Amino Acids (2008) 35: 485–494
DOI 10.1007/s00726-007-0551-y
Printed in The Netherlands

Amino Acids

Mass spectrometrical analysis of the processed metastasis-inducing *anterior gradient protein 2 homolog* reveals 100% sequence coverage

J.-K. Myung, T. Frischer, L. Afjehi-Sadat, A. Pollak, and **G. Lubec**

Department of Pediatrics, Medical University of Vienna, Vienna, Austria

Received January 28, 2007
Accepted March 23, 2007
Published online May 14, 2007;

Summary. *Anterior gradient protein 2 homolog* is a metastasis-inducing protein in a rat model of rat breast cancer and prognostic for outcome in hormonally treated breast cancer patients. Carrying out protein profiling in several mammalian cells and tissues, we detected this protein (synonym: secreted cement gland protein XAG-2 homolog) that was originally described in toad skin, in human bronchial epithelia.

Tissues obtained from biopsies were homogenised and extracted proteins were run on two-dimensional gel electrophoresis. Following in-gel digestion with proteases trypsin, AspN, LysC and chymotrypsin, mass spectrometrical analysis was carried out by MALDI-TOF/TOF.

The use of MS following multi-enzyme digestion of the protein resulted into 100% sequence coverage. MS/MS analysis enabled sequencing of 87% of the protein structure. This percentage does not include the signal peptide that was not observed in our protein due to processing. No post-translational modifications were detectable and no sequence conflicts were observed.

Complete analysis, unambiguous identification and characterisation of this biologically important protein could be shown, which is relevant for the definition of a marker protein that has been described so far by immunochemical methods only. Complete analysis is of importance as it forms the basis for all future work on this protein and, moreover, may serve as an analytical tool for further studies.

Keywords: *Anterior gradient protein 2 homolog* – Human bronchial epithelia – Sequencing – MALDI-TOF/TOF

Introduction

The *anterior gradient protein 2 homolog gene* (AGR2) is the human homologue of the *Xenopus laevis* cement gland gene, mapping to chromosome band 7p21.3 (Petek et al., 2000). Comparisons of this cloned cDNA sequences with the GenBank/EMBL databases using BLASTn indicated high similarity with the *Xenopus laevis* cement gland gene XAG-2 (Xenopus Anterior Gradient-2), with lower similarity to XAG-1 (Aberger et al., 1998; Sive et al., 1989).

The *anterior gradient protein 2 homolog*, hAG-2 (SwissProt Acc. No. O95994) consists of 175 amino acids with a molecular weight of approximately 19979.2 Da and the predicted p*I* value is 9.03. The protein sequence of hAG-2 was compared with *Xenopus laevis* homologue, the probable secreted protein XAG-2, (SwissProt Acc. No. P55869) and its mouse homologue, mAG-2 (SwissProt Acc. No. O88312), which was previously named GOB-4 (Komiya et al., 1999). The alignment revealed 54% identity and 71% similarity of AGR2 with XAG-2 and 91% identity and 96% similarity of AGR2 with mAG-2 protein. However, the predicted isoelectric point for probable secreted protein XAG-2 was 6.90 instead of 9.03 and 9.04 for both mAG-2 and AGR2, respectively.

XAG-2 has been reported to be a secretory protein (Aberger et al., 1998) and a signal peptide was identified in the first 18 amino acid residues and hAG-2 also has a predicted N-terminal cleavable secretory signal sequence (http://psort.nibb.ac.jp) (Adam et al., 2003). Computational analysis using SignalP (http://www.cbs.dtu.dk/services/SignalP) identified the first 20 amino acid residues of the protein as a secretory signal peptide for both AGR2 and mAG-2. The secretory property of AGR2 was investigated experimentally and the presence of a cleavable signal peptide was also supported (Zhang and Henzel, 2004; Zhang et al., 2005).

Although the precise biological roles and regulation of hAG-2 and mAG-2 are understood incompletely at this moment, some hints at their possible functions may be derived from the literature pertaining to the *Xenopus laevis* homologues XAG-2. In *Xenopus laevis*, XAG-2 has been shown to act as a signaling molecule and plays an important role in cement gland differentiation and ectodermal patterning (Aberger et al., 1998; Sive et al., 1989;

Sive and Bradley, 1996). XAG-2 is highly expressed in the cement gland, which consists of mucus-secreting cells functioning as endocrine organs. Overexpression of XAG-2 induces both, cement gland differentiation and expression of other anterior neural marker genes (Aberger et al., 1998; Zhang et al., 2005). hAG-2 has been found to be highly expressed in the cement gland, an ectodermal organ in the head associated with anteroposterior fate determination during early development (Fletcher et al., 2003), as well as in trachea, lung, stomach, colon, prostate and small intestine (Thompson and Weigel, 1998). It belongs to the AGR family and is known to interact with metastasis gene C4.4a and dystroglycan; and it is associated with estrogen receptor-positive breast tumours (Fletcher et al., 2003). By structural predictions hAG-2 may be involved in the epithelial barrier function from an evolutionary perspective considering of function as endocrine organs (Zheng et al., 2006) and in metastasis in a rat model of rat breast cancer and prognostic for outcome in hormonally treated breast cancer patients (Innes et al., 2006; Thompson and Weigel, 1998).

As the first step toward elucidating their functions in a species and providing a window into complex cellular regulatory networks, knowledge of the amino acid sequence and posttranslational modifications (PTMs) of the native isoforms is required. Chemical analysis methods focus on sequencing the N (Bailey, 1995) or C termini (Bergman et al., 2001; Li and Liang, 2002) but are unsuitable for rapid sample analysis.

With the development of matrix-assisted laser desorption/ionization (MALDI) (Karas and Hillenkamp, 1988) MS, proteolytic digestion, followed by MS and tandem mass spectrometry (MS/MS) this technique can be used either to measure simply the molecular mass of a polypeptide or to determine additional structural features including the amino acid sequence information or the site of attachment that facilitates primary structure confirmation and PTM mapping of known proteins as well as now allow rapid analysis of proteins with mass accuracies better than 0.02% (Domon and Aebersold, 2006).

Nucleotide sequences of hAG-2 were reported by Thompson and Weigel (1998), but hAG-2 was not fully characterised in protein chemical terms. Therefore, this study describes the first extensive mass spectrometric sequence mapping and characterization of full-length hAG-2. Proteins obtained from bronchial epithelial tissues by biopsies were separated by 2-DE and then subjected to proteolytic in-gel digestion with various proteases such as trypsin, LysC, AspN and chymotrypsin, and mass spectrometrical analysis was carried out by MALDI-TOF/TOF.

In this report, we were able to obtain complete sequences of hAG-2 with 100% of the protein sequence coverage by the use of multiple proteolytic cleavages. Complete analysis, unambiguous identification and characterisation of this biologically important protein could be shown. Complete analysis is of importance as it forms the basis for all future work on this protein and, moreover, may serve as an analytical tool for further studies.

