

Molecular dynamics simulations of the conformational states of a D-tryptophan-containing *Conus* venom peptide and its all-L-amino acid analog

Neil Andrew D. Bascos^{1,*} and Elsie C. Jimenez²

¹National Institute of Molecular Biology and Biotechnology, University of the Philippines Diliman, Quezon City 1101; ²Department of Physical Sciences, College of Science, University of the Philippines Baguio, Baguio City 2600, Philippines.

ABSTRACT

Molecular dynamics (MD) simulation is a powerful method for examining the conformational states of biomolecular systems. In the present work, MD simulations were employed to probe into the dynamic modes and conformational states of contryphan-Sm (a *Conus* venom peptide with D-Trp4) and its analog, [L-Trp4]contryphan-Sm, specifically focusing on the investigation of their structural differences. Molecular modeling showed that the basic cyclic structures of contryphan-Sm and [L-Trp4]contryphan-Sm were similar, with no steric clashes occurring among the amino acid residues. The MD simulations showed that contryphan-Sm assumed a more compact conformation compared to [L-Trp4]contryphan-Sm based on their maximum peptide dimensions and radii of gyration. After ~ 20 ns of MD simulations, the root-mean-square deviation (RMSD) values were lower for almost all of the amino acid residues in contryphan-Sm, with its D-Trp4 showing the highest difference in RMSD from L-Trp4 in [L-Trp4]contryphan-Sm, suggesting that contryphan-Sm had less structural variability. Energy measurements supported this finding, with contryphan-Sm consistently exhibiting lower kinetic energy values compared to [L-Trp4]contryphan-Sm throughout the MD simulations. The Ramachandran plots showed greater variations in phi or psi angles

in L-Trp4, Gln5, Pro6 and Trp7 in [L-Trp4]contryphan-Sm than the corresponding residues in contryphan-Sm at the start and end of the MD simulations. Contryphan-Sm showed less solvent accessibility than [L-Trp4]contryphan-Sm as shown by the measurements of their solvent-activated surface areas. Decreased solvent accessibility may be linked to the stacked conformation adopted by contryphan-Sm, aligning D-Trp4, Pro6 and Trp7. Despite the observed motions of the Trp side chains, both contryphan-Sm and [L-Trp4]contryphan-Sm structures do not support the occurrence of intramolecular covalent crosslinking between D/L-Trp4 and Trp7. The observed differences in dynamic modes and conformational states of contryphan-Sm and [L-Trp4]contryphan-Sm are correlated with the greater structural stability of the D-Trp-containing contryphan.

KEYWORDS: *Conus* venom peptide, contryphan, conotoxin, conopeptide, D-tryptophan, molecular modeling, molecular dynamics simulation.

INTRODUCTION

Studies of the conformational dynamics of biomolecules have steadily gained interest in the recent years. This work documents the use of molecular dynamics (MD) simulations to examine and compare the conformational states of D-Trp- and L-Trp-containing contryphan analogs that influence their structural stability.

*Corresponding author
neilandrew.bascos@mbb.upd.edu.ph

The contryphans [1, 2] constitute a distinct family of D-amino acid-containing *Conus* peptide toxins referred to as conotoxins or conopeptides [3, 4]. Contryphan-Sm was purified from venom of the piscivorous cone snail species, *Conus stercusmuscarum*. It has the amino acid sequence, GCOWQPWC*, where O is 4-hydroxyproline, W is D-Trp, and the asterisk indicates C-terminal amidation [5]. The presence of D-Trp4 in contryphan was initially discovered in venom of the piscivorous *Conus radiatus* [6]; this contryphan was later named contryphan-R [5]. The primary sequence of contryphan-Sm is highly identical to that of contryphan-R (GCOWEPWC*), with a single amino acid substitution in which Gln5 in contryphan-Sm replaces Glu5 in contryphan-R. The mRNA sequence of contryphan-R predicts a prepropeptide that has the mature contryphan at the C-terminus, with the Trp4 residue encoded by UGG [7].

In contryphans, the isomerization of a single L- to D-amino acid can cause a significant difference in chromatographic features, due to diverse conformational dynamics. Contryphan-R and contryphan-Sm showed two peaks under reverse-phase high-performance liquid chromatography (HPLC) conditions, indicating interconversion between two conformational states. On the other hand, [L-Trp4]contryphan-R (GCOWEPWC*) and [L-Trp4]contryphan-Sm (GCOWQPWC*) revealed single and unusually broad HPLC peaks that eluted later than contryphan-R and contryphan-Sm, respectively [5, 6, 8].

The 3-dimensional structures of contryphan-R and contryphan-Sm have been determined by 2D-NMR. These revealed the existence of disulfide-bonded octapeptides in two conformational states in solution. These states occur due to *cis-trans* isomerism about the Cys2-Hyp3 peptide bond. The solution structures of the major forms (with *cis* Cys2-Hyp3 peptide bonds) showed a definite fold with non-hydrogen-bonded N-terminal chain reversal including *cis*-Hyp3 and D-Trp4, and a C-terminal type I β -turn [9, 10].

The [L-Trp4]contryphan-R appeared to be less stable than contryphan-R. The L-Trp-containing analog was observed to readily form a purple to blue-violet derivative. This was prevented by treating the peptide solution with nitrogen, thereby inhibiting oxidation by eliminating oxygen in the system. Contryphan-R and the nitrogen-treated [L-Trp4]contryphan-R

showed comparable biological activities based on mouse bioassays [6]. The ease of degradation of [L-Trp4]contryphan-R compared to contryphan-R, in the absence of nitrogen treatment during chemical synthesis, indicated that the D-Trp4 in contryphan-R could contribute to the improved structural stability of the peptide.

The unusual structures of contryphans, presumably due to their high content of post-translationally modified amino acids, have made them the subject of interest in studies using MD simulations and NMR spectroscopy. The former provides predictions on possible conformational states, and the latter allows validation of these predictions *in vitro*. NMR spectroscopy and MD simulations showed the main conformational features of contryphan-Vn (GDCPWKPWC*) from *Conus ventricosus*, including a salt bridge between Asp2 and Lys6, and mainly *cis* and *trans* conformations of Pro4 and Pro7, respectively [11]. With MD simulation, a statistical mechanical description of contryphan-Vn has been reported, including the major structural fluctuations due to the concerted motions of D-Trp5 and Trp8 and its reduced flexibility due to the disulfide bridge between Cys3 and Cys9 [12]. More recently, MD simulations were done to characterize the *cis-trans* isomerization about the Xxx-Pro bond in contryphan In936 (GCVLYPWC*, where L is D-Leu) from *Conus inscriptus*, its analog [V3P]In936 (GCPLYPWC*), and contryphan Lo959 (GCPWDPWC*) from *Conus lorioisii* [13]. NMR spectroscopy and MD simulations have been carried out to investigate the *cis-trans* isomerization of Pro4 in [W8S]contryphan-Vn (GDCPWKPSC*) [14].

Several D-Trp-containing contryphans continue to be discovered recently [15-17]. The D-Trp3 in contryphan-Vc2 (CRWTPVC*) from *Conus victoriae* was shown to confer stability to proteolysis by trypsin while [L-Trp3]contryphan-Vc2 (CRWTPVC*) was susceptible to this enzyme [16].

To date, however, no MD simulations have been done to compare the conformational states of the D-Trp- and L-Trp-containing contryphan analogs. To elucidate further the conformational dynamics in contryphans, MD simulations were conducted to assess the conformational states and possible intramolecular interactions of contryphan-Sm in comparison with [L-Trp4]contryphan-Sm. Analyses of the conformational dynamics of these contryphans,

including a comparison of the biophysical features of the D/L-Trp4 residue in relation to the other amino acid residues, are reported. The MD simulations are particularly useful for investigating the full range of side-chain and loop motions in these small disulfide-bonded peptides. This study could help elucidate further the importance of D-Trp in contryphan-Sm and its effect on the structural stability of contryphan-Sm and other D-Trp-containing contryphans.

MATERIALS AND METHODS

Molecular modeling

The initial coordinates of the major form of contryphan-Sm were obtained from the Protein Data Bank (PDB) with accession code PDBID:1DFY [10]. The 20 models that represent the range of conformations for the NMR structure of 1DFY showed minimal variations when superimposed. Therefore, one structure (i.e., model 1) was deemed sufficient to represent the average structure and to serve as basis for the starting models for MD simulations.

