Context Dependence of Protein Secondary Structure Formation: The Three-Dimensional Structure and Stability of a Hybrid between Chymotrypsin Inhibitor 2 and Helix E from Subtilisin Carlsberg†

Peter Osmark, Poul Sørensen, and Flemming M. Poulsen*

Carlsberg Laboratorium, Kemisk Afdeling, Gamle Carlsberg Vej 10, DK-2500 Valby, Copenhagen, Denmark

Received February 22, 1993; Revised Manuscript Received June 15, 1993®

ABSTRACT: The loop region of chymotrypsin inhibitor 2 from barley has been employed as a scaffold for testing the intrinsic propensity of a peptide fragment to form a secondary structure. The helix formation of the nine amino acid residue segment Lys-Gln-Ala-Val-Asp-Asn-Ala-Tyr-Ala of helix E from subtilisin Carlsberg has been studied by the construction of a hybrid consisting of chymotrypsin inhibitor 2 (CI2) where part of the active loop has been replaced by the nonapeptide. An expression system for a truncated form of CI2 where the 19 structureless residues of the N-terminus have been removed and Leu20 replaced by methionyl was constructed from the entire 83-residue wild-type CI2 gene by polymerase chain reaction methodology. The gene encoding the hybrid was constructed from the truncated inhibitor gene. The stability of the truncated inhibitor and of the hybrid toward guanidinium chloride denaturation was examined. From these measurements, the energy of unfolding in pure water was extrapolated to 30.5 ● 1.0 kJ/mol for the truncated inhibitor and 10.9 ± 0.3 kJ/mol for the hybrid. These energies show that the stability of CI2 is unaffected by the N-terminal truncation but severely decreased by the loop mutations. The three-dimensional structure of the hybrid protein has been determined in solution by nuclear magnetic resonance spectroscopy using 893 distance restraints and 84 torsional angle restraints. The average root-mean-square deviation (rmsd) of 15 structures compared to their geometrical average was 0.8 ± 0.2 Å for heavy backbone atoms and 1.3 ± 0.2 Å for all heavy atoms. The inserted peptide segment does not form an α-helix in the new structural context whereas the structure of the CI2 scaffold turns out to be amazingly conserved.

A considerable number of questions remains to be answered before a complete understanding is available of all the mechanisms involved in the process that forms a globular protein from a polypeptide chain. One question is concerned with the folding pathway of the polypeptide chain. For many small globular proteins the observations have been that the folding process is a reversible one-step process, essentially instantaneous and autonomous. Recently, however, a number of millisecond resolved chemical kinetics studies have revealed that certain stretches in the peptide chain tend to form secondary structure elements, in particular helix- and turn-like structures prior to the final folding (Udgaonkar & Baldwin, 1988; Roder et al., 1988; Bycroft et al., 1990; Radford et al., 1992). These prefolding structures of the peptide chain are of considerable interest because it has been suggested that they could be initiation sites for protein folding. Therefore, in a recently published series of studies the ability to form secondary structure elements was examined for shorter peptides with sequences copied from segments of two proteins: myohemerythrin, a four-helical bundle, and plastocyanin, a β-sandwich protein (Dyson et al., 1992a,b). The results of these examinations suggested that the peptides of the α-helical regions had a much larger propensity of forming local transient α-helical and other conformations than those peptides of β-sheet segments.

Another question related to protein folding is concerned with the influence of the amino acid sequence (Anfinsen, 1973). The propensity of being part of a particular type of secondary structure for amino acid sequences and for specific amino acid types has for long been monitored, and a considerable material is becoming available as the number of protein structures and amino acid sequence data increase. Prediction schemes based on this information have made it possible to estimate the secondary structure of a given peptide with an accuracy of 65–70% (Garnier, 1990). This suggests that throughout the development of protein sequences certain amino acid residue types in certain sequential constellations have been selected. This is in keeping with the observations for myohemerythrin and plastocyanin which indicate that the peptide sequences selected during the development of the two proteins examined have achieved sequence constellations with a predisposition to form transient secondary structure elements similar to those in the finally folded form of the proteins. The entire amino acid sequence of a protein represents the result of a long genetic development, and it is therefore not a surprise that the productive protein folding requires the entire or at least a significant proportion of the peptide sequence of the protein present. As, however, the results of Dyson et al. (1992a,b) suggest that a given peptide of a protein structure carries a predisposition to form a particular secondary structure, it is of interest to examine whether this might be carried into the context of another amino acid sequence and the secondary structure of such an imported sequence can be adapted in the receiving structure.

At the outset of planning such an experiment it must be realized that the *de novo* design of proteins or, as suggested here, the artificial combination of structural elements which have not mutually adapted under the pressure of evolution cannot be compared to the gradual development of naturally

† The atomic coordinates of the structure have been deposited in the Brookhaven Protein Data Bank under Accession Number 1CIS.

* Author to whom correspondence should be addressed.

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

0006-2960/93/0432-11007$04.00/0

occurring proteins, in which all structural components have coevolved for millions of years. Successful transfers of loops and β-turns from one protein to another have, however, been reported (Hynes et al., 1989; Wolfson et al., 1991; Eijsink et al., 1992), but these mutations have all been made on the surface of large proteins, distant from the hydrophobic core. The information obtained from experiments of the suggested type is of interest because a systematic study of the mechanisms of protein folding will eventually have to include not only studies of proteins that were developed over millions of years to fold productively but also *de novo* designed proteins for the testing of the hypotheses of protein folding that are presently emerging.

In the present work we have employed the proteinase binding loop in the protein serine proteinase inhibitor from barley CI2[1] as a peptide segment in a protein that can be replaced with other peptide segments. CI2 is a relatively stable protein that contains a sheet of four β-strands which are sandwiched between an α-helix and the proteinase binding loop. In the native protein most of the loop is exposed to solvent; however, groups of the peptide backbone and amino acid residues of the loop form a number of contacts to side chains of the β-sheet. In the present study the loop has been replaced by nine residues of helix E in subtilisin Carlsberg (Bode et al., 1987), and NMR studies have been applied to examine the conformation of the introduced peptide. Furthermore, the effect of replacing the loop peptide on the structure and stability of the receiving globular domain of CI2 has been studied. It has been a prerequisite for this work that the ^{1}H NMR spectrum of the native protein has been assigned and structure determinations of CI2 in solution by NMR spectroscopy and in crystal by X-ray crystallography are available (Clore et al., 1987; Ludvigsen et al., 1991b; McPhalen et al., 1985; McPhalen & James, 1987).

