Disulfide Exchange Folding of Insulin-like Growth Factor I†

Sophia Hober,‡ Göran Forsberg,‡,§ Gunnar Palm,§ Maris Hartmanis,‡,§ and Björn Nilsson*,‡,§

Department of Biochemistry and Biotechnology, Royal Institute of Technology, S-100 44 Stockholm, Sweden, and Kabi Pharmacia KabiGen, S-112 87 Stockholm, Sweden

Received June 19, 1991; Revised Manuscript Received November 12, 1991

ABSTRACT: The disulfide exchange folding properties of insulin-like growth factor I (IGF-I) have been analyzed in a redox buffer containing reduced (10 mM) and oxidized (1 mM) glutathione. Under these conditions, the 3 disulfide bridges of the 70 amino acid peptide were not quantitatively formed. Instead, five major forms of IGF-I were detected, and these components were concluded to be in equilibrium as their relative amounts were similar starting from either reduced, native, or a mismatched variant of IGF-I containing two non-native disulfides. The different components in the mixtures were trapped by thiol alkylation using vinylpyridine and subsequently isolated by reverse-phase HPLC. The purified variants were further characterized using plasma desorption mass spectrometry and peptide mapping. Two of the five different forms were identified as native and mismatched IGF-I. One form was a variant with only one disulfide bond, and the other two major components had two disulfides formed. In a separate experiment, early refolding intermediates were trapped by pyridylethylation after only 90 s of refolding in the glutathione buffer, starting from reduced IGF-I. The intermediates were identical to the components observed at equilibrium, but at different relative concentrations. On the basis of the disulfide bond patterns of the different components in the equilibrium mixtures, we conclude that the disulfide between cysteines-47 and -52 in IGF-I is an unfavorable high-energy bond that may exist in the native molecule in a strained configuration.

Human insulin-like growth factor I (IGF-I)[1] is a single-chain peptide growth factor of 70 amino acids, originally isolated from serum [for a review, see Humbel (1990)]. IGF-I is positively regulated by growth hormone (GH) and shows mitogenic effects on many cell types. Therefore, IGF-I is thought to mediate many of the growth-promoting effects of GH (Spencer et al., 1988). In the regions of homology, IGF-I and insulin are 49% homologous, including the six cysteine residues, furnishing three disulfide bridges (Figure 1) (Rinderknecht & Humbel, 1978). The three-dimensional structure of IGF-I has been modeled on the basis of the X-ray structure of insulin (Figure 2) (Blundell et al., 1978), and this model has recently been confirmed in the disulfide bridge regions by distance constraints obtained by 2-D NMR spectroscopy of IGF-I (Cooke et al., 1991).

Human recombinant IGF-I has been produced as a secreted product in both *Escherichia coli* and *Saccharomyces cerevisiae*. In isolated material from both species, IGF-I is found mainly as misfolded forms with intermolecular disulfides (Elliott et al., 1990; Samuelsson et al., 1991). Surprisingly, at least two distinct monomeric forms, with differences in their disulfide bond patterns, have been identified. One of these two forms contains the disulfide bond topology expected from the insulin structure, and this form is biologically active (Meng et al., 1988; Iwai et al., 1989). The other monomeric form, designated "mis-matched" (disulfides 6–47,[2] 48–52 instead of the native 6–48, 47–52; see Figures 1 and 2), lacks IGF-I receptor affinity (Raschdorf et al., 1988; Iwai et al., 1989; Forsberg et al., 1990). Also, in vitro refolding of reduced IGF-I by oxygen has demonstrated that native, mismatched, and aggregated IGF-I accumulate, even under dilute refolding conditions (Meng et al., 1988; Iwai et al., 1989; Samuelsson et al., 1991). Thus, both the in vivo and the in vitro folding properties of IGF-I raise questions concerning thermodynamics and kinetics of disulfide bond formation of IGF-I.

The folding of proteins having disulfide bonds in the native state can be studied by trapping intermediates through thiol alkylation during the process of folding. Trapped intermediates are subsequently identified, the kinetics of interconversion are determined, and the different components are connected to a folding pathway [for a review, see Creighton (1986)]. The best-studied disulfide exchange folding pathway is that of bovine pancreatic trypsin inhibitor (BPTI), which is a 58 amino acid peptide containing 3 disulfide bonds. Most notably, in the rate-limiting step of the kinetically preferred folding pathway of BPTI, the molecule passes through an intermediate with a non-native disulfide (Creighton, 1978). From the analysis of mutant BPTI variants, it has been demonstrated that the formation of the native disulfide, resulting from the non-native disulfide intermediate, is about 3 orders of magnitude faster in the rearrangement folding pathway than by direct formation (Marks et al., 1987; Goldenberg, 1988).

Here, we report the first analysis of disulfide exchange folding of IGF-I under reversible redox conditions in a thiol–disulfide buffer (Saxena & Wetlaufer, 1970). It is shown that IGF-I, at equilibrium in a glutathione buffer, is not forming its three disulfide bonds quantitatively. Instead, at

† This project has been financially supported by the Swedish Natural Science Research Council (Grant K-KU 9396-306), the Swedish National Board for Technical Development, and Kabi Pharmacia KabiGen.

* Address correspondence to this author at the Department of Biochemistry, Kabi Pharmacia KabiGen.

‡ Royal Institute of Technology.

§ Kabi Pharmacia KabiGen.

[1] Abbreviations: BPTI, bovine pancreatic trypsin inhibitor; CD, circular dichroism; DTT, dithiothreitol; ER, endoplasmatic reticulum; GH, growth hormone; GSH, reduced glutathione; GSSG, oxidized glutathione; IAA, iodoacetic acid; IGF-I, insulin-like growth factor I; IGF-II, insulin-like growth factor II; PDI, protein disulfide isomerase; PDMS, plasma desorption mass spectrometry; PFPA, pentafluoropropionic acid; RP-HPLC, reverse-phase high-performance liquid chromatography; VP, vinylpyridine.

[2] In this paper, amino acid residues are numbered according to their position in the corresponding protein from its N-terminus. Similarly, disulfide bonds are designated as connected residue numbers, e.g., 18–61 as the disulfide connecting amino acid residues 18 and 61.

0006-2960/92/0431-1749$03.00/0

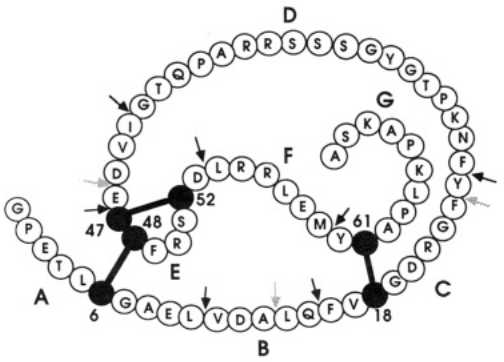

FIGURE 1: Two-dimensional representation of the structure of IGF-I. Amino acids are shown in one-letter code. The disulfide bridges are connected as determined in native IGF-I (Raschdorf et al., 1988). The major cleavage sites of pepsin are indicated with black arrows and the minor cleavage sites with gray arrows. Peptide fragments obtained after peptic cleavage are designated with letters (A–G).

