Assignment of protein disulphides by a computer method using mass spectrometric data

Carlo Caporale[a,*], Ciro Sepe[a], Carla Caruso[a], Piero Pucci[b], Vincenzo Buonocore[a]

[a] *Dipartimento di Agrobiologia ed Agrochimica, Università della Tuscia, Via S. Camillo de Lellis, 01100 Viterbo, Italy*
[b] *Dipartimento di Chimica Organica e Biologica, Università Federico II di Napoli and Servizio di Spettrometria di Massa del CNR, Napoli, Italy*

Received 18 July 1996; revised version received 2 August 1996

Abstract **We designed a computer program for the assignment of protein disulphides using mass spectrometric data. All the theoretical linear peptides containing from one to three cysteines are generated on the basis of the protein sequence. Their combination ways are determined in order to create all the possible disulphide-bridged fragments containing from two to six cysteines and to calculate their molecular weight. One, two and three S-S bridges per linked fragment were considered, taking into account the possibility that the fragments contain unabridged residues. The mass data obtained from the spectral analysis of peptide mixtures of the digested protein are then associated to the fitting structures of disulphide-bridged peptides, giving rise to the primary output. This output can then be screened by using information on the specificity of the proteolytic agent(s) used and/or any further mass data provided by Edman degradation and/or carboxypeptidase treatment of the peptide mixtures. The need for such a computer aid is discussed and examples of application are given.**

Key words: Disulfide bridge; Protein structure; Computer method; Mass spectrometry

1. Introduction

During the last decade, fast atom bombardment mass spectrometry (FAB-MS) has been widely used for the assignment of disulphide bridges in proteins [1–16]. The success of this procedure is mainly due to the advantage that the separation of complex peptide mixtures such as those generated by proteolytic digestion of native proteins is often unnecessary. Therefore, low amounts of the original protein are generally required and information for a rapid assignment can be obtained. The principle of the FAB-mapping strategy, first introduced by Morris and Pucci [1], consists in the direct identification of disulphide-bonded fragments in digestion mixtures. The attribution of the mass signals to the corresponding bridged peptides is usually confirmed by submitting the original peptide mixture to manual Edman degradation step(s) or carboxypeptidase treatment and observing the appearance of new signals in the mass spectra due to the expected loss of amino acids from the N- or C-terminus of the fragments. The focal point in this procedure consists in the unambiguous attribution of each mass signal to the corresponding structure of a cystine-linked fragment and/or cluster of peptides. The assignment procedure can be very complicated especially when aspecific proteases have to be used to digest the native protein with the aim of producing fragments containing a limited number of disulphide bridges (usually one or two). Moreover, depending on the size of the protein and the number of fragments, a single mass signal can be associated to many clusters of bridged peptides, with the ambiguity remaining still unsolved even after one step of chemical and/or enzymatic degradation [13].

This paper describes the design and application of a computer program analyzing the possible solutions and filtering the results on the basis of new data obtained after degradation and/or information on the specificity of the proteolytic agent(s). We also suggest that such a computer aid can be indispensable in many cases for the unambiguous assignment of disulphide bridges.

2. Materials and methods

The mass data and experimental procedures of the two proteins used in the application examples have already been reported [13,15]. Programs were written using Microsoft QuickBASIC (version 1.00b) and implemented on an Apple Macintosh LC 475 computer. The operative system was System 7.1. The compiled applications are compatible with all Apple Macintosh computers.