The crystal structure of contryphan-Sm has non-standard amino acid residues (Hyp3, D-Trp4, amidated Cys8 and disulfide-bonded Cys2-Cys8) that were built using the *xleap* and *antechamber* programs of the AMBER molecular simulation package [18-20]. The AMBER ff14SB force field was loaded to define the parameter of peptide motion. The molecular file for contryphan-Sm was then loaded onto *xleap* to generate the initial structure. Designations for the non-standard amino acids were encoded into the amino acid sequence either by using the library of structure names in *xleap*, or building the modified structures using the *antechamber* program. Modifications for certain structures, for e.g., the amidation of Cys8, were added using *antechamber*. Built structures were minimized prior to MD simulations. Minimization was done by selecting the built structure and using the *relax selection* function of *antechamber*. This process adjusted the bond and dihedral angles of the structure to attain lower energy states. Generation of [L-Trp4]contryphan-Sm structure was done through the same process. Models were rendered using either the RasTop molecular graphics visualization tool (<http://www.openrasmol.org>) [21] or the visual molecular dynamics (VMD) molecular viewer (<http://www.ks.uiuc.edu/Research/vmd>) [22].

Water box preparation for the simulation of an aqueous environment

The TIP3P intermolecular potential function [23] was employed to define the parameters of the simulated water movement. The MD simulations were done for either a single or multi-molecule system. The box size was designed to accommodate either a single molecule or 27 molecules with at least 0.5 nm of water on each side of the molecule based on standard protocols for NAMD (<http://www.ks.uiuc.edu/Training/Tutorials/namd-index.html>) [24]. The systems included peptides immersed in appropriately measured boxes with water that contained Cl⁻. For the multi-molecule system, 27 peptide molecules were equidistantly arranged in a cube, with 2.0 nm separation along the x, y, and z axes.

Molecular dynamics simulations

For each system, minimization, heating, and equilibration for 2 ns were carried out to prepare the files. For each peptide, 20 ns MD simulation run was performed using the NAMD 2.10 package [24]. Ten 2 ns production runs were done at 300 K and 101325 Pa (1 atm). The Langevin thermostat [25] was employed to regulate the temperature. To constrain all bonds involving hydrogen atoms the SHAKE algorithm [26] was used. The Particle mesh Ewald method [27] was applied to account for long-range interactions. To integrate the equations of motion a 2 fs time step was employed. Frames for data analysis were collected at 1 ps time intervals.

Extraction of molecular coordinates to analyze MD simulation data

The MD simulation was carried out in 10 successive rounds of 2 ns simulations for a total of 20 ns of molecular movement. The positions of each molecule in the system were recorded for every picosecond of the simulation. These represented 2000 data timepoints for every 2 ns of simulation. Each timepoint represented a 1ps timeframe within the 2 ns simulation. Data were stored and accessed as *.dcd files using the VMD viewer [21]. Structure files were generated from these for different timepoints in the simulation. By default, the VMD viewer labels the 2000 (1 ps) timeframes for each 2 ns simulation as frame 0 to frame 1999. For instance, to analyze the peptide structure at the beginning of the simulation, the peptide structure

in the first data frame of the first 2 ns of simulation run (run 1: frame 0) may be selected and saved as a molecular structure file (e.g., contryphan-Sm_0ns.pdb). The molecular structure of the same peptide at the end of the 20 ns of simulation may be observed by selecting the peptide in the last data frame of the 10th 2 ns of simulation run; total simulation time of 20 ns (run 10: frame 1999) and saving it as another molecular structure file (e.g., contryphan-Sm_20ns.pdb).

Peptide molecule dimension and radius of gyration

Data from the MD simulations were analyzed by tracking the features of the molecules throughout the simulation time (i.e., average maximum peptide dimension from 0 to 20 ns). The dimensions of the contryphan molecules were computed by measuring the maximum distance between atoms within the molecules (D_{\max}) and their radii of gyration (R_{gyr}). Data on the dimensions of the molecules provide information on the dynamics of the molecule conformations. The D_{\max} and R_{gyr} were used to approximate the size or conformation of the molecules.

To determine D_{\max} , the minimum and maximum coordinate values for the molecular structures were extracted using the *minmax* function of the VMD viewer [22]. The distance between the points defined by these maximum (x_{\max} , y_{\max} , z_{\max}) and minimum coordinates (x_{\min} , y_{\min} , z_{\min}) was designated as D_{\max} for the molecule. The distance between these points was computed based on the distance formula given in Equation 1.

The R_{gyr} for selected molecules were extracted using the *radgyr* function of the VMD viewer [22]. In VMD, the R_{gyr} of a selected molecule was computed as the root-mean-square distance of each atom of the molecule from its center of mass (r_{com}), as shown in Equation 2, where r_i is the position of the atom i and w_i is its weighted mass [28]. The coordinates for the weighted r_{com} were determined through the *measure center* command in the TK console of VMD. The coordinates of the r_{com} for each axis (i.e., x_{com} , y_{com} , z_{com}) are defined by Equations 3-5, where m_i is the atomic mass and x_i , y_i and z_i are the Cartesian coordinates of the atom i [28].

$$D_{\max} = \left[(x_{\max} - x_{\min})^2 + (y_{\max} - y_{\min})^2 + (z_{\max} - z_{\min})^2 \right]^{1/2} \quad (\text{Equation 1})$$

$$R_{\text{gyr}}^2 = \sum_{i=1}^n (w_i)(r_i - r_{\text{com}})^2 / \sum_{i=1}^n (w_i) \quad (\text{Equation 2})$$

$$x_{\text{com}} = \sum_{i=1}^n (m_i)(x_i) / \sum_{i=1}^n (m_i) \quad (\text{Equation 3})$$

$$y_{\text{com}} = \sum_{i=1}^n (m_i)(y_i) / \sum_{i=1}^n (m_i) \quad (\text{Equation 4})$$

$$z_{\text{com}} = \sum_{i=1}^n (m_i)(z_i) / \sum_{i=1}^n (m_i) \quad (\text{Equation 5})$$

Determination of Trp4-Trp7 separation

The MD simulation data were also analyzed by taking the distance between the C2's of the indole rings of Trp4 and Trp7 at 0 ns (e.g., contryphan-Sm_0ns.pdb) and 20 ns (e.g., contryphan-Sm_20ns.pdb) of a simulation. Distances between the reference C δ (C2) atoms of the indole rings of D/L-Trp4 and Trp7 in contryphan-Sm and [L-Trp4]contryphan-Sm were measured to compare their separation. The distance between two specific atoms within selected molecules was determined by extracting the coordinates of these atoms and applying the distance formula. A PERL script was used to extract the coordinates (x, y, z) assigned to reference atoms (e.g., atom 1 = C2 of Trp4; atom 2 = C2 of Trp7). Distances ($D_{1 \rightarrow 2}$) between the reference atoms 1 and 2 were computed based on the distance formula given in Equation 6.

$$D_{1 \rightarrow 2} = [(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2]^{1/2} \quad (\text{Equation 6})$$

Analyses of root-mean-square deviation value

To compare the relative stability of contryphan-Sm and [L-Trp4]contryphan-Sm, analysis was done using the root-mean-square deviation (RMSD) values for the amino acid residues. The RMSD value was used to approximate stability in the peptides because it monitors the observed motions and variability in position of the individual amino acid residues.

Determination of kinetic energy values

The data for the kinetic energy values were extracted from the log files of each of the MD simulation runs using the NAMD Plot extension of the VMD molecular viewer. These log files contained the coordinate and energy values for each timestep within the MD simulation. The extracted data were further processed using Microsoft Excel (2016) to determine their variability throughout each 2 ns simulation run. The average values and the standard deviations for the 2000 timesteps in each run were determined and compared.

Preparation of Ramachandran plots

Ramachandran plots present the values for the phi and psi angles between amino acids within a polypeptide sequence [29]. These allow monitoring of the existence of conformations that fall within standard types of secondary structures (beta-sheet and helical structures). Ramachandran plots were generated for each 2 ns timestep within the 20 ns of MD simulation. The plots were generated using the *Ramachandran Plot* extension of the VMD molecular viewer. The phi and psi angles for each residue were monitored in each timestep to assess their involvement in conformation variability.