Materials and methods

Patients

Three mucosal biopsies were taken under visual control from the right main stem bronchus or the carina of the right upper lobe and were immediately frozen in liquid nitrogen as described by Frischer et al. (2006). The freezing chain was never interrupted until analyses.

Sample preparation and tissue homogenisation

Bronchial tissue was suspended in 1.0 mL of sample buffer consisting of 7 M urea, 2 M thiourea, 4% CHAPS (Sigma, St. Louis, MO, USA), 65 mM DTT (Merck, Germany), 1 mM EDTA (Merck), and 1 mM PMSF and a mixture of protease inhibitors (protease inhibitor cocktail; Roche®, Basel, Switzerland). The suspension was sonicated for approximately 30 sec and the supernatant was centrifuged further at 14.000 rpm for 60 min to sediment undissolved material (Myung and Lubec, 2006).

Separation of proteins by 2-DE

Three hundred micrograms of protein were applied on immobilised pH 3–10 nonlinear gradient strips. Isoelectrofocusing step was performed until it reaches 100.000 Vh approximately as described by Frischer et al. (2006). After the first dimension, strips (18 cm) were equilibrated for 15 min in the buffer containing 6 M urea, 20% glycerol, 2% SDS, 2% DTT and then for 15 min in the same buffer containing 2.5% iodoacetamide instead of DTT. After equilibration, strips were loaded on 9–16% gradient SDS polyacrylamide gels ($180 \times 200 \times 1.5$ mm) and were run at 40 mA per gel for second-dimensional separation. Immediately after the second-dimension run, gels were fixed for 12 h in 50% methanol containing 10% acetic acid and stained with Colloidal Coomassie Blue (Novex, San Diego, CA) for 12 h on a rocking shaker (John et al., 2007).

Molecular masses were determined by running standard protein markers (BioRad Laboratories, Hercules, CA), covering the range 10–250 kDa. p*I* values were used as given by the supplier of the IPG strips (Amersham Bioscience, Uppsala, Sweden). Excess of dye was washed out from the gels with distilled water, and gels were scanned with Imagescanner (Amersham Bioscience).

Mass spectrometrical analysis

Sample preparation

Protein Spots stained by Colloidal Coomassie Blue were excised with a spot picker (PROTEINEER sp™, Bruker Daltonics, Bremen, Germany) with 1.5 mm diameter of cannulae and washed with 10 mM ammonium bicarbonate and 50% acetonitrile in 10 mM ammonium bicarbonate promoting metastable fragmentation. After washing, gel plugs were shrunk by addition of acetonitrile and dried by blowing out the liquid through the pierced well bottom.

Dried gel pieces were reswollen with four different kinds of enzymes and incubated depending on the conditions according to enzymes. First, the dried gel piece was digested with 40 µg/µL trypsin (Promega, Madison, WI) in 5 mM octyl-β-D-glucopyranoside (OGP) and 10 mM ammonium bicarbonate. The reaction was carried out at 30 °C for 4 h. The gel piece was also subjected to LysC treatment. 30 ng/µL of porcine LysC (sequencing grade; Roche Diagnostic, Mannheim, Germany) in 5 mM ammonium bicarbonate and 5 mM OGP was added to the gel piece and incubated for 18 h at 30 °C as described (John et al., 2006). Chymotrypsin and AspN digestion were also performed by addition of 25 ng/µL chymotrypsin (sequencing grade, Roche Diagnostic) and 25 ng/µL AspN (sequencing grade, Roche Diagnostic) in 25 mM ammonium bicarbonate containing 5 mM OGP and carried out for 2.5 h at 30 °C and for 18 h at 37 °C, respectively. Digested peptides were extracted with with 10 µL of 1% TFA (trifluoroacetic acid) in 5 mM OGP.

Protein identification by MALDI-TOF/TOF

Extracted peptides were directly applied onto a target (AnchorChip™, Bruker Daltonics) that was load with CHCA (α-cyano-4-hydroxy-cinnamic acid) (Bruker Daltonics) matrix thinlayer. The mass spectrometer used in this work was an Ultraflex™ TOF/TOF (Bruker Daltonics) operated in the reflector for MALDI-TOF peptide mass fingerprint (PMF) or LIFT mode for MALDI-TOF/TOF using the FlexControl™ software. An accelerating voltage of 25 kV was used for PMF. Calibration of the instrument was performed externally with $[M+H]^+$ ions of angiotensin I, angiotensin II, substance P, bombesin, and adrenocorticotropic hormones (clip 1–17 and clip 18–39). All MALDI spectra were internally calibrated with tryptic auto digest ions. Each spectrum was produced by accumulating data from 200 consecutive laser shots. Those samples which were analysed by PMF from MALDI-TOF were additionally analysed using MALDI-TOF/TOF from the same target. A maximum of precursor ions per sample were chosen for MS/MS analysis. In the TOF1 stage, all ions were accelerated to 8 kV under conditions promoting metastable fragmentation. After selection of jointly migrating parent and fragment ions in a timed ion gate, ions were lifted by 19 kV to high potential energy in the LIFT cell. After further acceleration of the fragment ions in the second ion source, their masses could be simultaneously analysed in the reflector with high sensitivity.

System for identification of proteins

Peptide masses were matched with the theoretical peptide masses of all proteins from human species. PMF and LIFT spectra obtained were interpreted manually or by database searching using the in-house licensed database search engine Mascot (John et al., 2007) (Matrix Science Ltd., London, UK) with the combination of PMF and MS/MS datasets via BioTools 2.2 software (Bruker Daltonics). The probability score calculated by the software was used as criteria for correct identification (http://www.matrixscience.com).

Results

Expression of the anterior gradient protein 2 homolog, hAG-2

Analysis of the protein sequences of the *anterior gradient protein 2 homolog*, hAG-2 predicted to have cleavable 20 amino acid signal peptides (Nakai and Kanehisa, 1992). The theoretical molecular weight of the unprocessed form is 19979.20 Da and the p*I* 9.03 (calculated in http://

