Biochemistry

 Volume 26, Number 5 March 10, 1987

## Accelerated Publications

# The Primary Structure of Thioredoxin from *Chromatium vinosum* Determined by High-Performance Tandem Mass Spectrometry†

Richard S. Johnson and Klaus Biemann*

*Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139*

*Received November 26, 1986; Revised Manuscript Received January 7, 1987*

ABSTRACT: The primary structure of thioredoxin, a redox protein isolated from *Chromatium vinosum*, was determined by high-performance tandem mass spectrometry, which permitted sequencing of the 14 peptides (ranging in length from 2 to 18 amino acids) generated by digestion with trypsin and of several peptides produced by *Staphylococcus aureus* protease. The mass spectrometrically determined molecular weights of the peptides from the latter digest were used to properly align the tryptic peptides, which could also be accomplished on the basis of the considerable homology with *Escherichia coli* thioredoxin. Finally, the molecular weight of the *Chromatium* thioredoxin was determined by mass spectrometry and found to be 11 748.0, in good agreement with 11 750.2 calculated for the proposed sequence. Although it was difficult to establish by mass spectrometry, five leucines and three isoleucines could be identified, leaving only eight undifferentiated.

In the recent past, a vast majority of primary structures of proteins have been determined either by the stepwise, automated Edman degradation or, indirectly, by the translation of the DNA sequence of the gene coding for the protein. Alternatives based on entirely different methods are always useful, if only to provide at least a way to overcome obstacles in the course of the conventional approach. Mass spectrometric peptide sequencing has been such an alternative (Biemann, 1980), which has been particularly useful for the sequencing of N-blocked peptides that are not suitable for the Edman method.

In the earlier work, peptides had to be chemically modified to make them suitable for conventional mass spectrometry, which required volatilization into the ion source of the spectrometer. More recently, Barber et al. (1981) developed fast atom bombardment (FAB)[1] mass spectrometry, which makes it possible to ionize and analyze relatively large polar molecules dissolved in a suitable matrix, such as glycerol. The mass spectrum is dominated by a $(M + H)^+$ ion, where M refers to the molecular weight of the dissolved peptide.

Fast atom bombardment is a "soft" ionization technique, which generates chiefly ions related to the molecular weight, such as $(M + H)^+$ or $(M + Na)^+$ etc., but these rather stable ions undergo little fragmentation and, therefore, contain little structural (i.e., sequence) information. However, the $(M + H)^+$ ions can be fragmented upon collision with neutral atoms, such as He, to produce relatively abundant product ("daughter") ions, which are then mass analyzed in a second mass spectrometer (tandem mass spectrometry or MS/MS) [for a review, see McLafferty (1983)]. Fragmentation at peptide bonds generates ions, the masses of which can be correlated with the sequence of the amino acids in the peptide under investigation [for a review, see Biemann and Martin (1987)].

This principle has also been demonstrated recently on a series of peptides derived from apolipoprotein B by using a

† This work was supported by grants from the National Institutes of Health (GM05472 and RR00317).

* Address correspondence to this author.

[1] Abbreviations: FABMS, fast atom bombardment mass spectrometry; $(M + H)^+$, protonated molecular ion; FABMS/MS, fast atom bombardment tandem mass spectrometry; E, electric field; B, magnetic field; MS-1, first of two high-resolution mass spectrometers in tandem; MS-2, second of two high-resolution mass spectrometers in tandem; Cys*, carbamidomethylated cysteine; HPLC, high-performance liquid chromatography; Xle, leucine or isoleucine; $m/z$, mass to charge ratio; TFA, trifluoroacetic acid.

0006-2960/87/0426-1209$01.50/0 

triple quadrupole mass spectrometer (Hunt et al., 1986). In such a system, the ions colliding with the neutral gas have a low kinetic energy (10–100 eV) and suffer multiple collisions. We use instead a high-performance magnetic deflection mass spectrometer where single collisions take place at 10-keV kinetic energy, resulting in simpler mass spectra that can be interpreted reliably without the necessity of obtaining a second set of spectra after esterification of the peptide mixture, as is apparently necessary with low-energy collision taking place in the triple quadrupole. Furthermore, the high resolving power of a double focusing spectrometer permits the selection of the $(M + H)^+$ ion consisting exclusively of $^{12}C$ isotopes, which further simplifies the resulting product ion spectra. One of the major advantages of high-performance tandem mass spectrometry (employing two double-focusing mass spectrometers) is that mixtures of peptides can be sequenced by focusing one $(M + H)^+$ ion after the other onto the collector slit of the first mass spectrometer (MS-1) and transmitting it through the collision gas cell into the second mass spectrometer of high resolving power (MS-2) (Biemann, 1986; Biemann et al., 1986). A triple quadrupole mass spectrometer performs, in principle, the same task except that a narrow mass range (5–10 mass units) rather than a single mass is transmitted into the collision region.

