

# CysMap and CysJoin: Database and tools for protein disulphides localisation

Carlo Caporale<sup>a,\*</sup>, Laura Bertini<sup>a</sup>, Piero Pucci<sup>b</sup>, Vincenzo Buonocore<sup>a</sup>, Carla Caruso<sup>a</sup>

<sup>a</sup> *Dipartimento di Agrobiologia ed Agrochimica, Università della Tuscia, via S. Camillo de Lellis, 01100 Viterbo, Italy*

<sup>b</sup> *Dipartimento di Chimica Organica e Biochimica, Università di Napoli Federico II, via Cinthia 6, 80126, Napoli, Italy*

Received 28 February 2005; revised 30 March 2005; accepted 20 April 2005

Available online 5 May 2005

Edited by Robert B. Russell

**Abstract** We have developed a computer program able to make user-customised databases derived from the public PIR non-redundant reference protein database. When the database of interest has been created, the user will generate the map of all the possible linear peptides containing one and two cysteines for each protein and combine them to calculate the mass of all the possible clusters of linear peptides linked by a disulphide bridge with a cysteine pair. It is also possible to create selected maps corresponding to peptides formed by the action of specific proteases. In this way, mass spectrometric data obtained from the hydrolysis of proteins of unknown sequence can be related to that contained in the database for quick disulphide assignment and protein identification. To confirm signal attribution, the program will also furnish the expected mass of cluster peptides after performing a cycle of Edman degradation. The utility of the program is discussed and examples of application are given.

© 2005 Federation of European Biochemical Societies. Published by Elsevier B.V. All rights reserved.

**Keywords:** Protein database; Mass spectrometry; Disulphide assignment; Software

## 1. Introduction

In the modern proteomics, mass spectrometry combined with electrophoresis is the elective technique for protein characterisation. In fact, the separation of proteins by electrophoresis followed by digestion with proteases is a common practice, the mass of the produced peptides being then determined by mass spectrometry analysis. These data are often used to perform similarity searches using public protein databases. To this purpose, many sites collecting useful links are present in the World Wide Web. As an example, the ExPasy site (<http://www.expasy.ch/tools>) offers a series of informatic tools based on the analysis of linear peptide mass fingerprints [1–4]. Anyhow, no software dedicated to disulphide bridges assignment is available at the moment. Since 1985 [5] and up to recent years [6–22], mass spectrometry has been widely used for the assignment of disulphide bridges due to the advantage that the separation of the peptide mixtures generated by proteolytic digestion of native proteins is often unnecessary. The technique allows the direct identification of cluster peptides

linked by disulphide bonds in the mass spectra of digestion mixtures by their unique mass values. The attributions of the mass signals to the corresponding bridged peptides can be confirmed by submitting the peptide mixture to manual Edman degradation step(s) and observing the appearance of new signals in the mass spectra due to the expected loss of amino acids from the *N*-terminus of the peptides [11,12]. The success of this procedure consists in the unambiguous attribution of a mass signal to the corresponding cystine-linked peptide containing a single S–S bridge, since the identification of cluster peptides containing more than two cysteines is not decisive for the disulphides assignment. In this case, further sub-digestions are necessary to produce useful peptides. In 1996, we developed a computer program generating all the possible pairings of Cys residues in a protein, followed by filtering the results on the basis of new data obtained after degradation and/or information on the specificity of the proteolytic agent(s) [23]. This program was developed for Macintosh computers and was able to work just on a single sequence inputted from the keyboard by the user. In this paper, we present a computer program able to work on entire databases extracted by the user from the PIR non-redundant reference protein database (PIR-NREF). In this way, mass spectrometric data obtained from the protein of interest can be used to perform similarity searches in the database for quick disulphide assignment and protein identification.

## 2. Materials and methods

Public PIR-NREF protein database was downloaded from the web site [ftp://ftp.pir.georgetown.edu/pir\\_databases/nref](ftp://ftp.pir.georgetown.edu/pir_databases/nref). The source code of the program was written using Real Basic 5.5.2 from Active Software. The applications have been developed on a Macintosh computer equipped with a double G5 processor (2.5 GHz), 2.5 GB RAM and compiled for Windows, Mac OS and Linux. The operative system was Mac OS 10.3.5. All the programs work both on Apple Macintosh and Windows PC computers. The output files can be read by the text editor included in the program or by any word processor.

