ELSEVIER

Journal of Chromatography A, 771 (1997) 1–7

JOURNAL OF
CHROMATOGRAPHY A

Application of sequential paired covariance to liquid chromatography–mass spectrometry data
Enhancements in both the signal-to-noise ratio and the resolution of analyte peaks in the chromatogram

David C. Muddiman, Baoming M. Huang, Gordon A. Anderson, Alan Rockwood, Steven A. Hofstadler, Mary S. Weir-Lipton, Andrew Proctor, Qinyuan Wu, Richard D. Smith*

Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA 99352, USA

Received 7 November 1996; revised 7 November 1996; accepted 13 January 1997

Abstract

The algorithm of sequential paired covariance (SPC) has been previously reported to dramatically enhance the signal-to-noise (*S/N*) ratio for on-line separations combined with mass spectrometry. That initial study focused on a limited number of data sets derived from the combination of capillary electrophoresis (CE) with time-of-flight mass spectrometry using an electrospray interface. Results from the initial study clearly demonstrated that a significant enhancement (almost two orders of magnitude) in the *S/N* ratio of the eluting peaks in the electropherogram could be obtained, facilitating identification of the analytes. In this report, the algorithm has been applied to liquid chromatography–mass spectrometry data obtained on a triple quadrupole instrument and we have evaluated the general applicability of the SPC approach to several types of microcolumn separations with mass spectrometric detection, including CE coupled with Fourier transform ion cyclotron resonance mass spectrometry. In all the cases we tested, we found the algorithm enhanced the *S/N* ratios of the resulting chromatograms or electropherograms to a similar extent. This report further demonstrates the SPC approach to enhance the resolution as well as the *S/N* ratio of the eluting peaks of a complex peptide mixture. While many variations of the algorithm are possible, we have also found higher order covariance (e.g., 3rd order) is useful for eliminating coincidental noise in sequential mass spectra, giving the potential to extract broad, low intensity analyte peaks. We also demonstrate the sequential covariance approach for enhancing the *S/N* ratio of mass spectra.

Keywords: Sequential paired covariance; Chemometrics; Signal-to-noise ratio; Resolution

1. Introduction

The combination of electrospray mass spectrometry (ESI-MS) with various separation techniques has proven to be a powerful tool for solving diverse bioanalytical problems [1–5]. One key step in the analysis of data from on-line liquid chromatography (LC) or capillary electrophoresis (CE) with MS is the effective reconstruction of chromatograms or electropherograms from a series of mass spectra, in order to identify analyte peaks in the chromatogram or electropherogram. The ability to locate a

*Corresponding author.

0021-9673/97/$17.00 Copyright © 1997 Elsevier Science B.V. All rights reserved
PII S0021-9673(97)00069-1

peak in a chromatographic trace is especially important for the analysis of unknown mixtures or mixtures having unknown elution times and containing analytes over wide concentration ranges. This reconstruction is usually accomplished by integrating parts or all of each mass spectrum to obtain a plot of ion current or the total ion current (TIC) vs. time (or scan number). In general, this reconstruction method has found widespread use and it should be noted that the TIC trace retains quantitative information (e.g., retention time, relative ion intensity) present in the mass spectra. However, when the signal-to-noise (S/N) ratio is poor, background noise could prevent detection of peaks in the chromatogram or electropherogram reconstructed in this manner.

The quality of a given data set can be improved by applying a variety of methods including spectral smoothing, deglitching, and background subtraction [6,7]. In the case where only a few spectra need to be processed, these are viable methods whether used individually or in combination. However, for high-performance separations coupled with MS, the generation of large data sets (>500 spectra) is not uncommon. In addition, it is often necessary to monitor the data in 'real time' (e.g., quality control). This precludes routine examination of each mass spectrum which often restricts user intervention to eliminate noise and/or reset the background. Thus, it is evident that the individual mass spectra constituting a data set are not initially examined (i.e., only after the peaks are identified in the chromatogram are the mass spectra studied). The low S/N ratio and/or poor resolution of analyte peaks in a chromatogram or electropherogram may prevent identification of a component(s) present, even though the mass spectrum from a particular chromatographic peak may be usable. Thus, approaches that enhance the S/N ratio and resolution of analyte peaks in chromatograms or electropherograms in high-performance separations coupled with MS, are of practical utility.

