PLOS ONE

Production of Recombinant Disulfide-Rich Venom Peptides for Structural and Functional Analysis via Expression in the Periplasm of *E. coli*

Julie K. Klint[9], Sebastian Senff[9], Natalie J. Saez, Radha Seshadri, Ho Yee Lau, Niraj S. Bende, Eivind A. B. Undheim, Lachlan D. Rash, Mehdi Mobli*, Glenn F. King*

Institute for Molecular Bioscience, The University of Queensland, St. Lucia, Australia

Abstract

Disulfide-rich peptides are the dominant component of most animal venoms. These peptides have received much attention as leads for the development of novel therapeutic agents and bioinsecticides because they target a wide range of neuronal receptors and ion channels with a high degree of potency and selectivity. In addition, their rigid disulfide framework makes them particularly well suited for addressing the crucial issue of *in vivo* stability. Structural and functional characterization of these peptides necessitates the development of a robust, reliable expression system that maintains their native disulfide framework. The bacterium *Escherichia coli* has long been used for economical production of recombinant proteins. However, the expression of *functional* disulfide-rich proteins in the reducing environment of the *E. coli* cytoplasm presents a significant challenge. Thus, we present here an optimised protocol for the expression of disulfide-rich venom peptides in the periplasm of *E. coli*, which is where the endogenous machinery for production of disulfide-bonds is located. The parameters that have been investigated include choice of media, induction conditions, lysis methods, methods of fusion protein and peptide purification, and sample preparation for NMR studies. After each section a recommendation is made for conditions to use. We demonstrate the use of this method for the production of venom peptides ranging in size from 2 to 8 kDa and containing 2–6 disulfide bonds.

Introduction

Animal venoms are gaining increased attention as a source of novel bioactive peptides [1,2]. The venoms of arthropod predators such as spiders, scorpions, and centipedes are essentially combinatorial peptide libraries that have been optimised for high potency and selectivity against their molecular targets over hundreds of millions of years of evolution [3]. These biochemical attributes have allowed some receptors to be pharmacologically characterized via their interaction with various animal toxins. Moreover, the inherent stability, specificity and potency of disulfide-rich venom peptides have made them an attractive source of lead compounds for the development of new therapeutic agents and bioinsecticides [2,4,5].

One of the major challenges when working with venom peptides is to obtain sufficient material for structural and functional characterisation. This is especially problematic when working with disulfide-rich venom peptides that may form non-native disulfide-bond isoforms [6]. For example, a peptide with three or four disulfide bonds could theoretically form 15 or 105 different disulfide-bond isomers, respectively.

Historically, most venom components were obtained via purification from native material [7]. This approach is only viable for abundant components and even then the small amount of final product often limits the amount of structure-function characterization that is possible. More recently, solid-phase peptide synthesis (SPPS) has become the dominant means by which venom peptides are produced [8–10]. This approach has the advantage of allowing the introduction of non-native amino acid residues and posttranslational modifications. However, in most cases, this approach necessitates extensive screening of *in vitro* folding conditions and it therefore remains an expensive means of producing venom peptides [8]. A less costly approach is recombinant production of venom peptides in a suitable host. The Gram-negative bacterium *Escherichia coli* has long been an attractive host for heterologous protein expression [11]. Heterologous proteins are generally expressed in the cytoplasm of this bacterium as it offers the advantage of high protein yields and simple plasmid constructs. However, a major challenge with intracellular expression of disulfide-rich peptides in *E. coli* are the low yields of correctly folded (native) protein due to the reducing environment in the intracellular space [11]. If allowed to accumulate within the cytoplasm, recombinant proteins are often

sequestered into aggregates known as inclusion bodies. Functional protein can be recovered using denaturant-induced solubilization, followed by optimization of refolding conditions [12]. This is often a laborious process, especially for disulfide-rich peptides, and finding a folding condition that will give high yield of the native fold is not guaranteed. Several approach have been introduced to make the cytoplasm of *E. coli* more suitable for expression of disulfide-rich proteins. These include making the cytoplasm less reducing by introducing mutations into the genes encoding glutathione reductase (*gor*) and thioredoxin reductase (*trxB*) (e.g. Origami™ strains) and by introducing a cytoplasmic disulfide isomerase protein (*DsbC*) to enhance disulfide bond formation (e.g. Shuffle™ strain).[13] An alternative approach to overcoming these problems is to bypass the cytoplasm altogether and have the nascent protein secreted into the periplasm of the bacterium, where the endogenous protein machinery for disulfide bond formation is located [14–16]. In essence, this allows one to hijack the existing *E. coli* refolding machinery in order to produce heterologous peptides with their native disulfide-bond arrangement.

The ability to produce recombinant disulfide-rich peptides in *E. coli* is not only cost effective, but it has the added benefit of enabling isotopic labelling of peptides for multidimensional, heteronuclear NMR studies [17]. NMR is the dominant approach for solving the structure of proteins smaller than 10 kDa, with ~80% of all structures of peptides <5 kDa having been solved using this approach [1,17]. Although homonuclear NMR approaches can be used to solve the structure of unlabelled peptides, the precision and stereochemical quality of the structure is generally better if the peptides are uniformly labelled with ^{15}N and ^{13}C and subjected to 3D/4D heteronuclear NMR experiments [17,18]. Isotopic labelling also facilitates study of the dynamic properties of the peptide [19,20].

Here we present a nine-step protocol for obtaining correctly folded disulfide-rich peptides for functional and structural characterization. This protocol is based on our experience in production of recombinant disulfide-rich venom peptides. Table 1 outlines the range of peptides that have been expressed using this system, which includes peptides ranging in size from 2 to 8 kDa and containing 2–6 disulfide bonds. The table includes both successful and failed attempts and reveals an overall success rate of 75%. Table 1 also includes several biophysical properties that may affect protein expression and folding but within this group of proteins no general trends can be discerned.

In the sections below, each of the 9 steps in this protocol has been divided into three sections: a discussion of what options are available, an explanation of what we do, and finally, based on our experience, what we recommend is the optimal approach.

Step 1 – What vector should I use for expressing disulfide-rich peptides?

What can you do? Vector design is potentially the most important step in the successful expression of any protein/peptide of interest. There are countless choices when it comes to *E. coli* expression vectors and the selection is dependent upon numerous parameters, including the conditions under which the protein/peptide will be induced and purified. Commercially available expression plasmids are an attractive starting point as they offer pre-optimized solutions for expression in specific strains and with certain fusion tags that aid in peptide/protein purification. Vector design or modification of an existing vector can help tailor the components for individual cases. Targeting the construct to the periplasm involves the insertion of a periplasmic export sequence

(or signal sequence) into the vector, such as the *MalE* signal sequence [21].

The addition of one or a combination of fusion tags can have a positive (or, if chosen poorly, negative) influence on the final yield, solubility and folding of the peptide of interest [22]. Popular tags include glutathione *S*-transferase (GST) [23,24], maltose-binding protein (MBP) [25,26], N-utilization substance A (NusA)[27,28], FLAG™ [29], biotin acceptor peptide (BAP) [30], hexahistidine (His$_6$) [31], streptavidin-binding peptide (STREP) [32,33], solubility-enhancing tag (SET) [34], thioredoxin A, and calmodulin-binding peptide (CBP) [35,36]. Among these, MBP has emerged as the preferred solubility tag for a range of diverse proteins. Although MBP is a periplasmic protein and is therefore particularly suited for such expression systems it has also found broad utility in cytoplasmic expression systems [26,37]. An alternative (and counter-intuitive) approach is to introduce a tag that reduces the solubility of the fusion protein, such as the very insoluble ketosteroid isomerase tag that directs the expressed protein into insoluble inclusion bodies [38]. This approach is particularly useful when the protein in question is toxic to the host when soluble and properly folded. It is in these cases more common to include a cleavage site that can be used under denaturing conditions such as cyanabromide cleavage after a methionine residue [39].

In the more common case of a soluble fusion tag it is desirable to remove the tag under native conditions so it does not interfere with downstream applications. Tobacco etch virus (TEV) protease is commonly used to cleave fusion proteins because of its high selectivity [40]; its canonical seven-residue recognition site is ENLYFQ/G, with cleavage prior to the C-terminal Gly residue. Thus, a potential disadvantage of TEV protease is that it leaves a vestigial Gly residue at the N-terminus of the recombinant protein/peptide. Fortunately, detailed studies of the specificity of TEV protease [40] have revealed that it can accommodate a wide range of amino acids in the P1' site, although it prefers short-chain amino acids (Ser, Ala, Gly). Other amino acids can be accommodated but at the expense of cleavage efficiency [40]. In most cases, however, it is possible when designing a plasmid to make the last residue of the TEV protease recognition site coincide with the first residue of the native protein/peptide sequence so that TEV protease cleavage of the recombinant fusion protein yields the native sequence with no vestigial N-terminal residues. Another potential disadvantage of TEV protease is that the addition of a TEV protease cleavage site can reduce the solubility of the expressed protein [41] although we have not found this to be an issue with expression of venom peptides.

Human rhinovirus endoprotease (PreScission™) is a potential alternative to TEV protease. It is highly specific, with an 8-residue recognition site, but it leaves two vestigial N-terminal residues (Gly-Pro) that, in contrast to TEV protease, cannot be substituted with other amino acid residues [42–44]. Thrombin is commonly used to remove fusion protein tags, but it leaves a vestigial Gly-Ser at the N-terminus of the protein/peptide that cannot be replaced with other amino acid residues. Moreover, its six-residue recognition site provides less specificity than with TEV and PreScission™ proteases and consequently there have been reports of non-canonical cleavage by thrombin [44,45]. Finally, Factor Xa and enterokinase can be used to generate native N-termini after digestion because their primary specificity determinants are N-terminal to the scissile bond. However, these protease have short 4/5-residue recognition sequences and thus, like thrombin, they can cleave at non-canonical sites [44,46].