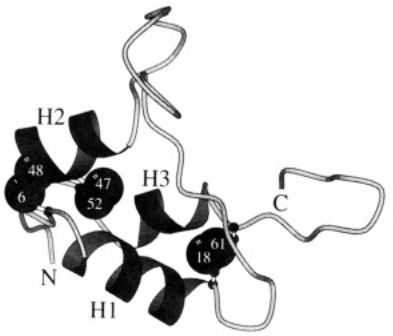

FIGURE 2: Main-chain trace representation of the model of the structure of IGF-I (Blundell et al., 1978), which also recently has been confirmed in the disulfide bond regions by NMR spectroscopy (Cooke et al., 1991). The three α-helices are shown as helical ribbons, which are marked by H1, H2, and H3 for helix 1, 2, and 3, respectively. The six cysteine residues are displayed by showing their sulfur atoms as black spheres according to their van der Waals radii, by showing residue numbers on each sulfur atom and by showing their C_α and C_β as ball-and-stick models. The N-terminus is shown by N, and the C-terminus is shown by C. The display was generated by the program MOLSCRIPT (Kraulis, 1991).

least five different forms of IGF-I, with differences only in their disulfide bond patterns, were identified. These IGF-I forms are connected to a pathway of interconversion. In addition, the implications of the folding properties of IGF-I, on the formation of the native disulfides of IGF-I in vivo in the eukaryotic cell, are discussed.

Materials and Methods

Preparation of Native, Mismatched, and Reduced Forms of IGF-I. Native and mismatched IGF-I were produced in *E. coli* as fusion proteins (Moks et al., 1987) and purified to homogeneity as described (Forsberg et al., 1990). Reduced IGF-I was prepared by incubating a mixture of native and mismatched IGF-I at 37 °C for 1 h at a concentration of 170 μM in a buffer containing 0.1 M Tris, pH 8.7, 0.2 M KCl, 1 mM EDTA, 8 M urea, and 10 mM reduced dithiothreitol (DTT). After reduction, the buffer was exchanged to 10 mM HCl by gel filtration using Sephadex G-25 (Pharmacia LKB Biotechnology, Uppsala Sweden). Aliquots of reduced IGF-I could be stored without detectable oxidation for months at −80 °C. Before use, the reduced state of the six thiols in the material was confirmed by alkylation followed by reverse-phase high-performance liquid chromatography (RP-HPLC) analysis as described below.

Protein Analysis. IGF-I concentrations were calculated from their absorbances at 280 nm using the specific absorption constant A_{280}(0.1%, 1 cm) = 2.1. UV absorbance spectra were determined in a Kontron 860 spectrophotometer (Kontron, Switzerland).

Circular dichroism (CD) spectra were collected in a J-720 spectropolarimeter (JASCO, Japan) at room temperature in 10 mM potassium phosphate buffer, pH 7.2, with or without 1 mM reduced DTT, or in 10 mM HCl. The scanning speed was 10 nm/min, and each spectrum was averaged from five individual scans. Cell path length was 1 mm, and protein concentration was 0.1 mg/mL. Secondary structure predictions were performed by fitting to reference spectra (Yang et al., 1986).

Amino-terminal sequencing was performed using an automated solid-phase Prosequencer 6600 (MilliGen). For covalent binding of carboxyl groups, arylamine membranes (Milligen) were used. After binding of the peptides to the membranes, disulfide bridges were reduced with tributylphosphine at pH 8.6. Free thiols were subsequently alkylated with iodoacetic acid (IAA), and the cysteine residues were detected as carboxymethylated or pyridylethylated residues.

Amino acid compositions were analyzed by first hydrolyzing each peptide in 6 M HCl at 155 °C for 45 min. Thereafter, the samples were analyzed on an ion-exchange column, and the amino acids were detected with ninhydrin. The composition analysis was done on a Beckman 6300 amino acid analyzer, equipped with a System Gold data handling system (Beckman).

Molecular masses were determined using a ^{252}Cf plasma desorption mass spectrometer (PDMS), Bio-Ion 20 (Applied Biosystems). The peptides were bound to a nitrocellulose-coated foil, dried by spinning the foil, and analyzed in the positive ion mode at an acceleration voltage of 18 kV (Forsberg et al., 1990). The errors in measured masses are estimated to be less than 0.1%.

Disulfide Exchange Reactions of IGF-I. The disulfide exchange reactions were carried out at 37 °C at an IGF-I concentration of 30 μM in a buffer containing 0.1 M Tris, pH 8.7, 0.2 M KCl, 1 mM EDTA, 10 mM reduced glutathione (GSH), and 1 mM oxidized glutathione (GSSG) (Creighton, 1986). Disulfide exchanges were terminated by alkylating free thiols using 160 mM vinylpyridine (VP) (Henschen, 1986). The pyridylethylation reaction was allowed to proceed for 15 min in the dark whereafter the buffer was exchanged to 10 mM HCl using gel filtration on Sephadex G-25 (Pharmacia LKB Biotechnology).

Separation of IGF-I Variants and IGF-I Peptide Fragments. Pyridylethylated variants of IGF-I were separated by RP-HPLC. The column used was a Kromasil C_8 with 7-μm particles having a pore diameter of 18 nm (Eka Nobel). The gradient used was 30–45% acetonitrile in 0.25% pentafluoropropionic acid (PFPA) over 30 min, at a flow rate of 1 mL/min and a temperature of 30 °C. The elution was monitored by a diode array detector and a fluorescence detector in series (Hewlett Packard).

Purified IGF-I variants were digested with porcine pepsin (Sigma) in 10 mM HCl, with a protease to IGF-I ratio of 1:10 (w/w). The peptic digestion was allowed to proceed for 3–4 h (Forsberg et al., 1990). Generated peptide fragments were separated by RP-HPLC on a Kromasil C_8 column with 7-μm particles having 10-nm pore diameter. The gradient used was 0–45% acetonitrile in 0.1% trifluoroacetic acid over 40 min at a temperature of 30 °C and a flow rate of 1 mL/min (Forsberg et al., 1990).

Relative free energies were calculated from equilibrium constants (Creighton, 1984). Molecular dynamics calculations

were performed using Discover with the CVFF force-field (Biosym, Inc., San Diego, CA).

