Journal of Chromatography, 646 (1993) 193–205
Elsevier Science Publishers B.V., Amsterdam

CHROMSYMP. 2819

Isolation of carboxyl-termini and blocked amino-termini of viral proteins by high-performance cation-exchange chromatography

Jeffrey J. Gorman*

Commonwealth Scientific and Industrial Research Organization, Australian Animal Health Laboratory, Geelong, Vic. (Australia) and Biomolecular Research Institute, 343 Royal Parade, Parkville, Vic. 3052 (Australia)

Brian J. Shiell

Commonwealth Scientific and Industrial Research Organization, Australian Animal Health Laboratory, Geelong, Vic. (Australia)

ABSTRACT

The strong cation-exchanger, PolySulfoethyl Aspartamide, has been assessed as a medium for isolation of carboxyl-terminal and blocked amino-terminal peptides from tryptic digests of small quantities of viral proteins.

Peptides with a single positive charge, the blocked amino-terminal peptides of ovalbumin and the Newcastle disease virus (NDV) matrix protein and carboxyl-terminal peptides of ovalbumin and the NDV nucleocapsid protein, eluted in early ion-exchange fractions and were readily isolated in homogeneous form by subsequent reversed-phase HPLC. Some early ion-exchange fractions also contained singly charged peptides derived by "chymotryptic-like" cleavage, whilst other peptides eluted in these fractions due to their highly acidic character. Terminal sequences with additional basic residues were isolated from later eluting ion-exchange fractions. Peptides with this property included the blocked amino-terminus of the NDV nucleocapsid protein and a portion of the carboxyl-terminus of the NDV matrix protein. Hitherto undescribed polymorphism in the amino-terminal region of ovalbumin was revealed in this study. Truncated peptides from the carboxyl-terminus of the NDV matrix protein were also detected. The presence of these peptides could be a reflection of carboxyl-terminal processing of the matrix protein.

The strategy described herein should be of general utility for selective microisolation of carboxyl-terminal peptides and blocked amino-terminal peptides from tryptic digests of proteins.

INTRODUCTION

Characterization of post-translational modifications of naturally occurring proteins is essential to fully define their active structures [1,2]. Analysis of proteins expressed *in vitro* is also required to verify that these proteins undergo the same post-translational modifications as their natural counterparts or acceptable alternate modifications [3–9]. Accordingly, strategies have been developed for determining a variety of post-translational modifications [4–8,10–13].

Modifications at the termini of mature proteins present a particularly difficult analytical challenge. These modifications include enzymatic cleavage and trimming of the primary translational products of genes and addition [14] or rearrangement reactions [15]. For example, amino-termini may be subjected to acetylation with or without preliminary removal of the initiating methionine [16]. In addition, amino-terminal glutamine residues may cyclize to the

* Corresponding author. Address for correspondence: Biomolecular Research Institute, 343 Royal Parade, Parkville, Vic. 3052, Australia.

0021-9673/93/$06.00 © 1993 Elsevier Science Publishers B.V. All rights reserved

pyrrolidone carboxyl, 5-oxoproline [15]. It has been estimated that around 80% of soluble mammalian proteins have acetylated amino-termini [17]. Similarly, carboxyl-terminal regions of proteins may be cleaved and/or trimmed and the resultant terminal α-carboxyl may or may not be converted to an amide or otherwise modified [1].

Edman degradation is ineffective for analysis of blocked (*e.g.* acetylated or cyclized) amino-termini and carboxyl-terminal sequences of intact proteins. Enzymatic [18,19] and chemical methods [20] have been described for deblocking [20] or removing acetylated amino-terminal amino acids [19] or cyclized amino termini [18]. Chemical manipulation has also been used to facilitate identification and isolation of blocked amino-terminal peptides from tryptic digests [21]. However, inherent limitations or additional sample handling may render these techniques ineffective, especially with limited quantities of proteins [19–22]. Furthermore, these techniques do not address isolation and characterization of the carboxyl-termini of proteins. Use of carboxypeptidases for carboxyl-terminal sequencing of intact proteins is also fraught with difficulties [23]. Hence, it is desirable to be able to selectively isolate fragments bearing amino terminal-blocking groups and carboxyl-terminal sequences of fully processed proteins. Once isolated, such fragments are amenable to a variety of structural analyses [19,21–23], such as enzymatic cleavage of the acetylated amino-terminal amino acid [19] or sequencing by mass spectrometric methods [5–7,12,21–25].

A strategy is described in this report for microisolation of carboxyl-terminal and blocked amino-terminal tryptic peptides using strong cation-exchange high-performance liquid chromatography (HPLC). This strategy exploits the fact that most tryptic peptides derived from the carboxyl-termini or blocked amino-termini of proteins have a single positive charge at acid pH and are barely retained by cation exchangers [26]. Other tryptic peptides have two or more positive charges under these conditions and are expected to be retained by cation exchangers. Thus, gradient elution is used to isolate terminal fragments with more than one positive charge. The efficacy of this strategy was assessed with an abundant protein and less available viral proteins isolated using microtechniques.

MATERIALS AND METHODS

Protein preparations

Hen egg white ovalbumin was obtained from Miles/ICN, Costa Mesa, CA, USA (Cat. No. 950512). This protein (10 mg in 1 ml of 0.1 *M* NH_4HCO_3–6 *M* guanidine hydrochloride) was reduced with 20 m*M* dithiothreitol for 16 h at 4°C followed by 2 h at 22°C. It was subsequently S-carboxamidomethylated by addition of iodoacetamide to 100 m*M* and allowing the reaction to proceed for 2 h at 22°C. Excess reagents were removed by methanol precipitation at −20°C followed by two washes of the precipitate with methanol.

The matrix protein of the avirulent V4 strain of Newcastle disease virus (NDV) was isolated from purified virions by detergent extraction according to an established method [27]. The apparently homogeneous protein, as judged by sodium dodecyl sulphate (SDS)–acrylamide gel electrophoresis (12%, w/v, acrylamide), was subsequently reduced and S-carboxamidomethylated using the conditions described above.

The nucleocapsid protein of the VRI-90/234 isolate of NDV [28] was obtained from purified virions which were disrupted with 1% (w/v) SDS and subjected to reduction with dithiothreitol and S-alkylated with 5-N-[(iodoacetamidoethyl)amino]naphthalene-1-sulphonic acid (5-IAEDANS) as described previously [12,29–31]. The fluorescently labeled proteins were subjected to preparative slab gel electrophoresis on SDS–acrylamide gradient (10–20%, w/v) gels, excised as individual fluorescent gel bands and recovered by electroelution [12,29–31].

