ELSEVIER

Capture and identification of folding intermediates of cystinyl proteins by cyanylation and mass spectrometry

J. Throck Watson, Y. Yang, and J. Wu

Departments of Biochemistry and Chemistry, Michigan State University, East Lansing, MI, USA

Trapping folding intermediates of cystinyl proteins by covalent modification of free sulfhydryl groups provides the opportunity for isolation, purification, and structural elucidation of individual species. The disulfide structure of the intermediates, coupled with their temporal abundance, provides a 'snapshot' of the pathway experienced by the refolding protein in a particular medium. Here, intermediates of cystinyl proteins containing free cysteines are trapped by cyanylation through reaction with an acidic (pH 3.0) solution of 1-cyano-4-dimethylamino-pyridinium (CDAP) tetrafluoroborate. The cyanylated species are separated by reversed-phase high-performance liquid chromatography, where the resulting chromatogram gives a visual indication of the distribution of intermediates at a designated time after commencing the refolding process. The disulfide structure of an intermediate can be determined by cleaving its cyanylated derivative and by mass mapping of the resulting fragments to the sequence of the original protein. Cleavage of a cyanylated species represented by any given peak in the chromatogram is achieved by treatment of that fraction with 1M NH_4OH at room temperature for 1 h; the resulting fragments are analyzed by matrix-assisted laser desorption ionization (MALDI) or electrospray mass spectrometry. Examples will be presented from in vitro refolding experiments with human epidermal growth factor (hEGF), for which more than 10 folding intermediates were isolated and identified at different time points, and a mutant of insulin-like growth factor-I, for which three intermediates were isolated and identified.

INTRODUCTION

Studies of the refolding behavior of cystinyl proteins offers the special opportunity to chemically trap folding intermediates by chemical methods based on the high and selective reactivity of sulfhydryl groups. Knowledge of the disulfide structure of various intermediates trapped in a time-course manner through the folding regime should provide clues to the transient folding pathway of the protein in the particular medium. The first well-documented disulfide folding pathway is that of bovine pancreatic trypsin inhibitor (BPTI), a small protein comprising 58 amino acids and three disulfides.[1] Characterization of folding intermediates has proven to be important in the elucidation of other protein-folding pathways.[2–5]

Of the several useful techniques available to study protein folding, the technique of pulsed-label NMR [6,7] permits identification of amide groups that are engaged in the structured components. However, intermediates trapped by pulsed NMR are not available for chromatographic purification because of the transient nature of folding intermediates.[8]

Chemical trapping of the intermediates provides the opportunity to isolate, purify, and determine the disulfide structure of individual species. The folding pathways of several proteins[9–12] have been studied in this way. Among them, BPTI [1,13–15] and ribonuclease A[8,16–19] have been most extensively characterized.

In the original model of BPTI folding,[1,20] eight well-populated 1- and 2-disulfide intermediates were identified. Five were shown to contain exclusively native disulfides, and those that adopted non-native disulfides were suggested to be kinetically important intermediates. This original BPTI model was reexamined using different separation and analytical methodologies.[14,15,21] It was concluded that there existed five species of well-populated intermediates (two 1-disulfide and three 2-disulfide species), all of them containing only native disulfide bonds. The inconsistency between the results of the two studies is believed to arise largely from differences in the trapping methods.

Proper trapping of folding intermediates is a critical step in elucidating a folding pathway and studying the associated kinetics. Some criteria proposed for a good trapping agent include that it block quickly, completely, and without modifying the protein at sites other than thiols.[16] They used 2-aminoethyl methanethiosulfonate [$(NH_2)C_2H_5SSO_2CH_3$] as

Corresponding author: J.T. Watson, Department of Biochemistry, Michigan State University, East Lansing, MI 48824, USA.

E-mail address: watsonj@msu.edu (J.T. Watson)

Journal of Molecular Graphics and Modelling 19, 119–128, 2001

655 Avenue of the Americas, New York, NY 10010

1093-3263/01/$–see front matter
PII S1093-3263(00)00127-3

a trapping agent because it reacts specifically and much more quickly than iodoacetate with thiol groups and thus prevents thiol/disulfide bond exchange. Another method uses phenylarsonous acid for trapping protein-folding intermediates by selective modification of bis-cysteine sulfhydryls, forming cyclic dithioesters over a wide pH range for two suitably spaced sulfhydryl groups of a peptide or protein.[22] However, derivatization by this reagent depends too much on the structural accessibility of the bis-cysteine SH groups, which mediate the reaction. Therefore, this reagent has limited practical use.

Traditionally, folding intermediates have been trapped by addition of iodoacetate, a reagent that alkylates free thiols to prevent oxidation or thiol–disulfide exchange. However, rearrangement of intermediates during trapping with iodoacetate has been observed for both BPTI[14] and ribonuclease A[16]; some rearrangement can occur because thiol–disulfide exchange apparently proceeds on the same time scale as alkylation by iodoacetate. This could be a severe problem for partially structured intermediates where steric hindrance might retard the rate of alkylation of some thiols.[23] Although a large molar excess of iodoacetate can be applied to minimize such a side reaction, modification of other functional groups by the high concentration of iodoacetate may provoke other problems.[24]

Acid quenching has been proposed as an alternative technique for providing 'snapshots' of the refolding of cystinyl proteins. Because the thiolate anion is the reactive species in thiol/disulfide exchange, it is possible to quench the folding process extremely rapidly by lowering the pH. Subsequent separation of acid-quenched intermediates by reversed-phase high-performance liquid chromatography (HPLC) at pH 2 can be achieved without introducing significant rearrangements of the trapped species. A practical advantage of acid quenching is its reversibility. As a result, it is possible to purify an acid-quenched intermediate and subsequently allow further rearrangement or folding to occur.[10,14] On the other hand, the intermediates trapped by acid must be chemically modified before the disulfide bond structure can be determined. In this respect, acid quenching is a cumbersome procedure.

In summary, disulfide bond structure elucidation of protein-folding intermediates is still a challenging problem in protein characterization. There is a demand for developing new techniques for trapping, isolating, and characterizing disulfide intermediates.