Solvent accessibility and solvent-activated surface area

The solvent accessibility of D/L-Trp4 was examined throughout the course of the MD simulations for contryphan-Sm and [L-Trp4]contryphan-Sm. The number of water molecules within 0.5 nm of D/L-Trp4 was determined for each contryphan isomer. Furthermore, determination of water molecules within 0.5 nm of Pro6 was done to validate a conformational feature observed with contryphan-Sm but not with [L-Trp4]contryphan-Sm. In addition, the solvent-activated surface area (SASA) was analyzed for both contryphan analogs and their corresponding amino acid residues at 0 and 20 ns. Total SASA for the whole peptide was computed using the *vmd.sasa.tcl* script (https://codedocs.xyz/mlund/faunus/vmd-sasa_8tcl_source.html) [30]. The SASA value for each amino acid was computed using the *sasa.tcl* script (http://www.ks.uiuc.edu/Research/vmd/mailling_list/vmd-l/att-18670/sasa.tcl) [31]. Details of the usage of each script are provided within their comments section.

RESULTS

Molecular models for contryphan-Sm and [L-Trp4]contryphan-Sm were generated for the analyses. The molecular models revealed that the basic cyclic structures were similar for contryphan-Sm and [L-Trp4]contryphan-Sm, with no steric clashes present in the residues. Images of these models in stereo view are provided in Figure 1.

The MD simulations were performed to probe the conformational states accessible in contryphan-Sm compared to [L-Trp4]contryphan-Sm. They

were conducted on models that represented D/L-Trp4-containing contryphan structures. The simulations did not document an interconversion between the L- and D-Trp forms, but showed their effects on the motions of their neighboring amino acid residues. The molecular motions (side chain and loop motions) were monitored throughout the MD simulations in the nanosecond timescale. The MD simulations covered these motions up to 20 ns, which is within the proposed optimal MD simulation time window [32]. The results provide further insight on the basis for the structural significance of D-Trp in contryphan-Sm.

Despite the retention of the cyclic structures for both contryphan analogs through the disulfide linkage, the range of motions observed in the MD simulations covered and surpassed the range observed in the NMR structure of contryphan-Sm (PDBID: 1DFY) [10]. Superimposition of the 0 and ~20 ns structures for contryphan-Sm using Gly1, Cys2 and Hyp3 as references showed that the shifts were concentrated in the movement of Trp4, Pro6 and Trp7 into the stacked conformation after the MD simulations. The attainment of the stacked conformation for contryphan-Sm resulted in a more compact and less extended structure compared to [L-Trp4]contryphan-Sm, as will be explained below.

Computations for maximum distance (D_{\max}) values were done using the *minmax* function of the VMD molecular viewer. The D_{\max} values were computed to quantify the difference in molecular dimensions for the two contryphans. The D_{\max} provided the maximum separation between atoms in the peptide molecule, and represented the most extended dimension of the molecule. The D_{\max} of contryphan-Sm decreased from 2.23 nm to 2.12 nm while that of [L-Trp4]contryphan-Sm increased from 2.44 nm to 2.64 nm by the end of the 20 ns of MD simulations. From 8 ns to 20 ns, contryphan-Sm had consistently shown lower D_{\max} than [L-Trp4]contryphan-Sm (Figure 2A). The difference in molecular dimensions for the two contryphans was consistently and distinctly observable within the 20 ns of MD simulations.

The radius of gyration (R_{gyr}) for each structure was measured based on the *radgyr* function of the VMD molecular viewer. The initial structure of contryphan-Sm had R_{gyr} of 0.56 nm, while [L-Trp4]contryphan-Sm had R_{gyr} of 0.58 nm.

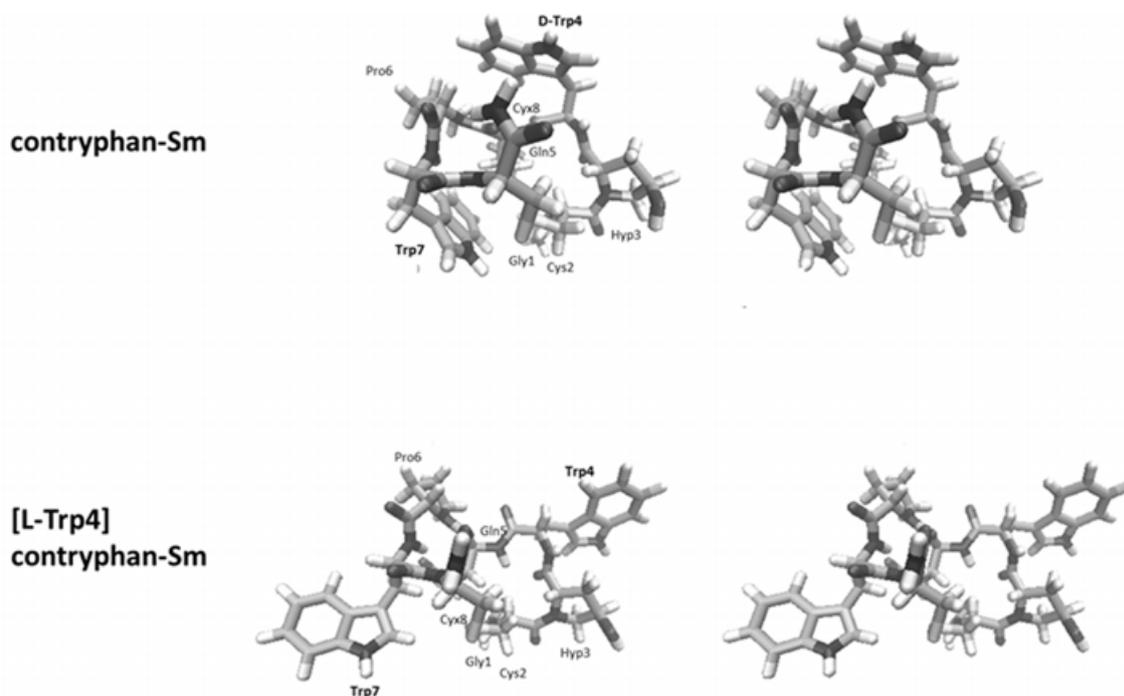


Figure 1. Stereo view of molecular models of contryphan-Sm and [L-Trp4]contryphan-Sm after 20 ns of MD simulations. The molecular model for contryphan-Sm was generated based on the structure elucidated by Pallaghy *et al.* (2000) and deposited in the Protein Data Bank, with PDBID:1DFY (www.pdb.org). The molecular model for [L-Trp4]contryphan-Sm with L-Trp4 instead of D-Trp4 was similarly generated.

From 6 ns until 20 ns, contryphan-Sm had persistently attained lower R_{gyr} values. By the end of the 20 ns MD simulations, contryphan-Sm assumed R_{gyr} of 0.53 nm compared to [L-Trp4]contryphan-Sm with R_{gyr} of 0.61 nm indicating a more compact structure for contryphan-Sm (Figure 2B). The Trp4, Pro6 and Trp7 residues adopted a stacked conformation in contryphan-Sm while [L-Trp4]contryphan-Sm remained extended at the end of the 20 ns of MD simulations.

Molecular models showing the progression of MD simulations at 0, 2 and 20 ns were prepared (Figure 3). The consistent and distinct conformations achieved by contryphan-Sm and [L-Trp4]contryphan-Sm allowed the characterization and comparison of their structures within the 20 ns of MD simulations.

