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 in which these proteins face an adaptive conflict, where mutations that enhance one activity compromise the other. Collectively, this work provides new structural and functional insights into a previously uncharacterized toxin superfamily in cone snails, illustrates how structural scaffolds can be repurposed for functions that diverge from the original while retaining their overall structure, and expands our understanding of the toxin arsenal available to venomous animals.

42 **INTRODUCTION**

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

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

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

64 “weaponized” (6). Through this process, these compact and stable structural scaffolds can be
65 repurposed for functions that diverge from the original, while retaining their overall structure.

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

76 A prominent example of a protease-inhibitor fold found in several venoms is the Kazal-
77 type domain that inhibits a broad range of serine proteases. Kazal-type venom peptides are found
78 in various venomous animals, including assassin bugs, shrew, honeybees, and wasps, where they
79 have been associated with functions like anti-coagulation (16-18), pro-coagulation (19),
80 antimicrobial activity (18), melanization in insects (20), and analgesia (19). In humans, the SPINK
81 family of Kazal-type protease inhibitors (KPIs) plays important functions in regulating cellular
82 protease activity. This is exemplified by the pancreatic trypsin inhibitor SPINK1, in which
83 sequence variants are associated with pancreatitis and pancreatic cancer (21). Another example of
84 a well-characterized KPI is greglin from the locust *Schistocerca gregaria*, which inhibits both
85 elastase and subtilisin with low nanomolar Ki-values (22). In general, KPIs function as substrate
86 analogs. They act by inserting their inhibitory P1 residue, the residue positioned immediately

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

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

106 Here, we surveyed venom gland transcriptomes to identify new macro-conotoxin families.
107 We describe the previously uncharacterized MKAVA family of conotoxins and show by NMR
108 spectroscopy that an MKAVA-family toxin from *Conus magus*, conkazal-M1, comprises a single
109 non-classical Kazal-type domain with both protease inhibitor activity and the ability to inhibit

110 KCl-induced calcium influx in a subset of dorsal root ganglia (DRG). We also uncover the
111 evolutionary origin of the MKAVA-family toxins and discuss potential mechanisms that shaped
112 their molecular evolution.

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). As illustrated in Figure 1 and S1, sequences from snails with a prey preference from all three feeding groups of venomous snails (fish, other molluscs, and worms) are present among the analyzed sequences. 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 sequence 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 (based on 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

136 K48) evolved under weak positive selection with >90% likelihood based on dN/dS ratios (see
137 Supplementary File 2).

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

151

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

154 To investigate the structure and function of the MKAVA superfamily, we selected a high-
155 expressing precursor sequence from the magician cone, *C. magus*, for recombinant expression and
156 characterization. This toxin, the first characterized member of its family, was found to adopt a KPI
157 structure (see below), and was therefore named "conkazal-M1" (where M stands for *magus*).

158 Accordingly, we will use the term "conkazal" rather than MKAVA to refer to this family of
159 conotoxins.

160 Conkazal-M1 was recombinantly expressed in *E. coli* using the csDisCoTune system (29).
161 This expression platform, a recent further development of the original CyDisCo system (30) and
162 our own csCyDisCo system (31), is specifically tailored for the expression of conotoxins and
163 allows for the production of correctly folded, fully oxidized disulfide-rich proteins in the bacterial
164 cytosol by co-expression of three foldases encoded by an auxiliary plasmid – the Erv1p oxidase,
165 human protein disulfide isomerase (hPDI), and a conotoxin-specific protein disulfide isomerase
166 (csPDI) (32). Conkazal-M1 was expressed as a fusion protein with an N-terminal ubiquitin (Ub)
167 tag containing an internal loop modified to contain ten consecutive histidine residues (His_{10}) and
168 a tobacco etch virus protease (TEVp) recognition site, allowing for release of the mature toxin
169 (33).

170 SDS-PAGE analysis of small-scale test expressions of Ub-His_{10}-conkazal-M1 alone, Ub-
171 His_{10}-conkazal-M1 with the csDisCoTune-encoded foldases, or with the foldases encoded by the
172 csCyDisCo expression platform clearly showed the beneficial effect of both expression systems
173 on Ub-His_{10}-conkazal-M1 production. Both systems resulted in the expression of soluble Ub-
174 His_{10}-conkazal-M1 fusion protein, in contrast to expression without the foldases, where equivalent
175 total protein levels were observed but the fusion protein was predominantly found in the insoluble
176 fraction in a reduced and/or partially oxidized state (Figure S3a). Because the csDisCoTune system
177 produced higher levels of oxidized Ub-His_{10}-conkazal-M1 compared to the csCyDisCo system,
178 large-scale expression was carried out with csDisCoTune. Cleavage of affinity-purified Ub-His_{10}-
179 conkazal-M1 by TEVp resulted in liberation of the 73-residue mature region of conkazal-M1,
180 which was purified (>95% purity) to a final yield of just over 5 mg per liter culture (Figure S3b).

181 Disulfide-bond formation in the purified toxin was verified by SDS-PAGE analysis, where
182 a clear downward mobility shift was observed upon addition of a reducing agent (Figure S3b), as
183 well as by Q-TOF mass spectrometry, which revealed the presence of two species in the purified
184 preparation. The primary species displayed a monoisotopic mass of 7990.54 Da, in accordance
185 with the theoretical mass (7990.55 Da) of the fully oxidized protein (Figure S3c). The secondary
186 component exhibited a monoisotopic mass of 7973.49 Da, i.e., 17 Da lower than the theoretical
187 mass. We attribute this reduction in mass to modification of the N-terminal Gln to a pyroglutamate,
188 a post-translational modification commonly found in conotoxins (34). This modification can be
189 enzymatically catalyzed by glutaminyl cyclase but is also known to occur spontaneously (35).

190

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

192 The three-dimensional structure of ^{15}N and ^{13}C stable isotope-labelled conkazal-M1 was solved in
193 solution using distance and torsion angle restraints obtained by NMR spectroscopy. We assigned
194 the chemical shifts of 79% of the backbone ^{1}H, ^{13}C, and ^{15}N resonances and 66% of sidechain
195 resonances. In total, 69% of the chemical shifts were assigned. We observed no resonances from
196 the first seven residues. 602 short distances assigned in NOESY spectra were combined with 100
197 dihedral angles calculated from chemical shifts and used as restraints in structure calculations
198 (Table 1). The structure is well ordered from residue 10 to the C-terminus with an RMSD value of
199 0.31Å for the positions of the backbone atoms in the 20 models that represent the structure (Figure
200 S4a).

201 The structure consists of a central three-stranded antiparallel β-sheet and two α-helices: α-
202 helix 1, which is parallel to the β-strands, and α-helix 2, which is almost perpendicular to α-helix

203 1 (Figure 2a and b). The N- and C-termini are tethered to the core structure via two disulfide
204 bridges. Located at each end of α-helix 1, Cys40 and Cys48 form disulfide bonds with the N-
205 terminal Cys14 and Cys4, respectively. The C-terminus is affixed to the short loop connecting β-
206 strand 2 and α-helix 1 by the bridge connecting Cys37 and Cys68. The last disulfide bond is formed
207 by Cys24 and Cys59, thus connecting β-strand 1 and α-helix 2.

208 As structural similarity often correlates with conserved functional properties, we used
209 FoldSeek (36) and PDB eFold (37) to identify proteins structurally similar to conkazal-M1. Both
210 search engines identified several Kazal-type protease inhibitors, such as greglin (38), among the
211 most similar structures to conkazal-M1. Conkazal-M1 and greglin (Figure 2a) superpose with an
212 RMSD value of 2.87Å over 48 residues (Figure 2b). Both structures harbor the KPI core structure
213 as well as an additional C-terminal α-helix not found in classical KPIs. The three characteristic
214 disulfide bridges of the classical KPIs are present in both conkazal-M1 and greglin and similarly
215 positioned in the two structures (blue arrowheads, Figure 2b). Both proteins harbor a fourth
216 disulfide bridge that tethers the C-terminal regions to the core structure of the proteins (Figure 2a;
217 orange). Finally, conkazal-M1 has a longer loop between the first and second cysteine residues.

218

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

220 Because conkazal-M1 adopts a Kazal fold, we speculated that the toxin may exhibit inhibitory
221 activity against serine proteases. Consistent with this hypothesis, Glu16 at the P1 position in
222 conkazal-M1 is solvent-exposed in all 20 of the lowest energy structures (Figure S4a). Still, we
223 note that although AlphaFold accurately predicts the overall fold of the conkazals, as illustrated
224 for Conkazal-S1 from *C. striatus* (Figure S4b), the Glu side chain at the P1 position points inward

225 in many structural models and is predicted to form hydrogen bonds with the main chain NH groups
226 from Tyr38 and Cys39. This observation informed our decision to determine the NMR structure
227 of the protein and provides an important reminder that AF does not always place side chains
228 accurately (39).

229 To experimentally determine a potential function of conkazal-M1 as a protease inhibitor,
230 we investigated the proteolytic activities of chymotrypsin, Endoproteinase GluC, and Subtilisin A
231 (SubA) in the presence of increasing amounts (0.25 μM – 3 μM) of conkazal-M1. GluC was chosen
232 as a possible candidate due to its preferential cleavage of peptide bonds C-terminal to glutamic
233 acid residues. SubA is a non-specific protease that favors cleavage after hydrophobic residues but
234 is not restricted to them. For instance, a Glu residue at the P1 position in the second domain of the
235 CrSPI two-domain Kazal-type protein from horseshoe crab is the primary residue for mediating
236 the contacts with subtilisin in the co-crystal of the complex between the two proteins (40).

