ELSEVIER

Journal of Chromatography A, 743 (1996) 91–98

JOURNAL OF CHROMATOGRAPHY A

Automated proteolytic mapping of proteins

Tim Nadler*, Christopher Blackburn, Jonathan Mark, Neal Gordon, Fred E. Regnier, George Vella

PerSeptive Biosystems Inc., 500 Old Connecticut Path, Framingham, MA 01701, USA

Abstract

A new, rapid, automated method for peptide mapping has been developed that requires less than two hours to complete. This method first (i) reduces the protein with dithiothreitol (DTT) or β-mercaptoethanol at 50°C, (ii) then alkylates it with an alkylating agent selected from iodoacetamide, iodoacetic acid, or vinylpyridine, (iii) digests the protein completely with immobilized, TPCK treated trypsin, and (iv) finally analyzes the tryptic fragments by high resolution, reversed-phase liquid chromatography (RPLC) in less than two hours. Reduction and alkylation are achieved in the autosampler of the instrument where the sample, reagents, and reaction protocol are specified by the operator in the system computer. Proteins with up to seven disulfide bridges were quantitatively reduced and alkylated by the system. Immobilized enzyme columns coupled in tandem with an RPLC column were shown to generate protein digests and reproducibly separate the fragments for many cycles of analysis. Based on the fact that any one of several alkylating agents could be used in the mapping process, it was demonstrated that a campaign of experiments could be executed automatically in a search for the optimum alkylating agent. The mapping technique was applied to five different proteins.

Keywords: Peptide mapping; Proteins

1. Introduction

Current analytical technology does not allow direct determination of protein primary structure. Fragmentation of proteins with proteolytic enzymes to produce smaller peptide fragments must be used to reduce the complexity of polypeptides to the extent that they are amenable to structure elucidation by Edman based sequencing technology. Although this rigorous approach is the most definitive method available for elucidating protein structure, it is time consuming and not always necessary, particularly in the case of proteins produced by recombinant DNA technology.

Proteins are widely synthesized today by inserting foreign DNA into a cell and inducing the cell to synthesize protein(s) for which that DNA codes. When the sequence of the coding DNA is known, one theoretically knows the protein sequence. There are however, cases in which the primary structure of the protein predicted from DNA sequence is different than that of the protein isolated from the cell. This can either be due to in vivo "errors" prior to or during biosynthesis or in vitro alterations during purification. Genetic drift during proliferation of the cell population can alter the DNA sequence coding for the protein causing incorporation of different amino acids at some positions [1,2]. Selectivity errors during transcription and translation can also cause aberrations in the primary structure of a protein during biosynthesis. Post-translational modification is yet another way that polypeptides are modified in vivo. Primary structure variants which arise in vitro, i.e. during purification, are the result of

*Corresponding author.

0021-9673/96/$15.00
PII S0021-9673(96)00359-7

post-synthetic modifications [3–6]. Proteolysis, methionine oxidation, deamidation, carbamylation, and hydrolysis are some examples of in vitro modification [7].

The analytical problem, when recombinant DNA technology is used to produce therapeutic proteins, is not to elucidate the structure of each lot of protein manufactured, it is rather to confirm that each lot of protein produced is the same as the structure of the protein used to determine therapeutic efficacy. There can be very large differences in the methodology used to *elucidate* the primary structure of a protein and the methodology to *confirm* that the primary structure of two lots of protein is the same. Primary structure elucidation, as described above, is lengthy, tedious, and costly. It is for this reason that Edman based sequencing is seldom used to confirm structures in the biotechnology industry. Ideally, methods to confirm structural equivalence would be rapid and automated.