What do we do? A synthetic gene encoding the venom peptide of interest is produced by Geneart AG (Regensburg,

Table 1. Summary of the diverse range of disulfide-rich venom peptides produced in our lab using periplasmic expression.

Toxin name	Organism	No. of residues	No. of SS bonds	Yield of correctly folded peptide[1]	No. of contiguous hydrophobic residues [total]	No. of charged residues [net charge]	pI[2]	Isoforms[3]	Ref.
U2-CUTX-As1a	spider	37	4	++	2 [7]	5 [+1]	7.69	2	[91]
DkTx	spider	76	6	+	4 [23]	22 [+2]	6.84	3	[92]
OAIP 1	spider	33	3	++	2 [6]	11 [+5]	8.35	1	[93]
OAIP 2	spider	33	3	++	2 [10]	8 [+4]	8.64	1	[93]
OAIP 3	spider	34	3	++	3 [9]	5 [+1]	7.77	1	[93]
U1-AGTX-Ta1a	spider	51	3	+++	3 [15]	9 [−1]	5.14	1	[94]
U1-CUTX-As1c	spider	76	4	—	4 [19]	14 [+4]	8.15	—	[95]
U1-PLTX-Pt1a	spider	46	5	—	2 [10]	9 [−1]	5.02	—	[96]
U1-TRTX-Pc1a	spider	33	3	++	3 [12]	5 [+4]	8.35	1	[97]
U2-SGTX-Sf1a	spider	46	4	+++	3 [13]	9 [+1]	6.01	1	[90]
U2-TRTX-Pc1a	spider	28	3	++	1 [5]	5 [+5]	9.22	1	[97]
β-TRTX-Cm1b	spider	33	3	+	2 [8]	12 [+4]	8.86	1	[98]
β-TRTX-Ps1a	spider	34	3	+++	3 [9]	10 [+4]	8.89	1	[98]
β/ω-TRTX-Tp1a	spider	35	3	—	3 [10]	8 [+4]	8.37	—	[99]
κ-TRTX-Gr3a	spider	34	3	+++	3 [11]	9 [+3]	8.66	1	[100]
κ-TRTX-Tb1a	spider	35	3	—	2 [9]	12 [+4]	8.64	—	[101]
κ-TRTX-Tb1b	spider	35	3	—	2 [8]	12 [+2]	8.32	—	[101]
µ-DGTX-Dc1a	spider	56	4	++	4 [16]	19 [+1]	7.71	2	[102]
µ-TRTX-Hhn1a	spider	35	3	+	2 [9]	9 [+4]	8.65	1	[103]
µ-TRTX-Hhn2a	spider	33	3	++	3 [7]	9 [+5]	8.63	1	[104]
µ-TRTX-Hhn2b	spider	33	3	++	3 [9]	9 [+3]	8.64	1	[105]
π-TRTX-Pc1a	spider	40	3	+++	1 [8]	16 [+4]	8.66	1	[20]
τ-TRTX-Pc1c	spider	34	3	—	3 [10]	11 [−3]	4.48	—	[106]
ω-CNTX-Pn4a	spider	55	6	++	2 [12]	15 [+5]	8.3	1	[107]
APETx2	sea anemone	42	3	+	5 [11]	6 [+2]	8.32	1	[8]
OD1	scorpion	65	4	—	4 [21]	15 [+1]	7.67	—	[108]
TxIA	conesnail	17	2	+++	1 [2]	5 [−1]	4.65	1	[109]
MVIA	conesnail	32	3	—	3 [12]	4 [−2]	4.03	—	[110]
κ-SLTPX-Ssm1a	centipede	52	3	+++	2 [11]	21 [+1]	6.46	1	[111]
κ-SLTPX-Ssm2a	centipede	31	3	++	2 [8]	7 [+5]	8.33	3	[111]
µ-SLPTX-Ssm1a	centipede	32	2	+	3 [9]	12 [+2]	7.91	3	[111]

Venom peptides with molecular weight (MW) ranging from 2000–8000 Da and with 2–6 disulfide (SS) bonds have been expressed from a phylogenetically diverse range of venomous animals. [1]Yield: +++, > 5 mg/litre; ++, 1–5 mg/litre; +, 0–1 mg/litre; —, no correctly folded protein obtained; [2]Calculated using ProtParam (http://web.expasy.org/protparam/accessed 20130220); [3]Number of disulfide isoforms evident in final RP-HPLC purification step. N/A = not available.

Germany), who utilize multi-parameter algorithms to optimize codon usage to obtain high levels of protein expression [47]. The venom-peptide gene is subsequently cloned into a variant of the pLic-MBP expression vector [48]. This vector encodes a MalE signal sequence (MalE$_{ss}$) for periplasmic export [49], a His$_6$ tag for affinity purification, a maltose binding protein (MBP) fusion tag to aid solubility [26], and a TEV protease recognition site directly preceding the codon-optimised venom-peptide gene (Fig. 1A). The plasmid is transformed into the protease-deficient *E. coli* strain BL21(λDE3) and expression of the venom-peptide gene is induced with β-D-1-thiogalactopyranoside (IPTG). This leads to export of the fusion protein to the periplasm where the machinery for disulfide-bond formation is located (Fig. 1B).

What do we recommend? We have had considerable success in producing a wide variety of disulfide-rich venom peptides using this variant of the pLic-MBP expression vector (see Table 1) and consequently we recommend trialling this vector system before any other. If possible, engineer the vector so that the C-terminal residue of the TEV cleavage site coincides with the first residue of your protein/peptide. If the N-terminal residue of your protein/peptide is predicted to give poor TEV protease cleavage [40], we recommend using an additional Gly as the N-terminal residue as it is the preferred P' residue for TEV protease and it is likely to have the least impact on the function of your protein/peptide.

Step 2 – What are the optimal growth conditions for maximum yield of fusion protein?

What can you do? The overexpression of fusion proteins mediated via a *lac*/T7 promoter allows for auto-induction [48]. Auto-induction is an attractive alternative to IPTG induction as less handling of the cultures is required and higher cell densities can be achieved, in addition to avoiding the cost of IPTG. A more conventional approach is to monitor the growth of the culture and add the inducer when an OD$_{600}$ of 0.8–1.0 has been attained (i.e. during the log-phase of growth). Induction temperature is also an important consideration. A decrease in temperature prior to induction can limit aggregation of over-expressed protein in the cell [50,51] and it also partially reduces the transcription of heat-shock proteases [52].

