ELSEVIER

Journal of Pharmaceutical and Biomedical Analysis
18 (1998) 545–554

JOURNAL OF
PHARMACEUTICAL
AND BIOMEDICAL
ANALYSIS

A rapid and sensitive procedure for the micro-purification and subsequent characterization of peptides and protein samples by N-terminal sequencing and matrix assisted laser desorption ionization time of flight mass spectrometry

R. Grimm [a], K.D. Grasser [b,1], J. Kubach [a], W.S. Hancock [c,*]

[a] *Hewlett-Packard, Chemical Analysis Group Europe, 76337 Waldbronn, Germany*
[b] *Institute for Biology III, University of Freiburg, Schänzlestrasse 1, 79104 Freiburg, Germany*
[c] *HP-Labs, Palo Alto, CA, USA*

Received 15 May 1998; received in revised form 29 July 1998; accepted 16 August 1998

Abstract

The characterization of the proteome, a key activity in the post-genome era, is made extremely challenging by the microheterogeneity introduced by post translational modifications such as glycosylation in the diverse set of proteins expressed in a cellular system. High resolution separation systems, such as 2D-gel electrophoresis and more recently liquid chromatography (LC) have been used to fractionate these complex mixtures, however, subsequent mass analysis is hindered by the low level of the purified components. Off-line coupling of matrix assisted laser desorption ionization time of flight mass spectrometry (MALDI-TOF/MS) is an attractive technique for the analysis of such samples, but suffers from sensitivity to the degree of salt contamination that is unavoidable in the isolation of low level protein samples from biological extracts. In this publication we will report on a novel application of a commercially available system for the micro-purification of peptides and proteins. In this procedure micro-columns (normally used for sequencing of electroblotted samples) were used to rapidly purify protein digests or crude extracts of proteins in sufficient amounts for further analyses by protein sequencing and MALDI-TOF/MS. To demonstrate the applicability of these techniques we isolated and performed structural analysis of the following samples: a high-mannose glycopeptide isolated from a digest of the glycoprotein rt-PA, a poly-His tagged recombinant DNA-binding protein isolated by Ni^{2+}-chelating agarose and a polyclonal antibody sample.

Keywords: Sample loading station; Micro-purification; Glycopeptide; Protein sequencing; MALDI-TOF/MS

* Corresponding author. Fax: +1-650-857-2860; e-mail: william_hancock@hp.com.
[1] Current address: Dade, Bonnstrasse 9, 3186 Düdingen, Switzerland.

0731-7085/98/$ - see front matter
PII: S0731-7085(98)00208-8

1. Introduction

A new era in pharmaceutical analysis has begun with the advent of the human genome and the subsequent emergence of a greatly increased need for characterisation of novel protein sequences. In the analysis of proteins and peptides, most native samples can only be obtained at the picomole level. Such samples are typically isolated from microbore or capillary columns with an inner diameter of 1 mm to 300 μm respectively [1–3] or isolated from 2D electrophoresis gels. The latter samples are usually available in significantly smaller amounts, with amounts of a hundred femtomoles or less as typical amounts. To perform reliable characterization of these small amounts of sample, new methods need to be developed to minimize sample loss and to reduce low molecular weight impurities to the level necessary for high sensitivity matrix assisted laser desorption ionization time of flight mass spectrometry (MALDI-TOF MS) analysis as well as N-terminal protein sequencing [4–6].

In addition to off-line analysis, high sensitivity chromatographic systems can work in an on-line mode with MS detectors at the low picomole and even in the high femtomole range, such as in microbore LC/MS, LC/MS-MS and CE-MS [7–9]. Even in these applications, however, efficient isolation and sample handling procedures are key to success in the subsequent hyphenated analysis.

In this study, we have adapted a commercially available sample loading station to provide a simple and very efficient micropurification device. In a straightforward one step procedure commercially available mini-columns can be packed with any column material. After a single step purification, subsequent characterization of the sample by techniques such as protein sequencing and MALDI-TOF/MS analysis is readily performed. To demonstrate the capability of the sample loading station a specific glycopeptide was purified from a tryptic digest of the glycoprotein rt-PA and subsequently identified. Furthermore, sufficient amounts of the poly-His tagged recombinant maize DNA-binding protein, HMGa, could be isolated and used to obtain reliable amino acid sequence results and mass information by MALDI-TOF/MS. After enzymatic digestion important internal protein sequence information could also be obtained. The micro-purification of polyclonal antibodies is another example presented in this study.