MATERIALS AND METHODS

DNA Construction. The cloning of the hybrid protein was performed in two stages: first an expression system was constructed for M20–CI2, i.e., CI2 without the 19 N-terminal residues and containing the point mutation Leu20→Met which was introduced to initiate translation. The pT7-7/pGPI-2 expression system which is based on the bacteriophage T7 RNA polymerase was used (Tabor & Richardson, 1985). The complete CI2 gene was available from the pAC4 plasmid (Longstaff et al., 1990). The gene encoding M20-CI2 was constructed from pAC4 by PCR (McFerrin et al., 1990) and ligated into the the *NdeI–SalI* fragment of the pT7-7 vector to give the plasmid pPO1.

The expression system for the hybrid protein was constructed on the basis of pPO1. The point mutation Pro52→Glu and Phe69→Ala were introduced by PCR, and a synthetic oligonucleotide cassette encoding the subtilisin Carlsberg nonapeptide segment Lys136–Ala144 was inserted into the *NcoI–PvuI* fragment of pPO1 to give the plasmid pPO2. Both pPO1 and pPO2 were transformed into the *Escherichia coli* strain K38/pGPI-2 (Tabor & Richardson, 1985) in order to overexpress the proteins.

Expression and Purification. M20–CI2 was obtained from a 10-L culture of the strain K38/pGP1-2/pPO1. The culture was grown in TB media (Sambrook et al., 1989) containing ampicillin and kanamycin at 30 °C. At an OD_{590} of 2, protein expression was induced by increasing the temperature to 42 °C for 30 min. After the induction the culture was grown for 4 h at 37 °C.

The cells were harvested by centrifugation (3000*g* for 15 min. at 4 °C), washed with 20 mM NaAc (pH 4.4, 4 °C), and resuspended in 1/25 culture volume of the NaAc buffer. The suspension was freeze-lysed by immersion in liquid nitrogen for 10 min. followed by thawing. After centrifugation (38000*g* for 15 min at 4 °C), the supernatant was collected, 1.25 volumes 20 mM NaAc was added, and the pH was adjusted to 4.8 with HCl.

The solution was loaded onto a 1-L CM-52 cation exchange column (5 × 50 cm) preequilibrated with 20 mM NaAc, pH 4.4, and the column was washed with the same buffer. To elute the protein, a 5-L salt gradient from 0 to 0.3 M NaCl in 20 mM NaAc was applied to the column. The fractions containing M20-CI2 (about 2 L) were identified by an inhibition color assay and collected. The pool was dialyzed against water (Spectrapor dialysis tube no. 3 with a cutoff of 3.5 kDa) and lyophilized to yield at least 20 mg of protein per liter of culture. The protein was subjected to gel filtration on a Sephadex G50 column.

The fermentation of K38/pGP1-2/pPO2 was performed in the same way, only the culture was grown at 30 °C overnight in the fermenter after heat induction because of thermal instability of the hybrid protein produced (see below). After that, the cells were harvested and lysed as before.

The hybrid protein was purified by a combination of gel filtration and chromatofocusing. The cell lysate was diluted to 5/8 culture volume with 20 mM NH_4Ac (pH 4.8). After a few minutes of incubation, the lysate was loaded onto a Sephadex G-25 column (2 L, fine grade; 5 × 100 cm) to remove low molecular components. The column was pre-equilibrated with 20 mM NH_4Ac, pH 4.8. The first 400 mL after the void volume of each run was collected and lyophilized. The concentrated protein mixture was redissolved in 45 mL of 20 mM NH_4Ac (pH 4.8) and applied onto a G-50 column (700 mL, super fine grade; 5 × 35 cm) equilibrated with the same buffer. The fractions giving rise to a distinct band on an SDS–PAGE gel corresponding to a protein of 7 kDa were collected and concentrated by lyophilization from around 150 to 10 mL. This 10 mL was loaded onto another, thinner G-50 column (410 mL super fine grade, 2.6 × 78 cm; same buffer). The pooled fractions containing hybrid protein were loaded onto a positively charged PBE94 chromatofocusing gel (150 mL; 1.5 × 85 cm) equilibrated with 25 mM imidazole buffer, pH 7.0. The protein was eluted with 400 mL of polybuffer 74 (pH 5.5). After this, the hybrid protein was dialyzed against water and lyophilized. A yield of 25 mg of hybrid protein from a 10-L fermentation was obtained.

Fluorescence Measurements. Guanidinium chloride induced unfolding was followed on a Perkin-Elmer LS50B Luminiscence spectrometer with a slit width of 1 mm and an integration time of 1 s with 2-s intervals. The measurements were made on 2.5 mL of ~3 μM protein samples equilibrated in 50 mM MES buffer (pH 6.3) containing various amounts of guanidinium chloride. The Trp24 fluroescence at 356 nm upon excitation at 280 nm was meausred at 25 °C.

Nuclear Magnetic Resonance and Structure Calculations. The 2D nuclear magnetic resonance (NMR) spectra were recorded on a Bruker 600 MHz AMX spectrometer at 298 K. The 2D ^{1}H NMR experiments DQF-COSY (Müller et al., 1986; Piantini et al., 1982; Rance et al., 1983; Shaka &

[1] Abbreviations: CI2, chymotrypsin inhibitor 2; DQF-COSY, double-quantum-filtered correlated spectroscopy; M20-CI2, (Δ1–19)CI2 with the point mutation L20M; NMR, nuclear magnetic resonance; NOESY, nuclear Overhauser enhancement spectroscopy; PCR, polymerase chain reaction; rmsd, root mean square deviation; TOCSY, total correlated spectroscopy; 2D, two dimensional.

