Proteomic characterization of two snake venoms: *Naja naja atra* and *Agkistrodon halys*

Shuting LI*†, Jingqiang WANG*†, Xumin ZHANG*†, Yan REN*†, Ning WANG*, Kang ZHAO*†, Xishu CHEN*†, Caifeng ZHAO*†, Xiaolei LI*†, Jianmin SHAO*†, Jianning YIN*†, Matthew B. WEST‡, Ningzhi XU*† and Siqi LIU*†‡¹

*Beijing Genomics Institute (BGI), Chinese Academy of Sciences, I-Zone, Shunyi, Beijing 101300, China, †Beijing Proteomics Institute (BPI), I-Zone, Shunyi, Beijing 101300, China, and ‡The Department of Medicine, University of Louisville, Louisville, KY 40202, U.S.A.

Snake venom is a complex mixture of proteins and peptides, and a number of studies have described the biological properties of several venomous proteins. Nevertheless, a complete proteomic profile of venom from any of the many species of snake is not available. Proteomics now makes it possible to globally identify proteins from a complex mixture. To assess the venom proteomic profiles from *Naja naja atra* and *Agkistrodon halys*, snakes common to southern China, we used a combination strategy, which included the following four different approaches: (i) shotgun digestion plus HPLC with ion-trap tandem MS, (ii) one-dimensional SDS/PAGE plus HPLC with tandem MS, (iii) gel filtration plus HPLC with tandem MS and (iv) gel filtration and 2DE (two-dimensional gel electrophoresis) plus MALDI-TOF (matrix-assisted laser desorption ionization–time-of-flight) MS. In the present paper, we report the novel identification of 124 and

74 proteins and peptides in cobra and viper venom respectively. Functional analysis based upon toxin categories reveals that, as expected, cobra venom has a high abundance of cardio- and neurotoxins, whereas viper venom contains a significant amount of haemotoxins and metalloproteinases. Although approx. 80 % of gel spots from 2DE displayed high-quality MALDI-TOF-MS spectra, only 50 % of these spots were confirmed to be venom proteins, which is more than likely to be a result of incomplete protein databases. Interestingly, these data suggest that post-translational modification may be a significant characteristic of venomous proteins.

Key words: *Agkistrodon halys*, *Naja naja atra*, proteome, snake, venom.

INTRODUCTION

Snake bites are a serious health problem in many tropical and sub-tropical regions [1]. For instance, approx. 15 000–20 000 people are estimated to die annually in India due to snake envenomation [2]. The development of novel medical therapies to treat venomous snake bites is of great importance, and biochemical characterization of snake venom is imperative, because the underlying treatable pathogenesis is dependent on the venom's composition. In addition, previous studies have shown that some components of snake venom have beneficial attributes in the treatment of various pathophysiological conditions. Enzymes from cobra venom show promise in the treatment and/or prevention of Parkinson's and Alzheimer's diseases [3,4], and the venom from snakes in the viper family has been shown to promote tumour reduction [5]. Hence the elucidation of specific proteomic profiles of snake venom could have vast implications for medicine.

Snake venom is composed mainly of proteins and peptides, which possess a variety of biological activities. Snake venom is broadly divided into three categories based on toxicity from envenomation. These categories are: (i) haemotoxins, which promote haemorrhaging primary to extensive local swelling and necrosis, (ii) neurotoxins, which disable muscle contraction and paralyse the heart as well as hinder respiration, and (iii) cardiotoxins, which elicit specific toxicity to cardiac and muscle cells, causing irreversible depolarization of cell membranes [6].

The cobra and viper are two of the world's most poisonous snakes and are indigenous to the countries of Asia and Oceania.

Cobra venom is mainly categorized as a cytotoxin [7], whereas vipers employ mostly haemotoxic venom [8]. The specific composition of snake venom varies considerably from species to species [9]. This combination of toxins is the determinant of venom toxicity. Obviously, to elucidate the mechanism of venom toxicity, it is necessary to identify and characterize venomous proteins in an individual snake. The global analysis of snake venom presents a challenge for analytical techniques, which demands an approach that effectively separates a complex mixture of proteins and peptides with varying size and charge. Traditionally, venomous proteins and peptides are measured either by antibody recognition or by peptide N-terminal sequencing. Both methods are inadequate for global analysis of snake venom, therefore a new strategy must have the power to universally separate and identify the proteins and the peptides in snake venoms.

During the last several years, the field of proteomics has evolved considerably [10]. With the availability of genomic sequences, the power of 2DE (two-dimensional electrophoresis) and multiple-dimensional chromatographies to separate complex mixtures of proteins, advances in MS, and the development of computational methods, it is now possible to globally identify proteins expressed in a cell under a given set of conditions. Several reports have described the analysis of snake venom using proteomic strategies [11–15]. At the first Swiss Proteomics Society congress (SPS'01), a total of eight laboratory groups participated in an exercise examining protein identification using different mass spectrometric approaches [16]. One of the samples for the

Abbreviations used: *Agki*, *Agkistrodon halys*; 1DE, one-dimensional gel electrophoresis (SDS/PAGE); 2DE, two-dimensional gel electrophoresis; DTT, dithiothreitol; 5'-END, 5'-ectonucleotidase; GF, gel filtration chromatography; HMK, haemorrhagic metalloproteinase kaouthiagin; LC-MS/MS, HPLC with tandem MS; MALDI-TOF MS, matrix-assisted laser-desorption ionization–time-of-flight; MM, molecular mass; *Naja*, *Naja naja atra*; TFA, trifluoroacetic acid.

¹ To whom correspondence should be addressed (email siqiliu@louisville.edu).

exercise was snake venom from the Brazilian snake, *Bothrops jararaca*, provided by Dr D. C. Pimenta (Institute Butantan, São Paulo, Brazil). A total of 12 different MS approaches and five different protein search engines were employed in these laboratories. Although databases for venom proteins and peptides are incomplete, the venomous toxins were identified successfully by all participants regardless of the methodology used to identify them. The results from this exercise revealed the scope and efficiency of the current proteomic techniques in identifying venom proteins.

