Automated Data Interpretation Based on the Concept of "Negative Signature Mass" for Mass-Mapping Disulfide Structures of Cystinyl Proteins

Jianfeng Qi* and Wei Wu
Department of Chemistry, Michigan State University, East Lansing, Michigan, USA

Chad R. Borges
Department of Biochemistry, Michigan State University, East Lansing, Michigan, USA

Dehua Hang, Matthew Rupp, and Eric Torng
Departments of Computer Science and Engineering, Michigan State University, East Lansing, Michigan, USA

J. Throck Watson
Departments of Chemistry and Biochemistry, Michigan State University, East Lansing, Michigan, USA

An efficient method for data processing and interpretation is needed to support and extend disulfide mass-mapping methodology based on partial reduction and cyanylation-induced cleavage to proteins containing more than four cystines. Here, the concept of "negative signature mass" is introduced as the novel feature of an algorithm designed to identify the disulfide structure of a cystinyl protein given an input of mass spectral data and an amino acid sequence. The "negative signature mass" process is different from the conventional approach in that it does not directly rule-in disulfide linkages, but rather eliminates linkages from a list of all possible theoretical linkages, with the goal of ruling out enough linkages so that only one disulfide structure can be constructed. The operating principles and the effectiveness of the algorithm are described in the context of analyzing ribonuclease A, a 124-residue protein containing eight cysteines in the form of four cystines (disulfides). (J Am Soc Mass Spectrom 2003, 14, 1032–1038) © 2003 American Society for Mass Spectrometry

Determination of the disulfide structure is a critical aspect of characterizing a cystinyl protein. During the last 25 years, most disulfide structures have been determined by a combination of Edman sequencing and mass mapping methodology based on proteolysis; however, this approach has serious limitations, such as not being applicable to cystinyl proteins where a cleavage site is not available between cysteines and being prone to artifacts due to disulfide scrambling [1–3]. In recent years, we have developed a novel chemical approach (based on partial reduction and cyanylation) for mapping the disulfide structure of cystinyl proteins that has been shown to be applicable to the analytical challenge of adjacent cysteines, while avoiding disulfide scrambling by functioning at low pH [4–6]. However, two key issues need to be addressed in order to scale up this technique so that it can handle proteins with more than four cystines: Eliminating the need to isolate partially reduced (and cyanylated) isoforms and automating data interpretation.

Here, we introduce the concept of a "negative signature mass algorithm" (NSMA) that addresses the issues identified above. In the following, we describe the operating principles of the NSMA in the context of mass analysis of all the experimentally observed partially reduced and cyanylated isoforms of native ribonuclease A (RNase A).

Methods

Synopsis of Cyanylation-Based Mass Mapping of Disulfide Bonds within Cystinyl Proteins

The method is based on selective cyanylation of cysteine; nucleophilic attack of this species causes cyanylation (CN)-induced cleavage of the peptide backbone

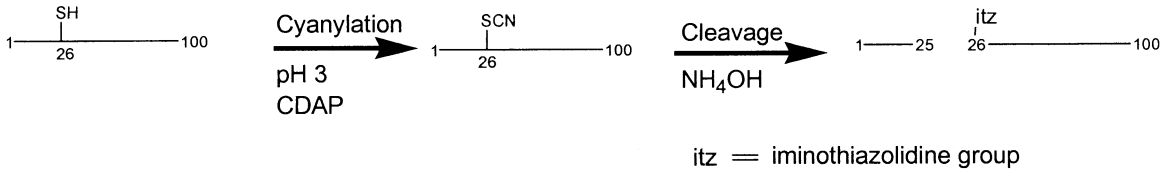
Published online July 31, 2003

Address reprint requests to Dr. J. T. Watson, Departments of Chemistry and Biochemistry, Michigan State University, 401 Biochemistry Building, East Lansing, MI 48824-1319. E-mail: watsonj@msu.edu

*Current address: EPIX Medical Inc., 71 Rogers St., Cambridge, MA 02142, USA.

1044-0305/03/$30.00
doi:10.1016/S1044-0305(03)00145-4

Received December 23, 2002
Revised February 25, 2003
Accepted February 25, 2003



Scheme 1. Selective cyanylation of sulfhydryls, and subsequent cleavage of the peptide bond on the N-terminal side of cyanylated cysteines.

