Biochemistry

 Volume 32, Number 42 October 26, 1993

Articles

Effect of Cavity-Creating Mutations in the Hydrophobic Core of Chymotrypsin Inhibitor 2†

Sophie E. Jackson,‡ Marco Moracci,§ Nadia elMasry, Christopher M. Johnson, and Alan R. Fersht*

MRC Unit for Protein Function and Design, Cambridge IRC for Protein Engineering, University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW, U.K.

Received April 29, 1993; Revised Manuscript Received August 12, 1993®

ABSTRACT: Hydrophobic residues in the core of a truncated form of chymotrypsin inhibitor 2 (CI2) have been mutated in order to measure their contribution to the stability of the protein. The free energy of unfolding of wild-type and mutants was measured by both guanidinium chloride-induced denaturation and differential scanning calorimetry. The two methods give results for the changes in free energy on mutation that agree to within 1% or 2%. The average change in the free energy of unfolding (± standard deviation) for an Ile → Val mutation is 1.2 ± 0.1 kcal mol^{-1}, for a Val → Ala mutation 3.4 ± 1.5 kcal mol^{-1}, and for either an Ile → Ala or a Leu → Ala mutation 3.6 ± 0.6 kcal mol^{-1}. This gives an average change in the free energy of unfolding for deleting one methylene group of 1.3 ± 0.5 kcal mol^{-1}. Two significant correlations were found between the change in the free energy of unfolding between wild-type and mutant, $\Delta\Delta G_{U-F}$, and the environment of the mutated residue in the protein. The first is between $\Delta\Delta G_{U-F}$ and the difference in side-chain solvent-accessible area buried between wild-type and mutant (correlation coefficient = 0.81, 10 points). The second and slightly better correlation was found between $\Delta\Delta G_{U-F}$ and N, the number of methyl/methylene groups within a 6-Å radius of the hydrophobic group deleted (correlation coefficient = 0.84, 10 points). The latter correlation is very similar to that found previously for barnase, suggesting that this relationship is general and applies to the hydrophobic cores of other globular proteins. The combined data for C12 and barnase clearly show a better correlation with N (correlation coefficient = 0.87, 30 points) than with the change in the solvent-accessible surface area (correlation coefficient = 0.82, 30 points). This indicates that the packing density around a particular residue is important in determining the contribution the residue makes to protein stability. In one case, Ile → Val76, a mutation which deletes the $C^{\delta 1}$ methyl group of a buried side chain, a surprising result was obtained. This mutant was found to be more stable than wild-type by 0.2 ± 0.1 kcal mol^{-1}. We have solved and analyzed the crystal structure of this mutant and find that there are small movements of side chains in the core, the largest of which, 0.7 Å, is a movement of the side chain that has been mutated. These small movements in some part compensate for the cavity created by the mutation.

The hydrophobic effect, the burial of hydrophobic residues, is generally considered to be the major driving force in protein folding [for a review, see Dill (1990)]. Extensive literature has been published of estimates of the net contribution of hydrophobic residues to protein stability [for reviews, see Rose et al. (1985) and Nakai et al. (1988); for experimental work, see Kellis et al. (1988, 1989), Alber et al. (1987), Yutani et al. (1987), Matsumara et al. (1988), Shortle et al. (1990), Sandberg and Terwilliger (1991), Serrano et al. (1992), and Fersht and Serrano (1993)]. Early studies on enzyme–substrate interactions have shown that the contribution of

† S.E.J. is a William Stone Research Fellow, Peterhouse, Cambridge. M.M. was supported by a CNR–MISM Fellowship.

* Author to whom correspondence should be addressed.

‡ Present address: Department of Chemistry, Harvard University, 12 Oxford St., Cambridge, MA 02138.

§ Present address: Istituto di Biochimica delle Proteine ed Enzimologia—CNR, Napoli, Italy.

® Abstract published in *Advance ACS Abstracts,* October 1, 1993.

0006-2960/93/0432-11259$04.00/0

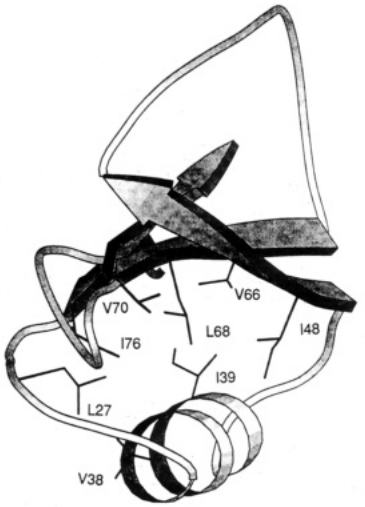

FIGURE 1: Secondary and tertiary structures of chymotrypsin inhibitor 2, showing the side chains of the hydrophobic core residues mutated in this study. The diagram was produced using the program *MolScript* (Kraulis, 1991).

hydrophobic groups to binding can be large and variable, up to 3.5 kcal mol^{-1} per methylene group [see Fersht (1985), Chapter 11]. Studies on the contribution of hydrophobic residues to protein stability have found that for mutations at a single position there can be a correlation with conventional values of hydrophobicity, e.g., Ile3 in T4 lysozyme (Matsumara et al., 1988) and position 49 in tryptophan synthase α-subunit (Yutani et al., 1987). Measurements on the hydrophobic core of barnase (Kellis et al., 1988, 1989) showed the effects to be larger and more variable than expected. These and more extensive studies show clearly that hydrophobicity alone does not account for the variation in interaction energies of hydrophobic residues (Shortle et al., 1990; Serrano et al., 1992). The contribution of a buried methylene group to protein stability is found to vary, for example, in barnase between 0.9 and 2.1 kcal mol^{-1} (Serrano et al., 1992), depending upon its local environment. Understanding how the environment of a hydrophobic residue affects its contribution to protein stability is, therefore, important for accurate prediction of the contribution of a residue to overall stability. This is essential as the free energy of folding of proteins is generally quite small, typically 5–15 kcal mol^{-1} (Privalov, 1979), which can be close to the free energy changes involved in the mutation of, e.g., Ile to Ala (up to 5 kcal mol^{-1}).

Here we present a study on hydrophobic core mutants of chymotrypsin inhibitor 2 (CI2). CI2 provides an excellent system for studying protein folding and stability. It is a single polypeptide chain of 64 amino acid residues. Yet, despite its small size, it has considerable secondary and tertiary structure (see Figure 1). Its three-dimensional structure has been determined both by X-ray crystallography (McPhalen & James, 1987) and by nuclear magnetic resonance (NMR) spectroscopy (Ludvigsen et al., 1992). It has no disulfide bridges, thereby minimizing residual structure in the unfolded state. We have previously shown that CI2 undergoes reversible denaturation mediated by chemical denaturants such as guanidinium chloride (GdnHCl), or heat, that follows a two-state process (Jackson & Fersht, 1991a). Using GdnHCl-induced denaturation and fluorescence spectroscopy to monitor the state of the protein, it is possible to measure the changes in the free energy of folding of the protein with a high degree of accuracy (Jackson & Fersht, 1991a). In this study, we use a truncated form of CI2 in which the N-terminal tail of 19 amino acid residues is removed from the parent 83-residue protein. This tail is disordered, as evidence from the solution structure determined by NMR indicates (Kjaer et al., 1987; Kjaer & Poulsen, 1987).

The hydrophobic core of CI2 is formed by the packing of a four-stranded, mixed-parallel and antiparallel β-sheet against the α-helix (McPhalen & James, 1987). Residues in the core of the protein that have been mutated in this study are shown in Figure 1. In order to minimize structural rearrangements, the mutations made are all designed to be "non-disruptive deletions"; that is, the side chains of residues in the core are truncated, thus removing interactions without introducing any new ones. Mutations made in this study were Ile → Val and Ile → Ala; Val → Ala; and Leu → Ala. In almost all cases, the side chains of the residues are completely buried. The effect of the removal of a hydrophobic side chain from the core of the protein was measured both by GdnHCl-induced denaturation and by differential scanning calorimetry (DSC). The contribution of the hydrophobic side chains to protein stability is analyzed in the context of the environment of the residue in the protein.

EXPERIMENTAL PROCEDURES

Materials

Chemicals. The buffer used in denaturation experiments was 2-(*N*-morpholino)ethanesulfonic acid (MES) purchased from Sigma. Guanidinium chloride was sequanal grade purchased from Pierce Chemicals. Water was purified to 15 MΩ resistance by an Elgastat system. Radiochemicals were from Amersham International. DEAE-Trisacryl was obtained from IBF, Villeneuve La Garenne, France. Plasmid pPO1 was a generous gift from Dr. Flemming Poulsen, Carlsberg Laboratories. Plasmid pTZ18U and the helper phage M13KO7 were obtained from Pharmacia, Sweden. Yeast extract and tryptone were purchased from Lab M, Bury, U.K. Ammonium sulfate was enzyme grade from BDH. All other chemicals were purchased from Sigma.

