

Venomics: a new paradigm for natural products-based drug discovery

Irina Vetter · Jasmine L. Davis · Lachlan D. Rash · Raveendra Anangi · Mehdi Mobli · Paul F. Alewood · Richard J. Lewis · Glenn F. King

Received: 11 January 2010 / Accepted: 4 February 2010 / Published online: 24 February 2010
© Springer-Verlag 2010

Abstract The remarkable potency and pharmacological diversity of animal venoms has made them an increasingly valuable source of lead molecules for drug and insecticide discovery. Nevertheless, most of the chemical diversity encoded within these venoms remains uncharacterized, despite decades of research, in part because of the small quantities of venom available. However, recent advances in the miniaturization of bioassays and improvements in the sensitivity of mass spectrometry and NMR spectroscopy have allowed unprecedented access to the molecular diversity of animal venoms. Here, we discuss these technological developments in the context of establishing a high-throughput pipeline for venoms-based drug discovery.

Keywords Drug discovery · Animal venoms · Venomics · Venom peptides · Natural products

Introduction

The drug discovery pipeline is critically dependent on access to diverse molecular libraries capable of providing high-quality leads. A poor choice of libraries at the discovery phase leads to wasted resources and no drugs. Therefore, it is surprising that modern drug discovery approaches have narrowly focussed on libraries of small molecules that attempt to adhere to Lipinski's rules for drug-like properties and which can be synthesized by combinatorial chemistry and screened in a high-throughput

manner (Koehn and Carter 2005). These libraries often lack structural diversity, despite the large numbers of molecules present, as they are often centered around a limited number of “privileged” molecular scaffolds that have previously been flagged as interesting and which are amenable to combinatorial chemistry (Macarron 2006).

The complex molecular scaffolds found in natural products constitute a vast chemical diversity that is unmatched by synthetic molecules (Clardy and Walsh 2004; Koehn and Carter 2005). Indeed, one can consider the chemical space available from natural sources as the culmination of a billion-year drug discovery program with unlimited resources. The value of natural products as sources of new drugs is highlighted by the fact that ~50% of all drugs in clinical use are of natural product origin (Paterson and Anderson 2005). The limited success of the combinatorial chemistry approach has, therefore, rekindled interest in natural products as a component of modern drug discovery efforts (Butler 2004; Ortholand and Ganesan 2004; Koehn and Carter 2005; Paterson and Anderson 2005).

Animal venoms are a rich source of natural compounds that have evolved high affinity and selectivity for a diverse range of biological targets, especially membrane proteins such as ion channels, receptors, and transporters (Lewis and Garcia 2003; Tedford et al. 2004; Fry et al. 2009). Venomics has therefore emerged as an important addition to modern drug discovery efforts (Escoubas et al. 2008; Escoubas and King 2009). In particular, the high potency and specificity of many venom-derived peptides, their ease of chemical synthesis and/or recombinant production, and the resistance of many disulfide-rich peptides to proteolytic degradation, are attributes that have made them attractive drug leads (Harvey 1995; Lewis and Garcia 2003; Olivera 2006). As early as the 1970s, the antihypertensive drug captopril was developed from lead peptides isolated from

I. Vetter · J. L. Davis · L. D. Rash · R. Anangi · M. Mobli · P. F. Alewood · R. J. Lewis · G. F. King (✉)
Institute for Molecular Bioscience, The University of Queensland, St. Lucia, QLD 4072, Australia
e-mail: glenn.king@imb.uq.edu.au

the Brazilian viper *Bothrops jararaca* (Cushman and Ondetti 1991). More recently, a number of novel drugs have emerged from venom-based drug discovery efforts. ω -conotoxin MVIIA (Ziconotide, Prialt®), a 25-residue peptide from the venom of the aquatic cone snail *Conus magus*, was approved by the FDA in 2004 as an analgesic for the treatment of chronic pain (Miljanich 2004), while exenatide (Byetta®), a 39-residue glucagon-like peptide-1 agonist derived from the saliva of the Gila monster (*Heloderma suspectum*), was approved in 2005 for the treatment of type 2 diabetes (Malone et al. 2009). Many other venom-derived peptides are in various stages of pre-clinical or clinical development (Table 1).

Despite their promise for drug discovery, most of the chemical diversity encoded within animal venoms remains uncharacterized. However, recent technological advances that facilitate high-throughput screening (HTS) and structural characterization of venoms and venom peptides promise to accelerate the venoms-based drug discovery pipeline. In the following sections, we discuss the various elements of this pipeline, which is outlined in Fig. 1. We review recent technological developments that promise to expedite venoms-based drug discovery, and provide recommendations and caveats based on the experience gained from our own drug discovery efforts.

High-throughput assays

The systematic isolation and characterization of bioactives from the mixture of peptides, proteins, small molecules,

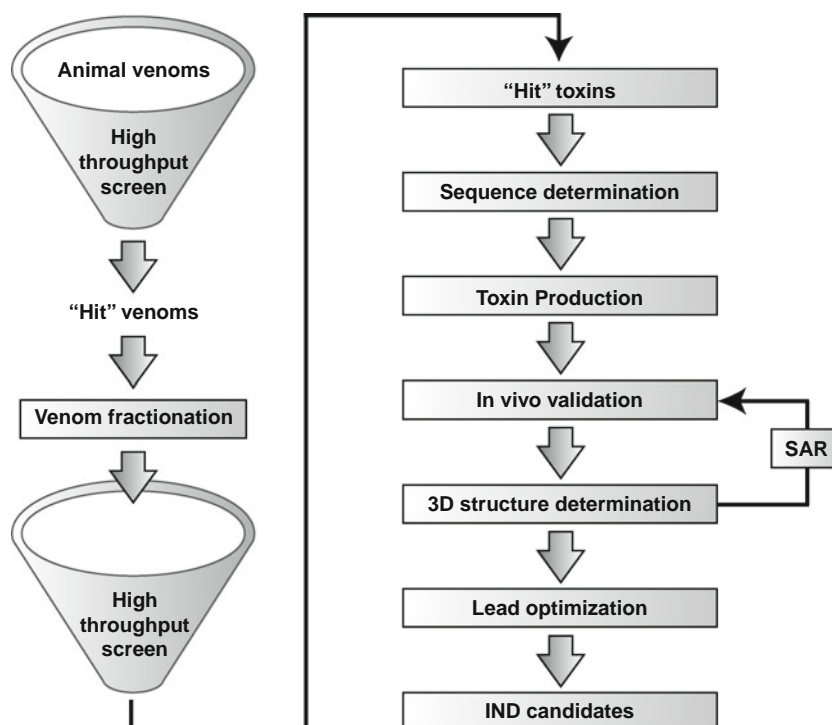
and salts present in venoms is often referred to as activity-guided fractionation. This process relies critically on the selectivity, specificity and capacity of high-throughput assays. The basic requirements of HTS include high-assay sensitivity and accuracy as well as high-assay robustness and reproducibility (e.g., high Z-factor, as defined by Zhang et al. (1999)). This is particularly important with respect to the screening of crude venoms or partially purified venom fractions. In contrast to combinatorial chemical libraries, which often comprise many thousands if not millions of compounds, natural product libraries frequently consist of crude or partially purified mixtures of compounds with diverse biological effects. This reduces the resource requirements for screening, but at the same time introduces the potential for interference from non-target-specific effects from other components of the venom. Thus, while the traditional goal for HTS in the pharmaceutical industry has been to increase screening capacity by increasingly automating and miniaturizing these assays, HTS in the context of venomomics arguably requires greater emphasis on data quality.

With the continuing definition of new potential drug targets, the repertoire of assays amenable to HTS has increased significantly in recent years. These include more traditional assays, such as electrophysiology, absorbance/fluorescence-based assays, radioligand binding, and ELISAs, as well as more recent developments, such as AlphaScreen and label-free technologies, and assays based on bioluminescence, fluorescence polarization, fluorescence-resonance energy transfer (FRET), bioluminescence resonance energy transfer, and scintillation proximity assays. In the following

Table 1 Examples of venom peptides or derivatives in clinical use or under development

Animal	Peptide/protein	Pharmacology	Indication	Stage
Snake	Captopril (Teprotide)	Inhibitor of angiotensin converting enzyme (ACE)	Hypertension	FDA approved (1981)
Snake	Eptifibatide (Integrilin™)	Inhibits fibrinogen binding to platelet glycoprotein IIb/IIIa receptor	Unstable angina	FDA approved (1998)
Cone snail	ω -Conotoxin MVIIA (Ziconotide, Prialt®)	Blocks Ca _v 2.2 voltage-gated calcium channel	Chronic pain	FDA approved (2004)
Lizard	Exenatide (Byetta®)	Insulin secretagogue (incretin mimetic)	Type 2 diabetes mellitus	FDA approved (2005)
Snake	Ancrod	Anti-coagulant	Ischemic stroke	Phase III
Cone snail	χ -Conotoxin MrIA (Xen2174)	Inhibits noradrenalin transporter	Chronic pain	Phase II
Scorpion	Chlorotoxin (TM-601)	Binds MMP2 on surface of glioma cells, impairing invasion ability	Glioma	Phase II
Spider	Psalmotoxin-1	Blocks acid-sensing-ion channel 1a (ASIC1a)	Inflammatory pain	Pre-clinical
Spider	Tx2-6	Nitric oxide release	Erectile dysfunction	Pre-clinical
Sea anemone	ShK	Blocks K _v 1.3 voltage-gated potassium channel	Autoimmune diseases including MS	Pre-clinical

Fig. 1 The key elements of a venoms-based drug discovery program. A robust high-throughput screen is essential to rapidly identify venoms with desired activity and to allow subsequent isolation of bioactive molecules. An efficient toxin production system is essential not only to produce sufficient toxin for functional and structural characterization, but also to facilitate structure–activity relationship (SAR) studies. Structural characterization is no longer a major bottleneck in the discovery pipeline due to recently developed high-throughput NMR methods



sections, we discuss the advantages and disadvantages of some of the most commonly used screening techniques in the context of a venoms-based drug discovery program.

