A test of the relationship between sequence and structure in proteins: Excision of the heme binding site in apocytochrome *b*₅

AILEEN J. CONSTANS, MICHAEL R. MAYER, STEVEN F. SUKITS,
AND JULIETTE T.J. LECOMTE

Department of Chemistry and the Center for Biomolecular Structure and Function, The Pennsylvania State University,
University Park, Pennsylvania, Pennsylvania 16802

(RECEIVED March 23, 1998; ACCEPTED May 22, 1998)

Abstract

The water-soluble domain of rat hepatic holocytochrome *b*₅ is an $\alpha\beta$ protein containing elements of secondary structure in the sequence $\beta 1-\alpha 1-\beta 4-\beta 3-\alpha 2-\alpha 3-\beta 5-\alpha 4-\alpha 5-\beta 2-\alpha 6$. The heme group is enclosed by four helices, $\alpha 2$, $\alpha 3$, $\alpha 4$, and $\alpha 5$. To test the hypothesis that a small *b* hemoprotein can be constructed in two parts, one forming the heme site, the other an organizing scaffold, a protein fragment corresponding to $\beta 1-\alpha 1-\beta 4-\beta 3-\Lambda-\beta 2-\alpha 6$ was prepared, where Λ is a seven-residue linker bypassing the heme binding site. The fragment ("abridged *b*₅") was found to contain α and β secondary structure by circular dichroism spectroscopy and tertiary structure by Trp fluorescence emission spectroscopy. NMR data revealed a species with spectral properties similar to those of the full-length apoprotein. This folded form is in slow equilibrium on the chemical shift time scale with other less folded species. Thermal denaturation, as monitored by circular dichroism, absorption, and fluorescence spectroscopy, as well as size-exclusion chromatography-fast protein liquid chromatography (SEC-FPLC), confirmed the coexistence of at least two distinct conformational ensembles. It was concluded that the protein fragment is capable of adopting a specific fold likely related to that of cytochrome *b*₅, but does not achieve high thermodynamic stability and cooperativity. Abridged *b*₅ demonstrates that the spliced sequence contains the information necessary to fold the protein. It suggests that the dominating influence to restrict the conformational space searched by the chain is structural propensities at a local level rather than internal packing. The sequence also holds the properties necessary to generate a barrier to unfolding.

Keywords: apocytochrome *b*₅; heme binding site; holocytochrome *b*₅; partially folded protein; protein folding; protein stability; structural specificity

The native fold of water-soluble globular proteins is inscribed in their amino acid sequence (Anfinsen, 1973). Although the stability of this fold is modest at best (Tanford, 1968, 1970; Privalov, 1979), it is nevertheless sufficient to insure that the folded form predominates under physiological conditions. Much experimental effort has been devoted to the study of the folding reaction during which the disordered polypeptide chain is transformed into its functional three-dimensional structure (Matthews, 1993; Ptitsyn, 1995; Miranker & Dobson, 1996; Shortle, 1996; Roder & Colón, 1997). The rules that underlie this transformation are of considerable fundamental and practical interest. Experiments fall in two categories: equilibrium studies, which inspect the reversible transformation of the native structure into less structured conformations, and relaxation kinetic studies, which follow in time the formation or disruption of structure when the solution conditions

are suddenly perturbed. It is often adequate to interpret the results of these experiments in terms of discrete thermodynamic states. The simplest proteins require only two, a native ensemble (*N*) and an unfolded ensemble (*U*); the difference in standard free energy between them (stability, $\Delta G^\circ(N \rightarrow U) = -RT \ln[U]_{eq}/[N]_{eq}$) ranges below 100 kJ/mol or, on average, a fraction of a kJ/mol per residue (Robertson & Murphy, 1997).

Over the last couple of decades, amino acid replacement investigations have revealed a consistent aspect of the folding reaction: the stability of the native fold is sensitive to sequence alterations, i.e., a few mutations can affect significantly the value of $\Delta G^\circ(N \rightarrow U)$, whereas the overall native fold is quite resilient and requires many replacements before losing its identity. Accordingly, Lattman and Rose (1993) have noted that within a two-state description, a protein for which $\Delta G^\circ(N \rightarrow U)$ is diminished still carries the specificity for a given native fold and none other; that is, the tendency of an amino acid sequence to adopt a particular fold is independent of its stability. This is illustrated not only in reversible denaturation experiments where

Reprint request to: Juliette T.J. Lecomte, Chemistry Department, The Pennsylvania State University, 152 Davey Laboratory, University Park, Pennsylvania 16802; e-mail: jtl1@psu.edu.

temperature or an additive is used to change $\Delta G^\circ(N \rightarrow U)$, but also under native conditions by proteins such as $\Delta 131\Delta$ staphylococcal nuclease (a mutant obtained by deletion of nine residues from both terminal regions; Alexandrescu et al., 1994; Zhang et al., 1997) and the isolated subtilisin BPN' prodomain (Bryan et al., 1995), which exhibit extensive disorganization and refold globally upon addition of the appropriate substrate. Thus, issues of stability can be separated to a large extent from issues of structural specificity.

A characteristic feature of native proteins is the efficient packing of their interior with nonpolar residues (Richards, 1977). Given a set of hydrophobic side chains dispersed in the sequence, good packing into a core implies restrictions on the backbone geometry; packing interactions have therefore been proposed to play a crucial role in determining the native structure (Kono & Doi, 1994; Dahiyat & Mayo, 1997; Lazar et al., 1997). An alternative view rests on the observation that given a fold, filling the interior of the protein with nonpolar side chain of complementary shape is a problem with multiple solutions (Behe et al., 1991; Desjarlais & Handel, 1995; Sauer, 1996). This de-emphasizes the importance of packing in the folding code and favors a stereochemical control dominated by local properties (Lattman & Rose, 1993). The nature of the message and its organization in the sequence become distinct questions to address in the protein folding problem.

To assess the contribution of internal packing in structural specificity, and ultimately to understand the relationship between sequence and structure, one approach is to explore the response of a marginally stable protein to primary structure modifications. These species are poised to undergo measurable excursions toward either stabilization or unfolding, and also toward either a contraction or an expansion of the sampled conformational space. The effects of sequence alteration should be readily detectable and report on the formulation of the folding message in the primary structure. The water-soluble fragment of cytochrome b_5 (cyt b_5) has the desired potential for partial unfolding and low stability, and is used in this study.

Cyt b_5 is a one-domain protein of unique α - β fold (Mathews et al., 1979). It contains five β -strands organized into a barrel completed by one helix (Figs. 1A, 2A). The heme is located on one side of the barrel and protected from solvent by a stretch of about 40 residues encompassing elements of secondary structure $\alpha 2$, $\alpha 3$, $\beta 5$, $\alpha 4$, and $\alpha 5$. The holoprotein has normal stability for a protein of this size and undergoes two-state thermal unfolding (Table 1; Pfeil, 1993). The low-resolution structure of the rat protein in the absence of the heme group (apocyt b_5) has been obtained by NMR methods (Falzone et al., 1996). The β topology is similar in the apo- and holo-protein forms but the absence of the heme loosens the $\beta 5$ strand from the sheet and alters the helices enclosing the heme (Figs. 1B, 2B) (Moore & Lecomte, 1993; Falzone et al., 1996). The structured segments define a part of the molecule that will eventually make few contacts with the prosthetic group. These regions organize a hydrophobic core and are responsible for the cooperative thermal unfolding exhibited by the apoprotein (Table 1; Pfeil, 1993).

