

Article

# Tripleurin XIIc: Peptide Folding Dynamics in Aqueous and Hydrophobic Environment Mimic Using Accelerated Molecular Dynamics

Chetna Tyagi <sup>1,2,\*</sup>, Tamás Marik <sup>1</sup>, András Szekeres <sup>1</sup>, Csaba Vágvölgyi <sup>1</sup>, László Kredics <sup>1</sup> and Ferenc Ötvös <sup>3</sup>

<sup>1</sup> Department of Microbiology, Faculty of Science and Informatics, University of Szeged, Szeged, Közép fasor 52, H-6726 Szeged, Hungary; mariktamas88@gmail.com (T.M.); szandras@bio.u-szeged.hu (A.S.); mucor1959@gmail.com (C.V.); kredics@bio.u-szeged.hu (L.K.)

<sup>2</sup> Doctoral School of Biology, Faculty of Science and Informatics, University of Szeged, Szeged, Közép fasor 52, H-6726 Szeged, Hungary

<sup>3</sup> Institute of Biochemistry, Biological Research Centre, Szeged, Temesvári krt. 62, H-6726 Szeged, Hungary; otvos@brc.hu

\* Correspondence: cheta231@gmail.com; Tel.: +36-62-544005

Received: 24 December 2018; Accepted: 18 January 2019; Published: 19 January 2019

**Abstract:** Peptaibols are a special class of fungal peptides with an acetylated *N*-terminus and a C-terminal 1,2-amino alcohol along with non-standard amino acid residues. New peptaibols named tripleurins were recently identified from a strain of the filamentous fungal species *Trichoderma pleuroti*, which is known to cause green mould disease on cultivated oyster mushrooms. To understand the mode of action of these peptaibols, the three-dimensional structure of tripleurin (TPN) XIIc, an 18-mer peptide, was elucidated using an enhanced sampling method, accelerated MD, in water and chloroform solvents. Non-standard residues were parameterized by the Restrained Electrostatic Potential (RESP) charge fitting method. The dihedral distribution indicated towards a right-handed helical formation for TPN XIIc in both solvents. Dihedral angle based principal component analysis revealed a propensity for a slightly bent, helical folded conformation in water solvent, while two distinct conformations were revealed in chloroform: One that folds into highly bent helical structure that resembles a beta-hairpin and another with an almost straight peptide backbone appearing as a rare energy barrier crossing event. The hinge-like movement of the terminals was also observed and is speculated to be functionally relevant. The convergence and efficient sampling is addressed using Cartesian PCA and Kullback-Leibler divergence methods.

**Keywords:** peptaibols; enhanced sampling; molecular dynamics simulations; solvent; principal component analysis; tripleurin

## 1. Introduction

Small bioactive peptides like peptaibols have piqued the interests of microbiologists owing to their antibacterial, antifungal, anti-viral, anti-helminth, and anti-tumor properties [1–5], as well as their abilities to elicit plant defense responses [6,7]. Peptaibols are linear, non-ribosomally produced amphipathic polypeptides of fungal origin, mostly comprising a high ratio of unusual amino acid content. Non-standard amino acid residues like aminoisobutyric acid (Aib), *D*-isovaline (Div), hydroxy-proline (Hyp), and C-terminal alcohol residues like phenylalaninol (Pheol), valinol, etc., along with an acetylated *N*-terminal (Ac) are characteristic for these peptides ranging seven–20 amino acid residues in length. They are synthesized by large modular enzymes called non-ribosomal peptide synthetases (NRPSs), where a single module contains multiple catalytic domains

responsible for the incorporation of a single amino acid residue into the peptide chain. Due to the relaxed specificity of NRPSs for residues, the peptaibol groups produced by a fungal strain show considerable heterogeneity [8]. For example, Raap et al. [9] reported that, upon addition of free Aib residues into the medium, the NRPS system of *Emericellopsis salmosynnemata* produces Zervamicin-IIa (containing only Aib residues) as the major secondary metabolite, while upon addition of L/D-Iva, the same NRPS produces Zervamicin-IIb with one D-Iva residue (replacing the first Aib residue in Zervamicin IIa), which indicates the lack of selectivity of NRPS for this substrate. Peptaibols show antibiotic properties by forming pores within biological membranes via aggregation, the implication of which has been widely studied [10–12]. Apart from studies on their bioactivity, substantial focus has been given to their three-dimensional structures and folding dynamics [13–16]. In the protein databank (PDB, [www.rcsb.org](http://www.rcsb.org)) about 15 experimentally (X-ray crystallography or NMR) known structures have been documented so far. Alamethicin is the most studied peptaibol so far, closely followed by antiameobin and zervamicin. Nagao et al. [17] studied the orientation of alamethicin in phospholipid bilayers and found that helical axes of N and C-terminals were tilted by 17° and 32° degrees, respectively, to the bilayer normal. The difference in tilt angles for the termini indicates towards a central kink and bends in the alamethicin structure. In fact, such observation has been made for all trans-membrane proteins by Hall et al. [18], who showed that 44% of trans-membrane helices contained a significant kink, proline being the cause 35% of the times. It can be speculated that the occurrence of this bend must have a crucial functional role either during membrane disruption or in the ion-channel formation.

In this study, we carried out structure elucidation of tripleurins, comprising a newly identified group of peptaibols produced by the fungus *Trichoderma pleuroti*, which causes a green mold disease in the cultivation of oyster mushroom (*Pleurotus ostreatus*) [19]. Tripleurins were also reported as potential growth inhibitors of oyster mushroom mycelia. Not much is known about their behavior and dynamics in different environments and this is the first study to understand their folding behavior in water and chloroform. The choice of chloroform as a solvent was made due to its low dielectric constant ( $\epsilon_r = 4.80$ ) that makes it a suitable mimic of the hydrophobic environment provided by biological membrane systems. It is commonly known that such anti-microbial peptides interact with host lipid bilayers and affect their integrity in several ways. Therefore, we were curious to observe any possible structural changes that might occur in such an environment. LC-MS (Liquid Chromatography-Mass Spectrometry) approaches could not distinguish between isobaric residues, which are therefore marked as Vxx for L-valine or L- and D-isovaline and Lxx for L-leucine or L-isoleucine. Similarly, their 1,2-amino alcohols are referred to as Vxxol and Lxxol. From a mixture of various tripleurins, tripleurin XIIc (referred to as TPN XIIc henceforth) has been selected based on its high yield and lowest number of ambiguous positions.

We also attempted to understand the evolution of peptide secondary structure based on Ramachandran plots (also referred as phi-psi plots/distribution) through the course of simulations carried out in different solvents. Hollingsworth and Karplus [20] described five main conformation clusters on the plot, namely,  $\alpha$ -helices,  $\beta$ -strands, polyproline II (PII) -spirals,  $\gamma$ -turns, and  $\gamma'$ -turns along with an additional  $\epsilon$ -region (later named PII' region) and the bridge region between  $\alpha$ -helices and  $\beta$ -strands ( $\delta$  and  $\delta'$ ). However, only three broad structure types occur in terms of linear groups, i.e., series of residues with the same repeating conformation. These groups are (1)  $\alpha/3_{10}$  helices, (2) a group that is largely made of  $\beta$ -strands, and (3) the last group that adopts a PII spiral conformation,  $\phi, \psi = (-65, +145)$ , while the left-handed helices may exist as short segments. They also mentioned  $\beta$ -turns, PII-spirals,  $\gamma$ - and  $\gamma'$ -turns,  $\zeta\alpha$ , and  $\zeta$ PII regions as less regular structures. The structural and dynamic study of peptaibols has been quite challenging due to the presence of non-standard residues, their stereochemistry and the lack of experimental structures. The question of simulation time length and adequate sampling remains a bottleneck in most cases involving unfolded structures. To overcome the sampling problem, an all-atom enhanced sampling technique called accelerated molecular dynamics (aMD) has been employed, which provides a non-negative boost to the potential, which reduces the energy barriers between transition states [21,22]. It has been demonstrated that a 500 ns long aMD simulation (using GPUs) could traverse an energy landscape

equivalent to a 1 ms long classical MD simulation [23]. Moreover, it does not require any predefined reaction coordinates, unlike other biased free energy calculation methods [24]. We carried out four independent simulations in water and chloroform solvents. The first three simulations were carried out for 500 ns, while the fourth was carried out for 1  $\mu$ s (1000 ns) using different boost parameters. The dynamics of TPN XIIc was studied in detail using principal component analysis on internal/dihedral angle coordinates and effective sampling was assessed using Cartesian coordinate-based PCA. A relatively new method, named Kullback-Leibler divergence, to measure extent of overlap between probability distribution [25], has been used to discuss adequate sampling and convergence.

We focused on the folding dynamics of a recently discovered 18-mer peptaibol, Tripleurin (TPN) XIIc, which is an important step to unfold the structural properties of tripleurins and an attempt to correlate their structure and function. It is interesting to analyze the behavior of individual residues of the peptide and their effects on the overall conformation. The primary goal was to obtain a stabilized folded conformation of TPN XIIc, as this knowledge can prove beneficial to design highly bioactive compounds.

## 2. Results and Discussion

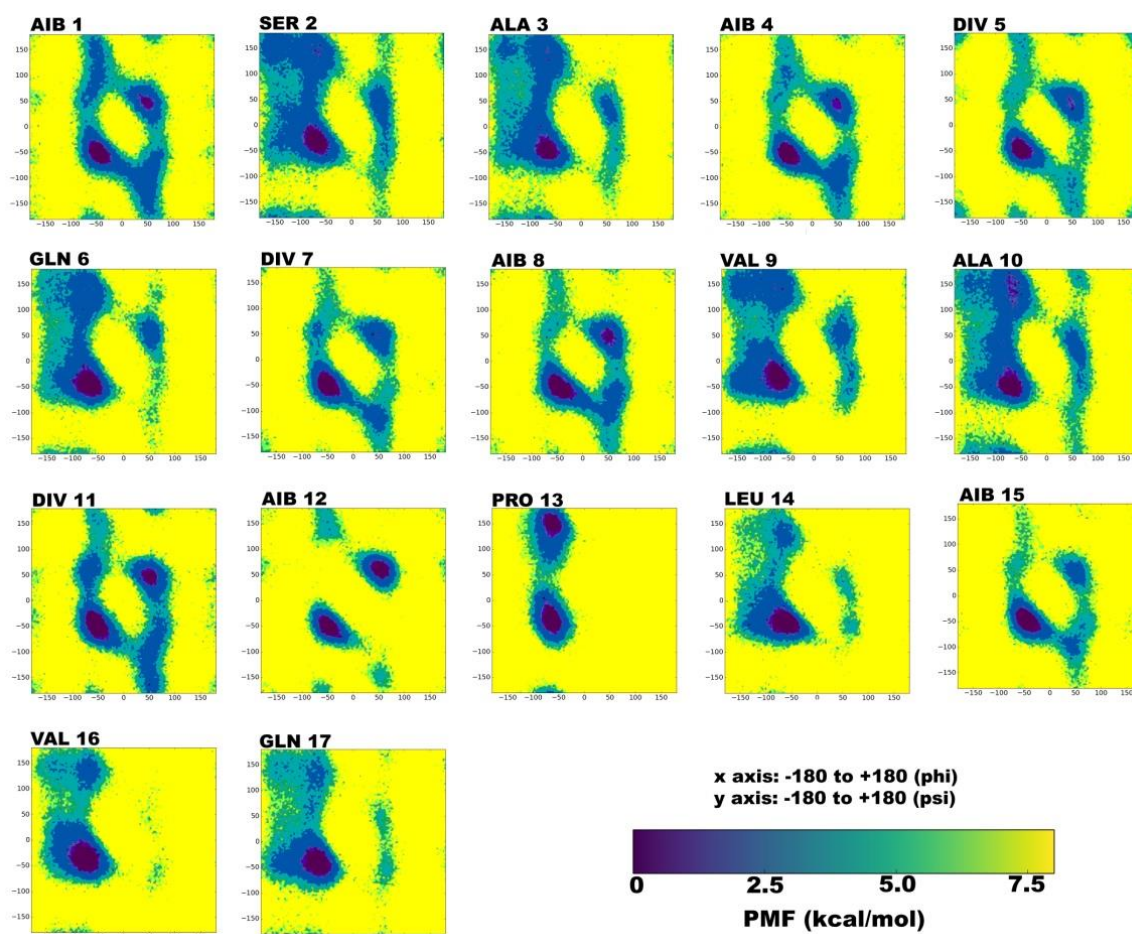
### 2.1. Secondary Structural Populations in Water and Chloroform Solvents