## 3. Results and discussion

### 3.1. User-customised databases

The PIR-NREF protein database includes all sequences from PIR-PSD, Swiss-Prot, TrEMBL, RefSeq, GenPept, and PDB. It is updated bi-weekly and grouped by sequence identity and taxonomy at the species level. At the moment, it contains 1.638166e+6 sequences. The file structure is very suitable for

\*Corresponding author. Fax +39 0761 357242.  
E-mail address: [caporale@unitus.it](mailto:caporale@unitus.it) (C. Caporale).

the development of computer applications since each protein is identified by just two records, the first one containing the identification (ID) number and biological information, the second containing the amino acid sequence. We developed a series of tools able to extract and split the sequences of interest by the ID number, keywords, sequence length and cysteine number. All these applications work independently and can also be used in combination to produce user-customised databases. As an example, Fig. 1 summarises the results in constructing databases by the keywords “thionin”, “disintegrin”, “lipase”, “maturase”, and “amylase inhibitor”. The time required to extract each database from the whole PIR-NREF is variable from 135 to 150 s. Of course, the user can also build its own database writing and saving a text file containing the sequences of interest. In this case, the user will input each sequence in a PIR-NREF compatible format, namely a first record beginning with the symbol “>” and containing information, plus a second record containing the amino acid sequence written in monoliteral code without spaces. As discussed below, this can be very useful in comparing sequences of homologous protein to determine both differences and identities in the mass of cluster peptides useful for the disulphide assignment.

### 3.2. Constructing the map of linear peptides

This application generates the sequences of all the linear peptides containing one and two cysteine residues from the sequence of each protein of the database. Their molecular weights are also calculated. Starting from each cysteine, the algorithm moves back and forth on the protein sequence add-

ing 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. The same logic is employed to create the peptides containing two cysteines. Although this algorithm is similar to that already described [23], two main differences have to be remarked. The actual program is able to work on the whole database and any type of mass data (integer, average or monoisotopic) can be chosen by the user to create the map file. Furthermore, the peptides can be generated without any information on possible proteolytic agent(s), giving rise to the full pattern, or indicating the hydrolysis sites expected for specific endoproteases, giving rise to selected maps. Fig. 2 shows a partial output of the file map constructed on the database “thionin” (40 sequences, Fig. 1) indicating “K” and “R” as potential trypsin hydrolysis sites and using integer variables for molecular weight calculation. The content of each record is indicated. All the useful data to combine the linear peptides to identify potential cluster containing disulphide bridges calculating their molecular weights are stored. Helpful information can be deduced reading this file. For example, it should be evident that no mass signal could be found to assign the disulphide bridges involving the cysteines in position 3 and 12 of the first protein, since the tryptic hydrolysis should not produce any useful linear peptide. Of course, these assignments could be done by choosing a different enzyme. As a matter of fact, the same file can be constructed without any indication of hydrolysis sites. In this case, the potential peptides useful for