Electrophysiology

Although patch-clamp or voltage-clamp electrophysiology is generally considered the gold-standard assay for assessing the functional activity of ion channels, it requires a high level of technical expertise and is generally used in low-throughput format. Recent advances include the development of automated platforms for single-cell electrophysiology studies, thus increasing throughput and simplifying assays (Bennett and Guthrie 2003). Commercially available automated high-throughput electrophysiology platforms such as the IonWorks Quattro (Molecular Devices), PatchXpress (Axon Instruments), and Flyscreen 8500 (Flyion GmbH) have enabled the development of high-throughput assays for a variety of ion channels of therapeutic interest (Hamelin et al. 2005; Dunlop et al. 2008). However, despite the advances in HTS offered by these platforms, problems relating to throughput and reproducibility still remain. These are partly due to inherent difficulties associated with using cells, such as the need for robust, high-quality cells with stable expression of the ion channel of interest and ideal patch-clamping characteristics (such as membrane seal and stability) (Dunlop et al. 2008).

AlphaScreen technology

Amplified Luminescent Proximity Homogeneous Assay or AlphaScreen technology is based on the principle of luminescent oxygen channeling (Ullman et al. 1994; Ullman et al. 1996). AlphaScreen assays were reported to have superior sensitivity and dynamic range when compared with time-resolved fluorescence and time-resolved FRET in the context of a nuclear receptor assay (Glickman et al. 2002). The ability to generate AlphaScreen beads specific to virtually any target of interest, second messenger or protein, has seen this technology applied to the detection of second messengers including cAMP and IP1, kinase activity, and receptor–ligand interactions (Taouji et al. 2009). These characteristics make AlphaScreen technology applicable to the HTS of venoms and venom components.

Label-free systems

Label-free systems such as the XCELLigence (Roche), BIND (SRU Biosystems), OWLS 210 (MicroVacuum), RAPid4 (TPP Labtech), and EPIC (Corning) platforms measure impedance, acoustic and optical resonance, or refractive index. While generally used for biochemical assays, label-free systems are increasingly being adapted to cell-based assays, where activation of ion channels and receptors, proliferation or apoptosis, infection with viruses, and other cellular events lead to microstructural changes in

the cell that are detected by these platforms. Label-free systems can avoid interference associated with the use of labels, and they enable detection of allosteric modulators with novel mechanisms of action. However, these systems may not be sensitive enough to detect small changes in cellular function, and they are potentially susceptible to interference from other venom components, yielding a higher number of false positives and negatives. Intuitively, this approach might be more suitable to more purified venom libraries to minimize such interference, although much more work is required to determine the place of label-free systems in HTS of venoms.

Radioligand binding

Radioligand binding assays have often been the method of choice for HTS of venoms as they are cheap and easily automated (Denyer et al. 1998). Radioligand binding assays rely on the competition between an unlabeled venom component and a radiolabeled ligand for the same or similar binding site on a target of interest. Consequently, radioligand binding assays are ill-suited to detection of allosteric modulators that do not share the binding site of traditional ligands (Noël et al. 2001). This is a significant caveat for the screening of voltage-gated ion channels, because multiple ligand binding sites have been described for many channels and consequently radioligand binding assays would be expected to yield a particularly high number of false negatives.

Fluorescence-based assays

Fluorescence-based assays are particularly amenable to HTS as they are robust and easily set up. Many platforms are available for fluorescence-based HTS, which usually differ in their ability to detect varying wavelengths, as well as their compound addition mechanics. The FLIPR^{TETRA} (Molecular Devices), Cell Lux (Perkin Elmer), and FDSS 6000 (Hamamatsu) in particular can be considered true high-throughput platforms as they are capable of measuring fluorescence from 96 wells simultaneously.

In recent years, a plethora of fluorescent dyes has become available for measurement of intracellular calcium (e.g. Fura-2, Fluo-3, Fluo4, and Calcium Green), sodium (e.g. SBFI, CoroNa Green, Sodium Green, and CoroNa Red), potassium (e.g. PBFI) and chloride (e.g. SPQ) ions. FRET assays based on voltage-sensitive dyes can also be used to monitor changes in membrane potential (Xu et al. 2001). These assays employ oxonol dyes as FRET acceptors, and coumarin-tagged phospholipids in the outer leaflet of the cell membrane as FRET donors (Zheng et al. 2004). The fluorescent properties of these dyes are altered by the binding of their cognate ions, or a change in membrane

potential, and thus they can be used to detect changes in ion concentration within the cell, or a voltage change across the cell membrane.

Calcium-sensitive fluorescent dyes generally give the most robust performance due to the large Ca^{2+} gradient across cells, and therefore are commonly used for HTS (Zheng et al. 2004). The ubiquitous nature of the Ca^{2+} signal can be utilized for the identification and isolation of venom components active at many targets, including voltage-gated and ligand-gated-ion channels (e.g. TRPV1, nicotinic acetylcholine receptors (nAChRs), and voltage-gated calcium channels) and G protein-coupled receptors (GPCRs) (Kitaguchi and Swartz 2005; Rivers et al. 2005). Although cell lines overexpressing targets of interest are typically used for HTS, it is also possible to exploit endogenously expressed ion channels or GPCRs for functional assessment of Ca^{2+} responses (Vetter and Lewis 2009). For example, we recently identified a novel α -conotoxin by screening fractionated crude venom for inhibition of endogenous $\alpha 7$ nAChR responses in human neuroblastoma cells (Fig. 2). Owing to the ubiquitous nature of the Ca^{2+} signal, fluorescent Ca^{2+} imaging also allows the isolation and identification of venom components with unknown targets or modes of action. Indeed, many venoms elicit increases in intracellular Ca^{2+} , and this approach has been used to study venom or venom components from the ectoparasitoid wasp *Nasonia vitripennis* (Rivers et al. 2005), the Portuguese Man-of-War *Physalia*

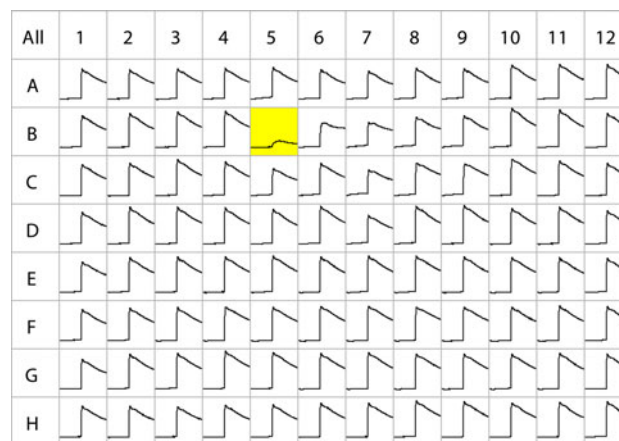


Fig. 2 Identification of a novel α -conotoxin from fractionated crude venom using a fluorescence-based HTS. SH-SY5Y human neuroblastoma cells were loaded with the calcium-sensitive fluorescent dye Fluo-4 and calcium responses were monitored for 550 s using a FLIPR^{TETRA}. After addition of crude venom fractions, endogenously expressed $\alpha 7$ nAChRs were stimulated with choline (30 μM) in the presence of the allosteric modulator PNU120596. Wells H1–H12 are control responses. Fraction 17 (well B5, highlighted) was identified as containing an $\alpha 7$ nAChR inhibitor. Further sub-fractionation and screening led to the isolation of a novel α -conotoxin

physalis (Edwards and Hessinger 2000), the duck-billed platypus *Ornithorhynchus anatinus* (Kita et al. 2009), and the Trinidad chevron tarantula *Psalmopoeus cambridgei* (Siemens et al. 2006).

In summary, successful screening requires a well-defined assay format that can readily and reliably identify actives (hits) preferably in crude venoms or alternatively in fractionated venoms. Crude venom “hits” need to be confirmed with fractionated venoms early in the isolation process as, depending on the venom and target, a high attrition rate can be encountered at this step. Recovery of activity also needs to be quantified at each step, since recovery of activity can be poor for some venom classes, especially those that elute either near the void volume or very late in reversed-phase HPLC (rpHPLC).

Isolation and sequencing of active toxins

In the context of drug discovery, the isolation and characterization of venom components is carried out for two main reasons: (1) to identify and characterize the component(s) responsible for the activity observed in a bioassay; or (2) to survey the complete structural diversity of a venom with the aim of discovering new sequences with novel structural scaffolds and pharmacologies, which can then be made in sufficient quantities and screened for activity in a panel of bioassays.