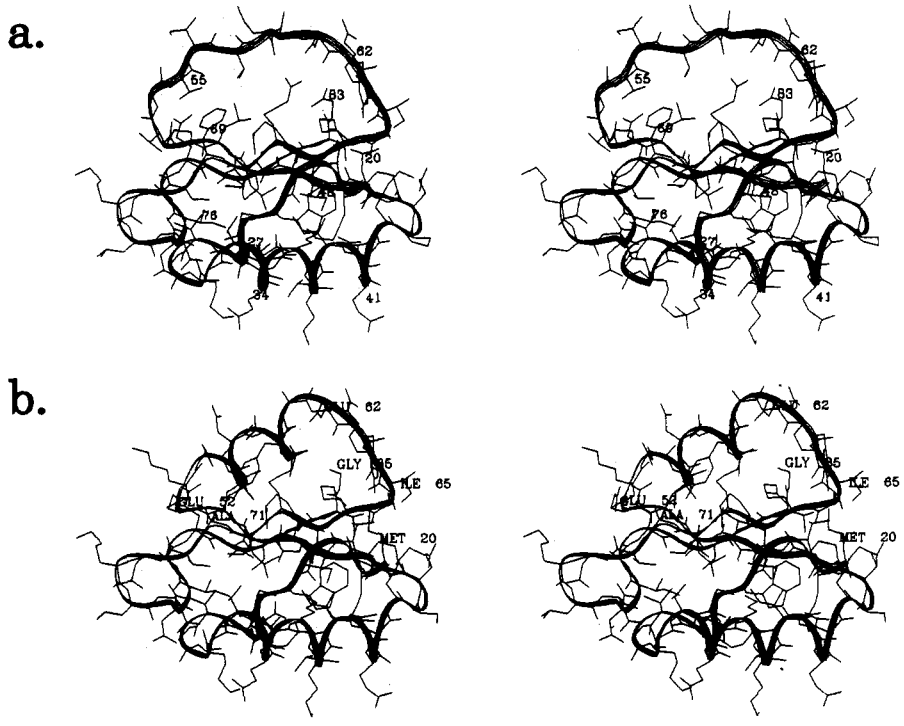

FIGURE 1: (a) Solution structure of (19–83)CI2 (Ludvigsen et al., 1991b). (b) Computer model of the hybrid between CI2 and helix E from subtilisin Carlsberg. The point mutations Pro→52Glu and Phe69→Ala are also indicated in the figure (as Glu52 and Ala71). Glu52 and Glu62 flank the subtilisin Carlsberg insert.

Freeman, 1983), NOESY (Anil-Kumar et al., 1980, 1981; Jeener et al., 1979; Macura & Ernst, 1980), and TOCSY (Braunschweiler & Ernst, 1983) with a DIPSI-2 mixing sequence (Rucker & Shaka, 1989) were applied. The NMR experiments were performed on ~5 mM protein samples, one in 10% D_2O/90% H_2O (pH 4.2) and one in 99.9% D_2O (pD 4.2; uncorrected pH-meter readings). NOESY spectra were recorded with mixing times of 80 and 200 ms. All spectra contain 2048 complex points in the t_2 dimension and 1024 points in the t_1 dimension. The spectral width is 8400 Hz. Fourier transformation, phase correction, and baseline correction of the spectra were performed with the MNMR (Pronto Software Development & Distribution, Copenhagen, Denmark) and UXNMR software (Bruker GmbH, Rheinstetten, Germany). All spectral analysis was carried out using the program PRONTO (Pronto Software Development & Distribution, Copenhagen, Denmark). Assigned NOEs were converted into distance constraints of three different categories according to their classification as either weak (1.8–5.0 Å), medium (1.8–3.3 Å), or strong (1.8–2.7 Å); 0.5 Å was added to the upper restraint limit for methyl groups (Williamson et al., 1985; Clore et al., 1985, 1986). $^3J_{H^NH^\alpha}$-coupling constants were measured by the method published by Ludvigsen et al. (1991a); the method is implemented in PRONTO. Coupling constants were grouped in the intervals 2–3 Hz [for which the ϕ angle was constrained to lie between −80° and −20° in accordance with the Karplus equation (Karplus, 1951; Pardi, 1984)], 3–4 Hz (−90° to −30°), 4–5 Hz (−95° to −35°), 5–6 Hz (−100° to −40°), 6–7 Hz (−105° to −45°; for one positive ϕ angle, 30° to 90°), 7–8 Hz (−170° to −70°), 8–9 Hz (−160° to −80°), and above 9 Hz (−150° to −90°). Positive ϕ angles can be distinguished by the presence of a very strong intraresidual H^N–H^α NOE (Ludvigsen & Poulsen, 1992a). No hydrogen-bond restraints were included in the structure calculations.

Initial structures were calculated by the program DIANA (Güntert & Wüthrich, 1991; Momany et al., 1975) and subjected to simulated annealing (Nilges et al., 1988), first for 3.75 ps at 1000 K and then for 2.8 ps where the protein was cooled to 300 K as described by Ludvigsen and Poulsen (1992b). The simulated annealing was followed by 2 ps of restrained molecular dynamics (Kirkpatrick et al., 1983) employing the full CHARMM potential (Brooks et al., 1983). Simulated annealing and molecular dynamics were performed using the X-PLOR version 2.1, which was also used for the statistical calculations (Brünger, 1990). All calculations were executed on a Silicon Graphics 320 Power station.

The computer program Insight from Biosym Technology Inc. (San Diego, CA) was used for design and display of protein structures.

The enzymes used were either from Promega, Madison, WI, or from Boehringer Mannheim GmbH, Germany. Gel materials were from Pharmacia, Uppsala, Sweden. Yeast extract from Bactotryptone was purchased from DIFCO Laboratories, Detroit, MI. PCR was performed with the GeneAmp DNA Amplification Reagent Kit from Perkin-Elmer Cetus, Norwalk, CT. Other chemicals used were from either SIGMA, Saint Louis, MO, or from Merck, Darmstadt, Germany.