We recently developed methodology to trap folding intermediates based on the cyanylation of thiol groups by 1-cyano-4-dimethylamino-pyridinium (CDAP) tetrafluoroborate under acidic conditions.[25] This approach has several unique advantages. First, cyanylation of thiol groups in acidic solution quenches the refolding process and minimizes thiol/disulfide exchange. Second, cyanylation of thiol is already part of our procedure for structural elucidation of the intermediates, which involves partial reduction, cyanylation, chemical cleavage, and mass mapping.[25,26] Thirdly, this methodology is fast, simple, and even applicable to disulfide structural analysis of proteins containing adjacent cysteines.[27]

Trapping Intermediates by Cyanylation in Acidic Media

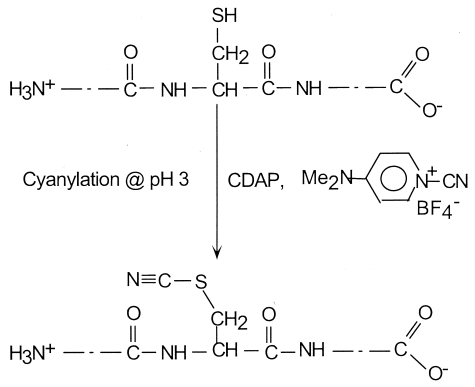
Both our trapping methodology and our structural elucidation procedure rely on the cyanylation of the protein. This involves covalent modification of free sulhydryl groups with a cyanylation reagent, namely, 1-cyano-4-dimethylamino-pyridinium (CDAP) tetrafluoroborate as shown in Scheme 1. The cyanylation reaction is carried out at pH 3 to minimize disulfide exchange during the chemical modification. Importantly, the acidic conditions also quench the folding process as the free sulhydryls are being cyanylated. In this way, the cyanylation procedure combines the advantages of acid quenching and covalent modification of reactive sulfhydryl groups in one step.

The cyanylation reaction is quantitative and selective for free sulhydryl groups. Methinine is not further modified, nor are serine or threonine, nor residues with acidic or basic side chains. The disulfide bond is not modified by the cyanylation reagent. A description of optimal use of the cyanylation reaction in combination with subsequent cleavage of the modified protein has been published elsewhere.[28]

We have employed freshly prepared 0.2 M CDAP to trap intermediates in the refolding of recombinant human epidermal growth factor (hEGF)[25] and insulin-like growth factor.[29] The trapping protocol for these studies is represented in the flowchart shown in Figure 1.

Refolding of hEGF with Trapping by Cyanylation: hEGF is a compact 53-residue protein of the following sequence: (NSDSECPLSHDGYCLHDGVCMYIEALDKYACNCVVGYIGERCQYRDLKWWELR), linked by three disulfide bonds (Cys6—Cys20, Cys14—Cys31, Cys33—Cys42); this potent protein stimulates the proliferation of epidermal cells. We undertook a folding study of hEGF to validate our developing alternative approach to trapping folding intermediates by comparing our results with those in the literature; thus, we followed the folding procedure described by Chang et al.[30]

The refolding of reduced and denatured hEGF (1.0 mg was dissolved in 0.2 ml of 0.1 M citrate buffer, pH 3.0, containing 6 M guanidine-HCl and 0.1 M TCEP reducing agent at 37°C for 2 h; HPLC was used to isolate the reduced/denatured species) was initiated by diluting the protein sample with 0.05 M Tris-HCl buffer (pH 8.5) to a final protein concentration of 1 mg/ml. The protein was subjected to folding under open air conditions to provide a valid basis for comparing our results with those in the literature.[30] Folding intermediates were trapped in a time-course manner by removing aliquots (0.1 ml) of the protein solution and mixing them with 10 μl of 0.5 M HCl solution containing freshly prepared 0.2 M CDAP. The pH of the solution was adjusted to 3, if necessary, and cyanylation of free sulfhydryl groups by CDAP proceeded at room temperature for 15 min.



Scheme 1

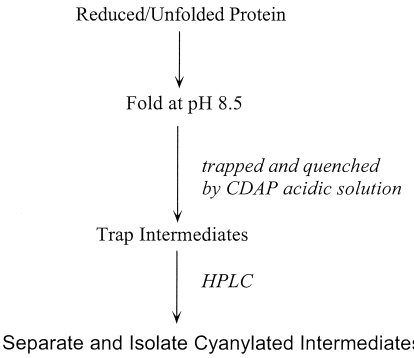


Figure 1. Scheme cyanylation approach for trapping intermediates during folding of cystinyl proteins.

The intermediates of refolding hEGF trapped by CDAP at designated time points were immediately separated by HPLC to give the profiles shown in Figure 2. The fractions were collected manually and analyzed by MALDI-TOF MS. Those with 0-Da, 50-Da, 100-Da, 150-Da increases over the mass of the reduced protein correspond to the 3-disulfide (non-native, III, or native, N), 2-disulfide (II-), 1-disulfide (I-), and 0-disulfide (R) species, respectively. The fractions corresponding to various intermediates were collected, dried in a speed-vac, and stored in a freezer at −80°C for subsequent cleavage, complete reduction, and analysis by mass spectrometry.

Refolding of IGF and a Mutant with Trapping by Cyanylation: Insulin-like growth factor-I (IGF-I)[31] is a single-chain polypeptide of 70 residues containing three intramolecular disulfide bonds, two of which involve adjacent cysteines

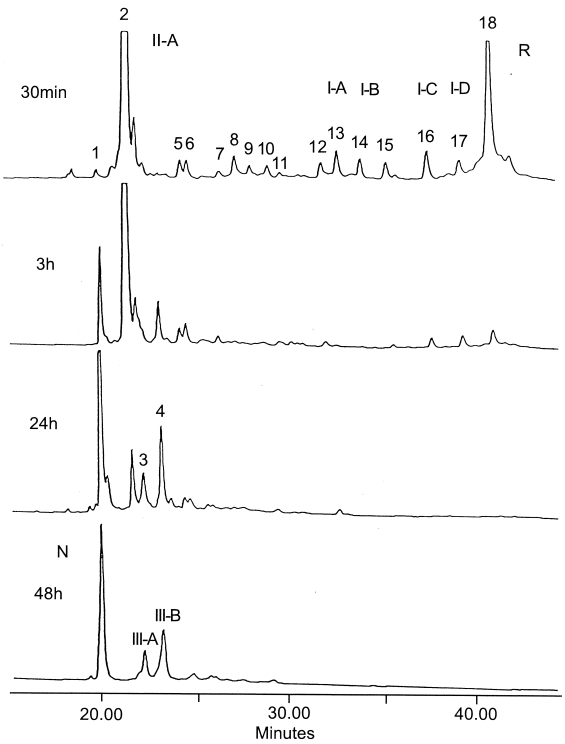


Figure 2. HPLC chromatogram of cyanylated intermediates of hEGF trapped after indicated periods of folding.

(Figure 3). Recombinant human long R^3 insulin-like growth factor-I (LR^3IGF-I) is a variant of human insulin-like growth factor-I (IGF-I) in which glutamate 3 is replaced by arginine, and a 13-residue extension appears at the N-terminus (Figure 3). LR^3IGF-I is substantially more potent than IGF-I in affecting carbohydrate metabolism and in stimulating the growth of fetal tissue in animals.[32]

Reduction of the proteins was carried out at 37 °C for 2 h by dissolving 0.1mg of LR^3IGF-I or IGF-I in 0.5 mL citrate buffer, pH 3.0, containing 6M guanidine-HCl and 0.1M TCEP reducing agent; the reduced/denatured LR^3IGF-I or IGF-I was purified by HPLC, dried under reduced pressure, and stored at −70°C. The refolding of LR^3IGF-I or IGF-I was initiated by diluting the reduced/unfolded protein sample with 0.10M Tris-HCl buffer (pH 8.7), containing 1mM GSSG, 10mM GSH, 0.2M KCl, and 1mM EDTA, to a final protein concentration of 0.1mg/ml.