Although easily defined, these tasks are not trivial due mainly to the vast complexity of animal venoms. Snake venoms are known to contain >100 unique components (Calvete et al. 2007) while venoms from scorpions, spiders, and cone snails have recently been shown to contain anywhere from 300 to over 1,000 unique molecular entities (Pimenta et al. 2001; Escoubas et al. 2006; Newton et al. 2007; Biass et al. 2009; Davis et al. 2009). It is this amazing complexity combined often with small amounts of material that poses the biggest challenge to venomomics. These challenges are further confounded by the paucity of genomic data and therefore the lack of theoretical protein databases that are instrumental for protein identification using classical proteomic approaches (Escoubas et al. 2008; Escoubas and King 2009).

The following sections discuss approaches taken to achieve the two goals outlined above, including recent technological developments that are having the greatest impact on this field.

Assay-guided fractionation

In an academic laboratory setting, once an active venom or fraction has been identified via bioassay, and given the amount of material available is sufficient, assay-guided

isolation and characterization of the active toxin is relatively straightforward. The tandem use of orthogonal separation techniques, such as rpHPLC to separate by hydrophobicity followed by ion exchange HPLC to separate by charge, is an efficient approach for obtaining pure peptides for amino acid sequence analysis, usually by Edman degradation. These processes, although effective, are time and resource consuming. If one is following up on only a few “hits” this may not be an issue, however, with the progress of HTS comes the scenario where many active fractions need to be purified and characterized to determine the most suitable leads for further study. The biggest challenge in this respect is rapid, reliable, and cost-effective peptide sequencing.

Whole-venom analysis: how many peptides?

As opposed to activity-guided fractionation, in which one is trying to isolate a limited number of components, whole-venom analysis is much a much more challenging prospect as it aims to determine the full structural diversity present in a given venom (i.e. the total number of components and their full, or at least partial, sequences). This task can be approached from two different, yet complementary, directions: proteomic/peptidomic and transcriptomic (genomic). In theory, both approaches should provide a similar, and hopefully complete, picture of a venom. The proteomic approach is based on the final product (i.e. the proteins and peptides secreted into the venom), while the transcriptomic approach is based on the genomic recipe book of potential toxins.

The first step in obtaining a holistic view of a venom is to determine exactly how complex it is, in other words, how many unique components it contains. In this respect, advances in the sensitivity, resolution, and accuracy of mass spectrometry have been the major contributors to our improved ability to determine venom complexity (Escoubas et al. 2008). The advent of soft ionization processes, including electrospray ionization (ESI) and matrix-assisted laser desorption ionization (MALDI) has allowed direct observation of molecules in crude venoms. MALDI-TOF MS is generally regarded as superior to ESI-MS for analyzing complex samples, whereas ESI, being liquid based, is more amenable to interfacing with online separation techniques such as rpHPLC. Despite this, MALDI-TOF MS of crude venoms typically results in relatively low mass counts due to ion suppression effects that prevent ionization of all molecules, particularly minor components (Escoubas et al. 2008). This problem can be largely obviated by combining one or more separation techniques with MS analysis either online (LC-ESI-MS) or offline (rpHPLC prior to MALDI-TOF MS). Naturally, this leads to the question as to which technique is best suited for the

task. Startling results have been achieved using both methods as highlighted by two recent studies. Escoubas et al. (2006) used offline rpHPLC combined with MALDI-TOF MS to reveal more than 600 peptide masses in the venom of male Sydney funnel-web spider *Atrax robustus*, and over 1,000 in the venom of the female Blue Mountains funnel-web spider *Hadronyche versuta*. Davis et al. (2009) used online rpHPLC ESI-MS to demonstrate that a single specimen of the cone snail *Conus textile* can contain as many as 1,117 unique masses. From a study of *Conus consors* venom that directly compared both analytical methods the authors concluded that the techniques are in fact complementary, with only 21% of the masses being common to both data sets (Biass et al. 2009).

Liquid chromatography is not the only way to achieve separation of complex mixtures, with more traditional proteomic approaches also having been used to successfully map venom landscapes. For example, Liang's group used size exclusion chromatography of crude venom followed by either two-dimensional (2D) gel electrophoresis, tryptic digestion, and MS/MS (for components >10 kDa) or ion exchange then rpHPLC and Edman degradation (for masses <10 kDa) to characterize the venom of two local tarantulas, *Haplopelma huwenum* (Yuan et al. 2007) and *Chilobrachys jingzhao* (Liao et al. 2007). Both venoms were found to contain >90 proteins (47 identified by MS/MS) and around 100–120 peptides with around half being fully or partially sequenced. A very similar picture (>80 proteins and 100–200 peptides) was obtained by applying a similar approach to the venom of Brazilian armed spiders (*Phoneutria* spp.) (Richardson et al. 2006).

Traditional gel-based proteomics approaches have been used mostly for snake venoms, which is not surprising given they are dominated by proteins larger than 10 kDa. Calvete et al. demonstrated the power of an initial rpHPLC step to resolve and recover most, if not all, venom components in the mass range usually reserved for gel electrophoresis (i.e. 7–150 kDa) as well as small peptides (0.4–7 kDa) (reviewed in Calvete et al. 2007, 2009). The use of rpHPLC with UV detection instead of gels has the added advantages of leaving the isolated components in solution and avoiding the requirement for de-staining and in-gel reactions and extractions, which are time consuming and can result in sample loss.

Determining amino acid sequences

Determining the complexity of venoms via mass counts provides only limited information in regards to biological activity. Masses can provide hints as to which toxin class a peptide might belong, but in order to more fully appreciate and exploit the diversity of chemical space present in animal venoms, we need to know the primary and tertiary

structure of a peptide and how this relates to its biological activity. Determination of the amino acid sequence of a venom peptide also enables it to be produced via chemical synthesis or recombinant methods (discussed in the next section), thus providing material for further characterization without overexploiting the natural source.

For the past 50 years, peptide sequencing has been performed predominantly by Edman degradation. Although reliable, this method is time consuming and expensive. Furthermore, Edman degradation is not possible when peptides are N-terminally blocked; for example, according to the ArachnoServer database [www.arachnoserver.org] (Wood et al. 2009) ~1.5% of all known spider toxins are N-terminally blocked with pyroglutamate. The development and refinement of tandem MS methods (MS/MS) has allowed the rapid and efficient identification of proteins via the determination of short peptide sequences which are then matched to protein databases. MS/MS requires very little material and mixtures of peptides can be sequenced rapidly. In addition, it enables detection of post-translational modifications (PTMs), which are typically not identified in Edman or genomic analyses. Confirmation of the presence of PTMs is important as they can contribute to the biological activity of the peptide and/or its resistance to enzymatic degradation (Jakubowski et al. 2006).

Given the vast array of biologically active peptides present in animal venoms and the lack of genomic information from which to build theoretical protein databases, it is easy to see why de novo sequencing by MS is one of the most important recent developments for venomomics studies. The theory and limitations of de novo sequencing by MS/MS and its application to venomomics have been thoroughly reviewed recently (Escoubas et al. 2008; Seidler et al. 2009) and hence the following will be a brief overview of the developments most applicable to the rapid determination of novel peptide sequences from venoms.

The most common MS/MS fragmentation technique, collision-induced dissociation (CID), involves the collision of an ionized peptide (preferably multiply charged, which makes this technique most applicable to ESI-MS, as MALDI-TOF MS results primarily in singly charged species) with an inert gas and subsequent fragmentation along the peptide backbone. The resulting mass spectra can be complex, containing multiple ion types (*b*, *y*, *a*, often $-\text{NH}_3$ or $-\text{H}_2\text{O}$, etc.), making them difficult to interpret, particularly if the peptide contains labile groups such as PTMs (Kjeldsen et al. 2007). Thus, for rapid, reliable and high-throughput sequence determination, alternative fragmentation techniques that result in simpler spectra but with comprehensive sequence coverage are required. Two recently reported techniques, electron transfer dissociation (ETD) and MALDI in-source decay (ISD), appear to hold the most promise for venomomics due to their speed, ease of

data interpretation, and their direct accessibility without the need for specialized or expensive high-end MS instruments (such as FT-ICR MS).

MALDI-ISD is the process whereby a peptide fragments in the time between ionization and ion extraction and typically results in *c* and *z* ions (fragmentation at the N–C α bond), as opposed to *b* and *y* ions (fragmentation at the peptide bond) seen in CID MS/MS. MALDI-ISD is facilitated by the use of 1,5-diaminonaphthelene (1,5-DAN) as the matrix (Fukuyama et al. 2006). Quinton et al. (2007) demonstrated the usefulness of this approach for characterizing disulfide-rich peptides from cone snail venom. They found that 1,5-DAN has two characteristics that greatly aid the analysis of intact (i.e. fully oxidized) venom peptides, thereby eliminating the need for reduction and alkylation: first, the resultant spectra are dominated by *c* ions which greatly facilitates analysis, and second, the number of disulfide bonds can be readily determined due to their in-source partial reduction. Another major advantage of this approach is that it is achieved in a single MS step and does not require tandem MS instruments. In our own hands, this approach enables generation of ten unambiguous sequence tags of 10–20 residues in approximately 30 min. The major limitations of MALDI-ISD are the requirement for pure peptides (>90%) and the lack of coverage at the N- and C-termini. However, it is an ideal method for rapid generation of sequence tags that can be used to design primers for extraction of complete transcript sequences from a venom-gland cDNA library (discussed below).

