




Article

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

Andrew J. Gregory<sup>1</sup>, Leah Voit-Ostricki<sup>1</sup>, Sándor Lovas<sup>2</sup> and Charles R. Watts<sup>1,3,\*</sup> 

<sup>1</sup> Department of Neurosurgery, Mayo Clinic Health System-Franciscan Healthcare in La Crosse, La Crosse, WI 54601, USA; ajgregory2@wisc.edu (A.J.G.); voit-ost.leah@uwlax.edu (L.V.-O.)

<sup>2</sup> Department of Biomedical Sciences, Creighton University, Omaha, NE 68178, USA; SandorLovas@creighton.edu

<sup>3</sup> Department of Neurologic Surgery, Mayo Clinic, Rochester, MN 55905, USA

\* Correspondence: charles.watts@parknicollet.com; Tel.: +1-952-993-3200; Fax: +1-952-993-7407

Received: 6 March 2019; Accepted: 10 March 2019; Published: 13 March 2019



**Abstract:** Chlorotoxin (CTX) is a 36–amino acid peptide with eight Cys residues that forms four disulfide 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- $\alpha$ -aminobutyric acid (Abu) have been previously reported. Using 4.2  $\mu$ s molecular dynamics, we compared the conformational and essential space sampling of CTX and analogs with selective substitution of the Cys residues and associated disulfide bonds with either Abu or Ser. The native and substituted peptides maintained a high degree of  $\alpha$ -helix propensity from residues 8 through 21, with the exception of substitution of the Cys<sup>5</sup>–Cys<sup>28</sup> residues with Ser and the Cys<sup>16</sup>–Cys<sup>33</sup> residues with Abu. In agreement with previous circular dichroism spectropolarimetry results, the C-terminal  $\beta$ -sheet content varied less from residues 25 through 29 and 32 through 36 and was well conserved in most analogs. The Cys<sup>16</sup>–Cys<sup>33</sup> and Cys<sup>20</sup>–Cys<sup>35</sup> disulfide-bonded residues appear to be required to maintain the  $\alpha\beta$  motif of CTX. Selective substitution with the hydrophilic Ser, may mitigate the destabilizing effect of Cys<sup>16</sup>–Cys<sup>33</sup> substitution through the formation of an inter residue H-bond from Ser<sup>16</sup>:O $\gamma$ H to Ser<sup>33</sup>:O $\gamma$ H bridged by a water molecule. All peptides shared considerable sampled conformational space, which explains the retained receptor binding of the non-native analogs.

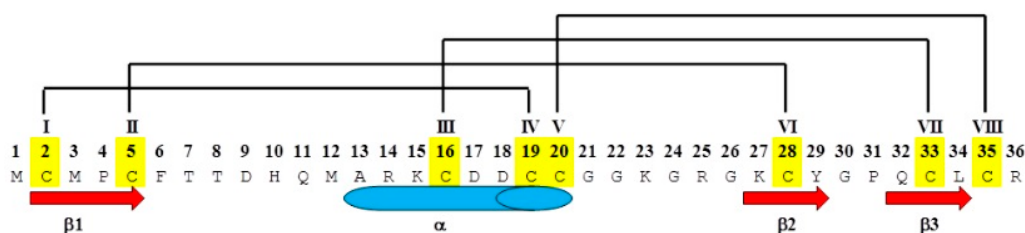
**Keywords:**  $\alpha\beta$  motif; Abu; chlorotoxin; Cys; disulfide bond; insectotoxin; isosteric substitution; L- $\alpha$ -aminobutyric acid; molecular dynamics; Ser

## 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 World Health Organization Grade IV intrinsic brain tumors, 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 eight Cys residues at positions 2, 5, 16, 19, 20, 28, 33, and 35 (I–VIII for homology modeling) that form four disulfide bonds between residues 2–19 (I–IV), 5–28 (II–VI), 16–33 (III–VII), and 20–35 (V–VIII) (Figure 1); Roman numerals in parentheses

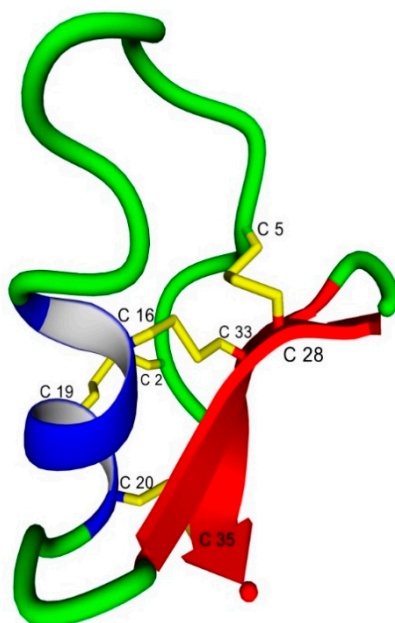
refer to the homology numbering [11–20]. The resulting secondary and tertiary structure are known as an  $\alpha\beta$  ( $\beta\alpha\beta\beta$ ) motif, a folding scaffold common to insectotoxins, insect defensins, plant  $\gamma$ -thionins, and inhibitory cystine knot peptides [11–20]. Multi sequence alignment of 20 scorpion toxin-derived peptides show that they share 49% to 88% sequence similarity [11–15]. Overlays of the three-dimensional structures of several of these peptides results in root-mean-square deviation (RMSD) less than 0.1 nm, which indicates the stability of the  $\alpha\beta$  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 (CTX), demonstrating the disulfide bond topology between Cys<sup>2</sup>–Cys<sup>19</sup> (I–IV), Cys<sup>5</sup>–Cys<sup>28</sup> (II–VI), Cys<sup>16</sup>–Cys<sup>33</sup> (III–VII), and Cys<sup>20</sup>–Cys<sup>35</sup> (V–VIII), shown as yellow with black brackets representing the associated disulfide bonds, and homology numbering shown as Roman numerals. The conserved secondary structural motifs of an  $\alpha$ -helix from Ala<sup>13</sup> to Cys<sup>20</sup>, a  $\beta$ -sheet from Cys<sup>2</sup> to Cys<sup>5</sup>, and 2 antiparallel  $\beta$ -sheets from Lys<sup>27</sup> to Tyr<sup>29</sup> and Lys<sup>32</sup> to Leu<sup>34</sup> are shown as a blue cylinder and red arrows, respectively.

The locations of the disulfide bonds within this family of peptides are important for maintaining secondary and tertiary structural features and protease resistance [16,22–26]. Despite the evolutionarily conserved disulfide bonds in this family of insectotoxins, the ability to selectively remove some of these bonds is well documented [24–28]. The II–VI bond of charybdotoxin and leiurotoxin I can be selectively removed by substitution of Cys residues with L- $\alpha$ -aminobutyric acid (Abu) without significant effects on oxidative folding [24–28]. Ojeda et al. demonstrated that selective substitution of Cys residues involving either the I–IV or II–VI disulfide bonds of CTX with Abu residues had little effect on peptide conformation, whereas the III–VII and V–VIII disulfide bonds were critical to the process of oxidative folding and obtaining natively like peptides [29]. Complete substitution of all disulfide-bonded Cys residues resulted in a peptide that maintained its biological activity with total loss of native secondary and tertiary structure and significantly increased susceptibility to serum proteases [29].

The tertiary structures of CTX in water (Protein Data Bank ID: 1CHL) has been determined using high-field <sup>1</sup>H nuclear magnetic resonance spectroscopy (NMR) (Figure 2) [30]. The presence of the N-terminal  $\beta$ -sheet from Cys<sup>2</sup> to Cys<sup>5</sup> is variable and dependent on the algorithm used to determine the structures [31,32]. In the current study, we used molecular dynamics (MD) simulations of CTX and its Abu- and Ser-substituted analogs to investigate the role of the disulfide bonds in stabilizing the  $\alpha\beta$  motif (Table 1). 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 structures.



**Figure 2.** Solution conformation of chlorotoxin (CTX) as determined by  $^1\text{H}$  nuclear magnetic resonance spectroscopy (NMR) (Protein Data Bank ID: 1CHL) demonstrating the  $\alpha$ -helix (Ala<sup>13</sup> to Cys<sup>20</sup>) and 2 antiparallel  $\beta$ -sheets (Lys<sup>27</sup> to Tyr<sup>29</sup> and Lys<sup>32</sup> to Leu<sup>34</sup>) with the four disulfide bonds between: Cys<sup>2</sup>-Cys<sup>19</sup> (I-IV), Cys<sup>5</sup>-Cys<sup>28</sup> (II-VI), Cys<sup>16</sup>-Cys<sup>33</sup> (III-VII), and Cys<sup>20</sup>-Cys<sup>35</sup> (V-VIII) residues, shown in yellow. The N-terminal  $\beta$ -sheet from Cys<sup>2</sup> to Cys<sup>5</sup> is not present in the lowest energy NMR conformation. Secondary structure elements are:  $\alpha$ -helix, blue;  $\beta$ -sheet, red; and  $\beta$ -turn/bend/coil, green. Cys side chains and associated disulfide bonds, yellow.

**Table 1.** Peptide naming convention for the Abu- and Ser-substituted chlorotoxin analogs. Native CTX sequence and disulfide bond homology numbering scheme are shown in Figures 1 and 2.

Peptide	Abu Substitution	Ser-Substitution	Homology Disulfide Bond(s) Removed
CTX			None
CTX1(A)	2, 19		I-IV
CTX1(S)		2, 19	I-IV
CTX2(A)	5, 28		II-VI
CTX2(S)		5, 28	II-VI
CTX3(A)	16, 33		III-VII
CTX3(S)		16, 33	III-VII
CTX4(A)	20, 35		V-VIII
CTX4(S)		20, 35	V-VII
CTX5(A)	2, 5, 16, 19, 20, 28, 33, 35		All
CTX5(S)		2, 5, 16, 19, 20, 28, 33, 35	All

## 2. Results

### 2.1. General Properties

The  $\text{RMSD}_{\text{CTX}}$ ,  $\text{RMSD}_{\text{AVG}}$ , and fractions of sampled DSSP secondary structure ( $\alpha$ -helix,  $\beta$ -sheet, and  $\beta$ -bend/turn), for CTX and Abu- and Ser-substituted analogs are given in Table 2. The values of  $\text{RMSD}_{\text{CTX}}$  and  $\text{RMSD}_{\text{AVG}}$  were similar between CTX and all analogs (Table 1), with the exception of CTX5(A) and CTX5(S), which had lower  $\text{RMSD}_{\text{AVG}}$  values.

**Table 2.** Calculated (mean  $\pm$  SD) C $\alpha$ -trace RMSDs between the peptide and the average conformation of CTX, between the Abu- or Ser-substituted analog and the average conformation of itself, and the fraction ( $\rho$ ) of sampled DSSP secondary structure.