The strategy developed for the determination of the primary sequence of a protein by tandem mass spectrometry is as follows: The reduced, S-carboxymethylated (or S-carbamidomethylated) protein is digested with a protease of high specificity, such as trypsin, the digest freed of remaining enzyme, buffer, etc. by reversed-phase HPLC and, at the same time, separated into fractions containing preferably one to three, but sometimes as many as ten, peptides, depending on the size of the protein and the complexity of the digest. A second set of overlapping peptides is generated by cleavage with an enzyme of different specificity [such as chymotrypsin or *Staphylococcus aureus* (strain V8) protease] or cyanogen bromide to allow the proper arrangement of the tryptic peptides.

We have used this approach to determine the primary structure of the thioredoxin from *Chromatium vinosum*, which is, to our knowledge, the first example of the sequencing of a protein entirely by tandem mass spectrometry. Finally, the molecular weight of the original molecule was determined to within ±2 daltons by using the high mass capability of the first stage of the tandem mass spectrometer. This result confirms that none of the tryptic peptides have escaped detection and is much more reliable than comparison of expected and determined amino acid composition, the conventional method used for this purpose.

## Experimental Procedures

*Preparation of Peptides.* Approximately 100 nmol of thioredoxin isolated from *Chromatium* was denatured in 6 M guanidinium chloride, 0.1 M tris(hydroxymethyl)aminomethane (Tris), and 1 mM ethylenediaminetetraacetic acid (EDTA) at pH 8.5 and reduced with a 50-fold molar excess of dithiothreitol for 1 h under nitrogen at 37 °C. The reduced thioredoxin was treated with a 2.6-fold molar excess of iodoacetamide over dithiothreitol at pH 8.5 and 37 °C for 1 h. The reaction mixture was desalted on a polyacrylamide gel column of molecular weight cutoff 6000 and evaporated to dryness on a Savant Speed Vac concentrator. Half of the product (corresponding to 50 nmol of starting protein) of the S-carbamidomethylated protein was digested with trypsin (50:1 substrate:enzyme) for 2 h in 0.1 M $NH_4HCO_3$ and 0.1 mM $CaCl_2$ at pH 8.5 and 37 °C. The tryptic digest was partially fractionated into ten fractions by reversed-phase high-performance liquid chromatography (HPLC). A portion of tryptic fraction 10 was treated with α-chymotrypsin (100:1 substrate:enzyme) for 45 min in the tryptic buffer at 37 °C and pH 8.5 and then fractionated by reversed-phase HPLC. Another one-fourth of the S-carbamidomethylated thioredoxin was digested with *S. aureus* (strain V8) protease (50:1 substrate:enzyme) in 0.1 M $NH_4HCO_3$ at 37 °C and pH 8.0 for 7 h and was separated into five fractions by reversed-phase HPLC.

*High-Performance Liquid Chromatography.* Reversed-phase HPLC was carried out with Waters Associates Model M6000A pumps, a Model 660 solvent programmer, a Rheodyne 7120 injector, and a Beckman Model 160 UV detector monitoring at 215 nm. A Waters μBondapak $C_{18}$ column was used at a flow of 1.5 mL/min with a linear gradient of $H_2O$/trifluoroacetic acid (TFA) (1:0.0005) to $CH_3CN$/$H_2O$/TFA (1:1:0.00042) over 30 min.

*Edman Degradations.* The manual Edman degradations were carried out according to the method described by Tarr (1977), modified for use in conjunction with FABMS.

*Fast Atom Bombardment Mass Spectrometry.* FABMS of the proteolytic peptides was carried out on the first (MS-1) of two mass spectrometers of a tandem high-resolution mass spectrometer (JEOL HX110/HX110) at an accelerating voltage of 10 kV and a resolution of 1:2200 and with 100-Hz filtering. For calibration, $(CsI)_nCs^+$ cluster ions were used. Single scans were acquired at a rate to scan from $m/z$ 200 to 3000 in about 2.5 min. A portion of each HPLC fraction was dissolved in glycerol at a concentration of 0.5–2.0 nmol/μL (based on the amount of initial protein) and 0.5–1.0 μL of the solution applied to a stainless steel probe tip. One-half microliter of 30% aqueous acetic acid alone or in combination with 5:1 dithiothreitol/dithioerythritol was then mixed with the sample on the probe tip. The JEOL FAB gun was operated at 6 kV with xenon as the FAB gas. For the determination of the molecular weight of the thioredoxin, the mass spectrometer (MS-1) was calibrated with $(CsI)_nCs^+$ ions ($n$ = 44–47) to cover the mass range from $m/z$ ~11 300 to 12 000. Approximately 0.8 nmol of the sample was deposited onto the sample probe and mixed with 1.5 μL of thioglycerol. The range from $m/z$ 11 600 to 11 800 was scanned in 10 s at a resolution of 1:500 and nine such scans were summed.