Enzymatic digestion

Trypsin digestion of reduced and alkylated proteins was performed at substrate concentrations of approximately 10 mg/ml in 0.1 *M* NH_4HCO_3. Trypsin (sequencing grade; Boehringer, Mannheim, Germany) was added to give

an enzyme to substrate ratio of 1:100 (w/w) and the digestion mixture maintained at 37°C for 2 h prior to increasing the enzyme to substrate ratio to 1:50 and allowing the digestion to proceed for a further 2 h at 37°C. Digests were desalted by drying *in vacuo*, reconstituted in water and redried *in vacuo*.

Strong cation-exchange chromatography

Strong cation-exchange HPLC was performed with PolySulfoethyl Aspartamide functionalized silica (5 μm particle size; 300 Å pore size) obtained from PolyLC, Columbia, MD, USA, packed in stainless-steel columns (200 mm × 4.6 mm I.D.). Conditions used to achieve optimal ion-exchange separation of tryptic peptides are described below. Separation was performed using a Waters chromatographic system, comprising two Model 510 pumps, a Model U6K injector, a Model 440 dual-channel fixed-wavelength detector set at 254 and 280 nm, a Model 481 variable-wavelength detector set at 210 nm and a Model 840 control and data acquisition system. Fractions were collected manually into 1.5-ml polypropylene tubes and solvent removed *in vacuo* using a Savant centrifugal concentrator prior to further fractionation.

Reversed-phase chromatography

Reversed-phase chromatography was performed with a chromatographic system similar to the one described above with modifications to minimize system delay volumes. These modifications included use of a Waters static mixer (part No. WATO 51518) and a Rheodyne Model 7125 injector with a 200-μl loop. An Applied Biosystems Model 1000S diode array detector was the sole detector on this system. Separation was performed on a 5-μm particle size Vydac octadecasilica column (250 mm × 4.6 mm I.D., cat. No. 218TP54; Separations Group, Hysperia, CA, USA) using conditions described below. Samples were reconstituted in 0.1% (v/v) trifluoroacetic acid (TFA) in water prior to reversed-phase fractionation and fractions were collected into 1.5-ml polypropylene tubes using a Pharmacia Frac-300 fraction collector under control of an ISCO Model 1650 peak separator. Solvent was removed from fractions *in vacuo* as described above.

Analytical methods

Samples were prepared for amino acid analysis in sealed evacuated tubes containing 200 μl of 6 *M* HCl, 0.3% (v/v) 2-mercaptoethanol and 0.1% (v/v) phenol by hydrolysis at 110°C. Hydrolysates were dried *in vacuo* and reconstituted in 0.1 *M* HCl and individual amino acids were separated and quantitated using an ion-exchange based HPLC system [30].

Fractions were reconstituted in 80% (v/v) CH_3CN in water containing 0.1% (v/v) TFA for automated stepwise protein sequence analysis [32]. Sequencing was performed with an automated sequenator (Applied Biosystems Model 470A) with vapour-phase delivery of critical reagents [33] in conjunction with a synchronized phenylthiohydantoin (PTH)-amino acid separation system (Applied Biosystems Model 120A PTH-amino acid analyser).

Fast atom bombardment (FAB) mass spectrometry [34] was performed using a VG 70/70E-HF double-focusing magnetic deflection mass spectrometer of Nier-Johnson geometry (VG Analytical, Manchester, UK) equipped with an Ion-Tech saddle field fast-atom gun. Scan control and data acquisition were performed with a VG 250J+ data system operated in multichannel analysis mode. Samples were reconstituted in 5 μl of 50% (v/v) CH_3CN in water containing 0.1% (v/v) TFA and 1.5-μl aliquots were applied to the probe tip containing 1 μl of a preapplied mixture of glycerol–thioglycerol (1:1). Xenon atoms at a discharge potential of 8 keV and current of 1 mA were used for sample bombardment. Spectra were recorded in the positive ion mode at 40 s/decade and a resolution of 1500. Five to ten scans were averaged and mass measured using an external calibration file created with CsI/RbI/NaI cluster ions.

RESULTS

Conditions found to be optimal for ion-exchange separation of peptides in the tryptic digest of reduced and carboxamidomethylated

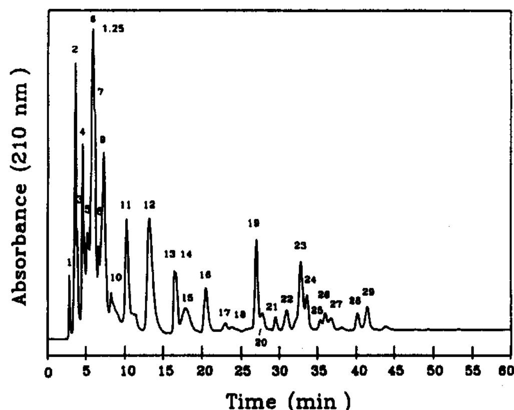

Fig. 1. Strong cation-exchange chromatography of unfractionated peptides derived by tryptic digestion of reduced and carboxamidomethylated ovalbumin. This separation was achieved using a buffer of 10 m*M* Na_2HPO_4–H_3PO_4 buffer (pH 3.0) containing 40% (v/v) acetonitrile as the starting buffer and buffer B comprised of starting buffer with Na_2SO_4 added to a concentration of 76 m*M*. The gradient protocol consisted of a linear increase of 0 to 20% buffer B over 50 min and 20 to 100% buffer B in a further 10 min. The eluent was subsequently held at 100% buffer B for 5 min, and then returned to starting buffer over a period of 5 min. A flow-rate of 1 ml/min was used throughout and the temperature was maintained at 22°C. The absorbance scale is indicated numerically adjacent to the apex of the highest absorbing peak.

