JOURNAL OF MASS SPECTROMETRY
J. Mass Spectrom. 2007; **42**: 781–792
Published online 18 May 2007 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/jms.1214

JMS

Improved protein identification efficiency by mass spectrometry using N-terminal chemical derivatization of peptides from *Angiostrongylus costaricensis*, a nematode with unknown genome

Ileana R. León,[1] Ana G. C. Neves-Ferreira,[1] Richard H. Valente,[1] Ester M. Mota,[2] Henrique L. Lenzi[2] and Jonas Perales[1]*

[1] Department of Physiology and Pharmacodynamics, Oswaldo Cruz Institute, FIOCRUZ, Rio de Janeiro, Brazil
[2] Department of Pathology, Oswaldo Cruz Institute, FIOCRUZ, Rio de Janeiro, Brazil

Received 20 October 2006; Accepted 13 March 2007

Matrix-assisted laser desorption ionization (MALDI), Peptide Mass Fingerprinting (PMF) and MALDI-MS/MS ion search (using MASCOT) have become the preferred methods for high-throughput identification of proteins. Unfortunately, PMF can be ambiguous, mainly when the genome of the organism under investigation is unknown and the quality of spectra generated is poor and does not allow confident identification. The post-source decay (PSD) fragmentation of singly charged tryptic peptide ions generated by MALDI-TOF/TOF typically results in low fragmentation efficiency and/or complex spectra, including backbone fragmentation ions (series b and y), internal fragmentation etc. Interpreting these data either manually and/or using *de novo* sequencing software can frequently be a challenge. To overcome this limitation when studying the proteome of adult *Angiostrongylus costaricensis*, a nematode with unknown genome, we have used chemical *N*-terminal derivatization of the tryptic peptides with 4-sulfophenyl isothiocyanate (SPITC) prior to MALDI-TOF/TOF MS. This methodology has recently been reported to enhance the quality of MALDI-TOF/TOF-PSD data, allowing the obtainment of complete sequence of most of the peptides and thus facilitating *de novo* peptide sequencing. Our approach, consisting of SPITC derivatization along with manual spectra interpretation and Blast analysis, was able to positively identify 76% of analyzed samples, whereas MASCOT analysis of derivatized samples, MASCOT analysis of nonderivatized samples and PMF of nonderivatized samples yielded only 35, 41 and 12% positive identifications, respectively. Moreover, *de novo* sequencing of SPITC modified peptides resulted in protein sequences not available in NCBInr database paving the way to the discovery of new protein molecules. Copyright © 2007 John Wiley & Sons, Ltd.

KEYWORDS: SPITC; MALDI TOF/TOF; PSD; *de novo* sequencing; *Angiostrongylus costaricensis*

INTRODUCTION

Peptide mass fingerprinting (PMF) using matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS) has become the preferred method for high-throughput identification of proteins.[1] The enzymatic digestion of a protein with specific enzymes (e.g. trypsin) produces a set of different peptides which are characteristic of each protein.[2,3] With the increasing number of completely sequenced genomes, the chance of correctly identifying the protein by PMF is likely high. However, the eventual presence of more than one protein in the same gel spot, as well as the occurrence of post-translational modifications, incomplete digestion and/or unspecific cleavage, can impair reliable identification. On the other hand, in the case of uncharacterized genomes, protein identification by PMF is virtually impossible.[4]

To identify a protein from a species with unknown genome, additional information regarding its primary structure is necessary. MALDI post-source decay (PSD) analysis, introduced by Kaufmann and coworkers,[5] has been used to determine amino acid sequences of peptides by recording in a reflectron TOF mass spectrometer those fragment ions formed after the ions have left the ionization source, but prior to their entrance into the reflectron. The fragmentation pattern in PSD reflectron TOF can be very complex due to the possible generation of several fragment ion types, such as *N*-terminal (a_n, b_n and c_n)[5,6] and C-terminal (x_n, y''_n and z_n)[7] series, as well as immonium[8] and other internal cleavage ions (for nomenclature, see Ref. 9). To complicate it even more, very often peptides show differences in their ability to fragment. As a consequence, interpreting these data either manually and/or using *de novo* sequencing software can

*Correspondence to: Jonas Perales, Av. Brasil, 4365, Manguinhos, 21040-900, Rio de Janeiro, RJ, Brazil. E-mail: jperales@ioc.fiocruz.br

frequently be a challenge. Precise sequencing of peptides is rarely possible, and in certain cases the partial amino acid sequence eventually retrieved from a PSD spectrum is too short or not specific enough to allow unambiguous protein identification because the number of proposed sequences resulting from a database search is too large.[10]

To overcome these limitations, Keough *et al*. introduced a procedure that dramatically improves the interpretation of PSD spectra.[11,12] This protocol includes an *N*-terminal derivatization with 4-sulphophenyl isothiocyanate (SPITC),[13] leading to the formation of a strongly negative *N*-terminal group by a simple sulfonation reaction. The presence of the newly formed acidic group greatly enhances the fragmentation efficiency of tryptic peptides. Moreover, the negative charge at the *N*-terminus neutralizes the positive charge of the captured proton in the b_n fragment ions, rendering them neutral and thus undetectable by mass spectrometry. The resulting spectra are mainly composed of a well-characterized series of y″ ions, making the interpretation easier and more efficient to analyze using different automatic software as well as manual *de novo* sequencing.

The ability to reliably interpret these spectra *de novo* facilitates both error-tolerant database searching[14] and sequence homology searching using commercially available search algorithms. This latter capability was used for definitive protein identification in a *Pseudomonas* species having an uncharacterized genome. Tryptic peptides containing up to three amino acid substitutions relative to known peptides from the NCBI database were unambiguously identified using this sequencing method.[15]

In the present study, we have compared the efficiency of *N*-terminal chemical derivatization combined with either MASCOT software or manual sequencing followed by BLAST algorithm for database searching to identify proteins with confidence. Control samples without derivatization were also analyzed using MASCOT. Proteins were extracted from our model organism, the nematode *Angiostrongylus costaricensis*. This parasite presents unknown genome, infects rodents and shows widespread distribution in the Central and South Americas, including Brazil.[16,17] A complete rediscription of the life history of *A. costaricensis*, based on the experimentally infected *Sigmodon hispidus*, was recently reported.[18] Humans are accidentally infected through the ingestion of contaminated food or water, developing a disease named abdominal angiostrongyliasis.[19,20] There is neither a diagnostic test nor a specific treatment for human angiostrongyliasis, and usual antihelminthic drugs are not effective against this nematode, and may worsen the clinical picture.[21] We expect that proteomic studies can ultimately contribute to the overall understanding of this uncharacterized worm, helping the development of an efficient treatment against this disease.