Construction of a Truncated Version of CI2. The naturally occurring form of chymotrypsin inhibitor 2 is 83 amino acids long. The first 19 amino acids, however, are completely unstructured, resulting in a random-coil N-terminal "tail" (Kjaer et al., 1987; Kjaer & Poulsen, 1987). It is known that these 19 amino acids do not contribute to either the stability or the activity of the protein. Previous folding studies have been carried out on the long form of the protein (Jackson & Fersht, 1991a,b). However, there are problems in the purification of such a form due to the proteolytic cleavage of the N-terminal tail which results in four isoforms. This makes the purification difficult and reduces the overall yield. To overcome these problems, a new plasmid, pCI2, was constructed in which the first 19 amino acids of the CI2 gene have been deleted and the leucine residue at position 20 has been replaced by a methionine. The pCI2 plasmid was constructed by subcloning the truncated CI2 gene into pTZ18U (Mead et al., 1986). The truncated CI2 gene (Met20–Gly83) was obtained as a 328 bp *Bgl*II/*Hin*dIII fragment by double digestion of pPO1 plasmid. This fragment also contained the T7 Φ 10 promoter and ribosome binding site. The fragment was isolated by electroelution and ligated into the *Bam*HI/*Hin*dIII site of the multiple cloning site on pTZ18U. The correct insertion of the isolated fragment was confirmed by sequencing the truncated wild-type gene of CI2, the Φ 10 promoter region, and the ribosome binding site. The new cloning and expression vector, pCI2, has both the bacteriophage f1 origin of replication and the bacteriophage

T7 promoter inserted into the β-galactosidase gene. This vector enables both single-stranded DNA production upon infection with helper phage M13KO7 and also expression of the CI2 gene, which is under the control of the lac and T7 polymerase promoters, and which can be induced with IPTG.

The truncated form of the protein was characterized by N-terminal sequencing and electrospray mass spectroscopy. The activity (K_i for inhibition of subtilisin BPN') and stability of the truncated protein were found to be similar to those of the naturally occurring 83 amino acid form of the protein. The numbering scheme is the same as for the original long form of the protein; i.e., the N-terminal residue (methionine) is residue 20. The "wild-type" in this paper refers to the truncated form of the protein. All site-directed mutagenesis was carried out on the truncated gene.

Mutagenesis. Single-stranded DNA was obtained from the modified pTZ18U, harbored in *Escherichia coli* TG2 after infection with the helper phage M13KO7 following the conditions described by Pharmacia. Site-directed mutagenesis was carried out using the method of Eckstein (Sayers et al., 1988) and the kit supplied by Amersham (using 0.2 of the recommended amounts of DNA and kit). Mutants were identified by directly sequencing the single-stranded DNA. In two cases, Ile → Val76 and Val → Ala66, the mutations were also confirmed on the purified protein by electrospray mass spectroscopy.

Expression and Purification of Wild-Type and Mutants. Wild-type and mutants were expressed and purified using the following protocol. The *E. coli* strain *TG2* containing *pCI2* was grown in 2×TY medium containing ampicillin (50 μg mL^{-1}) and IPTG (0.5 mM) for induction of expression. Cultures were grown overnight at 37 °C with vigorous shaking to an absorbance of 6–8 at 600 nm. Cells were harvested and concentrated by centrifugation (6500*g*), and the cell paste was resuspended in 10 mM Tris-HCl, pH 8.0, and 1 mM EDTA. The cells were lysed by three rounds of freeze–thawing, involving incubation of the resuspended cells in liquid N_2 for 10 min followed by incubation in a 65 °C water bath. Cell debris was removed by centrifugation (25000*g*) before the nucleic acids were removed by addition of poly(ethylenimine) to 1% (w/v). The precipitate was removed by centrifugation (25000*g*). CI2 was then precipitated from the supernatant by the addition of ammonium sulfate to 70% (w/v) and collected by centrifugation (25000*g*). Protein pellets were resuspended in 50 mM Tris-HCl, pH 8.6, and dialyzed against the same buffer until equilibrated. Unwanted proteins were then removed by batch-absorption with DEAE-Trisacryl. CI2 did not bind under the conditions used. The supernatant was filtered and applied to a HiLoad 26/60 Superdex G75 column (Pharmacia) preequilibrated in 50 mM Tris-HCl, pH 8.6, and 150 mM NaCl. The column was run at a flow rate of 2.6 mL min^{-1}, and CI2 eluted after approximately 75 min (200 mL). The protein was homogeneous as judged by SDS–PAGE and isoelectric focusing. Purified CI2 was dialyzed extensively against water to remove salt, flash-frozen, and stored at −70 °C. Yields were typically 20 mg of pure protein/L.

Methods

Spectroscopy. The intrinsic fluorescence of CI2 increases on denaturation which allows unfolding and refolding to be monitored by fluorescence spectroscopy. The maximal change in fluorescence upon denaturation is obtained with an excitation wavelength of 280 nm and an emission wavelength of 356 nm. A Perkin-Elmer LS5B luminescence spectrometer was used for the GdnHCl denaturation studies, with a band-pass of 10 nm.

Chemical Denaturation Experiments. Guanidinium chloride solutions were prepared gravimetrically in volumetric flasks. The guanidinium chloride solutions were divided into 800-μL aliquots with an SMI (American Hospital Supply) positive displacement pipetter with a repetitive pipetting attachment and stored at −20 °C. For each data point in the denaturation experiment, 100 μL of CI2 stock solution in 450 mM MES, pH 6.3, was diluted into 800 μL of the appropriate denaturant concentration, using an SMI positive displacement pipetter. The final CI2 concentration was approximately 2.5 μM. The protein/denaturant solutions were preequilibrated at 25 °C for approximately 1 h. Spectroscopic measurements were carried out in a thermostated cuvette holder at 25 °C, the temperature being monitored throughout the experiment by a thermocouple immersed in a neighbouring cuvette in the cell holder. Guanidinium chloride-mediated denaturation of CI2 has previously been shown to be completely reversible (Jackson & Fersht, 1991a).

Data Analysis

Guanidinium Chloride Denaturation. (i) Calculation of $[GdnHCl]_{50\%}$ and m. This analysis is for a two-state model of denaturation where only the native and the denatured states are populated. The equilibrium constant for unfolding, K_{U-F}, and free energy of unfolding, ΔG_{U-F}, in the presence of a denaturant may be calculated from eq 1 where F is the observed

$$K_{U-F} = (F_N - F)/(F - F_U) = \exp(-\Delta G_{U-F}/RT) \quad (1)$$

fluorescence, F_N and F_U are the values of the fluorescence of the native and denatured forms of the protein, respectively, and R is the gas constant, 8.314 J mol^{-1} K^{-1}. In some cases (Shortle et al., 1989; Santoro & Bolen, 1992), it has been found experimentally that the free energy of unfolding of proteins in the presence of guanidinium chloride is linearly related to the concentration of denaturant (Tanford, 1968; Pace, 1986):

$$\Delta G_{U-F}{}^{D} = \Delta G_{U-F}{}^{H_2O} - m[\text{denaturant}] \quad (2)$$

where $\Delta G_{U-F}{}^{D}$ is the free energy of unfolding at a particular denaturant concentration, $\Delta G_{U-F}{}^{H_2O}$ is the free energy of unfolding in water, and m is a constant that is proportional to the increase in the degree of exposure of the protein on denaturation. Assuming that both F_N and F_U are independent of guanidinium chloride concentration between 0 and 6 M, one obtains eq 3 by combining eq 1 and 2:

$$F = F_N - (F_N - F_U)\frac{\exp[(m[\text{GdnHCl}] - \Delta G_{U-F}{}^{H_2O})/RT]}{1 + \exp[(m[\text{GdnHCl}] - \Delta G_{U-F}{}^{H_2O})/RT]} \quad (3)$$

Alternatively, if it is assumed that both F_N and F_U are linearly dependent on guanidinium chloride concentration ($F_N = \alpha_N + \beta_N[\text{GdnHCl}]$, $F_U = \alpha_U + \beta_U[\text{GdnHCl}]$), then

$$F = \{(\alpha_N + \beta_N[\text{GdnHCl}]) + [(\alpha_U + \beta_U[\text{GdnHCl}]) \times \exp[(m[\text{GdnHCl}] - \Delta G_{U-F}{}^{H_2O})/RT]\}/\{1 + \exp[(m[\text{GdnHCl}] - \Delta G_{U-F}{}^{H_2O})/RT]\} \quad (4)$$

where α_N and α_U are the intercepts and β_N and β_U are the slopes of the base lines at the low (N) and high (U) guanidinium chloride concentrations. The entire data set from the fluorescence-monitored guanidinium chloride-induced dena-

turation experiments can be fitted to eq 3 or 4 using the nonlinear regression analysis program *Kaleidagraph* (version 2.1 Synergy Software, PCS Inc.) and values for $\Delta G_{U-F}^{H_2O}$ and m obtained with their standard errors. However, when comparing the stability of wild-type and mutant proteins, it is often more important to know the accuracy of $[GdnHCl]_{50\%}$, the concentration of guanidinium chloride at which 50% of the protein is denatured, and the accuracy of estimates of ΔG_{U-F}^{D} at various values of [GdnHCl]. At $[GdnHCl]_{50\%}$, it is apparent from eq 2 that $\Delta G_{U-F}^{H_2O} = m[GdnHCl]_{50\%}$; thus

$$\Delta G_{U-F}^{D} = m([GdnHCl]_{50\%} - [GdnHCl]) \quad (5)$$

So, from eq 4 and 5:

$$F = \{(\alpha_N + \beta_N[GdnHCl]) + (\alpha_U + \beta_U[GdnHCl]) \times \exp[m([GdnHCl] - [GdnHCl]_{50\%})/RT]\}/\{1 + \exp[m([GdnHCl] - [GdnHCl]_{50\%})/RT]\} \quad (6)$$

The data were fitted to this equation by nonlinear regression analysis using the general curve fit option in the Kaleidagraph program. This gives the calculated standard errors for individual experimental measurements of m and $[GdnHCl]_{50\%}$.