Refolding intermediates were trapped in a time-course manner by removing aliquots (0.1mL) of protein solution and mixing with approximately 0.02mL of 1.0M HCl containing freshly prepared 0.2M CDAP to give a solution of pH 2∼3. Cyanylation of free thiol groups by the CDAP proceeded at room temperature for 10 min. The trapped intermediates (covalently modified by cyanylation) were immediately separated by HPLC on a 10-μm particle size, 300-Å pore, 4.6 × 250mm Vydac C18 column (Solvent A was 0.1% aqueous TFA; Solvent B was acetonitrile/water [9:1, v/v] containing 0.1% TFA. The linear gradient was 30–50% solvent B in 45 min at a flow rate of 1mL/min). The HPLC fractions were collected manually and an aliquot of each analyzed by MALDI-TOF MS to determine the number (as indicated by the mass shift) of disulfide bonds as described above for hEGF.

The temporal distribution of intermediates during the refolding process is represented by the HPLC chromatograms (see Figure 4) of cyanylated species trapped at designated times. Figures 4a and 4b show an array of chromatograms of LR^3IGF-I and IGF-I intermediates trapped by reaction with CDAP under acidic conditions at various times after initiating refolding in the GSSG/GSH buffer. R and N represent the reduced and native protein; 1-, 2-, and 3-disulfide containing intermediates of LR^3IGF-I and IGF-I are represented by I′, I, II′, II, III′, and III, respectively.

Determination of Disulfide Structure by Cyanylation/Cleavage/Mass Mapping

The conventional approach for recognition of disulfide bond structures of folding intermediates is tedious and cumbersome, especially for large proteins where proteolytic digestion produces a large number of fragments that are irrelevant to the disulfide linkage. Further, the conventional approach, which is based on proteolysis, is impractical, if not unsuitable, for proteins containing closely spaced or adjacent cysteines, where proteolytic and/or chemical degradation cannot achieve cleavage between cysteine residues.[33]

We have developed a novel methodology for assigning disulfide bond pairings in proteins based on the mass mapping of chemical cleavage fragments to the sequence of the protein.[26] The key feature is specific chemical cleavage of the peptide backbone on the N-terminal side of a cyanylated cysteine as shown in Scheme 2. Analogous to the strategy used in proteolytic mapping of peptides, having a knowledge of the

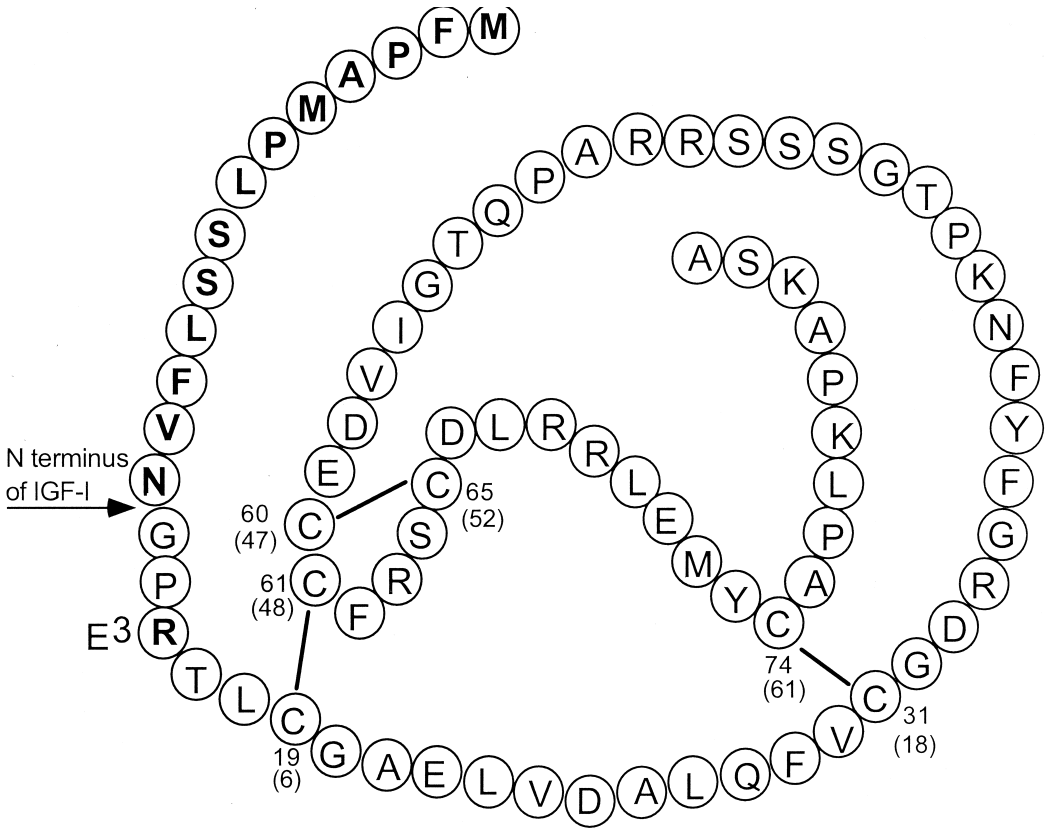


Figure 3. An overview of the amino acid sequence and disulfide structures of LR³IGF-I and IGF-I. The numbers indicate the location of cysteine residues in LR³IGF-I; those in parentheses indicate the locationof cysteine residues in IGF-I. Taking glycine at the N-terminus of IGF-I as a reference point, a 13-residue extension in LR³IGF-I is represented by the string of 13 bold letters. The bold **R** *in LR³IGF-I indicates the replacement of E at position 3 in IGF-I.*

sequence and the specificity of the cleavage reaction, it's possible to predict the number and mass of the cleavage fragments. In this case, the cleavage reaction, initiated by nucleophilic attack of OH^- on **n** cyanylated cysteines, results in the formation of **n+1** fragments, one of which is a simple truncated peptide together with **n** iminothiazolidine-blocked (itz) peptides.

