ELSEVIER

Journal of Chromatography B, 662 (1994) 279–290

JOURNAL OF CHROMATOGRAPHY B: BIOMEDICAL APPLICATIONS

Analysis of the primary structure and post-translational modifications of the *Schistosoma mansoni* antigen *Sm*p28 by electrospray mass spectrometry

Bernadette Bouchon[a,*], Michel Jaquinod[c,1], Klaus Klarskov[c,2], François Trottein[b], Michèle Klein[a], Alain Van Dorsselaer[c], Rainer Bischoff[a], Carolyn Roitsch[a]

[a]*Transgène S.A., 11 rue de Molsheim, 67082 Strasbourg Cédex, France*
[b]*Centre d'Immunologie et de Biologie Parasitaire, Institut Pasteur, 1 rue Pr. Calmette, BP245, 59019 Lille Cédex, France*
[c]*Laboratoire de Spectrométrie de Masse Bio-Organique, Faculté de Chimie, 11 rue Blaise Pascal, 67008 Strasbourg Cédex, France*

Abstract

The *Schistosoma mansoni* glutathione-S-transferase with an apparent molecular mass of 28 kDa, *Sm*p28, has a blocked N-terminus which has been elucidated with the aid of the cDNA sequence combined with mass spectrometry and amino acid composition analysis of the N-terminal tryptic peptide. The blocked N-terminal tryptic peptide (*m/z* 695.8) contained an equimolar ratio of E, G, H, A, I and K[3] upon amino acid composition analysis in agreement with its expected sequence AGEHIK, and showed a $\Delta m = +41.7$ Da compared to the predicted mass, which is consistent with the N-terminal alanine being acetylated ($\Delta m = +42.0$ Da). The mass of the complete molecule (23 744.5 ± 3.3 Da) determined by electrospray mass spectrometry showed a further mass increase of 14 Da with respect to *Sm*p28 containing an N-acetylated alanine. This result is consistent with one of the seven methionines being present as a methionine sulfoxide in ca. 90% of the *Sm*p28 molecules in this preparation. Tryptic mapping of *Sm*p28 showed five of the seven methionines to be partially oxidized by mass spectrometry. This is indicative of the ease with which this modification occurs. Two minor components were detected along with the intact molecule, corresponding to modified forms of the molecule, originating from reaction of the only cysteine residue either with itself forming a covalent dimer or with glutathione. On-line liquid chromatography–mass spectrometry has been compared with the off-line complete tryptic map of *Sm*p28 confirming 97% of the primary structure in less than 2 h.

1. Introduction

Schistosomiasis, a chronic debilitating disease, is one of the major parasitic diseases affecting approx. 200 million people in the world, and causing 500 000 deaths per year. Although effective chemotherapy exists, prevention of reinfection is still a problem [1].

Schistosoma mansoni p28 (*Sm*p28), which has been identified in adult worm extracts, is a protein with an apparent molecular mass of 28 000. Immunization of rats and mice with

* Corresponding author.
[1] Present address: Odense Universitet, Campusvej, 5230 Odense M, Denmark.
[2] Present address: Vakgroep Biochemie Fysiologie en Microbiologie, Ledeganckstraat 35, 9000 Gent, Belgium.
[3] The single letter code is used for amino acids.

0378-4347/94/$07.00
SSDI 0378-4347(94)00298-J

*Sm*p28 has been shown to provide significant levels of protection against schistosomiasis [2]. Using partial protein sequence data, the corresponding cDNA has been cloned [3] and the natural protein was subsequently isolated based on its glutathione-S-transferase activity (GST). Amino terminal sequence analysis showed *Sm*p28 to have a blocked N-terminus [4].

Here we report the determination of the complete primary structure of the protein including characterization of the N-terminal blocking group using electrospray mass spectrometry (ESMS), enzymatic digestion and chemical cleavage in combination with Edman degradation. In an extension of this study, coupling of reversed-phase high-performance liquid chromatography (HPLC) to electrospray mass spectrometry (LC–MS) using a post-column split according to Covey et al. [5] will be described rendering the technique applicable to rapid routine analyses performed in connection with the development of a recombinant version of this protein as a vaccine candidate against schistosomiasis.

2. Experimental

2.1. Chemicals

Chemicals were of analytical grade and solutions were prepared with MilliQ water (Millipore, Saint-Quentin-en-Yvelines, France). Reagents for amino acid composition and amino acid sequence analyses were from Applied Biosystems (Foster City, CA, USA). Sequencing grade reductively alkylated trypsin was from Promega (Madison, WI, USA); trifluoroacetic acid (TFA), N-ethyl morpholine, *n*-heptafluorobutyric acid (HFBA) and cyanogen bromide were from Pierce (Rockford, IL, USA); *n*-propanol was from Carlo Erba (Milano, Italy); formic acid was from Baker (Deventer, Netherlands) and Coomassie Brilliant Blue R250 was from Biorad (Ivry sur Seine, France).

2.2. Smp28 purification

*Sm*p28 was isolated from adult worms using GSH agarose affinity chromatography [4]. *Sm*p28 was further purified by reversed-phase HPLC using an Aquapore RP300 column (250 × 7 mm I.D., C_8, 7 μm particle size, 300 Å pore diameter, Applied Biosystems) on a Model HP1090 HPLC (Hewlett-Packard, Waldbronn, Germany). Solvent A was 0.1% aqueous TFA and solvent B *n*-propanol containing 0.1% TFA. A gradient was run from 0 to 20% of solvent B in 5 min, then from 20% to 40% solvent B in 50 min and from 40 to 80% solvent B in 5 min, at a flow-rate of 1.0 ml/min. Protein concentration was estimated by integration of the absorption at 205 nm.