For this purpose, we have recently developed a reconstruction algorithm which utilizes the similarity between sequential mass spectra [8,9]. In general, the sequential paired covariance (SPC) algorithm generates a series of 'virtual geometrically amplified' mass spectra. Each data point in each of the virtual mass spectra represents the dot product of two corresponding data points in sequential original spectra. The electropherogram or chromatogram is obtained by plotting the integral of ion current from each virtual mass spectrum vs. time (or scan number).

In the initial report, the SPC algorithm was successfully applied to CE–ESI time-of-flight MS data, and a significant enhancement in the S/N ratio of the analyte peaks in the SPC electropherogram was obtained [8]. In this communication, we report the application of the SPC algorithm to LC–MS data to further demonstrate the enhancement in the resolution as well as the S/N ratio in the chromatogram using the SPC approach. The general applicability of the algorithm has been tested with a wide variety of data formats (i.e., different instrumentation, including CE–ESI Fourier transform ion cyclotron resonance MS), and all of the data we examined (>30 data sets) resulted in nearly identical improvements. Thus, in order to avoid redundancy, a single data set obtained from an LC–MS run on a triple quadrupole instrument will be presented due to its complexity and ability to demonstrate the important features of the SPC approach. Finally, we introduce variations of the algorithm and discuss their importance.

2. Experimental

The LC–MS data were obtained using a microbore HPLC (Applied Biosystems, San Francisco, CA) with a 250 mm×1 mm I.D. C_{18} column (Phenomenex, Torrance, CA, USA) interfaced to a Finnigan TSQ 7000 triple quadrupole mass spectrometer (Finnigan MAT, San Jose, CA, USA) via an electrospray interface. A peptide digest solution was eluted using a linear gradient of 25–50% acetonitrile and 0.1% formic acid in 50 min at a flow-rate of 40 μl/min. Dry nitrogen (30 p.s.i.; 1 p.s.i.=6894.76 Pa) was used as co-axial sheath gas. More details about the preparation of the protein digest and LC–MS analysis can be found elsewhere [10–13].

Data analysis was conducted using a program developed at Pacific Northwest National Laboratory [14]. Profile data files from the Finnigan TSQ 7000 are processed using a command line program written in Borland C/C++ version 4.5. Profile data files contain a variable number of spectra and each spectrum contains a variable number of data points, depending on the number of peaks in the spectrum.

Each spectrum is defined by an array of X, Y (m/z, amplitude) data pairs. The SPC algorithm requires data arrays from two, or more, spectra to be multiplied. If the data sets contains the same X scale, and the same number of data points, this would simply require multiplying each Y value in spectrum one by the corresponding Y value in spectrum two, and so on. Since profile data files have a variable X axis, a different approach is required. When two spectra are multiplied, the Y value corresponding to an X axis value from the first spectrum is multiplied by the Y value corresponding to the same X axis value in the second spectrum. This particular X value may not exist, so a cubic spline interpolation is used to predict it. The command line program places all results in an ASCII data file. This file can be imported into a spread sheet program, such as Microsoft Excel, for data presentation.

A user interface to the command line program was developed using Microsoft Excel spread sheet application. Excel uses Visual Basic for Applications (VBA) as a macro language. VBA supports command buttons, text boxes, and many other user interface objects. This interface allows the user to request an operation by filling in a dialog box and pressing the start button. The Excel program will then call the command line application with the correct parameters for the requested operation. After the results are calculated, another button is pressed to graphically display the results. The program runs on a 75 MHz PC, and requires only a few seconds to calculate a SPC reconstructed chromatogram, and can be implemented on-line with nominal additional programming effort.

3. Results and discussion

3.1. General characteristics of the SPC algorithm

Fig. 1 provides a graphical representation that describes the fundamentals of the SPC algorithm. Two adjacent spectra are used (n and $n+1$) with the solid bars representing analyte signals, such as multiple charge states of the same molecular ion, and the dashed bars representing noise. In SPC, each data point in one mass spectrum is multiplied with the corresponding data point in the following mass spectrum, resulting in a geometrically amplified spectrum; the number of spectra used in each multiplication operation defines the order of the covariance algorithm. As demonstrated in Fig. 1, the SPC operation results in a virtual mass spectrum that amplifies the similarity (the signal) and suppresses or

