Anal. Chem. **2009,** *81,* 7314–7320

Methodology for Determining Disulfide Linkage Patterns of Closely Spaced Cysteine Residues

Bing Zhang and Steven L. Cockrill*

Analytical Sciences, Amgen, Inc., 4000 Nelson Road, Longmont, Colorado 80503

We report the development and application of a method for determining bonding patterns in disulfide-linked peptides containing closely spaced cysteine residues. Through the utility of classic N-terminal sequencing chemistry coupled with facile liquid chromatography and mass spectrometric analysis of the cleavage products, we report the ability to demonstrate unambiguous assignment of paired cysteine residues, using human insulin as a model protein. The conditions of the technique were selected and optimized to maintain disulfide integrity. In a forthcoming article, we will present the results of this method as applied to the complete elucidation of linkages in disulfide variants of a therapeutic monoclonal antibody of the IgG2 subclass.

Complete elucidation of the primary structural features of a protein of interest encompasses a broad array of analytical methods to assess specific and often separate structural features, including post-translational modifications such as glycosylation, processing of terminal variants, and pairing of cysteine residues to form disulfide bonds. The disulfide linkage pattern is a critical factor in determining the protein's conformational properties and may therefore impart significant impact to the biological function through the availability of active site(s).[1,2]

Typically, because of the large size of most proteins, cleavage of the polypeptide backbone using an appropriate endoprotease is necessary for characterization of the disulfide linkages. By comparison of the peptides under nonreducing and reducing conditions, it is possible to identify the constituent species of a disulfide-linked peptide. If these species contain only a single cysteine residue each, then elucidation of the linkage pattern is simply achieved by identification of those constituent peptides.

For species containing multiple disulfide linkages, additional cleavage between sequential cysteine residues is required; it is normally reasonable to affect cleavage between disulfide-linked cysteine residues due to the broad array of endoproteases and their associated specificities. However, the amino acid sequence of the protein of interest may contain closely spaced cysteine residues with no suitable intermediate cleavage site or even pairs of cysteine residues that are adjacent in the amino acid sequence. Such instances represent significant challenges to elucidation of the disulfide structure.

The contemporary literature contains several methods that may provide information for disulfide linkage analysis. Partial reduction and derivatization of the nascent free thiol group has been applied for determination of disulfide connectivity in a variety of proteins, including highly knotted substrates[3] and those with closely spaced cysteines.[4] However, for targeted analysis, this approach generally requires the ability to selectively reduce specific disulfide linkages. This may involve extensive optimization of the reduction conditions to afford the selective generation of isomers containing specific linkage reduction. Such optimization steps are often necessarily time-intensive, may consume significant quantities of a potentially limited sample, and are even more challenging when multiple similar linkages are present in a peptide, each with equivalent susceptibility to reduction. Additionally, the required denaturation of the protein prior to the partial reduction step may itself lead to disulfide scrambling.[5] While the data output is readily interpreted, the utility of multiple steps for each cycle of processing and identification typically requires several hours of manipulation and analysis.

Automated Edman sequencing has also been utilized for determining linkage assignments through the detection of di-PTH-cystine.[6] Indeed, this approach was used to demonstrate the presence of the "ladder" structure of cysteine bonding in the hinge region of therapeutic antibodies of both the IgG4[6] and IgG2[7] subclasses. However, this technique has limitations with respect to the inability to detect nonparallel linkages, and quantitative distinction between multiple disulfide variants present in a mixture cannot be made. Furthermore, issues with cycling efficiency and carryover are of concern, particularly with closely spaced cysteine residues.

Recently mass spectrometry-based approaches have been applied to disulfide characterization.[6,8,9] Techniques employing mass spectrometric detection and analysis are very attractive options due to the fact that there is typically no need for sample preparation and manipulation beyond standard digestion and separation (i.e., peptide mapping). This approach has been successfully employed in the analysis of various protein substrates, including therapeutic antibodies.[8,9] However, the data presented in these publications did not afford specific linkage determination between the four cysteine residues in the dimeric hinge peptide; it was not possible to differentiate the possible parallel or "crossed"

* Corresponding author. Steven L. Cockrill, e-mail: steven.cockrill@amgen.com. Fax: (303) 401-4403.

(1) Flynn, J. P.; Pallaghy, P. K.; Lew, M. J.; Murphy, R.; Angus, J. A.; Norton, R. S. *Biochim. Biophys. Acta* **1999**, *1434*, 177–190.

(2) Dillon, T. M.; Ricci, M. S.; Vezina, C.; Flynn, G. C.; Liu, Y. D.; Rehder, D. S.; Plant, M.; Henkle, B.; Li, Y.; Deechongkit, S.; Varnum, B.; Wypych, J.; Balland, A.; Bondarenko, P. V. *J. Biol. Chem.* **2008**, *283*, 16206–16215.

(3) Qi, J.; Wu, J.; Somkuti, G. A.; Watson, J. T. *Biochemistry* **2001**, *40*, 4531–4538.

(4) Martinez, T.; Guo, A.; Allen, M. J.; Han, M.; Pace, D.; Jones, J.; Gillespie, R.; Ketchem, R. R.; Zhang, Y.; Balland, A. *Biochemistry* **2008**, *47*, 7496–7508.

(5) Yen, T.-Y.; Joshi, R. K.; Yan, H.; Seto, N. O. L.; Palcic, M. M.; Macher, B. A. *J. Mass Spectrom.* **2000**, *35*, 990–1002.