Using this method, it is found that the errors in $[GdnHCl]_{50\%}$ are significantly smaller than those obtained for either $\Delta G_{U-F}^{H_2O}$ or m. This is because small errors in m can lead to large errors in the estimates of $\Delta G_{U-F}^{H_2O}$ due to the long extrapolation from the transition region, which is around 4 M GdnHCl, to 0 M GdnHCl. Values of $[GdnHCl]_{50\%}$ are found to be very reproducible, typically to within 0.02 M. Identical results for $[GdnHCl]_{50\%}$ are found using either eq 3, 4, or 6.

(ii) Calculation of $\Delta\Delta G_{U-F}$. It has been shown that the measurement of $[GdnHCl]_{50\%}$ is very reproducible. The value of $\Delta\Delta G_{U-F}^{[GdnHCl]_{50\%}}$, the difference in stability between wild-type and mutant proteins at [GdnHCl] = $0.5([GdnHCl]_{50\%} - [GdnHCl]'_{50\%})$, where $[GdnHCl]'_{50\%}$ is the midpoint of unfolding of mutant, can be obtained accurately by applying eq 7 where $\Delta[GdnHCl]_{50\%}$ is the difference in $[GdnHCl]_{50\%}$

$$\Delta\Delta G_{U-F}^{[GdnHCl]_{50\%}} = \langle m \rangle \Delta[GdnHCl]_{50\%} \quad (7)$$

of the wild-type and mutant proteins, $\Delta\Delta G_{U-F}^{[GdnHCl]_{50\%}}$ is their difference in the free energy of unfolding at a mean value of the $[GdnHCl]_{50\%}$ for the two proteins, and $\langle m \rangle$ is the average value of m. The midpoints of unfolding are found to be relatively insensitive to other variables such as sloping base lines, and can be measured reproducibly to within 0.02 M. For mutants of CI2 studied thus far, the m values for mutants and wild-type are the same, within experimental error. We have used a value of $\langle m \rangle$, obtained from measurements on a large number of mutant proteins and repetitive runs on wild-type, of 1.93 ± 0.025 kcal mol^{-2}. This use of a mean value for m allows calculation of $\Delta\Delta G_{U-F}^{[GdnHCl]_{50\%}}$ with a low standard error. $[GdnHCl]_{50\%}$ (± standard error) for wild-type is 3.99 ± 0.02 M, the average of four separate experiments.

A value for $\Delta\Delta G_{U-F}^{H_2O}$ can also be calculated without using an average m value using eq 8:

$$\Delta\Delta G_{U-F}^{H_2O} = \Delta G_{U-F}^{H_2O} - \Delta G'_{U-F}^{H_2O} \quad (8)$$

or at any denaturant concentration using the more general eq 9:

$$\Delta\Delta G_{U-F}^{[GdnHCl]} = m([GdnHCl]_{50\%} - [GdnHCl]) - m'([GdnHCl]'_{50\%} - [GdnHCl]) \quad (9)$$

where m' is the m value of mutant. The data were also fitted to eq 8 and 9 to compare with those values obtained using eq 7. It is apparent that the errors are much larger for this method.

Sloping versus Constant Base Lines and Accuracy of Data. The previous studies on chymotrypsin inhibitor 2 assumed that the fluorescence of the native and denatured forms of CI2 is independent of [GdnHCl]; i.e., the base lines are constant (Jackson & Fersht, 1991a). Studies on mutant proteins indicated that there appear to be small slopes of the initial and final base lines in plots of fluorescence against [GdnHCl] in regions in which the proteins are assumed to be either fully folded or fully unfolded. The sloping base lines at low guanidinium chloride concentrations are very small, while at high guanidinium chloride concentrations they are slightly higher. The slopes of both base lines vary slightly with mutation. Analyzing the data by either eq 3 or eq 4 gives, within experimental error, identical values for $[GdnHCl]_{50\%}$. All values calculated in this paper assume sloping base lines, and data are fitted to eq 4 or 6.

Calculation of Solvent-Accessible Surface Area. The difference in solvent-accessible surface area that is buried on folding between wild-type and mutant is calculated in the following way. The solvent-accessible surface area of the residue in the wild-type protein is calculated using the Lee and Richards (1971) algorithm. A probe radius of 1.4 Å was used; all other atomic radii are given in Chothia (1975). This value is then subtracted from the solvent-accessible surface area for that residue in a tripeptide (Miller et al., 1987). This gives a value A. In parallel, a theoretical mutation was performed by deleting the appropriate methylene group(s) from the coordinate file, and the solvent-accessible area of the mutated residue in the protein was calculated. This was subtracted from the value calculated for the mutated residue in a tripeptide to give a value B. The difference between the two, $A - B$, is the loss of solvent-accessible surface area buried in the folded protein upon mutation.

Calorimetry. Samples for calorimetry were dialyzed extensively against the appropriate buffer at 4 °C, and centrifuged at 13 000 rpm in an MSE microcentrifuge for 10 min before use. The same buffer stock was used for all proteins, and great care was taken to ensure complete equilibration on dialysis to ensure that all samples were prepared at the same pH and ionic strength. The final dialysis buffer was used for the DSC reference and for obtaining instrumental buffer base lines. The protein concentration was determined by absorbance measurements using ϵ = 6965 M^{-1} cm^{-1} at 282 nm. Extinction coefficients were also measured for several mutants, but the value was found to vary by less than 2%, and so the wild-type extinction coefficient was used throughout. DSC measurements were performed using a Microcal MC-2D instrument at a notional scan rate of 60 deg h^{-1}. Sample and reference solutions were degassed under vacuum with gentle stirring for 2–3 min before loading and were held at 2 atm of N_2 pressure during scanning to suppress bubble formation at high temperatures. Samples were generally rescanned after being cooled to below the transition temperature over a period of 50 min to check the reversibility of thermal unfolding. Repeated scans on the same sample were found to be consistent.

Analysis of Calorimetric Data. DSC thermograms were corrected by subtraction of appropriate buffer base lines and converted to excess heat capacity assuming a relative molecular mass of 7308 for all samples and using the cell volume supplied by the manufacturer. Changes in heat capacity between folded and unfolded states were eliminated using an interpolative base line between pre- and posttransition base-line slopes in

proportion to the progress through the transition. Analysis and fitting of thermograms as single-peak non-two-state transitions were performed using Microcal Origin software (version 1.16, Microcal Software Inc.) which is based on standard deconvolution models.

Calculation of ΔG from Calorimetric Data. From the transition curves, T_m (the transition temperature), $\Delta H_{cal}(T_m)$ (the molar calorimetric enthalpy of denaturation at T_m), and $\Delta H_{vH}(T_m)$ (the molar van't Hoff enthalpy of denaturation at T_m) can be calculated. $\Delta G_U(T)$, the free energy of unfolding, can be calculated at any temperature, T, using eq 10 which is derived elsewhere (Baldwin, 1986).

$$\Delta G_U(T) = \Delta H_U(T_m)\left(1 - \frac{T}{T_m}\right) + \Delta C_p\left[(T - T_m) - T\ln\frac{T}{T_m}\right] \quad (10)$$

$\Delta\Delta G_U(T)$ is simply the difference in $\Delta G_U(T)$ calculated for wild-type and mutant proteins at any temperature T.

ΔC_p was redetermined for the truncated form of the protein and the value found to be 0.72 kcal mol^{-1} K^{-1}. This compares with a value of 0.79 kcal mol^{-1} K^{-1} measured for the long form of the protein (Jackson & Fersht, 1991a); i.e., the N-terminal tail contributes little to the change in heat capacity of the protein on denaturation. We assume that ΔC_p is invariant with mutation, and a value of 0.72 kcal mol^{-1} K^{-1} is used to calculate ΔG_U and $\Delta\Delta G_U$ for all mutants.

Crystallization, Data Collection, and Structure Solution. Crystals of the mutant Ile → Val76 were grown under similar conditions to wild-type CI2 (McPhalen & James, 1987) and crystallized isomorphously (see Table IV). Data collection was carried out as indicated in Table IV. All data processing, scaling, molecular replacement, and electron density calculations were carried out with the CCP4 protein crystallography suite of programs (CCP4, 1979). Data processing parameters are shown in Table IV.

Structure of Chymotrypsin Inhibitor 2 and Description of Mutations. The major secondary structural features of CI2 are a central four-stranded mixed parallel and antiparallel "pseudo" β-sheet, flanked on one side by an α-helix of 3.6 turns, and on the other side by the reactive-site loop in an extended conformation. The hydrophobic core of CI2 is located at the interface of the α-helix and the β-sheet. It is comprised of 1 aromatic and 11 aliphatic residues. These residues include Trp24 (β-strand$_1$/type III turn); Leu27 (type III turn/type II turn); Ala35, Val38, and Ile39 (α-helix); Ala46 (type I turn); Ile48 (β-strand$_2$); Val66, Leu68, and Val70 (β–strand$_3$); and Ile76 and Pro80 (β-strand$_4$). The core is centered around Ile39. We have mutated eight of these residues and have introduced "non-disruptive" deletions, i.e., deleted one or more methylene groups from a side chain. Mutations fit into one of three classes: (i) Ile → Val mutations which remove one methyl group; (ii) Val → Ala mutations which remove two methyl groups; and (iii) Ile → Ala and Leu → Ala mutations which remove three methylene groups. The positions of the residues mutated, Leu27, Val38, Ile39, Ile48, Val66, Leu68, Val70, and Ile76 are shown in Figure 1. In addition to the single substitutions, we obtained the double-mutant Ile → Ala48/Ile → Val76, which is also analyzed in this paper. Table I summarizes the interactions and solvent exposure of the hydrophobic core residues analyzed by mutagenesis in this study.