Peptide	RMSD/nm		$\rho$		
	CTX <sup>a</sup>	AVG <sup>b</sup>	$\alpha$ -Helix	$\beta$ -Sheet	$\beta$ -Turn/Bend
CTX	0.73 $\pm$ 0.02	0.73 $\pm$ 0.02	0.31 $\pm$ 0.05	0.16 $\pm$ 0.02	0.17 $\pm$ 0.05
CTX1(A)	0.71 $\pm$ 0.02	0.74 $\pm$ 0.02	0.39 $\pm$ 0.09	0.14 $\pm$ 0.05	0.11 $\pm$ 0.09
CTX1(S)	0.73 $\pm$ 0.02	0.73 $\pm$ 0.02	0.27 $\pm$ 0.04	0.20 $\pm$ 0.03	0.26 $\pm$ 0.02
CTX2(A)	0.73 $\pm$ 0.02	0.73 $\pm$ 0.02	0.23 $\pm$ 0.07	0.16 $\pm$ 0.03	0.31 $\pm$ 0.04
CTX2(S)	0.73 $\pm$ 0.02	0.74 $\pm$ 0.02	0.34 $\pm$ 0.08	0.16 $\pm$ 0.08	0.08 $\pm$ 0.07
CTX3(A)	0.73 $\pm$ 0.02	0.76 $\pm$ 0.02	0.31 $\pm$ 0.07	0.19 $\pm$ 0.06	0.13 $\pm$ 0.14
CTX3(S)	0.73 $\pm$ 0.02	0.72 $\pm$ 0.02	0.29 $\pm$ 0.06	0.13 $\pm$ 0.06	0.22 $\pm$ 0.08
CTX4(A)	0.73 $\pm$ 0.02	0.72 $\pm$ 0.02	0.32 $\pm$ 0.08	0.16 $\pm$ 0.04	0.19 $\pm$ 0.07
CTX4(S)	0.72 $\pm$ 0.02	0.72 $\pm$ 0.02	0.30 $\pm$ 0.06	0.14 $\pm$ 0.04	0.27 $\pm$ 0.06
CTX5(A)	0.72 $\pm$ 0.02	0.61 $\pm$ 0.02	0.28 $\pm$ 0.09	0.03 $\pm$ 0.05	0.21 $\pm$ 0.14
CTX5(S)	0.73 $\pm$ 0.02	0.65 $\pm$ 0.02	0.30 $\pm$ 0.09	0.05 $\pm$ 0.06	0.09 $\pm$ 0.10

<sup>a</sup> C $\alpha$ -trace comparison between the peptide and the average (AVG) conformation of CTX; <sup>b</sup> C $\alpha$ -trace comparison between the peptide and the average (AVG) conformation of itself.

The fraction of  $\alpha$ -helix is retained across all analogs with the lowest fraction corresponding to CTX2(A) and the highest to CTX1(A) (Table 1). Likewise, the fraction of  $\beta$ -sheet is retained in the analogs where a single disulfide bond is substituted but markedly decreased in the CTX5(A) and CTX5(S) analogs. The greatest degree of variability in global secondary structure occurs for  $\beta$ -turn/bend. Although the fraction of  $\alpha$ -helix and  $\beta$ -turn/bend for CTX5(A) was 0.28 and 0.21, respectively, the fraction of random coil was 0.48, which is consistent with previously published CD spectra [29].

## 2.2. DSSP Secondary Structure

The residue fractions of sampled DSSP secondary structure for CTX and each of the Abu- and Ser-substituted analogs are shown in Figure 3. The secondary structure for CTX was consistent with its <sup>1</sup>H NMR solution conformation, having 3 prominent  $\beta$ -turns/bends in the N-terminal domain from residues 6 through 9 and 12 through 13, an  $\alpha$ -helix from residues 15 through 20, and 2 antiparallel  $\beta$ -sheets from residues 27 through 29 and 32 through 34, with intervening  $\beta$ -turns/bends from residues 21 through 24 and 30 through 31. The N-terminal predicted  $\beta$ -sheet from residues 2 through 5 was absent. For all single disulfide bond-substituted analogs, the  $\beta$ 2,  $\beta$ 3 antiparallel region is well preserved. The  $\alpha$ -helix, although present, is more variable in its location, length, and populated fraction. The CTX5(A) and CTX5(S) analogs demonstrate disruption of the native CTX  $\alpha\beta$  motif with elongation of the N-terminal  $\alpha$ -helix, loss of the  $\beta$ 2,  $\beta$ 3 antiparallel region and increases in  $\beta$ -turn/bend sampling throughout.

## 2.3. Structural Flexibility

The C $\alpha$ -trace RMSFs comparing each of the analogs with the average conformation of CTX are shown in Figure 3. There were only minor differences in the C $\alpha$ -trace RMSF between CTX and that of its average conformation. The most flexible regions of CTX correspond to the  $\beta$ -turn/bend regions from residues 8 through 9, 12 through 13, 21 through 24, and 30 through 31. All single disulfide bond Abu-substituted analogs demonstrated increased residue flexibility compared to native CTX, while the Ser-substituted analogs did so to a much lesser degree. The decrease in C $\alpha$ -trace RMSF was most marked for the CTX1(S) and CTX4(S) analogs corresponding to a lengthening of the  $\alpha$ -helix region. As expected, the CTX5(A) and CTX5(S) analogs show the greatest degree of C $\alpha$ -trace RMSF deviation from native CTX.

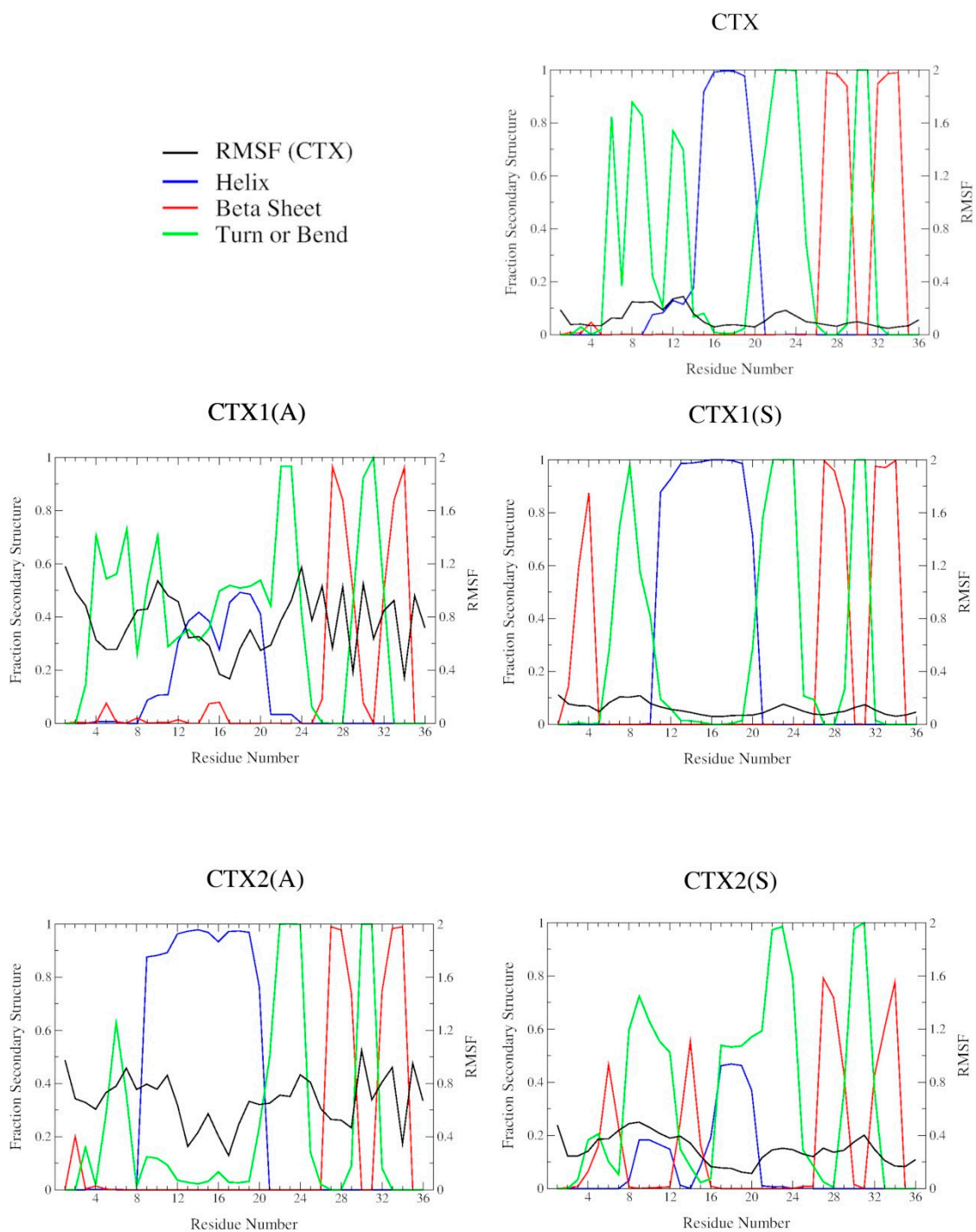
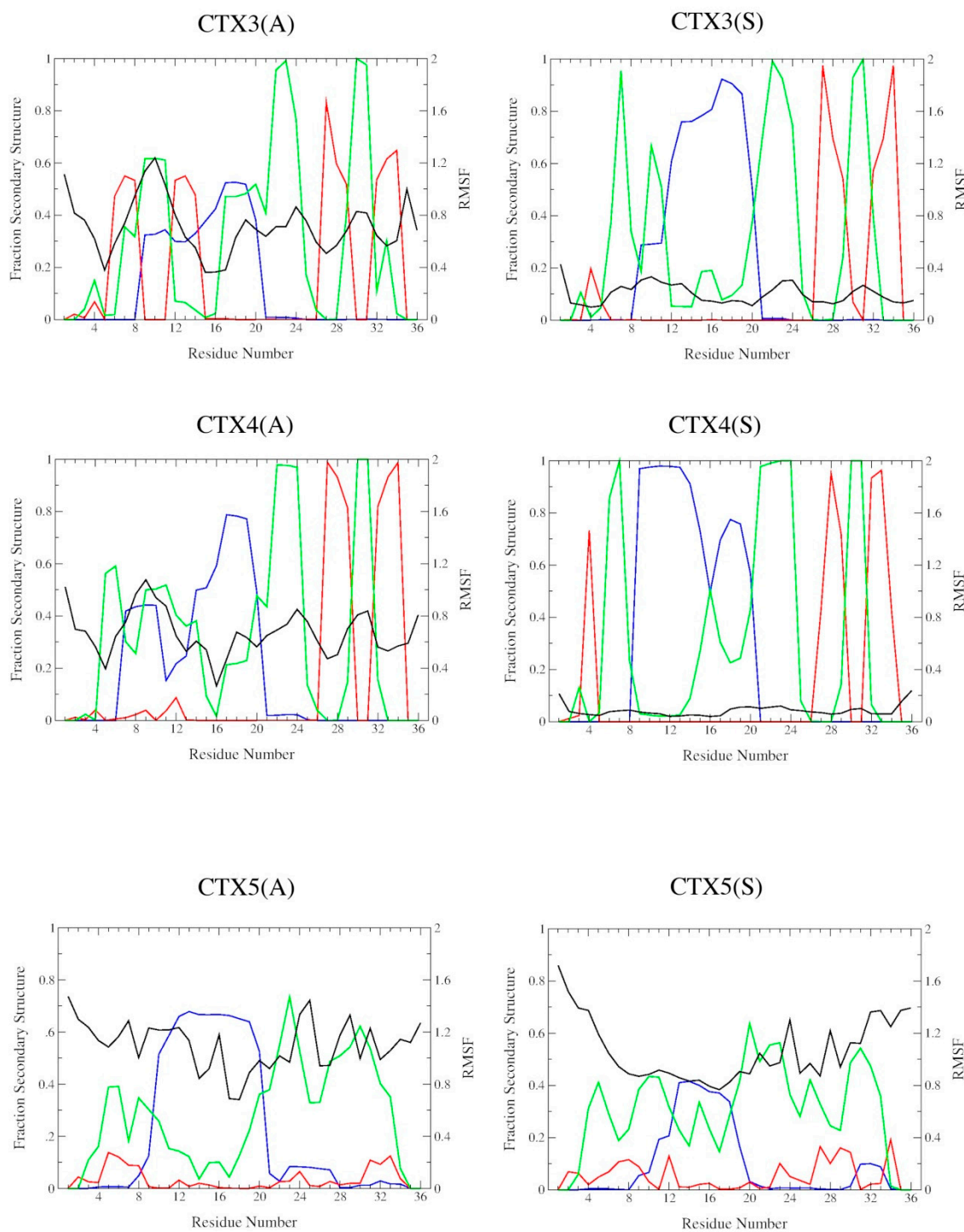


