Role of Native Disulfide Bonds in the Structure and Activity of Insulin-like Growth Factor 1: Genetic Models of Protein-Folding Intermediates†

Linda Owers Narhi,‡ Qing-Xin Hua,§ Tsutomu Arakawa,‡ G. Michael Fox,‡ Larry Tsai,‡ Robert Rosenfeld,‡ Paige Holst,‡ James A. Miller,*,‡ and Michael A. Weiss*,§,∥

Amgen Inc., Amgen Center, Thousand Oaks, California 91320, Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, and Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts 02114

Received December 14, 1992; Revised Manuscript Received February 10, 1993

ABSTRACT: Insulin and insulin-related proteins contain three motif-specific disulfide bonds. Here we examine the role of these disulfide bonds in the folding and function of one family member, human insulin-like growth factor 1 (IGF-1). Analogues containing pairwise Cys→Ser or Cys→Ala substitutions were expressed in *Escherichia coli*, purified, and analyzed with respect to receptor-binding, solution structure, and thermodynamic stability. An analogue lacking all three disulfide bonds (designated des-Cys-IGF-1) is inactive and unfolded. Introduction of the [18–61] disulfide bond, previously shown to occur in an early intermediate in oxidative refolding [Miller, J. A., Owers-Narhi, L., Hua, Q. X., Rosenfeld, R., Arakawa, T., Rohde, M., Prestrelski, S., Lauren, S., S. Stoney, K. S., Tsai, L., & Weiss, M. A. (1993) *Biochemistry* (preceding paper in this issue)], results in a compact partially folded state with low but significant biological activity. Additional but incomplete structural organization and biological activity are observed following introduction of either the [6–48] or the [47–52] disulfide bonds. Native function, structure, and stability require the presence of all three disulfide bonds. These analogues provide genetic models of IGF-1 protein-folding intermediates. Their characterization suggests that bifurcation of the IGF-1 folding pathway reflects alternative late steps in the folding of a molten-globule intermediate.

Disulfide bonds introduce nonlocal topological restraints in the folding of a polypeptide, and their formation provides a convenient assay to monitor kinetic stages of oxidative refolding (Creighton & Goldenberg, 1984; Goldenberg, 1989). A well-characterized model for such analysis is provided by bovine pancreatic trypsin inhibitor (BPTI),[1] which contains three disulfide bonds. In this system, the reduced protein (58 residues) is unfolded. Successive stabilization of native structural elements is observed in highly populated one- and two-disulfide intermediates (Creighton & Goldenberg, 1984; Stassinopoulou et al., 1984; States et al., 1984, 1987; Weissman & Kim, 1991; van Mierlo et al., 1991a,b, 1992). Native-like structure is also observed in a one-disulfide analogue, although the importance of the corresponding intermediate to oxidative refolding is not clear (Staley & Kim, 1992). Transient formation of nonnative disulfide bonds may also occur (Darby et al., 1992; Weissman & Kim, 1992a; Goldenberg, 1992). Investigation of partially reduced intermediates and corresponding genetic models complements stopped-flow ^{1}H-NMR and optical studies of kinetic intermediates (Roder & Wuthrich, 1986; Roder et al., 1988; Udgaonkar & Baldwin, 1988; Radford et al., 1992).

BPTI provides a classical model because oxidative refolding yields a unique native state (the presumed thermodynamic ground state) in accord with the Anfinsen concept (Anfinsen, 1973). In recent years, however, it has become clear that a more general model of protein folding must include contributions from accessory and regulatory proteins, including protein chaperones and enzymes such as the disulfide and proline isomerases (Krebs et al., 1983; Jaenicke et al., 1991; Martin et al., 1991; Gething & Sambrook, 1992). In addition, folding may be influenced by specific contributions from pro sequences (Zhu et al., 1989; Winther & Sorensen, 1991; Weissman & Kim, 1992b) and by differences in the kinetics of alternative pathways (Baker et al., 1992).

We and others have recently observed unusual refolding behavior in human IGF-1 (Hober et al., 1992; Miller et al., 1993), a globular protein of 70 residues that (like BPTI) contains 3 disulfide bonds (Rinderknecht & Humbel, 1978). Oxidative refolding yields two alternative disulfide isomers rather than a unique native state (Raschdorf et al., 1988; Meng et al., 1989). The two isomers (designated IGF-1 and IGF–swap) are of similar thermodynamic stability but exhibit distinct tertiary structures (Miller et al., 1993). Interconversion is observed in the presence of a substoichiometric concentration of a reducing agent (Hober et al., 1992) and requires passage through a large activation barrier ($E_a \approx 12$ kcal/mol; Miller et al., 1993). Interconversion from either native IGF-1 or IGF-swap leads to the same ratio of products as initial oxidative folding. IGF-1 thus provides an example of a bifurcating protein-folding pathway under thermodynamic control.

The primary structure of IGF-1 is shown in Figure 1, and its three-dimensional structure (Cooke et al., 1991) is shown

† Supported by Amgen Inc. and the National Institutes of Health (M.A.W.).

* Address correspondence to either author at Amgen Inc. (J.A.M.) or Harvard Medical School (M.A.W.).

‡ Amgen Inc.

§ Harvard Medical School.

∥ Massachusetts General Hospital.

[1] Abbreviations: BPTI, bovine pancreatic trypsin inhibitor; GdnHCl, guanidine hydrochloride; CD, circular dichroism; FTIR, Fourier-transform infrared; IGF-1, insulin-like growth factor 1; NMR, nuclear magnetic resonance; NOESY, nuclear Overhauser spectroscopy; OD, optical density; PBS, phosphate-buffered saline; photo-CIDNP, photochemical dynamic nuclear polarization; rp-HPLC, reverse-phase high-performance liquid chromatography; SDS–PAGE, sodium dodecyl sulfate–polyacrylamide gel electrophoresis; TOCSY, total correlation spectroscopy; 2D-NMR, two-dimensional NMR. Nomenclature: The refolded species with insulin-like disulfide pairing [6–48; 18–61; 47–52] is designated IGF-1; a refolded species with alternative disulfide pairing [6–47; 18–61; 48–52] is designated IGF-swap; analogues are otherwise designated by their disulfide pairing (e.g., an analogue with substitutions C47A and C52A is designated [6–48; 18–61]). Amino acids are designated in the text by three-letter abbreviations and in figures and tables by the standard single-letter code.