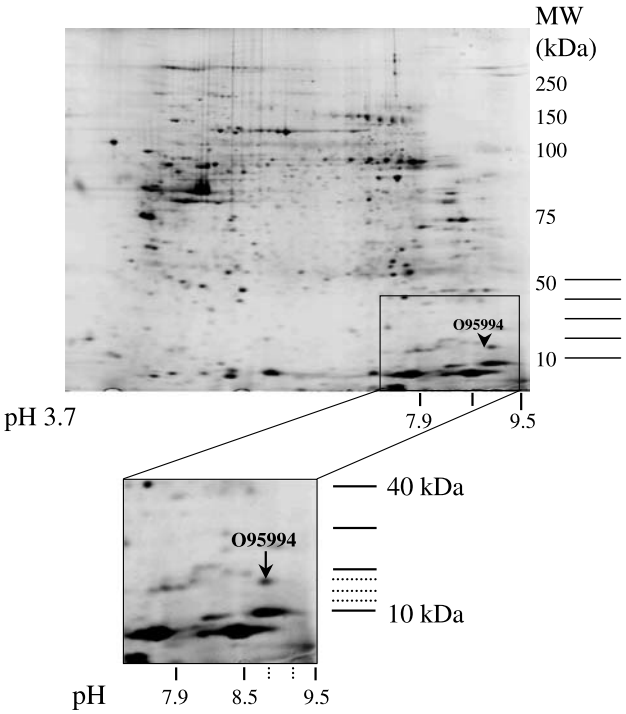

Fig. 1. The reference map of human bronchial epithelial tissue is presented. Anterior gradient protein 2 homolog, hAG-2, stained with coomassie-colloidal blue on a 9–16% SDS-PAGE is identified by mass spectrometry. The arrow indicates hAG-2 (SwissProt Acc. No. O95994)

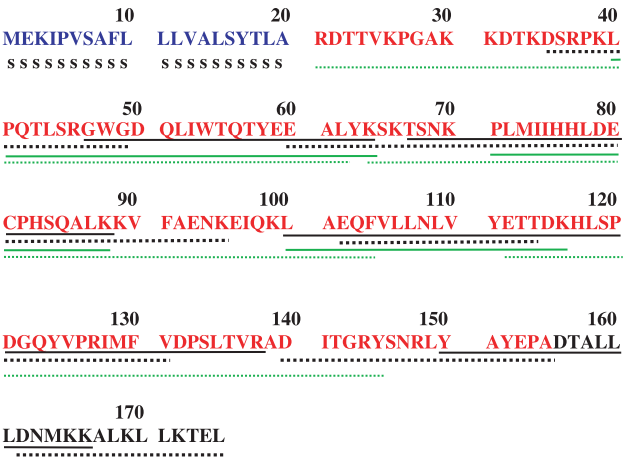

Fig. 2. Amino acid sequence of anterior gradient protein 2 homolog, hAG-2, from human (SwissProt Acc. No. O95994) obtained from chemical sequence analysis generated by trypsin, LysC, AspN and chymotrypsin digestion. The calculated and experimental masses are given in Table 1. Individual sequence coverages obtained by proteolytic degradation with trypsin (solid black line), LysC (solid green line), AspN (dot black line) and chymotrypsin (dot green line) are shown that together finally achieved to 100% (155/155 aa) of sequence coverage by MS and 88.6% (135/155 aa) by MSMS except signal peptide part. The sequences identified with MSMS are shown as red characters and the sequences of signal peptide are as blue and annotated "s"

www.expasy.org/tools/pi_tool.html). The predicted isoelectric point for the mature portions of the hAG-2 protein is 9.06 and the molecular weight is 17817.53 Da (calculated in http://www.expasy.org/tools/pi_tool.html). hAG-2 was expressed on 2D gel showing a similar theoretical molecular weight and p*I* values of the processed form (Fig. 1).

Mass spectrometric characterization of full-length hAG-2

To characterise the full-length sequence including the probable sites of PTMs in hAG-2, we digested the hAG-2 protein with trypsin, LysC, AspN or chymotrypsin and subjected the resulting peptides to MALDI-TOF and MALDI-TOF/TOF analysis. MS can give sequence data from the Colloidal Coomassie Blue stained gel spot. The protein is digested with various proteases in the gel and digested peptides are eluted and introduced into MALDI-TOF/TOF. The mass spectrometer determines the mass of the peptides and the sequence from the masses of the peptide fragments and sequence data is determined by comparison with known sequences.

The graphic display of hAG-2 sequence coverage obtained by proteolytic degradation by trypsin, LysC, AspN and chymotrypsin is shown (Fig. 2) revealing 100% of sequence coverage. A complete list of all hAG-2 peptides

Table 1. Measured and calculated protonated $[M+H]^+$ molecular masses and sequences of peptides determined from anterior gradient protein 2 homolog (hAG-2) by MALDI-TOF/TOF analysis

Start–end	Observed	Mr (expt)[1]	Mr (calc)[2]	Delta	Miss[3]	Sequence	Enz[4]
21–44	2680.41	2679.40	2679.50	−0.10	0	A.RDTTVKPGAKKDTKDSRPKLPQTL.S (MSMS)[5]	**C**
35–49	1697.88	1696.88	1696.90	−0.02	0	K.DSRPKLPQTLSRGWG.D (MSMS)	**A**
40–64	2996.49	2995.48	2995.51	−0.03	0	K.LPQTLSRGWGDQLIWTQTYEEALYK.S (MSMS)	**L**
45–54	1217.57	1216.56	1216.60	−0.03	2	L.SRGWGDQLIW.T	**C**
47–64	2201.10	2200.09	2200.05	0.04	0	R.GWGDQLIWTQTYEEALYK.S	**T**
53–62	1253.56	1252.56	1252.60	−0.04	2	L.IWTQTYEEAL.Y	**C**
59–78	2352.28	2351.27	2351.27	0.01	0	Y.EEALYKSKTSNKPLMIIHHL.D (MSMS)	**A**
59–78	2368.21	2367.20	2367.26	−0.06	1	Y.EEALYKSKTSNKPLMIIHHL.D + Oxidation (M) (MSMS)	**A**
64–87	2800.31	2799.30	2799.42	−0.11	2	Y.KSKTSNKPLMIIHHLDECPHSQAL.K + Oxidation (M)	**C**
67–88	2569.33	2568.32	2568.29	0.03	0	K.TSNKPLMIIHHLDECPHSQALK.K	**T**
67–88	2585.34	2584.33	2584.29	0.04	0	K.TSNKPLMIIHHLDECPHSQALK.K + Oxidation (M)	**T**
71–88	2139.06	2138.06	2138.08	−0.02	0	K.PLMIIHHLDECPHSQALK.K (MSMS)	**L**
71–88	2155.06	2154.06	2154.07	−0.01	0	K.PLMIIHHLDECPHSQALK.K + Oxidation (M)	**L**
73–87	1800.80	1799.79	1799.84	−0.05	1	L.MIIHHLDECPHSQAL.K	**C**
73–87	1816.79	1815.79	1815.84	−0.05	1	L.MIIHHLDECPHSQAL.K + Oxidation (M)	**C**
79–92	1629.78	1628.77	1628.80	−0.03	1	L.DECPHSQALKKVFA.E	**A**
79–95	2001.00	2000.00	1999.98	0.02	1	L.DECPHSQALKKVFAENK.E (MSMS)	**A**
88–104	2050.07	2049.06	2049.13	−0.07	2	L.KKVFAENKEIQKLAEQF.V (MSMS)	**C**
92–104	1547.77	1546.76	1546.80	−0.04	1	F.AENKEIQKLAEQF.V	**C**
100–116	1996.05	1995.04	1995.06	−0.01	0	K.LAEQFVLLNLVYETTDK.H (MSMS)	**T/L**
102–111	1237.68	1236.67	1236.68	−0.01	0	A.EQFVLLNLVY.E	**A**
102–114	1568.80	1567.79	1567.81	−0.02	1	A.EQFVLLNLVYETT.D	**A**
112–130	2250.01	2249.00	2249.08	−0.08	2	Y.ETTDKHLSPDGQYVPRIMF.V + Oxidation (M)	**C**
117–127	1268.64	1267.63	1267.63	0.00	0	K.HLSPDGQYVPR.I (MSMS)	**T**
121–131	1324.65	1323.64	1323.66	−0.02	0	P.DGQYVPRIMFV.D (MSMS)	**A**
121–131	1340.65	1339.64	1339.66	−0.02	0	P.DGQYVPRIMFV.D + Oxidation (M) (MSMS)	**A**
128–138	1293.67	1292.66	1292.68	−0.02	0	R.IMFVDPSLTVR.A + Oxidation (M)	**T**
131–145	1662.85	1661.84	1661.87	−0.04	1	F.VDPSLTVRADITGRY.S (MSMS)	**C**
140–152	1591.77	1590.76	1590.78	−0.02	0	A.DITGRYSNRLYAY.E (MSMS)	**A**
140–155	1888.88	1887.88	1887.91	−0.04	1	A.DITGRYSNRLYAYEPA.D (MSMS)	**A**
149–165	1940.93	1939.93	1939.96	−0.03	0	R.LYAYEPADTALLLDNMK.K	**T**
149–165	1956.94	1955.93	1955.95	−0.03	0	R.LYAYEPADTALLLDNMK.K + Oxidation (M)	**T**
149–166	2085.03	2084.02	2084.05	−0.03	0	LYAYEPADTALLLDNMKK + Oxidation (M)	**T**
162–175	1644.94	1643.94	1643.96	−0.03	1	L.DNMKKALKLLKTEL.-[6]	**A**

