



**HAL**  
open science

## **A Venomics Approach Coupled to High-Throughput Toxin Production Strategies Identifies the First Venom-Derived Melanocortin Receptor Agonists**

Steve Reynaud, Justyna Ciolek, Michel Degueldre, Natalie Saez, Ana Filipa Sequeira, Yoan Duhoo, Joana Brás, Herve Meudal, Miguel Cabo Díez, Victoria Fernández Pedrosa, et al.

### ► To cite this version:

Steve Reynaud, Justyna Ciolek, Michel Degueldre, Natalie Saez, Ana Filipa Sequeira, et al.. A Venomics Approach Coupled to High-Throughput Toxin Production Strategies Identifies the First Venom-Derived Melanocortin Receptor Agonists. *Journal of Medicinal Chemistry*, 2020, 63 (15), pp.8250-8264. 10.1021/acs.jmedchem.0c00485 . hal-02936788

**HAL Id: hal-02936788**

**<https://hal.science/hal-02936788>**

Submitted on 20 Nov 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## A Venomics approach coupled to high-throughput toxin production strategies identifies the first venom-derived melanocortin receptor agonists

Steve Reynaud, Justyna Ciolek, Michel Degueldre, Natalie J Saez, Ana Filipa Sequeira, Yoan Duhoo, Joana L A Brás, Herve Meudal, Miguel Cabo Diez, Victoria Fernández Pedrosa, Marion Verdenaud, Julia Boeri, Oscar Pereira Ramos, Frédéric Ducancel, Margot Vanden Driessche, Rudy Fourmy, Aude Violette, Gregory Upert, Gilles Mourier, Annette G. Beck-Sickinger, Karin Mörl, Céline Landon, Carlos M. G. A. Fontes, Rebeca Miñambres Herráiz, Ricardo Canek Rodríguez de la Vega, Steve Peigneur, Jan Tytgat, Loïc Quinton, Edwin De Pauw, Renaud Vincentelli, Denis Servent, and Nicolas Gilles

*J. Med. Chem.*, **Just Accepted Manuscript** • DOI: 10.1021/acs.jmedchem.0c00485 • Publication Date (Web): 30 Jun 2020

Downloaded from [pubs.acs.org](https://pubs.acs.org) on July 1, 2020

### Just Accepted

“Just Accepted” manuscripts have been peer-reviewed and accepted for publication. They are posted online prior to technical editing, formatting for publication and author proofing. The American Chemical Society provides “Just Accepted” as a service to the research community to expedite the dissemination of scientific material as soon as possible after acceptance. “Just Accepted” manuscripts appear in full in PDF format accompanied by an HTML abstract. “Just Accepted” manuscripts have been fully peer reviewed, but should not be considered the official version of record. They are citable by the Digital Object Identifier (DOI®). “Just Accepted” is an optional service offered to authors. Therefore, the “Just Accepted” Web site may not include all articles that will be published in the journal. After a manuscript is technically edited and formatted, it will be removed from the “Just Accepted” Web site and published as an ASAP article. Note that technical editing may introduce minor changes to the manuscript text and/or graphics which could affect content, and all legal disclaimers and ethical guidelines that apply to the journal pertain. ACS cannot be held responsible for errors or consequences arising from the use of information contained in these “Just Accepted” manuscripts.

	<p>Medicina Veterinária; NZYTech Lda, Genes &amp; Enzymes Miñambres Herráiz, Rebeca ; Sistemas Genómicos Ltd, Next-Generation Sequencing Laboratory Rodríguez de la Vega, Ricardo; Universite Paris-Sud, Ecologie, Systematique et Evolution Peigneur, Steve; KU Leuven, Toxicology and Pharmacology Tytgat, Jan ; KU Leuven, Toxicology and Pharmacology Quinton, Loïc; Université de Liège, Mass Spectrometry Laboratory De Pauw, Edwin; Université de Liège, Mass Spectrometry Laboratory Vincentelli, Renaud; Centre National de la Recherche Scientifique Servent, Denis; Université Paris Saclay, CEA, Département Médicaments et Technologies pour la Santé Gilles, Nicolas; Université Paris Saclay, CEA, Département Médicaments et Technologies pour la Santé</p>

SCHOLARONE™  
Manuscripts

# A Venomics approach coupled to high-throughput toxin production strategies identifies the first venom-derived melanocortin receptor agonists

Steve Reynaud<sup>†,‡</sup>, Justyna Ciolek<sup>‡</sup>, Michel Degueldre<sup>‡,ε</sup>, Natalie J. Saez<sup>ϑ,ψ</sup>, Ana Filipa Sequeira<sup>ϕ,ζ</sup>, Yoan Duhoo<sup>ϑ,ψ</sup>, Joana L. A. Brás<sup>ϕ,ζ</sup>, Hervé Meudal<sup>¶</sup>, Miguel Cabo Díez<sup>χ</sup>, Victoria Fernández Pedrosa<sup>χ</sup>, Marion Verdenaud<sup>‡</sup>, Julia Boeri<sup>‡</sup>, Oscar Pereira Ramos<sup>‡</sup>, Frédéric Ducancel<sup>‡</sup>, Margot Vanden Driessche<sup>‡</sup>, Rudy Fourmy<sup>δ</sup>, Aude Violette<sup>δ</sup>, Grégory Upert<sup>‡</sup>, Gilles Mourier<sup>‡</sup>, Annette G. Beck-Sickinger<sup>‡</sup>, Karin Mörl<sup>‡</sup>, Céline Landon<sup>¶</sup>, Carlos M. G. A. Fontes<sup>ϕ,ζ</sup>, Rebeca Miñambres Herráiz<sup>χ</sup>, Ricardo C. Rodríguez de la Vega<sup>§</sup>, Steve Peigneur<sup>θ</sup>, Jan Tytgat<sup>θ</sup>, Loïc Quinton<sup>‡</sup>, Edwin De Pauw<sup>‡</sup>, Renaud Vincentelli<sup>ϑ</sup>, Denis Servent<sup>‡</sup>, Nicolas Gilles<sup>‡\*</sup>

<sup>†</sup>Université Paris-Sud, 15 Rue Georges Clemenceau, Orsay, FR 91405.

<sup>‡</sup>Université Paris Saclay, CEA, Département Médicaments et Technologies pour la Santé, SIMoS, 91191 Gif-sur-Yvette, France.

<sup>¶</sup>Université de Liège, Mass Spectrometry Laboratory, Allée du six Aout 11, Quartier Agora, Liege, BE 4000.

<sup>ε</sup>UCB, Department of Analytical Science Biologicals, Chemin du Foriest, Braine L'Alleud, BE 1420.

<sup>ϑ</sup>Centre National de la Recherche Scientifique, Architecture et Fonction des Macromolécules Biologiques, Campus de Luminy, Marseille, FR 13288.

<sup>ψ</sup>The University of Queensland, Institute for Molecular Bioscience, Institute for Molecular Bioscience, The University of Queensland, St Lucia, QLD, AUS 4072.

<sup>ϕ</sup>Universidade de Lisboa, CIISA - Faculdade de Medicina Veterinária, Avenida da Universidade Técnica, Lisboa, PT 1300-477.

<sup>ζ</sup>NZYTEch Lda, Genes & Enzymes, Estrada do Paço do Lumiar, Campus do Lumiar, Edifício E - R/C, Lisboa, PT 1649-038.

<sup>χ</sup>Sistemas Genómicos Ltd, Next-Generation Sequencing Laboratory, Ronda de Guglielmo Marconi, 6, Paterna, ES 46980.

1  
2  
3 ‡ Université Paris Saclay, CEA, Département IDMIT, 18 route du Panorama, 92265 Fontenay-aux-Roses, France.

4  
5 † Alphabiotoxine Laboratory sprl, Barberie 15, Montreuil-au-bois, BE 7911.

6  
7 ‡ Universität Leipzig, Institute of Biochemistry, Leipzig, DE 04103.

8  
9 ¶ Centre National de la Recherche Scientifique, Centre de Biophysique Moléculaire, rue Charles Sadron, CS 80054  
10  
11 Orléans, FR 45071.

12  
13 § Université Paris-Sud, Ecologie, Systematique et Evolution, 15 rue Georges Clémenceau, Orsay, FR 91405.

14  
15 ¶ University of Leuven (KU Leuven), Toxicology and Pharmacology, Herestraat 49, Leuven, BE 3000.  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**ABSTRACT:** Animal venoms are rich in hundreds of toxins with extraordinary biological activities. Their exploitation is difficult due to their complexity and the small quantities of venom available from most venomous species. We developed a Venomics approach combining transcriptomic and proteomic characterization of 191 species and identified 20,206 venom toxin sequences. Two complementary production strategies based on solid-phase synthesis and recombinant expression in *E. coli* generated a physical bank of 3,597 toxins. Screened on hMC4R, this bank gave an incredible hit rate of 8%. Here, we focus on two novel toxins: N-TRTX-Preg1a, exhibiting an inhibitory cystine knot (ICK) motif, and N-BUTX-Ptr1a, a short scorpion-CS $\alpha$  $\beta$  structure. Neither N-TRTX-Preg1a nor N-BUTX-Ptr1a affect ion channels, the known targets of their toxin scaffolds, but bind to four melanocortin receptors with low micromolar affinities and activate the hMC1R/Gs pathway. Phylogenetically, these two toxins form new groups within their respective families and represent novel hMC1R agonists, structurally unrelated to the natural agonists.

## Introduction

It is estimated that more than 200,000 venomous species exist on Earth, found in almost all animal subgroups across a range of diverse habitats<sup>1</sup>. Many are terrestrial, such as some vertebrates (e.g., snakes, platypus, lizards) or arthropods (e.g., scorpions, spiders, centipedes, ticks, ants), some marine, such as cnidarians (e.g., sea anemones and jellyfish) or some mollusks (e.g., cuttlefish, octopus, marine cone snails, terebrids, squids), and some aerial including insects (e.g., mosquitoes, bugs, wasps, bees) and vampire bats. Venoms are used for defense and/or predation and are composed, among other components, of peptide toxins. Toxins are small proteins up to one hundred residues in size, reticulated by disulfide bridges. Number of toxins present in venoms depends of the species and of the method used to identify them. In the 1990's, publications mentioned few tens of toxins in *Conus* venom. Recently, almost 6,000 toxins were reported in the species *Conus lividus*<sup>2</sup>. All in all, we can estimate that at least 40 million toxins exist on the planet and of these, less than 5,000 have been

1  
2  
3 pharmacologically characterized. Despite this very small number, a plethora of activities have been  
4 reported. Toxins interacting with membrane receptors like voltage- and ligand-gated ion channels<sup>3,4</sup>  
5 and G-Protein Coupled receptors (GPCR)<sup>5-10</sup> have been frequently described whereas other toxins  
6 inhibit enzymes or bind membranes to exert hemolytic and/or antimicrobial actions<sup>11</sup>. Toxins are very  
7 often selective (with the exception of some lytic toxins) and make excellent pharmacological tools to  
8 better understand the physiological or pathophysiological roles of their targets. In some cases, the  
9 toxins themselves or their derivatives have applied clinical uses, such as exenatide from the beaded  
10 lizard), a treatment against type-1 diabetes, ziconotide from a cone snail, a treatment for chronic pain,  
11 captopril derived from a snake toxin and widely used to treat hypertension, but also some other snake  
12 toxins e.g. batroxobin with clotting properties used to stop bleeding<sup>12</sup>. In addition, more than 20 other  
13 toxins are currently investigated in clinical trials as biopharmaceutical and therapeutic tools.

14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
Venoms are a very promising source of selective bioactive molecules and considerable effort is being  
undertaken to discover new toxins with exploitable activities<sup>13,14</sup>. Classically, toxins are identified by  
bioguidage, i.e. successive cycles of venom deconvolution and screening tests against a given target.  
After defining its sequence, the novel toxin of interest must be produced to allow in-depth  
characterization. This strategy is long, costly and risky and requires large amounts of venoms,  
unavailable for most venomous animals which are very small in size. This strategy is not adapted for  
the vast majority of this fantastic biodiversity.

New sequencing strategies that appeared in the 21<sup>st</sup> century are revolutionizing the toxinological field.  
These technologies have been commonly applied to whole genome sequencing, however, eukaryote  
genomes are typically large with tandem repeats and only few genomes from venomous species were  
reported (several snakes, two scorpions, a handful of spiders and dozens of Hymenopteras) but none  
from cone snails. By contrast, many studies have taken advantage of these same strategies for the  
sequencing of various venom gland transcriptomes. Toxins are produced as precursor proteins by  
venom-apparatus specific cells and undergo several maturation steps before being secreted<sup>15</sup>. The

1  
2  
3 formation of disulfide bonds is the most important post-translational modification (PTM) and their  
4 arrangement is commonly used to classify toxins<sup>16</sup>. Sequenced transcripts are translated into their  
5 cognate amino acid sequences allowing identification of the endoplasmic reticulum signal peptide, the  
6 propeptide and mature peptide regions. While tools like SignalP often accurately predict the  
7 endoplasmic reticulum signal peptide boundary, relying on transcriptomics alone, it is far more  
8 complicated to obtain robust information regarding further processing to the mature peptide due to the  
9 lack of information on most pro-protein convertase specific cleavage sites. Moreover, it is important  
10 to note that using a purely transcriptomic strategy, the physical presence in the venoms of most of the  
11 sequences described in databases is almost never demonstrated. Alternatively, proteomic strategies  
12 study only secreted toxins and provide the opportunity to characterize post-translational modifications  
13 (PTMs). Thus, only the combination of these complementary proteomic and transcriptomic strategies  
14 leads to a higher level of confidence for mature peptide sequences. Consequently, it is expected that  
15 the toxin sequences present in these types of databases will become more and more accurate.

16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33 The last and most complicated challenge remains the biological characterization of these natural  
34 peptides. This step cannot be done by a simple sequence analysis, especially for peptides with a low  
35 sequence similarity with previously characterized toxins. It is necessary to build a physical bank of  
36 peptides ready to be screened. The FP7 Venomics Project is part of the biotechnology program “Small-  
37 to-Medium Enterprise (SME)-targeted research for developing tools and technologies for high-  
38 throughput research”. The total project budget was close to 10 million euros (equivalent to 11 million  
39 USD) shared by eight partners from five countries. The vision associated with the Venomics Project  
40 is not only to investigate the diversity of venom peptides through the development of “omics”  
41 technologies but also to replicate toxin diversity *in vitro* to generate novel peptide banks to be used in  
42 drug discovery programs<sup>17</sup>. This unique toxin bank was generated and screened on G-Protein Coupled  
43 Receptor (GPCR) families for which no toxin modulators have previously been described.



## Results

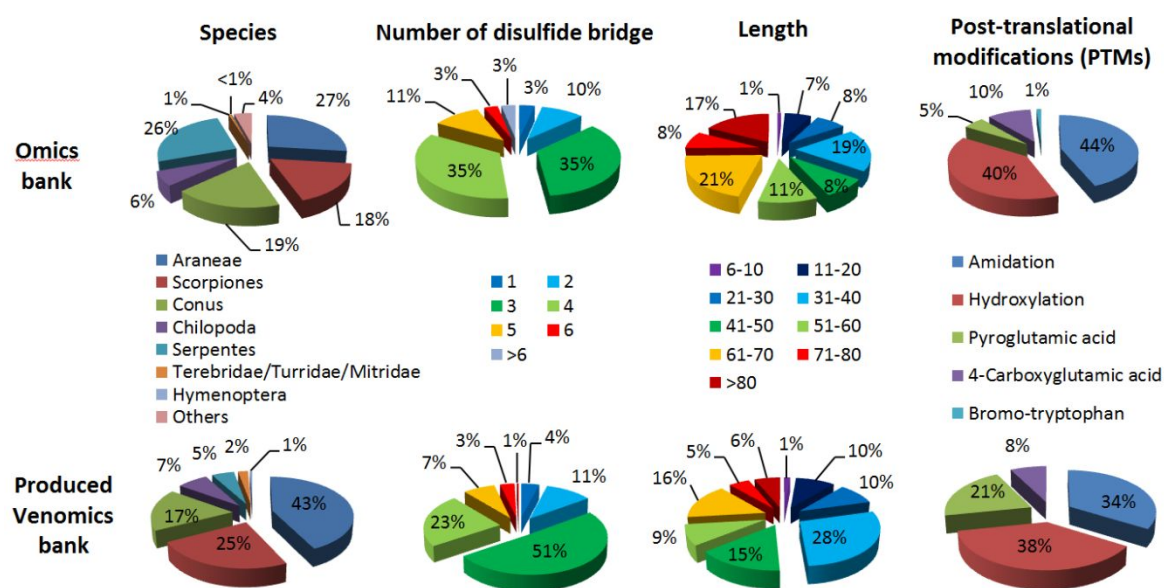
**A large venomous animal collection.** In total, 323 animal samples from 191 different species were gathered (Table 1). Almost every terrestrial and marine venomous clade was represented by at least one species such as ray (*Potamotrygon motoro*) or venomous lizard (*Heloderma horridum exasperatum*). Small terrestrial arthropods represented 95 species *i.e.* half of the sourcing. 55 species (29% of the total sourcing) were studied for the first time, including 44 new species from the most infamous venomous clades (Araneae, Conus, Scorpiones and Serpentes). For instance, 12 previously-unexplored species of Neogastropods were sourced including *Conus canonicus* from the Conidae family, the most studied group of Neogastropoda. The Turridae and Terebridae groups also belong to the Neogastropoda order but have been largely neglected compared to their counterpart<sup>18</sup>. For instance, among over 4,000 described species of Turridae, venom secretions of only twelve species have been studied so far. One of them, *Unedogemmula bisaya* was included in the sourcing. Concerning the Terebridae family, only seven of around 350 described species has been subject of venom analysis. In Venomics sourcing, two of them, *Terebra subulata* and *Terebra guttata*, were chosen. A third species, *Terebra maculata*, never explored before, completed the Terebridae species of the sourcing. Sourcing also included totally unexplored groups such as Mitridae. Two species considered non-venomous, the *Silurus glanis* fish and the *Trachycephalus resinifictrix* frog, but known to produce toxic skin secretions, were also part of the collection. The diverse collection of source animals encompassed various habitats, morphologies and animal sizes ranging from millimeters (*Manica rubida* ant) to several meters (*Ophiophagus hannah* snake).

