Rapid expansion of the protein disulfide isomerase gene family facilitates the folding of venom peptides

Helena Safavi-Hemami[a,b,1], Qing Li[c], Ronneshia L. Jackson[a], Albert S. Song[a], Wouter Boomsma[b], Pradip K. Bandyopadhyay[a], Christian W. Gruber[d,e], Anthony W. Purcell[f,g], Mark Yandell[c,h], Baldomero M. Olivera[a,1], and Lars Ellgaard[b]

[a]Department of Biology, University of Utah, Salt Lake City, UT 84112; [b]Department of Biology, University of Copenhagen, DK-2200 Copenhagen, Denmark; [c]Eccles institute of Human Genetics, University of Utah, Salt Lake City, UT 84112; [d]Center for Physiology and Pharmacology, Medical University of Vienna, A-1090, Vienna, Austria; [e]School of Biomedical Sciences, University of Queensland, St. Lucia, Queensland 4072, Australia; [f]Infection and Immunity Program, Monash Biomedicine Discovery Institute, Monash University, Clayton, Victoria 3800, Australia; [g]Department of Biochemistry and Molecular Biology, Monash University, Clayton, Victoria 3800, Australia; and [h]Utah Science Technology and Research Center for Genetic Discovery, University of Utah, Salt Lake City, UT 84112

Contributed by Baldomero M. Olivera, February 17, 2016 (sent for review November 30, 2015; reviewed by Johannes M. Herrmann and Maido Remm)

Formation of correct disulfide bonds in the endoplasmic reticulum is a crucial step for folding proteins destined for secretion. Protein disulfide isomerases (PDIs) play a central role in this process. We report a previously unidentified, hypervariable family of PDIs that represents the most diverse gene family of oxidoreductases described in a single genus to date. These enzymes are highly expressed specifically in the venom glands of predatory cone snails, animals that synthesize a remarkably diverse set of cysteine-rich peptide toxins (conotoxins). Enzymes in this PDI family, termed conotoxin-specific PDIs, significantly and differentially accelerate the kinetics of disulfide-bond formation of several conotoxins. Our results are consistent with a unique biological scenario associated with protein folding: The diversification of a family of foldases can be correlated with the rapid evolution of an unprecedented diversity of disulfide-rich structural domains expressed by venomous marine snails in the superfamily Conoidea.

protein disulfide isomerase | peptide folding | gene expansion | cone snail venom | conotoxins

Formation of correct disulfide bonds is essential for the structural stability and functional integrity of many secreted proteins and peptides, such as protease inhibitors, hormones, antimicrobial peptides, and toxins from venoms. Recent advances in nucleotide and protein sequencing have revealed that toxins from marine snails in the superfamily Conoidea, such as cone snails (*Conus*), comprise a remarkable diversity of cysteine-rich polypeptides (1, 2). Toxin expression and folding takes place in the endoplasmic reticulum (ER) of venom gland cells, where, at any given time, hundreds of distinct cysteine-rich peptides are properly folded and secreted in preparation for envenomation (3, 4). No other tissue type has been reported to produce such a high density and diversity of cysteine-rich peptides. Although a fraction of these peptides contain domains that are widely distributed in the animal and plant kingdom [e.g., the inhibitor cysteine knot (5), kunitz-type domains (6), and the insulin/relaxin-like fold (7)], the majority represent unique structural domains, expressed only in conoidean venoms. How these structural scaffolds are efficiently folded in the venom gland is not well understood, but it is clear that ER-resident helper proteins must be involved. For example, members of the well-characterized O superfamily of conotoxins contain six cysteine residues that can form three disulfide bonds. Despite the fact that these peptides could potentially adopt 15 different disulfide-bonded scaffolds, only one native fold is commonly found in cone snail venom (8). Conversely, in vitro folding of these toxins commonly results in low folding yields, as well as accumulation of misfolded or aggregated products (8), highlighting the need for a better understanding of the molecular processes guiding in vivo disulfide-bond formation.

Several common ER-resident foldases—including peptidyl prolyl *cis–trans* isomerase (PPI) (9), and protein disulfide isomerase (PDI) (10)—have been shown to assist in the oxidative folding of conotoxins. Whether specialized adaptations in the venom gland oxidative folding machinery have evolved to enable the folding of such a remarkably diverse set of cysteine-rich peptides has not been addressed.

Here, a systematic interrogation of 17 cone snail venom gland transcriptomes led to the identification and subsequent characterization of a large, previously undescribed PDI gene family that likely plays a critical role in the folding of conotoxins. Comparative sequence analysis revealed that this gene family arose by gene duplication and positive selection, complementing the rapid evolution of conotoxin-encoding genes. Thus, the evolution of the conotoxin-specific PDI (csPDI) family can be regarded as a key adaptation for the high-throughput production of cysteine-rich venom peptides.

Results

New PDI Sequence from *Conus geographus* Defines the First Member of a Diverse Gene Family. Analyses of the published venom gland transcriptome of *Conus geographus* (1) identified a sequence resembling other known cone snail PDIs (e.g., ∼96% identity to PDI from *Conus marmoreus*), but also revealed the presence of an additional related sequence sharing only ∼67% identity to

Significance

The majority of secreted proteins contain disulfide bonds that provide structural stability in the extracellular environment. The formation of correct disulfide bonds is assisted by the enzyme protein disulfide isomerase (PDI). Most secreted structural domains are ancient and widely distributed in all metazoans; in contrast, diverse sets of unique disulfide-rich structural domains have more recently evolved in venomous marine snails (superfamily Conoidea comprising >10,000 species). We have discovered a previously undescribed gene family encoding PDIs of unprecedented diversity. We suggest that these enzymes constitute an important part of the supporting molecular infrastructure required for properly folding the plethora of structural domains expressed in the venoms of snails in different conoidean lineages.