What do we do? We use conventional IPTG induction. The *lac* promoter is highly inducible and overexpression of the fusion protein (~45 kDa) is evident even at IPTG concentrations as low as 1 μM (see Fig. 2, step 2). Similarly high levels of expression are obtained at final IPTG concentrations of 10, 100, and 1000 μM. Uninduced cultures typically display some evidence of background ("leaky") expression, which is well characterized for T7 promoter-based vector systems [53]. IPTG is a costly chemical and here we show that IPTG concentrations as low as 10 μM are sufficient for high levels of fusion protein expression. We have found that venom peptides generally express better and are more soluble if the temperature is lowered to 16°C prior to induction (data not shown).
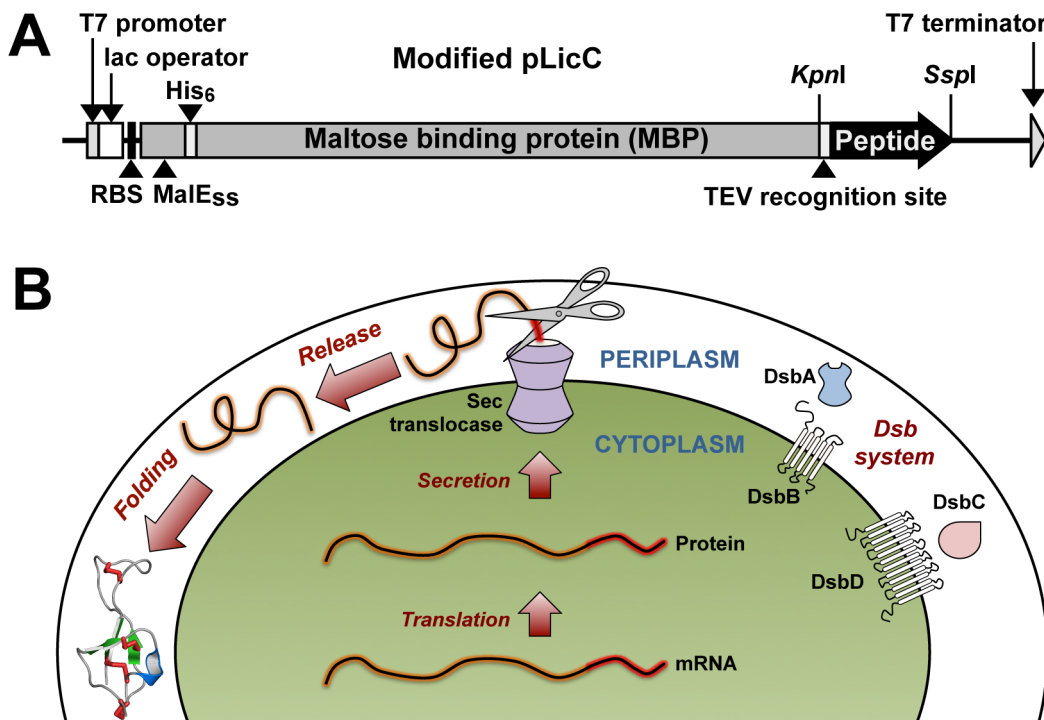
Figure 1. Design of expression vector. (**A**) Schematic representation of the pLic-MBP expression vector using for periplasmic expression of disulfide-rich peptides in *E. coli*. The coding region includes a MalE signal sequence (MalE$_{ss}$) for targeting the fusion protein to the periplasm, a His$_6$ tag for affinity purification, a MBP fusion tag to aid solubility, and a codon-optimised gene encoding the target peptide, with a TEV protease recognition site inserted between the MBP and target-peptide coding regions. The locations of key elements of the vector are shown, including the upstream ribosome binding site (RBS), T7 promoter, lac operator, and key restriction sites. (**B**) Schematic of the periplasmic expression system for production of disulfide-rich peptides in *E. coli*. After translation, the fusion protein is transported to the periplasm via the Sec translocase system [89]. The MalE signal sequence (red tube) is removed during this process, releasing the fusion protein (orange tube) into the periplasm where the Dsb machinery (DsbA, DsbB, DsbC, and DsbD) can assist with disulfide-bond formation.
doi:10.1371/journal.pone.0063865.g001
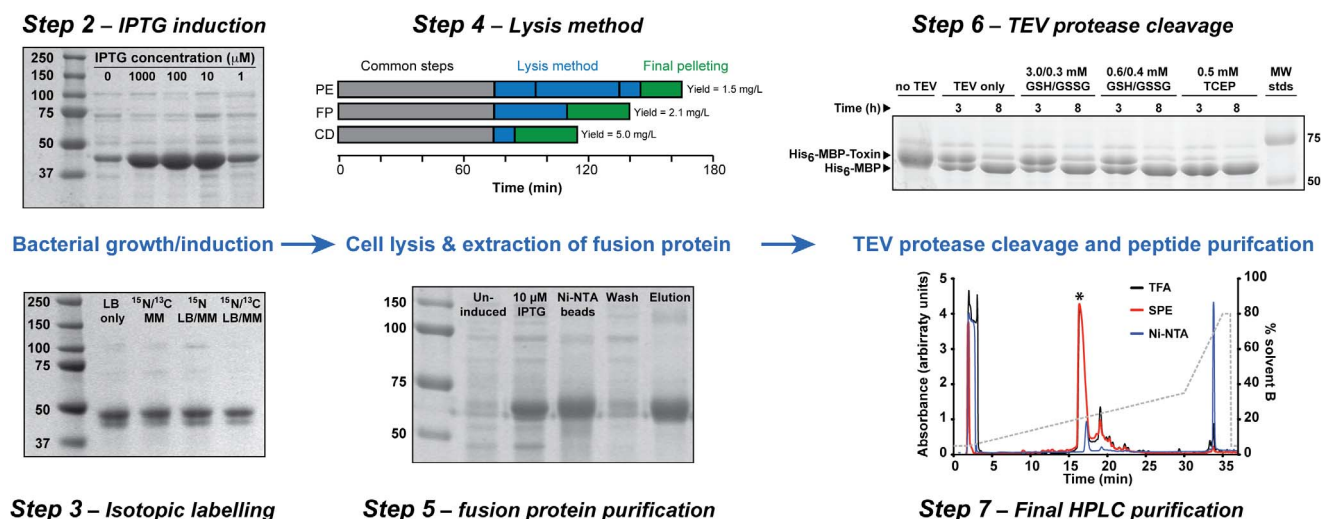
Figure 2. Workflow for obtaining a high yield of recombinant venom peptide. Coomassie stained SDS-PAGE gels illustrating various optimization steps. The molecular mass of the standards (in kDa) are indicated on the left or right of each gel. **Step 2**: The effect of increasing concentrations of IPTG on the level of fusion protein expression. Note that as little as 10 μM ITPG is required for induction of fusion protein expression. **Step 3**: Yields of isotopically labelled fusion protein obtaining using LB medium, minimal medium (MM), and using the dual media protocol (LB/MM). **Step 4**: Comparison of the yield of venom peptide obtained when cell pellets were obtained via periplasmic extraction (PE), French press (FP), or a constant-pressure cell disruptor (CD). **Step 5**: Ni-NTA purification of fusion protein. **Step 6**: Effect of various redox buffers on the efficiency of TEV protease cleavage of the MBP-venom peptide fusion protein. **Step 7**: RP-HPLC chromatograms comparing the efficacy of three different methods for removing His$_6$-tagged fusion protein (MBP) and TEV protease prior to the final peptide purification step: precipitation with 1% TFA (black); removal with a solid phase extraction (SPE) column (red); passage of the cleavage mixture through a Ni-NTA column (blue). The asterisked peak corresponds to the peptide of interest, while the dashed line shows the gradient of solvent B (0.043% TFA in 90% acetonitrile).
doi:10.1371/journal.pone.0063865.g002

What do we recommend? We recommend inducing cultures at a temperature of 16°C, starting with a final IPTG concentration of 100 μM, then titrating down to the lowest IPTG concentration consistent with maximal levels of protein expression.

Step 3 – How do I get maximum yield of ^{13}C and ^{15}N labelled fusion protein?

What can you do? Numerous methods have been reported for increasing the yield of isotopically labelled peptide/protein. They generally fall into two main categories: growing the entire culture in minimal medium (MM) [54] or initial growth in rich medium before transferring to MM (which we term the "*dual media approach*") [55]. Auto-induction is also a possiblity in MM but it necessitates use of costly ^{13}C-labelled glucose and glycerol [56]. MM supplements (e.g. trace metals, vitamin mixtures and commerically available enhancements) have also been shown to have a positive impact on growth and expression [57–59].

Not only is the dual media approach cost-effective but initial growth of the culture is faster compared to the traditional MM method. Maintaining short generation times for *E. coli* (~30 min) by growing cells in a rich medium such as Luria-Bertani medium (LB) during the logarithmic growth phase can result in a large increase in the number of ribosomes per cell [60]. This directly correlates to higher levels of protein expression and ensures that transcription rates do not outstrip those of translation. Additionally, exchanging the medium from LB to MM prior to induction can result in an increase in expression levels by removing by-products that inhibit growth and expression [55]. Exchanging the cells to a MM volume that is one quarter less than the LB volume used to grow the culture effectively results in a four-fold increase in cell density, which can contribute to higher expression levels [55,61]. Achieving similar cell densities using the tradition MM approach would require longer growth times.

What do we do? When comparing the yield of fusion protein between cultures grown in LB, MM, or the dual-media method, we find that the final yield of fusion protein remains essentially constant regardless of the choice of labelling method (Fig. 2, step 3). The dual media approach is more labour intensive and care needs to be taken when changing media in order to avoid unintentional lysis of cells due to the mechanical forces imparted when the cell pellet is resuspended in MM. For the dual media approach, we use a D-glucose concentration of 4 g/L [55]; higher concentrations can increase the yield of recombinant protein but not in a cost-effective manner [61]. This concentration of glucose equates to only a quarter of the isotopically labelled chemicals required for the traditional MM approach. This drastically reduces the cost associated with producing sufficient labelled material for NMR-based structural analysis.

What do we recommend? For producing uniformly isotopically labelled peptides we recommend the dual media protocol for both ^{15}N- and ^{15}N/^{13}C-labelled peptides; it is as efficient as growth in LB and more cost-effective compared to the traditional MM approach. Minor variations to this approach will be required for the incorporation of more exotic isotopic labels such as ^{77}Se-selenocysteine [62], which can aid in the assignment of disulfide bond connectivities [63].

Step 4 – How do I extract my fusion protein from *E. coli* cells?

What can you do? Following induction, the fusion protein can be recovered from the periplasm by selectively bursting the cell wall via osmotic shock [64] or more simply by whole-cell lysis. *E. coli* are easily lysed by several methods and for most laboratory set-ups sonication and freeze/thaw cycles are the method of choice [65]. Controlling the temperature in order to avoid overheating the sample is a problem with sonication, and this can be especially

challenging when dealing with large samples. Additionally, sonication is slow and laborious. Although freeze/thawing is a gentle method of lysing *E. coli* cells, it is often incomplete, and therefore it is often used in combination with chemical agents such as lysosyme (which degrades the bacterial cell wall) and/or Bugbuster™ which introduces extra steps and expensive reagents [66,67].

Several high-shear mechanical methods of cell lysis are also available [68]. In a French press, cells are lysed by passage through a narrow valve under high pressure (typically 150–200 MPa). This approach provides efficient lysis but it is cumbersome because of the size and weight of the apparatus, it is not well suited to processing large volumes, and the pressure cell has to be cleaned after each sample is processed. An alternative that is growing in popularity is constant-pressure cell disruption in which samples are instantly pressurized and passed at high velocity through a small orifice before being returned to atmospheric pressure.

What do we do? We compared periplasmic extraction by osmotic shock with two mechanical lysis methods, namely French press and constant-pressure cell disruption. We found that the time taken to perform the lysis was inversely proportional to the final yield of protein (Fig. 2, step 4). Constant-pressure cell disruption was not only the fastest lysis method but also resulted in higher yields (5.0 mg/L culture compared to 2.1 mg/L for the French press and 1.5 mg/L for periplasmic extraction). We found that several passages through the French press were required for complete cell lysis, and it is likely that incomplete lysis is the reason for the low yield obtained here. The even lower yield obtained using periplasmic extraction may be due to premature lysis when resuspending cells in a hyperosmotic medium prior to lysis [61]. This method is also more time consuming than the mechanical methods, requiring multiple centrifugation and resuspension steps. However, this approach has been used successfully with good yields [20] and it offers the advantage of potentially easier protein purification due to the absence of cytosolic proteins. Whole-cell lysis offers the advantage of shorter processing time and higher yields. Furthermore, we have not observed shuffling of disulfide bonds due to exposure of venom peptides to the cytosolic components of the cell after lysis. This is likely due to the dilution of intracellular glutathione (GSH) and other endogenous reducing agents when cells are lysed into the buffer.

What do we recommend? We recommend using a constant-pressure cell disruptor for cell lysis. In our hands, this approach achieves higher yields of fusion protein more rapidly and with less labor than other lysis methods.

Step 5 – How do I purify my fusion protein?