*Tandem Mass Spectrometry.* FABMS/MS was carried out by using all four sectors of the JEOL HX110/HX110, an instrument of $E_1B_1E_2B_2$ configuration. Collisional-induced fragmentation took place in the third field free region, thus operating both MS-1 ($E_1B_1$) and MS-2 ($E_2B_2$) as double-focusing instruments. Helium was used as the collision gas at a pressure sufficient to reduce the precursor ion signal by 50%. The FABMS/MS spectra were recorded at 30-Hz filtering, and the scan rate was the same as used for FABMS. Generally, for FABMS/MS the resolution of MS-1 was adjusted to transmit only the $^{12}C$ species of the $(M + H)^+$ ion to be analyzed. MS-2 was usually operated at a resolution of 1:1000; however, for the larger peptides [$(M + H)^+ > 2000$], the resolution was 1:500. MS-2 was calibrated with a mixture of CsI, NaI, KI, and LiCl. The FABMS and FABMS/MS spectra shown in Figures 1–3 are raw profile data of single scans and were recorded with a JEOL DA5000 data system.

## Results and Discussion

As a first step in the sequencing of the thioredoxin from *Chromatium*, a tryptic digest of the S-carbamidomethylated protein was partially separated by reversed-phase HPLC into ten fractions. (The modification of the cysteines is necessary

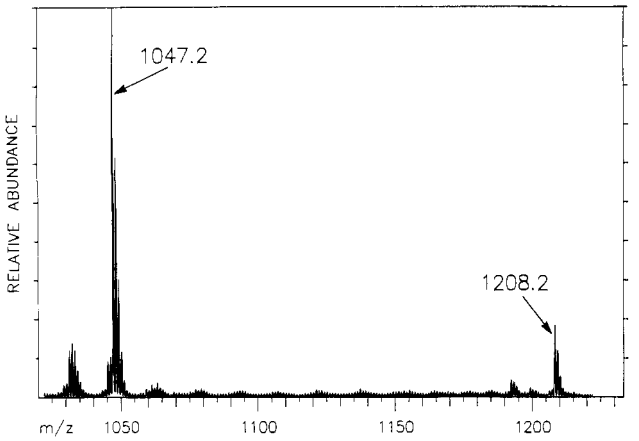
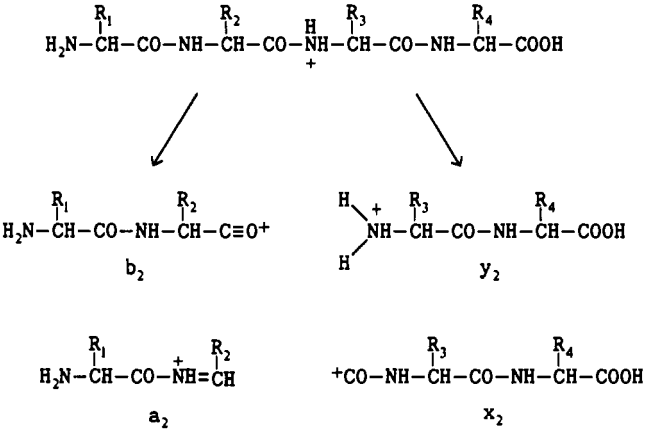



FIGURE 1: Segment of the FAB mass spectrum of HPLC fraction 9, indicating the presence of two peptides in the mass range of $m/z$ 1000 to 1250.

to make the protein susceptible to trypsin digestion.) The FABMS spectra of these ten fractions showed a total of 14 different protonated molecular ions [$(M + H)^+$], indicating the presence of 14 peptides. The $(M + H)^+$ ions of each of these were then subjected to collision with He, as described above, and the resulting FABMS/MS spectra recorded. As an example, Figure 1 shows the partial FAB mass spectrum of HPLC fraction 9, which indicates the presence of two peptides of $(M + H)^+ = m/z$ 1047.2, and $(M + H)^+ = m/z$ 1208.2. The FABMS/MS spectrum of the latter is shown in Figure 2. The accuracy of all mass assignments was ±0.3 dalton. The general structures of the fragment ions indicated in Figure 2 and Table I are shown in Scheme I. The labeling is a modification of the nomenclature proposed by Roepstorff and Fohlman (1984), except that we use lower case letters and $y_n$ instead of $Y_n''$.

The low abundance of sequence-specific fragment ions in conventional FAB mass spectra becomes a more severe problem at low mass where matrix ions are most abundant. However, in FABMS/MS this chemical noise is removed by MS-1 and only fragment ions derived from the selected precursor are observed. For instance, the sequence Val-Lys, the smallest tryptic peptide encountered, was deduced from the collision spectrum of $m/z$ 246.2 found in HPLC fraction 1, which showed the major sequence ions listed in Table I, as well

Scheme I: Some of the More Common Fragmentation Processes




as an ion resulting from the loss of the side chain of Val, and an ion at $m/z$ 84.0, the immonium ion of Lys ($^+H_2N{=}CHR_{Lys}$) cyclized by the elimination of $NH_3$.

