Journal of Protein Chemistry, Vol. 17, No. 8, 1998

A Computer Program to Compare Sequence Fingerprints of Homologous Proteins for the Rapid Assessment of Their Primary Structure Differences

Carlo Caporale,[1,2] Ciro Sepe,[1] Carla Caruso,[1] and Vincenzo Buonocore[1]

Received July 20, 1998

We have developed a computer program for the rapid assessment of the primary structure differences between a protein of unknown sequence and a homologous known protein. Both proteins are reduced, alkylated, and digested with the same hydrolytic agent. The unfractionated peptide mixtures are submitted to automatic sequence analysis. Based on the knowledge of the reference sequence, the program utilizes the analysis data to identify all the potential peptides present in the two mixtures, determining their primary structure, homology degree, and molecular weight calculated both as integer MH^+ and average mass variables. These fingerprints allow the user to easily identify the structural differences between the two proteins and clarify possible doubts by a mass spectrometric analysis of the two mixtures. In order to verify the utility of the program, we provide an application example using the already reported data of two homologous proteins.

KEY WORDS: Algorithm; computer; protein sequence; mass spectrometry; peptide mixtures.

1. INTRODUCTION

In recent years, sequence and mass spectrometric analyses have complemented each other in the rapid assessment of structural protein features since both techniques are able to analyze peptide mixtures. Automatic sequencers require a few picomoles of material to perform Edman degradation and problems related to different yields of phenylthiohydantoin (PTH)-amino acids and to the carryover of the reaction have been minimized in recent apparatus. Similarly, mass spectrometric analyses using modern ionization techniques are very sensitive. Of course, the data must be interpreted and many logical methods have been developed to solve specific problems quickly. The large spread of personal computers has inspired many authors to design algorithms for the interpretation and correlation of data. Methods based on knowledge of a sequence have been developed for the assignment of disulfide bridges (Caporale *et al.*, 1996*a*) and the determination of the protein hydrolysis pathway (Caporale *et al.*, 1996*b*, *c*; Petrilli *et al.*, 1994), while different approaches to assessing protein sequences have been reported that depend on the degree of knowledge of the structure of the target molecule. These programs are useful for determining a completely unknown sequence (Johnson and Walsh, 1992; Caporale *et al.*, 1993; Fernandez-de-Cossio *et al.*, 1995), establishing the alignment of a partially known sequence (Petrilli *et al.*, 1991; Caporale *et al.*, 1994), or identifying new proteins by comparing their mass fingerprints with those stored in structure databases (Mann *et al.*, 1993; Yates *et al.*, 1993, 1995; James *et al.*, 1993, 1994; Pappin *et al.*, 1993; Henzel *et al.*, 1993; Mann and Wilm, 1994; Cottrell, 1994; Griffin *et al.*, 1995; Patterson and Aebersold, 1995; Mortz *et al.*, 1996; Yates, 1996).

Noncomputerized strategies based on the analysis of peptide mixtures have been devised to determine the primary structure of a protein, given the sequence of a homologous protein (Caporale *et al.*, 1991, 1996*d*). The methods described in these papers are very useful for

[1] Dipartimento di Agrobiologia ed Agrochimica, Universita' della Tuscia, Viterbo, Italy.
[2] To whom correspondence should be addressed; e-mail: caporale @unitus.it.

0277-8033/98/1100-0867$15.00/0 © 1998 Plenum Publishing Corporation

establishing primary structure differences between isoforms of related proteins without performing expensive and time-consuming peptide purification steps and a large number of sequence and mass analyses. However, the correct interpretation of the data can be very difficult without appropriate computer aid since it depends on the protein size, the number of peptides in the mixtures, and the degree of identity of the two molecules. In this paper, we describe a program based on the combination and implementation of these strategies to analyze all the possible solutions in interpreting sequence and mass data.

2. EXPERIMENTAL PROCEDURES

2.1. Materials and Experimental Data

All the materials and procedures used in the application example of the program are described in Caporale *et al.* (1996*d*).

2.2. Hardware and Software

Program routines 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 self-running and compatible with all Apple Macintosh computers including PowerPC models equipped with System versions up to 7.5.3.