10.1021/ac901161e CCC: $40.75
Published on Web 07/23/2009

linkage patterns, i.e., bonds between Cys-229 and Cys-229 and Cys-232 and Cys-232 (parallel) or Cys-229 and Cys-232 and Cys-232 and Cys-229 (crossed). No fragmentation was observed between the closely spaced cysteines of each peptide chain, even with the adjacent proline residues.[9]

An example application of particular interest is in the hinge region of therapeutic antibody products, particularly those of the IgG2 subclass. Determination of the disulfide structure is of importance due to increasing expectations from regulatory agencies for a more complete understanding of a therapeutic molecule but also with potential concerns over product safety and efficacy. For example, recent publications presented the identification of previously undetermined disulfide variants.[2,7] While novel disulfide mediated structural variants were identified, the specific linkage patterns between the Fab arm and hinge species were not determined. The MS-based methodologies described above face significant challenges in achieving elucidation of the disulfide bonding patterns of such complex structures, given the large number of cysteine residues (up to 16) present in the signature nonreduced Lys-C peptides characteristic of each structural variant.[7] In a forthcoming paper, we will present complete elucidation of the disulfide structure through the application of a novel technique to the analysis of the disulfide-linked peptides unique to each variant.

In the current manuscript, we describe development of the novel methodology, with application to insulin as a model protein. Insulin was selected due to its convenient size, presentation of three disulfides (two interchain, one intrachain), combination of adjacent and widely spaced cysteine residues, and a long, established history of characterization. The interchain disulfide linkages are between residues Cys-7 and Cys-7 and Cys-20 and Cys-19 of the A and B chains, respectively. The intrachain disulfide exists between Cys-6 and Cys-11 of the A chain.[10] This structure of insulin is depicted in the Supporting Information for reference.

Insulin contains both widely spaced, interchain disulfide linkages with facile sites for proteolytic cleavage by Glu-C on both chains, as well as adjacent cysteine residues involved in separate disulfide bonds. As such, insulin is representative of an analytically challenging substrate that, conveniently, has been extensively characterized. Treatment of the model protein with endoprotease Glu-C following treatment with alkylating reagent under denaturing conditions simulates the typical proteolytic processing steps necessary for generation of peptides appropriate for efficient sequential release. The utility of an alkylating reagent such as *N*-ethylmaleimide (NEM) is employed to "scavenge" free sulfhydryl and thereby inherently prevent disulfide scrambling at the elevated temperatures and pH ranges typical of most sample preparation processes.[5] Additionally, the use of lower pH buffers (i.e., pH 6.5) for digestion conditions has been utilized with specific intent of assessing disulfide linkages.[5,12]

Literature reports also show that the expected disulfide structure of insulin may be perturbed under controlled redox conditions, resulting in scrambling of the linkages. Fortuitously, these disulfide variants are readily resolved by chromatographic means.[13]

In this paper, we describe a novel approach that provides facile analysis and confirmation of the connectivity of disulfide linkages using insulin as a model protein. The applications of this methodology are numerous but perhaps most relevant to the recently discovered disulfide-mediated structural variants of immunoglobulin G2.[7]

MATERIALS AND METHODS

Materials. Recombinant human Insulin was sourced from Invitrogen (Carlsbad, CA), and insulin oxidized B-chain was from Sigma-Aldrich (St. Louis, MO). Endoproteinase Glu-C (protease V 8, *Staphylococcus aureus* V 8), sequencing grade was purchased from Roche (Indianapolis, IN). Reagents for manual execution of Edman sequencing including trifluoroacetic acid (TFA), phenylisothiocyanate (PITC), and *N*-methylpiperidine/water/methanol solution were sequencing grade materials obtained from Applied Biosystems (Foster City, CA). Pyridine was ACS reagent grade from Fluka (Buchs, Switzerland). Preprepared mobile phases (0.1% TFA in water and 0.1% TFA in acetonitrile), as well as HPLC-grade water, were from J.T. Baker, (Phillipsburg, NJ).

Tris solution (1 M, pH 8.0) was from Calbiochem (La Jolla, CA), tris(2-carboxyethyl)phosphine (TCEP) was purchased from Pierce (Rockford, IL), and guanidine hydrochloride solution (8 M), iodoacetic acid (IAA), and *N*-ethylmaleimide (NEM) were obtained from Sigma-Aldrich (St. Louis, MO). NAP-5 Sephadex G-25 columns were sourced from Pharmacia Biotech (Little Chalfont, U.K.).

Chromatographic Analysis of Commercial Human Insulin. Commercial human insulin was assessed by reverse-phase HPLC in order to ascertain the potential presence of multiple disulfide structures. A 1 μg load of insulin (1 mg/mL in 2 M Gdn HCl, 100 mM Tris, pH 6.5) was applied to a reverse-phase column (Agilent Zorbax 300SB C8, 2.0 mm × 150 mm, 5 μm particle size) and separated using mobile phases consisting of (A) 20% acetonitrile in 200 mM ammonium sulfate and 50 mM sulfuric acid and (B) 40% acetonitrile in 200 mM ammonium sulfate and 50 mM sulfuric acid. Elution was accomplished using a 30 min gradient from 10% to 50% mobile phase B at a temperature of 40 °C and a flow rate of 0.2 mL/min, as reported previously.[11]

Enzymatic Digestion with Endoprotease Glu-C and Mapping of Insulin Peptides. Insulin was reconstituted in 400 μL of denaturation buffer (6 M guanidine HCl, 200 mM Tris, pH 6.5) and then treated with 100 μL HPLC grade water and 12 μL of 0.5 M iodoacetic acid for a final protein concentration of 2 mg/mL. The sample was incubated at room temperature in the dark for 30 min, then buffer exchanged to 50 mM Tris, pH 6.5 using an equilibrated NAP-5 Sephadex G-25 column, with a resulting protein concentration of 1 mg/mL.