ETD is a fragmentation method that evolved from electron-capture dissociation (ECD), but has the advantage that it can be performed on a more accessible mass spectrometer (e.g. linear 2D quadrupole ion trap, as opposed to FT-ICR-MS) and produces high-quality spectra from a single scan, making it amenable to online MS/MS. ETD also results in the generation of *c* and *z* ions, and it is particularly useful for venomomics investigations as it can produce fragmentation across the entire peptide backbone and is well suited to sequencing peptides with PTMs (Seidler et al. 2009). The power of ETD for venomomics was highlighted in a recent study where 31 complete peptide sequences were obtained using just 7% of the venom dissected from a single cone snail, which represented one-third of all masses detected (Ueberheide et al. 2009).

Current de novo peptide sequencing by MS/MS can be time consuming and requires considerable operator expertise to generate and analyse fragmentation spectra. However, advances in fragmentation methods such as MALDI-ISD and ETD (which are complementary to the widespread ESI-CID method) and improved bioinformatic tools for automated interpretation of the spectra are beginning to allow unprecedented, rapid access to the full landscape of venom peptides.

Transcriptomics

An alternative, yet complementary, approach to obtaining venom-peptide sequences is via the use of venom-gland cDNA libraries. This involves either global profiling of the venom via random sequencing of expressed sequence tags (ESTs) or fishing for transcript(s) encoding a specific toxin for which partial amino acid sequence information has already been obtained by Edman degradation and/or MS methods.

ESTs are short DNA sequences (~200–500 bp) generated by sequencing the ends of expressed genes (i.e. transcripts in a cDNA library). Given that most animal-venom peptides are <100 residues long, the sequencing of ESTs often yields the complete DNA sequence of a peptide-toxin precursor. The EST approach has been successfully applied to venom from scorpions (Schwartz et al. 2007; D'Suze et al. 2009), spiders (Kozlov et al. 2005; Jiang et al. 2008), and snakes (Francischetti et al. 2004). Typically, several thousand clones are sequenced, resulting in 100–1,000 s of cDNA sequences. This method has already resulted in the identification of many novel peptide-toxin sequences and, in combination with the powerful new de novo sequencing tools described above, it marks a new era in our understanding of the chemical and functional diversity of animal venoms.

In the context of a venoms-based drug discovery program, however, one is typically only concerned with obtaining the complete sequence of toxins (and potentially paralogs thereof) that were found to be hits in a HTS. It would be expensive and far from guaranteed to try to obtain the desired sequence from random sequencing of ESTs, especially if the toxin arises from a poorly expressed transcript. Instead, it is safer and cheaper to sequence the desired transcript using RACE (rapid amplification of cDNA ends) technology (Frohman 1993). This approach relies on first obtaining a short sequence tag from the mature toxin using Edman degradation and/or MS. Using a gene-specific primer designed from this sequence tag, 5' RACE is then employed to obtain upstream sequence information, which includes the 3' untranslated region (UTR), the signal sequence, and very often a propeptide sequence; 3' RACE is then used in conjunction with a primer derived from the signal sequence or 3' UTR to obtain downstream sequence information, including the complete sequence of the mature toxin sequence and 5' UTR. It has been shown that the signal sequence is highly conserved (i.e. under strong negative selection) for families of paralogous toxins (Sollod et al. 2005) and hence the 3' RACE will typically yield not just the desired toxin sequence but also the sequence of several paralogs, the activities of which can often provide useful structure–activity relationship (SAR) information.

Production of recombinant and synthetic toxins

Perhaps, the most significant bottleneck with regard to realizing the full potential of animal venoms in the context of a drug discovery pipeline is the rapid production of sufficient quantities of venom peptides for complete structural, functional, and in vivo characterization. Only rarely is sufficient native material available for such studies. The alternative is to produce these peptides by recombinant methods or solid-phase peptide synthesis (SPPS).

Production of venom peptides presents a significant technical challenge due to the large number of structurally important disulfide bonds. Regardless of the method of production, the problem of producing the native disulfide-bond architecture must be addressed as a toxin with three, four, or five disulfide bonds is theoretically capable of forming 15, 105, or 945 different disulfide-bond isomers, respectively. Coupling SPPS with native chemical ligation (NCL) turns this problem on its head by taking advantage of the high cysteine content of peptide toxins to increase the efficiency of chemical synthesis. NCL involves chemoselective reaction of an unprotected peptide- α -thioester with the N-terminal Cys residue of a second peptide fragment, giving rise to a thioester-linked intermediate that spontaneously rearranges to form a native amide bond at the ligation site (Dawson et al. 1994). The rationale behind NCL is that the yield of synthetic peptide can be significantly improved by dividing the peptide into two or more easily synthesized fragments that can subsequently be joined (Fig. 3). Owing to their high-cysteine content, disulfide-rich peptide toxins are ideal candidates for NCL (Jensen et al. 2009).

Venom peptides can also be produced by overexpression in bacteria, yeast, or insect cells. Recombinant protein production is generally more time and cost-effective than SPPS and it enables facile site-directed mutagenesis for SAR studies and isotopic labeling for multidimensional NMR (discussed in the following section). However, it may not be possible to introduce exotic PTMs found in the native toxin using this approach and one is generally limited to genetically encoded L-amino acids for SAR studies.

Escherichia coli, a rod-shaped gram-negative bacterium, is the most widely used recombinant expression system and generally the first choice for recombinant venom-peptide production. The recombinant peptide is generally directed to the cytoplasm but this has the significant disadvantage that it does not allow disulfide-bond formation as the intracellular thioredoxin and glutaredoxin disulfide-reducing pathways ensure that all cytoplasmic cysteine residues are kept in the reduced state (Prinz et al. 1997). Consequently, many venom peptides produced via this approach accumulate as reduced, unfolded proteins in cytoplasmic

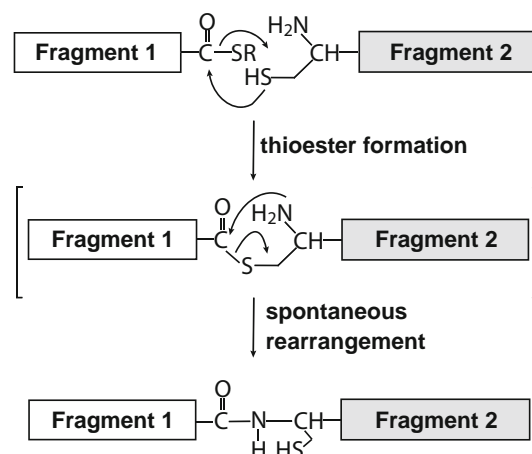


Fig. 3 Synthesis of disulfide-rich toxins using native chemical ligation (NCL). In this schematic example, the toxin has been divided into two fragments. The N-terminal fragment (Fragment 1) has an unprotected peptide- α -thioester at its C-terminus, whereas the C-terminal fragment (Fragment 2) has an N-terminal Cys residue, which forms part of the native sequence, at its N-terminus. Chemoselective reaction of the two fragments yields a thioester-linked intermediate that spontaneously rearranges to form a native amide bond at the ligation site (Dawson et al. 1994). With regard to peptide size, the same limitations apply to synthesis of NCL fragments as with standard SPPS; that is, the efficiency of fragment synthesis will be sequence dependent, but in general will not be problematic up to ~ 30 residues. The folding conditions for toxins produced by NCL will be no different from those produced by recombinant methods or traditional SPPS

inclusion bodies, and protein purification therefore requires time consuming denaturation and refolding steps (Zilberberg et al. 1996; Froy et al. 1999). An alternative approach that potentially obviates this problem is to use an *E. coli* strain with a defective thioredoxin reductase that allows cytoplasmic disulfide bond formation (Tedford et al. 2001). However, in our experience, a more effective approach is to direct the recombinant peptide to the *E. coli* periplasm where the machinery for disulfide-bond formation and isomerization is located. This method has been used previously to produce recombinant snake and scorpion toxins (Ducancel et al. 1989; Legros et al. 1997; Korolkova et al. 2001), and we now routinely use this approach for recombinant production of disulfide-rich spider toxins (see Table 2).