3. Results and discussion

3.1. Required data

The data required to run the program are the amino acid sequence of the target protein written in mono-literal code and the mass values corresponding to putative S-S bridged peptides, as provided by the mass analysis of protein digests. The specific algorithms described in this paper have been designed for the use of integer variables such as the nominal FAB-MS MH^+ data, since, up to now, these are the most used data for the assignment of protein disulphide bridges [1–16]. However, the average mass data usually provided by conventional instruments equipped for electrospray ionisation (ESI) or matrix-assisted laser desorption/ionisation (MALDI) can also be processed by the method. The corresponding algorithms are very similar to those described in this paper; although the calculation speed is reduced, the running time strictly depends on the calculator used. All the structures of disulphide-bridged peptides whose molecular weights match the experimental mass values are identified by these data. Some additional information concerning the specificity of the proteolytic agent(s) and/or the mass data obtained following Edman degradation or carboxypeptidase treatment of the peptide mixtures must be provided in order to filter the global results. In particular, when possible, the user should indicate the expected cleavage sites on the protein sequence by indicating the amino acid(s) recognized by the proteolytic agent(s) and the side, whether N- or C-terminal, where the cleavage is expected to occur. All these data are input from the keyboard and stored on the disk by simple input/edit routines. The program is divided into three sections, each corresponding

*Corresponding author. Fax: (39) (761) 357242.

PII S0014-5793(96)00894-0

Target sequence:

RVCMGKSQHHSFPCISDRLCSNECVKEEGGWTAGYCHLRYCRCQKAC

Total cysteines = 8 Positions: 3, 14, 20, 24, 36, 41, 43, 47

Constructing linear fragments containing one cysteine. Please, wait.....

File 3 : 33 possible fragments containing Cys 3
File 14: 66 possible fragments containing Cys 14
File 20: 24 possible fragments containing Cys 20
File 24: 48 possible fragments containing Cys 24
File 36: 60 possible fragments containing Cys 36
File 41: 10 possible fragments containing Cys 41
File 43: 8 possible fragments containing Cys 43
File 47: 4 possible fragments containing Cys 47

Constructing linear fragments containing two cysteines. Please, wait.....

File 3-14: 18 possible fragments containing Cys 3 and Cys 14
File 14-20: 44 possible fragments containing Cys 14 and Cys 20
File 20-24: 72 possible fragments containing Cys 20 and Cys 24
File 24-36: 20 possible fragments containing Cys 24 and Cys 36
File 36-41: 24 possible fragments containing Cys 36 and Cys 41
File 41-43: 20 possible fragments containing Cys 41 and Cys 43
File 43-47: 2 possible fragments containing Cys 43 and Cys 47

Constructing linear fragments containing three cysteines . Please, wait.....

File 3-14-20: 12 possible fragments containing Cys 3, Cys 14 and Cys 20
File 14-20-24: 132 possible fragments containing Cys 14, Cys 20 and Cys 24
File 20-24-36: 30 possible fragments containing Cys 20, Cys 24 and Cys 36
File 24-36-41: 8 possible fragments containing Cys 24, Cys 36 and Cys 41
File 36-41-43: 48 possible fragments containing Cys 36, Cys 41 and Cys 43
File 41-43-47: 5 possible fragments containing Cys 41, Cys 43 and Cys 47

Fig. 1. Identification of the linear fragments containing one, two and three cysteines in the γ-thionin sequence. The relative information is stored in the corresponding mapping files.

to a single application addressed to the solution of a specific problem, which will be illustrated separately.

3.2. Creating the information files

The first section of the program is devoted to generating all the linear peptides containing from one to three cysteine residues using the sequence of the target protein. Their molecular weights are calculated. These peptides are generated on a purely theoretical basis regardless the proteolytic agent(s) effectively used in the experiments. Their combination ways are then determined for the potentiality of creating all the possible disulphide-linked fragments containing up to six cysteines and of calculating the corresponding molecular weights. This represents the data map. Once stored on the disk, the data map will be used for any search of disulphide bridges using experimental mass data. Starting from each cysteine, the algorithm moves back and forth on the protein sequence adding one amino acid at a time until it reaches the previous Cys residue (or the N-terminus of the protein) on one side and the next Cys (or the C-terminus) on the other side. The initial and final positions of each peptide containing that particular cysteine are stored on the disk, generating a series of files. The same logic is employed to create all the theoretical fragments containing two and three cysteines. In order to increase the calculation speed, linear fragments with more than three Cys have not been considered, since the cystine-linked clusters deriving from the combination ways of these peptides are very rarely formed using an effective digestion method which