Higher RMSD in a residue suggests greater variability in position, and correspondingly, lower conformational stability for the peptide region in which it is located. Moreover, higher RMSD is associated with less compact and more flexible peptide structure. RMSD was calculated for each

amino acid based on variance (at each timeframe) from the initial time step (frame 0) of each production set. The average RMSD values for the production set (e.g., prod 1: 0-2 ns or prod 10: 18-20 ns) were then determined for the individual amino acids in the peptides. The RMSD values of the individual amino acids showed different behaviors in the early stages (0-2 ns) compared to the late stages (18-20 ns) of MD simulations. The individual residue RMSD values were almost similar in contryphan-Sm and [L-Trp4]contryphan-Sm during the first 2 ns of simulations, except in Trp7 which showed the highest difference in RMSD values (Figure 4A). For all time points at 18-20 ns MD simulation, almost all of the residues of contryphan-Sm had lower RMSD values (Figure 4B). The RMSD values of contryphan-Sm were significantly different from those in [L-Trp4]contryphan-Sm based on a paired Student's T-test ($p = 0.03$). The residues that were more stabilized in contryphan-Sm compared to [L-Trp4]contryphan-Sm comprised the entire cyclic structure linked by Cys2 and Cys8. The greatest variation between the

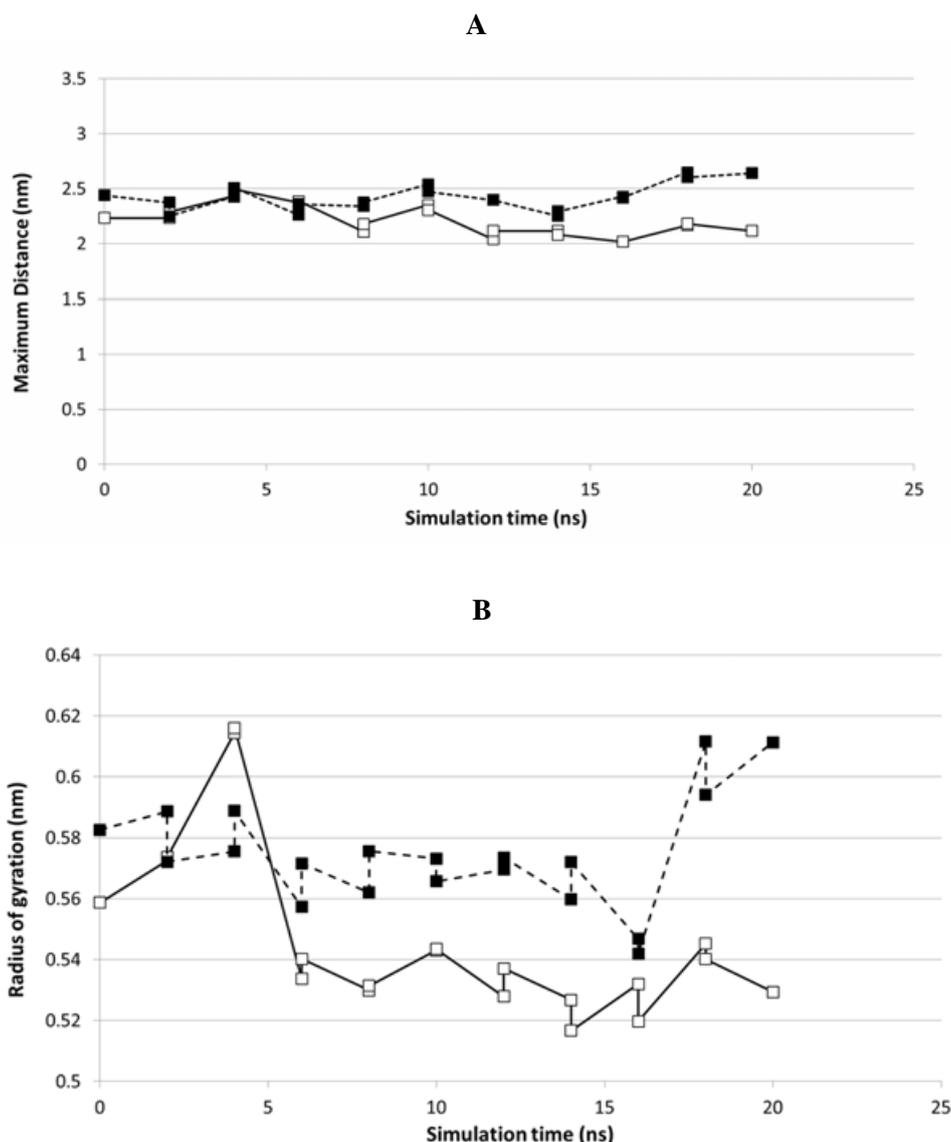


Figure 2. Contryphan dimensions tracked from 0 to 20 ns of molecular dynamics simulations. **(A)** Maximum distances (D_{\max}) were measured for contryphan-Sm and [L-Trp4]contryphan-Sm using the *minmax* function of visual molecular dynamics (VMD) viewer, with the minimum and maximum coordinates of peptide structures as references. **(B)** Radii of gyration (R_{gyr}) for the peptides were measured using the *radgyr* function of VMD viewer. -□- contryphan-Sm; -■- [L-Trp4]contryphan-Sm.

two populations occurred in D-Trp4 and L-Trp4, with RMSD difference of 2.721 Å. Thus, the greater stability observed in contryphan-Sm may be attributed mainly to D-Trp4. The lower average residue RMSD values observed for contryphan-Sm may be correlated with the behavior of the individual amino acid residues. The results suggest that contryphan-Sm attained a more stable structure than [L-Trp4]contryphan-Sm. It could be further

noted that for the 0-2 ns and 18-20 ns timepoints the RMSD values of Gly1 that is not part of the cyclic structure, did not vary between the two contryphan analogs.

The observed differences in RMSD values between contryphan-Sm and [L-Trp4]contryphan-Sm suggest greater flexibility in the latter structure. Increased flexibility allows for greater movement in the affected amino acid residues. Greater movement was indeed

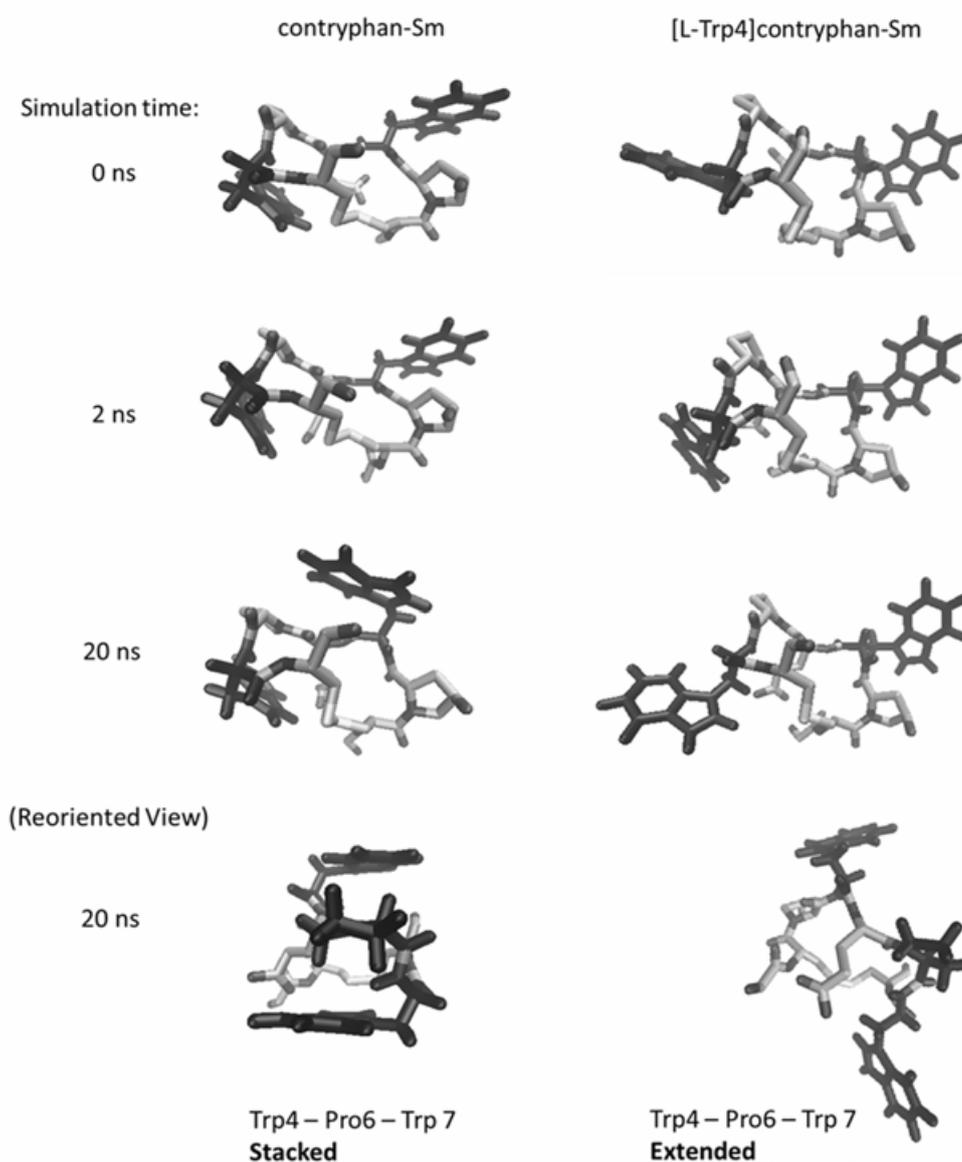


Figure 3. Molecular models showing the conformational states of contryphan-Sm and [L-Trp4]contryphan-Sm at 0, 2 and 20 ns of MD simulations. D/L-Trp4, Pro6 and Trp7 are blackened to highlight the stacking orientation in contryphan-Sm.