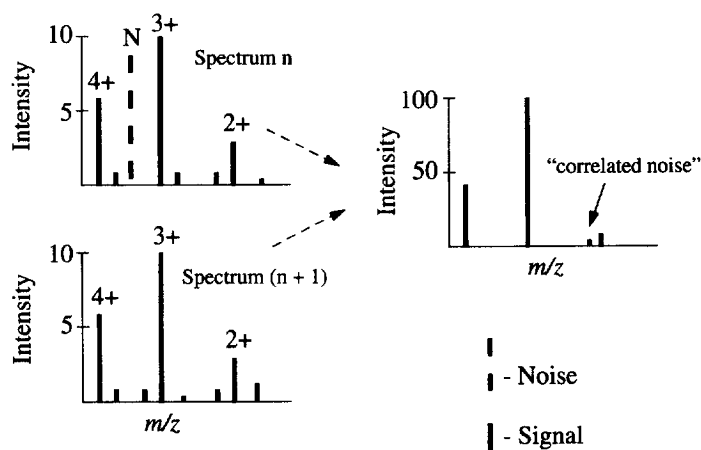

Fig. 1. Graphical representation showing the general characteristics of the sequential paired covariance (SPC) algorithm. Each data point in each original mass spectrum is multiplied with each corresponding data point in the following original mass spectrum, generating a series of geometrically amplified virtual mass spectra. The total ion intensities of individual virtual mass spectra are then plotted vs. scan numbers to reconstruct the chromatogram or the electropherogram. Between (or among) sequential mass spectra, the similarity (the signal) is amplified, and the dissimilarity (the noise) is suppressed or eliminated, and enhancements in both the S/N ratio and the resolution of analyte peaks in the chromatogram or the electropherogram are achieved.

eliminates the dissimilarity (the noise) between the original adjacent mass spectra. Specifically, the noise peak N in Fig. 1, which is of similar intensity as the 3+ charge state signal in spectrum n, is eliminated in the dot product mass spectrum because at that particular x-value, spectrum $n+1$ had no amplitude (i.e., the y-value in spectrum n was essentially multiplied by zero). The intensity in the reconstructed chromatogram is obtained by integration of each geometrically amplified mass spectrum and plotted vs. the scan number (i.e., time). Thus, dramatic enhancements of the S/N ratio and the resolution in the chromatogram are achieved using the SPC algorithm by amplifying the signal (similarity between adjacent mass spectra) and suppressing the noise (dissimilarity between adjacent mass spectra). It is worth noting that the S/N ratio and the resolution enhancements are achieved only when an analyte peak is present in n or more consecutive original spectra (where n is the order of the covariance algorithm as described above). Fig. 1 also illustrates how the relative intensities are distorted. For example, the 3+ charge state intensity changed from 10 in the original two spectra to 100 in the amplified mass spectrum while the 4+ charge-state increased from 6 to 36. Thus, the intensity ratio (3+/4+) has changed from 1.6 in the original spectra to 2.8 in the geometrically amplified spectrum.

3.2. TIC vs. SPC chromatogram reconstruction methods

Fig. 2A shows the TIC chromatogram of a tryptic digest obtained from an LC–MS analysis. This chromatogram is representative of a typical protein digest and thus serves to highlight several advantages of the SPC approach. Note that there are two relatively sharp peaks which elute first (peaks a and b) followed by a broad, unresolved peak (peak c) which contributes a significant portion of the TIC (based on peak area measurements). The SPC algorithm was used to generate a chromatogram from the same data set and is shown in Fig. 2B. Comparison of Fig. 2A with Fig. 2B reveals the two main attributes that signify the SPC algorithm: (1) enhancement of the S/N ratio of analyte peaks in the chromatogram; and (2) enhancement of the resolution of analyte peaks in the chromatogram. First, it is clear that the signal intensity of the eluting peaks has been enhanced by over 6 orders of magnitude. In addition, the noise has been reduced and a smoother baseline is observed. Thus, a significant enhancement in the S/N ratio of analyte peaks in the chromatogram is achieved with the SPC approach. This enhancement in S/N is helpful in identifying the retention time of weak analyte peaks, which would otherwise be buried in noise, to allow identification of the analyte via its mass spectrum.