**Table 1. Selected taxa for sourcing.**

Taxa	Main families	Total N° of species	Never studied species
Neogastropoda	<i>Conidae</i>	38	10
	Other ( <i>Turridae</i> , <i>Terebridae</i> & <i>Mitridae</i> )	5	2
Octopoda	<i>Octopodidae</i>	3	0
Chilopoda	<i>Scolopendridae</i>	6	2
Scorpiones	<i>Buthidae</i> , <i>Scorpionidae</i> , <i>Vaejovidae</i> , <i>Scorpionidae</i> , <i>Hemiscorpiidae</i> & <i>Euscorpiidae</i>	40	7
	<i>Theraphosidae</i> , <i>Ctenidae</i> , <i>Agelenidae</i> , <i>Araneidae</i> , <i>Dipluridae</i> , <i>Lycosidae</i> , <i>Nephilidae</i> , <i>Sparassidae</i> & <i>Theridiidae</i>	41	23
Hymenoptera	<i>Apidae</i> , <i>Vespida</i> , <i>Formicidae</i> & <i>Myrmicinae</i>	8	4
Chondrichthyes	<i>Potamotrygonidae</i>	1	0
Teleostei	<i>Synanceiidae</i> , <i>Plotosidae</i> & <i>Siluridae</i>	8	2
	Serpentes ( <i>Elapidae</i> , <i>Viperidae</i> & <i>Atractaspididae</i> )	39	4
Squamata	Anguimorpha ( <i>Helodermatidae</i> )	1	0
	Anura	<i>Hylidae</i>	1

**A huge sequence toxin diversity in animal venoms.** *De novo* transcriptomics opens the door to the study of Nature without any previous knowledge of it and with a very high resolution, allowing the unbiased identification and quantification of transcripts in a broad spectrum, even for those of low expression. Total RNA was extracted from venom glands. In order to include species diversity, sample requirements were decreased from 2 µg to 100 ng of RNA, being even as low as 60 ng in specific samples. Illumina specific 2x100 pair-end libraries were constructed from each RNA following manufacturer's protocols. Over 60 million reads were obtained per sample using Hiseq 2000 sequencing platform with an average length of 500 nucleotides.

Two proteomic strategies were used according to the size of the toxins<sup>19</sup>. The first one, for venoms presenting most of toxins with less than 25 amino acid residues (like Neogastropoda, Hymenoptera, Anura), analyzed the peptides post-disulfide-reduction by liquid chromatography coupled to tandem mass spectrometry. The second strategy, dedicated to the sequencing of the long venom toxins (> 25 amino acid residues (like Chilopoda, Scorpiones, Araneae, Squamata) required additional alkylation and trypsin digestion (short time and low enzyme concentration) steps after disulfide treatment leading to the generation of overlapping peptides (+/- 15 amino acids)<sup>19</sup>. These resulting peptides were then analyzed with the same mass spectrometry platform. The number of sequences varied between fifty to several hundred, depending on the venom. A high number of sequences are observed in Conidae venoms by proteomics. This diversity relies on the presence of post-translational modifications at different sites in a single toxin and has been discussed in previous publications<sup>2,20</sup>. A total of 21,300,000 sequence tags were obtained.



**Figure 1.** Richness of animal origin and structural features (number of disulfide bridges, length and post-translational modification) of Venomics bank (lower panel) compared to the original Omics bank (upper panel). Omics bank represents 20,206 sequences identified both in transcriptomic analysis of

1  
2  
3 the venom glands and proteomic analysis of the venoms. From it, 3,597 peptides were produced to  
4  
5 build the Venomics bank which closely replicates the origin and structural characteristics of the  
6  
7 primary Omics bank.  
8  
9

10  
11  
12 Proteomic and transcriptomic data generated from the same animal were cross-validated to extract  
13  
14 only sequences present in both the venom and venom gland transcripts, resulting in 20,206 validated  
15  
16 toxin sequences (defined here as peptide up to 110 residues with at least one disulfide bridge). Figure  
17  
18 1 describes these sequences in term of origins, disulfide bridges (1 to 9), sizes (6 to 110) and PTMs.  
19  
20 The quasi-totality of Conus sequences (98%) bears at least one PTM, in addition to disulfide bridges.  
21  
22 Amidation is found in 2,889 sequences (14%) from all groups studied except Terebridae, Turridae,  
23  
24 Mitridae and Teleostei. Hydroxylation on Pro, Met, Lys or Val residues was found in 2,630 sequences  
25  
26 (13%) from various groups with a great majority from Conus (70%). In our study, 4-carboxyglutamic  
27  
28 acid, pyroglutamic acid and bromo-tryptophan appeared to be Conus specific. Strangely, we did not  
29  
30 find pyroglutamic acid in Arachnid venoms (Scorpiones and Araneae), although its presence has been  
31  
32 described previously in many other studies<sup>21,22</sup>. The 20,206 toxin sequences were divided into 137  
33  
34 different cysteine patterns, 97 with an even number of cysteines. Among these 97 patterns, 38 present  
35  
36 new cysteine patterns never described in any venomous species. Regarding the 59 known patterns, C-  
37  
38 C-CC-C-C framework is the most represented comprising 19% of the bank (Table S1). It is related to  
39  
40 Inhibitory Cystine Knot Toxins (ICK), which is known to be particularly abundant in spider, snail and  
41  
42 scorpion venoms<sup>23-26</sup>. For the first time, ICK peptides were identified at protein level from Mitridae,  
43  
44 Turridae, *Manica rubida* ant, Chilopoda and the octopus *Hapalochlaena lunulata*.  
45  
46

47  
48 Interestingly, 874 sequences covering 40 cysteine patterns display an odd number of cysteine residues  
49  
50 (Table S2). Only thirteen were previously described. Two patterns (both lacking vicinal cysteine  
51  
52 residues) are particularly abundant: C-C-C-C-C-C-C with 130 sequences (Scorpiones, Serpentes,  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 Octopoda, Conus) and C-C-C-C-C specific to the Araneae taxon with 179 sequences. We did not  
4  
5 identify any sequences from Conus sharing the latter pattern, despite it has been previously described<sup>27</sup>.  
6  
7 Among these 20,206 sequences, 782 (4%) display a new cysteine pattern and 4,393 (22%) share less  
8  
9 than 50% sequence identity with any described peptide or protein in the UniprotKB/SwissProt database  
10  
11 using default settings (e-value max =1). It must be noted that the cysteine connectivity of the native  
12  
13 toxins can not be explicitly determined using this pipeline.  
14  
15

16  
17 **From sequences to a high throughput screening compatible toxin bank.** In order to allow screening  
18  
19 and characterization of peptide activity, we developed two high-throughput production protocols.  
20  
21 5,923 sequences representing the broad diversity of the sequence database were selected and produced  
22  
23 in the dedicated time of 18 months. Coming from 64 different species, they cover 91 different cysteine  
24  
25 patterns, including 14 with an odd number of cysteine residues (Tables S1 and S2).  
26  
27

28  
29 We selected 931 peptides shorter than 35 residues, with or without PTMs, to be produced by solid-  
30  
31 phase chemical synthesis<sup>28</sup>. The crude synthetic product was cleaved from the resin, desalted and  
32  
33 subjected to an optimized oxidation process, whatever the number or pattern of cysteines<sup>28</sup>. After a  
34  
35 final desalting step, liquid chromatography and mass analysis were performed to validate the presence  
36  
37 of the expected oxidized mass (Figure S1). 893 toxins from 9 to 34 residues with 1 to 3 disulfide  
38  
39 bridges were successfully produced, representing a success rate of 96%.  
40  
41

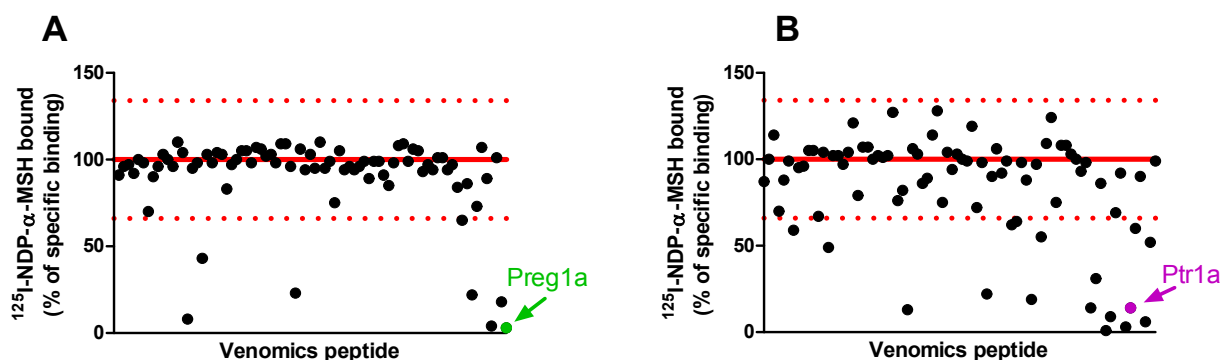
42  
43 The remaining 4,992 sequences containing more than 34 residues were selected for heterologous  
44  
45 expression in *E. coli*<sup>29-31</sup>. All synthetic genes encoding toxin sequences were subcloned into an  
46  
47 optimized periplasmic expression plasmid containing an N-terminal *E. coli* Disulfide bond Isomerase  
48  
49 C (DsbC) chaperone fusion tag, a 6xHIS tag for purification and a tobacco etch virus (TEV) protease  
50  
51 cleavage site to release the native peptide from the fusion protein. Expressed fusion proteins are  
52  
53 purified on a Nickel resin, subjected to overnight TEV cleavage and finally desalted. The correct mass  
54  
55 was confirmed by mass analysis (Figure S2). 2,704 toxins, from 35 to 110 residues with 1 to 9 disulfide  
56  
57 bridges were successfully produced representing a success rate of 54%.  
58  
59  
60

1  
2  
3 Determining disulfide connectivity in native toxins is a challenge and the Venomics project did not  
4 have the resources to carry out this task. In the absence of this information, cysteine residues were  
5 randomly oxidized and the vast majority of toxins presented a unique peak in LC, suggesting that they  
6 adopted only one conformation, hopefully the natural one. In the case where LC profiles presented two  
7 or three peaks with the good expected oxidized mass, we assumed that it could be due to the generation  
8 of peptide isomers<sup>32</sup> (Figures S1 and S2).  
9

10 Together, these two strategies resulted in a bank of 3,597 peptides composed of 71 different cysteine  
11 frameworks (Tables S1 and S2). In addition, 158 toxins (4%) display a novel cysteine pattern and  
12 1,026 (29%) share less than 50% sequence identity with any previously described peptide or protein  
13 in the UniprotKB/SwissProt database. All together, we have successfully translated the high  
14 complexity of the omics bank into a physical toxin bank via high throughput methods.  
15  
16

17 **First animal toxins on melanocortin receptors.** Animal toxins are mostly known for their bioactivity  
18 on ion channels. Few but very interesting toxins active on GPCRs have also been identified<sup>33</sup>. For  
19 example, the commercially available type 2 diabetes treatment, exenatide derived from a venomous  
20 lizard, targets a class A GPCR. Belonging to the same class, human melanocortin 4 receptor (hMC4R)  
21 regulates food intake and its genetic invalidation is responsible of morbid obesity. Screening of the  
22 Venomics bank on hMC4R gave an amazing high hit rate of 8%, 163 coming from chemical synthesis  
23 and 123 from recombinant expression. We decided to further investigate two hits displaying cysteine  
24 patterns related to two of the largest structural families in animal toxins, ICK-related pattern (-C-C-  
25 CC-C-C-) and the CS $\alpha\beta$ -related cysteine patterns (-C-C-C-C-C-K-C-), and primary renowned for their  
26 activities on ion channels. The ICK-related N-TRTX-Preg1a comes from the spider *Poecilotheria*  
27 *regalis* while the CS $\alpha\beta$ -related N-BUTX-Ptr1a from the scorpion *Parabuthus transvaalicus*. N letter  
28 (uppercase nu greek letter) has been chosen as the activity descriptor for modulators of hMCRs. In  
29 primary screening on hMC4R, N-TRTX-Preg1a and N-BUTX-Ptr1a both showed strong inhibition of  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

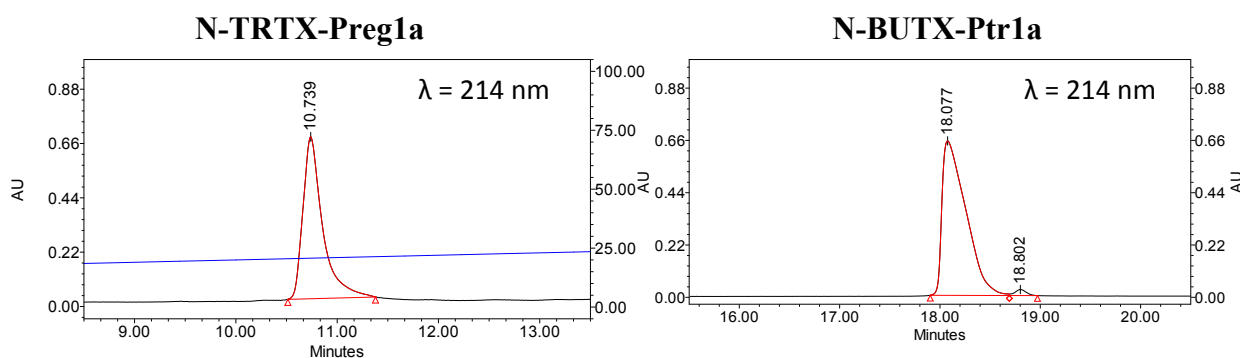
$^{125}\text{I}$ -NDP- $\alpha$ -MSH, 97% and 86% respectively (Figure 2). Further pharmacological investigations were then done with newly synthetically produced toxins of more than 95% purity (Figures 3 and S3).



**Figure 2.** Screening results on hMC4R from the two plates containing N-TRTX-Preg1a (A) and N-BUTX-Ptr1a (B) peptides in the Venomics bank. Screening on hMC4R, stably expressed in CHO cells, was based on competitive binding on the receptor of the tested product against the radio-labeled ligand,  $^{125}\text{I}$ -NDP- $\alpha$ -MSH. Full red line is the median and red dash lines are median plus or minus 3 median absolute deviation. Dots above or under dash lines are hits. Each dot corresponds to a tested Venomics peptide (80 peptides by plate). Coloured dot represents either N-TRTX-Preg1a (in green) or N-BUTX-Ptr1a (in purple). Preg1a: N-TRTX-Preg1a. Ptr1a: N-BUTX-Ptr1a.