Digestion using endoprotease Glu-C was accomplished by incubating 120 μL of 1 mg/mL insulin solution (in 50 mM Tris

(6) Wei, Z.; Marzilli, L. A.; Rouse, J. C.; Czupryn, M. J. *Anal. Biochem.* **2002**, *311*, 1–9.
(7) Wypych, J.; Li, M.; Guo, A.; Zhang, Z.; Martinez, T.; Allen, M. J.; Fodor, S.; Kelner, D. N.; Flynn, G. C.; Liu, Y. D.; Bondarenko, P. V.; Ricci, M. S.; Dillon, T. M.; Balland, A. *J. Biol. Chem.* **2008**, *283*, 16194–16205.
(8) Chelius, D.; Huff Wimer, M. E.; Bondarenko, P. V. *J. Am. Soc. Mass Spectrom.* **2006**, *17*, 1590–1598.
(9) Wu, S.-L.; Jiang, H.; Lu, Q.; Hancock, W. S.; Karger, B. L. *Anal. Chem.* **2009**, *81*, 112–122.
(10) Chang, S.-G.; Choi, K.-D.; Jang, S.-H.; Shin, H.-C. *Mol. Cells* **2003**, *16*, 323–330.
(11) Hua, Q.-X.; Jia, W.; Frank, B. H.; Phillips, N. F. B.; Weiss, M. A. *Biochemistry* **2002**, *41*, 14700–14715.
(12) Foley, S. F.; Sun, Y.; Zheng, T. S.; Wen, D. *Anal. Biochem.* **2008**, *377*, 95–104.
(13) Forsberg, G.; Palm, G.; Ekeracke, A.; Josephson, S.; Hartmanis, M. *Biochem. J.* **1990**, *271*, 357–363.

Table 1. Putative Cleavage Products of Insulin Digestion by Endoprotease Glu-C under Nonreducing Conditions[a]

peptide	structure/identity	theoretical mass (Da)	observed mass (Da)	mass error (ppm)
A1	GIVE	416.23	416.23	0
A2(s-s)B1	QCCTSICSLYQLE(s-s)FVNQHLCGSHLVE	2967.33	2967.32	3
A2	QCCTSICSLYQLE	1489.63	1489.63	0
B1	FVNQHLCGSHLVE	1481.71	1481.72	7
A3(s-s)B2	NYCN(s-s)ALYLVCGE	1376.59	1376.58	7
A3	NYCN	512.17	512.17	0
B2	ALYLVCGE	866.42	866.42	0
B3	RGFFYTPKT	1115.58	1115.58	0

[a] Masses are monoisotopic.

buffer, pH 6.5) with Glu-C (substrate to enzyme ratio 3:1) overnight at room temperature in foil-wrapped sample tubes. Digestion was quenched by addition of 12 μL of 10% TFA in water.

Separation of the resulting peptides was accomplished using a 1200 HPLC system (Agilent, Santa Clara, CA) equipped with a binary pump. A reverse-phase column (Agilent Zorbax 300SB C8, 2.1 mm × 150 mm, 5 μm particle size) was employed with mobile phases consisting of (A) 0.1% TFA in water and (B) 0.1% TFA in acetonitrile. A linear gradient between 2 − 50% mobile phase B was used, following a 5 min system equilibration step at 2% B, with a flow rate of 0.2 mL/min. Reduced mapping was performed under identical conditions following treatment of the digest products with 2 μL of Tris(2-carboxyethyl)phosphine (TCEP) for 10 min at room temperature.

For LC−MS analysis, the outlet of the HPLC separation was coupled to a MSD-TOF ESI-MS instrument (Agilent, Santa Clara, CA) running in positive ion mode. Source settings were a scan range of 200 − 3000 m/z, fragmentor, skimmer, octopole rf, and capillary voltages of 200, 60, 250, and 4 000 V, respectively. Drying gas was supplied at 10.0 L/min and 300 °C. Nebulizer pressure was 40 psig. For fraction collection, the MS instrument was bypassed, and the HPLC eluent was collected manually and dried by vacuum centrifugation prior to further processing and analysis.

Coupling Reaction. Derivatization of the primary amines was accomplished using phenyisothiocyanate (PITC) as traditional Edman degradation chemistry.[14] Briefly, the peptides resulting from Glu-C digestion of 4 nmol of insulin and isolated by peptide mapping as described above were reconstituted in 10 μL of 10 mM *N*-ethylmaleimide NEM in water, then incubated at room temperature in the dark for 30 min. A 40 μL volume of anhydrous pyridine, 5 μL of PITC, and a 5 μL volume of *N*-methylpiperidine/water/methanol solution were added under dry nitrogen. Coupling was achieved by incubation at various temperatures (room temperature, 50 °C) for various times (3−20 min). The solution was then dried completely by vacuum centrifugation. Completeness of coupling was assessed by resolubilization in 30% pyridine in water and analysis by LC−MS, using separation and detection conditions as described for execution of the peptide map.

Cleavage Reaction. Cleavage of the derivatized substrate was accomplished by acid treatment using anhydrous TFA. Various modes were assessed, including direct addition of a 10 μL aliquot of liquid TFA to the dried sample under dry nitrogen, in TFA vapor (introduced by a filter paper soaked in 20 μL of anhydrous TFA), and incubating at 50 °C or room temperature for 2−10 min. Completeness of cleavage was assessed by LC−MS, using the conditions noted above.

(14) Tarr, G. E. *Anal. Biochem.* **1975**, *63*, 361–370.

Conversion Reaction. The solution resulting from the cleavage step detailed above was cooled and pooled by centrifuging for 2 min at room temperature. A 15 μL aliquot of HPLC-grade water was added to reach a final TFA concentration of 25%, after which the solution was incubated at 64 °C for 10 min. After cooling, a portion of the samples was loaded directly onto the LC−MS system for analysis. Note that these conditions were designed to be identical to those of the automated Procise 494 instrument (Applied Biosystems, Foster City, CA).