2. Materials and methods

2.1. Materials

Trypsin, Concanavalin A Sepharose, methyl-a-D-mannopyranoside, Protein A linked to polyacrylamide beads were purchased from Sigma (St. Louis, MO). Ni^{2+}-chelating agarose was from Qiagen (Hilden, FRG). rt-PA was kindly provided by Genentech (San Francisco, CA). All chemicals used were p.a. quality.

2.2. Tryptic digestion of rt-PA

Reduction, alkylation and tryptic digestion of 1 mg rt-PA was carried out essentially as described in [7].

2.3. Expression of the His tagged HMGa protein in Escherichia coli

The coding region of the cloned maize HMGa cDNA [10] was cloned into a BamHI and HindIII digested pQE9 expression vector from Diagen (Hilden, Germany) (details will be reported elsewhere). Expression of the recombinant His-tagged HMGa protein using the *E. coli* strain M15 [pREP4] was performed as described in [11].

2.4. The chromatography system

The chromatography system consists essentially of the Hewlett Packard N-terminal protein sequencer loading station (HP G1001A). All chromatographic manipulations were carrier out using nitrogen pressure under control of a simple built-in manometer. Mini-columns were connected to a polypropylene funnel which has a capacity of 5 ml solution and which was placed in the loading station.

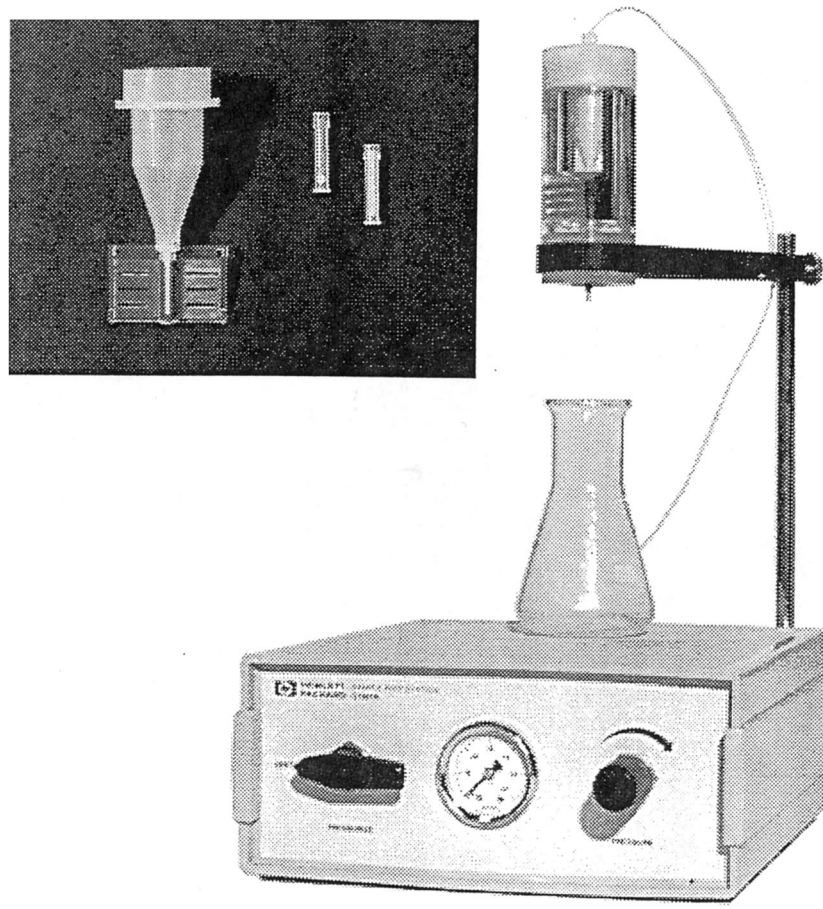

Fig. 1. The Hewlett-Packard sequencer sample loading station HP G1001A and mini-columns plus funnel used in this study.