N-TRTX-Preg1a and N-BUTX-Ptr1a display low micromolar affinities on membrane preparations of every studied subtype of melanocortin receptors stably expressed in CHO-K1 cells for hMC1R, hMC4R, hMC5R and HEK293 cells for hMC3R (Table 2, Figure 4A). A final check was performed by measuring the affinities of the linear toxins with cysteines alkylated by iodoacetamide treatment. Up to 100  $\mu\text{M}$  (1, 10 and 100  $\mu\text{M}$  concentrations were tested), unstructured N-TRTX-Preg1a and N-BUTX-Ptr1a did not inhibit the binding of  $^{125}\text{I}$ -NDP- $\alpha$ -MSH on hMC1R and hMC4R, demonstrating that the three-dimensional structures of the toxins are essential to their activities (Figure 4A). Their agonistic and antagonistic activities on the signalling pathways hMC1R/Gs and hMC4R/Gs were then assessed on HEK293 cells. As control, it was demonstrated that neither toxin induced any effect on

control hMCR-free HEK293. Both toxins tested at 100  $\mu$ M induced cAMP production on hMC1R, similar to 1 nM of  $\alpha$ -MSH, while no effect was observed on hMC4R (Figure 4B). In the presence of 1 nM of  $\alpha$ -MSH, the toxins slightly enhance hMC1R related cAMP production but did not reduce the cAMP production related to hMC4R (Figure 4B). Both toxins were also tested at 100  $\mu$ M on several human sodium channels (Nav1.1, Nav1.2, Nav1.3, Nav1.4, Nav1.5, Nav1.6, Nav1.7 and Nav1.8), potassium channels (Kv1.1, Kv1.2, Kv1.3, Kv1.4, Kv1.5, Kv1.6, Kv2.1, Kv3.1, Kv4.3, Kv7.1, Kv7.3, Kv10.1, Kv11.1(hERG), GIRK1), a voltage-gated calcium channel (Cav3.3) and ligand-gated ion channels nAChR  $\alpha$ <sub>1</sub> $\beta$ 1 $\gamma$  $\delta$  and nAChR  $\alpha$ 7. No activity could be detected (Figure S4 and S5).

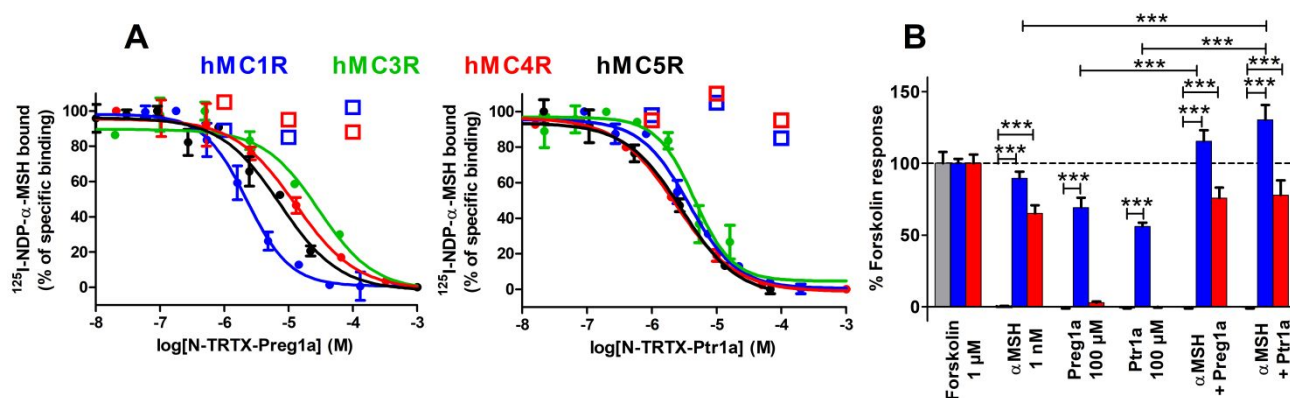


**Figure 3.** Analytical HPLC chromatogram of the final purified products of N-TRTX-Preg1a and N-BUTX-Ptr1a. Between 10 and 15  $\mu$ g of each toxin were analyzed on a X Bridge BEH 300Å 5  $\mu$ m 4.6 x 250 mm with a gradient of 10-50% CH<sub>3</sub>CN/TFA 0.1% in 40 min at a flow rate of 1 mL/min. A spontaneous transformation of the N-terminal glutamine (Q form) residue of the peptide N-BUTX-Ptr1a into a pyroglutamic acid (Z form) can be detected after crude peptide deprotection. According to LC-MS analysis performed on Elute Sp Pump HPG 700 (Bruker) coupled with AmaZon SL LC/MS (Bruker), the peak at 18.077 min corresponds to the Q form (average mass of 3795.70 m/z) and represents 96.5% of the total area. The peak at 18.802 min with an average mass of 3778.70 m/z seems to correspond to the Z form of the peptide and represent 3.5% of the total area.



**Table 2.** Inhibition constant ( $K_i$ ) values of N-TRTX-Preg1a, N-BUTX-Ptr1a on melanocortin receptors, mean ( $\mu\text{M}$ )  $\pm$  SEM (n=3-10).

	hMC1R	hMC3R	hMC4R	hMC5R
N-TRTX-Preg1a	1.8 $\pm$ 0.3	19.8 $\pm$ 3.5	7.1 $\pm$ 1.5	10.0 $\pm$ 4.4
N-BUTX-Ptr1a	2.9 $\pm$ 0.6	3.9 $\pm$ 0.7	2.6 $\pm$ 0.4	2.2 $\pm$ 0.8



**Figure 4.** Pharmacological characterization of N-TRTX-Preg1a and N-BUTX-Ptr1a. (A) Representative curves of binding of  $^{125}\text{I}$ -NDP- $\alpha$ -MSH on membrane preparations of hMC1R, hMC4R and hMC5R (stably expressed in CHO cells) and on hMC3R (stably expressing in HEK293 cells) in the presence of N-TRTX-Preg1a (left panel) or N-BUTX-Ptr1a (right panel). Mean (%)  $\pm$  SD (n=2-3). Open squares represent peptides with alkylated cysteines by iodoacetamide treatment, tested on hMC1R and hMC4R. (B) Functional assays on HEK293 (grey), HEK293 stably expressing hMC1R (blue) or HEK293 transitory expressing hMC4R (red). cAMP production expressed as % of the positive control (1  $\mu\text{M}$  forskolin). Preg1a: N-TRTX-Preg1a. Ptr1a: N-BUTX-Ptr1a. Mean (%)  $\pm$  SEM (n = 3), \*\*\*: p<0.001; one-way analysis of variance (ANOVA) followed by Dunnett's multiple comparison test.

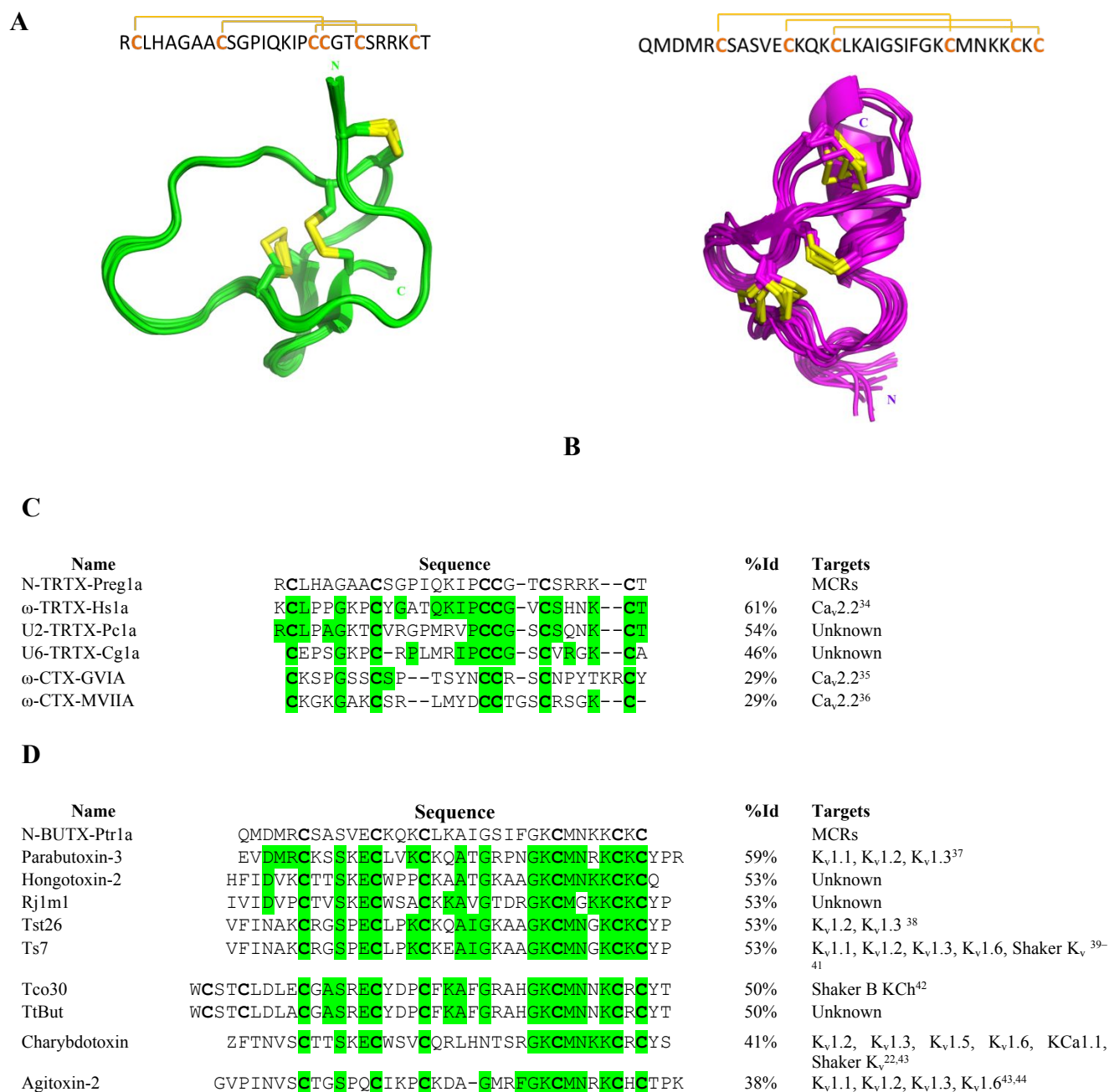
### Are N-TRTX-Preg1a and N-BUTX-Ptr1a real ICK-related and CS $\alpha$ -related scaffold toxins?

The two N-TRTX-Preg1a and N-BUTX-Ptr1a toxins have a cysteine pattern corresponding to the ICK-

1  
2  
3 related and the CS $\alpha$  $\beta$ -related structure with the following pairing C1-C4, C2-C5, C3-C6. As both  
4  
5 toxins display original pharmacological activities, we characterized both structures by NMR and  
6  
7 determined experimentally their cysteine connectivity. Both peptides are well-folded, illustrated by  
8  
9 their good dispersion of resonances. All the  $^1\text{H}$  (except for the methyl groups of Met2 and Met28 of  
10  
11 N-BUTX-Ptr1a), the  $^{15}\text{N}$  chemical shifts and the  $^{13}\text{C}$  resonances of C $\alpha$  and C $\beta$  were assigned.  
12  
13 Depositions in the BioMagResBank were done with the entry code 34117 for N-TRTX-Preg1a and  
14  
15 34118 for N-BUTX-Ptr1a.  
16  
17

18  
19 A set of 546 NOEs for N-TRTX-Preg1a were used in the final ARIA run, which includes 454  
20  
21 unambiguous restraints (Table S3). Additional constraints were introduced during the iterative ARIA  
22  
23 procedure: 38 dihedral angle restraints, 10 H-bond constraints observed within the  $\beta$ -strands and 3  
24  
25 covalent bonds between sulphur atoms (C2-C19, C9-C22, C18-C27). Indeed, when enough  
26  
27 experimental data had been introduced into the first runs of the iterative ARIA process, the C18-C27  
28  
29 disulfide bond is formed in two-thirds of the N-TRTX-Preg1a structures. The fine analysis of NOE  
30  
31 constraints and the evaluation of distance between sulfur atoms allowed discrimination between the  
32  
33 two remaining possibilities: C2-C19, C9-C22, C18-C27 (favorable) or C2-C22, C9-C19, C18-C27  
34  
35 (excluded). Therefore, the corresponding covalent bonds were added between sulfur atoms to refine  
36  
37 the N-TRTX-Preg1a structure in the last runs of the calculations. The 10 models representative of the  
38  
39 solution structure (Figure 5A) display a very low RMSD, small potential energy values, respect all the  
40  
41 experimental data and are in perfect agreement with standard covalent geometry with 64.8% of the  
42  
43 residues in the most favored regions of the Ramachandran plot, and 35.2% in additional allowed  
44  
45 regions (Table S3). The N-TRTX-Preg1a structure is very well defined, organized into the ICK motif  
46  
47 without ambiguities regarding cysteine pairing. It is composed of a short antiparallel  $\beta$ -sheet made up  
48  
49 of two short strands (22–23 and 26–27). The very short turn between these two strands, containing two  
50  
51 arginine residues (R24 and R25), is highly constrained. Note the presence of an extended region (7–  
52  
53 9), which faces the two-stranded  $\beta$ -sheet, resembling a third quasi- $\beta$ -strand. Proline residue 12 is in a  
54  
55  
56  
57  
58  
59  
60

majority trans configuration (about 80%). N-TRTX-Preg1a sequence produced only a handful of significant hits in public databases and shares only 61% identity with the closest toxin with a known activity, Hs1a from *Cyriopagopus schmidti* (Figure 5C), a calcium channel blocker<sup>34</sup>.



**Figure 5.** Structural characterization of N-TRTX-Preg1a and N-BUTX-Ptr1a. Primary structure and overlay of the 10 NMR models representative of the solution structure of N-TRTX-Preg1a (A) and N-

1  
2  
3 BUTX-Ptr1a (B). N and C termini are indicated. Comparison of amino acid sequences between N-  
4 TRTX-Preg1a (C) or N-BUTX-Ptr1a (D) and the most similar toxins obtained with NCBI Protein Blast  
5 program (default settings). Cysteine residues are in bold, conserved residues from N-TRTX-Preg1a or  
6 N-BUTX-Ptr1a are in green. Sequence alignment was performed with Clustal Omega (version 1.2.4  
7 from Emboss programs, EBlosum62 matrix for two pair alignment).  
8  
9  
10  
11  
12  
13  
14  
15  
16

17 Regarding structural determination of N-BUTX-Ptr1a, 839 NOEs (including 558 unambiguous  
18 distance restraints) and 4 hydrogen bonds present within the  $\beta$ -sheet were used in the final ARIA run  
19 (Table S4). 92% of the 250 final structures display a C1-C4, C2-C5, C3-C6 disulfide bridge array,  
20 without adding any constraint on the bonds. The 10 models representative of the solution structure  
21 (Figure 5B) display low root mean square deviations (RMSD), small potential energy values, and  
22 respect all the experimental data, as attested by the absence of residual violations (Table S4). The  
23 Ramachandran plot exhibits 53% of the ( $\phi, \psi$ ) angles in the most favored regions, and 40% in the  
24 additional allowed regions. The N-BUTX-Ptr1a structure scaffold is a well-known motif within  
25 scorpion toxins<sup>45</sup>, plant defensins<sup>46</sup> and insect defensins<sup>47</sup>, and was designated as a cysteine-stabilized  
26  $\alpha$ -helix  $\beta$ -sheet motif (CS $\alpha\beta$  motif)<sup>48</sup>. The structure is composed of an  $\alpha$ -helix (11–19) and a 2-stranded  
27 short  $\beta$ -sheet (26–28; 31–33), stabilized by three disulfide bridges (C6-C27, C12-C32, C16-C34)  
28 unambiguously assigned. N-BUTX-Ptr1a shares the same sequence pattern as the majority of  $\alpha$ -KTx  
29 toxins known to block potassium channels: C-x(4,5)-C-x(3)-C-x(6,9)-[GAS]-K-C-[IMQT]-x(3)-C-x-  
30 C (Prosite [PS01138](http://www.pdb.org)). Yet, N-BUTX-Ptr1a shares only 59% identity with the closest toxin with a known  
31 activity, Parabutoxin-3 from the same species *Parabuthus transvaalicus* (Figure 5D), a potassium  
32 channel blocker<sup>37</sup>. The coordinates of structures have been deposited in the Protein Data Bank  
33 (<http://www.pdb.org>) with the entry codes 6SAA and 6SAB for N-TRTX-Preg1a and N-BUTX-Ptr1a,  
34 respectively.  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 While N-TRTX-Preg1a and N-BUTX-Ptr1a display the canonical ICK and CS $\alpha\beta$  structure,  
4 respectively, their sequences diverged so much from most described toxins that homology searches or  
5 phylogenetic reconstruction failed to assign them into known toxin groups. On one hand, N-TRTX-  
6 Preg1a matched only a handful of spider toxins that could not be linked to the main group of spider  
7 ICK toxins. On the other hand, phylogenetic reconstruction with all known alpha-KTx (UniProt)  
8 placed N-BUTX-Ptr1a as sister to a clade consisting of several monophyletic alpha-KTx subfamilies  
9 (alpha-KTx 2, 4 and 12).  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20

