide backbone of human carbonic anhydrase II. Positions for amino acid substitutions and crystallografic temperature factors are indicated. The drawing is courtesy of Prof. A Liljas. (B) Consecutive β-strands 2–6 with Val-68 and its closest neighboring amino acid residues Ile-59, Phe-70, and Phe-93. (In addition, Phe-70 and Phe-93 are part of an aromatic cluster also containing Phe-66, -95, -176, -179, and -226 and Trp-97). This figure is drawn from coordinates kindly provided by A. E. Eriksson.

reactivity of SH groups toward alkylating reagents has been used by us in earlier studies of the refolding kinetics of HCAI and HCAII (Bergenhem et al., 1988; Carlsson et al., 1975; Freskgård et al., 1991) and by others in studies of phosphoglycerate kinase (Minard et al., 1989).

In the present work, five mutants as well as cloned HCAII were investigated and characterized; attention was focused on the central twisted β-sheet (Figure 1A). In the studied mutants, the original amino acid side chains are larger than or of approximately the same size as the introduced cysteine

residues. In summary, the results indicate that at moderate denaturing conditions HCAII exists in a nonnative, stable compact state and that residual structure is found even at 8 M GuHCl.

MATERIALS AND METHODS

Chemicals. Guanidine hydrochloride (sequential grade) was purchased from Pierce and was made metal-free by extraction with dithizone (7 mg/L) in carbon tetrachloride. Concentrations were confirmed by measuring the index of refraction (Nozaki, 1972). 7-Chloro-4-nitrobenzofurazan (NBD-Cl) was obtained from FLUKA, and iodo[2-^{14}C]acetic acid (54 mCi/nmol) was purchased from Amersham. All other chemicals were of reagent grade.

Protein Isolation and Purification. For in vitro site-directed mutagenesis, we used a method (Kunkel, 1985) based on host cell deficiency in the enzymes dUTPase (*dut*$^-$) and uracil *N*-glycosylase (*ung*$^-$). The expression plasmid pHCAII, previously described by us (Forsman et al., 1988), was used for production of the enzyme variants. All enzyme variants were purified by affinity chromatography (Khalifah et al., 1977), and purity was verified by sodium dodecyl sulfate–polyacrylamide gel electrophoresis.

Stability Measurements. After incubation of enzyme (8.5 μM) in various concentrations (0–5 M) of GuHCl, buffered with 0.1 M Tris–H_2SO_4, pH 7.5, for 24 h, denaturation was monitored by measuring A_{292} as a function of GuHCl concentration at 23 °C. A Perkin-Elmer Lambda 5 was used, in all spectrophotometrical work; all measurements were made in 1-cm quartz cells. The reference and the sample contained the same concentration of denaturant and buffer. To compensate for variation in enzyme concentration during spectrophotometric measurement, the A_{292}/A_{260} ratio was determined for each sample, since A_{260} has been shown to be unaffected by complete denaturation (Edsall et al., 1966) and is therefore a suitable internal standard (Mårtensson et al., 1992).

The inactivation was followed by CO_2 hydration activity measurements, as described elsewhere (Rickli et al., 1964; Mårtensson et al., 1992).

CD Measurements. CD spectra were registered at 23 °C on a Jasco-720 spectropolarimeter, which was calibrated using 10-camphorsulfonic acid (Hennessey & Johnson, 1982). Each spectrum was the result of five scans using a bandwidth of 1 nm. In the far-UV region, a 1-mm path length was used while it was 4 mm in the near-UV region. The observed ellipticities were converted to mean residue ellipticities [Θ] on the basis of a molecular weight of 29 300 and 259 amino acid residues (Henderson et al., 1976).

Chemical Labeling and Measurement of Radioactivity. To analyze the content of free sulfhydryl groups, the enzyme sample was denatured in 5 M GuHCl for 24 h and then labeled with 7-chloro-4-nitrobenzofurazan (NBD-Cl) using a 20-fold molar excess of reagent over enzyme (17.1 μM); the reaction was essentially completed within 30 min. Quantitation was done spectophotometrically using ϵ_{420} = 13 000 M^{-1} cm^{-1} for the NBD–sulfur adduct (Birkett et al., 1971). Carboxymethylation in various GuHCl concentrations was performed as follows. The enzyme (8.5 μM) was incubated in the GuHCl solutions (1.5 mL) buffered with 0.1 M Tris–H_2SO_4 at pH 7.5 for 24 h; thereafter radioactive iodoacetate (54 mCi/nmol, stock solution 2.44 mM) was added to a final concentration of 24.2 μM and the reaction was allowed to proceed for 1 h. The carboxymethylation was halted by adding 2-mercaptoethanol to a final concentration of 1.4 M, a concentration shown to completely terminate the alkylation process. The unreacted iodoacetate and 2-mercaptoethanol were removed by passing the protein sample through a Sephadex G-25 column (13 × 2.5 cm) equilibrated with 0.1 M Tris–H_2SO_4 buffer, pH 7.5, containing 2.2 M GuHCl; the GuHCl was included to prevent precipitation of the enzyme during analysis. The equilibration buffer was used for elution and the concentration of protein in the collected fractions (~ 1.5 mL) was determined from the absorption at 280 nm using an extinction coefficient of $A_{280}^{1\%}$ = 18.7. An aliquot (1.0 mL) from the peak fraction of the chromatography of the carboxymethylated enzyme samples was then dissolved in 15 mL of scintillation liquid (Liquid Scintillator Supersolve X, Zinsser Analytic), and the radioactivity was determined in a liquid scintillation counter (Beckman LS 1801). The fractions containing the modified protein and unreacted reagent were well separated since measurements on fractions in between these showed background radioactivity only.