RESULTS

Structural Implications of IGF-I Misfolding. Native IGF-I contains 70 amino acids including 6 cysteine residues furnishing 3 disulfide bonds (Figure 1). The three-dimensional structure of native IGF-I has been modeled on the basis of the insulin structure (Blundell et al., 1978) (Figure 2). The model structure was recently shown to be in agreement with distance geometries obtained by 2-D NMR spectroscopy (Cooke et al., 1991). There are three α-helices in the model, which are cross-linked by disulfide bonds. The disulfide 6–48 connects the N-terminus of helix 1 (helices are numbered from the N-terminus) to the C-terminus of helix 2. The 47–52 disulfide connects helix 2 to the loop region between helix 2 and helix 3. The 18–61 disulfide connects the C-terminus of helix 1 to the C-terminus of helix 3 (Figures 1 and 2).

A mismatched form of IGF-I, where the 6–48, 47–52 disulfides in native IGF-I are exchanged to 6–47, 48–52, appears both in recombinant and in in vitro refolded IGF-I (Raschdorf et al., 1988; Iwai et al., 1989; Elliott et al., 1990; Forsberg et al., 1990; Samuelsson et al., 1991). In the model of native IGF-I, the distance between the centers of the sulfur atoms of cysteine residues 6 and 47 is approximately 9.0 Å. Thus, a structural change, moving the two sulfur atoms 7 Å closer, is required for the formation of the "mis-matched" 6–47 bridge. In addition, a rotation in the polypeptide chain has to take place to allow the formation of an unstrained disulfide bond (Srinivasan et al., 1990). Similarly, the distance between cysteines-48 and -52 in the model structure of IGF-I is 9.5 Å, and here too, the torsion angles are suboptimal for a disulfide. Thus, mismatched IGF-I cannot be obtained from the model structure of native IGF-I by changing the disulfide bonds without seriously changing the three-dimensional structure. We have forced the mismatched disulfides into the native IGF-I model and analyzed the structural consequences by collecting a library of conformations using molecular dyanmics calculations in vacuum at elevated temperature (data not shown). In *all* generated conformations, the α-helical contents were reduced, compared to the model of native IGF-I. In contrast, when the molecular dynamics calculations were performed using the same parameters but starting from the model structure of native IGF-I, the secondary structure was essentially unaffected throughout the calculations (data not shown). In conclusion, this rough analysis implies that mismatched IGF-I cannot simply be generated from native IGF-I and that it should possess a significantly different structure compared to the model of native IGF-I.

Circular Dichroism Spectra of Native, Mismatched, and Reduced IGF-I. The dependence on disulfides to form the nativelike IGF-I structure was studied by comparing the circular dichroism (CD) spectra of native and reduced IGF-I. In addition, the CD spectrum of mismatched IGF-I was determined to compare the secondary structure content with native IGF-I (Figure 3). For native IGF-I, the circular dichroism spectrum shows a minimum at 208 nm and a shoulder at 222 nm in combination with a maximum between 185 and 190 nm, indicating α-helicity in the molecule (Johnson, 1990). Secondary structure predictions from the spectrum show that native IGF-I has approximately 22% α-helical content. This is in good agreement with the model of IGF-I (Blundell et al., 1978). In contrast, reduced IGF-I (Figure 3) resembles a CD spectrum close to random coil (Johnson, 1990). To avoid disulfide formation during spectroscopy, the spectrum of reduced IGF-I in Figure 3 was

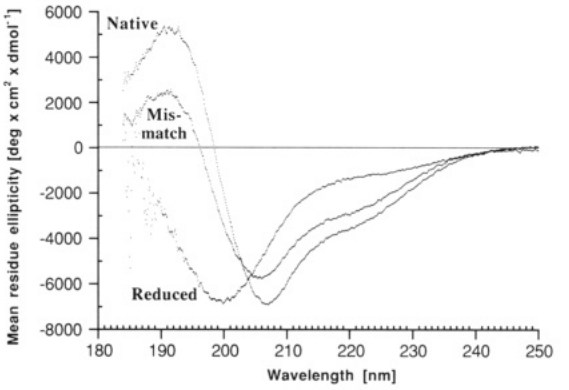

FIGURE 3: Far-UV circular dichroism spectra of reduced (Reduced), mismatched (Mis-match), and native (Native) IGF-I from 184 to 250 nm as described under Materials and Methods.

collected at pH 2, but the CD spectrum of reduced IGF-I at neutral pH in the presence of reduced DTT looks indistinguishable from the pH 2 spectrum from 250 to 215 nm. However, a complete far-UV spectrum could not be obtained in the presence of DTT due to the strong far-UV absorbance of DTT (not shown). Thus, without its three disulfides, IGF-I does not fold into a nativelike three-dimensional conformation. Mismatched IGF-I shows a significantly reduced α-helical content compared to native IGF-I. This is demonstrated both by the amplitudes and by the shift in the minimum at 208 nm toward a lower wavelength. Secondary structure predictions from the spectrum indicate that the α-helical content has been reduced to approximately 15%. These results support the lowered secondary structure contents of the generated models of mismatched IGF-I in the molecular dynamics calculations.

Equilibrium Studies of Reduced, Native, and Mismatched IGF-I. Native, mismatched, and reduced IGF-I, respectively, were equilibrated in a glutathione redox buffer containing 1 mM GSSG and 10 mM GSH. In a control experiment, in the absence of the reduced and oxidized glutathione, oxygen showed only minor oxidative effects on reduced IGF-I after 1 h (data not shown). Thus, we conclude that the glutathione buffer is responsible for essentially all of the observed disulfide exchange reactions under the conditions used. After 1 h, the components in the three different equilibrium mixtures were trapped by thiol alkylation using VP. Separation of the different IGF-I forms in the mixtures was performed on RP-HPLC. The chromatograms of the three different reactions are similar, and at least five major peaks were observed (Figure 4). The materials in these different peaks were collected. In a separate control experiment, native and reduced IGF-I were alkylated with VP and analyzed by RP-HPLC. No detectable alkylation could be observed in native IGF-I, as determined by the absence of the pyridylethyl chromophore at 254 nm, while six pyridylethyl groups per IGF-I molecule were detected in reduced IGF-I (data not shown). Thus, under the conditions used, VP will specifically alkylate free thiol groups in IGF-I.