## 21 **Discussion and conclusions**

22  
23 To our knowledge, this is the first time that a thorough ‘-omic’ analysis of venoms has been carried  
24 out on such a wide number and diversity of venomous species from various habitats (128 terrestrial,  
25 55 aquatic and 8 aerial) including two European species (*Manica rubida* ant and *Buthus occitanus*  
26 scorpion). The *Manica rubida* ant was the smallest species of the project, studied for the first time and  
27 35 sequences were validated. The number of validated sequences ranges from 0, due to experimental  
28 failures (some collecting conditions were not compatible with proper mRNA extraction), to 1,163, in  
29 the venom of the Mediterranean scorpion *Buthus occitanus*. Finally, no validated toxin sequences  
30 could be detected in the skin secretions of *Silurus glanis* fish and the *Trachycephalus resinifictrix* frog.  
31 An average of 106 toxin sequences per venom were validated, much less than what could be expected  
32 according to the assumed richness of venoms. This difference highlights the difficulties in obtaining  
33 full and unambiguous toxin sequences. The gotten sequence diversity also highlights our very limited  
34 state of knowledge of the toxin world. Indeed, almost one quarter of the toxins share less than 50%  
35 sequence identity with known toxin sequences, suggesting unpredictable and original biological  
36 activities. Furthermore, 65 new naturally occurring cysteine patterns have been found, suggesting 65  
37 new structural organizations. This is an opportunity to discover peptides with attractive biophysical  
38 properties in terms of engineering and thermodynamic stability for optimal drugability. Another level  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 of complexity, not investigated here, is the dimerization (homo- and/or hetero-) of toxins containing  
4 an odd number of cysteines. A specific proteomic analysis should be designed to address this point.  
5

6  
7 Converting a sequence database into physical peptides is challenging, in particular for the oxidation of  
8 disulfide bridges. Our recombinant expression strategy was able to directly produce oxidized peptide  
9 and has proven to be effective in terms of high throughput production. The two-step chemical synthesis  
10 strategy was also extremely efficient. This is mainly thanks to the use of a universal folding strategy  
11 applied on all toxins regardless of size or number of cysteines. These production strategies provided  
12 peptides with impurities, as no high-efficiency purification step was performed. While the presence of  
13 these contaminations could have caused numerous false positive results during the screening tests<sup>49</sup>  
14 we shown this was not the case as 90% of the tested hits were confirmed.  
15  
16

17 GPCRs constitute one of the largest and most versatile families of receptors. They are involved in a  
18 wide spectrum of physiological functions. Since the late 20<sup>th</sup> century, pharmaceutical companies have  
19 significantly diminished their natural product programs, partly because of the advances in both HTS  
20 and combinatorial synthesis, thereby the creation of large synthetic libraries of small but artificial  
21 molecules. Natural product libraries consist of extracts and partially purified fractions and are  
22 perceived to be incompatible with HTS. Unfortunately, this loss of interest in natural products  
23 correlated with the overall reduction in new leads in the drug development pipeline and the substantial  
24 decline in new drug approval<sup>50</sup>. The Venomics bank reconciles venoms, a particularly difficult  
25 exploitable resource of natural peptides, with HTS. The screening of the Venomics bank provided an  
26 extraordinary hit rate of 3%, much higher than the one reported in literature concerning artificial  
27 molecules and estimated at around 0.05%<sup>51,52</sup>. To our knowledge, this is the first large-scale  
28 demonstration of the power of natural peptide toxins to provide GPCR ligands. This strategy opens up  
29 an unexplored avenue for the discovery of new ligands on receptors linked to diseases with unmet  
30 medical needs such as obesity.  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

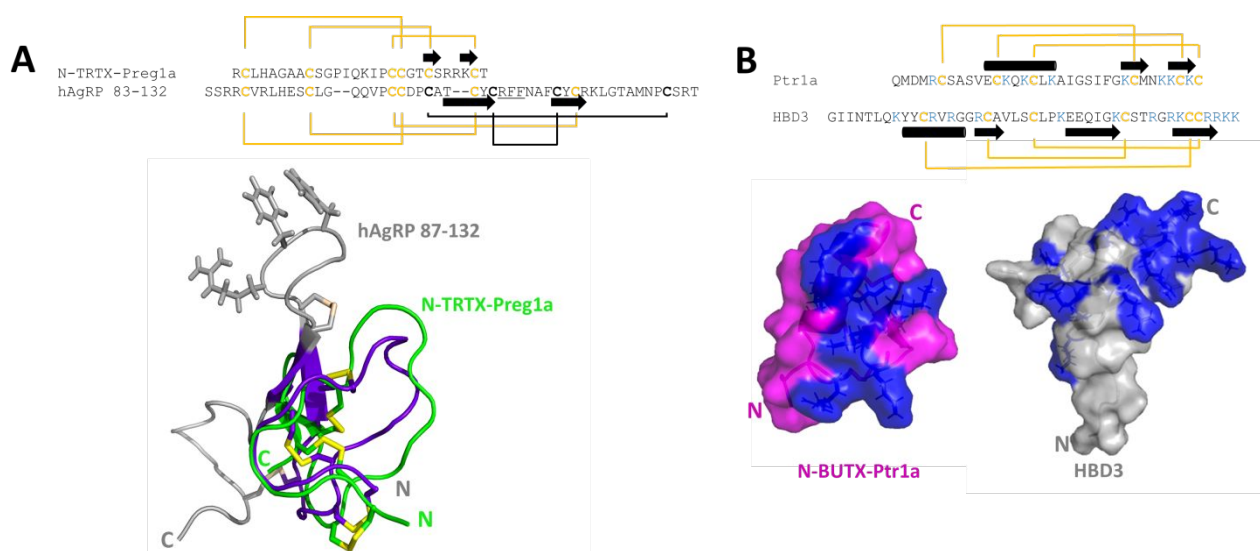
1  
2  
3 MC4R is mainly expressed in the paraventricular nucleus of the hypothalamus. Its activity is finely  
4 regulated by both anorexigenic agonists, the melanocyte stimulating hormones (MSH), and the  
5 orexigenic antagonist, Agouti-related protein (AgRP)<sup>53</sup>. Loss-of-function mutations on hMC4R are the  
6 primary cause of monogenic obesity (<http://www.orpha.net/>) making it a therapeutic target in eating  
7 disorders. Currently, only one MC4R ligand is in clinical trials (Setmelanotide,  
8 <https://clinicaltrials.gov/>). Subtype 1 (MC1R) is also of therapeutic interest with its protective role  
9 against melanoma. It elicits eumelanin production, a dark, UV radiation-absorbing, photoprotective  
10 pigment<sup>54</sup> and promotes nucleotide excision repair of DNA damaged by UV radiation<sup>55</sup>. There is an  
11 urgent need for development of innovative ligands to better understand MCR functions and to generate  
12 new medicines. N-TRTX-Preg1a and N-BUTX-Ptr1a are the first MCR venom-peptide ligands,  
13 structurally unrelated to natural endogenous MCR agonists, while similarities and differences exist  
14 with other natural ligands such as human Agouti-related protein (hAgRP) or human  $\beta$ -defensin 3  
15 (HBD3). As N-TRTX-Preg1a, hAgRP displays an ICK fold in its C-terminal part (hAgRP 83-132)  
16 representing its hypothesized mature form<sup>56</sup> with two additional disulfide bridges. As presented in  
17 Figure 6A, cysteine residues involved in ICK fold in N-TRTX-Preg1a and hAgRP 83-132 are well  
18 aligned with similarly  $\beta$ -sheets involved the two last cysteine residues of the ICK pattern. Composition  
19 and loop sizes are quite divergent between the two peptides especially for the third loop which is  
20 almost two times longer in hAgRP 83-132 (C3-C9). This loop contains the Arg-Phe-Phe  
21 pharmacophore of AgRP<sup>56,57</sup> that is lacking in N-TRTX-Ptr1a. This difference suggests that in spite of  
22 sharing the same structural ICK family, N-TRTX-Preg1a and hAgRP 83-132 have a different mode of  
23 action on MCRs. Regarding N-BUTX-Ptr1a, we noticed that it shared with HBD3 interesting features.  
24 They both display similar secondary structures composed of a short  $\alpha$ -helix and several antiparallel  $\beta$ -  
25 sheets reticulated by three disulfide bridges (Figure 6B). In addition, they are both rich in cationic  
26 residues with high basic pIs of 10 for HBD3 and 9.3 for N-BUTX-Ptr1a. This last remark was of  
27 particular interest as in a previous study, Nix *et al.* demonstrated the role of large patches of positively  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 charged residues in the binding of HBD3 on melanocortin receptors<sup>58</sup>. In the same study, the authors  
4 described nanomolar affinities on melanocortin receptors (42 nM on hMC1R and 110 nM on hMC4R),  
5  
6 but more recently, Ericson *et al.* published micromolar affinities, 7  $\mu$ M on hMC1R and 100  $\mu$ M on  
7  
8 hMC4R<sup>59</sup>. According to Ericson *et al.*, this difference may be explained by a discrepancy between  
9  
10 protocols. We used herein the same methodology as Ericson *et al.* but with a longer incubation time  
11  
12 (3 hours *versus* 1 hour). Without testing HBD3 in our conditions, we cannot be sure that HBD3 and  
13  
14 N-BUTX-Ptr1a share or not similar affinities on melanocortin receptors. Nevertheless, the micromolar  
15  
16 affinity of HBD3 on hMC1R was supported by functional results who described the peptide as a  
17  
18 micromolar agonist of the Gs pathway on hMC1R ( $EC_{50} = 0.40 \mu$ M)<sup>59</sup>, a feature shared with N-BUTX-  
19  
20 Ptr1a. Despite these similarities, important differences between the two peptides have also to be  
21  
22 noticed. Firstly, they do not share global three-dimensional structures, surely linked to the difference  
23  
24 in cysteine pattern (Figure 6B). Secondly, Nix *et al.* showed that the absence of disulfide bridges, by  
25  
26 cysteine replacement in serine in HBD3 sequence, decreased HBD3 affinities by less than 10 times<sup>58</sup>  
27  
28 contrary to the linear N-BUTX-Ptr1, which lost at least 100 times in affinity. Therefore, the structural  
29  
30 constraints induced by disulfide bridges may be more important in N-BUTX-Ptr1a binding than in  
31  
32 HBD3. These two structural differences may be linked in distinct interaction profile of the two peptides  
33  
34 with melanocortin receptors.  
35  
36  
37  
38  
39  
40

41  
42 N-TRTX-Preg1a and N-BUTX-Ptr1a also lack the HFRW pharmacophore, described in endogenous  
43  
44 agonists as the minimal active sequence<sup>62,63</sup>, suggesting that they activate hMC1R in a different way,  
45  
46 that would be interesting to deeply study. In addition, both toxins promote a synergistic effect with  
47  
48 endogenous agonist  $\alpha$ -MSH. This activity deserves to be finely studied as both ligands may activate  
49  
50 hMC1R by a new molecular mechanism. N-TRTX-Preg1a and N-BUTX-Ptr1a inhibit <sup>125</sup>I-NDP- $\alpha$ -  
51  
52 MSH binding on hMC4R but are not able to modulate activate or antagonize the hMC4R/Gs pathway,  
53  
54 under our experimental conditions. This discrepancy can come from our experimental procedure. We  
55  
56 perhaps did not incubate enough time as the toxins display low affinities and may need longer time to  
57  
58  
59  
60



bind the receptor. Indeed, functional tests are designed according to the pharmacological characteristic of the agonist, which is nM. Another possibility is that these toxins display biased functions, with activities on other pathways than the Gs one. A deeper exploration of the functions of the two toxins on various MC4R signaling pathways (Gq,  $\beta$ -arrestin, Kinases, etc.) will be performed to better understand their modes of action.



**Figure 6. Structure comparison of the spider and scorpion toxins with melanocortin receptor structured peptide ligands.** (A) N-TRTX-Pregla *versus* human Agouti-related protein 83-132 (hAgRP 83-132). Upper panel: primary (Needl, version 6.6.0 from Emboss programs, EBlosum62 matrix) and secondary structure (arrow:  $\beta$ -sheet) alignment. Cysteine residues and associated disulfide connectivity are in bold and orange when implied in the ICK motif or in bold and black otherwise. The hAgRP Arg-Phe-Phe is underlined. Lower panel: Overlay of N-TRTX-Pregla and hAgRP 87-132 (grey, 1HYK<sup>60</sup>) NMR structures. Aligned ICK part of hAgRP is represented in colors. The hAgRP pharmacophore residues (Arg-Phe-Phe) have developed side-chains. (B) N-BUTX-Ptr1a (Ptr1a) *versus* human  $\beta$ -defensin 3 (HBD3). Upper panel: primary and secondary structures (arrow:  $\beta$ -sheet, tube:  $\alpha$ -helix). Cysteine residues and associated disulfide connectivity are in orange. Basic residues are in blue. Lower panel: tertiary NMR structure comparison of N-BUTX-Ptr1a and HBD3 (1KJ6<sup>61</sup>). Structures

1  
2  
3 display the solvent exposed surface area with basic residues in blue. “C” and “N” correspond to the N-  
4 and C-termini of the peptides.  
5  
6  
7

8  
9  
10 N-TRTX-Preg1a exhibits a common ICK fold only described so far for activity on ion channels. Its  
11 sequence is close to calcium channel toxins but does not possess the tyrosine residue in position 16  
12 (N-TRTX-Preg1a numbering), described to be essential for blocking  $Ca_v2.2$  channel conductivity<sup>64</sup>.  
13  
14 N-BUTX-Ptr1a, close in sequence to potassium channel toxins, possesses a dyad composed of a basic  
15 and an aromatic residue<sup>65,66</sup> (F24 and K26, N-BUTX-Ptr1a numbering). This dyad, stated as essential  
16 for potassium activity<sup>65,66</sup> but we showed that is clearly not sufficient for N-BUTX-Ptr1a to bind  
17 potassium channels. N-TRTX-Preg1a and N-BUTX-Ptr1a sequences diverge so much within their own  
18 group that it was even impossible to phylogenetically anchor them into known toxin families. Thus,  
19 they constitute new clades for ICK and  $CS\alpha\beta$  families. Several other toxins of the produced bank  
20 sharing more than 80% of sequence identity with these two toxins are present in the Venomics  
21 database, suggesting that N-TRTX-Preg1a and N-BUTX-Ptr1a are not the only members of these new  
22 families. Production and pharmacological characterization of these natural variants should provide  
23 interesting insights with regard to the mode of action of these toxins on hMCRs. This may aid the  
24 generation of ligands with improved affinity and selectivity for development of new hMC1R agonists  
25 with applications in the treatment and/or prevention of melanoma.  
26  
27

28 To our knowledge, the Venomics bank is the largest one ever built. We demonstrated that it represents  
29 a real opportunity to identify ligands with unusual structures and pharmacological properties. Its high  
30 hit rate shows that quality can beat quantity. Two toxins, with classical toxin structures but original  
31 pharmacological activities highlight the diversity of toxin properties and open the door to the  
32 development of new hMCR modulators.  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57

## 58 **Experimental Section**

1  
2  
3 **Sourcing.** We paid important caution to respect the natural resources to which we had access. Thus,  
4  
5 we limited the number of collected specimens to the strict number necessary to ensure the amounts  
6  
7 and volumes of biological samples were compatible with the experiments to be carried out within the  
8  
9 Venomics consortium. For instance, during the Mayotte and Polynesia expeditions, supernumerary  
10  
11 cone snails collected during the day, were systematically dropped back to water. In the case of the  
12  
13 venomous lizard specimen (*Heloderma horridum exasperatum*), the lizard was anaesthetized while a  
14  
15 veterinary surgeon performed surgery to remove only one of the venomous salivary glands of the  
16  
17 animal. The animal fully recovered. This demonstrates the applicability of such a strategy of tissue  
18  
19 extraction that could certainly be applied with success to rare snakes too.  
20  
21  
22

