Scorpion Toxins: Positive Selection at a Distal Site Modulates Functional Evolution at a Bioactive Site

Limei Zhu,[1,2] Bin Gao,[†,1] Shouli Yuan,[†,1] and Shunyi Zhu*[,1]

[1]Group of Peptide Biology and Evolution, State Key Laboratory of Integrated Management of Pest Insects and Rodents, Institute of Zoology, Chinese Academy of Sciences, Chaoyang District, Beijing, China

[2]Key Laboratory of Tropical Marine Bio-resources and Ecology, Guangdong Provincial Key Laboratory of Applied Marine Biology, South China Sea Institute of Oceanology, Chinese Academy of Sciences, Guangzhou, China

[†]These authors contributed equally to this work.

***Corresponding author:** E-mail: zhusy@ioz.ac.cn.

Associate editor: Mary O'Connell

Abstract

The bioactive sites of proteins are those that directly interact with their targets. In many immunity- and predation-related proteins, they frequently experience positive selection for dealing with the changes of their targets from competitors. However, some sites that are far away from the interface between proteins and their targets are also identified to evolve under positive selection. Here, we explore the evolutionary implication of such a site in scorpion α-type toxins affecting sodium (Na^+) channels (abbreviated as α-ScNaTxs) using a combination of experimental and computational approaches. We found that despite no direct involvement in interaction with Na^+ channels, mutations at this site by different types of amino acids led to toxicity change on both rats and insects in three α-ScNaTxs, accompanying differential effects on their structures. Molecular dynamics simulations indicated that the mutations changed the conformational dynamics of the positively selected bioactive site-containing functional regions by allosteric communication, suggesting a potential evolutionary correlation between these bioactive sites and the distant nonbioactive site. Our results reveal for the first time the cause of fast evolution at nonbioactive sites of scorpion neurotoxins, which is presumably to adapt to the change of their bioactive sites through coevolution to maintain an active conformation for channel binding. This might aid rational design of scorpion Na^+ channel toxins with improved phyletic selectivity via modification of a distant nonbioactive site.

Key words: **bioactive site, distal site, toxin–channel interface, long-range communication, positive selection.**

Introduction

Proteins perform various biological functions through their bioactive sites to interact directly with their targets. Although these sites have been a focus of studies on the structure–function relationship of proteins and their evolution, the roles of the sites distant from the bioactive sites (herein termed distant sites) in modulating the catalytic activity and substrate specificity of enzymes are now being recognized (Campbell et al. 2016; Whitney et al. 2016; Ahuja et al. 2017). Directed evolution combined with structural and functional studies has revealed that some mutations leading to high activity are scattered remotely from the active sites of enzymes and especially mutations at these distant sites can markedly alter conformational dynamics of the catalytic residues of enzymes (Jiménez-Osés et al. 2014; Campbell et al. 2016; Whitney et al. 2016; Ahuja et al. 2017). However, whether mutations at distant sites can modulate functions of nonenzymatic proteins, and how they exert effects are poorly understood. To address these two questions, one needs to rapidly identify distant nonbioactive sites associated with the protein function change and then to study their communication with the active site. In this regard, distant sites in nonenzymatic proteins evolved under positive selection may be particularly attractive candidates, as accelerated substitutions at a protein site is generally related to functional adaptation and they can be easily detected by statistical methods in the genomic era (Yang 2002).

Positively selected sites (PSSs) refer to a class of amino acid sites with an excess of nonsynonymous to synonymous substitutions, which are indicators of adaptive evolution, arms races, or compensatory substitutions (Pál et al. 2006). PSSs have been identified in various protein families involved in a diversity of biological processes, such as immunity and predation (Sawyer et al. 2005; Zhu et al. 2005; Juarez et al. 2008; Tian et al. 2008; Puillandre et al. 2010; Kozminsky-Atias and Zilberberg 2012; Wong and Belov 2012; Vonk et al. 2013; Sironi et al. 2015; Pontremoli et al. 2016). In these examples, positive selection primarily acts on the bioactive sites to exert effects on the diversity of protein functions. A paradigm in this aspect is the major histocompatibility complex (MHC), a set of cell surface proteins essential for binding pathogen-derived oligopeptides (antigens), where nearly all PSSs fall in the antigen recognition site (ARS) domain (Hughes and Nei 1988; Yang and Swanson 2002). In addition to these positively selected bioactive sites, some sites far away from bioactive

Article

sites have also been detected to be under positive selection (Kapralov and Filatov 2007; Tian et al. 2008; Kozminsky-Atias and Zilberberg 2012; Wang et al. 2012; Sunagar et al. 2013; Ren et al. 2014; Zupunski and Kordis 2016). Because of their distant locations, few mutational experiments have been conducted to uncover their functional and evolutionary implications.

The α-ScNaTxs are a group of evolutionarily conserved neurotoxins specifically affecting voltage-gated sodium channels (Na_vs) (fig. 1A) (Possani et al. 1999; Bosmans and Tytgat 2007; Zhu et al. 2012), a class of transmembrane proteins responsible for the initiation of action potentials in neurons and other excitable cells (Catterall 2000). They adopt the cysteine-stabilized α-helix and β-sheet (CSαβ) fold cross-linked by four disulfide bridges (Fontecilla-Camps et al. 1980). As the essential lethal components of the venom of the Old World scorpions (Possani et al. 1999; Zhu et al. 2012), they are involved in predation of these scorpions on their prey (e.g., insects and spiders) and defense against their predators (e.g., birds, lizards, and small rodents) (Polis 1990; Zhang et al. 2015; Gao and Zhu 2018). These toxins function as modulators of voltage gating of Na_vs via binding to the receptor site 3 to prevent the outward movement of the S4 segment in DIV (fig. 1A) and thus slow down the rapid inactivation of Na_vs (Cestèle and Catterall 2000). According to their preference for insect and mammalian Na_vs, α-ScNaTxs can be divided into three subtypes: 1) Classical α-toxins that are highly toxic on mammals but nontoxic to insects; 2) Anti-insect α-toxins that are especially toxic on insects but weakly toxic on mammals; and 3) α-Like toxins that act on both mammalian and insect Na_vs (Possani et al. 1999; Zhu et al. 2012). Mutational analyses of different α-ScNaTxs have revealed a common bioactive surface that is primarily located on two loops (i.e., B-loop and J-loop) (termed "core-domain") for receptor binding. Another bioactive surface that makes up of the N-terminal five-residue turn and the C-tail (named "NC-domain") acts as an entirety only in some members, such as LqhαIT and BmKM1 (Wang et al. 2003; Karbat et al. 2004; Gordon et al. 2007; Zhu et al. 2016). In other α-ScNaTxs, for example, Lqh2, Lqh3, and MTα-5, the N-turn is not indispensable in receptor binding (Karbat et al. 2007; Kahn et al. 2009; Zhu et al. 2016). As the common bioactive surface of α-ScNaTxs, the core-domain contributes multiple hot-spot residues to interaction with the primary site of the receptor located on the voltage-sensing domain (VSD) in DIV (fig. 1). A short SS2–S6 loop in D1 that is structurally adjacent to the primary receptor site could provide the secondary site for toxin binding via their NC-domain (fig. 1A) (Wang et al. 2011; Zhang et al. 2015).

Because subtle variations at the bioactive surface of α-ScNaTxs can induce significant differences in potency and selectivity (Gordon and Gurevitz 2003; Gordon et al. 2007), this class of toxins has been a focus of molecular evolution studies. To date, there are at least five studies reporting PSSs of α-ScNaTxs with similar statistical methods but different data sets (Zhu et al. 2004; Weinberger et al. 2010; Kozminsky-Atias and Zilberberg 2012; Zhu et al. 2012; Zhang et al. 2015). All these studies indicate that positive selection frequently acts on the core domain (Zhu et al. 2012), in line with its functional importance in receptor binding. We have previously identified nine PSSs in the α-ScNaTxs from two *Mesobuthus* species (i.e., *M. eupeus* and *M. martensii*) with CodeML, a program for inference of positive selection through phylogenetic comparison of protein-coding genes (Zhu et al. 2012). By using site-directed mutagenesis, we defined the functional roles of these PSSs in MTα-5, an α-like toxin from *M. eupeus* (initially abbreviated as MT-5 [Zhu et al. 2016] and herein renamed MTα-5 to distinguish from the β-toxins in the same venom), in which sites I^{40} and L^{41} located at the core domain (B-loop) serve as two hot-spot residues directly interacting with the receptor. Remarkably, site 52 is the only PSS that is located at the opposite side of the toxin–channel interface, and its location belongs neither to the core domain, nor to the NC-domain (fig. 1B), thus representing an ideal candidate for studying the roles of a distant site in modulating the structure and function of a nonenzymatic protein.