Author contributions: H.S.-H., A.W.P., B.M.O., and L.E. designed research; H.S.-H., R.L.J., A.S.S., and P.K.B. performed research; C.W.G., A.W.P., and M.Y. contributed new reagents/analytic tools; H.S.-H., Q.L., and W.B. analyzed data; and H.S.-H., B.M.O., and L.E. wrote the paper.

Reviewers: J.M.H., University of Kaiserslautern; and M.R., University of Tartu.

The authors declare no conflict of interest.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. KT874559–KT874574).

[1]To whom correspondence may be addressed. Email: helena.safavi@utah.edu or olivera@biology.utah.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1525790113/-/DCSupplemental.

PDI from other cone snail species (Fig. 1). This previously undescribed sequence represented, to our knowledge, the first member of the csPDI gene family. Like canonical PDI, csPDI consists of four thioredoxin-like domains: an **a** and **a′** domain containing the active site CGHC motif and noncatalytic **b** and **b′** domain. Further transcriptome mining identified two variants of csPDI in the transcriptome dataset. Reverse-transcription PCR, cloning, and Sanger sequencing of *C. geographus* venom gland cDNA confirmed these variants and led to the identification of two additional csPDI sequences, previously undetected in the RNA sequencing (RNA-Seq) dataset.

Thus, a total of five distinct csPDI sequences that share 87–97% identity with each other and 61–65% identity with canonical PDI were retrieved (*SI Appendix*, Table S1). Notably, several of the variable amino acid residues were located between the conserved cysteine residues in the active-site motif (CXXC). These amino acids are known to affect the redox state of PDI and therefore the ability of the enzyme to form, reduce, and isomerize disulfide bonds in client proteins (11). Although canonical PDI in other organisms, including *C. geographus*, contains a glycine followed by histidine (CGHC), *C. geographus* csPDIs have a diverse set of residues in both catalytic domains: CGAC and CDAC in the **a** domain and CGLC and CEFC in the **a′** domain (*SI Appendix*, Fig. S1). How these changes may affect the oxidoreductase activity of these enzymes, especially in respect to conotoxin folding, is discussed further below.

csPDIs Are Hypervariable and Ubiquitously Expressed in the Venom Gland of Cone Snails. To determine whether csPDIs are expressed in the venom glands of other cone snail species, the venom gland transcriptomes of 15 additional cone snails were obtained. The published transcriptome of *Conus victoriae* (2) was also examined, and a close relative, *Conus textile*, was examined by RT-PCR (see *SI Appendix*, Table S2, for all species used in this study). All species examined were found to express csPDIs, demonstrating the importance and wide distribution of this protein family in the genus *Conus*. Sequences shared between 76.2% and 98.2% identify, with no exact matches even between very closely related species (e.g., *Conus praecellens* and *Conus andremenezi*).

Unlike the sequence diversity observed for *C. geographus* csPDIs, only one sequence per species was retrieved from assembled datasets, suggesting that *C. geographus* csPDIs were exceptionally diverse or that the true diversity of csPDI sequences was being missed using standard RNA-Seq assembly protocols. To better investigate the diversity of csPDIs, we applied a recently developed tool for next-generation sequencing read classification called Taxonomer (ref. 12; see *SI Appendix*, *SI Materials and Methods* for more details). Taxonomer specifically identified all RNA-Seq reads derived from the csPDI gene family before data assembly, thus enabling faster and more accurate assemblies of highly similar sequences. Taxonomer identified an average of 2.6 csPDI sequences per species, confirming the expansion of the csPDI family in cone snails. In total, 43 unique full-length and 4 partial csPDI sequences were identified from 18 species. Applying the same methodology, only one canonical PDI was identified per species with the exception of *Conus textile*, which expressed two distinct variants of PDI, a finding that has been reported (13).

Phylogenetic analysis clearly resolved the PDI and csPDI gene families and revealed that these enzymes have evolved by duplication from an ancestral gene (Fig. 2, black arrow). csPDI sequences resolved into two groups that correlate with the "primitive" and "complex" group of cone snails, previously described based on mitochondrial phylogenetics (14). However, within these groups, csPDI sequences from the same species do not group together,