2.5. Glycopeptide purification using Concavalin A affinity chromatography

After insertion of the empty column-funnel assembly into the loading station, 100 μl of affinity chromatography buffer, consisting of 20 mM Tris, pH 7.4, 0.5 M NaCl, 1 mM $MnCl_2$ and 1 mM $CaCl_2$, were added in the funnel by applying 10 psi nitrogen pressure for 2 s. Afterwards, 120 μl of the Concavalin A Sepharose slurry was filled into the funnel. A bubble-free packing of the column was subsequently achieved by applying less than 2 psi of pressure. After packing, the column was removed from the loading station and disconnected from the funnel. A small Teflon frit was finally placed carefully on top of the column.

Thereafter, the column together with the funnel were placed back into the loading station and the column was equilibrated with 5 ml affinity buffer. The tryptic digest was diluted by 1:1 with affinity buffer and placed into the funnel. After passing over of the digest, exhausted washing of the column was carried out with 2 ml of affinity buffer. Specific elution of the high-mannose containing glycopeptide from the r-tPA digest was performed with 200 μl 0.2M methyl-a-D-mannopyranosid solution.

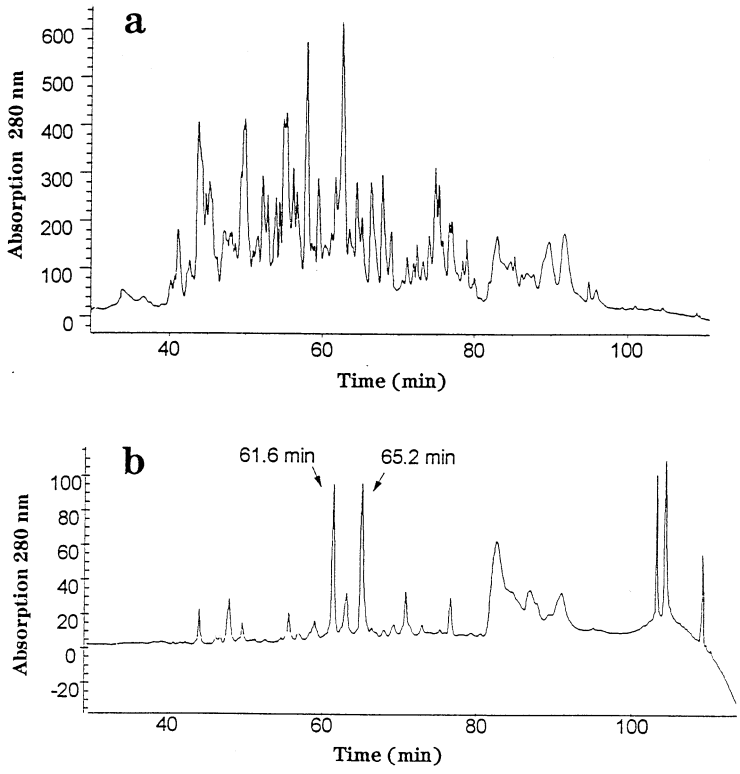

Fig. 2. Reversed-phase HPLC separation of: (a) the complete tryptic digest of rt-PA; and (b) of the eluate of Con A-Sepharose column obtained with methyl-a-D-mannopyranosid.

2.6. Purification of recombinant DNA-derived HMGa protein with Ni^{2+}-chelating agarose

Column packing with Ni^{2+}-chelating agarose was essentially carried out as described above, except that no frit was placed on top of the packed column at the end of the filling procedure. For filling and washing, column buffer A (50 mM Tris/HCl, 10 mM 2-mercapto-ethanol, 0.5 mM PMSF, 10% (v/v) glycerol, 0.5 M NaCl, 100 mg ml^{-1} bezamidine) was used.

After applying the crude *E. coli* extract to the column, it was washed with 5 ml of buffer A. Specific elution of the recombinant His-tagged protein was performed with buffer B (50 mM Tris/HCl, 10 mM 2-mercapto-ethanol, 0.5 mM PMSF, 100 mg ml^{-1} benzamidine, 25 mM EDTA).