Cysteine Number (CN) = **8**

A)

Intrachain bridge:

Combination ways for linear fragments containing 2 cysteines: CN-1 = **7**

Combination ways for linear fragments containing 3 cysteines: CN-2 = **6**

B)

Cluster bridges:

Combination ways 1+1 =	CN x (CN-1)/2 = **28**
Combination ways 2+1 =	(CN-1) x (CN-2) = **42**
Combination ways 2+2 =	(CN-2) x (CN-3)/2 = **15**
Combination ways 2+1+1 =	(CN-1) x (CN-2) x (CN-3)/2 = **105**
Combination ways 3+1 =	(CN-2) x (CN-3) = **30**
Combination ways 3+2 =	(CN-3) x (CN-4) = **20**
Combination ways 2+2+1=	(CN-2) x (CN-3) x (CN-4)/2 = **60**
Combination ways 3+3 =	(CN-4) x (CN-5)/2 = **6**
Combination ways 3+1+1 =	(CN-2) x (CN-3) x (CN-4)/2 = **60**
Combination ways 3+2+1 =	(CN-3) x (CN-4) x (CN-5) = **60**
Combination ways 2+2+2=	(CN-3) x (CN-4) x (CN-5)/6 = **10**
Combination ways 3+1+1+1 =	(CN-2) x (CN-3) x (CN-4) x (CN-5)/6 = **60**

Fig. 2. Type and number of the possible combination ways of the linear peptides generating disulphide-bridged fragments containing from 2 to 6 cysteine residues. The calculation regards any protein containing a total of eight cysteine residues.

is fundamental for the search of disulphides. However, the combination ways of linear fragments containing more than three cysteines can be easily implemented when needed. For the same reason, cystine-linked fragments containing more than six cysteines have not been taken into account, as they are not informative for the unambiguous assignment of S-S bridges and require further cleavages between the Cys residues [1]. The algorithm then calculates the mass values of the linear peptides, storing them in increasing order. Fig. 1 shows the results obtained when this application was employed to generate all the theoretical linear peptides using the sequence of a γ-thionin from *Sorghum bicolor*, a 48-residue protein containing eight cysteines at positions 3, 14, 20, 24, 36, 41, 43 and 47 [15]. As can be observed, the program identified all the possible linear peptides containing one, two and three Cys residues listed in separate groups in the figure. Since the nominal MH^+ values were the experimental data used in the localisation of the disulphide bridges of the protein [15], the program calculated the MH^+ values of all the above peptides. The information regarding the initial and final position of each peptide and its MH^+ value was stored. At this point, the algorithm calculated and identified all their possible combination ways, generating disulphide-bridged fragments containing from a minimum of two to a maximum of six cysteines. These combination ways can be separated into two different groups, as shown in Fig. 2. A disulphide bridge, in fact, can be located either within the same peptide (intrachain bridge) or formed between two different peptides (interchain or cluster bridge). The total number of the possible combination ways involving intrachain bridges in the linear peptides of a protein containing 8 cysteines, such as the γ-thionin, is shown in Fig. 2A. The possible combination ways of linear peptides when forming cluster bridges are shown in Fig. 2B. The simplest way of generating a fragment containing a single S-S bridge consists in joining two peptides each of which has a single Cys residue, leading to the combination ways indicated as 1+1 in the figure. However, a disulphide bridge can also be formed between a peptide containing 2 Cys and a fragment with only 1 Cys, the odd cysteine being present in the free SH form (combination ways 2+1). When the formation of two or three S-S bridges is considered, further combinations must be taken into account, leading to the entire list of possibilities shown in Fig. 2B. All the combination ways are stored in an information file. It should be emphasised that each combination way gives rise to a great number of possible disulphide-linked peptides. For example, to determine all the possible fragments linked by the Cys residues 3 and 14 (just one of the 28 possible combinations of type 1+1, Fig. 2), the 33 peptides containing Cys-3 have to be combined with the 66 peptides including Cys-14 (see Fig. 1). This generates a total of 2178 fragments. Once all the combination ways of the linear peptides have been identified, the mass values of all the possible cystine-linked fragments can easily be calculated by the application described below. For example, if the combination ways 1+1 are taken into consideration and the MH^+ values are utilized, the molecular mass of the S-S bridged peptides is given by the following formula