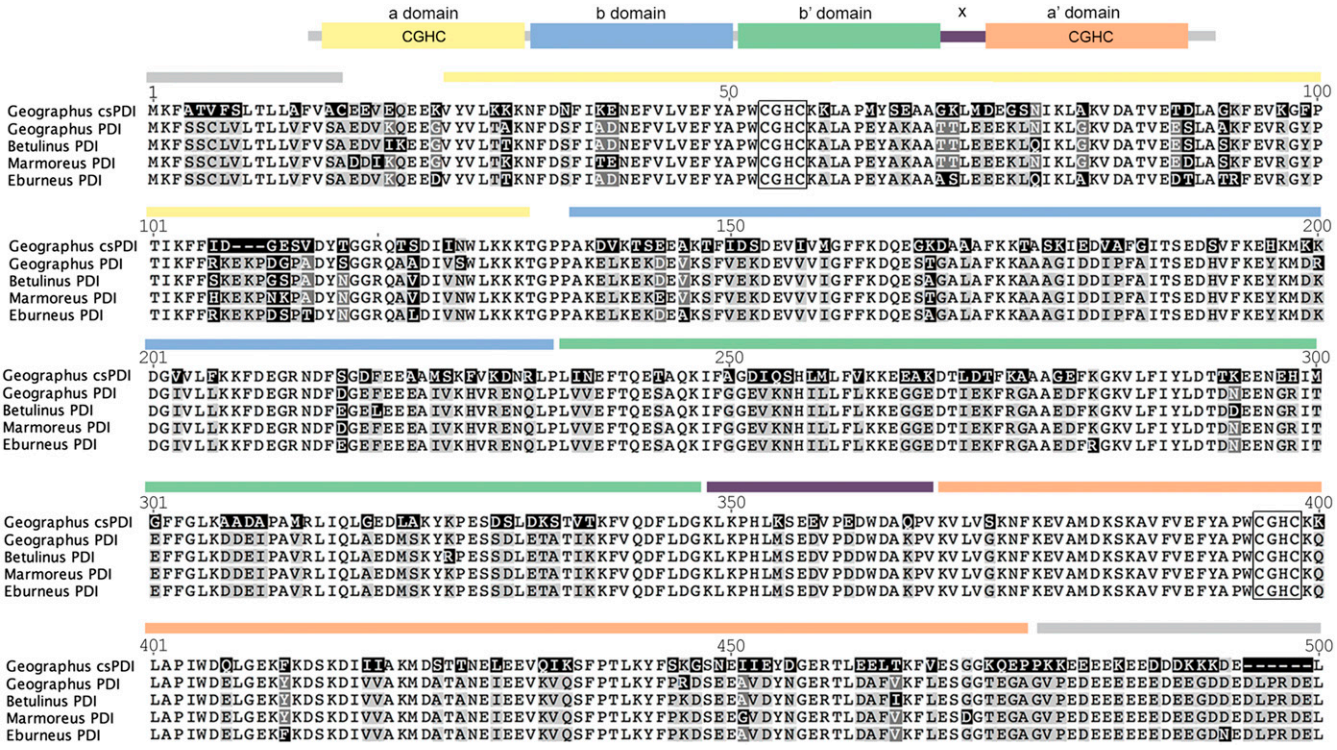


Fig. 1. Identification of a previously unidentified PDI sequence (csPDI) in the venom gland of *C. geographus* (GenBank accession no. KT874567). Multiple sequence alignment with canonical PDI from the same species (GenBank accession no. KT874559) and three additional species [*C. betulinus* (ADZ76593), *C. marmoreus* (ABF48564), and *C. eburneus* (ADZ76591)] identifies regions of divergence between csPDI and PDIs (white, 100% identity; light gray, 99–80% identity; dark gray, 80–60% identity; black: <60% identity). The alignment was performed in Geneious by using the Blosum62 similarity option for coloring (Version 8.1.3). csPDI and PDI sequences share 65–66% identity. Thioredoxin domain organization is depicted above the sequences and was predicted by using known boundaries for human PDI (31). Signal sequences (gray bar above sequence) were predicted by using InterProScan (32). The C terminus containing ER-retention motifs is also depicted with a gray bar. Active site CGHC motifs are boxed.

suggesting that the evolution of csPDIs is more closely tied to molecular function than phylogenetic relationships.

Similar to observations made for *C. geographus*, several csPDI sequences contained unusual variations in the active-site motifs. The most prominent difference was the replacement of a histidine for alanine in the **a** domain, a motif that was found in 13 of 18 species. Phylogenetic analysis suggested that this mutation evolved several times within the csPDI family (Fig. 2). Additional variations included CGIC in the **a** domain and CAHC and CEKC in the **a′** domain. Two partial sequences retrieved from the *C. victoriae* transcriptome (2) contained CEFC and CRPC variations in the **a′** domain. Hereafter, the two amino acids located between the active-site cysteines will be provided as superscript letters—e.g., $csPDI_{GA/GH}$, where the first two letters (GA) represent residues found in the **a** domain and the last two (GH) residues in the **a′** domain of the enzyme. BLAST searches could not retrieve a gene resembling the csPDI family from any other organism in the NCBI nonredundant protein and nucleotide collection, suggesting that the csPDI family specifically evolved in the genus *Conus*. Several other members of the PDI family were identified in transcriptome datasets (e.g., PDIA3 and PDIA5). Comparative phylogenetic analysis of these and *Conus* PDIs and csPDIs illustrated that the csPDIs are more closely related to PDI than to other PDI family members (*SI Appendix*, Fig. S2).

PDI and csPDIs Are Highly Expressed in the *Conus* Venom Gland and Among the Most Abundant Proteins. Analysis of transcriptomic data highlighted that PDI and csPDIs are highly expressed in the venom glands of all cone snail species, ranging from 0.03% to 0.27% of all sequenced reads. Expression ratios for csPDI/PDI ranged from 0.4 to 2.2 (mean ratio: 1.2), demonstrating that the csPDI family has similar expression values to PDI (*SI Appendix*, Fig. S3). Furthermore, csPDIs are preferentially expressed in the venom gland with very low expression levels in other tissues, as determined by quantitative real-time PCR (qPCR) and RNA-Seq analysis on the foot, venom bulb (a venom "pump" located at the inner end of the venom gland), esophagus, nerve ring, salivary gland, and venom gland of *C. geographus* and *C. rolani*, supporting a specialized role of the csPDI family in conotoxin folding (*SI Appendix*, Fig. S4; the esophagus and nerve ring were not available for *C. rolani*).

To investigate relative abundances of these proteins in the venom gland, the published proteome of *C. geographus* was revisited (15). Separation of venom gland proteins by 2D gel electrophoresis resolved two distinct gel areas that were identified as different isoforms of PDI in the original study (15). Reanalysis of mass spectrometric data by using a proteogenomic strategy revealed that these areas correspond to PDI and csPDIs (Fig. 3). Migration patterns are consistent with differences in the predicted isoelectric point (pI) for PDI (pI: 4.6) and members of the csPDI family (pI: 4.8–5.0).

Although gel analysis could not resolve individual csPDI members, matching of tryptic peptides obtained by mass spectrometry to csPDI sequences identified between one and nine unique peptides corresponding to each csPDI sequence (Fig. 3 and *SI Appendix*, Fig. S5). This finding strongly suggests that all *C. geographus* csPDI variants are translated into functional proteins.