RESULTS

Hybrid Design. The structure of CI2 has been determined previously by X-ray crystallography (McPhalen et al., 1985; McPhalen & James, 1987) and by NMR (Clore et al., 1987; Ludvigsen et al., 1991b); the protein consists of an α-helix overlaid by a mixed β-sheet which supports the active loop of the inhibitor (Figure 1a). The N-terminal 21 amino acid residues do not assume any ordered conformation (Kjær et al., 1987). The loop of CI2 is quite exposed to the solvent, and, seemingly, even severe mutations in this region should be possible without disturbing the overall folding of the protein core significantly. In a computer design, the seven residues Val53–Met59 (Val-Gly-Thr-Ile-Val-Thr-Met) were substituted by the nine residues Lys136–Ala144 (Lys-Gln-Ala-Val-Asp-Asn-Ala-Tyr-Ala) from subtilisin Carlsberg which in their

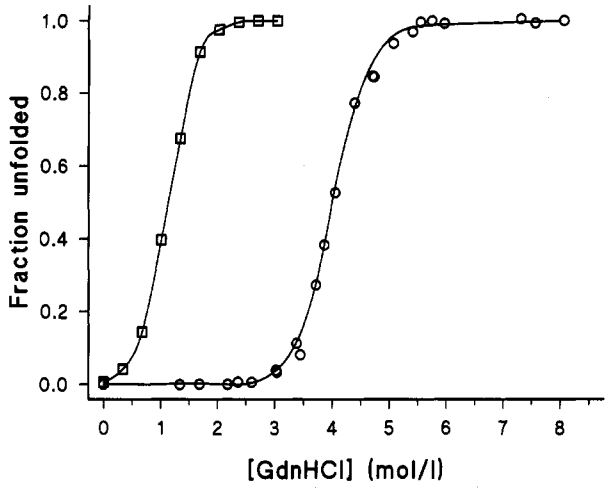

FIGURE 2: Denaturation curves for M20-C12 (○) and the hybrid (□). The relative amount of unfolded protein is plotted as a function of guanidinium chloride concentration ([GdnHCl]).

original environment are known to form an amphipathic α-helix (Bode et al., 1987; Figure 1b). By computer modeling the nonapeptide was built into a slightly modified structure of CI2. This was done only to ensure that the hybrid molecule presents no geometrical violations disfavoring the existence of such a structure. The helix, which constitutes a little more than two turns, was fitted so that the amphipathic orientation was retained. A turn structure in CI2 consisting of the $\alpha\alpha$-turn (Wilmot & Thornton, 1990) Arg62–Arg65, and a salt bridge between the Glu60 and Arg65 side chains was retained in the model to serve as a scaffold for the inserted helix. To accommodate the helix structure, two point mutations were introduced in the design: Phe69 was substituted by an alanine due to sterical considerations, and Pro52 was substituted by a glutamic acid because its backbone conformation is incompatible with α-helix formation; a glutamic acid was chosen for its ability to compensate for the positive charge induced by the helix dipole (Serrano & Fersht, 1989). In addition to these mutations, it was desirable to remove the N-terminal structureless residues as their presence does not affect the well-defined part of the protein structure and only tends to complicate the analysis of the NMR spectra; this was accomplished by the truncation of the first 19 codons of the CI2 gene and the alteration of Leu20 to a methionyl residue. The conservative Leu20→Met substitution should not influence the overall structure of the inhibitor. The model was subjected to molecular dynamics and energy minimization in order to ensure that structural stress did not prevent helix formation.

A gene encoding the amino acid sequence of the computer-modeled hybrid protein was constructed and cloned in a strain of *E. coli*, and 25 mg was purified. Also the truncated inhibitor M20–CI2 was cloned and purified; M20–CI2 was used as a reference to ensure that any changes in structure and stability of the hybrid compared to CI2 was not due to the N-terminal modifications of the protein.

Stability toward Guanidinium Chloride. The unfolding of wild-type CI2 by guanidinium chloride at pH 6.3 and 298 K has previously been studied fluorimetrically by Jackson and Fersht (1991). We measured the unfolding of M20–CI2 and of the hybrid protein under the same conditions (excitation at 280 nm and emission at 356 nm). In Figure 2 is shown the relative amount of folded protein as a function of guanidinium chloride concentration ([GdnHCl]); the much reduced stability of the hybrid is reflected in the unfolding of the protein at low

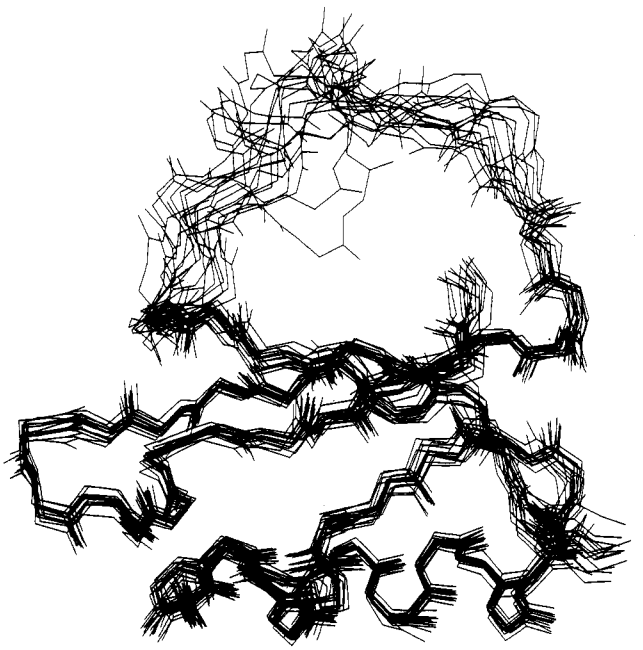

FIGURE 3: Backbone superposition of the 15 best RD structures of the hybrid calculated on the basis of the NMR spectra.

denaturant concentrations. The fluorescence as a function of guanidinium chloride concentration was fitted to the semiempiric expression

$$x_f = \frac{\exp[(m[\text{GdnHCl}] - \Delta G_{H_2O})/RT]}{1 + \exp[(m[\text{GdnHCl}] - \Delta G_{H_2O})/RT]} \quad (1)$$

where x_f is the relative amount of unfolded protein, ΔG_{H_2O} is the free energy of unfolding without denaturant, and m is a constant from the empirical relationship (Pace, 1986)

$$\Delta G_U = \Delta G_{H_2O} - m[\text{GdnHCl}] \quad (2)$$

where ΔG_U is the free energy of unfolding as a function of guanidinium chloride concentration. The fit was done with the program GraFit (Leatherbarrow, 1990). For M20–CI2 a ΔG_{H_2O} of 30.5 ± 1.0 kJ/mol and m value of 7.5 ± 0.3 kJL/mol^2 were found, which are almost the same values as those observed for wild-type CI2. For the hybrid the values were ΔG_{H_2O} = 10.9 ± 0.3 kJ/mol and m = 9.6 ± 0.3 kJL/mol^2.

Structure Calculations. DQF-COSY, NOESY, and TOCSY spectra of both proteins were recorded at 298 K and pH 4.2. Assignment of most of the M20–CI2 protons showed that the chemical shift values of the protons were essentially all unchanged compared to those of CI2 (Ludvigsen et al., 1991b), except for those of the two N-terminal residues. This proves that no significant structural changes have occurred by the Leu20→Met substitution.