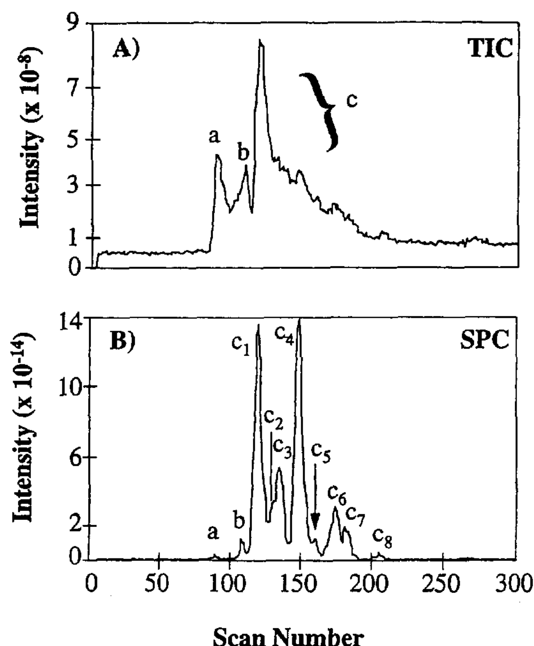

Fig. 2. Chromatograms reconstructed from the same LC–MS data set of a protein digestion products with: (A) the conventional total ion current (TIC) method; (B) the sequential paired covariance (n=2) approach. The quadrupole was scanned from m/z 105 to m/z 3000, and a portion (scans 1–300) of the data is shown. Enhancements in both the S/N ratio and resolution of analyte peaks are evident in Fig. Fig. 2B.

It was previously suggested that an effective enhancement in the chromatographic resolution should result from the application of the SPC algorithm [8]; however, this was not demonstrated in the initial work. It is evident that the unresolved peak (peak c) using the TIC reconstruction method in Fig. 2A, is resolved into at least 8 peaks (c_1–c_8) using the SPC method. This enhancement in effective resolution of analyte peaks allows one to rapidly access

the position (i.e., retention time) of a peak to retrieve its mass spectrum. All of the peaks labeled c_1–c_8 correspond to different analytes (based on their mass spectra) and three are shown in Fig. 3 with their corresponding peak labels (see Fig. 2). Without the enhancement of the resolution of analyte peaks in the chromatogram, it would be very difficult to locate a mass spectrum corresponding to a specific analyte in the region of peaks c_1–c_8.

3.3. Higher order covariance and sequential paired addition

The SPC algorithm can be applied in different ways depending on the application. One such variation involves higher order covariance; instead of geometrically amplifying two adjacent spectra, three or more sequential mass spectra are used. For example, in 3rd order covariance, spectra number 1, 2 and 3 are used to create the first virtual mass spectrum, which is then integrated; the second virtual mass spectrum is generated using spectra numbers 2, 3 and 4, and so on. The principle of the higher order sequential covariance (SC^n, with n representing the order) approach remains the same as that of the SPC (i.e., the similar features in multiple spectra are enhanced); however, in addition to further enhancement in the S/N ratio of the chromatogram, higher order covariance also reduces or eliminates 'accidental correlated noise' that may be present in two consecutive mass spectra but unlikely to be in three spectra. This can be understood because persistent noise may occur in two adjacent spectra, but the probability that a noise spike occurs in three sequential spectra is significantly lower. In the event that deterministic noise is present in each spectrum (e.g., 60 Hz noise) or if a buffer or solvent component gives rise to a reproducible peak in each spectrum, the background noise will also rise. This can be easily compensated for by eliminating discrete m/z values if the location of the peak is known, with the concomitant disadvantage of creating a 'blind spot' in the mass spectra.

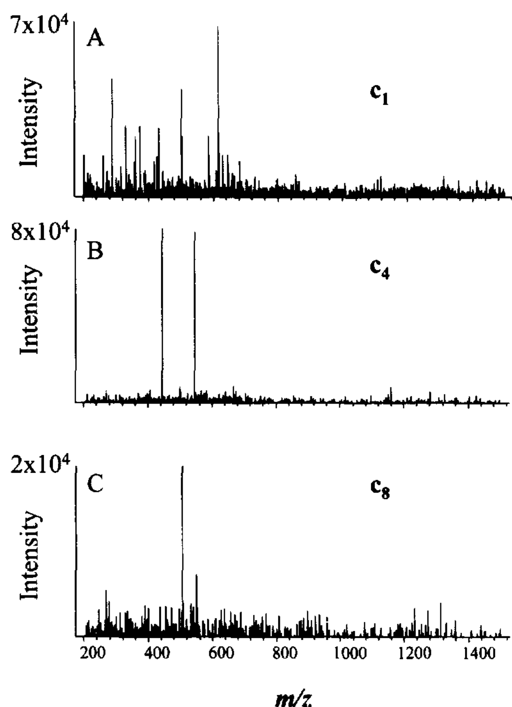

Fig. 3. Mass spectra of several analytes located in Fig. 2B, benefitted from the enhanced resolution and the enhanced S/N ratio of analyte peaks provided by the SPC algorithm. Fig. 3A–C correspond to mass spectra at the peaks of c1, c4 and c8, respectively. Locating these peaks would have been very difficult without the enhanced resolution and the enhanced S/N ratio of analyte peaks provided by the SPC approach.