0006-2960/93/0432-5214$04.00/0 © 1993 American Chemical Society

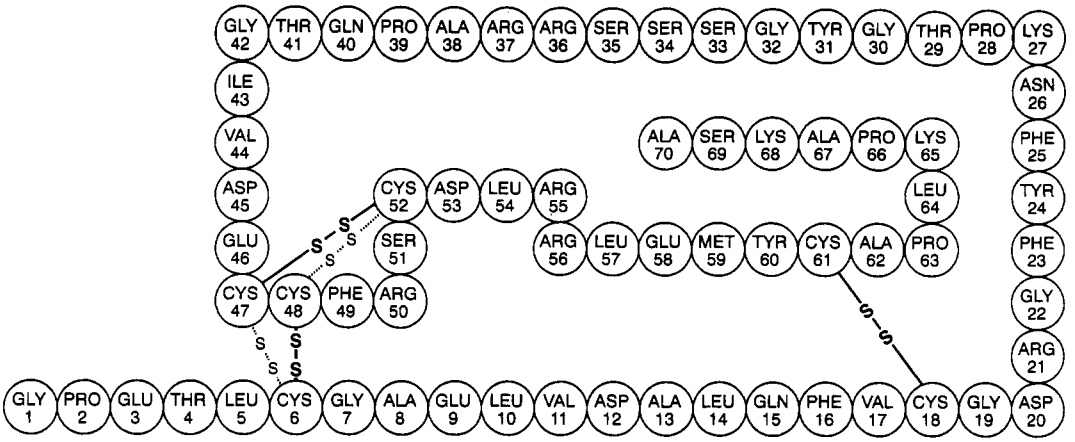

FIGURE 1: Sequence of human IGF-1. Native (insulin-like) disulfide bonds ([6–48; 18–61; 47–52]) are indicated by solid lines. Alternative disulfide bonds in IGF-swap ([6–48; 18–61; 47–52]; Raschdorf et al., 1988; Meng et al., 1989; Hober et al., 1992; Miller et al., 1993) are indicated by dotted lines.

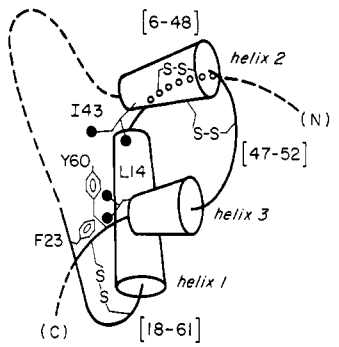

FIGURE 2: Cylinder model of the solution structure of IGF-1 (Cooke et al., 1991). Disordered regions are shown by dashed line. The positions of the three native disulfide bonds (6–48, 47–52, and 18–61) and key internal side chains are indicated. α-Helix 1 in IGF-1 corresponds to insulin B-chain helix B9–B19 in the crystallographic T-state; α-helices 2 and 3 correspond to insulin A-chain helices A1–A8 and A12–A18, respectively (Baker et al., 1988).

in schematic form in Figure 2. Although the IGF-1-folding pathway has not been characterized in detail, major intermediates include at least [18–61] (the disulfide shared in common between native IGF-1 and IGF-swap), [6–48; 18–61] (an immediate precursor of native IGF-1), and [18–61; 6–47] (an immediate precursor to IGF-swap with nonnative disulfide 6–47). Because of the conceptual importance of this model to studies of disulfide bond formation and protein folding (Creighton, 1984), we have investigated the structure, stability, and function of analogues containing zero, one, or two native disulfide bonds. These analogues contain combinations of Cys→Ala or Cys→Ser mutations and provide genetic models of protein-folding intermediates (Miller et al., 1993). This study extends previous studies of analogous BPTI mutants (Eigenbrot et al., 1990; Naderi et al., 1991; van Mierlo et al., 1991a,b, 1992; Staley & Kim, 1992) and represents a first step toward structural characterization of intermediates in IGF-1 refolding.

MATERIALS AND METHODS

Reagents and Buffers. Guanidine hydrochloride (>99% purity) was purchased from Amresco. PBS consists of 150 mM NaCl/10 mM sodium phosphate (pH 7.4) and was prepared by dilution from Dulbecco 10 × PBS. Buffer A consists of 50 mM potassium phosphate (pH 7.4). Buffer B consists of 20% deuterated acetic acid. Buffer C consists of 0.1% (v/v) aqueous trifluoroacetic acid. Buffer D consists of 88% acetonitrile, 2% 2-propanol, 9.9% water, and 0.1% trifluoroacetic acid.

Oligonucleotide-Directed Mutagenesis. The mature Lys-IGF-1 coding region was cloned into the phage vector M13mp19 and transformed into *Escherichia coli* strain JM101 for growth and preparation of single-stranded template DNA. Site-directed mutagenesis was performed using the Amersham oligonucleotide-directed in vitro mutagenesis system, version 2. Briefly, 5 μg of template and 2 pM kinased mutagenic primer were annealed by heating to 70 °C, slow-cooled to 37 °C, and placed on ice. Synthesis and ligation of mutant DNA were done at 15 °C overnight with Klenow fragment, T4 ligase, and a dNTP mix containing dCTPS instead of dCTP. Single-stranded DNA was removed by nitrocellulose filtration. The nonmutant strand of DNA was removed by nicking with *Nci*I and digestion with *Exo*III. Double-stranded closed circular DNA was then resynthesized at 15 °C overnight using DNA polymerase I and T4 ligase. This mixture was transformed into competent *E. coli* JM101 cells and plated on LB plates. Mutant phage were selected by hybridization of nitrocellulose replica filters with the appropriate ^{32}P-labeled mutagenic primer. Single-stranded phage DNA was prepared from positive plaques, and the sequence was verified using an Applied Biosystems automated DNA sequencer. Mutants requiring more than one change were modified by sequential rounds of mutagenesis or, where possible, by using two primers at once. All oligonucleotide sequences correspond to the sense strand of Lys-IGF-1.

Purification of Recombinant IGF-1 Analogues. Analogues were overexpressed and purified from *E. coli* as previously described (Miller et al., 1993). The oxidation–reduction status of one- and two-disulfide analogues was assessed by reactivity with iodoacetamide and by rp-HPLC (oxidized and reduced species exhibit different elution times). Formation of the 18–61 disulfide bond was verified in each case by observation of characteristic long-range NOEs in its immediate environment (see Results). The purified proteins were also characterized by SDS–PAGE, amino acid analysis, and amino-terminal sequence analysis. Proteins were dialyzed into 10 mM acetic acid, lyophilized, and stored until use.

HPLC. Reverse-phase HPLC separations were performed on a Vydac 214TP54 (The Separations Group, Hesperia, CA) column of butyl-derivatized end-capped, 5-μm silica. Base buffers for forming gradients were buffer C and buffer D. A Waters 600 gradient pump, 715 temperature-controlled autosampler, 481 UV monitor, and 740 computerized data acquisition system were also used as described (Miller et al., 1993).

Binding Assay. Relative binding affinities of IGF-1 analogues for the type I IGF-1 receptor (Ullrich et al., 1986)

Table I: Expression of IGF-1 Analogues[a]

analogue	substitution	level of expression	relative receptor binding[b] (%)
native IGF-1	none	++	100
[6–48; 18–61]	C47A, C52A	++	0.03
[18–61; 47–52]	C6S, C48S	++	0.3
[6–48; 47–52]	C18A, C61A	–	ND[c]
[18–61]	C6S, C47A, C48S, C52A	+	0.01
[47–52]	C6S, C18A, C48S, C61A	–	ND
[6–48]	C18A, C47A, C52A, C61A	–	ND
des-Cys-IGF-1	C6S, C18A, C47A, C48S, C52A, C61A	±	<0.001

[a] IGF-1 and analogues were expressed as fusion proteins as described (Miller et al., 1993). [b] Receptor-binding affinities are relative to native IGF-1, whose dissociation constant under these conditions is 0.11 nM. [c] ND, not determined.

were determined by radioreceptor assay using human placental membranes and radiolabeled recombinant human IGF-1 as described (Miller et al., 1993). The type 2 IGF receptor (Morgan et al., 1987) does not contribute to IGF-1 binding in this assay.