Iterative Coupling/Cleavage Cycles. For sequential cycles, the residual substrate following a given alkylation/coupling/cleavage/conversion cycle was collected from the chromatographic separation, dried by vacuum centrifugation, and then subjected to repetition of the alkylation, coupling, cleavage, and conversion steps followed by LC−MS analysis as described above.

RESULTS AND DISCUSSION

Assessment of Insulin Disulfide Integrity. The disulfide integrity of recombinant human insulin used as a model protein was assessed chromatographically. As reported previously,[11] scrambled products may be chromatographically resolved from the native structure using a reverse-phase separation. The recombinant insulin showed the presence of a single peak (shown in the Supporting Information). Because of the absence of multiple peaks in the reverse-phase profile, it is inferred that the model protein utilized for method development presented the expected disulfide arrangement.

Isolation of Disulfide-Linked Peptides from Recombinant Human Insulin. As shown in the Supporting Information, insulin contains three disulfide linkages between its six cysteine residues. Digestion with endoprotease Glu-C under nonreducing conditions putatively generates four proteolysis products, summarized in Table 1. Comparative peptide mapping of the digest under nonreducing and reducing conditions (Figure 1) affords identification of peaks corresponding to disulfide-containing peptides through either retention time shift, mass difference, or both. Two peptides were identified as containing disulfide linkages, peaks eluting at ∼29 and ∼36 min, as they were absent from the chromatogram upon treatment with reducing agent. The peak eluting at ∼29 min was tentatively identified as peptide A3(s-s)B2 by comparison of its observed and theoretical masses (1376.58 Da and 1376.59 Da, respectively, Table 1). Similarly, the peak eluting at ∼36 min was identified as peptide A2(s-s)B1 (2967.32 Da and 2967.33 Da for observed and theoretical masses, respectively, Table 1).

The disulfide connectivity of nonreduced peptides containing a single disulfide bond may be determined directly from confirma-

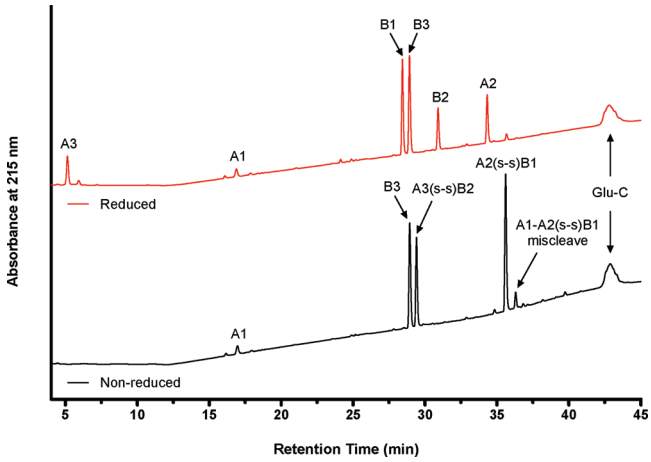

Figure 1. Comparative peptide maps of insulin Glu-C digests under nonreducing and reducing conditions.

tion of the constituent species. Peak A3(s-s)B2 eluting at ~29 min yielded two product peaks upon treatment with reducing agent (peaks A3 and B2 in Figure 1), and these species were identified by mass spectrometry as NYCN and ALYLVCGE (observed masses of 512.17 and 866.42 Da, respectively, Table 1). As the identified peptides contain only a single Cys residue each, the disulfide linkage of peptide A3(s-s)B2 must be through the two cysteine residues, i.e., Cys-20 in peptide A3 is bound to Cys-19 in peptide B2.

Peptide A2(s-s)B1 eluting at ~36 min also yielded two product peaks (peptides A2 and B1 in Figure 1). Mass spectrometry detection of the product peaks confirmed their identities as QCCTSICSLYQLE and FVNQHLCGSHLVE (observed masses of 1489.63 Da and 1481.72 Da, respectively, Table 1) through matching of the expected and observed masses. Because of the fact that one of the peptides contains three cysteine residues, confirmation of the specific linkage pattern cannot be inferred from the constituent peptides nor can the presence of coeluting disulfide variants. In order to ascertain the connectivity of the disulfide linkages in peptide A2(s-s)B1, a new methodology was developed, as described in the following sections.

Manual Edman of Glu-C Peptide A2(s-s)B1. *Principle of the Method.* The methodology comprises a convergence of traditional Edman chemistry executed manually with analysis using LC−MS. Employment of extracted ion chromatograms is a key aspect, on account of the fact that the UV trace is confounded by remnant reagent peaks. The mechanisms for Edman chemistry have been extensively documented.[14,15] Under basic conditions, the phenylisothiocyanate couples with α amino groups (i.e., the N-terminus) to form a phenylthiocarbamyl (PTC) group. This is cleaved by treatment with anhydrous acid, releasing the anilinothiazolinone (ATZ) derivative of the N-terminal amino acid and generating a new N-terminus for repetitive cycles. Conversion of the ATZ derivative to the more stable phenythiohydantoin (PTH) derivative results in the loss of water and a consequent reduction in mass.

Optimization Studies of Manual Edman Reaction. Optimization studies were accomplished using a simpler test molecule, insulin oxidized B-chain. Coupling and cleavage steps are summarized below.