Table I: Interactions and Solvent Exposure of the Hydrophobic Residues Analyzed by Mutagenesis

residue	solvent exposure of side chain (Å^2)	side chains within 4.5 Å
Leu27	1.4	W24, K30, A35, V38, I39, P80
Val38	52	W24, L27
Ile39	0	W24, L27, A46, I48, V66, L68, P80
Ile48	10.8	K36, I39, L40, A46, V50, L68
Val66	0	W24, I39, K43, A46, L68, P80, V82
Leu68	0	V32, A35, I39, I48, V66, I76, P80
Val70	10.1	V32, V50, L68, I76
Ile76	0.1	V32, A35, L68, V70, P80

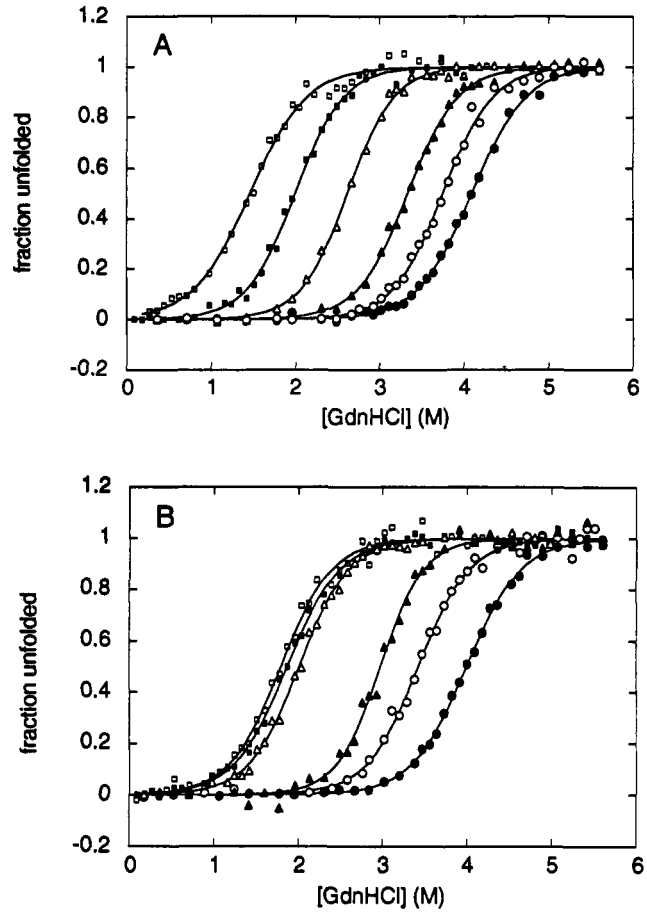

FIGURE 2: Guanidinium chloride-induced equilibrium unfolding of wild-type and mutant chymotrypsin inhibitor 2. (A) (Filled circles) IV76; (open circles) VA38; (filled triangles) IV39; (open triangles) LA27; (filled squares); IA48 (open squares) VA66. (B) (Filled circles) Wild-type; (open circles) IV48; (filled triangles) VA70; (open triangles) LA68; (filled squares) IA48IV76; (open squares) IA76. Unfolding was monitored by the intrinsic fluorescence of CI2 using an excitation wavelength of 280 nm and an emission wavelength of 356 nm. Under these conditions, the base lines of the native and denatured states are found to be dependent upon [GdnHCl]. These plots show the data normalized (fraction unfolded) taking into account the sloping base lines. The solid lines are the best fit of the data to a two-state process.

RESULTS

GdnHCl-Induced Denaturation. Wild-type and all mutants were found to follow two-state unfolding transitions. Typical normalized GdnHCl-induced denaturation curves are shown in Figure 2. Data from such transition curves were fitted to eq 6 to yield values for $[GdnHCl]_{50\%}$ and m for each mutant and associated errors (all errors are calculated from the best fit of the data and are not standard deviations from repetitive runs, unless otherwise stated). These values can then be used

Table II: Changes in the Free Energies of Unfolding upon Mutation As Determined by Reversible Guanidinium Chloride Denaturation Experiments

mutant	$[GdnHCl]_{50\%}$[a] (M)	m[a] (kcal mol^{-2})	$\Delta G_{U-F}^{H_2O}$[b] (kcal mol^{-1})	$\Delta\Delta G_{U-F}^{H_2O}$[c] (kcal mol^{-1})	$\Delta\Delta G_{U-F}^{[GdnHCl]_{50\%}}$[d] (kcal mol^{-1})	$\Delta\Delta G_{cal}(298)$ (kcal mol^{-1})	$\Delta\Delta G_{vH}(298)$ (kcal mol^{-1})
wild-type[e]	3.99 ± 0.02	1.89 ± 0.04	7.53 ± 0.16				
Leu → Ala27	2.62 ± 0.02	2.15 ± 0.13	5.63 ± 0.34	1.91 ± 0.38	2.64 ± 0.06	2.66	2.77
Val → Ala38	3.75 ± 0.03	2.01 ± 0.15	7.54 ± 0.57	0.00 ± 0.59	0.46 ± 0.07	0.46	0.52
Ile → Val39	3.33 ± 0.02	1.99 ± 0.12	6.63 ± 0.40	0.91 ± 0.43	1.27 ± 0.06	1.18[g]	1.10[g]
Ile → Val48[f]	3.43 ± 0.03	1.99 ± 0.18	6.80 ± 0.62	0.75 ± 0.64	1.09 ± 0.07	0.92	1.02
Ile → Ala48	2.00 ± 0.03	2.07 ± 0.19	4.14 ± 0.39	3.40 ± 0.42	3.84 ± 0.09	3.89	3.71
Val → Ala66	1.46 ± 0.10	1.76 ± 0.22	2.57 ± 0.37	4.97 ± 0.40	4.88 ± 0.21	4.91	5.00
Leu → Ala68	2.01 ± 0.03	1.98 ± 0.12	3.98 ± 0.25	3.56 ± 0.30	3.82 ± 0.09	3.74	3.72
Val → Ala70	2.98 ± 0.03	2.15 ± 0.19	6.41 ± 0.57	1.13 ± 0.59	1.95 ± 0.07	2.17	2.14
Ile → Val76[f]	4.10 ± 0.05	1.82 ± 0.12	7.47 ± 0.50	0.08 ± 0.53	−0.21 ± 0.10	−0.09	−0.02
Ile → Ala76	1.79 ± 0.05	1.93 ± 0.21	3.45 ± 0.39	4.09 ± 0.42	4.25 ± 0.12	4.28[g]	3.94[g]
Ile → Ala48/Ile → Val76[f]	1.90 ● 0.04	1.83 ± 0.14	3.46 ± 0.27	4.08 ± 0.32	4.05 ± 0.10	4.69	3.57

[a] Calculated by fitting GdnHCl-induced denaturation data to eq 6. [b] Calculated using eq 5. [c] Calculated using eq 8. [d] Calculated using eq 7. [e] Average of four separate experiments. [f] Average of two separate experiments. [g] Average of experiments at pH 3.0 and 3.5.

Table III: Results of Differential Scanning Microcalorimetry Experiments

protein	pH	T_m (°C)	$\Delta H_{cal}(T_m)$ (kcal mol^{-1})	$\Delta H_{vH}(T_m)$ (kcal mol^{-1})	$\Delta H_{vH}(T_m)/\Delta H_{cal}(T_m)$	$\Delta S_{cal}(T_m)$ (cal mol^{-1} K^{-1})	$\Delta\Delta G_{vH}(298)$[a] (kcal mol^{-1})	$\Delta\Delta G_{cal}(298)$[a] (kcal mol^{-1})	$\Delta\Delta G_{U}^{[GdnHCl]_{50\%}}$[b] (kcal mol^{-1})
wild-type[c]	3.0	64.0	61.0	61.2	1.00	181			
wild-type[c]	3.5	73.8	67.0	66.9	1.00	193			
Leu → Ala27	3.0	48.5	45.0	44.4	0.99	140	2.77	2.66	2.64
Val → Ala38	3.0	61.1	58.5	58.6	1.00	175	0.52	0.46	0.46
Ile → Val39	3.0	57.7	53.7	54.8	1.02	162	1.17	1.20	1.27
Ile → Val39	3.5	67.3	61.5	62.5	1.02	181	1.03	1.16	1.27
Ile → Val48	3.5	69.1	62.4	61.6	0.99	182	1.02	0.92	1.09
Ile → Ala48	3.5	52.5	44.9	46.9	1.05	138	3.71	3.89	3.84
Val → Ala66	3.5	45.3	37.3	35.8	0.96	117	5.00	4.91	4.88
Leu → Ala68	3.5	52.3	46.7	46.9	1.00	144	3.72	3.74	3.82
Val → Ala70	3.0	51.3	48.5	49.7	1.02	150	2.14	2.17	1.95
Ile → Val76[d]	3.0	65.4	60.5	60.5	1.00	179	−0.02	−0.09	−0.21
Ile → Ala76[d]	3.0	38.3	30.2	37.2	1.24	97	4.00	4.23	4.25
Ile → Ala76	3.5	50.3	41.0	46.8	1.14	127	3.87	4.34	4.25
Ile → Ala48/Ile → Val76	3.5	52.8	35.2	48.3	1.37	108	3.57	4.69	4.05