The two-state unfolding and the observation that the ordered part folds autonomously (Moore & Lecomte, 1993) led to the proposal that apocyt b_5 is made of two "modules," one for stability and pre-organization of the second, destined to bind the heme (Falzone et al., 1996). This hypothesis can be tested in various ways. A first level of inquiry concerns the ability of the organizing module to adopt its fold independently of the binding module. In this work, data are reported on an abridged form of the rat apo-

Fig. 1. The topology of the β sheet in (A) holocyt b_5 (Mathews et al., 1979) and (B) apocyt b_5 (Falzone et al., 1996). The heme binding site is formed by the stretch joining $\beta 3$ to $\beta 2$, which contains in succession $\alpha 2$, $\alpha 3$, $\beta 5$, $\alpha 4$, and $\alpha 5$. In the apoprotein, only the N-terminal turn of $\alpha 2$ and the C-terminal turn of $\alpha 5$ have a long lifetime and are well populated. The nomenclature is that generally used for cyt b_5 , where the strands are numbered according to the topology rather than sequentially. Scheme (C) depicts the intended topology of the construct, where the heme binding site fragment is bypassed by a short tether. The folding of this structure requires the insertion of the C-terminal $\beta 2$ strand between $\beta 1$ and $\beta 3$ of the same molecule.

cytochrome ("abridged b_5 ," Figs. 1C, 2C), designed to exclude the elements of structure associated only with the heme group and to retain the hydrophobic core of apocyt b_5 . To form abridged b_5 , 39 residues were excised from apocyt b_5 and replaced by a seven-residue tether.

The reduced stability of the full-length apoprotein and the folding specificity it manifests under native conditions make this molecule a starting point to analyze the interplay between the forces affecting $\Delta G^\circ(N \rightarrow U)$ and the propensity for a particular topology. The immediate question addressed with abridged b_5 is whether this summary protein assumes a structure reminiscent of cyt b_5 . If a specific fold is detected in the abridged version, an ancillary question concerns its stability and the characteristics of the $N \rightarrow U$ transition. As will be shown, abridged b_5 can serve to probe the organization and robustness of the structural message encoded in the amino acid sequence of cytochrome b_5 .

Fig. 2. **A:** The ribbon diagram of bovine liver holocyt *b*₅ (PDB structure 1cyo) in a view emphasizing the imperfect β -barrel. **B:** A representative ribbon diagram of rat liver apocyt *b*₅ (PDB structure 1iet) in a similar orientation. The numbers 34 and 74 point to the cleavage sites. T4 marks the turn used as part of the linker Δ . **A** and **B** were created with MOLSCRIPT (Kraulis, 1991). **C:** Primary structure of the soluble domain of rat liver cytochrome *b*₅ (Cyt *b*₅) and the derived abridged protein (Abr *b*₅). Numbering starts at residue -4 for consistency with the bovine sequence. Residue number 1 is identified with a star. The elements of secondary structure in the X-ray structure of the holoprotein are labeled and marked in bold face. A total of 39 residues were excised from the sequence and replaced by seven residues. The abridged construct ends with a Met (for later cleavage by CNBr) followed by six histidines as a tag for binding onto a nickel chelating column.

Results

Design of the abridged protein

Cyt *b*₅ contains several elements of secondary structure from N- to C-terminus, namely β 1, α 1, β 4, β 3, α 2, α 3, β 5, α 4, α 5, β 2, and

α 6 (Fig. 2A,C). The heme binding segment is located between β 3 and β 2. The abridged *b*₅ construct was designed by inspecting the structure of the parent full-length apoprotein. First the segment of the protein to be excised was defined by cleavage points no further apart than 10 Å in the average structure (Fig. 2B). They were

Table 1. Parameters for the thermal denaturation of abridged *b*₅ and parent proteins

	T_m^a (°C)	ΔH^{ob} (kJ mol ⁻¹)	ΔC_p^o (kJ mol ⁻¹ K ⁻¹)
Cyt <i>b</i> ₅ (rat) ^c	69.9 ± 0.5	320 ± 30	7 ± 1
Cyt <i>b</i> ₅ (rabbit) ^d	69.5	329.0	6.0
Apocyt <i>b</i> ₅ (rat) ^c	46 ± 1	136 ± 12	4 ± 1
Apocyt <i>b</i> ₅ (rabbit) ^d	48.5 ± 0.5	149 ± 7	4.2 ± 0.6
His80Ala/Pro81Ala cyt <i>b</i> ₅ ^c	62.7 ± 0.5	320 ± 30	7 ± 1
His80Ala/Pro81Ala apocyt <i>b</i> ₅ ^c	27 ± 2	90 ± 12	4 ± 1 ^e
Abridged <i>b</i> ₅	39 ± 3	66 ± 11	3 ± 1

^aMidpoint of the transition, under the two-state assumption.

^bEvaluated at the midpoint of the transition.

^cValues for the rat protein (wild-type or mutants), obtained by CD spectroscopy at pH 7.5.

^dValues reported by Pfeil (1993), obtained by differential scanning calorimetry on the rabbit protein at pH 7.4. The rat protein differ from the rabbit protein by four-residue extensions at the N- and C-termini and six replacements in the common part.

^eObtained by fixing the slope of the baselines (A_n and A_u in Equations 2 and 3) to the wild-type values. The heat capacity difference between the folded and unfolded states cannot be reliably measured by this method but is manifested in the appearance of cold denaturation.

selected outside of the binding loop, purposefully avoiding most of the residual structure in $\alpha 2$ and $\alpha 5$. The partial truncation of $\alpha 2$ and $\alpha 5$ is expected to reduce the stability of the fold, but conforms with the aim of this particular test. The excised portion spans residues 35 to 73. It is shorter than the 21-to-78 region defined as containing the heme in the holoprotein X-ray structure (Mathews, 1985). On the other hand, it contains all the residues participating in "core 2" of the protein. Core 2 is a structural hydrophobic set, made of Leu9, Ile12, Trp22, Ile24, Val29, Ile76, Leu79, and Ile87. Among these, only Trp22 is strictly invariant across known species (Mathews et al., 1979).

The second step was to choose a linker to bridge the cleavage points. The T4 turn of the holoprotein joins $\alpha 2$ and $\alpha 3$ and contains one of the heme coordinating histidines (His39). T4 is highly conserved and has been identified as part of a cyt b_5 signature (Bairoch, 1991; Attwood et al., 1994). With its adjacent residues, this turn can adopt a conformation with end-to-end span matching the gap between the extremities of the excised fragment. The sequence of the abridged protein (Fig. 2C) was completed with a string of six histidines at the C-terminal end to ease purification. A methionine was inserted prior to the histidine tag for CNBr cleavage, if desired.

In order to anticipate the structural properties of the abridged protein, a computer model was generated by subjecting the polypeptide to a crude distance geometry and simulated annealing protocol (Brünger, 1992) under the experimentally derived apoprotein restraints (Falzone et al., 1996), excluding those for residues 33 to 74. The structures so obtained showed no dramatic violation of structural principles: the hydrophilic linker can be accommodated and no additional hydrophobic residue is exposed upon removal of the binding site from the apoprotein. Not surprisingly the quality of the hypothetical structure so contrived, if measured against known structures in the database (Vriend & Sander, 1993; Hooft et al., 1996), is poor; it suffers from mediocre packing and unusual loop conformations. Database searches performed to recognize the three-dimensional shape of the module in other proteins of known structure (Holm & Sander, 1993) so far have yielded no indication that relatives of abridged b_5 served in contexts other than that of cyt b_5 . The abridged protein was prepared and purified. DNA sequencing of the gene and mass spectrometry of its product confirmed that the intended protein of 8,530 Da had been obtained.