The five different forms of IGF-I, isolated from the equilibrium mixtures, were further analyzed by PDMS. The masses reveal the possible number of covalently bound pyridylethyl or glutathione groups (Table I). The presence of pyridylethyl groups was also detected by its chromophore, showing a strong absorbance at 254 nm (Fullmer, 1984). The relative amounts of IGF-I in the different peaks were determined by integrating their fluorescence spectra (excitation at 280 nm and emission at 305 nm) (Table I). Pyridylethyl groups do not add to the fluorescence. In addition, to analyze possible differences in tyrosine fluorescence of the IGF-I forms in the different peaks, collected materials were also analyzed

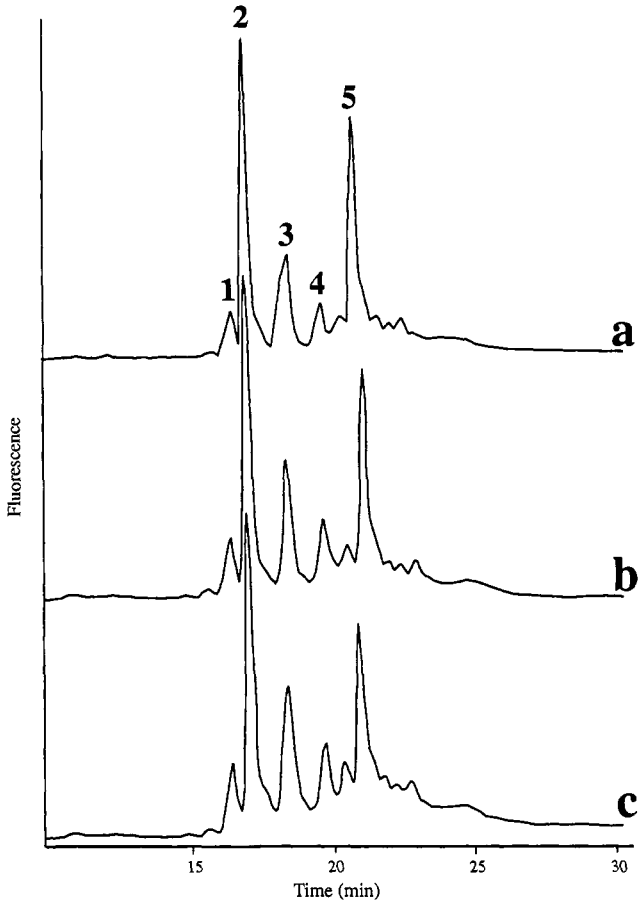

FIGURE 4: RP-HPLC separation of IGF-I components after incubation in a glutathione redox buffer and alkylating free thiols using VP. The gradient was linear, from 30 to 45% acetonitrile in 0.25% PFPA, over 30 min. The fluorescence (excitation at 280 nm and emission at 305 nm) of the different peaks was measured. The chromatograms show samples from incubations in the redox buffer starting from either (a) native IGF-I, (b) mismatched IGF-I, or (c) reduced IGF-I. The major peaks (1–5) are numbered according to their elution order.

Table I: Analysis of Separated IGF-I Components at Equilibrium in Glutathione Buffer (Figure 4)[a]

peak	designation	no. of VP	calcd mass (Da)	measured mass (Da)	rel free energies (kcal/mol)	amounts (%)
1	III′	0	7649.6	7650.0	0.5	8
2	II	2	7859.9	7859.6	−1.8	37
3	III	0	7649.6	7649.7	0	18
4	II″	2	7859.9	7861.0	−0.8	7
5	I	4	8070.2	8069.5	−3.1	30
6	0	6	8280.4	8280.9		

[a] The analysis was performed by PDMS and fluorescence integration from RP-HPLC (Figure 4) of the material in the different peaks. Percentage values were calculated from integrating the fluorescence of the chromatograms in Figure 4a. Relative free energies in the presence of 10^{-3} M GSH and 10^{-2} M GSSG were calculated relative to native IGF-I from calculated equilibrium constants. Peak 6 material was isolated from the 90-s refolding experiment as shown in Figure 6.

by amino acid composition analysis. The two methods differed by less than 5% in peak 1, 2, 3, and 5 and by less than 10% in peak 4. Thus, we conclude that these five peaks show similar tyrosine fluorescence under the conditions used and, thus, that integrated fluorescence values can be used for relative quantification. The relative numbers of pyridylethyl groups could be determined by dividing the integrated absorbance at 254 nm and the integrated fluorescence (data not shown). The number of free thiols of each IGF-I form calculated by this method was identical to the numbers derived from the molecular masses (Table I). Thus, among the five major forms of IGF-I in the equilibrium mixture, there was one form with four free thiols, two forms with two free thiols, and two forms lacking free thiols (Table I). The relative amounts were used to calculate free energies relative to native IGF-I from equilibrium constants (Table I).

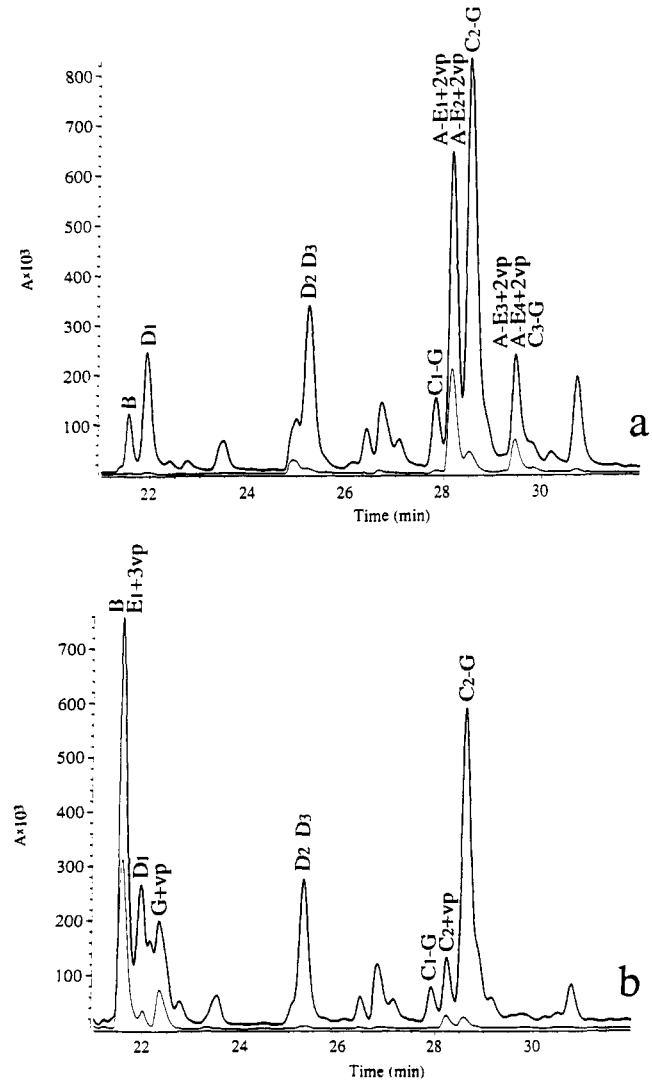

FIGURE 5: RP-HPLC separation of peptic cleavage of the materials from peak 2 (a) and peak 5 (b) in Figure 4. The absorbances were monitored at 220 nm (upper curve) and 254 nm (lower curve). The peaks were further analyzed (Table II), and the identities of some of the peaks are shown. Their identities are shown by letters as defined in Figure 1 and the number of bound VP(s) (vp).