Circular Dichroism. CD spectra of the IGF-1 analogues were determined in PBS at ambient temperature, using a Jasco J-500C spectropolarimeter. A cuvette with a path length of 1 cm was used in the near-UV region (320–240 nm), while a path length of 0.02 cm was used in the far-UV region (190–250 nm). The data are expressed as molar ellipticity, assuming a mean residue weight of 109. Protein concentrations were determined from the OD_{280}, assuming an extinction coefficient of 0.62 cm^{-1} for a 1% protein solution.

^{1}H-NMR Spectroscopy. Spectra were obtained at 25 and 37 °C at 500 MHz at Harvard Medical School. Two-dimensional experiments were performed in 20% deuterated acetic acid (Hua et al., 1991) and in PBS by the pure-phase method. ^{1}H-NMR assignments of native IGF-1 under these conditions are similar to those described by Sato et al. (1992) in 10% deuterated acetic acid and by Cooke et al. (1991) in aqueous solution at pH 3 and 50 °C.

Fluorescence Spectra. Spectra were determined using an Aminco SLM spectrofluorometer (cuvette path length 0.5 cm). Proteins diluted in PBS to an OD_{280} of 0.14 were excited at several wavelengths between 268 and 280 nm; emission spectra from 280 to 370 nm were recorded with a 5-nm slit width.

Guanidine Stability. Aliquots of protein in PBS were mixed with PBS and PBS-buffered 8 M GdnHCl to obtain solutions of 0.25 mg/mL protein and 0–7 M GdnHCl. The far-UV spectrum from 240 to 205 nm at each GdnHCl concentration was immediately recorded.

RESULTS

Overview of Experimental Design. To construct IGF-1 analogues containing zero, one, or two disulfide bonds, pairwise Cys→Ala and Cys→Ser substitutions were introduced by site-directed mutagenesis. Choice of alanine or serine was based on the solvent exposure of the cystine in IGF-1 (Cooke et al., 1991) or insulin (Baker et al., 1988). Internal disulfide bonds [47–52] and [18–61] were accordingly replaced by alanines, and the partially exposed [6–48] disulfide bond was replaced by serines. In each case only native disulfide bonds were removed (Figure 1).

Levels of expression differed for each construct as summarized in Table I. A general prerequisite for overexpression was the presence of Cys18 and Cys61. Because these cysteines are presumably reduced within the bacterium, the structural basis for this observation is not clear. Variability of expression restricted the present study to four analogues: [6–48; 18–61], [18–61; 47–52], [18–61], and des-Cys-IGF-1. The latter provides a model of reduced IGF-1 with enhanced solubility and stability to oxidation; its expression was poor but not

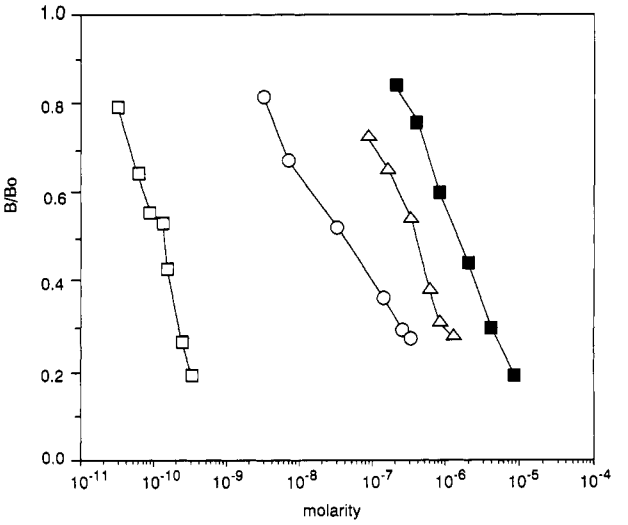

FIGURE 3: Dissociation constants (K_d) for specific binding of IGF-1 analogues to the type I IGF receptor measured by a radiodisplacement assay: native IGF-1 (□); [18–61; 47–52] (○); [6–48; 18–61] (△); [18–61] (■). No detectable binding was observed for the des-Cys analogue. The analogue concentration is shown along the horizontal axis (log scale); the fractional displacement of bound ^{125}I-labeled native IGF-1 is shown along the vertical axis.

negligible. The [6–48; 18–61] and [18–61] analogues correspond to acid-quenched protein-folding intermediates (Hober et al., 1992; Miller et al., 1993); the relationship of the [18–61; 47–52] analogue to the IGF-1-folding pathway is not presently clear.

IGF Receptor Binding. Competition displacement assays indicate that successive removal of native disulfide bonds by mutagenesis results in progressive loss of function (Figure 3 and Table I). Within the limits of the assay, no type I IGF receptor binding was exhibited by des-Cys-IGF-1. Of the two-disulfide analogues, [18–61; 47–52] is 10-fold more active than [6–48; 18–61]. The latter is 30-fold more active than [18–61]. Retention of specific receptor-binding activity by the one- and two-disulfide analogues provides indirect evidence for partial folding of corresponding receptor-binding surface(s).

Solution Structure. Probes of solution structure are provided by circular dichroism, tyrosine fluorescence, and ^{1}H-NMR spectroscopy, as described in turn.

(i) Circular Dichroism. Secondary structure was monitored by far-UV CD (Figure 4A). The spectrum of native IGF-1 (solid line) indicates an α-helix content (Greenfield & Fasman, 1969) of approximately 30–35% in accord with ^{1}H-NMR studies (Cooke et al., 1991; Sato et al., 1992). The α-helix contents of the one- and two-disulfide analogues are substantial (20–30%) but less than that of native IGF-1. No ordered secondary structure is observed in the des-Cys analogue in PBS (pH 7.4), in native IGF-1 in 7 M guanidine hydrochloride

