Oxidative Refolding of Insulin-like Growth Factor 1 Yields Two Products of Similar Thermodynamic Stability: A Bifurcating Protein-Folding Pathway†

James A. Miller,*,‡ Linda Owers Narhi,‡ Qing-Xin Hua,§ Robert Rosenfeld,‡ Tsutomu Arakawa,‡ Michael Rohde,‡ Steve Prestrelski,‡ Scott Lauren,‡ Kendall S. Stoney,‡ Larry Tsai,‡ 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 September 9, 1992; Revised Manuscript Received December 8, 1992

ABSTRACT: Can one protein sequence encode two structures? Oxidative folding of human insulin-like growth factor 1 (IGF-1), a globular protein of 70 residues, is shown to yield two products of similar thermodynamic stability. This observation is of particular interest in light of the recent demonstration that two of the three disulfide bonds in native IGF-1 rearrange in the presence of dithiothreitol [Hober, S., et al. (1992) *Biochemistry 31*, 1749–1756]. Kinetics of the IGF-1 folding pathway were monitored by high-performance liquid chromatography (rp-HPLC). Disulfide-pairing schemes of intermediates and products were established by peptide mapping. Two disulfide isomers were obtained as products: one with native insulin-like pairing [6–48; 18–61; 47–52] (designated native IGF-1; 60% yield) and the other with alternative pairing [6–47; 18–61; 48–52] (designated IGF-swap; 40% yield). The predominant early intermediate contains the single disulfide 18–61, which is shared in common by the two products. Relative yields of native IGF-1 and IGF-swap are independent of protein concentration under dilute conditions. In the absence of an added thiol reagent, each isomer is stable indefinitely at neutral pH; in the presence of an added thiol reagent, the two isomers interconvert with an Arrhenius activation barrier of 12 kcal/mol. Interconversion does not require complete reduction and yields the same ratio of products as initial folding, demonstrating thermodynamic control. Spectroscopic studies using circular dichroism (CD), infrared spectroscopy (FTIR), two-dimensional ^{1}H-NMR (2D-NMR), and photochemical dynamic nuclear polarization (photo-CIDNP) suggest that IGF-1 and IGF-swap adopt similar secondary structures but distinct tertiary folds. Implications of these observations for understanding the topology of protein-folding pathways are discussed.

Protein folding represents a coding problem of broad chemical and biological interest. The concept of a folding code emerged with the demonstration by Anfinsen and co-workers that the three-dimensional structure of a globular protein is uniquely determined by its primary structure (Anfinsen, 1973). Early steps in protein folding are proposed to involve nucleation of elements of secondary structure and stabilization of a compact partially folded state by solvent exclusion (the molten globule; Ptytsin, 1981; Kim & Baldwin, 1982; Kuwajima, 1989; Kuwajima et al., 1991). Support for this model has been provided by biophysical studies of isolated peptides (Dyson et al., 1992a,b) and of equilibrium models of partially folded proteins (Baum et al., 1989), and by ^{1}H nuclear magnetic resonance (^{1}H-NMR)[1] studies of the kinetics of amide proton exchange during refolding (Roder & Wüthrich, 1986; Roder et al., 1988; Udgaankar & Baldwin, 1988; Radford et al., 1992). Biochemical studies of disulfide pairing during reoxidation of bovine pancreatic trypsin inhibitor (BPTI) have revealed a sequence of preferred kinetic intermediates, which define a reaction pathway (Creighton, 1974, 1986; Creighton & Goldenberg, 1984). Reinvestigation of this pathway (Weissman & Kim, 1991) has emphasized the importance of nativelike structural elements in accord with ^{1}H-NMR studies of disulfide-trapped intermediates (States et al., 1984, 1987), genetic models of one- and two-disulfide intermediates (Eigenbrot et al., 1990; Darby et al., 1991; Staley & Kim, 1992; van Mierlo et al., 1992) and corresponding peptides (Oas & Kim, 1988; Staley & Kim, 1990). Restriction of a conformational search to a hierarchy of nativelike subdomains would result in simplification of the protein-folding problem (Kim & Baldwin, 1990). Of complementary interest is the issue of thermodynamic versus kinetic control (Creighton & Goldenberg, 1984; Baker et al., 1992; Creighton, 1992).

In this paper we reexamine a fundamental conclusion of the Anfinsen experiment: that the structure of a globular protein is uniquely determined by its amino acid sequence. Our studies focus on human insulin-like growth factor 1 (IGF-1), a globular protein of 70 residues which is homologous to proinsulin (Rinderknecht & Humbel, 1978) and similar in structure to insulin (Blundell et al., 1978; Cooke et al., 1991; Figure 1). The insulin motif defines an extended family of regulatory polypeptides that are highly conserved among metazoans; these peptides contain three characteristic disulfide bonds ([6–48; 18–61; 47–52] in human IGF-1). The oxidative refolding pathway of IGF-1 is of special interest in view of its anomalous disulfide exchange properties (Hober et al.,

† This work was supported by Amgen Inc. and by a grant from the National Institutes of Health to M.A.W. (DK-40949). M.A.W. is supported in part by the American Diabetes Association and Juvenile Diabetes Foundation International.

* Address correspondence to either author.
‡ Amgen Inc.
§ Harvard Medical School.
‖ Massachusetts General Hospital.

[1] Abbreviations: CD, circular dichroism; FTIR, Fourier-transform infrared; NMR, nuclear magnetic resonance; NOESY, nuclear Overhauser spectroscopy; photo-CIDNP, photochemical dynamic nuclear polarization; rp-HPLC, reverse-phase high-performance liquid chromatography; TOCSY, total correlation spectroscopy; TFA, trifluoroacetic acid; 2D-NMR, two-dimensional NMR; GdnHCl, guanidine hydrochloride. *Nomenclature*: The refolded species with insulin-like disulfide pairing [6–48; 18–61; 47–52] is designated IGF-1; the refolded species with alternative disulfide pairing [6–47; 18–61; 48–52] is designated IGF-swap. Amino acids are designated in the text by three-letter abbreviations and in figures by the standard single-letter code.

0006-2960/93/0432-5203$04.00/0

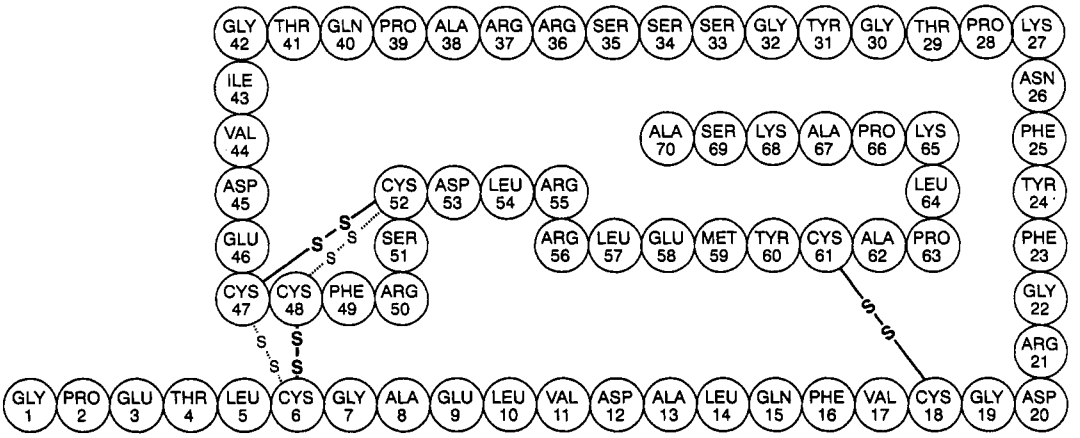

FIGURE 1: Sequence of human IGF-1. Native (insulin-like) disulfide pairing [6–48; 18–61; 47–52] is shown in solid line, IGF-swap pairing is shown in dotted line.

1992). We show that when reduced IGF-1 is refolded under native conditions to allow formation of disulfide bonds, two distinct pairing schemes are observed that differ by interchange ("swap") of the 6–48 and 47–52 disulfide bonds (Raschdorf et al., 1988; Meng et al., 1989). The nonnative product (designated IGF-swap) contains disulfide pairing [6–47; 18–61; 48–52]. Spectroscopic studies suggest that native IGF-1 and IGF-swap contain similar secondary structures with different tertiary orientations. The disulfide bond shared by IGF-1 and IGF-swap (18–61) is shown to be present in the major one-disulfide intermediate. In this pathway, the two products represent alternative thermodynamic ground states rather than kinetic traps (Haase-Pettingell et al., 1988). In the following paper in this issue the IGF-1 folding pathway is further investigated by construction of genetic models of protein-folding intermediates (Narhi et al., 1993). While these manuscripts were in preparation, a related study was published by Nillson and co-workers (Hober et al., 1992).

MATERIALS AND METHODS

Protein Chemistry

Buffers. PBS consists of 150 mM NaCl–10 mM sodium phosphate (pH 7.4). Buffer B consists of 50 mM potassium phosphate (pH 7.4). Buffer C consists of 20% acetic acid–sodium acetate (pH 3.0). Buffer D consists of 20% acetic acid. Buffer E consists of 0.1% (v/v) aqueous trifluoroacetic acid. Buffer F consists of 88% acetonitrile, 2% 2-propanol, 9.9% water, and 0.1% trifluoroacetic acid (all v/v). Buffer G is 0.1% diisopropylethylamine and 0.1% formate in 2% acetonitrile–water (v/v). Buffer H is 90% acetonitrile–water. Buffer I is 90% acetonitrile, 0.1% TFA in water (v/v). Buffer J is 80% acetonitrile, 0.1% TFA in water (v/v).