Coupling Reaction (Insulin Oxidized B Chain). Incorporation of the PITC label on primary amines was assessed by comparing 2% vs 20% of the total reaction volume at 50 °C for 20 min. Insulin oxidized B chain contains two primary amines: the free N-terminus and a lysine residue at position 29. Evaluation of extracted ion chromatograms for the unlabeled, singly-, and doubly labeled peptide demonstrated that under the coupling conditions, both 2% and 20% proportions generate essentially complete labeling at primary amines (data not shown). Therefore, in order to accommodate more complex disulfide-linked peptides containing more primary amines, an intermediate value of 10% PITC was selected.

A time-course study to evaluate the labeling requirements using 10% PITC and 50 °C reaction temperature was performed. Reaction times of 3, 5, and 20 min were assessed. In each case, 100% label incorporation (i.e., doubly labeled peptide) was observed; however at 20 min, a side reaction that generated a +12 Da adduct mass was observed. Therefore 5 min was selected as the optimal reaction time.

Incubation temperature was also probed. Automated Edman sequencing programs typically employ 50 °C reaction temperatures; however, in order to mitigate potential thermally induced disulfide rearrangements, a comparison of incubation at 50 °C and room temperature was employed for insulin oxidized B chain and also angiotensin II as a secondary test peptide. The results indicated that room temperature incubation was sufficient to yield complete labeling of primary amines (data not shown).

Automated programs for Edman sequencing employ washing steps with highly hydrophobic solvents (ethyl acetate and heptane) to remove the small molecule byproduct of the coupling reaction. However, it has been reported in the literature that exposure to organic solvents could potentially lead to disulfide scrambling.[16] In order to assess the impact on the peptide detectability of the washing step, the coupling reaction was performed using insulin oxidized B chain, with and without the washing step. The results demonstrated that although the reagent peaks were diminished with washing, recovery of the coupled peptide was reduced by approximately 50% with the washing step (peak areas of 2 973 728 vs 1 385 883 by the summed extracted ion chromatogram (EIC) of multiple charge states or 1 368.78 vs 675.20 by UV for exclusion or inclusion of the wash step, respectively, data not shown). The issues with compromised recovery as well as the potential for scrambling of disulfide-linkages was considered to be of sufficient risk to the value of the method that the washing step was omitted from the final protocol.

Therefore, the optimized coupling conditions were determined to be addition of a 40 μL volume of anhydrous pyridine, 5 μL of PITC (10% of the total reagent volume), and 5 μL of methylpiperidine under dry nitrogen to the reconstituted target peptide. Coupling was achieved by incubation at room temperature for 5 min, after which the solution containing coupled peptide was dried completely by vacuum centrifugation. Note that coupling was performed after reconstitution of the target peptide in 10 μL of 10 mM NEM in water and incubation at room temperature in the dark for 30 min in order to mitigate disulfide scrambling. Evaluation of the final coupling conditions on insulin Glu-C peptide

(15) Creighton, T. E. *Proteins: Structure and Molecular Properties*, 2nd ed.; W. H. Freeman and Co.: New York, 1993.

(16) Chang, J. Y.; Li, L. *J. Biol. Chem.* **2001**, *276*, 9705–9712.

A2(s-s)B1 showed that the expected incorporation of two PITC labels was essentially 100% complete.

Cleavage Reaction. Removal of the N-terminal amino acid residue following the coupling step was accomplished by acid cleavage, based upon the conditions used with automated instrumentation programs. Cleavage was found to be complete after incubating with 10 μL of neat anhydrous TFA for 2 min at 50 °C, i.e., no partial cleavage products were observed by LC−MS. The utility of room temperature incubation was found to be not sufficient for complete cleavage (data not shown). Vapor phase sequencing reactions have been reported in the literature to afford higher sensitivity;[17] however, in our assessment there was no discernible advantage using vapor phase TFA over liquid TFA, most likely due to the fact that the remaining steps of the degradation chemistry are performed in solution phase. Liquid phase TFA was employed in the final method on account of the fact that subsequent manipulations were more readily facilitated. The completeness of the final cleavage reaction conditions was also demonstrated using insulin peptide A2(s-s)B1 (data not shown).

Conversion. In automated Edman, a conversion step from the AZT derivative to the more stable PTH derivative is performed. This was accomplished simply by monitoring the masses of the resulting products; converted (PTH) and unconverted (AZT) differ by the loss of water and are therefore readily differentiated by MS detection. The conversion step is performed under identical conditions to that established for automated N-terminal sequencing using a Procise 494 instrument (Applied Biosystems, Foster City, CA). Conversion was found to be 100% complete after incubating with 25% TFA at 64 °C for 10 min and therefore was not optimized further.

Confirmation of Disulfide Integrity during Execution of Optimized Technique. A major concern with any methodology employed to probe disulfide bonding is the occurrence of linkage scrambling or rearrangement. Therefore an assessment of scrambling potential was made by performing the coupling, cleavage, and conversion steps in the presence of NEM. By incorporation of the alkylating reagent during the processing steps, any rearrangement of the disulfides would be captured through the labeling of the transitional free sulfhydryl and the resulting formation of unique products. Five distinct products would indicate the occurrence of scrambling: peptide B1* with a single NEM label, peptide A2* with a single NEM label, peptide A2* with three NEM labels, B1* dimer, or A2* dimer (where the * notation indicates the loss of the first N-terminal residue from the annotated peptide). Each of these species has a unique mass and likely a characteristic retention time.