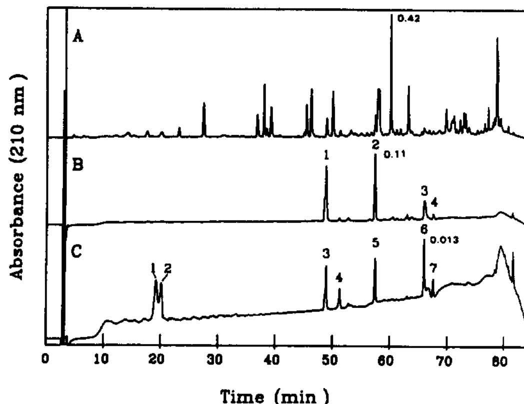

Fig. 2. Reversed-phase chromatography of unfractionated peptides derived by tryptic digestion of reduced and carboxamidomethylated ovalbumin (A) and tryptic peptides from ion-exchange fractions 2 (B) and 3 (C) of the tryptic digest (Fig. 1). Gradient elution was performed by varying the composition of acetonitrile in water whilst maintaining 0.1% (v/v) TFA with a flow-rate of 1 ml/min at 22°C. The gradient protocol consisted of linear increases in acetonitrile concentration from 0 to 24% (v/v) over 50 min, from 24 to 40% over the next 20 min and 40 to 80% over the final 15 min. The column was then returned to 0% CH_3CN over 5 min. Numbered fractions in chromatograms B and C were subjected to further analyses (Table I and Fig. 4). Absorbance scales for each chromatogram are indicated numerically adjacent to the apexes of their highest absorbing peaks.

ovalbumin (Fig. 1) employed 40% (v/v) aqueous acetonitrile in the sample and buffers in order to achieve a balance between optimal sample solubility and chromatographic separation. Peptides were eluted from the column using a Na_2SO_4 gradient [35]. The number of peaks observed corresponded closely to the theoretical number of tryptic peptides expected from this protein.

Ion-exchange fractions (Fig. 1) were collected and subjected to reversed-phase separation (Fig. 2). Analysis of reversed-phase fractions (Fig. 2) for amino acid compositions and, where appropriate, by stepwise sequence analysis and mass spectrometry (Table I) produced data consistent with the presence of the acetylated amino-terminal tryptic peptide (Fig. 3) in two reversed-phase peaks of both ion-exchange fractions 2 (Fig. 1 and Fig. 2B) and 3 (Fig. 1 and Fig. 2C). The carboxyl-terminal tryptic peptide (Fig. 3) was also located in two reversed-phase peaks of ion-exchange fraction 3 (Fig. 1 and Fig. 2C). The total amount of acetylated amino-terminal peptide (residues 1–16) was greater in ion-exchange fraction 2 than in fraction 3 which is consistent with incomplete resolution. However, there were some compositional differences between the reversed-phase fractions containing this sequence (see below) which depended upon which ion-exchange fraction they were derived from. One peptide present in both of these ion-exchange fractions [reversed-phase peaks 2 (Fig. 2B) and 5 (Fig. 2C)] was a singly charged portion of the ovalbumin sequence (residues 111–119) derived by "chymotryptic-like" cleavage of this protein as was the peptide (residues 62–72) in reversed-phase fraction 4 of ion-exchange fraction 3. Another portion of the sequence (residues 340–359) found in both of these ion-exchange fractions (peak 1; Fig. 2B and peak 3; Fig. 2C) appeared to be present due to its highly acidic content of 1 aspartate and 4 glutamates [36].

TABLE I

ANALYTICAL DATA ON TRYPTIC PEPTIDES OF OVALBUMIN

Amino acid[a]	Ion-exchange fraction 2				Ion-exchange fraction 3						
	Reversed-phase fractions				Reversed-phase fractions						
	1	2	3	4	1	2	3	4	5	6	7
CM Cys	–	–	1.02 (1)	1.01 (1)	1.09 (1)	1.13 (1)	–	–	–	0.91 (1)	0.54 (1)
Asp	1.05 (1)	–	1.04 (1)	1.01 (1)	–	–	1.03 (1)	1.09 (1)	–	1.00 (1)	0.94 (1)
Thr	–	–	–	–	–	–	–	–	–	–	–
Ser	3.19 (3)	–	2.14 (2)	2.12 (2)	1.09 (1)	1.03 (1)	2.90 (3)	1.00 (1)	–	2.14 (2)	2.0 (2)
Glu	4.22 (4)	2.06 (2)	1.25 (1)	1.22 (1)	–	–	4.18 (4)	2.30 (2)	2.43 (2)	1.51 (1)	1.26 (1)
Pro	–	ND (2)	–	–	ND (1)	ND (1)	–	ND (1)	ND (2)	–	–
Gly	2.16 (2)	–	2.08 (2)	2.18 (2)	–	–	2.01 (2)	1.91 (2)	–	2.66 (2)[c]	2.83 (2)[d]
Ala	4.33 (4)	–	2.08 (2)	1.33 (2)[b]	–	–	4.22 (4)	1.16 (1)	–	1.72 (2)	1.11 (2)[d]
Val	2.88 (4)	–	0.76 (1)	1.33 (1)[b]	0.82 (1)	0.84 (1)	3.40 (4)	–	–	0 (1)[c]	0.62 (1)[d]
Met	–	–	1.14 (1)	1.18 (1)	–	–	–	–	–	1.06 (1)	0.63 (1)
Ile	–	0.92 (1)	0.89 (1)	0.89 (1)	–	–	–	0.63 (1)	0.88 (1)	1.14 (1)	0.82 (1)
Leu	–	2.02 (2)	–	–	–	–	–	0.80 (1)	2.18 (2)	0.5 (0)	0.34 (0)
Tyr	–	1.99 (2)	–	–	–	–	–	–	1.54 (2)	–	–
Phe	0.98 (1)	–	2.62 (3)	2.63 (3)	–	–	1.08 (1)	1.09 (1)	–	2.40 (3)	1.77 (3)
Lys	–	–	1.00 (1)	0.87 (1)	–	–	–	–	–	1.38 (1)	0.91 (1)
His	–	–	–	–	–	–	–	–	–	–	–
Arg	1.17 (1)	–	–	–	–	–	1.15 (1)	–	–	–	–
Yield (pmol)	401	451	213	60	270	320	395	56	67	47	22
Sequence[e]	–	–	–	–	382–385	382–385	–	–	–	–	–
Mass[f]	–	1135.4 (1136.3)	1808.6 (1809.07)	1835.7[b] (1837.12)	462.0 (462.2)	462.0 (462.2)	–	–	–	–	–
Identity[g]	340–359	111–119	1–16	1–16	382–385	382–385	340–359	62–72	111–119	1–16	1–16

[a] Amino acid compositions of peptides are given as ratios calculated by dividing the quantity of each amino acid with the average value estimated for one amino acid residue, given as "Yield (pmol)". This average value was calculated by summing the pmoles of all amino acids detected for the particular peptide and dividing by the total number of residues expected for a sequence of ovalbumin appearing to correspond to the data. The ratios expected for each peptide are given in parentheses. Proline was not determined. Data on peptides isolated from ion-exchange fraction 2 and reversed-phase fractions 1 and 2 of ion-exchange fraction 3 are representative of three separate preparations. Other analyses were preformed on a single preparation.

[b] These data indicate variation of an alanine residue to a valine residue.

[c] These data suggest variation of a valine residue for a glycine residue.