2.3. Digestion of Smp28 with trypsin and purification of the derived peptides by HPLC

Tryptic digestion was performed in 250 m*M* N-ethyl morpholine, pH 8.5, containing 0.1 m*M*. calcium chloride and 4% (w/w) trypsin with a final protein concentration of 0.3 μg/μl. The mixture was allowed to react for 4 h at 37°C, and the reaction was stopped with 1% TFA (final concentration, 0.1%, v/v). HPLC separation of the peptides was performed using a 250 × 4.6 mm I.D. Vydac C_{18} column (5 μm particle size, 300 Å pore diameter, The Separation Group, Hesperia, CA, USA) at a flow-rate of 0.5 ml/min. Solvent A was 0.1% aqueous TFA, solvent B was acetonitrile containing 0.1% TFA and the gradient was run from 0 to 5% solvent B in 5 min, from 5 to 25% solvent B in 50 min, from 25 to 50% solvent B in 20 min, and from 50 to 80% solvent B in 10 min (separation system 1). The amount of protein was estimated by integration of the absorption at 205 nm.

In order to separate unresolved peaks, a different mobile phase was used on a 50 × 2.1 mm I.D. Vydac C_{18} column (5 μm particle size, 300 Å pore diameter). Solvent A was 10 m*M* HFBA in water, solvent B was 10 m*M* HFBA, 60% acetonitrile, 30% *n*-propanol in water.

Peptides were separated using a gradient from 5 to 40% solvent B in 60 min, and from 40 to 80% solvent B in 5 min at a flow-rate of 0.2 ml/min (separation system 2).

2.4. Polyacrylamide gel electrophoresis

Smp28 was analyzed by polyacrylamide gel electrophoresis in the presence of sodium dodecylsulfate (SDS-PAGE), according to Laemmli [6]. Staining of the protein bands in the gel was done with Coomassie Brilliant Blue R250.

2.5. N-Terminal sequence analysis

Amino acid sequences of protein and peptides were determined by automated Edman degradation performed on gas-phase (Model 470A), or pulsed liquid-phase (Model 477A) sequencers (Applied Biosystems) equipped with on-line phenylthiohydantoin amino acid (PTH-AA) analyzers (Model 120A), and microcomputers for data analysis.

2.6. Cyanogen bromide cleavage

Protein was dissolved in 70% aqueous formic acid. CNBr in 70% formic acid was added to a 300-fold molar excess per methionine residue. The reaction was performed at room temperature for 24 h in the dark under a nitrogen atmosphere, after which the reaction mixture was applied to a precycled biobrene-coated glass fiber filter and dried with nitrogen before N-terminal sequencing.

2.7. Amino acid analysis

Tryptic peptides purified with separation system 2 were submitted to acid hydrolysis (6 *M* HCl, 110°C for 18 h). Amino acid composition analysis was performed after derivatization to phenylthiocarbamyl amino acids (PTC-AAs) using a Model 420H derivatizer and on-line separation of the PTC-AAs on a Model 130A HPLC system (Applied Biosystems). Data acquisition, analyses and reports of the chromatographic data were performed on an ABI Model 920A data acquisition module. Norleucine was used as an internal standard for quantitation.

2.8. ESMS and LC–MS

ESMS was performed on a triple quadrupole mass spectrometer with a mass range of 4000 equipped with a pneumatically-assisted electrospray source (Fisons Bio-Q, VG-BioTech, Altrincham, UK). The samples were dissolved in a mixture of water–methanol (50:50, v/v) containing 1% acetic acid and 10-μl aliquots of such solutions containing 15–20 pmol/μl were introduced at a flow-rate of 4 μl/min and analyzed in the electrospray mode. Calibration was performed using multiply charged ions from a separate introduction of horse heart myoglobin (16 951.5 Da), giving a resolution of the m/z 998 myoglobin peak of 1.2 m/z units wide at 50% peak height. Scanning was performed from m/z 500 to m/z 1500 in 10 s in the multi-channel acquisition mode.

For LC–MS analysis, the mass spectrometer was operated according to specifications given by Klarskov et al. [7]. LC–MS was performed using a Model 140A HPLC pump (Applied Biosystems), and a 150 × 1.6 mm I.D. Vydac C_{18} microbore column (5 μm particle size, 300 Å pore diameter; MZ-Analysentechnik, Mainz, Germany) equipped with a 1/15 post-column split (Valco). Solvent A was 0.1% aqueous TFA and solvent B 0.09% TFA in 80% acetonitrile. The gradient was run for 5 min at 0% solvent B, then to 31% solvent B in 50 min, then to 62% solvent B in 20 min and then to 100% solvent B in 5 min, at a flow-rate of 100 μl/min. Eluted peptides were monitored at 214 nm (Waters Model 486, Millipore Corporation, Milford, MA, USA) and scanning was performed from m/z 200 to m/z 1600 in 5 s in the centroid mode. Molecular masses are given as average values based on the atomic masses of the elements

(C = 12.011, H = 1.00794, N = 14.0067, O = 15.9994, S = 32.06).

3. Results

3.1. Smp28 purification

*Sm*p28 was purified on glutathione (GSH: γ-glutamyl-cysteinyl-glycine) agarose based on its GST activity [4]. *Sm*p26, another *Schistosoma mansoni* protein with GST activity [8] coeluted with *Sm*p28. Direct sequencing of the obtained *Sm*GSTs allowed the identification of *Sm*p26 [9], while *Sm*p28 appeared to be N-terminally blocked as previously reported [4]. The sequencing yield indicated that *Sm*p26 represented roughly 10% of the total GSTs (data not shown). *Sm*p28 and *Sm*p26 were efficiently separated by HPLC on a C_8 column, using a gradient of *n*-propanol. Analysis by SDS-PAGE of the separated GSTs (Fig. 1, lane 3) indicated that the major fraction (lane 4) contained *Sm*p28 (purity >95%), while a minor fraction (lane 5) was enriched in *Sm*p26 as identified by N-terminal sequencing (data not shown). The fraction corresponding to lane 4 was subsequently used to further characterize *Sm*p28.