We are interested in developing an approach to distinguish snake species based upon proteomic characterization of their venom. In the present study, we have chosen two snakes, a cobra and a viper, the two major species of snake found in southern China [17]. The present study utilized four different proteomic strategies. The first involves direct tryptic digestion, followed by HPLC coupled with ion-trap tandem MS (shotgun-LC-MS/MS). The second uses 1DE [one-dimensional gel electrophoresis (SDS/PAGE)] to separate venomous proteins, then LC-MS/MS for identification of the separated proteins (1DE-LC-MS/MS). The third applies GF (gel filtration) to separate venom proteins followed by 1DE-LC-MS/MS. The last includes two steps for protein separation, GF (gel filtration chromatography) and 2DE, followed by protein identification with MALDI-TOF (matrix-assisted laser-desorption ionization-time-of-flight) MS (2DE-MALDI-TOF-MS). Using these strategies, we report here for the first time the complete proteomic profiles of venom from two different species of snake, in which 124 and 74 proteins and peptides, in cobra and viper venom respectively, have been verified. It is not surprising that the proteomic analysis provides the solid data to support the early observations that cobra venom contains abundant cardio- and neuro-toxins, and viper venom has high concentrations of haemotoxins and metalloproteinases. Furthermore, proteomic profiles offer a clear feature of protein distribution in both snake venoms, which is important for our understanding of envenomation mechanisms. In addition, the modifications of venomous proteins have been demonstrated by high resolution of 2DE techniques.

EXPERIMENTAL

Materials

Sephadex G-50, Immobiline DryStrips, ampholytes, TEMED (*N,N,N',N'*-tetramethylethylenediamine) and ammonium persulphate were purchased from Amersham Biosciences. Trypsin was from Promega. All other chemicals were from Merck. Deionized water was prepared with a tandem Milli-Q system and was used for the preparation of all buffers.

Snake venoms

Venom from *Naja naja atra* (*Naja*) from the cobra family and *Agkistrodon halys* (*Agki*) from the viper family were chosen as subjects for this study. Both snakes were purchased from Zhejiang Yiwu Snake Research Institute, Yiwu County, Zhejiang Province, China. The average length of the snake was 50 cm with an approximate age of 30 months. Snake venom was freshly drawn during the spring season and was frozen immediately at -80°C until use.

Preparation of venomous proteins for proteomic analysis

For shotgun or SDS/PAGE analysis, venom was diluted directly in trypsin digestion buffer or PBS. For 2DE analysis, venom proteins were precipitated by pre-cooled 10% TCA (trichloroacetic

acid)/acetone for 2 h at -20°C , followed by centrifugation at 13 000 *g* for 30 min. The precipitated proteins were washed once with acetone and twice with ethanol/ether (1:1, v/v), and finally dried by SpeedVac.

1DE

Venom samples (20 μg of protein) were reduced and denatured in SDS loading buffer and boiled for 3 min. The venomous proteins were resolved via SDS/PAGE using 15% polyacrylamide gels and a Hoefer electrophoresis device. The separated proteins were visualized by staining with Coomassie Brilliant Blue. Each lane on the gel was evenly excised to ten gel slices and was subjected to a complete tryptic digestion.

2DE

The precipitated venomous proteins were dissolved in 450 μl of rehydration solution containing 8 M urea, 2% (w/v) CHAPS, 20 mM DTT (dithiothreitol), 0.5% (v/v) IPG buffer, 0.002% (w/v) Bromophenol Blue. Commercial 24 cm IPG strips with a linear range of pH 3–10 were rehydrated overnight with 450 μl of venomous protein solution. Electrofocusing was carried out at 64 kV·h using IPGphor at 20°C following the manufacturer's instructions. Before the second dimension, the IPG strips were equilibrated by two equilibration steps: reduction buffer with 50 mM Tris/HCl, pH 8.8, 6 M urea, 30% (v/v) glycerol, 2% (w/v) SDS, a trace of Bromophenol Blue and 1% (w/v) DTT on a rocking table for 15 min; alkylation buffer with 50 mM Tris/HCl, pH 8.8, 6 M urea, 30% (v/v) glycerol, 2% (w/v) SDS, a trace of Bromophenol Blue and 2.5% (w/v) iodoacetamide for an additional 15 min. The electrophoresed strips were loaded and run on 10% and 15% polyacrylamide Laemmli gels (26 cm $\times$ 20 cm) for *Naja* and *Agki* respectively, using the Ettan DALT II system with a programmable power control, for 0.5 h at 0.5 W per gel, then at 15 W per gel until the dye front reached the bottom of the gel. The separated proteins were visualized by Coomassie Brilliant Blue staining.

GF

Venom samples (100 mg) were loaded on to a 8.0 mm $\times$ 250 mm Sephadex G-50 column pre-equilibrated with 10 mM Tris/HCl, pH 7.4, 10 mM DTT and 2 mM EDTA. Elution of venom was carried out using an equilibration buffer with a flow rate of 1.0 ml/min at room temperature (25°C). Protein elution was monitored at 280 nm. The eluted fractions were collected at 200 μl /tube, and the fractions were analysed further by SDS/PAGE to check separation efficiency. Based upon results from SDS/PAGE, the fractions containing venomous proteins with a MM (molecular mass) less than 10 kDa were pooled for direct tryptic digestion, and the remaining fractions were pooled for 2DE analysis.

Tryptic digestion

For direct digestion, venom samples or the pools from GF were diluted in a digestion solution containing 2 M urea, 50 mM ammonium bicarbonate and 1 mM CaCl_2 . The tryptic digestion was carried out by the addition of the modified trypsin to a final substrate/trypsin ratio of 40:1 (w/w) and by incubation at 37°C for 12 h. The enzymic digestion was stopped by acidification using 0.1% methanoic (formic) acid.

For in-gel digestion, gel slices from SDS/PAGE or gel spots from 2DE were carefully excised, and successively destained and dehydrated with 50% acetonitrile. The proteins were reduced

with 10 mM DTT at 56 °C for 1 h and alkylated by 55 mM iodoacetamide in the dark at room temperature for 45 min *in situ*. Finally, the gel slices or spots were thoroughly washed with 25 mM ammonium bicarbonate in water/acetonitrile (1:1, v/v) solution and were completely dried in a SpeedVac. Proteins were digested in 25 µl of modified trypsin solution (10 ng/µl in 25 mM ammonium bicarbonate) by incubation overnight at 37 °C. The peptides were released with vigorous shaking and were extracted in 50 µl of 50 % acetonitrile [containing 2.5 % TFA (trifluoroacetic acid)].