Mapping the Disulfides of the Different IGF-I Components. The identities of the five isolated major IGF-I components in the equilibrium mixtures, designated peaks 1–5 in Figure 4, were further analyzed. Peak 1 and peak 3 exhibit identical retention behavior in the RP-HPLC system to mismatched and native IGF-I, respectively (Forsberg et al., 1990) (data not shown). Also, the materials in these peaks did not contain any pyridylethyl groups, as concluded both from the absorbance at 254 nm and by their masses (Table I). Thus, peak 1 was identified as mismatched IGF-I, and peak 3 was identified as native IGF-I.

Peaks 2, 4, and 5 were further analyzed. The isolated IGF-I variants were digested with pepsin (Figure 1), and the generated fragments were separated on RP-HPLC (shown for peaks 2 and 5 in Figure 5). The absorbance maximum at 254 nm was monitored to identify fragments containing pyridyl-

Table II: Identification of Peptide Fragments by PDMS and Amino Acid Composition Analysis[a]

peptide fragment	amino acids	calcd mass (Da)	measured mass (Da) peak 2	peak 4	peak 5	confirmed with aa comp
A–E1+2VP	1–10, 47–53	2030.3	2029.4	2029.6		yes
A–E2+2VP	1–10, 46–53	2159.4	2158.6	2156.4		yes
A–E3+2VP	1–10, 43–53	2486.8	2486.4			yes
D1	25–42	1911.0			1910.6	yes
D2	25–45	2238.4	2237.6			yes[b]
D3	25–46	2367.5	2366.1			yes[b]
C1–G	16–23, 60–70	2046.4	2045.4		2045.6	yes[b]
C2–G	16–24, 60–70	2209.5	2208.6	2207.6	2208.7	yes
C3–G	14–23, 60–70	2287.7	2287.1			
E1+3VP	47–53	1148.4			1147.6	yes
G+VP	60–70	1253.5			1253.7[c]	
C2+VP	16–24	1168.3			1167.7	

[a] Peak numbers correspond to the designations in Figure 4. [b] From Forsberg et al. (1990). [c] Determined from peak 6.

ethyl groups. Most of the isolated peptide fragments were analyzed by PDMS and amino acid analysis (Table II).

On the basis of the mass (Table I) and absorbance at 254 nm, the IGF-I molecules in peak 2 are expected to contain two disulfides and two pyridylethylated thiols. In the different peptides that were generated by peptic cleavage (Figure 5a), an increase in the relative absorbance at 254 nm indicates the presence of pyridylethylated groups (lower curve, Figure 5a). The identities of the different collected peptides were determined by their molecular masses, which were individually analyzed by PDMS, and their respective amino acid composition. The results, which are summarized in Table II and in Figure 5a, clearly show that a disulfide bridge between cysteine residues 18 and 61 was formed. The isolated peptide that contained VP was fragment A attached to fragment E (Figure 1). The covalent linkage between peptides A and E was concluded to be the result of cysteine residue 6 participating in a disulfide bridge with either one of the three cysteine residues in fragment E: 47, 48, or 52 (Figure 1). The isolated A–E peptide was bound to an arylamine membrane and subsequently reduced and alkylated with IAA. Amino-terminal sequencing of the two overlapping A–E peptide sequences revealed that cysteine-6 was quantitatively linked to cysteine-48. Thus, IGF-I (6–48, 18–61) with pyridylethylated thiols at cysteines-47 and -52 elutes as peak 2.

The IGF-I molecules in peak 4 are also expected to contain two disulfide bridges (Table I). Peptides similar to the ones isolated from the analysis of peak 2 (data not shown) were obtained in the RP-HPLC separation of the peptic digest of peak 4. Also here, the 18–61 is present, and pepsin fragments A and E are covalently linked by a disulfide (Figure 1). Here, amino-terminal sequencing of the A–E peptide revealed a non-native disulfide between cysteine residues 6 and 52. Thus, peak 4 was concluded to contain IGF-I (6–52, 18–61) with pyridylethylated thiols at cysteines-47 and -48.

Peak 5 was subsequently analyzed. According to the mass and the pyridylethyl absorbance, this IGF-I form is expected to contain only one disulfide bond (Table I). After peptic cleavage, the different peptides were isolated and analyzed by PDMS and amino acid analysis. The results are summarized in Table II, and the identities of the peaks are marked in Figure 5b. In conclusion, the IGF-I form constituting peak 5 was identified as IGF-I (18–61) since attached C–G peptides (Figures 1 and 5b) were present and pyridylethyl groups were found in peptides on cysteine residues 47, 48, and 52, but in no case on cysteines-18 or -61. Even though the E peptide with three pyridylethyl groups could be isolated and analyzed, it should be mentioned that the expected pyridylethylated form of the short A peptide was not found in the chromatogram (Figure 1).

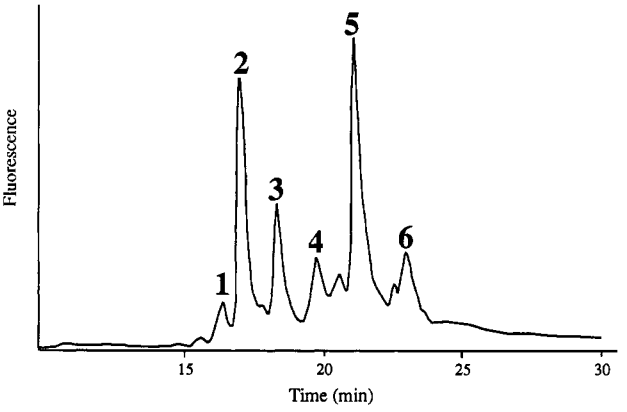

FIGURE 6: Folding intermediates of IGF-I after refolding from reduced protein for 90 s in 10 mM GSH and 1 mM GSSG. The disulfide exchange reactions were trapped by VP alkylation. Separation and detection of the different intermediates were performed as described in Figure 4.