2.7. Antibody purification using protein A linked to polyacrylamide beads

The filling of the column with protein A beads and washing of the packed column was performed as described above, except 50 mM Tris pH 7.4, 0.5 M NaCl was used as a buffer. A serum sample containing polyclonal antibodies was diluted 1:1 with buffer and loaded into the funnel. After passing over the serum sample, the column was extensively washed with buffer ($\sim$ 4 ml). Antibodies were eluted with 150 μl of 50 mM citrate buffer pH 3.0.

2.8. Reversed-phase HPLC

RP-HPLC separations were performed on the HP 1090 Series II HPLC system (Hewlett-Pack-

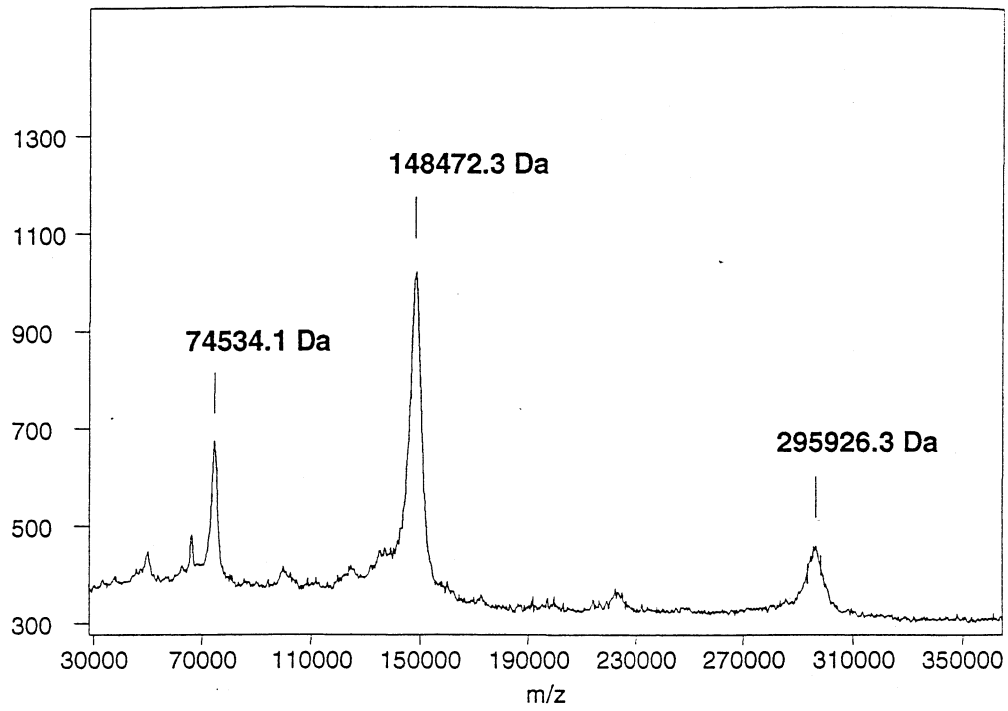

Fig. 3. MALDI-TOF/MS analysis of the HPLC peptide fraction eluting at 65.2 min.

ard, Waldbronn, Germany) using a Vydac C-18 column (1.0 × 250 mm). Separations were carried out at 40°C with a flow rate of 50 μl min^{-1}. Peptides were eluted with a linear gradient from 1 to 35% B within 80 min and to 98% B within 20 more min using 0.05% TFA as solvent A and 100% acetonitrile, 0.045% TFA as solvent B. Peptides were detected at 206 and 280 nm, respectively.

2.9. MALDI-TOF/MS analysis of collected peptide and protein fractions

All MALDI-TOF/MS measurements were performed using the HP G2025A instrument (Hewlett-Packard, Palo Alto, CA) which is equipped with a nitrogen laser. Aliquots (2 μl) of collected fractions were mixed with 3 μl of matrix and vacuum crystallized using the HP G2028A sample prep. accessory (Hewlett-Packard). For peptides, cyano-4-hydroxycinnamic acid was used as matrix and for proteins sinapinic acid with 0.1 M diammonium hydrogen citrate was used.