237 SubA activity was monitored using the chromogenic substrate N-succinyl-Ala-Ala-Pro-
238 Phe-pNA over a range of substrate concentrations in the presence and absence of conkazal-M1
239 (Figure 3). Initial rates were estimated from changes in absorbance at 405 nm over time and were
240 converted to reaction rates (μM/min/mg enzyme). Nonlinear regression of the initial rate data was
241 performed using Michaelis-Menten and competitive, non-competitive and uncompetitive
242 inhibition models. Model comparison by Akaike Information Criterion indicated that competitive
243 inhibition best described the data – as expected for a KPI – with estimated kinetic parameters of
244 Km=201 μM (Confidence Interval (C.I.): 173‑234 μM), Vmax= 12,888 μM/min/mg (C.I.: 12,357-
245 13,456 μM/min/mg) and Ki=1.1 μM (C.I.: 0.92-1.5 μM) (Figure 3). We performed similar
246 investigations with chymotrypsin and Endoproteinase GluC in the absence and presence (0.5 μM-
247 2 μM) of conkazal-M1, where no effect of the toxin on enzyme activity was discerned (Figure S5).

248

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

250 A wide variety of venom peptides function as neurotoxins that target neuronal ion channels.
251 Therefore, we next assessed the bioactivity of conkazal-M1 using the calcium imaging-based
252 constellation pharmacology assay (Teichert et al, 2012). This assay enables the monitoring of Ca^{2+}
253 influx in primary cell cultures obtained from mouse dorsal root ganglions (DRGs), which comprise
254 various types of sensory neurons, including nociceptors, mechanoreceptors, and proprioceptors,
255 that respond to different sensory modalities. During each experiment, intracellular calcium levels
256 were simultaneously monitored in 500-1,000 cells using the Fura-2-AM dye. Cells were assigned
257 to seven major classes based on their size, isolectin B4 staining (non-peptidergic nociceptors) and
258 calcitonin gene-related peptide (CGRP) expression (peptidergic nociceptors) (Fang et al, 2010).
259 To further characterize neuronal subtypes, a series of pharmacological agents (conotoxin *κ*M-
260 RIIIJ, allyl isothiocyanate [AITC], menthol, and capsaicin) were applied at the end of each
261 experiment (Giacobassi et al, 2020) (Figure 4a).

262 Dissociated DRG neurons were depolarized using 15 s pulses of high-potassium
263 extracellular solution (30 mM KCl), resulting in intracellular Ca^{2+} elevation, as indicated by the
264 rise in Fura-2 signal (Figure 4a). At the lowest concentration tested (1 μM), conkazal-M1 did not
265 produce significant effects on any neuronal subclasses. In contrast, application of 10 μM conkazal-
266 M1 reduced the KCl–evoked calcium signals in a subset of cells (12% of the 4,215 cells analyzed)
267 (Figure 4b). The blockade of calcium transients was predominantly observed in small (<400 μm^2)
268 peptidergic nociceptive neurons (33.9% of all affected cells) (identified by the expression of CGRP
269 genetically labeled with green fluorescent protein) (Figure 4b). Other neuronal subclasses affected
270 by 10 μM conkazal-M1 included large-diameter (>400 μm^2) mechanoreceptors (8%), a subset of

271 non-peptidergic nociceptors (33.1%), and C-low threshold mechanoreceptors (C-LTMR; 13.5%).
272 Thus, conkazal-M1 inhibits KCl-induced calcium influx in a subset of neuronal cells, albeit with
273 low affinity. To account for the observation that the frequency of neuronal cell types varies
274 considerably among DRG neurons (Figure S6), we quantified the block probability according to
275 cell type (Figure 4c). Most strikingly, considering that proprioceptors constitute only ~ 3% of all
276 cells (Figure S6), 27% of these were blocked by conkazal-M1 (Figure 4c), indicating that the target
277 of conkazal-M1 is likely most prevalent in this population.

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

284

285 **The conkazals are evolutionarily related to schistosomin**

286 To identify potential homologs of conkazals outside Conoidea and thereby gain insight into the
287 evolutionary history of these proteins, we used the conkazal-M1 precursor sequence to perform
288 homology searches against the NCBI non-redundant protein database and transcriptome shotgun
289 assemblies.

290 The searches revealed sequence similarity, including conservation of all eight cysteines,
291 with schistosomin-like proteins from several molluscs (Figures 5a and S7). Schistosomin is
292 seemingly absent in other phyla than molluscs and was first described in the freshwater pond snail,

293 *Lymnaea stagnalis*, as a peptide upregulated in response to infection with the schistosome parasite
294 *Trichobilharzia ocellata* (41) (see also Discussion). Overall, only modest sequence conservation
295 is observed across the conkazals and schistosomins. In addition to the cysteines, three residues
296 show high conservation. These comprise a glutamate residue located two residues after the second
297 cysteine in both groups of proteins, i.e. at the P1 position of KPIs, at position 34 in the alignment,
298 a proline immediately following the first Cys residue, and a positively charged residue at position
299 62 in the alignment (Figure 5a and S7).

300 A CLANS (CLuster ANalysis of Sequences) clustering analysis (42), which groups protein
301 families based on all-against-all pairwise protein sequence similarities, revealed a clear similarity
302 of the *Conus*-derived conkazal sequences with molluscan schistosomins from gastropods (Figure
303 5b; all sequences used in this analysis are provided in Supplementary File 3). Moreover, gene
304 structure analysis revealed a common gene structure of conkazals and molluscan schistosomins.
305 The genes have a common phase-1 intron located immediately downstream of the signal sequence
306 and a conserved phase-2 intron located between the sixth and seventh cysteine residues (Figure
307 5c). The common gene structures further support an evolutionary relationship between conkazal
308 conotoxins and schistosomins.

309 Given the conservation of all cysteine residues and the common evolutionary origin, it was
310 not surprising that the AlphaFold-predicted structure of a schistosomin from *Lymnaea stagnalis*
311 was found as the top hit in the FoldSeek search with conkazal-M1. We also identified an
312 unpublished crystal structure of schistosomin from another freshwater snail, *Biomphalaria*
313 *glabrata* (PDB ID: 9FDO), in the PDB. Conkazal-M1 and this schistosomin (Figure S8a)
314 superpose with an RMSD value of 1.9 Å over 38 Cα-atoms (Figure S8b). Both structures adopt
315 the KPI core structure as well as an additional C-terminal α-helix not found in classical KPIs. The

316 three characteristic disulfide bridges of the classical KPIs are present in both conkazal-M1 and
317 schistosomin and are similarly positioned in the two structures (light blue arrowheads, Figure S8b).
318 Both proteins harbor a fourth disulfide bridge that tethers the C-terminal regions to the core
319 structure of the proteins (Figure S8a).

320 While molluscan species exhibit a high degree of intrachromosomal gene translocation,
321 macrosynteny tends to be conserved. We therefore compared the genomic location of conkazal
322 genes from *C. ventricosus* with schistosomins from representative gastropods (Figure S9). The two
323 conkazal genes in *C. ventricosus* are adjacent to each other on scaffold 23, which represents one
324 of the chromosome-scale scaffolds in the cone snail genome. This chromosome shows homology
325 to chromosome NC_037594.1 of the gastropod *Pomacea canaliculata* and chromosome
326 NC_074711.1 of another gastropod, *Biomphalaria glabrata* (Figure S9, left). However, neither of
327 these chromosomes contains the schistosomin genes, which are located on NC_037595.1 and
328 NC_074716.1, respectively. Surprisingly, both *P. canaliculata* and *B. glabrata* have multiple
329 schistosomin gene copies located in tandem, though the neighboring genes show no syntenic
330 relationship (Figure S9, right). Overall, while conkazals show a high level of similarity to
331 schistosomins, strongly suggesting an evolutionary relationship, the genes are not part of a
332 common gene linkage group.

333

334

335 **DISCUSSION**

336 Here, we provide the first characterization of a toxin from the conkazal conotoxin family, conkazal-
337 M1. The conkazals constitute a large and widespread toxin family found across phylogenetically
338 diverse *Conus* species (from all three feeding groups) and several other species in the Conoidean
339 superfamily, indicating an important functional role of these peptides. The common evolutionary
340 history with the schistosomins indicates that an ancestral gene was recruited into the venom in
341 early Conoidea and then acquired neuroactive functions. However, given that the function of
342 schistosomins remains unclear, the exact process of recruitment is presently unknown.

343 Many conkazal transcripts are highly expressed in the venom gland, and we confirmed
344 their presence in venom by proteomic analysis, supporting a role as toxins. We show that conkazal-
345 M1 can perform dual functions as a protease inhibitor and as a partial inhibitor of KCl-stimulated
346 calcium influx in sensory neurons. Such dual activities have also been reported in Kunitz-type
347 venom proteins (11-14). In the context of constellation pharmacology, the partial suppression of
348 KCl-evoked calcium transients by conkazal-M1 provides preliminary evidence that this peptide
349 can modulate ion channel activity (directly or indirectly) in sensory neurons. The modest inhibition
350 at 10 μM indicates either a low-affinity or indirect mechanism, potentially consistent with dual-
351 function peptides in which ion-channel modulation is a secondary activity to protease inhibition,
352 as further discussed below. Future studies are needed to identify the target and mode of action of
353 conkazal-M1 in neuronal cells.