Early Intermediates. The significance of the identified components at equilibrium, in the process of folding, was investigated. Reduced IGF-I was allowed to refold in the glutathione buffer. After 90 s, the disulfide exchange reactions were terminated by VP alkylation of free thiols. The trapped reaction mixture was subsequently analyzed by RP-HPLC (Figure 6). The same five variants of IGF-I, based on RP-HPLC retentions and the presence of the VP chromophore, as were found in equilibrium (Figure 4) were also found as intermediates at this early stage of folding, but at different relative amounts. In addition, there was an additional major form, designated peak 6. The molecular mass and absorbance: fluorescence ratio as well as the retention time on RP-HPLC were identical to pyridylethylated reduced IGF-I (data not shown). Thus, peak 6 was identified as reduced IGF-I. The most predominant 90-s intermediate was the form containing only one disulfide bond, between cysteine residues 18 and 61.

Discussion

In this paper, disulfide exchange folding properties of IGF-I have been analyzed. In a redox buffer composed of reduced an oxidized glutathione, IGF-I did not fold quantitatively into its native conformation. Instead, an equilibrium was established between many different forms of which the five most predominant have been isolated and identified. In Figure 7, these identified IGF-I variants, as well as postulated intermediates, are connected to a pathway of interconversion. We

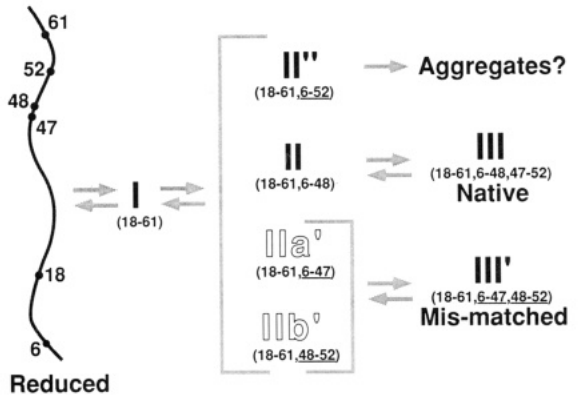

FIGURE 7: Disulfide exchange pathway of interconversion of IGF-I in glutathione. The number of disulfide bridges in the intermediates are indicated by their Roman numerals, and the cysteine bridges are shown below each intermediate. All non-native disulfides are underlined. A prime after the Roman numeral indicates a molecule where all formed disulfides are as the ones in "mis-matched" IGF-I. Two primes after the Roman numeral indicate a molecule having at least one disulfide which is found neither in "mis-matched" nor in "native" IGF-I. The absence of a prime or two primes after the Roman numeral indicates that there are only native disulfide(s) in the molecule. All forms that have been isolated, either at equilibrium or after 90 s of refolding, have their Roman numerals in black, and the postulated forms are shown with their Roman numerals outlined.

also believe that the pathway is relevant for the kinetics of refolding in glutathione since the same disulfide variants were found after only 90 s of refolding (Figure 6). In addition, we believe that the pathway can be used to explain the aggregation found in recombinant material from both *E. coli* and *S. cerevisiae* and in IGF-I which has been refolded in vitro (Elliott et al., 1990; Samuelsson et al., 1991) since the same components also show up as intermediates when reduced IGF-I is refolded by oxygen (S. Hober, unpublished experiments).

In the pathway, the first native disulfide to form is 18–61 (Figure 7, molecule I). In fact, in all isolated and analyzed forms of IGF-I containing at least one disulfide, the 18–61 bond is formed (Figure 5, Table II). Also, when reduced IGF-I was allowed to refold in the presence of oxygen, the first intermediate to form was IGF-I with only the 18–61 bond formed (S. Hober, unpublished results). Therefore, even though we have not characterized the kinetic role of this intermediate, we suggest not only that the 18–61 disulfide is the most stable disulfide bond (Table I) but also that this disulfide may be present in a region of early contacts of folding, similar to the first native disulfide (30–51) to form in BPTI (Creighton, 1974; Hurle et al., 1990).

From molecule I, there are six possible two-disulfide intermediates that can be formed. Four of these are shown in Figure 7. Two such intermediates were observed as major components in the equilibrium mixture or as major intermediates after 90-s refolding: 18–61, 6–48 (Figure 7, molecule II); and 18–61, 6–52 (Figure 7, molecule II''). In addition, molecule II is the most populated component in the equilibrium mixture (Figure 4, Table I).

Even though the IIa′ and IIb′ (Figure 7) forms of IGF-I have not been isolated, at least one of them must exist since they constitute a missing link in the pathway from any of the isolated variants to mismatched IGF-I (designated III′ in Figure 7). In contrast, the isolated intermediate II'' is not on the pathway to either III or III′. Even though disulfide bridges between neighboring cysteines are possible in peptides (Zhang & Snyder, 1989), they have not been found in any native protein to our knowledge, only in misfolded recombinant proteins produced in bacteria (Vértesy et al., 1991). Instead, we like to propose that the highly populated II'' is important for IGF-I aggregation during refolding by oxygen (Samuelsson et al., 1991).

Two conclusions can be made from the results that IGF-I does not fold quantitatively into its native conformation in a redox mixture containing oxidized and reduced glutathione (1 mM GSSG/10 mM GSH). First, there are energetically unfavorable disulfides in IGF-I. Glutathione is a mono thiol compound which, for entropic reasons, is expected to favor formation of protein disulfides unless they are energetically unfavorable, e.g., by being strained (Creighton, 1983; Wetlaufer et al., 1987). Under the conditions used, molecule I possesses the lowest free energy, while IGF-I variant II (Figure 7) is the mostly populated form. Thus, we conclude that the disulfide 47–52, perhaps also 6–48, is unfavorable in the native conformation of IGF-I and that the observed inability of IGF-I to quantitatively reach the native state is a thermodynamic property because of energetically unfavorable disulfide(s) in native IGF-I.

Second, the amino acid sequence of mature IGF-I does not contain all necessary information for its folding into a native state under the conditions used, since there are at least two separate monomeric three-disulfide forms of IGF-I in the mixture. Their relative amounts (Table I) show that their respective free energies in the redox buffer differ by only about 0.5 kcal/mol. This "non-Anfinsen" folding behavior is unexpected. Recently it was suggested that all protein molecules, even tissue plasminogen activator containing 34 cysteines, should contain all information within their amino acid sequence to quantitatively refold in vitro into their native conformation (Jaenicke, 1991). Our results would imply that IGF-I may not possess this ability.