2.10. N-terminal protein sequencing

All samples were subjected to automated Edman degradation using the HP G1005A protein sequencing system (Hewlett-Packard). All samples were loaded directly onto the hydrophobic part of the biphasic sequencing device using the HP protein sequencing loading station HP G1001A. The same loading station is also used in this study also as a chromatography system as described above. Sequencing was performed using the routine 3.0 chemistry according to the manufacturers protocol.

3. Results and discussion

The simple chromatography system we have used in this study is shown in Fig. 1. It essentially consists of a commercially available protein sequencer sample loading station. Samples are passed over hydrophobic mini-columns fitted at

Table 1
Characterisation of the glycopeptide fraction at 65.2 min

Masses from potential glycan structures (Da)	Mass derived from peptide (Da)	Total mass (Da)	Mass determined by MALDI-TOF/MS (Da)
$Man_4GlcNAc$: 1136	2112.2	3248.2	3248.1
$Man_5GlcNAc$: 1314	2112.2	3426.2	3426.2

the top with a polypropylene funnel of 5 ml volume via nitrogen pressure. Pressure regulation and hence flow rate control is achieved with a simple manometer. This allows sample loading, desalting and concentration in one step. For the N-terminal sequencing of electroblotted samples or for the C-terminal sequence analysis of samples, empty mini-columns are used which are filled with the membrane pieces. These commercially available mini-columns were used in this study (see also Fig. 1).

Bubble-free column filling is simply achieved by first placing 100–200 μl of buffer in the funnel and by applying nitrogen pressure for 2–3 s to fill the column with buffer. Afterwards, a slurry of column material is placed into the funnel and with ~2 psi pressure, columns can be properly packed.

All washing and elution procedures are easily performed by pipetting a certain volume of appropriate buffers into the funnel and applying pressure to the column–funnel assembly. The use of an inert gas like nitrogen also avoids any oxidation process of peptides and proteins.

The purification of a high-mannose specific glycopeptide from a tryptic digest of the recombinant glycoprotein rt-PA is shown in Fig. 2. The rt-PA protein contains three glycosylation sites of which one carries a high-mannose glycan structure and is attached at the Asn residue at position 117 of the amino acid sequence [8]. The two other glycosylation sites carry complex glycan structures [8] and were not studied in this investigation. In Fig. 2(a), the complex RP-HPLC separation of the entire tryptic digest of r-tPA is shown. The HPLC separation of the fraction, which was specifically eluted from the Concanavalin A-Sepharose column using methyl-D-mannopyranoside, contained 2 major peptides eluting at 61.6 and 65.2 min, respectively (Fig. 2(b)).

Both peptide fractions were collected and subjected subsequently to protein sequencing and to MALDI-TOF/MS analysis. For both peptide fractions the same N-terminal sequence was found starting with Ser-105: S T A E S G A E C T N W × S S A L A Q K. Both peptide fractions contained the glycosylation site Asn-117, which gave no detectable amino acid at this degradation cycle, a well-known characteristic of amino acids carrying glycan structures. From the sequencing data, it also became obvious that the peptide resulted from a loss of residues from the expected N-terminus of the peptide. An earlier study has demonstrated that this microheterogeneity was due to a chymotryptic-like activity present in most if not all trypsin preparations and which resulted in cleavage of this glycopeptide which should have started with Gly-102 [8].

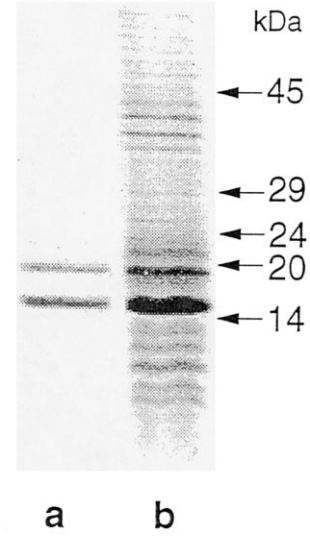

Fig. 4. SDS-PAGE of the crude *E. coli* extract (lane b) and the affinity purified poly-His tagged HMGa fraction (lane a).