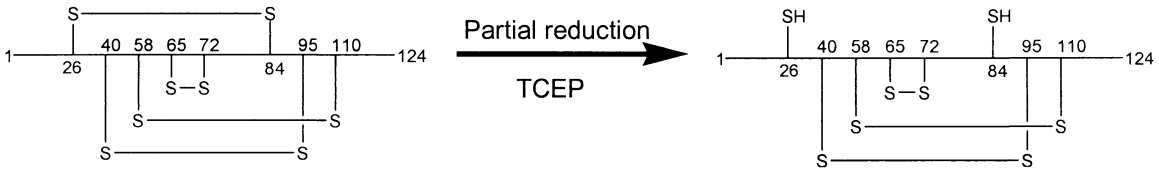
on the N-terminal side of the modified cysteine as represented in Scheme **1** [7]. Because cystines (disulfide bonds) are nonreactive with the cyanylating reagent, cyanodiaminopyridinium tetrafluoroborate (CDAP), a multi-cystinyl protein must be partially reduced (to generate specific pairs of free sulfhydryl groups) using tris-carboxyethylphosphine (TCEP), as illustrated in Scheme **2** for RNase A [8]. At the right in Scheme **2** is shown one of the four possible singly reduced isoforms of this 4-disulfide protein; in this case, the disulfide connecting cysteines 26 and 84 has been reduced.

Once cyanylated and subjected to nucleophilic attack in 1M NH_4OH, the singly reduced isoform of RNase A from Scheme **2** is cleaved at the peptide bond between residues 25 and 26, and also between residues 83 and 84. However, as shown in the middle of Scheme **3**, residual disulfide bonds hold two of the three cleavage products together in this case. Thus, the cleavage reaction mixture represented by the middle of Scheme **3** is treated with excess reducing reagent (TCEP) to reduce all residual disulfide bonds, thereby ensuring that all CN-induced cleavage products are free as shown at the end of Scheme **3**.

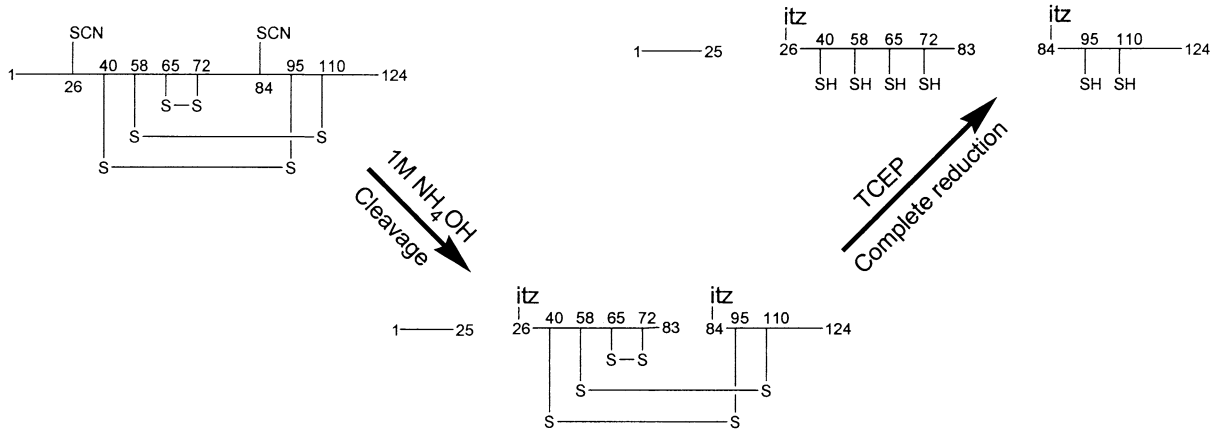
The cleavage reaction mixture represented at the right in Scheme **3** was analyzed by matrix-assisted laser-desorption/ionization (MALDI) mass spectrometry (MS) to produce the mass spectrum shown in Figure 1; the key experimental data are summarized in Table 1, together with calculated values for the indicated fragments, as obtained under condition described previously [4]. For this didactic example of manual data interpretation, the singly reduced and cyanylated isoforms of RNase A were separated by HPLC as four fractions [4]; one fraction was subjected to cleavage with 1 M NH_4OH and complete reduction by TCEP, and the resulting reaction mixture was analyzed by MALDI to produce the mass spectrum shown in Figure 1. The good agreement between the experimental data and the calculated masses identifies this fraction as a singly reduced isoform of native RNase A illustrated in Schemes **2** and **3**. From a knowledge of the amino acid sequence of RNase A, mass mapping of the observed cleavage products (peaks at *m/z* 2705.3, 4527.4, and 6548.5 in Figure 1 and Table 1) can be explained only by cleavage of the peptide bonds between residues 25 and 26, and between residues 83 and 84. Cyanylation-induced cleavage occurs only at cyanylated cysteines [7]; thus, Cys26 and Cys84 must have been available for cyanylation during the chemical processing. Because cysteines 26 and 84 are the only two free sulfhydryls available in this particular singly reduced isoform, they must have been connected to one another in the native structure of RNase A. The peak at *m/z* 9176.7 represents β-elimination of HSCN at Cys26 and cleavage at Cys84, and confirms the information available from peaks at *m/z* 2705.3 and 6548.5 that correspond to cleavage products 1–25 and itz-26–83, respectively; the peak at *m/z* 10,998.6 represents cleavage at Cys26 and β-elimination at Cys84, and confirms the information available from peaks at *m/z* 6548.5 and 4527.4, which correspond to cleavage products itz-26–83 and itz-84–124, respectively. Note that the sites of β-elimination in these fragments (except those with only one internal cysteine residue) cannot be deduced from the mass spectral data alone; these sites can only be indirectly inferred after the sites of cyanylation in the singly reduced and cyanylated isoform are identified. In the situation where singly reduced and cyanylated isoforms are not separated prior to the cleavage, it is impossible to pinpoint the sites of β-elimination. Therefore, these β-elimination peaks do not help deduce the disulfide linkages, but they can serve as an internal consistency check for the data interpretation.