In this study, we characterized the individual amino acid conformations, along with overall conformation based on the nomenclature described by Hollingsworth and Karplus [20]. The average  $\Phi$  and  $\Psi$  angle values for various conformations on the Ramachandran plot are given in Table 1. A detailed analysis for each residue is necessary for future evaluation of their contribution to the overall structure and bioactivity. For example, Aib has been established as the helix shape promoter during many investigations, and was thus used in many helical peptide-designing studies [26–28]. According to Mondal et al. [29], the addition of an extra  $\alpha$ -methyl group limits the range of allowed  $\varphi$ - $\psi$  (phi-psi) torsion angles of Aib in comparison with alanine and restricts the energetically accessible conformational space. It is clear that Aib constraints the backbone by reducing the degree of freedom of movement of the peptide. Another important aspect to be considered here is the achirality of Aib, which enables it to readily form right or left-handed helices depending on the screw-sense of preceding or following residues in the peptide chain. It has also been established that the presence of Aib renders the resulting peptide protease-resistant. Wada et al. [30] synthesized an amphipathic helix peptide using Aib, which showed potent resistance to trypsin and pronase. Additionally, D-residues in general have been shown in various cases to enhance the stability, activity [31], and resistance to proteolytic enzymes [32,33]. It is therefore beneficial to observe the effects of these residues exerted on the structure and function of peptides, and whether this effect is dependent on their number in the peptide chain.

**Table 1.** Average  $\Phi$  and  $\Psi$  angle values for various conformations on the Ramachandran plot (according to Hollingsworth and Karplus) <sup>13</sup>.

Type	$\Phi, \Psi$ angles
$\alpha$ (alpha helix)	-63, -43
$\beta$ (beta region)	-157.2, 161.9
PII-spirals	-65, +145
$\gamma$ -turns	+80, -80
$\gamma'$ -turns	-80, +80
$\delta$ regions	Extending at 45° angle to the left of $\alpha$ -helix region
$\delta'$ regions	Mirror image of $\delta$ region
$\zeta$ (pre-proline region)	-130, +80

The reweighted phi-psi plots have been constructed for each individual residue of TPN XIIc. The color scale in Figures 1–4 denotes potential of mean force or PMF ( $\text{kcal mol}^{-1}$ ) to characterize energetically favored conformations that were visited during simulations. The darkest violet regions denote the lowest energy minimum. In water solvent, the energy minima for non-standard residues, Aib1, Aib4, Div5, Aib8, Div11, and Aib12 lie in both right and left-handed  $\alpha$ -regions. This fluctuation is attributed to the achiral nature of Aib residue and the propensity of D-residues towards left-handed conformations. It has also been noted that chirality of isovaline (Iva) does not highly impact the screw sense of whole helical structure, possibly because of the limited difference in the length of its two side chains [34]. The Div residues are found in right-handed helical peptides, where the screw sense is governed by the L-isomeric amino acids present in their sequences. Another study reported that Iva-rich peptides favor beta-bends and  $3_{10}$  helices [35]. A study by Shenkarev et al. [36] compared structural and functional characteristics of Antiamoebin I (Aam-I) with Zervamicin IIb (Zrv-IIb). Both molecules acquire very similar structural topologies in membrane-mimicking environments, although they significantly differ at the N-terminal (1–8 res.) due to high presence of Aib, Div, and Gly achiral residues in Aam-I. On the other hand, Aam-I and ZRV-IIb show markedly different dynamic properties owing to the former's high 'motional' propensity, which may be caused by conformational exchange (characterized by positive and negative  $\phi/\psi$  torsion angles) undergone by Aib, Div, and Gly residues. They hypothesized that such conformational exchange may be responsible for its high solubility in water solvent, which may prevent Aam-I to effectively perturb lipid bilayers. It is experimentally proven that Aam-I shows lower binding affinity to lipid vesicles than Zrv-IIb, and therefore, also shows weaker bioactivity. This implies that although the presence of achiral residues like Aib is necessary for helix formation, the high fluctuation induced by them may render the peptide less bioactive. In case of TPN XIIc it is clear that the right-handed helical sense is energetically more favorable than the left-handed, although both right and left-handed helical states are visited in the case of Aib and Div residues. Due to such observations, it is difficult to believe that highly diverse peptaibol sequences are driven by evolution. The high heterogeneity within a single group of peptaibols indicates the lack of selectivity in NRPS proteins. On the other hand, the knowledge of correlation between the presence of certain residues and their functional relevance may help in the design and synthesis of relevant biomolecules that could be exploited as therapeutics and biocontrol agents.



**Figure 1.** Reweighted PMF (potential of mean force) phi-psi dihedral angle plots for each TPN XIIc residue during explicit water aMD simulation. The x and y axes range from -180 to +180. The darkest violet regions indicate toward minimum energy secondary structural regions favored by each residue during the simulation.

The energy minima for almost all standard residues like Ser2, Ala3, Gln6, Leu14, Val16, and Gln17 lie in the right-handed  $\alpha$ -region, while the other secondary structural states could only be reached at a difference of 2 kcal mol<sup>-1</sup>. One such state is the bridge between  $\alpha$ - and  $\beta$ -regions, known as the  $\delta$  region which represents  $\beta$ -turn formation in the peptide chain and continuous formation of  $\beta$ -turns gives rise to  $\beta$ -bend ribbon spirals. Most standard residues of TPN XIIc show considerable population in this region even if at a higher energy scale (~4 kcal mol<sup>-1</sup>). On the other hand, Pro13 traverses poly-proline II region, which is expected according to Brown et al. [37] as proline (Pro) shows the highest PII propensity.

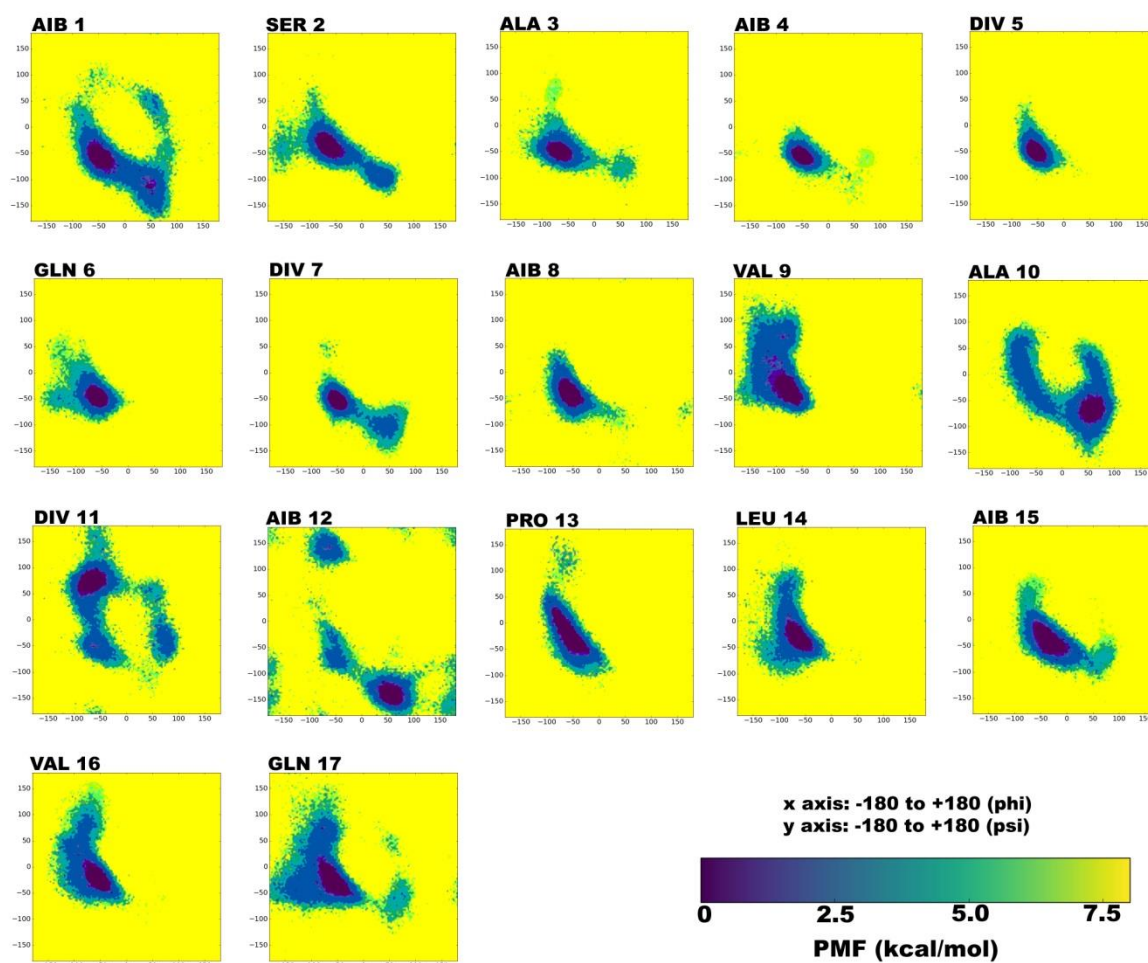
In water environment, TPN XIIc peptide shows fluctuating propensity for  $i + 3 \rightarrow i$  and  $i + 4 \rightarrow i$  type H-bonds (Table 2). Gln6  $\rightarrow$  Ala3 (19%), Val9  $\rightarrow$  Gln6 (15%), Ala10  $\rightarrow$  Div7 (11%), Aib12  $\rightarrow$  Val9 (4%), and Val16  $\rightarrow$  Pro13 (19%) are examples of  $i + 3 \rightarrow i$  type H-bonds, while Div7  $\rightarrow$  Ala3 (14%), Aib12  $\rightarrow$  Aib8 (12%), Ala10  $\rightarrow$  Gln6 (14%), Div11  $\rightarrow$  Div7 (14%), Gln17  $\rightarrow$  Pro13 (21%) and Phe18  $\rightarrow$  Leu14 (44%) are examples of  $i + 4 \rightarrow i$  type H-bonds. On the Ramachandran plot,  $\alpha$ - and  $3_{10}$  regions are present almost in the same subspace and are easily inter-convertible in terms of energy difference. Based on this analysis it can be deduced that a loose spiral-like conformation is favored instead of a strict helical conformation (except at C-terminal).