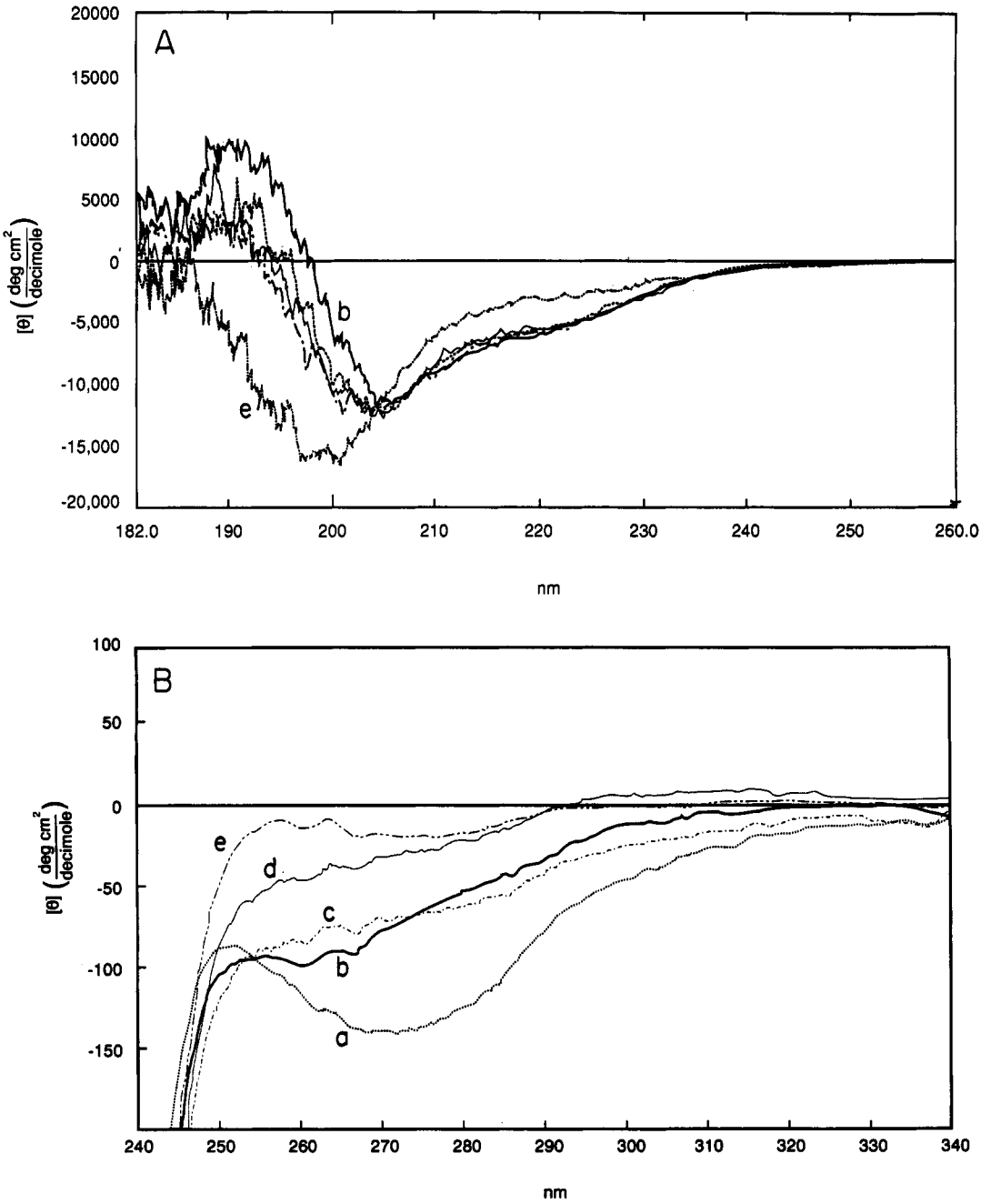

FIGURE 4: Far-UV (A) and near-UV (B) CD spectra of IGF-1 analogues. Panel A: native IGF-1 (b; thick solid line); [6–48; 18–61] (dotted line); [18–61] (thin solid line); [18–61; 47–52] (– • –); des-Cys IGF-1 (e; – •• –). Panel B: native IGF-1 (b; thick solid line); [18–61; 47–52] (c; – • –); [6–48; 18–61] (a; dotted line); [18–61] (d; thin line); des-Cys-IGF-1 (e; – – – –).

(pH 7.4), or in reduced IGF-1 solubilized in 0.01 N HCl (pH 2.0).

Of complementary interest is a comparison of near-UV CD spectra (Figure 4B). In this region ellipticity is due predominantly to disulfide bonds; additional contributions arise from aromatic rings. The near-UV spectrum of des-Cys-IGF-1 is essentially featureless and similar to that of native IGF-1 in 7 M guanidine hydrochloride (not shown). The spectra of the two-disulfide analogues are each distinct (despite having the same number of disulfide bonds), presumably due to structural differences. Contributions of individual chromophores cannot presently be resolved.

(ii) Tyrosine Fluorescence. IGF-1 contains three tyrosines (Tyr24, Tyr31, and Tyr60; Figure 1) and no tryptophan. Although the peak fluorescence emission wavelength of tyrosine (unlike that of tryptophan) is not sensitive to environment, its fluorescent intensity reflects the extent of solvation. In native IGF-1, Tyr31 is fully exposed, Tyr24 is partially exposed, and Tyr60 is buried (Figure 2), as monitored by photochemical dynamic nuclear polarization (photo-CIDNP; Miller et al., 1993). Such photo-CIDNP data predict that the average fluorescence intensity of the fully folded protein would be about 50% of that of the fully unfolded protein. This is indeed observed: the fluorescence emission intensity of native IGF-1 (at 304 nm) is 45–55% of that of des-Cys-IGF-1 (spectra a and e in Figure 5; see also Table II). The latter is similar to that of free tyrosine, adjusted to be equimolar in tyrosine (not shown). The fluorescence emission intensities of the one- and two-disulfide analogues are intermediate between those of native IGF-1 and des-Cys-IGF-1 (Figure 5 and Table II), suggesting intermediate states of protein organization. Whereas all three tyrosines are solvated in des-Cys-IGF-1, one or more tyrosine rings are progressively shielded in the order [18–61] < [6–48; 18–61], [18–61; 47–52] < native IGF-1.

