Available online at www.sciencedirect.com
SCIENCE DIRECT®
Analytical Biochemistry 329 (2004) 91–103
ANALYTICAL BIOCHEMISTRY
www.elsevier.com/locate/yabio

Algorithm-assisted elucidation of disulfide structure: application of the negative signature mass algorithm to mass-mapping the disulfide structure of the 12-cysteine transforming growth factor β type II receptor extracellular domain

Chad R. Borges,[a,*,1] Jianfeng Qi,[b,2] Wei Wu,[b] Eric Torng,[c] Andrew P. Hinck,[d] and Jack Throck Watson[a,b]

[a] *Department of Biochemistry, Michigan State University, East Lansing, MI 48824, USA*
[b] *Department of Chemistry, Michigan State University, East Lansing, MI 48824, USA*
[c] *Department of Computer Science and Engineering, Michigan State University, East Lansing, MI 48824, USA*
[d] *Department of Biochemistry, University of Texas Health Science Center at San Antonio, San Antonio, TX 78229-3900, USA*

Received 1 December 2003
Available online 30 April 2004

Abstract

The power of an algorithm-driven method for interpreting disulfide mass-mapping data is demonstrated in the context of determining the disulfide structure of the extracellular domain of the transforming growth factor β type II receptor, a 14-kDa cystinyl protein containing 12 cysteines in the form of six disulfide bonds. The disulfide mass-mapping methodology is based on partial reduction and cyanylation-induced cleavage of the cystinyl protein. Because the multiplicity of possible disulfide structures that must be considered grows rapidly with the number of cysteines, as does the difficulty in physically isolating each of the partially reduced and cyanylated isoforms of the analyte, manual data interpretation for disulfide mapping a cystinyl protein containing more than eight cysteines becomes unmanageable. Recently, we introduced the concept of a "negative signature mass algorithm" (NSMA) to determine the disulfide structure of a cystinyl protein by processing an input of its amino acid sequence and mass spectral data from analysis of its associated cyanylation-induced cleavage products. Here, we present experimental results to validate the NSMA concept. A key advantage of the NSMA, in addition to convenience and automation, is its capacity to interpret mass spectra from mixtures of cyanylation-induced cleavage fragments without separating the partially reduced isoforms of the cystinyl protein and without knowledge of the extent of partial reduction.

Keywords: Disulfide bonds; Partial reduction; Cyanylation; Data processing; Polypeptide cleavage; Mass spectra; Negative signature mass; Transforming growth factor β

A rapid and reliable method for identification of disulfide linkages will become increasingly important as disulfide proteomics develops [1]. For example, a convoluted disulfide bridge pattern usually introduces a set of constraints that determines the tertiary (and sometimes quaternary) structure of the protein. Furthermore, evolving disulfide patterns are often used as a means to track the protein folding pathways.

Determining the disulfide structure of a cystinyl protein by mass-mapping methodology can present a substantial analytical challenge, especially when the analyte contains a large number of cysteines, the cysteines are close to or next to one another in sequence, or both. In recent years, we have developed a cysteine-specific, cyanylation-based chemical approach [2] to controlled

* Corresponding author. Fax: 1-801-581-5034.
E-mail address: c.borges@utah.edu (C.R. Borges).
[1] Present address: Center for Human Toxicology, University of Utah, 20 South 2030 East, Rm. 490, Salt Lake City, UT 84112, USA.
[2] Present address: EPIX Medical, Inc., 71 Rogers Street, Cambridge MA 02142, USA.

0003-2697/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.ab.2004.01.037

degradation of a cystinyl protein for disulfide mass-mapping that overcomes limitations of the conventional proteolytic approach [3,4] related to disulfide scrambling or an insufficient number of proteolytic cleavage sites. Initial applications of the cyanylation-based mass-mapping methodology to a variety of cystinyl proteins containing up to four cystines [2], including those with adjacent cysteines [2,5], followed an analytical strategy in which the partially reduced and cyanylated isoforms of the analyte were physically isolated, an approach that had limitations of its own. Herein, we demonstrate that use of a specialized data processing algorithm obviates the need to isolate the individual partially reduced cyanylated isoforms, thereby extending the cyanylation-based disulfide mapping methodology to the analysis of proteins containing a large number of cysteines.

The number of disulfide structures that must be considered for a given cystinyl protein grows rapidly with the number of cysteines: there are just 105 possible disulfide structures for a protein with four disulfide linkages, however, there are 10,395 isomeric disulfide structures for a protein with six disulfide linkages.[3] Hence, manual processing of the mass-mapping data for a protein containing more than eight cysteines is impractical. For this reason, we introduced the concept of a "negative signature mass algorithm" (NSMA)[4] to deduce the disulfide structure of a cystinyl protein given an input of its amino acid sequence and mass spectral data from the partial reduction and cyanylation-induced cleavage mass-mapping experiment [6]. The concept of negative signature mass rules out the existence of theoretically possible linkages based on detecting certain structurally diagnostic cyanylation-induced cleavage fragments of the cystinyl protein. The algorithm is available at http://www.bch.msu.edu/facilities/massspec/disulfide.

In this article, we validate the applicability of the NSMA and describe practical guidelines for using it in processing mass spectral data obtained from analyzing complex (unseparated) mixtures of cyanylation-induced cleavage products of partially reduced, cyanylated isoforms of a 14-kDa protein containing 12 cysteines in the form of six disulfide bonds. For this didactic description, we chose the extracellular domain of the transforming growth factor β type II receptor (ecTβR2), derived through recombinant expression in *Escherichia coli* and oxidative refolding [7], because it represents a complicated disulfide structure as determined independently by X-ray diffraction methodology [8]. One of the fundamental assumptions of the NSMA algorithm is that the partially reduced isoforms are completely cyanylated. Hence, care must be taken to ensure that no undercyanylated partially reduced isoforms of the protein survive the chemical processing prior to analysis by mass spectrometry. Therefore, we also suggest and describe a chemical means for scavenging traces of incompletely cyanylated protein, which would otherwise interfere with proper functioning of the NSMA.

The NSMA greatly strengthens the cyanylation-based approach to mass mapping cystinyl proteins. Because it is no longer necessary to physically isolate the cyanylated partially reduced isoforms of the analyte, this methodology will become more attractive as a rapid and reliable means for elucidating disulfide structures.

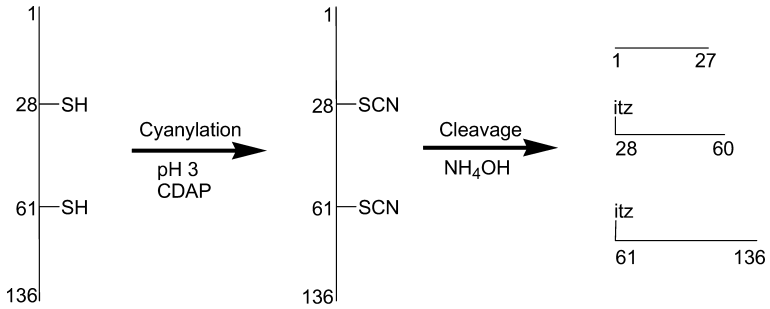
The cyanylation-based mass-mapping method is based on selective cyanylation of cysteine as illustrated in Scheme 1 for a hypothetical cysteinyl peptide having cysteines at positions 28 and 61 in its sequence. Subsequent nucleophilic attack by ammonia causes CN-induced cleavage of the peptide backbone on the N-terminal side of the modified cysteines.

The cyanylating reagent, cyanodiaminopyridinium (CDAP) tetrafluoroborate, is selective for free sulfhydryl groups and will not react with cystines. Thus, a given disulfide bond must be reduced so that its constituent cysteines become available for the obligatory cyanylation reaction. For this purpose, the technique of partial reduction [9], as shown in Scheme 2 for ecTβR2 using triscarboxyethylphosphine (TCEP), is the first step in modifying the analyte. There are, in principle, six possible singly reduced isoforms for a cystinyl protein containing six cystines. Only one of the six is shown in Scheme 2, namely that in which the disulfide linkage between Cys28 and Cys61 has been reduced. Although the conditions for partial reduction [9] can be optimized for some proteins to generate predominantly the singly reduced isoforms, it is a dynamic and structure-dependent process, and some doubly reduced isoforms and traces of the triply reduced isoforms typically will be present in the reaction mixture. For purposes of simplifying the description of the chemical treatment of a protein, only the singly reduced isoform shown in Scheme 2 will be pursued in this introductory example.