In addition, higher order covariance also has the potential for extracting broad, weak analyte peaks from a 'noisy' background. For example, a peak that elutes from a column with a long retention time is typically broad as a result of diffusion. Although the peak area is very large (due to the width) the peak height may be very small and buried in the background noise, and further aggravated by factors such as baseline drifts in gradient elution profiles. Invoking a higher order (e.g., 10) should enhance the S/N ratio of such a peak in the chromatogram substantially.

Another variation of the SPC algorithm addresses the fact that the SPC approach lacks quantitative information (e.g., distortion in relative ion intensities) due to the multiplicative nature of the algorithm. We have evaluated a similar approach, named sequential addition (SA^n, with n representing the order) which adds (instead of multiplies) two or more sequential mass spectra, to obtain a virtual mass spectrum. Fig. 4 shows a chromatogram reconstructed from the same data shown in Fig. 2 using the SA^5 approach. Again, there are enhancements in

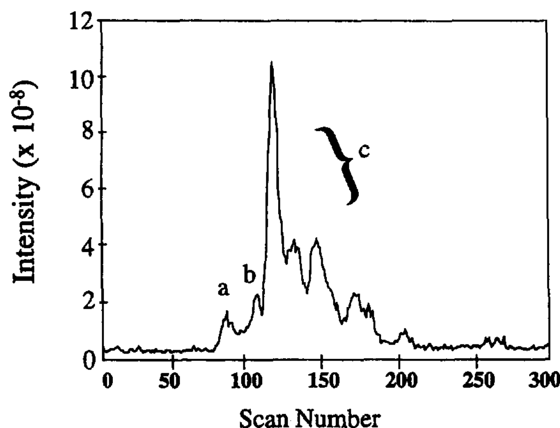

Fig. 4. Chromatogram resulting from the application of a 5th order sequential addition (SA^5) algorithm to the same data set shown in Fig. 2A. The SA approach preserves the quantitative information in the chromatogram while moderately increases the resolution of analyte peaks.

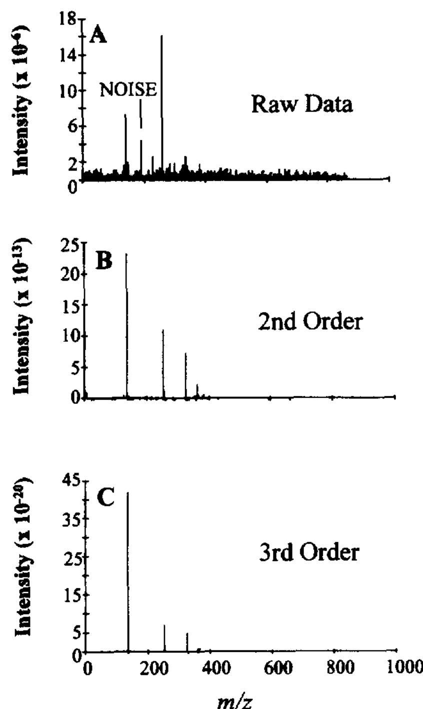

Fig. 5. Processing mass spectra with the sequential covariance (SC) algorithm. The SC approach facilitates signal recognition by providing enhancement in the S/N ratio of the mass spectra.

the S/N ratio and the resolution of analyte peaks (compared with the TIC approach; see Fig. 2A). These enhancements with the SA^5 approach are less dramatic, as compared with the SPC approach (Fig. 2B); however, relative intensities (and thus quantitative information) are preserved, and the resolution and the S/N ratio of analyte peaks in the chromatogram are moderately enhanced.