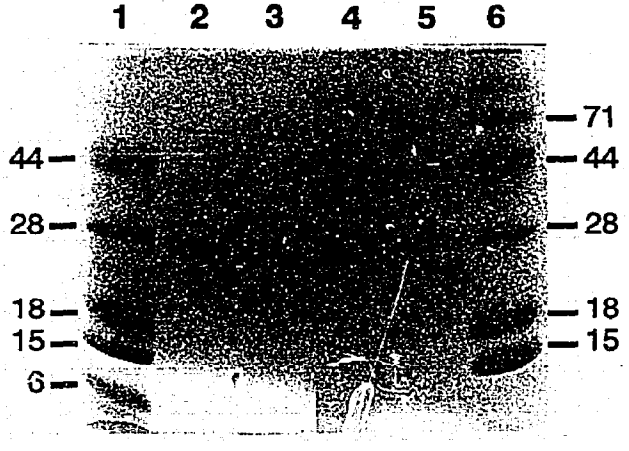

Fig. 1. Coomassie Blue-stained SDS-PAGE gel of *Sm*p28 at different stages of purification. Lanes 1 and 6: prestained molecular mass markers (BRL); lane 2: recombinant *Sm*p28 [26]; lane 3: protein fraction after GSH-agarose chromatography; lane 4: *Sm*p28 after RP-HPLC; lane 5: *Sm*p26 enriched fraction after HPLC. The center lanes of the original gel are omitted for simplicity.

3.2. ESMS of the complete molecule

Mass determination of the complete molecule by ESMS was repeated several times, and a typical spectrum is given in Fig. 2. The measured mass (23 744.5 ± 3.3 Da) was greater (Δm = + 56.2 Da) than the calculated mass without the methionine, but less (Δm = − 75.0 Da) than the calculated mass including an N-terminal methionine. Elimination of the N-terminal (methionine Δm = 131.2 Da; formylated methionine Δm = 159.2 Da) or C-terminal (phenylalanine Δm = 147.2 Da) amino acid could not individually account for the differences in mass, and in addition, the existence of an N-terminal blocking group had to be taken into account. However, at this point, the possibility of *Sm*p28 having an N-formyl methionine and perhaps missing its C-terminal phenylalanine (calculated mass: 23 700.3 Da) could not be entirely excluded as the measured mass was 44 Da in excess of this value, and partial oxidation of the eight methionines (+ 16 Da each) in the molecule could account for the excess mass.

A minor component representing less than 5% of the total signal in the spectrum had a mass of 47 497 ± 19 Da which is in agreement with dimer formation of the major component (expected mass = 47 487 Da). Dimer formation might be mediated through the single cysteine residue in *Sm*p28, and has been observed in SDS-PAGE under non-reducing conditions. Another component representing less than 10% of the observed signal had a mass of 24 046.4 ± 7.3 Da. This molecular mass was significantly different from the expected mass of *Sm*p26 (theoretical mass = 25 314.2 Da), and corresponded to the major component with the addition of a covalently bound GSH moiety (expected mass = 24 049.8 Da) as confirmed by tryptic peptide mapping (Table 1). Such a modification has also been observed for recombinant *Sm*p28 [7] and may be due to isolation on GSH agarose.

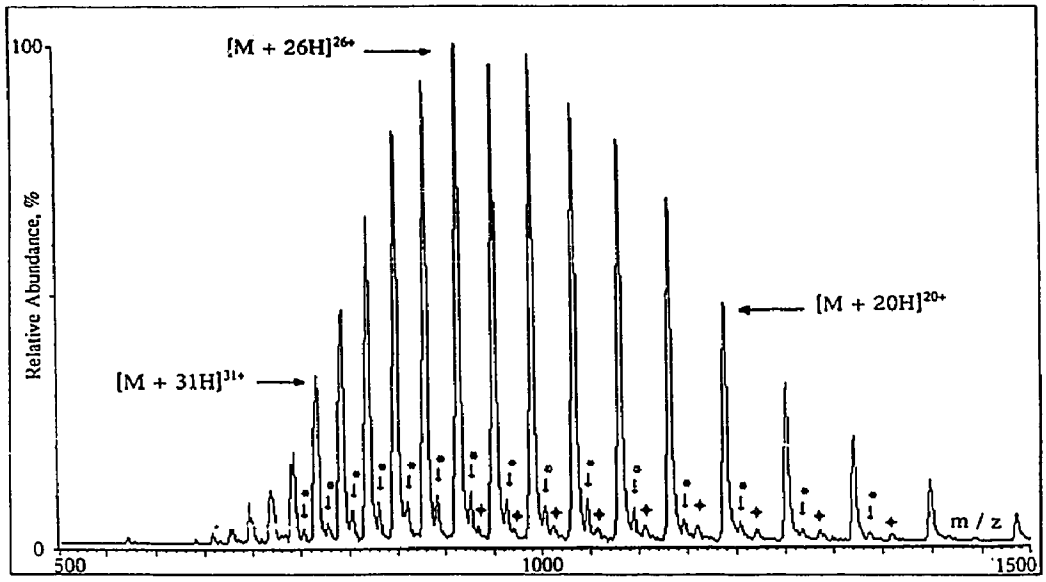

Fig. 2. Electrospray mass spectrum of *Sm*p28. The major series corresponds to a measured mass of 23 744.5 ± 3.3 Da (calculated mass of *Sm*p28 based on its cDNA sequence: 23 688.3 Da). The minor series marked ◆corresponds to a measured mass of 47 497 ± 19 Da, in accordance with the mass for a dimer of the major component. The minor series marked * has a mass of 24 046.4 ± 7.3 Da, and corresponds to the addition of glutathione (m = 305.3 Da) to *Sm*p28 (Δm = 301.9 Da) at C139.