The intensity of gel spots identified as PDI and csPDIs highlights that these enzymes are clearly among the most abundant soluble proteins present in the venom gland of *C. geographus* (Fig. 3).

csPDI Family Is Rapidly Evolving with High Sequence Variability at Functionally Important Sites. Several distinct csPDI sequences were identified for each cone snail species. This diversification suggests that the csPDI family is expanding and subject to strong positive selection. Additionally, comparative sequence alignments and phylogenetic analyses suggested that the genetic variability was higher for members of the csPDI gene family than for PDI-encoding genes. Evolutionary pressures can be quantified by the rates of substitutions at silent sites (dS), which are presumed neutral, relative to the rate of substitutions at nonsilent sites (dN), which possibly experience selection. To investigate whether the csPDI family contained sites that experience high positive selection rates, the mixed effects model of

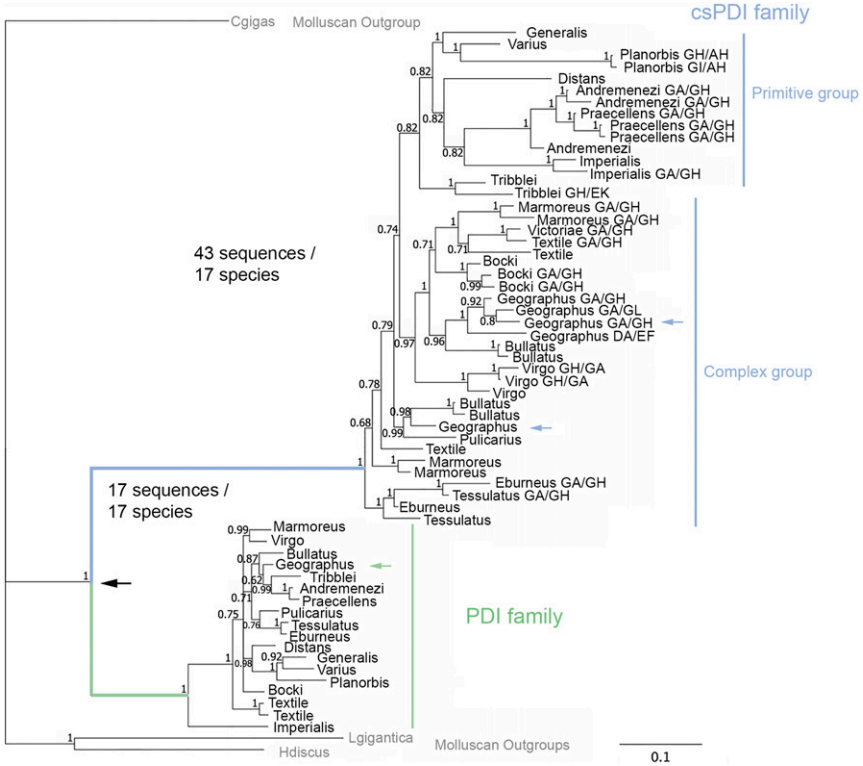
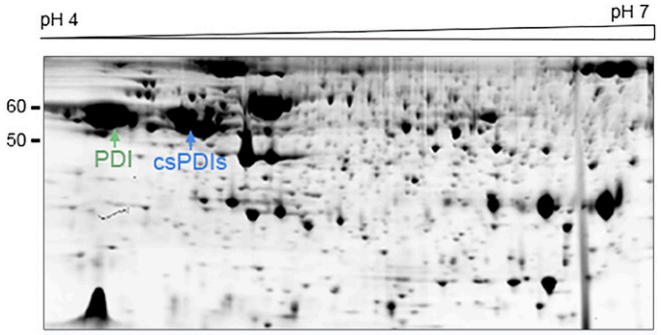


Fig. 2. Phylogenetic analysis of full-length PDI and csPDI protein sequences supports the presence of two gene families originating from an ancestral gene duplication event (black arrow, posterior probability: 1). Diversity and genetic variance for the csPDI family are apparent by more than one distinct csPDI sequence per species and long branch lengths, respectively. Two groups (the primitive and complex groups) resolved within the csPDI branch and correlate with molecular phylogenetics analyses (14). Amino acids of the active site motif are provided for sequences with divergent active sites. Arrows indicate *C. geographus* sequences selected for subsequent functional characterization. Bayesian tree was constructed using MrBayes (Version 3.2.2; ref. 33) with two runs each of four Markov chains sampling every 200 generations. The log-likelihood score stabilized after 1,100,000 generations. The consensus tree was calculated after omitting the first 25% of the samples as burn-in.



	MW	pI	Peptide Matches	Unique Matches
PDI	54034	4.6	26	17
csPDI GH/GH	53869	5.0	30	9
csPDI GA/GH	54023	4.9	23	2
csPDI GA/GH 2	53778	4.9	31	5
csPDI GA/GL	53894	4.8	21	2
csPDI DA/EF	54123	4.9	25	1

Fig. 3. Analysis of the venom gland proteome of *C. geographus* shows high abundances for PDI and csPDIs as determined by 2D gel electrophoresis and subsequent mass spectrometric protein identification. Data deposited in the original study (15) were revisited and examined for mass spectrometric peptide hits that matched PDI and csPDI sequences obtained in the present study. Gel spots identified as PDI and csPDIs are depicted and correlate with predicted molecular weights (MW) and isoelectric points (pI). The number of total and unique peptide matches obtained for PDI and different members of the csPDI family are provided (score > 99 using Protein Pilot; Version 3.0; AB SCIEX). Sequences and position of matched peptides onto the full-length sequences are provided in *SI Appendix*, Fig. S5. Reproduced from ref. 15, copyright the American Society for Biochemistry and Molecular Biology.