Protein Purification. Recombinant human IGF-1 was produced by overexpression of a synthetic gene in *Escherichia coli* (Elliott et al., 1990). Resultant inclusion bodies were solubilized in urea, oxidized by stirring with exposure to atmospheric oxygen, and purified by ion-exchange and preparative rp-HPLC chromatography. This protocol yielded IGF-1 and IGF-swap as resolved products (>98% purity as judged by SDS–PAGE and rp-HPLC). Amino acid compositions and N-terminal sequence analyses were as predicted. Proteins were dialyzed into 10 mM acetic acid and lyophilized for storage until use. Protein concentrations were determined from the optical density at 280 nm, assuming an extinction coefficient of 0.62 cm^{-1} for a 0.1% protein solution.

Proteolysis and Peptide Mapping. IGF-1 and IGF-swap were digested with S. endoproteinase Glu-C (Boehringer Mannheim) following the protocol of Canova-Davis et al. (1991). The HPLC column and the initial conditions used were the same as for the pepsin digest (see below). The gradient was a mixture of buffer G and buffer I, starting with an 18-min linear gradient to 21% solvent I, followed by a 17-min linear gradient to 23% solvent I and then a 15-min linear gradient to 30% solvent I. The eluent was monitored by diode array detector, so that peptides containing only phenylalanine could readily be detected. Peptides of interest were subjected to N-terminal sequence analysis using an ABI 477A protein sequencer.

IGF-1 and IGF-swap were also characterized by pepsin digestion. Various forms of IGF-1 conformers in 1% TFA (v/v) (Aldrich) were diluted 10-fold with 0.02 N HCl, and then 1 μg of pepsin (Boehringer Mannheim) was added at a weight ratio of 1 to 15. The samples were incubated at 37 °C for 24 h and subjected to mass spectral analysis and peptide mapping. Mass spectral analyses were carried out using a capillary liquid chromatogram (Michrom BioResources Ultrafast Microprotein Analyzer, Pleasanton, CA) linked to an ionspray interface of a triple quadrupole mass spectrometer (SCIEX API III, Toronto, Canada). Peptides were separated on a Michrom BioResources C-18 column (0.5 × 150 mm) packed with Reliasil 5-μm, 300-Å material. The gradient was a mixture of buffers G and H; a 30-min gradient from 2 to 60% buffer H provided suitable resolution. Mass spectra were analyzed using MacSpec version 3.0 software (SCIEX).

Peptides from pepsin digests were separated by rp-HPLC. Reverse-phase separations were performed on a Vydac 218TP54 (0.46 × 20 cm, 300 Å, 5-μm particle size) column on a Hewlett Packard 1090 liquid chromatography system and HP 79994A analytical workstation. The flow rate was 0.75 mL/min, and the initial conditions were 5 min at 100% solvent E. The gradient was a 35-min linear gradient to 42% solvent J. Peptides of interest were subjected to N-terminal sequence analysis.

Characterization of IGF-1 Folding Intermediates. Three protein-folding intermediates were isolated by rp-HPLC following acid-quenching of refolding reactions at successive time points (Table I). Peak I_d was reconstituted into 0.25 M Tris–1 mM EDTA, pH 8.5, with simultaneous addition of neat 4-vinylpyridine (Sigma), and then allowed to react for 15 min in the dark. Peak I_c was reconstituted into an ice-cold buffer of 0.15 M Tris–1 mM EDTA, pH 8.5, containing neat 4-vinylpyridine. The mixture was reacted in the dark at 0 °C for 60 min. The pyridylethylated peak I_d and peak I_c were purified by rp-HPLC and subjected to N-terminal sequence analysis.

Experimental Protocols

Disulfide Reassortment Experiments. Purified, lyophilized IGF-1 or IGF-swap was dissolved in either degassed 20 mM Hepes (pH 8.0 at 20 °C) or 50 mM sodium phosphate (pH 7.5 at 20 °C)–1 mM EDTA at a protein concentration of 100 μg/mL (1.3×10^{-5} M). The solution was then equilibrated to the desired temperature in a Waters 715 autosampler. To initiate the disulfide-exchange reaction, a freshly made stock solution of 0.02% (v/v) 2-mercaptoethanol was added to the protein solution to obtain 0.001% (v/v) 2-mercaptoethanol (0.17 mM). The molar ratio of 2-mercaptoethanol to IGF was hence 13:1. Reactions were sampled by programming the autosampler to inject an aliquot onto the HPLC column at the appropriate time and eluted as above. Reactions at 50 °C were monitored by removing aliquots from a sample held in a temperature-regulated water bath, manual quenching with 10 volumes of 0.1% TFA (v/v), and injection onto the HPLC column.

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 E and buffer F. A Waters 600 gradient pump, 715 temperature-controlled autosampler, 481 UV monitor, and 740 computerized data acquisition system were also used. A Hewlett Packard 1040 diode array detector was used to acquire and store chromatograms.

Initial rates of IGF-1→IGF-swap or IGF-swap→IGF-1 were estimated by integration of chromatograms in order to calculate the amount of IGF-1 or IGF-swap formed at the time in question. HPLC chromatograms were integrated in reference to a previously determined standard curve generated for IGF-1 at 215 nm. To calculate an initial rate, only the earliest chromatogram in each experiment was used in these estimates (in most cases, the amount of new species formed was less than 10% of the total amount of protein present).

HPLC Analysis of the Folding Pathway. One-half milligram of lyophilized purified IGF-1 was dissolved in 1 mL of 10 mM Tris (pH 8.0 at 20 °C), 4 M urea, 0.1 M sodium chloride, and 50 mM DTT. After 15 min at room temperature, 500 μL of this mixture was applied to a PD10 prepacked gel filtration column (Pharmacia, Piscataway, NJ) which had been previously equilibrated in degassed 50 mM sodium phosphate, pH 7.5. The column was eluted in 0.5-mL steps, collecting 0.5-mL fractions. The fraction containing the protein was found by monitoring the UV absorbance at 280 nm; the protein concentration was approximately 150 μg/mL; 0.5 mL of the sample was made 0.01% (w/v) in sodium azide and kept loosely capped at room temperature. At given times, a 20-μL aliquot was removed and diluted into 200 μL of 0.1% TFA in an HPLC minivial insert (Waters Associates, Milford, MA). The acid-quenched aliquots were stored at room temperature until all time points were collected, at which time they were analyzed by rp-HPLC. An identical procedure was used to follow the refolding of the reduced molecule in 8 M GdnHCl, except that the PD10 column was equilibrated and eluted with 50 mM sodium phosphate, pH 7.5, with 8 M GdnHCl.

For monitoring air-catalyzed folding of the fully reduced protein, the Vydac column was initially equilibrated in 80% buffer E and 20% buffer F at a flow rate of 1 mL/min (see above). After sample injection, the following gradients were employed: (i) a linear gradient to 30% buffer F over 3 min; (ii) a linear gradient to 38% F over the next 27 min; (iii) a linear gradient to 80% F over the next 20 min; (iv) wash and reequilibration cycles. For monitoring thiol-catalyzed rearrangement of isolated IGF-1 or IGF-swap, a shorter gradient was used: (i) 80% buffer E–20% buffer F at 1 mL/min; (ii) a 3-min linear gradient to 30% F; (iii) a 27-min linear gradient to 38% F, which resolved the various species; and (iv) a linear gradient to 100% F to wash the column, followed by reequilibration in 80% E–20% F.

IGF-1 Receptor Binding Assays. Relative affinities of IGF-1 and IGF-swap to type I receptors were performed using a competition assay with ^{125}I-IGF-1 and human placental membrane fragments (D'Ecrole et al., 1976).

Biophysical Methods

Circular Dichroism. IGF-1 and IGF-swap were analyzed in PBS or 20% acetic acid using a Jasco J-500C spectropolarimeter. A cuvette with a path length of 1 cm was used in the near-UV region (240–340 nm); a cuvette with a path length of 0.02 cm was used in the far-UV region (190–250 nm). The data are expressed as mean residue ellipticity, on the basis of a mean residue weight of 109. Percent α-helix was estimated by the method of Greenfield and Fasman (1969).

Fluorescence. Spectra were obtained using an Aminco SLM Model SPF-500 spectrofluorometer. A cuvette with a 0.5-cm path length was used. The proteins were excited at 275 nm; emission spectra from 280 to 370 nm were recorded, with slit widths set to provide 5-nm resolution.

GdnHCl Stability. To determine the relative stabilities of IGF-1 and IGF-swap, stock solutions of 1–2 mg/mL protein in PBS were mixed with buffered GdnHCl to final GdnHCl concentrations of 0–7 M in PBS. Far-UV CD spectra (205–240 nm) were obtained at each GdnHCl concentration. Ellipticity (in millidegrees) at 222 or 215 nm was plotted versus concentration of GdnHCl.

Gel Filtration. Aliquots of IGF-1 and IGF-swap in PBS were injected onto a TSK 2000 SW HPLC gel filtration column equilibrated in the same buffer. The protein was eluted using a flow rate of 0.5 mL/min and was monitored by the UV absorbance at 280 nm. Standard proteins (BSA, ovalbumin, RNase, aprotinin, insulin) were used to calibrate the column, and the apparent molecular weight of the IGF-1 species was determined from its retention time.

