Disulfide Mass Mapping in Proteins Containing Adjacent Cysteines Is Possible with Cyanylation/Cleavage Methodology

Ying Yang,[†] Jiang Wu,[†] and J. Throck Watson*[,†,‡]

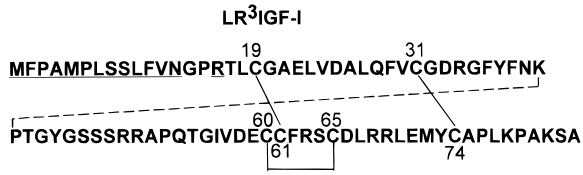
Departments of Chemistry and Biochemistry
Michigan State University
East Lansing, Michigan 48824-1319

Received October 16, 1997

Conventional methodology for determining the pairing in the disulfide bonding structure of proteins relies on the use of appropriate proteases to cleave the peptide backbone between the cysteine residues. Labor-intensive efforts are then pursued to isolate and sequence the various cysteinyl peptides in an effort to deduce the pairing of the disulfide bonds in the original protein. A major constraint is the requirement for a cleavage site for the enzyme between all cysteine residues, especially in cases where the cysteines reside close to one another in the sequence because the likelihood of a proteolytic cleavage site becomes less likely. Proteins containing adjacent cysteines are usually refractory to conventional methodology,[1] although dissociation of cystinyl peptide ions has found some success in deciphering cysteine connectivities in such cases.[2,3]

We have recently described a novel method for determining disulfide pairing by cyanylating the protein, which leads to selective cleavage of the peptide backbone at the N-terminal side of *S*-cyano-cysteine residues.[4] The cyanylation methodology used in conjunction with the technique of partial reduction[5] allows one to deduce disulfide bond pairing from mass analysis of the cleavage products.

Here, we demonstrate the applicability of our cyanylation/cleavage/mass mapping approach to the structural analysis of long insulin-like growth factor I (LR^3IGF-I), a protein consisting of 83 residues and three disulfide bonds, two of which involve adjacent cysteines. LR^3IGF-I is a mutant of insulin-like growth factor I (IGF-I), having a 13-residue (long, L) extension at the N-terminus (those underlined in the primary structure) and an arginine (R, also underlined) in position 3, *in lieu* of a glutamic acid residue in IGF-I. LR^3IGF-I is more potent than IGF-I in affecting carbohydrate metabolism and the growth of fetal tissue.[6] The disulfide bond assignments in LR^3IGF-I are presumed to be homologous with those in IGF-I, which were modeled to those in insulin after determination by X-ray crystallography.[7]



When the protein being characterized contains more than one disulfide bond, partial reduction is employed to generate singly reduced isoforms of the protein, each containing a different reduced disulfide bond. An overview of our analytical procedure

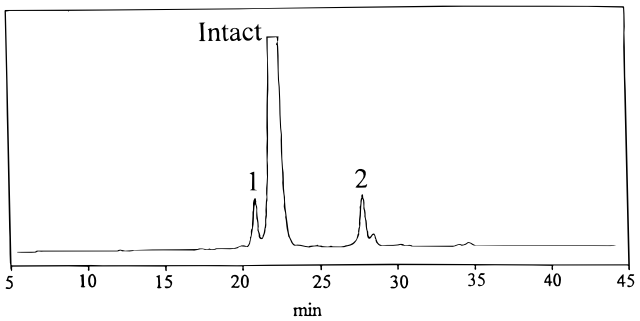
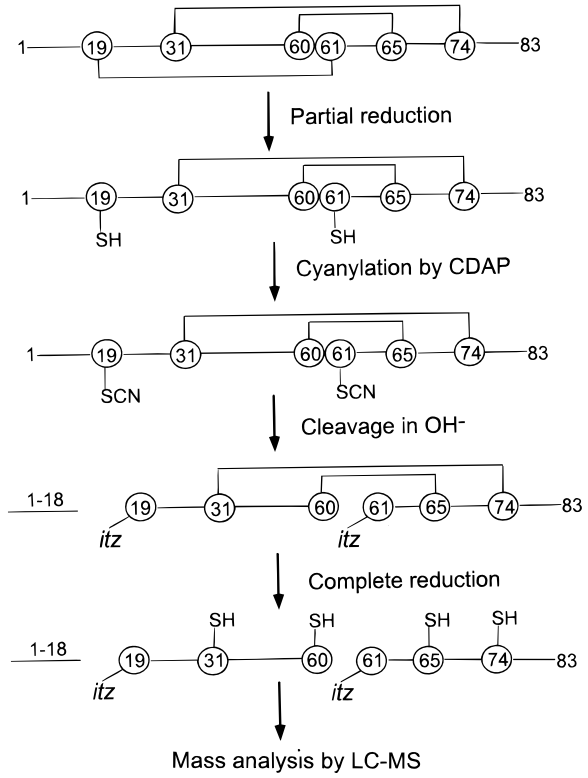


Figure 1. HPLC chromatogram of denatured LR^3IGF-I ("Intact") and its partially reduced/cyanylated isomers. Separation was carried out on a Vydac C18 column at a flow rate of 1.0 mL/min with a linear gradient of 30−50% B in 45 min, where $A = 0.1\%$ TFA in water and $B = 0.1\%$ TFA in CH_3CN. Peaks 1 and 2 represent singly reduced/cyanylated species as determined by electrospray (ESI).