High-performance reversed-phase liquid chromatography (RPLC) or coupled RPLC–mass spectrometry of proteolytic digests, generally from trypsin, are now well established with regulatory agencies [8,9] as methods to confirm primary structure equivalence of proteins. This is based on the fact that single amino acid substitutions or chemical alterations in retained tryptic peptides alter both their chromatographic behavior and mass. Any variation in the RP chromatogram, or "proteolytic map" as it has come to be known, of a proteolyzed sample of the target protein is an indicator that one or more amino acid substitutions or alterations has occurred.

Although much faster than Edman sequencing, RPLC mapping of protein fragments still requires proteolysis prior to the chromatographic separation. Although it has been shown that proteolytic fragments may be generated automatically by eluting proteins through beds of immobilized proteolytic enzyme [10–16], proteolysis of proteins must generally be preceded by disulfide reduction and alkylation of the resulting sulfhydryl groups to (i) allow proteolytic enzymes to gain easy access to the interior of the protein, (ii) increase solubility of the denatured protein, and (iii) preclude reformation of disulfide bridges during enzymatic digestion. This is a lengthy, labor intensive process requiring multiple manual manipulations over the course of a day or more. The objective of the work described in this paper was to automate this process and reduce analysis time at least 10 fold.

2. Experimental

2.1. Reagents

Lysozyme, bovine insulin, horse heart cytochrome *c*, ribonuclease A, human serum albumin, dithiothreitol (DTT), iodoacetamide, iodoacetic acid, vinylpyridine, tris(hydroxymethyl)aminomethane (Tris), calcium chloride and urea were obtained from Sigma (St. Louis, MO). Poroszyme Trypsin and Poroszyme Glu-C (both 30 mm×2.1 mm) as well as the PepMap C_{18} reversed-phase columns (250 mm×4.6 mm) were obtained from PerSeptive Biosystems (Framingham, MA). Trifluoroacetic acid (TFA) was purchased from Pierce (Rockford, IL) and acetonitrile from EM Science (Gibbstown, NJ).

2.2. Equipment

Analyses were automated on an Integral Micro-Analytical Workstation (PerSeptive Biosystems) in a tandem column configuration fitted with a built-in autosampler. Detection was achieved by monitoring absorbance at 214 nm. This system has an adjustable gradient delay volume which was plumbed to be about 350 μl counting the 100 μl injection loop and auxiliary mixer.

2.3. Reduction

Protein reduction was achieved in the autosampler with dithiothreitol (DTT) in reaction vessels consisting of a standard 12 mm×32 mm glass vials. The data system was instructed to withdraw aliquots of both the protein sample and reducing agent from their respective sample vials in the autosampler and inject them into the reduction vessel vial. The order of addition is not important. The final concentration of reducing agent was 4.5 m*M* and the final protein concentration was about 150 μM (about 2 mg/ml lysozyme). An aliquot of 8 *M* urea was also added to aid protein solubility. The mixture was then vortexed

by the autosampler in the reaction vessel and finally incubated at 50°C for 15 min.

2.4. Alkylation

A sample of the reduced protein was then automatically withdrawn by the autosampler and added to an empty sample vial which served as the alkylation vessel. Based on the volume of reduced protein, aliquots of alkylating agent, consisting of either iodoacetamide, iodoacetic acid, or vinylpyridine, and urea were then added to the alkylation vessel to final concentrations of 10 m*M* and 2.5 *M* respectively. The protein concentration was about 100 μM (about 1.4 mg/ml lysozyme) after this step. Subsequent to mixing by vortexing in the autosampler the mixture was allowed to react for approximately 5 min at room temperature.

An optional quenching step was added to the procedure at this point to deal with the possibility that excess alkylating reagent could either over alkylate the protein or react with proteolytic enzymes during proteolysis. A stoichiometric amount of the quenching agent cysteine or thiomalic acid was added to the alkylation mixture at the end of the alkylation reaction. Due to the way the samples were transferred to fresh vials as new reagents are added, the recovery of the initial sample was 50%. It is important to note that the recovery can be improved to 80% simply by switching to conical micro-vials for sample handling.