[d] These data suggest variation of an alanine residue for a valine residue in addition to the variation detected in peptide 6.

[e] Data obtained by Edman degradation.

[f] Values are protonated parent ion masses detected which correspond to a defined sequence of ovalbumin. Theoretical values are presented in parentheses.

[g] Identity of the peptide corresponding to residue numbers of ovalbumin accounted for by the analytical data set.

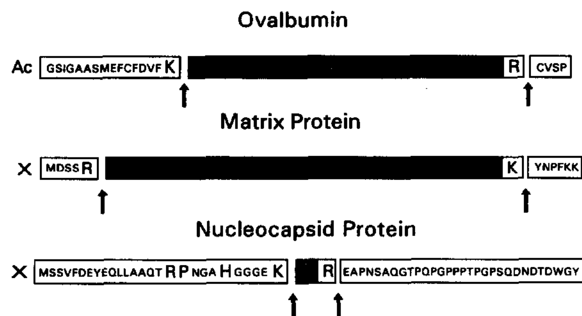

Fig. 3. Amino- and carboxyl-terminal sequences of ovalbumin [36], the NDV matrix protein [37,38] and the NDV nucleocapsid protein [39]. Ac indicates the presence of an acetyl group on the amino-terminus of ovalbumin [36] and X indicates that the amino-termini of the matrix and nucleocapsid proteins were refractory to Edman degradation. The fates of the initiating methionines on the matrix and nucleocapsid proteins were unknown, hence they are included. Arrows indicate the tryptic cleavage sites in these proteins giving rise to terminal fragments. Intervening portions of these proteins are indicated as shaded areas which are not drawn to scale.

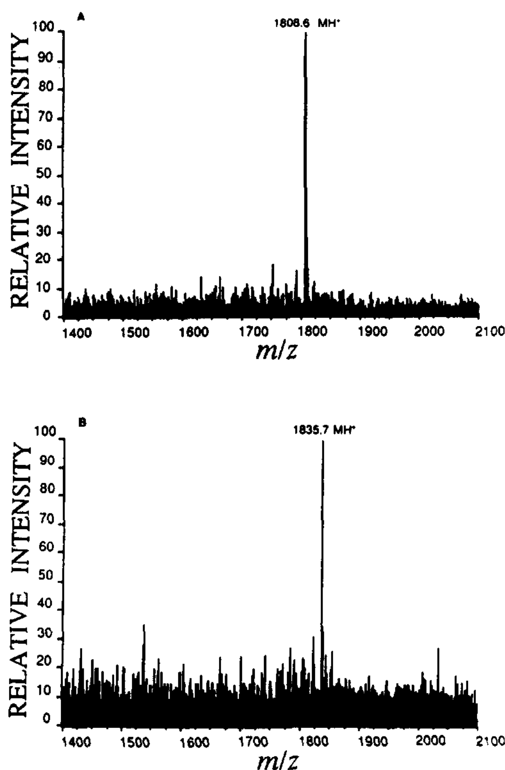

Fig. 4. Fast atom bombardment mass spectra of tryptic peptides contained in fractions 3 (A) and 4 (B) of reversed-phase chromatograms (Fig. 2B) of ion-exchange fraction 2 (Fig. 1) of the ovalbumin digest. MH^+ denotes the assignment of protonated parent ion mass to a particular ion based on correspondence of the mass to a discrete segment of the ovalbumin sequence.

The findings that amino-terminal and carboxyl-terminal sequences were both detected in two reverse phase chromatography fractions (Fig. 2B and C) of ion-exchange fractions 2 and 3 was investigated further. In the case of the amino-terminal sequence obtained from ion-exchange fraction 2 this was apparently due to polymorphism at residues five or six from the amino-terminus. Amino acid analysis data (Table I), obtained on reversed-phase fractions 3 and 4 (Fig. 2B), indicated replacement of one of the alanine residues at these positions with valine. This was supported by analysis of the molecular masses of the reversed-phase fractions by mass spectrometry (Fig. 4). Further polymorphism of this region of ovalbumin was also suggested by amino acid analysis of reverse phase fractions 6 and 7 of ion-exchange fraction 3 (Fig. 2C). It was apparent that this component of reversed-phase fractions 6 and 7 had a glycine in place of valine and that alanine five or six of the sequence isolated in fraction 7 had changed to valine. However, it should be noted that the data on fractions 6 and 7 were obtained on limited quantities of material and should be regarded as tentative at present. Unfortunately, these components were not isolated in sufficient quantities for further characterization by FAB-MS. Heterogeneity of the carboxyl-terminus was not defined by chemical analysis since both reverse phase fractions produced identical amino acid compositions and stepwise sequence data by Edman degradation and both had identical masses (Table I). This heterogeneity was reproducible with successive ion-exchange fractionations and probably represented *cis/trans* isomerism of the carboxyl-terminal proline residue.

The matrix protein of NDV was shown to have a blocked amino-terminus by application of Edman degradation to the intact protein (Fig. 3). Tryptic peptides of this protein were fractionated by ion-exchange chromatography (Fig. 5) and the early-eluting peaks subsequently subjected to reversed-phase HPLC (Fig. 6). Chemical

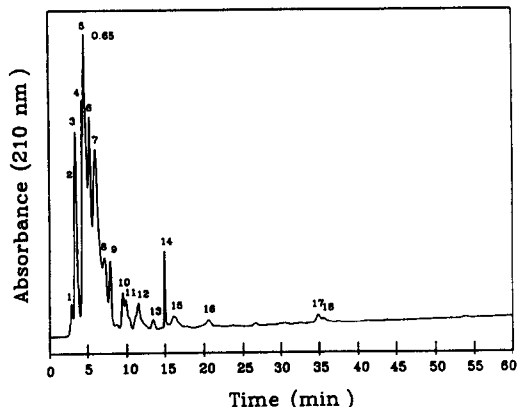

Fig. 5. Strong cation-exchange chromatography of unfractionated peptides derived by tryptic digestion of reduced and carboxamidomethylated NDV matrix protein. Conditions for chromatography as in Fig. 1 except that the CH_3CN concentration was 25% (v/v). The absorbance scale is indicated numerically adjacent to the apex of the highest absorbing peak.

analyses (Table II) of reversed-phase fraction 1 (Fig. 6B) of ion-exchange fraction 3 (Fig. 5) revealed that the amino-terminal peptide was acetylated on the initiating methionine. Reversed-phase fraction 6 (Fig. 6B) of this ion-exchange fraction produced data consistent with the carboxyl-terminal tryptic peptide which