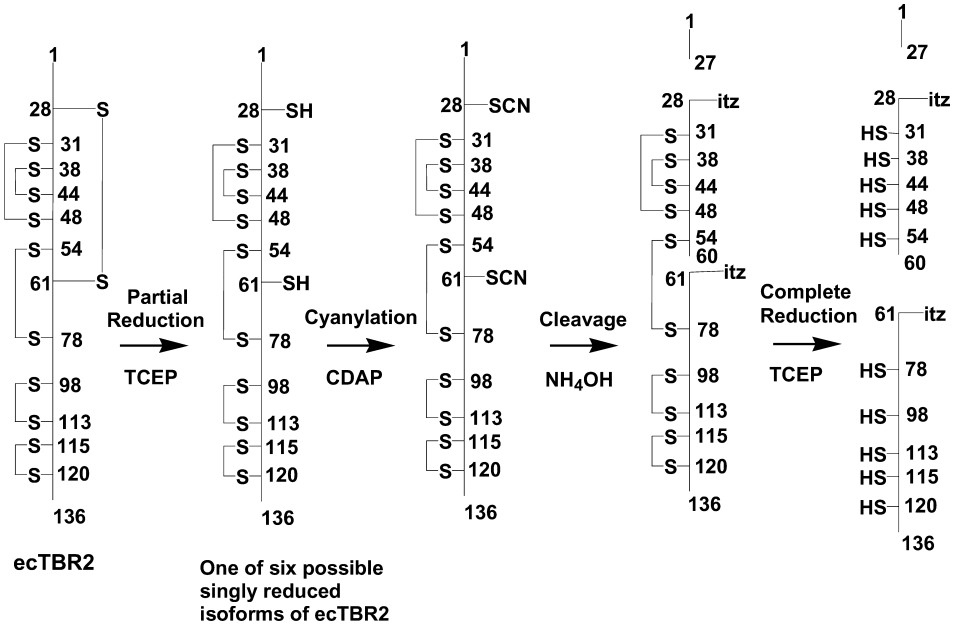
Once cyanylated and subjected to nucleophilic attack in 1 M NH_4OH, the singly reduced isoform of ecTβR2 can be cleaved at the peptide bonds between residues 27 and 28 and between residues 60 and 61, as shown in Scheme 2 (itz represents an iminothiazolidine-blocked N terminus). Depending on the disulfide structure, residual linkages may hold some of the CN-induced cleavage fragments together, despite the cleaved polypeptide backbone. This is illustrated in the middle of Scheme 2, where a residual disulfide bond between Cys54 and Cys78 holds two of the three cleavage products together. Thus, the cleavage reaction mixture represented in the middle of Scheme 2 is treated with excess reducing

[3] A disulfide linkage is defined as a disulfide bond between two specific cysteine residues. A disulfide structure is a particular combination of specific linkages within a protein.

[4] *Abbreviations used*: NSMA, negative signature mass algorithm; ecTβR2, extracellular domain of transforming growth factor β type II receptor.



Scheme 1.



Scheme 2.

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 2.

Direct manual analysis is simple if a partially reduced and cyanylated isoform, such as that represented at the left in Scheme 2, can be physically isolated from the others prior to cleavage. Ideally, analysis of a cleavage reaction mixture from a singly reduced isoform will show the presence of three specific fragments, as illustrated at the right in Scheme 2 for the selected singly reduced isoform of ecTβR2. In this case, CN-induced cleavage between residues 27 and 28 occurred because a free cysteine at position 28 was available for cyanylation; similarly, CN-induced cleavage between residues 60 and 61 occurred because a free cysteine at 61 was available for cyanylation. Because the analysis is carried out on a singly reduced isoform, i.e., a species containing two nascent free cysteines, the cysteines at positions 28 and 61 must have been produced by reduction of a disulfide bond between Cys28 and Cys61. In this way, analysis of the cleavage reaction mixture by mass spectrometry unambiguously identifies the singly reduced isoform of ecTβR2 represented in Scheme 2 and thus the connectivity of one of the six disulfide bonds in the original cystinyl protein, namely, that between Cys28 and Cys61.

The direct approach as demonstrated in this introductory example and as originally conceived with this methodology [2] however, is impractical for analysis of a protein containing six disulfide bonds because of the likely impossible task of physically separating all of the partially reduced and cyanylated isoforms completely. Isoforms must be physically separated from one another in this direct approach because otherwise it would be unknown, for example, whether cleavage products observed and identified by MALDI-MS arose from one or more doubly (or more) reduced isoforms, making direct assignment of disulfide bonds based on identified CN-induced cleavage fragments impossible.[5] The NSMA obviates the need to isolate the partially reduced (and cyanylated) isoforms and greatly simplifies data

[5] Isolation of a doubly reduced isoform allows inferences to be made about the presence of two disulfide bonds between four specific cysteine residues (see, for example, [10]), but if multiple isoforms are present in a cleavage mixture, even such inference becomes impossible.

interpretation as demonstrated in the following analysis of a recombinant form of ecTβR2.

The concept of the NSMA is based on detection of certain cyanylation-induced cleavage fragments that serve as negative signature masses (NSMs) to rule out certain disulfide linkages from a list of all theoretically possible linkages for the protein under consideration [6]. By general definition, a NSM corresponds to a CN-induced cleavage fragment that contains internal free cysteines, i.e., cysteines that lie between the N and C termini of the given CN-induced cleavage fragment. For example, consider itz-Cys28-Val60, a cleavage fragment at the right in Scheme 2 that contains five internal cysteines, namely, those at positions 31, 38, 44, 48, and 54. Detection of such a cleavage fragment rules out further consideration of a disulfide linkage between either of the cysteines defining the termini of the cleavage fragment, either Cys28 or Cys61 in this case, and one of its internal cysteines. Thus, detection of itz-Cys28-Val60 means that no disulfide bond in the original protein could have existed between Cys28 and Cys31 or Cys38 or Cys44 or Cys48 or Cys54, nor between Cys61 and Cys31 or Cys38 or Cys44 or Cys48 or Cys54. If such a disulfide bond had existed in the original protein, then a different CN-induced cleavage fragment would have resulted. For example, in consideration of the cysteine positions 31, 38, 44, 48, and 54 listed above, if a disulfide bond had existed between Cys28 and Cys44, its participation in partial reduction and cyanylation would have led to CN-induced cleavage at Cys28 and at Cys44, thereby producing the fragment itz-Cys28-43; under this hypothetical situation, the CN-induced cleavage fragment itz-Cys28-Val60 would not have been produced. In this way, detection of itz-Cys28-Val60 serves as a NSM for Cys28-Cys44 and 9 other (described above) of the 66 theoretically possible disulfide linkages (Table 4) for ecTβR2, a cystinyl protein containing 12 cysteines as six disulfide bonds.

In the above synopsis of our direct interpretation strategy [2], the cleavage fragment at the right of Scheme 2, itz-61-136, was used to help rule in the disulfide linkage between Cys28 and Cys61. Conversely, itz-61-136 is also a NSM by which the NSMA rules out the existence of disulfide linkages Cys61-Cys78, Cys61-Cys98, etc. These two examples illustrate the basis by which the NSMA eliminates theoretically possible linkages from consideration for the cystinyl protein under consideration, thereby facilitating data interpretation.