EXPERIMENTAL

Materials

3-[(3-Cholamidopropyl)dimethylammonio]1-propanesulfonate (CHAPS), dithiothreitol (DTT), iodoacetamide (IAA), urea, thiourea, tris base, Coomassie R-250, 4-sulphophenyl isothiocyanate (SPITC), ammonium bicarbonate and α-cyano hydroxycinaminic acid (CHCA) were purchased from Sigma-Aldrich (St Louis, USA). Sequencing grade modified trypsin was from Promega Corp. (Madison, USA). C18 ZipTip was purchased from Millipore Corp. (Billerica, USA). Trifluoroacetic acid (TFA) and sodium bicarbonate were purchased from Merck KGaA (Darmstadt, DE). Trichloroacetic acid (TCA) was from J. T. Baker (Phillipsburg, USA). Immobilized pH Gradient (IPG) strips, IPG buffer 4–7, Sample Grinding Kit and 2-D Quant Kit were from GE Healthcare (Chalfont St Giles, UK). All other reagents were either sequencing or HPLC grade.

Procedures

Protein extraction

Eighteen milligrams of isolated female adult *A. costaricensis* specimens[22] were rinsed with cold phosphate buffered saline and ground for 3 min into a 1.5 ml microcentrifuge tube containing abrasive resin (Sample Grinding Kit) and 300 µl of extraction solution made of 7 M urea, 2 M thiourea, 4% w/v CHAPS, 40 mM tris base, 60 mM DTT and 1% v/v IPG buffer 4–7. After centrifugation, the supernatant extract was analyzed for protein content (2-D Quant Kit) using bovine serum albumin as a standard.

2-D PAGE

One milligram of the protein extract was precipitated in 10% w/v TCA, on ice, for 2 h, washed twice with 90% acetone and dissolved overnight, at 4 °C, in 350 µl of the above-described extraction solution without tris. After centrifugation, the sample was applied to an 18 cm 3–10 linear strip using the sample in-gel rehydration method. Isoelectric focusing conditions in the IPGPhor (GE Healthcare) were: 30 V/12 h, 200 V/1 h, 500 V/1 h, 1000 V/1 h, up to 8000 V in 30 min and 8000 V/4 h. After focusing, the strip was incubated for 15 min in 10 ml of 50 mM tris-HCl buffer pH 8.8, 6 M urea, 30% v/v glycerol, 2% w/v sodium dodecyl sulfate, 0.002% bromophenol blue, 100 mg DTT, followed by a second incubation step in the same buffer solution, except for DTT, which was replaced by 400 mg IAA. After briefly rinsing the strip in tris-glycine electrode buffer,[23] it was placed directly on the top edge of a second-dimension slab gel (15% T) and ran in the DALTSix system (GE Healthcare) at 2.5 W/gel for 30 min and 100 W to the end. The gel was stained with Coomassie R-250 and digitalized using a conventional light scanner. Image analysis was performed using the ImageMaster 2-D Platinum software (GE Healthcare).

In-gel tryptic digestion

Digestion was performed as previously described[24] with modifications. Protein spots were excised from the gel, transferred to 0.5 ml tubes and cut into smaller pieces. Four hundred microliters of a 1 : 1 (v/v) 50 mM ammonium bicarbonate pH 8.0/acetonitrile solution was added to each tube, followed by shaking for 15 min, and then discarded. This washing procedure was repeated until destaining was complete. The washing solution was then removed and the gel dehydrated by the addition of 200 µl of acetonitrile

DOI: 10.1002/jms

for 5 min. The solvent was removed, and samples were completely dried in a vacuum centrifuge. Each sample was rehydrated with approximately 10 μl of ice-cold trypsin solution (20 ng/μl in 50 mM ammonium bicarbonate pH 8.0) and left on ice for 10 min. After gel reswelling, 20 μl of 50 mM ammonium bicarbonate pH 8.0 was added to the samples followed by incubation for 16 h at 37 °C. Following digestion, peptides were extracted by the addition of 15 μl of 50 mM ammonium bicarbonate pH 8.0 with ultrasonification for 10 min, and the solution was transferred to new 0.5 ml tubes. Each sample was divided into two equivolumetric aliquots and stored at −20 °C until use.

N-terminal chemical derivatization

The *N*-terminal chemical derivatization with SPITC was based on a previously described method[25] with some modifications. Briefly, a 5 to 10 μl aliquot of sample tryptic peptides (described above) was concentrated in a vacuum centrifuge to approximately 0.5 μl followed by the addition of 8.5 μl of reagent solution (10 mg/ml of SPITC in 20 mM $NaHCO_3$ pH 8.6) and incubated for 1 h at 56 °C. The reaction was stopped by addition of 1 μl 5% (v/v) TFA in water and stored at −20 °C until use.

Sample desalting

C18 ZipTip micropipette tips were used for the desalting of the derivatized and nonderivatized peptides. The tips were first activated with 50% acetonitrile in water and then equilibrated with 0.1% (v/v) TFA in water. The samples were aspirated and dispensed for eight cycles and the tips were washed with 0.1% (v/v) TFA in water three times. The peptides retained on the tips were eluted using 1 μl of a 50% acetonitrile, 0.1% (v/v) TFA in water. 0.3 μl of this eluate were immediately spotted on the ABI 192-target MALDI plate (Applied Biosystems, USA) by cocrystallization with 0.3 μl of CHCA matrix (10 mg/ml in 30% acetonitrile, 0.3% (v/v) TFA in water).