$$MH^+ \text{ first peptide} + MH^+ \text{ second peptide} - 3$$

since 2 H atoms are missing when two thiols are converted to a disulphide and one of the protons should be taken out as the peptides will carry a single positive charge. The time required to generate the data map for the γ-thionin is about 2 min using a Macintosh LC 475.

3.3. Searching for the disulphide bridges

The second section of the program was designed to associate the experimental mass values to the corresponding structure of disulphide-linked peptides. The mass values relative to disulphide-bridged peptides are identified in the mass spectra following the procedure already described [1,13,15,16] and inserted into the program. The algorithm will search for the experimental values by exploring all the possible combination ways of the linear fragments generated by the first section of the program. The mass value of each disulphide-linked fragment is calculated following the correct formula. In fact, the possible presence of multiple S-S bridges within the same cluster of peptides as well as the occurrence of intrachain together with interchain bridges has been considered. As an example, two fragments joined by following the combination ways 2+2, i.e. both containing two cysteine residues, can be linked by either one or two S-S bridges, giving rise to structures with different mass values. This aspect should be considered even if the native protein does not contain free thiol groups, as the reduction of S-S bridges can take place under FAB conditions [4,13,17]. The algorithm has been designed to consider even the presence of either homoserine or the corresponding lactone at the C-terminus of peptides derived from CNBr hydrolysis of the native protein. Furthermore, the possible cyclization of N-terminal Gln residues to form pyroglutamate has also been taken into account. When the molecular mass of a disulphide-containing fragment matches the experimental data, the structure of the fragment is identified and

Target sequence:RV**C**MGKSQHHSFP**C**ISDRL**C**SNE**C**VKEEGGWTAGY**C**HLRY**C**R**C**QKA**C**

Cysteine positions : 3, 14, 20, 24, 36, 41, 43, 47

A)

Original mass data: 883, 1708, 1938, 2888

MH^+ = 883: 22 different structures; MH^+ = 1708: 512 different structures; MH^+ = 1938: 754 different structures
MH^+ = 2888: 5230 different structures

1) 883 fragment 2- 4+13- 17 (bridge 3 - 14)
2) 883 fragment 2- 6+24- 26 (bridge 3 - 24)
3) 883 fragment 3- 3+32- 38 (bridge 3 - 36)
4) 883 fragment 2- 4+42- 45 (bridge 3 - 43)
5) 883 fragment 1- 4+43- 45 (bridge 3 - 43)
6) 883 fragment 1- 5+45- 47 (bridge 3 - 47)
7) 883 fragment 1- 6+46- 47 (bridge 3 - 47)
8) 883 fragment 14- 14+32- 38 (bridge 14 - 36)
9) 883 fragment 14- 15+32- 37 (bridge 14 - 36)
10) 883 fragment 13- 15+38- 41 (bridge 14 - 41)
11) 883 fragment 14- 17+43- 46 (bridge 14 - 43)
12) 883 fragment 14- 17+44- 47 (bridge 14 - 47)
13) 883 fragment 20- 20+32- 38 (bridge 20 - 36)
14) 883 fragment 19- 20+32- 37 (bridge 20 - 36)
15) 883 fragment 19- 23+45- 47 (bridge 20 - 47)
16) 883 fragment 24- 24+32- 38 (bridge 24 - 36)
17) 883 fragment 22- 25+35- 37 (bridge 24 - 36)
18) 883 fragment 32- 38+41- 41 (bridge 36 - 41)
19) 883 fragment 32- 38+43- 43 (bridge 36 - 43)
20) 883 fragment 32- 38+47- 47 (bridge 36 - 47)
21) 883 fragment 1- 3+41- 44 containing 1 bridge and 1 free cysteine
22) 883 fragment 34- 36+40- 43 containing 1 bridge and 1 free cysteine