3. RESULTS AND DISCUSSION

3.1. Sequencing Data

The validity of the program for determining primary structure differences between homologous proteins has been tested using the already reported experimental data of two α-amylase isoinhibitors from wheat coded 0.28 and 0.39 (Caporale *et al.*, 1996*d*). Inhibitor 0.39 was considered as the protein of unknown sequence, while the homologous inhibitor 0.28 was the reference protein of known sequence. Both the reduced and S-pyridylethylated proteins (PyEt–0.28 and PyEt–0.39) were digested with endoproteinase Glu-C, which hydrolyzes peptide bonds at the level of glutamic acid (E) leaving this residue at the C-terminus of the produced fragments. The unfractionated peptide mixtures were submitted to automatic sequence analysis. The sequence of the reference protein (single-letter code) and the residues identified at each of 36 Edman degradation steps performed on the two mixtures are reported in Fig. 1. No further residue was identified starting from the 37th step, meaning that the largest peptides present in the two mixtures were 36 residues long (Caporale *et al.*, 1996*d*). These are the data necessary to run the program.

3.2. Program Architecture

The program is divided into the following sections, each corresponding to a distinct application:

A. Input/edit section: all the data are inputted from the keyboard and stored on the disk by a simple routine. The user inputs the amino acid sequence of the reference protein (Fig. 1A) and all the amino acids identified at each step of Edman degradation performed on the digestion mixtures (Figs. 1B and 1C). Moreover, some additional data about the digesting agent(s) can be furnished. In particular, when a protease with a marked specificity has been used, the user indicates the amino acid(s) expected at the hydrolysis sites (E in this case) and their position in the produced peptides (C-terminal in this case). This information can be useful to simplify the identification of the fragments present in the mixtures, even though it is not indispensable.

B. Identifying the reference fragments of the PyEt–0.28 mixture: this application reconstructs the sequence of each possible fragment matching the target sequence of the reference protein (Fig. 1A) from the data of its mixture sequence analysis (Fig. 1B). First the residues identified at the first degradation step are localized on the target sequence. For example, S at the first step (Fig. 1B) is associated to possible fragments starting from S_1, S_5, S_{16}, S_{32}, S_{60}, S_{61}, S_{65}, S_{93}, and S_{106} (Fig. 1A). Then the algorithm tries to extend the fragment sequences utilizing the residues identified at the next steps until no sequence can be further expanded: the extensions S_1–E_{36}, S_{61}–S_{65}, S_{65}–V_{66}, S_{93}–P_{95}, and S_{106}–G_{107} are possible since the required residues are present at the next steps, while the extensions starting from S_5, S_{16}, S_{32}, and S_{60} are not possible owing to the absence at the second step of W, A, Q, and S, respectively (Fig. 1B). All the possible extensions (minimum two residues long) and their start–end positions in the target sequence are shown in Fig. 2. By utilizing them, the algorithm can identify the real fragments present in the mixture. To this purpose, the consecutive or overlapping extensions allowing the complete reconstruction of the target sequence are considered. While the extensions 1–36, 37–52, 53–71, 70–74, 75–78, 78–87, 86–96, and 97–123 are useful, the re-