The analytical strategy for determining the connectivity of cysteines in the folding intermediates containing only one disulfide bond is to focus attention on the location of the free cysteines, not directly on those involved in the disulfide bond. Cleavage is expected on the N-terminal side of each free cysteine as these will be cyanylated; no cleavage is expected directly at the site of the intact disulfide bond, as these cysteines are not available for cyanylation. For example, from mass analysis of an aliquot of the fraction represented by peak 13 in Figure 1, it could be deduced (from a mass shift of 100 Da from the mass of the completely reduced species R) that this folding intermediate of hEGF contained only one disulfide bond. It is known that hEGF contains six cysteines; hence, determining the location of the four free cysteines will allow the two cysteines involved in the disulfide bond to be deduced. For a protein containing six cysteine residues, there are fifteen possible 1-disulfide intermediates, each of which would yield five fragments after cleavage; the m/z values for each fragment corresponding to cleavage at cyanylated cysteine residues in each possible isomer can be calculated.[25]

The MALDI spectra of the cleavage products of four 1-disulfide intermediates (represented by peaks 13, 14, 16, and 17 in Figure 2) are shown in Figure 5. Peaks in Figure 5A can be assigned to cleavage products consisting of residues 1–13 as a simple truncated peptide, and to three iminothiazolidine-blocked (itz) peptides including itz-14–30, itz-33–41, and itz-42–53. The sequence of the hEGF (right column, page 120) reveals Cys at positions 6, 14, 20, 31, 33, and 42. Recognition of MALDI-MS peaks for itz-14–30, itz-33–41, and itz-42–53 indicates that residues 14, 33, and 42 are free cysteine residues in the intermediate. Ideally, a MALDI-MS peak for one of the other three Cys residues should have been observed. Cleavage at the cyanylated Cys31 would be the most likely as it is a complement to the formation of itz-14–30; such cleavage would generate the fragment itz-31–32, which could easily be hidden in the peaks for the matrix background or could be signal-suppressed due to its small mass. Further, observing peaks for 1–13, itz-14–30, itz-31–32, itz-33–41, and itz-42–53 would provide a complete set, covering the complete sequence of residues. This internal consistency expected in the data set allows recognition of 'imperfect data' such as the peak corresponding in mass to itz-33–53. Note that this peak is predominant in the next MALDI spectrum (I-B) in Figure 5; its presence in I-A is rationalized as cross-contamination of a species that is exceptionally sensitive to MALDI.[25]

Conspicuous by their absence in the MALDI spectrum (I-A) in the top panel of Figure 5 are peaks corresponding to cleav-

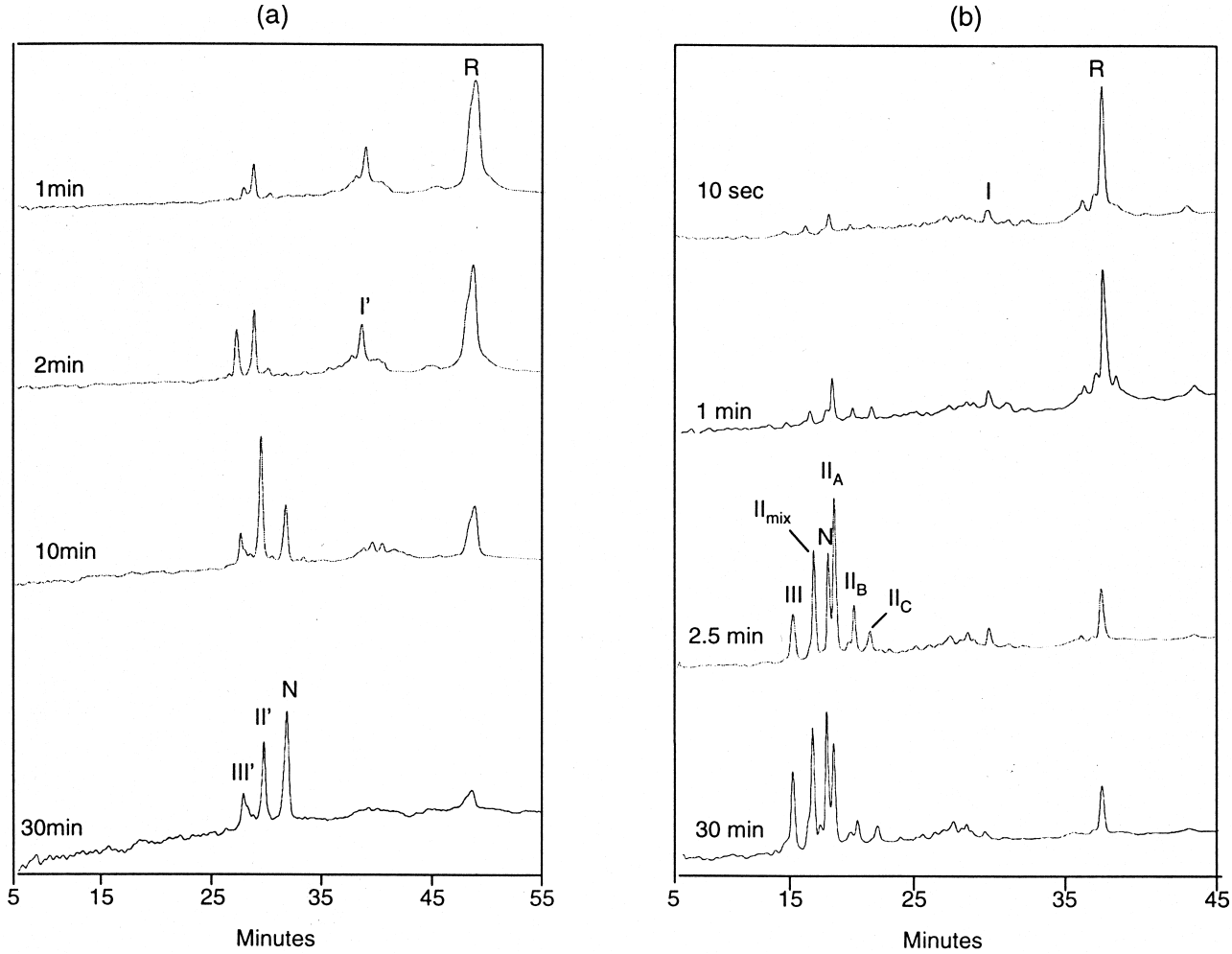
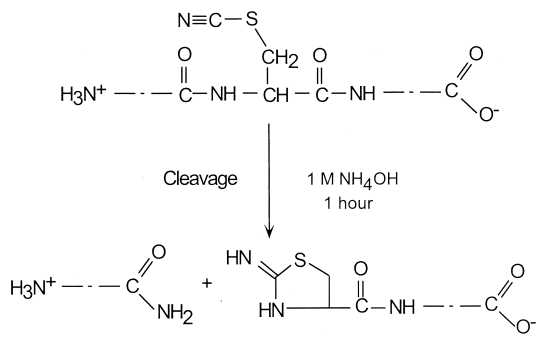


Figure 4. HPLC chromatogram showing the time-dependent distribution of CDAP-trapped intermediates during the refolding of (a) LR[3]IGF-I and (b) IGF-I. See experimental section for HPLC conditions.

age on the N-terminal side of Cys 6 or Cys 20. This observation, combined with observations leading to recognition of the location of four free cysteine residues, makes it possible to deduce the connectivity of Cys6—Cys20 for the only disulfide bond.