MALDI-TOF/TOF

MALDI-MS was performed on a 4700-Proteomics Analyzer with version 3.0 software (Applied Biosystems). MS spectra were acquired in positive ion reflector mode with 1250 laser shots per spot, processed with default calibration and the six most intense ions submitted to fragmentation. PSD spectra were acquired with 3000 laser shots and 1 keV collision energy with CID off (1×18^{-8} Torr). The MS data were analyzed by PMF using the MASCOT[26] software. The MS/MS data were analyzed both by running MASCOT as well as through manual analysis in order to obtain a larger number of *de novo* sequences to be compared with the NCBI nonredundant (NCBInr) database using the BLAST software[27] for short sequences.

MASCOT search parameters

PMF searches were carried out in the NCBInr database with no taxonomy restriction and a maximum peptide mass tolerance of 60 ppm, allowing one missing cleavage for trypsin and assuming carbamidomethylation of cysteine as a fixed modification. Variable modifications allowed were: methionine oxidation, *N*-terminal pyroglutamic acid and Ser/Thr phosphorylation. Identification was considered positive when the sequence coverage and MASCOT protein scores were bigger than 20 and 78% ($p < 0.05$), respectively. Furthermore, good correlation between experimental and theoretical molecular mass and pI values were taken into account.

MS/MS ion searches were carried out with the same parameters indicated above and included a 50 ppm maximum mass tolerance for fragment ions as well as SPITC as a variable modification. Positive identifications were accepted considering: (1) appearance of at least three to four consecutive fragment ions in spectrum; (2) precursor ion error smaller than 0.02 Da; (3) consistent fragment ion error pattern and (4) MASCOT individual ion scores indicating identity or extensive homology with $p < 0.05$.

RESULTS AND DISCUSSION

Figure 1 shows the bidimensional electrophoresis pattern of the proteins extracted from female adult *A. costaricensis*. We detected a total number of 434 spots from which we have selected 49 to be used in the present work. These spots are circled along with their respective numbers in the figure. Most of the spots selected are located in the 4–7 pI range, scattered all over the molecular mass interval and present variable staining intensities.

Figure 2(a) displays a traditional strategy for protein identification in proteome analysis currently used in our laboratory. This method is based on a MALDI-TOF/TOF running under PSD conditions, namely a 4700-Proteomics Analyzer from Applied Biosystems. Although it is technically

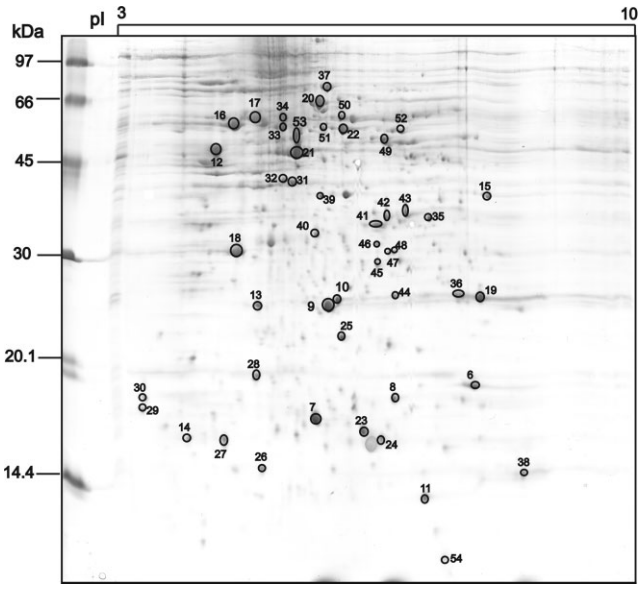

Figure 1. 2-D PAGE of 1 mg of *A. costaricensis* protein extract. In the first dimension, the sample was applied to an 18 cm 3–10 linear strip using the sample in-gel rehydration method. The second dimension was run in a 15% T slab gel, which was stained with Coomassie R-250 and analyzed with the ImageMaster 2-D Platinum software. Spots selected for MS analysis are outlined.

DOI: 10.1002/jms

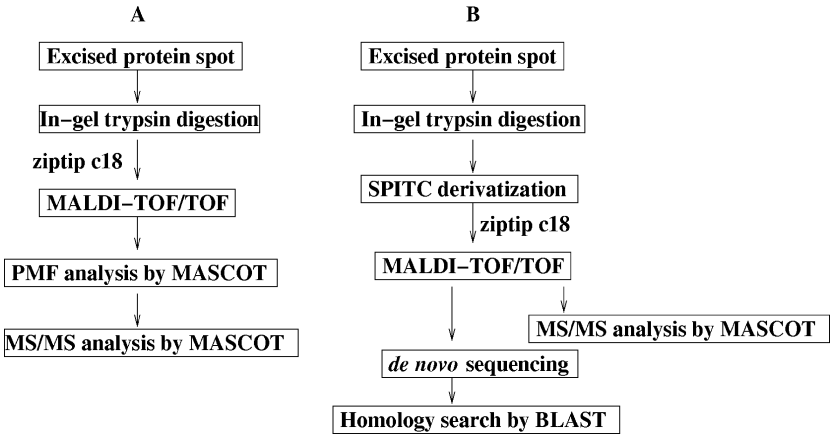

Figure 2. Schematic diagram showing: (A) traditional strategy for protein identification in proteome analysis, consisting of in-gel trypsin digestion, sample desalting/concentration and analysis by PMF followed by PSD of the six more intense peptides and identification by MS/MS ion search; (B) alternative approach consisting of in-gel trypsin digestion and peptide *N*-terminal sulfonation with SPITC, followed by desalting/concentration and analysis of MS/MS spectra both by MS/MS ion search using MASCOT and by manual interpretation associated to BLAST database search.

possible to run the instrument using collision-induced dissociation (CID), instead of PSD, it is our experience that the MS/MS spectra obtained are poorly informative, displaying mostly the low mass immonium ions (data not shown). On the other hand, PSD spectra are more complex and, hence, more difficult to interpret.[5–8] The traditional strategy depicted in Fig. 2(a) consists of initial sample desalting and concentration (using C_{18} reversed phase microcolumn) followed by two consecutive mass spectrometric stages: an initial MS broad (900–3500 Da) scan whose set of masses is used to carry out identification by PMF, followed by PSD analysis of the six more intense peptides and identification by MS/MS ion search.