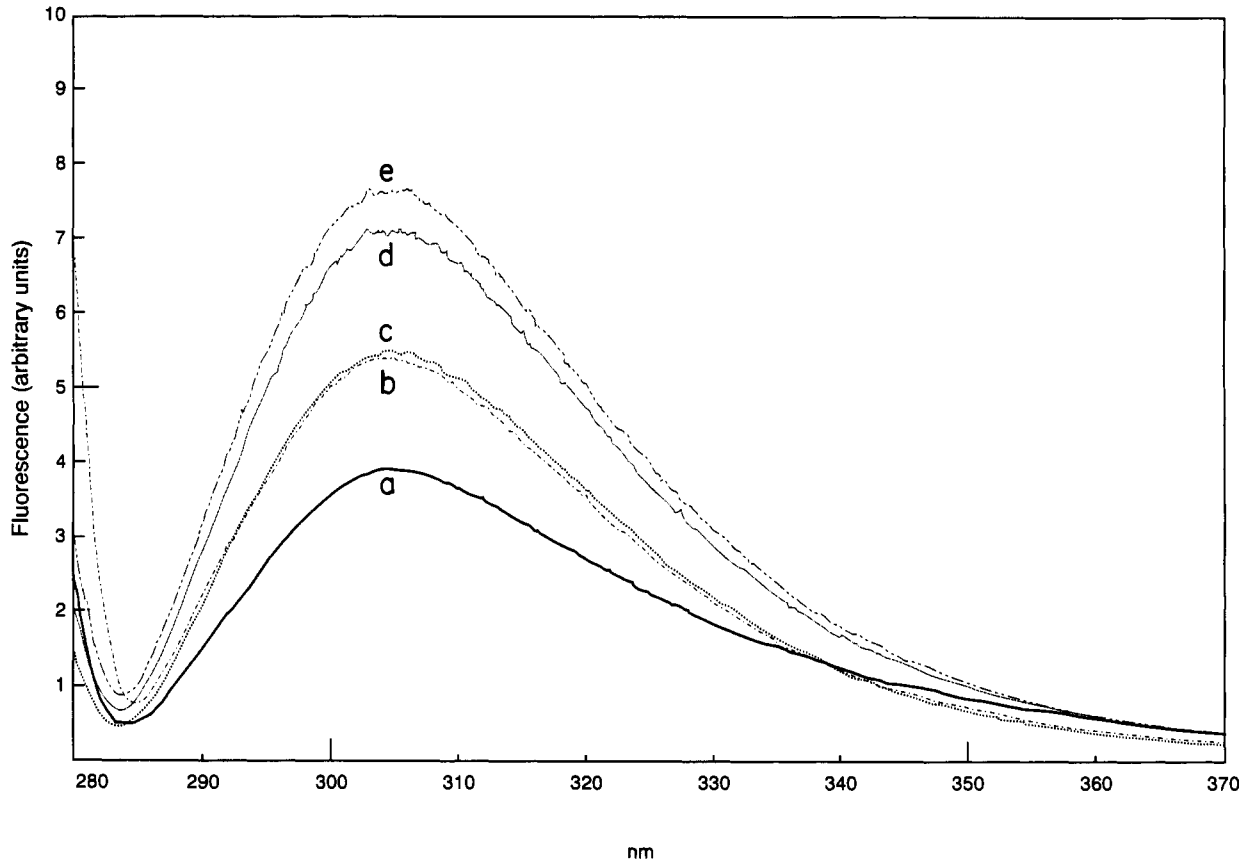

FIGURE 5: Tyrosine fluorescence emission spectra in PBS. In order of increasing emission intensity at 304 nm, the spectra are native IGF-1 (a, solid line), [18–61; 47–52] (b, dash–dot–dash), [6–48; 18–61] (c, dotted line), [18–61] (d, thin solid line), and des-Cys-IGF-1 (e, – ·· –). The emission spectrum of free tyrosine at an equivalent residue concentration is similar to spectrum e. The excitation wavelength was 275 nm, protein concentration 10 μM, temperature 25 °C, and slit width 5 nm.

Table II: Summary of Fluorescence and Protein Denaturation Studies

analogue	fluorescence intensity ratio[a] (%)	guanidine midpoints[b] (M)
native IGF-1	51	4.8
[6–48; 18–61]	71	3.6
[18–61; 47–52]	70	2.2
[18–61]	93	2.2
des-Cys-IGF-1	100	

[a] Fluorescence intensity ratio (±2%) is defined as the ratio of the tyrosine fluorescence intensity at 304 nm of the analogue to that of des-Cys-IGF-1, which is unfolded. [b] Guanidine midpoints are defined at that concentration of guanidine hydrochloride which in CD experiments induces 50% reduction in mean residue ellipticity at 222 nm.

The most significant contribution to quenching of tyrosine fluorescence in the native state presumably arises from desolvation of Tyr60. In IGF-1 (Cooke et al., 1991) and insulin (Baker et al., 1988; Hua et al., 1991), this side chain participates in long-range nonpolar contacts (Figure 2); its fluorescent intensity would thus be expected to provide a sensitive probe for tertiary structure. The wide range of tyrosine fluorescent intensities among analogues contrasts with the similarities of their far-UV CD spectra. Such discordance between probes of tertiary and secondary structure suggests that the one- and two-disulfide analogues have "molten" features, i.e., compact partially folded states (Kawajima et al., 1976; Brazhnikov et al., 1985; Dolgikh et al., 1985; Ptitsyn, 1987; Baum et al., 1989).

(iii) ^{1}H-NMR Spectroscopy. Chemical shift dispersion in ^{1}H-NMR spectroscopy provides an overall measure of the degree to which individual proton environments differ from their average values in a random coil. Spectra were obtained in PBS and in 20% deuterated acetic acid. PBS has the advantage of allowing direct comparison with the previous optical studies but incurs the disadvantage of partial protein self-association; in addition, amide line broadening and rapid base-catalyzed amide proton exchange preclude sequential assignment. In contrast, IGF-1 and its analogues are each monomeric in 20% deuterated acetic acid, and amide resonances are readily observed. Corresponding features are observed in the two solvents as previously described among insulin analogues (Weiss et al., 1991). Resonance assignment of native IGF-1 has recently been described in 10% deuterated acetic acid (Sato et al., 1992). Comparative ^{1}H-NMR studies in PBS provide a control for possible solvent perturbations under such conditions.

In Figures 6 and 7 are shown one-dimensional ^{1}H-NMR spectra in each solvent of native IGF-1 (A), IGF-swap (B), the [18–61; 47–52] analogue (C), the [6–48; 18–61] analogue (D), and the [18–61] analogue (E). In either solvent, the two three-disulfide forms (native IGF-1 and IGF-swap) exhibit marked dispersion of chemical shifts as previously described (Miller et al., 1993). Such dispersion is especially marked among methyl resonances, whose secondary shifts are sensitive to aromatic ring-current fields. Progressive loss of dispersion is seen with removal of one- and two-disulfide bonds. Consistent with optical studies, however, residual structure is apparent in ^{1}H-NMR spectra of the one- and two-disulfide analogues. These spectra have not been completely assigned due to severe resonance overlap in the $d_{\alpha N}$ NOESY "fingerprint" region. The data permit rigorous assignment of unique spin systems (such as Ile43, the only isoleucine) and provisional assignment of spin systems whose chemical shifts and interresidue NOEs are similar to those of native IGF-1.

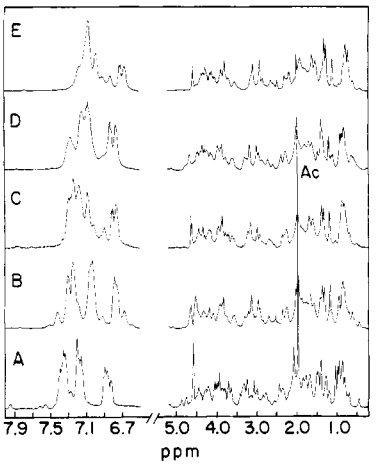

FIGURE 6: 500-MHz ^{1}H-NMR spectra in PBS (99.98% D_2O) at 25 °C of native IGF-1 (A), IGF-swap (B), the [18–61; 47–52] analogue (C), the [6–48; 18–61] analogue (D), and the [18–61] analogue (E). Ac indicates methyl resonance of acetic acid.

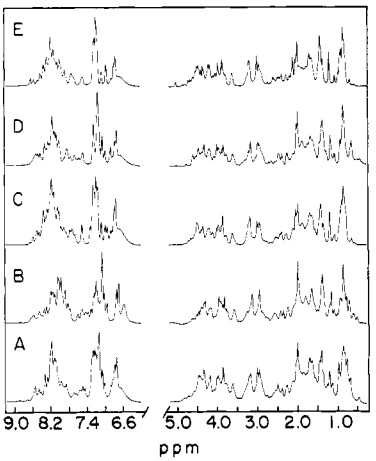

FIGURE 7: 500-MHz ^{1}H-NMR spectra in 20% deuterated acetic acid/80% H_2O at 25 °C of native IGF-1 (A), IGF-swap (B), the [18–61; 47–52] analogue (C), the [6–48; 18–61] analogue (D), and the [18–61] analogue (E).