[a] Calculated, at 298 K, from the measured van't Hoff [$\Delta\Delta G_{vH}(298)$] and calorimetric [$\Delta\Delta G_{cal}(298)$] enthalpy of unfolding using eq 10. [b] Calculated from the GdnHCl-induced denaturation experiments fitting data to eq 6 and 7. [c] Values are the average of four separate experiments. Standard deviations are T_m = 64.0 ± 0.2 °C, $\Delta H_{cal}(T_m)$ = 61.0 ± 2.3 kcal mol^{-1}, and $\Delta H_{vH}(T_m)$ = 61.2 ± 1.0 kcal mol^{-1}. [d] Values are the average of two separate experiments.

to calculate $\Delta G_{U-F}^{H_2O}$ ($\Delta G_{U-F}^{H_2O} = m[GdnHCl]_{50\%}$). It is clear from the summarized data in Table II that, within experimental error, the m values for wild-type and mutants are the same. Therefore, $\Delta\Delta G_{U-F}^{[GdnHCl]_{50\%}}$ can be calculated using an average m value according to eq 7. For comparison, $\Delta\Delta G_{U-F}^{H_2O}$ is also calculated, using individual m values, according to eq 8. Both these results are shown in Table II. It is clear that the errors associated with $\Delta\Delta G_{U-F}^{H_2O}$ are substantially larger than in calculating $\Delta\Delta G_{U-F}^{[GdnHCl]_{50\%}}$. This results from small errors in m leading to large errors in $\Delta G_{U-F}^{H_2O}$, because of the long extrapolation, typically from 4 to 0 M GdnHCl. If the m values are identical, then $\Delta\Delta G_{U-F}^{[GdnHCl]_{50\%}} = \Delta\Delta G_{U-F}^{H_2O}$; this is a reasonable assumption given that m varies little with mutation. Values for $\Delta\Delta G_{U-F}^{[GdnHCl]_{50\%}}$ are the most accurate and can usually be measured within 0.1 kcal mol^{-1}. These are the values used in the subsequent analysis. The values of $\Delta\Delta G_{U-F}^{[GdnHCl]_{50\%}}$ calculated represent values obtained in 2.2–4 M GdnHCl. To assess what effect the high concentration of salt has on $\Delta\Delta G_{U-F}^{[dnHCl]_{50\%}}$, calorimetry experiments were also performed.

Calorimetric Denaturation. The values for T_m, $\Delta H_{vH}(T_m)$, and $\Delta H_{cal}(T_m)$ calculated from the calorimetric unfolding curves are shown in Table III. In addition, the ratio of $\Delta H_{vH}(T_m)$ to $\Delta H_{cal}(T_m)$, $\Delta S_{cal}(T_m)$ [$\Delta S_{cal}(T_m) = \Delta H_{cal}(T_m)/T_m$], and the difference in the free energy of unfolding between wild-type and mutant at 298 K, calculated both from the calorimetric enthalpy of unfolding, $\Delta\Delta G_{cal}(298)$, and from the van't Hoff enthalpy of unfolding, $\Delta\Delta G_{vH}(298)$, using eq 10, are also shown.

Experiments were performed at either pH 3.0 or pH 3.5. Highly destabilized mutants were measured at pH 3.5. In addition to the wild-type protein, there were two mutants, Ile → Val39 and Ile → Ala76, for which measurements were performed at both pH 3.0 and pH 3.5. The results obtained at the two pHs were within experimental error. The experimental error involved in measuring $\Delta\Delta G_{cal}(298)$ or $\Delta\Delta G_{vH}(298)$ is estimated to be 10%. In most cases, the calorimetrically-determined $\Delta\Delta G_{U-F}$ is the same, within experimental error, as the fluorescence-monitored GdnHCl-induced denaturation. In one case, however, Ile → Ala48/Ile → Val76, the value for $\Delta\Delta G_{cal}(298)$ is slightly outside this error limit. This mutant is very destabilized, resulting in a low T_m, 52.8 °C. At these temperatures, the unfolding transition occurs at a point where the base line is strongly curved. Assigning base lines to the transition curves is much more difficult in this region and results in larger errors. Errors for this mutant are more likely to be on the order of 20%. Within these limits, the results from GdnHCl denaturation and calorimetry are the same.

The ratio of the van't Hoff enthalpy of denaturation to the calorimetric enthalpy of denaturation, $\Delta H_{vH}(T_m)/\Delta H_{cal}(T_m)$, gives a measure of how closely unfolding approaches a two-state transition. If the ratio is 1, the transition is a two-state

process; if it is greater than 1, then either there is an error in measuring $\Delta H_{vH}(T_m)$ or $\Delta H_{cal}(T_m)$ or there is an additional process which produces heat, such as aggregation. The ratios are determined with relatively large errors, contributions from both the calorimetric experiment and the determination of the protein concentration. In most cases, the ratio is close to 1 (see Table III), indicating that the transitions are two-state. However, Ile → Ala76 and the double-mutant Ile → Ala48/Ile → Val76 both have ratios which are significantly greater than 1. In these cases, the protein is substantially destabilized, between 4 and 5 kcal mol^{-1}, and the transition temperatures are between 38 and 53 °C. As previously discussed, base lines in this region are strongly dependent upon temperature. This means that it is difficult to assign base lines to transition curves. Errors in the base line can result in significant errors in $\Delta H_{vH}(T_m)$. Thus, it appears in these cases that $\Delta H_{vH}(T_m)$ is overestimated, resulting in a high $\Delta H_{vH}(T_m)/\Delta H_{cal}(T_m)$ ratio. $H_{cal}(T_m)$, which is not as sensitive to changes in the base lines as $H_{vH}(T_m)$, is, therefore, more accurate and used to calculate $\Delta\Delta G_{U-F}$. There is remarkable agreement between the values of $\Delta\Delta G_{U-F}$ determined from $\Delta H_{cal}(T_m)$ and those from GdnHCl denaturation using the simple eq 7. The results agree in general within 1% or 2% (Table II).

Double-Mutant Cycle Analysis of Ile → Ala48Ile → Val76. Double-mutant cycles can be used to measure the interaction energy, $\Delta\Delta G_{int}$, between two residues in a protein. The general theory of double-mutant cycles has been discussed extensively elsewhere (Horowitz & Fersht, 1990). The side chains of Ile48 and Ile76 are both buried in the hydrophobic core of CI2. However, they are not directly in contact; the closest distance between the two side chains is 8.0 Å. From the GdnHCl-induced equilibrium unfolding data for the single mutations (Ile → Ala48 and Ile → Val76) and for the double mutant (Ile → Ala48/Ile → Val76), the interaction energy between the two residues in the native state can be calculated, $\Delta\Delta G_{int} = 0.42$ kcal mol^{-1}. This indicates that the Ile → Ala48 mutation destabilizes the protein to a greater extent in the presence of the Ile → Val76 mutation than in the wild-type protein. This is probably the result of small structural rearrangements in the core of the protein on the creation of a cavity on mutation of Ile → Val76 (see next section).

Crystal Structure of Ile → Val76. Refinement was carried out using the Hendrickson–Konnert restrained least-squared refinement program PROLSQ (Hendrickson & Konnert, 1980). The final refinement parameters and results are shown in Table V. The side chain of the mutated residue, Val76, is well-defined (B-factor = 9.16 Å^2) and in a single conformation ($\chi_1 = -57°$). A comparison shows that the structure of Ile → Val76 is largely superimposable with that of wild-type. Differences in surface side chains were observed in regions involving crystal packing and in the large, partially disordered, reactive-site loop. This results in an rms difference between the side chains in wild-type and mutant of 0.56 Å (Table V). The rms difference of side chains in the core of the protein, however, is significantly smaller, 0.31 Å (Table V).

There are small structural rearrangements in the hydrophobic core resulting from the Ile → Val76 mutation. The largest shifts in atomic position are found at Val76 due to repacking of its side chain toward the cavity (C^β moves by 0.44 Å, $C^{\gamma 1}$ by 0.67 Å, and $C^{\gamma 2}$ by 0.50 Å) (see Figure 4). The Val76 side chain rotates very little, and the repacking toward the cavity is from a linear shift of the whole side chain. There are other small movements in the core which also contribute to repacking around the cavity ($C^{\gamma 2}$ of Val38 moves by 0.48 Å). The side chain of Ile48 also moves (C^β by 0.40 Å, $C^{\gamma 1}$ by 0.66 Å, $C^{\gamma 2}$ by 0.40 Å, and $C^{\delta 1}$ by 0.37 Å). It is not clear whether this is a result of the mutation (the closest contact between Ile48 and Ile76 is 8.0 Å) or whether it may result from small movements of an adjacent side chain, Leu40, which is a surface residue. There is no evidence, as in other studies (Eriksson et al., 1992; Buckle et al., 1993), that a solvent molecule occupies the cavity created by the mutation.