Chemical shifts for every proton in the hybrid were identified, except for the hydroxyl protons, some lysine Nζ- and arginine Nη-protons, and most of the protons of Met20. From the NOE spectra more than 1400 NOEs were assigned, 893 of which contained structural information. These were used as input to the structure calculation programs together with 57 ϕ angle, 25 χ^1 angle, and 2 χ^2 angle constraints. Thirty structures were calculated; the 15 of these with lowest energy are represented in Figure 3. The average root mean square deviation (rmsd) ± standard deviation between each structure and the mean structure is 0.8 ± 0.2 Å for backbone atoms, 1.3 ± 0.2 Å for heavy atoms, and 1.5 ± 0.2 Å for all

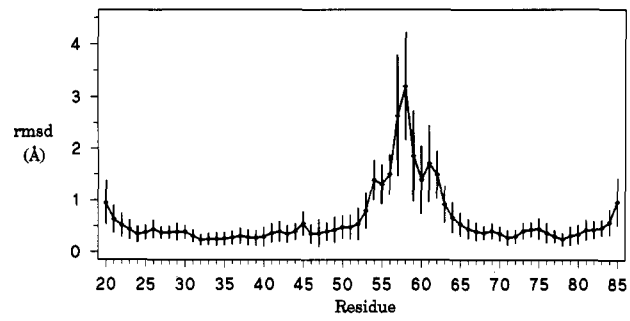

FIGURE 4: Backbone rmsd residue by residue between the 15 structures plotted ± twice the standard deviation (vertical bars).

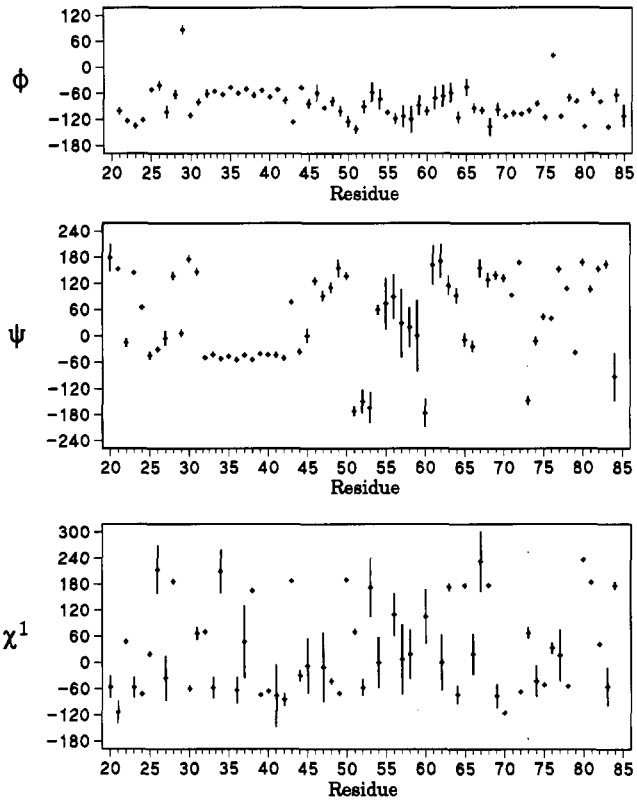

FIGURE 5: Average ϕ, ψ, and χ^1 angles for each amino acid residue between the 15 structures plotted ● twice the standard deviation (vertical bars). The location of the α-helix from S31 to K43 can be seen very distinctly. The structure of the inserted amino acid segment is poorly resolved. Note that 57 of the ϕ angles and 25 of the χ^1 angles have been constrained through the structure calculations, whereas no constraints have been imposed on the ψ angles.

atoms. The average energy is -3293.0 ± 92.6 kJ/mol. These structures show that no helix is formed by the inserted nonapeptide.

On average, 21 NOE constraints and 12 angle constraints have been violated and none more than 0.4 Å and 12°, respectively. The rmsd by residue for backbone atoms is shown in Figure 4 and the average values and rmsd's for ϕ angles, ψ angles, and χ^1 angles in Figure 5. From these plots it is clear that the best defined parts of the protein are the α-helix and β-sheet, whereas the loop from Glu52 to Glu62 is poorly defined. The well-resolved secondary structure elements are depicted in Figure 6. These elements have changed very little from CI2 to the hybrid; in the β-sheet two hydrogen bonds are missing in the hybrid compared to CI2 (Arg67O–Gly85H^N and Lys53H^N–Val72O), and a new hydrogen bond has been gained (Arg69O–Arg83H^N). This means that Lys53 and Gly85 are not part of the β-sheet in the hybrid as Val53 and Gly83 are in CI2.

The calculated solution structures of the hybrid reveal that no helix is formed by the amino acids in the sequence which has been inserted into the CI2 molecule. The overall conformation of the inhibitor, however, turns out to be remarkably conserved (Figures 7 and 8), even in the β-sheet underlying the loop, although the stability of the protein has decreased substantially. This destabilization is probably due to the reduced number of interactions in the loop and between the loop and the β-sheet reflected in the very low number of NOEs; the removal of the Phe69 phenyl ring, which in CI2 constitutes the centre of a hydrophobic compartment, may play an important role in the destabilization (Figure 9).

DISCUSSION

The insertion and replacements of amino acid residues in the sequence of CI2 represent a significant change of the primary structure of CI2, and this raises three questions regarding the structure of this constructed polypeptide. Does this hybrid peptide fold into a globular protein? Does the remaining CI2 sequence fold as it does in the native CI2? And does the helical sequence replacing the original sequence form an α-helix in the new context? The present work was set up to address these three questions, and it has provided an answer to each of these. The hybrid protein does fold into a globular protein where the remaining sequence of CI2 folds as in the native protein, and the α-helical sequence does not form an α-helix in the new context. In addition, it has been observed that the new derivative is less stable than the native molecule.

The fold of the hybrid peptide chain forming a structure as seen here where the secondary and tertiary structure are essentially identical with that of the native structure has implications for the understanding of the folding mechanism of the protein. The result suggests that both in the native and in the hybrid sequence neither the loop nor the inserted peptide is actively involved in the folding process. Rather, these results suggest that a peptide segment has to be present as a spacer sequence between the two parallel strands of the β-sheet to ensure that the native part of the two molecules can achieve the native folding.