354 In the context of a protease inhibitor function, subtilisin-targeting KPIs with a Glu at the
355 P1 position have previously been characterized (see e.g., (40, 43-45)). It is therefore unsurprising
356 that conkazal-M1, with its Glu16 at the P1 position, was able to inhibit subtilisin (although
357 additional residues in the reactive-site loop likely also contribute to substrate specificity). While

358 certain venoms appear to harbor microbiomes (46), bacterial subtilisin is an unlikely physiological
359 target of the protease-inhibitory function of conkazal-M1. This is also reflected in the relatively
360 high K_i value of conkazal-M1 for subtilisin when compared to many other subtilisin inhibitors,
361 such as greglin. In principle, as a protease inhibitor, conkazal-M1 could function either in the
362 venom, to regulate venom protease activity and maintain stability before delivery, or in the prey,
363 to exert a deleterious effect upon delivery. At present, we cannot distinguish between these
364 possibilities, although we note that we have found no evidence in the literature for an endogenous
365 regulatory function in the venom of protease inhibitors in other venomous organisms.

366 When comparing conkazal sequences with the schistosomins, the Glu residue at the P1
367 position is one of the few residues conserved across species (Figure 5). This observation supports
368 the hypothesis that the conkazals and mollusc schistosomins function as KPIs. However, Glu33 in
369 the crystal structure of schistosomin is not solvent-exposed (as would be expected of a P1 residue)
370 but points towards the interior of the structure (Figure S8b). Overall, the potential function of
371 schistosomin as a KPI remains unresolved. In addition to the eight cysteine residues, only two
372 other residues are as highly conserved in conkazals and schistosomins as the P1 Glu: a proline
373 residue immediately following the first Cys residue, likely playing a structural role, and a surface-
374 exposed lysine residue at a position corresponding to Lys43 in the mature conkazal-M1 sequence,
375 which is likely to have a functional role.

376 Schistosomiasis, a parasitic infection, is caused by flatworms known as schistosomes. In
377 the schistosomiasis life cycle, freshwater snails serve as intermediate hosts that release infectious
378 larvae into the water. These larvae can penetrate human skin, gaining access to the host, where the
379 subsequent release of eggs can cause a range of symptoms, including serious damage to organs
380 such as the liver and bladder. In the snail-schistosome system *L. stagnalis*-*T. ocellata*, schistosomin

381 was initially proposed to be a neuropeptide induced by parasite infection, but subsequent studies
382 showed that it is produced in connective tissue cells and in hemocytes (reviewed in (47)). In the
383 hemolymph, schistosomin was proposed to act as an antagonist of gonadotropic hormones, thereby
384 inhibiting reproductive activity in snails. However, in the *Biomphalaria glabrata-Schistosoma*
385 *mansoni* system, no correlation was found between parasite infection and schistosomin levels and
386 infection did not result in parasitic castration (48). Thus, the physiological function of schistosomin
387 remains unclear, and further investigation is warranted, particularly with its potential function as
388 a KPI in mind.

389 An ancestral protease inhibitor that acquires the ability to modulate neuronal activity while
390 retaining protease inhibition represents a functional change in which purifying selection preserves
391 the original function, while weak-to-moderate selection can favor the emergence of a novel activity
392 (6, 49). In contrast, a complete shift toward modulation of neuronal activity accompanied by partial
393 or total loss of protease activity can be explained by strong positive selection for the new function,
394 coupled with relaxed constraint on the ancestral activity. Because we detect no evidence of strong
395 positive selection, we favor the first evolutionary model. In this scenario, conkazals would be
396 "moonlighting" proteins, facing an adaptive conflict when mutations that enhance one activity
397 compromise the other. Such conflicts are typically resolved through gene duplication, followed by
398 subfunctionalization, which partitions the ancestral roles between paralogs, or through
399 neofunctionalization of one copy. Our data do not support a neofunctionalization event following
400 duplication. If such an event had occurred, we would expect to detect both an ancestral,
401 schistosomin-like gene and a duplicated conkazal gene. Instead, we only observe conkazal genes
402 when interrogating available cone snail genome data, consistent with a moonlighting role. This
403 interpretation is further supported by the expression of conkazals in tissues other than the venom

404 gland, including the venom bulb, nerve ring, and foot (see Materials and Methods section for SRA
405 entry numbers), where these peptides likely play an endophysiological function.

406 We note that multi-domain Kazal-type proteins are found in many venomous animals,
407 including the two-domain Kazal-type toxins BQTX from shrews (19) and Rhodniin from the
408 kissing bug *Rhodnius prolixus* (17, 50). In general, secreted cysteine-rich repeat-proteins (51) –
409 including a broad range of venom peptides – are common in nature and have been cataloged in the
410 ScrepYard database (52). Their multi-domain structure enables multivalent substrate interactions,
411 allowing them to bind targets strongly through an avidity effect. Indeed, Rhodniin binds its
412 substrate thrombin through multivalent interactions involving the reactive-site loop of one domain
413 and electrostatic interactions of the second domain, together conferring a picomolar affinity (17,
414 50). Analogous to the two-domain Kazal-type toxins described above, we have recently identified
415 a family of such conotoxins, the MKTAV family. As shown in Figure S10a-b, both Kazal domains
416 of these conotoxins, with few exceptions, contain a Glu or Asp residue three positions C-terminal
417 to the second Cys in the mature sequence. Although not located at the canonical P1 position, we
418 speculate that this residue could still perform the inhibitory function of a P1 residue. The
419 AlphaFold 3 predictions of the MKTAVs reveal two very similar Kazal-type domains connected
420 by a short linker, with each domain closely resembling the structure of conkazal-M1 (Figure S10c-
421 e).

422 In this study, we identify two previously unrecognized families of conotoxins, conkazal
423 and MKTAV, which encode single- and two-domain Kazal-type structures, respectively. We
424 uncover the dual function of conkazal-M1, acting both as a neuroactive peptide with the ability to
425 inhibit calcium transients in a subset of DRGs and as a protease inhibitor. Together, these
426 observations highlight the structural and functional diversity of Kazal-type toxins in *Conus*, extend

427 the current knowledge about the toxin complement available to venomous animals, and illustrate
428 the adaptive conflict that can shape the evolutionary trajectories of dual function toxins.

429

430

431

MATERIALS AND METHODS

Sequence mining and analysis

Multiple sequence alignments were generated with the MAFFT version 7 (53) online interface using the G-INSi method (54). Alignments were visualized in JalView (55). Schistosomin sequences were obtained using web version of blast to query against the non-redundant (nr) protein database and transcriptome shotgun assemblies with standard settings. Conotoxin sequences were obtained from (25). These sequences were used as input for CLANS (42), which is a sequence-based clustering program that assigns attractive values between sequences based on their BLOSUM62 blastp p-value. In a uniformly repelling force field this causes similar sequences to cluster together. Gene structure analyses were performed with using the protein2genome (p2g) model in Exonerate (v.2.76.2) (56). Macrosynteny analysis was performed with macrosyntR using OrthoFinder to assign orthologous proteins across *C. ventricosus* (GCA_018398815.2)*, P. canaliculata* (GCF_003073045.1)*,* and *B. glabrata* (GCF_947242115.1). The local gene neighborhoods were assessed and visualized in Figure S9. Maximum likelihood gene trees were constructed using alignments generated with MAFFT v.7.526 and the tree was constructed using IQ-TREE v. 2.2.2.3 (57) with UF bootstrap (58) calculation for 1000 replicates. The best model according to the Bayesian Information Criterion was calculated to be TMP3 + G4. The consensus tree was visualized using FigTree (59). Conkazal-like sequences were identified in the NCBI SRA database using the conkazal-M1 amino acid sequence as a query sequence in tblastn searches that retrieved the following SRA entries: SRX8002177, SRR11423812 (Drillidae), SRR26897535, SRR26897526 (Turridae), SRR10192879, SRR10192877 (Profundiconus), SRR2060989 (Terebridae), SRR16493587 (venom bulb), SRR16493599, SRR16493598 (nerve ring), and SRR16493591 (foot).

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

456 Proteomics files of reduced, alkylated and trypsin digested venom of *Conus geographus*
457 (accession: PXD052869) and *Conus marmoreus* (PXD038992) were retrieved from the PRIDE
458 ProteomeXChange repository and searched for MKAVA sequences using the Byonic software
459 (version 3.2.0 Protein metrics) with the following settings: cleavage site: RK, missed cleavage: 1,
460 Modifications: carbamidomethyl (C, +57.02146), Oxidation (M,P, +15.994915), pyro-Glu (Q,
461 +17.026549), Carboxy (E, +43.989829). High scoring peptide hits (score> 400) were manually
462 verified.

463

464 **Plasmid generation**

465 The codon-optimized sequence encoding mature conkazal-M1 was synthesized and inserted
466 (TWIST Bioscience) into the pET39_Ub19 expression vector (33). The resulting plasmid
467 (pLE657) encodes ubiquitin (Ub)-His_{10}-tagged conkazal-M1 (Ub-His_{10}-conkazal-M1) with a
468 recognition site for TEV protease located immediately N-terminally to the mature toxin sequence.

469

470 **Protein expression and preparation of lysate**

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

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

478 Initial small-scale expression tests were performed in 50 mL LB medium containing
479 appropriate antibiotics and supplemented with 0.05% glucose. Test cultures were inoculated with
480 2% overnight culture and grown at 37° C, 190 rpm until OD_{600} reached ~0.8. Expression was
481 induced by addition of isopropyl -D-1-thiogalactopyranoside (IPTG) to a final concentration of 1
482 mM and the cultures grown for 18 hours at 25° C with shaking at 190 rpm to allow protein
483 expression. Large-scale expression was performed similarly in 1 L LB medium. Expression
484 performed in 0.5 L M9 minimal medium for stable isotope labelling was performed as described
485 in (60).

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

494

495 **Protein purification**

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

498 lysate was applied to the column and washed with 30 column volumes (CVs) of lysis buffer before
499 elution of Ub-His_{10}-conkazal-M1 with a gradient of 0% to 100% elution buffer (50 mM Tris (pH
500 8.0), 300 mM NaCl, 400 mM imidazole) developed over 20 CVs. Pooled fractions were dialyzed
501 twice against 2 L cleavage buffer (50 mM Tris (pH 8.0), 300 mM NaCl).

