Conkazal-M1 from the MKAVA family of conotoxins – a dual-function protease inhibitor and neuroactive peptide

Celeste M. Hackney[1], Thomas Lund Koch[2,3], Nicklas Lund Ryding[1], Aymeric Rogalski[2], Kevin Chase[4], Matías Leonel Giglio[4], Samuel S. Espino[4], Zildjian G. Acyatan[2], Maren Watkins[4], Baldomero M. Olivera[4], Helena Safavi-Hemami[2,4], Kaare Teilum[1]*, and Lars Ellgaard[1]*

Affiliations:

[1] Department of Biology, Linderstrøm-Lang Centre for Protein Science, University of Copenhagen, Copenhagen, Denmark

[2] Department of Biochemistry, University of Utah, Salt Lake City, Utah, United States of America

[3] Department of Biomedical Sciences, University of Copenhagen, Denmark

[4] School of Biological Sciences, University of Utah, Salt Lake City, Utah, United States of America

***Correspondence:** Lars Ellgaard, University of Copenhagen, Department of Biology, Ole Maaløes Vej 5, DK-2200 Copenhagen N., Denmark. Email: lellgaard@bio.ku.dk

Kaare Teilum, University of Copenhagen, Department of Biology, Ole Maaløes Vej 5, DK-2200 Copenhagen N., Denmark. Email: kaare.teilum@bio.ku.dk

20 Keywords: Cone snails, conkazal, conotoxin, Kazal-type protease inhibitor, Kunitz domain,
21 molecular evolution, schistosomin

ABSTRACT

Marine cone snails produce a diverse array of bioactive peptides, known as conotoxins, in their venom. Given their high target potency and specificity, conotoxins are attractive compounds for the development of precision research tools and pharmacological agents. Here, we provide the first experimental characterization of a conotoxin from the MKAVA superfamily, conkazal-M1, from *Conus magus*. Using NMR spectroscopy, we show that conkazal-M1 adopts a fold characteristic of the Kazal-type protease inhibitor family, featuring a Glu residue at the inhibitory P1 position. Recombinantly expressed conkazal-M1 inhibits the proteolytic activity of Subtilisin A with an apparent Ki of 1.1 μM. In addition, conkazal-M1 partially inhibits calcium transients in mouse sensory neurons, suggesting a potential role in modulating ion-channel activity as seen for many other toxins. The dual function of conkazal-M1 in protease inhibition and neuroactivity is analogous to the dual function of several toxins harboring a Kunitz-type fold. The well-conserved sequence of the MKAVAs indicates an evolutionary trajectory where these proteins face an adaptive conflict when mutations that enhance one activity compromise the other. Collectively, this work provides new structural and functional insight into a previously uncharacterized toxin superfamily in cone snails and expands our knowledge about the toxin arsenal available to venomous animals.

41 **INTRODUCTION**

42 Peptide toxins from animal venoms target a broad range of biological processes to subdue prey or
43 deter predators (1). To exert these effects, toxins act by disrupting the cellular ion balance or by
44 interfering with signaling pathways mediated by transmembrane proteins such as ion channels, G
45 protein–coupled receptors (GPCRs), transporters, and tyrosine kinases (2). Other major classes of
46 toxins target hemostatic processes, such as blood coagulation (3). Due to the evolutionary
47 conservation of such key processes and pathways, many peptide toxins hold great potential as
48 research tools and potential pharmaceutical compounds.

49 The diverse functions of peptide toxins are enabled by a variety of molecular structures.
50 Many of these toxins fold into small, compact structures stabilized by multiple disulfide bonds.
51 While the cysteine residues involved in disulfide-bond formation are typically conserved, inter-
52 cysteine residues normally undergo rapid evolution that generate diversity (4, 5). Consequently,
53 toxin sequences – even within the same evolutionary family – often exhibit a high degree of
54 sequence variation.

55 Several prevalent toxins function as protease inhibitors, reflecting the critical role of
56 protease activity in the physiology of target animals. For example, venom protease inhibitors target
57 proteases in the blood-coagulation system (causing internal bleeding and hypotension), enzymes
58 of the cholinergic system (disrupting neurotransmitter metabolism), and the immune system
59 (modulating immune and inflammatory responses). Common protease-inhibitor folds present in
60 venoms include Kazal-, Kunitz- and cystatin-like domains, Snake Venom Serine Protease
61 Inhibitors, and Cysteine-Rich Secretory Proteins. As is common for other venom peptides, such
62 folds are often recruited into the venom gland from endogenous genes and are subsequently
63 "weaponized" (6). Through this process, these compact and stable structural scaffolds can be

64 repurposed to perform diverse functions that differ from the original, while retaining their overall
65 structure.

66 Well-characterized Kunitz-type toxins are known to have acquired new functions different
67 from the original protease-inhibitor function through evolution. For instance, the conkunitzins
68 from venomous marine cone snails (7, 8), as well as mambaquaretin-1 (9) and calcicludine (10)
69 from the Eastern green mamba, have evolved to inhibit ion channels or receptors. Moreover,
70 certain Kunitz-like toxins display a dual function as both protease inhibitors and ion channel
71 blockers (11-14). At the molecular level, the dual function of the tarantula Kunitz-type toxin
72 HWTX-X is achieved through two distinct binding interfaces (12). It has also been shown that
73 Kunitz-type protease inhibitors from venomous snakes undergo positive selection at sites that
74 cover 30-50% of the domain (15). These examples provide textbook illustrations of the sequence
75 plasticity often observed in toxins that allows them to adopt new functions using a preexisting
76 structural scaffold.

77 A prominent example of a protease-inhibitor fold found in several venoms is the Kazal-
78 type domain that inhibits a broad range of serine proteases. Kazal-type venom peptides are found
79 in various venomous animals including assassin bugs, shrew, honeybees, and wasps, where they
80 have been associated with functions like anti-coagulation (16-18), pro-coagulation (19),
81 antimicrobial activity (18), melanization in insects (20), and analgesia (19). In humans, the SPINK
82 family of Kazal-type protease inhibitors (KPIs) play important functions in regulating cellular
83 protease activity, as exemplified by the pancreatic trypsin inhibitor SPINK1, where sequence
84 variants are associated with pancreatitis and pancreatic cancer (21). Another example of a well-
85 characterized KPI is greglin from the locust *Schistocerca gregaria*, which inhibits both elastase
86 and subtilisin with low nanomolar Ki-values (22). In general, KPIs function as substrate analogs.

87 They act by inserting their inhibitory P1 residue, the residue positioned immediately adjacent to
88 the scissile bond, into the specificity pocket of the target protease (23). This residue is typically
89 located two positions C-terminally to the second cysteine in the sequence on the so-called reactive-
90 site loop (23). The nature of the P1 residue can vary considerably, in turn influencing substrate
91 specificity – for instance, trypsin-inhibiting KPIs often contain a lysine or arginine residue at the
92 P1 position to mimic physiological trypsin substrates (23). Classical KPIs harbor 40-60 amino acid
93 residues and fold into a structure that comprises an α-helix and a three-stranded β-sheet and is held
94 together by three disulfide bonds (23). KPIs that deviate from this definition (e.g. more or fewer
95 disulfides or varying disulfide patterns) are classified as non-classical KPIs (see Results for more
96 details).

97 The conotoxins produced by venomous marine cone snails are peptides encoded as precursor
98 sequences that comprise a well-conserved N-terminal signal sequence that directs the peptide to
99 the endoplasmic reticulum followed in some cases by a propeptide of intermediate sequence
100 conservation, and, finally, the mature conotoxin sequence. Approximately 70 different conotoxin
101 gene families have presently been identified (24-26). While most mature conotoxins are 15-45
102 residues in length, many larger peptides also exist (25). Many families of large conotoxins – also
103 known as macro-conotoxins (27) – remain uncharacterized due to the difficulty in producing large
104 disulfide-bonded peptides. These larger peptides can be expected to perform functions unavailable
105 to shorter conotoxins, e.g., as enzymes or as a consequence of interacting with their targets through
106 a larger binding interface.

107 Here, we surveyed venom gland transcriptomes to identify new macro-conotoxin families.
108 We describe the previously uncharacterized MKAVA family of conotoxins and show by NMR
109 spectroscopy that an MKAVA-family toxin from *Conus magus*, conkazal-M1, comprises a single

110 non-classical Kazal-type domain with both protease inhibitor activity and the ability to inhibit
111 KCl-induced calcium influx in a subset of dorsal root ganglia (DRG), as also seen for several
112 toxins harboring a Kunitz-type fold.

113

RESULTS

The MKAVA conotoxin superfamily is widely distributed among venomous cone snails

To characterize the MKAVA conotoxin superfamily, we mined publicly available venom gland transcriptomes from 49 cone snail species and identified 89 transcripts sharing the MKAVA consensus signal sequence (MKAVAVFLVVALAVAYG) with slight variations (Figure S1; all sequences provided in Supplementary File 1). In addition to their wide distribution across *Conus* species, we found that closely related MKAVA-like peptides are also found in other members of Conoidea, the larger superfamily of marine snails to which cone snails belong. Searches against the NCBI Sequence Read Archive (SRA) database identified similar sequences in species belonging to Drillidae, Turridae, Profundiconus, and Terebridae (see Materials and Methods section for SRA entry numbers). It should be noted that these sequences are incomplete, as they derive from unassembled short-read datasets.

In *Conus*, all identified sequences lack an apparent propeptide with the predicted mature toxin sequence (~73 residues) following immediately after the signal sequence. The toxins exhibit a high level of conservation (Figure 1a) – within the mature toxin sequence, 32 residues are >95% conserved (including all eight cysteines) and approximately 44 positions in an alignment of all identified full-length sequences show ≥75% identity (Figure S1). Selection analyses show that approximately 95% of sites evolved under negative to neutral selection (from comparisons of the M2 and M1a models and the M8 and M7 models in the Phylogenetic Analysis by Maximum Likelihood (PAML) package (28)), while only three sites (T38, K46 and K48) evolved under weak positive selection with >90% likelihood based on dN/dS ratios (see Supplementary File 2).

Generally, venom gland transcripts that are highly expressed across multiple species likely encode peptides that function as toxins. Analysis of the expression levels of MKAVA transcripts

137 revealed that most cone snail species express two transcripts: one with a high level of expression
138 (100–several thousand transcripts per million (tpm)) and the other with a low level of expression
139 (<100 tpm) (Figure 1b). Phylogenetic analysis of transcripts with high and low expression revealed
140 a well-resolved clustering pattern consistent with their relative expression (Figure 1c). The
141 bipartite nature of the gene tree is reflected at the sequence level, where specific positions
142 distinguish toxins encoded by either high- or low-expressing transcripts (Figure 1a). The high
143 expression of many MKAVA conotoxin-encoding transcripts support a function of these as toxins.
144 This is further corroborated by their presence in the venom. Using publicly available tandem mass
145 spectrometry data of the venom of *Conus geographus* and *Conus marmoreus*, we identify several
146 tryptic peptides corresponding to MKAVA sequences from these species (Figure S2).

147

148 **The MKAVAs are evolutionarily related to a schistosomin**

149 To identify possible homologs of MKAVAs outside of Conoidea, we used a high-expressing
150 precursor sequence from *C. magus* to perform homology searches against the NCBI non-redundant
151 protein database and transcriptome shotgun assemblies. Given that we find this toxin from *C.*
152 *magus* to display a KPI structure (see below), and that it is the first characterized protein from its
153 family, we name this toxin "conkazal-M1" (where M stands for *magus*). Likewise, we will use the
154 term "conkazal", rather than MKAVA, for this family of conotoxins from here on.