In addition to the FABMS/MS spectrum shown in Figure 2, the spectra of the other 13 peptides of the tryptic digest were recorded. The sequences deduced from all of these spectra are summarized in Table I. The nominal masses (C = 12.0, H = 1.0, etc.) of the major sequence defining ion series in each spectrum are shown directly above and below each sequence. For instance, the FABMS/MS spectrum of $(M + H)^+ = 560.3$ has a $y_4$ ion at $m/z$ 404 and a $b_3$ ion at $m/z$ 228. The notation Xle stands for leucine or isoleucine, whenever they are indistinguishable (see below).

With few exceptions, the ion series of the peptides listed in Table I define the entire sequence of each peptide. However, when glycine is at the N-terminus of a peptide, the corresponding N-terminal ions are not observed and the C-terminal ion $y_{n-1}$, which would reveal the loss of the glycine (57 mass units), is not only of very low abundance but is also indistinguishable from the loss of the side chain of Xle. Such an N-terminal glycine can be identified conclusively by subjecting the peptide to a single Edman cleavage and determining the mass of the shortened peptide.

Another problem arose in sequencing the peptide of $(M + H)^+ = 2165.0$. The two N-terminal and two C-terminal amino acids could not be determined from this spectrum, which has been reproduced in a recent review (Biemann, 1986). The $b_2$

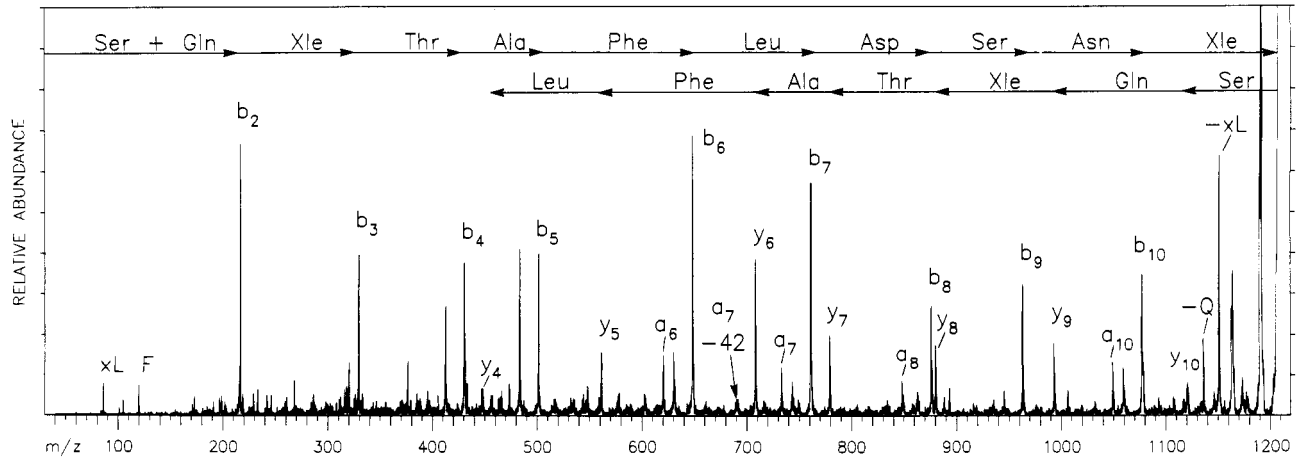



FIGURE 2: Tandem mass spectrum of the tryptic peptide of $(M + H)^+ = m/z$ 1208.2 (Figure 1) after collision and mass analysis in MS-2. See Scheme I for an explanation of the notation identifying the fragment ions. Peaks at $m/z$ 86 and 120 labeled xL and F are indicative of Leu and/or Ile and Phe, respectively, and those labeled –Q and –xL denote peaks due to loss of the side chains of Gln and Leu and/or Ile from the $(M + H)^+$ ion. The Xle in position 2 of this peptide was found to be Leu on the basis of the FABMS/MS spectrum of a peptide (positions 88–107) produced by *S. aureus* protease. Peaks 18 daltons lower than the $b_4$–$b_7$ ions are most likely due to the elimination of $H_2O$ (from Thr).