2.5. Proteolytic digestion

The reduced and alkylated protein sample was then injected into the Poroszyme cartridge for digestion. The length of digestion was controlled by adjusting the residence time in the Poroszyme cartridge with the flow-rate. Slower flow-rates led to longer digestion times. About 1–5 nmol of protein was digested on the Poroszyme cartridge in order to obtain good signal (peaks about 0.1 AU at 214 nm on a 4.6 mm I.D. reversed-phase column).

2.6. Reversed-phase chromatography

As the digested protein exited the Poroszyme cartridge, it was directed to the reversed-phase column (PepMap C_{18}) for analysis. Once the complete sample has been captured on the reversed-phase column, both columns were switched off-line and the workstation was purged with the reversed-phase solvents, 0.1% TFA in water and 80:20 (v/v) acetonitrile–water) with 0.1% TFA. Once the system was equilibrated with 0.1% TFA, only the reversed-phase column was switched back in-line to be eluted with a gradient of increasing acetonitrile concentration. Linear gradient elution was achieved in a solvent program ranging from 0.1% TFA (solvent A) to 0.1% TFA in 80% acetonitrile (solvent B) during the course of a fixed time, generally 60 min.

Reduced and alkylated peptides were also collected for mass spectral analysis. Peptides were either collected manually as their elution was indicated by the absorbance detector or collected automatically using the peak cutting mode of the software. In the peak cutting mode, the operator must define absorbance thresholds and slope values that will define the peaks which are to be collected. This method of collection is superior to simple time based collection because the thresholds can be set such that only the most concentrated portion of the peak will be collected. In most cases, the fractions were submitted directly for matrix assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF MS) analysis without further concentration.

2.7. Mass spectrometry

The peptides and proteins collected from the reversed phase column were analyzed by MALDI-TOF MS on a Voyager Biospectrometry Workstation (PerSeptive Biosystems), using alpha-cyano-4-hydroxy cinnamic acid (Sigma) as the matrix.

3. Results and discussion

3.1. General protocol

As noted above, tryptic mapping involves at least four steps; (i) reduction, (ii) alkylation, (iii) proteolysis, and (iv) RPLC. In the case where mass spectrometry is used in detection, there is a fifth step. Automation of reduction and alkylation was achieved

by using the autosampler of the liquid chromatograph (LC). [A schematic diagram of the analytical system is shown in Fig. 1.] An aliquot of the sample was withdrawn from a particular, computer specified sample vial and injected into a specified reaction vial with reducing agent, such as dithiothreitol (DTT) or β-mercaptoethanol, which had been obtained by the autosampler from a specific reducing agent vial. The position of all sample vials, reagent vials, reaction vials, and the sequence of events involved in carrying out specific reactions must be prescribed in the system computer by the operator. Following reduction, the autosampler was then used to add an aliquot of the reduced protein to a new empty vial with alkylating agent. In cases where excess reagents would either continue to react with the protein as it sits in the autosampler or destroy the proteolytic enzyme used to digest the protein, a quenching agent was added.

Reduction, alkylation, and quenching were achieved in parallel with the tandem column proteolysis-reversed-phase chromatography steps. The reactions may be performed in the autosampler while the reversed-phase column is being equilibrated with the aqueous starting solvent. For this reason, chemical reactions carried out in the autosampler did not add to the analysis time.

Following reduction and alkylation in the autosampler, all further steps of structure confirmation were achieved within the fluidics of a multi-valved liquid chromatograph fitted with an immobilized proteolytic enzyme column and a RPLC column. Proteolysis was achieved by withdrawing a sample of the reduced and alkylated protein from the reaction vial in the autosampler and introducing it into the LC where it was swept through the immobilized enzyme column. Reaction time was controlled by flow-rate and could be varied from a 1–600 s.