What can you do? Purification methods based on affinity tags offer a distinct advantage over traditional chromatographic purification methods as they enable selective purification of the protein of interest in a single step. The affinity tag chosen in Step 1 dictates which purification method will be required to "capture" the fusion protein from the cell lysate. For example, immobilized metal affinity chromatography (IMAC) is the method of choice for His$_6$ affinity tags [31], while amylose resin is commonly used for affinity purification of MBP fusion proteins [21,69]. Resins suitable for capturing specific fusion proteins are available for self-packed gravity fed columns. Alternatively, prepacked columns can be used in conjunction with a fast protein liquid chromatography (FPLC) system, which enables gradients of the mobile phase to be applied as well as real-time monitoring of the elution profile [70].

What do we do? The vector we use incorporates both His$_6$ and MBP fusion tags in the expressed fusion protein, and therefore IMAC and amylose affinity purification approaches could be employed individually or in tandem. We purify the His$_6$-MBP fusion construct using a gravity-fed column loaded with Ni-nitrilotriacetic acid (Ni-NTA) beads (Fig. 2, Step 5); in our experience the use of amylose resin either alone or in tandem offers no significant benefit over using IMAC alone (data not shown). As evidenced in lanes 2 and 3 of Fig. 2 (Step 5), the charged Ni-NTA resin retains essentially all of the His$_6$-MBP fusion protein in the cell lysate, with a negligible amount of fusion protein eluted during the column wash stage. Elution of the fusion protein with imidazole reveals the high specificity of this affinity purification approach as the eluate is devoid of other protein components that may have bound non-specifically to the resin (Fig. 2, Step 5, final lane).

What do we recommend? We recommend using Ni-NTA affinity chromatography as the initial purification step because it is simple, efficient, and provides high yield and purity.

Step 6 – How do I cleave my peptide from the fusion tag?

What can you do? We focus here on TEV protease cleavage since our vector encodes a TEV protease cleavage site immediately prior to the target peptide sequence. TEV protease is a cysteine protease that utilizes a cysteine thiol as the active site nucleophile. Thus, TEV protease is only active under reducing conditions; the disulfide-linked dimer is inactive. This necessitates special caution when working with disulfide-rich peptides. The recommended reaction conditions for TEV protease cleavage include 1 mM DTT or 0.5 mM TCEP as a reducing agent, or a redox pair (3.0 mM reduced GSH/0.3 mM oxidized GSH) for proteins containing disulfide bonds [71]. However, these concentrations of GSH and oxidized GSH (GSSG) will yield a solution redox potential of –260 mV (calculated using the Nernst equation [72]). This is a very low redox potential that may result in shuffling of the disulfide bonds into non-native configurations, thus defeating the benefits of targeting the peptides to the disulfide-bond machinery in the periplasm.

The redox potential required to maintain TEV activity is unknown. However, since the disulfide-bond in the protease dimer is non-native and unlikely to be stabilising, it is likely that its redox potential is close to or higher than that of free cysteine, which is about –220 mV [73]. In contrast, the disulfide bonds of venom peptides have evolved to be major contributors to peptide stability and their redox potential is likely to be lower than that of cysteine. Thus, we use TEV cleavage buffer with 0.6 mM GSH and 0.4 mM GSSG, resulting in a redox potential of –215 mV. This milder redox buffer should maintain TEV protease in an active, reduced state while minimizing the shuffling of the pre-formed disulfide bonds in the target peptide.

What do we do? To test this theory, fusion proteins were incubated with TEV protease under the recommended conditions of 0.5 mM TCEP and 3.0/0.3 mM GSH/GSSG or using the redox conditions we recommend of 0.6/0.4 mM GSH/GSSG. As can be seen from Fig. 2 (Step 6), all tested conditions had the same cleavage efficiency with the TCEP buffer yielding faster cleavage. Somewhat surprisingly, the fusion protein sample with no redox agents present was cleaved just as efficiently as when a reducing agent was added. This suggests that the dimerisation of TEV protease is slow and that sufficient amounts of monomeric TEV protease are available during the cleavage step.

What do we recommend? The source of TEV protease and the buffers used during its purification will influence its activity and the redox potential required for full activation. (We produce His$_6$-tagged TEV protease in-house, with final purification in a buffer containing no reducing agent). Thus, we recommend initially

performing TEV cleavage in a buffer containing 0.6/0.4 mM GSH/GSSG; the redox potential of the buffer can then be increased or decreased if shuffling of disulfide bonds or inefficient cleavage is observed, respectively.

Step 7 – How do I get rid of the affinity tags after fusion protein cleavage?

What can you do? TEV protease of the fusion protein yields a mixture of the target peptide, His_6-MBP, and in our case His_6-TEV protease. Thus, in principle, the peptide can be purified from the His_6-tagged MBP and TEV protease by passing the mixture through a Ni-NTA column as in the initial purification step. However, it should be noted that the backbone of a Ni-NTA bead has similar properties to a C_4 reverse-phase (RP) HPLC column [74] and thus it will retain hydrophobic peptides. Hence, this should not be the method of choice for removing fusion tags if your peptide is hydrophobic.

Using a syringe-driven solid-phase extraction (SPE) column offers a facile alternative for separation of the large hydrophobic components (TEV protease and MBP) from the target peptide, even if the latter is somewhat hydrophobic. The disadvantage of this method is that one needs to first determine the amount and volume of organic solvent needed to elute the peptide, but not the fusion tags. Furthermore, the peptide is eluted in a solution with a high content of organic solvent that may need to be removed before the final purification step. A simple alternative approach is to precipitate out the larger proteins by addition of 1% trifluoroacetic acid (TFA). However, the disadvantage of this method is the chance of co-precipitation of the peptide with a resultant decrease in the final yield.

What do we do? We compared the three methods just discussed (i.e., Ni-NTA column, SPE column, and TFA precipitation) with respect to the efficiency of MBP and TEV protease removal as well as recovery of target peptide (Fig. 2, Step 7). Although passage of the cleavage mixture over a Ni-NTA column works efficiently for some peptides [20], it can be seen from the blue trace in Fig. 2 (Step 7), that the yield of the target peptide chosen for this comparison (3.8 kDa, 3 disulfide bonds, spider peptide) is significantly lower when using the Ni-NTA column than for the two other methods tested. The SPE column and TFA extraction yielded similar amounts of the target peptide (Fig. 2, Step 7, black and red traces) and the removal of MBP and TEV protease was equally efficient.

After removal of MBP and TEV protease, a final chromatography step is typically (but not always) required to increase the purity of the target peptide to >95%, which is essential for most structural and functional analyses. We find that RP-HPLC is the most efficient method for the final purification step. The length of the alkyl chain on the RP-HPLC beads (i.e., C_{18}, C_8 or C_4) should be chosen in inverse proportion to the hydrophobicity of the target peptide. Using a C_{18} RP-HPLC column to purify a large and highly hydrophobic peptide will result in poor yields. In our experience, RP-HPLC is almost always sufficient to produce peptides with purity >95%. Only in rare circumstances will an orthogonal chromatography step, such as ion exchange chromatography, be required to increase peptide purity to the desired level.

What do we recommend? This step is highly dependent on the properties of the peptide being expressed. Unless your peptide is known to be highly hydrophobic (which would therefore exclude the Ni-NTA option), we recommend initially splitting the cleavage mixture into three batches and comparing the three methods described above for removal of the fusion tags.

Step 8 – What is the best way to quantify the concentration of recombinant peptide?

What can you do? Accurate quantitation of peptide concentration is very important for both functional and structural analyses. Quantitative amino acid analysis is the gold standard for accurately determining protein/peptide concentration as it is highly accurate and reproducible [75]. However, the procedure is complex, expensive, and generally outsourced to specialised laboratories [76]. For day-to-day experiments, simpler means of quantification are needed. Since the focus here is disulfide-rich venom peptides which typically contain multiple aromatic residues, the UV absorption at 280 nm (A_{280}) provides an efficient means of concentration estimation using the molar extinction coefficient (ε) and Beer-Lambert's law ($C_{peptide} = A_{280}/\varepsilon \cdot l_{path\ length}$) [77]. Estimates of ε based on the amino acid sequence of the target peptide can be obtained via online predictors such as ProtParam (http://web.expasy.org/protparam accessed 20130220). A_{280} measurements are commonly being made using instruments such as the NanoDrop (ThermoScientific) due to the ease of handling and the small sample volumes required. Cuvette-based spectrophotometers employ fixed cells that require larger volumes than the NanoDrop. However, the cuvette-based approach is non-destructive and the sample can be recovered after measurement.

Dye-based methods such as the Bradford protein assay [78] and the bicinchoninic acid (BCA) protein assay [79] can also be used for protein quantification. The Bradford and BCA assays are based on the shift in absorbance when a dye (Coomassie Brilliant Blue G-250 or BCA, respectively) binds to the protein. The popular Lowry protein assay [80] is a more chemically complex assay that relies on the reaction of Cu^+ ions, produced by oxidation of peptide bonds under alkaline conditions, with Folin-Ciocalteu reagent, leading to the oxidation of aromatic and cysteine residues; the concomitant reduction of Folin reagent can be measured at 750 nm. The major disadvantage of these methods is that the amount of dye bound, or Folin reagent reduced, depends on the sequence of the protein in an unpredictable manner. Hence, one always needs to concomitantly acquire a standard curve in order to estimate protein concentrations, and this can still lead to inaccurate concentration estimates if the property of the protein chosen as a standard differs significantly from the peptide/protein whose concentration is being measured. Moreover, in contrast with A_{280} measurement, these methods destroy the peptide/protein.

It should be noted that for most absorbance-based assays, the linear range is typically 0.1–1.0 absorbance units and hence readings outside this range will lead to inaccurate estimates of the protein concentration.