Analysis of aromatic contacts permits localization of nascent structure in the analogues. In Figure 8 are shown the aromatic spin systems of native IGF-1 (A), [18–61; 47–52] (B), [6–48; 18–61] (C), and [18–61] (D). Provisional assignments are as indicated. In each case, the putative aromatic resonances of Phe23 and Tyr60 exhibit significant secondary shifts, presumably due to stabilization of structure in the neighborhood of the [18–61] disulfide bond. In native IGF-1, these aromatic rings make key tertiary interactions (Figure 2; Cooke et al., 1991), which are observed as long-range NOEs (Figure 9A). Of particular interest are contacts between Phe23 and Leu14 (a β-strand/α-helix contact; Cooke et al., 1991), between Tyr60 and Ile43 (an α helix/helix contact), and between Tyr60 and Leu14 (an α-helix/helix contact). Additional long-range NOEs are observed between Tyr24 and Ala62 (dashed line in the figure). Corresponding NOESY spectra of the [6–48; 18–61] and [18–61] analogues are shown in panels B and C, respectively, of Figure 9. Progressive spectral simplication is seen, reflecting a reduction in chemical shift dispersion and number of NOE cross-peaks. Remarkably, however, analogous Phe23- and Tyr60-related NOEs (Phe23–Leu14, Tyr60–Ile43, Tyr60–Leu14, and Tyr24–Ala62) are observed in each case. These data suggest that formation of the [18–61] disulfide bond is sufficient to stabilize a portion of the hydrophobic core, helix 1, and adjoining β-turn (residues 19–23). Comparable NOESY spectra of the [6–48; 18–61] analogue were not obtained due to limited purification yield. ^{1}H-NMR spectra of des-Cys-IGF-1 were also not obtained.

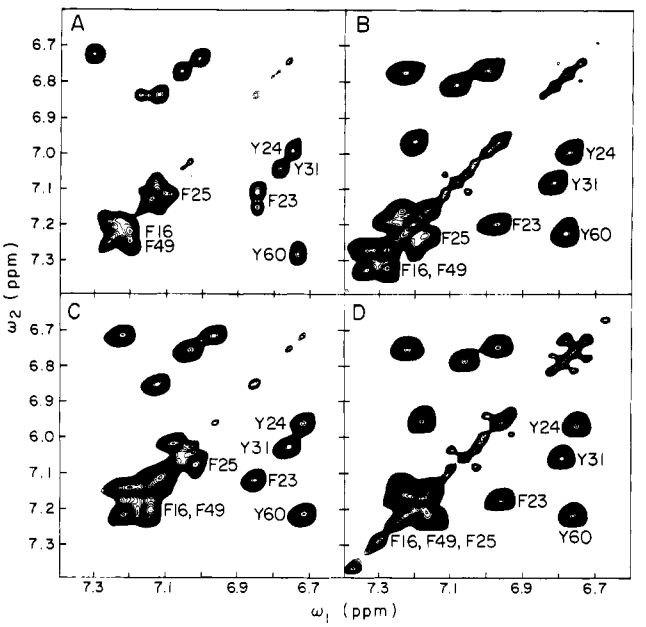

FIGURE 8: Aromatic regions of TOCSY spectra in 20% deuterated acetic acid/80% D_2O at 25 °C: (A) native IGF-1; (B) [18–61; 47–52]; (C) [6–48; 18–61]; (D) [18–61]. The TOCSY mixing time was 30 ms. Assignment of native IGF-1 is as described (Sato et al., 1992); presumptive assignments in analogues are as shown.

Protein Stability. Thermodynamic stabilities of IGF-1 analogues relative to their respective unfolded oxidized states were evaluated by protein denaturation in GdnHCl (as monitored by CD at 222 nm). Native IGF-1 exhibits a progressive unfolding transition over a broad range of GdnHCl concentrations (Miller et al., 1993). This transition is not highly cooperative, and so it is not valid to estimate ΔG_u by the two-state approximation. Nevertheless, the unfolding midpoint (in molar guanidine hydrochloride) provides a convenient marker for qualitative comparison of analogues (Table II). Whereas the unfolding midpoint of native IGF-1 is 4.8 M, the midpoints of the [18–61] and [18–61; 47–52] analogue are each 2.2 M. The unfolding midpoint of the [6–48; 18–61] analogue is intermediate between [18–61] and native IGF-1. These data do not pertain to thermodynamic stabilities relative to the reduced and unfolded state.

DISCUSSION

The mechanism of protein folding represents a fundamental problem in biochemistry (Kim & Baldwin, 1990) and cell biology (Gething & Sambrook, 1992). Of particular interest are oxidative refolding pathways since "cysteine trapping" provides an experimental probe for isolation of kinetic intermediates (Creighton, 1984; Goldenberg & Creighton, 1984; Weissman & Kim, 1991). Early studies of oxidative refolding of globular proteins indicated that the sequence of a protein uniquely determines its folded state, and, in particular, native disulfide pairing. Oxidative refolding in denaturant solutions yielded scrambled disulfides, directly demonstrating thermodynamic coupling between native folding and disulfide pairing (Anfinsen, 1973).

An influential paradigm has been established in an extensive series of studies of bovine pancreatic trypsin inhibitor (BPTI). Although presently under challenge (Weissman & Kim, 1991, 1992a,b), the BPTI oxidative refolding pathway requires transient formation of both native and nonnative disulfide

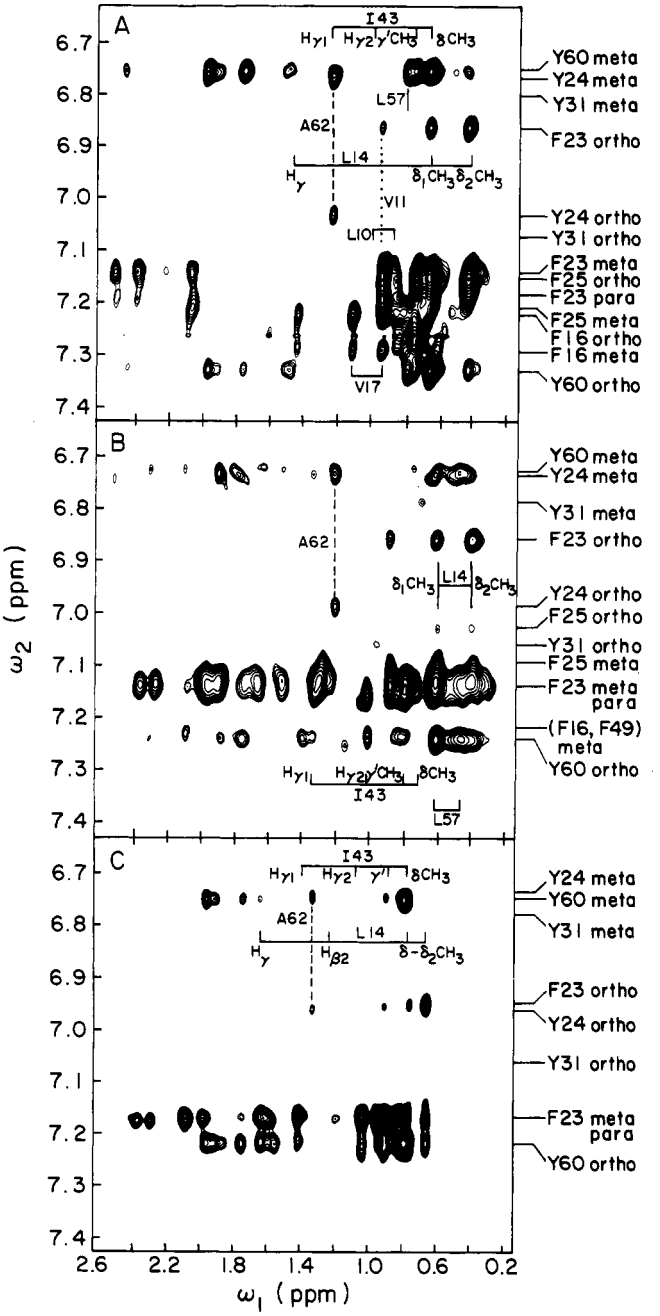

FIGURE 9: Aromatic–aliphatic NOEs in 20% deuterated acetic acid/80% H_2O at 25 °C: (A) native IGF-1; (B) [6–48; 18–61]; (C) [18–61]. The NOESY mixing time was 200 ms. Assignment of NOEs in native IGF-1 is as described (Cooke et al., 1991; Sato et al., 1992); presumptive assignments in analogues are as shown. Ile43 is a unique spin system (see Figure 1).