155 The searches revealed sequence similarity, including conservation of all eight cysteines,
156 with schistosomin-like proteins from several molluscs (Figures 2a and S3), but seemingly absent
157 in other phyla. Schistosomin was first described in the freshwater pond snail, *Lymnaea stagnalis*,
158 as a peptide upregulated in response to infection with the schistosome parasite *Trichobilharzia*

159 *ocellata* (29) (see also Discussion). Overall, modest sequence conservation is observed across the
160 conkazals and schistosomins. However, in addition to the cysteines, three residues show high
161 conservation. These comprise a glutamate residue located two residues after the second cysteine
162 in both groups of proteins, i.e. at the P1 position of KPIs, at position 34 in the alignment, a proline
163 immediately following the first Cys residue, and a positively charged residue at position 62 in the
164 alignment (Figure 2a and S3).

165 A CLANS (CLuster ANalysis of Sequences) clustering analysis (30), which groups protein
166 families based on all-against-all pairwise protein sequence similarities, revealed a clear similarity
167 of the *Conus*-derived conkazal sequences with molluscan schistosomins from gastropods (Figure
168 2b; all sequences used in this analysis are provided in Supplementary File 3). Moreover, gene
169 structure analysis revealed a common gene structure of conkazals and molluscan schistosomins.
170 The genes have a common phase-1 intron located immediately downstream of the signal sequence
171 and a conserved phase-2 intron located between the sixth and seventh cysteine residues (Figure
172 2c). The common gene structures further support an evolutionary relationship between conkazal
173 conotoxins and schistosomins.

174 While molluscan species exhibit a high degree of intrachromosomal gene translocation,
175 macrosynteny tends to be conserved. We therefore compared the genomic location of conkazal
176 genes from *C. ventricosus* with schistosomins from representative gastropods (Figure S4). The two
177 conkazal genes in *C. ventricosus* are adjacent to each another on scaffold 23, which represents one
178 of the chromosome-scale scaffolds in the cone snail genome. This chromosome shows homology
179 to chromosome NC_037594.1 of the gastropod *Pomacea canaliculata* and chromosome
180 NC_090294.1/ NC_090290.1 of another gastropod *Biomphalaria glabrata* (Figure S4). However,
181 neither of these chromosomes contain the schistosomin genes, which are located on NC_037595.1

182 and NC_074716.1, respectively. Surprisingly, both *P. canaliculata* and *B. glabrata* have multiple
183 schistosomin gene copies located in tandem though the neighboring genes show no syntenic
184 relationship (Figure S4b). Overall, while conkazals show a high level of similarity to
185 schistosomins, strongly suggesting an evolutionary relationship, the genes are not part of a
186 common gene linkage group.

187

188 **Conkazal-M1 can be produced in the fully oxidized form with the csDisCoTune *E. coli***
189 **expression platform**

190 To gain insight into the structure and function of the conkazal superfamily, conkazal-M1 from the
191 magician cone, *C. magus*, was selected for recombinant expression and characterization.
192 Conkazal-M1 was recombinantly expressed in *E. coli* using the csDisCoTune system (31). This
193 expression platform, a recent further development of the original CyDisCo system (32) and our
194 own csCyDisCo system (33), is tailored specifically for the expression of conotoxins and allows
195 for the production of correctly folded, fully oxidized disulfide-rich proteins in the bacterial cytosol
196 by co-expression of three foldases encoded by an auxiliary plasmid – the Erv1p oxidase, human
197 protein disulfide isomerase (hPDI), and a conotoxin-specific protein disulfide isomerase (csPDI)
198 (34). Conkazal-M1 was expressed as a fusion protein with an N-terminal ubiquitin (Ub) tag
199 containing an internal loop modified to contain ten consecutive histidine residues (His_{10}) and a
200 tobacco etch virus protease (TEVp) recognition site, allowing for release of the mature toxin (35).

201 SDS-PAGE analysis of small-scale test expressions of Ub-His_{10}-conkazal-M1 alone, Ub-
202 His_{10}-conkazal-M1 with the csDisCoTune-encoded foldases, or with the foldases encoded by the
203 csCyDisCo expression platform clearly showed the beneficial effect of both expression systems

204 on Ub-His_{10}-conkazal-M1 production. Both systems resulted in the expression of soluble Ub-
205 His_{10}-conkazal-M1 fusion protein, in contrast to expression without the foldases where equivalent
206 total protein levels were observed, but the fusion protein was predominantly found in the insoluble
207 fraction in a reduced and/or partially oxidized state (Figure S5a). Because the csDisCoTune
208 showed higher levels of oxidized Ub-His_{10}-conkazal-M1 compared to the csCyDisCo system,
209 large-scale expression was carried out with csDisCoTune. Cleavage of affinity-purified Ub-His_{10}-
210 conkazal-M1 by TEVp resulted in liberation of the 73-residue mature region of conkazal-M1,
211 which was purified (>95% purity) to a final yield of just over 5 mg per liter culture (Figure S5b).

212 Disulfide-bond formation in the purified toxin was verified by SDS-PAGE analysis, where
213 a clear downward mobility shift was observed upon addition of a reducing agent (Figure S5c), as
214 well as Q-TOF mass spectrometry, which revealed the presence of two species in the purified
215 preparation. The primary species displayed a monoisotopic mass of 7990.54 Da in accordance with
216 the theoretical mass (7990.55 Da) of the fully oxidized protein (Figure S5d). The secondary
217 component exhibited a monoisotopic mass of 7973.49 Da, i.e., 17 Da lower than the theoretical
218 mass. We attribute this reduction in mass to modification of the N-terminal Gln to a pyroglutamate,
219 a post-translational modification commonly found in conotoxins (36). This modification can be
220 enzymatically catalyzed by glutaminyl cyclase but is also known to occur spontaneously (37).

221

222 **The NMR structure of conkazal-M1 reveals a Kazal-type protease inhibitor fold**

223 The three-dimensional structure of ^{15}N and ^{13}C stable isotope-labelled conkazal-M1 was solved in
224 solution using distance and torsion angle restraints obtained by NMR spectroscopy. We assigned
225 the chemical shifts of 79% of the backbone ^{1}H, ^{13}C and ^{15}N resonances and 66% of sidechain

226 resonances. In total 69% of the chemical shifts were assigned. We observed no resonances from
227 the first seven residues. 602 short distances assigned in NOESY spectra were combined with 100
228 dihedral angles calculated from chemical shifts and used as restraints in structure calculations
229 (Table 1). The structure is well ordered from residue 10 to the C-terminus with an RMSD value of
230 0.31Å for the positions of the backbone atoms in the 20 models that represent the structure (Figure
231 S6).

232 The structure consists of a central three-stranded antiparallel β-sheet and two α-helices; α-
233 helix 1 that is parallel to the β-strands and α-helix 2 that is almost perpendicular to α-helix 1
234 (Figure 3a and b). The N- and C-termini are tethered to the core structure via two disulfide bridges.
235 Located at each end of α-helix 1, Cys40 and Cys48 form disulfide bonds with the N-terminal
236 Cys14 and Cys4, respectively. The C-terminus is affixed to the short loop connecting β-strand 2
237 and α-helix 1 by the bridge connecting Cys37 and Cys68. The last disulfide bond is formed by
238 Cys24 and Cys59, thus connecting β-strand 1 and α-helix 2.

239 As structural similarity often correlates with conserved functional properties, we used
240 FoldSeek (38) and PDB eFold (39) to identify proteins structurally similar to conkazal-M1. Both
241 search engines identified several Kazal-type protease inhibitors among the most similar structures
242 to conkazal-M1, with the AlphaFold-predicted structure of a schistosomin from *Lymnaea stagnalis*
243 found as the top hit in the FoldSeek search with conkazal-M1. We also identified an unpublished
244 crystal structure of schistosomin from another freshwater snail, *Biomphalaria glabrata* (PDB ID:
245 9FDO), in the PDB. Conkazal-M1 and this schistosomin (Figure 3a) superpose with an RMSD
246 value of 1.9 Å over 38 Cα-atoms (Figure 3b). Both structures harbor the KPI core structure as well
247 as an additional C-terminal α-helix not found in classical KPIs. The three characteristic disulfide

248 bridges of the classical KPIs are present in both conkazal-M1 and schistosomin and are similarly
249 positioned in the two structures (blue arrowheads, Figure 3b). Both proteins harbor a fourth
250 disulfide bridge that tethers the C-terminal regions to the core structure of the proteins (Figure 3a).
251 To illustrate the structural similarity between conkazal-M1 and a verified KPI, Figure S7 shows
252 an overlay of conkazal-M1 with the crystal structure of greglin (40). Similar to the other structures,
253 greglin contains a fourth disulfide bond, but its position differs from that in conkazal-M1 and
254 schistosomin (Figure S7; orange).

255

256 **Conkazal-M1 inhibits the proteolytic activity of Subtilisin A**

257 Because conkazal-M1 adopts a Kazal fold, we speculated that the toxin may exhibit inhibitory
258 activity against serine proteases. Consistent with this hypothesis, Glu15 at the P1 position in
259 conkazal-M1 is solvent exposed in all 20 of the lowest energy structures (Figure S6). We therefore
260 investigated the proteolytic activities of chymotrypsin, Endoproteinase GluC and Subtilisin A
261 (SubA) in the presence of increasing amounts (0.25 μM – 3 μM) of conkazal-M1. GluC was chosen
262 as a possible candidate due to its preferential cleavage of peptide bonds C-terminal to glutamic
263 acid residues. SubA is a non-specific protease that favors cleavage after hydrophobic residues but
264 is not restricted to them. For instance, a Glu residue at the P1 position in the second domain of the
265 CrSPI two-domain Kazal-type protein from horseshoe crab is the primary residue for mediating
266 the contacts with subtilisin in the co-crystal of the complex between the two proteins (41).

267 SubA activity was monitored using the chromogenic substrate N-succinyl-Ala-Ala-Pro-
268 Phe-pNA over a range of substrate concentrations in the presence and absence of conkazal-M1.
269 Initial rates were estimated from changes in absorbance at 405 nm over time and were converted

270 to reaction rates (μM/min/mg enzyme). Nonlinear regression of the initial rate data was performed
271 using Michaelis-Menten and competitive, non-competitive and uncompetitive inhibition models.
272 Model comparison by Akaike Information Criterion indicated that competitive inhibition best
273 described the data – as expected for a KPI – with estimated kinetic parameters of Km=201 μM
274 (Confidence Interval (C.I.): 173‑234 μM), Vmax= 12,888 μM/min/mg (C.I.: 12,357-13,456
275 μM/min/mg) and Ki=1.1 μM (C.I.: 0.92-1.5 μM) (Figure 4). We performed similar investigations
276 with chymotrypsin and Endoproteinase GluC in the absence and presence (0.5 μM-2 μM) of
277 conkazal-M1 where no effect of the toxin on enzyme activity was discerned (Figure S8).

278

279 **Conkazal-M1 modulates calcium transients in mouse sensory neurons**

280 A wide variety of venom peptides function as neurotoxins that target neuronal ion channels.
281 Therefore, we next assessed the bioactivity of conkazal-M1 using the calcium imaging-based
282 constellation pharmacology assay (Teichert et al, 2012). This assay enables the monitoring of Ca^{2+}
283 influx in primary cell cultures obtained from mouse DRGs, which comprise different types of
284 nociceptors, mechanoreceptors, and glia. During each experiment, intracellular calcium levels
285 were simultaneously monitored in 500-1,000 cells using the Fura-2-AM dye. Cells were assigned
286 to seven major classes based on their size, isolectin B4 staining (non-peptidergic nociceptors) and
287 calcitonin gene-related peptide (CGRP) expression (peptidergic nociceptors) (Fang et al, 2010).
288 To further characterize neuronal subtypes, a series of pharmacological agents (conotoxin *κ*M-
289 RIIIJ, allyl isothiocyanate [AITC], menthol, and capsaicin) were applied at the end of each
290 experiment (Giacobassi et al, 2020).