What do we do? We compared peptide concentrations estimated using a BCA assay or A_{280} measurements determined using a NanoDrop or conventional spectrophotometer (Table 2) under both native and denaturing conditions (GnHCl). Overall we found that for these peptides denaturation is not generally required as the hydrophobic core of these molecules is rather small compared to large proteins, and most hydrophobic residues are solvent exposed. Traditional spectrophotometric measurement using a quartz cuvette with a path length of 1 cm produced the most accurate and reproducible measurements. The lowest protein concentration that could be reliably measured using this approach was ~2 µM, and the standard error between readings was ~0.5 µM. In contrast, the lowest protein concentration that could be reliably measured using the NanoDrop was 20 µM and the standard error was 5 µM.

Table 2. Comparison of different methods for determining protein concentration.

Instrument	Solution	Assay Method	Concentration
Plate reader	BCA reagents	Abs at 562 nm	103.9 ± 13.9
Spectrophotometer	H_2O	Abs at 280 nm	107.0 ± 6.9
Spectrophotometer	GnHCl	Abs at 280 nm	94.2 ± 5.7
NanoDrop*	H_2O	Abs at 280 nm	88.0 ± 8.6
NanoDrop*	GnHCl	Abs at 280 nm	98.1 ± 11.6

Concentration determination was performed using five dilutions (1:1, 1:2, 1:5, 1:10 & 1:20) of a standard peptide solution (100 µM). The lowest dilution produced unreliable readings for all methods other than the BCA assay, and was not used for calculating the concentration or the standard deviation in those methods. Only the 1:1 and 1:2 dilutions produced reliable readings when using the NanoDrop; for this method, all other readings were omitted when calculating the concentration and standard deviation. * dilutions below 20 µM were unreliable and excluded.
doi:10.1371/journal.pone.0063865.t002

What do we recommend? If highly accurate determination of protein concentration is critical (e.g., for functional assays or circular dichroic spectropolarimetry), we recommend using a specialised laboratory for quantitative amino acid analysis. For all other purposes, we recommend determining peptide concentration via measurement of A_{280} with a conventional spectrophotometer. If the volume required is prohibitive (standard quartz cuvettes require ~300 µL, with specialised ones requiring as little as ~50 µL), a NanoDrop could be used, but due to the large variance between readings we recommend that the concentration estimate is based on triplicate measurements of at least at 3 dilutions of the protein.

Step 9 – How do I acquire the best NMR data from my peptide sample?

What can you do? The amount of peptide required for NMR studies will vary considerably depending on the NMR system available. Concentrations of 800–1000 µM will be suitable for most NMR instruments, with concentrations as low as 200 µM being sufficient for structural studies on a cryoprobe-equipped 900 MHz magnet [17,19]. Standard 5-mm NMR tubes require volumes of ~500 µL. If material is limited, one can use susceptibility-matched microtubes, which require ~300 µL. These tubes, although more expensive, have the added benefit of minimising problems associated with convection, which can lead to poor suppression of solvent resonances.

Once a sample of sufficient concentration has been obtained, solution conditions such as pH, temperature, and ionic strength should be explored in order to maximise NMR spectral quality. The buffer should be chosen primarily to ensure that the peptide is kept at a pH where it will not aggregate during the long period of time (days to weeks) required to acquire a complete NMR dataset. In general, the sample pH should be at least one pH unit away from the isoelectric point of the peptide. High solvent conductivity can have a detrimental effect on NMR signal intensities [81] and therefore low-conductivity buffers such as MES are preferred over high-conductivity buffers such as sodium phosphate. For the same reasons, high salt concentrations should also be avoided if possible. For samples that require high salt concentrations it is worth considering alternative NMR sample tubes such as shaped tubes or 3-mm tubes to reduce sensitivity problems [82]. Finally, a screen of additives can be conducted if the sample is unstable or not monodisperse; common additives include mild detergents or salts

[83-85]. Sample quality is most easily assessed by recording a 2D ^{1}H-^{15}N HSQC spectrum [17], but this requires a ^{15}N-labelled sample. For more extensive screening, unlabelled samples may be used, and 1D T_2 relaxation experiments [83] or pulsed-field gradient measurements of translational diffusion [84] can be used to estimate the aggregation state of the protein.

What do we do? Fortunately, disulfide-rich peptides are commonly secreted and thus inherently very soluble. The concentration of peptides in the venom gland far exceeds the concentration used in NMR experiments and functional assays. Our standard screen therefore involves acquisition of 2D ^{1}H-^{15}N HSQC spectra from a ^{15}N-labelled sample prepared in four different buffers: two at pH 6, one with 20 mM MES and the other with 20 mM sodium phosphate (low and high conductivity); one at pH 5 in 20 mM acetate buffer; and finally one at pH 4 in 20 mM citrate buffer. Where possible, we choose to work with one of the pH 6 buffers, as these are closest to physiological conditions.

Once the optimal buffer has been chosen, HSQC spectral quality is examined at a range of temperatures starting at 298 K and increasing up to 310 K (or higher if spectral quality is still improving with increases in temperature). In most cases, there is little difference between these temperatures, and hence the lowest temperature (298 K) is chosen for reasons of sample stability. In some instances, however, multiple peptide conformers are observed due to exchange processes such as *cis/trans* isomerisation of X-Pro peptide bonds [86] or conformational disulfide isomerism [87]. The presence of multiple conformers can greatly complicate NMR spectral assignment [86] and hence it is often worthwhile to try manipulating the equilibrium between conformers by subtle changes in temperature, pH, and solvent (see Fig. 3).

Finally, ^{2}H$_2$O (5%) and DSS [4,4-dimethyl-4-silapentane-1-sulfonic acid] (10 µM) are added for locking and referencing the NMR signals whilst additives such as EDTA, protease inhibitors, and NaN$_3$ (0.02%) are often added to remove paramagnetic metal ions, prevent peptide digestion, and inhibit bacterial growth respectively.

What do we recommend? We recommend preparing a ^{15}N-labelled sample in order to assess peptide yields in minimal medium and to screen NMR conditions. Buffer conditions can be screened by acquiring 2D ^{1}H-^{15}N HSQC spectra of the sample in the four buffers described above immediately after sample preparation and after one week at room temperature. If the sample is stable after one week, prepare a ^{13}C/^{15}N-labelled sample in the buffer that gives the highest quality spectra; aim for a final concentration of ≥ 300 µM in 300 µL. Add 10 µM DSS and 0.02% NaN$_3$, then filter the sample using a low-protein-binding Ultrafree-MC centrifugal filter (0.22 µm pore size). Details of our approach to NMR structure determination are described elsewhere [1,17,19].

Conclusion – Have faith and be persistent!

We have made recommendations for *E. coli*-based production of recombinant disulfide-rich peptides based on our extensive experience with expression of disulfide-rich venom peptides. This is not intended to be a universal protocol that is guaranteed to work for all disulfide-rich peptides. The level of protein expression and the solution properties of the recombinant fusion proteins and peptides can vary greatly despite small sequence variations; small optimizations will often needed for each individual case. Nevertheless, we hope that this article serves as a useful starting point for those interested in the challenge of producing recombinant peptides or proteins with multiple disulfide bonds.
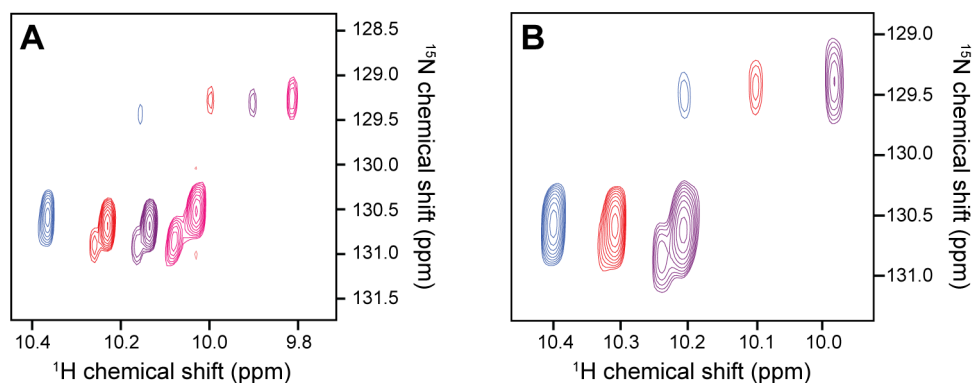
Figure 3. Effect of buffer, temperature, and pH on 2D ^{1}H-^{15}N HSQC NMR spectra of a disulfide-rich venom peptide (Step 9). (A) Overlays of the downfield region of 2D ^{1}H-^{15}N HSQC spectra of a spider-venom peptide (46 residues, 4 disulfide bonds) [90] acquired at 25°C using different buffers and pH: 20 mM MES pH 6 (pink); sodium phosphate, pH 6 (purple); 20 mM sodium acetate, pH 5 (red); 20 mM sodium citrate, pH 4 (blue). This region of the spectrum shows the sidechain ^{1}H-^{15}N correlation for the single Trp residue in this peptide. **(B)** Effect of temperature on the same resonance. Spectra were acquired in 20 mM citrate, pH 4 at the following temperatures: 10°C (purple); 25°C (red); 40°C (blue). At low pH and high temperature the equilibrium is shifted towards a single conformer, compared to the three conformers apparent at lower temperature.
doi:10.1371/journal.pone.0063865.g003

Materials and Methods

Steps 1 & 2: Recombinant protein production

Synthetic genes encoding the venom peptides with codons optimized for expression in *E. coli* were cloned into a variant of the pLic-MBP expression vector expression [48] by Geneart AG (Regenburg, Germany). Plasmids were transformed into *E. coli* strain BL21(λDE3) for recombinant toxin production. Cultures were grown at 37°C in LB medium supplemented with ampicillin (100 μg/mL) with shaking at 180 rpm. When the OD_{600} reached 0.8–1.0 the culture was cooled to 16°C and induced with 250 μM IPTG. Cells were harvested 12–14 h later by centrifugation for 15 min at 7741 g. Gel samples were mixed 1:1 with 2× Coomassie blue loading dye and boiled for 5 min. 15 μL of each sample was loaded on a 12.5% SDS-PAGE gel and 7 μL of Precision Plus Protein^{TM} standards was added to one lane to provide molecular mass markers. The gel was run for 60 min at 160 V.