Table 1 shows that the samples analyzed in the present work using the PMF approach yielded only 12% of positive identifications (spots 17, 18, 21, 33, 41 and 53). It is documented that for organisms with well annotated genomic sequences available in the databases most proteins can be identified by PMF with great confidence, except for heavily modified proteins or protein mixtures. In the case of an organism with uncharacterized genome, it is very difficult to identify proteins because most derived peptide masses will not match the theoretical masses calculated using similar genomes deposited in the databases.[4] This is the case for the present work where, basically, identification is being made by comparison to genomic data of a close organism that is well characterized, the nematode *Caenorhabditis elegans*. Figure 3(a) shows a MALDI-PSD spectrum of a tryptic peptide from spot 19. This spectrum may be considered as representative of the majority of spectra obtained by MALDI-PSD analysis in a TOF/TOF instrument. It confirms some limiting characteristics of this technology, regarding *de novo* sequencing, such as low fragmentation efficiency and complex spectra interpretation due to the presence of diverse fragment ions (b, y″, immonium and internal).[5] Consequently, these spectra are not ideal for *de novo* sequencing and what we do usually obtain is a short sequence tag (3–4 amino acid residues per spectrum). In most cases, this is not enough to obtain confident identification by homology searching using the BLAST software. However, extensive use of the PSD approach has revealed that the corresponding spectra contain the necessary information to identify proteins in the databases, whenever we are analyzing proteins with deposited sequences using MS/MS ion search strategy, such as the one used by the MASCOT software.[28] Hence, when we submitted the most intense peptides originated from MS mode to sequencing by PSD and analyzed the data using MASCOT's MS/MS ion search capability, we were able to identify 41% (20 spots) of the processed spots (Table 1 and Fig. 4).

A way to overcome PSD limitations, and hence get better identification yields, is to drive peptide fragmentation to favor the appearance of specific fragment ion series.[11,12,29] We used SPITC to introduce a fixed negative charge at the *N*-terminus[15] of all tryptic digests, desalted/concentrated the samples and analyzed the obtained MS/MS spectra both by MS/MS ion search using MASCOT and by manual interpretation associated with BLAST database search (Fig. 2(b)). As we can see in Fig. 3(b), the derivatized tryptic peptide from spot 19 has an addition of 215 Da corresponding to the mass of the modifying group. As expected, the major fragment ions present in the spectrum are y″-ions and, to a lesser degree, their z-ions. In this particular case, the whole y″-ion series is present, whereby ten consecutive amino acids could be *de novo* determined (Fig. 3(b)). The spectrum also shows the loss of the modifying group and sulfanilic acid.[30] When the MS/MS spectra of the SPITC modified peptides were analyzed manually (*de novo* sequencing) and submitted to BLAST homology search, our identification rate attained 76% (37 spots – Table 1 and Fig. 4). However, when the MS/MS spectra of the SPITC modified peptides were directly analyzed by MASCOT, only 17 spots (35%) could be identified (Table 1 and Fig. 4). This was not surprising since high identification yield by MS/MS ion search is also dependent on studied organism-specific genomic data.

Although no positive identifications were obtained for spots 8 and 49, we got good sequencing data using the

DOI: 10.1002/jms

Table 1. Proteins from *A. costaricensis* identified in this work[a]

Spot[b]	pI/*Mr*[c]	Method (PMF and/or MS/MS)	Sequence coverage (%)	MASCOT analysis No derivatization[d]	MASCOT analysis SPITC derivatization	De novo sequencing/BLAST SPITC derivatization[e,f]	Protein description
8	–	MS/MS	–	–	–	EENJDEFJSJR SCTVSR GVPIVR	NO IDENTIFICATION
9	5.80/24 435	MS/MS	–	HGEVCPAGWTPGK QJTVNDJPVGR JVQAFQFVDK DYGVJKDDEGJAYR	SVDETJR QJTVNDJPVGR DDEGJAYR	BJTVNDJPVGR DDEGJAYR SVDETJR	PeRoxireDoXin family member (prdx-2) [*C. elegans*]
10	5.94/25 065	MS/MS	–	–	GJFJJDPEGVVR QJTVNDJPVGR	GJFJJDPEGVVR . . .GVVR	PeRoxireDoXin family member (prdx-2) [*C. elegans*]
						NFGVJDEEAGVSFR	NO IDENTIFICATION
12	4.24/47 882	MS/MS	–	–	–	WEDYR WWADYDJDSGJDR	Calumenin-like protein [*C. elegans*]
13	4.83/24 698	MS/MS	–	–	–	NJQFFJGE**R**	Resistance to inhibitors of cholinesterase family member (ric-8)
15	8.04/38 598	MS/MS	–	–	–	DDJFNTNAGJVR FANA**J**VR	Malate dehydrogenase protein 1
16	4.48/55 368	MS/MS	–	AHAAETFEK KPEDWDER	–	*FYSR*	Calreticulin [*Necator americanus*]
17	4.76/57 682	MS/MS	–	VJDYTGDR JMEFFGJK SHNLLFVSK YADHENIIIAK	NFDQVAR VJDYTGDR	FD**E**GR FEVR NFDQVAR SWJQANR VJDYTGDR	Protein disulfide isomerase [*Ancylostoma caninum*]
		PMF	21				
18	4.53/29 951	MS/MS	–	VTEJGAEJSNEER SQQSYQEAFDJAK KVTEJGAEJSNEER MQPTHPJR	–	GDYYR J..QAER MQPTHPJR	Fourteen-three-three family member (ftt-2) [*C. elegans*]
		PMF	36				
19	7.93/25 355	MS/MS	–	–	–	JTYFNGR Q**PY**AJADQ**P**YED**V**R EJAQS**B**AJNR	Glutathione S-transferase [*Ancylostoma caninum*]

Table 1. (*Continued*)