Methods

Materials

TCEP hydrochloride, CDAP tetrafluoroborate, 4-dimethylaminophenylazophenyl-4'-maleimide (DABMI), and *p*-chloromercuribenzoic acid sodium salt (pCMB) were purchased from Sigma–Aldrich (St. Louis, MO). EAH Sepharose 4B was obtained from Amersham Biosciences (Uppsala, Sweden). pCMB was covalently linked to EAH Sepharose 4B to create an organomercury (OrgHg)-linked solid phase by methods described elsewhere [11]. Micro Bio-Spin 6 chromatography columns, with a size exclusion limit of 6000 Da were purchased from Bio-Rad (Hercules, CA). The ecTβR2 samples, containing an N19A mutation, were prepared as described earlier [7]; this 13,800-Da protein contains 12 cysteine residues involved in six disulfide bonds. The recombinant ecTβR2 used here is a truncated form in that it does not contain the first 13 residues of the native protein and, thus, is not strictly consistent with the didactic illustration in Scheme 2, but is consistent with the data in Tables 1–5 in which the N-terminus corresponds to (Met-mutated) residue 14 of the native protein. Mass spectrometry standards consisting of bovine bradykinin, bovine pancreatic insulin, and horse skeletal myoglobin were purchased from Sigma–Aldrich All other materials were of the highest purity commercially available.

Summary of experimental steps

The complexity of the partially reduced and cyanylated protein samples was reduced by running the sample through RP-HPLC; fractions corresponding to individual peaks were collected and analyzed by MALDI-MS to assess the state of reduction. (These analyses were performed here to illustrate the complexity and status of the sample at various stages of sample processing; neither preliminary analysis by HPLC nor determination of the state of analyte reduction by analysis by MALDI is a prerequisite for subsequent use of the NSMA for processing the mass spectral data from analysis of CN-induced cleavage products.[6]) The collected fractions were pooled, treated with an OrgHg solid-phase column, desalted by HPLC, cleaved with aqueous ammonia, and then fully reduced as described below. The final resulting mixture of CN-induced cleavage products was analyzed by MALDI-MS to provide input data for the NSMA algorithm to solve the disulfide structure.

Partial reduction and cyanylation

ecTβR2 (3.1 nmol) was solubilized in 10 μL of 0.1 M citrate buffer, pH 3.0, containing 6 M guanidine HCl. Par-

[6] The fundamental logic required for the NSMA does not require separation of partially reduced and cyanylated isoforms by HPLC or knowledge of the state of reduction of the isoforms. The NSMA is quite capable of solving complete disulfide structures, given all the appropriate input data. However, in cases where the input data set is insufficient to solve the complete disulfide structure, it may be helpful to supplement the output provided by the NSMA by obtaining direct disulfide connectivity information from one or two isolated isoforms.

tial reduction was accomplished by adding four equivalents of TCEP per cystine equivalent (i.e., 75 nmol TCEP from a 0.1 M stock solution for a conc. of 7.0 mM TCEP) and incubating at 37 °C for 15 min. Cyanylation is selective for cysteine, i.e., free sulfhydryls [12,13], and is accomplished by mixing in a 20-fold molar excess of CDAP over cysteine content (750 nmol from a 0.2 M stock solution for a final concentration of 54 mM CDAP) and incubating at room temperature for 15 min [10].

Organomercurial affinity chromatography

To ensure that the sample contained no incompletely cyanylated species, the pooled HPLC fractions containing the partially reduced/cyanylated protein isoforms were dried, redissolved in 10 μL of 6 M guanidine HCl ($pH \leqslant 5$), and passed through an OrgHg column. The solution was applied to a ~100-μL OrgHg column (a pasteur pipette plugged with glass wool) that had been equilibrated with >10 column volumes of 6 M guanidine HCl. Capacity testing had demonstrated that this volume of OrgHg resin could bind >90 nmol of free sulfhydryl equivalents, an amount far greater than the trace amounts of free sulfhydryl groups that the sample could possibly contain. Protein molecules lacking free sulfhydryl groups were eluted directly through the column with 500 μL of 6 M guanidine HCl, while protein molecules containing free sulfhydryl groups remained bound to the column. The flow-through sample was then desalted by RP-HPLC and dried.

Cleavage, complete reduction of the residual disulfide bonds, and final sample preparation

Each sample was dissolved in 100 μL of 4 M guanidine HCl containing 1 M NH_4OH. Cleavage was allowed to take place for 1 h at room temperature, after which time excess ammonia was removed under vacuum. Ammonium hydroxide promotes CN-induced cleavage of the peptide backbone on the N-terminal side of cyanylated cysteine residues, forming a carbamido C-terminal peptide[7] and an iminothiazolidine-blocked N-terminal peptide. β-elimination of the thiocyanate group can be a significant side reaction induced by 1 M NH_4OH [2,12]. On rare occasions, cleavage may be incomplete at the site of one of the cyanylated cysteines [14], leaving a trace of cyanylated, but incompletely cleaved peptide. These apparently sequence-dependent phenomena are not a serious problem because the algorithm recognizes such fragments via their calculated mass and can be programmed to use such information to corroborate detection of other fragments, thereby providing some analytical advantage.

[7] Occasionally, if a hydroxyl ion, which also exists in 1 M ammonia, acts as the nucleophile, a carboxylic acid C terminus will form [13,14].

Following cleavage, complete reduction of the remaining disulfide bonds was accomplished by bringing the sample to 0.05 M in TCEP and incubating for 30 min at 37 °C. Peptides (CN-induced cleavage products) were separated and collected in 12 fractions by microbore RP-HPLC. (This final HPLC step is not always necessary and may be omitted, especially in the case of proteins containing only two to four disulfide bonds, as long as the salt content of the final sample is dilute enough to provide good response by MALDI-MS [2,10].) Fractions were dried and reconstituted in 2 μL of 50/50 acetonitrile/0.1% trifluoroacetic acid saturated with α-cyano-4-hydroxycinnamic acid and then spotted onto a MALDI target using the thin-layer technique of Cadene and Chait [15].

Analysis by MALDI-MS

MALDI mass spectra were acquired on a Voyager DE-STR time-of-flight mass spectrometer (Perkin–Elmer Biosystems, Framingham, MA) equipped with a 337-nm nitrogen laser. Accelerating voltage was set to 20,000 V with grid voltage at 95%, guide wire at 0.05% and extraction delay time at 150 ns. Time-of-flight to mass conversion was achieved by external calibration using mass spectral peaks detected during analysis of standards of bradykinin (monoisotopic calculated mass for $MH^+ = 1060.57$ Da; average mass for $MH^+ = 1061.22$ Da), bovine pancreatic insulin (average calculated mass for $MH^+ = 5734.56$ Da; average calculated mass to charge ratio for $[M+2H]^{2+} = 2867.78$), and horse skeletal myoglobin (average mass for $MH^+ = 16{,}952$ Da).

Partial reduction and intentional under cyanylation

In separate experiments designed to test the effectiveness of sulfhydryl scavenging by an OrgHg solid phase, 5-nmol samples of partially reduced RNase A and ecTβR2 were intentionally undercyanylated desalted by HPLC, and split 60:40, the latter serving as a control. The 60% aliquot from each sample underwent OrgHg-column treatment followed by DABMI labeling (described below), while the 40% aliquot from each sample was directly labeled with DABMI. Following cleavage and complete reduction, the samples were analyzed by HPLC with detection by absorbance at 317 nm to detect DABMI-derivatized molecules.

Labeling of SH groups with DABMI, a hydrophobic, chromophoric maleimide, to facilitate detection of SH-containing species

Note: This step is not routinely performed and was done here only as part of control experiments to verify the effectiveness of the OrgHg column in scavenging free sulfhydryl-containing proteins. Protein samples were treated with 0.3 mM DABMI in 33% (v/v) acetonitrile in

6 M guanidine HCl / 0.1 M citrate buffer at pH 5.0 (70 μL total volume) [16]. The reaction proceeded for 1 h in the dark at room temperature. Excess DABMI was removed by size exclusion chromatography in a Micro Bio-Spin 6 chromatography column that had been equilibrated with 6 M guanidine HCl as the mobile phase; protein molecules were eluted immediately from the column by the mobile phase. Note: During CN-induced cleavage in aqueous ammonia, the DABMI moiety undergoes a base-driven ring opening hydrolysis reaction [17] to produce the chemically modified species called hydDABMI; during and after this process, the DABMI moiety remains covalently attached to the peptide [16].