^{1}H-NMR Spectroscopy. Spectra were obtained at 25 and 37 °C at 500 MHz at the Francis Bitter National Magnetic Laboratory and at Harvard Medical School. Two-dimensional experiments were performed by the pure-phase method (States et al., 1982). ^{1}H-NMR assignments of native IGF-1 at pH 3 and 50 °C have been reported by Cooke et al. (1991) and in 10% deuterated acetic acid by Sato et al. (1992). Spectra of IGF-1 and IGF-swap were otherwise obtained in PBS (pH 7.4) and in 20% deuterated acetic acid, and are similar under the three sets of conditions. Twenty percent acetic acid, in which insulin is natively folded (Weiss et al., 1989; Hua et al., 1991), permits partial sequential assignment, which is readily extended to PBS by analogy (Q.-X. Hua and M. A. Weiss, unpublished results).

Infrared Spectroscopy. IGF-1 and IGF-swap were prepared for infrared spectroscopy by dialyzing against pure water and lyophilizing. The lyophilized powders were dissolved in 20 mM imidazole in D_2O (Sigma Chemical Co., 99.9% isotopic purity). The pD of the buffer was 7.5, estimated by adding 0.4 to the direct pH reading. The protein concentration was 3% (w/v). The proteins were incubated in the D_2O buffer for 48 h to allow complete exchange of amide protons, as assayed by disappearance of the IR amide A band (N–H stretching, ca. 3300 cm^{-1}) and H_N ^{1}H-NMR resonances. Solutions were

placed in IR cells with 100-μm Teflon spacers and CaF_2 windows. IR spectra were collected on a Nicolet 800 FTIR spectrometer equipped with a Globar source, a GE-coated KBr beamsplitter, and a DTGS detector. For each spectrum, 512 double-sided interferograms were co-added and Fourier-transformed after application of a Happ–Genzel apodization function. Contributions from buffer components and residual H_2O vapor were subtracted using Nicolet FTIR software. Fourier self-deconvolution was performed by the method of Kaupinnen et al. (1981). Curve-fitting was done using the program PeakFit (Jandel Scientific Co.).

Photo-CIDNP. Riboflavin (0.4 mM) was added to a 1 mM solution of IGF-1 or IGF-swap in buffer B (99.9% D_2O). The light source was an Innova-6 continuous argon laser at 488 nm (Coherent, Inc., Palo Also, CA); timing of the laser pulse was gated by the NMR pulse program as previously described (Weiss et al., 1989, 1990). Dark spectra are the sum of 256 scans; 4 scans were obtained following laser irradiation.

RESULTS

Our results are presented in two parts. HPLC analyses of acid-quenched refolding reactions and disulfide reassortment experiments are presented in part I. Biophysical and biochemical characterization of IGF-1 and an alternatively refolded form are presented in part II.

(I) IGF-1 Purification and Refolding

Isolation of IGF-1 Isomers. The protein was expressed in *E. coli* and purified to homogeneity, as assayed by SDS–PAGE, sequence analysis, amino acid composition, and rp-HPLC (Figure 2B). Oxidative refolding of the reduced polypeptide results in two major rp-HPLC peaks and several minor ones (Figure 2A). As described below, the two major peaks have identical linear sequence but differ in disulfide pairing. The peak which elutes first is designated IGF-swap, and the second peak (identical to the naturally-occurring molecule) is designated IGF-1. Similar results are obtained under a broad range of solution conditions (pH 7–10 and [NaCl] 0–1 M) and in the presence of a redox buffer (oxidized and reduced glutathione). The two isomers are readily isolated (Figure 2B,C) and stored either in solution (pH 2.0–8.0) or as lyophilized powders. In the absence of added thiol reagents, the isolated proteins are stable indefinitely under these conditions.

Characterization of Disulfide Bonds. The disulfide-pairing schemes of IGF-1 and IGF-swap were established by S. endopeptidase Glu-C and pepsin digest mapping (see Materials and Methods) and found to be [6–48; 18–61; 47–52] and [6–47; 18–61; 48–52], respectively. Disulfide pairings present in predominant folding intermediates were also determined by proteolytic mapping and mass spectrometry as given in Table I. The most highly populated early intermediate (designated I_h) contains the single disulfide bond 18–61, which is shared in common between IGF-1 and IGF-swap.

Disulfide Rearrangement Experiments. Solutions of either IGF-1 or IGF-swap were found to be unstable in the presence of even small, substoichiometric amounts of thiol reagents such as cysteamine, 2-mercaptoethanol, or dithiothreitol; molar ratios of –SH to protein between 2:1 and 60:1 produced the same qualitative results (data not shown). Such reagents induced reequilibration into a distribution of species whose HPLC profile strongly resembles that observed in the course of initial refolding of the fully reduced and unfolded polypeptide (Figure 3A, 25 h, compared to Figure 4A, 68 h). Principal components were IGF-1 and IGF-swap, present in a ratio that was a function of temperature (Table II) and salt concentration (not shown) but independent of protein concentration. Disulfide rearrangement was observed at thiol concentrations low enough to prevent total reduction of the protein. The ratio of products was independent of reductant concentration and starting component; that is, IGF-1 rearranged into a similar ratio of components as did IGF-swap under the same conditions (Table II). This ratio is also similar to that obtained upon initial oxidative refolding. These observations demonstrate that IGF-1 and IGF-swap represent alternative thermodynamic ground states rather than kinetic traps.

The rate of rearrangement was sufficiently slow to permit monitoring of reaction kinetics by rp-HPLC (Figure 3A). Depending on temperature and thiol concentration, an equilibrium between IGF-1 and IGF-swap was reached within 3–36 h. Control experiments verified that acid-quenched aliquots reassorted no further for at least 24 h after quenching. Initial rates of formation of the alternative isomer are given in Table II. The temperature dependence of these rates was used to construct Arrhenius plots for IGF-1→IGF-swap and IGF-swap→IGF-1 (not shown). An activation energy (E_a) was calculated from these plots and found to be approximately

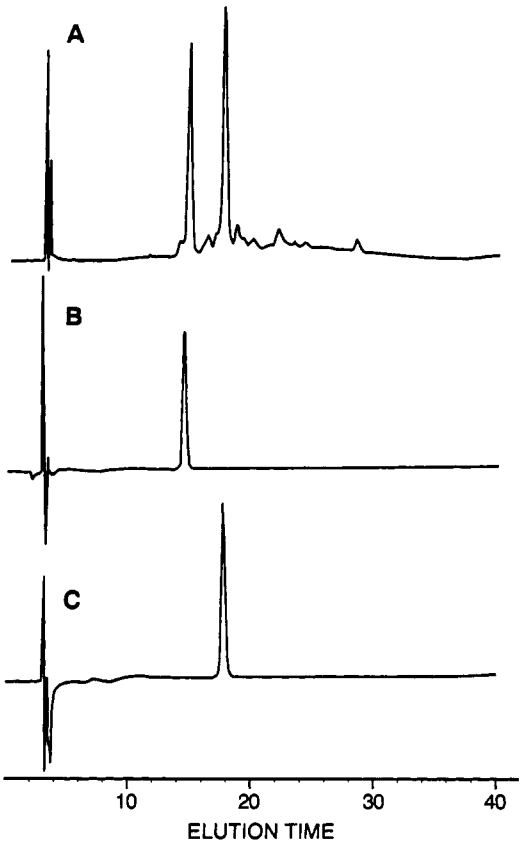

FIGURE 2: (A) rp-HPLC chromatogram of refolding reaction mixture after 36 h of air oxidation. (B) rp-HPLC chromatogram of purified IGF-swap (I_a). (C) rp-HPLC chromatogram of purified IGF-I (I_b).

Table I: Disulfide Structures of Intermediates and Products[a]

peak	disulfide bridges	unbridged cysteines
I_a ("IGF-swap")	6–47, 18–61, 48–52	
I_b (native IGF-1)	6–48, 18,61, 47–52	
I_c	6–48, 18–61	47, 52
I_d	6–47, 18–61	48, 52
I_h	18–61	6, 47, 48, 52

[a] Peaks correspond to elution position in rp-HPLC (Figure 4A).