observed for residues in [L-Trp4]contryphan-Sm compared to contryphan-Sm as monitored by the kinetic energy (KE) measurements in the MD simulations. The average KE values recorded for [L-Trp4]contryphan-Sm were 5117 ± 55 kcal mole⁻¹ at 0-2 ns and 5117 ± 56 kcal mole⁻¹ at 18-20 ns of simulations. In contrast, for contryphan-Sm significantly lower average KE values of 4855 ± 55 kcal mole⁻¹ and 4849 ± 52 kcal mole⁻¹ were obtained at the 0-2 ns and 18-20 ns timeframes,

respectively. Statistical analysis using Student's t-test showed significant difference between the initial and final states for contryphan-Sm ($p = 0.0003$), but not for [L-Trp4]contryphan-Sm ($p = 0.9751$). The average KE values in contryphan-Sm and [L-Trp4]contryphan-Sm for both 0-2 ns and 18-20 ns timeframes were significantly different based on Student's t-test ($p = 0$). These findings further provide support for greater structural stability in the presence of D-Trp4.

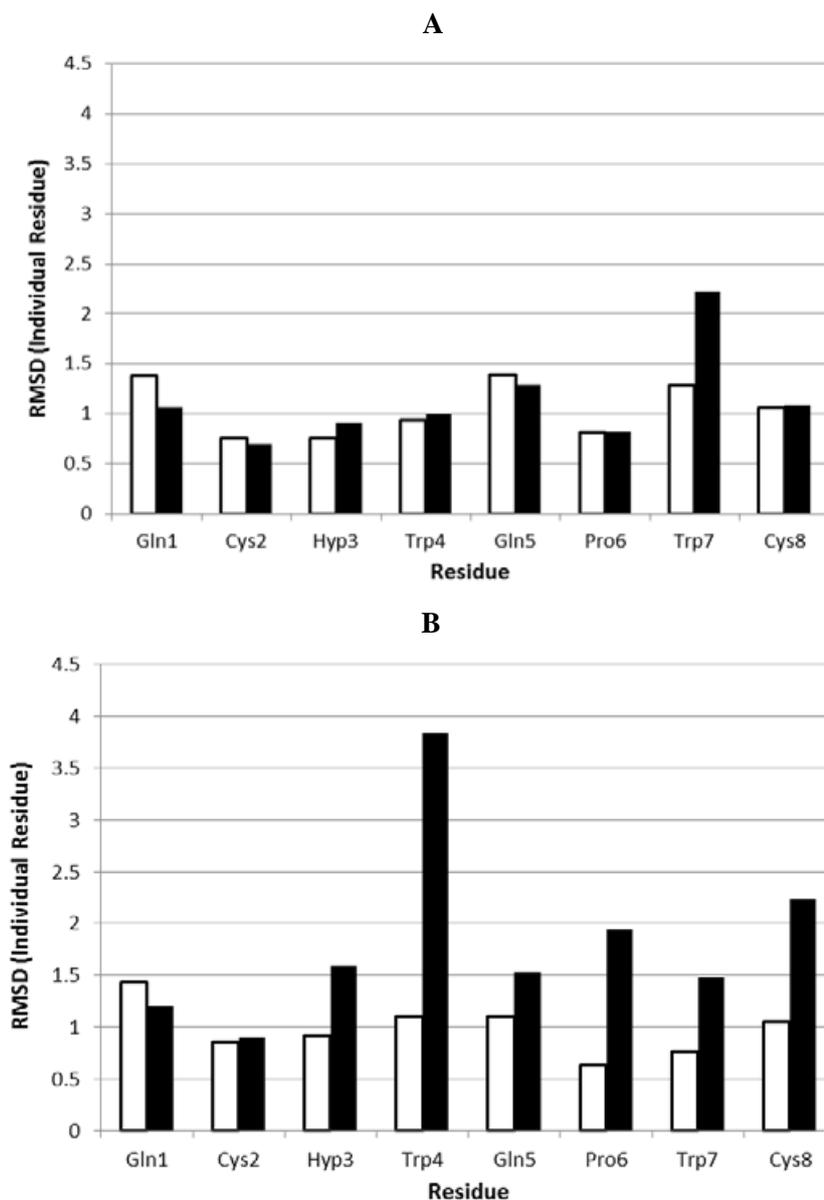


Figure 4. Individual amino acid residue stability. Average RMSD values (in Å) represent fluctuations in the individual residues of a single contryphan molecule. MD simulation (A) run 1 at 0-2 ns, and (B) run 10 at 18-20 ns. Each simulation run generated 2000 structures within each time frame of 2 ns (one structure/ps). White bars represent contryphan-Sm residues; black bars represent [L-Trp4]contryphan-Sm residues.

Ramachandran (phi-psi) plots were generated to further investigate the structural implications of the L/D-Trp4 in the contryphan analogs (Figure 5). Observations on the different phi (Φ) and psi (Ψ) angles for the different residues suggest minimal disruption of the alpha carbon backbone throughout the MD simulations. Most residue peaks were within the expected areas for secondary structures

(beta-sheet and helical structures) with the exceptions of Gly1, D-Trp4 and Cys8. Both Gly1 and Cys8 are placed outside the allowed regions, with 0 values for their Φ and Ψ angles, respectively, by virtue of their placement at the N and C termini of the peptide. Although outside the left-handed helix region, D-Trp4 in contryphan-Sm was nearer compared to the L-Trp4 in [L-trp4]contryphan-Sm

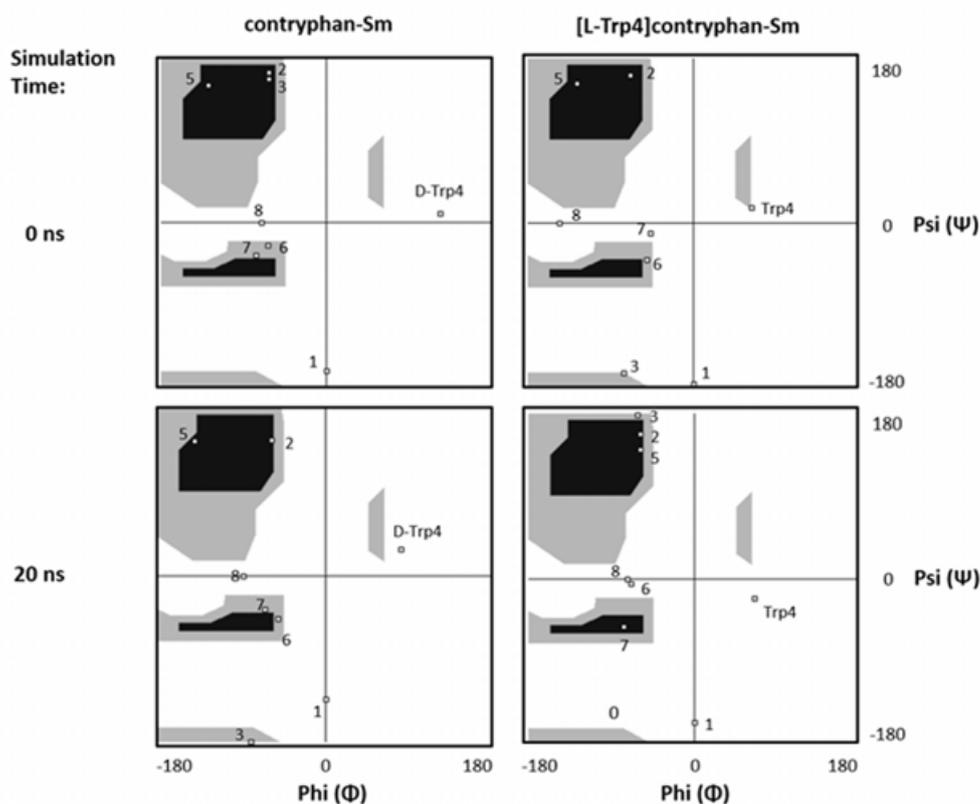


Figure 5. Ramachandran plots for contryphan-Sm and [L-Trp4]contryphan-Sm at 0 and 20 ns of MD simulations. Amino acid residues, except Trp4, are labeled using their residue numbers (1-8). Gray and black regions within the plot represent allowed combinations of Φ and Ψ angles in the common secondary structures (i.e., beta-sheet and helical structures).

