Peer-reviewed version available at Int. J. Mol. Sci. 2019, 20; doi:10.3390/ijms20061261

Research Article

Effects of Selective Substitution of Cysteine Residues on the Conformational Properties of Chlorotoxin Explored by Molecular Dynamics Simulations

Andrew J. Gregory [1], Leah Voit-Ostricki [2], Sándor Lovas [3], and Charles R. Watts [4,*]

[1] Department of Neurosurgery, Mayo Clinic Health System-Franciscan Healthcare in La Crosse, Wisconsin; ajgregory2@wisc.edu.
[2] Department of Neurosurgery, Mayo Clinic Health System-Franciscan Healthcare in La Crosse, Wisconsin; voit-ost.leah@uwlax.edu.
[3] Department of Biomedical Sciences, Creighton University, Omaha, Nebraska; SandorLovas@creighton.edu.
[4] Department of Neurosurgery, Mayo Clinic Health System-Franciscan Healthcare in La Crosse, Wisconsin and Department of Neurologic Surgery, Mayo Clinic, Rochester, Minnesota; charles.watts@parknicollet.com

* Correspondence: charles.watts@parknicollet.com; Tel.: +1-952-993-3200

Abstract: Chlorotoxin (CTX) is a 36–amino acid peptide with 8 Cys residues that forms four Cys-Cys bonds. It has high affinity for the glioma-specific chloride channel and matrix metalloprotease-2. Structural and binding properties of CTX analogs with various Cys residue substitutions with L-α-aminobutyric acid (Abu) have been previously reported. Using 4.2 μs molecular dynamics, we compared the conformational and essential space sampling of CTX and analogs [Abu/Ser^{2,19}]CTX, [Abu/Ser^{5,28}]CTX, [Abu/Ser^{16,33}]CTX, [Abu/Ser^{20,35}]CTX, and [Abu/Ser^{2,5,16,19,20,28,33,35}]CTX. The native and substituted peptides maintained a high degree of α-helix propensity from residues 8 through 21, with the exception of [Ser^{5,28}]CTX and [Abu^{16,33}]CTX. In agreement with previous circular dichroism spectropolarimetry results, the C-terminal β-sheet content varied less from residues 25 through 29 and 32 through 36 and was well conserved in all analogs, except: [Abu^{16,33}]CTX, [Ser^{20,35}]CTX, [Abu^{2,5,16,19,20,28,33,35}]CTX, and [Ser^{2,5,16,19,20,28,33,35}]CTX. The Cys^{16}-Cys^{33} and Cys^{20}-Cys^{35} Cys-Cys bonds appear to be required to maintain the αβ-motif of CTX. Their selective substitution with Ser^{16,33} however, may mitigate the destabilizing effect of Cys^{16}-Cys^{33} substitution by the formation of an inter residue H-bond from OγH of Ser^{16} to OγH of Ser^{33} bridged by a water molecule. All peptides shared considerable sampled conformational space, which explains the retained receptor binding of the nonnative analogs.

Keywords: αβ motif; Abu; chlorotoxin; Cys; disulfide bond; insectotoxin; isosteric substitution; L-α-aminobutyric acid; molecular dynamics; Ser

1. Introduction

Chlorotoxin (CTX) is a peptide toxin in the venom of the deathstalker scorpion (*Leiurus quinquestriatus*) [1,2]. The peptide binds with high affinity to chloride channels, causing paralysis in invertebrates, but it has minimal to no effect on vertebrates or mammals [2]. Because of its high affinity and selectivity, CTX was originally used as a tool to characterize the function of chloride channels in electrophysiology experiments. Pharmacologic interest in the peptide increased after CTX was shown to bind with high affinity to glioma-specific chloride channels on the surface of high-grade intrinsic brain tumors (glioblastoma multiforme), as well as other tumors of neuroectodermal embryologic origin [3-5]. CTX has been considered a potential lead for the

development of novel therapeutic agents, imaging adjuncts, and intraoperative optical imaging "tumor dyes/paints" [6-10].

CTX consists of 36 amino acids with 8 Cys residues at positions 2, 5, 16, 19, 20, 28, 33, and 35 (I-VIII for homology modeling) that form 4 Cys-Cys bonds: 2-19, 5-28, 16-33, and 20-35 (I-IV, II-VI, III-VII, and V-VIII) (Figure 1). The resulting secondary and tertiary structure are known as an αβ (βαββ) motif, a folding scaffold common to insectotoxins, insect defensins, plant γ-thionins, and inhibitory cystine knot peptides [11-20]. Multisequence alignment of 20 scorpion toxin–derived peptides shows that they share 49% to 88% sequence similarity [11-15]. Overlays of the 3-dimensional structures of several of these peptides have root-mean-square deviation (RMSD) less than 0.1 nm, which indicates the stability of the αβ motif [21]. The observed differences in channel blocking and molecular selectivity of the toxins are due to subtle differences in residue charge and sidechain conformation that affect surface electrostatic charge distribution and complementarity [21].

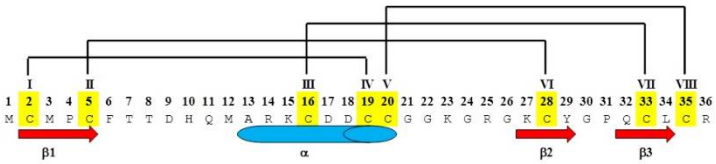


Figure 1. Amino acid sequence of chlorotoxin, demonstrating the Cys-Cys bond topology between Cys^{2}-Cys^{19}, Cys^{5}-Cys^{28}, Cys^{16}-Cys^{33}, and Cys^{20}-Cys^{35}, shown as yellow with black brackets representing disulfide bonds, and homology numbering shown as Roman numerals. The conserved secondary structural motifs of an α-helix from Ala^{13} to Cys^{20}, a β-sheet from Cys^{2} to Cys^{5}, and 2 antiparallel β-sheets from Lys^{27} to Tyr^{29} and Lys^{32} to Leu^{34} are shown as a blue cylinder and red arrows, respectively.

The locations of the Cys-Cys bonds within this family of peptides are important for maintaining secondary and tertiary structural features and providing a degree of protease resistance [16,22-26]. Despite the evolutionarily conserved Cys-Cys bonds in this family of insectotoxins, the ability to selectively remove some of these bonds is well documented [24-28]. Selective substitution of Cys residues with L-α-aminobutyric acid (Abu) in the charybdotoxin and leiurotoxin I peptides demonstrated successful substitution of the Cys^{II}-Cys^{VI} disulfide bond without significant effects on oxidative folding [24-28]. Previous work by Ojeda *et al.* demonstrated that selective substitution of the Cys^{2}-Cys^{19} or Cys^{5}-Cys^{28} bonds of CTX with Abu residues had little effect on peptide conformation, whereas the Cys^{16}-Cys^{33} and Cys^{20}-Cys^{35} disulfide bonds were critical to the process of oxidative folding and obtaining nativelike peptides [29]. The authors also demonstrated that complete substitution of all disulfide bonds resulted in a peptide that maintained its biological activity, but with complete disruption of its secondary and tertiary conformation and significantly increased susceptibility to serum proteases [29].

The tertiary structure of CTX in water (Protein Data Bank ID: 1CHL) has been determined using high-field ^{1}H nuclear magnetic resonance spectroscopy (NMR) (Figure 2) [30]. The presence of the N-terminal β-sheet from Cys^{2} to Cys^{5} is variable and dependent on the algorithm used to determine the structure [31,32]. In the current study, we used molecular dynamics (MD) simulations of the structures of CTX and Abu- and Ser-substituted CTX analogs to investigate the role of the Cys-Cys disulfide bonds in stabilizing the αβ motif. Emphasis was placed on investigating the role of hydrophobic (Abu) versus hydrophilic (Ser) isosteric substitutions, the distribution of the hydration shell around each respective residue substitution, and subsequent changes in secondary and tertiary conformation.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 December 2018 doi:10.20944/preprints201812.0214.v1

Peer-reviewed version available at Int. J. Mol. Sci. 2019, 20; doi:10.3390/ijms20061261

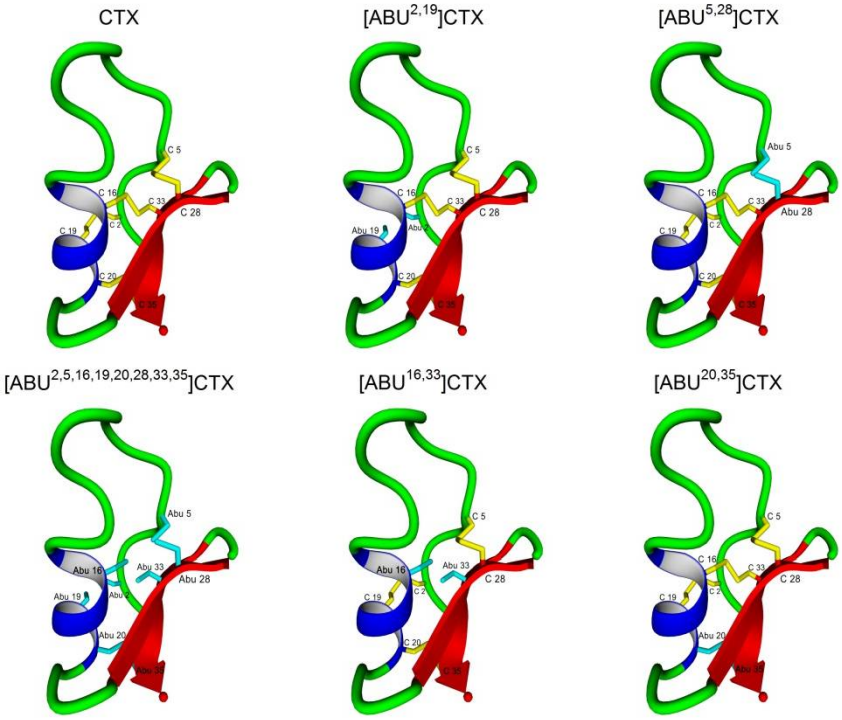


Figure 2. Solution conformation of chlorotoxin (CTX) as determined by ^{1}H nuclear magnetic resonance spectroscopy (NMR) (Protein Data Bank ID: 1CHL) demonstrating the α-helix (Ala^{13} to Cys^{20}) and 2 antiparallel β-sheets (Lys^{27} to Tyr^{29} and Lys^{32} to Leu^{34}) with the 4 Cys-Cys bonds: Cys^{2}-Cys^{19}, Cys^{5}-Cys^{28}, Cys^{16}-Cys^{33}, and Cys^{20}-Cys^{35}, shown in yellow. The N-terminal β-sheet from Cys^{2} to Cys^{5} is not present in the lowest energy NMR conformation. Location of the Cys-Cys substitutions with L-α-aminobutyric acid (Abu) or Ser (not shown) are in turquoise with the associated naming conventions: [Abu/Ser^{2,19}]CTX, [Abu/Ser^{5,28}]CTX, [Abu/Ser^{16,33}]CTX, [Abu/Ser^{20,35}]CTX, and [Abu/Ser^{2,5,16,19,20,28,33,35}]CTX.

2. Results

Details outlining the parameterization of the Abu residue are provided in Appendix A.

2.1 Biophysical Properties

The mean ± SD: R_g, $RMSD_{CTX}$, $RMSD_{AVG}$, fraction of sampled secondary structure (α-helix, β-sheet, and β-bend/turn), SASA, $SASA_{Hydrophobic}$, and $SASA_{Hydrophilic}$ values for each peptide are shown in Table 1, along with their tests of data homoscedasticity (Levene's test) and the associated Welch's ANOVA. The post hoc analysis results are shown in Table B1. Selective substitution of the Cys residues with either Ser or Abu caused small increases in the R_g, with the exception of [Ser^{2,19}]CTX, which was similar to the native peptide, and [Ser^{20,35}]CTX, which was more compact. Complete substitution of the Cys residues with either Ser or Abu resulted in significant increases in the R_g, with [Ser^{2,5,16,19,20,28,33,35}]CTX being less compact than [Abu^{2,5,16,19,20,28,33,35}]CTX. There were significant differences in the R_g values of all peptides, with the exception of pairwise comparisons between [Abu^{2,19}]CTX and [Abu^{16,33}]CTX and between [Abu^{5,28}]CTX and [Ser^{16,33}]CTX.

The calculated values of $RMSD_{CTX}$ and $RMSD_{AVG}$ were relatively similar between CTX and all peptide analogs (Table 1), with the exception of [Abu^{2,5,16,19,20,28,33,35}]CTX and [Ser^{2,5,16,19,20,28,33,35}]CTX, which had lower $RMSD_{AVG}$ values. There were significant differences in the $RMSD_{CTX}$ and $RMSD_{AVG}$ values, with the exception of 9 pairwise comparisons for $RMSD_{CTX}$ and 1 pairwise comparison for $RMSD_{AVG}$ (Table B1).

Table 1. Calculated biophysical properties with ANOVA

Peptide	Mean±SD Value								
	R_g, nm	$RMSD_{CTX}$, nm[a]	$RMSD_{AVG}$, nm[b]	α-helix, ρ	β-sheet, ρ	Bend/turn, ρ	SASA, nm²	$SASA_{Hydrophobic}$, nm²	$SASA_{Hydrophilic}$, nm²
CTX	0.88±0.02	0.73±0.02	0.73±0.02	0.31±0.05	0.16±0.02	0.17±0.05	29.4±1.3	13.9±1.1	15.5±1.3
[$Abu^{2,19}$]CTX	0.94± 0.03	0.71±0.02	0.74±0.02	0.39±0.09	0.14±0.05	0.11±0.09	30.2±1.5	11.0±0.8	19.2±1.1
[$Ser^{2,19}$]CTX	0.88±0.02	0.73±0.02	0.73±0.02	0.27±0.04	0.20±0.03	0.26±0.02	27.9±1.0	9.8±0.7	18.1±0.7
[$Abu^{5,28}$]CTX	0.90±0.02	0.73±0.02	0.73±0.02	0.23±0.07	0.16±0.03	0.31±0.04	28.7±1.1	11.2±0.8	17.5±0.8
[$Ser^{5,28}$]CTX	0.96±0.04	0.73±0.02	0.74±0.02	0.34±0.08	0.16±0.08	0.08±0.07	30.9±1.4	12.6±0.8	18.3±1.2
[$Abu^{16,33}$]CTX	0.94±0.04	0.73±0.02	0.76±0.02	0.31±0.07	0.19±0.06	0.13±0.14	30.1±1.5	11.8±0.9	18.3±1.1
[$Ser^{16,33}$]CTX	0.90±0.04	0.73±0.02	0.72±0.02	0.29±0.06	0.13±0.06	0.22±0.08	28.7±1.9	11.4±0.9	17.3±1.3
[$Abu^{20,35}$]CTX	0.91±0.03	0.73±0.02	0.72±0.02	0.32±0.08	0.16±0.04	0.19±0.07	29.8±1.5	10.8±1.1	19.0±1.0
[$Ser^{20,35}$]CTX	0.86±0.01	0.72±0.02	0.72±0.02	0.30±0.06	0.14±0.04	0.27±0.06	27.0±0.9	9.7±0.5	17.3±0.8
[$Abu^{2,5,16,19,20,28,33,35}$]CTX	1.07±0.15	0.72±0.02	0.61±0.02	0.28±0.09	0.03±0.05	0.21±0.14	33.4±3.2	9.9±1.6	23.5±2.1
[$Ser^{2,5,16,19,20,28,33,35}$]CTX	1.18±0.20	0.73±0.02	0.65±0.02	0.30±0.09	0.05±0.06	0.09±0.10	35.7±3.3	10.7±1.5	25.0±2.5
Levene's P Value	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
ANOVA F-Statistic	7,833.2	636.2	15,343	1,371.8	4,616.6	6,629.6	6,077.9	8,666.5	8,877.4
F (df1, df2)	(10, 17,326)	(10, 17,588)	(10, 17,583)	(10, 17,543)	(10, 17,447)	(10, 17,311)	(10, 17,537)	(10, 17,520)	(10, 17,537)
ANOVA P value	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001

Abbreviations: Abu, L-α-aminobutyric acid; ANOVA, analysis of variance; CTX, chlorotoxin; df1, degrees of freedom between the groups; df2, degrees of freedom within the groups; R_g, radius of gyration; RMSD, root-mean-square deviation; SASA, solvent-accessible surface area.

[a] Cα-trace comparison between the peptide and the average conformation of CTX.

[b] Cα-trace comparison between the peptide and the average conformation of itself.

The fraction of α-helix ranged from 0.23 ([Abu^{5,28}]CTX) to 0.39 ([Abu^{2,19}]CTX) (Table 1). The fraction of β-sheet in the single Cys-Cys substituted peptides ranged from 0.13 ([Ser^{16,33}]CTX) to 0.20 ([Ser^{2,19}]CTX). [Abu^{2,5,16,19,20,28,33,35}]CTX and [Ser^{2,5,16,19,20,28,33,35}]CTX demonstrated significant loss of β-sheet conformation, with only 0.03 and 0.05, respectively. There was more variation in the fraction of β-bend/turn, which ranged from 0.08 ([Ser^{5,28}]CTX) to 0.31 ([Abu^{5,28}]CTX). Although the fraction of α-helix and β-bend/turn for [Abu^{2,5,16,19,20,28,33,35}]CTX was 0.28 and 0.21, respectively, the fraction of random coil was 0.48, which is consistent with the previously published circular dichroism (CD) spectra [29].

SASA values for the single Cys-Cys substituted peptides ranged from 27.0 nm^2 ([Ser^{20,35}]CTX) to 30.9 nm^2 ([Ser^{5,28}]CTX) (Table 1). SASA values for [Abu^{2,5,16,19,20,28,33,35}]CTX and [Ser^{2,5,16,19,20,28,33,35}]CTX were higher: 33.4 nm^2 and 35.7 nm^2, respectively. The peptides with the lowest $SASA_{Hydrophobic}$ were [Ser^{20,35}]CTX, [Ser^{2,19}]CTX, and [Abu^{2,5,16,19,20,28,33,35}]CTX, whereas [Ser^{5,28}]CTX and CTX had the highest values. The peptide with the lowest $SASA_{Hydrophilic}$ was CTX, whereas [Abu^{2,5,16,19,20,28,33,35}]CTX and [Ser^{2,5,16,19,20,28,33,35}]CTX had the highest values. The SASA, $SASA_{Hydrophobic}$, and $SASA_{Hydrophilic}$ values showed significant differences, with the exception of several pairwise comparisons (Table B1).