23  
24 **Transcriptomic analysis.** The Illumina paired-end NGS platform was selected to assure sequence and  
25  
26 data quality. 203 RNA samples from different venomous species were sequenced through a low-input  
27  
28 method. Quality and quantity of total RNA was determined using a Bio-analyzer 2100 (Agilent,  
29  
30 California), and Qubit 2.0 (ThermoFisher, Massachussetts). The poly(A)+ mRNA fraction was  
31  
32 isolated from total RNA and cDNA libraries were obtained following Illumina's recommendations.  
33  
34 Briefly, poly(A)+ RNA was isolated on poly-T oligo-attached magnetic beads and chemically  
35  
36 fragmented prior to reverse transcription and cDNA generation. cDNA fragments then went through  
37  
38 an end repair process, the addition of a single 'A' base to the 3' end and then ligation of the adapters.  
39  
40 Finally, the products were purified and enriched with PCR to create the indexed final double stranded  
41  
42 cDNA library. Library quality was analyzed via a Bioanalyzer 2100 High Sensitivity assay and  
43  
44 quantity was determined by real-time PCR using a LightCycler 480 (Roche, Basilea, Suiza). The pool  
45  
46 of libraries was sequenced by paired-end sequencing (100 x 2) on an Illumina HiSeq 2000 sequencer  
47  
48 (Illumina, California). Bioinformatic analysis was a crucial step of the process to identify peptide  
49  
50 sequences. This analysis started with NGS raw data quality control, followed by exhaustive cleaning  
51  
52 of primers, adapters, indexes and low-quality bases. The next step consisted of joining the paired-end  
53  
54 reads and dereplicating all reads before entering the assembly step itself. To assure the quality of the  
55  
56  
57  
58  
59  
60

1  
2  
3 assembly two different assemblers were used; Velvet (v1.2.10)<sup>67</sup> and Trinity (v2.1.1)<sup>68</sup>, creating a  
4  
5 robust and comprehensive *de novo* assembly pipeline.  
6  
7

8 **Proteomic analysis.** Venoms predominantly containing toxins with less than 25 amino acid residues  
9  
10 such as cone snails and wasps, were treated with dithiothreitol (10 mM final concentration, 1 hour at  
11  
12 56 °C) and alkylated with iodoacetamide (20 mM final concentration, 1 hour at room temperature).  
13  
14 Reduced and alkylated peptidic toxins were separated by reverse phase chromatography (nanoLC-  
15  
16 UPLC Acquity, Waters, Milford, Massachusetts, United States) and analyzed by tandem mass  
17  
18 spectrometry (Q-Exactive, ThermoFisher, Waltham, Massachusetts, United States). This “top-down”  
19  
20 proteomics approach has the advantage of measuring the mass of the intact full-size toxins but also  
21  
22 allows for sequence determination using collision-based fragmentation mass spectrometry  
23  
24 experiments. Venoms containing large toxins, such as snake, spider and scorpion venoms, were, after  
25  
26 reduction and alkylation, enzymatically cleaved (by using a ratio of 1/50 protein/trypsin, with  
27  
28 incubation at 37°C over 1 hour). These cleavage conditions yield truncated peptides cleaved at  
29  
30 different positions, with several overlapping peptides covering the full toxin sequence. The peptides  
31  
32 were then analyzed with the same nanoUPLC-MS/MS platform, in the same way as the smaller toxins  
33  
34 permitting the full toxin sequences to be determined. As large toxins have been transformed into a  
35  
36 collection of smaller peptides, the fragmentation yield is then much better and MS/MS spectra can  
37  
38 provide useful sequence information.  
39  
40  
41  
42  
43

44 **Sequence validation.** For the transcriptomics data, different databases were generated: (1) the first  
45  
46 one including genes coding for peptidic toxins by using predictive bioinformatics tools and (2) the rest  
47  
48 of the venom gland transcriptome translated into the six possible open reading frames. The  
49  
50 bioinformatics data analysis has been supported by Peaks software (Bioinformatics Solutions Inc.,  
51  
52 Waterloo, Canada). In addition to classical data analysis, Peaks was also used to *de novo* sequence  
53  
54 peptides not identified from the databases, and to predict mutations with Spider algorithms. All the  
55  
56  
57  
58  
59  
60

1  
2  
3 peptides discovered was classified and only those with a false discovery rate of less than 0.1% were  
4  
5 considered for inclusion in the Venomics database.  
6

7  
8 **High-throughput peptide production by solid-phase synthesis.** The high-throughput chemical  
9  
10 production of toxins up to 34 residues with or without post-translational modifications (PTMs) was  
11  
12 undertaken by solid-phase peptide synthesis on an automated Prelude synthesizer (Protein  
13  
14 Technologies, USA) equipped with six reactions vessels using the Fmoc strategy. Amino acids were  
15  
16 coupled twice for 10 minutes using five or twenty equivalents for toxins up to 25 residues or between  
17  
18 26 and 34 residues, respectively, on a ChemMatrix<sup>®</sup> or pre-loaded Wang resins using HCTU as a  
19  
20 coupling reagent. Both strategies lead to toxins with excellent purities higher than 50%. After a  
21  
22 desalting step on homemade C18 columns and a drying step, the refolding of the different toxins was  
23  
24 performed directly on the crude material. A universal folding protocol was applied for each toxin as  
25  
26 already described<sup>28</sup> using 50 mM HEPES buffer pH 7.5, 20% acetonitrile, 1 mM cysteine, 0.1 mM  
27  
28 cystine, over 48 hours at 4 °C. After a desalting step and analytical characterizations by analytical  
29  
30 HPLC on a Waters e2695 separation module equipped with a XBridge BEH300 C18 column ( 5µm ,  
31  
32 4.6x250mm, Waters) and by mass spectrometry on a MALDI-TOF spectrometer (4800 MALDI TOF-  
33  
34 TOF Analyser from Applied Biosystems, USA), the toxins were distributed in 96 well plates at a  
35  
36 concentration between 2-120 µM.  
37  
38  
39  
40  
41

42 **High-throughput peptide production by recombinant expression.** Optimization of gene synthesis  
43  
44 and expression clone constructs for venom-peptide production have been published previously<sup>31</sup>, in  
45  
46 addition to the final protocols used in this project<sup>29</sup>. Briefly, genes encoding animal venom peptides  
47  
48 were designed with codon optimization for expression in *E. coli* using the ATGenium codon  
49  
50 optimization algorithm (NZYTech genes & enzymes, Portugal). Synthetic genes were produced in a  
51  
52 high-throughput pipeline and purified PCR products were subsequently cloned into the pHTP4  
53  
54 expression vector using the NZYEasy Cloning kit (NZYTech genes & enzymes, Portugal), according  
55  
56 to ligation independent cloning (LIC) technology protocols. In pHTP4, peptide genes are placed under  
57  
58  
59  
60

1  
2  
3 the control of a T7 promoter, and are produced as fusion proteins in the *E. coli* periplasmic space.  
4  
5 Fusions contain an N-terminal Disulphide bond isomerase C (DsbC) fusion, an internal hexa-histidine  
6  
7 tag for purification and a TEV protease recognition sequence to allow cleavage and isolation of the  
8  
9 native peptides. Optimization of expression and purification conditions<sup>30,31</sup> and the final recombinant  
10  
11 production protocols used herein<sup>29</sup> have been described previously. All steps were carried out in 24-  
12  
13 deep-well plates or in 96-deep-well plates on a Tecan Freedom EVO 200 robot (Switzerland)<sup>69</sup>. pHTP4  
14  
15 expression plasmids were transformed into BL21 (DE3) pLysS *E. coli* and cultured overnight at 25 °C  
16  
17 before harvesting. Cell pellets were lysed enzymatically and by sonication. DsbC-His-peptide fusions  
18  
19 were isolated from the soluble cell lysates by nickel affinity purification. Concentrations of the 96  
20  
21 purified DsbC-His-fusion proteins were estimated by UV absorbance at 280 nm in a microtiter plate  
22  
23 reader (GENios Plus, TECAN, Switzerland). TEVsh protease was added (1/10 w/w) directly to the  
24  
25 eluted fusion proteins. The mixture was supplemented with a final concentration of 0.1 mM DTT then  
26  
27 incubated overnight at room temperature in a Microtron shaking incubator (INFORS-HT, Switzerland)  
28  
29 (200 rpm). Finally, samples were acidified for 1 h with a final concentration of 5% acetonitrile and  
30  
31 0.1% formic acid (FA) at room temperature with mild agitation. TEV protease were precipitated during  
32  
33 this acidification process and the 96 recombinant toxins per plate were subjected to a final purification  
34  
35 step on C18 chromatographic resin by Solid-Phase Extraction (SPE). The target peptides were eluted  
36  
37 from the C18 resin, and analyzed by UHPLC and MS to confirm the identity and the quantity of the  
38  
39 toxin (Figure S2). The toxins were distributed in 96 well plates at a concentration between 1-10  $\mu$ M.

46  
47 **Production of N-TRTX-Preg1a and N-BUTX-Ptr1a toxins.** Linear peptide was obtained by solid-  
48  
49 phase chemical synthesis following the same protocol described above using Fmoc-Thr(tBu)-Wang  
50  
51 resin LL (Novabiochem<sup>®</sup>) for N-TRTX-Preg1a and H-Cys(Trt)-Trityl NovaPEG resin  
52  
53 (Novabiochem<sup>®</sup>) for N-BUTX-Ptr1a. After resin cleavage and ether precipitation, linear peptide was  
54  
55 then purified on Waters Preparative HPLC 600 with an XBridge BEH 300Å Prep C18 column (5  $\mu$ m,  
56  
57 19 x 250 mm, Waters). If disulfide bridges were to be prevented from forming, linear peptide was

1  
2  
3 treated in Guanidinium chloride 6M with iodoacetamide (ten times in excess) for 2h under agitation at  
4  
5 room temperature. The oxidation step was performed using Venomics oxidation conditions<sup>28</sup>.  
6  
7 Resulting oxidized peptide was finally purified on a Waters 600 controller using an XBridge BEH  
8  
9 300Å Prep C18 column (5 µm, 10 x 250 mm, Waters). Purity of the final product was above 95% and  
10  
11 was assessed separately by MALDI-TOF spectrometry (4800 MALDI TOF-TOF Analyser (Applied  
12  
13 Biosystems, USA)) and analytical HPLC on Waters e2695 separation module with an XBridge BEH  
14  
15 300Å C18 column (5 µm, 4.6 x 250 mm, Waters). Final quantity was measured by amino-acid  
16  
17 quantification using the PicoTag kit (Waters Associates, Milford, MA).  
18  
19

20  
21 **Sequence analysis and phylogenetic reconstruction.** Sequences were queried against UniProt  
22  
23 (release 2019\_02) and refseq\_protein (release-96) with blastp 2.29.1. Searching parameters were  
24  
25 adjusted for short-length sequences. Spider and scorpion venom proteins in UniProt were annotated as  
26  
27 in VenomZone (<https://venomzone.expasy.org/> release May, 2019). Multiple sequence alignments  
28  
29 were performed with mafft 7.310 --auto option. Phylogenetic trees were obtained with RAxML 8.2.11  
30  
31 --PROTGAMMAAUTO --# 100 options, and MrBayes v3.2.0 ngen=1000000 nchains=4 nruns=2.  
32  
33

34  
35 **NMR structures.** Solutions of 0.5 mM N-TRTX-Preg1a, and 1.2 mM N-BUTX-Ptr1a were prepared  
36  
37 in 3 mm tubes (respectively, 315 µg of N-TRTX-Preg1a and about 1 mg of N-BUTX-Ptr1a were  
38  
39 dissolved in 220 µL of H<sub>2</sub>O/D<sub>2</sub>O (95/5), at pH 4.5). All NMR spectra were acquired on a BRUKER 700  
40  
41 MHz NMR spectrometer equipped with a 5 mm TCI cryoprobe, at 298K. The standard set of 2D <sup>1</sup>H-  
42  
43 NMR experiments (COSY, 80 ms TOCSY, and 160 ms NOESY) were completed with natural  
44  
45 abundance <sup>15</sup>N- and <sup>13</sup>C-HSQC spectra. Processing, analyses, and calculations were successively  
46  
47 performed with Bruker's TopSpin3.2, CcpNMR<sup>70</sup> and Aria2.3<sup>71</sup> software. Chemical shift assignments  
48  
49 were carried out following the classical procedure, distance constraints were deduced from the volume  
50  
51 integration of NOE correlations and introduced into the standard iterative calculation process of the  
52  
53 Aria (ambiguous restraints for iterative assignment) software. For each Aria run, 500 structure  
54  
55 calculations were performed per iteration, and 250 structures were calculated in the final iteration in  
56  
57  
58  
59

1  
2  
3 water. For both samples, the first runs were performed allowing all combinations of disulfide pairing  
4 (ambiguous intersulfur constraints option). This option was kept for the N-BUTX-Ptr1a, for which the  
5  
6 calculations converged towards a unique array of disulfide bridges, and the only additional constraints  
7  
8 were the H-bonds observed in the  $\beta$ -sheet already formed in the first runs. For the N-TRTX-Preg1a  
9  
10 sample,  $(\varphi, \psi)$  angular restraints deduced by the TALOS software<sup>72</sup> from  $^{13}\text{C}$  and  $^1\text{H}$  chemical shifts  
11  
12 were added to NOE constraints. Covalent bonds between sulphur atoms involved in each defined  
13  
14 bridge were introduced in the last runs of the calculation. Finally, for each peptide, ten of the structures  
15  
16 refined in water in the final iteration were selected according to their agreement with experimental  
17  
18 data. The final sets of structures were displayed and analyzed using PyMol (The PyMOL Molecular  
19  
20 Graphics System, Version 2.0 Schrödinger, LLC), and their quality were evaluated using Promotif<sup>73</sup>  
21  
22 and PROCHECK3.5<sup>74</sup> software.

23  
24  
25  
26  
27  
28 **Screening.** The  $^{125}\text{I}$ -NDP- $\alpha$ -MSH and membranes expressing receptors were purchased from Perkin  
29  
30 Elmer (Courtaboeuf, France). Screening conditions were optimised in order to get a  $Z'$ -factor value  
31  
32 superior to 0.5.  $Z'$ -factor is defined as:  $1 - \frac{3\text{MAD}_{max} + 3\text{MAD}_{min}}{\text{Median}_{max} - \text{Median}_{min}}$ , where MAD equals Median  
33  
34 Absolute Deviation. On each 96-well plate, eight wells were dedicated to negative controls and  
35  
36 correspond to total binding of the radioligand in absence of any other ligand. Nonspecific binding was  
37  
38 measured in the presence of 1  $\mu\text{M}$  NDP- $\alpha$ -MSH (Tocris Bioscience) in eight other wells per plate.  
39  
40 Following screening, peptides with an associated binding value bigger or lower than Median –  
41  
42  $3 \times \text{MAD}_{\text{median}}$ <sup>75</sup> were classified as hits. Peptides were screened at concentrations between 0.1 to 100  
43  
44  $\mu\text{M}$  according to their production yields.

