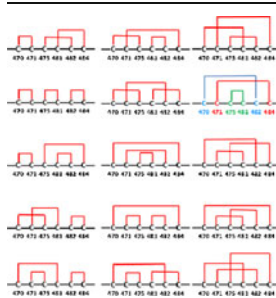


Complete Mapping of a Cystine Knot and Nested Disulfides of Recombinant Human Arylsulfatase A by Multi-Enzyme Digestion and LC-MS Analysis Using CID and ETD

Wenqin Ni,¹ Melanie Lin,² Paul Salinas,² Philip Savickas,² Shiao-Lin Wu,¹
Barry L. Karger¹

¹Barnett Institute of Chemical and Biological Analysis and Department of Chemistry and Chemical Biology, Northeastern University, Boston, MA 02115, USA

²Shire Human Genetic Therapies, Inc., Lexington, MA 02421, USA



Abstract. Cystine knots or nested disulfides are structurally difficult to characterize, despite current technological advances in peptide mapping with high-resolution liquid chromatography coupled with mass spectrometry (LC-MS). In the case of recombinant human arylsulfatase A (rhASA), there is one cystine knot at the C-terminal, a pair of nested disulfides at the middle, and two out of three unpaired cysteines in the N-terminal region. The statuses of these cysteines are critical structure attributes for rhASA function and stability that requires precise examination. We used a unique approach to determine the status and linkage of each cysteine in rhASA, which was comprised of multi-enzyme digestion strategies (from Lys-C, trypsin, Asp-N, pepsin, and PNGase F) and multi-fragmentation methods in mass spectrometry using electron transfer dissociation (ETD), collision induced dissociation (CID), and CID with MS³ (after ETD). In addition to generating desired lengths of enzymatic peptides for effective fragmentation, the digestion pH was optimized to minimize the disulfide scrambling. The disulfide linkages, including the cystine knot and a pair of nested cysteines, unpaired cysteines, and the post-translational modification of a cysteine to formylglycine, were all determined. In the assignment, the disulfide linkages were Cys138–Cys154, Cys143–Cys150, Cys282–Cys396, Cys470–Cys482, Cys471–Cys484, and Cys475–Cys481. For the unpaired cysteines, Cys20 and Cys276 were free cysteines, and Cys51 was largely converted to formylglycine (>70 %). A successful methodology has been developed, which can be routinely used to determine these difficult-to-resolve disulfide linkages, ensuring drug function and stability.

Key words: Cystine knot, ETD, Electron transfer dissociation, Nested disulfides, Free cysteines, Multi-enzymes

Received: 6 August 2012/Revised: 30 September 2012/Accepted: 30 September 2012/Published online: 4 December 2012

Introduction

Cystine knots are a structural motif with three disulfides (six cysteine residues in close proximity in a protein backbone), with one of the disulfides passing through a ring, formed by the other two disulfide bonds [1]. Cystine knots are known to enhance protein structural stability, and they can be found in many proteins with a wide range of

biological functions, such as inhibition, growth stimulation, and cyclization [2, 3]. However, the cystine-knot family shows low sequence homology, and it is therefore hard to predict cystine-knot signatures by sequence alignment. Furthermore, there are 15 ways to form three disulfides into a cystine knot, and the determination of the correct assignment of the disulfide bonds is a challenge. There is a clear need to develop methodology that can be used routinely to determine the disulfide linkages, thus providing structural information for protein function and stability.

Arylsulfatase A (ARSA), a lysosomal enzyme, contains a cystine knot. ARSA catalyzes the hydrolysis of cerebroside sulfate to cerebroside and sulfate. Deficiency of this enzyme cumulates cerebroside sulfate and leads to the destruction of myelin in the central and peripheral nervous systems,

Wenqin Ni and Melanie Lin have contributed equally to this project.

Electronic supplementary material The online version of this article (doi:10.1007/s13361-012-0510-z) contains supplementary material, which is available to authorized users.

Correspondence to: Shiao-Lin Wu; e-mail: si.wu@neu.edu; Barry L. Karger; e-mail: b.karger@neu.edu

resulting in a progressive demyelination disease known as metachromatic leukodystrophy (MLD) [4]. MLD is an autosomal recessive disease with late infantile, juvenile, and adult forms, and is a terminal illness. Most children with the infantile form die by age 5 years. Symptoms of the juvenile form progress with death occurring 10 to 20 years following onset, and those persons affected by the adult form typically die within 6 to 14 years following onset of symptoms.

Patients with MLD have been reported to have a disruption of the cystine knot by the mutation of Cys 470 to Arg [5]. Recent studies have shown that a partially or fully reduced cystine knot makes the protein susceptible to chemical or proteolytic degradation [6]. The conformation of ARSA, forming a homo-dimer protein at neutral pH and a homo-octamer at acidic pH (i.e., in the lysosome), requires proper disulfide linkages. The stability of the enzyme seems to relate to the dimer-to-octamer transition in the lysosomal milieu, in which formation of the octamer has been shown to be disrupted by the replacement of Cys282 with phenylalanine [7].

Recombinant human arylsulfatase A (rhASA) with the sequence homology to ARSA has been investigated for use in enzyme replacement therapy, a potential treatment for MLD patients [8, 9]. Thus, characterization of disulfides in rhASA is an important structure attribute for biopharmaceutical manufacturing to maintain drug function and stability. Currently, experimental approaches for the characterization of cystine knots include X-ray crystallography and nuclear magnetic resonance spectroscopy (NMR) [10], which can be time-consuming (X-ray), and inconclusive because of the near proximity of six cysteines (NMR). The development of an effective and robust methodology is needed for manufacturers' ability to routinely maintain and control the drug quality.