Data Analysis. Even though the course of enzyme unfolding is sequential, each transition can be treated as a two-state process. The free energies of unfolding to the intermediate state I(ΔG_{NI}) were calculated from the curves shown in Figure 2A,B. For inactivation, we used the equation

$$\Delta G_{NI} = -RT \ln \frac{a_N - a}{a - a_I} \quad (1)$$

where a is the observed enzyme activity and a_N and a_I are the activities of the native and intermediate states. Results from absorbance measurements were treated in a corresponding way, but in that case both transitions can be analyzed separately. For the first transition we used the equation

$$\Delta G_{NI} = -RT \ln \frac{(A_{292}/A_{260})_N - A_{292}/A_{260}}{A_{292}/A_{260} - (A_{292}/A_{260})_I} \quad (2)$$

and for the second transition

$$\Delta G_{IU} = -RT \ln \frac{(A_{292}/A_{260})_I - A_{292}/A_{260}}{A_{292}/A_{260} - (A_{292}/A_{260})_U} \quad (3)$$

where A_{292}/A_{260} is the observed ratio of absorbances and $(A_{292}/A_{260})_N$, $(A_{292}/A_{260})_I$, and $(A_{292}/A_{260})_U$ are the ratios for native, intermediate, and unfolded forms, respectively.

Since for both ΔG_{NI} and ΔG_{IU} the dependence on the concentration of GuHCl was linear, the net stabilities of the native and the intermediate states in the absence of denaturant ($\Delta G_{NI}^{H_2O}$ and $\Delta G_{IU}^{H_2O}$) were estimated by linear extrapolation by using the equations

$$\Delta G_{NI} = \Delta G_{NI}^{H_2O} - m_{NI}[\text{GuHCl}] \quad (4)$$

and

$$\Delta G_{IU} = \Delta G_{IU}^{H_2O} - m_{IU}[\text{GuHCl}] \quad (5)$$

From the linear regression analysis of the data, correlation coefficients were found to be 0.98 ± 0.02. In principle, one would prefer to compare the stabilities of wild-type and mutant proteins in the absence of denaturants. In practice, the necessity of extrapolating data, which can only be measured accurately in the transition zone, over a long range may lead to errors which are comparable to the differences in stability between wild-type and mutant protein. For example, Pace tested the validity of linear extrapolation and has found that deviations of up to 30% can occur near zero molar denaturant (Pace & Vanderburg, 1979). An alternative approach, which is strictly empirical, is to compare the free energies where

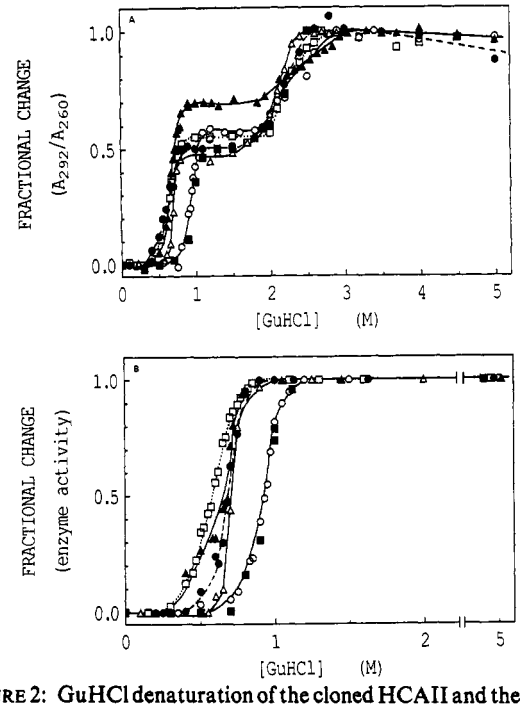

FIGURE 2: GuHCl denaturation of the cloned HCAII and the various mutants. The protein concentration was 0.25 mg/mL in 0.1 M Tris–H_2SO_4, pH 7.5, and incubation was performed for 24 h at 23 °C. Symbols: (○) cloned HCAII; (■) C206S; (□) I256C/C206S; (▲) W123C/C206S; (△) V68C/C206S; (●) S56C/C206S. A_{292}/A_{260} is plotted as a function of GuHCl concentration (A) and the fractional change in enzymic activity is plotted as a function of GuHCl concentration (B).