In this work, we carried out mutational analysis of site 52 in three α-ScNaTxs. In combination with molecular dynamics simulations (Yuan et al. 2017) and statistical coupling analysis (SCA) (Halabi et al. 2009), our experimental data demonstrated that the accelerated substitutions of site 52 is a result of adaptation to the evolutionary change of the bioactive PSSs, and that such coevolution may help the substituted bioactive sites to maintain an active conformation for receptor binding. Our study thus suggests that within one nonenzymatic protein, there exists also allosteric communication between its bioactive and nonbioactive distant sites.

Results

In this study, we chose three α-like toxins from *M. eupeus* (MTα-5, MTα-12 and MTα-13) to construct multiple site 52 mutants, including MTα-5(A52F), MTα-5(A52G), MTα-5(A52L), MTα-5(A52N), and MTα-5(A52P); MTα-12(E50V) (site 50 in this toxin structurally equivalent to site 52 in other toxins described here); and MTα-13(E52V). The introduction of a valine in MTα-12 and MTα-13 was inspired by the observation that two natural neurotoxins (MTα-4 and MTα-5) in the *M. eupeus* venom, differing by only one amino acid substitution at site 52 (V52A), exhibit different species selectivity (Zhu et al. 2012). This mutation was to test whether a valine at this site would evoke a shift of species selection to insects. For MTα-5, more mutations were introduced into site 52 in view of the availability of its experimental 3D structure (Zhu et al. 2012), allowing further analysis of the mutational effects by molecular dynamics simulations. These mutations included three different types of amino acids: 1) Hydrophobic leucine and proline; 2) Hydrophilic glycine and asparagine; 3) Aromatic phenylalanine. The choice of houseflies and a rat Na_v as targets for functional assays was based on an ecological consideration on the roles of α-ScNaTxs in scorpion predation and defense, where insects are the prey and rats are the predators (Polis 1990). According to the method previously described in Zhu et al. (2013), we prepared the recombinant products of these mutants, all of which were eluted as a single

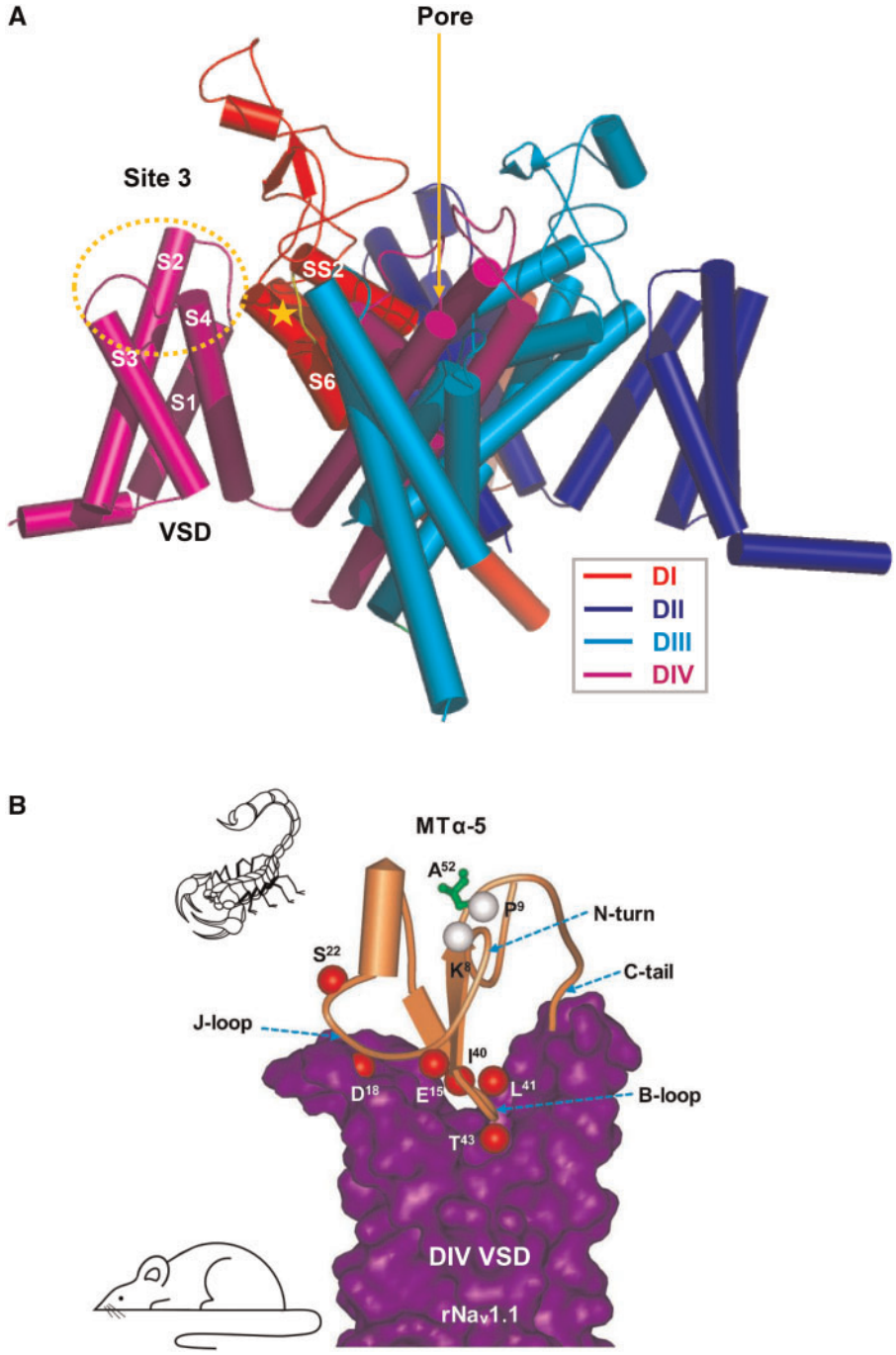

FIG. 1. Site 52 is a distal site far from the toxin-channel interface. (*A*) The structure of a representative Na_v (α-subunit) from *Periplaneta american* (pdb entry 5X0M) for showing positions of the receptor site 3 for α-ScNaTxs and the pore for Na^+ ions to pass through. Four repeat domains (DI–DIV) are shown in different colors where four transmembrane segments (S1–S4) of each domain comprise the voltage-sensing domain (VSD) and S5 and S6 of each domain commonly forms the pore module (Shen et al. 2017). The receptor is primarily located on the VSD of DIV, including two extracellular loops linking S1, S2 and S3, S4 fragments (the area circled by dotted line). For some toxins, the short SS2–S6 loop in D1 (shown in *yellow* and marked by a pentagram) is also likely involved in toxin binding (Cestèle and Catterall 2000; Wang et al. 2011). (*B*) A representative configuration of MTα-5 (orange) in complex with the DIV VSD of $rNa_v1.1$ (purple) based on our published work (Zhu et al. 2016). Site 52 is displayed as ball-and-stick models and colored in *green*; for other PSSs, only Cα atoms are shown as balls with those involved in toxin–channel interaction in *red*. In this complex model, two PSSs (I^{40} and L^{41}) in the B-loop bind directly to $Na_v1.1$, in agreement with our previous mutational data (Zhu et al. 2016).

peak at 22–25 min (see supplementary table S1, Supplementary Material online; fig. 2*A* as a representative of reverse phase high-performance liquid chromatography (RP-HPLC) chromatographic profiles of all recombinant toxins). Their experimental molecular weights (MWs) determined by MALDI-TOF matched their theoretical MWs (see supplementary table S1, Supplementary Material online), indicating that they have been oxidized to form four disulfide bridges and the initiation methionine has been automatically removed during purification (Zhu et al. 2013, 2016). All