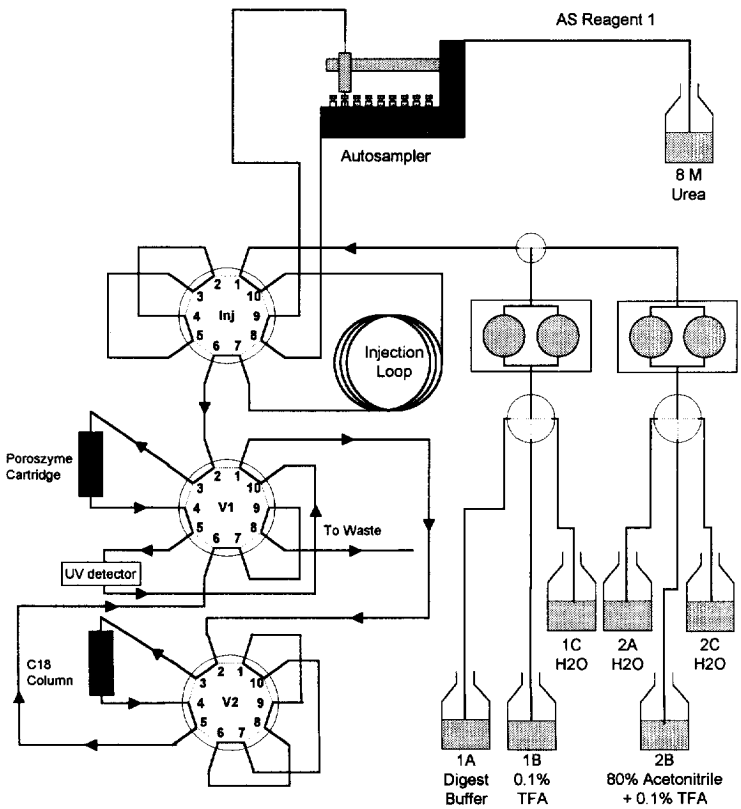

Fig. 1. A schematic diagram of the system used to automate reduction, alkylation, proteolysis, and reversed-phase chromatography of proteins.

Proteolytic fragments of the protein eluting from the enzyme column were directly transported to the RPLC column where they were concentrated. RPLC was achieved with an octadecylsilane derivatized, silica based column. At the completion of proteolysis, the RPLC column was gradient eluted to produce what has come to be known over the last decade as a "proteolytic map" of the protein.

This automated proteolytic mapping protocol was applied to five proteins during the course of this study; lysozyme, bovine insulin, cytochrome *c*, ribonuclease A and human serum albumin. Reproducible maps were produced in all cases.

3.2. Reductive alkylation

The individual steps of disulfide reduction and alkylation were carried out in the autosampler using lysozyme as a model protein. Following addition of 8 *M* urea, the sample protein was reduced by treatment with excess (4.5 m*M*) dithiothreitol (DTT) in the autosampler mixing port at 50°C for 15 min. [Addition of 8 *M* urea is optional, depending on the need to increase the solubility of reduced protein.] The reduced protein was then alkylated by treatment with either, iodoacetic acid, iodoacetamide, or 4-vinylpyridine at a concentration of 10 m*M* alkylating agent and ambient temperature for 5 min. As noted above, excess alkylating agent can be quenched to prevent overreaction with the protein or destruction of proteolytic enzymes. This was achieved by adding a quenching agent, such as cysteine or thiomalic acid, to the alkylation mixture at the end of the alkylation reaction. In the case of immobilized trypsin, excess alkylating agent was not observed to pose a problem and the quenching step was eliminated.

Complete alkylation of lysozyme was demonstrated by isolating the modified protein by reversed-phase chromatography from three separate alkylation reactions. A single protein peak was observed for each of the three reactions. The proteins were then analyzed by MALDI-TOF MS and shown in each case to be fully reduced and alkylated (i.e. the molecular weight corresponded to having eight alkylated thiols). As a control the reactions were also performed at low temperature and in this case, mixtures of native and fully alkylated lysozyme were observed (data not shown).