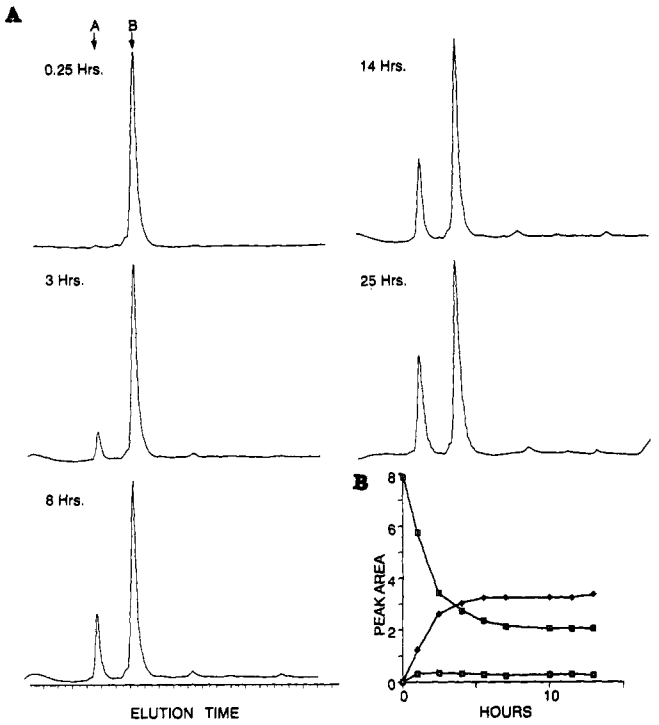

FIGURE 3: (A) rp-HPLC chromatogram of a reaction mixture in which pure IGF-I, of purity comparable to that shown in Figure 2C, was mixed with 0.001% (v/v) β-ME and incubated at 25 °C. At selected times after addition of β-ME, aliquots of the reaction were injected onto the reverse-phase column and chromatographed. IGF-swap and peak I_d form as the major equilibrium products aside from native IGF-I. (B) Plot of the time course of a typical β-ME-induced rearrangement, in this case of pure IGF-swap in 0.001% β-ME rearranging at 25 °C. IGF-swap peak area (⧈); IGF-I peak area (♦); I_d peak area (□).

12 kcal/mol in either direction.

A small temperature dependence of the IGF-swap⇔IGF-1 equilibrium constant was observed (Table II) with IGF-1 the more favored product at elevated temperatures. At or below 10 °C, the reactions failed to reach equilibrium within the time course of the studies. The plot of $\log K_{eq}$ vs $1/T$ produced a line whose slope indicated an endothermic enthalpy (ΔH) difference of approximately 2 kcal/mol for the IGF-swap→IGF-1 reaction (not shown). Since this is larger than the overall difference in Gibbs free energy (<0.25 kcal/mol in magnitude), formation of IGF-swap from native IGF-1 must be enthalpically driven.

Although the composition of the rearrangement reactions is dominated by IGF-1 and IGF-swap, minor components are also observed. These elute at higher concentrations of acetonitrile (Figure 3A). The principal minor component (3–5% of the total protein concentration) is I_d, the kinetics of whose formation are too rapid to monitor by rp-HPLC (Figure 3B). It was possible to isolate sufficient quantities of several of these intermediates and immediately quench free sulfhydryl groups using 4-vinylpyridine. Pepsin digestion and mapping of the subsequent peptides were performed in order to establish the location of disulfide bonds. These are listed in Table I. I_a and I_b were found to have the expected configuration. I_d, a metastable product, is a species with nonnative disulfide pairing [6–47; 18–61], leaving cysteines-48 and -52 free. Evidently, there is a kinetic barrier to its further oxidation to IGF-swap. The two predominant transient species contain native disulfide bonds: I_h ([18–61]) and I_c ([6–48; 18–61]).

Reassortment intermediates are also observed as kinetic intermediates in oxidative refolding reactions (below); presumably, the mechanism of reassortment recapitulates late stages of the protein folding pathway. Reassortment proceeds without detectable accumulation of the fully reduced species and is more rapid than initial oxidative refolding. These observations suggest that the IGF-1 and IGF-swap refolding pathways share a common intermediate, i.e., that folding from the fully reduced state can proceed via a branched pathway or pathways.

IGF-1 Oxidative Refolding Pathway. IGF-1 was reduced in 50 mM DTT and Tris buffer (pH 8.0 at 20 °C) in the presence of 6 M urea. rp-HPLC of this mixture demonstrated that the polypeptide eluted as a single species (designated I_r) at a high acetonitrile concentration (Figure 4A, 0 h). After reduction, the protein was quickly exchanged into 50 mM sodium phosphate buffer (pH 7.5) by gel filtration and allowed to refold. Experiments were also conducted at different pH (in the range 7–10) and NaCl concentrations (0–1 M); results were similar in each case.

The time course of appearance and disappearance of HPLC peaks is shown graphically in Figure 4B,C. Within 1 h, I_r had decreased and a partially resolved complex of peaks designated I_x formed. These are presumably partially folded intermediates (Figure 4A, 1 h). By 8 h (Figure 4A, 8 h), additional intermediates were observed, denoted I_c, I_e, I_f, and I_h (Table I). Also becoming apparent at 8 h were the three species which eventually dominate the refolding reaction, I_a (IGF-swap), I_b (IGF-1), and I_d.

After 24 h (Figure 4A, 24 h), I_r and I_x were gone, and intermediates I_{e-h} had decreased roughly in parallel. At this time point, I_c dominated the reaction mixture, and I_a, I_b, and I_d were also present in large amounts. This trend continued until at 53 h (Figure 4A, 53 h), when I_a and I_b completely dominated the composition of the refolding mixture such that $I_b > I_a > I_d$ and I_c was detectable only as a faint shoulder on I_b. Except for some small broad peaks (which may be disulfide-linked multimers), little else was present. The starting material I_r and intermediates I_x and I_{e-h} were not detectable.

The time course of the appearance and disappearance of folding intermediates is shown in Figure 4B. I_e (♦), I_f (□), and I_h (◇) appear in parallel and are at their maxima roughly 5 h after the initiation of refolding. I_c (⧈), begins to appear in significant quantities at 5 h and reaches a maximum at 20 h, after which time it slowly decreases. In Figure 4C is plotted the appearance of metastable or stable product peaks with time. I_d (□) appears quickly and reaches a steady state within 10 h, whereas the major products, I_a (⧈) and I_b (♦), are slower to appear. A quantitative model of precursor–product relationships has not been developed. An analogous refolding pathway is observed on a time scale of minutes (rather than hours) in the presence of 10 mM oxidized glutathione and 1 mM reduced glutathione as previously described (Hober et al., 1992).

Reoxidation under Denaturing Conditions. The 6 cysteines in the IGF-1 sequence can give rise to 75 possible species, including 15 isomers containing 3 disulfide bonds. Reoxidation in 8 M GdnHCl provides a control experiment to verify the expected thermodynamic coupling between protein folding and the nonrandom selection of disulfide bonds under native conditions. In contrast to refolding in physiologic buffers, a large number of products were observed; neither native IGF-1 nor IGF-swap predominate (data not shown). This result is in accord with the original studies of Anfinsen (1973) on refolding in denaturant solutions.

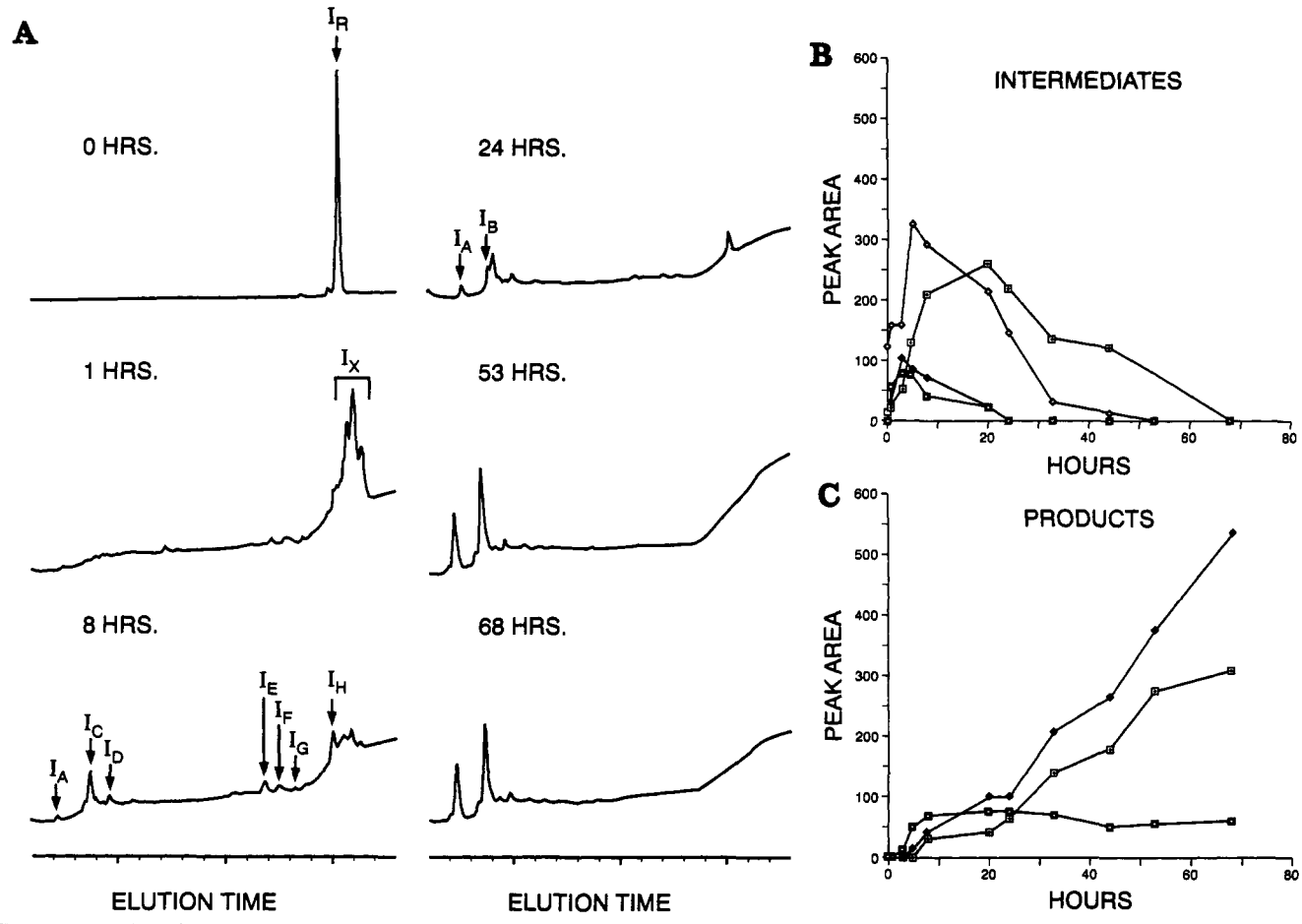

FIGURE 4: (A) Refolding of fully reduced insulin-like growth factor in phosphate buffer, pH 7.5, catalyzed by exposure to air at 25 °C, monitored by rp-HPLC at selected times. Labeled peaks are described in the text. The gradient used is longer than that in Figures 2 and 3 (see Materials and Methods). (B) Plot of peak areas of intermediate folding states versus time: (⊡) I_c (♦) I_e; (■) I_f; (◊) I_h. (C) Plot of peak areas versus time for late-forming peaks in refolding reaction: (⊡) I_a; (♦) I_b; (□) I_d. I_d contains two disulfide bonds ([6–47; 18–61]; Table I) and two free cysteines and is thus metastable under present conditions.