502 The Ub-His10-conkazal-M1 fusion protein was cleaved using His_6-tagged TEVp,
503 expressed, and purified essentially as described previously (27). A molar ratio Ub-His_{10}-conkazal-
504 M1: His_6-TEVp of 1:2 was used. To avoid reducing the disulfides in conkazal-M1, His_6-TEVp –
505 preactivated with 2 mM dithiothreitol (DTT) for 30 minutes at room temperature – was diluted to
506 approximately 0.002 mM DTT by 3 rounds of dilution/concentration in an Amicon Ultra 15 mL
507 10K Centrifugal Filter (Merck Millipore). TEVp cleavage was performed overnight at room
508 temperature.

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

517

518 **SDS-PAGE analysis**

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

523

524 **Protein concentration determination**

525 Concentration was determined by measuring absorbance at 280 nm and using the theoretical
526 extinction coefficient provided by the Expasy ProtParam tool available through the Expasy
527 bioinformatics resource web portal (61).

528

529 **Mass spectrometry**

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

533

534 **NMR spectroscopy**

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

540 HCCONH and HCCH-TOCSY spectra. Distance restraints were obtained from ^{15}N-NOESY-
541 HSQC and ^{13}C-NOESY-HSQC experiments recorded using a mixing time of 120 ms. The triple
542 resonance spectra were recorded with non-uniform sampling at 25% and reconstructed with
543 qMDD (62). All spectra were processed with nmrPipe (63) and analyzed in CCPNMR analysis
544 (64).

545

546 **NMR structure calculations**

547 Automated NOE assignment was performed using Cyana (65). The NOE list was manually refined.
548 XPLOR-NIH (66) was used for the final structural refinement including a torsion angle database
549 potential (67) and an implicit solvent mode (68). For the final ensemble representing the solution
550 structure of conkazal-M1, the 20 structures with the lowest energy were chosen from 100
551 calculated structures. The Ramachandran-plot statistics was calculated using PROCHECK (69)
552 and the coordinate precisions were calculated by PyMOL (DeLano Scientific). Structure
553 visualization was also performed with PyMOL. The structure has been deposited in the Protein
554 Data Bank (PDB, wwpdb.org) with PDB code 9SLR and chemical shifts and peaks lists from the
555 NOESY spectra have been deposited in the Biological Magnetic Resonance Bank
556 (http://www.bmrb.wisc.edu/) with ID 53313.

557

558 **Structural homology search**

559 Structures similar to conkazal-M1 were identified using FoldSeek (36), and structural
560 superpositions were executed manually in PyMOL (DeLano Scientific) utilizing the CEalign
561 command.

562

563 **Protease assays**

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

568 Inhibition assays with chymotrypsin and GluC were performed in assay buffer (50 mM
569 Tris, 2 mM $CaCl_2$, pH 7.8). Subtilisin A inhibition experiments were performed in assay buffer,
570 pH 8 supplemented with 5 mM $CaCl_2$. The proteases (3.6 nM SubA, 100 nM GluC and 400 nM
571 chymotrypsin) were incubated with varying concentrations of conkazal-M1 for 10 min at 29° C.
572 Proteolytic reactions were initiated by addition of the substrates, and absorbance at 405 nm was
573 recorded every 30 seconds for 30 minutes. The 96-well plate was shaken orbitally for 10 seconds
574 immediately prior to measurement. Initial reaction rates were estimated with the web-based tool,
575 IceKat (70). The absorbance-based rates for Subtilisin A were converted to reaction rates in
576 μM/min/mg using the extinction coefficient for AAPF-p-nitroaniline (ε = 8800 $M^{-1}cm^{-1}$) and an
577 estimated path length of 0.2 cm for 100 μL reaction volumes in 96-well plates. The absorbance-
578 based rates (Abs/min) for Chymotrypsin and GluC were plotted without conversion. The data were
579 fit to four models of enzyme inhibition: Michaelis-Menten (no inhibitor), competitive,
580 uncompetitive, and noncompetitive inhibition.

581 $$v = \frac{V_{max}[S]}{K_m + [S]}$$

582

583 $$v = \frac{V_{max}[S]}{K_m\left(1 + \frac{[I]}{K_i}\right) + [S]}$$

584

585 $$v = \frac{V_{max}[S]}{K_m + [S]\left(1 + \frac{[I]}{K_i}\right)}$$

586

587 $$v = \frac{V_{max}[S]}{K_m\left(1 + \frac{[I]}{K_i}\right) + [S]\left(1 + \frac{[I]}{K_i}\right)}$$

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

589

590 **Cell culture and constellation pharmacology**

591 Experiments were performed as previously described (71). Briefly, all lumbar DRG neurons from
592 C57BL/6 mice 30-40 days old of both genders were dissociated by trypsin digestion and
593 mechanical trituration and plated on polylysine-coated plates. The cells were incubated overnight
594 in a 5% CO_2 incubator at 37 °C in neuronal culture medium [minimal essential media
595 supplemented with 10% (vol/vol) fetal bovine serum, penicillin (100 U/mL), streptomycin (100
596 μg/mL), 1× Glutamax, 10 mM HEPES, and 0.4% (wt/vol) glucose, adjusted to pH 7.4]. One hour
597 before calcium imaging experiments, cells were loaded with 2.5 μM fura-2-acetoxymethyl ester
598 (Fura-2-AM, Sigma-Aldrich) and incubated at 37 °C. Cells loaded with Fura-2-AM dye were
599 excited intermittently with 340- and 380-nm light; fluorescence emission was monitored at 510
600 nm. An image was captured at each excitation wavelength, and the ratio of fluorescence intensities
601 (340/380 nm) was acquired every 2 s to monitor the relative changes in intracellular calcium
602 concentration in each cell as a function of time. Typically, 500-1,000 neurons were imaged for

603 each experiment. The calcium transients were elicited by 15 s applications of depolarizing stimulus
604 (30 mM KCl, except for stimulation prior to RIIIJ treatment, in which 20 mM KCl was used),
605 followed by four washes with extracellular solution [145 mM NaCl, 5 mM KCl, 2 mM CaCl2, 1
606 mM MgCl2, 1 mM sodium citrate, 10 mM HEPES, and 10 mM glucose, adjusted to pH 7.4]. For
607 activity screening, conkazal-M1 was applied in extracellular solution at 1 and 10 μM between KCl
608 pulses and activity was assessed by significant change in the KCl response after incubation with
609 conkazal-M1. Four different pharmacological agents were used for cell classification:
610 conotoxin *κ*M-RIIIJ (1 μM), AITC (100 μM), menthol (400 μM) and capsaicin (300 nM). At the
611 end of the experiment, cells were incubated with 0.7 mL of Hoechst stain (1000 μg/mL) and Alexa-
612 Fluor 647 Isolectin B4 (2.5 μg/mL) for 5 min. The data was acquired using NIS-Elements and
613 further processed with CellProfiler v.3.1.9 (Jones et al, 2008). Custom-built scripts in Python
614 v.3.7.2 and R were used for further data analysis and visualization. The experimental methods were
615 approved by the Institutional Animal Care and Use Committee (IACUC) of the University of Utah
616 (Protocol number: 17–05017).

617

618 **Statistical analysis of calcium imaging data**

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

624 *lm(Kauc15 ~ linear + toxin)*

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

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

Block responses were analyzed at the level of individual neurons using a generalized linear mixed-effects model (GLMM) implemented in R (version 4.x) with the lme4 and emmeans packages. For each cell, the binary outcome variable block10 μM (1 = blocked, 0 = unblocked) was modeled as a Bernoulli trial with a logit link, using the following structure:

$$\text{logit}(p_ij) = \beta 0 + \beta k[i] + uj$$

where p_ij is the probability that cell i in experiment j was blocked, βk[i] is the fixed effect associated with its neuronal class, and uj ~ N(0, σ^2_exp) is a random intercept capturing between-experiment variation. This formulation accounts for unequal numbers of cells across experiments and for correlations within experiments. Model fitting used the bobyqa optimizer with 200,000 iterations. Marginal means (estimated probabilities of block for each neuronal class, averaged across experiments) were obtained from the fitted model using emmeans with type = 'response', which returns predictions on the probability scale. The overall effect of neuronal cell type on block

648 probability was evaluated using a Wald chi-square test of the fixed effect in the GLMM. This test
649 indicated a highly significant difference among cell types ($\chi^2(8) = 276.98$, $p < 0.0001$). Pairwise
650 differences among cell types were evaluated with Tukey adjustment for multiple comparisons, and
651 odds ratios (ORs) with 95% confidence intervals were reported. For visualization, predicted
652 probabilities and confidence limits were plotted with ggplot2. Each black point represents the
653 model-estimated mean probability for a given neuronal class, error bars denote 95% Wald
654 confidence intervals, and faint gray points indicate observed block rates within individual
655 experiments. This approach provides both a population-level estimate and a visual impression of
656 inter-experiment variability.

657

658 **Fish behavioral assays**

659 All animal assays were conducted in accordance with the *Guide for the Care and Use of*
660 *Laboratory Animals* and were approved by the Institutional Animal Care and Use Committee
661 (IACUC) of the University of Utah (protocol #001629). Goldfish (*Carassius auratus*) of both
662 sexes (weight= 0.85 ± 0.11 g, mean ± SD) were obtained from a local supplier and used for the
663 experiments. An in-house prepared E3 Buffer (4.96 mM NaCl, 0.18 mM KCl, 0.33 mM
664 $CaCl_2 \cdot 2H_2O$, 0.4 mM $MgCl_2 \cdot 6H_2O$; pH 7.2) served as the vehicle. Groups of three fish each
665 received injections of either 13.5 µL E3 buffer alone (control) or E3 buffer containing the desired
666 concentration of conkazal-M1.