Graph-Theoretic Interpretation of the Problem

A mathematical graph G is composed of a set of nodes V and a set of edges E where each edge is a pair of nodes from V [9]. In the problem at hand, we can think of the set of cysteine residues as our set of nodes and we can think of the set of possible disulfide linkages as our



Scheme 2. Partial reduction of RNase A showing one of four possible singly reduced isoforms.



Scheme 3. Cleavage at cyanylated cysteines followed by complete reduction to yield three CN-induced cleavage products.

set of edges. A disulfide structure is defined as a set containing a specified number of possible disulfide linkages in which no cysteine is involved in more than one linkage. In graph theory, a disulfide structure corresponds to a "matching"; a matching is a set of edges that do not share any nodes. In this work, we assume that all cysteines are involved in disulfide linkages. Cysteines that are not involved in disulfide linkages can be easily identified [10] and thus removed from the set of nodes. In graph theory, a disulfide structure now corresponds to a maximum or perfect matching where all nodes are matched.

A key observation is that the number of possible edges (possible disulfide linkages) grows quadratically in the number of nodes (cysteines) while the number of possible perfect matchings grows exponentially in the number of nodes (cysteines) assuming all linkages are present. For example, for RNase A in which all eight cysteines are oxidized to four cystines, there are 105 possible perfect matchings (disulfide structures) while there are only 28 possible edges (disulfide linkages). Three of the 105 possible perfect matchings for RNase A are illustrated in Figure 2.

The Difficulty of Trying to Identify Singly Reduced Isoforms Directly

Manually interpreting mass spectral data obtained from direct analysis of singly reduced isoforms such as the one represented at the right in Scheme **2** required isolation by HPLC. This practice is time-consuming, and it may not be possible to achieve for more complex mixtures of partially reduced isoforms expected from multi-cystinyl proteins containing five or more disulfide bonds. Our development of the NSMA alleviates, in principle, the need for isolation of the partially reduced (and cyanylated) isoforms and greatly simplifies data interpretation.

The Concept of Negative Signature Mass

The NSMA operates by ruling out certain possible linkages (edges) based on the recognition of mass spectral peaks at mass values corresponding to protonated CN-induced cleavage fragments containing free cysteine residues (nodes); for example, there are four such residues in itz-26–83 in Scheme **4**. Masses of such fragments are deemed "negative signature masses". This process is different from the conventional approach in that it does not directly rule-in linkages (edges), but rather eliminates linkages (edges) from a list of all possible theoretical linkages, with the goal of ruling out enough linkages so that only one disulfide structure (perfect matching) can be constructed from the remaining list of potential disulfide linkages.