Results and discussion

Analysis of partially reduced and cyanylated isoforms of ecTβR2

In using ecTβR2, a protein containing 12 cysteines in the form of 6 cystines, as a model compound to demonstrate the effectiveness of the NSMA in interpreting data for disulfide mass mapping, the first step involved partial reduction [2,9]. (If the analyte contains free cysteines, it would have been appropriate to first alkylate these residues with an agent such as *N*-ethylmaleimide that is both reactive and specific at relatively low pH. Because this modification will cause a unique mass shift of the cysteine residue, blocking it from cyanylation and from subsequent cleavage, the alkylated cysteine residue can be designated a "user-defined" residue for recognition by the algorithm in its protocol for determining disulfide structure.[8]) Following treatment of the partially reduced mixture with a cyanylating reagent, CDAP [2], the sample was chromatographed and collected in fractions labeled A–L (see Fig. 1). As is evident, the HPLC fractions were not well separated. Furthermore, most of these incompletely separated fractions contained two or more partially reduced and cyanylated ecTβR2 isoforms, as determined by MALDI mass spectral data based on the mass shift caused by cyanylation (25 Da per modified cysteine). Such difficulty in isolating the partially reduced and cyanylated isoforms of a cystinyl protein makes it difficult to use our original strategy of direct analysis [2] to deduce the disulfide structure. However, because use of the NSMA [6] does *not* require separation of individual partially reduced, cyanylated isoforms (*nor* knowledge of their state of reduction; see footnote 8), the HPLC fractions were pooled (having been collected only to illustrate the complexity of the mixture), filtered

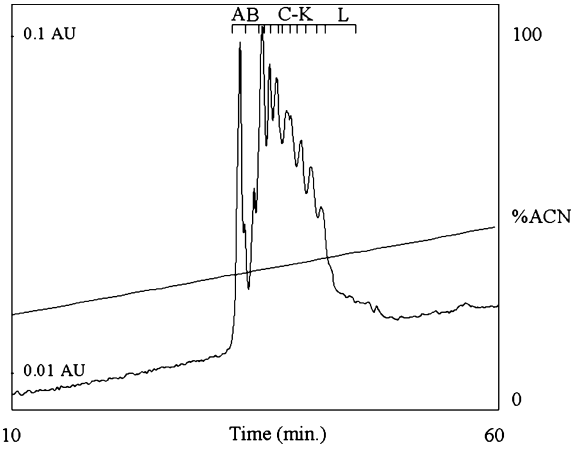


Fig. 1. HPLC chromatogram produced by absorbance at 214 nm of the partially reduced and cyanylated isoforms of ecTβR2 (precleavage mixture), il lustrating the complex unresolved mixture that can result from partial reduction of a protein containing six disulfide bonds. Fractions corresponding to individual peaks were collected (as labeled A–L), analyzed by MALDI-MS to determine their reduction state (for didactic purposes), and then pooled for further processing.

through an OrgHg column, cleaved in aqueous ammonia, and then completely reduced [2]. The resulting complex mixture of CN-induced cleavage fragments was partially separated by microbore HPLC (data not shown) and collected into 12 fractions. (By coincidence, the number of fractions collected here is the same as the number of chromatographic peaks shown in Fig. 1; the chromatogram of the mixture of CN-induced cleavage products (not shown) was entirely different in appearance from that in Fig. 1.) This partial separation of the

Table 1
Ion assignments for mass spectral peaks labeled in Fig. 2

Observed *m/z*	Peak assignment	Calculated *m/z*
851.2[a]	itz-CDVRFST-NH_2 (itz-31-37)	851.4
1493.4[a]	itz-CDVRFSTC_{β}DNQKS-NH_2 (itz-31-β@38-43)[b]	1493.7
1551.5[a]	itz-CDVRFSTC_{CN}DNQKS-NH_2 (itz-31-CN@38-43)[c]	1551.7
1963.1[d]	itz-CDVRFSTCDNQKSCMSN-NH_2 (itz-31-47)[e]	1963.2
1979.1[d]	itz-CDVRFSTCDNQKSCM_{ox}SN-NH_2 (itz-31-47)[f]	1979.2

Recall from Methods that a cyanylation-induced cleavage product may have either an amino or a carboxyl C terminus. With the exception of the C terminus of the protein (which always has a carboxyl C terminus), the status of the C terminus of a given CN-induced cleavage fragment was assigned in Tables 1–3 based on which calculated mass agreed most closely with an observed *m/z* value.

[a] Monoisotopic mass.

[b] The terminology "β@38" refers to a β-elimination event at Cys38 [2,12].

[c] The terminology "CN@38" refers to an incomplete cleavage event at Cys38 [14].

[d] Average mass.

[e] NSM.

[f] The methionine residue was oxidized to the sulfoxide.

[8] If the location(s) of the free cysteine residue(s) are unknown, the analyst could first use the cyanylation and cleavage methodology upon the nonreduced native protein to determine where the free cysteine(s) are located [18].

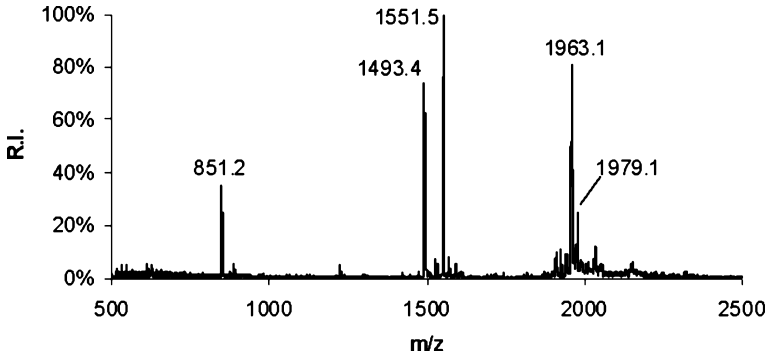


Fig. 2. MALDI mass spectrum of an HPLC fraction of CN-induced cleavage products from partially reduced and cyanylated ecTβR2. The peak at *m/z* 1963.1 represents a NSM, which corresponds to itz-31-47, a CN-induced cleavage fragment containing two internal cysteine residues. This NSM eliminates linkages between Cys31-Cys38, Cys31-Cys44, Cys38-Cys48, and Cys44-Cys48 (see Tables 1 and 3).

cleavage products was performed to reduce the sample complexity in an effort to minimize signal suppression during subsequent analysis by MALDI mass spectrometry.

A MALDI mass spectrum (labeled peaks are assigned in Table 1) from 1 of the 12 HPLC fractions of the CN-induced cleavage products described above is shown in Fig. 2. (The other 11 fractions were also analyzed by MALDI-MS.) The peak at *m/z* 1963.1 represents a NSM, which corresponds to itz-31–47, a CN-induced cleavage fragment containing two internal cysteine residues. This NSM eliminates linkages between Cys31-Cys38, Cys31-Cys44, Cys38-Cys48, and Cys44-Cys48 (see Tables 1 and 3). Of additional interest in this mass spectrum are peaks at *m/z* 1493.4 and *m/z* 1551.5; these peaks represent a β-elimination product [2,12] and an especially rare peptide containing a cyanylated cysteine residue which did not undergo cleavage [14], respectively.

Suggested criteria for selecting mass spectral data for processing by the NSMA

Sample components that would produce overt invalid NSMs (i.e., ions that should not be present based on the disulfide structure of the protein and expected processing chemistry) can be eliminated or minimized by chemical means or improved mass resolution as discussed below. It is also good practice to avoid selecting "satellite peaks," such as those likely to represent matrix or metal ion adducts or to prompt losses of ammonia or water, as these provide no useful structural information and might be coincidentally isomass with a NSM. Protection from covert sources of invalid NSMs, as would appear during the processing of a constituent having an isomeric disulfide structure, is much more tenuous.