3.3. Proteolysis

This automated mapping method has been examined with immobilized enzyme columns of trypsin, Glu-C (*Staph. aureous* V8 protease), pepsin, and papain, all of which give reproducible proteolysis in the flow-through digestion format. The data presented here will concentrate on trypsin and Glu-C.

Proteolysis may be achieved by passing the protein through the immobilized enzyme column under constant flow at room temperature or elevated temperatures [5]. In a reproducibility study, the enzyme column was connected in tandem with the RPLC column such that eluted proteolytic fragments were concentrated on the RPLC column. The enzyme column was held at 37°C during the reaction and valved out of the flow train prior to elution of the RPLC column with a mobile-phase gradient ranging from 0.1% TFA to 0.1% TFA in acetonitrile. A series of lysozyme maps obtained using the immobilized trypsin column at a digestion flow-rate of 0.5 ml/min shows (Fig. 2) that the method is highly reproducible. Each of these chromatograms was obtained by running the entire method, i.e. sampling, reduction, alkylation, proteolysis, and RPLC. Similar results were obtained using three different immobilized enzyme columns prepared from the same lot of immobilized enzyme (data not shown).

A reaction time study was conducted by varying the flow-rate of oxidized bovine insulin B chain through an immobilized Glu-C column from 0.05 ml/min to 0.5 ml/min (Fig. 3). Peptides eluting from the RPLC column were manually collected and analyzed by MALDI-TOF. Based on the fact that the sequence of insulin is known, it is possible to assign a sequence to each of the fragments based on the observed mass (Fig. 3). The peptide map varied with flow-rate, but in all cases was reproducible at a given flow-rate. All peptides obtained from the 0.05 ml/min reactor flow-rate were the result of total Glu-C digestion based on MALDI-TOF MS. In contrast, at higher flow-rate, i.e. shorter reaction time, an additional peak was observed at an intermediate retention time on the RPLC chromatogram. The mass of this

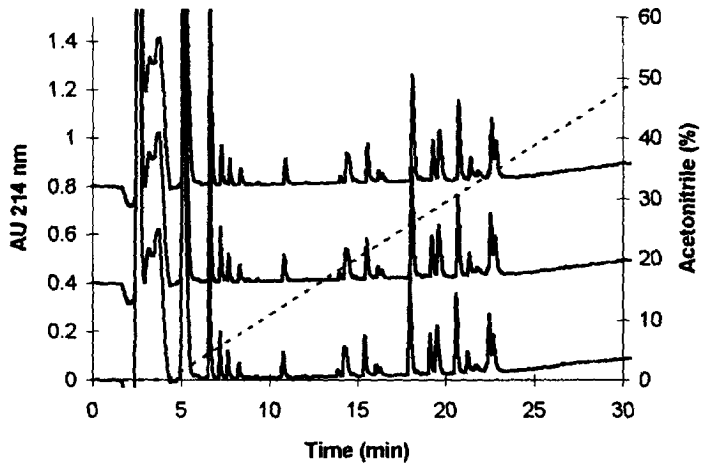

Fig. 2. Tryptic maps of lysozyme using an immobilized trypsin column. The flow-rate was 0.5 ml/min through a 30×2.1 mm Poroszyme column, producing a residence time in the enzyme reactor of approximately 12 s. Effluent from the Poroszyme column was transferred directly to a PepMap C_{18} reversed-phase liquid chromatography (RPLC) column by coupling the columns in tandem. Tryptic peptides were concentrated at the inlet of the RPLC column in this tandem mode. Following proteolysis of the protein by the immobilized trypsin column on valve V1 and concentration of peptides in the RPLC column on valve V2 the following sequence of events was initiated; (i) the protease column was removed from the mobile phase flow stream with valve V1, (ii) mobile phase was transported directly to the 4.6×250 mm PepMap C_{18} RPLC column, (iii) mobile phase velocity was increased to 1 ml/min, and (iv) the RPLC column was eluted with a linear gradient ranging from solvent 100% 1B to 70% 2B during the course of 30 min. Elution of peptide analytes was monitored at 214 nm.

