

Editorial

‘Venomics’ or: The venomous systems genome project

Keywords : Bioinformatics; Database; Drug discovery; Envenomation; Evolution; Genetics; Genomics; Proteomics; Transcriptomics; Toxinology; Venom; Venom gland; Venomous animals; Venomous function

1. Introduction

The objective of this editorial is to inform the members of the International Society on Toxinology (IST) of developments and progress on an ambitious program that aims to characterize the unique genes and gene products of venomous animals. The project consists in charting the genomes of a few selected animals and, in parallel, in analyzing their peptide and protein venoms and the mRNAs that encode these in venomous glands. The project has been discussed on several occasions by the nearly 500 IST members, particularly during recent IST meetings. A few IST members have embarked on an initiative to set up a worldwide consortium for this program. Aims and objectives have been discussed, major worldwide genome centers have been visited and ideas have been exchanged with experts in the relevant fields. This editorial focuses on the present status and the short- and long-term scientific, strategic and economic issues that will emerge from the program.

2. Background

The animal kingdom includes more than 100,000 venomous species spread through major phyla such as chordates (reptiles, fishes, amphibians, mammals), echinoderms (starfishes, sea urchins), mollusks (cone snails, octopi), annelids (leeches), nemertines, arthropods (arachnids, insects, myriapods) and cnidarians (sea anemones, jellyfish, corals). Venomous animals typically possess venom-producing exocrine glands coupled to a delivery system (e.g. fangs, needles, harpoons; [Mebs, 2002](#)).

The ‘venome’ is the sum of all natural venomous substances produced in the animal kingdom. Each individual venom is a unique cocktail of often more than 100 different peptides and proteins, making the venome a source

of millions of peptides and proteins naturally tailored to act on a myriad of exogenous targets, such as ion channels, receptors and enzymes within cells and on the plasma membrane. Venoms provide animals with a variety of advantages, including an ability to subdue and digest prey efficiently and to defend themselves against predators. Also, venoms often include protease inhibitors and stabilizing agents that protect them from internal and external (high temperature) detrimental effects, and hence preserve them in the glands for weeks.

The venome is of interest to human beings for various reasons. First, it is a source of basic tools to study complex physiological systems, including the central and peripheral nervous systems, the cardiovascular system, blood coagulation, homeostasis, the hormonal system, the complement system and more generally the immune system ([Ménez, 2002](#)). Second, the venome is a source of drug leads, approved drugs and diagnostic tools, with at least five drugs derived from venom components that are already present on the market, and dozens of related active pharmaceutical products that are currently being tested in pre-clinical or clinical trials ([Ménez et al., 2005](#)). Third, knowledge of the venome is anticipated to favor the development of improved protection against envenomations.

To elucidate the genetic basis of venomous function we are exploring the possibility of charting the full genomic maps of a few selected venomous animals and of analyzing in parallel the transcriptomes of their respective venom glands and proteomes of their venoms. The project, called Venomics, is expected to open up new horizons, including an understanding of the function and evolution of venomous systems in various phyla, a clarification of their genetic relationship with the general biology of venomous animals, clues for understanding the diversity and evolution of venom peptides and proteins, new developments in the field of drug discovery and improvements in the treatment of envenomations.

3. State of the art

To our knowledge, the only ongoing genome project focusing on a venomous species is a study of the honey bee (*Apis mellifera*), which is of major interest for economic and public health reasons. The annotation and analysis of the honeybee genome has started. The Baylor College of Medicine Human Genome Sequencing Center, Houston, Texas, has launched the v.2.0 assembly of the *Apis mellifera* genome (7.5× genome coverage). The honeybee is certainly one of the most studied venomous animals and the elucidation of its genome will be of major interest in viewing its impact on nutrition (honey), health care (royal jelly) and medicine (allergies to stings). However, features related to the venom gland may be of limited impact due to the relative paucity of the venom, which includes only a small number of peptides such as melittin and apamin, one major phospholipase A₂ and a few more compounds without real diversity in terms of isoforms. Nevertheless, the results of this genome project will certainly contribute to our understanding of venomous systems in general.