Eukaryotic expression systems have also been used for production of disulfide-rich toxins. However, mammalian and insect expression systems are typically costly, time consuming, produce low yields, and isotopic labeling for NMR studies is difficult and expensive (e.g. Escoubas et al. 2003; Ji et al. 2005). In contrast, yeast has proved to be an excellent system for expression of disulfide-rich animal toxins (Anangi et al. 2007). Yeast expression offers several advantages including high yields, low cost, the ability to incorporate PTMs, such as glycosylation,

Table 2 Recombinant toxins produced by expression in the periplasm of *E. coli*

Toxin	Origin	Number of residues	Number of disulfide bonds	Yield (mg/l)	References
A-erabutoxin	Snake	61	4	0.4–0.7	(Ducancel et al. 1989)
KTX2	Scorpion	63	3	0.2–0.3	(Legros et al. 1997)
AaHI/AaHII	Scorpion	64	4	ND	(Legros et al. 2001)
Iberiotoxin	Scorpion	37	3	1.0–1.5	a
PcTx1	Spider	40	3	3–5	a
TaITX-1	Spider	50	3	ND	a
SFI1	Spider	46	4	ND	a
PaurTx3	Spider	34	3	1.0	a

ND not determined

^a Current authors (unpublished data)

and cost-effective isotopic labeling (although not as facile or cheap as with *E. coli*) (Demain and Vaishnav 2009). Moreover, the ability to secrete folded toxins directly into the yeast growth medium simplifies toxin purification (Daly and Hearn 2005; Anangi et al. 2007). The two most utilized yeasts are *Saccharomyces cerevisiae* and *Pichia pastoris* (Cregg et al. 2000; Macauley-Patrick et al. 2005; Cregg 2007), with the later proving to be an excellent

system for the production of toxins with five or more disulfide bonds (Table 3).

Thus, although still challenging, toxin production is not the major bottleneck it once was for venoms-based drug discovery. In the context of our own drug discovery programs, we have found that periplasmic expression in *E. coli* works well for most toxins, with extracellular secretion in *Pichia pastoris* serving as a backup. SPPS is usually a last

Table 3 Disulfide-rich toxins successfully expressed in yeast

Toxin	Origin	Host	No. of residues	No. of disulfides	Yield (mg/l)	References
Huwentoxin	Spider	<i>S. cerevisiae</i>	55	3	10–12	(Peng et al. 2006)
SHL-Ib1b	Spider	<i>S. cerevisiae</i>	34	3	1.0	(Jiang et al. 2009)
GsMTx4	Spider	<i>P. pastoris</i>	34	3	100	(Park et al. 2008)
Jingzhaotoxin-34	Spider	<i>S. cerevisiae</i>	35	3	4.0	(Chen et al. 2009)
ProTx2	Spider	<i>P. pastoris</i>	30	3	ND	a
Rhodostomin	Snake	<i>P. pastoris</i>	68	6	10–15	(Guo et al. 2001)
Echistatin	Snake	<i>P. pastoris</i>	49	4	2–5	b
Halydin	Snake	<i>P. pastoris</i>	76	7	ND	(You et al. 2003)
Dendroaspin	Snake	<i>P. pastoris</i>	59	4	5–10	(Chen et al. 2006)
Erabutoxin	Snake	<i>P. pastoris</i>	61	4	12–18	b
α -Bungarotoxin	Snake	<i>P. pastoris</i>	75	5	0.1	(Levandoski et al. 2000)
γ -Bungarotoxin	Snake	<i>P. pastoris</i>	68	5	2–5	(Shiu et al. 2004)
κ -Bungarotoxin	Snake	<i>P. pastoris</i>	68	5	0.1	(Fiordalisi et al. 1996)
m1-Toxin1	Snake	<i>P. pastoris</i>	66	4	0.1	(Krajewski et al. 2001)
ButaIT	Scorpion	<i>P. pastoris</i>	38	4	1.0	(Pham Trung et al. 2006)
Bm α Tx14	Scorpion	<i>P. pastoris</i>	64	4	100–120	(Wang et al. 2006)
Bmk M1	Scorpion	<i>P. pastoris</i>	64	4	10	(Shao et al. 1999)
BmP05	Scorpion	<i>S. cerevisiae</i>	31	3	8–10	(Wu et al. 2002)
MgTx	Scorpion	<i>P. pastoris</i>	39	3	12–15	a
AgTX2	Scorpion	<i>P. pastoris</i>	38	3	15–18	a
APETx2	Anemone	<i>P. pastoris</i>	42	3	2–4	a

^a Anangi et al. (unpublished data)

^b Chuang et al. (unpublished data)

resort except when non-natural modifications or exotic PTMs are required. This makes SPPS particularly suitable for production of cone snail peptides since, according to ConoServer (Kaas et al. 2008), 77% of all structurally characterized conopeptides contain at least one PTM other than disulfide bonds.

High-throughput structure determination

Determining the structure of a lead toxin provides a platform for understanding the molecular basis of its interaction with its target and, moreover, the structure can be combined with SAR studies to develop a three-dimensional (3D) pharmacophore that can be used for mimetic design (Baell et al. 2004; Schroeder et al. 2004). The method of choice for high-throughput protein structure determination is generally X-ray crystallography, where much of the process has been automated, from sample handling to the actual structure determination process. However, for proteins smaller than 100 amino acid residues, which includes the vast majority of venom polypeptides, NMR spectroscopy is by far the dominant approach (Fig. 4a). The Protein Data Bank currently contains 170 structures of venom proteins smaller than 9 kDa, of which 82% were determined using NMR. Thus, significant increases in the speed of toxin structure determination will necessarily require advances in the speed of NMR data acquisition and analysis.

There are two established approaches for determining structures by NMR: the homonuclear approach relies exclusively on data from hydrogen nuclei (protons), whereas the heteronuclear approach additionally uses data from carbon and nitrogen nuclei (King and Mobli 2010). The classical homonuclear NMR approach (Wüthrich 1986) relies on 2D datasets that can typically be acquired in <3 days. However, these 2D spectra tend to be densely populated with peaks, often leading to ambiguities that must be manually resolved. As a result, resonance assignment tends to be rate limiting and structures can take months to determine. For this reason, homonuclear-based NMR structure determination is likely to be a major bottleneck in venoms-based drug discovery programs as it would take more than 1 year to determine the structure of even a modest number of hits (20–30) from a high-throughput assay.

For this reason, we advocate using a heteronuclear-based approach. Although 3D heteronuclear NMR datasets can take several weeks to acquire, interpretation of the resultant sparsely populated spectra is much easier as fewer ambiguities are present, data analysis can be completely automated, and the additional information content results in higher quality structures. The heteronuclear NMR

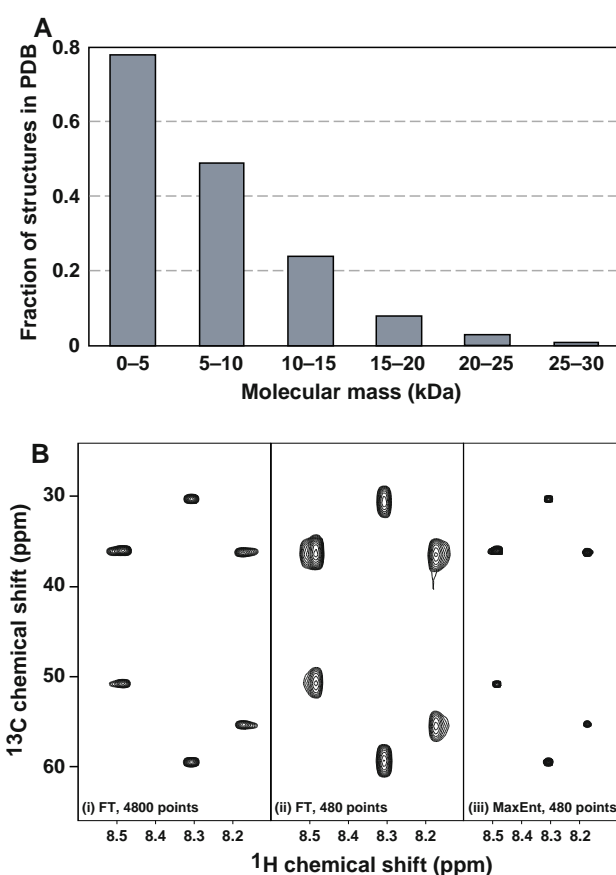


Fig. 4 **a** Fraction of structures in the Protein Data Bank (PDB) determined using NMR as a function of molecular mass. Proteins are grouped into 5-kDa bins. Almost 80% of all structures of proteins <5 kDa were determined using NMR. **b** Comparison of planes from 3D CBCA(CO)NH spectra of a 41-residue spider toxin acquired using either uniform sampling and Fourier transformation (spectra i and ii) or non-uniform sampling (NUS) and maximum entropy (MaxEnt) reconstruction (spectrum iii). Spectrum (i) was collected by uniform sampling of 4,800 complex data points over a period of 42 h. Spectrum (ii) was acquired by uniform sampling of 480 complex data points over a period of 4.2 h, which leads to much lower resolution than for spectrum (i). Spectrum (iii) was acquired by non-uniform sampling of 480 complex data points over 4.2 h, followed by MaxEnt reconstruction. The spectrum obtained via NUS/MaxEnt has significantly higher resolution (i.e., narrower peaks) than the conventional spectrum collected via uniform sampling/Fourier transformation despite being acquired ten times more quickly

approach requires the protein to be uniformly labeled with either ^{15}N or both ^{15}N and ^{13}C , which is now facile due to the development of efficient bacterial and yeast expression systems as described in the previous section. Moreover, numerous methods have been developed in recent years for increasing the speed at which heteronuclear NMR datasets can be acquired. Although several approaches have been proposed, they nearly all rely on the same basic principle, which is to acquire a subset of the complete dataset and then use computational approaches for extracting the relevant information from the “reduced” dataset. This mode

of data acquisition can generally be described as non-uniform sampling (NUS).