The nature of the problem with covert NSMs suggests that the best protection may be provided by an iterative use of the NSMA, first by processing masses associated with the most intense peaks in the mass spectrum. Depending on the number of candidate disulfide structures indicated after processing the initial selection of mass spectral data, masses corresponding to less intense peaks in the mass spectrum can be included in a subsequent analysis by the NSMA, which will likely eliminate additional linkages. Progressively more bold selections of mass spectral peaks may eventually eliminate too many linkages to comprise a disulfide structure, in which case, the analyst would revert to the previous result from the NSMA and try to distinguish among the candidate structures by our original approach of limited direct analysis [2], when possible as described below. In fact, such an example is described below. Whereas such a result may seem disappointing at first, the progressively more bold procedure suggested here forces maximum utilization of the computational power of the NSMA to reduce an overwhelming number of theoretically possible disulfide structures that must be considered to a manageable few for manual interpretation; in the case described below, the NSMA reduced the number of structures that must be considered for ecTβR2 from 10,395 to 3. The cautious approach recommended here for selection of mass spectral data necessarily causes the selection of the peak intensity threshold to be somewhat arbitrary. However, this recommended conservative approach to data selection is by no means a bottle neck to efficient determination of the disulfide structure of a cystinyl protein as the algorithm processes the selected batch of mass spectral data in less than 10 s. Conservative selection of the mass spectral data for the NSMA minimizes the chances of including data from covert NSMs, which would inappropriately eliminate a valid linkage.

Conservative selection of mass spectral data for ecTβR2

A second MALDI mass spectrum (labeled peaks assigned in Table 2) from another one of the 12 HPLC fractions is shown in Fig. 3. Conservatively selected mass spectral data (i.e., from all the mass spectra, only those peaks exceeding the threshold represented by the higher

Table 2
Ion assignments for mass spectral peaks labeled in Fig. 3

Observed *m/z*	Peak assignment	Calculated *m/z*
1358.3[a]	VTDNAGAVKFPQL-NH_2 (15-27)[b,c]	1358.7
1489.3[a]	MVTDNAGAVKFPQL-NH_2 (14-27)[b]	1489.8
1841.6[d]	itz-CIMKEKKKPGETFFM-NH_2 (itz-98-112)	1842.3
1857.6[d]	itz-CIMKEKKKPGETFFMox-NH_2 (itz-98-112)[e]	1858.3
2000.8[d]	itz-CNDNIIFSEEYNTSNPD-COOH (itz-120-136)	2001.1
2566.6[d]	itz-CDVRFSTCDNQKSCMSNCSITSI-NH_2 (itz-31-53)	2567.9

[a] Monoisotopic mass.
[b] Residue 14 is the N terminus of the truncated form of ecTβR2.
[c] A portion of the recombinant ecTβR2 sample lacked the N-terminal methionine residue.
[d] Average mass.
[e] One of the methionine residues was oxidized to the sulfoxide.

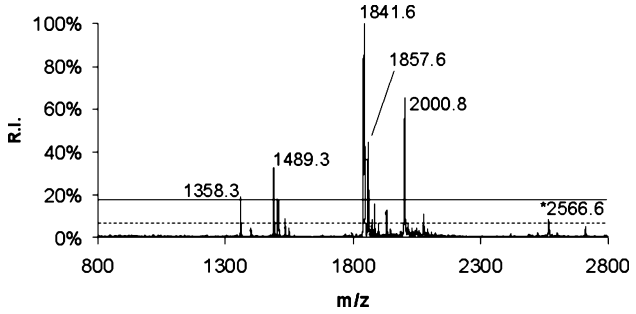


Fig. 3. MALDI mass spectrum of an HPLC fraction of CN-induced cleavage products from partially reduced and cyanylated ecTβR2. The solid horizontal line is a conservative threshold for peak intensity such that only the *m/z* values of peaks exceeding the line were entered into the NSMA. The dashed horizontal line represents a less conservative cutoff, which allows a larger number of *m/z* values to be processed by the NSMA. *In this case, the sample contained a minor component (see text for discussion relating to possible trace contamination by species of isomeric disulfide structure due to possible microheterogenity of the sample) as represented by a peak at *m/z* 2566.6 corresponding to itz-31-53, a NSM that eliminates a correct linkage, namely Cys31-Cys48.

(solid) line in Fig. 3 were selected) obtained from analysis of all 12 HPLC fractions of the CN-induced cleavage products were processed by the NSMA [6]. Each of the 12 mass spectra contain peaks representing CN-induced cleavage products, only some of which are NSMs. For example, whereas the peak at *m/z* 1963.1 in Fig. 2 represents a NSM (Table 1), the other peaks in Fig. 2 and all those above the solid line in Fig. 3 happen to represent non-NSMs (Table 2). From the cumulative mass spectra of all 12 HPLC fractions, the algorithm determines that 18 mass spectral peaks correspond to NSMs, as listed in Table 3, resulting in elimination of 37 of the 66 theoretically possible disulfide linkages from further consideration for the specific disulfide structure in this case. Among the remaining 29 disulfide linkages (Table 4), only 3 possible disulfide structures (unique arrangements of 6 of the 29 surviving theoretically possible linkages) remain as viable candidates, as listed in Table 5. The candidate structures were computed by the NSMA based on the algorithm-derived list of remaining disulfide linkages—a function automatically performed with every execution of the NSMA. All other arrangements of 6 linkages form invalid disulfide structures [6]. For example, one such candidate disulfide structure, Cys28-Cys44, Cys31-Cys48, Cys38-Cys54, Cys54-Cys78, Cys98-Cys120, Cys113-Cys115, consisting of disulfide linkages from Table 4, is obviously invalid because Cys54 is involved in more than 1 disulfide linkage.

Less conservative selection of mass spectral data for the NSMA for ecTβR2

The dashed line in Fig. 3 indicates a less conservative threshold of peak intensity for selecting mass spectral data for processing by the NSMA. The inclusion of masses for peaks exceeding this dashed threshold in all the mass spectra (including the one in Fig. 3) identifies 20 NSMs, including the 18 already listed in Table 3. One of the 2 additional NSMs, itz-CIMKIKKKPGET FFMCS-NH_2 (itz-98-114) as represented by a peak at *m/z* 2033.2 (in a mass spectrum not shown), eliminates Cys98-Cys113 and Cys113-Cys115; the other, itz-CDVRFSTCDNQKSCMSNCSITSI-NH_2 (itz-31-53), represented by a peak at *m/z* 2566.6 (seen in Fig. 3 and assigned in Table 2), eliminates Cys31-Cys38, Cys31-Cys44, Cys31-Cys48, Cys38-Cys54, Cys44-Cys54, Cys48-Cys54. Because some of the above listed linkages had already been eliminated by 1 or more of the 18 previously considered NSMs (see Table 3), the less conservatively selected mass spectral data cause a net increase of three more linkages to be eliminated, namely, Cys98-Cys113, Cys31-Cys48, and Cys38-Cys54.

Thus, the number of linkages eliminated increases from 37 to 40 as a result of processing the more boldly selected mass spectral data in Fig. 3. Of the 66 theoretically possible disulfide linkages, 26 survive. Of the 26 surviving disulfide linkages, an insufficient number of unique linkages (6 required) are available to define a disulfide structure [6]. Apparently, the more liberally (less conservatively) chosen set of mass spectral data includes at least one peak for a NSM (a covert NSM) which, when processed by the NSMA, eliminates a valid disulfide linkage. The relatively small peak marked with an asterisk in Fig. 3, representing itz-31-53, corresponds to such a NSM, and it is responsible for eliminating the valid disulfide linkage, Cys31-Cys48. Because the sample processing mixture was treated with OrgHg resin (at the stage of cyanylation, but before the complete reduction step), which scavenges free sulfhydryl-containing species as described later, it is unlikely that this NSM species derived from incomplete cyanylation during sample

Table 3
Negative signature masses (NSMs) observed for ecTβR2 and the corresponding theoretically possible disulfide linkages eliminated by the NSMA during processing of conservatively selected mass spectral data