Separation of digestive peptides by HPLC

Digested peptides were separated on reverse-phase (C_{18}) capillary columns [5 µm, 300 Å (1 Å = 0.1 nm) particles, 0.15 mm × 100 mm or 0.15 mm × 150 mm; MicroTech] using an Agilent 1100 Capillary LC system. Flow was maintained at 100 µl/min, and elutants were analysed with a diode array detector. HPLC buffer A composed of 0.1 % methanoic acid in water, and buffer B composed of 0.1 % methanoic acid in acetonitrile, were used for peptide elution. A linear gradient of 2–80 % buffer B was employed, but at different gradient periods, 360 min for separation of the shotgun digestion of total snake venom, and 70 min for the peptides generated from in-gel digestions.

MS

Most digested peptides, obtained from shotgun digestion, the digestion of gel slices from SDS/PAGE, or digestion of pooled fractions (MM < 10 kDa) from GF, were analysed by ion-trap MS (Agilent MSD Trap SL). The m/z range 200–1800 was scanned in 1.2 s, and the fragment amplitude was set at 1.15 V. The MSD ion-trap mass spectrometer was operated in a data-dependent mode for MS/MS for the most abundant ions.

For the digested peptides from 2DE, the samples were mainly analysed by MALDI-TOF MS. The digestions were mixed with a matrix solution consisting of α -cyano-4-hydroxycinnamic acid (12 mg/ml) in 50 % acetonitrile with 0.1 % TFA at ratio of 1:1. The suspensions were applied on to the target well, dried at room temperature and analysed by a Bruker AutoFlex MALDI-TOF MS. The mass spectrometer was operated under 19 kV accelerating voltage in the reflectron mode with a m/z range 600–4000. Some digestive products from 2DE spots were analysed with LC-MS/MS for further confirmation of amino acid sequence.

Protein identification

The ion spectra of peptides generated by MSD trap and monoisotopic peptide masses obtained from MALDI-TOF MS were interpreted by utilizing the Mascot search engine (http://www.matrixscience.com/search_form_select.html; Matrix Science). The proposed peptide sequences were compared with the non-redundant databases of snake venomous proteins generated from data compiled at the NCBI (National Center for Biotechnology Information) and the EBI (European Bioinformatics Institute).

RESULTS

Proteomic strategy to analyse snake venom proteomes

In spite of extensive developments in proteomic techniques, there is not a panacea in the field. One method may have strong advantages for the analysis of some proteomes, but may have disadvantages for others. A major concern for the proteomic analysis of snake venom is the variable range of molecular sizes of the proteins and peptides. To alleviate this problem, we have adopted

Scheme 1 Schematic overview of combination strategy for analysing the proteome of snake venom

two strategies: (i) utilizing techniques such as shotgun-LC-MS/MS and 1DE-LC-MS/MS, whereby direct tryptic digestion or electrophoresis could minimize loss of proteins and peptides, and (ii) using GF to separate small venom peptides from the larger proteins; the separated proteins and peptides treated further by 2DE or HPLC lead to better resolution for venom components. The experimental procedure is depicted in Scheme 1. Two techniques of MS, MALDI-TOF MS and LC-MS/MS, were used for protein identification in these experiments. In contrast with MALDI-TOF MS, LC-MS/MS offers the advantages of reverse-phase HPLC for more extensive separation of peptide mixtures, and ion-trap for better resolution of peptide sequences. Thus we conducted protein identification of 2DE gel spots using MALDI-TOF MS, but, in most cases, LC-MS/MS was the primary means for protein identification.

Comparison of snake venom proteomes using the shotgun digestion approach

A typical ion chromatogram for the tryptic digestion of *Agki* venom proteins is shown in Figure 1. The peptide chromatogram monitored at 214 nm in Figure 1(A) indicates that the loaded peptides are completely eluted at an acetonitrile concentration of 80 %. The reproducibility of the identified proteins was approx. 96 % among parallel experiments (three injections/sample). A full-scan mass spectrum from LC-MS/MS analysis of *Agki* venom is shown in Figure 1(B), a single MS survey scan at elution time of 82.13 min is represented in Figure 1(C) and a MS/MS scan of the precursor ion with a m/z value of 567.11 triggered by the survey scan is depicted in Figure 1(D). Depending on the population of the constituent peptides, approx. 4000 MS/MS spectra were automatically collected for each injection of *Naja* or *Agki* venom. After acquisition of MS/MS spectra, the proteins or peptides were identified by the software algorithm Mascot. Of the spectra, approx. 10 % gave reliable candidate peptides by searching NCBI and EBI protein databases of snake venom. A total of 78 and 28 unique proteins or peptides were identified from *Naja* and *Agki* venom respectively. All of the unique proteins were determined by more than two peptides matched. The information regarding the identified proteins, such as the matched peptide sequences, accession number, theoretical pI and MM, their charge states,

Figure 1 Analysis of *Agki* proteome by shotgun-LC/MS approach

(A) Elution profile of digestive peptides from *Agki* venom from HPLC monitored at 214 nm. (B) Peptide base peak ion chromatogram of digestive peptides from *Agki* venom. (C) An MS survey scan at time 82.13 min during LC-MS analysis. The parent ion 567.11 was selected further MS/MS analysis. (D) The MS/MS spectra for parent ion 567.11. The amino acid sequence, FVELVLVADK, was confirmed by analysing b- and y-ions derived from the peptide ion.

and *m/z* values, are given in Supplementary Table I (available at <http://www.BiochemJ.org/bj/384/bj3840119add.htm>).

To compare the biological functions of the identified proteins in this study, venomous proteins or peptides are broadly classified into three toxin groups classified by their toxicity to animals: (i) neurotoxin, (ii) cardiotoxin and (iii) haemotoxin. Two major protein families, phospholipase A₂ or metalloproteinase, generally exist in snake venoms; they are hence listed as specific

categories in this study. It is worth noting that the delineation of protein categories is not, in every case, clear cut due to the multiple toxic attributes of proteins.