Scheme 1



is illustrated in Scheme 1, which shows a singly reduced isoform of LR^3IGF-I in which the disulfide bond between cysteine 19 and cysteine 61 has been reduced. Once Cys 19 and Cys 61 are cyanylated, base-catalyzed cleavage can occur on the N-terminal side leading to the formation of a free peptide consisting of residues 1−18 and a much larger fragment in which Cys 19 and Cys 61 have been converted to an iminothiazolidine derivative represented as "itz" in the scheme. Upon complete reduction of the reaction mixture, three fragments result as illustrated at the bottom of Scheme 1, the mass analysis of which allows one to conclude that Cys 19 must have been connected to Cys 61 in the original protein.

Together with a knowledge of the sequence of LR^3IGF-I, the data in the following figures from treatment of 10 nmol of LR^3-IGF-I according to the protocol in Scheme 1 demonstrate that the difficult structural problem presented by adjacent cysteines is amenable to the cyanylation/cleavage/mass mapping methodol-

[†] Department of Chemistry.
[‡] Department of Biochemistry.

(1) Smith, D. L.; Zhou, Z. Strategies for locating disulfide bonds in proteins. In *Methods Enzymol.* **1990**, *193*, 374−389.
(2) Raschdorf, F.; Dahinden, R.; Maerki, W.; Richter, W. J.; Merryweather, J. P. *Biomed. Environ. Mass Spectrom.* **1988**, *16*, 3−8.
(3) Jones, M. D.; Patterson, S. D.; Lu, H. S. *Anal. Chem.* **1998**, *70*, 136−143.
(4) Wu, J.; Watson, J. T. *Protein Sci.* **1997**, *6*, 391−398.
(5) Gray, W. R. *Protein Sci.* **1993**, *2*, 1732−1748.
(6) Tomas, F. M.; Knowles, S. E.; Chandler, C. S.; Francis, G. L.; Owens, P. C.; Ballard, F. J. *J. Endocrinol.* **1993**, *137*, 413−421.

S0002-7863(97)03601-9 CCC: $15.00
Published on Web 05/29/1998

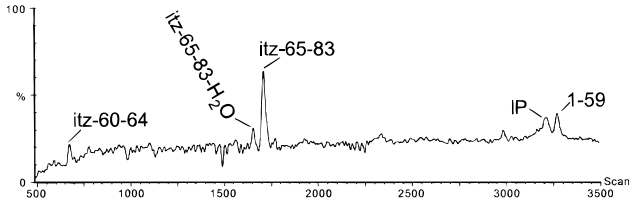


Figure 2. RTIC from LC-MS (ESI) of the peptide mixture resulting from cleavage of the cyanylated isoform represented by HPLC peak 1 in Figure 1. Separation was carried out on a Perkin-Elmer C18 capillary column at a flow rate of 7 μL/min with a linear gradient of 5−65% B in 135 min, where A = 0.085% TFA in water and B = 0.085% TFA in CH_3CN.

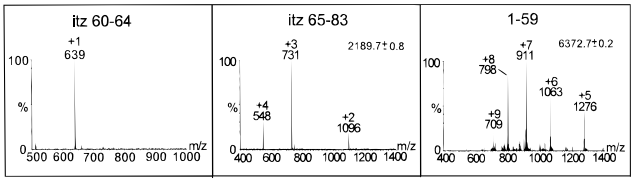


Figure 3. The ESI mass spectra of the peptides corresponding to the RTIC peaks in Figure 2. Mass spectra were obtained with a Micromass Platform-LC at a cone voltage of 30 V and a source temperature of 80 °C.

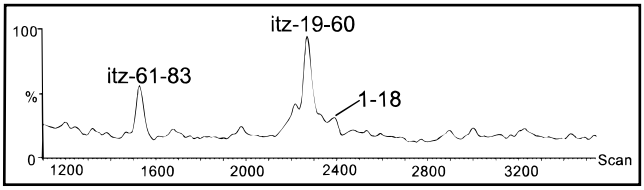


Figure 4. RTIC from LC-MS (ESI) of the peptide mixture resulting from cyanylation and cleavage of the fraction corresponding to HPLC peak 2 in Figure 1. Separation was carried out under the same conditions as for Figure 2.