4. What can we expect from the venomics project?

A key aspect of the project is that the venom and the venom glands from an animal can be physically isolated and studied independently by both proteomics and transcriptomics (transcript profiling). The resulting data will offer a powerful body of information to interpret genome sequencing data and in particular to determine the genes associated with the venomous system, to uncover their organization along a genome sequence, and hence to illuminate both fundamental and practical aspects associated with venomous function.

4.1. Functioning and evolutionary processes of venomous systems

At least four levels of complementary information are anticipated to emerge from the Venomics project.

First, as a large proportion of venom peptides and proteins adopt specific folds that are characterized by recognizable conserved cysteine patterns, in-depth genomic data-mining of their encoding genes should be facilitated. Usually, hypervariability in amino acid sequences occurs between the cysteines which lead to numerous peptide and protein isoforms. Such information should help to uncover the genomic organization of genes encoding venom peptides and proteins identified by transcriptomics and proteomics. This should also reveal how these genes are localized on chromosomes, how they may be duplicated and regulated, and how the associated introns are distributed.

Second, the cysteine patterns, and hence the associated folds that characterize the venom peptides and proteins, are also shared by non-toxic peptides, proteins or domains that

do not belong to the venome. Among various examples, this is the case of the CS α β fold which is shared by scorpion toxins and antimicrobial peptides present in the hemolymph of arthropods (Bontems et al., 1991), the SXC domain which is shared by sea anemone toxins and by other proteins from the nematode *C. elegans* (Blaxter, 1998), and the fold shared by snake venom sarafotoxins and endothelins (Mills et al., 1991). A genomic picture of the genes encoding structurally related endogenous homologs will thus emerge and should provide answers to a number of fundamental questions. What is the evolutionary relationship between these toxins and their non-toxin homologs? Is a gene encoding a non-toxin fold 'selected and expressed' at the level of the venom gland to become a toxin, or conversely have non-toxic peptides become toxins? Also, since not all possible disulfide-rich folds are exploited as toxin-folds by a venomous animal, is there any rationale for the toxin folds to have been selected? Answers to these questions will shed light on the biology and evolution of venomous systems. Also, identification of the functions of venom peptides and proteins and their homologs may help to determine the function of orphan proteins possessing structurally related domains in all organisms for which genome sequences are available, including the human genome.

Third, detailed proteomics and transcriptomics of the venom gland should reveal genes encoding proteins that have conventional cellular functions, and genes that are more specifically involved in the production of venom peptides and proteins. Thus, at least three biochemical and/or evolutionary aspects may be elucidated. Firstly, the genes encoding enzymes associated with post-translational modifications will be identified. These genes and the associated clusters will be located with respect to those encoding venom peptides and proteins, offering a genetic basis for the associated post-translational processes that occur in venomous glands. Secondly, the genes encoding venom peptides and proteins intriguingly undergo an abnormally high rate of mutations (Ohno et al., 1998) that might allow 'rapid' adaptations to changes in the availability of prey or expansion of the species that can be captured and consumed. Is there any 'biochemical machinery' responsible for this high rate of mutations (Conticello et al., 2001), or are the mutations simply the result of random or selected expression of 'sleeping genes' that can be induced by exogenous stress factors? Such putative 'machinery', would be quite original, and the Venomics project is expected to help clarify this remarkable system. Thirdly, it is generally believed that venom glands have evolved from other exocrine glands. For example, in the case of venomous snakes, it has been proposed that the venom glands derive from salivary glands. Comparison of the genes encoding proteins associated with the different types of exocrine glands of the venomous animal should help us understand their evolutionary relationship, if any.

Fourth, knowledge of the complete sequence of the genome of a venomous animal will provide specific

information on the animal itself (mollusk, reptile, arthropod, etc.). The information could set the stage for studies of the functions of various organs systems in the animals and even for examination of molecules that modulate their behaviors. This is of key importance in the project as it may illuminate how venom systems function and evolve with respect to the general biology of the animal, and hence to answer the key questions: are venomous systems genetically similar or different from one venomous animal to another and have they arisen through convergent or parallel evolution?

4.2. Practical issues of the Venomics project

The Venomics project offers unprecedented potential for drug discovery. It has been demonstrated that venom peptides and proteins constitute a unique source of drugs and drug leads for the treatment of pain, cancer, auto-immune diseases, allergies, hypertension, infectious diseases, neuronal diseases, as well as for the development of original diagnostic agents (Ménez et al., 2005). Remarkably, venom peptides and proteins are poorly immunogenic when injected in the absence of an adjuvant (Maillere et al., 1993), which reinforces their potential use as therapeutic drugs. Their high potency allows them to be used in minute amounts, so that production costs may not be a limiting factor. Another major advantage of toxins is their high specificity, which reduces the risk of adverse reactions. Furthermore, since they degrade into amino acids, there is reduced risk of metabolite toxicity.