291 Dissociated DRG neurons were depolarized using 15 s pulses of high-potassium
292 extracellular solution (30 mM KCl), resulting in intracellular Ca^{2+} elevation, as indicated by the
293 rise in Fura-2 signal. At the lowest concentration tested (1 μM), conkazal-M1 did not produce
294 significant effects on any neuronal subclasses. In contrast, application of 10 μM conkazal-M1
295 reduced the KCl–evoked calcium signals in a subset of cells (12.2% of the 1,494 cells analyzed)
296 (Figure 5a). The blockade of calcium transients was predominantly observed in small (<400 μm^2)
297 peptidergic nociceptive neurons (34.1% of affected cells) (identified by the expression of CGRP
298 genetically labeled with green fluorescent protein) (Figure 5b). Other neuronal subclasses affected
299 by 10 μM conkazal-M1 included large-diameter (>400 μm^2) mechanoreceptors (18.7%), a subset
300 of non-peptidergic nociceptors (15.9%), and C-low threshold mechanoreceptors (C-LTMR;
301 14.3%). Thus, conkazal-M1 inhibits KCl-induced calcium influx in a subset of neuronal cells,
302 albeit with low affinity.

303 Lastly, because conkazal-M1 originates from the venom of *C. magus*, a fish-hunting cone
304 snail, we hypothesized that the toxin might elicit measurable effects in a fish model. To test this,
305 we injected goldfish (*Carassius auratus*) with 5 or 10 nmol of toxin via both intraperitoneal and
306 intramuscular routes (N=3 for each group) and monitored them closely for a period of three hours
307 post-injection. No overt phenotypic effects were observed in either treatment group compared with
308 the controls (data not shown).

309

310 **DISCUSSION**

311 Here, we provide the first characterization of a toxin from the conkazal conotoxin family, conkazal-
312 M1. The conkazals constitute a large and widespread toxin family found across phylogenetically
313 diverse *Conus* species and several other species in the Conoidean superfamily, indicating an
314 important functional role of these peptides. The common evolutionary history with the
315 schistosomins indicates that an ancestral gene was recruited into the venom in early Conoidea and
316 then acquired neuroactive functions. However, given that the function of schistosomins remains
317 unclear, the exact process of recruitment is presently unknown.

318 Many conkazal transcripts are highly expressed in the venom gland, and we confirmed
319 their presence in venom by proteomic analysis, supporting a role as toxins. We show that conkazal-
320 M1 can perform dual functions as a protease inhibitor and as a partial inhibitor of KCl-stimulated
321 calcium influx in sensory neurons. Such dual activities have also been reported in Kunitz-type
322 venom proteins (11-14). In the context of constellation pharmacology, the partial suppression of
323 KCl-evoked calcium transients by conkazal-M1 provides preliminary evidence that this peptide
324 can modulate ion channel activity in sensory neurons. Although the effect was observed in only a
325 subset of cells, the preferential targeting of small-diameter, CGRP-positive nociceptors suggests
326 that conkazal-M1 may interact with molecular components enriched in these neurons, such as
327 high-voltage-activated calcium channels (*e.g.*, Cav2.2 or Cav2.3) or auxiliary subunits that shape
328 their kinetics (42-44). The modest inhibition at 10 μM indicates either a low-affinity or indirect
329 mechanism, potentially consistent with dual-function peptides in which ion-channel modulation is
330 a secondary activity to protease inhibition, as further discussed below. Future studies are needed
331 to identify the target and mode of action conkazal-M1 in neuronal cells.

332

333 In the context of protease inhibitor function, subtilisin-targeting KPIs with a Glu at the P1
334 position have previously been characterized (see e.g., (41, 45-47)). It is therefore unsurprising that
335 conkazal-M1 with its Glu15 at the P1 position was able to inhibit subtilisin (although additional
336 residues in the reactive-site loop are likely also to contribute to substrate specificity). While certain
337 venoms appear to harbor microbiomes (48), bacterial subtilisin is an unlikely physiological target
338 of the protease-inhibitory function of conkazal-M1. This is also reflected in the relatively high K_i
339 value of conkazal-M1 for subtilisin when compared to many other subtilisin inhibitors, such as
340 greglin. In principle, in its capacity as a protease inhibitor, conkazal-M1 could perform a function
341 either in the venom (to regulate the activity of venom proteases to maintain venom stability before
342 delivery) or in the prey to exert a deleterious effect upon delivery. At present, we cannot distinguish
343 between these possibilities although we note that we have not been able to find evidence in the
344 literature for an endogenous regulatory function of protease inhibitors in other venomous
345 organisms.

346 When comparing conkazal sequences with schistosomins from other molluscs, Glu15 at
347 the P1 position is one of the few residues conserved across species (Figure 2). This observation
348 supports to the hypothesis that the conkazals and mollusk schistosomins function as KPIs. In
349 addition to the eight cysteine residues, only two other residues are as highly conserved as Glu15:
350 a proline residue immediately following the first Cys residue, likely playing a structural role, and
351 a surface-exposed lysine residue at a position corresponding to Lys43 in the mature conkazal-M1
352 sequence, which is most likely to have a functional role.

353 Schistosomiasis, a parasitic infection, is caused by flatworms known as schistosomes. In
354 the schistosomiasis life cycle, freshwater snails serve as intermediate hosts that release infectious
355 larvae into the water. These larvae can penetrate human skin, gaining access to the host, where the

356 subsequent release of eggs can cause a range of symptoms, including serious damage to organs
357 such as the liver and bladder. In the snail-schistosome system *L. stagnalis*-*T. ocellata*, schistosomin
358 was initially proposed to be a neuropeptide induced by parasite infection, but subsequent studies
359 showed that it is produced in connective tissue cells and in hemocytes (reviewed in (49)). In the
360 hemolymph, schistosomin was proposed to act as an antagonist of gonadotropic hormones thereby
361 inhibiting reproductive activity in snails. However, in the *Biomphalaria glabrata*-*Schistosoma*
362 *mansoni* system, no correlation was found between parasite infection and schistosomin levels and
363 infection did not result in parasitic castration (50). Thus, the physiological function of schistosomin
364 remains unclear, and further investigation is warranted, particularly with its potential function as
365 a KPI in mind.

366 An ancestral protease inhibitor that acquires the ability to modulate neuronal activity while
367 retaining protease inhibition represents a functional change in which purifying selection preserves
368 the original function, while weak-to-moderate selection can favor the emergence of a novel activity
369 (6, 51). In contrast, a complete shift toward modulation of neuronal activity accompanied by partial
370 or total loss of protease activity can be explained by strong positive selection for the new function,
371 coupled with relaxed constraint on the ancestral activity. Because we detect no evidence of positive
372 selection, we favor the first evolutionary model. In this scenario, conkazals would be
373 "moonlighting" proteins that face an adaptive conflict when mutations that enhance one activity
374 compromise the other. Such conflicts are typically resolved through gene duplication followed by
375 subfunctionalization that lead to partitioning of the ancestral roles between paralogs or to
376 neofunctionalization of one copy. Our data do not support a neofunctionalization event. If such an
377 event had occurred, we would expect to detect both an ancestral, schistosomin-like gene and a
378 duplicated conkazal gene. Instead, we only observe conkazal genes when interrogating available

379 cone snail genome data, consistent with a moonlighting role. This interpretation is further
380 supported by the expression of conkazals in tissues other than the venom gland, including the
381 venom bulb, nerve ring, and foot (see Materials and Methods section for SRA entry numbers),
382 where these peptides likely retain an endophysiological, ancestral function.

383 We note that multi-domain Kazal-type proteins are found in many venomous animals,
384 including the two-domain Kazal-type toxins BQTX from shrews (19) and Rhodniin from the
385 kissing bug *Rhodnius prolixus* (17, 52). In general, secreted cysteine-rich repeat-proteins (53) –
386 including a broad range of venom peptides – are common in nature and have been cataloged in the
387 ScrepYard database (54). Their multi-domain structure enables multivalent substrate interactions,
388 allowing them to bind targets strongly through an avidity effect. Indeed, Rhodniin binds its
389 substrate thrombin through multivalent interactions involving the reactive-site loop of one domain
390 and electrostatic interactions of the second domain, together conferring a picomolar affinity (17,
391 52). Analogous to the two-domain Kazal-type toxins described above, we have recently identified
392 a family of such conotoxins, the MKIAV family. As shown in Figure S9a-b, both Kazal domains
393 of these conotoxins, with few exceptions, contain a Glu or Asp residue three positions C-terminal
394 to the second Cys in the mature sequence. Moreover, each MKIAV protein has a Glu/Asp at this
395 position in one of the two kazal-type domains. Although not located at the canonical P1 position,
396 we speculate that this residue could still perform the inhibitory function of a P1 residue. The
397 AlphaFold 3 predictions of the MKIAVs reveal two very similar Kazal-type domains connected
398 by a short linker, with each domain closely resembling the structure of conkazal-M1 (Figure S9c-
399 e).

400 In this study, we identify two previously unrecognized families of conotoxins, conkazal
401 and MKIAV, which encode single- and two-domain Kazal-type domain structures, respectively.

402 We uncover the dual function of conkazal-M1 as a neuroactive peptide with the ability to inhibit
403 calcium transients in a subset of DRGs and as a protease inhibitor. Together, these observations
404 highlight the structural and functional diversity of Kazal-type toxins in *Conus*, extend the current
405 knowledge about the toxin complement available to venomous animals, and illustrate the adaptive
406 conflict that can shape the evolutionary trajectories of dual function toxins.

407

408

409

410 **MATERIALS AND METHODS**

411 **Sequence mining and analysis**

412 Multiple sequence alignments were generated with the MAFFT version 7 (55) online interface
413 using the G-INSi method (56). Alignments were visualized in JalView (57). Schistosomin
414 sequences were obtained using web version of blast to query against the non-redundant (nr) protein
415 database and transcriptome shotgun assemblies with standard settings. Conotoxin sequences were
416 obtained from (25). These sequences were used as input for CLANS (30), which is a sequence-
417 based clustering program that assigns attractive values between sequences based on their
418 BLOSUM62 blastp p-value. In a uniformly repelling force field this causes similar sequences to
419 cluster together. Gene structure analyses were performed with using the protein2genome (p2g)
420 model in Exonerate (v.2.76.2) (58). Macrosynteny analysis was performed with macrosyntR using
421 OrthoFinder to assign orthologous proteins across *C. ventricosus* (GCA_018398815.2), *P.*
422 *canaliculata* (GCF_003073045.1), and *B. glabrata* (GCF_947242115.1). The local gene
423 neighborhoods were assessed and visualized in Figure 2C. Maximum likelihood gene trees were
424 constructed using alignments generated with MAFFT v.7.526 and the tree was constructed using
425 IQ-TREE v. 2.2.2.3 (59) with UF bootstrap (60) calculation for 1000 replicates. The best model
426 according to the Bayesian Information Criterion was calculated to be TMP3 + G4. The consensus
427 tree was visualized using FigTree (61). Conkazal-like sequences were identified in the NCBI SRA
428 database using the conkazal-M1 amino acid sequence as a query sequence in tblastn searches that
429 retrieved the following SRA entries: SRX8002177, SRR11423812 (Drillidae), SRR26897535,
430 SRR26897526 (Turridae), SRR10192879, SRR10192877 (Profundiconus), SRR2060989
431 (Terebridae), SRR16493587 (venom bulb), SRR16493599, SRR16493598 (nerve ring), and
432 SRR16493591 (foot).