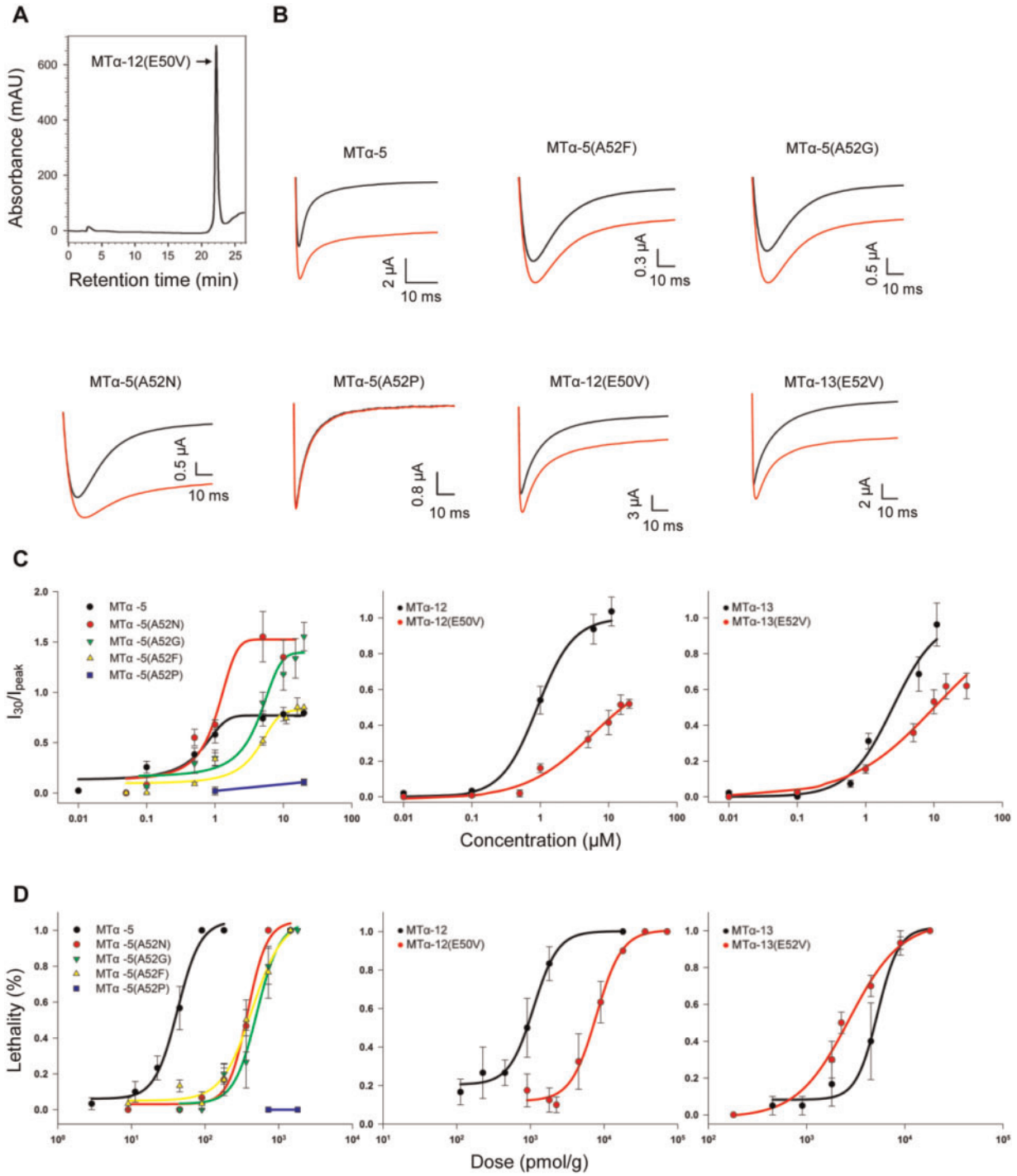

FIG. 2. Purification and functional characterization. (*A*) RP-HPLC showing retention time of MTα-12(E50V) (here as an example). C18 column (Agilent Zorbax 300SB-C18, 4.6 × 150 mm, 5 μM) was equilibrated with 0.05% trifluoroacetic acid in water, with a linear gradient of 0–60% acetonitrile within 40 min with a flow rate of 1 ml/min. (*B*) Representative whole cell Na^+ current traces of oocytes expressing cloned $rNa_v1.1$. Red line indicates the steady-state current peak amplitude in the presence of 1 μM MTα-5 and its mutants or 10 μM MTα-12(E50V) and MTα-13(E52V). Current traces were evoked by a depolarization step to −10 mV from a holding potential of −90 mV. (*C*) Dose–response curves of toxins and mutants on $rNa_v1.1$ expressed in *X. laevis* oocytes. The dose–response curves were obtained by plotting the relative I_{30ms}/I_{peak} values of the channels in function of the toxin concentrations. All the curves were fitted with the Hill equation. Each point represents mean ± SE ($n \geq 6$). (*D*) Comparison of the insect toxicity of wild-type toxins and their mutants. Dose–response curves were fitted with Hill equation, $y = (a-b)/[1 + (x/LD_{50})^n] + b$. Each point is the mean ± SE of three experiments.

mutants except for MTα-5(A52L) were dissolvable in aqueous solution. The poor solubility of this toxin hampered further functional assays.

$rNa_v1.1$, the most sensitive mammalian Na_v subtype reported for MTα-5, MTα-12, and MTα-13 (Zhu et al. 2013, 2016), was used to evaluate the pharmacological activity of

Table 1. The Toxicity of Recombinant *Mesobuthus* α-ScNaTxs and Their Mutants on $rNa_v1.1$ and Housefly Adults.

Toxin	$rNa_v1.1$ (EC_{50}, μM)	Fold Reduction	Toxicity on Houseflies (LD_{50}, nmol/g)	Relative Toxicity (%)
MTα-5	0.55 ± 0.11	—	0.42 ± 0.04	—
MTα-5(A52F)	3.92 ± 0.99	6.13	4.32 ± 0.80	9.72
MTα-5(A52G)	4.09 ± 0.82	6.44	5.05 ± 0.72	8.32
MTα-5(A52K)*	1.14 ± 0.21	1.53	1.22 ± 0.05	34.60
MTα-5(A52L)	i.d.	—	i.d.	—
MTα-5(A52N)	1.02 ± 0.21	0.85	3.78 ± 0.34	11.11
MTα-5(A52P)	n.a.	>35.36	>18.05	<2.33
MTα-12	0.91 ± 0.08	—	1.10 ± 0.11	—
MTα-12(E50V)	5.86 ± 4.7	5.4	7.86 ± 0.66	13.99
MTα-13	2.50 ± 0.44	—	5.27 ± 0.52	—
MTα-13(E52V)	10.1 ± 1.17	3	2.70 ± 0.45	195.19

NOTE.—Fold reduction of a mutant on $rNa_v1.1$ was calculated as (EC_{50} of a mutant−EC_{50} of its wild-type toxin)/EC_{50} of the wild-type toxin, and relative toxicity of a mutant on housefly adults was calculated as a percentage of (LD_{50} of a wild-type toxin/LD_{50} of its mutant). "n.a." means no activity observed at 20 μM. "i.d.", inability to determine due to insolubility of the peptide. "*", data from the reference (Zhu et al. 2016).

these toxins and their mutants on mammals (fig. 2*B* and *C*). Compared with MTα-5, MTα-5(A52N) had a similar activity at 1 μM, but MTα-5(A52G) and MTα-5(A52F) showed significantly decreased activity whereas MTα-5(A52P) totally lost the activity as no effect was observed even at 20 μM. MTα-12(E50V) and MTα-13(E52V) exhibited significantly decreased activity at 10 μM when compared with their respective wild-type peptides (fig. 2*B*). We further determined the concentration for 50% of maximal effect (EC_{50}) of each peptide except for MTα-5(A52P) (fig. 2*C* and table 1). Quantitatively, the activity of MTα-5(A52N) was comparable with that of MTα-5 (EC_{50}: 1.02 ± 0.21 μM vs. 0.55 ± 0.11 μM) whereas MTα-5(A52G) and MTα-5(A52F) both showed >6-fold reduced activity. When compared with their respective parental peptides, MTα-12(E50V) and MTα-13(E52V) both reduced their activity by 5.4-fold and 3-fold, respectively.