A typical strategy for cysteine characterization included proteolytic cleavage of the backbone using an appropriate endoprotease. Comparison of the reduced and non-reduced peptide maps leads to the identification of the bound and free cysteines [11, 12]. However, it is challenging to correctly assign the disulfide linkages for nested disulfides or cystine knots using this methodology. Recently, Wu et al. established a methodology using LC-MS with collision induced dissociation (CID), electron transfer dissociation (ETD), and CID of the isolated charge-reduced ions (MS^3) to determine complicated and intertwined disulfides including scrambling [13–15]. In this work, we applied a multi-enzyme digestion strategy combined with CID, ETD, and CID- MS^3 to characterize the complete cysteine status in rhASA. Several digestion protocols with different pH were evaluated to determine whether the free cysteines of the protein promote disulfide scrambling under alkaline conditions typically used for enzymatic digestion. By optimizing specific protocols for each cysteine, the status of all the cysteines in rhASA, including the disulfide linkages from the cystine knot and nested disulfide, were successfully determined.

Experimental

Samples

rhASA, GMP lot JPT11001, manufactured by Shire Human Genetic Therapies (Lexington, MA, USA), was provided at 39.1 mg/mL. The sample was aliquoted (10 μ L or 391 μ g per vial) and stored at -80°C before analysis.

Reagents

Sequencing-grade trypsin was purchased from Promega (Madison, WI, USA). Mass spectrometry grade lysyl endopeptidase (Lys-C) was obtained from Wako (Richmond, VA, USA). Pepsin was purchased from MP Biomedicals (Santa Ana, CA, USA). Asp-N, PNGase F, ammonium bicarbonate (NH_4HCO_3), and formic acid (FA) were from Sigma-Aldrich (St. Louis, MO, USA). LC-MS grade water was purchased from J.T. Baker (Phillipsburg, NJ, USA), and HPLC grade acetonitrile from ThermoFisher Scientific (Fairlawn, NJ, USA). Amicon centrifugal filter units (10 kDa MWCO) were obtained from Millipore (Bedford, MA, USA).

Enzymatic Digestion

The protein solution (2.5 μ L of 39.1 mg/mL) was buffer exchanged with 100 mM ammonium bicarbonate (pH 8) or 50 mM Tris-HCl buffer (pH 6.8) using a 10 kDa molecular weight cutoff filter and concentrated to 2 mg/mL (49 μ L). In a separate study, a slightly less than alkaline pH (pH 6.8) was used to examine the effect of pH on the formation of alternative disulfide linkages during the digestion procedure. If a difference was observed, pepsin digestion at pH 2 was used to eliminate scrambling that can occur under basic pH conditions. For pepsin digestion, the protein solution was buffer exchanged with 10 mM HCl (pH 2). Pepsin (1:10, wt/wt) was added to the protein solution and incubated at 37°C for 30 min. The reaction was quenched by adjusting the pH to 6 with 10 mM sodium hydroxide. For Lys-C plus trypsin digestion, the protein solution (pH 6.8 or 8) was incubated with endoproteinase Lys-C (1:50 wt/wt) and trypsin (1:50 wt/wt) for 8 h at room temperature. The enzyme was added a second time (1:50 wt/wt for each enzyme), and the digestion was allowed to occur for an additional 12 h at room temperature. For Lys-C plus trypsin, Asp-N, and PNGase F digestion, the protein solution (pH 6.8 or 8) was treated with the combination of endoproteinase Lys-C (1:50 wt/wt), trypsin (1:50 wt/wt), Asp-N (1:50 wt/wt), and PNGase F (1 units/10 μ g) for 8 h at room temperature; the mixture of enzymes was then added a second time (the same ratio for each enzyme), and the digestion was allowed to occur for an additional 12 h at room temperature. After checking the digestion efficiency, no differences could be observed for PNGase F added either to the same mixture or prior to the mixture. For simplicity, PNGase F is added in the same mixture. In all cases except pepsin digestion, digestion was terminated by the addition of 1 % formic acid. An aliquot of 2 μ g of the enzyme digest was analyzed per LC-MS run.

LC-MS

An Ultimate 3000 nano-LC pump (Dionex, Mountain View, CA, USA) and a self-packed C18 column (Magic C18, 200 Å pore, and 5 µm particle size, 75 µm i.d. × 15 cm) (Magic C18 particle from Michrom Bioresources, Auburn, CA, USA) was coupled to an LTQ-Orbitrap-ETD XL mass spectrometer (ThermoFisher Scientific, San Jose, CA, USA) equipped with a nanospray ion source (New Objective, Woburn, MA, USA). Mobile phase A was 0.1 % formic acid in water, and mobile phase B was 0.1 % formic acid in acetonitrile. The peptides were eluted at 200 nL/min using a linear gradient from 2 % to 60 % B in 90 min, followed by 60 % to 80 % B over 10 min. The LTQ-Orbitrap-ETD XL mass spectrometer was operated in the data-dependent mode to switch automatically between MS (scan 1 in the Orbitrap), CID-MS² (scan 2 in the LTQ), and ETD-MS² (scan 3 in the LTQ). Briefly, after a survey MS spectrum from m/z 300 to 2000, subsequent CID-MS², and ETD-MS² steps were performed on the same precursor ion with $\pm 2.5m/z$ isolation width, with ion/ion reaction duration time being maintained constant throughout the experiment at 100 ms [14]. CID-MS² and ETD-MS² spectra were repeated by targeting specific ions, in order to gain linkage information not obtained in the initial run. These targeted approaches, using the Orbitrap in scans 2 and 3 (if needed), were repeated (e.g., targeting multiple charges of a precursor ion or the same disulfide-linked peptide but with different enzymatic cleavage sites or missed cleavages) until the linkage information was complete. If necessary, the ions of interest obtained with ETD-MS² were targeted for CID-MS³.