Characterization of abridged b_5 by optical methods

If the abridged protein folds into a conformation related to that of the full-length apoprotein, it should contain nearly 20% of each α and β secondary structure (Fig. 2C). The circular dichroism spectrum of abridged b_5 at pH 6.8 and 15°C is shown in Figure 3. The abridged protein spectrum exhibits weak signatures of α and β structure. By raising the temperature, these nonrandom features are lost, and a thermal denaturation curve is obtained. The reaction is reversible and was monitored at two wavelengths. The data were globally fitted with a simple two-state denaturation model to obtain the parameters listed in Table 1. Conversion to an apparent fraction of unfolded protein vs. temperature (Fig. 4) allows for comparison with the denaturation behavior of the full-length wild-type apoprotein, and a double-mutant thereof, His80Ala/Pro81Ala.

In the holoprotein and the apoprotein, the N δ atom of His80 acts as a hydrogen bond acceptor for the backbone NH of Asp82 (Mathews et al., 1979; Lecomte & Moore, 1991). This interaction

Fig. 3. Far UV circular dichroism spectrum of abridged b_5 . Molar residual ellipticity (MRE) is plotted vs. wavelength. The protein concentration was 30 μ M, and the pH maintained with 20 mM phosphate buffer at pH 6.8. The temperature was 15°C. There exists a small but detectable extent of secondary structure under these conditions. The inset shows the thermal denaturation of a 35 μ M sample monitored at 215 nm.

caps the N-terminal end of $\alpha 6$ (Presta & Rose, 1988). The double-mutant protein is one of several mutants constructed to study the consequences of N-cap removal. The thermal denaturation curves of Figure 4 indicate that the stability of the full-length apoprotein can be significantly diminished by two amino acid replacements perturbing $\alpha 6$ and disrupting the interface between helix $\alpha 6$ and strand $\beta 1$. Within the two-state model, there is a significant lowering of the enthalpy of denaturation compared to the wild-type apoprotein. Apocyt b_5 , whose fold is of marginal stability, is strongly affected by these modifications to its structured portions. Nevertheless, the transition of the double-mutant appears two-state, and NMR data (not shown) suggest no change in the overall specificity of the fold.

The thermal transition of the abridged protein takes place over a wide range of temperatures. If the two-state representation applies, the shallow slope is a consequence of the yet lower enthalpy of denaturation, amounting to about half the value observed for the full-length wild-type apoprotein (Table 1). The ΔC_p° value is small, close to that of the full-length protein and has an observable effect at low temperatures. Given the unaffected ΔC_p° the reduction in enthalpy change indicates that a number of interactions have been severed. Also according to this interpretation of the data, the extent of secondary structure reaches a maximum at 15°C and this temperature was chosen for further experiments.

Fluorescence spectroscopy was used to probe the environment of the only tryptophan in the abridged protein. In the holo- and apoproteins, Trp22, on strand $\beta 4$, points into the barrel and is in contact with several residues (Fig. 2B). These contacts can be disrupted by addition of urea, and the results are an increase in Trp fluorescence intensity and a bathochromic shift (Huntley & Strittmatter, 1972; Moore & Lecomte, 1993). Emission spectra in the

Fig. 4. Thermal denaturation of abridged *b*₅ and parent proteins as monitored by the change in circular dichroism signal. The apparent fraction of unfolded protein evaluated with a two-state denaturation model (F_u) is plotted vs. temperature. Abridged *b*₅ followed at 215 nm ($\circ$, $\square$) and 222 nm (Δ). The protein concentration ranged from 20 to 30 μ M, with pH maintained at 6.8 with 20 mM phosphate buffer. The line results from global fitting to the three data sets. The transition temperature is 39°C. For comparison, the curve obtained for the full-length wild-type apoprotein is shown (*, $T_m = 46^\circ\text{C}$) as well as that of a double mutant in helix $\alpha 6$ (His80Ala/Pro81Ala cyt *b*₅) ($\times$, $T_m = 27^\circ\text{C}$). The full-length protein data were collected at pH 7.5, monitoring the signal at 220 nm. The abridged protein has a wide transition zone, indicating a low enthalpy of denaturation within the two-state interpretation of the data.

presence and absence of urea were collected for the abridged protein; they are shown in Figure 5. A marked increase in fluorescence intensity is measured when urea is added in denaturing concentration. More importantly, the fluorescence enhancement is accompanied by a red shift, from 355 to 362 nm. The latter data suggest that Trp22 is not fully exposed to solvent under native conditions. It should be noted that the maximum shift induced by the protein structure is smaller in the abridged protein (7 nm) than in the apoprotein (24 nm); thus, the Trp environments in abridged *b*₅ and apocyt *b*₅ are not identical over the whole population of protein molecules and some degree of local rearrangement takes place. The coexistence of multiple conformations at 15°C prevents a detailed analysis of the results; nevertheless, this experiment provides evidence that tertiary structure possibly related to core 2 is sampled in the absence of the heme binding loop.

When the fluorescence intensity is followed as a function of urea concentration in a titration from 0 to 6 M, the signal change is gradual over the whole range of concentration (not shown) and does not resemble a typical, sharp denaturation transition. A slight curvature is observed between 0 and 1.5 M, which might reflect unfolding around the Trp side chain. However, the baselines cannot be defined. The response at low urea is consistent with a protein that does not populate fully its folded conformation: the slightest amount of denaturant is effective in disrupting a core that is disordered a portion of the time even in the absence of the additive.

The tertiary structure and secondary structure behaviors described above prompt further inspection of the thermal unfolding. To test the two-state nature of the thermal denaturation, Trp fluorescence and aromatic UV absorption were used as techniques

Fig. 5. Fluorescence emission spectrum of Trp22 in abridged *b*₅. The protein concentration was 7 μ M; the sample was maintained at pH 6.8 with 20 mM phosphate buffer, and the temperature was 15°C. In the full-length protein, Trp22 (the only Trp) is part of hydrophobic core 2, located in the barrel (Fig. 2A,B). The bottom trace was recorded in the absence of urea, whereas the top trace corresponds to the same sample to which 6 M urea was added. The vertical at 355 nm is drawn through the maximum of the 0 M trace to guide the eye. The red shift upon denaturation indicates a hydrophobic environment for the Trp in the native state.

complementary to the CD method. The high temperature fluorescence spectrum of abridged *b*₅ displays a decrease in signal intensity, with a small change in maximum wavelength. The intensity change through the transition was monitored at 355 nm and the curve is shown in Figure 6. As in the CD thermal denaturation, a transition is observed. However, the fluorescence intensity of the native and the unfolded states is highly dependent on temperature and the fit of the data yields thermodynamic parameters of insufficient accuracy for meaningful comparison with the CD param-

Fig. 6. Thermal denaturation of abridged *b*₅ as monitored by Trp fluorescence spectroscopy. Emission intensity in arbitrary units is plotted vs. temperature. The protein concentration was 7 μ M; the sample was maintained at pH 6.8 with 20 mM phosphate buffer. The baselines in the folded and unfolded temperature ranges are too steep for an accurate melting point to be obtained.

eters. UV absorption between 260 and 300 nm also reflects a thermal transition. The data exhibit baselines whose slopes do not differ greatly from the slope in the transition region, and the curve cannot be fitted with a satisfactory degree of confidence.