they can be most accurately measured (Matthews, 1987; Cupo & Pace, 1983; Kellis et al., 1989). Therefore, we have chosen to calculate the differences in stability ($\Delta\Delta G$) at 0.75 M GuHCl and at 2.25 M GuHCl for the first and second transition, respectively. These concentrations are chosen from the middle of the overlapping data in the transition zones. The destabilization of the different mutants (MUT), as compared to that of the cloned HCAII (REF), was determined from the following equations:

at 0.75 M GuHCl for the first transition

$$\Delta\Delta G_{NI} = \Delta G_{NI}^{MUT} - \Delta G_{NI}^{REF} \quad (6)$$

and at 2.25 M GuHCl for the second transition

$$\Delta\Delta G_{IU} = \Delta G_{IU}^{MUT} - \Delta G_{IU}^{REF} \quad (7)$$

RESULTS

HCAII unfolds in two well-separated transitions when the GuHCl concentration is raised from 0 to 5 M, indicating the existence of a stable intermediate state (I) with residual structure (Figure 2A, Table I). The transitions were monitored by recording the change in absorbance at 292 nm, and the results are presented as the ratio A_{292}/A_{260} where A_{260} represents an internal standard for protein concentration (Mårtensson et al., 1992). This parameter reflects the degree of exposure of buried tryptophans to solvent during various stages of the unfolding process. There are a total of seven tryptophan residues situated in different parts of the molecule, and the ratio A_{292}/A_{260} thus reflects changes coming from the entire three-dimensional structure. However, in the native enzyme, the tryptophans are exposed to solvent to different extents (Lindskog & Nilsson, 1973; Eriksson et al., 1988); hence, each tryptophan might contribute differently to the change in absorbance at 292 nm during unfolding.

The enzymic activity vanishes completely during the first transition, at low concentrations of GuHCl (Figure 2B), showing that the intermediate state is inactive. The midpoint for this inactivation transition (Table II) correlates well with the observed first transition, when change in the absorbance at 292 nm is measured. The transition midpoints for cloned HCAII are 0.94 M (N–I) and 2.4 M (I–U) (Table I). Similar results were obtained for HCAII isolated from erythrocytes. Transition midpoint concentrations were calculated from the data using eqs 4 and 5.

Five substitutions in different parts of the structure were made by site-directed mutagenesis. In the C206S mutant we replaced cysteine 206 with a serine residue, obtaining an enzyme devoid of cysteine. Subsequently, four more positions in the C206S mutant were subjected to substitution: new cysteine residues were introduced in various parts of the structure to replace serine 56, valine 68, tryptophan 123, and isoleucine 256, resulting in the four double mutants S56C/C206S, V68C/C206S, W123C/C206S, and I256C/C206S, respectively (Figure 1A). These mutants unfold in two transitions like cloned HCAII when subjected to increasing concentrations of GuHCl, as monitored by absorbance at 292 nm (Figure 2A). The first transition was also demonstrated by the loss of enzymatic activity with increasing GuHCl concentrations. In some cases, the midpoint concentrations of one or both transitions were markedly influenced by the mutation. However, regarding the transition between the native and intermediate states, the stabilities of the single mutant (C206S) and the cloned HCAII are practically identical (Table I and II). Therefore, the decreased stability of the double mutants observed in the first transition is most probably due to the replacement with cysteine at the specified positions. A comparison with cloned HCAII shows that the native states of the double mutants are destabilized by 2.2–2.4 kcal/mol ($\Delta\Delta G_{NI}$) while the single mutant C206S is unaffected or slightly stabilized ($\Delta\Delta G_{NI}$ = +0.1 kcal/mol; Table I). Thus, the native states of the double mutants, as compared to that of the C206S single mutant, are significantly destabilized relative to their I states.

Variation in the relative stability of the I and U states is less striking (Figure 2A, Table I). The I states of S56C/C206S, I256C/C206S, and C206S are destabilized by 0.7 kcal/mol or less, while that of the V68C/C206S mutant is somewhat more destabilized (–1.2 kcal/mol).

Apparently, it is possible to vary the relative stability of the N and I states of HCAII over a broad range by single amino acid changes, whereas the stability of the I state relative to the unfolded state (U) seems to be less affected by mutations. This is even more pronounced for another mutant (S29C/C206S), included in previous experiments (Mårtensson et al., 1992), for which it was shown that the first transition can occur at a midpoint concentration of <0.1 M GuHCl, without any significant destabilizing effect on the second transition. CD spectra of the C206S mutant, in various GuHCl concentrations, were recorded in both the near-UV and far-UV regions (Figure 3).

To characterize the structure of the I state of HCAII in greater detail, the cysteine residues introduced in the different mutants were used to follow local conformational changes in the mutated regions. For each of these mutants, as well as for cloned HCAII, we measured the variation in relative exposure of the cysteine SH groups in different concentrations of GuHCl. After incubation of the protein in the GuHCl-

Table I: Stability Parameters for GuHCl-Induced Unfolding Measured by Absorbance Change[a]

protein	C_{mNI}	C_{mIU}	$\Delta G_{NI}^{H_2O}$	$\Delta G_{IU}^{H_2O}$	m_{NI}	m_{IU}	$\Delta\Delta G_{NI}$[b]	$\Delta\Delta G_{IU}$[c]
cloned HCAII	0.94	2.4	7.6	5.8	8.1	2.5		
C206S	0.96	2.2	6.8	11.9	7.0	5.5	0.1	–0.7
S56C/C206S	0.69	2.2	7.2	7.9	10.5	3.6	–2.2	–0.5
V68C/C206S	0.70	2.0	12.9	8.8	18.3	4.3	–2.4	–1.2
W123C/C206S	0.66	2.4	6.8	4.8	10.2	2.0	–2.4	0.0
I256C/C206S	0.63	2.2	4.6	8.6	7.4	3.8	–2.4	–0.3

[a] Units are as follows: ΔG, kcal mol^{-1}; m, kcal mol^{-1} M^{-1} (GuHCl). C_m represents transition midpoint concentration (GuHCl). [b] At 0.75 M GuHCl. [c] At 2.25 M GuHCl.