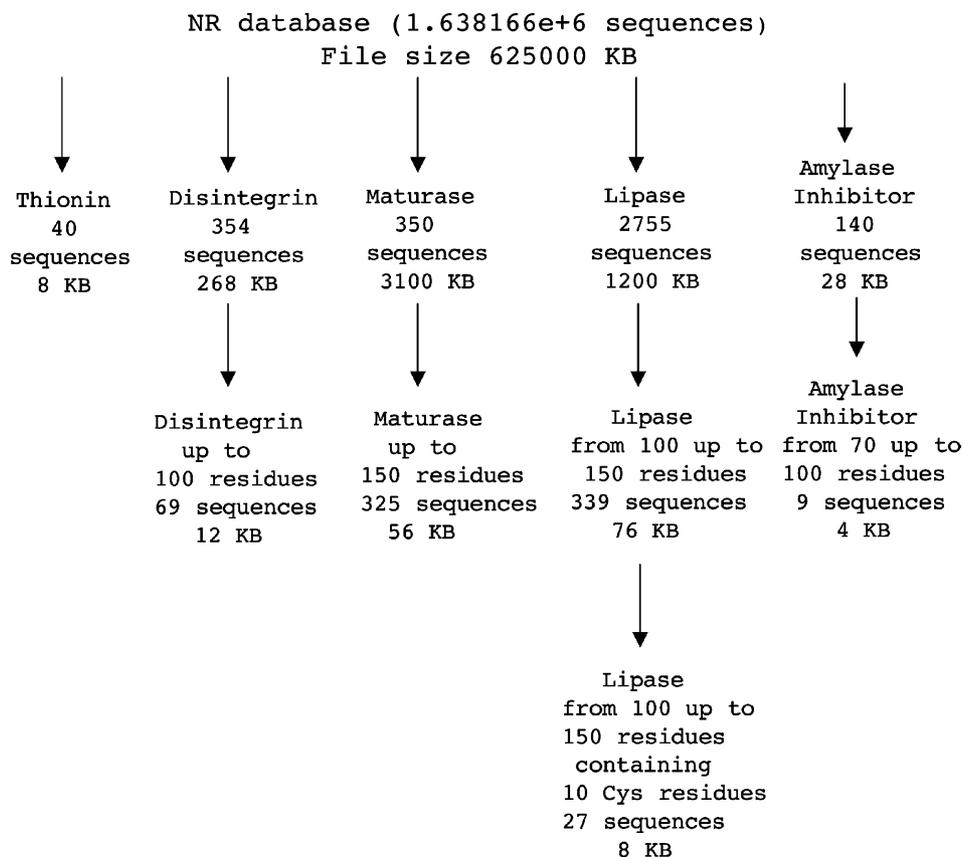


Fig. 1. Extraction of personalised databases from the PIR-Nref database by selecting keywords, sequence length and cysteine number.

```

40          Total sequences
>NF00739775 Thionin [Pyricularia pubera]          NREF-PIR information of the first sequence
KSCCRNTWARNCYNVCRLPGTISRREICAKKCDCKIISGTTCCPSDYPK    first sequence
8          Number of cysteines contained in the first sequence
3          Position of the first cysteine in the sequence (3)
a          Number of peptides containing cysteine 3 and ending with "K" or "R"(0)
4          Position of the second cysteine in the sequence (4)
2          Number of peptides containing cysteine 4 and ending with "K" or "R"(2)
C4-5R      Start-End residues and positions of the first peptide
278        Molecular weight of the first peptide
C4-10R     Start-End residues and positions of the second peptide
906        Molecular weight of the second peptide
12         Position of the third cysteine in the sequence (12)
0          Number of peptides containing cysteine 12 and ending with "K" or "R"(0)
16         Position of the fourth cysteine in the sequence (16)
8          Number of peptides containing cysteine 16 and ending with "K" or "R"(8)
C16-17R   Start-End residues and positions of the first peptide
278        Molecular weight of the first peptide
V15-17R   Start-End residues and positions of the second peptide
377        Molecular weight of the second peptide
N14-17R   Start-End residues and positions of the third peptide
491        Molecular weight of the third peptide
Y13-17R   Start-End residues and positions of the fourth peptide
654        Molecular weight of the fourth peptide
.....etc

----- Separating linear peptides containing two cysteines

36         Number of linear peptides of the first sequence containing two cysteines
C3-5R      Start-End residues and positions of the first peptide
381        Molecular weight of the first peptide
C31-34K    Start-End residues and positions of the second peptide
468        Molecular weight of the second peptide
S2-5R      Start-End residues and positions of the third peptide
468        Molecular weight of the third peptide
.....etc

```

Fig. 2. Partial file output of the linear tryptic peptides of the data base "thionin". Integer variables are utilised for molecular weight calculation.

the assignment of disulphides involving Cys3 and Cys12 are 3 and 32, respectively (not shown). The time necessary to produce these maps is about 30 s. As can be observed, the linear peptides are sorted in increasing molecular weight order. This is useful to enhance the calculation speed in finding the molecular weights of cluster peptides matching mass spectrometry data by the next application.