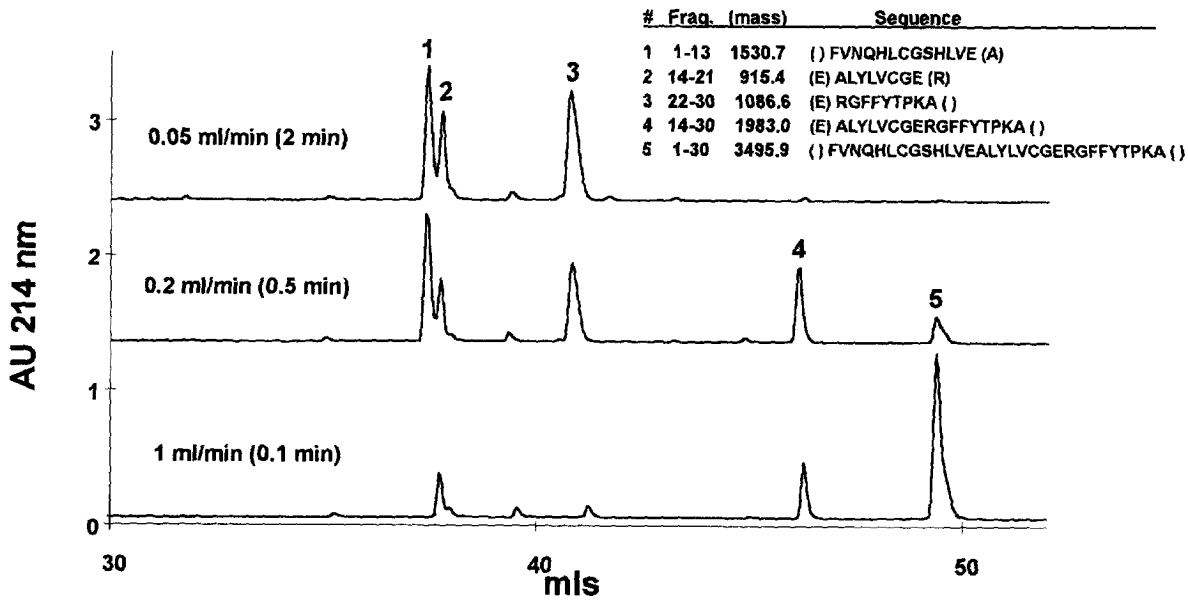

Fig. 3. Proteolytic reaction time study. Using 30×2.1 mm immobilized Glu-C column, reaction times of 120 s (0.05 ml/min flow-rate), 30 sec (0.2 ml/min flow-rate), and 12 s (0.5 ml/min flow-rate) were examined. The protocol and chromatographic conditions were identical to those used in Fig. 2.

peptide indicates that it is the result of partial digestion, i.e. it is a peptide containing uncleaved fragments 2 and 3. This indicates that the cleavage site between fragments 1 and 2 is clipped more rapidly by immobilized Glu-C than the site between fragments 2 and 3.

Reproducibility of tryptic maps as a function of large sample numbers was also examined. A single trypsin column was subjected to >200 injections of alkylation mixture without significant loss of activity. Figs. 2 and 4 are large sample number tryptic maps of lysozyme and bovine serum albumin, respectively.