Table II: Reaction Rates and Equilibrium Ratios

temp (°C)	rate $I_a \rightarrow I_b$ (mol s^{-1})	rate $I_b \rightarrow I_a$ (mol s^{-1})	rate $I_b \rightarrow I_a$/ rate $I_a \rightarrow I_b$	equilibrium ratio of IGF-swap:IGF-1 for $I_a \rightarrow I_b$ reaction	equilibrium ratio of IGF-swap:IGF-1 for $I_b \rightarrow I_a$ reaction
5	1.3×10^{-10}	ND	NA	NA	NA
10	2.5×10^{-10}	7.9×10^{-11}	0.32	NA	NA
12	1.9×10^{-10}	1.1×10^{-10}	0.58	0.70	0.60
18	4.7×10^{-10}	2.6×10^{-10}	0.54	0.75	0.57
25	5.4×10^{-10}	3.4×10^{-10}	0.63	0.62	0.57
40	1.6×10^{-9}	7.4×10^{-10}	0.46	0.52	0.49
50	3.5×10^{-9}	1.5×10^{-9}	0.43	0.41	0.40

(II) Comparative Biochemical and Biophysical Studies

IGF Type I Receptor Binding Assays. Relative affinities of IGF-1 and IGF-swap for human placental type I IGF-1 receptor were determined by a radioreceptor assay. The type I receptor (Ullrich et al., 1986) is the principal binding site observed in this assay; the type 2 receptor (Morgan et al., 1987) has substantially lower affinity for IGF-1 and does not contribute significantly to observed binding under these conditions. Native IGF-1 exhibits an apparent K_d of 1×10^{-10} M, and IGF-swap exhibits an apparent K_d of 1×10^{-9} M. Although the latter is reduced 10-fold, retention of high-affinity receptor binding strongly suggests that the disulfide isomers share at least one protein surface. A corresponding "swapped" isomer of human insulin exhibits similar bioactivity relative to native insulin (Sieber et al., 1978). Receptor-catalyzed disulfide reassortment has not been excluded.

Oligomeric States. Gel filtration chromatography in PBS demonstrates that both IGF-1 and IGF-swap have identical Stokes radii at protein concentrations of 0.5 mg/mL, eluting with an apparent molecular weight of 14 000. This value is consistent with formation of a dimer (data not shown). Dimerization is weak in each case, as determined by equilibrium ultracentrifugation ($K_D > 50$ μM; D. Yphantis, personal communication). The protein surfaces responsible for dimerization have not been identified in either case. The two isomers are monomeric and stably folded in 20% acetic acid (see below).

Infrared Spectroscopy. FTIR spectra of IGF-1 and IGF-swap in the amide I (predominantly C=O stretching, 1700–1620 cm^{-1}) region are shown in Figure 5. Deconvolution into component subspectra is shown in each panel in dashed line; positions and relative intensities of component bands are related to elements of secondary structure as described (Byler & Susi, 1986; Surewicz & Mantsch, 1988). Peak positions, relative intensities, and secondary structure assignments are given in Table III. These data indicate that the secondary structures of IGF-1 and IGF-swap are similar: each contains extended structure (~45%), α-helix (~40%), and reverse turn (~15%). Although these similarities suggest an overall correspondence of elements of secondary structure, spectral differences are also observed. The most significant difference occurs in the component band near 1662 cm^{-1} (reverse turns; see Table III). The intensity of this component is greater in the spectrum of IGF-swap than in that of native IGF-1. Conversely, the

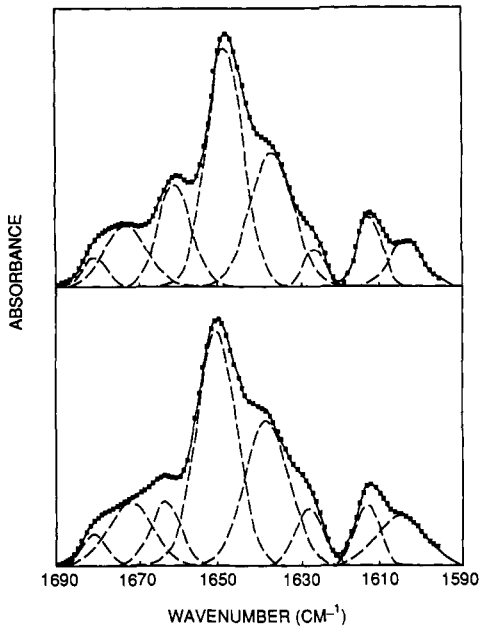

FIGURE 5: Deconvoluted spectra of IGF-1 isomers in the amide I region. Upper panel, IGF-swap (I_a); lower panel, native IGF-1 (I_b). (Each square represents one data point.) Dashed lines: individual Guassian components. The deconvolution was carried out using values of 25 cm^{-1} and 3 for the undeconvoluted half-width and resolution-enhancement factor, respectively. A Lorentzian line shape was assumed, and a Bessel apodization function was used. Two bands are fitted below 1620 cm^{-1} which are not amide I components, but are included in order to preclude the necessity of adding a sloping base-line parameter. Estimates of secondary structure contributions in each case are given in Table III.

Table III: Peak Positions, Relative Intensities, and Secondary Structure Assignments for Amide I Component Bands in IR Spectra of IGF-1 Peak I_a and I_b Fractions

IGF-1 ν (cm^{-1})	area (%)	IGF-swap ν (cm^{-1})	area (%)	assignment
1628	7	1626	4	extended strands
1638	28	1637	27	extended strands
1651	41	1648	40	α-helices
1663	8	1661	15	reverse turns
1672	12	1673	11	extended strands
1681	3	1681	3	reverse turns

sum of the intensities of components assigned to extended structure is lower in IGF-swap.

CD Spectroscopy. Far-UV CD provides a complementary probe for secondary structure and is especially useful for estimating α-helical content. The two disulfide isomers exhibit similar CD spectra in this region (Figure 6A). The α-helix content, as calculated by the Greenfield–Fasman equation (Greenfield & Fasman, 1969), is approximately 30–35% for native IGF-1; the α-helix content of IGF-swap is somewhat less, in accord with the results of Hober et al. (1992). More marked differences are observed in the near-UV region (Figure 6B), most likely reflecting different disulfide configurations and environments. Far- and near-UV CD spectra of IGF-1 and IGF-swap in 20% acetic acid were also determined, and are identical to those described above.

Spectra of reduced IGF-1 could not be obtained in PBS due to its limited solubility and rapid oxidation under these conditions. Such spectra are readily obtained in 10 mM HCl, however, and are essentially without evidence of ordered structure (Figure 6A). These data strongly suggest that disulfide bond formation is coupled to stabilization of ordered structure. This hypothesis is supported by analysis of IGF-1 analogues containing pairwise Cys→Ala or Cys→Ser substitutions (Narhi et al., 1993).

Thermodynamic Stability. The relative thermodynamic stabilities of IGF-1 and IGF-swap (relative to their respective denatured but oxidized states) were evaluated by CD studies of GdnHCl denaturation (Figure 6C,D). In each case, complex unfolding is observed that is not readily modeled by a single two-state cooperative transition. Nevertheless, the two disulfide isomers exhibit similar unfolding curves as a function of GdnHCl concentration. In each case, ellipticity at a helix-sensitive wavelength (222 nm) is 50% attenuated at approximately 4.4 M GdnHCl. In contrast, a mutant IGF-1 with native disulfide pairing (Phe23→Gly) exhibits 50% attenuation at 2.5 M GdnHCl (unpublished results), demonstrating the sensitivity of this assay to substitutions in the hydrophobic core (as defined by 2D-NMR; Cooke et al., 1991). Evidence that the thermodynamic stabilities of IGF-1 and IGF-swap are similar relative to the reduced and unfolded states is presented above.