Step 3: Recombinant production of isotopically-labelled peptide

Dual media approach (LB/MM): Cultures were grown at 37°C in LB media supplemented with ampicillin (100 μg/mL) with shaking at 180 rpm. When the OD_{600} reached 0.8–1.0 the cells were harvested by centrifugation for 10 min at 3000 g. LB medium was poured off and the cell pellet resuspended carefully in M9 salts medium (22 mM KH_2PO_4, 90 mM Na_2HPO_4, 17 mM NaCl. 1.6 mM $MgSO_4$. 80 nM $CaCl_2$, 18 nM NH_4Cl or $^{15}NH_4Cl$, 22 mM D-glucose or $^{13}C_6$-D-glucose. 2 μg/mL thiamine, 0.002% (v/v) vitamin solution, 100 μg/mL ampicillin). Cells were always resuspended in a volume of M9 medium equal to one fourth of the volume of LB used to grow the cells (e.g., 500 mL M9 medium would be used for a 2 L LB culture). The labelled culture was returned to a 37°C incubator with shaking at 180 rpm for 1 h to promote cell recovery, then cooled to 16°C and induced with 250 μM IPTG. Cells were harvested 12–14 h later by centrifugation for 15 min at 7741 g.

Minimal medium (MM) approach: Briefly, the medium composition was as follows: 16 nM $FeCl_2$, 0.66 nM $CaCl_2$, 0.41 nM H_3BO_4, 0.13 nM $MnCl_2$, 0.07 nM $CoCl_2$, 0.01 nM $CuCl_2$, 1 mM $ZnCl_2$, 1.2 mM Na_2MoO_4, 9.7 mM KH_2PO_4, 40.2 mM K_2HPO_4, 25 mM NaCl, 1.2 mM $MgCl_2$, 275 nM K_2SO_4, 18 mM NH_4Cl or $^{15}NH_4Cl$, and 22 mM D-glucose or

$^{13}C_6$-D-glucose. Cell cultures were grown, induced, and harvested as described for the dual media approach.

Step 4: Lysis of cells

Cell disruption Whole cells were resuspended in 40 mM Tris, 400 mM NaCl, pH 8 (TN buffer) then the His$_6$-MBP-toxin fusion protein was extracted from the bacterial periplasm by continuous flow cell disruption (TS Series Benchtop System, Constant Systems Ltd, Northants, UK) at a constant pressure of 30 kPa.

Periplasmic extraction The His$_6$-MBP-toxin fusion protein was extracted from the bacterial periplasm using osmotic shock. Briefly, the cell pellet was defrosted on ice to prevent lysis, resuspended in 30 mM Tris HCl, 2 mM EDTA, 40% sucrose, pH 7.2, then centrifuged at 17,418 for 20 min. The supernatant was discarded. The pellets were resuspended in ice-cold water and 20 mM $MgCl_2$, then incubated on ice for 10 min before centrifugation as before. The supernatant was collected and enough 2 mM Tris pH 8.0, 5 M NaCl and 100% glycerol was added to give a final concentration of 20 mM Tris, 200 mM NaCl, 10% glycerol pH 8.0 (TNG buffer). The sample was then subjected to Ni-NTA affinity chromatography.

French Press The His$_6$-MBP-toxin fusion protein was extracted from the bacterial periplasm using a French Pressure Cell System (Biolab, Scoresby, VIC, Australia). Briefly, the cell pellet was defrosted on ice to prevent lysis, resuspended in TNG buffer, then passaged three times through a pre-cooled cell at 1,000 psi. The flow-through was collected and centrifuged at 41,399 g for 30 min. The supernatant was subjected to Ni-NTA affinity chromatography.

Step 5 & 6: Purification and cleavage of fusion protein

The soluble lysate fraction was isolated by centrifugation at 41,399 g for 30 min and the His$_6$-MBP-toxin fusion protein was captured by passing the supernatant over Ni-NTA Superflow resin (QIAGEN, Valencia, CA, USA) followed by washing with 15 mM imidazole in TN buffer to remove nonspecifically bound proteins. The fusion protein was then eluted with 500 mM imidazole in TN buffer. The imidazole was removed by centrifugal filtration (Amicon ® Ultra, Milipore), then GSH and GSSG were added to 0.6 and 0.4 mM, respectively. Approximately 40 μg of recombinant His$_6$-tagged TEV protease (made in-house according

to [88]) was added per mg of fusion protein. The cleavage reaction was allowed to proceed at room temperature for 12 h with shaking.

Step 7: Purification of peptide

TFA precipitation The cleaved His$_6$-MBP and TEV protease were removed by precipitation with 1% v/v TFA followed by centrifugation of the precipitate for 41,399 g for 10 min. The supernatant was further purified using RP-HPLC using a Supelco apHera^{TM} C$_4$ analytical column (150×4.6 mm; particle size, 5 μm, pore size 300 Å), a flow rate of 1 ml/min and a gradient of 5 to 35% solvent B (0.043% TFA in 90% ACN) in solvent A (0.05% TFA in water) over 28 min.

Reapplication of cleavage mixture to Ni-NTA column The cleaved His$_6$-MBP and His$_6$-TEV protease were removed from the cleavage mixture by passing it over Ni-NTA Superflow resin (Qiagen, Valencia, CA). The initial flow-through was collected, filtered and further purified using RP-HPLC as described.

SPE column purification The cleavage mixture was added to solvent B to give a final concentration of 5% and filtered using a 0.22 μm syringe filter. The resulting filtrate was passed through a Maxi-Clean^{TM} C$_{18}$ large-pore SPE column (Grace Davison Discovery Sciences, Rowville, VIC, Australia) conditioned with methanol and equilibrated at 5% solvent B. Peptide, cleaved His$_6$-MBP, and TEV protease were separated and eluted over a range of solvent B concentrations (5 mL each of 5%, 10%, 20%, 30%, 40%, 50%, 60% and 100%). Eluted samples were lyophilized and further purified using RP-HPLC as described.

Step 8: Quantitation of peptide concentration

A ~100 μM stock solution of a U$_1$-AGTX-Ta1a was diluted 1:1, 1:2, 1:5, 1:10 and 1:20 in water. The A$_{280}$ absorbance was assessed in both water and guanidine hydrochloride (GnHCl). Briefly, 50 μL of the dilution stocks were made up to a final volume of 1 mL with GnHCl (6 M). A$_{280}$ measurements were recorded using either ~1 μL drops on a NanoDrop^{TM} or using a 300 μL quartz cuvette in a Varian Cary 50 UV/Vis-spectrophotometer (Agilent Technologies, Mulgrave, VIC, Australia). A standard BCA assay (Thermo Scientific, Waltham, MA) was also performed on all dilutions. The absorbance at 562 nm of samples from the BCA assay were measured using a standard 96-well plate reader (PowerWave XS, Bio-Tek, VT).

Author Contributions

Conceived and designed the experiments: NJS MM JKK SS GFK. Performed the experiments: JKK SS MM RS NJS HYL NSB. Analyzed the data: JKK SS MM. Contributed reagents/materials/analysis tools: JKK SS MM RS NJS HYL NSB EABU LDR GFK. Wrote the paper: JKK SS MM GFK.

References

1. Vetter I, Davis JL, Rash LD, Anangi R, Mobli M, et al. (2011) Venomics: a new paradigm for natural products-based drug discovery. Amino Acids 40: 15–28.
2. King GF (2011) Venoms as a platform for human drugs: translating toxins into therapeutics. Expert Opin Biol Ther 11: 1469–1484.
3. Sollod BL, Wilson D, Zhaxybayeva O, Gogarten JP, Drinkwater R, et al. (2005) Were arachnids the first to use combinatorial peptide libraries? Peptides 26: 131–139.
4. Lewis RJ, Garcia ML (2003) Therapeutic potential of venom peptides. Nat Rev Drug Discov 2: 790–802.
5. King GF, Hardy MC (2012) Spider-venom peptides: structure, pharmacology, and potential for control of insect pests. Annu Rev Entomol: 475–496.
6. Annis I, Hargittai B, Barany G (1997) Disulfide bond formation in peptides. Methods Enzymol 289: 198–221.
7. Tytgat J, Vandenberghe I, Ulens C, Van Beeumen J (2001) New polypeptide components purified from mamba venom. FEBS Lett 491: 217–221.
8. Jensen JE, Durek T, Alewood PF, Adams DJ, King GF, et al. (2009) Chemical synthesis and folding of APETx2, a potent and selective inhibitor of acid sensing ion channel 3. Toxicon 54: 56–61.
9. Walewska A, Zhang MM, Skalicky JJ, Yoshikami D, Olivera BM, et al. (2009) Integrated oxidative folding of cysteine/selenocysteine containing peptides: improving chemical synthesis of conotoxins. Angew Chem Int Ed 48: 2221–2224.
10. Morales RAV, Daly NL, Vetter I, Mobli M, Napier IA, et al. (2010) Chemical synthesis and structure of the prokineticin Bv8. ChemBioChem 11: 1882–1888.
11. Makrides SC (1996) Strategies for achieving high-level expression of genes in Escherichia coli. Microbiol Rev 60: 512–538.
12. Rudolph R, Lilie H (1996) In vitro folding of inclusion body proteins. FASEB J 10: 49–56.
13. Salinas G, Pellizza L, Margenat M, Fló M, Fernández C (2011) Tuned Escherichia coli as a host for the expression of disulfide-rich proteins. Biotechnol J 6: 686–699.
14. Choi JH, Lee SY (2004) Secretory and extracellular production of recombinant proteins using Escherichia coli. Appl Microbiol Biotechnol 64: 625–635.
15. Baneyx F, Mujacic M (2004) Recombinant protein folding and misfolding in Escherichia coli. Nat Biotechnol 22: 1399–1408.
16. Heras B, Shouldice SR, Totsika M, Scanlon MJ, Schembri MA, et al. (2009) DSB proteins and bacterial pathogenicity. Nat Rev Microbiol 7: 215–225.
17. Kwan AH, Mobli M, Gooley PR, King GF, Mackay JP (2011) Macromolecular NMR spectroscopy for the non-spectroscopist. FEBS J 278: 687–703.
18. King GF, Mobli M (2010) Determination of peptide and protein structures using NMR spectroscopy. In: Mander L, Liu H-W, editors. Comprehensive