at the end of 20 ns simulation. It is interesting to note the variance of the Φ and Ψ angles of adjacent residues linked to either D-Trp4 or L-Trp4 (i.e., Hyp3 and Gln5). Variance in Hyp3 is mainly observed for its Ψ value which is associated with its connection to the succeeding amino acid, D/L-Trp4. Similar behavior is seen for the changes in this angle in the two contryphan analogs. In contrast, variance in Gln5 is mainly observed with changes in its Φ angle that is associated with the preceding amino acid, D/L-Trp4. While the Φ angle of Gln5 is almost constant for contryphan-Sm, greater variance is observed for this angle of Gln5 in [L-Trp4]contryphan-Sm. Moreover, greater fluctuations were observed in the Ψ values for Pro6 and Trp7 of [L-Trp4]contryphan-Sm compared with those of contryphan-Sm. These results provide further support for a more stabilized structure in contryphan-Sm.

The stability of the structures attained may have functional implications due to changes in solvent

accessibility of key amino acid residues. The solvent exposure of D/L-Trp4 was observed throughout the course of the MD simulations. While solvent accessibilities of D-Trp4 and L-Trp4 were similar at 0 ns, the conformational change in contryphan-Sm led to a decrease in the water molecules surrounding D-Trp4 at the end of ~ 20 ns of MD simulation. This may be attributed to the formation of a hydrophobic pocket between D-Trp4 and Pro6 upon taking on a stacked conformation (Figure 6A). The Pro6 showed a similar decrease in its surrounding water molecules at the end of ~ 20 ns, with the adoption of a stacked conformation with D-Trp4 (Figure 6B). In contrast, the extended conformation of [L-Trp4]contryphan-Sm was retained throughout the simulation with no change in solvent accessibility of L-Trp4. It must be noted that data shown in Figure 6 involves structures after 19.5 ns of simulation (i.e., frame 1421/1999 in the 10th 2 ns production run). This timepoint was chosen as

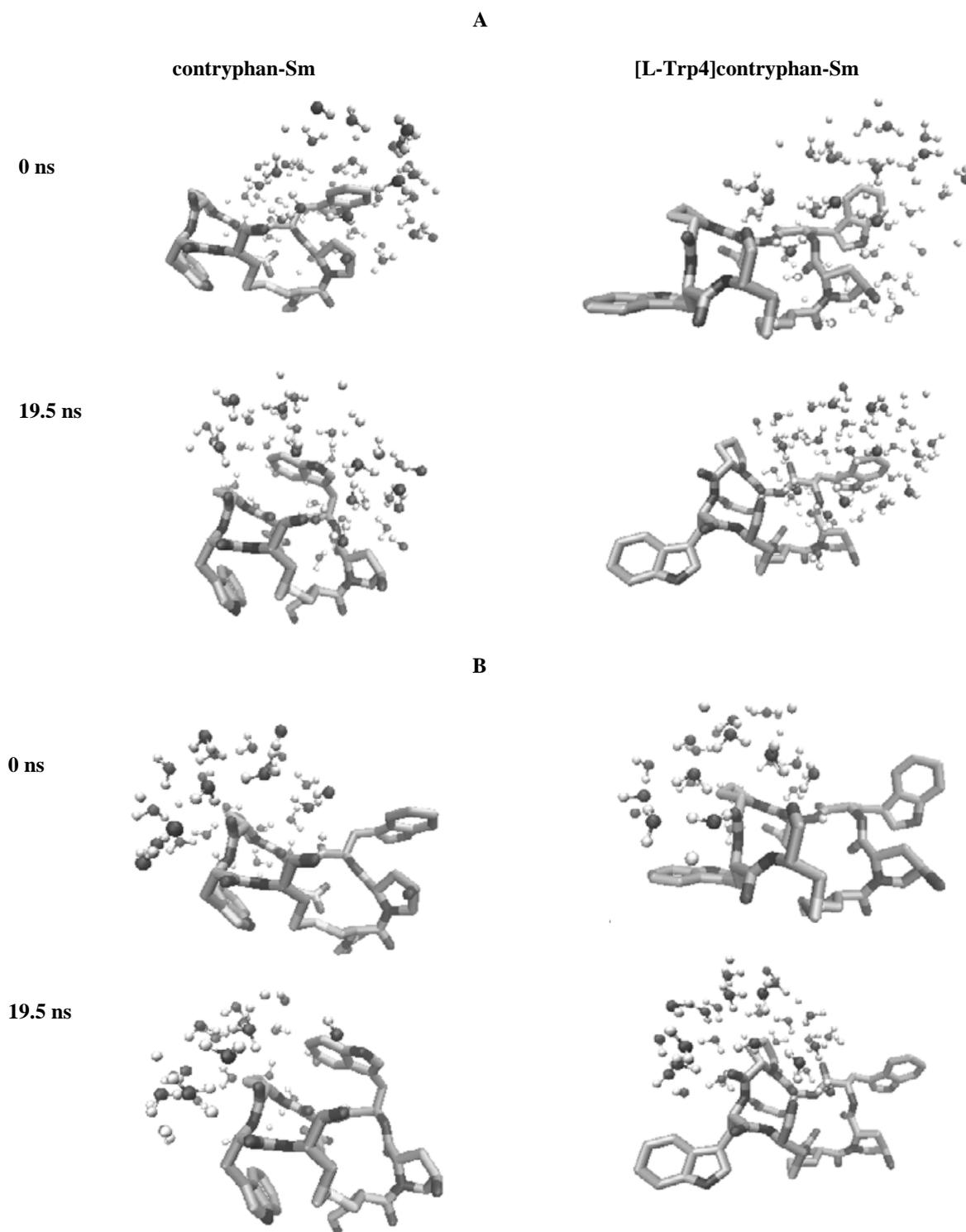


Figure 6. Comparison of solvent accessibilities of contryphan-Sm and [L-Trp4]contryphan-Sm. The images represent the distribution of water molecules within 0.5 nm of (A) D/L-Trp4 and (B) Pro6, in contryphan-Sm and [L-Trp4]contryphan-Sm. Data for each peptide was obtained at 0 ns and ~20 ns of MD simulation of each single contryphan molecule immersed in water. Water molecules are presented in ball and stick format to distinguish them from the contryphan structures.

it was the last for which both molecules were completely immersed in water. Later simulation timepoints had [L-Trp4]contryphan-Sm located at the edge of the water box, which prevented the uniform distribution of water molecules around its structure and the proper analysis of its solvent exposure.

To quantify solvent exposure, the solvent-activated surface areas (SASA) for the two contryphan analogs were computed and compared. For contryphan-Sm the SASA decreased from 10.37 to 9.72 nm², whereas for [L-Trp4]contryphan-Sm the SASA increased from 10.62 to 11.23 nm² after ~ 20 ns of MD simulations. The SASA for the corresponding amino acid residues were further compared (Table 1). In contryphan-Sm, the SASA decreased for most residues except for Hyp3 and Gln5 which increased from 1.01 to 1.43 nm² and from 0.66 to 0.79 nm², respectively. The large reductions in SASA for D-Trp4 (from 2.37 to 2.15 nm²), Pro6 (from 1.41 to 1.07 nm²) and Trp7 (from 2.31 to 1.99 nm²) coincided with the hydrophobic pocket described above. The movement of D-Trp4 and its stacking with Pro6 and Trp7 may explain the greater solvent exposure of Hyp3 and Gln5. In [L-Trp4]contryphan-Sm, a large increase in SASA was obtained for Trp7 (from 2.09 to 2.37 nm²). The SASA for L-Trp4 in [L-Trp4]contryphan-Sm were consistently high, recorded at 2.49 nm² at 0 ns, and 2.74 nm² after 20 ns of MD simulation.