[1] Mr (expt): Experimental molecular weight
[2] Mr (calc): Calculated molecular weight
[3] Miss: Missed cleavage number
[4] Enz: Used enzyme T, trypsin; A, AspN; L, LysC; C, chymotrypsin
[5] MS/MS: The peptides analysed by MALDI-TOF/TOF
[6] C-terminal amino acid sequence of hAG-2

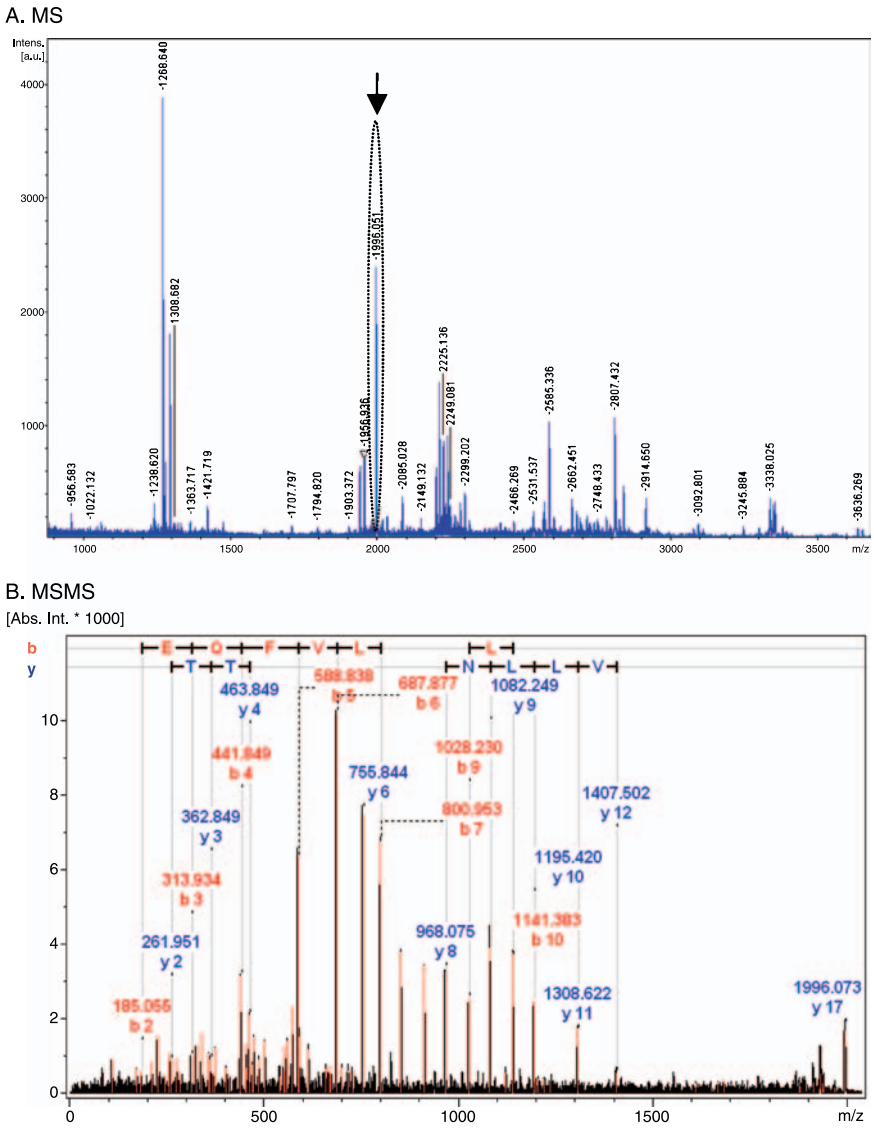

Fig. 3. MS-TOF spectrum of the anterior gradient protein 2 homolog, hAG-2 digested by trypsin (**A**) and LIFT-TOF/TOF spectrum of m/z 1996.051 (**B**) that unambiguously assigned the identified sequence LAEQFVLLNL VYETTDK to hAG-2 (SwissProt Acc. No. O95994) (for a colour reproduction of this figure, the reader is referred to the web version of this paper under www.springerlink.com)

that were sequenced and confirmed by MS with calculated and experimental masses is summarised in Table 1.

Trypsin digestion

To get the protein sequence, MALDI-TOF/TOF was used to analyze tryptic digests of hAG-2. Observed ions were searched with MASCOT against SwissProt database. As shown in Fig. 3A, protonated molecular ions, $[M+H]^+$, for 9 peptides were detected by MALDI-MS (i.e., PMF) for hAG-2 after tryptic digestion, yielding an overall protein sequence coverage of 55.0% (97/175 amino acid). Two peptides were analyzed by MALDI-TOF/TOF leading to confirmation of protein sequence by subsequent peptide mapping and one spectrum of m/z 1996.05 was shown in Fig. 3B.