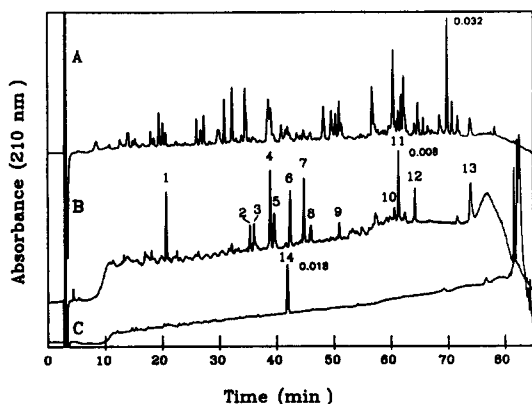

Fig. 6. Reversed-phase chromatography of unfractionated peptides derived by tryptic digestion of reduced and carboxamidomethylated NDV matrix protein (A) and tryptic peptides contained in ion-exchange fractions 3 (B) and 9 (C) of the tryptic digest (Fig. 6). Analytical data on numbered fractions are presented in Table II. Absorbance scales for each chromatogram are indicated numerically adjacent to the apexes of their highest absorbing peaks.

lacked two lysine residues (Fig. 3) predicted to be present on the protein by gene sequence analysis [37,38]. The other major reversed-phase fractions (4, 7 and 11) were identified as singly charged peptides derived in part by "chymotryptic-like" cleavage of the matrix protein (fractions 7 and 11) or the NDV fusion protein (fraction 4) which was present as a slight contaminant of the matrix protein preparation not apparent by SDS–polyacrylamide gel electrophoresis (PAGE). The minor peptides in reversed-phase fractions 2, 3, 5, 8, 9, 10, 12 and 13 also appeared to be "chymotryptic-like" peptides or baseline artifacts (peak 13) but were not present in sufficient quantities to enable their identification, however, they did not appear to be derived from the matrix or fusion proteins.

Ion-exchange fraction 9 (Fig. 5) was found to contain another part of the carboxyl-terminus of the matrix protein. This peptide apparently had two positive charges since it eluted much later in the ion-exchange chromatogram (peak 9; Fig. 5) and appeared to be homogeneous by reversed-phase HPLC (Fig. 6C). Amino acid analysis demonstrated that this peptide contained two amino groups, due to the α-amino group and an ϵ-amino group of one of the two lysine residues predicted to be at the carboxyl terminus (Fig. 3) of the matrix protein [37,38].

The nucleocapsid protein of NDV (Fig. 3) also has a blocked amino-terminus [12]. This protein was obtained by microisolation using SDS-PAGE and electroelution [12,29–31] and subjected to tryptic digestion. Fractionation of the tryptic digest isolated the carboxyl-terminal tryptic peptide of this protein (Fig. 3) in a single early eluting ion-exchange fraction (peak 3; Fig. 7). Chemical analysis (Table III) of peptides resolved by reversed-phase chromatography (Fig. 8B) showed that this ion-exchange fraction contained the carboxyl-terminal peptide (peak 1; Fig. 8B) and one other tryptic peptide (peak 2; Fig. 8B) which had a strongly acidic character due in part to presence of a sulfonic acid moiety of the AEDANS group attached to cysteine residue 139 and further influenced by the content of four aspartates and three glutamates [12,39].

The amino-terminus of the nucleocapsid protein was obtained in a homogeneous form from a

TABLE II

ANALYTICAL DATA ON TRYPTIC PEPTIDES OF THE NDV MATRIX PROTEIN

Amino acid[a]	Reversed-phase fractions					
	1	4	6	7	11	14
CM Cys	–	–	–	1.04 (1)	1.01 (1)	–
Asp	0.98 (1)	0.23 (0)	1.18 (1)	–	1.47 (1)	1.02 (1)
Thr	–	2.38 (3)	–	1.02 (1)	1.10 (1)	–
Ser	2.03 (2)	0.51 (0)	–	1.10 (1)	2.95 (3)	–
Glu	–	1.22 (1)	–	1.14 (1)	–	–
Pro	–	ND (1)	ND (1)	–	ND (2)	ND (1)
Gly	–	0.94 (1)	–	–	1.28 (1)	–
Ala	–	1.02 (1)	–	1.04 (1)	2.68 (3)	–
Val	–	–	–	–	–	–
Met	1.09 (1)	0.22 (0)	–	–	–	–
Ile	–	0.14 (0)	–	0.76 (1)	0.96 (1)	–
Leu	–	1.0 (1)	–	0.92 (1)	1.96 (2)	–
Tyr	–	0.8 (1)	0.96 (1)	–	0.96 (1)	1.01 (1)
Phe	–	–	0.87 (1)	–	1.67 (2)	0.96 (1)
Lys	–	–	–	–	–	1.01 (1)
His	–	–	–	–	–	–
Arg	0.9 (1)	–	–	–	–	–
Yield (pmol)	167	84	61	43	27	29
Identity[b]	1–5	375–383[c]	359–362	314–320	289–306	359–363

[a] Amino acid compositions of peptides are given as ratios calculated by dividing the quantity of each amino acid with the average value estimated for one amino acid residue, given as "Yield (pmol)". This average value was calculated by summing the pmoles of all amino acids detected for the particular peptide and dividing by the total number of residues expected for a sequence of the NDV matrix protein appearing to correspond to the data. The ratios expected for each peptide are given in parentheses. Proline was not determined. Data presented are representative of two separate preparations of all peptides.
[b] Identity of the peptide corresponding to residue numbers of the NDV matrix protein accounted for by the analytical data set.
[c] Data correspond to a sequence of the NDV fusion protein.

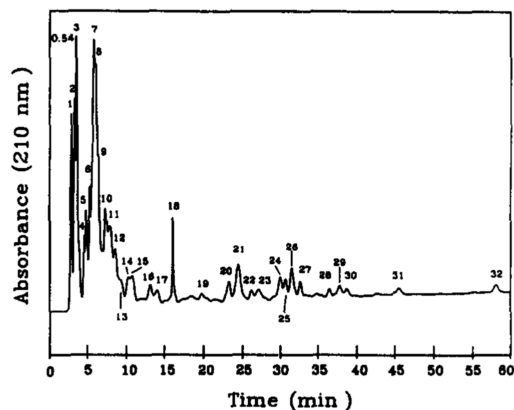

Fig. 7. Strong cation-exchange chromatography of unfractionated peptides derived by tryptic digestion of NDV nucleocapsid protein reduced and alkylated with IAEDANS and isolated by SDS-PAGE and electroelution. Conditions for chromatography were as in Fig. 1. The absorbance scale is indicated numerically adjacent to the apex of the highest absorbing peak.

late eluting ion-exchange fraction (peak 16; Fig. 7) as indicated by reversed-phase HPLC (Fig. 8C) of ion-exchange fraction 16 and chemical analysis (Table III) of the sole peptide apparent in the reversed-phase chromatogram (Fig. 8C). Amino acid compositional and mass spectrometric analyses indicated that this peptide contained three basic residues due to failure to cleave at an arginyl–prolyl bond between residues 17 and 18 and the presence of histidine at position 22, as predicted by gene sequence analysis [39]. This resistant bond is actually between residues 16 and 17 of the mature protein due to removal of the initiating methionine and acetylation of the subsequent serine during maturation of the protein [12], the present data confirmed this course of amino-terminal processing for the VRI-90/234 isolate of NDV [28].