The potential for the indole rings of D/L-Trp4 and Trp7 in contryphan-Sm and [L-Trp4]contryphan-Sm to form an intramolecular covalent crosslink was examined. The possibility of forming this bond would be affected by the accessibility of potential partner Trp residues. The atoms selected as anchor points for the Trp-Trp distance measurements were based on the potential interaction sites between C2s of the indole rings of the Trp side chains. The use of these points allowed the assessment of the likelihood that a Trp-Trp covalent crosslink could form in the conformations attained with the contryphan. Thus, the separation of C2 atoms of the indole rings was used to assess this possibility. After 20 ns of MD simulations, the average intramolecular distances between D/L-Trp4 and Trp7 for contryphan-Sm and [L-Trp4]contryphan-Sm were 1.036 ± 0.106 nm and 1.202 ± 0.111 nm, respectively. Similar results were observed for the 27-molecule system in which the average intramolecular distances between D/L-Trp4 and Trp7 for contryphan-Sm and [L-Trp4]contryphan-Sm were 1.006 ± 0.118 nm and 1.201 ± 0.101 nm, respectively. In both contryphan analogs, the measured distances between Trp residues do not favor intramolecular crosslinking as they are generally beyond the expected interaction radii for a C-C single covalent bond that would have a length of about 0.154 nm [33]. The predicted separations are beyond the 0.500 nm limit of detectable through-space interactions in NMR NOESY experiments [34].

Table 1. Solvent-activated surface areas (SASA) of amino acid residues for contryphan-Sm and [L-Trp4] contryphan-Sm at 0 and ~ 20 ns of MD simulations.

Amino acid residue	Contryphan-Sm SASA (nm ²)		[L-Trp4]contryphan-Sm SASA (nm ²)	
	0 ns	~ 20 ns	0 ns	~ 20 ns
Gly1	1.12	0.98	0.96	1.09
Cys2	0.66	0.43	0.62	0.61
Hyp3	1.01	1.43	1.14	1.18
Trp4	2.37	2.15	2.49	2.74
Gln5	0.66	0.79	0.97	0.87
Pro6	1.41	1.07	1.30	1.38
Trp7	2.31	1.99	2.09	2.37
Cys8	1.03	1.02	1.26	1.15

DISCUSSION

The molecular models prepared for contryphan-Sm and [L-Trp4]contryphan-Sm showed similar basic cyclic structures. Very recently, contryphan-Vc2 and [L-Trp3]contryphan-Vc2 were shown to have structural similarities as well [16]. The similarities in structures may explain the similar biological activities of contryphan-Vc2 and [L-Trp3]contryphan-Vc2 [16] in mouse bioassays. Similar biological activities based on mouse bioassays were also observed between contryphan-R and [L-Trp4]contryphan-R [6].

The more extended structure of [L-Trp4]contryphan-Sm than that of contryphan-Sm and the higher RMSD values of the amino acid residues may explain the longer retention time and broader peak observed in reverse-phase HPLC of [L-Trp4]contryphan-Sm compared to contryphan-Sm [8]. Similar observations were noted between [L-Trp4]contryphan-R and contryphan-R when subjected to reverse-phase HPLC [6]. The L-Trp4 (a hydrophobic amino acid) in [L-Trp4]contryphan-Sm, being more exposed would make the peptide behave to be more nonpolar, thus with stronger interaction towards the stationary phase. On the other hand, the D-Trp4 in contryphan-Sm appears to be 'buried' in the peptide fold due to its stacking with Pro6 and Trp7. Moreover, the broader peak observed for [L-Trp4]contryphan-Sm could be due to greater fluctuations in the molecular structure, as shown by higher RMSD values of the individual amino acids, particularly for L-Trp4.

The formation of purple to blue-violet derivative was observed during the chemical synthesis of [L-Trp4]contryphan-R without nitrogen treatment, but not in contryphan-R [6]. Nitrogen was used to prevent oxidation by eliminating the oxygen in the system. In [L-Trp4]contryphan-R and [L-Trp4]contryphan-Sm, only the Trp residues are expected to be modified to form blue-violet derivatives under the oxidizing conditions of chemical synthesis. Incidentally, the blue indigo dye has a double covalent bond between C2 atoms of two 3-oxindole rings. Although indigo is not normally synthesized from indole rings of two Trp residues, the structural similarity between these functional groups gave a lead for the further examination of the indole rings of Trp residues in contryphan.

Formation of Trp dimers *via* single covalent bond between C2 atoms of indole rings in Trp residues has been documented [35]. A single covalent bond between Trp6 and Trp9 was found in X-indolicidin (ILPWKWPWWPWR*), a byproduct formed during the chemical synthesis of indolicidin. Interestingly, X-indolicidin and indolicidin showed comparable antimicrobial activities [36]. Hence, it was investigated if C2-C2 single covalent bond between the indole rings of Trp4 and Trp7 could possibly occur in [L-Trp4]contryphan-Sm. A C2-C2 single bond formed between the indole rings of Trp4 and Trp7 would be about 0.154 nm [33]. Thus, the much longer distance (> 1.0 nm) between Trp4 and Trp7 in [L-Trp4]contryphan-Sm does not seem to favor a similar Trp-Trp interaction that could lead to the formation of an intramolecular covalent bond. The presence of a disulfide bond between Cys2 and Cys8 in contryphan-Sm or [L-Trp4]contryphan-Sm contributes to the structural stability of the peptide, presumably eliminating the possibility of an intramolecular Trp4-Trp7 covalent crosslink. In contrast, X-indolicidin does not have any disulfide bond, and the covalent crosslink between C2s of the indole rings of Trp6 and Trp9 may have caused a stabilizing effect on its structure, thereby maintaining its antimicrobial activity.

Consequently, the possibility of Trp4 oxidation in [L-Trp4]contryphan-R, as well as in [L-Trp4]contryphan-Sm, was analyzed. Because of the less compact structure, the greater RMSD value and greater solvent accessibility of [L-Trp4]contryphan-Sm as shown by MD simulations, the L-Trp4 in [L-Trp4]contryphan-Sm could be more prone to oxidative degradation, without treatment with nitrogen that is normally used to prevent oxidation. Oxidation involving the indole rings of Trp residues in peptides and proteins producing blue to violet products that include N-formylkynurenine and kynurenine has been documented. The Trp residue in monoclonal antibody drug formulation formed N-formylkynurenine, kynurenine and other colored products [37]. N-formylkynurenine was identified as a post-translationally modified product of Trp residues in cardiac mitochondrial proteins [38].

A scheme showing that the L- to D-Trp4 isomerization could occur prior to the proteolytic cleavage of the contryphan propeptide has been

proposed [7]. This appears consistent with the fact that despite the numerous contryphans that have been identified so far in *Conus* venoms, no all L-amino acid contryphan or contryphan with oxidized L-Trp has been isolated. On the other hand, the L-to D-Phe3 isomerization in the 72-residue crustacean hyperglycemic hormones, [D-Phe3]cHHA and [D-Phe3]cHHB, was shown to occur after the proteolytic cleavage of the propeptides. This coincided with the isolation of the corresponding all-L-amino acid hormone analogs, [L-Phe3]cHHA and [L-Phe3]cHHB [39, 40]. Although the Phe residue in protein may be labile to oxidative processes [41, 42], the much bigger [L-Phe3]cHHA and [L-Phe3]cHHB, the nature of the L-Phe3 side chain compared to L-Trp4 side chain in the much smaller [L-Trp4]contryphan-Sm and [L-Trp4]contryphan-R [5, 6], and the location of L-Phe3 in the 3-dimensional structures of the peptide hormones may provide protection for L-Phe3 from possible oxidative degradation.

CONCLUSION

The present work showed that contryphan-Sm has a more compact structure compared to [L-Trp4]contryphan-Sm. The D-Trp4 and other amino acid residues in contryphan-Sm showed less flexibility and variability. Moreover, the D-Trp4 exhibited less solvent accessibility. The results indicate that the D-Trp could confer greater structural stability for contryphan-Sm and possibly other D-Trp-containing contryphans.