B)

Specificity filter applied: Y as C-terminal site, K as C-terminal site, R as C-terminal site.
MH^+ mass data to be filtered: 883, 1708, 1938, 2888

The mass value of the following cluster fragments fitted the experimental data:

883 fragment 1- 6+46- 47 (bridge 3 - 47)
1708 fragment 19- 26+40- 42+43- 45 cluster containing 2 bridges (20-41 and 24-43 or 20-43 and 24-41)
1938 fragment 7- 18+36- 39 (bridge 14 - 36)
2888 fragment 7- 18+27- 39 (bridge 14 - 36)
2888 fragment 2- 6+19- 39 cluster containing 1 bridge (3-20 or 3-24 or 3-36) and 2 free cysteines

Fig. 3. Results obtained by searching for the disulphide bridges of the γ-thionin using the experimental data. (A) Synthesis of the global output; the 22 structures fitting the mass value 883 are shown. (B) Complete output obtained by applying the cleavage specificity filter.

stored in an output file. The searching routine is accelerated since the mass values of the linear peptides to be added up had been stored in increasing order. Therefore, when the sum of their mass values added, according to the correct formula, is higher than the experimental data, the algorithm does not execute the next unnecessary sums and passes to a different search. It should be remarked that the total number of combination ways for fragments containing six Cys residues (3+3, 3+2+1, 3+1+1+1, 2+2+2 in Fig. 2) greatly increases with the increase of the number of cysteines present in the protein. For this reason, when using a personal computer, the running time for a complete search including these combination ways is reasonable for proteins containing up to 17 cysteines. In any case, this should be a very high threshold, since, up to now, the assignment of disulphide bridges by mass spectrometry has been done for proteins containing a maximum of 10 cys residues [13]. However, excluding the search for fragments containing six Cys residues, it is possible to analyze proteins containing more than 17 cysteines.

Fig. 3A shows the results obtained when the experimental data for the γ-thionin reported by Nitti et al. [15] are inserted into the program. Each of the four mass signals recognized as corresponding to S-S bridged peptides can be theoretically associated to a large number of different structures. It should be remarked that the number of possible structures fitting a single mass value greatly increases with the increase of the mass value. These results show that no conclusive attribution of the mass signals can be achieved without filtering the primary output with suitable evaluation methods. However, the strength of the computer approach to the problem of assigning the S-S bridges is well illustrated here by the fact that all the possible cystine-linked fragments whose mass values fit the experimental data have been identified. The time required to perform the above search is about 4 min using a Macintosh LC 475.

3.4. Filtering the primary output

This section of the program allows the user to screen the primary output of disulphide-linked structures by using further experimental information which can lead to the unambiguous assignment of the S-S bridges. The various disulphide-containing structures can be discriminated on the basis of the specificity of the proteolytic agent(s) used and/or by using new mass data obtained following manual Edman degradation step(s) or carboxypeptidase treatment of the digestion mixtures. Fig. 3B shows the results of the filtering procedure applied to the primary output of the γ-thionin. Since the native protein had been digested with trypsin [15], the output was screened on the basis of the enzyme specificity by indicating the C-terminal side of Arg (R), Lys (K) and Tyr (Y) as the expected cleavage sites, the latter being an aspecific cleavage which occurs frequently in tryptic digestions. The number of possible explanations for each mass signal was drastically decreased. The signals at *m/z* 883 and