Table II: Stability Parameters for GuHCl-Induced Unfolding Measured by Enzyme Activity[a]

protein	C_{mNI}	$\Delta G_{NI}^{H_2O}$	m_{NI}	$\Delta\Delta G_{NI}$[b]
cloned HCAII	0.93	7.9	8.5	
C206S	0.95	7.9	8.2	0.2
S56C/C206S	0.66	6.9	10.4	–2.4
V68C/C206S	0.72	9.4	13.1	–2.0
W123C/C206S	0.65	6.7	10.3	–2.6
I256C/C206S	0.57	4.2	7.4	–2.9

[a] Units are as follows: ΔG, kcal mol^{-1}; m, kcal mol^{-1} M^{-1} (GuHCl). C_m represents transition concentration (GuHCl). [b] At 0.75 M GuHCl.

solutions for 24 h, the samples were allowed to react with a 3-fold molar excess of radioactive iodoacetate for a period of 1 h. The rate of incorporation of radioactivity was regarded as a measure of the exposure of the cysteine at the particular GuHCl concentration. To facilitate comparisons between different mutants, we calculated the relative rates of alkylation using the formula $(\nu - \nu_N)/(\nu_U - \nu_N)$ where ν is the measured rate and ν_N is the rate obtained for the native state and ν_U is the highest rate obtained for the denatured state. The relative rates of alkylation of the enzyme variants are shown in Figure 4. With the exception of the V68C/C206S mutant, the cysteines seem to be fully accessible to solvent within the range of the two transitions, considering measurements of change in absorbance at 292 nm. On the other hand, the cysteine at position 68 is not fully accessible to reagent, even in 8 M GuHCl.

In the experiments on iodoacetate alkylation of SH groups, a possible source for misinterpretation could be competing formation of disulfide bridges between the polypeptide chains. To exclude this possibility, titration of the SH group with NBD–Cl in 5 M GuHCl was performed. The choice of reaction conditions was based on the assumption that in 5 M GuHCl the SH groups are most accessible to reagent and to disulfide formation. A degree of modification of 85–90% was achieved in all cases, which clearly shows that the majority of the SH groups were in a reduced state free to react with the alkylating reagent.

DISCUSSION

The main theme of this report, which is mapping of the intermediate structure by measuring alkylation rates of purposely introduced SH groups, is treated after a discussion of stability of the native and intermediate structures, and after a discussion of CD spectra, which gives information of secondary structure.

Stability of the Native State. Cloned HCAII and double mutants apparently differ in regard to stability of the native state, judging from $\Delta\Delta G_{NI}$ values obtained from the first transition zone. The stability of the C206S mutant, in which the single cysteine was replaced by a serine, is practically indistinguishable from that of cloned HCAII. For the mutant C206S, as for all the other studied enzyme forms, two unfolding transitions were revealed when absorbance change was studied; hence, the two transitions cannot be due to disulfide formation. The Cys → Ser mutation is a change to a chemical and structural analog which is somewhat smaller in size and which should be easily accommodated in the native structure. Considering that residue 206 is not in close proximity to any of the other mutated residues, it is likely that the observed destabilization of all the double mutants with C206S can be attributed to changes at positions 56, 68, 123, and 256.

In the S56C/C206S mutant, we replaced a serine with a cysteine at position 56. The observed destabilizing effect in this case ($\Delta\Delta G_{NI}$ = –2.2 kcal/mol) might have been due to weaker hydrogen bonding with the incorporated SH group or to a steric effect in which the somewhat larger sulfur atom has unfavorable van der Waals interactions with the neighboring residues. Notably, a comparison of Ramachandran plots of cysteine and serine in model peptides demonstrates that the rotational freedom of the cysteine side chain is greatly reduced (Görbitz, 1990).

Assuming that differences in hydrophobicity represents a major cause of destabilization, one might expect that a mutation from isoleucine to cysteine would cause a larger destabilization than a valine to cysteine mutation which is contrary to the observed destabilizations of the mutants V68C/C206S ($\Delta\Delta G_{NU}$ = –3.6 kcal/mol) and I256C/C206S ($\Delta\Delta G_{NU}$ = –2.7 kcal/mol). A possible explanation may be found in an eloquent study by Eriksson et al. (1992) of Leu → Ala mutants made in the hydrophobic core of T4–lysozyme. They found that the observed destabilization can be approximated by a constant term, which is equal to the differences in the energy of transfer between water and organic solvent for leucine and alanine, plus a variable term, which increases linearly with the observed size of the cavity created by the mutation. The free energies of transfer between water and organic solvent (*N*-methylacetamide; Damodaran & Song, 1986) are –0.4 kcal/mol for the valine–cysteine pair and –1.0 kcal/mol for the isoleucine–cysteine pair. Because the side chains of valine and isoleucine have similar physicochemical properties, we assume that the relation between the transfer values will be roughly maintained (although the magnitudes might change) also for transfer between GuHCl solutions and organic solvent. Accordingly, the magnitude of the destabilization by the mutations might be explained by creation of cavities that are larger for the V68C/C206S than for the I256C/C206S mutant. The larger cavity around position 68 might reflect that this part of the protein is rather rigid.