**Table 2.** Backbone H-bonds of TPN XIc in explicit water solvent along with their frequency of occurrence given by fraction, average distance, and angle.

Acceptor	Donor	Fraction	Average Distance	Average Angle
Aib1	Div5	0.0560	2.89	162.38
Ser2	Gln6	0.0843	2.88	160.03
Ala3	Gln6	0.1911	2.87	154.95
Ala3	Div7	0.1478	2.88	161.59
Aib4	Div7	0.0296	2.90	155.38
Aib4	Aib8	0.0929	2.89	161.68
Div5	Aib8	0.0242	2.88	154.44
Div5	Val9	0.0271	2.89	161.79
Gln6	Val9	0.1597	2.87	154.92
Gln6	Aib8	0.0545	2.78	149.02
Aib8	Aib12	0.1254	2.89	159.16
Gln6	Ala10	0.1417	2.87	160.11
Div7	Val9	0.0611	2.79	147.89
Div7	Ala10	0.1154	2.88	152.30
Div7	Div11	0.1484	2.89	161.04
Val9	Aib12	0.0456	2.88	159.08
Ala10	Aib12	0.0434	2.80	148.52
Ala10	Leu14	0.0351	2.88	159.89
Div11	Leu14	0.1235	2.86	151.96
Div11	Aib15	0.1890	2.88	162.72
Pro13	Val16	0.1974	2.87	151.15
Pro13	Gln17	0.2110	2.87	159.68
Leu14	Gln17	0.1082	2.88	152.47
Leu14	Pol18	0.4430	2.86	160.43

On the contrary, reweighted phi-psi plots obtained from chloroform simulation indicate towards a considerably different folding pattern (Figure 2). The energy minima for Aib1, Aib4, Div5, Div7, Aib8, and Aib15 lie only in the right-handed  $\alpha$ -region, unlike in the water solvent, where both right and left-handed states are visited. A rather unusual behavior is exhibited by Ala10, which favors the classic gamma turn formation, while the consecutive Div11 residue favors inverse gamma turn region. This would probably bend the backbone in a slight 'S'-shape or an outward kink.



**Figure 2.** Reweighted PMF (potential of mean force) phi-psi dihedral angle plots for each TPN XIIc residue during explicit chloroform aMD simulation. The x and y axes range from -180 to +180. The darkest violet regions indicate toward minimum energy secondary structural regions favored by each residue during the simulation.

The corresponding H-bond Aib12  $\rightarrow$  Ala10, which occurs in 44% of the simulation time, also indicates towards a strong backbone reversal at this position due to formation of a gamma turn. The H-bonds formed during this simulation have been summarized in Table 3. This observation can be attributed to the non-polar nature of chloroform environment. A recent study on the behavior of alanine dipeptide in explicit chloroform and water solvents by Rubio-Martinez et al. [38] described that  $C7^{eq}$  conformation only appears as a low energy minimum in chloroform. Standard residues like Ser2, Ala3, Gln6, Val9, Leu14, Val16, and Gln17 show energy minima strictly in the right-handed  $\alpha$ -region.

**Table 3.** Backbone H-bonds of TPN XIIc in explicit chloroform solvent along with their frequency of occurrence given by fraction, average distance, and angle.

Acceptor	Donor	Fraction	Average Distance	Average Angle
Aib1	Div5	0.1538	2.90	162.68
Ser2	Aib4	0.0902	2.82	148.54
Ser2	Gln6	0.2475	2.88	160.23
Ala3	Div5	0.1023	2.80	149.61
Ala3	Div7	0.2394	2.90	162.90
Ala3	Gln6	0.0746	2.88	152.31
Aib4	Aib8	0.0904	2.90	163.41

Gln6	Val9	0.1050	2.89	154.32
Gln6	Div11	0.0767	2.90	152.28
Gln6	Ala10	0.1334	2.87	156.93
Div7	Val9	0.0451	2.91	162.20
Div7	Ala10	0.2448	2.88	152.99
Aib8	Ala10	0.0254	2.88	146.67
Val9	Div11	0.1822	2.83	147.37
Val9	Aib12	0.0186	2.88	161.66
Ala10	Aib12	0.4465	2.79	149.19
Div11	Leu14	0.1384	2.87	155.91
Div11	Aib15	0.0833	2.88	163.24
Aib12	Leu14	0.2099	2.85	151.67
Aib12	Aib15	0.0855	2.89	160.15
Pro13	Aib15	0.1378	2.84	148.97
Pro13	Val16	0.1997	2.88	158.57
Aib15	Pol18	0.2063	2.88	158.80
Leu14	Val16	0.1747	2.81	147.09
Leu14	Gln17	0.2091	2.88	159.58
Leu14	Pol18	0.1474	2.87	161.57
Val16	Pol18	0.0851	2.84	147.39

In conclusion, the TPN XIIc peptide shows higher propensity for spiral-like helix at the N-terminal and  $\alpha/3_{10}$  helix at the C-terminal with a slight backbone bend in water solvent, and for  $\gamma$ -turn in the central region that may induce backbone reversal in chloroform solvent. The C-terminal mostly folds into a  $3_{10}$  helix in both solvents, but is disrupted by few gamma turn-inducing bonds like Val16  $\rightarrow$  Leu14 (17% occurrence) in the chloroform solvent.

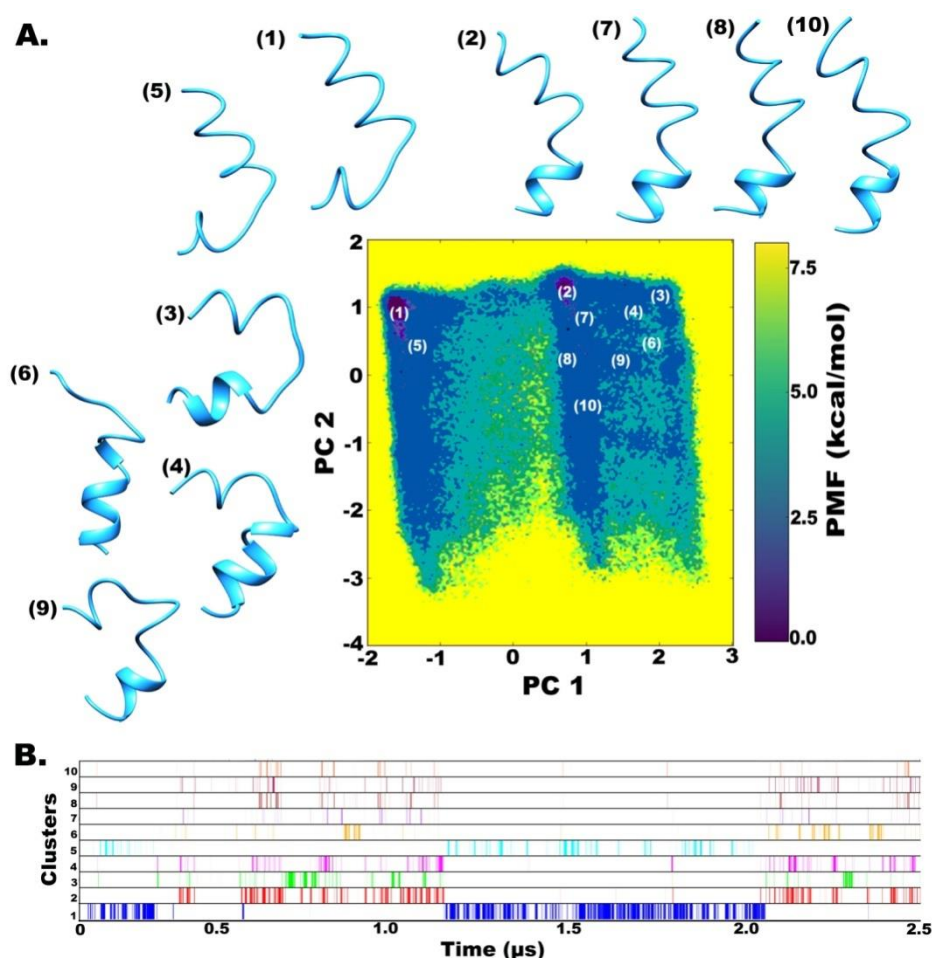
## 2.2. Clustering based on Free Energy Landscapes (FEL): Vision through Principal Component Analysis

In addition to the native structure, we were also interested in the folding properties of TPN XIIc under the effect of different solvents. Peptide folding is a dynamic process of evolution of intermediate ensemble states in a back-and-forth fashion, which should eventually result in a folded state. The results of the simulations were analyzed by principal component analysis (PCA) to reduce the dimensionality of data and to visualize the free-energy landscapes, resulting in the intermediate states and their path to achieve the final folded state [39]. A Cartesian coordinate PCA (cPCA) based on the overall motions of a peptide can distort this idea and present ambiguity in the spread of free-energy basins [40], therefore dihedral-angle based PCA (dPCA) [41] was employed to only include internal motions (defined by  $\phi$ ,  $\psi$  dihedral angles) for the peptide folding process. A free energy landscape based on internal motions (projected along first two principal components PC1 and PC2 using  $\mu(q_1, q_2) = -k_B T \ln P(q_1, q_2)$ ) provides accurate results of the minimum energy wells and barriers between them, especially for systems undergoing large changes [42]. All dihedral PCA-based FEL plots have been reweighted for calculation of PMF and accurate description of the energy minimum.

The two simulations were carried out for 500 ns  $\times$  3 using different starting structures along with another 1000 ns long simulation. All simulations have been combined (2.5  $\mu$ s) for the clustering procedure. The clustering procedure involves identification of isolated peaks in a three- or five-dimensional density map obtained from the trajectory's principal component distribution. Each point on this plot signifies all the structures that have PC values closest to that point. All peaks with density higher than a given threshold are selected, which correspond to a distinct cluster. The darkest violet regions on the FEL map (Figure 3a) show the lowest energy conformation clusters, which denote the metastable states of this peptide. As can be observed on the reweighted maps, at least two different regions of energy minima can be identified in the case of water solvent, while only one prominent energy minimum is revealed in chloroform.