### 3.3. Disulphides assignment

Linear peptides containing one and two cysteines can be combined in two different ways to furnish unambiguous information on the localisation of disulphide bridges. Peptides containing 1 cysteine can be linked together by an interchain bridge (way 1 + 1). In this case, the number "C" of possible combinations is:

$$C = N \times (N - 1) / 2,$$

$N$  being the number of cysteines in the sequence. It is also possible that a bridge links two cysteines contiguous in the sequence. In this case, the number "C" of possible combinations is:

$$C = N - 1.$$

Of course, each combination will produce a different number of potential cluster peptides containing a disulphide bridge. As an example, the combination 1 + 1 of cysteine 4 + cysteine

16 (Fig. 2) will produce  $2 \times 8 = 16$  potential clusters containing the bridge C4–C16. The search for the assignment of disulphides will be performed by calculating the mass of all the possible clusters and selecting those matching the mass data inputted. There are no limits to the maximal polypeptide length and to the number of cysteines that the program can handle. The user can decide to retrieve results deriving from perfect matches or included in a window of values up to a tolerance of 100 ppm. Of course, the size of the output file will drastically increase in the second case. The appropriate formula is used. For example, in the case of fast atom bombardment (FAB) integer mass data, the formula

$$MH^+_{\text{peptide 1}} + MH^+_{\text{peptide 2-3}}$$

will be applied, since 2H atoms are missing when two cysteines are converted to cystine and one of the protons must be taken out, as the cluster will carry a single positive charge. The possible cyclisation of  $N$ -terminal glutamine to pyroglutamic acid, as well as the transformation of methionine residues to homoserine and homoserine lacton after cleavage with CNBr, has also been taken into account for calculations. Fig. 3 shows a partial output of the program combining the tryptic linear peptides of the proteins in the database "thionin" (Fig. 2) for the identification of cluster peptides matching the mass signal 883. This  $MH^+$  value was useful to identify the peptide pairs 1–6 + 46–47 in the sequence

```

>NF00217389 Thionin like protein [Nicotiana tabacum]
8 cysteines
Positions:34 45 51 55 65 72 74 78
G62-66R+C74-76R
883
After Edman degradation the mass should be:
723
>NF00739826 Thionin [Viscum album]
14 cysteines
Positions:28 29 37 41 50 54 56 64 73 77 82 94 98 102
C29-30R+S79-84K
883
After Edman degradation the mass should be:
693
>NF00996511 Thionin Osth1 [Oryza sativa]
13 cysteines
Positions:10 30 31 43 53 59 66 88 92 97 116 120 124
D51-55K+A119-121R
883
After Edman degradation the mass should be:
697
>NF00856860 putative thionin protein [Arabidopsis thaliana]
12 cysteines
Positions:28 29 41 51 57 65 85 89 94 116 120 124
S55-58R+I123-126K
883
After Edman degradation the mass should be:
683
C85-86K+E114-118K
883
After Edman degradation the mass should be:
651
>NF00381349 Thionin class 1 precursor (Peptide) [Tulipa gesneriana]
11 cysteines
Positions:15 16 24 28 38 42 44 52 95 99 103
G43-45K+E101-105K
883
After Edman degradation the mass should be:
697
>NF00381343 Thionin class 4 precursor [Tulipa gesneriana]
11 cysteines
Positions:21 30 34 39 45 49 51 59 106 110 114
G50-52K+E112-116K
883
After Edman degradation the mass should be:
697
-----etc

```

Fig. 3. Partial output showing the cluster tryptic peptides of the proteins in the database “thionin” matching the mass signal 883.

of a  $\gamma$ -thionin from *Sorghum bicolor*, a 48-residue protein containing eight cysteines at positions 3, 14, 20, 24, 36, 41, 43 and 47, leading to the unambiguous identification of the bridge Cys 3–Cys 47 [6]. As can be observed, this mass value can be attributed to different proteins of the database, these molecules differing for sequence, cysteines content and position. So, if a protein of interest can be classified as a thionin and a signal at  $m/z$  883 is found in the spectrum of a tryptic digest, a first assignment of disulphide bridge and protein identification could be done (Fig. 3). The program also calculates the mass which should be observed after performing a manual Edman degradation step on the digestion mixtures. This is of great utility to refine the assignments by a further tool that we have added to the program and named “Confirm attribution”. For example, searching for the new mass value 693 obtained after an Edman degradation step, the user will identify the cluster