433 **Proteomics-based identification of MKAVA peptides in venom**

434 Proteomics files of reduced, alkylated and trypsin digested venom of *Conus geographus*
435 (accession: PXD052869) and *Conus marmoreus* (PXD038992) were retrieved from the PRIDE
436 ProteomeXChange repository and searched for MKAVA sequences using the Byonic software
437 (version 3.2.0 Protein metrics) with the following settings: cleavage site: RK, missed cleavage: 1,
438 Modifications: carbamidomethyl (C, +57.02146), Oxidation (M,P, +15.994915), pyro-Glu (Q,
439 +17.026549), Carboxy (E, +43.989829). High scoring peptide hits (score> 400) were manually
440 verified.

441

442 **Plasmid generation**

443 The codon-optimized sequence encoding mature conkazal-M1 was synthesized and inserted
444 (TWIST Bioscience) into the pET39_Ub19 expression vector (35). The resulting plasmid
445 (pLE657) encodes ubiquitin (Ub)-His_{10}-tagged conkazal-M1 (Ub-His_{10}-conkazal-M1) with a
446 recognition site for TEV protease located immediately N-terminally to the mature toxin sequence.

447

448 **Protein expression and preparation of lysate**

449 Chemically competent BL21(DE3) were co-transformed with pLE657 and the csDisCoTune
450 expression plasmid (pLE900) or, for comparison, alone or together with the csCyDisCo plasmid
451 (pLE577). Cells were plated on lysogeny broth (LB) agar supplemented with kanamycin (50
452 µg/mL) and – when co-transforming with pLE900/pLE577 – chloramphenicol (30 µg/mL). LB
453 medium containing the same type and concentration of antibiotic as in the transformation plates

454 was inoculated with a single colony and incubated for at 37° C for 18 hours at 190 rpm on an
455 orbital shaker.

456 Initial small-scale expression tests were performed in 50 mL LB medium containing
457 appropriate antibiotics and supplemented with 0.05% glucose. Test cultures were inoculated with
458 2% overnight culture and grown at 37° C, 190 rpm until OD_{600} reached ~0.8. Expression was
459 induced by addition of isopropyl -D-1-thiogalactopyranoside (IPTG) to a final concentration of 1
460 mM and the cultures grown for 18 hours at 25° C with shaking at 190 rpm to allow protein
461 expression. Large-scale expression was performed similarly in 1 L LB medium. Expression
462 performed in 0.5 L M9 minimal medium for stable isotope labelling was performed as described
463 in (62).

464 Induced cultures were harvested by centrifugation at 4,000 x g for 20 minutes at 4° C. The
465 cell pellets were resuspended in 20 mL lysis buffer (50 mM Tris (pH 8.0), 300 mM NaCl, 20 mM
466 imidazole) per liter culture. Cell lysis was performed using a UP200S ultrasonic processor
467 (Hielscher) while keeping the cells on ice. The cells were lysed with 8 x 30-second pulses at 90%
468 power with 30-second rests between each pulse. Cell debris was subsequently pelleted by
469 centrifugation at 30,000 x g for 45 minutes at 4° C. The cleared supernatant was reserved, and cell
470 pellets were resuspended in an equal volume lysis buffer containing 8 M urea for SDS-PAGE
471 analysis.

472

473 **Protein purification**

474 Ub-His_{10}-conkazal-M1 was affinity purified from the clarified lysate on an ÄKTA START system
475 equipped with a 5-mL prepacked HisTrap HP (Cytiva) column equilibrated in lysis buffer. The

476 lysate was applied to the column and washed with 30 column volumes (CVs) of lysis buffer before
477 elution of Ub-His_{10}-conkazal-M1 with a gradient of 0% to 100% elution buffer (50 mM Tris (pH
478 8.0), 300 mM NaCl, 400 mM imidazole) developed over 20 CVs. Pooled fractions were dialyzed
479 twice against 2 L cleavage buffer (50 mM Tris (pH 8.0), 300 mM NaCl).

480 The Ub-His10-conkazal-M1 fusion protein was cleaved using His_6-tagged TEVp,
481 expressed, and purified essentially as described previously (27). A molar ratio Ub-His_{10}-conkazal-
482 M1: His_6-TEVp of 1:2 was used. To avoid reducing the disulfides in conkazal-M1, His_6-TEVp –
483 preactivated with 2 mM dithiothreitol (DTT) for 30 minutes at room temperature – was diluted to
484 approximately 0.002 mM DTT by 3 rounds of dilution/concentration in an Amicon Ultra 15 mL
485 10K Centrifugal Filter (Merck Millipore). TEVp cleavage was performed overnight at room
486 temperature.

487 To remove uncleaved Ub-His_{10}-conkazal-M1, the Ub-His_{10} tag, and His_6-TEVp, the
488 cleavage mixture was applied to a 5 mL prepacked HisTrapHP (Cytiva) affixed an ÄKTA START
489 equilibrated in cleavage buffer. The flow-through and the first wash fraction were collected. The
490 presence of cleaved conkazal-M1 in the flow-through and wash fractions was investigated by
491 analysis on a 15% SDS-PAGE gel, and the protein-containing fractions pooled. Due to minor
492 remaining impurities, the cleaved conkazal-M1 was subjected to size exclusion chromatography
493 on a Superdex 75 Increase 10/300 GL column (Cytiva) equilibrated in 50 mM Tris (pH 7.8), 150
494 mM NaCl. Fractions containing purified conkazal-M1 were pooled and stored at 4 C.

495

496 **SDS-PAGE analysis**

497 Samples from bacterial expression and subsequent purification steps were separated on 15%
498 glycine SDS-PAGE gels and where indicated samples were treated with 40 mM DTT. Protein
499 bands were visualized with Coomassie brilliant blue stain and images were captured with a BioRad
500 Chemidoc Imaging System.

501

502 **Protein concentration determination**

503 Concentration was determined by measuring absorbance at 280 nm and using the theoretical
504 extinction coefficient provided by the Expasy ProtParam tool available through the Expasy
505 bioinformatics resource web portal (63).

506

507 **Mass spectrometry**

508 Mass spectrometry was performed on a Waters Xevo G3 QTof system equipped with a ACQUITY
509 UPLC C4 reversed-phase liquid chromatography column (300 Å, 1.7 µm, 2.1 mm × 50 mm) using
510 a sample of purified conkazal-M1 at 100 µM adjusted to pH 2 with trifluoroacetic acid as input.

511

512 **NMR spectroscopy**

513 A sample of 500 µM ^{13}C,^{15}N-labelled conkazal-M1 was prepared in 10 mM 2-(N-
514 morpholino)ethanesulfonic acid (MES), 200 mM NaCl, 5% D_2O, pH 6.0. For backbone chemical
515 shift assignments ^{15}N-HSQC, HNCA, HN(CO)CA, CBCA(CO)NH, CBCANH, CBCA(CO)NH,
516 HNCO, and HN(CA)CO spectra were recorded on a Bruker Avance III HD spectrometer operating
517 at 750 MHz with a triple resonance cryoprobe. Side chain assignments were achieved from

518 HCCONH and HCCH-TOCSY spectra. Distance restraints were obtained from ^{15}N-NOESY-
519 HSQC and ^{13}C-NOESY-HSQC experiments recorded using a mixing time of 120 ms. The triple
520 resonance spectra were recorded with non-uniform sampling at 25% and reconstructed with
521 qMDD (64). All spectra were processed with nmrPipe (65) and analyzed in CCPNMR analysis
522 (66).

523

524 **NMR structure calculations**

525 Automated NOE assignment was performed using Cyana (67). The NOE list was manually refined.
526 XPLOR-NIH (68) was used for the final structural refinement including a torsion angle database
527 potential (69) and an implicit solvent mode (70). For the final ensemble representing the solution
528 structure of conkazal-M1, the 20 structures with the lowest energy were chosen from 100
529 calculated structures. The Ramachandran-plot statistics was calculated using PROCHECK (71)
530 and the coordinate precisions were calculated by PyMOL (DeLano Scientific). Structure
531 visualization was also performed with PyMOL. The structure has been deposited in the Protein
532 Data Bank (PDB, wwpdb.org) with PDB code 9SLR and chemical shifts and peaks lists from the
533 NOESY spectra have been deposited in the Biological Magnetic Resonance Bank
534 (http://www.bmrb.wisc.edu/) with ID 53313.

535

536 **Structural homology search**

537 Structures similar to conkazal-M1 were identified using FoldSeek (38), and structural
538 superpositions were executed manually in PyMOL (DeLano Scientific) utilizing the CEalign
539 command.

540

541 **Protease assays**

542 Endoproteinase GluC, bovine pancreas α-chymotrypsin and subtilisin A and substrates for
543 chymotrypsin (N-Succinyl-Gly-Gly-Phe-*p*-nitroanilide) and subtilisin A (N-succinyl-Ala-Ala-
544 Pro-Phe-p-nitroanilide) were all purchased from Merck. Substrate for GluC (Z-Phe-Leu-Glu-pNA)
545 was purchased from Bachem.

546 Inhibition assays with chymotrypsin and GluC were performed in assay buffer (50 mM
547 Tris, 2 mM $CaCl_2$, pH 7.8). Subtilisin A inhibition experiments were performed in assay buffer,
548 pH 8 supplemented with 5 mM $CaCl_2$. The proteases (3.6 nM SubA, 100 nM GluC and 400 nM
549 chymotrypsin) were incubated with varying concentrations of conkazal-M1 for 10 min at 29° C.
550 Proteolytic reactions were initiated by addition of the substrates and absorbance at 405 nm was
551 recorded every 30 seconds for 30 minutes. The 96-well plate was shaken orbitally for 10 seconds
552 immediately prior to measurement. Initial reaction rates were estimated with the web-based tool,
553 IceKat (72). The absorbance-based rates for Subtilisin A were converted to reaction rates in
554 μM/min/mg using the extinction coefficient for AAPF-p-nitroaniline (ε = 8800 $M^{-1}cm^{-1}$) and an
555 estimated path length of 0.2 cm for 100 μL reaction volumes in 96-well plates. The absorbance-
556 based rates (Abs/min) for Chymotrypsin and GluC were plotted without conversion. The data were
557 fit to four models of enzyme inhibition: Michaelis-Menten (no inhibitor), competitive,
558 uncompetitive, and noncompetitive inhibition.