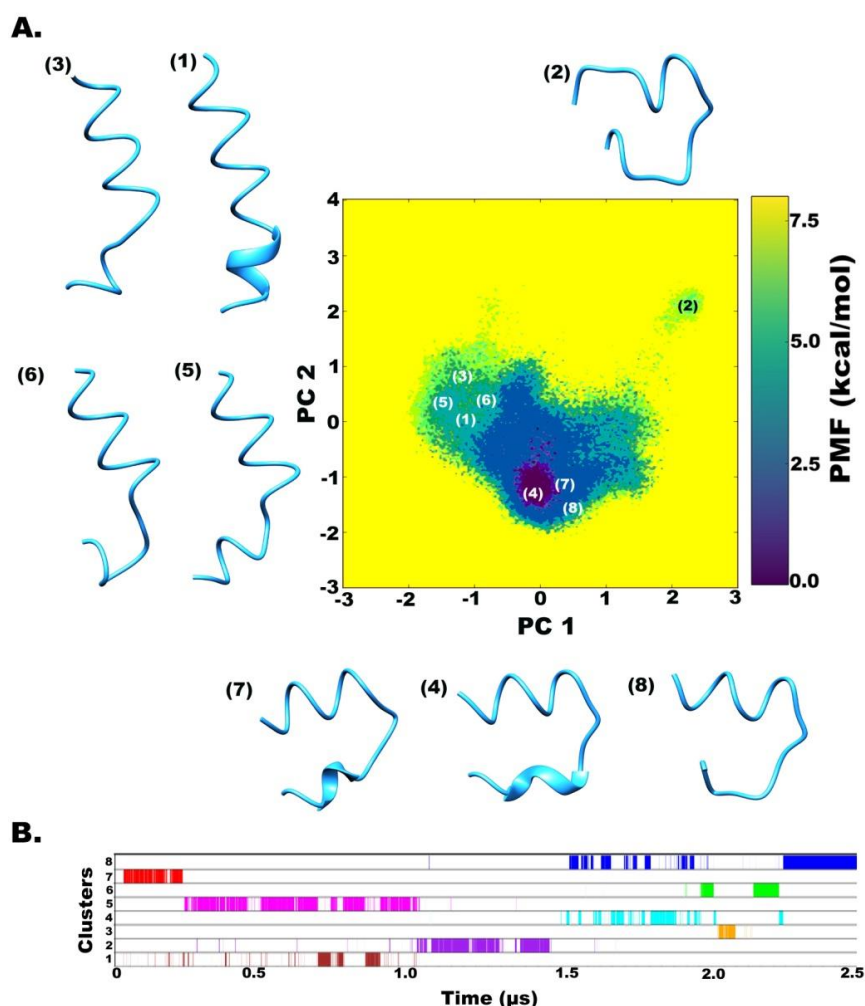
The full trajectory was clustered into 10 major representative groups (Figure 3a). The two prominent clusters, (1 and 2 based on 12% and 5% occurrence during the simulation, respectively) on the FEL map are separated by an energy barrier of at least 5 kcal mol<sup>-1</sup>. The representative structures show major differences in the C-terminal folding, which is a loose spiral in clusters 1 and 5, but a highly folded helix in cluster 2. A similar observation was made by calculating PMF (in kcal mol<sup>-1</sup>) for end-to-end distance of the peptide over whole trajectory where the distance from ~18 to 21 Å denotes a large energy minimum basin which means that a slightly bent peptide backbone is energetically favored (Supplementary file 1, Figure S1). Structures 7, 8, and 10 are closer to cluster 2 and represent a slightly bent helical folded structure unlike clusters 3, 4, 6, and 9, which are highly curved folded structures. Figure 3b shows a diagrammatic distribution of the inter-conversion between these clusters. The two main clusters denoting two deepest energy minima are revisited several times during the course of simulation but show an almost mutually exclusive occurrence with respect to each other. Cluster 5 occurs intermittently with cluster 1 while the rest of them show inter-conversion with Cluster 2. This suggests that the formation of an alpha-helical C-terminal fold is an uncommon energy barrier-crossing event and would not have been effectively sampled using short-timescale classical MD approaches. Another observation is the similarity of structures sampled in all independent simulations (clustered in the same group based on dihedral PCA), which signifies satisfactory sampling.



**Figure 3.** (A) Reweighted free energy landscape of the first two principal components calculated from dihedral angles, phi-psi, for better clustering based on internal motions. (B) Diagrammatic representation of cluster distribution along the simulation trajectory in water.

In the case of chloroform solvent (Figure 4a), eight different clusters were obtained. Interestingly, the major conformations obtained through the first three independent 500 ns long simulations showed a backbone curvature in the folded peptide that is observed as the region of the

deepest minimum denoted by Cluster 4 (13% occurrence throughout the combined trajectory). This is expected due to the presence of  $\gamma$ -turns that cause backbone reversal. It was determined that the Aib12  $\rightarrow$  Ala10 bond results in an energetically stable backbone reversal by  $i + 1^{\text{th}}$  residue Div11 which populates the  $\gamma'$ -region (inverse gamma turn or  $C_{7^{\text{eq}}}$  conformation) on the phi-psi distribution plot. An end-to-end distance of 14 Å (indicating backbone reversal) lies in an energy minimum, while the unfolded conformation lies in another minimum (Supplementary File 1, Figure S1). The last 1  $\mu\text{s}$  long simulation mainly resulted in a distinct linear backbone conformation represented by Cluster 1 (11% occurrence throughout the combined trajectory) that was observed for an insignificant period during previous simulations. The energy barrier between the two states (Cluster 4  $\rightarrow$  Cluster 1) is  $\sim 4 \text{ kcal mol}^{-1}$ , i.e., the conversion from a highly bent to a linear conformation is a rare event and requires a longer continuous simulation to be achieved with aggressive dihedral boost parameters. It is also supported by the end-to-end distance value of 27 Å (linear backbone) which is achieved after crossing a barrier of  $5 \text{ kcal mol}^{-1}$  (Supplementary File 1, Figure S1). There is a  $6 \text{ kcal mol}^{-1}$  energy barrier between the unfolded conformation represented by cluster 2 and folded conformation of cluster 4, which was achieved immediately after 200 ns. Clusters 4, 7, and 8, on the other hand, show interconversion quite a few times between 700 to 900 ns with respect to each other (Figure 4b).

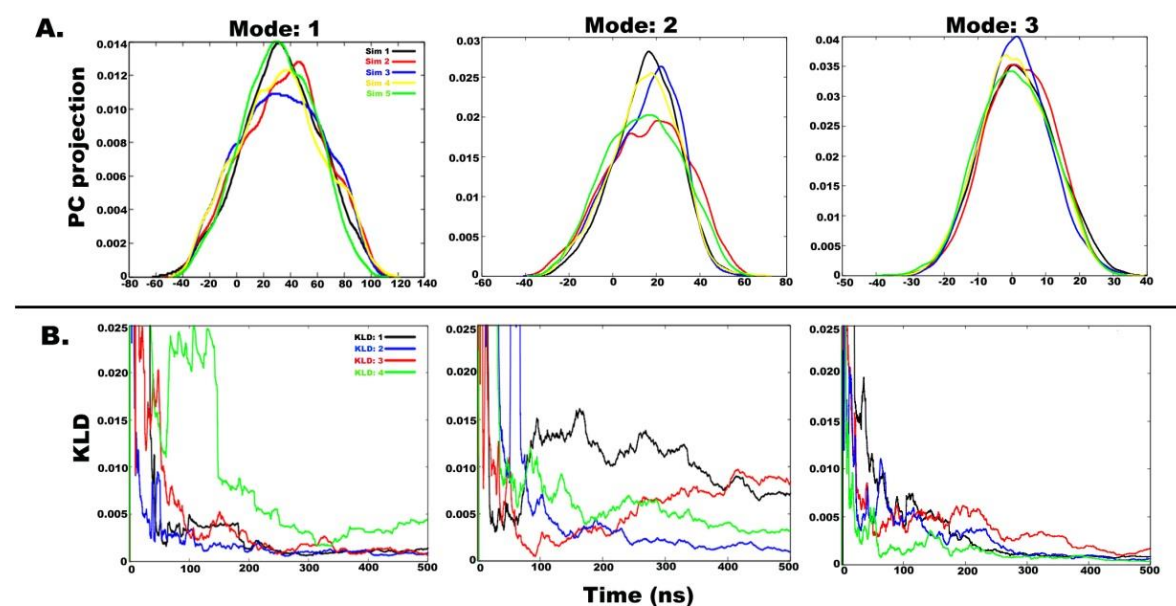


**Figure 4.** (A) Reweighted free energy landscape of the first two principal components calculated from dihedral angles, phi-psi, for better clustering based on internal motions. (B) Diagrammatic representation of cluster distribution along the simulation trajectory in chloroform.

It should be carefully considered that the energetically stable state of TPN XIIc in chloroform is a highly bent shape that is not ideal for membrane-spanning peptides. The ideal linear conformation was only obtained after crossing the energy barrier that remains stable afterwards.

### 2.3. Understanding the Dynamics of TPN XIIc through Cartesian PCA

In the previous section, we presented peptide dynamics based on free energy landscapes obtained through dihedral PCA, which only accounts for the internal motions. In this section, we shall uncover the overall motions of the peptide to attain various metastable states. The combined trajectory (2.5  $\mu$ s) was divided into five parts of equal time length, i.e.,  $\sim$ 500 ns. This means that each individual trajectory and the last 1  $\mu$ s long simulation divided in two equal parts were considered. Figure 5a shows the histograms for projection of coordinates along the first three eigenvectors (i.e., the first three modes with highest eigenvalues). The trajectories are named as “Sim1” to “Sim5” and the extent of PC overlap signifies convergence between independent runs. The degree of overlap suggests that the independent simulations sampled similar conformational space [36]. The essential PCs 1, 2, and 3 obtained from five separate trajectories accounting for 24%, 15%, and 11% of overall motions, respectively, show considerable overlap. A considerable overlap between first PC histograms (Mode 1) shows very similar motions of the peptide during the first and last simulations while the second PC histograms (Mode 2) show major similarity between the second and last simulations. The third PC histogram (Mode 3) is similar for all five independent trajectories.



**Figure 5.** (A) Histograms of projection of principal components 1, 2, and 3 for all four simulations in water, where the last 1  $\mu$ s long simulation is treated in 2 parts. Histograms were calculated using a Gaussian kernel density estimator. (B) A measure of overlap between histograms from independent simulations calculated using the Kullback-Leibler divergence method. The slope values lying below 0.025 indicate convergence between two independent runs.