It is customary to compare normalized denaturation curves to determine the number of states involved in the reaction. The secondary structure of the abridged protein, as viewed by CD, appears to undergo a thermal transition between two states, but it is not clear that its tertiary structure as monitored by Trp absorption and fluorescence follows the same trend. When fitting errors are taken into account, models involving more than two states cannot be ruled out.

Characterization of abridged *b5* by NMR methods

NMR spectroscopy experiments were carried out at 15 °C on several samples at low concentration (from 85 to 800 μ M) to probe the structural features of the abridged protein. The one-dimensional spectrum shown in Figure 7 has dispersed lines typical of a folded protein and reminiscent of the apoprotein spectrum (Moore & Lecomte, 1990): signals at 11.2 and -0.8 ppm suggest the formation of the N-capping interaction of helix $\alpha 6$ and the presence of aromatic-aliphatic contacts, respectively. The signal at -0.8 ppm in apocyt *b5* arises from Ile76, shifted by the nearby Trp22 (Moore & Lecomte, 1990), and is a good indicator of long-range side-chain interactions. Other resolved signals in the 5 to 6 ppm region point to the formation of β structure. Upon raising the temperature of the sample the resolved lines first sharpen, then broaden again to disappear between 32 and 35 °C. Cooling the sample restores the same spectrum characteristic of a mixture of conformations. This

Fig. 7. One-dimensional 600 MHz ^1H NMR spectrum of abridged *b5* (85 μ M) in 95% $\text{H}_2\text{O}/5\%$ $^2\text{H}_2\text{O}$ at 15 °C and pH 6.8 (20 mM phosphate buffer). A 5-Hz line broadening was applied. The insets present vertical expansions of the main trace. Signals marked "a" (11.2 ppm) and "b" (-0.8 ppm) correspond to Asp82 NH and Ile76 δCH_3 in the full-length apoprotein. Although the proportion of folded signals is difficult to estimate accurately, it appears lower than inferred from the CD data of Figure 4. Heating the sample to 35 °C removes all chemical shift dispersion and returning to 15 °C restores the spectrum.

reversibility test confirms that the specific fold is reproducibly achieved after disruption at higher temperatures.

Two-dimensional data were collected to characterize the structure further. Scalar connectivities are recognized for Trp22, at chemical shifts similar to those in apocyt *b5*. NOEs are observed between this side chain and the upfield shifted aliphatic residue, which gives rise to an isoleucine pattern. The simplest interpretation is that a core involving Trp22 is formed in the construct. Complete analysis of the data is made difficult by the presence of unfolded protein in equilibrium with the species of interest. For example, the downfield shifted signal attributed to Asp82 in apocyt *b5* and holoct *b5* is sampling at least two conformations on a slow time scale compared to the chemical shift difference (Fig. 8). This chemical exchange phenomenon is observed in apocyt *b5* (Lecomte & Moore, 1991) and is thought to reflect the lability of the N-capping interaction of helix $\alpha 6$. Interestingly, the double-mutant data suggest that the integrity of the cap plays a considerable role in the total stability of the protein, perhaps through non-local effects between $\alpha 6$ and $\beta 1$.

In addition to the exchange process attributed to $\alpha 6$, there are other exchange processes that may or may not be correlated. A one-dimensional saturation transfer experiment in which the shifted methyl group at -0.8 ppm is irradiated shows unequivocally that there is exchange to at least two other forms with chemical shift closer to the random coil value (0.77 and 0.65 ppm). Rotating

Fig. 8. Two-dimensional 500 MHz ^1H NMR NOESY data collected on abridged *b5* (400 μ M) in 95% $\text{H}_2\text{O}/5\%$ $^2\text{H}_2\text{O}$ at 15 °C and pH 6.8. The spectrum, whose amide region is shown, was collected with a mixing time of 100 ms and gradient suppression of the water line (Sklenár et al., 1993). The circled signal marked "a" is detected in TOCSY data and ROESY data where it has the same sign as the diagonal. It arises from a second conformation in slow exchange on the chemical shift time scale. This is directly comparable to the behavior of Asp82 in the full-length apoprotein (Lecomte & Moore, 1991). Three other exchange peaks are marked with "X." Also present in this region are backbone amide NH-NH connectivities indicating the sampling of helical structure.

frame experiments also allowed to distinguish the dipolar effects from the chemical exchange effects. Several residues were found in exchange, most significantly the unique Trp and residues in helical conformation. It is likely that the activation energy barrier separating the different conformations of Trp22 (β_4) and presumably Ile76 (β_2) is related to the formation of the β topology. A generous upper bound to the exchange rate experienced by the backbone is estimated at 600 s^{-1} since signals 100 Hz apart are seen to give effects of the same phase as the diagonal. To put this number in perspective, three kinetic phases are observed when core 2 of the apoprotein refolds after being exposed to 8 M urea and rapidly placed at 0.8 M urea. The fastest phase, accounting for 75% of the signal change, has a rate of 70 s^{-1} (Moore & Lecomte, 1993).

Characterization of abridged *b₅* by chromatography

SEC-FPLC offers a convenient means to characterize long-lived, stable intermediates of different compactness in folding reactions (Uversky, 1993). The elution of $30\text{ }\mu\text{M}$ abridged *b₅* yields an asymmetrical peak, which can be deconvoluted into two components of similar integrated intensity. The apparent molecular weights of the two fractions, as calibrated with globular protein standards, are approximately 10,000 and 14,000 Da. To verify that this is evidence for slow exchange among more than one monomeric form, the elution profile was recorded for protein solutions between $30\text{ }\mu\text{M}$ (value at which some of the CD data and absorbance data were collected) and $3\text{ }\mu\text{M}$ (value at which some of the fluorescence experiments were carried out). A 10-fold reduction of concentration from $30\text{ }\mu\text{M}$ would alter measurably the proportion of the two forms in favor of the monomer if the asymmetry was due to such an equilibrium. The elution profile was found independent of concentration, and the exchange is therefore a unimolecular process. The high apparent molecular weight for the two fractions is in agreement with both sets of conformations having a Stokes radius larger than that of the sphere of equivalent mass (Uversky, 1993).

Discussion

One aim of this work was to test the possibility that apocyt *b₅* contains an autonomously folding module capable of organizing the heme binding site. Abridged *b₅* encompasses this module and includes the 12 residues participating in a conserved hydrophobic core of the cytochrome (Mathews et al., 1979). It was found that a population of the protein fragment folds into a structure that has well-defined, long-lived secondary and tertiary interactions. These interactions resemble those in the full-length apoprotein but overall they are weak, and other conformations are adopted as well. The denaturation of the fragment can be monitored with conventional methods; the data reveal broad transitions and suggest a multi-state process. This is unlike the full-length proteins (wild-type and N-cap mutants), which show sharper transitions.