Target sequence:
SGPWSWCDPATGYKVSALTGCRAMVKLQCVGSQVPEAVLRDCCQQLADINNEWCRCGDL
SSMLRSVYQELGVREGKEVLPGCRKEVMKLTAASVPEVCKVPIPNPSGDRAGVCYGDWAAY
PDV
Cysteine positions : 7, 21, 29, 42, 43, 54, 56, 82, 98, 113

A)
Original mass data: 1232, 1467, 1900, 1999
MH^+ = 1232: 454 different structures; MH^+ = 1467: 428 different structures; MH^+ = 1900: 3785 different structures; MH^+ = 1999: 4939 different structures

B)
New mass data after one step of Edman degradation: 990, 1265, 1597, 1757
The MH^+ values of the following cluster fragments fitted the new mass data after one step of Edman degradation:
MH^+ = 990 : 3 different structures: 1 deriving from 1232; 2 from 1999
MH^+ = 1265 : 20 different structures: 2 deriving from 1467; 7 from 1900; 11 from 1999
MH^+ = 1597 : 43 different structures: 30 deriving from 1900; 13 from 1999
MH^+ = 1757 : 3 different structures deriving from 1999

C)
MH^+ = 990 : 1 structure deriving from 1232 justifying the bridge 29-82
MH^+ = 1265 : 2 different structures deriving from 1467 justifying the bridges 42-98 or 43-98
MH^+ = 1597 : 14 different structures deriving from 1900 justifying the bridges 21-42 or 21-43 or 7-42 or 7-43 or 42-113 or 43-113
MH^+ = 1757 : 3 different structures deriving from 1999 justifying the bridges 29-82 or 21-53 or 43-54 or 43-56

D)
New mass data after two steps of Edman degradation: 763, 1063, 1530
The following data fitted the new mass values after two steps of Edman degradation:

Original signal:	MH^+ = 1232	27- 30 + 77- 83 (1 bridge 29-82)
First Edman step:	MH^+ = 990	28- 30 + 78- 83 (1 bridge 29-82)
Second Edman step:	MH^+ = 763	29- 30 + 79- 83 (1 bridge 29-82)
Original signal:	MH^+ = 1467	41-46 + 93-99 (1 bridge; 1 free cysteine)
First Edman step:	MH^+ = 1265	42- 46 + 94- 99 (1 bridge; 1 free cysteine)
Second Edman step:	MH^+ = 1063	43- 46 + 95- 99 (1 bridge 43-98)
Original signal:	MH^+ = 1900	41- 46 + 19- 22 + 93- 99 (2 bridges; 43-98 and 21-42 or 42-98 and 21-43)
First Edman step:	MH^+ = 1597	42- 46 + 20- 22 + 94- 99 (2 bridges; 43-98 and 21-42 or) 42-98 and 21-43)
Second Edman step:	MH^+ = 1063	43- 46 + 95- 99 (1 bridge 43-98) and MH^+ = 278 (21- 22)
Original signal:	MH^+ = 1999	27- 38 + 77- 83 (1 bridge 29-82)
First Edman step:	MH^+ = 1757	28- 38 + 78- 83 (1 bridge 29-82)
Second Edman step:	MH^+ = 1530	29- 38 + 79- 83 (1 bridge 29-82)

Fig. 4. Results obtained by searching for the disulphide bridges of the α-amylase inhibitor 0.28 using the experimental data. (A) Synthesis of the global output. (B) Synthesis of the output filtered by the data obtained after one step of Edman degradation. (C) Further synthesis of the previous output obtained by excluding incompatible data. (D) Complete output filtered by the data obtained after a second step of Edman degradation.