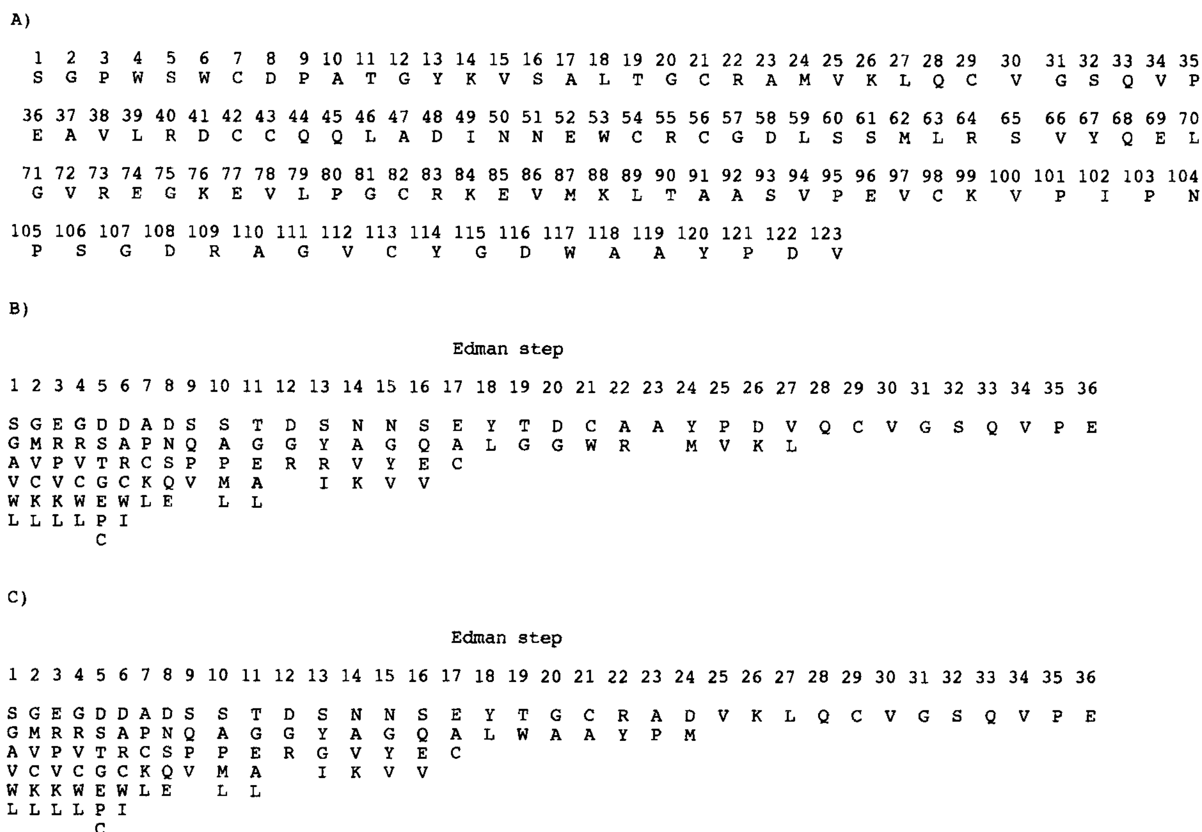

Fig. 1. (A) Reference amino acid sequence of inhibitor 0.28. (B) Amino acid residues identified at each sequencing step performed on the PyEt–0.28 peptide mixture. (C) Amino acid residues identified at each sequencing step performed on the PyEt–0.39 peptide mixture.

maining extensions marked with an asterisk are casual since their complementary sequences are missing. For example, the extension 20–22 is casual since no overlapping extension including T_{19} exists. As a consequence, also the extensions 23–25 and 25–27 are fortuitous. The algorithm consider, only the useful extensions to detect the real fragments present in the mixture. Consecutive extensions (e.g., 1–36, 37–52, and 53–71) identify the cleavages unambiguously (e.g., cleavages at level of E_{36} and E_{52}), while overlapping extensions should be further analyzed to detect the cleavages and identify the fragments. For example, the extension 70–74 starting from L_{70} indicates a cleavage at the level of E_{69}. Its overlapping extension 53–71 includes the trait L_{70}–G_{71} casually and the real fragment present in the mixture is 53–69. All the fragments identified in this way and their start–end positions in the target sequence are reported in Fig. 3. It is important to remark that all of them end with a residue of glutamic acid (E) despite the fact that information on the hydrolytic agent has not been exploited. In fact, the algorithm should not be dependent on the effectiveness of the hydrolytic agent(s) since partial or no cleavages could have been occurred at the expected sites. Anyway, the user can choose the utilization of the specificity information to speed up the identification of the fragments. In this way, the extension 53–71 would be truncated at the level of E_{69} and the fragment 53–69 identified immediately. It should be underlined that each peptide is unambiguously identified by those residues which can belong only to its sequence and are not in common with the sequence of other fragments. For example, V at the first step (Fig. 1B) is not useful for identifying a single peptide, since it is common to the sequence of the peptides 75–85, 86–96, and 97–123 (Fig. 3). L and M at the second step (Fig. 1B) are the first residues useful for ascertaining unambiguously the presence of the peptides 75–85 and 86–96, while V at the fourth step is the first residue distinguishing the peptide 97–123. The results show that endoproteinase Glu-C gave rise to complete cleavages