For the W123C/C206S mutant, we observe a destabilization of 2.4 kcal/mol, a surprisingly small value when considering the large difference in size of the side chains and that this tryptophan residue is evolutionarily invariant in all known isoenzymes of carbonic anhydrase. The residue is, however, partly exposed at the surface, which might explain why the substitution at hand has a relatively small effect on the stability.

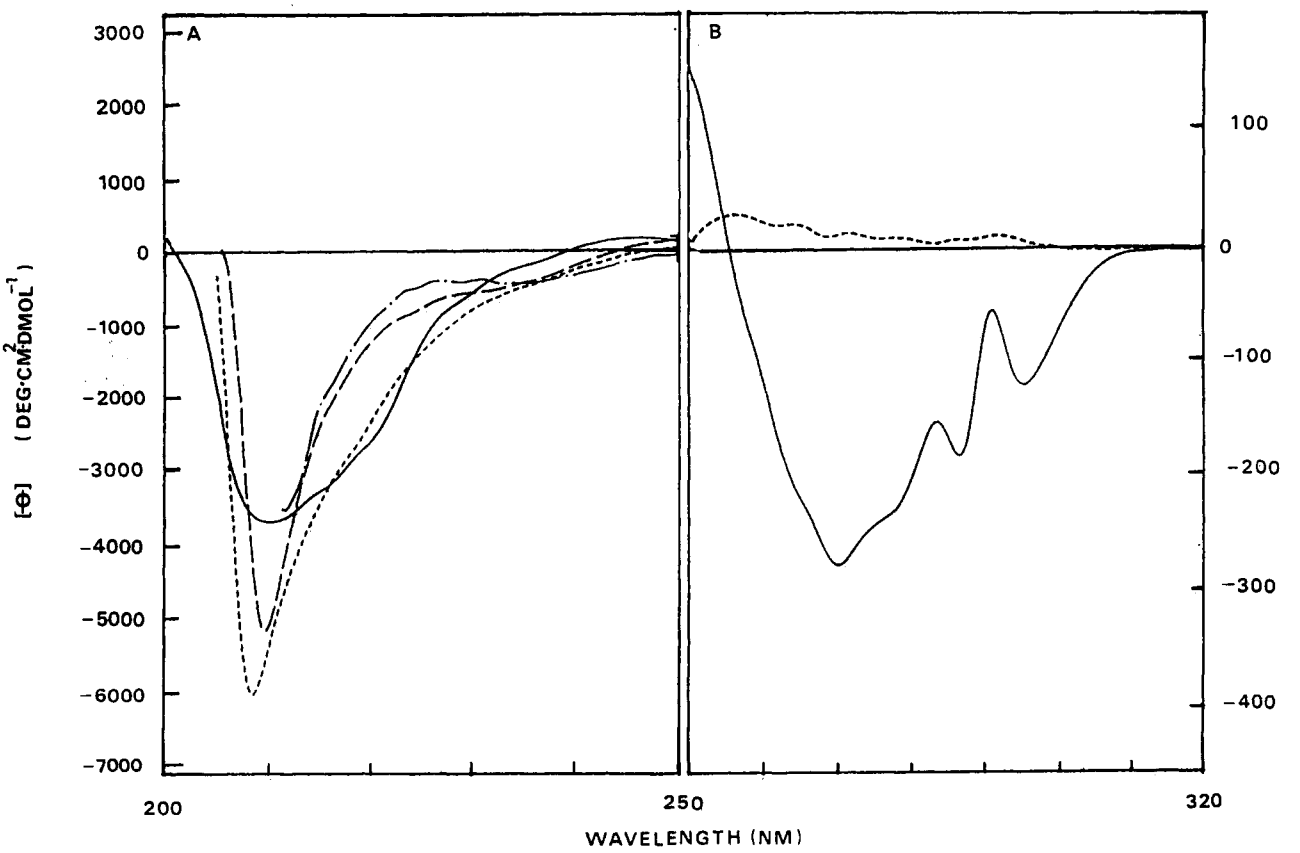

FIGURE 3: Far-ultraviolet (A) and near-ultraviolet (B) circular dichroism spectra of the mutant C206S in various concentrations of GuHCl at 23 °C. Protein concentrations were 0.5 mg/mL (far-UV) and 1.0 mg/mL (near-UV) and the corresponding cell lengths were 1 mm and 4 mm, respectively. All samples contained 10 mM phosphate buffer, pH 7.5. Symbols: (—) 0 M GuHCl, (- - -) 1.5 M GuHCl, (– –) 2.5 M GuHCl, (– • –) 5 M GuHCl.