The already-known *structural signatures* reflected in the conserved cysteine patterns of venom peptides and proteins will serve as guides, or at least as starting points, in identifying all proteins or genome domains that possess the same structural signature. We anticipate that there is a huge reservoir of bioactive components containing the same cysteine patterns. In addition, it is becoming clear that venom peptides and proteins tend to use simple *functional signatures* to recognize their physiological targets. These include, for example, the well-known 'RGD' motifs and the derived triplets that bind to integrins (Ruoslahti, 1996), work that has led to the development of anti-thrombotic drugs such as Integrilin[®], and the 'KY-dyad' which constitutes one of the functional signatures of toxins that bind to voltage-gated potassium channels (Dauplais et al., 1997). The combined use of these types of signatures to scan the genomes of venomous animals is expected to increase the number of possible candidates with a desired recognition function, which in turn may help in the search for homologs of potential therapeutic value in other genomes, including the human genome.

A genomics program to study venomous function cannot be considered without evaluating its impact on the treatment of human envenomations. It is believed that every year there are about 5 million snake bites that lead to some 100,000 deaths (Chippaux, 1998) and 100,000 scorpion stings that cause about 800 deaths. There are no epidemiological data on cone snail stings, spider bites or other envenomations,

but their potential to cause harmful and even lethal stings and bites is high. Current treatment of envenoming is based on immunotherapy, which is not entirely satisfactory. Genomic studies are expected to help generate new concepts, such as the identification of widespread generic structural patterns that may serve as a basis for the design of immunogens for antivenom production. It is also known that many venomous animals are resistant to their own venom, and sometimes this is due to the presence in their blood of potent neutralizing or inhibiting substances. An understanding of the genetic basis of such compounds may help us to develop new protective strategies against human envenomation.

Finally, a proper understanding of the genetic basis of venomous systems may serve as a lead for the design of artificial machinery to produce a range of highly post-translationally modified peptides and proteins.

5. How to proceed with the venomics project?

5.1. A consortium

The Venomous Systems Genome Project will need an Executive Committee of specialists to coordinate it, supported by an advisory board of worldwide experts. The Executive Committee will offer IST and other laboratories the opportunity to participate in the Venomics project and hence to constitute a consortium. Fundraising will be the consortium's primary objective. The idea of creating a non-profit foundation as a legal entity to deal with administrative issues and act as an interface between the IST and external partners is being considered. Information related to the project can be found on the following website: <http://www.toxinomics.org>.

5.2. Genomics

Various options are being considered regarding the sequencing of the genome of a venomous animal. One option consists in focusing on a single venomous animal and in completing its genome sequence (6–8× coverage). This will be a long and costly process considering the genome sizes of most venomous animals (see below). A second option under discussion is the concomitant elucidation of the genome sequences of several venomous animals, but at low coverages (1–2× coverage), while focusing on genome regions possessing genes encoding typical peptide and protein venom signatures. If successful, this approach should hasten identification of common and distinct genetic features associated with the function and evolution of venomous systems of phylogenetically unrelated animals. This option appears the most appropriate, but it is unclear whether comparison of phylogenetically distant animals will adequately meet our aims.

A key question is the choice of animal. The answer is not readily evident as each scientist has his or her own ideas on

the matter, usually conditioned by previous experience. Most probably it will be important to look at various practical considerations, such as the availability of the animal under international law, the complexity of the venom, the biology of the animal, the size of the genome, etc. Therefore, a case-by-case approach will be required when deciding which animal genomes should be sequenced. At present, several discussions are engaged to start the Venomics project with a cone snail. Funding agencies are contacted and applications are being submitted.