The insecticidal activity of the mutants on housefly adults is shown in figure 2*D* and table 1. Again, MTα-5(A52P) showed no activity even at a peptide dose up to 18.05 nmol/g and all other MTα-5 mutants showed a significantly decreased toxicity compared with the parental peptide. The median lethal doses (LD_{50}) determined were 0.42 ± 0.04 nmol/g for MTα-5, 4.32 ± 0.80 nmol/g for MTα-5(A52F), 5.05 ± 0.72 nmol/g for MTα-5(A52G), and 3.78 ± 0.34 nmol/g for MTα-5(A52N), indicating that these mutants only remained about 10% residual toxicity of MTα-5. MTα-12(E50V) retained 14% toxicity of MTα-12 (LD_{50} of 7.86 ± 0.66 vs. 1.10 ± 0.11 nmol/g). Intriguingly, MTα-13(E52V) showed an enhanced toxicity compared with MTα-13, as identified by their LD_{50} values (2.70 ± 0.45 vs. 5.27 ± 0.52 nmol/g), which was equivalent to 195.19% toxicity of MTα-13.

Given that site 52 is spatially far away from the bioactive face of α-ScNaTxs, the mutation-induced functional effects observed here may arise from structural alteration. To address this issue, we used circular dichroism (CD) spectroscopy, a technique classically used to assess the secondary structure of proteins (Greenfield 2006), to compare potential structural effects of the mutations. From the CD spectra shown in figure 3, we observed three categories of mutation-induced structural effects: 1) Mutations without destroying the structure of a parental peptide. These include A52F and A52N of MTα-5 and E50V of MTα-12. They all had a very similar CD spectrum to their respective parental toxins (fig. 3*A* and *D*); 2) Mutations destroying the structure of a parental peptide, such as A52L and A52P in MTα-5. For these two mutants, their CD spectra remarkably changed relative to the unmodified peptide which had a spectral signature at 208 nm (negative maximum) and 230 nm (positive maximum) (fig. 3*C*). This provides a rational explanation for MTα-5(A52L) aggregation in solution and the complete activity loss in MTα-5(A52P); 3) Mutations changing the stability of a parental peptide, such as A52G in MTα-5 and E52V in MTα-13. Although these two mutants had an overall similar CD spectrum to those of their respective wild-type toxins, their positive maximum at 230 nm markedly altered in intensity (fig. 3*B* and *E*). The intensity at 230 nm increased in A52G(MTα-5) but decreased in MTα-13(E52V), indicating that the mutation from Ala to Gly enhanced the stability of MTα-5 whereas the mutation from Glu to Val reduced the stability of MTα-13. This conclusion is further supported by our variable-temperature CD analysis of MTα-13 (fig. 3*F*). Obviously, its CD band at 230 nm decreased in intensity with increasing temperature and virtually disappeared at 60 °C (fig. 3*F*). Taken together, our CD data revealed a differential structural effect induced by mutations at site 52, which appears to depend on the amino-acid type introduced and toxin individuals used.

To answer the question of whether the activity change in the MTα-5 mutants with no visible structural perturbation is due to conformational alteration, we conducted molecular dynamics (MD) simulations on MTα-5 and its two mutants (A52F and A52N). The root-mean-square deviations (RMSDs) of their Cα atoms during 40 ns simulations are shown in figure 4*A*. Analysis of the root-mean-square fluctuation (RMSF) revealed more flexibility in several loops of A52F, including two bioactive loops (B-loop and J-loop) (fig. 4*B* and *C*). In the mutant A52N, more flexibility was observed in the C-tail (fig. 4*B* and *C*). For Cα atoms of the two hot-spot residues in the two mutants, their positions

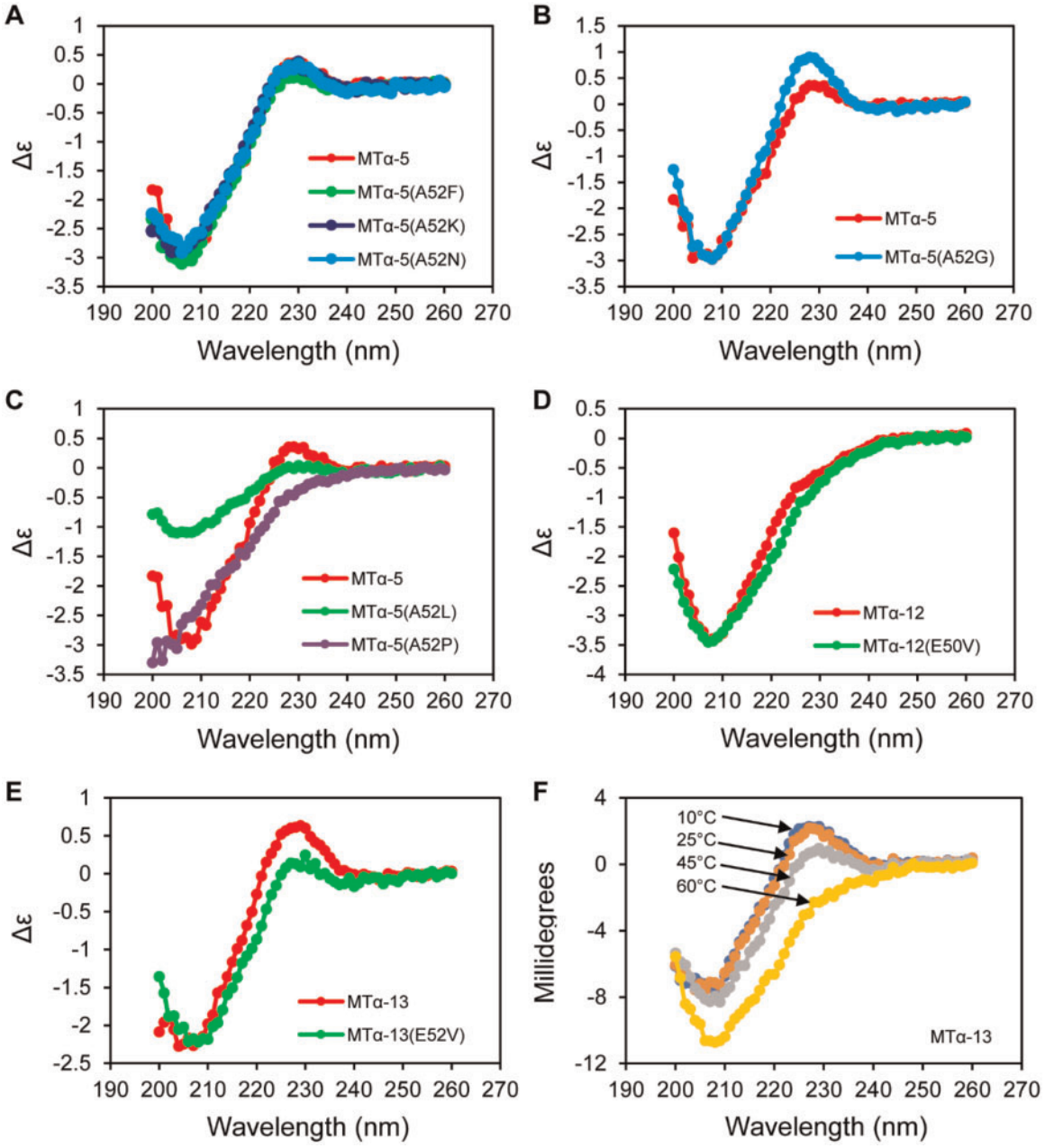

Fig. 3. Comparison of CD spectra between wild-type toxins and mutants. (*A*–*C*) MTα-5 and its mutants (Note: A52K [Zhu et al. 2016]); (*D*) MTα-12 and MTα-12(E50V); (*E*) MTα-13 and MTα-13(E52V); (*F*) variation of the CD spectra of MTα-13 with temperature.