3.4. Alkylating agent selection

The three alkylating agents, iodoacetic acid, iodoacetamide, and vinylpyridine, confer different hydrophobic and absorbance characteristics on the peptide fragments. This means that selectivity of the RPLC system, in terms of both chromatographic behavior and relative molar response of the detector, should vary widely between sulfhydryl and non-sulfhydryl containing peptides with different alkylating agents. This was investigated using lysozyme as a model system and the "MULTIMETHODS" software to construct three programs that could be coupled to add sequentially either iodoacetic acid, iodoacetamide, or vinylpyridine at the alkylation step during three automated tryptic mapping campaigns on the same sample (Fig. 5). Peptide peaks were also manually collected for mass spectral analysis. It was found that (i) chromatograms with the various alkylating reagents could be reproduced in multiple runs, (ii) elution time of sulfhydryl containing peptides, such as T9,T10, varied with alkylating agent while those of non-sulfhydryl peptides, such as T7 and T13, were constant, (iii) resolution of the digest was alkylating agent dependent, (iv) sulfhydryl containing peptides were easily recognized, (v) multiple peptide maps of the same sample resulting from the use of different alkylating agents provide further confirmation of

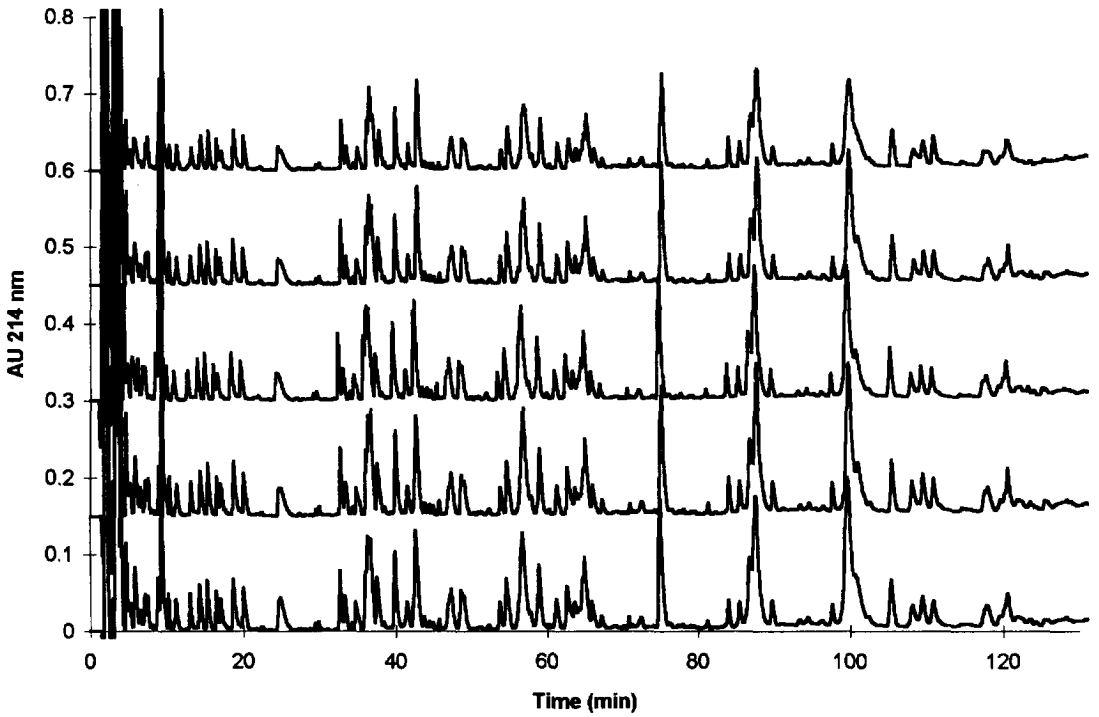

Fig. 4. Reproducibility of the mapping process with human serum albumin. Transport time through the 30×2.1 mm immobilized trypsin column was 120 s at 0.05 ml/min at 50°C. The protocol and chromatographic conditions were identical to those used in Fig. 2 with the exception that a 2 h gradient from 100% 1B to 70% 2B was used to elute the reversed-phase column.