evolution (MEME) implemented in datamonkey (16) was used. MEME analysis revealed a total of 12 and 35 sites ($P < 0.1$), with positive selection for PDI and csPDI families, respectively, demonstrating that the csPDI family experiences higher selection rates than PDI. Interestingly, for the PDI family, episodes of positive selection were absent in the **b′** domain, a domain known to be important for substrate binding (17), whereas the csPDI family displayed 13 positive selection events in this domain (*SI Appendix*, Fig. S6). To graphically illustrate protein sequence variation for PDIs and csPDIs, a sequence variation score was generated based on multiple sequence alignments for the two enzyme families (see *SI Appendix*: a sequence logo representation for the two multiple sequence alignments compared with the sequence of human PDI is provided in *SI Appendix*, Fig. S7). This score was subsequently converted into a red-white color range, where darker color represents more sequence variation, and mapped onto the crystal structure of full-length human PDI (18) (Fig. 4 *A* and *B*). Modeling revealed that PDIs and csPDIs show widely distributed sequence variations in the **a** and **b** domains. Notably, the PDIs show only very moderate variation in the **b′** and **a′** domains, compared with csPDIs. This difference is most pronounced for the **b′** domain, where by far most residues are strictly conserved among PDIs, but vary considerably in the csPDIs (Fig. 4*C*). This finding is consistent with the MEME analysis as discussed above. In human PDI, the **b′** domain harbors a hydrophobic patch known to bind substrates directly (17). In addition, we noticed that two equivalent positions in the **a** and **a′** domains, located two residues C-terminal of the second cysteine of the CXXC active-site motif, show high sequence variation that is restricted to the csPDIs (arrows in Fig. 4*B*). Overall, csPDIs show higher sequence variation compared with the PDIs on key positions implicated in substrate binding and found at or in close proximity to the active site of both redox active domains.

***C. geographus* csPDIs Have Distinct Effects on the Folding Kinetics of Conotoxin Substrates.** To determine whether csPDIs can assist in the folding of conotoxins, two csPDI family members from *C. geographus* containing the two most widely distributed active site motifs ($csPDI_{GH/GH}$ and $csPDI_{GA/GH}$) were expressed for oxidative folding studies (*SI Appendix*, Fig. S8). PDI was analyzed for comparison. Three O-superfamily conotoxins containing six cysteine residues were selected for oxidative folding studies based on their distinct folding characteristics (Fig. 5). Omega-GVIA, from the venom of *C. geographus*, folds rapidly with very little misfolded byproducts (19). PDI and both csPDIs significantly accelerated the folding of ω-GVIA compared with no-enzyme control reactions (half-time $t_{1/2}$ for accumulation of native product: 42.7 min). Folding was fastest in the presence of $csPDI_{GH/GH}$ (7.1 min) followed by $csPDI_{GA/GH}$ (8.7 min) and PDI (19.4 min) (Fig. 5*A*). The folding of μ-SmIIIA, a peptide with faster folding kinetics than ω-GVIA (20), was significantly accelerated only in the presence of $csPDI_{GH/GH}$ and $csPDI_{GA/GH}$ (Fig. 5*B*). PDI had no significant effect. Most remarkably, when the recombinant enzymes were tested on δ-PVIA, a member of the delta conotoxin family that is characterized by very slow in vitro folding kinetics (21), folding was accelerated by a factor of ~32 in the presence of $csPDI_{GH/GH}$ ($t_{1/2}$ = 9.7 min) compared with no-enzyme controls ($t_{1/2}$ = 315.8 min; Fig. 5*C*). PDI was slightly less efficient ($t_{1/2}$ = 20.6 min) followed by $csPDI_{GA/GH}$

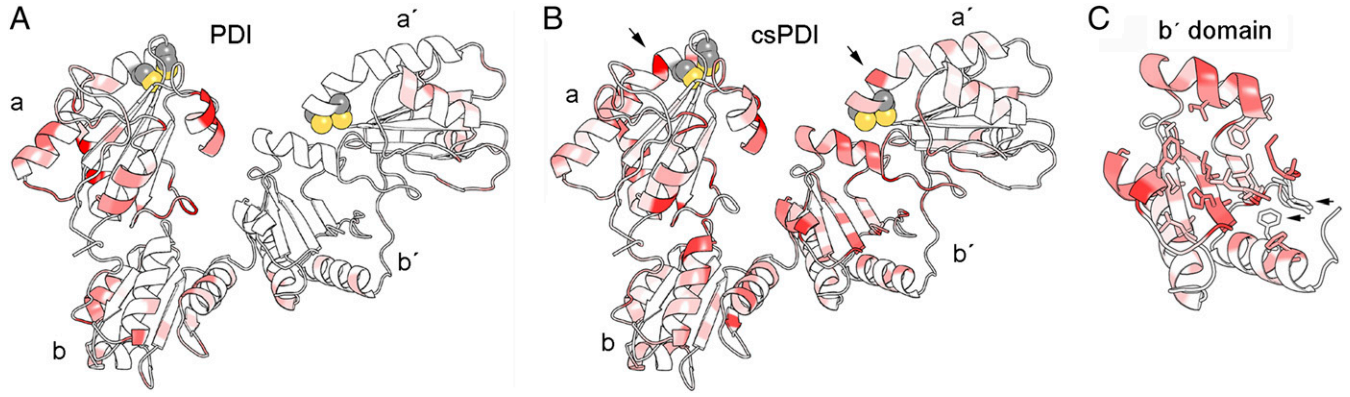


Fig. 4. (*A* and *B*) Sequence variation in cone snail PDIs (*A*) and csPDIs (*B*) mapped onto a representation of the crystal structure of full-length human PDI (Protein Data Bank ID code 4EKZ). Multiple sequence alignments of PDIs and csPDIs were used to assign a variation score for each position in the alignment. This score was then converted to a red–white color range, where darker color indicates higher sequence variation. Heavy atoms of active-site cysteines are depicted as space-filling models (gray, C; yellow, S). The four thioredoxin-like domains are indicated, and arrows point to the +2 position C-terminal of the CXXC motifs in the **a** and **a′** domains of the csPDIs, which shows sequence variation only in this group of enzymes and not in the PDIs (see text for details). (*C*) Residues of the hydrophobic patch of the **b′** domain are shown as stick models. Sequence variation is apparent in all but two of these residues (arrows).