Two obvious questions are how IGF-I folds in vivo in mammals and why mismatched IGF-I has not been found in materials isolated from serum. We find three possible scenarios that would answer these questions. First, mismatched IGF-I may in fact be formed in the lumen of the endoplasmatic reticulum (ER). After formation, it may possess a yet unknown function in circulation. Alternatively, it may be recognized as misfolded in the ER, cleared away, and subsequently degraded, as suggested to be a general mechanism to be found in the ER (Hurtley & Helenius, 1989). Second, the pro-form of IGF-I, containing a C-terminal peptide extension known as the E peptide (Rotwein, 1986), may show different thermodynamics of disulfide bond formation, favoring the formation of pro-IGF-I with native disulfide bonds. After processing of the E peptide, native IGF-I may be stable in serum in the absence of reductants. Third, IGF-I may fold by the aid of molecular chaperones in vivo (Nilsson & Anderson, 1991). These chaperones could help by specifically stabilizing only the native form. The protein disulfide isomerase (PDI), which is responsible for disulfide bond formation in the lumen of the ER in eukaryotic cells (Edman et al., 1985), seems to act as a true catalyst and is thus not expected to change the equilibrium between species (Nilsson & Anderson, 1991). However, we cannot totally rule out the possibility that PDI could favor the formation of the native disulfides of IGF-I in vivo. In principle, this could be the result if PDI would possess different accessibility to the various disulfide possibilities of IGF-I. In fact, native insulin, which is structurally related to IGF-I, can be formed from the scrambled molecule by PDI (Tang et al., 1988).

The proposed pathway of interconversion of IGF-I in glutathione (Figure 7) explains the formation of mismatched IGF-I in vivo in *E. coli* and *S. cerevisiae*, but not the formation

of aggregates. Aggregates held together by *inter*molecular disulfides account for more that 80% of the material isolated from both yeast (Elliott et al., 1990) and *E. coli* (Samuelsson et al., 1991). In contrast, in the glutathione buffer, we have been able to detect traces of dimeric materials in the equilibrium mixture (data not shown), but no aggregated forms of IGF-I have been found. We conclude that aggregates possess low stability and are therefore not populated in the equilibrium system. In the microbial in vivo systems as well as in the published in vitro folding systems (Meng et al., 1988; Iwai et al., 1989; Samuelsson et al., 1991), the oxidation by oxygen is irreversible, resulting in the formation of aggregates. Possibly, the highly populated components at equilibrium in glutathione (I, II, or II″) could possess long half-lifes when oxygen is used as oxidant and act as templates for aggregation. Indeed, when an equilibrium mixture of IGF-I in the glutathione buffer is incubated for an additional 24 h in the presence of oxygen, a majority of the IGF-I material forms aggregates (data not shown).

What role does the choice of oxidant have on the determined disulfide exchange folding properties of IGF-I? To address this, we have refolded reduced IGF-I using oxygen as the oxidizing agent, and the same intermediates were found under these conditions as at equilibrium in the glutathione buffer (data not shown). These results imply that the disulfide exchange folding properties of IGF-I described in this paper are mainly an intrinsic property of IGF-I and not a function of the specific oxidation system acting on the molecule.

The kinetics of pyridylethylation of thiols have not been determined. Even though the alkylation reaction, under the conditions used, is expected to be faster than folding-promoted disulfide rearrangements in proteins (Creighton, 1984; Henschen, 1986), it cannot be totally be ruled out that the pyridylethylation itself is the cause of the apparent similar content of IGF-I components independent of starting molecule (Figure 4). In fact, it was recently suggested that the high population of scrambled intermediates in the rate-limiting step of the BPTI folding pathway may be an artifact by the alkylation reaction which was used to stop the disulfide exchange (Weissman & Kim, 1991). However, there are two reasons why we beleive that the conclusions in this paper are valid. First, during refolding, the apparent population of the different intermediates changes from totally reduced to the components found at equilibrium (Figures 4 and 6; S. Hober, unpublished results). Second, we have alkylated free thiols in the different equilibrium mixtures using 100 mM IAA, which has a well-studied and rapid rate of thiol alkylation in proteins; the rate constant is reported to range from 1 to 16 s^{-1} M^{-1} and is often used to trap intermediates during disulfide exchange protein folding (Creighton, 1984). The mixtures were analyzed by an optimized electrophoresis system for IGF-I (S. Hober, unpublished results) dividing the components into groups according to the number of disulfides. These results were fully compatible with the results presented in Figure 4 and Table I (data not shown). Therefore, we conclude that the pyridylethylation reaction does not cause any major artifacts in the analysis of the equilibrium mixtures and, thus, that the basic conclusions in this paper are valid. However, for the assignments of detailed kinetic roles of the different intermediates, we suggest that the kinetics of the trapping reaction should be validated further, as suggested by the recent reexamination of the folding pathway of BPTI (Weissman & Kim, 1991).

IGF-I belongs to a family of proteins consisting of, in addition to IGF-I, insulin, insulin-like growth factor II (IGF-II), and relaxin (Blundell et al., 1978). These homologous proteins all have the three disulfide bridges in common. The question arises whether the other peptides of this family show the same property as IGF-I, with regard to unfavorable disulfides and two monomeric three-disulfide conformations of equal free energies. For IGF-II, there is a report that there are two different monomeric forms of recombinant IGF-II isolated from *E. coli* (Smith et al., 1989). Even though the disulfide bond patterns of these two different forms have not yet been characterized, they could very well correspond to mismatched and native IGF-II. Insulin and relaxin are not single-chain molecules, making a straightforward comparison difficult. One similarity between insulin (Holmgren, 1979) and IGF-I (Enberg & Holmgren, 1985) is their extreme sensitivity to reductants. However, limited reduction of insulin under some conditions results in the specific breakage of A7–B7 (Busse & Gattner, 1973), which is analogous to the 6–48 disulfide of IGF-I. Most interestingly, a Hg^{2+} ion can specifically be incorporated between the two sulfurs of the A6–A11 disulfide of insulin (Sperling & Steinberg, 1974). This disulfide corresponds to the 47–52 disulfide of IGF-I. Mercuration elongates a disulfide bond by about 3 Å. Many possible explanations for the unexpected specific mercuration of this buried disulfide have been presented, including stabilization of the mercury ion by surrounding amino acid residues (Sperling & Steinberg, 1974). On the basis of the results of IGF-I presented in this paper, we would like to add the hypothesis that the 3-Å elongation relaxes a strain in the molecule.

Is there a biological significance to energetically unfavorable 6–48, 47–52 disulfides? "High energy disulfides" may be present on protein surfaces for reactivity purposes, e.g., in oxidized thioredoxin or protein disulfide isomerase (Gleason & Holmgren, 1988). However, to our knowledge, a native energetically unfavorable buried disulfide, as 47–52 in IGF-I, has not previously been observed in a protein. We are speculating that this disulfide may function in the inactivation of the molecule in vivo. This hypothesis is based on the fact that proteolytic degradation of insulin in the lyzosome is preceded by a reduction of the disulfides (Griffiths & Lloyd, 1979). An alternative hypothesis is that the strained disulfide should act by a disulfide exchange reaction in the receptor binding step. Indeed, disulfide exchange may be significant in the binding of insulin to its receptor since a small fraction of bound insulin molecules were trapped linked through a disulfide bond to the receptor (Clark & Harrison, 1985). A strained disulfide could be very important in such a step.