Observed mass (Da)	Constituent residues of the corresponding cleavage fragment	Eliminated linkages
1231.0	itz-CKFCDVRFST-NH_2 (itz-28-37)	28-31, 31-38
1868.9	MVTDNAGAVKFPQLCKF-NH_2 (14-30)[a]	28-31
1897.8	itz-CMSNCSITSICEKPQEV-NH_2 (itz-44-60)	44-48, 44-54, 48-61, 54-61
1963.1	itz-CDVRFSTCDNQKSCMSN-NH_2 (itz-31-47)	31-38, 31-44, 38-48, 44-48
2677.4	MVTDNAGAVKFPQLCKFCDVRFST-NH_2 (14-37)[a]	28-38, 31-38
2711.4	itz-CSCSSDECNDNIIFSEEYNTSNPD-COOH (itz-113-136)	113-115, 113-120
2830.6	itz-CEKPQEVCVAVWRKNDENITLETV-COOH (itz-54-77)	54-61, 61-78
4294.1	itz-CVAVWRKNDENITLETVCHDPKLPYHDFILEDAASPK-NH_2 (itz-61-97)	61-78, 78-98
4311.5	itz-CHDPKLPYHDFILEDAASPKCIMKEKKKPGETFFMCS-NH_2 (itz-78-114)	78-98, 78-113, 98-115, 113-115
5712.2	itz-CSITSICEKPQEVCVAVWRKNDENITLETVCHDPKLPYHDF ILEDAA SPK-NH_2 (itz-48-97)	48-54, 48-61, 48-78, 54-98, 61-98, 78-98
6146.0	itz-CMSNCSITSICEKPQEVCVAVWRKNDENITLETVCHDPKLPYHDF ILEDAASPK-NH_2 (itz-44-97)	44-48, 44-54, 44-61, 44-78, 48-98, 54-98, 61-98, 78-98
6286.5	itz-CVAVWRKNDENITLETVCHDPKLPYHDFILEDAASPKCIMKEK KKPGETFFMCS-COOH (itz-61-114)	61-78, 61-98, 61-113, 78-115, 98-115, 113-115
6790.5	itz-CHDPKLPYHDFILEDAASPKCIMKEKKKPGETFFMCSCSSDECN DNIIFSEEYNTSNPD-COOH (itz-78-136)	78-98, 78-113, 78-115, 78-120
7702.6	itz-CSITSICEKPQEVCVAVWRKNDENITLETVCHDPKLPYHDFIL EDAASPKCIMKEKKKPGETFFMCS-NH_2 (itz-48-114)	48-54, 48-61, 48-78, 48-98, 48-113, 54-115, 61-115, 78-115, 98-115, 113-115
8141.7	itz-CMSNCSITSICEKPQEVCVAVWRKNDENITLETVCHDPKLPYH DFILEDAASPKCIMKEKKKPGETFFMCS-COOH (itz-44-114)	44-48, 44-54, 44-61, 44-78, 44-98, 44-113, 48-115, 54-115, 61-115, 78-115, 98-115, 113-115
10182.9	itz-CSITSICEKPQEVCVAVWRKNDENITLETVCHDPKLPYHDFIL EDAASPKCIMKEKKKPGETFFMCSCSSDECNDNII FSEEYNT SNPD-COOH (itz-48-136)	48-54, 48-61, 48-78, 48-98, 48-113, 48-115, 48-120
10621.0	itz-CMSNCSITSICEKPQEVCVAVWRKNDENITLETVCHDPKLPYHDFIL EDAASPKCIMKEKKKPGETFFMCSCSSDE CNDNIIFSEEYNT SNPD-COOH (itz-44-136)	44-48, 44-54, 44-61, 44-78, 44-98, 44-113, 44-115, 44-120
11447.4	MVTDNAGAVKFPQLCKFCDVRFSTCDNQKSCMSNCSITSICEKPQEVC VAVWRKNDENITLETVCHDPKLPYHDFILEDAASPKCIMKEKKKPGE TFFMCS-NH_2 (14-114)[a]	28-115, 31-115, 38-115, 44-115, 48-115, 54-115, 61-115, 78-115, 98-115, 113-115

Observed masses are within 0.05% of their calculated values.
[a] Residue 14 is the N terminus of the truncated form of ecTβR2.

Table 4
The 66 theoretically possible disulfide linkages for ecTβR2 showing which were eliminated by NSMs (Table 3) after processing of conservatively selected mass spectral data

~~28-31~~										
~~28-38~~	~~31-38~~									
28-44	~~31-44~~	***38-44***								
28-48	***31-48***	~~38-48~~	~~44-48~~							
28-54	31-54	38-54	~~44-54~~	~~48-54~~						
28-61	31-61	38-61	~~44-61~~	~~48-61~~	~~54-61~~					
28-78	31-78	38-78	~~44-78~~	~~48-78~~	***54-78***	~~61-78~~				
28-98	31-98	38-98	~~44-98~~	~~48-98~~	~~54-98~~	~~61-98~~	~~78-98~~			
28-113	31-113	38-113	~~44-113~~	~~48-113~~	54-113	~~61-113~~	~~78-113~~	***98-113***		
~~28-115~~	~~31-115~~	~~38-115~~	~~44-115~~	~~48-115~~	~~54-115~~	~~61-115~~	~~78-115~~	~~98-115~~	~~113-115~~	
28-120	31-120	38-120	~~44-120~~	~~48-120~~	54-120	61-120	~~78-120~~	98-120	~~113-120~~	***115-120***

Correct linkages (confirmed as described in the text) are in bold.

processing. Rather, it likely arose from an isomeric disulfide species of low abundance (estimated at 5% according to the relative intensity of the peak with the asterisk in Fig. 3 and an unpredictable MALDI response).

Reliability of the NSMA vs. mass measurement accuracy

Increased mass measurement accuracy will decrease the likelihood of wrongly identifying a random (background) mass as a NSM. For example, a background

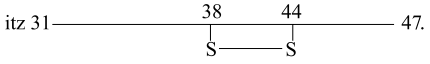
mass spectral peak at *m/z* 1000.5 would be wrongly identified as a NSM whose calculated mass is 1000.0 at mass measurement error limit of 1 Da, but at mass measurement error limit of 0.1 Da, it would be distinguished from a NSM of 1000.0 Da.

Consider the following computation, which highlights the stochastic risk of eliminating a correct linkage. If 20 random integer mass values (all between 500 Da and the mass of ecTβR2, ~13,800 Da) are entered into the NSMA, there is an approximately 8% chance that 1 of them will eliminate a correct linkage at a mass measurement error limit of 1 Da. This estimate is based on the following rationale. There are 52 NSMs (at unit mass resolution representing 52 integer mass values) between 500 and 13,800 Da (the mass of intact ecTβR2) that will eliminate correct linkages. Thus, rounding to the nearest whole number and requiring an exact match, there is a $(13{,}248/13{,}300)^{20} = 0.92$ or 92% chance that no correct linkages would be eliminated, leaving an 8% chance that a correct linkage might be eliminated. For a more accurate mass measurement, when the mass measurement error limit is 0.1 Da, the chance that 1 of 20 randomly detected masses would eliminate a correct linkage would decrease to 0.8% ($1-(132{,}948/133{,}000)^{20}$).

Complementary structural information from direct analysis

Returning to the three candidate disulfide structures (Table 5) resulting from processing the conservatively selected mass spectral data by the NSMA, direct analysis of singly reduced isoforms of ecTβR2 was pursued to determine which one of the three is the correct disulfide structure (see footnote 6). Analysis of an aliquot of the fractions represented in the HPLC chromatogram in Fig. 1 by MALDI-MS indicated that fractions A and B contained only the starting material or its singly reduced isoforms (indicated by a 50-Da shift caused by the addition of –CN groups to the two nascent sulfhydryls [2]). Fractions A and B were pooled and treated with 1 M NH_4OH to effect cleavage of the protein backbone on the N-terminal side of the cyanylated cysteines in the singly reduced isoforms [2].

Analysis of the cleavage reaction mixture from the chemical processing (not yet including complete reduction of the residual disulfides) of fractions A and B by MALDI-MS (data not shown) indicated a peak for an ion of 1961.3 Da, which is in good agreement with the calculated mass of 1961.2 Da for an expected cleavage fragment, itz-31-47, represented below:



Because this fragment was derived from a singly reduced isoform, a disulfide linkage must have been present originally between Cys31 and Cys48. Following complete reduction of the reaction mixture, reanalysis by mass spectrometry indicated that the mass of the protonated cleavage fragment described above shifted to 1963.1 Da (calculated 1963.2 Da), confirming the presence of a disulfide bond (between Cys38 and Cys44 by mass mapping) in the original putative cleavage fragment. Finding mass spectral evidence for disulfide bonds between Cys31-Cys48 and Cys38-Cys44 confirmed structure C as being correct among the three candidate disulfide structures (Table 5).