2.2 Secondary Structure

The per-residue fractions of sampled secondary structure for each of the simulated peptides are shown in Figure 3. The secondary structure for CTX was consistent with its ^{1}H NMR solution conformation, having 3 prominent β-turns/bends in the N-terminal domain from residues 6 through 9 and 12 through 13, α-helix from residues 15 through 20, and 2 antiparallel β-sheets from residues 27 through 29 and 32 through 34, with the expected intervening β-turns/bends from residues 21 through 24 and 30 through 31. Except for [Abu^{2,5,16,19,20,28,33,35}]CTX and [Ser^{2,5,16,19,20,28,33,35}]CTX, all other peptides maintained a high degree of secondary structure retention during simulations.

2.3 Structural Flexibility

The per-residue Cα-trace RMSFs comparing each of the simulated peptides with the average conformation of CTX and the average conformation of themselves are shown in Figure 3 and Figure B1. There were only minor differences in the per-residue RMSF between CTX and that of the average conformation. The most flexible regions of CTX correspond to the turn/bend regions from residues 8 through 9, 12 through 13, 21 through 24, and 30 through 31. [Abu^{2,19}]CTX showed increased flexibility from residues 6 through 14, which corresponded to lengthening and destabilization of the α-helix region. [Ser^{2,19}]CTX, [Abu^{5,28}]CTX, and [Ser^{20,35}]CTX were more stable on a per-residue RMSF than CTX and all other peptides. [Ser^{5,28}]CTX had increased flexibility from residues 1 through 14, which corresponded to disruption of the α-helical region and the development of 2 N-terminal antiparallel β-sheets. There was also a slight increase in the flexibility of the C-terminal region from residues 22 through 32, which corresponded to a slight decrease in stability of the 2 antiparallel β-sheets. The degree of Cα-trace RMSF for [Abu^{16,33}]CTX, [Ser^{16,33}]CTX, and [Abu^{20,35}]CTX was similar to CTX. [Abu^{2,5,16,19,20,28,33,35}]CTX and [Ser^{2,5,16,19,20,28,33,35}]CTX demonstrated significant flexibility throughout the peptide, as would be expected based on the significant disruption of peptide secondary structure that occurs with complete removal of all Cys-Cys disulfide bonds.

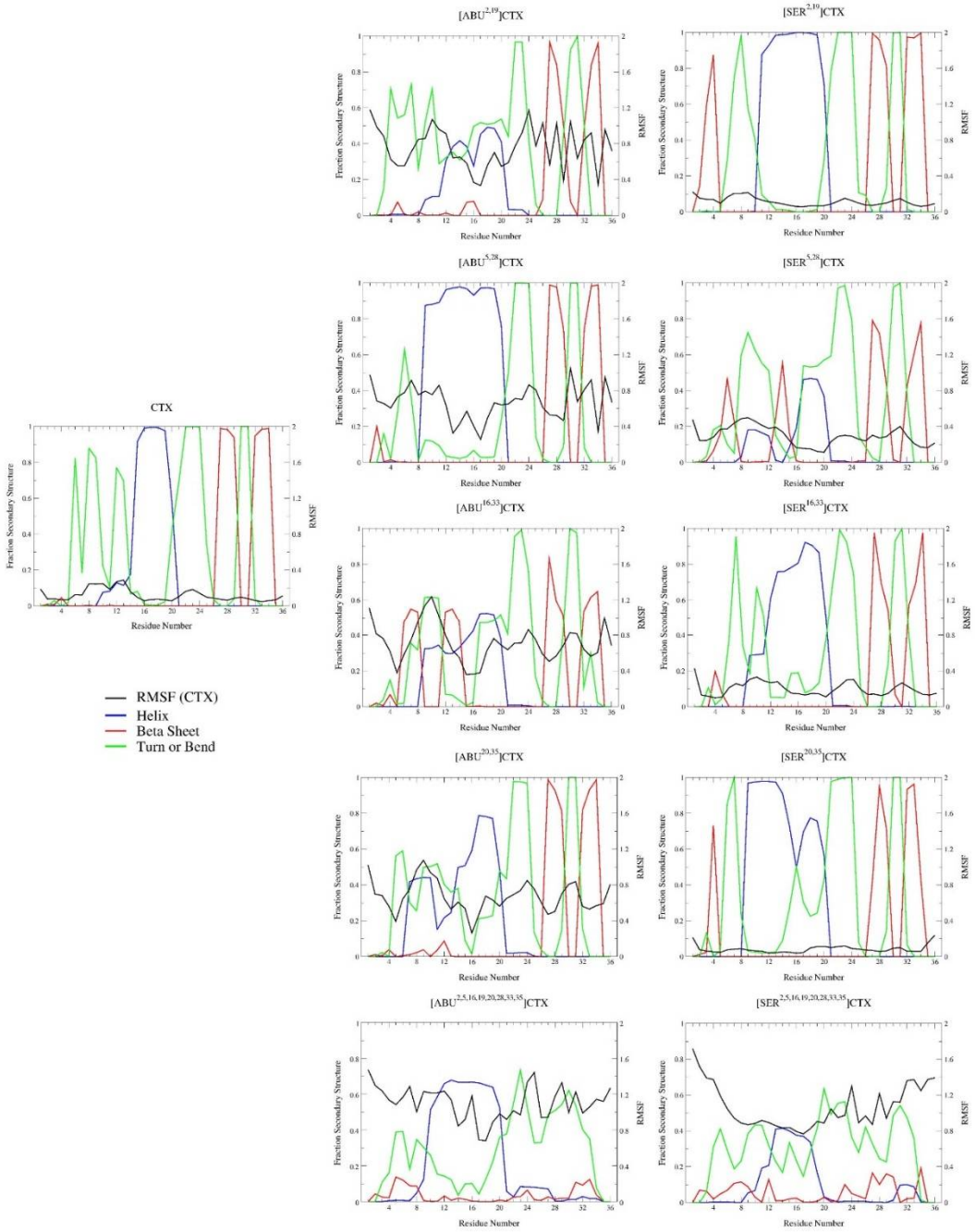


Figure 3. Graphs showing root-mean-square fluctuation from the time averegae conformation of CTX (RMSF, black) and fraction of sampled secondary structure (α-helix, blue; β-sheet, red; and β-turn/bend/loop, green) as a function of residue number for all 11 chlorotoxin (CTX) peptides.

2.4 Interactions

The per-residue H-bond donor/acceptor populations are shown in Figure B2. Residues with the highest fraction of H-bond donor-acceptor participation were predominantly in the N-terminal region, especially residues 8 through 20, which have the greatest propensity for α-helical formation. This pattern held for CTX, [Abu^{2,19}]CTX, [Ser^{2,19}]CTX, [Ser^{5,28}]CTX, [Abu^{16,33}]CTX, and [Abu^{20,35}]CTX. [Abu^{5,28}]CTX had a donor-acceptor peak from residues 23 through 25. This corresponds to elongation of the α-helical region and a β-turn from residues 22 through 24. [Ser^{16,33}]CTX, [Ser^{20,35}]CTX, and [Ser^{2,5,16,19,20,28,33,35}]CTX showed significant disruption of the N-terminus through

the α-helical H-bonding pattern with more uniform values throughout the residues. [Abu^{2,5,16,19,20,28,33,35}]CTX had peaks corresponding to the α-helical region and residues 25 and 29, which represent β-turn formation.

Per-residue relative SASA (rSASA) and ΔrSASA are shown in Figure B3. The most solvent-shielded residues for CTX were the Cys residues 5, 16, 19, 20, 28, and 33. The N-terminal and C-terminal Cys residues were significantly less solvent shielded. The central 16, 20, 28, and 33 residues are the most solvent shielded in the native peptide and all of the single Cys-Cys disulfide bond substituted peptides. In general, Abu substitution tends to favor solvent shielding, whereas the use of Ser tends to favor solvent exposure. Noted exceptions are [Abu^{2,5,16,19,20,28,33,35}]CTX and [Ser^{2,5,16,19,20,28,33,35}]CTX, in which residues at the substituted positions experienced the greatest degree of decreased rSASA.

The BB-BB and SC-SC contact probability maps are shown in Figure 4. For CTX, most of the high-probability contacts are SC-SC, with the expected long-distance contacts between the 2,19; 5,28; 16,33; or 20,35 disulfide bonded residues and the contacts between the N-terminal residues with the α-helix from residues 15 through 20 and the 2 C-terminal β-sheets from residues 27 through 29 and 32 through 34. The only long distance BB-BB contacts are those expected between the 2 antiparallel β-sheets and interactions between residues 4 and 5 with 31 through 33. [Abu^{2,19}]CTX allowed for an increase in both SC-SC and BB-BB contacts. The removal of this Cys-Cys disulfide bond allowed for increased interactions between the N-terminal region, the α-helix, and the C-terminal antiparallel β-sheets. There were also increased BB-BB interactions between the α-helix from residues 12 through 17 and the proximal β-sheet from residues 24 through 28 and between the N-terminal 8 residues and residues 28 through 32. The contact probability map for [Ser^{2,19}]CTX was much more similar to the native peptide for both SC-SC and BB-BB interactions. [Abu^{5,28}]CTX contact probabilities were very similar to those of CTX. [Ser^{5,28}]CTX was more similar to [Abu^{2,19}]CTX except that the interaction between the N-terminal residues and α-helix with the proximal β-sheet were much less pronounced and an interaction was present between the N-terminal residues 1 through 8 and the α-helix from residues 13 through 19. The [Abu^{16,33}]CTX and [Ser^{16,33}]CTX contact probability maps were similar and shared substantial similarity with that of [Ser^{5,28}]CTX. [Abu^{20,35}]CTX and [Ser^{20,35}]CTX were more similar to native CTX. [Abu^{2,5,16,19,20,28,33,35}]CTX showed almost no SC-SC interactions, with low-frequency BB-BB interactions between multiple residues. [Ser^{2,5,16,19,20,28,33,35}]CTX is the exact opposite, with few, if any, BB-BB interactions and sparsely populated SC-SC interactions.

Calculated CβHβ-to-CβHβ center-of-mass distances and probability of proximity within 0.26 nm are shown in Table 2, along with their tests of data homoscedasticity (Levene's test) and the associated Welch's ANOVA. The post hoc analysis results are shown in Table B3. The removal of a Cys-Cys disulfide bond in most of the substituted peptides resulted in significant increases in the CβHβ-to-CβHβ center-of-mass distances and decreases in the probabilities of contact, with the exception of [Ser^{16,33}]CTX and [Abu^{20,35}]CTX. The post hoc analysis revealed only minor variations in interaction distances, other than as noted above.

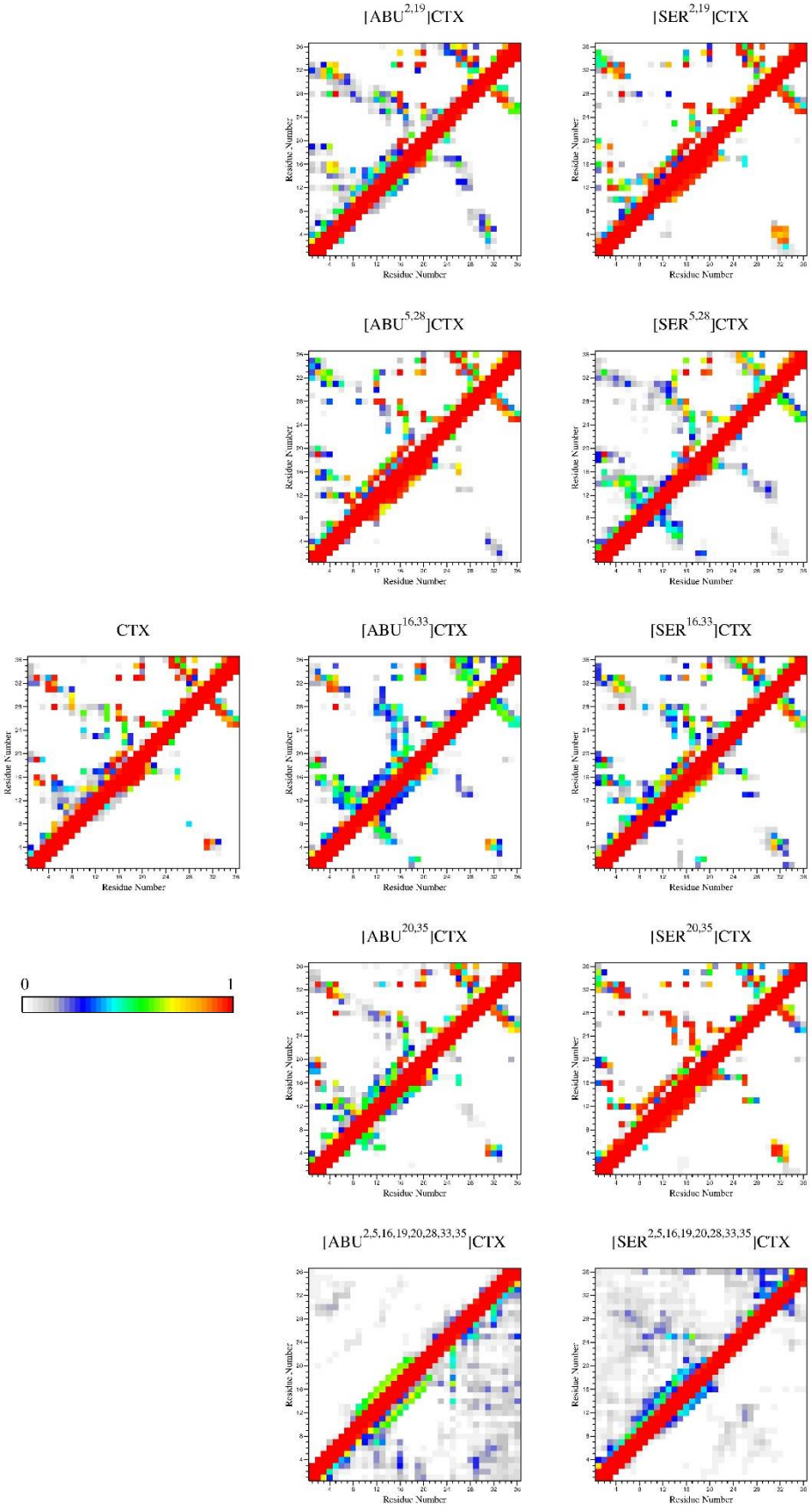


Figure 4. Contact probability maps for backbone-backbone contacts (below the diagonal) and sidechain-sidechain contacts (above the diagonal) as a function of residue number for all 11 chlorotoxin (CTX) peptides. Color scale shows the relative probability of contact.

Table 2. Calculated CβHβ-to-CβHβ center-of-mass distances and probability of proximity ≤0.26 nm for native CTX and Abu- and Ser- substituted peptides[a,b]

Peptide	Mean±SD, nm (Probability of Contact)			
	2-19	5-28	16-33	20-35
CTX	0.24±0.01 (0.98)	0.21±0.01 (0.99)	0.26±0.01 (0.71)	0.21±0.01 (1.00)
[Abu^{2,19}]CTX	0.45±0.08 (0.02)	0.22±0.02 (1.00)	0.21±0.01 (1.00)	0.21±0.01 (1.00)
[Ser^{2,19}]CTX	0.43±0.04 (0.00)	0.22±0.02 (1.00)	0.21±0.01 (1.00)	0.21±0.01 (1.00)
[Abu^{5,28}]CTX	0.22±0.02 (0.97)	0.49±0.08 (0.00)	0.24±0.02 (0.84)	0.21±0.01 (1.00)
[Ser^{5,28}]CTX	0.23±0.02 (0.89)	0.46±0.11 (0.01)	0.21±0.02 (0.98)	0.21±0.01 (1.00)
[Abu^{16,33}]CTX	0.24±0.02 (0.67)	0.22±0.01 (1.00)	0.33±0.07 (0.13)	0.21±0.01 (1.00)
[Ser^{16,33}]CTX	0.24±0.02 (0.77)	0.22±0.01 (0.99)	0.26±0.05 (0.70)	0.21±0.01 (1.00)
[Abu^{20,35}]CTX	0.22±0.02 (0.99)	0.22±0.01 (1.00)	0.24±0.02 (0.80)	0.22±0.03 (0.90)
[Ser^{20,35}]CTX	0.23±0.01 (0.99)	0.21±0.01 (1.00)	0.23±0.01 (1.00)	0.30±0.08 (0.39)
[Abu^{2,5,16,19,20,28,33,35}]CTX	0.62±0.28 (0.04)	0.65±0.22 (0.01)	0.55±0.18 (0.01)	0.62±0.19 (0.02)
[Ser^{2,5,16,19,20,28,33,35}]CTX	0.78±0.30 (0.01)	0.69±0.28 (0.05)	0.70±0.21 (0.00)	0.75±0.25 (0.00)
Levene's *P* value	<.001	<.001	<.001	<.001
ANOVA F-Statistic	14,501	9,781.4	16,558	4,418.3
F (df1, df2)	(10, 17,491)	(10, 17,445)	(10, 17,263)	(10, 17,498)
ANOVA *p* value	<.001	<.001	<.001	<.001

Abbreviations: Abu, L-α-aminobutyric acid; ANOVA, analysis of variance; CTX, chlorotoxin; df1, degrees of freedom between the groups; df2, degrees of freedom within the groups.

[a] The mean±SD calculated CβHβ-to-CβHβ center-of-mass distance for the (Acetyl-Ala-Cys-Ala-NH_2)$_2$ dipeptide, as described in Appendix A, is 0.23±0.03 nm.