45  
46  
47  
48  
49 **Radioligand binding assays.** The  $^{125}\text{I}$ -NDP- $\alpha$ -MSH and melanocortin receptors were purchased from  
50  
51 Perkin Elmer (Courtaboeuf, France). Binding experiments were performed using radioligand at 50 pM  
52  
53 in a 100- $\mu\text{L}$  reaction mixture at room temperature in buffer composed of 25 mM Hepes pH 7.4, 1 mM  
54  
55  $\text{MgCl}_2$ , 1 mM  $\text{CaCl}_2$ , 0.2 g/L BSA with various concentrations of competitors. Controls were the same  
56  
57  
58  
59  
60



1  
2  
3 as for the primary screening on hMC4R. After 3 hours of incubation time, samples were filtered  
4 through 96 well GF/C filter plates preincubated with 0.5% polyethylenimine. 25  $\mu$ L of MicroScint 0  
5  
6 was added onto each dry filter and the radioactivity was quantified on a TopCount beta counter (Perkin  
7  
8 Elmer, Courtaboeuf, France) with 45% yield. Competition binding data were fitted to a one-site/state  
9  
10 inhibition mass action curve using Prism software, v5.0.0 (GraphPad, La Jolla, CA, USA). IC<sub>50</sub> values  
11  
12 were converted to K<sub>i</sub>, using the Cheng–Prusoff equation<sup>76</sup>. K<sub>d</sub> values were 0.34 nM, 0.47 nM, 0.25  
13  
14 nM and 0.88 nM, respectively, for hMC1R, hMC3R, hMC4R and hMC5R. Data represent at least  
15  
16 three independent experiments performed in duplicate and are presented as mean  $\pm$  SEM.  
17  
18  
19  
20

21 **Functional assays.** Human embryonic kidney cells (HEK293, DSMZ, ACC305, authentication: full-  
22 matching STR reference profile of 293 in the database, authentic) and HEK293-MC1R (human  
23 embryonic kidney cells expressing the human melanocortin 1 receptor with an N-terminal HA-tag and  
24 C-terminal FLAG-tag due to stable transfection, selection marker: hygromycin, AG Beck-Sickinger,  
25 University Leipzig, Leipzig, Germany) were grown in Dulbecco's modified Eagle's medium (DMEM)  
26 with 4.5 g/L glucose and L-glutamine and Ham's F12 (1:1, v/v; Lonza) supplied with 15% (v/v) heat-  
27 inactivated fetal calf serum (FCS; Lonza) in a humidified atmosphere at 37 °C and 5% CO<sub>2</sub>. HEK293  
28 or stably transfected HEK293-HA\_MC1R\_FLAG cells were transfected in solution with 0.15  $\mu$ L  
29 Lipofectamine 2000 (Invitrogen) per well and 30 ng Plasmid pGL4.29[luc2P/CRE/Hygro] (Promega)  
30 per well and additionally 30 ng pcDNA3-hMC4R per well when needed. Lipofectamine and plasmid  
31 solutions were prepared in 6  $\mu$ L Opti-MEM medium (Gibco) per well. After 5 min incubation at RT  
32 the Lipofectamine solution was slowly added to the DNA solution and incubated for 30 min at RT.  
33 Transfection of 20,000 cells/well in 30  $\mu$ L medium (DMEM/Ham's F12/ 15% FCS) without antibiotics  
34 was performed in solution by slowly adding the DNA/Lipofectamine solution to the cells. After careful  
35 mixing, 42  $\mu$ L per well was plated on a 384-well microplate (Greiner bio-one, F-Bottom, white) and  
36 incubated overnight. The day after transfection, medium was removed and cells were stimulated with  
37 15  $\mu$ L peptide solution in serum free medium (1 nM  $\alpha$ -MSH (AG Beck-Sickinger, University Leipzig,  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 Leipzig, Germany) corresponding to  $\sim$ EC<sub>85</sub> at hMC1R or  $\sim$ EC<sub>75</sub> at hMC4R, 100  $\mu$ M N-TRTX-Preg1a  
4 or 100  $\mu$ M N-BUTX-Ptr1a) for 3 h at 37 °C, 5% CO<sub>2</sub>. As a positive control 1  $\mu$ M forskolin (Sigma-  
5 Aldrich) was used, while serum-free medium was used as the negative control (baseline). For  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
Luciferase detection 15  $\mu$ l OneGlo Substrate (Promega) was added at RT to each well. The 384-well  
plate was transferred directly to the plate reader (Tecan Infinite 200 PRO) and luminescence was  
measured after shaking for 5 sec (double orbital, amplitude: 3.5 mm) and 5 min incubation with an  
integration time of 1000 ms at room temperature. Data analysis was performed with GraphPad Prism.  
After baseline subtraction, data were normalized to stimulation with 1  $\mu$ M Forskolin. Data represent  
the mean  $\pm$  SEM from three independent experiments performed in triplicates. Statistical analysis was  
performed with GraphPad Prism 5.03 (GraphPad Software, Inc., San Diego, California, USA).  
Statistical significance was determined by one-way analysis of variance (ANOVA) followed by  
Dunnett's multiple comparison test. \*\*\* p < 0.001.

31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
**Electrophysiology.** For the expression of Na<sub>v</sub> channels (human (h)Na<sub>v</sub>1.1, rat (r)Na<sub>v</sub>1.2, rNa<sub>v</sub>1.3,  
rNa<sub>v</sub>1.4, hNa<sub>v</sub>1.5, mouse (m)Na<sub>v</sub>1.6, hNa<sub>v</sub>1.7, rNa<sub>v</sub>1.8 channels and the auxiliary subunits r $\beta$ 1, h $\beta$ 1),  
K<sub>v</sub> channels (rK<sub>v</sub>1.1, rK<sub>v</sub>1.2, hK<sub>v</sub>1.3, rK<sub>v</sub>1.4, rK<sub>v</sub>1.5, rK<sub>v</sub>1.6, rK<sub>v</sub>2.1, hK<sub>v</sub>3.1, rK<sub>v</sub>4.3, hK<sub>v</sub>7.2,  
hK<sub>v</sub>7.3, hK<sub>v</sub>10.1, hK<sub>v</sub>11.1), Ca<sub>v</sub> channels (rCa<sub>v</sub>3.3), hGIRK1 and nAChRs (h $\alpha$ 1 $\beta$ 1 $\gamma$  $\delta$  and h $\alpha$ 7)  
in *Xenopus laevis* oocytes, the linearized plasmids were transcribed using the T7 or SP6 mMACHINE-  
mMACHINE transcription kit (Ambion, Carlsbad, CA). The harvesting of stage V–VI oocytes from  
anesthetized female *Xenopus laevis* frogs was previously described<sup>77</sup>. Oocytes were injected with 50  
nl of cRNA at a concentration of 1 ng/nl using a microinjector (Drummond Scientific, Broomall, PA).  
The oocytes were incubated in a solution containing 96 mM NaCl, 2 mM KCl, 1.8 mM CaCl<sub>2</sub>, 2 mM  
MgCl<sub>2</sub>, and 5 mM HEPES pH 7.4, supplemented with 50 mg/L gentamycin sulfate.

54  
55  
56  
57  
58  
59  
60  
Two-electrode voltage-clamp recordings were performed at room temperature (18–22 °C) using a  
Geneclamp 500 amplifier (Molecular Devices, Downingtown, PA) controlled by a pClamp data  
acquisition system (Axon Instruments, Union City, CA). Whole cell currents from oocytes were

1  
2  
3 recorded 1–4 days after injection. Bath solution composition was 96 mM NaCl, 2 mM KCl, 1.8 mM  
4  
5 CaCl<sub>2</sub>, 2 mM MgCl<sub>2</sub>, and 5 mM HEPES pH 7.4. Toxins were applied directly to the bath. Resistances  
6  
7 of both electrodes were kept between 0.8 and 1.5 MΩ. The elicited currents were filtered at 2 kHz  
8  
9 (Na<sub>v</sub>), 0.5 kHz (K<sub>v</sub> & Ca<sub>v</sub>), 0.05 kHz (GIRK & nAChR) and sampled at 20 kHz (Na<sub>v</sub>), 2 kHz (K<sub>v</sub> &  
10  
11 Ca<sub>v</sub>) and 0.1 kHz (GIRK & nAChR) using a four-pole low-pass Bessel filter. Leak subtraction was  
12  
13 performed using a -P/4 protocol. For the electrophysiological analysis, a number of protocols were  
14  
15 applied from a holding potential of -90 mV. K<sub>v</sub>1.1–K<sub>v</sub>1.6 currents were evoked by 250 ms  
16  
17 depolarizations to 0 mV followed by a 250 ms pulse to -50 mV. K<sub>v</sub>2.1, K<sub>v</sub>3.1, and K<sub>v</sub>4.3 currents  
18  
19 were elicited by 250 ms pulses to +20 mV from a holding potential of -90 mV. K<sub>v</sub>7 and K<sub>v</sub>10.1  
20  
21 currents were evoked by 2 s depolarizing pulses to 0 mV from a holding potential of -90 mV. Current  
22  
23 traces of K<sub>v</sub>11.1 channels were elicited by applying a +40 mV prepulse for 2 s followed by a step to  
24  
25 -120 mV for 2 s. Sodium current traces were evoked by 100 ms depolarization to the voltage  
26  
27 corresponding to maximal sodium current in control conditions. For GIRK measurements, the oocytes  
28  
29 were clamped at -70 mV and superfused with a high K<sup>+</sup> solution (96 mM KCl, 2 mM NaCl, 1 mM  
30  
31 MgCl<sub>2</sub>, 1.8 mM CaCl<sub>2</sub>, 5 mM HEPES pH 7.5). Ca<sub>v</sub> channel currents, carried by 10 mM Ba<sup>2+</sup>, were  
32  
33 elicited by 200 ms test depolarizations applied from -90 mV to voltages ranging from -50 mV or +30  
34  
35 mV. For nAChRs, oocytes were clamped at a holding potential of -70 mV throughout all recordings.  
36  
37 To minimize interference of endogenous Ca<sup>2+</sup>-activated Cl<sup>-</sup> channels during the experiments, a Ca<sup>2+</sup>-  
38  
39 free bath perfusion solution was used, containing 82.5 mM NaCl, 2.5 mM KCl, 1 mM MgCl<sub>2</sub>, 2.5  
40  
41 mM BaCl<sub>2</sub>, and 5 mM HEPES pH 7.4. Current responses were evoked by 10 s applications of  
42  
43 increasing concentrations of acetylcholine (starting at 200 μM). All data were analyzed using pClamp  
44  
45 Clampfit 10.1 (Molecular Devices) and Origin 7.5 software (OriginLab, Northampton, MA).  
46  
47  
48  
49  
50  
51  
52

53 **Toxins nomenclature.** Newly characterized toxins were named according to the rational nomenclature  
54  
55 proposed by King *et al.*<sup>78</sup> with a greek letter associated with a biological function followed by the  
56  
57 generic name indicating from which taxon the toxin is derived, genus and species indicated  
58  
59  
60

1  
2  
3 respectively by uppercase and lowercase letters, a number to distinguish different toxins with similar  
4 activity and a lowercase letter for isoforms. In this study, uppercase nu greek letter ( $\nu$ ) was chosen as  
5 new descriptor of toxins active on melanocortin receptors.  
6  
7  
8  
9

## 10 11 12 **ASSOCIATED CONTENT**

### 13 14 **Supporting Information**

15 The Supporting Information is available free of charge on the ACS Publications website at DOI:  
16  
17 LC-MS analysis of five representative toxins with their cystein patterns, produced by solid phase  
18 chemical synthesis; analysis of 4 representative toxins with their cystein patterns, produced by  
19 recombinant expression; quality assessment at each stage of N-TRTX-Preg1a and N-BUTX-Ptr1a  
20 production in Fmoc solid-phase chemical synthesis; activity of 100  $\mu$ M N-TRTX-Preg1a on ion  
21 channels; activity of 100  $\mu$ M N-BUTX-Ptr1a on ion channels; cysteine pattern diversity for toxins  
22 containing an even number of cysteines identified in the Venomics bank; cysteine pattern diversity for  
23 toxins containing an odd number of cysteine residues identified in the Venomics bank; experimental  
24 NMR data used for the calculations, and final structural statistics for the ten models representative of  
25 the solution structure of N-TRTX-Preg1a; experimental NMR data used for the calculations, and final  
26 structural statistics for the ten models representative of the solution structure of N-BUTX-Ptr1a  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43

### 44 45 **AUTHOR INFORMATION**

46 Corresponding Author

47 \*E-mail: [Nicolas.gilles@cea.fr](mailto:Nicolas.gilles@cea.fr)  
48  
49  
50  
51  
52  
53

### 54 55 **ACKNOWLEDGMENT**

56 We specifically thank Dr. Philippe Robin, lecturer at the Université Paris-Sud, and Mrs. Marion  
57 Chaigneau for technical support.  
58  
59  
60

## ABBREVIATIONS

ANOVA, one-way analysis of variance; BUTX, butantoxin; Cav, voltage-gated calcium channel; CHO, chinese hamster ovary cells; COSY, correlation spectroscopy; CS $\alpha\beta$ , cysteine-stabilized  $\alpha$ -helix  $\beta$ -sheet; DsbC, Disulfide bond Isomerase C; DMEM, Dulbecco's modified Eagle's medium; Fmoc, fluorenylmethoxycarbonyl; GIRK, G protein-coupled inwardly-rectifying potassium channel; GPCR, G-Protein Coupled receptors; HCTU, H-6-Chlorobenzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate; HEK, human embryonic kidney cells; HEPES, hydroxyethyl piperazineethanesulfonic acid; HPLC, high-performance liquid chromatography; HT, high-throughput; HT, high-throughput screening; ICK, inhibitory cystine knot; Ki, inhibitory constant; Kv, voltage-gated potassium channel; LC, liquid chromatography; hMC1R, human melanocortin 1 receptor; hMC3R, human melanocortin 3 receptor; hMC4R, human melanocortin 4 receptor; hMC5R, human melanocortin 5 receptor; MALDI-TOF, Matrix Assisted Laser Desorption Ionisation - Time of Flight; MAD, median absolute deviation; MS, mass spectrometry; MSH, melanocyte stimulating hormone; N $^{\circ}$ , number; Nav, voltage-gated sodium channel; NDP- $\alpha$ -MSH, [Nle<sup>4</sup>,D-phe<sup>7</sup>]- $\alpha$ -melanocyte stimulating hormone; NGS, next-generation sequencing; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; NOESY, nuclear Overhauser effect spectroscopy; PTM, post-translational modification; RMSD, root-mean-square deviation; RT, room temperature; SD, standard deviation; SEM, standard error of the mean; Seq, sequence; TOCSY, total correlation spectroscopy; TRTX, theraphotoxin; UPLC, ultra-performance liquid chromatography; UV, ultraviolet.

## FUNDING

This work was supported by the European FP7 project 278346 2011-2015 and by the French Infrastructure for Integrated Structural Biology (FRISBI) ANR-10-INSB-05-01. A PhD grant (S.R.)

1  
2  
3 was supported by the University of Paris-Sud. JT was funded by GOC2319N and GOA4919N (F.W.O.  
4  
5 Vlaanderen) and CELSA/17/047 (BOF, KU Leuven). SP is a postdoctoral fellow supported by KU  
6  
7 Leuven funding (PDM/19/164).  
8  
9  
10  
11

## 12 REFERENCES

### 13 Edited Books:

- 14  
15  
16  
17  
18 (1) Jenner, R.; Undheim, E. *Venom: The Secrets of Nature's Deadliest Weapon*, The Natura.;  
19  
20 Smithsonian Books, 2017.  
21  
22

### 23 Journals:

- 24  
25 (2) Degueldre, M.; Verdenaud, M.; Legarda, G.; Minambres, R.; Zuniga, S.; Leblanc, M.; Gilles,  
26  
27 N.; Ducancel, F.; De Pauw, E.; Quinton, L. Diversity in Sequences, Post-Translational  
28  
29 Modifications and Expected Pharmacological Activities of Toxins from Four Conus Species  
30  
31 Revealed by the Combination of Cutting-Edge Proteomics, Transcriptomics and  
32  
33 Bioinformatics. *Toxicon* **2017**, *130*, 116–125. <https://doi.org/10.1016/j.toxicon.2017.02.014>.  
34  
35  
36 (3) Beraud, E.; Chandy, K. G. Therapeutic Potential of Peptide Toxins That Target Ion Channels.  
37  
38 *Inflamm. Allergy Drug Targets* **2011**, *10* (5), 322–342.  
39  
40  
41 (4) Lewis, R. J.; Dutertre, S.; Vetter, I.; Christie, M. J. Conus Venom Peptide Pharmacology.  
42  
43 *Pharmacol. Rev.* **2012**, *64* (2), 259–298. <https://doi.org/10.1124/pr.111.005322>.  
44  
45  
46 (5) Servent, D.; Fruchart-Gaillard, C. Muscarinic Toxins: Tools for the Study of the  
47  
48 Pharmacological and Functional Properties of Muscarinic Receptors. *J. Neurochem.* **2009**, *109*  
49  
50 (5), 1193–1202. <https://doi.org/10.1111/j.1471-4159.2009.06092.x>.  
51  
52  
53 (6) Ciolek, J.; Reinfrank, H.; Quinton, L.; Viengchareun, S.; Stura, E. A.; Vera, L.; Sigismeau, S.;  
54  
55 Mouillac, B.; Orcel, H.; Peigneur, S.; Tytgat, J.; Droctové, L.; Beau, F.; Nevoux, J.; Lombès,  
56  
57 M.; Mourier, G.; De Pauw, E.; Servent, D.; Mendre, C.; Witzgall, R.; Gilles, N. Green Mamba  
58  
59  
60