relative to those of the unmodified toxins changed (fig. 4*D*). Quantitatively, the values of I^{40} position changes were 3.24 ± 0.52 Å for A52N and 3.87 ± 0.61 Å for A52F; For L^{41}, the values were 4.65 ± 0.61 Å for A52N and 5.76 ± 0.76 Å for A52F (fig. 4*E*). Compared with A52N, A52F showed a greater position change in the two bioactive sites, highlighting the importance of a proper rigidity in these residues for the toxicity of MTα-5 on $rNa_v1.1$. The declined insecticidal activity in A52N also hints a putative functional role of the C-tail because its conformation was also influenced by the mutation. However, even without obvious change in the C-tail, the A52F mutant also showed a decreased insect toxicity, suggesting that the insect toxicity of ScNaTxs may require multiple functional regions in a proper rigidity. In parallel, we also calculated the position change of Cα atoms at site 52 of the two mutants, the values were 5.51 ± 1.05 Å for A52N and 3.70 ± 0.74 Å for A52F (fig. 4*E*), indicating that relative to the unmodified toxin, the two mutants showed some position changes at this site. Such changes may stem from local structural rearrangement in the context of different side chains, which is obviously different from the position changes observed in I^{40} and L^{41} where there exists no side-chain difference between the wild-type peptide and its two mutants. Since site 52 is not involved in a direct interaction with Na_vs, we assume that this kind of position changes could exert minor direct impact on the toxin function. All these

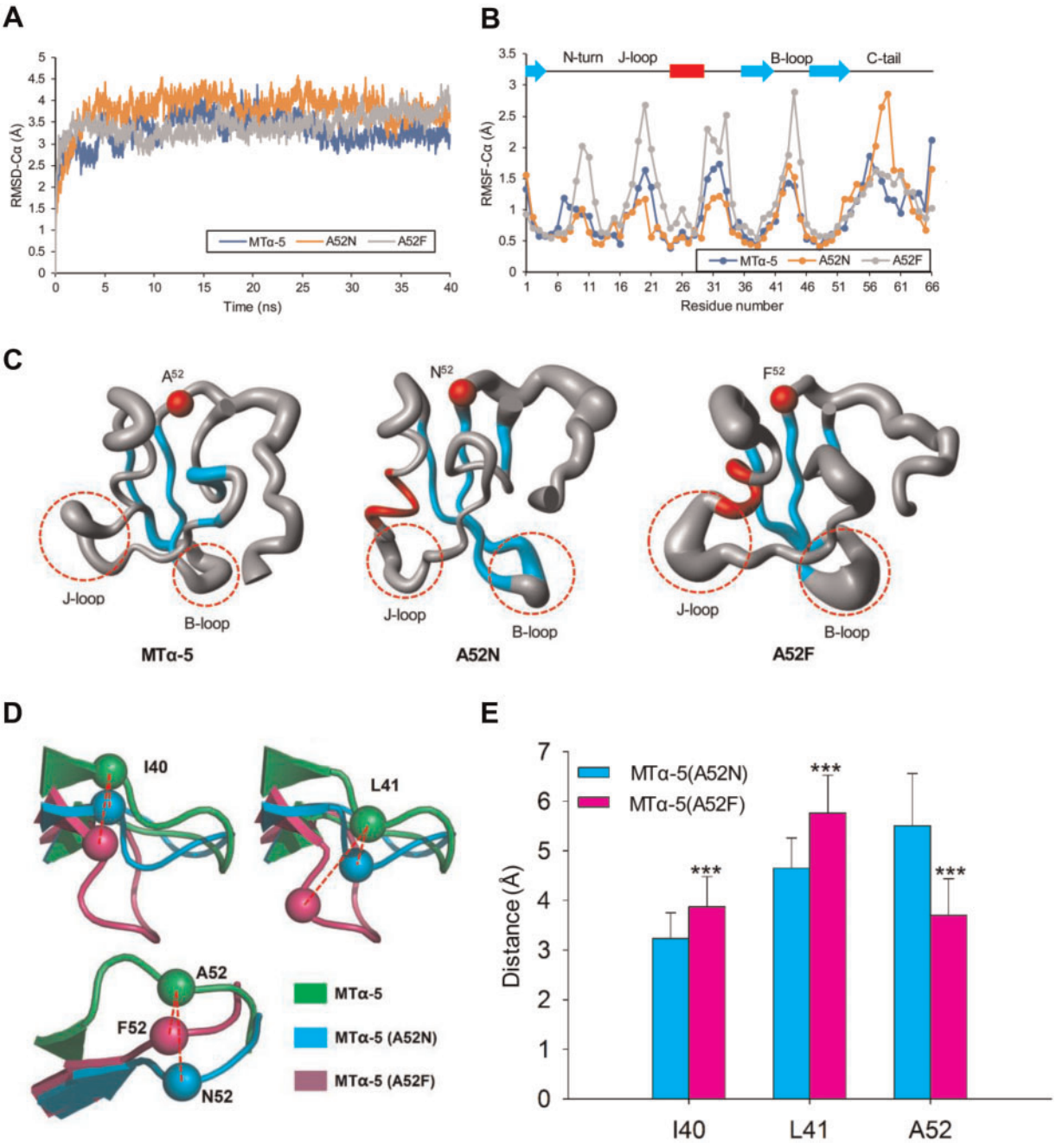

FIG. 4. MD simulations of MTα-5 and its mutants. (*A*) Cα-RMSDs during 40 ns simulations; (*B*) Cα-RMSFs of MTα-5 (30–40 ns), MTα-5(A52N), and MTα-5(A52F) (both 20 and 40 ns); (*C*) Conformational ensembles represented by the "sausage" model with MOLMOL (https://sourceforge.net/projects/molmol/files/, last accessed September 10, 2015). The Cα atom of site 52 for each toxin is displayed as a red sphere. (*D*) Site 52 mutation-induced position changes at sites 40 and 41, represented by their Cα atoms (red dotted lines). The position change of site 52 itself is also shown. (*E*) Quantitative analysis of the position changes in the Cα atom of I^{40} or L^{41} induced by the mutation from Ala to Asn (A52N) or Phe (A52F). The quantitative position change for site 52 itself is also presented. Data were extracted from conformational snapshots of the simulations every 100 ps and expressed as mean ± SD ($n = 201$). The differences between the groups were compared by unpaired t-test with SPSS19.0. $^{***}P < 0.001$.

observations indicated that mutations at a distant site can change the functional property of a ScNaTx through allosteric effect to affect the conformation of its functional regions.

The finding of the allosteric communication between site 52 and the functional loop of the toxin prompts us to study their coevolutionary relationship. To this end, SCA, a method for identifying groups of coevolving residues (called sectors) (Halabi et al. 2009), was applied to a multiple sequence alignment of α-ScNaTxs from two *Mesobuthus* species (see supplementary fig. S1, Supplementary Material online), which detected one sector comprising site 52 and ten other residues (see supplementary fig. S2, Supplementary Material online). Among these coevolving residues, fours (i.e., S^{21} form the J-loop; Q^{39} and I^{40} from the B-loop; and N^{56} from the C-tail segment), were found to be allosterically affected by mutations at site 52 in our MD simulations (fig. 4), suggesting that