LysC digestion

hAG-2 was digested with the protease LysC that cleaves C-terminal to Lys residue. Figure 4A and B show the MS and MS/MS fragmentation spectrum of precursor ion 2139.063 with the sequence PLMIIHHLDECPHSQALK to hAG-2. Peptide mapping and MS/MS measurements on the LysC digested fragments led to 34.3% (60/175 amino acid) sequence coverage and both trypsin and LysC digested peptides covered 59.4% (104/175 amino acid) sequence showing the overlapping of some peptides. Theoretical cleavage of LysC on hAG-2 gives the information of the sequence of N-terminal (4–26), IPVSAFLLLVALSYTLARDTTVK, at m/z 2491.4534 if the signal peptide was not processed but we could not obtain any fragment.

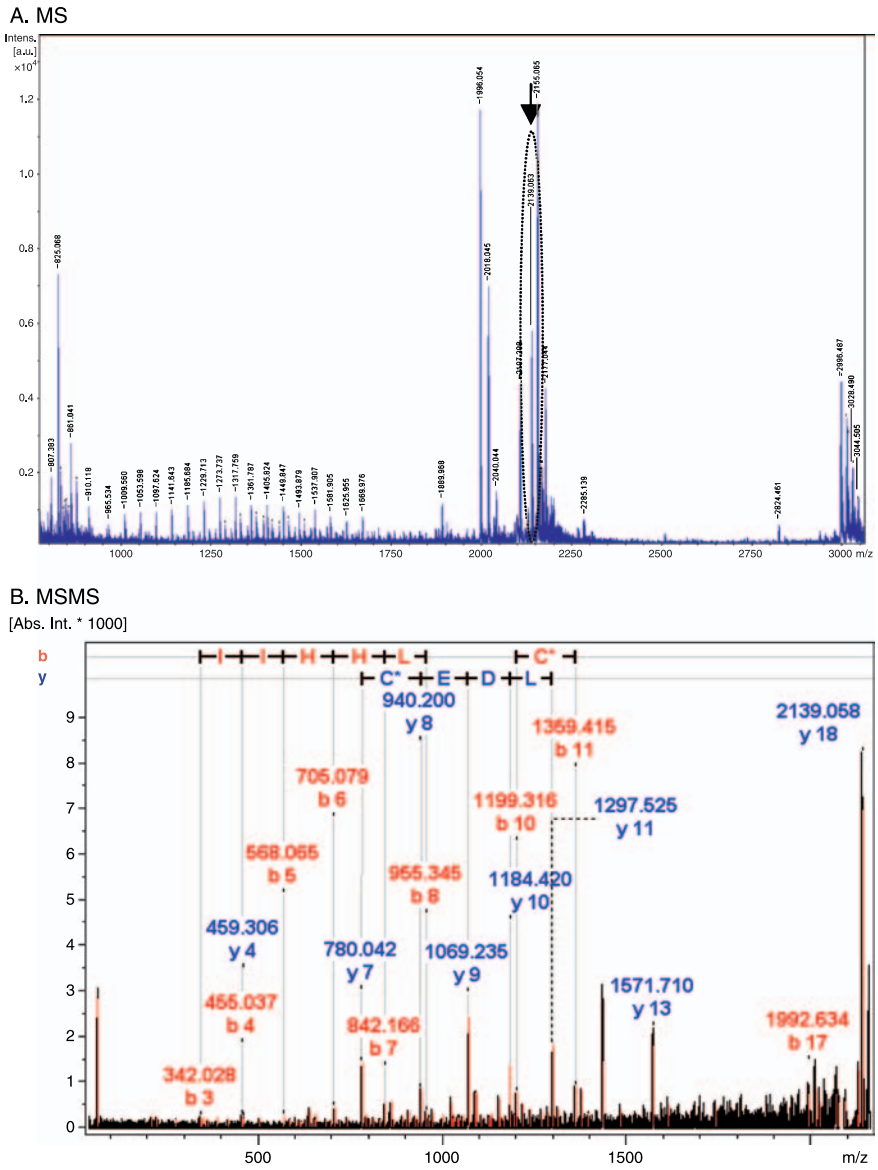

Fig. 4. MS-TOF spectrum of the anterior gradient protein 2 homolog, hAG-2 digested by LysC (**A**) and LIFT-TOF/TOF spectrum of m/z 2139.063 (**B**) that unambiguously assigned the identified sequence PLMIIHHLDECPHSQALK to hAG-2 (SwissProt Acc. No. O95994) (for a colour reproduction of this figure, the reader is referred to the web version of this paper under www.springerlink.com)

AspN digestion

Both trypsin and LysC digestion were not sufficient to provide the whole sequence information of hAG-2. N-terminal peptides of 39 amino acids including signal peptide part, small peptides (amino acids 65–66, 89–99 and 139–148) and C-terminal peptide (167–175) were missing. The protease AspN that cleaves N-terminal to Asp and Glu residues was used and identified most sequences that were previously unidentified. MALDI-TOF and MALDI-TOF/TOF TOF analysis resulted in 60.6% (106/175 amino acid) sequence coverage and three different enzymes made it possible to cover 77.7% (136/175 amino acid) sequences.

An oxidised peptide was also detected showing the change in mass of 16 Da. We could get the information of the oxidation of methionine residue by the detection of a 16 Da mass increase between m/z 1324.65 and m/z 1340.65 peptide. The MS-TOF spectrum of the hAG-2 digested by AspN is shown in Fig. 5A and three LIFT-TOF/TOF spectra including a methionine oxidised peptide are shown in Fig. 5B, C and D.

Chymotrypsin digestion

In order to enhance sequence coverage, especially for N-terminal sequences, chymotryptic digestion was carried out. The digestion was performed for 2.5 h to prevent peptides to be cleaved into short peptides and the masses of the peptide fragments were measured by MALDI-TOF/TOF. Chymotryptic digestion of hAG-2 generated

10 peptides yielding an overall protein sequence coverage of 66.9%. Though some peptides are overlapping with trypsin, LysC or AspN digested peptides, this overlapping of identification peptides with various enzymes gives more confirmation of sequence information. The spectra of MS and one MALDI-MS/MS were represented in Fig. 6A and B and observed fragments are listed in Table 1.

The processing (absence) of signal peptide was also confirmed by MALDI-MS/MS analysis. It was confirmed by sequencing of N-terminal peptide starting from 21 to 44 amino acids digested by chymotrypsin (m/z 2680.409). Theoretically, chymotrypsin preferentially cleaves at Trp, Tyr and Phe (high specificity) and to a lesser extent (taken into account when dealing with low specificity chymotrypsin) at Leu, Met and His in C-terminal (Keil, 1992). If the signal peptide is not processed, the sequence of peptide digested with chymotrypsin should start at the 18th amino acid, but MSMS data of m/z 2680.409 shows that the sequence of this protein starts at the 21th amino acid, which shows that the signal peptide is processed (Fig. 7).