Fluorescence Spectroscopy. Fluorescence spectra of IGF-1 and IGF-swap (excitation wavelength 280 nm) provide a probe of tertiary structure as shown in Figure 7. Due to the absence of tryptophan in IGF-1, fluorescence is due to tyrosine (emission maximum at 304 nm). Although the emission wavelength of tyrosine is not sensitive to environment, its fluorescent intensity is influenced by the degree of solvent exposure. The fluorescent tyrosine(s) in IGF-1 and IGF-swap appear to be in similar average environments, as the spectra of IGF-swap resulting from excitation at several wavelengths between 268 and 282 nm are identical to those of IGF-1 excited at the same set of wavelengths.

^{1}H-NMR Studies. One-dimensional NMR spectra of IGF-1 and IGF-swap in PBS are shown in Figure 8. The proteins are predominantly dimeric under these conditions. The spectra each exhibit a range of chemical shifts characteristic of a folded protein but otherwise display widespread differences in secondary shifts. Such differences provide evidence of a global change in tertiary structure. Nevertheless, the following three similarities are noteworthy. (i) The two spectra contain a similar pattern of sharp resonances, indicating that no gross difference exists in the extent of ordered or disordered substructure. (ii) In each spectrum, the aromatic resonances of the three tyrosines are inequivalent (panels C and D; resonances a–c and a′–c′), and those of Tyr60 (see below) exhibit large secondary shifts. (iii) In each spectrum leucine methyl resonances are shifted upfield. These are assigned in the case of IGF-1 to Leu14 (Cooke et al., 1991; Sato et al., 1992); the same assignment is obtained by sequential assignment of IGF-swap in 20% acetic acid (unpublished results). Interestingly, the homologous leucine in insulin (LeuB15) is also shifted upfield in 20% acetic acid (Hua & Weiss, 1991) and in PBS (Weiss et al., 1991). Its secondary shift is ascribed to the ring-current of PheB24 (Weiss et al., 1991; Kristensen et al., 1991), which is in close proximity to LeuB15 in crystal structures (Baker et al., 1988) and in solution (Hua et al., 1991).

Corresponding regions of NOESY spectra of IGF-1 and IGF-swap are shown in panels A and B, respectively, of Figure 9. This region contains NOEs between aromatic resonances (ω_2; vertical scale) and aliphatic resonances (ω_1; horizontal scale). On a qualitative level, the two spectra are comparably "crowded", indicating the presence of multiple interresidue contacts in each case. In native IGF, specific NOEs are observed between Leu14 and Tyr60 (cross-peaks l and p; arrows) and between Leu14 and Phe23 (cross-peaks n and o);

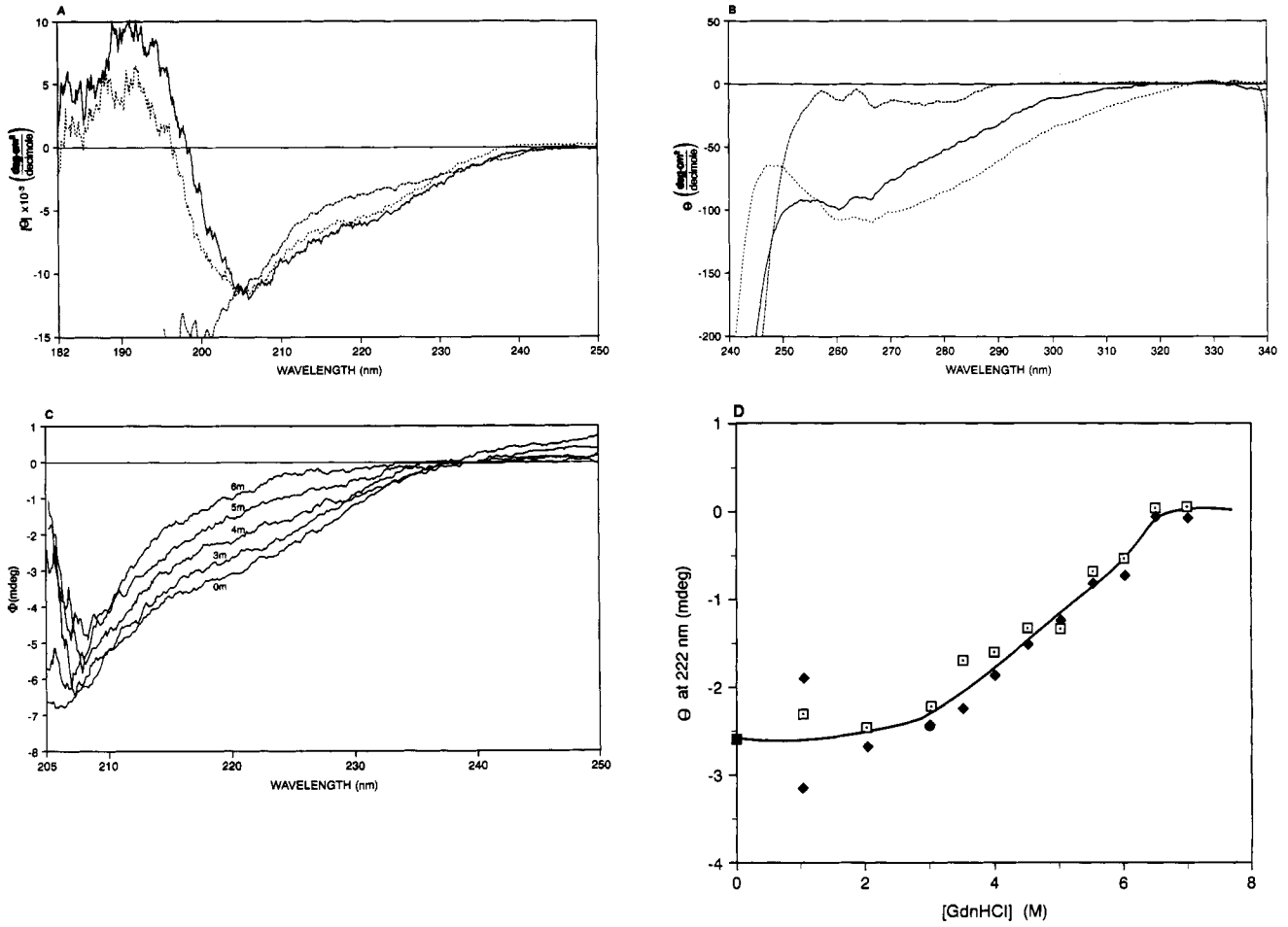

FIGURE 6: (A) Far-UV CD spectra of reduced IGF-1 in 10 mM HCl (labeled r; dashed line), IGF-swap in PBS (dotted line), and native IGF-1 in PBS (solid line). (B) Corresponding near-UV CD spectra demonstrate alteration in disulfide environments. (C) Representative guanidine hydrochloride denaturation experiment; the far-UV CD spectrum of native IGF-1 is shown at successive concentrations of guanidine hydrochloride (0, 1, 2, 3, 4, 5, 6, 7, and 8 M). Data obtained at wavelengths below 210 nm deteriorate at high concentrations of GdnHCl due to its absorbance. (D) Dependence of the observed ellipticity at a structure-sensitive wavelength $[(\theta)_{222}]$ on guanidine hydrochloride concentration: IGF-swap (□); native IGF-1 (♦). Experimental conditions: 100 μM protein concentration in phosphate-buffered saline and 25 °C.

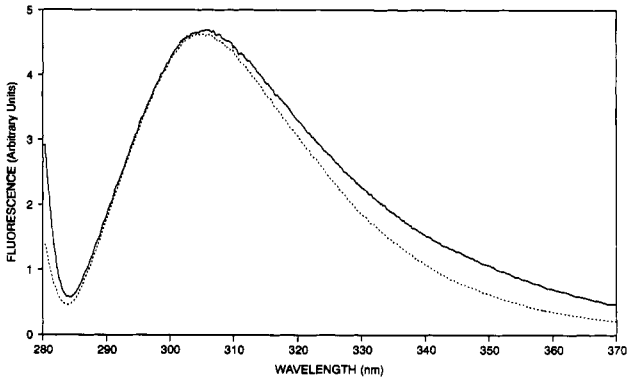

FIGURE 7: Fluorescence emission spectra of IGF-swap (dotted curve) and native IGF-1 (solid curve) in PBS. The excitation wavelength was 275 nm, the protein concentration was 10 μM, the temperature was 25 °C, and the slit width was 5 nm.

these data are in accord with the results of Cooke et al. (1991). In IGF-swap (panel B), analogous NOEs are observed between Leu14 and Tyr60 (cross-peaks l′ and p′) and Phe23 (n′, o′, and o″). Detailed differences are seen in relative NOE intensities. Analogous similarities and differences are observed under monomeric conditions in 20% acetic acid (data not shown). Tyr60 is adjacent to Cys61, which forms an identical disulfide bond (18–61) in the two isomers. These data suggest that the tertiary structures of IGF-1 and IGF-swap are similar in the neighborhood of their shared disulfide bridge.

Photo-CIDNP. Laser excitation of a triplet-state probe (riboflavin) enables the differential accessibility of tyrosine residues in IGF-1 and IGF-swap to be evaluated by ^{1}H-NMR spectroscopy (Kaptein, 1980; Muzkat et al., 1984); the protein contains no histidine or tryptophan residues, which may also exhibit photo-CIDNP enhancement. Photo-CIDNP spectra of IGF-1 and IGF-swap are shown in Figure 10. In each case, the difference spectrum (spectra C and C′) shows one tyrosine with significant enhancement (resonances a and a′), one tyrosine with partial enhancement (resonances b and b′), and one unenhanced tyrosine (resonances c and c′). In IGF-1, resonances a, b, and c are assigned to the meta ($H_{3,5}$) protons of Tyr31, Tyr24, and Tyr60, respectively (Cooke et al., 1991; Sato et al., 1992). A similar pattern is observed in IGF-swap, and analogous assignment of a′, b′, and c′ is suggested by sequential assignment in 20% acetic acid. The inaccessibility of Tyr60 in both IGF-1 and IGF-swap provides additional evidence for similar tertiary structrue in the immediate neighborhood of the shared [18–61] disulfide.