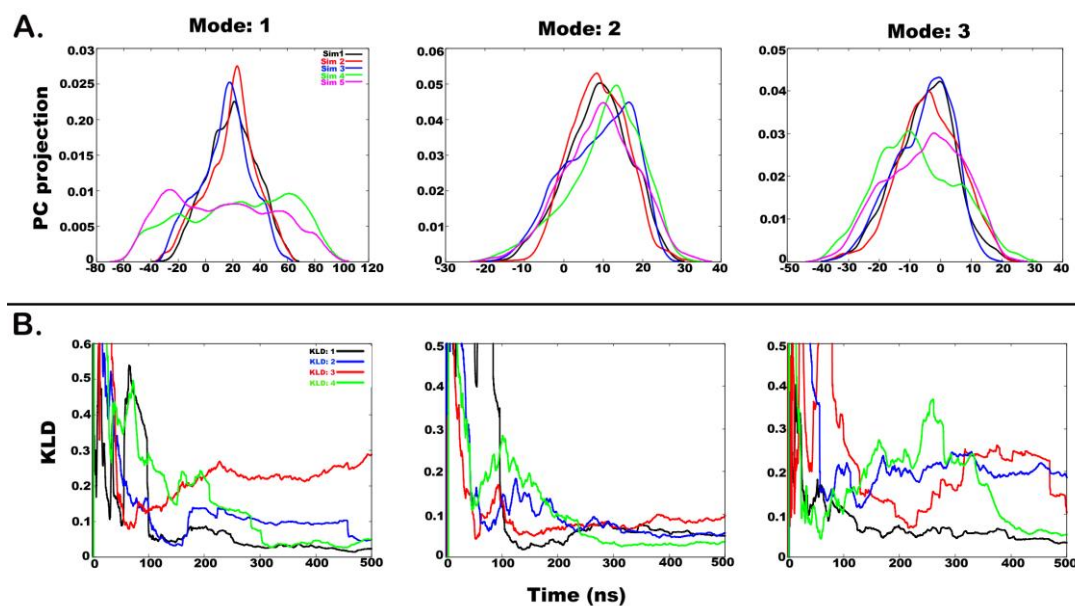
The quantification of convergence and statistical significance in macromolecular MD simulations often presents a challenging task, and more so, in case of peptide folding calculations. A quantitative measure of extent of overlap between any probability distribution is the Kullback-Leibler divergence (KLD) method that can indicate satisfactory sampling [25]. By measuring the overlap of PC histograms as a function of simulation time, we can assess the convergence of dynamic properties of simulations. The principal components give an idea of conformational states that have been visited during the simulation. Two independent simulations started from different configurations should eventually begin to sample the same conformations

[43]. The extent of overlap between such simulations will indicate that satisfactory sampling has been attained. The KLD method can be applied to measure the extent of overlap of any quantity (e.g. RMSD), but we have chosen PC histograms as they provide a clear idea of the dynamics of the peptide. A rapidly decreasing slope of KLD as a function of time indicates convergence between two independent simulations. As per previous studies, we have selected the KLD value of 0.025 as the cutoff for convergence. When the KLD slope hits below 0.025, the two simulations will have converged.

Figure 5b shows KLD as a function of time between subsequent histograms from five different simulations for PCs 1, 2, and 3 (accounting for 50% of total motion). “KLD:1” denotes divergence between Sim1 and Sim2, “KLD:2” between Sim2 and Sim3, “KLD:3” between Sim3 & Sim4 and “KLD:4” between Sim4 and Sim5. It is evident that the slope of KLD:1, KLD:2, and KLD:3 values for PC 1 does not change significantly after 200 ns, therefore, signifying convergence. KLD:4, on the other hand, shows a drop to 0.025 at around 300 ns before rising again at 350 ns which means that a wider conformational space was sampled during 1  $\mu$ s long simulation. The KLD values for PC 2 are highly divergent and show that convergence was not obtained for this mode of motion while KLD for PC 3 shows high convergence. This analysis also indicates that the minimum sampling time required for TPN XIIc peptide in water is 500 ns using accelerated MD. It can be safely stated that all major metastable states of TPN XIIc in aqueous medium have been sampled during independent simulations.

A similar analysis for chloroform solvent divulged some interesting results. Figure 6a shows the histograms for projection of coordinates along the first three eigenvectors (i.e., the first three modes with highest eigenvalues). The overlap of first PC histograms (mode 1 accounting for 53% of overall motions) shows similarity between the first three independent simulations, while the fourth 1  $\mu$ s long simulation shows a drastically different mode of motion. However, such a stark difference is not observed for the second PC (mode 2).

Figure 6b shows KLD as a function of time between subsequent histograms from five different simulations for PCs 1, 2, and 3 (accounting for 68% of total motion). It can be noted that the slope of KLD:1 and KLD:4 values for PC 1 does not change significantly after 300 ns, therefore, signifying convergence between Sim1 and Sim2 and Sim4 and Sim5, respectively. KLD:2 values for PC1 reach a value of 0.05 at around 150 ns before increasing again. This means that Sim2 and Sim3 initially sampled similar structures before diverging on separate conformational paths. KLD:3 values between Sim3 and Sim4 remain divergent, which is expected due to very different conformational spaces covered during the previous 500 ns long simulations and the latest 1  $\mu$ s long simulation. As mentioned before, when KLD slope rises beyond 0.025, it indicates that two independent simulations diverged in the conformational space. The occurrence of a linear backbone structure in the last 1  $\mu$ s simulation in contrast to a highly bent structure in previous simulations can be believed to be the reason for this difference. As the peptide undergoes significant motion around the Div11-Aib12-Pro13 hinge region to attain a straight backbone, a new metastable state is achieved, which is comparable to structures obtained from water simulation. Therefore, a combination of all simulations has resulted in a nearly complete conformational landscape for the peptide of interest in chloroform solvent. KLD values for PC2 show high convergence, while for PC3, the high KLD:4 values for the two parts of the 1  $\mu$ s long simulation shows that highly divergent internal motions have occurred in the peptide.



**Figure 6.** (A) Histograms of projection of principal component 1, 2, and 3 from all four simulations in chloroform where the last 1  $\mu$ s long simulation is treated in 2 parts. Histograms were calculated using a Gaussian kernel density estimator. (B) A measure of overlap between histograms from independent simulations calculated using Kullback-Leibler divergence method. The slope values lying below 0.025 indicate convergence between two independent runs.

In some instances, the KLD value rises again after hitting the cutoff once, which indicates that the two simulations diverged and sampled different conformational spaces as the simulation progressed. In the case of chloroform simulation, the last 1  $\mu$ s long simulation resulted in a state not observed in previous simulations. This is a high energy metastable state that could only be observed due to the use of aggressive boost parameters during accelerated MD. That means that the first three simulations did not sample the entire conformational space (shown by increasing slope of corresponding KLD: 3) and hence could not be considered converged. This puts a question mark on the use of KLD as a method of assessing convergence as if the fourth simulation had not been carried out, the previous three simulations would indicate perfect convergence and that is clearly not true. Our rationale lies in the fact that chloroform is a very restrictive solvent for peptide folding which is clear from our results. The conformational landscape covered in the fourth simulation is a higher energy state (Supplementary Figure S1, linear peptide) that could not have been achieved during shorter simulations. This shows that 500 ns long runs are not adequate to attain the higher energy state in chloroform. The peptide seems to be stuck in a single energy state in chloroform, while it is highly dynamic in water and jumps through various metastable states with relative ease. Irrespective of the starting configuration, all simulations must, at some point, begin to sample the same space which could not be observed in the case of chloroform. However, the combined trajectory of 2.5  $\mu$ s surely indicates adequate sampling. This is a drawback of using such a statistical method for addressing convergence, especially when enhanced sampling methods have been used in solvents other than water. Nevertheless, it is more reliable than previous methods of considering convergence criteria as RMSD.

Along with the discussion on convergence of PC space that conveys the dynamics of peptide folding, it was interesting to consider structural convergence as well [44]. Based on Good-Turing formalism [45–47] applied on the root-mean-square-deviation values it could be deduced, that slightly higher structural convergence is achieved in chloroform than in water (Supplementary file 1, Figure S2). This indicates that the folding of the TPN XIIc peptide is very dynamic in an aqueous environment and can probably result in many unfolded or intermediate states, even though all energetically stable states have already been achieved. A biological membrane mimicking



environment, like chloroform, has a stabilizing effect on peptide dynamics and requires longer time scales to be able to sample higher energy states. A similar study by Levy et al. [48], reported differences in the folding pattern of Ala<sub>12</sub> peptide in vacuum and water. The peptide was able to achieve its native state in vacuum, while it was destabilized in water and achieved a  $\beta$ -state as native. Additionally, a comparison between different folding and unfolding time scales for Ala<sub>12</sub> in hydrophilic and hydrophobic solvents was attributed to the higher entropy of native state in water.

For most proteins, their interactions with organic solvents depend on hydrophobic core destabilization, side-chain exposure, and favoring of hydrophilic protein surface. The flexibility of a peptide and its associated entropy are products of various intermolecular interactions with the solvent. For example, the interaction of water molecules with the peptide polar groups can lead to the destabilization of a strict helical structure and render it in coil-like formation to make the polar groups more accessible. Clearly, any peptide, including TPN XIIc shows higher flexibility in water than its organic counterpart, which was also evident by comparing the extent of conformational landscape covered. A similar observation was reported by Levy et al. [48], who explained that random coils are more prevalent in water as the loss of intra-molecular interaction is compensated by intermolecular interactions with water molecules. On the contrary, while they reported two distinct energy wells for Ala<sub>12</sub> falling in  $\alpha$ -helical and  $\beta$ -sheet regions, we did not observe such distinctive characterization in water or chloroform for TPN XIIc. This indicates that TPN XIIc possesses a stronger propensity for the helical conformation, and due to its amphipathic nature, it can access similar secondary structural states in hydrophilic and hydrophobic environments. The largest difference observed upon change of solvent is the overall bent in the peptide linearity. The peptide seems to undertake its biggest motion around the Div11-Aib12-Pro13 region, rendering either a hairpin like bend or a linear shape. In relation to that, it was noted that the highly bent, hairpin-like shape observed mainly in chloroform occurred due to burying of hydrophilic residues like Gln6 and Gln17 inside the curve, while the hydrophobic residues like Val9, Ala10, Pro13, Leu14, and Val16 are outside, facing the hydrophobic solvent. The same hydrophobic residues shall form the interior of an ion-channel with other peptide units. Thus, it can be stated that due to the effect of a hydrophobic solvent, all hydrophobic residues are aligned to interact with their surroundings. The two energetically stable states obtained in water are relatively linear, where the masking of hydrophobic residues is not possible, but inter-molecular interaction between water and polar residue Gln17 in Cluster 1 may explain the loss of helical structure and its replacement with a loose coil. Probably that is the reason of higher entropy and flexibility in the case of the TPN XIIc peptide. A detailed study of peptide folding characteristics and the corresponding ensemble identification with the help of enhanced molecular dynamics methods has been proven crucial in elucidating the mode of action of their bioactivities. Characterization of bioactive peptides is necessary to identify structural features that are crucial for bioactivity and utilize this information to synthetically design novel therapeutics.