Disulfide Assignment

The expected disulfide-linked tryptic or multi-enzyme digested peptide masses with different charges were first calculated and then matched to the observed masses in the LC-MS chromatogram. The matched masses (with <5 ppm mass accuracy for highly abundant ions and <20 ppm for low abundant ions, as determined by 10 % above or below the main peak) were further verified by analysis of the corresponding CID-MS² and ETD-MS² fragmentation spectra, as well as the CID-MS³ fragmentation spectra, as needed. Any internal cleavages (e.g., simultaneous cleavages at both the P1 and P2 polypeptides, or two cleavages within an intra-linked disulfide, were assigned manually. For these assignments, the other cleavages (i.e., the portion cleaved from the association with internal cleavages) should be found to confirm the assignment of the internal cleavages. In our experience, these internal cleavages for disulfide-linked peptides seem to occur more often than typical peptides (without disulfide linkages) and, thus, need to be paid more attention in assignment.

Results and Discussion

The primary structure of recombinant human arylsulfatase A (rhASA) with six disulfides linkages and three unpaired cysteines is shown in Figure 1. The crystal structure of human ARSA (without glycosylation) deduced these six disulfide linkages and three free cysteines, including one that is post-translationally modified to formylglycine [16]. The cystine knot is formed at the C-terminal end of the molecule from the six cysteine residues indicated in the figure. It is important to note that ARSA is a glycoprotein with glycosylation sites indicated in Figure 1. It is clear that the complexity of the protein structure makes the accurate determination of the status of all cysteines very challenging. Multiple strategies are necessary for the elucidation of cysteines, as described below.

Digestion Strategy

A multi-enzyme digestion strategy is clearly needed for the complicated disulfide and unpaired cysteine structure for rhASA. In principle, identification of a single disulfide linkage is straightforward because there is usually only one possibility for connection. Consequently, proteases that can cut proteins into peptides containing only a single disulfide are desired. However, intertwined disulfides or a cysteine-rich region in a protein such as in the case of rhASA may prevent enzyme digestion to the desired peptide size. Preferred peptide sizes are 1 to 5 kDa since peptide recovery and electrospray ionization efficiency can be a problem for larger peptides, while peptides less than 1 kDa may not retain well on a reversed-phase column. In some cases, the disulfide assignment will require further adjustment of peptide sizes to generate peptide lengths with sufficiently high-charge states for effective ETD fragmentation [13]. It should be noted that the enzymatic cleavages of the protein are the same using either trypsin or Lys-C plus trypsin. Nevertheless, the use of Lys-C plus trypsin seems to yield slightly higher digestion efficiency than trypsin alone. The reason could be that the protein size was reduced by Lys-C, leading to a more effective trypsin digestion. Thus, the selection of specific enzymes needs to be carefully considered. Also, for disulfide-linked peptides containing N-linked glycosylation, an additional PNGase F treatment should be considered to reduce the complexity of the mass spectra. For peptides containing free cysteines, the digestion pH for the selected enzymes needs also to be optimized to maintain sufficient enzyme activity while avoiding scrambling. In this study, after surveying several enzyme combinations (Lys-C, trypsin, Asp-N, pepsin, and PNGase F), several protocols were developed for the full cysteine status of rhARSA. Table 1 lists the various digestion protocols including the fragmentation methods for the specific assignments. A detailed description of these steps is in the following sections, beginning with the unpaired cysteines, followed

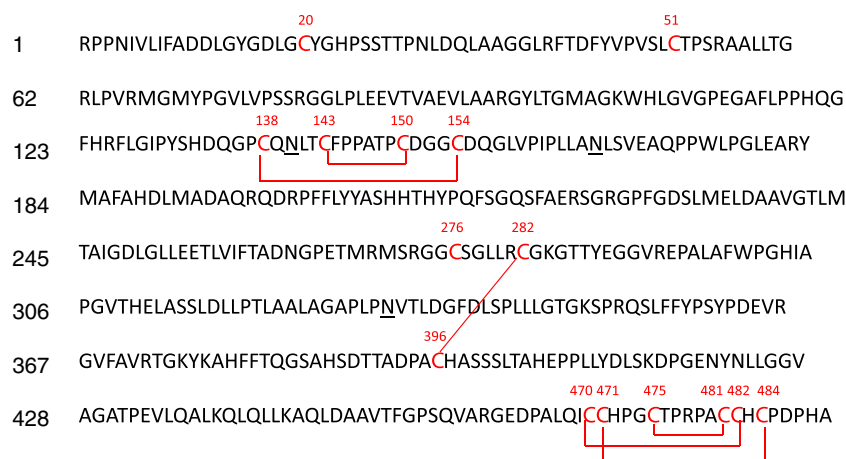


Figure 1. Primary structure of rhASA with disulfide linkages and unpaired cysteines. Note: the sites of disulfide linkages and unpaired cysteines were colored in red and the N-glycosylation sites were underlined

Table 1. Unpaired Cysteine and Disulfide Linkages in rhASA

#	Site	Status	Enzyme	Corresponding peptide sequence	I.D. (method)	Theoretical mass	Observed mass
1	C20	Free Cys	Pepsin	G ²⁰ CYGHPSSTTPNL (19-31)	CID	667.2957 (2+)	667.2965 (2+)
2	C51	Cys as fgly	Pepsin	YVPVSL ⁵¹ C(fgly)TPSRAAL (45-58)	CID	729.9012 (2+)	729.9020 (2+)
3	C51	Free Cys	Pepsin	YVPVSLCTPSRAAL (45-58)	CID	738.8976 (2+)	738.9059 (2+)
4	C276	Free Cys	Pepsin	RMSRGG ²⁷⁶ C ²⁸² SGL (270-279)	CID	512.2448 (2+)	512.2453 (2+)
5	C282-C396	1 disulfide linkage	Pepsin	LR ²⁸² C ³⁹⁶ GKGTTYEGGVRE (280-294) FTQGS ³⁹⁶ AHS ³⁹⁶ DTTADPACHASSSL (381-402)	CID/ETD	1271.9117 (3+)	1271.9153 (3+)
6	C138-C154 C143-C150	2 disulfide linkages	Lys-C+ Trypsin+ Asp-N+ PNGase F	DQG ¹³⁸ P ¹⁴³ C ¹⁵⁰ Q (134-139) DLT ¹³⁸ CFPPATPCDGG ¹⁵⁴ C (140-154)	CID/ETD	1069.8997 (2+)	1069.9006 (2+)
7	C470-C482 C471-C484 C475-C481	3 disulfide linkages	Pepsin	PALQI ⁴⁷⁰ C ⁴⁷¹ CH ⁴⁷⁵ PGCT ⁴⁸¹ PRPA ⁴⁸² C ⁴⁸⁴ CH ⁴⁸⁴ CPD PHA (465-489)	CID/ETD/ CIDMS3	874.3564 (3+)	874.3596 (3+)