To recognize negative signature masses, the algorithm calculates the masses of all possible CN-induced cleavage fragments based on the protein sequence information and the masses of the constituent amino acid residues, then compares this theoretical list to experimental mass spectral data. Matches, to within a user-specified error limit, indicate which fragments were present in the cleavage reaction mixture. If cysteine residues (nodes) containing a free sulfhydryl group are part of a recognized fragment, the algorithm rules out linkage (edge) between the terminal cysteines (those at which CN-induced cleavage occurred) and the internal cysteine residue(s) such as those shown at the top of Scheme **4**. For example, the CN-induced cleavage fragment itz-26–83 serves as a negative signature mass. Detection of the fragment itz-26–83, which as shown in Scheme **4** contains internal sulfydryls at residues 40, 58, 65, and 72, means that no disulfide bond (edge) in the original protein could have existed between Cys26 and Cys40 or Cys58 or Cys65 or Cys72, nor between Cys84 and Cys40 or Cys58 or Cys65 or Cys72.

Had there been a disulfide (edge) between Cys26 and Cys40, quite a different cleavage product would have resulted; partial reduction would have exposed

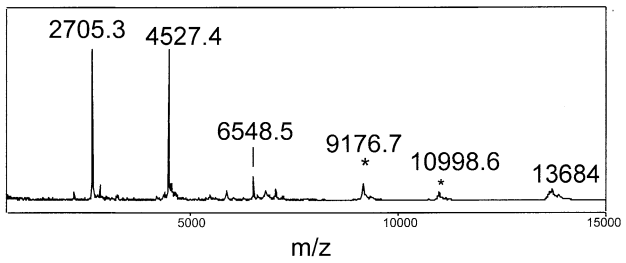


Figure 1. MALDI mass spectrum of the cleavage reaction mixture of a singly reduced isoform of RNase A [4].

Cys40 as a free sulfhydryl simultaneously with Cys26, and both Cys26 and Cys40 would have become cyanylated as shown near the bottom of Scheme **4**. Subsequent CN-induced cleavage would have produced the three fragments shown near the bottom of Scheme **4** (1–25, itz-26–39, and itz-40–124). Note that itz-26–83 is not among them. Thus, in this way, itz-26–83 serves as a negative signature mass for the otherwise theoretically possible linkage (edge) between Cys26–Cys40.

Similarly, detection of itz-26–83 (Scheme **4**) also rules out the existence of disulfides (edges) Cys26–Cys58, Cys26–Cys65, etc. as listed near the top of Scheme **4**. This is the basis by which the NSMA eliminates possible linkages (edges) from consideration, thereby simplifying data interpretation.

Software Development

It is possible to manually apply the logic of "negative signature mass" (Scheme **4**) to carry out analysis of the mass spectral data, but it is excessively tedious and time consuming. For this reason, a program was written using Microsoft Visual C++ 6.0 that allows the user to input protein sequence information and mass spectral data including error limits. The program then internally performs mass mapping, and, using the "negative signature mass" logic, returns a list of possible disulfide linkages (edges) and structures (matchings) that fit the input data.

Results and Discussion

The NSMA was tested with mass spectral data collected during analysis of RNase A using CN-induced cleavage products of singly reduced isoforms [4]. Heretofore, to facilitate manual data interpretation, it has been our practice to purify and physically isolate (by RP-HPLC) isomeric forms of the partially reduced protein (differing by number and position of disulfide bond reduction events). However, in the test of the NSMA described below, the mass spectral data from analysis of the various CN-induced cleavage products (listed in Table 3) were entered into the NSMA as a single batch even though each of the four singly reduced isoforms had been analyzed separately in previous work [4].

Table 1. Summary of key mass spectral data from Figure 1 and the corresponding cleavage product

Cleavage Product	Calc'd Mass	Expmt'l
1——25	2706.8 Da	2705.3 Da
itz-26 SH SH SH SH 40 58 65 72 83	6547.3 Da	6548.5 Da
itz-84 SH SH 95 110 124	4526.0 Da	4527.4 Da