Insulin peptide A2(s-s)B1 was pretreated with 10 mM NEM at room temperature for 30 min before execution of the manual Edman coupling, cleavage, and conversion cycles in order to identify the propensity for disulfide scrambling. Figure 2 shows the EIC of possible NEM derivatives of peptides A2* and B1*, the A2* dimer and B1* dimer species, as well as of the expected disulfide-linked structure of peptide A2*(s-s)B1*. The fact that no detectable quantities of the potential scrambled products were observed indicates that the disulfide structure was not compromised by execution of the coupling, cleavage, and conversion steps of the Edman reaction. The small signal in the A2* dimer (theoretical mass 2717.13 Da) EIC trace is related to the sodium

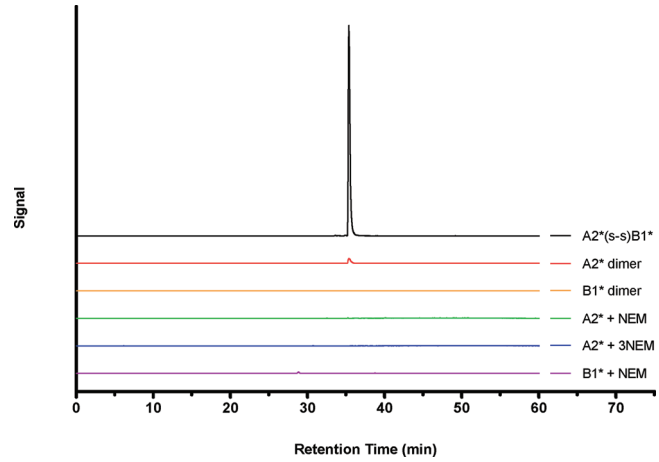

Figure 2. EIC traces for expected A2*(s-s)B1* and potential scrambled products after one cycle.

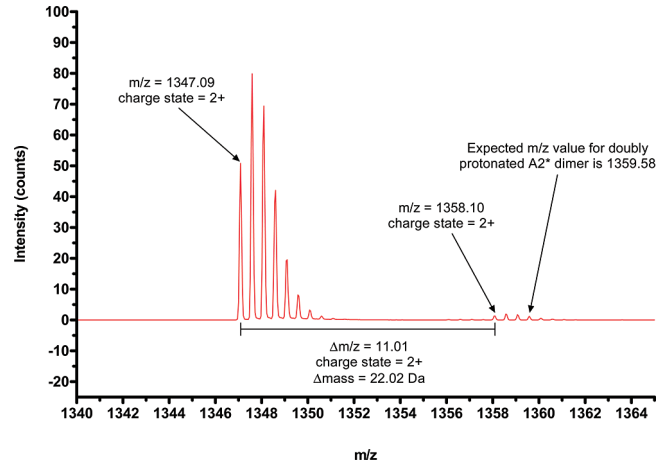

Figure 3. Mass spectrum for doubly charged species eluting at ~35 min.

adduct of the expected A2*(s-s)B1* structure (theoretical mass 2692.21 Da). The fourth isotope of the $[M + H + Na]^{2+}$ ion of the expected A2*(s-s)B1* structure exhibits the same m/z value as the first isotope of the $[M + 2H]^{2+}$ ion of the A2* dimer (Figure 3). The fact that the two signals occur at the same elution time suggests that they are related to the same structure (i.e., protonated vs sodiated), which is considerably more likely than coelution of different structures. The absence of the B1* dimer also supports this proposal.

Theoretical Distinction of Disulfide Variants for Glu-C Peptide A2(s-s)B1. Three putative structures (structures 1, 2, and 3) of peptide A2(s-s)B1 resulting from the Glu-C digestion of insulin are presented in Figure 4, which differ in the connectivity but all share the same theoretical mass of 2967.33 Da. In structure 1, the disulfide linkages are Cys-6 to Cys-11 (intrachain on chain A) and Cys-A7 to Cys-B7 (interchain). This represents the traditionally accepted linkage pattern.[10,11] The other two structures contain variations of the disulfide linkages; structure 2 contains an interchain linkage between Cys-6 of the A-chain to Cys-7 of the B-chain and an intrachain linkage between Cys-7 and Cys-11 of the A-chain. Structure 3 contains an intrachain linkage between Cys-6 and Cys-7 of the A-chain and an interchain linkage between Cys-11 of the A-chain and Cys-7 of the B-chain. Note that these represent purely hypothetical arrangements given the confirmed constituents of peptide A2(s-s)B1,

(17) Brandt, W. F.; Frank, G. *Anal. Biochem.* **1988**, *168*, 314–323.