fgly formylglycine, in which the side chain of cysteine, $-\text{CH}_2\text{SH}$ is converted to $-\text{CHO}$, Cys residues are in red as: C for free cysteine, C(*fgly*) for cysteine coupling with *fgly*, and C-C for disulfide linkage.

by the single disulfide, and then the nested disulfide, with the final section dealing with the complicated cystine knot.

Unpaired Cysteines: Cys20, Cys51, Cys276

When trypsin or Lys-C plus trypsin digestion was used to assign the unpaired cysteines, disulfide scrambling, which formed various disulfides mainly among the free cysteines, was observed using a standard digestion buffer at pH 8 (~40 %), and to a lesser extent at pH 6.8 (~5 %). As expected, we did not observe scrambled disulfides with pepsin digestion at pH 2. It should be noted that the scrambled disulfides obtained at higher pH (i.e., pH 8) provided us the types and linkage information. Thus, we could target these scrambled disulfides at the lower pH analysis. Although the amount of scrambled disulfides could be lower at the lower pH, the targeting approach (extraction of specific ions for targeted MSⁿ) should be more sensitive than in the initial discovery mode. Although pepsin digestion is nonspecific, major cleavages were found after

leucine residues (C-terminal side), followed by aromatic amino acids, proline, and glutamic acid residues. Thus, we could focus on these cleavages for potential scrambled disulfides as well. The major pepsin fragment containing the unpaired Cys20 was identified. The peptide with the corresponding mass and charge is shown in Figure 2a. As shown, the precursor ion scan was performed in the Orbitrap, and its monoisotopic mass accurately matched the theoretical peptide mass with an unmodified free cysteine, as m/z 667.2965 (observed) matched to m/z 667.2957 (theoretical) for the 2+ charge state. The site of the free cysteine was determined by CID-MS² of the precursor ion, as shown in Figure 2b.

The remaining unpaired cysteines were identified in a similar manner, as shown in the supplementary material, Figure S1A and B (Cys51 converted to formylglycine), Figure S2A and B (Cys51, a free cysteine), and Figure S3A and B (Cys276, a free cysteine). Table 1 (#1, #2, #3, and #4) summarizes the assignments for all the unpaired cysteines. At Cys51, it contains more than 70 % of the formylglycine