Extension	Position
SGPWSWCDPATGYKVSALTGCRAMVKLQCVGSQVPE	1-36
*SMLRS	61-65
*SV	65-66
*SVP	93-95
*SG	106-107
*GCR	20-22
*GVR	71-73
GKEV	75-78
*GCR	81-83
*GV	111-112
*AL	17-18
*AMV	23-25
AVLRDCCQQLADINNE	37-52
*AGCV	110-113
*VKL	25-27
*VG	30-31
*VLR	38-40
VLPGCRKEVM	78-87
VMKLTAASVPE	86-96
VCKVPIPNPSGDRAGVCYGDWAAYPDV	97-123
*VC	112-113
*WC	6-7
WCRCGDLSSMLRSVYQELG	53-71
LGVRE	70-74

Fig. 2. Extensions found in determining the sequence of the reference fragments present in the PyEt–0.28 mixture. Their start–end positions in the target sequence are indicated. The casual extensions are marked with an asterisk.

in the experimental conditions used. The application calculates the molecular weight of the produced fragments both as MH^+ and average mass variables (Fig. 3).

C. Identifying the fragments of the PyEt–0.39 mixture: this application checks the sequence of each possible fragment present in the mixture of the homologous protein of unknown sequence. To this purpose, the program utilizes the PyEt–0.39 mixture data (Fig. 1C) to attempt the reconstruction of the PyEt–0.28 fragment sequences (Fig. 3). First this algorithm checks whether the residues necessary to completely reconstruct the sequences are present at the useful degradation steps. The sequences 1–36, 37–52, 53–69, 70–74, 75–77, 78–85, and 86–96 are reconstructed starting from the first step and ending at the step equal to the fragment length (Fig. 1C), indicating that these peptides, marked with an asterisk in Fig. 3, are present also in the PyEt–0.39 mixture. This means that no primary structure difference between the two proteins exists in the region 1–96. Since the sequence of the reference fragment 97–123 cannot be completely reconstructed by perfect matches, the algorithm analyzes all the possibilities justifying the differences contained in this region. This is done by diversified searching, the results of which are summarized in Fig. 4. Starting from the first step in Fig. 1C, the following uncompleted reconstruction of the fragment sequence 97–123 is obtained: VCKVPIPNPSGD**X**AGVCY**XXXX**A**XXXX**, where **X** means that the matching residue is missing at the corresponding step of the sequence analysis. For example, the residue R present at the 13th step in Fig. 1B and belonging to the sequence of the reference fragment (position 13) is absent at the 13th step in Fig. 1C. Now the algorithm searches for possible substitutions, insertions, and deletions to replace the unknown residues X with the useful residues shown in Fig. 1C. Perfect matches of the sequence 97–123 were found from step 1–12 and 14–18, indicating that X at position 13 of the fragment is due to a substitution. The only difference between the 13th steps of the two analyses regards the residue R (Fig. 1B), which is replaced by G (Fig. 1C). Since the common residues S, Y, and I belong to the fragments 53–69, 1–36, and 37–52 which are present in both mixtures, the algorithm assumes that R_{109} in inhibitor 0.28 is substituted by G at the same position in inhibitor 0.39. Similarly, the residues G, D, W, and A at steps 19–22 are replaced by W, A, A, and Y; A at step 23 is present in both analyses; Y at step 24 is replaced by D; P at step 25 is replaced by V or not replaced by any residue; and D, V at steps 26–27 are not substituted at all. Consequently, it is possible that the fragment VCKVPIPNPSGDGAGVCYWAAYAD of 24 residues is present in the PyEt-0.39 mixture, indicating the following nine differences from inhibitor 0.28: substitutions R → G (position 109), G → W (position 115), D → A (position 116), W → A (position 117), A → Y (position 118), Y → D (position 120), and deletions of the residues 121, 122, and 123. Moreover, considering that P at the step 25 can also be replaced by V, the further fragment VCKVPIPNPSGDGAGVCYWAAYADV of 25 residues is compatible. As shown in Fig. 4 (alignment A), the algorithm calculates their molecular weight and homology degree with the 0.28 reference fragment 97–123. Of course, this is not the only search to be considered. In fact, some insertion or deletion could be responsible of the lack of perfect matches in the trait 115–123. So the algorithm shifts the residues W, A, A, Y, P, D, (V) present at steps 19–25 of the 0.39 analysis to steps 20–26 and 21–27 to determine the sequence of further compatible fragments. As one can observe, the best alignment is found by the double shift yielding the peptide VCKVPIPNPSGDGAGVCYWAAYPDV of 25 residues (Fig. 4, alignment C) which differs from the reference fragment 97–123 in one substitution at position 109 (R → G) and two deletions at positions 115 and 116 (G and D). Although the output shown in Fig. 4 strongly indicates that this is the real fragment present in the PyEt-0.39 mixture since it has the highest homology degree with the reference fragment (88.9%), the doubt can