3.3. CNBr cleavage

As the protein was N-terminally blocked, and cDNA information predicted the presence of eight methionine residues, including the initiator methionine (Tables 2 and 3), cleavage with CNBr was used to probe whether *Sm*p28 contains an N-terminally blocked methionine residue. A 300-fold molar excess of CNBr over methionine residues in *Sm*p28 resulted in efficient cleavage and six amino acid residues could be determined simultaneously in the first cycle of Edman degradation. Alignment of the cDNA sequence with the six N-termini that had been generated allowed sequences L68–M77, A78–M83 and K114–M137 to be totally assigned. The three larger fragments (T22–K44.., G85–V104.., I138–L160..) could not be completely sequenced but were positively identified (Table 2). No sequence corresponding to the expected N-terminus without the methionine could be determined, indicating that the methionine in position −1 (Table 3) had been removed during protein biosynthesis in the parasite.

3.4. Analysis of tryptic digest

For a more complete characterization of the protein, enzymatic digestion with trypsin was performed, and peptides were separated by HPLC (Fig. 3). Twenty-eight of the collected fractions were submitted to mass determination and N-terminal sequencing (Table 1) confirming the protein sequence. Peptides K80–K96 (fractions 31–32), W66–R75 (fraction 47) and V131–K143 (fractions 50 and 53) containing methionine residues were detected in both normal and oxidized forms (measured mass differences of approximately + 16Da). Peptide K80–K96 was primarily present in its non-oxidized form, and the oxidized form of the related H81–K96 was not detected at all. While peptides W66–R75 and K80–K96 appeared in their oxidized and non-oxidized forms in single HPLC fractions, peptide V131–K143 was found in two separate peaks with the oxidized form eluting significantly earlier. This strongly suggested that oxidation of V131–K143 occurred prior to HPLC analysis, while the other two pep-

Table 1
Analysis of HPLC fractions after tryptic digestion

Sequence obtained by Edman degradation	Position	Off-line fraction No.	Measured mass off-line[a]	Expected mass	Measured mass on-line[a]	On-line peak No.
LAK	198–200	9	330.1	330.4	330	1
ITK	122–124		360.2	360.2	361	
YMAK	76–79	13	511.3	511.6	512	2
VTDDHGHVK	57–65		1006.5 ± 0.2	1007.1	1007.5	
VTDDHGHVK	57–65	14	1006.7 ± 0.2	1007.1	1007.5	3
AESIR	16–20		574.4	574.6	574.5	4
*AGEHIK	1–6	17	695.4	695.8	696	5
LAVGDK	149–154	18	601.3	601.7	602	6
(TLMKPQEEK)	111–119		n.d.	1119.3[b]	1120	
EILNGK	125–130	20	673.5	672.8	673	7
LPAVK	52–56		527.4	526.7	n.d.	
LPAVK	52–56	21	526.7	526.7	526.5	8
TLMKPQEEK	111–119		n.d.	1103.3	n.d.	
TLMKPQEEK	111–119	22	1103.7 ± 0.4	1103.3	1103.5	9
TLMKPQEEKEK	111–121		1360.9 ± 0.5	1360.6	1362	
YPEIHEK	181–186		n.d.	785.9	n.d.	
YPEIHEK	181–186	23	785.9 ± 0.4	785.9	786	10
(GFLTGK)	175–180		622.8 ± 0.4	621.7	n.d.	
IKPTIPGGR	43–51	25	938.2 ± 0.6	938.1	938.5	11
HRENLLASSPR	187–197	28	1279.7 ± 0.4	1279.4	1280.5	12
ENLLASSPR	189–197		986.2	986.1	986	
(ENLLASSPR)	189–197	29	985.9 ± 0.3	986.1	n.d.	
LIGQAEDVEHEYHK	97–110	30	1668.2 ± 0.5	1667.8	1668	14
GFLTGK	175–180	31	621.9 ± 0.3	621.7	622	13
(KHHMMGETDEEYYSVEK)	80–96		2113.6 ± 0.6	2113.3	n.d.	
	80–96		2129.9 ± 0.6	2129.3[b]	2129	
KHHMMGETDEEYYSVEK	80–96	32	2113.6 ± 0.2	2113.3	2114.5	15
	80–96		2130.3 ± 1.1	2129.3[b]	n.d.	
HHMMGETDEEYYSVEK	81–96		1985.5 ± 0.9	1985.1	1986	
VIYFDGR	7–13	35	868.9 ± 0.4	869.0	869.5	16
YLSNRPATPF	201–210	37	1164.7 ± 1.0	1165.3	1165.5	17
MTLVAAGVDYEDER	21–34	39	1569.6 ± 1.4	1568.7	1569.5	18
GFLTGKYPEIHK	175–186		1390.1 ± 0.5	1389.6	n.d.	
ISFQDWPK	35–42	42	1020.1	1020.2	1020	19

Table 1. (*Continued*)

Sequence obtained by Edman degradation	Position	Off-line fraction No.	Measured mass off-line[a]	Expected mass	Measured mass on-line[a]	On-line peak No.
(VPVLLNMICESLK)	131–143	45	1765.7 ± 4.2	1764.1[c]	1766	22
WMLESLAIAR	66–75	47	1205.1 ± 0.4	1205.2[b]	1206	20
	66–75		1189.2 ± 0.3	1189.4	1190	21
(VPVLLNMICESLK)	131–143	50	1474.1 ± 1.2	1475.4[b]	n.d.	
VPVLLNMICESLK	131–143	53	1459	1458.8	1460	23
LAVGDKVTLADLVLI...PEIHK	149–186	54	4119.1 ± 0.4	4118.8	4120	24
LAVGDKVTLADLVLI...TDLDK	149–174	55	2746.9 ± 0.2	2747.2	2748	
LAVGDKVTLADLVLI...TDLDK	149–174	56	n.d.	2747.2	n.d.	
VTLADLVLIAVIDHVTDLDK	155–174		n.d.	2163.5	n.d.	
VTLADLVLIAVIDHVTDLDK	155–174	57	2163.6 ± 0.1	2163.5	2165	25
VTLADLVLIAVIDHVTDLDK	155–174	58	2163.7 ± 0.4	2163.5	n.d.	
(VPVLLNMICESLKGSTGK)	131–148		n.d.	1905.3[b]	1904	