In conclusion, we have demonstrated that mismatched and aggregated forms of IGF-I, found in recombinant and in vitro refolded material, might be explained by the existence of energetically unfavorable disulfides in the native structure. At present, we are isolating mutant forms of IGF-I lacking each of the three disulfide bridges, respectively, to study their relative biological activities, stabilities, and structures. These studies will give valuable information on the structural–functional significance of the findings presented in this paper.

ACKNOWLEDGMENTS

We are grateful to Drs. Stephen Anderson, Mathias Uhlén, Staffan Josephson, Lars Abrahmsén, and Tomas Moks for valuable discussions. We also thank an unknown and enthusiastic reviewer for helpful suggestions on the manuscript. Dr. Torsten Sejlitz is acknowledged for suggesting the use of vinylpyridine for specific thiol alkylation. Per Persson and Kristina Zachrisson are acknowledged for performing amino acid composition analysis and assisting in the PDMS analysis. Per Kraulis is acknowledged for providing the MOLSCRIPT program.

Registry No. IGF-I, 67763-96-6.

REFERENCES

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

Busse, W.-D., & Gattner, H.-G. (1973) *Hoppe-Seyler's Z. Physiol. Chem. 354*, 147–155.

Clark, S., & Harrison, L. C. (1985) *Biochem. J. 229*, 513–519.

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. (1978) *Prog. Biophys. Mol. Biol. 33*, 231–297.

Creighton, T. E. (1983) in *Functions of Glutathione: Physiological, Toxicological, and Clinical Aspects* (Larsson, A., Ed.) pp 205–213, Raven Press, New York.

Creighton, T. E. (1984) *Methods Enzymol. 107*, 305–329.

Creighton, T. E. (1986) *Methods Enzymol. 131*, 83–106.

Edman, J. C., Ellis, L., Blacher, R. W., Roth, R. A., & Rutter, W. J. (1985) *Nature 317*, 267–270.

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., & Arakava, T. (1990) *J. Protein Chem. 9*, 95–104.

Enberg, G., & Holmgren, A. (1985) *FEBS Lett. 183*, 52–54.

Forsberg, G., Palm, G., Ekebacke, A., Josephson, S., & Hartmanis, M. (1990) *Biochem. J. 271*, 357–363.

Fullmer, C. S. (1984) *Anal. Biochem. 142*, 336–339.

Gleason, F. K., & Holmgren, A. (1988) *FEMS Microbiol. Rev. 54*, 271–298.

Goldenberg, D. P. (1988) *Biochemistry 27*, 2481–2489.

Griffiths, P. A., & Lloyd, J. B. (1979) *Biochem. Biophys. Res. Commun. 89*, 428–434.

Henschen, A. (1986) in *Advanced Methods in Protein Microsequence Analysis* (Wittmann-Liebold, B., Salnikow, J., & Erdmann, V. A., Eds.) pp 244–255, Springer-Verlag, Berlin and Heidelberg.

Holmgren, A. (1979) *J. Biol. Chem. 254*, 9627–9632.

Humbel, R. E. (1990) *Eur. J. Biochem. 190*, 445–462.

Hurle, M. R., Marks, C. B., Kosen, P. A., Anderson, S., & Kuntz, I. D. (1990) *Biochemistry 29*, 4410–4419.

Hurtley, S. M., & Helenius, A. (1989) *Annu. Rev. Cell Biol. 5*, 277–307.

Iwai, M., Kobayashi, M., Tamura, K., Ishii, Y., Yamada, H., & Niwa, M. (1989) *J. Biochem. 106*, 949–951.

Jaenicke, R. (1991) *Biochemistry 30*, 3149–3161.

Johnson, C. W., Jr. (1990) *Proteins: Struct., Funct., Genet. 7*, 205–214.

Kraulis, P. (1991) *J. Appl. Crystallogr. 24*, 946–950.

Marks, C. B., Naderi, H., Kosen, P. A., Kuntz, I. D., & Anderson, S. (1987) *Science 235*, 1370–1373.

Meng, H., Burleigh, B. D., & Kelly, G. M. (1988) *J. Chromatogr. 443*, 183–192.

Moks, T., Abrahmsén, L., Holmgren, E., Bilich, M., Olsson, A., Uhlén, M., Pohl, G., Sterky, C., Hultberg, H., Josephson, S., Holmberg, A., Jörnvall, H., & Nilsson, B. (1987) *Biochemistry 26*, 5239–5244.

Nilsson, B., & Anderson, S. (1991) *Annu. Rev. Microbiol. 45*, 607–635.

Raschdorf, F., Dahinden, R., Maerki, W., Richter, W. J., & Merryweather, J. P. (1988) *Biomed. Mass Spectrom. 16*, 3–8.

Rinderknecht, E., & Humbel, R. E. (1978) *J. Biol. Chem. 253*, 2769–2776.

Rotwein, P. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 77–81.

Samuelsson, E., Wadensten, H., Hartmanis, M., Moks, T., & Uhlén, M. (1991) *Bio/Technology 9*, 363–366.

Saxena, V. P., & Wetlaufer, D. (1970) *Biochemistry 9*, 5015–5021.

Smith, M. C., Cook, J. A., Furman, T. C., & Occolowitz, J. L. (1989) *J. Biol. Chem. 264*, 9314–9321.

Spencer, E. M., Skover, G., & Hunt, T. K. (1988) in *Growth Factors and Other Aspects of Wound Healing: Biological and Clinical Implications* (Alan, R., Ed.) pp 103–116, Liss, Inc., New York.

Sperling, R., & Steinberg, I. Z. (1974) *Biochemistry 13*, 2007–2013.

Srinivasan, N., Sowdhamini, R., Ramakrishnan, C., & Balaram, P. (1990) *Int. J. Pept. Protein Res. 36*, 147–155.

Tang, J.-G., Wang, C.-C., & Tsou, C.-L. (1988) *Biochem. J. 255*, 451–455.

Vértesy, L., Tripier, D., Koller, K.-P., & Riess, G. (1991) *Biol. Chem. Hoppe-Seyler 372*, 187–192.

Weissman, J. S., & Kim, P. S. (1991) *Science 253*, 1386–1393.

Wetlaufer, D. B., Branca, P. A., & Chen, G.-X. (1987) *Protein Eng. 1*, 141–146.

Yang, J. T., Wu, C.-S. C., & Martinez, H. M. (1986) *Methods Enzymol. 130*, 208–269.

Zhang, R., & Snyder, G. H. (1989) *J. Biol. Chem. 264*, 18472–18479.