### 3. Materials and Methods

#### 3.1. Sequence Selection

A recent study identified and published a new group of 18-residue peptaibols produced by *T. pleuroti*, named as tripleurins [19]. Out of 24 reported tripleurin sequences, three compounds (Tripleurins VI, VIIIb, and XIIc) were present in the highest area percentage with 7.9%, 10.0%, and 12.0% on the mass spectrometric chromatogram extracted at  $m/z$  of sodiated molecular ions  $[M + Na]^+$ , respectively. Tripleurin XIIc, an 18-residue long sequence with five ambiguous Val/Div and one Leu/Ile position, was selected. Its primary structure is:

AcAib<sup>1</sup>-Ser<sup>2</sup>-Ala<sup>3</sup>-Aib<sup>4</sup>-Vxx<sup>5</sup>-Gln<sup>6</sup>-Vxx<sup>7</sup>-Aib<sup>8</sup>-Vxx<sup>9</sup>-Ala<sup>10</sup>-Vxx<sup>11</sup>-Aib<sup>12</sup>-Pro<sup>13</sup>-Lxx<sup>14</sup>-Aib<sup>15</sup>-Vxx<sup>16</sup>-Gln<sup>17</sup>-Pheol<sup>18</sup>

The ambiguous residue positions were predicted based on the sequence of non-ribosomal peptide synthetase (NRPS) proteins using the antiSMASH database server [49]. Positions Vxx5, Vxx7, and Vxx11 were predicted to be D-isovalines, Vxx9 and Vxx16 were predicted to be valines,

while Lxx14 was predicted as leucine. This prediction adds a high number of D-amino acid residues, which may have a strong influence on the overall structure and bioactivity.

### 3.2. Partial charge Calculation and Force Field Library Generation for Non-Standard Residues

As famously known, fungal peptaibols are characterized by their unusual amino acid content. In the selected sequence, Aib, Div, and Pheol are non-standard residues. A graphical representation of their 2D structures is provided in Figure 7. Both Aib and Div are derivatives of the alanine, in which one methyl group is present as the side-chain. An additional methyl group ( $-\text{CH}_3$ ) is attached to the  $\text{C}_\alpha$  carbon atom as the side-chain in Aib, while an ethyl group ( $-\text{CH}_2\text{CH}_3$ ) is attached in the D-Iva residue. The R.E.D server was used for calculation of their partial charges and creating force field libraries [50]. R.E.D stands for RESP ESP charge derive [51]. RESP (restrained electrostatic potential) was used to calculate the charges with a HF/6-31G(d) basis set and Gaussian09 [52] as quantum mechanical program interface. The charges for Aib and Div were calculated along with a few other standard amino acids. The charges calculated for standard residues were used to confirm with existing “leap” libraries. For each residue, two conformations, i.e., alpha helix ( $\Phi = -63.8$ ,  $\Psi = -38.3$ ) and beta sheet or  $\text{C}_5$  ( $\Phi = -157.2$ ,  $\Psi = 161.9$ ) were used. These were modified and generated using Avogadro [53] based on the strategy described by Cieplak [54]. The terminal residue phenylalaninol (Pheol) was also parameterized using a slightly different strategy, where two molecules, ethyl alcohol and phenylalanine were used to form the Pheol unit. The results include calculated charges and a script to make force field libraries for these forces. The TPN XIIc sequence was built by supplying residue units from scratch using “tleap”. The parameters for each non-standard residue are provided in Supplementary file 1.

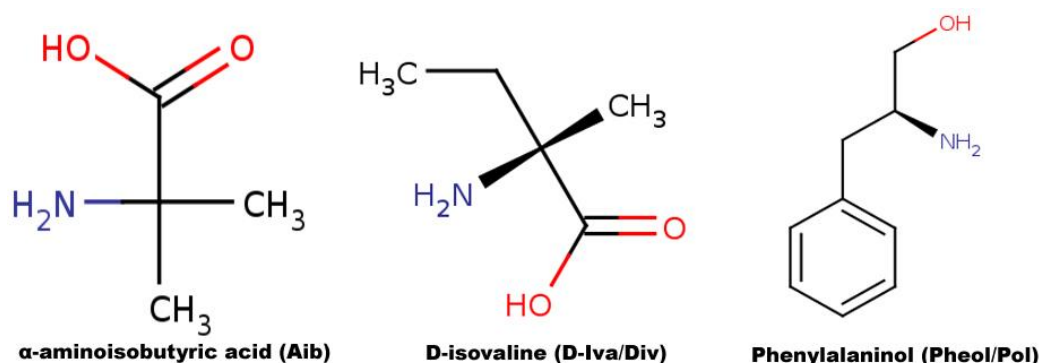


Figure 7. Graphical representation of 2D molecular structures of Aib, Div and Pheol/Pol.

### 3.3. Accelerated Molecular Dynamics Simulations of TPN XIIc

The unfolded peptide conformation prepared by using the ‘tleap’ module of AmberTools18 [55] was solvated in water and chloroform solvents. For water solvent, TIP3P water model was loaded from the AmberTools18 solvents library and 2657 water molecules were added with a box size of  $44.34 \times 59.11 \times 42.85 \text{ \AA}$  and a volume of  $112332.666 \text{ \AA}^3$ . The Amber solvent library was also loaded for chloroform, which added 1021  $\text{CHCl}_3$  molecules with a box size of  $49.608 \times 59.514 \times 67.986 \text{ \AA}$  and a volume of  $200719.862 \text{ \AA}^3$  to the system. The Amberff14SB force field was used to prepare and minimize all systems.

As classical molecular dynamics (cMD) offers limited utility in terms of shorter time scales, a relatively new approach named accelerated molecular dynamics (aMD) was adopted for this study to enhance sampling. It is a bias potential function introduced by Hamelberg et al. [21,22], which can be used to make the simulation “jump over” high energy barriers and to sample rare events. A detailed discussion is provided elsewhere [56].

All systems were prepared for aMD in six consecutive steps, i.e., (a) minimization (conjugate gradient followed by steepest descent method) of solvent for 20,000 cycles while keeping peptide under restraint, (b) water movement at 300 K under isothermal and isobaric (NTP) conditions while

keeping peptide under restraint, (c) minimization of the whole system for 20,000 cycles, (d) heating from 0 K to 300 K under isothermal & isovolumetric (NVT) conditions while keeping peptide under restraints, (e) relax the system at 300 K for 0.5 ns while keeping peptide heavy atoms under restraint, and (f) relax system at 300 K under NTP conditions for 5 ns. The temperature scaling was carried out using Langevin thermostat while the pressure was regulated using the default Berendsen barostat for all corresponding calculations. SHAKE bond length constraints were applied on all bonds involving hydrogen. A short classical MD run to obtain average dihedral and potential energies (in kcal mol<sup>-1</sup>) was also carried out for 100 ns at 300 K temperature and periodic boundary condition was used with constant pressure using Berendsen barostat in each case.

The two systems, unfolded peptide in water and chloroform, were simulated using aMD for 500 ns with three independent starting structures (500 ns × 3) along with a 1000 ns long simulation making a combined simulation time of 2.5 μs. All simulations were carried out at 300 K temperature, 2 fs time step, and energies and boost information was written down at every 1000 steps. The electrostatic interactions were calculated using PME (particle mesh Ewald summation) [57] and long-range interactions were also calculated with a cutoff of 10.0. The temperature scaling was carried out using Langevin thermostat without pressure scaling during aMD. The SHAKE algorithm was applied on all bonds involving hydrogen. The GPU machines available through the NIF clusters of Hungary were utilized for all aMD simulations. All simulations were carried out using the *pmemd.cuda* implementation of Amber14, which is also available at the cluster.

aMD can be carried out using three criteria (i) independently boosting the torsional terms of the potential (*iamd* = 2) or (ii) the whole potential at once (*iamd* = 1), and (iii) to boost the whole potential with an extra boost to torsions (*iamd* = 3). The third criterion seemed to be an appropriate choice, as dihedral-only aMD boost is known to enhance the convergence of the underlying free energy landscape by 5-fold in comparison to classical MD, but the dual boost option provides a better reweighting distribution [56].

The extra parameters  $E_{\text{dihed}}$ ,  $\alpha_{\text{dihed}}$ ,  $E_{\text{total}}$ , and  $\alpha_{\text{total}}$  were calculated, as required in Equation 1:

$$E_{\text{dihed}} = V_{\text{avg\_dihed}} + a_1 \times N_{\text{res}}, \alpha_{\text{dihed}} = a_2 \times N_{\text{res}}/5 \quad (1)$$

$$E_{\text{total}} = V_{\text{avg\_total}} + b_1 \times N_{\text{atoms}}, \alpha_{\text{total}} = b_2 \times N_{\text{atoms}}$$

where  $N_{\text{res}}$  is the number of peptide residues (19, with an addition of Ace at the *N*-terminal),  $N_{\text{atoms}}$  is the total number of atoms in the system which were 8243 and 7462 in the case of water and chloroform solvent systems, respectively.  $V_{\text{avg\_dihed}}$  and  $V_{\text{avg\_total}}$  are the average dihedral and total potential energies obtained from the 100 ns long cMD run in each solvent. The various parameters used for all aMD simulations have been summarized in Table 4.

**Table 4.** Summary of various accelerated molecular dynamics parameters.

Simulations	Time (ns)	Boost option	Vavg_dihed (kcal mol <sup>-1</sup> )	Vavg_total (kcal mol <sup>-1</sup> )
In water	2500 (500 × 3 + 1000 ns)	iamd = 3	210	-25429
In chloroform	2500 (500 × 3 + 1000 ns)	iamd = 2	206	-7535

In the water simulation, application of smaller boost coefficients  $b_1$ ,  $b_2 = 0.16$  results in very small boost (~5 kcal mol<sup>-1</sup>) and the folding event could not be observed for 500 ns. Therefore, comparably aggressive boost parameters, i.e.,  $b_1$ ,  $b_2 = 0.30$  and  $b_1$ ,  $b_2 = 0.20$  were applied to speed up the folding process. It shall also be noted that the total potential energies obtained for the chloroform system was much higher (~ -7335 kcal mol<sup>-1</sup>) than in water systems (~ -25000 kcal mol<sup>-1</sup>). The chloroform solvent-based simulation showed unexpected behavior when the boost was applied to the potential energy (*iamd* = 3 or *iamd* = 1). Even a small potential boost defined by coefficient values of  $b_1$ ,  $b_2 = 0.01$  and 0.08 (according to Equation 1) resulted in extremely high energy values and the calculation could not proceed. Similar results were obtained when a bigger system

comprised of more chloroform molecules was tested. It was suspected that the chloroform system already possesses a very high potential energy and even a small boost to the potential increase in the system energy to enormous levels. Subsequently, it was decided to apply only torsional boost (iamd =2) to this system. To the best of our knowledge, this is the first study on the application of accelerated MD on a peptide solvated in the chloroform solvent. Different values of coefficients a1, a2 and b1, b2 were tested and applied, as summarized in Table 5.

**Table 5.** Summary of coefficient a1, a2 and b1, b2 applied to consecutive simulations and the resulting average boost energy.