The Need for Partial Reduction of Multi-Cystinyl Proteins: When more than one disulfide bond is present in a protein, some measure is necessary to specifically reduce a particular disulfide bond so that the cyanylation and subsequent cleavage chemistry can be applied to the nascent free sulfhydryls, thereby revealing their location within the sequence (Scheme 3).



Scheme 2

An example of structural elucidation of a multi-cystinyl intermediate can be shown in the context of the LR[3]IGF-I refolding experiment. In Figure 4, the peak labeled II′ was determined to represent a 2-disulfide species by virtue of its mass being 52 Da greater than that of the native protein (i.e., the intermediate had two free cysteines). The analytical strategy called for identification of the location of the two free cysteines, followed by partial reduction of the 2-disulfide system to obtain information on connectivity.

Because the compound represented by peak II′ in Figure 4 is a cyanylated species, an aliquot of this fraction could be subjected directly to cleavage conditions (pH 12) to produce cleavage fragments (after total reduction of residual disulfides) that would reveal the location of the free (now cyanylated) cysteines. Analysis of the cleavage fragments by MALDI MS gave the mass spectrum shown in Figure 6a, which shows peaks corresponding to 1-59, itz-60-(65)-83, itz-60-(65*)-83, and itz-65-83. Recognizing from the sequence of LR[3]IGF-I that there are cysteines at positions 19, 31, 60, 61, 65, and 74, the

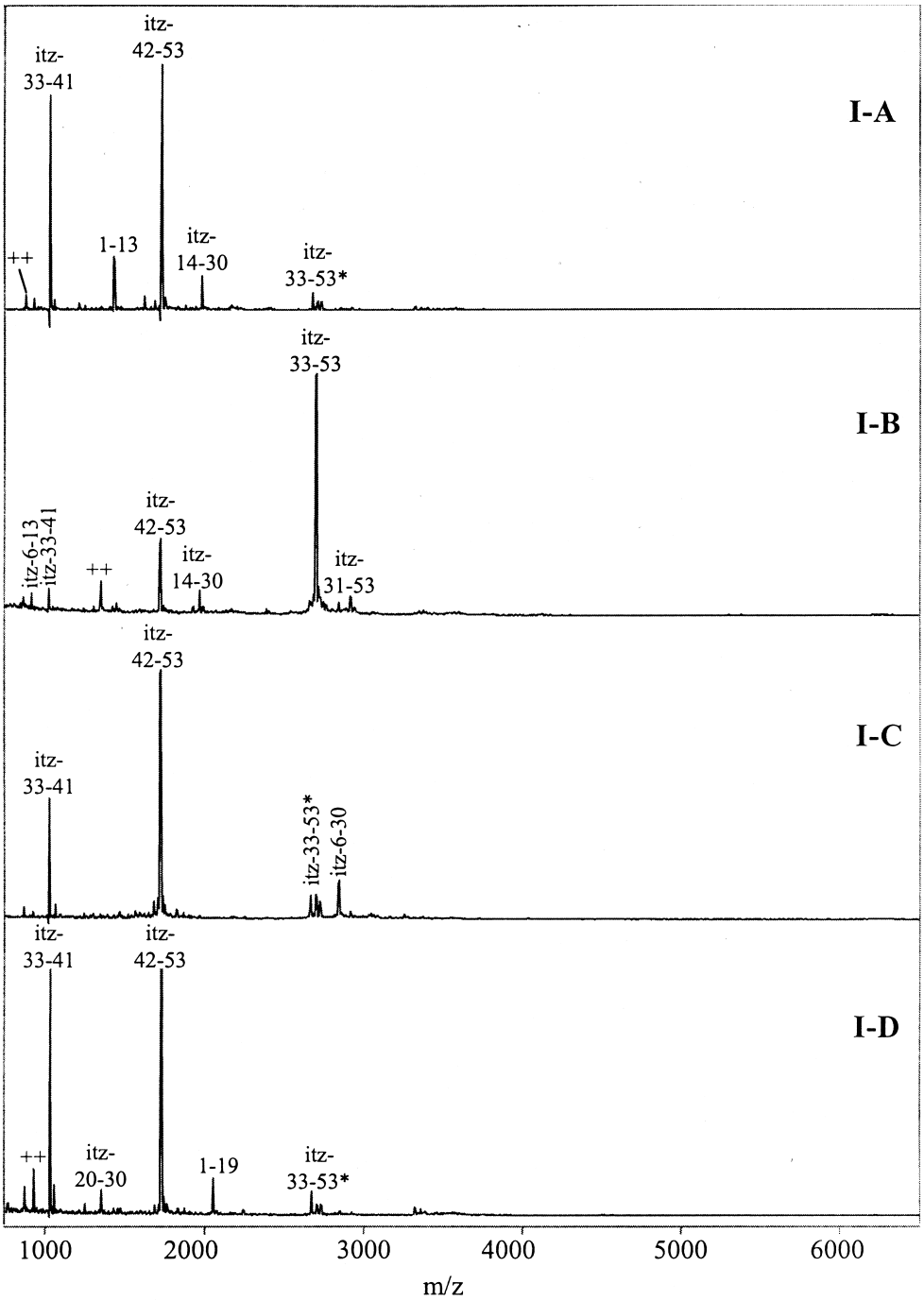
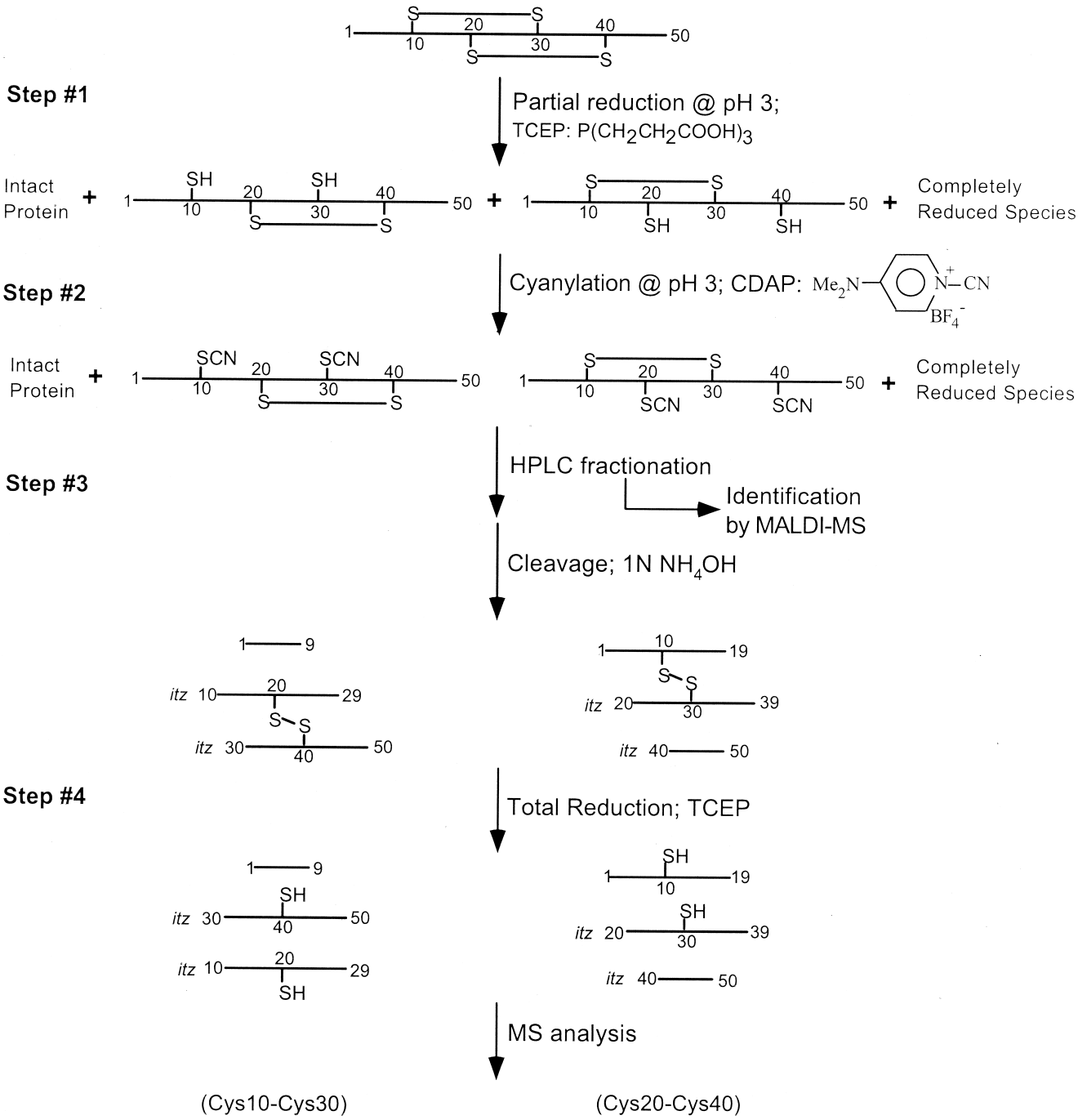