C29-30R + S79-84K present in the thionin >NF00739826 and identifying the bridge Cys 29–Cys 82 (Fig. 4A). On the contrary, new mass values at  $m/z$  651 and 683 could be useful to distinguish between clusters both belonging to the protein >NF00856860, but identifying different bridges (C85–C116 or C57–C124, respectively, Fig. 4B and C). Furthermore, the signal at  $m/z$  1938, which identified the bridge Cys 14–Cys 36 in the case of the  $\gamma$ -thionin from *S. bicolor* [6], is also useful to identify a possible bridge between the two contiguous cysteines 31 and 32 of the protein >NF00994508 (Fig. 4D). The time required to perform the above search is about 2 s using the tryptic map of linear peptides of the database “thionin” that is formed by proteins of quite different sequence and cysteine content. Furthermore, the program can be of utility also comparing homologous proteins preserving cysteines in the same sequence position. Although it is very probable that such molecules maintain the same disulphide bridges, the confirmation of this hypothesis should be important. As an example, we constructed the database of the pathogenesis-related proteins of class IV (PR4). These proteins participate to mechanisms of plant defence against biotic and abiotic injuries. Actually, five homologous mature proteins have been sequenced [24–27], but the assignment of the three

#### A

```

>NF00739826 Thionin [Viscum album]
14 cysteines
Positions:28 29 37 41 50 54 56 64 73 77 82 94 98 102
C29-30R+S79-84K
883
After Edman degradation the mass should be:
693

```

#### B

```

>NF00856860 putative thionin protein [Arabidopsis thaliana]
12 cysteines
Positions:28 29 41 51 57 65 85 89 94 116 120 124
C85-86K+E114-118K
883
After Edman degradation the mass should be:
651

```

#### C

```

>NF00856860 putative thionin protein [Arabidopsis thaliana]
12 cysteines
Positions:28 29 41 51 57 65 85 89 94 116 120 124
S55-58R+I123-126K
883
After Edman degradation the mass should be:
683

```

#### D

```

>NF00994508 Leaf thionin Asthi2 [Avena sativa]
13 cysteines
Positions:13 31 32 40 44 54 58 60 68 90 94 118 126
V16-33K
1938
After Edman degradation the mass should be:
1839

```

Fig. 4. Identification and disulphide assignment of the proteins in the data base “thionin” by new signals after one step of Edman degradation: (A) 883 to 693; (B) 883 to 651; (C) 883 to 683; (D) 1938 to 1839.

disulphide bridges has been carried out just for one of them [24]. We have generated the map of the linear peptides and used the mass values of the tryptic clusters obtained from the protein “barwin” [24] for the disulphide assignment of the other four. Fig. 5 summarises the results of the search. The signals to localise the disulphide bridges C66–C123 and C31–C63 in barwin could be useful to assign the same bridges in the proteins wheatwin2 and wheatwin4, also leading to the assignment of the remaining bridge C52–C86. On the contrary, no assignment could be done for the proteins wheatwin1 and wheatwin3 by these data. Of course, this is due to the fact that differences in the primary structure of the corresponding cluster peptides are present. To help the user to decide which mass values could be convenient for the assignment, we developed a further tool generating and recording the molecular weight of all the possible cluster peptides from the map of linear peptides. Using this application, it is simple to determine that the integer mass

value of the cluster A8-43R + G59-65K identifying the bond C31–C63 should be 4775 in the proteins wheatwin1 and wheatwin3. Similarly, it is not hard to conclude that, due to the substitution Arg/Gln at position 68, the tryptic cluster C66-68R + I102-125D localising the bridge C66–C123 does not exist in these proteins. The same situation occurs for the cluster S44-58R + I82-101K, useful for the identification of the bridge C52–C86 in barwin. In these cases, the user will easily identify different clusters and mass values effective for the assignment (not shown).