[b] The CβHβ-to-CβHβ center-of-mass distance representing the 95% 1-sided CI for the (Acetyl-Ala-Cys-Ala-NH_2)$_2$ dipeptide, as described in Appendix A, is ≤0.26 nm.

The probabilities of finding water molecules within 0.5 nm of the terminal γ-sidechain atoms are shown in Table 3. The terminal Sγ of CTX remained relatively solvent shielded with a low probability of adjacent water molecules, regardless of its position within the native peptide. Substitution of a single Cys-Cys disulfide bond at the 2,19 and 5,28 positions tended to increase the probability of adjacent water molecules at the involved residues. For the hydrophobic Abu

Table 3. Probability ($\rho_{g(r)}$) of finding water molecules within 0.5 nm of the terminal γ-sidechain atoms of Cys-, Abu-, and Ser- substituted residues (Sγ, CγHγ, and OγHγ, respectively)[a,b,c]

	Residue $\rho_{g(r)}$							
Peptide	**2**	**5**	**16**	**19**	**20**	**28**	**33**	**35**
CTX	0.066	0.044	0.013	0.076	0.020	0.049	0.019	0.094
[Abu^{2,19}]CTX	0.173	0.016	0.012	0.113	0.022	0.037	0.017	0.086
[Ser^{2,19}]CTX	0.147	0.070	0.028	0.121	0.018	0.087	0.041	0.081
[Abu^{5,28}]CTX	0.077	0.144	0.013	0.084	0.015	0.060	0.012	0.081
[Ser^{5,28}]CTX	0.086	0.186	0.033	0.066	0.027	0.104	0.032	0.079
[Abu^{16,33}]CTX	0.090	0.035	0.078	0.059	0.025	0.044	0.053	0.070
[Ser^{16,33}]CTX	0.079	0.032	0.050	0.061	0.029	0.048	0.060	0.086
[Abu^{20,35}]CTX	0.075	0.032	0.031	0.093	0.038	0.052	0.024	0.111
[Ser^{20,35}]CTX	0.082	0.006	0.024	0.076	0.046	0.026	0.025	0.165
[Abu^{2,5,16,19,20,28,33,35}]CTX	0.101	0.138	0.101	0.138	0.126	0.124	0.123	0.138
[Ser^{2,5,16,19,20,28,33,35}]CTX	0.203	0.194	0.156	0.166	0.202	0.172	0.182	0.204

Abbreviations: Abu, L-α-aminobutyric acid; CTX, chlorotoxin.

[a] The probability ($\rho_{g(r)}$) of finding water molecules within 0.5 nm of the terminal γ-sidechain atom of Cys (Sγ) in the $(Acetyl\text{-}Ala\text{-}Cys\text{-}Ala\text{-}NH_2)_2$ dipeptide, as described in Appendix A Materials, is 0.114.

[b] The probability ($\rho_{g(r)}$) of finding water molecules within 0.5 nm of the terminal γ-sidechain atoms of Abu (CγHγ) in the $Acetyl\text{-}Ala\text{-}Abu\text{-}Ala\text{-}NH_2$ peptide, as described in Appendix A, is 0.190.

[c] The probability ($\rho_{g(r)}$) of finding water molecules within 0.5 nm of the terminal γ-sidechain atoms of Ser (OγHγ) in the $Acetyl\text{-}Ala\text{-}Ser\text{-}Ala\text{-}NH_2$ peptide, as described in Appendix A, is 0.213.

substitutions, this effect tended to be more local, with the exception of [Abu^{16,33}]CTX, which had an increase for Cys^{2}, and [Abu^{20,35}]CTX, with an increase for Cys^{2} and Cys^{16}. For the hydrophilic Ser substitutions, the effects appeared to be more widespread, with changes in hydration for other nonsubstituted Cys residues. The [Abu^{2,5,16,19,20,28,33,35}]CTX, and [Ser^{2,5,16,19,20,28,33,35}]CTX peptides showed significant increases in adjacent water molecules through the substituted residues.

2.5 Conformational Analysis

The lowest energy conformations were determined by projecting each trajectory onto the first 2 dihedral principal components, as show in Figure 5. The optimal number of clusters was selected by using a combination of visual inspection, sum of squared error, and silhouette plots, as shown in Figure B4. The lowest energy conformation of CTX was consistent with the ^{1}H NMR structure: α-helix from residues 15-20, 2 antiparallel β-sheets from residues 27-29 and 32-34, and 2 intervening turns from residues 22-25 and 30-31. There were 4 salt bridges: Arg^{14} and Asp^{17}, Lys^{15} and Asp^{18}, Arg^{25} and the C-terminus, and Lys^{27} and the C-terminus.

The lowest energy conformations of [Abu^{2,19}]CTX, [Ser^{2,19}]CTX, [Abu^{5,28}]CTX, and [Ser^{16,33}]CTX maintain secondary and tertiary structure that are consistent with the αβ-motif and native CTX. This is despite the finding that the substituted residues' sidechains remain separated and do not interact (Table 2). The [Ser^{16,33}]CTX peptide is unique in that the 2 Ser residues are in close proximity (Table 2) and most likely interact by hydrogen bonding to each other or through an associated water molecule, given the significant increase in $\rho_{g(r)}$ (Table 3). The lowest energy conformations of [Ser^{5,28}]CTX, [Abu^{16,33}]CTX, [Abu^{20,35}]CTX, [Ser^{20,35}]CTX, [Abu^{2,5,16,19,20,28,33,35}]CTX, and [Ser^{2,5,16,19,20,28,33,35}]CTX deviate significantly from the αβ-motif and native CTX, with changes in both their secondary and tertiary structure.

2.6 Essential Subspace Analysis

Figure 6 shows a comparison of sampled essential subspace and its normalized values. Despite the substantial changes that occurred in secondary and tertiary conformations with some of the selective and global Cys substitutions, particularly [Ser^{5,28}]CTX, [Abu^{16,33}]CTX, [Ser^{20,35}]CTX, [Abu^{2,5,16,19,20,28,33,35}]CTX, and [Ser^{2,5,16,19,20,28,33,35}]CTX, none of the nRMSIP values were 0.65 or less. This indicates that a significant degree of sampled conformational space is shared across all substituted peptides. The results can be divided into 4 classes for comparison: $0.65 \leq$ nRMSIP < 0.75, $0.75 \leq$ nRMSIP < 0.85, and $0.85 \leq$ nRMSIP < 0.95, and $0.95 \leq$ nRMSIP. Peptide pairwise comparisons sharing the greatest degree of subspace conformational sampling were CTX to [Ser^{2,19}]CTX, CTX to [Abu^{16,33}]CTX, CTX to [Abu^{20,35}]CTX, [Ser^{2,19}]CTX to [Abu^{5,28}]CTX, [Ser^{2,19}]CTX to [Abu^{20,35}]CTX, and [Abu^{16,33}]CTX to [Abu^{20,35}]CTX.

3. Discussion

Selective pairwise replacement of CTX Cys-Cys disulfide bonds: 2-19, 5-28, 16-33, and 20-35 with Abu, and complete 2, 5, 16, 19, 20, 28, 33, and 35 replacements with Abu, have been performed [29]. Native CTX with intact disulfide bonds had an early-eluting peak on analytic high-performance liquid chromatography (HPLC) consistent with greater exposure of hydrophilic and polar residues in the folded structure. The [Abu^{2,19}]CTX and [Abu^{5,28}]CTX analogs also had defined early-eluting peaks on HPLC but with lower intensity than native CTX, which indicated that removal of the disulfide bond at either Cys^{2}-Cys^{19} or Cys^{5}-Cys^{28} had some minor effects on the folded structure. The [Abu^{16,33}]CTX and [Abu^{20,35}]CTX analogs did not form distinct elution peaks on HPLC, which indicated the formation of non–nativelike structures. Mass spectroscopy analysis of [Abu^{16,33}]CTX and [Abu^{20,35}]CTX also suggested multiple non–nativelike disulfide bonds, which would indicate that a

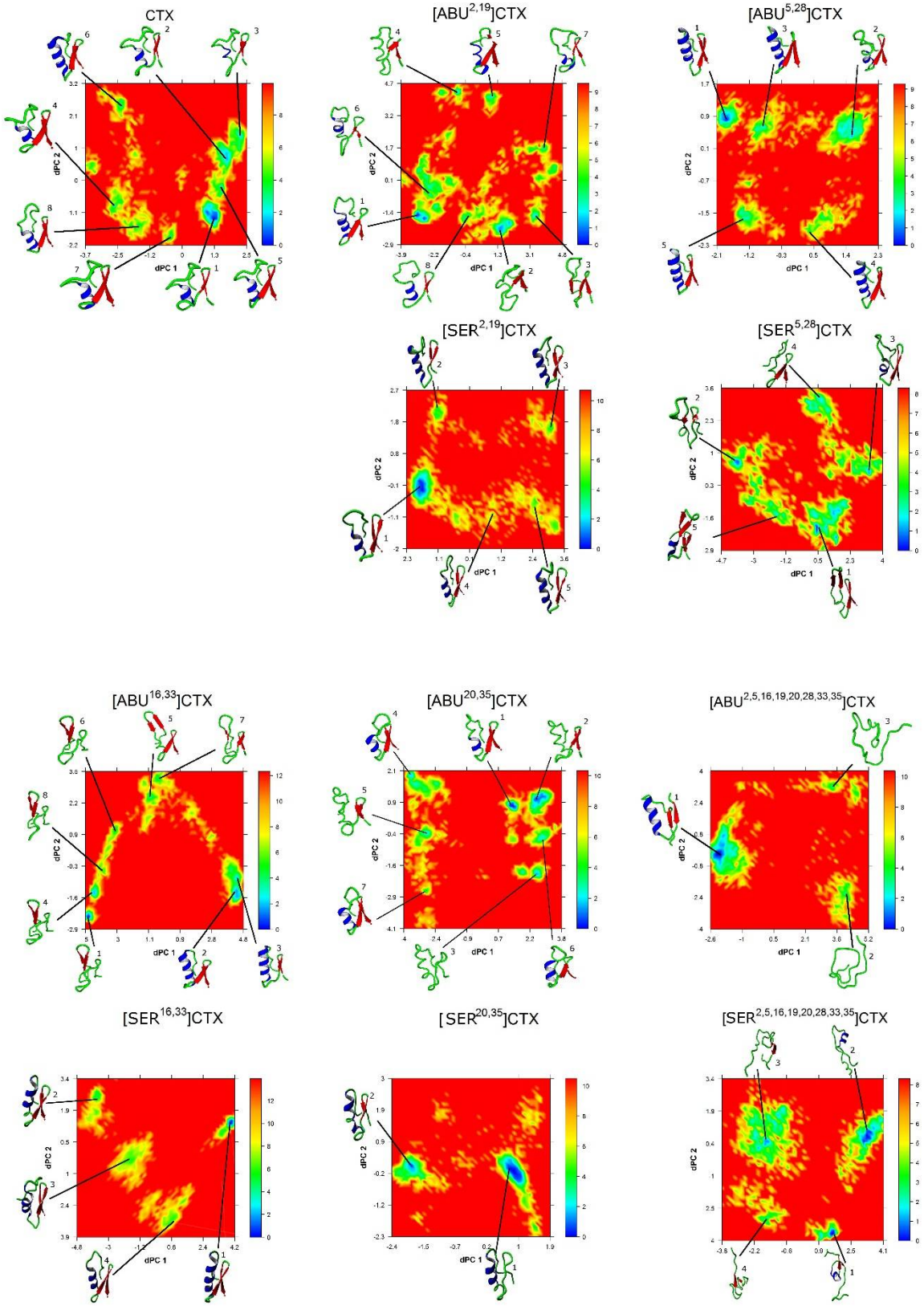


Figure 5. The free energy landscape (kJ•mol^{-1}) as a function of the first 2 components of the dihedral principal component analysis (dPC1 and dPC2). The lowest energy conformations, as determined by cluster analysis, are shown and numbered in order from lowest to highest relative energy. The conformations have been rotated to optimally display the secondary structure elements and interactions.

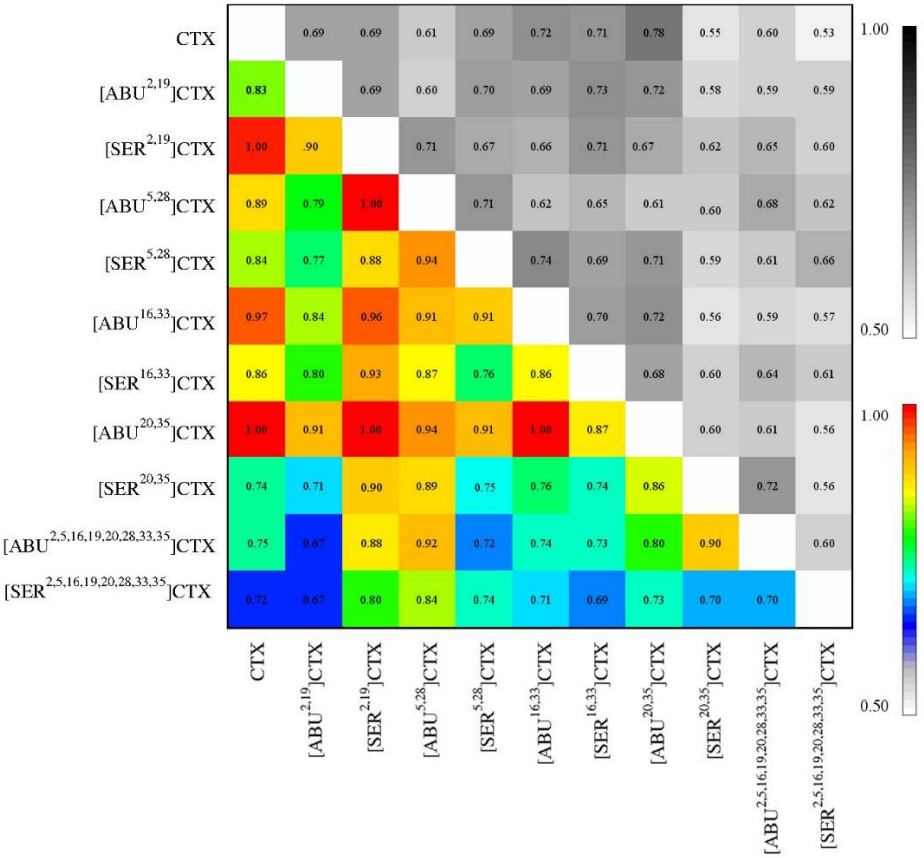


Figure 6. Comparison of essential subspace sampling. Root-mean-square inner product RMSIP (top right in grayscale) and normalized RMSIP (bottom left in color) overlap matrix of trajectories of molecular dynamics simulation of all 11 chlorotoxin (CTX) peptides. By definition, the diagonal values equal 1.0.

nativelike fold is not the most energetically favored conformation. CD measurements showed substantial structural similarity between CTX, [Abu^{2,19}]CTX, and [Abu^{5,28}]CTX suggestive of a folded αβ motif. The CD spectrum of [Abu^{2,5,16,19,20,28,33,35}]CTX was more consistent with significant random coil conformation [29]. The authors did not report deconvolution of the CD spectra with the fraction of populated α-helical, β-sheet, or bend/turn/coil conformations. Biological testing using a vascular endothelial growth factor–stimulated assay of cell migration indicated that CTX, [Abu^{2,19}]CTX, [Abu^{5,28}]CTX, and [Abu^{2,5,16,19,20,28,33,35}]CTX retain inhibitory activity, which suggested that the folded nativelike structure is not necessary for biological activity but does increase resistance to proteolytic degradation in vitro and in vivo [29].

The choice of Abu for substituting Cys residues has been driven by its isosteric and hydrophobic nature, with the goal of preserving the hydrophobic pocket in the peptide [27,85]. Because of the small and highly charged nature of these peptides, the surrounding environment of the Cys-Cys disulfide bond may not be as hydrophobic as previously considered. Examination of the rSASA plot of CTX in Figure B3 demonstrates that, although Cys residues are relatively solvent shielded, only the Cys^{2} and Cys^{5} residues are surrounded by either hydrophobic or aromatic residues. The remaining Cys residues are adjacent to charged residues such as Lys, Asp, or Arg, with the exception of Cys^{33} which is adjacent to a hydrophilic Gln^{32} and hydrophobic Leu^{34}. It is also

clear that Cys[2], Cys[19], and Cys[35] are significantly less solvent shielded, whereas Cys[16] and Cys[33] are the most solvent shielded as demonstrated by their rSASA values and their decreased probability of having water molecules within 0.5 nm (Table 3). In the current MD simulations, we were able to give detailed descriptions of the local hydrophobic/hydrophilic environment of each disulfide bond site.

Removal of the Cys[2]-Cys[19] disulfide bond and replacement with either Abu or Ser has only minor effects on CTX secondary and tertiary conformation. Removal of this bond is well tolerated because the remaining Cys[5]-Cys[28] bond maintains the N-terminus in close proximity with the antiparallel β-sheet. This interaction is also stabilized with an H-bond between the amine of Cys[5] and the carbonyl of Pro[31]. Subtle differences are present between the Abu and Ser substitutions. Whereas Abu substitution allows for the 2 Abu sidechains to widely separate and the N-terminus to drift away slightly from the α-helix and β-sheet, Ser substitution allows for close approximation of the N-terminus to the distal strand of the β-sheet, forming a 3-strand antiparallel β-sheet with prominent H-bonds between the amine of Ser[5] and the carbonyl of Pro[31], the amine of Cys[33] and the carbonyl of Met[3], the amine of Met[3] and the carbonyl of Cys[33], and NεH of Gln[32] and Oγ of Ser[3]. The sidechain of the Ser[19] residue on the α-helix is positioned so that it can interact with the surrounding solvent. Substitution of the Cys[2]-Cys[19] disulfide bond with Ser residues most likely removes some degree of tertiary strain within the peptide. Furthermore, it allowed more H-bonding interactions between the N-terminus and the antiparallel β-sheet. This is consistent with the disulfide bond between Cys[2]-Cys[19] being more solvent exposed and thus favoring potential hydrophilic substitutions.