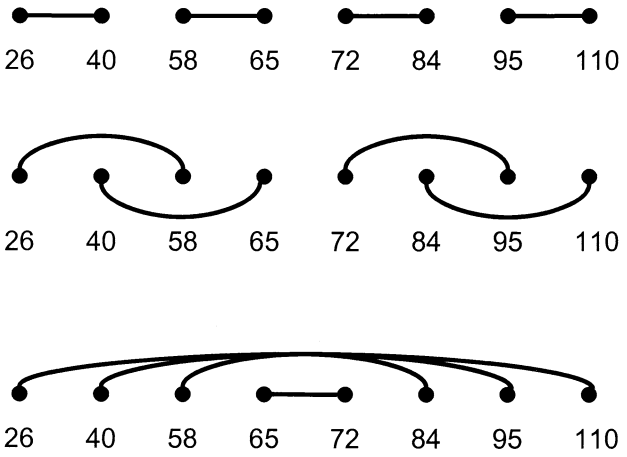
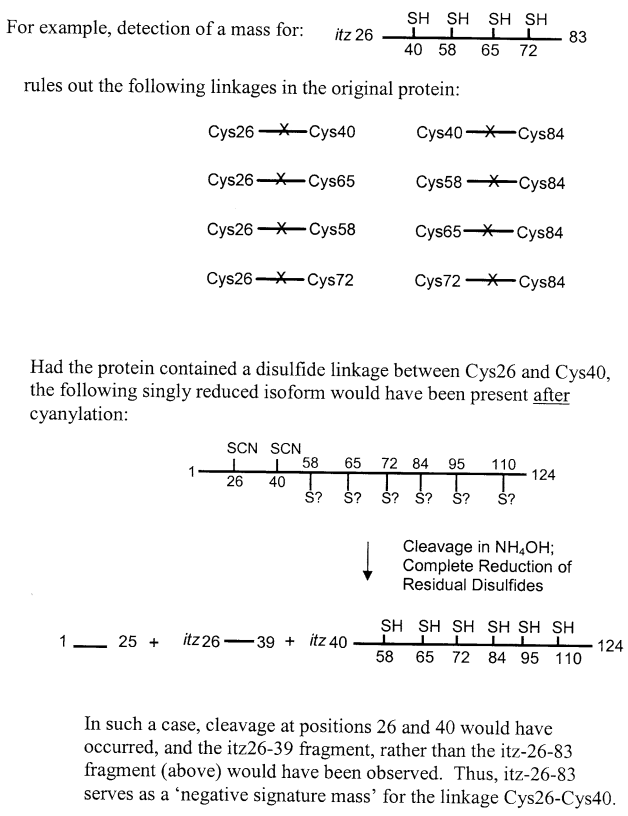


Figure 2. Graphical depiction of 3 of the 105 possible perfect matchings (disulfide structures) for RNase A. Each matching partitions the 8 nodes (cysteines) into 4 edges (disulfide linkages).

The NSMA first eliminates candidate disulfide linkages (edges) based on identified negative signature masses. A visual summary of the work done by the NSMA in the first step is provided in Table 2, which shows a list of the 28 potential linkages for RNase A in column 2. If a potential linkage has one X (or more) in its row, the NSMA eliminates the linkage (edge) based on recognition of the corresponding negative signature mass. For example, detection of a peak at *m/z* 7083.9 corresponds to the CN-induced cleavage fragment consisting of the first 64 residues of RNase A; this fragment contains Cys26 as a free sulfhydryl, which means that this residue could not have been connected to Cys65 in the original molecule (otherwise, it would have become available for cyanylation simultaneously with Cys65). Linkages without any Xs in their row remain as candidate disulfide linkages (edges). As indicated in Table 2, 21 of 28 potential disulfide linkages (edges) were eliminated based on the detection of peaks for nine negative signature masses.

After the disulfide linkage (edge) elimination step, we need to deduce valid disulfide structures (perfect matchings) from the remaining disulfide linkages (edges). In graph theory, this corresponds to the problem of finding all possible perfect matchings from the remaining edges. Finding a single perfect matching can be done in time polynomial in the number of nodes (cysteines). However, if we have not eliminated enough



Scheme 4. The logic of the NSMA.

edges (disulfide linkages) using the NSMA, the number of perfect matchings (disulfide structures) can be exponential in the number of nodes (cysteines) which means identifying and listing all possible disulfide structures will take time exponential in the number of cysteines. An exhaustive matching algorithm was implemented to enumerate all possible combinations of n (number of disulfides in the protein) disulfide linkages from the remaining ones that survived the elimination step by negative signature masses and check the validity of these combinations (disulfide structures). Obviously, not all combinations are valid disulfide structures because one cysteine may appear to form disulfide linkages with more than one cysteine in some of these combinations. These invalid combinations are ruled out, and possible disulfide structure(s) are found from among the remaining valid combinations. When enough negative signature masses are detected, only one disulfide structure, the correct one, will be deduced. Otherwise, several possible disulfide structures will be found.