Discussion

The major outcome of the study is to demonstrate the successful use of multi-enzyme digestion using trypsin, LysC, AspN and chymotrypsin for the complete sequence analysis of a given protein. This is relatively time con-

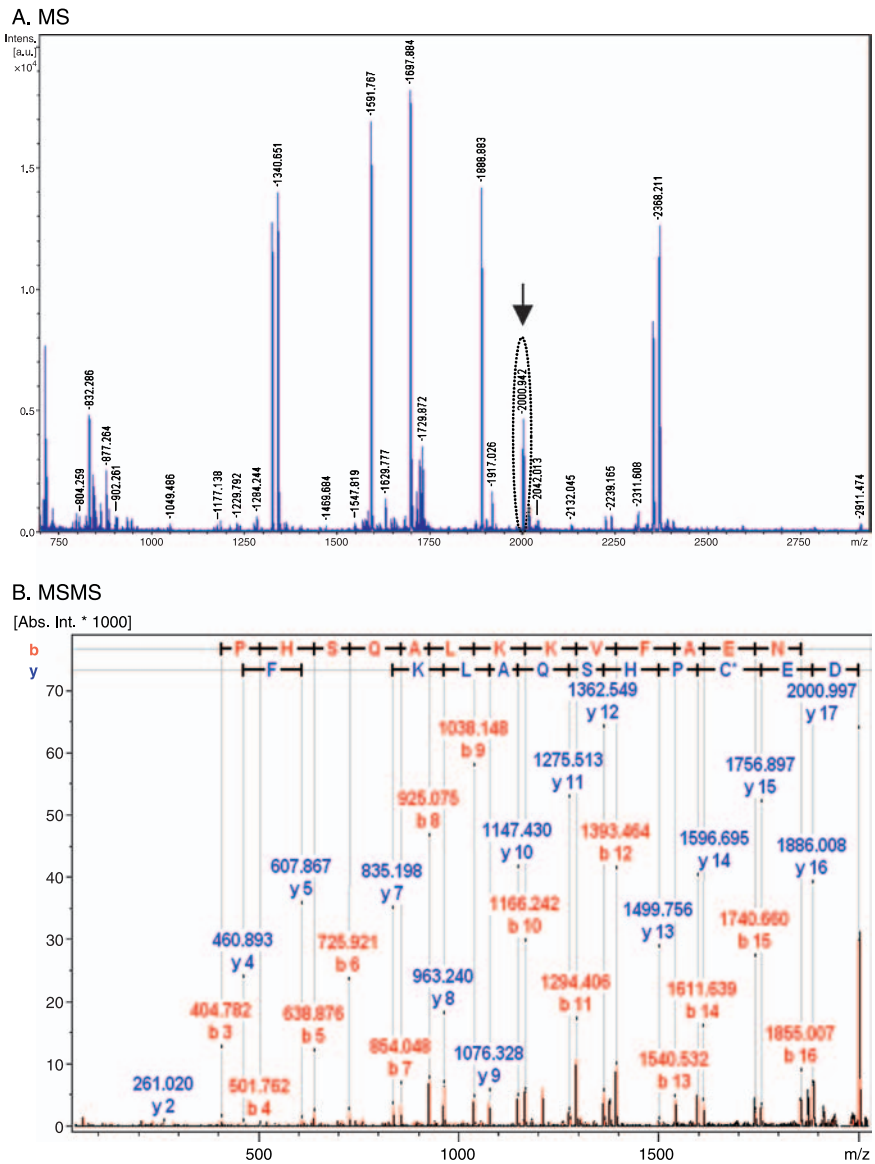

Fig. 5. MS-TOF spectrum of the anterior gradient protein 2 homolog, hAG-2 digested by AspN (**A**) and LIFT-TOF/TOF spectrum of m/z 2000.997 (**B**) that unambiguously assigned the identified sequence DECPHS QALKKVFAENK to hAG-2 (SwissProt Acc. No. O95994) (for a colour reproduction of this figure, the reader is referred to the web version of this paper under www.springerlink.com). LIFT-TOF/TOF spectra of m/z 1340.666 (**C**) and m/z 2352.293 (**D**) of the anterior gradient protein 2 homolog, hAG-2 (SwissProt Acc. No. O95994) digested by AspN, which shows the methionine oxidized peptide DGQYV PRIM*FV and the peptide EEALYKSKT SNKPLMIIHHL, respectively (for a colour reproduction of this figure, the reader is referred to the web version of this paper under www.springerlink.com)

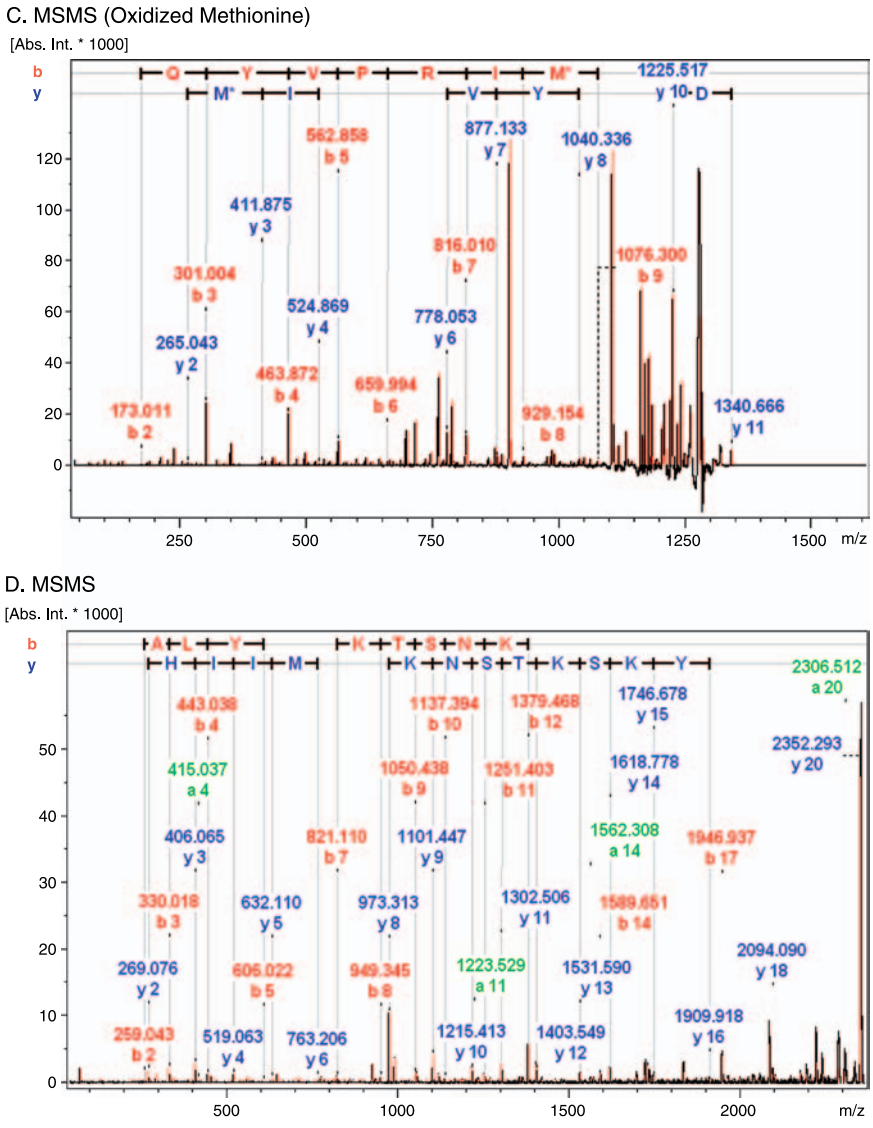

Fig. 5 (*continued*)