3.4. Application of the SPC algorithm for mass spectral processing

Since the analyte concentration is often decreased in liquid chromatography due to diffusion and other peak broadening phenomena, the S/N ratio of the mass spectra are often compromised in LC–MS (see Fig. 3). Signal averaging of a chromatographic peak is common practice in order to enhance the S/N ratio of the mass spectrum. When only qualitative information is desired (i.e., molecular mass determination), an additional facet of the SPC algorithm can be used to enhance the S/N ratio of mass spectra which has been previously demonstrated for matrix-assisted laser desorption/ionization MS [15]. Fig. 5A shows the mass spectrum obtained from scan number 140 (see Fig. 2). The S/N ratio is relatively poor and there appears to be an intense peak centered around m/z 195 as well as some additional peaks. Fig. 5B shows the result after geometrically amplifying scans numbers 140 and 141, and Fig. 5C shows the third order covariance of scans numbers 140, 141 and 142. The overall S/N ratio has been dramatically increased in both the 2nd and 3rd order amplifications as evidenced by the 14 order of magnitude difference in the y-scale of SC^n compared to the raw data (see Fig. 5). Interestingly, the apparent peak centered at m/z 195 in Fig. 5A is non-existent in Fig. 5B,C, and is attributed to noise.

4. Conclusions

The general applicability of the SPC algorithm for the effective reconstruction of chromatograms in LC–MS or electropherograms in CE–MS has been demonstrated, with both the example LC–MS data set discussed in this paper and the tests we performed on other instruments including CE–ESI Fourier transform ion cyclotron resonance MS. En-

hancements in both the S/N ratios and the resolution of analyte peaks in the chromatogram or the electropherogram are routinely achieved. In addition, variations of the SPC algorithm, such as higher order covariance ($n>2$), sequential addition, and geometrically amplified mass spectral processing, lead to elimination of artifact peaks and extraction of broad peaks, preservation of quantitative information while enhancing analyte resolution, and mass spectral peak recognition in noisy mass spectra, respectively. The intention of developing the SPC algorithm is not to replace the widely used conventional TIC method; rather, it has been developed to compliment TIC for particular applications where conventional TIC methods are inadequate. With additional programming effort, it is possible to implement both SPC (or SA) and TIC on line simultaneously (with two different windows or traces), and the comparison of the two traces should help to meet the demand of complex mixture analysis.

Acknowledgments

We thank the US Department of Energy, Office of Health and Environmental Research, for support of this research. Pacific Northwest National Laboratory is operated by Battelle Memorial Institute for the US Department of Energy through contract No. DE-AC06-76RLO 1830.

References

[1] L.J. Deterding, C.E. Parker, J.R. Perkins, M.A. Moseley, J.W. Jorgenson, K.B. Tomer, J. Chromatogr. 554 (1991) 329.
[2] T. Wachs, J.C. Conboy, F. Garcia, J.D. Henion, J. Chromatogr. Sci. 29 (1991) 357.
[3] J. Vandergreef, W.M.A. Niessen, Int. J. Mass Spectrom. Ion Proc. 118 (1992) 857.
[4] J.Y. Cai, J. Henion, J. Chromatogr. A 703 (1995) 667.
[5] R.D. Smith, J.H. Wahl, D.R. Goodlett, S.A. Hofstadler, Anal. Chem. 65 (1993) A574.
[6] E.R. Malinowski and D.G. Howery, Factor Analysis in Chemistry, Wiley, New York, 1980.
[7] G. Horlick and G.M. Hieftje, Contemporary Topics in Analytical and Clinical Chemistry, Plenum Press, New York, 1978.
[8] D.C. Muddiman, A.L. Rockwood, Q. Gao, J.C. Severs, H.R. Udseth, R.D. Smith, A. Proctor, Anal. Chem. 67 (1995) 4371.
[9] B.M. Huang, D.C. Muddiman, G.A. Anderson, A.L. Rockwood, R.D. Smith and A. Proctor, Presented at the ASMS conference on Mass Spectrometry and Allied Topics, Portland, OR, 1996.
[10] J. Houmard, G.R. Drapeau, Proc. Natl. Acad. Sci. USA 69 (1972) 3506.
[11] G.R. Drapeau, Can. J. Biochem. 56 (1978) 534.
[12] M.S.W. Lipton, D.L. Springer and C.G. Edmonds, Anal. Biochem., in preparation (1995).
[13] M.S. Weir, D.L. Springer, B.D. Thrall, A.F. Fuciarelli and C.G. Edmonds, in J.W. Crabb, (Editor), Techniques in Protein Chemistry VI 1994, p. 115.
[14] D.C. Muddiman and G.A. Anderson, Sequential Covariance v1.0 PNNL, copyright 1996, all rights reserved.
[15] R.W. Nelson and M.L. Vestal, Presente at the ASMS conference on Mass Spectrometry and Allied Topics, San Francisco, CA, May 31–June 3, 1993.