

Venoms proteomics: Novel approaches for drug discovery

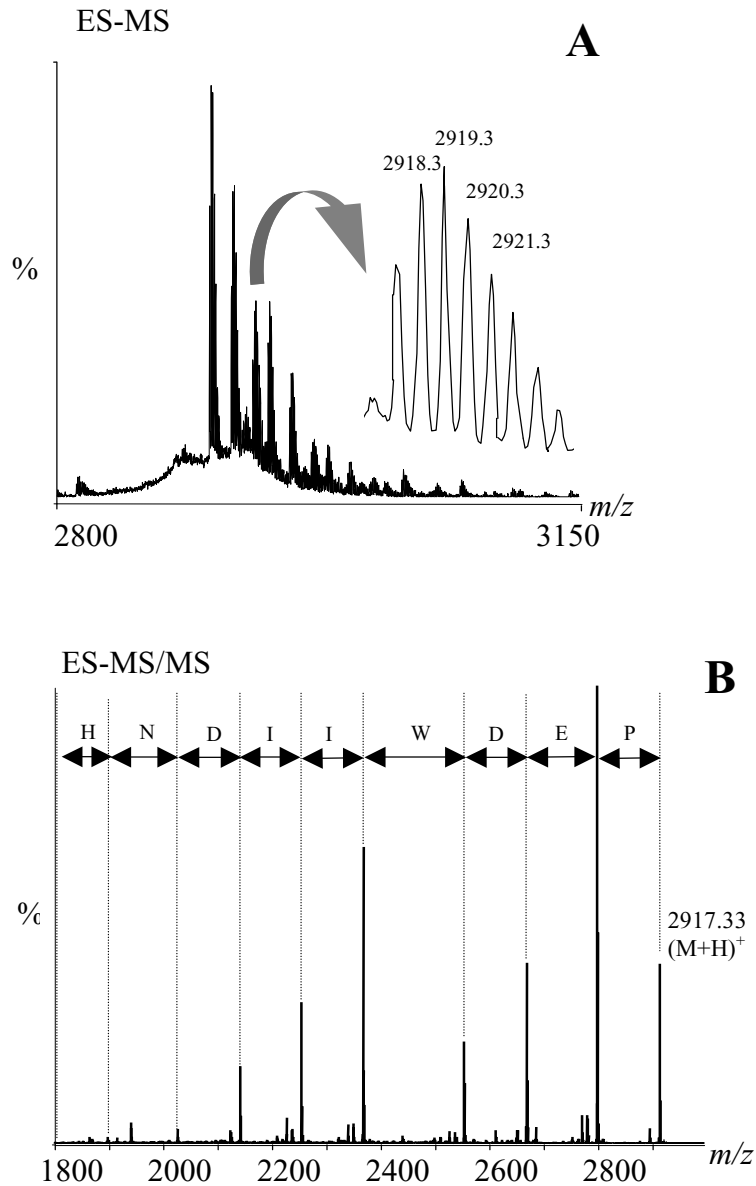
Philippe FAVREAU, Sophie MICHALET and Reto STÖCKLIN

*Atheris Laboratories, case postale 314, CH-1233 Bernex-Geneva, Switzerland
reto.stocklin@atheris.ch*

The discovery of naturally occurring bioactive compounds mainly relies on the observation of a particular biological activity. In a complementary approach, we focus on a backward strategy from structure to function based on a combination of genomics (sequencing of cDNA from venom glands), proteomics (systematic identification and characterization of the venom components) and biocomputing (development of a unique biochemical and biological database on venomous animals and their venoms).

Venoms are highly complex mixtures, usually made-up of more than a hundred different components, each with a specific biological activity. Some 1500 different peptidic toxins and enzymes of venomous origin have been isolated and characterized so far on the basis of their biological activity and usually with a time consuming approach. A combination of mass spectrometric techniques appears as a straightforward and efficient alternative method to characterize proteins from venoms.