suming and expensive but fragmentations carried out by instrumentation do not lead to complete sequence coverage in the majority of cases. Knowledge of the full protein sequence is mandatory to fully characterise and identify particular proteins that are in diagnostic use, as it is the case for *anterior gradient protein 2 homolog*, hAG-2 (Innes et al., 2006; Thompson and Weigel, 1998). It is virtually impossible to reliably design an antibody simply based upon prediction from nucleic acid sequences and, moreover, most proteins present with isoforms or splice variants that would not necessarily be detected at the protein level otherwise. In addition, PTMs would have to be considered as it is well-known and documented that antibodies may not react with modified proteins. In our case no PTMs were observed. This does not rule out, however, that the protein is not modified because it may well be that during the analytical procedure PTMs were destroyed or cleaved. As samples were processed under denaturing conditions removal of phosphorylation by phosphatases is highly unlikely, and, in addition analysis with and without phosphatase inhibitors do not show different protein phosphorylation (own preliminary results, unpublished).

Full sequencing information is also important, when database conflicts have to be considered and indeed, no data base conflicts were observed for this protein, thus representing database validation and confirmation. Full sequencing is also mandatory when hypothetical proteins, i.e. structures that have been reported at the nucleic acid sequence without experimental work, are analysed (Lubec et al., 2005).

Taken together, a potentially important marker protein has been analysed at the chemical level by a well-docu-

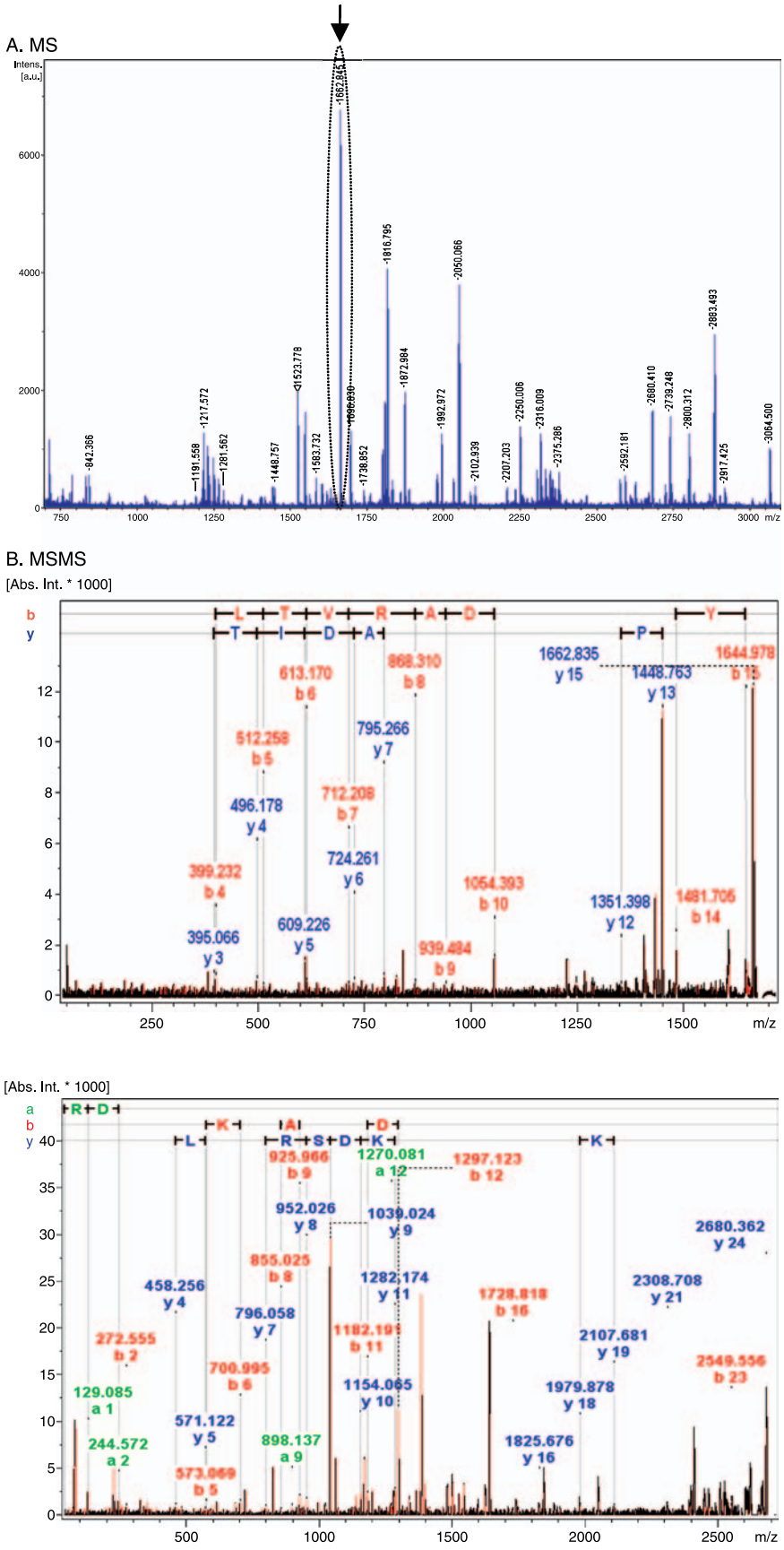

Fig. 6. MS-TOF spectrum of the anterior gradient protein 2 homolog, hAG-2 digested by chymotrypsin (**A**) and LIFT-TOF/TOF spectrum of m/z 1662.845 (**B**) that unambiguously assigned the identified sequence VDPSLTVRADITGRY to hAG-2 (SwissProt Acc. No. O95994) (for a colour reproduction of this figure, the reader is referred to the web version of this paper under www.springerlink.com)

Fig. 7. LIFT-TOF/TOF spectrum of m/z 2680.41 of the anterior gradient protein 2 homolog, hAG-2 digested by chymotrypsin that unambiguously assigned the identified sequence RDTTVKPGAKKDTKDSRPKL PQTL to N-terminal of hAG-2 showing the processing of signal peptide (for a colour reproduction of this figure, the reader is referred to the web version of this paper under www.springerlink.com)

mented and reliable method (Afjehi-Sadat et al., 2007; Gruber-Olipitz et al., 2006; John et al., 2007; Lubec et al., 2005; Pollak et al., 2006; Weitzdorfer et al., 2006) and the sequence can be reliably used for further studies, including immunochemical techniques. The value of multi-enzyme digestion with a combination of four proteases for reaching highest sequence coverage is recommended.