Figure 3. Cont.





**Figure 3.** Fractions of sampled DSSP secondary structure ( $\alpha$ -helix, blue;  $\beta$ -sheet, red; and  $\beta$ -turn/bend, green) as a function of residue number and  $C\alpha$ -trace root-mean-square fluctuation (RMSF) from the time average conformation of CTX (RMSF, black) for CTX and its Abu- and Ser-substituted analogs.

#### 2.4. Interactions

The backbone–backbone (BB–BB) and sidechain–sidechain (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 2–9; 5–28; 16–33; or 20–35 disulfide-bonded residues. The SC–SC contacts between the N-terminal residues with the  $\alpha$ -helix from residues 15 through 20 and the 2 C-terminal  $\beta$ -sheets from residues 27 through 29 and 32 through 34 are also present. The only

long-distance BB–BB contacts are between the 2 antiparallel  $\beta$ -sheets and interactions between residues 4 and 5 with 31 through 33. In CTX1(A) both SC–SC and BB–BB contacts are increased. The removal of the Cys<sup>2</sup>-Cys<sup>19</sup> disulfide bond facilitated increased interactions between the N-terminal region, the  $\alpha$ -helix, and the C-terminal antiparallel  $\beta$ -sheets. There were also increased BB–BB interactions between the  $\alpha$ -helix from residues 12 through 17 and the proximal  $\beta$ -sheet from residues 24 through 28 and between the N-terminal 8 residues and residues 28 through 32. The SC–SC and BB–BB contact probabilities in CTX1(S) and CTX2(A) were similar to those in CTX. Contacts in CTX2(S) were similar to those in CTX1(A) except that the interaction between the N-terminal residues and  $\alpha$ -helix with the proximal  $\beta$ -sheet were much less pronounced and an interaction was present between the N-terminal residues 1 through 8 and the  $\alpha$ -helix from residues 13 through 19. Contact probability maps for CTX3(A) and CTX3(S) were similar and shared similarity with CTX2(S). Likewise, CTX4(A) and CTX4(S) were similar to CTX. CTX5(A) had sparse SC–SC interactions and low-frequency BB–BB interactions between multiple residues. CTX5(S) was the exact opposite, with few, if any, BB–BB interactions and sparsely populated SC–SC interactions.

To determine the degree conformational stability conferred by each disulfide bond on CTX and its Abu- and Ser-substituted analogs, the  $C\beta H\beta^{1,2}-C\beta H\beta^{1,2}$  center-of-mass distances between residue pairs and probabilities of contact  $\leq 0.26$  nm are shown in Table 3. The  $C\beta H\beta^{1,2}-C\beta H\beta^{1,2}$  center-of-mass distance was used because with <sup>1</sup>H NMR, the method of structural determination used for CTX, S–S interatomic distances cannot be assigned in spectra. The distance geometry algorithm calculation of peptide conformations is therefore dependent on the presence of bonded CysH $\beta^{1,2}$ -CysH $\beta^{1,2}$  nuclear Overhauser effect spectral peaks. Removal of a disulfide bond in most of the substituted peptides resulted in significant increases in the  $C\beta H\beta^{1,2}-C\beta H\beta^{1,2}$  center-of-mass distances and decreases in the probabilities of contact, with the exception of CTX3(S) and CTX4(A).

**Table 3.** The  $C\beta H\beta^{1,2}-C\beta H\beta^{1,2}$  center-of-mass distances (D) in nm (mean $\pm$ SD) between residue pairs for CTX and its Abu- and Ser-substituted analogs given in nm. Probabilities ( $\rho$ ) of contact  $\leq 0.26$  nm are in parenthesis. The Abu- and Ser-substituted residues are highlighted with grey background <sup>a</sup>

Peptide	D ( $\rho$ )			
	2–19	5–28	16–33	20–35
CTX	0.24 $\pm$ 0.01 (0.98)	0.21 $\pm$ 0.01 (0.99)	0.26 $\pm$ 0.01 (0.71)	0.21 $\pm$ 0.01 (1.00)
CTX1(A)	0.45 $\pm$ 0.08 (0.02)	0.22 $\pm$ 0.02 (1.00)	0.21 $\pm$ 0.01 (1.00)	0.21 $\pm$ 0.01 (1.00)
CTX1(S)	0.43 $\pm$ 0.04 (0.00)	0.22 $\pm$ 0.02 (1.00)	0.21 $\pm$ 0.01 (1.00)	0.21 $\pm$ 0.01 (1.00)
CTX2(A)	0.22 $\pm$ 0.02 (0.97)	0.49 $\pm$ 0.08 (0.00)	0.24 $\pm$ 0.02 (0.84)	0.21 $\pm$ 0.01 (1.00)
CTX2(S)	0.23 $\pm$ 0.02 (0.89)	0.46 $\pm$ 0.11 (0.01)	0.21 $\pm$ 0.02 (0.98)	0.21 $\pm$ 0.01 (1.00)
CTX3(A)	0.24 $\pm$ 0.02 (0.67)	0.22 $\pm$ 0.01 (1.00)	0.33 $\pm$ 0.07 (0.13)	0.21 $\pm$ 0.01 (1.00)
CTX3(S)	0.24 $\pm$ 0.02 (0.77)	0.22 $\pm$ 0.01 (0.99)	0.26 $\pm$ 0.05 (0.70)	0.21 $\pm$ 0.01 (1.00)
CTX4(A)	0.22 $\pm$ 0.02 (0.99)	0.22 $\pm$ 0.01 (1.00)	0.24 $\pm$ 0.02 (0.80)	0.22 $\pm$ 0.03 (0.90)
CTX4(S)	0.23 $\pm$ 0.01 (0.99)	0.21 $\pm$ 0.01 (1.00)	0.23 $\pm$ 0.01 (1.00)	0.30 $\pm$ 0.08 (0.39)
CTX5(A)	0.62 $\pm$ 0.28 (0.04)	0.65 $\pm$ 0.22 (0.01)	0.55 $\pm$ 0.18 (0.01)	0.62 $\pm$ 0.19 (0.02)
CTX5(S)	0.78 $\pm$ 0.30 (0.01)	0.69 $\pm$ 0.28 (0.05)	0.70 $\pm$ 0.21 (0.00)	0.75 $\pm$ 0.25 (0.00)

<sup>a</sup> The  $C\beta H\beta^{1,2}-C\beta H\beta^{1,2}$  center-of-mass distance for the (Acetyl-Ala-Cys-Ala-NH<sub>2</sub>)<sub>2</sub> dipeptide, as described in Appendix A; D was 0.23  $\pm$  0.03 nm and one-sided Cl<sub>95</sub> was  $\leq 0.26$  nm.