[a] As obtained from Lab-base version 1.0 from VG Biotech.
[b] Calculated mass is based on the peptide containing one methionine sulfoxide.
[c] Calculated mass is based on the peptide containing one GSH.
* = N-acetylation; n.d. = not determined.
See Fig. 3 for off-line analysis and Fig. 5 for on-line analysis. Sequences in parentheses were identified by mass determination only. Off-line fraction 17 was characterized by amino acid analysis. Sequences with "..." in the middle are identical to the sequence given in Table 3.

tides were partially oxidized after HPLC separation.

Mass spectrometric analysis of the complete molecule indicated the presence of a small amount of *Sm*p28 containing a covalently-bound GSH moiety (see Fig. 2). Analysis of the separated tryptic fragments confirmed this assumption, since the peptide in fraction 45 had a mass which corresponded well to the expected mass of V131–K143 containing one GSH (Table 1). A similar modification has recently been found in recombinant *Sm*p28 which was also purified on a

Table 2
Summary of the data obtained by Edman degradation of *Sm*p28 after CNBr cleavage

Sequences	Position
----.----|----.----|----	
TLVAAGVDYEDERISFQDWPKIK..	22–44..
LESLAIARYM	68–77
AKKHHM	78–83
GETDEEYYSVEKLIGQAEDV....	85–104..
KPQEEKEKITKEILNGKVPVLLNM	114–137
ICESLKGSTGKLAVGDKVTLADL..	138–160..

Table 3
Protein sequence of *Smp*28 as derived from the cDNA sequence [3]

```
 1         |         |         |         |         |        5|0
MAGEHIKVIYFDGRGRAESIRMTLVAAGVDYEDERISFQDWPKIKPTIPGGRLPAV

    |         |         |         |        10|0        |
KVTDDHGHVKWMLESLAIARYMAKKHHMMGETDEEYYSVEKLIGQAEDVEHEYHK

         |         |         |        15|0        |
TLMKPQEEKEKITKEILNGKVPVLLNMICESLKGSTGKLAVGDKVTLADLVLIAV

    |         |         |        20|0        |
IDHVTDLDKGFLTGKYPEIHKHRENLLASSPRLAKYLSNRPATPF
```

Methionine residues are underlined;] indicates cleavage sites for trypsin.

GSH affinity matrix [7]. Due to partial tryptic digestion, several overlapping peptides were obtained, thus permitting primary structure confirmation (Table 1). Two peptides (G14–R15 and G144–S–T–G–K148) were not observed. The mass of the dipeptide (231.3 Da) was too low to be detected by ESMS since scans were started at *m/z* 300. The strong hydrophilicity of the pentapeptide was most likely responsible for it not being detected, as it may have eluted in the flow-through fraction of the HPLC analysis where the presence of low-molecular mass contaminants may have led to a decreased sensitivity for ESMS.

3.5. Identification of the N-terminal blocking group

Fraction 17 which was N-terminally blocked had a measured mass of 695.4 Da which did not correspond to the mass of any of the expected tryptic peptide fragments. It was thus submitted to amino acid composition analysis. Since the material in fraction 17 appeared to be slightly

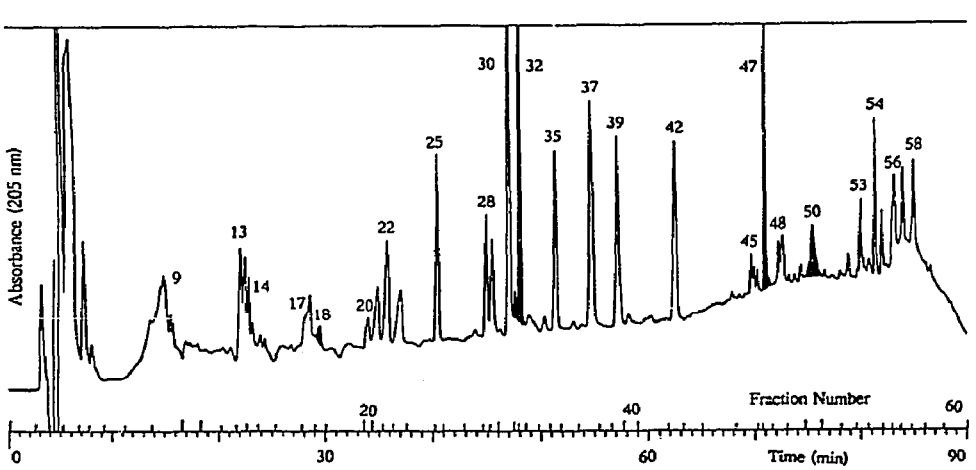

Fig 3. HPLC of a tryptic digest of *Smp*28 monitored at 205 nm. (see Experimental for details). Fractions are identified in Table 1. Shaded peaks correspond to fractions containing modified peptide fragments.

heterogenous (Fig. 3), it was subjected to an additional purification step by HPLC using a mobile phase containing HFBA (separation system 2) to eliminate some minor contaminants (data not shown). Quantitative amino acid composition analysis established that the purified fraction contained E, G, H, A, I and K in equimolar ratios (Fig. 4). This composition corresponded to the composition of the N-terminal peptide without methionine. The calculated mass for the potential N-terminal peptide based on this analysis is 653.7 Da (AGEHIK) while mass measurement of fraction 17 gave a mass of 695.4 Da ($\Delta m = +41.7$ Da). Of the most frequently occurring low molecular mass N-terminal blocking groups in eukaryotic cells, an acetyl group ($\Delta m = +42.0$ Da) corresponded best to the observed mass addition [10]. Taken together the results of amino acid composition analysis, mass spectrometry and the fact that the protein is N-terminally blocked are all in agreement with *Sm*p28 containing an N-acetylalanine residue.