According to the classification described above, the distributions of the identified proteins in five categories are listed in Table 1. As expected, *Naja* venom contains high amounts of cardio- and neuro-toxins, both occupying approx. 60% of the composition, and *Agki* has a high abundance of proteins involved in disruption of the haemostatic system, with 28% of haemotoxins and 40% of metalloproteinases. The content of phospholipase A₂ in both venoms is close, at 18% and 14% respectively. Although some identified proteins have diverse amino acid sequences, these proteins may belong to the same protein family, at least performing similar biological functions. For instance, five unique proteins identified in *Naja* by this approach are assigned as GI numbers 1000502, 1326087, 1661022, 1134871 and 299268 that are the members of cardiotoxin 1 family with over 80% identity of amino acid sequence (see Supplementary Tables at <http://www.BiochemJ.org/bj/384/bj3840119add.htm>).

Comparison of snake venom proteomes using a 1DE-LC-MS/MS approach

Figure 2 shows 15% polyacrylamide gels that were loaded and electrophoresed with 300 µg of snake venom proteins. As shown, the major components of *Naja* venom are distributed at a MM lower than 10 kDa, whereas *Agki* venom proteins are spread evenly through a wide range of MMs. To separate the proteins effectively, each SDS/PAGE gel was evenly divided into ten slices, approx. 5 mm/slice, and subsequently digested by trypsin and separated by reverse-phase HPLC followed by analysis of tandem mass spectrometry. The collected MS/MS spectra varied from 500 to 2000/slice. In total, 5500 and 9000 MS/MS spectra were collected for *Naja* and *Agki* venom respectively. The spectra matched 65 and 43 unique proteins or peptides for *Naja* and *Agki* venom respectively (see details in Supplementary Table I at <http://www.BiochemJ.org/bj/384/bj3840119add.htm>).

These identified venomous proteins or peptides are categorized in Table 1. In *Naja* venom, 74% of the identified proteins and peptides are cardiotoxins and neurotoxins, but only one haemotoxin was found in this venom. In contrast with *Naja*, *Agki* venom has 23% haemotoxin and 30% metalloproteinases, but only three cardiotoxins were detected, in which cardiotoxin-2c (GI number 3342766) is commonly found in both venoms. Comparing the differences between theoretical and apparent values of MM, 94% of the proteins identified from *Naja* venom display good correlations, whereas 46% of *Agki* proteins closely matched their theoretical predictions. Over 50% of *Agki* proteins migrated on SDS/PAGE with significantly low apparent MMs, suggesting that the degradation caused by proteases is common in *Agki* venom. Interestingly, none of the phospholipases A₂ in *Agki* were found degraded; on the contrary, more than 80% of metalloproteinases appeared as the reduced molecular sizes on SDS/PAGE.

Comparison of snake venom proteomes using GF-LC-MS/MS and GF-2DE-MALDI-TOF-MS approaches

The approach of 2DE is a powerful technique to separate and identify soluble proteins; however, it is limited to the analysis of small proteins and peptides. Therefore size-exclusion chromatography was employed to separate venom proteins into two crude fractions, small proteins or peptides (MM < 10 kDa) and larger proteins (MM > 10 kDa). The small proteins were subjected directly to tryptic digestion, and the large proteins were separated

Table 1 Summary of the identified venomous proteins from *Naja* (a) and *Agki* (b) determined by different proteomics approaches

The numbers in parentheses indicate how many proteins were identified by an approach.

(a)					
Approach	Shotgun-LC-MS/MS (%)	1DE-LC-MS/MS (%)	GF-LC-MS/MS (%)	GF-2DE-MALDI-TOF-MS (%)	All approaches (%)
Venom proteins					
Cardiotoxins	56 (44)	70 (46)	78 (60)	—	56 (68)
Neurotoxins	5 (4)	4 (2)	13 (10)	—	8 (11)
Haemotoxins	3 (2)	2 (1)	—	13 (2)	2 (3)
Phospholipases A ₂	18 (14)	7 (5)	4 (3)	—	11 (15)
Metalloproteinases	—	—	—	—	—
Other proteins	18 (14)	17 (11)	5 (4)	87 (14)	22 (27)
(b)					
Approach	Shotgun-LC-MS/MS (%)	1DE-LC-MS/MS (%)	GF-LC-MS/MS (%)	GF-2DE-MALDI-TOF-MS (%)	All approaches (%)
Venom proteins					
Cardiotoxins	7 (2)	7 (3)	8 (4)	—	8 (6)
Neurotoxins	—	—	2 (1)	—	1 (1)
Haemotoxins	28 (8)	23 (10)	32 (15)	27 (4)	35 (26)
Phospholipases A ₂	14 (4)	23 (10)	15 (7)	27 (4)	18 (13)
Metalloproteinases	40 (11)	30 (13)	30 (14)	33 (5)	27 (20)
Other proteins	11 (3)	16 (7)	13 (6)	13 (2)	11 (8)

Figure 2 Separation of the venomous proteins by SDS/15% PAGE with Coomassie Brilliant Blue staining

Lane M, protein ladder; lane A, *Naja* venomous proteins; lane B, *Agki* venomous proteins. MM values are given in kDa.

further by 2DE. As shown in Figure 3, the eluted fractions from Sephadex G-50 GF were examined by SDS/PAGE, in which the venomous proteins are roughly separated according to their molecular sizes. Obviously, the protein profile of *Naja* venom is different from that of *Agki*, owing to the abundance of small proteins and peptides eluted in later fractions. The initial fractions containing the large proteins (> 10 kDa) were pooled and subjected to 2DE (1 mg of protein/gel). As 2DE separation of venomous proteins shows in Figure 4, the striking difference between the two venom samples is the distribution pattern of proteins along the horizontal pH range. Notably, most proteins from the *Naja* sample migrated to basic regions (pH > 7), whereas most of the *Agki* venom proteins were located in the acidic pH region (pH < 7).