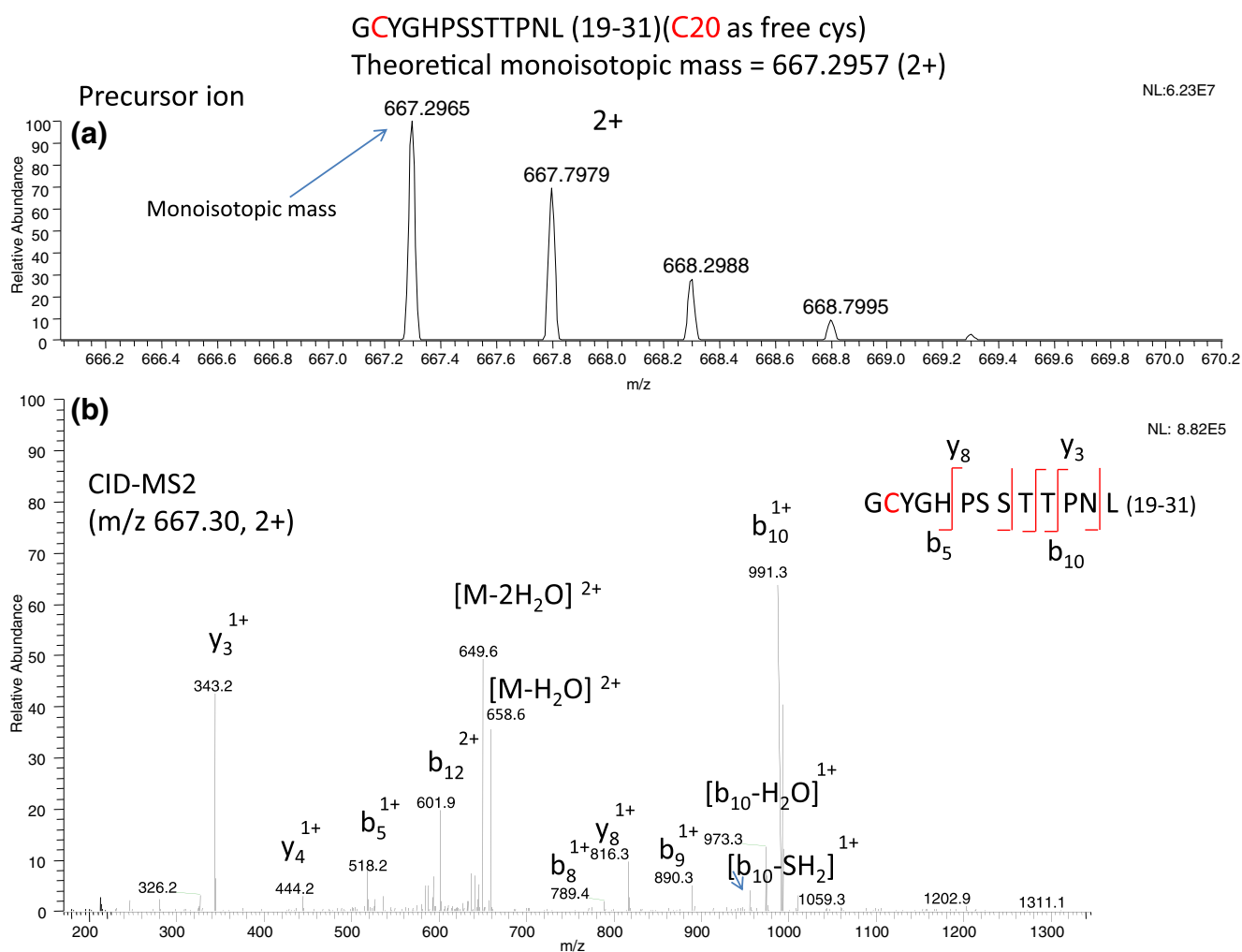


Figure 2. (a) Mass and charge of the pepsin-digested peptide with an unmodified (free) Cys20, and (b) CID-MS² spectrum of the precursor from (a). The sequence and theoretical mass of the peptide are indicated in the insert (a)

form. Since the ionization efficiency of the peptide containing the free cysteine could be different from that containing the formylglycine, the ratio of the two is a rough estimation of the formylglycine conversion.

Single Disulfide: Cys282–Cys396

For the peptide with a single disulfide (Cys282–Cys396, #5 in Table 1), the linkage assignment was straightforward. Although alkali pH (i.e., pH 8) should not cause the disulfide-linked cysteines to scramble, the other free cysteines in the protein could potentially cross react with the disulfide at alkali pH. Indeed, a minute amount of cross-reacted disulfides was observed using a digestion buffer at pH 8. As expected, no cross-reacted disulfides could be observed when trypsin at pH 6.8 or pepsin (pH 2) was used for the digestion. The assignment of the disulfide-linked peptide is illustrated by the pepsin digestion protocol in the supplementary material (Figure S4). In Figure S4A, the observed accurate mass matched the theoretical peptide mass with one disulfide (a loss of 2H from the backbone sequence). The corresponding CID-MS² spectrum, b and y ions in Figure S4B, verified the correct sequence. For the corresponding ETD-MS² spectrum, the disulfide bond was preferentially dissociated as expected [14, 15], resulting in two dissociated peptides designated as P1 and P2 (Figure S4C), which confirms that the two peptides are linked together.

In this disulfide linkage assignment, we used the low pH approach instead of alkylating the free cysteine to prevent scrambling since it is difficult to control the alkylation properly. Alkylation under denaturing conditions, the scrambling (particularly from free cysteine) can occur quickly before the alkylation step. If the protein is at native conditions, the alkylation step is often not optimized. Incomplete alkylation could cause the remaining free cysteine to scramble as well. Therefore, we used low pH to protonate (inactivate) the free cysteine to assign the disulfide linkages. In addition, with the use of low pH (not to alkylate the free cysteine), the free cysteine in the sequence can be assigned as well, as shown in Figure 2, as well as in Figure S2 and Figure S3 in the supplementary material.

Nested Disulfide: Cys138–Cys154 and Cys143–Cys150

As shown in Figure 1, the cysteines for the nested disulfides are located in Cys138–Cys154 and Cys143–Cys150. Since there are four cysteines, other potential linkages could be either as two separate disulfides (Cys138–Cys143 and Cys150–Cys154) as well as two crossed disulfides (Cys138–Cys150 and Cys143–Cys154) (see Figure S5 in the supplementary material). Furthermore, the complexity is increased by two N-linked glycosylation sites, one within, and the other next to the two disulfides (N underlined in Figure 1). To reduce the complexity, the two N-linked glycans were removed with PNGase F, converting Asn (N) to Asp (D). This conversion provided a target for Asp-N

digestion. Thus, in addition to Lys-C plus trypsin (to reduce the protein size), the addition of PNGase F and Asp-N enzymes effectively cut the disulfide-linked peptide to a suitable size for mass spectrometric analysis (see Figure S6 in the supplementary material). These nested disulfide bonds form a ring, which significantly reduces CID fragmentation efficiency for the amino acids inside the ring [13], thereby complicating the assignment for disulfide linkages inside the ring. Although ETD is effective to break the disulfides, the peptide length obtained by trypsin or pepsin alone is too large for effective fragmentation ($m/z > 1000$) [13]. The digestion protocol required four enzymes to obtain the proper size for effective fragmentation by mass spectrometry (see Figure S6 and Figure S7 in the supplementary material). The assignment of the disulfide-linked peptide based on the mass spectra is shown in Figure 3. In Figure 3a, the observed accurate mass matched the theoretical peptide mass with two disulfides (a loss of 4H from the backbone sequence). Since the ring structure formed by nested disulfides was broken by the additional Asp-N digestion, the disulfide linkages could be conclusively assigned as long as cleavages can be observed in the backbone between the CDGGC amino acid residues. As shown in Figure 3b, the y_1 , y_3 , b_{11} , and b_{12} fragments in the CID-MS² spectrum provide strong evidence for the linkages Cys138 with Cys154, and Cys143 with Cys150. In addition to the CID-MS² spectrum, the corresponding ETD-MS² spectrum (Figure S8 in the supplementary material) confirms that the two linked peptides (P1 and P2) are connected.