Table 5
Disulfide linkages in candidate structures from the NSMA

A	B	C
115-120	115-120	115-120
98-113	98-113	98-113
54-78	54-78	54-78
38-61	38-44	38-44
31-48	31-61	31-48
28-44	28-48	28-61

C was later confirmed as the correct disulfide structure with additional evidence from direct analysis.

Consequence of incomplete cyanylation

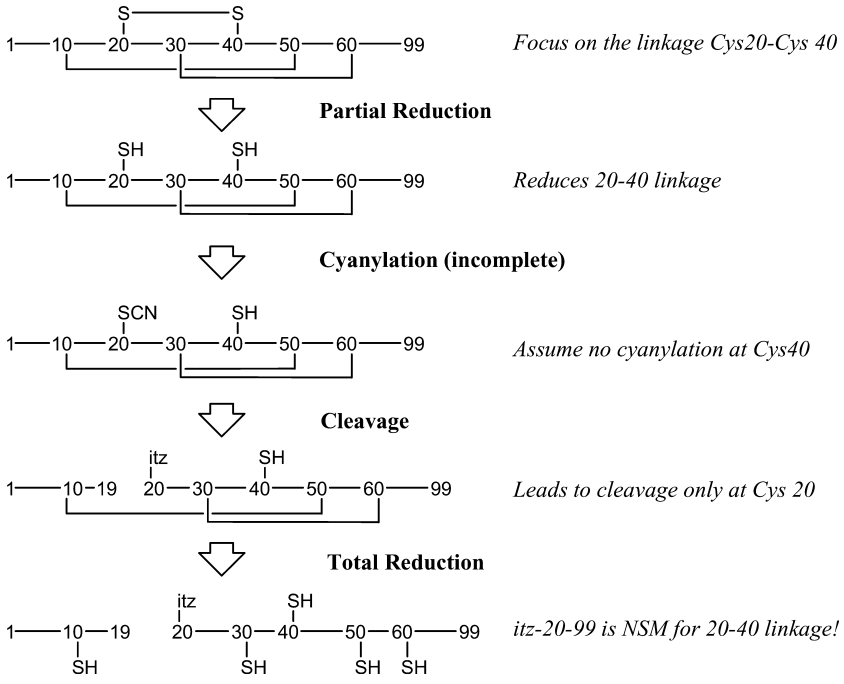
If cyanylation following partial reduction is incomplete, the resulting cleavage products from incompletely cyanylated species will, according to the logic of the NSMA, rule out the corresponding correct disulfide linkages [6]. The basis for this is illustrated in Scheme 3 with a hypothetical example.

Based on detection of the CN-induced cleavage product, itz-20-99 at the bottom of Scheme 3, containing free sulfhydryls at positions 30, 40, 50, and 60, the NSMA would indicate that no disulfide bond could have existed between 20 and 30 or 40 or 50 or 60. Of course, the first line of Scheme 3 shows that a disulfide bond between 20 and 40 did exist; the NSMA eliminated the 20-40 linkage because the undercyanylated cysteine at position 40 gave rise to the CN-induced cleavage fragment, itz-20-99, a NSM for the Cys20-Cys40 linkage.

In reality, the undercyanylation illustrated in Scheme 3 rarely (if at all) occurs under optimal conditions for cyanylation [12]. However, if the analyst suspects that undercyanylated (sulfhydryl-containing) species might be present in the processed sample, use of an OrgHg solid-phase column as described here is effective in removing such species.

Effectiveness of the organomercurial solid phase for scavenging sulfhydryl-containing species

DABMI derivatization of sulfhydryl-containing peptides and proteins facilitates recognition of these species in complex mixtures due to their specific UV absorption (at 317 nm) and mass spectrometric fragmentation



Scheme 3.

pattern [16]. HPLC chromatograms by absorbance at 317 nm of the CN-induced cleavage products from partially reduced, intentionally undercyanylated, DABMI-labeled (with or without prior OrgHg-column treatment), cleaved, and then finally fully reduced ribonuclease A (RNaseA) or ecTβR2 are shown in Fig. 4 (Figs. 4A and 4B are reaction mixtures relating to RNase A; Figs. 4C and 4D are analogous chromatograms relating to ecTβR2). The 60% aliquot of each sample underwent OrgHg treatment followed by DABMI labeling to enhance detection of peptides that contained free sulfhydryl groups (Figs. 4A and 4C), while the 40% aliquot of each sample was directly labeled with DABMI without OrgHg treatment (Fig. 4B and 4D).

The most intense peak (34 min) in the HPLC chromatogram (Fig. 4) represents excess DABMI reagent, and the three earlier eluting peaks (time frame of 27–32 min) represent DABMI-related degradation products as confirmed by injection of samples containing only DABMI into the HPLC (data not shown). Other peaks (in the time frame of 37–55 min) arise from absorbance by DABMI-labeled peptides (maximum of 3 nmol for Figs. 4A and 4C, and 2 nmol for panels Figs. 4B and 4D), i.e., peptides from intentionally undercyanylated partially reduced protein molecules. The chromatograms in Figs. 4B and 4D show many peaks representing DABMI-labeled peptides, indicating that sulfhydryl-containing species are present in the intentionally undercyanylated preparations of RNase A and ecTβR2. There is no trace of such DABMI peaks in Figs. 4A and 4C, which are chromatograms of parallel samples of intentionally undercyanylated preparations of RNase A and ecTβR2 that were stripped of sulfhydryl-containing species by OrgHg treatment prior to DABMI labeling; this result is especially reassuring because the chromatograms in Figs. 4A and 4C were obtained from 50% more protein than was used to produce the chromatograms in Figs. 4B and 4D. The OrgHg column selectively trapped the proteins containing free cysteines; a chromatogram recorded in parallel at 214 nm showed several peaks in the 37- to 55-min range for all samples (data not shown). These data demonstrate the effectiveness of the OrgHg treatment in scavenging incompletely cyanylated protein molecules and suggest that testing by DABMI labeling will not be required routinely.

Caveat on heterogeneous cystinyl protein samples

The concept of the NSMA is predicated on the expectation that the sample consists of a single disulfide-bonded structure of the given cystinyl protein. The presence of isomeric disulfide structures within the original protein sample will cause the NSMA to rule out too many linkages to produce even one disulfide structure. This occurs because CN-induced cleavage products from one disulfide structure serve as NSMs for certain linkages in the other isomers, and vice versa. We suspect that isomeric microheterogenity in the recombinant ecTβR2 sample, rather than incomplete cyanylation during sample processing, was responsible for the low-intensity mass spectral data that caused the inappropriate elimination of a disulfide linkage in the less conservative selection of mass spectral data described above. The problem caused by microheterogeneity of the sample is