Spot[b]	pI/*Mr*[c]	Method (PMF and/or MS/MS)	Sequence coverage (%)	MASCOT analysis No derivatization[d]	MASCOT analysis SPITC derivatization	*De novo* sequencing/BLAST SPITC derivatization[e,f]	Protein description
20	6.24/66 743	MS/MS	–	NVJJEQSWGSPK AAVEEGJVPGGGVAJJR		VTDAJCATR. . .AVE . . .AVEJR ED..JJR	Homologous to chaperonin protein [*C. elegans*]
21	5.34/46 819	MS/MS	–	AGFAGDDAPR GYSFTTTAER QEYDESGPSJVHR SYELPDGQVJTVGNER VAPEEHPVJJTEAPJNPK	AVFPSJVGR GYSFTTTAER QEYDESGPSJVHR	GYSFTTTAER BEYDESGPSJVHR JJAPJJR AVFPSJVGR	ACTin family member (act-4) [*C. elegans*]
		PMF	34				Similar to ACTin-5C [*Apis mellifera*]
22	6.00/53 570	MS/MS	–	–	–	A..QJJVVR ANAFEFJJGR	NO IDENTIFICATION
						SJYETJJR	CLC-type chloride channel family member (clh-2) [*C. elegans*]
23	6.29/16 467	MS/MS	–	–	–	E..KPNJEFVR E. . .GJMTR	F30A10.6 [*C. elegans*]
25	6.00/21 819	MS/MS	–	–	NFMJQGGDFTR	. . . SVVR NFMJQGGDFTR EDVVJA**DCGAJR**	Chain A, Cyclophilin_5 From [*C. elegans*]
26	4.88/14 895	MS/MS	–	–	–	V..VSJPR AAJVD**GTS**G**A**VWAR . . .**J**AVYEG**E**NEV**S**AQVR	TRYPSIN Profilin-1 [*C. elegans*]
27	4.36/16 128	MS/MS	–	EJFNJYDEEJDK	–	JDGTQ**J**GD**WR** HJJJAJGER JJ**S**AEJR	Myosin Light Chain family member (mlc-3) [*C. elegans*]
						V..VSJPR	TRYPSIN
28	4.81/19 225	MS/MS	–	–	–	WFR EHA**J**R	transposase [*C. elegans*]
						FQNJR	NO IDENTIFICATION
29	3.22/17 453	MS/MS	–	–	–	. . . GNJR VATVSJPR	TRYPSIN
31	5.32/40 933	MS/MS	–	APHFPQQPVAR FEVPQGAPTFTR JTGFSSPTFVEKPQJSSR	APHFPQQPVAR DDGQVMVMEFR FEVPQGAPTFTR	APHFPKKPVAR FEVP**B**GAP**T**FTR VJNVJR	Disorganized Muscle family member (dim-1) [*C. elegans*]

32	5.19/41 322	MS/MS	–	FEVPQGAPTFTR JTGFSSPTFVEKPQJSSR	–	APHFPQQPVAR DDGQVMVMEFR FEVPQGAPTFTR D... D**S**DAG**G**YR VJNVJR	Disorganized Muscle family member (dim-1) [*C. elegans*] NO IDENTIFICATION
33	5.18/55 295	MS/MS	–	FVSJEETJR JMNVJGEPJDER AHGGYSVFAGVGER VSJVYGQMNEPPGAR FTQAGSEVSAJJGR GJAELAJYPAVDPJDSTSR JJJEVSQHJGDNVVR	JJJEVSQHJGDNVVR JMNVJGEPJDER VSJVYGQMNEPPGAR FTQAGSEVSAJJGR FVSJEETJR	JJJEVSQHJGDNVVR JMNVJGEPJDER VSJVYGQMNEPPGAR FTQAGSEVSAJJGR TAJFGPNQNR FVSJEETJR	ATP synthase subunit family member (atp-2) [*C. elegans*]
		PMF	31				Atp synthase subunit protein 2 [*C. elegans*]
34	5.16/59 383	MS/MS	–	EDAANNYAR EJJDJVJDR	EDAANNYAR EJJDJVJDR	EDAANNYAR EJJDJVJDR ...VDEJR	TuBulin, Alpha family member (tba-8) [*C. elegans*]
35	7.22/34 943	MS/MS	–	–	–	JAYER QJJR SYJTJ**NR** JJATQNBYBJR JJTSJSNABR	NO IDENTIFICATION Annexin family protein 1 [*C. elegans*]
36	7.59/25 627	MS/MS	–	–	–	JTYFDGR EJAQS**B**AJNR GJAE**C**AR WJESR	Glutathione S-Transferase family member (gst-21) [*C. elegans*] NO IDENTIFICATION
37	5.79/72 183	MS/MS	–	DAGQJAGJNVJR VQQTVQDJFGR VJENAEGVR	VCQGER DAGQJAGJNVJR VJENAEGVR VQQTVQDJFGR		Heat shock protein 9B [*Danio rerio*]
						DAGBJAGJNVJR VJENAEGVR VQBTVQEVFGR	Heat Shock Protein family member (hsp-6) [*C. elegans*]
38	8.53/14 685	MS/MS	–	VHJTDAEK JHVDPENFR GTFASJSEJHCDK	–	VNHDAVGAEAJGR J..D..NFR	Hemoglobin beta subunit (Hemoglobin beta chain) (Beta-globin)

Table 1. (*Continued*)