The direct analysis of crude venoms using on-line LC-ES-MS and MALDI-TOF-MS has extensively been used in our laboratory. These two approaches permit to establish a specific mass map of the venom analyzed, that can be used like a "molecular fingerprint" representative of the species to which the venomous animal belongs, known as toxin mass maps or venom mass fingerprint. These sets of data can further be compared to molecular masses calculated from described toxins having their sequences listed in specialized databases such as Swiss-Prot (Bairoch and Apweiler 2000) or Venoms (Stöcklin and Cretton 2000) allowing easy targeting of potentially novel bioactive compounds in view to drug development. On-line LC-ES-MS was the first chemotaxonomic technique used to unambiguously classify venomous animals (cobra snakes of the genus *Naja*) by virtue of their venom mass fingerprints (Gillard *et al.* 1997; Stöcklin and Mebs 1995; Stöcklin *et al.* 2000). Recent investigation on *Tityus serrulatus* scorpion venom also revealed that the off-line MALDI-TOF-MS analysis of HPLC fractions can be a more powerful method for searching for low representative molecules in a highly complex mixture (Pimenta *et al.* 2001).



(A) High resolution mass spectrum of an HPLC fraction of crude *Atractaspis microlepidota microlepidota* snake venom. The inset presents an enlarged view of the isotopic envelope around 2920 m/z . **(B)** Part of the MS/MS spectrum obtained from of the triply-charged species at 973.11 m/z corresponding to a monoisotopic molecular mass of 2916.3 Da.

In a search for novel sarafotoxins (endothelins-like peptides), we have analyzed crude venom from the snake *Atractaspis microlepidota microlepidota*. Although more than 150 molecular masses were detected by on-line LC-ES-MS analysis of the crude venom, we were surprised not to find any mass in the expected sarafotoxin range between 2400 and 2600 Da. However, a series of some 20 masses around 2800-3000 Da drew our attention, and was thus submitted to *de novo* MS/MS amino acid sequencing (Figure). This analysis allowed C-terminal sequencing straight from LC-MS fractions, revealing a new family of 24 amino acid sarafotoxins with an additional C-terminal "Asp-Glu-Pro" pattern that follows the typical 21 amino acids consensus motif. These results confirmed those obtained in parallel by molecular biology through cDNA cloning of the venom gland of the same snake specimen, and allowed us to characterize five novel endothelin-type peptides (Ducancel *et al.* 1999). Biological activity and specificity are presently under investigation.

The flourishing of techniques related the field of mass spectrometry now provide a panel of methods facilitating protein identification and characterization, which can be successfully applied to the study of complex protein mixtures. The information such as venom's molecular mass fingerprint, *in vivo* biochemical processing and partial or complete sequence that can thus be obtained opens new perspectives in molecular toxinology and drug discovery.

We are grateful to Frédéric Ducancel and André Ménez (DSV/DIEP, CEA/Saclay, France); Badia Amekraz, Christophe Moulin and Philippe Pierrard (DEN/DPC/SPCA, CEA-Saclay, France); Amos Bairoch (Swiss Institute for Bioinformatics), Denis Hochstrasser (University Hospital), Robin Offord (University Medical Center) and Keith Rose (GeneProt Inc.) in Geneva; as well as Guillaume Cretton and Sylvie Stöcklin from our laboratories.

- Bairoch A, Apweiler R (2000) The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.*, **28**: 45-48.
- Ducancel F, Wery M, Hayashi MAF, Muller BH, Stöcklin R, Ménez A (1999) Les sarafotoxines de venins de serpent. In: Georges Cohen, ed. *Annales de l'Institut Pasteur / Actualités* : "Les venins" Institut Pasteur, Paris. **10** (2), 183-194.
- Gillard C, Virelizier H, Arpino P and Stöcklin R. (1997). Classification of the white *Naja* by on-line LC-ES-MS. In : *Eighteenth International Symposium on Capillary Chromatography* (ISSC Ed.), ISSC, Riva del Garda, Italy, **Vol. III** : 2192-2197.

- Pimenta AMC, Stöcklin R, Favreau P, Bougis PE, Martin-Eauclaire M-F (2001) Moving pieces in a proteomic puzzle: mass fingerprinting of toxic fractions from the venom of *Tityus serrulatus*. *Rapid Commun. Mass Spectrom.*, **15**:1562-1572.
- Stöcklin R, Mebs D (1995) Analysis and identification of snake venoms by mass spectrometry. In: Institut Pasteur, ed. *1st International Congress on Envenomations and their Treatments*. Paris: Institut Pasteur, 189 (abstract)
- Stöcklin R, Mebs D, Boulain JC, Panchaud PA, Virelizier H, Gillard-Factor C (2000) Identification of snake species by toxin mass fingerprinting of their venoms. *Methods Mol. Biol.*, **146**: 317-335.