667 All injections were performed with a 1 mL insulin syringe (31 G). For intraperitoneal
668 injections, fish were positioned ventral side up on a chilled sponge embedded with E3 Buffer under
669 a dissecting microscope, and the solution was delivered into the abdominal cavity. For
670 intramuscular injections, fish were positioned laterally on a hard surface moistened with E3 buffer

671 under a dissecting microscope, and the solution was administered into the lateral muscle below the
672 posterior end of the dorsal fin. Following injection, all fish were maintained under observation for
673 three hours.

674

675 **TABLES**

676 **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

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

678 **ABBREVIATIONS**

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

688

689 **ACKNOWLEDGMENTS**

690 We thank Signe A. Sjørup, SBiNLab, University of Copenhagen, for recording mass spectra and
691 Dr. Paula Florez Salcedo, Genentech Inc., for kindly providing illustrations to Figs. 1 and S1.

692

693 **FUNDING**

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

699

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., M.L.G., 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.L.G., S.E., Z.G.A., 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.

FIGURES

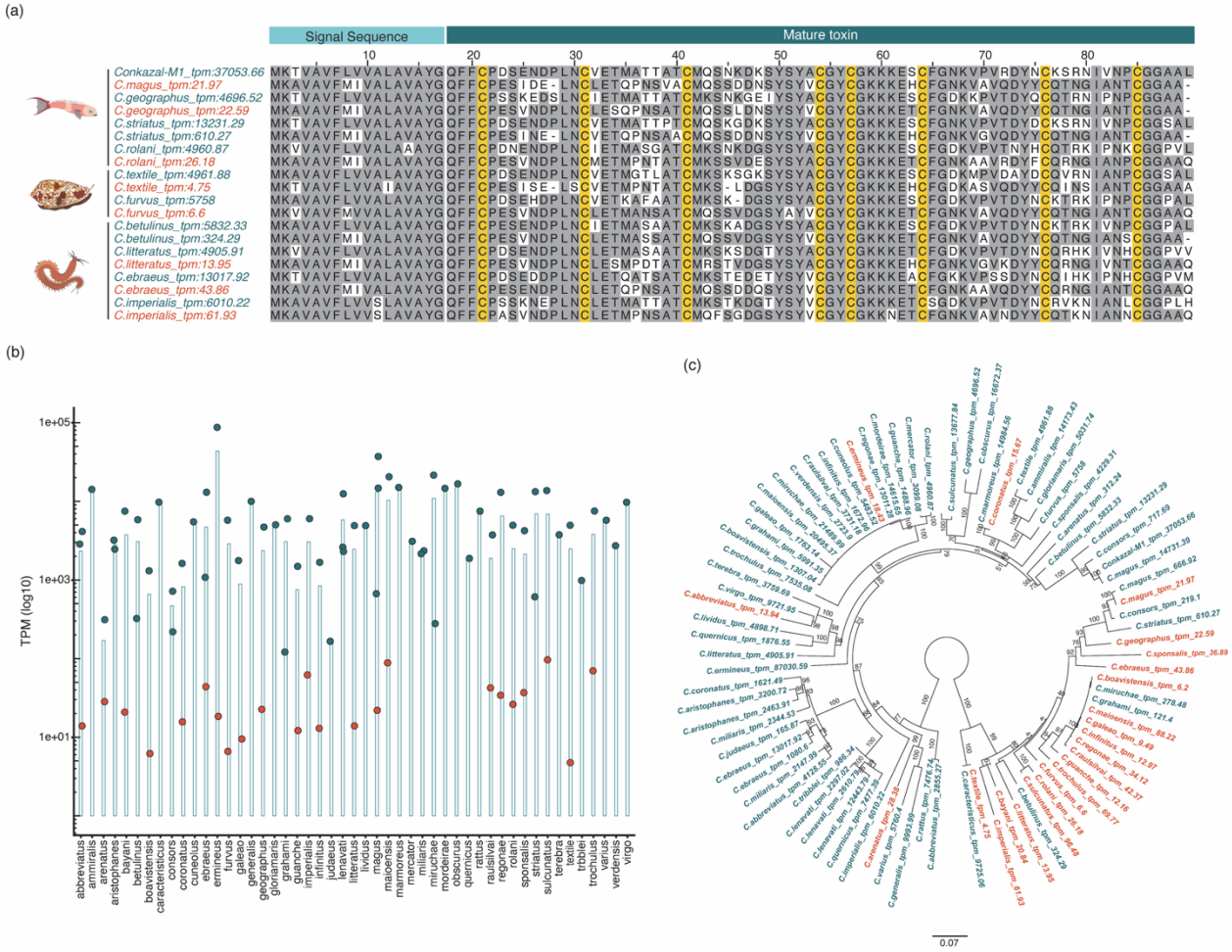

Figure 1. The MKAVAs are encoded by either low- or high-expressing transcripts. (a) Multiple sequence alignment of select *Conus* precursor sequences from the MKAVA superfamily. Sequences encoded by high- or low-expressing transcripts are indicated in dark teal and orange, respectively. The signal sequence and mature toxin regions are identified by light and dark teal bars above the alignment. The prey preference of each species (i.e. fish, other molluscs, or worms) is indicated by the illustrations, illustrations, kindly provided by Dr. Paula Florez Salcedo, on the left. **(b)** Gene expression levels of 85 MKAVA transcripts from 49 cone snail species. Dark teal and orange circles indicate high (>100 tpm) and low (<100 tpm) expressors, respectively, with light teal bars showing the average tpm-value of the transcripts for each species. **(c)** Phylogenetic

729 reconstruction of the MKAVA transcripts from (b). Maximum-likelihood was performed using IQ-
730 Tree's UF bootstrap with 1000 replicates, with best fit model TMP3 + G4 selected by the Bayesian
731 Information Criterion. High-expressing transcripts are shown in dark teal and low-expressing in
732 orange. Some bootstrap values were removed for visualization purposes but are shown in Figure
733 S11. The tree was rooted to separate the two paralogous genes.

734

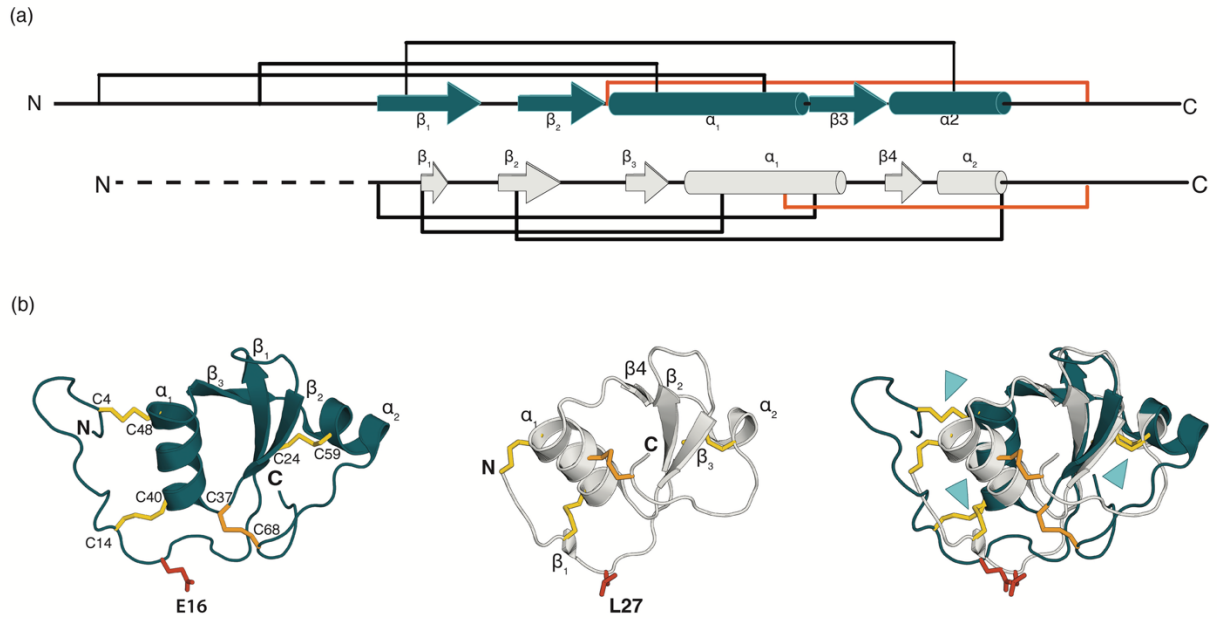

Figure 2. Conkazal-M1 harbors a KPI fold. (a). Graphic representations of Conkazal-M1 (teal) and the Kazal-type protease inhibitor greglin from the locust *Schistocerca gregaria* (gray). Disulfide bridges are represented as brackets. The additional disulfide bridge in conkazal-M1 and greglin (compared to classical KPIs) is shown in orange. The dotted line represents the first 19 residues of greglin not present in the sequence used for X-ray crystallography. **(b)**. Cartoon representation of the conkazal-M1 NMR structure (PDB: 9SLR) (left), cartoon representation of the greglin X-ray crystallography structure (PDB:4GI3) (middle), and superposition of conkazal-M1 and greglin with the three conserved disulfide bridges labeled with blue arrowheads (right). Disulfide bridges are colored yellow or orange in the case of the additional disulfide bride, and the amino acid residue at the P1 position is shown in red.