TABLE III

ANALYTICAL DATA ON TRYPTIC PEPTIDES OF THE NDV NUCLEOCAPSID PROTEIN

Amino acid[a]	Reversed-phase fractions		
	1	2	3
CM Cys	–	1.17 (1)	–
Asp	5.00 (5)	4.96 (5)	1.97 (2)
Thr	2.92 (3)	4.82 (4)[b]	1.03 (1)
Ser	2.28 (2)	1.31 (1)	1.99 (2)
Glu	4.53 (4)	2.96 (3)	4.97 (5)
Pro	ND (8)	ND (2)	ND (1)
Gly	4.34 (4)	2.15 (2)	4.05 (4)
Ala	2.07 (2)	2.08 (3)[b]	3.11 (3)
Val	–	1.48 (2)	0 (1)[c]
Met	–	–	–
Ile	–	0.92 (1)	0.94 (0)[c]
Leu	–	1.10 (1)	2.06 (2)
Tyr	0.92 (1)	–	0.87 (1)
Phe	–	0.87 (1)	0.95 (1)
Lys	–	–	0.94 (1)
His	–	–	0.96 (1)
Arg	–	1.19 (1)	1.10 (1)
Yield (pmol)	270	350	108
Sequence[d]	460–488 +	–	Nil
Mass[e]	Nil	3161.6 (3059.22)	2821.2 (2819.02)
Identity[f]	460–489	138–174	2–27

[a] Amino acid compositions of peptides are given as ratios calculated by dividing the quantity of each amino acid with the average value estimated for one amino acid residue, given as "Yield (pmol)". This average value was calculated by summing the pmoles of all amino acids detected for the particular peptide and dividing by the total number of residues for a sequence of the NDV V_4 strain nucleocapsid protein appearing to correspond to the data. The ratios expected for each peptide are given in parentheses. Proline was not determined. Data presented are representative of two separate preparations of all peptides.

[b] These data indicate variation of an alanine residue to a threonine residue in the 90/234 nucleocapsid protein.

[c] These data indicate variation of a valine residue to an isoleucine residue in the 90/234 nucleocapsid protein.

[d] Data obtained by Edman degradation; + indicates that no signal was detected after residue 488 and Nil indicates that no sequence was detected.

[e] Values are protonated parent ion masses detected which correspond to a defined sequence of the nucleocapsid protein allowing for strain variation; Nil indicates that no corresponding mass was detected. Theoretical values are presented in parentheses.

[f] Identity of the peptide corresponding to residue numbers of the NDV nucleocapsid protein accounted for by the analytical data set.

DISCUSSION

A strategy is described herein for selective isolation of carboxyl-terminal and blocked amino-terminal peptides from tryptic digests of proteins. This strategy involves ion-exchange fractionation and enrichment of these peptides from complex mixtures into simpler mixtures which can subsequently be fractionated into homogeneous peptides with relative ease by reversed-phase HPLC. The ion-exchange fractionation exploits the fact that most tryptic peptides have at least two positive charges at acid pH, contributed by the α-amino group of the amino-terminus and the side chain of lysine or arginine at the carboxyl-terminus. Blocked amino-terminal tryptic peptides, however, lack the positive charge of the α-amino group, and carboxyl-terminal tryptic peptides generally lack lysine or arginine. Exceptions to this rule will be if the sequences being sought contain a histidine and/or a lysine/arginine-proline sequence or if

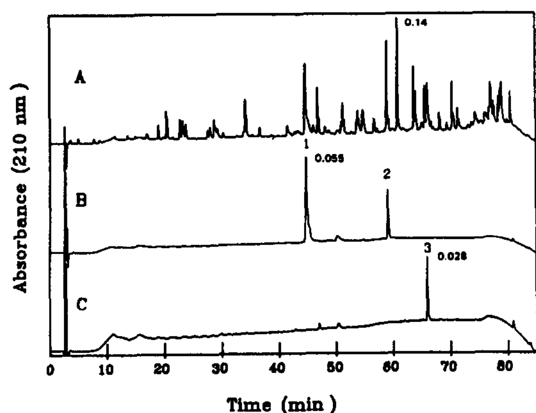

Fig. 8. Reversed-phase chromatography of unfractionated peptides derived by tryptic digestion of AEDANS-labelled NDV nucleocapsid protein (A) and tryptic peptides contained in ion-exchange fractions 3 (B) and 16 (C) of the tryptic digest (Fig. 8). Analytical data on numbered fractions are presented in Table III. Absorbance scales for each chromatogram are indicated numerically adjacent to the apexes of their highest absorbing peaks.

proteins of interest terminate with lysine or arginine at their carboxyl-termini. Basic amino acids are resistant to trypsin cleavage under these circumstances and thus will impart an additional positive charge to these peptides. Whilst this appears to detract from the strategy, it is not a severe deficiency since histidine and lysine/arginine–proline are relatively rare. Furthermore, in most cases this strategy would be utilized to document the structures of proteins whose sequences would have already been deduced by gene sequence analysis, thus, it would be possible to predict the likely success of the strategy in advance. If terminal sequences containing multiple positive charges were expected in tryptic digests, it might be possible to predict where they would elute during ion-exchange chromatography. The present study was similar in concept to a previous report [26], in which stepwise elution from a sulfopropyl substituted hydrophilic polyvinyl strong cation-exchange column was used to purify singly charged terminal tryptic fragments of proteins. However, isolation of terminal peptides with more than one positive charge, in homogeneous form, by gradient elution of retained peptides was not investigated in the earlier study [26]. Furthermore, the present study also evaluated the applicability of this concept to scarce viral proteins purified by microisolation procedures.