Figure 5. MALDI mass spectra of the 1-disulfide folding intermediates of hEGF indicated by the same code as in Figure 2. itz = iminothiazolidine derivative.

mass spectral data indicate that there are free cysteines at positions 60 and 65 in this intermediate.

The remainder of the LR^3IGF-I intermediate represented by II′ in Figure 4 was subjected to partial reduction and the resulting mixture of compounds then were cyanylated.[27] Analysis of the cyanylated partial reduction mixture by HPLC gave the chromatogram shown in Figure 7, in which peak 1 represents the intact cyanylated 2-disulfide intermediate (same as II′ in Figure 4) and peak 2 represents a cyanylated 1-disulfide species (as determined by a mass shift of 104 Da from the mass of LR^3IGF-I, data not shown). (It is generally expected that two singly reduced species will be produced from a precursor containing two disulfides; only one was detected in this case, presumably because one disulfide is so much more stable than the other.)

The fraction corresponding to peak 2 in Figure 7 (the cyanylated derivative of a singly-reduced isoform of II′ see Figure 4) was subjected to cleavage conditions, and the resulting mixture of cleavage products was analyzed by MALDI MS to give the mass spectrum shown in Figure 6b. Peaks in Figure 6b



Scheme 3

correspond to 1-18, itz-19-59, itz-61-83, itz-61-83-SH, and itz-65-83. All members of the sequence of LR^{3}IGF-I are represented in the mass spectrum in Figure 6b, except residue 60. It is not surprising that a peak is not observed for itz-60 because it would surely suffer interference from the matrix background. Further, from direct cleavage analysis of cyanylated II′ as described above, we know that Cys60 and Cys65 are free in the unreduced species. The new information from interpretation of the mass spectrum in Figure 6b of the cyanylated singly-reduced isoform of II′ is that Cys19 and Cys61 are also free in this species; these new data indicate that Cys19 and Cys61 are connected by a disulfide bond in II′. By default, the remaining two cysteines, Cys31 and Cys74, are also connected by a disulfide bond in II′.

FURTHER DISCUSSION

The chromatograms in Figure 4, which represent a temporal distribution of intermediates during the refolding of hEGF, agree well with those reported in the literature.[30] We have been able to determine the structure of ten intermediates (see Figure 8), of which I-A and I-B are newly reported.[25]

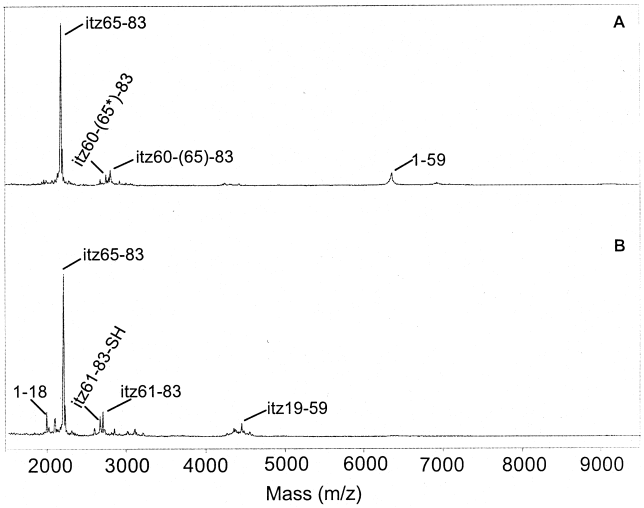


*Figure 6. The MALDI mass spectra of : (a) fragments resulting from cleavage of the purified/cyanylated intermediate represented by HPLC peak II′ in Figure 4a; (b) fragments resulting from cleavage of the purified/cyanylated singly-reduced isoform of II′ as represented by HPLC peak 2 in Figure 7. itz = iminothiazolidine derivative; * indicates beta-elimination product.*

The folding intermediates of IGF-I trapped by CDAP and their distribution pattern are similar to those obtained by acid trapping under the same refolding conditions.[34,35] By mass-mapping the cleavage products resulting from cyanylated intermediates, the disulfide structures of six well-populated intermediates were identified (as summarized in Table 1), including a native one-disulfide intermediate (I), a native two-disulfide intermediate (II_A); two non-native two-disulfide intermediates (II_B and II_C), a mismatched intact protein (III), and a mixed two-disulfide intermediate with glutathione (II_{mix}). The intermediates II_B and II_{mix} were not observed by vinylpyridine trapping.[12] Furthermore, intermediate I was the major form captured by pyridylethylation,[12] but it was the minor form in the distribution of the folding intermediates trapped by the CDAP approach described here, and also by acid trapping as reported elsewhere.[34,35]

Unlike the refolding of hEGF, which was conducted under 'open-air' conditions, the refolding of LR^3IGF-I or IGF-I was performed in the presence of a GSH/GSSH redox couple.