3.6. LC–MS

In order to perform routine verification of the amino acid sequence of *Sm*p28 and the corresponding recombinant protein, peptides derived from a tryptic digestion of *Sm*p28 were submitted to analysis by LC–MS (Fig. 5). Peptides were detected by absorbance at 214 nm and total ion current monitoring. Combined scans representing approximately 3–10 scans across the top of each peak were analyzed, permitting mass determination of thirty three peptides (Table 1). Comparison of the measured mass values with those expected for tryptic fragments of *Sm*p28 allowed their assignment which corroborated the results obtained by off-line analysis (Table 1), and resulted in confirmation of 97% of the *Sm*p28 sequence. Part of the peptide containing the only cysteine residue of the protein was identified as containing glutathione (V131–K143,

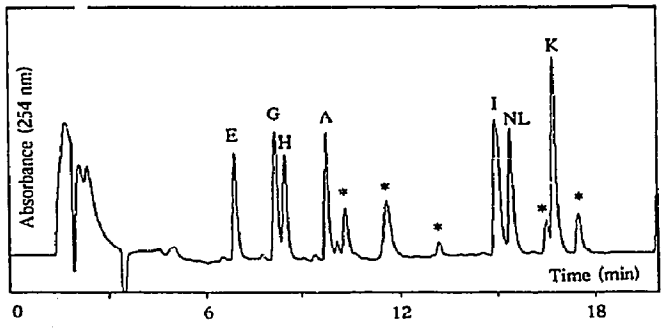

Fig. 4. Amino acid composition analysis of the N-terminal peptide after acid hydrolysis. Norleucine (NL) was used as an internal standard (200 pmol). The amino acids E, G, H, A, I and K were present in the peptide in equimolar ratios. Peaks marked * correspond to by-products of the derivatization with phenylisothiocyanate, and appeared also in a blank chromatogram.

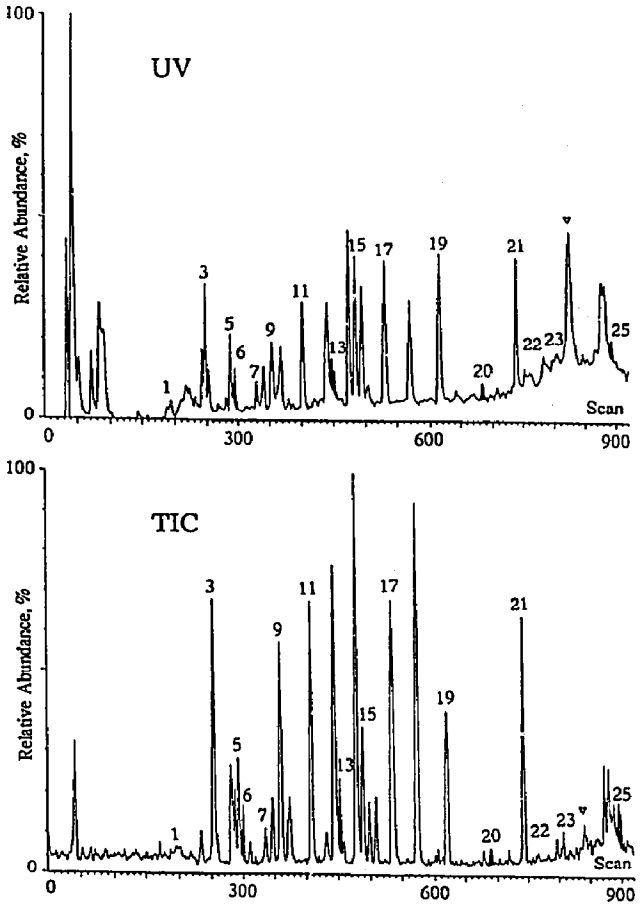

Fig. 5. On-line LC–MS analysis of *Sm*p28 tryptic peptides. UV monitoring at 214 nm (upper trace) and total ion current (TIC, bottom trace) as measured by electrospray mass spectrometry are shown. Peaks are identified in Table 1. Peak marked ∇corresponds to a solvent impurity. Shaded peaks correspond to fractions containing modified peptides. Time scale, 1 scan = 5 s.

Fig. 5, peak 22) in agreement with the earlier results obtained off-line (Fig. 3, fraction 45). Another fraction, peptide V131–K148 resulting from partial cleavage, contained a methionine sulfoxide (Fig. 5, peak 25) in agreement with results from off-line analysis (Fig. 3, fraction 50). In addition, peptide V131–K148 was of significance in that the off-line missing G144–K148 segment was identified here due to incomplete tryptic digestion. Peptides T111–K119, K80–K96 and W66–R75, containing one methionine residue each, were split into two different peaks (Fig. 5; peaks 6 and 9, 13 and 15, 20 and 21 and Table 1) showing that the protein was partially oxidized prior to LC–MS analysis. This oxidation pattern of *Sm*p28 did not show in all preparations indicating that it occurred during sample preparation and might be related to denaturation of the protein as a result of HPLC. Due to slight differences in the gradients used, several minor peaks containing modified peptides had different on-line retention times compared to off-line analysis (for oxidized peptides, on-line peak 13 eluted earlier than the corresponding off-line fraction 31, and for glutathionylated peptides, on-line peak 22 eluted earlier than the corresponding off-line fraction 45).