The fractions containing proteins with MM < 10 kDa were pooled, digested by trypsin and subjected to LC-MS/MS as described above for the shotgun approach. Approx. 3500 and

2500 MS/MS spectra were collected in a single operation from *Naja* and *Agki* venom respectively. Of these spectra, approx. 10% gave reliable candidate peptides by searching the snake venom protein database. Totals of 77 and 48 unique proteins or peptides were confirmed for *Naja* and *Agki* venom respectively (Table 1; details in Supplementary Table I at <http://www.BiochemJ.org/bj/384/bj3840119add.htm>). The proteins with apparent sizes more than 10 kDa were analysed by 2DE and subjected to MALDI-TOF-MS after staining by Coomassie Brilliant Blue and in-gel digestion as described above. A mass fingerprint of the unfractionated peptide mixture was obtained by MALDI-TOF-MS followed by Mascot protein identification mapping. To ensure the accuracy of protein identification, the following criteria were followed for each sample: (i) ranked in top two hits; (ii) at least four matched sequences; and (iii) over 10% protein sequence coverage. Totals of 190 and 169 spots from *Naja* and *Agki* respectively were picked from the corresponding gel. For *Naja* venom, 152 high-quality MALDI-TOF-MS spectra were collected and 100 spectra were matched to the correct peptides. From these samples, 16 unique proteins were assigned. For *Agki* venom, 133 high-quality MALDI-TOF-MS spectra were collected and 85 spectra matched the correct peptides. *Agki* venom had 15 unique confirmed proteins (Table 1; details in Supplementary Table II at <http://www.BiochemJ.org/bj/384/bj3840119add.htm>).

Following Sephadex G-50 GF, the pool of small venom proteins is expected to contain few, if any, large proteins (MM > 10 kDa). Peptide identification of *Naja* venom by LC-MS/MS indicated that approx. 90% of proteins were cardio- and neuro-toxins, and over 91% had a MM less than 10 kDa. Protein degradation was apparent in the pool collected from late fractions of GF from the *Agki* sample, where approx. 90% of the proteins identified by MS have the mass values of more than 10 kDa. Surprisingly, the number of unique proteins verified by MALDI-TOF-MS from 2DE gel spots was much less than that of the identified proteins, with an average of six spots/unique protein. These data suggest that degradation or modification were prevalent in both samples. For instance, in *Naja* venom, 5'-END (5'-ectonucleotidase) from *Rattus norvegicus* (GI number 11024643) has theoretical pI and MM values of 6.5 and 64 kDa respectively. A total of 28 spots from 2DE were verified as 5'-END, with MMs ranging from 29

Figure 3 Separation of the venomous proteins by GF (Sephadex G-50)

(A) *Naja*. (B) *Agki*. Upper panels, the elution profiles of snake venomous proteins monitored at 260 nm and 280 nm. Lower panels, the eluted fractions analysed by SDS/15 % PAGE with Coomassie Brilliant Blue staining. MM values are given in kDa. mAU, milli-absorbance units.

to 64 kDa and pIs ranging from 4.9 to 9.0. In *Agki* venom, a metalloproteinase from *Gloydius halys* (GI number 4106001) has theoretical pI and MM values of 5.3 and 70 kDa respectively; however, eight protein spots were identified as this enzyme on 2DE, with MMs ranging from 28 to 62 kDa and with pI values in the narrow range of 4.5 to 5.3. Interestingly, several venomous proteins were found to have higher MMs than their theoretical values. Haemorrhagic metalloproteinase kaouthiagin (GI number 32469675; 44 kDa) in *Naja* and salmobin (GI number 3668352; 29 kDa) in *Agki* migrated with MMs of 100 and 48 kDa respectively.

Summarized comparisons of snake venom proteomes

The overall comparison of venom proteins identified by the different approaches is summarized in Table 1. In total, 124 and 74 proteins were verified from *Naja* and *Agki* venoms respectively (see Supplementary Table III at <http://www.BiochemJ.org/bj/384/bj3840119add.htm>). Considering that current techniques limit the analysis of intact proteins by MS, generation of detectable peptides is crucial to the identification of proteins. In the present study, two different digestion approaches, shotgun and in-gel digestion, have been employed to achieve the peptide signals to a great extent. As described above, the samples from diluted snake venom or from the GF pool of small-mass proteins were chopped by shotgun tryptic digestion and subsequently analysed by LC-MS/MS. Combining the two approaches, a total of 110 and 56 proteins were identified from *Naja* and *Agki* respectively. With in-gel digestion after 1DE and 2DE, a total of 78 and 47 proteins were confirmed by MS from *Naja* and *Agki* respectively (see Supplementary Tables I and II at <http://www.BiochemJ.org/bj/384/bj3840119add.htm>).

As mentioned above, snake venom proteins are broadly categorized into three groups based on toxicity and two groups based on sequence homology. A general summary of the proteomes of *Naja* and *Agki* venom is briefly discussed below.

Haemotoxin

The haemotoxins are the major protein components in *Agki* venom, with a total of 26 confirmed proteins, including many involved in haemostatic regulation, such as serine proteases,

coagulation factors, haemorrhagic toxins, disintegrin and fibrinogen clotting inhibitor (detailed in Supplementary Table III at <http://www.BiochemJ.org/bj/384/bj3840119add.htm>). In these identified haemotoxins, some belong to the same protein family, for instance, the sequence identities of *Agki* platelet aggregation disintegrins (GI numbers 104444, 118651, 25527065, 25527067, 25527068 and 25527069) are over 70%. Only three are haemotoxins found in *Naja* venom: two haemorrhagic metalloproteinases (GI numbers 20530119 and 32469675) and one salmobin (GI number 3044080). The haemorrhagic metalloproteinases are specific for the cobra species, and the salmobin is identical with one found in *Agki* venom.

Neurotoxin

Naja venom contains 11 confirmed neurotoxins, all of which are cobra-specific. Some *Naja* neurotoxins share high identity in amino acid sequences, such as cobrotoxin III (GI number 4185752) being 93% identical with cobrotoxin IV (GI number 4185750), and neurotoxin 6 (GI number 85982) sharing 71% identity with neurotoxin 7 (GI 85983). Only one neurotoxin, acutolysin from *Deinagkistrodon acutus*, was detected in *Agki* venom, which does not overlap any amino acid sequences from *Naja* neurotoxins.

Cardiotoxin

As shown in Table 1, the cardiotoxins are the most abundant components of *Naja* venom, with MMs ranging from 6.7 to 10 kDa. Based upon the identities of amino acid sequences, the 68 *Naja* cardiotoxins could be broadly grouped into 14 different families, as illustrated in Supplementary Table III (available at <http://www.BiochemJ.org/bj/384/bj3840119add.htm>). Six cardiotoxins were verified in *Agki* venom, in which two cardiotoxins (GI numbers 1054811 and 3342766) are the same as *Naja* cardiotoxins, and the other four are viper-specific proteins.