#### 4. Conclusion

At the moment, this is the only program able to combine linear peptides of database-stored proteins for disulphide assignment. The described applications are integrated by a friendly interface (Fig. 6). In our intention, their usage should address

<b>A</b>						
PR4 protein	Average	Monoisotopic	Integer	Cluster	Bridge	
Barwin	3161.1	3157.457	3156	C66-68R+ I102-125D	C66-C123	
Wheatwin1	no match	no match	no match	no cluster	no assignment	
Wheatwin2	3161.1	3157.457	3156	C66-68R+ I102-125D	C66-C123	
Wheatwin3	no match	no match	no match	no cluster	no assignment	
Wheatwin4	3161.1	3157.457	3156	C66-68R+ I102-125D	C66-C123	
<b>B</b>						
PR4 protein	Average	Monoisotopic	Integer	Cluster	Bridge	
Barwin	4767.7	4761.212	4759	A8-43R+ G59-65K	C31-C63	
Wheatwin1	no match	no match	no match	no cluster	no assignment	
Wheatwin2	4767.7	4761.212	4759	A8-43R+ G59-65K	C31-C63	
Wheatwin3	no match	no match	no match	no cluster	no assignment	
Wheatwin4	4767.7	4761.212	4759	A8-43R+ G59-65K	C31-C63	
<b>C</b>						
PR4 protein	Average	Monoisotopic	Integer	Cluster	Bridge	
Barwin	3808.4	3804.777	3803	S44-58R+ I82-101K	C52-C86	
Wheatwin1	no match	no match	no match	no cluster	no assignment	
Wheatwin2	no match	no match	no match	no cluster	no assignment	
Wheatwin3	no match	no match	no match	no cluster	no assignment	
Wheatwin4	no match	no match	no match	no cluster	no assignment	

Fig. 5. Signal attribution and disulphide assignment of the proteins in the data base “PR4”.

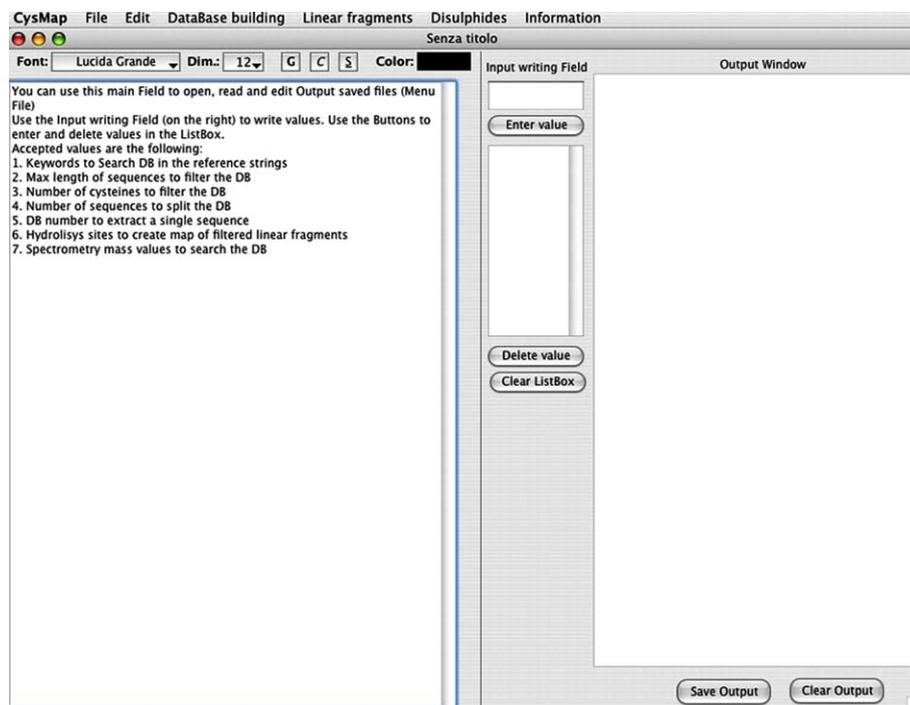


Fig. 6. Main window of the program.

two different goals. The first one is the achievement of information on possible disulphide localisation of proteins of unknown sequence. To this aim the user will perform quick similarity searches using the PIR-NREF database and mass spectrometry data. The second goal is the unambiguous attribution of mass signals obtained from digests of a protein of known sequence to cluster peptides containing disulphide bridges. In this view, the program can be considered as an upgrade of the previous one [23], since the code has been rewritten for both Macintosh and Windows modern computers. Furthermore, average and monoisotopic mass data as well as integer variables can be utilised. In conclusion, this software should be of great utility since the problem of disulphide assignment cannot be resolved without such a computer aid, being strictly dependent on the number of cysteine residues, the size of the protein and the hydrolytic agents used. The program is freeware and available from the authors. We are also planning the set-up of a web service.