4. Discussion

*Sm*p28 (210 amino acids), a natural GST of *Schistosoma mansoni* can be separated from another GST (*Sm*p26, 217 amino acids) by HPLC following GST isolation on a GSH-agarose column, using an *n*-propanol gradient, which gave better resolution than the previously used solvent acetonitrile [11]. The *Sm*p28 thus purified appears homogeneous on SDS-PAGE with an apparent molecular mass of 28 kDa from which its name was derived.

*Sm*p28 analysis by ESMS shows the major protein component to have a mass of 23 744 Da with minor components at 24 046 Da (< 10%) and at 47 497 Da (< 5%). The single cysteine residue (C139) gives *Sm*p28 the ability to covalently dimerize or to form a disulfide bond with other thiol-containing molecules. The mass of the 24.05 kDa component which does not correlate with the theoretical mass of *Sm*p26 (25 314.2 Da), was in agreement with the mass of a glutathionylated form of *Sm*p28 having an N-terminal N-acetylated alanine (expected mass: 24 035.7 Da, measured mass: 24 046.4 Da; see below). This is consistent with the observation that part of peptide V131–K143 was linked to GSH (Table 1, fraction 45 and peak 22) as identified by mass measurements. Glutathionylation may have occurred during purification of *Sm*p28 on GSH-agarose with GSH as the eluent. A similar modification has also been identified in the recombinant protein [7]. The 47.5 kDa component correlates well with a covalent *Sm*p28 dimer (theoretical mass = 47 458.8 Da) and is consistent with the fact that a low level of dimer has also been observed upon SDS-PAGE under non-reducing conditions.

Previous analyses of *Sm*p28 [4] permitted identification of several peptides which corresponded to the published cDNA sequence [3]. However, no detailed analysis of the natural protein has been performed up to now, and the sequence data on the protein covered less than half of the molecule, excluding the N-terminal amino acid which was shown to be blocked. Subsequent RNA analyses have shown sequence variations at positions 60 and 135 (D60/L135, N60/F135, N60/L135), with the major form encoding a protein having D60 and L135 (R. Pierce, personal communication) in agreement with the protein sequence determined in this manuscript. As the cDNA sequence alone did not provide information about post-translational modifications of the protein, it was of interest to perform a complete analysis of the natural protein purified from schistosomules by amino acid sequencing and mass determination after enzymatic and chemical cleavages. Analysis of *Sm*p28 after CNBr cleavage by amino acid sequencing indicated that the methionine residue corresponding to the start codon of the cDNA was not present in the protein (Table 2). Tryptic digestion resulted in ca. 30 major peaks upon HPLC, and analysis by ESMS and N-terminal sequencing allowed verification of 95% of the

expected sequence, including the C-terminal F210. A single peptide fragment however, with a measured molecular mass of 695.4 Da appeared to be inaccessible to Edman degradation indicating this to be the N-terminal fragment (expected mass: 695.8 Da). Mass measurement and amino acid composition analysis of this peptide were used to characterize the N-terminus of *Sm*p28 to be N-acetyl alanine. Although alanine is an amino acid commonly acetylated in eukaryotes [10], *Sm*p26, a GST which copurifies with *Sm*p28 on GSH-agarose and starts also with an alanine was shown not to be N-terminally blocked. A number of GSTs from different organisms have been cloned, and the full-length protein sequences deduced from the corresponding cDNAs [12]. However, only few data concerning direct analyses of the natural proteins have been presented. Two GSTs which were completely sequenced [13,14] appeared not to be blocked, while partial amino acid sequencing indicated that other GSTs may be blocked [15,16]. Although it can protect against degradation and affect the in vivo lifetime of proteins [10], the physiological role of this post-translational modification has not been clearly established [17]. In addition, there is no reliable procedure for removing an N-terminal acetyl group, in spite of several reported enzymatic [18,19] and chemical [20] approaches. Mass determination is thus one of the easiest and fastest ways to identify N-terminal blocking groups [21,22].

Mass measurement of *Sm*p28 (23 744.5 Da) and comparison with the expected value (23 730.4 Da) for the molecule with an N-acetyl alanine did not allow to assign the correct mass to the N-terminal blocking group, since a mass difference of +14.1 Da prevailed, indicating further modifications. Mass measurement on the same sample kept frozen for 6 days at −20°C in analysis buffer resulted in an additional mass shift of +10 Da (data not shown), indicating further modification of the protein which had to be considered as being responsible for the observed discrepancy. Partial oxidation of certain methionine residues might explain this shift, as the molecule appeared to be quite sensitive to oxidation (see Table 1). Oxidation of methionines is a common phenomenon giving rise to labile methionine sulfoxides which are not detected by automated N-terminal sequencing under standard conditions. However, this modification reduces the retention times in HPLC, thus permitting separation of oxidized and non-oxidized forms [23] (Fig. 3). Since the theoretical width of the isotope envelope for a 24 kDa protein (*Sm*p28) is >15 Da [24], two proteins with a difference in mass of 16 Da cannot be distinguished with certainty by average molecular mass measurements and partial oxidation would lead to an increase of the measured molecular mass due to a shift of the isotope envelope. If oxidation was the source for the 14.1 Da mass increase observed for *Sm*p28, ca. 90% of *Sm*p28 must contain one oxidation. Of the seven methionines present, M67, M83, M84, M113 and M137 have been identified in an oxidized state. From a three dimensional model of recombinant *Sm*p28 based on crystallographic data of another GST, M21 and M77 are hidden within the molecule, while the other methionine residues are more exposed (J.P. Mornon and J. Chomilier, personal communication) which supports the observation that M21 and M77 have never been found in their oxidized state in this study. Nevertheless, in previous experiments on recombinant *Sm*p28, an oxidation on M21 had been found, in addition to oxidations on the other five methionines [7]. Furthermore, it has to be kept in mind that *Sm*p28 was at least partially denatured and dissociated into its subunits following HPLC.