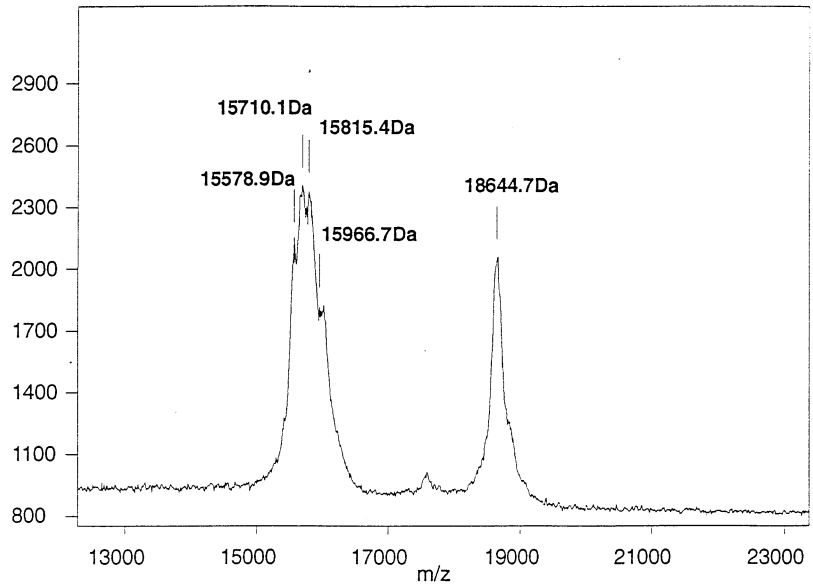

Fig. 5. MALDI-TOF/MS of the affinity purified HMGa preparation.

MALDI-TOF/MS analysis of both peptide fractions gave weak, but clearly detectable signals, especially for the peptide eluting at 65.2 min, as shown in Fig. 3. The two masses found for this peptide were 3248.1 and 3426.2 Da, respectively, suggesting that the difference in mass is due to one mannose subunit (178 Da). As summarized in Table 1, the mass information of the fraction eluting at 65.2 min is consistent with the presence of two glycopeptides which have the same amino acid sequence, but differ by one mannose residue: The following two glycosylation modifications of this peptide seems to be very likely, Man4 GlcNAc2 and Man5GlcNAc2, which is in good agreement with results previously published [6,9]

Another example of the use of the micro-column chromatography system was the micro-scale purification of the recombinant DNA-derived, binding HMGa (high mobility group) protein, with a poly-histidine tag. Generally, poly-His tagging of recombinant proteins allows a rapid, one-step purification via Ni^{2+}-chelating agarose [10]. HMGa was expressed in *E. coli* and the expected protein product would be a 18 600 Da polypeptide which contained 181 amino acids. The crude protein extract was passed over a Ni^{2+}-chelating agarose column to allow isolation of material that contained the poly-histidine tag. A Coommassie stained SDS-gel (Fig. 4, lane a and b) shows that the one step elution, using a buffer containing 25 mM EDTA, has allowed a selective purification of the His tagged recombinant HMGa protein. The analysis by gel electrophoresis, however, of both the crude cell lysate (part b) and the purified material (part a) shows two components (the full-length polypeptide of 18 kD and 3 kDa shorter polypeptide). Often rDNA processes result in undesired proteolytic processing of the product either from the N- or C-terminus and such a side reaction could explain the presence of two components. Of course, the second band could be due to the presence of an unrelated protein, although it would presumably need to also contain the poly-histidine sequence.

To further characterize the maize HMGa protein expressed in *E. coli*, an aliquot was subjected to N-terminal protein sequence analysis. The following single, expected amino–terminal amino acid sequence (M R G S H H H H H H H G S H M K G A K S K G A A K A D A) was obtained, which is consistent with an earlier sug-