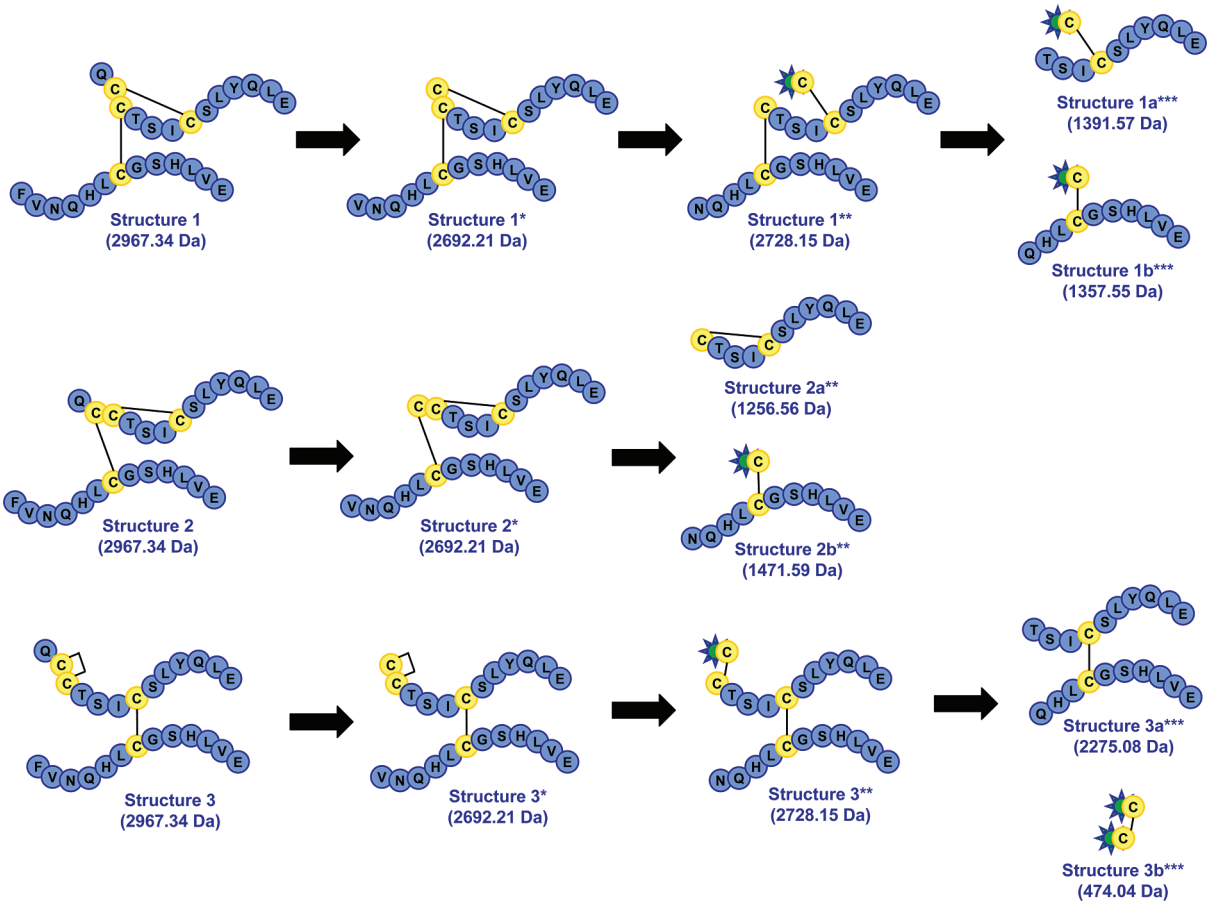

Figure 4. Differentiation of insulin disulfide variants through characteristic residual and leaving groups.

although both structures 2 and 3 have been generated under conditions of limited reduction and denaturation.[11]

The three species are differentiated by their respective residual and leaving groups following sequential cycles of Edman sequencing chemistry, as demonstrated in Figure 4. The first cycle does not permit distinction as each of the species simply loses the glutamine and valine residues from the A and B chains, respectively. While the products of structures 1 and 3 remain isobaric after the second cycle, structure 2 is clearly differentiated as two product species are formed of masses 1256.56 and 1471.59 Da. Three cycles are necessary to distinguish structures 1 and 3, after which the A and B chains of structure 1 are cleaved, resulting in the formation of two product species of mass 1391.57 and 1357.55 Da, respectively. Structure 3 is not cleaved into separate chains by three cycles of Edman chemistry, and as such the residual mass is 2275.08 Da.

Application of Repetitive Cycles to Insulin Peptide A2(s-s)B1. As noted in Figure 4, the presence of potential disulfide variants may be distinguished only after at least two cycles, due to the location of the first cysteine at the second amino acid position in peptide A2. Therefore multiple rounds of coupling, cleavage, and conversion steps were executed in sequential manner.

Cycle 1. The first cycle of N-terminal truncation does not result in distinguishing species as each of the three structures 1*, 2*, and 3* possess the same mass. The EIC for peptide A2*(s-s)B1 is shown in Figure 5. These data indicate that the expected species, A2*(s-s)B1* was detected. As noted above (Figure 2), potential scrambling products were not detected.

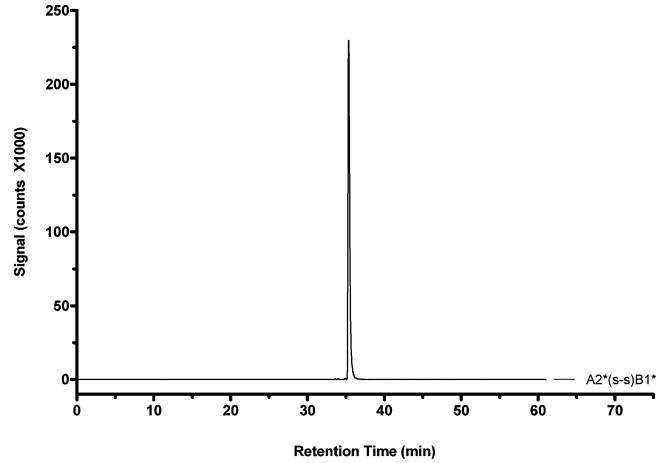

Figure 5. EIC trace for A2*(s-s)B1* peptide (structures 1, 2, and 3) after one cycle.

Cycle 2. The second cycle of N-terminal truncation is the first opportunity at which the discrete disulfide variant structures may be differentiated. As noted above, structures 1** and 3** (i.e., A2**(s-s)B1**), where ** indicates the completion of two Edman cycles, cannot be resolved by mass at this point, as they still possess the same constituent amino acid inventory. However, structures 2a** (peptide A2**) and 2b** (peptide B1**) are readily differentiated from structures 1** and 3**. EICs for these structures are presented in Figure 6. The traces demonstrate that the dominant species identified has a mass that corresponds to the expected structure (structure 1**), although clearly no

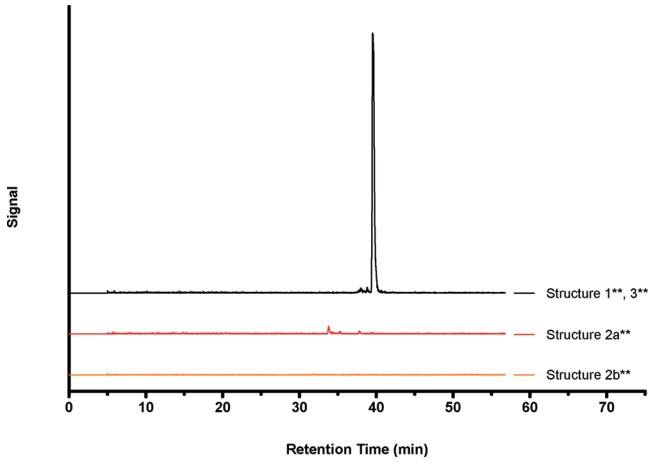

Figure 6. EIC traces for species resulting from structures 1, 2, and 3 after two cycles.