It should be noted that although Asp-N should cleave aspartic acid in the protein backbone, the aspartic acid residue adjacent to a cysteine (a disulfide) inside the ring was not cleaved (see Figure S6C in the supplementary material). For digestion at pH 8, a significantly scrambled disulfide was observed at a different LC retention time (see Figure S9A and B in the supplementary material), as the structure resembled to scramble 1 in Figure S5 (note: no scramble 2 could be observed). Without chromatographic separation, it would be difficult to determine the disulfide isomers since they are often isobaric. Nevertheless, the scrambled disulfides often have different configurations from the correct one. Thus, a different LC retention time or a shoulder peak with identical mass as the correct one is often the potential area for evaluation of scrambling. At pH 6.8, the scrambled disulfide was reduced to a trace amount and could not be observed at pH 2 with pepsin. Although the pepsin-digested disulfide could not be effectively fragmented by CID, the fragmentation did indirectly confirm the nested disulfide linkage (see Figure S10 in the supplementary material). ETD was also tested to fragment the pepsin-digested disulfide but was not successful, due to minimal fragmentation and mainly charge-reduced species in the ETD spectrum. Although CID-MS³ and even MS⁴ have been attempted to fragment the charge-reduced species, the fragmentation efficiency was still poor for the peptide of this size. As described above, the use of an additional enzyme (i.e., Asp-N) to obtain a proper size and configuration of the disulfides was critical.

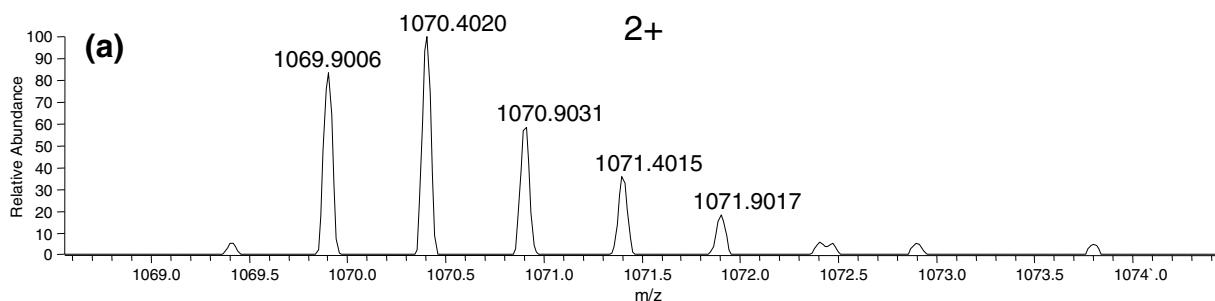
Precursor ion

DQGPCQ (134-139) (C138-C154)

DLTCFPPATPCDGGC (140-154) (C143-C150)

Theoretical monoisotopic mass = 1069.8997 (2+)

NL1.71E6



CID-MS2

(m/z 1069.91, 2+)

NL: 1.01E4

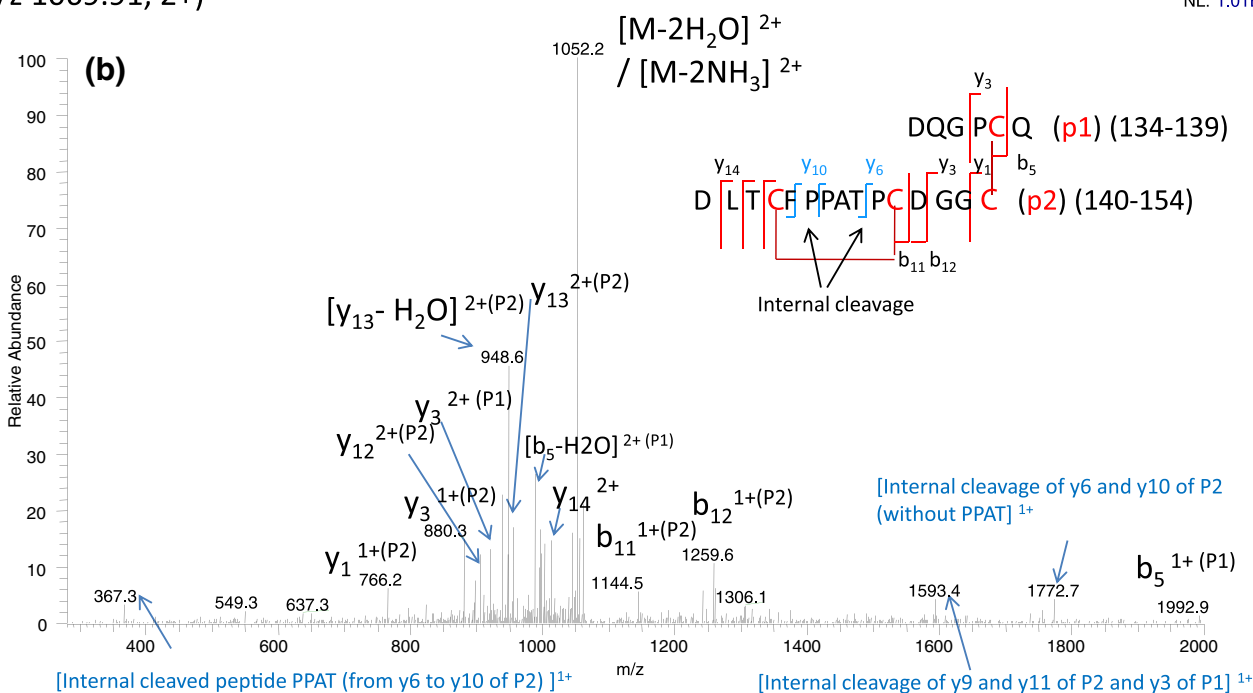


Figure 3. (a) Mass and charge of the Lys-C plus trypsin plus Asp-N plus PNGaseF-digested peptide with two disulfides (Cys138 with Cys154, and Cys143 with Cys150), and (b) CID-MS² spectrum of the 2+ charged precursor from (a). The sequence and theoretical mass of the peptide are indicated in the insert of (a)