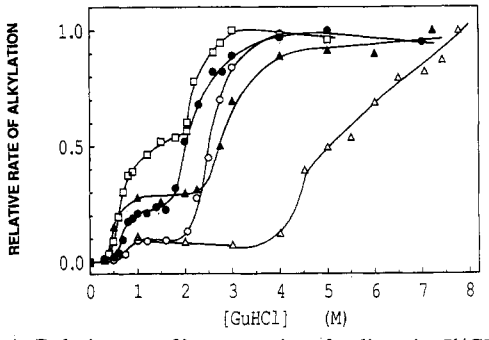

FIGURE 4: Relative rate of incorporation of radioactive $I^{14}CH_2COO^-$ by cysteines in the protein variants. The absolute rates of alkylation in cpm $nmol^{-1}$ h^{-1} in 5 M GuHCl concentrations are as follows: 3310 for cloned HCAII(C206), 4250 for S56C/C206S, 990 for V68C/C206S, 1130 for W123C/C206S, and 7940 for I256C/C206S. Symbols are as in Figure 2.

Stability of the Intermediate State. From Figure 2A and $\Delta\Delta G_{IU}$ values presented in Table I, it is obvious that most of the mutations presented in this report have only minor effects on the stability of the intermediate state as compared to the unfolded state; all except V68C/C206S are affected by less than 1 kcal/mol, whereas the native state of all double mutants is destabilized by 2.2–2.4 kcal/mol. In the mutant S29C/C206S, considered in a previous study (Mårtensson et al., 1992), we observed a destabilization of the native state by approximately 8 kcal/mol. In contrast, the observation that $\Delta\Delta G_{IU} = 0$ indicates that the stability of the intermediate was much less affected by the S29C mutation. Hence, the native state seems to be more sensitive to the present mutations than the intermediate state. The explanation for this might be that in the native state all side chains in positions to be mutated are involved in structure-dependent specific interactions. In the intermediate state, the structure is still rather compact (see below), but the motion of side chains might be less constrained, leading to fewer specific interactions. In our experiments, mutations were introduced in positions buried in the protein, and specific interactions in the native state are therefore inevitably affected. This in turn means that destabilization of the native state is highly probable. An intermediate state with a more flexible structure could, perhaps, more easily accommodate changes in side-chain volume. Notably, the 1.2 kcal/mol destabilization of the intermediate state caused by the V68C/C206S mutation is significantly larger than for the other studied mutants. This indicates that residue 68 is part of a structure that is highly ordered, even in the intermediate state. In addition, using the formalism of Pace (Pace et al., 1990), the m_{NI} value for the V68C mutant is significantly higher then the m_{NI} values for the other mutants, indicating that side chain exposure is greater in the V68C intermediate than in the intermediates of the other mutants. Consequently, with the original valine in position 68, the surrounding structure might be even more stable than is apparent considering our measurements on the V68C mutant. In the secondary structure, mutations which destabilize the native state but not the intermediate state have been proposed to indicate a molten-globule intermediate and not a framework intermediate (Hughson et al., 1991; Baldwin, 1991). Using this operational definition, we find that our variants with mutations in positions 56, 123, and 256 are indicative of a molten-globule intermediate, whereas the surroundings of position 68 seem to be more structured in the intermediate state than is allowed by the molten-globule model.

Secondary Structure in the Intermediate State. CD spectra in the near-UV region show that at 1.5 M GuHCl virtually all bands have disappeared, showing that in the intermediate state the aromatic residues are in a symmetric environment (Figure 3). Hence, we may conclude that bands from aromatic residues that contribute to the far-UV CD spectra of native protein disappear also from these spectra at 1.5 M GuHCl. Contributions from aromatic residues in this region normally have a positive sign (Rosenberg, 1966) and notably we observe a larger negative ellipticity for the intermediate than for the native state. At higher GuHCl concentrations, these bands decrease gradually. Taken together these results indicate that the β-structure is substantially intact in the intermediate state.

Mapping the Intermediate Structure. To obtain a more detailed picture of how different parts of the structure unfold, and to elucidate the hierarchy of interactions that are still intact in the intermediate state, we employed the cysteine reactivity approach. This method is used to determine the compactness of the intermediate in the area surrounding the mutation. The locations of the mutated positions are indicated in Figure 1A; crystallographic *B*-factors are indicated to enable comparison of rigidity in the native state with compactness in various parts of the intermediate structures.

Cloned HCAII: The naturally occurring Cys-206 is located in β-strand 7. According to hydrophobicity calculations, this β-strand and β-strand 6 represent the most hydrophobic regions in the molecule (Bergenhem et al., 1989). After labeling Cys-206 with [^{14}C]iodoacetate, we observed only a slight increase in accessibility in GuHCl concentrations causing the first transition (Figure 4). This shows that this position is practically completely buried in the intermediate state, and consequently, the second transition must be dominant.

I256C/C206S: X-ray data imply that Ile-256 is unexposed in the native state, which corresponds well with the low rate of labeling (30 cpm nmol^{-1} h^{-1}) of the inserted cysteine residue. Position 256 is situated in C-terminal β-strand 9. This strand is located next to the outermost β-strand in the dominant β-structure that extends throughout the entire molecule. The C-terminal strand is part of a knot topology (Eriksson et al., 1988; Creighton, 1984), which, according to current knowledge, is a feature unique to carbonic anhydrase (Figure 1A). Labeling at position 256 in GuHCl shows two transitions with about equal magnitude. Comparable results were obtained for the first transition when the rate of incorporation of radioactivity and absorbance change at 292 nm were measured, and the results for the single transition measured as loss of enzyme activity were also similar. This means that in the intermediate state position 256 is relatively exposed. This part of the intermediate should have a fluctuating structure, the secondary structure of which has also started to unfold. This substructure seems to be completely exposed in 3 M GuHCl.