An important factor in this strategy is the excellent separation of peptides that can be achieved with the commercially available strong cation-exchanger, PolySulfoethyl Aspartamide, in HPLC column form, based on net charge variation [35,40–45]. This factor has previously been exploited for net charge based isolation of tryptic peptides bearing other specific post-translationally derived structural features, such as disulfide bonds [35,43,46]. Charge variants of peptides arising from polymorphic proteins have also been resolved on this medium [45]. It has also been demonstrated that it is possible to resolve peptides of the same net charge on this ion-exchanger due to selectivity based on juxtaposition of charged residues [45]. It is conceivable that in some circumstances this selectivity might lead to separation of carboxyl-terminal and blocked amino-terminal peptides with a single positive charge. Parenthetically, the sulfopropyl column used in the above-mentioned stepwise elution method [26] was assessed at a preliminary stage of the present study but it failed to give satisfactory gradient elution of tryptic peptides from ovalbumin.

The currently described strategy has been shown to be effective with an abundant protein (*e.g.* ovalbumin), a relatively scarce protein isolated by conventional protocols (*e.g.* NDV matrix protein), and a protein isolated on a microscale by SDS-PAGE and electroelution (the NDV nucleocapsid protein). The strategy has also been applied to the NDV haemagglutinin/neuraminidase-protein, isolated by SDS-PAGE and electroelution, which is even less abundant than the proteins investigated herein and products of viral genes artificially expressed using the baculovirus expression system (unpublished data).

Ovalbumin is a readily available protein with a fully defined amino acid sequence [36]. This protein is known to have an acetylated amino-terminus [36,47] and a carboxyl-terminus which should yield tryptic peptides with a single positive charge [36]. Hence, it was chosen as a model for assessing conditions for achieving maximal

separation between tryptic peptides with single and double positive charges and optimal peak shapes. As predicted the carboxyl-terminal and blocked amino-terminal tryptic peptides of ovalbumin contained single positive charges and eluted in early ion-exchange fractions. It was of interest that heterogeneity was detected in both of the amino- and carboxyl-terminal sequences. The amino-terminal heterogeneity was due to protein polymorphism and it is likely that the carboxyl-terminal heterogeneity was due to *cis*/*trans* isomerism of the carboxyl-terminal proline [48]. This amino-terminal polymorphism of ovalbumin has not been described previously. The polymorphism probably reflects the fact that the commercially obtained protein was a product from a pool of eggs from many chickens with varying ovalbumin sequences. Polymorphism in proteins is not uncommon [45,49] and is known to occur at other sites in ovalbumin [50,51]. Interestingly, this polymorphism was not observed in a previous analysis of ovalbumin obtained from pooled eggs [36]. The amino-terminal tryptic peptides of ovalbumin are very hydrophobic and are barely apparent in the reversed-phase chromatogram of the total digest of ovalbumin (Fig. 2A). The minor variant peptide (Fig. 2B) is present in about one fourth the abundance of the major amino-terminal peptide. These peptides, especially the minor variant, may not have been isolated at all in the present study unless the terminal sequences had been subjected to relative enrichment by ion-exchange fractionation. The presence of acetonitrile in the sample and ion-exchange buffers may have facilitated their isolation by enhancing their solubilization in the presence of other tryptic peptides. The previously described step elution ion-exchange method [26] for isolation of terminal peptides did not incorporate use of an organic solvent. The carboxyl-terminal peptide was also only a minor component of the total tryptic digest of ovalbumin. Isolation of this fragment was also greatly facilitated by ion-exchange enrichment.

With the less available NDV matrix protein, the acetylated amino-terminal sequence was isolated with relative ease in an early ion-exchange fraction as a singly charged peptide, as was a portion of the carboxyl-terminus predicted from the gene sequence [37,38] less two residues from the terminus. In addition, another peptide was isolated which contained one of the two lysine residues predicted at the carboxyl-terminus. These data suggest processing of the carboxyl-terminus of the matrix protein prior to tryptic digestion. This is likely to be due to the action of a "carboxypeptidase-like" enzyme because it cannot be accounted for by the endoproteolytic specificity of trypsin for arginine and lysine residues. However, "chymotryptic-like" cleavage cannot be ruled out in the case of the singly charged peptide terminating at phenylalanine 362. The carboxyl-terminal sequence of the NDV nucleocapsid protein was also readily isolated as a singly charged tryptic peptide from an early eluting ion-exchange fraction. Elution of the acetylated amino-terminal peptide of the nucleocapsid protein was delayed, reflecting the influence of its additional basic residues.

In conclusion, the currently described strategy should prove to be a general method for microisolation of terminal fragments from both ends of proteins of interest in order to document their mature forms and to define post-translational biosynthetic processing of such proteins. This procedure may be of value for characterization and quality control of proteins produced in recombinant expression systems for both research and regulatory purposes [2–8]. Another application of this strategy may be for defining the termini of proteins in order to produce primers to enable polymerase chain reaction amplification of entire genes encoding proteins of interest. Where singly charged blocked amino-terminal peptides or carboxyl-terminal peptides are predicted to ensue from tryptic digestion of a protein, these peptides should be contained in early eluting ion-exchange fractions and be readily purified by reversed-phase HPLC using the currently described fractionation regime. Furthermore, where variations in the net charge state of the peptides of interest are expected, due to prior determination or deduction of their amino acid sequences, their elution positions can be anticipated to some degree according to their net charge states [35,40–44], particularly if the performance of the ion-exchange column has

been calibrated with peptide standards [42,44]. With thorough analysis of ion-exchange fractions it is also possible to identify and characterize portions of the carboxyl-termini of proteins which have unexpected net charge variation due to post-translational processing.

ACKNOWLEDGEMENTS

We gratefully acknowledge the acquisition of mass spectrometric and Edman degradation sequence data by Gary Corino and Gary Beddome, respectively. Constructive review of this manuscript by Neil McKern and Lindsay Sparrow is also greatly appreciated.