**Acknowledgement:** This research was supported by the Ministero dell'Istruzione, Università e Ricerca Scientifica (MIUR), project MIUR-FIRB (RBNE01KZE7).

## References

- [1] Gattiker, A., Bienvenut, W.V., Bairoch, A. and Gasteiger, E. (2002) FindPept, a tool to identify unmatched masses in peptide mass fingerprint protein identification. *Proteomics* 2, 1435–1444.
- [2] Wilkins, M.R., Gasteiger, E., Bairoch, A., Sanchez, J.C., Williams, K.L., Appel, R.D. and Hochstrasser, D.F. (1998) *Protocols protein identification and analysis tools in the ExPASy server in: 2-D Proteome Analysis* (Link, A.J., Ed.), Humana Press, New Jersey.
- [3] Wilkins, M.R., Gasteiger, E., Wheeler, C., Lindskog, I., Sanchez, J.C., Bairoch, A., Appel, R.D., Dunn, M.D. and Hochstrasser, D.F. (1998) Multiple parameter cross-species protein identification using MultiIdent – a world-wide web accessible tool. *Electrophoresis* 19, 3199–3206.
- [4] Wilkins, M.R. and Williams, K.L. (1997) Cross-species identification using amino acid composition: a theoretical evaluation. *J. Theor. Biol.* 186, 7–15.
- [5] Morris, H. and Pucci, P. (1985) A new method for rapid assignment of S–S bridges in proteins. *Biochem. Biophys. Res. Commun.* 126, 1122–1128.
- [6] Nitti, G., Orru, S., Bloch, C., Morhy, L., Marino, G. and Pucci, P. (1995) Amino acid sequence and disulphide-bridge pattern of three gamma-thionins from *Sorghum bicolor*. *Eur. J. Biochem.* 228, 250–256.
- [7] Harwig, S.S., Swiderek, K.M., Lee, T.D. and Lehrer, R.I. (1995) Determination of disulphide bridges in PG-2, an antimicrobial peptide from porcine leukocytes. *J. Pept. Sci.* 1, 207–215.
- [8] Yoshida, S., Takamatsu, T., Denda, S., Ohnuma, M., Tajima, M., Lerner, E.A. and Kanda, F. (1996) Structural characterization and location of disulphide linkages of a potent vasodilatory peptide, recombinant maxadilan, by a multiple mass spectrometric approach. *Rapid Commun. Mass Spectrom.* 10, 641–648.
- [9] Calvete, J.J., Schrader, M., Raida, M., McLane, M.A., Romero, A. and Niewiarowski, S. (1997) The disulphide bond pattern of bitistatin, a disintegrin isolated from the venom of the viper *Bitis arietans*. *FEBS Lett.* 416, 197–202.
- [10] Flahaut, C., Mizon, C., Aumercier-Maes, P., Colson, P., Bailly, C., Sautiere, P. and Mizon, J. (1998) Disulphide bonds assignment in the inter-alpha-inhibitor heavy chains-structural and functional implications. *Eur. J. Biochem.* 255, 107–115.
- [11] Scaloni, A., Monti, M., Angeli, S. and Pelosi, P. (1999) Structural analysis and disulfide-bridge pairing of two odorant-binding proteins from *Bombyx mori*. *Biochem. Biophys. Res. Commun.* 266, 386–391.
- [12] Poerio, E., Caporale, C., Carrano, L., Pucci, P. and Buonocore, V. (1991) Assignment of the five disulfide bridges in an  $\alpha$ -amylase inhibitor from wheat kernel by Fast Atom Bombardment Mass Spectrometry and Edman degradation. *Eur. J. Biochem.* 199, 595–600.
- [13] Calvete, J.J., Jurgens, M., Marcinkiewicz, C., Romero, A., Schrader, M. and Niewiarowski, S. (2000) Disulphide-bond pattern and molecular modelling of the dimeric disintegrin