In CI2 the peptide loop includes the peptide sequence that binds to the active site in serine proteinases. This site is structurally stabilized by a number of interactions in particular salt bridges and hydrogen bonds between the side chains of the globular part and the peptide backbone groups of the loop. Nevertheless, the loop peptide seems to play no direct role for the folding of the remaining sequence. Between the loop and the rest of the structure, the major interactions as seen in the X-ray structure are between the guanidino group of Arg65 and the carboxylate and carbonyl of Glu60 including also the contacts to the C-terminal carboxylate group of Gly83. These contacts are difficult to observe by NMR; however, the C-terminal Gly85^{hybrid} unlike the native form seems not to be involved in the hydrogen bond pattern of the antiparallel β-sheet. This indicates that several of the interactions between the loop and the globular structure do not form, and this is in spite of the C-terminal sequence from Glu62^{hybrid} to Gly85^{hybrid} being the same as in native CI2 with the exception of Ala71^{hybrid} replacing Phe69 in the native sequence. These observations may also explain the reduced stability of the hybrid. The lack of interactions between the loop and the globular structure and in addition the missing apolar interactions of the phenylalanine residue present in CI2 are both effects that are expected to result in a destabilization of the protein. However, these interactions seem to have no major influence on the folding process of the protein.

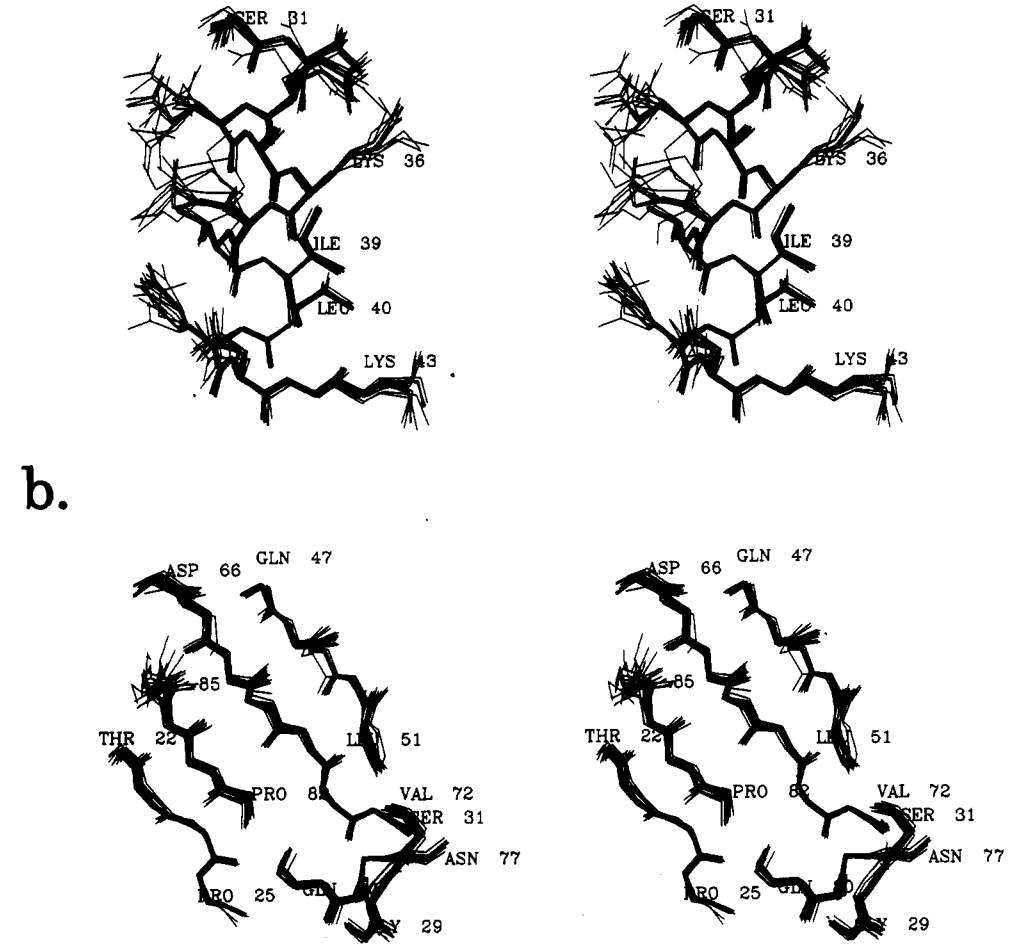

FIGURE 6: (a) Stereoview of the (S31–K43) α-helix and (b) the backbone of the β-sheet of the hybrid (15 best RD structures). The well-resolved side chains to the right of the helix are buried in the protein, whereas the side chains to the left are exposed to the solvent.

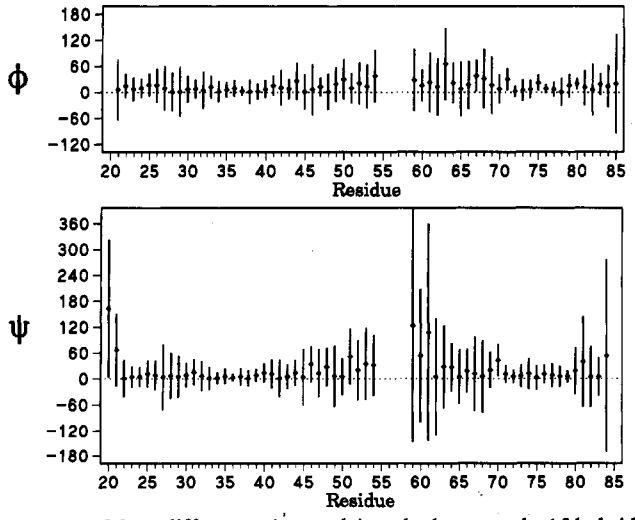

FIGURE 7: Mean differences in ϕ and ψ angles between the 15 hybrid structures and 20 structures of CI2 (Ludvigsen et al., 1991b). The vertical bars indicate twice the sum of the standard deviation of the hybrid structures and the standard deviation of the CI2 structures. The residues are numbered according to the hybrid so that a number of two must be subtracted from residues after 58 to give the position of the corresponding residue in CI2. Ala55–Asn58 of the hybrid and Thr55–Ile56 of CI2 have been omitted in the comparison. The residues in positions 20, 52–54, 59–61, and 71 in the figure are not identical in the two proteins.