Table IV: Crystal Parameters, Data Collection Conditions, and Data Processing Parameters for Ile → Val76

	Ile → Val76
precipitant	40% (w/v) $(NH_4)_2SO_4$
buffer	50 mM Tris HCl, pH 8.0
temp (°C) of data collection	4–6
intensity collection	
detector	MAR image plate
wavelength (Å)	1.5418
mounting axis	c*
oscillation range (deg)	0–60
oscillation step size (deg)	1
unit-cell dimensions	
$a = b$ (Å)	68.93
c (Å)	53.09
space group	*P622*
no. of molecules in asymmetric unit	1
resolution (Å)	2.2
reflections observed	35386
R-merge (highest resolution shell) (%)	10.7 (24.7)
unique reflections	3857
completeness (highest resolution shell) (%)	95.4 (95.2)
multiplicity (highest resolution shell)	7.0 (7.3)
Wilson scaling: B	24.19
overall $\langle F/\sigma F \rangle$ (highest resolution shell)	31.68 (15.37)

Table V: Final Refinement Parameters and Results

	Ile → Val76
refinement procedure	PROLSQ and molecular replacement[a]
crystallographic R factor (%)	17.1
no. of protein atoms	514
no. of solvent atoms	32
total no. of atoms	546
overall mean B-value (Å^2)	20.1
rms difference to wild-type structure (Å)	
all atoms[b]	
main chain	0.21
side chain	0.56
hydrophobic core only[c]	
main chain	0.20
side chain	0.31
model geometry (rms deviations from ideality)	
bond lengths (Å)	0.02
side-chain torsion angles (deg)	16.5
peptide Ω angles (deg)	3.2
B-value restraints (rms differences) (Å^2)	
main-chain bond-related	1.23
main-chain angle-related	1.96
side-chain bond-related	2.94
side-chain angle-related	4.63

[a] Hendrickson & Konnert (1980). [b] Calculated using the LSQKAB program, least-squares fit of all main-chain atoms from Lys21 to Gly83. [c] Calculated for the 12 hydrophobic core residues only (Trp24, Leu27, Ala35, Val38, Ile39, Ala46, Ile48, Val66, Leu68, Val70, Ile76, and Pro80).

DISCUSSION

Validity of Using $\langle m \rangle$ To Calculate $\Delta\Delta G_{U-F}$ and Comparison of GdnHCl Denaturation and Calorimetric Experiments. $\Delta\Delta G_{U-F}$ can be calculated from the GdnHCl denaturation data using two different methods. Individual m values can be used and $\Delta\Delta G_{U-F}$ calculated according to eq 8. This gives the value of $\Delta\Delta G_{U-F}^{H_2O}$ measured in water. It is clear that the errors associated with $\Delta\Delta G_{U-F}^{H_2O}$ are quite large, typically ±0.5 kcal mol^{-1} (see Table II). This is the

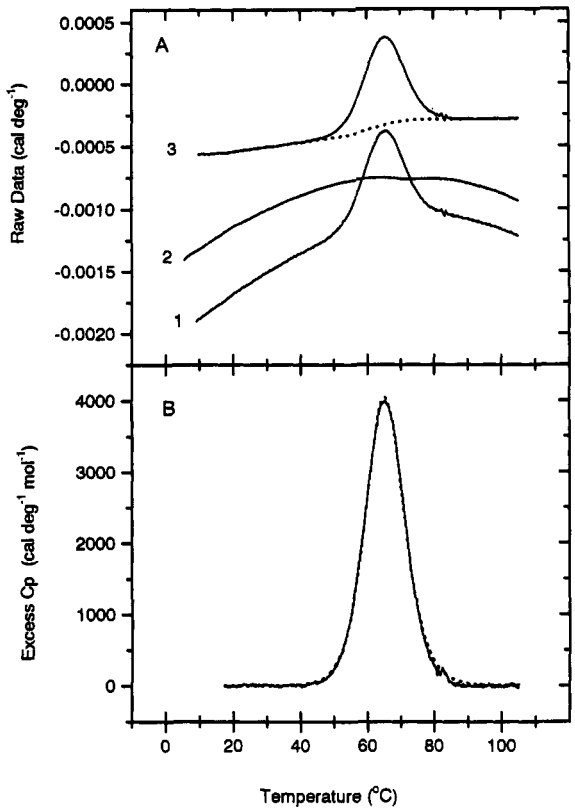

FIGURE 3: Typical DSC thermograms for the denaturation of CI2 mutant Ile → Val76 at a protein concentration of 1.07 mg mL^{-1}, 10 mM glycine hydrochloride, pH 3.0. (A) Raw data showing (1) CI2 and (2) control buffer scans with (3) CI2 data corrected for the buffer base line and indicating the interpolative base line (dotted line) used to eliminate the denaturation heat capacity changes. (B) Normalized data showing deconvolution according to a non-two-state transition model with the best fit to the data (dotted line). The T_m is 65.2 °C with a calorimetric enthalpy of 61.9 kcal mol^{-1} and a van't Hoff enthalpy of 59.5 kcal mol^{-1} [$\Delta H_{cal}(T_m)/\Delta H_{vH}(T_m)$ = 1.04].

result of small errors in m leading to large errors in $\Delta G_{U-F}^{H_2O}$ because of the long extrapolation from 2.2–4 M GdnHCl to water. The second method uses an average value of m, $\langle m \rangle$, obtained from measurements on a large number of mutant proteins and repetitive runs on wild-type. $\Delta\Delta G_{U-F}$, in this case, can be calculated according to eq 7, at the mean value of $[GdnHCl]_{50\%}$ for wild-type and mutant protein, *i.e.*, 2.2–4 M GdnHCl. As $\Delta\Delta G_{U-F}$ in this case is calculated at a concentration of GdnHCl at which the unfolding curve is measured, and does not involve a long extrapolation, the errors are considerably smaller, typically less than ±0.1 kcal mol^{-1} (Table II). This method is only valid if there are no significant changes in the m value on mutation. The m values calculated for the hydrophobic core mutants in this study do not differ significantly from the wild-type value, and are all within experimental error. Thus, there is strong evidence that this assumption, and the method, is valid.

A comparison of the values for $\Delta\Delta G_{U-F}$ from GdnHCl denaturation and from calorimetric experiments supports this assumption further. Comparing $\Delta\Delta G_{cal}(298)$ with the values for $\Delta\Delta G_{U-F}$ obtained using both eq 7, $\Delta\Delta G_{U-F}^{[GdnHCl]_{50\%}}$, and eq 8, $\Delta\Delta G_{U-F}^{H_2O}$ (see Table II), it is clear that the calorimetrically determined values are much closer to those determined using an average value for m, than those determined using individual m values. Thus, we have shown that the most accurate method of determining $\Delta\Delta G_{U-F}$ from GdnHCl data is by using an average m value and that the assumptions made in this method are valid. In addition, we have shown that the value of $\Delta\Delta G_{U-F}^{[GdnHCl]_{50\%}}$, which is calculated between 2.2 and 4 M GdnHCl depending upon mutation, is an accurate measure of $\Delta\Delta G_{U-F}$ in water. It should be noted that this applies to CI2 and need not apply elsewhere. In particular, m for staphylococcal nuclease changes dramatically on mutation (Shortle et al., 1990).

Comparison of Results with Other Proteins. Using guanidinium chloride-induced denaturation of chymotrypsin inhibitor 2, we have been able to measure, to a high degree of accuracy, the effect of cavity-creating mutations on the stability of a protein. These results allow us to estimate the energetic cost of deleting one (Ile → Val), two (Val → Ala), or three (Ile → Ala; Leu → Ala) methyl(ene) groups from within the core of a protein. For CI2, the average change in the free energy of unfolding (± standard deviation) for an Ile → Val mutation is 1.18 ± 0.09 kcal mol^{-1} (Ile → Val76 was not included in this analysis because of evidence that there are structural rearrangements on mutation), for a Val → Ala mutation 3.42 ± 1.47 kcal mol^{-1} (Val → Ala38 was not included in the average as this residue is not completely buried), and for either a Ile → Ala or a Leu → Ala mutation 3.64 ± 0.60 kcal mol^{-1}. These results can be compared to the results obtained for barnase (Serrano et al., 1992), staphylococcal nuclease (Shortle et al., 1990), and gene V protein (Sandberg & Terwilliger, 1991), which are 1.52 ± 0.58 kcal mol^{-1} for an Ile → Val mutation, 2.76 ± 0.48 kcal mol^{-1} for a Val → Ala mutation, and 4.38 ± 1.39 kcal mol^{-1} for either an Ile → Ala or a Leu → Ala mutation. Within the distributions, the results from the four different proteins agree. For CI2, the average change in the free energy of unfolding for deleting one methyl(ene) group is 1.33 ± 0.45 kcal mol^{-1}. Dividing the changes in stability by the changes in the solvent-accessible surface area for the hydrophobic side chains of Leu, Ile, Val, and Ala in solution (Miller et al., 1987) gives an average value of 55 cal mol^{-1} Å^{-2} for CI2. This value is 2–3 times the value found in model studies (Chothia, 1976; Eisenberg & McLachlan, 1986; Ooi et al., 1987) but is very similar to the value of 63 cal mol^{-1} Å^{-2}, an average value found for barnase, staphylococcal nuclease, and gene V protein.

$\Delta\Delta G_{U-F}$ and Statistical Correlations with Parameters of Residue Environment. It is apparent from Table II that identical mutations at different positions in the protein give rise to different values of $\Delta\Delta G_{U-F}$. These variations must be the result of variations in the environment surrounding the different residues as they cannot result from differences in the hydrophobicity of the side chain. In a similar manner to other studies (Shortle et al., 1990; Serrano et al., 1992), statistical correlations were sought between the values of $\Delta\Delta G_{U-F}$ and various parameters which were indicators of the environment surrounding a residue. These were calculated from the crystal structure of the wild-type protein (McPhalen & James, 1987). A number of strong correlations were found.