The cone snails of the genus *Conus* employ venom to catch prey (such as worm, mollusk or fish), guard against predators and dissuade competitors. Part of the success of these organisms may be attributed to their ability to thrive in the diverse, often extreme environments of the ocean floor. Each *Conus* species produces a large variety of conotoxins suited for its ecological targets in these varied locales [3, 4]. Increased structural stability for some of these conotoxins would enhance the survival of the cone snail in extreme conditions. D-amino acid-containing contryphans, with their increased conformational stability may help provide the sustained biological activity that allows the 700 living species of *Conus* to be among the most flourishing of marine animals.

ACKNOWLEDGEMENTS

We are grateful to Dr. Ricky Nellas of the Institute of Chemistry, University of the Philippines Diliman for the discussion regarding the computer programs and for technical assistance in the molecular modeling and MD simulations. We thank Adam Jo Elatico and Yeo Daniel Bascos for technical assistance. We appreciate the support of NIMBB-UPD (to NADB) and UP Baguio (to ECJ).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

REFERENCES

1. Jimenez, E. C. 2007, *Curr. Topics Peptide Protein Res.*, 8, 11.
2. Jimenez, E. C. 2017, *Phil. Sci. Lett.*, 10, 72.
3. Olivera, B. M. 2006, *J. Biol. Chem.*, 281, 31173.
4. Terlau, H. and Olivera, B. M. 2004, *Physiol. Rev.*, 84, 41.
5. Jacobsen, R., Jimenez, E. C., Grilley, M., Watkins, M., Hillyard, D., Cruz, L. J. and Olivera, B. M. 1998, *J. Peptide Res.*, 51, 173.
6. Jimenez, E. C., Olivera, B. M., Gray, W. R. and Cruz, L. J. 1996, *J. Biol. Chem.*, 271, 28002.
7. Jimenez, E. C., Craig, A. G., Watkins, M., Hillyard, D. R., Gray, W. R., Gulyas, J., Rivier, J. E., Cruz, L. J. and Olivera, B. M. 1997, *Biochemistry*, 36, 989.
8. Jacobsen, R., Jimenez, E. C., De la Cruz, R., Gray, W. R., Cruz, L. J. and Olivera, B. M. 1999, *J. Peptide Res.*, 54, 93.
9. Pallaghy, P. K., Melnikova, A. P., Jimenez, E. C., Olivera, B. M. and Norton, R. S. 1999, *Biochemistry*, 38, 11553.
10. Pallaghy, P. K., He, W., Jimenez, E. C., Olivera, B. M. and Norton, R. S. 2000, *Biochemistry*, 39, 12845.
11. Eliseo, T., Cicero, D. O., Romeo, C., Schininà, M. E., Massilia, G. R., Polticelli, F., Ascenzi, P. and Paci, M. 2004, *Biopolymers*, 74, 189.
12. D'Alessandro, M., Paci, M. and Amadei, A. 2004, *Biopolymers*, 74, 448.
13. Sonti, R., Gowd, K. H., Rao, K. N., Ragothama, S., Rodriguez, A., Perez, J. J. and Balaram, P. 2013, *Chemistry*, 19, 15175.
14. Nepravishta, R., Mandaliti, W., Melino, S., Eliseo, T. and Paci, M. 2014, *Amino Acids*, 46, 2841.

15. Rajesh, R. P. 2015, *J. Pept. Sci.*, 21, 20.
16. Drane, S. B., Robinson, S. D., MacRaild, C. A., Chhabra, S., Chittoor, B., Morales, R. A., Leung, E. W., Belgi, A., Espino, S. S., Olivera, B. M., Robinson, A. J., Chalmers, D. K. and Norton, R. S. 2017, *Toxicon.*, 129, 113.
17. Vijayasathy, M., Basheer, S. M., Franklin, J. B. and Balaram, P. 2017, *J. Proteome Res.*, 16, 763.
18. Case, D. A., Cheatham, T. E. III, Darden, T., Gohlke, H., Luo, R., Merz, K. M., Onufriev, A., Simmerling, C., Wang, B. and Woods, R. J. 2005, *J. Comput. Chem.*, 26, 1668.
19. Case, D. A., Babin, V., Berryman, J. T., Betz, R. M., Cai, Q., Cerutti, D. S., Cheatham, T. E. III, Darden, T. A., Duke, R. E., Gohlke, H., Goetz, A. W., Gusarov, S., Homeyer, N., Janowski, P., Kaus, J., Kolossváry, I., Kovalenko, A., Lee, T. S., LeGrand, S., Luchko, T., Lu, R., Madej, B., Merz, K. M., Paesani, F., Roe, D. R., Roitberg, A., Sagui, C., Salomon-Ferrer, R., Seabra, G., Simmerling, C. L., Smith, W., Swails, J., Walker, R. C., Wang, J., Wolf, R. M., Wu, X. and Kollman, P. A. 2014, *AMBER 14*, University of California, San Francisco.
20. Wang, J., Wang, W., Kollman, P. A. and Case, D. A. 2006, *J. Mol. Graph.*, 25, 247.
21. Bernstein, H. J., Duque, C., Grossman, G., Molinaro, M., Mueller, A., Salame, N., Sayle, R., and Valadon, P. 2004, *RasTop Molecular Visualization Software ver.*, 2.1.
22. Humphrey, W., Dalke, A. and Schulten, K. 1996, *J. Molec. Graphics*, 14, 33.
23. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. and Klein, M. L. 1983, *J. Chem. Phys.*, 79, 926.
24. Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kalé, L. and Schulten, K. J. 2005, *J. Comput. Chem.*, 26, 1781.
25. Adelman, S. A. and Doll, J. D. 1976, *J. Chem. Phys.*, 64, 2375.
26. Ryckaert, J. P., Ciccotti, G. and Berendsen, H. J. C. 1977, *J. Comput. Phys.*, 23, 327.
27. Darden, T., York, D. and Pedersen, L. 1993, *J. Chem. Phys.*, 98, 10089.
28. Falsafi-Zadeh, S., Karimi, Z. and Galehdari, H. 2012, *Bioinformation*, 8, 341.
29. Ramachandran, G. N. and Sasisekharan, V. 1968, *Adv. Protein Chem.*, 23, 283.
30. Lund, M. 2014, *vmd.sasa.tcl*. A solvent accessible surface area (SASA) computing script. (https://codedocs.xyz/mlund/faunus/vmd-sasa_8tcl_source.html)
31. Falsafi, S. and Karimi, Z. 2011, *sasa.tcl*. Asolvent accessible surface area (SASA) computing script. (http://www.ks.uiuc.edu/Research/vmd/mailling_list/vmd-l/att-18670/sasa.tcl)
32. Becker, O. M. and Watanabe, M. 2001, *Computational Biochemistry and Biophysics*, O. M. Becker, A. D. Jr., MacKerell, B. Roux and Watanabe, M. (Eds.), Marcel Decker, Inc., New York, 39.
33. Bartell, L. S. 1959, *J. Am. Chem. Soc.*, 81, 3497.
34. Wüthrich, K. 1986, *NMR of proteins and nucleic acids*. New York: Wiley, 3.
35. Stachel, S. J., Habeeb, R. L. and van Vranken, D. L. 1993, *J. Am. Chem. Soc.*, 118, 1225.
36. Ösapay, K., Tran, D., Ladokhin, A. S., White, S. H., Henscheni, A. H. and Selsted, M. E. 2000, *J. Biol. Chem.*, 275, 12017.
37. Li, Y., Polozova, A., Gruia, F. and Feng, J. 2014, *Anal. Chem.*, 86, 6850.
38. Taylor, S. W., Fahy, E., Murray, J., Capaldi, R. A. and Ghosh, S. S. 2003, *J. Biol. Chem.*, 278, 19587.
39. Soyez, D., Van Herp, F., Rossier, J., Le Caer, J-P., Tensen, C. P. and Lafont, R. 1994, *J. Biol. Chem.*, 269, 18295.
40. Soyez, D., Toullec, J. Y., Ollivaux, C. and Geraud, G. 2000, *J. Biol. Chem.*, 275, 37870.
41. Stadtman, E. R. 1993, *Annu. Rev. Biochem.*, 62, 797.
42. Silva, A. M., Marcal, S. L., Vitorino, R., Domingues, M. R. and Domingues, P. 2013, *Arch. Biochem. Biophys.*, 530, 23.