bonds en route to the final state (Darby et al., 1992). Highly populated intermediates in this pathway exhibit progressive increase in the extent of native structure, but passage through transition states may require partial protein unfolding with transient formation of nonnative interactions. A one-disulfide form that apparently does not contribute to the productive refolding pathway exhibits native structure (Staley & Kim, 1992).

Mature IGF-1 provides an unusual model for studies of protein folding (Miller et al., 1993). Unlike classic "Anfinsen" proteins (Anfinsen, 1973), oxidative refolding of IGF-1 yields two disulfide isomers (Raschdorf et al., 1988; Meng et al., 1989). Such anomalous refolding occurs under a broad range of solution conditions and with accelerated kinetics in the presence of a redox buffer. The two isomers differ by interchange of a strained disulfide bond (Hober et al., 1992). This bifurcating protein-folding pathway is under thermodynamic control (Miller et al., 1993), demonstrating that one polypeptide sequence can encode two distinct ground states. The extent of structural difference between the two isomers has not been defined in detail. Molecular modeling studies, based on the solution structure of native IGF-1 (Cooke et al., 1991), strongly suggest that the alternative disulfide pairing requires nonlocal changes in structure. Consistent with this suggestion, interconversion between isomers (in the presence of catalytic quantities of a reducing agent; Hober et al., 1992) incurs a large activation energy barrier (12 kcal/mol; Miller et al., 1993). Analogous isomers of human proinsulin (Steiner & Clark, 1968; Chance and R. DiMarchi, personal communication) or IGF-2 (unpublished results) are not seen upon oxidative refolding; disulfide pairing in insulin is controlled by sequences in the A- and B-chains (Wang et al., 1986; Tang et al., 1988). IGF-1 sequence elements that control pathway "branching" have not been identified, but presumably involve residues that differ between IGF-1, IGF-2, and proinsulin. It is not known whether refolding of preproIGF-1 or proIGF-1 would also yield two isomers; specific contributions of pro sequences to protein folding have been observed in other systems (Zhu et al., 1989; Winther & Sorensen, 1991; Weissman & Kim, 1992b). Alternative thermodynamic pathways of protein folding have previously been described that, unlike IGF-1, involve protein self-association (King et al., 1989).

In this paper we have described the engineering and biophysical characterization of genetic models of IGF-1 protein-folding intermediates (Miller et al., 1993). Analogues containing zero, one, or two native disulfide bonds were obtained by pairwise replacement of cysteines by alanine or serine. A similar strategy has previously been used to obtain models of BPTI intermediates (Eigenbrot et al., 1990; Naderi et al., 1991; van Mierlo et al., 1991a,b, 1992; Staley & Kim, 1992). Variability of bacterial expression (Table I) limited biophysical studies to four analogues: (i) des-Cys-IGF-1, which provides a model of the reduced and unfolded state; (ii) [18–61], which provides a model of the major one-disulfide intermediate; (iii) [6–48; 18–61], which provides a model of a major native two-disulfide intermediate; and (iv) [18–61; 47–52], whose relevance to the folding pathway is not clear (Miller et al., 1993). Models containing nonnative disulfide bonds were not constructed, although such analogues are of future interest in relation to formation of IGF-swap.

Our studies strongly suggest that successive formation of native disulfide bonds leads to stepwise stabilization of native structural elements. The des-Cys analogue, like fully reduced IGF-1 under acidic conditions (Miller et al., 1993), appears by CD and fluorescence to be a random coil; these data do not exclude formation of nascent structure in local segments of polypeptide (Dyson et al., 1992a,b; Neri et al., 1992). The one- and two-disulfide analogues exhibit significant secondary structure, as probed by CD, but incomplete tertiary structure, as probed by intrinsic tyrosine fluorescence and ^{1}H-NMR spectroscopy. In both [18–61] and [6–48; 18–61], a native-like microdomain is observed by 2D-NMR in the neighborhood of the 18–61 disulfide bond. This microdomain contains long-range NOEs (Leu14–Tyr60, Leu14–Phe23, and Ile43–Tyr60); in the native state, these contacts define key elements of the hydrophobic core. Their preservation in the analogues requires at least partial stabilization of helix 1 and the adjoining β-turn (Figure 2). Progressive desolvation of the Tyr60 chromophore with formation of native disulfide bonds presumably accounts for successive attenuation of tyrosine fluorescence among the analogues (Table II). The one- and two-disulfide analogues

may be viewed as molten globules, i.e., compact partially folded states containing defined elements of secondary structure with fluctuating tertiary orientation (Ptytsin, 1981; Baum et al., 1989; Kuwajima et al., 1991).

The present results provide a foundation for further analysis of genetic models of IGF-1 protein-folding intermediates by multidimensional NMR spectroscopy. Of particular interest will be the structure and dynamics of [18–61], since this intermediate defines a branch point of the IGF-1 pathway (Hober et al., 1992; Miller et al., 1993). Understanding of its nascent structure may enable rational design of mutants that stabilize one isomer or the other. Of complementary interest will be corresponding studies of the proIGF-1-folding pathway. The anomalies of this system may provide unique insight into the informational content of protein sequences.

ACKNOWLEDGMENT

We thank M. Karplus for discussion of protein folding, T. L. Blundell, R. E. Chance, R. DiMarchi, G. G. Dodson, B. H. Frank, and S. E. Shoelson for discussion of insulin chemistry, L. Ward and J. Steele for preparation of the figures, and J. Bennett for preparation of the manuscript.