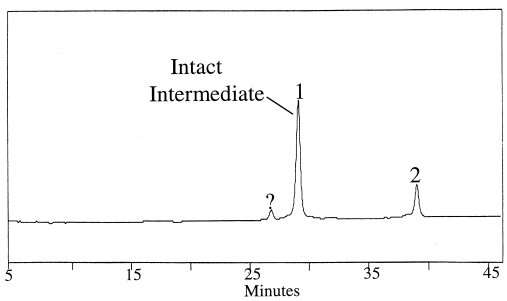


Figure 7. HPLC separation of intact intermediate II′ (peak 1) and one of its singly-reduced isoforms (Peak 2).

Table 1. Disulfide linkage of intermediates trapped during the refolding and reductive unfolding of IGF-I and LR^3IGF-I. The underlined residues are nonnative disulfides; primed Roman numerals relate to LR^3IGF-I

Protein	HPLC Peak	Disulfide Structure
IGF-I	I	18-61
	II_A	18-61, 6-48
	II_B	18-61, 6-47
	II_C	18-61, 6-52
	II_{mix}	Mixed two disulfides
	III	18-61, 6-47, 48-52
	N	18=61, 6-48, 47-52
LR^3IGF-I	III′	31-74
	II′	31-74, 19-61
	I′	31-4, 19-60, 61-65
	N	31-74, 19-61, 60-65

Adjustment of the GSH/GSSH redox ratio had a substantial effect on the equilibrium distribution of residual intermediates and endproducts as well as on the kinetics of their formation.[27]

SUMMARY

We have described here the integrated combination of two techniques, a disulfide trapping technique and our disulfide mapping technique, for the capture and identification of folding intermediates of cystinyl proteins. This integrated method provides several important advantages. First, the trapping reaction, occurring in acidic solution, greatly reduces the risk of sulfhydryl/disulfide exchange that may accompany iodoacetate trapping. Second, the CDAP quantitatively cyanylates free sulfhydryl groups to stop further folding. Unlike the acid-quenching technique, which is totally reversible and requires further chemical modification to determine the disulfide structure of the folding intermediates, the cyanylation technique captures intermediates in a form that can be subjected directly to HPLC and mass spectrometry for characterization of disulfide structures. Third, our disulfide bond mapping technique itself provides unique advantages, such as applicability to cystinyl proteins containing closely arranged and even adjacent cysteines, over conventional techniques, as described previously.[26] Our results for refolding hEGF are in good agreement with the literature and demonstrate the feasibility of the cyanylation methodology for trapping and identifying folding intermediates. Our results for the refolding of LR^3IGF-I and IGF-I not only provided a comparison of the mutant and wild type, but characterized the folding behavior of LR^3IGF-I for the first time. In general, the cyanylation methodology is shown to be simpler, faster, and more sensitive than the conventional approach of capturing and identifying folding intermediates of cystinyl proteins.

ACKNOWLEDGMENTS

This work was supported by the Biotechnology Research Technology Program (BRTP) of the National Center for Research Resources of the National Institutes of Health (RR-00480).

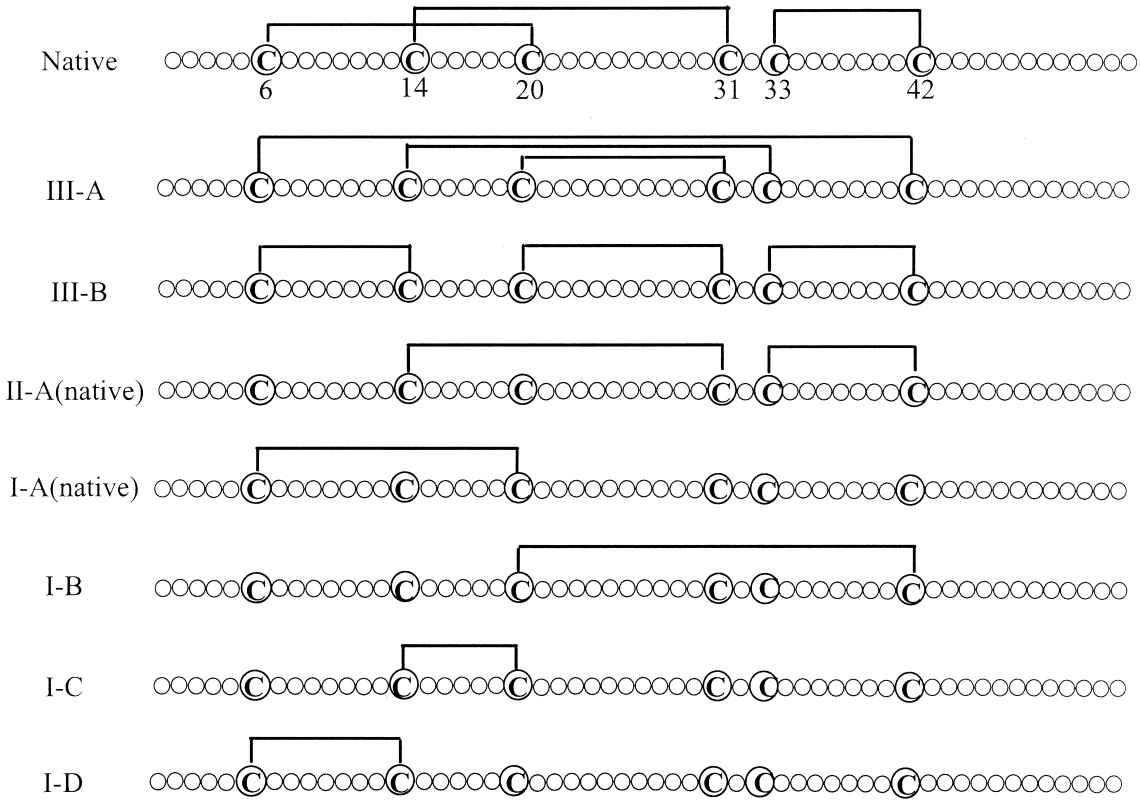


Figure 8. Disulfide structures of hEGF and its seven well-populated folding intermediates.