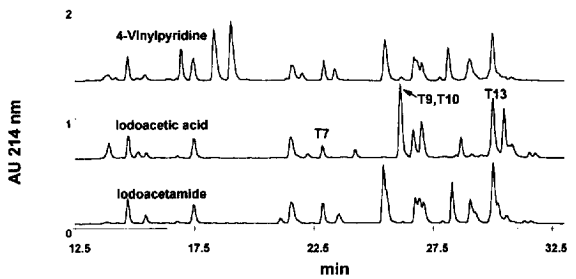

Fig. 5. Automated, sequential mapping of a protein using three different alkylating agents. The program used to carry out the individual automated analyses described in Fig. 2 was executed three times with the exception that a different alkylating agent was specified in each of the three cycles. Transport time through the 30×2.1 mm immobilized enzyme column was 120 s at 0.05 ml/min.

primary structure equivalency, and (vi) selection of the optimum alkylating agent may be achieved automatically.

4. Conclusions

The major conclusion to be derived from this work is that in confirming polypeptide primary structure by chromatographic mapping methods it is possible to automate the reduction, alkylation, proteolytic cleavage, and chromatographic phases of the mapping process on a single instrument platform. Moreover, the entire process takes only a few min more than the RPLC step alone, i.e. mapping may be achieved within 1–2 h. This is easily an order of magnitude faster than manual, non-integrated procedures.

It may be further concluded that this approach is highly reproducible and robust. Many determinations could be made with a single immobilized enzyme column without significant alteration of the proteolytic map.

Finally it is concluded that the methods described in this paper enable reproducible, automated, multi-dimensional chromatographic mapping in which different chromatographic maps may be produced by varying chemical characterization reactions performed on the sample protein before it is subjected to chromatography. Specific methods used in these studies were immobilized proteolytic enzymes to cleave polypeptides at different sites and multiple alkylating agents which convey different chromatographic and spectral properties to sulfhydryl containing peptides. This multi-variant approach will strengthen protein primary structure confirmation.

References

[1] F. Gilbert, Biologicals, 18 (1990) 355.
[2] J.V. O'Connor, Biologicals, 21 (1993) 111–117.
[3] J.L. Brown and W.K. Roberts, J. Biol. Chem., 251 (1975) 1009–1014.
[4] J.L. Brown, J. Biol. Chem., 254 (1979) 1447–1449.
[5] B.J. Sutton and D.C. Phillips, Biochem. Soc. Trans., 11 (1983) 130–132.
[6] F. Wold, Ann. Rev. Biochem., 50 (1981) 783.
[7] B.A. Johnson et al, J. Biol. Chem., 264 (1989) 14 262–14 271.
[8] Committee on Process Development and Manufacturing, Biopharm., 4 (1991) 22–27.
[9] R.L. Garnick, Develop. Biol. Standard, 76 (1992) 117–130.
[10] G.P. Royer and J.P. Andres, J. Macromol. Sci. Chem., A7 (1973) 1167.
[11] J. Emneus and G. Marko-Varga, J. Chromatogr. A, 703 (1995) 191–243.
[12] D.B. Kassel, T.G. Consler, M. Shalaby, P. Sekhri, N. Gordon and T. Nadler, in J.W. Crabb (Editor), Techniques in Protein Chemistry VI, Academic Press, San Diego, CA, 1995, pp. 39–46.
[13] J. Ronnenberg, B. Preitz, G. Wostemeier and Stephan Diekmann, J. Chromatogr. B, 656 (1994) 169–177.
[14] R.D. Voyksner, D.C. Chen and H.E. Swaisgood, Anal. Biochem., 188 (1990) 72–81.
[15] W-K. Chui and I.W. Wainer, Anal. Biochem., 201 (1992) 237–245.
[16] Y.L.F. Hsieh, H. Wang, C. Elicone, J. Mark, S.A. Martin and F. Regnier, Anal. Chem., 68 (1996) 455–462.