Substitution of the Cys[5]-Cys[28] disulfide bond allows the α-helix region to be lengthened from residues 9 through 20. The preserved Cys[2]-Cys[19] bond, however, maintains the N-terminal region in close proximity with the distal α-helix and β-sheet. Due to elongation of the α-helix, the N-terminal region is not oriented in a way that would allow H-bonding to the distal strand of the β-sheet. Abu[5] and Abu[28] reside within a hydrophobic cleft between the α-helix and the proximal strand of the anti-parallel β-sheet interacting with the sidechain of Met[12]. This preserves the hydrophobic region present in the native CTX but in a slightly less compact conformation. The conformational importance of this hydrophobic region is demonstrated by Ser substitution, which results in significant disruption of the α-helix with a large loop extending from Ser[5] through Lys[15] and a shift in the location of the antiparallel β-sheet from residues 25 through 29 and 32 through 36. This results in a conformation that exposes the involved residues to solvent and disrupts the native hydrophobic region.

Despite the high degree of solvent shielding at the disulfide bond between Cys[16]-Cys[33], substitution with the hydrophobic Abu was not tolerated and resulted in nonnative isomers in oxidative folding experiments [28,29]. There is significant disruption of the secondary and tertiary conformations of the [Abu[16,33]]CTX peptide. As noted above, the α-helix is disrupted and the N-terminus is dominated by antiparallel β-sheet formation in the lowest energy conformations. The lack of a disulfide bond between Cys[16]-Cys[33] disrupts the association between the α-helical region and the terminal strand of the antiparallel β-sheet. This is particularly important because the sidechains of the antiparallel β-sheet region are oriented away from the α-helix and toward the solvent, and with the exception of Abu[16] and Cys[20], none of the α-helix residues are aligned to interact with the back surface of the β-sheet. As shown in Figure B2, the lack of H-bond donor and acceptor groups between the α-helix and β-sheet regions also makes the association between the α-helix and β-sheet strongly dependent on the presence of the disulfide bond between residues Cys[16] and Cys[33]. This substitution also results in closer proximity of the Cys[2]-Cys[19], Cys[5]-Cys[28], and Cys[20]-Cys[35] disulfide bonds, which may explain why multiple isomers occur with synthesis [29]. Ser substitution of these residues, however, preserves the overall secondary and tertiary conformation. The elongation of the α-helix appears to stabilize interaction between the two Ser residues and an associated water molecule.

Substitution at the Cys[20] and Cys[35] residues appears to allow preservation of the overall secondary and tertiary conformation with some degree of elongating of the α-helical region. Although the overall fold of the peptide is preserved, multiple H-bonds occur between the α-helix

and β-sheets that do not occur in the native peptide, and the resulting fold places the Cys^{2}-Cys^{19}, Cys^{5}-Cys^{28}, and Cys^{16}-Cys^{33} residues in close proximity and, like the Cys^{16}-Cys^{33} substitution, may explain why multiple isomers occur with synthesis. The result for Ser substitution is similar, except that the C-terminal antiparallel β-sheet is disrupted.

4. Materials and Methods

4.1 Initial Peptide Structures

All MD simulations were performed with the GROMACS 5.1.2 software packages using the CHARMM36m force field parameters with the CHARMM36m consistent version of TIP3P water [33-44]. The Leonard-Jones and electrostatic potentials for the atoms of the Abu residue were assigned in a manner consistent with other noncyclic aliphatic residues (Ala, Leu, Ile, and Val) within the CHARMM36m force field and are provided in Appendix A[33-36]. The initial structure of CTX for the simulation was the first conformation of the published ^{1}H NMR solution structure (Protein Data Bank ID: 1CHL) [30]. Starting structures of the peptides [$Abu^{2,19}$]CTX, [$Ser^{2,19}$]CTX, [$Abu^{5,28}$]CTX, [$Ser^{5,28}$]CTX, [$Abu^{16,33}$]CTX, [$Ser^{16,33}$]CTX, [$Abu^{20,35}$]CTX, [$Ser^{20,35}$]CTX, [$Abu^{2,5,16,19,20,28,33,35}$]CTX, and [$Ser^{2,5,16,19,20,28,33,35}$]CTX were obtained by changing the respective residues from Cys to Abu or Ser using YASARA [45]. The protonation state and charges of all residues within the peptides were set to correspond to a pH of 7.0.

4.2 Molecular Dynamics

Peptides were solvated in dodecahedral boxes with TIP3P water with 150 mM NaCl. Additional Cl^{-} and Na^{+} ions were used to neutralize the charges of the systems. The minimum distance of the peptide to the edge of the dodecahedron was 1.4 nm, with the exception of [$Abu^{2,5,16,19,20,28,33,35}$]CTX and [$Ser^{2,5,16,19,20,28,33,35}$]CTX, which required a 2.0 nm minimum distance of the peptide to the edge of the dodecahedron to prevent interaction of the peptide with its periodic image. The solvated systems were subjected to 5,000 steps of steepest descent energy minimization without restraints, allowing all bond distances and angles to relax. NVT (constant number, volume, and temperature) simulations of the positionally restrained peptides (force constant, 1,000 kJ•mol^{-1}) were performed for 10 nanoseconds at 310 K, followed by constant number, pressure and temperature (NPT) simulations of the positionally restrained peptides for 10 nanoseconds at 310 K and 101.325 kPa. The temperature and pressure, respectively, were kept constant by the stochastic velocity-rescaling method of Bussi *et al.* and the method of Berendsen *et al.* [46,47]. The relaxation constant was 0.1 picosecond and the isothermal compressibility was 4.5×10^{-5} bar^{-1}. We used 2 femtoseconds for the integration step and the LINCS algorithm for constraining all bonds to their correct length; the warning angle was 30° [48,49]. The particle mesh Ewald method was used to calculate long-range electrostatic interactions; cutoff distances were 1.2 nm and 1.0 nm, and the Fourier spacing was 0.15 nm [50]. The switch method was used to calculate van der Waals interactions, the short-range and long-range cutoffs were 1.0 and 1.2 nm, respectively.

Production runs of 4.2-microsecond NPT simulations were performed at 310 K and 101.325 kPa. The peptides and solvent with ions were separately coupled to a Parrinello-Rahman barostat and the temperature of the peptides and solvent separately maintained by the stochastic velocity-rescaling method [46,51]. The integration step, bond and angle constraints, long-range electrostatic interactions, and van der Waals interactions were calculated as described above.

4.3 Trajectory Analysis

The first 200 nanoseconds of each peptide simulation was considered system equilibration, and the subsequent 4 microseconds was used for analysis with a sampling frequency of 0.1 nanosecond.

4.3.1 Biophysical Properties

The radius of gyration (R_g), Cα-trace RMSD from the ^{1}H NMR solution conformation of CTX ($RMSD_{CTX}$), Cα-trace RMSD from the average sampled peptide conformation ($RMSD_{AVG}$), per-residue Cα-trace root-mean-square fluctuation (RMSF) from the average sampled peptide conformation, global and per-residue fraction of sampled secondary structure (α-helix, β-sheet, and bend/turn), and solvent-accessible surface area (SASA), hydrophobic SASA ($SASA_{Hydrophobic}$), and hydrophilic SASA ($SASA_{Hydrophilic}$) components were calculated using *g_gyration, g_rmsd, g_rmsf, do_dssp,* and *g_sasa* utilities of GROMACS, respectively [31,43]. The SASA of the peptides was calculated using the atomic radii of Lee and Richards with a probe radius of 0.14 nm and 1,000 points-per-sphere resolution [52].

4.3.2 Interactions

The per-residue SASA was normalized to values of the maximal SASA as determined from MD simulations on Acetyl-Ala-Xaa-Ala-NH_2 (where Xaa is either Ser or Abu) and (Acetyl-Ala-Cys-Ala-NH_2)$_2$ (dimer with Cys-Cys disulfide bond) [53,54]. Since the published data sets do not contain corresponding values for Abu, control simulations and calculations were performed as outlined in Appendix A (Table A1 and Figures A1 through A5). The change in per-residue SASA (ΔSASA) between the native CTX and Abu- or Ser-substituted peptides was calculated as:

$$\Delta SASA = SASA_{Substituted} - SASA_{CTX} \quad (1)$$

The intrachain sidechain-sidechain (SC-SC) and backbone-backbone (BB-BB) contacts and H-bonds were calculated with the *g_mdmat* and *g_hbond* utilities of GROMACS [43]. SC-SC and BB-BB contacts were designated for interatomic distances (between the heavy atoms of a residue) of 0.5 nm or less; H-bonds were designated for a donor-acceptor radius of 0.35 nm or smaller and a donor-hydrogen-acceptor angle of 30° or less. ^{1}H NMR cannot measure the sulfur-to-sulfur interatomic distance, and assignment of the spectra and calculation of the peptide's conformations is dependent on the presence of a Cys-Hβ to bonded Cys-Hβ nuclear Overhauser effect spectral peak; therefore, interactions between bound Cys and Abu- or Ser-substituted residues were determined by measuring the CβHβ to CβHβ center-of-mass distance and calculating the probability of contact within 0.5 nm [30,55]. Hydration of the sidechains of Cys, Abu, and Ser was determined by integrating the radial distribution function for water molecules within 0.5 nm of the sidechain. The integral of this function is equal to the probability of finding water molecules within the defined radius [56,57].

4.3.2 Conformational Analysis

The sampled conformations of the peptides were analyzed using the dihedral principal component analysis method [58,59]. The time-dependent φ/ψ dihedral angles from residues 2 to 35 of the peptides were extracted from the trajectories using the *g_rama* utility of GROMACS, and an in-house python script was used to transform the data for the input to the dihedral principal component analysis program (provided by Dr. Yuguang Mu). We identified the lowest energy conformations by projecting the trajectories of the first 2 principal components onto a 2-dimensional free energy (ΔG) landscape:

$$\Delta G = -R \cdot T \cdot ln \frac{\rho_{x,y}}{\rho_{max}} \quad (2)$$

where R is the universal gas constant, T is the temperature, and x and y are the first 2 dihedral principal components from the trajectory. The ΔG landscape was calculated by dividing the principal component 1 – principal component 2 subspace into grids to create a 2-dimensional histogram of the sampled phase space and calculating the probability $\rho_{(x,y)}$ using an in-house python

script, with $\rho_{max(x,y)}$ corresponding to the grid with the maximum probability of occurrence. Results were visualized using the *scatterplot3D*, *akima*, and *latticeExtra* packages in the R software environment, with conformations and secondary structural elements rendered using YASARA [45,60-63]. Families of low-energy conformations were identified using k-means clustering as implemented in the *cluster* package in R, and the identified lowest energy conformations were extracted for further analysis [63,64]. The optimal number of clusters was determined by visual inspection, sum of squared error, average silhouette width, silhouette coefficient, and distribution plots, as shown in Figure B4 [65-68].

4.3.4 Essential Subspace Analysis

The overlap of the sampled essential subspace of trajectories for different peptides as a function of Cys-Cys substitution was determined by calculating the root-mean-square inner product ($RMSIP_{AB}$):

$$\boldsymbol{RMSIP_{AB}} = \left(\frac{\mathbf{1}}{\boldsymbol{N}} \cdot \sum_{\boldsymbol{i=1}}^{\boldsymbol{N}} \sum_{\boldsymbol{j=1}}^{\boldsymbol{N}} \left(\boldsymbol{\eta_i^A} \cdot \boldsymbol{\eta_j^B} \right)^{\mathbf{2}} \right)^{\frac{\mathbf{1}}{\mathbf{2}}} \quad (3)$$

comparing MD trajectories of peptides A and B, where $\boldsymbol{\eta_i^A}$ and $\boldsymbol{\eta_j^B}$ are the respective eigenvectors of the sampled essential subspace, and N is the total number of eigenvectors to be considered [69-72]. To correct for sampling errors and autocorrelation, the RMSIP can be normalized (nRMSIP):

$$\boldsymbol{nRMSIP} = \frac{\boldsymbol{RMSIP_{A \bullet B}}}{\sqrt{\boldsymbol{RMSIP_{A1 \bullet A2}} \cdot \boldsymbol{RMSIP_{B1 \bullet B2}}}} \quad (4)$$

$RMSIP_{A1 \bullet A2}$ and $RMSIP_{B1 \bullet B2}$ are comparisons between the first and second halves of MD trajectories for peptides A and B, respectively. The nRMSIP ranges from 0 to 1; nRMSIP values approaching 1 indicate that the sampled essential subspaces of peptides A and B are similar, whereas lower values indicate differences between the sampled essential subspace of peptides A and B that cannot be explained by sampling alone. An nRMSIP of 0 indicates that the sampled essential subspaces of peptides A and B are orthogonal [72].

4.4 Statistical Analysis

The biophysical properties R_g, $RMSD_{CTX}$, $RMSD_{AVG}$, fractions of populated secondary structure (α-helix, β-sheet, and β-bend/turn and coil), SASA, $SASA_{Hydrophobic}$, $SASA_{Hydrophilic}$, and Cys and substituted Cys CβHβ-to-CβHβ center-of-mass distances were characterized for each peptide and reported as mean ± standard deviation. The homoscedasticity (homogeneities of variances) within the data were determined with Levene's test and comparisons made with analysis of variance (ANOVA) or Welch's ANOVA where appropriate (Levene's $P<.05$).

There were a total of 55 pairwise comparisons (N):

$$\boldsymbol{N} = \frac{\boldsymbol{n(n-1)}}{\mathbf{2}} \quad (5)$$

where *n* is the number of peptides in the series. The ANOVA and Welch's ANOVA are statistically robust methods of identifying differences between the means of the sampled populations, but they cannot determine which pairwise comparison of the means is statistically significant. Therefore, we

used a *post hoc* analysis with the Games-Howell test as implemented in the *car* package in R [63, 73-84].

5. Conclusions

In CTX, the disulfide bonds between Cys^{16}-Cys^{33} and Cys^{20}-Cys^{35} are required to maintain the association between the α-helical region from residues 13 through 20 and the 2 antiparallel β-sheets from residues 27 through 29 and 32 through 34. This conformational dependence arises from the lack of SC-SC and BB-BB interactions between the 3 components of the motif, either from hydrophobic, aromatic, salt bridge, or H-bond formation. Of these 2 disulfide bonds, the bond between Cys^{16} and Cys^{33} appears to be more critical, allowing for wider separation of the α-helix and β-sheet, whereas replacement of the Cys^{20} and Cys^{35} residues causes less conformational disruption. Selective use of $Ser^{16,33}$ may mitigate this destabilizing effect, since an interaction may occur between the 2 residues and an associated water molecule. Although stepwise removal of disulfide bonds or all disulfide bonds increases protease degradation, it has a less marked effect on biological activity, as demonstrated by Ojeda *et al.* [29]. The binding region of CTX to its purported receptor may be independent of a restricted conformation. It should also be noted that even with removal of all 4 Cys-Cys disulfide bonds, the peptides still share a substantial degree of conformational space sampling, and short, appropriately ordered segments of the peptide may present successfully to the receptor to initiate binding.

Supplementary Materials: Supplementary materials can be found at www.mdpi.com/xxx/s1.

Author Contributions: conceptualization, C.R.W and S.L.; methodology, C.R.W and S.L.; software, C.R.W., A.J.G., L.V.-O., and S.L.; validation, C.R.W., A.J.G. and L.V.-O.; formal analysis, A.J.G. and L.V.-O.; investigation, A.J.G. and L.V.-O.; resources, C.R.W.; data curation, A.J.G. and L.V.-O.; writing—original draft preparation, C.R.W.; writing—review and editing, C.R.W, A.J.G, L.V.-O., and S.L.; visualization, A.J.G and L.V.-O.; supervision, C.R.W.; project administration, C.R.W.; funding acquisition, C.R.W.

Funding: Funding for this investigation was provided in part through a grant from Mayo Clinic Health System-Franciscan Healthcare Foundation Inc. and Mayo Foundation for Medical Education and Research.

Acknowledgments: The molecular dynamics simulations and subsequent analysis were completed using the High Performance Cluster, Resource Computing Services, Mayo Clinic, Rochester, Minnesota.

Conflicts of Interest: Dr. Charles R. Watts is a consultant for Medtronic Spine and Biologics. The remaining authors have disclosed that they do not have any conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results".