In subtilisin Carlsberg, the nonapeptide sequence used to replace the active site loop in CI2 is known to form an amphipathic α-helix. After translocation to CI2 this peptide segment was seen no longer to form an α-helix. All the NMR parameters measured for the peptide suggest that the peptide is in a random coil-like conformation. The chemical shifts measured for the ^{1}H resonances are either close or identical to reported random coil values. The $^3J_{H^NH^\alpha}$ measured for each of the residues except Lys54 are 6 or 7 Hz, suggesting that the ϕ angle undergoes conformational averaging. This is further supported by the observation of the intraresidue and sequential NOEs between, respectively, H^N and H^α in the loop region.

The studies of Dyson et al. (1992a,b) showed that the individual peptides of helical parts of myohemerythrin had an intrinsic predisposition to form local transient α-helical conformations of the peptide backbone. The present experiment represents an alternative method to address the issues of peptide folding propensities. By incorporating a peptide into the scaffold of a globular protein, it becomes possible to examine the intrinsic properties of the peptide to form secondary structures while the N- and C-terminal are fixed to the rest of the structure, a condition that is more similar to the real situation of protein folding than studies of the free peptides. In order to characterize the fold of the inserted peptide, we have examined by ^{1}H NMR spectroscopy the conformation of the structure using measurements of chemical shifts, coupling constants, and NOEs. Each of these measures provide a clear indication that the inserted peptide is not in a well-defined conformation in the CI2 loop context. This

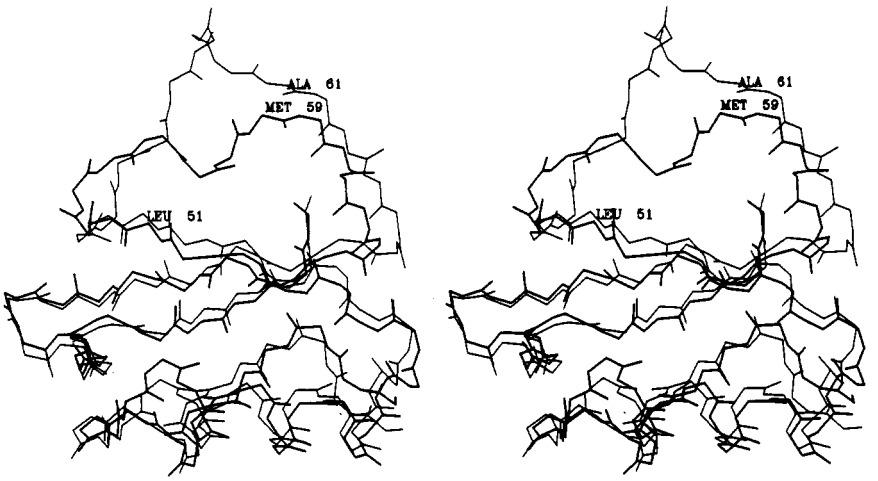

FIGURE 8: Backbone superposition in stereoview of a representative RD structure of the hybrid and of CI2 (bold). The superposition is based on the first 50 and last 20 residues of the proteins. The figure shows that the backbone conformation has been retained up until Leu51 on one side of the loop and apparently all the way to Ala61 in the hybrid (Met59 in CI2) from the C-terminal side of the loop.

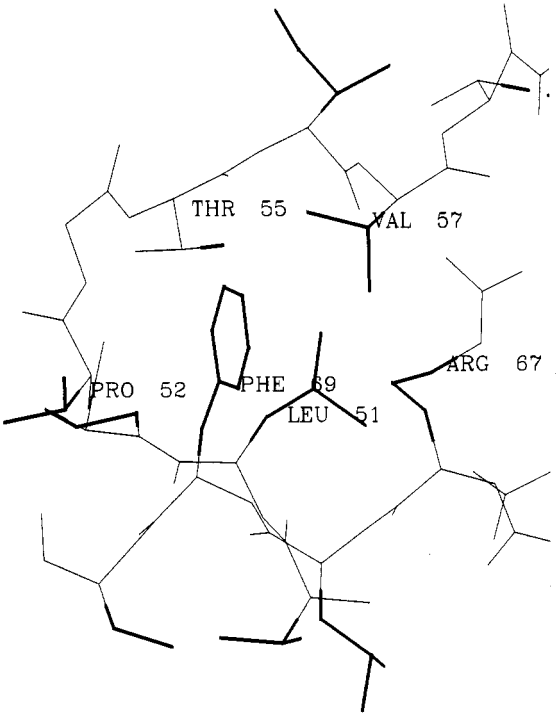

FIGURE 9: Hydrophobic pocket around Phe69 in CI2 (Ludvigsen et al., 1991b). Aliphatic and aromatic parts of side chains in the loop are represented with bold lines. Phe69 has been substituted with Ala71 in the hybrid protein, and the removal of this hydrophobic patch is probably partly responsible for the decrease in stability.

implies that the peptide in the sequential context of CI2 does not have an intrinsic preference for formation of the α-helix that it forms in subtilisin, suggesting that the helix formation is context dependent and relying clearly on other parts of the amino acid sequence in subtilisin to stabilize the α-helix formation.

The present experiment has shown that a peptide that folds as an α-helix in one structural context need not do so in a new structural context. It has also shown that a large unstructured region of the amino acid sequence in a protein can be replaced by a new sequence without severely affecting the overall structure, suggesting that the folding of a protein depends mostly on the formation and organization of well-defined secondary structure elements. Finally, the present work has outlined a test-bed system by which the predisposition of a peptide to assume a certain structure can be examined.

ACKNOWLEDGMENT

We thank Miss Pia Mikkelsen for skilled technical assistance. Dr. Robin J. Leatherbarrow is thanked for supplying us with the pAC4 plasmid and Dr. Ib Svendsen for determining the N-terminal sequence of M20–CI2 and the hybrid.

SUPPLEMENTARY MATERIAL AVAILABLE

Two figures showing the Ramachandran plot for the 15 hybrid protein structures and, schematically, the changes in the β-sheet from CI2 to the hybrid (2 pages). Ordering information is given on any current masthead page.