0.28 Reference fragments	Start-End degradation step	Start-End sequence position	Integer MW	Average MW
*SGPWSWCDPATGYKVSALTGCRAMVKLQCVGSQVPE	1-36	1-36	MH^+ = 4125	4128.815
*AVLRDCCQQLADINNE	1-16	37-52	MH^+ = 2014	2015.291
*WCRCGDLSSMLRSVYQE	1-17	53-69	MH^+ = 2241	2244.603
*LGVRE	1-5	70-74	MH^+ = 572	572.662
*GKE	1-3	75-77	MH^+ = 333	333.356
*VLPGCRKE	1-8	78-85	MH^+ = 1005	1007.232
*VMKLTAASVPE	1-11	86-96	MH^+ = 1145	1146.380
VCKVPIPNPSGDRAGVCYGDWAAYPDV	1-27	97-123	MH^+ = 3058	3060.497

Fig. 3. Sequences of the reference fragments present in the PyEt–0.28 mixture. Their start–end degradation steps and sequence positions are indicated. The molecular weights are calculated both as integer MH^+ and average mass variables. The molecular weight of Cys residues is calculated as PyEt-Cys. The sequences of the fragments present in the PyEt–0.39 mixture perfectly matching the reference fragments are marked with an asterisk.

		Homology degree	Integer MW	Average MW
0.28 Reference fragment 97-123	VCKVPIPNPSGD**R**AGVCY**GDWAAYPDV**	100%	MH^+ = 3058	3060.497
Uncompleted 0.39 Matching fragment	VCKVPIPNPSGD**X**AGVCY**XXXXAXXXX**	----	----------	--------
Compatible 0.39 fragments				
Alignment A	VCKVPIPNPSGD**G**AGVCY**WAAYAD---**	66.7%	MH^+ = 2662	2664.051
	VCKVPIPNPSGD**G**AGVCY**WAAYADV--**	66.7%	MH^+ = 2761	2763.183
Alignment B	VCKVPIPNPSGD**G**AGVCY**-WAAAPD--**	70.4%	MH^+ = 2596	2597.991
	VCKVPIPNPSGD**G**AGVCY**-WAAAPDV-**	70.4%	MH^+ = 2695	2697.124
Alignment C	VCKVPIPNPSGD**G**AGVCY--WAAYPD-	85.2	MH^+ = 2688	2690.089
	VCKVPIPNPSGD**G**AGVCY--WAAYPDV	88.9%	MH^+ = 2787	2789.221

Fig. 4. Sequences of the 0.39 peptides homologous to the 0.28 reference fragment 97–123 and compatible with the experimental data. The homology degree and the molecular weight are calculated for each possible peptide corresponding to alignments A, B, and C. The molecular weight of Cys residues is calculated as PyEt-Cys. The differences from the reference fragment are shown in boldface.

be easily resolved by performing a mass analysis since the algorithm will furnish complete molecular weight information of the peptides whose presence in the mixture is possible.