REFERENCES

1 Creighton, T.E., and Goldenberg, D.P. Kinetic role of a meta-stable native-like two-disulphide species in the folding transition of bovine pancreatic trypsin inhibitor. *J. Mol. Biol.* 1984, **179**, 497–526

2 Kim, P.S., and Baldwin, R.L. Intermediates in the folding reactions of small proteins. *Ann. Rev. Biochem.* 1990, **59**, 31–660

3 Matthews, C.R. Pathways of protein folding. *Ann. Rev. Biochem.* 1993, **62**, 653–683

4 Ptitsyn, O.B. Structures of folding intermediates. *Curr. Opin. Struct. Biol.* 1995,**5**, 74–78

5 Li, Y.J., Rothwarf, D.M., and Scheraga, H.A. Mechanism of reductive protein unfolding. *Nature Struct. Biol.* 1995, **2**, 489–494

6 Jennings, P.A., and Wright, P.E. Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin. *Science* 1993, **262**, 892–896

7 Balbach, J., Forge, V., van Nuland, N.A.J., Winder, S.L., Hore, P.J., and Dobson. C.M. Following protein folding in real time using NMR spectroscopy. *Nature Struct. Biol.* 1995, **2**, 865–870

8 Xu, X., Rothwarf, D.M., and Scheraga, H.A. Nonrandom distribution of the one-disulfide intermediates in the regeneration of ribonuclease A. *Biochemistry* 1996, **35**, 6406–6417

9 Creighton, T.E., and Ewbank, J.J. Disulfide-rearranged molten globule state of alpha-lactalbumin. *Biochemistry* 1994, **33**, 1534–1538

10 Chang, J.Y. The disulfide folding pathway of tick anticoagulant peptide (TAP), a Kunitz-type inhibitor structurally homologous to BPTI. *Biochemistry* 1996, **35**, 11702–11709

11 Glocker, M.O., Arbogast, B., Milley, R., Cowgill, C., and Deinzer, M.L. Disulfide linkages in the *in vitro* refolded intermediates of recombinant human macrophage-colony-stimulating factor: analysis of the sulfhydryl alkylation of free cysteine residues by fast-atom bombardment mass spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* 1994, **91**, 5868–5872

12 Hober, S., Forsberg, G., Palm, G., Hartmanis, M, and Nilsson B. Disulfide exchange folding of insulin-like growth factor I. *Biochemistry* 1992, **31**, 1749–1756

13 Darby, N.J., Norin, P.E., Talbo, G., and Creighton, T.E. Refolding of bovine pancreatic trypsin inhibitor via nonnative disulphide intermediates. *J. Mol. Biol.* 1995, **249**, 463–477

14 Weissman, J.S., and Kim, P.S. Reexamination of the folding of BPTI: predominance of native intermediates. *Science* 1991, **253**, 1386–1393

15 Weissman, J.S., and Kim, P.S. Kinetic role of nonnative species in the folding of bovine pancreatic trypsin inhibitor. *Proc. Natl. Acad. Sci. U.S.A.* 1992, **89**, 9900–9904

16 Rothwarf, D.M, and Scheraga, H.A. Regeneration of bovine pancreatic ribonuclease A. 1. Steady-state distribution. *Biochemistry* 1993, **32**, 2671–2679

17 Rothwarf, D.M., and Scheraga, H.A. Regeneration of bovine pancreatic ribonuclease A. 2. Kinetics of regeneration. *Biochemistry* 1993,**32**, 2680–2689

18 Rothwarf, D.M., and Scheraga, H.A. Regeneration of bovine pancreatic ribonuclease A. 3. Dependence on the

nature of the redox reagent. *Biochemistry* 1993, **32**, 2690–2697
19 Rothwarf, D.M., and Scheraga, H.A. Regeneration of bovine pancreatic ribonuclease A. 4. Temperature dependence of the regeneration rate. *Biochemistry* 1993, **32**, 2698–2703
20 Creighton, T.E. The disulfide folding pathway of BPTI. *Science* 1992, **256**, 111–114
21 Weissman, J.S., and Kim, P.S. The pro region of BPTI facilitates folding. *Cell* 1992, **71**, 841–851
22 Happersberger, P., Cowgill, C., Deinzer, M.L., and Glocker, M.O. *Proceedings of the 44th ASMS Conference on Mass Spectrometry and Allied Topics*, Portland, OR, May, 1996, p. 1050.
23 Goto, Y., and Hamaguchi, K. Formation of the intrachain disulfide bond in the constant fragment of the immunoglobulin light chain. *J. Mol. Biol.* 1981, **146**, 321–340
24 Torella, C., Ruoppolo, M., Marino, G., and Pucci, P. Analysis of RNase A refolding intermediates by electrospray/mass spectrometry. *FEBS Letters* 1994, **352**, 301–306
25 Wu, J., Yang, Y., and Watson, J.T. Trapping of intermediates during the refolding of recombinant human epidermal growth factor (hEGF) by cyanylation, and subsequent structural elucidation by mass spectrometry. *Protein Sci.* 1998, **7**, 1017–1028
26 Wu, J., and Watson, J.T. A novel methodology for assignment of disulfide bond pairings in proteins. *Protein Sci.* 1997, **6**, 391–398
27 Yang, Y., Wu, J., and Watson, J.T. Disulfide mass mapping in proteins containing adjacent cysteines is possible with cyanylation/cleavage methodology. *J. Am. Chem. Soc.* 1998, **120**, 5834–5835
28 Wu, J., and Watson, J.T. Optimization of the cleavage reaction for cyanylated cysteinyl proteins for efficient and simplified mass mapping. *Anal. Biochem.* 1998, **258**, 268–276
29 Yang, Y., Wu, J., and Watson, J.T. Probing the folding pathways of long (R^3)-insulin-like growth factor-I (LR^3IGF-I) and IGF-I via capture and identification of disulfide intermediates by cyanylation methodology and mass spectrometry. *J. Biol. Chem.* 1999, **274(53)**, 37598–37604
30 Chang, J.Y., Schindler, P., Ramseier, U., and Lai, P.H. The disulfide folding pathway of human epidermal growth factor. *J. Biol. Chem.* 1995, **270(16)**, 9207–9216
31 Rinderknetcht, E., and Humbel, R.E. *J. Biol. Chem.* 1978, **253**, 2769
32 Tomas, F.M., Knowles, S.E., Chandler, C.S., Francis, G.L., Owens, P.C., and Ballard, F.J. *J. Endocrinol.* 1995,**137**, 413–421
33 Smith, D.L., and Zhou, Z.R. Strategies for locating disulfide bonds in proteins. *Methods Enzymol.* 1990, **193**, 374–389
34 Miller, J.A., Narhi, L.O., Hua, Q.X, Rosenfeld, R., Arakawa, T., Rohde, S.P., Lauren, S., Stoney, K.S., Tsai, L., and Weiss, M.A. Oxidative refolding of insulin-like growth factor 1 yields: Two products of similar thermodynamic stability *Biochem.* 1993, **32**, 5203–5213
35 Hua, Q.X., Narhi, L., Jia W., Arakawa, T., Rosenfeld, R., Hawkins, N, Miller, J.A., and Weiss, M. Native and non-native structure in a protein-folding intermediate: Spectroscopic studies of partially reduced IGF-I and an engineered alanine model. *J. Mol. Biol.* 1996, **259**, 297–313