1938 could only be assigned to the peptide pairs 1–6+46–47 and 7–18+36–39, respectively, leading to the unambiguous identification of the disulphide bridges Cys-3–Cys-47 and Cys-14–Cys-36. This assignment also allowed us to discriminate between the two different structures which still match the mass value 2888, ruling out the cluster 2–6+19–39 and confirming the bridge Cys-14–Cys-36. However, it should be remarked that clusters containing one bridge and two free cysteines as the 2–6+19–39, very hard to identify without computer aid, could really be present in the digestion mixtures of other proteins, their mass signals being useful for S-S bridge assignment. Finally, the signal at *m/z* 1708 could be assigned only to a cluster of peptides bridged by two S-S bonds, with the ambiguity still unsolved concerning the pairings of the four Cys residues within the cluster (whether Cys-20–Cys-41 and Cys-24–Cys-43 or vice versa). These results are in agreement with those reported previously [15] and clearly indicate the usefulness of the computer program in the unambiguous attribution of the mass signals when specific proteolytic agents are used in protein digestions.

However, when aspecific proteases have to be used, a completely different approach has to be considered, since the uncertainty of the cleavage sites prevents the filtering of the primary output by specificity criteria. Fig. 4A shows the synthesis of the original output when the computer procedure was applied for the S-S bridge assignment of the α-amylase inhibitor 0.28, a 123-residue protein containing 10 cysteines [13]. The protein was digested with both elastase and trypsin [13] which prevented the use of specificity filtering. The primary output was screened by using new experimental mass data provided by submitting the unfractionated protein digest to manual Edman degradation and re-running FAB mass spectra after each degradation cycle [13]. The algorithm calculates the new mass values of each S-S containing structure following the removal of the N-terminal residue(s); the structures which do not match the new mass data are ruled out. Moreover, the possibility that a cluster fragment collapses when an N-terminal Cys is released by the chemical degradation together with the S-S linked peptide has also been considered. The mass values of both the fragments generated are then calculated.

Fig. 4B shows the synthesis of the results of this filtering procedure by following a single step of Edman degradation. The number of possible explanations for each mass signal is drastically decreased and further structures can be excluded on the basis of this output. In particular, the new signal at *m/z* 990 unambiguously derived from the signal at *m/z* 1232 since the signal at *m/z* 1232 disappeared after the step of degradation [13]. It could generate just the new signal at *m/z* 990. For the same reason, the signal at *m/z* 1265 could be due only to the degradation of one of the two different structures of MH^+ 1467. The signal at *m/z* 990 is associated to the cluster 28–30+78–83 containing a disulphide bridge between Cys-29 and Cys-82. Starting from this finding, it is possible to exclude from the filtered output further structures which are not compatible with this unambiguous assignment. In this way, the result shown in Fig. 4C is obtained. As can be observed, no further unambiguous attribution of the mass signals can be achieved in this case even after one step of Edman degradation. This demonstrates that without computer aid it is not possible to know in advance how many steps of degradation must be performed to provide unambiguous attributions of the mass signals.

Although the authors of the paper cited in [13] could not anticipate all the possibilities remaining after one step of degradation and summarized in Fig. 4C, they performed a second step of Edman degradation on the mixture in order to be certain of the assignments done [13]. Fig. 4D shows the complete results obtained by the program following the second Edman degradation cycle. Each mass signal can be, at this point, assigned to a single disulphide-bridged structure, leading to the identification of the disulphides Cys-21–Cys-42, Cys-29–Cys-82 and Cys-43–Cys-98 in agreement with the results reported previously [13].