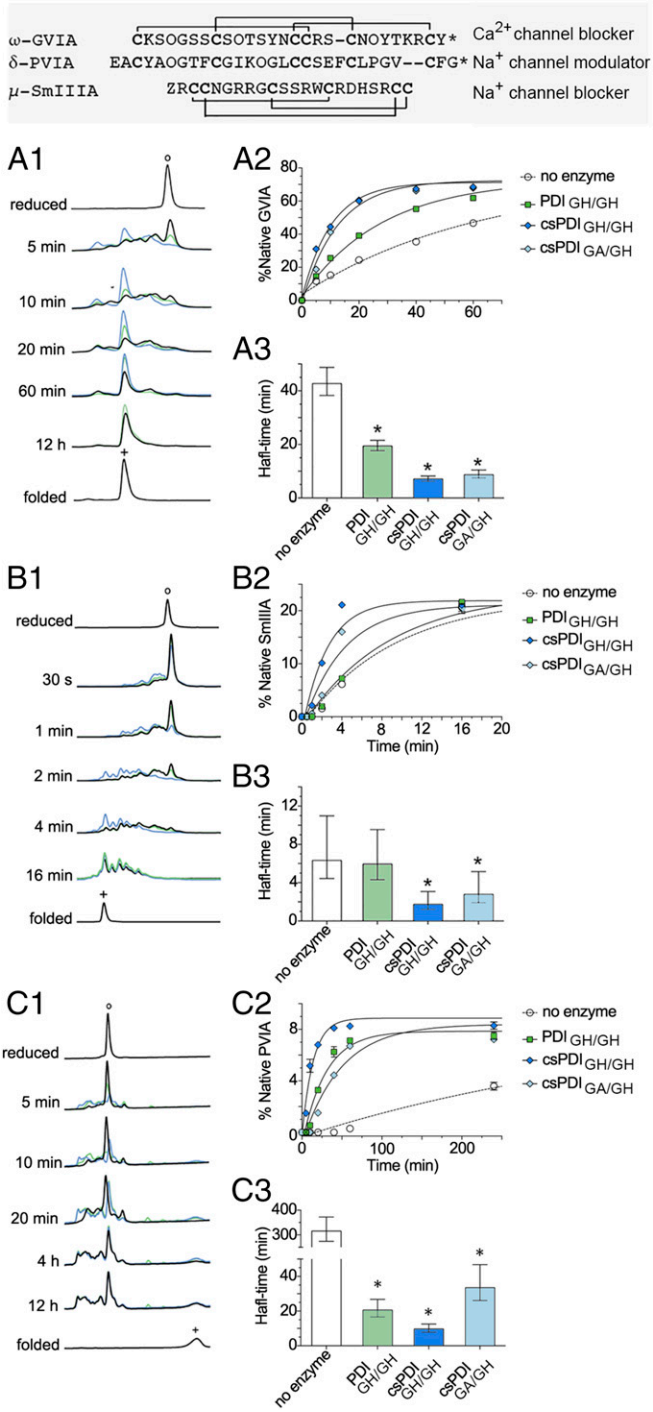


Fig. 5. Oxidative folding of conotoxin substrates in the presence of PDI and two members of the csPDI family from *C. geographus*. Sequences of the three conotoxin substrates tested are shown with their names, molecular targets, and disulfide connectivities. Amino acids: Z, pyroglutamate; O, hydroxyproline; *C-terminal amidation. (*A* and *B*) Folding assays for ω-GVIA (*A*) and μ-SmIIIA (*B*) were carried out at room temperature in the absence and presence of 2 μM enzyme in 100 mM Tris·HCl (pH 7.5), 1 mM EDTA, 0.4 mM reduced glutathione, and 0.2 mM oxidized glutathione. (*C*) Folding of δ-PVIA was performed at 4 °C and in the presence of 1% Tween-20. Reactions were initiated by adding 20 μM reduced toxin, quenched at different time points with formic acid (final 10% vol/vol), and analyzed by reverse-phase chromatography on a C_{18} column. Chromatograms of reactions without enzyme (black), with PDI (green), and with $csPDI_{GH/GH}$ (blue) are shown in *A1*, *B1*, and *C1*, respectively. The area under the curve was determined for the native, fully folded toxin and plotted against the reaction time (*A2*, *B2*, and *C2*). Half-time for the appearance of folded toxins (95% confidence values) was calculated in Prism (Version 6.0e; GraphPad) and is shown in *A3*, *B3*, and *C3*. Reactions that were significantly different from no-enzyme controls are indicated. *$P < 0.01$ (unpaired *t* test with Welch's correction).

($t_{1/2}$ = 33.5 min). Together, folding studies demonstrate that csPDIs are highly efficient in accelerating conotoxin folding and have distinct effects on the kinetics of disulfide-bond formation.

Discussion

Key evolutionary events can induce a rapid expansion and diversification of gene families to promote fitness and survival. An example is the parasitic liver fluke *Fasciola hepatica*. Cathepsin peptidases are important for the migration of the parasite thorough host tissue. The cathepsins in *F. hepatica* have greatly expanded and diverged to form multigenic families (22). These presumably play an important role at the host–parasite interface. The vast expansion of cathepsins was suggested to contribute to the high evolutionary potential of *F. hepatica* for infecting novel hosts and adapting to changes in the environment (22).

Similar observations have been made for venomous cone snails, in which a rapid expansion of multigene toxin families has facilitated exceptional rates of species diversification (14, 23). The molecular mechanisms behind the accelerated evolution of conotoxin-encoding genes are not fully understood, but high rates of gene duplication and positive selection have been repeatedly proposed (24, 25). Conotoxins are disulfide-rich peptides that highly selectively target a specific receptor or ion channel expressed in the nervous system of their prey, predators, or competitors. A conotoxin gene duplication could lead to advantageous neofunctionalization in one of the copies, which might act directly at the predator–prey interface, for which positive selection could be extremely high (26).