These results may be compared to previous photo-CIDNP studies of insulin (Muzkat et al., 1984; Weiss et al., 1989) and proinsulin (Weiss et al., 1990). The homologous residues in insulin are PheB25 (Tyr24) and TyrA19 (Tyr60); there is no aromatic residue in proinsulin analogous to Tyr31 in IGF-1. Like Tyr60 in IGF-1, TyrA19 is not enhanced in the photo-

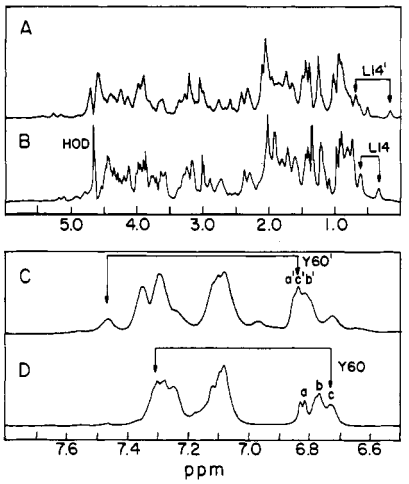

FIGURE 8: Aliphatic region of 500-MHz ^{1}H-NMR spectra of IGF-swap (panel A) and native IGF-1 (panel B); corresponding aromatic spectra are shown in panels C (IGF-swap) and D (IGF-1). Each spectrum exhibits dispersion of chemical shifts characteristic of a folded protein. The upfield-shifted methyl resonances of Leu14 are indicated by arrows in panels A and B; the respective aromatic spin systems of Tyr60 are indicated by arrows in panels C and D. Aromatic resonances a, b, and c (and a′, b′, and c′ in IGF-swap) are assigned to Tyr31, Tyr24, and Tyr60, respectively. Spectra were obtained in PBS at 25 °C at a protein concentration of 2 mM.

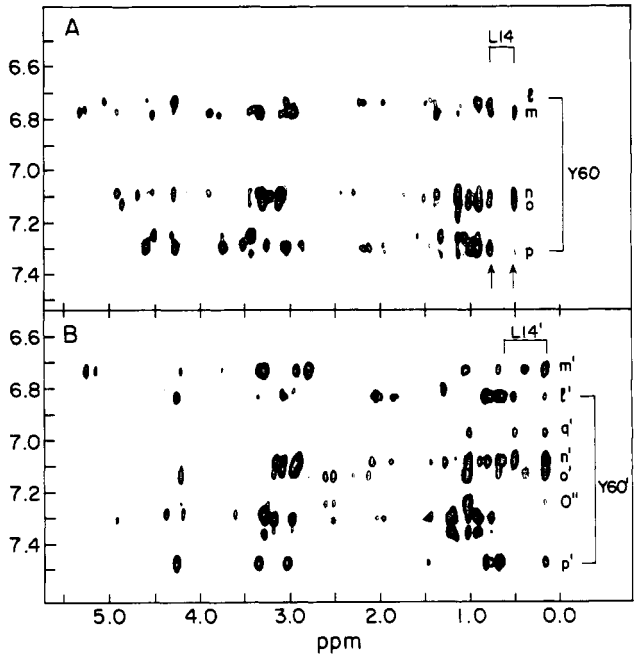

FIGURE 9: Two-dimensional NOESY spectra of native IGF-1 (panel A) and IGF-swap (panel B) in PBS. Each exhibits numerous aromatic–aliphatic contacts characteristic of tertiary structure; these include NOEs between Leu14 and Tyr60. Arrows in panel A indicate NOEs from Tyr60 to Leu14 (cross-peaks l and p). Additional labeled cross-peaks (m and m′ etc.) represent analogous NOEs from corresponding aromatic spin systems in the two spectra. Spectra were obtained under the conditions of Figure 8; the NOESY mixing time was 200 ms.

CIDNP spectrum of insulin or proinsulin. Although the aromatic resonances of PheB25 cannot be enhanced, the neighboring residue TyrB26 exhibits partial enhancement like that assigned to Tyr24 in IGF-1. TyrA14, which like Tyr31 in IGF-1 is exposed on a flexible portion of the protein surface, also exhibits the largest enhancement. There is thus a general similarity between the photo-CIDNP spectra of IGF-1 and proinsulin in accord with the similarity of their structures.

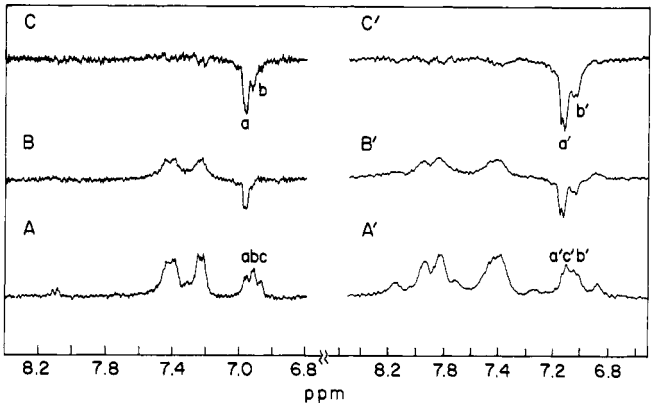

FIGURE 10: Photo-CIDNP enhancement spectra of IGF-1 (panels A, B, and C) and IGF-swap (panels A′, B′, and C′) in PBS. Spectra A and A′ were obtained in the absence of laser irradiation, spectra B and B′ were obtained following laser irradiation, and spectra C and C′ are calculated difference spectra. In each case, resonances a (a′), b (b′), and c (c′) are assigned to the 3,5 ring resonances of Tyr31, Tyr24, and Tyr60 in IGF-1 (IGF-swap), respectively. Tyr31 is fully enhanced, Tyr24 is partially enhanced, and Tyr60 is inaccessible.

DISCUSSION

The insulin-related family of proteins represents an ancestral motif of protein folding (Blundell et al., 1978). This motif is defined in part by three canonical disulfide bonds [6–48; 18–61; 47–52] (Figure 1). The distinct environments of these disulfide bonds are similar in crystal structures of insulin (Baker et al., 1988) and in the solution structure of native IGF-1 (Figure 1; Cooke et al., 1991). Molecular modeling suggests that interchange of any two disulfide bonds cannot be accommodated without nonlocal structural perturbations. Reduced proinsulin refolds to form predominantly the native pairing (A6–A11; A7–B7; A20–B19), and the necessary information is encoded in the sequences of the isolated A- and B-chains (Wang & Tsou, 1986; Tang et al., 1988). The refolding pathway of proinsulin (or any other member of this family) has not previously been characterized.

Unlike reduced proinsulin, reduced IGF-1 refolds in ambient atmosphere to yield two predominant products. One is native IGF-1; the other has the "swapped" disulfide pairing [6–47; 18–61; 48–52] (Raschdorft et al., 1988; Meng et al., 1989). Analogous results are obtained in the presence of a redox buffer (oxidized and reduced glutathione) but with accelerated kinetics (Hober et al., 1992). Each isomer is stable in the absence of an added thiol reagent, and interconverts in its presence (Hober et al., 1992). Kinetic analysis of the interconversion reaction indicates a substantial activation free-energy barrier ($E_a \approx 12$ kcal/mol) between the two forms. Barrier crossing presumably requires partial protein unfolding following reduction of at least one disulfide bond. Although complete reduction is apparently not required, the mechanism of interconversion involves population of one or more reassortment intermediates. The rp-HPLC profile of reassortment intermediates recapitulates that of kinetic folding intermediates. Of particular interest is the 18–61 single-disulfide species. We imagine that this intermediate occupies a branch point in the oxidative refolding pathway. Future biochemical and structural characterization of this intermediate may provide general insight into how the topology of a folding pathway is determined. A genetic model of this intermediate is described in the following paper (Narhi et al., 1993).

The fully reduced IGF-1 polypeptide appears to be a random coil: acquisition of structure is coupled to disulfide bond formation. That the ratio of IGF-1 to IGF-swap is the same

in reassortment reactions as in oxidative refolding demonstrates that the folding pathway is thermodynamically controlled. In particular, the thermodynamic stabilities of IGF-1 and IGF-swap must be similar relative to the reduced and denatured state, ($\Delta\Delta G$ < 0.25 kcal/mol in magnitude). The reaction IGF-swap→IGF-1 is entropically driven whereas the back-reaction IGF-1→IGF-swap is enthalpically driven. GdnHCl denaturation studies demonstrates that IGF-1 and IGF-swap also exhibit similar thermodynamic stabilities relative to their respective denatured and oxidized states. Why is the IGF-1 sequence ambiguous (i.e., cannot discriminate between two competing disulfide pairings)? Although the answer to this question is not known, DiMarchi et al. (1992) have established that isolated IGF-1 peptides corresponding to the insulin A- and B-chains recombine to give analogous isomers, thus localizing sites of ambiguity to the A and B domains.