Table I: Sequences of Peptides Derived from a Tryptic Digest of *C. vinosum* Thioredoxin

| HPLC[a] frac. | $(M+H)^+$ obs. | Position | Sequence and m/z[b] of sequence characteristic ions | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | 147 | y | | | | | | | | | | | | | | | | |
| 1 | 246.2 | 53-54 | Val | Lys | | | | | | | | | | | | | | | | | |
| | | | 72 | | a | | | | | | | | | | | | | | | | |
| | | | | 218 | 147 | y | | | | | | | | | | | | | | | |
| 2 | 331.2 | 55-57 | Xle | Ala | Lys | | | | | | | | | | | | | | | | |
| | | | 86 | 157 | | a | | | | | | | | | | | | | | | |
| | | | | 322 | 175 | y | | | | | | | | | | | | | | | |
| 4 | 435.3 | 80-82 | Leu | Phe | Arg | | | | | | | | | | | | | | | | |
| | | | 86 | 233 | | a | | | | | | | | | | | | | | | |
| | | | | 345 | | 175 | y | | | | | | | | | | | | | | |
| 4 | 508.3 | 70-73 | Tyr | Gly | Xle | Arg | | | | | | | | | | | | | | | |
| | | | | 371 | 314 | 201 | x | | | | | | | | | | | | | | |
| | | | | 461 | 404 | 333 | 234 | 147 | y | | | | | | | | | | | | |
| 3 | 560.3 | 91-96 | Val | Gly | Ala | Val | Ser | Lys | | | | | | | | | | | | | |
| | | | | 157 | 228 | 327 | 414 | | b | | | | | | | | | | | | |
| | | | | 574 | 461 | | 263 | | y | | | | | | | | | | | | |
| 6 | 631.3 | 74-79 | Gly | Ile | Pro | Thr | Leu | Met | | | | | | | | | | | | | |
| | | | | 171 | 268 | 369 | 482 | | b | | | | | | | | | | | | |
| | | | | 687 | 574 | | 376 | 263 | | y | | | | | | | | | | | |
| 8 | 744.6 | 74-80 | Gly | Ile | Pro | Thr | Leu | Met | Leu | | | | | | | | | | | | |
| | | | | 171 | 268 | 369 | 482 | 613 | | b | | | | | | | | | | | |
| | | | | 733 | 676 | 547 | 448 | 319 | 248 | 147 | y | | | | | | | | | | |
| 3 | 790.3 | 83-90 | Gly | Gly | Glu | Val | Glu | Ala | Thr | Lys | | | | | | | | | | | |
| | | | | 115 | 244 | 343 | 472 | 543 | 644 | | b | | | | | | | | | | |
| | | | | 990 | 877 | 780 | 679 | 566 | 435 | 322 | 175 | y | | | | | | | | | |
| 9 | 1047.2 | 74-82 | Gly | Ile | Pro | Thr | Leu | Met | Leu | Phe | Arg | | | | | | | | | | |
| | | | | 171 | 268 | 369 | 482 | 613 | 726 | | | b | | | | | | | | | |
| | | | | 1121 | 993 | 880 | 779 | 708 | 561 | 448 | | | | y | | | | | | | |
| 9 | 1208.2 | 97-107 | Ser | Gln | Leu | Thr | Ala | Phe | Leu | Asp | Ser | Asn | Xle | | | | | | | | |
| | | | | 216 | 329 | 430 | 501 | 648 | 761 | 876 | 963 | 1077 | | b | | | | | | | |
| | | | | 1266 | 1152 | 1039 | 924 | 795 | 681 | 584 | 470 | 369 | 272 | 175 | y | | | | | | |
| 5 | 1379.7 | 58-69 | Xle | Asn | Xle | Asp | Glu | Asn | Pro | Asn | Thr | Pro | Pro | Arg | | | | | | | |
| | | | | 228 | 341 | 456 | 585 | 699 | | | 1011 | | 1205 | | b | | | | | | |
| | | | | 1631 | 1518 | 1447 | | 1251 | 1138 | 1023 | 894 | 781 | 710 | 595 | 466 | 303 | 232 | 175 | y | | |
| 10 | 1762.9 | 37-52 | Met | Xle | Ala | Pro | Val | Xle | Asp | Glu | Ile | Ala | Asp | Glu | Tyr | Ala | Gly | Arg | | | |
| | | | | | 1544 | | 1376 | 1277 | 1164 | 1049 | 920 | 807 | 736 | 621 | | | | | x | | |
| | | | | 1961 | 1846 | | 1646 | 1547 | 1410 | 1311 | 1210 | 1095 | 980 | 893 | 746 | | | | | | y |
| 6 | 2049.0 | 1-18 | Ser | Asp | Ser | Ile | Val | His | Val | Thr | Asp | Asp | Ser | Phe | Glu | Glu | Glu | Val | Xle | Lys | |
| | | | | 203 | 290 | 403 | 502 | 639 | 738 | 839 | 954 | 1069 | 1156 | 1303 | 1432 | 1561 | 1690 | 1789 | 1902 | | b |
| | | | | | 1980 | 1865 | | 1669 | 1556 | 1457 | 1342 | 1179 | 993 | 922 | 807 | 621 | 461 | 404 | 307 | | y |
| 10 | 2165.0 | 19-36 | Ser | Pro | Asp | Pro | Val | Leu | Val | Asp | Tyr | Trp | Ala | Asp | Trp | Cys | Gly | Pro | Cys | Lys | |
| | | | | 185 | 300 | 397 | 496 | 609 | 708 | 823 | | | | | | | | | | | b |