To illustrate the process of eliminating invalid combinations of disulfide linkages, consider the 28 possible linkages available in a protein having the sequence for RNase A as listed in Table 2. After applying the NSMA, seven disulfide linkages remain (those in the shaded rows in Table 2), leaving 35 possible combinations of four linkages. Most of those combinations are invalid disulfide structures. For example, disulfide linkages 5, 6, 7, and 12 in Table 2 cannot form a valid disulfide structure because Cys26 would be linked to more than one cysteine. Examination of all 35 possible combinations by the exhaustive matching algorithm reveals that only one (5, 12, 18, 19), is valid; thus, it is the correct disulfide structure of RNase A.

At this stage of development of the NSMA, a standard maximum matching algorithm [11], more efficient than the exhaustive matching algorithm, has been implemented to find the valid disulfide structure for the cases in which enough negative signature masses are detected. We are in the process of extending the code to efficiently determine all valid disulfide structures for the cases in which not enough negative signature masses are detected.

The NSMA does not require the detection of mass spectral peaks for all possible negative signature masses to eliminate enough disulfide linkages to deduce the correct disulfide structure. This aspect can be seen in Table 2, where most rows have more than two Xs, indicating that the data set is highly redundant. This degenerate feature provides considerable flexibility in the partial reduction chemistry and subsequent analysis by mass spectrometry. For example, if only three, instead of nine, negative signature masses (namely, those of 6547.3 Da calculated for itz-26–83, 5767.4 Da for itz-58–109, and 6063.6 Da for itz-40–94) had been detected due to less successful chemistry or mass spectral analysis, the same 21 disulfide linkages would still have been eliminated. Thus, the test of the NSMA described here with data from the analysis of RNase A demonstrates the power and great analytical advantage of this approach to data interpretation. Namely, these results show that from batch processing the mass spectral data from all singly reduced and cyanylated isoforms, the disulfide structure of RNase A can be distinguished from the other 104 isomeric forms by detecting mass spectral peaks for only three negative signature masses: 6547.3 Da calculated for (itz-26–83), 5767.4 Da for (itz-58–109), and 6063.6 Da for (itz-40–94).

In spite of the degenerate nature of some data used by the NSMA, it is conceivable that signal suppression in MALDI could preclude detection of critically important cleavage products during batch analysis. In such cases, the cleavage reaction mixture could be simplified by HPLC to reduce the complexity of the mixture so that signal suppression would be less likely. Alternatively, analysis of the cleavage reaction mixture by electrospray may be useful; furthermore, limited LC/MS may be considered, not to achieve complete separation of components, but to reduce the complexity of the cleavage reaction mixture somewhat.

Manual interpretation of mass spectral data from analysis of cyanylation-induced cleavage products, including the direct ruling-in of disulfide linkages, is simple provided that the corresponding partially reduced and cyanylated isoform is physically separated from other partially reduced and cyanylated isoforms, and its degree of reduction is known. These conditions

Table 2. List of possible disulfide linkages for a protein having the sequence of RNase A and the associated *m/z* value for the corresponding negative signature mass

		Identified negative signature masses (Da)								
S-S No.	Potential disulfide linkages	4526.0 Itz-84-124	5906.5 Itz-72-124	7083.9 I-64	6547.3 Itz-26-83	3302.7 Itz-95-124	4413.9 I-39	6063.6 Itz-40-94	5767.4 Itz-58-109	6353.1 1-57
1	Cys26–Cys40				X		X			
2	Cys26–Cys58				X					X
3	Cys26–Cys65			X	X					
4	Cys26–Cys72				X					
5	**Cys26–Cys84**									
6	**Cys26–Cys95**									
7	**Cys26–Cys110**									
8	Cys40–Cys58							X		X
9	Cys40–Cys65			X				X		
10	Cys40–Cys72							X		
11	Cys40–Cys84				X			X		
12	**Cys40–Cys95**									
13	**Cys40–Cys110**									
14	Cys58–Cys65			X					X	
15	Cys58–Cys72								X	
16	Cys58–Cys84				X				X	
17	Cys58–Cys95							X	X	
18	**Cys58–Cys110**									
19	**Cys65–Cys72**									
20	Cys65–Cys84				X					
21	Cys65–Cys95							X		
22	Cys65–Cys110								X	
23	Cys72–Cys84		X		X					
24	Cys72–Cys95		X					X		
25	Cys72–Cys95		X						X	
26	Cys84–Cys95	X						X		
27	Cys84–Cys110	X							X	
28	Cys95–Cys110					X			X	