Within this general strategy, two different approaches have emerged. One approach is based on correlating data acquisition in two dimensions, which leads to reduced dimensionality projections of the higher dimensional data. Another approach samples at random points of the higher dimensional object and then attempts to approximate the higher dimensional object through a reconstruction process. Methods based on the first approach include G-Fourier transform (GFT) NMR (Szyperski et al. 1993), reduced dimensionality NMR (Ding and Gronenborn 2002), back projection reconstruction (BPR) (Kupče and Freeman 2004), automated projection spectroscopy (APSY) (Hiller et al. 2005), high-resolution iterative frequency identification (Eghbalnia et al. 2005) and projection decomposition (PRODECOMP) (Malmodin and Billeter 2005). The second approach includes methods such as non-uniform FT (Kazimierczuk et al. 2006), maximum entropy (MaxEnt) (Hoch and Stern 2002), multidimensional decomposition (MDD) (Luan et al. 2005), and BPR. Within the protein structure initiative (PSI) funded by the NIH, the Northeast Structural Genomics Consortium (NESG) have chosen to use the GFT approach, the Structural Genomics Consortium Toronto (SGC Toronto) are mainly using MDD, and the Joint Center for Structural Genomics (JCSG) have opted for APSY.

In our structural venomomics efforts at the University of Queensland we use automated MaxEnt (Mobli et al. 2007a; Mobli et al. 2007b; Mobli et al. 2010) (Fig. 4B) as it allows full automation of the sometimes time consuming processing step and, from theoretical analysis and the available experimental data, it is clear that MaxEnt works best in situations of low signal-to-noise. By combining NUS/MaxEnt with automated sequence-specific resonance assignment via the PINE web server (Bahrami et al. 2009) and automated NOESY spectral analysis and structure calculation using the CANDID module in CYANA (Herrmann et al. 2002; Güntert 2004), we can routinely determine the structures of venom proteins in less than one week using sample concentrations as low as 150 μ M (Mobli et al. 2009). Thus, using a heteronuclear NMR approach, 3D structure determination is no longer likely to be a major bottleneck in venoms-based drug discovery programs.

Concluding remarks

Animal venoms can be viewed as a pre-optimized combinatorial library of small peptides with high potency and specificity for a range of cell-surface targets, such as ion channels, transporters, and GPCRs (Sollod et al. 2005). Consequently, they have become a valuable resource for

the discovery of new pharmacological agents. Many of the venom peptides that have emerged from low-throughput assay-guided fractionation studies have proved useful as molecular probes and as lead compounds for drug and insecticide development (Lewis and Garcia 2003; Tedford et al. 2004; Escoubas and King 2009). Nevertheless, the vast majority of the chemical diversity present in animal venoms remains untapped.

Fortunately, the miniaturization of bioassays, advances in recombinant toxin production, and improvements in the sensitivity of mass spectrometry and NMR spectroscopy promise to usher in a new era of venoms-based drug discovery. Not only will these advances expedite the speed at which venoms can be screened and venom peptides characterized, they will facilitate the study of venoms that are only available in small quantities, thus dramatically increasing the biodiversity that can be accessed for drug discovery.

Acknowledgments We acknowledge financial support from the Australian Research Council (Discovery Grants DP0774245 to GFK and DP0878450 to GFK and PFA) and the Australian National Health & Medical Research Council (Project Grant 511067 to LDR, GFK, and PFA, Program Grant 569927 to RJL and PFA, Senior Research Fellowship to RJL, and Australian Biomedical Training Fellowship to IV).

References

- Anangi R, Chen C-Y, Cheng C-H, Chen Y-C, Chen C-C, Chu Y-P, Chang C-H, Jeng W-Y, Shiu J-H, Chuang W-J (2007) Expression of snake venom toxins in *Pichia pastoris*. *Toxin Rev* 26:169–187
- Baell JB, Duggan PJ, Lok YP (2004) ω -Conotoxins and approaches to their nonpeptide mimetics. *Aust J Chem* 57:179–185
- Bahrami A, Assadi AH, Markley JL, Eghbalnia HR (2009) Probabilistic interaction network of evidence algorithm and its application to complete labeling of peak lists from protein NMR spectroscopy. *PLoS Comput Biol* 5:e1000307
- Bennett PB, Guthrie HR (2003) Trends in ion channel drug discovery: advances in screening technologies. *Trends Biotechnol* 21:563–569
- Biaass D, Dutertre S, Gerbault A, Menou JL, Offord R, Favreau P, Stöcklin R (2009) Comparative proteomic study of the venom of the piscivorous cone snail *Conus consors*. *J Proteomics* 72:210–218
- Butler MS (2004) The role of natural product chemistry in drug discovery. *J Nat Prod* 67:2141–2153
- Calvete JJ, Juarez P, Sanz L (2007) Snake venomomics: strategy and applications. *J Mass Spectrom* 42:1405–1414
- Calvete JJ, Sanz L, Angulo Y, Lomonte B, Gutierrez JM (2009) Venoms, venomomics, antivenomics. *FEBS Lett* 583:1736–1743
- Chen CY, Cheng CH, Chen YC, Lee JC, Chou SH, Huang W, Chuang WJ (2006) Preparation of amino-acid-type selective isotope labeling of protein expressed in *Pichia pastoris*. *Proteins* 62:279–287
- Chen J, Zhang Y, Rong M, Zhao L, Jiang L, Zhang D, Wang M, Xiao Y, Liang S (2009) Expression and characterization of jingzhao-toxin-34, a novel neurotoxin from the venom of the tarantula *Chilobrachys jingzhao*. *Peptides* 30:1042–1048