The genomes of snakes (36–42 chromosomes) range from 1.32 pg (almost equivalent to giga base pairs) in *Crotalus durissus terrificus* to 3.69 pg for the sea snake *Laticauda* sp. Cone snail genomes are slightly larger, averaging around 2.6, 3.9 and 3.6 pg for *Conus californicus*, *Conus lividus* and *Conus pennaceus*, respectively. *Agele-nopsis* spiders are of great biomedical interest, but their genomes tend to be large, ranging from 3.6 to 4.2 pg. In general these are rather large genomes, and one has to realize that assembly of each novel genome is likely to require 6 to 8-fold sequence coverage. Despite rapidly evolving DNA sequencing technologies, no more than a dozen genome centers around the world presently have the capacity to sequence completely a genome of such size. Moreover, in spite of automated high-throughput instrumentation, each project will require months of sequencing work. Data will be delivered in a standardized database format and made available to the scientific community.

5.3. Transcriptomics

An obvious task will be transcriptomic investigation of the venom gland of the same specimen used for the genomic work. A program of design and sequence analysis of cDNA (complementary DNA) and EST (Expressed Sequence Tags) libraries will be conducted either by the center in charge of the genomic work, or through collaborations with other institutes or IST laboratories. Representative clones of the main families of precursors pre-identified by the EST strategy will be sequenced, which will also allow the design of recombinant expression systems. The genes that are expressed in the venom gland should be quite diverse. Some are required for 'housekeeping' functions, while others allow the gland to make venom precursors, respond to signals that trigger venom ejection, and finally deliver the venom into prey. The development, maintenance, and regulation of such an elaborate gland must be highly regulated, and we expect to find nuclear factors that are involved in these processes among the genes expressed in venom glands.

5.4. Proteomics

A full map of the bioactive components is needed, namely of peptides and proteins that constitute the venom, and this will be achieved through biochemical analysis. Venom from the specimen undergoing genome sequencing

will be fractionated to detect most of its components, which will be identified by mass spectrometry using de novo sequencing and conventional proteomic techniques, together with bioinformatics for matching against the theoretical sequences expected from the genomic and transcriptomic sequences. A large number of peptides and proteins of interest—both those that are found in venom extracts or those that are predicted from mRNA or genomic sequences—will be selected for further investigation. Since most of these new compounds will only be available in minute quantities, if at all, a high-throughput synthesis program will have to be setup.

5.5. Bioassays

The project will be coupled to a wide campaign of bioactivity screening to maximize the discovery of novel bioactive compounds of biomedical potential. To this end, high-throughput assays of a wealth of targets of biomedical relevance will be organized to screen the venom fractions and synthetic compounds. Bioassays will include electrophysiology screening of multiple neuroreceptors, ligand-gated and voltage-gated ion channels, binding assays of G-coupled protein receptors, cell-based assays, enzymatic activity and enzyme inhibitors, anti-microbial, anti-parasitic and anti-viral assays. Positive hits will be evaluated and matched against existing knowledge for assessment of their potential usefulness.

5.6. Database

The Consortium will enter data sets generated by the genome project in a common general annotated database available to the scientific community through a specifically developed web-based platform. Genomic, transcriptomic, peptidomic and proteomic information will be cross-linked and submitted to automated and manual annotation by specialists in the field using appropriate biocomputing tools. Venom fractions and a library of synthetic peptides and proteins available for screening will be described at this site, together with the results obtained from bioassays. References and patents will be included as well, covering a broad bibliographical survey of the fields of interest. Specific data- and text-mining and other dedicated biocomputing tools will serve as a central core for a user-friendly interface. This central knowledge database will also be cross-linked to other public or private databases of interest.

5.7. Innovations

Exploration of the abundant sources of peptides and proteins identified by the Venomics project will require methodological innovations. In particular, classical 'series' studies of each discovered peptide should ideally be replaced by large-scale parallel technologies. In this respect, chemical and recombinant synthesis of peptides and

proteins should be developed, as well as methods for simultaneous analysis of several biological activities. For example, injections of mixtures of labeled synthetic peptides or proteins into animal models followed by investigation of the target tissues and identification of the bound compounds by mass spectrometry may provide rapid information on the targets of action of the injected peptides (Dive, 2002). Also, in order not to solve the 3D structure of each individual homolog by time-consuming structural methods, robust modeling methods need to be developed. In addition, it will be of interest to search for new functional signatures in order to use them in functional analyses of all available genomes, including the human genome.