Phospholipase A₂

In contrast with other protein categories, phospholipase A₂ is a common protein family in both snake venoms with relatively high abundance, represented by 15 and 13 proteins in *Naja* and *Agki*

Figure 4 Separation of the venomous proteins by 2DE approach

(A) The *Naja* venomous proteins collected from G-50 GF with MM more than 10 kDa loaded on to SDS/10% PAGE. (B) The diluted *Agki* venomous proteins loaded on to SDS/12% PAGE. The pH gradient ranged from 3 to 10 and the gels were stained by Coomassie Brilliant Blue. MM values are given in kDa.

respectively. However, phospholipase A₂ is species-dependent, and no common phospholipase A₂ was identified in both species in this study. Of 15 *Naja* phospholipases A₂, 67% of these proteins have over 70% identity in amino acid sequence; however, these have low similarity to *Agki* phospholipase A₂.

Metalloproteinase

In the present study, no metalloproteinase has been identified in *Naja* venom, even though two haemorrhagic metalloproteinases were found and categorized as haemotoxins. However, metalloproteinase was confirmed a major protein component in *Agki* venom using all of the proteomic methods. The *Agki* metalloproteinases usually have 30–60 kDa theoretical MMs and slightly

acidic pI values, which are found in a good separation by 2DE. In Table 1(B), nevertheless, only five metalloproteinases were identified by 2DE-MALDI-TOF-MS; each of the other three approaches measured over ten metalloproteinases. Surprisingly, LC-MS/MS analysis monitored 14 metalloproteinases in the small MM pool (MM < 10 kDa) after G-50 GF. Furthermore, the image analysis indicates that multiple spots from a single metalloproteinase with smaller apparent MMs are common phenomenon on 2DE gels. For example, metalloproteinase (GI number 231997), having theoretical MM of 66 kDa and pI of 5.2, was found as ten Coomassie-Blue-staining spots on 2DE with the apparent mass from 30 to 40 kDa and pI from 4.7 to 6.3 (see Supplementary Table IIB at <http://www.BiochemJ.org/bj/384/bj3840119add.htm>). Taken together, the metalloproteinases were revealed to be strongly autocleaved either in the venom gland or during sample preparation.

Other proteins

This category represents, for the most part, species-specific proteins such as latisemin, ophanin and natrin specific for cobra, and saxatilin, catrocollastatin and mamushigin specific for viper. No common protein is shared by both snake venoms in this category.

DISCUSSION

Evolution has presented Nature with a tremendously diverse number of proteins and peptides in snake venom with equally diverse biological functions. With traditional analysis of snake venom, the primary focus has been on understanding the structural as well as functional aspects of individual proteins and peptides [18]. These studies have been successful in providing important information regarding specific components of venom as well as venom-induced pathologies. Nevertheless, these studies fall short in global identification of venom proteins, which is necessary for our understanding of the complexity of venom toxicity as well as elucidating the major biochemical differences among snake families. More than 40 years ago, researchers conducted electrophoresis experiments to globally isolate venom proteins [19]. Interestingly, what they noticed with starch electrophoresis was very similar to our proteomic observations: that on 2DE, *Naja* venom proteins appeared to be basic and *Agki* venom was mainly composed of acidic proteins. The limitations that existed at the time precluded more detailed analysis that is now possible with current technology. Advances in proteomic techniques have given researchers an edge in protein identification and characterization. Recently these methods have been used in several venom studies. Rioux et al. [11] separated the proteomes from *Laticauda colubrine* and *Vipera russelli* with 2DE and analysed the venom proteins by microsequencing of their N-terminal sequences and internal fragments. Using LC-MS, Fry et al. [14] measured the peptides from eight snake venoms and claimed that a hitherto unsuspected diversity of toxins was present in all lineages. Nawarak et al. [15] delineated the proteomes of eight snake venoms from elapidae and viperidae using multidimensional chromatographic approaches. The venom proteins from the elapidae family displayed similar distribution patterns on 2DE, whereby most protein spots had high pIs and low MMs, and the proteins obtained from the viperidae family were found in the neutral pH range with MMs of 30–45 kDa. Obviously, these investigations have revealed substantial information for the studies of venomous proteins; however, these data are short in convincingly demonstrating a complete proteomic profile for any species of snake venom. To our knowledge, this study

Table 2 Comparison of the identified venomous proteins and their sequence sources in database of snake proteins

All sequence data from snake venoms, cobra and viper, are from NCBI and EBI. The information of species sources are provided by submitters.

		Total identified proteins	Sequence source from cobra	Sequence source from viper	Sequence source from other species
<i>Naja</i>	Shotgun-LC-MS/MS	78	70 (90 %)	2 (2 %)	6 (8 %)
	1DE-LC-MS/MS	65	57 (87 %)	1 (1.5 %)	7 (10.7 %)
	GF-LC-MS/MS	77	76 (98.7 %)	—	1 (1.3 %)
	GF-2DE-MALDI-TOF-MS	16	5 (31 %)	1 (6.3 %)	10 (62.5 %)
<i>Agki</i>	Shotgun-LC-MS/MS	28	1 (3.5 %)	26 (93 %)	1 (3.5 %)
	1DE-LC-MS/MS	43	1 (2 %)	42 (98 %)	—
	GF-LC-MS/MS	47	1 (1.5 %)	46 (97 %)	1 (1.5 %)
	GF-2DE-MALDI-TOF-MS	15	—	14 (93 %)	1 (6 %)

is the first to document a complete proteomic profile of snake venom. Since the proteins and peptides in snake venom span a wide range of molecular size as well as biochemical properties, proteomic analysis with a single approach could certainly result in loss of proteins in separation as well as identification.