Water Simulation			Chloroform Simulation		
a1, a2	b1, b2	Avg. boost (kcal mol <sup>-1</sup> )	a1, a2	b1, b2	Avg. boost (kcal mol <sup>-1</sup> )
4.0	0.16	5	4.0	-----	6.5
3.5	0.30	45	4.5	-----	10
3.5	0.20	15	6.0	-----	30

### 3.4. Tools for Simulation Analysis

The calculation of dihedral angle populations and hydrogen bond calculations were carried out using the *cpptraj* [58] module. The dihedral based PCA and Cartesian coordinate PCA were also done using the *cpptraj* module. For dihedral PCA, the phi psi torsion angles are calculated for all residues and the covariance matrix is calculated. The eigenvectors were calculated based on the covariance matrix. The first two principal components are reweighted by the Maclaurin series expansion method. The Cartesian PCA was also carried out with RMS-fit (root-mean-square) to first frame to remove global translation/rotation followed by RMS-fit to average structure calculated in the previous step. The covariance matrix was calculated and diagonalized. The common eigenvectors were calculated and the individual trajectories obtained from water and chloroform were projected on PCA space. The principal component (PC) histograms were obtained for each of the four simulations (replicas) carried out in both solvents. As the last 1  $\mu$ s long simulation is being treated in two parts, we considered PC histograms from five distributions, i.e., Sim1 to Sim5. The extent of overlap between such replicas in the same solvent should provide an idea of convergence. The Kullback–Leibler divergence (KLD) [25] between the principal component histograms from two replicas (e.g., between Sim1 and Sim2, Sim2 and Sim3 and so on in each solvent) over time was calculated. The time-dependent KLD is calculated as:

$$\text{KLD}(t) = \sum_i P(t, i) \ln \left( \frac{P(t, i)}{Q(t, i)} \right) \quad (2)$$

where  $P(t, i)$  and  $Q(t, i)$  represent different probability distributions normalized to 1.0, with  $i$  representing a histogram bin index and  $t$  representing the time at which the histogram is being constructed (i.e., all data from time 0 to  $t$ ). Histograms were constructed using a Gaussian kernel density estimator to exclude the bins with no population. It has been used in previous studies as a measure of convergence [43,44]. *grcarma* [59,60] was used to generate the highest populated clusters using the top three principal components (PC) and write their representative structures in pdb format files. An important aspect of an aMD calculation is to reweight the distribution to remove the effect of boost applied to the system and to recover the original free energy landscapes. To recover this distribution, 'amd.log' is generated during the run that contains information about the extra boost added to each snapshot of the trajectory. Theoretically, aMD simulations can be reweighted by the Boltzmann factors of the corresponding boost potential (i.e.,  $e^{\Delta V/k_B T}$ ) and averaged over each bin of selected reaction coordinate(s) to obtain the canonical ensemble, a technique named as exponential average which suffers from large statistical noise especially if higher boost was applied. To overcome this problem, a method named Maclaurin series expansion was used, which

approximates the exponential Boltzmann factor and reduces energetic noise considerably [24]. A third algorithm named cumulant expansion (first, second, and third order) was reported to be the most accurate by Miao et al. [61]. While this study reported that cumulant expansion to the second order is most accurate when the boost potential follows near-Gaussian distribution, Jing et al. [62] argued that boost potential should exactly be Gaussian or else this method may show considerable deviation. Maclaurin series expansion up to the 10<sup>th</sup> order was found to be the most accurate reweighting procedure for all simulations.

#### 4. Conclusions

In this study, we have characterized an 18-mer peptaibol named TPN XIIc for its structural and dynamical properties. The peptide shows different folding behavior in water and chloroform solvents. The two largest low-energy clusters obtained from the water system show continuous beta turns, which form a beta-bend ribbon spiral at the N-terminal and a 3<sub>10</sub>/α-helical continuous structure at the C-terminal. In chloroform, a higher propensity to form γ-turn at the central Aib12 → Ala10 bond is observed which renders the backbone in a bent state comparable to a beta hairpin. A longer simulation revealed a linear backbone structure that could be attained after crossing a 4 kcal mol<sup>-1</sup> energy barrier. While addressing the convergence of these two simulations, it was observed that five independent simulations have sampled very similar conformations in water and hence show convergence. While in chloroform, the last 1 μs long simulation samples a very different conformational space not observed previously. The peptide shows a highly dynamic behavior in water compared to chloroform and structural convergence is achieved earlier in the latter. The structural information of these short peptides is crucial to mark functionally important segments. This study is the first step to link the structure to the biological function of tripleurins and to understand their folding mechanisms.

**Supplementary Materials:** The following are available online at [www.mdpi.com/1420-3049/24/2/358/s1](http://www.mdpi.com/1420-3049/24/2/358/s1): which includes detailed explanation for Good-Turing analysis and data of parameterized non-standard amino acid residues, Aib, Div, and Pheol. Figure S1: Potential-of-mean-force values calculated in kcal mol<sup>-1</sup> as a function of end-to-end distance of TPN XIIc calculated for each step in the trajectory. Figure S2: Estimation of extent of sampling based on Good-Turing formalism (black) water and (red) chloroform.

**Author Contributions:** Conceptualization, C.T. L.K. and F.Ö.; Methodology, C.T. T.M. and F.Ö.; Software, C.T.; Formal Analysis, C.T.; Investigation, C.T. T.M. and F.Ö.; Data Curation, C.T. and T.M.; Writing-Original Draft Preparation, C.T.; Writing-Review & Editing, A.S. C.V. L.K. and F.Ö.; Visualization, C.T.; Supervision, L.K. and F.Ö.; Funding Acquisition, C.V. and L.K.

**Funding:** This work was supported by grant [NKFI K-116475] (National Research, Development and Innovation Office) as well as by the Hungarian Government and the European Union within the frames of the Széchenyi 2020 Programme [grant GINOP-2.2.1-15-2016-00006]. LK is grantee of the János Bolyai Research Scholarship (Hungarian Academy of Sciences) and the Bolyai+ Scholarship [ÚNKP-18-4-SZTE-93].

**Acknowledgments:** The authors are thankful to National Information Infrastructure Development (NIIF) Programme in Hungary through which all state-of-the-art GPU machines were made available for carrying out all simulations.

**Conflicts of Interest:** The authors declare no conflict of interest.

#### References

1. Mohamed-Benkada, M.; Francois, P.Y.; Verite, P.; Pagniez, F.; Caroff, N.; Ruiz, N. Identification and biological activities of long-chain peptaibols produced by a marine-derived strain of *Trichoderma longibrachiatum*. *Chem. Biodivers.* **2016**, *13*, 521–530.
2. Berg, A.; Grigoriev, P.A.; Degenkolb, T.; Neuhof, T.; Hartl, A.; Schlegel, B.; Grafe, U. Isolation, structure elucidation and biological activities of trichofumins A, B, C and D, new 11 and 13mer peptaibols from *Trichoderma* sp. HKI 0276W. *J. Pept. Sci.* **2003**, *9*, 810–816.



3. Leitgeb, B.; Szekeres, A.; Manczinger, L.; Vagvolgyi, C.; Kredics, L. The history of alamethicin: A review of the most extensively studied peptaibol. *Chem. Biodivers.* **2007**, *4*, 1027–1051.
4. Yun, B.-S.; Yoo, I.-D.; Kim, Y.H.; Kim, Y.-S.; Lee, S.-J.; Kim, K.-S.; Yeo, W.-H. Peptaivirins A and B, two new antiviral peptaibols against TMV infection. *Tetrahedron Lett.* **2000**, *41*, 1429–1431.
5. Schiell, M.; Hofmann, J.; Kurz, M.; Schmidt, F.R.; Vertesy, L.; Vogel, M.; Wink, J.; Seibert, G. Cephaibols, new peptaibol antibiotics with anthelmintic properties from *Acremonium tubakii* DSM 12774. *J. Antibiot.* **2001**, *54*, 220–233.
6. Viterbo, A.; Wiest, A.; Brotman, Y.; Chet, I.; Kenerley, C. The 18mer peptaibols from *Trichoderma virens* elicit plant defence responses. *Mol. Plant Pathol.* **2007**, *8*, 737–746.
7. Engelberth, J.; Koch, T.; Kuhnemann, F.; Boland, W. Channel-Forming Peptaibols Are Potent Elicitors of Plant Secondary Metabolism and Tendril Coiling. *Angew. Chem. Int. Ed. Engl.* **2000**, *39*, 1860–1862.
8. Miller, B.R.; Gulick, A.M. Structural biology of nonribosomal peptide synthetases. In Evans B. (eds.) *Nonribosomal Peptide and Polyketide Biosynthesis*, Methods in Molecular Biology, vol 1401. Humana Press: New York, 2016; pp 3–29.
9. Raap, J.; Erkelens, K.; Ogrel, A.; Skladnev, D.A.; Brückner, H. Fungal biosynthesis of non-ribosomal peptide antibiotics and  $\alpha$ ,  $\alpha$ -dialkylated amino acid constituents. *J. Pept. Sci.* **2005**, *11*, 331–338.
10. Shenkarev, Z.O.; Balashova, T.A.; Efremov, R.G.; Yakimenko, Z.A.; Ovchinnikova, T.V.; Raap, J.; Arseniev, A.S. Spatial structure of zervamicin IIB bound to DPC micelles: Implications for voltage-gating. *Biophys J* **2002**, *82*, 762–771.
11. Karle, I.L.; Flippen-Anderson, J.L.; Agarwalla, S.; Balaram, P. Crystal structure of [Leu1]zervamicin, a membrane ion-channel peptide: Implications for gating mechanisms. *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 5307–5311.
12. Fox, R.O. Jr.; Richards, F.M. A voltage-gated ion channel model inferred from the crystal structure of alamethicin at 1.5-Å resolution. *Nature* **1982**, *300*, 325–330.
13. Balashova, T.A.; Shenkarev, Z.O.; Tagaev, A.A.; Ovchinnikova, T.V.; Raap, J.; Arseniev, A.S. NMR structure of the channel-former zervamicin IIB in isotropic solvents. *FEBS Lett* **2000**, *466*, 333–336.
14. Condamine, E.; Rebuffat, S.; Prigent, Y.; Ségalas, I.; Bodo, B.; Davoust, D. Three-dimensional structure of the ion-channel forming peptide trichorzianin TA VII bound to sodium dodecyl sulfate micelles. *Biopolymers* **1998**, *46*, 75–88.
15. Snook, C.F.; Woolley, G.A.; Oliva, G.; Pattabhi, V.; Wood, S.F.; Blundell, T.L.; Wallace, B.A. The structure and function of antiameobin I, a proline-rich membrane-active polypeptide. *Structure* **1998**, *6*, 783–792.
16. Kronen, M.; Gorls, H.; Nguyen, H.H.; Reissmann, S.; Bohl, M.; Suhnel, J.; Grafe, U. Crystal structure and conformational analysis of ampullosporin A. *J. Pept. Sci.* **2003**, *9*, 729–744.
17. Nagao, T.; Mishima, D.; Javkhlantugs, N.; Wang, J.; Ishioka, D.; Yokota, K.; Norisada, K.; Kawamura, I.; Ueda, K.; Naito, A. Structure and orientation of antibiotic peptide alamethicin in phospholipid bilayers as revealed by chemical shift oscillation analysis of solid state nuclear magnetic resonance and molecular dynamics simulation. *Biochim. Biophys. Acta* **2015**, *1848*, 2789–2798.
18. Hall, S.E.; Roberts, K.; Vaidehi, N. Position of helical kinks in membrane protein crystal structures and the accuracy of computational prediction. *J. Mol. Graph. Model.* **2009**, *27*, 944–950.
19. Marik, T.; Urban, P.; Tyagi, C.; Szekeres, A.; Leitgeb, B.; Vagvolgyi, M.; Manczinger, L.; Druzhinina, I.S.; Vagvolgyi, C.; Kredics, L. Diversity profile and dynamics of peptaibols produced by green mould trichoderma species in interactions with their hosts *Agaricus bisporus* and *Pleurotus ostreatus*. *Chem. Biodivers.* **2017**, *14*, e1700033.
20. Hollingsworth, S.A.; Karplus, P.A. A fresh look at the Ramachandran plot and the occurrence of standard structures in proteins. *Biomol. Concepts* **2010**, *1*, 271–283.
21. Hamelberg, D.; Mongan, J.; McCammon, J.A. Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules. *J. Chem. Phys.* **2004**, *120*, 11919–11929.
22. Hamelberg, D.; de Oliveira, C.A.; McCammon, J.A. Sampling of slow diffusive conformational transitions with accelerated molecular dynamics. *J. Chem. Phys.* **2007**, *127*, 155102.
23. Pierce, L.C.; Salomon-Ferrer, R.; Augusto, F. d. O. C.; McCammon, J.A.; Walker, R.C. Routine access to millisecond time scale events with accelerated molecular dynamics. *J. Chem. Theory Comput.* **2012**, *8*, 2997–3002.
24. Miao, Y.; Feixas, F.; Eun, C.; McCammon, J.A. Accelerated molecular dynamics simulations of protein folding. *J. Comput. Chem.* **2015**, *36*, 1536–1549.
25. Kullback, S.; Leibler, R.A. On information and sufficiency. *Ann. Math. Stat.* **1951**, *22*, 79–86.
26. Banerjee, R.; Basu, G. A short Aib/Ala-based peptide helix is as stable as an Ala-based peptide helix double its length. *ChemBioChem* **2002**, *3*, 1263–1266.