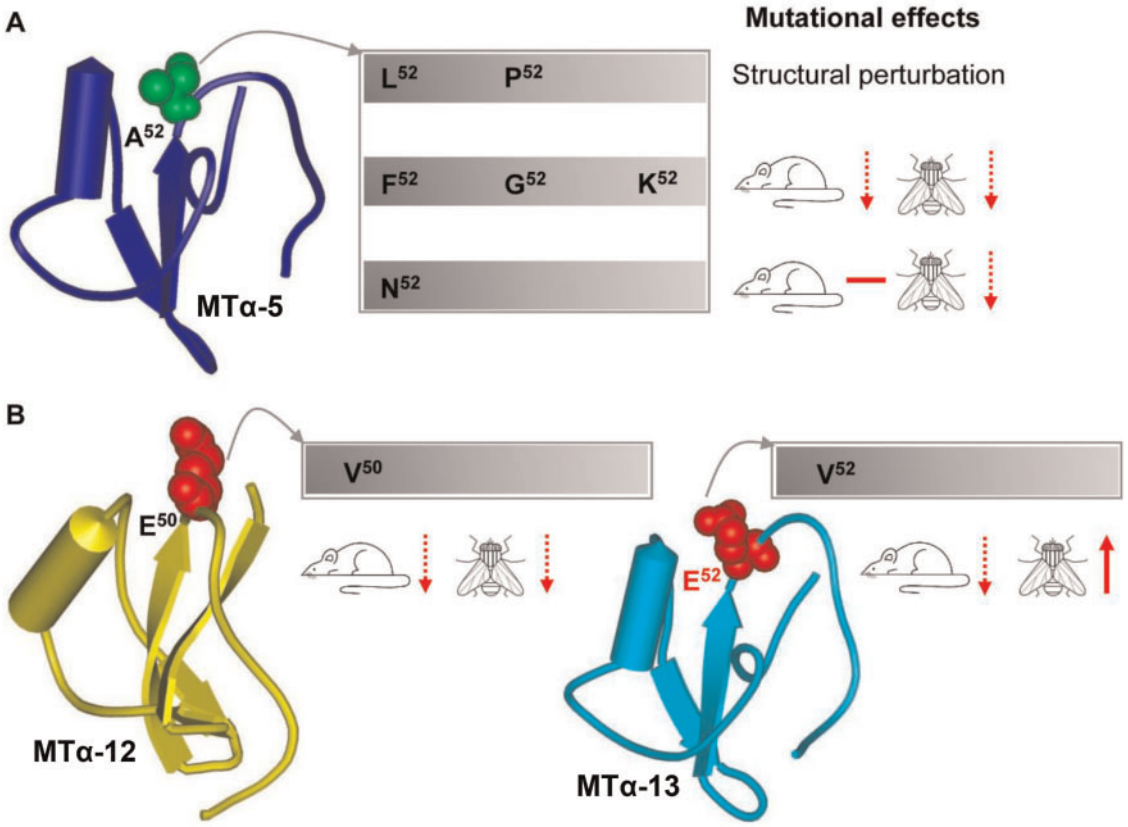

FIG. 5. Site 52 mutation-induced structural and functional changes of toxins. (*A*) Amino acid type-dependency; (*B*) toxin individual-dependency. Residues at site 52 are shown as small spheres and colored according to their side-chain nature (*red*, acidic; *green*, hydrophobic). Mutational effects are indicated by a red horizontal line (almost unchanged), dotted arrows (declined), or the solid arrow (enhanced).

their coevolution is required for longer-range energetic interactions (Lockless and Ranganathan 1999) among toxin sites.

Discussion

To date, no literature is reported with regard to the functional and evolutionary significances of a distant PSS in animal toxins. Here, we show that positive selection at a distal site of the scorpion α-type toxin family can modulate its functional evolution at a bioactive site. In our previous studies, we have observed a weak impact of the site 52 mutation (A52K) on the function of MTα-5 (Zhu et al. 2016). In this work, we report extensive mutational effects of this site on the structures and functions of different ScNaTxs through introduction of more mutations and using more toxins. We found that the site 52 mutation-induced changes in toxin structure and function are amino acid type- and toxin individual-dependent (fig. 5). For MTα-5, the introduction of Pro or Leu resulted in structural perturbation, suggesting their deleterious effect and highlighting a key role of this site in preservation of the global structure of α-ScNaTxs. Mutations by other residues (e.g., Phe and Asn) retained the structure but caused different functional consequences. For example, the toxicity of A52F on both $rNa_v1.1$ and houseflies diminished whereas the insecticidal activity of A52N significantly decreased but its toxicity on $rNa_v1.1$, by and large, remained. In addition, the stability of A52G increased but its toxicity decreased, suggesting a stability-function tradeoff. For MTα-12 and MTα-13, although the same mutation caused the activity decline on $rNa_v1.1$, an opposite effect was observed for their insecticidal activity. This difference might arise from epistatic interactions (Parera and Martinez 2014) between site 52 and other sites in different genetic contexts.

Especially absorbing in this work is the findings that the mutation of MTα-5 site 52 from Ala to Asn caused a decreased insecticidal activity but at the same time maintained the mammalian toxicity and the mutation of MTα-13 at the same site from Glu to Val caused a decreased mammalian toxicity but an enhanced insecticidal activity. This clearly shows a key role of this site in regulating the species selectivity of the toxins, as previously observed in two natural variants (Zhu et al. 2012). The discovery of the amino acid type- and toxin individual-dependency in determining the mutational effects of a distant site indicates more possibilities for toxin retrofitting. Considering that the number of nonbioactive sites occupies an absolute majority in a protein, it is practically impossible to analyze mutational effects of all the sites and to study their interactions with bioactive sites. Alternatively, statistical methods, such positive selection and SCA used here, can narrow down possibilities to be tested in lab to quickly find such distant sites.

Based on the observation that positively selected bioactive sites in ScNaTxs directly bind to the highly variable channel residues, it has been proposed that an essential selection

agent for these toxins comes from the evolutionary variability of the toxin-bound regions at Na_vs from both predators and prey of scorpions (Zhang et al. 2015; Zhu et al. 2016). However, site 52 is not a site directly involved in toxin–channel interactions, which may exclude the involvement of channels in providing a force driving the evolution of this distant site. The presence of allosteric communication between this site and the bioactive surface undergoing positive selection suggests that the evolution of site 52 is a consequence of dealing with the changes of the positively selected bioactive sites. Therefore, compensatory mutation-mediated coevolution (Pál et al. 2006) within one toxin molecule could be a more plausible explanation. In our opinion, when one amino acid substitution occurs at a toxin's bioactive site for adapting to the channel variation, its distant site 52 must follow change for maintaining an active conformation of the mutated bioactive site for channel binding. Their correlation is further revealed by our SCA, which identified two PSSs (I40 and A52) belonging to the same sector (i.e., a group of coevolving residues), suggesting that natural selection can drive accelerated substitutions of multiple sites simultaneously.

In summary, our results demonstrate that a distant PSS can contribute to its roles in toxins' structure and function via allosteric communication with the bioactive sites. Although scorpion toxins and enzymes are by no means comparable, we observed a strategic commonality in their functional evolution, both via interactions between bioactive and distant nonbioactive sites. These findings would promote more studies on the evolution of protein inner life (functional interactions among sites within a molecule) (Wagner 2012) for better understanding of their functional innovation. Our work may also be valuable in stimulating exploration of positively selected distant sites for engineering design of new proteins with improved functional properties.

Materials and Methods

Site-Directed Mutagenesis

Inverse PCR was employed to prepare all the mutants described here, in which pET-28a-MTα-5, -MTα-12, and -MTα-13 previously constructed (Zhu et al. 2013, 2016) were used as templates. To generate a mutant, two back-to-back primers were synthesized (see supplementary table S2, Supplementary Material online) and modified by 5′-end phosphorylation with T4 polynucleotide kinase (TOYOBO, Osaka) and ATP (Takara, Dalian). Template plasmids were amplified by two back-to-back primers and ExTaq DNA polymerase (Takara, Dalian). Ends of linear PCR products were polished by Pfu polymerase (newProbe, Beijing) and circularized by T4 DNA ligase (Takara, Dalian). Circularized products were transformed into *Escherichia coli* DH5α competent cells. Positive clones were confirmed by DNA sequencing.

Preparation and Identification of Recombinant Products

Expression, refolding, and purification of recombinant MTα-5, MTα-12, MTα-13 and their mutants were performed according to the method previously described in Turkov et al. (1997) and Zhu et al. (2013). In brief, a recombinant plasmid was transformed into *E. coli* BL21 (DE3) pLysS Cells. The induction was initiated by adding 0.5 mM IPTG at an OD_{600} of 0.2–0.3. Cells were harvested 4 h later by centrifugation and pellets were resuspended in Resuspension Buffer (0.1 M Tris-HCl, pH 8.5; 0.1 M NaCl) for cell disruption by sonication. All recombinant proteins studied here were accumulated as inclusion bodies and their renaturation was performed according to the procedures previously described in Zhu et al. (2013). Briefly, inclusion bodies were first washed with the isolation buffer (2 M urea and 2% Triton X-100 in the resuspension buffer) and then solubilized in the denaturation solution (6 M guanidine-HCl, 0.1 M Tris-HCl [pH 8.5], 1 mM EDTA, and 30 mM β-mercaptoethanol). Renaturation was initiated by 20-fold dilution in the refolding buffer (0.2 M ammonium acetate, pH 9.0) at room temperature for 36 h and the precipitates were removed by filtering. Refolded proteins were recovered by centrifugation after salting out by 80% saturation of solid ammonium sulfate. The precipitates were collected and dissolved in water. Recombinant peptides were further purified twice by RP-HPLC and effluent was monitored by measuring the absorbance at 225 nm. Fractions eluted at the corresponding retention time were collected. Molecular masses of mutants were determined by MALDA-TOF mass spectra and secondary structure features were studied by circular dichroism spectroscopy analysis on a Chirascan Plus spectropolarimeter (UK). Peptides were dissolved in 5 mM phosphate buffer to a final concentration of 0.1–0.3 mg/ml.