4. CONCLUSION

The case used to illustrate the program was already analyzed without the aid of a computer to demonstrate the validity of strategies based on the analysis of peptide mixtures (Caporale *et al.*, 1996*d*). This application example is elementary since the two proteins show a very high degree of identity (97.6%) leading to the simple outputs shown in Figs. 3 and 4. The minimal differences between these two molecules make it easier to explain the algorithms, but do not sufficiently show the utility of the program described in this paper. To this purpose, it should be underlined that the manual analysis of all the possible ways to detect fragment differences is very difficult when the homology degree is lower. In fact, all the procedures checking the substitution, deletion, and/or insertion of residues representing cleavage sites must be also considered. To provide an additional example of the program utility, we used simulated data of the 0.39 inhibitor sequence analysis corresponding to the introduction of nine further randomized differences in its primary structure. The sequence of no 0.28 reference fragment could be perfectly reconstructed using these data, although the homology degree between the two proteins remains high (90.2%). Each uncompleted 0.39 matching fragment generated a great number of compatible peptides, since all the correct searching ways were considered. The global output of the program provided the homology degree and molecular weight information of 873 fragments of inhibitor 0.39 whose presence in the mixture was possible. Those exhibiting the highest homology degree with the 0.28 reference sequences were reported at the top of the output list, suggesting the most probable differences (not shown). This complex case demonstrated that a computer is necessary to identify the real fragments and resolve doubts by a mass analysis.

The validity of this method is based on the knowledge of the primary structure of a reference protein. Fur-

thermore, the homology degree of the two proteins should be quite high. In fact, although no calculation limit exists in comparing sequence analyses, the complexity of the output is linked to the number of possible peptides of the protein of unknown sequence which are compatible with the reference fragments. This strictly depends on the homology degree between the two proteins. Of course, an enormous output deriving from poorly related proteins and showing a great number of compatible peptides of low homology degree cannot provide useful information, even if the molecular weights were calculated. Moreover, the proportions of the output are also due to the complexity of the mixtures, related to the size of the proteins and to the specificity of the hydrolytic agent utilized. In fact, the sequence analyses of few fragments yield few residues at each step, which can be easily compared. For this reason, the usefulness of the method can be evaluated case by case and not established in advance. Anyway, good results should be obtained with proteins of up to 150,000 daltons, with a homology degree greater than 75% and hydrolyzed with specific agents. Computer programs implementing different strategies should be used when the sequence to be determined is completely unknown or the homology degree is quite low (Fairwell *et al.*, 1970; Cannon and Lovins, 1972; Biemann, 1980; Matsuo *et al.*, 1981; Kitagishy *et al.*, 1981, 1982; Shimonishi *et al.*, 1981; Herlihy and Biemann, 1981; Erickson and Jardine, 1986; Petrilli *et al.*, 1991; Johnson and Walsh, 1992; Caporale *et al.*, 1993, 1994; Fernandez-de-Cossio *et al.*, 1995).

Finally, the choice of this approach instead of cDNA sequencing depends on the availability of the source material. If the purification procedure of the mature proteins is known and simple, the present method should be preferable since it needs just two sequence and mass analyses; cDNA sequencing could be advantageous when suitable libraries are available. In any event cDNA sequencing does not provide information on posttranslational events, while the identification of modified residues by Edman degradation just depends on the availability of the corresponding PTH residues utilized as standards in the sequence analyses. If such products are not available, the corresponding PTH derivatives of the two proteins can be compared as "unknown" residues showing the same retention time in the on-line HPLC chromatograms.

In conclusion, this computerized method should enhance the combined advantages of using strategies based on mass and sequence analyses of peptide mixtures to quickly assess protein structure, since it is a powerful tool for interpreting experimental data. The program is freeware and available from the authors.

ACKNOWLEDGMENTS

Research supported by MIRAAF, Piano Nazionale Biotecnologie Vegetali, Prog. N. 152.