[a] The fraction listed is that in which the peptide was most preponderant, but some of them were also present in adjacent fractions. [b] Although determined to within ±0.3 dalton, to save space only nominal masses of the product ions are shown.

ion of $m/z$ 185 (see Table I) reveals that the N-terminus could only contain Ser and Pro or Ala and Xle. This question was resolved by digestion of tryptic HPLC fraction 10 with $\alpha$-chymotrypsin and fractionation by reversed-phase HPLC. The FAB mass spectrum of these fractions indicated seven $(M + H)^+$ ions: $m/z$ 620.9, 631.0, 1004.1, 1190.4, 1478.3, 1562.1, and 1708.2. The FABMS/MS spectrum of $(M + H)^+ =$ 620.9 indicated the sequence Cys*-Gly-Pro-Cys*-Lys, where Cys* is carbamidomethylated cysteine, thus confirming the C-terminal sequence of the original peptide of $(M + H)^+ =$ 2165.0 that exhibited only a weak $y_2$ ion. The FABMS/MS spectrum of $(M + H)^+ = 1004.5$ had the sequence (Ser,Pro or Ala,Xle)-Asp-Pro-Val-Leu-Val-Asp-Tyr, identifying it as the peptide containing the nine N-terminal amino acids of $(M + H)^+ = 2165.0$, but again did not define the sequence of the two N-terminal amino acids. A single Edman cycle on this fraction indicated the loss of 87 daltons, which reveals that Ser is at the N-terminus. The complete sequence of this peptide is as shown in Table I.

One last difficulty was the differentiation of the isomeric amino acids isoleucine and leucine. This can be accomplished by searching for an ion 28 daltons (loss of $C_2H_4$) below an $a_n$ ion containing a C-terminal Ile, or an ion 42 daltons ($C_3H_6$) below an $a_n$ ion containing a C-terminal Leu (Martin & Biemann, 1986). Figure 3 represents the region of $m/z$ 530–615 of the FABMS/MS spectrum of the tryptic/chymotryptic peptide of $(M + H)^+ = m/z$ 1004.5 mentioned above, showing the $b_6$ ion ($m/z$ 609.4), the $a_6$ ion ($m/z$ 581.5), and an ion 42 daltons lower than the $a_6$ ion ($m/z$ 539.4). The sixth amino acid in this peptide must therefore be Leu. To make this identification, significant ions of type $a_n$ are required; even the $a_6$ ion in Figure 2 is accompanied by a detectable peak

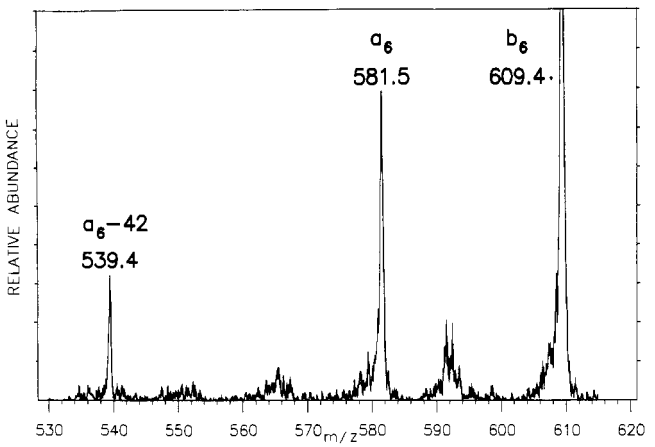



FIGURE 3: Expanded section of the tandem mass spectrum of the tryptic/chymotryptic peptide of $(M + H)^+ = m/z$ 1004.5 showing the loss of 42 daltons from the $a_6$ ion, indicating the amino acid in position 6 of this peptide is Leu, not Ile.

42 but not 28 daltons lower. FABMS/MS spectra lacking $a_n$ ions are thus not suitable for the differentiation of Leu and Ile. The FABMS/MS spectra of other peptides containing $a_n$ ions permitted the assignment of all but three leucines and five isoleucines (based on sequence and amino acid analysis).

An amino acid analysis of each peptide would allow the differentiation of Leu from Ile but only if (1) the peptide is completely separated from others and (2) there is only one of the two isomers present. Amino acid composition data would also reduce the number of possibilities to be considered in the interpretation of the spectra. However, this alone does not justify the effort and loss of material required for complete separation and purification of each peptide.