27. Banerjee, R.; Basu, G.; Roy, S.; Chène, P. Aib-based peptide backbone as scaffolds for helical peptide mimics. *J. Pept. Res.* **2002**, *60*, 88–94.
28. Tsuji, G.; Misawa, T.; Doi, M.; Demizu, Y. Extent of helical induction caused by introducing  $\alpha$ -aminoisobutyric acid into an oligovaline sequence. *ACS Omega* **2018**, *3*, 6395–6399.
29. Mondal, S.; Adler-Abramovich, L.; Lampel, A.; Bram, Y.; Lipstman, S.; Gazit, E. Formation of functional super-helical assemblies by constrained single heptad repeat. *Nat. Commun.* **2015**, *6*, 8615.
30. Wada, S.-i.; Tsuda, H.; Okada, T.; Urata, H. Cellular uptake of Aib-containing amphipathic helix peptide. *Bioorganic. Med. Chem. Lett.* **2011**, *21*, 5688–5691.
31. Li, H.; Anuwongcharoen, N.; Malik, A.A.; Prachayasittikul, V.; Wikberg, J.E.; Nantasenamat, C. Roles of D-amino acids on the bioactivity of host defense peptides. *Int. J. Mol. Sci.* **2016**, *17*, 1023.
32. Carmona, G.; Rodriguez, A.; Juarez, D.; Corzo, G.; Villegas, E. Improved protease stability of the antimicrobial peptide Pin2 substituted with D-amino acids. *Protein J.* **2013**, *32*, 456–466.
33. Tugyi, R.; Uray, K.; Ivan, D.; Fellingner, E.; Perkins, A.; Hudecz, F. Partial D-amino acid substitution: Improved enzymatic stability and preserved Ab recognition of a MUC2 epitope peptide. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 413–418.
34. De Zotti, M.; Biondi, B.; Crisma, M.; Hjorringgaard, C.U.; Berg, A.; Bruckner, H.; Toniolo, C. Isovaline in naturally occurring peptides: A nondestructive methodology for configurational assignment. *Biopolymers* **2012**, *98*, 36–49.
35. Formaggio, F.; Crisma, M.; Bonora, G.M.; Pantano, M.; Valle, G.; Toniolo, C.; Aubry, A.; Bayeul, D.; Kamphuis, J. (R)-isovaline homo-peptides adopt the left-handed 3(10)-helical structure. *Pept. Res.* **1995**, *8*, 6–15.
36. Shenkarev, Z.O.; Paramonov, A.S.; Lyukmanova, E.N.; Gizatullina, A.K.; Zhuravleva, A.V.; Tagaev, A.A.; Yakimenko, Z.A.; Telezhinskaya, I.N.; Kirpichnikov, M.P.; Ovchinnikova, T.V. Peptaibol antimioebin I: Spatial structure, backbone dynamics, interaction with bicelles and lipid-protein nanodiscs, and pore formation in context of barrel-stave model. *Chem. Biodivers.* **2013**, *10*, 838–863.
37. Brown, A.M.; Zondlo, N.J. A propensity scale for type II polyproline helices (PPII): Aromatic amino acids in proline-rich sequences strongly disfavor PPII due to proline-aromatic interactions. *Biochemistry* **2012**, *51*, 5041–5051.
38. Rubio-Martinez, J.; Tomas, M.S.; Perez, J.J. Effect of the solvent on the conformational behavior of the alanine dipeptide deduced from MD simulations. *J. Mol. Graph. Model.* **2017**, *78*, 118–128.
39. Maisuradze, G.G.; Liwo, A.; Scheraga, H.A. Principal component analysis for protein folding dynamics. *J. Mol. Biol.* **2009**, *385*, 312–329.
40. Sittel, F.; Jain, A.; Stock, G. Principal component analysis of molecular dynamics: On the use of Cartesian vs. internal coordinates. *J. Chem. Phys.* **2014**, *141*, 014111.
41. Altis, A.; Nguyen, P.H.; Hegger, R.; Stock, G. Dihedral angle principal component analysis of molecular dynamics simulations. *J. Chem. Phys.* **2007**, *126*, 244111.
42. 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.
43. Galindo-Murillo, R.; Roe, D.R.; Cheatham III, T.E. Convergence and reproducibility in molecular dynamics simulations of the DNA duplex d (GCACGAACGAACGAACGC). *Biochim. Biophys. Acta (BBA)* **2015**, *1850*, 1041–1058.
44. Roe, D.R.; Bergonzo, C.; Cheatham, T.E. 3rd, Evaluation of enhanced sampling provided by accelerated molecular dynamics with Hamiltonian replica exchange methods. *J. Phys. Chem. B* **2014**, *118*, 3543–3552.
45. Koukos, P.I.; Glykos, N.M. On the application of Good-Turing statistics to quantify convergence of biomolecular simulations. *J. Chem. Inf. Model.* **2014**, *54*, 209–217.
46. Serafeim, A.P.; Salamanos, G.; Patapati, K.K.; Glykos, N.M. Sensitivity of folding molecular dynamics simulations to even minor force field changes. *J. Chem. Inf. Model.* **2016**, *56*, 2035–2041.
47. Shao, J.; Tanner, S.W.; Thompson, N.; Cheatham, T.E. Clustering molecular dynamics trajectories: 1. characterizing the performance of different clustering algorithms. *J. Chem. Theory Comput.* **2007**, *3*, 2312–2334.
48. Levy, Y.; Jortner, J.; Becker, O.M. Solvent effects on the energy landscapes and folding kinetics of polyalanine. *Proc. Natl. Acad. Sci.* **2001**, *98*, 2188–2193.
49. Blin, K.; Medema, M.H.; Kottmann, R.; Lee, S.Y.; Weber, T. The antiSMASH database, a comprehensive database of microbial secondary metabolite biosynthetic gene clusters. *Nucleic Acids Res.* **2017**, *45*, D555–D559.
50. Vanquelef, E.; Simon, S.; Marquant, G.; Garcia, E.; Klimerak, G.; Delepine, J.C.; Cieplak, P.; Dupradeau, F.Y. R.E.D. Server: A web service for deriving RESP and ESP charges and building force field libraries for new molecules and molecular fragments. *Nucleic Acids Res.* **2011**, *39*, W511–W517.

51. Dupradeau, F.Y.; Pigache, A.; Zaffran, T.; Savineau, C.; Lelong, R.; Grivel, N.; Lelong, D.; Rosanski, W.; Cieplak, P. The RED. tools: Advances in RESP and ESP charge derivation and force field library building. *Phys. Chem. Chem. Phys.* **2010**, *12*, 7821–7839.
52. Frisch, M.J.; Trucks, G.W.; Schlegel, H.B.; Scuseria, G.E.; Robb, M.A.; Cheeseman, J.R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G.A.; et al. *Gaussian 09*, Gaussian Inc.: Wallingford, CT, USA, 2009.
53. Hanwell, M.D.; Curtis, D.E.; Lonie, D.C.; Vandermeersch, T.; Zurek, E.; Hutchison, G.R. Avogadro: An advanced semantic chemical editor, visualization, and analysis platform. *J. Cheminform.* **2012**, *4*, 17.
54. Cieplak, P.; Cornell, W.D.; Bayly, C.; Kollman, P.A. Application of the multimolecule and multiconformational RESP methodology to biopolymers: Charge derivation for DNA, RNA, and proteins. *J. Comput. Chem.* **1995**, *16*, 1357–1377.
55. Case, D.A.; Ben-Shalom, I.Y.; Brozell, S.R.; Cerutti, D.S.C. III, T.E.; Cruzeiro, V.W.D.; Darden, T.A.; Duke, R.E.; Ghoreishi, D.; Gilson, M.K.; et al. *AMBER 18*; University of California: San Francisco, CA, USA, 2018.
56. Wereszczynski, J.; McCammon, J.A. Using selectively applied accelerated molecular dynamics to enhance free energy calculations. *J. Chem. Theory Comput.* **2010**, *6*, 3285–3292.
57. Essmann, U.; Perera, L.; Berkowitz, M.L.; Darden, T.; Lee, H.; Pedersen, L.G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577–8593.
58. Roe, D.R.; Cheatham, T.E. 3rd, PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.* **2013**, *9*, 3084–3095.
59. Glykos, N.M. Software news and updates. Carma: A molecular dynamics analysis program. *J. Comput. Chem.* **2006**, *27*, 1765–1768.
60. Koukos, P.I.; Glykos, N.M. Grcarma: A fully automated task-oriented interface for the analysis of molecular dynamics trajectories. *J. Comput. Chem.* **2013**, *34*, 2310–2312.
61. Miao, Y.; Sinko, W.; Pierce, L.; Bucher, D.; Walker, R.C.; McCammon, J.A. Improved reweighting of accelerated molecular dynamics simulations for free energy calculation. *J. Chem. Theory Comput.* **2014**, *10*, 2677–2689.
62. Jing, Z.; Sun, H. A comment on the reweighting method for accelerated molecular dynamics simulations. *J. Chem. Theory Comput.* **2015**, *11*, 2395–2397.

**Sample Availability:** The producer of tripleurins, strain T. pleuroti SZMC 12454 is available from the Szeged Microbiology Collection ([www.szmc.hu](http://www.szmc.hu)).



© 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/>).