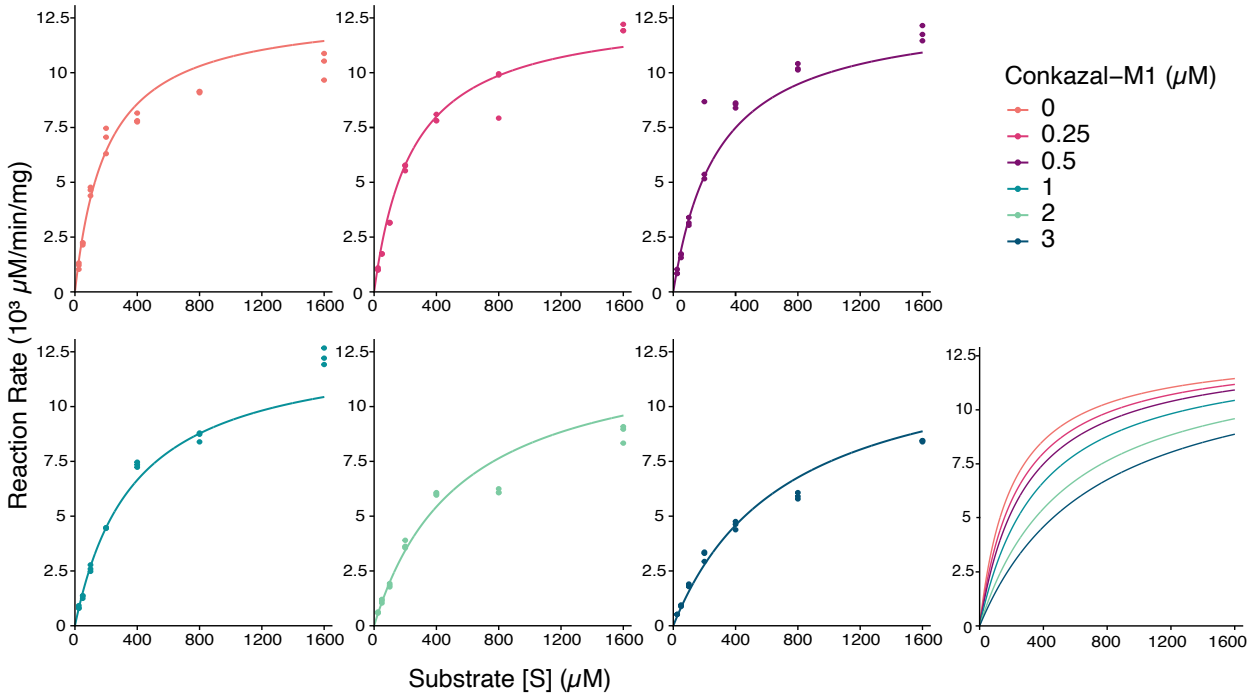

Figure 3. Conkazal-M1 inhibits proteolytic activity of Subtilisin A. Kinetic analysis of Subtilisin A in the presence of increasing concentrations of conkazal-M1. Reaction rates were measured across a range of substrate concentrations and were normalized to enzyme mass. The plots display experimental 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 the best-fit model (competitive inhibition). The lower right plot shows all fits overlaid without datapoints to more clearly illustrate the inhibitory effect of conkazal-M1.

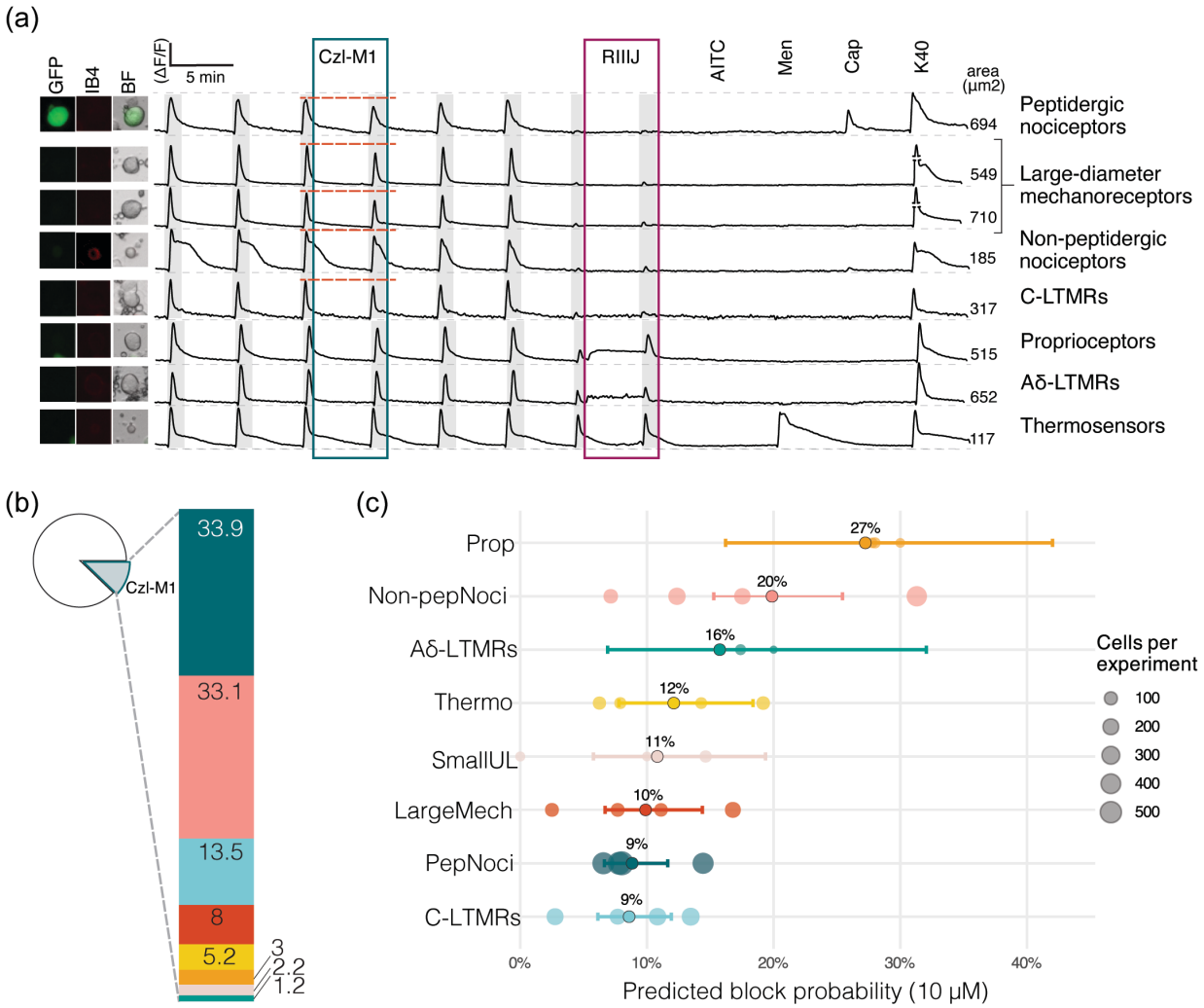

Figure 4. Effects of conkazal-M1 on mouse sensory neurons by constellation pharmacology. (**a**) Representative calcium-imaging traces of individual neurons responding with a partial block to applications of 10 μM conkazal-M1 (top 5 traces, conkazal-M1 application highlighted by blue box), and neurons that did not respond (bottom 3 traces). Light grey shading indicates KCl (30 mM, except for stimulation prior to RIIIJ treatment, in which 20 mM KCl was used) 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: conotoxin κM-RIIIJ (1 μM, burgundy box); allyl isothiocyanate (AITC, 100 μM), menthol (Men, 400 μM), and capsaicin (Cap, 300 nM). A

770 pulse of 40 mM KCl (K40) was used to assess cell viability at the end of the experiment. Cell size
771 is indicated on the right side of each trace. Aδ-LTMRs: Aδ–low-threshold mechanoreceptors;
772 CGRP: calcitonin-gene related peptide; C-LTMR: C-low threshold mechanoreceptors; GFP: green
773 fluorescent protein; N16: capsaicin-positive small-diameter neurons; RIIIJ: conotoxin *κ*M-RIIIJ.
774 (**b**) Pie chart representing the population of neurons analyzed (n = 4,215) across four independent
775 experiments, highlighting the 12% of conkazal-M1 (10 μM)-sensitive sensory neurons. Bar graph
776 showing the percentage of each class of conkazal-M1-sensitive DRG neurons. Coloration as
777 defined in Panel (c). (**c**) Estimated block probability by neuronal cell type. Each point represents
778 the model-estimated probability that neurons of a given class were blocked by the 10 μM test
779 compound, based on a generalized linear mixed-effects model (logistic link) with experiment as a
780 random intercept. Error bars denote 95% confidence intervals for the marginal means, and
781 translucent circles indicate observed block rates within individual experiments (point size ∝
782 number of cells). This analysis accounts for between-experiment variability and unequal sample
783 sizes, providing population-level estimates of block probability for each neuronal cell type. The
784 effect of cell type was highly significant ($\chi^2(8) = 276.98$, $p < 0.0001$), indicating robust differences
785 in block probability across neuronal classes. Source data for individual traces and quantifications
786 found in Supplementary File 4. Prop (orange): proprioceptors; Non-pepNoci (pink): non-
787 peptidergic nociceptors; Aδ-LTMRs (green): Aδ–low-threshold mechanoreceptors; Thermo
788 (yellow): Thermosensors; SmallUL (light pink): small-diameter unlabelled neurons; LargeMech:
789 (red) Large-diameter mechanoreceptors; PepNoci (dark teal): Peptidergic nociceptors; C-LTMR
790 (light blue): C-low threshold mechanoreceptors.