Electrophysiological Recordings

For expression of $rNa_v1.1$ in *Xenopus* oocytes, its linearized plasmid was transcribed by mMESSAGE-mMACHINE T7 transcription kits (Ambion, USA). The harvesting of stages V and VI oocytes from anesthetized female *Xenopus laevis* was previously described in Zhu et al. (2013). Oocytes were injected with Na_v cRNA by a micro-injector (Nanoliter 2000, WPI) and then incubated in ND96 solution with 50 mg/L gentamycin sulfate at 16 °C for 1–5 days. Two-electrode voltage-clamp recordings were performed at room temperature with an Oocyte Clamp Amplifier (OC-725C, Harvard Apparatus Company) controlled by a data acquisition system (Digidata 1440A, Axon CNS) dominated by pCLAMP 10.2. Leak subtraction was performed by a $-P/4$ protocol. Resistances of both electrodes were kept at 0.5–1.5 MΩ. The elicited currents were filtered at 1 kHz and sampled at 20 kHz with a four-pole low-pass Bessel filter. The maximum efficacy (slowing down of inactivation) represents the maximum α-effect at a given concentration which reflected by the ratio of the steady-state current measured at 30 ms (I_{30ms}) after depolarization to the control peak current (I_{peak}). The dose–response effect of the toxins (slowing down of inactivation) was calculated by plotting the ratio of the I_{30ms} to the I_{peak} as a function of toxin concentration. The data were fitted with the Hill equation. All data were presented as mean ± standard error (SE) of at least three-independent experiments. All data were analyzed by SigmaPlot 11.0.

Insect Toxicity Assays

Insecticidal activity was measured by injecting peptides dissolved in insect saline (200 mM NaCl, 3.1 mM KCl, 5.4 mM $CaCl_2$, 5 mM $MgCl_2$, 2 mM $NaHCO_3$, 0.1 mM NaH_2PO_4, pH 7.2) into housefly adults, as previously described in Wang et al. (1999) and Tedford et al. (2001). Experiments were performed in triplicates with a cohort of 10 flies for each concentration of toxins. Control files were injected with 1 μl of insect saline. All flies were kept at 4 °C during injections and subsequently transferred to room temperature. The median lethal dose (LD_{50}) was calculated by SigmaPlot 11.0 which fits the dose–response data with the Hill equation, $y = (a–b)/[1 + (x/LD_{50})^n] + b$, where y represents the percentage of deaths in the sample population at 12 h postinjection, x represents the toxin dose in pmol/g, n is the variable slope factor, a is maximum response, and b is the minimum response.

Molecular Dynamics (MD) Simulations

The experimental structure of MTα-5 (PDB entry 2LKB) was used to construct the models of mutants with the Swiss PDB Viewer (http://spdbv.vital.it.ch/, last accessed April 20, 2016). MD simulations (40 ns) were performed with GROMACS 5.1.4 with the OPLS-(AA)/L all-atom force field (Van Der Spoel et al. 2005). Each peptide was immersed in a cubic box extending to at least 1.0 nm from the protein surface and solvated with explicit SPC water molecular. Sodium and chloride ions were added to neutralize the total charge of the system, which was then simulated with periodic boundary conditions. Each solved structure was energy minimized for 5,000 steps of steepest descent minimization termination with a maximum force $<$1,000 KJ/mol/nm. The system was then subject to equilibration phase of 100 ps NVT (Number of particles, volume, and temperature) equilibration and 100 ps NPT (Number of particles, pressure, and temperature) equilibration. The velocity rescaling method (also called modified Berendsen thermostat) was used to maintain the temperature at 300 K, and the Parrinello–Rahman methods were employed to maintain the pressure at 1 bar. The particle mesh Ewald method was used for long-range electrostatic interactions and the Linear Constraint Solver (LINCS) algorithm constrained all bonds (Hess 2008). Trajectories were saved every 10 ps for analysis. RMSD, RMSF, and the distances of atoms were extracted to compare structural deviations between MTα-5 and its mutants after simulations.

Protein Sector Identification

The *Mesobuthus* ScNaTxs were collected from the GenBank database (https://www.ncbi.nlm.nih.gov/, last accessed August 15, 2017) by using MTα-5 as a query and aligned by ClustalX2 (http://www.clustal.org/Clustal2/, last accessed August 20 2017) with default parameters. The SCA toolbox was obtained from Rama Ranganathan. We modified the MATLAB script to analyze the test cases. The method of identifying protein sectors has been previously described in Halabi et al. (2009). In brief, the toolbox firstly calculated the degree of conservation of each amino acid in the alignment and then generated a positional correlation, a sequence correlation matrix, a spectral decomposition with analysis of the statistically significant eigenvalues and an independent component analysis. Finally, this method identified protein sectors and mapped them onto the primary structure of the toxin based on their position in the structure file.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

We would like to thank Prof. A. L. Goldin for gifting cDNA of $rNa_v1.1$. We are also grateful to Prof. X. Qiu for providing houseflies and Miss L. Jin for drawing the images of rat, scorpion, and housefly. This work was supported by the National Natural Science Foundation of China (31870766 and 31570773) to S. Z. and (31501862) to L. Z.

Author Contributions

S. Z. conceived and designed this study. L. Z. performed all the experiments except CD measurements and preparation of cRNAs of Na_v channels that were done by B. G.; L. Z. and S. Y. commonly performed molecular dynamics simulations and sequence analysis. All the authors jointly wrote the paper.

References

Ahuja LG, Kornev AP, McClendon CL, Veglia G, Taylor SS. 2017. Mutation of a kinase allosteric node uncouples dynamics linked to phosphotransfer. *Proc Natl Acad Sci U S A*. 114(6):931–940.

Bosmans F, Tytgat J. 2007. Voltage-gated sodium channel modulation by scorpion α-toxins. *Toxicon* 49(2):142–158.

Campbell E, Kaltenbach M, Correy GJ, Carr PD, Porebski BT, Livingstone EK, Afriat-Jurnou L, Buckle AM, Weik M, Hollfelder F, et al. 2016. The role of protein dynamics in the evolution of new enzyme function. *Nat Chem Biol*. 12(11):944–950.

Catterall WA. 2000. From ionic currents to molecular mechanisms: the structure and function of voltage-gated sodium channels. *Neuron*. 26(1):13–25.

Cestèle S, Catterall WA. 2000. Molecular mechanisms of neurotoxin action on voltage-gated sodium channels. *Biochimie* 82(9–10):883–892.

Fontecilla-Camps JC, Almassy RJ, Suddath FL, Watt DD, Bugg CE. 1980. Three-dimensional structure of a protein from scorpion venom: a new structural class of neurotoxins. *Proc Natl Acad Sci U S A*. 77(11):6496–6500.

Gao B, Zhu S. 2018. *Mesobuthus* venom-derived antimicrobial peptides possess intrinsic multifunctionality and differential potential as drugs. *Front Microbiol*. 9:320.

Gordon D, Gurevitz M. 2003. The selectivity of scorpion α-toxins for sodium channel subtypes is determined by subtle variations at the interacting surface. *Toxicon* 41(2):125–128.

Gordon D, Karbat I, Ilan N, Cohen L, Kahn R, Gilles N, Dong K, Stuhmer W, Tytgat J, Gurevitz M. 2007. The differential preference of scorpion α-toxins for insect or mammalian sodium channels: implications for improved insect control. *Toxicon* 49(4):452–472.

Greenfield NJ. 2006. Using circular dichroism spectra to estimate protein secondary structure. *Nat Protoc*. 1(6):2876–2890.

Halabi N, Rivoire O, Leibler S, Ranganathan R. 2009. Protein sectors: evolutionary units of three-dimensional structure. *Cell* 138(4):774–786.

Hess B. 2008. LINCS: a parallel linear constraint solver for molecular simulation. *J Chem Theory Comput*. 4(1):116–122.

Hughes AL, Nei M. 1988. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* 335(6186):167–170.

Jiménez-Osés G, Osuna S, Gao X, Sawaya MR, Gilson L, Collier SJ, Huisman GW, Yeates TO, Tang Y, Houk KN. 2014. The role of distant mutations and allosteric regulation on LovD active site dynamics. *Nat Chem Biol*. 10(6):431–436.