In spite of great advancements in recent years, proteomic analysis is still considered to be a prototype technique. To date, there is no unique method that can universally be utilized for all proteomic analyses [20]. In the present study, we have employed a combination strategy to profile the venom proteomes. This strategy includes three different combinations: (i) separation of proteins using HPLC and electrophoresis, (ii) digestion of proteins with trypsin, in-gel or aqueous, and (iii) identification of protein by ion-trap MS, tandem MS and MALDI-TOF MS. Using these combinations, it is reasonable that a comprehensive profile for venom proteomes can be established. In *Naja* venom, 31 proteins were detected by all of the proteomic approaches, whereas only 14 *Agki* proteins were found in the same way. There are a few proteins uniquely identified by a single approach. For instance, in *Naja* venom, 13 proteins were only detected by shotgun-LC-MS/MS, two by 1DE-LC-MS/MS and eight by GF-LC-MS/MS; in *Agki* venom, none was detected only by shotgun-LC-MS/MS, 15 by 1DE-LC-MS/MS and 12 by GF-LC-MS/MS. Moreover, the detection coverage of venomous proteins is compensated by each approach. GF-LC-MS/MS played well in the detection of cardiotoxins, but missed the signals of phospholipase A₂, and shotgun-LC-MS/MS and 1DE-LC-MS/MS performed sensitive identification for other large sizes of proteins. In the present study, most identified proteins were identified by at least two approaches; thus the combination strategy for proteomic analysis not only enhances the detection rate, but also ensures the accuracy of the measurements.

To generate a database of snake venom proteins, we collected all snake-venom-related data from NCBI and EBI. Interestingly, protein searches based upon mass signals from either MALDI-TOF-MS or LC-MS/MS consistently retrieve species-specific information. As summarized in Table 2, 91.9 ± 6.1 % of identified proteins from *Naja* (except for 2DE-MALDI-TOF-MS data) belong to the cobra species and 95.1 ± 2.8 % of the *Agki* proteins are from the viper species. For instance, phospholipase A₂ is commonly present in both venoms. Verifying biological sources, 96 % of the identified phospholipase A₂ from *Agki* venom in the present study matches well with the enzymes from vipers listed in the database, and 94 % of the proteins in *Naja* venom closely matched cobra proteins. These results demonstrate that the accuracy of protein identification based upon mass signals from the digestive peptides is robust and accurate. Furthermore, it is worthwhile to point out that proteomic analysis provides a set of fundamental data, such as how many proteins can be expressed in

a certain species or a specific period; however, it does not answer all of the questions related to venom proteins, especially for their biological reactivity. For example, four functional domains, preprosequence, metalloproteinase, disintegrin and cysteine-rich, may exist in snake venom metalloproteinase [21]. Herein the proteomic analysis has identified the peptide fragments from preprosequence, metalloproteinase and disintegrin by MS. Nevertheless, these data cannot give a certain answer as to whether all of the domains are in one compact molecule or are in free forms.

On 2DE images (Figure 4), totals of 190 and 169 protein spots are detected by Coomassie Blue staining from *Naja* or *Agki* respectively. Approximately half of these spots, 54 % for *Naja* and 50 % for *Agki*, were identified as venom proteins by MALDI-TOF-MS. However, the unique proteins are only 16 from *Naja* and 15 from *Agki* venom. Therefore it is reasonable to assume that these proteins must undergo a series of post-translational modifications in the snake venom gland. Comparing the apparent mass sizes of these modified proteins with the theoretically predicted values, 24 % (24/100) of *Naja* proteins and 61 % (52/85) of *Agki* proteins appear to be significantly reduced in size. This is logical due to the fact that *Agki* venom contains abundant proteinases, both metalloproteinases and serine proteases. In contrast with *Agki*, *Naja* has fewer proteases which may account for fewer degraded protein products. Interestingly, several venom proteins displayed significant increases in their apparent MMs. For instance, 27 gel spots, which represent almost 14 % of the detected spots on 2DE from *Naja* venom, were confirmed to be the same protein, HMK (haemorrhagic metalloproteinase kaouthiagin; GI number 32469675), with a theoretical MM of 44 kDa. The image analysis revealed that 89 % (24 out of 27) of this protein appeared to have higher MMs, ranging from 55 to more than 100 kDa on 2DE. Moreover, the similar conclusion was drawn from 1DE-LC-MS/MS data. Not only HMK, but also several venomous proteins, natrin (GI number 32492059), actin (GI number 113307), 5'-END (GI number 11024643), and nerve growth factor β (GI number 11275218) in *Naja*, and halysetin metalloproteinase (GI number 15076875), coagulation factor IX-binding protein (GI number 33638229), phospholipase A₂ (GI number 27151651) and L-amino acid oxidase (GI number 15887054) in *Agki*, have been observed to significantly increase the apparent masses. Obviously, modifications of venomous proteins are a common phenomenon in snake venoms. Of a number of mechanisms inducing chemical modification of protein, glycosylation is one of the most probable causes. Upon HMK amino acid sequence analysis by NetOglyc 3.0 and NetNGlyc 1.0 [22,23], one N-glycosylation site (112–115) is predicted, which just matches the observation of Ito et al. [24] that native HMK treated with peptide N-glycosidase resulted in the apparent MM decreasing. According to glycosylation prediction, all of the

proteins with increased apparent masses on 2DE contain at least one N-glycosylation consensus sequence. However, further biochemical experiments must be conducted to confirm the predictions. By searching the literature, L-amino acid oxidase [25], phospholipase A₂ [26], coagulation factor IX [27] and other snake venom proteins [28] have been reported as having glycosylation modifications in venom.

In conclusion, by combining four different proteomic approaches, more than 100 proteins or peptides have been identified in snake venoms of *Naja* and *Agki*. This comprehensive proteomic analysis reveals global expression of venomous proteins from *Naja* and *Agki*, defines the fundamentally differential proteomes from the two species and demonstrates a series of chemical modifications in the venomous proteins. These data certainly contribute to the current knowledge of venom proteins, and will benefit the studies on the mechanisms of snake envenomation.

We thank Haibing Yang, Haibo Zhang and Wei Leng for expert technical assistance. This work was supported by grant from the China National 973 Project (2001CB210501).