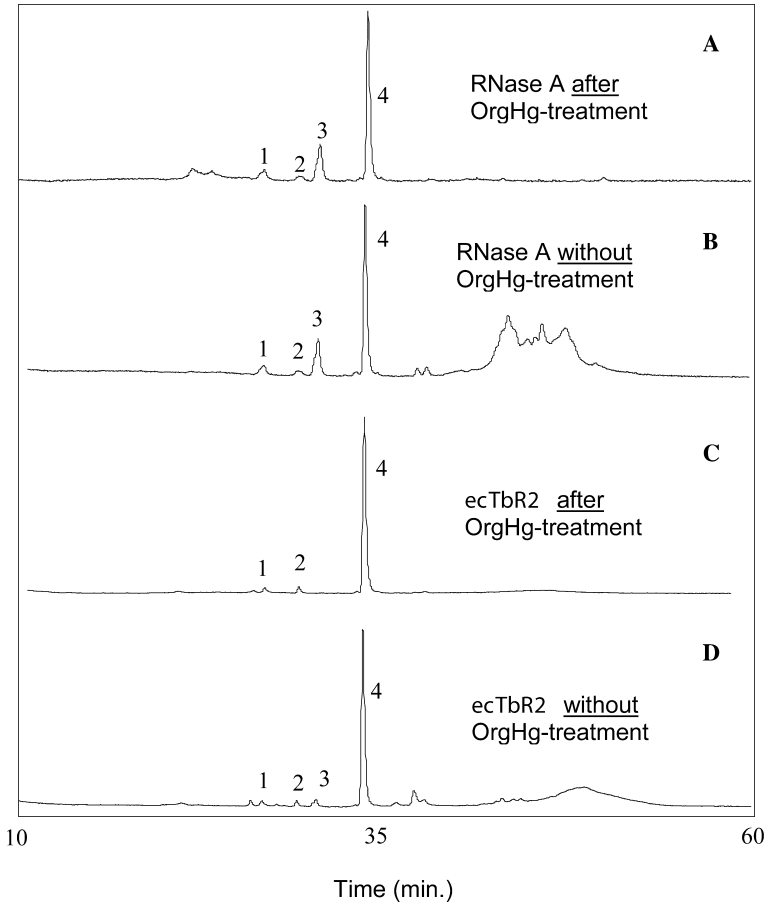


Fig. 4. HPLC chromatograms by absorbance at 317 nm of the cyanylation-induced cleavage products from partially reduced, intentionally undercyanylated ribonuclease A or ecTβR2 after (A and C) and without (B and D) scavenging of traces of sulhydryl-containing peptides by OrgHg treatment. All four samples (A–D) were treated with DABMI to enhance detection of sulfhydryl-containing peptides by UV absorption. Peaks 1-3 represent DABMI-related degradation products and peak 4 represents unreacted DABMI (unpublished data). Other peaks (in the time frame 37–55 min) arise from absorbance by DABMI-labeled peptides (maximum of 3 nmol for A and C and 2 nmol for B and D), i.e., peptides from protein molecules that were not completely cyanylated during intentional undercyanylation of the samples as described in the text.

not amenable to our chemical procedure with OrgHg and DABMI as described above, because there are no free sulfhydryl species in an isomer of ecTβR2 *nor* in the properly processed sample prior to complete reduction. Furthermore, the ecTβR2 sample was processed by the OrgHg procedure, which is a quantitative scavenger of sulfhydryl-containing species.

The major limitation of the NSMA approach to solving disulfide structures is the requirement for starting with one single, pure disulfide structure, free from other molecules with isomeric disulfide structures. Of course, this is a requirement for all methodologies used to solve disulfide structures because some conflicting evidence will always be produced by isomeric disulfide structures using any distinguishing technique. Thus, contaminating isomeric species cannot be tolerated at levels that allow their CN-induced cleavage fragments to be represented by significant mass spectral peaks, e.g., those greater than 5% relative intensity.

Notwithstanding the necessary precautions described above, the NSMA approach to solving disulfide structures using cyanylation-based mass-mapping methodology can be quite powerful, as demonstrated herein with the analysis of the complicated cystinyl ecTβR2 protein and previously with the well-studied Ribonuclease A protein [6].

Conclusions

A great advantage of the NSMA, in addition to convenience and automation, is its capacity to interpret mass spectra obtained during analysis of mixtures of CN-induced cleavage fragments without physically separating the partially reduced isoforms of the cystinyl protein and without knowledge of the extent of partial reduction. The power of the NSMA is demonstrated in processing mass spectral data obtained during the disulfide mapping of ecTβR2, a 13,800-Da protein cntaining six disulfide bonds, by partial reduction, cyanylation-induced cleavage, and mass-mapping methodology. Mass spectra from a complex mixture of

cleavage fragments from many (>10) chromatographically unresolved partially reduced isoforms were conservatively selected for processing by the NSMA. Despite apparent minor isomeric microheterogeneity, the NSMA substantially simplified the structure elucidation process by outputting only 3 candidate structures from a list of 10,395 theoretically possible disulfide structures. Direct manual analysis of mass spectral data of cleavage fragments from an ensemble of singly reduced isoforms of the protein confirmed the validity of 1 of the 3 candidate structures listed by the NSMA, in agreement with X-ray crystallography data for ecTβR2 [8].

References

[1] H. Yano, S. Kuroda, B.B. Buchanan, Disulfide proteome in the analysis of protein function and structure, Proteomics 2 (2002) 1090–1096.

[2] J. Wu, J.T. Watson, A novel methodology for assignment of disulfide bond pairings in proteins, Protein Sci. 6 (1997) 391–398.

[3] D.L. Smith, Z.R. Zhou, Strategies for locating disulfide bonds in proteins, Methods Enzymol. 193 (1990) 374–389.

[4] J.J. Gorman, T.P. Wallis, J.J. Pitt, Protein disulfide bond determination by mass spectrometry, Mass Spectrom. Rev. 21 (2002) 183–216.

[5] Y. Yang, J. Wu, J.T. Watson, Disulfide mass mapping in proteins containing adjacent cysteines is possible with cyanylation/cleavage methodology, J. Am. Chem. Soc. 120 (1998) 5834–5835.

[6] J. Qi, D. Hang, M. Rupp, C.R. Borges, W. Wu, E. Torng, J.T. Watson, Automated data interpretation based on the concept of 'negative signature mass' for mass-mapping disulfide structures of cystinyl proteins, J. Am. Soc. Mass Spectrom. 14 (2003) 1032–1038.

[7] A.P. Hinck, K.P. Walker, N.R. Martin, S. Deep, C.S. Hinck, D.I. Freedberg, Sequential resonance assignments of the extracellular ligand binding domain of the human TGF-beta type II receptor, J. Biomol. NMR 18 (2000) 369–370.

[8] P.J. Hart, S. Deep, A.B. Taylor, Z. Shu, C.S. Hinck, A.P. Hinck, Crystal structure of the human TbetaR2 ectodomain–TGF-beta3 complex, Nat. Struct. Biol. 9 (2002) 203–208.

[9] W.R. Gray, Disulfide structures of highly bridged peptides: a new strategy for analysis, Protein Sci. 2 (1993) 1732–1748.

[10] J. Qi, J. Wu, G.A. Somkuti, J.T. Watson, Determination of the disulfide structure of sillucin, a highly knotted, cysteine-rich peptide, by cyanylation/cleavage mass mapping, Biochemistry 40 (2001) 4531–4538.

[11] W. Jakoby (Ed.), Affinity chromatography, in: S.P. Colowick, N.O. Kaplan (Eds.), Methods in Enzymology, vol. 12, Academic Press, New York, 1971.

[12] J. Wu, J.T. Watson, Optimization of the cleavage reaction for cyanylated cysteinyl proteins for efficient and simplified mass mapping, Anal. Biochem. 258 (1998) 268–276.

[13] G.R. Jacobson, M.H. Schaffer, G.R. Stark, T.C. Vanaman, Specific chemical cleavage in high yield at the amino peptide bonds of cysteine and cystine residues, J. Biol. Chem. 248 (1973) 6583–6591.

[14] J. Qi, in: Ph.D. Dissertation, Dept. of Chemistry, Michigan State University, East Lansing, MI, 2001 pp. 71-77.

[15] M. Cadene, B.T. Chait, A robust, detergent-friendly method for mass spectrometric analysis of integral membrane proteins, Anal. Chem. 72 (2000) 5655–5658.

[16] C.R. Borges, J.T. Watson, Recognition of cysteine-containing peptides through prompt fragmentation of the 4-dimethylaminophenylazophenyl-4′-maleimide derivative during analysis by MALDI-MS, Protein Sci. 12 (2003) 1567–1572.

[17] Y. Ishii, S.S. Lehrer, Effects of the state of the succinimido-ring on the fluorescence and structural properties of pyrene maleimide-labeled alpha alpha-tropomyosin, Biophys. J 50 (1986) 75–89.

[18] J. Wu, D.A. Gage, J.T. Watson, 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. 235 (1996) 161–174.