Abbreviations

Abu	L-α-aminobutyric acid
ANOVA	analysis of variance
BB-BB	backbone-backbone
CTX	chlorotoxin
CD	circular dichroism
HPLC	high-performance liquid chromatography
MD	molecular dynamics
nRMSIP	normalized root-mean-square inner product

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 December 2018 doi:10.20944/preprints201812.0214.v1

Peer-reviewed version available at Int. J. Mol. Sci. 2019, 20; doi:10.3390/ijms20061261

NMR	nuclear magnetic resonance spectroscopy
NPT	constant number, pressure, and temperature
R_g	radius of gyration
rSASA	relative solvent-accessible surface area
RMSD	root-mean-square deviation
RMSF	root-mean-square fluctuation
RMSIP	root-mean-square inner product
SASA	solvent-accessible surface area
SC-SC	sidechain-sidechain
SD	standard deviation

Appendix A

A.1 Assignments of Force Field Parameters for L-α-Aminobutyric Acid (Abu)

The Leonard-Jones and electrostatic potentials for the atoms of the Abu residue were assigned in a manner consistent with other noncyclic aliphatic residues within the CHARMM36m force field (Ala, Leu, Ile, and Val) [33-36]. The following lines were added to the merged.rtp file in the charmm36-june2015.ff directory tree of GROMACS. Standard CMAP potentials were used for the ϕ,ψ dihedral angles and no modifications were made to the cmap.itp file within the charmm36-june2015.ff directory.

```
    [ ABU ]
      [ atoms ]
                N   NH1   -0.470  0
               HN     H    0.310  1
               CA   CT1    0.070  2
               HA   HB1    0.090  3
               CB   CT2   -0.180  4
              HB1   HA2    0.090  5
              HB2   HA2    0.090  6
               CG   CT3   -0.270  7
              HG1   HA3    0.090  8
              HG2   HA3    0.090  9
              HG3   HA3    0.090 10
                C     C    0.510 11
                O     O   -0.510 12
      [ bonds ]
               CB    CA
               CG    CB
                N    HN
                N    CA
                C    CA
                C    +N
               CA    HA
               CB   HB1
               CB   HB2
```

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 December 2018 doi:10.20944/preprints201812.0214.v1

Peer-reviewed version available at Int. J. Mol. Sci. 2019, 20; doi:10.3390/ijms20061261

```
              CG    HG1
              CG    HG2
              CG    HG3
               O     C
      [ impropers ]
               N    -C     CA     HN
               C    CA     +N      O
      [ cmap ]
              -C     N     CA      C     +N
```

A.2 Biophysical Properties of L-α-Aminobutyric Acid (Abu)

The solvent-accessible surface area (SASA) and radial distribution functions (Table A1) of the terminal methyl (Abu), hydroxyl (Ser), and disulfide (Cys-Cys) unfolded residues were determined in the following manner. The Acetyl-Ala-Xaa-Ala-NH_2 (where Xaa is either Ser or Abu) and (Acetyl-Ala-Cys-Ala-NH_2)$_2$ (bridged with a Cys-Cys disulfide bond) peptides were simulated using 0.5 μs NPT molecular dynamics runs at 310 K and 101.325 kPa. Fully extended initial conformations were created using the YASARA program [47]. The peptides were separately solvated in dodecahedral boxes with TIP3P water and 150 mM NaCl. The solvated conformations were energy-minimized by 5000 steps of steepest descent without restraint allowing all bond distances and angles to relax. NVT simulations of the positional restrained peptides (force constant of 1000 kJ•mol^{-1}) were performed for 10 ns at 310 K followed by NPT simulations of the positional restrained peptides for 10 ns at 310 K and 101.325 kPa stochastic velocity-rescaling method of Bussi for the temperature and the method of Berendsen for the pressure with separate temperature and pressure baths for the peptide and solvent and relaxation constants of 0.1 ps and 4.5 X 10^{-5} bar^{-1} isothermal compressibility [48,49]. The production runs of 500 ns NPT simulations were performed at 310 K and 101.325 kPa pressure. The peptide and solvent with ions were separately coupled to a 101.325 kPa Parrinello-Rahman barostats and the temperatures were maintained by separate coupling to stochastic thermostats using the velocity-rescaling method of Bussi-Parrinello [48,53]. The long-range electrostatic interactions were calculated using the PME method with a 1.2 nm cutoff distance and 0.15 nm Fourier spacing [52]. For the calculations of van der Waals interactions the switch method was used and the short-range and long-range cutoff, respectively, was 1.0 and 1.2 nm. The integration step was 2 fs; the LINCS algorithm was used to constrain all bonds to their correct length, with a warning angle of 30 [48,49].

Table A1. Residue solvent exposed surface area (SASA) Abu, Ser, and Cys-Cys; and probability ($\rho_{g(r)}$) of finding water molecules within 0.5 nm distance of the terminal methyl (Abu), hydroxyl (Ser), and disulfide (Cys) moieties. For the Cys-Cys disulfide bond the monomer value is given first and the dimer in parenthesis.

Residue	SASA /nm^2	$\rho_{g(r)}$
Abu	1.347 ± 0.121	0.190
Ser	1.219 ± 0.111	0.213
Cys-Cys	0.758 ± 0.142 (1.515 ± 0.283)	0.114

Standard quality assurance of the trajectories was performed and calculations of biophysical properties as outlined in the methods section of the manuscript Figures A1 through Figure A3. The first 100 ns of each peptide simulation was considered system equilibration and the subsequent 500 ns used for analysis with a sampling frequency of 0.1 ns.

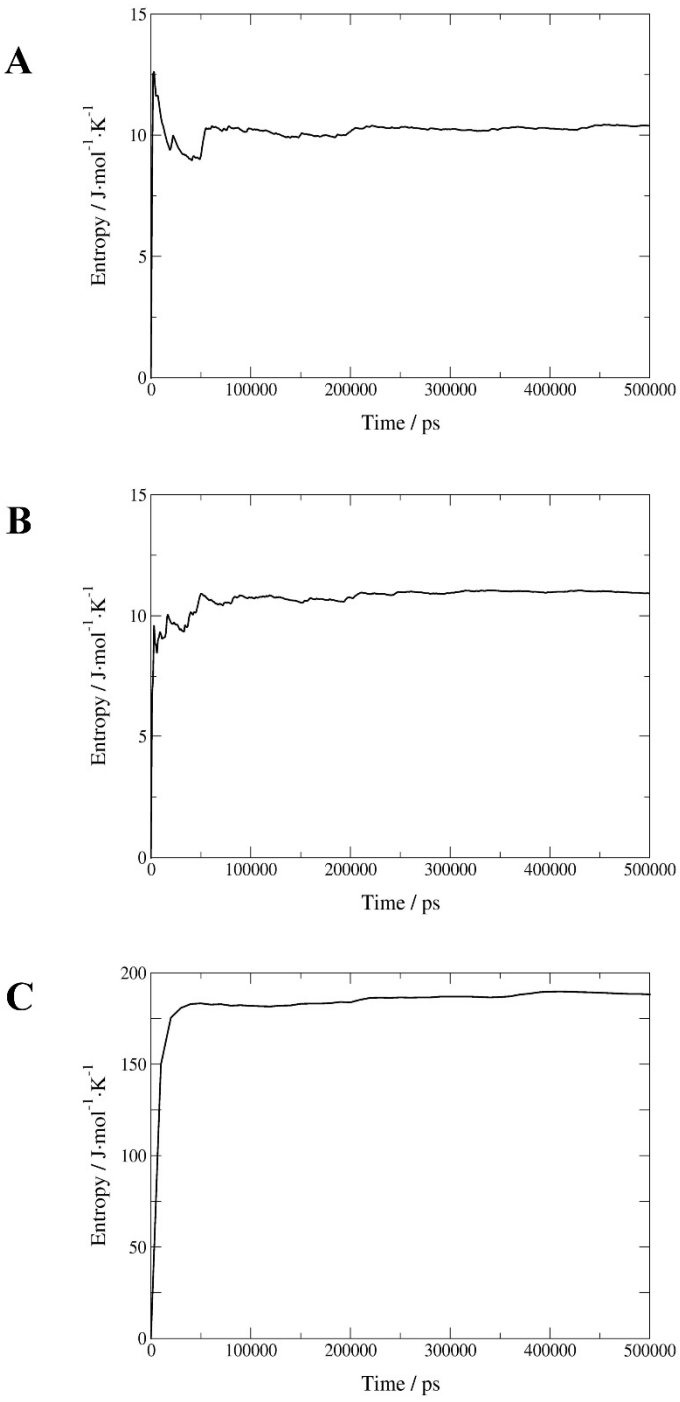


Figure A1. Cα-trace configurational entropy for Acetyl-Ala-Abu-Ala-NH_2 (**A**), Acetyl-Ala-Ser-Ala-NH_2 (**B**), and (Acetyl-Ala-Cys-Ala-NH_2)$_2$ (**C**) as a function of time.

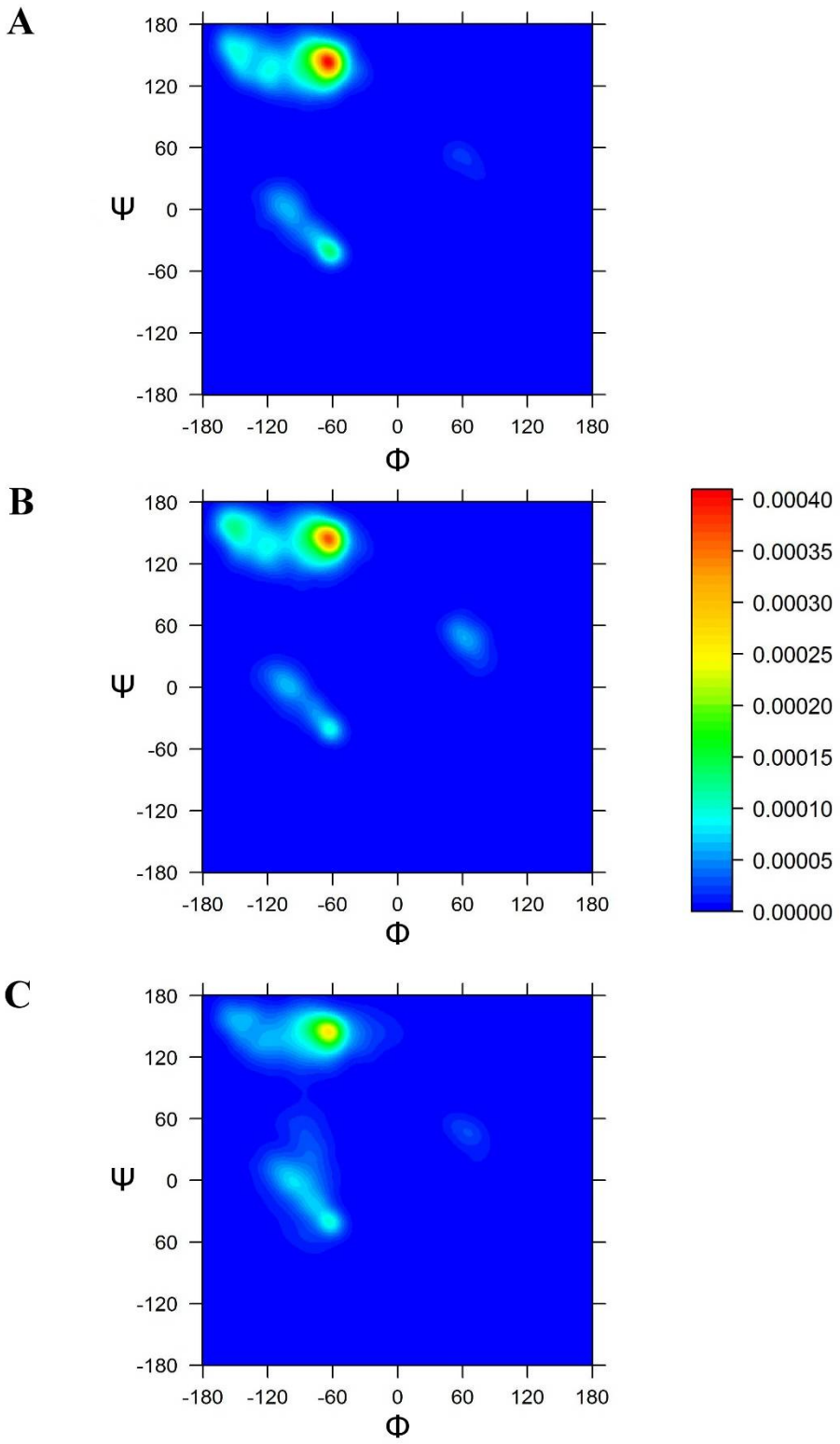
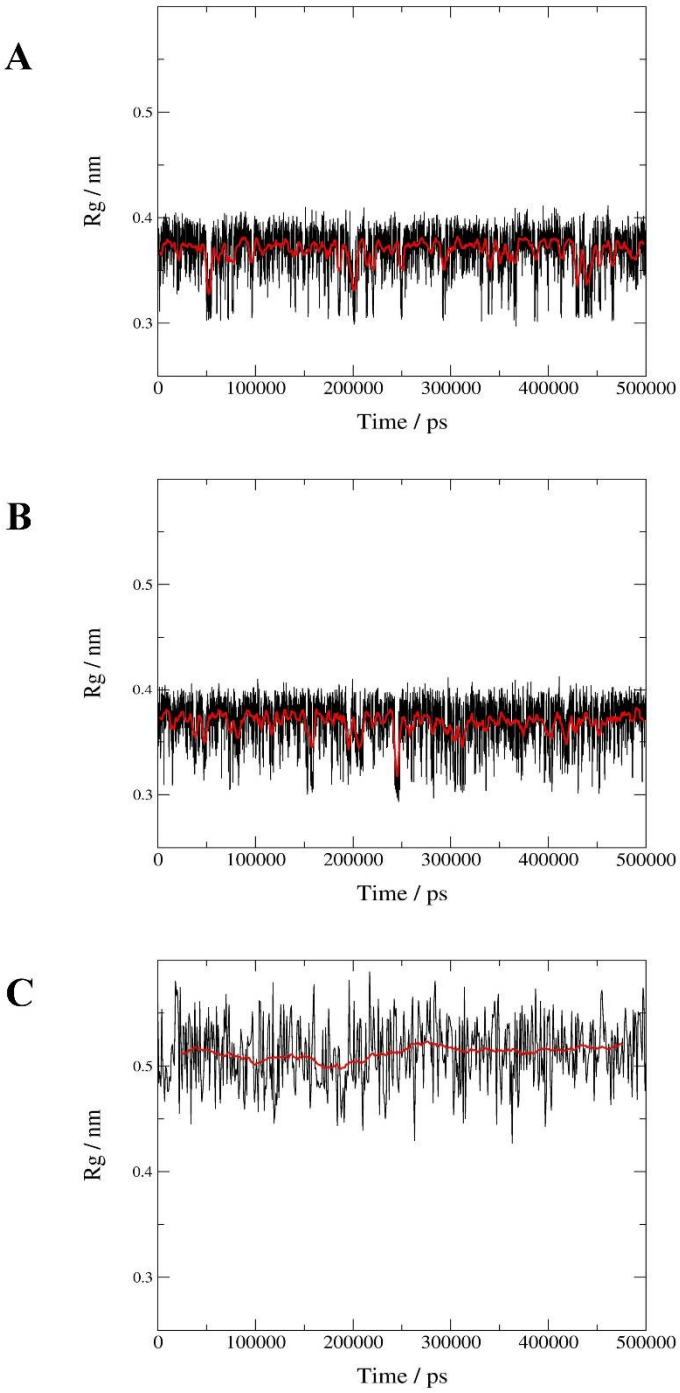


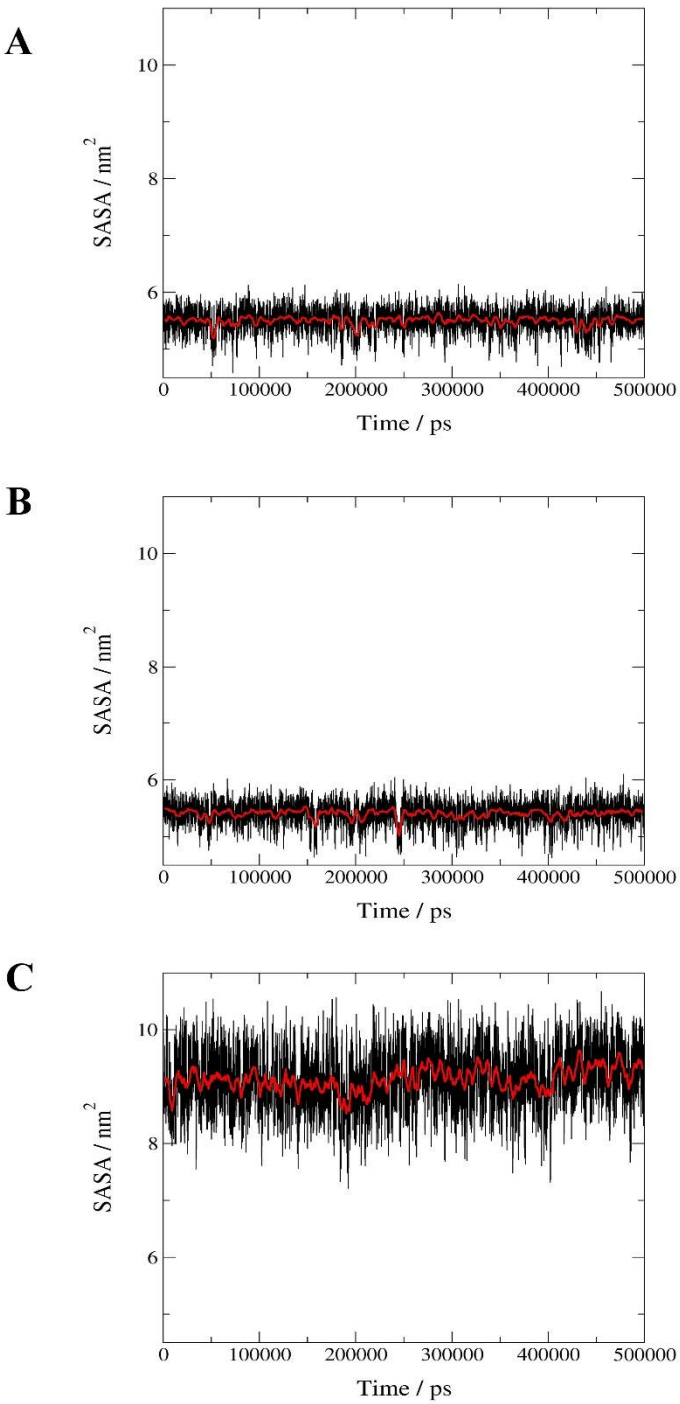
Figure A2. The sampled ϕ–ψ dihedral angles of Abu (**A**), Ser (**B**), and Cys-Cys disulfide bond (**C**). The color scale bars represent the global probability of a given ϕ–ψ dihedral angle.