REFERENCES

- Kini, R. M. and Evans, H. J. (1990) Effects of snake venom proteins on blood platelets. *Toxicon* **28**, 1387–1422
- Warrell, D. A. (1986) Tropical snake bite: clinical studies in South-East Asia. In *Natural Toxins: Animal, Plant and Microbial*, 1st edn. (Harris, J. B., ed.), pp. 20–25, Oxford Science Publications, Oxford
- Barker, R. A., Ratcliffe, E., McLaughlin, M., Richards, A. and Dunnett, S. B. (2000) A role for complement in the rejection of porcine ventral mesencephalic xenografts in a rat model of Parkinson's disease. *J. Neurosci.* **20**, 3415–3424
- Zhao, X., Yeh, J. Z. and Narahashi, T. (2001) Post-stroke dementia: nootropic drug modulation of neuronal nicotinic acetylcholine receptors. *Ann. N.Y. Acad. Sci.* **939**, 179–186
- Marcinkiewicz, C., Weinreb, P. H., Calvete, J. J., Kisiel, D. G., Mousa, S. A., Tuszynski, G. P. and Lobb, R. R. (2003) Obtustatin: a potent selective inhibitor of $\alpha_1\beta_1$ integrin *in vitro* and angiogenesis *in vivo*. *Cancer Res.* **63**, 2020–2023
- Hati, R., Mitra, P., Sarker, S. and Bhattacharyya, K. K. (1999) Snake venom hemorrhagins. *Crit. Rev. Toxicol.* **29**, 1–19
- Yang, C. C. (1996) Structure and function of cobra neurotoxin. *Adv. Exp. Med. Biol.* **391**, 85–96
- Matsui, T., Fujimura, Y. and Titani, K. (2000) Snake venom proteases affecting hemostasis and thrombosis. *Biochim. Biophys. Acta* **1477**, 146–156
- Tibballs, J. (1998) The cardiovascular, coagulation and haematological effects of tiger snake (*Notechis scutatus*) venom. *Anaesth. Intensive Care* **26**, 529–535
- Hanash, S. (2003) Disease proteomics. *Nature (London)* **422**, 226–232
- Rioux, V., Gerbod, M. C., Bouet, F., Menez, A. and Galat, A. (1998) Divergent and common groups of proteins in glands of venomous snakes. *Electrophoresis* **19**, 788–796
- Pimenta, A. M., Stocklin, R., Favreau, P., Bougis, P. E. and Martin-Eauclaire, M. F. (2001) Moving pieces in a proteomic puzzle: mass fingerprinting of toxic fractions from the venom of *Tityus serrulatus* (Scorpiones, Buthidae). *Rapid Commun. Mass Spectrom.* **15**, 1562–1572
- Kalapothakis, E., Jardim, S., Magalhaes, A. C., Mendes, T. M., De Marco, L., Afonso, L. C. and Chavez-Olortegui, C. (2001) Screening of expression libraries using ELISA: identification of immunogenic proteins from *Tityus bahiensis* and *Tityus serrulatus* venom. *Toxicon* **39**, 679–685
- Fry, B. G., Wuster, W., Ramjanm, S. F., Jackson, T., Martelli, P. and Kini, R. M. (2003) Analysis of Colubroidea snake venoms by liquid chromatography with mass spectrometry: evolutionary and toxinological implications. *Rapid Commun. Mass Spectrom.* **17**, 2047–2062
- Nawarak, J., Sinchaikul, S., Wu, C. Y., Liao, M. Y., Phutrakul, S. and Chen, S. T. (2003) Proteomics of snake venoms from Elapidae and Viperidae families by multidimensional chromatographic methods. *Electrophoresis* **24**, 2838–2854
- Binz, P. A., Abdi, F., Affolter, M., Allard, L., Barblan, J., Bhardwaj, S., Bienvenut, W. V., Bulet, P., Burgess, J., Carrette, O. et al. (2003) Proteomics application exercise of the Swiss Proteomics Society: report of the SPS'02 session. *Proteomics* **3**, 1562–1566
- Liu, Y. B. and Xu, K. (1990) Lack of the blocking effect of cobrotoxin from *Naja naja atra* venom on neuromuscular transmission in isolated nerve muscle preparations from poisonous and non-poisonous snakes. *Toxicon* **28**, 1071–1076
- Tu, A. T. (1995) Overview of snake venom chemistry. *Adv. Exp. Med. Biol.* **391**, 37–62
- Master, R. W. and Rao, S. S. (1961) Identification of enzymes and toxins in venoms of Indian cobra and Russell's viper after starch gel electrophoresis. *J. Biol. Chem.* **236**, 1986–1990
- Phizicky, E., Bastiaens, P. I., Zhu, H., Snyder, M. and Fields, S. (2003) Protein analysis on a proteomic scale. *Nature (London)* **422**, 208–215
- Bjarnason, J. B. and Fox, J. W. (1995) Snake venom metalloendopeptidases: repolysins. *Methods Enzymol.* **248**, 345–368
- Hansen, J. E., Lund, O., Tolstrup, N., Gooley, A. A., Williams, K. L. and Brunak, S. (1998) NetOglyc: prediction of mucin type O-glycosylation sites based on sequence context and surface accessibility. *Glycoconj. J.* **15**, 115–130
- Blom, N., Sicheritz-Ponten, T., Gupta, R., Gammeltoft, S. and Brunak, S. (2004) Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics* **4**, 1633–1649
- Ito, M., Hamako, J., Sakurai, Y., Matsumoto, M., Fujimura, Y., Suzuki, M., Hashimoto, K., Titani, K. and Matsui, T. (2001) Complete amino acid sequence of kaouthiagin, a novel cobra venom metalloproteinase with two disintegrin-like sequences. *Biochemistry* **40**, 4503–4511
- Geyer, A., Fitzpatrick, T. B., Pawelek, P. D., Kitzing, K., Vrieling, A., Ghisla, S. and Macheroux, P. (2001) Structure and characterization of the glycan moiety of L-amino-acid oxidase from the Malayan pit viper *Calloselasma rhodostoma*. *Eur. J. Biochem.* **268**, 4044–4053
- Lai, C. C. and Her, G. R. (2002) Analysis of N-glycosylation of phospholipase A₂ from venom of individual bees by microbore high-performance liquid chromatography-electrospray mass spectrometry using an ion trap mass spectrometer. *J. Chromatogr. B Anal. Biomed. Life Sci.* **766**, 243–250
- Sinha, U. and Wolf, D. L. (1993) Carbohydrate residues modulate the activation of coagulation factor X. *J. Biol. Chem.* **268**, 3048–3051
- Nawarak, J., Phutrakul, S. and Chen, S. (2004) Analysis of lectin-bound glycoproteins in snake venom from the Elapidae and Viperidae families. *J. Proteome Res.* **3**, 383–392

Received 4 March 2004/20 July 2004; accepted 29 July 2004

Published as BJ Immediate Publication 29 July 2004, DOI 10.1042/BJ20040354