In a majority of known cases, the replacement of a few amino acids in a protein causes a lowering of the standard free energy difference between folded and unfolded states. Disruptive N-cap mutations in helix α_6 of apocyt *b₅*, of which the His80Ala/Pro81Ala protein is one example, fall in this category. Also in many instances, the deletion of a few consecutive residues tends to be detrimental (Alexandrescu et al., 1994; Fetrow, 1995), and a reasonable extrapolation would predict a 30-residue excision to be

catastrophic. Remarkably, abridged *b₅* folds. This conformational bias shows that the module is a potential support for loops of various sizes and a useful building block for protein engineering (ElHawrani et al., 1994).

The folded state of abridged *b₅* has low thermodynamic stability. In exploratory experiments to define favorable conditions for structural work, the module was found to be sensitive to pH changes and to respond to sucrose addition (Lee & Timasheff, 1981). However, in order to determine whether the limited stability is an intrinsic property or an accident of the design, it will be necessary to study other constructs. Assuming similar core structures, the reduced ΔG° compared to the full-length protein can be explained by the removal of α_2 and α_5 , Λ composition and Λ size. To improve $\Delta G^\circ(N \rightarrow U)$ in abridged *b₅*, several schemes can be considered. For example, the core can be stabilized by rational site-specific mutations, as described, for example, for thermolysin-like proteases (Vriend & Eijssink, 1993), T4 lysozyme (Zhang et al., 1995), and subtilisin BPN' prodomain (Ruvinov et al., 1997). More interesting are loop optimization and local propensity adjustments as explained below.

Loops are known to affect the stability of a protein fold (Gekko et al., 1993; Hoedemaeker et al., 1993; Dion-Schultz & Howell, 1997; Fetrow et al., 1997). This is illustrated notably by circularly permuted proteins (Garrett et al., 1996; Uversky et al., 1996) where natural turns and artificial linkers can make the difference between a well-folded protein and a molten globule. The effects of loop length have been systematically investigated in variants of the α -spectrin SH3 domain, a small monomeric β protein (Viguera & Serrano, 1997). In this system, insertion of an increasing number of glycines into a flexible loop leads to a native state destabilization that levels off at a value of 4 kJ/mol for 10 residues. In apocyt *b₅* ($\Delta G^\circ = 8\text{ kJ/mol}$), Λ contains 39 residues; a few of these are involved in structure and most display enhanced motions on picosecond to microsecond time scales (Falzone et al., 1996; S. Bhattacharya, C.J. Falzone, J.T.J. Lecomte, in prep.). Thus, a high energetic price is likely associated with Λ closure, and abridged *b₅* ($\Lambda = 7$) might have a lower entropic penalty to pay (Poole et al., 1991). The current T4 loop may place inordinate strain on the β_2 strand as it inserts between strands β_1 and β_3 of the same molecule, and prevent efficient packing of core 2. The Λ tether could, therefore, be either altered to enhance flexibility, with the caveat of entropic costs, or trimmed further, with the caveat that tethers that are too short or unable to adopt the proper conformation may facilitate the transition from monomeric protein to oligomeric ones (Bennett et al., 1995; Green et al., 1995).

Judicious increases in the local propensity for a given native element of structure are known to result in stabilization (Villegas et al., 1995; Muñoz et al., 1996; Viguera et al., 1997). Secondary structure prediction algorithms such as PHD (Rost et al., 1994) identify in abridged *b₅* one helical segment, corresponding to α_1 , and extended chain preference in all β regions of cyt *b₅*. Turns or loops are predicted between α_1 and β_4 , β_3 , and β_2 , and in place of α_6 . According to the calculated propensities, α_1 and β_3 are elements to be strengthened by mutation. In contrast, strong non-native propensities in a sequence have detrimental effects (Wrabl & Shortle, 1996) but can be structurally overcome by nonlocal interactions. In one example, Serrano and coworkers engineered α propensities in a β strand of an SH3 domain and observed that the native fold was unchanged though its stability much reduced (Prieto et al., 1997). Similar effects are also detected in β -lactoglobulin, where during the refolding reaction an α helix forms transiently in

a region of high α propensity but of native β structure (Hamada et al., 1996; Hamada & Goto, 1997). Non-native local tendencies are noted by PHD and LINUS (Srinivasan & Rose, 1995) in the primary structure of rat liver cyt b_5 , particularly in the $\beta 4$ – $\beta 3$ turn (T3) and $\alpha 6$. It is possible that the deleted segment overrides inappropriate local propensities in apocyt b_5 , whereas the same uncompensated propensities carry more relative weight in abridged b_5 . These non-native stretches are good target for corrective mutagenesis attempts.

Significance

It is clear that abridged b_5 could be altered in several ways in order to increase $\Delta G^\circ(N \rightarrow U)$ and to investigate the stability issues of the protein folding problem. Although a wider energetic gap between folded and unfolded ensembles would facilitate the detailed description of the protein, it would also conceal interesting aspects that low stability reveals, namely the emergence of structural specificity. The abbreviated primary structure message from cyt b_5 codes for an artificial protein unconstrained by evolutionary demands on $\Delta G^\circ(N \rightarrow U)$, and yet displays a preference for a particular three-dimensional structure. Thus, the stereochemical code at work dictates a considerable restriction of the conformational space searched by the chain. If one considers the arbitrariness of the cleavage points and of the inserted tether, it can be speculated that the folding signals encoded in the summary sequence are robust, overdetermined, and distributed over the polypeptide length, as suggested by Lattman and Rose (1993). As a corollary, the encoding of the heme binding site helices and turns must be separate from that of the organizing module, and confined to the excised fragment in an independent scheme.

The NMR data show that several conformational ensembles are sampled by the chain. The ensemble yielding resolved lines is attributed to the native form (N). The unfolded ensemble (U) is also populated while the other set(s) (U') account for additional near-random coil signals and seem to contain various degrees of fluctuating secondary structure detectable by CD spectroscopy. These less structured ensembles can readily be accommodated in the thermal denaturation data by extension to a multistate model. The multistatedness of the process indicates a breakdown of cooperativity, which emphasizes the role of local interactions as opposed to long-range interactions in determining the fold. In this case, the specific side-chain side-chain contacts observed by NMR spectroscopy would result from the predispositions of the chain.

Abridged b_5 goes in and out of its native form on a time scale similar to that observed when apocyt b_5 refolds from the urea-denatured ensemble. It is conceivable that related pathways are followed by both proteins in establishing their structure. This appears to be the case at least at a segmental level illustrated by the conformational equilibrium of $\alpha 6$ and implies that not only the determinants of the final fold are coded into the reduced sequence, but also the routes to reach and escape from this fold. The conservation of barriers and kinetic traps for a designed fragment with low stability brings insight into the folding reaction. Two features emerge: the formation of an essential helix N-cap interaction, probably accompanied by docking of the helix on a strand, and the subordinate role played by hydrophobic core 2. This behavior is consistent with the dominance of local determinants in the formation of the apocyt b_5 fold.

Materials and methods

Reagents

All buffer and M9 media components were purchased from Sigma (St. Louis, Missouri) and Aldrich (Milwaukee, Wisconsin). Cell culture reagents were from Difco (Detroit, Michigan); bacterial cell strains and plasmids from Novagen (Madison, Wisconsin) and Clontech (Palo Alto, California). Enzymes used in polymerase chain reactions (PCR) were supplied by New England Biolabs (Beverly, Massachusetts). Column chromatography materials and PCR beads were purchased from Pharmacia Biotech (Uppsala, Sweden). DNA purification materials were obtained from Promega (Madison, Wisconsin) and Genomed (Research Triangle Park, North Carolina).