559 $$v = \frac{V_{max}[S]}{K_m + [S]}$$

560

561 $$v = \frac{V_{max}[S]}{K_m\left(1 + \frac{[I]}{K_i}\right) + [S]}$$

562

563 $$v = \frac{V_{max}[S]}{K_m + [S]\left(1 + \frac{[I]}{K_i}\right)}$$

564

565 $$v = \frac{V_{max}[S]}{K_m\left(1 + \frac{[I]}{K_i}\right) + [S]\left(1 + \frac{[I]}{K_i}\right)}$$

566 Model comparison using Akaike Information Criterion identified the best fit to the data.

567

568 **Cell culture and constellation pharmacology**

569 Experiments were performed as previously described (73). Briefly, all lumbar DRG neurons from
570 C57BL/6 mice 30-40 days old of both genders were dissociated by trypsin digestion and
571 mechanical trituration and plated on polylysine-coated plates. The cells were incubated overnight
572 in a 5% CO_2 incubator at 37 °C in neuronal culture medium [minimal essential media
573 supplemented with 10% (vol/vol) fetal bovine serum, penicillin (100 U/mL), streptomycin (100
574 μg/mL), 1× Glutamax, 10 mM HEPES, and 0.4% (wt/vol) glucose, adjusted to pH 7.4]. One hour
575 before calcium imaging experiments, cells were loaded with 2.5 μM fura-2-acetoxymethyl ester
576 (Fura-2-AM, Sigma-Aldrich) and incubated at 37 °C. Cells loaded with Fura-2-AM dye were
577 excited intermittently with 340- and 380-nm light; fluorescence emission was monitored at 510
578 nm. An image was captured at each excitation wavelength, and the ratio of fluorescence intensities
579 (340/380 nm) was acquired every 2 s to monitor the relative changes in intracellular calcium
580 concentration in each cell as a function of time. Typically, 500-1,000 neurons were imaged for

each experiment. The calcium transients were elicited by 15 s applications of depolarizing stimulus (30 mM KCl), followed by four washes with extracellular solution [145 mM NaCl, 5 mM KCl, 2 mM CaCl2, 1 mM MgCl2, 1 mM sodium citrate, 10 mM HEPES, and 10 mM glucose, adjusted to pH 7.4]. For activity screening, conkazal-M1 was applied in extracellular solution at 1 and 10 μM between KCl pulses and activity was assessed by significant change in the KCl response after incubation with conkazal-M1. Four different pharmacological agents were used for cell classification: conotoxin *κ*M-RIIIJ (1 μM), AITC (100 μM), menthol (400 μM) and capsaicin (300 nM). At the end of the experiment, cells were incubated with 0.7 mL of Hoechst stain (1000 μg/mL) and Alexa-Fluor 647 Isolectin B4 (2.5 μg/mL) for 5 min. The data was acquired using NIS-Elements and further processed with CellProfiler v.3.1.9 (Jones et al, 2008). Custom-built scripts in Python v.3.7.2 and R were used for further data analysis and visualization. The experimental methods were approved by the Institutional Animal Care and Use Committee (IACUC) of the University of Utah (Protocol number: 17–05017).

Statistical analysis of calcium imaging data

Each experiment had a series of KCl-induced responses. Their magnitude was measured as the maximum area under the curve for any 15 s interval. The deviation of this response after incubation with conkazal-M1 was used to estimate the effect of conkazal-M1. We estimated this effect on each neuron using the linear model (lm function in R) with the following:

$$lm(Kauc15 \sim linear + toxin)$$

601 where Kauc15 was the maximum area under the curve for any 15 s interval for each KCl pulse,
602 linear was the sequential trend coded as sequential integers, toxin was an indicator variable for the
603 pulse immediately following incubation with conkazal-M1.

604 To control for false positives and set appropriate thresholds for significance we used Monte
605 Carlo simulations. Each simulation generated random numbers drawn from a normal distribution
606 with mean and standard deviations of the actual data. The T-stats were estimated for all cells and
607 recorded. This is considered a null distribution. The T-stat estimates from 100 Monte Carlo
608 simulations were used to establish the thresholds for single test cell significance as well as per-
609 experiment significance.

610 We used the relative residual of the affected KCl pulse in the model as a ratio of the
611 observed to the model predicted magnitude of effect. This was calculated as:

612 $$(KPp-KPobs)/KPp,$$

613 where KPp was the predicted peak height of the pulse immediately after conkazal-M1 application
614 and KPobs was the observed KCl pulse.

615

616 **Fish behavioral assays**

617 All animal assays were conducted in accordance with the *Guide for the Care and Use of*
618 *Laboratory Animals* and were approved by the Institutional Animal Care and Use Committee
619 (IACUC) of the University of Utah (protocol #001629). Goldfish (*Carassius auratus*) of both
620 sexes (weight= 0.85 ± 0.11 g, mean ± SD) were obtained from a local supplier and used for the
621 experiments. An in-house prepared E3 Buffer (4.96 mM NaCl, 0.18 mM KCl, 0.33 mM
622 $CaCl_2 \cdot 2H_2O$, 0.4 mM $MgCl_2 \cdot 6H_2O$; pH 7.2) served as the vehicle. Groups of three fish each

623 received injections of either 13.5 μL E3 buffer alone (control) or E3 buffer containing the desired
624 concentration of conkazal-M1.

625 All injections were performed with a 1 mL insulin syringe (31 G). For intraperitoneal
626 injections, fish were positioned ventral side up on a chilled sponge embedded with E3 Buffer under
627 a dissecting microscope, and the solution was delivered into the abdominal cavity. For
628 intramuscular injections, fish were positioned laterally on a hard surface moistened with E3 buffer
629 under a dissecting microscope, and the solution was administered into the lateral muscle below the
630 posterior end of the dorsal fin. Following injection, all fish were maintained under observation for
631 three hours.

632

633 **TABLES**

634 **Table 1.** Statistics for the ensemble of 20 conformers representing conkazal-M1

Experimental restraints	
NOE based restraints	
Intraresidual (|i - j| = 0)	130
Sequential (|i - j| = 1)	183
Medium range (2 ≤ |i - j| ≤ 4)	92
Long range (|i - j| ≥ 5)	197
Total	602
Dihedral angles	100
Disulfide bonds	4
Restraint violation statistics[a]	
NOE-based distances (Å)	0.046 ± 0.004
Dihedral angles (°)	1.04 ± 0.09
Ramachandran plot statistics[b]	
Most favored regions	83.0%
Allowed regions	15.3%
Generously allowed regions	1.5%
Disallowed regions	0.3%
Coordinate precision (RMSD to the average)[c]	
Residues 10 - 73	
Backbone N, C^{α} and C'	0.31 ± 0.07 Å
All heavy atoms	2.4 ± 0.6 Å

635 [a]XPLOR-NIH, [b]PROCHECK, [c]PyMOL

636 **ABBREVIATIONS**

637 CLANS, CLuster ANalysis of Sequences); csPDI, conotoxin-specific PDI; DRG: dorsal root
638 ganglion, DTT, dithiothreitol; ER, endoplasmic reticulum; HEPES, 4-(2-hydroxyethyl)-1-
639 piperazineethanesulfonic acid; hPDI, human PDI; IMAC, immobilized metal affinity
640 chromatography; IPTG, isopropyl ß-D-1-thiogalactopyranoside; KPI: Kazal-type protease
641 inhibitor, LB, lysogeny broth: MSA, multiple sequence alignment; NCBI, National Center for
642 Biotechnology Information; PDB, Protein Data Bank; PDI, protein disulfide isomerase; RMSD,
643 root mean square deviation; SDS-PAGE, sodium dodecyl sulfate-polyacrylamide gel
644 electrophoresis; SRA, sequence read archive; TEVp, tobacco etch virus protease; TPM, transcripts
645 per million; Ub, ubiquitin; Ub-His_{10}, Ub containing 10 consecutive histidines.

646

647 **ACKNOWLEDGMENTS**

648 We thank Signe A. Sjørup, SBiNLab, University of Copenhagen, for recording mass spectra.

649

650 **FUNDING**

651 This work was supported by the Independent Research Fund Denmark (10.46540/3103-00126B to
652 L.E.), the Lundbeck Foundation (R500-2024-1904 to L.E.), the Novo Nordisk Foundation
653 (NNF18OC0032996 to K.T.), and a National Institutes of Health grant (GM144719 to B.M.O. and
654 H.S.-H). T.L.K. was supported by an international postdoc fellowship from the Independent
655 Research Fund Denmark (3102-00006B).

656

657 **AUTHOR CONTRIBUTIONS**

Conceptualization, C.M.H., T.L.K., H.S.-H., and L.E.; Data curation, C.M.H., T.L.K., A.R., K.C., M.W., K.T., and L.E.; Formal Analysis, C.M.H., T.L.K., A.R., K.C., and K.T.; Funding acquisition, T.L.K., B.O., H.S.-H., K.T., and L.E.; Investigation, C.M.H., T.L.K., N.L.R., M.W., A.R., K.T. and L.E.; Project administration, L.E.; Supervision, B.O., H.S.-H., and L.E.; Visualization, C.M.H., T.L.K., A.R., K.T., and L.E.; writing—original draft preparation, C.M.H., T.L.K., A.R., H.S.-H., K.T. and L.E.; writing—review and editing, all authors. All authors have read and agreed to the published version of the manuscript.

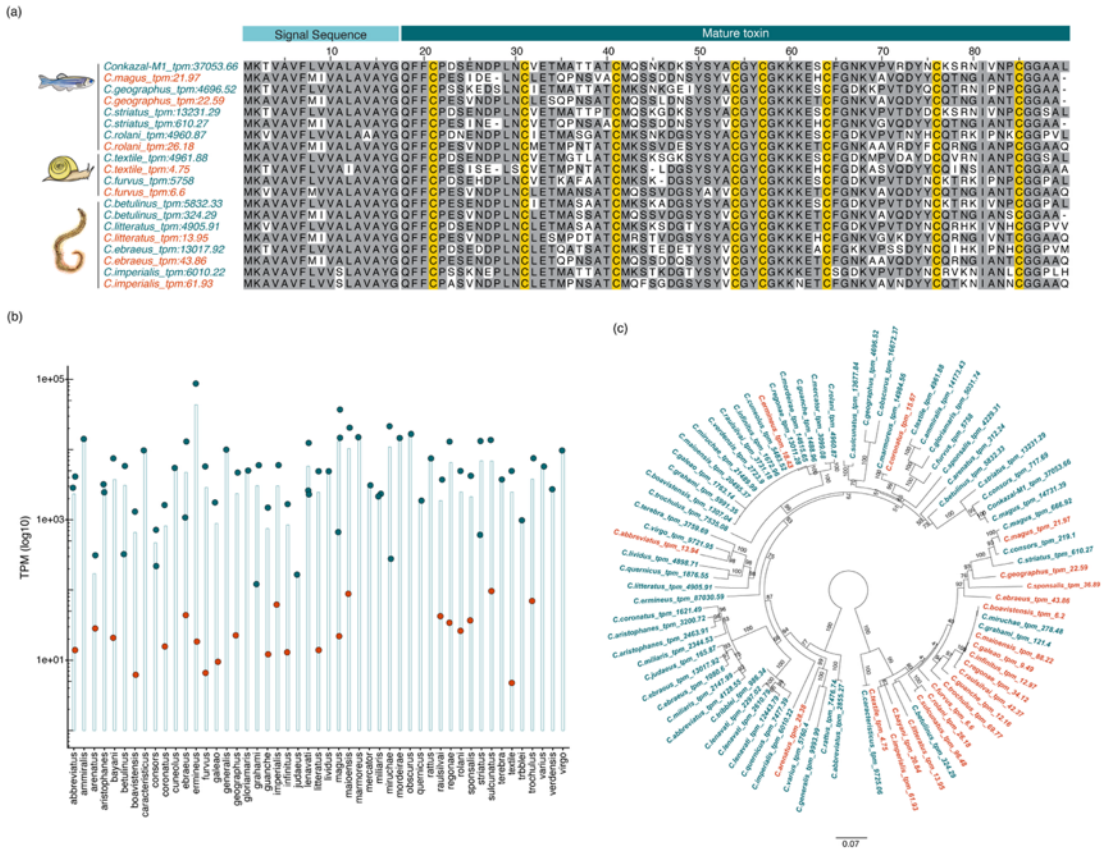
COMPETING INTERESTS

The authors declare that no competing interests exist.