- Clardy J, Walsh C (2004) Lessons from natural molecules. *Nature* 432:829–837
- Cregg JM (2007) Introduction: distinctions between *Pichia pastoris* and other expression systems. *Methods Mol Biol* 389:1–10
- Cregg JM, Cereghino JL, Shi J, Higgins DR (2000) Recombinant protein expression in *Pichia pastoris*. *Mol Biotechnol* 16:23–52
- Cushman DW, Ondetti MA (1991) History of the design of captopril and related inhibitors of angiotensin converting enzyme. *Hypertension* 17:589–592
- D'Suze G, Schwartz EF, Garcia-Gomez BI, Sevcik C, Possani LD (2009) Molecular cloning and nucleotide sequence analysis of genes from a cDNA library of the scorpion *Tityus discrepans*. *Biochimie* 91:1010–1019
- Daly R, Hearn MT (2005) Expression of heterologous proteins in *Pichia pastoris*: a useful experimental tool in protein engineering and production. *J Mol Recognit* 18:119–138
- Davis J, Jones A, Lewis RJ (2009) Remarkable inter- and intra-species complexity of conotoxins revealed by LC/MS. *Peptides* 30:1222–1227
- Dawson PE, Muir TW, Clark-Lewis I, Kent SBH (1994) Synthesis by native chemical ligation. *Science* 266:776–779
- Demain AL, Vaishnav P (2009) Production of recombinant proteins by microbes and higher organisms. *Biotechnol Adv* 27:297–306
- Denyer J, Worley J, Cox B, Allenby G, Banks M (1998) HTS approaches to voltage-gated ion channel drug discovery. *Drug Discov Today* 3:323–332
- Ding K, Gronenborn AM (2002) Novel 2D triple resonance NMR experiments for sequential resonance assignments of proteins. *J Magn Reson* 156:262–268
- Ducancel F, Boulain JC, Trémeau O, Ménez A (1989) Direct expression in *E. coli* of a functionally active protein A—snake toxin fusion protein. *Protein Eng* 3:139–143
- Dunlop J, Bowlby M, Peri R, Vasilyev D, Arias R (2008) High-throughput electrophysiology: an emerging paradigm for ion-channel screening and physiology. *Nat Rev Drug Discov* 7:358–368
- Edwards L, Hessinger DA (2000) Portuguese Man-of-war (*Physalia physalis*) venom induces calcium influx into cells by permeabilizing plasma membranes. *Toxicon* 38:1015–1028
- Eghbalian HR, Bahrami A, Tonelli M, Hallenga K, Markley JL (2005) High-resolution iterative frequency identification for NMR as a general strategy for multidimensional data collection. *J Am Chem Soc* 127:12528–12536
- Escoubas P, King GF (2009) Venomics as a drug discovery platform. *Expert Rev Proteomics* 6:221–224
- Escoubas P, Bernard C, Lambeau G, Lazdunski M, Darbon H (2003) Recombinant production and solution structure of PcTx1, the specific peptide inhibitor of ASIC1a proton-gated cation channels. *Protein Sci* 12:1332–1343
- Escoubas P, Sollod BL, King GF (2006) Venom landscapes: mining the complexity of spider venoms via a combined cDNA and mass spectrometric approach. *Toxicon* 47:650–663
- Escoubas P, Quinton L, Nicholson GM (2008) Venomics: unravelling the complexity of animal venoms with mass spectrometry. *J Mass Spectrom Rev* 43:279–295
- Fiordalisi JJ, James PL, Zhang Y, Grant GA (1996) Facile production of native-like κ -bungarotoxin in yeast: an enhanced system for the production of a neuronal nicotinic acetylcholine receptor probe. *Toxicon* 34:213–224
- Francischetti IM, My-Pham V, Harrison J, Garfield MK, Ribeiro JM (2004) *Bitis gabonica* (Gaboon viper) snake venom gland: toward a catalog for the full-length transcripts (cDNA) and proteins. *Gene* 337:55–69
- Frohman MA (1993) Rapid amplification of complementary DNA ends for generation of full-length complementary DNAs: thermal RACE. *Methods Enzymol* 218:340–356
- Froy O, Zilberberg N, Gordon D, Turkov M, Gilles N, Stankiewicz M, Pelhate M, Loret E, Oren DA, Shaanan B, Gurevitz M (1999) The putative bioactive surface of insect-selective scorpion excitatory neurotoxins. *J Biol Chem* 274:5769–5776
- Fry BG, Roelants K, Champagne DE, Scheib H, Tyndall JD, King GF, Nevalainen TJ, Norman JA, Lewis RJ, Norton RS, Renjifo C, de la Vega RC (2009) The toxicogenomic multiverse: convergent recruitment of proteins into animal venoms. *Annu Rev Genomics Hum Genet* 10:483–511
- Fukuyama Y, Iwamoto S, Tanaka K (2006) Rapid sequencing and disulfide mapping of peptides containing disulfide bonds by using 1, 5-diaminonaphthalene as a reductive matrix. *J Mass Spectrom* 41:191–201
- Glickman JF, Wu X, Mercuri R, Illy C, Bowen BR, He Y, Sills M (2002) A comparison of AlphaScreen, TR-FRET, and TRF as assay methods for FXR nuclear receptors. *J Biomol Screen* 7:3–10
- Güntert P (2004) Automated NMR structure calculation with CYANA. *Methods Mol Biol* 278:353–378
- Guo RT, Chou LJ, Chen YC, Chen CY, Pari K, Jen CJ, Lo SJ, Huang SL, Lee CY, Chang TW, Chung WJ (2001) Expression in *Pichia pastoris* and characterization by circular dichroism and NMR of rhodostomin. *Proteins* 43:499–508
- Hamelin M, Meng X, Cuddy M, Korsun K, Ondeyka J, Simpson PB, Cully DF, Priest BT (2005) A high-throughput assay for modulators of ligand-gated chloride channels. *Assay Drug Dev Technol* 3:59–64
- Harvey AL (1995) From venoms to toxins to drugs. *Chem Ind* 22:914–916
- Herrmann T, Güntert P, Wüthrich K (2002) Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J Mol Biol* 319:209–227
- Hiller S, Fiorito F, Wüthrich K (2005) Automated projection spectroscopy (APSY). *Proc Natl Acad Sci USA* 102:10876–10881
- Hoch JC, Stern AS (2002) Maximum entropy reconstruction, spectrum analysis and deconvolution in multidimensional nuclear magnetic resonance. *Methods Enzymol* 338:159–178
- Jakubowski JA, Kelley WP, Sweedler JV (2006) Screening for post-translational modifications in conotoxins using liquid chromatography/mass spectrometry: an important component of conotoxin discovery. *Toxicon* 47:688–699
- Jensen JE, Durek T, Alewood PF, Adams DJ, King GF, Rash LD (2009) Chemical synthesis and folding of APETx2, a potent and selective inhibitor of acid sensing ion channel 3. *Toxicon* 54:56–61
- Ji W, Zhang X, Hu H, Chen J, Gao Y, Liang S, An C (2005) Expression and purification of Huwentoxin-I in baculovirus system. *Protein Expr Purif* 41:454–458
- Jiang L, Peng L, Chen J, Zhang Y, Xiong X, Liang S (2008) Molecular diversification based on analysis of expressed sequence tags from the venom glands of the Chinese bird spider *Ornithoctonus huwena*. *Toxicon* 51:1479–1489
- Jiang L, Peng L, Zhang Y, Chen J, Zhang D, Liang S (2009) Expression, purification and characterization of a group of lectin-like peptides from the spider *Ornithoctonus huwena*. *Peptides* 30:669–674
- Kaas Q, Westermann JC, Halai R, Wang CK, Craik DJ (2008) ConoServer, a database for conopeptide sequences and structures. *Bioinformatics (Oxford, England)* 24:445–446
- Kazimierczuk K, Koźmiński W, Zhukov I (2006) Two-dimensional Fourier transform of arbitrarily sampled NMR data sets. *J Magn Reson* 179:323–328
- King GF, Mobli M (2010) Determination of peptide and protein structures using NMR spectroscopy. In: Mander L, Liu H-W

- (eds) Comprehensive natural products chemistry II. Elsevier, Amsterdam
- Kita M, Black DS, Ohno O, Yamada K, Kigoshi H, Uemura D (2009) Duck-billed platypus venom peptides induce Ca^{2+} influx in neuroblastoma cells. *J Am Chem Soc* 131:18038–18039
- Kitaguchi T, Swartz KJ (2005) An inhibitor of TRPV1 channels isolated from funnel web spider venom. *Biochemistry* 44:15544–15549
- Kjeldsen F, Giessing AM, Ingrell CR, Jensen ON (2007) Peptide sequencing and characterization of post-translational modifications by enhanced ion-charging and liquid chromatography electron-transfer dissociation tandem mass spectrometry. *Anal Chem* 79:9243–9252
- Koehn FE, Carter GT (2005) The evolving role of natural products in drug discovery. *Nat Rev Drug Discov* 4:206–220
- Korolkova YV, Kozlov SA, Lipkin AV, Pluzhnikov KA, Hadley JK, Filippov AK, Brown DA, Angelo K, Strobaek D, Jespersen T, Olesen SP, Jensen BS, Grishin EV (2001) An ERG channel inhibitor from the scorpion *Buthus eupeus*. *J Biol Chem* 276:9868–9876
- Kozlov S, Malyavka A, McCutchen B, Lu A, Schepers E, Herrmann R, Grishin E (2005) A novel strategy for the identification of toxin like structures in spider venom. *Proteins* 59:131–140
- Krajewski JL, Dickerson IM, Potter LT (2001) Site-directed mutagenesis of m1-toxin 1: two amino acids responsible for stable toxin binding to M_1 muscarinic receptors. *Mol Pharmacol* 60:725–731
- Kupče Ě, Freeman R (2004) Projection-reconstruction technique for speeding up multidimensional NMR spectroscopy. *J Am Chem Soc* 126:6429–6440
- Legros C, Feyfant E, Sampieri F, Rochat H, Bougis PE, Martin-Eauclaire MF (1997) Influence of a NH_2 -terminal extension on the activity of KTX2, a K^+ channel blocker purified from *Androctonus australis* scorpion venom. *FEBS Lett* 417:123–129
- Legros C, Kaabi H, El Ayeb M, Ceard B, Vacher H, Bougis PE, Martin-Eauclaire MF (2001) Use of fusion protein constructs to generate potent immunotherapy and protection against scorpion toxins. *Vaccine* 20:934–942
- Levandoski MM, Caffery PM, Rogowski RS, Lin Y, Shi Q-L, Hawrot E (2000) Recombinant expression of α -bungarotoxin in *Pichia pastoris* facilitates identification of mutant toxins engineered to recognize neuronal nicotinic acetylcholine receptors. *J Neurochem* 74:1279–1289
- Lewis RJ, Garcia ML (2003) Therapeutic potential of venom peptides. *Nat Rev Drug Discov* 2:790–802
- Liao Z, Cao J, Li S, Yan X, Hu W, He Q, Chen J, Tang J, Xie J, Liang S (2007) Proteomic and peptidomic analysis of the venom from Chinese tarantula *Chilobrachys jingzhao*. *Proteomics* 7:1892–1907
- Luan T, Orekhov VY, Gutmanas A, Billeter M (2005) Accuracy and robustness of three-way decomposition applied to NMR data. *J Magn Reson* 174:188–199
- Macarron R (2006) Critical review of the role of HTS in drug discovery. *Drug Discov Today* 11:277–279
- Macauley-Patrick S, Fazenda ML, McNeil B, Harvey LM (2005) Heterologous protein production using the *Pichia pastoris* expression system. *Yeast* 22:249–270
- Malmodin D, Billeter M (2005) Multiway decomposition of NMR spectra with coupled evolution periods. *J Am Chem Soc* 127:13486–13487
- Malone J, Trautmann M, Wilhelm K, Taylor K, Kendall DM (2009) Exenatide once weekly for the treatment of type 2 diabetes. *Expert Opin Investig Drugs* 18:359–367
- Miljanich GP (2004) Ziconotide: neuronal calcium channel blocker for treating severe chronic pain. *Curr Med Chem* 11:3029–3040
- Mobli M, Maciejewski MW, Gryk MR, Hoch JC (2007a) An automated tool for maximum entropy reconstruction of biomolecular NMR spectra. *Nat Methods* 4:467–468
- Mobli M, Maciejewski MW, Gryk MR, Hoch JC (2007b) Automatic maximum entropy spectral reconstruction in NMR. *J Biomol NMR* 39:133–139
- Mobli M, Mijonović T, Bermel W, King GF (2009) ASAP-NMR: a high-throughput pipeline for rapid protein structure determination. In: Proceedings of third Asia-Pacific NMR symposium
- Mobli M, Stern AS, Bermel W, King GF, Hoch JC (2010) A non-uniformly sampled 4D HCC(CO)NH-TOCSY experiment processed using maximum entropy for rapid protein sidechain assignment. *J Magn Reson*:in press
- Newton KA, Clench MR, Deshmukh R, Jeyaseelan K, Strong PN (2007) Mass fingerprinting of toxic fractions from the venom of the Indian red scorpion, *Mesobuthus tamulus*: biotope-specific variation in the expression of venom peptides. *Rapid Commun Mass Spectrom* 21:3467–3476
- Noël F, Mendonça-Silva DL, Quintas LE (2001) Radioligand binding assays in the drug discovery process: potential pitfalls of high throughput screenings. *Arzneimittel-Forschung* 51:169–173
- Olivera BM (2006) Conus peptides: biodiversity-based discovery and exogenomics. *J Biol Chem* 281:31173–31177
- Ortholand JY, Ganesan A (2004) Natural products and combinatorial chemistry: back to the future. *Curr Opin Chem Biol* 8:271–280
- Park SP, Kim BM, Koo JY, Cho H, Lee CH, Kim M, Na HS, Oh U (2008) A tarantula spider toxin, GsMTx4, reduces mechanical and neuropathic pain. *Pain* 137:208–217
- Paterson I, Anderson EA (2005) The renaissance of natural products as drug candidates. *Science* 310:451–453
- Peng K, Lin Y, Liang SP (2006) Nuclear magnetic resonance studies on huwentoxin-XI from the Chinese bird spider *Ornithoctonus huwena*: ^{15}N labeling and sequence-specific ^1H , ^{15}N nuclear magnetic resonance assignments. *Acta Biochim Biophys Sin (Shanghai)* 38:457–466
- Pham Trung N, Fitches E, Gatehouse JA (2006) A fusion protein containing a lepidopteran-specific toxin from the South Indian red scorpion (*Mesobuthus tamulus*) and snowdrop lectin shows oral toxicity to target insects. *BMC Biotechnol* 6:18
- Pimenta AM, Stocklin R, Favreau P, Bougis PE, Martin-Eauclaire MF (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
- Prinz WA, Åslund F, Holmgren A, Beckwith J (1997) The role of the thioredoxin and glutaredoxin pathways in reducing protein disulfide bonds in the *Escherichia coli* cytoplasm. *J Biol Chem* 272:15661–15667
- Quinton L, Demeure K, Dobson R, Gilles N, Gabelica V, De Pauw E (2007) New method for characterizing highly disulfide-bridged peptides in complex mixtures: application to toxin identification from crude venoms. *J Proteome Res* 6:3216–3223
- Richardson M, Pimenta AM, Bemquerer MP, Santoro MM, Beirao PS, Lima ME, Figueiredo SG, Bloch C Jr, Vasconcelos EA, Campos FA, Gomes PC, Cordeiro MN (2006) Comparison of the partial proteomes of the venoms of Brazilian spiders of the genus *Phoneutria*. *Comp Biochem Physiol C Toxicol Pharmacol* 142:173–187
- Rivers DB, Crawley T, Bauser H (2005) Localization of intracellular calcium release in cells injured by venom from the ectoparasitoid *Nasonia vitripennis* (Walker) (Hymenoptera: Pteromalidae) and dependence of calcium mobilization on G-protein activation. *J Insect Physiol* 51:149–160
- Schroeder CI, Smythe ML, Lewis RJ (2004) Development of small molecules that mimic the binding of ω -conotoxins at the N-type voltage-gated calcium channel. *Mol Divers* 8:127–134