Cystine Knot: Cys470–Cys482, Cys471–Cys484, and Cys475–Cys481

The cystine knot could not be broken by any of the enzymes or combination of enzymes employed. In addition, CID fragmentation could not produce backbone cleavages within the cystine knot. Thus, ETD was examined. For the amino acid sequence in this region, pepsin digestion was selected in order to obtain the proper peptide length with less

acidic residues for effective fragmentation by ETD (i.e., eliminated additional glutamic and aspartic acid residues as compared to the corresponding tryptic fragment). The corresponding mass and charge of the pepsin-digested peptide is shown in Figure S11 (in the supplementary material). The monoisotopic mass matched the expected peptide mass with three disulfides (a loss of 6H from the backbone sequence). Limited sequence information, as expected, was obtained by CID-MS² (Figure S12). Never-

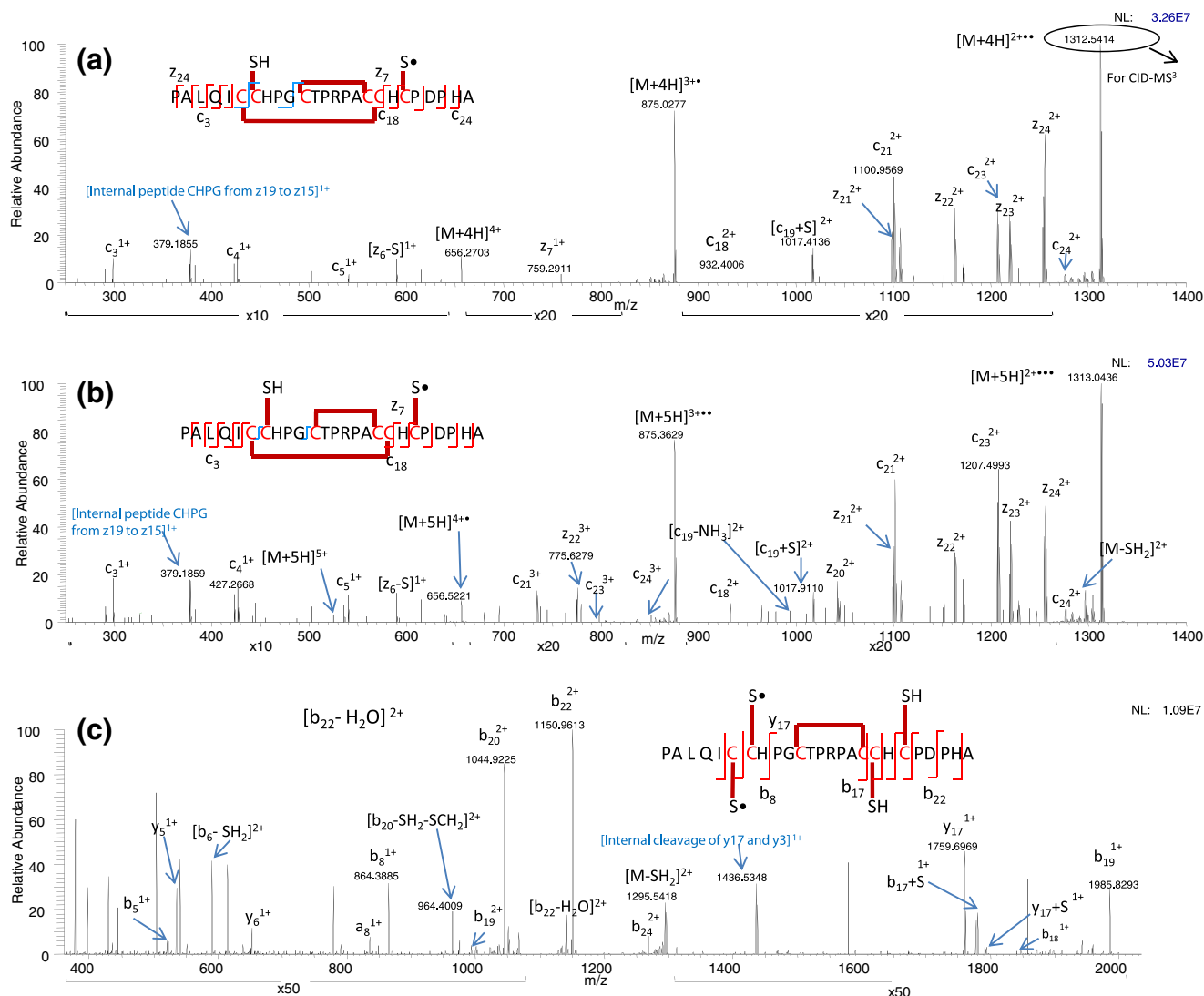


Figure 4. (a) ETD-MS² spectrum (using the Orbitrap) of the cysteine knot precursor (m/z 656.30, 4+), (b) ETD-MS² spectrum (using the Orbitrap) of the same peptide from (a) but with different charged precursor (m/z 525.20, 5+), and (c) CID-MS³ spectrum (using the Orbitrap) of m/z 1312.6 from (a). For the spectra measured in the Orbitrap (a), (b), and (c), only the highest abundant isotopic mass is shown for each ion. The monoisotopic mass of these individual ions are listed in Table S1

theless, and significantly, ETD-MS² dissociated the disulfides, which allowed cleavage of the peptide backbone, as shown in Figure 4a and b. The fragmentation of this disulfide-linked peptide but for two different charge states is shown in Figure 4a (m/z 656.30, 4+) and Figure 4b (m/z 525.20, 5+). The fragmentation data from the two different charge states demonstrates consistency with respect to cleavage sites and verifies that the linkage assignments are correct. Since the peptide was linked through three intertwined disulfides, a partial reduction of a particular disulfide (with mass shift by only 1 Da), the high resolution-accurate mass instrument (Orbitrap) with ETD provided even more convincing evidence for the disulfide bond assignments. As seen in both ETD spectra of Figure 4a and b, z₇ and c₁₈, along with the internal cleavages from the dissociated disulfide, confirm the connection between Cys471 and

Cys484. In addition, one of the charge-reduced species (m/z 1312.6, $[M + 4H]^{2+}$) in the ETD spectrum was further fragmented (CID-MS³ using the Orbitrap) as shown in Figure 4c. The MS³ spectra contain additional disulfide and backbone cleavages, such as y₁₇ and b₈, confirming the connection between Cys470 and Cys482. The fragmentation pattern and assignments were also observed with the same CID-MS³ spectrum generated in the LTQ ion trap (Figure S13 in the supplementary material), which makes the method applicable even with low resolution MS instruments. After assigning the two disulfide linkages, the non-dissociated (the third) disulfide was left with the only possible remaining connection, which was a linkage between Cys475 and Cys481. In summary, the combination of ETD-MS² and CID-MS³ mass spectral analysis confirms the linkage sites as Cys470 with Cys482, Cys471 with

Cys484, and Cys475 with Cys481. The theoretical and observed fragment ions are listed in the supplementary material (Table S1).