598

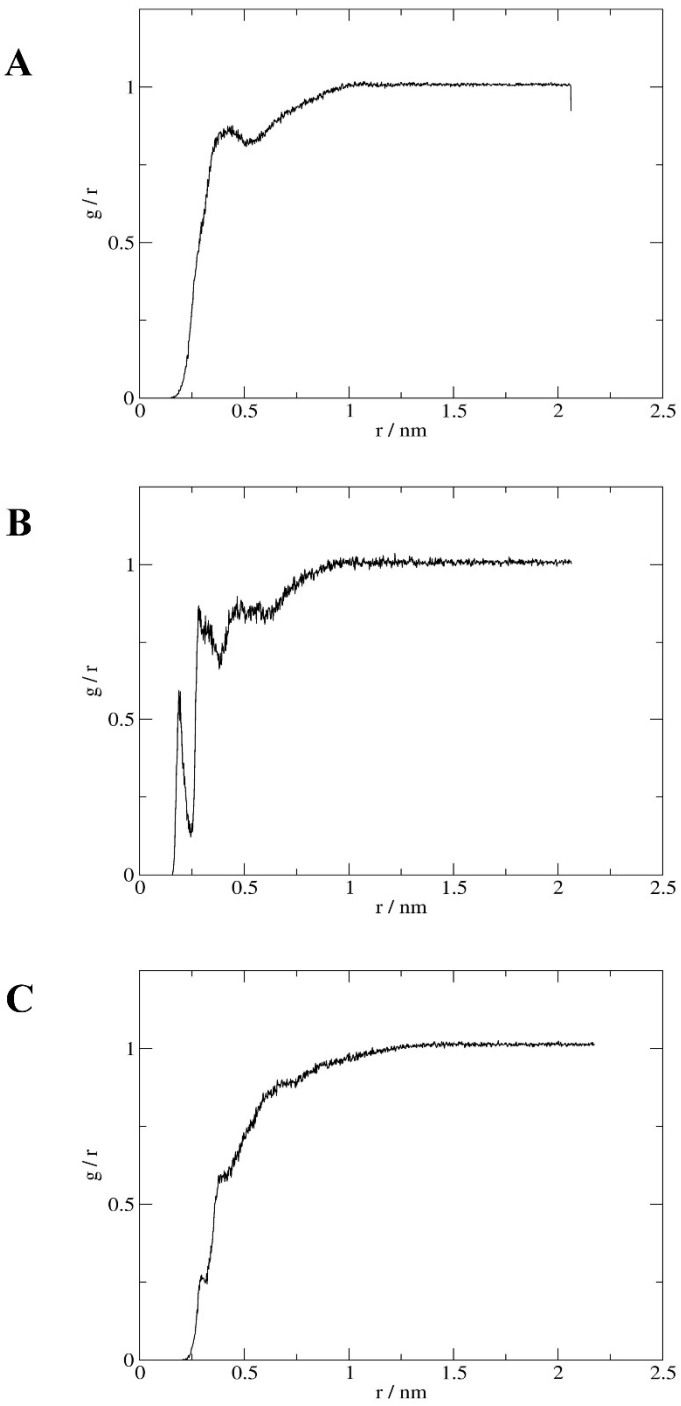
599 **Figure A3.** Radius of gyration (R_g) for Acetyl-Ala-Abu-Ala-NH_2 (**A**), Acetyl-Ala-Ser-Ala-
600 NH_2 (**B**), and (Acetyl-Ala-Cys-Ala-NH_2)$_2$ (**C**) as a function of time.

601 The time-averaged SASA and radial distribution plots representing the probability of finding
602 water molecules within 0.5 nm of the terminal moiety of Cys, Ser, and Abu are displayed
603 graphically in Figures A4 and A5.



604

605 **Figure A4.** Solvent Accessible Surface Area (SASA) for Abu (**A**), Ser (**B**), and Cys-Cys
606 disulfide bond (**C**) as a function of time.



607

608 **Figure A5.** Radial distribution function of water from the terminal methyl, hydroxyl and
609 disulfide moieties of Acetyl-Ala-Abu-Ala-NH_2 (**A**), Acetyl-Ala-Ser-Ala-NH_2 (**B**), and
610 (Acetyl-Ala-Cys-Ala-NH_2)$_2$ (**C**) respectively. The probability of finding water molecules
611 surrounding the respective moieties was determined by integrating the radial distribution
612 function within 0.5 nm distance.

613

Appendix B

Table B1 Non-statistically significant comparisons between calculated biophysical properties and the peptides as determined by Games-Howell *post-hoc* analysis, all other comparisons were considered statistically significant ($p < 0.05$).

Property	Compared Peptides		Mean Difference	95% CI	p-value
R_g /nm	[Abu^{2,19}]CTX	[Abu^{16,33}]CTX	0.0012	-0.0015 to 0.0039	0.94
	[Abu^{5,28}]CTX	[Ser^{16,33}]CTX	0.0009	-0.0013 to 0.0030	0.97
$RMSD_{CTX}$ /nm	[Abu^{5,28}]CTX	[Abu^{16,33}]CTX	0.0002	-0.0014 to 0.0018	1.00
	[Abu^{20,35}]CTX	[Ser^{2,19}]CTX	0.0004	-0.0010 to 0.0018	1.00
	[Abu^{20,35}]CTX	[Ser^{5,28}]CTX	0.0014	-0.0001 to 0.0028	0.10
	[Abu^{20,35}]CTX	[Ser^{2,5,16,19,20,28,33,35}]CTX	0.0006	-0.0009 to 0.0019	0.97
	[Ser^{2,19}]CTX	[Ser^{5,28}]CTX	-0.0010	-0.0004 to 0.0023	0.43
	[Ser^{5,28}]CTX	[Ser^{16,33}]CTX	0.0004	-0.0018 to 0.0010	1.00
	[Ser^{2,19}]CTX	[Ser^{2,5,16,19,20,28,33,35}]CTX	-0.0002	-0.0014 to 0.0011	1.00
	[Ser^{5,28}]CTX	[Ser^{2,5,16,19,20,28,33,35}]CTX	0.0008	-0.0005 to 0.0022	0.69
	[Ser^{16,33}]CTX	[Ser^{2,5,16,19,20,28,33,35}]CTX	0.0012	-0.0001 to 0.0025	0.09
$RMSD_{Avg}$ /nm	[Abu^{2,19}]CTX	[Ser^{5,28}]CTX	0.0013	-0.0001 to 0.0027	0.07
α-helix /ρ	CTX	[Abu^{20,35}]CTX	0.0047	-0.0001 to 0.0095	0.07
	[Abu^{16,33}]CTX	[Ser^{2,5,16,19,20,28,33,35}]CTX	-0.0047	-0.0106 to 0.0012	0.27
	[Ser^{20,35}]CTX	[Ser^{2,5,16,19,20,28,33,35}]CTX	-0.0045	-0.0099 to 0.001	0.24
β-sheet /ρ	[Abu^{5,28}]CTX	[Abu^{20,35}]CTX	-0.0022	-0.0047 to 0.0002	0.12
	[Abu^{20,35}]CTX	[Ser^{5,28}]CTX	-0.0005	-0.0048 to 0.0038	1.00
	[Abu^{5,28}]CTX	[Ser^{5,28}]CTX	0.0017	-0.0024 to 0.0058	0.97
Bend/turn /ρ	[Abu^{2,5,16,19,20,28,33,35}]CTX	[Ser^{16,33}]CTX	0.0039	-0.0041 to 0.0120	0.90
SASA /nm^2	[Abu^{2,19}]CTX	[Abu^{16,33}]CTX	0.066	-0.041 to 0.172	0.66
	[Abu^{5,28}]CTX	[Ser^{16,33}]CTX	-0.061	-0.175 to 0.053	0.82
$SASA_{Hydrophobic}$ /nm^2	[Abu^{2,5,16,19,20,28,33,35}]CTX	[Ser^{2,19}]CTX	-0.049	-0.14 to 0.039	0.78
$SASA_{Hydrophilic}$ /nm^2	[Abu^{16,33}]CTX	[Ser^{5,28}]CTX	-0.004	-0.089 to 0.081	1.00
	[Ser^{16,33}]CTX	[Ser^{20,35}]CTX	-0.002	-0.079 to 0.076	1.00

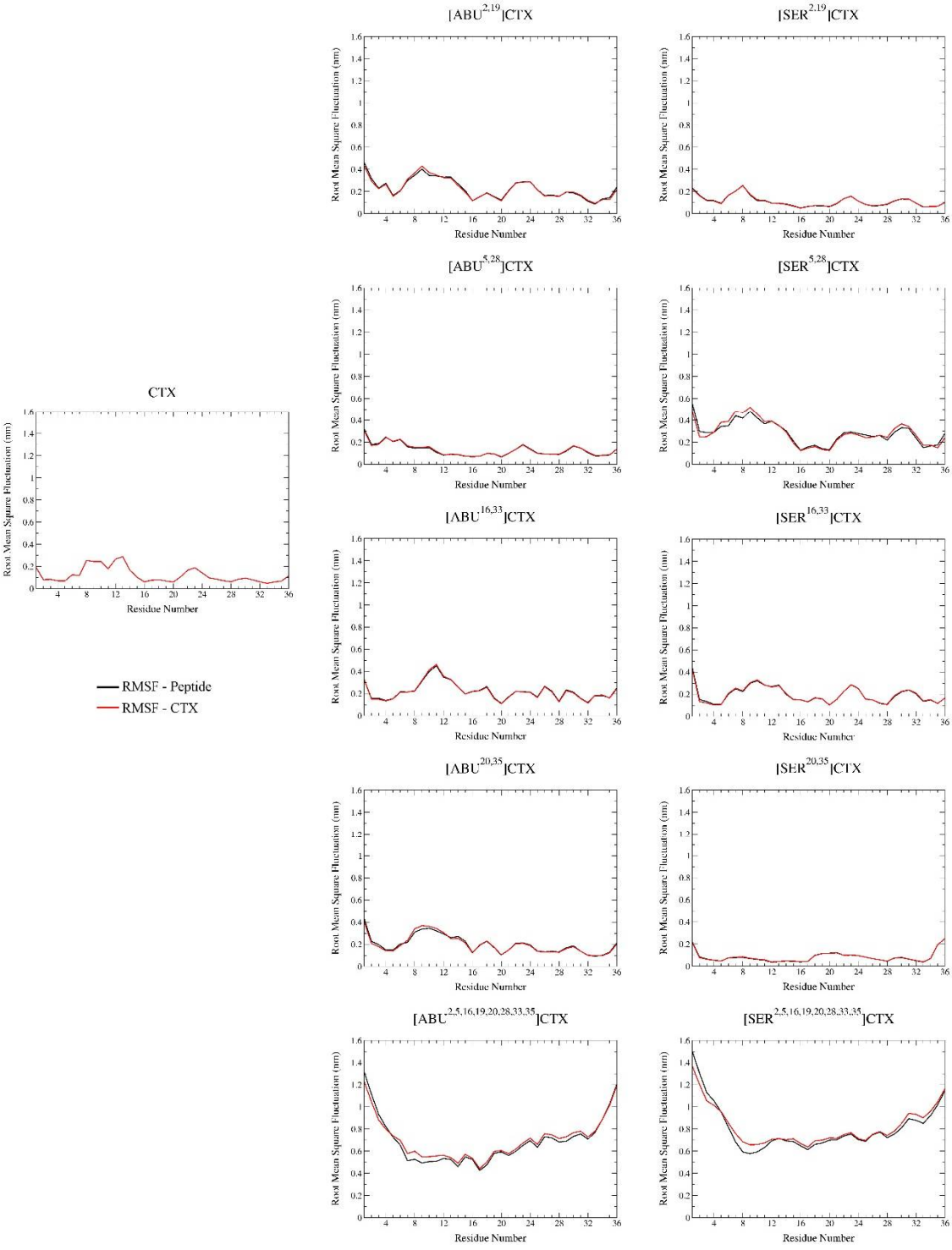


Figure B1. The per-residue Cα-trace RMSF (black) overlaid with the average conformation of the peptide itself (black) and with the average conformation of CTX (red).

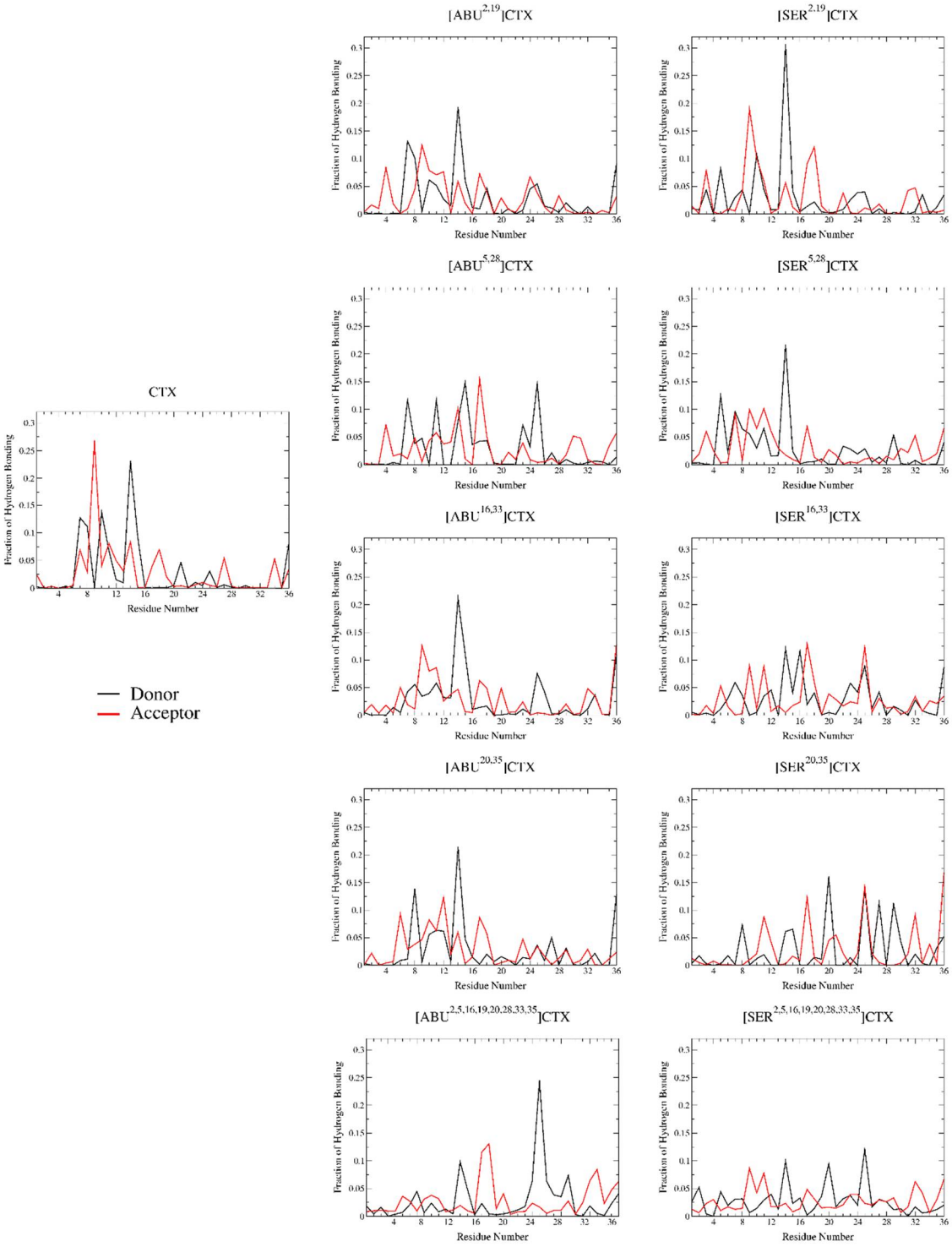


Figure B2. The fraction of populated H-bonds normalized to the total H-bonds observed during the simulation is plotted by residue.

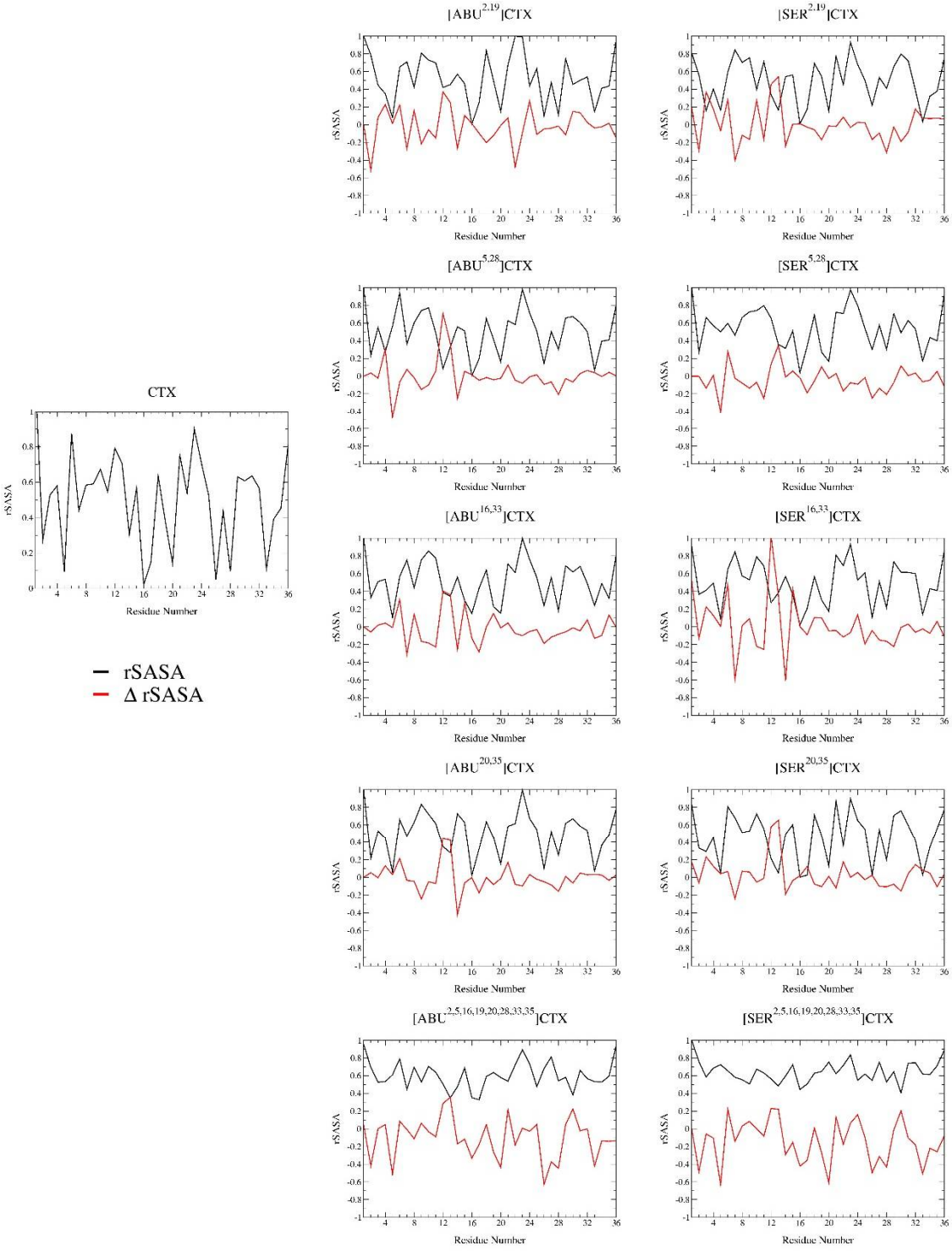


Figure B3. The relative solvent accessible surface area (rSASA in black) and change in relative solvent accessible surface area (ΔrSASA in red) compared to CTX as a function of residue number.