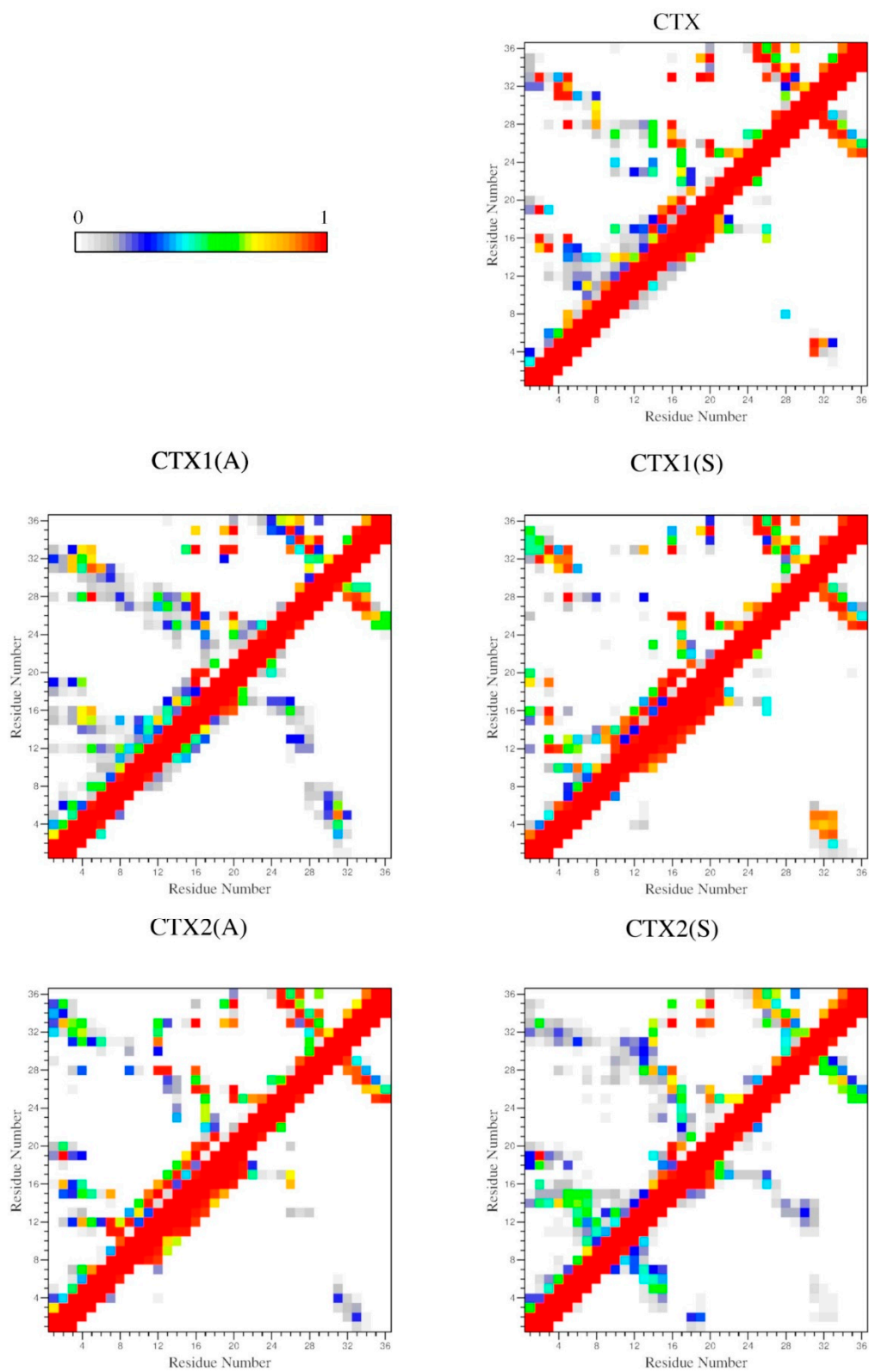
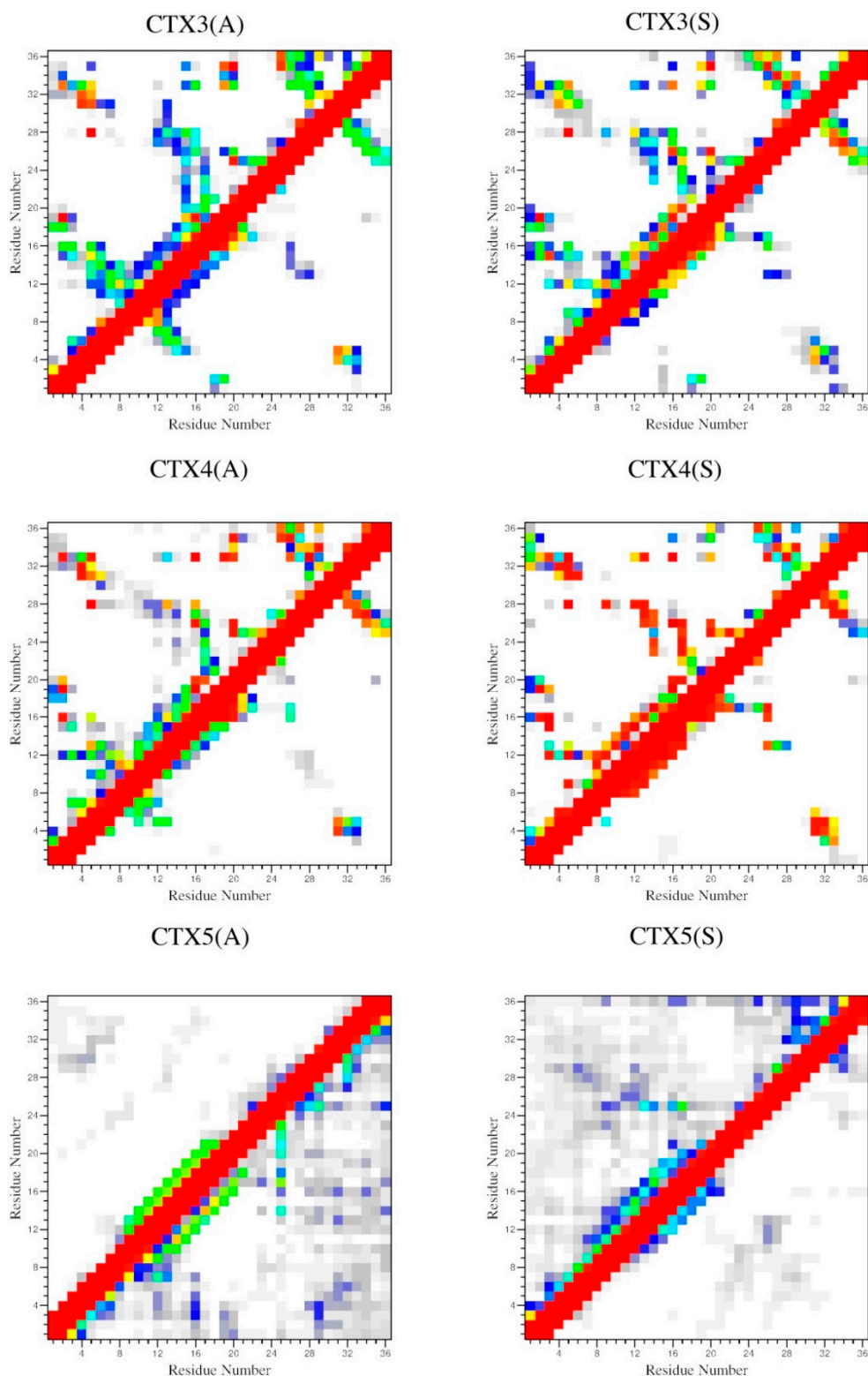


Figure 4. Cont.





**Figure 4.** Probability maps for backbone–backbone contacts (below the diagonal) and sidechain–sidechain contacts (above the diagonal) as a function of residue number for CTX and its Abu- and Ser-substituted analogs. The color scale represents the relative probability of contact.

The probabilities of finding water molecules within 0.5 nm of the terminal  $\gamma$ -sidechain atoms ( $S\gamma$ ,  $C\gamma H\gamma$ , and  $O\gamma H\gamma$ ) of Cys, Abu, and Ser respectively, and the residue relative solvent-accessible surface area (rSASA) of the residues are shown in Table 4. The terminal  $S\gamma$  of CTX remained relatively

solvent shielded, regardless of its position within the native peptide with the exception of the Cys<sup>2</sup>, Cys<sup>19</sup> and Cys<sup>35</sup> residues. The Cys<sup>2</sup> and Cys<sup>35</sup> residues are located at the more solvent-accessible N- and C-termini respectively while the Cys<sup>19</sup> residue is on the solvent exposed surface of the N-terminal  $\alpha$ -helix, Figure 2. Substitution of disulfide-bonded Cys residues with either a hydrophobic Abu or hydrophilic Ser at the 2,19 (I–IV), 5,28 (II–VI), and 20,35 (V–VIII) positions tended to increase the probability of adjacent water molecules at the involved residues and increase the rSASA. This local relationship does not however hold for the 16,33 (III–VII) disulfide bond. For the hydrophobic Abu-substituted CTX3(A) there is the expected increase in probability of adjacent water and rSASA for the substituted residues but also a slight increase for Cys<sup>2</sup> and more interestingly, decreases for Cys<sup>19</sup>. For the hydrophilic Ser-substituted CTX3(S), the substituted residues demonstrate an increase in the probability of an adjacent water molecules and a significant decrease in rSASA indicating that the interacting water molecules may be sequestered from the bulk solvent. The CTX5(A), and CTX5(S) peptides showed significant increases in the probability of adjacent water molecules and increases in rSASA through the substituted residues.

**Table 4.** Probability of finding water molecules within 0.5 nm of the terminal sidechain  $\gamma$  atoms of Cys-, Abu-, and Ser- substituted residues ( $S_\gamma$ ,  $C_\gamma H_\gamma$ , and  $O_\gamma H_\gamma$ , respectively). The relative solvent-accessible surface area (rSASA) of the residues are in parenthesis. The Abu- and Ser-substituted residues are highlighted with grey background <sup>a,b,c,d,e</sup>

Peptide	Residue							
	2	5	16	19	20	28	33	35
CTX	0.066 (0.27)	0.044 (0.09)	0.013 (0.02)	0.076 (0.38)	0.020 (0.14)	0.049 (0.10)	0.019 (0.11)	0.094 (0.45)
CTX1(A)	0.173 (0.79)	0.016 (0.08)	0.012 (0.01)	0.113 (0.50)	0.022 (0.15)	0.037 (0.11)	0.017 (0.15)	0.086 (0.44)
CTX1(S)	0.147 (0.56)	0.070 (0.16)	0.028 (0.01)	0.121 (0.55)	0.018 (0.15)	0.087 (0.41)	0.041 (0.04)	0.081 (0.38)
CTX2(A)	0.077 (0.24)	0.144 (0.56)	0.013 (0.01)	0.084 (0.42)	0.015 (0.16)	0.060 (0.30)	0.012 (0.07)	0.081 (0.41)
CTX2(S)	0.086 (0.27)	0.186 (0.50)	0.033 (0.04)	0.066 (0.27)	0.027 (0.17)	0.104 (0.30)	0.032 (0.17)	0.079 (0.40)
CTX3(A)	0.090 (0.33)	0.035 (0.10)	0.078 (0.15)	0.059 (0.23)	0.025 (0.15)	0.044 (0.18)	0.053 (0.24)	0.070 (0.32)
CTX3(S)	0.079 (0.37)	0.032 (0.09)	0.050 (0.02)	0.061 (0.30)	0.029 (0.17)	0.048 (0.22)	0.060 (0.14)	0.086 (0.41)
CTX4(A)	0.075 (0.22)	0.032 (0.06)	0.031 (0.02)	0.093 (0.46)	0.038 (0.15)	0.052 (0.25)	0.024 (0.07)	0.111 (0.49)
CTX4(S)	0.082 (0.33)	0.006 (0.05)	0.024 (0.00)	0.076 (0.48)	0.046 (0.13)	0.026 (0.20)	0.025 (0.03)	0.165 (0.56)
CTX5(A)	0.101 (0.69)	0.138 (0.61)	0.101 (0.35)	0.138 (0.64)	0.126 (0.58)	0.124 (0.54)	0.123 (0.53)	0.138 (0.59)
CTX5(S)	0.203 (0.76)	0.194 (0.72)	0.156 (0.44)	0.166 (0.65)	0.202 (0.75)	0.172 (0.53)	0.182 (0.62)	0.204 (0.71)