Here, we report on the expansion of a gene family involved in oxidative folding, a crucial step in conotoxin biosynthesis. Phylogenetic analysis revealed that this gene family evolved by gene duplication of an ancestral PDI gene. In humans, PDI is highly abundant and expressed in nearly all tissues types, where it serves in the formation, reduction, and isomerization of disulfide bonds (27). To date, no viable PDI knockout mouse has been reported demonstrating a crucial role of this enzyme in survival (28). The initial duplication of the PDI gene presumably allowed neofunctionalizations of the new PDI gene copy in the *Conus* venom gland, while maintaining the fundamental enzymatic properties of canonical PDI. As suggested by the presence of multiple csPDI variants in almost all cone snail species examined, the initial generation of the csPDI gene was followed by additional duplication events accompanied by high mutation rates that resulted in further gene specializations. Thus, csPDI expansion and diversification complemented the evolution of their conotoxin substrates, implying a rapidly changing need for oxidative folding of newly evolved disulfide-rich structural domains. A specialized function of the csPDIs in conotoxin folding is further supported by the finding that csPDIs are found in relatively very high abundance in the venom gland compared with other tissues (*SI Appendix*, Fig. S4).

Despite thousands of different conotoxin sequences, only a limited number of disulfide scaffolds are found in vivo, a phenomenon that has been referred to as the conotoxin folding puzzle (8). Conotoxins that significantly differ in their amino acid sequence efficiently adopt the same structural fold. However, in vitro, even toxins that contain the same cysteine framework often display an array of different folding properties and commonly adopt nonnative structures (8). To our knowledge, the csPDI family provides the first insight into addressing this biological conundrum. By guiding the folding of conotoxins into their native structural fold, csPDIs may eliminate the effects of extensive sequence variations observed in these peptide substrates.

Variation within the csPDI family is specifically found in regions that play an important role for enzyme activity and substrate binding. The greatest diversity was observed in *C. geographus*: Four

of the five csPDI enzymes had mutations in the two amino acids located between the active-site cysteine residues (CXXC) of the **a** and **a′** domain. Comparative alignment of all csPDI sequences further detected a conspicuous sequence variation on the +2 position C-terminal of the CXXC active-site motif in both **a** and **a′** domains. We are not aware of any systematic investigation of the potential functional consequence of mutating residues at this position in redox-active thioredoxin-like domains. Still, the close proximity to the active site could well indicate an influence of the residue at this position in modulating the active-site reduction potential and thereby the redox activity of the given enzyme. If so, the csPDIs could use sequence variation at this position to modulate their redox activity to assist the folding of specific conotoxins. In contrast, the active site motif of canonical PDI was conserved in all species, pointing to a more constrained role of this enzyme.

Analysis of position-specific sequence variations demonstrated that the **b′** domain showed pronounced variation in the csPDIs, but remains highly conserved for PDI (Fig. 4 and *SI Appendix*, Fig. S7). In human PDI, a hydrophobic patch in **b′** is important for domain–domain interactions between **b′** and **a′** and for binding substrates directly (29–31). This hydrophobic patch is clearly conserved in the *Conus* PDIs. Notably, many of the residues of the hydrophobic patch show sequence variation in the csPDIs (Fig. 4*C*). Despite sequence variation, the hydrophobic nature of this patch is kept intact in the csPDIs. Thus, we speculate that this region is also involved in substrate binding in csPDIs, but may have evolved to accommodate a more diverse set of substrate peptides.

Functional characterization was carried out with two csPDI variants from *C. geographus* that had active site motifs found in almost all other cone snail species: $csPDI_{GH/GH}$ and $csPDI_{GA/GH}$. Oxidative folding assays using several conotoxin substrates confirmed that these enzymes are highly efficient in accelerating conotoxin folding and showed distinct effects on the kinetics of disulfide bond formation compared with PDI.

In conclusion, the discovery and subsequent characterization of the csPDI gene family represents an evolutionary showcase for the dynamic interplay between enzymes and their hypervariable substrates and provide important insight into the complex folding machinery evolved in conoidean venoms.

Materials and Methods

Detailed material and methods are provided in *SI Appendix, SI Materials and Methods*. Briefly, transcriptomes were sequenced on an Illumina HiSeq instrument, assembled by using Trinity software, and annotated by BLASTx. Additional csPDI sequence variants were discovered by using the recently developed software tool Taxonomer. Sequences were confirmed from several species by RT-PCR. The 2D gel electrophoresis coupled with mass spectrometric analysis confirmed the presence of PDI and csPDIs in the venom gland of *C. geographus*. qPCR and RNA-Seq on different cone snail tissues demonstrated high expression levels of the csPDI family in the venom gland. Recombinant *C. geographus* enzymes were expressed in *Escherichia coli* and purified by metal affinity and size-exclusion chromatography. Oxidative folding studies were carried out by using Fmoc synthesized linear conotoxins. Folding reactions were analyzed by reverse-phase chromatography.

ACKNOWLEDGMENTS. We thank Dr. William Low for MS analysis of synthetic peptides; the sequencing and genomics core facility at the University of Utah for DNA sequencing; My Thi Thao Huynh for help with image presentation; and Dr. Alexander Fedosov for dissecting different cone snail tissues. This work was supported in part by National Institutes of Health Grants GM 48677 (to B.M.O. and P.K.B.) and GM 099939 (to M.Y. and P.K.B.); and European Commission International Outgoing Fellowship Grant CONBIOS 330486 (to H.S.-H. and L.E.). A.W.P. acknowledges fellowship support from the Australian National Health and Medical Research Council. C.W.G. is an Australian Research Council Future Fellow. W.B. is supported by the Villum Foundation.