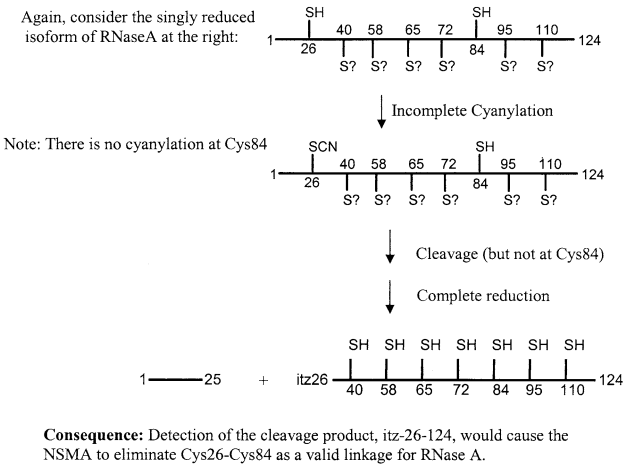
become difficult to meet as both the number of cysteines and protein size increase, and are nearly impossible to achieve for proteins with more than 10 cysteine residues because of the difficulty in separating all possible partially reduced isoforms. The NSMA eliminates the need for all of these conditions. The most useful feature of the NSMA approach to solving disulfide structures, besides automation, is its ability to select disulfide linkage and structural candidates based on mass spectral data obtained without physical separation nor determination of the state of reduction of partially reduced and cyanylated protein isoforms.

As with manual data analysis techniques, the success of an experiment that utilizes the NSMA for data interpretation depends on the prevailing chemistry and success in producing multiple partially reduced protein isoforms. In general, the NSMA is most effective and productive when dealing with singly reduced species because the peptides produced (CN-induced cleavage products) from these species are likely to have the greatest number of internal free cysteine residues, allowing the NSMA to eliminate the highest number of possible disulfide linkages. Because of the structural diversity among cystinyl proteins, several adjustments in conditions (pH, concentration, etc.) for partial reduction may be necessary before an optimal distribution of partially reduced protein isoforms is obtained.

A prerequisite for using the NSMA approach to solving disulfide structures is that the nascent free sulfhydryl groups formed during partial reduction are completely cyanylated prior to cleavage, full reduction,

Table 3. Mass spectral data from mass mapping the CN-induced cleavage products of singly reduced isoforms of RNase A [4]

m/z, Experimental	Calculated average mass, Da	CN-induced cleavage product
2705.3	2706.8	1-25
4527.4	4526.0	itz-84-124
6548.5	6547.3	itz-26-83
7083.8	7083.9	1-64
Not detected	789.8	itz-65-71
5907.7	5906.5	itz-72-124
6351.1	6353.1	1-57
5766.8	5767.4	itz-58-109
1659.8	1659.8	itz-110-124
4414.4	4413.9	1-39
6061.2	6063.6	itz-40-94
3303.7	3302.7	itz-95-124



Scheme 5. The problem associated with incomplete Cyanylation.

and mass spectral analysis. Otherwise, incorrect elimination of a candidate disulfide linkage may result due to an incorrectly identified negative signature mass. As shown in Scheme **5**, Cys26–Cys84 is a valid linkage, but if Cys84 had not been cyanylated, there would have been no cleavage there, and the detection of itz-26–124 as a negative signature mass would have wrongly eliminated Cys26–Cys84 as a valid linkage. To guard against the consequences of incomplete cyanylation, use of the NSMA approach will require additional chemical treatment to remove any possible incompletely cyanylated protein molecules.