Protein preparation and purification

For the wild-type protein, the modified pET-3d(*cytb₅*) plasmid was transformed into BL21(DE3) *E. coli* cell line as described earlier (Falzone et al., 1996). The double mutant gene for the His80Ala/Pro81Ala protein was prepared according to the megaprimer method of polymerase chain reaction (Kuipers et al., 1991) and otherwise treated as the wild-type gene. Growth and purification followed the usual procedure (Beck von Bodman et al., 1986; Falzone et al., 1996).

The gene for the abridged protein was constructed by taking advantage of *Xma*I sites in the *cytb₅* gene to remove the heme binding region, and inserting a short loop in its place by double ligation. A string of six histidines following a methionine was coded for at the C-terminus. The gene was inserted into the pET3d plasmid as described previously (Falzone et al., 1996). DNA sequencing was performed to insure that the desired protein would be obtained. Expression of the abridged protein was carried out in *E. coli* strain BL21(DE3) at 25 °C. Purification of abridged b_5 (theoretical pI 6.4) was performed on a chelating Sepharose column charged with Ni^{2+} and eluted with 20 mM phosphate, pH 7.9, 500 mM NaCl, 500 mM imidazole. It was necessary to proceed with a second purification step (Sephadex G-75 column, 20 mM phosphate buffer, pH 6.8) to obtain 95% pure protein by SDS-PAGE. MALDI mass spectrometry analysis was carried out at the Wistar Institute. The protein has the expected molecular weight of 8,350 Da. Small amounts of higher molecular weight impurities (10,000–11,000 Da) were present as detected in overloaded gels.

Absorption spectroscopy

The oxidized holoprotein extinction coefficients were taken to be 114 mM⁻¹ cm⁻¹ at 413 nm (Moore & Lecomte, 1993). The abridged protein and apoprotein extinction coefficients at 280 nm were estimated at 11 mM⁻¹ cm⁻¹ from the aromatic amino acid content of 1 Trp and 4 Tyr (Mach et al., 1992). Thermal denaturation was performed on 30 μ M protein samples (20 mM phosphate buffer, pH 6.8) in 1-cm pathlength cells by collecting a spectrum from 320 to 260 nm every 2 °C, from 0 to 80 °C with an equilibration time of 6 min. The maximum absorption change upon denaturation after baseline correction was found at 288.5 nm and an isosbestic point at 285.5 nm. In this and other thermal denaturation experiments, reversibility was assessed by returning the sample temperature to the lowest value and comparing the spectra. Better than 85% reversibility was typically attained. No evidence for aggregation was observed in the experiments reported here.

Circular dichroism

CD spectra were collected on an AVIV 62DS CD spectropolarimeter. The cell pathlength was 2 mm. Mean residual ellipticity (MRE) in deg cm² dmol⁻¹ was obtained through the equation

$$\text{MRE}(\lambda) = \Theta(\lambda)/(10lcn) \quad (1)$$

where $\Theta(\lambda)$ is the ellipticity in degrees at wavelength λ , l is the pathlength in cm, c is the concentration in M, and n is the number of residues in the protein (Schmid, 1989).

Thermal denaturation of the full-length proteins (wild-type and double mutant) were carried out at pH 7.5 (20 mM phosphate buffer) on 4 μ M protein samples. The process was monitored at 220 nm. Equilibration was 2 min, and signal averaging 30 s. Thermal denaturation of abridged *b₅* was monitored at 215 and 222 nm on samples containing 20 to 40 μ M protein (20 mM phosphate buffer, pH 6.8). The temperature was raised in steps of 2 °C, from 0 to 80 °C with an equilibration time of 5 min.

Fluorescence spectroscopy

The tryptophan fluorescence emission was recorded on a Spex Fluorolog 1681 fluorimeter with 295 nm excitation and a bandwidth of 5 nm. The effect of urea on the spectrum (Fig. 5) was determined on a 3 μ M protein sample (20 mM phosphate buffer, pH 6.8, at 15 °C). Urea titrations were performed on 3 and 6 μ M protein sample (20 mM phosphate buffer, pH 6.8, at 25 and 15 °C) with an automatic titrator (Aviv model ATF 105) delivering urea in 45 increments over the 0 to 6 M range. Urea concentrations were measured using the refractive index method (Pace, 1986). The thermal denaturation was monitored at 355 nm on a 7 μ M sample (20 mM phosphate buffer, pH 6.8). The temperature was raised in 2 °C steps with stirring during equilibration (6 min).

NMR spectroscopy

¹H NMR spectra were acquired on Bruker spectrometers (AMX2-500 and DRX-600) in 95% H₂O/5% ²H₂O with a modified WATERGATE sequence (Sklénár et al., 1993) to suppress the water signal. The mixing times were 100 ms (NOESY; Kumar et al., 1980) and 40 ms (TOCSY with relaxation-compensated DIPSI mixing scheme; Cavanagh & Rance, 1992). In addition, NOESY and rotating frame Overhauser enhancement data (Bothner-By et al., 1984) were collected with presaturation of the water line and a mixing time of 100 and 40 ms, respectively. The spin-lock field in the ROESY spectrum was centered on the water and the field strength was set to 5.0 kHz. The data sets were collected with a spectral width of 12,500 Hz in the acquisition dimension and 6,250 Hz in the indirect dimension, except for the ROESY data set, which was collected over 7,042 Hz in both dimensions. The number of t₁ increments varied between 380 and 512. Quadrature detection was achieved in the indirect ¹H dimension by using TPPI (Drobny et al., 1979). Processing was performed with Felix (Molecular Simulation Inc., San Diego, California) with squared sinebell (shift of 72–90°).

Size exclusion chromatography–fast protein liquid chromatography

All experiments were carried out on a Pharmacia-LKB FPLC system with a Superose 12 column maintained at 25 °C. For these

experiments, the protein was passed through a second G-75 column to eliminate high molecular weight impurities. The protein concentration ranged between 3 and 30 μ M (100 mM phosphate buffer, pH 6.8). All calibration standards (ubiquitin, insulin, myoglobin, carbonic anhydrase, and albumin) were prepared in the same buffer. A plot of the inverse of the elution volume vs. molecular weight (Uversky, 1993) was used to obtain an estimate of the apparent protein mass.

Structure calculations

Structures were calculated with X-PLOR (Brünger, 1992) by removing from the restraint files used for the apoprotein (Falzone et al., 1996) all constraints pertaining to residues 33 to 74. The procedure was otherwise as described in Falzone et al. (1996), starting with extended chaining and applying standard distance geometry and simulated annealing protocol to obtain 100 structures, which were refined by restrained minimization.