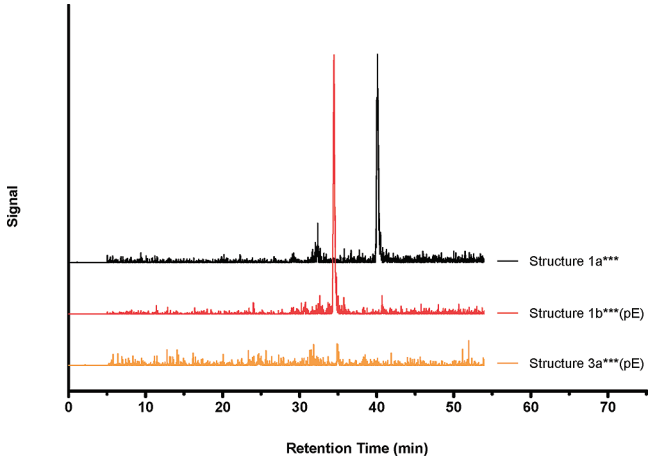

Figure 7. EIC traces for species resulting from structures 1 and 3 after three cycles.

distinction can be made regarding the presence of structure 3**. What is evident is the fact that no detectable levels of the species corresponding to structure 2 are observed; therefore, it is concluded that this particular unexpected disulfide variant is not present in the population of insulin molecules.

Cycle 3. The third cycle of N-terminal truncation provides differentiation of structures 1 and 3. After three cycles, structure 1 dissociates to discrete peptides from each chain (structures 1a*** and 1b*** are peptides A2*** and B1***, respectively). Completion of three cycles for structure 3 results in the loss of a di-PTH-cystine group (structure 3b***) and a residual group that contains peptides A2*** and B1*** bound through a disulfide linkage. EICs for these structures are presented in Figure 7. Note that the di-PTH-cystine is not reliably detected due to its small mass or poor retention on the column.

The traces demonstrate that the dominant species identified are structures 1a*** and 1b***, which correspond to the expected structure (structure 1). Note that structure 1b*** and 3a*** were detected as the pyroglutamylate derivative (with concomitant loss of 17.03 Da), as the N-terminal residue of each species is a glutamine. Very minor amounts of species derived from structure 3 were detected, again demonstrating that the native disulfide structure is essentially preserved throughout multiple rounds of manual Edman cycling.

One aspect to note is the reduction in signal-to-noise through sequential steps (Figures 5−7), which is recognized as a limitation of the present methodology. This is analogous to the limited cycling efficiency of automated Edman degradation, which typically demonstrates a repetitive yield of ~90%;[18] however, it is apparent that the losses in the manual methodology may be higher due to the need to recover the residual groups from multiple chromatographic steps.

CONCLUSIONS

We have presented here a summary of a novel methodology that affords unambiguous assignment of disulfide linkages in proteins/peptides containing closely spaced or adjacent cysteine residues. The method is rapid and sensitive and utilizes established methodologies including Edman sequencing chemistry. Data interpretation is facile, unlike contemporary methods such as MS/MS fragmentation of disulfide linked peptides. Scrambling was shown to be well controlled through optimization studies and judicious selection of reaction parameters consistent with conditions reported in the current literature.

A limitation of the method is the apparent cumulative losses through repetitive cycling. Therefore, the proximity of the first cysteine residue to the native N-terminus of the protein or peptide of interest will dictate the sample requirements for successful analysis without ambiguity. This highlights the need for judicious selection of hydrolytic reagent for cleavage as close to the cysteine residues as possible. Simply pooling multiple rounds of purifications may afford sufficient material for repetitive analysis. Nevertheless, it is recognized that the cycling efficiency is an area of potential improvement. The methodology presented in this manuscript utilizes traditional analytical scale separations. Improvements in yield and sensitivity may be afforded through coupling the separation with a nanospray ionization source capable of simultaneous fraction collection, such as the Advion NanoMate Triversa. Preliminary data in our laboratory suggests that this may afford significant sensitivity improvements due to the properties of the nanospray interface as well as throughput gains by eliminating the need for separate chromatographic steps for analysis and collection. In a forthcoming paper we will present the application of this methodology in the determination of linkage connectivity to the recently discovered disulfide variants of antibodies of the IgG2 subclass.

SUPPORTING INFORMATION AVAILABLE

Representation of the established insulin amino acid sequence and disulfide structure (peptide annotations (A1, etc.) refer to expected Glu-C peptide nomenclature for the corresponding A or B chain from the N-terminus). In addition a chromatographic assessment of insulin disulfide integrity by reverse-phase separation is shown. This material is available free of charge via the Internet at http://pubs.acs.org.

Received for review May 28, 2009. Accepted July 13, 2009.

AC901161E

(18) Miyashita, M.; Presley, J. M.; Buchholz, B. A.; Lam, K. S.; Lee, Y. M.; Vogel, J. S.; Hammock, B. D. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 4403–4408.