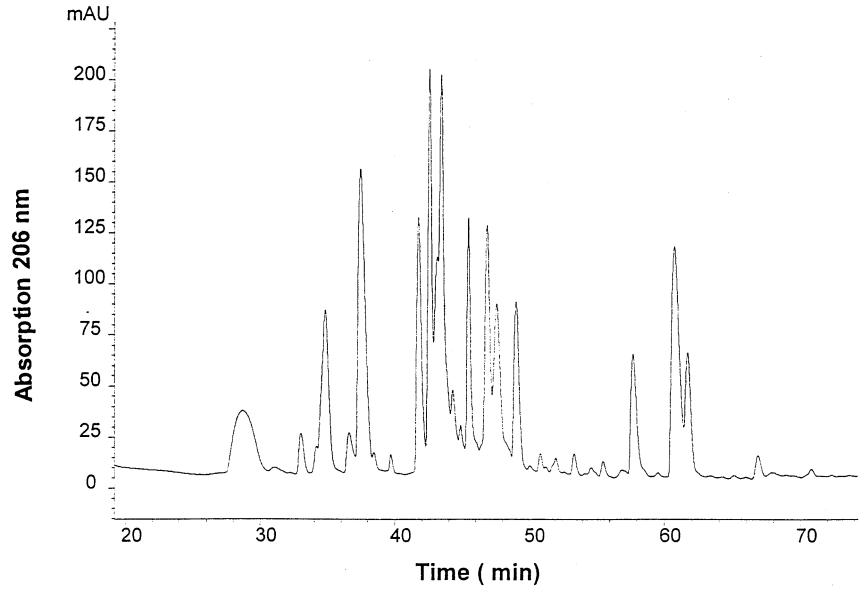

Fig. 6. Reversed-phase HPLC separation of the V8 digest of the recombinant HMGa protein preparation.

gestion (K.D. Grasser, unpublished data) of processing at the C-terminus of the HMGa protein.

MALDI-TOF/MS analysis of the purified HMGa fraction revealed two major components (Fig. 5). The larger protein fraction was found to have a molecular weight of 18644.7 Da which is ~10 Da less than the theoretical mass derived from the HMGa amino acid sequence. Since the measurement was performed with external calibration, the mass deviation of 0.05% is within accuracy of the method. The presence of a second component in the SDS-PAGE gel was confirmed by the presence of an additional peak at ~15700 Da in the MALDI-TOF/MS analysis. Interestingly, this peak was heterogeneous showing four different masses. This result could indicate C-terminal heterogeneity resulting from proteolytic processing of the HMGa protein at the C-terminal side. Polypeptide fragments containing deletions of 24, 25, 26 and 27 amino acid residues from the acidic C-terminus would be consistent with the mass data.

Table 2
Internal protein sequence analysis of HMGa after V8 cleavage

V8 fragment (min)	Sequence	Position
45.7	Y N K A I A A Y N K G E	102–113
57.9	K N P K N K S V A A V G K A A G D R W K S L E	62–85

The rest of the purified HMGa fraction was digested with endoproteinase V8 in order to obtain internal sequence information and confirm identification of the protein sample. Several peptide fractions were collected from the reversed-phase HPLC separation (Fig. 6) and two of the major subfractions could be completely sequenced to their C-terminal end. These peptides comprised positions 62–85 and 102–113 of the HMGa polypeptide chain (Table 2).

Another traditional purification method is the isolation of antibodies from serum by protein A Sepharose chromatography. This simple purification procedure can also be easily performed using the mini-column approach yielding highly purified antibodies, as can be seen from the MALDI-TOF/MS analysis (Fig. 7). The major mass peak of 148 kDa is characteristic of IgG antibodies, while the mass at 74 kDa corresponds to the