W123C/C206S: Trp-123 is partly exposed to the surface and is contained in β-strand 5, which is situated on the hydrophobic side of the active-site cavity. A relatively high rate of labeling of this inserted cysteine residue is obtained, even in the absence of a denaturing agent (360 cpm nmol^{-1} h^{-1}, as compared to 30–80 cpm nmol^{-1} h^{-1} for the other mutants). This indicates that the cavity that is created by replacing tryptophan with cysteine does not collapse in the native protein. Furthermore, in the intermediate state, the substructure around this position is not unfolded to the same extent as the region surrounding position 256, suggesting relative rigidity in the area of Trp-123; this part of the molecule may be stabilized by the large neighboring hydrophobic cluster. During the second transition, this region seems to be totally ruptured.

S56C/C206S: Ser-56 is located in β-strand 2, a side chain which must be shielded by β-strand 1 to remain unexposed. β-strands 1 and 10 are on the opposite rims of the β-sheet and at the surface of the native molecule. The labeling experiments show that residue 56 is largely protected in the intermediate, indicating that β-strand 1 has a native-like structure in 1–2 M GuHCl. This substructure becomes fully exposed during the second transition, as do the substructures around positions 206, 256, and 123. Notably, position 56 and 256 which are located in the outer β-strands (Figure 1A) become fully exposed in lower concentrations of GuHCl than do position 206 and 123, which are located in the middle of the protein molecule.

V68C/C206S: Residue 68 is found in β-strand 3 in the middle of one of the two hydrophobic clusters present in the molecule. This cluster is made up of 32 amino acid residues, 8 of which are aromatic, and it is located between the central β-sheet core and secondary structural elements on the surface. The low crystallographic B-values may indicate that the region around position 68 is relatively rigid or they may merely reflect that the protein moves as an independent rigid body in the crystal and therefore the displacements are smaller at the center of mass (Kuriyan & Weis, 1991). Position 68 is buried under the side chains of Ile-59, Phe-70, and Phe-93 and is part of a central β-sheet composed of β-strands 2–6, which are antiparallel and consecutive in the primary structure (Figure 1B). The side chain of Trp-97 in β-strand 4 is also contained in the same hydrophobic cluster as residue 68. According to Henkens and Oleksiak (1991), NMR measurements indicate that the structure around position 97 is nonrandom in 6.2 M GuHCl. Interestingly, our chemical labeling data on the Cys-68 variant (Figure 4) show that in 4 M GuHCl the enzyme does not form a random coil, as previously believed, but instead exhibits a compact residual structure. In fact, even in the very strongly denaturing environment represented by 8 M GuHCl (the solubility limit), a plateau value is not reached, indicating that this substructure is not completely ruptured. The labeling reaction was repeated using the neutral reagent iodoacetamide, but no difference in reactivity was noted, and a repulsion effect of the negatively charged iodoacetate can therefore be excluded. Stable residual structures have been suggested to act as "seeds" that initiate the folding process (Baldwin, 1986). Since the substructure around position 68 is unusually stable, it might function as such a "seed". Interestingly, the maximal absolute rates of cysteine alkylation in 5 M GuHCl vary depending on which mutant was studied and are in descending order: I256C/C206S > S56C/C206S > HCAII(C206) > W123C/C206S > V68C/C206S (data in legend to Figure 4). These rate differences may be understood by noting that the absolute rates are lower the closer the position is to the central antiparallel β-sheet with the adhering aromatic cluster. This is due to the fact that a cysteine that is exposed but still close to a compact structure undergoes a lower rate of alkylation since the reagent can diffuse toward the cysteine from a limited space only.

The results of our stability measurements are compatible with the view that a collapsed state is formed under mild denaturing conditions, an idea originally put forward by Dolgikh et al. (1984). This intermediate has previously been characterized using methods that probe global structural properties. With the type of approach we used, it is possible to analyze intermediate states in greater detail than previously possible. The present results indicate that HCAII assumes

an intermediate state with an ordered secondary structure in the central part of the twisted β-sheet, while the peripheral part of the β-sheet seems to be less ordered; parts of the large aromatic cluster might be rather rigid.

Kinetic data from a previous study also indicate the formation of an intermediate in the early stages of the folding process (Freskgård et al., 1991). Thus, as soon as the refolding is initiated, position 206 is rapidly hidden from an alkylating agent, and the present studies show that position 206 is buried in the intermediate (I) state but is totally exposed in the denatured (U) state (Figure 4). During the refolding process position 256 is buried with a half-time of 75 s, showing that the surroundings of this position are not structured in the initiation process. This is in agreement with the present finding that position 256 is largely exposed in the intermediate state. Thus, positions that are totally inaccessible in the equilibrium intermediate state (I) are probably also hidden very rapidly in the folding process.

CONCLUSIONS

Our results indicate that the major central twisted β-sheet of the intermediate state is native-like and that the largest degree of fluctuation in this sheet occurs in its outer parts. Furthermore, the dominating hydrophobic cluster is largely intact and remarkably stable, even in high GuHCl concentrations.

In conclusion, the described method of combined site-directed mutagenesis and specific labeling may be a powerful tool for mapping intermediate structures of proteins.