Spot[b]	pI/*Mr*[c]	Method (PMF and/or MS/MS)	Sequence coverage (%)	MASCOT analysis No derivatization[d]	MASCOT analysis SPITC derivatization	*De novo* sequencing/BLAST SPITC derivatization[e,f]	Protein description
39	5.71/38 671	MS/MS	–	–	–	ATVEWFR G. . . YGJDYR JY. . . JVEJPR	Inorganic PYroPhosphatase family member (pyp-1) [*C. elegans*]
40	5.64/32 608	MS/MS	–	–	–	VYVGTTYAR	Aspartyl protease inhibitor [*Parelaphostrongylus tenuis*]
41	6.50/34 028	MS/MS		NSJJSNEWGNEER GEWGKEER	EYEHR FASYAHR NSJJSNEWGNEER	FASYAHR EYEHR DGDJAJHFNPR NSJJSNEWGNEER	Galectin [*Haemonchus contortus*]
		PMF	38	–	–	–	GaLECtin family member [*C. elegans*]
42	6.60/35 109	MS/MS		SADFSGNDVPJHJSVR	SYPVPYR DYEHR	DYEHR	32-kDa galectin [*C. elegans*]
43	6.89/35 747	MS/MS	–	EFKDYEHR FTSFAHR SADFSGNDVPJHJSVR	SADFSGNDVPJHJSVR FTSFAHR	SADFSGNDVPJHJSVR NAJQANEWGNEER FTSFAHR	GaLECtin family member [*C. elegans*] Galectin [*Haemonchus contortus*]
44	6.73/25 564	MS/MS	–	–	–	JEQTENYJFTHSDEG. . . GEMJGDR VATVSJPR	NO IDENTIFICATION
						JTYFDGR EJAQSBAJNR	Glutathione S-Transferase family member (gst-21) [*C. elegans*]
46	6.49/31 336	MS/MS	–	–	–	TDVPGECTHR	Hypothetical protein R11E3.4 [*C. elegans*]
47	6.65/30 200	MS/MS	–	NQYDSDVTVWSPQGR	JQANTQYYGR FJQTECSSWR	JQANTQYYGR FJQTECSSWR YEEJQJTAGR	Hypothetical protein CBG09365 [*C. briggsae*] Proteasome alpha subunit protein 6 [*C. elegans*] KERATIN
48	6.72/30 286	MS/MS	–	–	JQANTQYYGR	JQANTQYYGR	Hypothetical protein CBG09365 [*C. briggsae*]
49	6.58/51 557	MS/MS	–	–	–	FJQTECSSWR JDHNR. . . DHVYDR	NO IDENTIFICATION
50	5.98/58 836	MS/MS	–	–	–	DGVESQAYDGPR TQGNJFQFEQR . . .NDVP. . .FEVR	Protein Disulfide Isomerase family member (pdi-3) [*C. elegans*]

51	5.73/54 859	MS/MS	–			. . .JDQQJNTJTR	NO IDENTIFICATION
				VAFTGSTEIGR	VAFTGSTEIGR	. . .TGSTEJGR	Aldehyde deHydrogenase family member (alh-1)
						JYDDFVER	[*C. elegans*]
53	5.35/52 658	MS/MS	–	AGFAGDDAPR	GYSFTTTAER	AVFPSJVGR	Actin [*C. elegans*]
				QEYDESGPSJVHR	QEYDESGPSJVHR	GYSFTTTAER	
				GYSFTTTAER	–	JJAPPER	ACTin family member (act-4) [*C. elegans*]
				SYEJPDGQVJTVGNER		QEYDESGPSJVHR	
				VAPEEHPVJJTEAPJNPK		–	
		PMF	31	–	–	–	PREDICTED: similar to Actin-5C [*Apis mellifera*]
54	7.42/11 630	MS/MS	–	EGJPPDQQR	EGJPPDQQR	E. . .PPDQQR	UBiQuitin family member (ubq-1) [*C. elegans*]
				ESTJHJVJR	ESTJHJVJR	E. . .HJVJR	
				TJTJEVEASDTJENVK	–	–	
				–			

[a] The following strategies were employed: (1) PMF followed by PSD of the six more intense peptides and direct identification using MASCOT; (2) *N*-terminal sulfonation of the tryptic peptides with SPITC, followed by analysis of PSD spectra both by MS/MS ion search using MASCOT and by manual *de novo* sequencing associated to Blast database search.

[b] Spot numbers refer to spots excised from gel shown in Fig. 1.

[c] Experimental pI and Mr values were calculated using the ImageMaster 2-D Platinum software.

[d] Residue J can be either Ile or Leu and residue B stands for either Gln or Lys.

[e] Residues outlined in bold differ from the protein sequence deposited in the NCBInr database.

[f] Residues in italic are present in the identified protein sequence but did not suffice for an unambiguous identification.