- EMF-10, a potent and selective integrin  $\alpha 5\beta 1$  antagonist from *Eristocophis macmahoni* venom. *Biochem. J.* 345, 573–581.
- [14] Calvete, J.J., Moreno-Murciano, M.P., Sanz, L., Jurgens, M., Schrader, M., Raida, M., Benjamin, D.C. and Fox, J.W. (2000) The disulfide bond pattern of catrocollastatin C, a disintegrin-like/cysteine-rich protein isolated from *Crotalus atrox* venom. *Protein Sci.* 9, 1365–1373.
- [15] Amoresano, A., Orru, S., Siciliano, R.A., De Luca, E., Napoleoni, R., Sirna, A. and Pucci, P. (2001) Assignment of the complete disulphide bridge pattern in the human recombinant follitropin beta-chain. *Biol. Chem.* 382, 961–968.
- [16] Cunsolo, V., Foti, S., Saletti, R., Ceraulo, L. and Di Stefano, V. (2001) Detection and localisation of disulphide bonds in a synthetic peptide reproducing the sequence 1–30 of Par j 1.0101 by electrospray ionisation mass spectrometry. *Proteomics* 1, 1043–1048.
- [17] Maltman, D.J., Simon, W.J., Wheeler, C.H., Dunn, M.J., Wait, R. and Slabas, A.R. (2002) Proteomic analysis of the endoplasmic reticulum from developing and germinating seed of castor (*Ricinus communis*). *Electrophoresis* 23, 626–639.
- [18] Amoresano, A., Pucci, P., Duro, G., Colombo, P., Costa, M.A., Izzo, V., Lamba, D. and Geraci, D. (2003) Assignment of disulphide bridges in Par j 2.0101, a major allergen of *Parietaria judaica* pollen. *Biol. Chem.* 384, 1165–1172.
- [19] Calvete, J.J., Moreno-Murciano, M.P., Theakston, R.D., Kisiel, D.G. and Marcinkiewicz, C. (2003) Snake venom disintegrins: novel dimeric disintegrins and structural diversification by disulphide bond engineering. *Biochem. J.* 372, 725–734.
- [20] Juarez, P., Sanz, L. and Calvete, J.J. (2004) Snake venomics: characterization of protein families in *Sistrurus barbouri* venom by cysteine mapping, N-terminal sequencing, and tandem mass spectrometry analysis. *Proteomics* 4, 327–338.
- [21] Cunsolo, V., Foti, S. and Saletti, R. (2004) Mass spectrometry in the characterization of cereal seed proteins. *Eur. J. Mass Spectrom.* 10, 359–370.
- [22] Ferranti, P. (2004) Mass spectrometric approach for the analysis of food proteins. *Eur. J. Mass Spectrom.* 10, 349–358.
- [23] Caporale, C., Sepe, C., Caruso, C., Pucci, P. and Buonocore, V. (1996) Assignment of protein disulphides by a computer method using mass spectrometric data. *FEBS Lett.* 393, 241–247.
- [24] Svensson, B., Svendsen, I., Hojrup, P., Roepstorff, P., Ludvigsen, S. and Poulsen, F.M. (1992) Primary structure of barwin: a barley seed protein closely related to the C-terminal domain of proteins encoded by wound-induced plant genes. *Biochemistry* 31, 8767–8770.
- [25] Caruso, C., Caporale, C., Poerio, E., Facchiano, A. and Buonocore, V. (1993) The amino acid sequence of a protein from wheat kernel closely related to proteins involved in the mechanisms of plant defence. *J. Protein Chem.* 12, 379–386.
- [26] Caruso, C., Caporale, C., Chilosi, G., Vacca, F., Bertini, L., Magro, P., Poerio, E. and Buonocore, V. (1996) Structural and antifungal properties of a pathogenesis-related protein from wheat kernel. *J. Protein Chem.* 15, 35–44.
- [27] Caruso, C., Nobile, M., Leonardi, L., Bertini, L., Buonocore, V. and Caporale, C. (2001) Isolation and amino acid sequence of two new PR-4 proteins from wheat. *J. Protein Chem.* 20, 327–335.