ACKNOWLEDGMENT

We thank Dr. N. Bergenhem for fruitful criticism during the progress of this work and Prof. R. Henkens for valuable discussions of the manuscript.

REFERENCES

Baldwin, R. L. (1986) *Trends Biochem. Sci. 11*, 6–9.
Baldwin, R. L. (1991) in *Protein Conformation* (Chadwick, D. J., & Widdows, K., Eds.) pp 190–205, John Wiley & Sons, New York.
Baum, F. M., Dobson, C. M., Evans, P. A., & Hanley, C. (1989) *Biochemistry 28*, 7–13.
Bergenhem, N., Carlsson, U., & Karlsson, J.-A. (1989) *Int. J. Pept. Protein Res. 33*, 140–145.
Birkett, D. J., Dwek, R. A., Radda, G. K., Richards, R. E., & Salmon, A. G. (1971) *Eur. J. Biochem. 20*, 494–508.
Carlsson, U., Henderson, L. E., & Lindskog, S. (1973) *Biochim. Biophys. Acta 310*, 376–387.
Carlsson, U., Aasa, R., Henderson, L. E., Jonsson, B.-H., & Lindskog, S. (1975) *Eur. J. Biochem. 52*, 25–36.
Creighton, T. E. (1984) in *Protein, Structures and Molecular Properties*, p 226, W. H. Freeman, New York.
Cupo, J. F., & Pace, C. N. (1983) *Biochemistry 22*, 2654–2658.
Damodaran, S., & Song, K. S. (1986) *J. Biol. Chem. 261*, 7220–7222.
Dolgikh, D. A., Kolomiets, A. P., Bolotina, I. A., & Ptitsyn, O. B. (1984) *FEBS Lett. 165*, 88–92.
Edsall, J. T., Mehta, S., Myers, D. V., & Armstrong, J. McD. (1966) *Biochem. Z. 345*, 9–36.
Eriksson, A. E., Jones, T. A., & Liljas, A. (1988) *Proteins: Struct., Funct., Genet. 4*, 274–282.
Eriksson, A. E., Baase, W. A., Zhang, X.-J., Heinz, D. W., Blaber, M., Baldwin, E. P., & Matthews, B. W. (1992) *Science 255*, 178–183.
Forsman, C., Behravan, G., Osterman, A., & Jonsson, B.-H. (1988) *Acta Chem. Scand. B42*, 314–318.
Freskgård, P.-O., Carlsson, U., Mårtensson, L.-G., & Jonsson, B.-H. (1991) *FEBS Lett. 289*, 117–122.
Görbitz, C. H. (1990) *Acta Chem. Scand. 44*, 584–590.
Henderson, L. E., Henriksson, D., & Nyman, P. O. (1976) *J. Biol. Chem. 251*, 5457–5463.
Henkens, R. W., & Oleksiak, T. P. (1991) in *Carbonic Anhydrase. From Biochemistry and Genetics to Physiology and Clinical Medicine* (Botrè et al., Eds.) pp 44–49, VCH, New York.
Henkens, R. W., Kitchell, D. B., Lottich, S. C., Stein, P. J., & Williams, T. J. (1982) *Biochemistry 21*, 5918–5925.
Hennessey, J. P., Jr., & Johnson, W. C., Jr. (1982) *Anal. Biochem. 125*, 177–188.
Hughson, F. M., Wright, P. E., & Baldwin, R. L. (1990) *Science 249*, 1544–1548.
Hughson, F. M., Barrich, D., & Baldwin, R. L. (1991) *Biochemistry 30*, 4113–4118.
Kellis, J. T., Nyberg, K., & Fersht, A. R. (1989) *Biochemistry 28*, 4914–4922.
Khalifah, R. G., Strader, D. J., Bryant, S. H., & Gibson, S. M. (1977) *Biochemistry 16*, 2241–2247.
Kunkel, P. A. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 488–492.
Kuriyan, J., & Weis, I. W. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 2773–2777.
Kuwajima, K., Nitta, K., Yoneyama, M., & Sugai, S. (1976) *J. Mol. Biol. 106*, 359–373.
Lindskog, S., & Nilsson, A. (1973) *Biochim. Biophys. Acta 295*, 117–130.
Mårtensson, L.-G., Jonsson, B.-H., Andersson, M., Kihlgren, A., Bergenhem, N., & Carlsson, U. (1992) *Biochim. Biophys. Acta 1118*, 179–186.
Matthews, C. R. (1987) *Methods Enzymol. 154*, 498–511.
Minard, P., Desmadril, M., Ballery, N., Perahia, D., Hall, L., & Yon, J. M. (1989) *Eur. J. Biochem. 183*, 419–423.
Nozaki, Y. (1972) *Methods Enzymol. 26*, 43–50.
Pace, C. N., & Vanderburg, K. E. (1979) *Biochemistry 18*, 288–292.
Pace, C. N., Laurents, D. V., & Thomson, J. A. (1990) *Biochemistry 29*, 2564–2572.
Rickli, E. E., Ghazanfar, S. A. S., Gibbons, B. H., & Edsall, J. T. (1964) *J. Biol. Chem. 236*, 1065–1078.
Rosenberg, A. (1966) *J. Biol. Chem. 241*, 5119–5125.