Natural Products Chemistry II: Chemistry and Biology. Oxford:Elsevier. pp. 279–325.
19. Bieri M, Kwan AH, Mobli M, King GF, Mackay JP, et al. (2011) Macromolecular NMR spectroscopy for the non-spectroscopist: beyond macromolecular solution structure determination. FEBS J 278: 704–715.
20. Saez NJ, Mobli M, Bieri M, Chassagnon IR, Malde AK, et al. (2011) A dynamic pharmacophore drives the interaction between psalmotoxin-1 and the putative drug target acid-sensing ion channel 1a. Mol Pharmacol 80: 796–808.
21. Riggs P (2001) Expression and purification of maltose-binding protein fusions. In: Ausubel FM, editor. Current Protocols in Molecular Biology.Beverly Massachusetts, USA John Wiley & Sons, Inc., Unit 16 .16.
22. Waugh DS (2005) Making the most of affinity tags. Trends Biotech 23: 316–320.
23. Smith DB, Johnson KS (1988) Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase. Gene 67: 31–40.
24. Guan KL, Dixon JE (1991) Eukaryotic proteins expressed in Escherichia coli: an improved thrombin cleavage and purification procedure of fusion proteins with glutathione S-transferase. Anal Biochem 192: 262–267.
25. di Guan C, Li P, Riggs PD, Inouye H (1988) Vectors that facilitate the expression and purification of foreign peptides in Escherichia coli by fusion to maltose-binding protein. Gene 67: 21–30.
26. Kapust RB, Waugh DS (1999) Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. Protein Sci 8: 1668–1674.
27. Davis GD, Elisee C, Newham DM, Harrison RG (1999) New fusion protein systems designed to give soluble expression in Escherichia coli. Biotech Bioeng 65: 382–388.
28. De Marco V, Stier G, Blandin S, de Marco A (2004) The solubility and stability of recombinant proteins are increased by their fusion to NusA. Biochem Biophys Res Commun 322: 766–771.
29. Einhauer A, Jungbauer A (2001) The FLAG peptide, a versatile fusion tag for the purification of recombinant proteins. J Biochem Biophys Methods 49: 455–465.
30. Schatz PJ (1993) Use of peptide libraries to map the substrate specificity of a peptide-modifying enzyme: a 13 residue consensus peptide specifies biotinylation in Escherichia coli. Biotechnology (NY) 11: 1138–1143.
31. Gaberc-Porekar V, Menart V (2001) Perspectives of immobilized-metal affinity chromatography. J Biochem Biophys Methods 49: 335–360.
32. Schmidt TG, Skerra A (1994) One-step affinity purification of bacterially produced proteins by means of the "Strep tag" and immobilized recombinant core streptavidin. J Chromatogr A 676: 337–345.

33. Schmidt TG, Skerra A (2007) The Strep-tag system for one-step purification and high-affinity detection or capturing of proteins. Nat Protocols 2: 1528–1535.

34. Zhang YB, Howitt J, McCorkle S, Lawrence P, Springer K, et al. (2004) Protein aggregation during overexpression limited by peptide extensions with large net negative charge. Protein Expr Purif 36: 207–216.

35. Stofko-Hahn RE, Carr DW, Scott JD (1992) A single step purification for recombinant proteins. Characterization of a microtubule associated protein (MAP 2) fragment which associates with the type II cAMP-dependent protein kinase. FEBS Lett 302: 274–278.

36. Zheng CF, Simcox T, Xu L, Vaillancourt P (1997) A new expression vector for high level protein production, one step purification and direct isotopic labeling of calmodulin-binding peptide fusion proteins. Gene 186: 55–60.

37. Dyson M, Shadbolt SP, Vincent K, Perera R, McCafferty J (2004) Production of soluble mammalian proteins in Escherichia coli: identification of protein features that correlate with successful expression. BMC Biotechnol 4: 32.

38. Young CL, Britton ZT, Robinson AS (2012) Recombinant protein expression and purification: A comprehensive review of affinity tags and microbial applications. Biotechnol J 7: 620–634.

39. Majerle A, Kidrič J, Jerala R (2000) Production of stable isotope enriched antimicrobial peptides in Escherichia coli: An application to the production of a ^{15}N-enriched fragment of lactoferrin. J Biomol NMR 18: 145–151.

40. Kapust RB, Tozser J, Copeland TD, Waugh DS (2002) The P1' specificity of tobacco etch virus protease. Biochem Biophys Res Commun 294: 949–955.

41. Kurz M, Cowieson NP, Robin G, Hume DA, Martin JL, et al. (2006) Incorporating a TEV cleavage site reduces the solubility of nine recombinant mouse proteins. Protein Expr Purif 50: 68–73.

42. Walker PA, Leong LE, Ng PW, Tan SH, Waller S, et al. (1994) Efficient and rapid affinity purification of proteins using recombinant fusion proteases. Biotechnology (NY) 12: 601–605.

43. Cordingley MG, Callahan PL, Sardana VV, Garsky VM, Colonno RJ (1990) Substrate requirements of human rhinovirus 3C protease for peptide cleavage in vitro. J Biol Chem 265: 9062–9065.

44. Waugh DS (2011) An overview of enzymatic reagents for the removal of affinity tags. Protein Expr Purif 80: 283–293.

45. Jenny RJ, Mann KG, Lundblad RL (2003) A critical review of the methods for cleavage of fusion proteins with thrombin and factor Xa. Protein Expr Purif 31: 1–11.

46. Forsberg G, Baastrup B, Rondahl H, Holmgren E, Pohl G, et al. (1992) An evaluation of different enzymatic cleavage methods for recombinant fusion proteins, applied on des(1-3)insulin-like growth factor I.J Prot Chem 11: 201–211.

47. Maertens B, Spriestersbach A, von Groll U, Roth U, Kubicek J, et al. (2010) Gene optimization mechanisms: a multi-gene study reveals a high success rate of full-length human proteins expressed in Escherichia coli. Protein Sci 19: 1312–1326.

48. Cabrita LD, Dai W, Bottomley SP (2006) A family of E. coli expression vectors for laboratory scale and high throughput soluble protein production. BMC Biotechnol 6: 12.

49. Bassford PJ, Jr. (1990) Export of the periplasmic maltose-binding protein of Escherichia coli. J Bioenerg Biomembr 22: 401–439.

50. Schein CH (1990) Solubility as a function of protein structure and solvent components. Biotechnology (NY) 8: 308–317.

51. Kiefhaber T, Rudolph R, Kohler HH, Buchner J (1991) Protein aggregation in vitro and in vivo: a quantitative model of the kinetic competition between folding and aggregation. Biotechnology (NY) 9: 825–829.

52. Chesshyre JA, Hipkiss AR (1989) Low temperatures stabilize interferon α-2 against proteolysis in Methylophilus methylotrophus and Escherichia coli. Appl Microbiol Biotech 31: 158–162.

53. Mertens N, Remaut E, Fiers W (1995) Tight transcriptional control mechanism ensures stable high-level expression from T7 promoter-based expression plasmids. Biotechnology (NY) 13: 175–179.

54. Sivashanmugam A, Murray V, Cui C, Zhang Y, Wang J, et al. (2009) Practical protocols for production of very high yields of recombinant proteins using Escherichia coli. Protein Sci 18: 936–948.

55. Marley J, Lu M, Bracken C (2001) A method for efficient isotopic labeling of recombinant proteins. J Biomol NMR 20: 71–75.

56. Tyler RC, Sreenath HK, Singh S, Aceti DJ, Bingman CA, et al. (2005) Auto-induction medium for the production of [U-^{15}N]- and [U-^{13}C, U-^{15}N]-labeled proteins for NMR screening and structure determination. Protein Expr Purif 40: 268–278.

57. Reilly D, Fairbrother WJ (1994) A novel isotope labeling protocol for bacterially expressed proteins. J Biomol NMR 4: 459–462.

58. Jansson M, Li YC, Jendeberg L, Anderson S, Montelione GT, et al. (1996) High-level production of uniformly ^{15}N- and ^{13}C-enriched fusion proteins in Escherichia coli. J Biomol NMR 7: 131–141.

59. Kainosho M (1997) Isotope labelling of macromolecules for structural determinations. Nat Struct Biol 4: 858–861.

60. Bremer H (1975) Parameters affecting the rate of synthesis of ribosomes and RNA polymerase in bacteria. J Theoret Biol 53: 115–124.

61. Berthold DA, Jeisy VJ, Sasser TL, Shea JJ, Fredricks HL, et al. (2012) Application Note 15: Top ten tips for producing ^{13}C ^{15}N protein in abundance. Andover, MA, , USA:Cambridge Isotope Laboratories.

62. Schaefer SA, Dong M, Rubenstein RP, Wilkie WA, Bahnson BJ, et al. (2013) ^{77}Se enrichment of proteins expands the biological NMR toolbox. J Mol Biol 425: 222–231.

63. Mobli M, de Araujo AD, Lambert LK, Pierens GK, Windley MJ, et al. (2009) Direct visualization of disulfide bonds through diselenide proxies using ^{77}Se NMR spectroscopy. Angew Chem Int Ed 48: 9312–9314.

64. Neu HC, Heppel LA (1965) The release of enzymes from Escherichia coli by osmotic shock and during the formation of spheroplasts. J Biol Chem 240: 3685–3692.