At this point, the sequences of all of the tryptic peptides were known, and they could be aligned by maximizing homology with *Escherichia coli* thioredoxin (Holmgren, 1968; Hoog et al., 1984). Without such a template the order of tryptic peptides can be established from the molecular weights of the peptides generated by another enzyme or chemical cleavage. The FABMS of HPLC fractions from a *S. aureus* protease digest of S-carbamidomethylated *Chromatium* thioredoxin indicated the presence of seven peptides of molecular weights 1180.8, 1559.1, 1707.9, 1787.2, 2049.3, 2585.5, and 3372.9. This makes it possible to order the tryptic peptides correctly as shown in Figure 4, where the *S. aureus* peptides correspond to the regions 16–26, 72–85, 1–15, 72–87, 88–107, 49–71, and 16–44. One exception is the region covered by $(M + H)^+ = 2586.5$ (positions 49–71). Given only the molecular weight of this peptide, the three complete tryptic peptides encompassed could be arranged six different ways. However, the FABMS/MS spectrum of $(M + H)^+ = 2586.5$ is compatible only with the sequence

```
     2422          2138      1911 1798 1727      1486
Tyr—Ala—Gly—Arg—Val—Lys—Xle—Ala—Lys—Xle—Asn—
              420  519       760  831           1186

1372      1144 1015 901           589  492                y
Xle—Asp—Glu—Asn —Pro—Asn—Thr—Pro—Pro—Arg—Tyr—Gly
1299 1414 1543 1657 1754 1868 1969 2066 2163 2319          a
```

The FABMS/MS spectra of most of the other *S. aureus* protease peptides were obtained. They confirmed the sequence determined from the tryptic peptides and exhibited a number of $a_n$ ions, which aided the differentiation of leucine and isoleucine, as mentioned above. The relatively high degree of homology with the *E. coli* thioredoxin increases the confidence in the structure of the *C. vinosum* protein even though the mass spectra of the tryptic peptides were sufficiently different (with the exception of 70–73, which is identical) that they had to be interpreted independently before the homologies could be recognized.

A comment about the sample requirement is in order. While 100 nmol of thioredoxin was used as the starting material, the products were split at two points and only about 0.5–1.0 nmol of peptide was used to obtain a FABMS/MS spectrum (see Experimental Procedures). We expect that, with proper care in sample handling and more experience in the methodology, much less material would suffice than what was used in the work described here.

The amino acids composing this sequence agree with the amino acid analysis, except that the latter showed nine prolines and six serines instead of eight and seven. It also indicated 8 leucines and 8 isoleucines, in agreement with the total of 16 required by the sequence. While this is generally considered adequate, we made use of the high mass range of the JEOL HX110 mass spectrometer (to mass 14 500 at 10-kV accelerating voltage) to determine the molecular weight of the thioredoxin from *Chromatium* directly. The value of 11 749.0 obtained for the weighted average of the isotope cluster of the

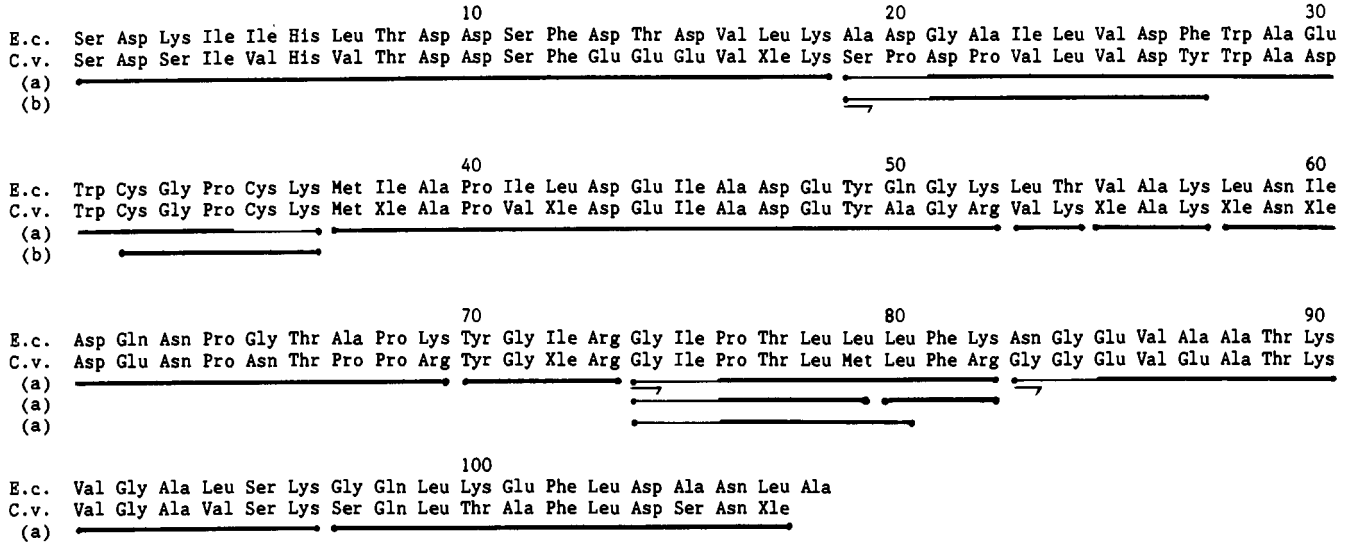



FIGURE 4: Amino acid sequence of the thioredoxin from *C. vinosum* (C.v.) aligned with that from *E. coli* thioredoxin (E.c.): (a) tryptic peptides; (b) further digestion with chymotrypsin. Heavy underlining: sequence derived from FABMS/MS spectra. Half-arrows: amino acid removed by manual Edman step.