Conclusions

In this study, in-depth LC-MS protocols have been developed to assign the status of all 15 cysteine residues in rhASA, including the disulfide linkages from the nested disulfide and cystine knot. Although both cystine knot and nested disulfides are difficult to resolve, strategies with a combination of different enzymes and MS fragmentation methods could successfully determine the assignments. The successful assignment of the disulfide linkages in the cystine knot demonstrates the power of the approach, which should be generally useful for other cystine knots. Using the described methods, it becomes feasible to monitor the disulfide linkages of recombinant rhASA.

Acknowledgments

The authors acknowledge support for this work in part by NIH grant GM 15847. Contribution number 1006 from the Barnett Institute.

References

- Lenguyen, D., Heitz, A., Chiche, L., Castro, B., Boigegrain, R.A., Favel, A., Colettipreviero, M.A.: Molecular recognition between serine proteases and new bioactive microproteins with a knotted structure. *Biochimie* **72**(6/7), 431–435 (1990)
- Alvarez, E.; Cahoreau, C.; Combarnous, Y.: Comparative structure analyses of cystine knot-containing molecules with eight aminoacyl ring including glycoprotein hormones (GPH) alpha and beta subunits and GPH-related A2 (GPA2) and B5 (GPB5) molecules. *Reprod. Biol. Endocrinol.* **7**(90), (2009)
- Daly, N.L., Craik, D.J.: Bioactive cystine knot proteins. *Curr Opin Chem Biol* **15**(3), 362–368 (2011)
- von Figura, K., Gieselmann, V., Jaeken, J. (eds.): Metachromatic leukodystrophy: Lysosomal disorders. McGraw-Hill, New York (2011)
- Coulter-Mackie, M.B., Gagnier, L.: Spectrum of mutations in the arylsulfatase A gene in a Canadian DNA collection including two novel frameshift mutations, a new missense mutation (C488R) and an MLD mutation (R84Q) in cis with a pseudodeficiency allele. *Mol Genet Metab* **79**(2), 91–98 (2003)
- Kolmar, H.: Alternative binding proteins: biological activity and therapeutic potential of cystine-knot miniproteins. *FEBS J* **275**(11), 2684–2690 (2008)
- Marcao, A., Azevedo, J.E., Gieselmann, V., Sa Miranda, M.C.: Oligomerization capacity of two arylsulfatase A mutants: C300F and P425T. *Biochem Biophys Res Commun* **306**(1), 293–297 (2003)
- Matzner, U., Lullmann-Rauch, R., Stroobants, S., Andersson, C., Weigelt, C., Eistrup, C., Fogh, J., D'Hooge, R., Gieselmann, V.: Enzyme replacement improves ataxic gait and central nervous system histopathology in a mouse model of metachromatic leukodystrophy. *Mol Ther* **17**(4), 600–606 (2009)
- Matzner, U., Herbst, E., Hedayati, K.K., Lullmann-Rauch, R., Wessig, C., Schroder, S., Eistrup, C., Moller, C., Fogh, J., Gieselmann, V.: Enzyme replacement improves nervous system pathology and function in a mouse model for metachromatic leukodystrophy. *Hum Mol Genet* **14**(9), 1139–1152 (2005)
- Yu, B., Millhauser, G.L.: Chemical disulfide mapping identifies an inhibitor cystine knot in the agouti signaling protein. *FEBS Lett* **581**(29), 5561–5565 (2007)
- Keck, R.G., Berleau, L., Harris, R., Keyt, B.A.: Disulfide structure of the heparin binding domain in vascular endothelial growth factor: characterization of posttranslational modifications in VEGF. *Arch Biochem Biophys* **344**(1), 103–113 (1997)
- Wu, J., Watson, J.T.: A novel methodology for assignment of disulfide bond pairings in proteins. *Protein Sci* **6**(2), 391–398 (1997)
- Wu, S.L., Jiang, H., Hancock, W.S., Karger, B.L.: Identification of the unpaired cysteine status and complete mapping of the 17 disulfides of recombinant tissue plasminogen activator using LC-MS with electron transfer dissociation/collision induced dissociation. *Anal Chem* **82**(12), 5296–5303 (2010)
- Wu, S.L., Jiang, H., Lu, Q., Dai, S., Hancock, W.S., Karger, B.L.: Mass spectrometric determination of disulfide linkages in recombinant therapeutic proteins using online LC-MS with electron-transfer dissociation. *Anal Chem* **81**(1), 112–122 (2009)
- Wang, Y., Lu, Q., Wu, S.L., Karger, B.L., Hancock, W.S.: Characterization and comparison of disulfide linkages and scrambling patterns in therapeutic monoclonal antibodies: using LC-MS with electron transfer dissociation. *Anal Chem* **83**(8), 3133–3140 (2011)
- Lukatela, G., Krauss, N., Theis, K., Selmer, T., Gieselmann, V., von Figura, K., Saenger, W.: Crystal structure of human arylsulfatase A: the aldehyde function and the metal ion at the active site suggest a novel mechanism for sulfate ester hydrolysis. *Biochemistry* **37**(11), 3654–3664 (1998)