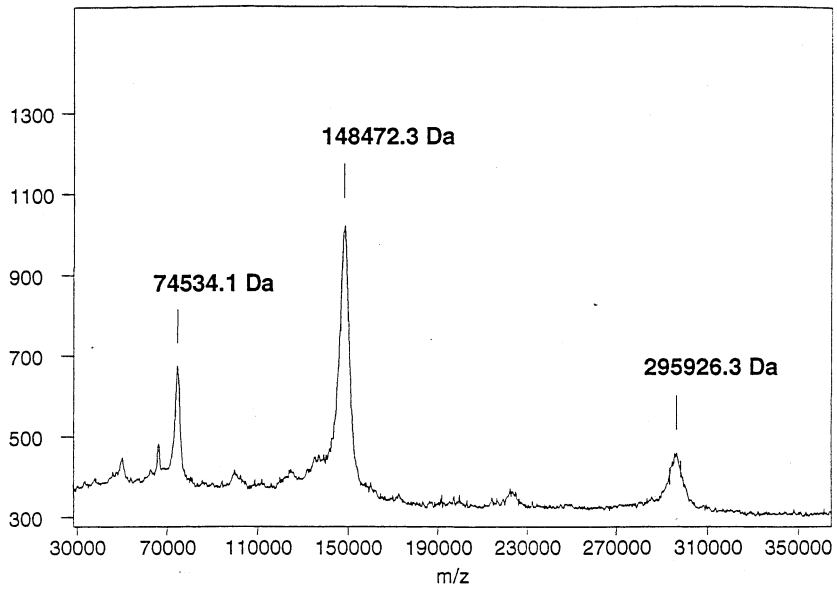

Fig. 7. MALDI-TOF/MS analysis of the purified antibody fraction.

doubly-charged species. The mass of ~300 kDa corresponds to a dimer of an IgG and is probably formed an artifact of the MS analysis. The mass at 65 kDa could be attributed to a minor impurity.

4. Conclusions

Using a wide range of samples, we have shown that this loading station can be used as an efficient, simple and cheap system for the micropurification of peptides and proteins. Columns can be easily self-packed using any column material. Flow rates can be regulated by varying the gas pressure via a simple manometer. Sufficient peptide and protein quantities can be obtained from single chromatographic procedures for further analysis by mass spectrometry, N-terminal sequence and internal sequence analysis. We feel that such technology will be important for characterization studies of post-translational modifications of proteins, particularly for samples with limiting starting material or for the study of low abundance proteins.

5. Further reading

K.D. Grasser, A.B. Krech, G. Feix, Plant J. 6 (1994) 351–358.

Acknowledgements

We are very thankful to Dr Billy Wu from Genentech (San Francisco, CA) for the generous gift of r-tPA. We also thank Drs Gordon Ross and Hubert Kaudewitz (Hewlett-Packard) for critical reading of the manuscript and Martin Verhoef (Hewlett-Packard) and John Michnowicz (Hewlett-Packard) for their continuous support to verify ideas. K.D. Grasser is supported by the Deutsche Forschungsgemeinschaft.

References

[1] J.P. Chervet, M. Ursem, J.P. Salzmann, R.W. Vannoort, J. High Resolut. Chromatogr. 5 (1989) 278–281.
[2] J.P. Chervet, C.J. Meijvogel, M. Ursem, J.P. Salzmann, LC.GC 2 (1992) 140–148.
[3] R.L. Moritz, R.J. Simpson, in: Methods in Protein Sequence analysis, Plenum, New York, 1993, pp. 3–10

[4] R. Grimm, M. Herold, J. Cap. Electrophor. 1 (1994) 74–77.
[5] R. Grimm, J. Cap. Electrophor. 2 (1995) 111–115.
[6] H.J. Boss, M.F. Rhode, R.S. Rush, Anal. Biochem. 230 (1995) 123–129.
[7] V. Ling, A.W. Guzzetta, E. Canova-Davis, W.S. Stults, W.S. Hancock, T.R. Covey, B.I. Shushan, Anal. Chem. 63 (1991) 2909–2915.
[8] W.S. Hancock, E. Canova-Davis, R.C. Chloupek, S.-I. Wu, I.P. Baldonado, J.E. Battersby, M.W. Spellman, L.J. Basa, J.A. Chakel, Therapeutic Peptides and Proteins: Assessing the New Technologies, Banbury Report 29, The Characterization of Degradation Products of Recombinant Human Growth Hormone, Cold Spring Harbor, 1988, pp. 95–117.
[9] A.W. Guzzetta, L.J. Basa, W.S. Hancock, B.A. Key, W.F. Bennett, Anal. Chem. 65 (1993) 2953–2962.
[10] D. Hess, T.C. Covey, R. Winz, R.W. Brownsey, R. Aebersold, Protein Sci. 2 (1993) 1342–1351.
[11] K.D. Grasser, G. Feix, Nucleic Acids Res. 19 (1991) 2573–2577.