JMS

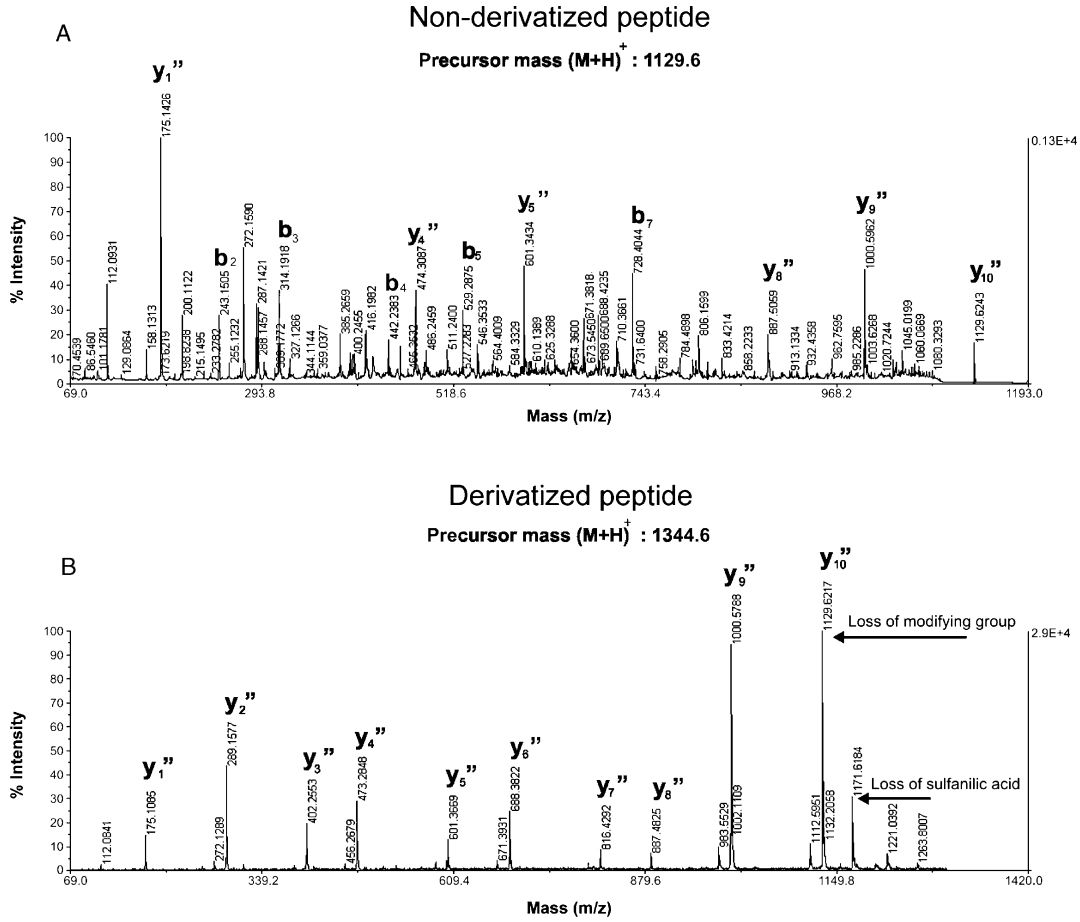

Figure 3. Example of PSD spectra, acquired on a 4700-Proteomics Analyzer mass spectrometer, of the same peptide, either (A) nonderivatized or (B) derivatized with SPITC. Determined peptide sequence was EJAQSBAJNR, where residue J can be either Ile or Leu and residue B stands for either Gln or Lys.

SPITC derivatization (Table 1). These data could be used for oligonucleotide synthesis in order to pursue protein identification by the use of molecular biology techniques.

Considering all the data obtained employing both strategies described in Fig. 2, we were able to achieve a 78% positive identification rate as well as generating, by *de novo* sequencing, protein sequences not available in NCBInr database. Moreover, even for those 20 nonderivatized spots readily identified by MASCOT, complementary SPITC-derivatized peptides data contributed to a higher sequence coverage since, in most cases, different peptides were sequenced (Table 1). It is important to note that this kind of information can be important when you need more structural data about the identified protein, as for example to determine the presence of different isoforms, as recently demonstrated,[31] or posttranslational modifications.

Another advantage of SPITC derivatization is the 215 Da increase in the precursor ion mass which, especially for tryptic peptides, promotes a mass shift of small peptides (3–4 amino acids) from the matrix cluster region (<900 Da) to a higher mass one, where they can be easily analyzed without matrix interference.

One drawback associated with SPITC is that because of the acidic nature of sulfonic groups that react with lysine's ε-amino group, MALDI-PSD analysis of derivatized peptides

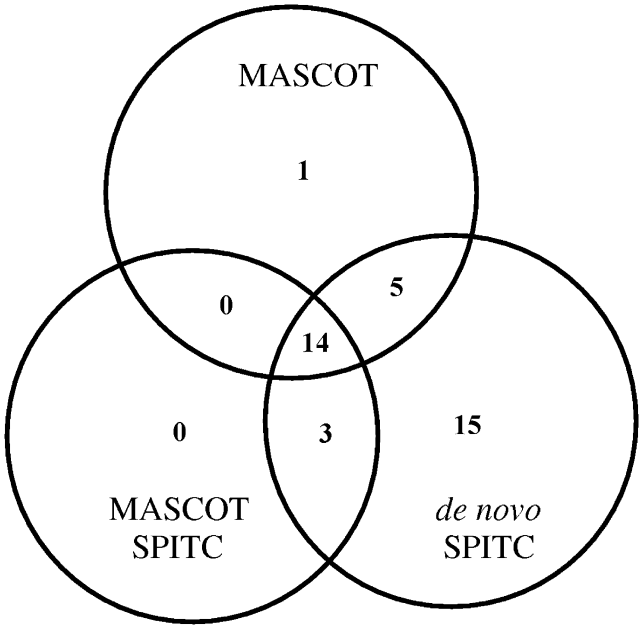

Figure 4. Venn diagram showing the number of proteins identified in this study by the following strategies: MASCOT analysis of nonderivatized samples, MASCOT analysis of SPITC modified samples and *de novo* sequencing of SPITC modified peptides.

DOI: 10.1002/jms

is insensitive to the detection of lysine-terminated peptides and, hence, only arginine-terminated peptides are detected. Furthermore, it should be noted that it has also been observed that peptides having unprotected lysine residues often form disulfonate products, which result in poor ionization and unpredictable fragmentations.[30]

To increase the positive identification rate using SPITC, one can analyze the lysine-terminated peptides by previously derivatizing them by guanidation[32] and thus avoiding unwanted sulfonation by SPITC. Alternatively, it has been recently demonstrated that SPITC derivatized peptides can be analyzed by reversed phase chromatography coupled to electrospray ionization (ESI) ion trap mass spectrometer. It should be noted, however, that the analysis of the spectra is not so straightforward since SPITC peptides submitted to these CID conditions lead to the formation of a major y_{n-1} ion along with low-intensity fragments that correspond to y''-series ions of the peptide.[30]

CONCLUSIONS

When analyzing 2D gel trypsin digested spots from an organism with unknown genome (*A. costaricensis*), the use of an approach consisting of SPITC derivatization along with manual spectra interpretation and BLAST analysis was able to positively identify 76% of analyzed samples, whereas MASCOT analysis of derivatized samples, MASCOT analysis of nonderivatized samples and PMF of nonderivatized samples yielded only 35, 41 and 12% positive identifications, respectively. Moreover, *de novo* sequencing of SPITC modified peptides resulted in protein sequences not available in NCBInr database, paving the way to the discovery of new molecules.

Acknowledgements

This study was supported by grants from Conselho Nacional de Desenvolvimento Científico and Tecnológico (CNPq), Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ) and Fundação Oswaldo Cruz-PDTIS, Brazil. IRL thanks Instituto Oswaldo Cruz for her Master's scholarship.