FOOTNOTES

The atomic coordinates for conkazal-M1 are available with the Protein Data Bank under accession number 9SLR and in the Biological Magnetic Resonance Data Bank under ID number 53313.

675 **FIGURES**



676

677 **Figure 1. The MKAVAs are encoded by either low or high expressing transcripts. (a)** Multiple
678 sequence alignment of select *Conus* precursor sequences from the MKAVA superfamily.
679 Sequences encoded by high or low expressing transcripts are indicated in dark teal and orange,
680 respectively. The signal sequence and mature toxin regions are identified by light and dark teal
681 bars above the alignment. The prey preference of each species (i.e. piscivorous, molluscivorous or
682 vermivorous) is indicated by the icons on the left. **(b)** Gene expression levels of 85 MKAVA
683 transcripts from 49 cone snail species. Dark teal and orange circles indicate high (>100 tpm) and
684 low (<100 tpm) expressors, respectively, with light teal bars showing the average tpm-value of the
685 transcripts for each species. **(c)** Phylogenetic reconstruction of the MKAVA transcripts from (b).
686 Maximum-likelihood was performed using IQ-Tree's UF bootstrap with 1000 replicates with best

687 fit model TMP3 + G4 selected by the Bayesian Information Criterion. High-expressing transcripts
688 are shown in dark teal and low-expressing in orange. Some bootstrap values were removed for
689 visualization purposes but can be seen in Figure S10. The tree was rooted to separate the two
690 paralogous genes.

691

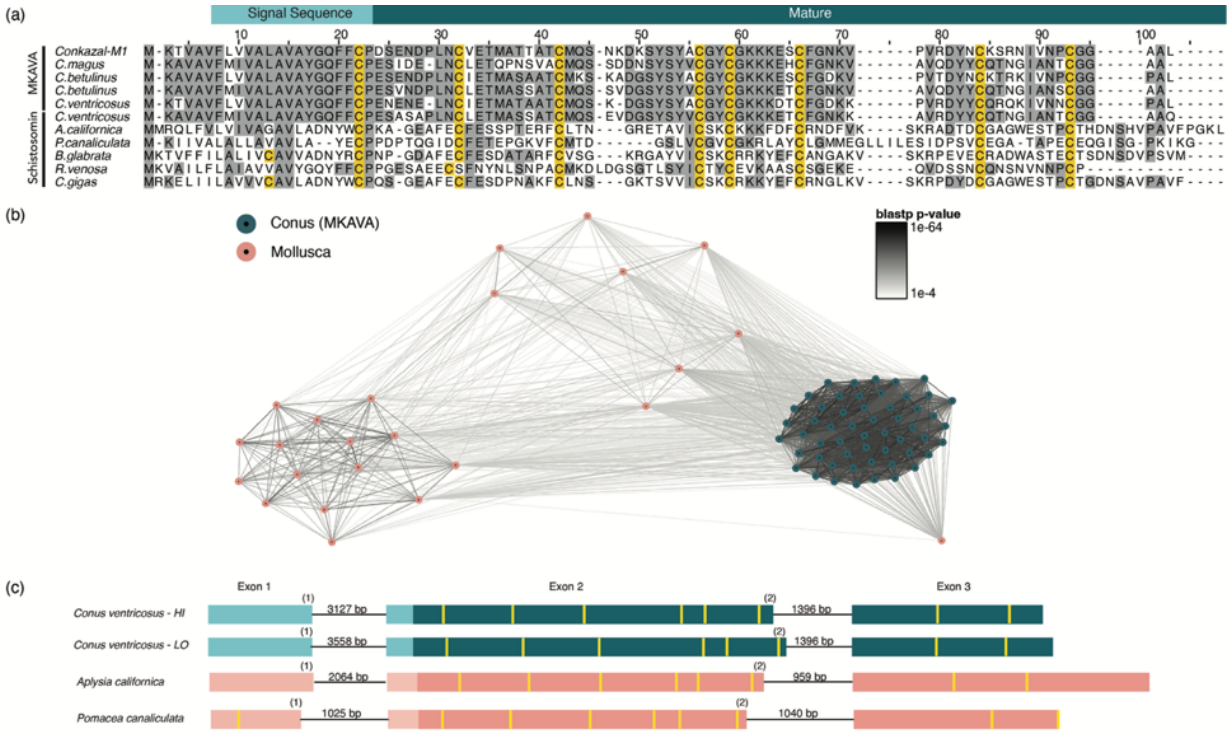


Figure 2. The MKAVAs are evolutionarily related to schistosomin. (a) Multiple sequence alignment of MKAVAs and schistosomin-like sequences from *Aplysia californica* (XP_005098364.1), *Pomacea cataticulata* (PVD28780.1), *Biomphilaria glabrata* (NP_001298209.1), *Rapana venosa* (GDIA01100366.1) and *Crassostrea gigas* (GHAT01224603.1). **(b)** CLANS clustering analysis of conotoxins and schistosomins. Deep teal: conotoxins from MKAVA superfamily; pink: molluscan schistosomin. Nodes depict individual precursor sequences, and the edges correspond to BLOSUM62 blastp p-values. **(c)** Gene structure analysis of toxins with high or low expression from *Conus ventricosus*, and of schistosomins from *A. californica* and *P. cataticulata*. The exons are depicted as boxes proportional to the length of the sequences, while introns are represented as thin lines (with sequence length labelled). The intron phases (0, 1, or 2) are indicated prior to the introns. The yellow lines indicate cysteine positions within the exons.

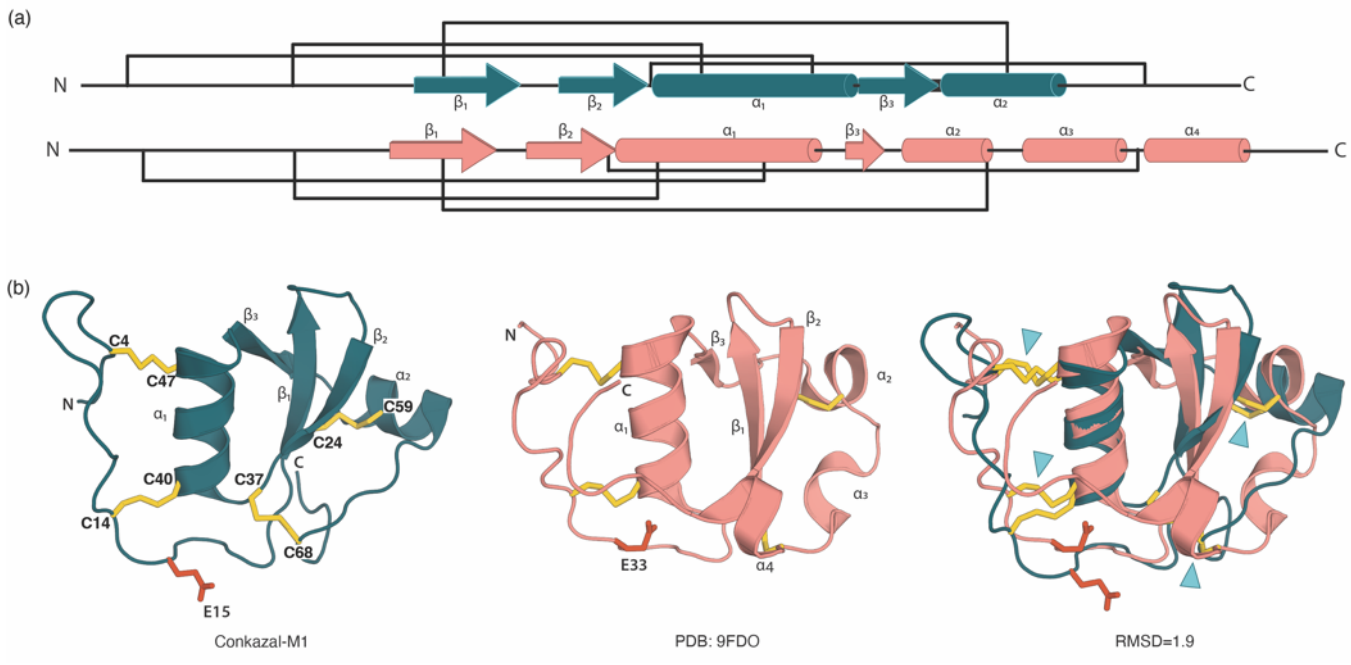


Figure 3. Conkazal-M1 harbors a similar fold to schistosomin. (a). Graphic representations of Conkazal-M1 (teal) and schistosomin from *Biomphalaria glabrata* (pink). α-helices and β-strands are represented by cylinders and block arrows, respectively, and disulfide bridges are represented as black brackets. **(b).** Cartoon representation of the conkazal-M1 NMR structure (PDB: 9SLR) (left), cartoon representation of the *B. glabrata* schistosomin structure solved by X-ray crystallography (PDB: 9FDO) (middle), and superposition of conkazal-M1 and schistosomin (right). Coloration as described in Panel (a). Disulfide bridges are colored yellow and the amino acid residue at the P1 position is shown in red. The RMSD value is derived from 38 α-carbons and determined using the PyMOL super command.

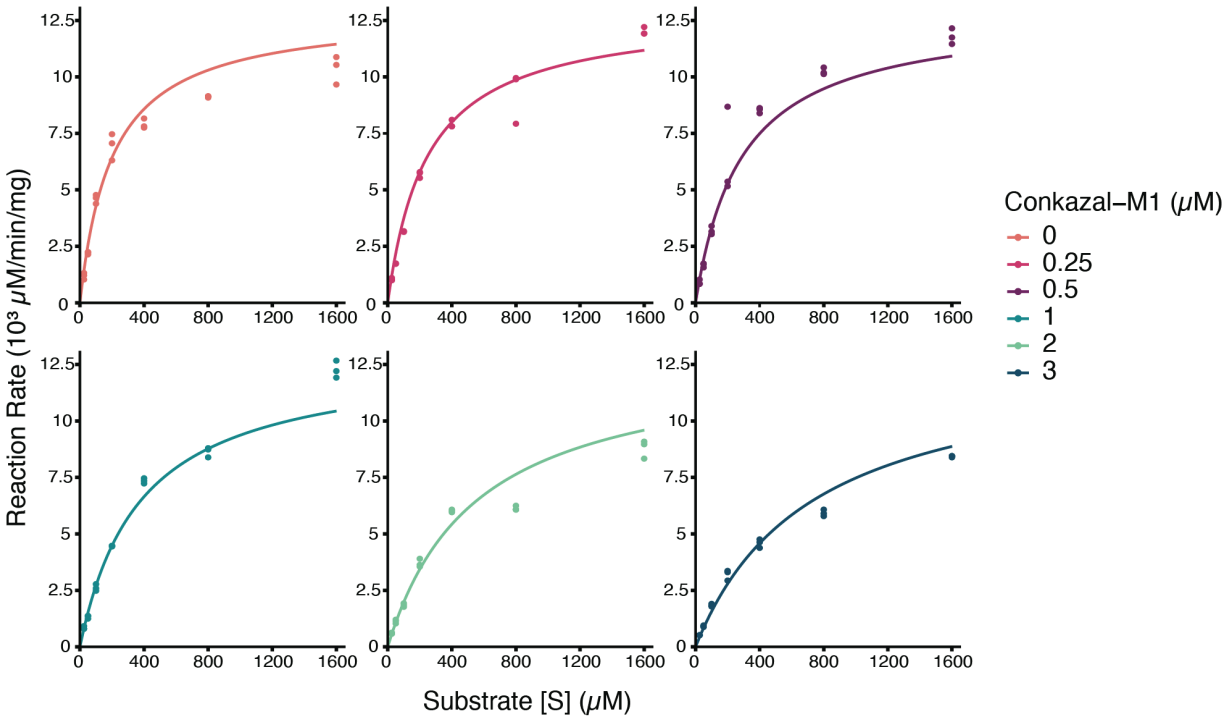


Figure 4. Conkazal-M1 inhibits proteolytic activity of Subtilisin A. Kinetic analysis of Subtilisin A in the presence of increasing concentrations Conkazal-M1. Reaction rates were measured across a range of substrate concentrations and were normalized to enzyme mass. The plots display experimental data points (colored according to inhibitor concentration: 0 μM (orange), 0.25 μM (pink), 0.5 μM (purple), 1 μM (light teal), 2 μM (cyan) and 3 μM (dark teal)) overlaid with lines showing showing the best-fit model (competitive inhibition).