- 1  
2  
3 Peptide Targets Type-2 Vasopressin Receptor against Polycystic Kidney Disease. *Proc. Natl.*  
4 *Acad. Sci. U. S. A.* **2017**, *114* (27), 7154–7159. <https://doi.org/10.1073/pnas.1620454114>.  
5  
6  
7  
8 (7) Petrel, C.; Hocking, H. G.; Reynaud, M.; Upert, G.; Favreau, P.; Biass, D.; Paolini-Bertrand,  
9 M.; Peigneur, S.; Tytgat, J.; Gilles, N.; Hartley, O.; Boelens, R.; Stocklin, R.; Servent, D.  
10 Identification, Structural and Pharmacological Characterization of  $\tau$ -CnVA, a Conopeptide That  
11 Selectively Interacts with Somatostatin Sst3 Receptor. *Biochem. Pharmacol.* **2013**, *85* (11),  
12 1663–1671. <https://doi.org/10.1016/j.bcp.2013.03.019>.  
13  
14  
15  
16  
17  
18  
19 (8) Rouget, C.; Quinton, L.; Maïga, A.; Gales, C.; Masuyer, G.; Malosse, C.; Chamot-Rooke, J.;  
20 Thai, R.; Mourier, G.; De Pauw, E.; Gilles, N.; Servent, D. Identification of a Novel Snake  
21 Peptide Toxin Displaying High Affinity and Antagonist Behaviour for the A<sub>2</sub>-Adrenoceptors.  
22 *Br. J. Pharmacol.* **2010**, *161* (6), 1361–1374. [https://doi.org/10.1111/j.1476-](https://doi.org/10.1111/j.1476-5381.2010.00966.x)  
23 [5381.2010.00966.x](https://doi.org/10.1111/j.1476-5381.2010.00966.x).  
24  
25  
26  
27  
28  
29  
30 (9) Quinton, L.; Girard, E.; Maïga, a.; Rekik, M.; Lluel, P.; Masuyer, G.; Larregola, M.; Marquer,  
31 C.; Ciolek, J.; Magnin, T.; Wagner, R.; Molgó, J.; Thai, R.; Fruchart-Gaillard, C.; Mourier, G.;  
32 Chamot-Rooke, J.; Ménez, a.; Palea, S.; Servent, D.; Gilles, N. Isolation and Pharmacological  
33 Characterization of AdTx1, a Natural Peptide Displaying Specific Insurmountable Antagonism  
34 of the  $\alpha$  1A- Adrenoceptor. *Br. J. Pharmacol.* **2010**, *159* (2), 316–325.  
35 <https://doi.org/10.1111/j.1476-5381.2009.00532.x>.  
36  
37  
38  
39  
40  
41  
42  
43  
44 (10) Jolkkonen, M.; Adem, A.; Hellman, U.; Wernstedt, C.; Karlsson, E. A Snake Toxin against  
45 Muscarinic Acetylcholine Receptors: Amino Acid Sequence, Subtype Specificity and Effect on  
46 Guinea-Pig Ileum. *Toxicon* **1995**, *33* (4), 399–410. [https://doi.org/10.1016/0041-](https://doi.org/10.1016/0041-0101(94)00102-e)  
47 [0101\(94\)00102-e](https://doi.org/10.1016/0041-0101(94)00102-e).  
48  
49  
50  
51  
52  
53 (11) Cologna, C. T.; Cardoso, J. dos S.; Jourdan, E.; Degueldre, M.; Upert, G.; Gilles, N.;  
54 Uetanabaro, A. P. T.; Costa Neto, E. M.; Thonart, P.; de Pauw, E.; Quinton, L. Peptidomic  
55 Comparison and Characterization of the Major Components of the Venom of the Giant Ant  
56  
57  
58  
59  
60

- 1  
2  
3 Dinoponera Quadriceps Collected in Four Different Areas of Brazil. *J. Proteomics* **2013**, *94*,  
4 413–422. <https://doi.org/10.1016/j.jprot.2013.10.017>.  
5  
6  
7  
8 (12) Kumar, V. A.; Wickremasinghe, N. C.; Shi, S.; Hartgerink, J. D. Nanofibrous Snake Venom  
9 Hemostat. *ACS Biomater. Sci. Eng.* **2015**, *1* (12), 1300–1305.  
10 <https://doi.org/10.1021/acsbiomaterials.5b00356>.  
11  
12  
13  
14 (13) Pennington, M. W.; Czerwinski, A.; Norton, R. S. Peptide Therapeutics from Venom: Current  
15 Status and Potential. *Bioorg. Med. Chem.* **2018**, *26* (10), 2738–2758.  
16 <https://doi.org/10.1016/j.bmc.2017.09.029>.  
17  
18  
19  
20 (14) Saez, N. J.; Herzig, V. Versatile Spider Venom Peptides and Their Medical and Agricultural  
21 Applications. *Toxicon* **2019**, *158*, 109–126. <https://doi.org/10.1016/j.toxicon.2018.11.298>.  
22  
23  
24  
25 (15) Mebs, D. Toxicity in Animals. Trends in Evolution? *Toxicon* **2001**, *39* (1), 87–96.  
26  
27  
28 (16) Kaas, Q.; Westermann, J.-C.; Craik, D. J. Conopeptide Characterization and Classifications: An  
29 Analysis Using ConoServer. *Toxicon* **2010**, *55* (8), 1491–1509.  
30 <https://doi.org/10.1016/j.toxicon.2010.03.002>.  
31  
32  
33  
34 (17) Gilles, N.; Servent, D. The European FP7 Venomics Project. *Futur. Med. Chem* **2014**, *6* (15),  
35 1611–1612. <https://doi.org/10.4155/FMC.14.85>.  
36  
37  
38  
39 (18) Turner, A. H.; Craik, D. J.; Kaas, Q.; Schroeder, C. I. Bioactive Compounds Isolated from  
40 Neglected Predatory Marine Gastropods. *Mar. Drugs* **2018**, *16* (4).  
41 <https://doi.org/10.3390/md16040118>.  
42  
43  
44  
45 (19) Morsa, D.; Baiwir, D.; La Rocca, R.; Zimmerman, T. A.; Hanozin, E.; Grifnée, E.; Longuespée,  
46 R.; Meuwis, M.-A.; Smargiasso, N.; Pauw, E. De; Mazzucchelli, G. Multi-Enzymatic Limited  
47 Digestion: The Next-Generation Sequencing for Proteomics? *J. Proteome Res.* **2019**, *18* (6),  
48 2501–2513. <https://doi.org/10.1021/acs.jproteome.9b00044>.  
49  
50  
51  
52 (20) Dutertre, S.; Jin, A.; Kaas, Q.; Jones, A.; Alewood, P. F.; Lewis, R. J. Deep Venomics Reveals  
53 the Mechanism for Expanded Peptide Diversity in Cone Snail Venom. *Mol. Cell. Proteomics*  
54  
55  
56  
57  
58  
59  
60



- 1  
2  
3       **2013**, *12* (2), 312–329. <https://doi.org/10.1074/mcp.M112.021469>.
- 4  
5  
6 (21) Rong, M.; Duan, Z.; Chen, J.; Li, J.; Xiao, Y.; Liang, S. Native Pyroglutamation of Huwentoxin-  
7  
8       IV: A Post-Translational Modification That Increases the Trapping Ability to the Sodium  
9  
10       Channel. *PLoS One* **2013**, *8* (6), e65984. <https://doi.org/10.1371/journal.pone.0065984>.
- 11  
12 (22) Lucchesi, K.; Ravindran, A.; Young, H.; Moczydlowski, E. Analysis of the Blocking Activity  
13  
14       of Charybdotoxin Homologs and Iodinated Derivatives against Ca<sup>2+</sup>-Activated K<sup>+</sup> Channels.  
15  
16       *J. Membr. Biol.* **1989**, *109* (3), 269–281.
- 17  
18 (23) Saez, N. J.; Senff, S.; Jensen, J. E.; Er, S. Y.; Herzig, V.; Rash, L. D.; King, G. F. Spider-Venom  
19  
20       Peptides as Therapeutics. *Toxins (Basel)*. **2010**, *2* (12), 2851–2871.  
21  
22       <https://doi.org/10.3390/toxins2122851>.
- 23  
24 (24) Gao, B.; Harvey, P. J.; Craik, D. J.; Ronjat, M.; De Waard, M.; Zhu, S. Functional Evolution of  
25  
26       Scorpion Venom Peptides with an Inhibitor Cystine Knot Fold. *Biosci. Rep.* **2013**, *33* (3), 513–  
27  
28       527. <https://doi.org/10.1042/BSR20130052>.
- 29  
30 (25) Robinson, S. D.; Norton, R. S. Conotoxin Gene Superfamilies. *Mar. Drugs* **2014**, *12* (12), 6058–  
31  
32       6101. <https://doi.org/10.3390/md12126058>.
- 33  
34 (26) Imperial, J. S.; Watkins, M.; Chen, P.; Hillyard, D. R.; Cruz, L. J.; Olivera, B. M. The  
35  
36       Augertoxins: Biochemical Characterization of Venom Components from the Toxoglossate  
37  
38       Gastropod *Terebra Subulata*. *Toxicon* **2003**, *42* (4), 391–398. [https://doi.org/10.1016/S0041-](https://doi.org/10.1016/S0041-0101(03)00169-7)  
39  
40       0101(03)00169-7.
- 41  
42 (27) Lavergne, V.; Harliwong, I.; Jones, A.; Miller, D.; Taft, R. J.; Alewood, P. F. Optimized Deep-  
43  
44       Targeted Proteotranscriptomic Profiling Reveals Unexplored Conus Toxin Diversity and Novel  
45  
46       Cysteine Frameworks. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112* (29), E3782-91.  
47  
48       <https://doi.org/10.1073/pnas.1501334112>.
- 49  
50 (28) Upert, G.; Mourier, G.; Pastor, A.; Verdenaud, M.; Alili, D.; Servent, D.; Gilles, N. High-  
51  
52       Throughput Production of Two Disulphide-Bridge Toxins. *Chem. Commun.* **2014**, *50* (61),  
53  
54  
55  
56  
57  
58  
59  
60

- 8408–8411. <https://doi.org/10.1039/c4cc02679a>.
- (29) Duhoo, Y.; Sequeira, A. F.; Saez, N. J.; Turchetto, J.; Ramond, L.; Peysson, F.; Brás, J. L. A.; Gilles, N.; Darbon, H.; Fontes, C. M. G. A.; Vincentelli, R. High-Throughput Production of Oxidized Animal Toxins in Escherichia Coli. *Methods Mol. Biol.* **2019**, *2025*, 165–190. [https://doi.org/10.1007/978-1-4939-9624-7\\_7](https://doi.org/10.1007/978-1-4939-9624-7_7).
- (30) Sequeira, A. F.; Turchetto, J.; Saez, N. J.; Peysson, F.; Ramond, L.; Duhoo, Y.; Blémont, M.; Fernandes, V. O.; Gama, L. T.; Ferreira, L. M. A.; Guerreiro, C. I. P. I.; Gilles, N.; Darbon, H.; Fontes, C. M. G. A.; Vincentelli, R. Gene Design, Fusion Technology and TEV Cleavage Conditions Influence the Purification of Oxidized Disulphide-Rich Venom Peptides in Escherichia Coli. *Microb. Cell Fact.* **2017**, *16* (1), 1–16. <https://doi.org/10.1186/s12934-016-0618-0>.
- (31) Turchetto, J.; Sequeira, A. F.; Ramond, L.; Peysson, F.; Brás, J. L. A.; Saez, N. J.; Duhoo, Y.; Blémont, M.; Guerreiro, C. I. P. D.; Quinton, L.; Pauw, E.; Gilles, N.; Darbon, H.; Fontes, C. M. G. A.; Vincentelli, R. High-Throughput Expression of Animal Venom Toxins in Escherichia Coli to Generate a Large Library of Oxidized Disulphide-Reticulated Peptides for Drug Discovery. *Microb. Cell Fact.* **2017**, *16* (1), 1–15. <https://doi.org/10.1186/s12934-016-0617-1>.
- (32) Massonnet, P.; Haler, J. R. N.; Upert, G.; Smargiasso, N.; Mourier, G.; Gilles, N.; Quinton, L.; De Pauw, E. Disulfide Connectivity Analysis of Peptides Bearing Two Intramolecular Disulfide Bonds Using MALDI In-Source Decay. *J. Am. Soc. Mass Spectrom.* **2018**, *29* (10), 1995–2002. <https://doi.org/10.1007/s13361-018-2022-y>.
- (33) Muratspahić, E.; Freissmuth, M.; Gruber, C. W. Nature-Derived Peptides: A Growing Niche for GPCR Ligand Discovery. *Trends Pharmacol. Sci.* **2019**, *40* (5), 309–326. <https://doi.org/10.1016/j.tips.2019.03.004>.
- (34) Liu, Z.; Dai, J.; Dai, L.; Deng, M.; Hu, Z.; Hu, W.; Liang, S. Function and Solution Structure of Huwentoxin-X, a Specific Blocker of N-Type Calcium Channels, from the Chinese Bird

- Spider *Ornithoctonus Huwena*. *J. Biol. Chem.* **2006**, *281* (13), 8628–8635.  
<https://doi.org/10.1074/jbc.M513542200>.
- (35) Favreau, P.; Gilles, N.; Lamthanh, H.; Bournaud, R.; Shimahara, T.; Bouet, F.; Laboute, P.; Letourneux, Y.; Ménez, A.; Molgó, J.; Le Gall, F. A New Omega-Conotoxin That Targets N-Type Voltage-Sensitive Calcium Channels with Unusual Specificity. *Biochemistry* **2001**, *40* (48), 14567–14575. <https://doi.org/10.1021/bi002871r>.
- (36) Wang, F.; Yan, Z.; Liu, Z.; Wang, S.; Wu, Q.; Yu, S.; Ding, J.; Dai, Q. Molecular Basis of Toxicity of N-Type Calcium Channel Inhibitor MVIIA. *Neuropharmacology* **2016**, *101*, 137–145. <https://doi.org/10.1016/j.neuropharm.2015.08.047>.
- (37) Huys, I.; Dyason, K.; Waelkens, E.; Verdonck, F.; van Zyl, J.; du Plessis, J.; Müller, G. J.; van der Walt, J.; Clynen, E.; Schoofs, L.; Tytgat, J. Purification, Characterization and Biosynthesis of Parabutoxin 3, a Component of *Parabuthus Transvaalicus* Venom. *Eur. J. Biochem.* **2002**, *269* (7), 1854–1865. <https://doi.org/10.1046/j.1432-1033.2002.02833.x>.
- (38) Papp, F.; Batista, C. V. F.; Varga, Z.; Herceg, M.; Román-González, S. A.; Gaspar, R.; Possani, L. D.; Panyi, G. Tst26, a Novel Peptide Blocker of Kv1.2 and Kv1.3 Channels from the Venom of *Tityus Stigmurus*. *Toxicon* **2009**, *54* (4), 379–389. <https://doi.org/10.1016/j.toxicon.2009.05.023>.
- (39) Rogowski, R. S.; Krueger, B. K.; Collins, J. H.; Blaustein, M. P. Tityustoxin K Alpha Blocks Voltage-Gated Noninactivating K<sup>+</sup> Channels and Unblocks Inactivating K<sup>+</sup> Channels Blocked by Alpha-Dendrotoxin in Synaptosomes. *Proc. Natl. Acad. Sci. U. S. A.* **1994**, *91* (4), 1475–1479. <https://doi.org/10.1073/pnas.91.4.1475>.
- (40) Cerni, F. A.; Pucca, M. B.; Peigneur, S.; Cremonez, C. M.; Bordon, K. C. F.; Tytgat, J.; Arantes, E. C. Electrophysiological Characterization of Ts6 and Ts7, K<sup>+</sup> Channel Toxins Isolated through an Improved *Tityus Serrulatus* Venom Purification Procedure. *Toxins (Basel)*. **2014**, *6* (3), 892–913. <https://doi.org/10.3390/toxins6030892>.