6. Conclusions

The Venomous Systems Genome Project aims to complete the genomics, transcriptomics and proteomics of 5–6 model species, including one hymenoptera that is already going on, one cone snail that might start soon, one or two venomous snakes, one scorpion and one spider, over a period of 5–10 years. It is expected to offer to the scientific community the largest body of information ever compiled on venoms and their evolution and on the genetics of venomous systems. It will offer IST members a unique opportunity to position their work in this innovative and rewarding endeavor, which may help toxicology to become a major discipline in the life sciences.

Acknowledgements

We wish to thank Dr Michael Brownstein from the *J. Craig Venter Institute* for his helpful comments and advices.

References

- Blaxter, M., 1998. *Caenorhabditis elegans* is a nematode. *Science* 282, 2041–2046.
- Bontems, F., Roumestand, C., Gilquin, B., Ménez, A., Toma, F., 1991. Refined structure of charybdotoxin: common motifs in scorpion toxins and insect defensins. *Science* 254, 1521–1523.
- Chippaux, J.P., 1998. Snake bites: appraisal of the global situation. *Bull. World Health Org.* 76, 515–524.

- Conticello, S.G., Gilad, Y., Avidan, N., Ben-Asher, E., Levy, Z., Fainzilber, M., 2001. Mechanisms for evolving hypervariability: the case of conopeptides. *Mol. Biol. Evol.* 18, 120–131.
- Dauplais, M., Lecoq, A., Song, J., Cotton, J., Jamin, N., Gilquin, B., Roumestand, C., Vita, C., de Medeiros, C.L., Rowan, E.G., Harvey, A.L., Ménez, A., 1997. On the convergent evolution of animal toxins. Conservation of a diad of functional residues in potassium channel-blocking toxins with unrelated structures. *J. Biol. Chem.* 272, 4302–4309.
- Dive, V., 2002. Patent No. 02-06-698 (31 May 2002). Screening methods of biomolecule libraries and their applications.
- Maillère, B., Cotton, J., Mourier, G., Leonetti, M., Leroy, S., Ménez, A., 1993. Role of thiols in the presentation of a snake toxin to murine T cells. *J. Immunol.* 150, 5270–5280.
- Mebs, D., 2002. *Venomous and Poisonous Animals*. Medpharm, Stuttgart.
- Ménez, A., Servent, D., Gasparini, S., 2002. The sites by which animal toxins bind their targets involve two components: a clue for selectivity, evolution and design proteins. In: Ménez, A. (Ed.), *Perspectives in Molecular Toxicology*. Wiley, Chichester, pp. 175–202.
- Ménez, A., Gillet, D., Grishin, E., 2005. Toxins: threats and benefits. In: Gillet, D., Johannes, L. (Eds.), *Recent Research Developments on Toxins from Bacteria and Other Organisms*. Research Signpost, Trivandrum, India.
- Mills, R.G., Atkins, A.R., Harvey, T., Junius, F.K., Smith, R., King, G.F., 1991. Conformation of sarafotoxin-6b in aqueous solution determined by NMR spectroscopy and distance geometry. *FEBS Lett.* 282, 247–252.
- Ohno, M., Ménez, R., Ogawa, T., Danse, J.M., Shimohigashi, Y., Fromen, C., Ducancel, F., Zinn-Justin, S., Le Du, M.H., Boulain, J.C., Tamiya, T., Ménez, A., 1998. Molecular evolution of snake toxins: is the functional diversity of snake toxins associated with a mechanism of accelerated evolution? *Prog. Nucleic Acid Res. Mol. Biol.* 59, 307–364.
- Ruoslahti, E., 1996. RGD and other recognition sequences for integrins. *Annu. Rev. Cell Dev. Biol.* 12, 697–715.

André Ménez*

*Département d'Ingénierie et d'Etude des Protéines,
CEA-Saclay, DSV/DIEP, Bât. 151,
F-91191 Gif-sur-Yvette, France
E-mail address: andre.menez@cea.fr*

Reto Stöcklin

*Atheris Laboratories, Case postale 314,
CH-1233 Bernex-Geneva, Switzerland*

Dietrich Mebs

*Zentrum der Rechtsmedizin,
Kennedyallee 104, D-60596 Frankfurt/Main, Germany*

Available online 3 February 2006

* Corresponding author. Tel.: +33 1 69 08 26 03; fax: +33 1 69 08 90 71.