It will be of considerable interest to compare the three-dimensional structures of IGF-1 and IGF-swap. Is the swap in disulfide bonds associated with a major change in structure or a minor local conformational adjustment (e.g., analogous to cis–trans isomerization of proline)? The answer to this question is of fundamental importance. If IGF-1 and IGF-swap have essentially the same overall structures, then the present studies relate primarily to the phenomonology of disulfide bond formation rather than to general principles of protein folding. On the other hand, if IGF-1 and IGF-swap exhibit global differences in structure, then this system provides a counterexample to the paradigm that an amino acid sequence encodes a unique three-dimensional structure. Here we have shown that IGF-1 and IGF-swap exhibit an overall correspondence in secondary structure, as viewed at low resolution by CD and FTIR. Fluorescence and photo-CIDNP studies demonstrate similar average tyrosine environments. In addition, the retention by IGF-swap of specific dimerization and of 10% receptor binding affinity further suggests that functional protein surfaces are in part retained. *Nevertheless, one-dimensional ^{1}H-NMR spectra of IGF-1 and IGF-swap are as different as two unrelated proteins.* Such differences in chemical shift presumably reflect a reordering of the hydrophobic core and, in particular, reorientation of aromatic ring-current fields. Further analysis of the three-dimensional structure of IGF-swap by multidimensional NMR is in progress and may permit the molecular mechanism of branched refolding to be analyzed.

The present study focuses on in vitro refolding properties of the mature IGF-1 sequence and does not address whether altenatively refolded forms of IGF-1 occur in vivo. Although considered unlikely in this system, a regulated switch between alternative folding pathways could in principle provide an additional biological control mechanism. We emphasize, however, that physiologic protein folding differs from chemical refolding studies in three respects. (i) Nascent protein folding may be contranslational in the cell and reflect the polarity of protein biosynthesis. (ii) The relevant folding species in vivo is pro-IGF and not the mature hormone. It is possible that C-terminal prosequences provide additional information to guide native disulfide pairing. (iii) Cellular chaperoning may stabilize or destabilize alternative protein-folding intermediates, thereby influencing the direction of the folding pathway (Ostermann et al., 1989; Jaenicke et al., 1991; Krebs et al., 1983). In the future, it would be of interest to examine these issues. In addition, comparison of the three-dimensional structures of IGF-1, IGF-swap, and folding intermediates may permit rational design of analogues that favor one branch or the other (King et al., 1989; Fein & King, 1991). To our knowledge, such "pathway engineering" has not been attempted in other systems and would have broad theoretical and practical implications. A structural foundation for such studies is provided in the following paper by analysis of genetic models of protein-folding intermediates (Narhi et al., 1993).

ACKNOWLEDGMENT

We thank Bruce Altrock and Dan Vapnek (Amgen Inc.) for support and encouragement; M. Karplus (Harvard University) and F. Richards (Yale University) for discussion of protein folding; T. L. Blundell, R. E. Chance, G. G. Dodson, B. H. Frank, and S. E. Shoelson for discussion of insulin chemistry; J. P. Lee (Harvard Medical School) for assistance with NMR measurements; and I. Khait and L. J. Neuringer (MIT) for generous assistance with photo-CIDNP measurements at the Francis Bitter National Magnet Laboratory. NMR spectra were otherwise obtained at the Harvard Medical School NMR Facility. We also thank L. Ward and Jason Steele for preparation of the figures and Joan Bennett for preparation of the manuscript.

ADDED IN PROOF

A study of oxidative refolding of Ala-Glu-IGF-1 in 25% ethanol solution has recently been published (Hejnaes et al., 1992).

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., Dodson, 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.

Blundell, T. L., Bedarkar, S., Rinderknecht, E., & Humbel, R. E. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 180–184.

Byler, D. M., & Susi, H. (1986) *Biopolymers 25*, 469–487.

Canova-Davis, E., Kessler, T. J., & Lind, V. T. (1991) *Anal. Biochem. 196*, 39–45.

Cooke, R. M., Harvey, T. S., & Campbell, I. D. (1991) *Biochemistry 30*, 5484–5491.

Creighton, T. E. (1974) *J. Mol. Biol. 87*, 603–624.

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.

D'Ercole, A. J., Underwood, L. E., Van Wyk, J. J., Decedue, C. J., & Foushee, D. B. (1976) in *Growth Hormones and Related Peptides* (Pecile, A., & Muller, E., Eds.) pp 190–201, Excerpta Medica, Amsterdam.

DiMarchi, R. D., Mayer, J. P., Fan, L., Brems, D. N., Frank, B. H., Green, L. K., Hoffmann, J. A., Howey, D. C., Long, H. B., Shaw, W. N., Shields, J. E., Slieker, L. J., Su, K. S. E., Sundell, K. L., & Chance, R. E. (1992) in *Peptides: Proceedings of the Twelfth American Peptide Symposium* (Smith, J. A., & Rivier, J. E., Eds.) pp 26–28, ESCOM Science Publishers B. V., Leiden, The Netherlands.

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.
Elliott, S., Fagin, K. D., Narhi, L. O., Miller, J. A., Jones, M., Koski, R., Peters, M., Hsieh, P., Sachdev, R., Rosenfeld, R. D., Rohde, M. F., & Arakawa, T. (1990) *J. Protein Chem. 9*, 95–104.
Fein, B., & King, J. (1991) *Genetics 127*, 263–277.
Goldenberg, D. P. (1988) *Biochemistry 27*, 2481–2489.
Greenfield, M., & Fasman, G. D. (1969) *Biochemistry 8*, 4108–4116.
Haase-Pettingell, C. A., & King, J. (1988) *J. Biol. Chem. 263*, 4977–4983.
Hejnaes, K. R., Bayne, S., Nørskov, L., Sørensen, H. H., Thomsen, J., Schäffer, L., Wollmer, A., & Skriver, L. (1992) *Protein Eng. 5*, 797–806.
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.
Kaptein, R. (1980) in *Photo-CIDNP Studies of Proteins in Biological Magnetic Resonance* (Berliner, W., & Rueben, J., Eds.) Vol. 4, pp 145–191, Plenum Press, New York.
Kato, S., Okamura, M., Shimamoto, N., & Utiyama, H. (1981) *Biochemistry 20*, 1080–1085.
Kauppinen, J. K., Moffatt, D. J., Mantsch, H. H., & Cameron, D. L. (1981) *Appl. Spectrosc. 35*, 271–277.
Kim, P. S., & Baldwin, R. L. (1982) *Annu. Rev. Biochem. 51*, 459–489.
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.
Kristenson, S. M., Jorgenson, A. M. M., & Led, J. J. (1991) *J. Mol. Biol. 218*, 221–231.
Kuwajima, K. (1989) *Proteins: Struct., Funct., Genet. 6*, 87–103.
Kuwajima, K., Garvey, E. P., Finn, B. E., Matthews, C. R., & Sugai, S. (1991) *Biochemistry 30*, 7693–7703.
Meng, H., Burleigh, B. D., & Kelly, G. M. (1989) *J. Chromatogr. 443*, 183–192.
Morgan, D., Edman, J., Standring, D., Fried, V., Smith, M., Roth, R., & Rutter, W. (1987) *Nature 329*, 301–307.
Muszkat, K. A., & Gilon, C. (1978) *Nature (London) 271*, 685–686.
Narhi, L. O., Hua, Q.-X., Arakawa, T., Fox, G. M., Tsai, L., Rosenfeld, R., Holst, P., Miller, J. A., & Weiss, M. A. (1993) *Biochemistry* (following paper in this issue).
Nilsson, B., & Anderson, S. (1991) *Annu. Rev. Microbiol. 45*, 607–635.
Oas, T. G., & Kim, P. S. (1988) *Nature (London) 336*, 42–48.
Ostermann, J., Horwich, A. L., Neupert, W., & Hartl, F.-U. (1989) *Nature 341*, 126–130.
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.
Shortle, D., & Lin, B. (1985) *Genetics 110*, 539–555.
Sieber, P., Eisler, K., Kamber, B., Riniker, B., Rittel, W., Marki, F., & deGasparo, M. (1978) *Hoppe-Seyler's Z. Physiol. Chem. 359*, 113–123.
Staley, J. P., & Kim, P. S. (1990) *Nature 344*, 685–688.
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., Haberkorn, R. A., & Ruben, D. J. (1982) *J. Magn. Reson. 48*, 286–292.
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.
Surewicz, W. R., & Mantsch, H. H. (1988) *Biochim. Biophys. Acta 952*, 115–130.
Tang, J.-G., Wang, C.-C., & Tsou, C.-L. (1988) *Biochem. 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 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., Nguyen, D., Khait, I., Inouye, K., Frank, B. H., Beckage, M., O'Shea, E., Shoelson, S. E., Karplus, M., & Neuringer, L. J. (1989) *Biochemistry 28*, 9855–9873.
Weiss, M. A., Frank, B. F., Khait, I., Pekar, A., Heinery, R., Shoelson, S. E., & Neuringer, L. J. (1990) *Biochemistry 29*, 8389–8401.
Weiss, M. A., Hua, Q.-X., Frank, B. H., Lynch, C., & Shoelson, S. E. (1991) *Biochemistry 30*, 7373–7389.
Weissman, J. S., & Kim, P. S. (1991) *Science 253*, 1386–1393.