- 1  
2  
3 (41) Rodrigues, A. R. A.; Arantes, E. C.; Monje, F.; Stühmer, W.; Varanda, W. A. Tityustoxin-  
4 K(Alpha) Blockade of the Voltage-Gated Potassium Channel Kv1.3. *Br. J. Pharmacol.* **2003**,  
5  
6 K(Alpha) Blockade of the Voltage-Gated Potassium Channel Kv1.3. *Br. J. Pharmacol.* **2003**,  
7  
8 139 (6), 1180–1186. <https://doi.org/10.1038/sj.bjp.0705343>.  
9
- 10 (42) Diego-García, E.; Batista, C. V. F.; García-Gómez, B. I.; Lucas, S.; Candido, D. M.; Gómez-  
11  
12 Lagunas, F.; Possani, L. D. The Brazilian Scorpion Tityus Costatus Karsch: Genes, Peptides  
13  
14 and Function. *Toxicon* **2005**, 45 (3), 273–283. <https://doi.org/10.1016/j.toxicon.2004.10.014>.  
15  
16
- 17 (43) Garcia, M. L.; Garcia-Calvo, M.; Hidalgo, P.; Lee, A.; MacKinnon, R. Purification and  
18  
19 Characterization of Three Inhibitors of Voltage-Dependent K<sup>+</sup> Channels from Leiurus  
20  
21 Quinquestriatus Var. Hebraeus Venom. *Biochemistry* **1994**, 33 (22), 6834–6839.  
22  
23
- 24 (44) Takacs, Z.; Toups, M.; Kollewe, A.; Johnson, E.; Cuello, L. G.; Driessens, G.; Biancalana, M.;  
25  
26 Koide, A.; Ponte, C. G.; Perozo, E.; Gajewski, T. F.; Suarez-Kurtz, G.; Koide, S.; Goldstein, S.  
27  
28 A. N. A Designer Ligand Specific for Kv1.3 Channels from a Scorpion Neurotoxin-Based  
29  
30 Library. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, 106 (52), 22211–22216.  
31  
32  
33 <https://doi.org/10.1073/pnas.0910123106>.  
34
- 35 (45) Kuzmenkov, A. I.; Grishin, E. V.; Vassilevski, A. A. Diversity of Potassium Channel Ligands:  
36  
37 Focus on Scorpion Toxins. *Biochem.* **2016**, 80 (13), 1764–1799.  
38  
39  
40 <https://doi.org/10.1134/s0006297915130118>.  
41
- 42 (46) Carvalho, A. de O.; Gomes, V. M. Plant Defensins--Prospects for the Biological Functions and  
43  
44 Biotechnological Properties. *Peptides* **2009**, 30 (5), 1007–1020.  
45  
46  
47 <https://doi.org/10.1016/j.peptides.2009.01.018>.  
48
- 49 (47) Koehbach, J. Structure-Activity Relationships of Insect Defensins. *Front. Chem.* **2017**, 5 (July),  
50  
51 45. <https://doi.org/10.3389/fchem.2017.00045>.  
52  
53
- 54 (48) Cornet, B.; Bonmatin, J. M.; Hetru, C.; Hoffmann, J. A.; Ptak, M.; Vovelle, F. Refined Three-  
55  
56 Dimensional Solution Structure of Insect Defensin A. *Structure* **1995**, 3 (5), 435–448.  
57  
58  
59 [https://doi.org/10.1016/S0969-2126\(01\)00177-0](https://doi.org/10.1016/S0969-2126(01)00177-0).  
60

- 1  
2  
3 (49) Feng, B. Y.; Shoichet, B. K. Synergy and Antagonism of Promiscuous Inhibition in Multiple-  
4 Compound Mixtures. *J. Med. Chem.* **2006**, *49* (7), 2151–2154.  
5  
6 <https://doi.org/10.1021/jm060029z>.  
7  
8  
9  
10 (50) Li, J. W.-H.; Vederas, J. C. Drug Discovery and Natural Products: End of an Era or an Endless  
11 Frontier? *Science* **2009**, *325* (5937), 161–165. <https://doi.org/10.1126/science.1168243>.  
12  
13  
14 (51) Bender, A.; Bojanic, D.; Davies, J. W.; Crisman, T. J.; Mikhailov, D.; Scheiber, J.; Jenkins, J.  
15 L.; Deng, Z.; Hill, W. A. G.; Popov, M.; Jacoby, E.; Glick, M. Which Aspects of HTS Are  
16 Empirically Correlated with Downstream Success? *Curr. Opin. Drug Discov. Devel.* **2008**, *11*  
17 (3), 327–337. <https://doi.org/18428086>.  
18  
19  
20 (52) Oprea, T. I. Current Trends in Lead Discovery: Are We Looking for the Appropriate Properties?  
21 *Mol. Divers.* **2002**, *5* (4), 199–208.  
22  
23  
24 (53) Andermann, M. L.; Lowell, B. B. Toward a Wiring Diagram Understanding of Appetite  
25 Control. *Neuron* **2017**, *95* (4), 757–778. <https://doi.org/10.1016/j.neuron.2017.06.014>.  
26  
27  
28 (54) Kobayashi, N.; Nakagawa, A.; Muramatsu, T.; Yamashina, Y.; Shirai, T.; Hashimoto, M. W.;  
29 Ishigaki, Y.; Ohnishi, T.; Mori, T. Supranuclear Melanin Caps Reduce Ultraviolet Induced  
30 DNA Photoproducts in Human Epidermis. *J. Invest. Dermatol.* **1998**, *110* (5), 806–810.  
31 <https://doi.org/10.1046/j.1523-1747.1998.00178.x>.  
32  
33  
34 (55) Kadekaro, A. L.; Leachman, S.; Kavanagh, R. J.; Swope, V.; Cassidy, P.; Supp, D.; Sartor, M.;  
35 Schwemberger, S.; Babcock, G.; Wakamatsu, K.; Ito, S.; Koshoffer, A.; Boissy, R. E.; Manga,  
36 P.; Sturm, R. A.; Abdel-Malek, Z. A. Melanocortin 1 Receptor Genotype: An Important  
37 Determinant of the Damage Response of Melanocytes to Ultraviolet Radiation. *FASEB J.* **2010**,  
38 *24* (10), 3850–3860. <https://doi.org/10.1096/fj.10-158485>.  
39  
40  
41 (56) Creemers, J. W. M.; Pritchard, L. E.; Gyte, A.; Le Rouzic, P.; Meulemans, S.; Wardlaw, S. L.;  
42 Zhu, X.; Steiner, D. F.; Davies, N.; Armstrong, D.; Lawrence, C. B.; Luckman, S. M.; Schmitz,  
43 C. A.; Davies, R. A.; Brennand, J. C.; White, A. Agouti-Related Protein Is Posttranslationally  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

- Cleaved by Proprotein Convertase 1 to Generate Agouti-Related Protein (AGRP)83-132: Interaction between AGRP83-132 and Melanocortin Receptors Cannot Be Influenced by Syndecan-3. *Endocrinology* **2006**, *147* (4), 1621–1631. <https://doi.org/10.1210/en.2005-1373>.
- (57) Tota, M. R.; Smith, T. S.; Mao, C.; MacNeil, T.; Mosley, R. T.; Van der Ploeg, L. H.; Fong, T. M. Molecular Interaction of Agouti Protein and Agouti-Related Protein with Human Melanocortin Receptors. *Biochemistry* **1999**, *38* (3), 897–904. <https://doi.org/10.1021/bi9815602>.
- (58) Nix, M. A.; Kaelin, C. B.; Ta, T.; Weis, A.; Morton, G. J.; Barsh, G. S.; Millhauser, G. L. Molecular and Functional Analysis of Human  $\beta$ -Defensin 3 Action at Melanocortin Receptors. *Chem. Biol.* **2013**, *20* (6), 784–795. <https://doi.org/10.1016/j.chembiol.2013.04.015>.
- (59) Ericson, M. D.; Singh, A.; Tala, S. R.; Haslach, E. M.; Dirain, M. L. S.; Schaub, J. W.; Flores, V.; Eick, N.; Lensing, C. J.; Freeman, K. T.; Smeester, B. A.; Adank, D. N.; Wilber, S. L.; Speth, R.; Haskell-Luevano, C. Human  $\beta$ -Defensin 1 and  $\beta$ -Defensin 3 (Mouse Ortholog MBD14) Function as Full Endogenous Agonists at Select Melanocortin Receptors. *J. Med. Chem.* **2018**, *61* (8), 3738–3744. <https://doi.org/10.1021/acs.jmedchem.8b00251>.
- (60) Bolin, K. A.; Anderson, D. J.; Trulson, J. A.; Thompson, D. A.; Wilken, J.; Kent, S. B.; Gantz, I.; Millhauser, G. L. NMR Structure of a Minimized Human Agouti Related Protein Prepared by Total Chemical Synthesis. *FEBS Lett.* **1999**, *451* (2), 125–131. [https://doi.org/10.1016/s0014-5793\(99\)00553-0](https://doi.org/10.1016/s0014-5793(99)00553-0).
- (61) Schibli, D. J.; Hunter, H. N.; Aseyev, V.; Starner, T. D.; Wiencek, J. M.; McCray, P. B.; Tack, B. F.; Vogel, H. J. The Solution Structures of the Human Beta-Defensins Lead to a Better Understanding of the Potent Bactericidal Activity of HBD3 against Staphylococcus Aureus. *J. Biol. Chem.* **2002**, *277* (10), 8279–8289. <https://doi.org/10.1074/jbc.M108830200>.
- (62) Hruby, V. J.; Wilkes, B. C.; Hadley, M. E.; Al-Obeidi, F.; Sawyer, T. K.; Staples, D. J.; de Vaux, A. E.; Dym, O.; Castrucci, A. M.; Hintz, M. F. Alpha-Melanotropin: The Minimal Active

- 1  
2  
3 Sequence in the Frog Skin Bioassay. *J. Med. Chem.* **1987**, *30* (11), 2126–2130.  
4  
5 <https://doi.org/10.1021/jm00394a033>.  
6  
7  
8 (63) Singh, A.; Haslach, E. M.; Haskell-Luevano, C. Structure-Activity Relationships (SAR) of  
9  
10 Melanocortin and Agouti-Related (AGRP) Peptides. *Adv. Exp. Med. Biol.* **2010**, *681*, 1–18.  
11  
12 [https://doi.org/10.1007/978-1-4419-6354-3\\_1](https://doi.org/10.1007/978-1-4419-6354-3_1).  
13  
14  
15 (64) Sousa, S. R.; Vetter, I.; Lewis, R. J. Venom Peptides as a Rich Source of Cav2.2 Channel  
16  
17 Blockers. *Toxins (Basel)*. **2013**, *5* (2), 286–314. <https://doi.org/10.3390/toxins5020286>.  
18  
19  
20 (65) Gasparini, S.; Danse, J. M.; Lecoq, A.; Pinkasfeld, S.; Zinn-Justin, S.; Young, L. C.; de  
21  
22 Medeiros, C. C.; Rowan, E. G.; Harvey, A. L.; Ménez, A. Delineation of the Functional Site of  
23  
24 Alpha-Dendrotoxin. The Functional Topographies of Dendrotoxins Are Different but Share a  
25  
26 Conserved Core with Those of Other Kv1 Potassium Channel-Blocking Toxins. *J. Biol. Chem.*  
27  
28 **1998**, *273* (39), 25393–25403. <https://doi.org/10.1074/jbc.273.39.25393>.  
29  
30  
31 (66) Dauplais, M.; Lecoq, A.; Song, J.; Cotton, J.; Jamin, N.; Gilquin, B.; Roumestand, C.; Vita, C.;  
32  
33 de Medeiros, C. L.; Rowan, E. G.; Harvey, A. L.; Ménez, A. On the Convergent Evolution of  
34  
35 Animal Toxins. Conservation of a Diad of Functional Residues in Potassium Channel-Blocking  
36  
37 Toxins with Unrelated Structures. *J. Biol. Chem.* **1997**, *272* (7), 4302–4309.  
38  
39  
40 (67) Zerbino, D. R.; Birney, E. Velvet: Algorithms for de Novo Short Read Assembly Using de  
41  
42 Bruijn Graphs. *Genome Res.* **2008**, *18* (5), 821–829. <https://doi.org/10.1101/gr.074492.107>.  
43  
44  
45 (68) Grabherr, M. G.; Haas, B. J.; Yassour, M.; Levin, J. Z.; Thompson, D. A.; Amit, I.; Adiconis,  
46  
47 X.; Fan, L.; Raychowdhury, R.; Zeng, Q.; Chen, Z.; Mauceli, E.; Hacohen, N.; Gnirke, A.;  
48  
49 Rhind, N.; di Palma, F.; Birren, B. W.; Nusbaum, C.; Lindblad-Toh, K.; Friedman, N.; Regev,  
50  
51 A. Full-Length Transcriptome Assembly from RNA-Seq Data without a Reference Genome.  
52  
53 *Nat. Biotechnol.* **2011**, *29* (7), 644–652. <https://doi.org/10.1038/nbt.1883>.  
54  
55  
56 (69) Saez, N. J.; Nozach, H.; Blemont, M.; Vincentelli, R. High Throughput Quantitative Expression  
57  
58 Screening and Purification Applied to Recombinant Disulfide-Rich Venom Proteins Produced  
59  
60

- 1  
2  
3 in E. Coli. *J. Vis. Exp.* **2014**, No. 89, e51464. <https://doi.org/10.3791/51464>.
- 4  
5 (70) Vranken, W. F.; Boucher, W.; Stevens, T. J.; Fogh, R. H.; Pajon, A.; Llinas, M.; Ulrich, E. L.;  
6  
7 Markley, J. L.; Ionides, J.; Laue, E. D. The CCPN Data Model for NMR Spectroscopy:  
8  
9 Development of a Software Pipeline. *Proteins* **2005**, *59* (4), 687–696.  
10  
11 <https://doi.org/10.1002/prot.20449>.
- 12  
13  
14 (71) Rieping, W.; Habeck, M.; Bardiaux, B.; Bernard, A.; Malliavin, T. E.; Nilges, M. ARIA2:  
15  
16 Automated NOE Assignment and Data Integration in NMR Structure Calculation.  
17  
18 *Bioinformatics* **2007**, *23* (3), 381–382. <https://doi.org/10.1093/bioinformatics/btl589>.
- 19  
20  
21 (72) Cornilescu, G.; Delaglio, F.; Bax, A. Protein Backbone Angle Restraints from Searching a  
22  
23 Database for Chemical Shift and Sequence Homology. *J. Biomol. NMR* **1999**, *13* (3), 289–302.
- 24  
25  
26 (73) Hutchinson, E. G.; Thornton, J. M. PROMOTIF--a Program to Identify and Analyze Structural  
27  
28 Motifs in Proteins. *Protein Sci.* **1996**, *5* (2), 212–220. <https://doi.org/10.1002/pro.5560050204>.
- 29  
30  
31 (74) Laskowski, R. A.; Rullmann, J. A.; MacArthur, M. W.; Kaptein, R.; Thornton, J. M. AQUA  
32  
33 and PROCHECK-NMR: Programs for Checking the Quality of Protein Structures Solved by  
34  
35 NMR. *J. Biomol. NMR* **1996**, *8* (4), 477–486.
- 36  
37  
38 (75) Zhang; Chung; Oldenburg. A Simple Statistical Parameter for Use in Evaluation and Validation  
39  
40 of High Throughput Screening Assays. *J. Biomol. Screen.* **1999**, *4* (2), 67–73.  
41  
42 <https://doi.org/10.1177/108705719900400206>.
- 43  
44  
45 (76) Cheng, Y.; Prusoff, W. H. Relationship between the Inhibition Constant (K<sub>1</sub>) and the  
46  
47 Concentration of Inhibitor Which Causes 50 per Cent Inhibition (I<sub>50</sub>) of an Enzymatic Reaction.  
48  
49 *Biochem. Pharmacol.* **1973**, *22* (23), 3099–3108. [https://doi.org/10.1016/0006-2952\(73\)90196-](https://doi.org/10.1016/0006-2952(73)90196-2)  
50  
51 [2.](https://doi.org/10.1016/0006-2952(73)90196-2)
- 52  
53  
54 (77) Peigneur, S.; Cheneval, O.; Maiti, M.; Leipold, E.; Heinemann, S. H.; Lescrinier, E.; Herdewijn,  
55  
56 P.; De Lima, M. E.; Craik, D. J.; Schroeder, C. I.; Tytgat, J. Where Cone Snails and Spiders  
57  
58 Meet: Design of Small Cyclic Sodium-Channel Inhibitors. *FASEB J.* **2019**, *33* (3), 3693–3703.  
59  
60



<https://doi.org/10.1096/fj.201801909R>.

- (78) King, G. F.; Gentz, M. C.; Escoubas, P.; Nicholson, G. M. A Rational Nomenclature for Naming Peptide Toxins from Spiders and Other Venomous Animals. *Toxicon* **2008**, *52* (2), 264–276.

<https://doi.org/10.1016/j.toxicon.2008.05.020>.

#### Web Sites:

Clinical Trial. <https://clinicaltrials.gov> (accessed on Sep 21, 2019).

Orphanet. <http://www.orpha.net> (accessed on Sep 21, 2019).

Protein Data bank. <http://www.pdb.org> (accessed on Sep 21, 2019).

UniProt. <https://www.uniprot.org/> (accessed on Oct 21, 2019).

VenomZone. <https://venomzone.expasy.org/> (accessed on Oct 21, 2019).

#### Table of Contents Graphic