65. Berrow NS, Bussow K, Coutard B, Diprose J, Ekberg M, et al. (2006) Recombinant protein expression and solubility screening in Escherichia coli: a comparative study. Acta Crystallogr D 62: 1218–1226.

66. Bell PA (2001) E. coli expression systems In: Gerstein AS, editor. Molecular Biology Problem Solver: A laboratory Guide Wiley-Liss, Inc. pp. 461–490.

67. Listwan P, Pedelacq JD, Lockard M, Bell C, Terwilliger TC, et al. (2010) The optimization of in vitro high-throughput chemical lysis of Escherichia coli. Application to ACP domain of the polyketide synthase ppsC from Mycobacterium tuberculosis. J Struct Funct Genomics 11: 41–49.

68. Goldberg S (2008) Mechanical/physical methods of cell disruption and tissue homogenization. Methods Mol Biol 424: 3–22.

69. Maina CV, Riggs PD, Grandea AG, Slatko BE, Moran LS, et al. (1988) An Escherichia coli vector to express and purify foreign proteins by fusion to and separation from maltose-binding protein. Gene 74: 365–373.

70. Gaberc-Porekar V, Menart V (2005) Potential for using histidine tags in purification of proteins at large scale. Chem Eng Tech 28: 1306–1314.

71. Waugh DS (2010) TEV protease FAQ. Macromolecular Crystallography Laboratory, USA, National Cancer Institute. (http://mcl1.ncifcrf.gov/waugh_tech/faq/tev.pdf accessed 20130220).

72. Schafer FQ, Buettner GR (2001) Redox environment of the cell as viewed through the redox state of the glutathione disulfide/glutathione couple. Free Radical Biol Med 30: 1191–1212.

73. Jocelyn PC (1967) The standard redox potential of cysteine-cystine from the thiol-disulphide exchange reaction with glutathione and lipoic acid. Eur J Biochem 2: 327–331.

74. QIAGEN (2003) The QIAexpressionist: a handook for high-level expression and purification of 6xHis-tagged proteins.

75. Ozols J (1990) Amino acid analysis. Methods Enzymol 182: 587–601.

76. Rutherfurd SM, Dunn BM (2011) Quantitative amino acid analysis In: Coligan JE, editor. Current Protocols in Protein Science John Wiley & Sons, Inc., Unit 3.2.

77. Aitken A, Learmonth M (2002) Protein determination by UV absorption. In: Walker JM, editor. The Protein Protocols Handbook:Humana Press. pp. 3–6.

78. Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem 72: 248–254.

79. Desjardins P, Hansen JB, Allen M (2009) Microvolume spectrophotometric and fluorometric determination of protein concentration In: Coligan JE, editor. Current Protocols in Protein Science John Wiley & Sons, Inc.

80. Lowry OH, Rosebrough NJ, Farr AL, Randall RJ (1951) Protein measurement with the Folin phenol reagent. J Biol Chem 193: 265–275.

81. Kelly AE, Ou HD, Withers R, Dötsch V (2002) Low-conductivity buffers for high-sensitivity NMR measurements. J Am Chem Soc 124: 12013–12019.

82. Robosky LC, Reily MD, Avizonis D (2007) Improving NMR sensitivity by use of salt-tolerant cryogenically cooled probes. Anal Bioanal Chem 387: 529–532.

83. Anglister J, Grzesiek S, Ren H, Klee CB, Bax A (1993) Isotope-edited multidimensional NMR of calcineurin B in the presence of the non-deuterated detergent CHAPS. J Biomol NMR 3: 121–126.

84. Dingley AJ, Mackay JP, Chapman BE, Morris MB, Kuchel PW, et al. (1995) Measuring protein self-association using pulsed-field gradient NMR spectroscopy: application to myosin light chain 2. J Biomol NMR 6: 321–328.

85. Golovanov AP, Hautbergue GM, Wilson SA, Lian L-Y (2004) A simple method for improving protein solubility and long-term stability. J Am Chem Soc 126: 8933–8939.

86. Robson SA, King GF (2006) Domain architecture and structure of the bacterial cell division protein DivIB. Proc Natl Acad Sci USA 103: 6700–6705.

87. Otting G, Liepinsh E, Wüthrich K (1993) Disulfide bond isomerization in BPTI and BPTI(G36S): an NMR study of correlated mobility in proteins. Biochemistry 32: 3571–3582.

88. Fang L, Jia KZ, Tang YL, Ma DY, Yu M, et al. (2007) An improved strategy for high-level production of TEV protease in Escherichia coli and its purification and characterization. Protein Expr Purif 51: 102–109.

89. Lycklama a Nijeholt JA, Driessen AJM (2012) The bacterial Sec-translocase: structure and mechanism. Philos Trans Roy Soc Lond – Series B, Biol Sci 367: 1016–1028.

90. Lipkin A, Kozlov S, Nosyreva E, Blake A, Windass JD, et al. (2002) Novel insecticidal toxins from the venom of the spider Segestria florentina. Toxicon 40: 125–130.

91. Bende N, Kang E, Herzig V, Bosmans F, Nicholson GM, et al. (2013) The insecticidal toxin Aps III is an atypical knottin peptide that potently blocks insect voltage-gated sodium channels. Biochem Pharmacol, in press.

92. Bohlen CJ, Priel A, Zhou S, King D, Siemens J, et al. (2010) A bivalent tarantula toxin activates the capsaicin receptor, TRPV1, by targeting the outer pore domain. Cell 141: 834–845.

93. Hardy MC, King GF (2012) Pest-controlling agents and uses therefor. International Patent Application.

94. Johnson JH, Bloomquist JR, Krapcho KJ, Kral RM, Jr., Trovato R, et al. (1998) Novel insecticidal peptides from *Tegenaria agrestis* spider venom may have a direct effect on the insect central nervous system. Arch Insect Biochem Physiol 38: 19–31.

95. Skinner WS, Dennis PA, Li JP, Quistad GB (1992) Identification of insecticidal peptides from venom of the trap-door spider, *Aptostichus schlingeri* (Ctenizidae). Toxicon 30: 1043–1050.

96. Quistad GB, Skinner WS (1994) Isolation and sequencing of insecticidal peptides from the primitive hunting spider, *Plectreurys tristis* (Simon). J Biol Chem 269: 11098–11101.

97. Choi SJ, Parent R, Guillaume C, Deregnaucourt C, Delarbre C, et al. (2004) Isolation and characterization of Psalmopeotoxin I and II: two novel antimalarial peptides from the venom of the tarantula *Psalmopoeus cambridgei*. FEBS Lett 572: 109–117.

98. Bosmans F, Rash L, Zhu S, Diochot S, Lazdunski M, et al. (2006) Four novel tarantula toxins as selective modulators of voltage-gated sodium channel subtypes. Mol Pharmacol 69: 419–429.

99. Middleton RE, Warren VA, Kraus RL, Hwang JC, Liu CJ, et al. (2002) Two tarantula peptides inhibit activation of multiple sodium channels. Biochemistry 41: 14734–14747.

100. Jung HJ, Lee JY, Kim SH, Eu YJ, Shin SY, et al. (2005) Solution structure and lipid membrane partitioning of VSTx1, an inhibitor of the KvAP potassium channel. Biochemistry 44: 6015–6023.

101. Ebbinghaus J, Legros C, Nolting A, Guette C, Celerier ML, et al. (2004) Modulation of Kv4.2 channels by a peptide isolated from the venom of the giant bird-eating tarantula *Theraphosa leblondi*. Toxicon 43: 923–932.

102. Krapcho KJ, Kral RM, Jr., Vanwagenen BC, Eppler KG, Morgan TK (1995) Characterization and cloning of insecticidal peptides from the primitive weaving spider *Diguetia canities*. Insect Biochem Mol Biol 25: 991–1000.

103. Xiao YC, Liang SP (2003) Purification and characterization of Hainantoxin-V, a tetrodotoxin-sensitive sodium channel inhibitor from the venom of the spider *Selenocosmia hainana*. Toxicon 41: 643–650.

104. Xiao Y, Liang S (2003) Inhibition of neuronal tetrodotoxin-sensitive Na$^+$ channels by two spider toxins: hainantoxin-III and hainantoxin-IV. Eur J Pharmacol 477: 1–7.

105. Li D, Xiao Y, Hu W, Xie J, Bosmans F, et al. (2003) Function and solution structure of hainantoxin-I, a novel insect sodium channel inhibitor from the Chinese bird spider *Selenocosmia hainana*. FEBS Lett 555: 616–622.

106. Siemens J, Zhou S, Piskorowski R, Nikai T, Lumpkin EA, et al. (2006) Spider toxins activate the capsaicin receptor to produce inflammatory pain. Nature 444: 208–212.

107. Cordeiro Mdo N, de Figueiredo SG, Valentim Ado C, Diniz CR, von Eickstedt VR, et al. (1993) Purification and amino acid sequences of six Tx3 type neurotoxins from the venom of the Brazilian 'armed' spider *Phoneutria nigriventer* (Keys). Toxicon 31: 35–42.

108. Maertens C, Cuypers E, Amininasab M, Jalali A, Vatanpour H, et al. (2006) Potent modulation of the voltage-gated sodium channel Nav1.7 by OD1, a toxin from the scorpion *Odonthobuthus doriae*. Mol Pharmacol 70: 405–414.

109. Dutertre S, Ulens C, Buttner R, Fish A, van Elk R, et al. (2007) AChBP-targeted α-conotoxin correlates distinct binding orientations with nAChR subtype selectivity. EMBO J 26: 3858–3867.

110. Bulaj G, DeLaCruz R, Azimi-Zonooz A, West P, Watkins M, et al. (2001) δ-conotoxin structure/function through a cladistic analysis. Biochemistry 40: 13201–13208.

111. Yang S, Liu Z, Xiao Y, Li Y, Rong M, et al. (2012) Chemical punch packed in venoms makes centipedes excellent predators. Mol Cell Proteomics 11: 640–650.