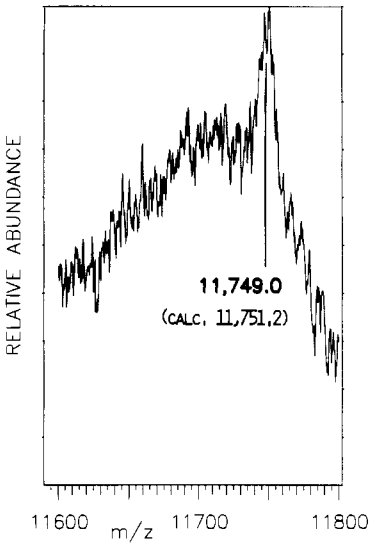



FIGURE 5: Unresolved isotopic multiplet of the $(M + H)^+$ ion of thioredoxin from *C. vinosum*. The mass value shown corresponds to the center of the peak.

$(M + H)^+$ ion agreed well with that calculated, 11 751.2 (Figure 5), lending further support to the correctness and completeness of the primary structure shown in Figure 4.

In conclusion, the sequence of *C. vinosum* thioredoxin, a protein of $M_r$ 11 750, was determined entirely by mass spectrometric methods, predominantly by tandem mass spectrometry. Only three leucines and five isoleucines remained undifferentiated. This technique is efficient and fast and, in contrast to conventional Edman methodology, does not require complete separation of proteolytic or chemically cleaved peptides. Although not a factor in the present case, N-terminally blocked peptides are equally suitable for mass spectrometric sequencing as demonstrated by Crabb et al. (1986) on a blocked tetradecapeptide from prostatropin. The capability of measuring the molecular weight of large peptides or the protein itself facilitates the alignment of proteolytic peptides and provides a check on the primary structure deduced from the sequence data.

ACKNOWLEDGMENTS

We thank Bob Buchanan for providing the sample of thioredoxin and express appreciation to Stephen A. Martin and Hubert A. Scoble for stimulating discussion of matters concerning mass spectrometry, to James E. Biller for some of the computational data processing and presentation techniques, and to W. Rodney Mathews, who provided some preliminary data prior to the work discussed here.

REFERENCES

Barber, M., Bordoli, R. S., Sedgwick, R. D., & Tyler, A. N. (1981) *J. Chem. Soc., Chem. Commun.*, 325–327.

Biemann, K. (1980) *Biochem. Appl. Mass Spectrom. (1st Suppl. Vol.)*, 469–525.

Biemann, K. (1986) *Anal. Chem. 58*, 1289A–1300A.

Biemann, K., & Martin, S. A. (1987) *Mass Spectrom. Rev. 6*, 1–76.

Biemann, K., Martin, S. A., Scoble, H. A., Johnson, R. S., Papayannopoulos, I. A., Biller, J. E., & Costello, C. E. (1986) in *Mass Spectrometry in the Analysis of Large Molecules—Proceedings of the Texas Symposium on Mass Spectrometry* (McNeal, C. J., Ed.) pp 131–149, Wiley, Sussex, England.

Crabb, J. W., Armes, L. G., Carr, S. A., Johnson, C. M., Roberts, G. D., Bordoli, R. S., & McKeehan, W. L. (1986) *Biochemistry 25*, 4988–4993.

Holmgren, A. (1986) *Eur. J. Biochem. 6*, 475–484.

Hoog, J.-O., Von Bahr-Lindstrom, H., Josephson, S., Wallace, B. J., Kushner, S. R., Jornvall, H., & Holmgren, A. (1984) *Biosci. Rep. 4*, 917–923.

Hunt, D. F., Yates, J. R., III, Shabanowitz, J., Winston, S., & Hauer, C. R. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 6233–6237.

Martin, S. A., & Biemann, K. (1986) presented at the 34th Annual Conference on Mass Spectrometry and Allied Topics, Cincinnati, OH, pp 854–855.

McLafferty, F. W., Ed. (1983) *Tandem Mass Spectrometry*, Wiley, New York.

Roepstorff, P., & Fohlman, J. (1984) *Biomed. Mass Spectrom. 11*, 601.

Tarr, G. A. (1977) *Methods Enzymol. 47*, 335–357.