There is a correlation between $\Delta\Delta G_{U-F}$ and the differences in the side-chain solvent-accessible area buried between wild-type and mutant (see Figure 5A). For CI2, the correlation coefficient of the linear fit is 0.79, and the number of data points, n, is 10. Similar results have been found for another protein, barnase (Serrano et al., 1992). In this case, 20 mutants were analyzed, and the correlation coefficient is 0.83. The data for CI2 and for barnase are found to fall on the same line (see Figure 5A), suggesting that this is a general relationship. A slightly better correlation is found between $\Delta\Delta G_{U-F}$ and N, the number of methylene side-chain groups surrounding the methylene groups that are deleted by mutation (see Figure 5B). The best correlation is found when a cutoff

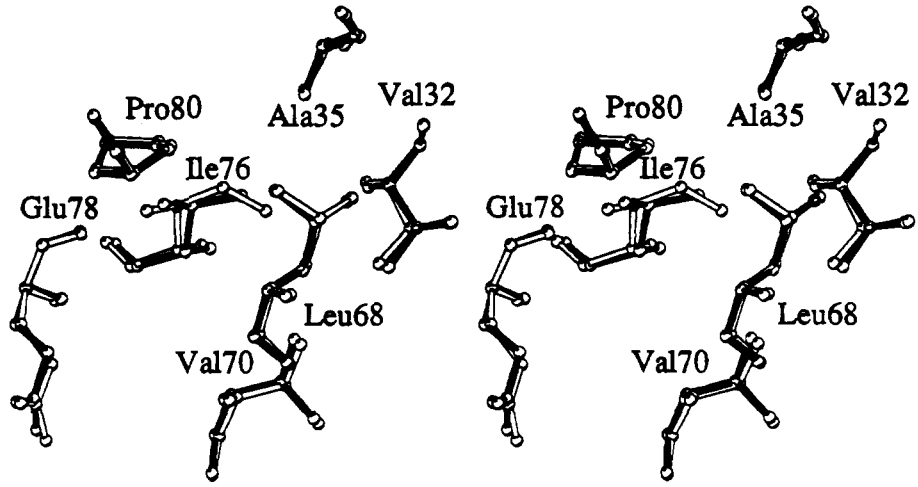

FIGURE 4: Stereo picture showing the superposition of the structures of wild-type CI2 (solid bonds) and the mutant Ile → Val76 (open bonds). Residues within a sphere of radius 4.5 Å of the mutation are shown.

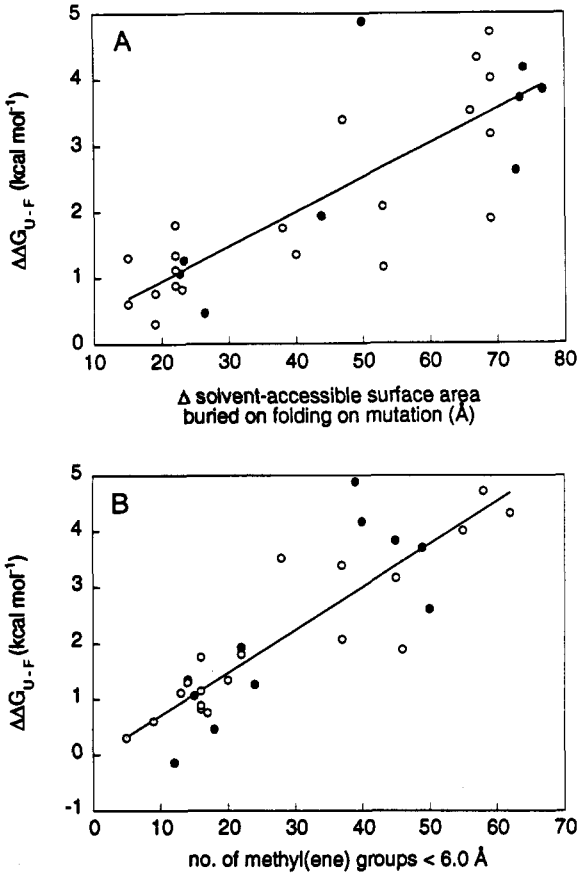

FIGURE 5: (A) Correlation between the difference in the solvent-accessible surface area that is buried on folding between wild-type and mutant side chains and the changes in the free energy of unfolding for mutations of hydrophobic residues for CI2 (filled circles) and barnase (open circles). The solid line shows the best fit of combined CI2 and barnase data to a linear equation (correlation coefficient = 0.82). (B) Correlation between the number of side-chain methylene groups, in a radius of 6 Å of the group deleted from wild-type, and the changes in the free energy of unfolding for mutations of hydrophobic residues for CI2 (filled circles) and barnase (open circles). The solid line shows the best fit of combined CI2 and barnase data to a linear equation (correlation coefficient = 0.87).

radius of 6 Å is used, at which point the correlation coefficient is 0.84, $n = 10$. For barnase, the correlation coefficient is 0.90, $n = 20$ (Serrano et al., 1992). A comparison of the data for CI2 and barnase (Figure 5B) again shows that the two sets of data are co-linear, suggesting that this is a general feature of the hydrophobic cores of globular proteins. The combined data for CI2 and barnase clearly show a better correlation with N (correlation coefficient of 0.87, $n = 30$) than with change in the solvent-accessible surface area (correlation coefficient of 0.82, $n = 30$). Since N depends to some extent on the number of groups deleted from the side chain, we checked for a correlation between $\Delta\Delta G_{U-F}$ and the number of methylene groups deleted on mutation. For CI2 and barnase, there is a weaker correlation with a correlation coefficient of 0.69. To check further that we are not just examining an effect due to just the size of the chain deleted, we examined a single set of mutations. We chose Ile → Val mutations as these are the most conservative. The correlations between $\Delta\Delta G_{U-F}$ and the difference in the side-chain solvent-accessible area buried between wild-type and mutant, and N, are shown in Figure 6A and Figure 6B, respectively. The correlation coefficients are 0.59 and 0.83, respectively ($n = 10$). Thus, we have shown that there is a strong correlation between $\Delta\Delta G_{U-F}$ and N irrespective of the size of the group deleted.

Pace (1992) found a correlation between $\Delta\Delta G_{U-F}$ and ΔG_{tr}, the free energy change on transferring an amino acid side chain from water to 1-octanol (Fauchere & Pliska, 1983), corrected to take into account the volume changes (Sharp et al., 1991). In this case, an average value of $\Delta\Delta G_{U-F}$ was used based on experiments on barnase, staphylococcal nuclease, T_4 lysozyme, and gene V protein. However, the standard deviation of the sample used was large. Following the same analysis on CI2, we find a much weaker correlation between $\Delta\Delta G_{U-F}$ and ΔG_{tr} (correlation coefficient is 0.67) compared with N.

For staphylococcal nuclease, the best statistical correlation was found to be the number of C^α within 10 Å of the C^α of the residue affected by mutation (Shortle et al., 1990). In this case, the correlation coefficient was 0.68 for mutations to alanine and 0.76 for mutations to glycine. Analyzing mutations to alanine only, a correlation is also found between $\Delta\Delta G_{U-F}$ and the number of C^α within 10 Å of the C^α of the residue affected by the mutation, for CI2. However, it is much weaker than the correlations discussed above; the correlation coefficient is 0.66. In addition, for staphylococcal nuclease a correlation was also found between $\Delta\Delta G_{U-F}$ and the crystallographic thermal B-factor of the C^β of the mutated residue. The correlation coefficient was 0.68 for mutations to alanine and 0.75 for mutations to glycine (Shortle et al., 1990). For CI2, we find no correlation between $\Delta\Delta G_{U-F}$ and B-factors (correlation coefficients = 0.04).

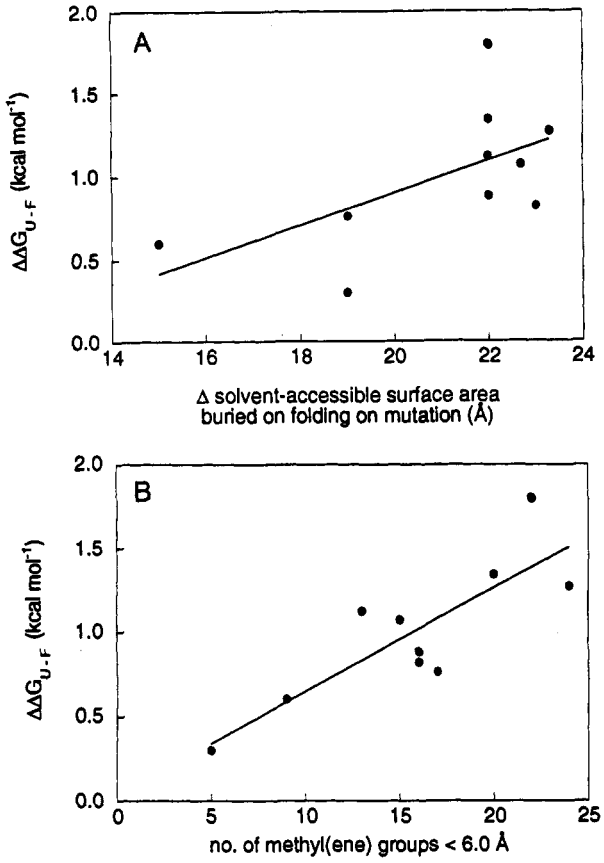

FIGURE 6: (A) Correlation between the difference in the solvent-accessible surface area that is buried on folding between wild-type and mutant side chains and the changes in the free energy of unfolding for Ile → Val mutations only for CI2 and barnase. The solid line shows the best fit of the data to a linear equation (correlation coefficient = 0.59). (B) Correlation between the number of side-chain methylene groups, in a radius of 6 Å of the group deleted from wild-type, and the changes in the free energy of unfolding for Ile → Val mutations only for CI2 and barnase. The solid line shows the best fit of combined CI2 and barnase data to a linear equation (correlation coefficient = 0.83).