REFERENCES

Biemann, K. (1980). In *Biochemical Applications of Mass Spectrometry* (Waller, G. R. ed.), Wiley, New York, pp. 469–525.
Cannon, L. E., and Lovins, R. E. (1972). *Anal. Biochem.* **46**, 33–44.
Caporale, C., Carrano, L., Nitti, G., Poerio, E., Pucci P., and Buonocore V. (1991). *Protein Seq. Data Anal.* **4**, 3–8.
Caporale, C., Caruso, C., Petrilli, P., Sepe, C., Poerio, E., and Buonocore, V. (1993). *Protein Seq. Data Anal.* **5**, 337–344.
Caporale, C., Sepe, C., Caruso, C., Petrilli, C., and Buonocore, V. (1994). *Comput. Appl. Biosci.* **10**, 489–494.
Caporale, C., Sepe, C., Caruso, C., Pucci, P., and Buonocore, V. (1996*a*). *FEBS Lett.* **393**, 241–247.
Caporale, C., Garzillo, A. M. V., Caruso, C., and Buonocore, V. (1996*b*). *Phytochemistry* **41**, 385–393.
Caporale, C., Sepe, C., Caruso, C., Garzillo, A. M. V., and Buonocore, V. (1996*c*). *Comput. Appl. Biosci.* **12**, 81–88.
Caporale, C., Caruso, C., and Buonocore, V. (1996*d*). *J. Protein Chem.* **15**, 405–412.
Cottrell, J. S. (1994). *Peptide Res.* **7**, 115–124.
Erickson, B. J., and Jardine, I. (1986). *Biomed. Environ. Mass. Spectrom.* **13**, 343–346.
Fairwell, T., Barnes, W. T., Richards, F. F., and Lovins, R. E. (1970). *Biochemistry* **9**, 2260–2267.
Fernandez-de-Cossio, J., Gonzalez, J., and Besada, V. (1995). *Comput. Appl. Biosci.* **11**, 427–434.
Griffin, P. R., MacCoss, M. J., Eng, J. K., Blevins, R. A., Aaronson, J. S., and Yates, J. R., III (1995). *Rapid Commun. Mass Spectrom.* **9**, 1546–1551.
Henzel, W. J., Billeci, T. M., Stults, J. T., Wong, S. C., Grimley, C., and Watanabe, C. (1993). *Proc. Natl. Acad. Sci. USA* **90**, 5011–5015.
Herlihy, W. C., and Biemann, K. (1981). *Biomed. Mass. Spectrom.* **8**, 70–77.
James, P., Quadroni, M., Carafoli, E., and Gonnet, G. (1993). *Biochem. Biophys. Res. Commun.* **195**, 58–64.
James, P., Quadroni, M., Carafoli, E., and Gonnet, G. (1994). *Protein Sci.* **3**, 1347–1350.
Johnson, R. S., and Walsh, K. A. (1992). *Protein Sci.* **1**, 1083–1091.
Kitagishy, T., Hong, Y., and Shimonishi, Y. (1981). *Int. J. Peptide Protein Res.* **17**, 436–443.
Kitagishy, T., Hong, Y., Takao, T., Aimoto, S., and Shimonishi, Y. (1982). *Bull. Chem. Soc. Jpn.* **55**, 575–580.
Mann, M., and Wilm, M. (1994). *Anal. Chem.* **66**, 4390–4399.
Mann, M., Hojrup, P., and Roepstorff, P. (1993). *Biol. Mass Spectrom.* **22**, 338–345.
Matsuo, T., Matsuda, H., and Katakuse, I. (1981). *Biomed. Mass. Spectrom.* **8**, 137–143.
Mortz, E., O'Conrior, P. B., Roepstorff, P., Kelleher, N. L., Wood, T. D., McLafferty, F. W., and Mann, M. (1996). *Proc. Natl. Acad. Sci. USA* **93**, 8264–8267.
Pappin, D., Hojrup, P., and Bleasby, A. (1993). *Curr. Biol.* **3**, 327–332.
Patterson, S. D., and Aebersold, R. (1995). *Electrophoresis* **16**, 1791–1814.

Petrilli, P., Sepe, C., and Pucci, P. (1991). *Biol. Mass. Spectrom.* **20**, 115–120.
Petrilli, P., Caporale, C., and Caruso, C. (1994). *Int. J. Pept. Protein Res.* **43**, 201–204.
Shimonishi, Y., Hong, Y., Katakuse, I., and Hara, S. (1981). *Bull. Chem. Soc. Jpn.* **54**, 3069–3075.
Yates, J. R. (1996). *Meth. Enzymol.* **271**, 351–377.
Yates, J. R., Speicher, S., Griffin, P. R., and Hunkapiller, T. (1993). *Anal. Biochem.* **214**, 397–408.
Yates, J. R., III, Eng, J. K., McCormack A. L., and Schieltz, D. (1995). *Anal. Chem.* **67**, 1426–1436.