REFERENCES

Anfinsen, C. B. (1973) *Science 181*, 223–230.
Anil-Kumar, Ernst, R. R., & Wüthrich, K. (1980) *Biochem. Biophys. Res. Comm. 95*, 1–6.
Anil-Kumar, Wagner, G., Ernst, R. R., & Wüthrich, K. (1981) *J. Am. Chem. Soc. 103*, 3654–3658.
Bode, W., Papamokos, E., & Musil, D. (1987) *Eur. J. Biochem. 166*, 673–692.
Braunschweiler, L., & Ernst, R. R. (1983) *J. Magn. Reson. 53*, 521–528.
Brooks, B. R., Bruccoleri, R., Olafson, B., States, D., Swaninathan, S., & Karplus, M. (1983) *J. Comput. Chem. 4*, 187–217.
Brünger, A. T. (1990) X-PLOR, version 2.1. A system for crystallography and NMR. Yale University, New Haven, CT.
Bycroft, M., Matouschek, A., Kellis, J. T., Serrano, L., & Fersht, A. R. (1990) *Nature 346*, 488–490.
Clore, G. M., Gronenborn, A. M., Brünger, A. T., & Karplus, M. (1985) *J. Mol. Biol. 186*, 435–455.
Clore, G. M., Brünger, A. T., Karplus, M., & Gronenborn, A. M. (1986) *J. Mol. Biol. 191*, 523–551.
Clore, G. M., Gronenborn, A. M., Kjær, M., & Poulsen, F. M. (1987) *Protein Eng. 1*, 305–311.
Dyson, H. J., Merutka, G., Waltho, J. P., Lerner, R. A., & Wright, P. (1992a) *J. Mol. Biol. 226*, 795–817.
Dyson, H. J., Sayre, J. R., Merutka, G., Shin, H.-C., Lerner, R. A., & Wright, P. (1992b) *J. Mol. Biol. 226*, 819–835.
Eijsink, V. G. H., Vriend, G., van den Burg, B., van der Zee, J. R., Veltman, O. E., Stulp, B. K., & Venema, G. (1992) *Protein Eng. 5*, 157–163.
Garnier, J. (1990) *Biochimie 92*, 513–524.
Güntert, P., & Wüthrich, K. (1991) *J. Mol. NMR 1*, 447–456.
Heinz, D. W., Hyberts, S. G., Peng, J. W., Priestle, J. P., Wagner, G., & Grütter, M. G. (1992) *Biochemistry 31*, 8755–8766.
Hynes, T. R., Kautz, R. A., Goodman, M. A., Gill, J. F., & Fox, R. O. (1989) *Nature 339*, 73–76.

Jackson, S. E., & Fersht, A. R. (1991) *Biochemistry 30*, 10428–10435.
Jeener, J., Meier, B. H., Backmann, P., & Ernst, R. R. (1979) *J. Chem. Phys. 71*, 4546–4553.
Karplus, M. (1951) *J. Phys. Chem. 30* 11–15.
Kirkpatrick, S., Gelatt, C. D., Jr., & Vecchi, M. P. (1983) *Science 220*, 671–680.
Kjær, M., Ludvigsen, S. J., Sørensen, O. W., Denys, L. A., Kindtler, J., & Poulsen, F. M. (1987) *Carlsberg Res. Commun. 52*, 327–354.
Leatherbarrow, R. J. (1990) *GraFit*, Version 2.0, Erithacus Software Ltd., Staines, U.K.
Longstaff, C., Campbell, A. F., & Fersht, A. R. (1990) *Biochemistry 29*, 7339–7349.
Ludvigsen, S. J., Andersen, K. V., & Poulsen, F. M. (1991a) *J. Mol. Biol. 217*, 731–736.
Ludvigsen, S. J., Shen, H., Kjær, M., Madsen, J. C., & Poulsen, F. M. (1991b) *J. Mol. Biol. 222*, 621–635.
Ludvigsen, S. J., & Poulsen, F. M. (1992a) *J. Biomol. NMR 2*, 227–233.
Ludvigsen, S. J., & Poulsen, F. M. (1992b) *Biochemistry 31*, 8783–8789.
Macura, S., & Ernst, R. R. (1980) *Mol. Phys. 41*, 95–117.
McFerrin, K. D., Terranova, P., Schreiber, S. L., & Verdine, G. L. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 1937–1941.
McPhalen, C. A., Svendsen, I., Jonassen, I., & James, M. N. G. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 7242–7246.
McPhalen, C. A., & James, M. N. G. (1987) *Biochemistry 26*, 261–269.
Momany, F. A., McGuire, R. F., Burgess, A. W., & Scheraga, H. A. (1975) *J. Phys. Chem. 79*, 2361–2381.
Müller, N., Ernst, R. R., & Wüthrich, K. (1986) *J. Am. Chem. Soc. 108*, 6482–6492.
Nilges, M., Clore, G. M., & Gronenborn, A. M. (1988) *FEBS Lett. 229*, 317–324.
Pace, C. N. (1986) *Methods Enzymol. 131*, 266–279.
Pardi, A., Billeter, M., & Wüthrich, K. (1984) *J. Mol. Biol. 180*, 741–751.
Piantini, U., Sørensen, O. W., & Ernst, R. R. (1982) *J. Am. Chem. Soc. 104*, 6800–6801.
Radford, S. E., Dobson, C. M., & Evans, P. A. (1992) *Nature 358*, 302–307.
Rance, M., Sørensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 177*, 479–485.
Roder, H., Elöve, G. A., & Englander, S. W. (1988) *Nature 335*, 700–704.
Rucker, S. P., & Shaka, A. J. (1989) *Mol. Phys. 68*, 509–517.
Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989) *Molecular Cloning*, Cold Spring Harbor Laboratory Press, New York.
Serrano, L., & Fersht, A. R. (1989) *Nature 342*, 296–299.
Shaka, A. J., & Freeman, R. (1983) *J. Magn. Reson. 51*, 169–173.
Tabor, S., & Richardson, C. C. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 1074–1078.
Udgaonkar, J. B., & Baldwin, R. L. (1988) *Nature 335*, 700–704.
Wilmot, C. M., & Thornton, J. M. (1990) *Protein Eng. 3*, 479–493.
Williamson, M. P., Havel, T. F., & Wüthrich, K. (1985) *J. Mol. Biol. 182*, 295–315.
Wolfson, A. J., Kanaoka, M., Lau, F. T. K., & Ringe, D. (1991) *Protein Eng. 4*, 313–317.