<sup>a</sup> The probability of finding water molecules within 0.5 nm of the terminal  $\gamma$ -sidechain atom of Cys ( $S_\gamma$ ) in the (Acetyl-Ala-Cys-Ala-NH<sub>2</sub>)<sub>2</sub> dipeptide, as described in Appendix A Materials, is 0.114; <sup>b</sup> The probability of finding water molecules within 0.5 nm of the terminal  $\gamma$ -sidechain atoms of Abu ( $C_\gamma H_\gamma$ ) in the Acetyl-Ala-Abu-Ala-NH<sub>2</sub> peptide, as described in Appendix A, is 0.190; <sup>c</sup> The probability of finding water molecules within 0.5 nm of the terminal  $\gamma$ -sidechain atoms of Ser ( $O_\gamma H_\gamma$ ) in the Acetyl-Ala-Ser-Ala-NH<sub>2</sub> peptide, as described in Appendix A, is 0.213; <sup>d</sup> The rSASA is the ratio of the residue solvent exposed surface area (SASA) to the residue maximum solvent exposed surface area (mSASA) in (Acetyl-Ala-Xaa-Ala-NH<sub>2</sub>) peptide where Xaa is either Abu or Ser as described in the Methods and Appendix A. The mSASA for Abu was:  $1.347 \pm 0.121 \text{ nm}^2$ . The mSASA for Ser was:  $1.219 \pm 0.111 \text{ nm}^2$ ; <sup>e</sup> The rSASA for Cys is the ratio of the residue solvent exposed surface area (SASA) to the residue maximum solvent exposed surface area (mSASA) in (Acetyl-Ala-Cys-Ala-NH<sub>2</sub>)<sub>2</sub> dipeptide as described in the Methods and Appendix A. The mSASA for Cys was:  $0.758 \pm 0.142 \text{ nm}^2$ .

### 2.5. Conformational Analysis

The lowest energy conformations were determined by projecting each trajectory onto the first 2 dihedral principal (dPC) components, as show in Figure 5. The lowest energy conformation of CTX was consistent with the  $^1\text{H}$  NMR structure as follows:  $\alpha$ -helix from residues 15–20, 2 antiparallel  $\beta$ -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<sup>14</sup> and Asp<sup>17</sup>, Lys<sup>15</sup> and Asp<sup>18</sup>, Arg<sup>25</sup> and the C-terminus, and Lys<sup>27</sup> and the C-terminus.

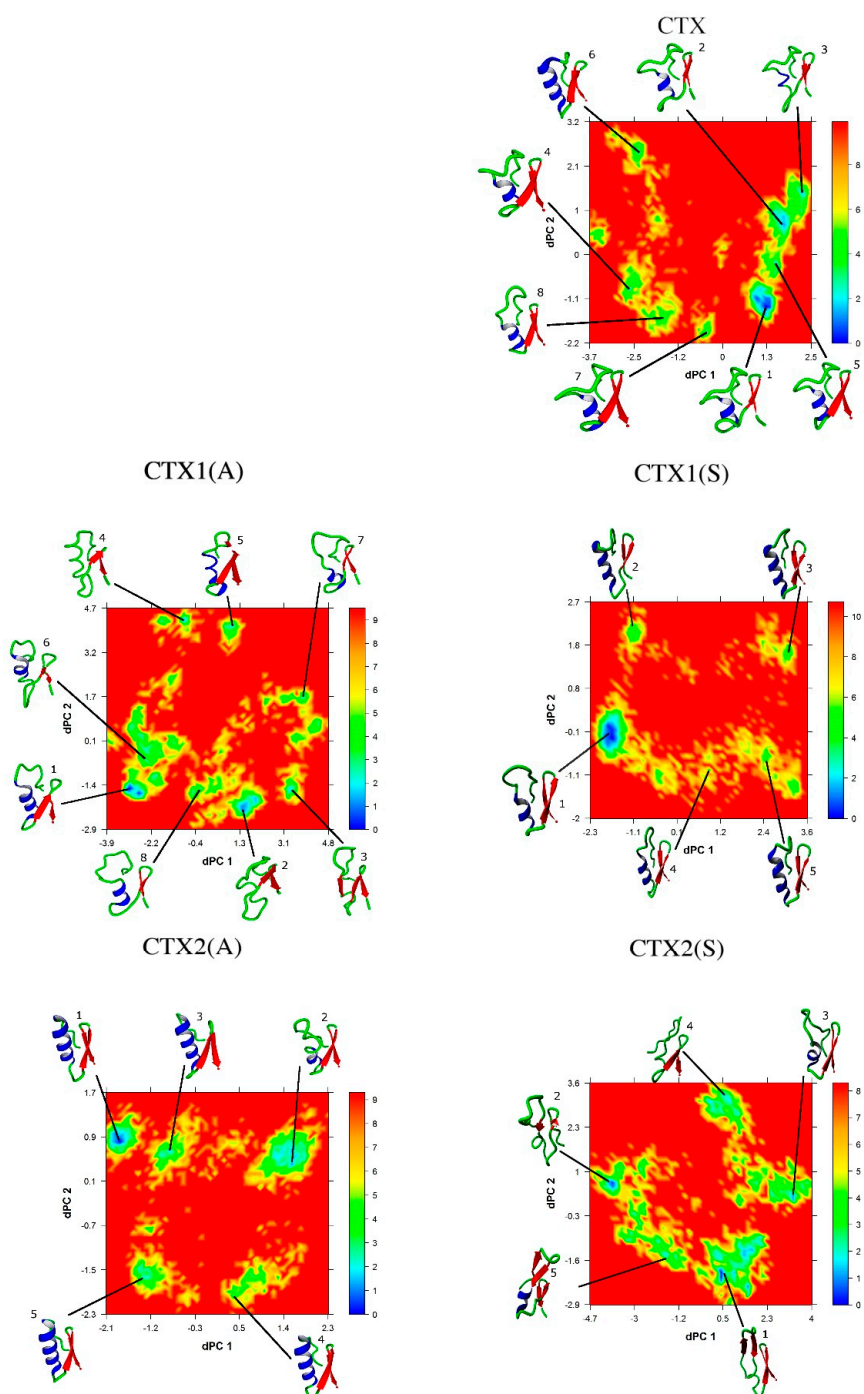
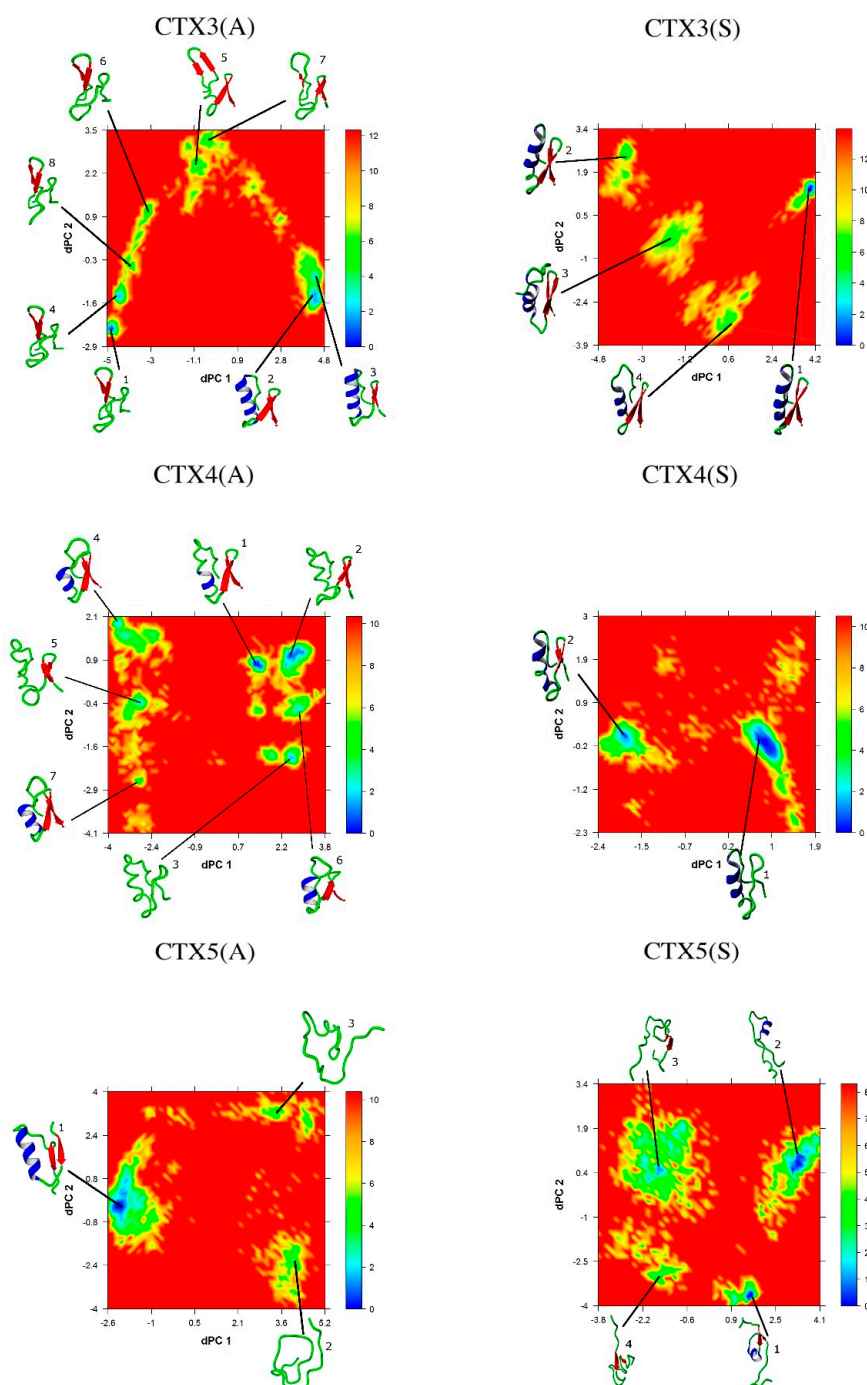


Figure 5. Cont.