To accelerate confirmation of the primary structure of *Sm*p28, LC–MS was performed on 1 nmole of protein, using a narrow-bore column and a post-column split. One fifteenth of the sample was introduced into the mass spectrometer for mass measurement while the rest was available for further analyses. It was thus possible to confirm 97% of the complete primary structure of *Sm*p28 in a single analysis within a few hours. Partial proteolysis resulted in overlapping peptide fragments, thus permitting identification of very hydrophilic peptides which would otherwise be found in the flow-through fraction of the HPLC. LC–MS analysis also facilitated N-

and C-terminal verification, an important criterium for structural integrity. Initial analyses have shown that the recombinant protein derived from the cDNA coding for *Sm*p28 exhibits heterogeneity upon isoelectric focusing [25], and LC–MS will facilitate the recombinant p28 protein analysis in view of its development as a vaccine candidate against schistosomiasis with the use of capillary columns increasing the sensitivity of the technique [7].

Acknowledgements

We are indebted to the late Dr. Jean-Pierre Lecocq for his encouragement in this and all aspects of the Pasteur Mérieux anti-schistosomiasis program. We thank Dr. Raymond J. Pierce and Dr. André Capron (Institut Pasteur, Lille, France) for the kind gift of the natural SmGSTs, Dr. Jean-Pierre Mornon and Dr. J. Chomilier (Université Paris VI, France) for communicating results to us prior to publication. Our gratitude goes to Dr. Jean Sabatié for his enthusiastic and efficient coordination of the Transgène recombinant *Sm*p28 program. The technical assistance of Luciana Delatre for amino acid analysis is greatly appreciated.

References

[1] J. Cherfas, *Science*, 251 (1991) 630.
[2] J.M. Balloul, J.M. Grzych, R.J. Pierce and A. Capron, *J. Immunol.*, 138 (1987) 3448.
[3] J.M. Balloul, P. Sondermeyer, D. Dreyer, M. Capron, J.M. Grzych, R.J. Pierce, D. Carvallo, J.P. Lecocq and A. Capron, *Nature*, 326 (1987) 149.
[4] J.B. Tayor, A. Vidal, G. Torpier, D.J. Meyer, C. Roitsch, J.M. Balloul, C. Southan, P. Sondermeyer, S. Pemble, J.P. Lecocq, A. Capron and B. Ketterer, *EMBO J.*, 7 (1988) 465.
[5] T.R. Covey, E.C. Huang and J.D. Henion, *Anal. Chem.*, 63 (1991) 1193.
[6] U.K. Laemmli, *Nature*, 227 (1970) 680.
[7] K. Klarskov, D. Roecklin, B. Bouchon, J. Sabatié, A. Van Dorsselaer and R. Bischoff, *Anal. Biochem.*, 216 (1994) 127.
[8] F. Trottein, M.P. Kieny, C. Verwaerde, G. Torpier, R.J. Pierce, J.M. Balloul, D. Schmitt, J.P. Lecocq and A. Capron, *Mol. Biochem. Parasitol.*, 41 (1990) 35.
[9] M.D. Wright, R.A. Harrison, A.M. Melder, G.R. Newport and G.F. Mitchell, *Mol. Biochem. Parasitol.*, 49 (1991) 177.
[10] R.L. Kendall, R. Yamada and R.A. Bradshaw, *Methods Enzymol.*, 185 (1990) 398.
[11] A.K. Ostlund Farrants, D.J. Meyer, B. Coles, C. Southan, A. Ken, P.J. Johnson and B. Ketterer, *Biochem. J.*, 245 (1987) 423.
[12] B. Mannervik and U.H. Danielson, *CRC Crit. Rev. Biochem.*, 23 (1988) 283.
[13] P. Ålin, B. Mannervik and H. Jörnvall, *Eur. J. Biochem.*, 156 (1986) 343.
[14] R. Morgenstern, J. DePierre and H. Jörnvall, *J. Biol. Chem.*, 260 (1985) 13976.
[15] J.D. Hayes, D.H. Judah, G.E. Neal and T. Nguyen, *Biochem. J.*, 285 (1992) 173.
[16] S.S. Singhal, M. Saxena, H. Ahmad, S. Awashti, A.K. Haque and Y.C. Awashti, *Arch. Biochem. Biophys.*, 299 (1992) 232.
[17] M.A. Romanos, C.A. Scorer and J.J. Clare, *Yeast*, 8 (1992) 423.
[18] R.G. Krishna, C.C.Q. Chin and F. Wold, *Anal. Biochem.*, 199 (1991) 45.
[19] H. Hirano, S. Komatsu, H. Takakura, F. Sakiyama and S. Tsunasawa, *J. Biochem.*, 111 (1992) 754.
[20] J.D. Hulmes and Y.C.E. Pan, *Anal. Biochem.*, 197 (1991) 368.
[21] F.M. Greer, H.R. Morris, J. Forstrom and D. Lyons, *Biomed. Environ. Mass Spectrom.*, 16 (1988) 191.
[22] D. Promé, Y. Blouquit, C. Ponthus, J.C. Promé and J. Ross, *J. Biol. Chem.*, 266 (1991) 13050.
[23] G. Teshima and E. Canova-Davis, *J. Chromatogr.*, 625 (1992) 207.
[24] A. Van Dorsselaer, F. Bitsch, B. Green, S. Jarvis, P. Lepage, R. Bischoff, H.V.J. Kolbe and C. Roitsch, *Biomed. Environ. Mass Spectrom.*, 19 (1990) 692.
[25] R. Bischoff, D. Roecklin and C. Roitsch, *Electrophoresis*, 13 (1992) 214.
[26] G. Loison, A. Vidal, A. Findelli, C. Roitsch, J.M. Balloul and Y. Lemoine, *Yeast*, 5 (1989) 497.