Thermal denaturation data analysis

Whenever possible, the thermal denaturation data were fitted to equations representing a two-state behavior (Schellman, 1987; Schmid, 1989; Dill, 1990). The observable, $Y(T)$, has a value between the fully folded (n) and fully unfolded (u) baselines represented by the linear equations

$$Y_n(T) = A_n T + B_n \quad (2)$$

$$Y_u(T) = A_u T + B_u \quad (3)$$

At any point in the transition, the apparent fraction of unfolded protein, F_u , is given by

$$F_u(T) = [Y(T) - Y_n(T)]/[Y_u(T) - Y_n(T)]. \quad (4)$$

In addition,

$$F_u(T) = K(T)/[1 + K(T)] \quad (5)$$

and

$$K(T) = \exp[-\Delta G^\circ(T)/RT]. \quad (6)$$

By expanding the Gibbs–Helmholtz relation, and using the midpoint T_m as the reference temperature, one obtains

$$\begin{aligned} \Delta G^\circ(T) = & \Delta H^\circ(T_m)[1 - T/T_m] \\ & - \Delta C_p^\circ[(T_m - T) + T \ln(T/T_m)]. \end{aligned} \quad (7)$$

The final equation combines all of the above to relate $Y(T)$ to $\Delta H^\circ(T_m)$, ΔC_p° , T_m , A_n , B_n , A_u , and B_u . The program NFIT was used to fit the data sets singly. Savuka (a global fitting program written by Drs. D.G. Lambright and O. Bilsel) was used to extract the parameters from the CD melts recorded at different wavelengths. The thermodynamic quantities obtained by optical methods in this work have limited precision and accuracy, in most cases because of inadequate definition of the baseline(s). The uncertainties reported in Table 1 are the standard deviation of the fits.

Acknowledgments

This work was supported by the National Institutes of Health grant GM-54217. The members of Dr. C.R. Matthews's research group, in particular Dr. Jill Zitzewitz, Katherine Bowers, Brandon Doyle, Peter Gualfetti, and John O'Neill, are thanked for technical assistance and discussions. The help of Dr. Kaye Speicher (mass spectrometry, Wistar Institute), Dr. Kristen MacLaughlin (protein preparation) and Shibani Bhattacharya (NMR spectroscopy) is also acknowledged. The numerous comments and suggestions of Drs. Christopher Falzone, C. Robert Matthews, and George Rose were essential in the presentation of the work.

References

- Alexandrescu AT, Abeygunawardana C, Shortle D. 1994. Structure and dynamics of a denatured 131-residue fragment of staphylococcal nuclease: A heteronuclear NMR study. *Biochemistry* 33:1063–1072.
- Anfinsen CB. 1973. Principles that govern the folding of protein chains. *Science* 181:223–230.
- Attwood TK, Beck ME, Bleasby AJ, Parry-Smith DJ. 1994. PRINTS: A database of protein motif fingerprints. *Nucleic Acids Res* 22:3590–3596.
- Bairoch A. 1991. PROSITE: A dictionary of sites and patterns in proteins. *Nucleic Acids Res* 19:2241–2245.
- Beck von Bodman A, Schuler MA, Jollie DR, Sligar SG. 1986. Synthesis, bacterial expression, and mutagenesis of the gene coding for mammalian cytochrome *b₅*. *Proc Natl Acad Sci USA* 83:9443–9447.
- Behe MJ, Lattman EE, Rose GD. 1991. The protein-folding problem: The native fold determines packing, but does packing determine the native fold? *Proc Natl Acad Sci USA* 88:4195–4199.
- Bennett MJ, Schlunegger MP, Eisenberg D. 1995. 3D domain swapping: A mechanism for oligomer assembly. *Protein Sci* 4:2455–2468.
- Bothner-By AA, Stephens RL, Lee J-M, Warren CD, Jeanloz RW. 1984. Structure determination of a tetrasaccharide: Transient nuclear Overhauser effects in the rotating frame. *J Am Chem Soc* 106:811–813.
- Brünger AT. 1992. *X-PLOR. Version 3.1. A system for X-ray crystallography and NMR*. New Haven: Yale University Press.
- Bryan P, Wang L, Hoskins J, Ruvinov S, Strausberg S, Alexander P, Almog O, Gilliland G, Gallagher T. 1995. Catalysis of a protein folding reaction: Mechanistic implications of the 2.0 Å structure of the subtilisin-prodomain complex. *Biochemistry* 34:10310–10318.
- Cavanagh J, Rance M. 1992. Suppression of cross-relaxation effects in TOCSY spectra via a modified DIPSI-2 mixing sequence. *J Magn Reson* 96:670–678.
- Dahiyat BI, Mayo SL. 1997. De novo protein design: Fully automated sequence selection. *Science* 278:82–87.
- Desjarlais JR, Handel TM. 1995. De novo design of the hydrophobic cores of proteins. *Protein Sci* 4:2006–2018.
- Dill K. 1990. Dominant forces in protein folding. *Biochemistry* 29:7133–7155.
- Dion-Schultz A, Howell EE. 1997. Effects of insertions and deletions in a β -bulge region of *Escherichia coli* dihydrofolate reductase. *Protein Eng* 10:263–272.
- Drobny G, Pines A, Sinton S, Weitekamp DP, Wemmer D. 1979. Fourier transform multiple quantum nuclear magnetic resonance. *Symp Faraday Soc* 13:49–55.
- ElHawrani AS, Foreton KM, Sessions RB, Clarke AR, Holbrook JJ. 1994. Engineering surface loops of proteins—A preferred strategy for obtaining new enzyme function. *Trends Biotechnol* 12:207–211.
- Falzone CJ, Mayer MR, Whiteman EL, Moore CD, Lecomte JTJ. 1996. Design challenges for hemoproteins: The solution structure of apocytochrome *b₅*. *Biochemistry* 35:6519–6526.
- Fetrow JS. 1995. Ω loops: Nonregular secondary structures significant in protein function and stability. *FASEB J* 9:708–717.
- Fetrow JS, Horner SR, Oehrl W, Schaak DL, Boose TL, Burton RE. 1997. Analysis of the structure and stability of Ω loop A replacements in yeast iso-1-cytochrome *c*. *Protein Sci* 6:197–210.
- Garrett JB, Mullins LS, Raushel FM. 1996. Are turns required for the folding of ribonuclease T1? *Protein Sci* 5:204–211.
- Gekko K, Yamagami K, Kunori Y, Ichihara S, Kodama M, Iwakura M. 1993. Effect of point mutation in a flexible loop on the stability and enzymatic activity of *Escherichia coli* dihydrofolate reductase. *J Biochem (Tokyo)* 113:74–80.
- Green SM, Gittis AG, Meeker AK, Lattman EE. 1995. One-step evolution of a dimer from a monomeric protein. *Nature Struct Biol* 2:746–751.
- Hamada D, Goto Y. 1997. The equilibrium intermediate of β -lactoglobulin with non-native α -helical structure. *J Mol Biol* 269:479–487.
- Hamada D, Segawa S, Goto Y. 1996. Non-native α -helical intermediate in the refolding of β -lactoglobulin, a predominantly β -sheet protein. *Nature Struct Biol* 3:868–873.
- Hoedemaeker FJ, vanEijsden RR, Diaz CL, Depater BS, Kijne JW. 1993. Destabilization of pea lectin by substitution of a single amino acid in a surface loop. *Plant Mol Biol* 22:1039–1046.
- Holm L, Sander C. 1993. Protein structure comparison by alignment of distance matrices. *J Mol Biol* 233:123–138.
- Hoof RWW, Vriend G, Sander C, Abola EE. 1996. Errors in protein structures. *Nature* 381:272.
- Huntley TE, Strittmatter P. 1972. The effect of heme binding on the tryptophan residue and the protein conformation of cytochrome *b₅*. *J Biol Chem* 247:4641–4647.
- Kono H, Doi J. 1994. Energy minimization method using automata network for sequence and side-chain conformation prediction from given backbone geometry. *Proteins Struct Funct Genet* 19:244–255.
- Kraulis PJ. 1991. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J Applied Cryst* 24:946–950.
- Kuipers OP, Boot HJ, de Vos WM. 1991. Improved site-directed mutagenesis method using PCR. *Nucleic Acids Res* 19:4558.
- Kumar A, Ernst RR, Wüthrich K. 1980. A 2D NOE experiment for the elucidation of complete proton-proton cross-relaxation networks in biological macromolecules. *Biochem Biophys Res Commun* 95:1–6.
- Lattman EE, Rose GD. 1993. Protein folding—What's the question? *Proc Natl Acad Sci USA* 90:439–441.
- Lazar GA, Desjarlais JR, Handel TM. 1997. De novo design of the hydrophobic core of ubiquitin. *Protein Sci* 6:1167–1178.
- Lecomte JTJ, Moore CD. 1991. Helix formation in apocytochrome *b₅*: The role of a neutral histidine at the N-cap position. *J Am Chem Soc* 113:9663–9665.
- Lee JC, Timasheff SN. 1981. The stabilization of proteins by sucrose. *J Biol Chem* 256:7193–7201.
- Mach H, Middaugh CR, Lewis RV. 1992. Statistical determination of the average values of the extinction coefficients of tryptophan and tyrosine in native proteins. *Anal Biochem* 200:74–80.
- Mathews FS. 1985. The structure, function and evolution of cytochromes. *Prog Biophys Molec Biol* 45:1–56.
- Mathews FS, Czerwinski EW, Argos P. 1979. The X-ray crystallographic structure of calf liver cytochrome *b₅*. In: Dolphin D, ed. *The porphyrins*. New York: Academic Press. pp 107–147.
- Mathews CR. 1993. Pathways of protein folding. *Annu Rev Biochem* 62:653–683.
- Miranker AD, Dobson CM. 1996. Collapse and cooperativity in protein folding. *Curr Opin Struct Biol* 6:31–42.
- Moore CD, Lecomte JTJ. 1990. Structural properties of apocytochrome *b₅*: Presence of a stable native core. *Biochemistry* 29:1984–1989.
- Moore CD, Lecomte JTJ. 1993. Characterization of an independent structural unit in apocytochrome *b₅*. *Biochemistry* 32:199–207.
- Muñoz V, Cronet P, Lopez-Hernandez E, Serrano L. 1996. Analysis of the effect of local interactions on protein stability. *Fold Design* 1:167–178.
- Pace CN. 1986. Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol* 131:266–280.
- Pfeil W. 1993. Thermodynamics of apocytochrome *b₅* unfolding. *Protein Sci* 2:1497–1501.
- Poole LB, Loveys DA, Hale SP, Gerlt JA, Stanszyk SM, Bolton PH. 1991. Deletion of the Ω -loop in the active site of staphylococcal nuclease. I. Effects on catalysis and stability. *Biochemistry* 30:3621–3627.
- Presta LG, Rose GD. 1988. Helix signals in proteins. *Science* 240:1632–1641.
- Prieto J, Wilmans M, Jimenez MA, Rico M, Serrano L. 1997. Non-native local interactions in protein folding and stability: Introducing a helical tendency in the all β -sheet α -spectrin SH3 domain. *J Mol Biol* 268:760–778.
- Privalov PL. 1979. Stability of proteins: Small globular proteins. *Adv Prot Chem* 33:167–241.
- Puitsyn OB. 1995. Molten globule and protein folding. *Adv Prot Chem* 47:83–229.
- Richards FM. 1977. Areas, volumes, packing and protein structure. *Ann Rev Biophys Bioeng* 6:151–171.
- Robertson AD, Murphy KP. 1997. Protein structure and the energetics of protein stability. *Chem Rev* 97:1251–1267.
- Roder H, Colón W. 1997. Kinetic role of early intermediates in protein folding. *Curr Opin Struct Biol* 7:15–28.
- Rost B, Sander C, Schneider R. 1994. PHD—An automatic mail server for protein secondary structure prediction. *Comp Appl Biosci* 10:53–60.
- Ruvinov S, Wang L, Ruan B, Almog O, Gilliland GL, Eisenstein E, Bryan PN. 1997. Engineering the independent folding of the subtilisin BPN' prodomain: Analysis of two-state folding versus protein stability. *Biochemistry* 36:10414–10421.
- Sauer RT. 1996. Protein folding from a combinatorial perspective. *Fold Design* 1:R45–R49.
- Schellman JA. 1987. The thermodynamic stability of proteins. *Annu Rev Biophys Biophys Chem* 16:115–137.
- Schmid FX. 1989. Spectral methods of characterizing protein conformation and