629 **Table B2**. Non-statistically significant comparisons between calculated center of mass
630 distances between CβHβ and CβHβ atoms for the peptides as determined by Games-
631 Howell post-hoc analysis, all other comparisons were considered statistically significant (p
632 < 0.05).

Cystine Bond	Compared Peptides		Mean Difference	95% CI	p-value
2-19	CTX	[Abu^{16,33}]CTX	-0.0006	-0.0020 to 0.0008	0.96
	[Abu^{5,28}]CTX	[Abu^{20,35}]CTX	0.0003	-0.0001 to 0.0015	1.00
5-28	CTX	[Ser^{20,35}]CTX	-0.0005	-0.0011 to 0.0003	0.62
	[Abu^{2,19}]CTX	[Ser^{2,19}]CTX	-0.0009	-0.0021 to 0.0002	0.27
	[Abu^{2,19}]CTX	[Ser^{16,33}]CTX	0.0004	-0.0008 to 0.0015	1.00
16-33	CTX	[Ser^{16,33}]CTX	0.0004	-0.0023 to 0.0031	1.00
	[Ser^{2,19}]CTX	[Ser^{20,35}]CTX	0.0001	-0.0008 to 0.0006	1.00
20-35	CTX	[Abu^{2,19}]CTX	0.0004	-0.0009 to 0.0002	0.58
	[Abu^{16,33}]CTX	[Ser^{2,19}]CTX	-0.0005	-0.0011 to 0.0001	0.31
	[Abu^{16,33}]CTX	[Ser^{16,33}]CTX	0.0002	-0.0004 to 0.0008	0.98

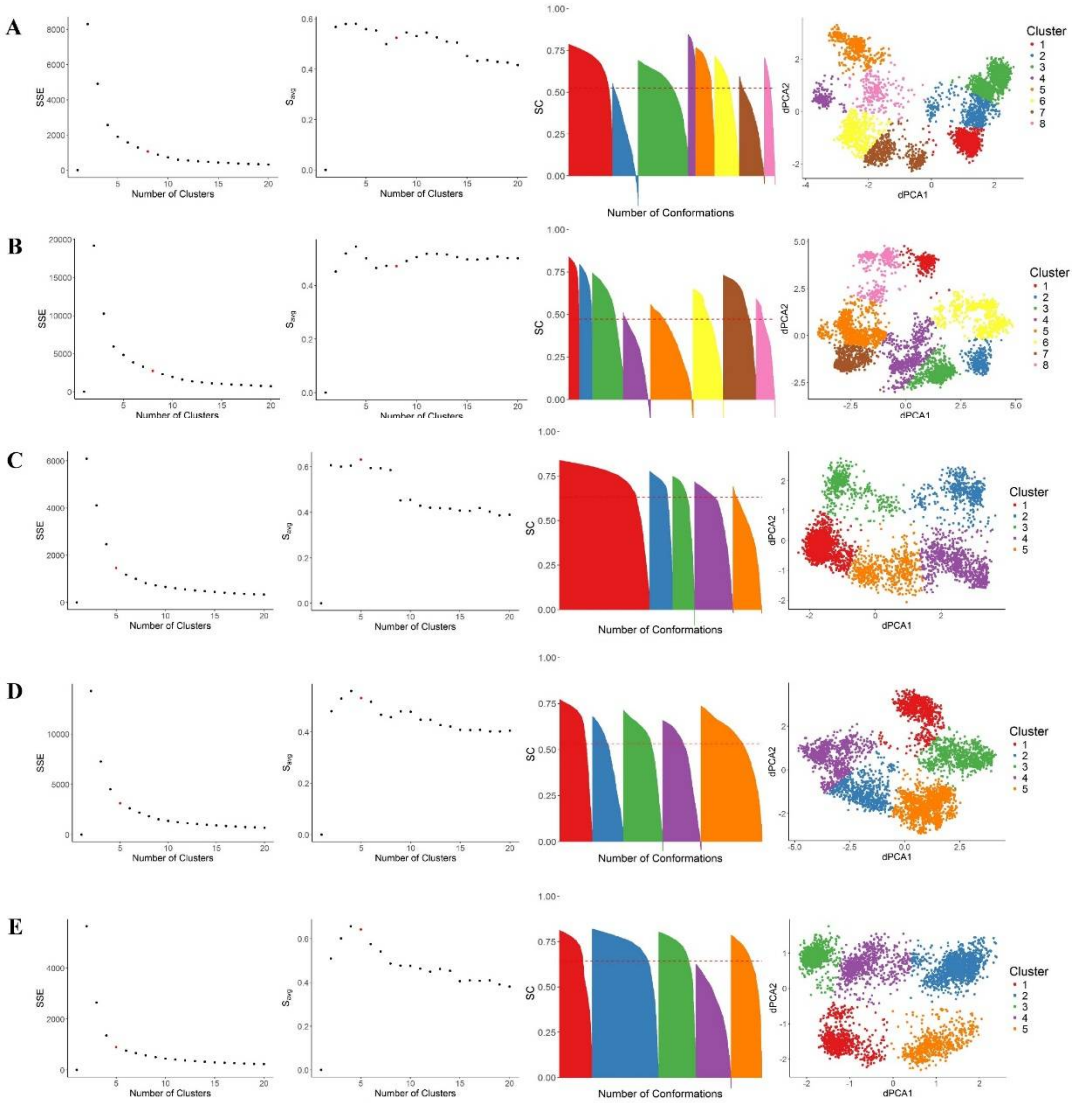


Figure B4. Sum of Squared Error (SSE), Average Silhouette width (S_{AVG}), Silhouette Coefficient (SC) and cluster distribution plots corresponding to the dihedral principle component conformational analysis presented in Figure 7 of the manuscript for: CTX (**A**), [Abu^{2,19}]CTX (**B**), [Ser^{2,19}]CTX (**C**), [Abu^{5,28}]CTX (**D**), [Ser^{5,28}]CTX (**E**), [Abu^{16,33}]CTX (**F**), [Ser^{16,33}]CTX (**G**), [Abu^{20,35}]CTX (**H**), [Ser^{20,35}]CTX (**I**), [Abu^{2,5,16,19,20,28,33,35}]CTX (**J**), and [Ser^{2,5,16,19,20,28,33,35}]CTX (**K**). The selected cluster number is shown in red with its corresponding average silhouette coefficient.

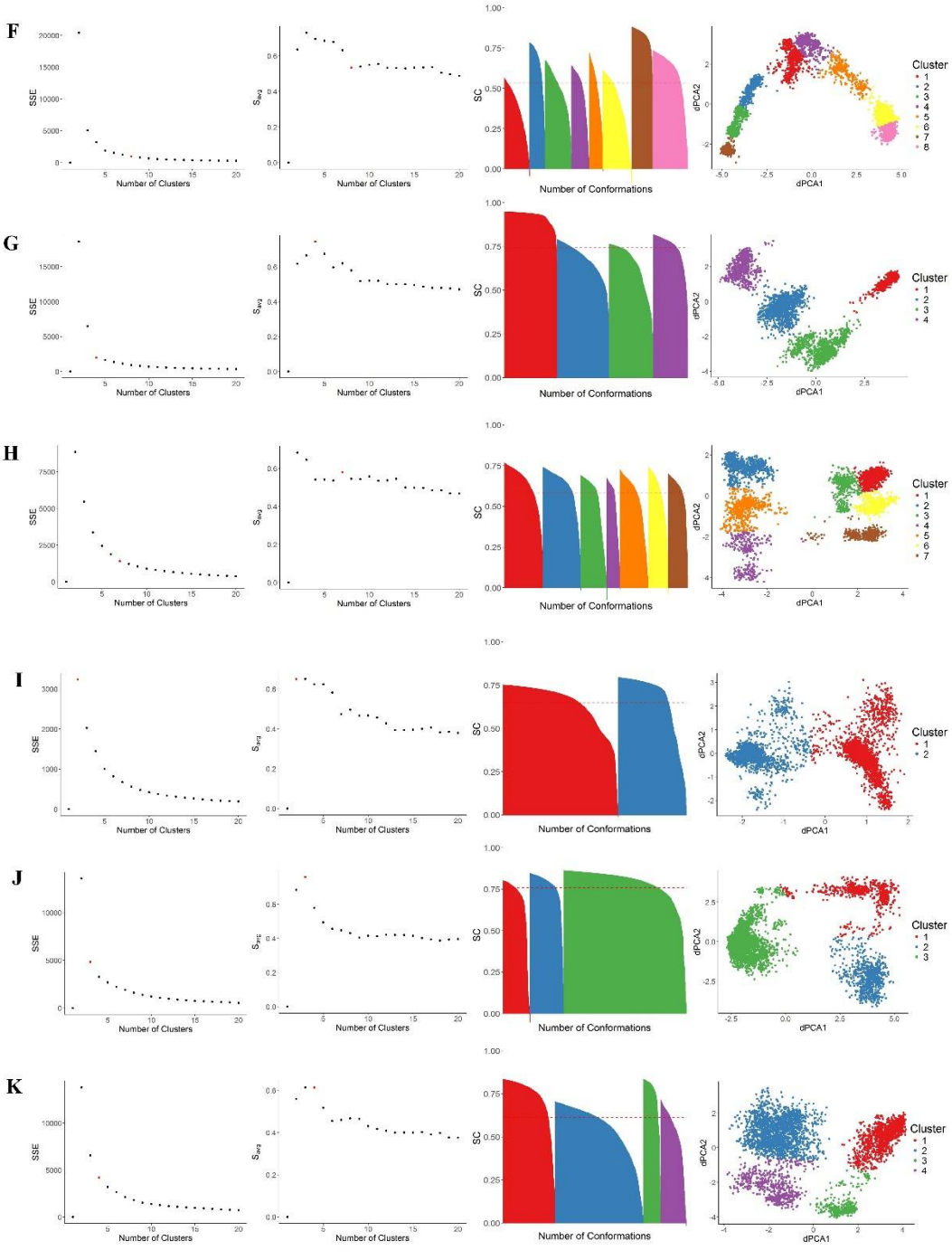


Figure B4. Sum of Squared Error (SSE), Average Silhouette width (S_{AVG}), Silhouette Coefficient (SC) and cluster distribution plots corresponding to the dihedral principle component conformational analysis presented in Figure 7 of the manuscript for: CTX (**A**), [Abu^{2,19}]CTX (**B**), [Ser^{2,19}]CTX (**C**), [Abu^{5,28}]CTX (**D**), [Ser^{5,28}]CTX (**E**), [Abu^{16,33}]CTX (**F**), [Ser^{16,33}]CTX (**G**), [Abu^{20,35}]CTX (**H**), [Ser^{20,35}]CTX (**I**), [Abu^{2,5,16,19,20,28,33,35}]CTX (**J**), and [Ser^{2,5,16,19,20,28,33,35}]CTX (**K**). The selected cluster number is shown in red with its corresponding average silhouette coefficient.

References

1. DeBin, J.A.; Strichartz, G.R. Chloride Channel Inhibition by the venom of the scorpion Leiurus Quinquestriatus. *Toxicon.* **1991**, *29(11)*, 1403-1408, DOI: 10.1016/0041-0101(91)90128-E.
2. DeBin, J.A.; Maggio, J.E.; Strichartz, G.R. Purification and characterization of chlorotoxin, a chloride channel ligand from the venom of the scorpion. *Am J Physiol.* **1993**, *264(2 Pt 1)*, C361-369, DOI: 10.1152/ajpcell.1993.264.2.C361.
3. Olsen, M.L.; Schade, S.; Lyons, S.A.; Amaral, M.D.; Sontheimer, H. Expression of voltage-gated chloride channels in human glioma cells. *J Neurosci.* **2003**, *23(13)*, 5572-5582, DOI: 10.1523/JNEUROSCI.23-13-05572.2003.
4. Soroceanu, L.; Gillespie, Y.; Khazaeli, M.B.; Sontheimer, H. Use of chlorotoxin for targeting of primary brain tumors. *Cancer Res.* **1998**, *58(21)*, 4871-4879, Available online: https://http://cancerres.aacrjournals.org/content/58/21/4871.
5. Lyons, S.A.; O'Neal, J.; Sontheimer, H. Chlorotoxin, a scorpion-derived peptide, specifically binds to gliomas and tumors of neuroectodermal origin. *Glia.* **2002**, *39(2)*, 162-173, DOI: 10.1002/glia.10083.
6. Mamelak, A.N.; Jacoby, D.B. Targeted delivery of antitumoral therapy to glioma and other malignancies with synthetic chlorotoxin (TM-601). *Expert Opin Drug Deliv.* **2007**, *4(2)*, 175-186, DOI: 10.1517/17425247.4.2.175.
7. Veiseh, O.; Gunn, J.W.; Kievit, F.M; Sun, C.; Fang, C.; Lee, J.S.H.; Zhang, M. Inhibition of tumor-cell invasion with chlorotoxin-bound superparamagnetic nanoparticles. *Small.* **2009**, *5(2)*, 256-264, DOI: 10.1002/smll.200800646.
8. Veiseh, M.; Gabikian, P.; Bahrami, S.B.; Veiseh, O.; Zhang, M.; Hackman, R.C.; Ravanpay, A.C.; Stroud, M.R.; Kusuma, Y.; Hansen, S.J.; Kwok, D.; Munoz, N.M.; Sze, R.W.; Grady, W.M.; Greenberg, N.M.; Ellenbogen, R.G.; Olson, J.M. Tumor paint: a chlorotoxin:Cy5.5 bioconjugate for intraoperative visualization of cancer foci. *Cancer Res.* **2007**, *67(14)*, 6882-6888, DOI: 10.1158/0008-5472.CAN-06-3948.
9. Akcan, M.; Stroud, M.R.; Hansen, S.J. Clark, R.J.; Daly, N.L.; Craik, D.; Olson, J.M. Chemical re-engineering of chlorotoxin improves bioconjugation properties for tumor imaging and targeted therapy. *J Med Chem.* **2011**, *54(3)*, 782-787, DOI: 10.1021/jm101018r.
10. Akcan, M.; Stroud, M.R.; Hansen, S.J.; Clark, R.J.; Daly, N.L.; Craik, D.; Olson, J.M. Correction to Chemical Re-engineering of Chlorotoxin Improves Bioconjugation Properties for Tumor Imaging and Targeted Therapy. *J Med Chem.* **2013**, *56(23)*, 9807, DOI: 10.1021/jm4016119.
11. Rosso, J.P.; Rochat, H. Characterization of ten proteins from the venom of the Moroccan scorpion Androctonus mauretanicus mauretanicus, six of which are toxic to the mouse. *Toxicon.* **1985**, *23(1)*, 113-125, DOI: 10.1016/0041-0101(85)90114-X.
12. Ali, S.A.; Stoeva, S.; Schütz, J.; Kayed, R.; Abbasi, A.; Zaidi, Z.H.; Voelter, W. Purification and primary structure of low molecular mass peptides from scorpion (Buthus sindicus) venom. *Comp Biochem Physiol A Mol Integr Physiol.* **1998**, *121(4)*, 323-332, DOI: 10.1016/S1095-6433(98)10140-X.
13. Possani, L.D.; Becerril, B.; Delepierre, M.; Tytgat, J. Scorpion toxins specific for Na+-channels. *Eur J Biochem.* **1999**, *264(2)*, 287-300, DOI: 10.1046/j.1432-1327.1999.00625.x.
14. Possani, L.D.; Merino, E.; Corona, M.; Bolivar, F.; Becerril, B. Peptides and genes coding for scorpion toxins that affect ion-channels. *Biochimie.* 2000, *82(9-10)*, 861-868, DOI: 10.1016/S0300-9084(00)01167-6.
15. Fry, B.G.; Roelants, K.; Champagne, D.E.; Scheib, H.; Tyndall, J.D.; King, G.F.; Nevalainen, T.J.; Norman, J.A.; Lewis, R.J.; Norton, R.S.; Renjifo, C.; de la Vega, R.C. The toxicogenomic multiverse: convergent recruitment of proteins into animal venoms. *Annu Rev Genomics Hum Genet.* **2009**, *10*, 483-511, DOI: 10.1146/annurev.genom.9.081307.164356.
16. Bontems, F.; Roumestand, C.; Gilquin, B.; Menez, A.; Toma, F. Refined structure of charybdotoxin: common motifs in scorpion toxins and insect defensins. *Science.* **1991**, *254(5037)*, 1521-1523, DOI: 10.1126/science.1720574.
17. Bonmatin, J.M.; Bonnat, J.L.; Gallet, X.; Vovelle, F.; Ptak, M.; Reichart, J.M.; Hoffman, J.A.; Keppi, E.; Legrain, M.; Achstetter, T. Two-dimensional 1H NMR study of recombinant insect defensin A in water: resonance assignments, secondary structure and global folding. *J Biomol NMR.* **1992**, *2(3)*, 235-256, DOI: 10.1007/BF01875319.
18. Cornet, B.; Bonmatin, J.M.; Hetru, C.; Hoffmann, J.A.; Ptak, M.; Vovelle, F. Refined three-dimensional solution structure of insect defensin A. *Structure.* **1995**, *3(5)*, 435-448, DOI: 10.1007/BF01875319.