Acknowledgements

We are highly indebted to the Verein zur Durchführung der wissenschaftlichen Forschung auf dem Gebiet der Neonatologie und Kinderintensivmedizin (“Unser Kind”).

References

Aberger F, Weidinger G, Grunz H, Richter K (1998) Anterior specification of embryonic ectoderm: the role of the Xenopus cement gland-specific gene XAG-2. Mech Dev 72: 115–130

Adam PJ, Boyd R, Tyson K, Fletcher GC, Stamps AC, Hudson L, Poyser HR, Redpath N, Griffiths M, Steers G, Harris AL, Patel S, Berry J, Loader JA, Townsend RR, Daviet L, Legrain P, Parekh R, Terrett JA (2003) Comprehensive proteomic analysis of breast cancer cell membranes reveals unique proteins with potential roles in clinical cancer. J Biol Chem 21: 6482–6489

Afjehi-Sadat L, Yang JW, Pollak A, Kim DW, Choi SY, Lubec G (2007) Structural and functional analysis of hypothetical proteins in mouse hippocampus from two-dimensional gel electrophoresis. J Proteome Res 6: 711–723

Bailey JM (1995) Chemical methods of protein sequence analysis. J Chromatogr A 705: 47–65

Bergman T, Cederlund E, Jornvall H (2001) Chemical C-terminal protein sequence analysis: improved sensitivity, length of degradation, proline passage, and combination with Edman degradation. Anal Biochem 290: 74–82

Domon B, Aebersold R (2006) Mass spectrometry and protein analysis. Science 312: 212–217

Fletcher GC, Patel S, Tyson K, Adam PJ, Schenker M, Loader JA, Daviet L, Legrain P, Parekh R, Harris AL, Terrett JA (2003) hAG-2 and hAG-3, human homologues of genes involved in differentiation, are associated with estrogen receptor-positive breast tumours and interact with metastasis gene C4.4a and dystroglycan. Br J Cancer 24: 579–585

Frischer T, Myung JK, Maurer G, Eichler I, Szepfalusi Z, Lubec G (2006) Possible dysregulation of chaperon and metabolic proteins in cystic fibrosis bronchial tissue. Proteomics 6: 3381–3388

Gruber-Olipitz M, Yang JW, Slavc I, Lubec G (2006) Nectin-like molecule 1 is a high abundance protein in cerebellar neurons. Amino Acids 30: 409–415

Innes HE, Liu D, Barraclough R, Davies MP, O’Neill PA, Platt-Higgins A, de Silva Rudland S, Sibson DR, Rudland PS (2006) Significance of the metastasis-inducing protein AGR2 for outcome in hormonally treated breast cancer patients. Br J Cancer 10: 1057–1065

John JP, Anrather D, Pollak A, Lubec G (2006) Mass spectrometrical verification of stomatin-like protein 2 (SLP-2) primary structure. Proteins 1: 543–551

John JP, Pintsov O, Petter-Puchner A, Redl H, Pollak A, Chen WQ, Lubec G (2007) Nitric oxide and oxygen radical attack on GDP-Dissociation Inhibitor 2 (GDI-2) in spinal cord injury of the rat. J Proteome Res 6: 1500–1509

Karas M, Hillenkamp F (1988) Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. Anal Chem 60: 2299–2301

Keil B (1992) Proteolysis data bank: specificity of alpha-chymotrypsin from computation of protein cleavages. Protein Seq Data Anal 1: 13–20

Komiya T, Tanigawa Y, Hirohashi S (1999) Cloning of the gene gob-4, which is expressed in intestinal goblet cells in mice. Biochim Biophys Acta 19: 434–438

Li J, Liang S (2002) C-terminal sequence analysis of peptides using triphenylgermanyl isothiocyanate. Anal Biochem 302: 108–113

Lubec G, Afjehi-Sadat L, Yang JW, John JP (2005) Searching for hypothetical proteins: theory and practice based upon original data and literature. Prog Neurobiol 77: 90–127

Myung JK, Lubec G (2006) Use of solution-IEF-fractionation leads to separation of 2673 mouse brain proteins including 255 hydrophobic structures. J Proteome Res 5: 1267–1275

Nakai K, Kanehisa M (1992) A knowledge base for predicting protein localisation sites in eukaryotic cells. Genomics 14: 897–911

Petek E, Windpassinger C, Egger H, Kroisel PM, Wagner K (2000) Localization of the human anterior gradient-2 gene (AGR2) to chromosome band 7p21.3 by radiation hybrid mapping and fluorescence in situ hybridisation. Cytogenet Cell Genet 89: 141–142

Pollak DD, John J, Hoeger H, Lubec G (2006) An integrated map of the murine hippocampal proteome based upon five mouse strains. Electrophoresis 27: 2787–2798

Sive H, Bradley L (1996) A sticky problem: the Xenopus cement gland as a paradigm for anteroposterior patterning. Dev Dyn 205: 265–280

Sive HL, Hattori K, Weintraub H (1989) Progressive determination during formation of the anteroposterior axis in Xenopus laevis. Cell 58: 171–180

Thompson DA, Weigel RJ (1998) hAG-2, the human homologue of the Xenopus laevis cement gland gene XAG-2, is coexpressed with estrogen receptor in breast cancer cell lines. Biochem Biophys Res Commun 9: 111–116

Weitzdorfer R, Hoger H, Pollak A, Lubec G (2006) Changes of hippocampal protein levels during postnatal brain development in the rat. J Proteome Res 5: 3205–3212

Zhang JS, Gong A, Cheville JC, Smith DI, Young CY (2005) AGR2, an androgen-inducible secretory protein overexpressed in prostate cancer. Gene Chromosome Canc 43: 249–259

Zhang Z, Henzel WJ (2004) Signal peptide prediction based on analysis of experimentally verified cleavage sites. Protein Sci 13: 2819–2824

Zheng W, Rosenstiel P, Huse K, Sina C, Valentonyte R, Mah N, Zeitlmann L, Grosse J, Ruf N, Nurnberg P, Costello CM, Onnie C, Mathew C, Platzer M, Schreiber S, Hampe J (2006) Evaluation of AGR2 and AGR3 as candidate genes for inflammatory bowel disease. Genes Immun 7: 11–18

Authors’ address: Prof. Dr. Gert Lubec, Department of Pediatrics, Medical University of Vienna, Währinger Gürtel 18, A-1090 Vienna, Austria, Fax: +43-1-40400-6065, E-mail: gert.lubec@meduniwien.ac.at