REFERENCES

1. Cottrell JS. Protein identification by peptide mass fingerprinting. *Peptide Research* 1994; **7**: 115.
2. Pappin DJ, Hojrup P, Bleasby AJ. Rapid identification of proteins by peptide-mass fingerprinting. *Current Biology* 1993; **3**: 327.
3. Mann M, Hojrup P, Roepstorff P. Use of mass spectrometric molecular weight information to identify proteins in sequence databases. *Biological Mass Spectrometry* 1993; **22**: 338.
4. Garbis S, Lubec G, Fountoulakis M. Limitations of current proteomics technologies. *Journal of Chromatography A* 2005; **1077**: 1.
5. Kaufmann R, Kirsch D, Spengler B. Sequencing of peptides in a time-of-flight mass spectrometer: evaluation of postsource decay following matrix-assisted laser desorption ionization (MALDI). *International Journal of Mass Spectrometry and Ion Processes* 1994; **131**: 355.
6. Yalcin T, Csizmadia IG, Peterson MR, Harrison AG. The structure and fragmentation of bn (n ≪= 3) ions in peptide spectra. *Journal of the American Society for Mass Spectrometry* 1996; **7**: 233.
7. Biemann K. Sequencing of peptides by tandem mass spectrometry and high-energy collision-induced dissociation. *Methods in Enzymology* 1990; **193**: 455.
8. Ambihapathy K, Yalcin T, Leung HW, Harrison AG. Pathways to immonium ions in the fragmentation of protonated peptides. *International Journal of Mass Spectrometry* 1997; **32**: 209.
9. Roepstorff P, Fohlman J. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomedical Mass Spectrometry* 1984; **11**: 601.
10. Chaurand P, Luetzenkirchen F, Spengler B. Peptide and protein identification by matrix-assisted laser desorption ionization (MALDI) and maldi-post-source decay time-of-flight mass spectrometry. *Journal of the American Society for Mass Spectrometry* 1999; **10**: 91.
11. Keough T, Lacey MP, Youngquist RS. Derivatization procedures to facilitate *de novo* sequencing of lysine-terminated tryptic peptides using postsource decay matrix-assisted laser desorption/ionization mass spectrometry. *Rapid Communications in Mass Spectrometry* 2000; **14**: 2348.
12. Keough T, Lacey MP, Youngquist RS. Solid-phase derivatization of tryptic peptides for rapid protein identification by matrix-assisted laser desorption/ionization mass spectrometry. *Rapid Communications in Mass Spectrometry* 2002; **16**: 1003.
13. Marekov LN, Steinert PM. Charge derivatization by 4-sulfophenyl isothiocyanate enhances peptide sequencing by post-source decay matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *International Journal of Mass Spectrometry* 2003; **38**: 373.
14. Mann M, Wilm M. Error-tolerant identification of peptides in sequence databases by peptide sequence tags. *Analytical Chemistry* 1994; **66**: 4390.
15. Keough T, Youngquist RS, Lacey MP. Sulfonic acid derivatives for peptide sequencing by MALDI MS. *Analytical Chemistry* 2003; **75**: 156A.
16. Ubelaker JE. Systematics of species referred to the genus *Angiostrongylus*. *Journal of Parasitology* 1986; **72**: 237.
17. Graeff-Teixeira C, Camillo-Coura L, Lenzi HL. Clinical and epidemiological aspects of abdominal angiostrongyliasis in southern Brazil. *Revista do Instituto de Medicina Tropical de São Paulo* 1991; **33**: 373.
18. Mota EM, Lenzi HL. *Angiostrongylus costaricensis*: complete redescription of the migratory pathways based on experimental *Sigmodon hispidus* infection. *Memórias do Instituto Oswaldo Cruz* 2005; **100**: 407.
19. Morera P, Cespedes R. *Angiostrongylus costaricensis* n. Sp. (nematoda: Metastrongyloidea), a new lungworm occurring in man in Costa Rica. *Revista de Biologia Tropical* 1970; **18**: 173.
20. Graeff-Teixeira C, Camillo-Coura L, Lenzi HL. Histopathological criteria for the diagnosis of abdominal angiostrongyliasis. *Parasitology Research* 1991; **77**: 606.
21. Morera P, Bontempo I. Acción de algunos antihelminticos sobre *Angiostrongylus costaricensis*. *Revista Médica del Hospital Nacional de Niños Dr. Carlos Saenz Herrera* 1985; **20**: 165.
22. Morera P. Life history and redescription of *Angiostrongylus costaricensis* Morera and Cespedes, 1971. *American Journal of Tropical Medicine and Hygiene* 1973; **22**: 613.
23. Laemmli UK. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 1970; **227**: 680.
24. Shevchenko A, Wilm M, Vorm O, Mann M. Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Analytical Chemistry* 1996; **68**: 850.
25. Wang D, Kalb SR, Cotter RJ. Improved procedures for N-terminal sulfonation of peptides for matrix-assisted laser desorption/ionization post-source decay peptide sequencing. *Rapid Communications in Mass Spectrometry* 2004; **18**: 96.
26. Perkins DN, Pappin DJ, Creasy DM, Cottrell JS. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 1999; **20**: 3551.
27. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* 1997; **25**: 3389.

28. Gevaert K, Demol H, Martens L, Hoorelbeke B, Puype M, Goethals M, Van Damme J, De Boeck S, Vandekerckhove J. Protein identification based on matrix assisted laser desorption/ionization-post source decay-mass spectrometry. *Electrophoresis* 2001; **22**: 1645.
29. Roth KD, Huang ZH, Sadagopan N, Watson JT. Charge derivatization of peptides for analysis by mass spectrometry. *Mass Spectrometry Reviews* 1998; **17**: 255.
30. Lee YH, Kim MS, Choie WS, Min HK, Lee SW. Highly informative proteome analysis by combining improved N-terminal sulfonation for *de novo* peptide sequencing and online capillary reverse-phase liquid chromatography/tandem mass spectrometry. *Proteomics* 2004; **4**: 1684.
31. Hjerno K, Alm R, Canback B, Matthiesen R, Trajkovski K, Bjork L, Roepstorff P, Emanuelsson C. Down-regulation of the strawberry Bet v 1-homologous allergen in concert with the flavonoid biosynthesis pathway in colorless strawberry mutant. *Proteomics* 2006; **6**: 1574.
32. Bonetto V, Bergman AC, Jornvall H, Sillard R. C-terminal sequence analysis of peptides and proteins using carboxypeptidases and mass spectrometry after derivatization of Lys and Cys residues. *Analytical Chemistry* 1997; **69**: 1315.

DOI: 10.1002/jms