19. Bruix, M.; Jimenez, M.A.; Santoro, J.; Gonzalez, C.; Colilla, F.J.; Mendez, E.; Rico, M. Solution structure of gamma 1-H and gamma 1-P thionins from barley and wheat endosperm determined by 1H-NMR: a structural motif common to toxic arthropod proteins. *Biochemistry.* **1993**, *32(2)*, 715-724, DOI: 10.1021/bi00053a041.
20. Craik, D.J.; Daly, N.L.; Waine, C. The cystine knot motif in toxins and implications for drug design. *Toxicon.* **2001**, *39(1)*, 43-60, DOI: 10.1016/S0041-0101(00)00160-4.
21. Ali, S.A.; Alam, M.; Abbasi, A.; Undheim, E.A.B.; Fry, B.G.; Kalbacher, H.; Voelter, W. Structure-Activity Relationship of Chlorotoxin-Like Peptides. *Toxins (Basel).* **2016**, *8(2)*, 36-52, DOI: 10.3390/toxins8020036.
22. Mouhat, S.; Jouirou, B.; Mosbah, A.; De Waard, M.; Sabatier, J.M. Diversity of folds in animal toxins acting on ion channels. *Biochem J.* **2004**, *378(Pt 3)*, 717-726, DOI: 10.1042/bj20031860.
23. Bulaj, G. Formation of disulfide bonds in proteins and peptides. *Biotechnol Adv.* **2005**, *23(1)*, 87-92, DOI: 10.1016/j.biotechadv.2004.09.002.
24. Sabatier, J.M.; Lecomte, C.; Mabrouk, K.; Darbon, H.; Oughideni, R.; Canarelli, S.; Rochat, H.; Martin-Eauclaire, M.F.; van Rietschoten, J. Synthesis and characterization of leiurotoxin I analogs lacking one disulfide bridge: evidence that disulfide pairing 3-21 is not required for full toxin activity. *Biochemistry.* **1996**, *35(33)*, 10641-10647, DOI: 10.1021/bi960533d.
25. Drakopoulou, E.; Vizzavona, J.; Neyton, J.; Aniort, V.; Bouet, F.; Virelizier, H.; Ménez, A.; Vita C. Consequence of the removal of evolutionary conserved disulfide bridges on the structure and function of charybdotoxin and evidence that particular cysteine spacings govern specific disulfide bond formation. *Biochemistry.* **1998**, *37(5)*, 1292-1301, DOI: 10.1021/bi9721086.
26. Zhu, Q.; Liang, S.; Martin, L.; Gasparini, S.; Menez, A.; Vita, C. Role of disulfide bonds in folding and activity of leiurotoxin I: just two disulfides suffice. *Biochemistry.* **2002**, *41(38)*, 11488-11494, DOI: 10.1021/bi026136m.
27. Song, J.; Gilquin, B.; Jamin, N.; Drakopoulou, E.; Guenneugues, M.; Dauplais, M.; Vita, C.; Ménez, A. NMR solution structure of a two-disulfide derivative of charybdotoxin: structural evidence for conservation of scorpion toxin alpha/beta motif and its hydrophobic side chain packing. *Biochemistry.* 1997, *36(13)*, 3760-3766, DOI: 10.1021/bi962720h.
28. Ojeda, P.G.; Wang, C.K.; Craik, D.J. Chlorotoxin: Structure, activity, and potential uses in cancer therapy. *Biopolymers.* **2016**, *106(1)*, 25-36, DOI: 10.1002/bip.22748.
29. Ojeda, P.G.; Chan, L.Y.; Poth, A.G.; Wang, C.K.; Craik, D.J. The role of disulfide bonds in structure and activity of chlorotoxin. *Future Med Chem.* **2014**, *6(15)*, 1617-1628, DOI: 10.4155/fmc.14.93.
30. Lippens, G.; Najib, J.; Wodak, S.J.; Tartar, A. NMR sequential assignments and solution structure of chlorotoxin, a small scorpion toxin that blocks chloride channels. *Biochemistry.* **1995**, *34(1)*, 13-21, DOI: 10.1021/bi00001a003.
31. Kabsch, W.; Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers.* **1983**, *22(12)*, 2577-2637, DOI: 10.1002/bip.360221211.
32. Frishman, D.; Argos, P. Knowledge-based protein secondary structure assignment. *Proteins.* **1995**, *23(4)*, 566-579, DOI: 10.1002/prot.340230412.
33. Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot B.L.; Grubmüller, H.; MacKerell, A.D. Jr. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat Methods.* **2017**, *14(1)*, 71-73, DOI: 10.1038/nmeth.4067.
34. Best, R.B.; Zhu, X.; Shim, J.; Lopes, P.E.; Mittal, J.; Feig, M.; Mackerell, A.D., Jr. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone phi, psi and side-chain chi(1) and chi(2) dihedral angles. *J Chem Theory Comput.* **2012**, *8(9)*, 3257-3273, DOI: 10.1021/ct300400x.
35. Mackerell, A.D., Jr.; Feig, M.; Brooks, C.L., 3rd. Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J Comput Chem.* **2004**, *25(11)*, 1400-1415, DOI: 10.1002/jcc.20065.
36. MacKerell, A.D., Jr., Bashford, D.; Bellott, M.; Dunbrack, R.L.; Evanseck, J.D.; Field, M.J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F.T.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D.T.; Prodhom, B.; Reiher, W.E.; Roux, B.; Schlenkrich, M.; Smith, J.C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B.* **1998**, *102(18)*, 3586-3616, DOI: 10.1021/jp973084f.

37. Abraham, M.J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J.C. Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*. **2015**, *1-2*, 19-25, DOI: 10.1016/j.softx.2015.06.001.
38. Páll, S.; Abraham, M.J.; Kutzner, C.; Hess, B.; Lindahl, E. Tackling Exascale Software Challenges in Molecular Dynamics Simulations with GROMACS. In: *Solving Software Challenges for Exascale: International Conference on Exascale Applications and Software* Proceedings of the EASC2014, Stockholm, Sweden, 2-3 April 2014; Markidis, S., Laure, E. Eds., Springer International Publishing, Heidelberg, Germany, 2015; pp. 3-27, DOI: 10.1007/978-3-319-15976-8_1.
39. Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J Chem Theory Comput.* **2008**, *4(3)*, 435-447, DOI: 10.1021/ct700301q.
40. van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A.E.; Berendsen, H.J. GROMACS: fast, flexible, and free. *J Comput Chem.* **2005**, *26(16)*, 1701-1718, DOI: 10.1002/jcc.20291.
41. Lindahl, E.; Hess, B.; van der Spoel, D. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J Mol Mod.* **2001**, *7(8)*, 306-317, DOI: 10.1007/s008940100.
42. Berendsen, H.J.C.; van der Spoel, D.; van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Comput Phys Commun.* **1995**, *91(1)*, 43-56, DOI: 10.1016/0010-4655(95)00042-E.
43. Abraham, M.J.; van der Spoel, D.; Lindahl, E.; Hess, B. and the GROMACS Development Team. GROMACS User Manual version 5.1.2. Available online: http://www.gromacs.org (Accessed on 1 June 2017).
44. Bjelkmar, P.; Larsson, P.; Cuendet, M.A.; Hess, B.; Lindahl, E. Implementation of the CHARMM Force Field in GROMACS: Analysis of Protein Stability Effects from Correction Maps, Virtual Interaction Sites, and Water Models. *J Chem Theory Comput.* **2010**, *6(2)*, 459-466, DOI: 10.1021/ct900549r.
45. Krieger, E.; Vriend, G. YASARA View - molecular graphics for all devices - from smartphones to workstations. *Bioinformatics.* **2014**, *30(20)*, 2981-2982, DOI: 10.1093/bioinformatics/btu426.
46. Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J Chem Phys.* **2007**, *126(1)*, 014101, DOI: 10.1063/1.2408420.
47. Berendsen, H.J.C.; Postma, J.P.M.; van Gunsteren, W.F.; DiNola, A.; Haak, J.R. Molecular dynamics with coupling to an external bath. *J Chem Phys.* **1984**, *81(8)*, 3684-3690, DOI: 10.1063/1.448118.
48. Hess, B.; Bekker, H.; Berendsen, H.J.C.; Fraaije, J.G.E.M. LINCS: A linear constraint solver for molecular simulations. *J Comput Chem.* **1997**, *18(12)*, 1463-1472, DOI: 10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H.
49. Hess, B. P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. *J Chem Theory Comput.* **2008**, *4(1)*, 116-122, DOI: 10.1021/ct700200b.
50. Essmann, U.; Perera, L.; Berkowitz, M.L. A smooth particle mesh Ewald method. *J Chem Phys.* **1995**, *103(19)*, 8577-8593, DOI: 10.1063/1.470117.
51. Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J Appl Phys.* **1981**, *52(12)*, 7182-7190, DOI: 10.1063/1.328693.
52. Lee, B.; Richards, F.M. The interpretation of protein structures: estimation of static accessibility. *J Mol Biol.* **1971**, *55(3)*, 379-400, DOI: 10.1016/0022-2836(71)90324-X
53. Zielenkiewicz, P.; Saenger, W. Residue solvent accessibilities in the unfolded polypeptide chain. *Biophys J.* **1992**, *63(6)*, 1483-1486, DOI: 10.1016/S0006-3495(92)81746-0.
54. Tien, M.Z.; Meyer, A.G.; Sydykova, D.K.; Spielman, S.J.; Wilke, C.O. Maximum allowed solvent accessibilities of residues in proteins. *PLoS One.* **2013**, *8(11)*, e80635, DOI: 10.1371/journal.pone.0080635.
55. Klaus, W.; Broger, C.; Gerber, P.; Senn, H. Determination of the disulfide bonding pattern in proteins by local and global analysis of nuclear magnetic resonance data. Application to flavoridin. *J Mol Biol.* **1993**, *232(3)*, 897-906, DOI: 10.1006/jmbi.1993.1438.
56. White, A.D.; Keefe, A.J.; Ella-Menye, J.R.; Nowinski, A.K.; Shao, Q.; Pfaendtner, J.; Jiang, S. Free energy of solvated salt bridges: a simulation and experimental study. *J Phys Chem B.* **2013**, *117(24)*, 7254-7259, DOI: 10.1021/jp4024469.
57. Nguyen, B.L.; Pettitt, B.M. Effects of Acids, Bases, and Heteroatoms on Proximal Radial Distribution Functions for Proteins. *J Chem Theory Comput.* **2015**, *11(4)*, 1399-1409, DOI: 10.1021/ct501116v.
58. Mu, Y, Nguyen, P.H.; Stock, G. Energy landscape of a small peptide revealed by dihedral angle principal component analysis. *Proteins.* **2005**, *58(1)*, 45-52, DOI: 10.1002/prot.20310.

59. Altis, A.; Nguyen, P.H.; Hegger, R.; Stock, G. Dihedral angle principal component analysis of molecular dynamics simulations. *J Chem Phys.* **2007**, *126(24)*, 244111, DOI: 10.1063/1.2746330.
60. Ligges, U.; Maechler, M. scatterplot3d - An R Package for Visualizing Multivariate Data. *J Stat Softw.* **2003**, *8(11)*, 1-20, Available online: https://cran.rproject.org/web/packages/scatterplot3d/ (accessed on 6 Aug 2018).
61. Akima, H.; Gebhardt, A.; Petzold, T.; Maechler, M. Akima: Interpolation of Irregularly and Regularly Spaced Data. Available online: https://cran.r-project.org/web/packages/akima/ (accessed on 6 Aug 6 2018).
62. Sarkar, D.; Andrews, F. latticeExtra: Extra Graphical Utilities Based on Lattice. Available online: https://cran.r-project.org/web/packages/latticeExtra/ (accessed on 6 Aug 2018).
63. R Core Team. R: A language and environment for statistical computing. Available online: http://www.R-project.org/ (accessed on 6 Aug 2018).
64. Maechler, M.; Rousseeuw, P.; Struyf, A.; Hubert, M.; Hornik, K.; Studer, M.; Roudier, P.; Gonzalez, J.; Kozlowski, K. cluster: methods for cluster analysis. Available online: https://cran.r-project.org/web/packages/cluster/ (accessed on 6 Aug 2018).
65. Tan, P.-N.; Steinbach, M.; Kumar, V. Cluster Analysis: Basic Concepts and Algorithms. In: Introduction to Data Mining. 2nd ed.; Pearson Addison Wesley, Boston, MA, USA 2005; pp. 487-568, ISBN 0133128903 978-0133128901.
66. Tan, P.-N.; Steinbach, M.; Kumar, V. Cluster Analysis: Additional Issues and Algorithms In: Introduction to Data Mining. 2nd ed.; Pearson Addison Wesley, Boston, MA, USA 2005; pp. 569-650, ISBN 0133128903 978-0133128901.
67. Rousseeuw, P.J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J Comput Appl Math.* **1987**, *20*, 53-65, DOI: 10.1016/0377-0427(87)90125-7.
68. Thinsungnoen, T.; Kaoungku, N.; Durongdumronchai, P.; Kerdprasop, K.; Kerdprasop, N. The Clustering Validity with Silhouette and Sum of Squared Errors. Proceedings of the 3rd International Conference on Industrial Application Engineering, Kitakyushu, Japan, 28-31 March 2015; Ehara, F., Nakashima, S. Eds.; Institute of Industrial Applications Engineers (IIAE), Kitakyushu, Japan, 2015, *3*, 44-51, DOI: 10.12792/iciae2015.012.
69. Andricioaei, I.; Karplus, M. On the calculation of entropy from covariance matrices of the atomic fluctuations. *J Chem Phys.* **2001**, *115(14)*, 6289-6292, DOI: 10.1063/1.1401821.
70. Hess, B. Convergence of sampling in protein simulations. *Phys Rev E Stat Nonlin Soft Matter Phys.* **2002**, *65(3 Pt 1)*, 031910, DOI: 10.1103/PhysRevE.65.031910.
71. D'Alessandro, M.; Paci, M.; Amadei, A. Characterization of the conformational behavior of peptide Contryphan Vn: a theoretical study. *Biopolymers.* **2004**, *74(6)*, 448-456, DOI: 10.1002/bip.20090.
72. Martin-Garcia, F.; Papaleo, E.; Gomez-Puertas, P.; Boomsma, W.; Lindorff-Larsen, K. Comparing molecular dynamics force fields in the essential subspace. *PLoS One.* **2015**, *10(3)*, e0121114, DOI: 10.1371/journal.pone.0121114.
73. Navidi, W. Sampling and Descriptive Statistics. In: Statistics for Engineers and Scientists. 4th ed.; McGraw Hill Education, New York, NY, USA, 2015; pp. 1-47, ISBN 0073401331 978-0073401331.
74. Navidi W. Factorial Experiments. In: Statistics for Engineers and Scientists. 4th ed.; McGraw Hill Education, New York, NY, USA, 2015; pp. 662-764, ISBN 0073401331 978-0073401331..
75. Fisher, R.A. The Correlation between Relatives on the Supposition of Mendelian Inheritance. *Trans R Soc Edinb.* **1918**, *52(2)*, 399-433, DOI: 10.1017/S0080456800012163.
76. Fisher, R.A. On the "probable error" of a coefficient of correlation deduced from a small sample. *Metron.* **1921**, *1*, 3-32, Available online: https://digital.library.adelaide.edu.au/dspace/handle/2440/15169.
77. Tiku, M.L. Power Function of the F-Test Under Non-Normal Situations. *J Am Stat Assoc.* **1971**, *66(336)*, 913-916, DOI: 10.1080/01621459.1971.1048237.
78. Levene, H. Robust tests for equality of variances. In: Contributions to probability and statistics : essays in honor of Harold Hotelling.; Olkin, I.; Ghurye, S.G.; Hoeffding, W.; Madow, W.G.; Mann, H.B. Eds.; Stanford University Press, Stanford, CA: Stanford University Press, 1960, pp. 278-292, ISBN 0804705968, 978-0804705967.
79. Scheffé H. The analysis of variance. Wiley, New York, NY, USA, 1959, ISBN 0471345059 978-0471345053.
80. Welch, B.L. The generalization of student's problems when several different population variances are involved. *Biometrika.* **1947**, *34(1-2)*, 28-35, DOI: 10.1093/biomet/34.1-2.28.

81. Welch, B.L. On the Comparison of Several Mean Values: An Alternative Approach. *Biometrika.* **1951**, *38(3-4)*, 330-336, DOI: 10.2307/2332579.
82. Games, P.A.; Keselman, H.J.; Clinch, J.J. Tests for homogeneity of variance in factorial designs. *Psychol Bull.* **1979**, *86(5)*, 978-984, DOI: 10.1037/0033-2909.86.5.978.
83. Tukey, J.W. Comparing individual means in the analysis of variance. *Biometrics.* **1949**, *5(2)*, 99-114, DOI: 10.2307/3001913.
84. Fox, J.; Weisberg, S.; Price, B.; Adler, D.; Bates, D.; Baud-Bovy, G.; Bolker, B.; Ellison, S.; Firth, D.; Friendly, M.; Gorjanc, G.; Graves, S.; Heiberger, R.; Laboissiere, R.; Maechler, M.; Monette, .; Murdoch, D.; Nilsson, H.; Ogle, D.; Ripley, B.; Venables, W.; Walker, S.; Winsemius, D.; Zeileis, A. car: Companion to Applied Regression. Available online: https://cran.r-project.org/web/packages/car (accessed on 7 Aug 2018).
85. Vita, C.; Aniort, V.; Menez, A.; Toma, F. Charybdotoxin Analogs Missing One Disulfide Bridge. In: Innovation and perspectives in solid phase synthesis : peptides, proteins and nucleic acids : biological and biomedical applications, Proceedings of the third international symposium; Epton R, Ed.; Oxford, England, UK, 31 Aug - 4 Sept 1993, Oxford, England, UK. Birmingham: Mayflower Worldwide; 1994, pp. 201-206, ISBN 0951573519 978- 0951573519.