Finally, as with any experiment based on mass mapping, one should refrain from entering the mass value corresponding to every observed peak into the NSMA. Because they are readily identifiable, masses resulting from β-elimination [4, 6, 12] of the thiocyanate group or other incomplete cleavage [6] will not rule out correct linkages. As is the case with manual interpretation, β-elimination products cannot be directly used in the NSMA to help deduce the valid disulfide structure, but they can be used to check the internal consistency of the data processing. Other masses, however, such as those of sodium adducts or those resulting from prompt water or ammonia losses or methionine oxidation might be isomass with sulfhydryl-containing cleavage products and thus, entering such data might unwittingly rule out correct linkages. A safe rule of thumb is to enter only the mass values corresponding to major peaks in the mass spectrum, and to ignore, at least initially, the small peaks that surround major peaks. Some information may be lost by this procedure, but wrong conclusions might be avoided by ensuring that correct linkages are not eliminated.

Conclusions

The NSMA provides an automated, powerful way to solve protein disulfide structures based on partial reduction, cyanylation-induced cleavage, and mass mapping—a technique proven capable of handling the most complicated of disulfide structures. The NSMA does not require physical separation nor determination of the state of reduction of partially reduced and cyanylated protein isoforms, making it especially useful in solving the disulfide structures of proteins containing many disulfide bonds. In most cases, highly redundant cross correlations exist between negative signature mass and the candidate disulfide linkages so that irrelevant linkage possibilities can be eliminated by detecting mass spectral peaks for only a fraction of the possible negative signature masses. Proper use of the NSMA requires complete cyanylation of nascent sulfhydryls, meaning that no partially reduced and cyanylated protein molecules undergoing analysis may contain free cysteine residues prior to cleavage and complete reduction of remaining disulfide bonds; chemical methodology to ensure this goal is under development.

Acknowledgments

This work was supported by a grant from NIH R01-GMS60576 (to JTW).

References

1. Smith, D. L.; Zhou, Z. R. Strategies for Locating Disulfide Bonds in Proteins. *Methods Enzymol.* **1990,** *193,* 374–389.
2. Zhou, Z. R.; Smith, D. L. Location of Disulfide Bonds in Antithrombin III. *Biomed. Environ. Mass Spectrom.* **1990,** *19,* 782–786.
3. Merewether, L. A.; Le, J.; Jones, M. D.; Lee, R.; Shimamoto, G.; Lu, H. S. Development of Disulfide Peptide Mapping and Determination of Disulfide Structure of Recombinant Human Osteoprotegerin Chimera Produced in Escherichia coli. *Arch. Biochem. Biophys.* **2000,** *375,* 101–110.
4. Wu, J.; Watson, J. T. A Novel Methodology for Assignment of Disulfide Bond Pairings in Proteins. *Protein Sci.* **1997,** *6,* 391–398.
5. Wu, J.; Watson, J. T. Assignment of Disulfide Bonds in Proteins by Chemical Cleavage and Peptide Mapping by Mass Spectrometry. *Methods Mol. Biol.* **2002,** *194,* 1–22.
6. Qi, J.; Wu, J.; Somkuti, G. A.; Watson, J. T. Determination of the Disulfide Structure of Sillucin, a Highly Knotted, Cysteine-Rich Peptide, by Cyanylation/Cleavage Mass Mapping. *Biochemistry* **2001,** *40,* 4531–4538.
7. Jacobson, G. R.; Schaffer, M. H.; Stark, G. R.; Vanaman, T. C. Specific Chemical Cleavage in High Yield at the Amino Peptide Bonds of Cysteine and Cystine Residues. *J. Biol. Chem.* **1973,** *248,* 6583–6591.
8. Gray, W. R. Disulfide Structures of Highly Bridged Peptides: A New Strategy for Analysis. *Protein Sci.* **1993,** *2,* 1732–1748.
9. Harary, F. *Graph Theory.* Addison-Wesley: New York, 1969 pp 8–23, 96.
10. Wu, J.; Gage, D. A.; Watson, J. T. A Strategy to Locate Cysteine Residues in Proteins by Specific Chemical Cleavage Followed by Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry. *Anal. Biochem.* **1996,** *235,* 161–174.
11. Gabow, H. N. An Efficient Implementation of Edmond's Algorithm for Maximum Matching on Graphs. *J. Assoc. Comput. Machin.* **1976,** *23,* 221–234.
12. Wu, J.; Watson, J. T. Optimization of the Cleavage Reaction for Cyanylated Cysteinyl Proteins for Efficient and Simplified Mass Mapping. *Anal. Biochem.* **1998,** *258,* 268–276.