REFERENCES

1 F. Wold, *Ann. Rev. Biochem.*, 50 (1981) 783.
2 S.C.B. Yan, B.W. Grinnel and F. Wold, *Trends Biochem. Sci.*, 14 (1989) 264.
3 F.A. Robey, in K.A. Walsh (Editor), *Methods in Protein Sequence Analysis*, Humana Press, Clifton, NJ, 1987, p. 67.
4 R.L. Garnick, N.J. Solli and P.A. Papa, *Anal. Chem.*, 60 (1988) 2546.
5 S.A. Carr, M.E. Hemling and G.D. Roberts, in D.H. Schlesinger (Editor), *Macromolecular Sequencing and Synthesis, Selected Methods and Applications*, A.R. Liss, New York, 1988, p. 83.
6 S.A. Carr, G.D. Roberts and M.E. Hemling, in C.N. McEwen and B.S. Larsen (Editors), *Mass Spectrometry of Biological Materials*, Marcel Dekker, New York, 1990, p. 87.
7 S.A. Carr, M.E. Hemling, M.F. Bean and G.D. Roberts, *Anal. Chem.*, 63 (1991) 2802.
8 V.R. Anicetti, B.A. Keyt and W.S. Hancock, *Trends Biotechnol.*, 7 (1989) 342.
9 D.T.-H. Liu, *Trends Biotechnol.*, 10 (1992) 114.
10 S.A. Carr and K. Biemann, *Methods Enzymol.*, 106 (1984) 29.
11 S.A. Carr and G.D. Roberts, *Anal. Biochem.*, 157 (1986) 396.
12 J.J. Gorman, G.L. Corino and B.J. Shiell, *Biomed. Environ. Mass Spectrom.*, 19 (1990) 646.
13 J.J. Gorman, *Trends Anal. Chem.*, 11 (1992) 96.
14 K. Narita, *Biochim. Biophys. Acta*, 28 (1958) 184.
15 E.M. Press, P.J. Piggot and R.R. Porter, *Biochem. J.*, 99 (1966) 356.
16 R.A. Bradshaw, *Trends Biochem. Sci.*, 14 (1989) 276.
17 J.L. Brown and W.K. Roberts, *J. Biol. Chem.*, 251 (1976) 1009.
18 R.F. Doolittle and R.W. Armentrout, *Biochemistry*, 7 (1968) 516.
19 R.G. Krishna, C.C.Q. Chin and F. Wold, *Anal. Biochem.*, 199 (1991) 45.
20 D. Wellner, C. Panneerselvam and B.L. Horecker, *Proc. Natl. Acad. Sci. U.S.A.*, 87 (1990) 1947.
21 D.L. Crimmins and R.L. Thoma, *J. Chromatogr.*, 634 (1993) 241.
22 A. Aitken, M.J. Geisow, J.B.C. Findlay, C. Holmes and A. Yarwood, in J.B.C. Findlay and M.J. Geisow (Editors), *Protein Sequencing: A Practical Approach*, IRL Press, Oxford, 1989, p. 43.
23 A. Yarwood, in J.B.C. Findlay and M.J. Geisow (Editors), *Protein Sequencing: A Practical Approach*, IRL Press, Oxford, 1989, p. 119.
24 K. Biemann, in J.B.C. Findlay and M.J. Geisow (Editors), *Protein Sequencing: A Practical Approach*, IRL Press, Oxford, 1989, p. 99.
25 K. Biemann and H.A. Scoble, *Science*, 237 (1987) 992.
26 H. Kawasaki, S. Imajoh and K. Suzuki, *J. Biochem.*, 102 (1987) 393.
27 A. Scheid and P.W. Choppin, *J. Virol.*, 11 (1973) 263.
28 J.J. Gorman, A.N. Hodder, P.W. Selleck and E. Hansson, *J. Virol. Meth.*, 37 (1992) 55.
29 J.J. Gorman, *Anal. Biochem.*, 160 (1987) 376.
30 J.J. Gorman, G.L. Corino and S.J. Mitchell, *Eur. J. Biochem.*, 168 (1987) 169.
31 J.J. Gorman, A. Nestorowicz, S.J. Mitchell, G.L. Corino and P.W. Selleck, *J. Biol. Chem.*, 263 (1988) 12522.
32 P. Edman and G. Begg, *Eur. J. Biochem.*, 1 (1967) 80.
33 R.M. Hewick, M.W. Hunkapiller, L.E. Hood and W.J. Dryer, *J. Biol. Chem.*, 256 (1981) 7990.
34 M. Barber, R.S. Bordoli, G.J. Elliot, R.D. Sedgwick and A.N. Tyler, *Anal. Chem.*, 54 (1982) 645A.
35 P.C. Andrews, in J.L. Villafranca (Editor), *Current Research in Protein Chemistry*, Academic Press, San Diego, CA, 1990, p. 95.
36 A.D. Nisbet, R.H. Saundry, A.J. Moir, L.A. Fothergill and J.E. Fothergill, *Eur. J. Biochem.*, 115 (1981) 335.
37 P. Chambers, N.S. Millar, S.G Platt and P.T. Emmerson, *Nucleic Acids Res.*, 14 (1986) 9051.
38 L.W. McGinnes and T.G. Morrison, *Virology*, 156 (1987) 221.
39 N. Ishida, H. Taira, T. Omata, K. Mizumoto, S. Hattori, K. Iwasaki and M. Kawakita, *Nucleic Acids Res.*, 14 (1986) 6551.
40 A.J. Alpert and P.C. Andrews, *J. Chromatogr.*, 443 (1988) 85.
41 D.L. Crimmins, J. Gorka, R.S. Thoma and B.D. Schwartz, *J. Chromatogr.*, 443 (1988) 63.
42 D.L. Crimmins, R.S. Thoma, D.W. McCourt and B.D. Schwartz, *Anal. Biochem.*, 176 (1989) 255.
43 D.L. Crimmins, *Peptide Res.*, 2 (1989) 395.
44 C.T. Mant and R.S. Hodges, in C.T. Mant and R.S. Hodges (Editors), *High-Performance Liquid Chromatography of Peptides and Proteins: Separation, Analysis and Conformation*, CRC Press, Boca Raton, FL, 1991, pp. 171–185.
45 P. Iadarola, M.C. Zapponi, L. Minchiotti, M.L. Meloni, M. Galliano and G. Ferri, *J. Chromatogr.*, 512 (1990) 165.

46 J.D. Sandy, C.R. Flannery, R.E. Boynton and P.J. Neame, *J. Biol. Chem.*, 265 (1990) 21108.
47 R.D. Palmiter, J. Gagon and K.A. Walsh, *Proc. Natl. Acad. Sci. U.S.A.*, 75 (1978) 94.
48 J.C. Gesquiere, E. Diesis, M.T. Cung and A. Tartar, *J. Chromatogr.*, 478 (1989) 121.
49 H. Jörnvall, in K.A. Walsh (Editor), *Methods in Protein Sequence Analysis: 1986*, Humana Press, Clifton, NJ, 1987, pp. 323–333.
50 R.L. Wiseman, J.E. Fothergill and L.A. Fothergill, *Biochem. J.*, 127 (1972) 775.
51 S.L.C. Woo, W.G. Beattie, J.F. Catterall, A. Dugaiczyk, R. Staden, G.G. Brownlee and B.W. O'Malley, *Biochemistry*, 20 (1981) 6437.