Juarez P, Comas I, Gonzalez-Candelas F, Calvete JJ. 2008. Evolution of snake venom disintegrins by positive Darwinian selection. *Mol Biol Evol*. 25(11):2391–2407.

Kahn R, Karbat I, Ilan N, Cohen L, Sokolov S, Catterall WA, Gordon D, Gurevitz M. 2009. Molecular requirements for recognition of brain voltage-gated sodium channels by scorpion α-toxins. *J Biol Chem*. 284(31):20684–20691.

Kapralov MV, Filatov DA. 2007. Widespread positive selection in the photosynthetic Rubisco enzyme. *BMC Evol Biol*. 7:73.

Karbat I, Frolow F, Froy O, Gilles N, Cohen L, Turkov M, Gordon D, Gurevitz M. 2004. Molecular basis of the high insecticidal potency of scorpion α-toxins. *J Biol Chem*. 279(30):31679–31686.

Karbat I, Kahn R, Cohen L, Ilan N, Gilles N, Corzo G, Froy O, Gur M, Albrecht G, Heinemann SH, et al. 2007. The unique pharmacology of the scorpion α-like toxin Lqh3 is associated with its flexible C-tail. *FEBS J*. 274(8):1918–1931.

Kozminsky-Atias A, Zilberberg N. 2012. Molding the business end of neurotoxins by diversifying evolution. *FASEB J*. 26(2):576–586.

Lockless SW, Ranganathan R. 1999. Evolutionarily conserved pathways of energetic connectivity in protein families. *Science* 286(5438):295–299.

Pál C, Papp B, Lercher MJ. 2006. An integrated view of protein evolution. *Nat Rev Genet*. 7(5):337–348.

Parera M, Martinez MA. 2014. Strong epistatic interactions within a single protein. *Mol Biol Evol*. 31(6):1546–1553.

Polis GA. 1990. The biology of scorpions. Stanford (CA): Stanford University Press.

Pontremoli C, Forni D, Cagliani R, Filippi G, De Gioia L, Pozzoli U, Clerici M, Sironi M. 2016. Positive selection drives evolution at the host-filovirus interaction surface. *Mol Biol Evol*. 33(11): 2836–2847.

Possani LD, Becerril B, Delepierre M, Tytgat J. 1999. Scorpion toxins specific for Na^+ channels. *Eur J Biochem*. 264(2):287–300.

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.

Ren LL, Liu YJ, Liu HJ, Qian TT, Qi LW, Wang XR, Zeng QY. 2014. Subcellular relocalization and positive selection play key roles in the retention of duplicate genes of populus class III peroxidase family. *Plant Cell* 26(6):2404–2419.

Sawyer SL, Wu LI, Emerman M, Malik HS. 2005. Positive selection of primate TRIM5α identifies a critical species-specific retroviral restriction domain. *Proc Natl Acad Sci U S A*. 102(8):2832–2837.

Shen H, Zhou Q, Pan X, Li Z, Wu J, Yan N. 2017. Structure of a eukaryotic voltage-gated sodium channel at near-atomic resolution. *Science* 355(6328):eaal4326.

Sironi M, Cagliani R, Forni D, Clerici M. 2015. Evolutionary insights into host-pathogen interactions from mammalian sequence data. *Nat Rev Genet*. 16(4):224–236.

Sunagar K, Jackson TN, Undheim EA, Ali SA, Antunes A, Fry BG. 2013. Three-fingered RAVERs: rapid accumulation of variations in exposed residues of snake venom toxins. *Toxins* 5(11):2172–2208.

Tedford HW, Fletcher JI, King GF. 2001. Functional significance of the beta hairpin in the insecticidal neurotoxin omega-atracotoxin-Hv1a. *J Biol Chem*. 276(28):26568–26576.

Tian C, Yuan Y, Zhu S. 2008. Positively selected sites of scorpion depressant toxins: possible roles in toxin functional divergence. *Toxicon* 51(4):555–562.

Turkov M, Rashi S, Noam Z, Gordon D, Ben Khalifa R, Stankiewicz M, Pelhate M, Gurevitz M. 1997. In vitro folding and functional analysis of an anti-insect selective scorpion depressant neurotoxin produced in *Escherichia coli*. *Protein Expr Purif*. 10(1):123–131.

Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ. 2005. GROMACS: fast, flexible, and free. *J Comput Chem*. 26(16):1701–1718.

Vonk FJ, Casewell NR, Henkel CV, Heimberg AM, Jansen HJ, McCleary RJ, Kerkkamp HM, Vos RA, Guerreiro I, Calvete JJ, et al. 2013. The king cobra genome reveals dynamic gene evolution and adaptation in the snake venom system. *Proc Natl Acad Sci U S A*. 110(51): 20651–20656.

Wagner GP. 2012. Genetics: the inner life of proteins. *Nature* 490(7421):493–494.

Wang CG, Gilles N, Hamon A, Le Gall F, Stankiewicz M, Pelhate M, Xiong YM, Wang DC, Chi CW. 2003. Exploration of the functional site of a scorpion α-like toxin by site-directed mutagenesis. *Biochemistry* 42(16):4699–4708.

Wang X, Jimenez-Vargas JM, Xu C, Possani LD, Zhu S. 2012. Positive selection-guided mutational analysis revealing two key functional sites of scorpion ERG K^+ channel toxins. *Biochem Biophys Res Commun*. 429(1–2):111–116.

Wang X, Smith R, Fletcher JI, Wilson H, Wood CJ, Howden ME, King GF. 1999. Structure–function studies of ω-atracotoxin, a potent antagonist of insect voltage-gated calcium channels. *Eur J Biochem*. 264(2):488–494.

Wang J, Yarov-Yarovoy V, Kahn R, Gordon D, Gurevitz M, Scheuer T, Catterall WA. 2011. Mapping the receptor site for alpha-scorpion toxins on a Na^+ channel voltage sensor. *Proc Natl Acad Sci U S A*. 108(37):15426–15431.

Weinberger H, Moran Y, Gordon D, Turkov M, Kahn R, Gurevitz M. 2010. Positions under positive selection-key for selectivity and potency of scorpion α-toxins. *Mol Biol Evol*. 27(5):1025–1034.

Whitney DS, Volkman BF, Prehoda KE. 2016. Evolution of a protein interaction domain family by tuning conformational flexibility. *J Am Chem Soc*. 138(46):15150–15156.

Wong ES, Belov K. 2012. Venom evolution through gene duplications. *Gene* 496(1):1–7.

Yang Z. 2002. Inference of selection from multiple species alignments. *Curr Opin Genet Dev*. 12(6):688–694.

Yang Z, Swanson WJ. 2002. Codon-substitution models to detect adaptive evolution that account for heterogeneous selective pressures among site classes. *Mol Biol Evol*. 19(1):49–57.

Yuan S, Gao B, Zhu S. 2017. Molecular dynamics simulation reveals specific interaction sites between scorpion toxins and $K_v1.2$ channel: implications for design of highly selective drugs. *Toxins* 9(11):354.

Zhang S, Gao B, Zhu S. 2015. Target-driven evolution of scorpion toxins. *Sci Rep*. 5:14973.

Zhu S, Bosmans F, Tytgat J. 2004. Adaptive evolution of scorpion sodium channel toxins. *J Mol Evol*. 58(2):145–153.

Zhu S, Gao B, Tytgat J. 2005. Phylogenetic distribution, functional epitopes and evolution of the CSαβ superfamily. *Cell Mol Life Sci*. 62(19–20):2257–2269.

Zhu S, Peigneur S, Gao B, Lu X, Cao C, Tytgat J. 2012. Evolutionary diversification of *Mesobuthus* α-scorpion toxins affecting sodium channels. *Mol Cell Proteomics* 11(1):M111.012054.

Zhu L, Peigneur S, Gao B, Tytgat J, Zhu S. 2013. Two recombinant α-like scorpion toxins from *Mesobuthus eupeus* with differential affinity toward insect and mammalian Na^+ channels. *Biochimie* 95(9):1732–1740.

Zhu L, Peigneur S, Gao B, Zhang S, Tytgat J, Zhu S. 2016. Target-driven positive selection at hot spots of scorpion toxins uncovers their potential in design of insecticides. *Mol Biol Evol*. 33(8):1907–1920.

Zupunski V, Kordis D. 2016. Strong and widespread action of site-specific positive selection in the snake venom Kunitz/BPTI protein family. *Sci Rep*. 6:37054.