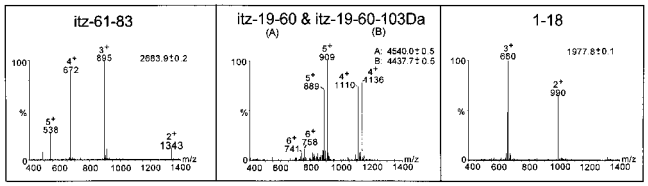


Figure 5. The ESI mass spectra of the peptides corresponding to the RTIC peaks in Figure 4.

Table 1. Calculated and Observed Masses for the Fragments from the Cleavage Reaction of Singly Reduced and Cyanylated LR^{3}IGF-I Isoforms

reduction of disulfide	residues in fragments	calcd mass (Da)	obsd mass (Da)
Cys 60−Cys 65	1−59	6371.2	6372.7
	itz-60−64	637.7	638.0
	itz-65−83	2188.6	2189.7
Cys 19−Cys 61	1−18	1978.4	1977.8
	itz-19−60	4538.0	4540.0
	itz-61−83	2681.1	2683.9

ogy. Because LR^{3}IGF-I consists of three disulfide bonds, use of the partial reduction technique should result in the production of three isoforms, each of which consists of a singly reduced isomer emanating from a different disulfide bond. However, as shown in the chromatogram (Figure 1), only two singly reduced isoforms were detected in addition to a major peak for the intact unreduced protein; the numbered peaks (minor peaks following peak 2 represent doubly and completely reduced forms of the protein, respectively) represent a singly reduced isoform as determined by mass analysis, showing that the molecular weight of each species has shifted by 52 Da (data not shown) from that of the intact protein (a 2-Da shift for conversion of a cystine to two cysteines, followed by a 50-Da shift for the replacement of a hydrogen on each of two nascent sulfhydryls with a cyano group). This result is not unexpected as some disulfide bonds are much more stable than others as reported in the results of a recent study of the stability of the disulfide bond consisting of Cys 18 to Cys 61 in IGF-I,[8] which is analogous to the disulfide linkage of Cys 31−Cys 74 in LR^{3}IGF-I. The Cys 18−Cys 61 disulfide bond in IGF-I is preferentially formed in the folding process, a feature that may explain the resistance of Cys 31−Cys 74 of LR^{3}IGF-I to reduction under our reaction conditions. In any case, in dealing with a protein known to contain three disulfide bonds, the determination of the pairing of two of the three disulfide bonds allows the remaining one to be determined by default.

Pursuing the analytical process, the singly reduced isoforms as represented in Figure 1 as peaks 1 and 2 were subjected to the chemical processing illustrated in Scheme 1. The cyanylation/cleavage products of the first isoform (peak 1 in Figure 1) were analyzed by LC-MS with electrospray (ESI) mass spectrometry to give the reconstructed total ion current (RTIC) chromatogram in Figure 2 and the corresponding mass spectra in Figure 3; similarly, LC-MS of the cyanylation/cleavage products of the second isoform (peak 2 in Figure 1) resulting from reduction of a different disulfide bond gave the RTIC chromatogram in Figure 4 and the corresponding electrospray spectra in Figure 5. Table 1 lists the calculated masses for possible fragments resulting from the cleavage reaction of the protein chain at designated pairs of cysteine sites.

The peaks in the RTIC in Figure 2 can be assigned to specific fragments 1−59, itz-60−64, itz-65−83, itz-65−83 minus H_2O, and the uncleaved protein according to the corresponding masses from ESI spectra. The ESI spectra in Figure 3 give masses 6372.7, 638.0 (single charge), and 2189.7 Da which are due to fragments 1−59, itz-60−64, and itz-65−83, respectively. From these data, one can deduce that Cys 60 is linked to Cys 65. The peaks in the RTIC in Figure 4 resulting from analysis of cleavage products of the second cyanylated isoform (peak 2 in Figure 1) can be assigned to fragments 1−18 (1977.8 Da), itz-19−60 (4540.98 Da), and itz-61−83 (2684.93 Da) according to peaks in the ESI spectra in Figure 5, which indicate Cys 19 was linked to Cys 61 as illustrated in Scheme 1. Peak B in the ESI spectrum in the middle panel of Figure 5 corresponds arithmetically to itz-19−60 minus 103 Da; we have no rational explanation for its origin at this time. Whether this peak corresponds to the expulsion of a cysteine residue from itz-19−60 awaits further study with other peptides or proteins containing adjacent cysteines. Overall, the disulfide bond structure containing two adjacent cysteines can be readily assigned. These results provide direct evidence for the disulfide bonding pattern in LR^{3}IGF-I that has hitherto only been presumed by homology to that in IGF-I.

Acknowledgment. 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).

JA9736018

(7) Milner, S. J.; Francis, G. L.; Wallace, J. C.; Magee, B. A.; Ballard, F. J. *Biochem. J.* **1995**, *308*, 865−871.
(8) Hober, S.; Forsberg, G.; Palm, G.; Hartmanis, M.; Nilsson, B. *Biochemistry* **1992**, *31*, 1749−1756.