1. Hu H, Bandyopadhyay PK, Olivera BM, Yandell M (2012) Elucidation of the molecular envenomation strategy of the cone snail *Conus geographus* through transcriptome sequencing of its venom duct. *BMC Genomics* 13(284):284.
2. Robinson SD, et al. (2014) Diversity of conotoxin gene superfamilies in the venomous snail, *Conus victoriae*. *PLoS One* 9(2):e87648.
3. Lu A, Yang L, Xu S, Wang C (2014) Various conotoxin diversifications revealed by a venomic study of *Conus flavidus*. *Mol Cell Proteomics* 13(1):105–118.
4. Tayo LL, Lu B, Cruz LJ, Yates JR, 3rd (2010) Proteomic analysis provides insights on venom processing in *Conus textile*. *J Proteome Res* 9(5):2292–2301.
5. Norton RS, Pallaghy PK (1998) The cystine knot structure of ion channel toxins and related polypeptides. *Toxicon* 36(11):1573–1583.
6. Bayrhuber M, et al. (2005) Conkunitzin-S1 is the first member of a new Kunitz-type neurotoxin family. Structural and functional characterization. *J Biol Chem* 280(25): 23766–23770.
7. Safavi-Hemami H, et al. (2015) Specialized insulin is used for chemical warfare by fish-hunting cone snails. *Proc Natl Acad Sci USA* 112(6):1743–1748.
8. Bulaj G, Olivera BM (2008) Folding of conotoxins: Formation of the native disulfide bridges during chemical synthesis and biosynthesis of *Conus* peptides. *Antioxid Redox Signal* 10(1):141–155.
9. Safavi-Hemami H, Bulaj G, Olivera BM, Williamson NA, Purcell AW (2010) Identification of *Conus* peptidylprolyl cis-trans isomerases (PPIases) and assessment of their role in the oxidative folding of conotoxins. *J Biol Chem* 285(17):12735–12746.
10. Safavi-Hemami H, et al. (2012) Modulation of conotoxin structure and function is achieved through a multienzyme complex in the venom glands of cone snails. *J Biol Chem* 287(41):34288–34303.
11. Ellgaard L, Ruddock LW (2005) The human protein disulphide isomerase family: Substrate interactions and functional properties. *EMBO Rep* 6(1):28–32.
12. Graf EH, et al. (2016) Unbiased detection of respiratory viruses using RNA-seq-based metagenomics: A systematic comparison to a commercial PCR panel. *J Clin Microbiol* 03060-15.
13. Bulaj G, et al. (2003) Efficient oxidative folding of conotoxins and the radiation of venomous cone snails. *Proc Natl Acad Sci USA* 100(Suppl 2):14562–14568.
14. Puillandre N, et al. (2014) Molecular phylogeny and evolution of the cone snails (Gastropoda, Conoidea). *Mol Phylogenet Evol* 78:290–303.
15. Safavi-Hemami H, et al. (2014) Combined proteomic and transcriptomic interrogation of the venom gland of *Conus geographus* uncovers novel components and functional compartmentalization. *Mol Cell Proteomics* 13(4):938–953.
16. Murrell B, et al. (2012) Detecting individual sites subject to episodic diversifying selection. *PLoS Genet* 8(7):e1002764.
17. Pirneskoski A, et al. (2004) Molecular characterization of the principal substrate binding site of the ubiquitous folding catalyst protein disulfide isomerase. *J Biol Chem* 279(11):10374–10381.
18. Wang C, et al. (2013) Structural insights into the redox-regulated dynamic conformations of human protein disulfide isomerase. *Antioxid Redox Signal* 19(1):36–45.
19. Price-Carter M, Gray WR, Goldenberg DP (1996) Folding of omega-conotoxins. 2. Influence of precursor sequences and protein disulfide isomerase. *Biochemistry* 35(48): 15547–15557.
20. Fuller E, et al. (2005) Oxidative folding of conotoxins sharing an identical disulfide bridging framework. *FEBS J* 272(7):1727–1738.
21. DeLa Cruz R, Whitby FG, Buczek O, Bulaj G (2003) Detergent-assisted oxidative folding of delta-conotoxins. *J Pept Res* 61(4):202–212.
22. Cwiklinski K, et al. (2015) The Fasciola hepatica genome: Gene duplication and polymorphism reveals adaptation to the host environment and the capacity for rapid evolution. *Genome Biol* 16:71.
23. Stanley SM (2008) Predation defeats competition on the seafloor. *Paleobiology* 34: 1–21.
24. Chang D, Duda TFJ, Jr (2012) Extensive and continuous duplication facilitates rapid evolution and diversification of gene families. *Mol Biol Evol* 29(8):2019–2029.
25. Puillandre N, Watkins M, Olivera BM (2010) Evolution of Conus peptide genes: Duplication and positive selection in the A-superfamily. *J Mol Evol* 70(2):190–202.
26. Kordis D, Gubensek F (2000) Adaptive evolution of animal toxin multigene families. *Gene* 261(1):43–52.
27. Okumura M, Kadokura H, Inaba K (2015) Structures and functions of protein disulfide isomerase family members involved in proteostasis in the endoplasmic reticulum. *Free Radic Biol Med* 83:314–322.
28. Hatahet F, Ruddock LW (2009) Protein disulfide isomerase: A critical evaluation of its function in disulfide bond formation. *Antioxid Redox Signal* 11(11):2807–2850.
29. Denisov AY, et al. (2009) Solution structure of the bb′ domains of human protein disulfide isomerase. *FEBS J* 276(5):1440–1449.
30. Nguyen VD, et al. (2008) Alternative conformations of the x region of human protein disulphide-isomerase modulate exposure of the substrate binding b′ domain. *J Mol Biol* 383(5):1144–1155.
31. Wang C, et al. (2012) Human protein-disulfide isomerase is a redox-regulated chaperone activated by oxidation of domain a′. *J Biol Chem* 287(2):1139–1149.
32. Zdobnov EM, Apweiler R (2001) InterProScan–an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* 17(9):847–848.
33. Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17(8):754–755.