4. Conclusion

Computer aids for calculating the mass values of fragment ions containing disulphide bridges have already been reported [6,18]. However, these programs are not designed to calculate and analyze the combination ways of the Cys residues. The approach described in this paper is totally different in that this program is devoted to the unambiguous assignment of disulphide bridges in proteins. All the theoretical fragments containing from two to six cysteines linked by one, two or three S-S bridges and fitting the experimental mass data are considered, thus ensuring that the correct pairings of disulphides are present in the primary output. These results can then be screened by using suitable methods based on further information available from the experimental procedures. Moreover, the program might also help in choosing the best strategy for the confident assignment of the S-S bridges. The assignments of the disulphide bridges reported in the application examples [13,15] were found to be correct, analyzing the experimental data by the computer method described in this paper. However, it should be clear that ambiguous assignments could be done in other cases without this type of computer aid. In fact, the complexity of the problem strictly depends on the size of the protein, the number of cysteine residues and the hydrolytic agents used. The program is available from the authors.

Acknowledgements: Research supported by National Research Council of Italy, Special Project RAISA, Sub-project N.2.

References

[1] Morris, H. and Pucci, P. (1985) Biochem. Biophys. Res. Commun. 126, 1122–1128.
[2] Yazdanparast, R., Andrews, P.C., Smith, D.L. and Dixon, J.E. (1987) J. Biol. Chem. 262, 2507–2513.
[3] Takao, T., Yoshida, M., Hong, Y.M., Aimoto, S. and Shimonishi, Y. (1984) Biomed. Mass Spectrom. 11, 549–556.
[4] Akashi, S., Hirayama, K., Seino, T., Ozawa, S., Fukuhara, K., Oouchi, N., Murai, A., Arai, M., Murao, S., Tanaka, K. and Nojima, I. (1988) Biomed. Environ. Mass Spectrom. 15, 541–546.
[5] Raschdorf, F., Dahinden, R., Maerki, W., Richter, W.J. and Merryweather, J.P. (1988) Biomed. Environ. Mass Spectrom. 16, 3–8.
[6] Sun, Y. and Smith, D.L. (1988) Anal. Biochem. 172, 130–138.
[7] Toren, P., Smith, D., Chance, R. and Hoffman, J. (1988) Anal. Biochem. 169, 287–299.
[8] Williamson, R.A., Marston, F.A., Angal, S., Koklitis, P., Panico, M., Morris, H.R., Carne, A.F., Smith, B.J., Harris, T.J.R. and Freedman, R.B. (1990) Biochem. J. 268, 267–274.

[9] Morris, H.R., Pucci, P., Panico, M. and Marino, G. (1990) Biochem. J. 268, 803–806.
[10] Ishibashi, Y., Kikuchi, T., Wakimasu, M., Mizuta, E. and Fujino, M. (1991) Biol. Mass Spectrom. 20, 703–708.
[11] Violand, B.N., Tou, J.S., Vineyard, B.D., Siegel, N.R., Smith, C.E., Pyla, P.D., Zobel, J.F., Toren, P.C. and Kolodziej, E.W. (1991) Int. J. Peptide Protein Res. 37, 463–467.
[12] Yoshino, K., Takao, T., Shimonishi, Y. and Suzuki, N. (1991) FEBS Lett. 294, 179–182.
[13] Poerio, E., Caporale, C., Carrano, L., Pucci, P. and Buonocore, V. (1991) Eur. J. Biochem. 199, 595–600.
[14] Glocker, M.O., Arbogast B., Schreurs, J. and Deinzer, M.L. (1993) Biochemistry 32, 482–488.
[15] Nitti, G., Orrù, S., Bloch, C., Morhy, L., Marino, G. and Pucci, P. (1995) Eur. J. Biochem. 228, 250–256.
[16] Vaccaro, A.M., Salvioli, R., Barca, A., Tatti, M., Ciaffoni, F., Maras, B., Siciliano, R., Zappacosta, F., Amoresano, A. and Pucci, P. (1995) J. Biol. Chem. 270, 9953–9960.
[17] Yazdanparast, R., Andrews, P., Smith, D.L. and Dixon, J.E. (1986) Anal. Biochem. 153, 348–353.
[18] Lee, T.D. and Vemuri, S. (1990) Biomed. Environ. Mass Spectrom. 19, 639–645.