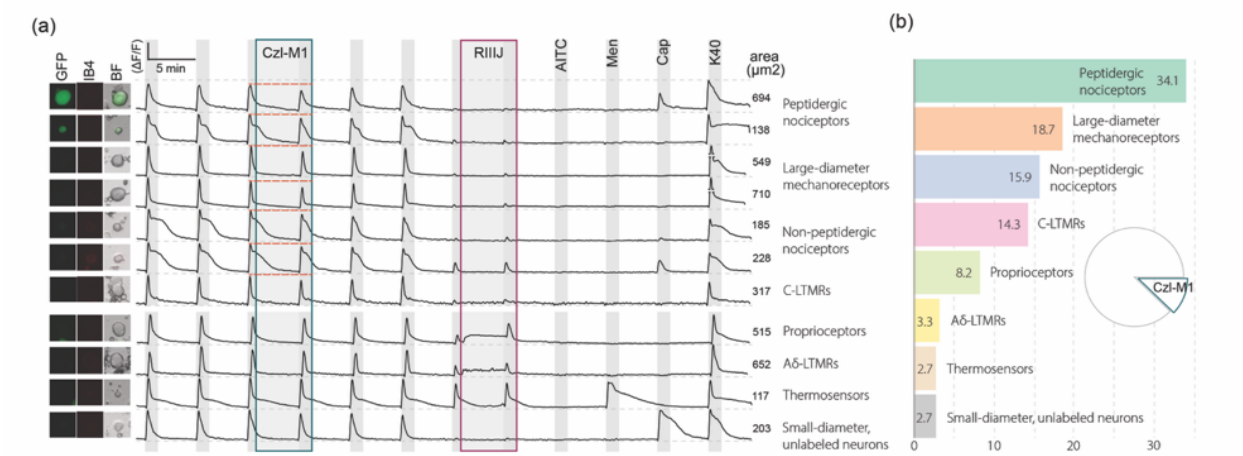


Figure 5. Effects of conkazal-M1 on mouse sensory neurons by constellation pharmacology. (**A**) Representative calcium-imaging traces of individual neurons responding with a partial block to applications of 10 μM conkazal-M1 (top 7 traces, conkazal-M1 application highlighted by blue box), and neurons that did not respond (bottom 4 traces). Light grey shading indicates KCl (30 mM) depolarization pulses. Each trace represents the calcium signal (ΔF/F) of the neuron pictured on the left (GFP: CGRP expression; IB4: isolectin B4 staining; BF: brightfield image; 10x magnification). Pharmacological agents used for cell classification: RIIIJ (1 μM, burgundy box); AITC (100 μM), menthol (400 μM, Men), and capsaicin (300 nM, Cap). A pulse of 40 mM KCl (K40) was used to assess cell viability at the end of the experiment. Cell size is indicated on the left side of each trace. (**B**) Bar graph showing the percentage of each class of conkazal-M1-sensitive DRG neurons. The pie chart represents the population of neurons analyzed (n = 1,494), highlighting the 12.2% of conkazal-M1 (10 μM)-sensitive sensory neurons. AITC: allyl isothiocyanate; Aδ-LTMRs: Aδ–low-threshold mechanoreceptors; CGRP: calcitonin-gene related peptide; C-LTMR: C-low threshold mechanoreceptors; GFP: green fluorescent protein; N16: capsaicin-positive small-diameter neurons; RIIIJ: conotoxin *κ*M-RIIIJ.

745 **REFERENCES**

746 1. Casewell, N. R., Wuster, W., Vonk, F. J., Harrison, R. A., and Fry, B. G. (2013) Complex
747 cocktails: the evolutionary novelty of venoms *Trends Ecol Evol* **28**, 219-229
748 2. Lewis, R. J., and Garcia, M. L. (2003) Therapeutic potential of venom peptides *Nat Rev*
749 *Drug Discov* **2**, 790-802
750 3. Sachetto, A. T. A., and Mackman, N. (2019) Modulation of the mammalian coagulation
751 system by venoms and other proteins from snakes, arthropods, nematodes and insects
752 *Thromb Res* **178**, 145-154
753 4. Woodward, S. R., Cruz, L. J., Olivera, B. M., and Hillyard, D. R. (1990) Constant and
754 hypervariable regions in conotoxin propeptides *The EMBO journal* **9**, 1015-1020
755 5. Sollod, B. L., Wilson, D., Zhaxybayeva, O., Gogarten, J. P., Drinkwater, R., and King, G.
756 F. (2005) Were arachnids the first to use combinatorial peptide libraries? *Peptides* **26**, 131-
757 139
758 6. Jackson, T. N. W., and Koludarov, I. (2020) How the Toxin got its Toxicity *Front*
759 *Pharmacol* **11**, 574925
760 7. Bayrhuber, M., Vijayan, V., Ferber, M., Graf, R., Korukottu, J., Imperial, J. *et al.* (2005)
761 Conkunitzin-S1 is the first member of a new Kunitz-type neurotoxin family: structural and
762 functional characterization *J Biol Chem* **280**, 23766-23770
763 8. Finol-Urdaneta, R. K., Remedi, M. S., Raasch, W., Becker, S., Clark, R. B., Struver, N. *et*
764 *al.* (2012) Block of Kv1.7 potassium currents increases glucose-stimulated insulin
765 secretion *EMBO molecular medicine* **4**, 424-434

766 9. Droctove, L., Ciolek, J., Mendre, C., Chorfa, A., Huerta, P., Carvalho, C. *et al.* (2022) A
767 new Kunitz-type snake toxin family associated with an original mode of interaction with
768 the vasopressin 2 receptor *Br J Pharmacol* **179**, 3470-3481
769 10. Schweitz, H., Heurteaux, C., Bois, P., Moinier, D., Romey, G., and Lazdunski, M. (1994)
770 Calcicludine, a venom peptide of the Kunitz-type protease inhibitor family, is a potent
771 blocker of high-threshold Ca2+ channels with a high affinity for L-type channels in
772 cerebellar granule neurons *Proc Natl Acad Sci USA* **91**, 878-882
773 11. Mishra, M. (2020) Evolutionary Aspects of the Structural Convergence and Functional
774 Diversification of Kunitz-Domain Inhibitors *J Mol Evol* **88**, 537-548
775 12. Yuan, C. H., He, Q. Y., Peng, K., Diao, J. B., Jiang, L. P., Tang, X. *et al.* (2008) Discovery
776 of a distinct superfamily of Kunitz-type toxin (KTT) from tarantulas *PloS one* **3**, e3414
777 13. Gladkikh, I., Peigneur, S., Sintsova, O., Lopes Pinheiro-Junior, E., Klimovich, A.,
778 Menshov, A. *et al.* (2020) Kunitz-Type Peptides from the Sea Anemone Heteractis crispa
779 Demonstrate Potassium Channel Blocking and Anti-Inflammatory Activities *Biomedicines*
780 **8**,
781 14. Peigneur, S., Billen, B., Derua, R., Waelkens, E., Debaveye, S., Beress, L. *et al.* (2011) A
782 bifunctional sea anemone peptide with Kunitz type protease and potassium channel
783 inhibiting properties *Biochem Pharmacol* **82**, 81-90
784 15. Zupunski, V., and Kordis, D. (2016) Strong and widespread action of site-specific positive
785 selection in the snake venom Kunitz/BPTI protein family *Sci Rep* **6**, 37054
786 16. Zdenek, C. N., Cardoso, F. C., Robinson, S. D., Mercedes, R. S., Raidjoe, E. R., Hernandez-
787 Vargas, M. J. *et al.* (2024) Venom exaptation and adaptation during the trophic switch to
788 blood-feeding by kissing bugs *iScience* **27**, 110723

789 17. Friedrich, T., Kroger, B., Bialojan, S., Lemaire, H. G., Hoffken, H. W., Reuschenbach, P.
790 *et al.* (1993) A Kazal-type inhibitor with thrombin specificity from Rhodnius prolixus *J*
791 *Biol Chem* **268**, 16216-16222
792 18. Kim, B. Y., Lee, K. S., Zou, F. M., Wan, H., Choi, Y. S., Yoon, H. J. *et al.* (2013)
793 Antimicrobial activity of a honeybee (Apis cerana) venom Kazal-type serine protease
794 inhibitor *Toxicon* **76**, 110-117
795 19. Liao, Z., Tang, X., Chen, W., Jiang, X., Chen, Z., He, K. *et al.* (2022) Shrew's venom
796 quickly causes circulation disorder, analgesia and hypokinesia *Cell Mol Life Sci* **79**, 35
797 20. Qian, C., Fang, Q., Wang, L., and Ye, G. Y. (2015) Molecular Cloning and Functional
798 Studies of Two Kazal-Type Serine Protease Inhibitors Specifically Expressed by Nasonia
799 vitripennis Venom Apparatus *Toxins* **7**, 2888-2905
800 21. Wang, Q. W., Zou, W. B., Masson, E., Ferec, C., Liao, Z., and Chen, J. M. (2025) Genetics
801 and clinical implications of SPINK1 in the pancreatitis continuum and pancreatic cancer
802 *Hum Genomics* **19**, 32
803 22. Brillard-Bourdet, M., Hamdaoui, A., Hajjar, E., Boudier, C., Reuter, N., Ehret-Sabatier, L.
804 *et al.* (2006) A novel locust (Schistocerca gregaria) serine protease inhibitor with a high
805 affinity for neutrophil elastase *Biochem J* **400**, 467-476
806 23. Rimphanitchayakit, V., and Tassanakajon, A. (2010) Structure and function of invertebrate
807 Kazal-type serine proteinase inhibitors *Dev Comp Immunol* **34**, 377-386
808 24. Safavi-Hemami, H., Foged, M. M., and Ellgaard, L. (2018) Evolutionary Adaptations to
809 Cysteine-rich Peptide Folding In *Oxidative Folding of Proteins: Basic Principles, Cellular*
810 *Regulation and Engineering*, Feige M, ed. Royal Society of Chemistry, 99-128