REFERENCES

Anfinsen, C. B. (1973) *Science 181*, 223–230.
Baker, D., Sohl, J. L., & Agard, D. A. (1992) *Nature 356*, 263–265.
Baker, E. N., Blundell, T. E., Cutfield, G. S., Cutfield, S. M., Dobson, E. J., Dodson, G. G., Hodgkin, D. M. C., Hubbard, R. E., Iassac, M. W., Reynolds, D. C., Sakabe, K. S., Sakabe, N., & Vjayan, N. M. (1988) *Philos. Trans. R. Soc. London 319*, 389–456.
Bashford, D., Chothia, C., & Lesk, A. M. (1987) *J. Mol. Biol. 196*, 199–216.
Baum, J., Dobson, C. M., Evans, P. A., & Hanley, C. (1989) *Biochemistry 28*, 7–13.
Cooke, R. M., Harvey, T. S., & Campbell, I. D. (1991) *Biochemistry 30*, 5484–5491.
Creighton, T. E. (1986) *Methods Enzymol. 131*, 83–106.
Creighton, T. E. (1992) *Nature 356*, 194–195.
Creighton, T. E., & Goldenberg, D. P. (1984) *J. Mol. Biol. 179*, 497–526.
Darby, N. J., van Mierlo, C. P. M., & Creighton, T. E. (1991) *FEBS Lett. 279*, 61–64.
Darby, N. J., van Mierlo, C. P. M., Scott, G. H., Neuhaus, D., & Creighton, T. E. (1992) *J. Mol. Biol. 224*, 905–911.
Dolgikh, D. A., Abaturov, L. V., Bolotina, I. A., Brazhnikov, E. V., Bychkova, V. E., Bushuev, V. N., Gilmanshin, R. I., Semisotnov, G. V., Tiktopulo, E. I., & Ptitsyn, O. B. (1985) *Eur. Biophys. J. 13*, 109.
Dyson, H., Merkuta, G., Waltho, J., Lerner, R., & Wright, P. (1992a) *J. Mol. Biol. 226*, 795–817.
Dyson, H., Sayre, J., Merkuta, G., Shin, H.-C., Lerner, R., & Wright, P. (1992b) *J. Mol. Biol. 226*, 819–835.
Eigenbrot, C., Randal, M., & Kossiakoff, A. (1990) *Protein Eng. 3*, 591–598.
Gething, H. F., & Sambrook, J. (1992) *Nature 335*, 33–45.
Goldenberg, D. P. (1989) *Biochemistry 27*, 2481–2489.
Goldenberg, D. P. (1992) *Trends Biochem. Sci. 17*, 257–261.
Greenfield, M., & Fasman, G. D. (1969) *Biochemistry 8*, 4108–4116.
Hober, S., Forsberg, G., Palm, G., Hartmanis, M., & Nilsson, B. (1992) *Biochemistry 31*, 1749–1756.
Hua, Q. X., Kochoyan, M., Shoelson, S. E., & Weiss, M. A. (1991) *Nature 354*, 238–241.
Jaenicke, R. (1991) *Biochemistry 30*, 3149–3161.
Kim, P. S., & Baldwin, R. L. (1990) *Annu. Rev. Biochem. 59*, 631–660.
King, J., Fane, B., Haase-Pettingell, C., Mitraki, A., Villa-Fane,R., & Yu, M.-H. (1989) in *Protein Folding* (Gierasch, L., & King, J., Eds.) pp 225–240, AAAS, Washington, D.C.
Krebs, H., Schmid, F. X., & Jaenicke, R. (1983) *J. Mol. Biol. 169*, 619–635.
Kuwajima, K., Garvey, E. P., Finn, B. E., Matthews, C. R., & Sugai, S. (1991) *Biochemistry 30*, 7693–7703.
Martin, J., Langer, T., Boteva, R., Schramel, A., Horwich, A. I., & Hartl, F.-U. (1991) *Nature 352*, 36–42.
Meng, H., Burleigh, B. D., & Kelly, G. M. (1989) *J. Chromatogr. 443*, 183–192.
Miller, J. A., Owers-Narhi, L., Hua, Q. X., Rosenfeld, R., Arakawa, T., Rohde, M., Prestrelski, S., Lauren, S., Stoney, K. S., Tsai, L., & Weiss, M. A. (1993) *Biochemistry* (preceding paper in this issue).
Morgan, D., Edman, J., Standring, D., Fried, V., Smith, M., Roth, R., & Rutter, W. (1987) *Nature 329*, 301–307.
Naderi, H. M., Thomason, J. F., Borgias, B. A., Anderson, S., James, T. L., & Kuntz, I. (1991) in *Conformations and Forces in Protein Folding* (Nall, B. T., & Dill, K. A., Eds.) pp 86–114, American Association for the Advancement of Science, Washington, D.C.
Neri, D., Billeter, M., Wider, G., & Wuthrich, K. (1992) *Science 257*, 1559–1563.
Radford, S. E., Dobson, C. M., & Evans, P. A. (1992) *Nature 358*, 302–307.
Raschdorf, F., Dahinden, R., Maerki, W., Richter, W., & Merryweather, J. (1988) *Biomed. Environ. Mass Spectrom. 16*, 3–8.
Rinderknecht, E., & Humbel, R. E. (1978) *J. Biol. Chem. 253*, 2769–2776.
Roder, H., & Wüthrich, K. (1986) *Proteins: Struct., Funct., Genet. 1*, 34–42.
Roder, H., Elvoe, G. A., & Englander, S. W. (1988) *Nature 335*, 700–704.
Sato, A., Nishimura, S., Ohkubo, T., Kyogoku, Y., Koyama, S., Kobayashi, M., Yasuda, T., & Kobayashi, Y. (1992) *J. Biochem. 111*, 529–536.
Staley, J. P., & Kim, P. S. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 1519–1523.
Stassinopoulou, C. I., Wagner, G., & Wüthrich, K. (1984) *Eur. J. Biochem. 145*, 423–430.
States, D. J., Dobson, C. M.; Karplus, M., & Creighton, T. E. (1984) *J. Mol. Biol. 174*, 411–418.
States, D. J., Creighton, T. E., Dobson, C. M., & Karplus, M. (1987) *J. Mol. Biol. 195*, 731–739.
Steiner, D. C., & Clark, J. L. (1968) *Proc. Natl. Acad. Sci. U.S.A. 60*, 622–629.
Tang, J.-G., Wang, C.-C., & Tsou, C.-L. (1988) *Biochemn. J. 255*, 451–455.
Udgaonkar, J. B., & Baldwin, R. L. (1988) *Nature 335*, 694–699.
Ullrich, A., Gray, A., Tam, A., Yang-Feng, T., Tsubokawa, M., Collins, C., Henzel, W., Le Bon, T., Kathoria, S., Chen, E., Jacobs, S., Francke, U., Ramachandran, J., & Fujita-Yamaguchi, Y. (1986) *EMBO J. 5*, 2503–2512.
Van Dyk, T. K., Gatenby, A. A., & Larossa, R. A. (1989) *Nature 342*, 451–453.
Van Mierlo, C. P. M., Darby, N., Neuhaus, D., & Creighton, T. E. (1991a) *J. Mol. Biol. 222*, 353–371.
Van Mierlo, C. P. M., Darby, N., Neuhaus, D., & Creighton, T. E. (1991b) *J. Mol. Biol. 222*, 373–390.
Van Mierlo, C. P. M., Darby, N., & Creighton, T. E. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 6775–6779.
Wang, C.-C., & Tsou, C.-L. (1986) *Biochemistry 25*, 5336–5340.
Weiss, M. A., Hua, Q.-X., Frank, B. H., Lynch, C., & Shoelson, S. E. (1991) *Biochemistry 30*, 7373–7398.
Weissman, J. S., & Kim, P. S. (1991) *Science 253*, 1386–1393.
Weissman, J. S., & Kim, P. S. (1992a) *Proc. Natl. Acad. Sci. U.S.A. 89*, 9330–9334.
Weissman, J. S., & Kim, P. S. (1992b) *Cell 71*, 841–851.
Winther, J. R., & Sorensen, P. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 9330–9334.
Zhu, X. L., Ohta, Y., Jordan, F., & Inouye, M. (1989) *Nature 339*, 483–484.