- conformational changes. In: *Protein structure a practical approach*. New York: IRL Press. pp 251–285.
- Shortle D. 1996. The denatured state (the other half of the folding equation) and its role in protein stability. *FASEB J* 10:27–34.
- Sklenár V, Píotko M, Leppik R, Saudek V. 1993. Gradient-tailored water suppression for ¹H-¹⁵N HSQC experiments optimized to retain full sensitivity. *J Magn Reson Ser A* 102:241–245.
- Srinivasan R, Rose GD. 1995. LINUS: A hierarchic procedure to predict the fold of a protein. *Proteins Struct Funct Genet* 22:81–99.
- Tanford C. 1968. Protein denaturation. *Adv Prot Chem* 23:121–282.
- Tanford C. 1970. Protein denaturation. *Adv Prot Chem* 24:1–95.
- Uversky VN. 1993. Use of fast protein size-exclusion liquid chromatography to study the unfolding of proteins which denature through the molten globule. *Biochemistry* 32:13288–13298.
- Uversky VN, Kutysheiko VP, Protasova N, Rogov VV, Vassilenko KS, Gudkov AT. 1996. Circularly permuted dihydrofolate reductase possesses all the properties of the molten globule state, but can resume functional tertiary structure by interaction with its ligands. *Protein Sci* 5:1844–1851.
- Viguera AR, Serrano L. 1997. Loop length, intramolecular diffusion and protein folding. *Nature Struct Biol* 4:939–946.
- Viguera AR, Villegas V, Avilés FX, Serrano L. 1997. Favourable native-like helical local interactions can accelerate protein folding. *Fold Design* 2:23–33.
- Villegas V, Viguera AR, Avilés FX, Serrano L. 1995. Stabilisation of proteins by rational design of α -helix stability using helix/coil transition theory. *Fold Design* 1:29–34.
- Vriend G, Eijssink V. 1993. Prediction and analysis of structure, stability and unfolding of thermolysin-like proteases. *J Comp-Aided Molec Design* 7:367–396.
- Vriend G, Sander C. 1993. Quality control of protein models: Directional atomic contact analysis. *J Appl Cryst* 26:47–60.
- Wrabl JO, Shortle D. 1996. Perturbations of the denatured state ensemble: Modeling their effects on protein stability and kinetics. *Protein Sci* 5:2343–2352.
- Zhang O, Kay LE, Shortle D, Forman-Kay JD. 1997. Comprehensive NOE characterization of a partially folded large fragment of staphylococcal nuclease D131D, using NMR methods with improved resolution. *J Mol Biol* 272:9–20.
- Zhang XJ, Baase WA, Shoichet BK, Wilson KP, Matthews BW. 1995. Enhancement of protein stability by the combination of point mutations in T4 lysozyme is additive. *Protein Eng* 8:1017–1022.