811 25. Koch, T. L., Robinson, S. D., Salcedo, P. F., Chase, K., Biggs, J., Fedosov, A. E. *et al.*
812 (2024) Prey Shifts Drive Venom Evolution in Cone Snails *Mol Biol Evol* **41**,
813 26. Robinson, S. D., Safavi-Hemami, H., McIntosh, L. D., Purcell, A. W., Norton, R. S., and
814 Papenfuss, A. T. (2014) Diversity of conotoxin gene superfamilies in the venomous snail,
815 Conus victoriae *PloS one* **9**, e87648
816 27. Hackney, C. M., Florez Salcedo, P., Müller, E., Koch, T. L., Kjelgaard, L. D., Watkins, M.
817 *et al.* (2023) A previously unrecognized superfamily of macro-conotoxins includes an
818 inhibitor of the sensory neuron calcium channel Cav2.3 *PLoS Biol* **21**, e3002217
819 28. Yang, Z. (1997) PAML: a program package for phylogenetic analysis by maximum
820 likelihood *Comput Appl Biosci* **13**, 555-556
821 29. Hordijk, P. L., de Jong-Brink, M., ter Maat, A., Pieneman, A. W., Lodder, J. C., and Kits,
822 K. S. (1992) The neuropeptide schistosomin and haemolymph from parasitized snails
823 induce similar changes in excitability in neuroendocrine cells controlling reproduction and
824 growth in a freshwater snail *Neurosci Lett* **136**, 193-197
825 30. Frickey, T., and Lupas, A. (2004) CLANS: a Java application for visualizing protein
826 families based on pairwise similarity *Bioinformatics* **20**, 3702-3704
827 31. Bertelsen, A. B., Hackney, C. M., Bayer, C. N., Kjelgaard, L. D., Rennig, M., Christensen,
828 B. *et al.* (2021) DisCoTune: versatile auxiliary plasmids for the production of disulphide-
829 containing proteins and peptides in the E. coli T7 system *Microb Biotechnol* 10.1111/1751-
830 7915.13895
831 32. Gaciarz, A., Khatri, N. K., Velez-Suberbie, M. L., Saaranen, M. J., Uchida, Y., Keshavarz-
832 Moore, E. *et al.* (2017) Efficient soluble expression of disulfide bonded proteins in the

833 cytoplasm of Escherichia coli in fed-batch fermentations on chemically defined minimal
834 media *Microb Cell Fact* **16**, 108
835 33. Nielsen, L. D., Foged, M. M., Albert, A., Bertelsen, A. B., Soltoft, C. L., Robinson, S. D.
836 *et al.* (2019) The three-dimensional structure of an H-superfamily conotoxin reveals a
837 granulin fold arising from a common ICK cysteine framework *J Biol Chem* **294**, 8745-
838 8759
839 34. Safavi-Hemami, H., Li, Q., Jackson, R. L., Song, A. S., Boomsma, W., Bandyopadhyay, P.
840 K. *et al.* (2016) Rapid expansion of the protein disulfide isomerase gene family facilitates
841 the folding of venom peptides *Proc Natl Acad Sci USA* **113**, 3227-3232
842 35. Rogov, V. V., Rozenknop, A., Rogova, N. Y., Lohr, F., Tikole, S., Jaravine, V. *et al.* (2012)
843 A universal expression tag for structural and functional studies of proteins *Chembiochem*
844 **13**, 959-963
845 36. Buczek, O., Bulaj, G., and Olivera, B. M. (2005) Conotoxins and the posttranslational
846 modification of secreted gene products *Cell Mol Life Sci* **62**, 3067-3079
847 37. Fischer, W. H., and Spiess, J. (1987) Identification of a mammalian glutaminyl cyclase
848 converting glutaminyl into pyroglutamyl peptides *Proc Natl Acad Sci USA* **84**, 3628-3632
849 38. van Kempen, M., Kim, S. S., Tumescheit, C., Mirdita, M., Lee, J., Gilchrist, C. L. M. *et al.*
850 (2024) Fast and accurate protein structure search with Foldseek *Nat Biotechnol* **42**, 243-
851 246
852 39. Krissinel, E., and Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for
853 fast protein structure alignment in three dimensions *Acta Crystallogr D Biol Crystallogr*
854 **60**, 2256-2268

855 40. Derache, C., Epinette, C., Roussel, A., Gabant, G., Cadene, M., Korkmaz, B. *et al.* (2012)
856 Crystal structure of greglin, a novel non-classical Kazal inhibitor, in complex with
857 subtilisin *FEBS J* **279**, 4466-4478
858 41. Shenoy, R. T., Thangamani, S., Velazquez-Campoy, A., Ho, B., Ding, J. L., and Sivaraman,
859 J. (2011) Structural basis for dual-inhibition mechanism of a non-classical Kazal-type
860 serine protease inhibitor from horseshoe crab in complex with subtilisin *PloS one* **6**, e18838
861 42. Rivier, J., Galyean, R., Gray, W. R., Azimi-Zonooz, A., McIntosh, J. M., Cruz, L. J. *et al.*
862 (1987) Neuronal calcium channel inhibitors. Synthesis of omega-conotoxin GVIA and
863 effects on 45Ca uptake by synaptosomes *J Biol Chem* **262**, 1194-1198
864 43. Qi, L., Iskols, M., Shi, D., Reddy, P., Walker, C., Lezgiyeva, K. *et al.* (2024) A mouse DRG
865 genetic toolkit reveals morphological and physiological diversity of somatosensory neuron
866 subtypes *Cell* **187**, 1508-1526 e1516
867 44. Bhuiyan, S. A., Xu, M., Yang, L., Semizoglou, E., Bhatia, P., Pantaleo, K. I. *et al.* (2024)
868 Harmonized cross-species cell atlases of trigeminal and dorsal root ganglia *Sci Adv* **10**,
869 eadj9173
870 45. Tian, M., Benedetti, B., and Kamoun, S. (2005) A Second Kazal-like protease inhibitor
871 from Phytophthora infestans inhibits and interacts with the apoplastic pathogenesis-related
872 protease P69B of tomato *Plant Physiol* **138**, 1785-1793
873 46. Tian, M., and Kamoun, S. (2005) A two disulfide bridge Kazal domain from Phytophthora
874 exhibits stable inhibitory activity against serine proteases of the subtilisin family *BMC*
875 *Biochem* **6**, 15

876 47. Li, X. C., Wang, X. W., Wang, Z. H., Zhao, X. F., and Wang, J. X. (2009) A three-domain
877 Kazal-type serine proteinase inhibitor exhibiting domain inhibitory and bacteriostatic
878 activities from freshwater crayfish Procambarus clarkii *Dev Comp Immunol* **33**, 1229-1238
879 48. Esmaeilishirazifard, E., Usher, L., Trim, C., Denise, H., Sangal, V., Tyson, G. H. *et al.*
880 (2022) Bacterial Adaptation to Venom in Snakes and Arachnida *Microbiol Spectr* **10**,
881 e0240821
882 49. De Jong-Brink, M. (1995) How schistosomes profit from the stress responses they elicit in
883 their hosts *Adv Parasitol* **35**, 177-256
884 50. Zhang, S. M., Nian, H., Wang, B., Loker, E. S., and Adema, C. M. (2009) Schistosomin
885 from the snail Biomphalaria glabrata: expression studies suggest no involvement in
886 trematode-mediated castration *Mol Biochem Parasitol* **165**, 79-86
887 51. Force, A., Lynch, M., Pickett, F. B., Amores, A., Yan, Y. L., and Postlethwait, J. (1999)
888 Preservation of duplicate genes by complementary, degenerative mutations *Genetics* **151**,
889 1531-1545
890 52. van de Locht, A., Lamba, D., Bauer, M., Huber, R., Friedrich, T., Kroger, B. *et al.* (1995)
891 Two heads are better than one: crystal structure of the insect derived double domain Kazal
892 inhibitor rhodniin in complex with thrombin *The EMBO journal* **14**, 5149-5157
893 53. Maxwell, M., Undheim, E. A. B., and Mobli, M. (2018) Secreted Cysteine-Rich Repeat
894 Proteins "SCREPs": A Novel Multi-Domain Architecture *Front Pharmacol* **9**, 1333
895 54. Liu, J., Maxwell, M., Cuddihy, T., Crawford, T., Bassetti, M., Hyde, C. *et al.* (2023)
896 ScrepYard: An online resource for disulfide-stabilized tandem repeat peptides *Protein Sci*
897 **32**, e4566

898 55. Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002) MAFFT: a novel method for rapid
899 multiple sequence alignment based on fast Fourier transform *Nucleic Acids Res* **30**, 3059-
900 3066
901 56. Katoh, K., and Standley, D. M. (2013) MAFFT multiple sequence alignment software
902 version 7: improvements in performance and usability *Mol Biol Evol* **30**, 772-780
903 57. Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M., and Barton, G. J. (2009)
904 Jalview Version 2--a multiple sequence alignment editor and analysis workbench
905 *Bioinformatics* **25**, 1189-1191
906 58. Slater, G. S., and Birney, E. (2005) Automated generation of heuristics for biological
907 sequence comparison *BMC Bioinformatics* **6**, 31
908 59. Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von
909 Haeseler, A. *et al.* (2020) IQ-TREE 2: New Models and Efficient Methods for Phylogenetic
910 Inference in the Genomic Era *Mol Biol Evol* **37**, 1530-1534
911 60. Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., and Vinh, L. S. (2018)
912 UFBoot2: Improving the Ultrafast Bootstrap Approximation *Mol Biol Evol* **35**, 518-522
913 61. Rambaut, A. (2018) FigTree (Version 1.4.4) Institute of Evolutionary Biology, University
914 of Edinburgh,
915 62. Khilji, M. S., Hackney, C. M., Koch, T. L., Hone, A. J., Rogalski, A., Watkins, M. *et al.*
916 (2025) Structural similarities reveal an expansive conotoxin family with a two-finger toxin
917 fold *Protein Sci* **34**, e70333
918 63. Wilkins, M. R., Gasteiger, E., Bairoch, A., Sanchez, J. C., Williams, K. L., Appel, R. D. *et*
919 *al.* (1999) Protein identification and analysis tools in the ExPASy server *Methods Mol Biol*
920 **112**, 531-552

921 64. Kazimierczuk, K., and Orekhov, V. Y. (2011) Accelerated NMR spectroscopy by using
922 compressed sensing *Angew Chem Int Ed Engl* **50**, 5556-5559
923 65. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe:
924 a multidimensional spectral processing system based on UNIX pipes *J Biomol NMR* **6**,
925 277-293
926 66. Skinner, S. P., Fogh, R. H., Boucher, W., Ragan, T. J., Mureddu, L. G., and Vuister, G. W.
927 (2016) CcpNmr AnalysisAssign: a flexible platform for integrated NMR analysis *J Biomol*
928 *NMR* **66**, 111-124
929 67. Güntert, P., and Buchner, L. (2015) Combined automated NOE assignment and structure
930 calculation with CYANA *J Biomol NMR* **62**, 453-471
931 68. Schwieters, C. D., Kuszewski, J. J., Tjandra, N., and Clore, G. M. (2003) The Xplor-NIH
932 NMR molecular structure determination package *J Magn Reson* **160**, 65-73
933 69. Bermejo, G. A., Clore, G. M., and Schwieters, C. D. (2012) Smooth statistical torsion angle
934 potential derived from a large conformational database via adaptive kernel density
935 estimation improves the quality of NMR protein structures *Protein Sci* **21**, 1824-1836
936 70. Tian, Y., Schwieters, C. D., Opella, S. J., and Marassi, F. M. (2017) High quality NMR
937 structures: a new force field with implicit water and membrane solvation for Xplor-NIH *J*
938 *Biomol NMR* **67**, 35-49
939 71. Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kaptein, R., and Thornton, J. M.
940 (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein
941 structures solved by NMR *J Biomol NMR* **8**, 477-486

942 72. Olp, M. D., Kalous, K. S., and Smith, B. C. (2020) ICEKAT: an interactive online tool for
943 calculating initial rates from continuous enzyme kinetic traces *BMC Bioinformatics* **21**,
944 186
945 73. Teichert, R. W., Smith, N. J., Raghuraman, S., Yoshikami, D., Light, A. R., and Olivera, B.
946 M. (2012) Functional profiling of neurons through cellular neuropharmacology *Proc Natl*
947 *Acad Sci USA* **109**, 1388-1395

948