791

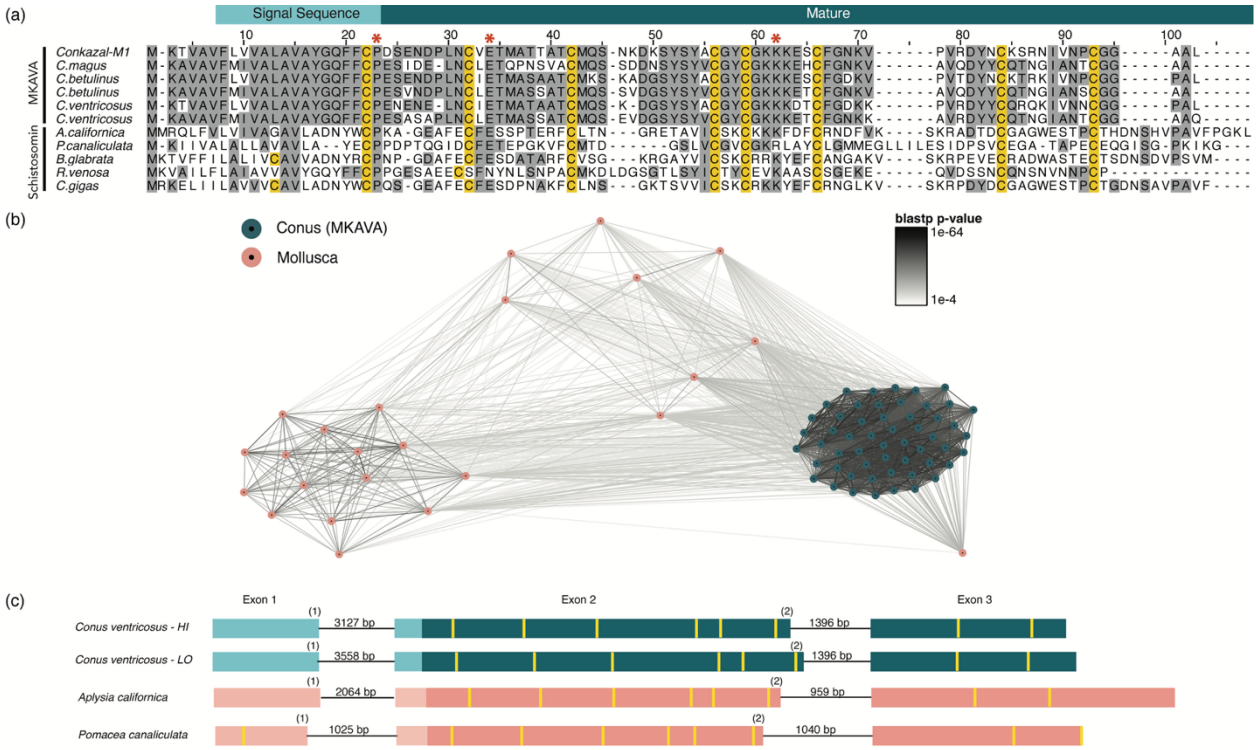

Figure 5. The MKAVAs are evolutionarily related to schistosomin. (a) Multiple sequence alignment of MKAVAs and schistosomin-like sequences from *Aplysia californica* (XP_005098364.1), *Pomacea canaliculata* (PVD28780.1), *Biomphalaria glabrata* (NP_001298209.1), *Rapana venosa* (GDIA01100366.1) and *Crassostrea gigas* (GHAT01224603.1). Red asterisks denote amino acid residues within the mature protein region exhibiting over 88% identity across all sequences analyzed (see Supplementary Figure S7). **(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. canaticulata*. 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.

806

807 **REFERENCES**

808 1. Casewell, N. R., Wuster, W., Vonk, F. J., Harrison, R. A., and Fry, B. G. (2013) Complex
809 cocktails: the evolutionary novelty of venoms *Trends Ecol Evol* **28**, 219-229

810 2. Lewis, R. J., and Garcia, M. L. (2003) Therapeutic potential of venom peptides *Nat Rev*
811 *Drug Discov* **2**, 790-802

812 3. Sachetto, A. T. A., and Mackman, N. (2019) Modulation of the mammalian coagulation
813 system by venoms and other proteins from snakes, arthropods, nematodes and insects
814 *Thromb Res* **178**, 145-154

815 4. Woodward, S. R., Cruz, L. J., Olivera, B. M., and Hillyard, D. R. (1990) Constant and
816 hypervariable regions in conotoxin propeptides *The EMBO journal* **9**, 1015-1020

817 5. Sollod, B. L., Wilson, D., Zhaxybayeva, O., Gogarten, J. P., Drinkwater, R., and King, G.
818 F. (2005) Were arachnids the first to use combinatorial peptide libraries? *Peptides* **26**, 131-
819 139

820 6. Jackson, T. N. W., and Koludarov, I. (2020) How the Toxin got its Toxicity *Front*
821 *Pharmacol* **11**, 574925

822 7. Bayrhuber, M., Vijayan, V., Ferber, M., Graf, R., Korukottu, J., Imperial, J. *et al.* (2005)
823 Conkunitzin-S1 is the first member of a new Kunitz-type neurotoxin family: structural and
824 functional characterization *J Biol Chem* **280**, 23766-23770

825 8. Finol-Urdaneta, R. K., Remedi, M. S., Raasch, W., Becker, S., Clark, R. B., Struver, N. *et*
826 *al.* (2012) Block of Kv1.7 potassium currents increases glucose-stimulated insulin
827 secretion *EMBO molecular medicine* **4**, 424-434

828 9. Droctove, L., Ciolek, J., Mendre, C., Chorfa, A., Huerta, P., Carvalho, C. *et al.* (2022) A
829 new Kunitz-type snake toxin family associated with an original mode of interaction with
830 the vasopressin 2 receptor *Br J Pharmacol* **179**, 3470-3481
831 10. Schweitz, H., Heurteaux, C., Bois, P., Moinier, D., Romey, G., and Lazdunski, M. (1994)
832 Calcicludine, a venom peptide of the Kunitz-type protease inhibitor family, is a potent
833 blocker of high-threshold Ca2+ channels with a high affinity for L-type channels in
834 cerebellar granule neurons *Proc Natl Acad Sci USA* **91**, 878-882
835 11. Mishra, M. (2020) Evolutionary Aspects of the Structural Convergence and Functional
836 Diversification of Kunitz-Domain Inhibitors *J Mol Evol* **88**, 537-548
837 12. Yuan, C. H., He, Q. Y., Peng, K., Diao, J. B., Jiang, L. P., Tang, X. *et al.* (2008) Discovery
838 of a distinct superfamily of Kunitz-type toxin (KTT) from tarantulas *PloS one* **3**, e3414
839 13. Gladkikh, I., Peigneur, S., Sintsova, O., Lopes Pinheiro-Junior, E., Klimovich, A.,
840 Menshov, A. *et al.* (2020) Kunitz-Type Peptides from the Sea Anemone Heteractis crispa
841 Demonstrate Potassium Channel Blocking and Anti-Inflammatory Activities *Biomedicines*
842 **8**
843 14. Peigneur, S., Billen, B., Derua, R., Waelkens, E., Debaveye, S., Beress, L. *et al.* (2011) A
844 bifunctional sea anemone peptide with Kunitz type protease and potassium channel
845 inhibiting properties *Biochem Pharmacol* **82**, 81-90
846 15. Zupunski, V., and Kordis, D. (2016) Strong and widespread action of site-specific positive
847 selection in the snake venom Kunitz/BPTI protein family *Sci Rep* **6**, 37054
848 16. Zdenek, C. N., Cardoso, F. C., Robinson, S. D., Mercedes, R. S., Raidjoe, E. R., Hernandez-
849 Vargas, M. J. *et al.* (2024) Venom exaptation and adaptation during the trophic switch to
850 blood-feeding by kissing bugs *iScience* **27**, 110723

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

873 25. Koch, T. L., Robinson, S. D., Salcedo, P. F., Chase, K., Biggs, J., Fedosov, A. E. *et al.*
874 (2024) Prey Shifts Drive Venom Evolution in Cone Snails *Mol Biol Evol* **41**,
875 26. Robinson, S. D., Safavi-Hemami, H., McIntosh, L. D., Purcell, A. W., Norton, R. S., and
876 Papenfuss, A. T. (2014) Diversity of conotoxin gene superfamilies in the venomous snail,
877 Conus victoriae *PloS one* **9**, e87648
878 27. Hackney, C. M., Florez Salcedo, P., Müller, E., Koch, T. L., Kjelgaard, L. D., Watkins, M.
879 *et al.* (2023) A previously unrecognized superfamily of macro-conotoxins includes an
880 inhibitor of the sensory neuron calcium channel Cav2.3 *PLoS Biol* **21**, e3002217
881 28. Yang, Z. (1997) PAML: a program package for phylogenetic analysis by maximum
882 likelihood *Comput Appl Biosci* **13**, 555-556
883 29. Bertelsen, A. B., Hackney, C. M., Bayer, C. N., Kjelgaard, L. D., Rennig, M., Christensen,
884 B. *et al.* (2021) DisCoTune: versatile auxiliary plasmids for the production of disulphide-
885 containing proteins and peptides in the E. coli T7 system *Microb Biotechnol* 10.1111/1751-
886 7915.13895
887 30. Gaciarz, A., Khatri, N. K., Velez-Suberbie, M. L., Saaranen, M. J., Uchida, Y., Keshavarz-
888 Moore, E. *et al.* (2017) Efficient soluble expression of disulfide bonded proteins in the
889 cytoplasm of Escherichia coli in fed-batch fermentations on chemically defined minimal
890 media *Microb Cell Fact* **16**, 108
891 31. Nielsen, L. D., Foged, M. M., Albert, A., Bertelsen, A. B., Soltoft, C. L., Robinson, S. D.
892 *et al.* (2019) The three-dimensional structure of an H-superfamily conotoxin reveals a
893 granulin fold arising from a common ICK cysteine framework *J Biol Chem* **294**, 8745-
894 8759