- Schwartz EF, Diego-Garcia E, Rodríguez de la Vega RC, Possani LD (2007) Transcriptome analysis of the venom gland of the Mexican scorpion *Hadrurus gertschi* (Arachnida: Scorpiones). *BMC Genomics* 8:119
- Seidler J, Zinn N, Boehm ME, Lehmann WD (2009) De novo sequencing of peptides by tandem mass spectrometry. *Proteomics* 10:634–649
- Shao F, Xiong Y-M, Zhu R-H, Ling M-H, Chi C-W, Wang D-C (1999) Expression and purification of the BmK M1 neurotoxin from the scorpion *Buthus martensii* Karsch. *Protein Expr Purif* 17:358–365
- Shiu J-H, Chen C-Y, Chang L-S, Chen Y-C, Chen Y-C, Lo Y-H, Liu YC, Chuang W-J (2004) Solution structure of γ -bungarotoxin: the functional significance of amino acid residues flanking the RGD motif in integrin binding. *Proteins* 57:839–849
- Siemens J, Zhou S, Piskrowski R, Nikai T, Lumpkin EA, Basbaum AI, King D, Julius D (2006) Spider toxins activate the capsaicin receptor to produce inflammatory pain. *Nature* 444:208–212
- Sollod BL, Wilson D, Zhaxybayeva O, Gogarten JP, Drinkwater R, King GF (2005) Were arachnids the first to use combinatorial peptide libraries? *Peptides* 26:131–139
- Szyperiski T, Wider G, Bushweller JH, Wüthrich K (1993) Reduced dimensionality in triple resonance NMR experiments. *J Am Chem Soc* 115:9307–9308
- Taouji S, Dahan S, Bossé R, Chevet E (2009) Current screens based on the AlphaScreen technology for deciphering cell signalling pathways. *Curr Genomics* 10:93–101
- Tedford HW, Fletcher JI, King GF (2001) Functional significance of the β -hairpin in the insecticidal neurotoxin ω -atracotoxin-Hv1a. *J Biol Chem* 276:26568–26576
- Tedford HW, Sollod BL, Maggio F, King GF (2004) Australian funnel-web spiders: master insecticide chemists. *Toxicon* 43:601–618
- Ueberheide BM, Fenyo D, Alewood PF, Chait BT (2009) Rapid sensitive analysis of cysteine rich peptide venom components. *Proc Natl Acad Sci USA* 106:6910–6915
- Ullman EF, Kirakossian H, Singh S, Wu ZP, Irvin BR, Pease JS, Switchenko AC, Irvine JD, Dafforn A, Skold CN, Wagner DB (1994) Luminescent oxygen channeling immunoassay: measurement of particle binding kinetics by chemiluminescence. *Proc Natl Acad Sci USA* 91:5426–5430
- Ullman EF, Kirakossian H, Switchenko AC, Ishkanian J, Ericson M, Wartchow CA, Pirio M, Pease J, Irvin BR, Singh S, Singh R, Patel R, Dafforn A, Davalian D, Skold C, Kurn N, Wagner DB (1996) Luminescent oxygen channeling assay (LOCI): sensitive, broadly applicable homogeneous immunoassay method. *Clin Chem* 42:1518–1526
- Vetter I, Lewis RJ (2009) Characterization of endogenous calcium responses in neuronal cell lines. *Biochem Pharmacol* 79:908–920
- Wang K, Yin S-J, Lu M, Yi H, Dai C, Xu X-J, Cao Z-J, Wu Y-L, Li W-X (2006) Functional analysis of the α -neurotoxin, Bm α TX14, derived from the Chinese scorpion, *Buthus martensii* Karsch. *Biotechnol Lett* 28:1767–1772
- Wood DL, Miljenovic T, Cai S, Raven RJ, Kaas Q, Escoubas P, Herzig V, Wilson D, King GF (2009) ArachnoServer: a database of protein toxins from spiders. *BMC Genomics* 10:375
- Wu J-J, He L-L, Zhou Z, Chi C-W (2002) Gene expression, mutation, and structure-function relationship of scorpion toxin BmP05 active on SK_{Ca} channels. *Biochemistry* 41:2844–2849
- Wüthrich K (1986) NMR of proteins and nucleic acids. Wiley, New York
- Xu J, Wang X, Ensign B, Li M, Wu L, Guia A (2001) Ion-channel assay technologies: *quo vadis?* *Drug Discov Today* 6:1278–1287
- You WK, Jang YJ, Chung KH, Kim DS (2003) A novel disintegrin-like domain of a high molecular weight metalloprotease inhibits platelet aggregation. *Biochem Biophys Res Commun* 309:637–642
- Yuan C, Jin Q, Tang X, Hu W, Cao R, Yang S, Xiong J, Xie C, Xie J, Liang S (2007) Proteomic and peptidomic characterization of the venom from the Chinese bird spider, *Ornithoctonus huwena* Wang. *J Proteome Res* 6:2792–27801
- Zhang JH, Chung TD, Oldenburg KR (1999) A simple statistical parameter for use in evaluation and validation of high throughput screening assays. *J Biomol Screen* 4:67–73
- Zheng W, Spencer RH, Kiss L (2004) High throughput assay technologies for ion channel drug discovery. *Assay Drug Dev Technol* 2:543–552
- Zilberberg N, Gordon D, Pelhate M, Adams ME, Norris TM, Zlotkin E, Gurevitz M (1996) Functional expression and genetic alteration of an alpha scorpion neurotoxin. *Biochemistry* 35:10215–10222