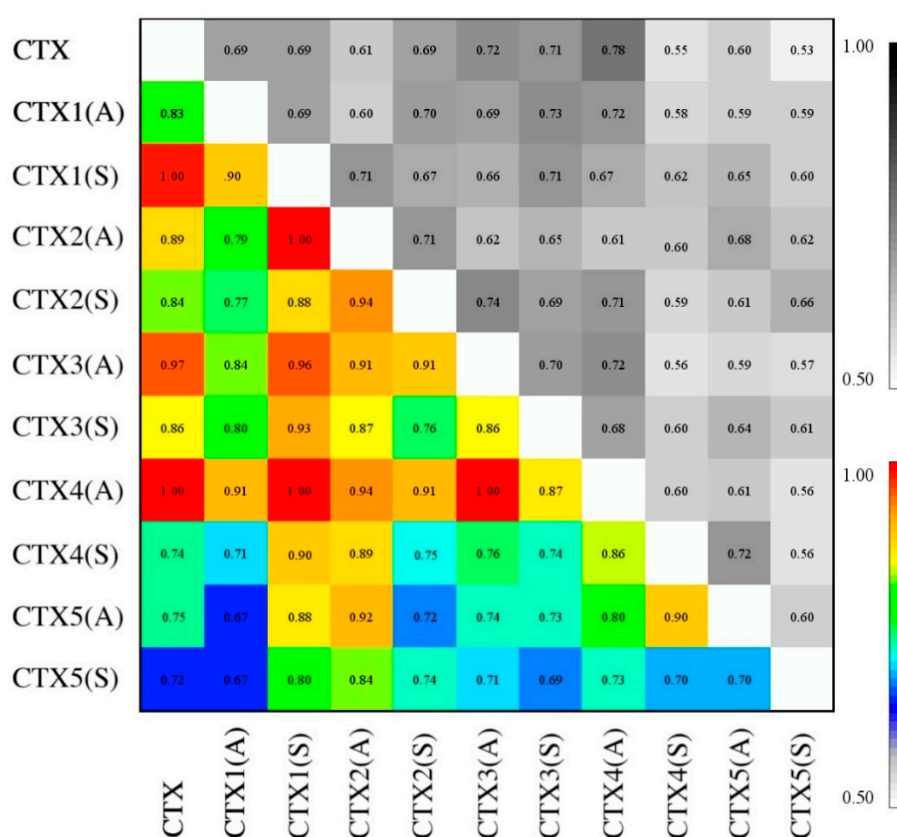
**Figure 5.** The free energy landscape ( $\text{kJ}\cdot\text{mol}^{-1}$ ) as a function of the first two dihedral principal components (dPC1 and dPC2). The lowest energy conformations, as determined by cluster analysis, are shown and numbered in order from lowest to highest relative energy. The conformations have been rotated to optimally display the secondary structure elements and interactions. Secondary structure elements show as: red  $\beta$ -sheet, blue  $\alpha$ -helix, and green turn/bend/coil.

The lowest energy conformations of CTX1(A), CTX1(S), CTX2(A), and CTX3(S) maintain secondary and tertiary structures that are consistent with the  $\alpha\beta$  motif of native CTX. This is despite the finding that the sidechains of substituted residues remain separated and do not interact (Table 3). The CTX3(S) peptide is unique in that the two Ser residues are in close proximity (Table 3) and most likely interact by hydrogen bonding to each other or through an associated water molecule that is sequestered from the solvent, given the significant increase in  $\rho_{g(r)}$  and low rSASA (Table 4). The lowest

energy conformations of CTX2(S), CTX3(A), CTX4(A), CTX4(S), CTX5(A), and CTX5(S) deviate significantly from the  $\alpha\beta$  motif and 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 CTX2(S), CTX3(A), CTX4(S), CTX5(A), and CTX5(S), 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:  $0.65 \leq \text{nRMSIP} < 0.75$ ,  $0.75 \leq \text{nRMSIP} < 0.85$ , and  $0.85 \leq \text{nRMSIP} < 0.95$ , and  $0.95 \leq \text{nRMSIP}$ . Peptide pairwise comparisons sharing the greatest degree of subspace conformational sampling were CTX to CTX1(S), CTX to CTX3(A), CTX to CTX4(A), CTX1(S) to CTX2(A), CTX1(S) to CTX3(A), and CTX3(A) to CTX4(A).



**Figure 6.** Comparison of essential subspace sampling. Root-mean-square inner product (RMSIP, top right in grayscale) and normalized RMSIP (bottom left in color) matrix of trajectories of molecular dynamics simulation of CTX and its Abu- and Ser- substituted analogs. By definition, the diagonal values equal 1.0, but not color-coded for clarity.

### 3. Discussion

We performed  $\mu\text{s}$ -scale MD simulations to compare the conformational and essential space sampling of CTX and its analogs with selective substitution of the Cys residues in disulfide bonds with either Abu or Ser. 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 [27,33]. Because of the small size and highly charged nature of these peptides, the surrounding environment of the disulfide bonds may not be as hydrophobic as previously considered. In the current MD simulations, we were able to give detailed descriptions of the local hydrophobic/hydrophilic environment of each



disulfide bond. Examination of the probability of water contact and rSASA data (Table 4) demonstrates that some of the Cys residues are relatively solvent shielded (5, 16, 20, 28, and 33) while others are more exposed (2, 19, and 35). For Cys<sup>2</sup> and Cys<sup>33</sup> the degree of solvent exposure and hydration is secondary to their location at the N- and C-termini respectively while Cys<sup>19</sup> is more solvent exposed due to its location on the outside surface of the N-terminal  $\alpha$ -helix. For Cys<sup>5</sup>, there may be some degree of local solvent shielding contribution secondary to the adjacent hydrophobic Pro<sup>4</sup> and aromatic Phe<sup>5</sup> residues. The remaining Cys residues (16 and 20) however are adjacent to charged residues such as Lys, Asp, or Arg, with the exception of Cys<sup>28</sup> which is adjacent to the charged Lys<sup>27</sup> and aromatic Tyr<sup>29</sup>, and Cys<sup>33</sup> which is adjacent to a hydrophilic Gln<sup>32</sup> and hydrophobic Leu<sup>34</sup>.

Replacement of the disulfide-bonded Cys<sup>2</sup> and Cys<sup>19</sup> residues with either Abu or Ser has only minor effects on CTX secondary and tertiary conformations. Removal of this bond is well tolerated because the remaining Cys<sup>5</sup>-Cys<sup>28</sup> disulfide bond maintains the N-terminus in close proximity with the antiparallel  $\beta$ -sheet. This interaction is also stabilized with an H-bond between the amine of Cys<sup>5</sup> and the carbonyl of Pro<sup>31</sup>. Subtle differences are present between the Abu- and Ser- substituted analogs; CTX1(A) and CTX1(S). Abu substitution facilitates wide separation of the N-terminus from the  $\alpha$ -helix and  $\beta$ -sheet. Ser-substitution allows for close approximation of the N-terminus to the distal strand of the  $\beta$ -sheet, forming a 3-strand antiparallel  $\beta$ -sheet. The CTX1(S) structure is stabilized by H-bonds between the amine of Cys<sup>5</sup> and the carbonyl of Pro<sup>31</sup>, the amine of Cys<sup>33</sup> and the carbonyl of Met<sup>3</sup>, the amine of Met<sup>3</sup> and the carbonyl of Cys<sup>33</sup>, and N $\epsilon$ H of Gln<sup>32</sup> and O $\gamma$  of Ser<sup>3</sup>. The sidechain of the Ser<sup>19</sup> residue on the  $\alpha$ -helix is positioned so that it can interact with the surrounding solvent. Substitution of the Cys<sup>2</sup>-Cys<sup>19</sup> disulfide bond with Ser residues most likely removes some degree of tertiary strain within the peptide. Furthermore, it allows more H-bonding interactions between the N-terminus and the antiparallel  $\beta$ -sheet. This is consistent with the disulfide bond between Cys<sup>2</sup>-Cys<sup>19</sup> being more solvent exposed and thus favoring potential hydrophilic substitutions.

Substitution of the disulfide-bonded Cys<sup>5</sup> and Cys<sup>28</sup> residues allows the  $\alpha$ -helix region to be lengthened from residues 9 through 20. The preserved disulfide bond between Cys<sup>2</sup> and Cys<sup>19</sup>, however, maintains the N-terminal region in close proximity with the C-terminal  $\beta$ -sheets. Due to elongation of the  $\alpha$ -helix, the N-terminal region is not oriented in a way that would allow H-bonding to the distal strand of the  $\beta$ -sheet. CTX2(A) places Abu<sup>5</sup> and Abu<sup>28</sup> residues within a hydrophobic cleft between the  $\alpha$ -helix and the proximal strand of the antiparallel  $\beta$ -sheets interacting with the sidechain of Met<sup>12</sup>. This preserves the hydrophobic region present in CTX but in a slightly less compact conformation. The conformational importance of this hydrophobic region is demonstrated by Ser-substitution, CTX2(S), which results in significant disruption of the  $\alpha$ -helix with a large loop extending from Ser<sup>5</sup> through Lys<sup>15</sup> and a shift in the location of the antiparallel  $\beta$ -sheets 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<sup>16</sup> and Cys<sup>33</sup>, substitution of these residues with the hydrophobic Abu is not tolerated. This is supported by the detected non-native isomers in oxidative folding experiments [28,29]. There is significant structural disruption in the CTX3(A) peptide. The lack of a disulfide bond between Cys<sup>16</sup> and Cys<sup>33</sup> disrupts the association between the  $\alpha$ -helical region and the terminal strand of the antiparallel  $\beta$ -sheets. This is particularly important because the sidechains of the antiparallel  $\beta$ -sheet residues are oriented away from the  $\alpha$ -helix and toward the solvent, and with the exception of Abu<sup>16</sup> and Cys<sup>20</sup>, none of the  $\alpha$ -helix residues are aligned to interact with the back surface of the  $\beta$ -sheet. A lack of H-bond donor and acceptor groups between the  $\alpha$ -helix and  $\beta$ -sheet regions also makes the association between the  $\alpha$ -helix and  $\beta$ -sheet strongly dependent on the presence of the disulfide bond between residues Cys<sup>16</sup> and Cys<sup>33</sup>. This substitution also results in closer proximity of the Cys<sup>2</sup>-Cys<sup>19</sup>, Cys<sup>5</sup>-Cys<sup>28</sup>, and Cys<sup>20</sup>-Cys<sup>35</sup> disulfide bonds, which may explain why multiple isomers occur with synthesis [29]. Ser-substitution of these residues, CTX3(S), however, preserves the overall structure.

The elongation of the  $\alpha$ -helix appears to stabilize interaction between the two Ser residues and an associated water molecule.