Ile → Val76. The interactions that the side chain of residue 76 makes in the crystal structure of wild-type CI2, both free inhibitor and complexed with subtilisin BPN′, and in the mutant structure are listed in Table VI. Using a 3.6-Å cutoff for van der Waals interactions, it can be seen that the Ile → Val76 mutation results in no net gain or loss of any contacts. The $C^{\delta 1}$ atom makes a single contact in the wild-type structure (see footnote to Table VI), indicating that the structure is not densely packed in this region. The small movements in side-chain positions partially compensate for the creation of a cavity within the core of the protein, thus reducing the destabilization energy. However, results from other studies (Eriksson et al., 1992; Buckle et al., 1993) suggest that these small rearrangements could not account for the value of $\Delta\Delta G_{U-F}$ measured for this mutation. Other factors which may account for this result include release of conformational strain from the side chain of Ile76 or removal of an unfavorable contact ($C^{\gamma 2}$ is 2.7 and 2.9 Å from O^{78} in the wild-type structure in the free inhibitor and in the complex, respectively, and 3.2 Å in the mutant structure). It is not clear to what extent these different factors affect the energy of the native state, and all may contribute to reducing $\Delta\Delta G_{U-F}$. It is interesting to note that Ile76 is a valine in most other homologous proteins (Sander & Schneider, 1991).

Conclusions. The results on the hydrophobic core mutants of chymotrypsin inhibitor 2 show that hydrophobic interactions are very important in determining protein stability. The energetics of the hydrophobic interactions are found to be quite variable and context-dependent. Correlations between $\Delta\Delta G_{U-F}$ and various parameters which are indicators of the environment around the mutated residue suggest that the most important factor in determining the effect of hydrophobic mutations on protein stability is the packing density in the area of the mutated residue. For residues in a densely packed region, structural rearrangements will be minimized because of the large number of contacts made with each other by the residues surrounding the site of mutation. In this case, the reorganization energy will be small. For residues which are not in a densely packed region, structural rearrangements on mutation can occur which will result in a decrease in $\Delta\Delta G_{U-F}$, e.g., Ile → Val76. For CI2, such a packing density seems to be correlated with changes in the side-chain solvent-accessible area buried between wild-type and mutant. These results are in agreement with studies on other proteins (Shortle et al., 1990; Serrano et al., 1992), suggesting that these correlations are general and apply to the hydrophobic cores of all globular proteins.

Table VI: Interactions within 4.0 Å of the Side Chain of Residue 76 in the Crystal Structure of the Wild-Type (Free Inhibitor), the Wild-Type Complexed with Subtilisin BPN′, and Ile → Val76

	wild-type CI2[a] (free inhibitor)	wild-type CI2[b] (complex with subtilisin BPN′)	Ile → Val76[c]
C^{β}	N^{77} 3.5 Å O^{27} 3.5 Å C^{75} 3.6 Å O^{30} 3.8 Å	O^{27} 3.4 Å N^{77} 3.5 Å C^{75} 3.8 Å $C^{\alpha 28}$ 3.9 Å O^{30} 3.9 Å	O^{27} 3.4 Å O^{30} 3.5 Å N^{77} 3.6 Å C^{75} 3.7 Å
$C^{\gamma 1}$	O^{30} 3.6 Å $C^{\beta 35}$ 3.8 Å O^{27} 4.0 Å	O^{30} 3.5 Å O^{27} 3.5 Å $C^{\beta 35}$ 3.7 Å	O^{78} 3.2 Å N^{77} 3.6 Å O^{27} 3.6 Å $C^{\gamma 80}$ 3.8 Å H_2O 3.9 Å
$C^{\gamma 2}$	O^{78} 2.7 Å N^{77} 3.4 Å $C^{\gamma 80}$ 3.7 Å N^{78} 3.9 Å C^{78} 3.9 Å $C^{\delta 80}$ 3.9 Å	O^{78} 2.9 Å N^{77} 3.5 Å $C^{\gamma 80}$ 3.7 Å $C^{\delta 80}$ 3.9 Å $C^{\alpha 28}$ 4.0 Å N^{78} 4.0 Å C^{78} 4.0 Å H_2O 4.0 Å	O^{30} 3.3 Å $C^{\beta 35}$ 3.7 Å C^{75} 4.0 Å
$C^{\delta 1}$	*d*	$C^{\beta 35}$ 3.5 Å	

[a] McPhalen & James (1987). [b] McPhalen et al. (1985). [c] Side chain of Val76 in the mutant structure has $\chi_1 = -57°$. [d] Contacts for this methyl group are not included in this analysis because the side chain of Ile76 is in an eclipsed conformation in this structure ($\chi_1 = -61°$, $\chi_2 = 4°$). This is a highly unlikely side-chain conformation, and, therefore, the position of the $C^{\delta 1}$ methyl group is probably not reliable. In this case, the position of the $C^{\delta 1}$ methyl group of Ile76 in the complex structure is taken as a more reliable guide to the contacts made by this group. In the structure of the complex, Ile76 has a standard side-chain conformation ($\chi_1 = -52°$, $\chi_2 = -51°$).

ACKNOWLEDGMENT

We thank Dr. Flemming Poulsen, Carlsberg Laboratories, for the generous gift of plasmid pPO1, Michael Rheinnecker for constructing the pCI2 plasmid, and P. Skelton for running the mass spectrometer. We are also very grateful to Dr. Kim Henrick and Ashley Buckle at the IRC for Protein Engineering for advice and help with the X-ray crystallography.

REFERENCES

Alber, A. (1989) *Annu. Rev. Biochem. 58*, 765–798.

Baldwin, R. L. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 8069–8072.
Buckle, A. M., Henrick, K., & Fersht, A. R. (1993) *J. Mol. Biol.* (in press).
Chothia, C. (1975) *Nature 254*, 304–308.
Chothia, C. (1976) *J. Mol. Biol. 105*, 1–14.
Dill, A. K. (1990) *Biochemistry 29*, 7133–7155.
Eisenberg, D., & McLachlan, A. D. (1986) *Nature 319*, 199–203.
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.
Fauchere, J. L., & Pliska, V. E. (1983) *Eur. J. Med. Chem. 18*, 369–375.
Fersht, A. R. (1985) *Enzyme Structure and Mechanism*, W. H. Freeman and Company, New York.
Fersht, A. R., & Serrano, L. (1993) *Curr. Opin. Struct. Biol. 3*, 75–83.
Hendrickson, W. A., & Konnert, J. H. (1980) *Computing* in *Crystallography* (Diamond, R., Ramesseshan, S., & Venkatesan, K., Eds.) pp 1301–1323, Indian Academy of Sciences, Bangalore.
Horovitz, A., & Fersht, A. R. (1990) *J. Mol. Biol. 214*, 613–617.
Jackson, S. E., & Fersht, A. R. (1991a) *Biochemistry 30*, 10428–10435.
Jackson, S. E., & Fersht, A. R. (1991a) *Biochemistry 30*, 10428–10435.
Kellis, J. T. J., Nyberg, K., Sali, D., & Fersht, A. R. (1988) *Nature 333*, 784–786.
Kellis, J. T. J., Nyberg, K., & Fersht, A. R. (1989) *Biochemistry 28*, 4914–4922.
Kjaer, M., & Poulsen, F. M. (1987) *Carlsberg Res. Commun. 52*, 355–362.
Kjaer, M., Ludvigsen, S. J., Sorenson, O. W., Denys, L. A., Kindtler, J., & Poulsen, F. M. (1987) *Carlsberg Res. Commun. 52*, 327–354.
Kraulis, P. (1991) *J. Appl. Crystallogr. 24*, 946–950.
Lee, B., & Richards, F. M. (1971) *J. Mol. Biol. 55*, 379–400.
Matsumara, M., Becktel, W. J., & Matthews, B. W. (1988) *Nature 334*, 406–410.
McPhalen, C. A., & James, M. N. G. (1987) *Biochemistry 26*, 261–269.
Miller, S., Janin, J., Lesk, A. M., & Chothia, C. (1987) *J. Mol. Biol. 196*, 641–656.
Nakai, K., Kidera, A., & Kanehisa, M. (1988) *Protein Eng. 2*, 93–100.
Ooi, T., Oobatake, M., Nemethy, G., & Scheraga, H. A. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 3086–3090.
Pace, C. N. (1986) *Methods Enzymol. 131*, 266–279.
Pace, C. N. (1992) *J. Mol. Biol. 226*, 29–35.
Privalov, P. L. (1979) *Adv. Protein Chem. 33*, 167–236.
Rose, G. D., Geselowitz, A. R., Lesser, G. J., Lee, R. H., & Zehfus, M. (1985) *Science 229*, 834–838.
Sandberg, W. S., & Terwilliger, T. C. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 1706–1710.
Sander, C., & Schneider, R. (1991) *Proteins: Struct., Funct., Genet. 9*, 56–68.
Santoro, M. M., & Bolen, D. W. (1992) *Biochemistry 31*, 4901–4907.
Sayers, J. R., Schmidt, W., & Eckstein, F. (1988) *Nucleic Acids Res. 16*, 791–802.
Serrano, L., Kellis, J. T., Jr., Cann, P., Matouschek, A., & Fersht, A. R. (1992) *J. Mol. Biol. 224*, 783–804.
Sharp, K. A., Nicholls, A., Friedman, R., & Honig, B. (1991) *Biochemistry 30*, 9686–9697.
Shortle, D., Meeker, A. K., & Gerring, S. L. (1989) *Arch. Biochem. Biophys. 272*, 103–113.
Shortle, D., Stites, W. E., & Meeker, A. K. (1990) *Biochemistry 29*, 8033–8041.
Tanford, C. (1968) *Adv. Protein Chem. 23*, 121–282.
Yutani, K., Ogasahara, K., Tsujita, T., & Sugino, Y. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 4441.