895 32. Safavi-Hemami, H., Li, Q., Jackson, R. L., Song, A. S., Boomsma, W., Bandyopadhyay, P.
896 K. *et al.* (2016) Rapid expansion of the protein disulfide isomerase gene family facilitates
897 the folding of venom peptides *Proc Natl Acad Sci USA* **113**, 3227-3232
898 33. Rogov, V. V., Rozenknop, A., Rogova, N. Y., Lohr, F., Tikole, S., Jaravine, V. *et al.* (2012)
899 A universal expression tag for structural and functional studies of proteins *Chembiochem*
900 **13**, 959-963
901 34. Buczek, O., Bulaj, G., and Olivera, B. M. (2005) Conotoxins and the posttranslational
902 modification of secreted gene products *Cell Mol Life Sci* **62**, 3067-3079
903 35. Fischer, W. H., and Spiess, J. (1987) Identification of a mammalian glutaminyl cyclase
904 converting glutaminyl into pyroglutamyl peptides *Proc Natl Acad Sci USA* **84**, 3628-3632
905 36. van Kempen, M., Kim, S. S., Tumescheit, C., Mirdita, M., Lee, J., Gilchrist, C. L. M. *et al.*
906 (2024) Fast and accurate protein structure search with Foldseek *Nat Biotechnol* **42**, 243-
907 246
908 37. Krissinel, E., and Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for
909 fast protein structure alignment in three dimensions *Acta Crystallogr D Biol Crystallogr*
910 **60**, 2256-2268
911 38. Derache, C., Epinette, C., Roussel, A., Gabant, G., Cadene, M., Korkmaz, B. *et al.* (2012)
912 Crystal structure of greglin, a novel non-classical Kazal inhibitor, in complex with
913 subtilisin *FEBS J* **279**, 4466-4478
914 39. Perrakis, A., and Sixma, T. K. (2021) AI revolutions in biology: The joys and perils of
915 AlphaFold *EMBO Rep* **22**, e54046

916 40. Shenoy, R. T., Thangamani, S., Velazquez-Campoy, A., Ho, B., Ding, J. L., and Sivaraman,
917 J. (2011) Structural basis for dual-inhibition mechanism of a non-classical Kazal-type
918 serine protease inhibitor from horseshoe crab in complex with subtilisin *PloS one* **6**, e18838
919 41. Hordijk, P. L., de Jong-Brink, M., ter Maat, A., Pieneman, A. W., Lodder, J. C., and Kits,
920 K. S. (1992) The neuropeptide schistosomin and haemolymph from parasitized snails
921 induce similar changes in excitability in neuroendocrine cells controlling reproduction and
922 growth in a freshwater snail *Neurosci Lett* **136**, 193-197
923 42. Frickey, T., and Lupas, A. (2004) CLANS: a Java application for visualizing protein
924 families based on pairwise similarity *Bioinformatics* **20**, 3702-3704
925 43. Tian, M., Benedetti, B., and Kamoun, S. (2005) A Second Kazal-like protease inhibitor
926 from Phytophthora infestans inhibits and interacts with the apoplastic pathogenesis-related
927 protease P69B of tomato *Plant Physiol* **138**, 1785-1793
928 44. Tian, M., and Kamoun, S. (2005) A two disulfide bridge Kazal domain from Phytophthora
929 exhibits stable inhibitory activity against serine proteases of the subtilisin family *BMC*
930 *Biochem* **6**, 15
931 45. Li, X. C., Wang, X. W., Wang, Z. H., Zhao, X. F., and Wang, J. X. (2009) A three-domain
932 Kazal-type serine proteinase inhibitor exhibiting domain inhibitory and bacteriostatic
933 activities from freshwater crayfish Procambarus clarkii *Dev Comp Immunol* **33**, 1229-1238
934 46. Esmaeilishirazifard, E., Usher, L., Trim, C., Denise, H., Sangal, V., Tyson, G. H. *et al.*
935 (2022) Bacterial Adaptation to Venom in Snakes and Arachnida *Microbiol Spectr* **10**,
936 e0240821
937 47. De Jong-Brink, M. (1995) How schistosomes profit from the stress responses they elicit in
938 their hosts *Adv Parasitol* **35**, 177-256

939 48. Zhang, S. M., Nian, H., Wang, B., Loker, E. S., and Adema, C. M. (2009) Schistosomin
940 from the snail Biomphalaria glabrata: expression studies suggest no involvement in
941 trematode-mediated castration *Mol Biochem Parasitol* **165**, 79-86
942 49. Force, A., Lynch, M., Pickett, F. B., Amores, A., Yan, Y. L., and Postlethwait, J. (1999)
943 Preservation of duplicate genes by complementary, degenerative mutations *Genetics* **151**,
944 1531-1545
945 50. van de Locht, A., Lamba, D., Bauer, M., Huber, R., Friedrich, T., Kroger, B. *et al.* (1995)
946 Two heads are better than one: crystal structure of the insect derived double domain Kazal
947 inhibitor rhodniin in complex with thrombin *The EMBO journal* **14**, 5149-5157
948 51. Maxwell, M., Undheim, E. A. B., and Mobli, M. (2018) Secreted Cysteine-Rich Repeat
949 Proteins "SCREPs": A Novel Multi-Domain Architecture *Front Pharmacol* **9**, 1333
950 52. Liu, J., Maxwell, M., Cuddihy, T., Crawford, T., Bassetti, M., Hyde, C. *et al.* (2023)
951 ScrepYard: An online resource for disulfide-stabilized tandem repeat peptides *Protein Sci*
952 **32**, e4566
953 53. Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002) MAFFT: a novel method for rapid
954 multiple sequence alignment based on fast Fourier transform *Nucleic Acids Res* **30**, 3059-
955 3066
956 54. Katoh, K., and Standley, D. M. (2013) MAFFT multiple sequence alignment software
957 version 7: improvements in performance and usability *Mol Biol Evol* **30**, 772-780
958 55. Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M., and Barton, G. J. (2009)
959 Jalview Version 2--a multiple sequence alignment editor and analysis workbench
960 *Bioinformatics* **25**, 1189-1191

961 56. Slater, G. S., and Birney, E. (2005) Automated generation of heuristics for biological
962 sequence comparison *BMC Bioinformatics* **6**, 31
963 57. Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von
964 Haeseler, A. *et al.* (2020) IQ-TREE 2: New Models and Efficient Methods for Phylogenetic
965 Inference in the Genomic Era *Mol Biol Evol* **37**, 1530-1534
966 58. Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., and Vinh, L. S. (2018)
967 UFBoot2: Improving the Ultrafast Bootstrap Approximation *Mol Biol Evol* **35**, 518-522
968 59. Rambaut, A. (2018) FigTree (Version 1.4.4) Institute of Evolutionary Biology, University
969 of Edinburgh,
970 60. Khilji, M. S., Hackney, C. M., Koch, T. L., Hone, A. J., Rogalski, A., Watkins, M. *et al.*
971 (2025) Structural similarities reveal an expansive conotoxin family with a two-finger toxin
972 fold *Protein Sci* **34**, e70333
973 61. Wilkins, M. R., Gasteiger, E., Bairoch, A., Sanchez, J. C., Williams, K. L., Appel, R. D. *et*
974 *al.* (1999) Protein identification and analysis tools in the ExPASy server *Methods Mol Biol*
975 **112**, 531-552
976 62. Kazimierczuk, K., and Orekhov, V. Y. (2011) Accelerated NMR spectroscopy by using
977 compressed sensing *Angew Chem Int Ed Engl* **50**, 5556-5559
978 63. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe:
979 a multidimensional spectral processing system based on UNIX pipes *J Biomol NMR* **6**,
980 277-293
981 64. Skinner, S. P., Fogh, R. H., Boucher, W., Ragan, T. J., Mureddu, L. G., and Vuister, G. W.
982 (2016) CcpNmr AnalysisAssign: a flexible platform for integrated NMR analysis *J Biomol*
983 *NMR* **66**, 111-124

984 65. Güntert, P., and Buchner, L. (2015) Combined automated NOE assignment and structure
985 calculation with CYANA *J Biomol NMR* **62**, 453-471
986 66. Schwieters, C. D., Kuszewski, J. J., Tjandra, N., and Clore, G. M. (2003) The Xplor-NIH
987 NMR molecular structure determination package *J Magn Reson* **160**, 65-73
988 67. Bermejo, G. A., Clore, G. M., and Schwieters, C. D. (2012) Smooth statistical torsion angle
989 potential derived from a large conformational database via adaptive kernel density
990 estimation improves the quality of NMR protein structures *Protein Sci* **21**, 1824-1836
991 68. Tian, Y., Schwieters, C. D., Opella, S. J., and Marassi, F. M. (2017) High quality NMR
992 structures: a new force field with implicit water and membrane solvation for Xplor-NIH *J*
993 *Biomol NMR* **67**, 35-49
994 69. Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kaptein, R., and Thornton, J. M.
995 (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein
996 structures solved by NMR *J Biomol NMR* **8**, 477-486
997 70. Olp, M. D., Kalous, K. S., and Smith, B. C. (2020) ICEKAT: an interactive online tool for
998 calculating initial rates from continuous enzyme kinetic traces *BMC Bioinformatics* **21**,
999 186
1000 71. Teichert, R. W., Smith, N. J., Raghuraman, S., Yoshikami, D., Light, A. R., and Olivera, B.
1001 M. (2012) Functional profiling of neurons through cellular neuropharmacology *Proc Natl*
1002 *Acad Sci USA* **109**, 1388-1395

1003