Substitution at the Cys<sup>20</sup> and Cys<sup>35</sup> residues appears to allow preservation of the overall structure with some degree of elongating of the  $\alpha$ -helical region. Although the overall fold of the peptide is preserved, multiple H-bonds occur between the  $\alpha$ -helix and  $\beta$ -sheets that do not occur in the native peptide, and the resulting fold places the Cys<sup>2</sup>–Cys<sup>19</sup>, Cys<sup>5</sup>–Cys<sup>28</sup>, and Cys<sup>16</sup>–Cys<sup>33</sup> residues in close proximity and, like the Cys<sup>16</sup>–Cys<sup>33</sup> substitution, may explain why multiple isomers occur with synthesis. The result for Ser-substitution is similar, except that the C-terminal antiparallel  $\beta$ -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 [34–46]. The Lennard-Jones, electrostatic potentials, bond, bond angle, and torsional parameters for the Abu residue were assigned based on the similarity and transferability of force field parameters for similar noncyclic aliphatic residues (Ala, Leu, Ile, and Val) within the CHARMM36m force field using the same methodology as the CGenFF program and are provided in Appendix A [34–38]. The initial structure of CTX for the simulation was the first conformation of the published <sup>1</sup>H NMR solution structure (Protein Data Bank ID: 1CHL) [30]. Starting structures of the CTX Abu- and Ser- substituted analogs as given in Table 1 were obtained by changing the respective residues from Cys to Abu or Ser using YASARA [47]. 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<sup>−</sup> and Na<sup>+</sup> 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 CTX5(A) and CTX5(S), 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, 1000 kJ·mol<sup>−1</sup>) 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. [48,49]. The relaxation constant was 0.1 picosecond and the isothermal compressibility was  $4.5 \times 10^{-5}$  bar<sup>−1</sup>. 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° [50,51]. 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 [52]. 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  $\mu$ s 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 [48,53]. 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 simulation were considered system equilibration, and the subsequent 4  $\mu$ s was used for analysis with a sampling frequency of 0.1 ns.

#### 4.3.1. Biophysical Properties

The C $\alpha$ -trace RMSD from the <sup>1</sup>H NMR solution conformation of CTX (RMSD<sub>CTX</sub>), C $\alpha$ -trace RMSD from the average sampled peptide conformation (RMSD<sub>AVG</sub>), per-residue C $\alpha$ -trace root mean square fluctuation (RMSF) from the average sampled peptide conformation, global and per-residue fraction of sampled secondary structure ( $\alpha$ -helix,  $\beta$ -sheet, and  $\beta$ -bend/turn *g\_rmsd*, *g\_rmsf*, and *do\_dssp*, utilities of GROMACS, respectively [31,45].

#### 4.3.2. Interactions

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 [44]. 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. <sup>1</sup>H NMR cannot measure the sulfur-sulfur interatomic distance, and assignment of the spectra and calculation of the peptide's conformations is dependent on the presence of bonded CysH $\beta^{1,2}$ –CysH $\beta^{1,2}$  nuclear Overhauser effect spectral peaks; therefore, interactions between bounded Cys, Abu-, or Ser-substituted residues were determined by measuring the C $\beta$ H $\beta^{1,2}$ –C $\beta$ H $\beta^{1,2}$  center-of-mass distances and calculating the probability of contacts within the one-sided 95% confidence interval (CI<sub>95</sub>) [30,54]. The SASAs of residues 2, 5, 16, 19, 20, 28, 33, and 35 of CTX and its Abu- and Ser-substituted analogs were calculated using the *g\_sasa* module of GROMACS using the atomic radii of Lee and Richards, a probe radius of 0.14 nm for water and 1000 points per sphere resolution [55]. The residue SASAs were normalized to values of their maximal solvent-accessible surface area as determined from MD simulations on Acetyl-Ala-Xaa-Ala-NH<sub>2</sub> (where Xaa is either Ser or Abu) and (Acetyl-Ala-Cys-Ala-NH<sub>2</sub>)<sub>2</sub> (dimer with Cys–Cys disulfide bond) as outlined in Appendix A to yield relative SASA (rSASA) [56,57]. Normalization of SASA to the rSASA has been shown to correlate to models of protein folding, stability, and structural determination [57]. 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 [58,59].

#### 4.3.3. Conformational Analysis

The sampled conformations of the peptides were analyzed using the dihedral principal component analysis method [60,61]. The time-dependent  $\varphi/\psi$  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 two-dimensional free energy ( $\Delta G$ ) landscape:

$$\Delta G = -R \cdot T \cdot \ln \frac{\rho_{x,y}}{\rho_{max}} \quad (1)$$

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  $\Delta G$  landscape was calculated by dividing the principal component 1 – principal component 2 subspace into grids to create a two-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 [47,62–65]. 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 [65,66]. The optimal number of clusters was determined by visual inspection, sum of squared error, average silhouette width, silhouette coefficient, and distribution plots [67–70].

#### 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<sub>AB</sub>):

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

comparing MD trajectories of peptides A and B, where  $\eta_i^A$  and  $\eta_j^B$  are the respective eigenvectors of the sampled essential subspace, and N is the total number of eigenvectors to be considered [71–74]. To correct for sampling errors and autocorrelation, the RMSIP can be normalized (nRMSIP):

$$nRMSIP = \frac{RMSIP_{A \bullet B}}{\sqrt{RMSIP_{A1 \bullet A2} \cdot RMSIP_{B1 \bullet B2}}} \quad (3)$$

RMSIP<sub>A1•A2</sub> and RMSIP<sub>B1•B2</sub> 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 [74].

## 5. Conclusions

In CTX, the disulfide bonds between Cys<sup>16</sup>–Cys<sup>33</sup> and Cys<sup>20</sup>–Cys<sup>35</sup> residues are required to maintain the association between the  $\alpha$ -helical region from residues 13 through 20 and the two antiparallel  $\beta$ -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 three components of the  $\alpha\beta$  motif, either from hydrophobic, aromatic, ionic interactions, or H-bond formation. Of these two disulfide bonds, the bond between the Cys<sup>16</sup> and Cys<sup>33</sup> residues appears to be more critical, allowing for wider separation of the  $\alpha$ -helix and  $\beta$ -sheet, whereas replacement of the Cys<sup>20</sup> and Cys<sup>35</sup> residues causes less conformational disruption. Selective use of Ser-substitution of residues 16 and 33 may mitigate this destabilizing effect, since an interaction may occur between the two residues and an associated water molecule. Although step-wise removal of individual 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 four 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.

**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:** 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- $\alpha$ -aminobutyric acid
AVG	average
BB–BB	Backbone–backbone
CTX	chlorotoxin
CD	circular dichroism
CI <sub>95</sub>	One-sided 95% confidence interval
HPLC	high-performance liquid chromatography
MD	molecular dynamics
nRMSIP	normalized root-mean-square inner product
NMR	nuclear magnetic resonance spectroscopy
NPT	constant number, pressure, and temperature
R <sub>g</sub>	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

### Appendix A.1. Assignments of Force Field Parameters for L- $\alpha$ -Aminobutyric Acid (Abu)

The Lennard-Jones, electrostatic potentials, bond, bond angle, and torsional parameters for the Abu residue were assigned based on the similarity and transferability of force field parameters for similar noncyclic aliphatic residues (Ala, Leu, Ile, and Val) within the CHARMM36m force field using the same methodology as the CGenFF program while taking care to use the appropriate backbone dihedral and CMAP potentials [33–38]. This method is consistent with that used by other authors for parameterization of other non-native aliphatic amino acid residues [75]. 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  $\phi, \psi$  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
      CG  HG1
      CG  HG2
      CG  HG3
      O   C

[ impropers ]
      N   -C   CA   HN
      C   CA   +N   O

[ cmap ]
      -C   N   CA   C   +N

```

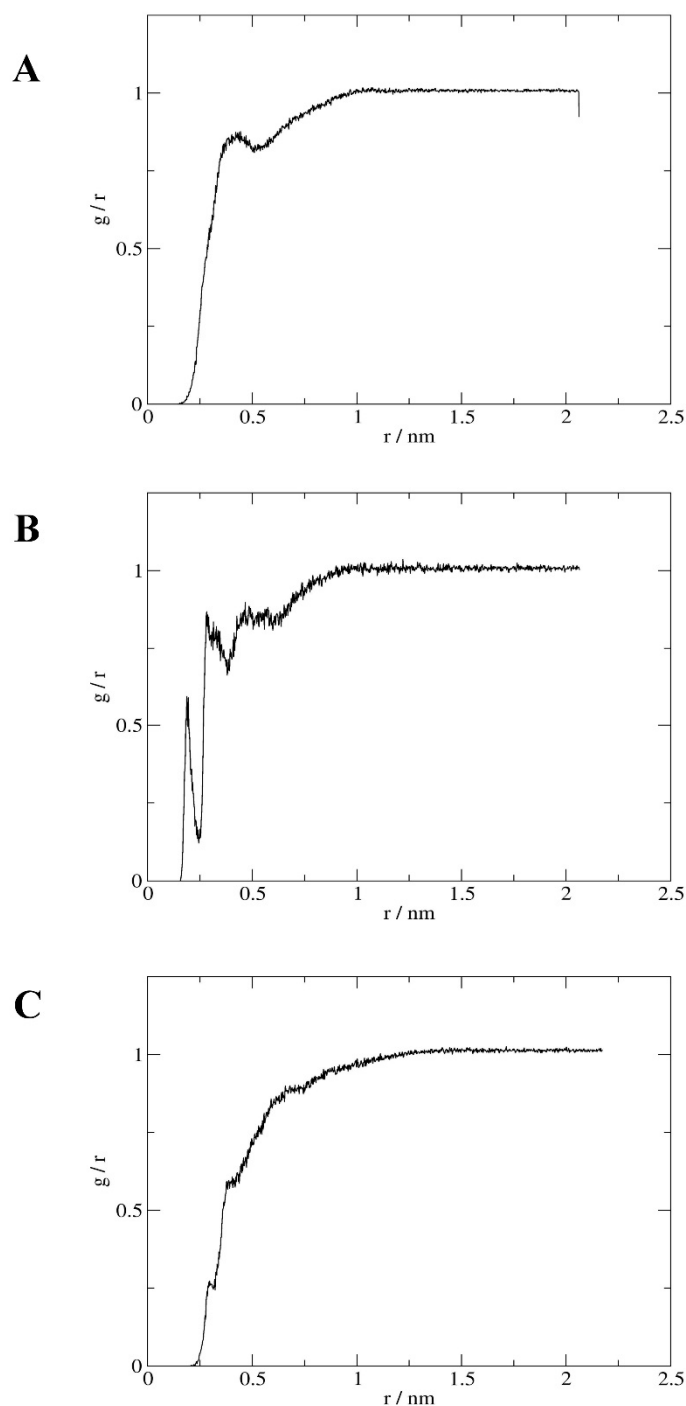
### Appendix A.2. MD Simulations of Model Peptides

The Acetyl-Ala-Xaa-Ala-NH<sub>2</sub> (where Xaa is either Ser or Abu) and (Acetyl-Ala-Cys-Ala-NH<sub>2</sub>)<sub>2</sub> (bridged with a Cys–Cys disulfide bond) peptides were simulated using 0.5  $\mu$ s 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<sup>-1</sup>) 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  $4.5 \times 10^{-5}$  bar<sup>-1</sup> 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].

The first 100 ns of each peptide simulation was considered system equilibration and the subsequent 400 ns used for analysis with a sampling frequency of 0.1 ns. The center-of-mass distance (mean  $\pm$  SD) of

the CysH $\beta^{1,2}$ –CysH $\beta^{1,2}$  groups and their probability of contact representing the one-sided 95% confidence interval (CI<sub>95</sub>) were calculated and are given in the footnotes to Table 3. The SASA of the residues were calculated using the *g\_sasa* module of GROMACS using the atomic radii of Lee and Richards, a probe radius of 0.14 nm for water and 1000 points per sphere resolution [54].

The radial distribution plots representing the probability of finding water molecules within 0.5 nm of the terminal moieties of Cys, Ser, and Abu are displayed graphically in Figure A1 and their numeric values given in the footnotes to Table 4 [57,58].



**Figure A1.** Radial distribution function of water from the terminal methyl, hydroxyl and disulfide moieties of Acetyl-Ala-Abu-Ala-NH<sub>2</sub> (A), Acetyl-Ala-Ser-Ala-NH<sub>2</sub> (B), and (Acetyl-Ala-Cys-Ala-NH<sub>2</sub>)<sub>2</sub> (C) respectively. The probability of finding water molecules surrounding the respective moieties was determined by integrating the radial distribution function within 0.5 nm distance.

## References

1. DeBin, J.A.; Strichartz, G.R. Chloride Channel Inhibition by the venom of the scorpion *Leiurus Quinquestriatus*. *Toxicon* **1991**, *29*, 1403–1408. [[CrossRef](#)]
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 Pt 1*, C361–C369. [[CrossRef](#)]
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*, 5572–5582. [[CrossRef](#)] [[PubMed](#)]
4. Soroceanu, L.; Gillespie, Y.; Khazaeli, M.B.; Sontheimer, H. Use of chlorotoxin for targeting of primary brain tumors. *Cancer Res.* **1998**, *58*, 4871–4879. [[PubMed](#)]
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*, 162–173. [[CrossRef](#)] [[PubMed](#)]
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*, 175–186. [[CrossRef](#)]
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*, 256–264. [[CrossRef](#)]
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.; et al. Tumor paint: A chlorotoxin: Cy5.5 bioconjugate for intraoperative visualization of cancer foci. *Cancer Res.* **2007**, *67*, 6882–6888. [[CrossRef](#)]
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*, 782–787. [[CrossRef](#)]
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*, 9807. [[CrossRef](#)]
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*, 113–125. [[CrossRef](#)]
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. *Comput. Biochem. Physiol. A Mol. Integr. Physiol.* **1998**, *121*, 323–332. [[CrossRef](#)]
13. Possani, L.D.; Becerril, B.; Delepierre, M.; Tytgat, J. Scorpion toxins specific for Na<sup>+</sup>-channels. *Eur. J. Biochem.* **1999**, *264*, 287–300. [[CrossRef](#)] [[PubMed](#)]
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*, 861–868. [[CrossRef](#)]
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.; et al. The toxicogenomic multiverse: Convergent recruitment of proteins into animal venoms. *Annu. Rev. Genomics Hum. Genet.* **2009**, *10*, 483–511. [[CrossRef](#)] [[PubMed](#)]
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*, 1521–1523. [[CrossRef](#)] [[PubMed](#)]
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 <sup>1</sup>H NMR study of recombinant insect defensin A in water: Resonance assignments, secondary structure and global folding. *J. Biomol. NMR* **1992**, *2*, 235–256. [[CrossRef](#)]
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*, 435–448. [[CrossRef](#)]
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 <sup>1</sup>H-NMR: A structural motif common to toxic arthropod proteins. *Biochemistry* **1993**, *32*, 715–724. [[CrossRef](#)] [[PubMed](#)]
20. Craik, D.J.; Daly, N.L.; Waine, C. The cystine knot motif in toxins and implications for drug design. *Toxicon* **2001**, *39*, 43–60. [[CrossRef](#)]
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* **2016**, *8*, 36. [[CrossRef](#)] [[PubMed](#)]

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. [[CrossRef](#)]
23. Bulaj, G. Formation of disulfide bonds in proteins and peptides. *Biotechnol. Adv.* **2005**, *23*, 87–92. [[CrossRef](#)] [[PubMed](#)]
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*, 10641–10647. [[CrossRef](#)] [[PubMed](#)]
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*, 1292–1301. [[CrossRef](#)] [[PubMed](#)]
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*, 11488–11494. [[CrossRef](#)] [[PubMed](#)]
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*, 3760–3766. [[CrossRef](#)] [[PubMed](#)]
28. Ojeda, P.G.; Wang, C.K.; Craik, D.J. Chlorotoxin: Structure, activity, and potential uses in cancer therapy. *Biopolymers* **2016**, *106*, 25–36. [[CrossRef](#)] [[PubMed](#)]
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*, 1617–1628. [[CrossRef](#)] [[PubMed](#)]
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*, 13–21. [[CrossRef](#)] [[PubMed](#)]
31. Kabsch, W.; Sander, C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983**, *22*, 2577–2637. [[CrossRef](#)] [[PubMed](#)]
32. Frishman, D.; Argos, P. Knowledge-based protein secondary structure assignment. *Proteins* **1995**, *23*, 566–579. [[CrossRef](#)]
33. 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, Oxford, England, UK, 31 August–4 September 1993*; Epton, R., Ed.; Mayflower Worldwide: Birmingham, UK, 1994; pp. 201–206, ISBN 0951573519-978-0951573519.
34. 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*, 71–73. [[CrossRef](#)] [[PubMed](#)]
35. 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*, 3257–3273. [[CrossRef](#)] [[PubMed](#)]
36. 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*, 1400–1415. [[CrossRef](#)] [[PubMed](#)]
37. 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.; et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616. [[CrossRef](#)]
38. Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; et al. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **2009**, *31*, 671–690. [[CrossRef](#)] [[PubMed](#)]
39. 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. [[CrossRef](#)]

40. 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. [CrossRef]
41. 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*, 435–447. [CrossRef]
42. 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*, 1701–1718. [CrossRef]
43. Lindahl, E.; Hess, B.; van der Spoel, D. GROMACS 3.0: A package for molecular simulation and trajectory analysis. *J. Mol. Mod.* **2001**, *7*, 306–317. [CrossRef]
44. Berendsen, H.J.C.; van der Spoel, D.; van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* **1995**, *91*, 43–56. [CrossRef]
45. Abraham, M.J.; van der Spoel, D.; Lindahl, E.; Hess, B.; the GROMACS Development Team. GROMACS User Manual Version 5.1.2. Available online: <http://www.gromacs.org> (accessed on 1 June 2017).
46. 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*, 459–466. [CrossRef] [PubMed]
47. Krieger, E.; Vriend, G. YASARA View-molecular graphics for all devices—from smartphones to workstations. *Bioinformatics* **2014**, *30*, 2981–2982. [CrossRef]
48. Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, 014101. [CrossRef]
49. 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*, 3684–3690. [CrossRef]
50. 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*, 1463–1472. [CrossRef]
51. Hess, B. P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 116–122. [CrossRef]
52. Essmann, U.; Perera, L.; Berkowitz, M.L. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577–8593. [CrossRef]
53. Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182–7190. [CrossRef]
54. 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*, 897–906. [CrossRef] [PubMed]
55. Lee, B.; Richards, F.M. The interpretation of protein structures: Estimation of static accessibility. *J. Mol. Biol.* **1971**, *55*, 379–400. [CrossRef]
56. Zielenkiewicz, P.; Saenger, W. Residue solvent accessibilities in the unfolded polypeptide chain. *Biophys. J.* **1992**, *63*, 1483–1486. [CrossRef]
57. 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*, e80635. [CrossRef] [PubMed]
58. 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*, 7254–7259. [CrossRef] [PubMed]
59. 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*, 1399–1409. [CrossRef] [PubMed]
60. Mu, Y.; Nguyen, P.H.; Stock, G. Energy landscape of a small peptide revealed by dihedral angle principal component analysis. *Proteins* **2005**, *58*, 45–52. [CrossRef] [PubMed]
61. Altis, A.; Nguyen, P.H.; Hegger, R.; Stock, G. Dihedral angle principal component analysis of molecular dynamics simulations. *J. Chem. Phys.* **2007**, *126*, 244111. [CrossRef] [PubMed]
62. Ligges, U.; Maechler, M. scatterplot3d—An R Package for Visualizing Multivariate Data. *J. Stat. Softw.* **2003**, *8*, 1–20. [CrossRef]
63. 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 August 2018).



64. Sarkar, D.; Andrews, F. latticeExtra: Extra Graphical Utilities Based on Lattice. Available online: <https://cran.r-project.org/web/packages/latticeExtra/> (accessed on 6 August 2018).
65. R Core Team. R: A language and environment for statistical computing. Available online: <http://www.R-project.org/> (accessed on 6 August 2018).
66. Maechler, M.; Rousseeuw, P.; Struyf, A.; Hubert, M.; Hornik, K.; Studer, M.; Roudier, P.; Gonzalez, J.; Kozłowski, K. Cluster: Methods for Cluster Analysis. Available online: <https://cran.r-project.org/web/packages/cluster/> (accessed on 6 August 2018).
67. 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.
68. 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.
69. Rousseeuw, P.J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **1987**, *20*, 53–65. [[CrossRef](#)]
70. Thinsungnoen, T.; Kaoungku, N.; Durongdumronchai, P.; Kerdprasop, K.; Kerdprasop, N. The Clustering Validity with Silhouette and Sum of Squared Errors. In *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; Volume 3, pp. 44–51. [[CrossRef](#)]
71. Andricioaei, I.; Karplus, M. On the calculation of entropy from covariance matrices of the atomic fluctuations. *J. Chem. Phys.* **2001**, *115*, 6289–6292. [[CrossRef](#)]
72. Hess, B. Convergence of sampling in protein simulations. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **2002**, *65 Pt 1*, 031910. [[CrossRef](#)]
73. D'Alessandro, M.; Paci, M.; Amadei, A. Characterization of the conformational behavior of peptide Contryphan Vn: A theoretical study. *Biopolymers* **2004**, *74*, 448–456. [[CrossRef](#)] [[PubMed](#)]
74. 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*, e0121114. [[CrossRef](#)] [[PubMed](#)]
75. Turpin, E.R.; Mulholland, S.; Teale, A.M.; Boyan, B.; Boneva, B.B.; Hirst, J.D. New CHARMM force field parameters for dehydrated amino acid residues, the key to lantibiotic molecular dynamics simulations. *RSC Adv.* **2014**, *4*, 48621–48631. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).